Introduction to Mathematical Physics 2 The Mathematics of Quantum Mechanics

Roderich Tumulka*

Spring 2011

^{*}Department of Mathematics, Rutgers University, 110 Frelinghuysen Road, Piscataway, NJ 08854-8019, USA. E-mail: tumulka@math.rutgers.edu

Contents

1	Overview, the Schrödinger equation	4
2	Hilbert space 2.1 L ² spaces 2.2 Completion 2.3 Orthonormal bases	8 9 11 11
3	More about Hilbert space3.1 Unitaries3.2 Projections3.3 Classification of Hilbert spaces3.4 Bounded operators	15 15 15 18 19
4	Fourier transformation	21
5	The free Schrödinger equation	27
6	Distributional solutions	29
7	Unitary 1-parameter groups and self-adjoint operators7.1Unitary 1-parameter groups	34 36 37 41
8	The spectral theorem for self-adjoint operators8.1The spectral theorem in terms of multiplication operators8.2The spectral theorem in terms of functional calculus8.3The spectral theorem in terms of PVMs	42 43 44 45
9	Positive operators 9.1 Bounded positive operators	51 52
10	The predictive formalism of quantum mechanics 10.1 Tensor product spaces 10.2 POVMs 10.3 The main theorem about POVMs 10.4 The momentum operator and the asymptotics of the Schrödinger equation for large times 10.5 Heisenberg's uncertainty relation 10.6 The quantum measurement problem	53 53 56 56 61 62 63

11 Density operators	65
11.1 Trace of an operator	65
11.2 The trace formula in quantum mechanics	67
11.3 Reduced density operators	69
11.4 Partial trace	70
12 More about the quantum formalism 12.1 Collapse in Bohmian mechanics 12.2 General formulation of the collapse rule	74 74 76
13 Spin and representations of the rotation group SO(3) 13.1 The Pauli equation	78 80

1 Overview, the Schrödinger equation

One of the fundamental laws of quantum mechanics is the Schrödinger equation

$$i\hbar\frac{\partial\psi}{\partial t} = -\sum_{i=1}^{N} \frac{\hbar^2}{2m_i} \nabla_i^2 \psi + V \psi \,. \tag{1.1}$$

We will spent the first few weeks of the course with a discussion of this partial differential equation, and will talk a lot about relevant facts and concepts from functional analysis.

The Schrödinger equation governs the time evolution of the wave function $\psi = \psi_t = \psi(t, \boldsymbol{x}_1, \boldsymbol{x}_2, \dots, \boldsymbol{x}_N)$. It can be expected to be valid only in the non-relativistic regime, i.e., when the speeds of all particles are small compared to the speed of light. In the general case (the relativistic case) it needs to be replaced by other equations, such as the Klein-Gordon equation and the Dirac equation that we will discuss later. We focus first on spinless particles and discuss the phenomenon of spin later. Eq. (1.1) applies to a system of N particles in \mathbb{R}^3 ; a configuration of N particles is a list of their positions; configuration space is thus, for our purposes, the Cartesian product of N copies of physical space, or \mathbb{R}^{3N} . The wave function of quantum mechanics, at any fixed time, is a function on configuration space, either complex-valued or spinor-valued; for spinless particles, it is complex-valued, so

$$\psi: \mathbb{R}_t \times \mathbb{R}_q^{3N} \to \mathbb{C}. \tag{1.2}$$

The subscript indicates the variable: t for time, $q = (\boldsymbol{x}_1, \ldots, \boldsymbol{x}_N)$ for the configuration. Note that i in (1.1) either denotes $\sqrt{-1}$ or labels the particles, $i = 1, \ldots, N$; m_i are positive constants, called the *masses* of the particles; $\hbar = h/2\pi$ is a constant of nature, h is called Planck's quantum of action or Planck's constant, $h = 6.63 \times 10^{-34} \text{ kg m}^2 \text{s}^{-1}$; ∇_i is the derivative operator with respect to the variable $\boldsymbol{x}_i, \nabla_i^2$ the corresponding Laplacian; V is a given function on configuration space, called the *potential energy* or just *potential*.

Fundamentally, the potential in non-relativistic physics is

$$V(\boldsymbol{x}_1, \dots, \boldsymbol{x}_N) = \sum_{1 \le i < j \le N} \frac{e_i e_j}{|\boldsymbol{x}_i - \boldsymbol{x}_j|} - \sum_{1 \le i < j \le N} \frac{Gm_i m_j}{|\boldsymbol{x}_i - \boldsymbol{x}_j|}, \quad (1.3)$$

where $|\mathbf{x}|$ denotes the Euclidean norm in \mathbb{R}^3 , e_i are constants called the *electric charges* of the particles (which can be positive, negative, or zero); the first term is called the *Coulomb potential*, the second term is called the *Newtonian gravity potential*, G is a constant of nature called Newton's constant of gravity $G = 6.67 \times 10^{-11} \text{ kg}^{-1} \text{m}^3 \text{s}^{-2}$, and m_i are again the masses. However, when the Schrödinger equation is regarded as an *effective equation* rather than as a fundamental law of nature then the potential V may contain terms arising from particles outside the system interacting with particles belonging to the system. That is why the Schrödinger equation is often considered for rather arbitrary functions V, also time-dependent ones. The operator

$$H = -\sum_{i=1}^{N} \frac{\hbar^2}{2m_i} \nabla_i^2 + V$$
 (1.4)

is called the Hamiltonian operator.

The Schrödinger equation is meant to determine the time evolution of ψ_t in that for a given initial wave function $\psi_0 = \psi(t = 0) : \mathbb{R}^{3N} \to \mathbb{C}$ it uniquely fixes ψ_t for any $t \in \mathbb{R}$; to confirm this picture, we will have to discuss existence and uniqueness of solutions.

So far I have not said anything about what this new physical object ψ has to do with the particles. One fundamental connection is

Born's rule. The system's configuration at time t is random with probability density

$$\rho_t(q) = \left|\psi_t(q)\right|^2. \tag{1.5}$$

For this rule to make sense, we need that

$$\int_{\mathbb{R}^{3N}} |\psi_t(q)|^2 \, dq = 1. \tag{1.6}$$

And indeed, the Schrödinger equation provides the resources to guarantee this relation; one says that the Schrödinger equation "conserves $|\psi|^2$." More precisely, it implies the continuity equation¹

$$\frac{\partial |\psi(t,q)|^2}{\partial t} = -\sum_{i=1}^N \nabla_i \cdot \boldsymbol{j}_i(t,q) \quad \text{with} \quad \boldsymbol{j}_i(t,q) = \frac{\hbar}{m_i} \text{Im}\Big(\psi^*(t,q)\nabla_i\psi(t,q)\Big)$$
(1.7)

because

$$\frac{\partial}{\partial t} \left(\psi^* \psi \right) = 2 \operatorname{Re} \left(\psi^* \frac{-i}{\hbar} H \psi \right)$$
(1.8)

$$= \frac{2}{\hbar} \operatorname{Im} \left(-\sum_{i=1}^{N} \frac{\hbar^2}{2m_i} \psi^* \nabla_i^2 \psi + \underbrace{V(q) |\psi|^2}_{\text{real}} \right)$$
(1.9)

$$= -\sum_{i=1}^{N} \frac{\hbar}{m_i} \operatorname{Im}\left(\psi^* \,\nabla_i^2 \psi + \underbrace{(\nabla_i \psi^*) \cdot (\nabla_i \psi)}_{\operatorname{real}}\right) = -\sum_{i=1}^{N} \nabla_i \cdot \boldsymbol{j}_i.$$
(1.10)

The continuity equation expresses that $|\psi|^2$ is locally conserved, taking it to flow with the current $(\mathbf{j}_1, \ldots, \mathbf{j}_N)$. Indeed, note that it asserts that the (3N + 1)-dimensional (configuration-space-time) vector field $\mathbf{j} = (|\psi|^2, \mathbf{j}_1, \ldots, \mathbf{j}_N)$ has vanishing divergence. By the Ostrogradski–Gauss integral theorem, the surface integral of a vector field equals the volume integral of its divergence, so the surface integral of a divergence-less vector field vanishes. Let the surface be the boundary of a (3N + 1)-dimensional cylinder $[0,T] \times S$, where $S \subseteq \mathbb{R}^{3N}$ is a ball or any set with smooth boundary ∂S . Then the surface integral of \mathbf{j} is

$$0 = -\int_{S} |\psi_{0}|^{2} + \int_{S} |\psi_{t}|^{2} + \int_{0}^{T} dt \int_{\partial S} dA \, n_{\partial S} \cdot j$$
(1.11)

 $^{{}^{1}}$ I don't know where this name comes from. It has nothing to do with being continuous. It should be called conservation equation.

with $n_{\partial S}$ the unit normal vector field in \mathbb{R}^{3N} on the boundary of S. That is, the amount of $|\psi|^2$ in S at time T differs from the initial amount of $|\psi|^2$ in S by the flux of j across the boundary of S during [0, T]—a conservation law. If (and we will see later that this is indeed the case) there is no flux to infinity, i.e., if the last integral becomes arbitrarily small by taking S to be a sufficiently big ball, then the total amount of $|\psi|^2$ remains constant,

$$\|\psi_T\|^2 = \|\psi_0\|^2 \tag{1.12}$$

with the L^2 norm

$$\|\psi\| = \left(\int_{\mathbb{R}^{3N}} dq \, |\psi(q)|^2\right)^{1/2}.$$
(1.13)

Thus, the Born rule is consistent with the Schrödinger equation, provided the initial datum ψ_0 has norm 1, which we will henceforth assume. The wave function ψ_t will in particular be square-integrable, and this makes the space $L^2(\mathbb{R}^{3N})$ of square-integrable functions a natural arena, even though one would believe that any physical wave function is a smooth function, $\psi \in C^{\infty}(\mathbb{R}^{3N})$, and L^2 contains non-smooth functions. We will soon discuss L^2 in detail.

Still, the Schrödinger equation and the Born rule together don't form a complete formulation of quantum mechanics. In fact, there is no consensus as to what the complete formulation of quantum mechanics should be. The main opposing attitudes could be called the *positivist attitude* and the *realist attitude*. The positivist wants to provide rules that allow us to compute, for any conceivable experiment, the possible outcomes and their respective probabilities. The realist wants to provide a model describing what actually happens. In the realist picture, the positivist rules are theorems; in the positivist framework, they are postulates. Many founding fathers of quantum mechanics, particularly Niels Bohr and his "Copenhagen school" (including Werner Heisenberg and Wolfgang Pauli), claimed from the 1920s onwards that a realist picture of quantum mechanics was impossible—a claim that we know now is false. Disagreement persists as to whether the positivist or the realist attitude is better scientific practice. I now give a brief outline of a positivist and a realist formulation of quantum mechanics.

The simplest realist version of quantum mechanics is *Bohmian mechanics*. According to this theory, every particle has a precise position $Q_i(t) \in \mathbb{R}^3$ at any time t and moves along a trajectory determined by Bohm's equation of motion

$$\frac{d\boldsymbol{Q}_i}{dt} = \frac{\boldsymbol{j}_i}{|\psi|^2}(t, Q(t)) = \frac{\hbar}{m_i} \mathrm{Im} \frac{\nabla_i \psi}{\psi}(t, Q(t))$$
(1.14)

with $Q(t) = (\mathbf{Q}_1(t), \dots, \mathbf{Q}_N(t))$ the configuration at time t. Eq. (1.14) is an ordinary differential equation depending on ψ , the same wave function we talked about before. Since (1.14) is of first order (specifying the velocity rather than the acceleration), the initial configuration Q(0) determines Q(t) for all t; thus, Bohmian mechanics is a deterministic theory. The state of the system at any time t is given by the pair $(Q(t), \psi_t)$. It is further assumed that the configuration Q(t) is random with the Born distribution (1.5). Because of the determinism, this distribution can be *assumed* only for one time, say t = 0; for any other time t, then, the distribution of Q(t) is fixed by (1.14). It is a theorem that the distribution of Q(t) is indeed given by $|\psi_t(q)|^2$. Thus, it is consistent to assume the Born distribution for every t. We will talk more about Bohmian mechanics at a later point. There are also other, inequivalent, realist theories of quantum mechanics, in particular theories of spontaneous wave function collapse and many-worlds theories.

With the positivist attitude, one avoids statements about what electrons actually are and do, and prefers statements about outcomes of experiments (so-called operational statements). Thus, one prefers to formulate Born's rule as: "If, at time t, an observer outside the system performs a position measurement on all particles of the system, then the outcomes will be random with joint distribution density $|\psi_t|^2$."² More generally, the positivist tries to provide a formula for the probability distribution of the outcome of *any* experiment. Such formulas involve the wave function ψ_t as well as operators called *observables*. A problem with this approach is that while a realist model can be specified completely in a few lines, it is very hard, indeed impossible, to formulate the positivist rules in purely operational terms. The rules will usually stay vague, implicit, and incomplete, and attempts to make them more precise also make them excessively complicated. We will study some such rules (though not formulated in purely operational terms) at a later point.

We will first focus on the Schrödinger equation. In order to study existence and uniqueness of solutions, we might focus on particularly nice initial wave functions (viz., smooth functions that quickly tend to zero at infinity), or we might on the contrary try to define the evolution for as big a class of initial wave functions as possible (viz., for tempered distributions), in particular for ψ_0 that are not twice differentiable so that the Laplace operator cannot be taken literally. We will follow both strategies and learn something from both. But we will first consider a class of functions that lies in the middle between these extremes, viz., $L^2(\mathbb{R}^{3N})$.

²The word "measurement" is somewhat inappropriate here because in normal English it suggests that the particle already *had* a position just before *t* which was merely *found out* by the observer—a suggestion that is against the positivist attitude.

2 Hilbert space

Definition 2.1. A *norm* on a complex vector space X is a mapping $\|\cdot\| : X \to [0, \infty)$ such that

$$\|\psi\| \neq 0 \text{ for every } \psi \in X \setminus \{0\}$$

$$(2.1)$$

$$\|c\psi\| = |c| \|\psi\| \text{ for every } c \in \mathbb{C} \text{ and } \psi \in X$$
(2.2)

$$\|\psi + \phi\| \le \|\psi\| + \|\phi\| \text{ for every } \psi, \phi \in X.$$
(2.3)

An inner product $\langle \psi | \phi \rangle$ on a complex vector space X is a mapping $X \times X \to \mathbb{C}$ that is conjugate-symmetric,

$$\langle \phi | \psi \rangle = \langle \psi | \phi \rangle^*, \tag{2.4}$$

sesqui-linear (Latin sesqui = one and a half),

$$\langle \psi | \lambda \phi \rangle = \lambda \langle \psi | \phi \rangle, \quad \langle \psi | \phi_1 + \phi_2 \rangle = \langle \psi | \phi_1 \rangle + \langle \psi | \phi_2 \rangle,$$
 (2.5)

and positive definite,

$$\langle \psi | \psi \rangle > 0 \text{ for } \psi \neq 0.$$
 (2.6)

An inner product always defines a norm by

$$\|\psi\| = \sqrt{\langle \psi | \psi \rangle}.$$
(2.7)

It is straightforward to check (2.1) and (2.2). To prove the triangle inequality (2.3), we first prove the *Cauchy–Schwarz inequality* which asserts that, in any vector space with an inner product,

$$\left|\langle\psi|\phi\rangle\right| \le \|\psi\| \,\|\phi\|\,. \tag{2.8}$$

Proof. If $\phi = 0$ or $\psi = 0$, both sides vanish. Suppose $\psi \neq 0$, set $\tilde{\psi} = \psi/||\psi||$, $\phi_{\parallel} = \langle \tilde{\psi} | \phi \rangle \tilde{\psi}$ and $\phi_{\perp} = \phi - \phi_{\parallel}$. Then $\langle \phi_{\parallel} | \phi_{\perp} \rangle = 0$ and thus $\| \phi \|^2 = \| \phi_{\parallel} \|^2 + \| \phi_{\perp} \|^2 \ge \| \phi_{\parallel} \|^2 = |\langle \tilde{\psi} | \phi \rangle|^2$.

Now the triangle inequality follows:

$$\|\psi + \phi\|^2 = \langle \psi + \phi | \psi + \phi \rangle = \langle \psi | \psi \rangle + \langle \phi | \phi \rangle + 2\operatorname{Re}\langle \psi | \phi \rangle$$
(2.9)

$$\leq \|\psi\|^{2} + \|\phi\|^{2} + 2|\langle\psi|\phi\rangle|^{2} \leq \|\psi\|^{2} + \|\phi\|^{2} + 2\|\psi\|\|\phi\|.$$
(2.10)

It also follows that from the Cauchy–Schwarz inequality that the inner product is continuous: If $\psi_n \to \psi$ then $\langle \psi_n | \phi \rangle \to \langle \psi | \phi \rangle$.

The polarization identity

$$\langle \psi | \phi \rangle = \frac{1}{4} \Big(\|\psi + \phi\|^2 - \|\psi - \phi\|^2 - i\|\psi + i\phi\|^2 + i\|\psi - i\phi\|^2 \Big), \tag{2.11}$$

valid in any vector space with an inner product, allows us to express inner products in terms of norms.

Definition 2.2. A *Hilbert space* is a complex vector space \mathscr{H} equipped with an inner product that is complete with respect to the norm defined by the inner product. A normed vector space is called *complete* if every Cauchy sequence converges; complete normed vector spaces are also called *Banach spaces*.

Example 2.3. \mathbb{C}^n with

$$\langle \psi | \phi \rangle = \sum_{i=1}^{n} \psi_i^* \phi_i \tag{2.12}$$

is a Hilbert space.

2.1 L^2 spaces

Theorem 2.4. For any measure space $(\Omega, \mathfrak{A}, \mu)$ is $L^2(\Omega, \mathfrak{A}, \mu)$ a Hilbert space.

 $L^2(\Omega, \mathfrak{A}, \mu)$ is defined to be the set of all equivalence classes of square-integrable complex-valued functions, i.e., of \mathfrak{A} -measurable functions $f: \Omega \to \mathbb{C}$ such that

$$\int_{\Omega} \mu(dx) |f(x)|^2 < \infty , \qquad (2.13)$$

where f, g are equivalent if $\mu \{ x \in \Omega : f(x) \neq g(x) \} = 0$. The inner product is defined by

$$\langle f|g\rangle = \int_{\Omega} \mu(dx) f(x)^* g(x) \,. \tag{2.14}$$

Before we turn to the proof of Theorem 2.4, let us look at some instances.

• The special case $\Omega = \mathbb{N}$, \mathfrak{A} all subsets, and $\mu(A) = \#A$ is known as the space ℓ^2 of all square-summable complex sequences (x_1, x_2, x_3, \ldots) . Its inner product is

$$\langle x|y\rangle = \sum_{n=1}^{\infty} x_n^* y_n \,. \tag{2.15}$$

• The special case of immediate relevance to the Schrödinger equation is $\Omega = \mathbb{R}^{3N}$, $\mathfrak{A} = \mathfrak{B}(\mathbb{R}^{3N})$ the Borel σ -algebra, and μ the Lebesgue measure; for this we simply write $L^2(\mathbb{R}^{3N})$. Since $|\psi|^2$ plays the role of probability density, the only $\psi \in$ $L^2(\mathbb{R}^{3N})$ that occur physically are those with $||\psi|| = 1$.

To prove Theorem 2.4, we need to show (i) that the integral (2.14) is finite and (ii) that L^2 is complete. Concerning (i), this follows from the *Cauchy–Schwarz inequality* for integrals: For any \mathfrak{A} -measurable functions $f, g: \Omega \to \mathbb{C}$,

$$\int_{\Omega} \mu(dx) |f(x)| |g(x)| \le \left(\int_{\Omega} \mu(dx) |f(x)|^2 \right)^{1/2} \left(\int_{\Omega} \mu(dx) |g(x)|^2 \right)^{1/2},$$
(2.16)

where the integrals may be ∞ . (If ||f|| = 0 and $||g|| = \infty$ then the left hand side is actually 0.) It follows that if both integrals on the right are finite, then also the integral on the left is finite; thus, (2.14) is well defined for any two elements of $L^2(\Omega, \mathfrak{A}, \mu)$.

Proof. of (2.16). If ||f|| = 0 then f = 0 μ -almost everywhere, so fg = 0 μ -almost everywhere, and the left hand side vanishes. If ||f|| > 0 and $||g|| = \infty$, then the right hand side is infinite and the inequality always true. Thus, we can assume $0 < ||f||, ||g|| < \infty$. Since, constant factors can be pulled out, it suffices to consider ||f|| = 1 = ||g||. We can further assume without loss of generality that f(x) and g(x) are real and ≥ 0 for every x.

For any $x \in \Omega$, we have that

$$f(x) g(x) \le \frac{f(x)^2 + g(x)^2}{2}$$
(2.17)

because $0 \leq (f(x) - g(x))^2$. Thus, the left hand side of (2.16) is

$$\int \mu(dx) fg \leq \frac{1}{2} \underbrace{\int \mu(dx) f^2}_{=1} + \frac{1}{2} \underbrace{\int \mu(dx) g^2}_{=1} = 1 = \|f\| \|g\|.$$
(2.18)

_		

Now for completeness:

Theorem 2.5. (Theorem of Riesz and Fischer) For any measure space $(\Omega, \mathfrak{A}, \mu)$ and any $1 \leq p \leq \infty$ is $L^p(\Omega, \mathfrak{A}, \mu)$ complete (and thus a Banach space).

Proof. We need and prove here only the case p = 2. Let $(f_n)_{n \in \mathbb{N}}$ be a Cauchy sequence in L^2 . We need to show that the sequence converges in L^2 , i.e., that there is $f \in L^2$ with

$$\lim_{n \to \infty} \|f_n - f\| = 0.$$
 (2.19)

It suffices to find a convergent subsequence because the convergence of a subsequence of a Cauchy sequence implies the convergence of the entire sequence. Choose a subsequence $(g_k) = (f_{n_k})$ such that $||g_k - g_{k+1}|| \leq 2^{-k}$. We claim that μ -almost everywhere, g_k converges to a function g. To this end, set

$$h_n(x) = \sum_{k=1}^{n-1} |g_k(x) - g_{k+1}(x)|, \quad h(x) := \lim_{n \to \infty} h_n(x) \le \infty.$$
(2.20)

By the triangle inequality (2.3),

$$\|h_n\| \le \sum_{k=1}^{n-1} \|g_k - g_{k+1}\| < 1.$$
(2.21)

By the Lebesgue monotone convergence theorem,

$$1 \ge \lim_{n \to \infty} \int \mu(dx) \, |h_n(x)|^2 = \int \mu(dx) \, \lim_{n \to \infty} |h_n(x)|^2 = \int \mu(dx) \, |h(x)|^2. \tag{2.22}$$

Thus, h(x) is finite μ -almost everywhere. As a consequence,

$$g_n = g_1 + \sum_{k=1}^{n-1} (g_{k+1} - g_k)$$
(2.23)

converges almost everywhere pointwise to a measurable function g. We now apply the dominated convergence theorem; to obtain a dominating function, note that $(a + b)^2 \leq 2a^2 + 2b^2$ for $a, b \in \mathbb{R}$ and thus

$$|g_k|^2 \le (|g_1| + h_n)^2 \le 2|g_1|^2 + 2h_n^2 \le 2|g_1|^2 + 2h^2, \qquad (2.24)$$

which is integrable. The dominated convergence theorem now yields that $g \in L^2$ and $||g_n - g|| \to 0$.

2.2 Completion

Theorem 2.6. A vector space X with inner product can always be completed; *i.e.*, there is a Hilbert space \mathscr{H} and a linear mapping $A : X \to \mathscr{H}$ that is one-to-one and preserves inner products,

$$\langle A\psi | A\phi \rangle_{\mathscr{H}} = \langle \psi | \phi \rangle_X, \tag{2.25}$$

and A(X) is dense in \mathcal{H} .

Proof. In fact, every metric space (X, d) can be completed (i.e., embedded in a complete metric space M as a dense subset). To this end, consider the Cauchy sequences of X, call two sequences (x_n) , (y_n) equivalent if $\lim_{n\to\infty} d(x_n, y_n) = 0$, and let M be the set of equivalence classes. One can show that for any two Cauchy sequences $\lim_{n\to\infty} d(x_n, y_n)$ exists and depends only on the equivalence classes of (x_n) and (y_n) . This limit defines a metric on M, M is complete, and $x \in X$ can be identified with the equivalence class of the constant sequence (x, x, x, \ldots) , a dense subset of M.

If X has an inner product, one can show that for any sequences $(x_n), (y_n)$ in X with $x_n \to x \in M, y_n \to y \in M, \langle x_n | y_n \rangle$ converges. Define $\langle x | y \rangle_M = \lim \langle x_n | y_n \rangle$, which in fact is independent of the choice of the sequences (x_n) and (y_n) ; $\langle \cdot | \cdot \rangle_M$ is an inner product on M whose metric is $d, \sqrt{\langle x - y | x - y \rangle} = d(x, y)$.

2.3 Orthonormal bases

Definition 2.7. A subset B of a Hilbert space \mathcal{H} is called *orthogonal* iff

$$\langle \psi | \phi \rangle = 0 \quad \text{for any } \psi, \phi \in B \text{ with } \psi \neq \phi.$$
 (2.26)

B is called *orthonormal* iff it is orthogonal and $\|\psi\| = 1$ for every $\psi \in B$. That is, $B = \{\phi_i : i \in \mathscr{I}\}$ is an orthonormal set (ONS) iff

$$\langle \phi_i | \phi_j \rangle = \delta_{ij}. \tag{2.27}$$

An orthonormal basis (ONB) or complete orthonormal set of \mathscr{H} is a maximal orthonormal set, i.e., an orthonormal set that is not contained as a proper subset in any other orthonormal set.

Theorem 2.8. Every Hilbert space \mathcal{H} (except $\{0\}$) has an orthonormal basis.

Proof. Let C be the set of all orthonormal sets in \mathscr{H} . We apply Zorn's lemma to C: The inclusion $B_1 \subseteq B_2$ defines a partial order \subseteq on C; C is non-empty because for any $\psi \in \mathscr{H} \setminus \{0\}, \{\psi/\|\psi\|\}$ is an ONS. The hypothesis of Zorn's lemma is that every linearly ordered subset $S \subseteq C$ has an upper bound; this is the case because $\bigcup_{B \in S} B$ is again an ONS containing each $B \in S$. Zorn's lemma then concludes that C has a maximal element, which is an ONS not properly contained in any other ONS.

The name "basis" conveys that every element of \mathscr{H} can be written as a linear combination of basis elements. This is justified by the following theorem.

Theorem 2.9. Let $B = \{\phi_i : i \in \mathscr{I}\}$ be an orthonormal basis of the Hilbert space \mathscr{H} . For each $\psi \in \mathscr{H}$,

$$\psi = \sum_{i \in \mathscr{I}} \langle \phi_i | \psi \rangle \phi_i \quad and \tag{2.28}$$

$$\langle \psi | \chi \rangle = \sum_{i \in \mathscr{I}} c_i^* d_i \quad with \quad c_i = \langle \phi_i | \psi \rangle , \quad d_i = \langle \phi_i | \chi \rangle.$$
(2.29)

Eq. (2.28) means that the sum converges (independent of order) in \mathscr{H} to ψ . If the index set \mathscr{I} is uncountable, then only countably many terms are nonzero. Conversely, if $\sum_{i \in \mathscr{I}} |c_i|^2 < \infty$, $c_i \in \mathbb{C}$, the $\sum_{i \in \mathscr{I}} c_i \phi_i$ converges to an element of \mathscr{H} .

A few remarks before the proof.

- If dim $\mathscr{H} < \infty$ then, for every ONB B, $\#B = \dim \mathscr{H}$; if \mathscr{H} has infinite dimension then B must be an infinite set.
- The sense of "linear combination" employed here is that of a convergent series. Convergent means that

$$\left\|\psi - \sum_{i=1}^{m} c_i \,\phi_i\right\| \xrightarrow{m \to \infty} 0\,. \tag{2.30}$$

- A basis in the sense that linear combinations are convergent series is called a *Schauder basis*, whereas a basis in the sense that linear combinations can only involve finitely many terms is a called a *Hamel basis* or *algebraic basis*. Using Zorn's lemma, one can show that every vector space has a Hamel basis, in much the same way as in the proof of Theorem 2.8: Consider the linearly independent sets (in the sense of finite linear combinations), use Zorn's lemma to show the existence of a maximal linearly independent set, and verify that such a set must be a Hamel basis.
- If \mathscr{I} is uncountable and $\sum_{i \in \mathscr{I}} |c_i|^2 < \infty$, then only countably many c_i can be nonzero because only finitely many $|c_i|$ can be greater than a given $\varepsilon > 0$.

• Fourier series as an example of an ONB: In $L^2([0, 2\pi])$ (where the interval $[0, 2\pi]$ is equipped with its Borel σ -algebra and the Lebesgue measure), the set

$$B = \{\phi_n(x) = (2\pi)^{-1/2} e^{inx} : n \in \mathbb{Z}\}$$
(2.31)

is an orthonormal set, as can be verified through computation. In fact, B is an orthonormal basis; the proof is not easy and will be omitted here. The expansion (2.28) in terms of B is the *Fourier series* of ψ . It may be surprising that $L^2([0, 2\pi])$ has a *countable* ONB in view of the fact that $[0, 2\pi]$ is an *uncountable* set and functions $\psi(x)$ can vary rather arbitrarily; of course, it plays a role that we identified functions differing on a null set and allowed infinite linear combinations.

Proof. of Theorem 2.9. For any finite ONS $\{\phi_1, \ldots, \phi_N\}$,

$$\|\psi\|^{2} = \sum_{n=1}^{N} |\langle \phi_{n} |\psi \rangle|^{2} + \left\|\psi - \sum_{n=1}^{N} \langle \phi_{n} |\psi \rangle \phi_{n}\right\|^{2}.$$
 (2.32)

Indeed, writing ψ as

$$\psi = \sum_{n=1}^{N} \langle \phi_n | \psi \rangle \phi_n + \left(\psi - \sum_{n=1}^{N} \langle \phi_n | \psi \rangle \phi_n \right), \tag{2.33}$$

one easily checks that the two terms are orthogonal, proving (2.32). A consequence is

$$\|\psi\|^2 \ge \sum_{n=1}^N |\langle \phi_n |\psi \rangle|^2, \qquad (2.34)$$

a relation known as *Bessel's inequality*. For the ONB B, we obtain that for any finite subset $\mathscr{I}' \subset \mathscr{I}$,

$$\sum_{i \in \mathscr{I}'} \left| \langle \phi_i | \psi \rangle \right|^2 \le \|\psi\|^2.$$
(2.35)

Thus, $\langle \phi_i | \psi \rangle \neq 0$ for at most a countable number of *i*'s in \mathscr{I} (which we write as $i = 1, 2, 3, \ldots$). Since, as a function of N, $\sum_{i=1}^{N} |\langle \phi_i | \psi \rangle|^2$ is increasing and bounded, it converges to a finite limit as $N \to \infty$. Let

$$\psi_n = \sum_{i=1}^n \langle \phi_i | \psi \rangle \phi_i.$$
(2.36)

Then for n > m,

$$\|\psi_n - \psi_m\|^2 = \left\|\sum_{i=m+1}^n \langle \phi_i | \psi \rangle \phi_i \right\|^2 = \sum_{i=m+1}^n |\langle \phi_i | \psi \rangle|^2.$$
(2.37)

Therefore, (ψ_n) is a Cauchy sequence and converges to an element ψ' of \mathscr{H} . Observe that

$$\langle \phi_m | \psi - \psi' \rangle = \lim_{n \to \infty} \left\langle \phi_m | \psi - \sum_{i=1}^n \langle \phi_i | \psi \rangle \phi_i \right\rangle$$
 (2.38)

$$= \langle \phi_m | \psi \rangle - \langle \phi_m | \psi \rangle = 0.$$
 (2.39)

For $i \in \mathscr{I}$ not among the countably many with $\langle \phi_i | \psi \rangle \neq 0$, we also have $\langle \phi_i | \psi - \psi' \rangle = 0$. Therefore $\psi - \psi'$ is orthogonal to all $\phi_i \in B$. Since B is an ONB, $\psi - \psi' = 0$ or

$$\psi = \lim_{n \to \infty} \sum_{i=1}^{n} \langle \phi_i | \psi \rangle \phi_i \tag{2.40}$$

or (2.28). Furthermore,

$$0 = \lim_{n \to \infty} \left\| \psi - \sum_{i=1}^{n} \langle \phi_i | \psi \rangle \phi_i \right\|^2$$
(2.41)

$$= \lim_{n \to \infty} \left(\|\psi\|^2 - \sum_{i=1}^n \left| \langle \phi_i |\psi \rangle \right|^2 \right)$$
(2.42)

$$= \|\psi\|^2 - \sum_{i \in \mathscr{I}} |\langle \phi_i |\psi \rangle|^2, \qquad (2.43)$$

proving

$$\|\psi\|^2 = \sum_{i \in \mathscr{I}} |\langle \phi_i | \psi \rangle|^2, \qquad (2.44)$$

known as *Parseval's relation* and corresponding to (2.29) for $\chi = \psi$. Using the polarization identity (2.11) expressing inner products in terms of norms, we obtain (2.29). The converse statement is easy to prove.

One sometimes speaks of a generalized orthonormal basis $\{\phi_k\}$ with a continuous parameter k and the intention that $\psi \in \mathscr{H}$ can be expanded according to

$$\psi = \int dk \, c(k) \, \phi_k \,. \tag{2.45}$$

For example, for the generalized orthonormal basis of the "momentum representation" one considers $\psi \in L^2(\mathbb{R}^d)$, $k \in \mathbb{R}^d$, $\phi_k(x) = e^{ik \cdot x}$ and c(k) the Fourier transform of ψ . However, the plane waves ϕ_k are not themselves square-integrable. We will come back to this concept later. It should be distinguished from an uncountable orthonormal basis in the sense of Definition 2.7.

3 More about Hilbert space

3.1 Unitaries

The "isomorphisms" of Hilbert spaces are called unitaries:

Definition 3.1. A linear mapping $U : \mathscr{H}_1 \to \mathscr{H}_2$ is called *unitary* iff it is bijective and preserves inner products,

$$\langle U\psi|U\phi\rangle = \langle \psi|\phi\rangle \quad \forall \psi, \phi \in \mathscr{H}_1.$$
(3.1)

By the polarization identity, the last condition can be replaced with preserving norms (= being isometric),

$$\|U\psi\| = \|\psi\| \quad \forall \psi \in \mathscr{H}_1.$$
(3.2)

The condition "bijective" can be replaced with "surjective" because any U preserving norms is injective. In case $\mathscr{H}_2 = \mathscr{H}_1 = \mathscr{H}$, U is called a *unitary operator*.

Since the Schrödinger equation entails the conservation of $|\psi|^2$, we expect that for initial data ψ_0 with $||\psi_0|| = 1$ also ψ_t has norm 1; by linearity, the evolution from time 0 to time t (if unique) preserves norms. At this point, however, it is not clear whether the evolution mapping $\psi_0 \mapsto \psi_t$ is surjective, and not even whether it is defined on all of $L^2(\mathbb{R}^{3N})$.

Unitaries can be used to define the notion of a generalized orthonormal basis in \mathscr{H} as a unitary $U : \mathscr{H} \to L^2(\Omega, \mathfrak{A}, \mu)$. This allows us to represent every ψ as a function $f(x) = U\psi(x)$ on the set Ω , with f(x) playing the role of the expansion coefficients c_i or c(k). For the momentum representation mentioned before, U corresponds to the Fourier transformation; we will see later that Fourier transformation indeed defines a unitary $U : L^2(\mathbb{R}^d) \to L^2(\mathbb{R}^d)$.

3.2 Projections

The closed subspaces of a Hilbert space \mathscr{H} are themselves Hilbert spaces (with the same inner product $\langle \cdot | \cdot \rangle$). In contrast, if a subspace $X \subset \mathscr{H}$ is not closed then it is a vector space with an inner product but not a Hilbert space. If \mathscr{H} is finite-dimensional, then all subspaces are closed. But not so if dim $\mathscr{H} = \infty$: For example, consider $\mathscr{H} = \ell^2$ and X the set of all sequences in which only finitely many terms are non-zero,

$$X = \bigcup_{n=1}^{\infty} \left\{ (x_1, \dots, x_n, 0, 0, 0, \dots) : x_1, \dots, x_n \in \mathbb{C} \right\}.$$
 (3.3)

Since X is closed under addition and scalar multiplication, it is a subspace. Its closure, however, is ℓ^2 and thus strictly bigger. Indeed, for any element $\psi = (x_1, x_2, ...)$ of ℓ^2 and any $\varepsilon > 0$, there is a $\phi \in X$ with $\|\phi - \psi\| < \varepsilon$; simply set $\phi = (x_1, ..., x_n, 0, 0, ...)$ with n so large that

$$\|\psi\|^{2} = \sum_{i=1}^{\infty} |x_{i}|^{2} < \sum_{i=1}^{n} |x_{i}|^{2} + \varepsilon^{2}, \qquad (3.4)$$

which exists because the series converges. Then $\|\phi - \psi\|^2 = \sum_{i=n+1}^{\infty} |x_i|^2 < \varepsilon^2$.

Proposition 3.2. For any set $S \subseteq \mathcal{H}$, its orthogonal complement

$$S^{\perp} = \{ \psi \in \mathscr{H} : \langle \psi | \phi \rangle = 0 \,\forall \phi \in S \}$$
(3.5)

is a closed subspace of \mathcal{H} .

Proof. With ψ_1 and $\psi_2 \in S^{\perp}$, also $c\psi_1 + \psi_2 \in S^{\perp}$ for any $c \in \mathbb{C}$; so S^{\perp} is a subspace. Now suppose $\psi_1, \psi_2, \ldots \in S^{\perp}$ and $\psi_n \to \psi$; then, for any $\phi \in S$, $0 = \langle \psi_n | \phi \rangle \to \langle \psi | \phi \rangle$, so $\psi \in S^{\perp}$.

Theorem 3.3. (Projection theorem) Let $X \subset \mathscr{H}$ be a closed subspace. Then every $\psi \in \mathscr{H}$ can be decomposed in a unique way as $\psi = \phi + \chi$ with $\phi \in X$ and $\chi \in X^{\perp}$.

Proof. Existence. Let $\psi \in \mathscr{H}$. We first show that there is a $\phi \in X$ that is closest to ψ . Let $d = \inf_{f \in X} \|\psi - f\|$. Choose a sequence $f_n \in X$ so that $\|\psi - f_n\| \to d$. We show that (f_n) is a Cauchy sequence. For this we use the *parallelogram law*

$$||u+v||^{2} + ||u-v||^{2} = 2||u||^{2} + 2||v||^{2}, \qquad (3.6)$$

which holds in any vector space with inner product (easy to check). So

$$||f_n - f_m||^2 = ||(f_n - \psi) - (f_m - \psi)||^2$$
(3.7)

$$= 2\|f_n - \psi\|^2 + 2\|f_m - \psi\|^2 - \|(f_n - \psi) + (f_m - \psi)\|^2$$
(3.8)

$$= 2\|f_n - \psi\|^2 + 2\|f_m - \psi\|^2 - 4\left\|-\psi + \frac{1}{2}(f_n + f_m)\right\|^2$$
(3.9)

$$\leq 2\|f_n - \psi\|^2 + 2\|f_m - \psi\|^2 - 4d^2 \tag{3.10}$$

$$\xrightarrow{m,n \to \infty} 2d^2 + 2d^2 - 4d^2 = 0. \tag{3.11}$$

Thus, (f_n) is a Cauchy sequence, so it converges to f, and $f \in X$. (This step would fail if X were not closed.) It follows that $\|\psi - f\| = d$.

Now set $\phi = f$ and $\chi = \psi - f$. Then $\psi = \phi + \chi$, $\phi \in X$, and it remains to show that $\chi \in X^{\perp}$. For any $g \in X$ and any $t \in \mathbb{R}$,

$$d^{2} \leq \left\|\psi - (f + tg)\right\|^{2} = \|\chi - tg\|^{2}$$
(3.12)

$$= \underbrace{\|\chi\|^2}_{=d^2} + t^2 \|g\|^2 - 2t \operatorname{Re} \langle \chi |g \rangle, \qquad (3.13)$$

so $0 \le t^2 ||g||^2 - 2\operatorname{Re} \langle \chi | g \rangle$ for all $t \in \mathbb{R}$, which implies $\operatorname{Re} \langle \chi | g \rangle = 0$. A similar argument using ti instead of t shows that $\operatorname{Im} \langle \chi | g \rangle = 0$. This completes the proof of existence.

Uniqueness. If $\phi + \chi = \psi = \phi' + \chi'$ with $\phi, \phi' \in X$ and $\chi, \chi' \in X^{\perp}$ then set $\Delta \phi = \phi - \phi' \in X$, $\Delta \chi = \chi - \chi' \in X^{\perp}$, note $0 = \Delta \phi + \Delta \chi$ and thus

$$0 = \langle \Delta \chi | 0 \rangle = \langle \Delta \chi | \Delta \phi + \Delta \chi \rangle = \underbrace{\langle \Delta \chi | \Delta \phi \rangle}_{0} + \| \Delta \chi \|^{2}$$
(3.14)

so $\Delta \chi = 0$; as a consequence, $\Delta \phi = 0$.

For two Hilbert spaces $\mathscr{H}_1, \mathscr{H}_2$, their *direct sum* or *orthogonal sum* $\mathscr{H}_1 \oplus \mathscr{H}_2$ is the Cartesian product $\mathscr{H}_1 \times \mathscr{H}_2$, equipped with the componentwise addition and scalar multiplication (i.e., the direct sum of vector spaces) and the inner product

$$\langle (\psi_1, \psi_2) | (\phi_1, \phi_2) \rangle_{\mathscr{H}_1 \oplus \mathscr{H}_2} = \langle \psi_1 | \phi_1 \rangle_{\mathscr{H}_1} + \langle \psi_2 | \phi_2 \rangle_{\mathscr{H}_2}, \qquad (3.15)$$

which implies $\|(\psi_1, \psi_2)\| = \sqrt{\|\psi_1\|^2 + \|\psi_2\|^2}$. One easily checks that $\mathscr{H}_1 \oplus \mathscr{H}_2$ is again a Hilbert space.

The projection theorem provides a canonical unitary isomorphism $\mathscr{H} \to X \oplus X^{\perp}$, $\psi \mapsto (\phi, \chi)$. It is common to neglect the difference between \mathscr{H} and $X \oplus X^{\perp}$ in the notation and write $\mathscr{H} = X \oplus X^{\perp}$.

Definition 3.4. For any closed subspace $X \subseteq \mathscr{H}$, the mapping $\psi \mapsto \phi$ is the orthogonal projection (or simply projection) to X and defines a linear operator $P_X : \mathscr{H} \to X$ (or, if we wish, $P_X : \mathscr{H} \to \mathscr{H}$). It has the properties

$$P_X^2 = P_X \tag{3.16}$$

and

$$\langle P_X u | v \rangle = \langle u | P_X v \rangle. \tag{3.17}$$

(This is easily visible from $\mathscr{H} = X \oplus X^{\perp}$.)

Corollary 3.5. It also follows from the projection theorem that if X is a closed subspace then $(X^{\perp})^{\perp} = X$. More generally, for any set $X \subseteq \mathscr{H}$, $(X^{\perp})^{\perp}$ is the smallest closed subspace containing X (i.e., the closure of the linear hull of X, $\overline{\operatorname{span}} X$).

Proof. of the second statement: $(\overline{\operatorname{span}} X)^{\perp} = X^{\perp}$, so $X^{\perp \perp} = (\overline{\operatorname{span}} X)^{\perp \perp} = \overline{\operatorname{span}} X$. \Box

Proposition 3.6. Every ONS $B \subset \mathcal{H}$ is an ONB of span B.

Proof. Clearly, $B = \{\phi_i : i \in \mathscr{I}\}$ is an ONS also in the Hilbert space $X = \overline{\text{span}} B$. If B were not maximal, then there would exist a unit vector $\psi \in X$ with $\langle \phi_j | \psi \rangle = 0$ for every $j \in \mathscr{I}$. Since

$$\psi = \sum_{i \in \mathscr{I}'} c_i \phi_i \tag{3.18}$$

with some countable set $\mathscr{I}' \subseteq \mathscr{I}$,

$$\langle \phi_j | \psi \rangle = \sum_{i \in \mathscr{I}'} c_i \langle \phi_j | \phi_i \rangle = c_j$$
 (3.19)

for every $j \in \mathscr{I}'$; thus, all $c_j = 0$ and $\psi = 0$, in contradiction to $\|\psi\| = 1$.

3.3 Classification of Hilbert spaces

The following theorem provides the classification of all Hilbert spaces (modulo unitary equivalence). There is exactly one Hilbert space for every cardinality.

Theorem 3.7. Let B_1 be an ONB of \mathscr{H}_1 and B_2 an ONB of \mathscr{H}_2 . There is a unitary isomorphism $\mathscr{H}_1 \to \mathscr{H}_2$ if and only if B_1 and B_2 have equal cardinality. Moreover, for any set S, an example of a Hilbert space with an ONB of the same cardinality as S is provided by $L^2(S, all subsets, \#)$ (with # the counting measure).

Lemma 3.8. All ONBs of a Hilbert space have the same cardinality.

Proof. Omitted. See N. Dunford and J. Schwartz, Linear Operators vol. 1, page 253.

Proof. of Theorem 3.7. Suppose $B_1 = \{\phi_i^{(1)} : i \in \mathscr{I}_1\}$ and $B_2 = \{\phi_i^{(2)} : i \in \mathscr{I}_2\}$ have equal cardinality, i.e., there is a bijection $\varphi : B_1 \to B_2$. Any $\psi \in \mathscr{H}_1$ can, by Theorem 2.9, be written as

$$\psi = \sum_{i \in \mathscr{I}_1} c_i \phi_i^{(1)}. \tag{3.20}$$

Define $U: \mathscr{H}_1 \to \mathscr{H}_2$ by

$$U\psi = \sum_{i \in \mathscr{I}_1} c_i \phi_i^{(2)}, \qquad (3.21)$$

which exists by Theorem 2.9. Define $V : \mathscr{H}_2 \to \mathscr{H}_1$ in the analogous way and observe $UV = I_{\mathscr{H}_2}$ and $VU = I_{\mathscr{H}_1}$, which shows that U is surjective. The preservation of inner products follows from (2.29). Thus, U is unitary.

Conversely, if $U : \mathscr{H}_1 \to \mathscr{H}_2$ is unitary then UB_1 is an ONB of \mathscr{H}_2 and by Lemma 3.8 has the same cardinality as B_2 .

For $L^2(S, \#)$, $B = \{\phi_s : s \in S\}$ with $\phi_s(x) = 1$ if x = s and $\phi_s(x) = 0$ if $x \neq s$ is an ONB.

Hilbert spaces whose ONBs are uncountable are rarely considered in quantum physics. A Hilbert space is called *separable* if its ONBs are either finite or countably infinite. So $L^2(\mathbb{R}^d)$ is separable. More generally, a metric space is said to be *separable* if there is a dense countable subset. To see that for Hilbert spaces these two definitions are equivalent, note first that if \mathscr{H} has a finite or countable ONB $\{\phi_n\}$ then the countable set

$$\left\{\sum_{n=1}^{N} c_n \phi_n \middle| N \in \mathbb{N}, c_n \in \mathbb{Q} + i\mathbb{Q}\right\}$$
(3.22)

is dense in \mathscr{H} . Conversely, if the sequence $(\tilde{\phi}_n)_{n\in\mathbb{N}}$ is dense in \mathscr{H} then dilute it to a linearly independent sequence $(\phi_n)_{n\in\mathbb{N}}$ with $\operatorname{span}\{\phi_n : n \in \mathbb{N}\} = \operatorname{span}\{\tilde{\phi}_n : n \in \mathbb{N}\}$. Then apply the Gram–Schmidt procedure of orthonormalization.

3.4 Bounded operators

Definition 3.9. A linear operator $L: X \to Y$ between normed vector spaces X and Y is called *bounded* iff there is $C < \infty$ with

$$\|L\psi\|_Y \le C \|\psi\|_X \quad \forall \psi \in X. \tag{3.23}$$

The operator norm of L is defined by

$$||L|| = \sup_{\|\psi\|_X = 1} ||L\psi||_Y.$$
(3.24)

That is, ||L|| is the smallest possible constant C in (3.23)

Example 3.10. Suppose $X = Y = \mathscr{H}$ with dim $\mathscr{H} = n < \infty$, so L can be regarded as an $n \times n$ matrix. Suppose further that L is diagonalizable with eigenvalues $\lambda_1, \ldots, \lambda_n \in \mathbb{C}$. Then $\|L\| = \max\{|\lambda_1|, \ldots, |\lambda_n|\}$.

Example 3.11. Projections P are bounded operators with $X = Y = \mathcal{H}$ and ||P|| = 1 (except for P = 0, which can be regarded as the projection to $\{0\}$).

Theorem 3.12. Let X, Y be normed vector spaces, $L : X \to Y$ linear. The following statements are equivalent:

(i) L is continuous at θ .

 $L\phi_n \not\rightarrow 0$, in contradiction to continuity at 0.

- (ii) L is continuous.
- (*iii*) L is bounded.

Proof. (iii) \Rightarrow (i): If $\|\psi_n\| \to 0$ then $\|L\psi_n\| \le \|L\| \|\psi_n\| \to 0$. (i) \Rightarrow (ii): Suppose $\|\psi_n - \psi\| \to 0$ and L is continuous at 0. Then $\|L\psi_n - L\psi\| = \|L(\psi_n - \psi)\| \to 0$. (ii) \Rightarrow (iii): Suppose L was not bounded. Then there is a sequence $\psi_n \in X$ with $\|\psi_n\| = 1$ and $\|L\psi_n\| \ge n$. Then $\phi_n := \psi_n / \|L\psi_n\|$ converges to 0 but $\|L\phi_n\| = 1$, so

Theorem 3.13. Let X be a normed space, Y a Banach space, $Z \subset X$ a dense subspace and $L: Z \to Y$ bounded linear. Then L possesses a unique bounded linear continuation $\tilde{L}: X \to Y$ with $\tilde{L}|_Z = L$ and $\|\tilde{L}\| = \|L\|$.

Proof. Let $x \in X$. By hypothesis there is a sequence $z_n \in Z$ with $||z_n - x||_X \to 0$. Since z_n converges, it is in particular a Cauchy sequence; because of $||Lz_n - Lz_m||_Y = ||L(z_n - z_m)||_Y \le ||L|| ||z_n - z_m||_X$ we also have that (Lz_n) is a Cauchy sequence in Y and thus converges, $Lz_n \to y \in Y$. Here, y does not depend on the choice of z_n (only of x): If (z'_n) is another sequence in Z with $||z'_n - x||_X \to 0$ then also the sequence $z_1, z'_1, z_2, z'_2, \ldots$ converges to x and, by the above argument, $Lz_1, Lz'_1, Lz_2, Lz'_2, \ldots$ converges to $\tilde{y} \in Y$. Since any subsequence must have the same limit, $y = \tilde{y}$. So we can set $\tilde{L}x := y$. By construction, \tilde{L} is linear. It is bounded since

$$\|\tilde{L}x\|_{Y} = \lim_{n \to \infty} \|Lz_{n}\|_{Y} \le \lim_{n \to \infty} \|L\| \|z_{n}\|_{X} = \|L\| \|x\|_{X}.$$
(3.25)

As a consequence, \tilde{L} is continuous, and a continuous mapping is uniquely determined by its restriction to a dense subset.

Many relevant operators in quantum mechanics are *not* bounded. The Coulomb potential V = -1/r is not bounded, the Laplacian $-\nabla^2$ is not bounded. If $\psi \in L^2(\mathbb{R}^d)$ then $V\psi$ is not necessarily square-integrable; likewise, if $\psi \in C^2(\mathbb{R}^d) \cap L^2(\mathbb{R}^d)$ then $-\nabla^2 \psi$ is not necessarily square-integrable. We thus describe *unbounded operators* as a *pair* (A, \mathscr{D}) , where $\mathscr{D} \subseteq \mathscr{H}$ is a subspace (usually a dense subspace), and $A : \mathscr{D} \to \mathscr{H}$ is a linear mapping; \mathscr{D} is called the *domain* of A. Unlike bounded operators, A cannot be continued in a natural way to $\tilde{A} : \mathscr{H} \to \mathscr{H}$. (However, if a linear mapping $R : \mathscr{H} \to \mathscr{H}$ is given, it may well be an unbounded operator; after all, \mathscr{H} has a Hamel basis, say $\{u_i : i \in \mathscr{I}\}$ (not orthonormal!) and R can be defined by choosing arbitrary $v_i \in \mathscr{H}$ for $i \in \mathscr{I}$ and setting $Ru_i = v_i$.)

4 Fourier transformation

The Schrödinger equation with V = 0 is called the *free* Schrödinger equation. We look for solutions $\psi : \mathbb{R}_t \times \mathbb{R}_x^d \to \mathbb{C}$ of the PDE

$$i\hbar\frac{\partial\psi}{\partial t} = -\sum_{i=1}^{d} \frac{\hbar^2}{2m_i} \frac{\partial^2\psi}{\partial x_i^2}.$$
(4.1)

(If x_1, x_2, x_3 are the coordinates of particle 1 then $m_1 = m_2 = m_3$ should be the mass of that particle. Our notation allows to consider a dimension $\neq 3$ of physical space, and does not specify the number of particles.)

By separation of variables, one obtains the special solutions

$$\psi_k(t,x) = e^{ik \cdot x} e^{-i\omega(k)t} \quad \text{with} \quad \omega(k) = \sum_{i=1}^d \frac{\hbar k_i^2}{2m_i}, \qquad (4.2)$$

the plane waves with wave vector $k \in \mathbb{R}^d$, starting from the initial plane wave

$$\phi_k(x) = e^{ik \cdot x} \,. \tag{4.3}$$

Since $|\psi_k(t,x)|^2 = 1$, $||\psi_k(t,\cdot)|| = \infty$. But the linearity of the Schrödinger equation suggests that if a (square-integrable) initial wave function $\psi(t = 0, \cdot)$ can be expressed as a "continuous linear combination"

$$\psi(0,x) = \int_{\mathbb{R}^d} \hat{\psi}(k) \, e^{ik \cdot x} \, dk \tag{4.4}$$

then $\psi(t, \cdot)$ might be expressed as

$$\psi(t,x) = \int_{\mathbb{R}^d} \hat{\psi}(k) \, e^{ik \cdot x} e^{-i\omega(k) \, t} dk \,. \tag{4.5}$$

In other words, we want to use Fourier transformation.

Definition 4.1. For $f \in L^1(\mathbb{R}^d)$, the Fourier transform of f is

$$\hat{f}(k) = (\mathscr{F}f)(k) := \frac{1}{(2\pi)^{d/2}} \int_{\mathbb{R}^d} f(x) e^{-ik \cdot x} dx$$
 (4.6)

and the *inverse Fourier transform* of f is

$$\check{f}(k) = (\mathscr{F}^{-1}f)(k) := \frac{1}{(2\pi)^{d/2}} \int_{\mathbb{R}^d} f(x) e^{ik \cdot x} dx \,. \tag{4.7}$$

(The name "inverse" and the notation \mathscr{F}^{-1} will be justified later.)

Lemma 4.2. Let $\Gamma \subset \mathbb{R}$ be an open interval and $f : \mathbb{R}^d \times \Gamma \to \mathbb{C}$ be such that $f(\cdot, \gamma) \in L^1(\mathbb{R}^d)$ for every $\gamma \in \Gamma$. Set $I(\gamma) = \int_{\mathbb{R}^d} f(x, \gamma) dx$.

- (i) If $\gamma \mapsto f(x, \gamma)$ is continuous for almost every $x \in \mathbb{R}^d$, and if there is a function $g \in L^1(\mathbb{R}^d)$ with $\sup_{\gamma \in \Gamma} |f(x, \gamma)| \leq g(x)$ for almost every $x \in \mathbb{R}^d$, then $I(\cdot)$ is continuous.
- (ii) If $\gamma \mapsto f(x, \gamma)$ is continuously differentiable for almost every $x \in \mathbb{R}^d$, and if there is a function $g \in L^1(\mathbb{R}^d)$ with $\sup_{\gamma \in \Gamma} |\partial_{\gamma} f(x, \gamma)| \leq g(x)$ for almost every $x \in \mathbb{R}^d$, then $I(\cdot)$ is continuously differentiable and

$$\frac{dI}{d\gamma} = \frac{d}{d\gamma} \int_{\mathbb{R}^d} f(x,\gamma) \, dx = \int_{\mathbb{R}^d} \frac{\partial}{\partial\gamma} f(x,\gamma) \, dx \,. \tag{4.8}$$

Proof. Both statements follow straightforwardly from Lebesgue's dominated convergence theorem, which asserts that if f_n is a sequence of real-valued measurable functions on a measure space $(\Omega, \mathfrak{A}, \mu)$ that converges pointwise almost everywhere to the measurable function f, and if there is $g \in L^1(\Omega, \mathfrak{A}, \mu)$ with $|f_n(x)| \leq g(x)$ for all $n \in \mathbb{N}$ and almost every $x \in \Omega$, then $\lim_{n \to \infty} \int_{\Omega} f_n(x) \, \mu(dx) = \int_{\Omega} f(x) \, \mu(dx)$.

We will first study Fourier transformation on particularly nice functions.

Definition 4.3. A multi-index for \mathbb{R}^d is $\alpha = (\alpha_1, \ldots, \alpha_d) \in \mathbb{N}_0^d$. We write $|\alpha| = \alpha_1 + \ldots + \alpha_d$ and, for $x \in \mathbb{R}^d$,

$$x^{\alpha} = x_1^{\alpha_1} x_2^{\alpha_2} \cdots x_d^{\alpha_d}$$
 and $\partial_x^{\alpha} = \frac{\partial^{|\alpha|}}{\partial x_1^{\alpha_1} \cdots \partial x_d^{\alpha_d}}.$ (4.9)

Definition 4.4. The Schwartz space or space of rapidly decreasing functions or space of Schwartz functions $\mathscr{S}(\mathbb{R}^d) \subset C^{\infty}(\mathbb{R}^d)$, named after Laurent Schwartz (1915–2002), is the set of those $f \in C^{\infty}(\mathbb{R}^d)$ with

$$\|f\|_{\alpha,\beta} := \|x^{\alpha}\partial_x^{\beta}f(x)\|_{\infty} < \infty$$
(4.10)

for all multi-indices $\alpha, \beta \in \mathbb{N}_0^d$. (Here, $\|f\|_{\infty} = \sup_{x \in \mathbb{R}^d} |f(x)|$.)

The functions in \mathscr{S} decrease at $|x| \to \infty$ faster than 1/P(x) for any polynomial P(x), and so do all of their partial derivatives. For example, the Gauss function $f(x) = \exp(-x^T C x)$ with any symmetric $d \times d$ matrix C with positive eigenvalues lies in $\mathscr{S}(\mathbb{R}^d)$. Note that $\mathscr{S}(\mathbb{R}^d) \subset L^1(\mathbb{R}^d)$ and that for $f \in \mathscr{S}$ also $x^{\alpha} \partial_x^{\beta} f \in \mathscr{S}$ for any $\alpha, \beta \in \mathbb{N}_0^d$.

Schwartz functions are square-integrable; thus, $\mathscr{S}(\mathbb{R}^d)$ can be regarded as a subspace of $L^2(\mathbb{R}^d)$: When a continuous function is square-integrable, then its equivalence class (modulo equality almost everywhere) contains only one continuous representative. It is thus common not to distinguish in the notation between the function (if it is continuous!) and its equivalence class. In this sense, one can also identify spaces of continuous functions with subspaces of L^2 (while spaces like $L^p \cap L^2$, $1 \leq p \leq \infty$, can be formed directly).

Definition 4.5. We say $f_n \to f$ in \mathscr{S} iff $||f_n - f||_{\alpha,\beta} \to 0$ for all $\alpha, \beta \in \mathbb{N}_0^d$.

With this notion of convergence, \mathscr{S} should no longer be identified with a subspace of $L^2(\mathbb{R}^d)$ —it has a different topology.

Proposition 4.6. Convergence in \mathscr{S} is equivalent to convergence with respect to the metric

$$d_{\mathscr{S}}(f,g) = \sum_{n=0}^{\infty} 2^{-n} \max_{|\alpha|+|\beta|=n} \left(\frac{\|f-g\|_{\alpha,\beta}}{1+\|f-g\|_{\alpha,\beta}} \right) \le 2.$$
(4.11)

Proof. We show that $d_{\mathscr{S}}$ is a metric; then the equivalence of the two types of convergence is straightforward. Positivity, $d_{\mathscr{S}}(f,g) \geq 0$, and symmetry, $d_{\mathscr{S}}(f,g) = d_{\mathscr{S}}(g,f)$, are obvious. So is definiteness, as $d_{\mathscr{S}}(f,g) = 0$ implies $0 = ||f - g||_{0,0} = ||f - g||_{\infty}$ and thus f = g. The triangle inequality follows from the facts that $|| \cdot ||_{\alpha,\beta}$ satisfies a triangle inequality, that h(x) = x/(1+x) is an increasing function for $x \geq 0$, and that $h(x+y) \leq h(x) + h(y)$.

We note that the metric space $(\mathscr{S}, d_{\mathscr{S}})$ is, in fact, complete; for the proof see, e.g., M. Reed and B. Simon, *Methods of Modern Mathematical Physics vol.* 1, page 133, Theorem V.9.

Lemma 4.7.
$$\mathscr{F}$$
 and \mathscr{F}^{-1} are continuous mappings $\mathscr{S} \to \mathscr{S}$ with
 $\widehat{xf}(k) = i\nabla_k \widehat{f}(k)$ and $\widehat{\nabla_x f}(k) = ik\widehat{f}(k)$, (4.12)

or, more generally,

$$\left((ik)^{\alpha}\partial_{k}^{\beta}\mathscr{F}f\right)(k) = \left(\mathscr{F}\partial_{x}^{\alpha}(-ix)^{\beta}f\right)(k).$$

$$(4.13)$$

Proof. By Lemma 4.2, for $f \in \mathscr{S}$,

$$(2\pi)^{d/2} \Big((ik)^{\alpha} \partial_k^{\beta} \mathscr{F} f \Big)(k) = \int_{\mathbb{R}^d} (ik)^{\alpha} \partial_k^{\beta} e^{-ik \cdot x} f(x) \, dx \tag{4.14}$$

$$= \int_{\mathbb{R}^d} (ik)^{\alpha} (-ix)^{\beta} e^{-ik \cdot x} f(x) \, dx \tag{4.15}$$

$$= \int_{\mathbb{R}^d} (-1)^{|\alpha|} (\partial_x^{\alpha} e^{-ik \cdot x}) (-ix)^{\beta} f(x) \, dx \qquad (4.16)$$

$$= \int_{\mathbb{R}^d} e^{-ik \cdot x} \partial_x^{\alpha} \left((-ix)^{\beta} f(x) \right) dx \tag{4.17}$$

$$= (2\pi)^{d/2} \Big(\mathscr{F} \partial_x^{\alpha} (-ix)^{\beta} f \Big)(k) \,. \tag{4.18}$$

This shows that \mathscr{F} maps \mathscr{S} to $C^{\infty}(\mathbb{R}^d)$, and that (4.13) holds. As a consequence,

$$\|\hat{f}\|_{\alpha,\beta} = \left\|k^{\alpha}\partial_{k}^{\beta}\hat{f}\right\|_{\infty} \le (2\pi)^{-d/2} \int_{\mathbb{R}^{d}} \left|\partial_{x}^{\alpha}x^{\beta}f(x)\right| \frac{(1+|x|^{2})^{d}}{(1+|x|^{2})^{d}} dx \tag{4.19}$$

$$\leq (2\pi)^{-d/2} \sup_{x \in \mathbb{R}^d} \left| (1+|x|^2)^d \partial_x^{\alpha} x^{\beta} f(x) \right| \underbrace{\int_{\mathbb{R}^d} \frac{dx}{(1+|x|^2)^d}}_{<\infty}$$
(4.20)

$$\leq C \sum_{j=0}^{m} \sup_{|\tilde{\alpha}|+|\tilde{\beta}|=j} \|f\|_{\tilde{\alpha},\tilde{\beta}}$$

$$(4.21)$$

for suitable $m \in \mathbb{N}$ and $0 < C < \infty$ that are independent of f. Thus, $\hat{f} \in \mathscr{S}$, and $f_n \to f$ in \mathscr{S} implies $\hat{f}_n \to \hat{f}$ in \mathscr{S} . In metric spaces, sequential continuity implies continuity, so $\mathscr{F} : \mathscr{S} \to \mathscr{S}$ is continuous. The argument for \mathscr{F}^{-1} is analogous. \Box

Theorem 4.8. (Fourier inversion theorem) The Fourier transformation $\mathscr{F} : \mathscr{S} \to \mathscr{S}$ is a continuous bijection, and \mathscr{F}^{-1} is its continuous inverse.

The fact that \mathscr{F} is bijective on \mathscr{S} will be relevant later for defining the Fourier transform of distributions ("generalized functions").

Proof. Given Lemma 4.7, we need to show only that $\mathscr{F}^{-1}\mathscr{F} = \mathrm{id}_{\mathscr{S}}$ and $\mathscr{F}\mathscr{F}^{-1} = \mathrm{id}_{\mathscr{S}}$; since the second relation can be proved in the same way as the first, we focus on the first. Since $\mathscr{F}^{-1}\mathscr{F}$ and $\mathrm{id}_{\mathscr{S}}$ are continuous, it suffices to show their equality on a dense subset.

Lemma 4.9. The set $C_0^{\infty}(\mathbb{R}^d)$ of smooth functions with compact support is dense in $\mathscr{S}(\mathbb{R}^d)$.

Proof. Choose a smooth "cut-off" function G with compact support and G(0) = 1, e.g.,

$$G(x) = \begin{cases} e^{1 - \frac{1}{1 - |x|^2}} & \text{for } |x| < 1\\ 0 & \text{otherwise.} \end{cases}$$
(4.22)

Then $f_n(x) := f(x)G(x/n)$ is a sequence in $C_0^{\infty}(\mathbb{R}^d)$ that converges to f in all semi-norms $\|\cdot\|_{\alpha,\beta}$.

We continue the proof of Theorem 4.8. Let $f \in C_0^{\infty}(\mathbb{R}^d)$. We will use Fourier series as Riemann sums approximating the Fourier integral. Let $C_m \subset \mathbb{R}^d$ be the cube centered at the origin of side length 2m with m so large that the support of f is contained in C_m . Let $L_m = \frac{\pi}{m} \mathbb{Z}^d$. Write f on C_m as a Fourier series

$$f(x) = \sum_{k \in L_m} f_k e^{ik \cdot x}, \quad f_k \in \mathbb{C}.$$
(4.23)

We use that the series converges pointwise; in fact, it converges uniformly:

Lemma 4.10. If the function $g : \mathbb{R} \to \mathbb{C}$ is 2π -periodic and continuously differentiable then its Fourier series $\sum_{n \in \mathbb{Z}} c_n e^{inx}$, $c_n = \langle e^{inx} | g \rangle / 2\pi$, converges uniformly to g.

Proof. Since g' is continuous and thus $g' \in L^2([0, 2\pi])$, its Fourier series $\sum_{n \in \mathbb{Z}} b_n e^{inx}$, $b_n = \langle e^{inx} | g' \rangle / 2\pi$, converges in L^2 , so $\sum |b_n|^2 < \infty$. By integration by parts, $b_n = inc_n$; thus, $\sum n^2 |c_n|^2 < \infty$. Apply the Cauchy–Schwarz inequality in $L^2(\mathbb{Z}, \#)$ to $\alpha_n = n |c_n|$ and $\beta_n = 1/n$ for $n \neq 0$ and $\beta_0 = 0$, $\sum |c_n| = \sum \alpha_n \beta_n \leq ||\alpha|| ||\beta|| < \infty$, or $(c_n)_{n \in \mathbb{Z}} \in L^1(\mathbb{Z}, \#)$. Now the uniform convergence follows. The uniform limit must be g because the L^2 limit is g.

We continue the proof of Theorem 4.8, considering a Fourier expansion of $f \in C_0^{\infty}(\mathbb{R}^d)$ with a very fine lattice L_m of k-values. The Fourier coefficients are

$$f_k = \frac{1}{\operatorname{vol}(C_m)} \int_{C_m} f(x) e^{-ik \cdot x} dx = \frac{1}{\operatorname{vol}(C_m)} \int_{\mathbb{R}^d} f(x) e^{-ik \cdot x} dx = \frac{(2\pi)^{d/2}}{(2m)^d} \hat{f}(k) \,. \tag{4.24}$$

Thus,

$$f(x) = \sum_{k \in L_m} \frac{\hat{f}(k)e^{ik \cdot x}}{(2\pi)^{d/2}} \left(\frac{\pi}{m}\right)^d,$$
(4.25)

which is a Riemann sum with $(\pi/m)^d$ the volume of a cube of side length π/m around a site in the lattice L_m . Since $\hat{f} \in \mathscr{S}$ is continuous, the Riemann sums converge to the Riemann integral, which exists (on every bounded cube) and equals the Lebesgue integral. That is,

$$f(x) = \lim_{m \to \infty} \sum_{k \in L_m} \frac{\hat{f}(k)e^{ik \cdot x}}{(2\pi)^{d/2}} \left(\frac{\pi}{m}\right)^d = \frac{1}{(2\pi)^d/2} \int_{\mathbb{R}^d} \hat{f}(k)e^{ik \cdot x} dk = (\mathscr{F}^{-1}\mathscr{F}f)(x). \quad (4.26)$$

Proposition 4.11. \mathscr{F} preserves L^2 norms: $\|\widehat{f}\| = \|f\|$ for every $f \in \mathscr{S}(\mathbb{R}^d)$.

Proof. By Fubini's theorem,

$$\int \left(\int e^{-ik \cdot x} f(k) \, dk \right) g(x) \, dx = \int \left(\int e^{-ik \cdot x} g(x) \, dx \right) f(k) \, dk \,, \tag{4.27}$$

so $\langle g^* | \hat{f} \rangle = \langle \hat{g}^* | f \rangle$. Since

$$(\mathscr{F}f)(k)^* = (2\pi)^{-d/2} \int \left(e^{-ik \cdot x} f(x) \right)^* dx = \mathscr{F}^{-1}(f^*)(k) , \qquad (4.28)$$

setting $g = \mathscr{F}^{-1}(f^*) = (\mathscr{F}f)^*$ yields $\langle \hat{f} | \hat{f} \rangle = \langle f | f \rangle$.

Theorem 4.12. (Plancherel theorem) The Fourier transformation $\mathscr{F} : \mathscr{S}(\mathbb{R}^d) \to \mathscr{S}(\mathbb{R}^d)$ possesses a unique bounded extension $\widetilde{\mathscr{F}} : L^2(\mathbb{R}^d) \to L^2(\mathbb{R}^d)$, which is a unitary operator.

For the proof, we need the fact that $\mathscr{S}(\mathbb{R}^d)$ is dense in $L^2(\mathbb{R}^d)$. This follows from **Theorem 4.13.** $C_0^{\infty}(\mathbb{R}^d)$ is dense in $L^2(\mathbb{R}^d)$.

Proof. The same reasoning that showed that in ℓ^2 the finite sequences form a dense subspace also shows, when applied to Fourier expansion, that in $L^2([0, 2\pi]^d)$ the finite Fourier series (or trigonometric polynomials, i.e., polynomials in $e^{ix_1}, \ldots, e^{ix_d}$) form a dense subspace. A fortiori, the smooth functions on \mathbb{R}^d that are 2π -periodic in every variable are dense. It is easy to check that every such function can be approximated with arbitrary accuracy in the norm of $L^2([0, 2\pi]^d)$ by smooth functions whose support in the open cube $(0, 2\pi)^d$ is compact; thus, $C_0^{\infty}((0, 2\pi)^d)$ is dense in $L^2([0, 2\pi]^d)$. Now we rescale the cube. Any $\psi \in L^2(\mathbb{R}^d)$ can be approximated by a function that vanishes outside some cube $[-a, a]^d$, which in turn can be approximated by a function in $C_0^{\infty}((-a, a)^d) \subset$ $C_0^{\infty}(\mathbb{R}^d)$. Proof of Theorem 4.12. Since $\mathscr{S}(\mathbb{R}^d) \subset L^2(\mathbb{R}^d)$, we can regard \mathscr{F} as $\mathscr{S} \to L^2$; since it preserves norms, it is bounded with $\|\mathscr{F}\| = 1$. By Theorem 3.13, it possesses a unique bounded extension $\tilde{\mathscr{F}}$ on L^2 . $\tilde{\mathscr{F}}$ preserves norms: Since $\tilde{\mathscr{F}}$ is continuous, for $\psi \in L^2$ and $\psi_n \in \mathscr{S}$ with $\psi_n \to \psi$ in the L^2 norm, $\mathscr{F}\psi_n = \mathscr{\tilde{F}}\psi_n \to \mathscr{\tilde{F}}\psi$ and thus (since the norm is continuous) $\|\psi_n\| = \|\mathscr{F}\psi_n\| \to \|\mathscr{\tilde{F}}\psi\|$ while $\|\psi_n\| \to \|\psi\|$, so $\|\mathscr{\tilde{F}}\psi\| = \|\psi\|$. $\mathscr{\tilde{F}}$ is surjective because $\mathscr{F} : \mathscr{S} \to \mathscr{S}$ is: For any $\psi \in L^2$, find $\psi_n \in \mathscr{S}$ with $\psi_n \to \psi$, set $\phi_n = \mathscr{F}^{-1}\psi_n \in \mathscr{S}$; then (ϕ_n) is a Cauchy sequence (because (ψ_n) is and \mathscr{F}^{-1} preserves norms), so $\phi_n \to \phi \in L^2$ and $\mathscr{\tilde{F}}\phi = \mathscr{\tilde{F}} \lim \phi_n = \lim \mathscr{\tilde{F}}\phi_n = \lim \psi_n = \psi$. Thus, $\mathscr{\tilde{F}}$ is unitary.

5 The free Schrödinger equation

Example 5.1. The Fourier transform of a Gauss function. For a > 0 let $f_a(x) = \exp(-ax^2/2)$. Then, using the substitution $y = \sqrt{\frac{a}{2}x}$,

$$\hat{f}_{a}(k) = \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} e^{-ax^{2}/2} e^{ik \cdot x} dx = \frac{1}{\sqrt{a\pi}} \int_{\mathbb{R}} e^{-y^{2} - iyk\sqrt{2/a}} dy$$
(5.1)

$$= \frac{e^{-k^{2}/2a}}{\sqrt{a\pi}} \underbrace{\int_{\mathbb{R}} e^{-(y+ik/\sqrt{2a})^{2}} dy}_{=\sqrt{\pi}} = \frac{1}{\sqrt{a}} e^{-k^{2}/2a} \,.$$
(5.2)

Now consider the free Schrödinger equation

$$i\hbar\frac{\partial\psi}{\partial t} = -\sum_{j=1}^{d} \frac{\hbar^2}{2m_j} \frac{\partial^2\psi}{\partial x_j^2}.$$
(5.3)

Taking the Fourier transform in x on both sides, we formally obtain

$$i\hbar \frac{\partial \widehat{\psi}}{\partial t} = \omega(k)\widehat{\psi}(t,k) \quad \text{with} \quad \omega(k) = \sum_{j=1}^{d} \frac{\hbar k_j^2}{2m_j}.$$
 (5.4)

For every fixed $k \in \mathbb{R}^d$, this is an ordinary differential equation of first order whose unique global solution is

$$\widehat{\psi}(t,k) = e^{-i\omega(k)t}\widehat{\psi}(0,k).$$
(5.5)

Now we Fourier transform back to obtain an expression for the solution to the free Schrödinger equation,

$$\psi(t,x) = \left(\mathscr{F}^{-1}e^{-i\omega(k)t}\mathscr{F}\psi_0\right)(x), \qquad (5.6)$$

with initial datum $\psi(0, x) = \psi_0(x)$. Here is the rigorous version of this reasoning:

Theorem 5.2. (Unique global solution of the free Schrödinger equation) Let $\psi_0 \in \mathscr{S}(\mathbb{R}^d)$. The unique global solution $\psi \in C^{\infty}(\mathbb{R}_t, \mathscr{S}(\mathbb{R}^d))$ of the free Schrödinger equation with $\psi(0, x) = \psi_0(x)$ is given by (5.6), or, equivalently, by

$$\psi(t,x) = \frac{1}{(2\pi i t)^{d/2}} \int_{\mathbb{R}^d} \exp\left(i\sum_{j=1}^d \frac{\hbar(x_j - y_j)^2}{2m_j t}\right) \psi_0(y) \, dy \,.$$
(5.7)

Moreover, $\|\psi(t,\cdot)\| = \|\psi_0\|$. (The exponential term, together with the pre-factor, is also known as the Green function for the free Schrödinger equation, named after George Green (1830).)

Proof. Define $\psi(t, x)$ by (5.6) and check that $\psi \in C^{\infty}(\mathbb{R}_t, \mathscr{S}(\mathbb{R}^d))$. We show as an example that $t \mapsto \psi(t)$ is differentiable in the sense claimed. Set

$$\varphi(t,x) = (\mathscr{F}^{-1}(-i\omega(k))e^{-i\omega(k)t}\mathscr{F}\psi_0)(x), \qquad (5.8)$$

which is our candidate for $\partial \psi / \partial t$. Obviously, $\varphi(t, \cdot) \in \mathscr{S}$. We have to show that

$$\lim_{h \to 0} \left\| \frac{\psi(t+h) - \psi(t)}{h} - \varphi(t) \right\|_{\alpha,\beta} = 0$$
(5.9)

for each α, β . By the continuity and linearity of \mathscr{F}^{-1} and \mathscr{F} , this is equivalent to

$$\lim_{h \to 0} \left\| \frac{\widehat{\psi}(t+h) - \widehat{\psi}(t)}{h} - \widehat{\varphi}(t) \right\|_{\alpha,\beta} = 0.$$
(5.10)

This follows from the smoothness of $e^{-i\omega t}$ and the decrease of $\widehat{\psi}_0(k)$ at infinity:

$$\left\|\frac{\widehat{\psi}(t+h) - \widehat{\psi}(t)}{h} - \widehat{\varphi}(t)\right\|_{\alpha,\beta} = \sup_{k \in \mathbb{R}^d} \left|k^{\alpha} \partial_k^{\beta} \left(\frac{e^{-i\omega(t+h)} - e^{-i\omega t}}{h} + i\omega(k)e^{-i\omega t}\right)\widehat{\psi}_0(k)\right|$$
(5.11)

converges to 0 as $h \to 0$.

Since (5.5) is the unique solution of (5.4), $\psi(t, x)$ is the unique solution in \mathscr{S} .

Now (5.7) can be verified by checking that it also defines a solution of the Schrödinger equation.

The preservation of norms follows from the facts that $\mathscr{F}, \mathscr{F}^{-1}$ preserve norms and that $|e^{-i\omega t}| = 1$.

Corollary 5.3. From (5.7) it follows that solutions of the free Schrödinger equation become small for large times,

$$\sup_{x \in \mathbb{R}^d} |\psi(t, x)| \le \frac{\|\psi_0\|_{L^1}}{(2\pi t)^{d/2}} \xrightarrow{t \to \infty} 0.$$
 (5.12)

Since the L^2 norm is preserved, the wave function has to spread more and more.

Example 5.4. Let $\psi_0(x) = e^{-ax^2+bx+c}$ with $a, b, c \in \mathbb{C}$, $\operatorname{Re} a > 0$. This is called a *Gaussian wave packet*. Obviously, $\psi_0 \in \mathscr{S}$. The solution stays a Gaussian packet for all times, but the coefficients become time-dependent, a(t), b(t), c(t). This can be seen either by verifying that the Fourier transform of a Gaussian packet is another Gaussian packet (and that $e^{-i\omega t}$ times a Gaussian packet is a Gaussian packet), or by inserting a Gaussian ansatz into the Schrödinger equation, which leads to the following ODEs:

$$\frac{da}{dt} = -2i\frac{\hbar}{m}a^2, \qquad \frac{db}{dt} = -2i\frac{\hbar}{m}ab, \qquad \frac{dc}{dt} = i\frac{\hbar}{m}(\frac{1}{2}b^2 - a). \tag{5.13}$$

Since these equations are locally Lipschitz, they have unique solutions that in fact exist for all times.

6 Distributional solutions

The algebraic dual space X^D of a complex vector space X is the space of all linear forms $X \to \mathbb{C}$. For a topological vector space X (i.e., a vector space with a topology in which addition and scalar multiplication are continuous), the *continuous dual space* X' of X is the space of all continuous linear forms $X \to \mathbb{C}$. Any normed vector space X is a topological vector space, and X', which consists of the bounded linear forms $X \to \mathbb{C}$, is equipped with the operator norm (and is, in fact, a Banach space).

Definition 6.1. The elements of the continuous dual space $\mathscr{S}'(\mathbb{R}^d)$ of Schwartz space are called *tempered distributions* (or sometimes *generalized functons*).

Example 6.2. (a) Suppose $f : \mathbb{R}^d \to \mathbb{C}$ is measurable and *polynomially* L^1 *bounded*, i.e., there are a polynomial P(x) and $g \in L^1(\mathbb{R}^d)$ such that

$$|f(x)| \le |P(x)| \, |g(x)| \tag{6.1}$$

for all $x \in \mathbb{R}^{d,3}$ Then $T_f : \mathscr{S} \to \mathbb{C}$ given by

$$T_f(\varphi) = \int_{\mathbb{R}^d} f(x) \,\varphi(x) \,dx \tag{6.2}$$

is linear and continuous, so $T_f \in \mathscr{S}'$.

Proof. Let $\varphi_n \to \varphi \in \mathscr{S}$. Then

$$|T_f(\varphi_n - \varphi)| \le \int_{\mathbb{R}^d} |f(x)| |\varphi_n(x) - \varphi(x)| dx$$
(6.3)

$$\leq \|g\|_{L^1} \left\| P(x) \left| \varphi_n - \varphi \right| \right\|_{\infty} \xrightarrow{n \to \infty} 0.$$
(6.4)

We will use the association $f \mapsto T_f$ to regard (e.g.) L^1 and \mathscr{S} as subsets of \mathscr{S}' . In the same way, $L^2(\mathbb{R}^d)$ can be regarded as a subset of \mathscr{S}' ; that is, for $f \in L^2$ is T_f as defined in (6.2) continuous: Indeed,

$$|T_f(\varphi_n - \varphi)| = \left| \langle f | \varphi_n - \varphi \rangle \right| \le ||f|| \, ||\varphi_n - \varphi|| \xrightarrow{n \to \infty} 0 \tag{6.5}$$

because $\|\varphi\| \le C \|(1+|x|^2)^{d/2}\varphi\|_{\infty}$ with $0 < C = (\int (1+|x|^2)^{-d} dx)^{1/2} < \infty$.

(b) The delta distribution $\delta : \mathscr{S} \to \mathbb{C}$ is defined by

$$\delta(\varphi) = \varphi(0) \,. \tag{6.6}$$

³Since every polynomial P(x) is bounded by $C(1+|x|^2)^n$ for some C > 0 and $n \in \mathbb{N}$, it suffices, in fact, to consider those special polynomials.

It is linear and continuous since

$$|\delta(\varphi_n - \varphi)| = |\varphi_n(0) - \varphi(0)| \le \|\varphi_n - \varphi\|_{\infty} \to 0, \qquad (6.7)$$

so $\delta \in \mathscr{S}'$.

Often one writes, in parallel to (6.2),

$$\delta(\varphi) = \int_{\mathbb{R}^d} \delta(x) \,\varphi(x) \,dx \tag{6.8}$$

and

$$\int_{\mathbb{R}^d} \delta(x-a) \,\varphi(x) \, dx = \varphi(a) \,. \tag{6.9}$$

Obviously, there is no genuine function $\delta : \mathbb{R}^d \to \mathbb{C}$ that would do this for every $\varphi \in \mathscr{S}$. But δ can be approximated in \mathscr{S}' by functions. For example, let $f \in L^1(\mathbb{R})$ with $\int f(x) dx = 1$ and

$$f_n(x) = n f(nx).$$
 (6.10)

Then, by the dominated convergence theorem, for every bounded and continuous φ (in particular, for $\varphi \in \mathscr{S}$),

$$T_{f_n}(\varphi) = \int_{\mathbb{R}} f_n(x) \,\varphi(x) \, dx = \int_{\mathbb{R}} f_n(x) \,\varphi(0) \, dx + \int_{\mathbb{R}} f_n(x) \,(\varphi(x) - \varphi(0)) \, dx \quad (6.11)$$

$$=\varphi(0) + \int_{\mathbb{R}} \underbrace{f(y)\left(\varphi(\frac{y}{n}) - \varphi(0)\right)}_{\to 0 \text{ pointwise}} dy \xrightarrow{n \to \infty} \varphi(0) = \delta(\varphi).$$
(6.12)

When defining $T \in \mathscr{S}'$ by defining $T(\varphi)$ for $\varphi \in \mathscr{S}$, the function φ is often called a *test function*.

Definition 6.3. Let X be a topological vector space. A sequence (φ_n) in X converges weakly to $\varphi \in X$ iff

$$\lim_{n \to \infty} T(\varphi_n) = T(\varphi) \quad \text{for all } T \in X'.$$
(6.13)

In this case one writes w-lim_{$n\to\infty$} $\varphi_n = \varphi$ or $\varphi_n \rightharpoonup \varphi$. A sequence (T_n) in X' converges weakly^{*} to $T \in X'$ iff

$$\lim_{n \to \infty} T_n(\varphi) = T(\varphi) \quad \text{for all } \varphi \in X.$$
(6.14)

In this case one writes w*-lim_{$n\to\infty$} $T_n = T$ or $T_n \stackrel{*}{\rightharpoonup} T$.

So $T_{f_n} \stackrel{*}{\rightharpoonup} \delta$ with f_n as in (6.10).

Theorem 6.4. If $A : \mathscr{S} \to \mathscr{S}$ is linear and continuous then

$$(A'T)(\varphi) = T(A\varphi) \qquad for \ all \ \varphi \in \mathscr{S}$$
(6.15)

defines a linear mapping $A': \mathscr{S}' \to \mathscr{S}'$ that is weakly^{*} continuous, called the adjoint of A.

Proof. We have that $A'T \in \mathscr{S}'$ because $A'T = T \circ A$ is the composition of continuous mappings and thus itself continuous. We need to establish the weak^{*} continuity. We first establish sequential continuity. Suppose $T_n \stackrel{*}{\rightharpoonup} T$, then, for every $\varphi \in \mathscr{S}$,

$$(A'T_n)(\varphi) = T_n(A\varphi) \to T(A\varphi) = (A'T)(\varphi), \qquad (6.16)$$

so $A'T_n \stackrel{*}{\rightharpoonup} A'T$. In general, sequential continuity does not imply continuity (as the weak* topology of \mathscr{S}' is not given by a metric). However, the same argument can be done for nets (a.k.a. Moore–Smith sequences) and net continuity implies here continuity.

Definition 6.5. For $T \in \mathscr{S}'$ the Fourier transform $\widehat{T} \in \mathscr{S}'$ is defined by $\widehat{T}(\varphi) = T(\widehat{\varphi})$ for all $\varphi \in \mathscr{S}$, or $\mathscr{F}_{\mathscr{S}'} = \mathscr{F}'_{\mathscr{S}}$.

Corollary 6.6. Fourier transformation $\mathscr{F}_{\mathscr{S}'}: \mathscr{S}' \to \mathscr{S}'$ is a weakly^{*} continuous bijection that extends the Fourier transformation on \mathscr{S} (and on L^1): for $f \in \mathscr{S}$ (or $f \in L^1$), $\widehat{T}_f = T_f$.

Proof. We have already that $\mathscr{F}_{\mathscr{S}'}$ is weakly^{*} continuous. Since $(\mathscr{F}^{-1'}\mathscr{F}'T)(f) = (\mathscr{F}'T)(\mathscr{F}^{-1}f) = T(\mathscr{F}\mathscr{F}^{-1}f) = T(f)$, the Fourier transformation on \mathscr{S}' must be bijective, and its (weakly^{*} continuous) inverse must be $(\mathscr{F}^{-1})'$. Finally, for $f \in L^1$,

$$\widehat{T}_f(\varphi) = T_f(\widehat{\varphi}) = \int f(x)\,\widehat{\varphi}(x)\,dx = \int \widehat{f}(x)\,\varphi(x)\,dx = T_{\widehat{f}}(\varphi) \tag{6.17}$$

using a relation we obtained in the proof of Proposition 4.11.

Example 6.7. The Fourier transform of the delta distribution is

$$\hat{\delta}(\varphi) = \delta(\hat{\varphi}) = \hat{\varphi}(0) = \frac{1}{(2\pi)^{d/2}} \int \varphi(x) \, dx = \int \frac{1}{(2\pi)^{d/2}} \varphi(x) \, dx = T_{(2\pi)^{-d/2}}(\varphi) \,. \tag{6.18}$$

That is, $\mathscr{F}(\delta)$ is the constant function $f(x) = (2\pi)^{-d/2}$.

The derivative $\partial_x^{\alpha} : \mathscr{S} \to \mathscr{S}$ is a continuous linear mapping as well, as

$$\|\partial_x^{\alpha}f\|_{\gamma,\beta} = \|x^{\gamma}\partial_x^{\beta}\partial_x^{\alpha}\|_{\infty} = \|f\|_{\gamma,\alpha+\beta}.$$
(6.19)

Thus, we can also extend the derivative from \mathscr{S} to \mathscr{S}' by taking more or less its adjoint:

Definition 6.8. For $T \in \mathscr{S}'$, the distributional derivative or weak derivative $\partial_x^{\alpha} T \in \mathscr{S}'$ is defined by

$$(\partial_x^{\alpha} T)(\varphi) = T\left((-1)^{|\alpha|} \partial_x^{\alpha} \varphi\right).$$
(6.20)

Corollary 6.9. The distributional derivative $\partial_x^{\alpha} : \mathscr{S}' \to \mathscr{S}'$ is weakly^{*} continuous and extends the derivative on \mathscr{S} (i.e., $\partial_x^{\alpha} T_f = T_{\partial_x^{\alpha} f}$ for all $f \in \mathscr{S}$).

Proof. It is continuous because it is essentially the adjoint. To see that it agrees with the ordinary derivative on \mathscr{S} , integrate by parts $|\alpha|$ times:

$$(\partial_x^{\alpha} T_f)(\varphi) = T_f((-1)^{|\alpha|} \partial_x^{\alpha} \varphi) = \int f(x) (-1)^{|\alpha|} \partial_x^{\alpha} \varphi(x) \, dx = \int \varphi(x) \, \partial_x^{\alpha} f(x) \, dx = T_{\partial_x^{\alpha} f}(\varphi)$$
(6.21)

Example 6.10. (a) For the Heaviside function $\Theta(x) = 1_{[0,\infty)}(x)$, the distributional derivative is $d\Theta/dx = \delta$. Indeed,

$$(d\Theta/dx)(\varphi) = -\Theta(d\varphi/dx) = -\int_0^\infty \frac{d\varphi}{dx} dx = \varphi(0) - \lim_{R \to \infty} \varphi(R) = \varphi(0). \quad (6.22)$$

- (b) The distributional derivative of f(x) = |x| with $x \in \mathbb{R}$ is $\Theta(x) \Theta(-x)$. Likewise, δ is the second derivative of the continuous function $f(x) = \max(0, x)$.
- (c) The derivatives of the delta distribution are

$$(\partial_x^{\alpha}\delta)(\varphi) = (-1)^{|\alpha|} \partial_x^{\alpha} f(0) .$$
(6.23)

 δ' can also be obtained as

$$w^* - \lim_{\varepsilon \to 0} \frac{\delta(x+\varepsilon) - \delta(x)}{\varepsilon} = w^* - \lim_{\varepsilon \to 0} \frac{\delta(x+\varepsilon) - \delta(x-\varepsilon)}{2\varepsilon} = w^* - \lim_{n \to \infty} \frac{d}{dx} \left(nf(nx) \right)$$
(6.24) for any $f \in \mathscr{S}(\mathbb{R})$ with $\int f(x) \, dx = 1$.

These examples are already rather typical:

Theorem 6.11. (Regularity theorem for distributions) Let $T \in \mathscr{S}'$. Then $T = \partial^{\alpha} f$ for some polynomially bounded continuous function f.

Proof. See M. Reed and B. Simon, Mathematical Methods of Modern Physics vol. I, page 139.

Let $C^{\infty}_{\text{poly}}(\mathbb{R}^d)$ be the space of all smooth functions f such that f and all of its derivatives are bounded by polynomials: For every $\alpha \in \mathbb{N}^d_0$,

$$|\partial^{\alpha} f| \le |P_{\alpha}(x)| \tag{6.25}$$

for a suitable polynomial P_{α} . For every such f, the multiplication operator $M_f : \mathscr{S} \to \mathscr{S}$ defined by

$$M_f(\varphi)(x) = f(x)\,\varphi(x) \tag{6.26}$$

is continuous. It is extended by its adjoint $M'_f : \mathscr{S}' \to \mathscr{S}'$; we can thus regard $M'_f T$ as the multiplication of $T \in \mathscr{S}'$ by f, and will simply write fT for it. Note, however, that the multiplication of two distributions is not defined. The extension of M_f (as well

as those of \mathscr{F} and ∂^{α}) are, in fact, unique, since \mathscr{S} is dense in \mathscr{S}' with respect to the weak^{*} topology.

* * *

Now consider the free Schrödinger equation, allow ψ to be a distribution in x, and regard the x derivatives as weak derivatives:

Theorem 6.12. For $\psi_0 \in \mathscr{S}'(\mathbb{R}^d)$, the unique global solution $\psi \in C^{\infty}(\mathbb{R}_t, \mathscr{S}'(\mathbb{R}^d))$ of the free Schrödinger equation in the distributional sense is given by

$$\psi(t) = \mathscr{F}^{-1} e^{-i\omega(k)t} \mathscr{F} \psi_0 \,. \tag{6.27}$$

Proof. We first note that $\psi(t)$ as defined by (6.27) lies in \mathscr{S}' . To see that $\psi(t)$ satisfies the free Schrödinger equation, let $\varphi \in \mathscr{S}(\mathbb{R}^d)$ be any test function:

$$i\hbar\frac{d}{dt}\psi(t)(\varphi) = i\hbar\frac{d}{dt}\psi_0\left(\mathscr{F}e^{-i\omega t}\mathscr{F}^{-1}\varphi\right)$$
(6.28)

$$= i\hbar \lim_{\varepsilon \to 0} \psi_0 \left(\mathscr{F} \frac{e^{-i\omega(t+\varepsilon)} - e^{-i\omega t}}{\varepsilon} \mathscr{F}^{-1} \varphi \right)$$
(6.29)

$$=i\hbar\psi_0\left(\mathscr{F}\lim_{\varepsilon\to 0}\frac{e^{-i\omega(t+\varepsilon)}-e^{-i\omega t}}{\varepsilon}\mathscr{F}^{-1}\varphi\right)$$
(6.30)

$$=\psi_0 \Big(\mathscr{F} e^{-i\omega t} \hbar \omega \mathscr{F}^{-1} \varphi \Big) \tag{6.31}$$

$$=\psi_0\Big(\mathscr{F}e^{-i\omega t}\mathscr{F}^{-1}H\varphi\Big) \tag{6.32}$$

$$=\psi_t\Big(H\varphi\Big)\tag{6.33}$$

$$=H\psi_t(\varphi)\,.\tag{6.34}$$

7 Unitary 1-parameter groups and self-adjoint operators

Since $L^2 \subset \mathscr{S}'$, the distributional solution of the free Schrödinger equation just defined provides, in particular, a solution $\psi(t)$ for every initial datum $\psi_0 \in L^2$. Since \mathscr{F} on \mathscr{S}' extends \mathscr{F} on L^2 (and so do multiplication operators),

$$\psi(t) = \mathscr{F}^{-1} e^{-i\omega t} \mathscr{F} \psi_0 \,. \tag{7.1}$$

(Recall that $\omega = \omega(k) = \sum_{j=1}^{d} \hbar k_j^2 / 2m_j$.) The operator mapping ψ_0 to $\psi(t)$,

$$U_t = \mathscr{F}^{-1} e^{-i\omega t} \mathscr{F}, \quad U_t : L^2(\mathbb{R}^d) \to L^2(\mathbb{R}^d),$$
(7.2)

is unitary (as it is the composition of 3 unitary operators) and known as the (free) propagator. In fact, the $U_t, t \in \mathbb{R}$ form a unitary 1-parameter group, i.e., a 1-parameter subgroup of the group of all unitary operators $L^2(\mathbb{R}^d) \to L^2(\mathbb{R}^d)$. To see this, we note the homomorphism property of $t \mapsto U_t$:

$$U_0 = I, \quad U_s U_t = U_{s+t}.$$
 (7.3)

The last equation is almost obvious from the meaning of U_t as "evolving by t time units" (and the time translation invariance, i.e., the fact that the evolution from time s to time s + t is the same as from 0 to t), but it can also be checked explicitly:

$$\mathscr{F}^{-1}e^{-i\omega s}\mathscr{F}\mathscr{F}^{-1}e^{-i\omega t}\mathscr{F} = \mathscr{F}^{-1}e^{-i\omega(s+t)}\mathscr{F}.$$
(7.4)

A question we need to ask about $\psi(t) = U_t \psi_0$ is: In which case is $t \mapsto \psi(t)$ actually differentiable, and what is its time derivative? We first note that it is continuous, as

$$\|\psi(t) - \psi(t_0)\|^2 = \|(U_t - U_{t_0})\psi_0\|^2 = \int_{\mathbb{R}^d} \left| e^{-i\omega t} - e^{-i\omega t_0} \right|^2 |\widehat{\psi}_0(k)|^2 \, dk \xrightarrow{t \to t_0} 0 \tag{7.5}$$

by the dominated convergence theorem. The statement $\frac{d\psi}{dt}(t_0) = \phi$ means that the following expression tends to zero as $t \to t_0$:

$$\left\|\frac{\psi(t) - \psi(t_0)}{t - t_0} - \phi\right\|^2 = \int_{\mathbb{R}^d} \left|\frac{e^{-i\omega t} - e^{-i\omega t_0}}{t - t_0} - \frac{\widehat{\phi}(k)}{\widehat{\psi}_0(k)}\right|^2 |\widehat{\psi}_0(k)|^2 \, dk \,. \tag{7.6}$$

This is the case if and only if

$$\widehat{\phi}(k) = -i\omega \, e^{-i\omega t_0} \, \widehat{\psi}_0(k) \tag{7.7}$$

for almost every k and, in addition, $\omega \widehat{\psi}_0 \in L^2$. In this case, ϕ actually is

$$\phi = i \sum_{j} \frac{\hbar}{2m_j} \frac{\partial^2 \psi_t}{\partial x_j^2} =: -\frac{i}{\hbar} H \psi_t$$
(7.8)

with weak derivatives. That is, if $\omega \widehat{\psi}_0 \in L^2$ then $t \mapsto \psi_t$ is differentiable at all times, $H\psi_t$ (with weak derivatives) exists, lies in L^2 , and $i\hbar d\psi_t/dt = H\psi_t$. One says that under this condition ψ_t solves $i\hbar d\psi_t/dt = H\psi_t$ in the L^2 sense. **Definition 7.1.** For $m \in \mathbb{Z}$, the *m*-th Sobolev space $H^m(\mathbb{R}^d) \subset \mathscr{S}'(\mathbb{R}^d)$ is the set of those $f \in \mathscr{S}'$ for which f is a measurable function and

$$(1+|k|^2)^{m/2}\hat{f} \in L^2(\mathbb{R}^d).$$
(7.9)

For $m \ge 0$, $H^m \subset L^2$.

So the initial conditions for which ψ_t solves the Schrödinger equation in the L^2 sense are those in the second Sobolev space H^2 .

Lemma 7.2. (Sobolev's lemma) Let $\ell \in \mathbb{N}_0$ and $f \in H^m(\mathbb{R}^d)$ with $m > \ell + \frac{d}{2}$. Then $f \in C^{\ell}(\mathbb{R}^d)$ and

$$\partial^a f \in C_{\infty}(\mathbb{R}^d) = \left\{ f \in C(\mathbb{R}^d) : \lim_{R \to \infty} \sup_{|x| > R} |f(x)| = 0 \right\}$$
(7.10)

for all $\alpha \in \mathbb{N}_0^d$ with $|\alpha| \leq \ell$.

Proof. We show that $k^{\alpha} \hat{f}(k) \in L^1$ for all α with $|\alpha| \leq \ell$. Then we use the Lemma of Riemann–Lebesgue to conclude $\partial^{\alpha} f \in C_{\infty}$.

Lemma 7.3. (Lemma of Riemann–Lebesgue) For $f \in L^1(\mathbb{R}^d)$, $\hat{f} \in C_{\infty}(\mathbb{R}^d)$.

Proof. For $f \in \mathscr{S}$ we know $\hat{f} \in \mathscr{S} \subset C_{\infty}$. Obviously, $\|\hat{f}\|_{\infty} \leq (2\pi)^{-d/2} \|f\|_{L^1}$. Thus, Fourier transformation is a bounded linear map from a dense set in L^1 to C_{∞} . By Theorem 3.13, it has a unique bounded linear extension $L^1 \to C_{\infty}$ (which must be \mathscr{F}).

We now prove Sobolev's lemma. Since $f \in H^m$, $(1 + |k|^2)^{m/2} \hat{f}(k) \in L^2$ and thus, for every α with $|\alpha| \leq \ell$,

$$\int_{\mathbb{R}^d} |k^{\alpha} \, \hat{f}(k)| \, dk \le \int (1 + |k|^2)^{\ell/2} \, |\hat{f}(k)| \, dk \tag{7.11}$$

$$= \int (1+|k|^2)^{m/2} |\hat{f}(k)| (1+|k|^2)^{(\ell-m)/2} dk$$
(7.12)

$$\leq \left\| (1+|k|^2)^{m/2} \, \hat{f}(k) \right\|_{L^2} \left(\int \frac{dk}{(1+|k|^2)^{m-\ell}} \right)^{1/2}, \tag{7.13}$$

using the Cauchy–Schwarz inequality. The last integral is finite iff $2(m-\ell) > d$. \Box

One may consider, instead of $t \mapsto \psi_t$, $t \mapsto U_t$. That is a mapping $\mathbb{R} \to \mathscr{B}(L^2)$, where $\mathscr{B}(L^2)$ is the space of bounded operators on L^2 . (For any normed spaces X, Y, the space $\mathscr{B}(X, Y)$ of bounded linear operators $X \to Y$ is again a normed space with the operator norm. If Y is complete (i.e., a Banach space) then so is $\mathscr{B}(X, Y)$.) However, $t \mapsto U_t$ is not continuous:

$$\|U_t - U_{t_0}\| = \left\| e^{-i\omega t} - e^{-i\omega t_0} \right\| = \sup_{k \in \mathbb{R}^d} \left| e^{-i\omega t} - e^{-i\omega t_0} \right| = 2 \quad \forall t \neq t_0.$$
(7.14)

7.1 Unitary 1-parameter groups

Definition 7.4. Let $A \in \mathscr{B}(\mathscr{H})$ and A_n be a sequence in $\mathscr{B}(\mathscr{H})$.

(i) A_n converges in norm to A (lim $A_n = A, A_n \to A$) iff

$$||A_n - A|| \to 0.$$
 (7.15)

(ii) A_n converges strongly to A (s-lim $A_n = A, A_n \xrightarrow{s} A$) iff

$$||A_n\psi - A\psi|| \to 0 \quad \text{for all } \psi \in \mathscr{H}.$$
(7.16)

(iii) A_n converges weakly to A (w-lim $A_n = A, A_n \xrightarrow{w} A$) iff

$$\langle \psi | (A_n - A)\phi \rangle \to 0$$
 for all $\psi, \phi \in \mathscr{H}$. (7.17)

The following implications hold:

norm convergence \Rightarrow strong convergence \Rightarrow weak convergence (7.18)

The converse implications are not generally valid.

So, U_t is a strongly continuous unitary 1-parameter group.

Definition 7.5. A densely defined operator H with domain $\mathscr{D}(H) \subseteq \mathscr{H}$ is the generator of a strongly continuous unitary 1-parameter group U_t iff

- (i) $\mathscr{D}(H) = \{ \psi \in \mathscr{H} : t \mapsto U_t \psi \text{ is differentiable} \}$
- (ii) $i\hbar dU_t \psi/dt = HU_t \psi$ for $\psi \in \mathscr{D}(H)$.

The generator of the physical time evolution is called the *Hamiltonian*. The Hamiltonian of the free Schrödinger is essentially the Laplace operator; more precisely, it is

$$H = \mathscr{F}^{-1}\hbar\omega\mathscr{F}, \quad \text{with } \mathscr{D}(H) = H^2(\mathbb{R}^d).$$
 (7.19)

We will see soon that exactly the self-adjoint operators are generators of strongly continuous unitary 1-parameter groups.

Proposition 7.6. Let H be a generator of a strongly continuous unitary 1-parameter group U_t .

- (i) $\mathscr{D}(H)$ is invariant under U_t , i.e., $U_t\mathscr{D}(H) = \mathscr{D}(H)$ for all $t \in \mathbb{R}$.
- (ii) H commutes with U_t , i.e.,

$$[H, U_t]\psi = HU_t\psi - U_tH\psi = 0 \quad for \ all \ \psi \in \mathscr{D}(H).$$

$$(7.20)$$

(iii) H is symmetric, i.e.,

$$\langle H\psi|\phi\rangle = \langle \psi|H\phi\rangle \quad \text{for all } \psi, \phi \in \mathscr{D}(H).$$
 (7.21)
(iv) U is uniquely determined by H.

(v) H is uniquely determined by U.

Proof. (i) $s \mapsto U_s U_t \psi = U_{s+t} \psi$ is differentiable iff $s \mapsto U_s \psi = U_{-t} U_{s+t} \psi$ is. (ii) For $\psi \in \mathscr{D}(H)$,

$$U_t H \psi = U_t i \hbar \frac{d}{ds} U_s \psi \Big|_{s=0} = i \hbar \frac{d}{ds} U_t U_s \psi \Big|_{s=0} = i \hbar \frac{d}{ds} U_s \underbrace{U_t \psi}_{\in \mathscr{D}(H)} \Big|_{s=0} = H U_t \psi \,. \tag{7.22}$$

(iii) This follows from the unitarity: For $\psi, \phi \in \mathscr{D}(H)$,

$$0 = \frac{d}{dt} \langle \psi | \phi \rangle = \frac{d}{dt} \langle U_t \psi | U_t \phi \rangle = \langle -\frac{i}{\hbar} H U_t \psi | U_t \phi \rangle + \langle U_t \psi | -\frac{i}{\hbar} H U_t \phi \rangle$$
(7.23)

$$= \frac{i}{\hbar} \langle U_t H \psi | U_t \phi \rangle - \frac{i}{\hbar} \langle U_t \psi | U_t H \phi \rangle = \frac{i}{\hbar} \Big(\langle H \psi | \phi \rangle - \langle \psi | H \phi \rangle \Big).$$
(7.24)

(iv) Suppose that H is also a generator of \tilde{U}_t . Then, by the symmetry of H,

$$\frac{d}{dt} \left\| (U_t - \tilde{U}_t)\psi \right\|^2 = 2\frac{d}{dt} \left(\|\psi\|^2 - \operatorname{Re}\langle U_t\psi|\tilde{U}_t\psi\rangle \right)$$
(7.25)

$$= -2\operatorname{Re}\left(\langle -\frac{i}{\hbar}HU_t\psi|\tilde{U}_t\psi\rangle + \langle U_t\psi| - \frac{i}{\hbar}H\tilde{U}_t\psi\rangle\right)$$
(7.26)

$$= -2 \operatorname{Re}\left(\frac{i}{\hbar} \langle HU_t \psi | \tilde{U}_t \psi \rangle - \frac{i}{\hbar} \langle U_t \psi | H\tilde{U}_t \psi \rangle\right)$$
(7.27)

$$=0 \tag{7.28}$$

for all $\psi \in \mathscr{D}(H)$. From $(U_0 - \tilde{U}_0)\psi = 0$ we can conclude $\tilde{U}|_{\mathscr{D}(H)} = U|_{\mathscr{D}(H)}$; since $\overline{\mathscr{D}(H)} = \mathscr{H}$, we obtain $\tilde{U} = U$ on all of \mathscr{H} .

(v) is immediate from the definition of H.

Example 7.7. Let $T_t : L^2(\mathbb{R}) \to L^2(\mathbb{R})$ be the group of translations, $T_t \psi(x) = \psi(x-t)$. It is strongly continuous and generated by $H = -i\hbar \frac{d}{dx}$, defined on $\mathscr{D}(H) = H^1(\mathbb{R})$.

7.2 Adjoint of a bounded operator

If $A: X \to Y$ is continuous, where X and Y are normed spaces, then the adjoint $A': Y' \to X'$ is defined by

$$(A'y')(x) = y'(Ax)$$
(7.29)

for $x \in X$ and $y' \in Y'$, and is continuous, too (with $||A'|| \leq ||A||$ because $||A'y'|| \leq ||y'|| ||A||$). We mention the fact that X' and Y' are automatically Banach spaces.

We will now explain that any Hilbert space \mathscr{H} can be regarded as its own dual space; this will allow us to regard the dual A' of an operator A on \mathscr{H} again as an operator A^* on \mathscr{H} .

For finite-dimensional spaces X, an inner product defines an identification between X and its dual space $X^D = X', J : X \to X^D$, by $J\psi = \langle \psi | \cdot \rangle$. We will now see that a Hilbert space can be identified with its continuous dual space in the same way.

Theorem 7.8. (*Riesz representation theorem*) Let \mathscr{H} be a Hilbert space and $T \in \mathscr{H}'$. Then there is a unique $\psi_T \in \mathscr{H}$ such that

$$T(\phi) = \langle \psi_T | \phi \rangle \quad \forall \phi \in \mathscr{H}.$$
(7.30)

Proof. Existence. Let $T \in \mathscr{H}'$ and M the kernel of T. (Note that if ψ_T exists then $M = \psi_T^{\perp}$.) If $M = \mathscr{H}$ then T = 0 and $\psi_T = 0$ does what was claimed. Now suppose $M \neq \mathscr{H}$. Then we want to show that M^{\perp} is one-dimensional. To see this, note that for any $\psi_0, \psi_1 \in M^{\perp} \setminus \{0\}$, by setting $\alpha = T(\psi_0)/T(\psi_1)$, we have that

$$T(\psi_0 - \alpha \psi_1) = T(\psi_0) - \alpha T(\psi_1) = 0$$
(7.31)

and thus

$$\psi_0 - \alpha \psi_1 \in M \cap M^\perp = \{0\} \tag{7.32}$$

or $\psi_0 = \alpha \psi_1$. Thus, M^{\perp} is 1-dimensional. Since T is continuous, M is closed. By the projection theorem, every $\phi \in \mathscr{H}$ can be written uniquely as

$$\phi = \phi_M + \phi_{M^{\perp}} = \phi_M + \frac{\langle \psi_0 | \phi \rangle}{\| \psi_0 \|^2} \psi_0 \,. \tag{7.33}$$

Now set $\psi_T = \frac{T(\psi_0)^*}{\|\psi_0\|^2} \psi_0$ and obtain

$$T(\phi) = T(\phi_M) + \frac{\langle \psi_0 | \phi \rangle}{\|\psi_0\|^2} T(\psi_0) = \langle \psi_T | \phi \rangle.$$
(7.34)

Uniqueness. This follows from the definiteness of the inner product.

Corollary 7.9. (Self-duality) The mapping

$$J\psi = \langle \psi | \cdot \rangle \tag{7.35}$$

is a bijection between \mathscr{H} and \mathscr{H}' . It is anti-linear, continuous, and isometric (i.e., preserves norms).

Proof. The range of J lies in \mathscr{H}' because $\langle \cdot | \cdot \rangle$ is continuous. By the Riesz theorem, J is surjective. By the Cauchy–Schwarz inequality, J is isometric and therefore injective and continuous.

Definition 7.10. For bounded $A : \mathscr{H} \to \mathscr{H}$, its *Hilbert-space-adjoint* A^* is defined by $J^{-1}A'J$. A is called *self-adjoint* if $A^* = A$.

Proposition 7.11. For $A \in \mathscr{B}(\mathscr{H})$,

$$\langle \psi | A\phi \rangle = \langle A^* \psi | \phi \rangle \quad \forall \psi, \phi \in \mathscr{H}.$$
 (7.36)

 A^* is uniquely determined by this property. A bounded operator $A : \mathcal{H} \to \mathcal{H}$ is selfadjoint iff it is symmetric, i.e.,

$$\langle \psi | A \phi \rangle = \langle A \psi | \phi \rangle \quad \forall \psi, \phi \in \mathscr{H}.$$
 (7.37)

Proof. By definition of A^* ,

$$\langle \psi | A\phi \rangle = (J\psi)(A\phi) = A'(J\psi)(\phi) = JJ^{-1}A'J\psi(\phi) = JA^*\psi(\phi) = \langle A^*\psi | \phi \rangle.$$
(7.38)

Uniqueness: Since the mapping $\phi \mapsto \langle \psi | A \phi \rangle$ is continuous and linear, the Riesz representation theorem guarantees the uniqueness of $\chi \in \mathscr{H}$ with $\langle \psi | A \phi \rangle = \langle \chi | \phi \rangle$.

The statement about self-adjointness now follows. We note already that, for unbounded operators A, being symmetric does not in general imply being self-adjoint. \Box

Example 7.12. Orthogonal projections P are self-adjoint operators. They are bounded, ||P|| = 1 (except P = 0), and satisfy $\langle P\psi|\phi\rangle = \langle \psi|P\phi\rangle$, see (3.17). Also, any finite linear combination of self-adjoint bounded operators with real coefficients is self-adjoint and bounded.

Theorem 7.13. (Properties of the adjoint) Let $A, B \in \mathscr{B}(\mathscr{H}), \lambda \in \mathbb{C}$. Then

- (i) $(A+B)^* = A^* + B^*$ and $(\lambda A)^* = \lambda^* A^*$.
- (*ii*) $(AB)^* = B^*A^*$
- (*iii*) $||A^*|| = ||A||$
- (*iv*) $A^{**} = A$
- (v) $||AA^*|| = ||A^*A|| = ||A||^2$
- (vi) ker $A = (\operatorname{im} A^*)^{\perp}$ and ker $A^* = (\operatorname{im} A)^{\perp}$.

Proof. (i)–(iii) follow from the corresponding properties of A' (||A'|| = ||A|| follows from the Hahn–Banach theorem), and (iv) from

$$\langle \psi | A\phi \rangle = \langle A^* \psi | \phi \rangle = \langle \phi | A^* \psi \rangle^* = \langle A^{**} \phi | \psi \rangle^* = \langle \psi | A^{**} \phi \rangle$$
(7.39)

for all $\psi, \phi \in \mathscr{H}$. Concerning (v) we observe that

$$||A\phi||^2 = \langle A\phi|A\phi\rangle = \langle \phi|A^*A\phi\rangle \le ||\phi||^2 ||A^*A||$$
(7.40)

and conclude

$$\|A\|^{2} = \sup_{\|\phi\|=1} \|A\phi\|^{2} \le \|A^{*}A\| \le \|A^{*}\| \|A\| = \|A\|^{2}.$$
(7.41)

Concerning (vi), $\phi \in \ker A$ iff $A\phi = 0$ iff $\langle \psi | A\phi \rangle = 0 \ \forall \psi$ iff $\langle A^*\psi | \phi \rangle = 0 \ \forall \psi$ iff $\phi \in (\operatorname{im} A^*)^{\perp}$.

Example 7.14. Let $L : \ell^2 \to \ell^2$ and $R : \ell^2 \to \ell^2$ be the left shift and right shift operators:

$$L(x_1, x_2, x_3, \ldots) = (x_2, x_3, \ldots), \quad R(x_1, x_2, x_3, \ldots) = (0, x_1, x_2, x_3, \ldots).$$
(7.42)

They are adjoints of each other, $R^* = L$ and $L^* = R$:

$$\langle x|Ry\rangle = \sum_{n=2}^{\infty} x_n y_{n-1} = \sum_{n=1}^{\infty} x_{n+1} y_n = \langle Lx|y\rangle.$$
(7.43)

Note that R is not unitary, although it is isometric: LR = I but $RL \neq I$.

Proposition 7.15. $U \in \mathscr{B}(\mathscr{H})$ is unitary iff $UU^* = I = U^*U$.

Proof. If U is unitary then

$$\langle U^*U\psi - \psi|\phi\rangle = \langle U\psi|U\phi\rangle - \langle \psi|\phi\rangle = 0 \quad \text{for all } \psi, \phi \in \mathscr{H}$$
(7.44)

and thus $U^*U = I$. It follows further that $UU^*U = U$ and, since U is surjective, $UU^* = I$.

Conversely, let $UU^* = I = U^*U$. Then U is surjective and

$$\langle U\psi|U\phi\rangle = \langle U^*U\psi|\phi\rangle = \langle \psi|\phi\rangle \quad \text{for all } \psi,\phi\in\mathscr{H}.$$
 (7.45)

Proposition 7.16. Let $H : \mathscr{H} \to \mathscr{H}$ be bounded and self-adjoint. Then

$$U_t = e^{-iHt/\hbar} = \sum_{n=0}^{\infty} \frac{(-iHt/\hbar)^n}{n!}$$
(7.46)

defines a strongly continuous (even norm continuous) unitary 1-parameter group whose generator is H with $\mathscr{D}(H) = \mathscr{H}$. (It is even true that $\mathbb{R} \to \mathscr{B}(\mathscr{H}) : t \mapsto e^{-iHt/\hbar}$ is differentiable.)

Proof. The series converges in the operator norm because it is Cauchy and $\mathscr{B}(\mathscr{H})$ is a Banach space. The claims can be verified in much the same way as for the ordinary exponential function. First check that when $A, B \in \mathscr{B}(\mathscr{H})$ and AB = BA then $e^{A+B} = e^A e^B = e^B e^A$. This implies the group property. Then check that $e^{A^*} = (e^A)^*$. This implies $U_{-t} = U_t^*$ and thus $U_t^* U_t = I = U_t U_t^*$, so U_t is unitary. $\|e^{tA} - I\| \leq \sum_{n=1}^{\infty} |t|^n \|A\|^n / n! = e^{|t| \|A\|} - 1 \to 0$ as $t \to 0$. Concerning dU_t/dt , $\|(e^{tA} - I)/t - A\| = \|\sum_{n=2}^{\infty} t^{n-1} A^n / n!\| \leq \sum_{n=2}^{\infty} |t|^{n-1} \|A\|^n / n! = [(e^{|t| \|A\|} - 1)/|t| - \|A\|] \to 0$ as $t \to 0$. □

7.3 Adjoint of an unbounded operator

Recall that an unbounded operator is a linear mapping $A : \mathscr{D}(A) \to \mathscr{H}, \ \mathscr{D}(A) \subseteq \mathscr{H}$. If $\mathscr{D}(A)$ is dense in \mathscr{H} then one says that A is *densely defined*. A is called *symmetric* iff

$$\langle A\psi|\phi\rangle = \langle \psi|A\phi\rangle \quad \text{for all } \psi, \phi \in \mathscr{D}(A) .$$
 (7.47)

(Example: the free Hamiltonian on the second Sobolev space is a densely defined symmetric operator.) $B: \mathscr{D}(B) \to \mathscr{H}$ is called an *extension* of A if $\mathscr{D}(A) \subseteq \mathscr{D}(B)$ and $B|_{\mathscr{D}(A)} = A$; in this case we write $A \subseteq B$.

Definition 7.17. The *adjoint operator* A^* of the densely defined operator $A : \mathscr{D}(A) \to \mathscr{H}$ has the domain

$$\mathscr{D}(A^*) = \left\{ \psi \in \mathscr{H} : \exists \chi \in \mathscr{H} \forall \phi \in \mathscr{D}(A) : \langle \psi | A\phi \rangle = \langle \chi | \phi \rangle \right\}$$
(7.48)

$$= \left\{ \psi \in \mathscr{H} : \phi \mapsto \langle \psi | A\phi \rangle \text{ is continuous on } \mathscr{D}(A) \right\}$$
(7.49)

and is on this domain defined by the relation

$$\langle \psi | A\phi \rangle = \langle A^* \psi | \phi \rangle \tag{7.50}$$

for all $\psi \in \mathscr{D}(A^*)$ and $\phi \in \mathscr{D}(A)$. (If $\mathscr{D}(A)$ were not dense then this relation would not uniquely determine A^* . A^* is a linear operator but not necessarily densely defined.) If $\mathscr{D}(A) = \mathscr{D}(A^*)$ and $A = A^*$ then A is called *self-adjoint*.

Theorem 7.18. Every strongly continuous unitary 1-parameter group has a generator, which is densely defined and self-adjoint (Stone's theorem). Conversely, a densely defined operator H is the generator of a strongly continuous unitary 1-parameter group iff it is self-adjoint.

We will not prove this theorem. The second statement follows from the *spectral* theorem for self-adjoint operators, which we will describe (but not prove) in the next chapter. As we will explain, the group generated by H can be written as $e^{-iHt/\hbar}$ also for unbounded H.

Example 7.19. (Multiplication operators) Let $V : \mathbb{R}^d \to \mathbb{C}$ be measurable, and let M_V be the multiplication operator $M_V : \mathscr{D}(M_V) \to L^2(\mathbb{R}^d)$,

$$(M_V\psi)(x) = V(x)\,\psi(x) \tag{7.51}$$

defined on

$$\mathscr{D}(M_V) = \left\{ \psi \in L^2(\mathbb{R}^d) : V\psi \in L^2(\mathbb{R}^d) \right\}.$$
(7.52)

 $\mathscr{D}(M_V)$ is always dense in L^2 and the adjoint operator M_V^* is given by

$$(M_V^*\psi)(x) = V(x)^*\psi(x), \quad \text{i.e., } M_V^* = M_{V^*},$$
(7.53)

on $\mathscr{D}(M_V^*) = \mathscr{D}(M_V)$. If V is real-valued then M_V is self-adjoint.

Example 7.20. The free Hamiltonian $H = -\sum_{j=1}^{d} \frac{\hbar^2}{2m_j} \partial_j^2$ on $H^2(\mathbb{R}^d)$ is self-adjoint because it is unitarily equivalent (via \mathscr{F}) to the multiplication operator $M_{\hbar\omega}$ on its maximal domain.

8 The spectral theorem for self-adjoint operators

Theorem 8.1. (The spectral theorem for self-adjoint operators in finite dimension) For every self-adjoint $n \times n$ matrix A there is an orthonormal basis of \mathbb{C}^n consisting of eigenvectors of A; every eigenvalue is real.

Also in infinite-dimensional \mathscr{H} , an *eigenvector* of $A : \mathscr{D}(A) \to \mathscr{H}$ is a $\psi \in \mathscr{D}(A) \setminus \{0\}$ such that

$$A\psi = \lambda\psi \tag{8.1}$$

for some $\lambda \in \mathbb{C}$; then λ is called the *eigenvalue* of ψ ; a number $\lambda \in \mathbb{C}$ is called an eigenvalue of A if there is a $\psi \in \mathscr{D}(A) \setminus \{0\}$ such that (8.1) holds. If $\mathscr{H} = L^2(\Omega)$ then eigenvectors are also called *eigenfunctions*. The set of eigenvectors with eigenvalue λ , together with the zero vector, forms a subspace, called the *eigenspace* with eigenvalue λ . If A is self-adjoint then its eigenvalues are real,

$$\lambda \langle \psi | \psi \rangle = \langle \psi | A \psi \rangle = \langle A \psi | \psi \rangle = \lambda^* \langle \psi | \psi \rangle, \qquad (8.2)$$

and eigenvectors corresponding to distinct eigenvalues are orthogonal:

$$(\lambda_i - \lambda_j) \langle \psi_i | \psi_j \rangle = \langle A \psi_i | \psi_j \rangle - \langle \psi_i | A \psi_j \rangle = 0.$$
(8.3)

In infinite-dimensional \mathscr{H} , operators do not necessarily have eigenvalues. For example, the free Hamiltonian has no eigenvalues, as it is unitarily equivalent to the multiplication operator $M_{\hbar\omega}$, which has no eigenvalues:

Example 8.2. The eigenvalues of a multiplication operator $M_V, V : \mathbb{R}^d \to \mathbb{C}$, are those values $\lambda \in \mathbb{C}$ such that the set $V^{-1}(\lambda) = \{x \in \mathbb{R}^d : V(x) = \lambda\}$ has positive measure.

Indeed, if $M_V \psi = \lambda \psi$ then $V(x)\psi(x) = \lambda \psi(x)$ for all x except in a set of measure zero, and so $V(x) = \lambda$ for all x with $\psi(x) \neq 0$ (except a null set). If $\psi \neq 0$ then $\{x : \psi(x) \neq 0\}$ must have positive measure. Conversely, if $V^{-1}(\lambda)$ has positive measure then any nonzero ψ that vanishes outside $V^{-1}(\lambda)$ is an eigenfunction of M_V with eigenvalue λ .

We thus need a notion of *generalized eigenvalues*; the set of generalized eigenvalues is called the *spectrum*.

Definition 8.3. The spectrum of the operator $A : \mathscr{D}(A) \to \mathscr{H}$ is

$$\sigma(A) = \left\{ z \in \mathbb{C} \left| (A - zI) : \mathscr{D}(A) \to \mathscr{H} \text{ is not bijective} \right\}.$$
(8.4)

One breaks down the spectrum as follows:

- σ_p(A) = {z ∈ C : A − zI is not injective} is called the *point spectrum*; it is the set of eigenvalues of A.
- $\sigma_c(A) = \{z \in \mathbb{C} : A zI \text{ is injective, not surjective, and has dense range}\}$ is called the *continuous spectrum*.

• $\sigma_{\rm r}(A) = \{z \in \mathbb{C} : A - zI \text{ is injective, not surjective, range not dense}\}$ is called the *residual spectrum*.

If dim $\mathscr{H} < \infty$ then $\sigma(A) = \sigma_{p}(A)$.

Example 8.4. The spectrum of a multiplication operator M_V , $V : \mathbb{R}^d \to \mathbb{R}$, is the essential range of V,

$$\left\{ y \in \mathbb{R} \middle| \forall \varepsilon > 0 : \mu \{ x \in \mathbb{R}^d : |V(x) - y| < \varepsilon \} > 0 \right\}$$
(8.5)

with μ the Lebesgue measure on \mathbb{R}^d . For the free Hamiltonian, $V(k) = \hbar \omega(k)$, the essential range is $\sigma(H) = [0, \infty)$.

Proof. Since V - zI is again a multiplication operator, we need to determine when a multiplication operator M_f is bijective. We have already shown that M_f is injective iff $\mu(f^{-1}(0)) = 0$. Suppose M_f is injective; it is surjective iff $\phi(x) = f(x)\psi(x)$ can be solved for $\psi \in L^2$ for any given $\phi \in L^2$, i.e., iff ϕ/f (which is well defined almost everywhere) is in L^2 for every $\phi \in L^2$; this is the case iff 1/f is bounded except on a null set, i.e., iff $\mu(f^{-1}[-\varepsilon,\varepsilon]) = 0$ for some $\varepsilon > 0$.

Theorem 8.5. Let A be self-adjoint and densely defined in \mathscr{H} . Then A has no residual spectrum, $\sigma(A) \subseteq \mathbb{R}$, and $||A|| = \sup_{\lambda \in \sigma(A)} |\lambda|$ (called the spectral radius).

Proof. See Reed and Simon, vol. I, Theorems VI.8 and VI.6.

8.1 The spectral theorem in terms of multiplication operators

Theorem 8.6. Let A be self-adjoint and densely defined in the separable Hilbert space \mathscr{H} . There is a measure space $(\Omega, \mathfrak{A}, \mu)$ with finite measure μ and a measurable function $h : \Omega \to \mathbb{R}$ such that A is unitarily equivalent to the multiplication operator M_h on $L^2(\Omega, \mathfrak{A}, \mu)$; i.e., there is a unitary $U : \mathscr{H} \to L^2(\Omega, \mathfrak{A}, \mu)$ such that $\psi \in \mathscr{D}(A)$ iff $hU\psi \in L^2(\Omega, \mathfrak{A}, \mu)$ and $UA\psi = hU\psi$ for all $\psi \in \mathscr{D}(A)$.

Proof. See Reed and Simon, vol. I, Theorem VIII.4.

There is a lot of freedom in the choice of $\Omega, \mathfrak{A}, \mu$, and h. Obivously, if $(\Omega', \mathfrak{A}', \mu')$ is isomorphic to $(\Omega, \mathfrak{A}, \mu)$ (i.e., if there is $\Phi : \Omega \to \Omega'$ that is measurable and bijective with measurable inverse and $\mu'(\Phi(\Delta)) = \mu(\Delta)$ for all $\Delta \in \mathfrak{A}$), then Ω can be replaced with Ω' and h with $h' = h \circ \Phi^{-1}$.

The space Ω can, in fact, be taken to be $\mathbb{R} \times \mathbb{N}$, or $\sigma(A) \times \mathbb{N}$, together with the function h(x, n) = x. The measure μ , however, is usually *not* the Lebesgue measure.

For $\mathscr{H} = \mathbb{C}^n$, Ω can be taken to be a set of n elements, $\mu = \#$, h any mapping whose values are the eigenvalues of A with appropriate multiplicity, and $\{U^{-1}e_i : i \in \Omega\}$, with $\{e_i\}$ the standard basis of $L^2(\Omega)$, an orthonormal basis of \mathscr{H} consisting of eigenvectors of A. A multiplication operator then means a diagonal matrix. (In the representation $\Omega = \mathbb{R} \times \mathbb{N}$, and $\mathscr{H} = \mathbb{C}^n$, μ must be taken to be concentrated on n points in Ω , each with first component equal to an eigenvalue of A, and each eigenvalue occurring with appropriate multiplicity; μ can be taken to give equal weight to each of those n points.)

8.2 The spectral theorem in terms of functional calculus

To define a functional calculus for an operator A means to define operators f(A) for all $f : \mathbb{R} \to \mathbb{C}$ or $f : \mathbb{C} \to \mathbb{C}$ or $f : \sigma(A) \to \mathbb{C}$ in some function space. If $f : \mathbb{C} \to \mathbb{C}$ is a polynomial, then it is obvious how to define f(A) for $A \in \mathscr{B}(\mathscr{H})$; for self-adjoint unbounded A, f(A) also makes sense but on a smaller domain:

$$\mathscr{D}(A^n) = \left\{ \psi \in \mathscr{H} : \psi \in \mathscr{D}(A), A\psi \in \mathscr{D}(A), A^2\psi \in \mathscr{D}(A), \dots, A^{n-1}\psi \in \mathscr{D}(A) \right\}$$
(8.6)

It follows from the spectral theorem 8.6 that

$$\mathscr{D}(A^n) = \left\{ \psi \in \mathscr{H} : h^n U \psi \in L^2(\Omega, \mathfrak{A}, \mu) \right\}.$$
(8.7)

For $A \in \mathscr{B}(\mathscr{H})$, one can also easily define f(A) if $f : \mathbb{C} \to \mathbb{C}$ is an entire holomorphic function, i.e., given by a power series of infinite radius of convergence: If

$$f(z) = \sum_{n=0}^{\infty} c_n z^n \text{ then } f(A) = \sum_{n=0}^{\infty} c_n A^n$$
(8.8)

is norm convergent.

In $\mathscr{H} = \mathbb{C}^n$, the functional calculus for self-adjoint A can be defined for arbitrary functions f by diagonalization: If $A = U^{-1} \operatorname{diag}(\lambda_1, \ldots, \lambda_n) U$ then set

$$f(A) = U^{-1} \operatorname{diag}(f(\lambda_1), \dots, f(\lambda_n)) U.$$
(8.9)

This functional calculus obviously extends the one for polynomials and power series.

So, different approaches allow us to define f(A) for different types of operators A and different classes of functions f. Let $\mathscr{L}^{\infty}(\mathbb{R})$ be the space of bounded measurable functions; note $L^{\infty} = \mathscr{L}^{\infty}/\sim$, where \sim means equality almost everywhere.

Theorem 8.7. (Spectral theorem in terms of functional calculus) Let A be self-adjoint and densely defined in the separable Hilbert space \mathscr{H} . Then there is a unique mapping $\mathscr{L}^{\infty}(\mathbb{R}) \to \mathscr{B}(\mathscr{H})$, denoted $f \mapsto f(A)$ and called the functional calculus of A, such that

(i) $f \mapsto f(A)$ is a homomorphism of algebras, i.e., linear and multiplicative:

$$(f + \lambda g)(A) = f(A) + \lambda g(A), \quad (fg)(A) = f(A) g(A) \quad \forall f, g \in E.$$
(8.10)

- (*ii*) $f^*(A) = f(A)^*$
- (iii) $||f(A)|| \leq ||f||_{\infty}$ (where $||\cdot||_{\infty}$ is the supremum, not the essential supremum)
- (iv) If a sequence $f_n \in \mathscr{L}^{\infty}(\mathbb{R})$ converges pointwise to x and $|f_n(x)| \leq |x|$ for all x and n then $\lim_{n\to\infty} f_n(A)\psi = A\psi$ for every $\psi \in \mathscr{D}(A)$.
- (v) If a sequence $f_n \in \mathscr{L}^{\infty}(\mathbb{R})$ converges pointwise to $f \in \mathscr{L}^{\infty}(\mathbb{R})$ and $\sup_n ||f_n||_{\infty} < \infty$ then s- $\lim_{n\to\infty} f_n(A) = f(A)$.

In addition:

- (vi) If $A\psi = \lambda \psi$ then $f(A)\psi = f(\lambda)\psi$.
- (vii) If f vanishes on the spectrum of A then f(A) = 0.

Proof. Uniqueness: see Reed and Simon vol. I, Theorems VII.1, VII.2, VIII.5.

Existence: follows from the first version of the spectral theorem 8.6 by setting

$$f(A) = U^{-1} M_{f \circ h} U \,. \tag{8.11}$$

Since $f \circ h$ is bounded, $M_{f \circ h} \in \mathscr{B}(\mathscr{H})$. Now the properties can be verified.

Corollary 8.8. Let \mathscr{H} be separable, H self-adjoint and densely defined, and $U_t = e^{-iHt/\hbar}$ as defined by the functional calculus. Then U_t is a strongly continuous unitary group with generator H.

Proof. The group property and $U_t^{-1} = U_{-t} = U_t^*$ follow from (i) and (ii). Furthermore, $e^{-ixt} \to 1$ pointwise as $t \to 0$ and $||e^{-ixt}||_{\infty} = 1$; by (v), s- $\lim_{t\to 0} e^{-iHt/\hbar} = I$; thus, U_t is a strongly continuous unitary group.

To see that H is the generator, let $U : \mathscr{H} \to L^2(\Omega, \mathfrak{A}, \mu)$ be as in the spectral theorem 8.6 for A = H. Then $U_t = U^{-1} e^{-iht/\hbar} U$ and

$$i\hbar \frac{d}{dt}U_t\psi = i\hbar U^{-1}\frac{d}{dt}e^{-iht/\hbar}U\psi = U^{-1}he^{-iht/\hbar}U\psi \in \mathscr{H} \Leftrightarrow hU\psi \in L^2(\Omega)$$
(8.12)

$$\Leftrightarrow H\psi \in \mathscr{H} \Leftrightarrow \psi \in \mathscr{D}(H).$$
(8.13)

Thus, H is a generator of U_t .

8.3 The spectral theorem in terms of PVMs

Another formulation of the spectral theorem in \mathbb{C}^n says that any self-adjoint A can be written as

$$A = \sum_{j=1}^{m} \lambda_j P_j \,, \tag{8.14}$$

where the λ_j are the eigenvalues of A, m is the number of eigenvalues, and P_j is the projection to the eigenspace of λ_j . In fact, this is the unique decomposition of A as a real-linear combination of projections that are mutually orthogonal. This formulation can be transferred to the infinite-dimensional case.

Definition 8.9. A projection-valued measure (PVM) P on a measurable space (Ω, \mathfrak{A}) acting on \mathscr{H} is a mapping $\mathfrak{A} \to \mathscr{B}(\mathscr{H})$ such that

- for every $\Delta \in \mathfrak{A}$, $P(\Delta)$ is a projection,
- $P(\Omega) = I$,

• P is σ -additive, i.e., for every sequence $\Delta_1, \Delta_2, \ldots \in \mathfrak{A}$ of mutually disjoint sets,

$$P\left(\bigcup_{n=1}^{\infty} \Delta_n\right) = \sum_{n=1}^{\infty} P(\Delta_n), \qquad (8.15)$$

where the series on the right hand side converges weakly, i.e., $\sum_{n} \langle \psi | P(\Delta_n) \psi \rangle$ converges for every $\psi \in \mathscr{H}$.

Corollary 8.10. (i) $P(\emptyset) = 0$.

- (ii) P is also finitely additive, e.g., $P(\Delta_1 \cup \Delta_2) = P(\Delta_1) + P(\Delta_2)$ if $\Delta_1 \cap \Delta_2 = \emptyset$.
- (iii) If $\Delta_1 \cap \Delta_2 = \emptyset$ then $P(\Delta_1)$ and $P(\Delta_2)$ correspond to orthogonal subspaces, $P(\Delta_1) P(\Delta_2) = 0.$

(*iv*)
$$P(\Delta_1 \cap \Delta_2) = P(\Delta_1) P(\Delta_2)$$

- (v) The series (8.15) also converges strongly, i.e., $\sum_{n} P(\Delta_n)\psi$ converges for every $\psi \in \mathscr{H}$.
- *Proof.* (i) Set $\Delta_1 = \Delta_2 = \ldots = \emptyset$ and obtain $P(\emptyset) = P(\emptyset) + P(\emptyset) + \ldots$, which can converge weakly only if $\langle \psi | P(\emptyset) \psi \rangle = 0$ for every $\psi \in \mathscr{H}$, which implies $P(\emptyset) = 0$.
 - (ii) Set $\Delta_3 = \Delta_4 = \ldots = \emptyset$.
- (iii) If the sum of two projections is again a projection, they must be mutually orthogonal: Recall that $P^2 = P$ for any projection P. If $P_1 + P_2$ is a projection then

 $P_1 + P_2 = (P_1 + P_2)^2 = P_1^2 + P_1P_2 + P_2P_1 + P_2^2 = P_1 + P_2 + P_1P_2 + P_2P_1 \quad (8.16)$

so $P_1P_2 + P_2P_1 = 0$. For any ψ_1 in the range of P_1 , it follows that $P_1P_2\psi = -P_2\psi$, but since $P_1\phi = -\phi$ implies $\phi = 0$, we have that $P_2\psi_1 = 0$. In the same way, we see $P_1\psi_2 = 0$ for any ψ_2 in the range of P_2 , so the two subspaces are orthogonal, and $P_1P_2 = 0 = P_2P_1$.

(iv) Set $A = \Delta_1 \cap \Delta_2$, $B = \Delta_1 \setminus A$, $C = \Delta_2 \setminus A$. Then

$$P(\Delta_1) P(\Delta_2) = P(A \cup B) P(A \cup C) = (P(A) + P(B))(P(A) + P(C))$$

= $P(A)^2 + P(B)P(A) + P(A)P(C) + P(B)P(C) = P(A)$. (8.17)

(v) Set
$$\Delta = \bigcup_{n=1}^{\infty} \Delta_n$$
 and $\tilde{\Delta}_N = \bigcup_{n=N+1}^{\infty} \Delta_n$.

$$\left\| \sum_{n=1}^{N} P(\Delta_n) \psi - P(\Delta) \psi \right\|^2 = \left\| P\left(\bigcup_{n=1}^{N} \Delta_n\right) \psi - P(\Delta) \psi \right\|^2 = \left\| P(\tilde{\Delta}_N) \psi \right\|^2$$

$$= \langle \psi | P(\tilde{\Delta}_N)^2 \psi \rangle = \langle \psi | P(\tilde{\Delta}_N) \psi \rangle = \sum_{n=1}^{N} \langle \psi | P(\Delta_n) \psi \rangle - \langle \psi | P(\Delta) \psi \rangle \xrightarrow{N \to \infty} 0.$$

Example 8.11. For any measure space $(\Omega, \mathfrak{A}, \mu)$, the *natural PVM* on Ω acting on $L^2(\Omega, \mathfrak{A}, \mu)$ is $P(\Delta) = M_{1_{\Delta}}$, i.e.,

$$\left(P(\Delta)\psi\right)(\omega) = \begin{cases} \psi(\omega) & \text{if } \omega \in \Delta\\ 0 & \text{if } \omega \notin \Delta. \end{cases}$$
(8.18)

Example 8.12. Suppose that Ω is a finite or countable set. Then $P(\Delta)$ can be expressed by singletons:

$$P(\Delta) = \sum_{j \in \Delta} P(\{j\}).$$
(8.19)

Thus, the PVM $P(\cdot)$ is determined by the $P_j = P(\{j\})$. Conversely, if $\{P_j : j \in \Omega\}$ is a family of mutually orthogonal projections with $\sum_{j\in\Omega} P_j = I$ then (8.19) defines a PVM. As a consequence, the diagonalization of a matrix $A : \mathbb{C}^n \to \mathbb{C}^n$ as in (8.14) defines a PVM on $\sigma(A)$ acting on \mathbb{C}^n ; it can also be regarded as a PVM on \mathbb{R} ; $P(\Delta)$ is the projection to the subspace spanned by the eigenspaces of all eigenvalues contained in Δ .

From the functional calculus it follows that with a self-adjoint operator A on a separable \mathscr{H} there is associated a *spectral PVM* $P(\cdot)$ on \mathbb{R} acting on \mathscr{H} by

$$P(\Delta) = 1_{\Delta}(A). \tag{8.20}$$

Theorem 8.13. (Spectral theorem in terms of PVMs) Let A be self-adjoint and densely defined in a separable \mathscr{H} . There is a unique PVM $P(\cdot)$ on \mathbb{R} acting on \mathscr{H} such that

$$A = \int_{\mathbb{R}} x P(dx) \,. \tag{8.21}$$

P is the spectral PVM of A. Conversely, the right hand side of (8.21) is always a self-adjoint and densely defined operator.

Definition 8.14. For a given PVM P, we define the operator

$$B_f = \int_{\Omega} f(\omega) P(d\omega)$$
(8.22)

for any measurable $f: \Omega \to \mathbb{C}$ by "weak integration": For any $\psi, \phi \in \mathscr{H}$,

$$\mu_{\psi,\phi}(\Delta) = \langle \psi | P(\Delta)\phi \rangle \tag{8.23}$$

defines a complex measure $\mu_{\psi,\phi}$ on Ω ; we define B_f by the property

$$\langle \psi | B_f \phi \rangle = \int_{\Omega} f(\omega) \, \mu_{\psi,\phi}(d\omega) \qquad \forall \psi, \phi \in \mathscr{D}(B_f)$$
(8.24)

on the domain

$$\mathscr{D}(B_f) = \left\{ \phi \in \mathscr{H} : \int_{\Omega} |f(\omega)|^2 \, \mu_{\phi,\phi}(d\omega) < \infty \right\}.$$
(8.25)

Proposition 8.15. $\mathscr{D}(B_f)$ is a dense subspace, and (8.24) defines a unique operator B_f on $\mathscr{D}(B_f)$.

Proof. We first show that $\mathscr{D}(B_f)$ is a subspace. Suppose $\psi, \phi \in \mathscr{D}(B_f)$; $\lambda \phi \in \mathscr{D}(B_f)$ is clear; we need to show that $\psi + \phi \in \mathscr{D}(B_f)$. Let $f_n(\omega) = \sum_{j=1}^n \lambda_{jn} \mathbb{1}_{\Delta_{jn}}(\omega)$ be a sequence of simple functions converging pointwise monotonically to $|f(\omega)|$.

$$\int |f|^{2} \mu_{\psi+\phi,\psi+\phi}(d\omega) = \lim_{n \to \infty} \int f_{n}^{2} \mu_{\psi+\phi,\psi+\phi}(d\omega)$$

$$= \lim_{n \to \infty} \sum_{j=1}^{n} \lambda_{jn}^{2} \langle \psi + \phi | P(\Delta_{jn}) | \psi + \phi \rangle$$

$$\leq \lim_{n \to \infty} \sum_{j=1}^{n} \lambda_{jn}^{2} \left(\langle \psi | P(\Delta_{jn}) | \psi \rangle + \langle \phi | P(\Delta_{jn}) | \phi \rangle + 2 \left| \langle \psi | P(\Delta_{jn}) | \phi \rangle \right| \right)$$

$$\leq \int |f|^{2} \mu_{\psi,\psi}(d\omega) + \int |f|^{2} \mu_{\phi,\phi}(d\omega) +$$

$$+ 2 \lim_{n \to \infty} \sum_{j=1}^{n} \lambda_{jn}^{2} \langle \psi | P(\Delta_{jn}) | \psi \rangle^{\frac{1}{2}} \langle \phi | P(\Delta_{jn}) | \phi \rangle^{\frac{1}{2}}$$

$$< \infty \qquad (8.26)$$

since, by the Cauchy–Schwarz inequality in \mathbb{R}^n ,

$$\lim_{n \to \infty} \sum_{j=1}^{n} \lambda_{jn}^{2} \langle \psi | P(\Delta_{jn}) | \psi \rangle^{\frac{1}{2}} \langle \phi | P(\Delta_{jn}) | \phi \rangle^{\frac{1}{2}}
\leq \lim_{n \to \infty} \left(\sum_{j=1}^{n} \lambda_{jn}^{2} \langle \psi | P(\Delta_{jn}) | \psi \rangle \right)^{\frac{1}{2}} \left(\sum_{j=1}^{n} \lambda_{jn}^{2} \langle \phi | P(\Delta_{jn}) | \phi \rangle \right)^{\frac{1}{2}}
= \left(\int |f|^{2} \mu_{\psi,\psi}(d\omega) \right)^{\frac{1}{2}} \left(\int |f|^{2} \mu_{\phi,\phi}(d\omega) \right)^{\frac{1}{2}} < \infty.$$
(8.27)

Thus, $\mathscr{D}(B_f)$ is a subspace. We now show it is dense: For any $\psi \in \mathscr{H}$ and $n \in \mathbb{N}$, let

$$\tilde{\Delta}_n = \{ \omega \in \Omega : |f(\omega)|^2 < n \}$$
(8.28)

and $\psi_n = P(\tilde{\Delta}_n)\psi$. Then $\psi_n \in \mathscr{D}(B_f)$ because μ_{ψ_n,ψ_n} is concentrated on $\tilde{\Delta}_n$ because

$$\mu_{\psi_n,\psi_n}(\Delta) = \langle \psi | P(\tilde{\Delta}_n) P(\Delta) P(\tilde{\Delta}_n) \psi \rangle = \langle \psi | P(\Delta \cap \tilde{\Delta}_n) \psi \rangle$$
(8.29)

by Corollary 8.10(iv). Furthermore, $\psi_n \to \psi$ by Corollary 8.10(v) for $\Delta_n = \{n - 1 \le |f|^2 < n\}$.

To see that $\int f \mu_{\psi,\phi}(d\omega)$ is well defined, we use polarization,

$$\mu_{\psi,\phi} = \frac{1}{4} \Big(\mu_{\psi+\phi,\psi+\phi} - \mu_{\psi-\phi,\psi-\phi} + i\mu_{\psi-i\phi,\psi-i\phi} - i\mu_{\psi+i\phi,\psi+i\phi} \Big) \,, \tag{8.30}$$

and note that for any $\chi \in \{\psi + \phi, \psi - \phi, \psi - i\phi, \psi + i\phi\}$ we have that $\chi \in \mathscr{D}(B_f)$ and thus

$$\int |f| \,\mu_{\chi,\chi}(d\omega) \leq \underbrace{\int \mathbf{1}_{\tilde{\Delta}_1} \,\mu_{\chi,\chi}(d\omega)}_{\langle \chi | P(\tilde{\Delta}_1)\chi \rangle} + \int |f|^2 \mu_{\chi,\chi}(d\omega) < \infty \,. \tag{8.31}$$

It now follows that the rhs of (8.24) is a sesquilinear form $S(\psi, \phi)$. We now show that for every $\phi \in \mathscr{D}(B_f)$ there is $C_{\phi} > 0$ such that, for all $\psi \in \mathscr{D}(B_f)$,

$$|S(\psi,\phi)| \le C_{\phi} \|\psi\|.$$
(8.32)

It then follows from the Riesz representation theorem, by regarding ϕ as fixed and ψ as a variable, that $S(\psi, \phi) = \langle \psi | \chi_{\phi} \rangle$ for some unique $\chi_{\phi} \in \mathscr{H}$; since χ_{ϕ} depends linearly on ϕ , it defines an operator $B_f \phi = \chi_{\phi}$. To see (8.32), choose simple functions $f_n = \sum_{j=1}^n \lambda_{jn} \mathbb{1}_{\Delta_{jn}}$ such that $f_n \to f$ pointwise and $|f_n| \leq |f|$. Then

$$|S(\psi,\phi)| = \left| \int f(\omega) \,\mu_{\psi,\phi}(d\omega) \right| \tag{8.33}$$

$$= \lim_{n \to \infty} \left| \int f_n(\omega) \,\mu_{\psi,\phi}(d\omega) \right| \tag{8.34}$$

$$= \lim_{n \to \infty} \left| \sum_{j=1}^{n} \lambda_{jn} \underbrace{\mu_{\psi,\phi}(\Delta_{jn})}_{\langle P(\Delta_{jn})\psi | P(\Delta_{jn})\phi \rangle} \right|$$
(8.35)

$$\leq \lim_{n \to \infty} \sum_{j=1}^{n} |\lambda_{jn}| \| P(\Delta_{jn})\psi \| \| P(\Delta_{jn})\phi \|$$
(8.36)

$$\leq \lim_{n \to \infty} \left(\sum_{j=1}^{n} \underbrace{\|P(\Delta_{jn})\psi\|^2}_{\langle \psi | P(\Delta_{jn})\psi \rangle} \right)^{1/2} \left(\sum_{j=1}^{n} |\lambda_{jn}|^2 \|P(\Delta_{jn})\phi\|^2 \right)^{1/2}$$
(8.37)

$$= \lim_{n \to \infty} \|\psi\| \left(\int |f_n|^2 \mu_{\phi,\phi}(d\omega) \right)^{1/2}.$$
 (8.38)

The functional calculus can be recovered from the spectral PVM of A by defining

$$f(A) = \int_{\mathbb{R}} f(x) P(dx) \,. \tag{8.39}$$

By a diagonalization of A, we mean either the unitary equivalence $U : \mathscr{H} \to L^2(\Omega, \mathfrak{A}, \mu)$ that carries A into a multiplication operator, or the PVM such that $A = \int x P(dx)$.

We say that the self-adjoint operators A_1, \ldots, A_n are simultaneously diagonalizable if there is one $U : \mathscr{H} \to L^2(\Omega, \mathfrak{A}, \mu)$ that will carry each A_i into a multiplication operator M_{h_i} ; equivalently, there is one PVM $P(\cdot)$ on \mathbb{R}^n such that

$$A_i = \int x_i P(dx) \tag{8.40}$$

for each *i*. We report that this is the case iff the A_i commute pairwise. For bounded operators, it is clear what that means: $A_iA_j = A_jA_i$. For unbounded operators, the meaning is less clear because $A_j\psi$ may not lie in the domain of A_i . That is why one says that two unbounded operators A_1, A_2 commute iff $P_1(\Delta_1)P_2(\Delta_2) = P_2(\Delta_2)P_1(\Delta_1)$ for all measurable sets $\Delta_1, \Delta_2 \subseteq \mathbb{R}$, with P_i the spectral PVM of A_i .

Can non-self-adjoint operators be diagonalized? Not necessarily. Let us consider only bounded operators. Note first that every $A \in \mathscr{B}(\mathscr{H})$ can be written in a unique way as

$$A = B + iC \tag{8.41}$$

with self-adjoint $B, C \in \mathscr{B}(\mathscr{H})$; indeed, $B = \frac{1}{2}(A + A^*)$ and $C = \frac{1}{2i}(A - A^*)$. A can be diagonalized iff B and C can be simultaneously diagonalized, which occurs iff B and C commute. This occurs iff A and A^* commute; such operators are called *normal*. For example, every unitary is normal because $U^*U = I = UU^*$.

A PVM can provide a notion of generalized orthonormal basis that we hinted at before. An ordinary orthonormal basis $\{\phi_n : n \in \mathbb{N}\}$ would correspond to the PVM Pon \mathbb{N} given by

$$P(\Delta) = \sum_{n \in \Delta} P_{\mathbb{C}\phi_n} , \qquad (8.42)$$

which is the projection to $\overline{\text{span}}\{\phi_n : n \in \Delta\}$. In Section 3.1 we defined a generalized ONB as a unitary $U : \mathscr{H} \to L^2(\Omega, \mathfrak{A}, \mu)$; define the PVM P on Ω by $P(\Delta) = U^{-1}M_{1\Delta}U$.

I will from now on take "self-adjoint" to include "densely defined."

9 Positive operators

Definition 9.1. The operator $A : \mathscr{D}(A) \to \mathscr{H}$ is called *positive* iff

$$\langle \psi | A \psi \rangle \ge 0 \qquad \forall \psi \in \mathscr{D}(A) \,.$$

$$\tag{9.1}$$

(In linear algebra, such operators are usually called "positive semi-definite.")

For example, projections are positive, as

$$\langle \psi | P\psi \rangle = \langle \psi | P^2 \psi \rangle = \langle P\psi | P\psi \rangle = \| P\psi \|^2 \ge 0.$$
(9.2)

Proposition 9.2. Let A be a self-adjoint operator in a separable \mathcal{H} . A is positive iff $\sigma(A) \subseteq [0, \infty)$.

Proof. By the spectral theorem 8.6, A is unitarily equivalent to a multiplication operator M_h , whose spectrum is the essential range of h. If that contains a negative number λ then for every $\varepsilon > 0$ (such as $0 < \varepsilon < |\lambda|$), the set $h^{-1}([\lambda - \varepsilon, \lambda + \varepsilon])$ has positive measure, so there is be a nonzero $\psi \in L^2$ concentrated on that set, and then

$$\langle \psi | M_h \psi \rangle = \int |\psi(\omega)|^2 h(\omega) \, \mu(d\omega) < (\lambda + \varepsilon) \|\psi\|^2 < 0 \,, \tag{9.3}$$

and A is not positive. Conversely, if the essential range of h is $[0, \infty)$ then for every $\lambda < 0$ there is $\varepsilon_{\lambda} > 0$ such that $h^{-1}(\lambda - \varepsilon_{\lambda}, \lambda + \varepsilon_{\lambda})$ is a null set; for any $n \in \mathbb{N}$, $h^{-1}[-n, -\frac{1}{n}]$ is a null set since $[-n, -\frac{1}{n}]$ is compact and thus covered by finitely many of the sets $(\lambda - \varepsilon_{\lambda}, \lambda + \varepsilon_{\lambda})$; thus, $h^{-1}(-\infty, 0) = \bigcup_{n} h^{-1}[-n, -\frac{1}{n}]$ is a null set. Therefore, $g = h \mathbf{1}_{h \ge 0}$ differs from h only on a null set, so $M_g = M_h$ and

$$\langle \psi | M_g \psi \rangle = \int |\psi(\omega)|^2 g(\omega) \, \mu(d\omega) \ge 0 \,.$$

$$(9.4)$$

As a consequence, if A is self-adjoint and positive then $P(-\infty, 0) = 0$ for the spectral PVM P of A. As another consequence, the free Hamiltonian is a positive operator.

Proposition 9.3. For every positive self-adjoint operator A in a separable \mathscr{H} there is a unique positive self-adjoint operator \sqrt{A} such that $\sqrt{A}^2 = A$.

Proof. We have seen that in $UAU^{-1} = M_h$, h can be taken to be non-negative. Define $\sqrt{A} = U^{-1}M_{\sqrt{h}}U$ on $\mathscr{D}(\sqrt{A}) = U\mathscr{D}(M_{\sqrt{h}})$, which we know is self-adjoint and positive. Uniqueness: Let $B^2 = A$, B self-adjoint and positive, and let P_A, P_B be the spectral PVMs of A, B. Then

$$P_A(\Delta) = 1_{\Delta}(B^2) = 1_{\Delta}\left(\int_0^\infty x^2 P_B(dx)\right) = \int_0^\infty 1_{\Delta}(x^2) P_B(dx) = P_B\{\sqrt{x} : x \in \Delta\}.$$
 (9.5)

9.1 Bounded positive operators

If $A \in \mathscr{B}(\mathscr{H})$ is positive then it is self-adjoint. In fact, it suffices that $\langle \psi | A\psi \rangle \in \mathbb{R}$ for all $\psi \in \mathscr{H}$: Polarization allows us to express $\langle \psi | A\phi \rangle$ in terms of diagonal elements $\langle \chi | A\chi \rangle$. Thus, if $\langle \psi | A\psi \rangle = \langle \psi | B\psi \rangle$ for all $\psi \in \mathscr{H}$ then A = B. Thus, if $\langle \psi | A\psi \rangle = \langle A\psi | \psi \rangle = \langle \psi | A^*\psi \rangle$ for all $\psi \in \mathscr{H}$ then $A = A^*$.

Definition 9.4. A partial order " \leq " is defined on $\mathscr{B}(\mathscr{H})$ by setting

$$A \le B \quad :\Leftrightarrow \quad (B-A) \text{ is positive.}$$
(9.6)

Recall that a partial order is a relation such that $A \leq A$ (reflexive); if $A \leq B$ and $B \leq A$ then A = B (anti-symmetric); if $A \leq B$ and $B \leq C$ then $A \leq C$ (transitive). Indeed, A - A = 0 is positive; if B - A is positive and A - B is positive, too, then A - B = 0; if B - A is positive and C - B is positive then C - A = (C - B) + (B - A) is positive, too.

It follows that $A \ge 0$ is another notation for saying that A is positive. The set of positive operators in $\mathscr{B}(\mathscr{H})$ forms a convex cone. For every $A \in \mathscr{B}(\mathscr{H})$ is A^*A a positive operator, as $\langle \psi | A^*A\psi \rangle = ||A\psi||^2 \ge 0$; AA^* is a possibly different positive operator.

10 The predictive formalism of quantum mechanics

We have considered the time evolution of the wave function but have hardly talked about the connection between the wave function and the visible world. In Bohmian mechanics, this connection lies in the fact that the particles move in a way that depends on the wave function. In orthodox quantum mechanics, one focuses on the connection between a system's wave function and the probability distribution of outcomes of experiments done on the system. The equations describing this connection are *postulated* in orthodox quantum mechanics but can be *derived* in Bohmian mechanics; the latter I will elucidate in this chapter. I first need to add a remark about the Schrödinger equation with potential, concerning the simplest case: bounded potentials.

Remark 10.1. Let H_0 be the free Hamiltonian. If $V : \mathbb{R}^d \to \mathbb{R}$ is (measurable and) bounded then $M_V \in \mathscr{B}(\mathscr{H})$ with $||M_V|| = ||V||_{L^{\infty}}$, and $H = H_0 + M_V$ is self-adjoint on $\mathscr{D}(H) = \mathscr{D}(H_0)$. Indeed, H is certainly defined and symmetric on $\mathscr{D}(H_0)$; we need to check $\mathscr{D}(H^*) = \mathscr{D}(H)$. Recall

$$\mathscr{D}(H^*) = \left\{ \psi \in \mathscr{H} : \phi \mapsto \langle \psi | H\phi \rangle \text{ is bounded for } \phi \in \mathscr{D}(H) \right\}.$$
(10.1)

If there is $C_{\psi} > 0$ such that, for all $\phi \in \mathscr{D}(H) := \mathscr{D}(H_0)$,

$$C_{\psi} \|\phi\| \ge \left| \langle \psi | H\phi \rangle \right| \ge \left| \langle \psi | H_0\phi \rangle \right| - \left| \langle \psi | V\phi \rangle \right|$$
(10.2)

$$\geq \left| \langle \psi | H_0 \phi \rangle \right| - \| \psi \| \| \phi \| \| V \|_{L^{\infty}} \tag{10.3}$$

then $(C_{\psi} + \|\psi\|\|V\|_{L^{\infty}})\|\phi\| \ge |\langle\psi|H_0\phi\rangle|$, so $\psi \in \mathscr{D}(H_0)$ since H_0 is self-adjoint. This shows that $\mathscr{D}(H^*) = \mathscr{D}(H)$.

Theorem 10.2. (Global existence and equivariance of Bohmian mechanics) Suppose that V is smooth and bounded, that $\psi_0 \in C^{\infty}(\mathbb{R}^d) \cap L^2(\mathbb{R}^d)$ and every $\partial_x^{\alpha} \psi_0 \in L^2(\mathbb{R}^d)$. Then the Bohmian trajectory $X(t) = (X_1(t), \ldots, X_d(t))$, i.e., the solution of

$$\frac{dX_i}{dt} = \frac{\hbar}{m_i} \operatorname{Im} \frac{\partial_i \psi}{\psi}(X(t)), \qquad (10.4)$$

exists for all $t \in \mathbb{R}$ for $|\psi_0|^2$ -almost every $X(0) \in \mathbb{R}^d$. Moreover, the $|\psi|^2$ distribution is equivariant, i.e., if X(0) is $|\psi_0|^2$ distributed then X(t) is $|\psi_t|^2$ distributed for any $t \in \mathbb{R}$.

Proof. See K. Berndl, D. Dürr, S. Goldstein, G. Peruzzi, N. Zanghì: "On the global existence of Bohmian mechanics," Commun. Math. Phys. **173**: 647–673 (1995), http://arXiv.org/abs/quant-ph/9503013; or S. Teufel, R. Tumulka: "Simple Proof for Global Existence of Bohmian Trajectories," Commun. Math. Phys. **258**: 349-365 (2005), http://arxiv.org/abs/math-ph/0406030. The condition on V can be relaxed.

10.1 Tensor product spaces

The operation of taking the tensor product $\mathscr{H}_1 \otimes \mathscr{H}_2$ of two Hilbert spaces is defined for arbitrary Hilbert spaces $\mathscr{H}_1, \mathscr{H}_2$, but in such a way that

$$L^{2}(\Omega_{1}) \otimes L^{2}(\Omega_{2}) = L^{2}(\Omega_{1} \times \Omega_{2}).$$

$$(10.5)$$

It comes together with an operation on vectors, $\psi_1 \otimes \psi_2 \in \mathscr{H}_1 \otimes \mathscr{H}_2$, defined such that for functions $\psi_1(\omega_1), \psi_2(\omega_2)$,

$$\psi_1 \otimes \psi_2(\omega_1, \omega_2) = \psi_1(\omega_1) \,\psi_2(\omega_2) \,. \tag{10.6}$$

Physically, the Hilbert space of two particles (or two systems) together is the tensor product of their Hilbert spaces. A wave function of two particles (or two systems) together is not always a tensor product (most functions of two variables are not a product of two one-variable functions); if it is then it is called *disentangled*, otherwise *entangled*.

The mathematical definition of the *tensor product of Hilbert spaces* will be the completion of the *algebraic tensor product*.

Definition 10.3. For $\phi_1 \in \mathscr{H}_1, \phi_2 \in \mathscr{H}_2$ let $\phi_1 \otimes \phi_2$ be the bilinear mapping $\mathscr{H}_1 \times \mathscr{H}_2 \to \mathbb{C}$ given by

$$(\phi_1 \otimes \phi_2)(\psi_1, \psi_2) = \langle \phi_1 | \psi_1 \rangle_{\mathscr{H}_1} \langle \phi_2 | \psi_2 \rangle_{\mathscr{H}_2} .$$
(10.7)

Let S be the space of finite linear combinations of such tensor products. On S, a scalar product is defined by

$$\left\langle \sum_{i} c_{i} \phi_{1i} \otimes \phi_{2i} \middle| \sum_{j} d_{j} \psi_{1j} \otimes \psi_{2j} \right\rangle_{S} = \sum_{i,j} c_{i}^{*} d_{j} \langle \phi_{1i} | \psi_{2j} \rangle_{\mathscr{H}_{1}} \langle \phi_{2i} | \psi_{2j} \rangle_{\mathscr{H}_{2}}.$$
(10.8)

 $\mathscr{H}_1 \otimes \mathscr{H}_2$ is the completion of S.

Proposition 10.4. Let $B_1 = \{\phi_{1i} : i \in \mathscr{I}_1\}$ and $B_2 = \{\phi_{2j} : j \in \mathscr{I}_2\}$ be orthnormal bases of \mathscr{H}_1 and \mathscr{H}_2 . Then $B_{12} = \{\phi_{1i} \otimes \phi_{2j} : i \in \mathscr{I}_1, j \in \mathscr{I}_2\}$ is an ONB of $\mathscr{H}_1 \otimes \mathscr{H}_2$.

Proof. Clearly, B_{12} is an ONS in $\mathscr{H}_1 \otimes \mathscr{H}_2$. We show that $\overline{\operatorname{span}}(B_{12})$ contains every $\psi_1 \otimes \psi_2$ with $\psi_1 \in \mathscr{H}_1$ and $\psi_2 \in \mathscr{H}_2$. This will imply that it also contains S and thus its closure $\mathscr{H}_1 \otimes \mathscr{H}_2$. To this end, let

$$\psi_k = \sum_{i \in \mathscr{I}_k} c_{ki} \phi_{ki}, \quad k = 1, 2$$
(10.9)

observe that only countably many terms are nonzero, so we can write

$$\psi_k = \sum_{i=1}^{\infty} c_{ki} \phi_{ki} \,, \tag{10.10}$$

and set

$$\psi_k(N) = \sum_{i=1}^N c_{ki} \phi_{ki} \,. \tag{10.11}$$

Then

$$\psi_1(N) \otimes \psi_2(N) = \sum_{i,j=1}^N c_{1i} c_{2j} \,\phi_{1i} \otimes \phi_{2j} \in \operatorname{span}(B_{12}) \tag{10.12}$$

and

$$\left\| \psi_1 \otimes \psi_2 - \psi_1(N) \otimes \psi_2(N) \right\|_S^2$$

$$= \left\langle \psi_1 \otimes \psi_2 - \psi_1(N) \otimes \psi_2(N) \right| \psi_1 \otimes \psi_2 - \psi_1(N) \otimes \psi_2(N) \right\rangle_S$$
(10.13)

$$= \|\psi_1\|^2 \|\psi_2\|^2 + \|\psi_1(N)\|^2 \|\psi_2(N)\|^2 - 2\operatorname{Re}\langle\psi_1(N)|\psi_1\rangle \langle\psi_2(N)|\psi_2\rangle \xrightarrow{N \to \infty} 0. \quad (10.14)$$

The universal property of the tensor product is that every bounded bilinear mapping $B: \mathscr{H}_1 \otimes \mathscr{H}_2 \to \mathscr{H}_3$ factors over \otimes , i.e., there is a bounded linear operator $L: \mathscr{H}_1 \otimes \mathscr{H}_2 \to \mathscr{H}_3$ such that $B(\psi_1, \psi_2) = L(\psi_1 \otimes \psi_2)$.

There are other ways of defining $\mathscr{H}_1 \otimes \mathscr{H}_2$: (i) One can say that the pair $(\mathscr{H}_1 \otimes \mathscr{H}_2, \otimes)$ consisting of a Hilbert space \mathscr{K} and a bilinear mapping $\otimes : \mathscr{H}_1 \times \mathscr{H}_2 \to \mathscr{K}$ is defined only up to unitary isomorphism by the universal property and the property

$$\left\langle \psi_1 \otimes \psi_2 \middle| \phi_1 \otimes \phi_2 \right\rangle = \left\langle \psi_1 \middle| \phi_1 \right\rangle \left\langle \psi_2 \middle| \phi_2 \right\rangle.$$
 (10.15)

(ii) One could make Proposition 10.4 the definition: Let $\mathscr{H}_1 \otimes \mathscr{H}_2$ be a Hilbert space with an ONB consisting of the formal symbols $\phi_{1i} \otimes \phi_{2j}$, and let, for $\psi_k = \sum c_{ki}\phi_{ki}$, (k = 1, 2)

$$\psi_1 \otimes \psi_2 = \sum_{i \in \mathscr{I}_1} \sum_{j \in \mathscr{I}_2} c_{1i} c_{2j} \phi_{1i} \otimes \phi_{2j} \,. \tag{10.16}$$

(iii) One could take the set S' of all formal symbols $\psi_1 \otimes \psi_2$, where $\psi_k \in \mathscr{H}_k$, consider the vector space V' formally spanned by it, and mod out the relations $(c\psi_1) \otimes \psi_2 \sim \psi_1 \otimes (c\psi_2) \sim c(\psi_1 \otimes \psi_2), (\psi_1 + \psi'_1) \otimes \psi_2 \sim \psi_1 \otimes \psi_2 + \psi'_1 \otimes \psi_2, \psi_1 \otimes (\psi_2 + \psi'_2) \sim \psi_1 \otimes \psi_2 + \psi_1 \otimes \psi'_2$. This leads essentially to S. Then proceed as with S.

Proposition 10.5.

$$L^{2}(\Omega_{1},\mathfrak{A}_{1},\mu_{1})\otimes L^{2}(\Omega_{2},\mathfrak{A}_{2},\mu_{2})=L^{2}(\Omega_{1}\times\Omega_{2},\mathfrak{A}_{1}\otimes\mathfrak{A}_{2},\mu_{1}\otimes\mu_{2}), \qquad (10.17)$$

$$\psi_1 \otimes \psi_2(\omega_1, \omega_2) = \psi_1(\omega_1)\psi_2(\omega_2),$$
 (10.18)

where $\mathfrak{A}_1 \otimes \mathfrak{A}_2$ is generated by product sets, $\mu_1 \otimes \mu_2$ is the product measure, and "=" means canonically isomorphic. (Alternatively, this is one of the unitarily equivalent ways of defining the tensor product via the universal property.)

Proof. Fubini's theorem.

10.2 POVMs

Definition 10.6. A positive-operator-valued measure (POVM) E on a measurable space (Ω, \mathfrak{A}) acting on \mathscr{H} is a mapping $\mathfrak{A} \to \mathscr{B}(\mathscr{H})$ such that

• for every $\Delta \in \mathfrak{A}$, $0 \leq E(\Delta)$,

•
$$E(\Omega) = I$$
,

• E is σ -additive (in the weak topology).

Remark 10.7. It follows that $E(\emptyset) = 0$, that E is finitely additive, that $E(\Delta) \leq I$, and that E is σ -additive also in the strong topology. $\int_{\Omega} f(\omega) E(d\omega)$ can be defined as for PVMs; for real-valued f, the integral is a self-adjoint operator A, but (unlike for PVMs on $\Omega = \mathbb{R}$) the POVM cannot be recovered from A.

Proof. The only part that does not follow in the same way as for PVMs is the strong convergence. By finite additivity,

$$E\left(\bigcup_{n=1}^{\infty} \Delta_n\right) - \sum_{n=1}^{N} E(\Delta_n) = E\left(\bigcup_{n=N+1}^{\infty} \Delta_n\right).$$
(10.19)

Call this operator T_n . By the definition of POVM, $T_n \ge 0$ and

$$0 \le E\left(\Omega \setminus \bigcup_{n=N+1}^{\infty} \Delta_n\right) = I - T_n.$$
(10.20)

That is, $0 \leq T_n \leq I$, and thus $T_n^2 \leq T_n$, so that $||T_n\psi||^2 = \langle \psi|T_n^2\psi\rangle \leq \langle \psi|T_n\psi\rangle \to 0$ as $n \to \infty$.

From a POVM $E(\cdot)$ on a set Ω one can create probability measures on Ω in the following way: Given any vector $\psi \in \mathscr{H}$ with $\|\psi\| = 1$, then

$$\mathbb{P}_{\psi}(\Delta) = \langle \psi | E(\Delta) | \psi \rangle \tag{10.21}$$

defines a probability measure $\mathbb{P}_{\psi}(\cdot)$ on Ω . To see this, note that $\langle \psi | E(\Delta) | \psi \rangle$ is a non-negative real number since $E(\Delta)$ is a positive operator, and

$$\mathbb{P}_{\psi}(\Omega) = \langle \psi | E(\Omega) | \psi \rangle = \langle \psi | I | \psi \rangle = \| \psi \|^2 = 1.$$
(10.22)

10.3 The main theorem about POVMs

It says: For every quantum physical experiment \mathscr{E} on a quantum system S whose possible outcomes lie in a space Ω , there exists a POVM $E(\cdot)$ on Ω such that, whenever S has wave function ψ at the beginning of \mathscr{E} , the random outcome Z has probability distribution given by

$$\mathbb{P}(Z \in \Delta) = \langle \psi | E(\Delta) | \psi \rangle.$$
(10.23)

Before we give a precise version and proof of this statement in the framework of Bohmian mechanics, some comments on the terminology. A POVM, when it plays the role of encoding the statistics of an experiment \mathscr{E} as described in the main theorem, is called a *generalized observable*. When E happens to be a PVM, it is called the *observable* associated with \mathscr{E} . For example, the natural PVM $P(\cdot)$ on $L^2(\mathbb{R}^d)$ is called the *position observable*, and the PVM $\mathscr{F}^{-1}P(\cdot)\mathscr{F}$ is called the *momentum observable*. When E happens to be a PVM on \mathbb{R} , so it corresponds (via the spectral theorem) to a self-adjoint operator A, then A is also called the *observable* associated with \mathscr{E} . It is also common to call \mathscr{E} a "measurement" of the observable A. The names "observable" and "measurement" are misleading, as they suggest that the outcome Z of \mathscr{E} is some quantity pertaining to S and defined already before \mathscr{E} ; however, as we will see, this is often not the case, and Z is often just a random value created by \mathscr{E} .

Definition 10.8. In Bohmian mechanics, a quantum experiment \mathscr{E} of run-time Δt on a system S consists of coupling S to an apparatus A at some time t_1 , letting $S \cup A$ evolve up to $t_1 + \Delta t$, and setting

$$Z = \zeta \left(Q_A(t_1 + \Delta t) \right), \tag{10.24}$$

where Q_A is the Bohmian configuration of A and ζ is called the *calibration function*.⁴ It is assumed that S and A are not entangled at the beginning of \mathscr{E} :

$$\Psi_{S\cup A}(t_1) = \psi_S(t_1) \otimes \phi_A(t_1) \,. \tag{10.25}$$

 \mathscr{E} is mathematically defined as a tuple $(\phi_A(t_1), H_{S\cup A}, \Delta t, \Omega, \mathfrak{A}, \zeta)$, where $\phi_A(t_1) \in L^2(\mathbb{R}^{d_A})$ with $\|\phi_A(t_1)\| = 1$ is called the "ready state" of A, $H_{S\cup A}$ is the Hamiltonian of $S \cup A$ during $[t_1, t_1 + \Delta t]$, $\Delta t > 0$, Ω is called the value space of \mathscr{E} , \mathfrak{A} is a σ -algebra on Ω , and $\zeta : \mathbb{R}^{d_A} \to \Omega$ is measurable.

Let $H^{\infty}(\mathbb{R}^d) \subset L^2(\mathbb{R}^d)$ be the space of $\psi \in C^{\infty}(\mathbb{R}^d) \cap L^2(\mathbb{R}^d)$ such that every $\partial^{\alpha} \psi \in L^2(\mathbb{R}^d)$. By the Sobolev lemma, this space is the intersection of all Sobolev spaces, and is the image under Fourier transformation of the space of functions that decrease more rapidly at infinity than 1/P(k) for any polynomial P. It is bigger than Schwartz space, though, and in particular dense in $L^2(\mathbb{R}^d)$. Note that $\psi \otimes \phi \in H^{\infty}(\mathbb{R}^{d_S+d_A})$ iff $\psi \in H^{\infty}(\mathbb{R}^{d_S})$ and $\phi \in H^{\infty}(\mathbb{R}^{d_A})$.

Theorem 10.9. (Main theorem about POVMs in Bohmian mechanics) Let \mathscr{E} be a quantum experiment on the system S such that: $H_{S\cup A} = H_0 + V$ in $L^2(\mathbb{R}^{d_S+d_A})$ with a bounded, smooth potential V, and $\phi_A(t_1) \in H^{\infty}(\mathbb{R}^{d_A})$ with $\|\phi_A(t_1)\| = 1$. Then there is a unique POVM $E(\cdot)$ on (Ω, \mathfrak{A}) such that for every $\psi_S(t_1) \in H^{\infty}(\mathbb{R}^{d_S})$ with $\|\psi_S(t_1)\| = 1$, the distribution of the outcome Z as defined in (10.24), assuming (10.25) and $Q(t_1) = (Q_S(t_1), Q_A(t_1)) \sim |\Psi(t_1)|^2$, is given by (10.23).

⁴In practice, the function ζ depends only on the macroscopic configuration Q_A , not on microscopic details. However, the arguments that follow apply to arbitrary calibration function.

Proof. Let $t_2 = t_1 + \Delta t$. By Theorem 10.2, the Bohmian trajectories exist for all times, and $Q(t_2) \sim |\Psi(t_2)|^2$. Using the notation $\tilde{P}(\cdot)$ for the natural PVM on $L^2(\mathbb{R}^{d_S+d_A})$, the last fact can be written as

$$\mathbb{P}(Q(t_2) \in \tilde{\Delta}) = \langle \Psi(t_2) | \tilde{P}(\tilde{\Delta}) | \Psi(t_2) \rangle$$
(10.26)

for all measurable $\tilde{\Delta} \subseteq \mathbb{R}^{d_S+d_A}$. Let $\tilde{\zeta}(Q_S, Q_A) = \zeta(Q_A)$ and $U = e^{-iH_{S\cup A}t/\hbar}$. Thus, for all measurable $\Delta \subseteq \Omega$,

$$\mathbb{P}(Z \in \Delta) = \mathbb{P}(Q(t_2) \in \tilde{\zeta}^{-1}(\Delta)) = \langle \Psi(t_2) | \tilde{P}(\tilde{\zeta}^{-1}(\Delta)) | \Psi(t_2) \rangle$$
(10.27)

$$= \langle \psi \otimes \phi | U^* \dot{P}(\hat{\zeta}^{-1}(\Delta)) U | \psi \otimes \phi \rangle = \langle \psi | E(\Delta) | \psi \rangle_{\mathbb{R}^{d_S}}$$
(10.28)

where the operator $E(\Delta)$ on $L^2(\mathbb{R}^{d_S})$ is defined by first mapping ψ to $\tilde{P}(\tilde{\zeta}^{-1}(\Delta)) \psi \otimes \phi$ and then taking the *partial inner product* with ϕ . The partial inner product of a function $\Psi(x, y)$ with the function $\phi(y)$ is a function of x defined as

$$\langle \phi | \Psi \rangle_y(x) = \int dy \, \phi^*(y) \, \Psi(x, y) \,. \tag{10.29}$$

More generally, the partial inner product with ϕ is the adjoint of $\psi \mapsto \psi \otimes \phi$, indeed the unique bounded linear mapping $L_{\phi} : \mathscr{H}_1 \otimes \mathscr{H}_2 \to \mathscr{H}_1$ such that

$$L_{\phi}(\psi \otimes \chi) = \langle \phi | \chi \rangle_2 \psi \,. \tag{10.30}$$

It has $||L_{\phi}|| = ||\phi||$ and satisfies

$$\langle \psi | L_{\phi} \Psi \rangle_{1} = \langle \psi \otimes \phi | \Psi \rangle_{1 \otimes 2} \,. \tag{10.31}$$

In this notation,

$$E(\Delta)\psi = L_{\phi} U^* \tilde{P}(\tilde{\zeta}^{-1}(\Delta)) U \psi \otimes \phi.$$
(10.32)

We check that $E(\cdot)$ is a POVM: For $\Delta = \Omega$ (the entire space), $\zeta^{-1}(\Omega) = \mathbb{R}^{d_A}$ and $\tilde{P}(\tilde{\zeta}^{-1}(\Omega)) = I$, and $E(\Omega) = I$ by (10.30). For every Δ , $E(\Delta)$ is clearly well defined and bounded, and positive by (10.31). The weak σ -additivity follows from that of $\tilde{P}(\cdot)$.

The uniqueness of $E(\cdot)$ follows from the fact that $E(\Delta)$ is uniquely determined by the values $\langle \psi | E(\Delta) | \psi \rangle$ for all ψ in a dense subspace such as H^{∞} : If $\langle \psi | T \psi \rangle = 0$ for all $\psi \in H^{\infty}$ then, by polarization, $\langle \psi | T \phi \rangle = 0$ for all $\psi, \phi \in H^{\infty}$, so $T\phi = 0$ on a dense subspace, and thus T = 0.

Remark 10.10. Factorization vs. permutation symmetry. You might worry that the factorization condition (10.25) never holds because of the symmetrization postulate (that we will discuss in detail in a later chapter): As soon as both the system and the apparatus contain electrons, the wave function has to be anti-symmetric in the electron variables q_i , in conflict with (10.25). However, (10.25) can hold nevertheless, as follows: For identical particles, the indices of the variables q_1, \ldots, q_N are mere mathematical labels, and the splitting into system and environment should not be based on these

unphysical labels but instead on regions of space, which requires considering systems with a variable number of particles (that we will also discuss in detail in a later chapter). I can give away that if $R \subseteq \mathbb{R}^3$ is a region of space such that both R and $\mathbb{R}^3 \setminus R$ have positive volume then $\mathscr{H}(\mathbb{R}^3) = \mathscr{H}(R) \otimes \mathscr{H}(\mathbb{R}^3 \setminus R)$, where $\mathscr{H}(S)$ is the fermionic (or bosonic) Fock space over $L^2(S)$, i.e., the Hilbert space of a variable number of identical particles.

Homework problem. What if factorization $\Psi_t = \psi \otimes \phi$ is not exactly satisfied, but only approximately? Then the probability distribution of the outcome Z is still approximately given by $\langle \psi | E(\cdot) | \psi \rangle$. To make this statement precise, suppose that

$$\Psi_t = c\psi \otimes \phi + \Delta \Psi, \qquad (10.33)$$

where $\|\Delta\Psi\| \ll 1$, $\|\psi\| = \|\phi\| = 1$, and $c = \sqrt{1 - \|\Delta\Psi\|^2}$ (which is close to 1). Show that, for any $\Delta \in \mathfrak{A}$,

$$\left|\mathbb{P}(Z \in \Delta) - \langle \psi | E(\Delta) | \psi \rangle \right| < 3 \|\Delta \Psi\|.$$
(10.34)

Many experiments have a run-time that is not prescribed in advance. For example, the experiment may involve waiting for a detector to click. Such experiments are covered by Theorem 10.9 as follows. It is reasonable to assume that there is an upper bound Δt on the run-time, and to expect that the apparatus will record and display the outcome Z (and the random run-time ΔT) until $t_1 + \Delta t$. Thus, we may read off Z and ΔT from $Q_A(t_1 + \Delta t)$, so the joint distribution of $(Z, \Delta T)$ is given by a POVM on $\Omega \times [0, \Delta t]$ (or perhaps on $\Omega \times [0, \Delta t] \cup \{$ "didn't finish" $\}$).

Example 10.11. Here are some examples of proper POVMs (i.e., POVMs that are not PVMs) that arise as generalized observables.

- Time of arrival: Send a particle towards a detector and measure the time at which the detector clicks. The distribution of outcomes is given by a POVM, but there is no "time operator" (while there are position operators, momentum operators, and the energy operator H), so the POVM is not a PVM.
- Position measurement with constraints: In some cases, not all square-integrable functions on \mathbb{R}^3 are possible as physical wave functions of a single particle, but only those from a suitable subspace \mathscr{H}_{phys} . For example, photon wave functions can be regarded as functions $\Psi : \mathbb{R}^3 \to \mathbb{C}^3$ obeying the constraint $\nabla \cdot \Psi = 0$. As another example, Dirac wave functions $\psi : \mathbb{R}^3 \to \mathbb{C}^4$ are usually regarded as physical only if they consist exclusively of Fourier components with positive energy, in other words, if they lie in the positive spectral subspace \mathscr{H}_{phys} of the Dirac Hamiltonian. In this case, the usual position operators and the natural PVM $P(\cdot)$ often map physical wave functions into unphysical ones, and are thus not defined as operators on the physical Hilbert space \mathscr{H}_{phys} . The problem is solved by replacing $P(\cdot)$ with $E(\cdot)$ defined by

$$E(\Delta) := P_{\text{phys}} P(\Delta) I_{\text{phys}}, \qquad (10.35)$$

where P_{phys} denotes the projection $\mathscr{H} \to \mathscr{H}_{\text{phys}}$ and I_{phys} the inclusion $\mathscr{H}_{\text{phys}} \to \mathscr{H}$. Then $E(\Delta)$ is an operator on $\mathscr{H}_{\text{phys}}$, and $E(\cdot)$ is a proper POVM on \mathbb{R}^3 .

• Fuzzy measurements: An ideal detector would determine exactly whether the particle is in the region $\Delta \subseteq \mathbb{R}^3$. Real detectors, however, have an inaccuracy, corresponding to a proper POVM $E(\cdot)$ that arises from the natural PVM $P(\cdot)$ by convolving with a "bump function" f (for example a Gaussian):

$$E(\Delta) = \int_{\Delta} dx \int_{\mathbb{R}^3} f(y-x) P(dy) \,. \tag{10.36}$$

Remark 10.12. Many physicists would say that a key part of the weirdness of quantum mechanics is its use of non-commuting operators as observables. It is often suggested that a coherent, "classical" reality should correspond to observables that are functions on the state space, which always commute, while non-commuting operators are a mathematical expression of something like a paradoxical reality. This impression contrasts starkly with the fact that in Bohmian mechanics, where we have a coherent, "classical" reality, operators occur in much the same way as in ordinary quantum mechanics. What is the resolution of this tension?

Part of the answer lies in the role that operators play: The word "measurement" suggests that the experiment is merely revealing a value (a "hidden variable") that was well defined already before the experiment. In orthodox quantum mechanics it is often emphasized that quantum experiments are not like that. Equally often in orthodox quantum mechanics, though, one speaks as if quantum experiments were like that. For example, when saying that the particle "was found to be at location x" or "was found to have energy E," or when using the word "measurement," or the word "observable" (or even "quantity") when talking about an operator. Often, orthodox physicists know that belief in this type of hidden variables can't be consistently entertained and nevertheless can't abandon that belief in their hearts.

In Bohmian mechanics, most experiments do not reveal pre-existing values but rather create random values (the exception being position measurements). The role of the operator, or the POVM, is to determine its probability distribution, encoding that information about the experiment (e.g., about ϕ , ζ , $H_{S\cup A}$, \tilde{P}) that is relevant for determining the distribution of the outcome Z, given any ψ . From this perspective, the mystery vanishes.

Corollary 10.13. There is no experiment with $Z = \psi$, i.e., that can measure the wave function of a given system.

Proof. Suppose there was. Then, for any given ψ , Z is deterministic, i.e., its probability distribution is concentrated on a single point, $\mathbb{P}(Z \in \Delta) = \delta_{\psi}(\Delta) = 1_{\psi \in \Delta}$. The dependence of this distribution on ψ is not quadratic, and thus not of the form $\langle \psi | E(\Delta) | \psi \rangle$ for any POVM E.

Corollary 10.14. There is no experiment in Bohmian mechanics that can measure the velocity of a particle with unknown wave function.

Proof. Again, the distribution of the velocity $\text{Im}\nabla\psi/\psi(Q)$ with $Q \sim |\psi|^2$ is not quadratic in ψ .

While it is common to call every self-adjoint operator an "observable," there is actually no guarantee that the mapping $\mathscr{E} \mapsto E$ is surjective, or that the image of this mapping contains all PVMs on \mathbb{R} (which would mean that every self-adjoint operator possesses a "measurement" procedure). For example, let Q and P be the position and momentum operators, $Q\psi(q) = q\psi(q)$ and $P\psi(q) = -i\hbar\partial\psi/\partial q$; then Q+P and QP+PQare symmetric operators on the appropriate domains and possess self-adjoint extensions, but it is completely unclear whether there are experiments that are "measurements" of these operators.

Different experiments can have the same POVM. This defines an equivalence relation between experiments; two equivalent experiments $\mathscr{E}, \mathscr{E}'$ will have the same distribution of their outcomes when applied to two systems with wave function ψ . In Bohmian mechanics, we can ask the question whether \mathscr{E} and \mathscr{E}' (with fixed $Q_A(t_1)$ and $Q'_A(t_1)$) would actually yield the same outcome, Z = Z', when applied to the *same* system. The answer is in general no. This fact shows that the outcome does not depend on the "observable" (bad, bad name) alone but depends on the experiment.

10.4 The momentum operator and the asymptotics of the Schrödinger equation for large times

"Position measurements" usually consist of detecting the particle. "Momentum measurements" usually consist of letting the particle move freely for a while and then measuring its position.⁵ We now explain how this method leads to the momentum operator.

The asymptotic Bohmian velocity is

$$AV = \lim_{t \to \infty} \frac{Q(t)}{t} \tag{10.37}$$

if it exists. Its distribution is

$$\mathbb{P}(AV \in \Delta) = \lim_{t \to \infty} \mathbb{P}(Q(t)/t \in \Delta) = \lim_{t \to \infty} \langle \psi_t | P(t\Delta) | \psi_t \rangle .$$
(10.38)

We leave aside the question whether AV exists and focus only on the last expression, the asymptotic distribution.

Theorem 10.15. Let $\psi(t, x)$ be a solution of the free Schrödinger equation and $\Delta \subseteq \mathbb{R}^d$ measurable. Then

$$\lim_{t \to \infty} \langle \psi_t | P(t\Delta) | \psi_t \rangle = \lim_{t \to \infty} \int_{t\Delta} |\psi(t,x)|^2 \, dx = \int_{m\Delta/\hbar} |\widehat{\psi}_0(k)|^2 \, dk \tag{10.39}$$

⁵Alternatively, one lets the particle collide with another particle, makes a "momentum measurement" on the latter, and makes theoretical reasoning about what the momentum of the former must have been.

Proof. Omitted. I've taken the theorem from Stefan Teufel's lecture notes on "Mathematical Physics 2." It is also discussed in Reed and Simon, Volume 2, page 60–61. \Box

That is, the probability density of the "momentum" p = mAV is $\hbar^{-d} |\hat{\psi}_0(\frac{p}{\hbar})|^2$. In particular, its expectation value is

$$\langle p \rangle := \int \hbar k |\widehat{\psi}_0(k)|^2 \, dk = \int \psi_0^*(x) (-i\hbar\partial_x) \psi_0(x) \, dx = \langle \psi_0| - i\hbar\partial_x |\psi_0\rangle \,, \qquad (10.40)$$

and the higher moments are

$$\langle p^{\alpha} \rangle = \int (\hbar k)^{\alpha} |\widehat{\psi}_0(k)|^2 \, dk = \langle \psi_0 | (-i\hbar\partial_x)^{\alpha} | \psi_0 \rangle \,. \tag{10.41}$$

These relations motivate calling $P = -i\hbar\partial_x$ the momentum operator.

10.5 Heisenberg's uncertainty relation

The variance of the momentum distribution for the initial wave function $\psi \in L^2(\mathbb{R})$ (in one dimension) is

$$\sigma_P^2 := \left\langle \left(p - \langle p \rangle \right)^2 \right\rangle = \left\langle \psi \middle| \left(P - \langle \psi | P | \psi \rangle \right)^2 \middle| \psi \right\rangle \tag{10.42}$$

The position distribution $|\psi(x)|^2$ has expectation

$$\langle Q(0) \rangle = \int x |\psi(x)|^2 \, dx = \langle \psi | X | \psi \rangle \tag{10.43}$$

with the position operator $X\psi(x) = x\psi(x)$ and higher moments

$$\langle Q(0)^{\alpha} \rangle = \int x^{\alpha} |\psi(x)|^2 \, dx = \langle \psi | X^{\alpha} | \psi \rangle \,, \tag{10.44}$$

so the variance of the position distribution $|\psi(x)|^2$ is

$$\sigma_X^2 := \int (x - \langle Q(0) \rangle)^2 |\psi(x)|^2 \, dx = \left\langle \psi \Big| \left(X - \langle \psi | X | \psi \rangle \right)^2 \Big| \psi \right\rangle. \tag{10.45}$$

More generally, let A be a self-adjoint operator in \mathscr{H} , $P_A(\cdot)$ its spectral PVM and $\psi \in \mathscr{H}$ with $\|\psi\| = 1$, then the probability distribution $\mathbb{P}_A(\Delta) = \langle \psi | P_A(\Delta) | \psi \rangle$ has expectation

$$E_A = \int_{\mathbb{R}} \lambda \langle \psi | P_A(\lambda) | \psi \rangle = \langle \psi | A | \psi \rangle , \qquad (10.46)$$

which exists iff $\psi \in \mathscr{D}(A)$, and variance

$$\sigma_A^2 = \int_{\mathbb{R}} \left(\lambda - (E_A)\right)^2 \langle \psi | P_A(\lambda) | \psi \rangle = \| (A - E_A) \psi \|^2, \qquad (10.47)$$

which also exists iff $\psi \in \mathscr{D}(A)$. When σ_A doesn't exist we set it equal to ∞ .

Theorem 10.16. (Heisenberg uncertainty relation) For $\psi \in L^2(\mathbb{R})$ with $\|\psi\| = 1$,

$$\sigma_X \, \sigma_P \ge \frac{\hbar}{2} \,. \tag{10.48}$$

For self-adjoint operators A and B in \mathscr{H} and $\psi \in \mathscr{D}(A) \cap \mathscr{D}(B)$ with $A\psi \in \mathscr{D}(B)$, $B\psi \in \mathscr{D}(A)$, and $\|\psi\| = 1$,

$$\sigma_A \sigma_B \ge \frac{1}{2} \left| \langle \psi | [A, B] | \psi \rangle \right|. \tag{10.49}$$

Proof. We first prove the second statement. Let $\Delta A := A - E_A$, $\Delta B = B - E_B$, $f = \Delta A \psi$, $g = \Delta B \psi$. Then $\sigma_A = ||f||$ an $\sigma_B = ||g||$. Since, for any $z \in \mathbb{C}$,

$$\left|\frac{1}{2i}(z-z^*)\right| = |\operatorname{Im} z| \le |z|,$$
 (10.50)

we have that

$$\left|\frac{1}{2i}\left(\langle f|g\rangle - \langle g|f\rangle\right)\right| \le \left|\langle f|g\rangle\right| \le \|f\| \,\|g\|\,,\tag{10.51}$$

which means

$$\frac{1}{2} \left| \langle \psi | [\Delta A, \Delta B] | \psi \rangle \right| \le \sigma_A \, \sigma_B \,. \tag{10.52}$$

Since E_A and E_B are numbers, we have that $[\Delta A, \Delta B] = [A, B]$.

For $\mathscr{H} = L^2(\mathbb{R})$, A = X, B = P, we have that $[X, P] = i\hbar$ wherever both XP and PX are defined. For $\psi \in \mathscr{D}(X) \cap \mathscr{D}(P)$ (but not necessarily $X\psi \in \mathscr{D}(P)$, $P\psi \in \mathscr{D}(X)$), it is still true that $\langle X\psi|P\psi \rangle - \langle P\psi|X\psi \rangle = i\hbar \langle \psi|\psi \rangle$ (integrate by parts). For $\psi \notin \mathscr{D}(X)$, $\sigma_X = \infty$, so (10.48) is still true.

10.6 The quantum measurement problem

This is a problem about orthodox quantum mechanics that illustrates the need for Bohmian mechanics or a similar theory behind quantum mechanics by showing that the orthodox attitudes become incoherent when analyzing an experiment.

Consider a "quantum measurement of the observable with operator A." Realistically, there are only finitely many possible outcomes, so A should have finite spectrum. Consider the system formed by the object together with the apparatus. I write Ψ for its wave function. This system is isolated, or we can make sure it is. So it should evolve according to the Schrödinger equation during the experiment, which begins (say) at t_1 and ends at t_2 . It is reasonable to assume that

$$\Psi(t_1) = \psi(t_1) \otimes \phi \tag{10.53}$$

with $\psi = \psi(t_1)$ the wave function of the object before the experiment and ϕ a wave function representing a "ready" state of the apparatus. Note that ψ can be written as a linear combination (superposition) of eigenfunctions of A,

$$\psi = \sum_{\alpha} \psi_{\alpha} \quad \text{with} \quad A\psi_{\alpha} = \alpha\psi_{\alpha} \,.$$
(10.54)

If the object's wave function is an eigenfunction (proportional to) ψ_{α} then the outcome is certain to be α . Set $\Psi_{\alpha}(t_1) = \psi_{\alpha} \otimes \phi$. Then $\Psi_{\alpha}(t_2)$ must represent a state in which the apparatus displays the outcome α . Since the Schrödinger equation is linear, the wave function of object and apparatus together at t_2 is

$$\Psi(t_2) = \sum_{\alpha} \Psi_{\alpha}(t_2) , \qquad (10.55)$$

a superposition of states corresponding to different outcomes—and not a random state corresponding to a unique outcome. This is the measurement problem. The upshot is that there is a conflict between the following assumptions:

- There is a unique outcome.
- The wave function is a complete description of a system's physical state.
- The evolution of the wave function of an isolated system is always given by the Schrödinger equation.

Thus, we have to drop one of these assumptions. If we drop the first, we opt for a many-worlds picture, in which all outcomes are realized, albeit in parallel worlds. If we drop the second, we opt for further variables as in Bohmian mechanics, where the state at time t is described by the pair (Q_t, ψ_t) . If we drop the third, we opt for replacing the Schrödinger equation by a non-linear evolution (as in the GRW = Ghirardi–Rimini–Weber approach).

11 Density operators

Let

$$\mathbb{S}(\mathscr{H}) = \{ \psi \in \mathscr{H} : \|\psi\| = 1 \}$$
(11.1)

denote the unit sphere in Hilbert space. It is common to write $\langle \psi |$ for the linear form $\phi \mapsto \langle \psi | \phi \rangle$, and $|\psi \rangle$ for ψ . (The latter notation also allows us to write $|1\rangle$, $|2\rangle$, ... instead of ψ_1, ψ_2, \ldots) Then $|\psi \rangle \langle \psi |$ is an operator, viz. $\phi \mapsto \langle \psi | \phi \rangle \psi$; for $\psi \in \mathbb{S}(\mathscr{H})$, this is the projection $P_{\mathbb{C}\psi}$ to the 1d subspace spanned by ψ . Note also that $\langle \psi |$ applied to $|\phi \rangle$ gives $\langle \psi | \phi \rangle$, and $\langle \psi | A | \phi \rangle = \langle \psi | A \phi \rangle$, so no ambiguity arises.

Suppose that (by whatever mechanism) we have generated a random wave function $\Psi \in \mathbb{S}(\mathscr{H})$ with distribution given by the probability measure μ on $\mathbb{S}(\mathscr{H})$. Then for any experiment \mathscr{E} with POVM $E(\cdot)$, the probability distribution of the outcome Z is

$$\operatorname{Prob}(Z \in \Delta) = \mathbb{E}\langle \Psi | E(\Delta) | \Psi \rangle = \int_{\mathbb{S}(\mathscr{H})} \mu(d\psi) \langle \psi | E(\Delta) | \psi \rangle = \operatorname{tr}(\rho_{\mu} E(\Delta)), \quad (11.2)$$

where \mathbb{E} means *expectation*,

$$\rho_{\mu} = \mathbb{E} |\Psi\rangle \langle \Psi| = \int_{\mathbb{S}(\mathscr{H})} \mu(d\psi) |\psi\rangle \langle \psi|$$
(11.3)

is called the *density operator* or *density matrix* (rarely: *statistical operator*) of the distribution μ , and tr means the trace. Equation (11.2) is called the *trace formula*.

11.1 Trace of an operator

Let \mathscr{H} be separable. The trace of an operator T is defined to be the sum of the diagonal elements of its matrix representation $T_{nm} = \langle n|T|m \rangle$ relative to an arbitrary orthonormal basis $\{|n\rangle\}$,

$$\operatorname{tr} T = \sum_{n=1}^{\infty} \langle n | T | n \rangle .$$
(11.4)

However, the series may not converge, or may converge for one orthonormal basis and not for another. That is why one splits the rigorous definition in two steps.

Step 1: If $T \in \mathscr{B}(\mathscr{H})$ is a positive operator then its trace is defined by (11.4), which is either a nonnegative real number or $+\infty$.

Proposition 11.1. This value does not depend on the choice of orthonormal basis. The trace has the following properties:

- (i) $\operatorname{tr}(A+B) = \operatorname{tr} A + \operatorname{tr} B$
- (*ii*) $\operatorname{tr}(\lambda A) = \lambda \operatorname{tr} A$ for all $\lambda \ge 0$
- (*iii*) $tr(UAU^{-1}) = tr A$ for any unitary operator U

(iv) If $0 \le A \le B$ then tr $A \le \text{tr } B$.

Proof. Let $\phi = \{\phi_n\}$ and $\psi = \{\psi_m\}$ be two ONBs of \mathscr{H} .

$$\operatorname{tr}_{\phi}(A) = \sum_{n=1}^{\infty} \langle \phi_n | A \phi_n \rangle = \sum_{n=1}^{\infty} \| A^{1/2} \phi_n \|^2$$
(11.5)

$$=\sum_{n=1}^{\infty} \left(\sum_{m=1}^{\infty} \left| \langle \psi_m | A^{1/2} \phi_n \rangle \right|^2 \right)$$
(11.6)

$$=\sum_{m=1}^{\infty} \left(\sum_{n=1}^{\infty} \left| \langle A^{1/2} \psi_m | \phi_n \rangle \right|^2 \right)$$
(11.7)

$$= \sum_{m=1}^{\infty} \|A^{1/2}\psi_m\|^2 = \sum_{m=1}^{\infty} \langle \psi_m | A\psi_m \rangle = \operatorname{tr}_{\psi} A \,. \tag{11.8}$$

Interchanging the sums is allowed because all terms are non-negative. For (iii), note that $\{U\phi_n\}$ is an ONB, too; (i), (ii), (iv) are obvious.

Example 11.2. If P is the projection to the closed subspace $X \subseteq \mathscr{H}$ then tr $P = \dim X \leq \infty$.

Step 2: This definition is extended to non-positive operators as follows.

Definition 11.3. An operator $T \in \mathscr{B}(\mathscr{H})$ belongs to the *trace class* \mathscr{I}_1 iff the positive operator $|T| = \sqrt{T^*T}$ has finite trace.

Proposition 11.4. \mathscr{I}_1 is a vector space. If $A \in \mathscr{I}_1$ and $B \in \mathscr{B}(\mathscr{H})$ then $AB, BA \in \mathscr{I}_1$ and $A^* \in \mathscr{I}_1$. If $A \in \mathscr{I}_1$ and $\{|n\rangle\}$ is any ONB, then tr $A := \sum_{n=1}^{\infty} \langle n|A|n \rangle$ converges absolutely and is independent of the ONB. $\mathscr{I}_1(\mathscr{H})$ is a Banach space with respect to the trace norm

$$||T||_1 := \operatorname{tr}\sqrt{T^*T} \,. \tag{11.9}$$

(i) tr is linear

(*ii*)
$$tr(A^*) = (tr A)^*$$

(iii) $\operatorname{tr}(AB) = \operatorname{tr}(BA)$ for $A \in \mathscr{I}_1$ and $B \in \mathscr{B}(\mathscr{H})$.

Proof. Reed and Simon, Volume 1, pages 207–211.

- **Remark 11.5.** If, for some ONB $\{|n\rangle\}$, $\sum_{n=1}^{\infty} |\langle n|A|n\rangle| < \infty$ then A does not have to be in the trace class.
 - If $T \in \mathscr{B}(\mathscr{H})$ is positive then |T| = T, and $T \in \mathscr{I}_1$ iff $\operatorname{tr} T < \infty$. The two definitions of trace (one for positive operators, one for trace class operators) obviously agree.

• By property (iii), the trace is invariant under cyclic permutation of any number of factors $A \in \mathscr{I}_1, B, \ldots, Z \in \mathscr{B}(\mathscr{H})$:

$$tr(AB\cdots YZ) = tr(ZAB\cdots Y). \qquad (11.10)$$

In particular tr(ABC) = tr(CAB), which is, however, not always the same as tr(CBA).

- If there exists an ONB of eigenvectors of A, then tr A is the sum of the eigenvalues, counted with multiplicity (= degree of degeneracy).
- The trace of a self-adjoint operator $A \in \mathscr{I}_1$ is real. A self-adjoint operator lies in the trace class if and only its spectrum is discrete and bounded, all nonzero eigenvalues have finite multiplicity, and the sum of the eigenvalues (with multiplicity) is finite (i.e., converges absolutely).

11.2 The trace formula in quantum mechanics

In order to verify the trace formula (11.2), note first that

$$\operatorname{tr}\left(|\psi\rangle\langle\psi|E(\Delta)\right) = \langle\psi|E(\Delta)|\psi\rangle \tag{11.11}$$

because, if we choose the basis $\{|n\rangle\}$ in (11.4) such that $|1\rangle = \psi$, then the summands in (11.4) are $\langle n|\psi\rangle\langle\psi|E(\Delta)|n\rangle$, which for n = 1 is $\langle\psi|E(\Delta)|\psi\rangle$ and for n > 1 is zero because $\langle n|1\rangle = 0$. By linearity, we also have that

$$\operatorname{tr}\left(\sum_{i=1}^{M} p_{i} |\psi_{i}\rangle\langle\psi_{i}| E(\Delta)\right) = \sum_{i=1}^{M} p_{i}\langle\psi|E(\Delta)|\psi\rangle, \qquad (11.12)$$

which yields (11.2) for any μ that is concentrated on finitely many points ψ_i on $\mathbb{S}(\mathscr{H})$. To allow arbitrary probability measures μ on $\mathbb{S}(\mathscr{H})$ (equipped with its Borel σ -algebra, which are the Borel sets in \mathscr{H} that are subsets of $\mathbb{S}(\mathscr{H})$), we need the following.

- **Proposition 11.6.** (i) The integral $\int \mu(d\psi) |\psi\rangle \langle \psi|$ is well defined as a weak integral, i.e., there is a unique $\rho \in \mathscr{B}(\mathscr{H})$ such that for all $\phi \in \mathscr{B}(\mathscr{H})$, $\langle \phi | \rho \phi \rangle = \int \mu(d\psi) \langle \phi | \psi \rangle \langle \psi | \phi \rangle$.
 - (ii) $\rho \ge 0$
- (iii) $\rho \in \mathscr{I}_1$ and tr $\rho = 1$.

(iv) For any
$$E \in \mathscr{B}(\mathscr{H})$$
, $\operatorname{tr}(\rho E) = \int \mu(d\psi) \langle \psi | E\psi \rangle$. This proves (11.2).

Proof. (i) The mapping $B(\phi, \chi) = \int \mu(d\psi) \langle \phi | \psi \rangle \langle \psi | \chi \rangle$ is well defined, is a sesqui-linear form $\mathscr{H} \times \mathscr{H} \to \mathbb{C}$, and is bounded, $|B(\phi, \chi)| \leq ||\phi|| ||\chi||$. By the Riesz representation theorem, there is an operator $\rho \in \mathscr{B}(\mathscr{H})$ such that $B(\phi, \chi) = \langle \phi | \rho | \chi \rangle$. We have seen before that bounded operators are uniquely determined by their diagonal elements.

(ii) follows from $\langle \phi | \rho \phi \rangle = \int \mu(d\psi) \langle \phi | \psi \rangle \langle \psi | \phi \rangle = \int \mu(d\psi) |\langle \phi | \psi \rangle|^2 \ge 0.$

(iii) $\rho \in \mathscr{I}_1$ follows when we have shown that tr $\rho < \infty$. By the Fubini–Tonnelli theorem,

$$\operatorname{tr} \rho = \sum_{n=1}^{\infty} \langle n | \rho | n \rangle = \sum_{n=1}^{\infty} \int \mu(d\psi) |\langle n | \psi \rangle|^2$$
(11.13)

$$= \int \mu(d\psi) \sum_{n=1}^{\infty} |\langle n|\psi\rangle|^2 = \int \mu(d\psi) \|\psi\|^2 = 1.$$
 (11.14)

(iv) By the Fubini theorem,

$$\operatorname{tr}(\rho E) = \sum_{n=1}^{\infty} \langle n | \rho E | n \rangle = \sum_{n=1}^{\infty} \int \mu(d\psi) \langle n | \psi \rangle \langle \psi | E | n \rangle$$
(11.15)

$$= \int \mu(d\psi) \sum_{n=1}^{\infty} \langle n|\psi\rangle \langle \psi|E|n\rangle$$
(11.16)

$$= \int \mu(d\psi) \operatorname{tr}\left(|\psi\rangle\langle\psi|E\right) = \int \mu(d\psi)\langle\psi|E|\psi\rangle.$$
(11.17)

To justify the interchange of summation and integration, we need that the integrand is in L^1 , i.e.,

$$\int \mu(d\psi) \sum_{n=1}^{\infty} \left| \langle n|\psi\rangle \langle \psi|E|n\rangle \right| \leq \int \mu(d\psi) \left(\sum_{n} |\langle n|\psi\rangle|^2 \right)^{1/2} \left(\sum_{n} |\langle E^*\psi|n\rangle|^2 \right)^{1/2}$$
(11.18)

$$= \int \mu(d\psi) \|\psi\| \|E^*\psi\| \le \|E\| < \infty.$$
 (11.19)

Now let us draw conclusions from the formula (11.2). It implies that the distribution of the outcome Z depends on μ only through ρ_{μ} . Different distributions μ_a, μ_b can have the same $\rho = \rho_{\mu_a} = \rho_{\mu_b}$; for example, if $\mathscr{H} = \mathbb{C}^2$ then the uniform distribution over $\mathbb{S}(\mathscr{H}) = \mathbb{S}^3$ has $\rho = \frac{1}{2}I$, and for every orthonormal basis $|\phi_1\rangle, |\phi_2\rangle$ of \mathbb{C}^2 the probability distribution

$$\frac{1}{2}\delta_{\phi_1} + \frac{1}{2}\delta_{\phi_2} \tag{11.20}$$

also has $\rho = \frac{1}{2}I$. Such two distributions μ_a, μ_b will lead to the same distribution of outcomes for any experiment, and are therefore *empirically indistinguishable*.

We can turn this result into an argument showing that there must be facts we cannot find out by experiment: Suppose I choose μ to be either μ_a or μ_b , then I choose $n = 10^4$ points ψ_i on $\mathbb{S}(\mathscr{H})$ at random independently with μ , then I prepare *n* systems with wave functions ψ_i , and then I hand these systems to you with the challenge to determine whether $\mu = \mu_a$ or $\mu = \mu_b$. As a consequence of (11.2), you cannot determine that by means of experiments on the *n* systems. On the other hand, nature knows the right answer, as nature must remember the wave function of each system; after all, I might keep records of each ψ_i and can predict that system *i* will in a quantum measurement of $|\psi_i\rangle\langle\psi_i|$ yield the outcome 1. Thus, there is a fact in nature (whether $\mu = \mu_a$ or $\mu = \mu_b$) that we cannot discover empirically. Nature can keep a secret.

If the random vector Ψ evolves according to the Schrödinger equation, $\Psi_t = e^{-iHt/\hbar}\Psi$, the distribution changes into μ_t and the density matrix into

$$\rho_t = e^{-iHt/\hbar} \rho e^{iHt/\hbar} \,. \tag{11.21}$$

In analogy to the Schrödinger equation, this can be written as a differential equation,

$$\frac{d\rho_t}{dt} = -\frac{i}{\hbar} [H, \rho_t], \qquad (11.22)$$

known as the von Neumann equation. ρ_t is weakly differentiable, i.e., $t \mapsto \langle \psi | \rho_t | \psi \rangle$ is differentiable, for $\psi \in \mathscr{D}(H)$, and (11.22) is true in the weak sense for such ψ .

If $\rho = |\psi\rangle\langle\psi|$ with $||\psi|| = 1$, then ρ is usually called a *pure quantum state*, otherwise a *mixed quantum state*. A probability distribution μ has $\rho = |\psi\rangle\langle\psi|$ if and only if μ is concentrated on $\mathbb{C}\psi$, i.e., $\Psi = e^{i\Theta}\psi$ with a random global phase factor.

As we have seen, a density matrix ρ is always a positive operator with $\operatorname{tr} \rho = 1$. Conversely, every positive operator ρ with $\operatorname{tr} \rho = 1$ is a density matrix, i.e., $\rho = \rho_{\mu}$ for some probability distribution μ on $\mathbb{S}(\mathscr{H})$. This is because any positive operator $A \in \mathscr{I}_1$ is bounded and thus self-adjoint; A has a discrete spectrum, and thus there is an orthonormal basis $\{|\phi_n\rangle : n \in \mathbb{N}\}$ of eigenvectors of ρ with eigenvalues $p_n \in [0, \infty)$, and

$$\sum_{n} p_n = \operatorname{tr} \rho = 1.$$
 (11.23)

Now let μ be the distribution that gives probability p_n to ϕ_n ; its density matrix is just the ρ we started with.

11.3 Reduced density operators

There is another way in which density matrices arise, leading to what is called the *reduced density matrix*. Suppose that the system under consideration consists of two parts, system a and system b, so that its Hilbert space is $\mathscr{H} = \mathscr{H}_a \otimes \mathscr{H}_b$, and that the experiment \mathscr{E} has a POVM of the form

$$E(\Delta) = E_a(\Delta) \otimes I_b , \qquad (11.24)$$

where I_b is the identity on \mathscr{H}_b .

Homework problem. Prove that in Bohmian mechanics, an experiment in which the apparatus interacts only with system a but not with system b has a POVM of the form

(11.24). To this end, adapt the proof of the main theorem of POVMs. Suppose the experiment \mathscr{E} begins at time t_1 and ends at time t_2 , and suppose the wave function of the apparatus, system a, and system b at time t_1 is $\Psi(t_1) = \phi \otimes \psi$ with $\psi \in \mathscr{H}_a \otimes \mathscr{H}_b$, so $\Psi(t_1) \in \mathscr{H}_{app} \otimes \mathscr{H}_a \otimes \mathscr{H}_b$. Assume further that the outcome Z is a function ζ of the configuration Q_{app} of the apparatus at time t_2 .

In the case (11.24), the distribution of the outcome is

$$\operatorname{Prob}(Z \in \Delta) = \langle \psi | E(\Delta) | \psi \rangle = \operatorname{tr}(\rho_{\psi} E_a(\Delta))$$
(11.25)

with the reduced density matrix of system a

$$\rho_{\psi} = \mathrm{tr}_b |\psi\rangle \langle\psi|\,,\tag{11.26}$$

where tr_b means the partial trace over \mathscr{H}_b .

11.4 Partial trace

Homework problem 11.7. For $T_a \in \mathscr{B}(\mathscr{H}_a)$ and $T_b \in \mathscr{B}(\mathscr{H}_b)$, there is a unique operator $T_a \otimes T_b \in \mathscr{B}(\mathscr{H}_a \otimes \mathscr{H}_b)$ satisfying

$$(T_a \otimes T_b)(\psi_a \otimes \psi_b) = (T_a \psi_a) \otimes (T_b \psi_b)$$
(11.27)

for all $\psi_a \in \mathscr{H}_a$ and $\psi_b \in \mathscr{H}_b$. It has the following properties.

- (i) $(T_a \otimes T_b)^* = T_a^* \otimes T_b^*$
- (ii) $(T_a \otimes T_b)(S_a \otimes S_b) = (T_a S_a) \otimes (T_b S_b)$
- (iii) If $T_a \ge 0$ and $T_b \ge 0$ then $T_a \otimes T_b \ge 0$. In that case, $\operatorname{tr}(T_a \otimes T_b) = (\operatorname{tr} T_a)(\operatorname{tr} T_b)$.
- (iv) If $T_a \in \mathscr{I}_{1,a} := \mathscr{I}_1(\mathscr{H}_a)$ and $T_b \in \mathscr{I}_{1,b}$ then $T_a \otimes T_b \in \mathscr{I}_{1,a\otimes b} := \mathscr{I}_1(\mathscr{H}_a \otimes \mathscr{H}_b)$. In that case, $\operatorname{tr}(T_a \otimes T_b) = (\operatorname{tr} T_a)(\operatorname{tr} T_b)$ and $||T_a \otimes T_b||_{1,a\otimes b} = ||T_a||_{1,a}||T_b||_{1,b}$.
- (v) $\mathscr{I}_{1,a\otimes b} = \overline{\operatorname{span}}\{T_a \otimes T_b : T_a \in \mathscr{I}_{1,a}, T_b \in \mathscr{I}_{1,b}\}$ (closure in the trace norm).

Example 11.8. When (and only when) the systems a, b do not interact, the Hamiltonian is of the form

$$H = H_a \otimes I_b + I_a \otimes H_b \,, \tag{11.28}$$

and the propagator $U_t = e^{-iHt/\hbar}$ is of the form

$$U_t = U_{a,t} \otimes U_{b,t} \tag{11.29}$$

with $U_{a/b,t} = e^{-iH_{a/b}t/\hbar}$.

Definition 11.9. tr_b is the unique linear mapping $\mathscr{I}_{1,a\otimes b} \to \mathscr{I}_{1,a}$ such that

$$\|\operatorname{tr}_b T\|_{1,a} \le \|T\|_{1,a\otimes b} \tag{11.30}$$

for all $T \in \mathscr{I}_{1,a\otimes b}$, and

$$\operatorname{tr}_b(T_a \otimes T_b) = \operatorname{tr}(T_b) T_a \,. \tag{11.31}$$

for all $T_a \in \mathscr{I}_{1,a}$ and $T_b \in \mathscr{I}_{1,b}$.

Here is an explicit construction of tr_b . Let $\{\phi_n^a\}$ be an ONB of \mathscr{H}_a and $\{\phi_m^b\}$ an ONB of \mathscr{H}_b . Then $\{\phi_n^a \otimes \phi_m^b\}$ is an ONB of $\mathscr{H}_a \otimes \mathscr{H}_b$. If $T \in \mathscr{I}_{1,a \otimes b}$ then

$$\operatorname{tr}_{b} T = \sum_{m=1}^{\infty} \langle \phi_{m}^{b} | T | \phi_{m}^{b} \rangle, \qquad (11.32)$$

where the inner product is a partial inner product, so that each term in the sum is an operator in $\mathscr{I}_{1,a}$,

$$\langle \phi_m^b | T | \phi_m^b \rangle \psi_a = L_{\phi_m^b} T \psi_a \otimes \phi_m^b \,, \tag{11.33}$$

and the series converges in the trace norm. Equivalently, we can characterize the operator $S = \operatorname{tr}_b T$ by its matrix elements $\langle \phi_n^a | S | \phi_k^a \rangle$:

$$\langle \phi_n^a | \operatorname{tr}_b T | \phi_k^a \rangle = \sum_{m=1}^{\infty} \langle \phi_n^a \otimes \phi_m^b | T | \phi_k^a \otimes \phi_m^b \rangle , \qquad (11.34)$$

where the inner products on the right hand side are inner products in $\mathscr{H}_a \otimes \mathscr{H}_b$.

The partial trace has the following properties:

- (i) $\operatorname{tr}(\operatorname{tr}_b(T)) = \operatorname{tr}(T)$. Here, the first tr symbol means the trace in \mathscr{H}_a , the second one the partial trace, and the last one the trace in $\mathscr{H}_a \otimes \mathscr{H}_b$. This property follows from (11.34) by setting k = n and summing over n.
- (ii) $\operatorname{tr}_b(T^*) = (\operatorname{tr}_b T)^*$.
- (iii) If $T \ge 0$ then $\operatorname{tr}_b T \ge 0$.
- (iv) $\operatorname{tr}_b \left[S(T_a \otimes I_b) \right] = (\operatorname{tr}_b S)T_a.$

From properties (iv) and (i) we obtain that

$$\operatorname{tr}\left[S(T_a \otimes I_b)\right] = \operatorname{tr}\left[(\operatorname{tr}_b S)T_a\right].$$
(11.35)

Setting $S = |\psi\rangle\langle\psi|$ and $T_a = E_a(\Delta)$, we find that $\operatorname{tr}_b S = \rho_{\psi}$ and

$$\langle \psi | E_a(\Delta) \otimes I_b | \psi \rangle = \operatorname{tr} \left[|\psi\rangle \langle \psi | (E_a(\Delta) \otimes I_b) \right] = \operatorname{tr} \left[\rho_{\psi} E_a(\Delta) \right], \qquad (11.36)$$

which proves (11.25).

From properties (i) and (iii) it follows also that ρ_{ψ} is a positive operator with trace 1. Conversely, every positive operator ρ on \mathscr{H}_a with tr $\rho = 1$ arises as a reduced density matrix. To see this, we use that ρ must have an orthonormal basis $\{\phi_n^a\}$ of eigenvectors with eigenvalues $p_n \in [0, \infty)$ such that $\sum p_n = 1$. Let $\{\phi_n^b\}$ be an arbitrary orthonormal basis of \mathscr{H}_b and set

$$\psi = \sum_{n} \sqrt{p_n} \, \phi_n^a \otimes \phi_n^b \,. \tag{11.37}$$

Then

$$\rho_{\psi} = \mathrm{tr}_b |\psi\rangle \langle \psi| \tag{11.38}$$

$$=\sum_{n,n',m} \langle \phi_m^b | \phi_n^a \otimes \phi_n^b \rangle \sqrt{p_n p_{n'}} \langle \phi_{n'}^a \otimes \phi_{n'}^b | \phi_m^b \rangle$$
(11.39)

$$=\sum_{n,n',m}\delta_{nm}|\phi_n^a\rangle\sqrt{p_np_{n'}}\langle\phi_{n'}^a|\delta_{n'm}$$
(11.40)

$$=\sum_{m} |\phi_m^a\rangle p_m \langle \phi_m^a| = \rho.$$
(11.41)

Statistical density matrices as in (11.3) and reduced density matrices can be combined: If $\Psi \in \mathscr{H}_a \otimes \mathscr{H}_b$ is random then set

$$\rho = \mathbb{E} \operatorname{tr}_{b} |\Psi\rangle \langle \Psi| = \operatorname{tr}_{b} \mathbb{E} |\Psi\rangle \langle \Psi|. \qquad (11.42)$$

Statistical and reduced density matrices sometimes get confused; here is an example. Consider again the wave function of the measurement problem,

$$\Psi = \sum_{\alpha} \Psi_{\alpha} \,, \tag{11.43}$$

the wave function of an object and an apparatus after a "quantum measurement" of the "observable" $A = \sum \alpha P_{\alpha}$. (In (11.43), Ψ is capitalized not because it is random—it isn't—but because it is the wave function of the "big" system including the apparatus.) Suppose that Ψ_{α} , the contribution corresponding to the outcome α , is of the form

$$\Psi_{\alpha} = c_{\alpha} \,\psi_{\alpha} \otimes \phi_{\alpha} \,, \tag{11.44}$$

where $c_{\alpha} = \|P_{\alpha}\psi\|$, ψ is the initial object wave function ψ , $\psi_{\alpha} = P_{\alpha}\psi/\|P_{\alpha}\psi\|$, and ϕ_{α} with $\|\phi_{\alpha}\| = 1$ is a wave function of the apparatus after having "measured" α . Since the ϕ_{α} have disjoint supports in configuration space, they are mutually orthogonal; thus, they are a subset of some orthonormal basis $\{\phi_n\}$. The reduced density matrix of the object is

$$\rho_{\Psi} = \operatorname{tr}_{b} |\Psi\rangle \langle \Psi| = \sum_{n} \langle \phi_{n} |\Psi\rangle \langle \Psi| \phi_{n} \rangle = \sum_{\alpha} |c_{\alpha}|^{2} |\psi_{\alpha}\rangle \langle \psi_{\alpha}|.$$
(11.45)
This is the same density matrix as the statistical density matrix associated with the probability distribution

$$\mu = \sum_{\alpha} |c_{\alpha}|^2 \,\delta_{\psi_{\alpha}} \,, \tag{11.46}$$

since

$$\rho_{\mu} = \sum_{\alpha} |c_{\alpha}|^2 |\psi_{\alpha}\rangle \langle\psi_{\alpha}|. \qquad (11.47)$$

It is sometimes claimed that this fact solves the measurement problem. The argument is: From (11.43) follows (11.45), which is the same as (11.47), which means that the system's wave function has distribution (11.46), so we have a random outcome α . This argument is incorrect, as the mere fact that two situations—one with Ψ as in (11.43), the other with random ψ' —define the same density matrix for the object does not mean the two situations are physically equivalent. And obviously from (11.43), the situation after a quantum measurement involves neither a random outcome nor a random wave function. As John Bell once put it, "and is not or."

It is often taken as the definition of *decoherence* that the reduced density matrix is (approximately) diagonal in the eigenbasis of the relevant operator A.

It is common to call a density matrix that is a 1-dimensional projection a *pure state* and otherwise a *mixed state*, even if it is a reduced density matrix and thus does not arise from a mixture (i.e., from a probability distribution μ).

Proposition 11.10. A reduced density matrix ρ_{ψ} is pure if and only if ψ is a tensor product, *i.e.*, there are $\chi^a \in \mathscr{H}_a$ and $\chi^b \in \mathscr{H}_b$ such that $\psi = \chi^a \otimes \chi^b$.

Proof. The "if" part is clear; to prove the "only if" part, suppose that $\rho_{\psi} = |\phi\rangle\langle\phi|$, set $\phi^a = \chi$, choose an orthonormal basis $\{\phi_n^a\}$ of \mathscr{H}_a such that $\phi_1^a = \chi$, choose an orthonormal basis $\{\phi_n^b\}$ of \mathscr{H}_b , and expand ψ in the product basis:

$$\psi = \sum_{nm} c_{nm} \,\phi_n^a \otimes \phi_m^b \,. \tag{11.48}$$

Then

$$\rho_{\psi} = \sum_{n,n'} \left(\sum_{m} c_{nm} c^*_{n'm} \right) |\phi^a_n\rangle \langle \phi^a_{n'}|, \qquad (11.49)$$

and since we know $\rho_{\psi} = |\phi_1^a\rangle \langle \phi_1^a|$, we can read off that

$$\sum_{m} c_{nm} c_{n'm}^* = \delta_{n1} \delta_{n'1} .$$
 (11.50)

By considering $n = n' \neq 1$ we obtain that $c_{nm} = 0$ for all m and all $n \neq 1$. Thus,

$$\psi = \phi_1^a \otimes \sum_m c_{1m} \, \phi_m^b \,, \tag{11.51}$$

which is what we wanted to show.

12 More about the quantum formalism

A usual formulation of the quantum formalism may read as follows.

- When a quantum system is isolated, its wave function ψ_t evolves according to the Schrödinger equation.
- Suppose that at time t an observer measures the observable \mathscr{A} on a quantum system with wave function $\psi = \psi_t$. With this observable is associated a self-adjoint operator A with finite spectrum. The possible outcomes of the measurement are the eigenvalues α of A; the actual outcome Z is random, and the probability that it assumes the value α is

$$\mathbb{P}(Z=\alpha) = \|P_{\alpha}\psi\|^2, \qquad (12.1)$$

where P_{α} is the projection to the eigenspace associated with α . In case $Z = \alpha$, the system's wave function immediately after the measurement is

$$\psi_{t+} = \lim_{\varepsilon \searrow 0} \psi_{t+\varepsilon} = \frac{P_{\alpha}\psi}{\|P_{\alpha}\psi\|}.$$
(12.2)

The part of the formalism we have not yet talked about is (12.2), known as the *projection postulate* or the *collapse of the wave function*. I will focus on two questions: (a) How does the collapse of the wave function come out of Bohmian mechanics without being included among the fundamental postulates of the theory? (b) In the situation described in the rules above, the POVM $E(\cdot)$ is given by a PVM, $E(\Delta) = \sum_{\alpha \in \Delta} P_{\alpha}$, the spectral PVM of A. Thus, the situation is a special case (called an *ideal measurement*). How can the quantum formalism be formulated in the general case?

12.1 Collapse in Bohmian mechanics

Let us use the same notation as in chapter 10.6: Let Ψ be the wave function of the object and the apparatus together, which evolves unitarily (that is, without collapse!), starting at the beginning of the experiment from

$$\Psi(t_1) = \psi(t_1) \otimes \phi(t_1) \tag{12.3}$$

with $\phi = \phi(t_1)$ a ready state of the apparatus. Write $\psi = \psi(t_1)$ as a superposition of eigenfunctions of A,

$$\psi = \sum_{\alpha} c_{\alpha} \psi_{\alpha} \quad \text{with} \quad A \psi_{\alpha} = \alpha \psi_{\alpha} , \quad \|\psi_{\alpha}\| = 1.$$
(12.4)

Suppose that if the object's wave function is an eigenfunction ψ_{α} then the outcome is certain to be α . Set $\Psi_{\alpha}(t_1) = \psi_{\alpha} \otimes \phi$. Suppose that $\Psi_{\alpha}(t_2)$ represents a state in which the apparatus displays the outcome α ; that is, $\Psi_{\alpha}(t_2)$ is concentrated in the region $\tilde{\Delta}_{\alpha}$

of configuration space in which the pointer particles point to the value α . By unitarity, the wave function of object and apparatus together at t_2 is

$$\Psi(t_2) = \sum_{\alpha} c_{\alpha} \Psi_{\alpha}(t_2) \tag{12.5}$$

and $\|\Psi_{\alpha}(t_2)\| = 1.$

In Bohmian mechanics, since the $\Psi_{\alpha} := \Psi_{\alpha}(t_2)$ have disjoint supports, and since the particle configuration Q has distribution $|\Psi(t_2)|^2$, the probability that Q lies in $\tilde{\Delta}_{\alpha}$ is

$$\mathbb{P}(Q \in \tilde{\Delta}_{\alpha}) = \int_{\tilde{\Delta}_{\alpha}} dx \, |\Psi(x)|^2 = \int_{\mathbb{R}^{3N}} dx \, |c_{\alpha}\Psi_{\alpha}(x)|^2 = |c_{\alpha}|^2 \,, \tag{12.6}$$

which agrees with the prediction of the quantum formalism for the probability of the outcome α .

Since the Ψ_{α} are macroscopically different they will not overlap significantly in the future (until the time when the universe reaches thermal equilibrium); this fact is called *decoherence*. If Q lies in the support of one among several disjoint packets then only the packet containing Q is relevant, by Bohm's law of motion (1.14), to determining dQ/dt. Thus, as long as the packets stay disjoint, only the packet containing Q is relevant to the trajectories of the particles, and all other packets could be replaced by zero without affecting the trajectories. That is why we can replace Ψ by Ψ_{α} , with α the actual outcome. Furthermore, if

$$\Psi_{\alpha} = \psi_{\alpha} \otimes \chi \tag{12.7}$$

then the object has wave function ψ_{α} , as claimed in the quantum formalism above.

Remark 12.1. Another example of a POVM is provided by a sequence of ideal measurements, first one corresponding to A_1 , then another corresponding to A_2 , and so on, up to A_n (each with finite spectrum). Suppose that these experiments are carried out one immediately after another, so that we can neglect the unitary time evolution in between. Note that the operators A_i need not commute with each other, as they are not "measured simultaneously," but in a specified order. The sequence of outcomes forms a vector in \mathbb{R}^n , whose distribution is given by a POVM $E(\cdot)$ that can be constructed from the spectral PVMs $P_i(\cdot)$ of the A_i as follows:

$$E\{(\lambda_1,\ldots,\lambda_n)\} = P_1\{\lambda_1\}\cdots P_n\{\lambda_n\}P_n\{\lambda_n\}\cdots P_1\{\lambda_1\}.$$
 (12.8)

In case the A_i commute with each other, $E(\cdot)$ is a PVM on \mathbb{R}^n . When the A_i do not commute then $E(\cdot)$ is a proper POVM. To make the setting more general, we can allow that the choice of the second operator A_2 depends on the outcome of the first experiment. To take this into account, replace $P_i\{\lambda_i\}$ in (12.8) by $P_{i,\lambda_1,\ldots,\lambda_{i-1}}\{\lambda_i\}$.

12.2 General formulation of the collapse rule

In a situation less idealized than the "ideal measurement," Ψ_{α} will not factorize as in (12.7) into a wave function of the object and one of the apparatus; the two will be entangled. If a second, later experiment acts on the object (but not on the first apparatus) then its statistics will be determined by the reduced density matrix of the object,

$$\rho' = \operatorname{tr}_A |\Psi_{\alpha}\rangle \langle \Psi_{\alpha}| \,. \tag{12.9}$$

Here Ψ_{α} is the (normalized) part of the wave function on the set $\tilde{\Delta}_{\alpha}$ of configurations in which the pointer points to α . We don't need the assumption that there exist wave functions ψ_{α} for which the outcome is certain. For any $\psi = \psi(t_1)$, we can obtain ρ' as follows: Form $\Psi(t_1) = \psi \otimes \phi$, evolve it to $\Psi(t_2) = U\Psi(t_1)$, apply the projection $\tilde{P}(\tilde{\Delta}_{\alpha})$ (\tilde{P} the natural PVM of object and apparatus together), trace out the apparatus, and normalize (to make the trace equal to 1):

$$\rho' = \frac{1}{\mathcal{N}} \operatorname{tr}_A \left(\tilde{P}(\tilde{\Delta}_{\alpha}) U | \psi \otimes \phi \rangle \langle \psi \otimes \phi | U^* \tilde{P}(\tilde{\Delta}_{\alpha}) \right) =: \frac{1}{\mathcal{N}} \mathscr{C}_{\alpha}(|\psi\rangle \langle \psi|)$$
(12.10)

with \mathcal{N} the normalizing factor, i.e., the trace of the expression that follows it. In fact,

$$\mathbb{P}(Z=\alpha) = \mathcal{N}, \qquad (12.11)$$

because

$$\mathcal{N} = \operatorname{tr} \mathscr{C}_{\alpha}(|\psi\rangle\langle\psi|) = \operatorname{tr}\left(\tilde{P}(\tilde{\Delta}_{\alpha})U|\psi\otimes\phi\rangle\langle\psi\otimes\phi|U^{*}\tilde{P}(\tilde{\Delta}_{\alpha})\right)$$
(12.12)

$$= \langle \psi \otimes \phi | U^* \tilde{P}(\tilde{\Delta}_{\alpha}) U | \psi \otimes \phi \rangle = \langle \psi | E(\alpha) | \psi \rangle.$$
(12.13)

The mapping

$$\mathscr{C}_{\alpha}(\rho) = \operatorname{tr}_{A}\left(\tilde{P}(\tilde{\Delta}_{\alpha})U\left[\rho \otimes |\phi\rangle\langle\phi|\right]U^{*}\tilde{P}(\tilde{\Delta}_{\alpha})\right)$$
(12.14)

can be defined for any density matrix ρ , not just $|\psi\rangle\langle\psi|$, in fact for any trace class operator ρ , and is a linear mapping $\mathscr{I}_1 \to \mathscr{I}_1$. Such a mapping is often called a *superoperator* because it maps density matrices to density matrices (up to a normalizing factor) rather than wave functions to wave functions. If ψ is random with distribution μ , then

$$\mathbb{P}(Z=\alpha) = \operatorname{tr} \mathscr{C}_{\alpha}(\rho_{\mu}), \qquad (12.15)$$

and the density matrix governing the distribution of a second, later experiment on the object is

$$\rho' = \frac{\mathscr{C}_{\alpha}(\rho_{\mu})}{\operatorname{tr}\mathscr{C}_{\alpha}(\rho_{\mu})}.$$
(12.16)

By comparison with the formalism above, an ideal measurement has

$$\mathscr{C}_{\alpha}(\rho) = P_{\alpha}\rho P_{\alpha} \,. \tag{12.17}$$

Both this \mathscr{C}_{α} and the one defined by (12.14) map positive operators to positive operators; they are even *completely positive*:

Definition 12.2. A mapping $\mathscr{C} : \mathscr{I}_1(\mathscr{H}_1) \to \mathscr{I}_1(\mathscr{H}_2)$ is completely positive iff for every $d = 1, 2, \ldots, \mathscr{C} \otimes I_d : \mathscr{I}_1(\mathscr{H}_1 \otimes \mathbb{C}^d) \to \mathscr{I}_1(\mathscr{H}_2 \otimes \mathbb{C}^d)$ maps positive operators to positive operators.

The main theorem about superoperators says: With every quantum physical experiment \mathscr{E} on a quantum system S whose possible outcomes lie in a finite set Ω and that does not act on any system that S is entangled with, there is associated a family of completely positive superoperators \mathscr{C}_{α} such that $\sum_{\alpha \in \Omega} \mathscr{C}_{\alpha}$ is trace-preserving. The POVM $E(\cdot)$ of \mathscr{E} is determined by $\operatorname{tr}(TE(\alpha)) = \mathscr{C}_{\alpha}(T)$ for all $T \in \mathscr{I}_1$. That is, if S has (reduced or statistical) density matrix ρ at the beginning of \mathscr{E} , the random outcome Z of \mathscr{E} has probability distribution given by

$$\mathbb{P}(Z \in \Delta) = \operatorname{tr} \mathscr{C}_{\alpha}(\rho) \,. \tag{12.18}$$

Moreover, conditional on $Z = \alpha$, the (reduced statistical) density matrix of S right after the experiment is

$$\rho' = \frac{\mathscr{C}_{\alpha}(\rho)}{\operatorname{tr} \mathscr{C}_{\alpha}(\rho)} \,. \tag{12.19}$$

13 Spin and representations of the rotation group SO(3)

In this chapter, we leave out most proofs and many details. A full discussion can be found in R. Sexl and H. Urbantke: *Relativity, Groups, Particles* (Springer-Verlag 2001).

SO(3) is the group of rotations in \mathbb{R}^3 around the origin. Its elements are 3×3 matrices R that are orthogonal, $R^t R = RR^t = I$ (where R^t means the transpose of R), and have det R = 1. SO(3) is a *Lie group*, i.e., a group and a differentiable manifold such that the group multiplication $(g, h) \mapsto gh$ and inversion $g \mapsto g^{-1}$ are C^{∞} mappings. SO(3) is compact and has (real) dimension 3.

When rotating the coordinate axes according to R, the wave function $\psi \in \mathscr{H} = L^2(\mathbb{R}^3)$ has to be replaced by

$$\psi'(x) = \psi(R^{-1}x). \tag{13.1}$$

The relation $\psi' = U_R \psi$ defines a unitary operator U_R on \mathscr{H} , and $R \mapsto U_R$ is a unitary representation of SO(3) on \mathscr{H} . Generally, a representation of a group G is a homomorphism $T: G \to GL(V)$, where V is a vector space (called the representation space) and GL(V) is the general linear group of V, i.e., the group of invertible endomorphisms of V, i.e., invertible linear operators $A: V \to V$. A Hamiltonian H is said to be rotationally symmetric or invariant under the action of SO(3) iff $U_R \mathscr{D}(H) = \mathscr{D}(H)$ and $U_R H U_R^{-1} = H$ for every $R \in SO(3)$.

The transformation law (13.1) can become more complicated in the following way. Consider, instead of $\psi : \mathbb{R}^3 \to \mathbb{C}$, a vector field $F : \mathbb{R}^3 \to \mathbb{R}^3$. Then, in the new coordinates, F has to be replaced by

$$F'(x) = RF(R^{-1}x). (13.2)$$

A tensor field M_{ij} , i.e., $M : \mathbb{R}^3 \to \mathbb{R}^3 \otimes \mathbb{R}^3 =: \mathbb{R}^{3,3}$ has to be replaced by

$$M'(x) = RM(R^{-1}x)R^t. (13.3)$$

The general pattern here is that

$$f'(x) = T_R(f(R^{-1}x)), \qquad (13.4)$$

where T is a representation of SO(3): In (13.1), T was the trivial representation $T_R = 1$; in (13.2), T was the defining representation of SO(3) on $V = \mathbb{R}^3$; in (13.3), T was the inherited representation on tensors (of the appropriate rank), say $V = \mathbb{R}^{3,3}$. For another example, V could be the space of anti-symmetric matrices (a subspace of $\mathbb{R}^{3,3}$ that is invariant under the action of SO(3)).

In quantum mechanics, if (13.4) applies with non-trivial T, then the particle is said to have *spin*. All known elementary particles (except the Higgs boson, if it exists and if it is elementary) have spin. The word is not to be taken literally; even in Bohmian mechanics, where the word "particle" is taken literally and particles have positions, they do not spin around their axes. Rather, the wave function is not a scalar; instead, it could be, e.g., a vector $(\psi : \mathbb{R}^3 \to \mathbb{C}^3)$ or tensor $(\psi : \mathbb{R}^3 \to \mathbb{C}^{3,3})$. In fact, these cases do not occur in nature; instead, other representation spaces V occur that are called *spin spaces*.

To find representation spaces, it is convenient to use *infinitesimal rotations*. For any Lie group, one defines the elements of the tangent space T_I at the neutral element I to be the *infinitesimal generators* of the group; from the group structure it inherits the structure of a *Lie algebra*, i.e., a vector space X together with a bilinear mapping $[\cdot, \cdot]: X \times X \to X$ satisfying the *Jacobi identity*

$$[[A, B], C] + [[C, A], B] + [[B, C], A] = 0.$$
(13.5)

For a Lie group contained in a $GL(\mathbb{R}^n)$, such as SO(3), its Lie algebra can be identified with an appropriate space of $n \times n$ matrices, and the operation $[\cdot, \cdot]$ is indeed the commutator.

Proposition 13.1. The Lie algebra so(3) of SO(3) consists of the anti-symmetric 3×3 matrices. That is, if R(t) is a smooth curve in SO(3) with R(0) = I then dR/dt(t = 0) is anti-symmetric.

Proof. Differentiate $I = R^t(t)R(t)$ to obtain $0 = \dot{R}^t R + R^t \dot{R}$, then set t = 0 to obtain $0 = \dot{R}^t + \dot{R}$.

It is convenient to use the following basis of so(3):

$$\Lambda_1 := \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & -1 \\ 0 & 1 & 0 \end{pmatrix}, \quad \Lambda_2 := \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ -1 & 0 & 0 \end{pmatrix}, \quad \Lambda_3 := \begin{pmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}.$$
(13.6)

Whenever a basis B of a Lie algebra is given, the coefficients in the relation

$$[B_i, B_j] = \sum_k c_{ijk} B_k \tag{13.7}$$

determine the operation $[\cdot, \cdot]$ completely. For so(3), we have the following fundamental commutation relations for the generators of the rotation group:

$$[\Lambda_i, \Lambda_j] = \sum_{k=1}^3 \varepsilon_{ijk} \Lambda_k \tag{13.8}$$

with ε_{ijk} anti-symmetric and $\varepsilon_{123} = 1$.

When a unitary representation U of SO(3) is given, it induces a representation u of so(3). In fact, for $A \in so(3)$, exp(At) is a 1-parameter subgroup of SO(3), which will be mapped to a 1-parameter group $U_{exp(At)} = e^{-iJt}$, whose generator is J = u(A). The three operators $J_i := u(\Lambda_i)$ are called the *angular momentum operators* and satisfy the commutation relations

$$[J_i, J_j] = \sum_{k=1}^3 i\varepsilon_{ijk} J_k \tag{13.9}$$

as a consequence of (13.8). For example, in the representation U corresponding to (13.1),

$$J_i = \sum_{j,k=1}^{3} \varepsilon_{ijk} x_j \left(-i\hbar\partial_k\right).$$
(13.10)

Theorem 13.2. Every continuous unitary representation of a compact Lie group in a Hilbert space is an orthogonal sum of irreducible subrepresentations. Every continuous irreducible representation of a compact Lie group in a Hilbert space is finite-dimensional.

Proof. See, e.g., J. A. Dieudonné, Treatise on Analysis, vol. 5, Academic Press (1977).

Proposition 13.3. There is, up to unitary equivalence, one irreducible representation of so(3) for every dimension $d = \dim \mathscr{H}$; it is called the spin-s representation with $s = \frac{d-1}{2}$.

Proof. See the book of Sexl and Urbantke, pages 187–189.

The spin- $\frac{1}{2}$ representation of so(3) is the one that applies to electrons and quarks; $\sigma_i = 2u(\Lambda_i)$ are self-adjoint complex 2 × 2 matrices known as the *Pauli spin matrices*,

$$\sigma_1 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad \sigma_2 = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \quad \sigma_3 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}.$$
(13.11)

This representation is called the *spinor representation*, and the elements of the representation space $V_{1/2} \cong \mathbb{C}^2$ are called *spinors* and denoted ψ_A , ϕ_B or the like.

There is one more twist in the story: The representations of so(3) do not directly correspond to representations of SO(3). That is because SO(3) is not simply connected; it is doubly connected, and its universal covering group is SU(2). The representations of su(2) = so(3) induce unitary representations of SU(2); for integer spin, they induce unitary representations of SO(3), but not for half-odd spin. However, they induce projective-unitary representations of SO(3).

In more detail, a rotation $R \in SO(3)$ can be written as $\exp(A)$ for some $A = \sum_i a_i \Lambda_i = \boldsymbol{a} \cdot \boldsymbol{\Lambda} \in so(3)$; A is unique up to addition of $2\pi nA/\|\boldsymbol{a}\|$, $n \in \mathbb{Z}$. Since so(3) = su(2), the same expression $\exp(A)$ can be interpreted in SU(2), but there $\exp(A + 2\pi A/\|\boldsymbol{a}\|) \neq \exp(A)$, only $\exp(A + 4\pi A/\|\boldsymbol{a}\|) = \exp(A)$. The spin- $\frac{1}{2}$ representation provides a unitary representation of SU(2), but with R it associates *two* operators on \mathbb{C}^2 that differ by a sign:

$$U_R = \pm \exp(\boldsymbol{a} \cdot \boldsymbol{\sigma}) \,. \tag{13.12}$$

13.1 The Pauli equation

A wave function of N electrons is a function $\psi : \mathbb{R}^{3N} \to (\mathbb{C}^2)^{\otimes N}$ and has 2^N complex components. It evolves according to the so-called *Pauli Hamiltonian*

$$H\psi(x) = \left(\frac{1}{2m}\sum_{k=1}^{N} \left(\boldsymbol{\sigma}_{(k)} \cdot \left(-i\hbar\nabla_{k} - \boldsymbol{A}(x_{k})\right)\right)^{2} + V(x)\right)\psi(x)$$
(13.13)

with V(x) the electric potential, A the magnetic vector potential, and

$$\boldsymbol{\sigma}_{(k)} = I \otimes I \otimes \cdots \bigotimes_{\substack{k \text{-th factor}}} \boldsymbol{\sigma} \otimes \cdots \otimes I.$$
(13.14)

The term

$$\sum_{k=1}^{N} \left(\boldsymbol{\sigma}_{(k)} \cdot \left(-i\hbar \nabla_{k} - \boldsymbol{A}(x_{k}) \right) \right)^{2}$$
(13.15)

in (13.13) can be re-written as

$$\sum_{k=1}^{N} \left(-i\hbar \nabla_k - \boldsymbol{A}(x_k) \right)^2 - \sum_{k=1}^{N} \hbar \boldsymbol{\sigma}_{(k)} \cdot \boldsymbol{B}(x_k)$$
(13.16)

with $\boldsymbol{B} = \nabla \times \boldsymbol{A}$ the magnetic field (see, e.g., C. Cohen-Tannoudji, B. Diu, and F. Laloë, *Quantum Mechanics, Volume II*, Wiley (1977), page 991).