

Chapter 2

General Concepts

2.1 Types of Second Order PDEs

For PDEs of order greater than one, there does not exist a general theory. Over the time there have been discovered different methods to solve several PDEs, in particular those PDEs which show up in physics. Afterwards these methods were extended to larger and larger classes of PDEs. It turned out that the successful methods of solving PDEs differ from each other substantially. As a result there does not exist one unified theory of PDEs, but there exist several islands of well understood families of PDEs inside the large set of all PDEs. It was Jacobi who formulated in his lectures on Dynamics in the years 1842–43 the following general recipe:

“The main obstacle for the integration of a given differential equations lies in the definition of adapted variables, for which there is no general rule. For this reason we should reverse the direction of our investigation and should endeavour to find, for a successful substitution, other problems which might be solved by the same.”

The strategy is to determine for any successful method all PDEs which can be solved by this method. We have seen that the method of characteristics is a more-or-less general method to solve first order PDEs. Now we investigate the second order PDEs. In this lecture we consider only second order linear PDEs. A general second order linear PDE has the following form

$$Lu(x) = \sum_{i,j=1}^n a_{ij}(x) \partial_i \partial_j u + \sum_{i=1}^n b_i(x) \partial_i u(x) + c(x)u(x) = 0.$$

By Schwarz’s Theorem for twice differentiable u this expression does not change if we replace a_{ij} by $\frac{1}{2}(a_{ij} + a_{ji})$. So we always assume that a_{ij} is symmetric.

It is a theorem of linear algebra that a real symmetric matrix can be diagonalised (by

an orthogonal matrix even) and the eigenvalues are all real. We use the signs of these eigenvalues as our classification.

Definition 2.1. *On an open domain $\Omega \subset \mathbb{R}^n$ a partial differential operator L*

$$Lu = \sum_{i,j=1}^n a_{ij}(x) \frac{\partial^2 u(x)}{\partial x_i \partial x_j} + \sum_{i=1}^n b_i(x) \frac{\partial u(x)}{\partial x_i} + c(x)u(x)$$

with symmetric coefficients $a_{ij} = a_{ji}$ is called

- *elliptic if all the eigenvalues of a are non-zero and have the same sign.*
- *parabolic if one eigenvalue of a is zero and the others have the same sign.*
- *hyperbolic if one eigenvalue of a has the opposite sign to all the others.*

This is not a complete classification, in the sense that there are PDEs which are not of these three types. For example, the a matrix of a PDE may have several positive, zero, and negative eigenvalues. If a is not constant, then of course the eigenvalues are functions of x and may be different at different points. An operator might only be elliptic on a subdomain of Ω . And this definition only applies to second order linear PDEs. None-the-less we will concentrate on PDEs of these three types; this is an introductory course.

Elliptic PDEs. As we have just stated, for an elliptic PDE the eigenvalues of a all have the same sign. Any solution of $Lu = 0$ is also a solution of $-Lu = 0$, so we can always arrange for the eigenvalues to be positive. There is an equivalent condition that is more often taken as the definition of ellipticity: the operator L is elliptic if and only if

$$v^T a v = \sum_{i,j=1}^n a_{ij}(x) v_i v_j > 0 \quad \text{for all } x \in \Omega \text{ and all } v \in \mathbb{R}^n \setminus \{0\}. \quad (2.1)$$

To see this equivalence, write $a = O^T D O$ for O an orthogonal matrix and $D = \text{diag}(\lambda_1, \dots, \lambda_n)$ the eigenvalues. Observe

$$v^T a v = v^T O^T D O v = (Ov)^T D (Ov) = w^T D w,$$

so a satisfies (2.1) if and only if D does. Now let $w = e_j$ be the j th standard basis vector. This yields $w^T D w = \lambda_j$. Hence (2.1) holds if and only if all the eigenvalues are positive.

Now we consider some concrete examples. If the matrix a_{ij} is the identity matrix and $b = 0, c = 0$, then this is the

Laplace equation.
$$\Delta u := \frac{\partial^2 u}{\partial x_1^2} + \dots + \frac{\partial^2 u}{\partial x_n^2} = 0.$$

Solutions of the Laplace equation are called harmonic functions. In Chapter 3 we present several tools which establish many properties of these harmonic functions. It turns out that many properties of the harmonic functions also apply to general solutions of $Lu = 0$, if the matrix a_{ij} is positive (or negative) definite. There has been done a lot of work to extend these tools to larger and larger classes of PDEs. One of the results is that the influence of the higher order derivatives on the properties of solutions is much more important than the influence of the lower order derivatives. We offer another lecture which presents many of these tools for such elliptic second order PDEs.

There are also non-linear PDEs to which these methods of elliptic PDEs apply. An important example whose investigation played a major role in the development of the elliptic theory is the

Minimal surface equation. $\nabla \cdot \frac{\nabla u}{\sqrt{1 + |\nabla u|^2}} = 0, \quad u : \Omega \rightarrow \mathbb{R}, \quad \Omega \subset \mathbb{R}^n \text{ open.}$

The graphs of solutions describe so called minimal surfaces. The area of such hypersurfaces in \mathbb{R}^{n+1} does not change with respect to infinitesimal variations. Soap bubbles are examples of such minimal surfaces. The boundary value problem of the minimal surface equation is called Plateau's problem. For the first proof of the existence of solutions of this Plateau problem in the 1930s, Jesse Douglas received the first Field's Medal. In this non-linear second order PDE the coefficients of the second derivatives also depend on the solution. A lot of work has been done to extend the tools of elliptic theory to elliptic PDEs whose coefficients belong to larger and larger functions spaces. This development induced the introduction of many new function spaces. In Section 2.4 we shall introduce the so called space of distributions. Many of the more advanced functions spaces are special subsets of the distributions.

Parabolic PDEs. For these linear PDEs the matrix a_{ij} considered as a symmetric bilinear form is only semi-definite and they belong to the boundary of the class of elliptic PDEs. Most of the methods of elliptic PDEs have an extension to this limiting case. So these limiting cases together with the class of elliptic PDEs form some extended class of elliptic PDEs. In spite of the deep relationship to the elliptic PDEs these equations have their own label. The simplest example is the:

Heat equation. $\dot{u} - \Delta u = 0.$

These parabolic PDEs describe diffusion processes. These are processes which level inhomogeneities of some quantity by some flow along the negative gradient of the quantity. A typical example for this quantity is the temperature from which the name for the heat equation originates. Many stochastic processes have this property. So the theory of parabolic PDEs has a deep relationship to the theory of stochastic processes. In this lecture we present in Chapter 4 this simplest example of linear parabolic PDE. We shall see how the tools for the Laplace equation can be applied in modified form to this heat equation. In case of the parabolic PDEs there also exists a non-linear example from the geometric analysis, whose investigation played a major role for the development of the

elliptic theory (the tensor fields g and R are defined below):

Ricci Flow. $\dot{g}_{ij} = -2R_{ij}$.

This PDE describes a diffusion-like process on Riemannian manifolds. It levels the inhomogeneities of the metric, namely the Riemannian metric g . In the long run the corresponding Riemannian manifolds converge to metric spaces with large symmetry groups. Richard Hamilton proposed (in the 1970s) a program that aims to prove the geometrization conjecture of Thurston with the help of these PDEs. It states that every three-dimensional manifold can be split into parts, which can be endowed with an Riemannian metric such that the isometry group acts transitively. This conjecture implies the Poincaré conjecture, which states that every simply connected compact manifold is the 3-sphere. Hamilton tries to control the long time limit of the Ricci flow on a general 3-dimensional Riemannian manifold. In 2003 the Russian mathematician Grisha Perelman published on the internet three articles which overcome the last obstacle of this program. This led to the first proof of one of the Millennium Problems of the American Mathematical Society and was a great success of geometric analysis.

Hyperbolic PDEs. Besides the elliptic PDEs (including the limiting cases) the other important class of linear PDEs are called hyperbolic. In this case the matrix a_{ij} has one eigenvalue of opposite sign than all other eigenvalues. The simplest example is the

Wave equation. $\frac{\partial^2 u}{\partial t^2} - \Delta u = 0$.

In Chapter 5 we present the methods how to solve this equation. We shall see that it describes the propagation of waves with constant finite speed. The solutions of general hyperbolic equations are similar to the solutions of this case, and many tools can be generalised to all hyperbolic PDEs. The investigation of these PDEs depends on the understanding of all trajectories, which propagate by the given speed. It was motivated by the theory of the electrodynamic fields, whose main system of PDEs are the

Maxwell equations.

$$\begin{aligned} \dot{E} - \nabla \times B &= -4\pi j & \dot{B} + \nabla \times E &= 0 \\ \nabla \cdot E &= 4\pi \rho & \nabla \cdot B &= 0. \end{aligned}$$

In this theory there is given a distribution of charges ρ and currents j on space time $\mathbb{R} \times \mathbb{R}^3$. The unknown functions are the electric magnetic fields E and B , which describe the electrodynamic forces induced by the given distributions of charges and currents ρ and j . The conservation of charge is formulated in the same way as in the scalar conservation law. So the change of the total charge contained in a spatial domain is described by the flux of the current through the boundary of the domain. By the divergence theorem this means that distributions of charge ρ and currents j obey

$$\dot{\rho} + \nabla \cdot j = 0.$$

Again there exists a non-linear version which stimulated the development of the theory:

Einstein's field equations of general relativity. $R_{ij} - \frac{1}{2}g_{ij}R = \kappa T_{ij}.$

Here for a given distribution of masses the energy stress tensor and the space time metric g_{ij} are the unknown functions. This metric is a symmetric bilinear form with one positive and three negative eigenvalues on the tangent space of space time. The corresponding Ricci curvature is denoted by R_{ij} and the scalar curvature by R :

$$\Gamma_{ij}^k := \frac{1}{2} \sum_{l=0}^3 g^{kl} \left(\frac{\partial g_{jl}}{\partial x^i} + \frac{\partial g_{il}}{\partial x^j} - \frac{\partial g_{ij}}{\partial x^l} \right), \quad (g^{ij}) := (g_{ij})^{-1} \text{ inverse metric}$$

$$R_{ij} := \sum_{k=0}^3 g^{kl} \left(\frac{\partial \Gamma_{ij}^k}{\partial x^k} - \frac{\partial \Gamma_{ik}^k}{\partial x^j} + \sum_{l=0}^3 (\Gamma_{lk}^k \Gamma_{ij}^l - \Gamma_{lj}^k \Gamma_{ik}^l) \right), \quad R := \sum_{i,j=0}^3 g^{ij} R_{ij}.$$

Integrable Systems with Lax operators. Finally I want to mention a smaller class of PDEs, which are the main objects of my research. They are non-linear PDEs which describe an evolution with respect to time which is very stable. This means that the solutions have in a specific sense a maximal number of conserved quantities. The theory of integrable systems belongs to the field of Hamiltonian mechanics, which originated from Newton's description of the motion of the planets. The Scottish Lord John Scott Russell got very excited in 1934 about the observation of an solitary wave in a Scottish channel and published a "Report on Waves". This report was quite influential. The two Dutch mathematicians Korteweg and De Vries translated his observation into a PDE describing the profile of water waves travelling along the channel:

Korteweg-de-Vries equation. $4\dot{u} - 6u \frac{\partial u}{\partial x} - \frac{\partial^3 u}{\partial x^3} = 0.$

First by numerical experiments in the 1950s with the first computers and latter in the 1970s by mathematical theory, the solutions of this PDE were shown to have exactly the properties which made Lord Russell so excited: they describe waves which propagate through each other without changing their shape. This led to the discovery of an hidden relation of the theory of integrable systems with the theory of Riemann surfaces, which is another field with a long history. A major step towards the discovery of this relation was the observation of Peter Lax that this equation can be written as

$$\dot{L} = [A, L] \quad \text{with} \quad L := \frac{\partial^2}{\partial x^2} + u \quad A := \frac{\partial^3}{\partial x^3} + \frac{3u}{2} \frac{\partial}{\partial x} + \frac{3}{4} \frac{\partial u}{\partial x}.$$

2.2 The Questions

In addition to the types of PDEs we will study, let us explain some of the questions that we are interested in answering. Broadly speaking they are *existence*, *uniqueness*, and *regularity*.

The first two you probably have some experience with, so let us begin with the third. The regularity of a solution of a differential equation refers to its properties. Most often this is its differentiability, for example, twice continuously differentiable. But it can also be about boundedness, integrability, or the growth/decay rate of the function. These regularity properties are usually expressed in the form of a function space, e.g. C^2 .

Many types of regularity can be ordered in a hierarchy. Later in this chapter we will introduce distributions. We say have the lowest regularity. Then comes the measurable functions. Some measurable functions are locally integrable, L^1_{loc} . A function f belongs to L^1_{loc} if for every compact set K in its domain $|f|_K$ has a finite integral. You might also know the Lebesgue norms. The elements of L^p_{loc} describe ever smaller families of functions, whose regularity increase with $p \in [1, \infty]$. The next smallest class are Sobolev functions whose k -th order partial derivatives belong to L^p_{loc} . The regularity further increases for the functions in C^k . Finally we end with the smooth functions and the analytic functions with the highest regularity.

Consider the analogy to the algebraic equation $x^2 + 1 = 0$. This has no solutions $x \in \mathbb{R}$ but two solutions $x \in \mathbb{C}$. Likewise, the number of solutions to a PDE can change depending on which functions we are considering. To be concrete, perhaps there are many C^2 solutions, but no bounded C^2 solutions. There is usually a natural level of regularity to require of a solution: the solution to a second order PDE should be twice differentiable. But as we have seen in the previous chapter, it is sometimes necessary to change the meaning of ‘solution’ and consider ‘non-differentiable solutions’ (weak solutions) to a PDE, even if that sounds like a contradiction.

Sometimes allowing lower regularity doesn’t increase the number of solutions of a differential equation. A classic example from ODE theory is $u' = u$. Suppose u is a differentiable solution to this equation. But then $u' = u$ tells us that u' is equal to a differentiable function. That means that u' is differentiable, i.e. that u is twice differentiable. Repeating this argument, we see that u is infinitely many times differentiable (smooth). We say that the solution has higher regularity than expected. The Laplace and heat equations both exhibit highly regular solutions.

A *problem* is a differential equation on some domain together with some additional (non PDE) conditions. A typical ODE has infinitely many solutions, but a typical partial differential equations has an infinite dimensional space of solutions. The idea is to give the right additional conditions so that the problem has a unique solution. A solution of an ordinary differential equations of m -th order is in many cases uniquely determined by fixing the values of the first m derivatives at $t = 0$. For partial differential equations the solutions are functions on higher dimensional domains $\Omega \subset \mathbb{R}^n$. A natural condition is the specification of the values of the solution and some of its derivatives on the boundary of the domain or on a hypersurface within the domain. The search for solutions which obey this further specification are called boundary value problems. When one of the variables represents time and we give conditions at $t = 0$, naturally we call this an initial value

problem.

We are most interested in problems that are *well-posed*. This means that (1) the problem has a solution, (2) the solution is unique, and (3) the solution depends continuously on the data. We have to balance the choice of regularity and how much data we give so that the problem has a unique solution, but not too much that it doesn't have any solution. If we determine all possible boundary values that have solutions, then the space of solutions is completely parameterised. Again to give an ODE analogy, the solutions of $u' = u$ are $u(t) = Ae^t$, so $A \in \mathbb{R}$ parameterises the solutions.

Finally, there is the question of existence. This is perhaps the most fundamental question, because it is about the definition of 'a solution'. We have already mentioned how this is affected by regularity (including weak solutions) and boundary value conditions. But proving the existence of solutions of PDEs is in general much more difficult than for ODEs, and there are not too many general theorems that we have. In fact, there is a famous example of a simple-looking PDE that does not have any solutions, not even locally. This example is a simplification by Nirenberg of an example of H. Lewy: there is no solution $u : \Omega \rightarrow \mathbb{C}$ on any open subset $\Omega \subset \mathbb{R}^2$ of

$$\frac{\partial u}{\partial x} + ix \frac{\partial u}{\partial y} = f(x, y),$$

where f is a specially constructed smooth function. Notice in particular that this is first order PDE, with analytic coefficients and a smooth inhomogeneous term. In previous years we gave a proof of this, but it requires certain facts from complex analysis (aka Funktionstheorie) that many students didn't understand. Interested students may ask me for a copy of the old script. Instead we will give a different example of a PDE without a solution in Section 3.5 using the techniques that fit the themes of this course.

2.3 Divergence Theorem

In this section we present a generalisation of the fundamental theorem of calculus to higher dimensions, namely the divergence theorem. This theorem has many important consequences. In this section we present two: First we generalise partial integration to higher dimensions. Second we explain in which sense the higher dimensional scalar conservation law describes a conserved quantity.

The divergence theorem is a statement about the integral over a submanifold of \mathbb{R}^n , so naturally we should define submanifolds and their integrals.

Definition 2.2. *The graph of a function $\lambda : U \subset \mathbb{R}^k \rightarrow \mathbb{R}^{n-k}$ is the k -dimensional subset*

$$\text{graph}(\lambda) = \{(x, y) \in U \times \mathbb{R}^{n-k} \mid y = \lambda(x)\} \subset \mathbb{R}^n.$$

We allow reordering of the components of \mathbb{R}^n . By this we mean that both $\{(x, y, x^2 + y^2) \in \mathbb{R}^3\}$ and $\{(x, x^2 + z^2, z) \in \mathbb{R}^3\}$ are graphs, for example.

A subset $A \subset \mathbb{R}^n$ is called a k -dimensional submanifold if it is a k -dimensional graph locally. That means for every point $x \in A$ there exists an open $O_x \subset \mathbb{R}^n$ such that $A_x := A \cap O_x$ is a k -dimensional graph.

We say that a graph or submanifold is C^l (or some other regularity) if it is locally the graph of a function $\lambda \in C^l$.

The classic, and for us important, example of a submanifold that is not a graph globally is the circle $\mathbb{S}^1 = \{(x, y) \in \mathbb{R}^2 \mid x^2 + y^2 = 1\}$. This is because $y = \pm\sqrt{1 - x^2}$ is not a function. However it can be written as the union of four graphs

$$\begin{aligned} \mathbb{S}^1 = & \{(x, +\sqrt{1 - x^2}) \mid x \in (-1, 1)\} \cup \{(x, -\sqrt{1 - x^2}) \mid x \in (-1, 1)\} \\ & \cup \{(+\sqrt{1 - y^2}, y) \mid y \in (-1, 1)\} \cup \{(-\sqrt{1 - y^2}, y) \mid y \in (-1, 1)\}. \end{aligned}$$

For practical calculation it is not always the best idea to reduce a submanifold to graphs. Often a parameterisation can cover more of the submanifold, which means less work.

Definition 2.3. A continuously differentiable injection $\Phi : U \subset \mathbb{R}^k \rightarrow A \subset \mathbb{R}^n$ is called a parameterisation of a submanifold A . It is called regular if the Jacobian Φ' has rank k at every point of U .

The Jacobian of Φ is an $n \times k$ matrix, so its rank cannot be greater than k . Thus a regular parameterisation is also called full-rank. A graph is a special type of parameterisation, one where k of the components of Φ are just the k input variables. In other words $\Phi(x) = (x, \lambda(x))$, or some rearrangement. This is always a regular parameterisation, because Φ' contains the $k \times k$ identity matrix. For an example of a non-regular parameterisation, consider the parameterisation $(x, y) \mapsto (x, 0, 0)$ of the x -axis in \mathbb{R}^3 . We see that y is not really playing any role and the submanifold is only one-dimensional. This is the reason we should consider regular parameterisations.

Definition 2.4. Let $A \subset \mathbb{R}^n$ be a subset with a regular parameterisation Φ and f a continuous function on A . We define

$$\int_A f \, d\sigma := \int_U f \circ \Phi \sqrt{\det((\Phi')^T \Phi')} \, d\mu_{\mathbb{R}^k}.$$

The symbol $d\sigma$ can be given a formal meaning, but for us it is just a reminder that it is a ‘submanifold integral’ and not an integral on a subset of \mathbb{R}^n in the usual sense. The k -dimensional parallelotope spanned by the k column vectors of a $n \times k$ -matrix A has the volume $\sqrt{\det(A^T A)}$. The motivation for the $\sqrt{\det}$ factor in the definition of the integral is that it measures the distortion of the parameterisation. This value turns out to be independent of the choice of regular parameterisation of A .

Lemma 2.5. *Let $A = \text{graph}(\lambda)$. Then the value of $\int_A f \, d\sigma$ does not depend on the choice of regular parameterisation.*

Proof. $\Phi(x) = (x, \lambda(x))$ is a regular parameterisation (without loss of generality we can relabel the coordinates to achieve this form). Suppose that we have another regular parameterisation $\Psi : V \rightarrow A$. Then define $\Upsilon = \Phi^{-1} \circ \Psi : V \rightarrow U$. We claim that Υ is continuously differentiable. This is not so clear, because Φ^{-1} is only defined on A , not on a euclidean space, and so we can't apply the chain rule directly. However, let $\Pi : \mathbb{R}^n \rightarrow \mathbb{R}^k$ be the projection $\Pi(x, y) = x$. Clearly $\Pi \circ \Phi(x) = \Pi(x, \lambda(x)) = x$. This shows that $\Phi^{-1} = \Pi|_A$. Therefore $\Upsilon = \Pi \circ \Psi$ is another formula for Υ . Now we can apply the chain rule and conclude Υ is continuously differentiable.

Now we can carry out an computation that connects the two integrals

$$\begin{aligned} & \int_V f \circ \Psi \times \sqrt{\det((\Psi')^T \Psi')} \, d\mu_{\mathbb{R}^k} \\ &= \int_V f \circ \Phi \circ \Upsilon \sqrt{\det(((\Phi \circ \Upsilon)')^T (\Phi \circ \Upsilon)')} \, d\mu_{\mathbb{R}^k} \\ &= \int_V \left(f \circ \Phi \sqrt{\det((\Phi')^T \Phi')} \right) \circ \Upsilon \, |\det \Upsilon'| \, d\mu_{\mathbb{R}^k} \\ &= \int_U f \circ \Phi \sqrt{\det((\Phi')^T \Phi')} \, d\mu_{\mathbb{R}^k}. \end{aligned}$$

In the last step we applied the transformation formula of Jacobi. This shows that using any parameterisation gives the same result as using the graph parameterisation. \square

This is a very practical definition in that it gives you a concrete integral to compute. However many submanifolds that we want to consider cannot be covered by a single parameterisation. The typical example is the sphere: any open set $U \subset \mathbb{R}^k$ is not compact and the sphere is compact, so there cannot exist a homeomorphism Φ between them. However if we use two parameterisations, then each can cover a part of the sphere and together they can cover the whole sphere. The trouble is now these A_i can overlap, so if we just integrate in each parameterisation then we will 'double-count' the points of A . The answer to this is an elegant theoretical tool, but one that is not practically useful: a so called *partitions of unity*.

Definition 2.6. (*Partition of Unity*) Let $\Omega \subset \mathbb{R}^n$ be covered by a countable family $(U_i)_{i \in \mathbb{N}}$ of open subsets of \mathbb{R}^n , i.e. $\bigcup_{i \in \mathbb{N}} U_i = \Omega$. A smooth partition of unity is a countable family $(h_i)_{i \in \mathbb{N}}$ of smooth functions $h_i : \Omega \rightarrow [0, 1]$ with the following properties:

- (i) Each $x \in \Omega$ has a neighbourhood on which all but finitely h_i vanish identically.
- (ii) For all $x \in \Omega$ we have $\sum_{i=1}^{\infty} h_i(x) = 1$.

(iii) Each h_i vanishes outside a compact subset of U_i .

For every countable family of open subsets of \mathbb{R}^n there exists a smooth partition of unity. A proof can be found in many textbooks and in Prof Schmidt's script of the lecture Analysis II.

Definition 2.7. Let $A \subset \mathbb{R}^n$ be compact k -dimensional submanifold. Because A is compact and $A \subset \bigcup O_x$, only finitely many parametrisations are needed to cover it. Choose a partition of unity $(h_i)_{i \in \mathbb{N}}$ subordinate to O_i . Let f be a continuous function on A . We define

$$\int_A f \, d\sigma = \sum_i \int_{A_i} h_i f \, d\sigma.$$

The idea of this definition is that we can write $f(x) = 1 \times f(x) = \sum_i h_i(x)f(x)$. Then each function $h_i f$ is zero outside of A_i so it is only necessary to integrate it on A_i , not on all of A . We assumed that A was compact so that the sum is finite and we avoid any issues of convergence. The restriction that A is compact is not necessary, but then one must deal with the convergence issues.

Lemma 2.8. The integral $\int_A f \, d\sigma$ neither depends on the choice of the partition of unity nor on the choice of the parametrizations.

Proof. Suppose that we have two covers of parameterising sets $A = \bigcup_i A_i = \bigcup_j B_j$ and correspondingly two partitions of unity h_i and g_j . Define a new cover $C_{i,j} = A_i \cap B_j$. It has a partition of unity $h_i g_j$. Each set $C_{i,j}$ can be parameterised by restricting the parameterisation Φ of A_i to $\Phi^{-1}[A_i \cap B_j]$. Observe

$$\sum_i \int_{A_i} h_i f \, d\sigma = \sum_i \int_{A_i} \left(\sum_j g_j \right) h_i f \, d\sigma = \sum_{i,j} \int_{A_i} g_j h_i f \, d\sigma = \sum_{i,j} \int_{C_{i,j}} g_j h_i f \, d\sigma.$$

The same calculation holds for the integral $\sum_j \int_{B_j} g_j f \, d\sigma$, showing that the two are equal. We have already seen that if we use two different parameterisations for the same set that the integral has the same value. Therefore we have shown that definition is independent of parameterisation and partition of unity. \square

Integrals over submanifolds have many of the same properties as the usual integral. This is because 'under the hood' it is the usual integral with a correction factor. An important property that does not carry over is the change of variables formula. Only certain changes of variables preserve the correction factor. These properties will be proved in the tutorials.

Lemma 2.9. The following properties hold for $a, b \in \mathbb{R}$ and $f, g \in C(A)$.

(i) *Linearity:* $\int_A a f + b g \, d\sigma = a \int_A f \, d\sigma + b \int_A g \, d\sigma$.

- (ii) *Order Preserving:* if $f \leq g$ on A then $\int_A f \, d\sigma \leq \int_A g \, d\sigma$.
- (iii) *Triangle Inequality:* $|\int_A f \, d\sigma| \leq \int_A |f| \, d\sigma$.
- (iv) *Transformation:* If $P : \mathbb{R}^n \rightarrow \mathbb{R}^n$ is a euclidean motion (translation, reflection, rotation) and $s \in \mathbb{R}^+$ is a scaling factor then $\int_A f \, d\sigma = s^k \int_{(sP)^{-1}[A]} f \circ (sP) \, d\sigma$. \square

We are almost ready to state the divergence theorem. In the divergence theorem we work with a bounded and open set $\Omega \subset \mathbb{R}^n$. In general such sets can have very complicated boundaries, for example fractals. We will require $\partial\Omega$ to be a $(n-1)$ -dimensional submanifold. The idea of requiring Ω to be bounded is that $\partial\Omega$ is compact. This idea appears in the proof when we say $\bar{\Omega}$ is compact.

There are three more formulas that we will need. First, just in case you missed the first tutorial, for a vector valued function F , the divergence of F is $\nabla \cdot F = \partial_1 F_1 + \cdots + \partial_n F_n$. The \cdot is meant to remind you of the formula for the dot product; it is not actually a dot product. Second, because a submanifold is locally a graph, it is possible to understand its geometry. In the situation of the theorem $\lambda : U \subset \mathbb{R}^{n-1} \rightarrow \mathbb{R}$ is a scalar valued function. Then $e_i + \partial_i \lambda e_n$ are tangent vectors to the submanifold and so $(-\nabla \lambda, 1)^T$ is perpendicular to the submanifold. Thus the unit length normal vector is

$$N = \pm \frac{1}{\sqrt{1 + |\nabla \lambda|^2}} \begin{pmatrix} -\nabla \lambda \\ 1 \end{pmatrix}.$$

We see that it is a smooth vector field, well-defined up to a choice of sign. The last formula is a simplification of the distortion factor for the graph parameterisation $\Phi(v) = (v, \lambda(v))^T$. We have $\Phi' = (I_{n-1} | \nabla \lambda)^T$ where I_{n-1} is the $(n-1) \times (n-1)$ identity matrix. Since the $(\Phi'(x))^T \Phi'(x)$ is the identity on the orthogonal complement of $\nabla \lambda(x)$ and acts as the multiplication with $|\nabla \lambda(x)|^2$ on $\nabla \lambda(x)^T$ the determinant is $\det(\Phi')^T \Phi' = 1 + |\nabla \lambda|^2$.

Theorem 2.10. (*Divergence Theorem*) Let $\Omega \subseteq \mathbb{R}^n$ be bounded and open with $\partial\Omega$ being a $(n-1)$ -dimensional submanifold of \mathbb{R}^n . Let $F : \bar{\Omega} \rightarrow \mathbb{R}^n$ be continuous and differentiable on Ω such that F' continuously extends to $\partial\Omega$. Then we have

$$\int_{\Omega} \nabla \cdot F \, d\mu = \int_{\partial\Omega} F \cdot N \, d\sigma$$

where N is the outward-pointing normal.

Proof. First we construct a cover for $\bar{\Omega}$. Because the boundary is a $(n-1)$ -dimensional submanifold for every point of $\partial\Omega$ we know that there is an open set O such that $\partial\Omega \cap O = \text{graph}(\lambda)$. By shrinking O , we can assume that O is a cube. Let the cover be Ω and all these cubes. Due to the compactness of $\bar{\Omega}$ we can find a finite subcover and choose a subordinate partition of unity. This decomposes F into a finite sum $\sum h_i F$. By linearity it suffices to show the statement for any $G = h_i F$ individually.

This leads to two cases. For the first case, suppose we have a term $G = h_l F$ corresponding to Ω . In particular G and $G' = (h_l F)'$ are zero on $\partial\Omega$. The right hand side of the divergence theorem is zero. By defining G to be zero outside of Ω we can extend G to a continuously differentiable function on \mathbb{R}^n . Choose a cube $[-R, R]^n$ which contains Ω . By Fubini we may integrate the i -th term of $\nabla \cdot G = \partial_1 G_1 + \dots + \partial_n G_n$ first over dx_i . Due to the fundamental theorem of calculus this integral is the difference of the values of G at two boundary points and vanishes (\widehat{dx}_i means without the factor dx_i):

$$\begin{aligned} \int_{\Omega} \partial_i G_i &= \int_{[-R, R]^n} \partial_i G_i \, dx = \int_{[-R, R]^{n-1}} \left(\int_{-R}^R \partial_i G_i \, dx_i \right) dx_1 \dots \widehat{dx}_i \dots dx_n \\ &= \int_{[-R, R]^{n-1}} \left[G_i \right]_{-R}^R dx_1 \dots \widehat{dx}_i \dots dx_n = \int_{[-R, R]^{n-1}} [0 - 0] dx_1 \dots \widehat{dx}_i \dots dx_n = 0. \end{aligned}$$

This shows that the left side of the divergence theorem also vanishes.

In the second case, we have a term $G = h_l F$ that corresponds to a set that covers the boundary. We assumed the set was a cube, so write it as $U \times (a, b)$. Relabel the coordinates so that the boundary is a graph $x_n = \lambda(x)$ and $\bar{\Omega} \cap (U \times (a, b)) = \{(\frac{u}{y}) \mid y \leq \lambda(u)\}$. We use the variables $u = (u_1, \dots, u_{n-1}) = (x_1, \dots, x_{n-1})$ and $y = x_n$ for convenience. We may assume that G and G' are zero on $\partial U \times (a, b)$ and $U \times \{a\}$, but not that it is zero on $\partial\Omega$. This is because $\partial\Omega$ is ‘inside’ the cube $U \times (a, b)$ and we only know that h_l vanishes on the outside of the cube.

Again, we handle the terms of $\nabla \cdot G = \partial_1 G_1 + \dots + \partial_n G_n$ one at a time. Suppose $1 \leq i < n$ and consider the function

$$u \mapsto \int_a^{\lambda(u)} G_i(u, y) \, dy.$$

It vanishes for $u \in \partial U$ as does its derivative

$$\frac{\partial}{\partial x_i} \int_a^{\lambda(u)} G_i(u, y) \, dy = \frac{\partial \lambda(u)}{\partial x_i} G_i(u, \lambda(u)) + \int_a^{\lambda(u)} \frac{\partial G_i(u, y)}{\partial x_i} \, dy.$$

Applying the same argument as in the first case, we see that the integral of ∂_i -derivative over U vanishes. Therefore

$$\begin{aligned} \int_{\bar{\Omega} \cap (U \times (a, b))} \frac{\partial G_i(u, y)}{\partial x_i} \, d\mu &= \int_U \int_a^{\lambda(u)} \frac{\partial G_i(u, y)}{\partial x_i} \, dy \, d^{n-1}u = - \int_U \frac{\partial \lambda(u)}{\partial x_i} G_i(u, \lambda(u)) \, d^{n-1}u \\ &= \int_U G_i(u, \lambda(u)) N_i \sqrt{1 + |\nabla \lambda|^2} \, d^{n-1}u = \int_{A'} G_i N_i \, d\sigma. \end{aligned}$$

Note that the signs required us to use the outward-pointing normal, which in this case means that the last component of the vector N is positive.

For $i = n$, we can just use the fundamental theorem of calculus on the inner integral

$$\begin{aligned} \int_{\Omega \cap (U \times (a,b))} \frac{\partial G_n(u, y)}{\partial x_n} d\mu &= \int_U \int_a^{\lambda(u)} \frac{\partial G_n(u, y)}{\partial x_n} dy d^{n-1}u = \int_U G_n(u, \lambda(u)) d^{n-1}u \\ &= \int_U G_n(u, \lambda(u)) N_n \sqrt{1 + |\nabla \lambda|^2} d^{n-1}u = \int_{A'} G_n N_n d\sigma \end{aligned}$$

Summing these terms together proves the theorem. \square

We consider now some special cases of the theorem that occur over and over in practice. For a scalar valued function f the divergence theorem implies for all $i = 1, \dots, n$

$$\int_{\Omega} \partial_i f d\mu = \int_{\partial\Omega} f N_i d\sigma$$

For two functions f and g whose product vanishes on the boundary $\partial\Omega$ and satisfies the corresponding assumptions of the divergence theorem we obtain by the Leibniz rule

$$\int_{\Omega} f \partial_i g d\mu = - \int_{\Omega} g \partial_i f d\mu \quad \text{for all } i = 1, \dots, n.$$

This is called integration by parts. Inductively we get for any multi-index γ

$$\int_{\Omega} f \partial^\gamma g d\mu = (-1)^{|\gamma|} \int_{\Omega} g \partial^\gamma f d\mu.$$

As a second application of the divergence theorem we can generalise the idea of the scalar conservation law to vector-valued functions. For any continuously differentiable function $F : \mathbb{R} \rightarrow \mathbb{R}^n$ we call

$$\dot{u}(x, t) + \nabla \cdot F(u(x, t)) = \dot{u}(x, t) + F'(u(x, t)) \cdot \nabla u(x, t) = 0$$

a conservation law. For open and bounded $\Omega \subset \mathbb{R}^n$ with $n - 1$ -dimensional submanifold $\partial\Omega$ of \mathbb{R}^n we obtain

$$\frac{d}{dt} \int_{\Omega} u(x, t) d^n x = \int_{\Omega} \dot{u}(x, t) d^n x = - \int_{\Omega} \nabla \cdot F(u(x, t)) d^n x = - \int_{\partial\Omega} F(u(x, t)) \cdot N(x) d\sigma(x).$$

This is the meaning of a conservation law: the change of the integral of $u(\cdot, t)$ over $\Omega \subset \mathbb{R}^n$ is equal to the integral of the flux $-F(u(\cdot, t)) \cdot N$ through the boundary $\partial\Omega$.

This idea also gives the following cute trick to calculate the surface area of a ball in relation to its volume. Let the volume of the n -dimensional unit ball be ω_n . By scaling, the volume of the ball $B(0, r)$ is $\omega_n r^n$. Let $\sigma_n(r)$ denote the area of $\partial B(0, r) \subset \mathbb{R}^n$. The divergence of $x \mapsto x$ is n , so by the divergence theorem we have

$$n\omega_n r^n = \int_{B(0, r)} \nabla \cdot x d\mu = \int_{\partial B(0, r)} x \cdot N(x) d\sigma(x) = \int_{\partial B(0, r)} x \cdot \frac{x}{|x|} d\sigma(x) = r\sigma_n(r).$$

In summary $\sigma_n(r) = n\omega_n r^{n-1}$.

The way that the divergence theorem relates an integral over a set to an integral over its boundary can be used to decompose the set into layers. The typical example, and the only relevant example for us, is that a ball $B(0, R)$ can be thought of as spheres $\partial B(0, r)$ for $0 < r < R$ (the origin is measure zero and can be ignored). The layers are the level-sets of a function ℓ . For the ball $\ell(x) = |x|$, which has $|\nabla \ell| = 1$ so the formula below simplifies further. There will also be an exercise that proves this formula for the ball directly from the definition of the submanifold integral.

Corollary 2.11 (Co-area Formula). *Let $\ell : \Omega \rightarrow \mathbb{R}$ be a C^1 non-negative function. Suppose for every $t \in [0, T]$ that $\Omega_t := \{x \in \Omega \mid \ell(x) < t\}$ is a domain to which the divergence theorem applies. For any $f \in C^1(\Omega)$,*

$$\int_{\Omega_T} f |\nabla \ell| \, dx = \int_0^T \left[\int_{\partial \Omega_t} f \, d\sigma \right] dt.$$

Proof. Let's make some additional definitions to simplify the working. The gradient of a function is always normal to its level set pointing in the direction of increase, so $N = |\nabla \ell|^{-1} \nabla \ell$ is the outward pointing unit normal of $\partial \Omega_t$. Define $F = fN$, a vector valued function. In particular $F \cdot N = fN \cdot N = f$.

Now we can begin. By the product rule for divergence

$$\begin{aligned} \nabla \cdot ((T - \ell)F) &= \nabla(T - \ell) \cdot F + (T - \ell) \nabla \cdot F \\ &= -\nabla \ell \cdot fN + (T - \ell) \nabla \cdot F \\ &= -f|\nabla \ell| + (T - \ell) \nabla \cdot F. \end{aligned}$$

Rearranging this and applying the divergence theorem shows

$$\begin{aligned} \int_{\Omega_T} f |\nabla \ell| \, dx &= \int_{\Omega_T} (T - \ell) \nabla \cdot F \, dx - \int_{\Omega_T} \nabla \cdot ((T - \ell)F) \, dx \\ &= \int_{\Omega_T} (T - \ell) \nabla \cdot F \, dx - \int_{\partial \Omega_T} (T - \ell) F \cdot N \, dx. \end{aligned}$$

On the boundary of Ω_T the function $\ell = T$, since the set $\Omega_T = \{\ell < T\}$ by definition. Hence $T - \ell = 0$ on $\partial \Omega_T$ and the second term on the right is zero. The next step is a 'magic trick':

$$\int_{\Omega_T} [T - \ell(x)] \nabla \cdot F(x) \, dx = \int_{\Omega_T} \left[\int_{\ell(x)}^T 1 \, dt \right] \nabla \cdot F \, dx = \int_{\Omega_T} \int_{\ell(x)}^T \nabla \cdot F \, dt \, dx.$$

We want to apply Fubini's theorem to change the order of this double integral. But the limits of the inner integral depends on the variable of the outer integral, so first we use

an indicator function to make them independent

$$\begin{aligned} \int_{\Omega_T} \int_{\ell(x)}^T \nabla \cdot F \, dt \, dx &= \int_{\Omega_T} \int_0^T \chi_{\{t \geq \ell(x)\}} \nabla \cdot F \, dt \, dx = \int_0^T \int_{\Omega_T} \chi_{\{t \geq \ell(x)\}} \nabla \cdot F \, dx \, dt \\ &= \int_0^T \int_{\Omega_t} \nabla \cdot F \, dx \, dt. \end{aligned}$$

We apply the divergence theorem one more time to get the result.

$$\int_0^T \int_{\Omega_t} \nabla \cdot F \, dx \, dt = \int_0^T \int_{\partial\Omega_t} F \cdot N \, d\sigma \, dt = \int_0^T \int_{\partial\Omega_t} f \, d\sigma \, dt. \quad \square$$

2.4 Distributions

For the transport equation we developed a solution that also seems to make sense when it is not differentiable. For the scalar conservation law we saw that there were in some situations no solutions, except if we generalised the notion of solution to include discontinuous functions. The lesson we draw from these examples is that the existence and uniqueness of solutions depends on the notion of solution we use. In order to say that these solutions solve the PDE, clearly all partial derivatives of a solution which occur in the partial differential equation have to exist. The trick is to come up with a new notion of partial derivative and interpret the PDE to be about these new derivatives.

In this section we introduce distributions (also called generalised functions) and a corresponding notion of differentiation. This notion is ‘backwards compatible’: if a differentiable function is considered as a distribution, the two types of derivatives are equal. Remarkably distributions can always be differentiated and indeed they can be differentiated infinitely many times. For this achievement we have to pay a price: these distributions cannot be multiplied with each other in general. Linear partial differential equations extend to well defined equations on such distributions. Distributions solving the linear partial differential equations are called *weak solutions* or solutions in the sense of distributions. There exist other notions of weak solutions which also apply to non-linear partial differential equations. The most prominent example is the notion of a Sobolev function, which are introduced in the course “Partial Differential Equations”, the sequel to this course. But Sobolev functions can be understood as a special type of distribution, so even if one is interested in Sobolev functions it is helpful to start with distributions.

First we need to define a special class of very well behaved functions. The support $\text{supp } f$ of a function f is the closure of $\{x \mid f(x) \neq 0\}$. On an open set $\Omega \subseteq \mathbb{R}^n$ let $C_0^\infty(\Omega)$ denote the algebra of smooth functions whose support is a compact subset of Ω . We call these *test functions* and say they have compact support in Ω , symbolically $\text{supp } f \Subset \Omega$.

Within the set of test functions there are a special families that we will often use called a *mollifier* or *approximate identity*. This is a family of non-negative test functions $(\lambda_\epsilon)_{\epsilon>0}$

with $\text{supp } \lambda_\epsilon = \overline{B(0, \epsilon)}$ and $\int \lambda_\epsilon \, d\mu = 1$. We construct a prototype: the function

$$\lambda(x) := \begin{cases} C \exp\left(\frac{1}{|x|^2-1}\right) & \text{for } |x| < 1 \\ 0 & \text{for } |x| \geq 1 \end{cases}$$

is a smooth function on \mathbb{R}^n , has support $\overline{B(0, 1)}$, and is non-negative. By the way, this example shows that test functions actually exist. We can choose the constant C such that its integral is 1. By rescaling x and λ we obtain

$$\lambda_\epsilon(x) = \epsilon^{-n} \lambda(x/\epsilon),$$

which has the required properties. This particular example of a mollifier is called the standard mollifier, but for our purposes it does not matter which mollifier we use. Any such family is called an approximate identity because of the following property. Take any continuous function f on Ω and suppose $0 \in \Omega$. By continuity f is approximately equal to $f(0)$ on a sufficiently small ball $B(0, \epsilon)$. Therefore

$$\int_{\Omega} f \lambda_\epsilon \, d\mu = \int_{B(0, \epsilon)} f \lambda_\epsilon \, d\mu \approx \int_{B(0, \epsilon)} f(0) \lambda_\epsilon \, d\mu = f(0) \int_{B(0, \epsilon)} \lambda_\epsilon \, d\mu = f(0).$$

In fact, as we will prove in the next lemma, this approximation becomes an equality in the limit $\epsilon \downarrow 0$.

Lemma 2.12. *Let $f \in C(\Omega)$ and $(\lambda_\epsilon)_{\epsilon>0}$ be a mollifier. The family of smooth functions*

$$f_\epsilon(x) := \int_{\Omega} f(y) \lambda_\epsilon(x - y) \, d^n y$$

converges uniformly on any compact subset of Ω to f as $\epsilon \downarrow 0$. For smooth functions the same holds for all derivatives of f .

Proof. Choose a compact subset of Ω . There is an ϵ such that for any point x in the compact set the ball $B(x, \epsilon)$ lies in Ω . For this ϵ or smaller we have

$$|f_\epsilon(x) - f(x)| = \left| \int_{\Omega} \lambda_\epsilon(x - y) (f(y) - f(x)) \, d^n y \right| \leq \sup_{y \in B(x, \epsilon)} |f(y) - f(x)|.$$

On compact sets, continuous functions are uniformly continuous. This shows the uniform convergence $\lim_{\epsilon \downarrow 0} f_\epsilon = f$.

Observe that if f is smooth, then we can compute the derivatives of f_ϵ in the following way. Choose any point $x_0 \in \Omega$ and let ϵ be small enough that $B(x_0, 2\epsilon) \subset \Omega$. Then for all points $x \in B(x_0, \epsilon)$

$$f_\epsilon(x) = \int_{B(x, \epsilon)} f(y) \lambda_\epsilon(x - y) \, d^n y = \int_{B(0, \epsilon)} f(x - z) \lambda_\epsilon(z) \, d^n z.$$

Therefore $\partial^\alpha f_\epsilon = (\partial^\alpha f)_\epsilon$ and the same convergence argument carries over to all partial derivatives of f . \square

The formula we see in the definition of f_ϵ turns out to be useful. We use it to define a type of product operator on $C_0^\infty(\mathbb{R}^n)$: the convolution

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

This product is bilinear, commutative, and associative (Exercise). One advantage of the convolution compared to pointwise multiplication is that it behaves nicely with differentiation. There is no Leibniz rule, rather

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

Furthermore convolution is well-behaved with respect to integral norms, which is useful in more advanced theory. We can consider the simplest case, where integral of $f * g$ is the product of the integrals of f and g . This follows by noticing that the coordinate transformation $z = y - x, y = y$ is volume preserving, thus

$$\begin{aligned} \int_{\mathbb{R}^n} (f * g)(x) \, dx &= \int_{\mathbb{R}^n} \int_{\mathbb{R}^n} f(x - y)g(y) \, dx \, dy = \int_{\mathbb{R}^n} \int_{\mathbb{R}^n} f(z)g(y) \, dz \, dy \\ &= \left(\int_{\mathbb{R}^n} f(z) \, dz \right) \left(\int_{\mathbb{R}^n} g(y) \, dy \right) \end{aligned}$$

Finally, we include a lemma that will be necessary later

Lemma 2.13. *Suppose that f and g are rotationally symmetric about a and b respectively. This means, for example for any orthogonal transformation O that $f(a + x) = f(a + Ox)$. Then the convolution of f and g is rotationally symmetric about $a + b$.*

Proof. The proof is just a sequence of coordinate transformations. We begin with the definition and make the euclidean motion $y = Oz + b$

$$(f * g)(a + b + Ox) = \int_{\mathbb{R}^n} f(a + b + Ox - y)g(y) \, dy = \int_{\mathbb{R}^n} f(a + O(x - z))g(b + Oz) \, dz.$$

It is important to see here that $dy = dz$ since O is orthogonal. Now we use the orthogonal properties of f and g to continue

$$= \int_{\mathbb{R}^n} f(a + x - z)g(b + z) \, dz = \int_{\mathbb{R}^n} f(a + x - y' + b)g(y') \, dy' = (f * g)(a + b + x). \quad \square$$

Now it is time to introduce distributions. We have seen in the previous lemma that the operation of integrating a continuous function against a test function somehow retains the information of the function. In this spirit each $f \in L_{\text{loc}}^1(\Omega)$ defines a linear map

$$F_f : C_0^\infty(\Omega) \rightarrow \mathbb{R}, \quad \phi \mapsto \int_{\Omega} f\phi \, d\mu.$$

We will see that the information of f is also retained in this linear form. The idea of distributions is consider not just functions integrated against test functions, but all linear forms F acting on $C_0^\infty(\Omega)$.

There is a technical matter to discuss at this point. The set of test functions should be given a different topology than the norm topology of $C^\infty(\Omega)$ with the supremum norm, but this other topology is tricky and only used in a few places in this course. We use the notation $\mathcal{D}(\Omega)$ for the set of test functions equipped with this topology. Instead of explaining the topology in full detail, let us give the criterion for when a sequence of test functions converges. We define for any compact subset $K \subset \Omega$ and every multi-index α the following seminorm:

$$\|\cdot\|_{K,\alpha} : C_0^\infty(\Omega) \rightarrow \mathbb{R}, \quad \phi \mapsto \|\phi\|_{K,\alpha} := \sup_{x \in K} |\partial^\alpha \phi(x)|.$$

We say that $f_n \rightarrow f$ in $\mathcal{D}(\Omega)$ if there is a compact subset $K \subset \Omega$ such that the supports of every f_n and f are contained in K and that $\|f_n - f\|_{K,\alpha} \rightarrow 0$ for every multi-index α (including $\alpha = 0$). This is a much stronger condition than convergence with respect to $\|\cdot\|_\infty = \|\cdot\|_{\Omega,0}$.

Definition 2.14. *On an open subset $\Omega \subseteq \mathbb{R}^n$ the space of distributions $\mathcal{D}'(\Omega)$ is defined as the vector space space of all linear maps $F : \mathcal{D}(\Omega) \rightarrow \mathbb{R}$ which are continuous with respect to the seminorms $\|\cdot\|_{K,\alpha}$; i.e. for each compact $K \subset \Omega$ there exist finitely many multi indices $\alpha_1, \dots, \alpha_M$ and constants $C_1 > 0, \dots, C_M > 0$ such that the following inequality holds for all test functions $\phi \in \mathcal{D}(\Omega)$ with compact support in K :*

$$|F(\phi)| \leq C_1 \|\phi\|_{K,\alpha_1} + \dots + C_M \|\phi\|_{K,\alpha_M}.$$

The \mathcal{D}' for distributions indicates (for the correctly defined topology) that they are the dual space of \mathcal{D} . Concretely the continuity condition yields the following convergence property for distributions: if $\phi_n \rightarrow \phi$ in $\mathcal{D}(\Omega)$ then the values $F(\phi_n)$ converges to $F(\phi)$. Similarly, a sequence of distribution F_n converges to F if $F_n(\phi) \rightarrow F(\phi)$ for all test functions ϕ .

As previously mentioned, any $f \in L^1_{\text{loc}}(\Omega)$ defines in a canonical way a distribution F_f . Let us verify now that it really meets the definition of distribution. For any compact subset $K \subset \Omega$ and $\phi \in \mathcal{D}(\Omega)$ with support K we have

$$|F_f(\phi)| \leq \int_K |f| |\phi| \, dx \leq \sup_{x \in K} |\phi(x)| \int_K |f| \, dx = \sup_{x \in K} |\phi(x)| \|f\|_{L^1(K)}.$$

Let us give another example of a distribution, one that does not correspond to an element of $L^1_{\text{loc}}(\mathbb{R}^n)$:

$$\delta : \mathcal{D}(\mathbb{R}^n) \rightarrow \mathbb{R} \quad \phi \mapsto \phi(0).$$

Intuitively (and we will prove rigorously soon) any corresponding $f \in L^1_{\text{loc}}(\mathbb{R}^n)$ would vanish on $\mathbb{R}^n \setminus \{0\}$ and would have a total integral one. Since $\{0\}$ has measure zero such a function does not exist. Distributions that come from $L^1_{\text{loc}}(\Omega)$ functions are called regular, and those that don't are non-regular. This distribution is called Dirac's δ -function. We can also show that it is the limit of the sequence of distributions corresponding to the mollifier λ_ϵ .

We now return to the question of whether the distribution F_f retains the information of f . The answer is yes.

Lemma 2.15. (*Fundamental Lemma of the Calculus of Variations*) *If $f \in L^1_{\text{loc}}(\Omega)$ obeys $F_f(\phi) \geq 0$ for all non-negative test functions $\phi \in C_0^\infty(\Omega)$, then f is non-negative almost everywhere. In particular the map $L^1_{\text{loc}}(\Omega) \rightarrow \mathcal{D}'(\Omega)$, $f \mapsto F_f$ is injective.*

Proof. It suffices to prove the local statement for $f \in L^1(\Omega)$. We extend f to \mathbb{R}^n by setting f on $\mathbb{R}^n \setminus \Omega$ equal to zero. The extended function is also denoted by f and belongs to $f \in L^1(\mathbb{R}^n)$. For a mollifier $(\lambda_\epsilon)_{\epsilon>0}$ we have

$$\begin{aligned} \|\lambda_\epsilon * f - f\|_1 &= \int_{\mathbb{R}^n} \left| \int_{B(0,\epsilon)} \lambda_\epsilon(y) f(x-y) \, d^n y - f(x) \right| d^n x \\ &\leq \int_{B(0,\epsilon)} \int_{\mathbb{R}^n} \lambda_\epsilon(y) |f(x-y) - f(x)| \, d^n x \, d^n y \leq \sup_{y \in B(0,\epsilon)} \|f(\cdot - y) - f\|_1. \end{aligned}$$

If f is the characteristic function of a rectangle, then the supremum on the right hand side converges to zero for $\epsilon \downarrow 0$. Due to the triangle inequality the same holds for step functions, i.e. finite linear combinations of such functions. Since step functions are dense in $L^1(\mathbb{R}^n)$ for each $f \in L^1(\mathbb{R}^n)$ this supremum becomes arbitrary small for sufficiently small ϵ . Hence the family of functions $(\lambda_\epsilon * f)_{\epsilon>0}$ converges in $L^1(\mathbb{R}^n)$ in the limit $\epsilon \downarrow 0$ to f .

Moreover, the functions $\lambda_\epsilon * f$ are non-negative. This is because the mollifiers are non-negative and we can write the convolution as the action of F_f on a test function

$$(\lambda_\epsilon * f)(x) = \int_{\mathbb{R}^n} \lambda_\epsilon(x-y) f(y) \, d^n y = F_f(\lambda_\epsilon(x - \cdot)) \geq 0$$

using the assumption on F_f .

So it remains to show that a limit in L^1 of a sequence of non-negative functions is also non-negative. In particular there exists a sequence $(\epsilon_n)_{n \in \mathbb{N}}$ which converges to zero, with $\|f_n - f\|_1 \leq 2^{-n}$ for all $n \in \mathbb{N}$ for $f_n = \lambda_{\epsilon_n} * f$. This ensures that the series $\sum_{n \in \mathbb{N}} |f_n - f|$ converges in $L^1(\mathbb{R}^n)$. So for almost every point x the series $\sum_{n \in \mathbb{N}} |f_n(x) - f(x)|$ is finite, and in particular the tail of the series converges to zero. In other words $\lim_{n \rightarrow \infty} f_n(x) = f(x)$. This indeed shows that f is a.e. non-negative.

In particular, if f belongs to the kernel of $f \mapsto F_f$, then both f and $-f$ are almost everywhere non-negative. So f vanishes almost everywhere. \square

Two definitions for functions carry over naturally to distributions. If $\Omega' \subset \Omega$ then every test function on Ω' extends to a test function on Ω . In this way we can think of any distribution on Ω as a distribution on Ω' , which we call the restriction. For regular distributions, this is really the restriction of functions. Using restriction we can give a definition of support. The complement of the support of a distribution is the union of all sets on which the restriction vanishes. In symbols

$$(\text{supp } F)^c = \bigcup \{ \Omega' \subset \Omega \mid F(\phi) = 0 \quad \forall \phi \in \mathcal{D}(\Omega') \}.$$

The support of the delta distribution is $\{0\}$, and the support of the distribution of a continuous function is its support in the normal sense.

We want to define as many operations on distributions as possible, in such a way that they extend operations on functions. Restriction and support are two examples where this is clear. The general strategy for making such definitions is to compare F_f to F_{Af} where A is the operation. If we can write the relation in a way that only depends on the distribution and not directly on the function, then it is suitable to make a generalised definitions. Let us consider the case of multiplication by a smooth function $g \in C^\infty(\Omega)$. Then for a regular distribution

$$F_{gf}(\phi) = \int_{\Omega} (gf)\phi = \int_{\Omega} f(g\phi) = F_f(g\phi).$$

The product of a distribution with a function $g \in C^\infty(\Omega)$ is defined as

$$gF : \mathcal{D}(\Omega) \rightarrow \mathbb{R}, \quad \phi \mapsto F(g\phi).$$

This product makes the embedding $C^\infty(\Omega) \hookrightarrow \mathcal{D}'(\Omega)$ to a homomorphism of modules over the algebra $C^\infty(\Omega)$. However, the product of a distribution with a non-smooth functions is not defined because then $g\phi$ is not a test function.

So we come to the most important operation on distributions. If f has a derivative, then by integration by parts we obtain

$$F_{\partial_i f} = \int_{\Omega} \partial_i f \phi \, d^n x = - \int_{\Omega} f \partial_i \phi \, d^n x = -F_f(\partial_i \phi).$$

Consequently for any distribution $F \in \mathcal{D}'(\Omega)$ we define the partial derivatives as

$$\partial_i F : \mathcal{D}(\Omega) \rightarrow \mathbb{R}, \quad \phi \mapsto -F(\partial_i \phi).$$

Here we see the advantage of choosing smooth test functions: test functions are always differentiable and so distributions have infinitely many derivatives. These two operations we have just defined, multiplication with a smooth function and partial differentiation, define new distributions. Clearly these new distributions are linear. We should check that they also obey the continuity condition, but we will skip this formality.

We also want to extend convolution to distributions. In order to extend it to a product between a smooth function and a distribution we calculate:

$$\begin{aligned} F_{g*f}(\phi) &= \int_{\mathbb{R}^n} (g * f)\phi \, d^n x = \int_{\mathbb{R}^n} \int_{\mathbb{R}^n} g(x-y)f(y)\phi(x) \, d^n y \, d^n x \\ &= \int_{\mathbb{R}^n} \left(\int_{\mathbb{R}^n} \phi(x)g(x-y) \, d^n x \right) f(y) \, d^n y = F_f(\phi * Pg), \end{aligned}$$

where $(Pg)(z) := g(-z)$ is the point-reflection operator. Therefore we define for $g \in C_0^\infty(\mathbb{R}^n)$ and $F \in \mathcal{D}'(\mathbb{R}^n)$

$$g * F : \mathcal{D}(\mathbb{R}^n) \rightarrow \mathbb{R}, \quad \phi \mapsto F(\phi * Pg).$$

Not only is this a well-defined distribution, the result of convolution is in fact always a regular distribution that corresponds to a smooth function!

Lemma 2.16. *The convolution $g * F$ of a test function $g \in C_0^\infty(\mathbb{R}^n)$ with a distribution $F \in \mathcal{D}'(\mathbb{R}^n)$ belongs to $C^\infty(\mathbb{R}^n)$. It is the function*

$$g * F : \mathbb{R}^n \rightarrow \mathbb{R}, \quad x \mapsto F(T_x Pg)$$

where $(T_x \phi)(y) := \phi(y-x)$ is the translation operator. The support of $g * F$ is contained in the pointwise sum $\text{supp}(g) + \text{supp}(F)$.

Proof. First we show that the function defined in the lemma exists and is smooth. The support of $T_x Pg$ is $\{y \in \mathbb{R}^n \mid x-y \in \text{supp}(g)\} = x - \text{supp}(g)$. Hence for every x the value $F(T_x Pg)$ is well defined for $F \in \mathcal{D}'(\Omega)$. Since continuous functions are uniformly continuous on compact sets, the map $x \mapsto T_x Pg$ is continuous with respect to the seminorms $\|\cdot\|_{K,0}$. Furthermore, the same holds for the seminorms $\|\cdot\|_{K,\alpha}$ since $\frac{T_{x+\epsilon h} - T_x}{\epsilon} g = T_x \frac{T_{\epsilon h} - \mathbf{1}}{\epsilon} g$ converges in the limit $\epsilon \rightarrow 0$ for all $g \in C_0^\infty(\mathbb{R}^n)$ uniformly on \mathbb{R}^n to $T_x (\sum_{i=1}^n -h_i \partial_i g)$. This shows $x \mapsto F(T_x Pg) \in C^\infty(\mathbb{R}^n)$ for $F \in \mathcal{D}'(\mathbb{R}^n)$.

Next we show this smooth function corresponds to the distribution $g * F$ we defined immediately before the lemma. For any $\phi \in \mathcal{D}(\mathbb{R}^n)$ appropriate Riemann sums define a sequence of finite linear combinations of functions in $\{T_x Pg \in C_0^\infty(\mathbb{R}^n) \mid x \in \text{supp}(\phi)\}$, which converges with respect to $\|\cdot\|_{K,\alpha}$ to $\int_{\mathbb{R}^n} T_x Pg \phi(x) \, d^n x$. Hence the linearity and continuity of F gives

$$\int_{\mathbb{R}^n} (g * F)(x)\phi(x) \, d^n x = \int_{\mathbb{R}^n} F(T_x Pg)\phi(x) \, d^n x = F\left(\int_{\mathbb{R}^n} T_x Pg \phi(x) \, d^n x\right) = F(Pg * \phi).$$

Finally, we consider the support. If $F(T_x Pg) \neq 0$, then $g(x-y) \neq 0$ for an element $y \in \text{supp } F$. Hence $x = y + (x-y) \subset \text{supp } F + \text{supp } g$ and $\text{supp}(x \mapsto F(T_x Pg)) \subset \text{supp } F + \text{supp } g$. \square

This Lemma implies that even the convolution of a distribution $F \in \mathcal{D}'(\mathbb{R}^n)$ with a distribution $G \in \mathcal{D}'(\mathbb{R}^n)$ with compact support $\text{supp } G$ is a well defined distribution:

$$F * G : \mathcal{D}(\Omega) \rightarrow \mathbb{R}, \quad \phi \mapsto F(\phi * PG) \text{ with } PG(\phi) := G(P\phi).$$

In particular, we can convolve any distribution with the δ -distribution. Remarkably this returns the same distribution, i.e. $F * \delta = F$ (Exercise). We say that δ is the identity element or neutral element of convolution.

Further details of the theory of distributions can be found in the short and lucid first chapter of the book of Lars Hörmander: “Linear Partial Differential Operators”.