Topics in Mathematical Physics

Chiara Saffirio

November 26, 2025

These notes summarize the class "Topics in Mathematical Physics", taking place at the University of Freiburg in the Fall Term of 2025. They are based on:

- E. H. Lieb, M. Loss. *Analysis 2*. Second Edition. Graduate Studies in Mathematics. Volume 14, American Mathematical Society.
- B. C. Hall. *Quantum Theory for Mathematicians*. Graduate Texts in Mathematics. Volume 267, Springer New York, 2013.
- E. H. Lieb, R. Seiringer. *The Stability of Matter in Quantum Mechanics*. Cambridge University Press, 2010.

Contents

1 Classical Mechanics

1.1 One-body Classical Mechanics

Newton's Equation. Consider a point particle moving on $\mathbb{R}^d$ at position $x \in \mathbb{R}^d$ with momentum $p \in \mathbb{R}^d$, where $p = mv$, with m the mass and $v = \dot{x}$ the velocity of the particle. The set $\{(x,p) \,|\, (x,p) \in \mathbb{R}^{2d}\}$ of all possible configurations of vectors (x,p) is referred to as the *phase space* of the system. The time evolution of the system is given by the map $t \mapsto (x(t), v(t))$, for all times $t \in \mathbb{R}$. Such map is defined as the solution to the Newton's equation

$$m\ddot{x}(t) = F(x(t)),$$

where $F : \mathbb{R}^d \to \mathbb{R}^d$ is the force field at point x. If $F = -\nabla K$, with K a real-valued function called potential, then Newton's equation can be written as

$$\begin{cases} \dot{x}(t) = & p(t)/m, \\ \dot{p}(t) = & -\nabla K(x(t)). \end{cases} \tag{1.1}$$

Observe that, under assumptions on the vector field $(p, -\nabla K) \in \mathbb{R}^d \times \mathbb{R}^d$ (e.g. $(p, -\nabla K) \in \mathrm{Lip}(\mathbb{R}^d)$) there exists a unique solution $(x,p) \in \mathbb{R}^d$ to (1.1) (cfr. standard theory of o.d.e.). In other words, given an initial configuration $(x,p) \in \mathbb{R}^{2d}$ at time $t = 0$, the configuration at time $t > 0$ is uniquely determined by the solution $(x(t), p(t))$ to (1.1), that means that classical mechanics is deterministic.

Hamiltonian systems. An important quantity that describes the point particle in the phase space is the Hamiltonian function $\mathcal{H} : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$

$$\mathcal{H}(x,p) = \frac{p^2}{2m} + K(x), \tag{1.2}$$

whose value represents the *energy* of the system.

Proposition 1.1.1 (Hamiltonian system)**.** *Given the differentiable Hamiltonian function* (1.2)*, Newton's equation* (1.1) *reads*

$$\begin{cases} \dot{x}(t) = & \frac{\partial \mathcal{H}}{\partial p}(x(t), p(t)), \\ \dot{p}(t) = & -\frac{\partial \mathcal{H}}{\partial x}(x(t), p(t)). \end{cases}$$

Proof. The statement follows by direct inspection. □

Corollary 1.1.2. *The energy is conserved by the time evolution.*

Proof. Proposition 1.1.1 yields

$$\frac{d}{dt}\mathcal{H}(x(t), p(t)) = \frac{\partial \mathcal{H}}{\partial x} \cdot \dot{x}(t) + \frac{\partial \mathcal{H}}{\partial p} \cdot \dot{p}(t) = \frac{\partial \mathcal{H}}{\partial x} \cdot \frac{\partial \mathcal{H}}{\partial p}(x(t), p(t)) - \frac{\partial \mathcal{H}}{\partial p} \cdot \frac{\partial \mathcal{H}}{\partial x}(x(t), p(t)) = 0.$$

Hence the Hamiltonian is preserved by the time-evolution. □

Definition 1.1.3. *A (possibly vector-valued) function A defined on the phase space $\mathbb{R}^{2d}$ is called observable.*

Definition 1.1.4. *Let A and B be two observables. The Poisson brackets $\{A, B\}$ of A and B are given by*

$$\{A, B\} := \frac{\partial A}{\partial x}\frac{\partial B}{\partial p} - \frac{\partial A}{\partial p}\frac{\partial B}{\partial x}.$$

Proposition 1.1.5. *Let A be an observable of the system with Hamiltonian $\mathcal{H}$. The time-evolution of A is given by the Poisson brackets of A and $\mathcal{H}$*

$$\frac{d}{dt}A(x(t), p(t)) = \{A, \mathcal{H}\}(x(t), p(t)).$$

Proof. The statement follows by direct inspection. □

Remark 1.1.6 (Conservation laws)**.** *Every observable A such that $\{A, \mathcal{H}\} = 0$ is preserved along the time-evolution. In particular, the conservation of energy follows easily from $\{\mathcal{H}, \mathcal{H}\} = 0$. In the specific case of a free particle (no force acting on the particle) $\mathcal{H}(x, p) = p^2/2m$ and therefore $\{p, \mathcal{H}\} = 0$, which means that momentum is preserved. The importance of conservation laws is due to the fact that they are often the main tool to solve Newton's equation.*

1.2 Many-body Classical Mechanics

Most physical systems are made of a big number of particles $N \gg 1$. Let N be the number of particles moving in $\mathbb{R}^d$. A configuration of the system is given by the collection of positions of the N particles $x = (x_1, \dots, x_N) \in \mathbb{R}^{dN}$ and velocities $v = (v_1, \dots, v_N) \in \mathbb{R}^{dN}$. The time evolution of the configuration $(x, v) \in \mathbb{R}^{2dN}$ is given by the system of Newton's equations

$$m_i \ddot{x}_i(t) = F_i(x(t)), \quad i = 1, \dots N,$$

where m_i is the mass of i-th particle, and $F_i : \mathbb{R}^{dN} \to \mathbb{R}^d$ is the force acting on particle i, which a priori depends on the whole configuration $x \in \mathbb{R}^{dN}$.

Coulomb potential. One special interaction in nature is the electrostatic interaction, or Coulomb potential, describing the interaction between two charged particles. To better understand what happens in many-body Coulomb systems, consider $N = 2$ and $d = 3$. The Coulomb potential then writes $K(x_1, x_2) = q_1 q_2/|x_1 - x_2|$, where q_i is the charge of particle i, $i = 1, 2$. The corresponding force field $F = (F_1, F_2) \in \mathbb{R}^d \times \mathbb{R}^d$, also called Coulomb force, is given by

$$F_i(x_1, x_2) = -\nabla_{x_i} K(x_1, x_2), \quad K(x_1, x_2) = \frac{q_1 q_2}{|x_1 - x_2|}, \quad i = 1, 2.$$

It describes the electrostatic attraction (if $q_1 q_2 < 0$) or repulsion (if $q_1 q_2 > 0$) between two charged particles with charges q_1 and q_2.
The Hamiltonian associated with such system is

$$\mathcal{H}(x, p) = \sum_{i=1}^{2} \frac{p_i^2}{2m} + \frac{q_1 q_2}{|x_1 - x_2|}.$$

Observe that there are two possible scenarios:

i) either $q_1 q_2 > 0$ (i.e. the two particles have charges of the same sign), then the total energy of the system (that is the Hamiltonian) is minimized by taking the particles as much as possible separated. This implies that the electrostatic repulsion prevents the system from forming bound states and the two particles flow away;

ii) or $q_1 q_2 < 0$ (i.e. the two particles have opposite signs), then the total energy of the system is minimized when the two particles are close to each other, in such case the Hamiltonian in unbounded below and the system is unstable.

Both situations fail to explain stability of atoms, that are made of neutrons (neutral particles), protons (positively charged particles) and electrons (negatively charged particles). As we shall see, stability of matter finds its rigorous justification in the theory of quantum mechanics.

2 Introduction to Quantum Mechanics

2.1 Basic Notions

Quantum mechanics is the microscopic theory of nature that describes physics on the scale of atoms and molecules. In analogy with classical mechanics, we will define evolution and observables in quantum mechanics. To this end, we will introduce basic functional analysis notions, which are collected in Appendix A.

Observables. Quantum systems are described on Hilbert spaces $\mathcal{H}$ (see Definition A.1.2), that are the analogue of classical mechanic phase spaces. Following the comparison with classical mechanics, states of the system are therefore elements of the Hilbert space, i.e. they are represented by vectors $\psi \in \mathcal{H}$, such that $\|\psi\| = 1$, while observables are associated with self-adjoint operators on $\mathcal{H}$. Here $\|\cdot\|$ denotes the norm induced by the inner product on $\mathcal{H}$. We shall see that $|\psi|^2$ can be interpreted as a probability, making quantum mechanics a non-deterministic theory. It therefore makes sense to consider the expectation of an observable described by the operator A in the state ψ by defining the inner product $\langle \psi, A\psi \rangle$.
In analogy with the classical case, consider a system consisting of a particle moving in $\mathbb{R}^d$, which can be described on the Hilbert space $\mathcal{H} = L^2(\mathbb{R}^d; \mathbb{C})$ of complex-valued, square integrable functions on $\mathbb{R}^d$. Physically, $|\psi(x)|^2$ is interpreted as the probability density for finding the particle at position x. Accordingly, for a measurable set $B \subset \mathbb{R}^d$, the probability that the particle is located in a region B is given by

$$\int_B |\psi(x)|^2 dx = \int \mathbf{1}_B(x) |\psi(x)|^2 dx = \langle \psi, \mathbf{1}_B(x)\psi \rangle$$

where $\mathbf{1}_B$ is the characteristic function of B. Similarly, the expectation value of the position of the particle is determined by

$$\int x |\psi(x)|^2 dx = \langle \psi \, , \, x\, \psi \rangle$$

The function $\psi \in L^2(\mathbb{R}^d)$ is often referred to as the wave function. It does not only determine the distribution of the position of the particle, but also every observable physical quantity is associated with a self-adjoint operator A acting on the Hilbert space $L^2(\mathbb{R}^d)$ so that the expectation value of the

observable is given by the inner product $\langle \psi, A\psi \rangle$ and the probability that a measurement of A gives a result in a set $B \subset \mathbb{R}^d$ is determined by $\langle \psi, \mathbf{1}_B(A)\psi \rangle$, where $\mathbf{1}_B(A)$ is defined through functional calculus.

The self-adjoint operator associated with the observable *position of the particle* is the multiplication operator $\psi(x) \mapsto x\psi(x)$. More precisely, for each component $i = 1, \dots, d$ of the vector $x \in \mathbb{R}^d$, we define the map $\psi(x) \mapsto x_i\psi(x)$ and define the vector made of the d components as the action of the position operator. Notice that the position operator, defined as multiplication through x, is an unbounded self-adjoint operator on $L^2(\mathbb{R}^d)$. For this reason, it is not defined on the full Hilbert space $L^2(\mathbb{R}^d)$ but only on the dense subspace consisting of functions $\psi \in L^2(\mathbb{R}^d)$ such that $x\psi(x)$ is again square integrable.

The observable *momentum of the particle* is associated with the self-adjoint operator $p = -i\hbar\nabla$. Notice that also p is a vector-valued observable with components $p_j = -i\hbar\partial_{x_j}$, $j = 1, \dots, d$ and it is an unbounded self-adjoint operator. Here $\hbar$ is the Planck constant, which we will henceforth assume to be $\hbar = 1$. The expected value of the momentum is therefore given by the expectation

$$\langle \psi, -i\nabla\psi \rangle = -i \int \overline{\psi(x)} \nabla\psi(x) dx$$

Since the Fourier transform diagonalizes the derivative, we can write

$$\langle \psi, -i\nabla\psi \rangle = \int k |\hat{\psi}(k)|^2 dk$$

In other words, $|\hat{\psi}(k)|^2$ is the probability density for finding the momentum close to k. This observation explains why it is important for ψ to be complex valued: while $|\psi(x)|^2$ determines the distribution of the position observable, it is the phase of ψ, the oscillations of ψ, which contributes to $|\hat{\psi}(k)|^2$ and determines the distribution of the momentum observable.

We moreover notice that the position operator is a multiplication operator and its distribution is determined directly by the probability density $|\psi(x)|^2$. The momentum operator is a differential operator, it is diagonal in Fourier space and its distribution is determined by the probability density $|\hat{\psi}(k)|^2$.

For a general self-adjoint operator A on $L^2(\mathbb{R}^d)$, its distribution is determined by the spectral decomposition of A. Suppose $\mathcal{H}$ to be finite dimensional. Then we know from Linear Algebra that the spectral theorem holds for self adjoint operators (i.e. symmetric matrices), that is for A self adjoint operator on the finite dimensional space $\mathcal{H}$,

$$A = \sum_j \lambda_j P_{\psi_j}$$

where $\lambda_j \in \mathbb{R}$ are the eigenvalues of A and P_{ψ_j} are orthogonal projections onto the eigenvector ψ_j. i.e.

$$P_{\psi_j}\psi = \langle \psi_j, \psi \rangle \psi_j \,.$$

Thus

$$\langle \psi, A\psi \rangle = \sum_j \lambda_j |\langle \psi, \psi_j \rangle|^2.$$

Moreover, by completeness of the orthonormal basis we have $\sum_j |\langle \psi, \psi_j \rangle|^2 = 1$.
Now, if $\mathcal{H}$ is infinite dimensional and A is self adjoint with discrete spectrum, an extension of the spectral theorem holds leading to $A = \sum_j \lambda_j P_{\psi_j}$, $\lambda_j \in \mathbb{R}$ eigenvalues and P_{ψ_j} orthogonal projections onto the eigenvectors ψ_j of A. Then, for $\psi \in \mathcal{H}$, $B \subset \mathbb{R}$, we have

$$\langle \psi, A\psi \rangle = \sum_j \lambda_j |\langle \psi, \psi_j \rangle|^2$$
$$\langle \psi, \mathbf{1}_B \psi \rangle = \sum_j \mathbf{1}_B(\lambda_j) |\langle \psi, \psi_j \rangle|^2 = \sum_{j \,:\, \lambda_j \in B} |\langle \psi, \psi_j \rangle|^2$$

This means that the observable A takes the values λ_j with probability $|\langle \psi, \psi_j \rangle|^2$. But what is the spectrum of A is not be purely discrete? If the spectrum of the self-adjoint operator A has a continuous component, let $(E_\lambda)_{\lambda \in \mathbb{R}}$ be the projection-valued measure associated with A. Then the generalization of the spectral theorem allows us to write

$$A = \int \lambda dE_\lambda$$

and we have

$$\langle \psi, A\psi \rangle = \int \lambda d\langle \psi, E_\lambda \psi \rangle$$
$$\langle \psi, \mathbf{1}_B(A)\psi \rangle = \int \mathbf{1}_B(\lambda) d\langle \psi, E_\lambda \psi \rangle = \int_B d\langle \psi, E_\lambda \psi \rangle$$

where $d\langle \psi, E_\lambda \psi \rangle$ is a Borel measure.

Heisenberg's uncertainty principle. Consider a quantum system described on a Hilbert space $\mathcal{H}$. States of the systems are associated to vectors $\psi \in \mathcal{H}$, normalized so that $\|\psi\| = 1$. Since observable quantities are given by expectations of the form $\langle \psi, A\psi \rangle$, for self-adjoint operators A, the vector ψ is only defined up to a phase (ie. ψ and $e^{i\theta}\psi$ describe the same state, if $\theta \in \mathbb{R}$ is a constant).
Taking into account that ψ determines simultaneously the distribution of all observables of the system, it follows that these distributions cannot be independent of each other. An important consequence of this observation is Heisenberg's uncertainty principle.

Theorem 2.1.1. *Let A, B be two self-adjoint operators acting on a Hilbert space $\mathcal{H}$ and let $\psi \in \mathcal{H}$ be a normalized (so that $\|\psi\| = 1$). We define the variance of A in the state ψ by*

$$\Delta A_\psi = \langle \psi, (A - \langle \psi, A\psi \rangle)^2 \psi \rangle = \langle \psi, A^2 \psi \rangle - \langle \psi, A\psi \rangle^2$$

Then, we have

$$\Delta A_\psi \Delta B_\psi \geq \frac{1}{4} |\langle \psi, [A, B]\psi \rangle|^2$$

Proof. Replacing A and B by $\widetilde{A} = A - \langle \psi, A\psi \rangle$ and $\widetilde{B} = B - \langle \psi, B\psi \rangle$, we can assume that $\langle \psi, A\psi \rangle = \langle \psi, B\psi \rangle = 0$. Then

$$\langle \psi, [A, B]\psi \rangle = \langle \psi, AB\psi \rangle - \langle \psi, BA\psi \rangle = 2i\mathrm{Im}\, \langle \psi, AB\psi \rangle$$

This implies that

$$|\langle\psi,[A,B]\psi\rangle| \leq 2|\langle\psi,AB\psi\rangle| = 2|\langle A\psi,B\psi\rangle| \leq 2\|A\psi\|\|B\psi\| = 2\Delta A_\psi^{1/2}\Delta B_\psi^{1/2}$$

and therefore that

$$\Delta A_\psi \Delta B_\psi \geq \frac{1}{4}|\langle\psi,[A,B]\psi\rangle|^2$$

□

Example. Consider a particle moving on $\mathbb{R}$, take $A = x$ and $B = -i\hbar\partial_x$. Then $[A,B] = i\hbar$. It follows that

$$\Delta x_\psi \Delta p_\psi \geq \frac{\hbar^2}{4}$$

In other words, x and p cannot be measured simultaneously with arbitrary precision. If ψ concentrates very much around a position x_0 (meaning that the variance Δx_ψ is very small), then $\hat{\psi}$ is flat, meaning that the variance Δp_ψ is large. Only commuting observables can be measured simultaneously.

The Hamilton operator and the time evolution. In every quantum system there is an observable that plays a particularly important role. The Hamilton operator, or the Hamiltonian, is a self-adjoint operator H on the Hilbert space $\mathcal{H}$, corresponding to the energy of the system. Through the Schrödinger equation

$$i\partial_t\psi(t) = H\psi(t) \tag{2.3}$$

for $\psi(t) \in \mathcal{H}$, the Hamilton operator generates the time evolution of the quantum system. Notice that

$$\begin{aligned}\frac{d}{dt}\|\psi(t)\|^2 &= \frac{d}{dt}\langle\psi(t),\psi(t)\rangle \\ &= i\left[\langle i\partial(t)\psi(t),\psi(t)\rangle - \langle\psi(t),i\partial_t\psi(t)\rangle\right] \\ &= i\left[\langle H\psi(t),\psi(t)\rangle - \langle\psi(t),H\psi(t)\rangle\right] \\ &= 0\end{aligned}$$

since H is self-adjoint. Hence, the Schrödinger equation defines a unitary evolution on $\mathcal{H}$, preserving the norm. Denoting by e^{-iHt} the unitary group generated by H, we can write the solution of the Schrödinger equation (2.3) with initial data $\psi(0)$ as $\psi(t) = e^{-iHt}\psi(0)$.

Suppose that $\varphi \in \mathcal{H}$ is a normalized eigenvector of H, ie $H\varphi = E\varphi$. Then, the solution of the Schrödinger equation with initial data $\psi(0) = \varphi$ is simply given by $\psi(t) = e^{-iEt}\varphi$. This implies that $\langle\psi(t),A\psi(t)\rangle = \langle\varphi,A\varphi\rangle$, independently of t. For this reason, eigenvectors of H are referred to as stationary states.

Consider now a quantum particle in $\mathbb{R}^d$, described on the Hilbert space $L^2(\mathbb{R}^d)$. As explained above, the observables depending on the position are associated with multiplication operators, while the momentum is associated with the differential operator $p = -i\nabla$. Recalling the classical energy $\mathcal{H} = p^2/(2m) + V(x)$, we can guess that the Hamilton operator for this system takes the form

$$H = -\frac{\Delta}{2m} + V(x) \tag{2.4}$$

where $\Delta = \nabla \cdot \nabla$ is the Laplace operator on $H^2(\mathbb{R}^d)$. Operators of the form (2.4) are known as Schrödinger operators. Understanding their spectral properties is crucial to understand properties of the time-evolution they generate.

Conservation laws. Let A be a self-adjoint operator on the Hilbert space $\mathcal{H}$. Let $\psi(t)$ denote the solution of the Schrödinger equation

$$i\partial_t \psi(t) = H\psi(t)$$

The variation of the expectation of the observable A has the form

$$i\frac{d}{dt}\langle \psi(t), A\psi(t)\rangle = \langle \psi(t), [A,H]\psi(t)\rangle$$

where $[A,H] = AH - HA$ is the commutator between the operators A and H. This should be compared with the Poisson brakets in the classical setting; the commutator replaces here the Poisson brackets.

In particular, for an observable A satisfying $[A,H] = 0$, the expectation $\langle \psi(t), A\psi(t)\rangle$ remains constant over time.

The free Schrödinger Equation. If there is no interaction and the quantum particle is moving freely, we talk about the free Schrödinger equation. In such case the Hamiltonian reads $H = -\Delta/2m$ on $H^2(\mathbb{R}^d)$ (this is the Hamilton operator of a free particle moving in $\mathbb{R}^d$), then $[H, -i\nabla] = 0$, ie. momentum is conserved. Similarly, the Hamilton operator $H = -\Delta/2m + V(|x|)$ acting on $H^2(\mathbb{R}^3)$, with a radial potential (ie. V only depends on the $\|x\|$), then $[H, x \wedge -i\nabla] = 0$, ie. (each component of) the angular momentum operator $L = -ix \wedge \nabla$ is conserved. We will use this conservation law below, to compute the spectrum of the Hamilton operator describing the hydrogen atom.

Example: the Stern-Gerlach experiment. In quantum mechanics, elementary particles have an internal degree of freedom, known as spin. The spin is an angular momentum; it has three components, σ_x, σ_y, σ_z. Experimentally, the existence of a spin can be verified with a Stern-Gerlach experiment, using the fact that spins couple to magnetic fields. In the Stern-Gerlach experiment, neutral particles (in the original experiment in 1922, silver atoms) are sent through a non-constant magnetic field, which deflects them according to their spin. Particles are then detected on a screen. The outcome of the experiment was, at the beginning, surprising (it led to the Nobel Prize for Stern in 1943); the screen does not show a continuous distribution but, instead, it reveals discrete accumulation points, showing that the spin (more precisely, the component of the spin in the direction parallel to the applied magnetic field) is quantized. Particles are associated with a spin number $n \in \mathbb{N}/2$; each spin component has then $2n+1$ possible values. In particular, for spin $1/2$ particles (like electrons, protons), each spin component has two possible values (spin up or spin down).

If we forget about all other degrees of freedom, we can describe the setting of the Stern-Gerlach experiment, restricting for simplicity our attention to spin $1/2$ particles, as a quantum system on $\mathbb{C}^2$. If the magnetic field in the Stern-Gerlach apparatus points in the z-direction, it is convenient to choose the vectors $(1,0), (0,1) \in \mathbb{C}^2$ to describe particles with the z-component of the spin pointing up and, respectively, down. In this representation of the system, the z-component of the spin is associated with the self-adjoint matrix

$$\sigma_z = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \tag{2.5}$$

The x- and the y-component of the spin are associated instead with the matrices

$$\sigma_x = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \qquad \sigma_y = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix} \tag{2.6}$$

The matrices $\sigma_x, \sigma_y, \sigma_z$ are known as Pauli matrices. They satisfy the commutation relation

$$[\sigma_x, \sigma_y] = 2i\sigma_z \tag{2.7}$$

and its cyclic permutations. In particular, notice that spins in different directions do not commute. Notice that we did not try to justify the choice of the matrices (2.5), (2.6). In fact, as we will discuss later, relations between operators associated to different spin components (in particular the commutator relations (2.7)), are determined by the properties of the group of rotations $SO(3)$.
The Stern Gerlach experiment is a measurement of the z-component of the spin. If the initial state of the system is described by the vector $\psi = (\alpha, \beta) \in \mathbb{C}^2$, normalized so that $|\alpha|^2 + |\beta|^2 = 1$, the measurement of σ_z will give the value $+1$ with probability $|\alpha|^2$ and the value -1 with probability $|\beta|^2$. In other words, after going through the Stern-Gerlach appartus, the particle will move up and hit the upper accumulation point on the screen with probability $|\alpha|^2$ and it will move down and hit the lower accumulation point on the screen with probability $|\beta|^2$. If the measurement of σ_z gives the value $+1$, after going through the apparatus the particle is described by the vector $(1, 0)$. In this case, if we let the particle go through a second Stern-Gerlach experiment, again with magnetic field pointing in the z direction, the particle will show spin up with probability one. On the other hand, if the measurement of σ_z gives the value -1, after going through the apparatus the particle is described by the vector $(0, 1)$. In this case, a second measurement of σ_z will give again the value -1 with probability one. It is also possible to insert, after the first Stern-Gerlach apparatus measuring the z-component of the spin, a Stern-Gerlach apparatus measuring, say, the x-component of the spin. Let us assume, for example, that the first measurement of σ_z returns the value $+1$ and therefore that after the first measurement the system is described by the vector $(1, 0)$. The expectation value of σ_x is given by

$$\left\langle (1,0), \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} (1,0) \right\rangle = 0$$

Since $\sigma_x^2 = 1$, the variance of σ_x is given by $\Delta\sigma_x = 1$. A different approach to reach the same conclusion consists in noticing that the matrix σ_x has the two eigenvalues ± 1 associated with the normalized eigenvectors $(1/\sqrt{2}, 1/\sqrt{2})$ and $(1/\sqrt{2}, -1/\sqrt{2})$. Since

$$(1,0) = \frac{1}{\sqrt{2}}(1/\sqrt{2}, 1/\sqrt{2}) + \frac{1}{\sqrt{2}}(1/\sqrt{2}, -1/\sqrt{2})$$

we conclude that the measurement of σ_x gives the value $+1$ and the value -1 both with probability $1/2$ (this explains also why the expectation of σ_x vanishes and its variance equals one). This result shows a well-known phenomenon in quantum systems. Non-commuting observables cannot be measured simultaneously. It is impossible to know the precise value of σ_z and of σ_x at the same time (mathematically, this means that two non-commuting self-adjoint operators cannot be diagonalized simultaneously).

Density matrices and mixed states. As explained above a quantum mechanical system is described on a Hilbert space $\mathcal{H}$. States of the system are associated with normalized vectors $\psi \in \mathcal{H}$. Observables are associated with self-adjoint operators A acting on $\mathcal{H}$. The expectation value for the measurement of the observable A in the state ψ is then given by the inner product $\langle \psi, A\psi \rangle$. The same state can be described through the orthogonal projection onto ψ, denoted by

$$\gamma = |\psi\rangle\langle\psi|$$

The expectation of the observable A is then given by

$$\langle \psi, A\psi \rangle = \mathrm{tr} A\gamma$$

In this case, describing the system through γ is fully equivalent to the description given by ψ. Moreover, observe that the description through γ shows that $e^{i\theta}\psi$ and ψ describe the same state. The description through γ has the advantage that it can be easily extended to mixed states.

Definition 2.1.2. *A density matrix over the Hilbert space $\mathcal{H}$ is a non-negative (in particular self-adjoint) trace class operator γ on $\mathcal{H}$, with $tr\gamma = 1$. A trace class operator is a compact operator, with the property that the sequence $\{\lambda_j\}_j$ of its eigenvalues is summable, ie. $\sum_j |\lambda_j| < \infty$. The trace of a trace class operator γ is defined by $tr\gamma = \sum_j \lambda_j$*

Given a density matrix γ over $\mathcal{H}$, we can always decompose it as

$$\gamma = \sum_j \lambda_j |\psi_j\rangle\langle\psi_j|$$

where $\{\psi_j\}_j$ is an orthonormal basis in $\mathcal{H}$ and where $\lambda_j \geq 0$ for all j, with $\sum_j \lambda_j = 1$. Such a density matrix describes the system, if it is in the state ψ_j with probability λ_j. If the density matrix γ is not an orthogonal projection, it is said to describe a mixed state (on the other hand, if $\gamma = |\psi\rangle\langle\psi|$ is an orthogonal projection, it is said to describe the pure state, which can be equivalently be described by the vector $\psi \in \mathcal{H}$).

There is an important consequence of being a pure or a mixed state, which is connected to the information we have on the quantum system. Consider pure states, given by linear combinations of vectors in $\mathcal{H}$, and mixed states, given by a density matrix that is a (convex) linear combination of orthogonal projections. Suppose for example ψ_1, ψ_2 are two orthogonal normalized vectors in $\mathcal{H}$. Then

$$\psi = \frac{1}{\sqrt{2}}(\psi_1 + \psi_2) \tag{2.8}$$

is again a normalized vector in $\mathcal{H}$. The expectation of an observable A in the pure state ψ is given by

$$\langle \psi, A\psi \rangle = \frac{1}{2}\left[\langle \psi_1, A\psi_1 \rangle + \langle \psi_2, A\psi_2 \rangle + \langle \psi_1, A\psi_2 \rangle + \langle \psi_2, A\psi_1 \rangle\right] \tag{2.9}$$

On the other hand, consider the mixed state, described by the density matrix

$$\gamma = \frac{1}{2}\left[|\psi_1\rangle\langle\psi_1| + |\psi_2\rangle\langle\psi_2|\right] \tag{2.10}$$

The expectation of A in the mixed state with density matrix γ is given by

$$\mathrm{tr} A\gamma = \frac{1}{2}\left[\langle\psi_1, A\psi_1\rangle + \langle\psi_2, A\psi_2\rangle\right]$$

In contrast to (2.9), there are here no "interference" terms of the form $\langle\psi_1, A\psi_2\rangle$, $\langle\psi_2, A\psi_1\rangle$. We call (2.8) a coherent superposition of the quantum states ψ_1, ψ_2. On the other hand, (2.10) is called an incoherent superposition of quantum states.
Mixed states are needed, when there is no complete information about the system. For example, equilibrium states at positive temperature are mixed.

2.2 Many-Body Quantum Mechanics

Bosonic and fermionic statistics. A system of N particles moving in the 3-dimensional space is described on the Hilbert space

$$\mathcal{H}_N = L^2(\mathbb{R}^{dN}, dx_1 \dots dx_N) = L^2(\mathbb{R}^3)^{\otimes N}$$

given by the tensor product of N copies of the one-particle space $L^2(\mathbb{R}^3)$. Wave functions $\psi_N \in \mathcal{H}_N$ are normalized so that

$$\|\psi_N\|^2 = \int |\psi_N(x_1, \dots, x_N)|^2 \, dx_1 \dots dx_N = 1$$

Accordingly, $|\psi_N(x_1, \dots, x_N)|^2$ is the probability density for finding particle 1 close to x_1, particle 2 close to x_2, and so on.
When considering systems of N indistinguishable particles, we need to restrict the Hilbert space to subspaces with appropriate statistics. Indistinguishability plays an important role in quantum mechanics, much more important than in classical mechanics. Classical particles follow trajectories and can be labeled. Quantum particles do not. Indistinguishability translates, in quantum system, into a restriction of the set of physical observables. Only observables that are invariant with respect to permutations of the N particles can be measured.
Let us denote by S_N the group of permutations π of the set $\{1, 2, \dots, N\}$. We define a unitary representation of S_N on the N-particle Hilbert space $L^2(\mathbb{R}^{3N})$ setting

$$(U_\pi \psi)(x_1, \dots, x_N) = \psi(x_{\pi 1}, \dots, x_{\pi N})$$

Physically meaningful observables of the N-particle system are self-adjoint operators A on $L^2(\mathbb{R}^{3N})$ with the property $U_\pi^* A U_\pi = A$ (or, equivalently, $[A, U_\pi] = 0$) for all $\pi \in S_N$.
The existence of such unitary representation is justified by the Abelianization of the group of permutations. More precisely, there exists a homeomorphism $e^{i\theta(\cdot)} : S_N \to U(1)$, where $U(1)$ is the unit circe, that is a one-dimensional representation of the symmetric group. We can therefore consider the Abelianization of S_N with the standard procedure[1] and construct the Abelian group

[1] By Abelianization of a group G we refer to the existence of a homeomorphism $h : G \mapsto G'$ such that G' is Abelian. Such homeomorphism can be constructed by considering the kernel given by the commutator subgroup $[G, G]$, which is the unique smallest normal subgroup of G such that the quotient group $G' = G/[G, G]$ is Abelian.

$S_N/[S_N, S_N]$. Observe that in dimension $d \geq 3$ this is topologically justified (whereas for $d = 2$ there are infinitely many choices for θ). The largest Abelian quotient is then made of two elements: the identity and the parity of the permutation. Since any one-dimensional representation must factor through the Abelianization, there are only two possibilities: for every $\pi \in S_N$, either $e^{i\theta(\pi)} = 1$ or $e^{i\theta(\pi)} = (-1)^{\mathrm{sgn}(\pi)}$, where $\mathrm{sgn}(\pi)$ is the sign of the permutation π. More precisely, if π consists of an even number of transpositions, $\sigma_\pi = -1$ if π consists of an odd number of transpositions). For each quantum particle, either completely symmetric or completely antisymmetric wave functions can be considered, not both at the same time. Particles described by wave functions that are symmetric w.r.t. permutations are called bosons. Particles described by antisymmetric wave function are called fermions. Systems of N bosons moving in the three dimensional space are described on the Hilbert space

$$L^2_s(\mathbb{R}^{3N}) = L^2(\mathbb{R}^3)^{\otimes_s N} = \{\psi \in L^2(\mathbb{R}^{3N}) : \psi(x_{\pi 1}, \dots, x_{\pi N}) = \psi(x_1, \dots, x_N)\}$$

Systems of N fermions on $\mathbb{R}^3$ are described, on the other hand, on the Hilbert space

$$L^2_a(\mathbb{R}^{3N}) = L^2(\mathbb{R}^3)^{\wedge N} = \{\psi \in L^2(\mathbb{R}^{3N}) : \psi(x_{\pi 1}, \dots, x_{\pi N}) = \sigma_\pi \psi(x_1, \dots, x_N)\}$$

The choice between bosonic and fermionic statistics is related with the internal degree of freedom of quantum particles, known as spin (already discussed in the Stern-Gerlach experiment). Particles with integer spin are bosons (they are described by wave functions that are symmetric w.r.t. permutations). Particles with half-integer spin are fermions (they must be described by wave functions that are antisymmetric w.r.t. permutations).

Examples of bosonic wave function: The simplest example of a bosonic wave function is the product of N copies of the same one-particle normalized wave function $\varphi \in L^2(\mathbb{R}^3)$, ie

$$\psi_N(x_1, \dots, x_N) = \prod_{j=1}^{N} \varphi(x_j)$$

which is clearly normalized and symmetric w.r.t. permutations.
More generally, if $\{\varphi_j\}_{j\in\mathbb{N}}$ is an orthonormal basis in $L^2(\mathbb{R}^3)$, we can describe the N-particle state with one particle in the state φ_{i_1}, one particle in the state φ_{i_2}, and so on, the symmetrized wave function

$$\psi_B(x_1, \dots, x_N) = \sqrt{\frac{1}{\ell_1!\ell_2! \dots \ell_k! N!}} \sum_{\pi \in S_N} \varphi_{i_1}(x_{\pi_1}) \varphi_{i_2}(x_{\pi 2}) \dots \varphi_{i_N}(x_{\pi N})$$

if the set $\{i_1, \dots, i_N\}$ contains k different indices, appearing respectively $\ell_1, \dots \ell_k$ times, with $\ell_1 + \ell_2 + \dots + \ell_k = N$.
Similarly, we can define construct fermionic wave functions. In this case, however, each one-particle state can be "occupied" by at most one-particle, otherwise the antisymmetrization is going to vanish (this is known as the Pauli principle). The fermionic state with one particle in the state φ_{i_1}, one particle in the state φ_{i_2}, and so on (with all different indices $i_1, \dots, i_N$) , can be described by the wave function

$$\psi_F(x_1, \dots, x_N) = \frac{1}{\sqrt{N!}} \sum_{\pi \in S_N} \sigma_\pi \varphi_{i_1}(x_{\pi 1}) \dots \varphi_{i_N}(x_{\pi N})$$

The wave function ψ_F can be written as

$$\psi_F(x_1,\dots,x_N)=\frac{1}{\sqrt{N!}}\det\,(\varphi_{i_j}(x_k))_{1\le j,k\le N}$$

Such wave functions are known as Slater determinants.

Reduced density matrices. The density matrix of a normalized N-particle wave function $\psi_N\in L^2(\mathbb{R}^{3N})$, with either bosonic or fermionic statistics, is defined as $\gamma_N=|\psi_N\rangle\langle\psi_N|$, ie. γ_N is the orthogonal projection onto ψ_N. More generally, mixed states of an N-particle system are described by non-negative trace class operators γ_N with $\mathrm{tr}\gamma_N=1$.
For a density matrix γ_N, we denote by $\gamma_N(x_1,\dots,x_N;y_1,\dots,y_N)$ the integral kernel of γ_N. This is defined so that

$$(\gamma_N\xi_N)(x_1,\dots,x_N)=\int\gamma_N(x_1,\dots,x_N;y_1,\dots,y_N)\xi_N(y_1,\dots,y_N)$$

Every trace class operator can be associated with an integral kernel (since every trace class operator is Hilbert Schmidt, the integral kernel is in fact a $L^2(\mathbb{R}^{3N}\times\mathbb{R}^{3N})$-function).
Using the integral kernel of γ_N, we can define its reduced density matrices. For $1\le k\le N$, we define the k-particle reduced density matrix associated with γ_N (and therefore with ψ_N) by

$$\gamma^{(k)}=\frac{N!}{(N-k)!}\mathrm{tr}_{k+1,k+2,\dots,N}\gamma_N$$

where $\mathrm{tr}_{k+1,\dots,N}\gamma_N$ denotes the partial trace of γ_N over the last $(N-k)$ particles. Taking the partial trace means integrating out the degrees of freedom associated with the last $(N-k)$ particles. In other words, $\gamma_N^{(k)}$ is defined through its integral kernel

$$\begin{aligned}&\gamma_N^{(k)}(x_1,\dots,x_k;y_1,\dots,y_k)\\&\quad=\frac{N!}{(N-k)!}\int\gamma_N(x_1,\dots x_k,x_{k+1},\dots,x_N;y_1,\dots y_k,x_{k+1},\dots,x_N)dx_{k+1}\dots dx_N\end{aligned}$$

By definition, it is easy to check that $\gamma_N^{(k)}$ is a non-negative operator, for all $1\le k\le N$. In fact, for an arbitrary $\varphi\in L^2(\mathbb{R}^{3k})$ we have

$$\begin{aligned}\langle\varphi,\gamma^{(k)}\varphi\rangle&=\int dx_1\dots dx_k\,dy_1\dots dy_k\bar\varphi(x_1,\dots,x_k)\gamma_N^{(k)}(x_1,\dots,x_k;y_1,\dots,y_k)\varphi(y_1,\dots,y_k)\\&=\int dx_{k+1}\dots dx_N\left|\int dx_1\dots dx_k\,\bar\varphi(x_1,\dots,x_k)\psi_N(x_1,\dots,x_k,x_{k+1},\dots,x_N)\right|^2\ge 0\end{aligned}$$

Moreover,

$$\mathrm{tr}\gamma_N^{(k)}=\frac{N!}{(N-k)!}\int dx_1\dots dx_k\gamma_N^{(k)}(x_1,\dots,x_k;x_1,\dots,x_k)=\frac{N!}{(N-k)!}$$

Reduced density matrices are useful, because they allow us to compute the expectation of observables that depend only on a fixed number k of particles. Let $\mathcal{O}^{(k)}$ be an operator on the k-particle space

$L^2(\mathbb{R}^{3k})$. Then $\mathcal{O}^{(k)} \otimes 1^{(N-k)}$ is an operator on $L^2(\mathbb{R}^{3N})$, acting as $\mathcal{O}^{(k)}$ on the first k particles and as the identity on the other $(N-k)$ particles. A simple computation shows that

$$\langle \psi_N, (\mathcal{O}^{(k)} \otimes 1^{(N-k)}) \psi_N \rangle = \frac{(N-k)!}{N!} \mathrm{tr}\, \mathcal{O}^{(k)} \gamma_N^{(k)}$$

where, on the r.h.s., tr denotes the trace over $L^2(\mathbb{R}^{3k})$. Interesting k-particle observables in N-body systems must be invariant w.r.t. permutations of the N particles and are therefore given by sums over all possible choices of the k indices, over which $\mathcal{O}^{(k)}$ must act non-trivially. Using the permutation invariance of ψ_N, their expectation is given by

$$\sum_{1 \leq i_1 < i_2 < \cdots < i_k \leq N} \langle \psi_N, \mathcal{O}^{(k)}_{i_1,\ldots,i_k} \psi_N \rangle = \sum_{1 \leq i_1 < i_2 < \cdots < i_k \leq N} \frac{(N-k)!}{N!} \mathrm{tr}\, \mathcal{O}^{(k)} \gamma_N^{(k)} = \frac{1}{k!} \mathrm{tr}\, \mathcal{O}^{(k)} \gamma_N^{(k)}$$

For example, the kinetic energy of an N-body system is the one-particle operator $\sum_{j=1}^N -\Delta_{x_j}$. Its expectation in the state ψ_N can be computed through the one-particle reduced density $\gamma_N^{(1)}$:

$$\langle \psi_N, \sum_{j=1}^N -\Delta_{x_j} \psi_N \rangle = \mathrm{tr}(-\Delta) \gamma_N^{(1)}$$

A two-body interaction has the form $\sum_{i<j}^N V(x_i - x_j)$. To compute its expectation we need its 2-particle reduced density:

$$\langle \psi_N, \sum_{i<j}^N V(x_i - x_j) \psi_N \rangle = \frac{1}{2} \mathrm{tr} V(x_1 - x_2) \gamma_N^{(2)}$$

Let us compute a couple of reduced density matrices, for simple example of N-particle states. In the bosonic setting, the simplest class of N-particle states are factorized wave functions of the form $\psi_N(x_1, \ldots, x_N) = \prod_{j=1}^N \varphi(x_j)$. In this case, it is very easy to compute reduced density matrices. We find

$$\gamma_N^{(k)} = \frac{N!}{(N-k)!} |\varphi\rangle\langle\varphi|^{\otimes k}$$

where $|\varphi\rangle\langle\varphi|$ is the orthogonal projection onto φ.
In the fermionic setting, the simplest class of N-particle states are Slater determinant, having the form

$$\psi_N(x_1, \ldots, x_N) = \frac{1}{\sqrt{N!}} \det\ (\varphi_i(x_j))_{1 \leq i,j \leq N}$$

where $\{\varphi_j\}_{j=1}^N$ is an orthonormal system on $L^2(\mathbb{R}^3)$. The one-particle reduced density matrix associated with ψ_N has the integral kernel

$$\begin{aligned} \gamma_N^{(1)}(x;x') &= \frac{N}{N!} \sum_{\pi,\pi' \in S_N} \sigma_\pi \sigma_{\pi'} \varphi_{\pi 1}(x) \varphi_{\pi' 1}(x') \int dx_2 \ldots dx_N \prod_{j=2}^N \varphi_{\pi j}(x_j) \varphi_{\pi' j}(x_j) \\ &= \frac{1}{(N-1)!} \sum_{\pi} \varphi_{\pi 1}(x) \varphi_{\pi 1}(x') = \sum_{j=1}^N \varphi_j(x) \varphi_j(x') \end{aligned}$$

Hence $\gamma_N^{(1)} = \sum_{j=1}^N |\varphi_j\rangle\langle\varphi_j|$ is the orthogonal projection onto the N-dimensional subspace of $L^2(\mathbb{R}^3)$, spanned by $\varphi_1, \ldots, \varphi_N$. Similarly, one can compute higher order reduced density matrices of the Slater determinant ψ_N. It turns out, they satisfy Wick's theorem, stating that

$$\gamma_N^{(k)}(x_1, \ldots x_k; x_1', \ldots, x_k') = \sum_{\pi \in S_k} \sigma_\pi \prod_{j=1}^k \gamma_N^{(1)}(x_j; x_{\pi j}')$$

In particular, the 2-particle reduced density is given by

$$\gamma_N^{(2)}(x_1, x_2; x_1', x_2') = \gamma_N^{(1)}(x_1; x_1')\gamma_N^{(1)}(x_2; x_2') - \gamma_N^{(1)}(x_1; x_2')\gamma_N^{(1)}(x_2; x_1')$$

Typical N-particle Hamilton operators have the form

$$H_N = \sum_{j=1}^N \big[-\Delta_{x_j} + V_{\text{ext}}(x_j)\big] + \sum_{i<j} V(x_i - x_j)$$

The first term is a one-particle operator, it describes the kinetic energy of the particles and their interaction with external fields. The second term is a two-body operator, describing the interaction among the particles. The energy of the N-particle system in a state ψ_N can be expressed in terms of the reduced density matrix $\gamma_N^{(2)}$ associated with ψ_N. We find

$$\langle \psi_N, H_N \psi_N \rangle = \text{tr}\,\big[-\Delta + V_{\text{ext}}\big]\gamma_N^{(1)} + \frac{1}{2}\text{tr}\, V(x_1 - x_2)\gamma_N^{(2)} \tag{2.11}$$

This identity suggests that a possible approach to compute the ground state energy of an N-particle system, ie. the infimum of $\langle \psi_N, H_N \psi_N \rangle$ over all normalized wave functions $\psi_N \in L^2(\mathbb{R}^{3N})$ with the appropriate statistics, consists in minimizing the r.h.s. of (2.11) over all possible choices of $\gamma_N^{(2)}$ (the one-particle density matrix $\gamma_N^{(1)}$ can be computed from $\gamma_N^{(2)}$, taking the partial trace over one of the two particles).

A Mathematical Background

A.1 L^p spaces and Hilbert spaces

Let $(\Omega, \mathcal{A}, \mu)$ be a measure space, where Ω is a set, $\mathcal{A}$ a σ-algebra on Ω and μ a measure on the σ-algebra $\mathcal{A}$. For $1 \leq p < \infty$, let

$$\mathcal{L}^p(\Omega, \mathcal{A}, \mu) = \left\{ f : \Omega \to \mathbb{C} \quad \mathcal{A}\text{-measurable s.t. } \int |f|^p d\mu < \infty \right\}$$

and, $f \in \mathcal{L}^p(\Omega, \mathcal{A}, \mu)$,

$$\|f\|_p = \left[\int |f|^p d\mu \right]^{1/p}.$$

For $p = \infty$, let

$$\mathcal{L}^\infty(\Omega, \mathcal{A}, \mu) = \{f : \Omega \to \mathbb{C} \text{ measurable s.t. } \exists\, C > 0 \text{ with } |f(x)| \leq C \text{ almost everywhere}\}$$

and, for $f \in \mathcal{L}^p(\Omega, \mathcal{A}, \mu)$ let

$$\|f\|_\infty = \inf\{C > 0 : |f(x)| \leq C \text{ almost everywhere}\}.$$

Hence $\|.\|_p : \mathcal{L}^p(\Omega, \mathcal{A}, \mu) \to [0; \infty)$, for any $1 \leq p \leq \infty$

$$\|\lambda f\|_p = |\lambda| \|f\|_p,$$

for all $\lambda \in \mathbb{C}$ and all $f \in \mathcal{L}^p(\Omega, \mathcal{A}, \mu)$. Moreover, for all $f, g \in \mathcal{L}^p(\Omega, \mathcal{A}, \mu)$ the triangular inequality holds:

$$\|f + g\|_p \leq \|f\|_p + \|g\|_p, \tag{A.12}$$

Nevertheless, $\|.\|_p$ does not define a norm on the vector space $\mathcal{L}^p(\Omega, \mathcal{A}, \mu)$, because $\|f\|_p = 0$ does not imply that $f = 0$ (it only implies that $f(x) = 0$ almost everywhere).
To make $\|.\|_p$ into a norm, we define the following equivalence relation on $\mathcal{L}^p(\Omega, \mathcal{A}, \mu)$, an equivalence relation. For $f, g \in \mathcal{L}^p(\Omega, \mathcal{A}, \mu)$ we write

$$f \sim g \text{ if and only if } f(x) = g(x) \ \mu\text{-almost everywhere.}$$

We can therefore define, for any $1 \leq p \leq \infty$,

$$L^p(\Omega, \mathcal{A}, \mu) = \mathcal{L}^p(\Omega, \mathcal{A}, \mu)/\sim = \{[f] : f \in \mathcal{L}^p(\Omega, \mathcal{A}, \mu)\}$$

where $[f] = \{g \in \mathcal{L}^p(\Omega, \mathcal{A}, \mu) : g \sim f\}$ is the equivalence class associated with f, which contains all functions that coincide with f almost everywhere. Switching from $\mathcal{L}^p(\Omega, \mathcal{A}, \mu)$ to $L^p(\Omega, \mathcal{A}, \mu)$, we make the space smaller, because we identify functions that coincide almost everywhere. By definition, elements of $L^p(\Omega, \mathcal{A}, \mu)$ are not functions, they are equivalence classes of functions.
On $L^p(\Omega, \mathcal{A}, \mu)$ we can define the structure of a vector space, by setting

$$[f] + [g] = [f + g], \qquad \lambda[f] = [\lambda f]$$

We can also define $\|.\|_p : L^p(\Omega, \mathcal{A}, \mu) \to [0; \infty)$ through $\|[f]\|_p := \|f\|_p$ ($\|.\|_p$ is well defined, because $\|f\|_p = \|g\|_p$ if $f \sim g$). When defined on $L^p(\Omega, \mathcal{A}, \mu)$, $\|.\|_p$ is indeed a norm, because $\|[f]\|_p = 0$ implies that $f(x) = 0$ almost everywhere, and therefore $[f] = [0]$. We conclude that, for every $1 \leq p \leq \infty$, $(L^p(\Omega, \mathcal{A}, \mu), \|.\|_p)$ is a normed vector space. As every norm, $\|.\|_p$ induces a metric on $L^p(\Omega, \mathcal{A}, \mu)$ (the metric is defined by $d(f, g) = \|f - g\|_p$), and thus a notion of convergence for sequences in $L^p(\Omega, \mathcal{A}, \mu)$.

Theorem A.1.1. *Let $(\Omega, \mathcal{A}, \mu)$ be a measure space, $1 \leq p \leq \infty$. Then $L^p(\Omega, \mathcal{A}, \mu)$, equipped with the norm $\|.\|_p$. is a Banach space, ie. it is a normed vector space, complete with respect to the metric induced by $\|.\|_p$. In other words, every Cauchy sequence on $L^p(\Omega, \mathcal{A}, \mu)$ converges.*

Definition A.1.2 (Hilbert space)**.** *Let $\mathcal{H}$ be a vector space over $\mathbb{C}$, equipped with a scalar product $\langle \cdot, \cdot \rangle$ and let $\| \cdot \|$ be the norm induced by this scalar product. If $\mathcal{H}$, equipped with this norm, is complete, then the pair $(\mathcal{H}, \langle \cdot, \cdot \rangle)$ is called a Hilbert space.*

A notable example of Hilbert space is the Lebesgue space $L^2(\Omega, \mathcal{A}, \mu)$. Indeed, we can define an inner product on $L^2(\Omega, \mathcal{A}, \mu)$. For $f, g \in L^2(\Omega, \mathcal{A}, \mu)$, we set

$$\langle f, g \rangle = \int \overline{f} g \, d\mu \tag{A.13}$$

It is easy to check that $\langle \cdot, \cdot \rangle$ is linear in its second argument, and antisymmetric (ie. $\langle g, f \rangle = \overline{\langle f, g \rangle}$). Moreover,

$$\langle f, f \rangle = \|f\|_2^2 \geq 0$$

Thus $\langle \cdot, \cdot \rangle$ is an inner product on $L^2(\Omega, \mathcal{A}, \mu)$ and it induces the norm $\| \cdot \|_2$. Hence, $L^2(\Omega, \mathcal{A}, \mu)$ is a Hilbert space. In particular, we have the Cauchy-Schwarz inequality

$$\left| \int \overline{f} g \, d\mu \right| = |\langle f, g \rangle| \leq \|f\|_2 \|g\|_2$$

for all $f, g \in L^2(\Omega, \mathcal{A}, d\mu)$.
The generalization of Cauchy-Schwarz inequality to L^p-spaces is called Hölder's inequality.

Theorem A.1.3 (Hölder's inequality)**.** *Let $1 \leq p, p' \leq \infty$, with $1/p + 1/p' = 1$. For $f \in L^p(\Omega, \mathcal{A}, \mu), g \in L^{p'}(\Omega, \mathcal{A}, \mu)$, we have*

$$\left| \int \overline{f} g \, d\mu \right| \leq \int |f||g| d\mu \leq \|f\|_p \|g\|_{p'} \tag{A.14}$$

The dual of $L^p(\Omega, \mathcal{A}, \mu)$. For an arbitrary normed vector space $(X, \|.\|)$ over $\mathbb{C}$, a linear functional on X is a linear map $L : X \to \mathbb{C}$. A linear functional L is continuous if and only if it is bounded, ie. if there exists a constant $C > 0$ such that $|L(f)| \leq C\|f\|$ for all $f \in X$. We define the dual space of X as

$$X^* = \{L : X \to \mathbb{C} : \ L \text{ is a continuous linear functional on } X\}$$

On X^* we can naturally introduce a sum and a multiplication with complex numbers. Hence X^* is a vector space. Since continuous functionals are bounded, we can also introduce a norm on X^*, by setting, for $L \in X^*$,

$$\|L\| = \sup_{f \in X : \|f\| \leq 1} |L(f)| = \sup_{f \in X : f \neq 0} \frac{|L(f)|}{\|f\|} \tag{A.15}$$

Equipped with this norm, X^* is a normed vector space. In fact, X^* is always complete (w.r.t. the norm (A.15), of course), and thus a Banach space.
In the following, we would like to characterize the dual space of $L^p(\Omega, \mathcal{A}, \mu)$. To this end, we observe that, choosing $1 \leq p' \leq \infty$ so that $1/p + 1/p' = 1$, for every $g \in L^{p'}(\Omega, \mathcal{A}, \mu)$, we can define a linear functional $L_g : L^p(\Omega, \mathcal{A}, \mu) \to \mathbb{C}$ by setting

$$L_g(f) = \int f g d\mu$$

Linearity of L_g is clear, continuity follows from Hölder's inequality. For $1 \leq p < \infty$, it turns out that all linear functionals on $L^p(\Omega, \mathcal{A}, \mu)$ have this form.

Theorem A.1.4. *Let $(\Omega, \mathcal{A}, \mu)$ be a measure space, $1 \leq p < \infty$ and $1 < p' \leq \infty$ with $1/p + 1/p' = 1$. If $p = 1$ and $p' = \infty$, we make the additional assumption that the measure space is sigma-finite. The map $\phi : L^{p'}(\Omega, \mathcal{A}, \mu) \to L^p(\Omega, \mathcal{A}, \mu)^*$ defined by $\phi(g) = \phi_g : L^p(\Omega, \mathcal{A}, \mu) \to \mathbb{C}$, with*

$$\phi_g(f) = \int f g d\mu$$

is an isometric isomorphism, meaning that it is linear, it preserves the norm, in the sense that $\|\phi_g\|_{L^p(\Omega,\mathcal{A},\mu)^*} = \|g\|_{L^{p'}}$, and it is a bijection.

Remark: If $p = \infty$ and $p' = 1$, the map $\phi : L^1(\Omega, \mathcal{A}, \mu) \to L^\infty(\Omega, \mathcal{A}, \mu)^*$ is still well defined, linear, isometric (and therefore injective), but it is (in general) not surjective.

A.2 Self-adjoint operators

On the Hilbert space $\mathcal{H}$ we define linear operators $A : \mathcal{H} \to \mathcal{H}$ as maps from the Hilbert space $\mathcal{H}$ to itself satisfying linearity, i.e.

$$A(\lambda f + \mu g) = \lambda A(f) + \mu A(g), \text{for all } f, g \in \mathcal{H}, \quad \forall\ \lambda, \mu \in \mathbb{R}$$

Such operators are called self-adjoint if

$$\langle f\ ,\ A\, g \rangle = \langle A\, f\ ,\ g \rangle \qquad \forall\, f, g \in \mathcal{H}.$$

An operator is called *bounded* if for all elements of the domain (in this case $\forall\ f \in \mathcal{H}$) there exists a positive constant C such that

$$\|Af\| \leq C\|f\|, \tag{A.16}$$

where $\|\cdot\|$ is the norm induced by the inner product on $\mathcal{H}$. The operator norm of a bounded operator is defined as the minimal constant C for which (A.16) holds true. Let $|||A|||$ denote the operator norm of the operator A, then the following characterization holds:

$$|||A||| = \sup_{f \in \mathcal{H}\ :\ \|f\| \leq 1} \frac{\|Af\|}{\|f\|}.$$

Proposition A.2.1. *For any bounded operator A, there is a unique bounded operator A^*, called the adjoint of A, such that*

$$\langle \phi, A\psi \rangle = \langle A^*\phi, \psi \rangle \qquad \forall \phi, \psi \in \mathcal{H}$$

The proof of the proposition replies on the celebrated Riesz Theorem:

Theorem A.2.2 (Riesz's Theorem)**.** *If $\xi : \mathcal{H} \to \mathbb{C}$ is a bounded linear functional, then there exists a unique $\chi \in \mathcal{H}$ sich that $\xi(\psi) = \langle \chi, \psi \rangle$ for all $\psi \in \mathcal{H}$.*

Proof of Proposition A.2.1. For each fixed $\phi \in \mathcal{H}$ consider the map $\psi \mapsto \langle \phi, A\psi \rangle$ for all $\psi \in \mathcal{H}$. For each fixed $\phi \in \mathcal{H}$, the map is a linear functional on $\mathcal{H}$. Therefore, by Riesz's Theorem, there exists a unique $\chi \in \mathcal{H}$ such that $\langle \chi, \psi \rangle = \langle \phi, A\psi, \rangle$ and we use the notation $\chi = A^*\phi$. Boundedness and linearity follow from the boundedness of A and by construction of the map. □

Definition A.2.3. *A is an unbounded operator on $\mathcal{H}$ if it is not a bounded operator and it is a linear map from a dense subspace $\mathcal{D}(A) \subset \mathcal{H}$ into $\mathcal{H}$.*

Definition A.2.4. *Let A be an unbounded operator on $\mathcal{H}$. A^* is defined as follows:*

i) $\phi \in \mathcal{H}$ belongs to $\mathcal{D}(A^*)$ if $\phi \mapsto \langle \phi, A\cdot \rangle$ defined on $\mathcal{H}$ is bounded;

ii) for each $\phi \in \mathcal{D}(A^*)$, $A^*\phi$ is the unique vector $\chi \in \mathcal{H}$ such that $\langle \chi, \psi \rangle = \langle \phi, A\psi \rangle$, for all $\psi \in \mathcal{D}(A)$.

Definition A.2.5. *An unbounded operator A on $\mathcal{H}$ is symmetric if $\langle \phi, A\psi \rangle = \langle A\phi, \psi \rangle$ for all $\psi, \phi \in \mathcal{D}(A)$.*
It is self-adjoint if $\mathcal{D}(A) = \mathcal{D}(A^*)$ and $A^*\phi = A\phi$ for all $\phi \in \mathcal{D}(A)$.

A.3 The Fourier Transform

For $f \in L^1(\mathbb{R}^d)$, we define the Fourier transform $\hat{f} : \mathbb{R}^d \to \mathbb{C}$ of f, setting

$$\hat{f}(k) = \frac{1}{(2\pi)^{d/2}} \int f(x) e^{-ik\cdot x} d\lambda_d(x), \tag{A.17}$$

where λ_d is the Lebesgue measure in $\mathbb{R}^d$.
The Fourier transform has the following important properties.

- *Boundedness and continuity.* Let $f \in L^1(\mathbb{R}^d)$. Then $\hat{f} \in L^\infty(\mathbb{R}^d)$ with

  $$\|\hat{f}\|_\infty \leq (2\pi)^{-d/2} \|f\|_1$$

  Moreover, $\hat{f}$ is continuous.

- *Riemann-Lebesgue Lemma.* Let $f \in L^1(\mathbb{R}^d)$. Then $\hat{f}(k) \to 0$, as $|k| \to \infty$.

- *Fourier transform of convolutions.* For $f, g \in L^1(\mathbb{R}^d)$ we have (from Young's inequality) $f * g \in L^1(\mathbb{R}^d)$. Then

  $$\widehat{f * g}(k) = (2\pi)^{d/2} \hat{f}(k) \hat{g}(k)$$

- *Fourier transform of a Gaussian.* Let $f_\alpha(x) = e^{-\alpha x^2/2}$. Then

  $$\hat{f}_\alpha(k) = \alpha^{-d/2} e^{-k^2/2\alpha}$$

- *Derivatives and Fourier transform.* Let $f \in L^1(\mathbb{R}^d)$, and $g : \mathbb{R}^d \to \mathbb{C}^d$ be defined through $g(x) = -ixf(x)$. If g is integrable, then $\hat{f} \in C^1(\mathbb{R}^d)$ with $\nabla \hat{f} = \hat{g}$.

- *Plancherel's identity.* Let $f \in L^1(\mathbb{R}^d) \cap L^2(\mathbb{R}^d)$. Then $\hat{f} \in L^2(\mathbb{R}^d)$ and $\|\hat{f}\|_2 = \|f\|_2$.

Plancherel's identity can be used to extend the Fourier transform to L^2. For an arbitrary $f \in L^2(\mathbb{R}^d)$, we can find a sequence $f_j \in L^1(\mathbb{R}^d) \cap L^2(\mathbb{R}^d)$ with $f_j \to f$ in L^2. Since $f_j \in L^1(\mathbb{R}^d)$, we can define the Fourier transform $\hat{f}_j$. From Plancherel's identity, it follows that $\hat{f}_j \in L^2(\mathbb{R}^d)$, with $\|\hat{f}_j\|_2 = \|f_j\|_2$. Moreover

$$\|\hat{f}_j - \hat{f}_m\|_2 = \|\widehat{f_j - f_m}\|_2 = \|f_j - f_m\|_2 \to 0$$

as $j, m \to \infty$. We conclude that $\hat{f}_j$ is a Cauchy sequence in $L^2(\mathbb{R}^d)$ and thus it must converge. We can therefore define

$$\hat{f} = \lim_{j\to\infty} \hat{f}_j. \tag{A.18}$$

It is important to observe that $\hat{f}$ does not depend on the choice of the sequence $f_j \in L^1(\mathbb{R}^d) \cap L^2(\mathbb{R}^d)$. The map $\hat{\cdot} : L^2(\mathbb{R}^d) \to L^2(\mathbb{R}^d)$ is clearly linear and isometric, i.e. $\|\hat{f}\|_2 = \|f\|_2$. It also preserves the inner product, i.e. $\langle \hat{f}; \hat{g} \rangle = \langle f; g \rangle$ for all $f, g \in L^2(\mathbb{R}^d)$. In fact, $\hat{\cdot} : L^2(\mathbb{R}^d) \to L^2(\mathbb{R}^d)$ is also invertible (the inverse is the map $\check{\cdot} : L^2(\mathbb{R}^d) \to L^2(\mathbb{R}^d)$, defined by $\check{f}(k) := \hat{f}(-k)$ for all $f \in L^2(\mathbb{R}^d)$). Thus, the Fourier transform $\hat{\cdot} : L^2(\mathbb{R}^d) \to L^2(\mathbb{R}^d)$ is a unitary map.

So far, we defined the Fourier transform as a linear map from $L^1(\mathbb{R}^d)$ to $L^\infty(\mathbb{R}^d)$ and as a linear map from $L^2(\mathbb{R}^d)$ to $L^2(\mathbb{R}^d)$. By interpolation, it can also be extended as a linear map from $L^p(\mathbb{R}^d)$ to $L^{p'}(\mathbb{R}^d)$, for all $1 < p < 2$, choosing $2 < p' < \infty$ so that $1/p + 1/p' = 1$. The extension is based on the Hausdorff-Young inequality (which replace the Plancherel's identity, for $p < 2$).

Theorem A.3.1. *Let $1 \leq p \leq 2$ and $2 \leq p' \leq \infty$ such that $1/p + 1/p' = 1$. Then there exists a constant $C > 0$ (depending on the dimension n and on p) such that $\hat{f} \in L^{p'}(\mathbb{R}^n)$ and*

$$\|\hat{f}\|_{p'} \leq C\|f\|_p \tag{A.19}$$

for any $f \in L^1(\mathbb{R}^d) \cap L^p(\mathbb{R}^d)$.

A.4 Distributions and Sobolev Spaces

Definition and basic properties. We start with the definition of the space of test functions.

Definition A.4.1. *Let $n \in \mathbb{N}\backslash\{0\}$ and $\Omega \subset \mathbb{R}^n$ be an open non-empty set ($\Omega = \mathbb{R}^n$ is allowed). We denote the space of test functions on Ω by $\mathcal{D}(\Omega)$. It consists of all functions in $C_c^\infty(\Omega)$ (the space of infinitely differentiable functions that are compactly supported inside Ω), equipped with the following notion of convergence. A sequence $\phi_m \in C_c^\infty(\Omega)$ converges in $\mathcal{D}(\Omega)$ to $\phi \in C_c^\infty(\Omega)$ if and only if there exists a fixed compact set $K \subset \Omega$ such that supp $(\phi_m - \phi) \subset K$ for all $m \in \mathbb{N}$ and*

$$D^\alpha \phi_m \to D^\alpha \phi$$

uniformly on K, for all multi-index $\alpha \in \mathbb{N}^n$. In other words,

$$\sup_{x\in K} \left| \frac{\partial^{|\alpha|}\phi_m(x)}{\partial x_1^{\alpha_1} \dots \partial x_n^{\alpha_n}} - \frac{\partial^{|\alpha|}\phi(x)}{\partial x_1^{\alpha_1} \dots \partial x_n^{\alpha_n}} \right| \to 0$$

as $m \to \infty$, for all $\alpha = (\alpha_1, \dots, \alpha_n) \in \mathbb{N}^n$ (with $|\alpha| = \alpha_1 + \dots + \alpha_n$).

Distributions are then continuous linear functionals on $\mathcal{D}(\Omega)$ (since the notion of convergence on $\mathcal{D}(\Omega)$ is very strong, it is very easy for a linear functional on $\mathcal{D}(\Omega)$ to be continuous).

Definition A.4.2. *A distribution T is a continuous linear functional on the space of test functions $\mathcal{D}(\Omega)$. In other words, it is a map $T : \mathcal{D}(\Omega) \to \mathbb{C}$, with*

$$T(\phi_1 + \lambda\phi_2) = T(\phi_1) + \lambda T(\phi_2)$$

for all $\lambda \in \mathbb{C}$, $\phi_1, \phi_2 \in \mathcal{D}(\Omega)$, such that $T(\phi_m) \to T(\phi)$ as $m \to \infty$, whenever $\phi_m \to \phi$ in $\mathcal{D}(\Omega)$. We denote by $\mathcal{D}'(\Omega)$ the space of distributions on Ω. Observe that $\mathcal{D}'(\Omega)$ has the structure of a vector space (distributions can be summed and multiplied with scalars $\lambda \in \mathbb{C}$).

An important observation is that every L^p-function defines a distribution. Since the behavior at infinity is not important, we define local L^p-spaces.

Definition A.4.3. *For $1 \leq p \leq \infty$, we define the space of locally p-th power integrable functions*

$$L^p_{loc}(\Omega) = \{f : \Omega \to \mathbb{C} \text{ measurable such that } \|f\|_{L^p(K)} < \infty \text{ for all } K \subset \Omega \text{ compact}\}$$

Notice that $L^p_{loc}(\Omega)$ is a vector space. In contrast with $L^p(\Omega)$, however, there is no natural norm on $L^p_{loc}(\Omega)$. Notice also that $L^q_{loc}(\Omega) \subset L^p_{loc}(\Omega)$ if $1 \leq p \leq q \leq \infty$. In particular, $L^1_{loc}(\Omega) \supset L^p_{loc}(\Omega) \supset L^p(\Omega)$ for all $1 \leq p \leq \infty$.

Every locally integrable function defines a distribution.

Lemma A.4.4. *Let $\Omega \subset \mathbb{R}^n$ open, non-empty, $f \in L^1_{loc}(\Omega)$. We define $T_f : \mathcal{D}(\Omega) \to \mathbb{C}$ setting*

$$T_f(\phi) = \int_\Omega f\phi \, d\lambda_n(x) \tag{A.20}$$

for all $\phi \in \mathcal{D}(\Omega)$. Then $T_f \in \mathcal{D}'(\Omega)$ is a distribution (it is the distribution associated with f).

While every locally integrable function $f \in L^1_{\text{loc}}(\Omega)$ defines a distribution $T_f \in \mathcal{D}'(\Omega)$, not every distribution in $\mathcal{D}'(\Omega)$ has the form T_f for an $f \in L^1_{\text{loc}}(\Omega)$. An example of a distribution which is not associated with a locally integrable function is the Dirac delta-distribution

$$\delta_x(\phi) = \phi(x) \tag{A.21}$$

for all $\phi \in \mathcal{D}(\Omega)$ and for a $x \in \Omega$.

Distributional derivatives and Sobolev spaces. The derivative of a distribution is defined as follows.

Definition A.4.5. *Let $\Omega \subset \mathbb{R}^n$ be open and non-empty, $T \in \mathcal{D}'(\Omega)$ a distribution. For a multi-index $\alpha = (\alpha_1, \ldots, \alpha_n) \in \mathbb{N}^n$, we define $D^\alpha T \in \mathcal{D}'(\Omega)$ through*

$$D^\alpha T(\phi) = (-1)^{|\alpha|} T(D^\alpha \phi) \tag{A.22}$$

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

Remark: Let $\alpha = (\alpha_1, \ldots, \alpha_n) \in \mathbb{N}^n$ be a multi-index and $f \in C^{|\alpha|}(\Omega)$. In this case, we have $D^\alpha T_f = T_{D^\alpha f}$. In other words, if we identify functions with the corresponding distributions, classical derivatives coincide, when they exist, with distributional derivatives. But, of course, distributional derivatives are much more general, since they can be applied to arbitrary locally integrable functions (for which classical derivatives do not need to exist) and, even more generally, to arbitrary distributions.

We can now define the Sobolev space $W^{m,p}(\Omega)$ as the space of all functions $f \in L^p(\Omega)$ whose distributional derivatives of order up to m (defined through the distribution T_f associated with f) are again functions in $L^p(\Omega)$.

Definition A.4.6. *For $\Omega \subset \mathbb{R}^n$ open, non-empty, $m \in \mathbb{N}$ and $1 \leq p \leq \infty$, we define*

$$W^{m,p}(\Omega) = \{f \in L^p(\Omega) : \forall\, \alpha \in \mathbb{N}^n \text{ with } |\alpha| \leq m \text{ there is } g_\alpha \in L^p(\Omega) \text{ with } D^\alpha T_f = T_{g_\alpha}\}$$

For $f \in W^{m,p}(\Omega)$ and $\alpha \in \mathbb{N}^n$ with $|\alpha| \leq m$, we define $D^\alpha f = g_\alpha \in L^p(\Omega)$ (the functions $D^\alpha f$ are known as the distributional or weak derivatives of f). For $1 \leq p < \infty$, we define

$$\|f\|_{W^{m,p}} = \left[\sum_{\alpha \in \mathbb{N}^n : |\alpha| \leq m} \|D^\alpha f\|_p^p\right]^{1/p}$$

For $p = \infty$, we set

$$\|f\|_{W^{m,\infty}} = \sum_{|\alpha| \leq m} \|D^\alpha f\|_\infty.$$

Then $\|.\|_{W^{m,p}}$ is a norm on $W^{m,p}(\Omega)$ and $(W^{m,p}(\Omega), \|.\|_{W^{m,p}})$ is a complete Banach space.

An important property of functions in the Sobolev space $W^{m,p}(\Omega)$, for $1 \leq p < \infty$, is the fact that they can be approximated by sequences of smooth functions. This is known as the Meyers-Serrin theorem.

Theorem A.4.7. *Let $m \in \mathbb{N}$, $1 \leq p < \infty$, $\Omega \subset \mathbb{R}^n$ open, non-empty. Let $f \in W^{m,p}(\Omega)$. Then there exists a sequence $f_j \in C^\infty(\Omega) \cap W^{m,p}(\Omega)$ such that*

$$\|f - f_j\|_{W^{m,p}(\Omega)} \to 0$$

as $j \to \infty$. If $\Omega = \mathbb{R}^n$, the sequence f_j can be chosen in $C_c^\infty(\mathbb{R}^n)$ (ie. to have compact support).

Remark. If $\Omega \subset \mathbb{R}^n$ has a boundary and $m \in \mathbb{N}\backslash\{0\}$, $C_c^\infty(\Omega)$ is not dense in $W^{m,p}(\Omega)$. The completion of $C_c^\infty(\Omega)$ with respect to the $W^{m,p}(\Omega)$ norm is a subspace of $W^{m,p}(\Omega)$, typically denoted by $W_0^{m,p}(\Omega)$, which is used to solve differential equations with Dirichlet boundary conditions.

The Hilbert spaces $H^m(\Omega)$. Let $\Omega \subset \mathbb{R}^n$ open and non-empty, $m \in \mathbb{N}$. We will use the notation $H^m(\Omega)$ for the space $W^{m,2}(\Omega)$. For $f, g \in H^m(\Omega)$, we define the inner product

$$\langle f; g\rangle_{H^m} = \sum_{\alpha \in \mathbb{N}^n : |\alpha| \leq m} \langle D^\alpha f; D^\alpha g\rangle_2$$

where $\langle .;. \rangle_2$ denotes the usual inner product on $L^2(\Omega)$. Clearly,

$$\|f\|^2_{W^{m,2}} \equiv \|f\|^2_{H^m} = \langle f; f \rangle_{H^m}.$$

Thus, $(H^m(\Omega), \langle .;. \rangle_{H^m})$ is a Hilbert space.

For $\Omega = \mathbb{R}^n$, there is a useful characterisaztion of $H^m(\Omega)$ in terms of Fourier transform.

Theorem A.4.8. *Let $f \in L^2(\mathbb{R}^n)$, and $\hat{f} \in L^2(\mathbb{R}^n)$ be its Fourier transform. Then $f \in H^m(\mathbb{R}^n)$ if and only if the function $k \to |k|^m \hat{f}(k)$ is in $L^2(\mathbb{R}^n)$. In this case, the distributional derivatives of f satisfy $\widehat{D^\alpha f}(k) = i^{|\alpha|} k^\alpha \hat{f}(k)$ for all $\alpha \in \mathbb{N}^n$ with $|\alpha| \leq m$ and therefore*

$$\|f\|^2_{H^m} = \sum_{\alpha \in \mathbb{N}^n : |\alpha| \leq m} \int_{\mathbb{R}^n} |k_1|^{2\alpha_1} |k_2|^{2\alpha_2} \dots |k_n|^{2\alpha_n} |\hat{f}(k)|^2 dk$$

Green's functions of the Laplacian. Let $G_2(x) = (2\pi)^{-1} \log|x - y|$ and, for $n \in \mathbb{N}$, $n \geq 3$,

$$G_n(x) = \frac{1}{(n-2)|S^{n-1}|} \frac{1}{|x|^{n-2}}$$

where $|S^{n-1}| = 2\pi^{n/2}/\Gamma(n/2)$ is the measure of the $(n-1)$-dimensional sphere with radius one. Then, we have, in the sense of distributions,

$$-\Delta G_n = \delta \tag{A.23}$$

where $\delta \in \mathcal{D}'(\mathbb{R}^n)$ is defined through $\delta(\phi) = \phi(0)$. Thus, if $f \in L^1_{\text{loc}}(\mathbb{R}^n)$ such that $y \to G_n(x-y)f(y)$ is integrable, for almost every $x \in \mathbb{R}^n$, then the function $u = G_n * f \in L^1_{\text{loc}}(\mathbb{R}^n)$ solves the Poisson equation $-\Delta u = f$, in the sense of distributions.

A.5 Sobolev Inequalities and Sobolev Embeddings

Sobolev inequalities for gradients. We begin with the standard Sobolev inequality for functions in $H^1(\mathbb{R}^n)$, $n \geq 3$.

Theorem A.5.1. *Let $n \in \mathbb{N}$, $n \geq 3$. Let $f \in H^1(\mathbb{R}^n)$. Then $f \in L^q(\mathbb{R}^n)$, for $q = 2n/(n-2)$. Moreover, there exists a constant C_n (depending only on the dimension n) such that*

$$\|f\|_q \leq C_n \|\nabla f\|_2 \tag{A.24}$$

for all $f \in H^1(\mathbb{R}^n)$.

In dimensions $n = 1, 2$, we cannot bound high L^q norms just with the L^2 norm of ∇f, we also need the L^2 norm of f.

Theorem A.5.2. *One-dimension: There exists a constant $C > 0$ such that*

$$\|f\|_q \leq C \|f\|_{H^1}$$

for all $f \in H^1(\mathbb{R})$ and all $2 \leq q \leq \infty$. Moreover, every $f \in H^1(\mathbb{R})$ has a representative satisfying

$$|f(x) - f(y)| \leq \|\partial f\|_2 |x - y|^{1/2}$$

for all $x, y \in \mathbb{R}$. In other words, H^1-functions in one-dimension are automatically Hölder continuous with exponent $\alpha = 1/2$.

Two-dimensions: For every $2 \leq q < \infty$ there exists $C > 0$ such that

$$\|f\|_q \leq C \|f\|_{H^1}$$

for all $f \in H^1(\mathbb{R}^2)$.

Remark: in two dimensions, the L^∞-norm cannot be bounded by the H^1-norm.

Compact Sobolev embedding. The Sobolev inequality implies that the unit ball in $H^1(\mathbb{R}^n)$ is a subset of $L^q(\mathbb{R}^n)$, if $q \leq 2n/(n-2)$ (or if $q < \infty$ in two dimensions, $q \leq \infty$ in one dimension). When restricting on sets with finite measure, it turns out that the unit ball of H^1 is sequentially compact subset of L^q. This follows from the following theorem.

Theorem A.5.3. *Let $\{f_j\}_{j \in \mathbb{N}}$ be a weakly convergent sequence in $H^1(\mathbb{R}^n)$, let f denote its weak limit. Let moreover $A \subset \mathbb{R}^n$ be a set of finite measure and χ_A the corresponding characteristic function. Then $\chi_A f_j \to \chi_A f$ strongly in $L^q(\mathbb{R}^n)$, for all $1 \leq q < 2n/(n-2)$, if $n \geq 3$, and for all $1 \leq q < \infty$, if $n = 1, 2$. For $n = 1$, the convergence is pointwise and uniform on every compact subset of $\mathbb{R}$.*

General Sobolev inequalities and embedding theorems. Sobolev inequalities can be extended to domains with the cone property, defined as follows.

Definition A.5.4. *For $r > 0$ and $\theta \in [0; 2\pi]$, let*

$$K_{r,\theta} := \{x \in \mathbb{R}^n : x \neq 0, 0 < x_n < |x| \cos\theta\} \cap B_r(0)$$

denote a finite cone of angle θ and length r. We say that an open domain $\Omega \subset \mathbb{R}^n$ has the cone property if there exists r, θ such that, for all $x \in \Omega$ there is a finite cone K_x congruent to $K_{r,\theta}$ which is contained in Ω and has vertex x.

We state now the general form of Sobolev inequality.

Theorem A.5.5. *Let $\Omega \subset \mathbb{R}^n$ be an open set, with the cone property for some r, θ. Let $1 \leq p \leq q \leq \infty$, $m, k \in \mathbb{N}$, with $m \geq 1$ and $k \leq m$. Then there exists a constant $C > 0$ (depending on m, k, q, p, θ, r) such that, for all $f \in W^{m,p}(\Omega)$, we have*

i) If $kp < n$,

$$\|f\|_{W^{m-k,q}(\Omega)} \leq C \|f\|_{W^{m,p}(\Omega)}$$

for all $p \leq q \leq np/(n-kp)$.

ii) If $kp = n$,

$$\|f\|_{W^{m-k,q}(\Omega)} \leq C\|f\|_{W^{m,p}(\Omega)}$$

for all $p \leq q < \infty$.

iii) If $kp > n$,

$$\|f\|_{W^{m-k,q}(\Omega)} \leq C\|f\|_{W^{m,p}(\Omega)}$$

for all $p \leq q \leq \infty$. In particular,

$$\max_{0\leq|\alpha|\leq m-k} \sup_{x\in\Omega} |D^\alpha f(x)| \leq C\|f\|_{W^{m,p}(\Omega)}$$

The next theorem, known as the Rellich-Kondrachov theorem, provides the general statement for compact Sobolev embeddings.

Theorem A.5.6. *Let $\Omega \subset \mathbb{R}^n$ be an open set, with the cone property for some r, θ. Let $1 \leq p \leq q \leq \infty$, $m, k \in \mathbb{N}$, with $m \geq 1$ and $k \leq m$. Let $f \in W^{m,p}(\Omega)$ and $\{f_j\}_{j\in\mathbb{N}}$ be a sequence in $W^{m,p}(\Omega)$ with $f_j \rightharpoonup f$ weakly in $W^{m,p}(\Omega)$ (meaning that $D^\alpha f_j \rightharpoonup D^\alpha f$ weakly in $L^p(\Omega)$, for all multiindex $\alpha \in \mathbb{N}^n$ with $0 \leq |\alpha| \leq m$). Let $\omega \subset \Omega$ be open and bounded (in particular, we can take $\omega = \Omega$, if $\Omega \subset \mathbb{R}^3$ is bounded). Then, we have*

i) If $kp < n$, $1 \leq q < np/(n-kp)$, then $\|f_j - f\|_{W^{m-k,q}(\omega)} \to 0$, as $j \to \infty$.

ii) If $kp = n$, $1 \leq q < \infty$, then $\|f_j - f\|_{W^{m-k,q}(\omega)} \to 0$, as $j \to \infty$.

iii) If $kp > n$, $1 \leq q \leq \infty$, then $\|f_j - f\|_{W^{m-k,q}(\omega)} \to 0$, as $j \to \infty$. In particular, taking $q = \infty$, we conclude that

$$\max_{0\leq|\alpha|\leq m-k} \sup_{x\in\omega} |D^\alpha f_j(x) - D^\alpha f(x)| \to 0$$

as $j \to \infty$.

Poincaré inequality. Finally, let us also mention a useful variant of the Sobolev inequality, known as the Poincaré inequality, which allows us to bound fluctuations of functions in Sobolev spaces.

Theorem A.5.7. *Let $\Omega \subset \mathbb{R}^n$ be a bounded, connected, open set with the cone property for some r, θ. Let $1 \leq p < \infty$, and let $g \in L^{p'}(\Omega)$ with $\int g d\lambda_n = 1$. Then there exists a constant $C = C(\Omega, g, p, q)$ such that*

$$\left\| f - \int_\Omega fg\, d\lambda_n \right\|_{L^q(\Omega)} \leq C\|\nabla f\|_{L^p(\Omega)}$$

for all $f \in W^{1,p}(\Omega)$ and all $1 \leq q < pn/(n-p)$ if $p < n$, all $1 \leq q < \infty$ if $p = n$, and all $1 \leq q \leq \infty$ if $p > n$.