# Distribution theory

J. B. Cooper
Johannes Kepler Universität Linz

## Contents

**1 The elementary theory** **2**
1.1 Introduction. . . . 2
1.2 Distributions on a compact interval . . . . 7
1.3 Operations on distributions . . . . 16
1.4 Recolleament des morceaux . . . . 17
1.5 Division of distributions . . . . 21
1.6 Distributions on open sets . . . . 24
1.7 Limits and integrals of distributions . . . . 27
1.8 Continuity at a point . . . . 28
1.9 Primitives and definite integrals . . . . 28
1.10 The order of growth of distributions . . . . 29
1.11 Distributions as functionals . . . . 31
1.12 Fourier series and distributions . . . . 32

**2 The Schwartzian theory** **35**
2.1 Unbounded operators and spaces of distributions . . . . 39

# 1 The elementary theory

## 1.1 Introduction.

The need for a theory of distributions arose from the inadequacy of certain methods of classical analysis with regard to some applications. Thus a number of procedures which were being employed with success by physicists could not be justified rigorously within the framework of existing theories. The most striking example, and one that we shall use as the guideline in our construction of distributions, is the differentiation of non-continuous functions—a famous example being the Dirac delta function which is the "derivative" of the Heaviside function (figure 1).

In fact, the theory of distributions can also cope with other procedures such as the changing of order in multiple integrals in situations where this is not justified by classical results and the use of divergent series. In addition, it allows a dramatic extension of the range of applicability of such methods as the Fourier transform in the solution of differential equations. Distributions have also paved the way for advances in pure mathematics, for example in the theory of existence and uniqueness of solutions of partial differential equations.

The origins of the theory of distributions can be traced back at least as far as the operational calculus of Heaviside. Perhaps the first step towards a modern mathematical presentation was made in the 1930's by Sobolev and Friedrichs. Dismayed by the fact that the wave equation

$$\frac{\partial^2 f}{\partial x^2} = \frac{\partial^2 f}{\partial t^2}$$

can only have solutions which are twice differentiable whereas physics require that all functions of the form

$$f(x,t) = u(x+t) + v(x-t)$$

are solutions, regardless of smoothness, they introduced an idea of a generalised solution of this equation which allowed non-differentiable functions as solutions.

The foundations of the theory of distributions proper is generally ascribed to L. Schwartz, who used the duality theory for locally convex spaces, at that time a recent generalisation of the theory of Banach spaces, to define distributions as functionals on suitable test spaces of smooth functions. Thus this definition generalises the notion of a measure (in the Bourbakian sense) rather than that of a function. It was soon realised that conceptually and technically simpler approaches to the theory were possible. For example, Mikusinski and Temple defined distributions as generalised limits of sequences of smooth functions i.e. used a topological completion process reminiscent of that used by Cauchy to define real numbers. On the other hand, H. König and J. Sebastião e Silva, independently returning to the work of Sobolev and Friedrichs, defined distributions as generalised derivatives of non-differentiable functions. In this approach, the role of distributions as generalised functions (rather than as functionals) is apparent.

We shall bring a slight variant of the method of Sebastião e Silva which lends itself to generalisations which will be exploited in later chapters to bring a unified approach to various spaces of distributions.

EXAMPLES. We begin with some simple examples which display the necessity of an extension of the classical concepts:

Let a particle of mass 1 move on the $x$-axis in such a way that its distance $x$ from the origin is a function of the time $t$. Consider the special motion represented by the graph of $x$ and the velocity $v$ as functions of $t$ as depicted in figure 2. The force $F$ on the particle (i.e. an instantaneous impulse at time $t = 0$) is given by the formula $F = \frac{dv}{dt}$. Hence we are faced with the problem of differentiating a non-continuous function. If we attempt to do this in the classical sense, then we see that the function $H$ which represents the velocity is differentiable at each point, with the exception of 0, and the derivative is zero. This is an inadequate solution in that the fundamental theorem of calculus fails — we cannot recover the function from its derivative. Despite this fact we shall proceed formally on the assumption that $H$ *has* a derivative (which we shall denote by $\delta$) and we shall try to "derive" some of its properties by formal manipulations. Suppose that $f$ is a smooth, real-valued function which vanishes at the points 1 and $-1$. Then, by integration by parts,

$$\begin{aligned}\int_{-1}^{1} f(t)\delta(t)\,dt &= \int_{-1}^{1} f(t)H'(t)\,dt \\ &= H(t)f(t)|_{-1}^{1} - \int_{-1}^{1} H(t)f'(t)\,dt \\ &= -\int_{0}^{1} f'(t)\,dt = f(0).\end{aligned}$$

This leads to the "definition" of the delta function as one which is zero everywhere except at the origin where it is infinitely large, this in such a fashion that we have the formula

$$\int_{\mathbf{R}} f(t)\delta(t)\,dt = f(0)$$

for suitable functions $f$.

We consider the one-dimensional wave equation

$$\frac{\partial^2 f}{\partial x^2} = \frac{\partial^2 f}{\partial t^2}$$

which describes the displacement $f$ of a particle at position $x$ on a string at time $t$. This equation is satisfied, for example, by a fictitious infinite, homogeneous, stretched string. By introducing new independent variable $x+t$ and $x-t$, the general solution can be calculated to have the form

$$f(x,t) = u(x+t) + v(x-t)$$

where $u$ and $v$ are twice differentiable functions of one variable. Geometrically, the first term (involving $u$) represents a plane wave travelling leftwards with uniform velocity (see figure 3). The general solution is thus the superposition of two such waves, moving with uniform velocity in opposite directions. Of course, this description of the solution is purely geometric and in no way involves the smoothness of the wave functions. Thus the wave depicted in figure 4 has the correct geometrical form for a solution but cannot be regarded as one within the classical theory since the expression $\dfrac{\partial^2 f}{\partial x^2}$ is meaningless. Of course, such solutions are physically relevant.

Consider once again the vibrating string, this time a finite one with fixed endpoints. Its motion is governed by the same differential equation, this time for an $f$ defined on the region $[0,\phi] \times [0,\infty[$. In addition, $f$ is subject to the boundary conditions

$$0 = f(0,t) = f(\phi,t) \text{ for each } t.$$

We treat the problem where the string lies along the $x$-axis at time $t=0$ and the particle at point $x$ has initial velocity $g(x)$ i.e. we have the initial conditions

$$f(x,0) = 0 \qquad \frac{\partial f}{\partial t}(x,0) = g(x) \quad (x \in [0,\pi]).$$

(See figure 5). A typical solution of this equation, without the initial condition, is

$$f(x,t) = \sin nx \sin nt$$

($n$ a positive integer). More generally, we can superpose such solutions to obtain ones of the form

$$f(x,t) = \sum_n b_n \sin nx \sin nt$$

(where the $b_n$'s are real numbers). The summation can be infinite—in this discussion we shall ignore questions of convergence. Hence if we expand the function $g$ in a Fourier sine series

$$g(x) = \sum_n a_n \sin nx$$

then we can calculate the coefficients $(b_n)$ by differentiating and comparing coefficients.

Now consider the case where the force is a sharp impulse at the point $x = \frac{\pi}{2}$. As an approximation, we can treat the case where $g$ is the characteristic function of an interval of length $\epsilon$ around $\frac{\pi}{2}$, normalised to have integral one (i.e. with height $\frac{1}{\epsilon}$) (see figure 6). This function has Fourier sine series

$$2\sum_{n=1}^{\infty} \frac{\sin n\pi\epsilon}{n\pi\epsilon} \sin 2nx.$$

In the limiting case of a sharp impulse, represented by a $\delta$-function, we get the equation

$$\delta(x - \frac{\pi}{2}) = 2\sum_{n=1}^{\infty} \sin 2nx$$

which provides the correct solution

$$f(x,t) = \sum_{n=1}^{\infty} \frac{1}{n} \sin 2nx \cos 2nt.$$

Notice the peculiar form of the equation for the Fourier expansion of the delta-function. On the left hand side, we have a function which doesn't exist in any classical sense and on the right hand side a series which does not converge. It is one of the aims of distribution theory to give such equations a meaning. We note in passing that we could have obtained the same equation by formally differentiating the Fourier series of the function pictured in figure 7.

We now consider the equation of the simple pendulum with external force $f$ i.e. the differential equation $x'' + x = f$. The special case $x'' + x = \delta(t-u)$ corresponds to a sudden impulse at time $t = u$. Intuitively it is clear that the solution in this case will be $x = 0$ until $t = u$. Thereafter, the pendulum will swing according to the initial conditions $x = 0$, $x' = 1$ (at $t = u$) i.e. its equation will be

$$x(t) = \begin{cases} 0 \quad (t \leq u) \\ \sin(t-u) \quad (t \geq u). \end{cases}$$

We denote this solution by $G(t,u)$ i.e. $G(t,u)$ is the displacement at time $t$ induced by an impulse at time $u$. Consider now the function

$$x(t) = \int G(t,u) f(u)\, du.$$

Manipulating formally once again, we have

$$\begin{aligned} x''(t) + x(t) &= \int [\frac{\partial^2 G}{\partial t^2} + G] f(t)\, dt \\ &= \int \delta(t-u) f(u)\, dt = f(t). \end{aligned}$$

Hence if we can solve the special case of the equation for an impulsive force, we can solve the general problem by approximating a continuous force by a series of impulses and superposing the resulting motions.

We consider again the differential equation $x'' + x = f$. A standard method for solving such equations is to introduce the Fourier transforms $X$ and $F$ defined by the equations

$$X(u) = \int x(t) e^{-itu}\, dt \qquad F(u) = \int f(t) e^{-itu}\, dt.$$

Formal manipulation gives the following equation for $X$:

$$X(u)(1-u^2) = F(u)$$

which no longer involves differentiation. Its solution is

$$X(u) = \frac{F(u)}{1-u^2}$$

and one recovers the solution $x$ of the original differential equation by applying the inverse Fourier transform to $X$. However, even in the simplest case where $f$ is the zero function, we will encounter the difficulty that the only $X$ which satisfies the derived equation is the zero function. This gives only the trivial solution of the differential equation and ignores the interesting ones which correspond to simple harmonic motion. However, one might conjecture that a distribution of the form $a\delta(u-1)+b\delta(u+1)$ (for constants $a$ and $b$), being zero everywhere except at 1 and $-1$, could be a solution of the algebraic equation. This is indeed the case and supplies the missing solutions.

## 1.2 Distributions on a compact interval

We begin with a definition which is suggested by the properties which one would expect of a space of distributions.

**Definition 1** *A* **space of distributions** *on the compact interval $I = [a, b]$ is a vector space $E$, together with a linear injection $i$ from $C(I)$ into $E$ and a linear mapping $\tilde{D}$ on $E$ so that*

*a) if $x \in C_1(I)$, then $\tilde{D}(ix) = iD(x)$;*

*b) if $y \in E$, there is an $x \in C(I)$ and a positive integer $p$ so that $y = \tilde{D}^p(ix)$;*

*c) if $y \in E$ and $\tilde{D}^p y = 0$, then there is an $x$ in $P_p(I)$ so that $y = ix$.*

The first condition says that functions are distributions (in other words, that distributions are generalised functions), the second that each distribution is a (repeated) derivative of some function. The third condition ensures that if differential equations have classical solutions, then we do not introduce any extra, unwanted distributional solutions.

In the above definition, we used the indefinite article in connection with spaces of distributions. This was because it is not yet clear either that such a space exists or that it is unique in a suitable sense. In fact, both of these conditions are satisfied i.e. there is precisely one such space - which we are then entitled to call *the* space of distributions on $I$.

PROOF. We begin with the uniqueness. Suppose that we have two distribution spaces $E$ and $E'$, with corresponding inclusions $i$ and $i'$ resp. differentiation operators $\tilde{D}$ and $\tilde{D}'$. Then we claim that there is a vector space isomorphism $U$ from $E$ onto $E'$ so that

a) for each $x$ in $C(I)$, $U(ix) = i'(x)$;

b) if $y$ is in $E$, then $U\tilde{D}(y) = \tilde{D}'(y)$.

This clearly means that $E$ and $E'$ are essentially the same spaces. In order to construct the above operator $U$, we consider a distribution $y$ in $E$. $y$ has the form $\tilde{D}^p(ix)$ for some continuous function $x$ on $I$. We simply define $Uy$ to be $(\tilde{D}')^p(i'(x))$. The reader will have no difficulty in checking that $U$ is well-defined and satisfies the above condition.

We now turn to the proof of the existence. Recall the sequence

$$C^\infty \subset \cdots \subset C^{n+1} \subset C^n \subset \cdots \subset C$$

of function spaces. We observe that $D$ is a continuous linear mapping from $C^{n+1}$ into $C^n$ for each $n$ (indeed the latter is a quotient space of the former with respect to this mapping). We will extend this scheme to the right to one of the form

$$C^\infty \subset \dots C^n \subset \dots \subset C \subset \dots C^{-n} \subset C^{-n-1} \subset \dots$$

where $C^{-n}$ will consist of those distributions which are the $n$-th derivatives of continuous functions.

We begin with some informal remarks before giving the precise construction. In order to construct the first space $C^{-1}$, we begin with the set of ordered pairs $(x, y)$ of continuous functions on $I$. Here a function is to be regarded as representing itself if it is in the first position. On the other hand, if it is in the second place, it is to represent a distribution, namely its derivative, whatever that may be. It is clear, however, that such a representation will not be unique. Thus on the unit interval, the pairs, $(t, 0), (0, \frac{t^2}{2})$ and $(\dfrac{t}{2}, \dfrac{t^2}{4})$ represent the same distribution. In order to eliminate this difficulty, we would like to say that the pair $(x, y)$ represents the zero distribution if $x + Dy = 0$. However, the fact that the operator $D$ is not everywhere defined on $C(I)$ leads to difficulties. For this reason it is better to integrate the above equation to obtain the condition $Ix + y = \ oxconstant$. (Recall that $I$ is a suitable integration operator). This leads to the definition of $C^{-1}$ as the quotient of the product space $C(I) \times C(I)/F_1$ where

$$F_1 = \{(x, y) : Ix + y \in P_1(n)\}.$$

(This definition is independent of the particular choice of base point for the integral operator).

This definition can be extended in the obvious way to obtain $C^{-2}$ as the quotient space $C(I) \times C(I) \times C(I)/F_2$ where

$$F_2 = \{(x, y, z) : I^2x + Iy + z \in P_2(I)\}.$$

For example, the delta-distribution on the interval $I = [-1, 1]$ could be represented by the triple $(0, 0, y)$ where $y$ is the function

$$t \mapsto \begin{cases} 0 & (t < 0) \\ t & (t \geq 0). \end{cases}$$

It should now be clear how the construction will proceed in the in the general case and we pass over to the formal details. We write $H_n$ for the $(n+1)$-fold product

$$C(I) \times C(I) \times \dots \times C(I)$$

whereby a typical element is written as $(x_0, \dots, x_n)$. $F_n$ is defined to be the subspace

$$\{(x_0, \dots, x_n) : I^n x_0 + I^{n-1} x_1 + \dots + x_n \in P_n(I)\}.$$

We remark in passing that $H_n$ is a Banach space and $F_n$ is a closed subspace (since it is the inverse image of the closed subspace $P_n(I)$ under a continuous linear mapping). Hence the space $C^{-n}(I)$, which is defined to be the quotient of $H_n$ by $F_n$, is a Banach space (and hence an LVS). Of course, $F_n$ (and so also $C^{-n}(I)$) is independent of the particular choice of integration operator. The proof of the existence of a space of distributions on $I$ now consists of the verification of a series of simple facts.

1. The natural injection

   $$(x_0, \dots, x_n) \mapsto (x_0, \dots, x_n, 0)$$

   from $H_n$ into $H_{n+1}$ induces a continuous injection from $C^{-n}(I)$ into $C^{-n-1}(I)$. This follows from the fact that the sequence $(x_1, \dots, x_n)$ is in $F_n$ if and only if $(x_0, \dots, x_n, 0)$ is in $F_{n+1}$. For the first condition means that $I^n x_0 + \dots + x_n$ is in $P^n(I)$ whereas the second is equivalent to the fact that $I^{n+1} x_0 l + \dots I x_n$ is in $P^{n+1}(I)$. It is clear that these conditions are equivalent.

2. The mapping

   $$(x_0, \dots, x_n) \mapsto (0, x_0, \dots, x_n)$$

   from $H_n$ into $H_{n+1}$ lifts to a continuous linear mapping $D_n$ from $C^{-n}(I)$ into $C^{-n-1}(I)$. Here it suffices to show that if the first term in $F_n$, then the second is in $F_{n+1}$. This is completely trivial.

By 1. above, we can regard $C^{-n}$ as a (vector) subspace of $C^{-n-1}$ i.e. we have the promised chain

$$C \subset C^{-1} \subset \dots \subset C^{-n} \subset C^{-n-1} \subset \dots$$

We define $C^{-\infty}$ to be the union of these spaces. We regard it as an LVS as the union of a sequence of Banach spaces. Thus a sequence is convergent in $C^{-\infty}(I)$ if and only it is contained in some $C^{-n}(I)$ and converges there. More precisely, this means:

On $C^{-\infty}$ we define an operator $D$ as follows: if $x$ is an element of the latter space, then it is in $C^{-n}$ for some $n$. We then define $Dx$ to be $D_n(x)$. That this is well-defined follows from the fact that the following diagram commutes.

Our claim is that $C^{-\infty}(I)$ with this operator is a (and therefore *the*) space of distributions on $I$. In order to do this, we must verify that the conditions a) - c) in the definition hold. This we proceed to do.

a) we must show that if $x$ is in $C^1(I)$, then the two elements $(0,x)$ and $(Dx,0)$ are equivalent i.e. that their difference $(Dx,-x)$ is in $F_1$. This means that $IDx - x$ should be in $P_1(I)$ which is clearly the case.

b) Consider an element $\pi_n(x_0,\dots,x_n)$ of $C^{-n}(I)$. Since

$$\pi_n(x_0,\dots,x_n) = \pi_n(0,\dots,I^n x_0 + \dots x_n),$$

the original distribution is $D^n X$ where

$$X = I^n x_0 + \dots + x_n.$$

c) Suppose that $y$ is in $C^{-\infty}$ with $D^p y = 0$. If $y$ is say in $C^{-n}(I)$, then the previous manipulation shows that it can be written in the form $\pi_n(0,\dots,0,X)$ with $X$ continuous. Then

$$D^p y = \pi_{n+p}(0,\dots,0,X)$$

and this is zero exactly when $X$ is in $P^{n+p}(I)$. In this case, $y = D^n X = D^n X \in P^p(I)$.

■

In light of the above facts, we shall call $C^{-\infty}$ **the space of distributions** on $I$. $C^{-n}(I)$ is called **the space of distributions of order $n$.** It consists of those distributions which are the $n-$th derivatives of continuous functions. In view of axiom a) we shall from now on drop the tilda on the $D$ and denote the derivative of a distribution $y$ simply by $Dy$ except in certain situations where it is important to distinguish between pointwise derivatives and distributional derivatives.

Examples. (of distributions)

I. Integrable functions: in our axiom system we demanded that continuous functions be distributions. In fact, the space of distributions contains most functions of any practical significance. In particular, we can regard Lebesgue integrable functions as distributions. In this case, however, we must distinguish between the pointwise derivative and the distributional derivative. Examples of this will be considered in more detail later. In order to regard such functions as distributions, we proceed as follows: we let $x$ be an (equivalence class of) a Lebesgue integrable function i.e. $x \in L^1(I)$. Then $x$ has an absolutely continuous primitive i.e. a function $X$ which is such that $x$ is equal almost everywhere to the derivative of $X$. Further, any two such functions differ by a constant. We define the distribution $T_x$ to be the distributional derivative $DX$ of $X$. The mapping $x \mapsto T_x$ is continuous and linear from $L^1(I)$ into $C^{-1}(I)$. It follows from elementary facts about the Lebesgue integral, that the mapping is injective.

Unless there is a particular reason for being more careful, we shall not normally differentiate between $x$ and $T_x$. Thus a statement such as "the distribution $z$ is an integrable function" means that there exists an $x$ in $L^1(I)$ so that $y = T_x$.

II. Measures: in a similar manner, we can regard Radon measures on $I$ as distributions. If $\mu$ is such a measure, then the function

$$x : s \to \mu(]-\infty, s] \cap I)$$

is of bounded variation on $I$ and so is integrable. Hence we can identify $\mu$ with the distribution $\tilde{D}(T_x)$. The use of the word "identify" is justified by the fact that the mapping which associates to $\mu$ the above distribution is a linear injection from the space of Radon measures on $I$ into $C^{-\infty}(I)$.

Examples. If $a$ is a point in the interior of $I$, then we denote by $H_a$ the **Heaviside function** with singularity at $a$ i.e. the function

$$s \mapsto \begin{cases} 0 & \text{if } s < a \\ 1 & \text{if } s \geq a. \end{cases}$$

(in other words, $H_a = \chi_{I \cap [a,\infty[}$).

Of course, $H_a$ is integrable and so can be regarded as a distribution on $I$. Its derivative $\tilde{D}H_a$, which is denoted by $\delta_a$, is the **Dirac distribution** with singularity at $a$. It is, in fact, a measure (which assigns to each set value 1 or 0 according as $a$ is or is not a member of the set). Its derivatives $\delta_a^{(n)}$ are then the $(n+1)$-th derivatives of $H_a$.

The functions $s^\lambda$, $s_-^\lambda$, $|s|^\lambda$. We suppose now that 0 is an interior point of the interval $I$. Consider the following functions (whereby $\lambda$ is a complex exponent):

$$s_+^\lambda : s \mapsto \begin{cases} 0 & \text{if } s < 0 \\ s^\lambda & \text{if } s \geq 0. \end{cases}$$

$$s_-^\lambda : s \mapsto \begin{cases} (-s)^\lambda & \text{if } s < 0 \\ 0 & \text{if } s \geq 0. \end{cases}$$

$$|s|^\lambda : s \mapsto \begin{cases} |s|^\lambda & \text{if } s \neq 0 \\ 0 & \text{if } s = 0. \end{cases}$$

(Of course, $|s|^\lambda = s_+^\lambda + s_-^\lambda$). If $\mathrm{Re}\lambda > -1$, then these functions are integrable and so represent distributions. If, on the other hand, the real part is less than or equal to $-1$, then they are well-defined, but not integrable. Hence they do not determine distributions in the above sense. Nevertheless we can regard them as distributions as follows. We consider firstly the case where the real part of $\lambda$, in addition to being at most $-1$, is not a negative integer. Then we choose a positive integer $m$ so that $\mathrm{Re}(\lambda + m) > 0$ and define the distributions of the title as follows:

$$\begin{aligned} s_+^\lambda &= \frac{D^m s_+^{\lambda+m}}{(\lambda+m)\dots(\lambda+1)} \\ s_-^\lambda &= \frac{(-1)^m D^m s_-^{\lambda+m}}{(\lambda+m)\dots(\lambda+1)} \\ |s|^\lambda &= s_+^\lambda + s_-^\lambda. \end{aligned}$$

Note that these do not depend on the choice $m$.

In order to be able to handle the case where the real part of $\lambda$ is a negative integer, we use the logarithm function. Define functions

$$\ln_+ : s \mapsto \begin{cases} \ln s & \text{if } s > 0 \\ 0 & \text{if } s \leq 0. \end{cases}$$

$$\ln_- : s \mapsto \begin{cases} 0 & \text{if } s > 0 \\ \ln(-s) & \text{if } s \leq 0. \end{cases}$$

$$\ln_{|\,|} : s \mapsto \begin{cases} \ln |s| & \text{if } s \neq 0 \\ 0 & \text{if } s = 0. \end{cases}$$

Since these functions are integrable, they define distributions on $I$. We can now define the distributions $s_+^{-m}$, $s^{-m}$ and $|s|^{-m}$ ($m$ a positive integer) as follows:

$$\begin{aligned} s_+^{-m} &= \frac{(-1)^m}{(m-1)!} D^m \ln_+ \\ s_-^{-m} &= -\frac{1}{(m-1)!} D^m \ln_- \\ |s|^{-m} &= s_+^{-m} + s_-^{-m}. \end{aligned}$$

Thus we have defined the distributions above for all complex values of the parameter $\lambda$. It follows immediately from these definitions that we always have the relationship

$$Ds_+^\lambda = \lambda s_+^{\lambda-1} \qquad (\lambda \in \mathbf{C}).$$

**The derivative of a piecewise smooth function** Consider the function $s_+^n$ where $n$ is a positive integer. The $n$-th derivative of this function is the Heaviside function $H_0$ (times the constant $n!$). Hence we have the equation

$$\tilde{D}^m s_+^n = (n!)\delta_0^{(m-n)}$$

for $m > n$. Hence, if $p$ is a polynomial of degree $n-1$, say

$$p : s \mapsto a_0 + a_1 s + \cdots + a_{n-1}s^{n-1},$$

and $p_+ = p.H_0$, then

$$\begin{aligned} \tilde{D}^n p_+ &= a_0\delta_0^{(n-1)} + a_1\delta_0^{(n-2)} + \cdots + (n-1)!a_{n-1}\delta_0 \\ &= p(0)\delta_0^{(n-1)} + p'(0)\delta_0^{(n-2)} + \cdots + p^{(n-1)}(0)\delta_0. \end{aligned}$$

Thus we see that jumps in a smooth function or its derivatives induce the appearance of $\delta$-type singularities in the derivatives. In particular, if a function is not continuous, then the almost everywhere pointwise derivative need not coincide with the distributional derivative. We shall use the above formula to derive corresponding ones for the distributional derivative of a piecewise $n$-times continuously differentiable function $x$ i.e. a function for which there is a finite sequence $(a_0, \dots, a_k)$ in the interior of $I$ so that

a) $x$ is $n$-times continuously differentiable on $I \backslash \{a_0, \dots, a_k\}$;

b) for each $i = 0, \dots, n$ and $j = 0, \dots, k$ $\lim_{s\to a_j+} x^{(i)}(s)$ and $\lim_{s\to a_j-} x^{(i)}(s)$ exist. Then we write

$$\sigma_j^i = \lim_{s\to a_j+} x^{(i)}(s) - \lim_{s\to a_j-} x^{(i)}(s).$$

Functions of the above form are integrable and so define distributions. We proceed to calculate the corresponding derivatives. In order to simplify the notation, we shall suppose that $x$ has only one point of singularity $a_0$. $x^{(i)}$ will denote the function which we obtain by differentiating $x$ $i$-times on the complement of $a_0$. We shall regard this as an $L^1$-function (and so as a distribution). Hence it will not be relevant that the function is not defined at $a_0$. Let $p$ be that polynomial of degree $n-1$ whose successive derivatives at $a_0$ are

$$p(a_0) = \sigma_0^0, \quad p'(a_0) = \sigma_0^1, \dots, p^{(n-1)}(a_0) = \sigma_0^{n-1}$$

and let $p_+$ be the function $H_{a_0}.p$. Then $x - p_+$ is $(n-1)$-times continuously differentiable on $I$ as the reader can easily verify. Now the distributional derivative of the latter function of order $n$ is $x^{(n)}$ since the latter is an integrable function which is the pointwise derivative of $x - p_+$ away from $a_0$. Hence

$$\begin{aligned} \tilde{D}^n x &= D^n(x - p_+) + D^n(p_+) \\ &= x^{(n)} + \sigma_0^0 \delta_{a_0}^{(n-1)} + \dots + \sigma_0^{(n-1)} \delta_{a_0}. \end{aligned}$$

The general expression for the derivative of the original $x$ (i.e. in the situation where we have $k$ singularities) is then

$$\tilde{D}^n x = x^{(n)} + \sum_{i,j} \sigma_i^j \delta_{a_i}^{(n-1-j)}.$$

Examples. For any positive $\epsilon$, the function

$$s \mapsto \ln(s + i\epsilon) = \ln|s + i\epsilon| + i \arg(s + i\epsilon)$$

is integrable on $I$. The pointwise limit as $\epsilon$ converges to zero is the function $\ln|s| + i\pi(1 - H_0(s))$ (for $s \neq 0$). Hence it is natural to define the distribution $\ln(s + i0)$ as

$$\ln|s| + i\pi(1 - H_0).$$

Similarly, we define distributions

$$\begin{aligned} \ln(s - i0) &= \ln|s| - i\pi(1 - H_0) \\ (s + i0)^{-1} &= D(\ln(s + i0)) = s^{-1} - i\pi\delta_0 \\ (s - i0)^{-1} &= D(\ln(s - i0)) = s^{-1} + i\pi\delta_0. \end{aligned}$$

(where $s^{-1}$ is the distribution

$$\tilde{D} \ln|s| = D \ln_+ s + D \ln_- s = s_+^{-1} - s_-^{-1}).$$

Now by the Lebesgue theorem on dominated convergence, $\ln(s + i\epsilon)$ converges to $\ln(s + i0)$ in the $L^1$-sense (and so in the distributional sense). Differentiating, we get:

$$(s + i0)^{-1} = \lim_{\epsilon \to 0_+} (s + i\epsilon)^{-1}.$$

Similarly,

$$(s - i0)^{-1} = \lim_{\epsilon \to 0_-} (s - i\epsilon)^{-1}.$$

Hence if we introduce the notation

$$\delta_0^+ = \frac{1}{2\pi i}(s - i0)^{-1} \qquad \delta_0^- = \frac{1}{2\pi i}(s + i0)^{-1}$$

we have the equation

$$\delta_0 = \delta_0^+ - \delta_0^-.$$

We remark that one can show that $(s + i0)^{-1}$ is the limit of the distributions $(s + i\epsilon)^{-1}$ without having recourse to Lebesgue's theorem by showing directly that the primitives of these functions converge in a suitable manner.

REMARK. For reasons which will become clear later, the distributions which we have here denoted by $s^{-m}$ ($m$ a positive integer) are often denoted by the symbol p.p.$\frac{1}{s^m}$ (for the principal part).

The methods used here allow one to regard any function on $I$ which is smooth except for a finite number of isolated poles—typically the restriction of a meromorphic function to $I$—as a distribution.

We conclude this list of examples by showing how to derive correctly the formula for the Fourier expansion of the delta-function referred to ion the Introduction. We shall work in the space $C^{-\infty}(I)$ where $I$ is the interval $[-\pi, \pi]$. Consider firstly the function $2H_0 - 1$ which is odd in the sense that its graph is point-symmetric about the origin. The latter has the Fourier expansion

$$\frac{1}{\pi} \sum_{n=1}^{\infty} \frac{1}{n} \sin(2nt)$$

as can easily calculated by elementary methods. This converges in the norm-topology of $L^1$ and so, *a fortiori*, in the distributional sense. Hence we can apply the continuous operator $\tilde{D}$ to both sides to get the desired equation:

$$\delta_0 = \frac{1}{\pi} \sum_{n=0}^{\infty} \cos(2nt).$$

More generally, we have the equations

$$\begin{aligned} \delta_0^{(k)} &= \frac{1}{\pi} \sum_{n=0}^{\infty} (2n)^k (-1)^{k/2} \cos 2nt \qquad (k \text{ even}) \\ \delta_0^{(k)} &= \frac{1}{\pi} \sum_{n=1}^{\infty} (2n)^k (-1)^{(k-1)/2} \sin 2nt \qquad (k \text{ odd}. \end{aligned}$$

As a second example, consider the function $x(t) = H(t)\sin t$. Using methods similar to those used above, once can show that $x$ is a solution of the differential equation $x'' + x = \delta_0$, more precisely

$$\tilde{D}^2 x + x = \delta_0$$

as claimed in the introduction.

## 1.3 Operations on distributions

We shall now show how to extend some standard operation on functions to distributions. The main problem is that such operations (for example, that of multiplication) can no longer be defined pointwise. This means that we are forced to seek less direct methods. One of the simplest is the content of the following Lemma:

**Lemma 1** *Let $S$ be a linear operator from $C(I)$ into $C(J)$ (where $I$ and $J$ are compact intervals) which commutes with $D$ i.e. are such that if $x$ is an element of $C^1(I)$, then $Sx \in C^1(J)$ and $\tilde{D}(Sx) = S(\tilde{D}x)$. Then $S$ possesses a unique extension $\tilde{S}$ to an operator from $C^{-\infty}(I)$ into $C^{-\infty}(J)$ which commutes with $\tilde{D}$. Further if $S$ is continuous, then so is $\tilde{S}$.*

Proof. Suppose firstly that a function $x$ in $C(I)$ is constant. Then $Dx = 0$ and so $\tilde{D}(Sx) = S(Dx) = 0$ i.e. $Sx$ is constant. Similarly, it follows from the equation $\tilde{D}^n Sx = S\tilde{D}^n x$ $(x \in C^n(I))$ which is a consequence of the commutativity; condition, that if $x$ is in $P^n(I)$, then $Sx$ is in $P^n(J)$. Also, applying the commutativity relation to $IX$ $(x \in C(I))$ and integrating, we get the equation

$$I(SDIx) = (ID(SIx)$$

and so

$$ISx = SIx + x_0$$

where $x_0$ is a constant. Similarly, for each $n$ in $\mathbf{N}$ and $x$ in $C(I)$,

$$I^n Sx - SI^n x \in P^n(J).$$

Consider now the mapping

$$S_n : (x_0, \dots, x_n) \mapsto (Sx_o, \dots, Sx_n)$$

from $H_n(I)$ into $H_n(J)$. Then $S_n(F_n(I)) \subset F_n(J)$. For if $(x_0, \dots, x_n)$ is in $F_n(I)$, then

$$I^n x_0 + \dots + x_n \in P^n(I)$$

and so

$$I^n Sx_0 + \cdots + Sx_n = S(I^n x_0) + \ldots Sx_n + p = S(I^n x_0 + \ldots x_n) + p$$

where $p \in P^n(J)$. Hence we can lift $S_n$ to a linear operator $\tilde{S}_n$ from $C^{-n}(I)$ into $C^{-n}(J)$, the former being continuous if the latter is. These mappings are compatible and so combine to define the mapping from $C^{-\infty}(I)$ into $C^{-\infty}$ which we are looking for.

■

It can easily be seen from the proof that $\tilde{S}$ is surjective (resp. injective, resp. bijective) if $S$ is.

Note that, in terms of representations of distributions by finite sequences of continuous functions, the above proof shows that $S$ acts componentwise.

■

As examples of operators which satisfy the conditions of the above Lemma, we consider translation and restriction: **Translation:** We know that the operator

$$\tau_h : ax \mapsto (s \mapsto x(s-h))$$

is continuous and linear from $C(I)$ into $C(I+h)$ and, of course, commutes with differentiation. Hence we see immediately that there exists a unique continuous linear operator $\tilde{\tau}_h$ from $C^{-\infty}(I)$ into $C^{-\infty}(I+h)$ which coincides with the usual translation for continuous functions and commutes with the operation of differentiation. **Restriction:** If $I$ and $J$ are compact intervals with $J$ contained in $I$, the restriction operation $\rho_{I,J}$ from $C(I)$ into $C(J)$ satisfies the conditions of the Lemma and so there is a uniquely determined continuous linear operator on the corresponding spaces of distributions which coincides with the latter for continuous functions and commutes with differentiation.

We call the corresponding distributions $\tilde{\tau}_h x$ and $\tilde{\rho}_{I,J} x$ the **h-translate** of $x$ resp. the **restriction** of $x$ to $J$

## 1.4 Recolleament des morceaux

We now consider the problem of piecing together distributions which are defined on subintervals of a given interval. We begin with the case of two such intervals i.e. we have the situation of a compact interval $K$ which is the union $I \cup J$ of two such intervals which have non-degenerate intersection. Suppose that $x$ is a distribution in $I$, $y$ one on J and that they coincide on $I \cap J$ i.e. that

$$\rho_{I,I\cap J} = \rho_{J,I\cap J}(y).$$

**Lemma 2** *Let $I$ and $J$ be compact intervals with $J$ a subset of $I$. Suppose further that $(y_0, \dots, y_n)$ is a representation of a distribution of a distribution on $J$ and let $x$ be a distribution in $C^{-n}(I)$ which agrees with $y$ on $J$. Then $x$ has a representation $(x_0, \dots, x_n)$ whereby $x_i = y_i$ on $J$ for each $i$.*

PROOF. $x$ has a representation of the form

$$(\tilde{x}_0, \dots, \tilde{x}_n).$$

Since $x$ and $y$ coincide on $J$, there is a polynomial $p$ of degree at most $n-1$ so that

$$I^n(\tilde{x}_0 - y_0) + \cdots + (\tilde{x}_n - y_n) = p$$

on $J$. We now extend the functions $y_0, \dots, y_{n-1}$ in an arbitrary manner to continuous functions on $I$ which we denote by $(x_0, \dots, x_{n-1})$. $x_n$ is then determined by the equation

$$x_n = -p + I^n(\tilde{x}_0 - x_0) + \cdots + I(\tilde{x}_{n-1} - x_{n-1}) + \tilde{x}_n.$$

Then $(x_0, \dots, x_n)$ is the required representation.

**Lemma 3** *Suppose that $K = I \cup J$ is a representation of the compact interval $K$ as above. Then if $x$ is a distribution on $I$ resp. $y$ one on $J$ which agree on $I \cap J$,there is a distribution $z$ on $K$ which agrees with $x$ on $I$ and with $y$ on $J$.*

PROOF. Let $(z_0, \dots, z_n)$ be a representation of $x$ (and so of $y$) on $I \cap J$. By a double application of the above Lemma, we can find representations $(x_0, \dots, x_n)$ of $x$ and $(y_0, \dots, y_n)$of $y$ on $J$ so that each $x_i$ and $y_i$ coincide with $z_i$ and so with each other on the intersection. Let $z_i$ be the continuous function on $K$ which is equal to $x_i$ on $I$ and so $y_i$ on $J$. Then $(z_0, \dots, z_n)$ is the required distribution.

■

We remark that the distribution $z$ is uniquely determined by the above condition. For if $I, J, K$ are as above and $z$ resp. $z_1$ are distributions which agree on $I$ land on $J$, then they agree on $K$. For let $(z_0, \dots, z_n)$ be a common representation of the two distributions on $I$. By the penultimate Lemma, we can find representations

$$(x_0, \dots, x_n) \text{ and } (y_0, \dots, y_n)$$

for $z$ and $z_1$ on $K$ where the $x_i$ and $y_i$ are extensions of $z_i$. Since these distributions agree on $J$, the function

$$I^n(x_0 - y_0) + \cdots + (x_n - y_n)$$

is a polynomial on $J$. However, this polynomial vanishes on the non-degenerate interval $I \cap J$ and so on the whole of $J$. Hence $z = z_1$.

A simple induction proof now yields the following:

**Proposition 1** *Let $(I_1, \dots, I_n)$ be a covering of the compact interval $I$ by such intervals so that each pair $I_j$ and $I_k$ is either disjoint or intersects in a non-degenerate interval. Then if we have a sequence of distributions $(x_i)_{i=1}^n$ where $x_i$ is defined on $I_i$ and $x_i$ coincides with $x_j$ on $I_i \cap I_j$ for each pair $i, j$, there is a unique distribution $x$ on $I$ which agrees with each $x_i$ on $I_i$.*

**Multiplication:** We now consider the problem of extending the operation of multiplication to distributions. Consider the following example which is due to Laurent Schwartz.

EXAMPLES. We "calculate" the product $s^{-1}.s.\delta_0$, firstly as $(s^{-1}.s).\delta_0$, then as $s^{-1}(s.\delta_0)$. In the first case, we have the calculation

$$\begin{aligned} s^{-1}\cdot &= (D \ln |s|) \cdot s = D(s \cdot \ln |s|) - \ln |s| \\ &= D(s \cdot \ln |s|) - D(s \cdot \ln |s| - s) = Ds = 1 \end{aligned}$$

and so the product above is equal to $\delta_0$.

On the other hand, if we write $X$ for that primitive of $H_0$ which vanishes at 0, then

$$H_0 = DX = D(s \cdot H_0) = H + s \cdot \delta_0$$

and so $s \cdot \delta_0 = 0$. Hence $s^{-1}(s \cdot \delta_0) = 0$.

Of course, these manipulations are purely formal (we have not yet defined the concept of the product of two distributions). What they show is that there is no multiplicative structure on $C^{-\infty}$ under which this space is a ring and the usual rules for the derivatives of sums and products hold. Hence we shall have to be more modest when attempting to define products of distributions. We shall commence by defining products $x.y$ where $x$ is in $C^n(I)$ and $y$ is in $C^{-n}(I)$.

**Lemma 4** *Suppose that $x$ is in $C^{n+1}$ and $y_1$ is in $C(I)$. Then*

$$\sum_{k=0}^{n+1} (-1)^k \binom{n+1}{k} D^{n+k}(D^k x \cdot y_1) = \sum_{k=0}^{n} (-1)^k \binom{n}{k} D^{n+1-k}(D^k x \cdot y_1) - \sum_{k=0}^{n} (-1)^k D^{n-k}(D^k x \cdot D y_1).$$

*Hence if $y_1$ is in $C^1(I)$, then*

$$\sum_{k=0}^{n+1} (-1)^k \binom{n+1}{k} D^{n+1-k}(D^k x \cdot z_1) = \sum_{k=0}^{n} (-1)^k \binom{n}{k} D^{n-k}(D^k x \cdot D y_1).$$

Proof. The first equation is a simple consequence of the relationship

$$\binom{n}{k} + \binom{n}{k-1} = \binom{n+1}{k}.$$

The second follows from the fact that

$$\sum_{k=0}^{n} (-1)^k D^{n+1-k}(D^k x \cdot y_1) = \sum_{k=0}^{n} (-1)^k \binom{n}{k} D^{n-k}(D^{k+1} x \cdot y_1 + D^k x \cdot D y_1).$$

■

Using this fact, we can now define the distribution $x.y$ where $x$ is in $C^n(I)$ and $y$ is in $C^{-n}(I)$ by putting

$$x \cdot y = \sum_{k=0}^{n} (-1)^k \binom{n}{k} D^{n-k}(D^k x \cdot y_1)$$

whereby $y = \tilde{D}^n y_1$ with $y_1$ a continuous function.

As usual, there are a number of things to be checked about this definition. Firstly, it follows immediately from the defining expression that the product depends linearly on each factor. We next note that if $x$ is in fact in $C^{n+1}(I)$, then the following version of Leibnitz' rule holds:

$$D(x \cdot D^n y_1) = Dx \cdot D^n y_1 + x \cdot D^{n+1} y_1.$$

In order to show that the product is well-defined, it obviously suffices to show that if $y_1$ is in $P^n(I)$ and $x$ is in $C_n(I)$, then the defining sum above vanishes. But by the result of the last Lemma, we have that this sum is equal to

$$\sum_{k=0}^{n-1} (-1)^k \binom{n-1}{k} D^{n+1-k}(D^k x \cdot D y_1)$$

i.e. the corresponding sum (up to $n-1$) for $D_1$, which is in $P^{n-1}(I)$. Hence we can prove that it vanishes by induction.

Examples. Suppose that $I$ is an interval which has 0 as an interior point and that $x$ is in $C(I)$. Then we claim that $x.\delta_0 = x(0).\delta_0$. More generally, we have the relationship

$$x \cdot \delta_0^n = \sum_{k=0}^{n} (-1)^k \binom{n}{k} x^{(k)}(0) \cdot \delta_0^{(n-k)}.$$

The first formula is obtained by differentiating $s.H_0$. Firstly, using the Leibnitz rule we have

$$D(x.H_0) = Dx.H_0 + x.\delta_0.$$

On the other hand, using the formulae for the derivatives of piecewise smooth functions, we have

$$D(x \cdot H_0) = Dx \cdot h_0 + x(0).\delta_0.$$

As a further example, we have the formula $s^{-1} \cdot s = 1$ which we leave to the reader to verify.

We remark briefly on some extensions of the definition of the product of distributions or functions which are often useful. Firstly, one can use this concept of the product of a continuous function and a measure to define products of the form $x.y$ where $x$ is in $C^n$ and $y$ is the $n$-th (distributional) derivative of a measure. We can also use the principle of *recollement des morceaux* to define the product of two distributions such as $\delta_a.\delta_b$ (where $a \neq b$) i.e. distributions where the singularities of one correspond to regions where the other factor is smooth. More precisely, suppose that $x$ and $y$ are two distributions on $I$ such that there is a covering $(I_1, \dots, I_n)$ by compact intervals of the type discussed above so that on each $I_i$ there is an $n_i$ so that on $I_i$ either $x$ is in $C^{n_i}(I)$ and $y$ in $C^{-n_i}(I)$ or vice versa. Then we can define the product $x.y$ on $I$ by defining it on each $I_i$ and then using the above principle.

## 1.5 Division of distributions

In the following, we consider the problem of division by distributions and show that the latter can be divided by non-vanishing polynomials. We begin with the simplest case—that where the polynomial is the identity function:

**Proposition 2** *Let $I$ be a compact interval which contains $0$ as an interior point. Then if $x$ is a distribution on $I$, there is a distribution $y$ thereon so that $s.y = x$.*

PROOF. We assume that $x$ has the form $\tilde{D}^n X$ where $X$ is continuous and introduce the function

$$Y : s \mapsto \begin{cases} s^n I_a(\frac{X}{t^{n+1}}) & (s < 0) \\ s^n I_b(\frac{X}{t^{n+1}}) & (s > 0) \end{cases}$$

(recall that $I_a$ and $I_b$ are the integration operators with base points $a$ and $b$ resp., the latter being the left and right-hand endpoints of $I$). The reader can check that $Y$ is Lebesgue integrable (in fact, we have estimates of the type

$$|Y(s)| \leq C \int_a^s -\frac{1}{t}\, dt \leq C(\ln(-a) - \ln(-s)) \quad (s\lambda < 0)$$

resp.

$$|Y(s)| \leq C(\ln b - \ln s) \quad (s > 0)$$

where $C = \sup |X(t)|$.

We claim that the distribution $y = \tilde{D}^{n+1}Y$ has the required property. Indeed, by the Leibnitz rule

$$D^{n+1}(s \cdot Y) = s \cdot D^{n+1}Y + (n+1)D^n y.$$

On the other hand, the function $sZ$ is piecewise smooth and so we have

$$\tilde{D}(s.Y) = (n+1)Y + X$$

by the formula on p. ??. Hence $\tilde{D}^{n+1}(s.Y) = (n+1)\tilde{D}^n Y + x$. Comparing the two equations, we see that $s.y = x$.

By induction, we see that if $x$ is a distribution on $I$ and if $r$ is a positive integer, then there is a distribution $y$ so that $s^r.y = x$. In fact, such a distribution is $\tilde{D}^{n+r}Y$ where $Y$ is the function

$$Y : s \mapsto \begin{cases} \frac{s^n}{(r-1)!} \int_a^s (s-t)^{r-1} \frac{X(t)}{t^{n+r}}\, dt & (s < 0) \\ \frac{s^n}{(r-1)!} \int_b^s (s-t)^{r-1} \frac{X(t)}{t^{n+r}}\, dt & (s > 0). \end{cases}$$

We remark here that $y$ is not uniquely determined by the equation $s^r.y = x$. Indeed, any distribution $y$ of the form

$$\sum_{j=0}^{r-1} c_j \delta_0^{(j)}$$

satisfies the equation $s^r.y = 0$ as we have seen above so that we can add such a distribution to a solution of the above kind and we still have a solution. We shall see later that these are the only solutions i.e. distributions $y$ so that $s^r.y = x$ are determined up to such combinations of derivatives of the delta-function.

Using this result and the principle of *recollement des morceaux* one can show that if $p$ is a smooth function on $I$ whose only zeros are isolated ones of finite order (typically the restriction to $I$ of polynomials or, more generally, analytic functions on the complex plane), then for each distribution $x$ on $I$ there is a distribution $y$ with $p.y = x$. We leave the details to the reader.

We now consider briefly the problem of changes of variable for distributions. We consider two compact intervals $I$ and $J$ with a smooth mapping $\phi$ from $I$ into $J$. We wish to define an operator $x \mapsto x \circ \phi$ from $C^{-\infty}(J)$ into $C^{-\infty}(J)$ in such a way that

a) if $x$ is a continuous function, then $x \circ \phi$ has its usual meaning;

b) the chain rule holds i.e. $\tilde{D}(x \circ \phi) = (\tilde{D}x \circ \phi)\phi'$. It is easy to see that if $x$ has the form $\tilde{D}^n X$ for some continuous function, then $x \circ \phi$ must be the distribution $(\frac{1}{\phi'}D)^n(X \circ \phi)$.

In fact, if we choose $\phi$ so that the latter expression is meaningful, then this can be used to define composition. More precisely, suppose that $\phi$ is in $C^{n+1}$ and that its derivative has no zeros in $I$. Then the above definition induces a continuous linear mapping $x \mapsto x \circ \phi$ from $C^{-nk}(J)$ into $C^{-n}(I)$. In order to prove this, it suffices to show that the composition $x \circ \phi$ is well-defined i.e. that if a distribution $x$ vanishes, then so does $x \circ \phi$. But in this case, $x = \tilde{D}^n X$ where $X$ is a polynomial of degree??? in which case $x \circ \phi = (??????)$ also vanishes (by the chain rule).

We remark that the operators $\tau_h$ and $\rho_{I,J}$ considered above can be assumed within the framework of this construction. They are just composition with the functions $t \mapsto t - h$ respectively the natural injection from $I$ into $J$ as the reader can easily verify.

Examples. We have the formula

$$\delta_0(kt) = \frac{1}{|k|}\delta_0$$

for any non-zero scalar $k$. More generally,

$$\delta_0^{(n)}(kt) = |k|^{-n}\delta_0$$

(where the expression $\delta_0(kt)$ refers to the delta-function composed with the dilation $t \mapsto kt$).

Remark. As in the case of multiplication, we can extend this definition if we use *recollement des morceaux.* For instance, it suffices to demand the following properties from $\phi$. We have a suitable covering $(I_1, \dots, I_n)$ of $I$ by compact intervals so that for each $i$, either

a) the restriction of $\phi$ to $I_i$ satisfies the conditions required above

or

b) the restriction of $x$ to $I_i$ is continuous in which case e it suffices that $\phi$ be continuous on $I_i$.

Under these conditions we can define $x \circ \phi$. Thus using this method we can define such expressions as $\delta_0(t^2 - a^2)$ (for $a \neq 0$).

## 1.6 Distributions on open sets

Our next task will be to define distributions on the whole real line $\mathbf{R}$ of, more generally, on open subsets resp. intervals thereof. One possible method would be to consider the operation of differentiation on a suitable class of functions on $\mathbf{R}$. Two natural choices are the Banach space $C^b(\mathbf{R})$ of bounded, continuous functions on the line resp. the space $C(\mathbf{R})$ of *all* continuous functions there. If we apply the methods of this chapter, we obtain different spaces of distributions and in neither case do we come up with the space of distributions of L. Schwartz.

EXAMPLES. Consider the "distributions:

$$\begin{aligned} T_1 &= \frac{1}{2}\delta_0 + \sum_{n=1}^{\infty}(-1)^n \delta_n; \\ T_2 &= \sum_{n=0}^{\infty} \delta_n; \\ T_3 &= \sum_{n=0}^{\infty} \delta_n^{(n)}. \end{aligned}$$

The first distribution is in the space generated by the bounded, continuous function (as the second derivative of such a function), the second is in that defined by $C(\mathbf{R})$ while the third is in neither.

In order to obtain *all* distributions, we shall use the method of projective limits i.e. we shall define a distribution on $\mathbf{R}$ to be a collection of local distributions defined on the compact subintervals of $\mathbf{R}$ which are compatible. The formal definition is as follows:

**Definition:** Let $U$ be an open interval of $\mathbf{R}$ and write $I(U)$ for the family of all non-degenerate subintervals of $U$. The latter is a directed set when ordered by inclusion. The family

$$\{C^{-\infty} : I \in \mathcal{I}(\mathcal{U})\}$$

together with the restriction mappings

$$\{\rho_{I,J} : J \subset I\}$$

is a projective system of vector space. The space $C^{-\infty}(U)$ of distributions on $U$ is by definition its projective limit. Thus a distribution on $U$ is a family $\{x_I : I \in \mathcal{I}(\mathcal{U})\}$ where $x_I$ is a distribution on $I$ which is compatible in the sense that whenever $I$ and $J$ are intervals with non-degenerate intersection, then $x_I$ and $x_J$ coincide on this intersection. As a projective limit of

LVS's $C^{-\infty}(U)$ itself has a limit structure, a sequence $(x^n)$ of distributions converging if and only if the sequence of components $(x_I^n)$ converges in the appropriate distribution space for each $I$.

We remark that we do not really need to specify a distribution on *all* compact subintervals of $U$. For example, a sequence $(x_n)$ where $x_n$ is a distribution on the interval $[-n, n]$ and the restriction of $x_{n+1}$ to $[-n, n]$ coincides with $x_n$ for each $n$ defines in the obvious way a distribution on $\mathbf{R}$.

We now list without comment some facts or constructions for distributions on $\mathbf{R}$ which follow immediately from the definition and the corresponding properties of distributions on compact intervals:

I. each locally integrable function on $\mathbf{R}$ (or on $U$) can be regarded as a distribution there. Similarly, a measure (not necessarily bounded) on the line can be regarded as a distribution.

II. we can extend the definition of the derivative of a distribution by defining the derivative of $x$ which is defined by the family $(x_I)$ to be the distribution defined by the family $(\tilde{D}x)_I$. Of course, we denote the latter distribution by $\tilde{D}x$.

III. The space of distributions on $\mathbf{R}$ satisfies the natural form of the axioms given at the beginning of this chapter with the exception of ???? Thus the distribution ??? is not the repeated derivative of any continuous function. A distribution which *is* the $n$-th derivative of a continuous function on the line for some integer $n$ is called a **distribution of finite order.**

IV. If $U_1$ is a subinterval of $U$, we can, in the obvious way define the restriction $\rho_{U,U_1}$ from the space of distributions on $U$ into the space of distributions in $U_1$. Similarly, we can define a translation operator $\tau_h$ from $C^{-\infty}(U)$ into $C^{\infty}(U + h)$.

We have the following version of the principle of *recollement des morceaux.*

**Proposition 3** *Let $(U_\alpha)_{\alpha \in A}$ be a covering of $U$ by open intervals and let $(x_\alpha : \alpha \in A)$ be a family of distributions where $x_\alpha$ is define on $U_\alpha$ and $x_\alpha$ coincides with $x_\beta$ on $U_\alpha \cap U_\beta$ for each pair $\alpha, \beta$. Then there is a unique distribution $x$ on $U$ which coincides with each $x_\alpha$ on $U_O\alpha$.*

PROOF. Consider a compact subinterval $I$ of $U$. Then the family $\{U_\alpha \cap I : \alpha \in |\}$ is an open covering of $I$ and so we can find a finite subcovering. We can then use *recollement des morceaux* for compact intervals to deduce the existence of a distribution $x_I$ on $I$ which agrees with $x_\alpha$ on $U_\alpha \cap I$ for each $\alpha$. (In fact, the result referred to here is not directly applicable since $U_\alpha \cap I$ is not compact. however, we can replace these relatively open intervals with slightly smaller closed ones which still cover $I$ and then apply the appropriate result). The distributions $(x_I)$ are clearly compatible and so define the required distribution on $U$.

■

We can use this principle to define the **support** of a distribution. We say that a distribution $x$ on $U$ has support in a closed set $C \subset U$ if the restriction of $x$ to the open $U \setminus C$ vanishes. There is a smallest such set which is obtained as follows. We say that $x$ **vanishes** on the open subset $U_1$ if its restriction there is the zero-distribution. It follows from the above principle that there is a largest such open set, namely the union of all open sets on which $x$ vanishes. The complement of this set is defined to be the **support** of $x$.

For example, the distribution $\delta_0$ and its derivatives have support in the one-point set $\{0\}$. In fact, they are the only such distributions. More precisely, if $T$ is a distribution on the line with support in $\{0\}$, then $T$ is a (finite) linear combination of $\delta_0$ and its derivatives. For the restriction of $x$ to the interval $[-1,1]$ has the form $D^r X$ for some continuous function and since $x$ vanishes on the open intervals $]0,1[$ and $]0,1[$, there are polynomials $p_-$ and $p_+$ (both of degree less than $r$) so that $X$ coincides with $p_-$ on $[-1,0[$ resp. with $p_+$ on $]0,1]$. Thus $X$ is a piecewise smooth function whose only singularity is at the origin. The result now follows from the formula for the distributional derivative of such a function.

From this we can deduce the fact which has already been mentioned that the only solutions of the equation $s^n \cdot x = 0$ are the elements in the linear span of $\delta_0$, $\delta_0^{(1)}$, ..., $\delta_0^{(n-1)}$. For recall that

$$s^r \cdot \delta_0^{(n-1)} = \sum_{k=0}^{n} (-1)^k \binom{n}{k} (s^r)^{(k)}(0) \cdot \delta_0^{(n-k)}$$

which vanishes if $r > n$ and has the value $(-1)^r \binom{n}{r} r! \delta_0^{(n-r)}$ otherwise. This implies the sufficiency of the above condition.

Now suppose that $s^n \cdot x = 0$. Clearly $s^r \cdot x = 0$ for $r \geq n$. Now it follows immediately that $x$ is supported by $\{0\}$ and is a linear combination of derivatives of $\delta$-functions. Suppose that the highest derivative which occurs is of order $r$ with $r \geq n$. Then if we substitute the expression $\sum_{j=0}^{r} \lambda_j \delta_0^{(j)}$ for

$x$ in the equation $s^r \cdot x = 0$ and use the above formula, we see that $\lambda_r = 0$ which is a contradiction.

Later we shall require the concept of the space of distributions with **compact support**. By definition, these are the distributions $x$ in $C^{-\infty}(\mathbf{R})$ for which there exists a compact interval $I$ so that $x$ is supported by $I$.

## 1.7 Limits and integrals of distributions

In order to define the limit of a distribution (at a point of the line or at infinity), we are forced to use a definition which is motivated by L'Hopital's rule. we begin with the case of a distribution $x$ on an interval $I$ which is unbounded on the right. We say that $x(s) \to 0$ as $s \to \infty$ if there exist an integer $p$, a $b \in I$ and a continuous function $X$ on $[b, \infty]$, so that

$$x = D^p X \text{ and } \lim_{s \to \infty} \frac{X(s)}{s^p} = 0.$$

We leave to the reader the verification of the fact that if this condition holds for one $p$-primitive $X$ of $x$, then it holds for each such primitive).

More generally, we write $\lim_{s \to \infty} x(s) = \lambda$ to mean that $\lim_{s \to \infty}(x(s) - \lambda) = 0$ i.e. that there is a $p$-primitive $X$ so that

$$\lim_{s \to \infty} \frac{X(s)}{s^p} = \frac{\lambda}{p!}.$$

For example, since $\cos t = D(\sin t)$ and $\dfrac{\sin t}{t} \to 0$, we see that $\lim_{t \to \infty} \cos t = 0$ in the distributional sense.

Of course, the usual rules for the calculations of limits are valid. Thus we have

$$\lim_{s \to \infty} (x(s) + y(s)) = \lim_{s \to \infty} x(s) + \lim_{s \to \infty} y(s)$$

provided both terms on the right hand side exits. If $I$ is unbounded on the left, we define $\lim_{s \to -\infty} x(s)$ in the analogous way.

Suppose now that $a$ is the (finite) left endpoint of the interval $I$. Then the condition $\lim_{s \to a+} x(s) = \lambda$ means that $x$ has a representation as $D^p X$ in a neighbourhood of $a$ so that

$$\lim_{s \to a+} \frac{X(s)}{(s-a)^p} = \frac{\lambda}{p!}.$$

We define the limit $\lim_{s \to b-} x(s)$ analogously, where $b$ is the right hand endpoint. Finally, if $c$ is an interior point of the interval, then $\lim_{s \to c} x(s) = \lambda$ means that $\lim_{s \to a-} x(s) = \lambda = \lim_{s \to a+} x(s)$.

For example, it is clear that

$$\lim_{s\to 0} \delta_0^{(k)}(s) = 0$$

for any derivative of the delta-function.

In contrast to the case of limits at infinity, the above condition on the primitive $X$ of $x$ clearly implies that it has limit 0 at $c$. hence the defining condition will only be satisfied by *one* primitive of $x$.

## 1.8 Continuity at a point

In general, we cannot talk about the value of a distribution at a point. however, it is clear that the expression "the value of the delta-function at the point 1 is zero" should be a valid statement. This and more subtle phenomena are covered by the following definition. Let $x$ be a distribution on the interval $I$ which contains the point $a$ as an interior point. Then we say that $x$ is continuous at $a$ with value $\lambda$ if $x$ has a representation $D^p X$ near $a$ whereby $X$ is continuous and

$$\lim_{s\to a} \frac{X(s)}{(s-a)^p} = \frac{\lambda}{p!}.$$

Note that this condition is stronger than the mere existence of the limit as $s$ tends to $a$. Thus $\delta_0$ is not continuous at 0 (although $\lim_{s\to 0} \delta_0(s)$ exists). On the other hand, the distribution $\cos \frac{1}{s}$ *is* continuous at 0 (with value 0 there). This follows from the representation

$$\cos \frac{1}{s} = 2s \cdot \sin \frac{1}{s} - D(s^2 \cdot \sin \frac{1}{s}).$$

## 1.9 Primitives and definite integrals

It is clear that any distribution $x$ on a compact interval has a primitive i.e. a distribution $X$ with $DX = x$. For if $x$ has a representation $D^p Y$ it suffices to take $D^{p=1} Y$. Further any two such distributions differ by a constant. The same holds true for a distribution on an open interval $U$. For $U$ can be covered by an increasing sequence $(I_n)$ of compact intervals. If $x$ is a distribution on $U$, we can find a distribution $X_1$ on $I_1$ so that $DX_1 = x$ on $I_1$. Similarly, there is a distribution $X_2$ on $I_2$ with $DX_2 = x$ on $I_2$. Then $X_1$ and $X_2$ differ at most by a constant on $I_1$. By subtracting this constant from $X_2$ we can ensure that $X_1$ and $X_2$ coincide on $I_1$. Proceeding in this manner, we can construct a sequence $(X_n)$ where $X_n$ is a distribution on $I_N$ with derivative $x$. Furthermore, the $X_n$ are compatible and so define a distribution on $U$ which is a primitive for $x$.

If we now combine the notions of primitive and limit of a distribution, we can carry over the classical definition of the definite integral of a function to distributions. Note however that we must be more careful about specifying behaviour at the endpoints.

Let $I$ be an interval with endpoints $a$ and $b$ (the latter need not be in $I$ i.e. we are not assuming that the interval is closed). We say that a distribution $x$ in $C^{-\infty}(I)$ is integrable (more precisely, that $\int_{a+}^{b-} x(s)\,ds$ exists) if $x$ has a primitive $X$ for which $\lim_{s\to a+} X(s)$ and $\lim_{s\to b-} X(s)$ exist. Of course, the integral is then defined to be the difference of the two limits. Similarly, we define such expressions as

$$\int_{a-}^{b-} \qquad \int_{a-}^{b+} \qquad \int_{a+}^{\infty} \qquad \int_{a-}^{\infty} \qquad \int_{-\infty}^{b+}$$

etc. We remark that in order to define say $\int_{a-}^{b+} x(s)\,ds$, we must assume that $x$ is defined on an interval which contains $a$ and $b$ as *interior* points.

For example, the reader may check that

$$\begin{aligned}
\int_{-\infty}^{\infty} \delta_0(s)\,ds &= \int_{0-}^{0+} \delta_0(s)\,ds = 1 \\
\int_{-\infty}^{\infty} \delta_0^{(k)}(s)\,ds &= 0 \quad (k > 0) \\
\int_{-\infty}^{\infty} e^{ist}\,dt &= 0 \quad \text{if} \quad s \neq 0.
\end{aligned}$$

The last equation follows from the fact that

$$e^{ist} = D_t\left(\frac{1}{is}e^{ist}\right)$$

and $\lim_{t\to\pm\infty} e^{ist} = 0$.

We note here that if a distribution $x$ on the line is supported in a compact interval $I = [a,b]$, then it is integrable and its integral over the line is equal to $\int_{a-}^{b+} x(s)\,ds$. This follows from the fact that there exist constants $c_1$ and $c_2$ so that $X - c_2$ on $]b,\infty[$ and $X = c_1$ on $]-\infty,a[$ ($X$ a primitive of $x$). Then both of these integrals are equal to $c_2 - c_1$.

## 1.10 The order of growth of distributions

Recall that if $\pi$ is a positive, continuous function defined on some neighbourhood $[a,\infty[$ of infinity, then a function $x$ (also defined in a neighbourhood of infinity) is said to be **of order** $\pi$ (written $x \in O(\pi)$) if there is a $b > a$ so that the quotient $\frac{x}{\pi}$ is bounded on $[b, \infty[$. $x$ is **small of order** $\pi$ if $\lim_{t\to\infty} \dfrac{x(t)}{\pi(t)} = 0$ (written $x \in o(\pi)$ as $t \to \infty$).

In fact, we shall only be concerned with the case where $\pi$ is a power of $t$.

We can extend these definition to distributions as follows:

**Definition 2** *Let $I$ be an interval of the form $[a, \infty[$, $\pi$ as above. Then we say*

- *a distribution $x$ in $C^{-\infty}(I)$ is bounded near infinity f(written $x \in O(1)$ as $t \to \infty$) if $x$ has a representation $x = \tilde{D}^n X$ on a neighbourhood of infinity, where $X$ is a continuous function which is $O(t^n)$ in the classical sense;*

- *we say that $x \in O(\pi)$ as $t \to \infty$ if $x = y \cdot \pi$ where $y$ is a distribution which is bounded near infinity. (In this definition, we are tacitly assuming that $\pi$ is smooth);*

- *that $x \in o(\pi)$ as $t \to \infty$ if $x = y \cdot \pi$ where $\lim_{t\to\infty} y(t)$ in the distributional sense.*

*In the same way, we can define the concepts of functions which are $O(\pi)$ resp. $o(\pi)$ as $t$ tends to $a-$, $b+$ or $-\infty$.*

One of the interesting consequences of this definition is that integrability for a distribution is more closely related to its rate of growth than is the case for functions. In fact, we have the following properties:

**Proposition 4** *If $x \in O(t^\alpha)$ as $t \to \infty$, then $Dx \in O(t^{\alpha-1})$.*

For we have the equation

$$Dx = t^{-1}(tD^{n+1}) = t^{-1}(D^{n+1}(tX)) - (n+1)D^n X$$

where $x = D^n X$. Now if $x$ is bounded, we can find such a representation where $X$ is $O(t^n)$. It follows easily then that $DX$ is $O(t^{n-1})$.

**Proposition 5** *If $x$ is integrable, then $x$ is $O(t^{-1}$.*

For if $X$ is a primitive of $x$, then $\lim_{t\to\infty} X(t)$ exists and hence $X$ is bounded near infinity. The result now follows from the first Proposotion.

**Proposition 6** *If $x \in O(t^\alpha)$ for some $\alpha < 1$, then $x$ is integrable.*

For the hypothesis means that $x = t^\alpha D^n X$ near infinity, where $\frac{X(t}{t^n}$ is bounded. now we can expand the right hand side as

$$\sum_{k=0}^{n} c_k D^{n-k}(t^{\alpha-k}X)$$

for suitable coefficients $c_k$. This has primitive

$$\sum_{k=0}^{n-1} c_k D_{n-k-1}(t^{\alpha-k}X) + c_n \int_0^t s^{\alpha-n} X(s)\, ds$$

and

$$\left|\frac{X(s)}{s^n}\right| \leq K.$$

## 1.11 Distributions as functionals

As mentioned in the introduction, distributions were defined by Schwartz as functionals on spaces of test functions. The fact that his theory is equivalent to the one given here is a consequence of the fact that distributions in the Schwartzian sense satisfy the axioms used here. however, we shall now display a more direct connection between the two theories by using integration theory to show tl how elements of $C^{-\infty}$-spaces can be regarded as functionals. **Definition:** We say that the distribution **acts on the distributions** $y$ (where both are in $C^{-\infty}(I)$ for some interval $I$) if the product $x \cdot y$ is defined and the definite integral $\int_I x(s)y(s)\, ds$ exists. In this case we denote the integral by $T_x(y)$.

EXAMPLES. If the interval $I$ is compact, then each distribution on $I$ acts on *each* function in $C_0^\infty(I)$. For suppose that $x = D^n X$ where $X$ is continuous. Then

$$x \cdot y = \sum_{k=0}^{n} (-1)^k \binom{n}{k} D^{n-k}(X \cdot y^{(k)}).$$

Now all of the terms on the right hand sum have as primitive a distribution which vanishes at the endpoints of $I$, with the exception of the final term $\pm X \cdot y^{(n)}$. Hence $x \cdot y$ is integrable and its integral is $(-1)^n \int_I X(s)y^{(n)}(s)\, ds$.

Since any distribution with compact support is integrable over $\mathbf{R}$, it follows that $T_x(y)$ exists whenever $x \cdot y$ has compact support, in particular when

- $y$ is in $C_0^\infty(\mathbf{R})$ and $x$ is any distribution on $\mathbf{R}$;
- $y$ is in $C^\infty(\mathbf{R})$ and $x$ is a distribution with compact support.

## 1.12 Fourier series and distributions

We say that a real number $h$ is a **period** of a distribution $x$ in $C^{-\infty}(\mathbf{R})$ if $x = \tau_h x$. We write $C^{-\infty}_{-p}(\mathbf{R})$ for the space of distributions with period 1. This is a closed vector subspace of $C^{-\infty}(\mathbf{R})$ as the kernel of the operator $(\tau_n - \mathrm{Id})$. The usual argument shows that the set of periods of a distribution, being a subgroup of $\mathbf{R}$, is either a dense set or is the set of integral multiples of a positive number $h$. In the former case it is then all of $\mathbf{R}$ (by continuity) and so the distribution is constant.,

Examples of distributions in $C^{-\infty}_p(\mathbf{R})$ are, of course, the exponential functions $e^{2\pi int}$ ($n \in \mathbf{Z}$) and so, more generally, trigonometrical polynomials i.e. functions of the form $\sum_{n=n_1}^{n_2} c_n e^{2\pi int}$.

Even more generally, if $(c_n)$ is an absolutely summable sequence indexed by $\mathbf{Z}$, then $\sum_{n\in\mathbf{Z}} c_n e^{2\pi int}$ converges uniformly and so defines a periodic distribution. In particular, this is the case if $(c_n)$ is $)(|n|^{-2})$. We shall now show that if $(c_n)$ is $O(|n|^k)$ for some positive $k$, then this series converges in the distributional sense to a periodic distribution. For consider the series

$$\sum_{n\in\mathbf{Z}\setminus\{0\}} c_n (2\pi in)^{-k-2} e^{2\pi int}.$$

By the above, this series converges (uniformly) to a continuous, periodic function $X$. Differentiating $k+2$ times, we see that the original series converges to the distribution $D^{k+2}X$ (up to a constant).

Examples. Consider the function which maps $s$ onto $s^2 - s$ on $[0,1]$ and is extended periodically to a continuous function on $\mathbf{R}$. Its Fourier series is

$$-\frac{1}{6} - \sum_{n\neq 0} \frac{1}{2\pi^2 n^2} e^{2\pi ins}$$

which converges uniformly on the line. We can use this to obtain the Fourier series of the delta-function, more precisely of its periodic version

$$\delta_0^p = \sum_{n\in\mathbf{Z}} \delta_n.$$

In fact, if we differentiate twice, we get

$$\sum_{n\in\mathbf{Z}} \delta_n = \sum_{n\in\mathbf{Z}} e^{2\pi int}.$$

If we continue differentiating we get the formulae:

$$\sum_{n\in\mathbf{Z}} \delta_n^{(k)} = \sum_{n\in\mathbf{Z}} (2\pi in)^k e^{2\pi int}.$$

We shall now show that, conversely, any distribution $x$ in $C_p^{-\infty}(\mathbf{R})$ has a series representation $\sum_{n\in\mathbf{Z}} c_n e^{2\pi int}$ where the Fourier coefficients $(c_n)$ satisfy the above growth condition. For the restriction of $x$ to $[0,2]$ has the form $D^k X$ for a continuously differentiable function on $[0,2]$ (we take the primitive of a continuous function which represents $x$). Since $x$ is periodic, $X - \tau_1 X$ is a polynomial of degree at most $n-1$ on $[1,2]$ i.e.

$$X|_{[1,2]} = \tau_1(X|_{[0,1]}) + H_1 p$$

for some $p \in P^n([1,2])$.

If we denote by $\tilde{X}$ the periodic extension of $X$ to a function on $\mathbf{R}$ (which is piecewise continuously differentiable), then $x = D^k\tilde{X} + y$ where $y$ is a linear combination of distributions of the form $\sum_{n\in\mathbf{Z}} \delta_n^{(r)}$ $(r = 0, 1, \dots, k-1)$ by the formula for the derivative of piecewise smooth functions. Hence it suffices to prove the existence of a Fourier series representation for distributions of type $D^k\tilde{X}$ (which follows by differentiating the series for $\tilde{X}$, the latter converging uniformly) and of type $y$, which was covered in the above example.

Examples. (The Poisson summation formula). Using the formula for the Fourier series of the periodic delta-function, we have, for each test function $\pi$ in $C_0^\infty(\mathbf{R})$, the identity

$$\sum_{n\in\mathbf{Z}} \pi(n) = 2\pi \sum_{n\in\mathbf{Z}} \hat{\pi}(n)$$

where $\hat{\pi}$ denotes the Fourier transform

$$t \mapsto \frac{1}{2\pi}\int_{\mathbf{R}} \pi(s)e^{2\pi st}\, ds$$

of $\pi$, which is known as the **Poisson summation formula**. For the left hand side is the value of the functional defined by the distribution $\sum_{n\in\mathbf{Z}}$ at $\pi$, while the right hand side is the value of $\sum_{n\in\mathbf{Z}} e^{2\pi int}$.

We remark the easily proved fact that if $\alpha$ is a smooth periodic function and $x$ is a periodic distribution (i.e. $\alpha \in C_p^\infty(\mathbf{R})$ and $x \in C_p^{-\infty}(\mathbf{R})$), then $\alpha x$ is also periodic and its Fourier series is obtained by formal multiplication of those of $x$ and $\alpha$.

If $x$ is an odd function i.e. such that $x(-t) = -x(t)$ for $t \in \mathbf{R}$, then its Fourier coefficients $(c_n)$ satisfy the conditions $c_{-k} = -c_k$ and so the Fourier representation of $x$ can be manipulated into the form

$$\sum_{n=1}^{\infty} b_n \sin(2\pi nt)$$

where $b_n =?$ just as in the case of functions.

Similarly *even* distributions have cosine representations. For example the series for the periodic delta function can be rewritten in the form

$$1 + 2\sum_{n=1}^{\infty} \cos(2\pi n t).$$

As a further example, we calculate the Fourier representation of the distribution cosec $(2\pi 2)$. The latter can be regarded as a periodic distribution since it is a meromorphic function whose only singularities are simple poles at the points $2\pi n \quad (n \in \mathbf{Z})$. It can be defined directly by the formula

$$\operatorname{cosec}(2\pi t) = \frac{1}{2\pi} D(\ln|\tan(\pi t)|),$$

the function $t \mapsto \ln|\tan(\pi t)|$ being locally integrable and 1-periodic. We now assume that the Fourier series of the cosec-function is $\sum_{n\in\mathbf{Z}} c_n E^{2\pi n t}$ and use the fact that

$$(\sin 2\pi t) \cdot (\operatorname{cosec} 2\pi t) = 1.$$

Of course, this equation holds pointwise, except at the singularities. In fact it holds in the sense that the product of the smooth function represented by the first term with the distributional second term is the constant function "one". We leave it to the reader to check that this is in fact, the case.

Now the Fourier series of $\sin 2\pi t$ is $\frac{1}{2\pi i}(e^{2\pi t} - e^{-2\pi t})$. Hence we can multiply the two series out to get

$$(\sum_{n\in\mathbf{Z}} c_n e^{2\pi i n t}) \cdot (e^{2\pi i t} - e^{-2\pi i t}) = 2i$$

i.e.

$$\sum_{n\in\mathbf{Z}} c_n (e^{2\pi i (n+1) t} - e^{2\pi i (n-1) t}) = 2i$$

i.e.

$$\sum_{n\in\mathbf{Z}} (c_{n-1} - c_{n+1}) e^{2\pi n t} = 2i.$$

Comparing coefficients we see that

$$c_{-1} - c_1 = 2i \quad \text{and} \quad c_{n-1} - c_{n+1} = 0 \quad (n \neq 0).$$

Since the distribution we are analysing is odd, we have further that the even coefficients vanish. Hence

$$c_1 = c_3 = \cdots = -i$$

and

$$c_{-1} = c_{-3} = \cdots = -i,$$

the other coefficients vanishing. Thus the Fourier expansion is

$$2 \sum_{n=0}^{\infty} \sin((2n+1)2\pi t).$$

# 2 The Schwartzian theory

We now use the duality theory for locally convex spaces to develop the Schwartzian theory of distributions. In view of the fact that the material has been covered in the first chapter from a different point of view and will be treated again in more depth in later chapters, we shall merely give a brief overview of the theory in the present chapter.

We begin with the elementary case of distributions on a compact interval $I = [a, b]$ of the line. We shall show here that the $C^{-\infty}$-spaces which we constructed in the first chapter can also be defined by means of duality. Our starting point is the Banach spaces $C_0^n(I)$. This is the space of $C^n$-functions in $I$ which vanish, together with all derivatives up to order $n$, at the endpoints of the interval. Note that these can also be described as those functions $x$ for which the extension to a function on the line which is obtained by setting $x$ equal to zero outside of $I$ is also $C^n$. Of course, $C_0^n(I)$ is a closed subspace of $C^n(I)$ and so is a Banach space with the norm

$$|| \, ||^n : x \mapsto \max\{||x^{k)}||_\infty : k = 1, \dots, n\}$$

The dual of $C_0^n(I)$ is called the **space of distributions of order** $n$ on $I$ and denoted by $D^n(I)$. Note that this space does not coincide with $C^{-n}(I)$ (it consists of the distributional $n$-th derivatives of measures rather than of continuous functions as we shall see shortly). However, the union of the $D^n$-spaces, which we denote temporarily by $D(I)$, will be shown to satisfy the axioms of chapter I and so *is* the space of distributions on $I$.

We remark firstly that it follows from one of the versions of the Weierstraß approximation theorem that $C_0(I)$ is dense in each $C_0^n(I)$ and so that $C_0^{n+1}(I)$ is dense there. Hence we can regard $D^n(I)$ as a subspace of $D^{n+1}(I)$ (more precisely, the adjoint of the embedding of $C_0^{n+1}(I)$ into $C_0^n(I)$ is an injection of $D^n$ into $D^{n+1}$).

We now show how to differentiate distributions. The mapping $D$ is continuous and linear form $C_0^{n+1}(I)$ into $C_0^n(I)$. Hence its transpose $D'$ maps $D^n(I)$ into $D^{n+1}(I)$. We define the **derivative** of a distribution $f$ in $D^n(I)$

to be the distribution $-D'(f)$ in $D^{n+1}(I)$. (As a temporary notation, we shall denote distributions in this section by letters such as $f$, $g$ etc. to emphasise that they are elements of the dual spaces of Banach spaces of smooth functions). The rather mysterious minus sign in the above definition comes from the formula for integration by parts as we shall see below. In order to distinguish between the operators $D$, $\tilde{D}$ and the one introduced above, we shall denote the latter by $\Delta$. Once again, this is a temporary notation to avoid confusion in the following.

Our first task will be to show how to regard continuous functions as distributions i.e. as functionals on the $C_0^n$-spaces. Of course, this is done by means of integration i.e. we identify the continuous function $y$ with the distribution

$$T_y : x \mapsto \int_I x(t)y(t)dt$$

which is a linear form on each $C_0^n(I)$.

We now verify that the distributional derivative of a smooth function coincides with its classical derivative. In terms of the notation which we have just introduced, this means that for $x$ in $C^1(I)$, $T_{x'} = \Delta T_x$. This is an exercise in integration by parts. For if $y$ is in $C_0^1(I)$, we have

$$\begin{aligned} T_{x'}(y) &= \int_I x'(t)y(t)dt \\ &= x(t)y(t)|_a^b - \int_I x(t)y'(t)dt \\ &= -T_x(y') = \Delta(T_x)(y). \end{aligned}$$

Of course, this explains the minus sign in the definition of $\Delta$.

In order to show that the space of distributions defined here coincides with that of the first chapter, it only remains to show that each distribution is a repeated derivative of a continuous function. We remark firstly that this is the case for a measure.

We now turn to the general theory of Schwartzian distributions. In order to do this we introduce the following menu of locally convex spaces:

I. Test spaces on a compact interval. We begin with the space of test functions on $I = [a, b]$. Here we define

$$\mathcal{E}([\dashv, \lfloor]) = \varprojlim_{\backslash} \mathcal{E}^{\backslash}([\dashv, \lfloor])$$

where $\mathcal{E}^n([a, b])$ is the space of functions on $[a, b]$ which are $n$-times continuously differentiable. This is a Banach space under the norm

$|| \, ||_\infty^n$ defined above and so $\mathcal{E}([\dashv, \rfloor])$ is a Fréchet space. Hence so is its closed subspace

$$\mathcal{D}([\dashv, \rfloor]) = \varprojlim_{\setminus} \mathcal{D}_{\prime}^{\setminus}([\dashv, \rfloor])$$

which consists of those function all of whose derivatives vanish at the endpoints $a$ and $b$. $\mathcal{D}_{\mathrm{per}}(\mathbf{R})$ is defined to be the closed subspace of $\mathcal{E}([\dashv, \rfloor])$ consisting of those functions $x$ which satisfy the boundary conditions

$$x^{(n)}(a) = x^{(n)}(b)(n = 0, 1, 2, \dots).$$

This space can be identified with that of those smooth functions on $\mathbf{R}$ which are periodic (with period the length of the interval i.e. $(b - a)$).

$\mathcal{D}_{[\dashv, \rfloor]}(\mathbf{R})$ is the closed subspace of $\mathcal{E}([\dashv, \rfloor])$ consisting of those functions which satisfy the boundary conditions

$$x^{(n)}(a) = x^{(n)}(b) \qquad (n = 0, 1, 2, \dots).$$

This can be identified with the space of those smooth functions on the line which have support in $[a, b]$. The corresponding dual spaces are

$\mathcal{D}'(I)$ - the space of distributions on $I$;

$\mathcal{D}'_I(\mathbf{R})$ - the space of distributions on the line with support in $I$;

$\mathcal{D}'_p(I)$ - the space of periodic distributions on $I$.

II. Test functions and distributions on open subsets of $\mathbf{R}$. If $U$ is an open interval in $\mathbf{R}$, we can represent it as a union $\bigcup I_n$ of an increasing sequence of compact intervals. For example, if $U = ]0, 1[$, we take $I_n = ]\frac{1}{n}, 1 - \frac{1}{n}[$. If $U = \mathbf{R}^+$, we take $I_n = ]\frac{1}{n}, n[$. If $U = \mathbf{R}$, we take $I_n = ]-n, n[$.

    We then define $\mathcal{E}(\mathcal{U})$ to be the projective limit

    Thus $\mathcal{E}(\mathcal{U})$ is a Fréchet space, $\mathcal{D}(\mathcal{U})$ is an LF-space. We have the following particular cases:

    More generally, if $U$ is an open subset of the line, then $U$ has a unique representation $U = \bigcup_n U_n$ as a countable union of disjoint open intervals. We then define

III. Subsets of $\mathbf{R}^n$: If $U$ is a bounded, open subset of $\mathbf{R}^n$ and $K$ is its closure, then we define the dual space of

If $\Omega$ is an open subset of $\mathbf{R}^n$, we define

$$\mathcal{D}(\otimes) = \lim \mathcal{D}_{\mathcal{K}_\setminus}(\otimes)$$

where $\Omega = \bigcup \Omega_n$ is a covering of $\Omega$ by a sequence of relatively compact, open subsets of $\Omega$, whereby we assume that $\overline{\Omega_n} \subset \Omega_{n+1}$ for each $n$. $K_n$ is the closure of $\Omega_n$. (The space $\mathcal{D}(\otimes)$ , together with its topology, is independent of the choice of the $\Omega_n$ as the reader can check).

Examples of distributions on $\Omega$: As in the case of distributions on compact intervals, we can regard the following mathematical objects as distributions on the open subset $\Omega$ of the line:

locally summable functions;

measures.

Since the space $\mathcal{D}(\otimes)$ has a partition of unity, the following properties hold as immediate consequences of the abstract theory of inductive and projective limits with partitions of unity:

$\mathcal{D}(\otimes)$ is an $LF$-space;

$\mathcal{D}(\otimes)$ is complete;

a subset $B$ of $\mathcal{D}(\otimes)$ is bounded if and only if there is a compact subset $K$ of $\Omega$ so that

$B$ is contained and bounded in $\mathcal{D}_{\mathcal{K}}(\otimes)$;

$\mathcal{D}(\otimes)$ is a Montel space;

a sequence $(\phi_n)$ of test functions converges to zero in $\mathcal{D}(\otimes)$ if and only if there is a compact subset $K$ of $\Omega$ so that the supports of the $\phi_n$ are all in $K$ and $\phi_n^{(k)} \to 0$ uniformly on $K$ for each $k \in \mathbf{N}$.

We define the differentiation operators for spaces of distributions as follows: consider the operator

$$D^p = D_1^{p-1} \dots D_n^{p-n}$$

on $\mathcal{D}(\otimes)$ where $p = (p_1, \dots, p_n)$ is a multi-index. Of course, this is a continuous, linear operator on the latter space and we define the operator $\Delta^p$ on the corresponding space of distributions to be $(-1)^p$ times its transpose (where $(-1)^p$ is, by definition, $(-1)^{p_1 + \dots p_n}$). Of course, $D^p$ and $\Delta^p$ can also be regarded as operators on the following spaces:

The dual of the space $\mathcal{E}(\mathcal{U})$ is called the **space of distributions with compact support**. The nomenclature is justified by the following considerations. We say that a distribution **vanishes**

## 2.1 Unbounded operators and spaces of distributions

Our starting point will be an unbounded, self-adjoint p.l.o. $A$ on the Hilbert space $H$. We shall assume that $A \geq I$ (the significance of this assumption will be discussed later). For each $k$ in $\mathbf{N}$ we introduce a new Hilbert space $H^k$ which is none other than the domain of the operator $A^k$. This is provided with the scalar product

$$(x|y)_k = (A^k x|A^k y).$$

That this is a Hilbert space (i.e. complete) can be seen immediately from the spectral theorem. For we can suppose that $A$ is a multiplication operator $M_x$ defined on an $L^2$-space, where $x$ is a measurable function with $x \geq 1$. Then $H^k$ is the space of functions $y$ in $L^2$ so that $x^k y$ is also in $L^2$, under the scalar product

$$(y|z)_k = \int x^{2k} yz \, d\mu$$

i.e. it is just $L^2(x^{2k}\mu)$–the $L^2$-space associated with the measure $\nu$ which has density $x^{2k}$ with respect to $\mu$. The Hilbert spaces $H^k$ decrease as $k$ increases and we define $H^\infty$ to be the intersection $\bigcap H^k$, with the corresponding locally convex structure. Thus $H^\infty$ is a reflexive Fréchet space as a limit of a sequence of Hilbert spaces. We define $H^{-k}$ to be the dual space of $H^k$. We have the following natural representation of the latter:hb We note that $H^k$ is a dense subspace of $H$. Hence, if we take the adjoint of the injection from $H^k$ into $H$, we obtain a (continuous) injection from $H'$ into the dual of $H^{-k}$. We can identify the first space with $H$ via the Riesz theorem and so we see that $H$ is a subspace of $H^{-k}$ and we have the symmetric diagram

$$H^k \subset H \subset H^{-k}.$$

We remark that this discussion show that $H^{-k}$ can be regarded as the completion of $H$ with respect to the **negative norm** $\| \, \|_{-k}$ which is defined as follows:

$$\|y\|_{-k} = \sup \{(y|z) : z \in B_{H^k}\}.$$

(Of course, the scalar product is that of $H$).

In terms of our concrete representation of $H$ as $L^2(\mu)$ resp. $H^k$ as $L^2(x^{2k}\mu)$, it is easy to check that $H^{-k}$ is just $L^2(x^{-2k}\mu)$, where the duality between $H^k$ and $H^{-k}$ is established by the scalar product in $L^2(\mu)$. we then define $H^{-\infty}$ to be the union $\bigcup H^{-n}$ of the spaces with negative norms. It follows from ??? that $H^{-\infty}$ is the dual of $H^\infty$, in particular, it is a reflexive (DF)-space and hence complete, barrelled and bornological.

We have thus generated an infinite scale

$$H^\infty \subset \cdots \subset H^n \subset \cdots \subset H_0 \subset \cdots \subset H^{-n} \subset \cdots \subset H^{-\infty}$$

of Hilbert spaces (where we have denoted $H$ by $H_0$ for reasons of symmetry), flanked by the limiting cases $H^\infty$ and $H^{-\infty}$.

We remark that if $H$ is a Hilbert space and $H_1$ is a dense subspace with a Hilbert structure defined by an inner product $(\,|\,)_1$ which dominates the scalar product in $H$ (i.e. is such that $(x|x) \leq (x|x)_1$ for each $x$ in $H_1$), then this can be used to generate such a scale.

The space $H^\infty$ and $H^{-\infty}$ constructed above are, when regarded in the $L62$-model, none other than the Nachbin spaces

$$H^\infty = \bigcap_n L^2(x^2 n\mu) \qquad H^{-\infty}(x^{-2n}\mu).$$

In order to investigate the structure of these spaces we use the following very simple result:

**Proposition 7** *If $T$ is a continuous linear operator on $H$ which commutes with $A$, then $T$ maps $H^\infty$ continuously into itself and has a unique extension to a continuous linear operator on $H^{-\infty}$.*

We denote these operators by $T^\infty$ and $T^{-\infty}$ respectively.