The Theory of Distributions
Robert G. Muncaster
Summer 1997

Preliminary Ideas

When you are presented with the task of solving a linear partial differential equation (PDE) written in the form

\[ Lu = f, \quad L = \sum_{|\alpha| \leq m} a_\alpha(x) D^\alpha, \tag{1} \]

you must not only worry about finding a candidate for a solution, but in addition you need to worry about whether the function \( u \) involved has enough derivatives for \( Lu \) to make sense. Mathematicians invented the theory of distributions so that these two worries, i.e. finding a solution and proving that it is smooth, could be pursued separately from each other. The idea was to introduce a space of objects called distributions, also referred to as generalized functions, that had all the smoothness properties that we might need, and to search first for a solution of the equations in this space. Thereafter one would ask whether this generalized function was a true function, and, if so, whether it was smooth and hence a classical solution.

If \( f \) is a locally integrable function on a domain \( \Omega \subset \mathbb{R}^n \) (i.e. the integral of \( f \) over any bounded subset of \( \Omega \) exists and is finite), then we write \( L^1_{\text{loc}}(\Omega) \). With such a function we associate a linear mapping \( F_f \) that operates on test functions \( \phi \in C_0^\infty(\Omega) \). The value of \( F_f \) at \( \phi \) is written in the form \( \langle F_f, \phi \rangle \) and has the explicit value

\[ \langle F_f, \phi \rangle = \int_{\Omega} f(x)\phi(x)dx. \tag{2} \]

We can think of \( F_f \) as giving us the averages of \( f \) over any conceivable subset of \( \Omega \), however small, simply by choosing a \( \phi \) that is 1 on this subset and drops quickly to zero outside the subset in question. So the essential question is this: what is the difference between knowing a function \( f \) point by point vs knowing its average value over each and every set? As we assess this question, it is useful to consider the way in which we measure quantities associated with physical systems. For example, if we want to determine the stress that is present at certain points of a bar or rod or other body, generally we attach to that body strain gauges at the points in question. However, those gauges provide at best an average of the stress over the region of attachment. In essence we are inferring the stress pointwise by working with its local averages. Most other measurements in physics are of the same kind — averages rather than point wise calculations — and so we are generally content in physics to define functions in terms of their local averages. For this reason we are already familiar with the conception of a distribution, even though it has not been pinned down so far.

The correspondence between a function \( f \) and the distribution \( F_f \) defining its averages is extremely important for several reasons. First, we will see shortly that the concept of a distribution is sufficiently general that there are distributions that cannot be represented as a function (the dirac distribution introduced later is an example, though less formal approaches to distributions try very hard to view this distribution as a function). Nevertheless, when solving differential equations using distributions, we ultimately come back to the important question of whether the distribution that represents our solution is an actual function. The only way to answer this is to verify (or disprove that) the value of the solution-distribution at a test function \( \phi \) has the form of the right-hand side of (2)
for some function \( f \). (Hint: to do this, the value of \( F_f \) at \( \varphi \) must at the very least involve an integration operation.)

The second way in which the function to distribution correspondence is important is that it is used to motivate definitions for more general distributions. That is, if we can write down a property of integrals (of the type (2)) in the notation of distributions, then those properties should hold for distributions in general. Let us consider as illustration some examples. First note to following formula that represents integration by parts:

\[
\int_{-\infty}^{\infty} f'(x) \varphi(x) dx = -\int_{-\infty}^{\infty} f(x) \varphi'(x) dx
\]

We have used here the fact that \( \varphi \) vanishes for \(|x| \) large. Written in distribution notation this becomes:

\[
\langle F_f', \varphi \rangle = -\langle F_f, \varphi' \rangle
\]

(3)

Given that we would like the distribution corresponding to \( f \) to be differentiable if \( f \) is differentiable, and in this case to have the derivative of \( F_f \) agree with the distribution corresponding to \( f' \), i.e. \( (F_f)' = F_{f'} \), we would need:

\[
\langle (F_f)', \varphi \rangle = -\langle F_f, \varphi' \rangle
\]

(4)

More generally, for higher derivatives and functions defined on \( \mathbb{R}^n \), we have the corresponding formula:

\[
\langle D^\alpha F_f, \varphi \rangle = (-1)^\alpha \langle F_f, D^\alpha \varphi \rangle
\]

(5)

This will form the basis later for how we define the derivatives of a distribution. The bottom line is this: for a definition in terms of distributions to be meaningful, that definition should reduce to natural properties of integrals when specialized to a distribution that is an actual function.

Consider another example. Here is a trivial property of integrals, as well as its statement in distribution form:

\[
\langle F_{vf}, \varphi \rangle = \int_{\Omega} (v(x)f(x))\varphi(x) dx = \int_{\Omega} f(x)(v(x)\varphi(x)) dx = \langle F_f, v\varphi \rangle
\]

(6)

If we would like to define the product of a distribution such as \( F_f \) by a function \( v \), then we would hope that any such definition would give, naturally, the distribution corresponding to \( vf \), i.e. \( vF_f = F_{vf} \). This would mean that (7) would become

\[
\langle vF_f, \varphi \rangle = \langle F_f, v\varphi \rangle
\]

(7)

We will use this later to define what we will mean by the product of a distribution and a function. It will ensure that for functions the distribution corresponding to \( vf \) is \( v \) times the distribution corresponding to \( f \).

In later sections of these notes we will see other examples of the use of this correspondence (see, in particular, the definition of the Fourier transform of a distribution), but let us now proceed to the theory.

**Definition of a Distribution**

For functions of \( n \) variables we generally work on a domain \( \Omega \subset \mathbb{R}^n \) (though \( \Omega \) might be all of \( \mathbb{R}^n \)). The space of test functions is

\[
D(\Omega) = C^\infty_0(\Omega) = \text{space of infinitely differentiable functions whose support is compact and contained in } \Omega
\]

The space of distributions on \( \Omega \) is denoted \( D'(\Omega) \) (i.e. in the notation of advanced analysis this is the dual space of \( D(\Omega) \)). It is defined as follows:
DEFINITION. A distribution is a real-valued mapping on $D(\Omega)$ that is linear and continuous.

[A section is needed here on what topology is used in $D(\Omega)$ that allows us to justify the use of the word continuous and verify continuity in specific examples.]

Since an integral with weight function $f$, such as (2), is linear, it certainly qualifies as a distribution. The matter of it being continuous is a technical requirement that we will avoid at this stage. All the examples we introduce here satisfy the continuity requirement (though the proof is not always trivial). Here are some more explicit examples:

EXAMPLE. (The Heaviside Distribution) Let $H(x)$ be the Heaviside function

$$H(x) = \begin{cases} 1 & \text{for } x \geq 0 \\ 0 & \text{for } x < 0 \end{cases}$$  \hfill (9)

Since this is a locally integrable function, its corresponding distribution $F_H$ is defined by (2). Writing this out explicitly we get:

$$\langle F_H, \varphi \rangle = \int_0^\infty \varphi(x)dx$$  \hfill (10)

Hence the Heaviside distribution gives the integral of the test function over the positive $x$ axis.

EXAMPLE. When looking at weak solutions of the wave equation $u_{xx} - u_{yy} = 0$ we considered the function

$$u(x,y) = \begin{cases} (x+y)^2 & \text{for } x \leq -y \\ 0 & \text{for } x > -y \end{cases}$$  \hfill (11)

Since this is a locally integrable function, its corresponding distribution is given by (2) again. Using the explicit formula for $u$, this distribution becomes:

$$\langle F_u, \varphi \rangle = \int_{x \leq -y} (x+y)^2 \varphi(x,y)dxdy$$  \hfill (12)

Later we will show that $F_u$ satisfies the wave equation in the sense of distributions.

EXAMPLE. (The Dirac Distribution) We define the Dirac distribution $\delta$ by

$$\langle \delta, \varphi \rangle = \varphi(0)$$  \hfill (13)

You should easily convince yourself that this function is linear in $\varphi$. Note that this definition makes sense if we are working in $\mathbb{R}^n$, not just $\mathbb{R}$. The distribution $\delta$ is sometimes referred to as the Dirac distribution centered at 0. In a similar way we can define the Dirac distribution $\delta_y$ centered at $y \in \mathbb{R}^n$ by:

$$\langle \delta_y, \varphi \rangle = \varphi(y)$$  \hfill (14)

The Dirac distributions, by their definitions, do not involve an integral operation such as in (2). For this reason they represent examples of distributions that cannot be represented as a true function. All that is important about the Dirac distribution is contained in (13), i.e. its value on a test function $\varphi$.

EXAMPLE. (Convolutions) If $f$ is locally integrable, then a different type of distribution associated with it is $F_y^f$, $y \in \mathbb{R}$, given by:
\[
\langle F^y, \phi \rangle = \int_{-\infty}^{\infty} f(y-x)\phi(x)dx
\] 
You should recognize this to be the standard one-dimensional convolution of \( f \) and \( \phi \).

EXAMPLE. Consider the distribution
\[
\langle E, \phi \rangle = \frac{1}{2} \int_{-\infty}^{\infty} f(x)(\phi(x) + \phi(-x))dx
\] 
Might this be regarded as the even part of \( f \)? Or the odd part of \( f \)? Can you use this idea to define what we might mean by the even or odd part of a general distribution? Is the Dirac distribution centered at 0 both even and odd?

EXAMPLE. In general, for any smooth function \( \phi \) defined on \( \mathbb{R} \) the new function \( \phi(x)/x \) has a singularity at \( x = 0 \) sufficiently strong that the integral of \( \phi(x)/x \) over any interval containing 0 in its interior necessarily diverges. However, the following formula defines a perfectly reasonable distribution (it is the Cauchy-principal value of the divergent integral — see any calculus text dealing with improper integrals):
\[
\langle C, \phi \rangle = \lim_{h \to 0} \left( \int_{-\infty}^{-h} \frac{\phi(x)}{x}dx + \int_{h}^{\infty} \frac{\phi(x)}{x}dx \right)
\] 
Clearly this expression is a linear function of \( \phi \).

**Distributions With Variables Separable**

One way to build up a set of examples of distributions is to use simple instances to construct more complex ones. For this the concept of **variables separable** is useful. This same notion will also be useful to us later as a mechanism for turning differential equations (for distributions) in multiple space dimensions into simpler one-dimensional problems (as in the method of separation of variables for solving boundary value problems).

We want a simple notation for denoting a function of the form \( \phi(x)\psi(y) \) in which the variables are separated. We will denote this product by \( \phi \otimes \psi \). The precise form of its definition is heavy but the idea is still rather simple:

**DEFINITION.** Given sets \( \Omega_1 \subset \mathbb{R}^{n_1} \) and \( \Omega_2 \subset \mathbb{R}^{n_2} \), and functions \( \phi \) and \( \psi \) defined on \( \Omega_1 \) and \( \Omega_2 \), respectively, the tensor product \( \phi \otimes \psi \) is the function defined on \( \Omega = \Omega_1 \times \Omega_2 \subset \mathbb{R}^n \), \( n = n_1 + n_2 \), whose values are given by:
\[
\phi \otimes \psi(x,y) = \phi(x)\psi(y)
\] 
You should be able to convince yourself that if \( \phi \in \mathcal{C}_0^\infty(\Omega_1) \) and \( \psi \in \mathcal{C}_0^\infty(\Omega_2) \), then \( \phi \otimes \psi \in \mathcal{C}_0^\infty(\Omega) \). Moreover, given any multi-index \( \alpha \), we can write \( \alpha = \alpha_1 + \alpha_2 \) where the multi-index \( \alpha_1 \) covers all indices associated with the \( x \) variable and \( \alpha_2 \) covers all those associated with \( y \). Then we have the simple relation
\[
D^{\alpha}(\phi \otimes \psi) = (D^{\alpha_1} \phi) \otimes (D^{\alpha_2} \psi).
\] 
This is the general version of a simple statement like \( D_{xy}(\phi(x)\psi(y)) = \frac{\partial}{\partial x} \phi(x)\psi(y) \).

Now let’s mimic these ideas for distributions. The driving idea is that the distribution corresponding to the tensor product of two functions should be the tensor product of the distributions corresponding to the separate functions. That is, our
DEFINITION. If $F$ and $G$ are distributions on $\Omega$, their tensor product $F \otimes G$ is the distribution on $\Omega$ whose values on tensor products of test functions satisfy
\[
\langle F \otimes G, \phi \otimes \psi \rangle = \langle F, \phi \rangle \langle G, \psi \rangle
\]
(20)

(Strictly speaking, when we define a distribution we need to give its values for all test functions. Here we have specified the values of $F \otimes G$ only for test functions that are tensor products. The extra piece of information that helps us is the technical fact that we can approximate any test function depending on both $x$ and $y$ arbitrarily close by a linear combination of tensor product test functions. This means that the definition (20) is sufficient to fully define $F \otimes G$.)

At the level of distributions this is a good definition because the right-hand side, which must be a real number, is simply the product of two real numbers (i.e. the values of the separate distributions at the separate test functions). Now let's make sure that our preliminary identity is true. Here is a proof:
\[
\langle F_f \otimes F_g, \phi \otimes \psi \rangle = \langle F_f, \phi \rangle \langle F_g, \psi \rangle = \left( \int_{\Omega_1} f(x)\phi(x)dx \right) \left( \int_{\Omega_2} g(y)\psi(y)dy \right) = \int_{\Omega_1 \times \Omega_2} f(x)\phi(x)g(y)\psi(y)dxdy
\]
\[
= \int_\Omega f \otimes g(x,y)\phi \otimes \psi(x,y)dxdy = \langle F_f \otimes g, \phi \otimes \psi \rangle
\]
(21)

Since this is valid for all tensor products $\phi \otimes \psi$ (and these products are dense in the set of test functions in the variables $(x,y)$), we conclude that $F_f \otimes F_g = F_f \otimes g$. Now to some examples.

EXAMPLE. Let us denote the Dirac distribution on $\mathbb{R}^n$ by $\delta^n$ so as to emphasize the dimension of the space. Then
\[
\delta^n = \delta^1 \otimes \cdots \otimes \delta^1 \quad (n \text{ copies}).
\]
(22)
This will be left as an exercise. Later we will be interested in finding distributions that satisfy the differential equation $Lu = \delta^n$ and (22) will help us to attack this problem using separation of variables.

EXAMPLE. The two dimensional Heaviside function is $H \otimes H$, i.e.
\[
H \otimes H(x,y) = \begin{cases} 1 & \text{for } x \ge 0, y \ge 0 \\ 0 & \text{elsewhere} \end{cases}
\]
(23)
or the function that is 1 on the positive $xy$ quadrant and zero elsewhere. Using either (2) or the definition (20), the associated two dimensional Heaviside distribution has the following equivalent forms:
\[
\langle F_{H \otimes H}, \varphi \rangle = \langle F_H \otimes F_H, \varphi \rangle = \int \int \varphi(x,y)dxdy
\]
(24)

Properties of Distributions
There are four main definitions that give additional properties to distributions, namely, the definitions of a) addition, b) multiplication by a function, c) differentiation, and d) convolution. (Later we will add to these the Fourier transform.) In this section we
consider the first three. They are modeled on relations (6) and (8) that hold for true functions and integration.

**DEFINITION.** If $F$ and $G$ are distributions on $\Omega$, then $F + G$ is the distribution whose values are given by

$$
\langle F + G, \varphi \rangle = \langle F, \varphi \rangle + \langle G, \varphi \rangle
$$

This generalizes the usual sum of functions since it implies that $F_{f+g} = F_f + F_g$ for distributions that come from functions.

**DEFINITION.** If $F$ is a distribution on $\Omega$ and $v$ is a $C^\infty$ function, then $vF$ is the distribution whose values are given by

$$
\langle vF, \varphi \rangle = \langle F, v\varphi \rangle
$$

The important technical issue here involves noting that if $\varphi \in C^\infty_0(\Omega)$ and $v \in C^\infty(\Omega)$, then $v\varphi \in C^\infty_0(\Omega)$ and hence the right-hand side of (26) makes sense.

**DEFINITION.** If $F$ is a distribution on $\Omega$, then $D^\alpha F$ is the distribution with values

$$
\langle D^\alpha F, \varphi \rangle = (-1)^\alpha \langle F, D^\alpha \varphi \rangle
$$

Recall that this simply reflects the integration by parts formula for integrals.

Now we look at a number of special cases and examples and other properties that follow from these definitions.

**EXAMPLE.**

$$
\frac{d}{dx} F_H = \delta \text{ on } \mathbb{R}
$$

Proof. For all test functions $\varphi$

$$
\langle \frac{d}{dx} F_H, \varphi \rangle = -\langle F_H, \frac{d\varphi}{dx} \rangle = -\int_0^\infty \frac{d\varphi}{dx} \, dx = -\varphi|_0^\infty = \varphi(0) = \langle \delta, \varphi \rangle
$$

since $\varphi$ vanishes at $\infty$. Hence (28) is true.

**EXAMPLE.** $D_{x_i} \delta$ is the distribution with values

$$
\langle D_{x_i} \delta, \varphi \rangle = -\varphi_{x_i}(x)
$$

**LEMMA.** If $v$ is a $C^\infty$ function such that $v(0) = 0$, then $v\delta = 0$, i.e. $v\delta$ is the zero distribution. (The proof is left to you. You simply need to show that $\langle v\delta, \varphi \rangle = 0$ for all test functions $\varphi$.)

**LEMMA.** (Product Rule) If $F$ is a distribution and $v$ is a $C^\infty$ function, then

$$
D(vF) = (Dv)F + vDf
$$

(We leave to you the proof of this also. It is useful to recognize that each of the terms on the right-hand side of (30) is a $C^\infty$ function times a distribution, and so the sum is again a distribution.)

**LEMMA.** (Derivatives of Separable Distributions) Given any multi-index $\alpha$, let us write $\alpha = \alpha_1 + \alpha_2$ where the multi-index $\alpha_1$ covers all indices associated with the $x$ variable
and \(\alpha_2\) covers all those associated with \(y\). Then the derivative \(D^\alpha\) of a tensor product \(F \otimes G\) of distributions satisfies the natural analogue of (19), namely,

\[
D^\alpha(F \otimes G) = (D^\alpha_1 F) \otimes (D^\alpha_2 G). \tag{31}
\]

By combining sums with multiplication by \(C^\infty\) functions with differentiation, we can define differential operators for distributions. In particular, the operator \(L\) defined in (1) can be applied to any distribution. The result is the following simple result that exhibits a new appearance of the adjoint operator:

**LEMMA.** \(\langle LF, \varphi \rangle = \langle F, L' \varphi \rangle\) \(\tag{32}\)

**Proof.**

\[
\langle LF, \varphi \rangle = \sum_{|\alpha| \leq m} \langle a_\alpha(x) D^\alpha F, \varphi \rangle = \sum_{|\alpha| \leq m} \langle D^\alpha F, a_\alpha(x) \varphi \rangle = (-1)^{|\alpha|} \sum_{|\alpha| \leq m} \langle F, D^\alpha (a_\alpha(x) \varphi) \rangle = \langle F, L' \varphi \rangle \tag{33}\]

This lemma permits us to introduce the idea of a distributional solution of a differential equation:

**DEFINITION.** A distribution \(U\) satisfies the partial differential equation \(Lu = f\) in the sense of distributions if

\[
LU = Ff \tag{34}\]

In general this means that we should regard the non-homogeneous term \(f\) as at minimum a locally integrable function.

The following result demonstrates the importance of taking a distributional solution and going the extra step (if possible) of showing that the distribution is represented by an actual function:

**LEMMA.** If \(U\) satisfies \(Lu = f\) in the sense of distributions, and if \(U\) is representable as a locally integrable function \(u\), i.e. \(U = Fu\), then \(u\) is a weak solution of \(Lu = f\).

**Proof.** By the definition of a weak solution, we need to show that

\[
\int_\Omega u L' \varphi dx = \int_\Omega f \varphi dx \tag{35}\]

for all test functions \(\varphi\). Equivalently we can use the representation of \(U\) and distribution notation to write this as

\[
\langle U, L' \varphi \rangle = \langle Ff, \varphi \rangle \tag{36}\]

and then the preceding lemma allows us to write this as

\[
\langle LU, \varphi \rangle = \langle Ff, \varphi \rangle \tag{37}\]

But this is satisfied because \(U\) is a distributional solution, i.e. satisfies (34). Hence \(u\) is a weak solution.

**Convolutions, Fundamental Solutions, and the Non-Homogeneous Problem**

The convolution product of functions arises in many areas of applied mathematics where integral transforms and integral operators are present. Generally speaking, the interaction of an integral operator and a convolution product leads to important simplifications in the formulation of a problem. Given that distributions, at least those
that arise from locally integrable functions, are integral operators, it should not be surprising that convolutions will be important in our discussion here.

The convolution product of two functions \( f \) and \( g \) defined on \( \mathbb{R}^n \) is the function \( f \ast g \) whose values are given by

\[
(f \ast g)(x) = \int_{\mathbb{R}^n} f(x-y)g(y)dy
\]  

(38)

Generally, in order for this integral to converge we require that at least one of the two functions have compact support. If either function is a test function, we are fine, but otherwise we will have to be careful. This will pose an interesting issue when we try to generalize the convolution product to distributions.

The generalization of the convolution to distributions has a computational side and also a technical side. Let’s consider the former first, so as to get to a definition of the convolution. We begin by generating the distribution \( F_{f \ast g} \) corresponding to \( f \ast g \) and then rearranging its definition so that it is expressed completely in terms of \( F_f \) and \( F_g \). Here we go, using (2):

\[
\langle F_{f \ast g}, \varphi \rangle = \int_{\mathbb{R}^n} (f \ast g)(x)\varphi(x)dx
\]

\[
= \int_{\mathbb{R}^n} \int_{\mathbb{R}^n} f(x-y)g(y)\varphi(x)dydx
\]

\[
= \int_{\mathbb{R}^n} \int_{\mathbb{R}^n} f(z)g(y)\varphi(y+z)dydz
\]

(39)

Now let us define a translation operator \( T \) by

\[
T_z \varphi(y) = \varphi(y+z)
\]

(40)

For each \( z \), \( T_z \varphi \) is a test function in the variable \( y \), and therefore we can view the \( y \) integration in (39) as the value of the distribution \( F_g \) at \( T \varphi \):

\[
\langle F_{f \ast g}, \varphi \rangle = \int_{\mathbb{R}^n} f(z)\langle F_g, T_z \varphi \rangle dz
\]

(41)

In this form we see that the integration with respect to \( z \) can be rewritten as a value of the distribution \( F_f \). This gives us the final result

\[
\langle F_{f \ast g}, \varphi \rangle = \langle F_f, \langle F_g, T \varphi \rangle \rangle
\]

(42)

A careful look at the preceding calculation shows that we could have switched the roles of \( f \) and \( g \) throughout.

If we formulate the definition of convolution for distributions correctly, the left side of (42) should be the convolution of \( F_f \) and \( F_g \). This provides us with motivation for the following definition:

**DEFINITION.** The convolution \( F \ast G \) of two distributions on \( \mathbb{R}^n \) is the distribution with values

\[
\langle F \ast G, \varphi \rangle = \langle F, \langle G, T \varphi \rangle \rangle = \langle G, \langle F, T \varphi \rangle \rangle.
\]

(43)

Before we look at the technical issues surrounding this definition, let us explore some of the properties of convolutions and find out why they are useful. We begin by noting directly from (43) that \( F \ast G = G \ast F \). This agrees with the corresponding property for functions. Next we note the following:

**LEMMA.** \( F \ast \delta = \delta \ast F = F \) The Dirac distribution is the identity element with respect to the convolution product.

**Proof.** For all test functions \( \varphi \) we have

so on half is verified. The other follows by symmetry of the convolution product, but here is a direct proof:

\[ \langle \delta * F, \varphi \rangle = \langle \delta, F \rangle = \langle F, T \varphi \rangle = \langle F, T \varphi \rangle \big|_{z=0} = \langle F, T_0 \varphi \rangle = \langle F, \varphi \rangle \]  

(45)

**Lemma.** \[ D^\alpha (F * G) = (D^\alpha F) * G = F * D^\alpha (G) \]  

(46)

**Proof.** We simply use the definitions of the convolution and the derivative of distributions:

\[ \langle D^\alpha (F * G), \varphi \rangle = (-1)^{\alpha_1} \langle F * G, D^\alpha \varphi \rangle = \langle F, (G, T D^\alpha \varphi) \rangle \]  

(47)

But \( T_z D^\alpha \varphi(y+z) = D^\alpha \varphi(y+z) \), so this last steps becomes

\[ \langle D^\alpha (F * G), \varphi \rangle = (-1)^{\alpha_1} \langle F, (G, D^\alpha T \varphi) \rangle = \langle F, (G, D^\alpha T \varphi) \rangle = \langle F, D^\alpha (F * G), \varphi \rangle \]  

(48)

This completes one half of the identity. The other part is proved in the same way.

If \( v \) is a \( C^\infty \) function, then it is generally not true that \( v(F * G) = (vF) * G = F * (vG) \). That is, we cannot carry a general function inside the members of a convolution. You can easily convince yourself, using (38), that this identity doesn’t even hold for functions. However, it is true if \( v \) is a constant. This follows from the fact that any expression of the form \( \langle F, \varphi \rangle \) is linear in \( F \). Given this fact, we can do special things with linear constant coefficient operators. In particular, generalizing (46), we have

\[ L(F * G) = (LF) * G = F * (LG) \]  

where \( L = \sum_{|\alpha| \leq m} c_\alpha D^\alpha \)  

(49)

This brings us to the most important property of convolutions:

**Definition.** A fundamental solution of the constant coefficient PDE \( Lu = f \) is a distribution \( U \) satisfying \( LU = d \).

**Theorem.** If \( U \) is a fundamental solution for \( L \), then \( u = U * F_f \) is a distributional solution of \( Lu = f \), i.e. \( Lu = F_f \).

**Proof.**

\[ Lu = L(U * F_f) = (Lu) * F_f = \delta * F_f = F_f \]  

(50)

Therefore, simply by taking convolutions with a fundamental solution we can generate a distributional solution of the non-homogeneous equation for a general right-hand side. In the next section we will investigate the problems involved in finding and using fundamental solutions.

Before moving on, let us at least recognize the technical issues involved in the definition of the convolution. Let us define a function \( \psi \) by:

\[ \psi(z) = \langle G, T_z \varphi \rangle \]  

(51)

That is, \( \psi \) is the result of applying the distribution \( G \) to the \( z \) translate of the test function \( \varphi \). Before the second expression in the definition (43) can make sense, we need to be sure that \( \psi \) is a test function, i.e. lies in \( D(R^n) = C_0^\infty (R^n) \). There are two sides to this question, namely, a) is the function infinitely differentiable and b) does it have compact support. The differentiability question proceeds from a calculation:
\[
D_{z_k} \psi = \lim_{h \to 0} \frac{1}{h} (\psi(z + he_k) - \psi(z)) \quad \text{where } e_k \text{ is a unit vector in the } k^{th} \text{ direction}
\]

\[
= \lim_{h \to 0} \frac{1}{h} \langle G, T_{z+he_k} \phi - T_z \phi \rangle \quad \text{since we have linearity in the second argument}
\]

\[
= \langle G, \lim_{h \to 0} \frac{1}{h} (T_{z+he_k} \phi - T_z \phi) \rangle \quad \text{since } G \text{ is a continuous operator}
\]

\[
= \langle G, D_{z_k} T_z \phi \rangle
\]

\[
= \langle G, T_z D_{y_k} \phi \rangle \quad \text{where } \phi \text{ is viewed as a function of } y
\]

We conclude that the differentiability of \( \psi \) follows directly from that of \( \phi \), and the linearity and continuity assumed in the definition of a distribution.

Now what do we do about the matter of compact support? For this we need to force it to be true by making an additional assumption. Here is a related definition:

**DEFINITION.** A distribution \( F \) has support in a compact set \( K \) if \( \langle F, \phi \rangle = 0 \) for any test function that vanishes outside of \( K \). That is, only when part of the support of \( \phi \) intersects \( K \) will \( F \) give a non-zero value at \( \phi \).

**EXAMPLE.** Demonstrate that the support of the Dirac distribution \( \delta \) is the single point \( x = 0 \).

Let us look back now at (51) and require that the distribution have support in a compact set \( K \). The effect of the translation operation \( T_z \) is to translate the support of the test function \( \phi \) around \( \mathbb{R}^n \). If we let \( |z| \to \infty \), we will move the support of \( \phi \) out sufficiently far that it no longer intersects \( K \), and this implies that \( \psi(z) = 0 \). Therefore, in order for the definition of the convolution product of two distributions to make sense, it is enough to require that at least one of those distributions has compact support. Since our most crucial use of the convolution is in the identity \( F * \delta = \delta * F = F \), the fact that \( \delta \) has compact support justifies the calculations.

**Practical Calculations Using Distributions**

Are there any practical strategies for looking for distributional solutions of a PDE? The answer to this question is contingent on one’s experience, but certainly some rules of thumb can be given. Here are a few that I have found in my years of study:

**Simple candidates.** In one-dimensional problems, try the simplest candidates first. Generally this means a trying functions that may possibly be discontinuous or have a singularity. For example, in considering the equation

\[
u' = \delta
\]

we would look for a (possibly discontinuous) function which, when differentiated, gives \( \delta \). Experience tells us that the Heaviside distribution has this property, so start there. For the problem

\[
u'' = \delta
\]

we need to have a function whose second derivative gives \( \delta \). A possibility is the integral of the Heaviside function, namely the function \( xH(x) \), so start with this as a candidate. And remember: a fundamental solution is only unique up to the addition of a solution of the homogeneous problem.
Reduction to functions. If the right-hand side is a distribution that behaves like a function in regions, then solve for function-solutions in those regions and then paste these solutions together. In both (52) and (53) we might view $\delta$ as behaving like the function $0$ for $x < 0$ and for $x > 0$, so solve $u'' = 0$ in each region and form a (discontinuous function) by piecing them together. Then, as a final step, see if there are any further restrictions by trying to satisfy (52) or (53) directly in a distributional sense. This approach is perhaps seen better in the case of the equation

$$u' - au = \delta \tag{54}$$

From $x < 0$ and for $x > 0$ we look at the equation $u' - au = 0$ and get the solution $ce^{ax}$. So try:

$$u(x) = \begin{cases} c_1 e^{ax} & \text{for } x < 0 \\ c_2 e^{ax} & \text{for } x > 0 \end{cases} \tag{55}$$

The next step is to write out the distributional form of (54), namely,

$$\langle u, -\phi' - a\phi \rangle = \phi(0) \tag{56}$$

(see where this comes from?) and then to substitute your candidate $u$ using (since $u$ is a locally integrable function) the representation (2). This gives us:

$$\int_{-\infty}^{0} c_1 e^{ax} (-\phi' - a\phi) dx + \int_{0}^{\infty} c_2 e^{ax} (-\phi' - a\phi) dx = \phi(0) \tag{57}$$

Now it is a matter of integration by parts and simplification. Remember that this equation must be satisfied for all test functions $\phi$, so you must choose $c_1$ and $c_2$ so that $\phi$ eventually drops out. This same approach can be used to attack the second-order problem

$$u'' - a^2 u = \delta \tag{58}$$

since the homogeneous equation $u'' - a^2 u = 0$ can be readily solve for $x < 0$ and for $x > 0$ (though two constants arise now for each region). You should also be able to attack a problem such as

$$u' - au = F_H \tag{59}$$

using the same strategy (i.e. the right hand-side is the Heaviside distribution).

Building up more examples. At this stage our grab bag of distributions is small, so we might benefit from constructing more examples. Here is a new one you should try. Consider the function $u(x) = 1/\sqrt{|x|}$. This function is locally integrable but has a singularity at $x = 0$. So what does its distributional derivative look like? How about its second derivative? Perhaps you will find a new set of problems that you can attack using such new distributions.

Separation of variables in higher dimensional problems. Separation of variable is a common way of reducing a multi-dimensional problem to one-dimensional problems. For the equation

$$u_{xy} = \delta \tag{60}$$

the right-hand side is the two-dimensional Dirac distribution. We know we can write it as a tensor product of one-dimensional versions to get:

$$u_{xy} = \delta^1 \otimes \delta^1 \tag{61}$$

Now look for a separated distributional solution by assuming $u = v \otimes w$ where $v$ is a distribution on $\mathbb{R}$ in $x$ and $w$ is a distribution on $\mathbb{R}$ in $y$. Using the properties of distributions presented in previous sections, we see that (61) becomes
\[ v_x \otimes w_y = \delta^1 \otimes \delta^1 \]  \hspace{1cm} (62)

You can satisfy this equation by pursuing the separate one-dimensional problems

\[ v_x = k\delta^1 \]
\[ w_y = (1/k)\delta^1 \]  \hspace{1cm} (63)

where \( k \) is a constant (the choice of \( k \) may be used as an added degree of flexibility in finding solutions). Note that not all problems have variables separable so this method has its limitations.

*Changes of variables to simplify higher dimensional problems.* It would be nice to have a theory of coordinate transformations within the theory of distributions. This would permit us to attack, for example, a hyperbolic problem by reducing it to canonical variable. Stay tuned. I will see if I can come up with an approach. Or perhaps you have one? Here is a very simple problem to get you thinking about it. \( K(x) = H(-x) \) is the function that is 1 for \( x < 0 \) and 0 for \( x > 0 \). How is the distribution associated with \( K \) related to the Heaviside distribution? Can you work with more general transformations, i.e. \( H(ax) \), etc.? Now keep going!