APPROXIMATION BY SMOOTH FUNCTIONS AND DISTRIBUTIONS

S.Berhanu

http://www.math.temple.edu/~berhanu

December 26, 2001

0.Introduction

This article introduces the reader to an important class of distributions called **tempered distributions**. These distributions are continuous linear functionals on the Schwartz space of rapidly decreasing functions. Future articles will illustrate the applications of distribution theory to partial differential equations and Fourier analysis. The first two sections contain some basic material on smooth functions of compact support and it is shown how such functions can be used to approximate L^p functions and continuous functions. Section 3 begins with the space of rapidly decaying functions, introduces tempered distributions and presents their basic properties.

The prerequisite for understanding this article is some familiarity with measure theory and integration.

The results discussed here will provide a good background for future introductory articles on Fourier Analysis and its applications.

Notations

We will use the following standard notations and definitions. Functions will generally be defined on subsets of $\mathbb{R}^n$ and will be complex-valued. By a multi-index $\alpha = (\alpha_1, \cdots, \alpha_n)$ we mean an $n-$tuple of nonnegative integers. If $\alpha = (\alpha_1, \cdots, \alpha_n)$ is a multi-index, $\partial_x^\alpha f$ or $\dfrac{\partial^\alpha f}{\partial x^\alpha}$ will denote the derivative $\left(\dfrac{\partial}{\partial x_1}\right)^{\alpha_1} \cdots \left(\dfrac{\partial}{\partial x_n}\right)^{\alpha_n} f$, $|\alpha| = \alpha_1 + \alpha_2 + \cdots + \alpha_n$, and $x^\alpha =$ the monomial $x_1^{\alpha_1} \cdots x_n^{\alpha_n}$. $C(\mathbb{R}^n)$ will denote the space of continuous functions on $\mathbb{R}^n$.

For a positive integer k, the space of functions that are k times continuously differentiable is denoted by

$$C^k(\mathbb{R}^n) = \{f \in C(\mathbb{R}^n) : \partial_x^\alpha f \in C(\mathbb{R}^n) \text{ for each } \alpha \text{ with } |\alpha| \leq k\}$$

The notation $C^0(\mathbb{R}^n)$ will sometimes be used for $C(\mathbb{R}^n)$.

$$C^\infty(\mathbb{R}^n) = \{f : f \in C^k(\mathbb{R}^n) \text{ for every } k\}.$$

For $p \in \mathbb{R}^n$ and $r > 0$, the open ball centered at p with radius r will be denoted by $B_r(p)$. The support of a function f denoted by $\text{supp}(f)$ is the closure of $\{x : f(x) \neq 0\}$. Equivalently, if $W =$ the union of all open sets where f vanishes, then $\text{supp}(f) = \mathbb{R}^n \backslash W$.

Note that $\text{supp}(f)$ is a closed set and if f is continuous, it is zero on its boundary. Clearly, $\text{supp}(f)$ may be strictly larger than the set where $f \neq 0$.

Example 0.1. *Let*

$$f(x) = \begin{cases} 1, & \textit{if } x \in [0,1] \cap Q \\ 0, & \textit{otherwise} \end{cases}$$

where $Q =$ the rational numbers. Then, supp$(f) = [0,1]$.

For k a nonnegative integer,

$$C_0^k(\mathbb{R}^n) = \big\{f \in C^k(\mathbb{R}^n) : \text{supp}(f) \text{ is compact}\big\}$$

$$C_0^\infty(\mathbb{R}^n) = \{f \in C^\infty(\mathbb{R}^n) : \text{supp}(f) \text{ is compact}\}$$

For $1 \le p < \infty$, $L^p(\mathbb{R}^n)$ denotes the usual space of measurable functions whose p^{th} power is Lebesgue integrable. $L^\infty(\mathbb{R}^n)$ denotes the space of essentially bounded functions.

1. Examples of functions in C^k but not in C^{k+1}

It is clear that $C^{k+1} \subseteq C^k$ for each k. We will now show that this inclusion is strict. We will also present examples that are in C_0^k but not in C_0^{k+1}. First, we focus on $\mathbb{R}$. The function

$$f_0(x) = \begin{cases} x, & x \ge 0 \\ 0, & x < 0 \end{cases}$$

is in $C(\mathbb{R})$ but not in $C^1(\mathbb{R})$ since f_0 is not differentiable at 0. Next, observe that

$$f_1(x) = \begin{cases} x^2, & x \ge 0 \\ 0, & x < 0 \end{cases}$$

is in $C^1(\mathbb{R})$ but $f_1 \notin C^2(\mathbb{R})$ since $f_1' = 2f_0$ is not in C^1. In general,

$$f_k(x) = \begin{cases} x^{k+1}, & x \ge 0 \\ 0, & x < 0 \end{cases}$$

is in C^k but not in C^{k+1}.

To get analogous examples of compact support, it suffices to take

$$g_k(x) = f_k(x)\, f_k(1-x)$$

which is in $C_0^k(\mathbb{R})$ but not in $C_0^{k+1}(\mathbb{R})$.

Exercise 1.1. *Give an example of* $f \in C^k(\mathbb{R}^n)\,(n>1)$ *that is not in* $C^{k+1}(\mathbb{R}^n)$.

Exercise 1.2. *Give an example of* $f \in C_0^k(\mathbb{R}^n)\,(n>1)$ *that is not in* $C_0^{k+1}(\mathbb{R}^n)$.

2. Approximation by C^∞ functions of compact support

This section considers first the construction of smooth functions supported in a given open set. It turns out that $C_0^\infty(\mathbb{R}^n)$ is dense in $L^p(\mathbb{R}^n)$ for $1 \le p < \infty$. For $p = \infty$, there is a substitute result. Namely, suppose we equip $C(\mathbb{R}^n)$ with its usual topology where a sequence $\{f_k\}_k$ in $C(\mathbb{R}^n)$ is said to converge to f if on every compact set K, the sequence converges uniformly to f on K. Then when $C(\mathbb{R}^n)$ is given this topology $C_0^\infty(\mathbb{R}^n)$ is dense in $C(\mathbb{R}^n)$. These density results will be proved in section 2. The fact that functions in $L^p(\mathbb{R}^n)$ or $C(\mathbb{R}^n)$ can be approximated by C_0^∞ functions means that in order to prove many results involving $L^p(\mathbb{R}^n)$ or $C(\mathbb{R}^n)$, we only have to establish the results for C_0^∞. Working with C_0^∞ functions is very convenient since they are infinitely differentiable and vanish outside a compact set—making it easy to apply many operations of analysis such as differentiation, integration, differentiation under the integral, Fourier transformation, etc. Later sections will illustrate this. Before proceeding further, the reader is advised at this point to spend some time trying to construct a nontrivial element of $C_0^\infty(\mathbb{R}^n)$, say for instance, an example of a $\varphi \in C_0^\infty(\mathbb{R}^n)$ such that $\varphi(x) \equiv 1$ for $|x| \le 1$. In order to construct a nontrivial example in $C_0^\infty(\mathbb{R})$, we will be guided by the examples f_k and g_k of section 1. Note that for each k, the function f_k has the following properties:

(1) $f_k \equiv 0$ for $x \le 0$;

(2) $f_k(x) > 0$ for $x > 0$;

(3) At 0, the j^{th} derivative $f_k^{(j)}(0) = 0$ for $0 \le j \le k$

Exercise 2.1. *Suppose $h \in C^k(\mathbb{R})$. Show that $h^{(j)}(0) = 0$ for $0 \le j \le k$ iff there is $C > 0$ such that $|h(x)| \le C|x|^{k+1}$ for $x \in (0,1)$.*

It therefore seems reasonable to first construct $f \in C^\infty(\mathbb{R})$ satisfying

i) $f^{(j)}(0) = 0 \quad \forall j$; and

ii) f is not identically 0 on $\mathbb{R}$.

By exercise 2.1, such an f must satisfy the following: for each k, there is $C_k > 0$ such that $|f(x)| \le C_k|x|^k$ for $x \in (0,1)$. A standard example satisfying these properties is given by

Lemma 2.1. *Define*

$$f(x) = \begin{cases} e^{-\frac{1}{x}}, & x > 0 \\ 0, & x \le 0 \end{cases}$$

Then $f \in C^\infty(R)$, and $f^{(j)}(0) = 0 \quad \forall j = 1, 2, \cdots$.

Proof. Clearly f is C^∞ on $\mathbb{R}\backslash\{0\}$. Therefore, we only have to show that $D^k f$ exists at 0 for each k. Indeed, this would then imply $D^m f$ exists and is continuous on $\mathbb{R}$ for each m. Recall that $e^t = \sum_{m=0}^{\infty} \frac{t^m}{m!}$, and so for each $m = 1, 2, \cdots$,

$$\left| e^{-\frac{1}{x}} \right| \le m!\, x^m \ \text{ for } \ x > 0.$$

It follows that f is continuous at 0 and so $f \in C(\mathbb{R})$. Observe next that for any $m = 1, 2, \cdots$

$$\left|\frac{f(x)}{x}\right| \leq (m+1)!\,|x|^m \quad \text{for} \quad x \neq 0.$$

Therefore, f is differentiable at 0 and $f'(0) = 0$. We then get

$$f'(x) = \begin{cases} \frac{1}{x^2}\, e^{-\frac{1}{x}}, & x > 0 \\ 0, & x \leq 0 \end{cases}$$

Proceeding by induction, assume the k^{th} derivative of f exists at 0 and

$$f^{(k)}(x) = \begin{cases} P\left(\frac{1}{x}\right)\, e^{-\frac{1}{x}}, & x > 0 \\ 0, & x \leq 0 \end{cases}$$

where $P(y)$ is a polynomial of degree $k+1$. Then, there is a $C > 0$ such that for $|x| \leq 1$, $\left|P\left(\frac{1}{x}\right)\right| \leq \frac{C}{|x|^{k+1}}$. Using the inequality $e^{-\frac{1}{x}} \leq (k+3)!\, x^{k+3}$ for $x > 0$, we then get

$$\left|\frac{f^{(k)}(x)}{x}\right| \leq C\ (k+3)!\ |x| \ \text{for} \ \ |x| < 1.$$

Hence $f^{(k+1)}$ exists at 0 and $f^{(k+1)}(0) = 0$.

Remark 2.1. *The function f in Lemma 2.1 may be familiar to the reader who may have seen it in elementary calculus. f is an example of an infinitely differentiable function which is not equal to its Taylor series centered at* 0.

We can now use Lemma 2.1 to easily construct several examples of functions in $C_0^\infty(\mathbb{R}^n)$ for any n.

Example 2.1. *Let f be as in Lemma 2.1. Then $g(x) = f(x)\, f(1-x) \in C_0^\infty(\mathbb{R})$. $g(x) > 0$ iff $x \in (0,1)$ and supp$(g) = [0,1]$.*

Example 2.2. *Let g be as above, $p \in \mathbb{R}$ and $\delta > 0$. Then*

$$h_\delta(x) = g\left(\frac{1}{2}\left(\frac{x-p}{\delta} + 1\right)\right) \in C_0^\infty(\mathbb{R}),$$

supp$(h_\delta) = [p-\delta, p+\delta]$, and $h_\delta > 0$ on $(p-\delta, p+\delta)$.

Example 2.3. *By example 2.2 we can get $h \in C_0^\infty(\mathbb{R})$ such that supp$(h) = [-1, 1]$, $h > 0$ on $(-1, 1)$. Then $h(|x|^2) = h(x_1^2 + \cdots + x_n^2) \in C_0^\infty(\mathbb{R}^n)$ and it is supported in the closure of the unit ball $B_1(0)$.*

Our next objective is to show that in fact C_0^∞ functions exist in some abundance. More precisely, we will show that given a compact set $K \subseteq \mathbb{R}^n$, and an open neighborhood V of K, there is $\varphi \in C_0^\infty(\mathbb{R}^n)$ such that (i) $\varphi \equiv 1$ on K, and (ii) supp $\varphi \subseteq V$. This result will imply that we can approximate the characteristic function $\chi_K(x)$ of the set K by a sequence $\{\varphi_j\}$ of C_0^∞ functions. Indeed, let $V_j = \left\{x : \text{ dist } (x, K) < \dfrac{1}{j}\right\}$. V_j is an open set containing K and so we have $\varphi_j \in C_0^\infty(\mathbb{R}^n)$ such that $0 \leq \varphi_j \leq 1$, $\varphi_j|_K \equiv 1$ and $\text{supp}(\varphi_j) \subseteq V_j$. It is clear that

$$(1) \quad \varphi_j(x) \longrightarrow \chi_K(x) \text{ pointwise on } \mathbb{R};$$

$$(2) \quad \varphi_j \longrightarrow \chi_K \text{ in } L^p \text{ for } 1 \leq p < \infty.$$

Once we have approximated the characteristic functions χ_K by elements of C_0^∞, the density of C_0^∞ in L^p $(1 \leq p < \infty)$ won't come as a surprise since the set of step functions $\left\{\sum_{j=1}^{m} a_j \, \chi_{K_j} : a_j \in \mathbb{C}, K_j \text{ compact}\right\}$ is dense in L^p. To achieve our goal of approximating χ_K by C_0^∞ functions, we will use the following basic result which supplies a sufficient condition for the boundedness (i.e. continuity) in L^p of integral operators.

Theorem 2.2. *Let $k(x, y)$ be a measurable function on $\mathbb{R}^n \times \mathbb{R}^n$ such that for some $c > 0$,*

$$\int_{\mathbb{R}^n} |k(x, y)| \, dx \leq c \quad \forall y \in \mathbb{R}^n, \text{ and}$$

$$\int_{\mathbb{R}^n} |k(x, y)| \, dy \leq c \quad \forall x \in \mathbb{R}^n.$$

Then for $1 \le p \le \infty$ *and* $f \in L^p(\mathbb{R}^n)$, *the function* $Tf(x) = \displaystyle\int_{\mathbb{R}^n} k(x,y)\, f(y)\, dy$ *is in* $L^p(\mathbb{R}^n)$ *and* $\|Tf\|_p \le c\, ||f||_p$, *where* c *is as in the hypothesis.*

Remark 2.2. *It may not be clear at the outset why the integral*

$$Tf(x) = \int_{\mathbb{R}^n} k(x,y) f(y)\, dy$$

makes sense for $f \in L^p$. *The proof will show that* $\int_{\mathbb{R}^n} |k(x,y)f(y)|\, dy < \infty$ *for almost every* x, *implying that* $Tf(x)$ *is defined a.e.*

Proof. Observe that the theorem is trivial for $p = \infty$, and so assume $1 \le p < \infty$. Let $\frac{1}{q} + \frac{1}{p} = 1$. Then,

$$\begin{aligned} \int |k(x,y)|\, |f(y)|\, dy &= \int |k(x,y)|^{\frac{1}{q}} . \left(|k(x,y)|^{\frac{1}{p}} . |f(y)| \right) dy \\ &\le \left(\int |k(x,y)|\, dy \right)^{\frac{1}{q}} \left(\int |k(x,y)|\, |f(y)|^p dy \right)^{\frac{1}{p}} \\ &\qquad \text{(by Hölder's inequality)} \\ &\le c^{\frac{1}{q}} \left(\int |k(x,y)|\, |f(y)|^p\, dy \right)^{\frac{1}{p}} \end{aligned}$$

Integrating in x we get:

$$\begin{aligned} \int \left(\int |k(x,y)|\, |f(y)|\, dy \right)^p dx = &\le c^{\frac{p}{q}} \int \left(\int |k(x,y)|\, |f(y)|^p\, dy \right) dx \\ &\le c^{\frac{p}{q}} \int \left(\int |k(x,y)|\, dx \right) |f(y)|^p\, dy \\ &\qquad \text{(by Tonelli)} \\ &\le c^{\frac{p}{q}+1}\, \|f\|_p^p \end{aligned}$$

Hence $\|Tf\|_p \le c\, \|f\|_p$.

Definition 2.1. *If f and g are measurable functions on $\mathbb{R}^n$, we define the convolution of f and g, denoted $f * g$ by*

$$f * g(x) = \int_{\mathbb{R}^n} f(x-y)\, g(y)\, dy$$

whenever the integral exists.

By a simple change of variable, one easily gets $f * g = g * f$. Clearly, $f * g$ is defined if say $f \in L^1(\mathbb{R}^n)$ and g is a bounded function. More examples together with an estimate on the norm of $f * g$ follow from:

Corollary 2.3. *(Young's inequality) If $f \in L^1(\mathbb{R}^n)$ and $g \in L^p(\mathbb{R}^n)$ $(1 \leq p \leq \infty)$, then $f * g \in L^p(\mathbb{R}^n)$ and $\|f * g\|_p \leq \|f\|_1 \, \|g\|_p$.*

Proof. Fix $f \in L^1$ and for $g \in L^p$ define

$$Tg(x) = \int_{\mathbb{R}^n} f(x-y)\, g(y)\, dy.$$

The corollary then follows from Theorem 2.2 upon setting $k(x,y) = f(x-y)$.

Exercise 2.2. *Prove the following generalization of Young's inequality: Suppose $\frac{1}{p} + \frac{1}{q} = 1 + \frac{1}{r}$. If $f \in L^q$ and $g \in L^p$, show that $f * g \in L^r$ and that $||f * g||_r \leq ||f||_q ||g||_p$*

Proposition 2.4. *Suppose f and g are continuous functions of compact support. Then $f * g$ is a continuous function of compact support and*

$$supp(f * g) \subseteq supp(f) + supp(g)$$

(Here, for two sets A and B in $\mathbb{R}^n$, $A + B = \{a + b : a \in A,\, b \in B\}$.)

Proof. Note that f is uniformly continuous on $\mathbb{R}^n$ and hence

$$f * g(x) = \int f(x-y) g(y)\, dy$$

is also uniformly continuous on $\mathbb{R}^n$. Suppose now $p \notin \text{supp}(f) + \text{supp}(g)$. Since $\text{supp}(f)$ and $\text{supp}(g)$ are compact, the sum $\text{supp}(f) + \text{supp}(g)$ is also compact. Therefore, there is a ball $B_r(p)$ disjoint from $\text{supp}(f) + \text{supp}(g)$. Now if $x \in B_r(p)$, we claim that $f * g(x) = 0$. Otherwise, $\displaystyle\int f(x-y)\,g(y)\,dy \neq 0$, implying the existence of y such that $f(x-y)\,g(y) \neq 0$. That is, $x - y \in \text{supp}(f)$, $y \in \text{supp}(g)$ and so $x = (x-y)+y \in \text{supp}(f)+ \text{supp}(g)$, contrary to our choice of $B_r(p)$. Hence $f * g = 0$ on $B_r(p)$ and so $p \notin \text{supp}\ (f * g)$.

A central application of convolution is in regularizing or approximating a function by smoother functions. For this, we need to know first that $f * g$ belongs to a differentiability class that is at least as good as that of f and g. Roughly speaking, if $f \in C^k$ and $g \in C^m$, then $f * g \in C^{k+m}$. More precisely, we have :

Theorem 2.5.

(1) *If* $f \in C_0^k(\mathbb{R}^n)$ *and* $g \in L^1_{loc}(\mathbb{R}^n)$, *then* $f * g \in C^k(\mathbb{R}^n)$ *and* $\partial_x^\alpha(f * g) = \partial_x^\alpha f * g$ *for* $|\alpha| \leq k$.

(2) *If* $f \in C_0^k(\mathbb{R}^n)$ *and* $g \in C^m(\mathbb{R}^n)$, *then* $f * g \in C^{k+m}(\mathbb{R}^n)$.

Proof. We will prove (1) in two steps. Step(i): Suppose first $k = 0$. We will show $f * g \in C(\mathbb{R}^n)$. Fix $x \in \mathbb{R}^n$ and let $x_j \longrightarrow x$. Let $\epsilon > 0$. Since f is continuous and compactly supported, it is uniformly continuous and so there is $\delta > 0$ such that $|p - q| < \delta \implies |f(p) - f(q)| < \epsilon$. Let $|x_j| \leq M$ for all j. Then, the functions $f_j(y) = f(x_j - y)$ are all supported in a fixed compact set K. Therefore, if we choose j_o so that $|x_j - x| < \delta$ whenever $j \geq j_o$, we then have:

$$\begin{aligned}|f * g(x_j) - f * g(x)| &\leq \int |f(x_j - y) - f(x - y)|\,|g(y)|\,dy \\ &\leq \epsilon \int_K |g(y)|\,dy\end{aligned}$$

This shows that $f * g$ is continuous.

Step(ii): Suppose now $f \in C_0^1$. We will show that $f * g \in C^1$. Let $x \in \mathbb{R}^n$ and $e_j = (0, \cdots, 1, \cdots, 0)$ with 1 in the j^{th} position, and 0 otherwise. For $t > 0$ we have

$$\begin{aligned}\frac{f * g(x + te_j) - f * g(x)}{t} &= \frac{1}{t} \int [f(x - y + te_j) - f(x - y)]\, g(y)\, dy \\ &= \int \left[\int_0^1 \frac{\partial f}{\partial x_j}(x - y + s\,t\,e_j)\, ds \right] g(y)\, dy\end{aligned}$$

where in the latter we applied the fundamental theorem of calculus to write

$$f(z + t\,e_j) - f(z) = \int_0^1 \frac{d}{ds} f(z + s\,t\,e_j)\, ds$$

It follows that

$$\begin{aligned}&\frac{f * g(x + te_j) - f * g(x)}{t} - \frac{\partial f}{\partial x_j} * g(x) \\ &= \int \left(\int_0^1 \frac{\partial f}{\partial x_j}(x - y + s\,t\,e_j)\, ds - \frac{\partial f}{\partial x_j}(x - y) \right) g(y)\, dy \\ &= \int \left(\int_0^1 \left[\frac{\partial f}{\partial x_j}(x - y + s\,t\,e_j) - \frac{\partial f}{\partial x_j}(x - y) \right] ds \right) g(y)\, dy\end{aligned}$$

Since $\frac{\partial f}{\partial x_j}$ has compact support, there is $M > 0$ such that for any s and t in $[0, 1]$, the support of

$$y \longmapsto \frac{\partial f}{\partial x_j}(x - y + ste_j) - \frac{\partial f}{\partial x_j}(x - y)$$

is contained in the ball $B_M(0)$. Moreover $\frac{\partial f}{\partial x_j}$ is uniformly continuous and so given $\epsilon > 0$, there is $\delta > 0$ such that $\left| \frac{\partial f}{\partial x_j}(u) - \frac{\partial f}{\partial x_j}(z) \right| < \epsilon$ whenever $|u - z| < \delta$. Therefore, for $|t| < \delta$, we get

$$\begin{aligned}&\left| \frac{f * g(x + te_j) - f * g(x)}{t} - \frac{\partial f}{\partial x_j} * g(x) \right| \\ &\le \int_K \left(\int_0^1 \left| \frac{\partial f}{\partial x_j}(x - y + ste_j) - \frac{\partial f}{\partial x_j}(x - y) \right| ds \right) |g(y)| dy \\ &\le \epsilon \int_K |g(y)|\, dy\end{aligned}$$

Hence, $\frac{\partial(f*g)}{\partial x_j} = \frac{\partial f}{\partial x_j} * g$. Furthermore, since $\dfrac{\partial f}{\partial x_j}$ is continuous, by step (i) we see that $\dfrac{\partial}{\partial x_j}(f*g)$ is continuous, that is, $f*g \in C^1(\mathbb{R}^n)$.

We can now iterate easily this C^1 result to conclude that if $f \in C_0^k$, then $f*g \in C^k$ and that for any α with $|\alpha| \leq k$, $\partial_x^\alpha(f*g) = \partial_x^\alpha f * g$. To prove (2), note first that by (1), for any α with $|\alpha| \leq k$, $\partial_x^\alpha(f*g) = \partial_x^\alpha f * g$. Fix $|\alpha| \leq k$ and consider $\partial_x^\alpha(f*g) = \partial_x^\alpha f * g$. The argument in step (ii) shows that $\partial_x^\alpha f * g \in C^m$ and that if $|\beta| \leq m$,

$$\partial_x^\beta \partial_x^\alpha (f*g) = \partial_x^\beta (\partial_x^\alpha f * g) = \partial_x^\alpha f * \partial_x^\beta g.$$

Indeed, although the support of g may not be compact, in the integrals, the values of y vary only in a compact set since f is compactly supported. Thus, $f*g \in C^{m+k}$.

Lemma 2.6. *Let $1 \leq p < \infty$ and $f \in L^p(\mathbb{R}^n)$. For $x \in \mathbb{R}^n$, set $f^x(y) = f(y-x)$. Then $\lim_{x\to 0} \|f^x - f\|_p = 0$.*

Remark 2.3. *Lemma 2.6 is not valid for $p = \infty$. In fact,*

$$\lim_{x\to 0} \|f^x - f\|_\infty = 0 \Longleftrightarrow f \text{ is uniformly continuous on } \mathbb{R}^n.$$

Proof of Lemma 2.6. We will first show that $C_0^0(\mathbb{R}^n)$ is dense in L^p. Let $g \in C_0^0(\mathbb{R})$ such that $0 \leq g \leq 1$, $g \equiv 1$ on $(-1,1)$ and supp $g \subseteq [-2,2]$. For example, we can take

$$g(x) = \begin{cases} 1, & |x| < 1 \\ 2-x, & 1 \leq x \leq 2 \\ x+2, & -2 \leq x \leq -1 \end{cases}$$

and $g(x) \equiv 0$ outside $(-2,2)$. For each $m \in \mathbb{N}$, define $h_m : \mathbb{R}^n \to R$ by $h_m(x) = g\left(\dfrac{|x|}{m}\right)$. Then $0 \leq h_m \leq 1$, $h_m(x) \equiv 1$ when $|x| \leq m$, and $h_m(x) \equiv 0$ for $|x| \geq 2m$.

Let $h \in C(\mathbb{R}^n) \cap L^p$. Then $h\, h_m \longrightarrow h$ pointwise, $|h(x)\, h_m(x)| \leq |h(x)|$. Therefore, by the dominated convergence theorem, $h\, h_m \longrightarrow h$ in L^p. Therefore, $C_0^0(\mathbb{R}^n)$ is dense (in L^p norm) in $C(\mathbb{R}^n) \cap L^p$. Since $C(\mathbb{R}^n) \cap L^p$ is dense in L^p, it follows that $C_0^0(\mathbb{R}^n)$ is dense in L^p. Consider now $f \in L^p$. Let $\epsilon > 0$. Choose $\psi \in C_0^0(\mathbb{R}^n)$ such that $\|f - \psi\|_p < \epsilon$. Now since ψ is uniformly continuous on $\mathbb{R}^n$ and has compact support, there is $\delta > 0$ such that $|x| < \delta \Rightarrow \|\psi^x - \psi\|_p < \epsilon$. Hence when $|x| < \delta$:

$$\begin{aligned}\|f^x - f\|_p &\leq \|f^x - \psi^x\|_p + \|\psi^x - \psi\|_p + \|\psi - f\|_p \\ &= 2\|f - \psi\|_p + \|\psi^x - \psi\|_p \\ &< 3\epsilon\end{aligned}$$

Theorem 2.7. *Let* $g \in L^1(\mathbb{R}^n)$ *with* $\int g(x)\, dx = 1$. *For* $\epsilon > 0$, *set* $g_\epsilon(x) = \frac{1}{\epsilon^n} g\left(\frac{x}{\epsilon}\right)$. *Then*

a) if $f \in L^p$, $1 \leq p < \infty$, *then* $f * g_\epsilon \longrightarrow f$ *in* L^p *as* $\epsilon \to 0^+$;

b) if $f \in C(\mathbb{R}^n)$ *and is bounded, then* $f * g_\epsilon \longrightarrow f$ *uniformly on compact sets as* $\epsilon \to 0^+$.

Proof of (a). By changing variables $z = \frac{y}{\epsilon}$, we can write

$$f * g_\epsilon(x) = \frac{1}{\epsilon^n} \int_{\mathbb{R}^n} f(x-y)\, g\left(\frac{y}{\epsilon}\right)\, dy = \int_{\mathbb{R}^n} f(x - \epsilon z) g(z)\, dz$$

Recalling also that $\displaystyle\int g(z)\, dz = 1$, we get:

$$f * g_\epsilon(x) - f(x) = \int_{\mathbb{R}^n} (f(x - \epsilon z) - f(x))\, g(z)\, dz$$

We apply the integral version of Minkowski's inequality to the integral on the right which leads to:

$$\begin{aligned}\|f * g_\epsilon - f\|_{L^p} &\leq \int_{\mathbb{R}^n} \left(\int_{\mathbb{R}^n} |f^{\epsilon z}(x) - f(x)|^p\, dx \right)^{\frac{1}{p}} |g(z)|\, dz \\ &= \int_{\mathbb{R}^n} \|f^{\epsilon z} - f\|_p |g(z)|\, dz.\end{aligned}$$

By Lemma 2.6, pointwise $\|f^{\epsilon z} - f\|_p \, |g(z)| \longrightarrow 0$, and $\|f^{\epsilon z} - f\|_p \, |g(z)| \leq 2\|f\|_p \, |g(z)|$. Therefore, by the dominated convergence theorem, $f * g_\epsilon \to f$ in L^p. The proof of (b) is left as an exercise.

Corollary 2.8. *C_0^∞ is dense in L^p $(1 \leq p < \infty)$.*

Proof. Suppose $f \in L^p$. For $m \in N$, let

$$f_m(x) = \begin{cases} f(x), & |x| \leq m \\ 0, & |x| > m. \end{cases}$$

Then $f_m \longrightarrow f$ in L^p and f_m is of compact support. Let $g \in C_0^\infty(\mathbb{R}^n)$ such that $\displaystyle\int g(x)\,dx = 1$, and $g_\epsilon(x) = \dfrac{1}{\epsilon^n} g\left(\dfrac{x}{\epsilon}\right)$. Now $f_m * g_\epsilon \in C_0^\infty(\mathbb{R}^n)$ by Theorem 2.5 and Proposition 2.4. Moreover, by Theorem 2.7, $f_m * g_\epsilon \longrightarrow f_m$ in L^p as $\epsilon \to 0^+$. This proves the corollary.

The family g_ϵ as in Theorem 2.7 is called an **approximate identity** family since convolution with this family tends to the identity operator as ϵ goes to 0. Theorem 2.7 provides the basic result on approximation of L^p or continuous functions by smooth functions of compact support. The key ingredient in these approximations is convolution. We will next try to gain some insight into this operation by taking special cases of the function g in the theorem. Theorem 2.5 and its proof showed that $f * g$ is smoother (ie. belongs to a higher differentiability class) than f and g since when we take a derivative, the derivative may fall on f or g, say if both are differentiable. If we approximate $f * g(x)$ by a Riemann sum

$$\sum f(x - y_j) g(y_j) \Delta y$$

we see that $f * g(x)$ is a superposition of translates of f with the weights $g(y_j)\Delta y$, and so it is reasonable to expect the convolution to inherit good properties of f such as differentiability provided that the size of g is somehow controlled. By symmetry, the properties of g will also be inherited. Even when neither f nor g is continuous, $f * g$ will be continuous in many important situations. This happens for example if $f \in L^1$ and g is bounded. The point is that although f and g may not be continuous, $f * g(x)$ is an integral and so is less sensitive to a small change in x. For a simple example illustraing this, take $f(x) =$ the characteristic function of $[0, 1]$. By taking cases depending on whether x is in $(0, 1)$, $(1, 2)$ or outside $(0, 2)$, one easily gets

$$f * f(x) = \begin{cases} x, & 0 < x < 1 \\ 2 - x, & 1 < x < 2 \\ 0, & \text{otherwise} \end{cases}$$

Thus, although f has discontinuities, $f * f$ has none. The reader is encouraged to compute $f * f * f$ and see how we get a further improvement. Consider next a function g as in Theorem 2.7. To get an idea of why for small ϵ the function $f * g_\epsilon$ approximates f, let us look at an easy example in $\mathbb{R}$. Let

$$g(x) = \begin{cases} \frac{1}{2}, & |x| \le 1 \\ 0, & |x| > 1. \end{cases}$$

Note that g satisfies the hypotheses of the theorem. For $\epsilon > 0$, we have

$$g_\epsilon(x) = \frac{1}{\epsilon} g\left(\frac{x}{\epsilon}\right) = \begin{cases} \frac{1}{2\epsilon}, & |x| \le \epsilon \\ 0, & |x| > \epsilon. \end{cases}$$

It is easy to see that

$$f * g_\epsilon(x) = \frac{1}{2\epsilon} \int_{x-\epsilon}^{x+\epsilon} f(y)\, dy$$

which is just the average value of f on the interval $[x - \epsilon, x + \epsilon]$. If f is continuous at x, this average value approximates $f(x)$ when ϵ is small and in fact converges to $f(x)$ as $\epsilon \to 0$. In general, given g as in Theorem 2.7, and $f \in C(\mathbb{R}^n)$ and bounded, we have

$$f * g_\epsilon(x) = \int f(x - \epsilon y) g(y)\, dy$$

Since $g \in L^1$, a substantial part of its integral lives on bounded sets. More precisely, for each positive integer k, there is $M_k > 0$ such that

$$\int_{|y|>M_k} |g(y)|\, dy < \frac{1}{k}$$

Therefore,

$$\left| f * g_\epsilon(x) - \int_{|y|<M_k} f(x-\epsilon y) g(y)\, dy \right| < \frac{1}{k} ||f||_{L^\infty}$$

As ϵ tends to 0, the integral

$$\int_{|y|<M_k} f(x-\epsilon y) g(y)\, dy$$

converges to

$$f(x) \int_{|y|<M_k} g(y)\, dy$$

and the latter converges to $f(x)$ as k tends to ∞.

In an article on Fourier Analysis, we will see that the Fourier transform converts convolution to the much simpler operation of multiplication. This will shed more light on the interpretation and properties of convolution. Convolutions arise widely in science and engineering with the interpretation depending on the context. In optics, interference and image formation can be formulated using convolution. In electrical circuits, the circuit response can be expressed using the concept of convolution. In probability, if X and Y denote two independent random variables with respective probability density functions f and g, then the convolution $f * g$ is the probability density function of the random variable $X + Y$.

We will next show how to approximate the characteristic function χ_K for K compact by smooth functions of compact support.

Theorem 2.9. *Let K be a compact set in $\mathbb{R}^n$ and V an open neighborhood of K. Then, there exists $\varphi \in C_0^\infty(\mathbb{R}^n)$, $0 \le \varphi \le 1$, satisfying $\varphi \equiv 1$ on K and supp$(\varphi) \subseteq V$. Moreover, φ can be chosen so that for any multi-index α,*

$$\|\partial_x^\alpha \varphi\|_{L^\infty} \le \frac{C_\alpha}{dist(K, \partial V)^{|\alpha|}},$$

where the constants C_α are independent of K and V.

Proof. We may assume $V \neq \mathbb{R}^n$. Let $d = dist(K, \partial V)$. For $0 < \epsilon < d$, define $K_\epsilon = \{x : d(x, K) < \epsilon\}$. Then $K_\epsilon = \bigcup_{x \in K} B_\epsilon(x)$ and $K \subseteq K_\epsilon \subseteq V$. Fix ϵ such that $0 < 2\epsilon < d$ and let $h^\epsilon(x) =$ the characteristic function of K_ϵ. Let $\varphi \in C_0^\infty(\mathbb{R}^n)$ supp $\varphi \subseteq B_1(0)$ and $\int \varphi(x)\, dx = 1$. Set $\varphi_\epsilon(x) = \frac{1}{\epsilon^n} \varphi\left(\frac{x}{\epsilon}\right)$. Consider now the function $\varphi_\epsilon * h^\epsilon$ for $0 < 2\epsilon < d$:

$$
\begin{aligned}
\varphi_\epsilon * h^\epsilon(x) &= \int_{\mathbb{R}^n} \varphi_\epsilon(y)\, h^\epsilon(x-y)\, dy \\
&= \int_{|y|<\epsilon} \varphi_\epsilon(y)\, h^\epsilon(x-y)\, dy
\end{aligned}
$$

since supp $\varphi_\epsilon \subseteq B_\epsilon(0)$. If $x \in K$ and $|y| < \epsilon$, then $h^\epsilon(x-y) = 1$, while if $x \notin K_{2\epsilon}$ and $|y| < \epsilon$, $x - y \notin K_\epsilon$. Hence $\varphi_\epsilon * h^\epsilon \equiv 1$ on K and $\varphi_\epsilon * h^\epsilon \equiv 0$ outside $K_{2\epsilon}$. Since $\varphi_\epsilon * h^\epsilon \in C_0^\infty(\mathbb{R}^n)$, the first part of the theorem is proved. Next note that

$$
\begin{aligned}
\partial_x^\alpha \left(\varphi_\epsilon * h^\epsilon\right)(x) &= \left(\partial_x^\alpha \varphi_\epsilon\right) * h^\epsilon(x) \\
&= \int \partial_y^\alpha \left(\varphi_\epsilon(y)\right)\, h^\epsilon(x-y)\, dy \\
&= \frac{1}{\epsilon^{n+|\alpha|}} \int \left(\partial_y^\alpha \varphi\right)\left(\frac{y}{\epsilon}\right) h^\epsilon(x-y)\, dy \\
&= \frac{1}{\epsilon^{|\alpha|}} \int \left(\partial_y^\alpha \varphi\right)(z) h^\epsilon(x - \epsilon z)\, dz
\end{aligned}
$$

Hence if we choose $\epsilon = \frac{d}{4}$,

$$
\begin{aligned}
\left|\partial_x^\alpha \left(\varphi_\epsilon * h^\epsilon\right)(x)\right| &\leq \frac{1}{\epsilon^{|\alpha|}} . \|\partial_y^\alpha \varphi\|_{L^1} \\
&= \frac{C_\alpha}{dist(K, \partial V)^{|\alpha|}},
\end{aligned}
$$

with $C_\alpha = 4^{|\alpha|} \left\|\partial_y^\alpha \varphi\right\|_{L^1}$, and hence independent of K and V.

Exercise 2.3. *Recall the Weierstrass approximation theorem : if f is continuous on $\mathbb{R}^n$ and K is compact, there is a sequence of polynomials P_k that converges uniformly to f on K. Prove this theorem by applying Theorem 2.7 to the function $g(x) = c_n e^{-|x|^2}$ (c_n a constant chosen appropriately). If $f \in C^k$, prove that the convergence is valid in the C^k topology.*

Exercise 2.4. *Consider the estimate on the partial derivatives of φ given by Theorem 2.9. Show that this estimate is sharp, ie. for any α, show that there is a point p such that $|\partial_x^\alpha \varphi(p)|$ is comparable to $\frac{1}{dist(K,\partial V)^{|\alpha|}}$. Hint: since $\varphi \equiv 1$ on K and 0 on the boundary of V, somewhere any partial derivative $\partial_i \varphi$ has to be comparable to $\frac{1}{dist(K,\partial V)}$.*

3. The Schwartz Space and Tempered Distributions

In this section we will introduce the reader to an important class of distributions called **tempered distributions**. The aim is to give some ideas of what distributions are all about. References for a more thorough treatment are presented at the end of this article.

The theory of distributions in the form widely used today was created by Laurent Schwartz at the end of the 1940's. Over the past 50 years, this theory has played a central role in the analysis of linear partial differential equations. In a forthcoming article on Fourier Analysis, we will indicate how distributions can be used to study the three archetypes of linear partial differential equations: Laplace's equation, the heat equation and the wave equation.

Definition 3.1. *The Schwartz Space $\mathcal{S} = \mathcal{S}(\mathbb{R}^n)$ of rapidly decreasing functions is the set*

$$\mathcal{S} := \{f \in C^\infty(\mathbb{R}^n) : \textit{for any multi-indices } \alpha \textit{ and } \beta, \ \sup_{x \in \mathbb{R}^n} |x^\alpha \partial^\beta f(x)| < \infty\}.$$

In other words, if $u \in \mathcal{S}$, then u and any derivative $\partial^\alpha u$ die out faster than the reciprocal of any polynomial at ∞. That is, $u \in \mathcal{S}$ iff for any β and $M > 0$ there is a constant $C = C(\beta, M)$ such that

$$|\partial^\beta u(x)| \leq \frac{C}{(1+|x|)^M}.$$

In particular, $\lim_{x\to\infty} \partial^\beta u(x) = 0$ for any β.

Example 3.1.

$$C_0^\infty(\mathbb{R}^n) \subset \mathcal{S}.$$

Example 3.2. *For any $a \in \mathbb{C}$ with $Re\, a > 0$,*

$$e^{-a|x|^2} \in \mathcal{S}.$$

Exercise 3.1. *Show that $\mathcal{S}$ is an algebra and that if $f \in \mathcal{S}$, then $\partial^\beta f \in \mathcal{S}$.*

Exercise 3.2. *(a). Show that whenever $p(x)$ is a polynomial and $f \in \mathcal{S}$, then*

$$p(x)f(x) \in \mathcal{S}$$

(b). Show that if $f \in \mathcal{S}$, then f and hence $\partial^\beta f$ is uniformly continuous.

For multi-indices α and β, define

$$\|u\|_{\alpha,\beta} := \sup_{x\in\mathbb{R}^n} |x^\alpha \partial^\beta u(x)|$$

on $\mathcal{S}$. It is easy to see that $\{\|.\|_{\alpha,\beta}\}$ forms a countable collection of seminorms on $\mathcal{S}$ which gives $\mathcal{S}$ a natural topology. That is, $\mathcal{S}$ is topologized by the requirement that a sequence $\{f_m\}$ converges to f in $\mathcal{S}$ iff for each α, β,

$$\|f_m - f\|_{\alpha,\beta} \to 0$$

as $m \to \infty$.

Proposition 3.1. *The space $\mathcal{S}$ equipped with the seminorms $\{\|.\|_{\alpha,\beta}\}$ is a Fréchet space, i.e. it is a complete, metrizable, locally convex space.*

Exercise 3.3. *Prove Proposition 3.1.*

Often, it will be convenient to use the equivalent directed family of seminorms

$$\rho_{k,m}(g) = \|g\|_{k,m} = \sum_{|\alpha|\le k,\, |\beta|\le m} \|g\|_{\alpha,\beta}, \quad \text{for } k \text{ and } m \text{ in } \mathbb{N}.$$

Since C_0^∞ is dense in L^p (Corollary 2.8) for $1 \le p < \infty$, it is clear that $\mathcal{S}$ is dense in L^p. The space $\mathcal{S}$ has several properties not shared by C_0^∞. One of these turns out to be the fact that the Fourier transform is an isomorphism from $\mathcal{S}$ onto $\mathcal{S}$ while the Fourier transform of $f \in C_0^\infty$ is never in C_0^∞ except in the trivial case when $f \equiv 0$.

Exercise 3.4. *Show that C_0^∞ is dense in $\mathcal{S}$.*

Hint: Let $\phi \in C_0^\infty(\mathbb{R}^n)$, supp$(\phi) \subset B_2(0)$, $\phi \equiv 1$ on $B_1(0)$. For $m \in \mathbb{N}$, let $\phi_m(x) = \phi(x/m)$. If $\psi \in \mathcal{S}$, show that $\phi_m\psi \to \psi$ in $\mathcal{S}$.

Definition 3.2. *The dual space of $\mathcal{S}$, denoted by $\mathcal{S}'$ is called the space of* **tempered distributions***.*

If $u \in \mathcal{S}'$, by definition, $u : \mathcal{S} \to \mathbb{C}$ is linear and whenever $f_j \to f$ in $\mathcal{S}$, then $u(f_j) \to u(f)$. Recall that if X is a normed vector space, then a linear map $T : X \to \mathbb{C}$ is continuous iff there is $C > 0$ such that

$$\|Tx\| \le C\|x\|, \ \forall x \in X.$$

Likewise, if $u : \mathcal{S} \to \mathbb{C}$ is linear, $u \in \mathcal{S}'$ iff there are m and $C > 0$ such that

$$|u(\phi)| \le C\rho_{m,m}(\phi), \quad \forall \phi \in \mathcal{S}.$$

It is easy to see that if the linear map u satisfies $|u(\phi)| \le C\rho_{m,m}(\phi), \ \ \forall \phi \in \mathcal{S}$, then $u \in \mathcal{S}'$. Conversely, suppose $u \in \mathcal{S}'$. If we cannot find C, m such that

$$|u(\phi)| \le C\rho_{m,m}(\phi), \ \forall \phi \in \mathcal{S},$$

then for each $m \in \mathbb{N}$, there is $\phi_m \in \mathcal{S}$ such that

$$|u(\phi_m)| > m^2 \rho_{m,m}(\phi_m).$$

Let

$$f_m = \frac{\phi_m}{m\rho_{m,m}(\phi_m)}.$$

Note that $f_m \to 0$ in $\mathcal{S}$. Hence since $u \in \mathcal{S}',\ u(f_m) \to u(0) = 0$. But

$$|u(f_m)| = \frac{|u(\phi_m)|}{m\rho_{m,m}(\phi_m)} > m,$$

which is a contradiction. We have thus proved

Theorem 3.2. *A linear map $u : \mathcal{S} \to \mathbb{C}$ is a tempered distribution iff there exist C, m such that*

$$|u(\phi)| \le C\rho_{m,m}(\phi), \ \ \forall \phi \in \mathcal{S}.$$

This theorem enables us to exhibit several examples of tempered distributions. The space $\mathcal{S}'$ is topologized by using the following pointwise convergence.

Definition 3.3. *We say that a sequence $\{u_k\}$ in $\mathcal{S}'$ converges to u in $\mathcal{S}'$ if for each $\phi \in \mathcal{S}$,*

$$u_k(\phi) \to u(\phi).$$

Example 3.3. *Let $g \in \mathcal{S}$. Then g defines an element $u_g \in \mathcal{S}'$ by means of the action*

$$u_g(\phi) := \int g(x)\phi(x)\, dx\,, \quad \phi \in \mathcal{S}.$$

u_g is linear and

$$|u_g(\phi)| \le \|g\|_{L^1} \sup_{x \in \mathbb{R}^n} |\phi(x)| = \|g\|_{L^1}\rho_{0,0}(\phi).$$

Thus $u_g \in \mathcal{S}'$. We will often denote $u_g(\phi)$ by $\langle g, \phi \rangle$, and when $u \in \mathcal{S}'$ and $\psi \in \mathcal{S}$ we will also use the notation $\langle u, \psi \rangle$ for $u(\psi)$.

Observe that the mapping $\mathcal{S} \ni g \longmapsto u_g \in \mathcal{S}'$ is one-to-one and continuous. To see the one-to-oneness, suppose $g \in \mathcal{S}$ and

$$\int g(x)\phi(x)\,dx = 0, \quad \forall \phi \in \mathcal{S}.$$

Assume $g(p) \neq 0$. Without loss of generality we may assume $g(p) > 0$, in which case by continuity there exist $\epsilon > 0$ and $r > 0$ such that $g(x) > \epsilon$ for every $x \in B_r(p)$. Let $\phi \in C_0^\infty(\mathbb{R}^n)$ such that $\operatorname{supp}\phi \subseteq B_r(p)$, $\phi \geq 0$ and $\phi(p) > 0$. Then we are led to

$$0 = \int g(x)\phi(x)\,dx = \int_{B_r(p)} g(x)\phi(x)\,dx > 0,$$

which is a contradiction.

We thus have a continuous embedding of $\mathcal{S}$ into $\mathcal{S}'$ which allows us to view $\mathcal{S}$ as a subset of $\mathcal{S}'$. In fact, it will turn out that $\mathcal{S}$ is dense in $\mathcal{S}'$ (see Example 3.10).

Example 3.4. *For any $1 \leq p \leq \infty$, $L^p \subseteq \mathcal{S}'$. This shouldn't come as a surprise since the embedding $\mathcal{S} \hookrightarrow L^q$ is continuous, and so taking duals we get a continuous embedding $L^p = (L^q)' \hookrightarrow \mathcal{S}'$ $(1/p + 1/q = 1)$. Explicitly, suppose $f \in L^p$. Define*

$$\langle f, \phi \rangle = \int f(x)\phi(x)\,dx, \quad \text{for } \phi \in \mathcal{S}.$$

Then

$$\begin{aligned} |\langle f, \phi \rangle| &= \int \frac{|f(x)|}{(1+|x|)^M}(1+|x|)^M |\phi(x)|\,dx, \\ &\leq C\|f\|_p \|(1+|x|)^{-M}\|_q \rho_{M,0}(\phi), \end{aligned}$$

where q is the Hölder conjugate of p, and $M = (n+1)/q$.

Example 3.5. *Let $a < n$, and define $g(x) = 1/|x|^a$. Note that $g \in L^1_{\text{loc}}(\mathbb{R}^n)$. Then $g \in \mathcal{S}'$. The proof is left as an exercise.*

Example 3.6. *The Dirac delta distribution δ_p: For $p \in \mathbb{R}^n$, define $\delta_p : \mathcal{S} \to \mathbb{C}$ by $\delta_p(\phi) = \phi(p)$. Then $\delta_p \in \mathcal{S}'$. Note that if $\operatorname{supp}\phi \subseteq \mathbb{R}^n \setminus \{p\}$, then $\delta_p(\phi) = 0$ and so away from p, δ_p agrees with the zero function. However, there is no $g \in L^1$ such that $\langle g, \phi\rangle = \langle \delta_p, \phi\rangle \ \ \forall \phi \in \mathcal{S}$. Thus δ_p is an example of a distribution that doesn't come from a function. Often δ will denote δ_0.*

Example 3.7. *The Principal value of $1/x$, $PV(1/x)$:*

For $\phi \in \mathcal{S}(\mathbb{R})$, define

$$\langle PV(\frac{1}{x}), \phi\rangle = \lim_{\epsilon \to 0^+} \int_{|x|>\epsilon} \frac{\phi(x)}{x}\, dx.$$

To see why the limit exists, note that

$$\int_{|x|>\epsilon} \frac{\phi(x)}{x}\, dx = \int_{|x|\geq 1} \frac{\phi(x)}{x}\, dx + \int_{\epsilon}^{1} \frac{\phi(x) - \phi(-x)}{x}\, dx.$$

The first integral on the right is dominated by

$$\int_{|x|\geq 1} \frac{|\phi(x)|}{|x|}\, dx \leq \int_{|x|\geq 1} \frac{|x\phi(x)|}{x^2}\, dx \leq \left(\int_{|x|\geq 1} \frac{1}{x^2}\, dx\right) \sup_{x\in\mathbb{R}} |x\phi(x)|.$$

$$\frac{\phi(x) - \phi(-x)}{x}$$

is continuous at 0, and so

$$\lim_{\epsilon \downarrow 0} \int_{\epsilon}^{1} \frac{\phi(x) - \phi(-x)}{x}\, dx = \int_{0}^{1} \frac{\phi(x) - \phi(-x)}{x}\, dx.$$

We have

$$\left|\int_{\epsilon}^{1} \frac{\phi(x) - \phi(-x)}{x}\, dx\right| = \left|\int_{\epsilon}^{1} \frac{1}{x} \int_{-x}^{x} \phi'(t)\, dt\, dx\right| \leq 2 \sup_{t\in\mathbb{R}} |\phi'(t)|.$$

This shows that $PV(1/x) \in \mathcal{S}'(\mathbb{R})$. Observe that for $\phi \in \mathcal{S}(\mathbb{R})$ supported in $\mathbb{R}\setminus\{0\}$,

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

That is away from 0, $PV(1/x)$ *agrees with* $1/x$.

In general, it may not make sense to talk about the pointwise value of a distribution. For example, we cannot assign a value to the Dirac δ at 0. However, it is meaningful to talk about whether two distributions are equal on a given open set. That is, we will say a distribution u vanishes on an open set Ω if $\forall \phi \in \mathcal{S}$, $\mathrm{supp}(\phi) \subseteq \Omega$, we have $\langle u, \phi \rangle = 0$. This is motivated by the fact that when $u \in \mathcal{S}$, then $u(x) = 0$ for every $x \in \Omega$ iff $\langle u, \phi \rangle = 0$, $\forall \phi \in \mathcal{S}$ with $\mathrm{supp}(\phi) \subseteq \Omega$. We can thus define the support of $u \in \mathcal{S}'$, denoted $\mathrm{supp} u$ to be the complement of the largest open set where u vanishes. This then makes precise the assertion that $\delta = \delta_0$ equals 0 away from the origin. That is, $\mathrm{supp}(\delta) = \{0\}$.

Exercise 3.5. *Suppose* $\{u_k\}$ *is a convergent sequence in* $\mathcal{S}'$. *Then show that there are* $C > 0$ *and* $m \in \mathbb{N}$ *such that for all* k

$$|u_k(\phi)| \leq C\rho_m(\phi), \ \ \forall \phi \in \mathcal{S}.$$

Hint: Review the proof of the analogue for a sequence of pointwise convergent bounded linear maps on a Banach space.

Exercise 3.6. *Let*

$$\phi \in L^1(\mathbb{R}^n), \quad \int \phi(x)\, dx = 1, \quad \text{and for } \epsilon > 0, \quad \phi_\epsilon(x) = \frac{1}{\epsilon^n}\phi\left(\frac{x}{\epsilon}\right).$$

Note that each $\phi_\epsilon \in L^1 \subseteq \mathcal{S}'$. *Prove that* $\phi_\epsilon \to \delta_0$ *as* $\epsilon \to 0^+$. *More generally, show that* $\phi_\epsilon(. - x_0) \to \delta_{x_0}$ *as* $\epsilon \to 0^+$.

Exercise 3.7. *Use Exercise 3.6 to show that*

$$\lim_{\epsilon \downarrow 0} \frac{\epsilon}{(x - x_0)^2 + \epsilon^2} = \pi \delta_{x_0}$$

in $\mathcal{S}'(\mathbb{R})$.

In a future article on the Fourier transform, we will see that the function

$$P(x, y) = \frac{y}{\pi(x^2 + y^2)}$$

that appears in Exercise 3.7 arises in the study of boundary value problems for the Laplace equation in the upper half plane $\{(x, y) : y > 0\}$. $P(x, y)$ is called the Poisson Kernel for the upper half plane.

Exercise 3.8. *Let*

$$K(x,t) = \frac{1}{\sqrt{4\pi t}} e^{\frac{-x^2}{4t}}, \quad \text{for } t > 0, \text{ and } x \in \mathbb{R}.$$

This function is called the Heat Kernel and it arises in the study of the heat equation as we will see in the future. Note that for each t, $x \mapsto K(x,t)$ is in $L^1(\mathbb{R})$. Prove that $\{K(.,t)\}_{t>0}$ converges to δ_0 in $\mathcal{S}'(\mathbb{R})$ as $t \to 0^+$.

Operations on Tempered Distributions.

Many operations on $\mathcal{S}$ such as convolution, differentiation and Fourier transformation can also be extended to $\mathcal{S}'$. The basic philosophy behind such extensions is as follows. Suppose we have a continuous linear map $T_1 : \mathcal{S} \to \mathcal{S}$ and an associated linear map $T_2 : \mathcal{S} \to \mathcal{S}$ such that

$$\langle T_1 f, g \rangle = \langle f, T_2 g \rangle, \quad \forall f, g \in \mathcal{S}.$$

Then we can extend T_1 to a continuous linear map $T_1 : \mathcal{S}' \to \mathcal{S}'$ by defining

$$\langle T_1 u, \phi \rangle = \langle u, T_2 \phi \rangle, \quad \forall u \in \mathcal{S}', \ \forall \phi \in \mathcal{S}.$$

So defined, $T_1 u \in \mathcal{S}'$. A very useful application of this philosophy leads to the concept of differentiation of distributions. Let $T_1 : \mathcal{S} \to \mathcal{S}$ be defined by

$$T_1(\phi) = \frac{\partial \phi}{\partial x_1}.$$

Integration by parts shows that

$$\langle T_1(f), g \rangle = -\langle f, T_1(g) \rangle, \quad \forall f, g \in \mathcal{S}.$$

That is

$$\langle T_1(f), g\rangle = \langle f, T_2(g)\rangle, \quad \text{where } T_2(\phi) = -\frac{\partial \phi}{\partial x_1}.$$

Therefore we can extend the concept of differentiation to $\mathcal{S}'$ by defining

$$\frac{\partial}{\partial x_j} : \mathcal{S}' \to \mathcal{S}' \quad \text{as}$$

$$\langle \frac{\partial u}{\partial x_j}, \phi\rangle = -\langle u, \frac{\partial \phi}{\partial x_j}\rangle, \quad \forall u \in \mathcal{S}', \ \forall \phi \in \mathcal{S}.$$

Likewise, if α is a multi-index and $u \in \mathcal{S}'$, $\partial^\alpha u$ will denote the element of $\mathcal{S}'$ defined by

$$\langle \partial^\alpha u, \phi\rangle = (-1)^{|\alpha|}\langle u, \partial^\alpha \phi\rangle, \quad \forall \phi \in \mathcal{S}.$$

Definition 3.4. *If $u \in \mathcal{S}'$ and α is a multi-index, $\partial^\alpha u$ is called a derivative of u in the distribution sense. It is also refered to as a weak derivative of u.*

Exercise 3.9. *Suppose $u_k \to u$ in $\mathcal{S}'$. For any multi-index α, show that*

$$\partial^\alpha u_k \to \partial^\alpha u$$

in $\mathcal{S}'$.

Example 3.8. *Define*

$$H(x) = \begin{cases} 1, & x > 0 \\ 0, & x < 0 \end{cases}.$$

The function H is called the Heaviside function. H is a bounded function and so $H \in \mathcal{S}'$. We wish to compute the distribution derivative dH/dx. Clearly $dH/dx = 0$ away from 0. So dH/dx as an element of $\mathcal{S}'$ "lives" at 0. That is, when applied to $\phi \in \mathcal{S}$, we expect dH/dx to detect the values of ϕ only near 0.

Indeed, for $\phi \in \mathcal{S}$

$$\begin{aligned}
\left\langle \frac{dH}{dx}, \phi \right\rangle &= -\langle H, \frac{d\phi}{dx} \rangle, \qquad \left(\text{by definition of } \frac{dH}{dx}\right) \\
&= -\int_{-\infty}^{\infty} H(x)\phi'(x)\,dx \\
&= -\int_{0}^{\infty} \phi'(x)\,dx \\
&= \phi(0) \\
&= \langle \delta_0, \phi \rangle.
\end{aligned}$$

Hence

$$\frac{dH}{dx} = \delta_0.$$

Thus although δ_0 is not a function, it is the weak derivative of a function. Notice also that H is not differentiable everywhere in the classical sense. In fact, it is not even continuous. However, it has a weak derivative which is a linear functional on $\mathcal{S}$. Moreover, where H has a classical derivative, that is, on $\mathbb{R} \setminus \{0\}$, the weak derivative dH/dx agrees with the classical derivative in the sense that for $\phi \in \mathcal{S}$, $\operatorname{supp} \phi \subseteq \mathbb{R} \setminus \{0\}$,

$$\left\langle \frac{dH}{dx}, \phi \right\rangle = 0.$$

Exercise 3.10. *Let*

$$G(x) = \begin{cases} x, & x > 0 \\ 0, & x < 0 \end{cases}$$

Prove that

$$\frac{dG}{dx} = H$$

in the sense of distributions and that

$$\frac{d^2G}{dx^2} = \delta_0.$$

We next consider convolution. If $f, g \in \mathcal{S}$, recall that

$$f * g(x) = \int f(x-y)g(y)\,dy.$$

By Exercise 3.2, f is uniformly continuous and so $f * g$ is uniformly continuous on $\mathbb{R}^n$. In fact, $f * g \in \mathcal{S}$. To show that $f * g$ is in C^∞, we can use the argument of Theorem 2.5 and conclude that

$$\partial^\alpha (f * g)(x) = ((\partial^\alpha f) * g)(x) = (f * \partial^\alpha g)(x).$$

Next observe that

$$\begin{aligned}
|x^\alpha \partial^\beta (f * g)(x)| &= \left| x^\alpha \int \partial^\beta f(x-y) g(y)\, dy \right| \\
&\le |x|^{|\alpha|} \int |\partial^\beta f(x-y)||g(y)|\, dy \\
&\le \int (|x-y| + |y|)^{|\alpha|} |\partial^\beta f(x-y)||g(y)|\, dy \\
&\le 2^{|\alpha|} \left(\int |x-y|^{|\alpha|} |\partial^\beta f(x-y)||g(y)|\, dy \right. \\
&\left. + \int |y|^{|\alpha|} |g(y)||\partial^\beta f(x-y)|\, dy \right) \\
&\le 2^{|\alpha|} \left(\|g\|_{L^1} \sup_z |z|^{|\alpha|} |\partial^\beta f(z)| + \|\partial^\beta f\|_{L^1} \sup_z |z|^{|\alpha|} |g(z)| \right) < \infty.
\end{aligned}$$

We have shown that $f * g \in \mathcal{S}$ whenever f and g are in $\mathcal{S}$. To define the convolution $u * f$ for $u \in \mathcal{S}'$ and $f \in \mathcal{S}$, we need to first express $f * g$ for f and g in $\mathcal{S}$ in terms of the action of f when viewed as a distribution.

$$f * g(x) = \int f(y) g(x-y)\, dy = \langle f, g_x \rangle,$$

where $g_x(y) = g(x-y)$. That is, when f and g are in $\mathcal{S}$, $f * g(x)$ equals the value of the tempered distribution f acting on the Schwartz function g_x. This suggests:

Definition 3.5. *For $u \in \mathcal{S}'$ and $f \in \mathcal{S}$, we define*

$$u * f(x) = \langle u, f_x \rangle.$$

Exercise 3.11. *Show that if* $u \in \mathcal{S}'$ *and* $f \in \mathcal{S}$*, then* $u * f \in \mathcal{S}$*.*

We can now define the convolution $u * v$ of $u, v \in \mathcal{S}'$. Again note that when f and g are in $\mathcal{S}$, then $f * g$ viewed as an element of $\mathcal{S}'$ has the following effect:

$$\langle f * g, h \rangle = \langle f, \tilde{g} * h \rangle, \quad \forall h \in \mathcal{S},$$

where $\tilde{g}(x) = g(-x)$. Note that $\tilde{g} \in \mathcal{S}$. It therefore seems reasonable to define $u * v$ for $u, v \in \mathcal{S}'$ by setting

$$\langle u * v, h \rangle = \langle u, \tilde{v} * h \rangle.$$

How should $\tilde{v}$ be defined for $v \in \mathcal{S}'$? Well, if $g \in \mathcal{S}$ then

$$\langle \tilde{g}, \phi \rangle = \langle g, \tilde{\phi} \rangle, \quad \forall \phi \in \mathcal{S},$$

and so we define

$$\langle \tilde{v}, \phi \rangle = \langle v, \tilde{\phi} \rangle, \quad \phi \in \mathcal{S}.$$

It is easily checked that $\tilde{v} \in \mathcal{S}'$. Moreover, for $h \in \mathcal{S}$, by Exercise 3.11, we have $\tilde{v} * h \in \mathcal{S}$ and so $\langle u, \tilde{v} * h \rangle$ makes sense.

Exercise 3.12. *If* $u, v \in \mathcal{S}'$*, show that* $u * v \in \mathcal{S}'$*.*

Exercise 3.13. *Let* $u_k \to u$ *in* $\mathcal{S}'$ *and* $w \in \mathcal{S}'$*. Show that* $u_k * w \to u * w$ *in* $\mathcal{S}'$*.*

Example 3.9. *Let* $\phi \in \mathcal{S}$*. Then* $\delta * \phi = \phi$*. Indeed,*

$$\delta * \phi(x) = \langle \delta, \phi_x \rangle = \phi(x).$$

Moreover, if $u \in \mathcal{S}'$*, then for any* $\phi \in \mathcal{S}$*,*

$$\langle u * \delta, \phi \rangle = \langle u, \tilde{\delta} * \phi \rangle = \langle u, \delta * \phi \rangle = \langle u, \phi \rangle,$$

and so $u * \delta = u$*.*

Example 3.10. *Let*

$$\phi \in C_0^\infty(\mathbb{R}^n), \quad \int \phi(x)\,dx = 1.$$

Let $\phi_\epsilon(x) = \frac{1}{\epsilon^n}\phi(x/\epsilon)$*. By Exercise 3.6,* $\phi_\epsilon \to \delta$ *in* $\mathcal{S}'$*. Therefore, for any* $u \in \mathcal{S}'$*, by Exercise 3.13, we get*

$$u * \phi_\epsilon \to u * \delta = u$$

in $\mathcal{S}'$*. Since each* $u * \phi_\epsilon \in \mathcal{S}$ *we now see that* $\mathcal{S}$ *is dense in* $\mathcal{S}'$*. By Exercise 3.4,* C_0^∞ *is also dense in* $\mathcal{S}'$*. Once again, just as in Theorem 2.7, the approximation is achieved by using the approximate identity family* $\{\phi_\epsilon\}_{\epsilon>0}$*.*

The delta distribution was invented by the electrical engineer Oliver Heaviside who manipulated it as if it were a function – to the horror of the mathematicians of his time. In the 40's Laurent Schwartz's theory of distributions provided a rigorous setting for Heaviside's computations. Even today many books on partial differential equations contain notations of the form

$$\int_{-\infty}^{\infty} \delta(x - x_0) f(x)\,dx = f(x_0), \quad \text{and} \quad \int_{-\infty}^{\infty} \delta'(x - x_0) f(x)\,dx = -f'(x_0).$$

For engineers, δ is the "unit impulse function" which when integrated against a smooth function f,

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

recovers $f(0)$. If we limit our scope only to functions, we will face a dilemma since no function can have the effect of δ. Distribution theory resolves this problem by expanding our scope and viewing δ not as a function, but rather as a functional acting on a space of functions. The notation

$$\int_{-\infty}^{\infty} \delta(x - x_0) f(x)\,dx$$

then really symbolizes the action of δ_{x_0} on f. Likewise,

$$\int_{-\infty}^{\infty} \delta'(x - x_0) f(x)\,dx$$

denotes

$$\left\langle \frac{d}{dx}\delta_{x_0}, f \right\rangle.$$

Although δ cannot be represented by a function, Example 3.10 shows us how to approximate it by functions. For example, if $g_n(x) = n$ for $|x| \leq 1/n$, and $g_n(x) = 0$ when $|x| > 1/n$, then the sequence

$$g_n \to \delta$$

in the sense of distributions, that is

$$\langle g_n, \phi \rangle \to \langle \delta, \phi \rangle, \quad \forall \phi \in \mathcal{S}.$$

Characterization of tempered distributions

We now wish to describe the structure of any tempered distribution. If f is a measurable bounded function and $p(x)$ is a polynomial, then $p(x)f(x) \in \mathcal{S}'$ because any $\phi \in \mathcal{S}'$ decays more rapidly than the reciprocal of any polynomial making it possible to integrate $p(x)f(x)\phi(x)$. The decay also holds for any partial derivative of ϕ and so the weak derivatives $\partial^\beta(p(x)f(x)) \in \mathcal{S}'$. As the following theorem shows, any tempered distribution is a finite sum of such distributions. Thus although a distribution may not be a function, it is always the weak derivative of a function of polynomial growth.

Theorem 3.3. *Let $u \in \mathcal{S}'(\mathbb{R}^n)$. Then there is k and functions $f_{\alpha\beta} \in L^\infty$ for $|\alpha| \leq k, \quad |\beta| \leq k$ such that*

$$u = \sum_{|\alpha| \leq k} \sum_{|\beta| \leq k} \partial^\beta (x^\alpha f_{\alpha\beta})$$

That is, for any $\phi \in \mathcal{S}$,

$$\langle u, \phi \rangle = \sum_{\alpha,\beta} (-1)^{|\beta|} \int x^\alpha f_{\alpha,\beta}(x) \partial^\beta \phi(x)\, dx$$

Proof. We will give the proof when $n = 1$ and leave as an exercise the case $n > 1$. Notice first that if $\phi \in \mathcal{S}$, then

$$|\phi(x) - \phi(y)| = \left| \int_y^x D\phi(t)\, dt \right| \leq ||D\phi||_{L^1}$$

Letting $y \mapsto \infty$, we see that

$$||\phi||_{L^\infty} \leq ||D\phi||_{L^1}, \qquad \text{for any } \phi \in \mathcal{S}$$

Therefore, for any m and k, there is a constant C independent of $\phi \in \mathcal{S}$ such that

$$||x^m D^k \phi||_{L^\infty} \leq C \left(||x^{m-1} D^k \phi||_{L^1} + ||x^m D^{k+1} \phi||_{L^1} \right)$$

If $u \in \mathcal{S}'$, it follows that there exist M and C such that

$$|\langle u, \phi \rangle| \leq C \sum_{k \leq M, m \leq M} ||x^m D^k \phi||_{L^1}$$

Embed $\mathcal{S}$ into the direct sum

$$V = \bigoplus_{j=1}^{(M+1)^2} L^1_j \qquad (\text{ each } L^1_j = L^1(\mathbb{R}))$$

by means of

$$\phi \longmapsto \left\{ x^l D^j \phi \right\}_{l=0, j=0}^{M+1}$$

Observe that V is a finite direct sum of Banach spaces, and so it is equipped with a natural norm $||(g_1, \ldots, g_{(M+1)^2})|| = \sum ||g_j||_{L^1}$. Thus u defines a linear map on the image of $\mathcal{S}$ in V and $|u(h)| \leq C||h||$ for all h in this image. By the Hahn-Banach theorem, u extends to a continuous linear map on V which we still call u and then since the dual of V is $\bigoplus L^\infty_j$ ($L^\infty_j = L^\infty(\mathbb{R})$), the theorem follows when $n = 1$.

Remark. *We remark here that many of the results about distributions become easier to prove when we use Theorem 3.3.*

Exercise 3.14. *Suppose $u, v \in \mathcal{S}'$. Show that*

$$\partial^\alpha (u * v) = \partial^\alpha u * v = u * \partial^\alpha v$$

Applications of Distributions

We will now briefly indicate the application of distribution theory to differential equations. Future articles will provide the details.

Consider the problem of finding a solution u to the linear partial differential equation (pde)

$$(\bigstar) \qquad \sum_{|\alpha|\leq m} a_\alpha(x)\partial^\alpha u(x) = f(x)$$

where f and a_α are C^∞ functions. In the classical framework, one has to find a C^∞ function u such that ($\bigstar$) is satisfied at every point. The fact that solutions are searched for only in the space C^∞ is a serious constraint. However, instead of insisting that equation ($\bigstar$) holds pointwise, we could look for distribution solutions. In some cases, even the distribution solutions turn out to be smooth functions. The advantage of working with distributions is that they are easier to obtain than classical solutions. An illustration of this occurs in a first pde course (undergraduate level) where one uses the method of separation of variables to solve an initial-boundary value problem. This method yields C^∞ solutions $\{f_k\}$ for the pde and some of the side conditions. To solve the complete problem, one considers a series

$$\sum_{k=0}^{\infty} c_k f_k$$

that formally solves it. To prove that such a series actually represents a solution, one needs to confront issues of convergence for the series and the differentiated series. However, if we use distribution theory, the series can easily be shown to be a solution to the pde since it is the limit in $\mathcal{S}'$ of partial sums which are in fact classical solutions (see Exercise 3.9). Thus, distribution theory frees us from the painful convergence questions we face when we insist on solutions in the pointwise sense.

For a more concrete example, consider solving the equation $\Delta u = f$ where

$$\Delta = \sum_{j=1}^{n} \frac{\partial^2}{\partial x_j^2}$$

is the Laplace operator. A distribution E is called a **fundamental solution** of the Laplace operator if $\Delta E = \delta$. Now if such an E is given, and say $f \in \mathcal{S}$, then $u = E * f$ solves $\Delta u = f$ (by Example 3.9 and Exercise 3.11). When $n = 1$, a fundamental solution was given in Exercise 3.10. For $n = 2$ the function $\frac{1}{2\pi} \ln(|x|)$ is a fundamental solution while for $n > 2$, there is a fundamental solution of the form $\frac{c_n}{|x|^{n-2}}$ for some dimensional constant c_n. Fundamental solutions will be studied in the future.

References for a further reading

For a comprehensive exposition of distribution theory and other related topics, here are two excellent books:

(1) L.Hormander, "The analysis of linear partial differential operators I", Springer Verlag, 1983
(2) F.Treves, "Topological vector spaces, distributions and kernels", Academic Press, New York, 1967