2 PDEs and Kinetic Theory

In the previous section some basic properties of PDEs (and how to solve them as infinite dimensional ODEs) were discussed along with some examples of prototypical evolution equations. In this section we shall discuss additional evolution equations, namely *transport equations*, methods for converting them into infinitely many ODEs, and we'll recall the definitions and properties of Fourier transforms and Sobolev spaces. We start with some motivation from physics.

2.1 Introduction to Kinetic Theory

Kinetic theory is concerned with the statistical description of many-particle systems (gasses, plasmas, galaxies) on a *mesoscopic scale*. [We typically denote the number of particles by N, thought to be large: $N \gg 1$]. This is an intermediate scale, complementing the two well-known scales:

- *Microscopic scale.* This is the naïve Newtonian description where one keeps track of each particle, and the evolution is due to all binary interactions. Already for $N = 3$ this description becomes highly nontrivial (the three body problem), and this is certainly true when considering realistic systems with $N \sim 10^{23}$ particles (Avogadro number).
- *Macroscopic scale.* This a description on the level of observables, taking into consideration conservation laws (mass, momentum, energy). One thus obtains a hydrodynamic description, with equations such as Euler's equations or the Navier-Stokes equations.

In contrast, in the **mesoscopic scale** the system is described by a *probability distribution function* (pdf), so that one does not care about each individual particle, but rather the *statistics* of all particles. More precisely, we introduce a function

$$f = f(t, x, p)$$

which measures the *density* of particles that at time $t \geq 0$ are located at the point $x \in \mathbb{R}^n$ and have momentum $p \in \mathbb{R}^n$ (n is the dimension, and is typically 1, 2 or 3). The pdf f is not an observable, but its moments in the momentum variable are. Let us mention the first two:

$$\rho(t,x) = \int_{\mathbb{R}^n} f(t,x,p)\,\mathrm{d}p = \text{particle density},$$
$$u(t,x) = \frac{1}{\rho(t,x)} \int_{\mathbb{R}^n} f(t,x,p)p\,\mathrm{d}p = \text{mean velocity}.$$

In these lectures we shall always take the spatial domain (i.e. the x variable) to be unbounded, although one could always restrict the domain and impose appropriate boundary conditions. The momentum domain is almost always taken to be unbounded, as there is no *a priori* reason why particle momenta must remain bounded.

2.1.1 The Vlasov equation

Informally speaking, Liouville's equation asserts that the full time derivative of f (also known as the *material derivative*) is nonzero only if there are collisions between particles. That is, as long as collisions are negligible, the pdf f is *transported* and the chain rule gives:

$$0 = \frac{\mathrm{d}f}{\mathrm{d}t} = \frac{\partial f}{\partial t} + \dot{x} \cdot \nabla_x f + \dot{p} \cdot \nabla_p f$$

which is called the *Vlasov equation* and is written (applying Newton's second law $\dot{p} = \mathbf{F}$):

$$\frac{\partial f}{\partial t} + v \cdot \nabla_x f + \mathbf{F}[f] \cdot \nabla_p f = 0 \tag{2.1}$$

where $v = v(p)$ is the velocity and $\mathbf{F}$ is a driving force and depends on the physics of the problem at hand. We use the convention that

$$\nabla_x = (\partial_{x_1}, \dots, \partial_{x_n}) \quad \text{and} \quad \nabla_p = (\partial_{p_1}, \dots, \partial_{p_n}).$$

To simplify presentation, we take all constants that typically appear (such as the mass of particles, or the speed of light which will be relevant in the relativistic case) to be 1. Hence we have

$$\begin{aligned} \text{classical case} \qquad & v = p \\ \text{relativistic case} \qquad & v = \frac{p}{\sqrt{1 + |p|^2}}. \end{aligned}$$

2.1.2 The Vlasov-Poisson system

Consider first the case of a plasma, i.e. a gas of ions or electrons. In the electrostatic case where all interactions are due to the instantaneous electric fields generated by the particles, the force $\mathbf{F} = \mathbf{E}$ is simply the electric field generated by the bulk. The system is hence closed by coupling *Poisson's equation* to the Vlasov equation (2.1):

$$\nabla \cdot \mathbf{E} = \rho. \tag{2.2}$$

Note that the resulting interaction is repulsive (this makes sense: all charged particles have the same charge, and are thus repelling one another). We note that this system can also be modified to be *attractive* by simply flipping the sign of $\mathbf{F}$. This results in a model for *galactic dynamics*. This can be summarised as follows:

$$\mathbf{F} = \gamma \mathbf{E}, \qquad \gamma = \begin{cases} +1 & \text{plasma problems} \\ -1 & \text{galactic dynamics} \end{cases} \tag{2.3}$$

2.1.3 The Vlasov-Maxwell system

In the plasma case, one can consider a more accurate model in which all electro*magnetic* interactions are taken into account, hence coupling the Vlasov equation (2.1) to *Maxwell's equations*

$$\nabla \cdot \mathbf{E} = \rho, \quad \nabla \cdot \mathbf{B} = 0, \quad \nabla \times \mathbf{E} = -\frac{\partial \mathbf{B}}{\partial t}, \quad \nabla \times \mathbf{B} = \mathbf{j} + \frac{\partial \mathbf{E}}{\partial t}, \tag{2.4}$$

through the *Lorentz force*

$$\mathbf{F} = \mathbf{E} + v \times \mathbf{B} \tag{2.5}$$

where $\mathbf{B}(t, x)$ is the magnetic field, and

$$\mathbf{j}(t, x) = \int_{\mathbb{R}^n} f(t, x, p) v \, \mathrm{d}p = \text{current density}.$$

It is well known that Maxwell's equations are hyperbolic (i.e. have the form of a wave equation). This can be seen, for instance, by taking a time derivative. Alternatively, letting ϕ and $\mathbf{A}$ be the electric and magnetic potentials satisfying

$$\nabla \phi = -\mathbf{E} \qquad \text{and} \qquad \nabla \times \mathbf{A} = \mathbf{B}$$

and imposing the Lorenz gauge condition $\nabla\cdot\mathbf{A}+\frac{\partial\phi}{\partial t}=0$ Maxwell's equations are transformed into the system

$$\frac{\partial^2}{\partial t^2}\phi-\Delta\phi=\rho,$$
$$\frac{\partial^2}{\partial t^2}\mathbf{A}-\Delta\mathbf{A}=\mathbf{j}.$$

2.2 Linear Transport Equations: The Method of Characteristics

As we have seen before, Liouville's equation implies that the pdf f is transported by the Vlasov equation (in the case where collisions are negligible). In this section our goal is to study first order (linear) PDEs. Letting the unknown be $u=u(t,y):\mathbb{R}\times\mathbb{R}^n\to\mathbb{R}$, we consider equations of the form

$$\frac{\partial u}{\partial t}(t,y)+w(t,y)\cdot\nabla_y u(t,y)=0 \tag{2.6}$$

where $w:\mathbb{R}\times\mathbb{R}^n\to\mathbb{R}^n$ is given, and ask whether given some initial data

$$u(0,y)=u_0(y) \tag{2.7}$$

solutions exist, and, if so, are they unique?[9] In what follows, this question is converted into a question on infinite dimensional ODEs. Intuitively, instead of thinking of (2.6) as transporting the initial data u_0 (this is also known as the **Eulerian viewpoint**), we write down an ODE for each "particle" (this is also known as the **Lagrangian viewpoint**).

Definition 2.1 (Characteristics). *A* characteristic *for* (2.6) *is a function $X\in C^1(I;\mathbb{R}^n)$ where $I\subset\mathbb{R}$ is an interval (that is, $X:I\to\mathbb{R}^n$ and $I\ni s\mapsto X(s)\in\mathbb{R}^n$ is C^1) satisfying*

$$\dot{X}(s)=w(s,X(s)).$$

Theorem 2.2 (Existence, uniqueness and regularity of characteristics). *Assume that $w(t,y)$ satisfies the two following conditions:*

$$(\mathrm{H1}):\ w\in C([0,T]\times\mathbb{R}^n;\mathbb{R}^n)\ \textit{and}\ D_y w\in C([0,T]\times\mathbb{R}^n;\mathrm{M}_n(\mathbb{R}^n)).$$

$$(\mathrm{H2}):\ \exists c>0\ \textit{such that}\ |w(t,y)|\le c(1+|y|)\ \textit{for all}\ (t,y)\in[0,T]\times\mathbb{R}^n.$$

Then the following holds:

1. *For all $(t,y)\in[0,T]\times\mathbb{R}^n$ there exists a unique characteristic $X(s;t,y)$[10] defined on $[0,T]$ such that $X(t;t,y)=y$. Moreover, $X\in C^1([0,T]\times[0,T]\times\mathbb{R}^n;\mathbb{R}^n)$.*

2. *For all $(s,t,y)\in[0,T]\times[0,T]\times\mathbb{R}^n$ and all $j=1,\dots,n$ the mixed partial derivatives $\partial_{x_j}\partial_s X$ and $\partial_s\partial_{x_j}X$ exist and are equal, and are $C([0,T]\times[0,T]\times\mathbb{R}^n;\mathbb{R}^n)$.*

Remark 2.3. We use the convention that a gradient of a scalar function is denoted ∇ (or ∇_y to make clear that the gradient is with respect to $y=(y_1,\dots,y_n)$) and the matrix of partial derivatives of a vector-valued function is denoted D (or D_y).

[9] Note that this is a *linear* problem, while the Vlasov-Poisson and Vlasov-Maxwell equations are *nonlinear* (the nonlinear term is the forcing term $\mathbf{F}[f]\cdot\nabla_p f$).

[10] X is thought of as a function of s, with t and y being parameters.

Proof. Part (1). The assumption (H1) ensures that there exists a unique local solution to the initial value problem

$$\dot{z}(s) = w(s, z(s)), \qquad z(t) = y, \qquad t \in [0,T], y \in \mathbb{R}^n$$

due to *Picard's theorem*. To show that this solution can be extended to $[0,T]$ write $X(s;t,y) = y + \int_t^s w(\zeta, X(\zeta;t,y))\,\mathrm{d}\zeta$ and show that blowup of $|X(s;t,y)|$ cannot occur in $[0,T]$ due to (H2) and *Gronwall's lemma*.

Part (2). This follows from the continuous dependence of solutions to ODEs on parameters and initial data. $\square$

Theorem 2.4 (Properties of the characteristics). *Assume that $w(t,y)$ satisfies (H1) and (H2) and let X be as defined in Definition 2.1. Then:*

1. *For all $t_1, t_2, t_3 \in [0,T]$ and $y \in \mathbb{R}^n$, $X(t_1; t_2, X(t_2; t_3, y)) = X(t_1; t_3, y)$.*
2. *For $(s,t) \in [0,T] \times [0,T]$, $\mathbb{R}^n \ni y \mapsto X(s;t,y)$ is a C^1-diffeomorphism on $\mathbb{R}^n$ (denoted $X(s;t,\cdot)$) with inverse $X(t;s,\cdot)$. Furthermore, $X(s;t,)$ is orientation preserving.*
3. *Define $J(s;t,y) := \det(D_y X(s;t,y))$. Then J solves*
$$\begin{cases} \dot{J}(s;t,y) = \nabla_y \cdot w(s, X(s;t,y)) J(s;t,y), \quad y \in \mathbb{R}^n,\ s,t \in [0,T], \\ J(t;t,y) = 1. \end{cases}$$
4. *If $\nabla_y \cdot w(t,y) = 0$ for all $y \in \mathbb{R}^n, t \in [0,T]$ then $X(s;t,\cdot)$ leaves the Lebesgue measure on $\mathbb{R}^n$ invariant, i.e.*
$$\int_{\mathbb{R}^n} \psi(X(s;t,y))\,\mathrm{d}y = \int_{\mathbb{R}^n} \psi(y)\,\mathrm{d}y, \qquad \forall s,t \in [0,T],\ \forall \psi \in C_0(\mathbb{R}^n). \tag{2.8}$$

Proof. Part (1). This is a simple consequence of the uniqueness of solutions.

Part (2). Due to the smooth dependence on initial data $X \in C^1([0,T] \times [0,T] \times \mathbb{R}^n)$. In particular $y \mapsto X(s;t,y)$ and $y \mapsto X(t;s,y)$ are C^1. Due to Part (1), $X(t;s,\cdot) \circ X(s;t,\cdot) = X(s;t,\cdot) \circ X(t;s,\cdot) = \mathrm{Id}_{\mathbb{R}^n}$ hence $X(s;t,\cdot)$ is one-to-one and onto with inverse $X(t;s,\cdot)$. The fact that they are both orientation preserving relies on the next part (in particular, it relies on the fact that $J(s;t,y) > 0$ for all $s,t \in [0,T], y \in \mathbb{R}^n$).

Part (3). Recall *Jacobi's formula*:

$$\frac{\mathrm{d}}{\mathrm{d}t}(\det A(t)) = \det(A)\,\mathrm{trace}\left(A^{-1}(t)\frac{\mathrm{d}}{\mathrm{d}t}(A(t))\right).$$

which implies that

$$\dot{J}(s;t,y) = J(s;t,y)\,\mathrm{trace}(D_y X(s;t,y)^{-1} \partial_s D_y X(s;t,y)).$$

In addition, due to Theorem 2.2 we know that $\partial_s D_y X = D_y \partial_s X \in C([0,T] \times [0,T] \times \mathbb{R}^n; \mathrm{M}_n(\mathbb{R}^n))$, hence we have

$$\begin{aligned}
\dot{J}(s;t,y) &= J(s;t,y)\,\mathrm{trace}(D_y X(s;t,y)^{-1} D_y \partial_s X(s;t,y)) \\
&= J(s;t,y)\,\mathrm{trace}(D_y X(s;t,y)^{-1} D_y(w(s, X(s;t,y)))) \\
&= J(s;t,y)\,\mathrm{trace}(D_y X(s;t,y)^{-1} (D_y w)(s, X(s;t,y)) D_y X(s;t,y)) \\
&= J(s;t,y)\,\mathrm{trace}((D_y w)(s, X(s;t,y))) \\
&= J(s;t,y) \nabla_y \cdot w(s, X(s;t,y)).
\end{aligned}$$

Since $X(t;t,y) = y$ we have $D_y X(t;t,y) = I_{\mathbb{R}^n}$ so that $J(t;t,y) = 1$.

Part (4). From the previous part we know that $J(s;t,y) = \exp\left(\int_t^s \nabla_y \cdot w(\tau, X(\tau;t,y))\mathrm{d}\tau\right) > 0$. If $\nabla_y \cdot w \equiv 0$ then $J \equiv 1$. ☐

Theorem 2.5 (Solution to the transport equation). *Let $w(t,y)$ satisfy (H1) and (H2) and let $u_0 \in C^1(\mathbb{R}^n)$. Then the Cauchy problem*

$$\begin{cases} \partial_t u(t,y) + w(t,y) \cdot \nabla_y u(t,y) = 0, & y \in \mathbb{R}^n, t \in (0,T) \\ u(0,y) = u_0(y) \end{cases}$$

has a unique solution $u \in C^1([0,T] \times \mathbb{R}^n)$ which is given by the formula

$$u(t,y) = u_0(X(0;t,y)), \qquad \forall t \in [0,T], y \in \mathbb{R}^n. \tag{2.9}$$

Finally, if u_0 is compactly supported then so is $u(t,\cdot)$.

Proof. **Uniqueness.** Uniqueness follows from the uniqueness of characteristics (the PDE is reduced to an ODE along the characteristic), as follows. Since the map $t \mapsto X(t;0,z)$ is C^1 and $u \in C^1$ (we show this in the next part), so is the map $t \mapsto u(t, X(t;0,z))$. Applying the chain rule one easily sees that

$$\frac{\mathrm{d}}{\mathrm{d}t} u(t, X(t;0,z)) = 0.$$

Hence $t \mapsto u(t, X(t;0,z))$ is constant on $[0,T]$ and $u(t,X(t;0,z)) = u(0, X(0;0,z)) = u_0(z)$. Letting $y = X(t;0,z)$ we have $z = X(0;t,y)$ so that

$$u(t,y) = u_0(X(0;t,y)), \qquad \forall (t,y) \in [0,T] \times \mathbb{R}^n.$$

Existence. a) The formula (2.9) defines a C^1 function. This is clear due to Theorem 2.2, since u is defined as the composition of the C^1 maps u_0 and $(t,y) \mapsto X(0;t,y)$. This function obviously satisfies $u|_{t=0} = u_0$.

b) We need to verify that (2.9) defines a solution to the transport equation. That is, we need to show that

$$(\partial_t + w(t,y) \cdot \nabla_y)\, u_0(X(0;t,y)) = 0.$$

By the chain rule, the left hand side is

$$(\partial_t + w(t,y) \cdot \nabla_y)\, u_0(X(0;t,y)) = \nabla u_0(X(0;t,y)) \cdot (\partial_t + w(t,y) \cdot \nabla_y)\, X(0;t,y)$$

and we claim that

$$(\partial_t + w(t,y) \cdot \nabla_y)\, X(0;t,y) = 0, \qquad \forall t \in [0,T], y \in \mathbb{R}^n.$$

In fact, let us show that the following stronger identity holds:

$$(\partial_t + w(t,y) \cdot \nabla_y)\, X(s;t,y) = 0, \qquad \forall s,t \in [0,T], y \in \mathbb{R}^n.$$

Recall the identity $X(t_1; t_2, X(t_2; t_3, y)) = X(t_1; t_3, y)$ from Theorem 2.4. Differentiating with respect to t_2 one obtains[11]

$$\begin{aligned} &\partial_t X(t_1; t_2, X(t_2;t_3,y)) + D_y X(t_1;t_2, X(t_2;t_3,y)) \cdot \dot{X}(t_2;t_3,y) \\ &\quad = \partial_t X(t_1;t_2,X(t_2;t_3,y)) + D_y X(t_1;t_2,X(t_2;t_3,y)) \cdot w(t_2, X(t_2;t_1,y)) = 0. \end{aligned}$$

[11] We denote $\dot{X}$ the partial derivative of X with respect to the first variable, and $\partial_t X$ the partial derivative of X with respect to the parameter t (the second "variable"), and $D_y X$ the matrix of partial derivatives with respect to the second parameter (third "variable").

Letting $t_2 = t_3 = t$ and $t_1 = s$ this identity becomes

$$\begin{aligned}
&\partial_t X(s;t,X(t;t,y)) + D_y X(s;t,X(t;t,y)) \cdot \dot{X}(s;t,y) \\
&\quad = \partial_t X(s;t,y) + D_y X(s;t,y) \cdot \dot{X}(s;t,y) = 0
\end{aligned}$$

which proves the claim.

Compact support. This is a simple consequence of the formula

$$\operatorname{supp}(u(t,\cdot)) = \operatorname{supp}(u_0(X(0;t,\cdot))) = X(0;t,\operatorname{supp}(u_0))$$

and the fact that characteristics remain bounded in $[0,T]$. □

2.3 The Fourier Transform

The Fourier transform is an essential tool when studying PDEs for many reasons. Notably, it relates *differentiation* to *multiplication*, which is much easier to analyse in many cases. There are many ways to define the Fourier transform (they usually differ from one another by where one inserts the factor of 2π). We choose a definition that renders the transform an *isometry.*

Definition 2.6. *Given $u \in L^1(\mathbb{R}^n)$ we define its Fourier transform $\mathcal{F}u$ to be*

$$(\mathcal{F}u)(\xi) = \hat{u}(\xi) := \frac{1}{(2\pi)^{n/2}} \int_{\mathbb{R}^n} e^{-ix\cdot\xi} u(x)\,\mathrm{d}x, \quad \xi \in \mathbb{R}^n. \tag{2.10}$$

Given $v \in L^1(\mathbb{R}^n)$ we define the inverse Fourier transform[12] *to be*

$$(\mathcal{F}^{-1}v)(x) = \check{v}(x) := \frac{1}{(2\pi)^{n/2}} \int_{\mathbb{R}^n} e^{ix\cdot\xi} v(\xi)\,\mathrm{d}\xi, \quad x \in \mathbb{R}^n. \tag{2.11}$$

Lemma 2.7. *The Fourier transform and its inverse embed $L^1(\mathbb{R}^n)$ into $L^\infty(\mathbb{R}^n)$.*

Theorem 2.8 (Riemann-Lebesgue Lemma). *If $u \in L^1(\mathbb{R}^n)$ then $\hat{u}$ is continuous and $|\hat{u}(\xi)| \to 0$ as $|\xi| \to \infty$.*

Theorem 2.9 (Plancherel). *The Fourier transform can be defined as an operator on $L^2(\mathbb{R}^n)$, and $\mathcal{F} : L^2(\mathbb{R}^n) \to L^2(\mathbb{R}^n)$ is a unitary isomorphism.*

Proposition 2.10 (Properties of the Fourier transform). *Some of the important properties of $\mathcal{F}$ are:*

1. $\mathcal{F}[u(\cdot - x_0)] = e^{-ix_0\cdot\xi}\mathcal{F}u$
2. $\mathcal{F}[u(\cdot/\lambda)] = |\lambda|^n (\mathcal{F}u)(\lambda\cdot)$
3. $\mathcal{F}[u * v] = \mathcal{F}u\mathcal{F}v$
4. $\mathcal{F}[\partial_{x_j} u] = -i\xi_j \mathcal{F}u$

[12] Strictly speaking, one cannot *define* the inverse transform; rather, one has to *show* that the inverse of $\mathcal{F}$ is indeed given by (2.11). However, this requires a detailed discussion of domains and ranges of the transform, a topic which we do not attempt to cover here.

2.4 Sobolev Spaces

Sobolev spaces are an essential tool in the analysis of PDEs. It is well known that working with L^2 spaces is highly beneficial due to their Hilbert space structure. Sobolev spaces are meant to adapt this to the study of PDEs: these are functional spaces (with a Hilbert space structure) that include functions that are L^2, as well as some of their derivatives.

Definition 2.11 (The Sobolev spaces $H^k(\mathbb{R}^n)$). *Let $u \in L^2(\mathbb{R}^n)$. We say that $u \in H^k(\mathbb{R}^n)$ (where $k \in \mathbb{N}$), if $\partial^\alpha u \in L^2(\mathbb{R}^n)$,*[13] *for any $|\alpha| \le k$.*[14] *The norm on $H^k(\mathbb{R}^n)$ is given by*

$$\|u\|^2_{H^k(\mathbb{R}^n)} := \sum_{|\alpha|\le k} \|\partial^\alpha u\|^2_{L^2(\mathbb{R}^n)}.$$

When there is no room for confusion, we may replace $\|\cdot\|_{H^k(\mathbb{R}^n)}$ by $\|\cdot\|_{H^k}$ or by $\|\cdot\|_k$.

In other words, $H^k(\mathbb{R}^n)$ is the space of all functions $u \in L^2(\mathbb{R}^n)$ such that all possible partial derivatives of u (including mixed partial derivatives), up to (and including) order k, are square integrable.

In Proposition 2.10 we saw how the Fourier transform relates differentiation to multiplication: ∂_{x_j} becomes $-i\xi_j$. This provides a tool for defining the Sobolev spaces H^k in a more efficient way. Moreover, this will allow us to replace the discrete parameter k by a *continuous* parameter, typically denoted s, and hence providing us with a far richer family of spaces that include functions that have “fractional derivatives” that are square integrable.

Theorem 2.12 (Definition of $H^k(\mathbb{R}^n)$ in terms of the Fourier transform). *$u \in H^k(\mathbb{R}^n)$ if and only if $(1+|\xi|^2)^{k/2}\hat{u} \in L^2(\mathbb{R}^n)$ and the norm $\|u\|_k^2$ defined above is equivalent to the norm*

$$\int_{\mathbb{R}^n} (1+|\xi|^2)^k |\hat{u}(\xi)|^2 \, \mathrm{d}\xi.$$

In this definition there is no apparent reason why k must be discrete, and indeed we may extend this definition:

Definition 2.13 (Sobolev spaces $H^s(\mathbb{R}^n)$ with continuous parameter). *We define the Sobolev space $H^s(\mathbb{R}^n)$, where $s \ge 0$,*[15] *to be the space of functions $u \in L^2(\mathbb{R}^n)$ such that $(1+|\xi|^2)^{s/2}\hat{u} \in L^2(\mathbb{R}^n)$. The associated norm $\|u\|_s^2$ is defined as*

$$\|u\|_s^2 := \int_{\mathbb{R}^n} (1+|\xi|^2)^s |\hat{u}(\xi)|^2 \, \mathrm{d}\xi.$$

Theorem 2.14 (Hilbert space structure). *$H^s(\mathbb{R}^d)$ is a Hilbert space with inner product*

$$(u,v)_s := \int_{\mathbb{R}^n} (1+|\xi|^2)^s \hat{u}(\xi)\overline{\hat{v}(\xi)} \, \mathrm{d}\xi.$$

Remark 2.15 (Inequalities and embeddings). Later in this course (as the need arises) we shall get a glimpse into a vast field of inequalities and embeddings, relating different

[13] Partial derivatives of u are defined *in the sense of distributions* or *weakly*, meaning that they are defined by integrating by parts against test functions.

[14] α is a multi-index $\alpha = (\alpha_1, \dots, \alpha_n)$ with $\alpha_j \ge 0$, $|\alpha| = \alpha_1 + \cdots + \alpha_n$, and $\partial^\alpha = \partial_{x_1}^{\alpha_1} \cdots \partial_{x_n}^{\alpha_n}$

[15] We can consider also values $s < 0$, however in that case u is not an element of L^2, rather u must be taken to be a *tempered distribution*.

spaces (Sobolev spaces, L^p spaces, Hölder spaces ...) to one another. For instance, it is often desirable to know whether a function belonging to H^s has certain continuity or boundedness properties. Furthermore, if we knew that H^s were *compactly* embedded in some C^k space, say, then we could conclude that a bounded sequence of elements in H^k has a convergent subsequence in C^k.