Canadian Mathematical Congress, Lecture Series, No. 1

INTRODUCTION TO THE THEORY OF DISTRIBUTIONS

by

ISRAEL HALPERIN
Associate Professor of Mathematics, Queen's University

based on the lectures given by

LAURENT SCHWARTZ
Professor of Mathematics, University of Nancy

UNIVERSITY OF TORONTO PRESS
TORONTO

Printed in Canada
London: Oxford University Press
Reprinted, 1960, 1968

PREFACE

THE Theory of Distributions was the subject of a course of lectures given in the Seminar of the Canadian Mathematical Congress held in Vancouver, August-September 1949. Since then my books have appeared and it does not seem useful to give a summary reproducing exactly my Canadian lectures. Instead, this pamphlet gives a detailed introduction, in terms of classical analysis, for applied mathematicians and physicists. Further study, with my books, requires some knowledge of functional-theoretic analysis. This explains the length of the development of the basic ideas and the brief mention of convolution and Fourier series.

The pamphlet was written by Professor Halperin whom I thank very much. He was led to think through the main problems again and many conceptions here are more Professor Halperin's than mine.

LAURENT SCHWARTZ

CONTENTS

THE THEORY OF DISTRIBUTIONS

§1. Introduction

Since the introduction of the operational calculus at the end of the last century, many formulae have been used which have not been adequately clarified from the mathematical point of view. For instance, consider the Heaviside function $Y(x)$ which vanishes for values of x not exceeding zero and is equal to 1 for positive x. It is said that the derivative of this function is the Dirac delta-function $\delta(x)$ which has the following (mathematically impossible!) properties: it vanishes everywhere except at the origin where its value is so large that

$$\int_{-\infty}^{+\infty} \delta(x)dx = 1.$$

This "function" and its successive "derivatives" have been used with considerable success.

It has been suggested by Dirac himself that the delta-function could be avoided by using instead a limiting procedure involving ordinary (mathematically possible) functions. However, the delta-function can be kept and made rigorous by defining it as a measure, that is, as a set-function in place of an ordinary point-function. This suggests that the notion of point-function be enlarged to include new entities[1] and that the notion of derivative be correspondingly generalized so that within the new system of entities, every point-function should have a rigorously defined derivative. This is done in the theory of distributions. The new system of entities, which we call distributions, includes all continuous functions, all Lebesgue locally summable functions and new objects of which a simple example is the Dirac measure function mentioned above. The more general (but rigorous) process of derivation assigns to every distribution a derivative which is again a distribution and so every distribution, including every locally summable point-function, has derivatives of all orders. The derivative of a locally summable point-function is always a distribution although not, in general, a point-function. However, it coincides with the classical derivative when the latter exists and is locally summable.

This theory of distributions gives rigorous content and validity to the formulae of operational calculus mentioned above. It can be developed not only for functions of one variable but also for functions of several variables and it provides a simple but more complete theory of such topics as Fourier Series and Integrals, Convolutions, and Partial Differential Equations.

[1] Just as the notion of rational number was enlarged by Dedekind to include all real numbers.

A systematic exposition of the theory of distributions is given in *Théorie des Distributions* by Laurent Schwartz, published by Hermann et Cie, Paris as Nos. 1091 (Tome I, 1950) and 1122 (Tome II, 1951) of the series, Actualités Scientifiques et Industrielles. The following pages may however serve as a useful introduction to some of the basic ideas.

§2. Point-functions as functionals

The following discussion will lead to the precise definition of distributions given later in this section.

Let (a, b) be a finite closed interval. A continuous $f(x)$ can be considered as a point-function[2] but it can also be considered in another way:[3] it defines a *functional* $F(\phi)$ by the formula

$$F(\phi) = \int_a^b f(x)\phi(x)dx$$

where $F(\phi)$ is a number[4] defined for every continuous $\phi(x)$. And this functional is linear, that is

$$F(c_1\phi_1 + c_2\phi_2) = c_1F(\phi_1) + c_2F(\phi_2).$$

Since there are linear functionals which cannot be expressed in this way in terms of any continuous (or even merely Lebesgue summable $f(x)$), our first suggestion is that the distributions be defined as the (arbitrary) linear functionals $F(\phi)$ over some suitable set S of continuous $\phi(x)$. The $\phi(x)$ chosen will be called testing functions.

If $f(x)$ is absolutely continuous and has a derivative $f'(x)$, the derivative will also define a linear functional, namely,

$$\int_a^b f'(x)\phi(x)dx.$$

This new functional can be expressed, for some ϕ, in terms of the original $F(\phi)$ by use of integration by parts. The formula

$$\int_a^b f'(x)\phi(x)dx = -\int_a^b f(x)\phi'(x)dx = -F(\phi')$$

is valid for all $\phi(x)$ which vanish at a and at b and have a continuous first derivative. This suggests: first, that we allow S to include only such $\phi(x)$

[2] Point-functions need be defined only almost everywhere and two functions are identified if they are equal almost everywhere. Thus we say that $f(x)$ is identically zero if it has the value zero almost everywhere, that $f(x)$ is continuous if it can be identified with a continuous function; the upper bound of $f(x)$ will mean the essential upper bound, the derivative of $f(x)$ will mean the derivative of that function which can be identified with $f(x)$ and which has a derivative if such a function exists, and so on.

[3] Just as every rational number was considered by Dedekind in a new way: as defining a cut in the set of all rationals.

[4] Numbers may be taken to mean real numbers or complex numbers.

and not all continuous $\phi(x)$; second, that for every linear $F(\phi)$, whether it comes from an $f(x)$ possessing a derivative $f'(x)$ or not, a derivative of F, to be written as $F'(\phi)$ (a linear functional on S), be defined by

$$F'(\phi) = -F(\phi').$$

But if $F'(\phi)$ is to be defined for all ϕ in S we will have to restrict S to include only such $\phi(x)$ as have $\phi'(x)$ also in S. This leads to the condition: $\phi(x)$ shall have derivatives of all orders and they, as well as $\phi(x)$ itself, shall vanish at a and at b.

Important examples of such $\phi(x)$ are the functions

$$\{\phi_{c,d}(x)\}^{1/n}$$

obtained by choosing any positive integral n and $a \leqslant c < d \leqslant b$ and defining

$$\begin{aligned} \phi_{c,d}(x) &= e^{-\left(\frac{1}{x-c}+\frac{1}{d-x}\right)} && \text{for } c < x < d, \\ &= 0 && \text{for all other } x. \end{aligned}$$

We shall not find it necessary to restrict S further. However, in restricting S we must be careful of one point: we intend to "identify" the point-function $f(x)$ and the functional $F(\phi)$ which it defines. Thus we want S to include sufficiently many testing functions so that functions $f(x)$ which are different on (a, b) will be identified with different $F(\phi)$. What is the same thing, we do not want

$$\int_a^b f(x)\phi(x)dx$$

to vanish for all ϕ in S except when $f(x)$ is identically zero on (a, b). Now this requirement is satisfied if S includes all

$$\{\phi_{c,d}(x)\}^{1/n},$$

described above, for

$$\lim_{n\to\infty} \int_a^b f(x)\{\phi_{c,d}(x)\}^{1/n}dx = \int_c^d f(x)dx,$$

and if

$$\int_c^d f(x)dx = 0$$

for all $a \leqslant c < d \leqslant b$ then $f(x)$ is identically zero on (a, b).

Another point is this: the $F(\phi)$ which are defined by point-functions, and all derivatives of such (the $F^{(n)}(\phi)$) have a certain continuity property and it is desirable to require this continuity property in defining distributions.[5]

[5]As we shall show in §5, this gives the smallest enlargement or completion of the system of locally summable point-functions which permits unrestricted derivation.

Finally, with a view to later applications, we shall define distributions on open intervals. The closed interval gives the mathematically simpler situation and indeed the open interval will be discussed in terms of its closed subintervals. However, we shall reserve the word "distribution" for the open interval and use the terminology "continuous linear functional" for the closed interval.

Our precise definitions now follow.

Definition of distribution. For any closed finite interval (a, b) let $S_{(a,b)}$ consist of all continuous $\phi(x)$ possessing derivatives $\phi^{(n)}(x)$ of all orders which, along with $\phi(x)$ itself, vanish at a and at b and for x outside (a, b). A linear functional $F(\phi)$, defined for all ϕ in $S_{(a,b)}$, is called a c.l.f. on (a, b) if it has the following continuity property; whenever all ϕ, ϕ_m are in $S_{(a,b)}$ and the $\phi_m(x)$ converge uniformly to $\phi(x)$, and, for each n, the derivatives $\phi_m{}^{(n)}(x)$ converge uniformly to $\phi^{(n)}(x)$, then $F(\phi_m)$ shall converge to $F(\phi)$.

For an arbitrary open interval I, finite or infinite, a distribution on I is a linear functional $F(\phi)$ such that for every closed finite interval (a, b) contained in I, $F(\phi)$ is defined for all ϕ in $S_{(a,b)}$ and, when restricted to these ϕ, defines a c.l.f. on (a, b).

Identification of distribution and point-function. A distribution $F(\phi)$ on an interval I is to be identified with a point-function $f(x)$ if for every closed finite interval (a, b) contained in I, $f(x)$ is summable on (a, b) and

$$F(\phi) = \int_a^b f(x)\phi(x)dx$$

for all ϕ in $S_{(a,b)}$. We shall sometimes use the notation $f(\phi)$ to denote the distribution identified with the point-function $f(x)$.

Identification of distribution and measure function. A distribution $F(\phi)$ on an interval I is to be identified with the Stieltjes measure $d\psi(x)$ if for every closed finite interval (a, b) contained in I, $\psi(x)$ is of bounded variation on (a, b) and

$$F(\phi) = \int_a^b \phi(x)d\psi(x)$$

for all ϕ in $S_{(a,b)}$.

Definition of derivative of a distribution. For any distribution $F(\phi)$ on I, the derivative-distribution $F'(\phi)$ is defined by

$$F'(\phi) = -F(\phi').$$

It is easily verified that this F' satisfies the conditions for a distribution on I. (The same formula defines derivative for a c.l.f.)

The distributions defined above include all continuous and even all Lebesgue (locally) summable point-functions, all Stieltjes measures, and, as we shall see, a variety of new mathematical entities. Within the system of distributions each distribution has a derivative and consequently, derivatives of all orders:

$$F^{(n)}(\phi) = (-1)^n F(\phi^{(n)}).$$

The derivative of a point-function $f(x)$ may be a point-function or a Stieltjes measure or a more general distribution. It will be a point-function $g(x)$ if and only if $f(x)$ is absolutely continuous on every finite closed (a, b) contained in I and then $g(x)$ is the ordinary point-derivative $f'(x)$ which then must be defined (almost everywhere). The derivative of $f(x)$ is a Stieltjes measure $d\psi(x)$ if and only if $f(x)$ is of bounded variation on every finite closed (a, b) contained in I and then $\psi(x)$ differs from $f(x)$ by an additive (arbitrary) constant.

In particular, the derivative δ of the Heaviside function $Y(x)$ is now rigorously defined (I being any open interval which contains the origin) and is the measure function $dY(x)$; this is the proper mathematical description of the Dirac delta-function which has no meaning as a point-function. We may now use the notation: $\delta = Y'$.

In applications to certain physical problems a locally summable $f(x)$ may be thought of as representing a distribution of mass or electric charge on the axis of x:

$$\int_c^d f(x)dx$$

is then the total (algebraic) charge on (c, d) and $f(x)$ is the density of charge at a particular x. From this point of view the Dirac δ represents the concentration of unit charge at a single point, the origin, δ' represents a dipole and higher derivatives of δ represent more complicated multiple-layers.

An interesting distribution which is quite different from the Dirac δ and its derivatives is this: let $f(x)$ have the value $x^{-\frac{1}{2}}$ for positive x and vanish for all other x. Then its derivative f' exists (as a distribution) although it is not quite a Stieltjes measure. Roughly speaking, f' corresponds to negative mass continuously distributed on the positive x axis with an infinite quantity in every neighbourhood of the origin, together with an infinite positive mass at the origin, in such a way that there is finite total algebraic mass on every finite closed interval. This is shown by the formula, for ϕ in $S_{(a,b)}$ with $a < 0 < b$:

$$f'(\phi) = -\int_a^b f(x)\phi'(x)dx = -\int_0^b f(x)\phi'(x)dx$$

$$= -\lim_{\epsilon\to 0}\int_\epsilon^b x^{-\frac{1}{2}}\phi'(x)dx$$

$$= -\lim_{\epsilon\to 0}\left[(x^{-\frac{1}{2}}\phi(x))_\epsilon^b + \int_\epsilon^b \tfrac{1}{2}x^{-3/2}\phi(x)dx\right]$$

$$= \lim_{\epsilon\to 0}\left[\frac{\phi(\epsilon)}{\sqrt{\epsilon}} + \int_\epsilon^b (-\tfrac{1}{2}x^{-3/2})\phi(x)dx\right]$$

$$= \lim_{\epsilon\to 0}\left[\frac{\phi(0)}{\sqrt{\epsilon}} + \int_\epsilon^b (-\tfrac{1}{2}x^{-3/2})\phi(x)dx\right]$$

since

$$\frac{\phi(\epsilon)-\phi(0)}{\sqrt{\epsilon}} = \sqrt{\epsilon}\,\frac{\phi(\epsilon)-\phi(0)}{\epsilon} \to 0.\phi'(0) = 0 \text{ as } \epsilon \to 0.$$

Although $-\frac{1}{2}x^{-3/2}\phi(x)$ is not summable on $(0, b)$ for arbitrary ϕ in $S_{(a,b)}$, yet the bracket as a whole always has a finite limit, which has been called by Hadamard the "finite part" of the divergent integral. We use the notation

$$f'(\phi) = \text{Fp.}\int_0^b f'(x)\phi(x)dx$$

where $f'(x)$ is the non-summable point-derivative of $f(x)$ for positive x. Finite parts of divergent integrals have been studied in great detail by Hadamard.[6]

Similarly, the linear functional

$$F(\phi) = \lim_{\epsilon\to 0}\left[\left(\int_a^{-\epsilon} + \int_\epsilon^b\right)\frac{1}{x}\phi(x)dx\right]$$

corresponds to a continuously distributed mass with infinite positive mass along the positive x axis together with infinite negative mass along the negative x axis in such a way that in every neighbourhood of the origin there is finite total algebraic mass. The limit of the bracket is another case of an Hadamard finite part of a divergent integral, in this case coinciding with the Cauchy principal value. The distribution itself is the derivative of the point-function $\log |x|$.

[6] J. Hadamard, *Le problème de Cauchy et les équations aux dérivées partielles linéaires hyperboliques*, Paris, Hermann et Cie, 1932.

§3. THE CALCULUS OF DISTRIBUTIONS

MULTIPLICATION of a distribution by a constant and addition of two distributions are defined by the formulae:

$$(cF)(\phi) = cF(\phi), \qquad (F_1 + F_2)(\phi) = F_1(\phi) + F_2(\phi).$$

It is easily verified that the usual rules of addition and subtraction hold, even when derivation is involved. Thus

$$(c_1F_1 + c_2F_2)' = c_1F_1' + c_2F_2'.$$

The point-function which is a constant, $f(x) = k$, is identified with constant-distribution $k(\phi)$ which has the characteristic property:

$$k(\phi) = \int_a^b k\phi(t)dt = k\int_a^b \phi(t)dt$$

for every ϕ in $S_{(a,b)}$. In particular, when $k = 0$ the corresponding distribution is called the zero-distribution. It is easily verified that if F is a constant-distribution F' is the zero-distribution. To prove the converse, it is useful to establish the following *expansion* lemma.

Lemma. *Let $\theta(x)$ be any function in $S_{(c,d)}$ for which*

$$\int_c^d \theta(x)dx = 1$$

(for instance $\theta(x)$ could be

$$\left\{\int_c^d \phi_{c,d}(t)dt\right\}^{-1} \phi_{c,d}(x)$$

where $\phi_{c,d}(x)$ is the function defined in §2). Let n be any positive integer. Then any $\phi(x)$ in any $S_{(a,b)}$ with $a \leqslant c < d \leqslant b$ can be expressed in the form:

$$\phi(x) = a_0\theta(x) + a_1\theta'(x) + \ldots + a_n\theta^{(n)}(x) + \rho_n^{(n+1)}(x)$$

where $a_0, a_1, \ldots, a_n$ are uniquely determined constants and $\rho_n(x)$ is in $S_{(a,b)}$.

This lemma can be proved by induction on n if we observe that a particular $\phi(x)$ in $S_{(a,b)}$ can be expressed in the form $\rho'(x)$ for some $\rho(x)$ in $S_{(a,b)}$ if and only if

$$\int_a^b \phi(t)dt = 0.$$

In the expansion for a general $\phi(x)$ in $S_{(a,b)}$,

$$a_0 = \int_a^b \phi(t)dt.$$

Now using this representation with $n = 1$, we argue that if $F' = 0$ then $F(\rho_1') = -F'(\rho_1) = 0$ for all ρ_1, and hence

$$F(\phi) = F(a_0\theta + \rho_1') = a_0F(\theta) + F(\rho_1') = a_0k$$

where k is a constant, the value of $F(\theta)$. Hence

$$F(\phi) = \int_a^b k\phi(t)dt,$$

proving that F is indeed a constant-distribution.

We shall say that the distribution G is a *primitive* of F if $G' = F$. From the preceding paragraph it follows that two primitives of the same F differ by a constant-distribution.

As is well known, in the case of point-functions there exists an infinity of different (point-function) primitives and in order to specify one of them it is sufficient to give its value at some particular point. In the case of an arbitrary distribution F there exists again an infinity of primitives (distributions!) and a particular one can be specified by giving its value for any particular testing-function $\theta(x)$ as described above; this follows immediately from the relationship:

$$G(\phi) = G(a_0\theta + \rho_1') = a_0G(\theta) + G(\rho_1') = a_0G(\theta) - F(\rho_1)$$

which can be used to define $G(\phi)$ for every ϕ in terms of an arbitrary (but fixed) $G(\theta)$, and given F; that this G is a distribution can be deduced from the relation

$$\rho_0(x) = \int_a^x \phi(t)dt - \left\{\int_a^b \phi(t)dt\right\}\int_a^x \theta(t)dt.$$

§4. MULTIPLICATION OF DISTRIBUTIONS

THE product F_1F_2 is *not* defined for arbitrary distributions. This reflects the fact that the product $f_1(x)f_2(x)$ of two locally summable point-functions may not be locally summable.

However, we do define F_1F_2 in certain cases. For example, if F_1, F_2 can be identified with $f_1(x), f_2(x)$ respectively and the product $f_1(x)f_2(x)$ is locally summable then F_1F_2 is defined to be the distribution which is identified with $f_1(x)f_2(x)$. This is a special case of the following general definition.

Definition of product of two c.l.f.'s on a finite closed interval. Suppose for some n that $F_1^{(n)}$ is a point-function $f_1(x)$ and that F_2 is the nth derivative of a point-function $f_2(x)$:

$$F_1^{(n)} = f_1, \quad F_2 = f_2^{(n)}.$$

Suppose, too, that the product $f_1(x)f_2(x)$ is summable. Then we define F_1F_2 by the formula:

$$F_1F_2 = F_1f_2^{(n)} = (F_1f_2)^{(n)} - \binom{n}{1}(F_1'f_2)^{(n-1)} + \binom{n}{2}(F_1^{(2)}f_2)^{(n-2)}$$
$$+ \dots + (-1)^r\binom{n}{r}(F_1^{(r)}f_2)^{(n-r)} + \dots + (-1)^nF_1^{(n)}f_2.$$

Each term on the right is a c.l.f. since $F_1^{(n)}f_2$ is $f_1(x)f_2(x)$ which was assumed summable and for $r < n$, $F_1^{(r)}f_2$ is the product of an absolutely continuous (hence bounded) point-function and f_2.

Definition of product of distributions on an open interval I. Let $F_{(a,b)}$ denote the distribution F restricted to the testing functions in $S_{(a,b)}$. If now F_1, F_2 are distributions on I such that $F_{1(a,b)}F_{2(a,b)}$ is defined on every (a, b) contained in I, then F_1F_2 is defined to be the distribution on I which, when restricted to any $S_{(a,b)}$, coincides with $F_{1(a,b)}F_{2(a,b)}$.

It is not difficult to verify that these definitions give a unique F_1F_2 whenever they define F_1F_2 at all and that our formula above for the product F_1F_2 is equivalent to:

$$(F_1F_2)(\phi) = (-1)^n\int_a^b f_2(x)(F_1(x)\phi(x))^{(n)}dx.$$

It can also be verified that the product law $(F_1F_2)' = F_1'F_2 + F_1F_2'$ is valid and that the three products which occur in this statement are necessarily defined whenever one of the products on the right is defined.

There is a special but most important case in which our product rule can be greatly simplified. Suppose that F_1 is a continuous point-function $\alpha(x)$ possessing point-function derivatives of all orders so that $F_1^{(n)}$ is a continuous point-function for each n. Then our product rule defines F_1F_2 for every F_2 which, when restricted to an $S_{(a,b)}$, can be put in the form $f_2^{(n)}$ for some f_2 and n which might depend on a and b. Our rule simplifies in this case to:

$$(F_1F_2)(\phi) = (\alpha F_2)(\phi) = F_2(\alpha\phi)$$

for every testing function ϕ. (Observe that $\alpha\phi$ is a testing function along with ϕ.)

The preceding paragraph suggests that we could use the relation

$$(\alpha F_2)(\phi) = F_2(\alpha\phi)$$

to *define* αF_2 whenever $\alpha(x)$ has (ordinary) derivatives of all orders but F_2 is an *arbitrary* distribution. It is noteworthy that such a definition would not give anything new since *every* distribution F_2 when restricted to an $S_{(a,b)}$, can be put in the form $f_2^{(n)}$ for some point-function $f_2(x)$ and some n, which may depend on a and b. This theorem will be proved in the next section and it is a consequence of the continuity condition in our definition of distribution.

We note that the original definition of cF, where c is a constant, is included in the general definition of product of distributions if c is considered as a constant-distribution.

We note also that when the Dirac δ and its derivatives $\delta^{(n)}$ are multiplied by an $\alpha(x)$ with derivatives $\alpha^{(n)}(x)$ of all orders, we obtain:

$$\alpha(x)\delta = \alpha(0)\delta,$$

$$\alpha(x)\delta = (\alpha\delta)' - \alpha'\delta = \alpha(0)\delta' - \alpha'(0)\delta,$$

and in general,

$$\alpha(x)\delta^{(n)} = \alpha(0)\delta^{(n)} - n\alpha'(0)\delta^{(n-1)} + \binom{n}{2}\alpha''(0)\delta^{(n-2)} + \ldots + (-1)^n\alpha^{(n)}(0)\delta.$$

§5. THE ORDER CLASSIFICATION OF DISTRIBUTIONS

THE system of c.l.f.'s on a given closed, finite interval includes, of course, every summable $f(x)$ and all its derivatives. We shall show in this section that *there are no other c.l.f.'s*, that is, every c.l.f. is either $f(\phi)$ or $f^{(n)}(\phi)$ for some suitable summable $f(x)$ and some finite n. (By using a higher n we can prove this with $f(x)$ continuous or even absolutely continuous.)

Let F be a c.l.f. on (a, b). F will be said to have *finite order* on (a, b) either if it can be identified with a summable point-function $f(x)$ or if, for some finite r, F is the rth order derivative of some such $f(x)$: $F = f^{(r)}$. The smallest possible r will be called the order of F.[7]

Clearly if F can be identified with a summable $f(x)$ then it has order 0 and its derivative F' has order either 0 or 1. If F has order r greater than 0 its derivative has order $r + 1$ precisely.

If F has order r and $s \geqslant r$ then for some f which depends on F and s, $F = f^{(s)}$; if $s > r$, f is absolutely continuous; if $s = 0$, f is uniquely determined to within Lebesgue equivalence, but if $s > 0$, f is determined only to within an additive (arbitrary) polynomial in x of degree $s - 1$.

We wish to show that F must have finite order. The definition of c.l.f. requires that $F(\phi_m)$ shall converge to $F(\phi)$ whenever all ϕ, ϕ_m are in $S_{(a,b)}$ and, for each $n \geqslant 0$, the $\phi_m^{(n)}(x)$ converge uniformly to $\phi^{(n)}(x)$. We shall now show that this implies the following apparently stronger condition. For some finite r which depends on F,

(C_r) $F(\phi_m)$ converges to $F(\phi)$ whenever all ϕ, ϕ_m are in $S_{(a,b)}$ and, for all n with $0 \leqslant n \leqslant r$, the $\phi_m^{(n)}(x)$ converge uniformly to $\phi^{(n)}(x)$.

Indeed, suppose if possible, that (C_r) is false for every r. We can then define a sequence of testing functions ϕ_m such that for each m,

(i) $|\phi_m^{(n)}| < 2^{-m}$ for all $n \leqslant m$,

(ii) $F(\phi_m) > 1$.

(We are using the notation $|\phi| = \max\,[|\phi(x)|;\, a \leqslant x \leqslant b]$.) Then for every n, the $\phi_m^{(n)}(x)$ converge uniformly to the zero function and since F is a c.l.f.,

[7] In *Théorie des Distributions* I, p. 25, a slightly different definition is used: the order of F is defined there to be the least r for which $F = F_0^{(r)}$ with F_0 a Stieltjes measure.

this should imply that $F(\phi_m) \to 0$. But this contradicts (ii) above and so (C_r) can not be false for every r, that is, (C_r) holds for some finite r.

But for $n < r$,

$$\phi^{(n)}(x) = \int_a^x \frac{(x-t)^{r-n-1}}{(r-n-1)!} \phi^{(r)}(t)dt$$

so that $|\phi^{(n)}| \leqslant K|\phi^{(r)}|$ for all $n < r$, for some finite K which depends only on r, a and b. It follows that the condition (C_r) is equivalent to the condition

(B_r) $F(\phi_m)$ converges to $F(\phi)$ whenever all ϕ, ϕ_m are in $S_{(a,b)}$ and the ${\phi_m}^{(r)}(x)$ converge uniformly to $\phi^{(r)}(x)$.

We now show that the condition (B_r) implies the condition

(A_r) $|F(\phi)| \leqslant |F|_r |\phi^{(r)}|$

for all ϕ in $S_{(a,b)}$ with $|F|_r$ a finite constant (we shall let $|F|_r$ actually denote the smallest possible such constant).

Indeed if (A_r) were false we could define a sequence ϕ_m with $F(\phi_m) > m|{\phi_m}^{(r)}|$. Then the functions $\mu_m(x) = |{\phi_m}^{(r)}|^{-1}m^{-1}\phi_m(x)$ would be in $S_{(a,b)}$ and the ${\mu_m}^{(r)}(x)$ would converge uniformly to zero since $|{\mu_m}^{(r)}| = m^{-1}$. But $F(\mu_m) > 1$, contradicting (B_r). Thus (A_r) cannot be false.

To summarize our results: if F is a c.l.f. on (a, b) there is a finite r for which $|F|_r < \infty$ and

$$|F(\phi)| \leqslant |F|_r |\phi^{(r)}|$$

for all ϕ in $S_{(a,b)}$.

Now we construct a new functional L, which is defined for functions which can be put in the form $\phi^{(r)}$, by the formula $L(\phi^{(r)}) = F(\phi)$. The functional L is, by the preceding paragraph, a bounded linear functional on the linear space of functions of the form $\phi^{(r)}$ with norm $|\phi^{(r)}|$. The Hahn-Banach procedure[8] can be used to extend this functional L to all continuous functions on (a, b) without increasing the bound of L. Then the representation theorem of F. Riesz[9] applies and shows that

$$L(\phi^{(r)}) = \int_a^b \phi^{(r)}(x)d\psi(x)$$

for some $\psi(x)$ of total variation equal to $|F|_r$ and $\psi(x)$ can be assumed to satisfy $|\psi(x)| \leqslant |F|_r$. Thus

[8]See S. Banach, *Théorie des Opérations Linéaires*, Warsaw, 1932, Théorème 2, p. 55
[9]See *Théorie des Opérations Linéaires*, pp. 59-61.