Spectral theory in Hilbert spaces (ETH Zürich, FS 09)

E. Kowalski

ETH Zürich – D-MATH, Rämistrasse 101, 8092 Zürich, Switzerland
E-mail address: kowalski@math.ethz.ch

Contents

Chapter 1. Introduction and motivations 1
1.1. What is spectral theory 1
1.2. Examples 2
1.3. Motivation for spectral theory 8
1.4. Prerequisites and notation 9

Chapter 2. Review of spectral theory and compact operators 16
2.1. Banach algebras and spectral theory 16
2.2. Compact operators on a Hilbert space 20

Chapter 3. The spectral theorem for bounded operators 34
3.1. Continuous functional calculus for self-adjoint operators 35
3.2. Spectral measures 40
3.3. The spectral theorem for self-adjoint operators 42
3.4. Projection-valued measures 48
3.5. The spectral theorem for normal operators 55

Chapter 4. Unbounded operators on a Hilbert space 57
4.1. Basic definitions 57
4.2. The graph, closed and closable operators 60
4.3. The adjoint 63
4.4. Criterion for self-adjointness and for essential self-adjointness 68
4.5. Basic spectral theory for unbounded operators 70
4.6. The spectral theorem 74

Chapter 5. Applications, I: the Laplace operator 79
5.1. Definition and self-adjointess issues 79
5.2. Positive operators and the Friedrichs extension 80
5.3. Two basic examples 82
5.4. Survey of more general cases 93

Chapter 6. Applications, II: Quantum Mechanics 99
6.1. Postulates and foundations 99
6.2. Stone's theorem 101
6.3. The free particle 108
6.4. The uncertainty principle 111
6.5. The harmonic oscillator 112
6.6. The interpretation of quantum mechanics 119

Bibliography 125

CHAPTER 1

Introduction and motivations

This script follows up on a standard course in Functional Analysis and builds on the principles of functional analysis to discuss one of the must useful and widespread among its applications, the analysis, through spectral theory, of linear operators $T : H_1 \to H_2$ between Hilbert spaces.

The emphasis of the course is on developing a clear and intuitive picture, and we intend a leisurely pace, with frequent asides to analyze the theory in the context of particularly important examples.

1.1. What is spectral theory

The goal of spectral theory, broadly defined, can be described as trying to "classify" all linear operators, and the restriction to Hilbert space occurs both because it is much easier – in fact, the general picture for Banach spaces is barely understood today –, and because many of the most important applications belong to this simpler setting. This may seem like luck, but recall that Hilbert spaces are distinguished among Banach spaces by being most closely linked to plane (and space) Euclidean geometry, and since Euclidean geometry seems to be a correct description of our universe at many scales, it is not so surprising, perhaps, that whenever infinite-dimensional arguments are needed, they should also be close to this geometric intuition.

One may think of different ways of "classifying" linear operators. Finite-dimensional linear algebra suggests that two linear maps T_1, $T_2 : H_1 \to H_2$ which are linked by a formula

$$T_2 \circ U_1 = U_2 \circ T_1, \tag{1.1}$$

for some invertible operators $U_i : H_i \to H_i$, share many similar properties. In the finite-dimensional case, this is because the U_i correspond to changing basis in H_i, which should be an operation that does not affect the intrinsic properties of the operators. This interpretation fails in general for infinite-dimensional spaces where no good theory of bases exists, but the definition still has interest, and one may try to describe all operators $H_1 \to H_2$ up to such relations.

Similarly, one can refine this idea if $H_1 = H_2 = H$, a single space, by considering that two operators $T_1, T_2 : H \to H$ should be in the same class if there is a single invertible $U : H \to H$ such that

$$T_2 \circ U = U \circ T_1, \qquad \text{i.e.} \qquad T_2 = UT_1U^{-1}. \tag{1.2}$$

Again, the interpretation of U as a change of basis is not available, but the notion is natural.

In linear algebra, the classification problem is successfully solved by the theory of eigenvalues, eigenspaces, minimal and characteristic polynomials, which leads to a canonical "normal form" for any linear operator $\mathbf{C}^n \to \mathbf{C}^n$, for $n \geqslant 1$.

We won't be able to get such a general theory for H of infinite dimension, but it turns out that many operators of greatest interest have properties which, in the finite-dimensional case, ensure an even simpler description: they may belong to any of the special classes of operators defined on a Hilbert space by means of the *adjoint* operation $T \mapsto T^*$: normal operators, self-adjoint operators, positive operators, or unitary operators. For these classes, if $\dim H = n$, there is always an orthonormal basis $(e_1, \ldots, e_n)$ of *eigenvectors* of T with eigenvalues λ_i, and in this bases, we can write

$$T(\sum_i \alpha_i e_i) = \sum_i \alpha_i \lambda_i e_i \tag{1.3}$$

(corresponding to a diagonal matrix representation).

In the infinite-dimensional case, we can not write things as easily in general, as one sees in the basic theory of the spectrum in the Banach algebra $L(H)$. However, there is one interpretation of this representation which turns out to be amenable to great generalization: consider the linear map

$$U \begin{cases} H \to \mathbf{C}^n \\ e_i \mapsto (0, \ldots, 0, 1, 0, \ldots, 0), \text{ with a 1 in the } i\text{-th position.} \end{cases}$$

This map is a *bijective isometry*, by definition of an orthonormal basis, if $\mathbf{C}^n$ has the standard inner product, and if we define

$$T_1 \begin{cases} \mathbf{C}^n \to \mathbf{C}^n \\ (\alpha_i) \mapsto (\alpha_i \lambda_i) \end{cases}$$

then (1.3) becomes

$$T_1 \circ U = U \circ T. \tag{1.4}$$

This is obvious, but we interpret this as follows, which gives a slightly different view of the classification problem: for any finite-dimensional Hilbert space H, and normal operator T, we have found a *model space and operator* $(\mathbf{C}^n, T_1)$, such that – in the sense of the previous formula – (H, T) is *equivalent* to $(\mathbf{C}^n, T_1)$ (in fact, *unitarily equivalent*, since U is isometric).

The theory we will describe in the first chapters will be a generalization of this type of "normal form" reduction, which is the point of view emphasized in [**RS1**, Ch. VII]. This is very successful because the model spaces and operators are indeed quite simple: they are of the type $L^2(X, \mu)$ for some measure space (X, μ) (the case of $\mathbf{C}^n$ corresponds to $X = \{1, \ldots, n\}$ with the counting measure), and the operators are multiplication operators

$$T_g : f \mapsto gf$$

for some suitable function $g : X \to \mathbf{C}$.

1.2. Examples

In order to focus the coming discussions with important examples, here are some types of operators in Hilbert space.

Example 1.1 (Multiplication operators). We have already hinted at these examples: they will indeed serve as model for all (normal) operators on Hilbert space. Let (X, μ) be a finite measure space (i.e., we have $\mu(X) < +\infty$), and let $g \in L^\infty(X, \mu)$ be a bounded function. Then we have a continuous linear map

$$M_g : \begin{cases} L^2(X,\mu) \to L^2(X,\mu) \\ f \mapsto gf \end{cases}$$

indeed, we have

$$\int_X |g(x)f(x)|^2 d\mu(x) \leqslant \|g\|_\infty^2 \|f\|^2,$$

so that M_g is well-defined and is continuous with norm $\|M_g\| \leqslant \|g\|_\infty$.

Note moreover that

$$\langle M_g(f_1), f_2\rangle = \int_X g(x) f_1(x)\overline{f_2(x)} d\mu(x) = \langle f_1, M_{\bar{g}}(f_2)\rangle,$$

for all f_1, $f_2 \in L^2(X, \mu)$, and therefore the adjoint of M_g is given by

$$M_g^* = M_{\bar{g}},$$

showing among other things that M_g is self-adjoint if and only if g is (almost everywhere) real-valued.

For g_1, $g_2 \in L^\infty(X, \mu)$, we also have the obvious relation

$$M_{g_1}(M_{g_2}(f)) = g_1(g_2 f) = g_2(g_1 f) = M_{g_2}(M_{g_1}(f)),$$

so all the operators M_g for $g \in L^\infty(X, \mu)$ commute; in particular, they are all normal.

If $X \subset \mathbf{C}$ is a bounded measurable set for the Lebesgue measure μ, the case $g(x) = x$ is particularly important.

The next lemma is here for future reference: it shows that one can not construct more general (bounded) multiplication operators than those associated with bounded multipliers:

Lemma 1.2. *Let (X, μ) be a finite measure space and let g be a measurable function $X \to \mathbf{C}$. If $\varphi \mapsto g\varphi$ maps $L^2(X, \mu)$ to $L^2(X, \mu)$, not necessarily continuously, then $g \in L^\infty(X)$.*

Proof. We may first observe that the assumption implies, in fact, that $T : \varphi \mapsto g\varphi$ is continuous, by the Closed Graph Theorem: indeed, if $(\varphi_n, g\varphi_n)$ is a convergent sequence in the graph of T, so that $\varphi_n \to \varphi$, $g\varphi_n \to \psi$ in $L^2(X, \mu)$, we can extract a subsequence where both convergence hold μ-almost everywhere. But then $g\varphi_n$ converges almost everywhere to $g\varphi$ and to ψ, so that $g\varphi = \psi \in L^2(X, \mu)$, i.e., (φ, ψ) lies also in the graph of T.

Now, knowing that T is bounded, we know there is a constant $C \geqslant 0$ such that

$$\|g\varphi\|^2 \leqslant C\|\varphi\|^2$$

for any $\varphi \in L^2(X, \mu)$. Consider φ to be the characteristic function of the set

$$X_A = \{x \mid |g(x)| \geqslant A\}$$

for some $A > 0$; we obtain

$$A^2\mu(X_A) \leqslant \int_X |g(x)|^2 |\varphi(x)|^2 d\mu(x) \leqslant C \int_X |\varphi(x)|^2 d\mu(x) = C\mu(X_A).$$

If we select A so that $A^2 > C$, this implies $\mu(X_A) = 0$, i.e., g is almost everywhere $\leqslant A$, which means $g \in L^\infty(X, \mu)$. ☐

Example 1.3 (Hilbert-Schmidt operators). One standard example is given by compact operators of Hilbert-Schmidt type. Recall that for a measure space (X, μ), given a *kernel function*

$$k : X \times X \to \mathbf{C}$$

such that $k \in L^2(X \times X, \mu \times \mu)$, we have a bounded operator

$$\begin{cases} L^2(X,\mu) \to L^2(X,\mu) \\ f \mapsto T_k(f) \end{cases}$$

where

$$T_k f(x) = \int_X k(x,y) f(y) d\mu(y).$$

One knows that T_k is a compact operator (see the next chapter for a review of the definitions involved), and that its adjoint is given by $T_k^* = T_{\tilde{k}}$, where

$$\tilde{k}(x,y) = \overline{k(y,x)}$$

and in particular T_k is self-adjoint if k is real-valued and symmetric. Examples of this are $k(x,y) = |x-y|$ or $k(x,y) = \max(x,y)$ for $X = [0,1]$.

Example 1.4 (Unitary operator associated with a measure-preserving transformation). (See [**RS1**, VII.4] for more about this type of examples). Let (X, μ) be a finite measure space again, and let now

$$\phi : X \to X$$

be a *bijective* measurable map which *preserves the measure* μ, i.e., we have

$$\mu(\phi^{-1}(A)) = \mu(A) \quad \text{for all measurable set } A \subset X. \tag{1.5}$$

This setting is the basis of *ergodic theory*, and there is an associated linear map

$$U_\phi \begin{cases} L^2(X,\mu) \to L^2(X,\mu) \\ f \mapsto f \circ \phi, \end{cases}$$

which is unitary (i.e., $U_\phi^* = U_\phi^{-1}$). Indeed, U_ϕ is well-defined, and is an isometry, since for any $f \in L^2(X,\mu)$, we have

$$\int_X |U_\phi f(x)|^2 d\mu(x) = \int_X |f(\phi(x))|^2 d\mu(x) = \int_X |f(x)|^2 d\mu(x)$$

where we use the fact that (1.5) is well-known to be equivalent with the "change of variable" formula

$$\int_X f(x) d\mu(x) = \int_X f(\phi(x)) d\mu(x)$$

for all integrable (or non-negative) functions f. Checking unitarity is similar: for any $f_i \in L^2(X,\mu)$, we have

$$\begin{aligned} \langle U_\phi(f_1), f_2 \rangle &= \int_X f_1(\phi(x)) \overline{f_2(x)} d\mu(x) \\ &= \int_X f_1(y) \overline{f_2(\phi^{-1}(y))} d\mu(y) \end{aligned}$$

which shows that

$$U_\phi^* = U_{\phi^{-1}},$$

while we have also $U_{\phi^{-1}} \circ U_\phi = U_\phi \circ U_{\phi^{-1}} = \mathrm{Id}$.

As concrete examples, among many, we have:

(1) the *Bernoulli shift*: here $X = \{0,1\}^{\mathbf{Z}}$, the space of doubly-infinite sequences $(x_n)_{n\in\mathbf{Z}}$, with $x_n \in \{0,1\}$, which can be equipped with a natural infinite product Radon measure[1] μ characterized by the property that for any $k \geqslant 1$, any $\{n_1, \ldots, n_k\} \in \mathbf{Z}$ distinct and $\varepsilon_j \in \{0,1\}$ for $1 \leqslant j \leqslant k$, we have

$$\mu(\{(x_n) \mid x_{n_j} = \varepsilon_j, \text{ for } 1 \leqslant j \leqslant k\}) = \frac{1}{2^k}$$

(in particular $\mu(X) = 1$). The map

$$B \begin{cases} X \to X \\ (x_n)_{n\in\mathbf{Z}} \mapsto (x_{n+1})_{n\in\mathbf{Z}} \end{cases}$$

is bijective and preserves the measure μ (as one can check easily by the characterization above).

(2) the *rotations of the circle*: let $X = \mathbf{S}^1 = \{z \in \mathbf{C} \mid |z| = 1\}$ be the unit circle, and fix some $\theta \in \mathbf{R}$. Then the corresponding rotation is defined by

$$R_\theta \begin{cases} \mathbf{S}^1 \to \mathbf{S}^1 \\ z \mapsto e^{i\theta} z, \end{cases}$$

and if $\mathbf{S}^1$ is equipped with the Lebesgue measure dz such that

$$\int_{\mathbf{S}^1} f(z)dz = \int_0^1 f(e^{2i\pi t})dt,$$

it is immediate that R_θ is a bijection preserving the measure dz.

Example 1.5 (The Fourier transform). Let $n \geqslant 1$ be an integer. The *Fourier transform* is defined first as a linear map

$$\mathcal{F} \begin{cases} L^1(\mathbf{R}^n, dx) \to L^\infty(\mathbf{R}^n, dx) \\ f \mapsto \Bigl(t \mapsto \int_{\mathbf{R}^n} f(x) e^{-2i\pi xt} dx\Bigr) \end{cases}$$

between Banach spaces, which is clearly continuous with norm $\| \mathcal{F} \| \leqslant 1$. Remarkably, although the definition does not make sense a priori for $f \in L^2(\mathbf{R}^n, dx)$ (such a function may not be integrable), one can show the Parseval formula

$$\int_{\mathbf{R}^n} |f(x)|^2 dx = \int_{\mathbf{R}^n} |\mathcal{F} f(t)|^2 dt$$

for any f which is smooth and compactly supported on $\mathbf{R}^n$. This implies that, using the L^2-metric, $\mathcal{F}$ extends uniquely by continuity to an isometric linear map on the closure of the space of smooth compactly supported functions. A standard fact of integration theory is that this closure is the whole $L^2(\mathbf{R}^n, dx)$, and this provides us with an isometric linear operator

$$\mathcal{F} : L^2(\mathbf{R}^n, dx) \to L^2(\mathbf{R}^n, dx).$$

Moreover, one can show this extension is surjective (e.g., from the Fourier inversion formula

$$f = \mathcal{F} \tilde{f} \text{ where } \tilde{f}(x) = f(-x)$$

which is valid for any integrable f for which $\mathcal{F} f$ is also integrable, and the fact that such functions are dense in $L^2(\mathbf{R}^n, dx)$). Thus the Fourier transform is an example of

[1] I.e., for the natural product topology on X, μ is finite on compact sets.

a bijective isometry on $L^2(\mathbf{R}^n, dx)$, and the Plancherel formula generalizes the Parseval formula to show that the adjoint $\mathcal{F}^*$ is the inverse Fourier operator

$$f \mapsto \mathcal{F}\tilde{f},$$

i.e., $\mathcal{F}$ is unitary.

If the facts we mentioned above are not familiar, complete proofs can be found in many places, for instance [**RS1**, Ch. IX] or [**W**, V.2].

A few additional properties of the Fourier transform will be needed in later chapters; they are summarized in the section on notation and prerequisites below.

Example 1.6 (The Laplace operator). Consider the linear differential operator

$$\Delta : f \mapsto \frac{\partial^2 f}{\partial x^2} + \frac{\partial^2 f}{\partial y^2}$$

for $f : \mathbf{R}^2 \to \mathbf{C}$. Of course, it is not defined for all L^2 functions, but if we assume that f_1, f_2 are C^2 (or more) and both vanish outside a big disc (i.e., they have compact support), we can at least compute the inner product in $L^2(\mathbf{R}^2, dxdy)$:

$$\begin{aligned}\langle \Delta f_1, f_2\rangle &= \int_{\mathbf{R}^2} \Big(\frac{\partial^2 f_1}{\partial x^2} + \frac{\partial^2 f_1}{\partial y^2}\Big)\overline{f_2(x,y)}dxdy \\ &= \int_{\mathbf{R}}\int_{\mathbf{R}} \frac{\partial^2 f_1}{\partial x^2}(x,y)\overline{f_2(x,y)}dxdy + \int_{\mathbf{R}}\int_{\mathbf{R}} \frac{\partial^2 f_2}{\partial y^2}(x,y)\overline{f_2(x,y)}dydx.\end{aligned}$$

Integrating by parts twice each of the inner integral (in x for the first one, in y for the second), the compact support condition ensures that the "integrated" terms all vanish, and we obtain

$$\langle \Delta f_1, f_2\rangle = \langle f_1, \Delta f_2\rangle$$

for such functions. It is therefore very tempting to say that Δ is self-adjoint, but this is not warranted by the usual definition since Δ is not defined from and to a Hilbert space.

Still, temptation is very hard to resist, especially if one considers the following additional fact: let M be, formally again, the multiplication operator

$$M = M_{-4\pi^2(x^2+y^2)} : f \mapsto -4\pi^2(x^2+y^2)f,$$

which is defined and linear also for (say) compactly supported functions on $\mathbf{R}^2$. Then, using the Fourier transform

$$\mathcal{F} : L^2(\mathbf{R}^2, dxdy) \to L^2(\mathbf{R}^2, dxdy),$$

we can check by elementary integration by parts again that if f is smooth and compactly supported, we have

$$\mathcal{F}(\Delta f) = M(\mathcal{F} f), \tag{1.6}$$

or abusing notation, $\Delta = \mathcal{F}^{-1} M \mathcal{F}$, which is very similar to (1.4) with the model operator being multiplication by $-4\pi^2(x^2+y^2)$.

As we will see in Chapter 4, this type of formal computation can be given a rigorous basis: this is the theory of "unbounded" (non-continuous) linear operators between Hilbert spaces, which can be developed from the principle of looking at operators which are linear and defined on a dense subspace of a Hilbert space and satisfy certain properties, from which it is in particular possible to define an adjoint, and for which a suitable spectral theorem can be obtained. One can guess that it is necessary to be quite careful, and the following fact reinforces this: as we know (and will recall), if $T : H \to H$ is a continuous operator on a Hilbert space, and T is for instance self-adjoint, then any

$\lambda \in \sigma(T)$ – in particular any eigenvalue λ – is *real*. However, for the Laplace operator, observe that formally we can take

$$f(x, y) = e^{ax+by}$$

for arbitrary complex parameters a and b, obtaining a smooth function with

$$\Delta f = (a^2 + b^2)f,$$

and of course $a^2 + b^2$ runs over the whole complex plane as (a, b) range over $\mathbf{C}$!

It can already be guessed on our example, and it will also be seen, in the next section and in later chapters, that extending spectral theory to unbounded operators is not just a question of seeking maximal generality. For instance, the mathematical formalism of Quantum Mechanics (which we will discuss as an application of spectral theory) depends on this extension.

Example 1.7 (Diagonal operators). Another class of operators, which are in fact special cases of multiplication operators, are the diagonal operators obtained by fixing an orthonormal basis $(e_i)_{i\in I}$ of a Hilbert space H, and defining

$$T \begin{cases} H \to H \\ e_i \mapsto \alpha_i e_i \end{cases}$$

for $i \in I$, where the scalars α_i are chosen so that $\max |\alpha_i| < +\infty$. To relate this to a multiplication operator, take

$$X = (I, \text{counting measure}),$$

so that $L^2(X)$ is identified with H through the isometry

$$\begin{cases} L^2(X) \to H \\ f \mapsto \sum_{i\in I} f(i)e_i, \end{cases}$$

and (α_i) is identified with a function $g \in L^\infty(X)$ for which we have $T = M_g$.

Example 1.8 (Unitary representations of topological groups). In this final example, we consider a topological group G, i.e., a group G which is given a topology (assumed to be Hausdorff) for which the group operations

$$(x, y) \mapsto xy, \qquad x \mapsto x^{-1}$$

are both continuous. Important examples of such groups are $\mathbf{R}$ or $\mathbf{C}$ (with addition), $\mathbf{R}^\times$ or $]0, +\infty[$ (with multiplication), $GL(n, \mathbf{R})$ (with product of matrices) or its subgroup $SL(n, \mathbf{R}) = \{g \in GL(n, \mathbf{R}) \mid \det(g) = 1\}$.

A *unitary representation* of G is a group homomorphism

$$\rho : G \to U(H)$$

where H is a Hilbert space and $U(H)$ is the group of unitary operators on H. The latter is given with the *strong topology* and ρ is then assumed to be continuous for this topology, which means that for any vector $v \in H$, the map

$$\begin{cases} G \to H \\ g \mapsto \rho(g)(v) \end{cases}$$

is assumed to be continuous.

We see that a representation ρ gives us a large family of unitary operators on H, parametrized by the group G. The spectral theory, applied to those, can lead to a better understanding of ρ, and then of G. Even for G looking "simple enough", as in the examples given, the fact that $U(H)$ is a group of *unitary* operators, whereas (say) $GL(n, \mathbf{R})$ does not have this feature, means that this may bring to light new information concerning G.

As an example, assume that G has a non-trivial measure $d\mu(g)$ which is *left-invariant*: for any integrable function f and any fixed $h \in G$, we have

$$\int_G f(hg)d\mu(g) = \int_G f(g)d\mu(g),$$

(it is known that such a measure always exists, and is unique up to multiplication by a non-zero positive real number, if G is locally compact, like all the examples we gave; quite often, this measure also satisfies

$$\int_G f(gh)d\mu(g) = \int_G f(g)d\mu(g),$$

but this is not always true). Then one can show (continuity requires some epsilon-management) that defining

$$\rho(g) \begin{cases} L^2(G, d\mu) \to L^2(G, d\mu) \\ f \mapsto (h \mapsto f(g^{-1}h)) \end{cases}$$

gives a unitary representation of G. Notice that, for a given g, this operator is a special case of Example 1.4.

Another basic example arises when $G = \mathbf{R}$ with the group law given by the usual addition law and the standard topology. Then a unitary representation $\mathbf{R} \to U(H)$ is also called a *one-parameter unitary group of operators*. We will see a parametrization of all such one-parameter groups in Section 6.2, in terms of *unbounded* self-adjoint operators, and we will see that those one-parameter groups are intimately related to the mathematical formalism of Quantum Mechanics

1.3. Motivation for spectral theory

Now let's come back to a general motivating question: why should we want to classify operators on Hilbert spaces (except for the fact that the theory is quite beautiful, and that it is especially thrilling to be able to say something deep and interesting about *non continuous* linear maps)?

The basic motivation comes from the same source as functional analysis does: for applications, we often need (or want) to solve linear equations

$$T(v) = w,$$

between Banach spaces, in particular Hilbert spaces. For this purpose, having an explicit classification in terms of simple models can be very useful: first, if we have a relation like (1.1), with U_i invertible, we have

$$T_1(v) = w \text{ if and only if } T_2(v_1) = w_1, \text{ with } v_1 = U_1(v),\ w_1 = U_2(w).$$

So if we understand the "model" operator T_2 and the invertible maps U_1, U_2, we can reduce the solution of linear equations involving T_1 to those involving T_2. Similarly with (1.2) or (1.4).

Now notice that for model multiplication operators (Example 1.1) $T_2 = M_g$ on $L^2(X, \mu)$, the solution of $M_g(f) = h$ is (formally at least) immediate, namely $f = h/g$; this corresponds intuitively to diagonal systems of linear equations, and of course requires more care than this: the function g might have zeros, and the ratio h/g might not be in $L^2(X,\mu)$.

In particular, still formally, note how the Fourier transform together with (1.6) strongly suggests that one should try to solve equations involving the Laplace operator, like $\Delta f = g$, by "passing to the Fourier world". This is indeed a very fruitful idea, but obviously requires even more care since the operators involved are not continuous.

Besides this rough idea, we will see other reasons for trying to develop spectral theory the way it is done. However, this should be sufficient, at least, to launch with confidence on the work ahead...

1.4. Prerequisites and notation

Prerequisites for the script are the basic principles of functional analysis for Hilbert and Banach spaces, though for the most part it is the theory of Hilbert spaces which is most important. For many examples, it is also important to have a good understanding of integration theory on a general measure space.

Below, the most important facts are recalled and, in the next chapter, we survey and summarize the basic facts of general spectral of Banach algebras that we will use, as well as the basic theory of compact operators on a Hilbert space. However, we start with describing some notation.

1.4.1. Notation. Most notation is very standard, and we only summarize here the most common. We write $|X|$ for the cardinality of a set, and in particular $|X| = +\infty$ means that X is infinite, with no indication on the infinite cardinal involved.

In topology, we use the French traditional meaning of a compact space: a compact space is always assumed to be Hausdorff.

By $f \ll g$ for $x \in X$, or $f = O(g)$ for $x \in X$, where X is an arbitrary set on which f is defined, we mean synonymously that there exists a constant $C \geqslant 0$ such that $|f(x)| \leqslant Cg(x)$ for all $x \in X$. The "implied constant" is any admissible value of C. It may depend on the set X which is always specified or clear in context. The notation $f \asymp g$ means $f \ll g$ and $g \ll f$. On the other hand $f(x) = o(g(x))$ as $x \to x_0$ is a topological statement meaning that $f(x)/g(x) \to 0$ as $x \to x_0$. We also use the $O()$ notation in other types of expressions; the meaning should be clear: e.g., $f(x) \leqslant g(x) + O(h(x))$ for $x \in X$, means that $f \leqslant g + h_1$ in X for some (non-negative) function h_1 such that $h_1 = O(h)$. (For instance, $x \leqslant x^2 + O(1)$ for $x \geqslant 1$, but it is not true that $x - x^2 = O(1)$).

In measure theory, we will use sometimes *image measures*: if (X, Σ, μ) and (Y, Ω, ν) are two measure spaces and $f : X \to Y$ is a measurable map, the image measure $f_*(\mu)$ on (Y, Ω) is the measure such that

$$f_*(\mu)(B) = \mu(f^{-1}(B)) \tag{1.7}$$

for all $B \in \Omega$; the measurability of f shows that this is well-defined, and it is easy to check that this is a measure. The formula above generalizes easily to the general integration formula

$$\int_Y \varphi(y) df_*(\mu)(y) = \int_X \varphi(f(x)) d\mu(x) \tag{1.8}$$

for any measurable function φ on Y (in the sense that whenever one of the two integrals exists, the other also does, and they are equal).

We will denote integrals by either of the following notation:

$$\int_X f d\mu = \int_X f(x) d\mu(x).$$

In the particular case of the space $X = \mathbf{R}^n$, $n \geqslant 1$, we will denote simply by dx the Lebesgue measure on $\mathbf{R}^n$.

We also recall the definition of the support $\operatorname{supp} \mu$ of a Borel measure μ on a topological space X: this is defined by

$$\text{(1.9)} \qquad \operatorname{supp} \mu = \{x \in X \mid \mu(U) > 0 \text{ for all open neighborhood } U \text{ of } x\},$$

and is also known to be the complement of the largest open subset $U \subset X$ (for inclusion) with $\mu(U) = 0$.

1.4.2. Reminders. As often in analysis, some important arguments in spectral theory will depend on *approximation arguments*: to define various objects, in particular linear maps between Banach or Hilbert spaces (say $V \to W$), we start by performing the construction for vectors in a suitable "nice" subspace $V_0 \subset V$, and then claim that it extends to the whole space automatically. This requires two properties to be checked: the first is that the "nice" subspace V_0 be dense in V (for some norm or topology), and the second is that the construction over V_0 be continuous for the same norm on V and some norm on W for which the latter is complete.

Note that the first point (the density of V_0) is potentially independent of the desired construction and of W, so that the same choice of V_0 might be suitable for many purposes. Here are some well-known examples which we will use:

- The most famous example is probably the Weierstrass approximation theorem; we recall the following general version:

  THEOREM 1.9 (Stone-Weierstrass). *Let X be a compact topological space. Let $A \subset C(X)$ be any subalgebra of the space of complex-valued continuous functions on X such that*

  $$1 \in A, \quad f \in A \Rightarrow \bar{f} \in A,$$

  and for any two points z_1, $z_2 \in X$, there exists $f \in A$ with $f(z_1) \neq f(z_2)$. Then A is dense in $C(X)$ for the L^∞ norm $\|f\|_{C(X)} = \max_{x \in X} |f(x)|$.

  REMARK 1.10. The condition that A be stable under conjugation can sometimes be omitted, but it is necessary in general; for instance, the space A of continuous functions on the closed unit disc D of $\mathbf{C}$ which are holomorphic inside the disc forms a proper closed subalgebra of $C(D)$ which satisfies all conditions except the stability under conjugation.

  The following corollary is used particularly often:

  COROLLARY 1.11. *Let $X \subset \mathbf{R}$ be a compact set of real numbers with the induced topology. Then the space of (restrictions to X of) polynomial functions on X is dense in $C(X)$ for the L^∞ norm.*

  PROOF. It is clear that the space A of polynomial functions on X is a subalgebra of $C(X)$, containing 1, and that it separates points (the single function $x \mapsto x$ already does, and is in A). Moreover, if

  $$f(x) = \sum_{0 \leqslant j \leqslant d} \alpha(j) x^j \in A,$$

we have

$$\bar{f}(x) = \sum_{0 \leqslant j \leqslant d} \bar{\alpha}(j) x^j,$$

since $\bar{x} = x$, which means that $\bar{f} \in A$ also. The Stone-Weierstrass concludes the proof. □

Another corollary is the following:

Corollary 1.12. *The linear combinations of the functions*

$$\begin{cases} \mathbf{R} \to \mathbf{C} \\ x \mapsto \frac{1}{x-\lambda} \end{cases}$$

where λ ranges over non-real elements of $\mathbf{C}$ is dense in the space $C_0(\mathbf{R})$ of continuous functions on $\mathbf{R}$ with limit 0 at $\pm\infty$, for the supremum norm.

Proof. Consider the compact space $X = \mathbf{R} \cup \{\infty\}$, with the usual topology on $\mathbf{R}$ and with neighborhoods of ∞ defined as the complements of compact subsets (i.e., the one-point compactification of $\mathbf{R}$). It is easy to check that functions $f \in C_0(\mathbf{R})$ extend to continuous functions on X with $f(\infty) = 0$, and that

$$C(X) \simeq C_0(\mathbf{R}) \oplus \mathbf{C}$$

where the second summand corresponds to constant functions. Now the linear combinations of functions as described forms an algebra stable A_0 under conjugation (by partial-fraction decomposition), separating points, and $A = A_0 \oplus \mathbf{C} \subset C(X)$ is an algebra stable under conjugation which separates points and contains the constants. By the Stone-Weierstrass Theorem, A is therefore dense in $C(X)$. Now consider any function $f \in C_0(\mathbf{R})$; from the above, we can write f as a uniform limit over X of functions $f_n \in A$. We write

$$f_n = g_n + c_n, \qquad g_n \in A_0, \quad c_n \in \mathbf{C};$$

then we have $f_n(\infty) = c_n \to f(\infty) = 0$, and therefore $g_n = f_n - c_n \in A_0$ also converges uniformly on X to f, in particular g_n converges to f in $C_0(\mathbf{R})$. □

- The next example is a standard fact of measure theory. Recall that a Radon measure μ on a locally compact topological space X is a Borel measure such that $\mu(K) < +\infty$ for all compact subsets $K \subset X$. Then the following holds:

  Lemma 1.13. *Let X be a locally compact topological space. For any p such that $1 \leqslant p < +\infty$, the space $C_c(X)$ of compactly-supported continuous functions on X is dense in the space $L^p(X, \mu)$, for the L^p-norm. In particular, this space is dense in $L^2(X, \mu)$.*

  As a special case, this implies that the characteristic function of a subset with finite measure in X can be approached arbitrarily closely (in any L^p norm with $1 \leqslant p < +\infty$) by continuous functions with compact support.
- In the case where $X = U \subset \mathbf{R}^n$ is an open subset of euclidean space, for some $n \geqslant 1$, we have the distinguished Lebesgue measure dx, and we also have the notion of smoothness. The approximation by continuous functions with compact support can be refined:

Lemma 1.14. *Let $U \subset \mathbf{R}^n$ be an open subset of $\mathbf{R}^n$ for some $n \geqslant 1$. Let $C_c^\infty(U)$ be the space of compactly-supported functions on U which have compact support. Then for $1 \leqslant p < +\infty$, the space $C_c^\infty(U)$ is dense in $L^p(U, dx)$ for the L^p-norm.*

We also recall the Riesz-Markov theorem identifying measures with positive linear functionals on spaces of continuous functions:

Theorem 1.15. *Let X be a locally compact topological space, let $C_b(X)$ be the Banach space of bounded functions on X, with the supremum norm. Let $\ell : C_b(X) \to \mathbf{C}$ be a linear map such that $f \geqslant 0$ implies $\ell(f) \geqslant 0$. Then there exists a unique Radon measure μ such that*

$$\int_X f(x) d\mu(x) = \ell(f)$$

for all $f \in C_b(X)$.

See, e.g., [**RS1**, Th. IV.4], for the proof. We note in passing the well-known useful fact that a positive linear functional ℓ of this type is necessarily continuous: indeed, positivity implies that if f and g are real-valued and $f \leqslant g$, we have $\ell(f) \leqslant \ell(g)$. Then $|f| \leqslant \|f\|_\infty$ leads to

$$\ell(|f|) \leqslant \ell(1)\|f\|_\infty$$

while

$$-|f| \leqslant f \leqslant |f|$$

gives $|\ell(f)| \leqslant \ell(|f|) \leqslant \ell(1)\|f\|_\infty$ for f real-valued, and finally

$$|\ell(f + ig)| \leqslant |\ell(f)| + |\ell(g)| \leqslant 2\ell(1)\|f + ig\|_\infty$$

for general functions.

We end by recalling two facts concerning the Fourier transform, in addition to the definition of Example 1.5. First of all, consider the functions

$$g_{m,\sigma}(x) = \exp\left(-\tfrac{(x-m)^2}{2\sigma^2}\right)$$

for $x \in \mathbf{R}$, $m \in \mathbf{R}$, $\sigma > 0$. These are standard "gaussians", and their Fourier transforms are well-known: we have

$$\mathcal{F}\, g_{m,\sigma}(y) = \sigma\sqrt{2\pi} \exp(-2\pi^2\sigma^2 y^2 - 2i\pi m y). \tag{1.10}$$

Since this computation will be used in Chapter 6, we recall quickly how one may proceed: by an affine change of variable, one reduces to the case of $g(x) = g_{0,(2\pi)^{-1/2}}(x) = e^{-\pi x^2}$, and in that case we have

$$\mathcal{F}\, g(y) = g(y),$$

as follows by differentiating

$$\mathcal{F}\, g(y) = \int_{\mathbf{R}} e^{-\pi x^2 - 2i\pi x y} dx$$

under the integral sign: after a further integration by parts, one derives that

$$\mathcal{F}\, g'(y) = -2\pi y\, \mathcal{F}\, g(y),$$

a differential equation with solution given by

$$\mathcal{F}\, g(y) = \mathcal{F}\, g(0) e^{-\pi y^2}.$$

Finally, the value

$$\mathcal{F} g(0) = \int_{\mathbf{R}} e^{-\pi x^2} dx = 1$$

is classical.

The last fact we will use (again in Chapter 6, in particular Section 6.5) is the definition of the Schwartz space $S(\mathbf{R})$ and the fact that $\mathcal{F}$, restricted to $S(\mathbf{R})$, is a linear isomorphism with inverse given by the inverse Fourier transform. This space $S(\mathbf{R})$ is defined to be the space of C^∞ functions φ such that, for any $k \geqslant 0$ and $m \geqslant 0$, we have

$$\lim_{|x| \to +\infty} (1 + |x|)^m \varphi^{(k)}(x) = 0.$$

This space is dense in $L^2(\mathbf{R})$ for the L^2 metric (and in $L^1(\mathbf{R})$ for the L^1 metric), since it contains the dense space $C_c^\infty(\mathbf{R})$ of smooth compactly supported functions (Lemma 1.14). The usefulness of $S(\mathbf{R})$ is precisely that, however, the Fourier transform does not send compactly supported functions to compactly supported functions (in fact, no non-zero function f has the property that both f and $\mathcal{F} f$ are compactly supported).

The gaussian functions $g_{m,\sigma}$ give standard examples of functions in $S(\mathbf{R})$ which are not in $C_c^\infty(\mathbf{R})$.

We note that it is sometimes useful (and necessary in comparing with other books or papers) to use other normalizations of the Fourier transform, e.g., to consider the map

$$\mathcal{F}_1 \begin{cases} L^1(\mathbf{R}^n, dx) \to L^\infty(\mathbf{R}^n, dx) \\ f \mapsto \Bigl(t \mapsto \frac{1}{\sqrt{2\pi}} \int_{\mathbf{R}^n} f(x) e^{-i\pi x t} dx\Bigr). \end{cases}$$

We have thus

$$\mathcal{F}_1(f)(t) = \frac{1}{\sqrt{2\pi}} \mathcal{F} f\Bigl(\frac{t}{2\pi}\Bigr)$$

and one gets in particular that $\mathcal{F}_1$ is also unitary for the L^2-metric (say on $S(\mathbf{R})$) and extends to a linear isometry of $L^2(\mathbf{R})$ with inverse given by replacing $-ixt$ by ixt in the definition.

When we deal with unbounded operators on a Hilbert space, it will be useful to remember the Closed Graph Theorem:

THEOREM 1.16. *Let V, W be Banach spaces and $T : V \to W$ a linear map, not necessarily continuous. Then in fact T* is *continuous, if and only if, the graph*

$$\Gamma(T) = \{(v, w) \in V \times W \mid w = T(v)\} \subset V \times W$$

is the closed subset of $V \times W$, relative to the product metric.

A similar but slightly less well-known result will also be useful: the Hellinger-Toeplitz Theorem; however, we will state it in the course of Chapter 4.

1.4.3. Integration of Hilbert-space valued functions. The reader may skip this section in a first reading, since we will only require its contents during the proof of Stone's Theorem in Section 6.2. Precisely, we will need to use integrals of the type

$$\int_{\mathbf{R}} A(t) dt$$

where A is a function on $\mathbf{R}$ with values in a Hilbert space H, typically infinite-dimensional. Since such integrals are not usually considered in standard classes of measure theory, we explain here briefly how they are defined, and describe a few of their properties. All of the latter can be summarized by saying "the integral behaves like an integral"...

Proposition 1.17 (Integral of Hilbert-space valued functions). *Let H be a Hilbert space and let*

$$A : \mathbf{R} \longrightarrow H$$

be a function which is weakly measurable, *in the sense that for all $v \in H$, the function*

$$\begin{cases} \mathbf{R} \longrightarrow \mathbf{C} \\ t \mapsto \langle v, A(t) \rangle \end{cases}$$

is measurable, and the function

$$t \mapsto \|A(t)\|$$

is measurable.

Assume further that A is bounded *and* compactly supported, *i.e., $A(t) = 0$ outside of a compact set. Then there exists a unique vector $x \in H$ such that*

$$\langle v, x \rangle = \int_{\mathbf{R}} \langle v, A(t) \rangle dt$$

for all $v \in H$.

The vector given by this proposition is denoted

$$x = \int_{\mathbf{R}} A(t) dt.$$

Proof. This is a simple consequence of the Riesz Representation Theorem for Hilbert spaces: we define a linear form

$$\ell_A : H \longrightarrow \mathbf{C}$$

by

$$\ell_A(v) = \int_G \langle v, A(t) \rangle dt.$$

The integral is well-defined because of the assumptions on A: the integrand is measurable, bounded and compactly supported, hence integrable. Moreover

$$|\ell_A(v)| \leqslant C\|v\|$$

with

$$C = \int_G \|A(t)\| dt < +\infty$$

(again because A is compactly supported), so that ℓ_A is continuous. By the Riesz theorem, there exists therefore a unique vector $x \in H$ such that

$$\ell_A(v) = \langle v, x \rangle$$

for all $v \in H$, and this is precisely what the proposition claims. □

The integral notation is justified by the following properties:
– The integral is linear with respect to A. (This is clear.)
– For all $w \in H$, we have

$$\langle x, w \rangle = \int_{\mathbf{R}} \langle A(t), w \rangle dt \tag{1.11}$$

More generally, for any continuous linear map $T : H \rightarrow H_1$, where H_1 is another Hilbert space, we have

$$T\Big(\int_{\mathbf{R}} A(t) dt\Big) = \int_{\mathbf{R}} T(A(t)) dt. \tag{1.12}$$

(Indeed, let $y = T(x)$ be the left-hand side; then for $v \in H_1$, we have

$$\langle v, y\rangle = \langle v, T(x)\rangle = \langle T^* v, x\rangle = \int_{\mathbf{R}} \langle T^* v, A(t)\rangle dt = \int_{\mathbf{R}} \langle v, T(A(t))\rangle dt$$

which is the defining property of the right-hand side; because of the unicity of the integral, the two sides must be equal in H_1.)

– For any fixed $t_0 \in \mathbf{R}$, we have

$$\int_{\mathbf{R}} A(t + t_0) dt = \int_{\mathbf{R}} A(t) dt \tag{1.13}$$

(indeed, if x denotes the left-hand side, we have

$$\langle v, x\rangle = \int_{\mathbf{R}} \langle v, A(t + t_0)\rangle dt = \int_{\mathbf{R}} \langle v, A(t)\rangle dt$$

for all $v \in H$, since the Lebesgue measure is invariant under translation; hence again the identity holds.)

– Suppose $A_n : \mathbf{R} \longrightarrow H$ is a sequence of weakly measurable H-valued functions, which have a common compact support and are uniformly bounded; if there exists a weakly-measurable function A such that

$$A_n(t) \longrightarrow A(t), \qquad \text{for all } t \in \mathbf{R},$$

then

$$\lim_{n\rightarrow+\infty} \int_{\mathbf{R}} A_n(t) dt = \int_{\mathbf{R}} A(t) dt \tag{1.14}$$

(indeed, replacing A_n with $A_n - A$, we can assume that $A = 0$; then let x_n be the integral of A_n, which we must show tends to 0 as $n \rightarrow +\infty$; we use the formula

$$\|x_n\| = \sup_{\|v\|\leqslant 1} |\langle v, x_n\rangle,$$

and note that if $v \in H$ has norm at most 1, we have

$$|\langle v, x_n\rangle| = \Big|\int_{\mathbf{R}} \langle v, A_n(t)\rangle dt\Big| \leqslant \int_{\mathbf{R}} \|A_n(t)\| dt,$$

hence by taking the supremum over v, we get

$$\|x_n\| \leqslant \int_{\mathbf{R}} \|A_n(t)\| dt,$$

and the right-hand side converges to 0 by the dominated convergence theorem in view of the assumptions, e.g., if $K \subset \mathbf{R}$ is a common compact support of the A_n, C a common upper bound for $\|A_n(t)\|$, we have

$$\|A_n(t)\| \leqslant C\chi_K(t),$$

for all t and all n, and this is an integrable function on $\mathbf{R}$.)

Acknowledgments. Thanks to the assistants of the Funktionalanalysis course and of this course, for helping in making it work "fluently". And, of course, thanks to the students for their interest and for sending me lists of corrections and suggestions for improvements, in particular F. Bachmair, G. Oberdieck and E. Widmer.

CHAPTER 2

Review of spectral theory and compact operators

In this chapter, we review for completeness the basic vocabulary and fundamental results of general spectral theory of Banach algebras over $\mathbf{C}$. We then recall the results concerning the spectrum of compact operators on a Hilbert space, and add a few important facts, such as the definition and standard properties of *trace-class operators*.

2.1. Banach algebras and spectral theory

A Banach algebra (over $\mathbf{C}$, and which we always assume to have a unit) is a Banach space A with a multiplication operation

$$(a, b) \mapsto ab$$

which obeys the usual algebraic rules (associativity, distributivity with addition), is $\mathbf{C}$-linear, and interacts with the norm on A by the two conditions

$$\|ab\| \leqslant \|a\| \, \|b\|, \qquad \|1\| = 1.$$

Of course, by induction, one gets

$$\|a_1 a_2 \cdots a_k\| \leqslant \|a_1\| \cdots \|a_k\|$$

(even for $k = 0$, if the usual meaning of an empty product in an algebra is taken, namely that the empty product is the unit 1, either in A or in $\mathbf{C}$).

The basic examples are $A = L(V)$, the Banach space of continuous linear maps $V \to V$, where V is itself a Banach space, with the usual operator norm

$$\|T\|_{L(V)} = \sup_{\|v\| \leqslant 1} \|T(v)\| = \sup_{v \neq 0} \frac{\|T(v)\|}{\|v\|}$$

and the composition as product. In particular, this includes the case $A = L(H)$, where H is a Hilbert space, in which case there is an additional structure on A, the adjoint operation

$$T \mapsto T^*,$$

which is characterized by

$$\langle Tv, w \rangle = \langle v, T^* w \rangle, \qquad \text{for all } v, w \in H.$$

This adjoint satisfies the following rules: it is additive, involutive (i.e., we have $((T^*)^* = T)$, conjugate-linear, i.e. $(\alpha T)^* = \bar{\alpha} T^*$, and

$$\|T^*\| = \|T\| = \sqrt{\|T^* T\|} = \sqrt{\|T T^*\|}. \tag{2.1}$$

Given an arbitrary Banach algebra A, we have the *spectrum* of elements of A, defined by

$$\sigma(a) = \{\lambda \in \mathbf{C} \mid \lambda \cdot 1 - a \notin A^{\times}\},$$

the complement of which is the *resolvent set*. Here $A^{\times}$ is the group (under multiplication) of invertible elements of A, those a such that there exists b with $ab = ba = 1$.

The crucial feature of Banach algebras, in which the submultiplicativity of the norm and the completeness of the underlying vector space both interact, is the following lemma:

Lemma 2.1. *Let A be a Banach algebra. Then $A^\times$ is open in A, and more precisely, for any $a_0 \in A^\times$, for instance $a_0 = 1$, there exists $\delta > 0$ such that for any $a \in A$ with $\|a\| < \delta$, we have $a_0 + a \in A^\times$ with*

$$(a_0 + a)^{-1} = a_0^{-1} \sum_{k \geqslant 0} (-1)^k (a_0^{-1} a)^k,$$

where the series converges absolutely. If $a_0 = 1$, one can take $\delta = 1$.

Proof. We have

$$\|(-1)^k (a_0^{-1} a)^k\| \leqslant \|a_0^{-1} a\|^k \leqslant \|a_0^{-1}\|^k \|a\|^k,$$

so the series converges absolutely in A if $\|a\| < \|a_0^{-1}\|^{-1}$. The fact that its a_0^{-1} times its sum is the inverse of $a_0 + a$ is then checked by the usual geometric series argument. $\square$

The basic property of $\sigma(a) \subset \mathbf{C}$ is that it is a compact, *non-empty* subset of $\mathbf{C}$ – the last part, which is the most delicate, depends on using the field of complex numbers, as can be seen already in the simplest case of finite-rank matrices.

Example 2.2. A simple example shows that the restrictions just described are the only general conditions on the spectrum: any compact subset $K \subset \mathbf{C}$ can occur as the spectrum of some element of some Banach algebra.

Indeed, consider first more generally a compact topological space X and let $A = C(X)$ with the supremum norm. This Banach space is a Banach algebra with the pointwise product of functions. Now given $f \in A$, we have

$$\sigma(f) = f(X) \subset \mathbf{C}, \tag{2.2}$$

since a function $g \in A$ is invertible if and only if $0 \notin g(X)$ (the inverse of a continuous function which has no zero is itself continuous).

If we now fix $K \subset \mathbf{C}$ a non-empty compact subset, we can take $X = K$ and $f : x \mapsto x$ in $C(X) = C(K)$, and deduce that $\sigma(f) = K$ in that case.

In the situation of interest to us, where $A = L(H)$ with H a Hilbert space, two refinement are possible: one can isolate, using the adjoint, special subclasses of operators with particular spectral properties (and this will be very important), and on the other hand, one can partition the spectrum in three disjoint subsets.

We start with this second idea: given $T \in L(H)$ and $\lambda \in \sigma(T)$, one says that:

– λ is in the *point spectrum* $\sigma_p(T)$ if $\mathrm{Ker}(\lambda - T)$ is not zero, i.e., if $\lambda - T$ is not injective; this is called an eigenvalue;

– λ is in the *residual spectrum* $\sigma_r(T)$ if $\lambda - T$ is injective, and the closure of the image of $\lambda - T$ is not equal to H;

– λ is in the *continuous spectrum* $\sigma_c(T)$ of $\lambda - T$ is injective, not surjective, but its image is dense in H.

It is is consequence of the Banach Isomorphism Theorem that these three possibilities exhaust the cases where $\lambda - T$ is not invertible in $L(H)$. All three types of spectrum can occur, though any one of them could be empty for a given operator T.

Remark 2.3. We consider in particular the *continuous* spectrum: if $\lambda \in \sigma_c(T)$, we can define an inverse mapping

$$(\lambda - T)^{-1} \begin{cases} \mathrm{Im}(\lambda - T) \to H \\ v \mapsto \text{ the only } w \in H \text{ with } \lambda w - T(w) = v, \end{cases}$$

which is an *unbounded* linear operator, defined on a dense subspace of H: these will be studied more systematically in Chapter 4, and it will turn out that it is often possible to study such unbounded operators by seeing them as "inverses" of injective operators T which have dense image (see, for instance, the proof of the Spectral Theorem 4.42).

Note also that $(\lambda - T)^{-1}$ is surjective (again by definition of the continuous spectrum), so from the point of view of solving linear equations, it corresponds to a case where we can solve uniquely the equation with arbitrary right-hand side, but *we can not do it in a continuous way*, or equivalently, we can not do it with an at-most-linear control of the norm of the solution in terms of the norm of the parameter.

Special classes of operators on a Hilbert space are defined as follows:
– T is *positive* if we have

$$\langle T(v), v\rangle \geqslant 0 \text{ for all } v \in V,$$

– T is *normal* if

$$T^*T = TT^*,$$

and this is equivalent with

$$\|T(v)\| = \|T^*(v)\| \text{ for all } v \in V, \tag{2.3}$$

which may be a more natural condition;
– T is *self-adjoint* if

$$T^* = T,$$

and this is equivalent with

$$\langle T(v), v\rangle \in \mathbf{R} \text{ for all } v \in V, \tag{2.4}$$

which shows that any positive operator is also self-adjoint;
– T is *unitary* if it is invertible and

$$T^{-1} = T^*,$$

so that in particular a unitary T is normal.

A basic analogy to keep in mind compares $L(H)$ with $\mathbf{C}$, as follows, based on the analogy of the adjoint with the complex conjugation:
– T self-adjoint corresponds to $z \in \mathbf{R}$;
– T unitary corresponds to $|z| = 1$;
– T positive corresponds to $z \in [0, +\infty[$.

One may also notice the following useful fact: for any $T \in L(H)$, the operators TT^* and T^*T are both self-adjoint and positive: indeed, we have

$$\langle TT^*(v), v\rangle = \|T(v)\|^2 \text{ for all } v.$$

We recall the proofs that (2.3) and (2.4) correspond to normal and self-adjoint operators, respectively; clearly in either case, the operator definition implies the stated relations for $\langle T(v), v\rangle$ or $\|T(v)\|$. Conversely, notice that both (2.3) and (2.4) can be expressed using the adjointness as

$$\langle A(v), v\rangle = \langle B(v), v\rangle \text{ for all } v,$$

for some operators A and B for which $A = B$ is the desired conclusion (e.g., $A = TT^*$, $B = T^*T$ for the normal property). So we need only recall the next lemma:

Lemma 2.4. *Let A, B in $L(H)$ be such that*

$$\langle A(v), v\rangle = \langle B(v), v\rangle \text{ for all } v.$$

Then we have $A = B$; the result also holds if the condition is assumed to be valid only for $\|v\| = 1$. The converse is of course true.

Proof. By linearity, one can assume $B = 0$. Then for all fixed v, $w \in H$ and any α, $\beta \in \mathbf{C}$, expanding leads to

$$0 = \langle A(\alpha v + \beta w), \alpha v + \beta w\rangle = \alpha\bar{\beta}\langle A(v), w\rangle + \bar{\alpha}\beta\langle A(w), v\rangle.$$

This, as an identity valid for all α and β, implies the coefficients $\langle A(v), w\rangle$ and $\langle A(w), v\rangle$ are both zero. But if this is true for all v and w, putting $w = A(v)$ gives $A = 0$. □

This lemma also call to mind the following useful property of the spectrum in $L(H)$: we have

$$\sigma(T) \subset \overline{N(T)},$$

the closure of the *numerical range* defined by

$$N(T) = \{\langle T(v), v\rangle \mid \|v\| = 1\}.$$

Another important feature of Hilbert spaces is the formula

$$r(T) = \|T\| \tag{2.5}$$

for the *spectral radius*

$$r(T) = \max\{|\lambda| \mid \lambda \in \sigma(T)\} = \inf_{n\geqslant 1} \|T^n\|^{1/n} = \lim_{n\to+\infty} \|T^n\|^{1/n},$$

of a *normal* operator (the equalities in this definition are themselves non-obvious results); indeed, (2.5) follows from the limit formula and the simpler fact that

$$\|T^2\| = \|T\|^2 \tag{2.6}$$

for a normal operator in $L(H)$.

One can see easily that

$$\lambda \in \sigma_p(T) \text{ if and only if } \bar{\lambda} \in \sigma_r(T^*) \tag{2.7}$$

(because if e_0 is a non-zero eigenvector of $\lambda - T$, we have

$$\langle(\bar{\lambda} - T^*)(v), e_0\rangle = \langle v, (\lambda - T)e_0\rangle = 0,$$

we see that the image of $\bar{\lambda} - T^*$ is included in the closed kernel of the non-zero linear functional $v \mapsto \langle v, e_0\rangle$), and in particular it follows that the residual spectrum of a *self-adjoint* operator T is empty.

The continuous spectrum can be difficult to grasp at first; the following remark helps somewhat: if $\lambda \in \mathbf{C}$ is not in $\sigma_p(T) \cup \sigma_r(T)$ then $\lambda \in \sigma_c(T)$ if and only if

(2.8) there exist vectors v with $\|v\| = 1$ and $\|(T - \lambda)v\|$ arbitrarily small.

Indeed, the existence of a sequence of such vectors with $\|(T-\lambda)v_n\| \to 0$ is equivalent with the fact that the map (already described)

$$\begin{cases} \mathrm{Im}(T - \lambda) \to H \\ w \mapsto v \text{ such that } (T - \lambda)v = w \end{cases}$$

is *not* continuous (not bounded on the unit sphere). Intuitively, these sequences are "almost" eigenvectors.

2.2. Compact operators on a Hilbert space

As a general rule, we can not expect, even for self-adjoint operators, to have a spectral theory built entirely on eigenvalues, as is the case for finite-dimensional Hilbert spaces. There is, however, one important special class of operators for which the eigenvalues essentially suffice to describe the spectrum: those are *compact* operators, of which the Hilbert-Schmidt operators of Example 1.3 are particular cases.

2.2.1. Reminders. We recall the two equivalent definitions of compact operators: $T \in L(H)$ is compact, denoted $T \in K(H)$, if either (1) there exists a sequence of operators $T_n \in L(H)$ with $\dim \mathrm{Im}(T_n) < +\infty$ for all n, and

$$\lim_{n\to+\infty} T_n = T, \text{ in the norm topology on } L(H);$$

or (2) for any bounded subset $B \subset H$ (equivalently, for the unit ball of H), the image $T(B) \subset H$ is *relatively compact*, meaning that its closure is compact.

Although the first definition might suggest that compact operators are plentiful, the second shows this is not so: for instance, in any infinite dimensional Hilbert space, since the closed unit ball of H is *not* compact,[1] it follows that the identity operator is not in $K(H)$. On the other hand, the first definition quickly shows that if $T \in K(H)$ is compact and $S \in L(H)$ is *arbitrary*, we have

$$ST, \quad TS \in K(H),$$

or in algebraic language, $K(H)$ is a *two-sided ideal* in the Banach algebra $L(H)$.

It follows also from the first definition that $K(H)$ is closed in $L(H)$. In fact, it is known that if H is separable, $K(H)$ is the only closed two-sided ideal in $L(H)$, with the exception of 0 and $L(H)$ of course.

Here is the basic spectral theorem for a compact operator in $L(H)$, due to Riesz, Fredholm, Hilbert and Schmidt.

Theorem 2.5 (Spectral theorem for compact operators). *Let H be an infinite dimensional Hilbert space, and let $T \in K(H)$ be a compact operator.*

(1) *Except for the possible value* 0*, the spectrum of T is entirely point spectrum; in other words*

$$\sigma(T) - \{0\} = \sigma_p(T) - \{0\}.$$

(2) *We have $0 \in \sigma(T)$, and $0 \in \sigma_p(T)$ if and only if T is not injective.*

(3) *The point spectrum outside of* 0 *is countable and has finite multiplicity: for each $\lambda \in \sigma_p(T) - \{0\}$, we have*

$$\dim \mathrm{Ker}(\lambda - T) < +\infty.$$

(4) *Assume T is normal. Let $H_0 = \mathrm{Ker}(T)$, and $H_1 = \mathrm{Ker}(T)^{\perp}$. Then T maps H_0 to H_0 and H_1 to H_1; on H_1, which is separable, there exists an orthonormal basis $(e_1, \ldots, e_n, \ldots)$ and $\lambda_n \in \sigma_p(T) - \{0\}$ such that*

$$\lim_{n\to+\infty} \lambda_n = 0,$$

and

$$T(e_n) = \lambda_n e_n \textit{ for all } n \geqslant 1.$$

[1] Any sequence (e_n) of vectors taken from an orthonormal basis satisfies $\|e_n - e_m\|^2 = 2$ for all $n \neq m$, so it has no Cauchy subsequence, and a fortiori no convergent subsequence!

In particular, if $(f_i)_{i\in I}$ is an arbitrary orthonormal basis of H_0, which may not be separable, we have

$$T\Big(\sum_{i\in I}\alpha_i f_i + \sum_{n\geqslant 1}\alpha_n e_n\Big) = \sum_{n\geqslant 1}\lambda_n\alpha_n e_n$$

for all scalars α_i, $\alpha_n \in \mathbf{C}$ for which the vector on the left-hand side lies in H, and the series on the right converges in H. This can be expressed also as

$$T(v) = \sum_{n\geqslant 1}\lambda_n\langle v, e_n\rangle e_n. \tag{2.9}$$

In Example 3.20, we will prove the self-adjoint case using the general spectral theorem for bounded self-adjoint operators.

Note that (2.9) will be the most commonly used version of this statement for T a normal compact operator, where the (e_n) form an orthonormal basis of $\mathrm{Ker}(T)^\perp$ and $T(e_n) = \lambda_n e_n$.

Corollary 2.6 (Fredholm alternative). *Let H be a Hilbert space, let $T \in K(H)$, and let $\lambda \in \mathbf{C} - \{0\}$. If there is no non-trivial solution $0 \neq v \in H$ to*

$$T(v) - \lambda v = 0,$$

then for every $w \in H$, there is a unique $v \in H$ such that

$$T(v) - \lambda v = w.$$

Moreover, this unique solution is bounded by $\|v\| \leqslant C\|w\|$ for some constant $C \geqslant 0$.

Proof. Indeed, the assumption is that $\lambda \notin \sigma_p(T) - \{0\}$, and this means $\lambda \notin \sigma(T)$ since the non-zero spectrum is purely made of eigenvalues. Thus $(T-\lambda)$ is invertible, □

Part (4) of the Theorem has a type of converse:

Proposition 2.7. *Any diagonal operator defined on a separable Hilbert space H by fixing an orthonormal basis $(e_n)_{n\geqslant 1}$ and putting*

$$T(e_n) = \lambda_n e_n$$

for some sequence (λ_n) of complex numbers with $\lambda_n \to 0$ is a compact operator on $L(H)$, and its spectrum is the set of values of (λ_n), with the addition of 0 if H is infinite-dimensional.

Note there may be multiplicities of course. For the proof, note that Part (4) means that any compact operator on a separable Hilbert space has such a "diagonal" model, which can be considered to be defined (via the orthonormal basis) on the space $\ell_2(\mathbf{N})$.

2.2.2. First applications. We now show a simple application which is the prototype of many similar statements further on, and shows the potential usefulness of the spectral theorem.

Proposition 2.8. *Let H be a Hilbert space and let $T \in K(H)$ be a positive compact operator. Then there exists a positive operator $S \in K(H)$ such that $S^2 = T$, which is denoted $\sqrt{T}$ or $T^{1/2}$. It is unique among positive bounded operators.*

Proof. Consider the orthogonal decomposition

$$H = H_0 \oplus H_1$$

as in Part (4) of Theorem 2.5 (which applies since positive operators are self-adjoint, hence normal), and an arbitrary choice of orthonormal basis (e_n) of H_1 made of eigenvectors of T with $T(e_n) = \lambda_n e_n$. We have

$$\lambda_n = \lambda_n \langle e_n, e_n \rangle = \langle \lambda_n e_n, e_n \rangle = \langle T(e_n), e_n \rangle \geqslant 0$$

for all n by the positivity assumption. Then we define simply

$$S(v_0 + v_1) = \sum_{n \geqslant 1} \sqrt{\lambda_n} \alpha_n e_n$$

for any $v_0 \in H_0$ and

$$v_1 = \sum_{n \geqslant 1} \alpha_n e_n \in H_1.$$

This is a diagonal operator with coefficients $(\sqrt{\lambda_n})$ which tend to zero, to we know that $S \in K(H)$ is well-defined and compact.

Then we compute simply that for all $v \in H$ as above, we have

$$S^2(v) = S(S(v_0 + v_1)) = S\Big(\sum_{n \geqslant 1} \sqrt{\lambda_n} \alpha_n e_n\Big) = \sum_{n \geqslant 1} \lambda_n \alpha_n e_n = T(v),$$

using the continuity of S.

We now show unicity. Let $S \in L(H)$ be such that $S^2 = T$, $S \geqslant 0$. Then $ST = S \cdot S^2 = S^2 \cdot S = TS$, so S and T commute, and it follows that for any non-zero eigenvalue λ_n of T, S induces a positive operator

$$S_n : \operatorname{Ker}(T - \lambda_n) \to \operatorname{Ker}(T - \lambda_n),$$

on the *finite-dimensional* λ_n-eigenspace of T. This finite-dimensional operator S_n satisfies $S_n^2 = \lambda_n \mathrm{Id}$, hence its only eigenvalues are among $\pm\sqrt{\lambda_n}$. Since it is positive, $\sqrt{\lambda_n}$ is in fact the only eigenvalue, and this implies that $S_n = \sqrt{\lambda_n}\mathrm{Id}$.

In addition, we have

$$\|S(v)\|^2 = \langle S(v), S(v) \rangle = \langle T(v), v \rangle,$$

and so $\operatorname{Ker}(S) = \{v \mid \langle T(v), v \rangle = 0\}$. By the expression

$$\langle T(v), v \rangle = \sum_{n \geqslant 1} \lambda_n |\langle v, e_n \rangle|^2$$

in terms of an orthonormal basis of eigenvectors (e_n) (see (2.9)), and the positivity $\lambda_n > 0$ of the eigenvalues, we have $\langle T(v), v \rangle = 0$ if and only if v is perpendicular to the span of (e_n), i.e., by the construction, if and only if $v \in \operatorname{Ker}(T)$. Thus the positive operator S with $S^2 = T$ is uniquely determined on each eigenspace of T, and on the kernel of T. By the spectral theorem, this implies that S is unique. □

Remark 2.9. (1) This positive solution to $S^2 = T$ is in fact unique among all bounded operators, but this is not yet obvious.

Also, we can clearly do the same thing to construct many more such operators: essentially, for any suitable function $f : \mathbf{C} \to \mathbf{C}$, we can define $f(T)$ by putting $f(\lambda_n)$ as eigenvalue for e_n. This type of general construction will indeed be of great importance later, but we keep a complete treatment for the spectral theory of bounded (and then unbounded) operators.

(2) This proposition is in fact true for all positive bounded operators $T \in L(H)$, as will be shown in the next chapter.

Since, for any compact operator $T \in K(H)$, we have $T^*T \geqslant 0$, this justifies the following definition:

DEFINITION 2.10. Let $T \in K(H)$ be an arbitrary compact operator. The compact positive operator $\sqrt{T^*T}$, as defined by the previous proposition, is denoted $|T|$.

Note that if $T \geqslant 0$, then of course $|T| = T$. We use this proposition to find a generalization of the Spectral Theorem to all compact operators.

REMARK 2.11. It is *not* true that $|T_1 + T_2| \leqslant |T_1| + |T_2|$.

PROPOSITION 2.12. *Let H be a Hilbert space and let $T \in K(H)$. There exist two orthonormal systems (e_n) and (f_n) in H, and positive real numbers s_n with $s_n \to 0$ as $n \to +\infty$, such that*

$$T(v) = \sum_{n\geqslant 1} s_n \langle v, e_n\rangle f_n,$$

for all $v \in H$. Moreover the s_n are the non-zero eigenvalues of the positive compact operator $|T|$.

Note that the (e_n) and (f_n) may span a proper subspace of H. Also neither are eigenvectors of T in general: for any $m \geqslant 1$, we have

$$T(e_m) = s_m f_m, \qquad T(f_m) = \sum_{n\geqslant 1} s_n \langle f_m, e_n\rangle f_n.$$

PROOF. The operator T^*T is a compact positive operator, and so we can find $S \in K(H)$ with $S^2 = T^*T$. Moreover, we have

$$\|S(v)\|^2 = \langle S(v), S(v)\rangle = \langle T^*T(v), v\rangle = \|T(v)\|^2 \tag{2.10}$$

for all $v \in H$. This means in particular that $T(v) = 0$ if and only if $S(v) = 0$ (i.e., $\mathrm{Ker}(S) = \mathrm{Ker}(T)$), and implies that the linear map

$$U \begin{cases} \mathrm{Im}(S) \to H \\ v \mapsto T(w) \quad \text{where } S(w) = v, \end{cases}$$

is well-defined, and then that it is isometric, and in particular continuous. In shorthand, the definition can be expressed as

$$U(S(w)) = T(w)$$

for $w \in H$.

We can then extend U by continuity to the closure $H_1 = \overline{\mathrm{Im}(S)}$, and (if needed) by 0 to

$$H_1^{\perp} = \mathrm{Im}(S)^{\perp} = \mathrm{Ker}(S),$$

and we have the relation $US = T$ by the remark above.

In addition, apply the Spectral Theorem to S to find a sequence (s_n) of eigenvalues of H and a countable orthonormal system (e_n) such that

$$S(v) = \sum_{n\geqslant 1} s_n \langle v, e_n\rangle e_n.$$

If we let $f_n = U(e_n)$, the isometry of U means that (f_n) is still an orthonormal system, and we have by definition

$$T(v) = U \circ S(v) = \sum_{n\geqslant 1} s_n \langle v, e_n\rangle f_n.$$

□

Remark 2.13. The decomposition of T as US, where U is an isometry and S is positive is an analogue of the polar decomposition $z = re^{i\theta}$ of a complex number $z \in \mathbf{C}$ with $r > 0$ and $\theta \in \mathbf{R}$. We will see it come into play in other contexts.

2.2.3. Variational computations of eigenvalues. We now present an important result which sometimes can be used to compute or estimate eigenvalues of a compact operator. It is, again, a generalization of a result valid for the finite-dimensional case, but since those are not as well-known as some of the others, it is of some interest even in that case.

Let H be a Hilbert space and $T \in K(H)$ a positive operator. The positive eigenvalues of T, counted with multiplicity, form either a finite sequence (e.g., if $\dim H < +\infty$), or a sequence tending to 0. We can in any case enumerate them in decreasing order

$$\lambda_1 \geqslant \lambda_2 \geqslant \cdots \geqslant \lambda_n \cdots > 0,$$

where the sequence may terminate if $\dim H < +\infty$. This is an unambiguous definition of functions

$$\lambda_k : T \mapsto \lambda_k(T)$$

for $k \geqslant 1$ and $T \in K(H)$ positive.

Proposition 2.14 (Courant-Raylegh minimax principle). *With notation as above, we have*

$$\lambda_k = \max_{\dim V = k} \min_{v \in V - \{0\}} \frac{\langle T(v), v \rangle}{\|v\|^2}, \tag{2.11}$$

and

$$\lambda_k = \min_{\dim V = k-1} \max_{v \in V^{\perp} - \{0\}} \frac{\langle T(v), v \rangle}{\|v\|^2} \tag{2.12}$$

for any $k \geqslant 1$ such that λ_k is defined, and in particular, those maxima and minima exist. In both cases, V runs over subspaces of H of the stated dimension, and in the first case it is assumed that $V \subset \mathrm{Ker}(T)^{\perp}$.

Proof. Fix an orthonormal basis (e_n) of $\mathrm{Ker}(T)^{\perp}$ such that

$$v = \sum_{n \geqslant 1} \langle v, e_n \rangle e_n, \qquad T(v) = \sum_{n \geqslant 1} \lambda_n \langle v, e_n \rangle e_n$$

for all $v \in \mathrm{Ker}(T)^{\perp}$; the second formula is in fact valid for all $v \in V$, and orthonormality of the basis gives the formulas

$$\langle T(v), v \rangle = \sum_{n \geqslant 1} \lambda_n |\langle v, e_n \rangle|^2, \qquad \|v\|^2 = \sum_{n \geqslant 1} |\langle v, e_n \rangle|^2,$$

for all $v \in V$. Since $\lambda_n > 0$ for all $n \geqslant 1$, each term of the sums is non-negative, and by positivity we can write

$$\langle T(v), v \rangle \geqslant \sum_{1 \leqslant n \leqslant k} \lambda_n |\langle v, e_n \rangle|^2,$$

and then

$$\langle T(v), v \rangle \geqslant \lambda_k \sum_{1 \leqslant n \leqslant k} |\langle v, e_n \rangle|^2,$$

since the eigenvalues are ordered in decreasing order. If we restrict v to the k-dimensional subspace V_k spanned by $e_1, \ldots, e_k$, we note first that $V_k \subset \mathrm{Ker}(T)^{\perp}$, and then that the above gives

$$\langle T(v), v \rangle \geqslant \lambda_k \|v\|^2, \qquad v \in V_k,$$

and therefore

$$\inf_{v \in V_k - \{0\}} \frac{\langle T(v), v\rangle}{\|v\|^2} \geqslant \lambda_k,$$

and in fact this is a minimum and an equality, since $v = e_k \in V_k$ achieves equality.

So the right-hand side of (2.11) – assuming the maximum is attained – is at least equal to λ_k. Now let V be any subspace of $\mathrm{Ker}(T)^\perp$ with $\dim V = k$. Consider the restriction

$$P_k : V \to V_k$$

of the orthogonal projection onto V_k. Its kernel is $V \cap V_k^\perp$; if it is non-zero, then for $v \in V_k^\perp \cap V$, we have

$$\langle T(v), v\rangle = \sum_{n \geqslant k+1} \lambda_n |\langle v, e_n\rangle|^2 \leqslant \lambda_{k+1}\|v\|^2 \leqslant \lambda_k \|v\|^2,$$

and if $V_k^\perp \cap V = 0$, P_k must be bijective since V and V_k have dimension k. Thus we can find $v \in V$ with

$$v = e_k + w, \qquad w \in V_k^\perp,$$

and then

$$\langle T(v), v\rangle = \lambda_k + \langle T(w), w\rangle \leqslant \lambda_k + \lambda_{k+1}\|w\|^2 \leqslant \lambda_k \|v\|^2.$$

It follows that the right-hand side of (2.11) is $\leqslant \lambda_k$, showing the equality holds. In the other direction, with V_{k-1} as before, we see that

$$\langle T(v), v\rangle \leqslant \lambda_k \|v\|^2$$

for $v \in V_{k-1}^\perp - \{0\}$, with equality if $v = e_k$, so that the right-hand side of (2.12) is $\leqslant \lambda_k$, with equality for $V = V_{k-1}$. If V is any subspace with $\dim V = k-1$, the restriction to V_k, which has dimension k, of the orthogonal projection on V must have non-zero kernel, so we find a non-zero vector $v \in V^\perp \cap V_k$. Thus

$$\langle T(v), v\rangle = \sum_{1 \leqslant j \leqslant k} \lambda_j |\langle v, e_j\rangle|^2 \geqslant \lambda_k \|v\|^2$$

showing that

$$\max_{v \in V^\perp - \{0\}} \frac{\langle T(v), v\rangle}{\|v\|^2} \geqslant \lambda_k,$$

and hence that the right-hand side of (2.12), with an infimum instead of minimum, is $\geqslant \lambda_k$. Hence the infimum is a minimum indeed, and is equal to λ_k. □

Example 2.15. (1) For $k = 1$, the minimax characterization (2.12) gives simply

$$\lambda_1 = \max_{v \neq 0} \frac{\langle T(v), v\rangle}{\|v\|^2} = r(T),$$

which was already known.

(2) For $k = 2$, we obtain

$$\lambda_2 = \min_{w \neq 0} \max_{\substack{v \neq 0 \\ v \perp w}} \frac{\langle T(v), v\rangle}{\|v\|^2},$$

and in fact the proof shows that the minimum over $w \in H - \{0\}$ can be restricted to a minimum over vectors $w \in \mathrm{Ker}(T - \lambda_1)$. In some important applications, the maximal eigenvalue λ_1 is known and it is also known that its eigenspace has dimension 1, and is spanned by some *explicit* vector e_1. In that case, the next-largest eigenvalue is given by

$$\lambda_2 = \max_{\substack{v \neq 0 \\ v \perp e_1}} \frac{\langle T(v), v\rangle}{\|v\|^2},$$

which can be a very convenient expression.

The following corollary shows the usefulness of this characterization of eigenvalues. We recall that $\lambda_k(T)$ designates the k-th eigenvalue of T, in decreasing order.

Corollary 2.16 (Monotonicity and continuity). *Let H be a Hilbert space.*

(1) *If T_1, $T_2 \in K(H)$ are positive compact operators such that*

$$\langle T_1(v), v\rangle \leqslant \langle T_2(v), v\rangle$$

for all $v \in V$, then we have

$$\lambda_k(T_1) \leqslant \lambda_k(T_2) \text{ for all } k \geqslant 1.$$

(2) *Let T_1, T_2 be positive compact operators such that*

$$\|T_1 - T_2\|_{L(H)} \leqslant \varepsilon,$$

for some $\varepsilon \geqslant 0$. Then for all $k \geqslant 1$, we have

$$|\lambda_k(T_1) - \lambda_k(T_2)| \leqslant \varepsilon.$$

In particular, from (2) it follows that if (T_n) is a sequence of positive compact operators such that $T_n \to T$ as $n \to \infty$, in the topology for $L(H)$, then T (which is also a positive compact operator) satisfies

$$\lim_{n \to +\infty} \lambda_k(T_n) = \lambda_k(T)$$

for all $k \geqslant 1$.

Proof. (1) For any subspace $V \subset H$ with $\dim V = k-1$, we have by assumption

$$\max_{v \in V^\perp - \{0\}} \frac{\langle T_1(v), v\rangle}{\|v\|^2} \leqslant \max_{v \in V^\perp - \{0\}} \frac{\langle T_2(v), v\rangle}{\|v\|^2},$$

and hence (2.12) shows immediately that $\lambda_k \leqslant \mu_k$.

(2) The assumption implies that

$$|\langle T_1(v), v\rangle - \langle T_2(v), v\rangle| = |\langle (T_1 - T_2)(v), v\rangle| \leqslant \varepsilon \|v\|^2$$

for all $v \in H$, by the Cauchy-Schwarz inequality, or in other words that for $v \in H$, we have

$$\langle T_2(v), v\rangle - \varepsilon\|v\|^2 \leqslant \langle T_1(v), v\rangle \leqslant \langle T_2(v), v\rangle + \varepsilon\|v\|^2.$$

Hence, for any fixed $k \geqslant 1$ and any $V \subset H$ with $\dim V = k-1$, we have

$$\max_{v \in V^\perp - \{0\}} \frac{\langle T_2(v), v\rangle}{\|v\|^2} - \varepsilon \leqslant \max_{v \in V^\perp - \{0\}} \frac{\langle T_1(v), v\rangle}{\|v\|^2} \leqslant \max_{v \in V^\perp - \{0\}} \frac{\langle T_2(v), v\rangle}{\|v\|^2} + \varepsilon,$$

and then from (2.12) we derive

$$\lambda_k(T_2) - \varepsilon \leqslant \lambda_k(T_1) \leqslant \lambda_k(T_2) + \varepsilon, \qquad \text{i.e. } |\lambda_k(T_1) - \lambda_k(T_2)| \leqslant \varepsilon,$$

as desired. ∎

2.2.4. Trace class operators. An important operation for finite-dimensional operators is the *trace*. This, indeed, has the important feature of being both easy to compute when the operator is expressed as a matrix in a certain basis, and of existing independently of the choice of any basis: in other words, the trace of two operators in $L(\mathbf{C}^n)$ is the same whenever two operators are "equivalent" in the sense of conjugacy under $GL(n, \mathbf{C})$ – so, for instance, it can be used to check that some operators are *not* equivalent.

It naturally seems desirable to extend the definition of the trace to more general operators, but clearly requires great care. One may hope that compact operators would make a good setting, since their spectrum is (apart from 0, which should not be a problem for a trace...) entirely composed of eigenvalues, and one may hope to define

$$\mathrm{Tr}(T) = \sum_{n \geqslant 1} \lambda_n, \tag{2.13}$$

where (λ_n) are the non-zero eigenvalues of $T \in K(H)$, with multiplicity, if the series makes sense. Unfortunately, we may recall that *any* sequence (λ_n) of non-zero complex numbers such that $\lambda_n \to 0$ can be the set of (non-zero) eigenvalues of a compact operator on some Hilbert space, and thus the series (2.13) can not be expected to converge better than a general series of complex numbers...

The series suggests, however, a special class where things should be better, in analogy with integration theory: if $T \geqslant 0$, the series makes sense, provided it is accepted that its value lies in $[0, +\infty]$ (e.g., if $\lambda_n = \frac{1}{n}$).

Definition 2.17. Let H be a Hilbert space and let $T \geqslant 0$ be a positive compact operator with non-zero eigenvalues $(\lambda_n(T))_{n \geqslant 1}$. The *trace of* T is the sum of the series

$$\sum_{n \geqslant 1} \lambda_n(T)$$

in $[0, +\infty]$.

Here are the basic properties of this trace (note that even the additivity is not at all clear from such a definition!)

Proposition 2.18. *Let H be a separable Hilbert space.*

(1) *If $T \geqslant 0$ is a positive compact operator, and $(e_n)_{n \geqslant 1}$ is any orthonormal basis of H, then we have*

$$\mathrm{Tr}(T) = \sum_{n \geqslant 1} \langle T(e_n), e_n \rangle. \tag{2.14}$$

(2) *If T_1, T_2 are positive and compact, and α, β are in $[0, +\infty[$, we have*

$$\mathrm{Tr}(\alpha T_1 + \beta T_2) = \alpha\,\mathrm{Tr}(T_1) + \beta\,\mathrm{Tr}(T_2).$$

If $T_1 \leqslant T_2$, in sense that $T_2 - T_1 \geqslant 0$, we have

$$\mathrm{Tr}(T_1) \leqslant \mathrm{Tr}(T_2).$$

(3) *If $T \geqslant 0$ and T is compact, and if $U \in L(H)$ is unitary, then UTU^{-1} is compact and positive, and*

$$\mathrm{Tr}(UTU^{-1}) = \mathrm{Tr}(T).$$

Note that the last property can not be stated (yet) as $\mathrm{Tr}(UT) = \mathrm{Tr}(TU)$ because the operators UT and TU are not necessarily positive (of course, we know they are compact).

Proof. (1) [Note that (2.14) is intuitively reasonable: if we think in matrix terms, the right-hand side is the sum of the diagonal coefficients of the infinite matrix representing T in the orthonormal basis (e_n)]

First, by positivity of T, the series indicated also makes sense, in $[0,+\infty]$, as a series with non-negative terms. Second, if (e_n) is chosen specially so that it is a basis of eigenvectors of T, with $T(e_n) = \lambda_n(T)e_n$, then we have obviously $\langle T(e_n), e_n\rangle = \lambda_n(T)$, and hence (2.14) holds for such a basis. To conclude, it is therefore enough to show that the right-hand side is independent of the orthonormal basis chosen for the computation (and in fact of the order of the vectors in the basis).

Let therefore $(f_m)_{m\geqslant 1}$ be another (or the same, for what we care) orthonormal basis; all the steps in the following computation are justified by the non-negativity of the terms involved: we have

$$\begin{aligned}\sum_{m\geqslant 1}\langle T(f_m), f_m\rangle &= \sum_{m\geqslant 1}\sum_{n\geqslant 1}\lambda_n(T)|\langle f_m, e_n\rangle|^2\\ &= \sum_{n\geqslant 1}\lambda_n(T)\sum_{m\geqslant 1}|\langle f_m, e_n\rangle|^2 = \sum_{n\geqslant 1}\lambda_n(T)\sum_{m\geqslant 1}|\langle e_n, f_m\rangle|^2\\ &= \sum_{n\geqslant 1}\lambda_n(T)\|e_n\|^2 = \mathrm{Tr}(T).\end{aligned}$$

where we use the fact that (f_m) is an orthonormal basis to compute $1 = \|e_n\|^2$ by Parseval's identity.

(2) Both parts are now clear; for instance, since obviously $\alpha T_1 + \beta T_2$ is positive and compact, we can compute the trace in a fixed orthonormal basis (e_n) of H, where we have

$$\sum_{n\geqslant 1}\langle(\alpha T_1+\beta T_2)(e_n), e_n\rangle = \alpha\sum_{n\geqslant 1}\langle T_1(e_n), e_n\rangle + \beta\sum_{n\geqslant 1}\langle T_2(e_n), e_n\rangle.$$

(3) This is also clear because

$$\langle UTU^{-1}(e_n), e_n\rangle = \langle T(f_n), f_n\rangle$$

with $f_n = U^{-1}(e_n) = U^*(e_n)$ by unitarity, and because if (e_n) is an orthonormal basis, then so is (f_n) (again because U is unitary), so that (1) gives the result. Alternatively, one can check that the eigenvalues of T and UTU^{-1} coincide *with multiplicity*. □

We proceed to define general *trace class compact operators* as suggested by integration theory. We recall the definition of $|T|$ from Definition 2.10.

Definition 2.19. Let H be a separable Hilbert space. A compact operator $T \in K(H)$ is said to be *of trace class* if

$$\mathrm{Tr}(|T|) = \mathrm{Tr}(\sqrt{T^*T}) < +\infty.$$

Note that if $s_n(T)$ are the non-zero eigenvalues of $|T|$, we have by definition

$$\mathrm{Tr}(|T|) = \sum_{n\geqslant 1} s_n(T). \tag{2.15}$$

Proposition 2.20. *Let H be a separable Hilbert space.*

(1) *If $T \in K(H)$ is of trace class, then for any orthonormal basis $(e_n)_{n\geqslant 1}$ of H, the series*

$$\sum_{n\geqslant 1}\langle T(e_n), e_n\rangle \tag{2.16}$$

converges absolutely in H, and its sum is independent of the chosen basis. It is called the trace $\mathrm{Tr}(T)$ *of T.*

(2) *The set of trace-class operators is a linear subspace $TC(H)$ of $K(H)$, and it is indeed a two-sided ideal in $L(H)$ containing the finite-rank operators and dense in $K(H)$; the trace map*

$$\mathrm{Tr} : TC(H) \to \mathbf{C}$$

is a linear functional on $TC(H)$, which coincides with the trace previously defined for positive compact operators of trace class. It is not *continuous.*

(3) *For any* normal *operator $T \in TC(H)$ with non-zero eigenvalues $(\lambda_n)_{n \geqslant 1}$, the series*

$$\sum_{n \geqslant 1} \lambda_n$$

converges absolutely, and we have

$$\mathrm{Tr}(T) = \sum_{n \geqslant 1} \lambda_n.$$

(4) *For any $S \in L(H)$ and any $T \in TC(H)$, we have ST, $TS \in TC(H)$ and*

$$\mathrm{Tr}(ST) = \mathrm{Tr}(TS).$$

Remark 2.21. One can show that, conversely, if an arbitrary operator $T \in L(H)$ is such that

$$\sum_{n \geqslant 1} \langle T(e_n), e_n \rangle$$

converges absolutely for all orthonormal bases (e_n), then T is compact. It is then of course of trace class.

Proof. (1) We use the structure for T given by Proposition 2.12, namely

$$T(v) = \sum_{n \geqslant 1} s_n \langle v, e_n \rangle f_n,$$

where (s_n) is the sequence of non-zero eigenvalues of the positive compact operator $|T|$ and (e_n), (f_m) are orthonormal systems; of course, the series converges in H. In particular, we have by assumption

$$\sum_{n \geqslant 1} s_n = \mathrm{Tr}(|T|) < +\infty.$$

Now, if $(\varphi_k)_{k \geqslant 1}$ is an arbitrary orthonormal basis, we have for all $k \geqslant 1$ the formula

$$\begin{aligned}\langle T(\varphi_k), \varphi_k \rangle &= \langle \sum_{n \geqslant 1} s_n \langle \varphi_k, e_n \rangle f_n, \varphi_k \rangle \\ &= \sum_{n \geqslant 1} s_n \langle \varphi_k, e_n \rangle \langle f_n, \varphi_k \rangle,\end{aligned}$$

where the convergence in H and continuity of the inner product justify the computation.

Now we see that, provided the interchange of the series is justified, we get after summing over k that

$$\begin{aligned}\sum_{k\geqslant 1}\langle T(\varphi_k),\varphi_k\rangle &= \sum_{k\geqslant 1}\sum_{n\geqslant 1} s_n\langle\varphi_k,e_n\rangle\langle f_n,\varphi_k\rangle,\\ &= \sum_{n\geqslant 1} s_n\sum_{k\geqslant 1}\langle\varphi_k,e_n\rangle\langle f_n,\varphi_k\rangle\\ &= \sum_{n\geqslant 1} s_n\langle f_n,e_n\rangle\end{aligned}$$

which is, indeed, independent of the basis! So we proceed to justify this computation, by checking that the double series

$$\sum_{k\geqslant 1}\sum_{n\geqslant 1} s_n\langle\varphi_k,e_n\rangle\langle f_n,\varphi_k\rangle$$

is absolutely convergent. Since, by the above, we also have

$$\sum_{k\geqslant 1}|\langle T(\varphi_k),\varphi_k\rangle| \leqslant \sum_{k\geqslant 1}\sum_{n\geqslant 1}|s_n\langle\varphi_k,e_n\rangle\langle f_n,\varphi_k\rangle|,$$

this will also check that the series of $\langle T(\varphi_k),\varphi_k\rangle$ converges absolutely for a trace class operator.

Now to work: since $s_n\geqslant 0$, we get by Cauchy-Schwarz inequality (in $\ell_2(\mathbf{N})$)

$$\sum_{n\geqslant 1}|s_n\langle\varphi_k,e_n\rangle\langle\varphi_k,f_n\rangle| \leqslant \Bigl(\sum_{n\geqslant 1} s_n|\langle\varphi_k,e_n\rangle|^2\Bigr)^{1/2}\Bigl(\sum_{n\geqslant 1} s_n|\langle\varphi_k,f_n\rangle|^2\Bigr)^{1/2}$$

for any $k\geqslant 1$. We now sum over k, and apply once more the Cauchy-Schwarz inequality in $\ell_2(\mathbf{N})$:

$$\sum_{k\geqslant 1}\sum_{n\geqslant 1}|s_n\langle\varphi_k,e_n\rangle\langle\varphi_k,f_n\rangle| \leqslant \Bigl(\sum_{k\geqslant 1}\sum_{n\geqslant 1} s_n|\langle\varphi_k,e_n\rangle|^2\Bigr)^{1/2}\Bigl(\sum_{k\geqslant 1}\sum_{n\geqslant 1} s_n|\langle\varphi_k,f_n\rangle|^2\Bigr)^{1/2},$$

but (since for series with positive terms there is no problem exchanging the two sums, and because (φ_k) is an orthonormal basis) we have

$$\sum_{k\geqslant 1}\sum_{n\geqslant 1} s_n|\langle\varphi_k,e_n\rangle|^2 = \sum_{n\geqslant 1} s_n\|e_n\|^2 = \sum_{n\geqslant 1} s_n < +\infty,$$

and exactly the same for the other term. This proves the absolute convergence we wanted, and in fact

$$\sum_{n\geqslant 1}\sum_{k\geqslant 1}|s_n\langle\varphi_k,e_n\rangle\langle\varphi_k,f_n\rangle| \leqslant \sum_{n\geqslant 1} s_n(T).$$

(2) Since it is easy to see that $|\lambda T|=|\lambda||T|$ if $\lambda\in\mathbf{C}$ and $T\in K(H)$ (where the two $|\cdot|$ are quite different things...), it follows that $\lambda T\in TC(H)$ if $T\in TC(H)$, and that $\mathrm{Tr}(\lambda T)=\lambda\,\mathrm{Tr}(T)$. So we must show that if T_1, T_2 are trace-class operators, then so is $T_3=T_1+T_2$.

We now denote by $s_n(T)$ the eigenvalues, in decreasing order, of $|T|$ for $T\in K(H)$. We will bound $s_n(T_3)$ in terms of suitable values $s_{n_1}(T_1)$, $s_{n_2}(T_2)$, using the variational characterization (2.12).

Let $V\subset H$ be a subspace of dimension $n-1$, and let

$$\rho_i(V)=\max_{\substack{v\perp V\\ v\neq 0}}\frac{\langle (T_i^*T_i)(v),v\rangle}{\|v\|^2}=\max_{\substack{v\perp V\\ v\neq 0}}\frac{\|T_i(v)\|^2}{\|v\|^2},\qquad \text{for } 1\leqslant i\leqslant 3.$$

By (2.12), $s_n(T_3)$ is the minimum of $\rho_3(V)^{1/2}$ over all subspaces of dimension $n-1$ in H. Now for $n \geqslant 1$, write

$$n - 1 = n_1 - 1 + n_2 - 1, \text{ where } n_1 \geqslant 1, \quad n_2 \geqslant 1, \tag{2.17}$$

and for any V with $\dim V = n-1$, let V_1 and V_2 be subspaces of V of dimension $n_1 - 1$ and $n_2 - 1$ respectively. Then

$$V^\perp \subset V_1^\perp,\ V_2^\perp,$$

and since we have

$$\|T_3(v)\| \leqslant \|T_1(v)\| + \|T_2(v)\|, \qquad v \in V^\perp,$$

we get

$$\rho_3(V)^{1/2} \leqslant \rho_1(V_1)^{1/2} + \rho_2(V_2)^{1/2}.$$

By (2.12), we derive

$$s_n(T_3) \leqslant \rho_1(V_1)^{1/2} + \rho_2(V_2)^{1/2},$$

an inequality which is now valid for *any* choices of subspaces V_1, V_2 with dimension $n_1 - 1$, $n_2 - 1$ respectively. Consequently

$$s_n(T_3) \leqslant s_{n_1}(T_1) + s_{n_2}(T_2)$$

whenever (2.17) holds (this inequality is due to K. Fan).

Next, for given n, we select n_1 and n_2 as close as possible to $n/2$; summing then over $n \geqslant 1$, we see that at most two possible values of n give a given value of n_1, n_2, and thus we derive

$$\sum_{n\geqslant 1} s_n(T_3) \leqslant 2\sum_{n\geqslant 1} s_n(T_1) + 2\sum_{n\geqslant 1} s_n(T_2) < +\infty.$$

The rest of part (2) is easy: the linearity of the trace follows from the definition (1) once it is known it is well-defined for a linear combination of trace-class operators. Moreover, the finite-rank operators are sums of rank 1 operators, and for these T^*T has a single non-zero eigenvalue, so they, and their sums by what precedes, are of trace class. Since finite-rank operators are dense in $K(H)$, so are the trace-class ones. If $T \geqslant 0$ is in $TC(H)$, we have $|T| = T$ and (1) shows the trace defined for $TC(H)$ coincides with the trace for positive operators. And finally, the trace is not continuous, since there exist positive compact operators with norm 1 and arbitrarily large trace (use diagonal operators with eigenvalues decreasing from 1 but defining convergent series with larger and larger sums, e.g.,

$$\sum_{n\geqslant 1} \frac{1}{n^\sigma} \to +\infty, \qquad \text{as } \sigma > 1 \to 1.$$

(3) If $T \in TC(H)$ is normal, we can use an orthonormal basis (e_n) for which $T(e_n) = \lambda_n e_n$ to compute the trace (2.16), and we obtain

$$\sum_{n\geqslant 1} \langle T(e_n), e_n\rangle = \sum_{n\geqslant 1} \lambda_n,$$

which is thus a convergent series with value $\mathrm{Tr}(T)$.

(4) If S is unitary, it is clear that ST, TS are in $TC(H)$ for $T \in TC(H)$ (indeed, $(ST)^*(ST) = T^*T$, $(TS)^*(TS) = S^{-1}T^*TS$, which have the same eigenvalues as T^*T). Further, if (e_n) is an orthonormal basis, we get

$$\sum_{n\geqslant 1} \langle (ST)e_n, e_n\rangle = \sum_{n\geqslant 1} \langle S(TS)f_n, e_n\rangle = \sum_{n\geqslant 1} \langle (TS)f_n, f_n\rangle,$$

where $f_n = S^{-1}(e_n)$ is another orthonormal basis (because S is unitary!), the independence of the trace from the choice of basis when computed as in (2.16) leads to $\mathrm{Tr}(ST) = \mathrm{Tr}(TS)$ for unitary S.

Now the point is that we can use the linearity to deduce from this that, given $T \in TC(H)$, the set of $S \in L(H)$ for which ST and TS are in $TC(H)$ and

$$\mathrm{Tr}(TS) = \mathrm{Tr}(ST)$$

is a *linear subspace* of $L(H)$ containing the unitaries. But the lemma below shows that this subspace is *all* of $L(H)$, and the desired result follows. □

Here is the lemma we used.

LEMMA 2.22. *Let H be a Hilbert space. For any $T \in L(H)$, there exist U_i, $1 \leqslant i \leqslant 4$, unitary, and $\alpha_i \in \mathbf{C}$, $1 \leqslant i \leqslant 4$, such that*

$$T = \alpha_1 U_1 + \cdots + \alpha_4 U_4.$$

REMARK 2.23. If this sounds surprising, recall that in $\mathbf{C}$, any complex number is the sum of at most two real multiples of elements of modulus 1 (namely, $z = \mathrm{Re}(z) \cdot 1 + \mathrm{Im}(z) \cdot i$).

PROOF. Imitating the real-imaginary part decomposition, we can first write

$$T = \frac{T + T^*}{2} + i\frac{T - T^*}{2i},$$

and each of the two terms is self-adjoint, so it is enough to show that any self-adjoint $T \in L(H)$ is a combination (with complex coefficients) of two unitary operators. We may even assume that $\|T\| \leqslant 1$, by scaling properly. We then use the fact stated in Remark 2.9, (2) [which will be proved in the next chapter, independently of this section] that a square root exists for any positive operator. Here, we have

$$\langle (\mathrm{Id} - T^2)(v), v\rangle = \|v\|^2 - \|T(v)\|^2 \geqslant 0$$

for all v (recall $T = T^*$), hence the operator $S = \sqrt{\mathrm{Id} - T^2}$ exists. We have

$$T = \frac{T + iS}{2} + \frac{T - iS}{2},$$

and the proof is then completed by checking that $(T \pm iS)$ are unitary: indeed, we have

$$(T + iS)(T - iS) = T^2 + S^2 = \mathrm{Id},$$

so both are invertible with $(T + iS)^{-1} = T - iS$, and since S and T are self-adjoint, we have also $(T + iS)^* = T - iS$. □

The proof brings the following useful additional information:

COROLLARY 2.24. *Let H be a Hilbert space and let $T \in TC(H)$ be a trace-class operator. Then we have*

$$|\,\mathrm{Tr}(T)| \leqslant \mathrm{Tr}(|T|).$$

Moreover, if T is given by the decomposition

$$T(v) = \sum_{n \geqslant 1} s_n \langle v, e_n\rangle f_n$$

a described by Proposition 2.12, we have

$$\mathrm{Tr}(T) = \sum_{n \geqslant 1} s_n \langle f_n, e_n\rangle.$$

Remark 2.25. One can show that, in fact, we have

$$\mathrm{Tr}(|T_1 + T_2|) \leqslant \mathrm{Tr}(|T_1|) + \mathrm{Tr}(|T_2|),$$

from which one deduces that $TC(H)$ is a Banach space with the norm

$$\|T\|_{TC} = \mathrm{Tr}(|T|).$$

In fact (see, e.g., [**RS1**, Th. VI.26]), one can show that the trace leads to isomorphisms

$$(TC(H), \|\cdot\|_{TC}) \simeq K(H)',$$

where $K(H)'$ is the dual of the Banach space $K(H)$ (seen as closed subspace of $L(H)$), by sending $T \in TC(H)$ to the linear functional

$$S \mapsto \mathrm{Tr}(TS)$$

on $K(H)$. Similarly, we have

$$L(H) \simeq (TC(H), \|\cdot\|_{TC})'$$

by the same map sending $T \in L(H)$ to the linear functional $S \mapsto \mathrm{Tr}(TS)$ for $S \in TC(H)$.

Example 2.26. Let (X, μ) be a finite measure space and $T = T_k$ a Hilbert-Schmidt operators associated to a kernel function $k : X \times X \to \mathbf{C}$. Then we have the following important property of the non-zero eigenvalues (λ_n) of T:

$$\sum_{n\geqslant 1} |\lambda_n|^2 = \int_{X\times X} |k(x,y)|^2 d\mu(x) d\mu(y) < +\infty.$$

This follows from the expansion of k given by

$$k(x,y) = \sum_{n\geqslant 1} \lambda_n e_n(x)\overline{e_n(y)}, \tag{2.18}$$

where $T(e_n) = \lambda_n e_n$, which is valid in $L^2(X \times X)$.

This means that the positive compact operator $S = T^*T$, which has eigenvalues $|\lambda_n|^2$, has finite trace given by

$$\mathrm{Tr}(S) = \int_{X\times X} |k(x,y)|^2 d\mu(x) d\mu(y).$$

Now assume that in addition that X is a compact metric space, μ a finite Borel measure on X, and that the kernel function k is real-valued, symmetric (so that T_k is self-adjoint), and continuous on $X \times X$ (for the product topology). Then one can show (Mercer's Theorem, see [**W**, Satz VI.4.2] for the special case $X = [0, 1]$, μ the Lebesgue measure) the following:

– The eigenfunctions e_n of T_k are continuous functions on X;
– The expansion (2.18) is valid pointwise and uniformly for $(x, y) \in X \times X$;
– The operator T_k is of trace class and we have

$$\mathrm{Tr}(T_k) = \sum_{n\geqslant 1} \lambda_n = \int_X k(x,x) d\mu(x).$$

This integral over the diagonal can be interpreted as an analogue of the expression of the trace as sum of diagonal elements of a matrix.

CHAPTER 3

The spectral theorem for bounded operators

In this chapter, we start generalizing the results we proved for compact operators to the class of normal bounded operators $T \in L(H)$. We follow mostly [**RS1**, Ch. VII].

If $T \in L(H)$ is a normal, bounded but not compact, operator, it may well not have sufficiently many eigenvalues to use only the point spectrum to describe it up to (unitary) equivalence. For instance, the (self-adjoint) multiplication operators M_x on $L^2([0,1])$ (see Example 1.1 in Chapter 1) has no eigenvalues in $L^2([0,1])$: $f \in \mathrm{Ker}(\lambda - M_x)$ implies

$$(x-\lambda)f(x) = 0 \text{ almost everywhere,}$$

hence $f(x) = 0$ for almost all $x \neq \lambda$, i.e., $f = 0$ in L^2.

So we must use the full spectrum to hope to classify bounded operators. This does indeed give a good description, one version of which is the following, that we will prove in this chapter:

Theorem 3.1. *Let H be a separable Hilbert space and $T \in L(H)$ a continuous* normal *operator. There exists a finite measure space (X, μ), a unitary operator*

$$U : H \to L^2(X, \mu)$$

and a bounded function $g \in L^\infty(X, \mu)$, such that

$$M_g \circ U = U \circ T,$$

or in other words, for all $f \in L^2(X, \mu)$, we have

$$(UTU^{-1})f(x) = g(x)f(x),$$

for (almost) all $x \in X$.

Example 3.2. Let T be a compact normal operator T, and assume for simplicity that all its eigenvalues are non-zero and distinct. Then the representation

$$T(v) = \sum_{n \geqslant 1} \lambda_n \langle v, e_n \rangle e_n$$

in terms of an orthonormal basis of eigenvectors (e_n) corresponds to Theorem 3.1 in the following way: the measure space is

$$(X, \mu) = (\{\text{eigenvalues of } T\}, \text{counting measure}),$$

with U given by

$$U(v) = (\langle v, e_\lambda \rangle)_\lambda, \qquad \text{where } T(e_\lambda) = \lambda e_\lambda \text{ for } \lambda \in X,$$

and the function g is defined by

$$g(\lambda) = \lambda.$$

To prove Theorem 3.1, we must see how to associate a measure space to T. The example of compact operators suggests that X should be (related to) the spectrum $\sigma(T)$

of T. As for the measure, it will ultimately be obtained from the theorem of Markov-Riesz identifying measures as elements of the dual of the Banach space $(C(X), \|\cdot\|_\infty)$ of continuous functions on a compact space $X = \sigma(T)$, using linear functionals defined by

$$\ell(f) = \langle f(T)v, v\rangle,$$

where f is a continuous function on $\sigma(T)$. (The idea is that the measure detects how much of a vector is "concentrated" around a point in the spectrum). To make sense of this, we must first define what is meant by $f(T)$. This will be the content of the first section; note that in the previous chapter, we saw already how useful such constructions can be (see Proposition 2.8).

3.1. Continuous functional calculus for self-adjoint operators

The goal of this section is to show how to define $f(T)$ for a self-adjoint operator $T \in L(H)$ and $f \in C(\sigma(T))$. Theorem 3.1 will be deduced in the next section for these operators, and then generalized to normal ones using some tricks.

The definition of what $f(T)$ should be is clear for certain functions f: namely, if

$$p(z) = \sum_{j=0}^{d} \alpha_j z^j$$

is a polynomial in $\mathbf{C}[X]$, restricted to $\sigma(T)$, then the only reasonable definition is

$$p(T) = \sum_{j=0}^{d} \alpha_j T^j \in L(H).$$

Remark 3.3. In fact, this definition makes sense for any $T \in L(H)$, not only for T normal, but there is a small technical point which explains why only normal operators are really suitable here: if $\sigma(T)$ is finite, polynomials are not uniquely determined by their restriction to $\sigma(T)$, and the definition above gives a map $\mathbf{C}[T] \rightarrow L(H)$, not one defined on $C(\sigma(T))$. We can not hope to have a functional calculus only depending on the spectrum if this apparent dependency is real. And indeed, sometimes it is: consider the simplest example of the operator given by the 2×2 matrix

$$A = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} \in L(\mathbf{C}^2).$$

We then have $\sigma(A) = \{0\}$, but although the polynomials $p_1 = X$ and $p_2 = X^2$ coincide on $\{0\}$, we have $p_1(A) = A \neq p_2(A) = 0$.

However, if we assume that T is normal, the problem disappears as one observes the following fact: if $p \in \mathbf{C}[T]$ is a polynomial such that p is zero on $\sigma(T)$, and T is normal, then $p(T)$, as defined above, is zero. Indeed, this follows from the relation

$$\|p(T)\|_{L(H)} = \|p\|_{C(\sigma(T))} = \max_{\lambda \in \sigma(T)} |p(\lambda)|,$$

which is proved in Lemma 3.6 below – it will be clear that there is no circularity in applying this here.

This suggests a general definition, using the approximation principles discussed in Section 1.4.2: from the Weierstrass Approximation Theorem (Corollary 1.11), applied to $X = \sigma(T) \subset \mathbf{R}$ for T self-adjoint, we know that functions $f \in C(\sigma(T))$ can be approximated uniformly by polynomial functions. This suggests to define

$$f(T) = \lim_{n \rightarrow +\infty} p_n(T), \tag{3.1}$$

where (p_n) is a sequence of polynomials such that $\|f - p_n\|_{C(X)} \to 0$.

This definition is indeed sensible and possible, and the basic properties of this construction are given in the following theorem. Roughly speaking, any operation on (or property of) the functions f which is reasonable corresponds to analogue operation or property of $f(T)$.

THEOREM 3.4 (Continuous functional calculus). *Let H be a Hilbert space and $T \in L(H)$ a* self-adjoint *bounded operator. There exists a* unique *linear map*

$$\phi = \phi_T : C(\sigma(T)) \to L(H),$$

also denoted $f \mapsto f(T)$, with the following properties:

– (0) *This extends naturally the definition above for polynomials, i.e., for any $p \in \mathbf{C}[X]$ as before, we have*

$$\phi(p) = p(T) = \sum_{j=0}^{d} \alpha(j) T^j.$$

– (1) *This map is a Banach-algebra isometric homomorphism, i.e., we have*

$$\phi(f_1 f_2) = \phi(f_1)\phi(f_2) \text{ for all } f_i \in C(\sigma(T)), \qquad \phi(\mathrm{Id}) = \mathrm{Id},$$

and

$$\|\phi(f)\| = \|f\|_{C(\sigma(T))}. \tag{3.2}$$

In addition, this homomorphism has the following properties:

(2) *For any $f \in C(\sigma(T))$, we have $\phi(f)^* = \phi(\bar{f})$, i.e., $f(T)^* = \bar{f}(T)$, and in particular $f(T)$ is normal for all $f \in C(\sigma(T))$. In addition*

$$f \geqslant 0 \Rightarrow \phi(f) \geqslant 0. \tag{3.3}$$

(3) *If $\lambda \in \sigma(T)$ is in the point spectrum and $v \in \mathrm{Ker}(\lambda - T)$, then $v \in \mathrm{Ker}(f(\lambda) - f(T))$.*

(4) *More generally, we have the* spectral mapping theorem*:*

$$\sigma(f(T)) = f(\sigma(T)) = \sigma(f), \quad \textit{where } \sigma(f) \textit{ is computed for } f \in C(\sigma(T)). \tag{3.4}$$

REMARK 3.5. Given (1), the property (0) is implied solely by $\mathbf{C}$-linearity of ϕ and by the fact that $\phi(z \mapsto z) = T$.

PROOF. As already observed, the essence of the proof of existence of ϕ is to show that (3.1) is a valid definition. In fact, if we can prove (3.2) for $f = p \in \mathbf{C}[X]$, we can deduce that the map

$$\Phi : (\mathbf{C}[X], \|\cdot\|_{C(\sigma(T))}) \to L(H)$$

which maps p to $p(T)$ is linear, continuous, and in fact isometric. Hence it extends uniquely by continuity to $C(\sigma(T))$, and the extension remains isometric. By continuity, the properties

$$\phi(f_1 f_2) = \phi(f_1)\phi(f_2), \qquad \phi(f)^* = \phi(\bar{f}),$$

which are valid for polynomials (using $T = T^*$ for the latter), pass to the limit and are true for all f. It follows that

$$f(T)^* f(T) = \phi(\bar{f})\phi(f) = \phi(\bar{f} f) = \phi(f \bar{f}) = f(T) f(T)^*,$$

so $f(T)$ is always normal (and self-adjoint if f is real-valued). Moreover, if $f \geqslant 0$, we can write

$$f = (\sqrt{f})^2 = g^2$$

where $g \geqslant 0$ is also continuous on $\sigma(T)$. Then $g(T)$ is well-defined, self-adjoint (because g is real-valued), and for all $v \in V$, we have

$$\langle f(T)v, v\rangle = \langle g(T)^2 v, v\rangle = \|g(T)v\|^2 \geqslant 0,$$

which shows that $f(T) \geqslant 0$.

Hence we see that the following lemma is clearly of the essence:

Lemma 3.6. *Let H be a Hilbert space.*

(1) *For $T \in L(H)$ arbitrary and any polynomial $p \in \mathbf{C}[X]$, define $\Phi(p) = p(T) \in L(H)$ as before. Then we have*

$$\sigma(\phi(p))) = p(\sigma(T)). \tag{3.5}$$

(2) *Let $T \in L(H)$ be normal and let $p \in \mathbf{C}[X]$ be polynomial. Then we have*

$$\|\phi(p)\|_{L(H)} = \|p\|_{C(\sigma(T))}. \tag{3.6}$$

Proof of the lemma. For (1), consider an arbitrary $\lambda \in \mathbf{C}$; we can factor the polynomial $p(X) - \lambda$ in $\mathbf{C}[X]$:

$$p(X) - \lambda = \alpha \prod_{1 \leqslant i \leqslant d} (X - \lambda_i),$$

for some $\alpha \in \mathbf{C}^\times$ and complex numbers $\lambda_i \in \mathbf{C}$ (not necessarily distinct). Since the map $p \mapsto p(T)$ is an algebra homomorphism, it follows that

$$p(T) - \lambda \cdot \mathrm{Id} = \alpha \prod_{1 \leqslant i \leqslant d} (T - \lambda_i).$$

If λ is not in $p(\sigma(T))$, the solutions λ_i to $p(z) = \lambda$ are *not* in $\sigma(T)$; hence the $T - \lambda_i$ are then all invertible, hence so is $p(T) - \lambda$. In other words, we have (by contraposition)

$$\sigma(p(T)) \subset p(\sigma(T)).$$

Conversely, if $\lambda \in p(\sigma(T))$, one of the λ_i is in $\sigma(T)$. Because the factors commute, we can assume either $i = 1$, if $T - \lambda_i$ is not surjective, in which case neither is $p(T) - \lambda_i$; or $i = d$, in case $T - \lambda_i$ is not injective, in which case neither is $p(T) - \lambda_i$. In all situations, $\lambda \in \sigma(p(T))$, proving the converse inclusion.

For (2), we note first that $\phi(p) = p(T)$ is normal if T is. By the Spectral Radius formula (2.5), we have

$$\|p(T)\| = r(p(T)) = \max_{\lambda \in \sigma(p(T))} |\lambda|,$$

and by (3.5), we get

$$\|p(T)\| = \max_{\lambda \in p(\sigma(T))} |\lambda| = \max_{\lambda \in \sigma(T)} |p(\lambda)|,$$

as desired. ☐

It remains to check that additional properties (3) and (4) of the continuous functional calculus $f \mapsto f(T)$ hold. For (3), let $v \in \mathrm{Ker}(\lambda - T)$. Write f as a uniform limit of polynomials $p_n \in \mathbf{C}[X]$; since $T(v) = \lambda v$, we have also by induction and linearity

$$p_n(T)v = p_n(\lambda)v,$$

and by continuity we get $f(T)(v) = f(\lambda)v$.

For (4), we first recall (Example 2.2) that for a compact topological space X, $C(X)$, with the supremum norm, is itself a Banach algebra, for which the spectrum is given by $\sigma(f) = f(X)$ for $f \in C(X)$. So for $f \in C(\sigma(T))$, we have indeed $\sigma(f) = f(\sigma(T))$.

We prove first by contraposition that $\lambda \in \sigma(f(T))$ implies that $\lambda \in f(\sigma(T))$. Indeed, if the latter does not hold, the function

$$g(x) = \frac{1}{f(x) - \lambda}$$

is a continuous function in $C(\sigma(T))$, and therefore the bounded operator $S = \phi(g)$ is defined by what precedes. The relations

$$g \cdot (f - \lambda) = (f - \lambda) \cdot g = 1,$$

valid in $C(\sigma(T))$, imply by (1) that

$$g(T)(f(T) - \lambda) = (f(T) - \lambda)g(T) = \mathrm{Id},$$

i.e., that $S = (f(T) - \lambda)^{-1}$, so that λ is not in the spectrum of $f(T)$, as expected.

Conversely, let $\lambda = f(\lambda_1)$ with $\lambda_1 \in \sigma(T)$. We need to check that $\lambda \in \sigma(f(T))$. We argue according to the type of λ_1; if $\lambda_1 \in \sigma_p(T)$, then (3) shows that $f(\lambda_1) \in \sigma_p(f(T))$, as desired.

Since $\sigma_r(T) = \emptyset$ for T self-adjoint, we are left with the case $\lambda_1 \in \sigma_c(T)$. We use the observation (2.8), applied first to $T - \lambda_1$, and transfer it to $f(T) - \lambda$. Let v be an arbitrary vector with $\|v\| = 1$ and $p \in \mathbf{C}[X]$ an arbitrary polynomial. We write

$$\begin{aligned}
\|(f(T) - \lambda)v\| &\leqslant \|(f(T) - p(T))v\| + \|(p(T) - \lambda)v\| \\
&\leqslant \|(f(T) - p(T))v\| + \|(p(T) - p(\lambda_1))v\| + |(p(\lambda_1) - \lambda|\|v\| \\
&\leqslant \|f(T) - p(T)\|_{L(H)} + |p(\lambda_1) - \lambda| + \|(p(T) - p(\lambda_1))v\| \\
&\leqslant \|f - p\|_{C(\sigma(T))} + |p(\lambda_1) - \lambda| + \|(p(T) - p(\lambda_1))v\|.
\end{aligned}$$

Now write

$$p(X) - p(\lambda_1) = q(X)(X - \lambda_1),$$

so that

$$\|(f(T) - \lambda)v\| \leqslant \|f - p\|_{C(\sigma(T))} + |p(\lambda_1) - \lambda| + \|q\| \; \|(T - \lambda_1)v\|,$$

for all v in the unit sphere of H. Now fix $\varepsilon > 0$ arbitrary; we can find a polynomial p such that $\|f - p\|_{C(\sigma(T))} < \varepsilon/3$ and such that

$$|p(\lambda_1) - \lambda| < \frac{\varepsilon}{3}.$$

Then for all v with $\|v\| = 1$, we have

$$\|(f(T) - \lambda)v\| \leqslant \frac{2\varepsilon}{3} + \|q\| \; \|(T - \lambda_1)v\|,$$

where q is now fixed by the choice of p. Since $\lambda_1 \in \sigma_c(T)$, we can find (see (2.8)) a vector v with $\|v\| = 1$ and $\|(T - \lambda_1)v\| < \varepsilon/3\|q\|$, and then deduce that

$$\|(f(T) - \lambda)v\| < \varepsilon.$$

As $\varepsilon > 0$ was arbitrarily chosen, this implies by (2.8) that $\lambda \in \sigma_c(f(T))$. At this point, we have proved completely that $f(\sigma(T)) \subset \sigma(f(T))$, and this concludes the proof of (4), which was the last item remaining to conclude the proof of Theorem 3.4. ☐

The next corollary generalizes Proposition 2.8, where the existence part was restricted to compact operators.

Corollary 3.7. *Let $T \in L(H)$ be a positive operator. For any $n \geqslant 1$, there exists a positive normal operator $T^{1/n}$ such that*

$$(T^{1/n})^n = T.$$

Note that such an operator is unique, but we will prove this only a bit later. (At the moment, we do not know enough about how the spectrum helps describe *how* the operator acts).

Proof. Since $T \geqslant 0$, we have $\sigma(T) \subset [0, +\infty[$, hence the function $f : x \mapsto x^{1/n}$ is defined and continuous on $\sigma(T)$. Using the fact that

$$f(x)^n = x,$$

the functional calculus implies that $f(T)^n = T$. Moreover $f \geqslant 0$, and hence $f(T) \geqslant 0$ by Theorem 3.4. $\square$

Here is another simple corollary which will be generalized later, and which shows how the functional calculus can be used to "probe" the spectrum.

Corollary 3.8. *Let H be a Hilbert space and $T \in L(H)$ a bounded self-adjoint operator. Let $\lambda \in \sigma(T)$ be an* isolated point*, i.e., for some $\varepsilon > 0$, $\sigma(T) \cap]\lambda-\varepsilon, \lambda+\varepsilon[= \{\lambda\}$. Then λ is in the point spectrum.*

Proof. The fact that λ is isolated implies that the function $f : \sigma(T) \to \mathbf{C}$ which maps λ to 1 and the complement to 0 is a *continuous* function on $\sigma(T)$. Hence we can define $P = f(T) \in L(H)$. We claim that P is non-zero and is a projection to $\mathrm{Ker}(T - \lambda)$; this shows that λ is in the point spectrum.

Indeed, first of all $P \neq 0$ because $\|P\| = \|f\|_{C(\sigma(T))} = 1$, by the functional calculus. Next, we have $f = f^2$ in $C(\sigma(T))$, and therefore $P = f(T) = f(T)^2 = P^2$, which shows that P is a projection. Next, we have the identity

$$(x - \lambda) f(x) = 0, \qquad \text{for all } x \in \sigma(T),$$

of continuous functions on $\sigma(T)$, hence by applying the functional calculus, we get

$$(T - \lambda)P = 0,$$

which shows that $0 \neq \mathrm{Im}(P) \subset \mathrm{Ker}(T - \lambda)$. $\square$

Example 3.9. Let H be a separable Hilbert space, and $T \in K(H)$ a compact self-adjoint operator. Writing

$$T(v) = \sum_{n \geqslant 1} \lambda_n \langle v, e_n \rangle$$

where (λ_n) are the non-zero (real) eigenvalues of T and e_n are corresponding eigenvectors, then $f(T)$ is defined on $\sigma(T) = \{0\} \cup \{\lambda_n\}$ by

$$f(T)v = f(0)P_0(v) + \sum_{n \geqslant 1} f(\lambda_n) \langle v, e_n \rangle e_n$$

where $P_0 \in L(H)$ is the orthogonal projection on $\mathrm{Ker}(T)$.

Example 3.10. Here is an example which is both very simple and extremely important; indeed, in view of the spectral theorem (Theorem 3.1) proved in this chapter, this example describes all self-adjoint bounded operators on a Hilbert space.

Let $H = L^2(X, \mu)$ for a finite measure space (X, μ), and let $g \in L^\infty(X)$ be a real-valued bounded function. The multiplication operator M_g (see Example 1.1) is then self-adjoint on H. The spectrum $\sigma(M_g)$ is the *essential range* of g, defined as follows:

$$\sigma(M_g) = \{x \in \mathbf{R} \mid \mu(g^{-1}(]x - \varepsilon, x + \varepsilon[)) > 0 \text{ for all } \varepsilon > 0\}.$$

Indeed, we have first $M_g - \lambda = M_{g-\lambda}$; formally first, we can solve the equation $(M_g - \lambda)\varphi = \psi$, for any $\psi \in L^2(X, \mu)$, by putting

$$\varphi = \frac{\psi}{g - \lambda},$$

and this is in fact the only solution *as a set-theoretic function* on X. It follows that we have $\lambda \in \rho(M_g)$ if and only if the operator

$$\psi \mapsto (g - \lambda)^{-1}\psi$$

is a bounded linear map on $L^2(X, \mu)$. By Lemma 1.2, we know this is equivalent to asking that $1/(g - \lambda)$ be an L^∞ function on X. This translates to the condition that there exist some $C > 0$ such that

$$\mu(\{x \in X \mid |(g(x) - \lambda)^{-1}| > C\}) = 0,$$

or equivalently

$$\mu(\{x \in X \mid |g(x) - \lambda| < \frac{1}{C}\}) = 0,$$

which precisely says that λ is not in the essential range of g.

Comparing with (1.9) and the definition (1.7) of an image measure, it is convenient to observe that this can also be identified with the *support* of the image measure $\nu = g_*(\mu)$ on $\mathbf{R}$:

$$\sigma(M_g) = \operatorname{supp} g_*(\mu).$$

In particular, if X is a bounded subset of $\mathbf{R}$ and $g(x) = x$, the spectrum of M_x is the support of μ.

For $f \in C(\sigma(M_g))$, the operator $f(M_g)$ is given by $f(M_g) = M_{f \circ g}$. Here, the composition $f \circ g$ is well-defined in $L^\infty(X)$, although the image of g might not lie entirely in $\sigma(M_g)$, because the above description shows that

$$\mu(x \mid g(x) \notin \sigma(M_g)) = \nu(\mathbf{R} - \sigma(M_g)) = 0,$$

(the complement of the support being the largest open set with measure 0) so that, for almost all x, $g(x)$ does lie in $\sigma(M_g)$ and therefore $f(g(x))$ is defined for almost all x (of course we can define arbitrarily the function on the zero-measure subset where $g(x) \notin \sigma(M_g)$, and this does not change the resulting multiplication operator denoted $M_{f \circ g}$).

3.2. Spectral measures

Using the functional calculus, we can clarify now how the spectrum "represents" an operator T and its action on vectors $v \in H$.

Proposition 3.11. *Let H be a Hilbert space, let $T \in L(H)$ be a self-adjoint operator and let $v \in H$ be a fixed vector. There exists a unique positive Radon measure μ on $\sigma(T)$,* depending on T and v, *such that*

$$\int_{\sigma(T)} f(x) d\mu(x) = \langle f(T)v, v\rangle$$

for all $f \in C(\sigma(T))$. In particular, we have

$$\mu(\sigma(T)) = \|v\|^2, \tag{3.7}$$

so μ is a finite measure.

This measure is called the spectral measure *associated to v and T.*

Proof. This is a direct application of the Riesz-Markoc Theorem (Theorem 1.15); indeed, we have the linear functional

$$\ell \begin{cases} C(\sigma(T)) \to \mathbf{C} \\ f \mapsto \langle f(T)v, v\rangle \end{cases}.$$

This is well-defined and positive, since if $f \geqslant 0$, we have $f(T) \geqslant 0$, hence $\langle f(T)v, v\rangle \geqslant 0$ by definition.By the Riesz-Markov theorem, therefore, there exists a unique Radon measure μ on $\sigma(T)$ such that

$$\ell(f) = \langle f(T)v, v\rangle = \int_{\sigma(T)} f(x)d\mu(x)$$

for all $f \in C(\sigma(T))$.

Moreover, taking $f(x) = 1$ for all x, we obtain (3.7) (which also means that $\|\ell\| = \|v\|^2$). □

Example 3.12. Let H be a separable, infinite-dimensional Hilbert space and let $T \in K(H)$ be a compact self-adjoint operator expressed by (2.9) in an orthonormal basis (e_n) of $\mathrm{Ker}(T)^{\perp}$ consisting of eigenvectors for the non-zero eigenvalues $\lambda_n \neq 0$ of T. We have then, by Example 3.9, the formula

$$f(T)v = f(0)P_0(v) + \sum_{n\geqslant 1} f(\lambda_n)\langle v, e_n\rangle e_n$$

for all $v \in H$, where P_0 is the orthogonal projection on $\mathrm{Ker}(T)$. Thus, by definition, we have

$$\int_{\sigma(T)} f(x)d\mu(x) = f(0)\|P_0(v)\|^2 + \sum_{n\geqslant 1} f(\lambda_n)|\langle v, e_n\rangle|^2$$

for all continuous functions f on $\sigma(T)$. Note that

$$\sigma(T) = \{0\} \cup \{\lambda_n \mid n \geqslant 1\},$$

and that, since $\lambda_n \to 0$ as $n \to +\infty$, f is thus entirely described by the sequence $(f(\lambda_n))$, with

$$f(0) = \lim_{n\to+\infty} f(\lambda_n).$$

Hence the formula above means that, as a measure on $\sigma(T)$, μ is a series of Dirac measures at all eigenvalues (including 0) with weight

$$\mu(0) = \|P_0(v)\|^2, \qquad \mu(\lambda_n) = \sum_{\lambda_m=\lambda_n} |\langle v, e_m\rangle|^2,$$

(the sum is needed in case there is an eigenvalue with multiplicity > 1). Equivalently, to be concise in all cases: for all $\lambda \in \sigma(T)$, $\mu(\lambda)$ is equal to $\|v_\lambda\|^2$, where v_λ is the orthogonal projection of v on the eigenspace $\mathrm{Ker}(T - \lambda)$.

This example indicates how, intuitively, one can think of μ in general: it indicates how the vector v is "spread out" among the spectrum; in general, any individual point $\lambda \in \sigma(T)$ carries a vanishing proportion of the vector, because $\mu(\{\lambda\})$ is often zero. But if we consider a positive-measure subset $U \subset \sigma(T)$, $\mu(U) > 0$ indicates that a positive density of the vector is in "generalized eigenspaces" corresponding to that part of the spectrum.

EXAMPLE 3.13. Let (X, μ) be a finite measure space and $T = M_g$ the multiplication operator by a real-valued bounded function on $L^2(X, \mu)$. Consider $\varphi \in L^2(X, \mu)$. What is the associated spectral measure? According to Example 3.10, the functional calculus is defined by $f(T) = M_{f \circ g}$ for f continuous on $\sigma(T)$ (which is the support of the measure $\nu = g_*(\mu)$ on $\mathbf{R}$). We have therefore

$$\langle f(T)\varphi, \varphi \rangle = \int_X f(g(x))|\varphi(x)|^2 d\mu(x) = \int_{\mathbf{R}} f(y) d\tilde{\nu}(y)$$

where

$$\tilde{\nu} = g_*(|\varphi|^2 d\mu),$$

by the standard change of variable formula for image measures. Since the support of $\tilde{\nu}$ is contained in the support of ν, this can be written as

$$\langle f(T)\varphi, \varphi \rangle = \int_{\sigma(T)} f(y) d\tilde{\nu}(y)$$

which means, of course, that the spectral measure associated with φ is the measure $\tilde{\nu}$, restricted to $\sigma(T)$.

Notice the following interesting special cases: if $\varphi = 1$, the spectral measure is simply ν; and if, in addition, $X \subset \mathbf{R}$ is a bounded subset of the real numbers and $g(x) = x$, then the spectral measure is simply μ itself.

3.3. The spectral theorem for self-adjoint operators

Using spectral measures, we can now understand how the spectrum and the functional calculus interact to give a complete description of a self-adjoint operator in $L(H)$.

To see how this works, consider first $v \in H$ and the associated spectral measure μ_v, so that

$$\langle f(T)v, v \rangle = \int_{\sigma(T)} f(x) d\mu_v(x)$$

for all continuous functions f defined on the spectrum of T. In particular, if we apply this to $|f|^2 = f\bar{f}$ and use the properties of the functional calculus, we get

$$\|f(T)v\|^2 = \int_{\sigma(T)} |f(x)|^2 d\mu_v(x) = \|f\|^2_{L^2(\sigma(T), \mu_v)}.$$

In other words, the (obviously linear) map

$$\begin{cases} (C(\sigma(T)), \|\cdot\|_{L^2}) \to H \\ f \mapsto f(T)v \end{cases}$$

is an isometry. The fact that μ_v is a Radon measure implies that continuous functions are dense in the Hilbert space $L^2(\sigma(T), \mu_v)$, and so there is a continuous (isometric) extension

$$U : L^2(\sigma(T), \mu_v) \to H.$$

In general, there is no reason that U should be surjective (think of the case where $v = 0$). However, if we let $H_v = \mathrm{Im}(U) \subset H$, the subspace H_v is closed, and it is stable under T: indeed, the closedness comes from the fact that U is an isometry, and to show that $T(H_v) \subset H_v$, it is enough to show that

$$T(U(f)) \in H_v$$

for $f \in C(\sigma(T))$, or even for f a polynomial function, since the image of the those functions is dense in H_v. But we have indeed

$$T(U(f)) = T(f(T)v) = T\Big(\sum_j \alpha(j)T^j\Big) = \sum_j \alpha(j)T^{j+1}v = (xf)(T)(v) \tag{3.8}$$

which lies in H_v.

We see even more from this last computation. Denote by T_v (for clarity) the restriction of T to H_v:

$$T_v : H_v \to H_v$$

so that U is now an isometric isomorphism

$$U : L^2(\sigma(T_v), \mu_v) \to H_v.$$

We can therefore interpret T_v as (unitarily equivalent to) an operator $S = U^{-1}T_vU$ on $L^2(\sigma(T), \mu_v)$, and since[1] we have

$$T_v(U(f)) = (xf)(T)(v) = U(xf) \tag{3.9}$$

by (3.8), extended by continuity from polynomials to $L^2(\sigma(T), \mu_v)$, it follows that

$$S(f)(x) = xf(x)$$

(in L^2), i.e., that S is simply the multiplication operator M_x defined on $L^2(\sigma(T), \mu_v)$. This is therefore a special case of Theorem 3.1 which we have now proved, for the case where some vector v is such that $H_v = H$.

Remark 3.14. It is extremely important in this reasoning to keep track of the measure μ_v, which depends on the vector v, and to remember that L^2 functions are defined up to functions which are zero almost everywhere. Indeed, it could well be that v has support "outside" a fairly sizable part of the spectrum, and then the values of a continuous function f *on this part* are irrelevant in seeing f as an L^2 function for μ_v: the map

$$C(\sigma(T)) \to L^2(\sigma(T), \mu_v)$$

is not necessarily injective.

The standard general terminology is the following:

Definition 3.15. Let H be a Hilbert space and let $T \in L(H)$. A vector $v \in H$ is called a *cyclic vector for* T if the vectors $T^n(v)$, $n \geqslant 0$, span a dense subspace of H. In particular, H is then separable.

By the density of polynomials in $C(\sigma(T))$, a vector v is cyclic for a self-adjoint operator if and only if $H_v = H$ in the notation above.

It is not always the case that T admits a cyclic vector. However, we have the following lemma which allows us to reduce many questions to this case:

Lemma 3.16. *Let H be a Hilbert space and let $T \in L(H)$ be a self-adjoint operator. Then there exists a family $(H_i)_{i\in I}$ of non-zero, pairwise orthogonal, closed subspaces of H such that H is the orthogonal direct sum of the H_i, $T(H_i) \subset H_i$ for all i, and T restricted to H_i is, for all i, a self-adjoint bounded operator in $L(H_i)$ which has a cyclic vector.*

[1] We write x for the function $x \mapsto x$ to simplify notation.

Proof. It is clearly a matter of iterating the construction of H_v above (since T_v, by definition, admits v as cyclic vector), or rather, given the possibly infinite nature of the induction, it is a matter of suitably applying Zorn's Lemma. We sketch this quite quickly, since the details are very straightforward and close to many of its other applications. First of all, we dispense with the case $H = 0$ (in which case one takes $I = \emptyset$). Note also that since $v \in H_v$ as defined above, we have $H_v = 0$ if and only if $v = 0$.

Let $\mathcal{O}$ be the set of subsets $I \subset H - \{0\}$ such that the spaces H_v for $v \in I$ are pairwise orthogonal, ordered by inclusion. We can apply Zorn's Lemma to $(\mathcal{O}, \subset)$: indeed, if $\mathcal{T}$ is a totally ordered subset of $\mathcal{O}$, we define (as usual...)

$$J = \bigcup_{I \in \mathcal{T}} I \subset H,$$

and if v, w are in J, they belong to some I_1, I_2, in $\mathcal{T}$, respectively, and one of $I_1 \subset I_2$ or $I_2 \subset I_1$ must hold; in either case, the definition of $\mathcal{O}$ shows that H_v and H_w are (non-zero) orthogonal subspaces. Consequently, J is an upper bound for $\mathcal{T}$ in $\mathcal{O}$.

Now, applying Zorn's Lemma, we get a maximal element $I \in \mathcal{O}$. Let

$$H_1 = \bigoplus_{v \in I} H_v,$$

where the direct sum is orthogonal and taken in the Hilbert space sense, so elements of H_1 are sums

$$v = \sum_{i \in I} v_i, \qquad v_i \in H_i,$$

with

$$\|v\|^2 = \sum_{i \in I} \|v_i\|^2 < +\infty.$$

To conclude the proof, we must show that $H_1 = H$. Because H_1 is closed (by definition of the Hilbert sum), if it is not the case, there exists $v_0 \in H_1^{\perp} - \{0\}$, and then

$$I' = I \cup \{v_0\}$$

clearly lies in $\mathcal{O}$ and is strictly larger than I. So by maximality, we must have $H_1 = H$ indeed. □

Note that if H is separable, the index set in the above result is either finite or countable, since each H_i is non-zero.

We can now prove Theorem 3.1 for self-adjoint operators.

Theorem 3.17 (Spectral theorem for self-adjoint operators). *Let H be a separable Hilbert space and $T \in L(H)$ a continuous self-adjoint operator. There exists a finite measure space (X, μ), a unitary operator*

$$U : H \to L^2(X, \mu)$$

and a bounded function $g \in L^\infty(X, \mu)$, such that

$$M_g \circ U = U \circ T.$$

Proof. Consider a family $(H_n)_{n \geqslant 1}$ (possibly with finitely many elements only) of pairwise orthogonal non-zero closed subspaces of H, spanning H, for which $T(H_n) \subset H_n$ and T has a cyclic vector $v_n \neq 0$ on H_n. By replacing v_n with $2^{-n/2}\|v_n\|^{-1}v_n$, we can assume that $\|v_n\|^2 = 2^{-n}$. Let $\mu_n = \mu_{v_n}$ be the spectral measure associated with v_n (and T), so that

$$\mu_n(\sigma(T)) = \|v_n\|^2 = 2^{-n}.$$

By the argument at the beginning of this section, we have unitary maps

$$U_n : L^2(\sigma(T), \mu_n) \to H_n \subset H,$$

such that $U_n^{-1} T U_n = M_x$, the operator of multiplication by x. Now define

$$X = \{1, 2, \ldots, n, \ldots\} \times \sigma(T),$$

with the product topology, and the Radon measure defined by

$$\mu(\{n\} \times A) = \mu_n(A)$$

for $n \geqslant 1$ and $A \subset \sigma(T)$ measurable. It is easily checked that this is indeed a measure. In fact, functions on X correspond to sequences of functions (f_n) on $\sigma(T)$ by mapping f to (f_n) with

$$f_n(x) = f(n, x),$$

and

$$\int_X f(x) d\mu(x) = \sum_{n \geqslant 1} \int_{\sigma(T)} f_n(x) d\mu_n(x)$$

whenever this makes sense (e.g., if $f \geqslant 0$, which is equivalent with $f_n \geqslant 0$ for all n, or if f is integrable, which is equivalent with f_n being μ_n-integrable for all n). In particular

$$\mu(X) = \sum_{n \geqslant 1} \mu_n(\sigma(T)) = \sum_{n \geqslant 1} 2^{-n} < +\infty,$$

so (X, μ) is a finite measure space. Moreover, the map

$$V \begin{cases} L^2(X, \mu) \to \bigoplus_{n \geqslant 1} L^2(\sigma(T), \mu_n) \\ f \mapsto (f_n) \end{cases}$$

is then clearly a surjective isometry. We construct U by defining

$$U\Big(\sum_{n \geqslant 1} w_n\Big) = V^{-1}\Big(\sum_{n \geqslant 1} U_n^{-1}(w_n)\Big)$$

for all $w_n \in H_n$; since all H_n together span H, this is a linear map defined on all of H, and it is a unitary map with inverse

$$U^{-1}(f) = \sum_{n \geqslant 1} U_n(f_n).$$

Then consider

$$g \begin{cases} X \mapsto \mathbf{C} \\ (n, x) \mapsto x \end{cases}.$$

which is bounded and measurable, and finally observe that the n-th component of $U(v)$, for v expressed as

$$v = \sum_{n \geqslant 1} U_n(f_n),$$

is $U_n^{-1}(U_n(f_n)) = f_n$, hence the n-th component of $U(T(v))$ is

$$T(f_n) = x f_n,$$

which means exactly that $U \circ T = M_g \circ U$. □

This spectral theorem is extremely useful; it immediately implies a number of results which could also be proved directly from the continuous functional calculus, but less transparently.

Note that the method of proof (treating first the case of cyclic operators, and then extend using Zorn's Lemma) may also be a shorter approach to other corollaries, since in the cyclic case one knows that the multiplication function can be taken to be the identity on the spectrum.

Example 3.18. We continue with the example of a multiplication operator $T = M_g$ associated with a bounded function g, acting on $H = L^2(X, \mu)$ for a finite measure space (X, μ). For a given $\varphi \in H$, it follows from the previous examples that H_φ is the subspace of functions of the type $x \mapsto f(g(x))\varphi(x)$ for $f \in C(\sigma(T))$, the spectrum being the support of $g_*(\mu)$. If we select the special vector $\varphi = 1$, this is the space of functions $f(g(x))$. This may or may not be dense; for instance, if $X \subset \mathbf{R}$ and $g(x) = x$, this space is of course dense in H; if, say, $X = [-1, 1]$, μ is Lebesgue measure and $g(x) = x^2$, this is the space of even functions in L^2, which is not dense, so φ is not a cyclic vector in this case.

Corollary 3.19 (Positivity). *Let H be a separable Hilbert space and let $T \in L(H)$ be a self-adjoint operator. For $f \in C(\sigma(T))$, we have $f(T) \geqslant 0$ if and only if $f \geqslant 0$.*

Proof. Because of (3.3), we only need to check that $f(T) \geqslant 0$ implies that $f \geqslant 0$. But two unitarily equivalent operators are simultaneously either positive or not, so it suffices to consider an operator $T = M_g$ acting on $L^2(X, \mu)$ for a finite measure space (X, μ). But then we have $f(M_g) = M_{f\circ g}$ by Example 3.10, hence

$$\langle f(M_g)\varphi, \varphi\rangle = \int_X f(g(x))|\varphi(x)|^2 d\mu(x)$$

for all vectors $\varphi \in L^2(X, \mu)$.

The non-negativity of this for all φ implies that $f \geqslant 0$ everywhere, as desired: take

$$\varphi(x) = \chi(g(x)),$$

where χ is the characteristic function of $A = \{y \mid f(y) < 0\}$, to get

$$\int_{\sigma(T)} f(y)\chi(y)d\nu(y) = \int_A f(y)d\nu(y) \geqslant 0, \qquad \nu = g_*(\mu),$$

since $\sigma(T)$ is the support of the image measure. It follows that $\nu(A) = 0$, so f is non-negative almost everywhere on $\sigma(T)$, and since it is continuous (and the support of ν is the whole spectrum), this means in fact that $f \geqslant 0$ everywhere. □

Example 3.20 (Compact operators). We illustrate further the use of the spectral theorem by showing how to derive from it the spectral theorem for *compact* self-adjoint operators (Theorem 2.5). Of course, the latter is more general and can be proved directly more easily, but this is nevertheless a good indication of the fact that Theorem 3.17 can be used as a black-box encapsulating the basic properties of self-adjoint operators.

Thus we assume given a separable Hilbert space H and a compact self-adjoint operator T on H. We wish to show that, except for 0, the spectrum of T consists of eigenvalues with finite multiplicity. After applying Theorem 3.17, we can assume that $H = L^2(X, \mu)$ and that $T = M_g$ for some finite measure space (X, μ) and bounded real-valued function g on X.

Let $\varepsilon > 0$ be given. We want to show that (up to sets of measure 0), g only takes finitely many values with absolute value $\geqslant \varepsilon$. Let

$$A_\varepsilon = \{x \mid |g(x)| \geqslant \varepsilon\} \subset X,$$

and let

$$L_\varepsilon = \{\varphi \in L^2(X,\mu) \mid \varphi = 0 \text{ almost everywhere outside } A_\varepsilon\} \subset L^2(X,\mu).$$

We will first show that $\dim L_\varepsilon < +\infty$. First, the subspace L_ε is a closed subspace of $L^2(X,\mu)$, since $\varphi \in L_\varepsilon$ if and only if

$$\int_{X-A_\varepsilon} |\varphi(x)|^2 d\mu(x) = 0.$$

Now let C_ε be the ball centered at 0 with radius ε^{-2} in $L^2(X,\mu)$. We claim that the unit ball of L_ε is contained in $M_g(C_\varepsilon)$. Indeed, if $\varphi \in L_\varepsilon$ has norm $\leqslant 1$, defining h_ε to be the function

$$h_\varepsilon(x) = \begin{cases} 1/g(x) & \text{if } x \in A_\varepsilon, \\ 0 & \text{otherwise.} \end{cases},$$

we see that $M_g(h_\varepsilon\varphi) = \varphi$, with

$$\|h_\varepsilon\varphi\|^2 = \int_{A_\varepsilon} \frac{1}{|g(x)|^2} |\varphi(x)|^2 d\mu(x) \leqslant \varepsilon^{-2}.$$

By compactness, it follows that the unit ball of L_ε is relatively compact, and therefore that $\dim L_\varepsilon < +\infty$.

Now we wish to deduce from this that the intersection of the spectrum of M_g with $\{y \in \mathbf{R} \mid |y| \geqslant \varepsilon\}$ is a finite set. Let σ_ε be this intersection. We consider the measure $\nu = g_*\mu$, so that the spectrum of M_g is the support of ν, and the restriction ν_ε of ν to σ_ε. Let then

$$\tilde{L}_\varepsilon = \Big\{\psi \in L^2(\sigma(M_g),\nu) \mid \int_{\sigma-\sigma_\varepsilon} |\psi(x)|^2 d\nu(x) = 0\Big\} = L^2(\sigma_\varepsilon, \nu_\varepsilon).$$

Note that the linear map

$$\begin{cases} \tilde{L}_\varepsilon & \longrightarrow L_\varepsilon \\ \psi & \mapsto g \circ \psi \end{cases}$$

is well-defined and injective (it is isometric, by the integration formula for image measures), and therefore we also have $\dim \tilde{L}_\varepsilon < +\infty$. Now we use the following easy lemma:

Lemma 3.21. *Let $B \subset \mathbf{R}$ be a compact set, let ν be a Radon measure with support equal to B. If $\dim L^2(B,\nu)$ is finite, then B is a finite set and $\nu(\{y\}) > 0$ for all $y \in B$.*

Assuming this, it follows that there are finitely many elements

$$\{y_1, \ldots, y_{n(\varepsilon)}\} \subset \sigma_\varepsilon$$

such that $\mu(g^{-1}(y_i)) > 0$ and such that $g(x) \in \{y_1, \ldots, y_{n(\varepsilon)}\}$ for μ-almost every $x \in A_\varepsilon$. Since the y_i are isolated points of the spectrum, they are eigenvalues, and since the corresponding eigenspaces are contained in L_ε, they are finite-dimensional.

Letting $\varepsilon \to 0$ along some sequence, we conclude that the non-zero spectrum of M_g consists of (at most) countably many eigenvalues with finite multiplicity.

Proof of the lemma. Since ν is a Radon measure, the space $C(B)$ of continuous functions on B is dense in $L^2(B,\nu)$, and since the latter has finite dimension and the support of ν is all of B, we have $C(B) = L^2(B,\nu)$, and they are isomorphic as Banach spaces. For $x \in B$, the linear map $\psi \mapsto \psi(x)$ is continuous on $C(B)$, hence on $L^2(B,\nu)$. Since this linear form is non-zero (e.g., for the constant function 1), this implies that $\nu(\{x\}) > 0$. Thus every $x \in B$ has positive measure, and

$$L^2(B,\nu) = \{\psi \,:\, B \longrightarrow \mathbf{C} \mid \sum_{x\in B} \nu(x)|\psi(x)|^2 < +\infty\}.$$

The condition $\dim L^2(B,\nu) < +\infty$ therefore clearly implies that $|B| < +\infty$. □

3.4. Projection-valued measures

In this section, we describe another version of the Spectral Theorem 3.17, still for self-adjoint operators, which is essentially equivalent but sometimes more convenient. Moreover, it allows us to introduce some new concepts in a well-motivated way.

The idea is to generalize the following interpretation of the spectral theorem for a compact self-adjoint operator $T \in K(H)$: if we denote by P_λ the orthogonal projection onto $\mathrm{Ker}(T-\lambda)$, for $\lambda \in \mathbf{R}$, then we have the relations

$$v = \sum_{\lambda\in\mathbf{R}} P_\lambda(v), \qquad T(v) = \sum_{\lambda\in\mathbf{R}} \lambda P_\lambda(v),$$

valid for all $v \in H$, where the series are well-defined because $P_\lambda = 0$ for $\lambda \notin \sigma(T)$. To generalize this, it is natural to expect that one must replace the series with integrals. Thus some form of integration for functions taking values in H or $L(H)$ is needed. Moreover, $\mathrm{Ker}(T-\lambda)$ may well be zero for all λ, and the projections must be generalized. We consider these two questions abstractly first:

Definition 3.22 (Projection-valued measure). Let H be a Hilbert space and let $P(H)$ denote the set of orthogonal projections in $L(H)$. A *(finite) projection valued measure* Π on H is a map

$$\begin{cases} \mathcal{B}(\mathbf{R}) \to P(H) \\ A \mapsto \Pi_A \end{cases}$$

from the σ-algebra of Borel subsets of $\mathbf{R}$ to the set of projections, such that the following conditions hold:

– (1) $\Pi_\emptyset = 0$, $\Pi_{\mathbf{R}} = \mathrm{Id}$;
– (2) For some constant $R > 0$, we have $\Pi_{[-R,R]} = \mathrm{Id}$;
– (3) If A_n, $n \geqslant 1$, is an arbitrary sequence of pairwise disjoint Borel subsets of $\mathbf{R}$, let

$$A = \bigcup_{n\geqslant 1} A_n \in \mathcal{B}(\mathbf{R}),$$

and then we have

$$\Pi_A = \sum_{n\geqslant 1} \Pi_{A_n} \tag{3.10}$$

where the series converges in the "strong operator topology" of H, which is by definition the topology on $L(H)$ described by the seminorms

$$p_v \begin{cases} L(H) \to [0,+\infty[\\ T \mapsto \|T(v)\| \end{cases}$$

for $v \in H$.

REMARK 3.23. Of course, if the sequence (A_n) is finite, say $n \leqslant N$, the equality

$$\Pi_A = \sum_{n=1}^{N} \Pi_{A_n}$$

holds in $L(H)$.

The following are easy properties of the strong operator topology: (i) it is Hausdorff; (ii) if is weaker than the Banach-space topology given by the operator norm; (iii) a sequence (T_n) converges to T in the strong operator topology if and only if $T_n(v) \to T(v)$ for all v. Thus (3.10) means that

$$\Pi_A(v) = \sum_{n \geqslant 1} \Pi_{A_n}(v), \qquad \text{for all } v \in H.$$

If $(e_n)_{n\geqslant 1}$ is an orthonormal basis of a separable Hilbert space H, note that the projection P_n onto $\mathbf{C} \cdot e_n \subset H$ are such that, for any $v \in H$, we have

$$P_n(v) = \langle v, e_n \rangle e_n \to 0$$

as $n \to +\infty$ (since

$$\|v\|^2 = \sum_{n \geqslant 1} |\langle v, e_n \rangle|^2,$$

the coefficients $\langle v, e_n \rangle$ converge to 0 as $n \to +\infty$). Thus P_n converges strongly to 0, but of course $\|P_n\| = 1$ for all n, so (P_n) does not converge in $L(H)$ for the operator norm.

This definition resembles that of a (finite) Borel measure on $\mathbf{R}$. The following elementary properties are therefore not surprising:

LEMMA 3.24. *Let H be a Hilbert space and Π a projection-valued measure on H. Then:*

(1) *For $A \subset B$ measurable, $\Pi_A \leqslant \Pi_B$ and $\Pi_A \Pi_B = \Pi_B \Pi_A = \Pi_A$.*

(2) *For A, B measurable subsets of $\mathbf{R}$, we have $\Pi_{A\cap B} = \Pi_A \Pi_B = \Pi_B \Pi_A$. In particular, all projections Π_A commute, and if $A \cap B = \emptyset$, we have $\Pi_A \Pi_B = 0$.*

PROOF. (1) We have $B = A \cup (B - A)$, a disjoint union, hence by (3.10) we have

$$\Pi_B = \Pi_A + \Pi_{B-A},$$

and since Π_{B-A} is an orthogonal projection, it is $\geqslant 0$ (since $\langle P(v), v \rangle = \|P(v)\|^2$ for any orthogonal projection P). Moreover, we recall that whenever P_1, P_2 are orthogonal projections on H_1, H_2, respectively, we have

$$P_1 \geqslant P_2 \Rightarrow H_2 \subset H_1 \Rightarrow P_1 P_2 = P_2 P_1 = P_2$$

(indeed, since $P_1 \geqslant P_2$, we derive that $H_2 \cap H_1^{\perp} = 0$, since if v belongs to this subspace, we have

$$0 = P_1(v) = P_2(v) + (P_1 - P_2)(v) = v + (P_1 - P_2)(v),$$

hence (by positivity of the second term)

$$\|v\|^2 \leqslant \|v\|^2 + \langle (P_1 - P_2)v, v \rangle = 0,$$

so that $v = 0$; then $H_2 \cap H_1^{\perp} = 0$ implies $H_2 \subset H_1$; then from $H_2 \subset H_1$, we have directly that $P_2(P_1(v)) = P_2(v)$ and $P_1(P_2(v)) = P_2(v)$ for all $v \in H$).

In our case, with $P_1 = \Pi_B$, $\Pi_2 = \Pi_A$, this gives

$$\Pi_A \Pi_B = \Pi_B \Pi_A = \Pi_A.$$

(2) We start with the case $A \cap B = \emptyset$. Then we have

$$\Pi_{A\cup B} = \Pi_A + \Pi_B,$$

and multiplying by Π_A, since $A \subset A \cup B$, the first part gives

$$\Pi_A = \Pi_A \Pi_{A\cup B} = \Pi_A^2 + \Pi_A\Pi_B = \Pi_A + \Pi_A\Pi_B,$$

so that $\Pi_A\Pi_B = 0$. Similarly, of course, we have $\Pi_B\Pi_A = 0$.

Next, for any A and B, notice that we have a disjoint intersection

$$A = (A - B) + (A \cap B),$$

hence

$$\Pi_A = \Pi_{A-B} + \Pi_{A\cap B},$$

and multiplying by Π_B this time gives

$$\Pi_B\Pi_A = \Pi_B\Pi_{A-B} + \Pi_B\Pi_{A\cap B} = \Pi_{A\cap B}$$

because $B \cap (A - B) = \emptyset$ (and we apply the special case just proved) and $A \cap B \subset B$ (and we apply (1) again). Similarly, we get $\Pi_A\Pi_B = \Pi_{A\cap B}$. □

As expected, the point of projection-valued measures is that one can integrate with respect to them, and construct operators in $L(H)$ using this formalism.

Proposition 3.25. *Let H be a Hilbert space and let Π be a projection-valued measure on H. Then, for any bounded Borel function $f : \mathbf{R} \to \mathbf{C}$, there exists a unique operator $T \in L(H)$ such that*

$$\langle T(v), v\rangle = \int_{\mathbf{R}} f(\lambda) d\mu_v(\lambda) \tag{3.11}$$

for all $v \in H$, where μ_v is the finite Borel measure given by

$$\mu_v(A) = \langle \Pi_A(v), v\rangle, \qquad \textit{for } A \in \mathcal{B}(\mathbf{R}). \tag{3.12}$$

This operator is denoted

$$T = \int_{\mathbf{R}} f(\lambda) d\Pi(\lambda) = \int_{\mathbf{R}} f(\lambda) d\Pi_\lambda.$$

We have moreover

$$T^* = \int_{\mathbf{R}} \overline{f(\lambda)} d\Pi(\lambda),$$

and the operator T is normal. It is self-adjoint if f is real-valued and positive if f is non-negative.

If T is the self-adjoint operator associated to a projection valued measure Π, it is also customary to write

$$f(T) = \int_{\mathbf{R}} f(\lambda) d\Pi(\lambda). \tag{3.13}$$

If f is continuous, this coincides with the functional calculus for T.

Proof. Let Π be a projection valued measure on H, and let v be any fixed vector. We define μ_v on Borel subsets of $\mathbf{R}$ as indicated by (3.12), and we first check that it is indeed a Borel measure. Since any orthogonal projection is positive, μ_v takes non-negative; we also have

$$\mu_v(\emptyset) = 0, \quad \mu_v(\mathbf{R}) = \|v\|^2$$

by the first property defining projection-valued measures, and if A_n is a sequence of disjoint Borel subsets and A their union, we have

$$\mu_v(A) = \langle \Pi_A(v), v\rangle = \sum_n \langle \Pi_{A_n} v, v\rangle = \sum_n \mu_v(A_n),$$

by (3.10) and the definition of strong convergence of sequences. So we do have a finite measure. In particular, if f is a bounded measurable function on $\mathbf{R}$, the integral

$$\int_{\mathbf{R}} f(\lambda) d\mu_v(\lambda)$$

exists for all $v \in H$. If there exists an operator $T \in L(H)$ such that

$$\langle T(v), v\rangle = \int_{\mathbf{R}} f(\lambda) d\mu_v(\lambda),$$

for all v, we know that it is uniquely determined by those integrals, and this gives the unicity part of the statement.

To show the existence, we simply parallel the construction of integration with respect to a measure (one could be more direct by showing directly that the right-hand side of the equality above is of the form $\langle T(v), v\rangle$ for some $T \in L(H)$, but the longer construction is instructive for other reasons anyway).

We start by defining

$$\int_{\mathbf{R}} \chi_A(\lambda) d\Pi(\lambda) = \Pi_A$$

for any Borel subset $A \subset \mathbf{R}$, where χ_A is the characteristic function of A. The definition (3.12) exactly means that this definition is compatible with our desired statement (3.11) for $f = \chi_A$, i.e., we have

$$\langle \Big(\int_{\mathbf{R}} f d\Pi \Big) v, v\rangle = \int_{\mathbf{R}} f(\lambda) d\mu_v(\lambda), \tag{3.14}$$

for all $v \in H$.

We then extend the definition by linearity for *step functions*

$$f = \sum_{1 \leqslant i \leqslant N} \alpha_i \chi_{A_i}$$

where $A_i \subset \mathbf{R}$ are disjoint measurable sets, namely

$$\int_{\mathbf{R}} f d\Pi = \sum_{1 \leqslant i \leqslant N} \alpha_i \Pi_{A_i} \ ;$$

again, linearity ensures that (3.14) holds for such f, and unicity ensures that the resulting operator does not depend on the representation of f as a sum of characteristic functions.

Next, for $f \geqslant 0$, bounded and measurable, it is well-known that we can find step functions $s_n \geqslant 0$, $n \geqslant 1$, such that (s_n) converges uniformly to f: indeed, if $0 \leqslant f \leqslant B$, one can define

$$s_n(x) = \frac{iB}{n}, \text{ where } 0 \leqslant i \leqslant n-1 \text{ is such that } f(x) \in [iB/n, (i+1)B/n[$$

(and $s_n(x) = B$ if $f(x) = B$), so that $|f(x) - s_n(x)| \leqslant B/n$ for all x.

We will show that

$$T_n = \int_{\mathbf{R}} s_n d\Pi$$

converges in $L(H)$ to an operator T such that (3.14) holds. Indeed, for any step function s, we can write

$$s = \sum_{1 \leqslant i \leqslant N} \alpha_i \chi_{A_i}, \qquad A_i \text{ disjoint}, \qquad s^2 = \sum_{1 \leqslant i \leqslant N} \alpha_i^2 \chi_{A_i}$$

and by definition we get

$$\Bigl\| \Bigl(\int_{\mathbf{R}} s d\Pi \Bigr) v \Bigr\|^2 = \sum_{1 \leqslant i \leqslant N} \alpha_i^2 \|\Pi_{A_i}(v)\|^2 \leqslant \max |\alpha_i|^2 \|v\|^2, , \tag{3.15}$$

for all v (using Lemma 3.24, (2)), so

$$\Bigl\| \int_{\mathbf{R}} s d\Pi \Bigr\|_{L(H)} \leqslant \|s\|_{L^\infty}.$$

Applied to $s = s_n - s_m$, this shows that the sequence (T_n) is a Cauchy sequence in $L(H)$, hence it does converge to some operator $T \in L(H)$. We can then argue that, by continuity, we have

$$\langle T(v), v \rangle = \lim_{n \to +\infty} \langle T_n(v), v \rangle = \lim_{n \to +\infty} \int_{\mathbf{R}} s_n d\mu_v = \int_{\mathbf{R}} f d\mu_v$$

by the dominated convergence theorem (since $0 \leqslant s_n \leqslant f$ which is bounded, hence integrable with respect to a finite measure). This means that T satisfies (3.14), as desired.

We are now essentially done: given a bounded complex-valued function $f : \mathbf{R} \to \mathbf{C}$, we write

$$f = (\mathrm{Re}(f)_+ - \mathrm{Re}(f)_-) + i(\mathrm{Im}(f)_+ - \mathrm{Im}(f)_-),$$

where each of the four terms is $\geqslant 0$, and we define $\int f d\Pi$ by linearity from this expression. Again, (3.14) holds trivially.

To conclude the proof of the proposition, we note first that

$$\langle T^*(v), v \rangle = \overline{\langle T(v), v \rangle} = \overline{\int_{\mathbf{R}} f(\lambda) d\mu_v(\lambda)} = \int_{\mathbf{R}} \overline{f(\lambda)} d\mu_v(\lambda),$$

which shows that

$$T^* = \int_{\mathbf{R}} \overline{f(\lambda)} d\Pi(\lambda).$$

Generalizing (3.15) one shows that, for all f, we have

$$\Bigl\| \Bigl(\int_{\mathbf{R}} f d\Pi \Bigr) v \Bigr\|^2 = \int_{\mathbf{R}} |f|^2 d\mu_v = \Bigl\| \Bigl(\int_{\mathbf{R}} \bar{f} d\Pi \Bigr) v \Bigr\|^2,$$

and since T is normal if and only if $\|T(v)\|^2 = \|T^*(v)\|^2$ for all v, we deduce that any of the operators of the type $\int_{\mathbf{R}} f d\Pi$ is normal. Finally, the self-adjointness for f real-valued, and the positivity for $f \geqslant 0$, are clear from the construction. ☐

Example 3.26. Let Π be a projection-valued measure. We then have

$$\mathrm{Id} = \int_{\mathbf{R}} d\Pi(\lambda).$$

Corollary 3.27. *Let H be a Hilbert space, Π a finite projection-valued measure on H. Let (f_n) be a sequence of bounded measurable functions $\mathbf{R} \to \mathbf{C}$ such that*

$$f_n(x) \to f(x)$$

with f measurable and bounded, for all $x \in \mathbf{R}$ and with $\|f_n\|_\infty$ bounded for $n \geqslant 1$. Then

$$\int_{\mathbf{R}} f_n(\lambda) d\Pi(\lambda) \to \int_{\mathbf{R}} f(\lambda) d\Pi(\lambda)$$

in the strong topology.

Proof. Let T be the self-adjoint operator defined by Π. The left and right-hand sides are well-defined operators by Proposition 3.25, given by $f_n(T)$ and $f(T)$ respectively (using the notation 3.13), and it is enough to prove that, for any vector $v \in H$, we have

$$\langle f_n(T)v, v\rangle \to \langle f(T)v, v\rangle.$$

By definition, the left-hand side is

$$\int_{\mathbf{R}} f_n(\lambda) d\mu_v(\lambda)$$

where the measure μ_v, as in (3.12), is uniquely determined by Π and v. But the assumptions imply that $f_n(\lambda) \to f(\lambda)$ pointwise, and that $|f_n| \leqslant C$ where C is some constant independent of n; since μ_v is a finite measure on $\mathbf{R}$, this constant is μ_v-integrable and therefore we can apply the dominated convergence theorem to conclude that

$$\int_{\mathbf{R}} f_n(\lambda) d\mu_v(\lambda) \to \int_{\mathbf{R}} f(\lambda) d\mu_v(\lambda) = \langle f(T)v, v\rangle,$$

as desired. □

Now we have a new version of the spectral theorem:

Theorem 3.28 (Spectral theorem in projection-valued measure form). *Let H be a separable Hilbert space and let $T = T^*$ be a bounded self-adjoint operator on H. There exists a unique projection-valued measure Π_T such that*

$$T = \int_{\mathbf{R}} \lambda d\Pi_T(\lambda),$$

where the integral is extended to the unbounded function λ by defining it as

$$\int_{\mathbf{R}} \lambda \; d\Pi_T(\lambda) = \int \lambda\chi \; d\Pi_T,$$

where χ is the characteristic function of some interval $I = [-R, R]$ for which $\Pi_{T,I} = \mathrm{Id}$.

Moreover, if f is any continuous function on $\sigma(T)$, we have

$$f(T) = \int_{\sigma(T)} f(\lambda) d\Pi_T(\lambda).$$

Proof. By the Spectral Theorem (Theorem 3.17), we can assume that $T = M_g$ is a multiplication operator acting on $H = L^2(X, \mu)$ for some finite measure space (X, μ), and g a real-valued function in $L^\infty(X, \mu)$. For $A \subset \mathbf{R}$ a Borel subset, we then define

$$\Pi_{T,A} = M_{\chi_A \circ g},$$

the multiplication operator by $\chi_A \circ g \in L^\infty(X, \mu)$. We now check that

$$A \mapsto \Pi_{T,A}$$

is a projection-valued measure.

It is clear that

$$\Pi_{T,A}^2 \varphi = \chi_A(g(x))^2 \varphi(x) = \chi_A(g(x))\varphi = \Pi_{T,A}\varphi,$$

for every $\varphi \in L^2(X,\mu)$, so each $\Pi_{T,A}$ is a projection operator, and since it is self-adjoint (each $\chi_A \circ g$ being real-valued), it is an orthogonal projection. The properties

$$\Pi_{T,\emptyset} = 0, \qquad \Pi_{T,\mathbf{R}} = \Pi_{T,[-R,R]} = \mathrm{Id}$$

are clear if $|g| \leqslant R$. If (A_n) is a sequence of pairwise disjoint Borel subsets of $\mathbf{R}$, denoting their union by A, we have

$$\chi_A(y) = \sum_{n\geqslant 1} \chi_{A_n}(y)$$

for any $y \in \mathbf{R}$, where the series contains at most a single non-zero term. Hence

$$\Pi_{T,A}\varphi(x) = \chi_A(g(x))\varphi(x) = \sum_{n\geqslant 1} \chi_{A_n}(g(x))\varphi(x),$$

for any $\varphi \in L^2(X,\mu)$, showing that

$$\Pi_{T,A} = \sum_{n\geqslant 1} \Pi_{T,A_n},$$

in the strong topology on $L(H)$.

So we have constructed a projection-valued measure from T. Consider then the operator

$$S = \int_{\mathbf{R}} \lambda\chi d\Pi \in L(H),$$

where χ is as described in the statement of the theorem. We will check that $S = T$ as follows: let $\varphi \in H$ be given; we then have by (3.11) that

$$\langle S(\varphi), \varphi\rangle = \int_{\mathbf{R}} \lambda\chi(\lambda) d\mu_\varphi(\lambda)$$

where the measure μ_φ is defined by

$$\mu_\varphi(A) = \langle \Pi_{T,A}\varphi, \varphi\rangle = \int_X \chi_A(g(x))|\varphi(x)|^2 d\mu(x) = \int_{\mathbf{R}} \chi_A(y) d\nu(y),$$

where $\nu = g_*(|\varphi|^2 d\mu)$ is the spectral measure associated to T and φ (see Example 3.13).

On the other hand, we have

$$\langle T(\varphi), \varphi\rangle = \int_X g(x)|\varphi(x)|^2 d\mu(x) = \int_{\mathbf{R}} \lambda d\nu(\lambda).$$

Now, we know that the support of ν is the essential image of g, and hence by the choice of I, we have

$$\langle S(\varphi), \varphi\rangle = \int_{\mathbf{R}} \lambda\chi(\lambda) d\nu(\lambda) = \int_{\mathbf{R}} \lambda d\nu(\lambda) = \langle T(\varphi), \varphi\rangle$$

for all $\varphi \in H$. This means that $S = T$, as desired. □

Intuitively, and the proof illustrates this clearly, $\Pi_{T,A}$ is the orthogonal projection on the subspace of H which is the direct sum of those where T acts "by multiplication by some $\lambda \in A$".

The following lemma will be useful in the next section.

LEMMA 3.29. *Let H be a separable Hilbert space, and let T_1, T_2 be self-adjoint operators in $L(H)$ which commute, with associated projection valued measures Π_1 and Π_2. Then, for bounded measurable functions f and g, the operators*

$$S_1 = \int_{\mathbf{R}} f d\Pi_1, \qquad S_2 = \int_{\mathbf{R}} g d\Pi_2$$

also commute.

PROOF. It is enough to consider the case where g is the identity, so $S_2 = T_2$, because if it holds, we get first that S_1 commutes with T_2, and then the same argument with (T_1, T_2, f) replaced by (T_2, S_1, g) gives the desired conclusion.

Next, a simple limiting argument shows that it is enough to consider the case where f is the characteristic function of a measurable set, so $S_1 = \Pi_{1,A}$ is a projection, and we must show that $\Pi_{1,A}T_2 = T_2\Pi_{1,A}$.

Now, we argue as follows: the assumption implies immediately, by induction, that

$$T_1^n T_2 = T_2 T_1^n \quad \text{for all } n \geqslant 0,$$

so T_2 commutes with all polynomials $p(T_1)$. By continuity of multiplication in $L(H)$, T_2 commutes with all operators $\varphi(T_1)$, $\varphi \in C(\sigma(T_1))$. We know there exists a sequence (φ_n) of such continuous functions with

$$\varphi_n(x) \to \chi_A(x)$$

for all x. By strong convergence (Corollary 3.27), it follows that

$$\varphi_n(T_1) = \int_{\mathbf{R}} \varphi_n d\Pi_1 \to \Pi_A$$

strongly. Then we get

$$T_2(\Pi_A(v)) = \lim_{n\to+\infty} T_2(\varphi_n(T_1)v) = \lim_{n\to+\infty} \varphi_n(T_1)(T_2(v)) = \Pi_A(T_2(v)).$$

□

3.5. The spectral theorem for normal operators

Using the following simple lemma, we are now in a position to extend the spectral theorem and the continuous functional calculus to normal operators.

LEMMA 3.30. *Let H be a Hilbert space and $T \in L(H)$ a normal bounded operator. There exist two self-adjoint operators T_1, $T_2 \in L(H)$ such that $T = T_1 + iT_2$, and $T_1T_2 = T_2T_1$.*

PROOF. Write

$$T_1 = \frac{T + T^*}{2}, \qquad T_2 = \frac{T - T^*}{2i},$$

so that $T = T_1 + iT_2$, and observe first that both are obviously self-adjoint, and then that

$$T_1T_2 = T_2T_1 = \frac{T^2 - (T^*)^2}{4i}$$

because T is normal. □

We now have the basic result for normal operators.

PROPOSITION 3.31. *Let H be a separable Hilbert space and let $T \in L(H)$ be a normal bounded operator. There exists a finite measure space (X, μ), a bounded measurable function $g \in L^\infty(X, \mu)$ and a unitary isomorphism*

$$U : H \to L^2(X, \mu)$$

such that $M_g \circ U = U \circ T$.

Sketch of the proof. Write $T = T_1 + iT_2$ with T_1, T_2 both self-adjoint bounded operators which commute, as in the lemma. Let Π_1 (resp. Π_2) denote the projection valued measure for T_1 (resp. T_2). The idea will be to first construct a suitable projection valued measure associated with T, which must be defined on $\mathbf{C}$ since $\sigma(T)$ is not (in general) a subset of $\mathbf{R}$.

We first claim that all projections $\Pi_{1,A}$ and $\Pi_{2,B}$ commute; this is because T_1 and T_2 commute (see Lemma 3.29). This allows us to define

$$\tilde{\Pi}_{A\times B} = \Pi_{1,A}\Pi_{2,B} = \Pi_{2,B}\Pi_{1,A},$$

which are orthogonal projections. By basic limiting procedures, one shows that the mapping

$$A \times B \mapsto \tilde{\Pi}_{A\times B}$$

extends to a map

$$\mathcal{B}(\mathbf{C}) \to P(H)$$

which is a (finite) projection valued measure *defined on the Borel subsets of* $\mathbf{C}$, the definition of which is obvious. Repeating the previous section allows us to define normal operators

$$\int_{\mathbf{C}} f(\lambda) d\tilde{\Pi}(\lambda) \in L(H),$$

for f bounded and measurable defined on $\mathbf{C}$. In particular, one finds again that

$$T = \int_{\mathbf{C}} \lambda d\tilde{\Pi}(\lambda),$$

where the integral is again defined by truncating outside a sufficiently large compact set.

So we get the spectral theorem for T, expressed in the language of projection-valued measures.

Next one gets, for $f \in C(\sigma(T))$ and $v \in H$, the fundamental relation

$$\Big\|\Big(\int f d\tilde{\Pi}\Big)v\Big\|^2 = \int |f|^2 d\mu_v,$$

where μ_v is the associated spectral measure. This allows, again, to show that when T has a cyclic vector v (defined now as a vector for which the span of the vectors $T^n v$, $(T^*)^m v$, is dense), the unitary map

$$L^2(\sigma(T), \mu_v) \to H$$

represents T as a multiplication operator M_z on $L^2(\sigma(T), \mu_v)$. And then Zorn's lemma allows us to get the general case. □

CHAPTER 4

Unbounded operators on a Hilbert space

This chapter describes the basic formalism of unbounded operators defined on a dense subspace of a Hilbert space, and uses this together with the spectral theorem for bounded operators to prove a very similar spectral theorem for self-adjoint unbounded operators.

It should be emphasized here that, although one can develop a fairly flexible formalism, unbounded operators remain somewhat "wild", and that it is perfectly possible to make important mistakes if one forgets the fairly subtle conditions that define, for instance, a self-adjoint operator in this setting.

4.1. Basic definitions

Motivating examples of unbounded operators have already been described: these were the Laplace operator in Example 1.6, and the operators defined as the "inverses" of $T-\lambda$ for a bounded operator $T \in L(H)$ and $\lambda \in \sigma_c(T)$ (so that $\mathrm{Im}(T-\lambda)$, on which the inverse is defined, is a *dense* subset of H): see Remark 2.3. The other class of basic examples is that of multiplication operators by general (measurable) functions g on a space $L^2(X, \mu)$, where g is not necessarily bounded (see Example 4.6 below). In fact, the basic goal of this chapter is to show that this last class represents all self-adjoint operators up to unitary equivalence – in other words, we want an extension of Theorem 3.1 for *unbounded* operators.

We now formalize the type of situation described by those examples.

Definition 4.1. Let H be a Hilbert space. A *(densely) defined* operator on H is a *pair* $(D(T), T)$ where $D(T) \subset H$ is a dense subspace of H, called the *domain* of the operator, and

$$T : D(T) \to H$$

is a linear map. We denote by $DD(H)$ the set of densely defined operators on H.[1]

Remark 4.2. The linear map T, defined on $D(T)$ is usually *not continuous*. Indeed, if T happens to be in $L(D(T), H)$, the fact that $D(T)$ is dense ensures that T extends in a unique way to a bounded operator $\tilde{T} \in L(H)$, and there is no gain of generality in looking at dense domains.

One may wonder why one does not study, instead of densely defined operators which are not continuous, those which are (still) not continuous but defined on the whole of H – indeed, Zorn's lemma easily shows that any densely defined operator $(D(T), T)$ can be seen as the restriction (in the sense explained below) of a linear map defined on H itself. However, constructing these extensions must involve the Axiom of Choice if T was not continuous on $D(T)$, and this means that it is basically impossible to say what $T(v)$ is, for any vector *except* those in $D(T)$.

It is of great importance to emphasize immediately that, although we will be tempted to write things like "let $T \in DD(H)$ be given", this is a shorthand notation: the domain $D(T)$ is part of the data. In particular, if (D_1, T_1) and (D_2, T_2) are such that $D_1 \subset D_2$

[1] This notation is not standard, though it will be convenient.

and T_2 restricted to D_1 is equal to T_1, we *do not* identify them. This leads us to the next definition.

DEFINITION 4.3. Let H be a Hilbert space.

(1) If $(D(T), T)$ is a densely defined linear operator on H, and $D_1 \subset D(T)$ is a subspace of $D(T)$ which is still dense in H, we call

$$(D_1, T|D_1)$$

the *restriction* of the operator to D_1.

(2) An extension of a densely defined operator $(D(T), T)$ is a $(D_1, T_1) \in DD(H)$ such that $D(T) \subset D_1$ and $(D(T), T)$ is the restriction of (D_1, T_1) to $D(T)$.

REMARK 4.4. In the shorthand notation commonly used, if (D_1, T_1) is the restriction of (D_2, T_2) to D_1, one may write

$$T_1 = T_2|D_1, \quad \text{or} \quad T_1 \subset T_2,$$

the latter notation emphasizing the fact that there is an inclusion of domains underlying this restriction. Note that $T_1 = T_2$ (meaning the domains coincide, and the operators are the same on it) if and only $T_1 \subset T_2$ and $T_2 \subset T_1$.

REMARK 4.5. The kernel of $(D(T), T)$ is

$$\mathrm{Ker}(T) = \{v \in D(T) \mid Tv = 0\},$$

and the image is

$$\mathrm{Im}(T) = \{Tv \mid v \in D(T)\}.$$

Note that both *depend* on the domain: if $D(T)$ is replaced by a smaller (dense) subset, then the kernel may become smaller, and similarly for the image.

EXAMPLE 4.6. (1) As in the case of bounded operators (Example 1.1), multiplication operators give fundamental examples of densely-defined unbounded operators. Let (X, μ) be a finite measure space.

For a measurable function $g : X \to \mathbf{C}$, the map

$$M_g : \varphi \mapsto g\varphi$$

is linear and well-defined on the space of measurable functions defined on X, but it is not well-defined as a linear map acting on $L^2(X, \mu)$, since $g\varphi$ may well fail to be square-integrable.

For instance, if we take $X = \mathbf{R}$ with the Lebesgue measure and $\varphi(x) = x$, we have

$$\frac{x}{1 + |x|} \notin L^2(\mathbf{R})$$

although $1/(1 + |x|) \in L^2(\mathbf{R})$. However, we can define

$$D(M_g) = \{\varphi \in L^2(X, \mu) \mid g\varphi \in L^2(X, \mu)\}$$

so that M_g is a well-defined linear map

$$D(M_g) \to L^2(X, \mu).$$

Here is a small **warning**: the definition of $D(M_g)$ is *not* equivalent with $g\varphi \in L^2(X, \mu)$: the condition that φ itself be square integrable must not be forgotten; for instance, the function

$$\varphi(x) = \begin{cases} \frac{1}{\sqrt{|x|}} & \text{for} \quad |x| \leqslant 1, \ x \neq 0 \\ \frac{1}{x^2} & \text{for } |x| > 1, \end{cases}$$

is such that

$$\int_{\mathbf{R}} |x\varphi(x)|^2 dx = \int_{-1}^{1} |x| dx + 2\int_{1}^{+\infty} \frac{dx}{x^2} < +\infty,$$

but $\varphi \notin L^2(\mathbf{R})$, so φ is not in $D(M_x)$ on $L^2(\mathbf{R})$.

We have the following simple but important fact:

Lemma 4.7. *Any multiplication operator $(D(M_g), M_g)$ acting on $L^2(X, \mu)$, where $\mu(X) < +\infty$, is densely defined.*

Proof. Let $\varphi \in L^2(X, \mu)$. To approximate it using elements of $D(M_g)$, define

$$\chi_R(x) = \begin{cases} 1 & \text{if } |g(x)| \leqslant R \\ 0 & \text{if } |g(x)| > R, \end{cases}$$

for $R > 0$. Note that $\varphi\chi_R \to \varphi$ in $L^2(X, \mu)$ since

$$\|\varphi\chi_R - \varphi\|^2 = \int_{|g(x)|>R} |\varphi(x)|^2 d\mu(x) \to 0,$$

since $\varphi \in L^2(X, \mu)$ and the sets $|g(x)| > R$ "decrease" to $\emptyset$ (note that here we use the fact that $\mu(X) < +\infty$). Now, since

$$\int_X |g|^2 |\varphi\chi_R|^2 d\mu \leqslant R^2 \|\varphi\|^2 < +\infty,$$

we have $\varphi\chi_R \in D(M_g)$ for all R, and we have the desired approximations. □

A particular example of multiplication operator is $(D(M_x), M_x)$ on $L^2(\mathbf{R})$; this operator is called the *position operator* because of its interpretation in Quantum Mechanics, as we will see in Chapter 6.

(2) Here is another important example, which is a prototype for many other similar operators: let $H = L^2(\mathbf{R})$ again and consider

$$D(\partial_x) = C_0^1(\mathbf{R}),$$

the set of continuously differentiable functions with compact support. Note that $D(\partial_x)$ is dense in H (e.g., because the smaller space of smooth functions with compact support is dense in $L^2(\mathbf{R})$, see Lemma 1.14). Then define

$$\partial_x : \begin{cases} D(\partial_x) \to H \\ f \mapsto f'. \end{cases}$$

Since f' is continuous with compact support, $\partial_x(f)$ does indeed belong to H for $f \in D(\partial_x)$. Thus $(C_0^1(\mathbf{R}), \partial_x)$ is in $DD(L^2(\mathbf{R}))$.

Because of the need to take the domain into account, many constructions which are straightforward for bounded operators require some care if one wants to extend them to densely-defined operators.

One which is easy is "transport through isomorphisms": let H, H' be Hilbert spaces

$$A : H \to H'$$

an isomorphism (not necessarily isometric), and $(D(T), T)$ in $DD(H)$. Then

$$(A(D(T)), w \mapsto A(T(A^{-1}(w))))$$

is in $DD(H')$, and is commonly denoted ATA^{-1}. In particular, if A is a unitary isomorphism, we say at T and ATA^{-1} are *unitarily equivalent*.

Another important remark is that if $(D(T), T) \in DD(H)$ and $S \in L(H)$, there is an obvious definition of the sum $T + S$ as $(D(T), T + S)$. This will be used to define the spectrum of an unbounded operator.

EXAMPLE 4.8. Let $H = L^2(\mathbf{R})$ and $U : H \to H$ the Fourier transform (see Example 1.5): for f integrable on $\mathbf{R}$, we have

$$Uf(x) = \int_{\mathbf{R}} f(t) e^{-2i\pi xt} dt.$$

If f is in the domain $D(\partial_x) = C_0^1(\mathbf{R})$ of the operator of derivation defined in the previous example, we claim that $U(f) \in D(M_{2i\pi x}) = D(M_x)$, the domain of the operator of multiplication by x. Indeed, writing

$$Uf(x) = \int_{\mathbf{R}} f(t) e^{-2i\pi xt} dt,$$

we get

$$2i\pi x Uf(x) = \int_{\mathbf{R}} (2i\pi x) f(t) e^{-2i\pi xt} dt = \int_{\mathbf{R}} f'(t) e^{-2i\pi xt} dt \tag{4.1}$$

by integration by parts, which is justified given that f is C^1 with compact support. Thus $M_x(U_f)$ is the Fourier transform of the function f', which is continuous of compact support, hence in L^2, proving our claim.

We see, furthermore, that for $f \in D(\partial_x)$, we have

$$M_{2i\pi x}(Uf) = U(\partial_x(f)),$$

but since $U(D(\partial_x))$ is not the whole of $D(M_x)$ (e.g., because the function $g : x \mapsto e^{-\pi x^2}$ is in $D(M_x)$, but $Ug = g \notin C_c^1(\mathbf{R})$) , we can *not* write

$$U^{-1} M_{2i\pi x} U = \partial_x,$$

but only

$$\partial_x \subset U^{-1} M_{2i\pi x} U.$$

Note that, in this case, the domain of the multiplication operator seems much more "natural" than the one for the derivation (since there are many other differentiable L^2 functions than those in $C_c^1(\mathbf{R})$); the Fourier transport allows us to define what is, in a sense, the best domain as $U^{-1}(D(M_x))$. This space, and other similar ones, can be identified with *Sobolev spaces*.

Other operations on densely defined operators may be much more tricky. For instance, note that it is perfectly possible to have dense subspaces H_1, $H_2 \subset H$, such that $H_1 \cap H_2 = 0$. If this is the case, and T_1, T_2 are defined on H_1, H_2, respectively, then it will clearly be essentially impossible to define, for instance, the sum $T_1 + T_2$ in any sensible way. So $DD(H)$ is *not* a vector subspace in a reasonable sense.

4.2. The graph, closed and closable operators

Because of the frequent ambiguity in the choice of a domain for an unbounded operator, it is of importance to isolate a class of such operators where, in a natural way, there is a somewhat natural domain defined for an extension of T which is canonical in some way. This is naturally some type of closure operation.

Definition 4.9. Let H be a Hilbert space and $(D(T), T) \in DD(H)$. The *graph* $\Gamma(T)$ of $(D(T), T)$ is the linear subspace

$$\Gamma(T) = \{(v, w) \in H \times H \mid v \in D(T) \text{ and } w = T(v)\}$$

of $H \times H$.

The operator $(D(T), T) \in DD(H)$ is said to be *closed* if $\Gamma(T)$ is closed in $H \times H$ when the latter is seen as a Hilbert space with the inner product

$$\langle (v_1, w_1), (v_2, w_2) \rangle_{H \times H} = \langle v_1, v_2 \rangle_H + \langle w_1, w_2 \rangle_H.$$

The operator is said to be *closable* if there exists a closed extension of T.

Note in passing that $T \subset T_1$, for densely defined operators, can also be rephrased equivalently as $\Gamma(T) \subset \Gamma(T_1)$, where the inclusion is now the standard inclusion of subspaces.

Remark 4.10. At this point it is useful to recall the Closed Graph Theorem (Theorem 1.16): this shows again the link between T being unbounded and $D(T)$ being a proper subset of H, since it shows that any closed operator with $D(T) = H$ is in fact continuous.

Example 4.11. Let H be a Hilbert space and let $S \in L(H)$ be a bounded operator such that that 0 lies in the continuous spectrum of S, i.e., S is injective and $\mathrm{Im}(S)$ is dense but distinct from H. Then, as already mentioned, we can construct the densely defined operator inverse to S, namely

$$(D(T), T) = (\mathrm{Im}(S), S^{-1}).$$

Then this operator is closed: indeed, its graph $\Gamma(T)$ is simply given by $\Gamma(T) = \{(v, w) \in H \times H \mid (w, v) \in \Gamma(S)\}$. The graph of S is closed since S is continuous, and obviously switching the coordinates is also a continuous operator, so $\Gamma(T)$ is closed.

Allowing some wiggle room, this construction is very general: if $(D(T), T)$ is a closed unbounded operators, one can find some $\lambda \in \mathbf{C}$ such that $T - \lambda$ is the inverse of some bounded operator, provided the resolvent set of T is not empty (see Section 4.5 below, but note already that it is possible for the resolvent set to be empty, as shown in Example 4.37.

This example shows that a good way to prove spectral properties of unbounded operators may be to reduce to bounded operators by this trick; this is indeed what will be done in the proof of the spectral theorem for unbounded self-adjoint operators, where it will be shown that the spectrum (unsurprisingly) is real, so any non-real complex number lies in the resolvent set and can be used in the previous construction.

Remark 4.12. We see, concretely, that T is closed if and only if, for any sequence (v_n) of vectors in $D(T)$, if there exist vectors v_0, $w_0 \in H$ such that

$$v_n \to v_0, \qquad T(v_n) \to w_0,$$

then we have $(v_0, w_0) \in \Gamma(T)$, i.e., $v_0 \in D(T)$ and $w_0 = T(v_0)$. Note that since $D(T)$ is *not* closed (except in the case $D(T) = H$, which corresponds to T bounded), the fact that $v_0 \in D(T)$ is itself non-trivial, and depends on the property that $T(v_n)$ converges.

If T is closable, we also see that $D(\bar{T})$ can be described concretely as the space of vectors $v \in H$ which are limits of a sequence of vectors $v_n \in D(T)$, *with the condition* that $T(v_n)$ also converges (its limit being then $T(v)$).

The following lemma is simple but useful: it justifies that closable operators have a natural extension, with a domain that may be hoped to be possibly simpler to handle.

Lemma 4.13. *Let H be a Hilbert space, $(D(T), T) \in DD(H)$. Then T is closable if and only if $\overline{\Gamma(T)} \subset H \times H$ is the graph of some operator $(D(\bar{T}), \bar{T}) \in DD(H)$. This operator is called the* closure *of T, and we have $\bar{T} \subset S$ for any closed extension S of T, i.e., it is the smallest closed extension of T.*

Proof. If $S \supset T$ is closed then we have

$$\overline{\Gamma(S)} = \Gamma(S) \supset \overline{\Gamma(T)},$$

from which it follows immediately that $\overline{\Gamma(T)}$ is a graph (because it can not contain (v, w) and (v, w') with $w \neq w'$), and then that it is indeed the graph of a closed extension of T, which is also consequently the smallest possible. □

Example 4.14. The operator M_x of multiplication by x on $L^2(\mathbf{R})$ is closed, when $D(M_x)$ is as defined previously. Indeed, consider a sequence of functions (φ_n) in $D(M_x)$, such that

$$\varphi_n \to \varphi_0, \qquad x\varphi_n \to \psi_0,$$

where $\varphi_0,\ \psi_0 \in L^2(\mathbf{R})$. We need to show that $x\varphi_0 = \psi_0$, and the difficulty is (of course) that multiplying by x is not continuous in the L^2-norm (which is the whole point...) To do this, recall that there exists a subsequence (φ_{n_k}) such that the convergence $\varphi_{n_k}(x) \to \varphi_0(x)$ holds almost everywhere on $\mathbf{R}$. Then clearly $x\varphi_{n_k}(x)$ converges almost everywhere to $x\varphi_0(x)$, on the one hand, and on the other hand this must be $\psi_0(x)$ by unicity of the limit in L^2. Hence, almost everywhere, we do have $x\varphi_0(x) = \psi_0(x)$, from which it follows that $\varphi_0 \in D(M_x)$ and $M_x(\varphi_0) = \psi_0$.

Note that if we had considered the smaller operator $(C_c(\mathbf{R}), S : f \mapsto xf)$, where $C_c(\mathbf{R})$ is the space of compactly supported continuous functions, we would get a non-closed operator with $S \subset M_x$, and $\bar{S} = M_x$. (This is left as an easy exercise).

Example 4.15. The following illustrates a simple example of a non-closable operator: $\Gamma(T)$ is not a graph. It will be seen that the construction is somewhat artificial, and indeed most operators we will encounter further on, especially for applications, are closable.

Let $H = L^2([0, 1])$, and let $e_n(x) = e^{2i\pi nx}$, for $n \in \mathbf{Z}$, denote the classical orthonormal basis, used in the theory of Fourier series. Fix any function $\varphi_0 \in H$ such that $\langle \varphi_0, e_n \rangle \neq 0$ for infinitely many n (for instance, the function $\varphi(x) = x$ will do).

Now let

$$D(T) = \{f = \alpha\varphi_0 + P \mid \alpha \in \mathbf{C}, \quad P = \sum_{|n| \leqslant N} a_n e_n\},$$

the sum of the line spanned by φ_0 and the space of trigonometric polynomials (finite combinations of the basis vectors). Note that $D(T)$ is dense in H, because the polynomials already are, and that the decomposition $f = \alpha\varphi_0 + P$ is unique (i.e., the coefficient α is uniquely determined by f; indeed, if $|n| > N$, we have

$$\langle f, e_n \rangle = \alpha \langle \varphi_0, e_n \rangle$$

and if n is such that $\langle \varphi_0, e_n \rangle \neq 0$, we can determine α from this identity).

Now let

$$T(f) = \alpha\varphi_0 \text{ for } f = \alpha\varphi_0 + P \in D(T).$$

By the previous remark, T is well-defined on $D(T)$, and it is obviously linear, so $(D(T), T) \in DD(H)$. We claim that this operator is not closable.

Indeed, notice first that $\varphi_0 \in D(T)$ with $T(\varphi_0) = \varphi_0$ gives $(\varphi_0, \varphi_0) \in \Gamma(T) \subset \overline{\Gamma(T)}$. Next, let

$$\varphi_N = \sum_{|n| \leqslant N} \langle \varphi_0, e_n \rangle e_n$$

which is a trigonometric polynomial, hence lies in $D(T)$, with $T(\varphi_N) = 0$. Since $\varphi_N \to \varphi_0$ in $L^2([0,1])$, we have

$$(\varphi_0, 0) = \lim_{N \to +\infty} (\varphi_N, T(\varphi_N)) \in \overline{\Gamma(T)}.$$

Thus (φ_0, φ_0) and $(\varphi_0, 0)$ both lie in $\overline{\Gamma(T)}$, and this shows it is not a graph.

The following lemma is quite simple but will be important in defining the spectrum of a closed operator.

LEMMA 4.16. *Let H be a Hilbert space, $(D(T), T) \in DD(H)$ a closed operator. Then, for any bounded operator $S \in L(H)$, the operator $(D(T), S + T)$ is closed, and*

$$S + T \,:\, D(T) \to H$$

has a bounded inverse $(S+T)^{-1} \in L(H)$ if and only if it is a bijection.

PROOF. Consider any sequence $(v_n, w_n) \in \Gamma(S + T)$ with $v_n \to v_0$, $w_n \to w_0$ as $n \to +\infty$. We have then $w_n = Sv_n + Tv_n$ for all n, and since S is continuous and $v_n \to v_0$, we get that Sv_n converges to Sv_0. Now consider the sequence $(v_n, w_n - Sv_n)$. It lies in $\Gamma(T)$, and we have $v_n \to v_0$ and $w_n - Sv_n \to w_0 - Sv_0$. Since T is closed, it follows that $(v_0, w_0 - Sv_0) \in \Gamma(T)$, so that $v_0 \in D(T)$ and $Tv_0 = w_0 - Sv_0$, which means that $(v_0, w_0) \in \Gamma(S + T)$, showing that $S + T$ is closed.

Now if we assume that $S + T$ is a (linear, of course) bijection from $D(T)$ to H, the graph of its inverse $(S + T)^{-1}$ is

$$\Gamma((S+T)^{-1}) = \{(v, w) \in H \times D(H) \mid v = (S + T)w\},$$

so it is simply the graph of $(D(T), S + T)$ "with coordinates switched":

$$\Gamma((S+T)^{-1}) = \{(v, w) \in H \times H \mid (w, v) \in \Gamma(S + T)\}.$$

This implies of course that $(S + T)^{-1}$ has a closed graph, and hence it is continuous by the Closed Graph Theorem. □

4.3. The adjoint

We now come to the definition of the adjoint for unbounded operators. Of course, this can only be expected to exist (at best) itself as an unbounded operator, so we must first consider the question of the domain. We want (of course) to have

$$\langle Tv, w \rangle = \langle v, T^* w \rangle \tag{4.2}$$

whenever this makes sense, and this means in particular that

$$|\langle Tv, w \rangle| \leqslant \|T^* w\| \|v\|,$$

which implies that $v \mapsto \langle Tv, w \rangle$ would in fact be *continuous* on $D(T)$, and therefore would extend by continuity to H (in a unique way, since $D(T)$ is dense). This is something that may, or may not, be true, and it clearly defines a natural domain.

Definition 4.17. Let H be a Hilbert space and $(D(T), T) \in DD(H)$. The *domain of the adjoint* $D(T^*)$ is defined to be the set of $w \in H$ such that the linear map

$$\lambda_w^* \begin{cases} D(T) \to \mathbf{C} \\ v \mapsto \langle Tv, w \rangle \end{cases}$$

is continuous, i.e., those w such that equivalently, λ_w^* extends uniquely to a linear functional $\lambda_w^* \in H'$, or there exists a constant $C \geqslant 0$ with

$$|\langle Tv, w\rangle| \leqslant C\|v\|, \qquad \text{for } v \in D(T).$$

The *adjoint* is the linear map

$$\begin{cases} D(T^*) \to H \\ w \mapsto \text{the unique vector } T^*w \text{ such that } \lambda_w^*(v) = \langle v, T^*w\rangle, \end{cases}$$

where the existence of the vector is given by the Riesz Representation Theorem for Hilbert spaces.

From the definition, we see that the relation (4.2) holds for $v \in D(T)$, $w \in D(T^*)$, as desired. Intuitively, the domain $D(T^*)$ is thus (spanned by) the set of "coordinates" along which the operator T is in fact continuous. From this description, it is not clear whether $D(T^*)$ is dense in H or not. If it is, we obtain a densely defined operator $(D(T^*), T^*) \in DD(H)$, which is of course called the adjoint of $(D(T), T)$. But in fact, it may be the case that $D(T^*)$ is not dense.

Example 4.18. Let $H = L^2(\mathbf{R})$, and consider the densely-defined operator

$$(D(T), T) = \Big(L^1(\mathbf{R}) \cap L^2(\mathbf{R}), \varphi \mapsto \Big(\int_{\mathbf{R}} \varphi(t)dt\Big)\varphi_0\Big),$$

for φ_0 any element of H, for instance $\varphi_0(x) = e^{-\pi x^2}$. Given $\psi \in H$, we have $\psi \in D(T^*)$ if the linear map

$$\lambda_\psi^* : \varphi \mapsto \int_{\mathbf{R}} T(\varphi)\bar{\psi}dt$$

is continuous on $D(T)$. But

$$\lambda_\psi^*(\varphi) = \Big(\int_{\mathbf{R}} \varphi(t)dt\Big)\langle \varphi_0, \psi\rangle,$$

and we see that this is continuous (in L^2 norm) if and only if the inner product $\langle \varphi_0, \psi\rangle$ vanishes, so that $D(T^*) = \varphi_0^\perp$ is not dense in H.

It turns out, however, that the condition that $D(T^*)$ be dense is often satisfied. We will denote by $DD^*(H)$ the set of densely-defined operators on H for which $D(T^*)$ is dense. As it turns out, this class has already been introduced implicitly:

Lemma 4.19. *Let $(D(T), T)$ be an unbounded densely-defined operator on a Hilbert space H. Then $T \in DD^*(H)$ if and only if T is closable.*

This is Part (2) of the following proposition, which lists other easy properties of adjoints and their domains.

Proposition 4.20. *Let H be a Hilbert space and let $(D(T), T) \in DD(H)$. Denote by $(D(T^*), T^*)$ the adjoint of T, which is defined on the not-necessarily dense subset $D(T^*)$.*

(1) *The graph $\Gamma(T^*)$ is closed in $H \times H$. In particular, if T^* is densely defined, it is closed.*

(2) *The subspace $D(T^*)$ is dense in H if and only if T is closable.*

(3) *If T is closable, then $T^{**} = (T^*)^*$ is in $DD(H)$ and $T^{**} = \bar{T}$ is the closure of T, while $\bar{T}^* = T^*$.*

(4) *If T_1 and T_2 are densely defined operators with $T_1 \subset T_2$, we have $T_2^* \subset T_1^*$.*

Proof. (1) By definition $\Gamma(T^*)$ is the set of $(v, w) \in H \times H$ such that $v \in D(T^*)$ and $w = T^*v$. These two conditions mean exactly that, for all $x \in D(T)$, we have

$$\langle Tx, v\rangle = \langle x, w\rangle. \tag{4.3}$$

Indeed, this is the defining relation (4.2) if $(v, w) \in \Gamma(T^*)$, but also conversely: if (4.3) holds for all $x \in D(T)$, the Cauchy-Schwarz inequality implies that

$$|\langle Tx, v\rangle| \leqslant \|w\|\|x\|$$

so that $v \in D(T^*)$, and from (4.3), $w = T^*v$. Since the equations (4.3), parametrized by x, are linear and continuous, the set of solutions $\Gamma(T^*)$ is closed. In fact, one can write $\Gamma(T^*) = W^\perp$, where

$$W = \{(Tx, -x) \in H \times H \mid x \in D(T)\} \tag{4.4}$$

is a subspace of $H \times H$.

(2) Assume first that $D(T^*)$ is dense in H. Then $T^{**} = (T^*)^*$ is defined, and we claim that $T \subset T^{**}$. By (1), this implies that T^{**} is a closed extension of T, which is therefore closable. To see this, observe that if $v \in D(T)$, the map

$$w \mapsto \langle T^*w, v\rangle$$

is the same as $w \mapsto \langle w, Tv\rangle$, by definition (4.2). This map is continuous on $D(T^*)$ with norm $\leqslant \|Tv\|$, and this means exactly that $v \in D(T^{**})$ with $T^{**}(v) = Tv$. Since v is arbitrary in $D(T)$, this shows indeed that $T \subset T^{**}$.

Conversely, assume that T is closable. Let $v_0 \in D(T^*)^\perp$ be a vector perpendicular to the domain of T^*; we must show $v_0 = 0$ to prove that $D(T^*)$ is dense. This is a bit tricky: we note first that $(v_0, 0) \in \Gamma(T^*)^\perp$ in $H \times H$, and then notice that, from the computation in (1), we have

$$\Gamma(T^*)^\perp = (W^\perp)^\perp = \overline{W},$$

with W given by (4.4).

From this, it is clear that

$$\overline{W} = \{(w, -v) \mid (v, w) \in \overline{\Gamma(T)}\}, \tag{4.5}$$

hence we find that $(0, v_0) \in \overline{\Gamma(T)}$. Since $\overline{\Gamma(T)}$ is a graph, we must have $v_0 = \bar{T}(0) = 0$, which concludes the proof.

(3) If T is closable, T^* is densely defined and at the beginning of (2), we checked that T^{**} is densely defined and is a closed extension of T, so $\bar{T} \subset T^{**}$. In fact, there is equality, because we can use the computation in (1), applied to T^* instead of T, to determine the graph of T^{**}: $\Gamma(T^{**}) = V^\perp$ where

$$V = \{(T^*x, -x) \in H \times H \mid x \in D(T^*)\} \subset H \times H.$$

This V is obtained from the graph of T^* as $A(\Gamma(T^*))$, where A is the linear isometry $(v, w) \mapsto (w, -v)$. In (2) we saw that $\Gamma(T^*)^\perp = \overline{W}$, and we see by (4.5) that $A(\overline{W}) = \overline{\Gamma(T)}$, so this gives

$$\Gamma(T^{**}) = (A(\Gamma(T^*)))^\perp = A(\Gamma(T^*)^\perp) = A(\overline{W}) = \overline{\Gamma(T)}.$$

Finally, for T closable, we have

$$\bar{T}^* = (T^{**})^* = (T^*)^{**} = \overline{T^*} = T^*$$

since T^* is closed.

(4) This is clear from the defining relation (4.2) and the assumption $D(T_1) \subset D(T_2)$, with T_2 acting like T_1 on $D(T_1)$. ☐

We can now make the following important definition:

DEFINITION 4.21. Let H be a Hilbert space and $(D(T), T) \in DD(H)$ a closable operator.

(1) The operator $(D(T), T)$ is *symmetric* or *Hermitian* if it is closable and $T \subset T^*$, i.e., equivalently, if

$$\langle Tv, w\rangle = \langle v, Tw\rangle$$

for all v, $w \in D(T)$.[2]

(2) The operator $(D(T), T)$ is *self-adjoint* if it is closable and $T = T^*$, i.e., if it is symmetric and *in addition* $D(T^*) = D(T)$.

(3) A *self-adjoint extension* of $(D(T), T)$ is a self-adjoint operator T_1 such that $T \subset T_1$.

(4) An *essentially self-adjoint* operator $(D(T), T)$ is a symmetric operator such that $\bar{T}$ is self-adjoint.

REMARK 4.22. (1) The point of the last part of the definition is that, in general, there may exist more than one self-adjoint extension of a symmetric operator (or even none at all) – concrete examples will arise in Example 5.8 of Chapter 5. If $(D(T), T)$ is essentially self-adjoint, however, $\bar{T}$ is the unique self-adjoint extension of T: indeed, for any such S, we have $\bar{T} \subset S$ since S is a closed extension of T, and conversely, using (4) of Proposition 4.20, since $T^{**} \subset S$, we have $S^* \subset (T^{**})^*$ so $S \subset \bar{T}^* = \bar{T}$.

(2) If T is symmetric, we have

$$T \subset T^{**} \subset T^*$$

since T^* is then a closed extension of T. It may happen that T is closed and symmetric, but not self-adjoint, in which case $T = T^{**} \subset T^*$. This means in particular that if T is symmetric and closed, then T^* is self-adjoint if and only if T^* is symmetric.

REMARK 4.23. Issues of domains concerning symmetric operators are emphasized by the Hellinger-Toeplitz Theorem: if $(D(T), T)$ is a symmetric operator with $D(T) = H$, then T is automatically bounded.

EXAMPLE 4.24. (1) Let $H = L^2(\mathbf{R})$ and $(D(T), T) = (D(M_x), M_x)$ the operator of multiplication by x. Then we claim that T is self-adjoint. Indeed, we first determine $D(T^*)$ as follows: we have

$$\lambda^*_\psi(\varphi) = \langle T(\varphi), \psi\rangle = \int_{\mathbf{R}} x\varphi(x)\overline{\psi(x)}dx$$

for $\varphi \in D(M_x)$, $\psi \in H$. It is clear that if $\psi \in D(M_x)$, the integral can be expressed as

$$\langle \varphi, T(\psi)\rangle,$$

so it is obvious at least that T is symmetric.

Now suppose $\psi \in D(T^*)$, so there exists C such that

$$|\langle T(\varphi), \psi\rangle|^2 \leqslant C\|\varphi\|^2$$

[2] Indeed, if this identity holds, it follows that $D(T) \subset D(T^*)$, so $T \in DD^*(H)$ is closable, and T^* coincides with T on $D(T)$.

for all $\varphi \in D(M_x)$. We take φ to be $x\psi\chi_R$, where χ_R is the characteristic function of $[-R, R]$. This is in $L^2(\mathbf{R})$ since $|x\psi(x)i\chi_R(x)|^2 \leqslant R^2|\psi(x)|^2$, and the inequality becomes

$$\Bigl(\int_{\mathbf{R}} x^2|\psi(x)|^2\chi_R(x)dx\Bigr)^2 \leqslant C\int_{\mathbf{R}} x^2|\psi(x)|^2\chi_R(x)dx,$$

so

$$\int_{\mathbf{R}} x^2|\psi(x)|^2\chi_R(x)dx \leqslant C$$

for all $R \geqslant 0$. As $R \to +\infty$, the monotone convergence theorem implies that these integrals converge to

$$\int_{\mathbf{R}} x^2|\psi(x)|^2dx,$$

so we obtain $x\psi \in L^2(\mathbf{R})$, i.e., $\psi \in D(M_x)$. Hence, we have $D(T^*) = D(T)$, which finishes the proof that T is self-adjoint.

In fact, any multiplication operator $(D(M_g), M_g)$ on a finite measure space (X, μ) is self-adjoint if g is real-valued; this may be checked by a similar computation, or using the self-adjointness criterion discussed below, as we will see in Example 4.31.

(2) Consider the operator $(D(T), T) = (C_c^1(\mathbf{R}), i\partial_x)$ of (i times) differentiation on $L^2(\mathbf{R})$. We claim this operator is symmetric, but not self-adjoint. The symmetry, as before, reduces to a formal computation: let φ, $\psi \in C_c^1(\mathbf{R})$, then by integration by parts we get

$$\begin{aligned}\langle i\partial_x\varphi, \psi\rangle &= i\int_{\mathbf{R}} \varphi'(x)\overline{\psi}(x)dx\\ &= \Bigl[i\varphi\psi\Bigr]_{-\infty}^{+\infty} + \int_{\mathbf{R}} \varphi(x)\overline{i\psi(x)}dx = \langle \varphi, i\partial_x\psi\rangle,\end{aligned}$$

as desired. However, we can see for instance that $\varphi_0(x) = e^{-\pi x^2}$ is in $D(T^*) = D((i\partial_x)^*)$: we have

$$\lambda^*_{\varphi_0}(\varphi) = \int_{\mathbf{R}} i\varphi'(x)e^{-\pi x^2}dx = \int_{\mathbf{R}} \varphi(x)(-2\pi ix)e^{-\pi x^2}dx$$

by integration by parts (again), and this is continuous on $C_c^1(\mathbf{R})$ since $x \mapsto xe^{-\pi x^2}$ is still in $L^2(\mathbf{R})$.

It is important to notice, however, that the same computation shows that $T^*\varphi_0 = i\varphi_0'$, i.e., the adjoint of this differentiation operator acts on φ_0 "as the same differential operator".

Remark 4.25. We do not define normal operators here, because this is a tricky issue. The difficulty is that if T is closable, so T^* is densely defined, it is not clear what the intersection $D(T) \cap D(T^*)$ is, and this is the only obvious space on which $T(T^*v)$ or $T^*(Tv)$ make sense in order to be compared...

However, note that one of the motivation to study normal operators was to have a spectral theorem for unitary operators, and unitary operators are always bounded, so we do not need to enlarge our setting to accommodate them with unbounded operators.

The following lemma is again very simple, but worth pointing out as it is used constantly.

Lemma 4.26. *Let H be a Hilbert space and let $(D(T), T) \in DD^*(H)$ be a closable operator. For any $S \in L(H)$, $(D(T), S+T)$ is closable and its adjoint is $(D(T^*), S^*+T^*)$.*

Proof. For $w \in H$, the linear functional we need to consider is

$$\langle Tv, w\rangle + \langle Sv, w\rangle$$

and since the second one is continuous on H for all w, it follows that $D(T^*) \subset D((S+T)^*)$, and moreover it is clear that $(S+T)^*w = S^*w + T^*w$ (by unicity in the Riesz Representation Theorem). □

4.4. Criterion for self-adjointness and for essential self-adjointness

The self-adjointness of an unbounded operator is very sensitive to the choice of the domain, and it may well be difficult to determine which is the right one (as shown by the case of differential operators). This sensitivity, as we will see, persists when spectral theory is concerned: for instance, a symmetric operator which is not self-adjoint operator *will* have a spectrum which is not a subset of $\mathbf{R}$. So it is very important to have a convenient way to decide if a symmetric operator is self-adjoint (or, more ambitiously, if it has self-adjoint extensions, in which case a classification of these is also useful).

It turns out there is a very simple such criterion, which one can understand spectrall (although it doesn't refer explicitly to the spectrum in the formulation we give here):

Proposition 4.27 (Self-adjointness criterion). *Let H be a Hilbert space and let $(D(T), T) \in DD^*(H)$ be a densely defined operator. Assume T is symmetric. Then the following properties are equivalent:*

(1) *The operator T is self-adjoint.*

(2) *The operator is closed and* $\mathrm{Ker}(T^* + i) = \mathrm{Ker}(T^* - i) = 0$*, where the kernels are of course subspaces of $D(T^*)$.*

(3) *We have* $\mathrm{Im}(T + i) = \mathrm{Im}(T - i) = H$*, where the image refers of course to the image of the subspace $D(T)$.*

The point is that (2) is more or less obviously necessary since we expect a self-adjoint operator to not have any non-real eigenvalue; the fact that it is also sufficient may be surprising, however.

Either (2) or (3) are used in practice. For instance, (2) is sometimes very convenient to show that an operator is *not* self-adjoint: it is enough to exhibit an element $v \in H$ which is in $D(T^*)$, and for which $T^*v = iv$ or $-iv$ (see for instance Example 5.8 in Chapter 5). On the other hand, if we can solve the equations $Tv \pm iv = w$ for arbitrary $w \in H$, with a solution in $D(T)$, then T is self-adjoint (assuming it is known to be symmetric); this can be useful because this condition does not refer explicitly to the nature of T^* at all. For an example, see the proof of Lemma 5.4.

The proof will use the following facts, which are useful enough to be stated separately:

Lemma 4.28. *Let H be a Hilbert space and $(D(T), T) \in DD^*(H)$ a closed symmetric operator. Then:*

(1) *The subspaces* $\mathrm{Im}(T + i)$ *and* $\mathrm{Im}(T - i)$ *are closed in H;*

(2) *We have*

$$\mathrm{Ker}(T^* + i) = \mathrm{Im}(T - i)^{\perp}, \qquad \mathrm{Ker}(T^* - i) = \mathrm{Im}(T + i)^{\perp}.$$

Proof. (1) We prove that $\mathrm{Im}(T - i)$ is closed, the other case being of course similar. For $v \in D(T)$, we note first that

$$\begin{aligned} \|(T-i)v\|^2 &= \langle Tv - iv, Tv - iv\rangle \\ &= \|Tv\|^2 + \|v\|^2 - i\langle v, Tv\rangle + i\langle Tv, v\rangle = \|Tv\|^2 + \|v\|^2 \end{aligned} \tag{4.6}$$

by symmetry of T. Now if $w_n = (T-i)v_n$ is a sequence in $\mathrm{Im}(T-i)$ which converges to $w_0 \in H$, this formula applied to $v_n - v_m$ shows that the sequences (v_n) and (Tv_n) are both Cauchy sequences in H, hence (v_n, Tv_n) converges to some (v_0, v_0'). Because T is supposed to be closed, we have in fact $v_0 \in D(T)$ and $v_0' = Tv_0$, and hence

$$w_n = Tv_n - iv_n \to v_0' - iv_0 = (T-i)v_0,$$

showing that the limit w_0 belongs to $\mathrm{Im}(T-i)$, as desired.

(2) We have $(T-i)^* = T^* + i$, defined on $D(T^*)$ (by Lemma 4.26). The basic relation

$$\langle (T-i)v, w\rangle = \langle v, (T^*+i)w\rangle, \qquad v \in D(T),\ w \in D(T^*), \tag{4.7}$$

gives $\mathrm{Ker}(T^*+i) \subset \mathrm{Im}(T-i)^\perp$ immediately. The converse inequality holds since $D(T)$ is dense, so that (for a given $w \in D(T^*)$) the vanishing of the left-hand side of (4.7) for all $v \in D(T)$ implies that $(T^*+i)w = 0$. □

Proof of Proposition 4.27. We first prove that (1) implies (2) (this is similar to Proposition 4.38 below). First, if T is self-adjoint, it is closed, and moreover if $v \in D(T^*) = D(T)$ satisfies $T^*(v) = iv$, we get $Tv = iv$ by self-adjointness, and

$$i\|v\|^2 = \langle Tv, v\rangle = \langle v, Tv\rangle = -i\|v\|^2$$

so that $v = 0$. The same is true if $T^*v = -iv$, by similar computations.

Next assume that (2) is true. Then we know first by the previous lemma that $\mathrm{Im}(T-i)$ is closed. The second part also shows that

$$0 = \mathrm{Ker}(T^*+i) = \mathrm{Im}(T-i)^\perp,$$

which means $\mathrm{Im}(T-i) = H$. Again, the surjectivity of $T+i$ is proved similarly.

Thus there remains to prove that (3) implies (1), which is the crucial part. By the symmetry of T, we know $T \subset T^*$, so we must show that $D(T^*) = D(T)$. Let $w \in D(T^*)$ be given. The assumption (3) allows us to write

$$(T+i)w = (T+i)v$$

for some $v \in D(T)$. Since $T \subset T^*$, this gives $(T^*+i)(w-v) = 0$. By (4.7) again, we find that $v - w \in \mathrm{Im}(T-i)^\perp = 0$, so $w = v \in D(T)$, which concludes the proof. □

Remark 4.29. The proof clearly shows that there is nothing particularly special in using $T \pm i$: there similar statements with $\pm i$ replaced by λ, $\bar{\lambda}$, for any fixed $\lambda \in \mathbf{C}$ which is not real, are also valid.

Similarly, we obtain:

Proposition 4.30. *Let H be a Hilbert space and $(D(T), T) \in DD^*(H)$ a symmetric operator. Then the following properties are equivalent:*

(1) *The operator T is essentially self-adjoint.*

(2) *We have $\mathrm{Ker}(T^*+i) = \mathrm{Ker}(T^*-i) = 0$, where the kernels are of course subspaces of $D(T^*)$.*

(3) *The subspaces $\mathrm{Im}(T+i)$ and $\mathrm{Im}(T-i)$ are dense in H.*

Proof. (1) implies (2): if $\bar{T}$ is self-adjoint, we have $\bar{T}^* = T^* = \bar{T}$, so the previous proposition gives $\mathrm{Ker}(T^* \pm i) = \mathrm{Ker}(\bar{T}^* \pm i) = 0$.

(2) implies (3): apply again the second part of Lemma 4.28.

(3) implies (1): the assumption shows that $\mathrm{Im}(\bar{T} \pm i)$ is dense in H, and Lemma 4.28, (1), implies that $\mathrm{Im}(\bar{T} \pm i)$ is closed, so we can apply the corresponding result of Proposition 4.27 for $\bar{T}$. □

Example 4.31. Here is the probably simplest application of this criterion:

Lemma 4.32. *Let (X, μ) be a finite measure space, and $(D(M_g), M_g)$ be a multiplication operator on $L^2(X, \mu)$, for some* real-valued *measurable function $g : X \to \mathbf{R}$.*

Proof. It is obvious, for g real-valued, that M_g is symmetric. Now we will use Part (3) of Proposition 4.27 to show it is self-adjoint. Consider the operators $M_g \pm i$, which are just M_h for $h = g \pm i$. If $\varphi \in L^2(X, \mu)$ is arbitrary, we have

$$\varphi = h\varphi_1,$$

with $\varphi_1(x) = \varphi(x)/h(x)$: this function is well-defined since h has no zero (because g is real-valued). We now claim that $\varphi_1 \in D(M_h) = D(M_g)$, which then gives $\varphi = M_h\varphi_1 \in \mathrm{Im}(M_h)$, showing the surjectivity required.

This claim is very easy to check: since

$$\Big|\frac{1}{h(x)}\Big| = \frac{1}{1 + |g(x)|^2} \leqslant 1,$$

we have $\varphi_1 \in L^2(X, \mu)$; in addition, we have $h\varphi_1 = \varphi \in L^2(X, \mu)$ so $\varphi_1 \in D(M_h) = D(M_g)$. □

Remark 4.33. As already mentioned, there are more general results which classify (in some way) all self-adjoint extensions of a given symmetric operator $(D(T), T) \in DD^*(H)$. The result is the following: let

$$H_+ = \mathrm{Ker}(T^* - i), \qquad H_- = \mathrm{Ker}(T^* + i)$$

be the so-called *deficiency subspaces* of T, which already occur in Proposition 4.27. Then one shows first that T has self-adjoint extensions if and only if $\dim H_+ = \dim H_-$ (including the possibility that both be infinite, in which case they must be equal in the sense of infinite cardinals, i.e., for instance, countable dimension is strictly less than uncountable dimension), and if that is the case then there is a one-to-one correspondance between self-adjoint extensions $(D(S), S)$ of T and unitary maps $U : H_+ \to H_-$. This correspondance is given by the following construction (see [**RS2**,] for details): for a given U, we let

$$D = \{v \in H \mid v = v_0 + w + Uw, \text{ where } v_0 \in D(T),\ w \in H_+\},$$

and we define S on D, in terms of expressions $v = v_0 + w + Uw$, by

$$S(v_0 + w + Uw) = T(v_0) + iw - iUw.$$

4.5. Basic spectral theory for unbounded operators

We have already observed that bounded operators can lead to unbounded ones if one considers the resolvent $(T - \lambda)^{-1}$ for λ in the continuous spectrum. This connection can be reversed, and it explains why the following definitions make sense. They are the same as for bounded operators (with the necessary subtlety that one must remember that "injective" means "injective on $D(T)$").

Definition 4.34. Let H be a Hilbert space, $(D(T), T) \in DD(H)$ a *closed* operator, and let $\lambda \in \mathbf{C}$.

(1) The point λ is in the resolvent set $\rho(T) = \rho((D(T), T))$ of T if $\lambda - T$ is a bijection[3]

$$D(T) \to H$$

[3] There is no difficulty in speaking of $\lambda - T$ since the identity is continuous.

with bounded inverse, i.e., $\lambda - T$ is surjective and for some constant $c > 0$, we have

$$\lambda v - Tv = w \Rightarrow \|v\| \leqslant c\|w\|.$$

(2) The complement of the resolvent set of $(D(T), T)$ is the spectrum $\sigma((D(T), T))$ of T. The point spectrum $\sigma_p(T)$ is the set of λ for which $\lambda - T$ is not injective on $D(T)$; the residual spectrum $\sigma_r(T)$ is the set of those λ for which $\lambda - T$ is injective on $D(T)$ but has non-dense image; the continuous spectrum $\sigma_c(T)$ is the remainder of the spectrum.

Note the restriction to closed operators in the definition. One relaxes the definition to $T \in DD^*(H)$, a closable operator, by *defining* the spectrum (and its subsets) and the resolvent set to be those of the closure $\bar{T}$.

The closedness is useful because it implies that, for $\lambda \in \mathbf{C}$, the operator

$$\lambda - T \,:\, D(T) \to H$$

is invertible *if and only* if it is bijective, by Lemma 4.16 (with $S = \lambda\mathrm{Id}$). This, in turn, shows that the spectrum is indeed partitioned into its point, continuous and residual parts as defined above.

Proposition 4.35. *Let H be a Hilbert space and $(D(T), T) \in DD(H)$ a closed operator.*

(1) *The resolvent set is open in $\mathbf{C}$ and $\lambda \mapsto R_\lambda(T) = (\lambda - T)^{-1}$ is an analytic $L(H)$-valued function defined on $\rho(T)$.*

(2) *All resolvent operators $R_\lambda(T)$ for $\lambda \in \rho(T)$ commute, and satisfy*

$$R_\lambda(T) - R_\mu(T) = (\mu - \lambda)R_\mu(T)R_\lambda(T).$$

Remark 4.36. Note (and we will give examples below) that the spectrum *may be empty* and that it may also be equal to $\mathbf{C}$ (in which case the resolvent set is empty), or other unbounded subsets of $\mathbf{C}$. So the spectrum behaves differently than in the case of bounded operators. However, for self-adjoint operators, it will be seen that the spectrum is still rich enough to give a classification of unbounded self-adjoint operators.

Proof. One can check that the usual proof for bounded operators goes through without difficulty, but we give (some) details for completeness. Thus let $\lambda_0 \in \rho(T)$ be given, and let $S = (\lambda_0 - T)^{-1}$ be its (bounded) inverse. For any $\lambda \in \mathbf{C}$, we write

$$\begin{aligned}\lambda - T &= (\lambda - \lambda_0) + (\lambda_0 - T)\\ &= (\lambda - \lambda_0)S(\lambda_0 - T) + (\lambda_0 - T) = (\mathrm{Id} + (\lambda - \lambda_0)S)(\lambda_0 - T),\end{aligned}$$

where all these steps make sense as operators defined on $D(T)$.

For λ such that $|\lambda - \lambda_0| < \|S\|^{-1}$, we know the familiar geometric series expansion

$$(\mathrm{Id} + (\lambda - \lambda_0)S)^{-1} = \sum_{k \geqslant 0} (-1)^k (\lambda - \lambda_0)^k S^k \in L(H),$$

which shows that $\mathrm{Id} + (\lambda - \lambda_0)S$ is invertible in $L(H)$. It follows that $\lambda - T$ is a linear bijection with inverse

$$S \sum_{k \geqslant 0} (-1)^k (\lambda - \lambda_0)^k S^k$$

so $\lambda \in \rho(T)$, which proves that the resolvent set is an open set. Moreover, since this series is a convergent power-series expansion in λ, this also means that the resolvent on $\rho(T)$ is indeed an analytic function.

The commutativity of the resolvents and the resolvent formulas are proved just as usual, and left as exercises. ☐

EXAMPLE 4.37. We give examples where the spectrum is either empty, or the whole complex plane.

(1) Let $H = L^2(\mathbf{C})$, the square-integrable functions on $\mathbf{C}$ for the 2-dimensional Lebesgue measure. Consider the operator in $T \in DD(H)$ given by

$$(D, M_z),$$

where $D = \{\varphi \in H \mid z \mapsto z\varphi(z) \in H\}$, and $M_z\varphi = z\varphi$. Just like for multiplication by x on $L^2(\mathbf{R})$, this is a closed operator. But we claim its spectrum is $\mathbf{C}$. Indeed, let λ_0 be arbitrary in $\mathbf{C}$; consider the function f which is the characteristic function of a disc of radius 1 around λ_0. Then $f \in H$, but we claim that f is not in the image of $M_z - \lambda_0$, so that $\lambda_0 \in \sigma(T)$, as claimed. To see this, note that $(\lambda_0 - M_z)\varphi = f$ implies that

$$\varphi(z) = \frac{1}{\lambda_0 - z}$$

for almost all z in the disc of radius 1 around λ_0. But the right-hand side is not square integrable on this disc, so of course φ can not exist.

(2) Here is an example with $\sigma(T) = \emptyset$. Consider $H = L^2([0, 1])$, and the subspace $D = \mathrm{Im}(V)$, where V (often called the Volterra operator) is the bounded linear map such that

$$V\varphi(x) = \int_0^x \varphi(t)dt.$$

It is clear that D is dense in H, since for instance it clearly contains all C^1 functions with compact support in $]0, 1[$ and $L^2(]0, 1[) = L^2([0, 1])$. Moreover, V is injective (e.g., because it is easily checked that $V\varphi$ is differentiable almost everywhere with $(V\varphi)'(x) = \varphi(x)$ for almost all x), so 0 is in the continuous spectrum of V, and we can define (D, T) as the inverse of V, as in Example 4.11: $T\varphi = \psi$, for $\varphi \in D$, if and only if $V\psi = \varphi$ (intuitively, T is just the derivative operator, defined on the space D).

As in Example 4.11, (D, T) is closed. Now, let $\lambda_0 \in \mathbf{C}$ be given. We claim that $\lambda_0 - T$ is invertible, so that $\lambda_0 \notin \sigma(T)$. Note that this is by construction for $\lambda_0 = 0$. In general, we can guess a formula for the inverse of $\lambda_0 - T$ by remarking that – at least formally – it is a matter of solving the differential equation

$$-y' + \lambda_0 y = \varphi$$

for a given $\varphi \in H$. This is a linear, non-homogeneous, first order Ordinary Differential Equation (with constant coefficients), the solution of which is well-known. The homogeneous equation $y' - \lambda_0 y = 0$ has solutions

$$y(x) = Ce^{\lambda_0 x}, \qquad C \in \mathbf{C},$$

and if we use the method of variation of constants to look for solutions of the non-homogeneous equation, this leads to

$$y(x) = C(x)e^{\lambda_0 x}, \qquad y'(x) = C'(x)e^{\lambda_0 x} + \lambda_0 y(x),$$

so that we need $C' = e^{-\lambda_0 x}\varphi$ to solve the equation. So, in other words, we just construct the inverse $(\lambda_0 - T)^{-1}$ by

$$S_{\lambda_0}(\varphi)(x) = -e^{\lambda_0 x} \int_0^x \varphi(t)e^{-\lambda_0 t}dt.$$

It is clear that S_{λ_0} thus defined is a bounded linear operator on H. Differentiating under the integral sign, we obtain

$$S_{\lambda_0}(\varphi)'(x) = \lambda_0 S_{\lambda_0}(\varphi)(x) - \varphi(x),$$

in $L^2([0,1])$, which confirms that S_{λ_0} is inverse to $\lambda_0 - T$. Moreover, a simple integration by parts shows that

$$V(\lambda_0 S_{\lambda_0}\varphi - \varphi) = S_{\lambda_0}(\varphi)$$

so that S_{λ_0} maps H to D, and satisfies

$$TS_{\lambda_0}\varphi = \lambda_0 S_{\lambda_0}\varphi - \varphi,$$

as desired.

The next fact does not look surprising, but Proposition 4.27 and the example following the statement shows it is not a formality. It will be important to permit the construction of Example 4.11 to be applied to relate unbounded self-adjoint operators to bounded operators.

PROPOSITION 4.38. *Let H be a Hilbert space and let $(D(T), T) \in DD^*(H)$ be a self-adjoint operator. Then $\sigma(T) \subset \mathbf{R}$.*

PROOF. Let $\lambda_0 = x_0 + iy_0 \in \mathbf{C} - \mathbf{R}$, so that $y_0 \neq 0$. First, note that the operator $(D(T), (T - x_0)/y_0)$ is also self-adjoint by Lemma 4.26. Then Proposition 4.27 applied to it shows that $\mathrm{Im}((T - x_0)/y_0 - i) = H$, and consequently, $\mathrm{Im}(T - \lambda_0) = H$. Applied to $-\bar{\lambda}_0$, this also gives that $\mathrm{Im}(T + \bar{\lambda}_0) = H$. Next, if $v \in \mathrm{Ker}(T - \lambda_0)$, we get from the relation

$$0 = \langle (T - \lambda_0)v, w\rangle = \langle v, (T^* - \bar{\lambda}_0)w\rangle = \langle v, (T - \bar{\lambda}_0)w\rangle$$

for all $w \in D(T)$ that $v \in \mathrm{Im}(T + \bar{\lambda}_0)^{\perp} = 0$, so $T - \lambda_0$ is injective on $D(T)$. So it is indeed bijective from $D(T)$ to H, and $\lambda_0 \notin \sigma(T)$. ☐

EXAMPLE 4.39. Let (X, μ) be a finite measure space, g a real-valued measurable function on X, and let $(D(M_g), M_g)$ be the corresponding multiplication operator acting on $L^2(X, \mu)$ (Example 4.6). Then, as in the bounded case, we have

$$\sigma(M_g) = \mathrm{supp}\, g_*(\mu),$$

i.e., the spectrum of M_g is the support of the image measure $g_*\mu$ (or equivalently the essential range of g in $\mathbf{R}$). Indeed, the argument found in Example 3.10 did not use the boundedness of the function g (boundedness was only necessary for $1/(g - \lambda)$, to obtain bounded inverses, which is exactly what we need here again), and so it applies verbatim here.

REMARK 4.40. If $(D(T), T)$ is symmetric, and not self-adjoint, one can show that only the three following possibilities exist for $\sigma(T)$: either $\sigma(T) = \mathbf{C}$, $\sigma(T) = \mathbf{H}$ or $\sigma(T) = \bar{\mathbf{H}}$, where

$$\mathbf{H} = \{z \in \mathbf{C} \mid \mathrm{Im}(z) > 0\}$$

is the upper-half plane in $\mathbf{C}$ (see, e.g., [**RS2**, X]).

We conclude with a lemma which seems obvious, but where (again) the unboundedness requires some care. (This lemma also explains in part why spectral theory really makes sense only for closable operators).

LEMMA 4.41. *Let $(D(T), T) \in DD(H)$ be a closed densely-defined operator on a Hilbert space H. Then for every $\lambda \in \mathbf{C}$, the λ-eigenspace*

$$\mathrm{Ker}(\lambda - T) = \{v \in D(T) \mid Tv = \lambda v\}$$

is closed in H.

Proof. Of course, the point is that one can not simply say that this subspace is the inverse image of the closed set $\{0\}$ by a continuous map. However, we can write

$$\operatorname{Ker}(\lambda - T) = \pi^{-1}(\Gamma(T) \cap \Gamma_\lambda)$$

where $\pi : H \times H \to H$ is the first projection and Γ_λ is the graph of the multiplication by λ (i.e., $(v, w) \in \Gamma_\lambda$ if and only if $w = \lambda v$). Since $\Gamma(T)$ and Γ_λ are both closed and π *is* continuous, it follows that the eigenspace is also closed. $\square$

4.6. The spectral theorem

We now state the spectral theorem for unbounded self-adjoint operators, in the form similar to Theorem 3.1.

Theorem 4.42 (Spectral theorem for unbounded operators). *Let H be a separable Hilbert space and let $(D(T), T) \in DD^*(H)$ be a self-adjoint operator on H. Then there exists a finite measure space (X, μ), a measurable real-valued function $g : X \to \mathbf{R}$, and a unitary map $U : L^2(X, \mu) \to H$ such that $U^{-1}TU = M_g$, the operator of multiplicative by g, i.e., such that*

$$U^{-1}(D(T)) = D(M_g) = \{\varphi \in L^2(X, \mu) \mid g\varphi \in L^2(X, \mu)\},$$

and

$$U(g\varphi) = T(U(\varphi)) \tag{4.8}$$

for $\varphi \in D(M_g)$.

Moreover, we have

$$\sigma(T) = \sigma(M_g) = \operatorname{supp} g_*(\mu),$$

which is the essential range of g.

The only difference with the bounded case is therefore that the spectrum, and the function g, are not necessarily bounded anymore.

Proof. The basic idea is to represent T as the (unbounded) inverse of a suitable normal *bounded* operator $S \in L(H)$, and to apply the spectral theorem for the latter.[4]

The auxiliary bounded operator can be chosen in many ways. We select a fairly standard one, namely the inverse of $T + i$. Indeed, since T is self-adjoint, we know that $-i$ is in the resolvent set $\rho(T)$ (by Proposition 4.38) so that $S = (T + i)^{-1}$ is continuous.

We next check that S is normal; this is intuitively clear, since we should have $S^* = (T - i)^{-1}$ (which is also a bounded operator by the same reasoning), and Proposition 4.35 shows that the resolvents $(T + i)^{-1}$ and $(T - i)^{-1}$ commute. To confirm this, we simply write

$$\langle Sv, w\rangle = \langle v_1, (T - i)w_1\rangle = \langle (T + i)v_1, w_1\rangle = \langle v, (T - i)^{-1}w\rangle$$

for $v = (T + i)v_1$ with $v_1 = Sv$ and $w = (T - i)w_1$ with $w_1 \in D(T)$, where we used the self-adjointness of T at the second step.

Having done so, we deduce from Theorem 3.1 (for normal operators) that there exists a finite measure space (X, μ), $U : L^2(X, \mu) \to H$ a unitary isomorphism, and a bounded function $h \in L^\infty(X, \mu)$ such that

$$S = UM_hU^{-1} \tag{4.9}$$

in terms of the multiplication by h (note h is not real-valued here since h is merely normal).

[4] Intuitively, one may think of this as an extension of the functional calculus to some *unbounded* functions on the spectrum.

It is now intuitively clear that we should have

$$T = UM_gU^{-1}, \quad \text{where} \quad g(x) = \frac{1}{h(x)} - i$$

(so that, of course, we have $(h+i)^{-1} = g$).

To make sense of (and prove) this, we observe first that $h(x) \neq 0$ μ-almost everywhere, allowing us to defined g (indeed, the characteristic function χ of the set $\{x \mid h(x) = 0\}$ is in the kernel of M_h, so that $U\chi \in \mathrm{Ker}(S) = 0$, hence χ is zero almost-everywhere since U is unitary).

We thus have the densely-defined multiplication operator $(D(M_g), M_g)$ defined as in Example 4.6. We claim next that

$$U^{-1}(D(T)) = D(M_g),$$

which means simply that U, restricted to $D(M_g)$, is an isomorphism $D(M_g) \simeq D(T)$. Indeed, we have first

$$D(M_g) = \mathrm{Im}(M_h), \tag{4.10}$$

because

$$(\varphi \in L^2,\ g\varphi \in L^2) \Leftrightarrow (\varphi \in L^2,\ (1/h - i)\varphi \in L^2) \Leftrightarrow (\varphi \in L^2,\ h^{-1}\varphi \in L^2),$$

and hence $\varphi = M_h(h^{-1}\varphi) \in \mathrm{Im}(M_h)$ for any $\varphi \in D(M_g)$, with the converse equally clear since $gh\varphi = \varphi - ih\varphi \in L^2$ for any $\varphi \in L^2$. Next, the unitary equivalence of S and M_h means in particular that

$$U^{-1}(D(T)) = U^{-1}(\mathrm{Im}(S)) = \mathrm{Im}(M_h) = D(M_g),$$

as desired.

Now the unitary equivalence (4.8) is easy to check: it suffices to note that M_{g+i} is defined on $D(M_g)$ and is a bijection $M_{g+i} : D(M_g) \to L^2(X, \mu)$ with inverse

$$M_{g+i}^{-1} = M_h,$$

so that we have a square

$$\begin{array}{ccc} L^2(X,\mu) & \xrightarrow{U} & H \\ M_h \downarrow & & \downarrow\ S \\ D(M_g) & \xrightarrow{U} & D(T) \end{array}$$

where all maps are linear and bijective, and the square commutes; purely set-theoretically, the reciprocal square

$$\begin{array}{ccc} L^2(X,\mu) & \xrightarrow{U} & H \\ M_{g+i} \uparrow & & \uparrow\ T+i \\ D(M_g) & \xrightarrow{U} & D(T) \end{array}$$

also commutes, which means that $UM_{g+i} = (T+i)U$ as operators defined on $D(M_g)$, and by subtracting $i\mathrm{Id}$, we get the conclusion.

The final formula for the spectrum of M_g is just a reminder of Example 4.39. Finally, because the spectrum of T is a subset of $\mathbf{R}$, so must be the spectrum of M_g, and this implies that g is real-valued μ-almost everywhere (by definition of the support). ☐

Using the spectral theorem, we now define explicitly the functional calculus for unbounded self-adjoint operators; indeed, we define the functional calculus for all bounded functions on $\mathbf{R}$ (something we did not do explicitly for bounded operators in the previous chapter).

Corollary 4.43. *Let H be a Hilbert space and $(D(T),T)$ a self-adjoint operator on H. There exists a unique map*

$$\begin{cases} L^\infty(\mathbf{R}) \to L(H) \\ f \mapsto f(T) \end{cases}$$

continuous with norm $\leqslant 1$, with the following properties:

(i) *This is a ring homomorphism;*

(ii) *We have $f(T)^* = \bar{f}(T)$ for all f;*

(iii) *We have $f(T) \geqslant 0$ if $f \geqslant 0$;*

(iv) *If f_n converges pointwise to $f \in L^\infty(\mathbf{R})$ and $\|f_n\|_\infty$ is bounded, then $f_n(T) \to f(T)$ strongly;*

(v) *If $f_n \in L^\infty$ converge pointwise to the identity function $x \mapsto x$, and $|f_n(x)| \leqslant |x|$, we have*

$$f_n(T)v \to Tv$$

for all $v \in D(T)$.

Proof. The existence is easy: using the Spectral Theorem, one finds a unitary equivalence

$$UM_gU^{-1} = T$$

for some multiplication operator M_g on $L^2(X,\mu)$, (X,μ) being a finite measure space. Then we *define*

$$f(T) = UM_{f\circ g}U^{-1},$$

where the multiplication operator, being associated to the *bounded* function $f \circ g$, is bounded, so that $f(T) \in L(H)$.

With this definition, Properties (i) to (v) are easily checked to hold. For instance, consider (iv): for any vector $v \in H$ first, let $\varphi = U^{-1}v$; then we have by definition

$$f_n(M_g)\varphi = (f_n \circ g) \cdot \varphi,$$

and the assumption gives

$$|f_n(g(x))\varphi(x) - f(g(x))\varphi(x)|^2 \to 0, \qquad x \in X,$$

and (with C such that $\|f_n\|_\infty \leqslant C$ for all n) the domination relation

$$|f_n(g(x))\varphi(x) - f(g(x))\varphi(x)|^2 \leqslant 2(C + \|f\|_\infty)\|\varphi(x)\|^2 \in L^1(X,\mu),$$

so that the dominated convergence theorem leads to

$$f_n(M_g)\varphi \to f(M_g)\varphi,$$

and after applying U, we get $f_n(T)v \to f(T)v$.

The unicity of this functional calculus is not entirely obvious because (contrary to the case of the functional calculus for bounded operators) the conditions (i) to (v) do not supply us with $f(T)$ for any f except $f = 1$, and (asymptotically) for sequences of bounded function converging to the identity function $x \mapsto x$.

Let ϕ be any map $f \mapsto \phi(f) = f(T)$ with the properties above. Using the fact that ϕ is a ring homomorphism, we will first prove that $\phi(f)$ is the resolvent $R_\lambda(T)$ when $f(x) = 1/(\lambda - x)$, where $\lambda \notin \mathbf{R}$ is any non-real number; note that f is then bounded on $\mathbf{R}$.

To show this property, consider the functions χ_n which are the characteristic functions of $[-n,n]$, and denote the identity function by $X : x \mapsto x$. For every n, we have the relation

$$f(\lambda - X)\chi_n = \chi_n$$

(a relations as functions on $\mathbf{R}$). Since $(\lambda - X)\chi_n$ is bounded, we obtain

$$\phi(f)(\lambda I_n - X_n) = I_n, \qquad I_n = \phi(\chi_n), \quad X_n = \phi(X\chi_n)$$

in $L(H)$ after applying ϕ.

Since $\chi_n(x) \to 1$ for all x and $\|\chi_n\| \leqslant 1$ for all n, and since moreover $x\chi_n(x) \to x$ and $|x\chi_n(x)| \leqslant |x|$ for all x, properties (iv) and (v) of the map ϕ show that for any vector $v \in D(T)$, we have

$$I_n(v) \to v, \qquad X_n(v) \to Tv,$$

so that (since $\phi(f) \in L(H)$ is continuous) we have $\phi(f)(\lambda v - Tv) = v$, i.e., $\phi(f)(\lambda - T) = 1$ on $D(T)$. Since $\lambda - T$ is invertible, it follows that $\phi(f) = R_\lambda(T)$, as claimed.

Now we recall (Corollary 1.12) that the closure, in the L^∞ norm, of the span of functions of the type

$$x \mapsto \frac{1}{\lambda - x}, \qquad \lambda \notin \mathbf{R},$$

is the space $C_0(\mathbf{R})$ of continuous functions on $\mathbf{R}$ which tend to 0 at $\pm\infty$. So any two maps $f \mapsto \phi(f)$ of the type described must coincide on $C_0(\mathbf{R})$. But then condition (iv) implies that they coincide on all functions which can be obtained as a pointwise limit of continuous functions $f_n \in C_0(\mathbf{R})$ with bounded L^∞ norms. This space is known to be $L^\infty(\mathbf{R})$: indeed, first any characteristic function of a bounded measurable set lies in it by the regularity properties of the Lebesgue measure, then simple functions (which take only finitely many values) also, and these are dense in $L^\infty(\mathbf{R})$ for pointwise bounded convergence. □

We can also define the spectral measures; it is important to note that they also exist for any vector, including those not in $D(T)$.

Proposition 4.44. *Let H be a Hilbert space, $(D(T), T)$ a self-adjoint operator on H and $v \in H$. There exists a unique Borel measure $\mu_{v,T}$ on $\mathbf{R}$, called the spectral measure of v with respect to T, such that*

$$\langle f(T)v, v\rangle = \int_{\mathbf{R}} f(x) d\mu_{v,T}(x)$$

for all f bounded and continuous on $\mathbf{R}$. In particular, $\mu_{v,T}$ is a finite measure with $\mu_{v,T}(\mathbf{R}) = \|v\|^2$.

Moreover, if $v \in D(T)$, we have

$$\int_{\mathbf{R}} |x|^2 d\mu_{v,T}(x) < +\infty, \quad \int_{\mathbf{R}} |x|^2 d\mu_{v,T}(x) < +\infty,$$

and in that case we have

$$\int_{\mathbf{R}} x d\mu_{v,T}(x) = \langle Tv, v\rangle.$$

Proof. By the functional calculus the map

$$f \mapsto \langle f(T)v, v\rangle$$

is a well-defined positive linear map on the space of bounded continuous functions on $\mathbf{R}$, and thus is given by integration of f with respect to some Borel measure $\mu_{v,T}$, by the Riesz-Markov Theorem (Theorem 1.15).

To prove the last property, we represent T as a multiplication operator M_g. Then we have

$$\langle f(M_g)\varphi, \varphi\rangle = \int_X f(g(x))|\varphi(x)|^2 d\mu(x),$$

and it follows that μ_{φ,M_g} is – as in the bounded case – given by $g_*(|\varphi|^2 d\mu)$.

Now if $\varphi \in D(M_g)$, we have

$$\int_{\mathbf{R}} x^2 d\mu_{\varphi,M_g}(x) = \int_{\mathbf{R}} |g(x)|^2 |\varphi(x)|^2 d\mu(x) < +\infty.$$

By Cauchy-Schwarz, since μ_{φ,M_g} is a finite measure, we deduce that

$$\int_{\mathbf{R}} |x| d\mu_{\varphi,M_g} < +\infty.$$

The last formula is then immediate by the dominated convergence theorem and Property (v) of Corollary 4.43. □

Note that, by induction, one can check the following: if v is such that $v \in D(T)$, $Tv \in D(T)$,..., $T^{n-1}v \in D(T)$, then

$$\int_{\mathbf{R}} |x|^n d\mu_{v,T}(x) < +\infty,$$

and

$$\langle T^j v, v\rangle = \int_{\mathbf{R}} x^j d\mu_{v,T}(x)$$

for all $0 \leqslant j \leqslant n$.

CHAPTER 5

Applications, I: the Laplace operator

In this first chapter describing applications of spectral theory, we highlight one of the most important unbounded (differential) operators, the Laplace operator. Although there are many generalizations beyond the setting we will use, we restrict our attention to the Laplace operator on open subsets in euclidean space $\mathbf{R}^d$.

5.1. Definition and self-adjointess issues

Let $U \subset \mathbf{R}^d$ be an open subset, assumed to be non-empty, and let $H = L^2(U)$, the space of square integrable functions on U, with respect to Lebesgue measure. We will consider the Laplace operator $(D(\Delta), \Delta)$, defined as follows:
– The domain is $D(\Delta) = C_c^\infty(U)$, the space of smooth and compactly supported functions on U; this is known to be dense in H for the L^2 norm.
– For $\varphi \in D(\Delta)$, we put

$$\Delta\varphi = -\sum_{j=1}^{n} \frac{\partial^2 \varphi}{\partial x_j^2}.$$

This is again a smooth, compactly supported function, and therefore it is bounded and lies in $L^2(U)$.

From previous examples of differential operators, we can not expect that, on such a small domain, Δ will be self-adjoint. However, we will now proceed to show that, at least, there are self-adjoint extensions.

Proposition 5.1. *Let $U \subset \mathbf{R}^d$ be a non-empty open subset, and let $(D(\Delta), \Delta)$ be the Laplace operator defined above.*

(1) *The Laplace operator is symmetric on $D(\Delta)$, i.e., we have*

$$\langle \Delta\varphi, \psi \rangle = \langle \varphi, \Delta\psi \rangle$$

for all φ, $\psi \in C_c^\infty(U)$.

(2) *The Laplace operator is* positive *on $D(\Delta)$, i.e., we have*

$$\langle \Delta\varphi, \varphi \rangle \geqslant 0$$

for all $\varphi \in C_c^\infty(U)$.

Proof. Using integration by parts twice and the fact that the functions in the domain are compactly supported with respect to any fixed coordinate, we obtain

$$\begin{aligned}\langle \Delta\varphi, \psi \rangle &= \sum_{j=1}^{n} \int_U -\partial_{x_j}^2 \varphi(x) \overline{\psi}(x) dx \\ &= \sum_{j=1}^{n} \int_U \partial_{x_j} \varphi(x) \overline{\partial_{x_j} \psi(x)} dx \\ &= \langle \varphi, \Delta\psi \rangle\end{aligned}$$

for φ, $\psi \in D$, so that Δ is symmetric. Moreover, the intermediate step leads to

$$\langle \Delta\varphi, \varphi\rangle = \sum_{j=1}^{m} \int_U |\partial_{x_j}\varphi(x)|^2 dx = \int_U \|\nabla\varphi\|^2 dx \geqslant 0$$

for all $\varphi \in D$, so that the operator is positive (note that if the minus sign was omitted from the definition, the operator would be negative). Here we used the gradient $\nabla\varphi$, which is the (densely defined) operator defined on D, with values in H^n, by

$$\nabla\varphi = (\partial_{x_j}\varphi)_j.$$

□

Then we will now deduce:

Corollary 5.2. *Let $U \subset \mathbf{R}^d$ be a non-empty open subset. Then the Laplace operator admits at least one self-adjoint extension.*

As we will see, it is not the case that Δ is essentially self-adjoint in general, so we can not use the self-adjointness criterion of Proposition 4.30. However, we first sketch a non-constructive proof, using the results stated in Remark 4.33:

Although this proves Corollary 5.2, this is not necessarily a good proof since it does not provide a good description of any self-adjoint extension of Δ. So in the next section, we give another proof by constructing a *specific* self-adjoint extension, called the Friedrichs extension.

5.2. Positive operators and the Friedrichs extension

Corollary 5.2 is true for all positive operators on a Hilbert space.

Theorem 5.3 (Friedrichs). *Let H be a Hilbert space, $(D(T), T) \in DD^*(H)$ a* symmetric *operator which is* positive*, i.e., such that*

$$\langle T(v), v\rangle \geqslant 0,$$

for all $v \in D(T)$. Then T admits at least one self-adjoint extension $(D(S), S)$, called the Friedrichs extension, such that

$$\langle S(v), v\rangle \geqslant 0 \text{ for all } v \in D(S).$$

It should be noted that it is not the case that *all* self-adjoint extensions of a positive symmetric operator are positive.

Lemma 5.4. *Let H_1, H_2 be Hilbert spaces, and let $J : H_1 \to H_2$ be in $L(H_1, H_2)$, injective, with $\mathrm{Im}(J)$ dense in H_2. Then the map*

$$JJ^* : H_2 \to H_2$$

is injective, its image is dense in H_2 and $(D(S), S) = (\mathrm{Im}(JJ^*), (JJ^*)^{-1})$ is self-adjoint in $DD^*(H_2)$.

Proof. Since $\mathrm{Im}(J)$ is dense in H_2, $\mathrm{Ker}(J^*) = \mathrm{Im}(J)^\perp = 0$, so J^* is injective, and hence so is JJ^*. Moreover, $JJ^* \in L(H_2)$ is self-adjoint, so $\mathrm{Im}(JJ^*)^\perp = \mathrm{Ker}(JJ^*) = 0$, so JJ^* has dense image, as stated. This means the operator $(D(S), S)$ is well-defined in $DD(H_2)$. Because it is the inverse of a bounded operator, its graph is closed, so $S \in DD^*(H_2)$ is closed.

We now show that S is symmetric: indeed, starting with

$$\langle JJ^*v, w\rangle_{H_2} = \langle v, JJ^*w\rangle_{H_2},$$

for v, $w \in H_2$, we see that $v_1 = JJ^*v$ ranges over all $D(S)$, so

$$\langle v_1, w\rangle_{H_2} = \langle Sv_1, JJ^*w\rangle_{H_2}$$

for $v_1 \in D(S)$, $w \in H_2$. Now writing $w_1 = JJ^*w \in D(S)$, we get the desired formula

$$\langle v_1, Sw_1\rangle_{H_2} = \langle Sv_1, w_1\rangle_{H_2}$$

for v_1, $w_1 \in D(S)$.

To conclude that S is in fact self-adjoint, we apply Proposition 4.27, (3). Thus let $w \in H$ be given, and we try to solve the equation

$$Sv + iv = w,$$

with $v \in D(S) = \mathrm{Im}(JJ^*)$. To guess how to proceed, apply JJ^* to both sides: this leads to the necessary condition

$$v + iJJ^*v = w_1, \qquad \text{where} \qquad w_1 = JJ^*w \in D(S).$$

Now since we know that JJ^* is self-adjoint, the same criterion (or the simpler fact that $-i \notin \sigma(JJ^*)$, since $JJ^* \in L(H_2)$ is bounded) shows there exists $v \in H$ solving this second equation. In particular, we then have

$$v = w_1 - iJJ^*v = JJ^*(w - iv) \in D(S),$$

so we can apply S to both sides of the equation to get $Sv + iv = w$ with $v \in D(S)$, as desired. Of course, a similar argument applies to show that $\mathrm{Im}(S - i) = H$, and we conclude that S is self-adjoint as claimed. □

Proof of Theorem 5.3. Consider the operator $(D, \tilde{T}) = (D(T), T + \mathrm{Id})$. It is clearly symmetric and satisfies

$$\langle \tilde{T}v, v\rangle = \langle Tv, v\rangle + \|v\|^2 \geqslant \|v\|^2. \tag{5.1}$$

We will show that $\tilde{T}$ has a self-adjoint extension, say $(D(S), S)$, and it will then follow immediately that $(D(S), S - \mathrm{Id})$ is a self-adjoint extension of T, as desired (see Lemma 4.26). Moreover, we will show that

$$\langle S(v), v\rangle \geqslant \|v\|^2 \tag{5.2}$$

still holds for $v \in D(S)$, which gives the positivity of $S - \mathrm{Id}$.

To construct the self-adjoint extension, we construct first a new Hilbert space: the map

$$\begin{cases} D(T) \times D(T) \to \mathbf{C} \\ (v, w) \to \langle \tilde{T}v, w\rangle \end{cases}$$

is a positive definite inner product on $D(T)$. Of course, $D(T)$ may not be complete with respect to this inner product, but we can define the completion, which we denote by H_1, with its inner product which we also denote $\langle \cdot, \cdot\rangle_1$. Thus, by definition, we have $D(T) \subset H_1$ and $D(T)$ is dense in H_1, and moreover

$$\langle v, w\rangle_1 = \langle \tilde{T}v, w\rangle = \langle Tv, w\rangle + \langle v, w\rangle,$$

if v, $w \in D(T)$.

From this perspective, the inequality (5.1) can be interpreted as stating that the (linear) inclusion map

$$J \begin{cases} D(T) \to H \\ v \mapsto v \end{cases}$$

is in $L(D(T), H)$ with norm $\leqslant 1$. Consequently, there is a unique continuous extension, still denoted J, in $L(H_1, H)$, which satisfies

$$\|Jv\|^2 \leqslant \|v\|_1^2 \tag{5.3}$$

for $v \in H_1$.

We are going to apply Lemma 5.4 to J, so we must check that J is injective and has dense image. The latter is clear, since $\mathrm{Im}(J)$ contains $D(T)$. For the injectivity, we rewrite the definition of the inner product

$$\langle v, w\rangle_1 = \langle \tilde{T}v, Jw\rangle$$

for v, $w \in D(T)$. This extends by continuity to the same formula for $v \in D(T)$, $w \in H_1$, and it follows that $\mathrm{Ker}(J) \subset D(T)^{\perp_1} = 0$ (orthogonal for the inner product on H_1).

Thus we can indeed apply Lemma 5.4, and deduce that the densely defined operator

$$(D(S), S) = (\mathrm{Im}(JJ^*), (JJ^*)^{-1})$$

is self-adjoint in $DD^*(H)$. We now claim that this self-adjoint operator is an extension of $\tilde{T}$.

Indeed, we use again the above formula to get

$$\langle v, w\rangle_1 = \langle \tilde{T}v, Jw\rangle = \langle J^*\tilde{T}v, w\rangle_1$$

for $v \in D(T)$, $w \in H_1$, and therefore $v = J^*\tilde{T}v$ for $v \in D(T) \subset H_1$. Since J is the identity on $D(T) \subset H_1$, this even leads to

$$v = Jv = JJ^*\tilde{T}v \in \mathrm{Im}(JJ^*) = D(S),$$

so that $D(\tilde{T}) \subset D(S)$, and also $\tilde{T}v = Sv$ for $v \in D(T)$. So S is indeed an extension of $\tilde{T}$.

The last thing to check is (5.2): let v in $D(S)$, then by definition we get

$$\langle Sv, v\rangle = \langle w, JJ^*w\rangle,$$

where $w = (JJ^*)^{-1}v$. Further, we obtain

$$\langle Sv, v\rangle = \|J^*w\|_1^2 \geqslant \|JJ^*w\|^2 = \|v\|^2,$$

(applying (5.3) to J^*w) which is the desired inequality. □

Remark 5.5. The construction of the Friedrichs extension is explicit and unambiguous: if the operator has more than one self-adjoint extension, the Friedrichs extension is a particular, specific, one, and there is no choice involved that might lead to different extensions. However, the domain of the Friedrichs extension is somewhat abstractly defined, so it is not necessarily easy to describe it more explicitly, and in particular to identify it among all self-adjoint extensions. In Proposition 5.14, we will do so in one particular case.

5.3. Two basic examples

In this section, we give two examples of the Laplace operator, acting on the whole space $\mathbf{R}^n$ and on the "cube" $]0, 1[^n$, and we discuss their spectral properties by finding explicit representations of self-adjoint extensions of Δ as multiplication operators. The two examples diverge in two respects: (1) on $\mathbf{R}^n$, Δ is essentially self-adjoint (which, we recall, means that its closure is self-adjoint and is the unique self-adjoint extension of Δ), and its spectrum is purely continuous and as large as it can be (given that Δ is also positive): $\sigma(\Delta) = [0, +\infty[$ for Δ acting on $\mathbf{R}^n$; (2) on $]0, 1[^n$, Δ has distinct self-adjoint extensions, and we discuss two of them in particular: the Dirichlet and the Neumann extensions, which are distinguished by specific *boundary conditions*. On the other hand, on this relatively compact open set, Δ turns out to have purely point spectrum (and these eigenvalues can be computed fairly explicitly – although there remain some mysteries about them).

Example 5.6. Let first $U = \mathbf{R}^n$. Then the explicit representation of $(\Delta, D(\Delta))$ on U is given by the following:

Proposition 5.7. *The closure of the Laplace operator on* $\mathbf{R}^n$ *is unitarily equivalent with the multiplication operator* (D, T) *acting on* $L^2(\mathbf{R}^n)$*, where*

$$D = \{\varphi \in L^2(\mathbf{R}^n) \mid (x \mapsto \|x\|^2\varphi(x)) \in L^2(\mathbf{R}^n)\}$$
$$T\varphi(x) = (2\pi)^2\|x\|^2\varphi(x),$$

where $\|x\|$ *is the Euclidean norm* $(x_1^2 + \cdots + x_n^2)^{1/2}$ *on* $\mathbf{R}^n$*.*

In particular, Δ *is essentially self-adjoint on* $\mathbf{R}^n$*, its spectrum is equal to* $[0, +\infty[$ *and it is entirely continuous spectrum.*

Proof. The main tool in this argument is the Fourier transform (as it was in the earlier Example 4.8 with the operator of differentiation). In our setting of $\mathbf{R}^n$, this is the unitary operator

$$U : L^2(\mathbf{R}^n) \to L^2(\mathbf{R}^n)$$

such that

$$U\varphi(x) = \int_{\mathbf{R}^n} \varphi(t)e^{-2i\pi\langle x,t\rangle}dt \tag{5.4}$$

for $f \in L^2(\mathbf{R}^n) \cap L^1(\mathbf{R}^n)$, where $\langle x, t\rangle$ is the standard (Euclidean) inner product on $\mathbf{R}^n$. The basic fact is the relation

$$U(\partial_{x_j}\varphi)(x) = 2i\pi x_j U\varphi(x)$$

valid for any j, $1 \leqslant j \leqslant n$, and any $\varphi \in D(\Delta)$. Indeed, this is an easy integration by parts, similar to (4.1); from this, we derive that

$$U(\Delta\varphi)(x) = (2\pi)^2\|x\|^2 U(\varphi)(x)$$

for $\varphi \in D(\Delta)$.

We will first prove that Δ is essentially self-adjoint, using the last criterion from Proposition 4.30. Let $z = i$ or $-i$; we show that $\mathrm{Im}(\Delta + z)$ is dense in $L^2(\mathbf{R}^n)$ by showing its orthogonal complement is zero. Thus let $\varphi \in L^2(\mathbf{R}^n)$ be such that

$$\langle\varphi, (\Delta + z)\psi\rangle = 0 \text{ for all } \psi \in D(\Delta) = C_c^\infty(\mathbf{R}^n).$$

Using the unitarity of the Fourier transform, we derive

$$0 = \langle\varphi, (\Delta + z)\psi\rangle = \langle U\varphi, U((\Delta + z)\psi\rangle = \langle U\varphi, (4\pi^2\|x\|^2 + z)U\psi\rangle,$$

for all $\psi \in D(\Delta)$. Writing this down as an integral, this means that

$$0 = \langle(4\pi^2\|x\|^2 + \bar{z})U\varphi, U\psi\rangle$$

for all $\psi \in D(\Delta)$. However, since $D(\Delta)$ is dense in $L^2(\mathbf{R}^n)$, so is $UD(\Delta)$, and we deduce that

$$(4\pi^2\|x\|^2 + \bar{z})U\varphi = 0,$$

hence $U\varphi = 0$ and thus $\varphi = 0$.

This provides our proof of essential self-adjointness. Next the formula above shows that $(D(\Delta), \Delta)$ is unitarily equivalent with the multiplication operator

$$M_{4\pi^2\|\cdot\|^2} : \varphi \mapsto 4\pi^2\|x\|^2\varphi,$$

defined on $UD(\Delta)$. The latter is therefore essentially self-adjoint (this is actually what we checked!). But the multiplication operator can be defined on

$$D = \{\varphi \in L^2(\mathbf{R}^n) \mid \|x\|^2\varphi \in L^2(\mathbf{R}^n)\},$$

and indeed $(D, M_{4\pi^2\|x\|^2})$ is self-adjoint, and so is the closure of $(UD(\Delta), M_{4\pi^2\|x\|^2})$. Using the inverse Fourier transform again, it follows that the closure of Δ is unitarily equivalent with $(D, M_{4\pi^2\|x\|^2})$.

Finally, since the range (or the essential range) of the multiplier

$$x \mapsto 4\pi^2\|x\|^2$$

is $[0, +\infty[$, it follows that $\sigma(\Delta) = [0, +\infty[$. The spectrum is purely continuous spectrum, since it is clear that there is no eigenvalue of the multiplication operator. ☐

Intuitively, the "generalized eigenfunctions" are the complex exponentials

$$e_t(x) = e^{2i\pi\langle x,t\rangle}$$

for $t \in \mathbf{R}^n$, since (Δ being seen as a differential operator, see also the next example) we have

$$\Delta e_t = 4\pi^2\|t\|^2 e_t.$$

However $e_t \notin L^2(\mathbf{R}^n)$, so these are not eigenfunctions of Δ. However, note that – formally – the Fourier inversion formula

$$f(x) = \int_{\mathbf{R}^n} Uf(t)e_t(x)dx \tag{5.5}$$

looks very much like an integral form of decomposition in this "orthonormal basis" parameterized by t, with "coefficients"

$$Uf(t) = \int_{\mathbf{R}^n} f(x)\overline{e_t(x)}dx \text{ “=” } \langle f, e_t\rangle$$

(but these are really only formal expressions; their rigorous versions are given by the spectral decomposition of Δ, which is obtained by the Fourier transform).

Example 5.8. We now look at $U =]0, 1[^n$. Here the Fourier transform can be replaced by Fourier expansions, which (in contrast with (5.5)) corresponds to expanding functions $\varphi \in L^2(\mathbf{R}^n)$ in terms of a true orthonormal basis of $L^2(U)$, namely that formed by complex exponentials

$$e_k : x \mapsto e^{2i\pi\langle x,k\rangle}$$

for $k = (k_1, \ldots, k_n) \in \mathbf{Z}^n$. Those indeed belong to $L^2(U)$, and it is not difficult to show that they form an orthonormal basis of this space (e.g., using the case $n = 1$, one shows that the closure of the subspace they span contains all functions of the type

$$x \mapsto \varphi_1(x_1) \cdots \varphi_n(x_n)$$

and this type of functions is dense in $L^2(U)$, by the Stone-Weierstrass theorem for instance, since the continuous functions are dense in $C(\bar{U})$ which is dense in $L^2(\bar{U}) = L^2(U)$).

Note that $e_m \in C^\infty(U)$; moreover, if we see Δ simply as a differential operator, using the product "separation of variables"

$$e_k(x_1, \ldots, x_n) = e^{2i\pi k_1 x_1} \cdot e^{2i\pi k_2 x_2} \cdot e^{2i\pi k_n x_n},$$

and the relation

$$y'' = \alpha^2 y \text{ for } y(x) = e^{\alpha x}$$

we see that

$$\Delta e_k = (2\pi)^2 \|k\|^2 e_k.$$

However, note that e_k does not have *compact support* in U, so these L^2 eigenfunctions do not belong to the domain of the Laplace operator as we defined it. In fact, there are many more formal eigenfunctions of the Laplace operator: for any $\alpha = (\alpha_1, \ldots, \alpha_n) \in \mathbf{C}^n$, we also have

$$\Delta f_\alpha = (\alpha_1^2 + \cdots + \alpha_n^2) f_\alpha$$

for

$$f_\alpha(x) = e^{\alpha_1 x_1 + \cdots \alpha_n x_n}.$$

Using this, we can at least quickly confirm that $(D(\Delta), \Delta)$ is not essentially self-adjoint, using Proposition 4.30, (2). Namely, for any complex vector α with $\alpha_1^2 + \cdots + \alpha_n^2 = \pm i$ (for instance $\alpha = (e^{i\pi/4}, 0, \ldots, 0)$ will do for i), the function f_α is (formally) an eigenfunction of Δ with eigenvalue $\pm i$. We now check that this function is in the domain $D(\Delta^*)$ of the adjoint of $(D(\Delta), \Delta)$, and further satisfies

$$\Delta^* f_\alpha = \pm i f_\alpha.$$

Indeed, we have

$$\langle \Delta\varphi, f_\alpha \rangle = \int_U \Delta\varphi(x) e^{\langle \alpha, x \rangle} dx$$

for $\varphi \in D(\Delta)$, and the same integration by parts that shows that Δ is symmetric on $D(\Delta)$ shows that this is

$$\langle \varphi, \Delta f_\alpha \rangle$$

because the fact that φ has compact support (despite the fact that this is not so for f_α) is sufficient to ensure that the boundary terms vanish. By definition, this formula shows the desired properties of f_α, and prove that $(D(\Delta), \Delta)$ is not self-adjoint.

The computation itself does not provide us with examples of different self-adjoint extensions of $(D(\Delta), \Delta)$. However, it suggests that those should have something to do with the behavior of the boundary terms if we write integration by parts for more general functions than those with compact support. And this is indeed the case: the idea is to allow for more general functions than those with compact support, but in such a way that the boundary behavior on $\partial\bar{U}$ not only still force the integration by parts to "come out right" (to get symmetric operators), but in a way which can only be satisfied by the functions with the same boundary conditions (to get self-adjointness).

It is simpler to describe extensions of $(D(\Delta), \Delta)$ which are *essentially* self-adjoint – the actual self-adjoint extensions are their closures, the domains of which may be more delicate to describe.

It is also simpler to understand the situation first in the case $n = 1$, as one can guess by the fact that the space of solutions to the differential equation

$$-y'' = \pm iy$$

on $]0, 1[$ is two-dimensional in this case, and infinite dimensional for $n \geqslant 2$. Then one can (in many natural ways) extend the descriptions to all n.

For $n = 1$, we define three subspaces of $L^2([0, 1])$ containing $D(\Delta) = C_c^\infty(]0, 1[)$ as follows: first, define $\tilde{D}$ to be the space of function $\varphi \in C^\infty(]0, 1[)$ for which every derivative $\varphi^{(j)}$, $j \geqslant 0$, extends to a continuous function on $[0,1]$, where we denote by $\varphi^{(j)}(0)$ and $\varphi^{(j)}(1)$ the corresponding values at the boundary points. Then let

$$D_1 = \{\varphi \in \tilde{D} \mid \varphi(0) = \varphi(1),\ \varphi'(0) = \varphi'(1),\ \ldots\},$$
$$D_2 = \{\varphi \in \tilde{D} \mid \varphi(0) = \varphi(1) = 0\},$$
$$D_3 = \{\varphi \in \tilde{D} \mid \varphi'(0) = \varphi'(1) = 0\}.$$

Another interpretation of the first space is that it is really the space of smooth functions on $\mathbf{R}/\mathbf{Z}$, seen as smooth 1-periodic functions on $\mathbf{R}$: $\varphi \in D_1$ means that the function defined for x real by

$$\tilde{\varphi}(x) = \varphi(x-n) \text{ if } n \leqslant x < n+1, \text{ with } n \in \mathbf{Z},$$

is in $C^\infty(\mathbf{R})$. Because the values of derivatives at 0 and 1 may not match, D_2 is not a subspace of D_1, so the three subspaces are distinct and we have corresponding Laplace operators (D_i, Δ), each extending $(D(\Delta), \Delta)$. (The basis functions $e_k(x) = e^{2i\pi kx}$, for $k \in \mathbf{Z}$, are all in D_1, but not in D_2 or D_3, though $e_k - 1 \in D_2$ for all k).

The same integration by parts used to prove the symmetry of Δ on $D(\Delta)$ applies here to show that these extensions of $(D(\Delta), \Delta)$ are symmetric: we have

$$\begin{aligned}\langle \Delta\varphi, \psi\rangle &= \big[-\varphi'\bar{\psi}\big]_0^1 + \int_0^1 \varphi'\overline{\psi'}dt\\ &= \varphi'(0)\overline{\psi(0)} - \varphi'(1)\overline{\psi(1)} + \int_0^1 \varphi'\overline{\psi'}dt = \int_0^1 \varphi'\overline{\psi'}dt,\end{aligned}$$

the boundary terms vanishing in all three cases if φ, $\psi \in D_j$. This is a symmetric expression of φ and ψ, so Δ_j is symmetric.

The common terminology is that D_1 is the Laplace operator with *periodic* boundary conditions, D_2 is the one with *Dirichlet* boundary conditions, and D_3 is the one with *Neumann* boundary conditions.

Proposition 5.9. *The three operators (D_j, Δ), $1 \leqslant j \leqslant 3$, are essentially self-adjoint extensions of the Laplace operator defined on $D(\Delta)$. Moreover, all three $\sigma(\Delta_j) = \sigma_p(\Delta_j)$ – pure point spectrum – and the eigenvalues are given by the following:*

$$\sigma(D_1) = \{0, 4\pi^2, 16\pi^2, \ldots, 4\pi^2k^2, \ldots\},$$
$$\sigma(D_2) = \{\pi^2, 4\pi^2, \ldots, k^2\pi^2, \ldots\},$$
$$\sigma(D_3) = \{0, \pi^2, 4\pi^2, \ldots, k^2\pi^2, \ldots\}.$$

The spectrum is simple, i.e., the eigenspaces have dimension 1, for D_2 and D_3. For D_1, we have

$$\dim \operatorname{Ker}(D_1) = 1, \qquad \dim \operatorname{Ker}(D_1 - (2\pi k)^2) = 2, \text{ for } k \geqslant 1.$$

It will be clear from the proof that those three examples by no means exhaust the list of self-adjoint extensions of Δ.

Proof. Formally, this result is quite intuitive, because we will check that the boundary conditions, once they have been imposed, imply that none of the (non-zero) eigenfunctions for $\pm i$ belong to D_j. Indeed, for the eigenvalue i, those are the functions

$$\varphi(x) = ae^{\alpha x} + be^{-\alpha x}, \quad \text{with} \quad \varphi'(x) = \alpha(ae^{\alpha x} - be^{-\alpha x}),$$

where $\alpha^2 = i$. If we consider for instance D_3, we have the conditions

$$\begin{cases} a\alpha - b\alpha = 0 \\ a\alpha e^{\alpha} - b\alpha e^{-\alpha} = 0 \end{cases}$$

for φ to be in D_3. These clearly imply $a = b = 0$, as claimed, and the other two cases are similar.

Computing the eigenfunctions and eigenvalues is similarly intuitive: again, for any $\alpha \in \mathbf{C}$, the functions written above are the formal eigenfunctions of the Laplace differential operator with eigenvalue $\alpha^2 \in \mathbf{C}$, and the same computations can easily determine for which α, a and b, the boundary conditions of the operators D_j are satisfied. Consider now for instance D_2: the linear system becomes

$$\begin{cases} a + b = 0 \\ ae^{\alpha} + be^{-\alpha} = 0, \end{cases}$$

which has a non-zero solution if and only if $e^{2\alpha} = 1$, so $\alpha = ik\pi$ with $k \in \mathbf{Z}$. The solution is then given by

$$\varphi(x) = 2ia \sin(k\pi x),$$

which leads to a one-dimensional eigenspace with eigenvalue $\pi^2 k^2 \geqslant 0$ for $k \geqslant 1$ (note that the value $k = 0$ gives the zero function, k and $-k$ give the same solutions). Since those eigenfunctions are in the domain of D_2, they give elements of the point spectrum. (And similarly for D_1 and D_3 with eigenfunctions

$$\varphi(x) = ae^{2i\pi kx} + be^{-2i\pi kx}, \quad \varphi(x) = a\cos(k\pi x)$$

with $k \in \mathbf{Z}$ and $k \geqslant 0$, respectively, the eigenvalues being $4\pi^2 k^2$ and $\pi^2 k^2$).

To finish, we apply the useful Lemma 5.10 below, since we have already observed that, among the eigenfunctions of D_1 that we have found, we can find an orthonormal basis of $L^2(U)$ (namely $x \mapsto e^{2i\pi kx}$ for $k \in \mathbf{Z}$), and since the corresponding statement is not hard to check for the others.

Indeed, for D_2 and D_3, we can pick the functions

$$s_k \,:\, x \mapsto \sqrt{2}\sin(k\pi x), \quad k \geqslant 1, \qquad c_0 = 1, \quad c_k \,:\, x \mapsto \sqrt{2}\cos(k\pi x), \quad k \geqslant 1,$$

respectively. It is first clear that those functions are orthonormal in $L^2(U)$, and we next check that in fact they are orthonormal bases. Then, note that (by rescaling) the functions

$$f_k \,:\, \begin{cases}]-1, 1[\to \mathbf{C} \\ x \mapsto e^{i\pi kx} \end{cases}$$

for $k \in \mathbf{Z}$ form an orthonormal basis of $H_1 = L^2(]-1, 1[)$ for the inner product

$$\langle \varphi, \psi \rangle_1 = \frac{1}{2}\int_{-1}^{1} \varphi(t)\overline{\psi(t)}dt.$$

Decomposing $\varphi \in H_1$ in even and odd parts, we obtain an orthogonal direct sum

$$H_1 = H_1^+ \oplus^\perp H_1^-,$$

with

$$H_1^{\pm} = \{\varphi \in H_1 \mid \varphi(-x) = \pm\varphi(x) \text{ for almost all } x\},$$

and with orthogonal projections

$$\begin{cases} H_1 \to H_1^{\pm} \\ \varphi \mapsto \frac{1}{2}(\varphi(x) \pm \varphi(-x)). \end{cases}$$

By restriction, we obtain unitary isomorphisms

$$\begin{cases} H_1^{\pm} \to L^2([0,1]) \\ \varphi \mapsto \varphi \text{ restricted to } [0,1], \end{cases}$$

with inverse given by extending a function to be even or odd, as needed.

Now if we start with $\varphi \in L^2([0,1])$ and extend it to the function $\tilde{\varphi}$ on $[-1,1]$ which is even (say), we can therefore write

$$\tilde{\varphi} = \sum_{k\in\mathbf{Z}} \langle \tilde{\varphi}, f_k\rangle_1 f_k, \tag{5.6}$$

with

$$\begin{aligned} \langle \tilde{\varphi}, f_k\rangle_1 &= \frac{1}{2}\int_{-1}^{1} \tilde{\varphi}(t) e^{-ik\pi t} dt = \frac{1}{2}\int_0^1 \varphi(t)e^{-ik\pi t}dt + \frac{1}{2}\int_0^1 \varphi(t)e^{ik\pi t}dt \\ &= \begin{cases} \langle \varphi, \frac{1}{\sqrt{2}} c_k\rangle, & \text{if } k \neq 0, \\ \langle \varphi, c_0\rangle, & \text{if } k = 0. \end{cases} \end{aligned}$$

Hence, restricting (5.6) to $[0,1]$ gives

$$\varphi(x) = \sum_{k\in\mathbf{Z}} \langle \tilde{\varphi}, f_k\rangle_1 f_k = \langle \varphi, c_0\rangle + 2\sum_{k\geqslant 1} \frac{1}{\sqrt{2}} \langle \varphi, c_k\rangle c_k$$

which proves that the (c_k) generate $L^2(U)$ in the Hilbert sense, and similarly for the s_k. □

LEMMA 5.10. *Let H be a separable Hilbert space and $(D(T),T)$ a positive symmetric unbounded operator on H. Assume that there exists an orthonormal basis (e_j) of H such that $e_j \in D(T)$ for all j and which are eigenfunctions of T, i.e., $Te_j = \lambda_j e_j$ for some $\lambda_j \geqslant 0$ for all $j \geqslant 1$.*

Then T is essentially self-adjoint and its closure is unitarily equivalent with the multiplication operator (D,M) on $\ell_2(\mathbf{N})$ given by

$$D = \{(x_j) \in \ell^2(\mathbf{N}) \mid \sum_{j\geqslant 1} \lambda_j^2 |x_j|^2 < +\infty\}, \qquad M((x_j)) = (\lambda_j x_j).$$

In particular, the spectrum of $(D(T),T)$ is the closure *of the set of eigenvalues $\{\lambda_j\}$.*

Note that if the sequence (λ_j) has an accumulation point λ, the spectrum is not the same as the set of eigenvalues (this already occurred for compact operators, where 0 may belong to the spectrum even if the kernel is trivial).

PROOF. We already know that the multiplication operator is self-adjoint and has the spectrum which is described. So it is enough to show that the closure of $(D(T),T)$ is unitarily equivalent with (D,M), and of course the requisite unitary isomorphism is given by the Hilbert basis (e_j):

$$U(x_j) = \sum_{j\geqslant 1} x_j e_j \in H.$$

To see that $\bar{T} = UMU^{-1}$, start with $(v, w) \in \Gamma(T)$. Then for any j, the symmetry of the operator (in particular $D(T) \subset D(T^*)$) gives

$$\langle w, e_j \rangle = \langle Tv, e_j \rangle = \langle v, Te_j \rangle = \lambda_j \langle v, e_j \rangle,$$

and since (e_j) is an orthonormal basis, we have

$$w = \sum_j \langle w, e_j \rangle e_j = \sum_j \lambda_j \langle v, e_j \rangle e_j$$

and in particular

$$\sum_j \lambda_j^2 |\langle v, e_j \rangle|^2 = \|w\|^2 < +\infty,$$

so $U^{-1}D(T) \subset D(M)$, and $UMU^{-1}v = w = Tv$ for $v \in D(T)$, or in other words $T \subset UMU^{-1}$.

To conclude we must show that $\Gamma(T)$ is dense in $\Gamma(UMU^{-1})$. For this, let $(v, w) \in \Gamma(UMU^{-1})$; in the basis (e_j) we have

$$v = \sum_j \langle v, e_j \rangle e_j, \qquad w = \sum_j \lambda_j \langle v, e_j \rangle e_j$$

and we now define

$$v_n = \sum_{1 \leqslant j \leqslant n} \langle v, e_j \rangle e_j, \quad w_n = Tv_n = \sum_{1 \leqslant j \leqslant n} \langle v, e_j \rangle e_j$$

(which is permitted because $e_j \in D(T)$, the sum is finite and T linear). Now, in the norm of $H \times H$, we have

$$\|(v, w) - (v_n, w_n)\|^2 = \|v - v_n\|^2 + \|w - w_n\|^2,$$

and both terms tend to zero as $n \to +\infty$ (the second because of the definition of $D(M)$, of course). This proves the desired conclusion. ☐

Going back to the case of general n, we can extend the boundary conditions in many ways. The two most obvious ones are the extensions of the periodic and Dirichlet boundary conditions (D_1 and D_3), but one could use different conditions in the various directions.

Arguing as above one gets:

Corollary 5.11. *Let $U =]0, 1[^n$ with $n \geqslant 1$. Consider the operators $\Delta_p = (D_p, \Delta)$ and $\Delta_d = (D_d, \Delta)$ extending $(D(\Delta), \Delta)$ with domains given, respectively, by D_p which is the space of restrictions of C^∞ functions on $\mathbf{R}$ which are $\mathbf{Z}^n$-periodic, and D_d which is the space of function $\varphi \in C^\infty(U)$ for which every partial derivative of any order $\partial_\alpha \varphi$ extends to a continuous function on $\bar{U}$, and moreover such that $\varphi(x) = 0$ for $x \in \partial U$, where we use the same notation φ for the function and its extension to $\bar{U}$.*

Then Δ_p and Δ_d are essentially self-adjoint. Their closures have pure point spectra, given by the real numbers of the form

$$\lambda = 4\pi^2(k_1^2 + \cdots + k_n^2), \qquad k_i \in \mathbf{Z},$$

with the condition $k_i \geqslant 1$ for D_d. The multiplicity of a given λ is the number of $(k_1, \ldots, k_n) \in \mathbf{Z}^n$ with

$$\lambda = 4\pi^2(k_1^2 + \cdots + k_n^2)$$

for Δ_p, and is 2^{-n} times that number for Δ_d.

Remark 5.12. Although these open sets $]0,1[^n$ are particularly simple from the point of view of the Laplace operator, the result should not be considered "trivial". For instance, for $n \geqslant 2$, the description of the spectrum is not very explicit as a set of real numbers, and to obtain a more precise one requires fairly delicate tools related to number theory (properties of sums of squares of integers). One can show the following facts:

- If $n = 2$, a real number $\lambda = 4\pi^2 m$, $m \geqslant 1$, occurs in the spectrum of Δ_p (i.e., m is of the form $m = k_1^2 + k_2^2$ with k_1, $k_2 \geqslant 0$) if and only if, after factoring
$$m = p_1^{\nu_1} \cdots p_s^{\nu_s}$$
in terms of prime powers with $p_i \neq p_j$ if $i \neq j$ and $\nu_j \geqslant 1$, all exponents corresponding to primes p_i such that $p_i \equiv 3 \pmod 4$ are *even*. (This is due essentially to Fermat and Euler). In that case, the multiplicity, up to changes of signs of k_1 or k_2, is 2^t, with t the number of primes p_i with $p_i \equiv 1 \pmod 4$. For instance
$$4\pi^2 \cdot 17 = 4\pi^2(1^2 + 4^2) = 4\pi^2(4^2 + 1^2)$$
and the corresponding eigenfunctions are
$$e^{2i\pi x_1 + 4i\pi x_2}, \quad e^{8i\pi x_1 + 2i\pi x_2}.$$
The *number* $N(X)$ of $\lambda \leqslant X$ which are eigenvalues, not counting multiplicity, is asymptotically given by
$$N(X) \sim c\frac{X}{\sqrt{\log X}}$$
for some constant $c > 0$ (which can be written down); this is due to Landau (early 20th Century).
- For $n = 3$, a real number $\lambda = 4\pi^2 m$, $m \geqslant 1$, occurs in the spectrum of Δ_p if and only if m is not of the form $m = 4^a(8b \pm 7)$ with $a \geqslant 0$, $b \geqslant 0$; this is due to Gauss and is very delicate.
- For $n \geqslant 4$, every real number $\lambda = 4\pi^2 m$, $m \geqslant 1$ is an eigenvalue of Δ_p: this amounts to proving that every positive integer is the square of (at most) four squares of positive integers, which is a result of Lagrange. The multiplicities can then also be estimated, and are quite large. For instance, Jacobi proved that the number of representations
$$m = k_1^2 + \cdots + k_4^2$$
is given by four times the sum of odd (positive) divisors of m.

On the other hand, in all cases, we obtain a nice asymptotic for the total number of eigenvalues, counted with multiplicity:

Proposition 5.13 ("Weyl law"). *Let $n \geqslant 1$ be an integer and let*
$$M(X) = \sum_{\lambda \leqslant X} \dim(\lambda - \Delta_p)$$
be the counting function for the number of eigenvalues of the periodic Laplace operator, counted with multiplicity. Then we have
$$M(X) \sim c_n \frac{X^{n/2}}{(2\pi)^n}$$

as $X \to +\infty$, where c_n is the volume of the ball of dimension n and radius 1 in $\mathbf{R}^n$, namely

$$c_n = \frac{\pi^{n/2}}{\Gamma(1 + \frac{n}{2})}$$

where $\Gamma(z)$ is the Gamma function.

Recall that the Gamma function can be defined by

$$\Gamma(z) = \int_0^{+\infty} e^{-t} t^{z-1} dt$$

for $z > 0$, and it satisfies the following properties

$$\Gamma(z+1) = z\Gamma(z) \text{ for } z > 0,$$
$$\Gamma(n+1) = n! \text{ for } n \geqslant 0 \text{ integer},$$
$$\Gamma(1/2) = \sqrt{\pi},$$

which show that $\Gamma(1+n/2)$ can be computed easily by induction, and is always a rational number for n even, and $\sqrt{\pi}$ times a rational for n odd.

Proof. According to the corollary above, we get

$$M(X) = \sum_{\substack{k_1,\ldots,k_n \in \mathbf{Z}^n \\ 4\pi^2(k_1^2+\cdots+k_n^2) \leqslant X}} 1$$

which is the number of points $k \in \mathbf{Z}^n$ inside the *ball* S_r of radius $r = (2\pi)^{-1}\sqrt{X}$ in $\mathbf{R}^n$. Then the idea is to show that this number is approximately the volume of this sphere, because, seen from far away, this discrete set seems to fill up the ball quite well. To confirm this, a trick of Gauss is used.

Namely, for each point $k \in \mathbf{Z}^n$, consider the associated "hyper-cube"

$$C_k = \{k + x \mid x \in]0,1[^n\} \subset \mathbf{R}^n,$$

and note that $C_k \cap C_l = \emptyset$ if k and l are in $\mathbf{Z}^n$ and distinct, and that (of course) each C_k has volume 1 with respect to Lebesgue measure on $\mathbf{R}^n$ (which we denote here by $\mathrm{Vol}(\cdot)$). Thus we have

$$M(X) = \sum_{k \in \mathbf{Z}^n \cap S_r} 1 = \sum_{k \in \mathbf{Z}^n \cap S_r} \mathrm{Vol}(C_k) = \mathrm{Vol}\Big(\bigcup_{k \in \mathbf{Z}^n \cap S_r} C_k \Big)$$

Next, note that if $k \in S_r$, we have $C_r \subset S_{r+\sqrt{n}}$, since

$$\|k + x\| \leqslant \|k\| + \|x\| \leqslant \|k\| + \sqrt{n}$$

for $x \in]0,1[^n$. So we get the upper bound

$$M(X) \leqslant \mathrm{Vol}(S_{r+\sqrt{n}}).$$

Conversely, for every point $y \in S_r$, there is a unique $k \in \mathbf{Z}^n$ for which $y \in C_k$ (generalizing the "integral part" of a real number, and computed componentwise in this way). If $\|y\| \leqslant r - \sqrt{n}$, then $\|k\| \leqslant r$ since $k = y - x$ with $\|x\| \leqslant \sqrt{n}$. Thus the union of the C_k with $k \in \mathbf{Z}^n \cap S_r$ contains (at least) the whole ball of radius $r - \sqrt{n}$ – up to some parts of the boundaries of the C_k, which are of measure 0 –, and we get the upper bound

$$M(X) \geqslant \mathrm{Vol}(S_{r-\sqrt{n}}).$$

Finally, putting things together and using the homogeneity of the volume, we get the successive *encadrements*

$$c_n(r-\sqrt{n})^n = \mathrm{Vol}(S_{r-\sqrt{n}}) \leqslant M(X) \leqslant \mathrm{Vol}(S_{r+\sqrt{n}}) = c_n(r+\sqrt{n})^n$$

and since

$$c_n(r \pm \sqrt{n})^n = c_n r^n \Bigl(1 \pm \frac{\sqrt{n}}{r}\Bigr)^n \sim c_n r^n, \qquad \text{as } X \to +\infty,$$

recalling that $r = (2\pi)^{-1}\sqrt{X}$, we obtain the conclusion. □

It is quite remarkable (and surprising at first!) that this result (usually called "Weyl law") actually extends to much more general operators than this specific case, even though one can not compute the spectrum explicitly in general.

We conclude this example by identifying the Friedrichs extension:

Proposition 5.14. *Let $n \geqslant 1$ and $U =]0,1[^n$. Then the Friedrichs extension of $(D(\Delta),\Delta)$ on U is the Laplace operator with Dirichlet boundary, i.e., the closure of (D_d,Δ_d) as defined above.*

Proof. We prove this only for $n = 1$. Intuitively, this is because the domain of the Friedrichs extension is obtained from the norm defined on the space of C^1-functions (and in particular on $D(\Delta)$) by

$$\|\varphi\|_1^2 = \langle \Delta\varphi, \varphi\rangle + \|\varphi\|^2 = \|\varphi'\|^2 + \|\varphi\|^2,$$

and the domain of the Friedrichs extension is closed in this norm, while the closure of $D(\Delta)$ for it is contained in a space of continuous functions; hence the boundary conditions $\varphi(0) = \varphi(1) = 0$ which are valid for $\varphi \in D(\Delta)$ remain so for the Friedrichs extension.

For the details, let (D_F,Δ_F) be the Friedrichs extension of Δ. It is in fact enough to show that $\Delta_d \subset \Delta_F$, since it will follow that $\Delta_F^* = \Delta_F \subset \Delta_d^* = \overline{\Delta_d}$, so $\Delta_F = \overline{\Delta_d}$ as both are closed.

We recall from the proof of Theorem 5.3 that the domain D_F is given by

$$D_F = JJ^*(H_1)$$

where $(H_1, \|\cdot\|_1)$ is the completion of $D(\Delta)$ for the norm

$$\|\varphi\|_1^2 = \langle \Delta\varphi, \varphi\rangle + \langle \varphi, \varphi\rangle = \|\varphi'\|^2 + \|\varphi\|^2,$$

and $J : H_1 \to L^2(U)$ is the continuous extension of the inclusion map $D(T) \xrightarrow{J} H$. We first notice that $\mathrm{Im}(JJ^*) = \mathrm{Im}(J)$ because J^* is surjective: indeed, this follows from the fact that J is injective and $\mathrm{Im}(J^*)^\perp = \mathrm{Ker}(J) = 0$.

Now consider $\varphi \in D_d$; we claim there exists a sequence $(\varphi_n) \in D(\Delta)$ such that

$$\|\varphi_n - \varphi\|_1 \to 0. \tag{5.7}$$

From this, we deduce that $J(\varphi_n) \in H_1$ converges to some $\tilde{\varphi} \in H_1$. But we also deduce that φ_n converges to φ in $L^2(U)$ (since $\|\varphi_n - \varphi\| \leqslant \|\varphi_n - \varphi\|_1$). By unicity, it follows that

$$\varphi = J(\tilde{\varphi}) \in D_F,$$

so we get the first inclusion $D_d \subset D_F$. To check our claim is quite easy: consider any sequence ψ_n in $D(\Delta)$ such that

$$0 \leqslant \psi_n \leqslant 1, \qquad \psi_n(x) = 1 \text{ for } \frac{1}{n} \leqslant x \leqslant 1 - \frac{1}{n},$$

and with $\|\psi_n'\|_\infty \leqslant Cn$ for some constant $C > 0$. Then define

$$\varphi_n(x) = \varphi(x)\psi_n(x),$$

for which the claim (5.7) is a simple computation (the fact that $\varphi \in D_d$ enters in showing that

$$\lim_{n\to+\infty} n^2 \int_0^{1/n} |\varphi(x)|^2 dx = 0,$$

using a Taylor expansion $\varphi(x) = \varphi'(0)x + O(x^2)$ for x close to 0).

Now, to show that Δ_F extends Δ_d on D_d, we look at an eigenfunction

$$c(x) = \sin k\pi x, \qquad k \geqslant 1,$$

of Δ_d, and we claim that $\Delta_F c$, which is well-defined since $D_d \subset D_F$, is also given by $\Delta_F c = \pi^2 k^2 c$. From this, Lemma 5.10 allows us to deduce that Δ_F does extend Δ_d. To check this, let $\psi \in D(\Delta)$ be arbitrary; we then have

$$\begin{aligned}\langle \Delta_F c, \psi\rangle &= \langle c, \Delta_F \psi\rangle \\ &= \langle c, \Delta\psi\rangle \\ &= \langle c, \Delta_d \psi\rangle \\ &= \langle \Delta_d c, \psi\rangle\end{aligned}$$

where we used the fact that both Δ_F and Δ_d extend Δ on $D(\Delta)$ and are symmetric. From the density of $D(\Delta)$ in H, we conclude that

$$\Delta_F c = \Delta_d c = k^2\pi^2 c,$$

as desired. □

5.4. Survey of more general cases

The previous section has shown that the Laplace operator may exhibit different features depending on the open set U which is considered. It turns out that the qualitative results obtained in the very special cases that were considered are fairly representative of more general situations. However, the arguments must usually be different, because there is, in general, no explicit diagonalizing operator (like the Fourier transform) and the spectrum is not explicitly known (as was the case for $U =]0,1[$) to apply Lemma 5.10.

However, if one can check that there are no *smooth* eigenfunctions for the differential operator Δ corresponding to the eigenvalues $\pm i$, the following powerful technique is often very successful in checking rigorously that the $(\pm i)$-eigenspaces for the actual adjoint of Δ_j (which is defined on the bigger space $D(\Delta_j^*)$) are still trivial.

The point is that the condition

$$\Delta_j^* \varphi = \pm i\varphi$$

is sufficient to imply that φ is a *weak* solution to

$$\Delta\varphi = \pm i\varphi,$$

on U in the sense of distribution theory, i.e., that

$$\int_U \varphi(\Delta\psi \mp i\psi)dx = 0$$

for all $\psi \in C_c^\infty(U)$. Then one can apply the following statement, often known as Weyl's Lemma:

LEMMA 5.15. *Let $\lambda \in \mathbf{C}$, let U be an open set of $\mathbf{R}^n$ and let $\varphi \in L^2(U)$ – in fact, φ could just be a distribution – be such that $(\Delta+\lambda)\varphi$, computed in the sense of distribution, is a C^∞ function on U. Then φ is in fact also C^∞.*

Concretely, this means the following: if there exists $\varphi_0 \in C^\infty(U)$ such that

$$\int_U \varphi(\Delta\psi \mp i\psi)dx = \int_U \varphi_0\psi dx$$

for all $\psi \in C_c^\infty(U)$, then the function φ coincides almost everywhere with a smooth function. This fact, applied to the C^∞ function $\varphi_0 = 0$, implies that the hypothetical eigenfunctions φ are smooth.

If $U \subset \mathbf{R}^n$ is a bounded open set, the example of $]0,1[^n$ suggests that $(D(\Delta),\Delta)$ will not be essentially self-adjoint, and that the description of self-adjoint extensions will involve imposing various boundary conditions on the functions in the domain. These boundary conditions may well be complicated by the regularity (or lack of regularity) of the boundary of U: for instance, the Neumann boundary condition take the form of asking that the directional derivative of the functions vanish in the direction normal to the boundary, and this only makes sense (at least in an obvious way) if there exists a tangent space at all points of the boundary. However, this may not be true. For instance, consider a continuous function

$$f : [0,1] \to \mathbf{R}$$

which is continuous, $\geqslant 0$, but nowhere differentiable. Then the open set

$$U = \{(x,y) \in \mathbf{R}^2 \mid 0 < x < 1, \quad -f(x) < y < f(x)\}$$

will have an extremely wild boundary!

There is however a good theory when the boundary ∂U is assumed to be smooth enough (e.g., for $n=2$, if it is the image of a parameterization $\gamma : [0,1] \to \mathbf{R}^2$ where γ is C^1). But to simplify the discussion, we discuss only the simplest self-adjoint extension, namely the Dirichlet boundary condition (which is also the Friedrichs extension, though we won't prove this in full generality).

DEFINITION 5.16. Let $n \geqslant 1$ and let $U \subset \mathbf{R}^n$ be a bounded open set with boundary ∂U. The *Dirichlet laplacian* on U, or Laplace operator with Dirichlet boundary condition, is the extension (D_d, Δ_d) of $(D(\Delta),\Delta)$ defined by $\varphi \in D_d$ if and only if $\varphi \in C^\infty(U)$ is such that all partial derivatives of all order of φ, say $\partial_\alpha(\varphi)$, are the restrictions to U of functions φ_α continuous on $\bar{U}$, and if moreover we have $\varphi_0(x) = 0$ for $x \in \partial U$, where φ_0 is the continuous extension of φ itself to $\bar{U}$. The operator Δ_d acts like the usual differential operator.

One can then prove the following statement, which generalizes Corollary 5.11:

THEOREM 5.17. *Let $n \geqslant 1$ and let $U \subset \mathbf{R}^n$ be a bounded open set, with (D_d, Δ_d) the Laplace operator with Dirichlet boundary conditions on U. Then Δ_d is essentially self-adjoint, and has only point spectrum, in fact there is an orthonormal basis $(\varphi_n)_{n\geqslant 1}$ of $L^2(U)$ such that*

$$\Delta_d\varphi_n = \lambda_n\varphi_n, \qquad \lambda_n > 0,$$

with $\lambda_n \to +\infty$.

An equivalent way of phrasing the last conclusion is that Δ_d has *compact resolvent*, which by definition means that the resolvents $R_\lambda(\Delta_d) = (\lambda - \Delta_d)^{-1}$ for $\lambda \in \rho(\Delta_d)$ are compact operators. In fact, the resolvent identity

$$R_\lambda(\Delta_d) - R_{\lambda_0}(\Delta_d) = (\lambda_0 - \lambda)R_{\lambda_0}(\Delta_d)R_\lambda(\Delta_d),$$

and the fact that the space $K(H)$ of compact operators is an ideal of $L(H)$ implies that it is enough that one resolvent $R_{\lambda_0}(\Delta_d)$ be compact.

Note that having compact resolvent involves a bit more than simply having only point spectrum: all eigenspaces must also be finite-dimensional.

We take for granted that Δ_d is self-adjoint (if one does not wish to do so, redefine Δ_d to be the Friedrichs extension; the arguments below will remain valid, but the domain D_d is then possibly changed), and we show the remaining part. The idea is to exploit the knowledge of the special case $U =]0,1[^n$, some comparison results between spectra, and a characterization of eigenvalues which generalizes the min-max characterizations of eigenvalues of positive compact operators (Proposition 2.14). Indeed, we have the following theorem:

Theorem 5.18. *Let H be a Hilbert space and let $(D(T),T) \in DD^*(H)$ be a positive symmetric operator on H. For $k \geqslant 1$, define*

$$\lambda_k = \inf_{\dim V = k} \sup_{v \in V - \{0\}} \frac{\langle T(v), v\rangle}{\|v\|^2}, \tag{5.8}$$

where V runs over all subspaces $V \subset D(T)$ with $\dim V = k$.

(1) *If T is essentially self-adjoint, then the operator T has compact resolvent if and only if the increasing sequence*

$$\lambda_1 \leqslant \lambda_2 \leqslant \cdots \leqslant \lambda_n \leqslant \cdots$$

is unbounded. *In that case, the (λ_n) are the eigenvalues of T, repeated with multiplicity* $\dim(T - \lambda_n)$.

(2) *Let (D_F, T_F) be the Friedrichs extension of $(D(T),T)$, and let μ_k be defined by*

$$\mu_k = \inf_{\substack{\dim V = k \\ V \subset D_F}} \sup_{v \in V - \{0\}} \frac{\langle T(v), v\rangle}{\|v\|^2}. \tag{5.9}$$

Then we have $\mu_k = \lambda_k$, and in particular D_F has compact resolvent if and only if (λ_k) is unbounded.

To apply part (2) of this theorem in the situation of Theorem 5.17, we fix $R > 0$ large enough so that

$$U \subset U_0 =]-R, R[^n.$$

Let $(D(\Delta_0), \Delta_0)$ be the Laplace operator for U_0. Observing that the map

$$j \begin{cases} D(\Delta) \to D(\Delta_0) \\ \varphi \mapsto \varphi \chi_U \end{cases}$$

(where $\chi_U \in L^2(V)$ is the characteristic function of U) is isometric and satisfies

$$\Delta_0(j(\varphi)) = j(\Delta(\varphi)),$$

we see that $V \mapsto j(V)$ is a map sending k-dimensional subspaces V of $D(\Delta)$ to k-dimensional subspaces of $D(\Delta_0)$ in such a way that

$$\sup_{\varphi \in j(V) - \{0\}} \frac{\langle \Delta_0(\varphi), \varphi\rangle}{\|\varphi\|^2} = \sup_{\varphi \in V - \{0\}} \frac{\langle \Delta(\varphi), \varphi\rangle}{\|\varphi\|^2}.$$

From this and (5.8), we get that $\lambda_k(\Delta) \geqslant \lambda_k(\Delta_0)$ for all $k \geqslant 1$. But the $\lambda_k(\Delta_0)$, by the last part of Theorem 5.18, are the eigenvalues of the Friedrichs extension of $(D(\Delta_0), \Delta_0)$, and from Corollary 5.11, we have $\lambda_k(\Delta_0) \to +\infty$ as $k \to +\infty$. Hence we also have

$\lambda_k(\Delta) \to +\infty$, and from Theorem 5.18 again, we deduce that Δ_d, the Friedrichs extension of Δ, has compact resolvent.

Proof of Theorem 5.18. (1) Let us assume first that T is self-adjoint. Using the Spectral Theorem 4.42, we may assume that $(D(T), T) = (D(M), M)$ is a multiplication operator

$$\varphi \mapsto g\varphi, \qquad \varphi \in D(M) = \{\varphi \in L^2(X, \mu) \mid g\varphi \in L^2(X, \mu)\},$$

on $L^2(X, \mu)$, where (X, μ) is a finite measure space and $g \geqslant 0$ is a real-valued (measurable) function on X.

We first show that if the sequence (λ_k) is unbounded, the operator must have compact resolvent. For this, we first observe the following: let $c \geqslant 0$ be given, and let

$$Y_c = g^{-1}([0, c]) \subset X.$$

Let now W_c be the subspace of $L^2(X, \mu)$ of functions φ with $\varphi(x) = 0$ if $x \notin Y_c$ (up to subsets of measure zero). For $\varphi \in W_c$, we have

$$\langle M(\varphi), \varphi\rangle = \int_X g(x)|\varphi(x)|^2 d\mu(x) = \int_{Y_c} g(x)|\varphi(x)|^2 d\mu(x) \leqslant c\|\varphi\|^2,$$

and so

$$\max_{\varphi \in W_c} \frac{\langle M\varphi, \varphi\rangle}{\|\varphi\|^2} \leqslant c.$$

Now, *if*, for some value of c, the space W_c is infinite-dimensional, it follows that for any k we can find a subspace $V \subset W_c$ of dimension k, and thus by definition we get that $\lambda_k \leqslant c$ for all k. By contraposition, the assumption that (λ_k) be unbounded implies $\dim W_c < +\infty$ for all $c \geqslant 0$.

From this, we now deduce that the spectrum of M is a countable discrete subset of $[0, +\infty[$, and is unbounded (unless g is itself bounded). Let again $c \geqslant 0$ be given, and fix any finite subset $\{x_1, \ldots, x_m\} \subset \sigma(M) \cap [0, c]$. There exists $\varepsilon > 0$ such that the intervals $I_i =]x_i - \varepsilon, x_i + \varepsilon[$ are disjoint; since $\sigma(M) = \operatorname{supp} g_*(\mu)$, the definition of the support implies that $\mu(g^{-1}(I_i)) > 0$; then the characteristic functions $f_i = \chi_{g^{-1}(I_i)}$ are in W_c, and are linearly independent (because of the disjointess). Hence we get $m \leqslant \dim W_c$, which is finite; since the finite set was arbitrary, this means that $\sigma(M) \cap [0, c]$ itself must be finite, of cardinality $\leqslant \dim W_c$. So $\sigma(M)$ is a discrete unbounded set.

Now let $\lambda \in \sigma(M)$ be given; since λ is isolated, the set $Z_\lambda = g^{-1}(\{\lambda\})$ satisfies $\mu(Z_\lambda) > 0$. Then

$$\operatorname{Ker}(M - \lambda) = \{\varphi \in D(M) \mid \varphi(x) = 0 \text{ if } x \notin Z_\lambda)$$

and this is non-zero since the characteristic function of Z_λ belongs to it. Hence, we find that $\lambda \in \sigma_p(M)$.

Moreover, with the previous notation, we have $\operatorname{Ker}(M - \lambda) \subset W_\lambda$, so the λ-eigenspace must also be finite-dimensional. This means finally that M has a spectrum which can be ordered in a sequence

$$0 \leqslant \mu_1 \leqslant \mu_2 \leqslant \ldots \leqslant \mu_k \leqslant \cdots, \to +\infty$$

with finite-dimensional eigenspaces. This precisely says that M has compact resolvent, since the functional calculus implies that the bounded operator $(M+1)^{-1} = R_{-1}(M)$, for instance, is unitarily equivalent to the diagonal operator on $\ell^2(\mathbf{N})$with eigenvalues $1/(\mu_k + 1) \to 0$, which is compact (e.g., by Proposition 2.7).

This preliminary argument has the consequence that the theorem will be entirely proved in the case of $(D(T), T)$ self-adjoint once it is shown that, for T having compact resolvent, the numbers λ_k are indeed the eigenvalues, repeated with multiplicity. This is exactly parallel to the proof of the max-min principle (2.11) for compact operators (in fact, it can be deduced from that statement for the compact resolvent $(M-i)^{-1}$), and we leave the details to the reader.

Now we come back to the case where $(D(T), T)$ is only essentially self-adjoint; let $\bar{T}$ be its self-adjoint closure. Since the spectrum of a closable operator is defined as that of its closure, the only thing to prove is that

$$\lambda_k(\bar{T}) = \lambda_k(T),$$

for $k \geqslant 1$, the only subtlety being that $\lambda_k(T)$ is defined as an infimum over a *smaller* set of subspaces than $\lambda_k(\bar{T})$, so that we might have $\lambda_k(\bar{T}) < \lambda_k(T)$. Let $\varepsilon > 0$ be given, and let $V \subset D(\bar{T})$ be a subspace of dimension k with

$$\lambda_k(\bar{T}) \leqslant \max_{v \in V - \{0\}} \frac{\langle \bar{T}v, v\rangle}{\|v\|^2} \leqslant \lambda_k(\bar{T}) + \varepsilon.$$

The map

$$Q_V \quad \begin{cases} V \to [0, +, \infty[\\ v \mapsto \langle \bar{T}v, v\rangle \end{cases}$$

is a positive quadratic form on this finite dimensional space, and there exists a vector $v_1 \in V$ with norm 1 such that

$$Q_V(v_1) = \max_{v \in V - \{0\}} \frac{\langle \bar{T}v, v\rangle}{\|v\|^2} \in [\lambda_k(\bar{T}), \lambda_k(\bar{T}) + \varepsilon].$$

Complete v_1 to an orthonormal basis $(v_1, \ldots, v_k)$ of V. By definition of the closure, there exist sequences $(v_{j,n})$ in $D(T)$ such that

$$(v_{j,n}, Tv_{j,n}) \to (v_j, \bar{T}v_j), \quad \text{in} \quad H \times H,$$

for $1 \leqslant j \leqslant k$. We have by continuity

$$\langle v_{i,n}, v_{j,n}\rangle \to \langle v_i, v_j\rangle = \delta(i,j),$$

and

$$\frac{\langle Tv_{j,n}, v_{j,n}\rangle}{\|v_{j,n}\|^2} \to \frac{\langle \bar{T}v_j, v_j\rangle}{\|v_j\|^2},$$

so that if we denote by $V_n \subset D(T)$ the space generated by the $(v_{j,n})$, we see first that $\dim V_n = k$ for n large enough (the required determinant converging to that of the basis (v_j) of V, which is non-zero), and then that

$$\max_{v \in V_n - \{0\}} \frac{\langle Tv, v\rangle}{\|v\|^2} \to \frac{\langle Tv_{1,n}, v_{1,n}\rangle}{\|v_{1,n}\|^2} = Q_V(v_1),$$

so that for n large enough we obtain a subspace of $D(T)$ of dimension k with

$$\lambda_k(T) \leqslant \max_{v \in V_n - \{0\}} \frac{\langle Tv, v\rangle}{\|v\|^2} \leqslant \lambda_k(\bar{T}) + 2\varepsilon.$$

Taking $\varepsilon \to 0$, it follows that $\lambda_k(T) \leqslant \lambda_k(\bar{T})$, and this is the desired conclusion which finishes the proof of part (1) of the theorem.

(2) Now, we abandon the assumption that T is essentially self-adjoint, and consider the Friedrichs extension (D_F, T_F). Again, since $D(T) \subset D_F$, we have a priori the inequality

$$\mu_k = \lambda_k(T_F) \leqslant \lambda_k(T),$$

and must show it is an equality. But we can proceed exactly as in the last approximation argument using now the fact that for any vector $v \in D_F$, one can find a sequence (v_n) in $D(T)$ for which

$$v_n \to v, \quad \langle T v_n, v_n \rangle \to \langle T_F v, v \rangle,$$

a fact which is a consequence of the definition of D_F and T_F. □

Finally, we state the analogue of the Weyl Law of Proposition 5.13:

PROPOSITION 5.19. *Let $U \subset \mathbf{R}^n$ be a bounded open set and (D_d, Δ_d) the Laplace operator on U with Dirichlet boundary conditions. Let*

$$0 \leqslant \lambda_1 \leqslant \lambda_2 \leqslant \ldots \leqslant \lambda_n \leqslant \ldots$$

be the eigenvalues of Δ_d repeated with multiplicity. Then we have

$$\sum_{\lambda_n \leqslant X} 1 \sim c_n V \frac{X^{n/2}}{(2\pi)^n}$$

as $X \to +\infty$, where V is the Lebesgue measure of U.

SKETCH OF PROOF. □

CHAPTER 6

Applications, II: Quantum Mechanics

In this chapter, we survey some elementary and important features of the use of operators on Hilbert spaces as a foundation for Quantum Mechanics. It should be said immediately (and it will clear in what follows) that the author is *not* an expert in this field.

6.1. Postulates and foundations

Quantum Mechanics is a refinement of classical mechanics, as it was developped in particular by Newton, Euler, Lagrange, Hamilton. To indicate the parallel and striking differences, we quickly explain a bit what classical mechanical systems "look like". As we will only discuss, later on, very simple cases, we take a very simple-minded approach.

In the Newtonian description, the position $x(t) \in \mathbf{R}^3$ at time t of a punctual particle P with mass m is dictated by the second-order differential equation

$$mx''(t) = \text{sum of forces acting on } P.$$

The trajectory $x(t)$, $t \geqslant 0$, can therefore be found (in principle) if the forces are known. If there are more than a single particle, each of the trajectories satisfies the Newton equations, but of course the forces acting on the different particles will change; in fact, there will usually be *interactions* between the particles that make the whole system of equations rather complicated.

One remark that follows from the fact that the equation is of second order is that (at least intuitively!) the position $x(t)$ and the velocity $x'(t)$ at a fixed time t_0 are enough to determine the whole trajectory (and velocity) for $t \geqslant t_0$. Instead of using coordinates $(x(t), x'(t))$, it is customary to use $(x(t), p(t))$ where $p(t) = mx'(t)$ is the *momentum*. Thus the particle's state at t is determined by the pair $(x(t), p(t)) \in \mathbf{R}^6$, and any observed quantity will be a well-defined function $f(x, p)$ of $(x, p) \in \mathbf{R}^6$.

Here are three fundamental examples:

Example 6.1 (Free particle). A "free" particle is one on which no force is exerted; thus Newton's equations are simply $mx''(t) = 0$, i.e., the movement follows a straight line in $\mathbf{R}^3$, and the velocity $x'(t)$ is constant.

Example 6.2 (Attraction). If the particle is only submitted to the influence of gravitation from a body of mass 1 situated at the origin (and if $x(t) \neq 0$), the force is given by

$$-\gamma \frac{x(t)}{\|x(t)\|^3},$$

where $\gamma > 0$ is the gravitational constant (and $\|\cdot\|$ is the Euclidean distance in $\mathbf{R}^3$). Thus the equation of motion becomes

$$mx''(t) + \frac{\gamma x(t)}{\|x(t)\|^3} = 0.$$

Example 6.3 (Harmonic oscillator). If the particle is attached with a spring to the origin and forced to move on a one-dimensional axis $x(t) \in [0, +\infty[$, the spring exerts a force of the type

$$-kx(t)$$

for some constant $k > 0$, and the equation of motion becomes

$$mx''(t) + kx(t) = 0,$$

which is of course a very simple ordinary differential equation.

Classical mechanics is much richer than the simple imposition of Newton's equations as a *deus ex machina*. The Lagrangian formalism and the Hamiltonian formalism give a more fundamental a priori derivation of those equations from principles like the "principle of least action". We refer to the first chapter of [**T**] for a description of this, in particular of the wealth of mathematical concepts which arise (e.g., symplectic manifolds).

Quantum Mechanics builds on a very different formalism, which was definitely "passing strange" when it was first introduced. As we present it, we can hide the physics behind an axiomatic mathematical description that can be studied, like any set of axioms, in a type of vacuum (as far as applications or relevance to the natural world is concerned). Not surprisingly, this abstract mathematical theory is quite interesting. However, since it is a physical theory, one must not only make contact with the natural world, but also make a connection between classical mechanics and quantum mechanics. This leads to further mathematical questions of great interest.

Lacking again time and competence to discuss the origins of Quantum Mechanics, we describe the mathematical framework (basically due to J. von Neumann in 1927–1928, concluding the earlier crucial work of Heisenberg, Schrödinger, Jordan and Born) for the case of a single system P, e.g., a single electron:

- The state of P is described by a vector $\psi \in V$, where V is a (separable) Hilbert space, typically $\psi \in L^2(\mathbf{R}^3)$, such that $\|\psi\| = 1$;
- An observable (relative to the quantum system described by H) is a *self-adjoint* operator T on V, often unbounded;
- The dynamical evolution of the system, given that it is initially in the state ψ_0 and then evolves with time to be in state ψ_t, $t \geqslant 0$, is given by the *Schrödinger equation*

$$i\frac{h}{2\pi}\frac{d}{dt}\psi_t = H\psi_t \tag{6.1}$$

  where H is a self-adjoint operator (the *Hamiltonian*) on V describing the interactions and forces acting on the particle, and h is a physical constant, the *Planck constant* (in standard units, namely Joule (energy)-second (time), it is currently known to be equal to $6.62606896(33) \times 10^{-34}$ J s). One commonly writes $\hbar = \frac{h}{2\pi}$.
- Given the state ψ of the system and an observable T, one can not produce a (or speak of the) "value" of T at ψ (similar to the well-defined position or momentum of a classical particle); however, one can describe the distribution of "values" of T for the state ψ, i.e., describe a Borel probability measure $\mu_{T,\psi}$ such that $\mu_{T,\psi}(E)$ is interpreted as the "probability that the value of the observable lies in E in the state ψ". This measure is the spectral measure for T and ψ, namely it is the probability measure defined by

$$\int_{\mathbf{R}} f(x) d\mu_{T,\psi}(x) = \langle f(T)\psi, \psi\rangle$$

for f continuous and bounded, $f(T)$ given by the functional calculus of Corollary 4.43; more directly, if T is unitarily equivalent to multiplication by g on $L^2(X,\mu)$, with ψ corresponding to a function $\varphi \in L^2(X,\mu)$, we have

$$\mu_{T,\psi}(E) = \int_E g_*(|\varphi|^2 d\mu), \tag{6.2}$$

as in Example 3.13. (Note that even though T is not everywhere defined, the spectral measure is well-defined for all ψ because $f(T)$ is bounded if f is bounded).

- The last item is relevant to the physical interpretation and is not necessary for a purely mathematical description: one further expects that repeated, identical, experiments measuring the observable T will lead to values in $\mathbf{R}$ distributed according to the spectral measure of the previous item. Note an important consequence, since the support of the spectral measure $\mu_{v,T}$ is contained in the spectrum of T: *the result of any measurement is an element of the spectrum of the relevant observable.*

Notice that self-adjoint operators play two roles here: one as observables, and one as Hamiltonians in the Schrödinger equation.

6.2. Stone's theorem

Mathematically, one can take the prescriptions above as describing some questions worth studying. A first reasonable question is to understand the meaning and the solutions of Schrödinger's equation for various choices of Hamiltonian. The following important theorem makes rigorous the formal solution

$$\psi_t = e^{-\frac{i}{\hbar}tH}(\psi_0)$$

of (6.1) – which is almost immediately justified for a finite-dimensional Hilbert space –, and highlights once more how natural the condition of self-adjointness is.

Theorem 6.4 (Stone's Theorem). *Let H be a separable Hilbert space.*

(1) *If $(D(T),T)$ is a self-adjoint operator on T, define*

$$U(t) = e^{itT}, \quad t \in \mathbf{R},$$

for $t \in \mathbf{R}$, using the functional calculus. Then $U(t)$ is bounded on $D(T)$ and extends by continuity to a unitary operator on H. We have

$$U(t+s) = U(t)U(s), \qquad \text{for all } s,t \in \mathbf{R},$$

and U is strongly continuous: for any $v \in H$ and $t_0 \in \mathbf{R}$, we have

$$\lim_{t\to t_0} U(t)v = U(t_0)v. \tag{6.3}$$

Moreover, we have

$$D(T) = \{v \in H \mid t^{-1}(U(t)v - v) \text{ has a limit as } t \to 0\}, \tag{6.4}$$

and $t \mapsto U(t)v$ is a solution of the normalized Schrödinger equation, in the sense that for any $v \in D(T)$ and $t \in \mathbf{R}$, we get

$$\frac{d}{dt}U(t)v = \lim_{s\to t} \frac{U(s)v - U(t)v}{s-t} = iT(U(t)v).$$

In particular, for $v \in D(T)$, we have

$$Tv = \frac{1}{i} \lim_{t \to 0} \frac{U(t)v - v}{t}. \tag{6.5}$$

(2) *Conversely, suppose given a map*

$$U : \mathbf{R} \to L(H)$$

such that $U(t+s) = U(t)U(s)$ for all t and s, that $U(t)$ is unitary for all t, and such that U is strongly continuous. Define

$$D = \{v \in H \mid t^{-1}(U(t)v - v) \text{ has a limit as } t \to 0\}, \tag{6.6}$$

$$Tv = \frac{1}{i} \lim_{t \to 0} \frac{U(t)v - v}{t}, \quad v \in D(T). \tag{6.7}$$

Then D is dense in H, (D, T) is a self-adjoint operator on H, and we have $U(t) = e^{itT}$ for all $t \in \mathbf{R}$.

Note that the definition of the maps U in part (2) corresponds exactly to saying that U is a *unitary representation* of the topological group $(\mathbf{R}, +)$ on the Hilbert space H, as defined in Example 1.8 of Chapter 1. As in that example, we will also use the standard terminology and say that U is a one-parameter group of unitary operators.

Stone's Theorem shows that the theory of unbounded self-adjoint operators is, in some sense, "forced" on us even if only bounded (in fact, unitary) operators were of principal interest. In fact, in part (2), the condition of strong continuity can be weakened:

Proposition 6.5 (Von Neumann). *Let H be a separable Hilbert space and*

$$U : \mathbf{R} \to L(H)$$

such that $U(t+s) = U(t)U(s)$ for all t and s and $U(t)$ is unitary for all t. Then U is strongly continuous if and only if the maps

$$\begin{cases} \mathbf{R} \to \mathbf{C} \\ t \mapsto \langle U(t)v, w \rangle \end{cases}$$

are measurable for all v, $w \in H$.

Proof of Theorem 6.4. The first part is, formally, quite easy: first, having defined $U(t) = e^{itT}$ by the functional calculus (Corollary 4.43) as $f_t(T)$ where $f_t(s) = e^{its}$ for $s \in \mathbf{R}$, the multiplicative relations

$$f_{t_1} f_{t_2} = f_{t_1+t_2}, \quad f_t^{-1} = f_{-t} = \overline{f_t},$$

lead by functional calculus to

$$U(t_1)U(t_2) = U(t_1 + t_2), \quad U(t)^{-1} = U(-t) = \overline{U(t)},$$

which show that each $U(t)$ is unitary and that the group-homomorphism property holds.

To show the strong continuity (6.3), we can apply (iv) in the definition of the functional calculus (Corollary 4.43): for any sequence t_n converging to t_0, we have $f_{t_n} \to f_{t_0}$ pointwise, and of course $\|f_{t_n}\|_\infty = 1$ for all n, so we get $U(t_n) \to U(t_0)$ strongly, and because this is true for all sequences, we get more generally that $U(t) \to U(t_0)$ in the strong topology, as $t \to t_0$.

We next observe the following fact: let t_n be any sequence of real numbers converging to 0. Consider the functions

$$g_n(x) = \frac{1}{it_n}(e^{it_n x} - 1), \quad x \in \mathbf{R}.$$

The sequence (g_n) is a sequence of bounded functions on $\mathbf{R}$, and $g_n(x) \to x$ pointwise as $n \to +\infty$; moreover $|g_n(x)| \leqslant |x|$ for all n and x, as is easily checked, so that (v) in Corollary 4.43 shows that

$$\lim_{n\to+\infty} g_n(T)v = Tv$$

for all $v \in D(T)$. This translates to the existence and value of the limit

$$\lim_{t\to 0} \frac{U(t)v - v}{t} = iTv.$$

This proves

$$D(T) \subset D_1 = \Big\{ v \mid \text{the limit } \lim_{t\to 0} \frac{U(t)v - v}{t} \text{ exists} \Big\},$$

and also shows that the limit formula (6.5) holds for $v \in D(T)$. Moreover, a simple computation (like the one below for the proof of Part (2)) implies that (D_1, T_1) defined by (6.6), (6.7) is a symmetric extension of $(D(T), T)$. Since T is self-adjoint, this implies that in fact $(D_1, T_1) = (D(T), T)$. The unitarity of $U(t)$ immediately implies that $U(t)D(T) = U(t)D_1 \subset D_1 = D(T)$ and that

$$T(U(t)v) = U(t)(T(v)), \quad v \in D(T), \quad t \in \mathbf{R}.$$

Thus we get

$$\frac{d}{dt} U(t)v = \lim_{h\to 0} \frac{U(t+h)v - U(t)v}{h} = iT_1(U(t)v) = iT(U(t)v),$$

which is the normalized version of Schrödinger's equation.

We must now prove part (2) for an arbitrary map $t \mapsto U(t)$ satisfying the group-homomorphism property and the strong continuity. We first assume that D is dense in H, and we will check this property at the end. From it, it is clear that the operator T defined on D by (6.7) is a densely defined linear operator on H. We first check that it is symmetric (and hence closable): for any v, $w \in D$, and $t \in \mathbf{R}$, we have

$$\langle \frac{U(t)v - v}{it}, w \rangle = \langle v, \frac{U(-t)w - w}{-it} \rangle$$

by unitarity: $U(t)^* = U(t)^{-1} = U(-t)$. As $t \to 0$, since $-t \to 0$ also, we obtain

$$\langle Tv, w \rangle = \langle v, Tw \rangle,$$

as desired. We also observe that the definition of D and the property $U(t)U(t_0) = U(t + t_0) = U(t_0)U(t)$ gives

$$\frac{U(t)U(t_0)v - U(t_0)v}{t} = U(t_0)\Big(\frac{U(t)v - v}{t}\Big)$$

and therefore it follows that D is stable under $U(t_0)$ for any $t_0 \in \mathbf{R}$, and moreover that

$$T(U(t_0)v) = U(t_0)Tv, \qquad v \in D. \tag{6.8}$$

Now, we use the criterion of Proposition 4.30 to show that (D, T) is essentially self-adjoint. Let $w \in H$ be such that $T^*w = iw$; we look at the functions

$$\psi_v(t) = \langle U(t)v, w \rangle$$

defined for $v \in D$ and $t \in \mathbf{R}$. We claim that ψ_v is differentiable on $\mathbf{R}$, and satisfies $\psi_v' = \psi_v$. Indeed, continuity of ψ_v is immediate, and for any $t_0 \in \mathbf{R}$ and $h \neq 0$, we have

$$\frac{\psi_v(t_0 + h) - \psi_v(t_0)}{h} = \langle U(t_0)\Big(\frac{U(h)v - v}{h}\Big), w \rangle$$

and therefore, by definition of D, the limit as $h \to 0$ exists, and is equal to

$$\begin{aligned}\psi_v'(t_0) &= \langle iU(t_0)(Tv), w\rangle \\ &= \langle iT(U(t_0)v), w\rangle \qquad \text{by (6.8)} \\ &= \langle iU(t_0)v, T^*w\rangle = \langle iU(t_0)v, iw\rangle = \psi_v(t_0)\end{aligned}$$

using $T^*w = w$. It follows from this computation and the solution of the ODE $y' = y$ that for $t \in \mathbf{R}$ we have

$$\langle U(t)v, w\rangle = \psi_v(t) = \psi_v(0)e^t = \langle v, w\rangle e^t.$$

But by unitarity we also have $|\psi_v(t)| \leqslant \|v\|\|w\|$, so ψ_v is bounded, and the formula above is only compatible with this property if $\langle v, w\rangle = 0$. As this holds then for all $v \in D$, which is dense in H, we have $w = 0$. A similar reasoning shows that there is no solution to $T^*w = -iw$, and we conclude that (D, T) is essentially self-adjoint. Let $(\bar{D}, \bar{T})$ be the closure of (D, T), which is self-adjoint. We will check that $e^{it\bar{T}} = U(t)$ for $t \in \mathbf{R}$; from this, the formula (6.6) of Part (1), applied to $(\bar{D}, \bar{T})$, will imply that $\bar{D} = D$.

Define $V(t) = e^{it\bar{T}}$; this is (by Part (1)) a strongly continuous group of unitary operators on H. Let $v \in D$ be any vector; we consider the continuous map

$$g_v(t) = \|v(t)\|^2, \qquad v(t) = U(t)v - V(t)v, \quad t \in \mathbf{R}$$

and we proceed to prove that $g_v(t)$ is constant; since $g_v(0) = 0$, it will follow that $U(t) = V(t)$ since they coincide on the dense subspace D.

For any $t_0 \in \mathbf{R}$, we have

$$\begin{aligned}\frac{U(t)v - U(t_0)v}{t - t_0} &\to U(t_0)(iTv) = iT(U(t_0)v), \qquad \text{as } t \to t_0, \\ \frac{V(t)v - V(t_0)v}{t - t_0} &\to V(t_0)(iTv) = iT(V(t_0)v), \qquad \text{as } t \to t_0,\end{aligned}$$

by (6.8) and Part (1) (noting that $\bar{T}v = Tv$ since $v \in D$). Thus we get

$$\frac{v(t) - v(t_0)}{t} \to iT(U(t_0)v - V(t_0)v) = iT(v(t_0)),$$

(and all vectors for which T is applied belong to D). From this, writing

$$g_v(t) = \langle v(t), v(t)\rangle,$$

the standard differentiation trick for products gives the existence of $g_v'(t)$ with

$$g_v'(t) = \langle v(t), iT(v(t))\rangle + \langle iT(v(t)), v(t)\rangle = i(-\langle v(t), T(v(t))\rangle + \langle T(v(t)), v(t)\rangle = 0$$

by symmetry of T on D.

All this being done, we see that it only remains to check that the subspace D is indeed dense in H. This is not at all obvious, and this is where the strong continuity will be used most crucially.

The idea is to construct elements of D by *smooth averages* of vectors $U(t)v$. For this, we use the integral of Hilbert-space valued functions described in Section 1.4.3 (the reader does not necessarily need to read this before continuing; it may be helpful to first convince oneself that the expected formal properties of such an integral suffice to finish the proof).

Precisely, consider a function $\varphi \in C_c^\infty(\mathbf{R})$ and $v \in H$, and define the vector

$$v_\varphi = \int_{\mathbf{R}} \varphi(t)(U(t)v)dt \in H$$

as in Proposition 1.17: this is well-defined, since the integrand $\varphi(t)U(t)v$ is compactly supported, bounded by $|\varphi(t)|\|v\|$, and satisfies the measurability assumption of the proposition). We are going to show that the space D_1 spanned by such vectors (as φ and v vary) is dense in H, and then that $D_1 \subset D$, which will finish the proof.

First of all, the density of D_1 is easy: if $w \in H$ is orthogonal to D_1, we have

$$0 = \langle v_\varphi, w\rangle = \int_{\mathbf{R}} \varphi(t)\langle U(t)v, w\rangle dt$$

for all $\varphi \in C_c^\infty(\mathbf{R})$, by (1.11). Since $C_c^\infty(\mathbf{R})$ is dense in $L^2(\mathbf{R})$, this implies that $\langle U(t)v, w\rangle = 0$ for all $t \in \mathbf{R}$ and $v \in H$, and hence that $w = 0$.

Now we can finally check that $D_1 \subset D$. We fix v and φ. First, we have

$$U(t_0)v_\varphi = \int_{\mathbf{R}} \varphi(t)U(t_0 + t)dt = \int_{\mathbf{R}} \varphi(t - t_0)U(t)vdt$$

by (1.12) and (1.13). Thus, by linearity, we derive

$$\frac{U(t_0)v_\varphi - v_\varphi}{t} = \int_{\mathbf{R}} \frac{\varphi(t - t_0) - \varphi(t)}{t}U(t)vdt$$

As $t_0 \to 0$, with (say) $|t_0| \leqslant 1$, the integrands

$$t \mapsto \frac{\varphi(t - t_0) - \varphi(t)}{t}U(t)v$$

converge pointwise to

$$-\varphi'(t)U(t)v,$$

and they have a common compact support, say $K \subset \mathbf{R}$, and are bounded uniformly by

$$\Big\|\frac{\varphi(t - t_0) - \varphi(t)}{t}U(t)v\Big\| \leqslant \|v\| \sup_{s\in K} |\varphi'(s)|.$$

Thus we can apply the limit result (1.14) for H-valued integrals (adapted to a continuous limit), getting

$$\lim_{t_0\to 0} \int_{\mathbf{R}} \frac{\varphi(t - t_0) - \varphi(t)}{t}U(t)vdt = -\int_{\mathbf{R}} \varphi'(t)U(t)vdt = v_{-\varphi'}.$$

This means that v_φ is indeed in D, and we are done. □

Example 6.6. (1) [Baby example] The theorem is interesting even in the simple case where H is finite-dimensional, say $H = \mathbf{C}$. Then it states that any continuous group homomorphism

$$\chi : \mathbf{R} \longrightarrow U(\mathbf{C}^\times) = \{z \in \mathbf{C}^\times \mid |z| = 1\}$$

is of the form

$$\chi(t) = e^{iat}$$

for some $a \in \mathbf{R}$ (in that case T is the self-adjoint operator $z \mapsto az$ on $\mathbf{C}$). This result can also be proved more directly, but nevertheless, it is not “obvious”.

(2) [The position operator] Consider the self-adjoint operator $Q = M_x$ acting of $L^2(\mathbf{R})$. By construction, the associated unitary group is given by

$$U(t)\varphi(x) = e^{itx}\varphi(x),$$

for all $\varphi \in L^2(\mathbf{R})$ (in particular it is completely transparent how this is defined on the whole space, and not just on $D(M_x)$).

(3) [The momentum operator] Now consider the self-adjoint operator $P = i^{-1}\partial_x$ of differentiation with respect to x on $L^2(\mathbf{R})$, precisely (see Example 4.8), we consider either the closure of the essentially self-adjoint operator

$$(C_c^\infty(\mathbf{R}), \frac{1}{i}\partial_x),$$

or we define it directly by

$$D = \{\varphi \in L^2(\mathbf{R}) \mid \int_{\mathbf{R}} |x|^2 |\hat{\varphi}(x)|^2 dx < +\infty\},$$

where $\hat{\varphi} = U\varphi$ is the Fourier transform and

$$P\varphi = U^{-1} M_{2i\pi x} U\varphi.$$

We can of course compute the unitary group using this explicit multiplication representation:

$$e^{itP} = U^{-1} M_{e^{2i\pi xt}} U,$$

and since $U^{-1}\varphi(x) = \hat{\varphi}(-x)$ by the inverse Fourier transform, we get (at leat for φ smooth enough to justify the integration definition) that

$$e^{itP}\varphi(x) = \int_{\mathbf{R}} e^{2i\pi st}\hat{\varphi}(s)e^{2i\pi sx} ds = \int_{\mathbf{R}} \hat{\varphi}(s)e^{2i\pi(x+t)s} ds = \varphi(x+t).$$

By continuity, this shows that e^{itP} is simply the *translation operator* mapping φ to $x \mapsto \varphi(x+t)$. Indeed, this illustrates clearly Stone's Theorem, because it is not difficult to check *a priori* that

$$U(t)\varphi(x) = \varphi(x+t)$$

does define a one-parameter unitary group which is strongly continuous (indeed, the condition of von Neumann is quite obvious since

$$\langle U(t)\varphi, \psi \rangle = \int_{\mathbf{R}} \varphi(x+t)\overline{\psi(x)} dx,$$

which is a simple convolution operation), and it is quite natural to recover that

$$\frac{1}{i}\lim_{t\to 0} \frac{U(t)\varphi - \varphi}{t} = \frac{1}{i}\lim_{t\to 0} \frac{\varphi(\cdot + t) - \varphi(\cdot)}{t} = \frac{1}{i}\partial_x\varphi.$$

Here is yet another intuitive argument to obtain this, which also illustrates the subtlety of the functional calculus: if one tries to apply formally the power series

$$e^{itx} = \sum_{n\geqslant 0} \frac{(itx)^n}{n!},$$

one is led to the formal computation:

$$\begin{aligned} e^{itP}\varphi(x) &= \sum_{n\geqslant 0} \frac{(it)^n P^n \varphi(x)}{n!} \\ &= \sum_{n\geqslant 0} \frac{(it)^n}{n!}\Big(\frac{1}{i}\Big)^n (\partial_x\varphi)^n(x) \\ &= \sum_{n\geqslant 0} \frac{t^n}{n!}\varphi^{(n)}(x) = \varphi(x+t) \end{aligned}$$

by an (unjustified) application of a Taylor expansion formula!

Of course this can not be justified, except for the very special vectors $\varphi \in L^2(\mathbf{R})$ which represent analytic functions (such as functions $x \mapsto f(x)e^{-x^2}$, where f is a polynomial).

Physically, the correct normalization of the momentum observable is to take the operator $i\hbar\partial_x$.

Let (D, T) be a self-adjoint operator on H and $\psi \in D$ a "state" (i.e., a normalized vector, $\|\psi\| = 1$) which is an eigenfunction (one also says simply that ψ is an *eigenstate*), with $T\psi = \lambda\psi$. Then the time evolution (associated with Hamiltonian T) with initial condition ψ is particularly simple: the solution to

$$i\hbar\frac{d}{dt}\psi_t = T\psi_t, \qquad \psi_0 = \psi$$

is given by

$$\psi_t = e^{-i\lambda t/\hbar}\psi_0$$

(indeed, it is clear from the functional calculus that, for any continuous bounded function f, $f(T)\psi_0 = f(\lambda)\psi_0$).

In particular, the spectral measure μ_t associated with ψ_t is identical with the measure μ_0 associated with ψ_0: indeed, we have

$$\int_{\mathbf{R}} f(x)d\mu_t(x) = \langle f(T)\psi_t, \psi_t\rangle = \langle f(T)\psi_0, \psi_0\rangle = \int_{\mathbf{R}} f(x)d\mu_0(x).$$

This property explains why such vectors are called *stationary states*. If the Hamiltonian T has only eigenvalues (with finite multiplicities), we can find an orthonormal basis (ψ_j) of H consisting of stationary states; if we then write

$$\psi_0 = \sum_{j\geqslant 1} \langle \psi, \psi_j\rangle\psi_j$$

the solution to the Schrödinger equation with initial condition ψ_0 is given (by linearity and continuity) by

$$\psi_t = \sum_{j\geqslant 1} \langle \psi, \psi_j\rangle e^{-i\lambda_j t/\hbar}\psi_j.$$

There is a kind of converse to these assertions:

Proposition 6.7. *Let (D, T) be a self-adjoint operator on H, and let $\psi_0 \in H$ with $\|\psi_0\| = 1$ be such that $e^{itT}\psi_0 = \psi_t$ are proportional for all t. Then ψ_0 is in D and is an eigenvector for T.*

Proof. It follows from the assumption that there exists eigenvalues λ_t of $U(t) = e^{itT}$ such that

$$\psi_t = \lambda_t\psi_0.$$

Since $\lambda_t = \langle U(t)\psi_0, \psi_0\rangle$, the strong continuity and the Schödinger equation imply that

$$\lambda \begin{cases} \mathbf{R} \longrightarrow \mathbf{C}^\times \\ t \mapsto \lambda_t \end{cases}$$

is a continuous homomorphism, such that $|\lambda_t| = 1$ for all t. By the "baby case" of Stone's Theorem, there exists $a \in \mathbf{R}$ such that $\lambda_t = e^{iat}$ for all $t \in \mathbf{R}$. But then from $\psi_t = e^{iat}\psi_0$, we deduce both that $\psi_0 \in D$ and that $T\psi_0 = a\psi_0$ by Stone's Theorem. ∎

6.3. The free particle

We now discuss the quantum analogues of the classical examples described in the first section, starting with the "free particle" of Example 6.1. For a particle in $\mathbf{R}^3$ with mass $m > 0$ and no interactions, the dynamics in quantum mechanics is given by the Hamiltonian operator (D, H) where D is the domain of the Laplace operator Δ on $L^2(\mathbf{R}^3)$ (described in the previous chapter: precisely, it is the closure of the essentially self-adjoint operator $(C_0^\infty(\mathbf{R}^3), \Delta)$ described in Example 5.6) and

$$H\varphi = \frac{\hbar^2}{2m}\Delta\varphi.$$

Because we proved the self-adjointness by finding an explicit spectral representation as a multiplication operator, through the Fourier transform, we can fairly easily derive an explicit representation of the corresponding unitary group $U(t) = e^{-it\Delta/\hbar}$. More generally, in $\mathbf{R}^n$, after normalization, we get:

Proposition 6.8. *Let $n \geqslant 1$ be an integer, and let (D, Δ) be the closure of the Laplace operator $(C_c^\infty(\mathbf{R}^n), \Delta)$. For any function $\varphi \in L^1(\mathbf{R}^n) \cap L^2(\mathbf{R}^n)$, we have*

$$e^{-it\Delta}\varphi(x) = (4it)^{-n/2}\int_{\mathbf{R}^n} e^{i\|x-y\|^2/4t}\varphi(y)dy,$$

where $\sqrt{i}$ is given by $e^{i\pi/4}$ if n is odd.

Proof. Formally, this is easy: if U is the Fourier transform $L^2(\mathbf{R}^n) \rightarrow L^2(\mathbf{R}^n)$ given by (5.4), we have seen in the proof of Proposition 5.7 that

$$\Delta = U^{-1}M_{4\pi^2\|x\|^2}U,$$

and hence

$$e^{-it\Delta} = U^{-1}M_tU,$$

where M_t is the unitary multiplication operator by

$$g_t(x) = e^{-4i\pi^2t\|x\|^2}.$$

But it is well-known that the Fourier transform (and the inverse Fourier transform) exchange multiplication and convolution, in the sense that (under suitable conditions of existence) we have

$$U^{-1}(fg)(x) = \int_{\mathbf{R}^n}(U^{-1}f)(x-y)(U^{-1}g)(y)dy.$$

So we can expect that

$$e^{-it\Delta}\varphi(x) = U^{-1}(g_t(U\varphi))(x) = \int_{\mathbf{R}^n}(U^{-1}g_t)(x-y)\varphi(y)dy.$$

The functions g_t are not in $L^2(\mathbf{R}^n)$, so a priori $U^{-1}g_t$ does not make sense. However, it is well-known that if we consider

$$h_t(x) = e^{-4\pi^2t\|x\|^2}.$$

for some $t \in \mathbf{C}$ with $\mathrm{Re}(t) > 0$, we have (obviously) $h_t \in L^1 \cap L^2$ and

$$U^{-1}h_t(x) = \int_{\mathbf{R}^n} e^{-4\pi^2t\|y\|^2}e^{2i\pi\langle x,y\rangle}dy = \prod_{j=1}^{n}\int_{\mathbf{R}} e^{-4\pi^2ty^2}e^{2i\pi x_jy}dy.$$

By differentiating and then integrating by parts we see that

$$f(x) = \int_{\mathbf{R}} e^{-4\pi^2 ty^2} e^{2i\pi xy} dy$$

satisfies $f'(x) = -x(2t)^{-1} f(x)$, and so

$$\int_{\mathbf{R}} e^{-4\pi^2 ty^2} e^{2i\pi xy} dy = \Bigl(\int_{\mathbf{R}} e^{-4\pi^2 ty^2} dy\Bigr) e^{-x^2/(4t)} = (4\pi t)^{-1/2} e^{-x^2/(4t)}.$$

Formally, $g_t = h_{it}$, and this leads to

$$e^{-it\Delta}\varphi(x) = \int_{\mathbf{R}^n} (U^{-1}g_t)(x-y)\varphi(y)dy = (4\pi it)^{-n/2} \int_{\mathbf{R}^n} e^{i\|x-y\|^2/(4t)}\varphi(y)dy.$$

To justify this rigorously, the first thing to notice is that the formula certainly makes sense (i.e., is well-defined) for $\varphi \in L^1(\mathbf{R}^n) \cap L^2(\mathbf{R}^n)$, and that the above reasoning proves that

$$e^{-z\Delta}\varphi = (4\pi z)^{-n/2} \int_{\mathbf{R}^n} e^{-\|x-y\|^2/(4z)}\varphi(y)dy$$

for $\mathrm{Re}(z) > 0$. Using the strong convergence in the functional calculus, we know that if $z_n \to it$ with $\mathrm{Re}(z_n) > 0$, we have

$$e^{-z_n\Delta}\varphi \to e^{-it\Delta}\varphi$$

in $L^2(\mathbf{R}^n)$. Passing to a subsequence, it follows that for almost all x, we have

$$e^{-z_n\Delta}\varphi(x) = (4\pi z_n)^{-n/2} \int_{\mathbf{R}^n} e^{-\|x-y\|^2/(4z_n)}\varphi(y)dy \to e^{-it\Delta}\varphi(x).$$

Then using the dominated convergence theorem, we get the formula as stated. □

One may now ask in which way this formula reflects the physics of a classical particle following a uniform motion in space (i.e., following a line with constant velocity)? Some answers are given by the following examples.

Example 6.9. One can first try to "see" concretely the evolution for particularly simple initial states ψ_0. Consider $n = 1$ to simplify and an initial state for a particle of mass $m > 0$ where the spectral measure of the position observable is given by a centered Gaussian probability distribution

$$\mu_0 = |\psi_0|^2 dx = \frac{1}{\sqrt{2\pi\sigma_0^2}} e^{-\frac{x^2}{2\sigma_0}} dx,$$

for some $\sigma_0 > 0$ (so that the variance is σ_0^2). One can take for instance

$$\psi_0(x) = \frac{1}{(2\pi\sigma_0^2)^{1/4}} \exp(-\tfrac{x^2}{4\sigma_0} + ip_0 x)$$

where $p_0 \in \mathbf{R}$ is arbitrary. (Such states are called *Gaussian wave-packets*).

The Schrödinger equation (with $\hbar = 1$ for simplicity) can be solved explicitly, either by following the steps above and using the fact the the Fourier transform of a gaussian is a gaussian with explicit parameters in terms of the original one (see (1.10)). A more clever trick[1] is to represent the initial state as

$$\psi_0(x) = \frac{\sqrt{\sigma_0}}{(2\pi)^{1/4}} \exp(-\sigma_0^2 p_0^2) f(\sigma_0^2, x - 2i\sigma_0^2 p_0)$$

[1] Found by S. Blinder.

where, for $z > 0$, we put

$$f(z, y) = z^{-1/2} \exp\Big(-\frac{y^2}{4z}\Big).$$

The point of this substitution is that a direct computation reveals the identity

$$\frac{\partial^2 f}{\partial y^2} = \frac{\partial f}{\partial z}$$

Thus, the analyticity of f with respect to y and z makes permissible the formal computation

$$\begin{aligned} e^{-it/(2m)\ \Delta} f(z, y) &= \sum_{n \geqslant 0} \frac{(-it)^n}{(2m)^n n!} (-\partial_y^2 f)^n (z, y) \\ &= \sum_{n \geqslant 0} \frac{(-it)^n}{(2m)^n n!} (-\partial_z f)^n (z, y) \\ &= f\Big(z + \frac{it}{2m}, y\Big) \end{aligned}$$

and hence

$$\begin{aligned} \psi_t(x) &= \frac{\sqrt{\sigma_0}}{(2\pi)^{1/4}} \exp(-\sigma_0^2 p_0^2) f\Big(\sigma_0^2 + \frac{it}{2m}, x - 2i\sigma_0^2 p_0\Big) \\ &= \frac{\sqrt{\sigma_0}}{(2\pi)^{1/4}\sqrt{\sigma_0^2 + i(t)/(2m)}} \exp(-\sigma_0^2 p_0^2) \exp\Big(-\frac{(x - 2i\sigma_0^2 p_0)^2}{4(\sigma_0^2 + it/(2m))}\Big). \end{aligned}$$

If we now put

$$\sigma_t = \sigma_0 \sqrt{1 + \frac{t^2}{4m^2\sigma_0^4}},$$

we find after some painful computation that the probability density for (the x observable for) ψ_t is another Gaussian with density now given by

$$|\psi_t(x)|^2 = \frac{1}{\sqrt{2\pi\sigma_t^2}} \exp\Big(-\frac{(x - p_0 t/m)^2}{2\sigma_t^2}\Big).$$

The center of this gaussian has changed: it is now $x(t) = p_0 t/m$. To interpret this, we can compute the average value of the momentum observable $P = i^{-1}\partial_x$ for ψ_0. For this, we observe that the spectral measure of a state ψ for P is given by

$$d\mu_{\psi,P} = |\hat{\psi}(t/2\pi)|^2$$

and one can compute that its average is equal to p_0. Thus $x(t)$ describes the classical trajectory of a particle with mass m and initial momentum p_0.

With this classical analogy at hand, we can discuss another new feature of Quantum Mechanics: namely, although the "centers" of the Gaussians follow the classical trajectory, their variance evolves; in fact, for $|\psi_t|^2$, we have the variance

$$\sigma_t \sim \frac{t}{\sigma_0}, \quad t \to +\infty$$

which is *growing*. This means intuitively that the quantum particle's position becomes increasingly hard to determine accurately.

6.4. The uncertainty principle

Consider again a free particle in one dimension. The state vector lies in $L^2(\mathbf{R})$, and the position and momentum are "given" by the operators P and Q of multiplication by x and derivation. Note that both act on the same function, and their actions are related through Fourier transform. So the probability densities for position and momentum are not entirely free; in fact, one essentially determines the other, and although the precise relationship is not completely straightforward, the following consequence can be deduced as a striking phenomenon of quantum mechanics:

PROPOSITION 6.10 (Uncertainty relation for position and momentum). *Let $\psi \in L^2(\mathbf{R})$ be such that $\psi \in D(P) \cap D(Q)$, and $P\psi$, $Q\psi \in D(P) \cap D(Q)$, for instance $\psi \in C_c^\infty(\mathbf{R})$. Let μ_P, μ_Q be the spectral measures for ψ with respect to P and Q. Let*

$$\sigma_P^2 = \int_{\mathbf{R}} (x - p_0)^2 d\mu_P(x), \quad \sigma_Q^2 = \int_{\mathbf{R}} (x - q_0)^2 d\mu_Q(x),$$

where p_0, q_0 are the average of μ_P, and μ_Q. Then we have

$$\sigma_P^2 \sigma_Q^2 \geqslant \frac{1}{4}.$$

This is the result when $\hbar = 1$; the actual physical result is

$$\sigma_P^2 \sigma_Q^2 \geqslant \frac{\hbar^2}{4}. \tag{6.9}$$

Intuitively, σ_P measures the possible precision of a measurement of the momentum of the particle, and σ_Q the same for the measurement of the position. Thus the result is that *it is impossible to measure arbitrarily accurately both position and momentum.* On the other hand, the inequality is optimal: see (6.13) for examples where equality holds.

A more general form of the uncertainty principle applies to any pair of observables which do not commute.

THEOREM 6.11 (Heisengerg uncertainty principle). *Let H be a Hilbert space, and let $(D(T), T)$ and $(D(S), S)$ be self-adjoint operators on H. Let $v \in D(T) \cap D(S)$ be such that Tv, $Sv \in D(T) \cap D(S)$ also. Let μ_T, μ_S be the spectral measures for v with respect to T and S. Let*

$$\sigma_T^2 = \int_{\mathbf{R}} (x - t_0)^2 d\mu_T(x), \quad \sigma_S^2 = \int_{\mathbf{R}} (x - s_0)^2 d\mu_S(x),$$

where t_0, s_0 are the average of μ_T, and μ_S. Then we have

$$\sigma_T^2 \sigma_S^2 \geqslant \frac{1}{4}(\langle iw, v\rangle)^2.$$

where $w = TSv - STv = [T, S]v$.

PROOF. From the fact that $Tv \in D(T)$, one can check first that

$$\int_{\mathbf{R}} x^2 d\mu_T(x) = \langle T^2 v, v\rangle = \|Tv\|^2 < +\infty,$$

and from the fact that $v \in D(T)$, that

$$t_0 = \int_{\mathbf{R}} x\mu_T(x) = \langle Tv, v\rangle$$

(see Proposition 4.44 and the remark afterwards).

Thus one gets the expression

$$\sigma_T^2 = \int_{\mathbf{R}} x^2 \mu_T(x) - \Big(\int_{\mathbf{R}} x d\mu_T(x)\Big)^2 = \|Tv\|^2 - \langle Tv, v\rangle^2,$$

and a similar expression for σ_S^2.

Moreover, observe that if we replace T by $T - t_0$, S by $S - s_0$, the values of σ_T, σ_S do not change, the assumptions remain valid for v and these two operators, and w is also unaltered (because $[T - t_0, S - s_0] = [T, S]$). This means we may assume $t_0 = s_0 = 0$.

From this we can use the classical Cauchy-Schwarz technique: for any t, $s \in \mathbf{C}$, we have

$$\|(tT + sS)v\|^2 \geqslant 0,$$

which translates to

$$t^2\|Tv\|^2 + s^2\|Sv\|^2 + t\bar{s}\langle Tv, Sv\rangle + \bar{t}s\langle Sv, Tv \geqslant 0,$$

i.e., to

$$|t|^2\|Tv\|^2 + |s|^2\|Sv\|^2 + t\bar{s}\langle Tv, Sv\rangle + \bar{t}s\langle Sv, Tv \geqslant 0,$$

for all s, $t \in \mathbf{C}$. If we take for instance $t = 1$, $s = i\alpha$ with $\alpha \in \mathbf{R}$, it follows that

$$\alpha^2\|Sv\|^2 + \|Tv\|^2 + \alpha\langle iw, v\rangle \geqslant 0$$

for all v, and taking the discriminant gives the result. □

Since, for φ in $D(P) \cap D(Q)$ with $P\varphi$, $Q\varphi \in D(P) \cap D(Q)$, we have

$$P(Q\varphi) = \frac{1}{i}(x\varphi)' = \frac{1}{i}x\varphi' + \frac{1}{i}\varphi,$$

and

$$Q(P\varphi) = \frac{1}{i}Q\varphi',$$

we have in the case of the position and momentum operator that

$$i[P, Q]\varphi = \varphi,$$

showing that the general case indeed specializes to Proposition 6.10. (For the physical case, $P = i\hbar\partial_x$, we have instead $[P, Q] = i^{-1}\hbar$).

6.5. The harmonic oscillator

We now come to another example, the quantum-mechanical analogue of the harmonic oscillator of Example 6.3. Just as the classical case is based on imposing an extra force to a free particle on a line, the corresponding hamiltonian for quantum mechanics takes the form

$$H = \Delta + M_V,$$

where M_V is the multiplication operator by a *potential* V, which is a real-valued function on $\mathbf{R}$. Or at least, this is the case formally, and indeed many other classical systems are "quantized" in this manner, which various potentials V representing diverse physical situations. Intuitively, this should make sense from the point of view of our desired formalism, since Δ and M_V are both self-adjoint on their respective domains. However, because those domains are not usually identical (or even comparable in the sense of inclusion), it is not clear in general how to even *define* a sum of two self-adjoint operators as an unbounded operator, and then how to check that (or if) it is self-adjoint or essentially so, on a suitable domain.

Here, we may of course check that $C_c^\infty(\mathbf{R}) \subset D(\Delta) \cap D(M_V)$ *if* M_V is locally in L^2, and then of course $(C^\infty(\mathbf{R}), \Delta + M_V)$ is at least symmetric. However, it is not essentially self-adjoint in general.

EXAMPLE 6.12. This example concerns $L^2(]0, +\infty[)$ instead of $L^2(\mathbf{R})$, but it is of course still illustrative. Take

$$V(x) = \frac{c}{x^2}, \quad x > 0.$$

Then one can show (see, e.g., [**RS2**, Th. X.10]) that $\Delta + M_V$ is essentially self-adjoint on $C_c^\infty([0, +\infty[)$ if and only if $c \geqslant 3/4$.

For the harmonic oscillator with a particle of mass $m > 0$, one takes the hamiltonian

$$\frac{\hbar}{2m}\Delta + M_V, \qquad V(x) = \frac{m\omega^2 x^2}{2}.$$

The situation is then very well understood.

THEOREM 6.13. *Let (D, T) be the operator $\frac{1}{2}\Delta + M_V$, where $V(x) = \omega^2 x^2/2$, acting on $D = S(\mathbf{R})$, where the Schwartz space $S(\mathbf{R})$ is defined in Section 1.4.*

Then T is essentially self-adjoint and has compact resolvent. Its eigenvalues are given by

$$\lambda_n = (n + 1/2)\omega, \qquad n \geqslant 0,$$

and the eigenspaces are 1*-dimensional.*

More precisely, let H_n for $n \geqslant 0$ be the polynomial such that

$$H_n(x) = (-1)^n e^{x^2} \frac{d^n}{dx^n} e^{-x^2}.$$

Then, the function

$$\psi_n(x) = (2^n n!)^{-1/2} (\omega/\pi)^{1/4} H_n(\sqrt{\omega} x) \exp(-\tfrac{1}{2}\omega x^2)$$

is of norm 1 *and spans the eigenspace with eigenvalue λ_n, and $(\psi_n)_{n \geqslant 0}$ is an orthonormal basis of $L^2(\mathbf{R})$.*

Note that the use of the Schwartz space $S(\mathbf{R})$ as domain is just for convenience: ψ_n is in $S(\mathbf{R})$, but is not compactly supported.

REMARK 6.14. The first few Hermite polynomials H_n are:

$$H_0(x) = 1, \quad H_1(x) = 2x, \quad H_2(x) = 4x^2 - 2,$$
$$H_3(x) = 8x^3 - 12x, \quad H_4(x) = 16x^4 - 48x^2 + 12,$$

and they satisfy the recursion relation

$$H_{n+1}(x) = 2xH_n(x) - H_n'(x)$$

so that it is clear that H_n is of degree n and that the leading term is $2^n x^n$. Moreover, it is also clear from the definition that H_n is even if n is even, odd if n is odd.

The shapes of the graphs of ψ_n for $n = 0$ (grey), 1 (blue), 2 (red), 3 (green) and 7 (black) are displayed in Figure 1 (they correspond roughly to $\omega = 1$, up to the factor $\pi^{-1/4}$ which is omitted).

PROOF. We give a direct but not entirely enlightening proof, and will sketch afterwards a more conceptual one which explains very well the specific shape of the construction of eigenfunctions.

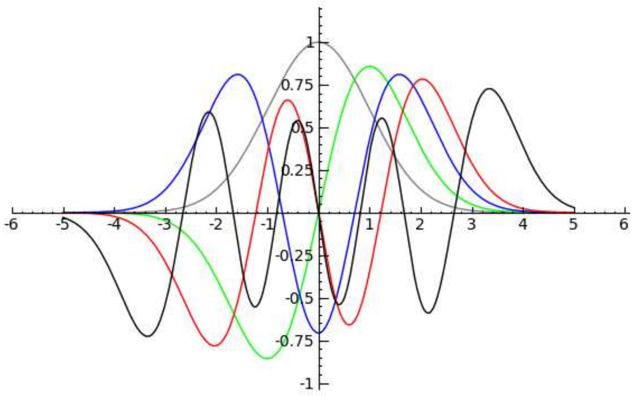

FIGURE 6.1. First stationary states

The operator T is positive on D. Given that $\psi_n \in D$, and given our claim about the ψ_n, we see that we will be done (by Lemma 5.10) if we can show that ψ_n, $n \geqslant 0$, is an eigenfunction of T, and if moreover (ψ_n) is an orthonormal basis of $L^2(\mathbf{R})$.

We start with the second part: first of all, we can assume $\omega = 1$ by rescaling. Define for the moment

$$\varphi_n(x) = H_n(x)e^{-x^2/2}.$$

Then we observe that the recurrence relation for H_{n+1} translates to

$$\varphi'_{n+1}(x) = x\varphi_n(x) - \varphi'_n(x) = T\varphi_n$$

where T is the operator $M_x - \partial_x$, densely defined on $S(\mathbf{R})$. The adjoint of T is also defined on $S(\mathbf{R})$, and satisfies $T^* = M_x + \partial_x$ on $S(\mathbf{R})$. Let then n, $m \geqslant 0$ be given, first with $n > m$. We write

$$\langle \varphi_n, \varphi_m \rangle = \langle T^n \varphi_0, \varphi_m \rangle = \langle \varphi_0, (T^*)^n \varphi_m \rangle.$$

But we also have the relation

$$T^*(p(x)e^{-x^2/2}) = (xp(x) + p'(x) - xp(x))e^{-x^2/2} = p'(x)e^{-x/2}$$

for any polynomial p, from which it follows that

$$(T^*)^n(\varphi_m) = H_m^{(n)}(x)e^{-x^2/2} = 0$$

because $n > m$ and H_m is a polynomial of degree m.

Having thus checked that (φ_n) is an orthogonal system (and it is also clear that $\|\varphi_n\|^2 > 0$), we proceed to the proof that the (Hilbert space) span is $L^2(\mathbf{R})$. For this, we notice that by Taylor expansion, we have

$$\sum_{n \geqslant 0} \frac{H_n(x)}{n!} a^n = \exp(-x^2 + 2ax) \tag{6.10}$$

for all $a \in \mathbf{R}$, where the convergence is uniform over compact subsets of $\mathbf{R}$. Then if $\varphi \in L^2(\mathbf{R})$ is orthogonal to all ψ_n, we obtain

$$\int_{\mathbf{R}} \varphi(x)e^{-x^2/2} \sum_{n \geqslant 0} \frac{H_n(x)}{n!} a^n dx = 0$$

for all $a \in \mathbf{R}$, and after completing the square, we get in particular that

$$\int_{\mathbf{R}} \varphi(x)e^{-(a-x)^2/2} dx = 0$$

for all $a \in \mathbf{R}$. Taking the Fourier transform of this convolution identity, we derive that

$$\hat{\varphi}\hat{\psi} = 0$$

where $\psi(x) = e^{-x^2/2}$. But $\hat{\psi}$, which is also a Gaussian, has no zero on $\mathbf{R}$, and thus $\varphi = 0$.

We can now check that the ψ_n are eigenfunctions of the harmonic oscillator by fairly simple computations. First, for any polynomial $p(x)$, we let

$$f(x) = p(x)e^{-\frac{1}{2}x^2},$$

and we get

$$f'(x) = p'(x)e^{-\frac{1}{2}x^2} - xp(x)e^{-\frac{1}{2}x^2},$$
$$f''(x) = e^{-\frac{1}{2}x^2}\Big\{p''(x) - 2xp'(x) - p(x) + x^2p(x)\Big\}$$

so that the eigenvalue equation $Tf = \lambda f$ becomes

$$p + 2xp' - p'' = 2\lambda p.$$

Thus our claim that ψ_n is an eigenvalue with $\lambda = \lambda_n$ becomes the fact that the Hermite polynomial H_n satisfies the second order differential equation

$$y'' - 2xy' + 2ny = 0. \tag{6.11}$$

We check this as follows (this is a standard argument about orthogonal polynomials): first, we claim that the H_n satisfy the second-order recurrence relation

$$H_n - 2xH_{n-1} = -2(n-1)H_{n-2}, \qquad n \geqslant 2. \tag{6.12}$$

Taking this for granted, the differential relation $H_{n-1} = 2xH_{n-2} - H'_{n-2}$ and its derivative can be inserted to eliminate the H_n and H_{n-1} in terms of H_{n-2} and its first and second derivatives: it leads to (6.11) for H_{n-2}.

To prove (6.12), we observe that $H_n - 2xH_{n-1}$ is a polynomial of degree at most $n-1$, since the leading terms cancel. In fact, it is of degree $\leqslant n-2$ because H_n is even and H_{n-1} is odd, so xH_{n-1} has no term of degree $n-1$. Since the H_n are of degree n, any polynomial p is a combination of those H_n with degree $\leqslant \deg(p)$, and hence we can write

$$H_n(x) - 2xH_{n-1}(x) = \sum_{0\leqslant\nu\leqslant n-2} c(\nu)H_\nu.$$

But now multiply by $\varphi_j e^{-x^2/2}$, for $j < n-2$, and integrate: by orthogonality, we get

$$c(j)\|\varphi_j\|^2 = \int_{\mathbf{R}} (H_n(x) - 2xH_{n-1}(x))H_j(x)e^{-x^2}dx$$
$$= \langle\varphi_n, \varphi_j\rangle - 2\langle\varphi_{n-1}, \tilde{\varphi}\rangle$$

where

$$\tilde{\varphi}(x) = xH_j(x)e^{-x^2/2}.$$

But $xH_j(x)e^{-x^2/2}$, for $j < n-2$, is itself a linear combination of $\psi_\nu(x)$ with $\nu < n-1$. So, by orthogonality, $\langle\varphi_{n-1}, \tilde{\varphi}\rangle = 0$, and we get

$$H_n(x) - 2xH_{n-1}(x) = c(n-2)H_{n-2},$$

which is (6.12), up to the determination of the constant $c(n-2)$. For this, we assume first n to be even; then evaluating at 0 with the help of (6.10) leads to

$$c(n-2) = \frac{H_n(0)}{H_{n-2}(0)}$$

with

$$H_n(0) = (-1)^{n/2}\frac{n!}{(n/2)!}$$

so $c(n-2) = -2(n-1)$, as desired. For n odd, we compute the derivative at 0, which satisfies the two relations

$$H_n'(0) = -H_{n+1}(0),$$
$$H_{n+2}'(0) - 2H_{n+1}(0) = c(n)H_n'(0),$$

which lead to $c(n) = 2 + H_{n+3}(0)/H_{n+1}(0) = -2(n+1)$, again as claimed.

Finally, we can use this to check the stated normalization: indeed, for $n \geqslant 2$, multiply the recurrence relation by $H_{n-2}(x)e^{-x^2}$ and integrate over $\mathbf{R}$: we get

$$\langle \varphi_n, \varphi_{n-1}\rangle - \int_{\mathbf{R}} 2xH_{n-1}(x)H_{n-2}(x)e^{-x^2}dx + 2(n-1)\|\varphi_{n-2}\|^2 = 0.$$

By orthogonality, the first term is zero. Moreover, as already observed, $2xH_{n-2} = H_{n-1} + p$ where p is of degree $< n-1$; so, again by orthogonality, we have

$$\int_{\mathbf{R}} 2xH_{n-1}(x)H_{n-2}(x)e^{-x^2}dx = \|\varphi_{n-1}\|^2,$$

and we finally obtain the relation

$$\|\varphi_{n-1}\|^2 = 2(n-1)\|\varphi_{n-2}\|^2,$$

which leads to

$$\|\varphi_n\|^2 = 2^n n!\|\varphi_0\|^2 = 2^n n! \int_{\mathbf{R}} e^{-x^2}dx = \sqrt{\pi}|2^n n!.$$

This means that $\psi_n = \varphi_n/\|\varphi_n\|$, which confirms the orthonormality of this system. □

Classically, for a particle of mass $m > 0$, the solution of

$$mx''(t) + kx(t) = 0$$

with initial conditions $x(0) = x_0$, $mx'(0) = p_0$, is typically expressed in the form

$$x(t) = A\cos(\omega t + \varphi)$$

where $\omega = \sqrt{k/m}$ and A, φ are constants determined by the initial conditions:

$$\begin{cases} x_0 &= A\cos\varphi \\ p_0 &= -A\sqrt{km}\sin\varphi. \end{cases}$$

One may note a few features of this solution: (1) the amplitude $|x(t)|$ is bounded, indeed $|x(t)| \leqslant A$ (with equality achieved for $t = (\nu\pi - \varphi)/\omega$, $\nu \in \mathbf{Z}$); (2) the energy $E(t) = \frac{1}{2}kx(t)^2/2 + \frac{1}{2}mx'(t)^2$ is *constant*: indeed, we have

$$E'(t) = kx(t)x'(t) + mx'(t)x''(t) = x'(t)(mx''(t) + kx(t)) = 0.$$

Its value is therefore given by

$$E(t) = E = \tfrac{1}{2}kx_0^2 + \tfrac{1}{2}m(p_0/m)^2 = \tfrac{1}{2}(kA^2\cos^2\varphi + m^{-1}A^2km\sin^2\varphi) = \tfrac{1}{2}kA^2$$

(it follows that the maximal amplitude can be expressed as $\sqrt{2E/k}$).

For the quantum analogue, we first renormalize the solution of the theorem to incorporate the mass of the particle and the Planck constant $\hbar$. Then, for a particle of mass m, the stationary states are given by

$$\psi_n(x) = (2^n n!)^{-1/2}\Big(\frac{m\omega}{\pi\hbar}\Big)^{1/4} H_n\Big(\sqrt{\frac{m\omega}{\hbar}}x\Big)\exp\Big(-\frac{m\omega x^2}{2\hbar}\Big)$$

for $n \geqslant 0$, which are eigenfunctions with eigenvalue $(n + 1/2)\hbar\omega$.

One can now see some striking differences between the classical and quantum oscillators, where the two cases are compared after fixing the same mass and parameter $k > 0$. The correspondance is then well-determined if we understand the quantum analogue of the energy. This is simply given by the observable which "is" the Hamiltonian – and this is natural from its expression compared with the definition of E. We can then compare a classical oscillator with energy E and a quantum one where the state ψ satisfies

$$\int_0^{+\infty} x d\mu_{\psi,H}(x) = E.$$

The simplest states are those corresponding to ψ_n; they are indeed particularly relevant physically, and we have the first striking feature that, if we observe those states, then *the energy E does not take arbitrary positive values*; the only possibilities are $E = \omega\hbar(n + 1/2)$ with $n \geqslant 0$ integers. This appearance of "discrete" energy levels, differing by integral multiples of the "quantum" $\omega\hbar$, is one of the historical reasons for the name "quantum mechanics".

Moreover, the energy has the further feature that it can not be arbitrarily small: it must always be at least equal to the lowest eigenvalue $\hbar\omega/2$. So a quantum oscillator, even at rest, carries some energy (furthermore, the state ψ_0, which is called the *ground state*, is the only state with minimal energy).

Another striking feature of the quantum particle is that it is not restricted to lie in the classical domain $|x(t)| \leqslant \sqrt{2E/k}$. This reflects the fact that the eigenfunctions ψ_n are not compactly supported: indeed, in terms of the position operator Q, which is multiplication by x as usual, the spectral measure associated to a state ψ_n is the measure

$$\mu_n = |\psi_n(x)|^2 dx = \frac{1}{2^n n!}\Big(\frac{m\omega}{\pi\hbar}\Big)^{1/2} H_n\Big(\sqrt{\frac{m\omega}{\hbar}}x\Big)^2 \exp\Big(-\frac{m\omega x^2}{\hbar}\Big),$$

in particular μ_0 (for the ground state) is the standard Gaussian with mean 0 and variance

$$\sigma_P^2 = \int_{\mathbf{R}} x^2 \mu_0(x) = \tfrac{\hbar}{2m\omega}.$$

In that case, we can compute the probability density that the particle lies outside of the classical area: we have

$$\frac{2E}{k} = \frac{\hbar}{m\omega}$$

so this is given by

$$\int_{|x|>\sqrt{\hbar(m\omega)^{-1}}} d\mu_0(x) = \frac{1}{\sqrt{\pi}}\int_{|x|>1} e^{-x^2} dx = 0.157...$$

(for the ground state).

What about the momentum observable? As already observed in the case of the free particle, the spectral measure for the oscillator in the state ψ_n with respect to the momentum observable P is obtained from the Fourier transform of ψ_n (after suitable normalization). Here we have the following lemma:

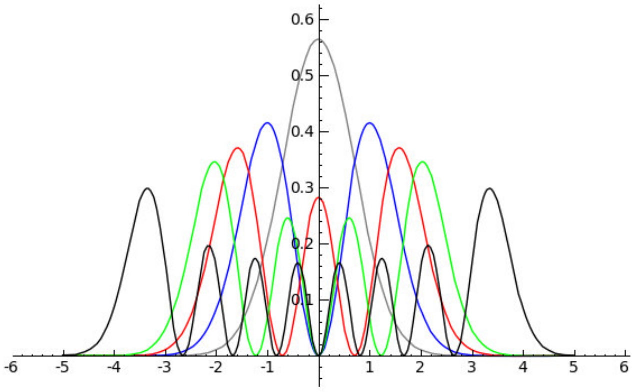

FIGURE 6.2. Densities of the first stationary states

LEMMA 6.15. *Let $n \geqslant 0$ be an integer. Let*

$$\varphi_n(x) = H_n(x)e^{-x^2/2}, \qquad \tilde{\varphi}_n(y) = \frac{1}{\sqrt{2\pi}} \int_{\mathbf{R}} \varphi_n(x)e^{-ixy}dx.$$

Then we have

$$\tilde{\varphi}_n(y) = i^n \varphi_n(y).$$

PROOF. We proceed by induction on n; for $n = 0$, $\varphi_0(x) = e^{-x^2/2}$ and we have

$$\tilde{\varphi}_0(y) = \frac{1}{\sqrt{2\pi}} \int_{\mathbf{R}} e^{-x^2/2-ixy}dx = e^{-y^2/2},$$

by the classical Fourier transform formula for a Gaussian.

Now, we have

$$T\varphi_n = \varphi_{n+1}, \qquad T^*\varphi_n = \varphi_{n-1}$$

with $T = M_x - \partial_x$, $T^* = M_x + \partial_x$, as already observed during the previous proof, and hence

$$U\varphi_{n+1} = U(T(\varphi_n)),$$

with U designating the variant of the Fourier transform in the current lemma: $\tilde{\varphi}_n = U\varphi_n$. This transform satisfies the commutation relations

$$UM_x = i\partial_y U, \qquad U\partial_x = -iM_y U,$$

on $S(\mathbf{R})$, with y denoting the "Fourier" variable. Hence we get

$$UT = (i\partial_y + iM_y)U = iT^*U,$$

and therefore

$$U\varphi_{n+1} = iT^*U\varphi_n,$$

so that if we assume by induction that $U\varphi_n = i^n\varphi_n$, we get

$$\tilde{\varphi}_{n+1} = i^{n+1}T^*\varphi_n = i^{n+1}\varphi_{n+1},$$

which concludes the proof. □

Taking into account the physical scaling, we obtain

$$\nu_n = \mu_{\psi_n,P} = \frac{1}{2^n n!}\Big(\frac{1}{\pi\hbar m\omega}\Big)^{1/2} H_n\Big(\sqrt{\frac{1}{\hbar m\omega}}x\Big)^2 \exp\Big(-\frac{x^2}{\hbar m\omega}\Big)dx$$

For $n = 0$, we find here that

$$\sigma_P^2 = \frac{m\omega\hbar}{2}.$$

Thus, in particular, we have

$$\sigma_P^2 \sigma_Q^2 = \frac{\hbar^2}{4}, \tag{6.13}$$

which is an extremal case for the Heisenberg uncertainty principle (6.9).

6.6. The interpretation of quantum mechanics

Quantum Mechanics, as we have (only partially) described it from the mathematical viewpoint leads to well-defined problems and questions about operators, spectra, and so on. We can try to solve these mathematically using Spectral Theory or other tools, as was done in the previous section for the quantum harmonic oscillator. In this respect, the situation is comparable to classical mechanics when expressed in any of the existing formalisms for it (Newtonian, Lagrangian or Hamiltonian).

However, this mathematical approach can not hide the fact that Classical and Quantum Mechanics are also supposed to be physical theories that can explain *and predict* natural phenomena in their respective domains of validity. For classical mechanics, there is little difficulty with this because the mathematical model coincide almost perfectly with the physical quantities of interest as they are intuitively grasped (e.g., for a single particle evolving in space, seen as point $(q, p) \in \mathbf{R}^6$ in phase space).

In the case of Quantum Mechanics, things are much more delicate. Observable quantities being identified with self-adjoint operators, and the state of a system with a vector in a Hilbert space from which only probability distributions about the results of experiments can be extracted, seemingly paradoxal conclusions can be reached when trying to *interpret* the results of Quantum Mechanical computations in terms of actual testable phenomena.

These difficulties have led to many controversies among the discoverers and users of Quantum Mechanics ever since M. Born gave the standard description (6.2) of the spectral measure associated with a state vector and an observable. One problem can be stated as follows: to check the validity of Quantum Mechanics, and to ascertain that it is not merely a mathematical game with arbitrary physical terminology, one must make experiments; the results of those experiments can only be some numbers (known with some finite precision, moreover), not a full-fledged Hilbert space with an operator and a state vector.

If we measure, say, the x-position of a particle, the theory says we will get a number, and – from the theory described in this chapter – the only thing we can predict is some distribution of values, which would reveal itself more fully if "identical" measurements were to be performed many times (this is a form of the "Law of Large Numbers"). One problem is that, if – almost immediately afterwards – another position measurement is made, we intuitively expect that the particle should be very close to the "random" position first measured, independently of the first result. Indeed, experiments do confirm this. But this must mean that after (or because of) the first measurement, the state vector has been altered to be, e.g., a wave packet with the spectral measure of the position observable very close to a Dirac mass located at this random position. To obtain a more complete theory, it seems necessary then to have a description of the effect of this measurement process on the state vector. This is somewhat awkward, because there might be many ways to measure a given observable, and the measuring apparatus is usually much larger (i.e., made of a gigantic number of atoms) than the particles being observed, and thus almost impossible to describe precisely in purely Quantum Mechanical terms.

Another puzzling problem of Quantum Mechanics is that of "entanglement". This problem, in a nutshell, is a consequence of the fact that for a system involving more than one particle, say two of them, denoted P_1 and P_2, the state vector may be one for which there is no way to isolate what constitutes the first and the second particle. Consider for instance systems made of two (otherwise identical) particles on a line; the natural Hilbert space in the position representation is $H = L^2(\mathbf{R}^2)$ (with Lebesgue measure) with

$$Q_1\psi(x, y) = x\psi(x, y), \quad \text{resp.} \quad Q_2\psi(x, y) = y\psi(x, y)$$

being the observable "position of P_1" (resp. "position of P_2").[2] If the state vector is of the type

$$\psi(x, y) = \psi_1(x)\psi_2(y),$$

with ψ_1, $\psi_2 \in L^2(\mathbf{R})$, each of norm 1, then the system can easily be interpreted as being two independent particles. The spectral measure corresponding to Q_1 will be the one corresponding to the position of P_1, and that of Q_2 will correspond to the position of P_2.

But there are many vectors of norm one in H which are not of this simple form, for instance something like

$$\psi(x, y) = \frac{1}{\sqrt{2}}(\psi_1(x)\psi_2(y) + \tilde{\psi}_1(x)\tilde{\psi}_2(y)),$$

where ψ_1, $\tilde{\psi}_1$ are *different* vectors of norm 1. In such a case, one can not isolate a "state" of P_1 and one of P_2 so that the global system (described by ψ) results from putting both independently together: the two particles are intrinsically entangled.

Such states are very paradoxical from a classical point of view. For instance, the "position of P_1" (as revealed by the spectral measure relative to the observable Q_1) might be like a wave packet located very far from the "position of P_2"; but any measurement which we seem to perform at this position really acts on the whole system of two particles. Since we have mentioned that some measurements seem to alter the whole state vector, it follows that in that case some physical operation at one position has some "effect" very far away...

This type of issues were raised most notably by A. Einstein in the 1930's and later, and expressed in forms leading him and his collaborators to the opinion that Quantum Mechanics (and in particular the insistence on the probabilistic interpretation of M. Born) is *incomplete*: a finer theory should exist where position and momentum are well-defined, but their usual measurements are done by averaging some type of "hidden variables", resulting in the distribution of values corresponding to the spectral measure. Einstein's playful words in a letter to M. Born in 1926 capture the idea quite well:

> "Die Quantenmechanik ist sehr Achtung gebietend. Aber eine innere Stimme sagt mir, dass das noch nicht der wahre Jakob ist. Die Theorie liefert viel, aber dem Geheimnis des Alten bringt sie uns kaum näher. Jedenfalls bin ich überzeugt, dass der Alte nicht würfelt."

which translates (rather literally) to

> "Quantum mechanics is very imposing. But an inner voice tells me, that this is not yet the real McCoy. The theory provides a lot, but it brings

[2] More abstractly, we would interpret H as the "tensor product" of the Hilbert spaces corresponding to the two particles.

us little closer to the secrets of the Old Man. At least I am certain, that the Old Man doesn't play dice."

This hypothesis of hidden variables has had varying fortunes since that time – but all indications available today seem to show they do not exist. There are various mathematical arguments (the first one going back to von Neumann; we will describe one such below) that indicate that, within the usual framework of Quantum Mechanics, hidden variables can not exist. But for physicists, of course, this is not sufficient since after all *if* a deeper theory were to exist, there is no reason it should be possible to describe it in this standard framework. However, J. Bell observed in the 1960's that intuitive assumptions on the existence of hidden variables lead in some situations to different predictions than does standard Quantum Mechanics. This led to experimental tests; and all such tests (as of today) have given results which are consistent with the probabilistic interpretation of the standard form of Quantum Mechanics.

The author has no competence to go further in describing these impressive facts, and after all this is a course on spectral theory. But to conclude, we will describe a nice and simple argument of S. Kochen and E.P. Specker [**KS**] that shows that hidden variables can not exist for a certain type of physical observable.

More precisely, they consider the (quantum analogue of) the "angular momentum" of certain systems; these are intuitively somewhat analogue of the usual angular momentum of rotating classical bodies, but have the peculiar property that the possible "values" take only finitely many values. This corresponds to the fact that the corresponding self-adjoint operators have only finitely many eigenvalues (since the result of any concrete measurement is an element of the spectrum of the observable, and isolated points in the spectrum are eigenvalues).

For the Kochen-Specker argument, one of the simplest cases is considered: systems where the angular momentum can take only the values -1, 0 and 1. However, the angular momentum can be considered with respect to any direction; for each such direction, the value may differ, but still it should be in $\{0, \pm 1\}$.

Now, crucially, Kochen and Specker consider further systems where it is possible to measure simultaneously the angular momentum for any choice of three orthogonal directions x, y, z; in spectral terms, this means that the three corresponding observables A_x, A_y, A_z commute. They do discuss a particuler physical system where this holds (a certain state of Helium; see bottom of [**KS**, p. 73]), but from the mathematical point of view, the consistency of these assumptions is ensured by writing down the following very simple self-adjoint operators acting on $\mathbf{C}^3$:

$$A_x = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & -1 \end{pmatrix}, \quad A_y = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix}$$

$$A_z = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 & -i & 0 \\ i & 0 & -i \\ 0 & i & 0 \end{pmatrix}$$

One checks easily the commutation relations as well as the fact that the three matrices have characteristic polynomial $T(T^2 - 1)$. Moreover, in this case as well as in the physical case, the "total angular momentum" (which is the observable $A = A_x^2 + A_y^2 + A_z^2$) is *always* equal to 2; indeed, one checks the matrix identity

$$A_x^2 + A_y^2 + A_z^2 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

What this means physically is that, whenever the three angular momenta are measured simultaneously for a system of this type, leading to three values a_x, a_y, a_z in $\{0, \pm 1\}$, we must have

$$a_x^2 + a_y^2 + a_z^2 = 2,$$

or in other words, the angular momentum must be 1 for two directions, and 0 for the other.

Here is now the Kochen-Specker argument: *if* there existed hidden variables h_δ described a concrete value of all the angular momenta in all directions δ (each being in $\{0, \pm 1\}$), such that the result of any experiment measuring the angular momentum in the direction δ has as result the corresponding value of h_δ, then there must exist a map

$$KS \begin{cases} \mathbf{S}^2 \to \{0, 1\} \\ x \mapsto h_x^2 \end{cases}$$

from the sphere $\mathbf{S}^2 \subset \mathbf{R}^3$ (identified with the set of directions in $\mathbf{R}^3$, up to identification of a point and its opposite) to integers 0 or 1, with the property that whenever $(x, y, z) \in \mathbf{S}^2$ are pairwise orthogonal, two of the components of $(KS(x), KS(y), KS(z))$ are 1 and one is zero, for all choices of (x, y, z). But we have the following elegant result:

Proposition 6.16 (Kochen-Specker). *There is no map $\mathbf{S}^2 \xrightarrow{f} \{0, 1\}$ with the property that $(f(x), f(y), f(z))$ always has two components equal to 1 and one equal to 0 for any triplet (x, y, z) of pairwise orthogonal vectors.*

Proof. One can, in fact, already show that there are finite subsets S of $\mathbf{S}^2$ for which no map $f : S \to \{0, 1\}$ satisfies the required condition. Kochen and Specker do this for a (somewhat complicated) set of 117 (explicit) directions. A. Peres found a much simpler and more symmetric set of 33 directions (see [**CK**, Figure 1]). Although such a proof is in fact much better as far as the physical interpretation is concerned (being indeed susceptible to a complete test), we give a proof due to R. Friedberg (see [**J**, p. 324]) which is more conceptual but involves potentially infinitely many tests.

The idea of the proof is of course to use contradiction, so we assume f to have the stated property. We will exploit the following fact: if we can show that there exists $\theta > 0$ such that, whenever two directions x, $y \in \mathbf{S}^2$ form an angle θ, we have $h(y) = 0$ if $h(x) = 0$, then it would follow that $h(x) = 0$ for *all* x, a contradiction. Indeed, the point is that for any fixed direction x_0, chosen so that $h(x_0) = 0$, and any other direction y, we can find finitely many directions

$$x_0, \quad x_1, \quad \dots, \quad x_m = y$$

where the angle between x_i and x_{i+1} is always equal to θ; thus $h(x_i) = 0$ by induction from $h(x_0) = 0$, leading to $h(y) = 0$.

So the goal is to construct the required θ. First consider a triplet (x, y, z) of pairwise orthogonal directions; observe that

$$(\frac{1}{\sqrt{2}}(x + y), \frac{1}{\sqrt{2}}(x - y), z)$$

are also pairwise orthogonal, and so are

$$(\frac{1}{\sqrt{2}}(x+z), \frac{1}{\sqrt{2}}(x-z), y).$$

From the properties of f, we deduce that it is not possible that

$$h(w) = 1 \qquad \text{for } w = \frac{1}{\sqrt{2}}(x \pm y), \quad w = \frac{1}{\sqrt{2}}(x \pm z);$$

indeed, if that were so, the two triples would allow us to conclude that $h(y) = h(z) = 0$, contradicting the pairwise orthogonality of (x, y, z).

Now we look again at the four directions

$$\frac{1}{\sqrt{2}}(x \pm y), \quad \frac{1}{\sqrt{2}}(x \pm z);$$

each of them is orthogonal to one of

$$w_1 = \frac{1}{\sqrt{3}}(x+y+z), \quad w_2 = \frac{1}{\sqrt{3}}(-x+y+z);$$

this implies that one of $h(w_1)$ and $h(w_2)$ is equal to 1, since otherwise we would obtain that the four directions w have $h(w) = 1$, which we already disproved.

Because the choice of (x, y, z) was arbitrary, it follows immediately that the angle θ_1 between the two directions w_1 and w_2 has the property that for *any two* directions d_1, d_2 with angle θ_1, it is not possible that $h(d_1) = h(d_2) = 0$.

Now we are almost done; let $\theta = \pi/2 - \theta_1$, and fix x and y with angle equal to θ, and with $h(x) = 0$. Let z be such that z is perpendicular to the plane spanned by x and y. We can moreover find w in this plane, perpendicular to y, such that the angle between w and x is θ_1. Since $h(x) = 0$, we have $h(z) = 1$ (using a third vector to complete the perpendicular vectors (x, z)). Because of the angle between w and x, we also have $h(w) = 1$. Thus the triple (w, y, z) implies that $h(y) = 0$.

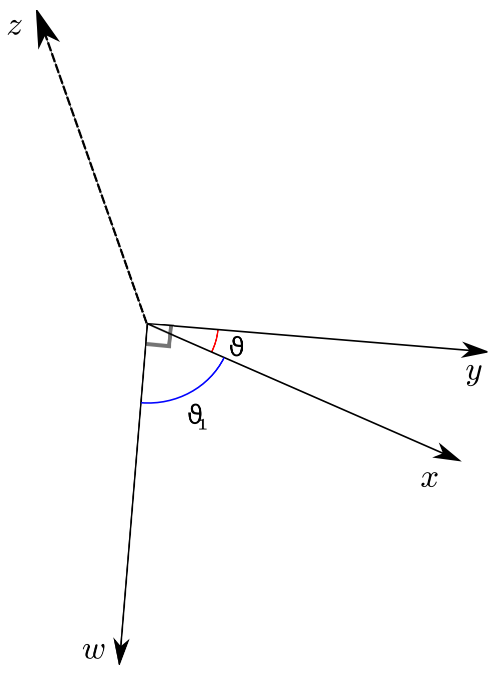

FIGURE 6.3. The last draw

This shows the existence of the angle θ claimed at the beginning of the proof, and concludes it... $\square$

Since experimental evidence is entirely on the side of the measurements of angular momentum satisfying the condition predicted by Quantum Mechanics, the Kochen-Specker argument leaves little room for the possibility of a purely deterministic refinement of it. Recently, Conway and Kochen have combined this construction with the entanglement (thought) experiments suggested by Einstein, Podolsky and Rosen to obtain a result they call the "Free Will theorem" which they interpret as saying that the behavior of particles obeying the conditions of the Kochen-Specker argument can not be predicted at all from earlier information about the state of the Universe (see [**CK**]). They even suggest that this "Free Will" property may, ultimately, play a role in the Free Will of human beings... Such a remarkable conclusion seems the best time to end this course![3]

[3] Except for a footnote: J. Conway describes the Free Will theorem and its consequences and interpretations in a series of lectures available on video at `http://www.princeton.edu/WebMedia/flash/lectures/2009_03_04_conway_free_will.shtml`.

Bibliography

[CK] J. Conway and S. Kochen: *The strong free will theorem*, Notices of the A.M.S. 56 (2009), 226–232.

[J] M. Jammer: *The philosophy of Quantum Mechanics: the interpretations of Quantum Mechanics in historical perspective*, Wiley, 1974.

[KS] S. Kochen and E.P. Specker: *The problem of hidden variables in quantum mechanics*, J. of Math. and Mech. 17 (1967), 59–87.

[W] E. Kowalski: these lecture notes, online at `https://www.math.ethz.ch/education/bachelor/lectures/fs2009/math/hilbert/`

[RS1] M. Reed and B. Simon: *Methods of modern mathematical physics, I: Functional analysis*, Academic Press 1980.

[RS2] M. Reed and B. Simon: *Methods of modern mathematical physics, II: Self-adjointness and Fourier theoretic techniques*, Academic Press 1980.

[T] L. Takhtajan: *Quantum mechanics for mathematicians*, A.M.S Grad. Studies in Math. 95, 2008.

[W] D. Werner: *Funktionalanalysis*, 6. Auflage, Springer 2007.