B4.3 Distribution Theory

Jan Kristensen

Michaelmas Term 2022

Contents

1 Background

The history of the theory of distributions is closely connected with the theory of PDEs. Probably the first to use notions resembling that of distributions in mathematics were Fourier (1822), Kirchhoff (1882) and Heaviside (1898). More rigorous treatments followed, notably by Bochner (1932) who used, though implicitly, a notion of distribution in connection with his treatment of the Fourier transformation, and in their studies of the Cauchy problem, Hadamard (1932) and M. Riesz (1949) considered certain special distributions. The first to rigorously define distributions as linear functionals was Sobolev (1936). The closely related concept of weak derivatives, that arises naturally in the study of PDEs by variational methods, was also used by Friedrichs (1939). However, it is only in the final form of Schwartz (1945–50), where also the Fourier transformation is an essential part, that distribution theory has become such a convenient and efficient tool for the analysis of PDEs. This course and its sequel B4.4 *Fourier Analysis* give an introduction to these topics.

1.1 Why Distributions?

The classical calculus for functions of several variables is inadequate if one seeks a simple and general theory of PDEs. Borrowing an example from Hörmander (1963) we consider the two PDEs

$$\frac{\partial^2 u}{\partial x \partial y} = 0 \quad \text{and} \quad \frac{\partial^2 u}{\partial y \partial x} = 0$$

for the real-valued function $u = u(x, y)$ of two variables. The PDEs are equivalent for twice continuously differentiable functions u, but they are not equivalent for more general functions: the first PDE is satisfied by every function $u = h(x)$ that depends on x alone, whereas the second PDE does not make sense for such functions when $h(x)$ is not differentiable. This is somewhat unnatural and indicates the need for supplementing functions by new objects, distributions, so that differentiation is always possible and we get a better general notion of solution. In doing so it is important that we only add what is strictly necessary and that the new objects obey, as close as possible, the usual calculus rules. In order to motivate the formal definition we consider the PDE

$$\frac{\partial^2 u}{\partial x \partial y} = f \text{ in } \mathbb{R}^2 \tag{1}$$

where we assume that $f = f(x, y)$ is a given continuous function. Assume first that u is a twice continuously differentiable solution of (1) and let $\varphi \colon \mathbb{R}^2 \to \mathbb{R}$ be a twice continuously differentiable function vanishing outside a bounded set. If we multiply (1) by φ, next integrate over $\mathbb{R}^2$ and perform two integrations by parts on the left-hand side (note the boundary terms disappear because φ vanishes outside a bounded set) we arrive at

$$\iint u \frac{\partial^2 \varphi}{\partial y \partial x} \, \mathrm{d}x\mathrm{d}y = \iint f\varphi \, \mathrm{d}x\mathrm{d}y \tag{2}$$

Note that we can recover (1) again from (2) when, as above, u is twice continuously differentiable, so that for such functions (1) and (2) are in fact equivalent. The advantage of (2) is that it makes sense also when u is not twice continuously differentiable, for example it would suffice to assume that u is continuous (or merely locally integrable). It is also not difficult to see, that for a given u as above, the identity (2) cannot hold for more than one continuous function f (compare the *Fundamental Lemma of the Calculus of Variations* in Subsection 3.4 below). This makes it natural to define $\partial^2 u/\partial x \partial y = f$ in the *weak sense* if the identity (2) holds for all functions φ that are twice continuously differentiable and vanish off a bounded set. Since $\partial^2 \varphi/\partial x \partial y = \partial^2 \varphi/\partial y \partial x$ for twice continuously differentiable functions φ it is clear that the PDEs $\partial^2 u/\partial x \partial y = f$ and $\partial^2 u/\partial y \partial x = f$ then become equivalent in the weak sense. The theory of distributions goes a step further and considers the linear functional

$$\varphi \mapsto \iint u \frac{\partial^2 \varphi}{\partial x \partial y} \, \mathrm{d}x\mathrm{d}y$$

as a representation for $\partial^2 u/\partial x \partial y$ even when there is no continuous function f for which (2) holds. In order to be able to study PDEs of any order we are thus led to consider linear functionals on the set of functions vanishing outside bounded sets and having continuous derivatives of any order.

Perhaps it is still not clear from the above why we should bother to introduce objects, distributions, that can always be differentiated. Our next example is taken from Strichartz's book *A Guide to Distribution Theory and Fourier Transforms.*

1.1.1 One Dimensional Wave Equation

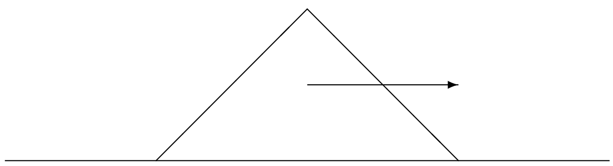
The equation

$$\frac{\partial^2 u}{\partial t^2}(x,t) = k^2 \frac{\partial^2 u}{\partial x^2}(x,t) \tag{3}$$

can be used to model a vibrating string. A function given by

$$u(x,t) = f(x - kt),$$

where f is a function of one variable, represents a travelling wave with shape $f(x)$ moving to the right with velocity k. When f is twice differentiable, one can check that u is a solution to (3). However, there is no physical reason for the shape of the travelling wave to be twice differentiable. For instance, the triangular profile



moving with speed k to the right is perfectly fine! We do not want to throw away physically meaningful solutions because of technicalities. Looking at the example above, one could think that if we accepted as solutions to differential equations any function that satisfies the differential equation except for some points (finitely many, say), where it fails to be differentiable, then all would be fine. But this would be a much too simplistic general principle, as the next example shows.

1.1.2 Laplace's Equation

In the plane $\mathbb{R}^2$ we have Laplace's equation

$$\Delta u := \frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} = 0. \tag{4}$$

A solution to the above equation has the physical interpretation of an electric potential in a region with no external charges. From physical experience it is known that such potentials should be smooth. However, as you may have seen last year,

$$u = G_0 := \frac{1}{4\pi} \log\left(x^2 + y^2\right),$$

is a solution in $\mathbb{R}^2 \setminus \{(0,0)\}$. Clearly it cannot be extended to the origin in a smooth manner, and so it should not be considered as a solution to (4) on the full plane.

Distribution theory allows us, among many other things, to distinguish between the case of the one dimensional wave equation (3) and Laplace's equation (4). Indeed, the standing wave satisfies the one dimensional wave equation in the sense of distributions for any continuous profile f, whereas

$$\Delta G_0 = \delta_0$$

as distributions, where δ_0 is Dirac's delta function, a distribution.

But there are in fact many other reasons to study distributions, and most of them are only really appreciated after the fact. Many physical quantities are naturally *not* defined pointwise. For instance, being able to measure temperature at a given point in space and time is an idealization – see the discussion in Strichartz's book *A Guide to Distribution Theory and Fourier Transforms*, §1. Similarly, in the theory of Lebesgue integration as discussed in the Part A Integration course you encountered L^p functions. Strictly speaking they are not functions, but equivalence classes of functions under the equivalence relation *equal almost everywhere*. Nonetheless, for $f \in \mathrm{L}^p(\mathbb{R}^n)$ and each measurable subset A of $\mathbb{R}^n$ the bracket

$$\langle f, \mathbf{1}_A \rangle := \int_A f(x)\,\mathrm{d}x \quad \text{ where now } x = (x_1\,,\,\dots\,,\,x_n) \tag{5}$$

is well-defined and does not depend on the particular representative used to calculate the integral. Note that if we know that f has a continuous representative, then we can uniquely determine the value of this continuous representative at all points $x \in \mathbb{R}^n$ from the values of the integrals (5) for all measurable subsets A of $\mathbb{R}^n$. In fact, we do not need the values in (5) for all measurable sets. For instance, it would suffice to know them for all open balls $B_r(x_0)$ since we have (denoting the continuous representative again by f) that

$$\frac{1}{\mathscr{L}^n(B_r(x_0))} \int_{B_r(x_0)} f(x)\,\mathrm{d}x \to f(x_0) \;\text{ as }\; r \searrow 0 \tag{6}$$

for all $x_0 \in \mathbb{R}^n$. On the other hand, for a general L^p function f, knowing the values of the integrals (5) for all measurable subsets A of $\mathbb{R}^n$ determines $f(x)$ uniquely almost everywhere, or more precisely, uniquely as an L^p function. (In fact, the assertion (6) remains true for almost all $x_0 \in \mathbb{R}^n$ when f is a general L^p function: the limit of the left-hand side exists in $\mathbb{R}$ for almost all $x_0 \in \mathbb{R}^n$ and defines a representative for the L^p function. This is a consequence of *Lebesgue's Differentiation Theorem.*) Note that here the indicator function of the set A,

$$\mathbf{1}_A(x) = \begin{cases} 1 & \text{if } x \in A, \\ 0 & \text{if } x \notin A \end{cases}$$

acts as a test function, or measurement of f, and instead of thinking about f as the basic object we could equally well consider the functional $\mathbf{1}_A \mapsto \langle f, \mathbf{1}_A \rangle$ to be the basic object. (*Terminology*: *Functional* means the same as *function*, but is often used instead if we have a real or complex-valued function defined on other functions.) It turns out that taking very nice test functions here is a good idea that allows us to extend aspects of differential calculus to L^p functions and beyond. This leads to the theory of distributions. Before getting there we must spend some time developing our notion of a test function. It is worth mentioning that there is not just one class of test functions. In this course and its sequel *Fourier Analysis* we shall systematically investigate two such classes: the compactly supported ones and the Schwartz ones. But for particular problems it is often the case that it is more natural to use other classes of test functions. To each class of test functions there corresponds a class of distributions. The principle to keep in mind here is that the nicer the test functions are, the rougher the corresponding distributions are allowed to be and vice versa. The general principle behind all of this is that of *duality*. You encountered it in a purely algebraic form already in *Linear Algebra* and if you follow the *Functional Analysis* courses this year you will see it again there.

1.2 A brief review of some Calculus

We start by fixing some notation. On $\mathbb{R}^d$ we shall usually employ the standard euclidean norm that is defined by the dot product:

$$|x| := \sqrt{x \cdot x} = \sqrt{x_1^2 + \ldots + x_d^2}.$$

Here the dimension d is understood from context and not emphasized in our notation. The open ball with centre x_0 and radius r is

$$B_r(x_0) := \bigl\{ x \in \mathbb{R}^d : \, |x - x_0| < r \bigr\}.$$

The corresponding closed ball is denoted with a bar on top, $\overline{B_r(x_0)}$. Similarly, the closure of any subset A of $\mathbb{R}^d$ is $\overline{A}$. Occasionally we shall also write $B_r(A)$ which is taken to mean the r-metric neighbourhood of the set A:

$$B_r(A) := \bigcup_{a \in A} B_r(a).$$

For two arbitrary subsets A and B of $\mathbb{R}^d$ we define the distance between them to be

$$\mathrm{dist}(A, B) := \inf \bigl\{ |a - b| : \, a \in A, \, b \in B \bigr\}.$$

We use the convention that the infimum of the empty set is $+\infty$ so that the distance between A and B is always defined, but possibly $+\infty$. We record in particular that if K is a compact subset of an open set V in $\mathbb{R}^d$, then the distance

$$\mathrm{dist}(K, \mathbb{R}^d \setminus V) := \inf \bigl\{ |x - y| : \, x \in K, \, y \in \mathbb{R}^d \setminus V \bigr\} > 0. \tag{7}$$

For two open subsets U and V of $\mathbb{R}^d$ we write $U \Subset V$ if $\overline{U}$ is compact and $\overline{U} \subset V$. Clearly we then also have that $\operatorname{dist}(U, \mathbb{R}^d \setminus V) > 0$ when $U \Subset V$.

When $A = \{x\}$ and B is a general subset we write $\operatorname{dist}(x, B) := \operatorname{dist}(\{x\}, B)$ and it is not difficult to check that the function $\mathbb{R}^d \ni x \mapsto \operatorname{dist}(x, B)$ is 1-Lipschitz: $|\operatorname{dist}(x, B) - \operatorname{dist}(y, B)| \leqslant |x - y|$ holds for all $x,\ y \in \mathbb{R}^d$.

1.2.1 Classical derivatives

Next we recall that a function $f \colon S \to \mathbb{R}$ defined on a subset S of $\mathbb{R}$ is said to be differentiable at a point x_0 if x_0 is an *interior point* of S and the difference quotient

$$\frac{f(x) - f(x_0)}{x - x_0},\ x \in S \setminus \{x_0\},$$

has a limit in $\mathbb{R}$ as $x \to x_0$. Of course this limit is the differential quotient of f at x_0, denoted as usual by $f'(x_0)$. The above generalizes in a straight forward manner to the case where the function f is complex valued, so $f \colon S \to \mathbb{C}$, and to the case where it is $\mathbb{R}^k$-valued, so $f \colon S \to \mathbb{R}^k$. In all cases we retain the notation $f'(x_0)$ for the differential quotient when it exists.

It is well-known that we can relate the value of a differentiable function $f \colon (a, b) \to \mathbb{R}$ to its derivative via the Mean Value Theorem. If f is continuously differentiable, then we even have that

$$f(x) = f(x_0) + \int_{x_0}^{x} f'(t)\,\mathrm{d}t$$

for all $x,\ x_0 \in (a, b)$ by the Fundamental Theorem of Calculus. While the latter remains true also for vector-valued functions, the Mean Value Theorem breaks down in that case (for instance try $f(x) = \mathrm{e}^{\mathrm{i}x}$, $x \in \mathbb{R}$, between 0 and 2π). The following weaker result can then sometimes be used instead.

Proposition 1.1. *(The Mean Value Inequality)*
Let $I \subseteq \mathbb{R}$ be an open interval and assume that $f \colon I \to \mathbb{R}^d$ is differentiable. Then

$$|f(y) - f(x)| \leqslant |y - x| \sup_{t \in (0,1)} \bigl|f'\bigl(x + t(y - x)\bigr)\bigr|$$

holds for all $x,\ y \in I$.

Proof. There is nothing to prove if the supremum on the right-hand side is $+\infty$ (we use the convention that the supremum of a set that is not bounded above is $+\infty$). Assume therefore that it is finite and fix

$$M > \sup_{t \in (0,1)} \bigl|f'\bigl(x + t(y - x)\bigr)\bigr|. \tag{8}$$

Now to prove the desired inequality it suffices to do it for the case where $x, y \in I$ satisfy $x > y$. We fix such a pair and define the set

$$E = \big\{ t \in [0,1] : |f(x+t(y-x)) - f(x)| \leqslant Mt|x-y| \big\}.$$

Clearly $0 \in E$ and because f is continuous the set E must be closed relative to $[0,1]$. It follows that it has a largest element, say $s = \max E$. Because I is open and $x + s(y-x) \in I$ we have by differentiability that for $t > s$ with $t - s$ sufficiently small,

$$|f(x+t(y-x)) - f(x+s(y-x))| \leqslant M(t-s)|y-x|.$$

Consequently we find

$$\begin{aligned} |f(x+t(y-x)) - f(x)| &\leqslant |f(x+t(y-x)) - f(x+s(y-x))| + |f(x+s(y-x)) - f(x)| \\ &\leqslant M(t-s)|y-x| + Ms|y-x| \\ &= Mt|y-x|, \end{aligned}$$

hence $s = 1$. Because M was arbitrary in (8) the proof is complete. $\square$

Corollary 1.2. *Let $I \subseteq \mathbb{R}$ be an open interval and C a closed subset of I. Suppose that $f \colon I \to \mathbb{R}^d$ is continuous, differentiable on $I \setminus C$ and $f(x) = 0$ for $x \in C$. If $x \in C$ and*

$$f'(y) \to 0 \ \text{ as } \ I \setminus C \ni y \to x, \tag{9}$$

then $f'(x)$ exists and equals 0.

Proof. If $y \in C$, then $f(y) - f(x) = 0$ and all is fine. We then consider the case $y \in I \setminus C$ and can assume that $y > x$ as the situation when $y < x$ is entirely similar. Fix such y and let z be the point in $C \cap [x,y]$ that is closest to y (why does it exist?). Then as f is differentiable on the interval (z,y) we get from the Mean Value Inequality

$$|f(y) - f(x)| = |f(y) - f(z)| \leqslant |y-z| \sup_{t \in (0,1)} |f'(z+t(y-z))|$$

and so

$$\frac{|f(y) - f(x)|}{|y-x|} \leqslant \sup_{t \in (0,1)} |f'(z+t(y-z))|.$$

Observe that $|z + t(y-z) - x| \leqslant |y-x|$ for all $t \in (0,1)$ so using the assumption (9), given $\varepsilon > 0$ we find $\delta > 0$ with the property that

$$|f'(a)| < \varepsilon \ \text{ for all } a \in (x, x+\delta) \cap I \setminus C$$

and the conclusion follows. $\square$

Example 1.3. Let $P(x) \in \mathbb{C}[x]$ be a polynomial and define the function

$$f(x) = \begin{cases} P\big(\frac{1}{x}\big)\mathrm{e}^{-\frac{1}{x}} & \text{if } x > 0 \\ 0 & \text{if } x \leqslant 0. \end{cases}$$

Then f is easily seen to be continuous everywhere and differentiable for $x \neq 0$ with

$$f'(x) = \begin{cases} \mathrm{e}^{-\frac{1}{x}}\big(P\big(\frac{1}{x}\big) - P'\big(\frac{1}{x}\big)\big)/x^2 & \text{if } x > 0 \\ 0 & \text{if } x < 0. \end{cases}$$

This is of course just the same form as that of f and so we conclude from Corollary 1.2 (or simply by use of the definition) that $f'(0)$ exists and equals 0. An easy induction argument now shows that f is infinitely often differentiable.

For functions of several variables we have partial derivatives and directional derivatives. Let $\{e_j\}_{j=1}^n$ be the standard basis for $\mathbb{R}^n$ and assume that $f\colon S \to \mathbb{R}$ is a real-valued function (or it could be complex or vector valued), where now S is a subset of $\mathbb{R}^n$. We then say that f has a partial derivative with respect to x_j at x_0 if x_0 is an interior point of S and the difference quotient

$$\frac{f(x_0 + he_j) - f(x_0)}{h}, \quad h \in \mathbb{R} \setminus \{0\} \text{ and } x_0 + he_j \in S$$

has a limit in $\mathbb{R}$ as $h \to 0$. This limit is the partial derivative at x_0 that we shall denote by various different symbols, including

$$\partial_j f(x_0) = \partial_{x_j} f(x_0) = D_j f(x_0) = \frac{\partial f}{\partial x_j}(x_0) = f_{x_j}(x_0) = \ldots\ldots$$

A slight variation of the above yields directional derivatives: replace e_j by a general vector $v \in \mathbb{R}^n \setminus \{0\}$ to define $\partial_v f(x_0)$. When f has all first order partial derivatives at the point x_0 we can collect them in a vector:

$$\nabla f(x_0) := \big(\partial_1 f(x_0)\ \ldots\ , \partial_n f(x_0)\big) = \sum_{j=1}^n \partial_j f(x_0) e_j.$$

This is the *gradient* of f at x_0. When f is $\mathbb{R}^d$-valued, say $f = (f_1, \ldots, f_d)^\dagger$ (we think of f as a column vector, hence the transpose) it is customary to collect the first order partial derivatives (when they exist) in a matrix called the *Jacobi matrix* for f at x_0:

$$\nabla f(x_0) = Df(x_0) := \big[\partial_1 f(x_0) \ldots \partial_n f(x_0)\big] = \begin{bmatrix} \partial_1 f_1(x_0) \ldots \partial_n f_1(x_0) \\ \cdots\cdots\cdots\cdots\cdots\cdots \\ \cdots\cdots\cdots\cdots\cdots\cdots \\ \partial_1 f_d(x_0) \ldots \partial_n f_d(x_0) \end{bmatrix} \in \mathbb{R}^{d\times n}.$$

It is perfectly possible for a function, say $f\colon \mathbb{R}^2 \to \mathbb{R}$, to have directional derivatives in *all* directions at $(0,0)$ and at the same time be discontinuous there. It can also happen that $\partial_v f(0,0) \neq \nabla f(0,0) \cdot v$ for some vectors v. However, these pathological situations are excluded if the function is *continuously differentiable*, meaning that $\partial_1 f, \ldots, \partial_n f$ all exist and are continuous. More precisely we have the following:

Lemma 1.4. *Let S be a subset of $\mathbb{R}^n$ and let x_0 be an interior point of S. Assume the function $f\colon S \to \mathbb{R}^d$ has partial derivatives $\partial_1 f(x), \ldots, \partial_n f(x)$ for all x in a neighbourhood of x_0 and that they are all continuous at x_0, then f is continuous at x_0 and we have that $\partial_v f(x_0) = \nabla f(x_0)v$ for all $v \in \mathbb{R}^n \setminus \{0\}$.*

The proof is an exercise.

The higher order partial derivatives are defined inductively. For instance, if we say that the mixed partial derivative $\partial^2 f/\partial x_j \partial x_k$ exists at the point x_0 for the function $f\colon S \to \mathbb{R}^d$, then it means that x_0 is an interior point of S, that $\partial f/\partial x_k$ exists in a neighbourhood of x_0 and has a partial derivative with respect to x_j at x_0. Note that in this generality it is important that we pay attention to the order in which we partially differentiate. We are actually not interested in such situations and shall mostly be working with classes of functions where this is no issue.

Definition 1.5. *Let Ω be a non-empty open subset of $\mathbb{R}^n$ and let $k \in \mathbb{N}$. Then a function $f\colon \Omega \to \mathbb{R}$ (or complex or vector-valued) is said to be k times continuous differentiable if all partial derivatives of f up to and including order k exist and are continuous throughout Ω.*

Example 1.6. The function f is continuously differentiable if f is continuous and all the partial derivatives $\partial f/\partial x_1, \ldots, \partial f/\partial x_n$ exist and are continuous on Ω. Note in particular that we require, for example, $\partial f/\partial x_1$ to be jointly continuous in $x = (x_1, x_2, \ldots, x_n) \in \Omega$.

Likewise, f is twice continuously differentiable if f is continuous and all the partial derivatives $\partial f/\partial x_i$, $i = 1, \ldots, n$, $\partial^2 f/\partial x_j \partial x_k$, $j, k = 1, \ldots, n$ exist and are continuous on Ω.

In this connection the following notation is standard.

$$\mathrm{C}(\Omega) \coloneqq \{u\colon \Omega \to \mathbb{R} \,:\, u \text{ is continuous}\}.$$

Similarly, for $k \in \mathbb{N}$ we define

$$\mathrm{C}^k(\Omega) \coloneqq \{u : \Omega \to \mathbb{R} \,:\, u \text{ is } k \text{ times continuously differentiable}\}.$$

We say that such functions are C^k functions. When $k = 0$ we write also $\mathrm{C}^0(\Omega) = \mathrm{C}(\Omega)$.

Note that $\mathrm{C}^k(\Omega)$ is descending in k, $\mathrm{C}^{k+1}(\Omega) \subsetneq \mathrm{C}^k(\Omega)$. We define

$$\mathrm{C}^\infty(\Omega) \coloneqq \bigcap_{k=0}^{\infty} \mathrm{C}^k(\Omega),$$

the class of infinitely differentiable functions on Ω. We remark that under the natural pointwise definitions of addition, multiplication by scalars, and multiplication, these classes all form commutative rings with unity and vector spaces.

The same notation will be used for complex and vector-valued functions (where of course in the vector-valued case the spaces cease to be rings).

C^k functions are well-behaved:

Lemma 1.7. *If $f\colon \Omega \to \mathbb{R}$ is* C^2, *then*

$$\frac{\partial^2 f}{\partial x_j \partial x_k} = \frac{\partial^2 f}{\partial x_k \partial x_j} \quad (1 \leqslant j,\, k \leqslant n)$$

on Ω. Hence the order in which we take (two) partial derivatives is unimportant for C^2 *functions.*

Proof. Let $\{e_j\}_{j=1}^n$ be the standard basis for $\mathbb{R}^n$ and denote by

$$\triangle_h f(x) := f(x+h) - f(x)$$

the increment in f at x corresponding to the increment h in x, where we must assume $x,\, x+h \in \Omega$. Observe that $\triangle_{se_j}\triangle_{re_k} f(x) = \triangle_{re_k}\triangle_{se_j} f(x)$ holds for $x \in \Omega$ and $s,\, r \in \mathbb{R}$ with $|s|,\, |r|$ sufficiently small. Because f is C^2 we may apply the Fundamental Theorem of Calculus twice whereby we find

$$\frac{1}{sr}\triangle_{se_j}\triangle_{re_k} f(x) = \int_0^1\int_0^1 \frac{\partial^2 f}{\partial x_j \partial x_k}(x + \sigma s e_j + \rho r e_k)\,\mathrm{d}\sigma\,\mathrm{d}\rho,$$

and hence

$$\left|\frac{1}{sr}\triangle_{se_j}\triangle_{re_k} f(x) - \frac{\partial^2 f}{\partial x_j \partial x_k}(x)\right| \to 0$$

as $(r, s) \to (0, 0)$. □

We can extend this result to C^k functions and arbitrary k-th order partial derivatives for $k \geqslant 2$ by induction, and so for such functions we do not have to worry about the order in which we partially differentiate. When there are many independent variables we shall often rely on multi-index notation.

1.2.2 Multi-index Notation

A multi-index α is an (ordered) n-tuple of non-negative integers, $\alpha = (\alpha_1, \ldots, \alpha_n) \in \mathbb{N}_0^n$. The length (or *order*) of α is

$$|\alpha| := \alpha_1 + \cdots + \alpha_n.$$

If $\alpha,\, \beta \in \mathbb{N}_0^n$ and $j \in \mathbb{N}_0$, then also $\alpha + j\beta \in \mathbb{N}_0^n$. For a multi-index $\alpha \in \mathbb{N}_0^n$ we define its factorial as

$$\alpha! := \alpha_1!\alpha_2! \ldots \alpha_n!$$

and for an n-tuple $x = (x_1,\, \ldots\,,\, x_n)$ of real or complex numbers we write

$$x^\alpha := x_1^{\alpha_1} \cdot\, \ldots\, \cdot x_n^{\alpha_n}.$$

Hence a complex polynomial in n indeterminates $x = (x_1, \dots, x_n)$, say $P(x) \in \mathbb{C}[x]$ can be written as

$$P(x) = \sum_{|\alpha| \leqslant m} c_\alpha x^\alpha$$

where $c_\alpha \in \mathbb{C}$ and $m \in \mathbb{N}_0$. When $u \in \mathrm{C}^k(\Omega)$ and $|\alpha| \leqslant k$, we write

$$\partial^\alpha u(x) = D^\alpha u(x) := \frac{\partial^{|\alpha|} u}{\partial x_1^{\alpha_1} \dots \partial x_n^{\alpha_n}}(x)$$

and by convention set $\partial^0 u(x) = D^0 u(x) := u(x)$.

Example 1.8. For $\alpha = (1, 2)$, $\beta = (0, 2)$ and $u \in \mathrm{C}^3(\Omega)$, where $\Omega \subset \mathbb{R}^2$, we have

$$\partial^\alpha u = \frac{\partial^3 u}{\partial x_1 \partial x_2^2}, \qquad \partial^\beta u = \frac{\partial^2 u}{\partial x_2^2}.$$

Likewise, when $u \in \mathrm{C}^5(\Omega)$,

$$\partial^{\alpha+\beta} u = \frac{\partial^5 u}{\partial x_1 \partial x_2^4}.$$

Note that by lemma 1.7, $\partial^{\alpha+\beta} u = \partial^\alpha(\partial^\beta u) = \partial^\beta(\partial^\alpha u) = \partial^{\beta+\alpha} u$. In a sense, lemma 1.7 justifies using multi-index notation for partial derivatives.

A convenient fact about multi-index notation is that it makes many calculus formulas for functions of several variables appear as when there is only one variable. Here we record two such instances.

Theorem 1.9 (Taylor's Formula)**.** *Assume $f \in \mathrm{C}^k(B_r(x_0))$. Then for $x \in B_r(x_0)$ we have*

$$f(x) = \sum_{|\alpha|<k} \frac{\partial^\alpha f(x_0)}{\alpha!}(x - x_0)^\alpha + k \int_0^1 (1-t)^{k-1} \sum_{|\alpha|=k} \frac{\partial^\alpha f\big(x_0 + t(x - x_0)\big)}{\alpha!}(x - x_0)^\alpha \, \mathrm{d}t.$$

The proof follows by applying, for a fixed $x \in B_r(x_0)$, the Chain Rule and the 1-dimensional Taylor formula to the function $s \mapsto f(x_0 + s(x - x_0))$. We record the following application that is often useful:

Corollary 1.10. *If $f \in \mathrm{C}^k(B_r(x_0))$ for some $k \in \mathbb{N}$, then $f(x) - f(x_0) = \sum_{j=1}^n (x - x_0)_j f_j(x)$, where $f_j \in \mathrm{C}^{k-1}(B_r(x_0))$, $\partial^\alpha f_j(x_0) = \partial^\alpha \partial_j f(x_0)/(1 + |\alpha|)$ and*

$$\sup_{B_r(x_0)} |\partial^\alpha f_j| \leqslant \sup_{B_r(x_0)} |\partial^\alpha \partial_j f|$$

for each $|\alpha| < k$.

Proof. We use the formula for $k = 1$, which actually amounts to the Fundamental Theorem of Calculus, whereby the Corollary is seen to hold with

$$f_j(x) = \int_0^1 (\partial_j f)\big(x_0 + t(x - x_0)\big)\,\mathrm{d}t.$$

The assertions about f_j all follow by inspection. $\square$

Theorem 1.11 (Generalized Leibniz Rule)**.** *Let $f, g \in \mathrm{C}^k(\Omega)$. Then $fg \in \mathrm{C}^k(\Omega)$ and for $\alpha \in \mathbb{N}_0^n$, $|\alpha| \leqslant k$, we have*

$$\partial^\alpha(fg) = \sum_{\beta \leqslant \alpha} \binom{\alpha}{\beta} \partial^\beta f \partial^{\alpha-\beta} g,$$

where $\beta \leqslant \alpha$ means $\beta_i \leqslant \alpha_i$ for all $i = 1, \ldots n$,

$$\binom{\alpha}{\beta} := \frac{\alpha!}{\beta!(\alpha - \beta)!}.$$

Proof. This can be proven by induction on the order $|\alpha|$ of differentiation, using the Leibniz rule

$$\partial_j(fg) = g\partial_j f + f\partial_j g$$

in the induction step. $\square$

Remark 1.12. The set $\{\alpha \in \mathbb{N}_0^n : |\alpha| = k\}$ of multi-indices of length $k \in \mathbb{N}_0$ is clearly a finite set. It is not too difficult to count its elements and show that the cardinality is

$$\mathbf{m}(n, k) = \binom{n + k - 1}{k}$$

and consequently that the cardinality of the set $\{\alpha \in \mathbb{N}_0^n : |\alpha| \leqslant k\}$ of multi-indices of length at most $k \in \mathbb{N}_0$ is

$$\mathbf{M}(n, k) = \binom{n + k}{k}.$$

2 Test functions

2.1 Support of a continuous function

For $u \in \mathrm{C}(\Omega)$ we define the support of u as

$$\operatorname{supp}(u) := \Omega \cap \overline{\{x \in \Omega \ : \ u(x) \neq 0\}},$$

that is, the closure of the set $\{u \neq 0\}$ relative to Ω. As such, $\operatorname{supp}(u)$ is closed in Ω, but need not be closed in $\mathbb{R}^n$.

Example 2.1. Define $u_1 \colon \mathbb{R} \to \mathbb{R}$ by

$$u_1(x) = \begin{cases} 1 - |x|, & |x| < 1, \\ 0, & |x| \geqslant 1. \end{cases}$$

Then $\operatorname{supp}(u_1) = [-1, 1]$. If instead we consider the restriction of u_1 to $\Omega = (-1, 1)$, that is, $u_2(x) = 1 - |x|$, $x \in (-1, 1)$, then $\operatorname{supp}(u_2) = (-1, 1)$.

One sees that the support of a function u depends on the domain Ω, and we could emphasize this and instead write $\operatorname{supp}_\Omega(u)$. However, for our purposes it will suffice to write $\operatorname{supp}(u)$, where Ω will be understood from context.

In the following we shall be particularly interested in having compact support.

Definition 2.2. *Let Ω be a non-empty open subset of $\mathbb{R}^n$. Then*

$$\mathscr{D}(\Omega) := \big\{ u \in \mathrm{C}^\infty(\Omega) \ : \ \operatorname{supp}(u) \textit{ is compact} \big\}$$

is the class of smooth compactly supported test functions.

Remark 2.3. Note that for $u \in \mathscr{D}(\Omega)$ we always have $\operatorname{dist}(\operatorname{supp}(u), \partial\Omega) > 0$, see (7).

We also write

$$\mathrm{C}_c^k(\Omega) := \{ u \in \mathrm{C}^k(\Omega) \ : \ \operatorname{supp}(u) \text{ is compact} \}$$

for $k \in \mathbb{N}_0 \cup \{\infty\}$. So in fact $\mathscr{D}(\Omega) = \mathrm{C}_c^\infty(\Omega)$. As before, we can define ring operations in the standard way, making $\mathrm{C}_c^k(\Omega)$ and $\mathscr{D}(\Omega)$ into commutative rings (*without* unity) and vector spaces (over $\mathbb{R}$ or $\mathbb{C}$).

We have defined a *test function* to be any smooth and compactly supported function, but so far we have seen no example. If we take the polynomial $P(x) = 1$ in Example 1.3, then we get the C^∞ function $f(x) = \mathrm{e}^{-1/x}$ for $x > 0$ and $f(x) = 0$ for $x \leqslant 0$. Now put $\mathcal{B}(x) = f\big(1 - |x|^2\big)$, $x \in \mathbb{R}^n$. Clearly $\mathcal{B}$ is C^∞ by the chain rule and its support is $\overline{B_1(0)}$, so $\mathcal{B}$ is a test function on $\mathbb{R}^n$.

2.2 Construction of test functions.

For later reference we record our first non-trivial test function:

Lemma 2.4. *The function*

$$\mathcal{B}(x) = \begin{cases} \exp\left(\frac{1}{|x|^2 - 1}\right), & |x| < 1, \\ 0, & |x| \geqslant 1 \end{cases}$$

is in $\mathrm{C}^\infty(\mathbb{R}^n)$ *with* $\operatorname{supp}(\mathcal{B}) = \overline{B_1(0)}$. *In particular,* $\mathcal{B} \in \mathscr{D}(\mathbb{R}^n)$.

Note that

$$\mathcal{B} \geqslant 0 \text{ and } 0 < \mathcal{B}(x) \leqslant \mathcal{B}(0) = \tfrac{1}{\mathrm{e}} \text{ for } |x| < 1.$$

For this reason we sometimes refer to $\mathcal{B}$ as a bump function. We also record that $\mathcal{B}$ is a *radial function*, meaning that its value $\mathcal{B}(x)$ at x only depends on $|x|$.

We can now easily produce more bump functions:

Example 2.5. We want a bump in $B_r(x_0)$ and put

$$\varphi(x) := \mathcal{B}\left(\frac{x - x_0}{r}\right), \quad x \in \mathbb{R}^n.$$

By the Chain Rule, we see that $\varphi \in \mathrm{C}^\infty(\mathbb{R}^n)$. Clearly, $\mathrm{supp}(\varphi) = \overline{B_r(x_0)}$, and so $\varphi \in \mathscr{D}(\mathbb{R}^n)$.

It is remarkable that once we have just one nontrivial bump function, then we can construct all the test functions we need to get Distribution Theory going. In order to make more refined constructions, still using the bump $\mathcal{B}$ from Lemma 2.4 as a building block, we shall use the operation of *convolution* that you encountered in Integration.

2.2.1 Convolution of functions

Recall that when $f,\, g \in \mathrm{L}^1(\mathbb{R}^n)$, then

$$(f * g)(x) := \int_{\mathbb{R}^n} f(x-y)g(y)\,\mathrm{d}y \tag{10}$$

is well-defined for almost every $x \in \mathbb{R}^n$, and $f * g \in \mathrm{L}^1(\mathbb{R}^n)$. This follows from the Fubini-Tonelli theorems. Let us more precisely recall how it goes. First we choose representatives, again denoted by f and g. Then the product σ-algebra was defined so that the function $(x,y) \mapsto \bigl|f(x-y)g(y)\bigr|$ is measurable and by Tonelli's theorem we may then calculate (note that we also use that the Lebesgue measure is translation invariant)

$$\begin{aligned}\int_{\mathbb{R}^n \times \mathbb{R}^n} \bigl|f(x-y)g(y)\bigr|\,\mathrm{d}(x,y) &= \int_{\mathbb{R}^n}\int_{\mathbb{R}^n} \bigl|f(x-y)g(y)\bigr|\,\mathrm{d}x\,\mathrm{d}y \\ &= \|f\|_1 \|g\|_1 < \infty.\end{aligned}$$

At this stage we can then use Fubini's theorem. Accordingly the function $y \mapsto f(x-y)g(y)$ is for almost all $x \in \mathbb{R}^n$ integrable and the integral

$$x \mapsto \int_{\mathbb{R}^n} f(x-y)g(y)\,\mathrm{d}y \tag{11}$$

is defined almost everywhere and (assigning arbitrary values in the points of the null set where it is not defined) is an integrable function. We emphasize that the integral in (11) is defined at $x \in \mathbb{R}^n$ precisely when

$$\int_{\mathbb{R}^n} \bigl|f(x-y)g(y)\bigr|\,\mathrm{d}y < \infty$$

and that this condition is independent of the chosen representatives used to calculate the integral. We therefore see that the set of points $x \in \mathbb{R}^n$ where the convolution (10) is defined and its value only depend on the L^1 functions f and g, and not on the particular representatives used to calculate it. In this connection we also record that (again using the translation invariance of Lebesgue measure)

$$\int_{\mathbb{R}^n} \big| f(x-y)g(y) \big| \, \mathrm{d}y = \int_{\mathbb{R}^n} \big| f(z)g(-z+x) \big| \, \mathrm{d}z$$

holds for *all* $x \in \mathbb{R}^n$, in the sense that either both sides are defined and equal at x or both are undefined there. Since addition is commutative on $\mathbb{R}^n$ (so $-z+x = x-z$ above) we conclude that also $(g*f)(x) = (f*g)(x)$ holds for all $x \in \mathbb{R}^n$ in the sense that either both sides are defined and equal at x or both are undefined there.

There are many other situations where it is possible to define convolution by the formula (10). One such instance is when one function is continuous and compactly supported and the other is an L^p function for some $p \in [1, \infty]$ (the case $p = 1$ is of course already covered by the remark above about convolutions of L^1 functions). We shall explore this and other possibilities in the next sections.

Example 2.6. If $f \in \mathrm{L}^p$, $g \in \mathrm{L}^q$ and $\frac{1}{p} + \frac{1}{q} = 1$, then the convolution $(f*g)(x)$ is well-defined for *all* $x \in \mathbb{R}^n$ since by Hölder's inequality and translation-invariance of Lebesgue measure we have

$$\int_{\mathbb{R}^n} \big| f(x-y)g(y) \big| \, \mathrm{d}y \leqslant \|f\|_p \|g\|_q < +\infty$$

In fact, this situation is special and more can be said about the convolution of f and g in this case. See Problem Sheet 1.

Exercise. Let $f \in \mathrm{C}^j(\mathbb{R})$, $g \in \mathrm{C}^k_c(\mathbb{R})$, where j, $k \in \mathbb{N}_0$. Show that the convolution $f*g$ is well-defined everywhere and that it is of class C^{j+k}. Next show that if f is also compactly supported, then

$$\operatorname{supp}(f*g) \subseteq \operatorname{supp}(f) + \operatorname{supp}(g)$$

holds. Generalize the results of this exercise to n dimensions.

2.2.2 The Standard Mollifier in $\mathbb{R}^n$.

Notice that for all $x \in \mathbb{R}^n$,

$$\mathrm{e}^{-\frac{4}{3}} \mathbf{1}_{B_{1/2}(0)}(x) \leqslant \mathcal{B}(x) \leqslant \mathrm{e}^{-1} \mathbf{1}_{B_1(0)}(x),$$

so we obviously have that

$$c_n := \int_{\mathbb{R}^n} \mathcal{B}(x) \, \mathrm{d}x$$

is well-defined and $\mathrm{e}^{-\frac{4}{3}}\mathscr{L}^n(B_{1/2}(0)) \leqslant c_n \leqslant \mathrm{e}^{-1}\mathscr{L}^n(B_1(0))$. Here we record that one can show that the n-dimensional Lebesgue measure of a ball of radius $r > 0$ is

$$\mathscr{L}^n\big(B_r(0)\big) = \mathscr{L}^n\big(B_1(0)\big)r^n = \frac{\omega_{n-1}}{n}r^n = \frac{2\pi^{\frac{n}{2}}}{n\Gamma\left(\frac{n}{2}\right)},$$

where ω_{n-1} is the $(n-1)$-dimensional surface area of $\mathbb{S}^{n-1}$ and Γ is Euler's gamma function. Note in particular that $\omega_0 = 2$, $\omega_1 = 2\pi$ and $\omega_2 = 4\pi$ in dimensions one, two and three, respectively.

Define

$$\rho(x) := \frac{1}{c_n}\mathcal{B}(x), \quad x \in \mathbb{R}^n. \tag{12}$$

We refer to ρ as the *standard mollifier kernel* and record the properties: $\rho \in \mathscr{D}(\mathbb{R}^n)$, $\rho \geqslant 0$, $\operatorname{supp}(\rho) = \overline{B_1(0)}$ and

$$\int_{\mathbb{R}^n} \rho(x)\,\mathrm{d}x = 1.$$

For each $\varepsilon > 0$ we put

$$\rho_\varepsilon(x) := \frac{1}{\varepsilon^n}\rho\left(\frac{x}{\varepsilon}\right), \quad x \in \mathbb{R}^n.$$

Then $\rho_\varepsilon \in \mathscr{D}(\mathbb{R}^n)$, $\rho_\varepsilon \geqslant 0$, $\operatorname{supp}(\rho_\varepsilon) = \overline{B_\varepsilon(0)}$ and

$$\int_{\mathbb{R}^n} \rho_\varepsilon(x)\,\mathrm{d}x = 1.$$

Definition 2.7. *We call the family of functions $(\rho_\varepsilon)_{\varepsilon>0}$* the standard mollifier *on $\mathbb{R}^n$.*

We also emphasize that the functions ρ_ε are all radial, meaning that the value $\rho_\varepsilon(x)$ only depends on $|x|$.

Proposition 2.8. *Let $1 \leqslant p < \infty$ and $u \in \mathrm{L}^p(\Omega)$. Define u to be zero outside Ω. Then*

(i) $\rho_\varepsilon * u \in \mathrm{C}^\infty(\Omega)$*,*

(ii) $\|\rho_\varepsilon * u\|_p \leqslant \|u\|_p$*, and*

(iii) $\|\rho_\varepsilon * u - u\|_p \to 0$ *as* $\varepsilon \searrow 0$*.*

We require the following auxiliary results for the proof.

Lemma 2.9. *Let $1 \leqslant p \leqslant \infty$, $\varphi \in \mathscr{D}(\Omega)$, and $u \in \mathrm{L}^p(\Omega)$. Define u to be zero outside of Ω. Then $\varphi * u \in \mathrm{C}^1(\Omega)$ and for each $1 \leqslant j \leqslant n$,*

$$\frac{\partial(\varphi * u)}{\partial x_j} = \left(\frac{\partial \varphi}{\partial x_j}\right) * u.$$

Proof. This is straight forward and we omit the details. $\square$

The next is an approximation result.

Lemma 2.10. *Let Ω be an open nonempty subset of $\mathbb{R}^n$ and $1 \leqslant p < \infty$. Then $\mathrm{C}_c(\Omega)$ is dense in $\mathrm{L}^p(\Omega)$.*

The proof is elementary but a little technical.

Proof. (The proof is not examinable.) We first consider the case where $\Omega = \mathbb{R}^n$ and start by recording that the space of simple functions

$$\mathrm{L}^{\mathrm{simple}}(\mathbb{R}^n) := \mathrm{span}\{\mathbf{1}_A : A \subset \mathbb{R}^n \text{ is measurable and } \mathscr{L}^n(A) < \infty\}$$

is a dense subspace of $\mathrm{L}^p(\mathbb{R}^n)$. Indeed in the process of defining the (Lebesgue) integral one shows that any nonnegative measurable function can be approximated pointwise from below by nonnegative simple functions. Hence given $f \in \mathrm{L}^p(\mathbb{R}^n)$ we select a representative, again denoted by f, write it in its positive and negative parts, $f = f^+ - f^-$, and find sequences of nonnegative simple functions (s_j^+), (s_j^-) with $s_j^+ \nearrow f^+$, $s_j^- \nearrow f^-$ as $j \nearrow \infty$. Now $s_j = s_j^+ - s_j^- \in \mathrm{L}^{\mathrm{simple}}(\mathbb{R}^n)$ and

$$\|f - s_j\|_p^p = \int_{\mathbb{R}^n} \left((f^+ - s_j^+)^p + (f^- - s_j^-)^p \right) \mathrm{d}x \searrow 0 \text{ as } j \nearrow \infty$$

by Lebesgue's monotone convergence theorem. It therefore suffices to show that we can approximate an indicator function for a measurable subset of finite Lebesgue measure by compactly supported continuous functions in $\mathrm{L}^p(\mathbb{R}^n)$. Fix a measurable subset A of $\mathbb{R}^n$ with $\mathscr{L}^n(A) < \infty$ and let $\varepsilon > 0$ be given. We first show that we can find a finite number of closed cubes $Q_1, \dots, Q_N$ in $\mathbb{R}^n$ such that

$$\mathscr{L}^n(A \Delta \bigcup_{j=1}^N Q_j) < \varepsilon$$

where Δ denotes symmetric set-difference: $S \Delta T := (S \setminus T) \cup (T \setminus S)$. This is essentially a matter of how we defined Lebesgue measure and what it means to be measurable. Recall that the outer Lebesgue measure on $\mathbb{R}^n$ is defined for any subset E of $\mathbb{R}^n$ by

$$\mathscr{L}_*^n(E) := \inf \sum_{j=1}^\infty \mathrm{vol}(Q_j)$$

where the infimum is taken over all countable families $\{Q_j\}_{j \in \mathbb{N}}$ of closed cubes with $E \subset \bigcup_{j \in \mathbb{N}} Q_j$. A closed cube Q has the form $Q = [a_1, a_1 + \ell] \times \dots \times [a_n, a_n + \ell]$ and its n-dimensional volume is $\mathrm{vol}(Q) = \ell^n$. Recall that the set E is (Lebesgue) measurable if for any $t > 0$ we can find an open set O in $\mathbb{R}^n$ so $E \subseteq O$ and $\mathscr{L}_*^n(O \setminus E) < t$. It was shown in the *Integration* course that the family of measurable sets in $\mathbb{R}^n$ is a σ-algebra that contains the open sets and that $\mathscr{L}_*^n$ is σ-additive on this σ-algebra. The Lebesgue measure $\mathscr{L}^n$ is the restriction of $\mathscr{L}_*^n$ to the σ-algebra of measurable sets. We now return to the measurable set A and choose a countable family of closed cubes so that

$$A \subset \bigcup_{j \in \mathbb{N}} Q_j \text{ and } \sum_{j=1}^\infty \mathrm{vol}(Q_j) \leqslant \mathscr{L}^n(A) + \frac{\varepsilon}{2}.$$

Because $\mathscr{L}^n(A) < \infty$, the series converges and we can find $N \in \mathbb{N}$ such that the tail $\sum_{j=N+1}^\infty \mathrm{vol}(Q_j) < \varepsilon/2$. Now with $B = \bigcup_{j=1}^N Q_j$ we have by additivity and monotonicity of $\mathscr{L}^n$ (measurability is used in the third line):

$$\begin{aligned}
\mathscr{L}^n(A \Delta B) &\leqslant \mathscr{L}^n(A \setminus B) + \mathscr{L}^n(B \setminus A) \\
&\leqslant \mathscr{L}^n(\bigcup_{j=N+1}^\infty Q_j) + \mathscr{L}^n(\bigcup_{j=1}^\infty Q_j \setminus A) \\
&\leqslant \sum_{j=N+1}^\infty \mathrm{vol}(Q_j) + \sum_{j=1}^\infty \mathrm{vol}(Q_j) - \mathscr{L}^n(A) \\
&\leqslant \varepsilon.
\end{aligned}$$

Because $\|\mathbf{1}_A - \mathbf{1}_B\|_p^p = \mathscr{L}^n(A \Delta B)$ we see that it suffices to show that we can approximate the indicator function of a closed rectangle R in $\mathbb{R}^n$ by $\mathrm{C}_c(\mathbb{R}^n)$ functions in $\mathrm{L}^p(\mathbb{R}^n)$. In the one-dimensional case $R = [a, b]$ and we can take a continuous piecewise linear function ψ defined by

$$\psi(x) = \begin{cases} 1 & \text{if } a \leqslant x \leqslant b, \\ 0 & \text{if } x \leqslant a - \varepsilon \text{ or } x \geqslant b + \varepsilon, \end{cases}$$

and with ψ linear on the intervals $[a - \varepsilon, a]$ and $[b, b + \varepsilon]$. Then $\|\mathbf{1}_R - \psi\|_p < (2\varepsilon)^{\frac{1}{p}}$. In the general n-dimensional case, it suffices to note that the indicator function of a closed rectangle is the product of indicator functions of closed intervals. Thus, the desired $\mathrm{C}_c(\mathbb{R}^n)$ function is simply the product of functions like ψ above.

Finally we must deal with the case of a general nonempty open set Ω. Let $f \in \mathrm{L}^p(\Omega)$ and $\varepsilon > 0$. We may consider $f \in \mathrm{L}^p(\mathbb{R}^n)$ simply by defining $f = 0$ on $\mathbb{R}^n \setminus \Omega$. By the above we can then find $g \in \mathrm{C}_c(\mathbb{R}^n)$ so $\|f - g\|_p < \varepsilon$. For each $j \in \mathbb{N}$ we consider the set of points in Ω whose distance to the boundary is larger than $1/j$:

$$\Omega_j = \{x \in \Omega : \operatorname{dist}(x, \partial\Omega) > 1/j\}.$$

Observe that by Lebesgue's monotone convergence theorem (since $f = 0$ off Ω)

$$\|f \mathbf{1}_{\mathbb{R}^n \setminus \Omega_j}\|_p \to 0 \text{ as } j \to \infty.$$

Put $\eta_j(x) = \operatorname{dist}(x, \mathbb{R}^n \setminus \Omega_j)$ and $\eta_{j,k}(x) = \min\{1, k\eta_j(x)\}$ for $x \in \mathbb{R}^n$. Then $\eta_{j,k}$ is continuous and has support $\overline{\Omega}_j$. Furthermore $\eta_{j,k}(x) \nearrow \mathbf{1}_{\Omega_j}(x)$ pointwise in $x \in \mathbb{R}^n$ and $j \in \mathbb{N}$ as $k \nearrow \infty$, so by another application of Lebesgue's monotone convergence theorem

$$\|f - \eta_{j,k} g\|_p \to \|f - \mathbf{1}_{\Omega_j} g\|_p \text{ as } k \to \infty$$

for each $j \in \mathbb{N}$. Take $j \in \mathbb{N}$ so $\|f \mathbf{1}_{\mathbb{R}^n \setminus \Omega_j}\|_p < \varepsilon$ and next $k \in \mathbb{N}$ so $\|f - \eta_{j,k} g\|_p < \|f - \mathbf{1}_{\Omega_j} g\|_p + \varepsilon$. Check that $\varphi = \eta_{j,k} g \in \mathrm{C}_c(\Omega)$ and

$$\begin{aligned} \|f - \varphi\|_p &< \|f - \mathbf{1}_{\Omega_j} g\|_p + \varepsilon \\ &\leqslant \|f \mathbf{1}_{\mathbb{R}^n \setminus \Omega_j}\|_p + \|\mathbf{1}_{\Omega_j}(f - g)\|_p + \varepsilon \\ &< 3\varepsilon. \end{aligned}$$

□

Proof of Proposition 2.8. Part *(i)* follows by applying Lemma 2.9 inductively. For part *(ii)*, we use Hölder's inequality. Let

$$\frac{1}{p} + \frac{1}{q} = 1$$

and write for each x and almost every y,

$$|\rho_\varepsilon(x - y) u(y)| = \rho_\varepsilon(x - y)^{\frac{1}{q}} \rho_\varepsilon(x - y)^{\frac{1}{p}} |u(y)|.$$

Integrating over $y \in \mathbb{R}^n$ and using Hölder's inequality,

$$\begin{aligned} \int_{\mathbb{R}^n} |\rho_\varepsilon(x - y) u(y)| \,\mathrm{d}y &\leqslant \left(\int_{\mathbb{R}^n} \rho_\varepsilon(x - y) \,\mathrm{d}y \right)^{\frac{1}{q}} \left(\int_{\mathbb{R}^n} \rho_\varepsilon(x - y) |u(y)|^p \,\mathrm{d}y \right)^{\frac{1}{p}} \\ &= \left(\int_{\mathbb{R}^n} \rho_\varepsilon(x - y) |u(y)|^p \,\mathrm{d}y \right)^{\frac{1}{p}}. \end{aligned}$$

Integrating over $x \in \Omega$,

$$\begin{aligned} \int_\Omega |(\rho_\varepsilon * u)(x)|^p \,\mathrm{d}x &\leqslant \int_\Omega \int_{\mathbb{R}^n} \rho_\varepsilon(x - y) |u(y)|^p \,\mathrm{d}y \,\mathrm{d}x \\ &\overset{\dagger}{=} \int_{\mathbb{R}^n} |u(y)|^p \int_\Omega \rho_\varepsilon(x - y) \,\mathrm{d}x \,\mathrm{d}y \\ &\leqslant \int_{\mathbb{R}^n} |u(y)|^p \int_{\mathbb{R}^n} \rho_\varepsilon(x - y) \,\mathrm{d}x \,\mathrm{d}y = \|u\|_p^p, \end{aligned}$$

where in † we used Fubini–Tonelli. For *(iii)*, let $\tau > 0$ and use Lemma 2.10 to find $v \in \mathrm{C}_c(\Omega)$ such that $\|u - v\|_p \leqslant \tau$. Using Minkowski's inequality, we have

$$\begin{aligned}
\|\rho_\varepsilon * u - u\|_p &\leqslant \|\rho_\varepsilon * (u - v)\|_p + \|\rho_\varepsilon * v - v\|_p + \|v - u\|_p \\
&\overset{(ii)}{\leqslant} 2\|v - u\|_p + \|\rho_\varepsilon * v - v\|_p \\
&< 2\tau + \|\rho_\varepsilon * v - v\|_\infty \mathscr{L}^n \left(\overline{B_\varepsilon(\mathrm{supp}(v))}\right)^{\frac{1}{p}}.
\end{aligned}$$

Because v is continuous and compactly supported, so in particular uniformly continuous, we can find $\varepsilon_0 = \varepsilon_0(\tau) > 0$ such that

$$\|\rho_\varepsilon * v - v\|_\infty \mathscr{L}^n \left(\overline{B_\varepsilon(\mathrm{supp}(v))}\right)^{\frac{1}{p}} < \tau$$

for all $\varepsilon \in (0, \varepsilon_0]$. Consequently, $\|\rho_\varepsilon * u - u\|_p < 3\tau$ for $\varepsilon \in (0, \varepsilon_0]$. □

Remark 2.11. The result of Proposition 2.8 (i) and (ii) are also true when $p = \infty$, however (iii) is false (why?)

We are now ready to prove two useful technical results.

2.2.3 Cut-off Functions and Partitions of Unity

Theorem 2.12. *Let K be a compact subset of Ω. There exists $\phi \in \mathscr{D}(\Omega)$ such that $0 \leqslant \phi \leqslant 1$ and $\phi \equiv 1$ on K. We refer to ϕ as a* cut-off function *between K and $\mathbb{R}^n \setminus \Omega$.*

Proof. Put $d \coloneqq \mathrm{dist}(K, \partial\Omega) > 0$ and fix $\delta \in \left(0, \frac{d}{4}\right]$. Put $\tilde{K} = \overline{B_{2\delta}(K)}$. Recall that by definition,

$$\tilde{K} = \{x \in \mathbb{R}^n \ : \ \mathrm{dist}(x, K) \leqslant 2\delta\}.$$

Let $(\rho_\varepsilon)_{\varepsilon>0}$ be the standard mollifier and put $\phi \coloneqq \rho_\delta * \mathbf{1}_{\tilde{K}}$. Then $\phi \in \mathrm{C}^\infty(\mathbb{R}^n)$, $\mathrm{supp}(\phi) \subset \overline{B_\delta(\tilde{K})} = \overline{B_{3\delta}(K)}$, and since $\delta \leqslant \frac{d}{4}$, then $\mathrm{supp}(\phi) \subset \Omega$ and hence $\phi \in \mathcal{D}(\Omega)$. Next, $0 \leqslant \phi \leqslant 1$, and for $x \in K$ we have $\overline{B_\delta(x)} \subset \tilde{K}$, so

$$\phi(x) = \int_{\mathbb{R}^n} \rho_\delta(x - y) \mathbf{1}_{\tilde{K}}(y)\,\mathrm{d}y = \int_{\mathbb{R}^n} \rho_\delta(x - y)\,\mathrm{d}y = 1.$$

Note that $\rho_\delta(x - y)$ is supported in $\overline{B_\delta(x)}$. □

Remark 2.13. For a multi-index α we have

$$\begin{aligned}
|D^\alpha \phi(x)| &= \left| \int_{\mathbb{R}^n} \delta^{-|\alpha|} (D^\alpha \rho)_\delta (x - y) \mathbf{1}_{\tilde{K}}(y)\,\mathrm{d}y \right| \\
&\leqslant \delta^{-|\alpha|} \int_{\mathbb{R}^n} |(D^\alpha \rho)_\delta (x - y)|\,\mathrm{d}y \\
&= \delta^{-|\alpha|} \|D^\alpha \rho\|_1,
\end{aligned}$$

hence if we take $\delta = d/4$, then we arrive at the bound

$$|D^\alpha \phi| \leqslant c_\alpha d^{-|\alpha|},$$

where $c_\alpha = c_\alpha(n, \rho) = 4^{|\alpha|} \|D^\alpha \rho\|_1$, a constant that only depends on α, n and ρ.

The next result refines Theorem 2.12 and is used in connection with localization of distributions.

Theorem 2.14. *Let $\Omega = \bigcup_{j=1}^m \Omega_j$, where $\Omega_1, \ldots, \Omega_m$ are open, non-empty, potentially overlapping sets. For $K \subset \Omega$ compact there exist $\phi_1, \ldots, \phi_m \in \mathscr{D}(\Omega)$ satisfying $\operatorname{supp}(\phi_j) \subset \Omega_j$, $0 \leqslant \phi_j \leqslant 1$,*

$$\sum_{j=1}^m \phi_j \leqslant 1 \quad \text{on} \quad \Omega \tag{13}$$

and

$$\sum_{j=1}^m \phi_j = 1 \quad \text{on} \quad K. \tag{14}$$

We refer to $\phi_1, \ldots, \phi_m$ as a smooth partition of unity *on K subordinate to the cover $\Omega_1, \ldots, \Omega_m$.*

Proof. (*The proof is not examinable.*) Let $x \in K \cap \Omega_j$. Because Ω_j is open, we can find $r_j(x) > 0$ such that $\overline{B_{r_j(x)}(x)} \subset \Omega_j$. The set

$$\left\{ B_{r_j(x)}(x) \ : \ x \in K, \ 1 \leqslant j \leqslant m \right\}$$

is an open cover of K, so by compactness it admits a finite subcover, say

$$\left\{ B_s := B_{r_{j_s}(x_s)}(x_s), \ 1 \leqslant s \leqslant N \right\}.$$

Put $J_j := \{ s \ : \ j_s(x_s) = j \}$, so that

$$\bigcup_{s \in J_j} \overline{B_s} \subset \Omega_j.$$

Now $K_j = K \cap \left(\bigcup_{s \in J_j} \overline{B_s} \right)$ is compact, $K_j \subset \Omega_j$ and $K = \bigcup_{j=1}^m K_j$. We now apply Theorem 2.12 to each K_j, Ω_j to find corresponding cut-off functions $\psi_j \in \mathscr{D}(\Omega_j)$ satisfying $0 \leqslant \psi_j \leqslant 1$ and $\psi_j \equiv 1$ on K_j. We extend ψ_j to $\Omega \setminus \Omega_j$ by zero and, denoting this extension again by ψ_j, have $\psi_j \in \mathscr{D}(\Omega)$. Now define $\phi_1 := \psi_1$, $\phi_2 := \psi_2(1 - \psi_1)$, $\ldots$, $\phi_m := \psi_m \prod_{j=1}^{m-1}(1 - \psi_j)$. By repeated use of the Leibniz rule we see that $\phi_1, \ldots, \phi_m \in \mathrm{C}^\infty(\Omega)$. Clearly $\operatorname{supp}(\phi_j) \subset \Omega_j$, and $0 \leqslant \phi_j \leqslant 1$. Finally, we have by induction on m that

$$\sum_{j=1}^m \phi_j - 1 = -\prod_{j=1}^m (1 - \psi_j)$$

and this easily implies (13) and (14). □

2.3 Convergence in the sense of test functions

Before defining distributions corresponding to smooth compactly supported test functions we must first discuss a notion of convergence in $\mathscr{D}(\Omega)$. Later when other notions of test functions are introduced we shall be more precise and refer to the present mode of convergence as $\mathscr{D}(\Omega)$-convergence.

Definition 2.15. *Let (ϕ_j) be a sequence in $\mathscr{D}(\Omega)$ and $\phi \in \mathscr{D}(\Omega)$. We say*

$$\phi_j \to \phi \quad \textit{in } \mathscr{D}(\Omega)$$

if there exists a compact set $K \subset \Omega$ such that $\operatorname{supp}(\phi), \operatorname{supp}(\phi_j) \subset K$ *for all j, and for each multi-index α*

$$\sup_K |\partial^\alpha(\phi_j - \phi)| \to 0.$$

In words, $\phi_j \to \phi$ in $\mathscr{D}(\Omega)$ if and only if all supports are contained in a fixed compact set in Ω, and we have uniform convergence of $\phi_j - \phi$ together with all partial derivatives to the zero function.

Remark 2.16. Convergence in $\mathscr{D}(\Omega)$ is a strong requirement. The requirement of all supports being contained in a fixed compact set is needed to ensure that $\phi(x - j)$ does not converge to zero in $\mathscr{D}(\mathbb{R})$ when $\phi \neq 0$.

Remark 2.17. It is possible to define a topology $\mathcal{T}$ on $\mathscr{D}(\Omega)$ in such a way that $\phi_j \to \phi$ in $\mathscr{D}(\Omega)$ corresponds to $\phi_j \to \phi$ in the topological space $(\mathscr{D}(\Omega), \mathcal{T})$. Furthermore, in this topology the vector space operations can be shown to be continuous so that $(\mathscr{D}(\Omega), \mathcal{T})$ is an example of a *topological vector space*. It can furthermore be shown that the topology $\mathcal{T}$ is not metrizable: there does not exist a metric d on $\mathscr{D}(\Omega)$ such that $\mathcal{T}$ is the family of open sets corresponding to d. The fact that the convergence can be defined in terms of a topology plays no direct role in this course.

Example 2.18. Let $\varphi \in \mathscr{D}(\mathbb{R}^n)$ and $v \in \mathbb{R}^n \setminus \{0\}$. Then

$$\frac{\Delta_{tv}\varphi}{t} \to \partial_v \varphi \text{ in } \mathscr{D}(\mathbb{R}^n) \text{ as } t \to 0.$$

We simply check the definition: Put $K = \overline{B_1(\operatorname{supp}(\varphi))}$. Then K is compact and for $0 < |t| < 1/|v|$ we have that $\Delta_{tv}\varphi/t$ and $\partial_v\varphi$ are supported in K. Furthermore we have for any multi-index α that

$$\partial^\alpha \frac{\Delta_{tv}\varphi}{t} = \frac{\Delta_{tv}\partial^\alpha\varphi}{t} \text{ and } \partial^\alpha\partial_v\varphi = \partial_v\partial^\alpha\varphi$$

and it is not difficult to see that

$$\max_K \left| \frac{\Delta_{tv}\partial^\alpha\varphi}{t} - \partial_v\partial^\alpha\varphi \right| \to 0 \text{ as } t \to 0.$$

Example 2.19. Let $(\rho_\varepsilon)_{\varepsilon>0}$ be the standard mollifier and $\varphi \in \mathscr{D}(\mathbb{R}^n)$. Then

$$\rho_\varepsilon * \varphi \to \varphi \text{ in } \mathscr{D}(\mathbb{R}^n) \text{ as } \varepsilon \searrow 0.$$

We leave it as an exercise to check the definition.

3 Distributions

3.1 The definition.

Definition 3.1. *(*Laurent Schwartz 1940s*) A functional $u\colon \mathscr{D}(\Omega) \to \mathbb{R}$ (or $u\colon \mathscr{D}(\Omega) \to \mathbb{C}$) is a* distribution *on Ω if*

(i) u is linear: $u(\phi + t\psi) = u(\phi) + tu(\psi)$ for ϕ, $\psi \in \mathscr{D}(\Omega)$, $t \in \mathbb{R}$ (or $\mathbb{C}$), and

(ii) u is continuous in the sense that $u(\phi_j) \to u(\phi)$ whenever $\phi_j \to \phi$ in $\mathscr{D}(\Omega)$.

The set of all distributions on Ω is denoted by $\mathscr{D}'(\Omega)$.

Remark 3.2. Firstly, because of linearity, the continuity condition (ii) holds if and only if it holds at $\phi = 0$. Indeed, if $u(\phi_j) \to 0$ whenever $\phi_j \to 0$ in $\mathscr{D}(\Omega)$ and $\psi_j \to \psi$ in $\mathscr{D}(\Omega)$, then we take $\phi_j = \psi_j - \psi$ and note that $\phi_j \to 0$ in $\mathscr{D}(\Omega)$. Then by assumption, $u(\phi_j) \to 0$. But u is linear, so $u(\phi_j) = u(\psi_j) - u(\psi)$ and so $u(\psi_j) \to u(\psi)$. In the following we shall often refer to the continuity condition (ii) as *$\mathscr{D}$-continuity*.

Secondly, when $u\colon \mathscr{D}(\Omega) \to \mathbb{R}$ is linear (and defined everywhere on $\mathscr{D}(\Omega)$), then chances are that u is $\mathscr{D}$-continuous and thus is a distribution on Ω. Indeed, the only counterexamples I know are obtained by use of a Hamel basis for $\mathscr{D}(\Omega)$. In turn that such Hamel bases exist follows from the Axiom of Choice.

Bracket notation. When $u \in \mathscr{D}'(\Omega)$ and $\phi \in \mathscr{D}(\Omega)$, we often use the bracket notation and write $\langle u, \phi\rangle$ instead of $u(\phi)$:

$$\langle u, \phi\rangle := u(\phi).$$

Note that $\mathscr{D}'(\Omega)$ becomes a vector space (over $\mathbb{R}$ or $\mathbb{C}$ depending on whether we consider real- or complex-valued distributions) by the definition:

$$(u + tv)(\phi) := u(\phi) + tv(\phi) \text{ for each } \phi \in \mathscr{D}(\Omega)$$

for u, $v \in \mathscr{D}'(\Omega)$ and $t \in \mathbb{R}$ (or $t \in \mathbb{C}$).

Example 3.3. If $f \in \mathrm{L}^p(\Omega)$, $p \in [1, \infty]$, then

$$\langle T_f, \phi\rangle = \int_\Omega f(x)\phi(x)\,\mathrm{d}x, \quad \phi \in \mathscr{D}(\Omega)$$

defines a distribution on Ω. Linearity follows from linearity of the integral, and continuity follows from the Dominated Convergence Theorem. In particular we note that each test function $\varphi \in \mathscr{D}(\Omega)$ also defines a distribution $T_\varphi \in \mathscr{D}'(\Omega)$.

Note that since each $\phi \in \mathscr{D}(\Omega)$ has compact support in Ω and since we defined convergence in $\mathscr{D}(\Omega)$ by requiring all supports to be in a fixed compact set in Ω, the above distribution T_f would also be well-defined if f was only *locally* in L^p.

3.2 Local Lebesgue Spaces.

Definition 3.4. *Let $p \in [1, \infty]$. Then a function $f\colon \Omega \to \mathbb{R}$ (or $f\colon \Omega \to \mathbb{C}$) is locally* L^p *on* Ω *if f is measurable and for each compact subset K of Ω we have that*

$$\begin{cases} \displaystyle\int_K |f(x)|^p \,\mathrm{d}x < \infty & \textit{when } p < \infty \\ \operatorname{ess.sup}_K |f| < \infty & \textit{when } p = \infty. \end{cases}$$

The space $\mathrm{L}^p_{\mathrm{loc}}(\Omega)$ is now defined as the usual L^p spaces, so is the space of equivalence classes of functions $f\colon \Omega \to \mathbb{R}$ that are locally L^p on Ω for the equivalence relation equal almost everywhere. Like the usual L^p spaces also $\mathrm{L}^p_{\mathrm{loc}}(\Omega)$ are vector spaces.

Example 3.5. The function $x^{-1} \notin \mathrm{L}^1(0, \infty)$, but $x^{-1} \in \mathrm{L}^1_{\mathrm{loc}}(0, \infty)$ and, in fact, $x^{-1} \in \mathrm{L}^p_{\mathrm{loc}}(0, \infty)$ for all $p \in [1, \infty]$. Note that Ω determines what *local* means. For example, $x^{-1} \in L^1_{\mathrm{loc}}(0, \infty)$, but $x^{-1} \notin \mathrm{L}^1_{\mathrm{loc}}(-1, 1)$.

Example 3.6. Summarizing a previous discussion, each $f \in \mathrm{L}^p_{\mathrm{loc}}(\Omega)$, $p \in [1, \infty]$, gives rise to a distribution on Ω via

$$\langle T_f, \phi \rangle = \int_\Omega f(x)\phi(x)\,\mathrm{d}x$$

for each $\phi \in \mathscr{D}(\Omega)$.

Example 3.7 (Dirac's delta function at $x_0 \in \Omega$). The map

$$\phi \mapsto \langle \delta_{x_0}, \phi \rangle := \phi(x_0)$$

for $\phi \in \mathscr{D}(\Omega)$ is clearly a distribution on Ω. Furthermore, so is $\phi \mapsto (D^\alpha \phi)(x_0)$ for any multi-index α.

Example 3.8. Let μ be a locally finite Borel measure on Ω (so μ is a countably additive set function defined on the Borel subsets of Ω such that $\mu(K) < \infty$ for all compact subsets K of Ω). Then

$$\langle T_\mu, \varphi \rangle := \int_\Omega \varphi \,\mathrm{d}\mu,\ \varphi \in \mathscr{D}(\Omega)$$

defines a distribution on Ω. Linearity is clear and the continuity condition follows, for instance, from the Dominated Convergence Theorem.

3.3 The boundedness property and the order of a distribution.

While the continuity condition (ii) in Definition 3.1 often is not an issue, it is nonetheless useful to reformulate it using linearity as follows.

Theorem 3.9. *A linear functional $u\colon \mathscr{D}(\Omega) \to \mathbb{R}$ (or $\mathbb{C}$) is a distribution if and only if for every compact set $K \subset \Omega$ there exist constants $c = c(K) > 0$ and $m = m(K) \in \mathbb{N}_0$ such that*

$$|\langle u, \phi \rangle| \leqslant c \sum_{|\alpha| \leqslant m} \sup_K |D^\alpha \phi| \tag{15}$$

for all $\phi \in \mathscr{D}(K) := \{\phi \in \mathscr{D}(\Omega) \ : \ \operatorname{supp}(\phi) \subset K\}$.

Proof. If $\phi_j \to 0$ in $\mathscr{D}(\Omega)$, then for some compact set $K \subset \Omega$ we have $\phi_j \in \mathscr{D}(K)$ for all j. Then by assumption we can find $c = c(K) > 0$ and $m = m(K) \in \mathbb{N}_0$ such that (15) holds. But then

$$|\langle u, \phi_j\rangle| \leqslant c \sum_{|\alpha|\leqslant m} \sup_K |\partial^\alpha \phi_j| \to 0.$$

For the converse, we argue by contradiction. Assume there exists $u \in \mathscr{D}(\Omega)$ and a compact set $K \subset \Omega$ such that (15) is violated for all choices of c and m. In particular, for $c = m = j$ we can find $\phi_j \in \mathscr{D}(K)$ with

$$|\langle u, \phi_j\rangle| > j \sum_{|\alpha|\leqslant j} \sup_K |\partial^\alpha \phi_j|.$$

Put $\lambda_j = \langle u, \phi_j\rangle$. Then $\lambda_j \neq 0$, $\psi_j := \frac{\phi_j}{\lambda_j} \in \mathscr{D}(K)$, $\big|\langle u, \psi_j\rangle\big| = 1$, and

$$1 > j \sum_{|\alpha|\leqslant j} \sup_K |\partial^\alpha \psi_j|.$$

Thus $|D^\alpha \psi_j| < j^{-1}$ on Ω for $j \geqslant |\alpha|$, and in particular $\psi_j \to 0$ in $\mathscr{D}(\Omega)$. But $\langle u, \psi_j\rangle = 1$, which does not converge to zero. □

Definition 3.10. *Let $u \in \mathscr{D}'(\Omega)$. If there exists an $m \in \mathbb{N}_0$ with the property that for all compact subsets $K \subset \Omega$ there exists a constant $c = c_K > 0$ such that*

$$|\langle u, \phi\rangle| \leqslant c \sum_{|\alpha|\leqslant m} \sup_K |\partial^\alpha \phi|$$

for all $\phi \in \mathscr{D}(K)$, then we say u has order at most m. *The set of these distributions is denoted*

$$\mathscr{D}'_m(\Omega) := \big\{u \in \mathscr{D}'(\Omega) : u \text{ has order at most } m\big\}.$$

We say u has order 0 *if u has order at most 0. For $m \in \mathbb{N}$ we say u* has order m *if u has order at most m, but not order at most $m - 1$.*

We say u has order infinity *if u does not have order at most m for any $m \in \mathbb{N}_0$.*

Note that by Theorem 3.9, any distribution has locally finite order.

Example 3.11. Let $f \in \mathrm{L}^1_{\mathrm{loc}}(\Omega)$. Then the corresponding distribution has order 0. Indeed, if $K \subset \Omega$ is compact and $\varphi \in \mathscr{D}(K)$, then

$$\begin{aligned} |\langle T_f, \varphi\rangle| &= \left|\int_\Omega f(x)\varphi(x)\,\mathrm{d}x\right| \\ &\leqslant \int_K |f||\varphi|\,\mathrm{d}x \\ &\leqslant \sup_K |\varphi| \int_K |f|\,\mathrm{d}x, \end{aligned}$$

and so $|\langle T_f,\, \varphi\rangle| \leqslant c\sup_K |\varphi|$. The same calculation applies when μ is a locally finite Borel measure on Ω and shows that also the distribution T_μ has order 0.

Example 3.12. Let $x_0 \in \Omega$ and $\alpha \in \mathbb{N}_0^n$. Define

$$\langle T,\, \varphi\rangle := (\partial^\alpha \varphi)(x_0)$$

for $\varphi \in \mathscr{D}(\Omega)$. Then $T \in \mathscr{D}'(\Omega)$ as T is clearly linear and for compact $K \subset \Omega$ and $\varphi \in \mathscr{D}(K)$,

$$|\langle T,\, \varphi\rangle| = |(\partial^\alpha \varphi)(x_0)| \leqslant \sup_K |\partial^\alpha \varphi|.$$

It also shows that T has order at most $|\alpha|$. If $\alpha = 0$ so that $T = \delta_{x_0}$, we see that T has order 0. Assume $|\alpha| > 0$. We shall prove that T has order $|\alpha|$. Suppose, for contradiction, that T has order at most $|\alpha| - 1$. Take $r \in (0, \min\{1, \operatorname{dist}(x_0, \partial\Omega)\})$ and put $K = \overline{B_r(x_0)}$. Then $K \subset \Omega$ is compact. By assumption, we can then find $c = c_K \geqslant 0$ such that

$$|\langle T,\, \varphi\rangle| = |(\partial^\alpha \varphi)(x_0)| \leqslant c \sum_{|\beta| \leqslant |\alpha|-1} \sup_K |\partial^\beta \varphi| \tag{16}$$

for all $\varphi \in \mathscr{D}(K)$. Take $\psi \in \mathscr{D}(B_1(0))$ with $\psi(0) = 1$ (for instance $\psi(x) = \rho(x)/\rho(0)$ with ρ the standard mollifier kernel will do) and define, for $\varepsilon \in (0, r)$,

$$\varphi(x) := \frac{(x - x_0)^\alpha}{\alpha!} \psi\left(\frac{x - x_0}{\varepsilon}\right), \quad x \in \Omega.$$

Note that φ is C^∞ and $\operatorname{supp}(\varphi) \subseteq \overline{B_\varepsilon(x_0)} \subset K$, so that $\varphi \in \mathscr{D}(K)$. Also,

$$\partial^\beta \left(\frac{(x - x_0)^\alpha}{\alpha!}\right)\bigg|_{x=x_0} = \begin{cases} 1 \text{ if } \beta = \alpha \\ 0 \text{ if } \beta \neq \alpha \end{cases},$$

so that $\partial^\alpha \varphi(x_0) = 1$. If $\beta \in \mathbb{N}_0^n$ is any multi-index with length $|\beta| \leqslant |\alpha| - 1$ and γ is a multi-index with $\gamma \leqslant \beta$, then for $x \in B_\varepsilon(x_0)$ we have

$$\left|\partial_x^\gamma \left(\frac{(x - x_0)^\alpha}{\alpha!}\right)\right| \leqslant \varepsilon^{|\alpha| - |\gamma|}.$$

This follows because when $\gamma \leqslant \alpha$, then

$$\left|\partial_x^\gamma \left(\frac{(x - x_0)^\alpha}{\alpha!}\right)\right| \leqslant |(x - x_0)^{\alpha - \gamma}|,$$

whereas if $\gamma_j > \alpha_j$ for some j, then $\partial_x^\gamma (x - x_0)^\alpha = 0$. We can now estimate for $x \in K$ using the generalized Leibniz rule and noticing that terms involving $\psi(\frac{x - x_0}{\varepsilon})$ vanish when $|x - x_0| \geqslant \varepsilon$:

$$\begin{aligned} |\partial^\beta \varphi(x)| &\leqslant \sum_{\gamma \leqslant \beta} \binom{\beta}{\gamma} \left|\partial_x^\gamma \left(\frac{(x - x_0)^\alpha}{\alpha!}\right)\right| \left|\partial_x^{\beta - \gamma} \psi\left(\frac{x - x_0}{\varepsilon}\right)\right| \\ &\leqslant \sum_{\gamma \leqslant \beta} 2^{|\beta|} \varepsilon^{|\alpha| - |\gamma|} \varepsilon^{|\gamma| - |\beta|} \max\bigl\{|\partial^\zeta \psi(x)| :\, |\zeta| \leqslant |\alpha| - 1,\, x \in \overline{B_\varepsilon(x_0)}\bigr\} \\ &\leqslant c_{\psi,\alpha} \varepsilon^{|\alpha| - |\beta|}, \end{aligned}$$

where we defined the constant

$$c_{\psi,\alpha} := 2^{|\alpha|}\big(|\alpha| + n\big)\max\big\{|\partial^\zeta \psi(x)| : |\zeta| \leqslant |\alpha| - 1,\, x \in B_1(0)\big\}.$$

We plug this φ into (16) and use the above estimates to get

$$\begin{aligned} 1 = |\langle T, \varphi\rangle| &\leqslant c \sum_{|\beta|\leqslant|\alpha|-1} \sup_K |\partial^\beta \varphi| \\ &\leqslant c \sum_{|\beta|\leqslant|\alpha|-1} c_{\psi,\alpha}\varepsilon^{|\alpha|-|\beta|} \\ &\leqslant cc_{\psi,\alpha}\big(|\alpha| + n\big)\varepsilon = \overline{c}\varepsilon, \end{aligned}$$

where we introduced the new constant

$$\overline{c} := cc_{\psi,\alpha}\big(|\alpha| + n\big)$$

and used that $\varepsilon \in (0,1)$. The contradiction is reached if we take $\varepsilon \in (0, r)$ so $\overline{c}\varepsilon < 1$.

A generalization of the above example goes as follows. Let x_j, where $j \in J$ is a countable or finite index set, be distinct points in Ω so that the set $\{x_j \,|\, j \in J\}$ has no limit points in Ω (that is, if there are any limit points, then they must be on $\partial\Omega$). For any set of multi-indices $\alpha_j \in \mathbb{N}_0^n$, $j \in J$, put

$$\langle T, \varphi\rangle := \sum_{j\in J} (\partial^{\alpha_j}\varphi)(x_j)$$

for $\varphi \in \mathscr{D}(\Omega)$. Then $T \in \mathscr{D}'(\Omega)$ and it can be shown that the order of T is $\sup_{j\in J} |\alpha_j|$.

The next is an extension result for distributions of finite order.

Theorem 3.13. *Let $u \in \mathscr{D}'_m(\Omega)$. Then u can be uniquely extended to a linear functional $\bar{u}\colon \mathrm{C}_c^m(\Omega) \to \mathbb{C}$ with the boundedness property: for each compact subset K of Ω there exists a constant $c = c(K)$ so*

$$\big|\langle \bar{u}, \varphi\rangle\big| \leqslant c \sum_{|\alpha|\leqslant m} \sup_K |\partial^\alpha \varphi| \tag{17}$$

holds for all $\varphi \in \mathrm{C}_c^m(\Omega)$.

Notation. We shall usually also denote this unique extension by u.

Proof. Existence: Let $\psi \in \mathrm{C}_c^m(\Omega)$. Take a compact set $K \subset \Omega$ with $\mathrm{supp}(\psi) \subseteq K$. Put $d = \mathrm{dist}(K, \partial\Omega)/2$ and $\psi_j = \rho_{1/j} * \psi$, where $(\rho_\varepsilon)_{\varepsilon>0}$ is the standard mollifier. If $\tilde{K} = \overline{B_d(K)}$, then $\tilde{K} \subset \Omega$ is compact and $\psi_j \in \mathscr{D}(\tilde{K})$ when $j > 1/d$. Because u is of order at most m we can find a constant $c_{\tilde{K}}$ so

$$\big|\langle u, \varphi\rangle\big| \leqslant c_{\tilde{K}} \sum_{|\alpha|\leqslant m} \sup_K |\partial^\alpha \varphi| \tag{18}$$

holds for all $\varphi \in \mathscr{D}(\tilde{K})$. Take $\varphi = \psi_j - \psi_k$ with $j,\, k > 1/d$ we see that

$$\big|\langle u, \psi_j\rangle - \langle u, \psi_k\rangle\big| \leqslant c_{\tilde{K}} \sum_{|\alpha| \leqslant m} \sup_{\tilde{K}} |\partial^\alpha \psi_j - \partial^\alpha \psi_k|.$$

Now $\partial^\alpha \psi$ is uniformly continuous when $|\alpha| \leqslant m$, so $\partial^\alpha \psi_j \to \partial^\alpha \psi$ uniformly as $j \to \infty$ and hence $(\partial^\alpha \psi_j)$ is a uniform Cauchy sequence. But then it follows that

$$\lim_{j,k\to\infty} \big|\langle u, \psi_j\rangle - \langle u, \psi_k\rangle\big| = 0,$$

so $\big(\langle u, \psi_j\rangle\big)$ is a Cauchy sequence in $\mathbb{C}$ and we may define

$$\langle \bar{u}, \psi\rangle = \lim_{j\to\infty} \langle u, \psi_j\rangle.$$

It is easy to see that hereby $\bar{u}\colon \mathrm{C}_c^m(\Omega) \to \mathbb{C}$ is linear and that (17) holds with $c(K) = c_{\tilde{K}}$.

Uniqueness: This is a straight forward exercise. $\square$

Definition 3.14. *A linear functional* $u\colon \mathrm{C}_c(\Omega) \to \mathbb{C}$ *with the boundedness property: for any compact set* $K \subset \Omega$ *we can find a constant* $c_K \geqslant 0$ *such that*

$$\big|\langle u, \varphi\rangle\big| \leqslant c_K \sup_K |\varphi|$$

for all $\varphi \in \mathrm{C}_c(\Omega)$ *with* $\operatorname{supp}(\varphi) \subseteq K$*, is called a* Radon measure *on* Ω*.*

Corollary 3.15. *A distribution of order* 0 *on* Ω *extends uniquely to a Radon measure on* Ω*.*

The next result is important and it also justifies the terminology *Radon measure*:

Theorem 3.16. ***(Riesz-Markov representation theorem)*** *Let* $u\colon \mathrm{C}_c(\Omega) \to \mathbb{C}$ *be a Radon measure on* Ω *and assume* u *is positive:*

$$\textit{if } \varphi \in \mathrm{C}_c(\Omega) \textit{ and } \varphi \geqslant 0, \textit{ then } \langle u, \varphi\rangle \geqslant 0.$$

Then there exists a unique locally finite Borel measure μ *on* Ω *so*

$$\langle u, \varphi\rangle = \int_\Omega \varphi \,\mathrm{d}\mu,\ \varphi \in \mathrm{C}_c(\Omega).$$

We omit the proof.

Theorem 3.17. *Let* $u \in \mathscr{D}'(\Omega)$ *be a* positive distribution*:*

$$\langle u, \varphi\rangle \geqslant 0 \ \textit{ if } \ \varphi \in \mathscr{D}(\Omega) \ \textit{ and } \ \varphi \geqslant 0.$$

Then there exists a unique locally finite Borel measure μ *on* Ω *so* $u = T_\mu$*.*

Proof. It suffices to check that u has order 0. To that end we fix a compact subset $K \subset \Omega$. Put $d = \operatorname{dist}(K, \partial\Omega)/4$ and $\psi = \rho_d * \mathbf{1}_{\overline{B_d(K)}}$, where $(\rho_\varepsilon)_{\varepsilon>0}$ is the standard mollifier. If $\varphi \in \mathscr{D}(K)$ is real-valued, then $0 \leqslant \|\varphi\|_\infty \psi \pm \varphi \in \mathscr{D}(\Omega)$, hence by positivity and linearity of u also

$$0 \leqslant \langle u, \|\varphi\|_\infty \psi \pm \varphi \rangle = \|\varphi\|_\infty \langle u, \psi \rangle \pm \langle u, \varphi \rangle$$

and consequently, $|\langle u, \varphi \rangle| \leqslant c_K \|\varphi\|_\infty$ with $c_K := \langle u, \psi \rangle$. When $\varphi \in \mathscr{D}(K)$ is complex valued, we apply the above to the real and imaginary parts of φ and use linearity: if $\varphi_1 = \operatorname{Re}(\varphi)$, $\varphi_2 = \operatorname{Im}(\varphi)$, then clearly $\varphi_1, \varphi_2 \in \mathscr{D}(K)$ and

$$\begin{aligned} \big|\langle u, \varphi \rangle\big| &\leqslant \big|\langle u, \varphi_1 \rangle\big| + \big|\langle u, \varphi_2 \rangle\big| \\ &\leqslant c_K\big(\|\varphi_1\|_\infty + \|\varphi_2\|_\infty\big) \leqslant 2c_K \|\varphi\|_\infty. \end{aligned}$$

□

3.4 The fundamental lemma of the Calculus of Variations.

When $f \in \mathrm{L}^p_{\mathrm{loc}}(\Omega)$, then as remarked before

$$\langle T_f, \phi \rangle = \int_\Omega f(x)\phi(x)\,\mathrm{d}x, \quad \phi \in \mathscr{D}(\Omega)$$

defines a distribution on Ω. It is natural to ask if the distribution T_f determines f, that is, if for $f, g \in \mathrm{L}^p_{\mathrm{loc}}(\Omega)$ we have $T_f = T_g$, must it be the case that $f = g$ almost everywhere? The answer is affirmative and relies on the following.

Lemma 3.18 (The Fundamental Lemma of the Calculus of Variations)**.** *If $f \in \mathrm{L}^1_{\mathrm{loc}}(\Omega)$ and*

$$\int_\Omega f(x)\phi(x)\,\mathrm{d}x = 0$$

for all $\phi \in \mathscr{D}(\Omega)$, then $f = 0$ almost everywhere.

Remark 3.19. The result is also sometimes called the *Du Bois-Reymond Lemma.*

Proof. Let $\mathcal{O}$ be a non-empty open subset of Ω such that $\overline{\mathcal{O}}$ is compact and $\overline{\mathcal{O}} \subset \Omega$. (Recall that we use the short-hand $\mathcal{O} \Subset \Omega$ for this situation.)

Put $g = f\mathbf{1}_{\mathcal{O}}$ and extend g to $\mathbb{R}^n \setminus \Omega$ by zero. Because $\overline{\mathcal{O}} \subset \Omega$ is compact, we have $g \in \mathrm{L}^1(\mathbb{R}^n)$. For the standard mollifier $(\rho_\varepsilon)_{\varepsilon>0}$ we know, by proposition 2.8, that $\|\rho_\varepsilon * g - g\|_1 \to 0$ as $\varepsilon \searrow 0$. Now note that for $x \in \mathcal{O}$,

$$\begin{aligned} (\rho_\varepsilon * g)(x) &= \int_{\mathbb{R}^n} \rho_\varepsilon(x-y)g(y)\,\mathrm{d}y \\ &= \int_{\mathcal{O}} \rho_\varepsilon(x-y)f(y)\,\mathrm{d}y. \end{aligned}$$

If we take $x \in \mathcal{O}$ and $\varepsilon \in (0, \operatorname{dist}(x, \partial\mathcal{O}))$, then, denoting $\phi^x(y) := \rho_\varepsilon(x-y)$ for $y \in \Omega$, we have $\phi^x \in \mathrm{C}^\infty(\Omega)$ and $\operatorname{supp}(\phi^x) = \overline{B_\varepsilon(x)} \subset \mathcal{O} \subset \Omega$, so $\phi^x \in \mathscr{D}(\Omega)$. By assumption,

$$0 = \int_\Omega f(y)\phi^x(y)\,\mathrm{d}y = \int_{\mathcal{O}} f(y)\rho_\varepsilon(x-y)\,\mathrm{d}y = (\rho_\varepsilon * g)(x).$$

It follows that $(\rho_\varepsilon * g)(x) \to 0$ as $\varepsilon \searrow 0$ pointwise in $x \in \mathcal{O}$. From Fatou's Lemma, we therefore get

$$\begin{aligned}\int_{\mathcal{O}} |g|\,\mathrm{d}x &\leqslant \liminf_{\varepsilon\searrow 0} \int_{\mathcal{O}} |\rho_\varepsilon * g - g|\,\mathrm{d}x \\ &\leqslant \lim_{\varepsilon\searrow 0} \int_{\mathbb{R}^n} |\rho_\varepsilon * g - g|\,\mathrm{d}x = 0.\end{aligned}$$

Thus $f = g = 0$ almost everywhere in $\mathcal{O}$ and since $\mathcal{O} \Subset \Omega$ was arbitrary, we conclude that $f = 0$ almost everywhere. □

Notation and terminology. When $f \in \mathrm{L}^p_{\mathrm{loc}}(\Omega)$, we shall also use f to denote the distribution T_f. Thus we simply identify T_f with f:

$$T_f = f.$$

This is of course an abuse of notation, but it is convenient and should not cause too much trouble. Furthermore, we refer to the distributions that correspond to an $\mathrm{L}^p_{\mathrm{loc}}(\Omega)$ function as *regular distributions* on Ω.

Lemma 3.20. *If μ and ν are two locally finite Borel measures on Ω and $T_\mu = T_\nu$, then $\mu = \nu$.*

Proof. It suffices to prove that $\mu(K) = \nu(K)$ for all compact subsets K of Ω, so fix such a set K. For each $\varepsilon \in (0, \operatorname{dist}(K, \partial\Omega))$ we define $\varphi_\varepsilon = \rho_\varepsilon * \mathbf{1}_K$, where $(\rho_\varepsilon)_{\varepsilon>0}$ is the standard mollifier. Then $\varphi_\varepsilon \in \mathscr{D}(\Omega)$ and so by hypothesis

$$\int_\Omega \varphi_\varepsilon\,\mathrm{d}\mu = \langle T_\mu, \varphi_\varepsilon\rangle = \langle T_\nu, \varphi_\varepsilon\rangle = \int_\Omega \varphi_\varepsilon\,\mathrm{d}\nu.$$

The conclusion follows if we take $\varepsilon \searrow 0$ and use the Dominated Convergence Theorem. □

Notation. In view of this lemma we shall also identify the distribution T_μ with μ from now on.

3.5 Convergence in the sense of distributions.

Later when other notions of distributions are introduced we shall be more precise and refer to the present mode of convergence as $\mathscr{D}'(\Omega)$-convergence.

Definition 3.21. *Let (u_j) be a sequence in $\mathscr{D}'(\Omega)$ and let $u \in \mathscr{D}'(\Omega)$. We say u_j converges to u in the sense of distributions on Ω and write*

$$u_j \longrightarrow u \text{ in } \mathscr{D}'(\Omega)$$

if

$$\langle u_j,\, \varphi\rangle \longrightarrow \langle u,\, \varphi\rangle$$

for each $\varphi \in \mathscr{D}(\Omega)$.

Remark 3.22. As with convergence in $\mathscr{D}(\Omega)$, one can define a topology $\mathcal{T}'$ on $\mathscr{D}'(\Omega)$ so that $u_j \to u$ in $\mathscr{D}'(\Omega)$ corresponds to $u_j \to u$ in the topological space $(\mathscr{D}'(\Omega), \mathcal{T}')$. As was the case for the space of test functions it can be shown that also this topology is not metrizable. Furthermore, $(\mathscr{D}'(\Omega), \mathcal{T}')$ is a so-called topological vector space, in fact exactly the dual space of $(\mathscr{D}(\Omega), \mathcal{T})$. We shall not pursue this abstract viewpoint here as it is not really necessary for the work with distributions that we cover in this course.

Whereas convergence in the sense of test functions was an extremely strong condition, convergence in the sense of distributions is an extremely weak condition. We illustrate this with an example.

Example 3.23. Let $p \in [1, \infty]$ and $f_j, f \in \mathrm{L}^p(\Omega)$. If $f_j \to f$ in $\mathrm{L}^p(\Omega)$, then $f_j \to f$ in $\mathscr{D}'(\Omega)$. This is easy to see. The converse, however, is false:

(i) Let $f_j(x) = \sin(jx)$, $x \in (0,1)$. Then $f_j \to 0$ in $\mathscr{D}'(0,1)$, but $f_j \not\to 0$ in $\mathrm{L}^p(0,1)$ for any $p \in [1, \infty]$.

(ii) Let $g_j(x) = g(jx)$, $x \in (0,1)$, where g is T-periodic and on $(0, T]$ is given by

$$g = -117\mathbf{1}_{(0,\frac{T}{2}]} + 117\mathbf{1}_{(\frac{T}{2},T]} = \begin{cases} -117 & \text{on } (0, \frac{T}{2}] \\ +117 & \text{on } (\frac{T}{2}, T] \end{cases}$$

Clearly $\|g_j\|_1 = 117 \not\to 0$. On Problem Sheet 3 you will be asked to prove that $g_j \to 0$ in $\mathscr{D}'(0,1)$.

(iii) Let $h_j(x) = jg(jx)$, $x \in (0,1)$, where g is as in (ii). Then $h_j \to 0$ in $\mathscr{D}'(0,1)$.

Example 3.24. Let $v \in \mathrm{C}_c(\mathbb{R}^n)$ and for $x_0 \in \Omega$ and $\varepsilon > 0$ put

$$v_\varepsilon(x) := \varepsilon^{-n} v\left(\frac{x - x_0}{\varepsilon}\right),\ x \in \Omega.$$

Then $v_\varepsilon \in \mathscr{D}'(\Omega)$ and $v_\varepsilon \to \int_{\mathbb{R}^n} v \,\mathrm{d}x\, \delta_{x_0}$ in $\mathscr{D}'(\Omega)$ as $\varepsilon \searrow 0$. Indeed, it is clear that $v_\varepsilon \in \mathscr{D}'(\Omega)$ for all $\varepsilon > 0$, and for $\varphi \in \mathscr{D}(\Omega)$ we have

$$\begin{aligned}
\langle v_\varepsilon,\, \varphi\rangle &= \int_\Omega \varepsilon^{-n} v\left(\frac{x - x_0}{\varepsilon}\right) \varphi(x)\,\mathrm{d}x \\
&= \int_{\mathbb{R}^n} v(y)\varphi(x_0 + \varepsilon y)\,\mathrm{d}y \\
&\underset{\varepsilon \searrow 0}{\longrightarrow} \int_{\mathbb{R}^n} v(y)\varphi(x_0)\,\mathrm{d}y = \int_{\mathbb{R}^n} v\,\mathrm{d}y\, \langle \delta_{x_0},\, \varphi\rangle,
\end{aligned}$$

where in the second line we made the change of variables $y = \varepsilon^{-1}(x - x_0)$.

When $\int_{\mathbb{R}^n} v \,\mathrm{d}x = 1$ and we take $x_0 = 0$ above, then the family $(v_\varepsilon)_{\varepsilon>0}$ is called an *approximate unit* (or an *approximate identity*). In particular, the standard mollifier $(\rho_\varepsilon)_{\varepsilon>0}$ is therefore an approximate unit and it has the property that

$$\rho_\varepsilon \to \delta_0 \text{ in } \mathscr{D}'(\mathbb{R}^n) \text{ as } \varepsilon \searrow 0.$$

4 Operations on distributions

4.1 Adjoint identities.

The title of this subsection refers to a principle that often allows us to extend well-known operations on test functions to corresponding operations on distributions. In more technical terms, it amounts to defining operations on distributions from operations on test functions by duality.

Let T be an *operation* on test functions, that is, assume $T\colon \mathscr{D}(\Omega) \to \mathscr{D}(\Omega)$ is a linear map. We would like to extend T to distributions on Ω. Suppose there exists a linear map $S\colon \mathscr{D}(\Omega) \to \mathscr{D}(\Omega)$ satisfying

$$\int_\Omega T(\varphi)\psi \,\mathrm{d}x = \int_\Omega \varphi S(\psi) \,\mathrm{d}x \tag{19}$$

for all φ, $\psi \in \mathscr{D}(\Omega)$. We call (19) an *adjoint identity.* If S is $\mathscr{D}$-continuous, meaning that, $S(\psi_j) \to S(\psi)$ in $\mathscr{D}(\Omega)$ whenever $\psi_j \to \psi$ in $\mathscr{D}(\Omega)$, then we can extend T to distributions u by the rule

$$\langle \bar{T}(u),\, \psi\rangle := \langle u,\, S(\psi)\rangle, \quad \psi \in \mathscr{D}(\Omega). \tag{20}$$

Because S is linear and $\mathscr{D}$-continuous, it follows that $\bar{T}(u) \in \mathscr{D}'(\Omega)$, and therefore that we have defined a map $\bar{T}\colon \mathscr{D}'(\Omega) \to \mathscr{D}'(\Omega)$. We can view $\mathscr{D}(\Omega)$ as a subspace of $\mathscr{D}'(\Omega)$ (recall that we decided to identify φ with its corresponding distribution T_φ) so from (19) and our definition of $\bar{T}$ follow that $\bar{T}|_{\mathscr{D}(\Omega)} = T$, so we have indeed an extension of T. We record that the definitions immediately give that $\bar{T}$ is linear: $\bar{T}(u + \lambda v) = \bar{T}(u) + \lambda\bar{T}(v)$ for u, $v \in \mathscr{D}'(\Omega)$, $\lambda \in \mathbb{C}$, since for $\psi \in \mathscr{D}(\Omega)$,

$$\langle \bar{T}(u + \lambda v),\, \psi\rangle = \langle u + \lambda v,\, S(\psi)\rangle = \langle u,\, S(\psi)\rangle + \lambda\langle v,\, S(\psi)\rangle = \langle \bar{T}(u),\, \psi\rangle + \lambda\langle \bar{T}(v),\, \psi\rangle.$$

The definitions also give that $\bar{T}$ is $\mathscr{D}'$-continuous, meaning that if $u_j \to u$ in $\mathscr{D}'(\Omega)$, then also $\bar{T}(u_j) \to \bar{T}(u)$ in $\mathscr{D}'(\Omega)$. Indeed for each $\psi \in \mathscr{D}(\Omega)$ we have

$$\langle \bar{T}(u_j),\, \psi\rangle = \langle u_j,\, S(\psi)\rangle \longrightarrow \langle u,\, S(\psi)\rangle = \langle \bar{T}(u),\, \psi\rangle.$$

We can now apply this procedure, that we shall refer to as the *adjoint identity scheme*, and extend some well-known operations to distributions.

Example 4.1. *(Differentiation)* Let $T = \frac{\mathrm{d}}{\mathrm{d}x} = \mathrm{D}$ on $\mathscr{D}(\mathbb{R})$. For $\varphi,\, \psi \in \mathscr{D}(\mathbb{R})$ integration by parts yields

$$\int_{\mathbb{R}} \varphi' \psi \,\mathrm{d}x = [\varphi\psi]_{-\infty}^{+\infty} - \int_{-\infty}^{\infty} \varphi\psi' \,\mathrm{d}x = \int_{\mathbb{R}} \varphi(-\psi') \,\mathrm{d}x,$$

hence we have an adjoint identity with $S = -\mathrm{D}$. Clearly, $S\colon \mathscr{D}(\mathbb{R}) \to \mathscr{D}(\mathbb{R})$ is linear and $\mathscr{D}$-continuous, so we may extend differentiation to distributions $u \in \mathscr{D}'(\mathbb{R})$ by the rule

$$\langle \bar{\mathrm{D}}u,\, \psi\rangle = \langle u,\, -\mathrm{D}\psi\rangle, \quad \psi \in \mathscr{D}(\mathbb{R})$$

We know this is consistent for test functions in the sense that $\bar{\mathrm{D}}\varphi = \mathrm{D}\varphi$ when $\varphi \in \mathscr{D}(\mathbb{R})$ ($\bar{\mathrm{D}}$ extends D), but what about for C^1 functions? Suppose $u \in \mathrm{C}^1(\mathbb{R})$ and consider also u as an element of $\mathscr{D}'(\mathbb{R})$. We recall again that we identify u with the corresponding distribution T_u, but perhaps it is useful in this example to be more precise and temporarily use the notation T_u again for the corresponding distribution. In such terms we would like to know the relation between the distributional derivative $\bar{\mathrm{D}}T_u$ defined above and the distribution corresponding to the usual derivative $T_{\mathrm{D}u}$. We have by our definitions:

$$\langle \bar{\mathrm{D}}T_u,\, \psi\rangle = \langle u,\, -\mathrm{D}\psi\rangle = \int_{-\infty}^{\infty} u(-\mathrm{D}\psi)\,\mathrm{d}x = -\,[u\psi]_{-\infty}^{+\infty} + \int_{-\infty}^{\infty} \psi \mathrm{D}u, \mathrm{d}x = \langle T_{\mathrm{D}u}\, \psi\rangle$$

for all $\psi \in \mathscr{D}(\mathbb{R})$, and so $\bar{\mathrm{D}}T_u = T_{\mathrm{D}u}$. That is, they are the same! We are therefore justified in identifying the distributional and usual derivatives for C^1 functions and accordingly we will write $\bar{\mathrm{D}}u = \mathrm{D}u$ when $u \in \mathrm{C}^1(\mathbb{R})$.

Example 4.2. *(Multiplication by smooth functions)* For $f \in \mathrm{C}^\infty(\mathbb{R})$ define $T(\varphi) := f\varphi$ for each $\varphi \in \mathscr{D}(\mathbb{R})$. Clearly $T\colon \mathscr{D}(\mathbb{R}) \to \mathscr{D}(\mathbb{R})$ is linear and $S = T$ trivially yields the adjoint identity:

$$\int_{\mathbb{R}} f\varphi\psi \,\mathrm{d}x = \int_{\mathbb{R}} \varphi f\psi \,\mathrm{d}x$$

for $\varphi,\, \psi \in \mathscr{D}(\mathbb{R})$. It is clear that $S\colon \mathscr{D}(\mathbb{R}) \to \mathscr{D}(\mathbb{R})$ is linear and $\mathscr{D}$-continuous (checked by Leibniz), so we may extend T to distributions by the rule

$$\langle fu,\, \psi\rangle := \langle u,\, f\psi\rangle$$

for $u \in \mathscr{D}'(\mathbb{R})$, $\psi \in \mathscr{D}(\mathbb{R})$. Clearly we have consistency here: when $u \in \mathrm{L}^1_{\mathrm{loc}}(\mathbb{R})$, then $fu \in \mathrm{L}^1_{\mathrm{loc}}(\mathbb{R})$ and fu can be identified with the above distribution. What the consistency means more precisely can be expressed as

$$T_{fu} = fT_u \quad \text{when } u \in \mathrm{L}^1_{\mathrm{loc}}(\mathbb{R}).$$

Example 4.3. Many other useful operations admit extensions to distributions. We list some elementary operations:

Translation: $T = \tau_h$ defined by $\tau_h\varphi(x) = \varphi(x+h)$ yields adjoint identity with $S = \tau_{-h}$. Thus for $u \in \mathscr{D}'(\mathbb{R})$, $\tau_h u \in \mathscr{D}'(\mathbb{R})$ is defined by the rule

$$\langle \tau_h u,\, \psi\rangle := \langle u,\, \tau_{-h}\psi\rangle$$

for $\psi \in \mathscr{D}(\mathbb{R})$.

Dilation: $T = d_r$ defined by $d_r\varphi(x) = \varphi(rx)$, $r > 0$, yields the adjoint identity with $S = \frac{1}{r}d_{\frac{1}{r}}$. Thus for $u \in \mathscr{D}'(\mathbb{R})$, $d_r u \in \mathscr{D}'(\mathbb{R})$ is defined by the rule

$$\langle d_r u,\, \psi\rangle := \left\langle u,\, \frac{1}{r}d_{\frac{1}{r}}\psi\right\rangle$$

for $\psi \in \mathscr{D}(\mathbb{R})$.

Reflection through the origin: $(T\varphi)(x) = \tilde{\varphi}(x) = \varphi(-x)$ admits the adjoint identity with $S = T$. Thus for $u \in \mathscr{D}'(\mathbb{R})$, $\tilde{u} \in \mathscr{D}'(\mathbb{R})$ is defined by the rule

$$\langle \tilde{u},\, \psi\rangle := \langle u,\, \tilde{\psi}\rangle$$

for $\psi \in \mathscr{D}(\mathbb{R})$.

Push-forward/composition by C^∞ diffeomorphism: A function $\Phi\colon \mathbb{R} \to \mathbb{R}$ is called a C^∞ diffeomorphism if Φ is C^∞, bijective and $\Phi'(x) \neq 0$ for all $x \in \mathbb{R}$. Using the usual substitution formula we have for test functions φ, $\psi \in \mathscr{D}(\mathbb{R})$ that

$$\int_{-\infty}^{\infty} \varphi(\Phi(x))\psi(x)\,\mathrm{d}x = \int_{-\infty}^{\infty} \varphi(y)\frac{\psi(\Phi^{-1}(y))}{\left|\Phi'\left(\Phi^{-1}(y)\right)\right|}\,\mathrm{d}y.$$

Since $\psi \circ \Phi^{-1}/\left|\Phi' \circ \Phi^{-1}\right| \in \mathscr{D}(\mathbb{R})$ we have an adjoint identity with $T(\varphi) = \varphi \circ \Phi$ and $S(\psi) = \psi \circ \Phi^{-1}/\left|\Phi' \circ \Phi^{-1}\right|$. We sometimes also denote this operation by $\Phi_*\varphi$, that is, $\Phi_*\varphi = \varphi \circ \Phi$ Using the chain and Leibniz rules we check that $S\colon \mathscr{D}(\mathbb{R}) \to \mathscr{D}(\mathbb{R})$ is $\mathscr{D}(\mathbb{R})$-continuous and so may extend Φ_* to distributions $u \in \mathscr{D}'(\mathbb{R})$ by the rule

$$\langle \Phi_* u, \varphi\rangle := \langle u, \frac{\varphi \circ \Phi^{-1}}{\left|\Phi' \circ \Phi^{-1}\right|}\rangle$$

for $\varphi \in \mathscr{D}(\mathbb{R})$. Note that we have consistency for $u \in \mathrm{L}^1_{\mathrm{loc}}(\mathbb{R})$ in the sense that $\Phi_* T_u = T_{u\circ\Phi}$. Indeed for $\varphi \in \mathscr{D}(\mathbb{R})$ we have by integration by substitution:

$$\begin{aligned}\langle \Phi_* T_u, \varphi\rangle &= \int_{-\infty}^{\infty} u\frac{\varphi \circ \Phi^{-1}}{\left|\Phi' \circ \Phi^{-1}\right|}\,\mathrm{d}x\\ &= \int_{-\infty}^{\infty} u \circ \Phi\varphi\,\mathrm{d}y = \langle T_{u\circ\Phi}, \varphi\rangle.\end{aligned}$$

We recover translation, dilation and reflection through the origin as special cases when we take $\Phi(x)$ to be $x + h$, rx and $-x$, respectively.

Example 4.4. *(Convolution with a test function)* For $\theta \in \mathscr{D}(\mathbb{R})$, $T\varphi = \theta * \varphi$ admits an adjoint identity with $S\psi = \tilde{\theta} * \psi$. Indeed, by Fubini,

$$\begin{aligned}\int_{-\infty}^{\infty} (\theta * \varphi)(x)\psi(x)\,\mathrm{d}x &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \theta(x-y)\varphi(y)\,\mathrm{d}y\psi(x)\,\mathrm{d}x \\ &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \theta(x-y)\psi(x)\,\mathrm{d}x\varphi(y)\,\mathrm{d}y \\ &= \int_{-\infty}^{\infty} (\tilde{\theta} * \psi)(y)\varphi(y)\,\mathrm{d}y.\end{aligned}$$

Thus for $u \in \mathscr{D}'(\mathbb{R})$, $\theta * u \in \mathscr{D}'(\mathbb{R})$ is defined by the rule

$$\langle \theta * u,\, \psi\rangle := \langle u,\, \tilde{\theta} * \psi\rangle$$

for $\psi \in \mathscr{D}(\mathbb{R})$. On Sheet 3 you will be asked to prove that $\theta * u \in \mathrm{C}^{\infty}(\mathbb{R})$.

Let us highlight the definitions of differentiation, multiplication by smooth functions and convolution with test function in n dimensions.

4.1.1 Differentiation in the sense of distributions

Definition 4.5. *Let Ω be a non-empty open subset of $\mathbb{R}^n$. Let $u \in \mathscr{D}'(\Omega)$ and $j \in \{1, \ldots, n\}$. The j-th partial derivative of u, $\bar{D}_j u$, in the sense of distributions is defined by the rule*

$$\langle \bar{D}_j u,\, \varphi\rangle := \langle u,\, -D_j\varphi\rangle$$

for $\varphi \in \mathscr{D}(\Omega)$.

Note that D_j fits into the adjoint identity scheme with $T = D_j$ and $S = -D_j$, and so is well-defined. Also note that $\bar{D}_j$ is continuous in the sense that if $u_k \to u$ in $\mathscr{D}'(\Omega)$, then $\bar{D}_j u_k \to \bar{D}_j u$ in $\mathscr{D}'(\Omega)$. As in the one-dimensional case, when $u \in \mathrm{C}^1(\Omega)$ the distributional $\bar{D}_1 u, \ldots, \bar{D}_n u$ and the classical partial derivatives $D_1 u, \ldots, D_n u$ coincide. In view of this we shall henceforth use the same notation for distributional derivatives as for the corresponding classical derivatives, so any of the symbols $D_j u = \partial_j u = \partial u/\partial x_j = u_{x_j}$ etc can stand for both the usual partial derivative and for the distributional partial derivative. What is meant exactly will be clear from context (and if it is not, then it will not matter because the derivatives can be identified). Moreover, note that since for $\varphi \in \mathscr{D}(\Omega)$ we have

$$\frac{\partial^2 \varphi}{\partial x_j \partial x_k} = \frac{\partial^2 \varphi}{\partial x_k \partial x_j},$$

we also have $\partial_j\partial_k u = \partial_k\partial_j u$ for $u \in \mathscr{D}'(\Omega)$. We can therefore also use multi-index notation for distributional derivatives. For $u \in \mathscr{D}'(\Omega)$ and $\alpha \in \mathbb{N}_0^n$ we have

$$\langle \partial^{\alpha} u,\, \varphi\rangle = (-1)^{|\alpha|}\langle u,\, \partial^{\alpha}\varphi\rangle$$

for $\varphi \in \mathscr{D}(\Omega)$, where we recall that $\alpha = (\alpha_1, \dots, \alpha_n)$, $|\alpha| = \alpha_1 + \cdots + \alpha_n$, and

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

We emphasize that $u \mapsto \partial^\alpha u$ is $\mathscr{D}'(\Omega)$-continuous.

When $u \in \mathscr{D}'(\Omega)$ we define its *distributional gradient* to be

$$\nabla u := \big(\partial_1 u,\, \dots\,,\, \partial_n u\big).$$

Thus ∇u is an example of a vector valued distribution, meaning it is of the form $v = \big(v_1,\, \dots\,,\, v_n\big)$ where each component $v_j \in \mathscr{D}'(\Omega)$ is a distribution. The set of vector valued distributions is denoted by $\mathscr{D}'(\Omega)^n$. For such distributions v we can define the *distributional divergence* as

$$\mathrm{div} v := \partial_1 v_1 + \, \dots \, + \partial_n v_n.$$

As in the usual vector calculus we have the relation that $\Delta u = \mathrm{div}\nabla u$ when $u \in \mathscr{D}'(\Omega)$, where now $\Delta u := \partial_1^2 u + \, \dots \, + \partial_n^2 u$ is the *distributional Laplacian.*

4.1.2 Multiplication by smooth function

Definition 4.6. *Let u be a distribution and f be a smooth function. Then the product fu in the sense of distributions is defined by the rule*

$$\langle fu,\, \varphi\rangle := \langle u,\, f\varphi\rangle$$

for $\varphi \in \mathscr{D}(\Omega)$.

This definition also fits into the adjoint identity scheme with $T(\phi) = f\phi = S(\phi)$ and so is well-defined. It is clearly consistent for $\mathrm{L}^1_{\mathrm{loc}}$ functions, as in the one-dimensional case. It is clear that we can define the product uf too and that we always have $fu = uf$.

Example 4.7. The *Heaviside function* is the function

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

Note that the value of $H(x)$ at $x = 0$ is not particularly important and is sometimes taken to be 0 instead (or in some other contexts even $\frac{1}{2}$). Clearly $H \in \mathrm{L}^1_{\mathrm{loc}}(\mathbb{R})$, so $H \in \mathscr{D}'(\mathbb{R})$ and we have $H' = \delta_0$. We check the latter and calculate for $\varphi \in \mathscr{D}(\mathbb{R})$:

$$\begin{aligned}
\langle H',\, \varphi\rangle &= \langle H,\, -\varphi'\rangle \\
&= \int_{-\infty}^{\infty} H(x)(-\varphi'(x))\,\mathrm{d}x \\
&= -\int_0^{\infty} \varphi'(x)\,\mathrm{d}x \\
&\overset{\mathrm{FTC}}{=} -\left[\varphi(x)\right]_{x=0}^{x\to\infty} = \varphi(0) = \langle \delta_0,\, \varphi\rangle.
\end{aligned}$$

Note also that by iteration of the above we find for $m \in \mathbb{N}$ that

$$\left\langle \frac{\mathrm{d}^m}{\mathrm{d}x^m}\delta_0,\, \varphi \right\rangle = \langle \delta_0,\, (-1)^m \varphi^{(m)} \rangle = (-1)^m \varphi^{(m)}(0).$$

A slight extension of the above formula for H' is obtained by differentiation of a piecewise C^1 function

$$h(x) = \begin{cases} f(x) & x < 0 \\ g(x) & x \geqslant 0, \end{cases}$$

where $f,\, g \in \mathrm{C}^1(\mathbb{R})$. This will be addressed on Sheet 2.

Example 4.8. We can define $\triangle_h = \tau_h - \mathbf{1}$ for $h \in \mathbb{R}$ on distributions $u \in \mathscr{D}'(\mathbb{R})$ by the adjoint identity scheme: for $\varphi \in \mathscr{D}(\mathbb{R})$ put

$$\langle \triangle_h u,\, \varphi \rangle = \big\langle u,\, \big(\tau_{-h} - \mathbf{1}\big)\varphi \big\rangle\,,$$

where $\big(\tau_{-h} - \mathbf{1}\big)\varphi(x) = \varphi(x-h) - \varphi(x)$. If $u \in \mathrm{C}^1(\mathbb{R})$, then clearly we have convergence

$$\frac{\triangle_h u(x)}{h} = \frac{u(x+h) - u(x)}{h} \underset{h \to 0}{\longrightarrow} u'(x)$$

locally uniformly in x. What happens when $u \in \mathscr{D}'(\mathbb{R})$? Recall that according to Example 2.18 we have

$$\frac{\tau_{-h} - \mathbf{1}}{h}\varphi \underset{h \to 0}{\longrightarrow} -\varphi' \quad \text{in } \mathscr{D}(\mathbb{R})$$

and therefore $\triangle_h u / h \underset{h \to 0}{\longrightarrow} \mathrm{D}u$ in $\mathscr{D}'(\mathbb{R})$.

Theorem 4.9 (Leibniz Rule)**.** *If $u \in \mathscr{D}'(\Omega)$, $f \in \mathrm{C}^\infty(\Omega)$, and $j \in \{1, \ldots, n\}$, then*

$$\partial_j(fu) = (\partial_j f)u + f\partial_j u$$

in $\mathscr{D}'(\Omega)$. In fact, the Generalized Leibniz Rule also holds for distributions: for a multi-index $\alpha \in \mathbb{N}_0^n$,

$$\partial^\alpha(fu) = \sum_{\beta \leqslant \alpha} \binom{\alpha}{\beta} \partial^\beta f \partial^{\alpha - \beta} u.$$

Proof. We only prove the basic case, the general case can be proved by induction, or simply by using the formula for test functions. First note that $\partial_j(fu)$, $(\partial_j f)u + f\partial_j u \in \mathscr{D}'(\Omega)$ and that for $\varphi \in \mathscr{D}(\Omega)$:

$$\langle \partial_j(fu),\, \varphi \rangle = \langle fu,\, -\partial_j \varphi \rangle = \langle u,\, -f\partial_j \varphi \rangle,$$

$$\begin{aligned} \langle (\partial_j f)u + f\partial_j u,\, \varphi \rangle &= \langle (\partial_j f)u,\, \varphi \rangle + \langle f\partial_j u,\, \varphi \rangle \\ &= \langle u,\, (\partial_j f)\varphi \rangle + \langle \partial_j u,\, f\varphi \rangle \\ &= \langle u,\, (\partial_j f)\varphi \rangle + \langle u,\, -\partial_j(f\varphi) \rangle \\ &= \langle u,\, (\partial_j f)\varphi - \partial_j(f\varphi) \rangle \\ &= \langle u,\, -f\partial_j \varphi \rangle, \end{aligned}$$

and we are done. □

4.1.3 Convolution with test function

Definition 4.10. *Let $u \in \mathscr{D}'(\mathbb{R}^n)$ and $\theta \in \mathscr{D}(\mathbb{R}^n)$. Then the convolution $u * \theta$ is defined by*

$$\langle u * \theta, \varphi \rangle := \langle u, \tilde{\theta} * \varphi \rangle, \quad \varphi \in \mathscr{D}(\mathbb{R}^n),$$

where, as in the 1-dimensional case, we define $\tilde{\theta}(x) := \theta(-x)$.

Using the adjoint identity scheme it is easy to check that hereby $u * \theta \in \mathscr{D}'(\mathbb{R}^n)$. Because addition is commutative on $\mathbb{R}^n$ we have that $u * \theta = \theta * u$ (with the obvious definition of the right-hand side).

Example 4.11. If we follow the trail of definitions we easily obtain the following result: for $u \in \mathscr{D}'(\mathbb{R}^n)$, $\theta \in \mathscr{D}(\mathbb{R}^n)$ and any multi-index $\alpha \in \mathbb{N}_0^n$ we have in the sense of distributions on $\mathbb{R}^n$ that

$$\partial^\alpha (u * \theta) = (\partial^\alpha u) * \theta = u * (\partial^\alpha \theta).$$

4.2 Mollification and approximation of distributions.

If we convolve the distribution $u \in \mathscr{D}'(\mathbb{R}^n)$ with ρ_ε from the standard mollifier we obtain the so-called *mollified distribution* $u * \rho_\varepsilon$:

$$\langle u * \rho_\varepsilon, \varphi \rangle = \langle u, \rho_\varepsilon * \varphi \rangle, \quad \varphi \in \mathscr{D}(\mathbb{R}^n) \tag{21}$$

Note that we have used $\tilde{\rho}_\varepsilon = \rho_\varepsilon$ to simplify the above formula.

Lemma 4.12. *Let $u \in \mathscr{D}'(\mathbb{R}^n)$ and $\left(\rho_\varepsilon\right)_{\varepsilon > 0}$ be the standard mollifier on $\mathbb{R}^n$. Then, for each $\varepsilon > 0$, $u * \rho_\varepsilon \in \mathrm{C}^\infty(\mathbb{R}^n)$,*

$$(u * \rho_\varepsilon)(x) = \big\langle u, \rho_\varepsilon(x - \cdot) \big\rangle,$$

that is, u acting on the test function $y \mapsto \rho_\varepsilon(x - y)$. Furthermore, $u * \rho_\varepsilon \to u$ in $\mathscr{D}'(\mathbb{R}^n)$ as $\varepsilon \searrow 0$.

Remark 4.13. Inspection of the proof below reveals that the only properties of the standard mollifier kernel ρ on $\mathbb{R}^n$ that are really needed are that $\rho \in \mathscr{D}(\mathbb{R}^n)$ and $\int_{\mathbb{R}^n} \rho \,\mathrm{d}x = 1$. Thus the result of Lemma 4.12 remains true if we replace the standard mollifier by the family $\left(\theta_\varepsilon\right)_{\varepsilon > 0}$, where

$$\theta_\varepsilon(x) := \frac{1}{\varepsilon^n} \theta\left(\frac{x}{\varepsilon}\right)$$

and $\theta \in \mathscr{D}(\mathbb{R}^n)$ and $\int_{\mathbb{R}^n} \theta(x) \,\mathrm{d}x = 1$.

Proof. The proof starts by observing that, for each fixed $\varepsilon > 0$, the convolution $\rho_\varepsilon * \varphi$ appearing on the right-hand side of (21) can be calculated as a Riemann integral and therefore obtained as a limit of Riemann sums in a very strong sense. In order to make this precise we shall introduce some convenient notation. For $k \in \mathbb{N}$ we understand by a *k-th generation dyadic cube* a cube Q of the form

$$Q = c^Q + (0, 2^{-k}]^n = (c_1^Q, c_1^Q + 2^{-k}] \times \ldots \times (c_n^Q, c_n^Q + 2^{-k}]$$

where its left cornerpoint c^Q belongs to the dilated integer grid $2^{-k}\mathbb{Z}^n$. The collection of all k-th generation dyadic cubes is denoted by $\mathcal{D}_k$ and we clearly have that $\mathbb{R}^n = \bigcup \mathcal{D}_k$ as a disjoint union. Now for each $x \in \mathbb{R}^n$ and $k \in \mathbb{N}$ the (finite!) sum

$$R_k(x) := \sum_{Q \in \mathcal{D}_k} \rho_\varepsilon(x - c^Q)\varphi(c^Q)\mathscr{L}^n(Q)$$

is a Riemann sum for $(\rho_\varepsilon * \varphi)(x)$. It is easy to check, using uniform continuity, that $R_k(x) \to (\rho_\varepsilon * \varphi)(x)$ uniformly in $x \in \mathbb{R}^n$ as $k \to \infty$. We assert that R_k are test functions and that the convergence in fact takes place in the sense of test functions. It is clear that R_k are C^∞ and if we let $K = \overline{B_\varepsilon(\mathrm{supp}(\varphi))}$, then K is a compact set with the property that $\mathrm{supp}(R_k) \subset K$ for all $k \in \mathbb{N}$. Next if $\alpha \in \mathbb{N}_0^n$ is a multi-index, then we have

$$(\partial^\alpha R_k)(x) = \sum_{Q \in \mathcal{D}_k} (\partial^\alpha \rho_\varepsilon)(x - c^Q)\varphi(c^Q)\mathscr{L}^n(Q) \to \big((\partial^\alpha \rho_\varepsilon) * \varphi\big)(x)$$

uniformly in $x \in \mathbb{R}^n$ as $k \to \infty$ by uniform continuity of $\partial^\alpha \rho_\varepsilon$ and φ. But this is exactly the asserted convergence in $\mathscr{D}(\mathbb{R}^n)$. Now note that by linearity of u we have

$$\langle u, R_k \rangle = \sum_{Q \in \mathcal{D}_k} \big\langle u, \rho_\varepsilon(\cdot - c^Q)\big\rangle \varphi(c^Q)\mathscr{L}^n(Q). \tag{22}$$

Because u is $\mathscr{D}'(\mathbb{R}^n)$-continuous we know that the limit of the left-hand side equals the right-hand side in (21). In order to deal with the right-hand side of (22) we rely on the following.

Auxiliary Lemma. *Let $u \in \mathscr{D}'(\mathbb{R}^n)$ and $\theta \in \mathscr{D}(\mathbb{R}^n)$. Then the function*

$$h(x) := \langle u, \theta(x - \cdot) \rangle$$

is C^1 and $\partial_j h(x) = \langle u, (\partial_j \theta)(x - \cdot) \rangle$ for each $1 \leqslant j \leqslant n$.

Sketch of proof for Auxiliary Lemma. The details are left as an exercise (or see Theorem 5.9 below for a general result of this type). Let $(e_j)_{j=1}^n$ be the standard basis for $\mathbb{R}^n$ and consider the difference quotient

$$\frac{h(x + te_j) - h(x)}{t} = \left\langle u, \frac{\theta(x + te_j - \cdot) - \theta(x - \cdot)}{t} \right\rangle \quad t \in \mathbb{R} \setminus \{0\}.$$

Show that for fixed $x \in \mathbb{R}^n$,

$$\frac{\theta(x + te_j - \cdot) - \theta(x - \cdot)}{t} \to (\partial_j \theta)(x - \cdot) \text{ in } \mathscr{D}(\mathbb{R}^n) \text{ as } t \to 0.$$

Next observe that $x \mapsto (\partial_j\theta)(x-\cdot)$ is continuous from $\mathbb{R}^n$ into $\mathscr{D}(\mathbb{R}^n)$. $\square$

We return to the right-hand side of (22) and apply an induction argument using the Auxiliary Lemma for the base case and in each induction step. Hereby we deduce that $x \mapsto \langle u, \rho_\varepsilon(x-\cdot)\rangle = \langle u, \rho_\varepsilon(\cdot-x)\rangle$ is C^k for any $k \in \mathbb{N}$ with $\partial_x^\alpha \langle u, \rho_\varepsilon(x-\cdot)\rangle = \langle u, (\partial^\alpha \rho_\varepsilon)(x-\cdot)\rangle$ for all multi-indices α. Finally, since $x \mapsto \langle u, \rho_\varepsilon(x-\cdot)\rangle\varphi(x)$ is uniformly continuous and compactly supported we may pass to the limit $k \to \infty$ in (22) whereby we find

$$\langle u, \rho_\varepsilon * \varphi\rangle = \int_{\mathbb{R}^n} \langle u, \rho_\varepsilon(x-\cdot)\rangle\varphi(x)\,\mathrm{d}x.$$

This concludes the proof. $\square$

In fact we can do a little better and even approximate a general distribution by test functions:

Lemma 4.14. *Let $u \in \mathscr{D}'(\mathbb{R}^n)$. Then we can find a sequence (u_j) of test functions so $u_j \to u$ in $\mathscr{D}'(\mathbb{R}^n)$.*

Proof. We have from Lemma 4.12 that $u * \rho_\varepsilon \in \mathrm{C}^\infty(\mathbb{R}^n)$ and $u * \rho_\varepsilon \to u$ in $\mathscr{D}'(\mathbb{R}^n)$ as $\varepsilon \searrow 0$. The only issue is that $u * \rho_\varepsilon$ will not in general have compact support. We fix this as follows: By virtue of Theorem 2.12 we may find $\chi \in \mathscr{D}(B_2(0))$ with $\chi = 1$ on $B_1(0)$. We extend χ to $\mathbb{R}^n \setminus B_2(0)$ by 0 and put $\chi^\varepsilon(x) := \chi(\varepsilon x)$, $x \in \mathbb{R}^n$. Define

$$u_\varepsilon := \chi^\varepsilon u * \rho_\varepsilon.$$

Clearly $\operatorname{supp}(u_\varepsilon) \subseteq \overline{B_{2/\varepsilon}(0)}$ so $u_\varepsilon \in \mathscr{D}(\mathbb{R}^n)$ and since for $\varphi \in \mathscr{D}(\mathbb{R}^n)$ we have

$$\langle u_\varepsilon, \varphi\rangle = \langle u * \rho_\varepsilon, \chi^\varepsilon\varphi\rangle$$

and $\chi^\varepsilon\varphi = \varphi$ when $\operatorname{supp}(\varphi) \subset B_{1/\varepsilon}(0)$ the proof is finished. (The sequence (u_j) can evidently be obtained from the family (u_ε) if we take $u_j := u_{\varepsilon_j}$ for any choice of null sequence $\varepsilon_j \searrow 0$.) $\square$

The last result can be elaborated further to apply also to distributions defined on arbitrary open sets.

Theorem 4.15. *Let Ω be a non-empty open subset of $\mathbb{R}^n$ and $u \in \mathscr{D}'(\Omega)$. Then there exists a sequence (u_j) in $\mathscr{D}(\Omega)$ so $u_j \to u$ in $\mathscr{D}'(\Omega)$.*

We have all the necessary ingredients for the proof but omit the details here. (In fact, from the perspective of Functional Analysis in the context of topological vector spaces this density result is not surprising at all. But since we have preferred to give a hands-on approach and avoid the abstract general theory we of course have to work a bit more occasionally.)

It is however interesting to return to the adjoint identity scheme for extending operations on test functions to operations on distributions. Recall that the starting point is that we have an operation on test functions, a linear map $T\colon \mathscr{D}(\Omega) \to \mathscr{D}(\Omega)$, that we would like to extend to distributions $u \in \mathscr{D}'(\Omega)$. We saw that such an extension was indeed possible if we had another linear and $\mathscr{D}(\Omega)$-continuous map $S\colon \mathscr{D}(\Omega) \to \mathscr{D}(\Omega)$ satisfying the adjoint identity (19). The extension $\bar{T}$ was then given in (20). As remarked before, this extension $\bar{T}\colon \mathscr{D}'(\Omega) \to \mathscr{D}'(\Omega)$ is linear and $\mathscr{D}'(\Omega)$-continuous. In fact, it follows now from Theorem 4.15 that $\bar{T}$ is the extension of $T = \bar{T}|_{\mathscr{D}(\Omega)}$ to $\mathscr{D}'(\Omega)$ by $\mathscr{D}'(\Omega)$-continuity. This gives rise to an often useful procedure for proving results about distributions: first establish the desired result for test functions (or for smooth functions) and then extend it to distributions by approximation. We shall see an example of this when we prove the constancy theorem for distributions in Section 5.

4.3 The Gauss-Green formula and some of its consequences.

Most of you will have seen a statement of the divergence theorem in a multi-variate calculus course, but these courses often do not provide any proof of the result. Here we also do not give a proof. Instead the aim is to give a precise statement of a basic case that suffices for the applications in this course. We start by giving some requisite definitions.

Definition 4.16. *Let Ω be a non-empty open subset of $\mathbb{R}^n$. An open subset ω of Ω is called a* C^1 subset of Ω *if there exists $\Psi \in \mathrm{C}^1(\Omega)$ so*

$$\omega = \{x \in \Omega : \Psi(x) < 0\}, \quad \Omega \cap \partial\omega = \{x \in \Omega : \Psi(x) = 0\}$$

and $\nabla\Psi(x) \neq 0$ for all $x \in \Omega \cap \partial\omega$.

The function Ψ is called a defining function for ω. *Its gradient $\nabla\Psi(x)$ is an outward pointing normal to ω at each point $x \in \Omega \cap \partial\omega$, hence*

$$\nu(x) := \frac{\nabla\Psi(x)}{|\nabla\Psi(x)|}$$

is the outward unit normal to ω *at each point $x \in \Omega \cap \partial\omega$.*

Remark 4.17. *Comment on surface integrals.* The condition on the defining function Ψ that $\nabla\Psi(y) \neq 0$ at each $y \in \Omega \cap \partial\omega$ is essential and implies (by the Implicit Function Theorem) that in some small ball $B_r(y) \Subset \Omega$ the boundary $B_r(y) \cap \partial\omega$ admits a C^1 parametrization. More precisely, if for instance $\partial_1\Psi(y) \neq 0$, then writing $x = (x_1, x')$ we can find a C^1 function $\psi = \psi(x')$ so

$$\begin{aligned} B_r(y) \cap \omega &= \{x \in \mathbb{R}^n : |x - y| < r \text{ and } x_1 < \psi(x')\}, \\ B_r(y) \cap \partial\omega &= \{(\psi(x'), x') : |x' - y'| < r\}. \end{aligned}$$

In terms of this parametrization we define for a continuous function $f\colon \overline{B_r(y)} \to \mathbb{C}$ the *surface integral*

$$\int_{B_r(y)\cap\partial\omega} f(x)\,\mathrm{d}S_x := \int_{\{x' \in \mathbb{R}^{n-1}\colon |x'|<r\}} f(\psi(x'), x')\sqrt{1 + |\nabla\psi(x')|^2}\,\mathrm{d}x'. \tag{23}$$

Using a change-of-variables formula it can be shown that the value of the integral on the right-hand side does not depend on the particular parametrization used. The surface integral of a continuous and compactly supported function $f\colon \Omega \to \mathbb{C}$ is then defined by use of a partition of unity to add up the finitely many contributions (23) in the support of f (again, one can show that this will not depend on how we cover the support of f with small balls where we have local parametrizations or on the used partition of unity).

Example 4.18. The open ball $B_R(y)$ is a C^1 subset of $\mathbb{R}^n$: we can use $\Psi(x) = |x-y|^2 - R^2$, $x \in \mathbb{R}^n$, as a defining function. Note that the outward unit normal is

$$\nu(x) = \frac{x-y}{R} \quad x \in \partial B_R(y).$$

In this case one can also use *polar coordinates* about y to calculate the surface integral over the sphere $\partial B_R(y)$.

The ring domain $A = B_R(y) \setminus \overline{B_r(y)}$, $0 < r < R$, is a C^1 subset of $\mathbb{R}^n$: we can use $\Psi(x) = \big(|x-y|^2 - R^2\big)\big(r^2 - |x-y|^2\big)$, $x \in \mathbb{R}^n$, as a defining function. Note that the outward unit normal is

$$\nu(x) = \begin{cases} \dfrac{x-y}{R} & \text{for } |x-y| = R \\[2ex] -\dfrac{x-y}{r} & \text{for } |x-y| = r. \end{cases}$$

Theorem 4.19. *(The Gauss-Green Formula.) Let Ω be a non-empty open subset of $\mathbb{R}^n$ and ω a C^1 subset of Ω. Then for each $V \in \mathscr{D}(\Omega)^n$ we have*

$$\int_\omega \mathrm{div} V \, \mathrm{d}x = \int_{\partial\omega} V \cdot \nu \, \mathrm{d}S_x.$$

Remark 4.20. The result is sometimes also called the divergence theorem, and in fact it holds in a much more general form. However, the above will suffice for the applications we have in mind. Note that for $n = 1$ and $\omega = (a,b)$ it reduces to the fundamental theorem of calculus.

We end this subsection by calculating some important distributional derivatives.

Example 4.21. *The distributional derivative of the indicator function of a C^1 subset ω of Ω:*

$$\nabla \mathbf{1}_\omega = -\nu \, \mathrm{d}S_x. \tag{24}$$

Of course we can express this in coordinates as $\partial_j \mathbf{1}_\omega = -\nu_j \, \mathrm{d}S_x$ for each $1 \leqslant j \leqslant n$. Indeed, if $\varphi \in \mathscr{D}(\Omega)$ we use the divergence theorem to calculate with $V = \varphi e_j$ that

$$\begin{aligned} \langle \partial_j \mathbf{1}_\omega, \varphi \rangle &= -\langle \mathbf{1}_\omega, \partial_j \varphi \rangle \\ &= -\int_\omega \mathrm{div} V \, \mathrm{d}x \\ &= -\int_{\partial\omega} V \cdot \nu \, \mathrm{d}S_x \\ &= -\int_{\partial\omega} \varphi \nu_j \, \mathrm{d}S_x \end{aligned}$$

and hence the conclusion.

Example 4.22. Let $y \in \mathbb{R}^n$ and denote

$$G_y(x) = G_y^n(x) := \begin{cases} -\frac{1}{(n-2)\omega_{n-1}}|x-y|^{2-n} & \text{if } n \in \mathbb{N} \setminus \{2\} \\ \frac{1}{\omega_1}\log|x-y| & \text{if } n = 2, \end{cases}$$

where we recall that ω_{n-1} denotes the $(n-1)$-dimensional area of the unit sphere $\mathbb{S}^{n-1}$ in $\mathbb{R}^n$ (so that $\omega_0 = 2$, $\omega_1 = 2\pi$, $\omega_2 = 4\pi$). We emphasize the 2-dimensional case

$$G_y^2(x) = \frac{1}{4\pi}\log\bigl((x_1-y_1)^2 + (x_2-y_2)^2\bigr)$$

that is known as the *logarithmic potential* and the 3-dimensional case

$$G_y^3(x) = -\frac{1}{4\pi}\frac{1}{\sqrt{(x_1-y_1)^2 + (x_2-y_2)^2 + (x_3-y_3)^2}}$$

that is known as the *Newtonian potential*.

Clearly, regardless of the dimension n, G_y is C^∞ away from y and it is routine to check that $\Delta G_y = 0$ there. If we use polar coordinates with centre at y it is not difficult to see that $G_y \in \mathrm{L}^1_{\mathrm{loc}}(\mathbb{R}^n)$, so we can consider G_y as a (regular) distribution on $\mathbb{R}^n$. We assert that

$$\Delta G_y = \delta_y \quad \text{in } \mathscr{D}'(\mathbb{R}^n).$$

We can assume that $y = 0$ and must then prove that

$$I := \int_{\mathbb{R}^n} \Delta\varphi G_0 \,\mathrm{d}x = \varphi(0)$$

holds for all $\varphi \in \mathscr{D}(\mathbb{R}^n)$. Fix $\varphi \in \mathscr{D}(\mathbb{R}^n)$. First we note that if for each $r > 0$ we put

$$I(r) := \int_{|x|>r} \Delta\varphi G_0 \,\mathrm{d}x,$$

then $I(r) \to I$ as $r \searrow 0$. Next, because φ has compact support we may find $R > 0$ so $\mathrm{supp}(\varphi) \subset B_R(0)$, hence with $A := \{x \in \mathbb{R}^n : r < |x| < R\}$ we have

$$I(r) = \int_A \Delta\varphi G_0 \,\mathrm{d}x.$$

Now G_0 is C^∞ on A and $\Delta G_0 = 0$ there, so using the divergence theorem and that $\nabla\varphi(x) = 0$ for $|x| = R$ we find

$$\begin{aligned} I(r) = \int_A \Delta\varphi G_0 \,\mathrm{d}x &= -\int_A \nabla\varphi \cdot \nabla G_0 \,\mathrm{d}x + \int_{|x|=r} G_0 \nabla\varphi \cdot \nu \,\mathrm{d}S_x \\ &= -\int_{|x|=r} \varphi \nabla G_0 \cdot \nu \,\mathrm{d}S_x + \int_{|x|=r} G_0 \nabla\varphi \cdot \nu \,\mathrm{d}S_x, \end{aligned}$$

where ν is the outward unit normal on the boundary of A. On the sphere $|x| = r$, we have $\nabla G_0 \cdot \nu = -\frac{1}{\omega_{n-1}} r^{1-n}$ and therefore

$$-\int_{|x|=r} \varphi \nabla G_0 \cdot \nu \, \mathrm{d}S_x = \tfrac{1}{\omega_{n-1}} \int_{\mathbb{S}^{n-1}} \varphi(rx) \, \mathrm{d}S_x \to \varphi(0) \text{ as } r \searrow 0.$$

The last integral vanishes in the limit $r \searrow 0$:

$$\begin{aligned} \left| \int_{|x|=r} G_0 \nabla \varphi \cdot \nu \, \mathrm{d}S_x \right| &\leqslant \tfrac{1}{|n-2|\omega_{n-1}} r^{2-n} \sup |\nabla \varphi| \omega_{n-1} r^{n-1} \\ &= \tfrac{\sup |\nabla \varphi|}{|n-2|} r \to 0. \end{aligned}$$

Example 4.23. The Cauchy-Riemann differential operators

$$\frac{\partial}{\partial z} := \frac{1}{2}\left(\frac{\partial}{\partial x} - \mathrm{i}\frac{\partial}{\partial y}\right), \quad \frac{\partial}{\partial \bar{z}} := \frac{1}{2}\left(\frac{\partial}{\partial x} + \mathrm{i}\frac{\partial}{\partial y}\right)$$

act on distributions defined on open subsets of $\mathbb{C}$. As usual we identify $\mathbb{C} \simeq \mathbb{R}^2$ corresponding to $x + \mathrm{i}y \simeq (x, y)$. Since $\Delta = 4\partial^2/\partial\bar{z}\partial z$ and $\Delta\big(\log|z|\big) = 2\pi\delta_0$ in $\mathscr{D}'(\mathbb{C})$ we calculate

$$\begin{aligned} \pi\delta_0 &= \frac{\partial^2}{\partial\bar{z}\partial z}\big(\log|z|^2\big) \\ &= \frac{\partial}{\partial\bar{z}}\left(\frac{\partial}{\partial z}\log(z\bar{z})\right) \\ &= \frac{\partial}{\partial\bar{z}}\left(\frac{1}{z}\right) \end{aligned}$$

and consequently,

$$\frac{\partial}{\partial\bar{z}}\left(\frac{1}{\pi z}\right) = \delta_0. \tag{25}$$

Likewise we can show that

$$\frac{\partial}{\partial z}\left(\frac{1}{\pi\bar{z}}\right) = \delta_0. \tag{26}$$

5 Some calculus for distributions

5.1 The basic theorems.

These are the constancy theorem and the fundamental theorem of calculus. We start with the former.

5.1.1 The constancy theorem

Theorem 5.1. *(The constancy theorem) Let Ω be a non-empty open and connected subset of $\mathbb{R}^n$. If $u \in \mathscr{D}'(\Omega)$ and*

$$\nabla u = 0 \ \text{ in } \mathscr{D}'(\Omega)^n,$$

then $u = c$ for some constant $c \in \mathbb{C}$.

Proof. We only give the details for the case $\Omega = \mathbb{R}^n$. The general case can be obtained along the same lines, but is more technical and therefore omitted here.

Let $\left(\rho_\varepsilon\right)_{\varepsilon>0}$ be the standard mollifier. Then we have $\nabla(\rho_\varepsilon * u) = \rho_\varepsilon * \nabla u = 0$ in $\mathscr{D}'(\mathbb{R}^n)^n$. According to Lemma 4.12, $\rho_\varepsilon * u \in \mathrm{C}^\infty(\mathbb{R}^n)$ and $\rho_\varepsilon * u \to u$ in $\mathscr{D}'(\mathbb{R}^n)$ as $\varepsilon \searrow 0$. By consistency of distributional derivatives for C^1 functions we have that $\nabla(\rho_\varepsilon * u) = 0$ on $\mathbb{R}^n$ in the usual sense. But then the usual constancy theorem implies that $\rho_\varepsilon * u$ is constant: $\rho_\varepsilon * u = c_\varepsilon$ on $\mathbb{R}^n$ for some constant $c_\varepsilon \in \mathbb{C}$. Now we calculate for the standard mollifier kernel $\rho\,(= \rho_1)$

$$\begin{aligned} c_\varepsilon = \int_{\mathbb{R}^n} c_\varepsilon \rho \,\mathrm{d}x &= \int_{\mathbb{R}^n} (\rho_\varepsilon * u)\rho \,\mathrm{d}x \\ &= \langle \rho_\varepsilon * u, \rho \rangle \to \langle u, \rho \rangle \ \text{ as } \varepsilon \searrow 0. \end{aligned}$$

Hence if we put $c := \langle u, \rho\rangle$ and take $\varphi \in \mathscr{D}(\mathbb{R}^n)$, then

$$\begin{aligned} c_\varepsilon \int_{\mathbb{R}^n} \varphi \,\mathrm{d}x &= \int_{\mathbb{R}^n} (\rho_\varepsilon * u)\varphi \,\mathrm{d}x \\ &= \langle u, \tilde{\rho}_\varepsilon * \varphi \rangle \to \langle u, \varphi \rangle \ \text{ as } \varepsilon \searrow 0, \end{aligned}$$

and thus $\langle u, \varphi\rangle = c \int_{\mathbb{R}^n} \varphi \,\mathrm{d}x$ as required. □

Example 5.2. *We prove the constancy theorem in a special case by construction of a suitable test function. Variants of this technique are quite common and used also in other contexts.* Assume $u \in \mathscr{D}'(\mathbb{R})$ and that $u' = 0$ in $\mathscr{D}'(\mathbb{R})$. The assumption means, by definition, that

$$0 = \langle u', \varphi \rangle = -\langle u, \varphi' \rangle$$

holds for all $\varphi \in \mathscr{D}(\mathbb{R})$. Let $\rho \in \mathscr{D}(\mathbb{R})$ be the standard mollifier kernel on $\mathbb{R}$. Now for $\varphi \in \mathscr{D}(\mathbb{R})$ we put

$$c_\varphi := \int_{\mathbb{R}} \varphi \,\mathrm{d}x$$

and

$$\psi(x) := \int_{-\infty}^{x} (\varphi(t) - c_\varphi \rho(t)) \,\mathrm{d}t, \quad x \in \mathbb{R}.$$

Take $a, b \in \mathbb{R}$ with $a < b$ so that $\varphi(t) = 0 = \rho(t)$ for $t \leqslant a$ or $t \geqslant b$. Then clearly $\psi(x) = 0$ for $x \leqslant a$, while for $x \geqslant b$,

$$\psi(x) = \int_{-\infty}^{x} (\varphi(t) - c_\varphi \rho(t)) \,\mathrm{d}t = \int_{-\infty}^{\infty} (\varphi(t) - c_\varphi \rho(t)) \,\mathrm{d}t = 0.$$

By the fundamental theorem of calculus, ψ is C^1 with $\psi'(x) = \varphi(x) - c_\varphi \rho(x)$ and hence it follows that in fact ψ is C^∞. Since also $\operatorname{supp}(\psi) \subseteq [a, b]$, we have shown that $\psi \in \mathscr{D}(\mathbb{R})$. Now

$$\begin{aligned}\langle u,\, \varphi\rangle &= \langle u,\, \psi' + c_\varphi \rho\rangle \\ &= \langle u,\, \psi'\rangle + c_\varphi \langle u,\, \rho\rangle \\ &= \langle -u',\, \psi\rangle + c_\varphi \langle u,\, \rho\rangle \\ &= c_\varphi \langle u,\, \rho\rangle,\end{aligned}$$

so

$$\langle u,\, \varphi\rangle = \langle u,\, \rho\rangle \int_{\mathbb{R}} \varphi(x)\,\mathrm{d}x = c \int_{\mathbb{R}} \varphi(x)\,\mathrm{d}x,$$

where we denoted $c = \langle u,\, \rho\rangle$.

5.1.2 The fundamental theorem of calculus for distributions

Theorem 5.3. *(The fundamental theorem of calculus for distributions.) Let $f \in \mathscr{D}'(a, b)$. Then there exists $F \in \mathscr{D}'(a, b)$ such that $F' = f$ in $\mathscr{D}'(a, b)$. Furthermore, if $u \in \mathscr{D}'(a, b)$ and $u' = f$ in $\mathscr{D}'(a, b)$, then $u = F + c$ for some constant $c \in \mathbb{C}$.*

Remark 5.4. We emphasize the following formula, found in the proof below, for a distributional primitive to $f \in \mathscr{D}'(a, b)$:

$$\langle F, \varphi\rangle := -\langle f, E(\varphi)\rangle \quad \varphi \in \mathscr{D}(a, b),$$

where $E(\varphi)$ is defined at (27).

Proof. We start by choosing $\chi \in \mathscr{D}(a, b)$ with $\int_a^b \chi\,\mathrm{d}x = 1$ ($\chi(x) = \rho_\varepsilon(x - x_0)$ for suitable $\varepsilon > 0$ and $x_0 \in (a, b)$ will do.) Define for $\varphi \in \mathscr{D}(a, b)$,

$$E(\varphi)(x) := \int_a^x \varphi(t)\,\mathrm{d}t - \int_a^b \varphi(t)\,\mathrm{d}t \int_a^x \chi(t)\,\mathrm{d}t, \quad x \in (a, b). \tag{27}$$

From the fundamental theorem of calculus it follows that $E(\varphi)$ is C^1 with

$$E(\varphi)'(x) = \varphi(x) - \int_a^b \varphi(t)\,\mathrm{d}t\chi(x).$$

The last expression shows that $E(\varphi)$ is C^∞. By assumptions of compact supports in (a, b) we can find a compact interval $[c, d] \subset (a, b)$ that contains the supports of both χ and φ. Then we clearly have $E(\varphi)(x) = 0$ for $x \in (a, c)$. For $x \in (d, b)$ we calculate

$$E(\varphi)(x) = \int_a^b \varphi(t)\,\mathrm{d}t - \int_a^b \varphi(t)\,\mathrm{d}t \int_a^b \chi(t)\,\mathrm{d}t = 0$$

since $\varphi = \chi = 0$ on (d, b). Thus $\operatorname{supp}(E(\varphi)) \subseteq [c, d]$ and we conclude that $E(\varphi) \in \mathscr{D}(a, b)$. By inspection the map $E\colon \mathscr{D}(a, b) \to \mathscr{D}(a, b)$ is linear. We assert that it is also $\mathscr{D}(a, b)$-continuous: if $\varphi_j \to 0$ in $\mathscr{D}(a, b)$, that is, for some compact interval $[\bar{a}, \bar{b}] \subset (a, b)$,

$$\begin{cases} \operatorname{supp}(\varphi_j) \subset [\bar{a}, \bar{b}], \\ \sup |\varphi_j^{(m)}| \to 0 \text{ as } j \to \infty \quad \text{for each } m \in \mathbb{N}. \end{cases}$$

Without loss in generality we may assume that also $\operatorname{supp}(\chi) \subset [\bar{a}, \bar{b}]$ and have then that $\operatorname{supp} E(\varphi_j) \subset [\bar{a}, \bar{b}]$ for all j too. We also have as $j \to \infty$,

$$\sup |E(\varphi_j)| \leqslant (\bar{b} - \bar{a})\big(1 + \int_a^b |\chi| \,\mathrm{d}t\big) \sup |\varphi_j| \to 0$$

and for each $k \in \mathbb{N}$,

$$\sup |E(\varphi_j)^{(k)}| \leqslant \sup |\varphi_j^{(k-1)}| + (\bar{b} - \bar{a}) \sup |\chi^{(k-1)}| \sup |\varphi_j| \to 0.$$

Thus $E(\varphi_j) \to 0$ in $\mathscr{D}(a, b)$ as asserted.

In view of the fundamental theorem of calculus, E is a left-inverse to $\frac{\mathrm{d}}{\mathrm{d}x}$ on $\mathscr{D}(a, b)$:

$$E(\varphi') = \varphi \quad \text{for } \varphi \in \mathscr{D}(a, b).$$

If therefore we define $\langle F, \varphi \rangle = -\langle f, E(\varphi) \rangle$, then clearly $F \in \mathscr{D}'(a, b)$ and $F' = f$. Finally, the statement about u now follows from the constancy theorem. $\square$

Corollary 5.5. *Let $f \in \mathrm{C}(a, b)$. If $u \in \mathscr{D}'(a, b)$ and $u' = f$ in $\mathscr{D}'(a, b)$, then $u \in \mathrm{C}^1(a, b)$ and we have $u' = f$ in the usual sense on (a, b).*

Proof. Take $x_0 \in (a, b)$ and put $F(x) = \int_{x_0}^x f(t) \,\mathrm{d}t$, $x \in (a, b)$. Then F is C^1 and $F' = f$ by the fundamental theorem of calculus. Now $u - F \in \mathscr{D}'(a, b)$ and $(u - F)' = 0$ in $\mathscr{D}'(a, b)$ so $u - F = c$ for some constant $c \in \mathbb{C}$ by the constancy theorem. $\square$

Example 5.6. Let $f \in \mathrm{C}^\infty(a, b)$ and $g \in \mathscr{D}'(a, b)$. We seek the general solution to the ODE

$$y' + fy = g$$

in $\mathscr{D}'(a, b)$. Fix $x_0 \in (a, b)$ and put $F(x) = \int_{x_0}^x f(t) \,\mathrm{d}t$, $x \in (a, b)$. Clearly, $\mathrm{e}^F \in \mathrm{C}^\infty(a, b)$ is an integrating factor, so that if $y \in \mathscr{D}'(a, b)$ is a solution, then

$$\frac{\mathrm{d}}{\mathrm{d}x}\big(\mathrm{e}^F y\big) = \mathrm{e}^F\big(y' + fy\big) = \mathrm{e}^F g.$$

From the fundamental theorem of calculus for distributions we infer that (in the notation of its proof) $\langle \mathrm{e}^F y, \phi \rangle = -\langle \mathrm{e}^F g, E(\phi) \rangle$ for $\phi \in \mathscr{D}(a, b)$, and hence that for some constant $c \in \mathbb{C}$,

$$\langle y, \phi \rangle := -\big\langle g, \mathrm{e}^F E\big(\mathrm{e}^{-F}\phi\big)\big\rangle + c \int_a^b \mathrm{e}^{-F}\phi \,\mathrm{d}x \quad (\phi \in \mathscr{D}(a, b))$$

Conversely we check that each of these distributions satisfy the ODE, so that we have indeed found the general solution. Note in particular that if $g \in \mathrm{C}^\infty(a,b)$ then also all solutions belong to $\mathrm{C}^\infty(a,b)$. It is not difficult to show that if the distribution g has order at most 1, then all solutions have order 0 (exercise!).

Example 5.7. We seek the general solution to the ODE $y'' + 4y' + 3y = 0$ in $\mathscr{D}'(a,b)$. The characteristic equation has roots -1 and -3, so we have the factorization

$$\frac{\mathrm{d}^2}{\mathrm{d}x^2} + 4\frac{\mathrm{d}}{\mathrm{d}x} + 3\mathrm{I} = \left(\frac{\mathrm{d}}{\mathrm{d}x} + \mathrm{I}\right)\left(\frac{\mathrm{d}}{\mathrm{d}x} + 3\mathrm{I}\right) \quad (\mathrm{I} = \text{ identity map})$$

We therefore consider first the ODE $z' + z = 0$ in $\mathscr{D}'(a,b)$. Multiply by e^x, use Leibniz' rule and the constancy theorem to find

$$z = c\mathrm{e}^{-x}$$

where $c \in \mathbb{C}$. Next, consider the ODE $y' + 3y = c\mathrm{e}^{-x}$ in $\mathscr{D}'(a,b)$. Multiply by e^{3x}, use Leibniz' rule and the constancy theorem to find

$$y = c_1\mathrm{e}^{-3x} + \frac{c}{2}\mathrm{e}^{-x}$$

where $c_1 \in \mathbb{C}$. In turn it is easy to check that this is a solution, so that we have shown that the general solution in $\mathscr{D}'(a,b)$ is $y = A\mathrm{e}^{-3x} + B\mathrm{e}^{-x}$, where A, $B \in \mathbb{C}$ are arbitrary.

Corollary 5.5 has an n-dimensional version too:

Proposition 5.8. *Let Ω be a non-empty open subset of $\mathbb{R}^n$. If $u \in \mathscr{D}'(\Omega)$ and $\nabla u \in \mathrm{C}(\Omega)^n$, then $u \in \mathrm{C}^1(\Omega)$.*

Proof. (The proof is not examinable.) We only give the proof in the special case $\Omega = \mathbb{R}^n$, where it is quite similar to that of the constancy theorem. For the standard mollifier $(\rho_\varepsilon)_{\varepsilon>0}$ we have by Lemma 4.12, $\rho_\varepsilon * u \in \mathrm{C}^\infty(\mathbb{R}^n)$ and $\rho_\varepsilon * u \to u$ in $\mathscr{D}'(\mathbb{R}^n)$ as $\varepsilon \searrow 0$. Now by the fundamental theorem of calculus and the chain rule we have for each $x \in \mathbb{R}^n$ and $\varepsilon > 0$,

$$(\rho_\varepsilon * u)(x) = (\rho_\varepsilon * u)(0) + \int_0^1 (\rho_\varepsilon * \nabla u)(tx) \cdot x \,\mathrm{d}t. \tag{28}$$

Multiply by $\rho(x)$ and integrate over $x \in \mathbb{R}^n$ to get

$$\langle \rho_\varepsilon * u, \rho \rangle = (\rho_\varepsilon * u)(0) + \int_{\mathbb{R}^n}\int_0^1 (\rho_\varepsilon * \nabla u)(tx) \cdot x \,\mathrm{d}t \rho(x) \,\mathrm{d}x.$$

Here we have, as $\varepsilon \searrow 0$,

$$\langle \rho_\varepsilon * u, \rho \rangle \to \langle u, \rho \rangle$$

and $\rho_\varepsilon * \nabla u \to \nabla u$ locally uniformly on $\mathbb{R}^n$, hence

$$\int_{\mathbb{R}^n}\int_0^1 (\rho_\varepsilon * \nabla u)(tx) \cdot x \,\mathrm{d}t \rho(x) \,\mathrm{d}x \to \int_{\mathbb{R}^n}\int_0^1 \nabla u(tx) \cdot x \,\mathrm{d}t \rho(x) \,\mathrm{d}x.$$

Consequently,

$$(\rho_\varepsilon * u)(0) \to c := \langle u, \rho \rangle - \int_{\mathbb{R}^n}\int_0^1 \nabla u(tx) \cdot x \,\mathrm{d}t \rho(x) \,\mathrm{d}x$$

as $\varepsilon \searrow 0$. Hence, returning to (28) we see that, as $\varepsilon \searrow 0$, $(\rho_\varepsilon * u)(x)$ converges locally uniformly in $x \in \mathbb{R}^n$, and since it also converges distributionally to u it is not hard to see that u is a continuous function and that the local uniform convergence in $x \in \mathbb{R}^n$ is in fact to $u(x)$. To see that u is C^1 we write $x = (x_j, x') \in \mathbb{R}^n$ and note that for fixed $x' \in \mathbb{R}^{n-1}$ the fundamental theorem of calculus yields

$$(\rho_\varepsilon * u)(x_j, x') = (\rho_\varepsilon * u)(0, x') + \int_0^{x_j} (\rho_\varepsilon * \partial_j u)(t, x') \,\mathrm{d}t$$

for each $x_j \in \mathbb{R}$. Pass to the limit $\varepsilon \searrow 0$ to get

$$u(x_j, x') = u(0, x') + \int_0^{x_j} (\partial_j u)(t, x')\,\mathrm{d}t$$

for each $x_j \in \mathbb{R}$. Since the distributional derivative $\partial_j u$ is continuous we infer that the partial function $x_j \mapsto u(x_j, x')$ is differentiable with derivative $(\partial_j u)(x_j, x')$. But then u must be C^1 as we wanted to show. □

The last result of this subsection is technical and we have encountered it before in special cases. The proof is merely a formalization of the proof we have sketched for the special case of mollification of a distribution.

Theorem 5.9. *(Differentiation behind the distribution sign) Let Ω, Ξ be nonempty open subsets of $\mathbb{R}^n$, $\mathbb{R}^m$, respectively. Assume $\Phi \in \mathrm{C}^\infty(\Omega \times \Xi)$ and that there exists a compact set $K \subset \Omega$ such that $\Phi(x, y) = 0$ whenever $x \notin K$, $y \in \Xi$. Then the function*

$$y \mapsto \big\langle u, \Phi(\cdot, y)\big\rangle$$

is a C^∞ function on Ξ for each $u \in \mathscr{D}'(\Omega)$, and for each multi-index $\alpha \in \mathbb{N}_0^m$,

$$\partial_y^\alpha \big\langle u, \Phi(\cdot, y)\big\rangle = \big\langle u, \big(\partial_y^\alpha \Phi\big)(\cdot, y)\big\rangle.$$

Proof. (The proof is not examinable.) For fixed $y \in \Xi$ we get by Taylor's formula

$$\Phi(x, y + h) = \Phi(x, y) + \sum_{j=1}^m \frac{\partial \Phi}{\partial y_j}(x, y) h_j + R(x, y, h),$$

where the remainder term

$$R(x, y, h) = 2\int_0^1 (1 - t) \sum_{|\alpha|=2} \frac{\partial^2 \Phi}{\partial y^\alpha}(x, y + th) h^\alpha \,\mathrm{d}t.$$

Fix $d > 0$ so $d < \operatorname{dist}(y, \partial\Xi)$. Then we have for each multi-index $\beta \in \mathbb{N}_0^n$ that

$$\sup_{x \in \Omega} \big|\partial_x^\beta R(x, y, h)\big| \leqslant c_\beta |h|^2$$

for all $|h| < d$, where we have defined the constant

$$c_\beta := \sum_{|\alpha|=2} \sup_{K \times B_d(y)} |\partial_x^\beta \partial_y^\alpha \Phi|$$

that by the assumptions on Φ is finite. Using the boundedness property of u on K we find $c_K \geqslant 0$, $m_K \in \mathbb{N}_0$ so

$$|\langle u, \varphi\rangle| \leqslant c_K \sum_{|\gamma| \leqslant m_K} \sup_K |\partial^\gamma \varphi|$$

holds for all $\varphi \in \mathscr{D}(K)$. We apply this with $\varphi = R(\cdot, y, h) \in \mathscr{D}(K)$ whereby we arrive at

$$\big|\langle u, R(\cdot, y, h)\rangle\big| \leqslant c_K \sum_{|\gamma| \leqslant m_K} c_\gamma |h|^2 =: \bar{c}|h|^2$$

for all $|h| < d$. Consequently we get (using Landau notation to simplify)

$$\langle u, \Phi(\cdot, y + h)\rangle - \langle u, \Phi(\cdot, y)\rangle = \sum_{j=1}^m \left\langle u, \frac{\partial \Phi}{\partial y_j}(\cdot, y)\right\rangle h_j + \mathrm{O}(|h|^2)$$

and it follows that all first order partial derivatives exist at y and equal the coefficients to h_j on the right-hand side. We also see that they must be continuous in y. Using induction on the order of differentiation we can apply the above argument in the induction step to $\partial_y^\alpha \Phi$ to conclude the proof. □

5.1.3 Characterization of monotone functions

Many other results from calculus have analogues in distribution theory. A rather pleasant one is the following. Recall that a function $u\colon (a,b) \to \mathbb{R}$ is *increasing* provided $u(x) \leqslant u(y)$ when $x,\ y \in (a,b)$ and $x \leqslant y$. If u is differentiable, then the mean value theorem implies that u is increasing if and only if $u'(x) \geqslant 0$ for all $x \in (a,b)$. A similar result of course holds for decreasing functions (where a function u is decreasing if $-u$ is increasing).

Theorem 5.10. *Let $u \in \mathscr{D}'(a,b)$. Then u is defined by an increasing function if and only if $u' \geqslant 0$ in $\mathscr{D}'(a,b)$.*

Remark 5.11. We clarify aspects of the statement of the theorem and its proof below. First, note that a monotone function $u\colon (a,b) \to \mathbb{R}$, meaning either an increasing or a decreasing function, is a regular distribution: Indeed, u is (Borel-) measurable since for all $y \in \mathbb{R}$, $u^{-1}(y,\infty)$ is an interval (in the wide sense) and u is locally bounded since for $[c,d] \subset (a,b)$ we have

$$\sup_{[c,d]} |u| \leqslant \max\bigl\{|u(c)|, |u(d)|\bigr\},$$

so that in particular, $u \in \mathrm{L}^1_{\mathrm{loc}}(a,b)$. Next, note that a regular distribution f on (a,b) is positive if and only if $f(x) \geqslant 0$ for almost all $x \in (a,b)$.

Proof. We suppose for simplicity that $(a,b) = \mathbb{R}$ and use mollification (the general case is omitted, but can be done along similar lines). Let $\bigl(\rho_\varepsilon\bigr)_{\varepsilon>0}$ be the standard mollifier on $\mathbb{R}$.

Step 1. Assume that $u \in \mathrm{C}^\infty(\mathbb{R})$.

We clearly have that u is increasing if and only if $u'(x) \geqslant 0$ for all x, and since distributional derivatives are just the usual ones for C^∞ functions, we have that the latter holds if and only if $u' \geqslant 0$ in $\mathscr{D}'(\mathbb{R})$.

Note that the shifted function $\rho^+(x) = \rho(x-1)$, $x \in \mathbb{R}$, satisfies

$$\rho^+ \geqslant 0,\ \int_{\mathbb{R}} \rho^+ \,\mathrm{d}x = 1 \ \text{ and } \ \mathrm{supp}(\rho^+) = [0,2].$$

We put as usual $\rho^+_\varepsilon(x) := \varepsilon^{-1}\rho\bigl(\varepsilon^{-1}x - 1\bigr)$, $x \in \mathbb{R}$ and $\varepsilon > 0$. Then $\bigl(\rho^+_\varepsilon\bigr)_{\varepsilon>0}$ is an approximate unit with essentially the same properties as the standard mollifier on $\mathbb{R}$, the only difference is that we have shifted the support to have $\mathrm{supp}(\rho^+_\varepsilon) = [0,2\varepsilon]$. Now, because of the support of ρ^+, we have

$$(u * \rho^+_\varepsilon)(x) = \int_0^2 u(x - \varepsilon y)\rho^+(y)\,\mathrm{d}y$$

is an increasing function of $x \in \mathbb{R}$ and a decreasing function of $\varepsilon > 0$.

Step 2. The general case $u \in \mathscr{D}'(\mathbb{R})$.

Assume $u' \geqslant 0$ in $\mathscr{D}'(\mathbb{R})$. Then $\rho_\varepsilon * u \in \mathrm{C}^\infty(\mathbb{R})$ and $\bigl(\rho_\varepsilon * u\bigr)' = \rho_\varepsilon * u' \geqslant 0$ so $\rho_\varepsilon * u$ is an increasing function on $\mathbb{R}$. Consider for $\bar{\varepsilon} > 0$ and $x \in \mathbb{R}$ the function $(\rho_\varepsilon * u) * \rho^+_{\bar{\varepsilon}}(x)$. This function is then increasing in $x \in \mathbb{R}$ and decreasing in $\bar{\varepsilon} > 0$. Since, as $\varepsilon \searrow 0$,

$$\big((\rho_\varepsilon * u) * \rho_{\bar{\varepsilon}}^+\big)(x) \to (u * \rho_{\bar{\varepsilon}}^+)(x)$$

pointwise in $x \in \mathbb{R}$ and $\bar{\varepsilon} > 0$, it follows that also $(u * \rho_{\bar{\varepsilon}}^+)(x)$ is increasing in $x \in \mathbb{R}$ and decreasing in $\bar{\varepsilon} > 0$. Define

$$\sup_{\bar{\varepsilon}>0}(u * \rho_{\bar{\varepsilon}}^+)(x) =: u_0(x) \in \mathbb{R} \cup \{\infty\}$$

and observe that it is a monotone increasing limit as $\bar{\varepsilon} \searrow 0$. Hence the function $u_0 \colon \mathbb{R} \to \mathbb{R} \cup \{\infty\}$ is increasing, meaning in particular, that if for some $x_0 \in \mathbb{R}$ we have $u_0(x_0) = \infty$, then $u_0(x) = \infty$ for all $x \geqslant x_0$. But we also have for $\varphi \in \mathscr{D}(\mathbb{R})$ with $\varphi \geqslant 0$ that

$$\langle u * \rho_{\bar{\varepsilon}}^+, \varphi \rangle \to \langle u, \varphi \rangle \ \text{ as } \ \bar{\varepsilon} \searrow 0.$$

By Lebesgue's monotone convergence theorem also

$$\langle u * \rho_{\bar{\varepsilon}}^+, \varphi \rangle = \int_{\mathbb{R}} u * \rho_{\bar{\varepsilon}}^+, \varphi \,\mathrm{d}x \nearrow \int_{\mathbb{R}} u_0 \varphi \,\mathrm{d}x \ \text{ as } \ \bar{\varepsilon} \searrow 0,$$

so we can exclude the value ∞ and have then that $u_0 \colon \mathbb{R} \to \mathbb{R}$ is a real-valued increasing function. Clearly also $u = u_0$ concluding the proof of this direction. Conversely, if $u \in \mathscr{D}'(\mathbb{R})$ is defined by an increasing function (that we denote u again), then so is $\rho_\varepsilon * u$:

$$(\rho_\varepsilon * u)(x) = \int_{\mathbb{R}} u(x - \varepsilon y)\rho(y) \,\mathrm{d}y \ \text{ is increasing in } \ x \in \mathbb{R}.$$

But also $\rho_\varepsilon * u \in \mathrm{C}^\infty(\mathbb{R})$ so $0 \leqslant \big(\rho_\varepsilon * u\big)' = \rho_\varepsilon * u'$ and taking $\varphi \in \mathscr{D}(\mathbb{R})$, $\varphi \geqslant 0$ we get $0 \leqslant \langle \rho_\varepsilon * u', \varphi \rangle \to \langle u', \varphi \rangle$ as $\varepsilon \searrow 0$, as required. □

Remark 5.12. Recall that a positive distribution by the Riesz-Markov representation theorem is a locally finite Borel measure, so the distributional derivative of an increasing function is a locally finite Borel measure. The next example explores the converse of this.

Example 5.13. Let μ be a locally finite Borel measure on (a, b). Fix $x_0 \in (a, b)$ and put

$$u(x) = \begin{cases} \mu\big([x_0, x]\big) & \text{if } x \in [x_0, b) \\ -\mu\big([x, x_0)\big) & \text{if } x \in (a, x_0). \end{cases} \tag{29}$$

Then u is an increasing function (hence a regular distribution) and we will show that $u' = \mu$ in $\mathscr{D}'(a, b)$.

Simply calculate for $\varphi \in \mathscr{D}(a, b)$ using Fubini's theorem to swap integration orders:

$$\begin{aligned} \langle u', \varphi \rangle &= -\langle u, \varphi' \rangle \\ &= -\int_{x_0}^{b} \mu\big([x_0, x]\big)\varphi'(x) \,\mathrm{d}x + \int_a^{x_0} \mu\big([x, x_0)\big)\varphi'(x) \,\mathrm{d}x \\ &= -\int_{[x_0,b)} \int_t^b \varphi'(x) \,\mathrm{d}x \,\mathrm{d}\mu(t) + \int_{(a,x_0)} \int_a^t \varphi'(x) \,\mathrm{d}x \,\mathrm{d}\mu(t) \\ &= \int_a^b \varphi \,\mathrm{d}\mu \end{aligned}$$

where we also used the usual fundamental theorem of calculus. (Note that the function u defined above satisfies $u(y) - u(x) = \mu(x, y]$ for all $x < y$ in (a, b) and compare it to how one can define Lebesgue-Stieltjes measures on (a, b). In particular, we emphasize that when μ is a probability measure, so $\mu(a, b) = 1$, then the cumulative distribution function is usually defined as $u(x) := \mu(a, x]$, $x \in (a, b)$, and the above calculation applies and gives $u' = \mu$ in $\mathscr{D}'(a, b)$.)

Returning to the general case (29), it is easy to check that the function u is continuous at the point $x \in (a, b)$ precisely when $\mu(\{x\}) = 0$. Hence u is a continuous function precisely when the measure μ is atom-free, meaning that $\mu(\{x\}) = 0$ for all $x \in (a, b)$.

Example 5.14. Let $f \in \mathrm{L}^1_{\mathrm{loc}}(a, b)$. What are the distributional primitives of f? We could of course employ the fundamental theorem of calculus for distributions, but here it is easier to just repeat the calculation with Fubini's theorem done in Example 5.13. Hereby we find

$$F(x) = \int_{x_0}^{x} f(t)\,\mathrm{d}t + c,$$

where $x_0 \in (a, b)$ and $c \in \mathbb{C}$. In particular we record that the fundamental theorem of calculus holds for F if by F' we understand the distributional derivative and we assume that it is a regular distribution $F' \in \mathrm{L}^1_{\mathrm{loc}}(a, b)$:

$$F(d) - F(c) = \int_{c}^{d} F'(t)\,\mathrm{d}t \quad \text{ whenever } [c, d] \subset (a, b).$$

When $f \in \mathrm{L}^1_{\mathrm{loc}}(a, b)$ and $f \geqslant 0$, then in the set-up of Example 5.13 we have the measure

$$\mu(A) = \int_{A} f\,\mathrm{d}t \quad \text{ for Borel sets } A \subset (a, b).$$

Measures that admit such a representation are called *absolutely continuous* (with respect to Lebesgue measure $\mathscr{L}^1$ on (a, b)).

5.2 Sobolev functions.

The functions encountered in the example of the previous subsection are important and have a name:

Definition 5.15. *A function $F\colon (a, b) \to \mathbb{C}$ is* absolutely continuous *if there exists $f \in \mathrm{L}^1(a, b)$ so*

$$F(x) = F(x_0) + \int_{x_0}^{x} f(t)\,\mathrm{d}t, \quad x,\, x_0 \in (a, b).$$

It is locally absolutely continuous *if only $f \in \mathrm{L}^1_{\mathrm{loc}}(a, b)$ above.*

Remark 5.16. (*Not examinable.*) It is possible to characterize absolute continuity as follows: the function $F\colon (a, b) \to \mathbb{C}$ is absolutely continuous if and only if for each $\varepsilon > 0$ there exists $\delta > 0$ so

$$\sum_{j=1}^{m} |F(b_j) - F(a_j)| < \varepsilon$$

whenever $m \in \mathbb{N}$ and $(a_1, b_1], \dots, (a_m, b_m]$ are disjoint subintervals of (a, b) with total length $\sum_{j=1}^{m}(b_j - a_j) < \delta$. In fact, this is usually taken as the definition of absolute continuity.

It is not difficult to see that a complex function is absolutely continuous precisely when its real and imaginary parts are absolutely continuous. It is possible to characterize absolute continuity of real-valued functions in terms of usual differentiability as follows:

Theorem A: A function $F\colon (a, b) \to \mathbb{R}$ is locally absolutely continuous if and only if the following three conditions are satisfied:

(i) F is continuous,

(ii) F is differentiable almost everywhere in (a, b) and $F' \in \mathrm{L}^1_{\mathrm{loc}}(a, b)$,

(iii) F has the *Luzin (N) property*: $\mathscr{L}^1(F(N)) = 0$ whenever $N \subset (a, b)$ and $\mathscr{L}^1(N) = 0$.

If we strengthen condition (ii) above and require differentiability *everywhere* then (iii) can be shown to follow and we have the following sufficient condition for local absolute continuity:

Theorem B: A function $F\colon (a, b) \to \mathbb{R}$ is locally absolutely continuous if F is differentiable at *all* points of (a, b) and $F' \in \mathrm{L}^1_{\mathrm{loc}}(a, b)$.

The condition that the usual derivative must be locally integrable is necessary and not automatic. For example

$$F(x) = \begin{cases} x \sin \frac{1}{x} & \text{if } x \in (-1, 1) \setminus \{0\} \\ 0 & \text{if } x = 0 \end{cases}$$

is differentiable everywhere in $(-1, 1)$, but $F' \notin \mathrm{L}^1_{\mathrm{loc}}(-1, 1)$.

The signed versions of the functions encountered in Example 5.13 are important and have a name:

Definition 5.17. *A function $F\colon (a, b) \to \mathbb{C}$ is of* locally bounded (essential) variation *if $F \in \mathrm{L}^1_{\mathrm{loc}}(a, b)$ and $F' \in \mathscr{D}'_0(a, b)$ (so the distributional derivative is of order 0).*

Remark 5.18. (*Not examinable.*) A function $F\colon (a, b) \to \mathbb{C}$ is of *bounded variation* if its total variation on (a, b), $\mathrm{TV}(F, (a, b))$ is finite. The total variation is defined as

$$\mathrm{TV}(F, (a, b)) := \sup_{\mathcal{P}} \sum_{j=0}^{n-1} |F(a_{j+1}) - F(a_j)|$$

where the supremum is taken over all partitions

$$\mathcal{P} : a = a_0 < a_1 < \dots < a_n = b, \quad n \in \mathbb{N},$$

of (a, b). The function $F\colon (a, b) \to \mathbb{C}$ is of *locally bounded variation* if it has bounded variation on each compact subinterval $[c, d] \subset (a, b)$.

It is easy to check that a complex valued function is of locally bounded variation precisely when its real and imaginary parts are. That real-valued functions of locally bounded variation is closely related to monotone functions is established in the following theorem.

Theorem C: The function $F\colon (a, b) \to \mathbb{R}$ is of locally bounded variation if and only if there exist two increasing functions $F_+, F_- : (a, b) \to \mathbb{R}$ such that $F = F_+ - F_-$.

It is interesting to compare this result with the *Hahn decomposition* of a Radon measure (see Definition 3.14): any real-valued Radon measure $u\colon \mathrm{C}_c(\Omega) \to \mathbb{R}$ can be written as a difference $u = u_+ - u_-$ of two positive Radon measures u_+, u_-.

From Theorems 5.10 and C we infer that $F\colon (a, b) \to \mathbb{C}$ is of locally bounded (essential) variation precisely when it admits a representative of locally bounded variation.

The generalization of the above to higher dimensions is very important, but we can only cover very little in this direction here. However, we can give a central definition:

Definition 5.19. *Let Ω be a nonempty open subset of $\mathbb{R}^n$, $m \in \mathbb{N}$ and $p \in [1, \infty]$. Then any $u \in \mathrm{L}^p(\Omega)$ for which $\partial^\alpha u \in \mathrm{L}^p(\Omega)$ for all multi-indices $\alpha \in \mathbb{N}_0^n$ of length $|\alpha| \leqslant m$ is called a* $\mathrm{W}^{m,p}$ Sobolev function. *The set of all these is denoted by*

$$\mathrm{W}^{m,p}(\Omega) := \left\{ u \in \mathrm{L}^p(\Omega) : \partial^\alpha u \in \mathrm{L}^p(\Omega), \, |\alpha| \leqslant m \right\}$$

and is called a Sobolev space. *It is not difficult to check that it is a vector subspace of* $\mathrm{L}^p(\Omega)$ *and that*

$$\|u\|_{\mathrm{W}^{m,p}} := \begin{cases} \left(\sum_{|\alpha|\leqslant m} \|\partial^\alpha u\|_p^p\right)^{\frac{1}{p}} & \text{if } p \in [1,\infty) \\ \max_{|\alpha|\leqslant m} \|\partial^\alpha u\|_\infty & \text{if } p = \infty \end{cases}$$

is a norm on $\mathrm{W}^{m,p}(\Omega)$ *in the same sense that* $\|\cdot\|_p$ *is a norm on* $\mathrm{L}^p(\Omega)$ *(so it is a norm on the vector space of equivalence classes under the equivalence relation* equal almost everywhere*). One can show that hereby* $\mathrm{W}^{m,p}(\Omega)$ *is an example of a Banach space (and a Hilbert space when* $p = 2$*). These general notions are important and discussed in Functional Analysis 1 and 2. However, they do not play any immediate role in this course.*

Example 5.20. ($\mathrm{W}^{1,p}$ *Sobolev functions on a bounded interval* (a,b))

The Sobolev space

$$\mathrm{W}^{1,p}(a,b) = \left\{u \in \mathrm{L}^p(a,b) : \, u' \in \mathrm{L}^p(a,b)\right\}$$

is normed by

$$\|u\|_{\mathrm{W}^{1,p}} := \begin{cases} \left(\int_a^b |u|^p \,\mathrm{d}x + \int_a^b |u'|^p \,\mathrm{d}x\right)^{\frac{1}{p}} & \text{if } p \in [1,\infty) \\ \max\left\{\|u\|_\infty, \|u'\|_\infty\right\} & \text{if } p = \infty \end{cases}$$

and because (a,b) is assumed bounded the family is strictly descending in the exponent p:

$$\mathrm{W}^{1,1}(a,b) \supset \mathrm{W}^{1,p}(a,b) \supset \mathrm{W}^{1,q}(a,b) \supset \mathrm{W}^{1,\infty}(a,b)$$

when $1 < p < q < \infty$.

Comparing our definitions we see that an integrable function $u\colon (a,b) \to \mathbb{C}$ is $\mathrm{W}^{1,1}$ Sobolev precisely when it has an absolutely continuous representative. For this absolutely continuous representative the fundamental theorem of calculus holds in the sense that

$$u(y) - u(x) = \int_x^y u'(t) \,\mathrm{d}t \tag{30}$$

holds for every x, $y \in (a,b)$, where u' is the distributional derivative (but see also Theorem A in Remark 5.16 above).

When the exponent p increases we get better continuity properties of $\mathrm{W}^{1,p}$ Sobolev functions. Indeed if $u \in \mathrm{W}^{1,p}(a,b)$ and u is its absolutely continuous representative, then we can apply Hölder's inequality on (30) with Hölder conjugate exponents p, q (and for the ensuing calculation assume $x < y$):

$$\begin{aligned} |u(y) - u(x)| \leqslant& \|\mathbf{1}_{[x,y]}\|_q \|u'\|_p \\ =& |x-y|^{1-\frac{1}{p}} \|u'\|_p \end{aligned}$$

We express this by saying that u is $(1-\frac{1}{p})$-*Hölder continuous* on (a,b). In the endpoint case $p=\infty$ we get

$$|u(y)-u(x)| \leqslant |x-y| \|u'\|_{\infty}$$

which of course is *Lipschitz continuity.* It can be shown that $\mathrm{W}^{1,\infty}(a,b)$ is exactly the space of $\mathrm{L}^{\infty}(a,b)$ functions that admit a representative that is Lipschitz continuous (see Problem Sheet 4). One cannot characterize the $\mathrm{W}^{1,p}$ Sobolev functions in such simple terms when $p<\infty$.

Before the next example we need a convenient result about distributional partial derivatives.

Lemma 5.21. *Assume the dimension $n \geqslant 2$ and let $f \in \mathrm{C}^1(\mathbb{R}^n \setminus \{0\}) \cap \mathrm{L}^1_{\mathrm{loc}}(\mathbb{R}^n)$. If the usual partial derivatives $\partial_j f \in \mathrm{L}^1_{\mathrm{loc}}(\mathbb{R}^n)$ for each direction $1 \leqslant j \leqslant n$, then we have*

$$\int_{\mathbb{R}^n} \partial_j f \varphi \,\mathrm{d}x = -\int_{\mathbb{R}^n} f \partial_j \varphi \,\mathrm{d}x$$

for all $\varphi \in \mathscr{D}(\mathbb{R}^n)$. In other words, the usual partial derivatives coincide with the distributional partial derivatives.

You will be asked to prove the result on Problem Sheet 4.

Example 5.22. (*Sobolev functions can be discontinuous in higher dimensions*)

For $\alpha>0$ and $n \geqslant 2$ we consider the function

$$u(x) := |x|^{-\alpha},\, x \in B_1(0) \subset \mathbb{R}^n.$$

Of course it is not really defined for $x=0$ but that is not a problem when discussing membership of L^p or $\mathrm{W}^{1,p}$.

Claim: $u \in \mathrm{L}^p(B_1(0))$ if and only if $\alpha < \frac{n}{p}$.

We calculate in polar coordinates:

$$\begin{aligned}\int_{B_1(0)} |x|^{-\alpha p} \,\mathrm{d}x &= \int_0^1 \int_{\partial B_r(0)} |x|^{-\alpha p} \,\mathrm{d}S_x \,\mathrm{d}r \\ &= \omega_{n-1} \int_0^1 r^{-\alpha p + n - 1} \,\mathrm{d}r < \infty\end{aligned}$$

precisely when $n - \alpha p > 0$, which is the claim.

Next we assume $\alpha < \frac{n}{p}$ so that u is regular distribution.

Claim: $u \in \mathrm{W}^{1,p}(B_1(0))$ precisely for $\alpha < \frac{n}{p}-1$ when $p \in [1,n)$ (and never when $\alpha>0$ and $p \geqslant n$).

Calculate the usual partial derivatives $\partial_j u = -\alpha |x|^{-\alpha-2} x_j$ for each direction $1 \leqslant j \leqslant n$. Here $\partial_j u \in \mathrm{L}^p(B_1(0))$ for each $1 \leqslant j \leqslant n$ precisely when

$$\sum_{j=1}^{n} |\partial_j u| \in \mathrm{L}^p(B_1(0)). \tag{31}$$

We simplify the latter using the inequalities

$$\left(\sum_{j=1}^{n} |y_j|^2\right)^{\frac{1}{2}} \leqslant \sum_{j=1}^{n} |y_j| \leqslant n^{\frac{1}{2}} \left(\sum_{j=1}^{n} |y_j|^2\right)^{\frac{1}{2}}.$$

Indeed (31) holds therefore precisely when

$$\left(\sum_{j=1}^{n} |\partial_j u|^2\right)^{\frac{1}{2}} = \alpha |x|^{-\alpha-1} \in \mathrm{L}^p(B_1(0)).$$

Again we calculate in polar coordinates:

$$\begin{aligned}\int_{B_1(0)} |x|^{-(\alpha+1)p}\,\mathrm{d}x &= \int_0^1 \int_{\partial B_r(0)} |x|^{-(\alpha+1)p}\,\mathrm{d}S_x\,\mathrm{d}r \\ &= \omega_{n-1} \int_0^1 r^{-(\alpha+1)p+n-1}\,\mathrm{d}r < \infty\end{aligned}$$

exactly for $n - (\alpha+1)p > 0$, whereby, in view of Lemma 5.21, we have found that $u \in \mathrm{W}^{1,p}(B_1(0))$ precisely if $\alpha < \frac{n}{p} - 1$ and $p \in [1, n)$. It can be shown that for $p > n$ all $\mathrm{W}^{1,p}$ Sobolev functions admit a continuous representative (this follows from the so-called Sobolev embedding theorems).

5.3 Localization of distributions

We would like to think about distributions as *generalized functions*. While it is not possible in general to assign pointwise values to a distribution $u \in \mathscr{D}'(\Omega)$, it is easy to define its restriction to an open subset ω of Ω. Indeed, any test function $\varphi \in \mathscr{D}(\omega)$ can be extended to a test function on Ω by simply defining $\varphi(x) = 0$ for $x \in \Omega \setminus \omega$ so we may consider $\mathscr{D}(\omega)$ as a subspace of $\mathscr{D}(\Omega)$ and then define *the restriction of u to ω*, denoted by $u|_\omega$, to be

$$\langle (u|_\omega), \varphi \rangle := \langle u, \varphi \rangle, \quad \varphi \in \mathscr{D}(\omega).$$

Proceeding a bit more formally, if $e_{\omega,\Omega} \colon \mathscr{D}(\omega) \to \mathscr{D}(\Omega)$ is the map extending each $\varphi \in \mathscr{D}(\omega)$ to $e_{\omega,\Omega}(\varphi) \in \mathscr{D}(\Omega)$ by 0, then $e_{\omega,\Omega}$ is clearly linear and $\mathscr{D}(\omega)$-continuous. The map $r_{\Omega,\omega} \colon \mathscr{D}'(\Omega) \to \mathscr{D}'(\omega)$ restricting each $u \in \mathscr{D}'(\Omega)$ to ω is then given as

$$\langle r_{\Omega,\omega} u, \varphi \rangle = \langle u, e_{\omega,\Omega} \varphi \rangle, \quad \varphi \in \mathscr{D}(\omega).$$

What is interesting is that distributions are in fact locally determined: if u, $v \in \mathscr{D}'(\Omega)$ and each $x \in \Omega$ admits an open neighbourhood ω_x in Ω so $u|_{\omega_x} = v|_{\omega_x}$, then $u = v$. This is the content of

Theorem 5.23. *If $u \in \mathscr{D}'(\Omega)$ and for each $x \in \Omega$ there exists an open neighbourhood ω_x of x in Ω so $u|_{\omega_x} = 0$, then $u = 0$.*

Proof. This is a typical smooth partition of unity argument. Let $\varphi \in \mathscr{D}(\Omega)$. Since we clearly have

$$\operatorname{supp}\varphi \subset \bigcup\{\omega_x \,:\, x \in \Omega\},$$

the compactness of $\operatorname{supp}\varphi$ means that we can find a finite subcover, say

$$\operatorname{supp}\varphi \subset \omega_1 \cup \cdots \cup \omega_m,$$

where we wrote $\omega_j = \omega_{x_j}$. Using Theorem 2.14 we find a smooth partition of unity $\phi_1, \dots, \phi_m \in \mathscr{D}(\Omega)$ with $\operatorname{supp}\phi_j \subset \omega_j$, $0 \leqslant \phi_j \leqslant 1$, and $\sum_{j=1}^m \phi_j = 1$ on $\operatorname{supp}\varphi$. Thus

$$\langle u, \varphi\rangle = \left\langle u, \sum_{j=1}^m \varphi\phi_j \right\rangle = \sum_{j=1}^m \langle u, \varphi\phi_j\rangle = 0.$$

□

Example 5.24. (*Cauchy's integral formula*) Let Ω be a nonempty open subset of $\mathbb{C}$ and $\omega \Subset \Omega$ be a C^1 subset. Then for $\phi \in \mathscr{D}(\Omega)$ and $z_0 = x_0 + \mathrm{i}y_0 \in \omega$:

$$\phi(x_0, y_0) = \frac{1}{2\pi\mathrm{i}} \int_{\partial\omega} \frac{\phi(x,y)}{z - z_0}\,\mathrm{d}z - \frac{1}{\pi}\int_\omega \frac{\partial\phi/\partial\bar{z}}{z - z_0}\,\mathrm{d}(x,y) \tag{32}$$

where the usual identification $\mathbb{C} \simeq \mathbb{R}^2$ via $z = x + \mathrm{i}y \simeq (x,y)$ is in force and the $\partial\omega$ in the contour integral is traversed so that ω is to the left of $\partial\omega$. (We note that the assumptions on ω mean that its boundary consists of a finite number of closed simple C^1 curves.)

There are many ways to derive this formula. Here we use distributions and their localization. First recall from Example 4.23 that $\frac{\partial}{\partial\bar{z}}\left(\frac{1}{\pi z}\right) = \delta_0$ and consequently that $\frac{\partial}{\partial\bar{z}}\left(\frac{1}{\pi(z-z_0)}\right) = \delta_{z_0}$. Now $u = \mathbf{1}_\omega/\pi(z - z_0) \in \mathrm{L}^1_{\mathrm{loc}}(\Omega)$ and since $z_0 \in \omega$ this distribution is locally in Ω of the form *distribution times* C^∞ *function*, so we may apply the Leibniz rule to calculate its derivatives:

$$\begin{aligned}\frac{\partial u}{\partial\bar{z}} =& \frac{\partial}{\partial\bar{z}}\left(\frac{1}{\pi(z-z_0)}\right)\mathbf{1}_\omega + \frac{1}{\pi(z-z_0)}\frac{\partial}{\partial\bar{z}}(\mathbf{1}_\omega)\\ =&\delta_{z_0} + \frac{1}{\pi(z-z_0)}\frac{\partial}{\partial\bar{z}}(\mathbf{1}_\omega).\end{aligned}$$

For $\varphi \in \mathscr{D}(\Omega)$ calculate:

$$\begin{aligned}\langle\frac{\partial}{\partial\bar{z}}(\mathbf{1}_\omega), \varphi\rangle &= -\int_\omega \frac{\partial\varphi}{\partial\bar{z}}\,\mathrm{d}(x,y)\\ &= -\frac{1}{2}\int_\omega \big(\partial_x\varphi + \mathrm{i}\partial_y\varphi\big)\,\mathrm{d}(x,y).\end{aligned}$$

For simplicity let us assume that the boundary $\partial\omega$ consists of only one closed C^1 curve that is then traversed in the anti-clockwise direction. Fix a point $p \in \partial\omega$ and let s denote arc length on $\partial\omega$ measured from p anti-clockwise. Then, if ℓ is the length of $\partial\omega$, $[0,\ell] \ni s \mapsto \big(x(s), y(s)\big)$ is a parametrization of $\partial\omega$. The derivative $\tau = \big(x'(s), y'(s)\big)$ is a unit tangent to $\partial\omega$ and $\tau^\perp = \big(-y'(s), x'(s)\big) = -\nu$ is the unit inward normal on $\partial\omega$. We are lined up for use of the divergence theorem: put $V = (\varphi, \mathrm{i}\varphi) \in \mathscr{D}(\Omega)^2$ so that $\mathrm{div}V = \partial_x\varphi + \mathrm{i}\partial_y\varphi$ and we get

$$\begin{aligned}
-\frac{1}{2}\int_\omega \big(\partial_x\varphi + \mathrm{i}\partial_y\varphi\big)\,\mathrm{d}(x,y) &= -\frac{1}{2}\int_{\partial\omega} V\cdot\nu\,\mathrm{d}S \\
&= -\frac{1}{2}\int_0^\ell \big(\varphi\nu_1 + \mathrm{i}\varphi\nu_2\big)\,\mathrm{d}s \\
&= -\frac{1}{2}\int_0^\ell \varphi\big(x(s), y(s)\big)\big(y'(s) - \mathrm{i}x'(s)\big)\,\mathrm{d}s \\
&= \frac{1}{2}\int_0^\ell \mathrm{i}\varphi\big(x(s), y(s)\big)\big(x'(s) + \mathrm{i}y'(s)\big)\,\mathrm{d}s \\
&= \frac{\mathrm{i}}{2}\int_{\partial\omega} \varphi\,\mathrm{d}z.
\end{aligned}$$

Thus

$$\begin{aligned}
\left\langle \frac{\partial}{\partial\bar{z}}\left(\frac{\mathbf{1}_\omega}{\pi(z-z_0)}\right), \phi\right\rangle &= \phi(x_0,y_0) + \left\langle \frac{1}{\pi(z-z_0)}\frac{\partial}{\partial\bar{z}}(\mathbf{1}_\omega), \phi\right\rangle \\
&= \phi(x_0,y_0) + \left\langle \frac{\partial}{\partial\bar{z}}(\mathbf{1}_\omega), \frac{\phi}{\pi(z-z_0)}\right\rangle \\
&= \phi(x_0,y_0) + \frac{\mathrm{i}}{2}\int_{\partial\omega}\frac{\phi(x,y)}{\pi(z-z_0)}\,\mathrm{d}z
\end{aligned}$$

and this rearranges easily to (32).

5.3.1 Support and singular support of a distribution

Definition 5.25. *Let $u \in \mathscr{D}'(\Omega)$. Then the* support of u, *denoted by* $\mathrm{supp}(u)$, *is the set of points $x \in \Omega$ having no open neighbourhood to which the restriction of u is 0. Thus $\Omega \setminus \mathrm{supp}(u)$ is the open set of points having an open neighbourhood in which u vanishes. By Theorem 5.23, $\Omega \setminus \mathrm{supp}(u)$ contains all open subsets of Ω where u vanishes, hence must be the largest such open subset of Ω.*

We record that the support of $u \in \mathscr{D}'(\Omega)$ is a relatively closed subset of Ω, and that for $\varphi \in \mathscr{D}(\Omega)$,

$$\langle u, \varphi\rangle = 0 \text{ whenever } \mathrm{supp}(u) \cap \mathrm{supp}(\varphi) = \emptyset. \tag{33}$$

As an easy exercise one can check that when $u \in \mathrm{C}(\Omega)$, then the support of u as a continuous function coincides with the support of u as a distribution. We are therefore justified in using the same notation for both.

Closely related to the notion of support is the notion of singular support:

Definition 5.26. *Let $u \in \mathscr{D}'(\Omega)$. Then the* singular support *of u, denoted by* sing.supp(u), *is the set of points $x \in \Omega$ having no open neighbourhood to which the restriction of u is a C^∞ function. By Theorem 5.23, $\Omega \setminus \operatorname{sing.supp}(u)$ contains all open subsets of Ω where u is a C^∞ function, hence must be the largest such open subset of Ω.*

We record that the singular support of $u \in \mathscr{D}'(\Omega)$ is a relatively closed subset of Ω. Since the zero function in particular is C^∞ we also have the inclusion $\operatorname{sing.supp}(u) \subseteq \operatorname{supp}(u)$.

Example 5.27. Let $x_0 \in \Omega$ and ω be an open subset of Ω. Then for the distributions δ_{x_0} and $\mathbf{1}_\omega$ we have $\operatorname{supp}(\delta_{x_0}) = \{x_0\} = \operatorname{sing.supp}(\delta_{x_0})$ and $\operatorname{supp}(\mathbf{1}_\omega) = \Omega \cap \overline{\omega}$, $\operatorname{sing.supp}(\mathbf{1}_\omega) = \Omega \cap \partial\omega$.

In Theorem 3.13 it was shown that a distribution of order at most m can be uniquely extended to the space of C_c^m functions as a linear functional with the boundedness property (17). We now return to the theme of extension of the domain for a distribution using the above ideas and start by considering regular distributions: When $u \in \mathrm{L}^1_{\mathrm{loc}}(\Omega)$, then the integral $\int_\Omega u\varphi \,\mathrm{d}x$ is well-defined for all $\varphi \in \mathrm{C}^\infty(\Omega)$ with $\operatorname{supp}(u) \cap \operatorname{supp}(\varphi)$ compact. In fact, the set

$$\left\{ \varphi \in \mathrm{C}^\infty(\Omega) : \operatorname{supp}(u) \cap \operatorname{supp}(\varphi) \text{ is compact } \right\}$$

is a linear subspace of $\mathrm{C}^\infty(\Omega)$ and the map $\varphi \mapsto \int_\Omega u\varphi \,\mathrm{d}x$ is well-defined and linear there. This extension procedure can be generalized:

Theorem 5.28. *Let $u \in \mathscr{D}'(\Omega)$ and let A be a relatively closed subset of Ω so $\operatorname{supp}(u) \subseteq A$. Then there exists a unique linear functional*

$$U \colon \left\{ \varphi \in \mathrm{C}^\infty(\Omega) : A \cap \operatorname{supp}(\varphi) \text{ is compact } \right\} \to \mathbb{C}$$

so $U(\varphi) = \langle u, \varphi \rangle$ for $\varphi \in \mathscr{D}(\Omega)$ and $U(\varphi) = 0$ for $\varphi \in \mathrm{C}^\infty(\Omega)$ with $A \cap \operatorname{supp}(\varphi) = \emptyset$.

Remark 5.29. The domain of U is largest when we take $A = \operatorname{supp}(u)$, but we need the uniqueness part of the result also for more general sets. We shall denote the unique such extension U by u again.

Proof. It is clear that the set $\left\{ \varphi \in \mathrm{C}^\infty(\Omega) : A \cap \operatorname{supp}(\varphi) \text{ is compact } \right\}$ is a linear subspace of $\mathrm{C}^\infty(\Omega)$, so the statement makes sense.

Uniqueness: Let $\varphi \in \mathrm{C}^\infty(\Omega)$ with $A \cap \operatorname{supp}(\varphi) =: K$ compact. Take a cut-off function $\psi \in \mathscr{D}(\Omega)$ so $\psi = 1$ near K. Then $\psi\varphi \in \mathscr{D}(\Omega)$ and $A \cap \operatorname{supp}\big((1-\psi)\varphi\big) = \emptyset$, so if U is an extension of u with the asserted property, then we must have

$$\begin{aligned} U(\varphi) &= U\big(\psi\varphi + (1-\psi)\varphi\big) \\ &= U(\psi\varphi) + U\big((1-\psi)\varphi\big) \\ &= U(\psi\varphi) = \langle u, \psi\varphi \rangle. \end{aligned}$$

Thus there is only one possible such extension.

Existence: If $\varphi \in \mathrm{C}^\infty(\Omega)$ with $A \cap \mathrm{supp}(\varphi)$ compact, then we can write $\varphi = \psi\varphi + (1-\psi)\varphi$ as above. Assume that we have another such representation, say that also $\varphi = \psi_0\varphi + (1-\psi_0)\varphi$. Then $\psi\varphi - \psi_0\varphi \in \mathscr{D}(\Omega)$ and since

$$A \cap \mathrm{supp}(\psi\varphi - \psi_0\varphi) = A \cap \mathrm{supp}\big((1-\psi)\varphi - (1-\psi_0)\varphi\big) = \emptyset$$

we must have by (33) that $\langle u, \psi\varphi - \psi_0\varphi\rangle = 0$, hence $\langle u, \psi\varphi\rangle = \langle u, \psi_0\varphi\rangle$. Consequently, we may consistently define $U(\varphi) := \langle u, \psi\varphi\rangle$. □

5.3.2 Compactly supported distributions

It is interesting to apply the extension result from the previous subsection in the case of compact support. Assume $u \in \mathscr{D}'(\Omega)$ has compact support. Then we can define $\langle u, \varphi\rangle$ for *all* $\varphi \in \mathrm{C}^\infty(\Omega)$: simply take a cut-off function $\psi \in \mathscr{D}(\Omega)$ with $\psi = 1$ near $\mathrm{supp}(u)$. According to the above we must then define for $\varphi \in \mathrm{C}^\infty(\Omega)$ that

$$\langle u, \varphi\rangle := \langle u, \psi\varphi\rangle.$$

Invoking the boundedness property of u on $K := \mathrm{supp}(\psi)$ we find $c_K \geqslant 0$, $m_K \in \mathbb{N}_0$ so

$$\begin{aligned}\big|\langle u, \varphi\rangle\big| &\leqslant c_K \sum_{|\alpha| \leqslant m_K} \sup_K \big|\partial^\alpha(\psi\varphi)\big| \\ &\leqslant c \sum_{|\alpha| \leqslant m} \sup_K \big|\partial^\alpha\varphi\big|\end{aligned}$$

where the last inequality follows by use of the Generalized Leibniz rule and we wrote $m = m_K$. Note that the inequality holds for all $\varphi \in \mathrm{C}^\infty(\Omega)$. We also note that a compactly supported distribution always has a finite order.

Conversely suppose that $u\colon \mathrm{C}^\infty(\Omega) \to \mathbb{C}$ is linear and for some compact subset $L \subset \Omega$ we can find $c \geqslant 0$, $m \in \mathbb{N}_0$ such that

$$\big|u(\varphi)\big| \leqslant c \sum_{|\alpha| \leqslant m} \sup_L |\partial^\alpha\varphi|$$

holds for all $\varphi \in \mathrm{C}^\infty(\Omega)$. Then the restriction of u to $\mathscr{D}(\Omega)$ is a distribution with support contained in L and clearly $u(\varphi) = 0$ for $\varphi \in \mathrm{C}^\infty(\Omega)$ with $\mathrm{supp}(\varphi) \cap L = \emptyset$. By the uniqueness part of Theorem 5.28 we conclude that u is the unique such extension of $u|_{\mathscr{D}(\Omega)}$.

In order to summarize, denote by $\mathscr{E}'(\Omega)$ the set of all linear functionals $u\colon \mathrm{C}^\infty(\Omega) \to \mathbb{C}$ for which there exists a compact subset $K = K_u$ of Ω and constants $c = c_u \geqslant 0$, $m = m_u \in \mathbb{N}_0$ such that

$$\big|\langle u, \varphi\rangle\big| \leqslant c \sum_{|\alpha| \leqslant m} \sup_K \big|\partial^\alpha\varphi\big| \tag{34}$$

holds for all $\varphi \in \mathrm{C}^\infty(\Omega)$. We have then shown the following result:

Theorem 5.30. *The set of distributions with compact support in* Ω *coincides with* $\mathscr{E}'(\Omega)$ *in the following sense:*

(i) each distribution with compact support in Ω *admits a unique extension to a linear functional on* $\mathrm{C}^\infty(\Omega)$ *satisfying a bound of the form (34) for a compact neighbourhood* K *of its support;*

(ii) the restriction of a functional in $\mathscr{E}'(\Omega)$ *to* $\mathscr{D}(\Omega)$ *is a distribution with compact support in* Ω *(if the bound (34) holds, then* K *will contain the support).*

By a compact neighbourhood of a set A we mean a compact set K such that A is contained in the interior of K.

Example 5.31. *It is in general not possible to take* $K = \mathrm{supp}(u)$ *in (34):* Define for $\varphi \in \mathscr{D}(\mathbb{R})$

$$\langle u, \varphi \rangle := \sum_{j=1}^{\infty} \frac{\varphi\left(\frac{1}{j}\right) - \varphi\left(-\frac{1}{j}\right)}{j}.$$

It is not difficult to see that for $\varphi \in \mathscr{D}(\mathbb{R})$ we always have

$$\left| \frac{\varphi\left(\frac{1}{j}\right) - \varphi\left(-\frac{1}{j}\right)}{j} \right| \leqslant 2 \max_{[-1,1]} |\varphi'| j^{-2}$$

and consequently that

$$\bigl|\langle u, \varphi \rangle\bigr| \leqslant \left(2 \sum_{j=1}^{\infty} \tfrac{1}{j^2} \right) \max_{[-1,1]} |\varphi'|.$$

It follows that $u \in \mathscr{D}'(\mathbb{R})$ has order at most 1. By inspection we see that

$$\mathrm{supp}(u) = \left\{ \tfrac{1}{j} : j \in \mathbb{Z} \setminus \{0\} \right\} \cup \{0\}$$

so that in particular u has compact support. Suppose that (34) holds for u with $K = \mathrm{supp}(u)$. Let $s \in \mathbb{N}$ and take $\varphi \in \mathscr{D}(\mathbb{R})$ so $\varphi = 1$ near $\{1, 1/2, \ldots, 1/s\}$ and $\varphi = 0$ near the complement $\mathrm{supp}(u) \setminus \{1, 1/2, \ldots, 1/s\}$. Then we have $\langle u, \varphi \rangle = \sum_{j=1}^{s} 1/j$ and from (34) for some fixed constant c the bound $\sum_{j=1}^{s} 1/j \leqslant c$, which is impossible for large s.

The compact set K in (34) can be any compact neighbourhood of the support $\mathrm{supp}(u)$, but if we shrink K then we have to enlarge the corresponding constant c.

We end this section with a characterization of distributions whose support is a singleton:

Theorem 5.32. *Let* $u \in \mathscr{D}'(\Omega)$ *and* $x_0 \in \Omega$*. If* $\mathrm{supp}\, u = \{x_0\}$*, then* $u \in \mathrm{span}\{\partial^\alpha \delta_{x_0} : \alpha \in \mathbb{N}_0^n\}$*.*

Proof. (The proof is not examinable.) For each $\varepsilon > 0$ we put $\psi_\varepsilon = \rho_\varepsilon * \mathbf{1}_{B_{2\varepsilon}(0)}$, where $(\rho_\varepsilon)_{\varepsilon>0}$ is the standard mollifier in $\mathbb{R}^n$. Then $\psi_\varepsilon \in \mathscr{D}(\mathbb{R}^n)$ is supported in $\overline{B_{3\varepsilon}(0)}$ and $\psi_\varepsilon = 1$ on $B_\varepsilon(0)$. In view of the discussion above we then have

$$\langle u, \varphi \rangle = \langle u, \psi_\varepsilon \varphi \rangle$$

for all $\varphi \in \mathrm{C}^\infty(\mathbb{R}^n)$ and all $\varepsilon > 0$. We also know that u has a finite order, say $m \in \mathbb{N}_0$, and that there exists a constant $c \geqslant 0$ so

$$|\langle u, \varphi \rangle| \leqslant c \sum_{|\alpha| \leqslant m} \sup_{B_1(0)} |\partial^\alpha \varphi| \tag{35}$$

holds for all $\varphi \in \mathrm{C}^\infty(\mathbb{R}^n)$. Now for each multi-index $\alpha \in \mathbb{N}_0^n$ of length $|\alpha| \leqslant m$ we put $c_\alpha = (-1)^{|\alpha|}\langle u, (\cdot - x_0) \rangle / \alpha!$ and assert that

$$u = \sum_{|\alpha| \leqslant m} c_\alpha \partial^\alpha \delta_{x_0}.$$

To prove it we fix $\varphi \in \mathscr{D}(\mathbb{R}^n)$ and get from Taylor's formula

$$\varphi(x) = \sum_{|\alpha| \leqslant m} \frac{\partial^\alpha \varphi(x_0)}{\alpha!}(x - x_0)^\alpha + R_m(x),$$

where

$$R_m(x) = (m+1) \int_0^1 (1-t)^m \sum_{|\alpha| = m+1} \frac{\partial^\alpha \varphi(x_0 + t(x - x_0))}{\alpha!}(x - x_0)^\alpha \, \mathrm{d}t.$$

Now

$$\begin{aligned} \langle u, \varphi \rangle &= \sum_{|\alpha| \leqslant m} \left\langle u, \frac{\partial^\alpha \varphi(x_0)}{\alpha!}(\cdot - x_0)^\alpha \right\rangle + \langle u, R_m \rangle \\ &= \left\langle \sum_{|\alpha| \leqslant m} c_\alpha \partial^\alpha \delta_{x_0}, \varphi \right\rangle + \langle u, \psi_\varepsilon R_m \rangle \end{aligned}$$

for each $\varepsilon > 0$. To see that the last term is 0 we use (35) and the Leibniz rule:

$$|\langle u, \psi_\varepsilon R_m \rangle| \leqslant c \sum_{|\beta| \leqslant m} \sup_{B_1(x_0)} |\partial^\beta (\psi_\varepsilon R_m)|$$

and

$$\begin{aligned} |\partial^\beta (\psi_\varepsilon R_m)| &= \left| \sum_{\gamma \leqslant \beta} \binom{\beta}{\gamma} \partial^\gamma \psi_\varepsilon \partial^{\beta - \gamma} R_m \right| \\ &\leqslant \sum_{\gamma \leqslant \beta} \binom{\beta}{\gamma} \mathcal{O}(\varepsilon^{-|\gamma|}) \mathcal{O}(\varepsilon^{m+1-|\beta|+|\gamma|}) \\ &\leqslant \mathcal{O}(\varepsilon), \end{aligned}$$

where for convenience we used Landau notation. Taking $\varepsilon \searrow 0$ concludes the proof. □

6 Convolution of distributions and fundamental solutions

6.1 Convolution of distributions.

In this subsection we define the convolution of two distributions one of which must have compact support. It is done in three steps. The first is the familiar convolution of a distribution and a test function that was defined using the adjoint identity scheme: if $u \in \mathscr{D}'(\mathbb{R}^n)$, $\theta \in \mathscr{D}(\mathbb{R}^n)$, then $u * \theta \in \mathscr{D}'(\mathbb{R}^n)$ was defined by the rule

$$\langle u * \theta, \varphi \rangle = \langle u, \tilde{\theta} * \varphi \rangle, \quad \varphi \in \mathscr{D}(\mathbb{R}^n).$$

Proceeding as in the proof of Lemma 4.12 (see also Remark 4.13 and Problem Sheet 3 for the one-dimensional case) we establish:

Lemma 6.1. *If $u \in \mathscr{D}'(\mathbb{R}^n)$, $\theta \in \mathscr{D}(\mathbb{R}^n)$, then $u * \theta \in \mathrm{C}^\infty(\mathbb{R}^n)$, $(u*\theta)(x) = \langle u, \theta(x-\cdot)\rangle$, $x \in \mathbb{R}^n$, and $\partial^\alpha(u*\theta) = (\partial^\alpha u) * \theta = u * (\partial^\alpha \theta)$ for all $\alpha \in \mathbb{N}_0^n$. Furthermore,*

$$\operatorname{supp}(u*\theta) \subseteq \operatorname{supp}(u) + \operatorname{supp}(\theta).$$

Proof. We only give the details for the assertion about the supports since the rest has been covered before (note also that we can refer to Theorem 5.9 for the C^∞ and differentiation behind the distribution sign). The definitions of *reflection in the origin* and *translation* given in Example 4.3 carry over to the n-dimensional case without modification. In these terms we then have

$$(u*\theta)(x) = \langle u, \theta(x-\cdot)\rangle = \langle u, \tau_{-x}\tilde{\theta}\rangle \tag{36}$$

for all $x \in \mathbb{R}^n$. Since $u*\theta$ in particular is a continuous function we have that $\operatorname{supp}(u*\theta)$ is the closure of the set $\{x \in \mathbb{R}^n : (u*\theta)(x) \neq 0\}$. In view of (36) and the definition of support for a distribution we have

$$(u*\theta)(x) = 0 \ \text{ whenever } \ \operatorname{supp}(u) \cap \operatorname{supp}(\tau_{-x}\tilde{\theta}) = \emptyset.$$

Note that $\operatorname{supp}(\tau_{-x}\tilde{\theta}) = x - \operatorname{supp}(\theta)$ and that since $\operatorname{supp}(\theta)$ is compact and $\operatorname{supp}(u)$ is closed, the set $A := \left\{x \in \mathbb{R}^n : \operatorname{supp}(u) \cap \operatorname{supp}(\tau_{-x}\tilde{\theta}) = \emptyset\right\}$ is open. We also have that $u*\theta = 0$ on A, so $\operatorname{supp}(u*\theta) \subseteq \mathbb{R}^n \setminus A$. In order to conclude we note that if $x \notin \operatorname{supp}(u) + \operatorname{supp}(\theta)$, so $x - y \notin \operatorname{supp}(u)$ for all $y \in \operatorname{supp}(\theta)$, then $\big(x - \operatorname{supp}(\theta)\big) \cap \operatorname{supp}(u) = \emptyset$, hence $x \in A$. □

Our next task is to define the convolution of a compactly supported distribution and a general C^∞ function: if $v \in \mathscr{E}'(\mathbb{R}^n)$ and $\psi \in \mathrm{C}^\infty(\mathbb{R}^n)$, then we define $v * \psi$ by the rule

$$\langle v*\psi, \varphi\rangle := \langle v, \tilde{\psi} * \varphi\rangle, \quad \varphi \in \mathscr{D}(\mathbb{R}^n). \tag{37}$$

To see that hereby $v*\psi \in \mathscr{D}'(\mathbb{R}^n)$ we first note that $\tilde{\psi}*\varphi$ is well-defined by

$$(\tilde{\psi}*\varphi)(x) = \int_{\mathbb{R}^n} \psi(y-x)\varphi(y)\,\mathrm{d}y, \quad x \in \mathbb{R}^n.$$

It is not difficult to check that $\tilde{\psi}*\varphi \in \mathrm{C}^\infty(\mathbb{R}^n)$ with $\partial^\alpha(\tilde{\psi}*\varphi) = (\partial^\alpha\tilde{\psi})*\varphi = \tilde{\psi}*(\partial^\alpha\varphi)$ for all multi-indices $\alpha \in \mathbb{N}_0^n$. Consequently, $v*\psi\colon \mathscr{D}(\mathbb{R}^n) \to \mathbb{C}$ is well-defined and linear. In order to see that it is also $\mathscr{D}(\mathbb{R}^n)$-continuous we check that it has the boundedness property. Fix a compact set K in $\mathbb{R}^n$ and let $\varphi \in \mathscr{D}(\mathbb{R}^n)$ with support in K. If L is a compact neighbourhood of the support of v, then the boundedness property (34) ensures existence of constants $c \geqslant 0$, $m \in \mathbb{N}_0$ so

$$\begin{aligned}\big|\langle v*\psi, \varphi\rangle\big| &= \big|\langle v, \tilde{\psi}*\varphi\rangle\big| \\ &\leqslant c \sum_{|\alpha|\leqslant m} \sup_L \big|\partial^\alpha(\tilde{\psi}*\varphi)\big| \\ &= c \sum_{|\alpha|\leqslant m} \sup_L \big|(\partial^\alpha\tilde{\psi})*\varphi\big|.\end{aligned}$$

Here we estimate for each $x \in L$:

$$\begin{aligned}\big|\big((\partial^\alpha \tilde{\psi}) * \varphi\big)(x)\big| &\leqslant \int_{\mathbb{R}^n} |\partial^\alpha \tilde{\psi}(y) \varphi(x-y)| \, \mathrm{d}y \\ &\leqslant \int_{L-K} |\partial^\alpha \tilde{\psi}(y)| \, \mathrm{d}y \sup_K |\varphi| \\ &= \int_{K-L} |\partial^\alpha \psi| \, \mathrm{d}y \sup_K |\varphi|\end{aligned}$$

and consequently

$$\begin{aligned}\big|\langle v * \psi, \varphi\rangle\big| \leqslant & c \sum_{|\alpha| \leqslant m} \int_{K-L} |\partial^\alpha \psi| \, \mathrm{d}y \sup_K |\varphi| \\ =& C \sup_K |\varphi|,\end{aligned}$$

where we have defined

$$C := c \sum_{|\alpha| \leqslant m} \int_{K-L} |\partial^\alpha \psi| \, \mathrm{d}y.$$

Because $K-L$ is compact (as image under the continuous map $(x,y) \mapsto x-y$ of the compact set $K \times L$) we see that $C \in [0,\infty)$. Note that we have shown that $v * \psi \in \mathscr{D}_0'(\mathbb{R}^n)$. But in fact the distribution is much more regular as we show next:

Lemma 6.2. *If $v \in \mathscr{E}'(\mathbb{R}^n)$, $\psi \in \mathrm{C}^\infty(\mathbb{R}^n)$, then $v * \psi \in \mathrm{C}^\infty(\mathbb{R}^n)$, $(v * \psi)(x) = \langle v, \psi(x-\cdot)\rangle$, $x \in \mathbb{R}^n$, and $\partial^\alpha(v * \psi) = (\partial^\alpha v) * \psi = v * (\partial^\alpha \psi)$ for all $\alpha \in \mathbb{N}_0^n$. Furthermore,*

$$\operatorname{supp}(v * \psi) \subseteq \operatorname{supp}(v) + \operatorname{supp}(\psi).$$

We omit the details of the proof that are very similar to those of Lemma 6.1. Before proceeding to the third and final step of the convolution definition we make some preliminary observations. First, if φ, $\theta \in \mathscr{D}(\mathbb{R}^n)$, $\psi \in \mathrm{C}^\infty(\mathbb{R}^n)$, then by change of integration order we establish the associative rule:

$$(\psi * \theta) * \varphi = \psi * (\theta * \varphi)$$

and since we already know that the commutative rule holds we can permute the terms in all possible combinations above. Next, if $u \in \mathscr{D}'(\mathbb{R}^n)$ and $v \in \mathscr{E}'(\mathbb{R}^n)$, then we have

$$(u * \theta) * \varphi = u * (\theta * \varphi) \tag{38}$$

and

$$(v * \psi) * \varphi = v * (\psi * \varphi). \tag{39}$$

Let us prove (39) and leave the entirely similar proof for (38) as an exercise:

$$\begin{aligned}\langle (v*\psi)*\varphi,\theta\rangle =&\langle v*\psi,\tilde{\varphi}*\theta\rangle\\ =&\langle v,\tilde{\psi}*(\tilde{\varphi}*\theta)\rangle\\ =&\langle v,(\tilde{\psi}*\tilde{\varphi})*\theta\rangle\\ =&\langle v,(\psi*\varphi)\tilde{}*\theta\rangle\\ =&\langle v*(\psi*\varphi),\theta\rangle\end{aligned}$$

as required. We now come to the third step:

Definition 6.3. *Let $u\in\mathscr{D}'(\mathbb{R}^n)$, $v\in\mathscr{E}'(\mathbb{R}^n)$. Then we define $u*v$, $v*u\in\mathscr{D}'(\mathbb{R}^n)$ by the rules*

$$\langle u*v,\varphi\rangle := \langle u,\tilde{v}*\varphi\rangle,\quad \varphi\in\mathscr{D}(\mathbb{R}^n)$$

and

$$\langle v*u,\varphi\rangle := \langle v,\tilde{u}*\varphi\rangle,\quad \varphi\in\mathscr{D}(\mathbb{R}^n).$$

It follows from Lemmas 6.1 and 6.2 that $\tilde{u}*\varphi\in\mathrm{C}^\infty(\mathbb{R}^n)$ and $\tilde{v}*\varphi\in\mathscr{D}(\mathbb{R}^n)$, respectively, and it is then not difficult to check that $u*v$ and $v*u$ are both well-defined distributions. We omit the definition chasing. Instead we prove

$$(u*v)*\theta = u*(v*\theta) \tag{40}$$

and

$$(v*u)*\theta = v*(u*\theta) \tag{41}$$

hold for all $\theta\in\mathscr{D}(\mathbb{R}^n)$. We prove (40) and leave the proof of (41) as an exercise. For $\varphi\in\mathscr{D}(\mathbb{R}^n)$ we have

$$\begin{aligned}\big\langle (u*v)*\theta,\varphi\big\rangle =&\big\langle u*v,\tilde{\theta}*\varphi\big\rangle\\ =&\big\langle u,\tilde{v}*(\tilde{\theta}*\varphi)\big\rangle\\ \overset{(39)}{=}&\big\langle u,(\tilde{v}*\tilde{\theta})*\varphi\big\rangle\\ =&\big\langle u,(v*\theta)\tilde{}*\varphi\big\rangle\\ =&\big\langle u*(v*\theta),\varphi\big\rangle\end{aligned}$$

as required. Next we calculate for φ, $\theta\in\mathscr{D}(\mathbb{R}^n)$:

$$\begin{aligned}(u*v)*(\varphi*\theta) \overset{(40)}{=}&u*\big(v*(\varphi*\theta)\big)\\ \overset{(39)}{=}&u*\big((v*\varphi)*\theta\big)\\ =&u*\big(\theta*(v*\varphi)\big)\\ \overset{(40)}{=}&(u*\theta)*(v*\varphi)\\ =&(v*\varphi)*(u*\theta)\end{aligned}$$

where we also used the commutative rule for convolution of a test function with a C^∞ function twice. For the standard mollifier on $\mathbb{R}^n$, $\big(\rho_\varepsilon\big)_{\varepsilon>0}$, we take first $\varphi = \rho_\varepsilon$ above and then pass to the limit $\varepsilon \searrow 0$ whereby we get

$$(u * v) * \theta = v * (u * \theta).$$

Here we take $\theta = \rho_\varepsilon$ and then pass to the limit $\varepsilon \searrow 0$ to get

$$u * v = v * u. \tag{42}$$

Above we have used that $u \mapsto u * v$ is continuous in the sense that if $u_j \to u$ in $\mathscr{D}'(\mathbb{R}^n)$, then $u_j * v \to u * v$ in $\mathscr{D}'(\mathbb{R}^n)$, and likewise that $v \mapsto u * v$ is continuous in the sense that if $v_j \to v$ in $\mathscr{D}'(\mathbb{R}^n)$ and for some fixed compact set K in $\mathbb{R}^n$ all the supports of v_j, v are contained in K, then $u * v_j \to u * v$. Both results are straight forward consequences of definitions and so we leave the verifications as an exercise. Instead we turn to the important

Theorem 6.4. *Let $u \in \mathscr{D}'(\mathbb{R}^n)$, $v \in \mathscr{E}'(\mathbb{R}^n)$. Then $\partial^\alpha(u * v) = (\partial^\alpha u) * v = u * (\partial^\alpha v)$ for all $\alpha \in \mathbb{N}_0^n$. Furthermore,*

$$\operatorname{supp}(u * v) \subseteq \operatorname{supp}(u) + \operatorname{supp}(v) \tag{43}$$

and

$$\operatorname{sing.supp}(u * v) \subseteq \operatorname{sing.supp}(u) + \operatorname{sing.supp}(v). \tag{44}$$

Proof. Take $\varphi = \rho_s$, $\theta = \rho_t$ $(s,\, t > 0)$ above to get

$$(u * v) * (\rho_s * \rho_t) = (\rho_s * u) * (\rho_t * v)$$

so by Lemmas 6.1 and 6.2 we get

$$\begin{aligned}\partial^\alpha(u * v) * (\rho_s * \rho_t) &= (\rho_s * \partial^\alpha u) * (\rho_t * v)\\ &= (\rho_s * u) * (\rho_t * \partial^\alpha v).\end{aligned}$$

Taking first $s \searrow 0$ and then $t \searrow 0$ we arrive at $\partial^\alpha(u * v) = (\partial^\alpha u) * v = u * (\partial^\alpha v)$. From (40), $(u * v) * \rho_\varepsilon = u * (v * \rho_\varepsilon)$ and so by Lemmas 6.1 and 6.2 again:

$$\begin{aligned}\operatorname{supp}\big(u * (v * \rho_\varepsilon)\big) &\subseteq \operatorname{supp}(u) + \operatorname{supp}(v * \rho_\varepsilon)\\ &\subseteq \operatorname{supp}(u) + \operatorname{supp}(v) + \operatorname{supp}(\rho_\varepsilon)\\ &= \operatorname{supp}(u) + \operatorname{supp}(v) + \overline{B_\varepsilon(0)}.\end{aligned}$$

If $\varphi \in \mathscr{D}(\mathbb{R}^n)$ and $\big(\operatorname{supp}(u) + \operatorname{supp}(v)\big) \cap \operatorname{supp}(\varphi) = \emptyset$, then because $\operatorname{supp}(u) + \operatorname{supp}(v)$ is a closed set (see Problem Sheet 4) and $\operatorname{supp}(\varphi)$ is compact we have for small $\varepsilon_0 > 0$ that

$$\big(\operatorname{supp}(u) + \operatorname{supp}(v) + \overline{B_{\varepsilon_0}(0)}\big) \cap \operatorname{supp}(\varphi) = \emptyset$$

too. But then for $\varepsilon \in (0, \varepsilon_0)$,

$$0 = \big\langle (u * v) * \rho_\varepsilon, \varphi \big\rangle = \langle u * v, \rho_\varepsilon * \varphi \rangle \\ \to \langle u * v, \varphi \rangle \text{ as } \varepsilon \searrow 0.$$

Therefore $u * v$ vanishes on the open set $\mathbb{R}^n \setminus \big(\operatorname{supp}(u) + \operatorname{supp}(v)\big)$ and (43) follows. We turn to (44) and assume first that both u, v have compact support. Write $u_1 = u$, $u_2 = v$ and $K_i = \operatorname{sing.supp}(u_i)$. Let $\varepsilon > 0$, put $A_i = B_\varepsilon(K_i)$ and take $\psi_i = \rho_\varepsilon * \mathbf{1}_{A_i}$. Then $\psi_i \in \mathscr{D}(\mathbb{R}^n)$, $\psi_i = 1$ on K_i and $\operatorname{supp}(\psi_i) \subseteq \overline{B_{2\varepsilon}(K_i)} = K_i + \overline{B_{2\varepsilon}(0)}$. Now

$$\begin{aligned} u_1 * u_2 =& \big(\psi_1 u_1 + (1 - \psi_i) u_1\big) * \big(\psi_2 u_2 + (1 - \psi_2) u_2\big) \\ =& (\psi_1 u_1) * (\psi_2 u_2) + (\psi_1 u_1) * \big((1 - \psi_2) u_2\big) \\ &+ \big((1 - \psi_1) u_1\big) * (\psi_2 u_2) + \big((1 - \psi_1) u_1\big) * \big((1 - \psi_2) u_2\big). \end{aligned}$$

Since $(1 - \psi_i) u_i \in \mathscr{D}(\mathbb{R}^n)$ it follows from Lemma 6.1 that the last three terms are C^∞ functions. Therefore

$$\begin{aligned} \operatorname{sing.supp}(u_1 * u_2) &= \operatorname{sing.supp}\big((\psi_1 u_1) * (\psi_2 u_2)\big) \\ &\subseteq \operatorname{supp}\big((\psi_1 u_1) * (\psi_2 u_2)\big) \\ &\overset{(43)}{\subseteq} \operatorname{supp}(\psi_1 u_1) + \operatorname{supp}(\psi_2 u_2) \\ &\subseteq \operatorname{supp}(\psi_1) + \operatorname{supp}(\psi_2) \\ &\subseteq K_1 + K_2 + \overline{B_{4\varepsilon}(0)}. \end{aligned}$$

Because $K_1 + K_2$ is compact the inclusion (44) follows by taking the intersection over all $\varepsilon > 0$ on the above right-hand side.

We return to the general case and note that (44) will follow if we can show

$$\operatorname{sing.supp}(u * v) \cap B_1(x) \subseteq \big(\operatorname{sing.supp}(u) + \operatorname{sing.supp}(v)\big) \cap B_1(x)$$

holds for all $x \in \mathbb{R}^n$. Fix $x \in \mathbb{R}^n$ and take $R \geqslant 1$ so large that $\operatorname{supp}(v) \subset B_R(0)$ and $R \geqslant |x|$. Put

$$\psi = \rho_R * \mathbf{1}_{B_{5R}(0)},$$

whereby $\psi \in \mathscr{D}(\mathbb{R}^n)$, $\psi = 1$ on $B_{4R}(0)$ and $\operatorname{supp}(\psi) \subseteq \overline{B_{6R}(0)}$. If we let $u_1 = \psi u$ and $u_2 = (1 - \psi) u$, then $u = u_1 + u_2$ and $\operatorname{supp}(u_1) \subseteq \overline{B_{6R}(0)}$, $\operatorname{supp}(u_2) \subseteq \mathbb{R}^n \setminus B_{4R}(0)$. From (43) we have

$$\operatorname{supp}(u_2 * v) \subseteq \big(\mathbb{R}^n \setminus B_{4R}(0)\big) + B_R(0) \subset \mathbb{R}^n \setminus B_{3R}(0)$$

and since also $B_1(x) \subset B_{2R}(0)$ it follows that $u_2 * v = 0$ on $B_1(x)$. Because u_1, v have compact supports we can apply the first part of the proof to conclude that

$$\operatorname{sing.supp}(u_1 * v) \subseteq \operatorname{sing.supp}(u_1) + \operatorname{sing.supp}(v),$$

and consequently

$$\begin{aligned}\text{sing.supp}(u*v)\cap B_1(x) &= \text{sing.supp}(u_1*v)\cap B_1(x)\\ &\subseteq \big(\text{sing.supp}(u_1)+\text{sing.supp}(v)\big)\cap B_1(x)\\ &\subseteq \big(\text{sing.supp}(u)+\text{sing.supp}(v)\big)\cap B_1(x)\end{aligned}$$

as required. □

Example 6.5. For all $u \in \mathscr{D}'(\mathbb{R}^n)$ we have $u * \delta_0 = u$ and hence for each $\alpha \in \mathbb{N}_0^n$ we then get

$$\partial^\alpha u = u * \partial^\alpha \delta_0.$$

A PDE

$$\sum_{|\alpha|\leqslant k} c_\alpha \partial^\alpha u = f \text{ in } \mathscr{D}'(\Omega)$$

can therefore be equivalently expressed as a convolution equation $g * u = f$ in $\mathscr{D}'(\Omega)$, where

$$g = \sum_{|\alpha|\leqslant k} c_\alpha \partial^\alpha \delta_0.$$

6.2 Fundamental solutions.

A linear differential operator with constant coefficients can conveniently be expressed in multi-index notation as

$$p(\partial) = \sum_{|\alpha|\leqslant k} c_\alpha \partial^\alpha \qquad (c_\alpha \in \mathbb{C})$$

If for some multi-index α of length k we have $c_\alpha \neq 0$, then $p(\partial)$ is said to be of order k. If also the coefficients c_α satisfy $c_\alpha = 0$ for all $|\alpha| < k$, then we say that $p(\partial)$ is homogeneous of order k. Thus the Laplace operator Δ is homogeneous of order 2 and the Cauchy-Riemann operators $\partial/\partial\bar{z}$ and $\partial/\partial z$ are both homogeneous of order 1.

In the study of questions of existence and regularity of solutions to PDEs the following notion is often useful:

Definition 6.6. *A distribution $E \in \mathscr{D}'(\mathbb{R}^n)$ is called a* fundamental solution *of the differential operator $p(\partial)$ if $p(\partial)E = \delta_0$*

The importance of fundamental solutions is due to the following two consequences of the definition and Theorem 6.4:

$$E * p(\partial)u = u \quad \text{for } u \in \mathscr{E}'(\mathbb{R}^n), \tag{45}$$

$$p(\partial)(E * f) = f \quad \text{for } f \in \mathscr{E}'(\mathbb{R}^n). \tag{46}$$

Thus convolution with E is both a left and a right inverse of $p(\partial)$ on $\mathscr{E}'(\mathbb{R}^n)$ (mapping into $\mathscr{D}'(\mathbb{R}^n)$), and it follows that the PDE $p(\partial)u = f$ has a solution $u = E * f \in \mathscr{D}'(\mathbb{R}^n)$ for every $f \in \mathscr{E}'(\mathbb{R}^n)$. The identity (45) makes it possible to obtain information on say the singularities of u from those of f.

In Examples 4.22 and 4.23 we found fundamental solutions to the Laplacian and the Cauchy-Riemann operators, respectively. We shall have more to say about the topic of fundamental solutions in the course B4.4 *Fourier Analysis*, but at this stage we record that the fundamental solutions E we found for the Laplace and Cauchy-Riemann operators satisfy $\text{sing.supp}(E) = \{0\}$ and therefore that the following result applies in these cases:

Theorem 6.7. *Assume that E is a fundamental solution to the differential operator $p(\partial)$ and that $\text{sing.supp}(E) = \{0\}$. Then for any nonempty open subset Ω of $\mathbb{R}^n$ we have*

$$\text{sing.supp}(u) = \text{sing.supp}(p(\partial)u)$$

for all $u \in \mathscr{D}'(\Omega)$.

Proof. On Problem Sheet 3 we saw that it is always true that $\text{sing.supp}(p(\partial)u) \subseteq \text{sing.supp}(u)$. For $u \in \mathscr{E}'(\mathbb{R}^n)$ we obtain from (45) and Theorem 6.4 that

$$\text{sing.supp}(u) = \text{sing.supp}\big(E * (p(\partial)u)\big) \subseteq \text{sing.supp}\big(p(\partial)u\big)$$

showing that the assertion is valid for $u \in \mathscr{E}'(\mathbb{R}^n)$. In the general case we fix an open subset $\omega \Subset \Omega$ and take $\psi \in \mathscr{D}(\Omega)$ so $\psi = 1$ on ω. It follows that $\psi u \in \mathscr{E}'(\mathbb{R}^n)$ if we define $\langle \psi u, \varphi \rangle := \langle u, \psi\varphi \rangle$ for $\varphi \in \mathrm{C}^\infty(\mathbb{R}^n)$, and hence

$$\begin{aligned}\omega \cap \text{sing.supp}\big(p(\partial)u\big) &= \omega \cap \text{sing.supp}\big(p(\partial)(\psi u)\big)\\ &= \omega \cap \text{sing.supp}(\psi u)\\ &= \omega \cap \text{sing.supp}(u)\end{aligned}$$

as required. □

6.3 Elliptic regularity

The next result is classical and is due to Hermann Weyl. He published it in 1940 in a form that in the terminology of this course would correspond to considering only regular distributions. The extension to general distributions is however routine and is also named after him:

Theorem 6.8 (Weyl's Lemma)**.** *Assume $u \in \mathscr{D}'(\Omega)$ and $\Delta u = 0$ in $\mathscr{D}'(\Omega)$. Then $u \in \mathrm{C}^\infty(\Omega)$ and u is harmonic.*

Corollary 6.9. *Let $\Omega \subset \mathbb{C}$ be open and assume $f \in \mathscr{D}'(\Omega)$ satisfies*

$$\frac{\partial f}{\partial \bar{z}} = 0$$

in $\mathscr{D}'(\Omega)$. Then f is holomorphic.

Proof. This is clear since we obviously also have that

$$\Delta f = 4\frac{\partial}{\partial z}\left(\frac{\partial}{\partial \bar{z}} f\right) = 0$$

in $\mathscr{D}'(\Omega)$. Weyl's Lemma then implies that f is C^∞, in which case distributional and classical derivatives coincide. Thus f satisfies the usual Cauchy–Riemann equations and is a holomorphic function. □

Note that in view of Example 4.22 we can deduce Weyl's lemma from Theorem 6.7. Here we shall give a proof that is independent of this and in the spirit of Weyl's original proof.

Proof of Weyl's Lemma. Let $(\rho_\varepsilon)_{\varepsilon>0}$ be the standard mollifier. Fix $\Omega' \Subset \Omega$ and put $\varepsilon_0 = \mathrm{dist}(\Omega', \partial\Omega)$. For each $x \in \Omega'$ and $\varepsilon \in (0, \varepsilon_0)$ the function

$$y \longmapsto \rho_\varepsilon(x-y)$$

belongs to $\mathscr{D}(\Omega)$ and so we may consider $\langle u, \rho_\varepsilon(x-\cdot)\rangle$. We assert that it is independent of $\varepsilon \in (0, \varepsilon_0)$. To prove it we calculate $\frac{\mathrm{d}}{\mathrm{d}\varepsilon}\rho_\varepsilon(x-y)$ for $x, y \in \mathbb{R}^n$. Recall that

$$\rho_\varepsilon(x-y) = \varepsilon^{-n}\rho\left(\frac{x-y}{\varepsilon}\right)$$

where the standard mollifier kernel ρ on $\mathbb{R}^n$ was defined at (12). If we put

$$\theta(t) = \begin{cases} \frac{1}{c_n}\mathrm{e}^{\frac{1}{t-1}} & \text{if } t < 1, \\ 0 & \text{if } t \geqslant 1, \end{cases}$$

then $\rho(x) = \theta(|x|^2)$. Clearly $\theta \in \mathrm{C}^\infty(\mathbb{R})$ satisfies $\theta(t) = 0$ for $t \geqslant 1$. Now calculate

$$\begin{aligned} \frac{\mathrm{d}}{\mathrm{d}\varepsilon}\left(\varepsilon^{-n}\rho\left(\frac{x-y}{\varepsilon}\right)\right) &= -n\varepsilon^{-n-1}\rho\left(\frac{x-y}{\varepsilon}\right) - \varepsilon^{-n}\nabla\rho\left(\frac{x-y}{\varepsilon}\right)\cdot\frac{x-y}{\varepsilon^2} \\ &= -\frac{1}{\varepsilon^{n+1}}\left(n\rho\left(\frac{x-y}{\varepsilon}\right) + \nabla\rho\left(\frac{x-y}{\varepsilon}\right)\cdot\frac{x-y}{\varepsilon}\right). \end{aligned}$$

Put $K(x) = -n\rho(x) - \nabla\rho(x)\cdot x$ so that

$$\frac{\mathrm{d}}{\mathrm{d}\varepsilon}\left(\varepsilon^{-n}\rho\left(\frac{x-y}{\varepsilon}\right)\right) = \frac{1}{\varepsilon^{n+1}}K\left(\frac{x-y}{\varepsilon}\right).$$

In terms of $\rho(x) = \theta(|x|^2)$ we get

$$K(x) = -\operatorname{div}\left(\rho(x)x\right) = -\operatorname{div}\left(\theta(|x|^2)x\right)$$

and if we set

$$\Theta(t) = \frac{1}{2}\int_t^\infty \theta(s)\,\mathrm{d}s,$$

then $\Theta \in \mathrm{C}^\infty(\mathbb{R})$ with $\Theta(t) = 0$ for $t \geqslant 1$, and $\Theta'(t) = -\frac{1}{2}\theta(t)$. Consequently

$$-\theta(|x|^2)x = \nabla\left(\Theta(|x|^2)\right),$$

and so $K(x) = \operatorname{div}\nabla\left(\Theta(|x|^2)\right) = (\Delta\Phi)(x)$, where $\Phi(x) = \Theta(|x|^2)$. Observe that $\Phi \in \mathscr{D}(\overline{B_1(0)})$, and

$$\begin{aligned} -\frac{1}{\varepsilon^{n+1}}\left(n\rho\left(\frac{x-y}{\varepsilon}\right) + \nabla\rho\left(\frac{x-y}{\varepsilon}\right)\cdot\frac{x-y}{\varepsilon}\right) &= \frac{1}{\varepsilon^{n+1}}\Delta_y\left(\Phi\left(\frac{x-y}{\varepsilon}\right)\right) \\ &= \Delta_y\left(\varepsilon^{1-n}\Phi\left(\frac{x-y}{\varepsilon}\right)\right). \end{aligned}$$

Here $y \mapsto \varepsilon^{1-n}\Phi\left(\frac{x-y}{\varepsilon}\right)$ is supported in $\overline{B_\varepsilon(x)} \subset \Omega$, and so by assumption

$$\left\langle u,\, \Delta_y\left(\varepsilon^{1-n}\Phi\left(\frac{x-y}{\varepsilon}\right)\right)\right\rangle = 0.$$

Now by considering difference quotients we see that

$$\frac{\mathrm{d}}{\mathrm{d}\varepsilon}\langle u,\, \rho_\varepsilon(x-\cdot)\rangle = \left\langle u,\, \frac{\mathrm{d}}{\mathrm{d}\varepsilon}\rho_\varepsilon(x-\cdot)\right\rangle.$$

Indeed, for $\varepsilon,\, \varepsilon' > 0$ we have

$$\begin{aligned} \frac{\rho_{\varepsilon+\varepsilon'}(x-y) - \rho_\varepsilon(x-y)}{\varepsilon'} &\overset{\text{FTC}}{=} \int_0^1 \frac{\mathrm{d}}{\mathrm{d}t}\rho_{\varepsilon+t\varepsilon'}(x-y)\,\mathrm{d}t \\ &\underset{\varepsilon'\searrow 0}{\longrightarrow} \left.\frac{\mathrm{d}}{\mathrm{d}s}\right|_{s=\varepsilon}\rho_s(x-y) \end{aligned}$$

in $\mathscr{D}'(\Omega)$ with respect to y, provided $x \in \Omega'$ and $0 < \varepsilon < \varepsilon_0$ (since we may differentiate both sides with respect to y). But then $\frac{\mathrm{d}}{\mathrm{d}\varepsilon}\langle u,\, \rho_\varepsilon(x-\cdot)\rangle = 0$, and so $\langle u,\, \rho_\varepsilon(x-\cdot)\rangle = \langle u,\, \rho_{\varepsilon_1}(x-\cdot)\rangle$ for all $\varepsilon \in (0,\varepsilon_0)$, where $\varepsilon_1 \in (0,\varepsilon_0)$. Now let $\varphi \in \mathscr{D}(\Omega')$. Then, by the usual trick when convolving distributions with test functions,

$$\begin{aligned} \int_{\Omega'}\langle u,\, \rho_\varepsilon(x-\cdot)\rangle\varphi(x)\,\mathrm{d}x &= \left\langle u,\, \int_{\Omega'}\rho_\varepsilon(x-\cdot)\varphi(x)\,\mathrm{d}x\right\rangle \\ &= \langle u,\, \rho_\varepsilon * \varphi\rangle, \end{aligned}$$

and so for $\varepsilon \in (0,\varepsilon_1)$ we have

$$\langle u,\, \rho_\varepsilon * \varphi\rangle = \int_{\Omega'}\langle u,\, \rho_{\varepsilon_1}(x-\cdot)\rangle\varphi(x)\,\mathrm{d}x.$$

Hence, as $\rho_\varepsilon * \varphi \to \varphi$ in $\mathscr{D}(\Omega)$ as $\varepsilon \searrow 0$, we get

$$\langle u,\, \varphi\rangle = \int_{\Omega'}\langle u,\, \rho_{\varepsilon_1}(x-\cdot)\rangle\varphi(x)\,\mathrm{d}x.$$

Consequently $u|_{\Omega'} \in \mathrm{C}^\infty(\Omega')$, and since Ω' was arbitrary, we are done. □

Remark 6.10. The above proof is inspired by the mean value property that is known to characterize harmonic functions in the following sense. Let $h \in \mathrm{C}(\Omega)$. Then h is harmonic in the usual sense (so $h \in \mathrm{C}^2(\Omega)$ and $\Delta h = 0$) if and only if for all balls $B_r(x_0) \Subset \Omega$ we have

$$h(x_0) = \frac{1}{\omega_{n-1} r^{n-1}} \int_{\partial B_r(x_0)} h(x) \, \mathrm{d}S_x.$$

Using polar coordinates about x_0 we see that when h is harmonic, then for $B_r(x_0) \Subset \Omega$

$$h(x_0) = (\rho_r * h)(x_0).$$