Chapter 5

The Delta "Function" and Distributions in One Space Dimension

"*Thus, physicists lived in a fantastic universe which they knew how to manipulate admirably, and almost faultlessly, without ever being able to justify anything. They deserved the same reproaches as Heaviside: their computations were insane by standards of mathematical rigor, but they gave absolutely correct results, so one could think that a proper mathematical justification must exist. Almost no one looked for one, though.*
I believe I heard of the Dirac function for the first time in my second year at the ENS. I remember taking a course, together with my friend Marrot, which absolutely disgusted us, but it is true that those formulas were so crazy from the mathematical point of view that there was simply no question of accepting them. It didn't even seem possible to conceive of a justification. These reflections date back to 1935*, and in* 1944*, nine years later, I discovered distributions. This at least can be deduced from the whole story: it's a good thing that theoretical physicists do not wait for mathematical justifications before going ahead with their theories!*"

- **Laurent Schwartz**[a]

[a]**Laurent Schwartz (1915–2002)** was a French mathematician who made fundamental contributions to the mathematical analysis of PDEs, particularly the theory of distributions. While several aspects of the theory were previously presented by the great Russian mathematician **Sergei Sobolev (1908–1989)**, it was Schwartz whose full development led to his winning the Fields Medal in 1950. His published two-volume treatise is simply titled "*Théorie des Distributions*". This quote is taken from pages 217–218 of his book *A Mathematician Grappling with His Century*, Birkhäuser, 2001. The translation from the original French was by Leila Schneps.

Throughout most of this text we are interested in finding solutions to PDEs which are inarguably functions. However, one of the key objects central to studying a large class of PDEs is *something* which is actually not a function and which encapsulates a rather central phenomenon: **concentration**. This *object* was not officially introduced by a mathematician, but rather by the Nobel Prize winning physicist **Paul Dirac**[1]. Dirac clearly understood what this object was and, more importantly, how to use it. However, for us easily confused mortals, it is most unfortunate that it has come to be known as the **Dirac delta function** because as we shall see, it is **not** a function or, at least, not a classical function as we know them. In the 1940s, the mathematician **Laurent Schwartz** developed a beautiful mathematical theory which gave such objects precise meaning by generalizing the notion of a function. These **generalized functions** are called **distributions** and are the focus of this chapter. The *Theory of Distributions* is extensive and, incidentally, Schwartz received the Fields Medal ("the Nobel Prize of mathematics") in 1950, in part for the development and advancement of this theory.

In this text it is our hope to convince you, the reader, that the **basic** rudiments of the theory of distributions are accessible, digestible, and tremendously useful and insightful. To this end, we will primarily address the following two types of distributions:

- distributions associated with point concentrations (delta "functions"),
- distributions which are generated by classical functions with discontinuities

from two points of view: **differentiation** and **limits**.

Much of this chapter (cf. sections with bullets) is fundamental to material we will cover in the subsequent chapters. Spending time and effort to appreciate and digest the delta "function" and, further, what it means to interpret a function as a distribution will serve all readers well, regardless of which scientific discipline they endeavor to pursue.

We begin by recalling what exactly a function is, with the underlying goal of a fundamental shift in mindset away from **pointwise values** to **integral values (averages)**. This shift in mindset will lead us naturally into the spirit of distributions.

5.1. • Real-Valued Functions

5.1.1. • What Is a Function? We recall that a **function** f is a well-defined map or **input-output machine** from one set of numbers into another. In other words, a function f takes an input number and outputs another number:

$$f\,:\ \big(\text{an input number}\big) \longrightarrow \text{the output number.}$$

If f is a function of several variables, the input is a vector of numbers. Real numbers, or vectors of real numbers, are ideal for presenting either the input or the output variables. They are ideal for parametrizing either positions in physical space and/or time. Hence, throughout this text we will deal with functions defined on values in $\mathbb{R}^N$ (or some domain in $\mathbb{R}^N$) with N usually 1, 2, or 3. Specifically, if we **input** a value for $\mathbf{x} \in \mathbb{R}^N$, the function f **outputs** a real (or complex) number $f(\mathbf{x})$. For example, here are two

[1]**Paul Dirac (1902–1984)** was an English physicist who made fundamental contributions to both quantum mechanics and quantum electrodynamics. In 1932, Paul Dirac published his treatise *The Principles of Quantum Mechanics*, in which the delta function was officially singled out and widely used. However, the notion of such an object can be traced back much earlier to several mathematicians, in particular **Cauchy**.

functions with $N = 2$ and $N = 1$, respectively:

$$f_1(x_1, x_2) = x_1 + 4(x_2)^2, \qquad f_2(x) = \begin{cases} x & \text{if } x < 0, \\ 2 & \text{if } x = 0, \\ 4 & \text{if } x > 0. \end{cases}$$

One often associates a geometric picture to such a function by considering its **graph**, the locus (set) of points in $\mathbb{R}^{N+1}$ of the form $(\mathbf{x}, f(\mathbf{x}))$. In the case $N = 1$, the graph is a curve in $\mathbb{R}^2$. At this point in your studies you have spent a good deal of time learning tools for *analyzing* functions of one or more real variables based upon differential and integral calculus.

5.1.2. • Why Integrals (or Averages) of a Function Trump Pointwise Values.

In what follows, we loosely use the terms
integrals and **averages** as synonyms.

From a physical (and mathematical) point of view, functions often denote **densities**. The density at any given point has no significance; rather, it is the value of its **integrals** (or **averages**) over parts of the domain which have physical significance, and it is these values which can be physically observed and measured. We give four examples.

(i) Consider the **mass density** of a certain quantity, such as water, in the atmosphere. For the length scales of which we are interested, we assume that the mass density is a function which can continuously change over space. We cannot directly measure mass density at a particular point; after all, any single point should have zero mass. What one can measure is the mass of regions with nonzero volume; in other words, compute integrals of the mass density over regions of space. In fact even the physical dimensions of mass density, namely mass per volume, suggest the underlying notion of a region of space, rather than a point.

(ii) Consider the **temperature** (as a measure of the amount of heat) in a room or in the atmosphere. Can one measure the temperature (heat) at exactly one particular point in space? No, one measures temperature over a small region of space where some measure of heat energy can be registered. Moreover, temperature by its very definition entails **averaging** over velocities of a particle and over the particles themselves.

So while one might consider a temperature function $T(\mathbf{x})$ defined over points in space, we should actually focus our attention not on its pointwise values, but on its integral over regions of space. Indeed, we should think of the temperature function $T(\mathbf{x})$ as giving a temperature density, with physical dimensions of temperature per unit volume.

(iii) You may recall from a preliminary course in probability and statistics the notion of a **continuous probability density function** $p(x)$ defined for $x \in \mathbb{R}$, for example, the normal probability density function associated with the infamous bell-shaped curve. This is directly connected with the probability of **events**. However, note that the probability of x being any given value, say, $x = 3$, is not $p(3)$; rather it is zero! What has nonzero probability is that x lies in some interval I, and this probability is given by

$$\int_I p(x)\,dx.$$

(iv) Here is an example of a function which you probably would not refer to as a density. Consider the **velocity** of a falling ball. Because of gravity, this velocity will continuously change with time. How would we measure the velocity v at a given time t_1? We would fix a small time interval I including t_1 of size $\triangle t$, measure how far the ball traveled during this time interval, and then divide by $\triangle t$. This approximation to the actual velocity at t_1 is an **average** velocity over the time interval chosen:

$$\frac{1}{\triangle t}\int_I v(t)\,dt.$$

An Important Remark: The premise of these examples, demonstrating that integral values "trump" pointwise values, might seem rather contradictory and indeed a bit hypocritical, given that the focus of this text is on PDEs — pointwise equations relating the values of partial derivatives. Phrasing basic laws of physics/mathematics in the form of pointwise equations (PDEs) can indeed be very useful. However, the reader will soon see that many PDEs, especially those modeling some aspect of **conservation**, are derived from integral laws: equations relating integrals of an unknown function. These include[2] the diffusion equation, Laplace's equation, and Burgers's equation. Via integral to pointwise (IPW) theorems (cf. Section A.7.1), we can reduce these integral laws to pointwise laws (PDEs); see, for example, Section 7.1. This reduction is based upon the assumption of an underlying smooth solution. On the other hand, the ability to incorporate **discontinuities/singularities** into solutions (or initial values) to these PDEs is tremendously useful. These singularities in solutions/initial values can only be captured and analyzed at the level of the integrals, not at the level of the pointwise values. The purpose of this chapter, as well as Chapter 9, is to provide a new way of interpreting the solution to these PDEs in a manner focused on integrals (or averages). The new idea here will be "distributions" and as we shall eventually see, interpreting these PDEs in the "sense of distributions" amounts to returning to, or rather being **faithful to**, their original integral-based derivation.

5.1.3. • Singularities of Functions from the Point of View of Averages. We just made a case that in dealing with functions describing a physical variable which depends on space and/or time, **averages (integrals)** of the function are more relevant than their **pointwise values**. These averages will characterize the essence of the function. In particular one could ask, if we know the value of

$$\int_S f(x)\,dx,$$

for all sets S, do we know exactly the function? The answer is yes modulo a "small" set which has negligible size, where "negligible" is from the point of view of integration. This means a set sufficiently small enough such that any function integrated over it is zero. An example of such a set would be a finite set of points; however, there are infinite sets which also have this property. While we will not need them in this text,

[2]In fact, even the wave equation was derived based upon the calculation of integrals.

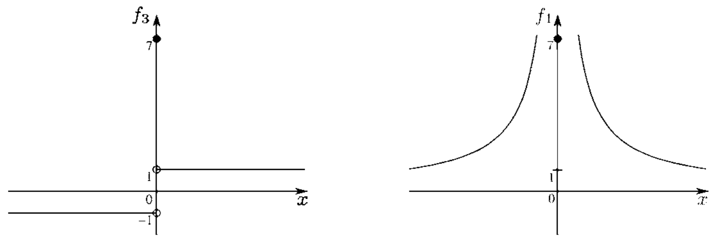

Figure 5.1. Graphs of the functions defined in (5.1) with, respectively, a jump and a blow-up singularity at $x = 0$.

it should be noted that there is a beautiful mathematical theory called *measure theory* which makes such notions very precise.

Thus, changing the values of a function at a finite number of inputs has no effect on any of its integrals. For example, from the point of view of integral calculus and its applications, the following two functions are the same:

$$f_1(x) = x^2 \qquad \text{and} \qquad f_2(x) = \begin{cases} x^2, & x \neq 2, \\ 7, & x = 2. \end{cases}$$

Certainly the values of their integrals over any interval are equal. On the other hand, there are other "singularities" that a function can possess which do have effects on integrals.

A **singularity** of a function is simply an input point x_0 where the function fails to be well-behaved (in terms of continuity or differentiability). Certain singularities, often labeled as **removable**, like that of $f_2(x)$ above, are for the most part irrelevant to the structure of the function. Other singularities, often labeled as **essential**, like those of

$$f_3(x) = \begin{cases} -1, & x < 0, \\ 1, & x > 0, \\ 7, & x = 0, \end{cases} \qquad \text{or} \qquad f_4(x) = \begin{cases} \frac{1}{|x|}, & x \neq 0, \\ 7, & x = 0, \end{cases} \tag{5.1}$$

are indeed essential to the structure of the function. Both of these functions have a singularity at $x = 0$. From the perspective of integrals and averages, the value of the function at $x = 0$ is indeed irrelevant. However, the behavior around the singularity — a jump and blow-up discontinuity, respectively — is relevant. Note that there is no redefinition of the value at $x = 0$ which would change the essential behavior around the singularity. As you will come to appreciate, both of these two types of singularities are something one would "*detect or see*" via integrals. They are central to the function's character. **Green's functions**, the subject of Chapter 10, are functions which exhibit such a singular behavior.

In this chapter we will consider piecewise smooth functions of **one variable** which have either jump or blow-up discontinuities. However, we will analyze these functions from the point of view of **integrals/averages** and, in doing so, unlock the true effect that these singularities have from the important perspective of **differentiation**.

With our focus on integration and averages of a function, we introduce (cf. Section A.4.3) two large classes of functions which will frequently appear in this and subsequent chapters.

A function $f(x)$ defined on $\mathbb{R}$ is **integrable** if

$$\int_{-\infty}^{\infty} |f(x)|\, dx \;<\; \infty.$$

This means that the improper integral exists and is a finite number. A function $f(x)$ defined on $\mathbb{R}$ is **locally integrable** if the integral of the absolute value over any finite interval is finite; i.e., for any $a < b$ with $a, b \in \mathbb{R}$,

$$\int_{a}^{b} |f(x)|\, dx \;<\; \infty.$$

5.2. • The Delta "Function" and Why It Is Not a Function. Motivation for Generalizing the Notion of a Function

You may have already encountered the **Dirac delta "function"** $\delta_0(x)$ which can be loosely defined as

$$\delta_0(x) = \begin{cases} 0, & x \neq 0, \\ \text{“suitably infinite”}, & x = 0, \end{cases}$$

where **"suitably infinite"** means that for any $a > 0$,

$$\text{“}\int_{-a}^{a} \delta_0(x)\, dx \;=\; 1\,\text{”}. \tag{5.2}$$

Alternatively, we can reformulate the above as follows: Given any function ϕ which is continuous on $[-a, a]$,

$$\text{“}\int_{-a}^{a} \delta_0(x)\,\phi(x)\, dx \;=\; \phi(0)\,\text{”}. \tag{5.3}$$

Thus, somehow the infinite density at $x = 0$ has resulted in the fact that multiplying the function $\phi(x)$ by the delta "function" and integrating *picks out* the value of ϕ at $x = 0$. **There is no function, that is, no input-output machine defined on the real numbers, that can do this.** This is why we make use of quotation marks in (5.2) and (5.3). Even if we decided to allow $+\infty$ as a possible output value for the function and consider a definition like

$$\text{“}\;\delta_0(x) = \begin{cases} 0, & x \neq 0, \\ +\infty, & x = 0, \end{cases}\;\text{”} \tag{5.4}$$

there is no unambiguous way to interpret its integral and, hence, enforce either (5.2) or (5.3). To achieve this one would have to make sense of the ambiguous product: $0 \times +\infty$. Moreover, we think you would agree that $2 \times +\infty$ should equal $+\infty$, but then by definition (5.4), should we conclude that

$$\text{“}\,\delta_0(x) \;=\; 2\delta_0(x)\,\text{”}\,?$$

In fact, the same reasoning would suggest that $\delta_0(x) = C\delta_0(x)$, for any $C > 0$.

From these observations, one might be tempted to conclude that this delta "function" is an abstract fabrication[3] with little connection to reality which, up to this point, seems to be well described by the standard calculus of *honest-to-goodness* functions. Well, nothing could be further from the truth; the delta "function" comes up naturally, even when one just deals with functions. As we shall see, it plays a **central role** in two fundamental areas:

- linear second-order partial differential equations,
- Fourier analysis (Fourier series and the Fourier transform).

Indeed, we stress that it is one of the most important "*players*" in this book. To give a small hint of this, we present in the next subsection two simple examples involving honest-to-goodness functions where this, as yet ill-defined, delta "function" arises. These examples will help provide the intuition behind both the essential character of the delta "function" and its eventual definition as a distribution. The key phenomenon prevalent in both examples is **concentration**.

5.2.1. • The Delta "Function" and the Derivative of the Heaviside Function. Consider the Heaviside[4] function defined by

$$H(x) := \begin{cases} 0, & x < 0, \\ 1, & x \geq 0. \end{cases} \tag{5.5}$$

What is its derivative? If we look *pointwise*, i.e., fix a point x and ask what $H'(x)$ is, we see that

$$H'(x) = \begin{cases} 0, & x \neq 0, \\ undefined, & x = 0. \end{cases}$$

Can we ignore the fact that $H'(0)$ is undefined? Since $H'(x) = 0$ except at one point, our previous discussion about values at a single point might suggest that the derivative H' behaves like the zero function. But then $H(x)$ should be a constant, which is not the case: $H(x)$ changes from 0 to 1 but does so *instantaneously*. Hence perhaps we should assert that the derivative is *entirely concentrated at* $x = 0$ and define $H'(0)$ to be $+\infty$, associated with the instantaneous change from 0 up to 1 at $x = 0$. In other words, take

$$H'(x) = \begin{cases} 0, & x \neq 0, \\ +\infty, & x = 0. \end{cases} \tag{5.6}$$

Unfortunately there is ambiguity in this $+\infty$; for example, in order to reconstruct H (up to a constant) from its derivative we would need to make sense of the recurring product "$0 \times +\infty$". Further to this point, suppose we considered a larger jump, say

$$H_7(x) = \begin{cases} 0, & x < 0, \\ 7, & x \geq 0, \end{cases}$$

---

[3] In fact even one of the greatest mathematicians / computer scientists of the 20th century, **John von Neumann (1903–1957)**, was so adamant that Dirac's delta function was "*mathematical fiction*" that he wrote his monumental monograph, published in 1932, *Mathematische Grundlagen der Quantenmechanik* (Mathematical Foundations of Quantum Mechanics) in part to explain quantum mechanics with absolutely no mention of the delta "function". John von Neumann described this artifact in the preface as an "*improper function with self-contradictory properties*". This counterpart to Dirac's treatise was based on functional analysis and Hilbert spaces.

[4] Named after the British self-taught scientist and mathematician **Oliver Heaviside (1850–1925)**, a rather interesting person who was quite critical about contemporary mathematical education. Look up online his "Letter to Nature".

would its derivative still be (5.6)? The functions $H(x)$ and $H_7(x)$ are clearly different functions whose structure is **lost** in the derivative definition (5.6).

Let us pursue this further by focusing on **integration**. Fix a smooth (C^1) function ϕ which is identically 0 if $|x| \geq 1$. Consider the function $f(x)$ (which, as we will later see, is called a convolution) defined by

$$f(x) \coloneqq \int_{-\infty}^{\infty} H(x-y)\,\phi(y)\,dy,$$

where H is the Heaviside function of (5.5). While the definition of $f(x)$ is rather "convoluted", it presents a perfectly well-defined function. Indeed, despite the discontinuity in H, f is continuous and, in fact, differentiable at all x. While, as we shall see, it is easy to prove these assertions (cf. (5.8) and (5.9) below), you should try to convince yourself of this before proceeding on.

Next, let us attempt to compute $f'(x) = \frac{df(x)}{dx}$ in two different ways:

First way: On one hand, we could "attempt" to bring the x-derivative inside[5] the integral (with respect to y) to find

$$f'(x) = \frac{df(x)}{dx} = \int_{-\infty}^{\infty} \frac{dH(x-y)}{dx}\,\phi(y)\,dy. \tag{5.7}$$

Should we take $\frac{dH(x-y)}{dx}$ as a function of y to be identically 0? After all, it is 0 except at the one point $y = x$. But then this would imply that $f'(x) = 0$ at all points x. If we try to incorporate the behavior at $y = x$ and claim

$$\frac{dH(x-y)}{dx} \text{ is infinity when } y = x,$$

then we are back to the same question on how to compute the integral.

Second way: On the other hand, by definition of the Heaviside function note that

$$H(x-y) = \begin{cases} 0 & \text{if } y > x, \\ 1 & \text{if } y \leq x. \end{cases}$$

Hence $f(x)$ can be conveniently written as

$$f(x) = \int_{-\infty}^{\infty} H(x-y)\,\phi(y)\,dy = \int_{-\infty}^{x} \phi(y)\,dy. \tag{5.8}$$

But now things are more transparent. By the Fundamental Theorem of Calculus, f is continuous and differentiable with

$$f'(x) = \phi(x). \tag{5.9}$$

Now there is absolutely no ambiguity, and the answer is not, in general, 0.

So what happened in (5.7)? What occurred was the derivative of the translated Heaviside function,

$$\frac{dH(x-y)}{dx} \quad \text{viewed as a function of } y,$$

[5]Note here that the hypotheses for differentiation under the integral sign (Theorem A.10) fail to hold true.

did what a delta "function" (concentrated at $y = x$) was supposed to do; i.e., when integrated against $\phi(y)$, it picked out the value of $\phi(y)$ at $y = x$. Note that, because of the translation, the concentration was moved from $y = 0$ to $y = x$.

5.2.2. • The Delta "Function" as a Limit of a Sequence of Functions Which Concentrate.

The delta "function" also appears when we look at what happens to **sequences of functions which concentrate**. Suppose that f_n is defined by

$$f_n(x) = \begin{cases} n - n^2 x, & 0 < x < 1/n, \\ n + n^2 x, & -1/n < x \le 0, \\ 0, & |x| \ge 1/n. \end{cases} \tag{5.10}$$

Do not be put off by the formula, as these are simply steeper and steeper spike functions as illustrated in Figure 5.2. Note that for all $n \in \mathbb{N}$,

$$\int_{-\infty}^{\infty} f_n(x)\, dx \;=\; 1. \tag{5.11}$$

However, for all $x \neq 0$, $f_n(x)$ (as a sequence in n) is eventually 0. So what happens to these functions as $n \to \infty$? If we look at this from a pointwise perspective, it would seem as if $f_n(x)$ tends to 0 as $n \to \infty$ except at $x = 0$. On the other hand, $f_n(0)$ tends to $+\infty$. If we include $+\infty$ as a possible output and write the pointwise limit as the function

$$f(x) = \begin{cases} 0, & x \neq 0, \\ +\infty, & x = 0, \end{cases} \tag{5.12}$$

then what is

$$\int_{-\infty}^{\infty} f(x)\, dx?$$

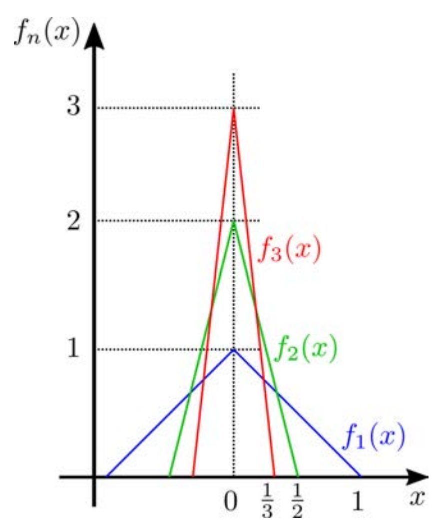

Figure 5.2. Plots of functions in the sequence $\{f_n(x)\}_{n\in\mathbb{N}}$ defined by (5.10) for $n = 1, 2, 3$.

Given that for every n, $\int_{-\infty}^{\infty} f_n(x)\,dx = 1$, one would expect it to be 1. But again, there is no way to make sense of this integral without having a nonambiguous value for $0 \times +\infty$. Indeed, the **same pointwise limit** (5.12) would occur with the sequence of functions $2f_n(x)$ or $7f_n(x)$, which differ from $f_n(x)$ in that the area under the curve is no longer always 1, but rather 2 and 7, respectively. This area information is **lost** by simply assigning the value of $+\infty$ at $x = 0$ in (5.12).

These points should sound familiar! As we just did for the derivative of the Heaviside function, let us investigate this from the perspective of integration or, more precisely, integration *against* another function. Suppose that ϕ is a continuous function on $\mathbb{R}$ and then consider what happens to

$$\int_{-\infty}^{\infty} f_n(x)\phi(x)\,dx \quad \text{as } n \to \infty.$$

Due to the concentration property of this sequence, we have

$$\int_{-\infty}^{\infty} f_n(x)\phi(x)\,dx = \int_{-\frac{1}{n}}^{\frac{1}{n}} f_n(x)\phi(x)\,dx.$$

When n is very large, the continuous function ϕ is approximately constant on the tiny interval $\left[-\frac{1}{n}, \frac{1}{n}\right]$; i.e., for $x \in \left[-\frac{1}{n}, \frac{1}{n}\right]$, $\phi(x) \sim \phi(0)$, where the notation $\sim$ simply means "close to" or "approximately equal to". Thus, when n is very large we have

$$\int_{-\frac{1}{n}}^{\frac{1}{n}} f_n(x)\phi(x)\,dx \quad \sim \quad \int_{-\frac{1}{n}}^{\frac{1}{n}} f_n(x)\phi(0)\,dx = \phi(0)\int_{-\frac{1}{n}}^{\frac{1}{n}} f_n(x)\,dx = \phi(0),$$

where we use (5.11) as well as the concentration property of the sequence for the last equality. Thus, it would seem that

$$\lim_{n\to\infty} \int_{-\infty}^{\infty} f_n(x)\phi(x)\,dx = \phi(0),$$

and as before, what the f_n seem to be converging to is this, as yet ill-defined, delta "function".

The sequence of functions represented in (5.10) is a very good way to *visualize* the delta "function" in terms of honest-to-goodness functions. In fact, this is exactly how Dirac thought of it[6]:

> *"To get a picture of $\delta(x)$, take a function of the real variable x which vanishes everywhere except inside a small domain, of length ϵ say, surrounding the origin $x = 0$, and which is so large inside this domain that its integral over this domain is unity. The exact shape of the function inside this domain does not matter, provided there are no unnecessarily wild variations (for example provided the function is always of order ϵ^{-1}). Then in the limit $\epsilon \to 0$ this function will go over to $\delta(x)$."*

[6]From page 58 of **P. Dirac**, *The Principles of Quantum Mechanics*, Oxford at the Clarendon Press, third edition, 1947.

Conclusion/Punchline of This Section

- Dirac's delta "function" pops up in instances when (i) differentiating the Heaviside function to capture a derivative which concentrates at one point and (ii) when considering limits of certain functions which seem to concentrate.
- In both cases, we were able to make sense of what was happening, not via pointwise values of x, but by **focusing on integration** involving a generic function $\phi(x)$.

True functions have significance without any notion of integration **but** the delta "function" **does not**. In the next section we will provide a precise definition of this delta "function" as a generalized function or **distribution**. After this, we will agree to keep the name and dispense with quotation marks in "function".

5.3. • Distributions (Generalized Functions)

Hopefully the previous section has given you some motivation for *generalizing* the notion of a function to include the delta "function" and to emphasize the importance of averages of classical "honest-to-goodness" functions. This generalization should preserve the important character of a classical function from the point of view of calculus. The basic object of our generalization will be a **distribution**.

Unlike with classical functions where we **understand/capture/characterize** the object as an input/output machine on the real numbers, we will **understand/capture/characterize** a distribution by "what it does" to **test functions**. In other words, a distribution will remain an input/output machine, but the difference is we now input a test function and output a real number. In this scenario, where the inputs are functions themselves, we often call the input/output machine a **functional**. Once we have a precise notion of a distribution, we can readily provide a precise meaning to the delta "function", transforming it into a proper mathematical object upon which we have a firm and solid foundation to explore its presence and uses.

5.3.1. • The Class of Test Functions $C_c^\infty(\mathbb{R})$. The class of test functions will consist of **localized smooth functions** where here, by *smooth*, we will mean infinitely differentiable (that is, very smooth!). By *localized* we mean that they are only nonzero on a bounded set. Precisely, we consider $\phi : \mathbb{R} \to \mathbb{R}$ in the class $C_c^\infty(\mathbb{R})$, where the "c" in the subindex position denotes **compact support**. Recall that a continuous function of one variable has compact support if it is identically zero outside of some closed finite interval.

The definition of $C_c^\infty(\mathbb{R})$ may seem very restrictive and, indeed, you might wonder whether or not such a function even exists; remember, we want a perfectly smooth function (no singularities in any derivative) which is eventually flat in either direction. In other words the function becomes identically flat (zero) at some finite value in a perfectly smooth fashion (cf. Figure 5.3). While no polynomial function can do this, here is the generic example which is called the **bump function** (illustrated in Figure

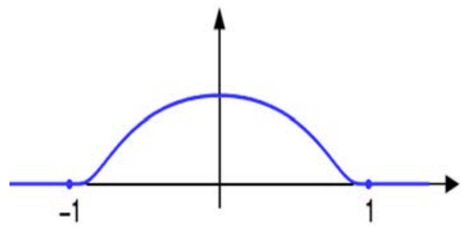

Figure 5.3. A canonical function in $C_c^\infty(\mathbb{R})$: graph of the *bump function* defined by (5.13) with $a = 1$.

5.3): For any $a > 0$, let

$$\phi_a(x) \;:=\begin{cases} e^{-\frac{1}{a^2-x^2}} & \text{if } |x| < a, \\ 0 & \text{if } |x| \geq a. \end{cases} \tag{5.13}$$

Clearly the support of this function is contained in the interval $[-a, a]$. In Exercise **5.1**, you are asked to further show that this function is $C^\infty(\mathbb{R})$; the only obstacle is at the joining points $x = \pm a$. You may view this example as rather contrived or special. However, even though we may not have simple explicit formulas for them, there are many more functions in $C_c^\infty(\mathbb{R})$. Indeed, in Section 6.3.4 we will show how the notion of convolution of functions enables one to use the bump function to "infinitely smooth out" any function with compact support, generating designer functions in $C_c^\infty(\mathbb{R})$. This will provide a way to "approximate" any function with compact support with a function in $C_c^\infty(\mathbb{R})$. In mathematical analysis one can phrase and prove precise statements about such approximations and, in doing so, establish that the set of $C_c^\infty(\mathbb{R})$ comprises a **dense** subset of many larger function spaces. **Conclusion: There are "lots and lots" of test functions.**

Lastly, note that the sum of any two test functions in $C_c^\infty(\mathbb{R})$ is also a test function in $C_c^\infty(\mathbb{R})$. The same is true for the product of a constant multiplied by a test function. Moreover,

$$\text{if } \phi \in C_c^\infty(\mathbb{R}), \text{ then } \phi' = \frac{d\phi}{dx} \in C_c^\infty(\mathbb{R}).$$

Consequently, derivatives of all orders of test functions are also test functions; this was one of the reasons why we imposed the C^∞ smoothness criterion.

5.3.2. • The Definition of a Distribution.

We are now at a point where we can define the basic object of this chapter — a distribution. To this end, let us first state the definition and then explain several of the terms used.

Definition of a Distribution

Definition 5.3.1. A **distribution** F (also known as a **generalized function**) is a rule, assigning to each test function $\phi \in C_c^\infty(\mathbb{R})$ a real number, which is linear and continuous.

By a rule, we simply mean a map (or functional) from the space of test functions $C_c^\infty(\mathbb{R})$ to $\mathbb{R}$. We denote by

$$F : \phi \in C_c^\infty(\mathbb{R}) \longrightarrow \mathbb{R}$$

such a functional on $C_c^\infty(\mathbb{R})$ and adopt the notation

$$\langle F, \phi \rangle$$

for the action of the distribution F on ϕ. That is, for each $\phi \in C_c^\infty(\mathbb{R})$, $\langle F, \phi \rangle$ is a real number. We require this functional F to be linear and continuous in the following sense:

Linearity means

$$\langle F, a\phi + b\psi \rangle = a \langle F, \phi \rangle + b \langle F, \psi \rangle \qquad \text{for all } a, b \in \mathbb{R} \text{ and } \phi, \psi \in C_c^\infty(\mathbb{R}).$$

Continuity can be phrased in two ways; one "loose" and informal, the other precise.

Loose and informal description of continuity: One way is to assert that if two test functions are "close", then the associated respective actions of the distribution (two real numbers) are also "close". Closeness of two real numbers is clear and unambiguous, but we need to specify what closeness means for two test functions. In the context of the test functions, we (loosely) say two test functions ϕ_1 and ϕ_2 in $C_c^\infty(\mathbb{R})$ are close if

$$\max_{x\in\mathbb{R}} |\phi_1(x) - \phi_2(x)| \qquad \text{is small,}$$

and, further, the same is deemed true for their derivative functions of any order; i.e., for any $k = 1, 2, \ldots$

$$\max_{x\in\mathbb{R}} |\phi_1^{(k)}(x) - \phi_2^{(k)}(x)| \qquad \text{is small,}$$

where $\phi_i^{(k)}$ is the k-th derivative of ϕ_i. Geometrically, two test functions ϕ_i would satisfy the first criterion of closeness if the graph of ϕ_2 can fit in a small **bar-neighborhood** of the graph of ϕ_1, as depicted in Figure 5.4. They would satisfy the second criterion of closeness if the same holds true for the respective functions obtained by taking the k-th derivative.

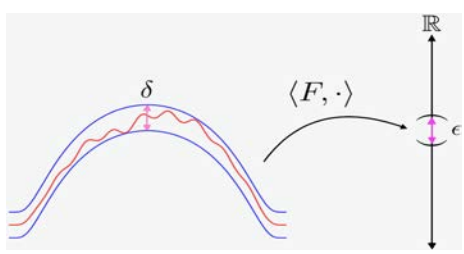

Figure 5.4. On the left is a δ bar-neighborhood in $C_c^\infty(\mathbb{R})$. For δ small, any two functions lying in the δ bar-neighborhood are deemed "close". Note, however, that for closeness in $C_c^\infty(\mathbb{R})$, we require each of the respective derivative functions to also lie in a δ bar-neighborhood. Any distribution F should map a function in this bar-neighborhood to an interval of width ϵ.

Precise description of continuity: Alternatively, we can rephrase continuity in the following sequential statement: If ϕ_n “converges to” ϕ in $C_c^\infty(\mathbb{R})$, then

$$\langle F, \phi_n \rangle \xrightarrow{n\to\infty} \langle F, \phi \rangle. \tag{5.14}$$

Since the second convergence relates to real numbers, there should not be any ambiguity here. But what about convergence of the test functions? While we were informal with the notion of closeness of functions, let us give here a precise definition of convergence of test functions. It relies on the notion of uniform convergence (cf. Section 11.7). We say

$$\phi_n \xrightarrow{n\to\infty} \phi \qquad \text{in } C_c^\infty(\mathbb{R}) \qquad \text{if}$$

(i) there exists an $a > 0$ such that all the functions $\phi_n(x)$ vanish for $|x| \geq a$ and

(ii) for any $k = 0, 1, 2, \ldots,$

$$\phi_n^{(k)}(x) \ \text{ converges uniformly to } \phi^{(k)}(x), \tag{5.15}$$

where $\phi_n^{(k)}(x)$ denotes the k-th derivative of ϕ_n (with the understanding that $\phi_n^{(0)}(x) = \phi_n(x)$). In the context of test functions in $C_c^\infty(\mathbb{R})$, the uniform convergence, i.e., (5.15), is equivalent to

$$\max_{x\in\mathbb{R}} |\phi_n^{(k)}(x) - \phi^{(k)}(x)| \xrightarrow{n\to\infty} 0, \qquad \text{for all } k = 0, 1, 2, \ldots.$$

Hence the precise definition of continuity is the following condition: **If** $\phi_n \xrightarrow{n\to\infty} \phi$ in $C_c^\infty(\mathbb{R})$, **then** (5.14) holds true.

Before continuing, let us provide one example of a distribution.

Example 5.3.1. Consider the functional F_1 on $C_c^\infty(\mathbb{R})$ which assigns to each $\phi \in C_c^\infty(\mathbb{R})$ its integral over $\mathbb{R}$; that is, the rule is given by

$$\langle F_1, \phi \rangle = \int_{\mathbb{R}} \phi(x)\, dx.$$

We claim that F_1 is a distribution. It should be clear that this rule (or functional) is linear (check this). It should also be intuitively clear that if the functions are close (in the above sense), then their integrals are close.[7]

The set (class) of distributions[8] forms a vector space over $\mathbb{R}$. This essentially means:

(i) We can multiply any distribution F by a scalar $a \in \mathbb{R}$ and generate a new distribution aF defined by

$$\langle aF, \phi \rangle := a\langle F, \phi \rangle \qquad \text{for all } \phi \in C_c^\infty.$$

(ii) We can add two distributions F and G to form a new distribution $F + G$ defined by

$$\langle F + G, \phi \rangle := \langle F, \phi \rangle + \langle G, \phi \rangle \qquad \text{for all } \phi \in C_c^\infty.$$

[7] A precise argument can be achieved via Theorem A.8.

[8] Readers familiar with the theory of linear algebra may recognize this formalism in the context of **dual** vector spaces. Indeed, one may view the space of distributions over the class $C_c^\infty(\mathbb{R})$ as the dual space of $C_c^\infty(\mathbb{R})$. A finite-dimensional vector space X is coupled with an additional vector space of the same dimension which is a vector space of linear functionals on X. The study of duality on function spaces (infinite-dimensional vector spaces) is far more involved and this lies in the realm of functional analysis.

To conclude, a distribution is characterized by how it **acts** on test functions. One should never ask what the *value* of a distribution F is at a point $x \in \mathbb{R}$, but rather what the value of F is "at" a test function ϕ. Put another way, you may think of the action (value) of F on a test function as a way of **sampling the distribution**. However, be warned to not directly identify these notions of sampling and distributions with their common uses in statistics; while they are related, they are not exactly the same.

In the next two subsections, we will present two fundamental types/examples of distributions. There are many other far more complicated distributions beyond these two types. While we will discuss one such class in Section 5.8, for most purposes in this book, these two types will suffice. The first type is indicative of a distribution **generalizing the notion of a function** and encapsulates the statement that every function can be regarded (or captured) as a distribution in a **natural way**.

5.3.3. • Functions as Distributions.

Any Locally Integrable Function as a Distribution

Let $f(x)$ be **a locally integrable function** on $\mathbb{R}$. Then $f(x)$ can be interpreted as the distribution F_f where

$$\langle F_f, \phi \rangle := \int_{-\infty}^{\infty} f(x)\, \phi(x)\, dx \quad \text{for any } \phi \in C_c^\infty(\mathbb{R}). \tag{5.16}$$

Think of F_f as **the distribution generated** by the function $f(x)$.

Why do we use the phrase "in a natural way"? Well, it is based upon the paradigm, argued in Section 5.1.2, that integrals (averages) of a function are fundamentally more important than its pointwise values. Indeed, the definition is only based upon the integration properties of the function f (averaging properties); hence, if we were to change the value of $f(x)$ at a finite number of points, we would still generate the **same** distribution[9]. For example, the two functions

$$f_1(x) = x^3 \qquad \text{and} \qquad f_2(x) = \begin{cases} x^3 & \text{if } x \neq 7, \\ 5 & \text{if } x = 7 \end{cases}$$

generate the same distribution. One could ask if, by only "recording" these weighted integrals of $f(x)$ (i.e., $\langle F_f, \phi \rangle$), do we lose information about f as a function? Stated a different way, do all the values of $\langle F_f, \phi \rangle$ uniquely determine the function $f(x)$ at all points $x \in \mathbb{R}$? The answer is yes modulo a negligible set of real numbers which has no effect on integration, i.e., zero measure in the sense of measure theory [**17**].

Lastly, note that our first example of a distribution (Example 5.3.1) was exactly the distribution generated by the function which was identically 1. This distribution assigns to every test function $\phi \in C_c^\infty(\mathbb{R})$ its integral over all $\mathbb{R}$.

Notation: The notation F_f becomes cumbersome over time so we will usually dispense with it, just using f **but** making it clear in words that we are "thinking" of f

[9] In fact, you can change the value of the functions on any set *which will have no effect on integration*. This is made precise if one studies a bit of *measure theory* [**17**].

as a distribution. Therefore, when one makes a statement about a function f and adds the phrase *in the sense of distributions*, one is speaking about F_f.

5.3.4. • The Precise Definition of the Delta "Function" as a Distribution.

We now give the delta "function" δ_0 **official status**. Indeed, while it is **not** a function, it most certainly **is** a distribution. Whereas part of our motivation in Section 5.1.2 was to capture functions from their "averages" rather than their pointwise values, here it will be crucial that **our test functions** are continuous and their value at a point (say $x = 0$) represents the local environment of the test function. Indeed, the following definition simply states that δ_0 is a distribution whose action on any test function picks out its value at 0.

Definition of the Delta Function

Definition 5.3.2. The delta "function" δ_0 is the distribution defined as follows:

$$\langle \delta_0, \phi \rangle = \phi(0) \qquad \text{for any } \phi \in C_c^\infty(\mathbb{R}).$$

We can also consider delta functions *concentrated at* point $x_0 \in \mathbb{R}$; that is,

$$\langle \delta_{x_0}, \phi \rangle = \phi(x_0) \qquad \text{for any } \phi \in C_c^\infty(\mathbb{R}).$$

With Definition 5.3.2 in hand, let us now agree to **abandon the quotation marks** around the word *function* and yield to tradition, partially admitting defeat, by simply referring to the distribution δ_0 as **the delta function**. However, we will usually suppress the explicit functional dependence on x and not write "$\delta_0(x)$".

Two important remarks are in order:

(i) **A delicate and subtle point on integration and the delta function:** There is **no integral** in the above definition of the delta function; we did **not** ever write (or need to write)

$$\text{“} \int \delta_0 \, \phi(x) \, dx \text{”}.$$

On the other hand, recall the punchline of Section 5.2.2 where we concluded by claiming that while functions have significance without any notion of integration, the delta function does not. So where is the notion of integration in Definition 5.3.2? The subtle answer is that integration is implicitly embedded in the definition because of the way in which we envision a function as a distribution. More precisely, to address δ_0 in any scenario involving functions, we must view the functions as objects to be integrated against test functions.

(ii) **Functional notation and the use of quotation marks in the sequel:** Occasionally, we may use the functional notation for the delta function to focus directly on the intuition; however, in these cases, we will also use quotation marks when viewing an object as a function which is the **delta function in disguise**, for example, "$H'(x)$", the derivative of the Heaviside function. Another important example of the delta function in disguise is the following indefinite integral:

$$\text{“}\, \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-ixy}\, dy. \,\text{”} \tag{5.17}$$

Here, i is the complex number $\sqrt{-1}$ (cf. Section 6.1) and x is a real parameter. It is natural to consider (5.17) as a function of x. However, this function of x makes no sense; for any fixed x, how do you make sense of the indefinite integral which can be thought of as periodically traversing the unit circle in the complex plane infinitely many times! On the other hand we shall see later, when studying the Fourier transform, that (5.17) is one of the most important and famous *disguises* for the delta function. That is, this improper integral can be interpreted in the sense of distributions and is the same distribution as the delta function. In fact, (5.17) is often referred to as an *integral representation of the delta function*.

To conclude in a nutshell: $\delta_0\,\cancel{(x)}$. Well, at least for now! As just mentioned, sometimes it is useful to guide us by informally treating the delta function as a function and manipulating the argument of the function, i.e., the $(\cdot)$. We will on occasion also do this but will always use the quotation marks. See Section 5.6 for more on this.

5.4. • Derivative of a Distribution

5.4.1. • Motivation via Integration by Parts of Differentiable Functions.

Given a function $f \in C^1$, we can find the derivative $f'(x)$ at any $x \in \mathbb{R}$ as

$$\lim_{h \to 0} \frac{f(x+h) - f(x)}{h}.$$

The definition highlights that this pointwise derivative is an *instantaneously local* quantity associated with the function f. We can think of $f'(x)$ itself as a continuous function on $\mathbb{R}$.

Let us explore the role that this pointwise derivative (or classical derivative) function plays from the perspective of integrating (or averaging) against a test function $\phi \in C_c^\infty(\mathbb{R})$. The key notion here is **integration by parts**. Indeed, since $\phi(x)$ has compact support, there exists some finite interval, say, $[-L, L]$, such that $\phi(x) = 0$ for all x with $|x| \geq L$. Hence, via integration by parts, we have

$$\begin{aligned}
\int_{-\infty}^{\infty} f'(x)\,\phi(x)\,dx \quad &= \quad \int_{-L}^{L} f'(x)\,\phi(x)\,dx \\
&\overset{\text{integrations by parts}}{=} \quad -\int_{-L}^{L} f(x)\,\phi'(x)\,dx \,+\, \Big[f(x)\phi(x) \Big]\Big|_{-L}^{L} \\
&= \quad -\int_{-L}^{L} f(x)\,\phi'(x)\,dx \,+\, f(L)\,\phi(L) \,-\, f(-L)\,\phi(-L) \\
&= \quad -\int_{-\infty}^{\infty} f(x)\,\phi'(x)\,dx.
\end{aligned}$$

Note, since $\phi(-L) = 0 = \phi(L)$, the boundary terms after the integration by parts are zero. The choice of L depends on the particular test function ϕ but, regardless, we still have

$$\text{for all } \phi \in C_c^\infty(\mathbb{R}), \qquad \int_{-\infty}^{\infty} f'(x)\,\phi(x)\,dx = -\int_{-\infty}^{\infty} f(x)\,\phi'(x)\,dx. \tag{5.18}$$

This simple formula tells us that, with respect to integration, we can place the derivative on either f or ϕ, at the expense of a minus sign.

5.4.2. • The Definition of the Derivative of a Distribution.

With (5.18) as motivation, we now define the derivative for **any** distribution.

Definition of the Derivative of a Distribution

Definition 5.4.1. Let F be any distribution. Then F' (its derivative) is also a distribution defined by

$$\langle F', \phi \rangle := -\langle F, \phi' \rangle \qquad \text{for } \phi \in C_c^\infty(\mathbb{R}). \tag{5.19}$$

Note that the definition makes sense since if $\phi \in C_c^\infty(\mathbb{R})$, then $\phi' \in C_c^\infty(\mathbb{R})$.

Thus every distribution has a distributional derivative. Since every locally integrable function generates a distribution, it follows that every locally integrable function, even those which are not classically differentiable, has a distributional derivative. But now let me pose the essential question with respect to **consistency**. Above we have given a definition for the distributional derivative. But is it a "good" definition in the sense that it is **consistent** with our classical notion of differentiability? To address this, consider a classically differentiable function $f \in C^1$. Its pointwise derivative f' exists and is a continuous function on $\mathbb{R}$. Hence, both f and f' are locally integrable and generate (can be thought of as) distributions via (5.16). For consistency to hold, we require that

$$(F_f)' = F_{f'}. \tag{5.20}$$

This says the following:

The distributional derivative of the distribution generated by a smooth function is simply the distribution generated by the classical derivative.

You should read this sentence over and over again until you have fully digested its meaning (this might take some time so please be patient!).

It is easy to see that it is precisely (5.18) which makes (5.20) hold true. To this end, the equality of distributions means that for any $\phi \in C_c^\infty(\mathbb{R})$,

$$\big\langle (F_f)', \phi \big\rangle = \big\langle F_{f'}, \phi \big\rangle.$$

But this follows directly from (i) the definition of the distributional derivative (Definition 5.4.1), (ii) the interpretation of a function as a distribution (5.16), and (iii) the integration by parts formula (5.18). Indeed,

$$\big\langle (F_f)', \phi \big\rangle \;=\; -\big\langle F_f, \phi' \big\rangle \;=\; -\int_{-\infty}^{\infty} f(x)\,\phi'(x)\,dx \;=\; \int_{-\infty}^{\infty} f'(x)\,\phi(x)\,dx \;=\; \big\langle F_{f'}, \phi \big\rangle .$$

Hence, the classical notion of differentiation is **carried over** to our generalized setting of distributions. However, Definition 5.4.1 also applies to **any** distribution. We now give a few words on the two important types of distributions (distributions generated by functions and the delta function).

Derivative of a Locally Integrable Function in the Sense of Distributions. Even if a locally integrable function f is not differentiable in the classical pointwise sense, we have a derivative in the sense of distributions. This derivative in the sense of distributions is a distribution G which may, or may not, be generated by another locally integrable function (cf. the next subsection). In terms of language, if f is a locally integrable function and G is a distribution, we say the derivative of f equals G **in the sense of distributions** if $(F_f)' = G$. To repeat, when we speak of a derivative of a function f in the sense of distributions, we always mean the distributional derivative of F_f.

The Derivative of the Delta Function. Consider the delta function δ_0. As we have discussed at length, it is not a function but it certainly is a distribution, and hence we can find its distributional derivative. This distributional derivative, conveniently denoted as δ_0', is the distribution defined by

$$\langle \delta_0', \phi \rangle \;\coloneqq\; -\langle \delta_0, \phi' \rangle \;=\; -\phi'(0) \qquad \text{for } \phi \in C_c^\infty(\mathbb{R}).$$

In other words, it is the distribution which assigns to each test function ϕ the negative value of its derivative at the point $x = 0$.

Finally, let us note that differentiation of distributions immediately carries over to **higher-order derivatives.** One can repeat this process of distributional differentiation (5.19) to define any number of derivatives of a distribution; this was the advantage of choosing C^∞ test functions. The sign on the right-hand side needs to be adjusted in accordance with the number of derivatives. The definition is again a consequence of the integration by parts formula which motivated definition (5.19). Indeed, if n is any positive integer, we define the n-th derivative of a distribution F to be the new distribution (denoted by $F^{(n)}$) defined by

$$\boxed{\langle F^{(n)}, \phi \rangle \;=\; (-1)^n \, \langle F, \phi^{(n)} \rangle \qquad \text{for } \phi \in C_c^\infty(\mathbb{R}).}$$

Here $\phi^{(n)}$ denotes the n-th derivative of the function $\phi(x)$.

5.4.3. • Examples of Derivatives of Piecewise Smooth Functions in the Sense of Distributions. Let us begin by differentiating some piecewise smooth functions in the sense of distributions. Why? Because we shall now see the effect of jump discontinuities in the distributional derivatives.

Example 5.4.1 (**the Heaviside Function**). Let us return to the Heaviside function (5.5) and differentiate it as a distribution; i.e., differentiate in the sense of distributions. We have for any test function ϕ

$$\begin{aligned} \langle (F_H)', \phi \rangle \;=\; -\langle F_H, \phi' \rangle \;&=\; -\int_{-\infty}^{\infty} H(x)\,\phi'(x)\,dx \\ &=\; -\int_{0}^{\infty} \phi'(x)\,dx \\ &=\; -\big(\phi(+\infty) - \phi(0)\big) \\ &=\; \phi(0) \\ &=\; \langle \delta_0, \phi \rangle . \end{aligned}$$

Note that the term $\phi(+\infty)$ vanished because of compact support. Thus, $(F_H)' = \delta_0$ or, in other words, the derivative of the Heaviside function in the sense of distributions is the Dirac delta function.

Note: We stress that at no point did we write, or need to write,

$$\text{“}\int_{-\infty}^{\infty} H'(x)\,\phi(x)\,dx\text{ ”}.$$

This does not make sense since $H'(x)$ is a distribution which cannot be generated by a function. Only distributions which are generated by a function have the property that their application to a test function can be written as a regular integral. However, it is common practice to write down expressions like

$$\text{“}\int_{-\infty}^{\infty} \underbrace{H'(x)}_{\delta_0}\;\phi(x)\,dx\text{”}.$$

Even though such expressions are technically incorrect, we now know how to make sense of them using the theory of distributions. In the rest of this text, we will sometimes write expressions like this for intuition and motivation. However, in these cases, we will always write them within parentheses "… " .

Example 5.4.2. Let $f(x) = |x|$. First, let us ask what $f'(x)$ is. Pointwise, we would say

$$f'(x) = \begin{cases} -1, & x < 0, \\ undefined, & x = 0, \\ 1, & x > 0. \end{cases}$$

So, again, we have the issue of undefined at $x = 0$. Is this "undefined" important? In other words, if we set the value of $f'(0)$ to be some number (say, 45), would we miss something? Let us do this differentiation in the sense of distributions. We have

$$\big\langle (F_f)', \phi \big\rangle = -\big\langle F_f, \phi' \big\rangle \;=\; -\int_{-\infty}^{\infty} |x|\,\phi'(x)dx \;=\; -\int_{-\infty}^{0} -x\,\phi'(x)dx - \int_{0}^{\infty} x\,\phi'(x)dx$$

and integrating by parts in each integral, we find that

$$\begin{aligned}
\langle (F_f)', \phi \rangle &= -\int_{-\infty}^{0} \phi(x)\,dx + [x\phi(x)]\Big|_{-\infty}^{0} + \int_{0}^{\infty} \phi(x)\,dx - [x\phi(x)]\Big|_{0}^{+\infty} \\
&= -\int_{-\infty}^{0} \phi(x)\,dx + \int_{0}^{\infty} \phi(x)\,dx \\
&= \int_{-\infty}^{\infty} g(x)\,\phi(x)\,dx = \langle F_g, \phi \rangle,
\end{aligned}$$

where g is the function

$$g(x) = \begin{cases} -1, & x < 0, \\ 1, & x \geq 0. \end{cases}$$

Note here how we dispensed with the boundary terms. They were trivially zero at $x = 0$ and were zero at $\pm\infty$ because ϕ has compact support; for example,

$$\big[x\,\phi(x)\big]\Big|_{-\infty}^{0} \coloneqq \lim_{L\to\infty} \big[x\,\phi(x)\big]\Big|_{-L}^{0} = \lim_{L\to\infty} L\,\phi(-L) = 0,$$

since $\phi \equiv 0$ outside some fixed interval. In conclusion, f' is simply g *in the sense of distributions*. The function g is known as the **signum function** (usually abbreviated by sgn) since it effectively gives the sign (either ± 1) of x; that is,

$$\operatorname{sgn}(x) := \begin{cases} -1, & x < 0, \\ 1, & x > 0, \\ 0, & x = 0. \end{cases}$$

Note that in this instance we took the value at $x = 0$ to be 0 but could just as well have taken it to be 45. The value at one point has no effect on *sgn as a distribution*.

Now, suppose we want to find f''? From our previous example of the Heaviside function, we know that doing this pointwise will **not** give us the right answer. So let us work in the sense of distributions. We have

$$\begin{aligned}
\langle (F_f)'', \phi \rangle &= \langle F_f, \phi'' \rangle \\
&= \int_{-\infty}^{\infty} |x|\,\phi''(x)\,dx \\
&= \int_{-\infty}^{0} (-x)\,\phi''(x)\,dx + \int_{0}^{\infty} x\,\phi''(x)\,dx \\
&\overset{\substack{\text{integration} \\ \text{by parts}}}{=} \int_{-\infty}^{0} \phi'(x)\,dx - \big[x\,\phi'(x)\big]\Big|_{-\infty}^{0} - \int_{0}^{\infty} \phi'(x)\,dx + \big[x\,\phi(x)\big]\Big|_{0}^{\infty} \\
&= \int_{-\infty}^{0} \phi'(x)\,dx - \int_{0}^{\infty} \phi'(x)\,dx \\
&= 2\phi(0).
\end{aligned}$$

Thus, $f'' = g' = 2\delta_0$ *in the sense of distributions.*

Example 5.4.3. Let

$$f(x) = \begin{cases} x^2, & x \geq 0, \\ 2x+3, & x < 0. \end{cases}$$

We find f' in the sense of distributions. Let $\phi \in C_c^\infty(\mathbb{R})$ be a test function; then we have

$$\begin{aligned}
\langle (F_f)', \phi \rangle &= -\langle F_f, \phi' \rangle \\
&= -\int_{-\infty}^{\infty} f(x)\, \phi'(x)\, dx \\
&= -\int_{-\infty}^{0} (2x+3)\, \phi'(x)\, dx \; - \; \int_0^\infty x^2\, \phi'(x)\, dx \\
&\overset{\substack{\text{integration} \\ \text{by parts}}}{=} \int_{-\infty}^{0} 2\,\phi(x)\, dx - \left[(2x+3)\,\phi(x)\right]\Big|_{-\infty}^{0} \\
&\qquad\qquad + \int_0^\infty 2x\,\phi(x)\, dx - \left[x^2\,\phi(x)\right]\Big|_0^\infty \\
&= \int_{-\infty}^{0} 2\,\phi(x)\, dx \; - \; 3\phi(0) \; + \; \int_0^\infty 2x\,\phi(x)\, dx \\
&= \int_{-\infty}^{\infty} g(x)\,\phi(x)\, dx - 3\phi(0),
\end{aligned}$$

where

$$g(x) = \begin{cases} 2x, & x \geq 0, \\ 2, & x < 0. \end{cases}$$

We used, as before, the fact that ϕ had compact support to dismiss the boundary terms at $\pm\infty$. Thus the derivative of f in the sense of distributions is the distribution $g - 3\delta_0$, or more precisely the sum of the two distributions, $F_g - 3\delta_0$. The $3\delta_0$ is a result of the jump discontinuity in f at $x = 0$.

To summarize this section:

- The notion of the derivative of a distribution allows us to differentiate any distribution to arrive at another distribution.
- We are primarily concerned with distributions which are either generated by functions or are constructed using delta functions. Whereas we can now consider the derivative of a distribution generated by **any** integrable function, we focused on distributions generated by piecewise smooth functions. In certain cases, this distributional derivative was simply the distribution generated by the pointwise derivative function, wherein we ignored the points at which the derivative function was undefined. In these cases, the function was either continuous or had a removable discontinuity at the singularity. The true singularity was in the first derivative and, hence, played no role in the distributional first derivative. In cases where the function had an essential jump discontinuity, the singularity was important and gave rise to a delta function in the distributional derivative.

In the examples presented here we have only addressed differentiating piecewise smooth functions in the sense of distributions.[10] One might thus be tempted to conclude that this new machinery of distributions was a bit of overkill. From the perspective of differentiating functions with discontinuities, the power and richness of this new machinery is really only seen in functions of several variables. We will begin to explore this in Chapter 9.

5.5. • Convergence in the Sense of Distributions

5.5.1. • The Definition.

We now define a notion of convergence for distributions and then apply this definition to sequences of functions interpreted as distributions.

Definition of Convergence of a Sequence of Distributions

Definition 5.5.1. A sequence F_n of distributions converges to a distribution F if

$$\langle F_n, \phi \rangle \xrightarrow{n\to\infty} \langle F, \phi \rangle \quad \text{for all } \phi \in C_c^\infty(\mathbb{R}).$$

In other words, for every $\phi \in C_c^\infty(\mathbb{R})$, the sequence of real numbers $\langle F_n, \phi \rangle$ converges to the real number $\langle F, \phi \rangle$. This convergence is written $F_n \to F$ in the sense of distributions.

Given a sequence of functions f_n, we may consider these functions as distributions and then apply the above definition. This leads to the following:

Definition 5.5.2. We say a sequence of locally integrable functions $f_n(x)$ **converges in the sense of distributions** to a distribution F if

$$\left\langle F_{f_n}, \phi \right\rangle \xrightarrow{n\to\infty} \langle F, \phi \rangle \quad \text{for all } \phi \in C_c^\infty(\mathbb{R}),$$

that is, if

$$\boxed{\int_{-\infty}^{\infty} f_n(x)\,\phi(x)\,dx \xrightarrow{n\to\infty} \langle F, \phi \rangle \quad \text{for all } \phi \in C_c^\infty(\mathbb{R}).} \tag{5.21}$$

Please note that occasionally we will leave out the "$n \to \infty$". It is understood!

A sequence of functions may converge in the sense of distributions to either another function or a more general distribution not generated by a function, e.g., a delta function. Specifically, $f_n(x)$ converging in the sense of distributions to a function $f(x)$ means that

$$\int_{-\infty}^{\infty} f_n(x)\,\phi(x)\,dx \xrightarrow{n\to\infty} \int_{-\infty}^{\infty} f(x)\,\phi(x)\,dx \quad \text{for all } \phi \in C_c^\infty(\mathbb{R}).$$

On the other hand, $f_n(x)$ converging in the sense of distributions to δ_0 means that

$$\int_{-\infty}^{\infty} f_n(x)\,\phi(x)\,dx \xrightarrow{n\to\infty} \phi(0) \quad \text{for all } \phi \in C_c^\infty(\mathbb{R}).$$

[10]**Warning:** In general, the distributional derivative of an integrable function can have a very complex structure, for example, a sample path of *Brownian motion* (cf. Section 7.3.4) which is continuous but nowhere differentiable, or what is called *the devil's staircase or the Cantor function*. These functions have a distributional derivative **but** this distributional derivative is neither generated by a function nor a combination of delta functions.

Automatic Distributional Convergence of Derivatives. The following result is a direct consequence of the definitions of differentiation and convergence in the sense of distributions (cf. Exercise **5.5**).

Proposition 5.5.1. *Suppose $\{F_n\}$ is a sequence of distributions which converges to a distribution F. Then*

$$F_n' \to F' \quad \text{in the sense of distributions.}$$

The same holds true for higher derivatives.

5.5.2. • Comparisons of Distributional versus Pointwise Convergence of Functions.

Suppose we are given a sequence of functions $f_n(x)$ and a limit function $f(x)$. In this scenario, **pointwise convergence** would mean the following: For every $x \in \mathbb{R}$, $f_n(x)$ as a sequence of numbers converges to the number $f(x)$. Pointwise convergence is, in principle, a stronger notion of convergence in the following sense:

Proposition 5.5.2. *Suppose $f_n(x)$ is a sequence of locally integrable functions which converges pointwise to a locally integrable function f. Further, suppose that there exists a locally integrable function g such that*

$$|f_n(x)| \leq g(x) \qquad \text{for all } n \text{ and for all } x \in \mathbb{R}.$$

Then f_n converges to f in the sense of distributions.

The proof is a simple consequence of Theorem A.8. As an example, the sequence of functions

$$f_n(x) = \frac{1}{n}\, e^{-\frac{x^2}{4n}}$$

converges pointwise to 0, and hence by Proposition 5.5.2, in the sense of distributions to the zero function $f(x) \equiv 0$. On the other hand, it is possible for a sequence of functions $f_n(x)$ to converge in the sense of distributions to function $f(x)$, but not in **any** pointwise sense. In fact, as we will discuss shortly in Section 5.5.5, the sequence $f_n(x) = \sin(nx)$ converges in the sense of distributions to the zero function $f(x) \equiv 0$; yet, for any fixed $x \in \mathbb{R}$ (different from integer multiples of π), the sequence $f_n(x) = \sin(nx)$ does not converge to any number.

5.5.3. • The Distributional Convergence of a Sequence of Functions to the Delta Function: Four Examples.

The sequence of spikes, defined by (5.10), was previously used in Section 5.2.2 to motivate the appearance of the delta function. In this and the next subsection, we address the distributional convergence of (5.10) and three other sequences of functions to δ_0, summarizing the essential properties shared by all of them and providing several proofs. For convenience of notation, let $\sigma_n := \frac{1}{n}$. Consider the following four sequences of functions:

1. **The sequence of hats**

$$f_n(x) = \begin{cases} \frac{n}{2} & \text{if} \quad |x| \le \sigma_n, \\ 0 & \text{otherwise.} \end{cases} \tag{5.22}$$

2. **The sequence of spikes**

$$f_n(x) = \left(n - n^2|x|\right)\chi_{[-\sigma_n,\sigma_n]} = \begin{cases} n - n^2 x & \text{if} \quad 0 < x < \sigma_n, \\ n + n^2 x & \text{if} \quad -\sigma_n < x \le 0, \\ 0 & \text{if} \quad |x| \ge \sigma_n. \end{cases} \tag{5.23}$$

3. **The sequence involving the derivative of arctangent**

$$f_n(x) = \frac{1}{\pi}\frac{\sigma_n}{x^2 + \sigma_n^2}. \tag{5.24}$$

4. **The sequence of Gaussians**

$$f_n(x) = \frac{1}{\sqrt{2\pi\,\sigma_n}}\, e^{-\frac{x^2}{2\sigma_n}}. \tag{5.25}$$

Figure 5.5 shows plots of these functions for increasing n.

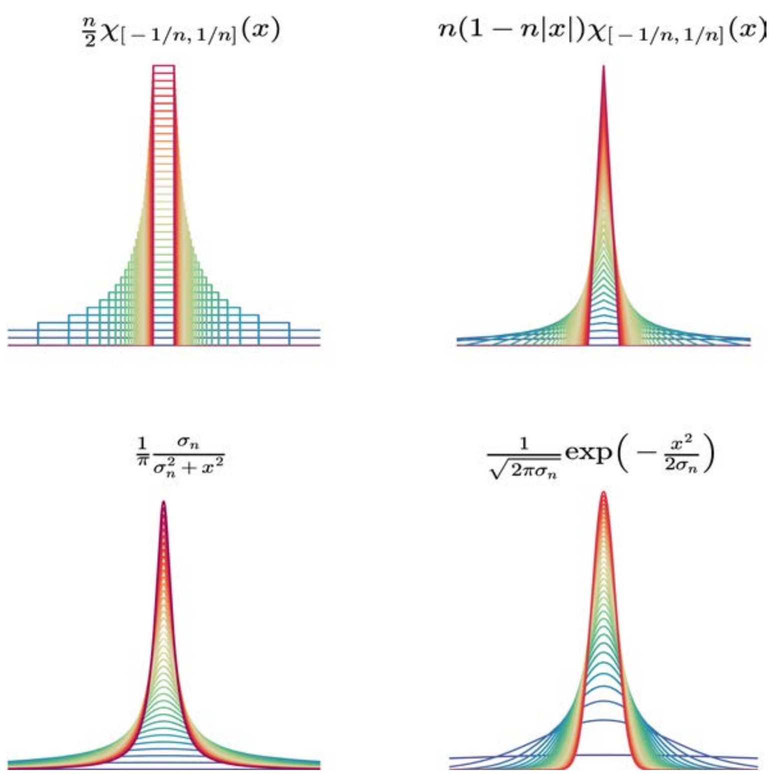

Figure 5.5. Plots for increasing n of the respective four sequences of functions which converge in the sense of distributions to δ_0.

All four sequences of functions converge in the sense of distributions to the distribution δ_0 (the delta function). This means, for any $\phi \in C_c^\infty(\mathbb{R})$, we have

$$\int_{-\infty}^{\infty} f_n(x)\,\phi(x)\,dx \;\to\; \phi(0) \qquad \text{as } n \to \infty. \tag{5.26}$$

Mathematicians sometimes call such a sequence **an approximation to the identity** and, as we previously pointed out (cf. the quote at the end of Section 5.2.2), such sequences were central to Dirac's vision of his delta function. There are **three properties** shared by all of these sequences that are responsible for their distributional convergence to δ_0:

Properties 5.5.1. *The three properties for $f_n(x)$ are:*

(1) ***Nonnegativity:*** The functions f_n are always nonnegative; i.e.,

$$\text{for all } x \in \mathbb{R} \text{ and } n \in \mathbb{N}, \text{ we have } f_n(x) \geq 0.$$

(2) ***Unit mass (area under the curve):*** The functions f_n all integrate to 1. That is, for each $n = 1, 2, 3, \ldots$

$$\int_{-\infty}^{\infty} f_n(x)\,dx \;=\; 1.$$

(3) ***Concentration at 0:*** As $n \to \infty$, the functions concentrate their **mass** (area under the curve) around $x = 0$. This loose condition can be made precise (see Exercise **5.15**).

Observe from Figure 5.5 that each of our four sequences appears to satisfy these three properties. We claim that Properties 5.5.1 are sufficient for (5.26) to hold. Let us first intuitively argue why. Properties (1) and (3) imply that if we consider $f_n(x)$ for very large n, there is a very small $\delta > 0$ such that the mass of f_n (i.e., the area under the curve) is concentrated in $[-\delta, \delta]$. Hence, by the second property, for large n we have

$$\int_{-\infty}^{\infty} f_n(x)\,dx \;\sim\; \int_{-\delta}^{\delta} f_n(x)\,dx \;\sim\; 1,$$

and for any continuous function ϕ,

$$\int_{-\infty}^{\infty} f_n(x)\,\phi(x)\,dx \;\sim\; \int_{-\delta}^{\delta} f_n(x)\,\phi(x)\,dx.$$

Here, again we use $\sim$ to loosely mean *close to* or *approximately equal to*. Thus for large n, we may focus our attention entirely on the small interval $x \in [-\delta, \delta]$. Since ϕ is continuous, it is essentially constant on this small interval; in particular, $\phi(x)$ on the interval $[-\delta, \delta]$ is close to $\phi(0)$. Thus, for n large,

$$\begin{aligned}\int_{-\infty}^{\infty} f_n(x)\,\phi(x)\,dx \;\sim\; \int_{-\delta}^{\delta} f_n(x)\,\phi(x)\,dx \quad &\sim \quad \int_{-\delta}^{\delta} f_n(x)\,\phi(0)\,dx \\ &\sim \quad \phi(0)\int_{-\delta}^{\delta} f_n(x)\,dx \\ &\sim \quad \phi(0).\end{aligned} \tag{5.27}$$

For any particular sequence with these properties, one can make these statements precise (i.e., formulate a proof) via an "ϵ vs. N" argument. In fact, a proof for the hat sequence (5.22) follows directly from the Averaging Lemma (Exercise **5.6**). We will now present the "ϵ vs. N" proofs for two sequences: the spike functions (defined by (5.23)) and sequence involving the derivative of arctangent (defined by (5.24)). While the above heuristics for this convergence are indeed rather convincing, we recommend all students read the first proof — even if you have not taken a class in mathematical analysis. Both proofs are actually relatively straightforward and serve as excellent examples to digest and appreciate the estimates behind a convergence proof.

5.5.4. ϵ vs. N Proofs for the Sequences (5.23) and (5.24). We begin with the sequence of spikes (5.23); here, the proof is straightforward since the $f_n(x)$ are identically zero on $|x| \geq \sigma_n = \frac{1}{n}$.

Theorem 5.1. *Let $f_n(x)$ be defined by (5.23). Then $f_n \to \delta_0$ in the sense of distributions.*

Proof. Fix any $\phi \in C_c^\infty$. We need to prove that

$$\int_{-\infty}^{\infty} f_n(x)\,\phi(x)\,dx \xrightarrow{n\to\infty} \phi(0). \tag{5.28}$$

In other words, the sequence of real numbers

$$\int_{-\infty}^{\infty} f_n(x)\,\phi(x)\,dx$$

converges to $\phi(0)$. To be precise, this means for any $\epsilon > 0$, there exists $N \in \mathbb{N}$ such that

$$\text{if } n > N, \qquad \text{then} \qquad \left|\left(\int_{-\infty}^{\infty} f_n\,\phi\,dx\right) - \phi(0)\right| < \epsilon.$$

To this end, we first note that since the functions f_n all integrate to 1, we have (rather trivially) that

$$\phi(0) = \phi(0)\,1 = \phi(0)\int_{-\infty}^{\infty} f_n(x)\,dx = \int_{-\infty}^{\infty} f_n(x)\,\phi(0)\,dx.$$

Thus,

$$\begin{aligned}\left|\int_{-\infty}^{\infty} f_n(x)\,\phi(x)\,dx - \phi(0)\right| &= \left|\int_{-\infty}^{\infty} f_n(x)\,\phi(x)\,dx - \int_{-\infty}^{\infty} f_n(x)\,\phi(0)\,dx\right| \\ &= \left|\int_{-\infty}^{\infty} f_n(x)\,(\phi(x)-\phi(0))\,dx\right| \\ &\leq \int_{-\infty}^{\infty} f_n(x)\,|\phi(x)-\phi(0)|\,dx.\end{aligned} \tag{5.29}$$

Since ϕ is continuous, **there exists** $\delta > 0$ **such that** if $|x| < \delta$, then $|\phi(x) - \phi(0)| < \epsilon$. Now **choose** $N \in \mathbb{N}$ sufficiently large such that $1/N < \delta$. Then if $n > N$, we have

$$\sigma_n = \frac{1}{n} < \frac{1}{N} < \delta$$

and

$$\begin{aligned} \int_{-\infty}^{\infty} f_n(x)|\phi(x)-\phi(0)|dx &= \int_{-\frac{1}{n}}^{\frac{1}{n}} f_n(x)|\phi(x)-\phi(0)|\,dx \\ &< \epsilon \underbrace{\int_{-\frac{1}{n}}^{\frac{1}{n}} f_n(x)\,dx}_{=1} \\ &= \epsilon. \end{aligned}$$

Combining this with (5.29), we have for all $n > N$

$$\left| \left(\int_{-\infty}^{\infty} f_n(x)\phi(x)\,dx \right) - \phi(0) \right| < \epsilon. \qquad \square$$

In the previous sequence of functions, the functions become identically zero on more and more of the domain as n tends to infinity: As a result, the proof never used the precise formula for f_n; that is, the exact shape of the spike was irrelevant. This is a rather strong form of the concentration property (3). Many sequences of functions still concentrate their mass around zero but are never actually equal to zero at a particular x. This is the case with (5.24) and (5.25) where $f_n(x) > 0$ for any n and any $x \in \mathbb{R}$. However, just by plotting $f_5(x)$, $f_{10}(x)$, and $f_{50}(x)$ for either case, one can readily see the concentration of mass (area under the curve). We now present a proof for (5.24).

First, note that property (2) holds; indeed,[11] for any n we have

$$\int_{-\infty}^{\infty} \frac{\sigma_n}{x^2+\sigma_n^2}\,dx = \sigma_n \int_{-\infty}^{\infty} \frac{1}{x^2+\sigma_n^2}\,dx = \sigma_n \frac{1}{\sigma_n} \tan^{-1}\left(\frac{x}{\sigma_n}\right)\Big|_{-\infty}^{\infty} = \frac{\pi}{2} - (-\frac{\pi}{2}) = \pi.$$

Hence, the same heuristics behind (5.27) suggest the distributional convergence to δ_0. The following theorem and proof present the precise argument which, as you will observe, is only slightly more involved than the proof of Theorem 5.1.

Theorem 5.2. *Let $f_n(x)$ be defined by* (5.24)*. Then $f_n \to \delta_0$ in the sense of distributions.*

Proof. Fix any $\phi \in C_c^\infty$. We need to prove that for any $\epsilon > 0$, there exists $N \in \mathbb{N}$ such that

$$\text{if } n > N, \qquad \text{then} \qquad \left| \int_{-\infty}^{\infty} f_n(x)\,\phi(x)\,dx - \phi(0) \right| < \epsilon.$$

Let $\epsilon > 0$. Due to the fact that the functions f_n are nonnegative and always integrate to 1 (i.e., satisfy properties (1) and (2)), we can repeat the initial steps leading up to (5.29) verbatim to find

$$\left| \int_{-\infty}^{\infty} f_n(x)\,\phi(x)\,dx - \phi(0) \right| \le \int_{-\infty}^{\infty} f_n(x)\,|\phi(x)-\phi(0)|\,dx.$$

[11] Recall from calculus the fact that for any $a > 0$,

$$\int \frac{1}{x^2+a^2}\,dx = \frac{1}{a}\tan^{-1}\left(\frac{x}{a}\right) + C.$$

Now comes the extra work associated with the fact that $f_n(x)$ is not identically zero on most of $\mathbb{R}$. Since ϕ is continuous, **there exists** $\delta > 0$ **such that** if $|x| < \delta$, then

$$|\phi(x) - \phi(0)| \; < \; \frac{\epsilon}{2}.$$

You will see shortly why we need the one-half on the right.

Splitting the integral into two pieces, we have

$$\int_{-\infty}^{\infty} f_n(x)\,|\phi(x) - \phi(0)|\,dx \; = \; \int_{-\delta}^{\delta} f_n(x)\,|\phi(x) - \phi(0)|\,dx + \int_{\{|x|\ge\delta\}} f_n(x)\,|\phi(x) - \phi(0)|\,dx. \tag{5.30}$$

For the first integral, we note that with our choice of δ we have

$$\int_{-\delta}^{\delta} f_n(x)\,|\phi(x) - \phi(0)|\,dx \; < \; \frac{\epsilon}{2}\int_{-\delta}^{\delta} f_n(x)\,dx \; \le \; \frac{\epsilon}{2}. \tag{5.31}$$

In the last inequality, we used the fact that f_n was positive and integrated over $\mathbb{R}$ to 1; hence, the integral of f_n on any subinterval must be less than or equal to 1.

For the second integral, we need to work a little harder. First, note that since ϕ has compact support and is continuous, it must be bounded. This means there exists a constant C such that $|\phi(x)| \le C$ for all $x \in \mathbb{R}$. Hence,

$$|\phi(x) - \phi(0)| \; \le \; |\phi(x)| + |\phi(0)| \; \le \; 2C,$$

and

$$\int_{\{|x|\ge\delta\}} f_n(x)\,|\phi(x) - \phi(0)|\,dx \; \le \; 2C\int_{\{|x|\ge\delta\}} f_n(x)\,dx. \tag{5.32}$$

Since for each n the function $f_n(x)$ is an even function,

$$2C\int_{\{|x|\ge\delta\}} f_n(x)\,dx \; = \; 4C\int_{\delta}^{\infty} f_n(x)\,dx \; = \; \frac{4C}{\pi}\int_{\delta}^{\infty} \frac{\sigma_n}{x^2 + \sigma_n^2}\,dx. \tag{5.33}$$

Note that this is the first time we are actually writing down the precise formula for f_n. Until now, all we required were properties (1) and (2). Now we need to interpret property (3) by means of choosing n large enough that this tail integral in (5.33) can be made small. How small? Less than $\epsilon/2$. To this end, note that

$$\int_{\delta}^{\infty} \frac{\sigma_n}{x^2 + \sigma_n^2}\,dx \; = \; \frac{\pi}{2} \; - \; \tan^{-1}\left(\frac{\delta}{\sigma_n}\right).$$

Now comes a slightly delicate argument. We have fixed an ϵ and, consequently, have established the existence of a δ (depending on ϵ) which gave the estimate (5.31). With this fixed δ, we note that since

$$\lim_{\theta\to\infty} \tan^{-1}\theta \; = \; \frac{\pi}{2} \qquad \text{and} \qquad \lim_{n\to\infty} \sigma_n \; = \; 0,$$

we can make $\tan^{-1}\left(\frac{\delta}{\sigma_n}\right)$ as close as we like to $\frac{\pi}{2}$ by choosing n sufficiently large. In our present context, this means we can **choose** N **such that** if $n > N$, then

$$\frac{\pi}{2} \; - \; \tan^{-1}\left(\frac{\delta}{\sigma_n}\right) \; < \; \frac{\pi}{4C}\,\frac{\epsilon}{2}. \tag{5.34}$$

Hence, combining (5.32)–(5.34), we find that for $n > N$,

$$\int_{\{|x|\geq\delta\}} f_n(x)\,|\phi(x)-\phi(0)|\,dx \;<\; \frac{\epsilon}{2}.$$

Bringing this together with (5.30) and (5.31), we have shown the following: For every $\epsilon > 0$, we can find (there exits) N such that

$$\text{if } n > N, \text{ then } \quad \left|\int_{-\infty}^{\infty} f_n(x)\,\phi(x)\,dx \;-\; \phi(0)\right| < \epsilon.$$

Note that the parameter δ was needed only as an intermediate variable in the course of the proof. □

This technique of splitting an integral into multiple parts and controlling each part to be less than the appropriate fraction of epsilon is quite common in analysis. The method of control (i.e., ensuring smallness) for each term can be distinct but must be mutually compatible. In this proof, our two methods of control were (a) continuity of the test function at $x = 0$ and (b) the precise notion that as n gets arbitrarily large, more and more of the unit mass of our functions gets pulled into the origin; in the previous proof, this entailed controlling the residual mass of the tails, $\pi/2 - \tan^{-1}(\delta/\sigma_n)$. These two methods were mutually compatible because after fixing a neighborhood of the origin of a given length 2δ in which we controlled the fluctuations of ϕ, we were able to control the residual mass at the tail determining the cutoff index, N, in terms of our window length δ.

In a very similar fashion (cf. Exercise **5.14**) one can prove that the sequence of Gaussians (5.25) converges to δ_0 in the sense of distributions as $n \to \infty$. One may ask why it is necessary to provide individual proofs for each particular sequence; why not just prove that the three properties of Properties 5.5.1 are sufficient for distributional convergence to the delta function? The issue here is that the third condition (concentration at 0) needs to be made precise. Exercise **5.15** addresses a general result in this direction. Another general result is the following theorem whose proof is left for Exercise **5.16**.

Theorem 5.3. *Let $f(x)$ be **any** nonnegative integrable function on $\mathbb{R}$ which integrates to one. For $n = 1, 2, \ldots$ define*

$$f_n(x) \;:=\; n\,f(nx).$$

Then

$$f_n \longrightarrow \delta_0 \quad \textit{in the sense of distributions.}$$

5.5.5. • The Distributional Convergence of $\sin nx$**.** We will now present an example of a sequence of functions which converge in the sense of distributions to another function (in fact the function which is identically zero). What is new here is that this sequence of functions does **not** converge in any pointwise or regular sense that we have seen before. This is an important example and it is worth spending some time to digest it. Consider the sequence of functions

$$\boxed{f_n(x) = \sin(nx).}$$

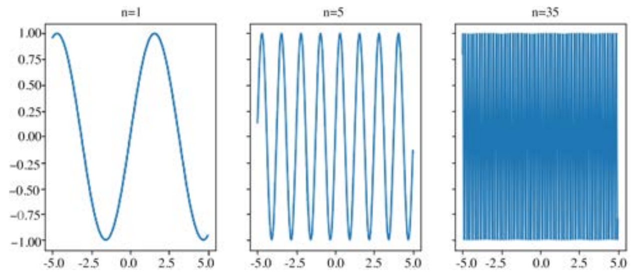

Figure 5.6. Plots of $\sin nx$ for $n = 1, 5, 35$.

What happens to f_n as $n \to \infty$? The functions oscillate between 1 and -1, increasingly as n gets larger (i.e., their period of oscillation tends to 0). In particular, except when x is an integer multiple of π, the sequence of numbers $f_n(x)$ does not converge[12]. So there exists no pointwise limit. On the other hand, what happens in the sense of distributions? In other words, if ϕ is a test function, what happens to

$$\int_{-\infty}^{\infty} \sin(nx)\, \phi(x)\, dx? \tag{5.35}$$

Plots of $\sin nx$ for $n = 1, 5, 35$ are given in Figure 5.6. What is important to note is that there are two fundamental properties of $\sin(nx)$ which stem from the mother function $\sin x$:

- **Periodicity:** The function $\sin x$ is periodic with period 2π. Hence, the function $\sin(nx)$ has period $\frac{2\pi}{n}$. As $n \to \infty$, this period tends to 0 and, hence, the larger the value of n the more oscillations of $\sin(nx)$ per unit length in x.
- **Averaging out over a period:** The function $\sin x$ integrates to 0 ("averages out") over any interval of length 2π. Note that this is not only for, say, the interval $[0, 2\pi]$, but for any interval $[a, b]$ with $b - a = 2\pi$. Hence, the function $\sin(nx)$ integrates to 0 (averages out) over any interval of size $\frac{2\pi}{n}$.

So in (5.35), we are multiplying a function, which oscillates more and more rapidly between 1 and -1, by ϕ and then integrating. For very large n, think of blowing up on the scale of the period $\frac{2\pi}{n}$, and then focus for a moment on one such interval. The continuous function ϕ is essentially constant on this tiny interval (period); hence, if we integrate $\phi(x) \sin nx$ over this interval, we are effectively integrating a constant times $\sin\, nx$ over a period interval, where the end result is "almost" *zero*. The full integral (5.35) can be viewed as a sum over all the period intervals and, consequently, as $n \to \infty$, (5.35) "should" tend to 0.

Does this sound convincing? Well, be careful! Yes, it is true that the integrals over the smaller and smaller periods will get closer to 0 but there are more and more of them. If this argument convinced you that the sum of all of them should also tend to 0, then we fear that we have convinced you that every Riemann integral is zero! In the

[12]Test this on your computer: Compute $\sin n$ (so $x = 1$) for $n = 1, \dots, 1{,}000$ and see if you detect convergence.

present case, with ϕ smooth and compactly supported, one can quantify the analysis more precisely by estimating the size of the integral over each period subinterval of size $\frac{2\pi}{n}$ versus the number of such small intervals (essentially, a constant times n). While this calculation would indeed yield that the full integral (5.35) tends to 0 as $n \to \infty$, here we provide a very quick and slick proof which only uses integration by parts!

Theorem 5.4.

$$\boxed{\sin(nx) \xrightarrow{n\to\infty} 0 \quad \textit{in the sense of distributions.}}$$

Note that the same distributional convergence holds for the sequence $\cos(nx)$.

Proof. Fix $\phi \in C_c^\infty(\mathbb{R})$. Our goal is to prove that for every $\epsilon > 0$, there exists N such that if $n \geq N$, then

$$\left| \int_{-\infty}^{\infty} \sin(nx)\,\phi(x)\,dx \right| < \epsilon.$$

We first note that since ϕ has compact support, there exists some $L > 0$ such that

$$\left| \int_{-\infty}^{\infty} \sin(nx)\,\phi(x)\,dx \right| = \left| \int_{-L}^{L} \sin(nx)\,\phi(x)\,dx \right|,$$

with $\phi(\pm L) = 0$. Let $C = \max_{x\in[-L,L]} |\phi'(x)|$. Then we have,

$$\begin{aligned}
\left| \int_{-L}^{L} \sin(nx)\,\phi(x)\,dx \right| &= \left| \int_{-L}^{L} \left(-\frac{\cos nx}{n}\right)' \phi(x)\,dx \right| \\
&\overset{\text{integrations by parts}}{=} \left| \frac{1}{n} \int_{-L}^{L} \cos nx\, \phi'(x)\,dx \right| \\
&\leq \frac{1}{n} \int_{-L}^{L} |\cos nx\, \phi'(x)|\; dx \\
&\leq \frac{1}{n} \int_{-L}^{L} |\phi'(x)|\; dx \;\leq\; \frac{2LC}{n}.
\end{aligned}$$

Thus we simply need to choose $N > \frac{2LC}{\epsilon}$ and we are done. □

In this proof, we conveniently exploited two facts about ϕ: compact support and C^1. However, the result is true for ϕ with far less restrictive assumptions and goes by the name of **the Riemann-Lebesgue Lemma**. It is connected with Fourier analysis (the Fourier transform and Fourier series) and will resurface in Chapters 6 and 11.

5.5.6. • The Distributional Convergence of the Sinc Functions and the Dirichlet Kernel: Two Sequences Directly Related to Fourier Analysis. Here we look at two more examples, without delving into their proofs, which are fundamentally related to the Fourier transform (Chapter 6) and to Fourier series (Chapter 11), respectively. They are, again, examples of a sequence of functions converging in the sense of distributions to the delta function; however, the functions now take both negative and positive values. They are, respectively, the (rescaled and normalized) **Sinc functions** and what is known as **the Dirichlet kernel**:

$$S_n(x) \ := \frac{\sin(nx)}{\pi x}, \qquad K_n(x) \ := \frac{\sin\left(\left(n+\frac{1}{2}\right)x\right)}{\sin\left(\frac{x}{2}\right)}.$$

While it is not immediately obvious, we will see later in Section 11.5.2 that for each n, $K_n(x)$ is a periodic function with period 2π. Also note that both S_n and K_n have a removable discontinuity at $x = 0$ (in the case of K_n, at all integer multiples of 2π). Hence redefining $S_n(0) = \frac{n}{\pi}$ and K_n at any integer multiple of 2π to be $1 + 2n$ yields smooth functions of $x \in \mathbb{R}$.

Let us begin with S_n, plots of which are presented in Figure 5.7. Again, we ask what happens as $n \to \infty$? With a bit of work, one can verify that for each n, the following holds:

$$\int_{-\infty}^{\infty} S_n(x)\,dx \ = \ 1. \tag{5.36}$$

However, in this instance, there is a major difference with our previous examples (5.22) –(5.25). The functions S_n are **no longer nonnegative** and, consequently, there are cancellation effects in (5.36). Figure 5.7 suggests that there is concentration near $x = 0$. Moving away from zero, the functions oscillate more and more but with an amplitude which tends to 0. With some mathematical analysis, one can indeed prove that

$$S_n(x) \xrightarrow{n\to\infty} \delta_0 \qquad \text{in the sense of distributions.}$$

We will encounter these sinc functions again in the next chapter, Chapter 6, on the Fourier transform. Moreover, Exercise **6.25** asks for a direct proof of the distributional convergence of the sinc functions S_n.

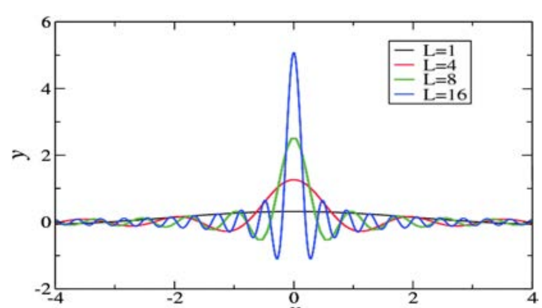

Figure 5.7. Sinc functions $S_n(x)$ for $n = L = 1, 4, 8, 16$. Note that the values at $x = 0$ are taken to be the $\lim_{x\to 0} S_n(x) = \frac{n}{\pi}$.

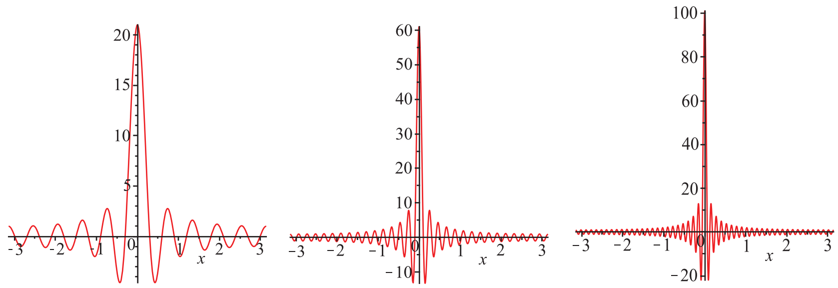

Figure 5.8. Plots of the Dirichlet kernel $K_n(x)$ on $[-\pi, \pi]$ for $n = 10, n = 30$, and $n = 50$ from left to right. Note that the values at $x = 0$ are taken to be the $\lim_{x\to 0} K_n(x) = 1 + 2n$.

The functions K_n are also no longer nonnegative and their convergence is even more subtle, entailing the focus of Section 11.5. First off, note that these functions are all periodic with period 2π, and hence we may restrict our attention to $x \in [-\pi, \pi]$. With a bit of simple trigonometry and algebra (cf. Section 11.5), one can verify that over any period, for each n the functions K_n integrate to 2π; i.e.,

$$\int_{-\pi}^{\pi} K_n(x)\, dx \;=\; 2\pi.$$

Figure 5.8 shows plots of K_n on $[-\pi, \pi]$ and, again, there appears to be concentration around $x = 0$ as n gets larger. As with S_n, there are increasing oscillations in the (here, finite) tails. However, the plots for K_n are deceiving, in that at first sight they seem to suggest that as $n \to \infty$, the amplitude of oscillation in the finite tails is tending to 0. This was the case for the infinite tails of S_n, but **not** for K_n. For K_n, the tail amplitudes appear to be tending to 0 because of the scaled y-axis. In fact, as n increases, the kernel's tail oscillates more and more frequently but at a fixed positive amplitude. As $n \to \infty$, these tails **do** tend to 0 in the sense of distributions (similar to the distributional convergence of $\sin(nx)$) but **not** in any pointwise sense. In the end, we have

$$\boxed{K_n(x) \xrightarrow{n\to\infty} 2\pi\delta_0 \quad \text{in the sense of distributions on } (-\pi, \pi),}$$

and this will be proven in Section 11.5. Note here that we are restricting our attention to test functions defined on the interval $(-\pi, \pi)$ (cf. Section 5.7.1). Instead, if we wanted the statement for test functions $\phi \in C_c^\infty(\mathbb{R})$, we would have distributional convergence to an infinite sum of delta functions with respective concentrations at $x = 2\pi m$ for all $m \in \mathbb{Z}$.

It is important to emphasize that the convergence of K_n is far more subtle than in **any** of the previous examples. Indeed, in the earlier examples, we only needed the test functions to be continuous. For K_n, continuity alone is insufficient for the test functions, and this fact is responsible for making the pointwise convergence analysis of Fourier series an amazingly difficult and delicate subject.

5.6. Dirac's Intuition: Algebraic Manipulations with the Delta Function

> "*Dirac is of course fully aware that the δ function is not a well-defined expression. But he is not troubled by this for two reasons. First, as long as one follows the rules governing the function (such as using the δ function only under an integral sign, meaning in part not asking the value of a δ function at a given point), then no inconsistencies will arise. Second, the δ function can be eliminated, meaning that it can be replaced with a well-defined mathematical expression. However, the drawback in that case is, according to Dirac, that the substitution leads to a more cumbersome expression that obscures the argument. In short, when pragmatics and rigor lead to the same conclusion, pragmatics trumps rigor due to the resulting simplicity, efficiency, and increase in understanding.*"
>
> From the article by Fred Kronz and Tracy Lupher: *Quantum Theory and Mathematical Rigor*, The Stanford Encyclopedia of Philosophy (Fall 2019 Edition), Edward N. Zalta (ed.), `https://plato.stanford.edu/archives/fall2019/entries/qt-nvd/`.

We have gone to great lengths to emphasize why the delta function is not a function and agreed not to write expressions like "$\delta_0(x - a)$" to represent a delta function concentrated at $x = a$, but rather δ_a. However, we strongly feel that it is not contradictory to also claim that there is merit and insight in Dirac's informal approach, and there are times when "pragmatics trumps rigor". In the physics, engineering, and even mathematics literature, one often encounters informal calculations which treat the delta function as a true function with argument "(x)" of real numbers. With the correct guidance, such informal calculations will steer us to make correct conclusions. This will be the case, for example, with the higher-dimensional delta function and its use in finding the Green's function with Neumann boundary conditions for a ball (cf. Section 10.5.2) and in formulating Green's functions for the wave equation (cf. Sections 13.2.3 and 13.2.4). See also Section 9.7.3 and Exercise **10.22**.

In this section, we review a few informal algebraic calculations with delta functions based upon manipulations of the argument. We will show that, with the right guidance, these manipulations are justifiable and, hence, acceptable (or, at the very least, tolerable). We repeat that, in this text, **we will place within quotation marks all such informal expressions**.

5.6.1. Rescaling and Composition with Polynomials.

Example 1: Consider first the following:

$$\text{“}\delta_0(2x)\text{”}.$$

While we will **not** give this expression **precise meaning as a distribution**, let us at least try to make some sense of it. The delta function is basically unitary mass concentration at $x = 0$. Since x and $2x$ are both 0 at $x = 0$, does that mean "$\delta_0(2x) = \delta_0(x)$"?

No, this type of pointwise reasoning will get us nowhere with the delta function. In fact, we will now argue that

$$\text{“}\delta_0(2x) = \frac{1}{2}\delta_0(x)\text{”}. \tag{5.37}$$

What is important to remember is that the delta function concentrates *unit mass* at a point and **mass** is captured by **integration**. The best way to proceed is to follow Dirac's lead (cf. the quote at the end of Section 5.2.2) and to view the delta function as a limit (in the sense of distributions) of a sequence of nonnegative functions f_n such that the following hold:

- For all n, $\int_{-\infty}^{\infty} f_n(x)\,dx = 1$.
- The supports of the f_n (i.e., the set of x in the domain for which $f(x) > 0$) get smaller and smaller as $n \to \infty$ and concentrate about 0.

Fix one of these sequences, say, (5.10), and let us **think of** the delta function "$\delta_0(x)$" as the true function $f_n(x)$ for n very large. We can then address "$\delta_0(2x)$" by considering $f_n(2x)$ for n very large; more precisely, what does $f_n(2x)$ converge to in the sense of distributions as $n \to \infty$? Let $\phi \in C_c^\infty(\mathbb{R})$ be a test function. Then by letting $y = 2x$, we have

$$\int_{-\infty}^{\infty} f_n(2x)\,\phi(x)\,dx = \frac{1}{2}\int_{-\infty}^{\infty} f_n(y)\,\phi(y/2)\,dy.$$

Since $f_n \to \delta_0$ in the sense of distributions and noting that $\psi(y) := \phi(y/2)$ is also a test function in $C_c^\infty(\mathbb{R})$, we have

$$\frac{1}{2}\int_{-\infty}^{\infty} f_n(y)\,\psi(y)\,dy \longrightarrow \frac{1}{2}\psi(0) \quad \text{as} \quad n \to \infty.$$

Since $\frac{1}{2}\psi(0) = \frac{1}{2}\phi(0)$, we have shown that for any test function ϕ,

$$\int_{-\infty}^{\infty} f_n(2x)\,\phi(x)\,dx \longrightarrow \frac{1}{2}\phi(0) \quad \text{as} \quad n \to \infty.$$

But this means that $f_n(2x)$ for a very large n behaves like $\frac{1}{2}\delta_0$, and so informally (5.37) holds true.

We can generalize this observation with respect to the composition of a delta function with certain C^1 functions. To this end, let $g(x)$ be a C^1 function which is zero exactly at $x = a$ with $g'(a) \neq 0$. Then we claim,

$$\text{“}\delta_0(g(x)) = \frac{1}{|g'(a)|}\delta_0(x-a)\text{”}. \tag{5.38}$$

The details follow exactly the same steps as above for $g(x) = 2x$ and are left as an exercise (Exercise **5.22**). The assumption that $g'(a) \neq 0$, which insured that g was one-to-one in a neighborhood of a, is critical in order to make sense of "$\delta_0(g(x))$". For example, we **cannot** make sense of "$\delta_0(x^2)$". Try to!

Example 2: Let us now justify the following informal equality for $a < b$:

$$\text{“}\delta_0[(x-a)(x-b)] = \frac{1}{|a-b|}\left(\delta_0(x-a) + \delta_0(x-b)\right)\text{”}.$$

To this end, consider any sequence $f_n(\cdot)$ which converges to δ_0 in the sense of distributions. Our goal is to determine what $f_n((x-a)(x-b))$ converges to in the sense of distributions. This means we need to determine what happens to

$$\int_{-\infty}^{\infty} f_n((x-a)(x-b))\,\phi(x)\,dx,$$

as $n \to \infty$ for any test function ϕ. In particular, we want to show it tends to

$$\frac{\phi(a)}{|a-b|} + \frac{\phi(b)}{|a-b|}.$$

If we let $y = (x-a)(x-b)$, we have an issue here because unlike with the example "$\delta_0(2x)$", we cannot solve uniquely for x as a function of y. Let $x = c$ be the point at which the curve

$$y(x) = (x-a)(x-b)$$

attains its minimum (sketch this curve). On either interval $x \in (-\infty, c)$ or (c, ∞), we can solve uniquely for x as a function of y. Let us call these two branches $x_1(y)$ and $x_2(y)$, respectively. Since $f_n(y)$ will concentrate its support (mass) when y is close to 0, $f_n(y(x))$ on $x \in (-\infty, c)$ will concentrate around $x = a$. Similarly, $f_n(y(x))$ on $x \in (c, \infty)$ will concentrate around $x = b$.

We now have

$$\begin{aligned}\int_{-\infty}^{\infty} f_n((x-a)(x-b))\,\phi(x)\,dx &= \int_{-\infty}^{c} f_n((x-a)(x-b))\,\phi(x)\,dx \\ &\quad + \int_{c}^{\infty} f_n((x-a)(x-b))\,\phi(x)\,dx.\end{aligned}$$

With the substitution $y = (x-a)(x-b)$, $dy = (2x-(a+b))dx$ in each of the two integrals, we focus on the first integral to find that

$$\begin{aligned}\int_{-\infty}^{c} f_n((x-a)(x-b))\,\phi(x)\,dx &= \int_{+\infty}^{y(c)} f_n(y)\frac{\phi(x_1(y))}{2x_1(y)-(a+b)}\,dy \\ &= -\int_{y(c)}^{+\infty} f_n(y)\frac{\phi(x_1(y))}{2x_1(y)-(a+b)}\,dy.\end{aligned}$$

Noting that $y(c) < 0$ and, hence, viewing

$$\frac{\phi(x_1(y))}{2x_1(y)-(a+b)}$$

as a test function in y, the effect of the integral as $n \to \infty$ will be to pick out the value of the test function at $y = 0$. Since by definition $x_1(0) = a$, this indicates that the integral will approach

$$-\frac{\phi(a)}{a-b} = \frac{\phi(a)}{b-a} = \frac{\phi(a)}{|a-b|}.$$

The analogous argument on the second integral yields the convergence to $\frac{\phi(b)}{|a-b|}$.

5.6.2. Products of Delta Functions in Different Variables. It is also commonplace to use **products** of delta functions in different variables[13] in order to denote the delta function in more than one variable. For example, in two dimensions we can define (cf. Chapter 9) the delta function $\delta_{\mathbf{0}}$ where $\mathbf{0} = (0,0)$ over test functions ϕ on $\mathbb{R}^2$ by

$$\langle \delta_{\mathbf{0}}, \phi(x,y)\rangle = \phi(0,0) \qquad \text{for all } \phi \in C_c^\infty(\mathbb{R}^2).$$

The distribution $\delta_{\mathbf{0}}$ is often written as

$$\text{“}\delta_{\mathbf{0}} = \delta_0(x)\,\delta_0(y)\text{”}. \tag{5.39}$$

In Exercise **9.2** we will present a precise definition of such **direct products of distributions**. Here, let us just argue that (5.39) is *intuitively correct*.

Let $f_n(x)$ be the particular approximation of δ_0 given by (5.10). We also consider $f_n(y)$ as an approximation of δ_0 in the y variable. We trivially extend these two sequences of functions on $\mathbb{R}$ to sequence of functions on $\mathbb{R}^2$ by

$$g_n(x,y) := f_n(x) \qquad \text{and} \qquad h_n(x,y) := f_n(y).$$

Now, consider the product sequence

$$g_n(x,y)h_n(x,y).$$

By plotting this product function for $n = 10, 50, 100$, you should be able to convince yourself that $g_n(x,y)h_n(x,y) \longrightarrow \delta_{\mathbf{0}}$ in the sense of distributions. Hence, the product is a good approximation to $\delta_{\mathbf{0}}$, the 2D delta function.

In Section 9.7.3, we will solve the 1D wave equation with an instantaneous source concentrated at $x = 0$ and $t = 0$. This equation is often written in the informal, but customary, form

$$\text{“}u_{tt} - cu_{xx} = \delta(x)\delta(t)\text{”}.$$

5.6.3. Symmetry in the Argument. It is also common, and often quite useful, to invoke intuitive symmetry properties of the delta function. For example, one often encounters equations such as

$$\text{“}\delta_0(x) = \delta_0(-x)\text{”} \qquad \text{and} \qquad \text{“}\delta_0(x-y) = \delta_0(y-x)\text{”}.$$

Since, loosely speaking, the delta function is 0 except when the argument is 0, these two equations seem plausible. Let us focus on the first equation, "$\delta_0(x) = \delta_0(-x)$". First note that this equation is justified by viewing the delta function as a limit of functions f_n (for example (5.10)) since each of the functions f_n is itself symmetric. Hence we would expect that symmetry is preserved in any reasonable limit. However, unlike in the previous examples, we can make complete sense of this equality in the sense of distributions. To this end, given any test function ϕ, note that

$$\phi^-(x) := \phi(-x)$$

is also a test function. For any distribution F, we define the distribution F^- as

$$\langle F^-, \phi\rangle := \langle F, \phi^-\rangle.$$

[13] Note that we are **not** considering products of distributions over **the same** underlying independent variable. In general, we **cannot** make sense of such products.

As with differentiation, this presents a reasonable definition if it is the case that for any integrable function f,

$$F_{f^-} = (F_f)^-.$$

You can readily verify that this is indeed the case.

With these definitions, one immediately sees that the informal equation "$\delta_0(x) = \delta_0(-x)$" is simply a way of expressing the fact that

$$\text{if } F = \delta_0, \text{ then } \quad F = F^-.$$

5.7. • Distributions Defined on an Open Interval and Larger Classes of Test Functions

In this section we briefly discuss two natural extensions to the theory of distributions.

5.7.1. • Distributions Defined over a Domain. We often encounter problems in which the independent variables do not lie in all of $\mathbb{R}$ but rather some domain Ω in $\mathbb{R}$, i.e., an open interval $(-r, r)$ for some $r > 0$. To this end it is useful to consider distributions where the underlying space of the test functions is not $\mathbb{R}$ but, rather, some open interval $(-r, r)$. In these cases the test functions ϕ must have compact support in $(-r, r)$, and we denote this space as $C_c^\infty\big((-r, r)\big)$.

But what exactly does compact support in $(-r, r)$ mean? We mean a C^∞ function defined on $\mathbb{R}$ (yes, $\mathbb{R}$) for which there exists a (proper) subset $K \subset (-r, r)$ which is bounded and closed in $\mathbb{R}$ such that

$$\phi(x) = 0 \qquad \text{for all } x \notin K.$$

So certainly such a test function must vanish (i.e., be 0) outside of the interval $(-r, r)$ and at the two boundary points $x = \pm r$. However more is true; as one approaches any boundary point, ϕ must be zero before reaching the boundary, that is, at some nonzero distance from the boundary. So if ϕ_a is the blip function defined by (5.13), then $\phi_a \in C_c^\infty\big((-r, r)\big)$ if $a < r$ (but not if $a = r$). Loosely speaking, think of an end point as a point which we never "reach" or "touch" with our test functions.

We give one example without proof for $\Omega = (-\pi, \pi)$.

Example 5.7.1 (**An Infinite Series**). Recall that an infinite series of functions is precisely a limit of partial (finite) sums of the functions. Let $x \in \Omega = (-\pi, \pi)$. Then

$$\sum_{k \text{ is odd}} \frac{2}{\pi} \cos(kx) = \delta_0 \quad \text{in the sense of distributions on } \Omega. \tag{5.40}$$

This means that if we define the sequence of functions

$$f_n(x) = \sum_{k=0}^{n} \frac{2}{\pi} \cos\big((2k+1)x\big),$$

then as $n \to \infty$, we have

$$\int_{-\pi}^{\pi} f_n(x)\, \phi(x)\, dx \longrightarrow \phi(0) \qquad \text{for all } \phi \in C_c^\infty\big((-\pi, \pi)\big).$$

5.7.2. • Larger Classes of Test Functions. A statement in the sense of distributions involves a class of test functions which, so far, have been conveniently taken to be infinitely smooth localized functions. However, it is not always the case that we require the infinite smoothness or even the compactness of the support and, therefore, we can often **extend the statement** to a **wider (larger)** class of test functions. This means that the statement (i.e., the equality) holds for more general test functions than just C_c^∞. Note that the **larger** the class of test functions, the **stronger** (meaning the more general) the statement becomes. We give a few examples:

(i) We can define the delta function with test functions ϕ which are simply continuous functions, not necessarily smooth and not necessarily with compact support. That is, $\langle \delta_0, \phi \rangle = \phi(0)$ for all continuous functions ϕ.

(ii) The statement that $H'(x) = \delta_0$ in the sense of distributions (cf. Example 5.4.1) means that for every $\phi \in C_c^\infty(\mathbb{R})$,

$$-\int_{-\infty}^{\infty} H(x)\,\phi'(x)\,dx \;=\; \phi(0).$$

This actually holds true for all ϕ which are simply C^1 with compact support. In fact, it would also hold for any C^1 function which is integrable.

(iii) The statement that f_n, defined by (5.10), converges to δ_0 in the sense of distributions holds for all continuous functions ϕ; i.e., (5.28) holds for all continuous functions ϕ.

(iv) The sequence of functions $\sin nx$ converges to 0 in the sense of distributions. However, it turns out that the test functions can be taken to be just **integrable** (not necessarily continuous or possessing any derivatives). Precisely, we have

$$\int_{-\infty}^{\infty} \sin(nx)\,\phi(x)\,dx \;\longrightarrow\; 0,$$

for any function ϕ, such that $\int_{-\infty}^{\infty} |\phi(x)|\,dx \,<\, \infty$. This is an important result in Fourier analysis and is known as the **Riemann-Lebesgue Lemma**.

(v) In the next chapter we will see that, in order to extend the Fourier transform to distributions, we need a larger class of distributions stemming from a larger class of test functions. These test functions will still be C^∞ but we will relax the condition on compact support to the point where we only require the functions to decay *rapidly* to 0 as $|x| \to \infty$. This larger class of test functions is called the **Schwartz class** and the distributions, which can act not only on C_c^∞ test functions but also on Schwartz functions, are called **tempered distributions**.

5.8. Nonlocally Integrable Functions as Distributions: The Distribution PV $\frac{1}{x}$

This section is optional and should be skipped on first reading.

Throughout this text we will primarily be interested in distributions that are either generated by locally integrable functions or involve delta functions. Yet, there are important distributions which are neither. As we shall see, their Fourier transforms are very useful in solving many PDEs and one such distribution, called the principle value of $\frac{1}{x}$, is the building block of the **Hilbert transform**, which plays a central role in harmonic analysis. Consider the function

$$f(x) = \frac{1}{x}$$

which is **not** locally integrable around $x = 0$; that is, $\int_I \frac{1}{x}\, dx$ diverges for any interval I whose closure contains 0. We cannot directly interpret the function f as a distribution; however, there are several distributions which encapsulate the essence of this function. The key here is the idea of approximating the nonlocally integrable function $1/x$ by a sequence of functions in such a way as to exploit cancellation effects; we call this process **regularization**.

5.8.1. Three Distributions Associated with the Function $\frac{1}{x}$. We discuss three ways to regularize $1/x$, each giving rise to a slightly different distribution.

1. **The principle value (PV)**. Because the function $\frac{1}{x}$ is odd, one can **exploit the cancellation effects** and define the distribution PV $\frac{1}{x}$ as follows: For any $\phi \in C_c^\infty(\mathbb{R})$,

$$\boxed{\left\langle \mathrm{PV}\,\frac{1}{x}\,,\,\phi\right\rangle \;:=\; \lim_{\epsilon\to 0+}\int_{\{x\,|\,|x|>\epsilon\}} \frac{1}{x}\,\phi(x)\,dx.} \tag{5.41}$$

One can readily verify that this definition gives PV $\frac{1}{x}$ the status of a distribution. Note that we are defining this distribution as a limit of distributions stemming from the truncated functions:

$$\left\langle \mathrm{PV}\,\frac{1}{x}\,,\,\phi\right\rangle \;=\; \lim_{n\to\infty}\int_{-\infty}^{\infty} f_n(x)\,\phi(x)\,dx,$$

where for $\sigma_n = \frac{1}{n}$,

$$f_n \;=\; \begin{cases} \frac{1}{x} & \text{if } |x| > \sigma_n, \\ 0 & \text{if } |x| \le \sigma_n. \end{cases} \tag{5.42}$$

In other words, PV $\frac{1}{x}$ is the distribution limit of the functions f_n viewed as distributions.

2. The distribution

$$\frac{1}{x+i0}$$

is defined by

$$\boxed{\left\langle \frac{1}{x+i0}\,,\,\phi\right\rangle \;:=\; \lim_{\epsilon\to 0+}\int_{-\infty}^{\infty} \frac{1}{x+i\epsilon}\,\phi(x)\,dx.} \tag{5.43}$$

3. The distribution
$$\frac{1}{x - i0}$$
is defined by
$$\boxed{\left\langle \frac{1}{x - i0}, \phi \right\rangle := \lim_{\epsilon \to 0+} \int_{-\infty}^{\infty} \frac{1}{x - i\epsilon} \phi(x)\, dx.} \tag{5.44}$$

Note the appearance of the complex number $i = \sqrt{-1}$ in the above distributions 2 and 3. This imaginary number demonstrates its utility by translating the singularity away from the real axis of integration. These two distributions are hence complex-valued distributions and, as we will see, important in the Fourier transform of certain functions.

While all three distributions
$$\mathrm{PV}\,\frac{1}{x}, \qquad \frac{1}{x + i0}, \qquad \frac{1}{x - i0}$$
are related to the function $1/x$, they should all be understood only as distributions and **not** as functions in any pointwise sense. They represent different distributions; indeed, we will prove shortly that in the sense of distributions[14],
$$\boxed{\frac{1}{x + i0} = \mathrm{PV}\,\frac{1}{x} - i\pi\delta_0 \quad \text{and} \quad \frac{1}{x - i0} = \mathrm{PV}\,\frac{1}{x} + i\pi\delta_0,} \tag{5.45}$$
and, hence,
$$\mathrm{PV}\,\frac{1}{x} = \frac{1}{2}\left(\frac{1}{x + i0} + \frac{1}{x - i0}\right).$$
Surely you are curious about the appearance of the delta function in (5.45). It comes from the behavior of the (nonintegrable) function $1/x$ around $x = 0$. We will prove (5.45) shortly; it will be a simple consequence of another possible way (sequence) to regularize $1/x$ and define the same distribution $\mathrm{PV}\,\frac{1}{x}$.

5.8.2. A Second Way to Write $\mathrm{PV}\,\frac{1}{x}$.

Consider the function
$$g(x) = \begin{cases} \log|x|, & x \neq 0, \\ 0, & x = 0. \end{cases}$$
This function is locally integrable (indeed, it has a finite integral about 0) and, therefore, one can consider it in the sense of distributions (i.e., as the distribution F_g). What is its derivative in the sense of distributions? The answer cannot be the distribution generated by $1/x$; the function is not locally integrable and, hence, cannot be directly viewed as a distribution.

We shall now see that
$$(\log|x|)' = \mathrm{PV}\,\frac{1}{x} \quad \text{in the sense of distributions} \tag{5.46}$$

[14]These equations are related to what is known as the Sokhotski-Plemelj Theorem in complex analysis.

and, in doing so, we will provide an alternate but equivalent definition of PV $\frac{1}{x}$. Consider the sequence of functions, where as usual $\sigma_n = \frac{1}{n}$,

$$g_n(x) := \begin{cases} \log|x| & \text{if } |x| > \sigma_n, \\ 0 & \text{if } |x| \le \sigma_n. \end{cases}$$

Then by Proposition 5.5.2, we have

$$g = \lim_{n\to\infty} g_n \quad \text{in the sense of distributions.} \tag{5.47}$$

This means that for all $\phi \in C_c^\infty(\mathbb{R})$,

$$\int_{-\infty}^{\infty} g_n(x)\,\phi(x)\,dx \longrightarrow \int_{-\infty}^{\infty} g(x)\,\phi(x)\,dx \qquad \text{as } n \to \infty. \tag{5.48}$$

With (5.47) in hand, it follows from Proposition 5.5.1 that g_n' converges in the sense of distributions to g'. However, g_n' as a distribution is simply f_n, defined in (5.42), as a distribution. Thus, for any test function ϕ,

$$\left\langle (\log|x|)', \phi \right\rangle = \lim_{n\to\infty} \int_{-\infty}^{\infty} f_n(x)\,\phi(x)\,dx.$$

Hence, by definition of PV $\frac{1}{x}$, (5.46) holds true.

Now here is a trick: We can repeat the previous argument using a different approximating sequence:

$$g_n(x) = \log\sqrt{x^2 + \sigma_n^2}.$$

Again, by Proposition 5.5.2 , this sequence converges in the sense of distributions to $\log|x|$. Hence, its derivative

$$\frac{x}{x^2 + \sigma_n^2}$$

must also converge in the sense of distributions to $(\log|x|)'$. We have thus given a different way of writing PV $\frac{1}{x}$; namely it is the distributional limit

$$\text{PV}\,\frac{1}{x} = \lim_{n\to\infty} \frac{x}{x^2 + \sigma_n^2}.$$

This means that

$$\boxed{\left\langle \text{PV}\,\frac{1}{x}, \phi \right\rangle := \lim_{n\to\infty} \int_{-\infty}^{\infty} \frac{x}{x^2 + \sigma_n^2}\,\phi(x)\,dx.} \tag{5.49}$$

With (5.49) in hand, we can now prove (5.45).

Proof of (5.45)**.** First note that for any $x \in \mathbb{R}$ and $\sigma_n = \frac{1}{n}$,

$$\frac{1}{x + i\sigma_n} = \left(\frac{1}{x + i\sigma_n}\right)\left(\frac{x - i\sigma_n}{x - i\sigma_n}\right) = \frac{x}{x^2 + \sigma_n^2} - \frac{i\sigma_n}{x^2 + \sigma_n^2}.$$

Hence, by (5.43) and (5.49), for any test function ϕ we have that

$$\begin{aligned}\left\langle \frac{1}{x+i0}, \phi \right\rangle &:= \lim_{n\to\infty} \int_{-\infty}^{\infty} \frac{1}{x+i\sigma_n}\, \phi(x)\, dx \\ &= \lim_{n\to\infty} \int_{-\infty}^{\infty} \frac{x}{x^2+\sigma_n^2}\, \phi(x)\, dx \;-\; \lim_{n\to\infty} \int_{-\infty}^{\infty} \frac{i\sigma_n}{x^2+\sigma_n^2}\, \phi(x)\, dx \\ &= \left\langle \mathrm{PV}\,\frac{1}{x}, \phi \right\rangle \;-\; \lim_{n\to\infty} \int_{-\infty}^{\infty} \frac{i\sigma_n}{x^2+\sigma_n^2}\, \phi(x)\, dx.\end{aligned}$$

Now recall the sequence (5.24) from Section 5.5.3

$$\frac{1}{\pi}\, \frac{\sigma_n}{x^2+\sigma_n^2}$$

which converged in the sense of distributions to δ_0 as $n \to \infty$. Since

$$\frac{i\sigma_n}{x^2+\sigma_n^2} \;=\; i\pi \left(\frac{1}{\pi}\, \frac{\sigma_n}{x^2+\sigma_n^2} \right),$$

we have

$$\left\langle \frac{1}{x+i0}, \phi \right\rangle \;=\; \left\langle \mathrm{PV}\,\frac{1}{x}, \phi \right\rangle \;-\; i\pi\phi(0) \qquad \text{for all } \phi \in C_c^\infty.$$

This proves the first equation in (5.45). In a similar fashion, one proves the second equation.

5.8.3. A Third Way to Write $\mathrm{PV}\,\frac{1}{x}$**.** There is yet another way to write the distribution $\mathrm{PV}\,\frac{1}{x}$, which highlights the issue surrounding the loss of integrability at $x = 0$. First, we show that if the test function ϕ happened to satisfy $\phi(0) = 0$, then we would simply have

$$\left\langle \mathrm{PV}\,\frac{1}{x}, \phi \right\rangle \;=\; \int_{-\infty}^{\infty} \frac{1}{x}\, \phi(x)\, dx.$$

To see this, recall from calculus the Mean Value Theorem, which applied to ϕ says that

$$\phi(x) = \phi(x) - \phi(0) = \phi'(\eta)x, \qquad \text{for some } \eta \in (0, x).$$

Moreover, since $\phi \in C_c^\infty$, there exists an L such that ϕ and ϕ' vanish outside the interval $[-L, L]$, and ϕ' is bounded by some constant C on $[-L, L]$. Then the integrand in the definition

$$\left\langle \mathrm{PV}\,\frac{1}{x}, \phi \right\rangle \;:=\; \lim_{\epsilon\to 0+} \int_{\{x\,|\,|x|>\epsilon\}} \frac{1}{x}\, \phi(x)\, dx$$

is bounded; that is,

$$\frac{1}{x}\phi(x) \;=\; \frac{1}{x}\phi'(\eta)x \;=\; \phi'(\eta) \;\le\; C.$$

Thus, the integrand is integrable around 0 and there is no need to take the limit over the truncations on the set $\{x \,|\, |x| > \epsilon\}$.

The issue with turning $1/x$ into a distribution pertains to the behavior of the test function at $x = 0$. Hence, one might wonder what happens if we replace the test function ϕ with $\phi(x) - \phi(0)$. We now show that we indeed get the same distribution; that is,

$$\boxed{\left\langle \mathrm{PV}\,\frac{1}{x}, \phi \right\rangle = \int_{-\infty}^{\infty} \frac{1}{x}\,(\phi(x) - \phi(0))\,dx \quad \text{for all } \phi \in C_c^\infty.} \tag{5.50}$$

So we have yet another way to write the distribution PV $\frac{1}{x}$.

To prove (5.50), fix $\phi \in C_c^\infty$. For any $\epsilon > 0$ and any $L > \epsilon$, we have

$$\begin{aligned}\int_{\{x\,|\,|x|>\epsilon\}} \frac{1}{x}\,\phi(x)\,dx \quad &= \quad \int_{\{x\,|\,|x|>L\}} \frac{1}{x}\,\phi(x)\,dx + \int_{\{x\,|\,\epsilon<|x|<L\}} \frac{1}{x}\,(\phi(x) - \phi(0))\,dx \\ &\quad + \int_{\{x\,|\,\epsilon<|x|<L\}} \frac{1}{x}\,\phi(0)\,dx.\end{aligned}$$

For any ϵ and L, the third integral on the right is always 0 since $1/x$ is an odd function. The previous argument using the Mean Value Theorem implies that

$$\lim_{\epsilon\to 0^+} \int_{\{x\,|\,\epsilon<|x|<L\}} \frac{1}{x}\,(\phi(x) - \phi(0))\,dx = \int_{-L}^{L} \frac{1}{x}\,(\phi(x) - \phi(0))\,dx.$$

Thus we have for any $L > 0$,

$$\begin{aligned}\left\langle \mathrm{PV}\,\frac{1}{x}, \phi \right\rangle \quad &:= \quad \lim_{\epsilon\to 0+} \int_{\{x\,|\,|x|>\epsilon\}} \frac{1}{x}\,\phi(x)\,dx \\ &= \quad \int_{\{x\,|\,|x|>L\}} \frac{1}{x}\,\phi(x)\,dx + \int_{-L}^{L} \frac{1}{x}\,(\phi(x) - \phi(0))\,dx.\end{aligned}$$

If we choose L large enough so that $[-L, L]$ encompasses the support of ϕ, then the first integral above vanishes and we arrive at (5.50).

The presence of $\phi(0)$ in (5.50) is suggestive of a delta function. Informally this is the case, but with the caveat that the delta function comes with an infinite constant factor! Indeed, if we let

$$\text{“}C_\infty = \int_{-\infty}^{\infty} \frac{1}{x}\,dx\text{”}$$

denote the "infinite" constant, then informally (5.50) says

$$\text{“PV}\,\frac{1}{x} = \frac{1}{x} - C_\infty \delta_0\text{”}.$$

What we are witnessing here is a particular cancellation of the infinite behavior (surrounding $x = 0$) and this is made precise by the right-hand side of (5.50).

5.8.4. The Distributional Derivative of PV $\frac{1}{x}$.

We end this section by computing

$$\left(\mathrm{PV}\,\frac{1}{x}\right)' \qquad \text{in the sense of distributions.}$$

One would expect that the distributional derivative has something to do with the function $-\frac{1}{x^2}$. However, we should be careful. As with the function $\frac{1}{x}$, $\frac{1}{x^2}$ is not locally integrable and hence does not directly generate a distribution. Moreover, unlike with the odd function $\frac{1}{x}$, there is no cancellation effect for x negative and positive and hence, we cannot define a principal value in a similar way to the regularizations used to define PV $\frac{1}{x}$ in (5.41) and (5.49); they were all based on the fact that the function in question was odd. The key is to cancel off the infinite behavior around $x = 0$, as was seen in the third definition (5.50).

Consider the approximating sequence of truncated functions f_n defined in (5.42), which converge in the sense of distributions to PV $\frac{1}{x}$. By Proposition 5.5.1, we must have $f_n' \longrightarrow \left(\text{PV}\,\frac{1}{x}\right)'$ in the sense of distributions. In other words, for every $\phi \in C_c^\infty(\mathbb{R})$,

$$\left\langle \left(\text{PV}\,\frac{1}{x}\right)', \phi \right\rangle \;=\; \lim_{n\to\infty} \langle f_n', \phi\rangle .$$

We find via integration by parts that

$$\begin{aligned}
\langle f_n', \phi\rangle &= -\langle f_n, \phi'\rangle \\
&= -\int_{\{x\,|\,|x|>\sigma_n\}} \frac{1}{x}\,\phi'(x)\,dx \\
&\overset{\substack{\text{integration}\\ \text{by parts}}}{=} \int_{\{x\,|\,|x|>\sigma_n\}} -\frac{1}{x^2}\,\phi(x)\,dx \;+\; \frac{\phi(\sigma_n)+\phi(-\sigma_n)}{\sigma_n} \\
&= \int_{\{x\,|\,|x|>\sigma_n\}} -\frac{1}{x^2}\,\phi(x)\,dx \;+\; \frac{2\phi(0)}{\sigma_n} \\
&\qquad + \left(\frac{\phi(\sigma_n)+\phi(-\sigma_n)-2\phi(0)}{\sigma_n}\right).
\end{aligned}$$

Now, we take the limit as $n \to \infty$, to find

$$\begin{aligned}
\left\langle \left(\text{PV}\,\frac{1}{x}\right)', \phi \right\rangle &= \lim_{n\to\infty} \langle f_n', \phi\rangle \\
&= \lim_{n\to\infty} \left(\int_{\{x\,|\,|x|>\sigma_n\}} -\frac{1}{x^2}\,\phi(x)\,dx \;+\; \frac{2\phi(0)}{\sigma_n} \right) \\
&\qquad + \lim_{n\to\infty} \left(\frac{\phi(\sigma_n)+\phi(-\sigma_n)-2\phi(0)}{\sigma_n}\right).
\end{aligned}$$

But, by Taylor's Theorem applied to the C^∞ function ϕ, the second limit[15] is 0. For the first limit, note that

$$\begin{aligned}
\int_{\{x\,|\,|x|>\sigma_n\}} \frac{1}{x^2}\phi(0)\,dx &= \phi(0)\left(\int_{-\infty}^{-\sigma_n} \frac{1}{x^2}\,dx + \int_{\sigma_n}^{\infty} \frac{1}{x^2}\,dx \right) \\
&= \phi(0)\left(\frac{1}{\sigma_n} + \frac{1}{\sigma_n}\right) = \frac{2\phi(0)}{\sigma_n}.
\end{aligned}$$

Thus

$$\boxed{\left\langle \left(\text{PV}\,\frac{1}{x}\right)', \phi \right\rangle = \lim_{n\to\infty} \int_{\{x\,|\,|x|>\sigma_n\}} -\frac{1}{x^2}\big(\phi(x) - \phi(0)\big)\,dx.}$$

---

[15]By Taylor's Theorem, we have $\phi(\sigma_n) - \phi(0) = \sigma_n\phi'(0) + \frac{\sigma_n^2}{2}\phi''(0) + O(\sigma_n^3)$ and $\phi(-\sigma_n) - \phi(0) = -\sigma_n\phi'(0) + \frac{\sigma_n^2}{2}\phi''(0) + O(\sigma_n^3)$. Add these, divide by σ_n, and let $n \to \infty$.

This defines $\left(\mathrm{PV}\,\frac{1}{x}\right)'$ as a distribution. Informally, we can write this as

$$\text{“}\left(\mathrm{PV}\,\frac{1}{x}\right)' \;=\; -\frac{1}{x^2} \;+\; C_\infty \delta_0\text{”},$$

where C_∞ is the "infinite constant" "$C_\infty \;=\; \int_{-\infty}^{\infty} \frac{1}{x^2}\,dx$".

5.9. Chapter Summary

- Distributions are defined by **their effect (or action) on test functions**. While the choice of test functions can vary, we have focused on the convenient class of **infinitely smooth functions with compact support**. In this chapter we only considered test functions which depended on one variable and, as such, only addressed distributions in one space dimension.
- **The theory of distributions** allows us to both
  - interpret and analyze (i.e., do calculus with) classical functions, not in a pointwise input/output sense, but from the point of view of averaging and integration;
  - give precise meaning to, and find occurrences of, objects like δ_0, which cannot be captured by a pointwise-defined classical function.
- In general, there are many types of distributions with very complicated structures. For our purposes, we are concerned with three important classes:
  (1) Distributions **generated by locally integrable functions**: If $f(x)$ is a locally integrable function on $\mathbb{R}$, then f can be thought of as the distribution F_f where

  $$\langle F_f, \phi\rangle \;:=\; \int_{-\infty}^{\infty} f(x)\,\phi(x)\,dx \quad \text{for any} \quad \phi \in C_c^\infty(\mathbb{R}).$$

  When we speak of the function f in the sense of distributions we mean F_f.
  (2) Distributions which **concentrate at points**; for example the **delta function** δ_0, which is defined by $\langle \delta_0, \phi\rangle = \phi(0)$ for any test function ϕ.
  (3) Distributions related to **nonlocally integrable functions**; for example PV $\frac{1}{x}$ which was the focus of Section 5.8.

  The first two classes of distributions will prevail throughout the sequel of this text. The PV $\frac{1}{x}$ will be important in using the Fourier transform.
- **Convergence in the sense of distributions** allows us to capture and calculate the limiting behavior of functions which do the following:
  - **Concentrate** their "mass" at a point; for example, f_n defined by (5.10), which concentrates at $x = 0$ and converges in the sense of distributions to δ_0.
  - **Oscillate** more and more about a fixed number; for example, $f_n(x) = \sin nx$, whose values oscillate more and more about 0.

  As Dirac demonstrated, it is useful to *visualize* the delta function δ_0 with one of these sequences which concentrate at $x = 0$; one can think of δ_0 as f_n, for some very large value of n. With $\sigma_n = 1/n$, we discussed six different examples of such sequences of functions. They are, roughly ordered in terms of increasing complexity,

$$\text{(i)}\ \ f_n(x) = \begin{cases} \frac{n}{2}, & |x| \le \sigma_n, \\ 0, & \text{otherwise,} \end{cases} \qquad \text{(ii)}\ \ f_n(x) = \begin{cases} n - n^2 x & \text{if} \quad 0 < x < \sigma_n, \\ n - n^2 x & \text{if} \ -\sigma_n < x < 0, \\ 0 & \text{if} \quad |x| \ge \sigma_n, \end{cases}$$

$$\text{(iii)}\ \ f_n(x) = \frac{1}{\pi}\,\frac{\sigma_n}{x^2 + \sigma_n^2}, \qquad \text{(iv)}\ \ f_n(x) = \frac{1}{\sqrt{4\pi\,\sigma_n}}\, e^{-\frac{x^2}{4\sigma_n}},$$

$$\text{(v)}\ \ f_n(x) = \frac{\sin nx}{\pi x}, \qquad \text{(vi)}\ \ f_n(x) = \frac{\sin\left(n + \frac{1}{2}\right)x}{\sin\left(\frac{x}{2}\right)}.$$

The first two denote hats and spikes, respectively, and are best viewed in terms of their simple graphs (Figure 5.5). As we shall see, the sequence of Gaussians (iv) are key to solving the diffusion equation, and the sequences (v) and (vi) play a fundamental role in the Fourier transform and Fourier series, respectively. Sequence (vi) consists of periodic functions with period 2π, and hence their distributional convergence to δ_0 is restricted to test functions with compact support in $(-\pi, \pi)$.

- We can take the **derivative** of any distribution by considering the action of the respective derivative on the test function. Such a derivative is also a distribution over the same class of test functions. For our purposes, we are concerned with derivatives of **distributions generated by functions**, that is, derivatives of functions in the sense of distributions. Moreover, we are primarily concerned with differentiating functions which are smooth except for certain **singularities**. We focused on the differentiation of a piecewise smooth function of one variable with a jump discontinuity. In doing so, we saw **two possible components** for the distributional derivatives of these functions: functions and delta functions. Important examples where distributional derivatives of functions (derivatives in the sense of distributions) yield the delta function are (in 1D)

$$\frac{d}{dx}H(x) = \delta_0 \quad \text{and} \quad \frac{d^2}{dx^2}\left(\frac{|x|}{2}\right) = \delta_0, \quad \text{in the sense of distributions.}$$

  In these examples there is no nonzero functional component to the distributional derivatives but, rather, only the delta function.

- In order to **find** derivatives in the sense of distributions of these functions with singularities, we do the following:
  (i) Fix a test function and write down the integral of the function multiplied by the derivative of the test function.
  (ii) **Isolate** the singularity (singularities) by breaking up the domain in $\mathbb{R}$ into disjoint intervals.
  (iii) On regions of integration where the original function is smooth, we perform **integration by parts** to place the derivatives back on the original function, at the expense of additional boundary terms.

Exercises

5.1 (a) Show that the test function (5.13) is in $C_c^\infty(\mathbb{R})$.
(b) Give an explicit function in $C_c^\infty(\mathbb{R})$ whose support is equal to a generic bounded interval $[a, b]$.

5.2 Let $f(x) = x$ if $x \geq 0$ and $f(x) = 0$ if $x < 0$. Find f' and f'' in the sense of distributions.

5.3 Let $f(x) = |x|$ for $-1 < x \leq 1$. Extend f to all of $\mathbb{R}$ by periodicity. In other words, repeat the function from 1 to 3, from 3 to 5, from -3 to -1, and so forth. Note that the extension f will satisfy $f(x+2) = f(x)$ for all $x \in \mathbb{R}$. Find f' and f'' in the sense of distributions.

5.4 Give an example of a function $f(x)$ whose first and second derivatives in the sense of distributions are both distributions generated by functions, but f''' in the sense of distributions is δ_0.

5.5 Prove Proposition 5.5.1.

5.6 Prove that the sequence of hat functions defined by (5.22) converges *in the sense of distributions* to δ_0.

5.7 Consider the following function:

$$f(x) = \begin{cases} 0 & \text{if } x \leq 0, \\ e^x & \text{if } 0 < x \leq 1, \\ x & \text{if } 1 < x \leq 2, \\ 0 & \text{if } x > 2. \end{cases}$$

What is its derivative in the sense of distributions? What is its second derivative in the sense of distributions?

5.8 Let $f(x) = e^{-x}$ if $x > 0$ and $f(x) = -e^x$ if $x \leq 0$. Find f' in the sense of distributions and show that $f'' = 2\delta_0' + f$ in the sense of distributions.

5.9 Find a function $f(x)$ whose derivative in the sense of distributions is the distribution $x^2 + 4x + \delta_2$. Note that this means a function f such that $(F_f)' = (F_g) + \delta_2$, where $g(x) = x^2 + 4x$ and δ_2 is the delta function with concentration at $x = 2$.

5.10 (**Summarizing the Distributional Derivative of a Piecewise Smooth Function**) This exercise generalizes (and summarizes) many of the previous exercises and examples. Suppose f is a C^1 (continuously differentiable) function except at a point x_0. Define the function

$$g(x) = \begin{cases} f'(x) & \text{if } x \neq x_0, \\ 0 & \text{if } x = x_0. \end{cases}$$

(a) If f is continuous at $x = x_0$, prove (show) that $f' = g$ in the sense of distributions.
(b) Suppose f has a jump discontinuity at x_0 with $a = \lim_{x \to x_0^-} f(x)$ and $b = \lim_{x \to x_0^+} f(x)$. Show that $f' = g + (b - a)\delta_{x_0}$ in the sense of distributions.

5.11 Is $\sum_{n=1}^{\infty} \delta_n$ a well-defined distribution? Note that to be a well-defined distribution, its action on any test function must be a finite number. Provide an example of a function $f(x)$ whose derivative in the sense of distributions is $\sum_{n=1}^{\infty} \delta_n$.

5.12 (a) Consider the function $f(x)$ defined on $[0, 1)$ to be -1 if $0 \le x \le \frac{1}{2}$ and 1 if $\frac{1}{2} < x < 1$. Extend f to the entire real line periodically (with period 1). For each $n = 1, 2, \ldots$, define $f_n(x) := f(nx)$. Now look at f_n as a sequence of distributions. Does $f_n(x)$ converge as $n \to \infty$ **in the sense of distributions**? If so, to what? If not, why? You do not need to prove your answer but, instead, provide some explanation. You should start by sketching a few of the f_n: say, $f_1(x)$, $f_2(x)$, $f_4(x)$.
(b) Repeat part (a) with $f(x)$ now defined on $[0, 1)$ to be -1 if $0 \le x \le \frac{1}{2}$ and 5 if $\frac{1}{2} < x < 1$.

The next five exercises are geared towards students with some exposure to mathematical analysis.

5.13 While we discussed sequential limits of functions, one can also consider continuum limits of functions, for example, the distributional limit of f_ϵ as $\epsilon \to 0^+$. Let f be any integrable function such that $\int_{\mathbb{R}} f(x)\,dx = 1$. For any fixed $x_0 \in \mathbb{R}$, prove that

$$\frac{1}{\epsilon} f\left(\frac{x - x_0}{\epsilon}\right) \xrightarrow{\epsilon \to 0^+} \delta_{x_0} \qquad \text{in the sense of distributions.}$$

Hint: Use the continuum analogue of Theorem A.8 in the Appendix.

5.14 Prove that the sequence $f_n(x)$ of Gaussians (5.25) converges to δ_0 in the sense of distributions.

5.15 From the basic structure of the proofs of Theorems 5.1 and 5.2 and Exercise **5.14**, it is natural to ask for one *mother* proof which shows that the three properties of Properties 5.5.1 are sufficient for distributional convergence to the delta function. The issue here is that the third property (concentration at 0) needs to be made precise. Use the following two conditions for property (3): (i) For any $a > 0$, $f_n(x)$ converges uniformly to 0 on the interval $|x| \ge a$; (ii) for any $a > 0$,

$$\lim_{n \to \infty} \int_{|x| \ge a} f_n(x)\,dx \;=\; 0,$$

to prove that the sequence of functions $f_n(x)$ converges to the delta function in the sense of distributions. Prove that the sequences (5.23), (5.24), and (5.25) all satisfy this precise concentration property.

5.16 Prove Theorem 5.3. Extra: Is the hypothesis on non-negativity necessary?

5.17 Let $f(x)$ be any integrable, periodic function with period l such that its integral over any interval of length l (i.e., over a period) is A for some $A \in \mathbb{R}$. Define $f_n(x) := f(nx)$, and prove that $f_n(x)$ converges to $f(x) \equiv \frac{A}{l}$ (the average) in the sense of distributions.

5.18 Find the distributional limit of the sequence of distributions $F_n = n\delta_{-\frac{1}{n}} - n\delta_{\frac{1}{n}}$. Hint: F_n is the distributional derivative of some function.

5.19 Consider the distribution δ_0', that is, the derivative of the 1D delta function in the sense of distributions. Find a sequence of functions $f_n(x)$ such that f_n converges to δ_0' in the sense of distributions. Sketch $f_n(x)$ for $n = 1, 5, 10, 20$.

5.20 Consider the equation (5.40). (a) By looking up references online, give either a partial or full (i.e., a proof) justification for this equation. (b) Find a situation in mathematics or physics where this sum turns up. (c) Suppose we considered the sum not on $\Omega = (-\pi, \pi)$, but on the whole real line $\mathbb{R}$. In that case, what would it converge to in the sense of distributions?

5.21 Let $y \in \mathbb{R}$ and consider the distribution δ_y (the delta function with concentration at $x = y$). As usual, x will denote the underlying independent variable of this distribution (i.e., the variable of the test functions). Suppose we wish to differentiate δ_y with respect to the source point y (not as we have previously done with respect to the underlying independent variable x). Rather than trying to make a precise distributional definition, simply justify the informal equation:

$$\text{“}\frac{\partial}{\partial y}\delta_y = -\delta_y' \left(= -\frac{\partial}{\partial x}\delta_y\right)\text{”.}$$

5.22 (a) Justify (5.38).
(b) Suppose $P(x)$ is a cubic polynomial with distinct real roots a, b, c such that the derivative $P'(x)$ is not zero at any of these roots. Justify the informal equation

$$\text{“}\delta_0[P(x)] = \frac{1}{|P'(a)|}\delta_0(x-a) + \frac{1}{|P'(b)|}\delta_0(x-b) + \frac{1}{|P'(c)|}\delta_0(x-c)\text{”.}$$

(c) (**Change of Variables in the Delta Function**) Let $y(x)$ be a smooth function. Justify the following informal change of variables statement: For all $a \in \mathbb{R}$, “$\delta_0(x-a) = y'(a)\,\delta_0(y - y(a))$”.
(d) Justify the informal statement “$\delta_0(\sin \pi x) = \sum_{n=-\infty}^{\infty} \delta_0(x-n)$”. The right-hand side is actually the distribution $\sum_{n=-\infty}^{\infty} \delta_n$ which is known as the **Dirac comb**. See Exercise **6.29** for more on the Dirac comb.

5.23 Justify the informal equations

$$\text{“}\int_{-\infty}^{a} \delta_0(x)\,dx = \begin{cases} 1 & \text{if } a > 0, \\ 0 & \text{if } a < 0 \end{cases}\text{”} \quad \text{and} \quad \text{“}\int_{a}^{\infty} \delta_0(x)\,dx = \begin{cases} 0 & \text{if } a > 0, \\ 1 & \text{if } a < 0. \end{cases}\text{”}$$

5.24 Let $g \in C^\infty(\mathbb{R})$. If F is any distribution, we can define the product of g with F as a new distribution defined by

$$\langle gF, \phi \rangle = \langle F, g\phi \rangle \qquad \text{for any } \phi \in \mathcal{D}.$$

(a) Why is this a reasonable definition?
(b) Prove (show) that in the sense of distributions, $x\,\delta_0 = 0$ and $x\,(\delta_0)' = -\delta_0$.
(c) If for any $n = 1, 2, \ldots$, $\delta_0^{(n)}$ denotes the n-th distributional derivative of δ_0, prove (show) that

$$x^n\,\delta_0^{(n)} = (-1)^n\, n!\, \delta_0.$$

(d) In the context of this question, show that the informal equation (written by Dirac on page 60 of *The Principles of Quantum Mechanics*)

$$\text{“}f(x)\delta(x-a) = f(a)\delta(x-a)\text{”}$$

has full justification in the sense of distributions.

5.25 Justify the informal equation written by Dirac on page 60 of *The Principles of Quantum Mechanics*:

$$\text{“}\left(\int \delta_0(x-a)\right)\delta_0(x-b) = \delta_0(a-b)\text{”}.$$

5.26 Consider the function $f(x) = x^{1/3}$. Show that $f(x)$ is locally integrable on $\mathbb{R}$ and, hence, can be considered as a distribution. Find f' in the sense of distributions. Note that the pointwise derivative of f is not locally integrable.

5.27 Consider the informal equation written by Dirac on page 61 of *The Principles of Quantum Mechanics*:

$$\text{“}\frac{d}{dx}\log x = \frac{1}{x} - i\pi\delta(x)\text{”}. \tag{5.51}$$

By $\frac{1}{x}$ above, Dirac is talking about PV $\frac{1}{x}$. However, recall from (5.46) that the distributional derivative of $\log|x|$ was simply PV $\frac{1}{x}$. In (5.51), Dirac considers the logarithm for all $x \in \mathbb{R}$, interpreting the log via its extension to the complex plane. With this interpretation of log, show (5.51) as a precise statement in the sense of distributions.