Robert S Strichartz

A Guide to Distribution Theory and Fourier Transforms

World Scientific

A Guide to Distribution Theory and Fourier Transforms

Robert S Strichartz
Cornell University, USA

World Scientific
New Jersey • London • Singapore • Hong Kong

Published by

World Scientific Publishing Co. Pte. Ltd.
5 Toh Tuck Link, Singapore 596224
USA office: Suite 202, 1060 Main Street, River Edge, NJ 07661
UK office: 57 Shelton Street, Covent Garden, London WC2H 9HE

British Library Cataloguing-in-Publication Data
A catalogue record for this book is available from the British Library.

First published in 1994 by CRC Press

A GUIDE TO DISTRIBUTION THEORY AND FOURIER TRANSFORMS

ISBN 981-238-421-9
ISBN 981-238-430-8 (pbk)

Printed in Singapore by World Scientific Printers (S) Pte Ltd

A Guide to Distribution Theory and Fourier Transforms

Preface

Distribution theory was one of the two great revolutions in mathematical analysis in the 20th century. It can be thought of as the completion of differential calculus, just as the other great revolution, measure theory, (or Lebesgue integration theory), can be thought of as the completion of integral calculus. There are many parallels between the two revolutions. Both were created by young, highly individualistic French mathematicians (Henri Lebesgue and Laurent Schwartz). Both were rapidly assimilated by the mathematical community, and opened up new worlds of mathematical development. Both forced a complete rethinking of all mathematical analysis that had come before, and basically altered the nature of the questions that mathematical analysts asked. (This is the reason I feel justified in using the word "revolution" to describe them). But there are also differences. When Lebesgue introduced measure theory (circa 1903), it almost came like a bolt from the blue. Although the older integration theory of Riemann was incomplete—there were many functions that did not have integrals—it was almost impossible to detect this incompleteness from within, because the non-integrable functions really appeared to have no well defined integral. As evidence that the mathematical community felt perfectly comfortable with Riemann's integration theory, one can look at Hilbert's famous list (dating to 1900) of 23 unsolved problems that he thought would shape the direction of mathematical research in the 20th century. Nowhere is there a hint that completing integration theory was a worthwhile goal. On the other hand, a number of his problems do foreshadow the developments that led to distribution theory (circa 1945). When Laurent Schwartz came out with his theory, he addressed problems that were of current interest, and he was able to replace a number of more complicated theories that had been developed earlier in an attempt to deal with the same issues.

From the point of view of this work, the most important difference is that in retrospect, measure theory still looks hard, but distribution theory looks easy. Because it is relatively easy, distribution theory should be accessible to a wide audience, including users of mathematics and mathematicians who specialize in other fields. The techniques of distribution theory can be used, confidently and effectively—just like the techniques of calculus are used—without a complete knowledge of the formal mathematical foundations of the subject. The aim of this book is thus very similar to the aim of a typical calculus textbook: to explain the techniques of the theory with precision, to provide an intuitive discussion of the ideas that underline the techniques, and to offer a selection of problems applying the techniques.

Because the Lebesgue theory of integration preceded distribution theory historically, and is required for the rigorous mathematical development of the theory, it might be thought that a knowledge of the Lebesgue theory would have to be a prerequisite for studying distribution theory. I do not believe that this is true, and I hope this book makes a good case for my point of view. When you see an integral sign in this book, you are free to interpret it in the sense of any integration theory you have learned. If you have studied the Lebesgue theory in any form, then of course think of the integrals as Lebesgue integrals. But if not, don't worry about it. Let the integral mean what you think it means.

Distribution theory is a powerful tool, but it becomes an even more powerful tool when it works in conjunction with the theory of Fourier transforms. One of the main areas of applications is to the theory of partial differential equations. These three theories form the main themes of this book. The first two chapters motivate and introduce the basic concepts and computational techniques of distribution theory. Chapters three and four do the same for Fourier transforms. Chapter five gives some important and substantial applications to particular partial differential equations that arise in mathematical physics. These five chapters, part I of the book, were written with the goal of getting to the point as quickly as possible. They have been used as a text for a portion of a course in applied mathematics at Cornell University for more than 10 years.

The last three chapters, part II of the book, return to the three themes in greater detail, filling in topics that were left aside in the rapid development of part I, but which are of great interest in and of themselves, and point toward further applications. Chapter six returns to distribution theory, explaining the notion of continuity, and giving the important structure theorems. Chapter seven covers Fourier analysis. In addition to standard material, I have included some topics of recent origin, such as quasicrystals and wavelets. Finally, Chapter eight returns to partial differential equations, giving an introduction to the modern theory of general linear equations. Here the reader will meet Sobolev spaces, a priori estimates, equations of elliptic and hyperbolic type, pseudodifferential operators, wave front sets, and the ideology known as microlocal analysis. Part II was written for this book, and deals with somewhat more abstract material. It was not designed for use as a textbook, but more to satisfy the curiosity of those readers of part I who want to learn in greater depth about the material. I also hope it will serve as an appetizer for readers who will go on to study these topics in greater detail.

The prerequisites for reading this book are multidimensional calculus and an introduction to complex analysis. A reader who has not seen any complex analysis will be able to get something out of this book, but will have to accept that there will be a few mystifying passages. A solid background in multi-dimensional calculus is essential, however, especially in part II.

Recently, when I was shopping at one of my favorite markets, I met a graduate of Cornell (who had not been in any of my courses). He asked me what I was doing, and when I said I was writing this book, he asked sarcastically "do you guys enjoy writing them as much as we enjoy reading them?" I don't know what other books he had in mind, but in this case I can say quite honestly that I very much enjoyed writing it. I hope you enjoy reading it.

Acknowledgments:

I am grateful to John Hubbard, Steve Krantz and Wayne Yuhasz for encouraging me to write this book, and to June Meyermann for the excellent job of typesetting the book in LaTeX.

Ithaca, NY
October 1993

Contents

Chapter 1

What are Distributions?

1.1 Generalized functions and test functions

You have often been asked to consider a function $f(x)$ as representing the value of a physical variable at a particular point x in space (or space-time). But is this a realistic thing to do? Let us borrow a perspective from quantum theory and ask: What can you measure?

Suppose that $f(x)$ represents temperature at a point x in a room (or if you prefer let $f(x,t)$ be temperature at point x and time t.) You can measure temperature with a thermometer, placing the bulb of the thermometer at the point x. Unlike the point, the bulb of the thermometer has a nonzero size, so what you measure is more an average temperature over a small region of space (again if you think of temperature as varying with time also, then you are also averaging over a small time interval preceeding the time t when you actually read the thermometer). Now there is no reason to believe that the average is "fair" or "unbiased." In mathematical terms, a thermometer measures

$$\int f(x)\varphi(x)\,dx$$

where $\varphi(x)$ depends on the nature of the thermometer and where you place it—$\varphi(x)$ will tend to be "concentrated" near the location of the thermometer bulb and will be nearly zero once you are sufficiently far away from the bulb. To say this is an "average" is to require

$$\varphi(x) \geq 0 \text{ everywhere, and}$$
$$\int \varphi(x)\,dx = 1 \text{ (the integral is taken over all space).}$$

However, do not let these conditions distract you. With two thermometers you can measure

$$\int f(x)\varphi_1(x)\,dx \qquad \text{and} \qquad \int f(x)\varphi_2(x)\,dx$$

and by subtracting you can deduce the value of $\int f(x)[\varphi_1(x) - \varphi_2(x)]\,dx$. Note that $\varphi_1(x)-\varphi_2(x)$ is no longer nonnegative. By doing more arithmetic you can even compute $\int f(x)(a_1\varphi_1(x) - a_2\varphi_2(x))\,dx$ for constants a_1 and a_2, and $a_1\varphi_1(x) - a_2\varphi_2(x)$ may have any finite value for its integral.

The above discussion is meant to convince you that it is often more meaningful physically to discuss quantities like $\int f(x)\varphi(x)\,dx$ than the value of f at a particular point x. The secret of successful mathematics is to eliminate all unnecessary and irrelevant information—a mathematician would not ask what color is the thermometer (neither would an engineer, I hope). Since we have decided that the value of f at x is essentially impossible to measure, let's stop requiring our functions to have a value at x. That means we are considering a larger class of objects. Call them *generalized functions.* What we will require of a generalized function is that something akin to $\int f(x)\varphi(x)\,dx$ exist for a suitable choice of averaging functions φ (call them *test functions*). Let's write $\langle f, \varphi\rangle$ for this something. It should be a real number (or a complex number if we wish to consider complex-valued test functions and generalized functions). What other properties do we want? Let's recall some arithmetic we did before, namely

$$a_1 \int f(x)\varphi_1(x)\,dx - a_2 \int f(x)\varphi_2(x)\,dx = \int f(x)(a_1\varphi_1(x) - a_2\varphi_2(x))\,dx.$$

We want to be able to do the same sort of thing with generalized functions, so we should require $a_1\langle f, \varphi_1\rangle - a_2\langle f, \varphi_2\rangle = \langle f, a_1\varphi_1 - a_2\varphi_2\rangle$. This property is called *linearity.* The minus sign is a bit silly, since we can get rid of it by replacing a_2 by $-a_2$. Doing this we obtain the condition

$$a_1\langle f, \varphi_1\rangle + a_2\langle f, \varphi_2\rangle = \langle f, a_1\varphi_1 - a_2\varphi_2\rangle.$$

Notice we have tacitly assumed that if φ_1, φ_2 are test functions then $a_1\varphi_1 + a_2\varphi_2$ is also a test function. I hope these conditions look familiar to you—if not, please read the introductory chapter of any book on linear algebra.

You have almost seen the entire definition of *generalized functions.* All you are lacking is a description of what constitutes a test function and one technical hypothesis of continuity. Do not worry about continuity—it will always be satisfied by anything you can *construct* (wise-guys who like using the axiom of choice will have to worry about it, along with wolves under the bed, etc).

So, now, what are the test functions? There are actually many possible choices for the collection of test functions, leading to many different theories of generalized functions. I will describe the space called $\mathcal{D}$, leading to the theory of distributions. Later we will meet other spaces of test functions.

The underlying point set will be an n-variable space $\mathbb{R}^n$ (points x stand for $x = (x_1, \dots, x_n)$) or even a subset $\Omega \subset \mathbb{R}^n$ that is open. Recall that this means every point $x \in \Omega$ is surrounded by a ball $\{y : |x - y| < \epsilon\}$ contained in Ω, where ϵ depends on x, and

$$|x - y| = \sqrt{(x_1 - y_1)^2 + \cdots + (x_n - y_n)^2}.$$

Of course $\Omega = \mathbb{R}^n$ is open, as is every open ball $\{y : |x - y| < r\}$. Intuitively, an open set is just a union of open balls.

The class of test functions $\mathcal{D}(\Omega)$ consists of *all functions $\varphi(x)$ defined in Ω, vanishing outside a bounded subset of Ω that stays away from the boundary of Ω, and such that all partial derivatives of all orders of φ are continuous.*

For example, if $n = 1, \Omega = \{0 < x < 1\}$, then

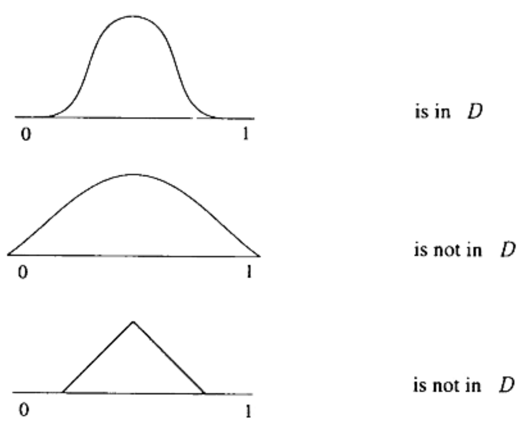

Figure 1.1

The second example fails because it does not vanish near the boundary point 0, and the third example fails because it is not differentiable at three points. To actually write down a formula for a function in $\mathcal{D}$ is more difficult. Notice that no analytic function (other than $\varphi \equiv 0$) can be in $\mathcal{D}$ because of the vanishing requirement. Thus any formula for φ must be given "in pieces." For example, in $\mathbb{R}^1$

$$\psi(x) = \begin{cases} e^{-1/x^2} & x > 0 \\ 0 & x \leq 0 \end{cases}$$

has continuous derivatives of all order:

$$\left(\frac{d}{dx}\right)^k e^{-1/x^2} = \frac{\text{polynomial in } x}{\text{polynomial in } x} e^{-1/x^2}$$

and as $x \to 0$ this approaches zero since the zero of e^{-1/x^2} beats out the pole of $\frac{1}{\text{polynomial in } x}$. Thus $\varphi(x) = \psi(x)\psi(1-x)$ has continuous derivatives of all orders (we abbreviate this by saying φ is C^∞) and vanishes outside $0 < x < 1$, so $\varphi \in \mathcal{D}(\mathbb{R}^1)$; in fact, $\varphi \in \mathcal{D}(a < x < b)$ provided $a < 0$ and $b > 1$ (why not $a \leq 0$ and $b \geq 1$?). Once you have one example you can manufacture more by

1. moving it about $\varphi(x + x_0)$
2. changing vertical scale $a\varphi(x)$
3. changing horizontal scale $\varphi(ax)$
4. taking linear combinations $a_1\varphi_1(x) + a_2\varphi_2(x)$ if $\varphi_1, \varphi_2 \in \mathcal{D}$.
5. taking products $\varphi_1(x_1, \ldots, x_n) = \varphi(x_1)\varphi(x_2)\ldots\varphi(x_n)$ to obtain examples in higher dimensions.

Exercise: Draw the pictures associated with operations 1–4.

These considerations should convince you that you can make a test function in $\mathcal{D}$ do anything you can draw a picture of a smooth function doing. You can make it take on prescribed values at a finite set of points, make it vanish on any open set and even take on a constant value (say 1) on a bounded open set away from the boundary (this requires a little more work).

O.K. That is what $\mathcal{D}(\Omega)$ is. We can now define the class of distributions on Ω, denoted $\mathcal{D}'(\Omega)$, to be all continuous *linear functionals* on $\mathcal{D}(\Omega)$. By *functional* I mean a real (or complex) -valued function on $\mathcal{D}(\Omega)$, written $\langle f, \varphi \rangle$. By *linear* I mean it satisfies the identity

$$a_1\langle f, \varphi_1 \rangle + a_2\langle f, \varphi_2 \rangle = \langle f, a_1\varphi_1 + a_2\varphi_2 \rangle.$$

(Yes, Virginia, $a_1\varphi_1 + a_2\varphi_2$ is in $\mathcal{D}(\Omega)$ if φ_1 and φ_2 are in $\mathcal{D}(\Omega)$.) By *continuous* I mean that if φ_1 is close enough to φ then $\langle f, \varphi_1\rangle$ is close to $\langle f, \varphi\rangle$—the exact definition can wait until later. Continuity has an intuitive physical interpretation—you want to be sure that different thermometers give approximately the same reading provided you control the manufacturing process adequately. Put another way, when you repeat an experiment you do not want to get a different answer because small experimental errors get magnified. Now, whereas discontinuous *functions* abound, *linear functionals* all tend to be continuous. This happy fact deserves a bit of explanation. Fix φ and φ_1 and call the difference $\varphi_1 - \varphi = \varphi_2$. Then $\varphi_1 = \varphi + \varphi_2$. Now perhaps $\langle f, \varphi\rangle$ and $\langle f, \varphi_1\rangle$ are far apart. So what? Move φ_1 closer to φ by considering $\varphi + t\varphi_2$ and let t get small. Then $\langle f, \varphi + t\varphi_2\rangle = \langle f, \varphi\rangle + t\langle f, \varphi_2\rangle$ by linearity, and as t gets small this gets close to $\langle f, \varphi\rangle$. This does *not* constitute a proof of continuity, since the definition requires more "uniformity," but it should indicate that a certain amount of continuity is built into linearity. At any rate, all linear functionals on $\mathcal{D}(\Omega)$ you will ever encounter will be continuous.

1.2 Examples of distributions

Now for some examples. Any function gives rise to a distribution by setting $\langle f, \varphi\rangle = \int_\Omega f(x)\varphi(x)\,dx$, at least if the integral can be defined. This is certainly true if f is continuous, but actually more general functions will work. Depending on what theory of integration you are using, you may make f discontinuous and even unbounded, provided the improper integral converges *absolutely*. For instance $\int_{|x|\le r} |x|^{-t}\,dx$ converges for $t < n$ (in n dimensions) and diverges for $t \ge n$. Thus the function $|x|^{-t}$ for $t < n$ gives rise to the distribution in $\mathcal{D}'(\mathbb{R}^n)\langle f, \varphi\rangle = \int_{\mathbb{R}^n} \varphi(x)|x|^{-t}\,dx$ (the actual range of integration is bounded since φ vanishes outside a bounded set).

A different sort of example is the Dirac δ-function: $\langle \delta, \varphi\rangle = \varphi(0)$. In this case we have to check the linearity property, but it is trivial to verify:

$$a_1\langle \delta, \varphi_1\rangle + a_2\langle \delta, \varphi_2\rangle = a_1\varphi_1(0) + a_2\varphi_2(0) = \langle \delta, a_1\varphi_1 + a_2\varphi_2\rangle.$$

An even wilder example is δ' (now in $\mathcal{D}'(\mathbb{R}^1)$) given by $\langle \delta', \varphi\rangle = -\varphi'(0)$.

Exercise: Verify linearity.

These examples demand a closer look, some pictures, and an explanation of the minus sign. Consider any function $f_k(x)$ (for simplicity we work in one dimension) that satisfies the conditions

1. $f_k(x) = 0$ unless $|x| \le \frac{1}{k}$

2. $\int_{-1/k}^{1/k} f_k(x)\,dx = 1$.

The simplest examples are

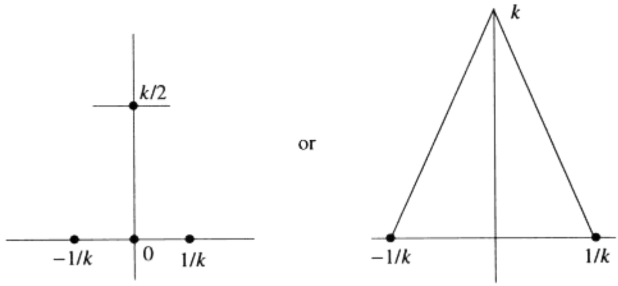

Figure 1.2

but we may want to take f_k smoother, even in $\mathcal{D}$

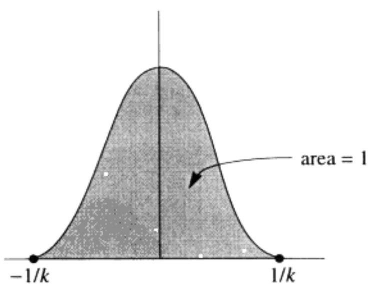

Figure 1.3

Now the distribution

$$\langle f_k, \varphi \rangle = \int f_k(x)\varphi(x)\,dx$$

is an average of φ near zero, so that if φ does not vary much in $-1/k \leq x \leq 1/k$, it is close to $\varphi(0)$. Certainly in the limit as $k \to \infty$ we get

$\langle f_k, \varphi\rangle \to \varphi(0) = \langle \delta, \varphi\rangle$. Thus we may think of δ as $\lim_{k\to\infty} f_k$. (Of course, pointwise

$$\lim_{k\to\infty} f_k(x) = \begin{cases} 0 & \text{if} \quad x \neq 0 \\ +\infty & \text{if} \quad x = 0 \end{cases}$$

for suitable choice of f_k, but this is nonsense, showing the futility of pointwise thinking.)

Now suppose we first differentiate f_k and then let $k \to \infty$?

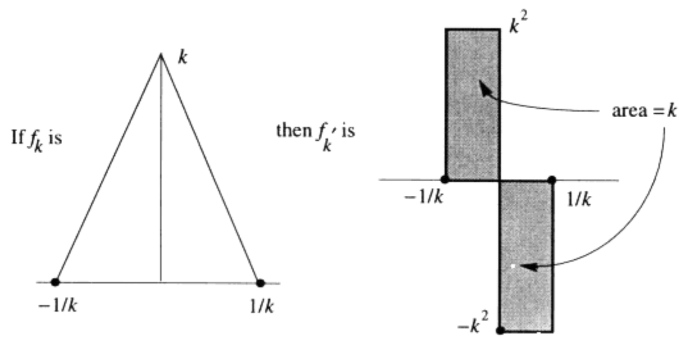

Figure 1.4

and

$$\langle f_k', \varphi\rangle \approx \frac{\varphi\left(-\frac{1}{2k}\right) - \varphi\left(\frac{1}{2k}\right)}{1/k}$$

(the points $-\frac{1}{2k}$ and $\frac{1}{2k}$ are the midpoints of the intervals and the factor $(1/k)^{-1} = k$ is the area) which approaches $-\varphi'(0)$ as $k \to \infty$. We obtain the same answer formally by integrating by parts:

$$\int f_k'(x)\varphi(x)\,dx = -\int f_k(x)\varphi'(x)\,dx = -\langle f_k, \varphi'\rangle \to -\varphi'(0)$$

as $k \to \infty$. Here we might assume that f_k is continuously differentiable. Note that there are no boundary terms in the integration-by-parts formula because φ vanishes for large values of x. Thus if $f_k \to \delta$ then $f_k' \to \delta'$, which justifies the notation of δ' as the derivative of δ.

1.3 What good are distributions?

Enough examples for now. We will have more later. Let us pause to consider the following question: Why this particular (perhaps peculiar) choice of test functions $\mathcal{D}$? The answer is not easy. In fact, the theory of generalized functions was in use for almost twenty years before Laurent Schwartz proposed the definition of $\mathcal{D}$. So it is not possible to use physical or intuitive grounds to say $\mathcal{D}$ is the only "natural" class of test functions. However, it does yield an elegant and useful theory—so that after the fact we may thank Laurant Schwartz for his brilliant insight. You can get some feel for what is going on if you observe that the smoother you require the test functions φ to be, the "rougher" you can allow the generalized functions f to be. To define $\langle\delta,\varphi\rangle$, φ must be at least continuous, and to define $\langle\delta',\varphi\rangle$, you must require φ to be differentiable. Later I will show you how to define derivatives for any distribution—the key point in the definition will be the ability to differentiate the test functions.

The requirement that test functions in $\mathcal{D}$ vanish outside a bounded set and near the boundary of Ω is less crucial. It allows distributions to "grow arbitrarily rapidly" as you approach the boundary (or infinity). Later we will consider a *smaller* class of distributions, called *tempered distributions*, which cannot grow as rapidly at infinity, by considering a *larger* class of test functions that have weaker vanishing properties.

Another question you should be asking at this point is: What good are distributions? Let me give a hint of one answer. Differential equations are used to construct models of reality. Sometimes the reality we are modeling suggests that some solutions of the differential equation need not be differentiable! For example, the "vibrating string" equation

$$\frac{\partial^2 u(x,t)}{\partial t^2} = k^2 \frac{\partial^2 u(x,t)}{\partial x^2}$$

has a solution $u(x,t) = f(x-kt)$ for any function of one variable f, which has the physical interpretation of a "traveling wave" with "shape" $f(x)$ moving at velocity k.

There is no physical reason for the "shape" to be differentiable, but if it is not, the differential equation is not satisfied at some points. But we do not want to throw away physically meaningful solutions because of technicalities. You might be tempted therefore to think that if a function satisfies a differential equation except for some points where it is not differentiable, it should be admitted as a solution. The next example shows that such a simplistic idea does not work.

Laplace's equation $\Delta u = 0$ where Δ (called the *Laplacian* and some-