SNAP 2017: Differentiating the Non-Differentiable
Northwestern University, Summer 2017

Aaron Peterson

Last updated: July 31, 2017

Contents

1 Motivation and Background

> I turn away with fright and horror from this dreadful plague of continuous functions which do not have derivatives.
>
> Charles Hermite

Students of analysis sympathize with Hermite. Not only are there continuous functions which do not have derivatives, but this phenomenon has lead to a proliferation of concepts intended to describe exactly how regular (or 'smooth') a function is: continuity, absolute continuity, Lipschitz continuity, Hölder continuity, differentiability, and so on. Indeed, this web of labels is necessitated by our desire to evaluate functions at specific points.

In physics one sees a slightly different picture: many functions representing physical quantities cannot be 'evaluated' in the traditional sense. Rather we may only observe the average or aggregate behavior of a function. For example, even though one cannot measure the velocity of a particle one *can* measure its average velocity over a small time interval. These averages are a decent substitute for evaluation. In this minicourse we take our cue from the physicists—treating functions as quantities that can only be measured in aggregate (rather than pointwise).

In particular, we will measure the local aggregate behavior of f by multiplying f by an appropriate smooth function ϕ and then integrating. By varying the function ϕ we will be able to get a complete picture of the aggregate behavior of f. For this reason the functions ϕ will be called **test functions**, since each one of them is used to test the behavior of f. We will ultimately generalize this idea in a way that does not require any sort of integration.

The two major references for this course are the two excellent books by Rudin [3] and Hörmander [2]. The book of Folland [1] offers a third treatment which is more focused on problems and applications. Much of this material can be found in those texts, albeit at a much higher level of generality.

The theory of generalized functions is most easily studied in the context of Lebesgue integration. However, this minicourse assumes that you have no familiarity with Lebesgue integration. Rather than begin our short course with an even shorter course on the Lebesgue integral, we will work with a definition of integrability that is based on Riemann's integral. In everything that follows, we assume that the sets involved are sufficiently 'nice' for the definitions to make sense. Everything we use in the course (both in theory and examples) will cause no problems in this regard.

We first introduce a bit of notation. For $\boldsymbol{a} \in \mathbb{R}^n$ and $\epsilon > 0$, define

$$\mathrm{Box}_\epsilon^n(\boldsymbol{a}) = \prod_{i=1}^n (a_i - \epsilon, a_i + \epsilon), \quad \text{so that} \quad \overline{\mathrm{Box}_\epsilon^n(\boldsymbol{a})} = \prod_{i=1}^n [a_i - \epsilon, a_i + \epsilon].$$

We will use the standard notation $B_R^n(\boldsymbol{a}) = \{\boldsymbol{x} \in \mathbb{R}^n \;:\; |\boldsymbol{x} - \boldsymbol{a}| < R\}$ to denote the (metric) ball of radius R centered at $\boldsymbol{a}$.

Definition 1.1. We say that $f : \mathbb{R}^n \to \mathbb{R}$ is **integrable** on (reasonably nice set) $A \subseteq \mathbb{R}^n$ if f is continuous on $\mathbb{R}^n$ except for possibly finitely many points, and if $\displaystyle\int_A |f| dV_n < \infty$ in the sense of (perhaps improper) Riemann integration. We say that f is **locally integrable** if f is integrable on K for every (reasonably nice) compact set $K \subseteq \mathbb{R}^n$.

Remark 1.2. If you are familiar with the Lebesgue integral, then you may replace the definition of the statement 'f is integrable on A' with the notion '$f \in L^1(A)$. The statement that 'f is locally integrable on $\mathbb{R}^n$' can also be replaced by the statement that '$f \in L^1_{loc}(\mathbb{R}^n)$'. If we use the Lebesgue integral, then we replace the loose condition that A be 'reasonably nice' with the mere assumption that A is measurable, and we completely remove the condition on the compact set K since such K are Borel.

Example 1.3. As an example, every continuous function $f : \mathbb{R}^n \to \mathbb{R}$ is locally integrable. To see this, for each compact set $K \subseteq \mathbb{R}^n$ the Extreme Value Theorem implies that there exists $M_K > 0$ with $|f(\boldsymbol{x})| \leq M_K$ for all $\boldsymbol{x} \in K$. Moreover, since K is compact there exists $R_K > 0$ with $K \subseteq \overline{\mathrm{Box}^n_{R_K}(\mathbf{0})}$. But then

$$\int_K |f(\boldsymbol{x})| dV_n(\boldsymbol{x}) \leq M_K \mathrm{Vol}(\overline{\mathrm{Box}^n_{R_K}(\mathbf{0})}) = M_K(2R_K)^n < +\infty.$$

Exercise 1. Prove that if $f : \mathbb{R}^n \to \mathbb{R}$ is locally integrable and if $\phi : \mathbb{R}^n \to \mathbb{R}$ is continuous, then $f\phi : \mathbb{R}^n \to \mathbb{R}$ is locally integrable.

Exercise 2. Let $f : \mathbb{R}^n \to \mathbb{R}$ be continuous at all but perhaps finitely many points. Prove that f is locally integrable if and only if $\int_{\overline{B^n_R(\mathbf{0})}} |f| dV_n < +\infty$ for all $R > 0$.

Exercise 3. Define $f : \mathbb{R}^n \to \mathbb{R}$ by

$$f(\boldsymbol{x}) = \begin{cases} 0 & \text{if } \boldsymbol{x} = \mathbf{0}, \\ \frac{1}{2}\ln(x_1^2 + \cdots + x_n^2) & \text{if } \boldsymbol{x} \neq \mathbf{0}. \end{cases}$$

Prove that f is locally integrable on $\mathbb{R}^n$ when $n = 1, 2$. (If you are familiar with spherical coordinates in $\mathbb{R}^n$ for $n \geq 3$, then you should prove the result for these values of n as well.)

Exercise 4. Define $f : \mathbb{R} \to \mathbb{R}$ by $f(x) = \frac{1}{x}$ $(x \neq 0)$ and $f(0) = 0$. Prove that f is not locally integrable on $\mathbb{R}$.

2 Test Functions

To define the class of smooth functions that we will use to 'test' the aggregate behavior of our generalized functions, we make the following definitions.

Definition 2.1. The **support** of a function $\phi : \mathbb{R}^n \to \mathbb{R}$, $\mathrm{supp}(\phi)$, is defined to be

$$\mathrm{supp}(\phi) = \overline{\{x \in \mathbb{R}^n \ : \ \phi(x) \neq 0\}}.$$

If $\mathrm{supp}(\phi)$ is compact, then we say that ϕ has **compact support**, or is **compactly supported**.

Remark 2.2. We can also think of $\mathrm{supp}(\phi)$ as the complement of the largest open subset of $\mathbb{R}^n$ on which $\phi = 0$.

Definition 2.3. Define $C^0(\mathbb{R}^n) := \{\phi : \mathbb{R}^n \to \mathbb{R} \ : \ \phi \text{ is continuous}\}$. For $k \in \mathbb{N}$, we recursively define

$$C^k(\mathbb{R}^n) := \{\phi : \mathbb{R}^n \to \mathbb{R} \ : \ \text{For each } i = 1, \ldots, n, \ \frac{\partial \phi}{\partial x_i} \text{ exists on } \mathbb{R}^n \text{ and } \frac{\partial \phi}{\partial x_i} \in C^{k-1}(\mathbb{R}^n)\}.$$

Finally, let $C^\infty(\mathbb{R}^n) := \displaystyle\bigcap_{k=0}^{\infty} C^k(\mathbb{R}^n)$.

From linearity of differentiation, and since linear combinations of continuous functions are continuous, one easily checks that $C^\infty(\mathbb{R}^n)$ (and, for that matter, each $C^k(\mathbb{R}^n)$ for $k \in \mathbb{Z}_{\geq 0}$) is a real vector space.

Definition 2.4. The space of **test functions** on $\mathbb{R}^n$, $\mathscr{D}(\mathbb{R}^n)$, is defined as

$$\mathscr{D}(\mathbb{R}^n) := \{\phi \in C^\infty(\mathbb{R}^n) \ : \ \mathrm{supp}(\phi) \text{ is compact}\}.$$

Remark 2.5. Some authors use the notation $C_c^\infty(\mathbb{R}^n)$ instead of $\mathscr{D}(\mathbb{R}^n)$. (The '$c$' denotes that the functions have compact support.) This is fine notation, but is somewhat more cumbersome.

Remark 2.6. Note that $\mathscr{D}(\mathbb{R}^n)$ always contains the zero function (which we will denote by 0), since 0 is smooth and $\mathrm{supp}(0) = \emptyset$ is compact.

In fact, $\mathscr{D}(\mathbb{R}^n)$ is a real vector space. Since $\mathscr{D}(\mathbb{R}^n)$ is a nonempty subset of the real vector space $C^\infty(\mathbb{R}^n)$, to show that $\mathscr{D}(\mathbb{R}^n)$ is a vector space it suffices to show that $\mathscr{D}(\mathbb{R}^n)$ is closed under function addition and scalar multiplication. Let $\phi, \psi \in \mathscr{D}(\mathbb{R}^n)$ and $\lambda \in \mathbb{R}$. Then

$$\mathrm{supp}(\lambda\phi) = \begin{cases} \emptyset & \text{if } \lambda = 0, \\ \mathrm{supp}(\phi) & \text{if } \lambda \neq 0, \end{cases}$$

and therefore $\mathrm{supp}(\lambda\phi)$ is compact since $\emptyset$ and $\mathrm{supp}(\phi)$ are compact. For sums we note that

$$\mathrm{supp}(\phi + \psi) \subseteq \mathrm{supp}(\phi) \cup \mathrm{supp}(\psi),$$

and therefore since $\mathrm{supp}(\phi)$ and $\mathrm{supp}(\psi)$ are compact and the finite union of compact sets is compact, $\mathrm{supp}(\phi + \psi)$ is a closed subset of a compact set (and hence is compact). Since $a\psi, \psi + \phi \in C^\infty(\mathbb{R}^n)$ as well, the proof is complete.

Bump Functions

The space $\mathscr{D}(\mathbb{R}^n)$ is very large. To demonstrate this we now construct a large class of test functions on $\mathbb{R}^n$ with numerous theoretical applications: the so-called **bump functions**.

Proposition 2.7. *For each* $\boldsymbol{a} \in \mathbb{R}^n$ *there exists a family of functions* $\{\eta_{\epsilon,\boldsymbol{a}}\}_{\epsilon>0} \subseteq \mathscr{D}(\mathbb{R}^n)$ *satisfying*

(i) $\operatorname{supp}(\eta_{\epsilon,\boldsymbol{a}}) = \overline{\operatorname{Box}_\epsilon^n(\boldsymbol{a})}$,

(ii) $0 < \eta_{\epsilon,\boldsymbol{a}}(\boldsymbol{x}) \le \eta_{\epsilon,\boldsymbol{a}}(\boldsymbol{a})$ *for all* $\boldsymbol{x} \in \operatorname{Box}_\epsilon^n(\boldsymbol{a})$, *and*

(iii) $\displaystyle\int_{\mathbb{R}^n} \eta_{\epsilon,\boldsymbol{a}} dV_n = 1.$

Moreover, $\eta_{\epsilon,\boldsymbol{0}}$ is even for all $\epsilon > 0$.

Proof. Recall[1] that

$$s : \mathbb{R} \to \mathbb{R}, \quad s(x) = \begin{cases} 0 & \text{if } x \le 0, \\ e^{-1/x} & \text{if } x > 0 \end{cases}$$

satisfies $s \in C^\infty(\mathbb{R})$ and $\operatorname{supp}(s) = [0, +\infty)$.

Now define $t_1 : \mathbb{R} \to \mathbb{R}$ by $t_1(x) = cs(1+x)s(1-x)$, where $c \in (0, \infty)$ will be chosen momentarily. Note immediately that t_1 is even, and that

(i) $t_1 \in C^\infty(\mathbb{R})$,

(ii) $\operatorname{supp}(t_1) = \overline{\operatorname{Box}_1^1(0)}$, and

(iii) $0 < t_1(x) \le t_1(0)$ for $x \in \operatorname{Box}_1^1(0)$.

Since $x \mapsto s(1+x)s(1-x)$ is continuous, non-negative, and strictly positive when $x = 0$, we have

$$0 < \int_{\mathbb{R}} s(1+x)s(1-x)dx =: c^{-1}.$$

This choice of c yields

(iv) $\displaystyle\int_{\mathbb{R}} t_1(x)dx = 1.$

For $n \in \mathbb{N}$ define $t_n : \mathbb{R}^n \to \mathbb{R}$ by $t_n(x_1, \ldots, x_n) := t_1(x_1)t_1(x_2)\cdots t_1(x_n)$. Then t_n is even and satisfies

(i) $t_n \in C^\infty(\mathbb{R}^n)$,

(ii) $\operatorname{supp}(t_n) = \overline{\operatorname{Box}_1^n(\boldsymbol{0})}$,

(iii) $0 < t_n(\boldsymbol{x}) \le t_n(\boldsymbol{0})$ for $\boldsymbol{x} \in \operatorname{Box}_1^n(\boldsymbol{0})$, and

(iv) $\displaystyle\int_{\mathbb{R}^n} t_n dV_n = 1,$

[1] This is a standard exercise in real analysis courses, so you may have already seen this. If not, then you should treat this as an exercise!

where (iv) holds by Fubini's Theorem. Since $\overline{\text{Box}_1^n(\mathbf{0})}$ is compact for each $n \in \mathbb{N}$, $t_n \in \mathscr{D}(\mathbb{R}^n)$ for each n.

Finally, for $\epsilon \in (0, \infty)$ and $\boldsymbol{a} \in \mathbb{R}^n$ define

$$\eta_{\epsilon,\boldsymbol{a}} : \mathbb{R}^n \to \mathbb{R}, \quad \eta_{\epsilon,\boldsymbol{a}}(\boldsymbol{x}) = \epsilon^{-n} t_n(\epsilon^{-1}(\boldsymbol{x} - \boldsymbol{a})).$$

Then for each ϵ and $\boldsymbol{a}$,

(i) $\eta_{\epsilon,\boldsymbol{a}} \in C^\infty(\mathbb{R}^n)$,

(ii) $\text{supp}(\eta_{\epsilon,\boldsymbol{a}}) = \overline{\text{Box}_\epsilon^n(\boldsymbol{a})}$,

(iii) $0 < \eta_{\epsilon,\boldsymbol{a}}(\boldsymbol{x}) \leq \eta_{\epsilon,\boldsymbol{a}}(\boldsymbol{a})$ for $\boldsymbol{x} \in \text{Box}_\epsilon^n(\boldsymbol{a})$, and

(iv) $\displaystyle\int_{\mathbb{R}^n} \eta_{\epsilon,\boldsymbol{a}}(\boldsymbol{x}) dV_n(\boldsymbol{x}) = \int_{\mathbb{R}^n} t_n(\boldsymbol{y}) dV_n(\boldsymbol{y}) = 1,$

where in (iv) we made the change of variable $\boldsymbol{x} = \epsilon\boldsymbol{y} + \boldsymbol{a}$, so that $\epsilon^{-1}(\boldsymbol{x} - \boldsymbol{a}) = \boldsymbol{y}$ and $dV_n(\boldsymbol{x}) = \epsilon^n dV_n(\boldsymbol{y})$. Finally, the evenness of t_n guarantees that $\eta_{\epsilon,\mathbf{0}}$ is even for all $\epsilon > 0$, and the proof is complete. □

The following example illustrates not only the utility of the class of test functions we've constructed, but also the suitability of 'testing' as a substitute for pointwise evaluation.

Example 2.8. Let $f : \mathbb{R}^n \to \mathbb{R}$ be locally integrable, and define

$$L_f : \mathscr{D}(\mathbb{R}^n) \to \mathbb{R}, \quad \langle L_f, \phi \rangle = L_f(\phi) := \int_{\mathbb{R}^n} f(\boldsymbol{x})\phi(\boldsymbol{x}) dV_n(\boldsymbol{x}).$$

By Exercise 1, $f\phi$ is locally integrable for all $\phi \in \mathscr{D}(\mathbb{R}^n)$, and since $\text{supp}(f\phi) \subseteq \text{supp}(\phi)$ is compact the integral $\langle L_f, \phi \rangle$ converges. One easily verifies that L_f is a linear functional[2] on $\mathscr{D}(\mathbb{R}^n)$.

The quantity $\langle L_f, \phi \rangle$ can be considered a snapshot of the aggregate behavior of f as measured using the function ϕ. The big question for us is the following: How is L_f related to f?

By definition, the locally integrable function f completely determines the functional L_f. Therefore, full knowledge of f leads directly to full knowledge about L_f. For the other direction, full knowledge about L_f (i.e. knowing the value of $\langle L_f, \phi \rangle$ for every $\phi \in \mathscr{D}(\mathbb{R}^n)$) allows us to recover the (pointwise) values of f at every point at which f is continuous.

Exercise 5. Let $f : \mathbb{R}^n \to \mathbb{R}$ be locally integrable and suppose that f is continuous at $\boldsymbol{a} \in \mathbb{R}^n$. Prove that

$$\lim_{\epsilon \to 0+} \langle L_f, \eta_{\epsilon,\boldsymbol{a}} \rangle = f(\boldsymbol{a}).$$

(Suggestion: It may be easier to show that $\lim_{\epsilon \to 0+} |\langle L_f, \eta_{\epsilon,\boldsymbol{a}} \rangle - f(\boldsymbol{a})| = 0$.)

[2] If V is a vector space over a field $\mathbb{F}$, then a linear transformation $L : V \to \mathbb{F}$ is called a **linear functional** on V. To facilitate computations, we will write $\langle L, v \rangle$ instead of $L(v)$. The branch of mathematics known as Functional Analysis is largely concerned with the study of functionals (both linear and nonlinear).

In other words, if f is locally integrable then the linear functional $L_f : \mathscr{D}(\mathbb{R}^n) \to \mathbb{R}$ allows us to recover the values of f at all but perhaps finitely many points. From a physical standpoint this amounts to complete knowledge about f, for if $g : \mathbb{R}^n \to \mathbb{R}$ is locally integrable and if $f = g$ at all but finitely many points, then $f - g = 0$ at all but finitely many points, so that $\int_{\mathbb{R}^n}(f - g)\phi dV_n = 0$ for all $\phi \in \mathscr{D}(\mathbb{R}^n)$ and therefore $\int_{\mathbb{R}^n} f\phi dV_n = \int_{\mathbb{R}^n} g\phi dV_n$ for all test functions ϕ. That is, under every possible 'test' of aggregate behavior the functions f and g give exactly the same result, and therefore are physically indistinguishable.

Remark 2.9. The phenomenon described at the end of the previous example is exactly why we do not study individual functions in the theory of Lebesgue integration. Instead we study equivalence classes of functions, where two functions are equivalent if they agree except for on a set of measure zero.

Example 2.8 suggests that linear functionals on $\mathscr{D}(\mathbb{R}^n)$ might be used to generalize functions on $\mathbb{R}^n$. The only additional detail we need in order to make this theory workable is an appropriate notion of continuity for these linear functionals. Our story becomes rather technical here because the vector space $\mathscr{D}(\mathbb{R}^n)$ is infinite dimensional, and therefore the notion of continuity for a linear map $L : \mathscr{D}(\mathbb{R}^n) \to \mathbb{R}$ is not an automatic consequence of linearity (as in the finite dimensional case). We establish some notation before stating the relevant topological facts.

Notation 2.10. A **multi-index** is a tuple $\alpha = (\alpha_1, \ldots, \alpha_n) \in \mathbb{Z}^n_{\geq 0}$. If α is a multi-index, then we define $|\alpha| = \alpha_1 + \cdots + \alpha_n$. *If $\alpha, \beta \in \mathbb{Z}^n_{\geq 0}$, then we define $\alpha + \beta = (\alpha_1 + \beta_1, \ldots, \alpha_n + \beta_n)$.*

Notation 2.11. Let $\alpha \in \mathbb{Z}^n_{\geq 0}$ be a multi-index. We define the partial differential operator ∂^α to be

$$\partial^\alpha = \frac{\partial^{|\alpha|}}{\partial x_1^{\alpha_1} \cdots \partial x_n^{\alpha_n}}.$$

Remark 2.12. Recall that for $k \in \{2, 3, \ldots, \}$ and $\phi \in C^k(\mathbb{R}^n)$, Clairaut's Theorem implies that all mixed partial derivatives of ϕ of order no more than k commute. For example, if $\phi \in C^2(\mathbb{R}^n)$ then $\frac{\partial^2\phi}{\partial x_i \partial x_j}(\boldsymbol{x}) = \frac{\partial^2\phi}{\partial x_j \partial x_i}(\boldsymbol{x})$ for all $i, j \in \{1, \ldots, n\}$ and all $\boldsymbol{x} \in \mathbb{R}^n$. This property is sometimes called **symmetry of mixed partial derivatives** (up to order k).

For an arbitrary mixed partial derivative of a smooth (C^∞) function we therefore need only track how many derivatives were taken in each variable (and not in what order these derivatives were taken). Multi-indices are perfect for this task, since an order $|\alpha| = \alpha_1 + \cdots + \alpha_n$ mixed partial derivative of $\phi \in C^\infty(\mathbb{R}^n)$ consisting of α_j derivatives with respect to x_j (for $j = 1, \ldots, n$), regardless of the order in which these derivatives were computed, can be expressed as $\partial^\alpha \phi = \frac{\partial^{|\alpha|}\phi}{\partial_{x_1}^{\alpha_1} \cdots \partial_{x_n}^{\alpha_n}}$.

The topology on $\mathscr{D}(\mathbb{R}^n)$ is closely tied to a family of norms, each of which tracks of a different level of smoothness.

Notation 2.13. Let $\phi \in \mathscr{D}(\mathbb{R}^n)$. Then for each $N \in \mathbb{Z}_{\geq 0}$ we define

$$\|\phi\|_N = \sum_{|\alpha| \leq N} \sup_{\boldsymbol{x} \in \mathbb{R}^n} |\partial^\alpha \phi(\boldsymbol{x})|.$$

Note that for $\phi \in \mathscr{D}(\mathbb{R}^n)$ the Extreme Value Theorem and the compactness of $\mathrm{supp}(\phi)$ guarantee that $\|\phi\|_N < \infty$ for every $N \in \mathbb{Z}_{\geq 0}$.

Convergence with respect to one of the norms $\| \bullet \|_N$ is equivalent to uniform convergence of partial derivatives up to order N, in the following sense.

Proposition 2.14. *Let (ϕ_j) be a sequence in $\mathscr{D}(\mathbb{R}^n)$ and $N \in \mathbb{Z}_{\geq 0}$. Then $\|\phi_j\|_N \to 0$ as $j \to \infty$ if and only if for every multi-index α with $|\alpha| \leq N$, $\partial^\alpha \phi_j \to 0$ uniformly on $\mathbb{R}^n$ as $j \to \infty$.*

We are now ready to describe the topological structure of $\mathscr{D}(\mathbb{R}^n)$. The construction of this topology and the proof that it is the 'natural' one for our purposes are highly technical and would take us far afield; for all of the gory (but aestecially satisfying) details, see [3]. Henceforth we assume that $\mathscr{D}(\mathbb{R}^n)$ has this special topology, the relevant properties of which are summarized in the following theorem.

Theorem 2.15. *There exists a topology on $\mathscr{D}(\mathbb{R}^n)$ with the following properties.*

(a) For every sequence (ϕ_j) in $\mathscr{D}(\mathbb{R}^n)$ and every function $\phi \in \mathscr{D}(\mathbb{R}^n)$, $\phi_j \to \phi$ in $\mathscr{D}(\mathbb{R}^n)$ if and only if the following two conditions hold:

- *(i) There exists a compact set $K \subset \mathbb{R}^n$ with $\operatorname{supp}(\phi_j) \subseteq K$ for all $j \in \mathbb{N}$, and*
- *(ii) For each $N \in \mathbb{Z}_{\geq 0}$ we have $\|\phi_j - \phi\|_N \to 0$ as $j \to \infty$.*

(b) Let $L : \mathscr{D}(\mathbb{R}^n) \to \mathbb{R}$ be linear. Then the following are equivalent:

- *(i) L is continuous (i.e. $L^{-1}(U)$ is open in $\mathscr{D}(\mathbb{R}^n)$ if $U \subseteq \mathbb{R}$ is open).*
- *(ii) For all $\phi_j, \phi \in \mathscr{D}(\mathbb{R}^n)$ $(j \in \mathbb{N})$ with $\phi_j \to \phi$ in $\mathscr{D}(\mathbb{R}^n)$, $\lim_{j\to\infty} \langle L, \phi_j \rangle = \langle L, \phi \rangle$.*
- *(iii) For all $\phi_j \in \mathscr{D}(\mathbb{R}^n)$ $(j \in \mathbb{N})$ with $\phi_j \to 0$ in $\mathscr{D}(\mathbb{R}^n)$, $\lim_{j\to\infty} \langle L, \phi_j \rangle = 0$.*
- *(iv) For every compact set $K \subset \mathbb{R}^n$ there exists $C > 0$ and $N \in \mathbb{Z}_{\geq 0}$ such that for all $\phi \in \mathscr{D}(\mathbb{R}^n)$ with $\operatorname{supp}(\phi) \subseteq K$, it follows that $|\langle L, \phi \rangle| \leq C\|\phi\|_N$.*

Example 2.16. Let (ϕ_j) be a sequence in $\mathscr{D}(\mathbb{R}^n)$ with $\phi_j \to 0$ in $\mathscr{D}(\mathbb{R}^n)$. Then for every multi-index α, $\partial^\alpha \phi_j \to 0$ in $\mathscr{D}(\mathbb{R}^n)$ as well.

To see why, note that if β is any multi-index, then $\partial^\beta(\partial^\alpha \phi_j) = \partial^{\beta+\alpha}\phi_j \to 0$ uniformly on $\mathbb{R}^n$ since $\phi_j \to 0$ in $\mathscr{D}(\mathbb{R}^n)$. Note also that there exists a compact set $K \subset \mathbb{R}^n$ with $\operatorname{supp}(\phi_j) \subseteq K$, and hence $\operatorname{supp}(\partial^\alpha \phi_j) \subseteq K$ for all j. This completes the proof.

Example 2.17. For a more involved example, fix $\phi \in \mathscr{D}(\mathbb{R}^n)$ and $\boldsymbol{z} \in B_1^n(\mathbf{0}) \subset \mathbb{R}^n$. Then $\phi(\boldsymbol{x} - \boldsymbol{z}) \to \phi(\boldsymbol{x})$ in $\mathscr{D}(\mathbb{R}^n)$ as $\boldsymbol{z} \to \mathbf{0}$.

To see why, note first that since $\operatorname{supp}(\phi)$ is compact there exists $R > 0$ with $\operatorname{supp}(\phi) \subseteq \overline{B_R^n(\mathbf{0})}$. But then since $|\boldsymbol{z}| \leq 1$ for all $\boldsymbol{z} \in B_1^n(\mathbf{0})$ we have $\operatorname{supp}(\phi(\bullet - \boldsymbol{z})) \subseteq \overline{B_{R+1}^n(\mathbf{0})}$ for all $\boldsymbol{z} \in B_1^n(\mathbf{0})$.

It remains to show that for each multi-index α, $\partial^\alpha\phi(\bullet - \boldsymbol{z}) \to \partial^\alpha \phi(\bullet)$ uniformly as $\boldsymbol{z} \to \mathbf{0}$. But we merely note that for $\boldsymbol{x} \in \mathbb{R}^n$ the Mean Value Theorem implies that there exists $\boldsymbol{c}$ between $\boldsymbol{x}$ and $\boldsymbol{x} - \boldsymbol{z}$ with

$$|\partial^\alpha\phi(\boldsymbol{x} - \boldsymbol{z}) - \partial^\alpha\phi(\boldsymbol{x})| = |\nabla(\partial^\alpha\phi)(\boldsymbol{c}) \bullet (\boldsymbol{x} - \boldsymbol{z} - \boldsymbol{x})| \leq |\nabla(\partial^\alpha\phi)(\boldsymbol{c})||\boldsymbol{z}| \leq \sqrt{n}\|\boldsymbol{\phi}\|_{|\alpha|+1}|\boldsymbol{z}| \to 0$$

as $\boldsymbol{z} \to \mathbf{0}$, uniformly in $\boldsymbol{x}$. This completes the proof.

Exercise 6. Let $\phi \in \mathscr{D}(\mathbb{R}^n)$ and fix $i \in \{1, \ldots, n\}$. For $h \in [-1, 1]$ with $h \neq 0$ define $\phi_{h,i}(\boldsymbol{x}) = \frac{\phi(\boldsymbol{x}+h\boldsymbol{e}_i) - \phi(\boldsymbol{x})}{h}$. Prove that $\phi_{h,i} \to \frac{\partial\phi}{\partial x_i}$ in $\mathscr{D}(\mathbb{R}^n)$ as $h \to 0$.

For the next exercise, recall that if $A, B \subseteq \mathbb{R}^n$, then $A+B := \{\boldsymbol{x}+\boldsymbol{y} \,:\, \boldsymbol{x} \in A \text{ and } \boldsymbol{y} \in B\}$.

Exercise 7. For $\psi \in C^\infty(\mathbb{R}^n)$ and $\phi \in \mathscr{D}(\mathbb{R}^n)$ we define the **convolution** of ψ and ϕ, $\psi * \phi$, by

$$\psi * \phi : \mathbb{R}^n \to \mathbb{R}, \quad (\psi * \phi)(\boldsymbol{x}) = \int_{\mathbb{R}^n} \psi(\boldsymbol{x} - \boldsymbol{y})\phi(\boldsymbol{y})dV_n(\boldsymbol{y}).$$

(a) Prove that $(\psi * \phi)(\boldsymbol{x}) = (\phi * \psi)(\boldsymbol{x})$.

(b) Prove that $\operatorname{supp}(\psi * \phi) \subseteq \operatorname{supp}(\psi) + \operatorname{supp}(\phi)$.

(c) Prove that $\psi * \phi \in C^\infty(\mathbb{R}^n)$ with $\partial^\alpha(\psi * \phi) = (\partial^\alpha \psi) * \phi = \psi * (\partial^\alpha \phi)$ for every multi-index α, and therefore $\psi * \phi \in C^\infty(\mathbb{R}^n)$. As a consequence, show that $\psi * \phi \in \mathscr{D}(\mathbb{R}^n)$ in the special case where $\operatorname{supp}(\psi)$ is compact.

(d) For $h > 0$, consider the Riemann sum

$$S_h(\boldsymbol{x}) = \sum_{\boldsymbol{m} \in \mathbb{Z}^n} \psi(\boldsymbol{x} - h\boldsymbol{m})\phi(h\boldsymbol{m})h^n.$$

(i) Prove that for each $h > 0$ all but finitely many terms in the sum S_h are identially zero (and therefore the sum is finite).

(ii) Now assume that $\psi \in \mathscr{D}(\mathbb{R}^n)$. Prove that $S_h \in \mathscr{D}(\mathbb{R}^n)$ for all $h > 0$ and that $S_h \to \psi * \phi$ in $\mathscr{D}(\mathbb{R}^n)$ as $h \to 0+$.
(Suggestion: For $\boldsymbol{x} \in \mathbb{R}^n$, first write

$$(\psi * \phi)(\boldsymbol{x}) = \sum_{\boldsymbol{z} \in \mathbb{Z}^n} \int_{h\boldsymbol{m}+[0,h]^n} \psi(\boldsymbol{x} - \boldsymbol{y})\phi(\boldsymbol{y})dV_n(\boldsymbol{y}),$$

and then, for each multi-index α, estimate $|\partial^\alpha S_h(\boldsymbol{x}) - \partial^\alpha(\psi * \phi)(\boldsymbol{x})|$.)

Exercise 8. Prove that if $\phi \in \mathscr{D}(\mathbb{R}^n)$, then $\eta_{\epsilon,\mathbf{0}} * \phi \to \phi$ in $\mathscr{D}(\mathbb{R}^n)$ as $\epsilon \to 0+$.

3 Distributions (Generalized Functions)

In light of Theorem 2.15 we now give our definition of generalized functions on $\mathbb{R}^n$.

Definition 3.1. A **distribution** (or **generalized function**) on $\mathbb{R}^n$ is a continuous linear functional $L : \mathscr{D}(\mathbb{R}^n) \to \mathbb{R}$. The space of all distributions on $\mathbb{R}^n$ is denoted by $\mathscr{D}'(\mathbb{R}^n)$.

Remark 3.2. The 'prime' notation $\mathscr{D}'(\mathbb{R}^n)$ is standard, and signifies that $\mathscr{D}'(\mathbb{R}^n)$ is the continuous dual space of $\mathscr{D}(\mathbb{R}^n)$.

Remark 3.3. $\mathscr{D}'(\mathbb{R}^n)$ is a real vector space, with sums and scalar multiplication of real-valued functions defined in the usual way. To see this, note that it is trivial to show that linear combinations of linear functions are linear. For continuity, let $L, S \in \mathscr{D}'(\mathbb{R}^n)$ and $a \in \mathbb{R}$, and assume that (ϕ_j) is a sequence in $\mathscr{D}(\mathbb{R}^n)$ with $\phi_j \to 0$ in $\mathscr{D}(\mathbb{R}^n)$ as $j \to \infty$. But then since L and S are continuous, Theorem 2.15 implies that $\langle L, \phi_j \rangle \to 0$ and $\langle S, \phi_j \rangle \to 0$ as $j \to \infty$, and therefore

$$\langle aL, \phi_j \rangle = a \langle L, \phi_j \rangle \to 0 \quad \text{and} \quad \langle L + S, \phi_j \rangle = \langle L, \phi_j \rangle + \langle S, \phi_j \rangle \to 0$$

as $j \to \infty$. By Theorem 2.15, aL and $L + S$ are also continuous.

Example 3.4. If f is locally integrable on $\mathbb{R}^n$, then

$$L_f : \mathscr{D}(\mathbb{R}^n) \to \mathbb{R}, \quad \langle L_f, \phi \rangle = \int_{\mathbb{R}^n} f(\boldsymbol{x})\phi(\boldsymbol{x}) dV_n(\boldsymbol{x})$$

is a distribution on $\mathbb{R}^n$.

Linearity is immediate from linearity of function multiplication and integration. For continuity, let $K \subset \mathbb{R}^n$ be compact and let $C_K = \int_K |f(\boldsymbol{x})| dV_n(\boldsymbol{x})$. Then we have

$$|\langle L_f, \phi \rangle| \leq \int_{\mathbb{R}^n} |f(\boldsymbol{x})||\phi(\boldsymbol{x})| dV_n(\boldsymbol{x}) = \int_K |f(\boldsymbol{x})||\phi(\boldsymbol{x})| dV_n(\boldsymbol{x}) \leq \|\phi\|_0 \int_K |f(\boldsymbol{x})| dV_n(\boldsymbol{x}) = C_K \|\phi\|_0$$

for all $\phi \in \mathscr{D}(\mathbb{R}^n)$ with $\operatorname{supp}(\phi) \subseteq K$. By part (b) of Theorem 2.15, L_f is continuous, and therefore $L_f \in \mathscr{D}'(\mathbb{R}^n)$.

Definition 3.5. If $L \in \mathscr{D}'(\mathbb{R}^n)$ and if there exists a locally integrable function $f : \mathbb{R}^n \to \mathbb{R}$ such that $\langle L, \phi \rangle = \int_{\mathbb{R}^n} f(\boldsymbol{x})\phi(\boldsymbol{x}) dV_n(\boldsymbol{x})$ for all $\phi \in \mathscr{D}(\mathbb{R}^n)$, then L is called a **regular** distribution. If L is not regular, then L is called **singular**.

Remark 3.6. When a distribution L is regular (say $\langle L, \phi \rangle = \int f \phi dV_n$ for some locally integrable $f : \mathbb{R}^n \to \mathbb{R}$), then it is customary to write f in place of L (i.e. $\langle f, \phi \rangle$ instead of $\langle L, \phi \rangle$). When we perform some operation on f which only makes sense to perform on the distribution L, then we say that we are performing the operation on f **in the sense of distributions**.

Exercise 9. For $L \in \mathscr{D}'(\mathbb{R}^n)$ and $g \in C^\infty(\mathbb{R}^n)$, define

$$gL : \mathscr{D}(\mathbb{R}^n) \to \mathbb{R}, \quad \langle gL, \phi \rangle := \langle L, g\phi \rangle .$$

Prove that $gL \in \mathscr{D}'(\mathbb{R}^n)$.

Exercise 10. Find all $L \in \mathscr{D}'(\mathbb{R})$ that satisfy $xL = 0$ in the sense of distributions. (Suggestion: First produce a fixed $\psi \in \mathscr{D}(\mathbb{R})$ with $\psi(x) = 1$ for $x \in [-1, 1]$. For $\phi \in \mathscr{D}(\mathbb{R})$, write $\phi = \psi\phi + (1 - \psi)\phi$. Show that $\langle L, (1 - \psi)\phi \rangle = 0$, and then use an appropriate Taylor expansion (with remainder) on ϕ to compute $\langle L, \psi\phi \rangle$.)

Exercise 11. For $g \in C^0(\mathbb{R}^n)$ and $\boldsymbol{x}_0 \in \mathbb{R}^n$ define $\tau_{\boldsymbol{x}_0} g(\boldsymbol{x}) := g(\boldsymbol{x} - \boldsymbol{x}_0)$.

(a) Prove that for $f \in C^0(\mathbb{R}^n)$, $\phi \in \mathscr{D}(\mathbb{R}^n)$, and $\boldsymbol{x}_0 \in \mathbb{R}^n$ we have

$$\int_{\mathbb{R}^n} (\tau_{\boldsymbol{x}_0} f)(\boldsymbol{x})\phi(\boldsymbol{x}) dV_n(\boldsymbol{x}) = \int_{\mathbb{R}^n} f(\boldsymbol{x})(\tau_{-\boldsymbol{x}_0}\phi)(\boldsymbol{x}) dV_n(\boldsymbol{x}).$$

(b) To generalize (a), let $L \in \mathscr{D}'(\mathbb{R}^n)$ and $\boldsymbol{x}_0 \in \mathbb{R}^n$. Prove that the map

$$\tau_{\boldsymbol{x}_0} L : \mathscr{D}(\mathbb{R}^n) \to \mathbb{R}, \quad \langle \tau_{\boldsymbol{x}_0} L, \phi \rangle := \langle L, \tau_{-\boldsymbol{x}_0}\phi \rangle$$

is a distribution on $\mathbb{R}^n$.

The next natural question, of course, is whether or not there exist singular distributions. We answer this question positively by introducing a distribution with widespread applications throughout physics, partial differential equations, and mathematical analysis.

Example 3.7. Define

$$\delta : \mathscr{D}(\mathbb{R}^n) \to \mathbb{R}, \quad \langle \delta, \phi \rangle = \phi(\mathbf{0}).$$

Linearity of δ is immediate from the definition of function addition and scalar multiplication, and continuity follows easily since $\langle \delta, \phi_j \rangle = \phi_j(\mathbf{0}) \to 0$ for every sequence (ϕ_j) in $\mathscr{D}(\mathbb{R}^n)$ with $\phi_j \to 0$ in $\mathscr{D}(\mathbb{R}^n)$. For historical reasons, we call $\delta \in \mathscr{D}'(\mathbb{R}^n)$ the **Dirac delta function**. We say 'historical' because there is no locally integrable function $f : \mathbb{R}^n \to \mathbb{R}$ with $\langle \delta, \phi \rangle = \int_{\mathbb{R}^n} f(\boldsymbol{x})\phi(\boldsymbol{x}) dV_n(\boldsymbol{x})$ for all $\phi \in \mathscr{D}(\mathbb{R}^n)$ (and therefore the word 'function' is inappropriate here).

To see why, suppose to the contrary that such a function f exists. Define[3]

$$\phi_j(\boldsymbol{x}) = \frac{1}{\eta_{1/j,\mathbf{0}}(\mathbf{0})} \eta_{1/j,\mathbf{0}}(\boldsymbol{x}) \in \mathscr{D}(\mathbb{R}^n), \quad j \in \mathbb{N}.$$

Then

$$\operatorname{supp}(\phi_j) = \overline{\operatorname{Box}^n_{1/j}(\mathbf{0})} \subseteq \overline{\operatorname{Box}^n_1(\mathbf{0})} \quad \text{and} \quad 0 < \phi_j(\boldsymbol{x}) \le \phi_j(\mathbf{0}) = 1$$

for every $j \in \mathbb{N}$ and $\boldsymbol{x} \in \mathbb{R}^n$. Since $\operatorname{supp}(\phi_j) = \overline{\operatorname{Box}^n_{1/j}(\mathbf{0})}$ we have

$$\phi_j(\boldsymbol{x}) \to \begin{cases} 1 & \text{if } \boldsymbol{x} = \mathbf{0}, \\ 0 & \text{if } \boldsymbol{x} \ne \mathbf{0}, \end{cases} \quad \text{and therefore} \quad (f\phi_j)(\boldsymbol{x}) \to \begin{cases} f(\mathbf{0}) & \text{if } \boldsymbol{x} = \mathbf{0}, \\ 0 & \text{if } \boldsymbol{x} \ne \mathbf{0}. \end{cases}$$

Since $\displaystyle\int_{\overline{\operatorname{Box}^n_1(\mathbf{0})}} |f| dV_n < \infty$ and since $|f(\boldsymbol{x})\phi_j(\boldsymbol{x})| \le |f(\boldsymbol{x})|$ for all $\boldsymbol{x} \in \mathbb{R}^n$, Lebesgue's Dominated Convergence Theorem implies that

$$\int_{\mathbb{R}^n} f(\boldsymbol{x})\phi_j(\boldsymbol{x}) dV_n(\boldsymbol{x}) = \int_{\overline{\operatorname{Box}^n_1(\mathbf{0})}} f(\boldsymbol{x})\phi_j(\boldsymbol{x}) dV_n(\boldsymbol{x}) \to 0 \quad \text{as } j \to \infty.$$

But $\int_{\mathbb{R}^n} f(\boldsymbol{x})\phi_j(\boldsymbol{x}) dV_n(\boldsymbol{x}) = \langle \delta, \phi_j \rangle = \phi_j(\mathbf{0}) = 1$ for all j, a contradiction.

[3] Here we need the fact that $t_n(\mathbf{0}) > 0$.

Exercise 12. For each $\epsilon > 0$, define $L_\epsilon : \mathscr{D}(\mathbb{R}) \to \mathbb{R}$ by

$$\langle L_\epsilon, \phi \rangle = \int_{|x|\geq\epsilon} \frac{\phi(x)}{x} dx = \int_{-\infty}^{-\epsilon} \frac{\phi(x)}{x} dx + \int_{\epsilon}^{\infty} \frac{\phi(x)}{x} dx.$$

(a) Prove that for each $R > 0$ and for every $\phi \in \mathscr{D}(\mathbb{R}^n)$ with $\operatorname{supp}(\phi) \subseteq [-R, R]$,

$$\langle L_\epsilon, \phi \rangle = \int_{\epsilon\leq|x|\leq R} \frac{\phi(x) - \phi(0)}{x} dx.$$

(b) Produce $C = C(R)$, independent of ϵ, so that for every $\phi \in \mathscr{D}(\mathbb{R})$ with $\operatorname{supp}(\phi) \subseteq [-R, R]$,

$$|\langle L_\epsilon, \phi \rangle| \leq C\|\phi\|_1.$$

(c) For each $\phi \in \mathscr{D}(\mathbb{R})$, prove that $\langle L, \phi \rangle := \lim\limits_{\epsilon\to0+} \langle L_\epsilon, \phi \rangle$ exists. Prove that $L : \mathscr{D}(\mathbb{R}) \to \mathbb{R}$ is linear and satisfies $|\langle L, \phi \rangle| \leq C\|\phi\|_1$ for all $\phi \in \mathscr{D}(\mathbb{R})$ with $\operatorname{supp}(\phi) \subseteq [-R, R]$, where $C = C(R)$ is the constant from (b), and conclude that $L \in \mathscr{D}'(\mathbb{R})$.

L is the closest thing we have to a regular distribution corresponding to the (not locally integrable) function $\frac{1}{x}$. The distribution L is called the **principal value** of $\frac{1}{x}$, and is denoted $L = p.v.\left(\frac{1}{x}\right)$.

4 Limits of Distributions

Now that we have defined distributions, the next step is to analyze their properties. Our goal here is to extend many of the standard calculus operations on functions into our general setting. We already know that $\mathscr{D}'(\mathbb{R}^n)$ is a real vector space, and we know how to multiply a distribution L by a smooth function g (i.e. $\langle gL, \phi \rangle = \langle L, g\phi \rangle$). We next consider the surprisingly easy operation of computing limits of sequences of distributions.

Theorem 4.1. *Let (L_j) be a sequence in $\mathscr{D}'(\mathbb{R}^n)$. Assume for each $\phi \in \mathscr{D}(\mathbb{R}^n)$ the sequence $(\langle L_j, \phi \rangle)$ converges (in $\mathbb{R}$), and define*

$$L : \mathscr{D}(\mathbb{R}^n) \to \mathbb{R}, \quad \langle L, \phi \rangle = \lim_{j\to\infty} \langle L, \phi_j \rangle.$$

Then $L \in \mathscr{D}'(\mathbb{R}^n)$.

Proof. Linearity of L follows immediately from the linearity of the L_j. Continuity of L is not so easy and relies on a fundamental result from functional analysis: the Uniform Boundedness Principle (also known as the Banach-Steinhaus Theorem); see [2] or [3] for the details. $\square$

Therefore, sequences of distributions which converge 'pointwise' on $\mathscr{D}(\mathbb{R}^n)$ (where the 'points' here are test functions) converge to distributional limits!

Example 4.2. In the previous exercise we showed by hand that $p.v.\left(\frac{1}{x}\right)$ is a distribution on $\mathbb{R}$. In light of this last theorem, we could also have noted that since $p.v.\left(\frac{1}{x}\right) = \lim\limits_{\epsilon\to0+} L_\epsilon$ in the sense of distributions, it is also a distribution.

Exercise 13. For $k \in \mathbb{N}$, let $f_k : \mathbb{R} \to \mathbb{R}$ be $f_k(x) = \cos(kx)$. Prove that $\lim_{k\to\infty} f_k = 0$ in the sense of distributions.

Exercise 14. Show that $\lim_{\epsilon\to 0+} \eta_{\epsilon,\mathbf{0}} = \delta$ in the sense of distributions.

5 Distributional Derivatives

We now come to differentiation. To motivate the definition of distributional derivatives, we consider the following (regular) example.

Example 5.1. Let $f \in C^1(\mathbb{R}^n)$ and fix $i \in \{1, \ldots, n\}$. Let $\phi \in \mathscr{D}(\mathbb{R}^n)$. Since $f \in C^1(\mathbb{R}^n)$ we have $\frac{\partial f}{\partial x_i} \in C^0(\mathbb{R}^n)$ and therefore $\frac{\partial f}{\partial x_i}\phi \in C^0(\mathbb{R}^n)$. Then

$$\left\langle \frac{\partial f}{\partial x_i}, \phi \right\rangle = \int_{\mathbb{R}^n} \frac{\partial f}{\partial x_i}(\boldsymbol{x})\phi(\boldsymbol{x})dV_n.$$

Moreover, $\operatorname{supp}(\frac{\partial f}{\partial x_i}\phi) \subseteq \operatorname{supp}(\phi) \subseteq \overline{\operatorname{Box}_R^n(\mathbf{0})}$ for some large $R > 0$. But then we apply Fubini's Theorem and integrate by parts to see that

$$\begin{aligned}
&\int_{\mathbb{R}^n} \frac{\partial f}{\partial x_i}(\boldsymbol{x})\phi(\boldsymbol{x})dV_n(\boldsymbol{x}) \\
= \;& \int_{\overline{\operatorname{Box}_R(\mathbf{0})}} \Big(\frac{\partial f}{\partial x_i}\phi\Big)(\boldsymbol{x})dV_n(\boldsymbol{x}) \\
= \;& \int_{[-R,R]^{n-1}} \left[\int_{-R}^{R} \Big(\frac{\partial f}{\partial x_i}\phi\Big)(x_1, \ldots, x_n)dx_i\right] dV_{n-1}(x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_n) \\
= \;& \int_{[-R,R]^{n-1}} \Bigg[(f\phi)(x_1, \ldots, R, \ldots, x_n) - (f\phi)(x_1, \ldots, -R, \ldots, x_n) \\
&\qquad\qquad - \int_{-R}^{R} \Big(f\frac{\partial \phi}{\partial x_i}\Big)(x_1, \ldots, x_n)dx_i\Bigg] dV_{n-1}(x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_n) \\
= \;& \int_{[-R,R]^{n-1}} \left[0 - 0 - \int_{-R}^{R} \Big(f\frac{\partial \phi}{\partial x_i}\Big)(x_1, \ldots, x_n)dx_i\right] dV_{n-1}(x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_n) \\
= \;& -\int_{\overline{\operatorname{Box}_R(\mathbf{0})}} f(\boldsymbol{x})\frac{\partial \phi}{\partial x_i}(\boldsymbol{x})dV_n(\boldsymbol{x}) \\
= \;& -\int_{\mathbb{R}^n} f(\boldsymbol{x})\frac{\partial \phi}{\partial x_i}(\boldsymbol{x})dV_n(\boldsymbol{x}),
\end{aligned}$$

where in the first, fourth, and final steps we used $\operatorname{supp}(f\phi) \subseteq \overline{\operatorname{Box}_R^n(\mathbf{0})}$. Therefore $\left\langle \frac{\partial f}{\partial x_i}, \phi \right\rangle = -\left\langle f, \frac{\partial \phi}{\partial x_i} \right\rangle$.

For $L \in \mathscr{D}'(\mathbb{R}^n)$, the 'integration by parts' formula

$$\left\langle \frac{\partial L}{\partial x_i}, \phi \right\rangle = -\left\langle L, \frac{\partial \phi}{\partial x_i} \right\rangle, \quad \phi \in \mathscr{D}(\mathbb{R}^n)$$

could therefore be a reasonable definition for $\frac{\partial L}{\partial x_i}$. We should first check, though, that this definition is always meaningful.

Proposition 5.2. *Let $L \in \mathscr{D}'(\mathbb{R}^n)$ and $i \in \{1, \dots, n\}$. Then*

$$S : \mathscr{D}(\mathbb{R}^n) \to \mathbb{R}, \quad \langle S, \phi \rangle := -\left\langle L, \frac{\partial \phi}{\partial x_i} \right\rangle$$

is a distribution on $\mathbb{R}^n$.

Proof. Linearity of S follows by noting that for every $\phi, \psi \in \mathscr{D}(\mathbb{R}^n)$ and all $a \in \mathbb{R}$ we have

$$\begin{aligned}\langle S, a\phi + \psi \rangle &= -\left\langle L, \frac{\partial(a\phi + \psi)}{\partial x_i} \right\rangle = -\left\langle L, a\frac{\partial \phi}{\partial x_i} + \frac{\partial \psi}{\partial x_i} \right\rangle = -a\left\langle L, \frac{\partial \phi}{\partial x_i} \right\rangle - \left\langle L, \frac{\partial \psi}{\partial x_i} \right\rangle \\ &= a\langle S, \phi \rangle + \langle S, \psi \rangle .\end{aligned}$$

For continuity, note that if (ϕ_j) is a sequence of test functions with $\phi_j \to 0$ in $\mathscr{D}(\mathbb{R}^n)$, then by Example 2.16 we have $\frac{\partial \phi_j}{\partial x_i} \to 0$ in $\mathscr{D}(\mathbb{R}^n)$. Since $L \in \mathscr{D}'(\mathbb{R}^n)$, Theorem 2.15 implies $\left\langle L, \frac{\partial \phi_j}{\partial x_i} \right\rangle \to 0$ (in $\mathbb{R}$), and therefore $\langle S, \phi_j \rangle = -\left\langle L, \frac{\partial \phi_j}{\partial x_i} \right\rangle \to 0$ (in $\mathbb{R}$). Another application of Theorem 2.15 concludes the result. □

We therefore give the following definition for distributional derivatives.

Definition 5.3. Let $L \in \mathscr{D}'(\mathbb{R}^n)$. For $i \in \{1, \dots, n\}$ we define the **distributional derivative of L with respect to** x_i to be

$$\frac{\partial L}{\partial x_i} : \mathscr{D}(\mathbb{R}^n) \to \mathbb{R}, \quad \left\langle \frac{\partial L}{\partial x_i}, \phi \right\rangle := -\left\langle L, \frac{\partial \phi}{\partial x_i} \right\rangle.$$

Proposition 5.2 guarantees that $\frac{\partial L}{\partial x_i} \in \mathscr{D}'(\mathbb{R}^n)$.

Notation 5.4. Partial derivatives only make sense when $n \geq 2$. As in calculus, we use the standard notation $\frac{dL}{dx}$ or L' for the distributional derivative of $L \in \mathscr{D}'(\mathbb{R})$.

Exercise 15. Here's another motivation for our definition of distributional derivatives, this time via the definition of partial derivatives from multivariable calculus. Recall the operator $\tau_{\boldsymbol{x}}$ from Exercise 11. Fix $i \in \{1, \dots, n\}$, and for $L \in \mathscr{D}'(\mathbb{R}^n)$ and $h \in \mathbb{R}\backslash\{0\}$ define

$$\Delta_h L = \frac{1}{h}(\tau_{-h\boldsymbol{e}_i} L - L) \in \mathscr{D}'(\mathbb{R}^n).$$

Prove that $\lim\limits_{h \to 0} \Delta_h L = \frac{\partial L}{\partial x_i}$ in the sense of distributions.

6 Properties of Distributional Derivatives

The theory of differentiation for distributions is much cleaner than that for ordinary functions. The next few remarks point out some of the highlights in this regard.

Remark 6.1. By Proposition 5.2 and the definition of distributional derivatives, a short induction argument implies that if $L \in \mathscr{D}'(\mathbb{R}^n)$, then L has distributional derivatives of all orders. This is quite different from the case of ordinary functions, where the number of derivatives one can take is usually limited. (This is, after all, the reason why we even define the spaces $C^0(\mathbb{R}^n)$, $C^1(\mathbb{R}^n)$, $C^2(\mathbb{R}^n)$, etc.)

Remark 6.2. Not only do distributions have derivatives of all orders, but we also have symmetry of mixed partial derivatives! To see this, note that if $L \in \mathscr{D}'(\mathbb{R}^n)$ and if $i, j \in \{1, \dots, n\}$, then for all $\phi \in \mathscr{D}(\mathbb{R}^n)$ we have

$$\left\langle \frac{\partial^2 L}{\partial x_i \partial x_j}, \phi \right\rangle = -\left\langle \frac{\partial L}{\partial x_j}, \frac{\partial \phi}{\partial x_i} \right\rangle = \left\langle L, \frac{\partial^2 \phi}{\partial x_j \partial x_i} \right\rangle = \left\langle L, \frac{\partial^2 \phi}{\partial x_i \partial x_j} \right\rangle = -\left\langle \frac{\partial L}{\partial x_i}, \frac{\partial \phi}{\partial x_j} \right\rangle = \left\langle \frac{\partial^2 L}{\partial x_j \partial x_i}, \phi \right\rangle.$$

Thus, as distributions, $\frac{\partial^2 L}{\partial x_i \partial x_j} = \frac{\partial^2 L}{\partial x_j \partial x_i}$. An induction argument establishes the analogous result for iterated derivatives of arbitrary order.

Example 6.3. In your real analysis or multivariable calculus course you undoubtedly encountered a function similar to

$$f : \mathbb{R}^2 \to \mathbb{R}, \quad f(x,y) = \begin{cases} 0 & \text{if } (x,y) = (0,0), \\ xy\left(\frac{x^2-y^2}{x^2+y^2}\right) & \text{if } (x,y) \neq (0,0). \end{cases}$$

It is a standard exercise to show that $f \in C^1(\mathbb{R}^2)$, that $f \in C^2(\mathbb{R}^2 \backslash \{(0,0)\})$, and that $\frac{\partial^2 f}{\partial x \partial y}(0,0)$ and $\frac{\partial^2 f}{\partial y \partial x}(0,0)$ exist but do not have the same value.

On the other hand, $\frac{\partial^2 f}{\partial x \partial y} = \frac{\partial^2 f}{\partial y \partial x}$ in the sense of distributions. This is entirely consistent with the pointwise discrepancy above since $\frac{\partial^2 f}{\partial x \partial y}(x,y) = \frac{\partial^2 f}{\partial y \partial x}(x,y)$ agree for all $(x,y) \neq (0,0)$, and two locally integrable functions which agree at all but a finite number of points are indistinguishable as distributions.

The preceeding remarks and an induction argument yield the following corollary.

Corollary 6.4. *Let $L \in \mathscr{D}'(\mathbb{R}^n)$. Then L has distributional derivatives of all orders and we have symmetry of mixed partial derivatives of all orders. Moreover, for every multi-index α we have*

$$\langle \partial^\alpha L, \phi \rangle = (-1)^{|\alpha|} \langle L, \partial^\alpha \phi \rangle \quad \textit{for all } \phi \in \mathscr{D}(\mathbb{R}^n).$$

Example 6.5. If $f \in C^k(\mathbb{R}^n)$ for some $k \in \mathbb{N}$, then the integration by parts argument from Example 5.1 shows that for all multi-indices α with $|\alpha| \leq k$, the distibutional derivative $\partial^\alpha f$ is exactly the regular distribution corresponding to the function $\partial^\alpha f$.

Example 6.6. For every multi-index α we have $\langle \partial^\alpha \delta, \phi \rangle = (-1)^{|\alpha|} (\partial^\alpha \phi)(\mathbf{0})$.

Exercise 16. (Product Rule) Prove that if $g \in C^\infty(\mathbb{R}^n)$, $L \in \mathscr{D}'(\mathbb{R}^n)$, and $i \in \{1, \ldots, n\}$, then $\frac{\partial}{\partial x_i}(gL) = \frac{\partial g}{\partial x_i} L + g \frac{\partial L}{\partial x_i}$.

Exercise 17. The **Heaviside function** is defined as

$$H : \mathbb{R} \to \mathbb{R}, \quad H(x) = \begin{cases} 0 & \text{if } x \leq 0, \\ 1 & \text{if } x > 0. \end{cases}$$

Prove that $H' = \delta$ in the sense of distributions on $\mathbb{R}$.

Exercise 18. Prove that $\frac{d}{dx} \ln|x| = p.v.\frac{1}{x}$ in the sense of distributions on $\mathbb{R}$.

Exercise 19. Prove the following result using the outline provided.

Theorem 6.7 (Antiderivatives)**.** *Let $L \in \mathscr{D}'(\mathbb{R})$. There exists $L^{(-1)} \in \mathscr{D}'(\mathbb{R})$ such that $\frac{dL^{(-1)}}{dx} = L$. Moreover, if $S \in \mathscr{D}'(\mathbb{R})$ satisfies $\frac{dS}{dx} = L$, then there exists a constant $C \in \mathbb{R}$ with $S = L^{(-1)} + C$ (as distributions). In particular, if $L \in \mathscr{D}'(\mathbb{R})$ with $\frac{dL}{dx} = 0$, then L is given by integration against a constant function.*

Fix $\omega \in \mathscr{D}(\mathbb{R})$ with $\int_{\mathbb{R}} \omega(x)dx = 1$. For $\phi \in \mathscr{D}(\mathbb{R})$ define

$$\Phi(x) = \int_{-\infty}^{x} \left[\phi(t) - \omega(t) \int_{\mathbb{R}} \phi(s)ds\right] dt.$$

(a) Prove that $\Phi \in \mathscr{D}(\mathbb{R})$ whenever $\phi \in \mathscr{D}(\mathbb{R})$.

(b) Define

$$L^{(-1)} : \mathscr{D}(\mathbb{R}) \to \mathbb{R}, \quad \left\langle L^{(-1)}, \phi \right\rangle := -\left\langle L, \Phi \right\rangle.$$

Prove that $L^{(-1)} \in \mathscr{D}'(\mathbb{R})$ and $\frac{dL^{(-1)}}{dx} = L$.

(c) Finish the proof.

The previous exercise has many applications, one of which proves that the distributional solutions of linear first-order ODEs with smooth coefficients must be regular. This result can be extended to ODEs of arbitrary order.

Exercise 20. Suppose that $L \in \mathscr{D}'(\mathbb{R})$ and that $g \in C^\infty(\mathbb{R})$. If $L' + gL = f \in C^0(\mathbb{R})$ in the sense of distributions, then there exists $u \in C^1(\mathbb{R})$ such that $L = L_u$ and $u' + gu = f$.

The next example is fundamental (one might say) to the theory of harmonic functions.

Example 6.8. Let $N : \mathbb{R}^2 \to \mathbb{R}$ be defined by $N(x,y) = \frac{1}{4\pi}\ln(x^2+y^2)$ when $(x,y) \neq 0$, and define $N(0,0) = 0$. Then $\Delta N = \delta$ in the sense of distributions, where $\Delta = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2}$ is Laplace's operator.

You proved that N is locally integrable on $\mathbb{R}^2$ in Exercise 3. For $\phi \in \mathscr{D}(\mathbb{R}^2)$, fix $R > 0$ so large that $\operatorname{supp}(\phi) \subseteq \overline{B_R^n(\mathbf{0})}$. We first compute that

$$\langle \Delta N, \phi\rangle = \langle N, \Delta\phi\rangle = \int_{\mathbb{R}^2} N(x,y)\Delta\phi(x,y)dV_2 = \lim_{\epsilon\to 0+}\int_{R\geq|(x,y)|\geq\epsilon} N(x,y)\Delta\phi(x,y)dV_2.$$

But note that on $\mathbb{R}^2\backslash\{(0,0)\}$ we have

$$N\Delta\phi = N\nabla\bullet(\nabla\phi) = \nabla\bullet(N\nabla\phi) - \nabla N\bullet\nabla\phi = \nabla\bullet(N\nabla\phi) - \nabla\bullet(\phi\nabla N) + \phi\Delta N.$$

A routine computation shows that for $(x,y) \neq (0,0)$ we have $\Delta N(x,y) = 0$, and therefore

$$\langle \Delta N, \phi\rangle = \lim_{\epsilon\to 0+}\Big[\int_{R\geq|(x,y)|\geq\epsilon}\nabla\bullet(N\nabla\phi)(x,y)dV_2 - \int_{R\geq|(x,y)|\geq\epsilon}\nabla\bullet(\phi\nabla N)(x,y)dV_2\Big].$$

We start with the first term. Since N and ϕ are each smooth away from $(0,0)$, for each $\epsilon \in (0, \min(1,R))$ we apply Green's Theorem to see that

$$\int_{R\geq|(x,y)|\geq\epsilon}\nabla\bullet(N\nabla\phi)(x,y)dV_2 = \int_{|(x,y)|=R}(N\nabla\phi)(x,y)\bullet\boldsymbol{n}d\sigma_R + \int_{|(x,y)|=\epsilon}(N\nabla\phi)(x,y)\bullet\boldsymbol{n}d\sigma_\epsilon,$$

where $d\sigma_t$ is the standard arclength measure on the circle $|(x,y)| = t$, and $\boldsymbol{n} = \boldsymbol{n}(x,y)$ denotes the outward-pointing (relative to the annulus) unit vector normal to the circle at the point (x,y). Since $\nabla\phi(x,y) = \mathbf{0}$ for $|(x,y)| \geq R$,

$$\int_{|(x,y)|=R}(N\nabla\phi)(x,y)\bullet\boldsymbol{n}d\sigma_R = 0.$$

On the other hand, we estimate the second integral as

$$\begin{aligned}
\Big|\int_{|(x,y)|=\epsilon}(N\nabla\phi)\bullet\boldsymbol{n}d\sigma_\epsilon\Big| &\leq \int_{|(x,y)|=\epsilon}|(N\nabla\phi)\bullet\boldsymbol{n}|d\sigma_\epsilon \\
&\leq \int_{|(x,y)|=\epsilon}|N(x,y)||\nabla\phi(x,y)|d\sigma_\epsilon \\
&\leq -\|\phi\|_1\ln(\epsilon)\epsilon \\
&\to 0 \text{ as } \epsilon\to 0+.
\end{aligned}$$

This proves that

$$\langle \Delta N, \phi \rangle = \lim_{\epsilon \to 0+} -\int_{R \geq |(x,y)| \geq \epsilon} \nabla \bullet (\phi \nabla N)(x,y) dV_2.$$

We again apply Green's Theorem and use the fact that $\mathrm{supp}(\phi) \subseteq \overline{B^n_R(\mathbf{0})}$ to see that

$$-\int_{R \geq |(x,y)| \geq \epsilon} \nabla \bullet (\phi \nabla N)(x,y) dV_2 = -\int_{|(x,y)|=\epsilon} \phi(x,y) \nabla N(x,y) \bullet \boldsymbol{n}(x,y) d\sigma_\epsilon.$$

Since $\boldsymbol{n}(x,y) = -\frac{1}{|(x,y)|}\left[\begin{smallmatrix} x \\ y \end{smallmatrix}\right]$ and since $\nabla N(x,y) = \frac{1}{2\pi |(x,y)|^2}\left[\begin{smallmatrix} x \\ y \end{smallmatrix}\right]$, this last integral simplifies to

$$\int_{|(x,y)|=\epsilon} \phi(x,y) \frac{1}{2\pi |(x,y)|} d\sigma_\epsilon = \frac{1}{2\pi\epsilon} \int_{|(x,y)|=\epsilon} \phi(x,y) d\sigma_\epsilon.$$

But since $\frac{1}{2\pi\epsilon}\int_{|(x,y)|=\epsilon} 1 d\sigma_\epsilon = 1$, we have

$$\begin{aligned}
\left| \frac{1}{2\pi\epsilon} \int_{|(x,y)|=\epsilon} \phi(x,y) d\sigma_\epsilon - \phi(0,0) \right| &= \left| \frac{1}{2\pi\epsilon} \int_{|(x,y)|=\epsilon} (\phi(x,y) - \phi(0,0)) d\sigma_\epsilon \right| \\
&\leq \frac{1}{2\pi\epsilon} \int_{|(x,y)|=\epsilon} |\phi(x,y) - \phi(0,0)| d\sigma_\epsilon \\
&\leq \left(\sup_{|(x,y)|=\epsilon} |\phi(x,y) - \phi(0,0)| \right) \frac{1}{2\pi\epsilon} \int_{|(x,y)|=\epsilon} 1 d\sigma_\epsilon \\
&= \sup_{|(x,y)|=\epsilon} |\phi(x,y) - \phi(0,0)| \to 0
\end{aligned}$$

as $\epsilon \to 0+$ by continuity of ϕ. We conclude that

$$\langle \Delta N, \phi \rangle = \lim_{\epsilon \to 0+} -\int_{R \geq |(x,y)| \geq \epsilon} \nabla \bullet (\phi \nabla N)(x,y) dV_2 = \phi(0,0) = \langle \delta, \phi \rangle.$$

Since $\phi \in \mathscr{D}(\mathbb{R}^2)$ was arbitrary, $\Delta N = \delta$ in the sense of distributions.

Since $\Delta N = \delta$, N is called a **fundamental solution** of the linear partial differential operator Δ. Later we'll see how to use a fundamental solution of Δ to solve the problem $\Delta u = f$ for $f \in \mathscr{D}(\mathbb{R}^2)$.

Exercise 21. Produce a fundamental solution for Laplace's operator $\Delta = \frac{d^2}{dx^2}$ on $\mathbb{R}$. That is, produce a locally integrable function $N : \mathbb{R} \to \mathbb{R}$ such that $\frac{d^2 N}{dx^2} = N'' = \delta$.

7 Regularization

We have seen examples of regular distributions and examples of singular distributions. One might wonder how 'wild' singular distributions really are. For example, the singular distribution δ is not too pathological since we can write $\lim\limits_{\epsilon \to 0+} \eta_{\epsilon,\mathbf{0}} = \delta$ in the sense of distributions, and therefore δ can be approximated (in the sense of distributions) by smooth functions. This is—perhaps surprisingly—a standard feature of distributions: *every* distribution is the (distributional) limit of a sequence of smooth functions. This is proved by **regularizing** or **mollifying** the distribution—both fancy words that mean 'smoothing out'. In this section we show how convolution allows us to do this.

Recall that the **convolution** of $\psi \in C^\infty(\mathbb{R}^n)$ and $\phi \in \mathscr{D}(\mathbb{R}^n)$ is given by

$$\psi * \phi : \mathbb{R}^n \to \mathbb{R}^n, \quad (\psi * \phi)(\boldsymbol{x}) = \int_{\mathbb{R}^n} \psi(\boldsymbol{x}-\boldsymbol{y})\phi(\boldsymbol{y})dV_n(\boldsymbol{y}) = \int_{\mathbb{R}^n} \psi(\boldsymbol{y})\phi(\boldsymbol{x}-\boldsymbol{y})dV_n(\boldsymbol{y}). \quad (1)$$

The elementary properties of the convolution are found in Exercises 7 and 8.

We desire a notion of convolution between a distribution and a test function. To this end, for $\psi \in \mathscr{D}(\mathbb{R}^n)$ we define $\tilde{\psi}(\boldsymbol{y}) = \psi(-\boldsymbol{y})$. One can check that $\tilde{\psi} \in \mathscr{D}(\mathbb{R}^n)$ whenever $\psi \in \mathscr{D}(\mathbb{R}^n)$. Recall (from Exercise 15) that for $\psi \in \mathscr{D}(\mathbb{R}^n)$ and $\boldsymbol{x}_0 \in \mathbb{R}^n$ we define the translation $\tau_{\boldsymbol{x}_0}\psi(\boldsymbol{y}) = \psi(\boldsymbol{y}-\boldsymbol{x}_0)$. To generalize (1) to distributions, we make the following definition.

Definition 7.1. Let $L \in \mathscr{D}'(\mathbb{R}^n)$ and $\psi \in \mathscr{D}(\mathbb{R}^n)$. Define

$$L * \psi : \mathbb{R}^n \to \mathbb{R}, \quad (L * \psi)(\boldsymbol{x}) = \left\langle L, \tau_{\boldsymbol{x}}\tilde{\psi} \right\rangle.$$

Example 7.2. As a quick example, we show that $\delta * \psi = \psi$ for all $\psi \in \mathscr{D}(\mathbb{R}^n)$. To see this, we merely note that for $\boldsymbol{x} \in \mathbb{R}^n$,

$$(\delta * \psi)(\boldsymbol{x}) = \left\langle \delta, \tau_{\boldsymbol{x}}\tilde{\psi} \right\rangle = \tau_{\boldsymbol{x}}\tilde{\psi}(\boldsymbol{0}) = \tilde{\psi}(-\boldsymbol{x}) = \psi(\boldsymbol{x}).$$

Analogously to Exercise 7 we have the following result.

Proposition 7.3. *Let $L \in \mathscr{D}'(\mathbb{R}^n)$ and $\psi \in \mathscr{D}(\mathbb{R}^n)$. Then $L * \psi \in C^\infty(\mathbb{R}^n)$ and for every multi-index α we have*

$$\partial^\alpha(L * \psi) = (\partial^\alpha L) * \psi = L * (\partial^\alpha \psi).$$

Proof. By induction, it suffices to show that $L * \psi$ is continuous and that $\frac{\partial}{\partial x_i}(L * \psi)$ exists at each point and is given by the desired formulas.

For continuity, fix $\boldsymbol{x} \in \mathbb{R}^n$ and choose a compact subset K containing

$$\operatorname{supp}(\tau_{\boldsymbol{x}}\tilde{\psi}) + \overline{B_1^n(\boldsymbol{0})} = \boldsymbol{x} - \operatorname{supp}(\psi) + \overline{B_1^n(\boldsymbol{0})},$$

and choose $N \in \mathbb{N}$ and $C > 0$ such that $|\langle L, \phi\rangle| \le C\|\phi\|_N$ for all $\phi \in \mathscr{D}(\mathbb{R}^n)$ with $\operatorname{supp}(\phi) \subseteq K$. Note that for all $\boldsymbol{z} \in B_1^n(\boldsymbol{0})$, $\tau_{\boldsymbol{x}+\boldsymbol{z}}\tilde{\psi}(\bullet) \in \mathscr{D}(\mathbb{R}^n)$ and $\operatorname{supp}(\tau_{\boldsymbol{x}+\boldsymbol{z}}\tilde{\psi}(\bullet)) \subseteq K$. Then the Mean Value Theorem implies that

$$|(L*\psi)(\boldsymbol{x}+\boldsymbol{z}) - (L*\psi)(\boldsymbol{x})| = \left|\left\langle L, \tau_{\boldsymbol{x}+\boldsymbol{z}}\tilde{\psi} - \tau_{\boldsymbol{x}}\tilde{\psi} \right\rangle\right| \le C\|\tau_{\boldsymbol{x}+\boldsymbol{z}}\tilde{\psi} - \tau_{\boldsymbol{x}}\tilde{\psi}\|_N \le C\sqrt{n}\|\psi\|_{N+1}|\boldsymbol{z}|,$$

and therefore $\lim_{\boldsymbol{z}\to\boldsymbol{0}}(L*\psi)(\boldsymbol{x}+\boldsymbol{z}) = (L*\psi)(\boldsymbol{x})$, proving continuity.

For differentiability we start by noting that for $i \in \{1, \ldots, n\}$, $\boldsymbol{x} \in \mathbb{R}^n$, and $h \neq 0$,

$$\begin{aligned} \frac{(L*\psi)(\boldsymbol{x}+h\boldsymbol{e}_i) - (L*\psi)(\boldsymbol{x})}{h} &= \left\langle L, \frac{\tilde{\psi}(\bullet - (\boldsymbol{x}+h\boldsymbol{e}_i)) - \tilde{\psi}(\bullet - \boldsymbol{x})}{h} \right\rangle \\ &= \left\langle L, \frac{\tilde{\psi}(\bullet - h\boldsymbol{e}_i - \boldsymbol{x}) - \tilde{\psi}(\bullet - \boldsymbol{x})}{h} \right\rangle. \end{aligned}$$

But by Exercise 6 we have

$$\frac{\tilde{\psi}(\bullet - h\boldsymbol{e}_i - \boldsymbol{x}) - \tilde{\psi}(\bullet - \boldsymbol{x})}{h} \to -\frac{\partial(\tau_x\tilde{\psi})}{\partial y_i}(\bullet) = -\tau_x\frac{\partial\tilde{\psi}}{\partial y_i}(\bullet) = \tau_x\widetilde{\frac{\partial\psi}{\partial y_i}}(\bullet) \quad \text{in } \mathscr{D}(\mathbb{R}^n) \text{ as } h \to 0.$$

It follows that, as $h \to 0$,

$$\frac{(L * \psi)(\boldsymbol{x} + h\boldsymbol{e}_i) - (L * \psi)(\boldsymbol{x})}{h} \to \left\langle L, \tau_{\boldsymbol{x}}\widetilde{\frac{\partial\psi}{\partial y_i}}\right\rangle = -\left\langle L, \frac{\partial(\tau_{\boldsymbol{x}}\tilde{\psi})}{\partial y_i}\right\rangle = \left\langle \frac{\partial L}{\partial y_i}, \tau_{\boldsymbol{x}}\tilde{\psi}\right\rangle.$$

This shows that $L * \psi$ is differentiable with respect to x_i at each $\boldsymbol{x} \in \mathbb{R}^n$, and

$$\frac{\partial}{\partial x_i}(L * \psi)(\boldsymbol{x}) = \Big(\frac{\partial L}{\partial y_i} * \psi\Big)(\boldsymbol{x}) = \Big(L * \frac{\partial\psi}{\partial y_i}\Big)(\boldsymbol{x}).$$

□

We need one additional lemma.

Lemma 7.4. *Let $L \in \mathscr{D}'(\mathbb{R}^n)$ and fix $\psi, \phi \in \mathscr{D}(\mathbb{R}^n)$. Then*

$$(L * \psi) * \phi = L * (\psi * \phi).$$

Remark 7.5. We've already shown that, under the hypotheses of the lemma, $\psi * \phi \in \mathscr{D}(\mathbb{R}^n)$ and $(L * \psi) \in C^\infty(\mathbb{R}^n)$, and therefore all of the convolutions in the conclusion are defined.

Proof. Fix $\boldsymbol{x} \in \mathbb{R}^n$. By Exercise 7(d), the sequence of Riemann sums

$$\sum_{\boldsymbol{m}\in\mathbb{Z}^n} \psi(\bullet - h\boldsymbol{m})\phi(h\boldsymbol{m})h^n \to (\psi * \phi)(\bullet) \text{ in } \mathscr{D}(\mathbb{R}^n) \text{ as } h \to 0.$$

It follows that

$$\sum_{\boldsymbol{m}\in\mathbb{Z}^n} \psi(\boldsymbol{x} - \bullet - h\boldsymbol{m})\phi(h\boldsymbol{m})h^n \to (\psi * \phi)(\boldsymbol{x} - \bullet) \text{ in } \mathscr{D}(\mathbb{R}^n) \text{ as } h \to 0.$$

But then, using the fact that for each $h > 0$ the sum $\sum_{\boldsymbol{m}\in\mathbb{Z}^n} \psi(\boldsymbol{x} - \bullet - h\boldsymbol{m})\phi(h\boldsymbol{m})h^n$ consists of only finitely many nonzero terms,

$$\begin{aligned}
(L * (\psi * \phi))(\boldsymbol{x}) &= \left\langle L, \tau_{\boldsymbol{x}}\widetilde{(\psi * \phi)}\right\rangle \\
&= \langle L, (\psi * \phi)(\boldsymbol{x} - \bullet)\rangle \\
&= \lim_{h\to 0}\left\langle L, \sum_{\boldsymbol{m}\in\mathbb{Z}^n} \psi(\boldsymbol{x} - \bullet - h\boldsymbol{m})\phi(h\boldsymbol{m})h^n\right\rangle \\
&= \lim_{h\to 0}\sum_{\boldsymbol{m}\in\mathbb{Z}^n} \langle L, \psi(\boldsymbol{x} - \bullet - h\boldsymbol{m})\phi(h\boldsymbol{m})h^n\rangle \\
&= \lim_{h\to 0}\sum_{\boldsymbol{m}\in\mathbb{Z}^n} \langle L, \psi(\boldsymbol{x} - \bullet - h\boldsymbol{m})\rangle\, \phi(h\boldsymbol{m})h^n \\
&= \lim_{h\to 0}\sum_{\boldsymbol{m}\in\mathbb{Z}^n} \left\langle L, \tau_{\boldsymbol{x}-h\boldsymbol{m}}\tilde{\psi}(\bullet)\right\rangle \phi(h\boldsymbol{m})h^n \\
&= \lim_{h\to 0}\sum_{\boldsymbol{m}\in\mathbb{Z}^n} (L * \psi)(\boldsymbol{x} - h\boldsymbol{m})\phi(h\boldsymbol{m})h^n \\
&= \int_{\mathbb{R}^n} (L * \psi)(\boldsymbol{x} - \boldsymbol{y})\phi(\boldsymbol{y})dV_n(\boldsymbol{y}) \\
&= ((L * \psi) * \phi)(\boldsymbol{x}).
\end{aligned}$$

□

Lemma 7.4 and several previous results can now be used to quickly prove that every $L \in \mathscr{D}'(\mathbb{R}^n)$ is the distributional limit of a sequence of smooth functions (in the sense of distributions).

Corollary 7.6. *Let $L \in \mathscr{D}'(\mathbb{R}^n)$. Then there exists a sequence (f_j) in $C^\infty(\mathbb{R}^n)$ with $f_j \to L$ in the sense of distributions.*

Proof. For each $j \in \mathbb{N}$, let $f_j = L * \eta_{1/j,\mathbf{0}}$. Since $\eta_{1/j,\mathbf{0}} \in \mathscr{D}(\mathbb{R}^n)$, $f_j \in C^\infty(\mathbb{R}^n)$. For fixed $\phi \in \mathscr{D}(\mathbb{R}^n)$, we have $\eta_{1/j,\mathbf{0}} * \phi \to \phi$ in $\mathscr{D}(\mathbb{R}^n)$ as $j \to \infty$. It follows that

$$\langle f_j, \phi \rangle = \left\langle L * \eta_{1/j,\mathbf{0}}, \phi \right\rangle = ((L * \eta_{1/j,\mathbf{0}}) * \tilde{\phi})(\mathbf{0}) = (L * (\eta_{1/j,\mathbf{0}} * \tilde{\phi}))(\mathbf{0}) = \left\langle L, \widetilde{\eta_{1/j,\mathbf{0}} * \tilde{\phi}} \right\rangle .$$

But

$$\widetilde{\eta_{1/j,\mathbf{0}} * \tilde{\phi}} = \tilde{\eta}_{1/j,\mathbf{0}} * \tilde{\tilde{\phi}} = \tilde{\eta}_{1/j,\mathbf{0}} * \phi = \eta_{1/j,\mathbf{0}} * \phi,$$

where the last equality follows from the fact that $\eta_{1/j,\mathbf{0}}$ is even. We therefore apply Exercise 8 to see that $\eta_{1/j,\mathbf{0}} * \phi \to \phi$ in $\mathscr{D}(\mathbb{R}^n)$ as $j \to \infty$, and therefore we have

$$\lim_{j\to\infty} \langle f_j, \phi \rangle = \lim_{j\to\infty} \left\langle L, \eta_{1/j,\mathbf{0}} * \phi \right\rangle = \langle L, \phi \rangle ,$$

which completes the proof of the corollary. $\square$

Exercise 22. Regularization gives us yet another way to motivate distributional derivatives. Let $L \in \mathscr{D}'(\mathbb{R}^n)$, and fix $i \in \{1, \ldots, n\}$. Choose functions $f_j \in C^\infty(\mathbb{R}^n)$ with $f_j \to L$ in the sense of distributions. Prove that $\frac{\partial f_j}{\partial x_i} \to \frac{\partial L}{\partial x_i}$ in the sense of distributions as well.

Our final exercise shows how to use distributions (in particular, fundamental solutions) in order to solve PDE.

Exercise 23. Let $f \in \mathscr{D}(\mathbb{R}^2)$. Prove that there exists $u \in C^\infty(\mathbb{R}^2)$ satisfying $\Delta u = f$.

References

[1] Gerald Folland, *Real Analysis: Modern Techniques and Their Applications*, 2nd edition, Wiley, 1999.

[2] Lars Hörmander, *The Analysis of Linear Partial Differential Operators I*, Springer-Verlag, 1983.

[3] Walter Rudin, *Functional Analysis*, 2nd edition, International Series in Pure and Applied Mathematics, McGraw-Hill, 1991.