LECTURE NOTES FOR AN INTRODUCTORY COURSE IN ANALYSIS, M2AAG, SEPTEMBER 2023

G. David

September 2023

Preliminary comment: I fixed some mistakes in this version. I am so happy that I managed to ut it online. I'll fix more mistakes and replace soon, now that I know how to do it!

I'll try to do these notes in english (I have to take the decision before the classes start); sorry if I guessed wrong. Also, this is only my second attempt, so be sure that there will be lots of mistakes, typos, and omissions.

I have been using a lot the notes of Frédéric Paulin called "Compléments de théorie spectrale et d'analyse harmonique", still on his web page, that correspond to a course he did some years ago for the "magistère 2nde année". He had much more time, so I will need to skip many things. Said in another way, if what I say here is not clear, probably the answer to your question is in his text.

The general idea will be to explain the first results for spectral theory, for operators defined on a Banach space (often a Hilbert space). One of the goals is to be able to compute functions of the operator, as you probably did already when you used the Cayley-Hamilton theorem to compute $P(A)$ when $A \in M_n$ is a matrix and P is a polynomial.

We'll do the theory for bounded operators, but curiously our main example will be the Laplacian Δ on a domain of $\mathbb{R}^n$, which is an unbounded operator. There is a difference (sorry about that), but also many things go through, and there are ways to apply the theory of bounded operators. So the difference is not enormous. But I don't want to have to worry about domains of operators, or to give the correct definition of a self-adjoint (unbounded) operator.

We'll try to discuss how to use spectral theory (in our special case, the eigenfunctions of Δ) to do a little bit of harmonic analysis, or solve PDE's.

For the harmonic analysis, I will do some things connected to the application in mind, but this will also be a good excuse to mention results like the Hardy-Littlewood maximal theorem, the Lebesgue density theorem, and Poincaré inequalities, which I think everyone should know.

Concerning these notes, the general principle is that I try to be self-contained, and that if you do not know some of the ingredients that I use here (for instance, Banach-Alaoglu), it is probably even more important for you to check these out than the class itself, because you are more likely to need them often.

I will try to correct some mistakes as time passes and put the new one on line. Not sure exactly at which speed.

Contents

1 Compact operators

1.1 Notations and Definitions

For the moment we consider a bounded operator $u : E \to F$, where E and F are Banach spaces. That is, E and F are complete normed vector spaces, and for the moment E and F can be real or complex, we won't see the difference. For me operator means linear mapping (and I am mostly thinking of spaces E and F of infinite dimensions). I recall the definition of bounded soon below.

We'll denote by B_E (or B when E is obvious from the context) the open unit ball in E. That is, $B_E = \big\{x \in E\,;\, ||x||_E < 1\big\}$. We also denote by $\overline{B}_E = \big\{x \in E\,;\, ||x||_E < 1\big\}$ the closed ball, and often we just write $||x||$ instead of $||x||_E$.

More generally, $B_E(x,r) = \big\{y \in E\,;\, ||y-x||_E < r\big\}$ and $\overline{B}_E(x,r) = \big\{y \in E\,;\, ||y-x||_E \leq r\big\}$. I will write $B(x,r)$ instead of $B_E(x,r)$ when the context allows.

Recall that a linear mapping $u : E \to F$ is bounded when there is a constant $C \geq 0$ such that $||u(x)||_F \leq C||x||_E$ for all $x \in E$. The smallest such constant C is then called the norm of u, and denoted by $|||u|||$.

Recall that a bounded set (of E, say) is just a set which is contained in some ball $B(0,R)$ in E.

Recall that for a linear mapping $u : E \to F$, u is bounded IFF (my way of saying if and only if) the image of B_E by u is a bounded set, IFF the image of any bounded set of E by u is a bounded set, IFF u is continuous at 0, IFF u is continuous at some point $x \in E$, IFF u is continuous at every point $x \in E$.

This should be an easy exercise for you. I still recall that for instance if f is continuous at 0, then (taking $\varepsilon = 1$ in the usual definition of continuity at 0), there is $\delta > 0$ such that $||u(x)||_F \leq 1$ when $x \in B_E(0,\delta)$. Then u is bounded, with norm at most δ^{-1}, and more generally $u(A) \subset B_F(0,\delta^{-1}R)$ for any $A \subset E$ such that $A \subset B_E(0,R)$. The rest of the verification is of the same type; I skip.

Definition 1.1. *Let $u : E \to F$ be a bounded operator. We way that u is compact when $u(B_E)$ is a relatively compact subset of F, which means that its closure (in F) is compact.*

When I say compact, it is for the topology of F given with its norm. We'll rarely use weak topologies in this text. Bounded is needed here; otherwise $u(B_E)$ is not bounded, hence $u(\overline{B}_E)$ cannot be compact.

And if u is compact, $u(A)$ is relatively compact (i.e., its closure is compact) in F for every bounded set $A \subset E$; indeed $A \subset B_E(0,R)$ for some R, hence $u(A) \subset u(B_E(0,R)) = Ru(\overline{B}_E)$ is relatively compact.

Because of the Bolzano-Weierstrass theorem, we can also say that u is compact if, for any sequence $\{x_n\}$ in E, we can find a subsequence $\{x_{n_k}\}$ such that the sequence of images $\{u(x_{n_k})\}$ converges (in F). This will be very convenient.

The next proposition is an example of how easy it is fm me to fall directly from a cliff (check what I say). In my earlier version I forgot to require E is reflexive [Later I may try to check whether this could also work when it is the dual of some Banach space E_0. But let me take no further risk and try not to confuse matters too much.]

Proposition 1.2. *Let $u : E \to F$ be a bounded operator, and suppose that E is reflexive. Then $u(\overline{B}_E)$ is closed. If in addition u is compact, then $u(\overline{B}_E)$ is compact.*

Since it is clear that u is compact if $u(\overline{B}_E)$ is compact, we see that **if E is reflexive,**

(1.1) The bounded operator $u : E \to F$ is compact if and only if $u(\overline{B}_E)$ is compact.

I said Proposition because it was not as obvious as it could have been. And I'll need to add a counterexample for some case when E is not reflexive The main point is the first sentence about closure; the other one is an easy consequence because the definition says that $u(\overline{B}_E)$ is relatively compact.

The beginning of the proof is what it should be: given y in the closure of the image (we have to prove that it lies in the image), we find a sequence $\{x_n\}$ in $\overline{B}_E$ such that $y_n = u(x_n)$ converges to y.

Now we want to find $x \in \overline{B}_E$ such that $u(x) = y$. Of course E is not locally compact in general, so we cannot find a subsequence that converges in E. But we can use the fact that E is reflexive, i.e., E is the topological dual of its topological dual $F = E^*$. The **Banach-Alaoglu** theorem says that the closed unit ball of the topological dual F^* of a space F (here, we take $F = E^*$ is weakly-$*$ compact.

Here the sequence $\{x_n\}$ lies in the closed unit ball of $E = F^*$, so we can extract a subsequence, which I will denote $\{x_{n_k}\}$, that converges weak-$*$ to some $x \in E$. This means that the effect of x_{n_k} on any $\varphi \in F$ converges to $x(\varphi)$. Or, using that F is the dual of E (and the pairing is the same), that

$$\varphi(x_{n_k}) \text{ converges to } \varphi(x) \tag{1.2}$$

for every $\varphi \in F = E^*$.

Anyway, we found an $x \in E$, and this should be a good sign. We now want to check that $u(x) = y$, or equivalently, that $\psi(u(x)) = \psi(y)$ for any $\psi \in E^*$. [We'll probably use again this fact that if all $\psi \in E^*$ vanish on some $z \in E$, then $z = 0$, which we see as a consequence of the Hahn-Banach theorem.]

So we pick ψ, and apply (1.2) with $\varphi = \psi \circ u$, i.e., $\varphi(z) = \psi(u(z))$ for $z \in E$. Notice that φ is continuous because u is bounded. We get that

$$\lim_{k \to +\infty} \psi(u(x_{n_k})) = \psi(u(x)). \tag{1.3}$$

Now $u(x_{n_k}) = y_{n_k}$, which converges to y in F, so (since ψ is continuous), $\lim_{k\to+\infty} \psi(u(x_{n_k})) = \psi(y)$. Thus $\psi(u(x)) = \psi(y)$ for every ψ, hence $y = u(x)$, and this completes the proof of the proposition. $\square$

Because of my earlier lack of attention, I now need to convince that $u(\overline{B}_E)$ is not always closed. And if we trust the proposition, it is better to try this in a space like ℓ_1 or L^1. Here is, I think, an example. Consider $L^1 = L^1(\mathbb{R}, dx)$, where dx is the Lebesgue measure. Consider the intervals $I_n = [-2^{-n}, 2^{-n}]$. To each $f \in L^1$ associate the sequence $u(f) = \{c_n\} \in \ell^\infty(\mathbb{N})$, where $c_n = (1+n)^{-1} \int_{I_n} f(x)dx$. I want to check that $u(\overline{B})$ is not closed in ℓ^∞, where B is the closed ball in L^1. First the image contains any sequence $\{c_n\}$, where $c_n = (1+n)^{-1}\alpha_n$ and $\{\alpha_n\}$ is any decreasing sequence such that $0 \leq \alpha_n \leq 1$ and that tends to 0. But the closure of this set of sequences also cotains the sequence $\{(1+n)^{-1}\}$. This last one is not autorized, because it actually belongs to the Dirac mass at 0, i.e., there is no function in L^1 such that $\int_{I_n} f(x)dx = 1$ for all n.

If we only wanted the image $u(E)$ to be closed, the situation was even worse, because the inclusion of a dense subspace E of some space F is typically a counterexample. For instance the inclusion of $W^{12}(\mathbb{R})$ below in $L^2(\mathbb{R})$ is continuous with a dense image.

Some examples.

When E is of finite dimension, every linear operator is continuous (take a basis, and use the formula $u(\sum \lambda_i e_j) = \sum \lambda_i u(e_i)$, and also compact (because $\overline{B}_E$ is compact and u is continuous).

Most bounded operators $u : E \to F$ are not compact (if E and F are of infinite dimensions), and for instance the identity $I : E \to E$ is not compact because $\overline{B}_E$ is not compact.

The simplest examples of compact operators are the bounded **finite rank operators**, i.e., the bounded $u = E \to F$ such that $u(E)$ is of finite dimension. This is a little more general than the example where the dimension of E is finite, but not much.

I claim we should understand the bounded finite rank operators in a fairly brutal way. Let u be such an operator. First select a finite collection of vectors $e_i \in E$ such that the $u(e_i)$ span $u(E)$. We can take the set with the smallest possible number of elements, and then the $u(e_i)$ are independent (otherwise we can organize a more efficient set of e_i), hence are a basis of $u(E)$. We also use the boundedness of u to say that the kernel $K = \big\{x \in E\,;\, u(x) = 0\big\}$ is a closed vector subspace of E.

Next we check that E is the direct sum of K and the space L spanned by the e_i. Obviously, $u_{|L} : L \to I$ is a bijection, where we set $I = u(E) = u(L)$. Next every $x \in E$ has a decomposition $x = x_K + x_L$: we pick $x_L \in L$ with $u(x_L) = u(x)$ (it is even unique), and obviously $x_K = x - x_L$ lies in K. It is also clear that if $0 = x_K + x_L$, then $u(x_L) = 0$, then $x_L = 0$; because of this the sum is direct (the decomposition is unique).

Call π the induced projection from E to L (parallel to K). We also claim that π is continuous, as the composition of $u : E \to I$, followed by $u_{|L}^{-1} : I \to L$ (linear in finite dimensions).

Finally, u is just obtained as the composition $u_{|L} \circ \pi$ of a continuous linear projection on L and a finite-dimensional mapping.

We will see later that in Hilbert spaces the simplest example is in fact enough to get all the compact operators, by taking limits.

We'll also see other examples of compact operators soon. But let us state the stability results first.

Theorem 1.3.

- *If $u, v : E \to F$ are compact operators, then $u + v$ is compact too.*
- *If $u : E \to F$ is compact and $v : F \to G$ is bounded, then $v \circ u$ is compact.*
- *If $u : E \to F$ is bounded and $v : F \to G$ is compact, then $v \circ u$ is compact.*
- *If the sequence $\{u_k\}$ of compact operators from E to F converges for the operator norm, and if each u_k is compact, then the limit is compact too.*

I should probably have replaced the first point with the more comprehensive "the set of compact operators from E to F is a vector space, in fact a Banach space when you put the operator norm on it." And for the Banach space story, the last point of the theorem is useful.

The simplest is probably to use the criterion with sequences for each of the four items. For instance, for the third one, we start from a bounded sequence $\{x_n\}$ in E; since u is bounded, the sequence $\{u(x_n)\}$ is bounded too. And since v is compact, there is a subsequence of $\{v(u(x_n))\}$ that converges.

For the last item with limits, let $\{u_k\}$ be a norm-convergent sequence of compact operators, and choose any bounded sequence $\{x_\ell\}$ in E. Since u_1 is compact we can extract a sequence x_{ℓ_j} such that the $u_1(x_{\ell_j})$ converge. Since u_2 is compact, we can extract again, and find a new subsequence, which we shall still call x_{ℓ_j} to save notation, such that the $u_2(x_{\ell_j})$ converge too. And so on. In fact, by taking a diagonal subsequence, we even can find x_{ℓ_j} (always denoted the same way), so that $\{u_k(x_{\ell_j})\}_{j\geq 1}$ converges for every k.

Let v denote the limit of the u_k. Thus $|||v - u_k|||$ tends to 0. We want to show that $\{v(x_{\ell_j})\}_{j\geq 1}$ converges. Or that it is a Cauchy sequence. So we estimate, for j, $m \geq 0$

$$\begin{aligned} ||v(x_{\ell_j}) - v(x_{\ell_m})|| &\leq ||u_k(x_{\ell_j}) - u_k(x_{\ell_m})|| + ||v(x_{\ell_j}) - u_k(x_{\ell_j})|| + ||v(x_{\ell_m}) - u_k(x_{\ell_m})|| \\ &\leq ||u_k(x_{\ell_j}) - u_k(x_{\ell_m})|| + 2|||v - u_k|||M, \end{aligned} \tag{1.4}$$

where M is an upper bound for the $||x_\ell||$. We conclude as usual: given $\varepsilon > 0$, we can choose k so that $2|||v - u_k|||M < \varepsilon/2$ and then we use the convergence of the $u_k(x_{\ell_j})$ to prove that for j and m large enough, $||u_k(x_{\ell_j}) - u_k(x_{\ell_m})|| \leq \varepsilon/2$. □

Some things will be easier when we deal with Hilbert spaces, for instance the following.

Corollary 1.4. *Let E be a Banach space and F a Hilbert space, and let u be a bounded operator from E to F. Then u is compact if and only if u is the limit, for the operator norm, of a sequence of finite rank operators.*

Or in other words, if for each $\varepsilon > 0$, we can find a finite rank operator v such that $|||u - v||| \leq \varepsilon$.

We have seen in the theorem that limits of finite rank operators are compact, so now we take a compact operator (with values in a Hilbert space) and an $\varepsilon > 0$, and we try to find a finite rank operator at distance at most ε.

We know that $K = u(\overline{B}_E)$ is compact, and in particular it is completely bounded (in french, précompact). That is, for each $\varepsilon > 0$ there is a finite set $Y \subset K$ such the balls

$B(y, \varepsilon)$, $y \in Y$, cover K. [Otherwise, we could find an infinite sequence $\{y_j\}$ in K such that the balls $B(y_j, \varepsilon)$ are disjoint, and no subsequence of $\{y_j\}$ would ever converge.] Call L the linear span of Y (a finite-dimensional vector space in F), and π the orthogonal projection on L. Let us try the finite rank operator $v = \pi \circ u$. Let us check that $|||v - u||| \leq 2\varepsilon$; this will be enough (we could have started with $\varepsilon/2$).

Pick $x \in B_E$; we want to check that $||u(x) - \pi(u(x))|| \leq 2\varepsilon$. But $u(x) \in B(y, \varepsilon)$ for some $y \in Y$, and now $||\pi(u(x)) - u(x)|| \leq ||\pi(u(x)) - \pi(y)|| + ||\pi(y) - y|| + ||y - u(x)||$. The second term vanishes, and the other two are less than $||u(x) - y|| \leq \varepsilon$; the result follows. □

Amusing fact: the corollary is wrong for general Banach spaces, and Paulin refers to the book of Lindenstrauss and Tzafiri (Springer 79), which I think is a very nice book on the geometry of Banach spaces anyway. If wr try the proof above, it is true that we fall on a continuous projection on the space L, but its norm may go wild when L gets large, and L large is needed when ε gets small.

Yet we'll consider that compact operators are the next best thing after finite rank. If we were to compare with classes of functions, finite rank operators could correspond to polynomials, or Hermite functions, or your preferred class of functions with some definite algebraic property, and compact operators with continuous functions which tend to 0 at infinity (or something like this). Then there are more precise classes of compact operators (like Schatten classes) which would correspond to functions with some given smoothness and/or decay at infinity.

1.2 Duality: transposed operators

Now we rapidly discuss duality and transposed operators.

In this paragraph again, E and F are Banach spaces (real or complex, but we'll say complex and you can replace with real and remove the bars if you want).

The have (topological) duals E^* and F^*. Thus E^* is the (complex) Banach space composed of all the bounded linear forms $\varphi : E \to \mathbb{C}$. The norm on E^* is the operator norm, i.e., $||\varphi||_{E^*} = \sup \big\{|\varphi(x)| \,;\, x \in B_E\big\}$.

And now every bounded operator $u : E \to F$ has a unique **transposed operator** ${}^t u : F^* \to E^*$, defined by the fact that for $\psi \in F^*$, ${}^t u(\psi)$ is the element $\varphi \in E^*$ such that

$$\varphi(x) := {}^t u(\psi)(x) = \psi(u(x)) \quad \text{for every } x \in E. \tag{1.5}$$

For my personal convenience, and at the price of maybe slightly abusing notation, I tend to prefer writing this as

$$\langle {}^t u(\psi), x\rangle_{E^*,E} = \langle \psi, u(x)\rangle_{F^*,F} \tag{1.6}$$

where $\langle \cdot, \cdot \rangle$ denotes a linear form applied to a vector. Here I made it explicit by mentioning as a subscript in the bracket notation the different spaces where the duality occurs, but you can guess that the idea is to get rid of these subscript when we can. This way also, the job will be done when we switch to Hilbert spaces, where brackets will simply be scalar products.

It is standard (and we leave as an exercise) that $|||{}^t u||| = |||u|||$. Or in other words, the transposition is an isometry from the Banach space $\mathcal{L}(E, F)$ of bounded linear operators from E to F, onto the space $\mathcal{L}(F^*, E^*)$.

Recall that $E \subset E^{**}$ (with a natural inclusion). Then ${}^t({}^t u) : E^{**} \to F^{**}$ and its restriction to $E \subset E^{**}$ is equal to u. Indeed, for $x \in E$, ${}^t({}^t u)(x)$ lies in F^{**}, it is a form on F^*, and its effect on $\psi \in F^*$ is

$$\langle {}^t({}^t u)(x), \psi \rangle_{F^{**}, F^*} = \langle x, {}^t u(\psi) \rangle_{E^{**}, E^*} = \langle {}^t u(\psi), x \rangle_{E^*, E} = \langle \psi, u(x) \rangle_{F^*, F}$$

where the second part acknowledges the inclusion $E \subset (E^*)^*$, and the last one is the definition of the transpose. The last term can also be written $\langle u(x), \psi \rangle_{F^{**}, F^*}$ because $F \subset (F^*)^*$, and we get the result. Hope I did not lose myself in the notation.

Here is one last stability result.

Theorem 1.5. *(Shauder) Let $u : E \to F$ be a compact operator between Banach spaces. Then ${}^t u : F^* \to E^*$ is compact too.*

The proof may look a little ugly to you. But remember that we want to do compactness, in a context where we have sets of functions (the linear forms), so it is not shocking to use Arzela-Ascoli as below.

Set $K = u(\overline{B}_E) \subset F$; this is a compact metric space, and now call $C(K)$ the set of continuous mappings $f : K \to \mathbb{C}$ (or $\mathbb{R}$ if our Banach spaces are real). We put the sup norm (the norm of uniform convergence on K) on $C(K)$, and this makes it a Banach space too.

Now consider $\mathcal{A} \subset C(K)$, the subset of functions f that come from elements of $\overline{B}_{F^*}$. That is, $\mathcal{A}$ is the set of restrictions to $K \subset F$ of continuous linear forms on F, of norm ≤ 1. In particular, all the elements of $\mathcal{A}$ are 1-Lipschitz on K (so they are equicontinuous). Also, they take values in $\overline{B}(0, M) \subset \mathbb{C}$, with $M = |||u|||$, because $K \subset B_F(0, M)$.

The **Arzela-Ascoli Theorem** says that $\mathcal{A}$ is a relatively compact subset of $C(K)$.

We are now ready to prove that ${}^t u$ is compact. Recall it acts on F^*. Take a bounded sequence $\{\psi_k\}$ in F^*; we may as well suppose that $\psi_k \in B_{F^*}$. We extract a subsequence so that the restrictions of the ψ_{k_j} to K (they lie in $\mathcal{A}$) have a convergent subsequence in $C(K)$. By the Cauchy criterion, this means that $a_{i,j} = ||\psi_{k_j} - \psi_{k_i}||_{L^\infty(K)}$ tends to 0 when i and j tend to $+\infty$. But

$$\begin{aligned} a_{i,j} = \sup_{y \in K} |\psi_{k_j}(y) - \psi_{k_i}(y)| &= \sup_{x \in \overline{B}_E} |\psi_{k_j}(u(x)) - \psi_{k_i}(u(x))| = \sup_{x \in \overline{B}_E} |{}^t u(\psi_{k_j})(x) - {}^t u(\psi_{k_i})(x)| \\ &= \sup_{x \in \overline{B}_E} \left| [{}^t u(\psi_{k_j}) - {}^t u(\psi_{k_i})](x) \right| = |||{}^t u(\psi_{k_j}) - {}^t u(\psi_{k_i}||| \end{aligned}$$

and since $a_{i,j}$ tends to 0, we get that the ${}^t u(\psi_{k_j})$ converge in E^*, as needed. $\square$

Comment: the converse is also true: if $u \in \mathcal{L}(E, F)$ and ${}^t u$ is compact, then u is compact as well. We already know from the theorem that $\widetilde{u} = {}^t({}^t u)$ is compact, and we have seen that the restriction of $\widetilde{u}$ to E is u; we get that the image $u(\overline{B}_E)$ is relatively compact in $B_{F^{**}}$. Now $u(\overline{B}_E) \subset F$, and F is closed in F^{**} (the image of F by an isometry), so $u(\overline{B}_E)$ is also relatively compact in B_F, as needed.

1.3 First spectral properties of compact operators

Now we consider a complex Banach space, and compact operators from E to E.

Set $\mathcal{L}(E) = \mathcal{L}(E, E)$ for convenience (bounded endomorphisms). We start with the definition of the spectrum of $u \in \mathcal{L}(E, E)$, then we go as fast as we can to simple spectral properties of compact operators. Later we return to the spectrum and the resolvent. In what follows, I (or some times I_E denotes the identity operator $E \to E$.

Definition 1.6. *For $u \in \mathcal{L}(E)$, the* **spectrum** *of u, denoted by $Sp(u)$, is the set of $\lambda \in \mathbb{C}$ such that $\lambda I - u$ is not invertible. The resolvent of u is the mapping $R_u : \mathbb{C} \setminus Sp(u) \to \mathcal{L}(E)$ defined by $R_u(\lambda) = (u - \lambda I)^{-1}$.*

We put the definition of R_u because it often goes with the spectrum, but we'll not use it soon. We will comment a little bit on the spectrum, then say something for compact operators, then return to the spectrum and resolvent in the following subsections.

When E is of <u>finite dimension</u>, $Sp(u)$ is the set of eigenvalues of u, because if $u - \lambda I$ is not invertible, is neither injective nor surjective, and in particular it has a nontrivial kernel. Hence in that case $Sp(u)$ is the set of roots of the characteristic polynomial $P(\lambda) = \det(\lambda I - u)$.

Simple examples. In infinite dimensions, things are more complicated. We still have a theorem of Banach that says that if $\lambda I - u$ is bijective, then its inverse $u^{-1} : E \to E$ is bounded as well. So there is no confusion in the definition: invertible means that u is bijective, and it also means that there is a bounded operator $u^{-1} : E \to E$ such that $u^{-1} \circ u = u \circ u^{-1} = I$.

1. But u (or $u - \lambda I$) can be injective but not surjective: in the Hilbert space

$$\ell^2 = \ell^2(\mathbb{N}) = \big\{ \{x_n\}_{n \geq 0} \, ; \, \sum_n |x_n|^2 < +\infty \big\} \tag{1.7}$$

the **direct shift** S is an example. It is given by $S(x) = y$, where $y_0 = 0$ and $y_n = x_{n-1}$ when $n \geq 1$. Here and below, the notation is that $x = \{x_n\}$ and $y = \{y_n\}$. It is an isometry, but its image is the proper closed subset $\big\{ \{x_n\} \in \ell^2 \, ; \, x_0 = 0 \big\}$.

For $u = S$, $\lambda = 0$ lies in the spectrum, but it is not an eigenvalue, since u has no kernel.

2. It can also happen that u (or $u - \lambda I$) is surjective, but not injective. This is even easier to realize, but anyway the **backwards shift** T is an example, where T is defined by $T(x) = y$, where $y_n = x_{n+1}$ for $n \geq 0$.

3. Our favorite diagonal example in $\ell^2 = \ell^2(\mathbb{N})$. The operator S is brutally non surjective, but it can also happen that u is injective, not surjective, but its image is dense. Denote by $\{e_n\}_{n \geq 0}$ the usual orthonormal basis of ℓ^2. Thus e_n has only one non-zero component, which is 1 standing at the n-th coordinate.

Let $\Lambda = \{\lambda_n\}_{n \geq 0}$ be a bounded sequence, and define u_Λ by

$$u_\Lambda(e_n) = \lambda_n e_n. \tag{1.8}$$

In other words, $u_\Lambda(x) = y$, where $x = \{x_n\}$, $y = \{n_n\}$, and $y_n = \lambda_n x_n$ for all n.

If none of the λ_k is zero, u_Λ is injective and in addition the image of u is dense. However, if λ_n tends to 0, or even if it is not bounded from below, the following things happen. First, u cannot be invertible, because if it were we would have $u^{-1}(e_n) = \lambda_n^{-1} e_n$ (recall there is no choice because u is injective); this does not define a bounded mapping u^{-1}. In addition, the image $u(\ell^2)$ is not closed, because we know it is dense, and if it were equal to ℓ^2 our mapping would be bijective. It is not hard either to find points $x = \{x_n\}$ that do not lie in the image: just make sure that $\sum_n \lambda_n^{-2} x_n^2$ diverges. Finally, we have no bound $c > 0$ such that $||u(x)|| \geq c||x||$ for $x \in \ell^2$, i.e., the injectivity does not come with bounds.

All these bad properties are related: if we could find $c > 0$ such that $||u(x)|| \geq c||x||$ for $x \in \ell^2$, we would get that $u(\ell^2)$ is closed and $u : \ell^2 \to u(\ell^2)$ is invertible, and would rather be as in our first example of shift.

Exercise. Check that the operator $u_\Lambda : \ell^2 \to \ell^2$ is compact in and only if $\lim_{n\to+\infty} \lambda_n = 0$. Notice that this is not shocking, here E is a Hilbert space, and we have a nice way to approximate u_Λ by finite rank operators.

We'll return to the spectrum later, but let us state the promised simple spectral properties of compact operators.

Theorem 1.7. *Let $u : E \to E$ be a compact operator. Then*

- *The kernel of $I - u$ has a finite dimension;*
- *The image of $I - u$ is closed;*
- *If $I - u$ is injective, then it is surjective (and $I - u$ is invertible).*

Denote by N the kernel of $v = I - u$. If $x \in N$, then $u(x) = x$ and so $N \cap \overline{B}_E \subset u(\overline{B}_E)$. Thus $N \cap \overline{B}_E$ is a compact set (we don't need Proposition 1.2 for this, we know that N is closed because u is continuous). This forces N to be finite-dimensional.

Incidentally, since we shall be doing this sort of argument repeatedly, let us check this fact that locally compact implies finite dimensional. Suppose E is not finite dimensional, and define by induction a subspace V_n of dimension n and a vector $y_n \in E \setminus V_n$ (so that we can continue with the space V_{n+1} spanned by y_n and V_n). First, V_n is closed (because it is finite-dimensional; use a basis to find a bijection ψ_n from V_n to $\mathbb{C}^n$, then notice that ψ_n and ψ_n^{-1} are continuous because of the equivalence of norms, and complete the argument as you like). So $d_n = \operatorname{dist}(y_n, V_n) > 0$. Pick a point z_n of V_n such that $||y_n - z_n|| = d_n$ (or is even close to d_n). Set $e_n = (y_n - z_n)/||y_n - z_n||$. This new point of V_{n+1} has a unit norm, and the distance from e_n to V_n is $\operatorname{dist}(e_n, V_n) \sim d_n^{-1} \operatorname{dist}(y_n - z_n, V_n) = d_n^{-1} \operatorname{dist}(y_n, V_n) = 1$ because $z_n \in V_n$. Now we cannot extract a converging subsequence of the e_n, because each one lies at distance $\geq 1/2$ from the space V_n that contain the previous ones.

We'll use this sort of argument repeatedly. Some times it is confusing because it looks like we project y_n on V_n, but we know that maybe there is no projection on V_n of norm 1 (or with uniform bounds in its norm), and yet we manage. $\square$ for the subresult

Next we show that L, the image of $v = I - u$, is closed. Let $w \in \overline{L}$ be given; we want to show that $w \in L$, and we may assume that $w \neq 0$. Let $\{y_k\}$ be a sequence in L that converges to w; by definition we may write

$$y_k = x_k - u(x_k) = v(x_k), \tag{1.9}$$

for some $x_k \in E$, and where we set $v = I - u$.

A little bit of psychology first. First observe that there is usually more than one solution to (1.9). In fact, if x'_k is another solution of (1.9), then $v(x'_k) - v(x_k) = 0$, hence $x'_k - x_k \in N$. So the set of solutions to (1.9) is just our initial point x_k, plus the vector space N.

We may want to choose x_k correctly (or if we did not do that, replace it with another solution), because if we can make the sequence $\{x_k\}$ converge to some limit x, then we will get that $y_k = v(x_k)$ converges to $y = v(x)$, which will prove that y lies in the image of $v = I - u$, as desired. Of course we could add huge elements of N to the x_k and ruin everything.

Pick any x_k as above; we may have to replace it with another point. Set $d_k = \operatorname{dist}(x_k, N) = \inf_{z \in N} ||z - x_k||$. Since N is a vector space of finite dimension, we know that we can find $z_k \in N$ such that $d_k = ||x_k - z_k||$. [Hint: we can restrict our attention to the set $K = \big\{z \in N\,;\, ||z - x_k|| \leq d_k + 1\big\}$, and then use the continuity of the distance and the compactness of K.] We claim that using the point $x'_k = x_k - z_k$ may be more clever, but let us first get rid of an unpleasant case.

Suppose $d_k = ||x_k - z_k||$ tends to $+\infty$. Set $w_k = d_k^{-1}(x_k - z_k) = \frac{x_k - z_k}{||x_k - z_k||}$; this is a unit vector, and by definition of a compact operator, we can replace $\{w_k\}$ with a subsequence so that $\{u(w_k)\}$ converges to some limit ξ. As "usual", we now replace sequences with subsequences without even changing their names: we should be writing $u(w_{k_j})$ but we won't. Notice that

$$w_k - u(w_k) = v(w_k) = d_k^{-1} v(x_k - z_k) = d_k^{-1} v(x_k)$$

because $z_k \in N$. But $v(x_k) = y_k$ has a limit by construction, hence $w_k - u(w_k)$ tends to 0. Recall that $u(w_k)$ tends to ξ, so w_k tends to ξ too, and now we also get that $u(w_k)$ tends to $u(\xi)$, so $u(\xi) = \xi$ and $\xi \in N$.

Yet $w_k = d_k^{-1}(x_k - z_k)$ so $\operatorname{dist}(w_k, N) = d_k^{-1} \operatorname{dist}(x_k - z_k, N) = d_k^{-1} \operatorname{dist}(x_k, N + z_k) = 1$ because N is a vector space and $z_k \in N$. This is impossible because w_k tends to $\xi \in N$.

So $d_k = ||x_k - z_k||$ does not tend to $+\infty$, and by replacing $\{x_k\}$ with a subsequence we may assume that d_k stays bounded. Since u is compact, we can extract again so that $u(x_k - z_k)$ has a limit ζ. But by (1.9),

$$y_k = v(x_k) = v(x_k - z_k) = (x_k - z_k) - u(x_k - z_k). \tag{1.10}$$

Now y_k tends to y and $u(x_k - z_k)$ tends to ζ; so $x_k - z_k$ tends to $y + \zeta$, and now $v(x_k - z_k)$ tends to $v(y + \zeta)$. By the first part of (1.10), $y = v(y + \zeta)$ and hence $y \in v(E)$. This proves that $v(E)$ is closed, as needed for the second part.

Now we suppose that v is injective and prove it is surjective. Suppose not. Set $E_0 = E$ and $E_1 = v(E_0)$; by assumption E_1 is strictly smaller than E_0. We proved that it is closed.

Next we want to show that $u_{|E_1}$ satisfies the assumptions of the theorem. This first means that $u(E_1) \subset E_1$. And indeed, if $x \in E_1$, $u(x) = (I - v)(x) = x - v(x)$ lies in E_1. In addition,

$u(B_E \cap E_1)$ is contained in the compact set $\overline{u(B_E)}$, which is also contained in E_1 (because it is closed). So it is a compact subset of E_1, as needed.

We also claim that $E_2 = v(E_1)$, which is closed for the same reason as E_1 was, is strictly contained in E_1. Suppose for a minute that $E_2 = E_1$. Consider $x \in E_0 \setminus E_1$; since $v(x) \in E_1 = E_2$, we can find $y \in E_1$ such that $v(y) = v(x)$, contradicting the injectivity of v. So E_2 is strictly contained in E_1.

At this point we can iterate: the mappings u and v preserve E_2, the restriction of u to E_2 is compact, and $v_{|E_2}$ is injective. By induction, there is a strictly decreasing sequence of nested closed sets $E_n = v(E_{n-1}) \subsetneq E_{n-1}$. We want to show that this cannot happen.

For each integer n, choose $x_n \in E_n \setminus E_{n+1}$. Notice that $\operatorname{dist}(x_n, E_{n+1}) > 0$ (because E_{n+1} is closed), so we can find $y_k \in E_{n+1}$ such that $||y_n - x_n|| \leq 2 \operatorname{dist}(x_n, E_{n+1})$. Then set $z_n = x_n - y_n$ and $\widetilde{z}_n = z_n/||z_n||$.

Notice that for $m < n$, we have $\zeta := \widetilde{z}_m - v(\widetilde{z}_m) + v(\widetilde{z}_n) \in E_{n+1}$ just because $E_{n+1} = v(E_n)$ and then this set is preserved by u and v. Then

$$\begin{aligned}||u(\widetilde{z}_n) - u(\widetilde{z}_m)|| &= ||[\widetilde{z}_n - v(\widetilde{z}_n)] - [\widetilde{z}_m - v(\widetilde{z}_m)]|| = ||\widetilde{z}_n - \zeta|| \\ &\geq \operatorname{dist}(\widetilde{z}_n, E_{n+1}) = ||z_n||^{-1} \operatorname{dist}(z_n, E_{n+1})\end{aligned}$$

because $u = I - v$, and by definitions. Now $\operatorname{dist}(z_n, E_{n+1}) = \operatorname{dist}(x_n - y_n, E_{n+1}) = \operatorname{dist}(x_n, E_{n+1})$ because $y_n \in E_{n+1}$, so finally $||u(\widetilde{z}_n) - u(\widetilde{z}_m)|| \geq 1/2$. Then there is no way we can extract a converging subsequence from $\{u(\widetilde{z}_n)\}$, which is a contradiction because the $\widetilde{z}_n$ lie on the unit sphere and u is compact. This contradiction shows that v was in fact surjective, and this completes the proof of the theorem. □

I can't resist commenting on the proof. It will happen often in these early results that each step of the proof is easy and the whole proof is a little confusing. It could be that I am saying this because I am not used to the story, but a priori I want to blame it on the unfair strength and mystery of elementary algebra.

Next we check properties of compact operators more directly related to the spectrum.

Theorem 1.8. *Let $u : E \to E$ be a compact operator, and assume E is a Banach space of infinite dimension* <u>*over $\mathbb{C}$*</u> *(in finite dimension, we already know enough). Then*

- $0 \in Sp(u)$ *(but it is not always an eigenvalue)*
- *If $\lambda \in Sp(u) \setminus \{0\}$, then it is an eigenvalue, and the kernel of $u - \lambda I$ is finite-dimensional*
- *Each $\lambda \in Sp(u) \setminus \{0\}$ is an isolated point in $Sp(u)$*
- *$Sp(u)$ is a compact subset of $\mathbb{C}$.*

Thus either $Sp(u)$ is finite, or it is composed of a sequence (of eigenvalues) in $\mathbb{R} \setminus \{0\}$ that tends to 0, plus 0 (we will see soon that $Sp(u)$ is closed).

Of course, many of the properties above are no longer true with general bounded operators. For instance, the operators u_Λ, where Λ is a bounded sequence, ot the operators in L^2 of multiplication by a function, already give some examples.

The simplest example of u_Λ on ℓ^2 (where $\lim_{n\to+\infty}\lambda_n = 0$ because u is compact) gives a realistic idea of what u is doing spectrally. Of course here we do not say that there is an orthonormal basis of eigenvectors; we do not even have a scalar product and even in Hilbert spaces this is wrong in general.

The case when $u = u_\Lambda$ and $\lambda_n \neq 0$ for all n shows that 0 can be in the spectrum without being an eigenvalue.

If E is a Banach space over the reals, we cannot diagonalize so easily, but we can always consider the complex extension of E and the complex extension $\widetilde{u}$ of u (defined by $\widetilde{u}(x+iy) = u(x) + iu(y)$), which is compact too, and apply the theorem to $\widetilde{u}$.

Let us prove all this. Suppose $0 \notin Sp(u)$. Then u is invertible, and $B_E = u(u^{-1}(B_E)) \subset u(B(0,M))$ for $M = |||u^{-1}|||$. This forces B_E to be relatively compact (in E), which is wrong in infinite dimension. So $0 \in Sp(u)$.

Next let $\lambda \in Sp(u) \setminus \{0\}$, and assume that it is not an eigenvalue. Apply the previous proposition to $\lambda^{-1}u$. Since the kernel of $I - \lambda^{-1}u$ is $\{0\}$, $I - \lambda^{-1}u$ is injective, so it is invertible. Then $\lambda I - u$ is invertible too, a contradiction.

So $\lambda \in Sp(u) \setminus \{0\}$ is an eigenvalue, and Theorem 1.7 says that the kernel of $I - \lambda^{-1}u$ (which is also the kernel of $\lambda I - u$) is finite-dimensional.

Now we need to check that λ is isolated in $Sp(u)$. Suppose not; then there is an injective sequence $\{\lambda_n\}$ in $Sp(u)$ that converges to λ. We have seen that λ_n is an eigenvalue, so we can find $e_n \in E$, of norm 1, such that $u(e_n) = \lambda_n e_n$. As before, we want to construct a sequence of unit vectors whose images by u are far from each other, so we modify the e_n.

Call E_n the vector space spanned by $e_1, \dots, e_n$. We need to know that

(1.11) eigenvectors associated to different eigenvalues are always independent,

so let us check this (standard proof, sorry). Let $e_1, ..., e_n$ be eigenvectors associated to the different eigenvectors $\lambda_1, ..., \lambda_n$. Suppose $e_1, ..., e_{n-1}$ are independent, but $e_n = \sum_{j<n} \mu_j e_j$; we want to check that this is impossible. Apply u; this gives $\lambda_n e_n = \sum_{j<n} \mu_j \lambda_j e_j$. But also $\lambda_n e_n = \sum_{j<n} \lambda_n \mu_j e_j$. By independence, $\lambda_n \mu_j = \lambda_j \mu_j$ for all $j < n$; hence $\mu_j = 0$ for all j, a contradiction which proves (1.11).

Now $\{E_n\}$ is a strictly increasing sequence of subspaces (by (1.11)), which are closed because they are finite-dimensional.

Take $e'_1 = 0$ and, for $n \geq 2$ choose $e'_n \in E_{n-1}$ such that $\operatorname{dist}(e_n, E_{n-1}) \leq \operatorname{dist}(e_n, e'_n) \leq 2\operatorname{dist}(e_n, E_{n-1})$ (notice that $\operatorname{dist}(e_n, E_{n-1}) > 0$). Set $f_n = e_n - e'_n$ and $\widetilde{f_n} = f_n/||f_n||$.

Pick $m > n$ and consider $\zeta = \frac{e'_n}{\lambda_n ||f_n||} + \frac{\widetilde{f_m}}{\lambda_m}$. By construction, $u(\zeta) \in E_{n-1}$, so

$$
\begin{aligned}
||u(\frac{\widetilde{f_n}}{\lambda_n}) - u(\frac{\widetilde{f_m}}{\lambda_m})|| &= ||u(\frac{e_n}{\lambda_n ||f_n||} - \frac{e'_n}{\lambda_n ||f_n||} - \frac{\widetilde{f_m}}{\lambda_m})|| = ||u(\frac{e_n}{\lambda_n ||f_n||}) - u(\zeta)|| \\
&= ||\frac{e_n}{||f_n||} - u(\zeta)|| \geq \operatorname{dist}(\frac{e_n}{||f_n||}, E_{n-1}) = \frac{1}{||f_n||}\operatorname{dist}(e_n, E_{n-1}) \geq \frac{1}{2}
\end{aligned} \tag{1.12}
$$

Now the sequence $\{\frac{\widetilde{f_n}}{\lambda_n}\}$ lies in a fixed ball of E (recall that $\lim_{n\to+\infty}\lambda_n = \lambda > 0$), so, since

u is compact, we should be able to extract a convergent subsequence from $\{u(\frac{\tilde{f}_n}{\lambda_n})\}$; this is clearly impossible by (1.12); this contradiction proves that λ is isolated.

This almost concludes the proof of Theorem1.8; we still need to check that $Sp(u)$ is compact; we'll see a general easy proof for bounded operators soon; of course we could deduce "closed" from the points above, but let us not even bother. □

2 More about the spectrum in general

We start with a few additional definitions concerning the spectrum of a bounded operator on the Banach space E. It will be more convenient to assume immediately that E is a complex Banach space; in the real case, the situation is not bad anyway: we can always extend E and u to be complex, use the results below on the extension, and maybe return to u if needed.

2.1 The resolvent

Definition 2.1. *Let $u \in \mathcal{L}(E)$, where E is a complex Banach space.*

A **regular value** *of u is a $\lambda \in \mathbb{C}$ such that $u - \lambda I$ is invertible. Thus the set of regular values is $\mathbb{C} \setminus Sp(u)$. We'll see soon that it is an open set and $Sp(u)$ is compact.*

The **resolvent** *of u is the mapping $R_u : \mathbb{C} \setminus Sp(u) \to \mathcal{L}(E)$ defined by*

$$R_u(\lambda) = (u - \lambda I)^{-1} \quad \text{for } \lambda \in \mathbb{C} \setminus Sp(u). \tag{2.1}$$

The resolvent looks complicated (how do you compute it exactly?), but is is nonetheless very useful, for instance to compute functions of u (we will probably not do this here, but there are formulas for computing $f(u)$ by integrating $R_u(\lambda)d\lambda$ on a path against some appropriate function). And at least, on its domain, R_u is a nice function. The main result here is the analyticity of R_u on $\mathbb{C} \setminus Sp(u)$.

Theorem 2.2. *Let E be a Banach space over $\mathbb{C}$ and $u \in \mathcal{L}(E)$. The spectrum $Sp(u)$ is a nonempty compact subset of $\mathbb{C}$ and $R_u : \mathbb{C} \setminus Sp(u) \to \mathcal{L}(E)$ is analytic.*

Maybe you are surprised by the notion of analytic functions f with values in the Banach space $F = \mathcal{L}(E)$, but this is not too complicated.

One way to feel better about this notion, with a low cost, is to say that f is analytic as soon as, for any linear form Φ on F, the mapping $\Phi \circ f$ (with values in $\mathbb{C}$) is analytic. But we do not need to worry, the weak definition of analytic given above is equivalent to a stronger one, where we say that f is given, near every point of its (open) domain of definition, by a normally convergent power series. This is not too hard to prove, because (if f is continuous, say) we can use the Cauchy formula, first on $\Phi \circ f$, to estimate the size of all the derivatives of any $\Phi \circ f$. This is then enough to prove the existence of the local description by convergent power series (and even with precise estimates on the size of the terms). We don't need any of this for the moment because in our situation we'll get the power series description directly for R_u.

Anyway, let $u \in \mathcal{L}(E)$ and $\lambda_0 \in \mathbb{C} \setminus Sp(u)$ be given. We want to show that there exists a small $r > 0$ such that $B(\lambda_0, r)$ does not meet the spectrum and $R_u(\lambda)$ is given by a converging power series on $B(\lambda_0, r)$.

We want to reduce to the following basic case where the inverse is given by a Neumann series.

Lemma 2.3. *Let $v \in \mathcal{L}(E)$ be such that $|||v||| < 1$. Then $I - v$ is invertible, and its inverse is given by the series*

$$(I - v)^{-1} = \sum_{n \geq 1} v^n \tag{2.2}$$

where the series converges for the operator norm.

The proof is easy: the series converges normally because $|||v^n||| \leq |||v|||^n$; then the fact that $(I - v)\Big(\sum_{n \geq 1} v^n\Big) = \Big(\sum_{n \geq 1} v^n\Big)(I - v) = I$ is proved by simple manipulations on norm-convergent power series. □

Return to $\lambda_0 \in \mathbb{C} \setminus Sp(u)$. Set $v_0 = (u - \lambda_0 I)^{-1}$, which exists by assumption, and try $r = |||v_0|||^{-1}$. Let $\lambda \in B(\lambda_0, r)$; we want to invert $u - \lambda I$, so we write

$$(u - \lambda I)v_0 = (u - \lambda_0 I)v_0 - (\lambda - \lambda_0)v_0 = I - (\lambda - \lambda_0)v_0. \tag{2.3}$$

We are lucky: $v = (\lambda - \lambda_0)v_0$ has a norm $|||v||| \leq (\lambda - \lambda_0)|||v_0||| < 1$, so we can invert $I - v$ and $[(u - \lambda I)v_0]^{-1} = (I - v)^{-1} = \sum_{n \geq 0} (\lambda - \lambda_0)^n v_0^n$. Then of course $u - \lambda I = [(u - \lambda I)v_0]v_0^{-1}$ is invertible too, and its inverse is

$$(u - \lambda I)^{-1} = v_0[(u - \lambda I)v_0]^{-1} = \sum_{n \geq 0} (\lambda - \lambda_0)^n v_0^{n+1}. \tag{2.4}$$

This is the power series expansion that we were waiting for. So far we have that $Sp(u)$ is closed and R_u is analytic on its complement.

Let us check that $|\lambda| \leq |||u|||$ for $\lambda \in Sp(u)$ (so that $Sp(u)$ is bounded). For λ such that $|\lambda| > |||u|||$, $\lambda I - u = \lambda(I - \lambda^{-1}u)$, and since $|||\lambda^{-1}u||| < 1$ by assumption, we can apply Lemma 2.3, and get that $I - \lambda^{-1}u$ and hence also $\lambda I - u$ are invertible. So $\lambda \notin Sp(u)$.

Finally we need to know that $Sp(u) \neq \emptyset$. First observe that

$$\lim_{\lambda \to \infty} |||R_u(\lambda)||| = 0 \tag{2.5}$$

because we just saw that for $|\lambda|$ large, $R_u(\lambda) = (\lambda I - u)^{-1} = \lambda^{-1}(I - \lambda^{-1}u)^{-1}$, and then $|||(I - \lambda^{-1}u)^{-1}||| \leq 2$ as soon as $|||\lambda^{-1}u||| \leq 1/2$, by (2.2).

Now, if $Sp(u) = \emptyset$, R_u is an analytic function on $\mathbb{C}$ that tends to 0 at ∞, and by Liouville's theorem it must be zero. Don't worry about the fact that it is valued in $\mathcal{L}(E)$, because for any bounded linear form φ on $\mathcal{L}(E)$, we get by composition that $\varphi \circ R_u$ is a (usual) nalytic function on $\mathbb{C}$ that vanishes at ∞, hence $\varphi \circ R_u(\lambda) = 0$ for every $\lambda \in \mathbb{C}$ and every φ, and the conclusion follows. $\square$

Another important object is the

$$\textbf{spectral radius } \rho(u) = \sup_{\lambda \in Sp(u)} |\lambda| \leq |||u|||. \tag{2.6}$$

The last inequality is what we checked at the end of the lemma above, with the Neumann series trick.

The inequality can be strict, even in the Euclidean $\mathbb{R}^2$ and even if u is diagonalizable (but not in an orthonormal basis). For instance take $u(e_1) = e_1 + e_2$, $u(e_2) = ae_2$ with a very close to 1. The matrix is diagonalizable, with two different eigenvalues 1 and a, but the norm is at least $||u(e_1)|| \geq \sqrt{2}$.

In fact there is a formula for the spectral radius:

$$\rho(u) = \lim_{n \to +\infty} ||u^n||^{1/n} = \inf_{n > 0} ||u^n||^{1/n}, \tag{2.7}$$

which is even what we'll use as a definition when we discuss C^*-algebras (and there is no notion of eigenvalue).

Now why do we get the second inequality and why do (2.7) and (2.6) give the same number?

For the second inequality, this uses the fact that the sequence $\{a_k\}$, where $a_k = |||u^k|||$, is not any sequence at randon, it has some regularity, which comes from the fact that

$$a_{k+l} \leq a_k a_l \ \text{ for } 0 \leq k, l < +\infty. \tag{2.8}$$

In the context of u here, this is just the fact that $|||u^k \circ u^l||| \leq |||u^k|||\,|||u^l|||$; in the later case of C^* algebras, this will be the same thing, coming from the basic rule $||uv|| \leq ||u||\,||v||$ for the norm. I won't do the full detail, but then the simplest to understand what is going on is probably to take $f(k) = \ln(a_k)$, so that (2.8) becomes the subadditivity rule $f(k+l) \leq f(k) + f(l)$.

I pass the study of such sequences, and the fact that they behave in a vaguely linear way, so that for instance $\frac{1}{n} f(n)$ tends to be decreasing and go to a limit. This eventually yields that $\lim_{n \to +\infty} ||u^n||^{1/n} = \inf_{n > 0} ||u^n||^{1/n}$ as above.

I also pass the proof of the fact that the two definitions of ρ coincide. If I recall, this is a not too complicated argument, using the discussion above and the Cauchy formula for the analytic function R_u. At least you can guess that if $|||u^k||| < A^k$ for some k, then the Neumann series trick and the control on the $|||u^n|||$, n large, that we get, shows that $\lambda I - u$ is invertible as soon as $|\lambda| > A$. Then you can let k tend to $+\infty$ and get one estimate.

Let us give a simple example instead of proving. On $\mathbb{C}^m$ we can already have a good idea of what happens, because u is equivalent to a linear matrix with a Jordan decomposition; the spectral radius is the largest size for an eigenvalue, and it is easy to see that the leading term for the norm of a large power of a Jordan matrix with eigenvalue λ is just $|\lambda|^n$.

2.2 Hilbert spaces and the adjoint

We will continue with the case when E is a Hilbert space (we will keep it complex for the same reasons as above). We exclude the case when $E = \{0\}$ to avoid trouble with some of the statements (for instance, when $E = \{0\}$, we have one operator u and its spectrum is empty. I think. But truly I don't want to think about that.)

Let us denote by $\langle u, v\rangle$ the scalar product of the two vectors $u, v \in E$. Let us adopt the notation that $\langle u, v\rangle$ is linear in u and semilinear in v. That is, $\langle u, \lambda v\rangle = \overline{\lambda}\langle u, v\rangle$. Typical example on $\mathbb{C}^n$: $\langle u, v\rangle = \sum_{j=1}^n u_j \overline{v}_j$.

Recall that by a theorem of Riesz, E can be identified to its topological dual (written E^* above). The identification mapping, we shall call it Ψ for a minute, and then we'll forget it, is given by $\Psi(v) = \varphi_v$, where for $v \in E$, φ_v is the continuous linear form given by $\varphi_v(u) = \langle u, v\rangle$. The main point of the Riesz theorem is that every element of E^* is a φ_v. We also know (and this is easy) that $|||\varphi_v||| = ||v||$ (Ψ is an isometry), but alas Ψ is only semilinear: $\Psi(\lambda v) = \overline{\lambda}\Psi(v)$.

The proof of the theorem of Riesz is easy (very little structure is needed), but this is misleading: **the Riesz representation theorem is very powerful**. One illustration among others: try to prove directly (without Riesz), even in the simple case of finite Borel measures on $\mathbb{R}$, the Radon-Nikodym theorem that says that any finite measure ν that is absolutely continuous with respect to the finite measure μ is given by a density $f \in L^1(\mu)$.

Return to $u \in \mathcal{L}(E)$ on a Hilbert space. When we identify E^* with E, the transposed operator becomes what we call an adjoint. Let us give the definition anyway. The **adjoint of** u is the operator $u^* \in \mathcal{L}(E)$ such that for every choice of $x, y \in E$,

$$\langle u^*(x), y\rangle = \langle x, u(y)\rangle \tag{2.9}$$

Note first that the existence of $u^*(x)$ comes from the fact that $y \to \langle x, u(y)\rangle$ is continuous semi-linear form (if you don't like, take the conjugates), so it is given by the scalar product (here on the right) by some element of E, which we call $u^*(x)$. This element is unique. Then u^* is linear: easy because both sides of (2.9) are linear, and by uniqueness of $u^*(x)$.

It is now also easy to check that $u \to u^*$ is a semi-linear mapping, that $(u^*)^* = u$, and that $|||u^*||| = u$. Also, $(u \circ v)^* = v^* \circ u^*$. Needless to say that these properties are proved like the same ones for transposed operators (in fact, it is the same thing). Things are simpler here because Hilbert spaces are reflexive (by Riesz again!).

For the same sort of reasons, let us leave as an exercise to check that u^* is compact if and only if u is compact.

Notice that u^* is invertible iff u is invertible, with $(u^*)^{-1} = (u^{-1})^*$ This is now easy to check because I just gave the formula.

Recall that in $\mathbb{C}^n$, taking the adjoint of u amounts by taking the adjoint of its matrix M, where the adjoint is $M^* = \overline{{}^tM} = {}^t\overline{M}$.

Exercise. Prove that $|||u \circ u^*||| = |||u^* \circ u||| = |||u|||^2$. The operators $u \circ u^*$ and $u^* \circ u$ are often nicer to use because they are self-adjoint.

Exercise. Check that the adjoint of the operator u_Λ of (1.8) is $u_{\Lambda'}$, where $\lambda'_n = \overline{\lambda}_n$ for $n \geq 0$.

Definition 2.4. *Let $u \in \mathcal{L}(E)$ (a complex Hilbert space) be given.*

• *We say that u is self-adjoint when $u = u^*$;*

• *We say that u is normal when $uu^* = u^*u$;*

• *We say that u is unitary when $uu^* = I = u^*u$ (we need both, think about the shift!); then u is obviously invertible, but it is also an isometry: $||u(x)|| = ||x||$ for $x \in E$.*

• *u is a projection if $u^2 = u$.*

The point of "normal" is that many properties of self-adjoint operators also hold for normal operators. For instance the fact that eigenvectors associated to different eigenvalues are orthogonal, or (hence) the diagonalization in finite dimensions.

Exercise. Check that every unitary u is an isometry. Check that $u \in \mathcal{L}(E)$ is unitary as soon as it is an isometry. [Hint: there is a way to compute a bilinear form from its quadratic form, and similarly for sesquilinear forms.]

Exercise. Check that if $u \in \mathcal{L}(E)$ is a projection, its image L is closed, there is a direct decomposition $E = L + N$, where N is the kernel of u, and $u(x + y) = x$ when $x \in L$ and $y \in N$. When in addition $u^* = u$, L is orthogonal to N (and u is the orthogonal projection on L).

2.3 First spectral properties of self-adjoint operators

We start with simple properties of general $u \in \mathcal{L}(E)$ (which will simplify when $u^* = u$). First

$$Sp(u^*) = \overline{Sp(u)} \quad \text{(complex conjugation)}. \tag{2.10}$$

This is easy: if $\lambda \notin Sp(u)$, i.e., $u - \lambda I$ is invertible, then its adjoint $u^* - \overline{\lambda} I$ is invertible, and so $\overline{\lambda} \notin Sp(u^*)$; the other direction is the same for u^*. Next

$$Ker(u^*) = Im(u)^\perp \quad \text{and } Ker(u) = Im(u^*)^\perp \tag{2.11}$$

where we denote by $Ker(u)$ the kernel of u, and by $Im(u)$ its image. Since $(u^*)^* = u$, we just need to check the first part. If $x \in Ker(u^*)$ and $y \in Im(u)$, we can write $y = u(z)$ for some $z \in E$ and then $\langle x, y\rangle = \langle x, u(z)\rangle = \langle u^*(x), y\rangle = 0$. So the spaces are orthogonal. Conversely, if $x \in E$ is orthogonal to $Im(u)$, then for all $y \in E$ $\langle u^*(x), y\rangle = \langle x, u(y)\rangle = 0$, so $u^*(x)$ is orthogonal to the whole word, hence $u^*(x) = 0$, as needed.

Obviously, it follows that

$$\text{if } u = u^*, \text{ then } Ker(u) = Im(u)^\perp. \tag{2.12}$$

Next we check that

$$\text{if } u = u^*, \text{ then } Sp(u) \subset \mathbb{R}. \tag{2.13}$$

We already know that $Sp(u)$ is symmetric with respect to the real axis, but this is not enough!

We first check the easier fact that eigenvalues for u are real. Indeed, if $u(x) = \lambda x$ for some $x \neq 0$, then $\lambda ||x||^2 = \langle u(x), x \rangle = \langle x, u^*(x) \rangle = \langle x, u(x) \rangle = \overline{\lambda} ||x||^2$; we divide by $||x||^2 \neq 0$ and get that λ is real.

Note that we'll often see the quantity $\langle u(x), x \rangle$ for self-adjoint operators; in particular studying it gives a good idea of where the spectrum is (see near (2.19)).

Now we can check that if $\lambda \in \mathbb{C}$ is not real, then $v = u - \lambda I$ is invertible. Observe that for all $x \in E$, and as above, $\langle u(x), x \rangle = \langle x, u^*(x) \rangle = \langle x, u(x) \rangle = \overline{\langle u(x), x \rangle}$, so $\langle u(x), x \rangle$ is always real. For $x \in E$,

$$\langle v(x), x \rangle = \langle u(x), x \rangle - \lambda ||x||^2. \tag{2.14}$$

Write $\lambda = a + ib$, with $b \neq 0$. Since $\langle u(x), x \rangle$ is real, so $|\langle v(x), x \rangle|$ is at least as large as its imaginary part, which is $|b| ||x||^2$. That is, we proved that

$$||v(x)|| ||x|| \geq |\langle v(x), x \rangle| \geq |b| ||x||^2 \tag{2.15}$$

hence, for $x \neq 0$,

$$||v(x)|| \geq |b| ||x||. \tag{2.16}$$

Thus, not only v is injective, but with uniform bounds that imply more: $v(E)$ is a closed space and $v : E \to v(E)$ is invertible (with norm at most $|b|^{-1}$). We leave the details as an exercise.

Now, since v also is self-adjoint, by (2.12) the orthogonal of $v(E)$ is the kernel of v, which is $\{0\}$. So in fact $v(E) = E$, and we have shown that v is invertible; (2.13) follows.

We continue with the list of properties of self-adjoint operators. Let us check that

$$\text{if } u = u^* \text{ and the vector space } F \text{ is such that } u(F) \subset F \text{ then } u(F^\perp) \subset F^\perp. \tag{2.17}$$

That is, $F^\perp$ also is invariant under u. Indeed if $x \in F^\perp$, then for all $y \in F$, $\langle u(x), y \rangle = \langle x, u^*(y) \rangle = \langle x, u(y) \rangle = 0$ because $u(y) \in F$. It follows that $x \in F^\perp$, as announced. Notice that we did not require F to be closed, but if F is invariant, then its closure too, because u is continuous.

The orthogonality of different eigenspaces is of constant use: for $u \in \mathcal{L}(E)$, $x, x' \in E$, $\lambda, \lambda' \in \mathbb{C}$,

$$\text{if } u = u^*,\ u(x) = \lambda x \text{ and } u(x') = \lambda' x', \text{ with } \lambda' \neq \lambda, \text{ then } \langle x, y \rangle = 0. \tag{2.18}$$

This is because $\lambda \langle x, y \rangle = \langle u(x), y \rangle = \langle x, u(y) \rangle = \langle x, \lambda' y \rangle = \overline{\lambda}' \langle x, y \rangle = \lambda' \langle x, y \rangle$, where I used the fact that since λ' is an eigenvalue for u, it lies in the spectrum, so it is real. We get that $\lambda \langle x, y \rangle = \lambda' \langle x, y \rangle$, hence $\langle x, y \rangle = 0$ as needed.

Next the "variational" characterization of the two extremities of the spectrum (we will see a generalization later, with Rayleigh quotients). Suppose $u^* = u$, and set

$$m = \inf\big\{\lambda\,;\,\lambda \in Sp(u)\big\} \ \text{ and } \ M = \sup\big\{\lambda\,;\,\lambda \in Sp(u)\big\} \tag{2.19}$$

(this makes sense since $Sp(u) \subset \mathbb{R}$). Then m and M lie in the spectrum $Sp(u)$, and

$$m = \inf\big\{\langle u(x), x\rangle\,;\, x \in E, ||x|| = 1\big\} \ \text{ and } \ M = \sup\big\{\langle u(x), x\rangle\,;\, x \in E, ||x|| = 1\big\}. \tag{2.20}$$

Let M' denote the supremum at the end of (2.20). We first check that if $\lambda \in (M', +\infty)$, then $u - \lambda I$ is invertible; this will show that $M \leq M'$. We proceed as for (2.13), but from the right instead of from above. Set $v = u - \lambda I$; then for $x \in E$,

$$\langle v(x), x\rangle = \langle u(x), x\rangle - \lambda ||x||^2 \leq -(\lambda - M')||x||^2. \tag{2.21}$$

By Cauchy-Schwarz, we now get that

$$||v(x)||||x|| \geq |\langle v(x), x\rangle| \geq (\lambda - M')||x||^2, \tag{2.22}$$

and then $||v(x)|| \geq |b|||x||$ (as in (2.16)). Again this forces v to be an isomorphism on its closed image $v(E)$; it also implies that $Ker(v) = \{0\}$, hence by (2.12) $v(E)$ is dense. As before, v is invertible, as needed for the proof that $M \leq M'$.

Next we check that $M' \in Sp(u)$; once we do this, we get that $M = M'$, the same proof (or replacing u with $-u$) gives the first part of (2.20), and we already know that m and M lie in the spectrum, because it is closed.

Set $v = M'I - u$. Then for $||x|| = 1$, $\langle v(x), x\rangle = M' - \langle u(x), x\rangle \geq 0$ by definition of M'. Consider the sesquilinear form Φ on $E \times E$ defined by $\Phi(x, y) = \langle v(x), y\rangle$. We just said that $\Phi(x, x) \geq 0$ for all x, and by Cauchy-Schwarz for this form,

$$|\langle v(x), y\rangle|^2 \leq \langle v(x), x\rangle\langle v(y), y\rangle \tag{2.23}$$

[if you are afraid that it does not hold for degenerate forms, either check or add $\varepsilon\langle x, y\rangle$ to Φ and then let ε go to 0]. Then use (2.23) to write

$$||v(x)||^2 = \sup_{||y||=1} |\langle v(x), y\rangle|^2 \leq \langle v(x), x\rangle \sup_{||y||=1} \langle v(y), y\rangle$$

and now by the usual Cauchy-Schwarz

$$||v(x)||^2 \leq \langle v(x), x\rangle\, |||v|||. \tag{2.24}$$

Now where is the problem? By definition of M' as a sup, we can find a sequence $\{x_n\}$ in the unit sphere of E such that $\langle v(x_n), x_n\rangle = M' - \langle u(x_n), x_n\rangle$ tends to 0. By (2.24), $||v(x_n)||$ tends to 0 too. This is not possible if we assumed that $M' \notin Sp(u)$, because v is invertible and so $1 = ||x_n|| = ||v^{-1}(x_n)|| \leq |||v^{-1}|||\, ||x_n||$. This concludes our proof of (2.20).

Some further comments. Concerning (2.20), it is curious that there is no more direct proof of the fact that M' lies in the spectrum. Somehow the proof uses the scaling of the size of the errors, rather than just their sign.

It can happen that M (and m would be the same) is not an eigenvalue. This is the case, for instance, if $u = u_\Lambda$ (as in (1.8)), for an increasing sequence such that $\lim_{n\to+\infty} \lambda_n = M$, but M is never reached.

We won't resist mentioning a last result (for the moment), called **Weyl's criterion**. Let us say that λ is an approximate eigenvalue for u when we can find a sequence $\{x_n\}$ in the unit sphere of E, such that

$$\lim_{n\to+\infty} ||u(x_n) - \lambda x_n|| = 0. \tag{2.25}$$

Let us even call $\sigma(u)$ (the Weyl spectrum of u) the set of approximate eigenvalue for u. In general, we have that

$$\sigma(u) \subset Sp(u), \tag{2.26}$$

because when $\lambda \in \mathbb{C} \setminus Sp(u)$, $u - \lambda I$ is invertible, so $||u(x) - \lambda x|| \geq c||x||$ for some $c > 0$ and (2.25) cannot happen. The amusing thing is that when $u \in \mathcal{L}(E)$ is self-adjoint, the converse is true:

$$\text{if } u^* = u, \text{ then } \sigma(u) = Sp(u). \tag{2.27}$$

Let us check this, as usual by contradiction, so let us assume that $u^* = u$ and $\lambda \in Sp(u) \setminus \sigma(u)$. Notice that λ is real (by (2.13)), and is not an eigenvalue (because eigenvalues are approximate eigenvalues!). So $v = u - \lambda I$ is injective, hence by (2.12) $v(E)$ is dense. On the other hand, since $\lambda \notin \sigma(u)$, we can find $\varepsilon > 0$ such that $||v(x)|| \geq \varepsilon$ for all x in the unit sphere. We have seen that already: then v is an isomorphism on its image, which is closed, hence $v(E) = E$ (it was dense), and v is invertible, a contradiction that proves (2.27).

Exercise. Let $u \in \mathcal{L}(E)$ be self-adjoint. Prove that u is nonnegative (i.e., $\langle u(x), x\rangle \geq 0$ for all $x \in E$) if and only if $Sp(u) \subset [0, +\infty)$.

Exercise Consider $E = L^2(\mathbb{R}^n, d\mu)$, with a measure μ that may as well be the Lebesgue measure. Let $u = u_f$ denote the operator of multiplication by the given bounded function f. In some way, this is a variant of the example above with ℓ^2, and in a way u already comes as diagonalized as possible.

1. Check that $\lambda = 0$ is an eigenvalue for u if and only if $\mu(f^{-1}(\{0\})) \neq 0$.
2. Check that u is invertible if and only if there is an $r > 0$ such that $\mu(f^{-1}(B(0,r)) = 0$.
3. Check that u is self-adjoint if and only if $f(x) \in \mathbb{R}$ for μ-almost every $x \in \mathbb{R}^n$.

Exercise. Consider the Fourier transform $\mathcal{F} : L^2(\mathbb{R}^n) \to L^2(\mathbb{R}^n)$, which I choose to be defined so that

$$\mathcal{F}f(\xi) = (2\pi)^{-n/2} \int_{x\in\mathbb{R}^n} e^{-i\langle x,\xi\rangle} f(x)dx \tag{2.28}$$

for $f \in L^1 \cap L^2$, and the normalization is chosen so that $\mathcal{F}$ is an isometry of L^2. Recall that its inverse $\mathcal{F}^{-1}$ is given by the same formula, but with $e^{i\langle x,\xi\rangle}$. Do the exercise only if you already know about $\mathcal{F}$.

1. Find a formula for $\langle \mathcal{F}f, g\rangle$ when f, g are in a dense class.
2. What is the adjoint of $\mathcal{F}$?
3. Is $\mathcal{F}$ unitary?
4. Can you prove that 1 is an eigenvalue for $\mathcal{F}$?

It looks like we should be able to say more (from the question 3), but this is not so clear to me at this point.

Exercise Use the two previous exercises to construct an exercise on Fourier multipliers (defined by $u : g \mapsto \mathcal{F}^{-1}(m\mathcal{F}(g))$).

2.4 Spectral decomposition of self-adjoint compact operators

We add more properties and get a simpler description of u. In what follows, E is a Hilbert space (certainly not $\{0\}$, but the interesting case is when E is infinite-dimensional).

And $u \in \mathcal{L}(E)$ is now assumed to be both self-adjoint and compact. Let us already recall what we know about the spectrum of u.

First, $Sp(u) \subset [-|||u|||, |||u|||] \subset \mathbb{R}$ (because $u^* = u$).

When E is infinite-dimensional, $0 \in Sp(u)$, but it does not need to be an eigenvalue.

Each point of $Sp(u) \setminus \{0\}$ is an isolated point of $Sp(u)$ and is an eigenvalue with a finite multiplicity (see Theorem 1.8).

Hence the set of positive spectral values (or here eigenvalues) is either empty, or finite, or can be organized as a decreasing sequence that tends to 0. Similarly, the set of negative spectral values is either empty, or finite, or can be organized as an increasing sequence that tends to 0.

We add two new pieces of information here. For $\lambda \in Sp(u)$, call $E_\lambda = Ker(u - \lambda I)$ the vector space spanned by the corresponding eigenfunction. It could happen (as in the case of u_Λ when the λ_n are positive and tend to 0) that $E_0 = \{0\}$. Anyway,

$$E \text{ is the direct orthogonal sum of the } E_\lambda,\ \lambda \in Sp(u). \tag{2.29}$$

This is a nice decomposition of E adapted to u. In finite dimensions, this just means that when u is self-adjoint, we can diagonalize u in an orthonormal basis (take an orthonormal basis of each E_λ).

Let us check this. Notice that E_λ is closed because it is the kernel of a bounded operator. Then $u(E_\lambda) \subset E_\lambda$ by definition. Also, $E_\lambda \perp E_\mu$ when $\lambda \neq \mu$, by (2.18). Call F the linear span of the E_λ; let us show that F is dense. Suppose not; then $F^\perp$ is not reduced to $\{0\}$. Since $u(F) = F$, we also get that $u(F^\perp) \subset F^\perp$ (see (2.17)). So we can consider $\widetilde{u}$, the restriction of u to $F^\perp$.

By construction, $\widetilde{u}$ has no eigenvalue. It is also a compact self-adjoint operator (the formula that defines self-adjoint is still valid on a smaller space), so we know that its spectrum is $\{0\}$ (recall that the empty set is impossible because of Liouville).

Now we apply (2.20) and find that $u(x) \perp x$ for all $x \in F^\perp$. Let $x \in F^\perp$, and set $y = u(x)$ and $z = u^2(x)$; then

$$0 = \langle u(x+y), x+y\rangle = \langle y+z, x+y\rangle = ||y||^2 + \langle z, x\rangle = ||y||^2 + \langle u^2(x), x\rangle = ||y||^2 + ||u(x)||^2$$

because $x \perp y$ and $y \perp z$, and then u is sel-adjoint. In particular $u(x) = 0$ on $F^\perp$, which is the desired contradiction because $F^\perp$ was not supposed to contain eigenvectors. So F is dense (we cannot expect it to be the whole E because we only took F to be the vector linear span of the E_λ (I mean, with finite sums and no series).

Now what do we mean by this direct orthogonal sum of E_λ? Call π_λ the orthogonal projection on E_λ. By Pythagorus, if $x \in E$, then $\sum_{\lambda \in Sp(u)} ||\pi_\lambda(x)||^2 \leq ||x||^2$, because this is true for any finite sum. Notice that we are lucky here: $Sp(u)$ is at most countable, so we could write sums as series in a natural way. Anyway, we can take limits of finite sums, and see that $\Pi(x) = \sum_{\lambda \in Sp(u)} \pi_\lambda(x)$ exists, and then that $\Pi(x) - x \in F^\perp$. That is, $\Pi(x) = x$ (because $F^\perp = \{0\}$; in general, $\Pi(x)$ would be the projection of x on $\overline{F}$). This gives an orthonormal decomposition of x as $\sum_\lambda \pi_\lambda(x)$.

If we want, we can also take an orthonormal basis of each E_λ, $\lambda \neq 0$, and then use the union $\{e_\mu\}$ of the bases, and we can write $x \in E$ as $x = \pi_0(x) + \sum_\mu x_\mu e_\mu$, with $x_\mu = \langle x, e_\mu\rangle$. The last sum is again at most countable. For π_0, if E is not separable, we need an uncountable Hilbertian basis of E_0 to decompose $\pi_0(x)$ (but this is all right).

Next we mention a very nice feature of self-adjoint (compact) operators: the use of so-called **Rayleigh quotients** to find the successive eigenfunctions of u. It is already interesting (and used a lot) in finite dimensions.

Let u be self-adjoint and compact, and assume that it has some positive eigenvalues. How do we find them? We start with the largest one, which we call M. we saw in (2.20) that the supremum of the spectrum is

$$M = \sup\big\{\langle u(x), x\rangle\,;\, x \in E, ||x|| = 1\big\}. \tag{2.30}$$

and if $M > 0$ we know that M is an eigenvalue. So (2.30) gives the formula for the largest eigenvalue. And incidentally the sup in (2.30) is a maximum, since $\langle u(x), x\rangle = M$ when x is a unit eigenvector of E_M (i.e., $||x|| = 1$ and $u(x) = Mx$).

Call $M = \lambda_0$. Suppose there are eigenvalues λ such that $0 < \lambda < M$, and call λ_1 the largest such eigenvalue. Then we claim that

$$\lambda_1 = \sup\big\{\langle u(x), x\rangle\,;\, x \in E, ||x|| = 1 \text{ and } x \perp E_{\lambda_0}\big\}. \tag{2.31}$$

Certainly the supremum is at least as large as λ_1, because if x is a unit eigenvector associated to λ_1, then $x \perp E_{\lambda_0}$ and $\langle u(x), x\rangle = \lambda_1$. For the other direction, we consider the restriction $\widetilde{u}$ of u to $V_1 = E_{\lambda_0}^\perp$; we know that $u(V) \subset V$ because $u(E_{\lambda_0}) \subset E_{\lambda_0}$, and $\widetilde{u}$ is again self-adjoint and compact. Now λ_1 is the largest eigenvalue for $\widetilde{u}$ (because $E_{\lambda_1} \subset V$ and the eigenvalue λ_0 is now forbidden on V), and (2.31) is exactly the same as (2.30) for $\widetilde{u}$ on V.

As you guessed, we can iterate: if $\lambda_k > 0$ is the k-th largest eigenvalue for u (assuming there are that many positive eigenvalues), then

$$\lambda_k = \sup \big\{ \langle u(x), x \rangle \, ; \, x \in E, ||x|| = 1 \text{ and } x \perp E_{\lambda_j} \text{ for } 0 \le j < k \big\}. \tag{2.32}$$

If you prefer to count eigenvalues with their multiplicity, and at the same time construct an orthonormal basis composed of eigenvectors, you can also do the following. Consider first $\mu_0 = \sup \big\{ \langle u(x), x \rangle \, ; \, x \in E, ||x|| = 1 \big\}$ (the same formula as in (2.30)). If $\mu_0 > 0$, we have seen that the sup is a max, and we can find a unit vector e_0 such that $u(e_0) = \mu_0 e_0$. If $\lambda_0 = \mu_0$ has a multiplicity larger than 1, we have a lot of choice for e_0, but this does not matter.

Then consider $V_1 = e_0^\perp$. It is easy to see that it is stable by u. So we do the same as above, but for the restriction of u to V_1. That is, we now consider $\mu_1 = \sup \big\{ \langle u(x), x \rangle \, ; \, x \in V_1, ||x|| = 1, \big\}$. If $\mu_1 > 0$, it is an eigenvalue for $u_{|V_1}$, the supremum is a maximum, and we can find a unit eigenvector e_1 for $u_{|V_1}$, associated to μ_1 (and orthogonal to e_0 since it lies in V_1). We can continus like this, up to the moment when we get a number $\mu_k = 0$, and then we know that we exhausted all the positive eigenvalues and eigenvectors.

Of course the negative eigenvalues can be treated the same way (or just use $-u$).

Comment. Even when E is not separable, taking a compact u puts us in the separable word. We can see this in the spectral decomposition above (u vanishes on E_0, and the rest has an at most countable Hilbertian basis), but we could also have used the fact that for every $\varepsilon = 2^{-k}$, u is within ε of a finite rank operator. When we add the dimensions of all the corresponding images, we still get an at most countable number. Compare (if you wish) with the fact that if the family $\{u_i\}_{i \in I}$ is summable, then its support (the set of i such that $u_i \neq 0$) is at most countable, even though I could a priori be really huge.

3 Spectral results, self-adjoint operators

3.1 Preparation to the functional calculus: C^*-algebras

For a matrix, we know how to compute polynomials in M (such as $M^3 - 17M$), and the theorem of Cayley-Hamilton says that $P(M) = 0$ when P is the characteristic polynomial P_0 of M, or of course the product of this polynomial by any other polynomial. Thus if we want to compute $P(M)$, it is advisable to first make an Euclidean division of P by P_0, i.e., write $P = QP_0 + R$ for some polynomials Q and R, and observe that $P(M) = R(M)$ and compute.

Here we shall consider extensions of the simpler situation where M is self-adjoint, hence diagonalizable in an orthonormal basis. In this case $P(M)$ is even easier to compute: in this orthonormal basis (i.e., after conjugating with an orthogonal matrix), it becomes a diagonal matrix with diagonal terms $P(\lambda_j)$, where the λ_j are the eigenvalues. In particular, $P(M)$ only depends on (M and) the values of P on the spectrum, and it makes sense to define $f(M)$ for any function f defined on the spectrum. We want to do something like this for operators, and for instance define $\sqrt{u}$ when u is compact, self-adjoint, and its spectrum is contained in $[0, +\infty)$.

You should be warned that there are lots of way to do functional calculus on u (i.e. define correctly $f(u)$ for suitable functions f), and lots of assumptions on u and its spectrum that make it possible.

We start with an unpleasant (for me; many people love it) **definition** part, but it will make it easier to formalize things later. So we are about to define **C*-algebras** (in french, C^*-algèbres, or even algèbres stellaires (avec une étoile)). Our model, however, is $\mathcal{L}(E)$, where E is a complex Hilbert space. We will use the complex structure, the fact that we have a nice involution $u \to u^*$, and a multiplication law, which is the composition of operators. So it is not assumed to be commutative.

A C^*-algebra is an involutive complex Banach algebra. Complex Banach aglebra means that it is a Banach space, with also a product law $u, v \to uv$ (not necessarily commutative) with the following three properties:

• It is associative:

$$(uv)w = u(vw) \text{ for } u, v, w \in A; \tag{3.1}$$

• the product is linear:

$$(u+v)w = uv + uw, u(v+w) = uv + uw, \text{ and } \lambda(uv) = (\lambda u)v = u(\lambda v) \quad \text{for } \lambda \in \mathbb{C} \text{ and } u, v, w \in A; \tag{3.2}$$

• The following norm inequality holds:

$$||uv|| \leq ||u||||v|| \text{ for } u, v \in A. \tag{3.3}$$

The third condition could seem weird to you, but without it we would not get that far, and more importantly we have interesting examples, so we keep it.

Something confusing to me: when you multiply $|| \cdot ||$ by $a < 1$, you make this last condition $||uv|| \leq ||u||||v||$ easier (so suppose you only had $||uv|| \leq C||u||||v||$, $a|| \cdot ||$ will have the desired property if $aC \leq 1$). This trick won't work as such with the next property.

And "**involutive**" means that we also add an involution, denoted bu $u \to u^*$, with the properties that $(u^*)^* = u$, but also which is semilinear (i.e., $(\lambda u + \mu v)^* = \overline{\lambda} u^* + \overline{\mu} v^*$), and such that $(uv)^* = v^* u^*$ for $u, v \in A$.

• Also, we add for "involutive normed" (and hence "involutive Banach") the condition that

$$||uu^*|| = ||u||^2. \tag{3.4}$$

Then

$$||u^*|| = ||u|| \tag{3.5}$$

because $||u||^2 = ||uu^*|| \leq ||u||||u^*||$, so $||u|| \leq ||u^*||$, and similarly $||u^*|| \leq ||u||$.

Finally, I'll follow a tradition (of many people but not all) that **the algebra A contains a unit,** which I'll denote by I and satisfies $Iu = uI = u$ for $u \in A$). In French we way that A is unifère, I think.

That's it. I know, that makes a long list. But we have **two simple examples**. The first one is the set $C(K)$ of continuous functions $f : K \to \mathbb{C}$, where K is a compact set, and with the sup norm (the involution is $f \to \overline{f}$ and the identity is the constant function 1); this is the model for a commutative C^*-algebra with a unit. The second one (non commutative) is $\mathcal{L}(E)$ mentioned above. For this case, (3.4) holds because $||u^*|| = ||u||$ by definition of self-adjoint, $||uu^*|| \leq ||u||^2$ is easy from the definitions (or (3.3)), and $||uu^*|| \geq ||u||^2$ because $||u(x)||^2 = \langle u(x), u(x)\rangle = \langle u^*u(x), x\rangle \leq ||u^*u(x)||\,||x||$.

From now on, unless otherwise specified, A is a C^*-algebra (with all the properties mentioned above and a unit). Some properties below do not need as much but I think we'll be able to manage for the rest of the course with this simplification.

Some definitions and basic properties mentioned for $\mathcal{L}(E)$ go through. The spectrum of $u \in A$ is the set of $\lambda \in \mathbb{C}$ such that $u - \lambda I$ is not invertible (in A of course).

So for $\mathcal{L}(E)$ this is the same as above. In the simpler case of $C(K)$, the function $f \in C(K)$ is invertible precisely when it does not vanish, so $f - \lambda I$ is invertible when f does not take the value λ, and hence $Sp(f) = f(K)$.

The spectral radius of an element makes sense, and is given by

$$\rho(u) = \lim_{n\to+\infty} ||u^n||^{1/n} = \inf_{n>0} ||u^n||^{1/n} \tag{3.6}$$

(the reason for the equality is the same as below: $||u^m u^n|| \leq ||u^m||\,||u^n||$, which is in the list of definitions above). And for instance we have the following analogue of properties above.

Proposition 3.1. *Let A be a C^*-algebra with a unit (so $A \neq \{0\}$), and $u \in A$. Then*
- *$Sp(u)$ is a nonempty compact subset of $\mathbb{C}$;*
- *$\rho(u^*) = \rho(u) \leq ||u||$;*
- *$\rho(u) = \sup_{\lambda\in Sp(u)} |\lambda|$;*
- *if $u^* = u$, then $\rho(u) = ||u||$.*

The unpleasant part is the third one, whose details we did not do above for $u \in \mathcal{L}(E)$, and we still don't do them now. But the idea is still to use analytic functions, then size estimates on the coefficients of power series expansions, and finally the submultiplicativity of the norm. Then for the last point we observe that if $u = u^*$, then $||u^2|| = ||u||^2$, and by induction $||u^{2^m}|| = ||u||^{2^m}$. We take these special numbers in the definition of $\rho(u)$ and get $\rho(u) = \lim_{m\to+\infty} ||u^{2^m}||^{2^{-m}} = ||u||$.

Now a little bit of morphisms. I let you guess what is a morphism of C^*-algebras (many things to check though). The idea is that it has to respect a certain number of things; let me state three in particular: $\varphi(u+v) = \varphi(u) + \varphi(v)$, $\varphi(uv) = \varphi(u)\varphi(v)$, and $\varphi(u^*) = \varphi(u)^*$. Here we require a unit for A and B, so we also require that $\varphi(I_A) = I_B$. Incidentally, this means that we can see $\mathbb{C}$ as a subset of A and of B, and $\varphi(\lambda I_A) = \lambda I_B$, so the restriction of φ to $\mathbb{C}I_A$ is trivial. We also require φ to be countiuous (bounded). I do not mention anything about the norm, but see the second item of Lemma 3.2 below. The fist morphism that we'll construct (continuous functional calculus) will preserve the norm though.

Here are simple examples.

1. In the $C(K)$, choose a continuous function h from K to L, and set $\varphi_h(u) = u \circ h$.

2. In $\mathcal{L}(E)$, choose a unitary $h \in \mathcal{L}(E)$ and set $\varphi_h(u) = h^{-1}uh$ (conjugation is always a good thing to try when we look for examples).

3. Closer to what we want to do, given a diagonal matrix $M \in M_n(\mathbb{C})$ with real coefficients, associate to a function f defined on the spectrum K of M the new matrix $f(M) \in M_n(\mathbb{C})$. Notice that $M_n(\mathbb{C})$ is a C^*-algebra too, it is morally the same as $\mathcal{L}(\mathbb{R}^n)$. And please check all the above before using it.

Lemma 3.2. *Let $\varphi : A \to B$ a morphism of C^*-algebras. Then*

- *$Sp(\varphi(u)) \subset Sp(u)$ for $u \in A$;*
- *$||\varphi(u)|| \leq ||u||$ for $u \in A$.*

For the first part, $Sp_B(\varphi(x)) \subset Sp_A(u)$ should be a more correct way to say it. Especially since it could happen that we consider the same u that lies in two algebras that contain each other, and $u - I$ could be invertible in the large one but not in the small one. The units are the same though.

Anyway, if $\lambda \notin Sp(u)$, then $u - \lambda I_A$ is invertible, hence $\varphi(u) - \lambda I_B = \varphi(u) - \lambda\varphi(I_A) = \varphi(u - \lambda I_A)$ is invertible too, because $\varphi((u - \lambda I_A)^{-1})$ is an inverse. So $\lambda \notin Sp(\varphi(u))$; the first part follows.

For the second part we write

$$||\varphi(u)||^2 = ||\varphi(u)\varphi(u)^*|| = ||\varphi(uu^*)|| = \rho(\varphi(uu^*)) \leq \rho(uu^*) \leq ||uu^*|| \leq ||u||^2$$

by the property $||v||^2 = ||vv^*||$, the morphism property, the fact that uu^* is self-adjoint, and so $\varphi(uu^*)$ is self-adjoint too, and hence its spectral radius is equal to its norm, then the inclusion that we just proved, then easy stuff again. □

Much much more exists, and many references too (see Paulin first).

3.2 Continuous functional calculus for bdd self-adjoint operators

We want to give a sense to $f(u)$ when $u = u^* \in \mathcal{L}(E)$ (a Hilbert space). But without the compactness assumption, so that we don't really have a decomposition of E into eigenspaces that we could play with.

Theorem 3.3. *Let E be a complex Hilbert space (not $\{0\}$) and $u \in \mathcal{L}(E)$ be self-adjoint. Set $K = Sp(u) \subset \mathbb{C}$ and denote by $C(K)$ the C^*-algebra of continuous functions on K with complex values. There is a unique morphism $\varphi = \varphi_u$ from $C(K)$ to $\mathcal{L}(E)$ such that $\varphi(id) = u$, where id is the identity mapping on K (i.e., $id(\lambda) = \lambda$ for $\lambda \in K$).*

Let us stop here to comment, but this is not all: φ_u has lots of interesting properties that we'll mention later. For the moment, we just say that for each continuous function $f : K \to \mathbb{C}$, we found a nice way to define $f(u) \in \mathcal{L}$, so that it has the algebraic properties that we would want. For instance, $fg(u) = f(u)g(u)$ (the composition), or $(f+g)(u) = f(u)+g(u)$, or (the involution) $\overline{f}(u) = f(u)^*$.

Here and below I take some liberties: I decided to call $f(u)$ what the theorem calls $\varphi_u(f) \in \mathcal{L}(E)$.

Notice that $C(K)$ is commutative, so its image by φ will be too. That is, $f(u)g(u) = g(u)f(u)$: all our functions of u commute with each other. We knew this already for polynomials, so it is not a surprise.

We now list additional properties of $\varphi = \varphi_u$ (or the mappings $f(u)$). We will prove all that later. First a little more theory:

(3.7) $\quad \varphi$ is an isometry from $C(K)$ to the smallest C^*-algebra that contains u.

Not so surprising, but we'll see. The smallest C^*-algebra in question (also called the C^*-algebra generated by u) is also the intersection of the C^*-algebras $A \subset \mathcal{L}(E)$ that contain u (which makes its existence easy to check, but due to the large number of laws we omit the proof). It is commutative; the definition by intersection does not make this obvious, but we'll have a more constructive proof where all the objects commute. Given that our algebra can be obtained from u by adding all the polynomials in u (that commute), then doing diverse closure operations, and taking limits, this makes sense. Next, a very important information (also called spectral theorem):

$$Sp(f(u)) = f(Sp(u)) \ \text{ for } f \in C(K) \tag{3.8}$$

(or, with the notation of the theorem, $Sp(\varphi(f)) = f(Sp(u))$. This is true when we compute functions of a diagonalizable matrix (or operator) in finite dimension, but it fails in general (when u is not self-adjoint), and it is quite useful. The proof is not really hard, but something has to be done there.

(3.9) $\ f(u)$ is self-adjoint iff (if and only if) f is real-valued; $f(u)$ is nonnegative iff $f \geq 0$;

$$f(u) \text{ is invertible iff } f \neq 0 \text{ on } K, \text{ and when this happens } f(u)^{-1} = \frac{1}{f}(u); \tag{3.10}$$

$$f(u) = 0 \text{ iff } f = 0 \text{ on } K; \tag{3.11}$$

$$\begin{aligned} &\text{if } \lambda \text{ is an eigenvalue for } u, \text{ then } f(\lambda) \text{ is an eigenvalue for } f(u), \\ &\qquad\qquad \text{and } Ker(u - \lambda I) \subset Ker(f(u) - f(\lambda)I). \end{aligned} \tag{3.12}$$

Some of these things are easy, but we'll try to discuss them after the basic construction, which will take some time.

First, $K = Sp(u)$ is a nonempty compact subset of the plane, so $C(K)$ is a C^*-algebra and the statement makes sense.

We can define $P(u)$ when P is a polynomial with complex coefficients, and the morphism $P \to P(u)$, from $\mathbb{C}[X]$ to $\mathcal{L}(E)$, is a morphism that preserves sums, products, and the involution. We use the fact that $u = u^*$ here, because otherwise we would have $P(u)^* = \overline{P}(u^*)$.

We want to find a way to use this "morphism" (the term is not appropriate though because $\mathbb{C}[X]$ is not complete, so it is not a C^*-algebra), essentially with an extension by continuity, and we will need some estimates to do that cleanly.

Lemma 3.4. *For $P, Q \in \mathbb{C}[X]$,*

$$Sp(P(u)) = P(Sp(u)) = P(K); \tag{3.13}$$

$$|||P(u)||| = \sup_{\lambda \in K} |P(\lambda)|; \tag{3.14}$$

$$P(u) = Q(u) \text{ as soon as } P \text{ and } Q \text{ coincide on } K. \tag{3.15}$$

First we check that $P(K) \subset Sp(P(u))$. Let $\lambda \in K$; we want to check that $P(\lambda) \in Sp(P(u))$. Write $P(X) - P(\lambda) = (X - \lambda)Q(X)$ for some polynomial Q. Then $P(u) - P(\lambda)I = (u - \lambda I) \circ Q(u) = Q(u) \circ (u - \lambda I)$. If $(u - \lambda I)$ is not injective, we use the second formula and find that $P(u) - P(\lambda)I$ is not injective. Otherwise, $(u - \lambda I)$ is not surjective, and we use the second formula to find that $P(u) - P(\lambda)I$ is not surjective.

Conversely, let $\mu \in Sp(P(u))$; we want to show that $\mu \in P(K)$. If $P(X) \equiv \mu$, this follows from the fact that $K \neq \emptyset$. Otherwise, call $\lambda_1, \dots, \lambda_k$ the roots of $P - \mu$ (and k its degree), and write $P(X) - \mu = a \prod_j (X - \lambda_j)$, with $a \neq 0$ because P is not constant. Hence $P(u) - \mu I = aU$, where U is the composition of the $u - \lambda_j I$. If none of the λ_j lies in K, then the composition is invertible, which contradicts the fact that $\mu \in Sp(P(u))$. Otherwise, some λ_j lies in K, and $P(\lambda_j) - \mu = 0$, so $\mu \in P(K)$ as desired. So (3.13) holds.

For (3.14), we observe that

$$\begin{aligned} |||P(u)|||^2 = |||P(u)P(u^*)||| &= \sup_{\lambda \in Sp(P(u)P(u)^*)} |\lambda| \\ &= \sup_{\lambda \in Sp((P\overline{P})(u))} |\lambda| = \sup_{\lambda \in K} |P\overline{P}(\lambda)| = \sup_{\lambda \in K} |P(\lambda)|^2 \end{aligned} \tag{3.16}$$

(true in general, then because u is self-adjoint, then by Proposition 3.1 (and because $P(u)P(u)^*) = (P\overline{P})(u)$, then by functional calculus, then (3.13), then calculus; (3.14) follows.

For (3.15), if $P = Q$ on K, then $P - Q$ vanishes on K and now $|||P(u) - Q(u)||| = 0$ by (3.14). □

We may now return to our morphism $P \to P(u)$, from $\mathbb{C}[X]$ to $\mathcal{L}(E)$; let us call it φ_0. First observe that if we want a morphism φ as in the theorem, then it should be that if f

is the restriction to K of a polynomial P, we have to take $\varphi(f) = P(u)$ [because $\varphi(id) = u$ and then φ should preserve "products", sums, and so on].

Because of (3.15), we are lucky, and $P(u)$ depends only on the restriction of P to K, so we can consider that φ is already defined on these restrictions (and equal to φ_0, modulo the correct identifications).

Now we want to extend this from the class C_P of restrictions of polynomials to the whole $C(K)$, and obviously we shall use the theorem of Weierstrass, that says that C_P is dense on $C(K)$ (for the uniform norm: we are not cheating). Now $P \to P(u)$ is uniformly continuous from C_P to $\mathcal{L}(E)$, (3.14) even says that it is an isometry. So φ_0 has a unique continuous extension to $C(K)$. In particular, φ will be unique (recall from Lemma 3.2 that morphisms are continuous).

We are finished with the construction of φ. Now we need to check its numerous properties. First, the extension too is an isometry. This allows us to check the various algebraic (morphism) conditions by taking limits. Let us skip the details.

At this point we have Theorem 3.3, and we start to check the various extra properties.

For (3.7), we see that the image of φ is in the closure of the image of φ_0; this is clearly the smallest C^*-algebra that contains u (because $P(u)$ has to lie in this C^*-algebra for every polynomial P). Also, the operators $P(u)$ commute with each other, so our smallest C^*-algebra is commutative too (take limits).

We already proved (3.8) when f is a polynomial. We have to extend this to $f \in C(K)$. Suppose first that $\lambda \notin f(K)$; then we can define a continuous function g on K by $g(t) = (f(t) - \lambda)^{-1}$, so that $g(t)(f(t) - \lambda) \equiv 1$ on K. Then $g(u) \circ (f(u) - \lambda I) = [g \cdot (f - \lambda)](u) = I$ and $(f(u) - \lambda I) \circ g(u) = I$ as well. That is, $f(u) - \lambda I$ is invertible and $\lambda \notin Sp(f(u))$.

Conversely, assume that $\lambda \notin Sp(f(u))$, so $f(u) - \lambda I$ is invertible and, by the same proof with Neumann series as when we checked that $Sp(u)$ is closed, we can find $\varepsilon > 0$ such that v is invertible for any $v \in \mathcal{L}(E)$ such that $|||v - f(u) + \lambda I|| \leq \varepsilon$. Let f_n be polynomial functions that tend to f uniformly on K; then we know that the $f_n(u)$ converge in norm to $f(u)$. In particular, we get that for n large enough, $v = f_n(u) - \lambda' I$ is invertible as soon as $|\lambda' - \lambda| \leq \varepsilon/2$. Then $\lambda' \notin Sp(f_n(u)) = f_n(K)$. In other words, $f_n(K)$ stays at distance at least $\varepsilon/2$ from λ, and hence $\lambda \notin f(K)$; (3.8) follows. [Check this proof because Paulin uses a different one.]

Next consider (3.9). If f is real-valued (on K), $f(u)$ is self-adjoint because $f(u)^* = \overline{f}(u) = f(u)$. If $f(u)$ is self-adjoint then its spectrum is real and since the spectrum is $f(K)$, f is real (on K). For the positivity statement, it looks like we have to give the solution of the exercise below (2.27). First suppose $f \geq 0$. Then $v = f(u)$ is self-adjoint, and its spectrum is contained in $[0, +\infty)$. Then by (2.20) $\langle v(x), x \rangle \geq 0$ for all $x \in E$, which is the definition of v nonnegative. Conversely, first observe that

$$\text{if } v \in \mathcal{L}(E) \text{ is nonnegative, then it is self-adjoint} \tag{3.17}$$

This requires a small computation. We know that $\langle v(x), x \rangle \geq 0$ for all x, and we want to check that $\langle v(x), y \rangle = \langle x, v(y) \rangle$ for all x and y. Set $a(x, y) = \langle v(x), y \rangle$ to save notation.

We want $a(x,y) = \overline{a}(y,x)$. By assumption, $a(x+y,x+y) \geq 0$. We expand and get that $a(x,y)+a(y,x) \in \mathbb{R}$. The same with $a(ix+y)$ yields $a(ix,y)+a(y,ix) = i[a(x,y)-a(y,x)] \in \mathbb{R}$. An easy exercise now. We apply (3.17) to $v = f(u)$ and get that it is self-adjoint (and nonnegative). Then by (2.20) its spectrum lies in $[0,+\infty)$. But the spectrum is $f(K)$, so $f \geq 0$ on K.

For (3.10), if f does not vanish, we have seen that by functional calculus $\frac{1}{f}(u)$ is an inverse for u. Conversely, if $f(u)$ is invertible, 0 does not lie in its spectrum, which is $f(K)$.

Next (3.11) follows, for instance, from the fact that φ is an isometry.

We are left with the story (3.12) about eigenvalues and eigenspaces. Suppose λ is an eigenvalue for u and choose any eigenvector x (so $u(x) = \lambda x$). Then for any polynomial, $P(u)(x) = P(\lambda)x$. And by taking limits, this stays true for continuous functions on K: $f(u)(x) = f(\lambda)x$; (3.12) follows.

[End of the spectral theorem for bounded self-adjoint operators] $\square$

3.3 Essential spectrum of a self-adjoint operator

The other, mysterious piece of the spectrum (eigenvalues are not mysterious, but they don't always exist).

Here and below, E is a (nontrivial) Hilbert space and $u \in \mathcal{L}(E)$ is self-adjoint.

Definition 3.5. *The essential spectrum of u is the set $Sp_{ess}(u)$ of $\lambda \in \mathbb{C}$ such that there exists a sequence $\{x_k\}$ in the unit sphere of E, such that*

$$\lim_{k\to+\infty} u(x_k) - \lambda x_k = 0 \tag{3.18}$$

and

$$\{x_k\} \text{ has no convergent subsequence.} \tag{3.19}$$

We need to explain (3.19). Certainly (3.19) forbids us to take $\{x_k\}$ in a fixed vector subspace of finite dimension. On the other hand, if λ is an eigenvalue with infinite multiplicity, it is fairly easy to see that $u = \lambda I$ on a closed subspace of infinite dimension, and then we can take for $\{x_k\}$ an orthonormal system and $\lambda \in Sp_{ess}(u)$.

On the other hand, (3.18) has some flexibility: the x_k are not really required to be eigenvectors, but only approximate eigenvectors.

Exercise. Check that if there is an invertible operator $v \in \mathcal{L}(E)$ such that $u_1 = v^{-1}uv$ then u and u_1 have the same essential spectrum. So the notion has a good chance to have a geometrical meaning. Also, we have the following.

Theorem 3.6. *For $u \in \mathcal{L}(E)$ self-adjoint, denote by $Vp(u)$ the set of eigenvalues for u. Then*

$$Sp(u) = Vp(u) \cup Sp_{ess}(u).$$

In addition, $Sp(u) \setminus Sp_{ess}(u)$ is the set of eigenvalues of u that have a finite multiplicity and are isolated points of $Sp(u)$.

We have seen already (3.18), because we said that λ is an approximate eigenvalue when there is a sequence $\{x_k\}$ in the unit sphere that satisfies (3.18). We even called Weyl's spectrum, and denoted by $\sigma(u)$, the set of approximate eigenvalues. So we immediately know (from (2.26)) that $Sp_{ess}(u) \subset \sigma(u) \subset Sp(u)$. Recall also that in fact $\sigma(u) = Sp(u)$ by (2.27) and because u is self-adjoint.

Let us check that every point of $Sp(u) \setminus Sp_{ess}(u)$ is isolated in $Sp(u)$. We take a point $\lambda \in Sp(u)$ which is not isolated and show that $\lambda \in Sp_{ess}(u)$. Let $\{\lambda_k\}$ be a sequence in $Sp(u) \setminus \{\lambda\}$ that tends to λ; we want to find a nice unit vector x_k for every k.

Since $u = u^*$, Weyl's theorem mentioned in (2.27) (he proved a few other theorems) says that here $\sigma(u) = Sp(u)$. So λ_k is an approximate eigenvalue for u. In particular we can pick a unit vector x_k such that $||u(x_k) - \lambda_k x_k|| \leq 2^{-k}|\lambda - \lambda_k|$, say. So clearly we have (3.18), but we still need to show that no subsequence of $\{x_k\}$ converges. So we suppose that some subsequence converges, which we immediately denote by $\{x_k\}$ to simplify the notation. Set $x = \lim_{k\to+\infty} x_k$; we get that $u(x) = \lambda x$ (since u is continuous), and also $||x|| = 1$. Now we shall use the fact that the x_k are so close to being eigenvalues for different eigenvectors that they cannot be close to each other:

$$\begin{aligned}(\lambda - \lambda_k)\langle x_k, x\rangle &= (\lambda - \lambda_k)\langle x_k, x\rangle + \langle x_k, [u - \lambda I]x\rangle \\ &= (\lambda - \lambda_k)\langle x_k, x\rangle + \langle [u - \lambda I]x_k, x\rangle \\ &= \langle u(x_k) - \lambda_k x_k, x\rangle\end{aligned} \tag{3.20}$$

by self-adjointness. Now by Cauchy-Schwarz

$$|\lambda - \lambda_k|\,|\langle x_k, x\rangle| \leq ||u(x_k) - \lambda_k x_k|| \leq 2^{-k}|\lambda - \lambda_k|$$

and hence $|\langle x_k, x\rangle| \leq 2^{-k}$. This does not happen for a sequence of unit vectors that tends to x. So there was no convergent subsequence, and finally our non-isolated λ lies in the essential spectrum.

Next we need to check that for self-adjoint bounded operators,

$$\text{every isolated point of } Sp(u) \text{ is an eigenvalue} \tag{3.21}$$

[but it could have infinite multiplicity and then it lies in the essential spectrum too]. Let us be brutal. Let μ be an isolated eigenvalue, and define f on K by $f(\mu) = 1$ and $f(\lambda) = 0$ on the rest of K. This is a continuous function on K, so we have $f(u)$. In addition, $(u-\mu I)\circ f(u)$ corresponds to the function $(t - \mu)f(t) = 0$ is null. But $f(u) \neq 0$ because $f(\mu) = 1$. So it has a nontrivial image, and this image is contained in the kernel of $u - \mu I$. So μ is an eigenvalue.

We are about finished. We checked that $Sp(u) = Vp(u) \cup Sp_{ess}(u)$. Then we said that points of $Sp(u)\setminus Sp_{ess}(u)$ are isolated, hence eigenvalues. The multiplicity is finite (otherwise $\lambda \in Sp_{ess}(u)$), so we have a direction.

We still need to see that if λ is an eigenvalue with finite multiplicity and is isolated in the spectrum, then it does not lie in $Sp_{ess}(u)$. Notice that $E_\lambda = Ker(u-\lambda I)$ is finite-dimensional (and closed). Since u preserves it, it also preserves $V = E_\lambda^\perp$, and we can study $v = u_{|V}$.

Observe first that $Sp(v) \subset Sp(u)$. Indeed, suppose that $u - \lambda I$ is invertible, then we claim that its restriction to V is invertible too. It is clearly injective, and for the surjectivity we observe that if $y \in V$ and $x \in E$ is such that $(u - \lambda I)(x) = y$, the part of x that lies in $V^\perp = E_\lambda$ does not contribute to $(u - \lambda I)(x) = y$. Here we are using that E_λ is closed and the sum is direct.

Return to our λ with finite multiplicity. Suppose first that $\lambda \in Sp(v)$. Then it is isolated in $Sp(v)$ (it is already isolated in the larger $Sp(u)$), hence it is an eigenvalue; impossible by definition of V. So $\lambda \notin Sp(v)$, and $u - \lambda I$ is invertible in V. In particular there exists $c > 0$ such that $||u(x) - \lambda x|| \geq c||x||$ for $x \in V$.

Recall we need to check that λ does not lie in $Sp_{ess}(u)$. So let $\{x_n\}$ be a sequence in the unit sphere such that $||u(x_n) - \lambda x_n||$ tends to 0; we just need to show that we can extract a convergent subsequence of $\{x_n\}$. Write $x_n = y_n + z_n$, with $y_n \in E_\lambda$ and $z_n \in V$. Since E_λ is finite dimensional (and $||y_n|| \leq ||x_n|| = 1$), we can extract a subsequence for which y_n converges to a limit y. Also, $u(z_n) - \lambda z_n = u(x_n) - \lambda x_n$ tends to 0 too, but since $||u(z_n) - \lambda z_n|| \geq c||z_n||$, we see that z_n tends to 0. So our subsequence converges, and this completes our proof of Theorem 3.6. □

3.4 Spectral resolution of a self-adjoint operator

At this point let us switch notation and call H (instead of E) our favorite Hilbert space. We may need the letter E for other things.

At the end of this subsection we get a nice description of any self-adjoint operator, in "spectral terms". The idea is still as when we diagonalize, to cut E into orthogonal subspaces where u acts simply, but we may have uncountably many pieces so we'll use measures.

Notation: when $E \subset H$ is a closed subspace, we call p_E the orthonormal projection on E.

Recall that we only consider continuous projections, i.e., operators $p \in \mathcal{L}(H)$ such that $p^2 = p$, and then there are two closed spaces $F = p(H)$ and $N = Ker(p)$, and p is the projection on F in the direction N; that is, $p = 0$ on N and $p = I$ on F. We shall mostly consider **orthogonal projections**, i.e., projections p such that $N = F^\perp$. Those are also the projection p such that $p^* = p$ (exercise).

Definition 3.7. *A* **resolution of the identity** *on the Hilbert space H is is a family P_λ, $\lambda \in \mathbb{R}$, of orthogonal projections, such that*

$$P_\lambda \circ P_\mu = P_{\min(\lambda,\mu)} \quad \text{for } \lambda, \mu \in \mathbb{R} \tag{3.22}$$

$$P_\lambda = 0 \text{ for } \lambda \text{ small enough, and } P_\lambda = I \text{ for } \lambda \text{ large enough;} \tag{3.23}$$

$$\lim_{\mu \to \lambda_+} P_\mu(x) = P_\lambda(x) \quad \text{for } \lambda \in \mathbb{R} \text{ and every } x \in H. \tag{3.24}$$

Comments. For me resolution of the identity means that you cut I into sums of projections $\sum_j [P_{\lambda_{j+1}} - P_{\lambda_j}]$ (with as many pieces as you want).

Calling $F_\lambda = P_\lambda(H)$ the image, (3.22) is a different way of asking that $F_\lambda \subset F_\mu$ for $\lambda < \mu$. We authorize $P_\lambda = P_\mu$ for $\lambda < \mu$ (and then $P_t = P_\lambda$ for $\lambda \leq t \leq \mu$).

Next, (3.24) can be expressed as "$P_\mu(x)$ converges strongly to P_λ" when μ tends to λ_+.

Finally we simplified (3.23) because we are only considering bounded operators, but the true definition that I like would only require that P_λ tends strongly to 0 when λ tends to $-\infty$, and P_λ tends strongly to I when λ tends to $+\infty$. This will make no difference here.

In (3.24) we chose the continuity on the right. This is, I think, due to a standard practice in probability/measure theory. But if we wanted left continuity we could consider $I - P_\lambda$ and reverse the time.

Basic example (but to be refined later): let $u \in \mathcal{L}(H)$ be compact and self-adjoint. For each $\lambda \in \mathbb{R}$, call E_λ the eigenspace for λ, and then for $\mu \in \mathbb{R}$, call F_μ the orthogonal sum of all the E_λ, $\lambda \leq \mu$. Call P_μ the orthogonal projection on F_μ. It is easy to check that this is a resolution of identity. This uses the "diagonalization" of u proved above (because the sum of all E_λ is the whole space!) We could not do that for general bounded self-adjoint u, because the eigenspaces leave out a big chunk of H; we'll find something else.

Now we want to do a bit of measure theory. Here is a theorem on **Stiljes measures**.

Theorem 3.8. *Let $F : \mathbb{R} \to \mathbb{R}$ be bounded, nondecreasing, and continuous from the right. Assume in addition that $\lim_{t\to-\infty} F(t) = 0$ (a normalization). Then there is a unique positive Borel measure μ on $\mathbb{R}$ such that*

$$\mu((-\infty, a]) = F(a) \quad \text{for } a \in \mathbb{R}. \tag{3.25}$$

Comments. If you constructed Lebesgue's measure, probably you know how to prove this. But we won't. Notice that F is allowed to have (positive) jumps, and then μ has Dirac masses. Here we will have $F(m) = 0$ for some m and $F(t) = F(M)$ for $t \geq M$ (for some $M \geq m$), and then it is easy to check that $\mu(\mathbb{R} \setminus [m, M]) = 0$. We call μ the Stiljes measure associated to F (and some times denoted by dF, but beware that it may have singular parts). There is an easy converse: every finite positive measure on $\mathbb{R}$ is a Stiljes measure (can be written like this). Obviously, take $F(a) = \mu((-\infty, a])$ and check the properties above. The right continuity is mostly a convention (much preferred by the probabilists, so let us follow).

We would like something similar, but with projections, so we proceed as follows. Let $\{P_\lambda\}$ be a resolution of the identity. For every $x \in H$, we define the function F_x by $F_x(\lambda) = \langle P_\lambda(x), x\rangle$. Notice that $F_x(\lambda)$ tends to 0 when $\lambda \to -\infty$, and to $||x||^2$ when $\lambda \to +\infty$, and more importantly that F_x is nondecreasing, because for $\lambda < \mu$,

$$\langle P_\lambda(x), x\rangle = ||P_\lambda(x)||^2 \leq ||P_\mu(x)||^2 = \langle P_\mu(x), x\rangle,$$

where we used the self-adjointness of the projections (or just brutal computations with orthogonal decompositions). So for each x there is a Stiljes measure associated to x (and in fact with total mass $||x||^2$). Let us elaborate a little and construct operators from these measures (where we integrate a function f).

Proposition 3.9. *Let $\{P_\lambda\}$ be a resolution of the identity on the Hilbert space H. Then for each $f \in C_b(\mathbb{R})$ (bounded and continuous on $\mathbb{R}$) there is a unique bounded operator $w = w_f$ such that*

$$\langle w_f(x), x\rangle = \int_{\lambda\in\mathbb{R}} f(\lambda) dF_x(\lambda) \tag{3.26}$$

(the integral against f of the Stiljes measure associated to $F_x(\lambda) = \langle P_\lambda(x), x\rangle$). Moreover, w is self-adjoint when f is real-valued and nonnegative when $f \geq 0$.

This is nice. But only called a proposition because to be honest, the converse will be more interesting. If we assume (3.23), the measures F_x are all supported on some interval $[m, M]$, and we don't need f to be bounded. That is, w_f is defined when $f \in C(\mathbb{R})$. Not surprising, because the values of f on $\mathbb{R} \setminus [m, M]$ won't matter. The following notation will help when we apply the proposition:

$$w_f = \int_{\mathbb{R}} f(\lambda) dP_\lambda. \tag{3.27}$$

We did not define this directly, because we fear to do operator-valued measures, but what we defined instead was a list of "projections" of this operator: all the numbers $\langle w_f(x), x\rangle$, which at the end amounts to the same thing, but makes sense more easily. Notice that if u is linear and you know all the $\langle u(x), x\rangle$, then you can also compute all the $\langle u(x), y\rangle$, and this gives you u. Exercise: do the computation; this looks like computing a bilinear form from its quadratic form, and if you are not used to this start when H and u are real. Hint: You will need $\langle u(x + y), x + y\rangle$ and $\langle u(x + iy), x + iy\rangle$.

We are ready for the proof. The uniqueness follows from what we just said (also called polarisation). For the existence, we start from what we have: for each $x \in H$, call $q(x) = \int f(\lambda)dF_x(\lambda) = \int f(\lambda)d\mu_x(\lambda)$ (if you prefer this notation). We also know that $|q(x)| \leq \int |f| d\mu_x \leq ||f||_\infty \mu_x(\mathbb{R}) \leq ||f||_\infty ||x||^2$. Then we check linearity: we set

$$a(x, y) = \frac{1}{2}\big(q(x + y) - q(x) - q(y)\big) + \frac{i}{2}\big(q(x + iy) - q(x) - q(y)\big);$$

this is a sesquilinear (linear in x, semilinear in y) form. This is unpleasant, but if we have the right formulas it cannot fail. The points is that we can expand things like $F_{x+y}(\lambda) = \langle P_\lambda(x + y), x + y\rangle$ using the linearity of each P_λ, and then compute. Once we have the sesquilinearity (and boundedness) of a, we can conclude that it is given by a linear operator (use the Riesz theorem to define $w(x)$ by its effect on all $y \in H$).

Finally the fact that w is self-adjoint when f is real and nonnegative when $f \geq 0$ is easy to check, because we are precisely given the $\langle w(x), x\rangle$ (and can compute the $\langle w(x), y\rangle$ if needed). The reader guessed that we'd skip the details. □

Comments.

• The operator $w = w_f$ of the proposition can be seen as roughly diagonalized by the description (3.27), in the sense that for instance, if f is close to the constant α on the interval $I = (a, b]$, then on the space W which is the orthonormal complement of $P_a(H)$ in $P_b(H)$, the operator looks a lot like αI, as we can check by computing $\langle w(x), x\rangle$ for $x \in W$. Notice that then we only integrate on $(a, b]$.

• We only took f continuous in (3.26) and (3.27), but we could take f bounded (but Borel-measurable to make sure it will be measurable for all the μ_x) and we still get a bounded operator w_f, defined for $f \in \mathcal{L}^\infty(\mathbb{R})$. For me, $\mathcal{L}^\infty(\mathbb{R})$ will be the set of bounded Borel functions on $\mathbb{R}$, with the sup norm. I'll refrain to say L^∞ here, because usually one works modulo functions that vanish almost everywhere, and here it is a little delicate because we are using a lot of measures μ_x at the same time. But this is not a serious issue.

Now we head for the converse of the proposition: the resolution of the identity associated to a self-adjoint operator $u \in \mathcal{L}(H)$, and then how to write u and functions of u as in the proposition above. We cut the statement in three, but the theorems are linked and we will prove them together.

Theorem 3.10 (Spectral resolution of u)**.** *Let $u \in \mathcal{L}(H)$ be self-adjoint. There is a unique resolution of the identity $\{P_\lambda\}$, $\lambda \in \mathbb{R}$ (on H) such that for $f \in C(Sp(u))$,*

$$f(u) = \int_{\lambda \in Sp(u)} f(\lambda) dP_\lambda. \tag{3.28}$$

Here the left-hand side is the continuous functional calculus on u (defined in Theorem 3.3), and the right-hand side is using the formula (3.27).

We integrated on $Sp(u)$ because f is not defined on the rest of $\mathbb{R}$, but we'll see that dP_λ (and we should say the $d\langle P_\lambda x, x\rangle$) never charge(s) $\mathbb{R} \setminus Sp(u)$.

We call $\{P_\lambda\}$ the spectral resolution of u. We'll see soon that it allows us to compute bounded functions of u (and not only continuous ones).

How do we get it? We will find it more convenient to prove a good part of the following thing first:

Theorem 3.11 (Bounded functional calculus for u, definition)**.** *Let $u \in \mathcal{L}(H)$ be self-adjoint. Call $K = Sp(u)$. There is a unique morphism ψ from the C^*-algebra $\mathcal{L}^\infty(K)$ (the set of bounded Borel functions on K, with the sup norm) to $\mathcal{L}(H)$, such that $\psi(id) = u$, and which has the following continuity property*

(3.29) *if $g \in \mathcal{L}^\infty(K)$ is the pointwise limit (everywhere) of a bounded sequence $\{g_k\}$ in $\mathcal{L}^\infty(K)$, then the $\psi(g_k)$ converge strongly to $\psi(g)$.*

Some comments before we continue with the properties of ψ. We are still using the space $\mathcal{L}^\infty$ of bounded functions (with the strong sup norm), which means that we do not try yet to determine whether two bounded functions are equivalent, but I feel ready to announce that two functions that are equivalent for all the measures $\mu_x = d\langle P_\lambda, x\rangle$ associated to the resolution of the identity of Theorem 3.10 will give the same $\psi(f)$, hence we could improve a tiny bit the continuity (3.29).

You know that $C(K)$ is not dense at all in $\mathcal{L}^\infty$ (or the reasonable L^∞), so the norm continuity of ψ is not enough to prove the uniqueness of ψ. And also, it is important to be allowed to use (3.29) to compute lots of $\psi(f)$.

Recall, the strong convergence means that $\psi(g_k)(x)$ converges (in H) to $\psi(g)(x)$ for every $x \in H$ (but $|||\psi(g_k)(x) - \psi(g)(x)|||$ does not need to tend to 0).

Finally,

Theorem 3.12 (Bounded functional calculus, properties)**.** *Let $u \in \mathcal{L}(H)$ be self-adjoint, and set $K = Sp(u)$. Let ψ be as in Theorem 3.11. Then*

(3.30) ψ *extends the continuous functional calculus φ of Theorem 3.3;*

$$|||\psi(f)||| \leq \sup_{t \in K} |f(t)|; \tag{3.31}$$

(3.32) *$\psi(f)$ is self-adjoint (resp. non-negative) when f is real-valued (resp. ≥ 0) on K;*

$$\psi(f) \circ v = v \circ \psi(f) \text{ for all } v \in \mathcal{L}(H) \text{ such that } v \circ u = u \circ v; \tag{3.33}$$

$$\begin{aligned} &\text{if } \lambda \text{ is an eigenvalue for } u, \text{ then } \psi(\lambda) \text{ is an eigenvalue for } \psi(f) \\ &\qquad\qquad \text{and } Ker(u - \lambda I) \subset Ker(\psi(f) - f(\lambda)I). \end{aligned} \tag{3.34}$$

No big surprise here, I hope. From the end of the proof on, we may again write $f(u)$ instead of $\psi(f)$, but for the moment let us not do that and reserve the notation for the continuous functional calculus.

The functorial identities of ψ (many things are preserved) are the same as for φ a long time ago; we added the better convergence theorem to accommodate bounded functions.

The uniqueness of ψ comes from the fact that it must be an extension of φ (by uniqueness of φ), our convergence condition, and the fact that every bounded Borel function is the pointwise limit of some sequence of (bounded) continuous functions. Note however that this fact requires some proof, which we will not do here!!

How do we define ψ? For $x, y \in H$ and $f \in C(K)$, we have a definition of $\langle f(u)(x), y \rangle$ (coming from the continuous functional calculus). For fixed x and y, this defines a continuous linear form on $\mathbb{C}(K)$, hence there is a finite Borel measure $\mu = \mu_{x,y}$ such that

$$\int_K f d\mu_{x,y} = \langle f(u)(x), y \rangle \text{ for } f \in C(K).$$

Then we can say a bit more about the $\mu_{x,y}$. First,

$$||\mu_{x,r}|| \leq ||x|| \, ||y|| \tag{3.35}$$

because $||\mu_{x,r}||$ (the total variation of the measure) is the norm of the linear form, and by functional calculus; then μ has the expected sesquilinear and symmetry dependence on

x and y (by functional calculus again). Because of this, we may now observe that for $g \in \mathcal{L}^\infty = \mathcal{L}^\infty(K)$, the application

$$(x, y) \to \int_K g d\mu_{x,y}$$

is a bounded sesquilinear form on H, so by Riez (or Riez-Fréchet?) it is given by a (unique) operator $\psi(g) \in \mathcal{L}(H)$. That is,

$$\int_K g d\mu_{x,y} = \langle \psi(g)(x), y \rangle \text{ for } x, y \in H.$$

This gives a definition of ψ. Then there are some various things to verify: linearity, complex conjugation, norm, etc. Let me just mention a few.

We put stress on the continuity condition. Now we have to check it: we suppose that $g \in \mathcal{L}^\infty(K)$ is the pointwise limit of a bounded sequence $\{g_k\}$ in $\mathcal{L}^\infty(K)$, and we want to prove the strong convergence of the $\psi(g_k)$ to $\psi(g)$. What we can easily get is that for $x, y \in H$, the $\langle \psi(g_n)(x), y \rangle = \int_K g_n d\mu_{x,y}$ converge to $\int_K g d\mu_{x,y} = \langle \psi(g)(x), y \rangle$ (by dominated convergence for a single measure). This is "weak convergence". We shall use the following classical lemma, which is often useful.

Lemma 3.13. *Let $\{z_k\}$ be a sequence in a Hilbert space H, which converges weakly to $z \in H$, in the sense that*

$$\lim_{k \to +\infty} \langle z_k, y \rangle = \langle z, y \rangle \text{ for every } y \in H. \tag{3.36}$$

Suppose in addition that we lose no mass, i.e.,

$$\lim_{k \to +\infty} ||z_k|| = ||z||. \tag{3.37}$$

Then $\lim_{k \to +\infty} ||z_k - z|| = 0$.

Easy proof, since $||z_k - z||^2 = ||z_n||^2 + ||z||^2 - 2Re\langle z_k, z \rangle$, so we even only need $y = z$. $\square$

Returning to our sequence $\{\psi(g_n)(x)\}$, we see that it is enough to check that $||\psi(g_n)(x)||$ tends to $||\psi(g)(x)||$; fortunately we can use the same trick again:

$$||\psi(g_n)(x)||^2 = \langle \psi(g_n)(x), \psi(g_n(x)) \rangle = \langle \psi(g_n)^* \psi(g_n)(x), x \rangle = \langle \psi(|g_n|^2)(x), x \rangle,$$

and the last term goes to the limit well, because $|g_n|^2$ converges pointwise to $|g|^2$, and then we can proceed backwards.

Let me leave out the proofs of (3.31) and (3.32). Maybe you were surprised by the statement of (3.33), but the point is that the statement is true for continuous functions f, even though we did not state it this way, because it holds for polynomial functions and then extends by continuity. Here it is the same, but we use the strong continuity of the definition of ψ. Finally (3.34) goes as before (true for continuous functions, then go to the limit).

Now we worry about Theorem 3.10. We start with the uniqueness. Consider the formula (3.28). More precisely it means that for all $f \in C(K)$ and $x, y \in H$,

$$\langle f(u)(x), x\rangle = \int_{\lambda\in K} f(\lambda) d\langle P_\lambda(x), x\rangle.$$

It follows that the Stiljes measure $\mu_x = d\langle P_\lambda(x), x\rangle$ is uniquely determined, or equivalently $\langle P_\lambda(x), x\rangle$ is uniquely determined and again by polarisation we can recover $\langle P_\lambda(x), y\rangle$, which yields P_λ.

Now the existence. We waited all this time because now we have ψ, and it is natural to take $P_\lambda = \psi(\chi_\lambda)$, where χ_λ is the characteristic function of $(-\infty, \lambda]$. If we did not want to wait, we could have guessed that we should take for P_λ a limit of operators $f(u)$, where $f(t) = 1$ for $t \leq \lambda$ and $f(t) = 0$ for t a little larger than λ. The remaining verification (that this works) is not surprising; we have the functional calculus, so for instance $P_\lambda \circ P_\mu$, which corresponds to the product $\chi_\lambda \chi_\mu$, is equal to $P_{\min(\lambda,\mu)}$. □

3.5 What next?

We'll stop here because of lack of time, but there are many more things to say. First we studied self-adjoint bounded operators (we have examples later), but other classes of operators are interesting. Such as unitary or (see later) unbounded operators.

You'd be surprised: there are many circumstances where you want to define and use functions of an operator (or even, in some cases, of a few operators (say, that commute with each other)).

Same general comments with C^*-algebra: there is a whole interesting world to discover for those who want. Good if you are bored with usual functions because they commute, but not only. See for instance books of Dixmier, Connes, Yoshida.

Related but not discussed here, semigroups (I have a good but tough memory of Yoshida).

Let me mention representations results, which I never know exactly how to interpret. The simplest says that every Hilbert space is isometric to an $L^2(I, \mu)$, and even with a measure on I which is a counting measure. I think it means that you won't ever get more information on your space H by knowing that it is an $L^2(\mu)$ (except of course if you are interested in results that relate to this μ precisely).

The second one says that any commutative C^*-algebra is in fact a $C_0(A)$ (continuous functions on a locally compact space A that tend to 0), if you want to allow algebras without a unit, and a $C(K)$, K compact if you have a unit. In this case this looks simpler (but the theorem says that any proof in $C(K)$ has to extend).

The last case is a theorem of Gelfand-Naimark which says that if A is a C^*-algebra, it is isomorphic (through an isometry) to a closed subalgebra of a $\mathcal{L}(H)$. Maybe the right way to interpret is that the structure of C^*-algebra is well concieved (general results on the structure of $\mathcal{L}(H)$ should have a C^*-algebra proof). Unless the devil is in the "closed subalgebra" part.

4 Some results of measure theory/analysis

Suddently you'll be surprised to see some sections in French. There are good reasons for this: the fact that preparing a class takes some time (hence an incentive to talk about what I know a bit more) and using some already written notes (in French) saves even more time. So the following sections are brutal inclusions of some other text. But contrary to the part above where I relied on Paulin's notes and added errors, I own the copyright for what follows (errors included). I'll still try to butcher the notes a bit, to make the result shorter; the original is on my web page : https://www.imo.universite-paris-saclay.fr/∼gdavid/CoursM2.pdf

4.1 Lusin's theorem

Theorem 4.1. *On se donne un espace métrique localement compact E (par exemple, $E = \mathbb{R}^n$), une mesure finie μ sur E, muni des boréliens, et une fonction f définie sur E, borélienne, et à valeurs dans $\mathbb{R}^m$, ou $\mathbb{C}^m$, ou $\overline{\mathbb{R}}$. Mais si elle est à valeurs dans $\overline{\mathbb{R}}$, on la suppose finie presque-partout. Alors pour tout $\varepsilon > 0$, il existe g continue (et finie partout), telle que $\mu(\{x \in E\,;\, f(x) \neq g(x)\}) \leq \varepsilon$.*

Pourquoi j'aime bien: on vous dit tout le temps qu'une fonction Borélienne, ou même un ensemble Borélien, ça peut être très compliqué. Oui mais, si on a le droit de jeter à la poubelle un morceau aussi petit qu'on veut, ce qui reste est simple.

On commence par quelques réductions faciles. On peut supposer que f est finie partout (la remplacer par 0 sur un ensemble de mesure nulle ne change pas la conclusion), puis à valeurs réelles (appliquer l'argument à chaque coordonnée et avec ε/n, puis prendre l'union des mauvais ensembles), puis même à valeurs dans $]-M, M[$ (choisir M assez grand pour que $\mu(\{x \in E\,;\, |f(x)| \geq M\}) \leq \varepsilon/2$), et enfin à valeurs dans $[0, 1[$ (considérer $[M + f(x)]/2M$.)

Par un argument standard de théorie de la mesure, on peut écrire $f = \sum_{k\geq 1} 2^{-k}\,\mathbb{1}_{E_k}$, avec des ensembles E_k boréliens. Les sommes partielles sont les approximations standard de f par des fonctions à valeurs dans $2^{-k}\mathbb{N}$, et une manière d'écrire la somme directement est de dire que pour tout x, $f(x) = \sum_{k\geq 1} 2^{-k}\,\mathbb{1}_{E_k}$ est l'écriture standard de $f(x)$ en base 2. Et sauf erreur de calcul, $\mathbb{1}_{E_k}(x) = e(2^k f(x)) - 2e(2^{k-1} f(x))$, où $e(y)$ est la partie entière de y; d'où la mesurabilité.

Maintenant, la mesure μ étant régulière, on sait que pour tout k, on peut trouver un compact K_k et un ouvert V_k tels que $K_k \subset E_k \subset V_k$ et $\mu(V_k \setminus K_k) < \varepsilon_k$, où l'on peut choisir $\varepsilon_k > 0$ très petit, par exemple, $\varepsilon_k = 2^{-k}\varepsilon$.

Par un théorème de topologie (Urysson, je crois), il existe une fonction continue g_k telle que $g_k(x) = 1$ sur K_k, $g_k(x) = 0$ hors de V_k, et $0 \leq g_k(x) \leq 1$ sur $V_k \setminus K_k$. D'ailleurs, ici l'espace est métrique et il est facile de donner une formule pour $g_k(x)$ en fonction de la distance de x au complémentaire de V_k. La série $\sum_k 2^{-k} g_k$ converge normalement, et sa somme g est continue. Il reste à voir que $g(x) = f(x)$ hors d'un ensemble Z tel que $\mu(Z) \leq \varepsilon$.

Mais on peut prendre $Z = \cup_k [V_k \setminus K_k]$, et utiliser le fait que $\mu(V_k \setminus K_k) < \varepsilon_k$. □

Le théorème a un corollaire facile: la fonction f est une limite simple presque-partout de fonctions continues. Mais à mon sens c'est moins impressionnant.

4.2 Egorov's theorem

Theorem 4.2. *Soit $(E, \mathcal{A}, \mu)$ un espace de probabilité, et soit $\{\varphi_n\}$ une suite de fonctions mesurables, disons de E dans $\mathbb{R}$, telle que $\lim_{n\to+\infty} \varphi_n(x) = 0$ pour μ-preque tout $x \in E$. Alors, pour tout $\varepsilon > 0$, il existe $F \subset E$ mesurable tel que $\mu(E \setminus F) \leq \varepsilon$ et*

$$\lim_{n\to+\infty} \varphi_n(x) = 0 \text{ uniformément sur } F.$$

J'ai pris une limite nulle pour simplifier. Si la suite $\{\varphi_n\}$ tend vers $\varphi \neq 0$, on a le même résultat de convergence uniforme sur un grand ensemble en appliquant le théorème à $\varphi_n - \varphi$.

Donc on doit jeter un petit ensemble pour rendre la limite uniforme. C'est sans doute à cause de cette ressemblance avec Lusin ci-dessus que je confonds souvent.

Quitte à remplacer les φ_n par 0 sur un ensemble de mesure nulle, et ajouter cet ensemble à $E \setminus F$, on peut supposer que $\lim_{n\to+\infty} \varphi_n(x) = 0$ partout. Posons $\alpha_k = 2^{-k}$. Fixons k et notons

$$A_{N,k} = \big\{x \in E\,;\, |\varphi_n(x)| \geq \alpha_k \text{ pour au moins un } n \geq N\big\}$$

pour $N \geq 1$. Puisque $\lim_{n\to+\infty} \varphi_n(x) = 0$ partout, l'intersection (visiblement décroissante) des $A_{N,k}$ est vide. Donc il existe $N(k)$ tel que $\mu(A_{N(k),k}) \leq \varepsilon 2^{-k-1}$.

Ensuite posons $Z = \cup_k A_{N(k),k}$, et $F = E \setminus Z$. Alors $\mu(E \setminus F) = \mu(Z) \leq \sum \mu(A_{N(k),k}) \leq \varepsilon$, et il ne reste plus qu'à vérifier (1).

On doit montrer que pour tout $\alpha > 0$, il existe N tel que $|\varphi_n(x)| \leq \alpha$ pour $x \in F$ et $n \geq N$. On choisit k tel que $\alpha_k < \alpha$ et $N = N(k)$. Alors si $x \in F$, x n'est pas dans $A_{N(k),k} \subset Z$, donc $\varphi_n(x) \leq \alpha_k < \alpha$ pour tout $n \geq N$, come souhaité. □

4.3 Maximal functions and he Lebesgue differentiation theorem

On commence par la **fonction maximale de Hardy-Littlewood**.

On se donne une mesure de référence μ, borélienne et positive sur $\mathbb{R}^n$ (penser à la mesure de Lebesgue), et qu'on va aussi prendre localement finie (de Radon). On se donne aussi une seconde mesure (borélienne positive) ν. On pose

$$M_\mu(\nu(x)) = \sup_{r>0} \frac{1}{\mu(B(x,r))} \int_{B(x,r)} d\nu \in [0, +\infty],$$

avec la convention que $0/0 = 0$ quand $\mu(B(x,r)) = 0$. C'est la fonction maximale (de Hardy-Littlewood, ou est-ce seulement dans le cas ou μ est la mesure de Lebesgue qu'ils avaient regardé?) de ν.

On n'utilisera que le cas où μ est la mesure de Lebesgue, mais il n'y a pas de grosse différence tout de suite. Je ne regarde pas la mesurabilité (je vous laisse vérifier; avec la mesure de Lebesgue, qui charge toutes les boules, il n'y a pas de problème).

Le cas qui arrive le plus souvent est celui d'une mesure de densité, quand $d\nu(x) = |f(x)|d\mu(x)$, avec f localement intégrable pour μ, et alors cela donne

$$M_\mu f(x) = \sup_{r>0} \frac{1}{\mu(B(x,r))} \int_{B(x,r)} |f|\, d\mu.$$

Le plus souvent, $d\mu$ est la mesure de Lebesgue, et $M_\mu(\nu)$ est noté ν^*.

On se demande si ces fonctions sont intégrables ou dans des $L^p(d\mu)$. Ca sera utile parce que pas mal de problèmes avec des sup ou des limites dépendent de fonctions maximales; voir plus loin pour au moins un exemple.

Lemma 4.3. *Si ν est une mesure finie, $M_\mu(\nu) \in L^1_{faible}(d\mu)$, avec*

$$\mu\big(\{x \in \mathbb{R}^n\,; M_\mu(\nu(x)) > \lambda\}\big) \leq \frac{C_n \nu(\mathbb{R}^n)}{\lambda} \quad \textit{pour tout } \lambda > 0.$$

Notation: pour $p > 0$, on note

$$||g||_{L^p_{faible}(d\mu)} = \sup_{\lambda>0} \big\{\lambda^p\, \mu\big(\{x \in \mathbb{R}^n\,; |g(x)| > \lambda\}\big)\big\}^{1/p},$$

et ensuite $L^p_{faible}(d\mu)$ est la classe des fonctions mesurables g telles que $||g||_{L^p_{faible}(d\mu)} < +\infty$. C'est bien un espace vectoriel complet, mais attention, $||g||_{L^p_{faible}(d\mu)}$ n'est pas une norme, l'inégalité triangulaire n'est pas vérifiée avec la constante 1. Bien sûr, $L^p \subset L^p_{faible}$, par Tchebyshev.

Noter que le lemme devient faux quand on remplace L^1_{faible} par L^1: prendre la mesure de Lebesgue sur $\mathbb{R}$, et la fonction $f = \mathbf{1}_{[0,1]}$ dont la fonction maximale est en $1/x$ à l'infini.

Corollary 4.4. *Si $f \in L^1(d\mu)$, alors $M_\mu f \in L^1_{faible}(d\mu)$, avec $||M_\mu f||_{L^1_{faible}(d\mu)} \leq C_n ||f||_{L^1(d\mu)}$.*

Démonstration. Le corollaire est immédiat dès qu'on a le lemme. Pour le lemme, on pose $O_\lambda = \{x \in \mathbb{R}^n\,; M_\mu(\nu(x)) > \lambda\}$, et pour tout $x \in O_\lambda$, on se donne une boule $B_x = B(x,r)$ centrée en x telle que $\nu(B_x) > \lambda\mu(B_x)$. On veut appliquer le lemme de recouvrement de Besicovitch, [Comment for the english version. In the french notes, I was trying to use the Besicovitch covering lemma because I like it, but there is a simpler proof with the usual 5-covering lemma of Vitali. See near (4.2) below. Anyway some covering lemma is used. Not described here.] donc on va se contenter de recouvrir $\Omega = O_\lambda \cap B(0,R)$ (pour n'importe quel $R > 0$) par une collection au plus dénombrable de boules B_x, $x \in X \subset \Omega$, qui sont de recouvrement borné comme dans le lemme. Alors

$$\begin{aligned} \mu(\Omega) &\leq \mu\big(\bigcup_{x\in X} B_x\big) \leq \sum_{x\in X} \mu(B_x) \leq \lambda^{-1} \sum_{x\in X} \nu(B_x) = \lambda^{-1} \sum_{x\in X} \int_{\mathbb{R}^n} \mathbf{1}_{B_x} d\nu \\ &= \lambda^{-1} \int_{\mathbb{R}^n} \Big(\sum_{x\in X} \mathbf{1}_{B_x}\Big) d\nu \leq \lambda^{-1} \int_{\mathbb{R}^n} C_1 d\nu = C_1 \lambda^{-1} \nu(\mathbb{R}^n) \end{aligned} \tag{4.1}$$

par définition des boules, puis le corollaire de Beppo-Lévi sur les séries, et enfin le fait que $\sum_{x\in X} \mathbf{1}_{B_x} \leq C_n$ partout. Il ne reste plus qu'à faire tendre R vers $+\infty$ pour récupérer le fait que $\mu(O_\lambda) \leq C_1\lambda^{-1}\nu(\mathbb{R}^n)$. $\square$

On peut vérifier (pas ici) que si $\mu(B(x,r))$ est une fonction continue de (x,r), alors $M_\mu(\nu)$ est semi-continue inférieurement. Autrement dit, que si l'on pose $O_\lambda = \{x \in \mathbb{R}^n\,; M_\mu(\nu(x)) > \lambda\}$ comme ci-dessus, alors O_λ est ouvert pour tout $\lambda > 0$.

Une variante de M_μ est la fonction maximale non centrée définie par

$$M'_\mu(\nu(x)) = \sup_B \frac{1}{\mu(B)} \int_B d\nu \in [0,+\infty],$$

où le sup est pris sur toutes les boules B qui contiennent x. Mattila prend les boules fermées, mais sauf erreur de ma part ou pourrait les prendre ouvertes aussi sans rien changer. La fonction $M'_\mu f = M'_\mu(f d\mu)$ est définie pareillement. Si la mesure μ est doublante, c.-à.-d. s'il existe $C \geq 1$ tel que

$$\mu(B(x,2r)) \leq C\mu(B(x,r)) \text{ pour tout } x \in \mathbb{R}^n \text{ et tout } r > 0,$$

on a le même genre de résultat que plus haut:

$$||M'_\mu(\nu)||_{L^1_{faible}(d\mu)} \leq C(n,\mu)\,\nu(\mathbb{R}^n),$$

et pareil pour les fonctions de L^1:

$$||M'_\mu f||_{L^1_{faible}(d\mu)} \leq C(n,\mu)||f||_{L^1(d\mu)}.$$

Démonstration sans se fatiguer: si B est une boule qui contient x, et si r est son diamètre, elle est contenue dans $B(x,2r)$, donc

$$\frac{1}{\mu(B)} \int_B d\nu \leq \frac{1}{\mu(B)} \int_{B(x,2r)} d\nu \leq C^2 \frac{1}{\mu(B(x,2r))} \int_{B(x,2r)} d\nu \leq C^2 M_\mu(\nu(x)),$$

car $B(x,2r) \subset 3B \subset 4B$, donc $\mu(B(x,2r)) \leq C^2\mu(B)$. On prend le sup et on trouve que $M'_\mu(\nu(x)) \leq C^2 M_\mu(\nu(x))$ partout.

La vraie démonstration raisonnable est d'utiliser le lemme de 5-recouvrement (qui est plus facile à démontrer et marche dans d'autres contextes), et de faire comme plus haut en passant par les $5B$. En plus, ça donne une inégalité plus précise (restricted type inequality):

$$\mu\big(\{x \in \mathbb{R}^n\,; M'_\mu(\nu(x)) > \lambda\}\big) \leq C_n\lambda^{-1}\nu\big(\{x \in \mathbb{R}^n\,; M'_\mu(\nu(x)) > \lambda\}\big).$$

Voici en gros comment ça marche [and this is probably the best proof of Lemma 4.3 for you]. On pose $O'_\lambda = \{x \in \mathbb{R}^n\,; M'_\mu(\nu(x)) > \lambda\}$, et on recouvre O'_λ par des boules $B_j = B(x_j, r_j)$ telles que $\nu(B_j) > \lambda\mu(B_j)$. [Chaque $x \in O'_\lambda$ est contenu dans une telle boule par définition.] Pour ne pas avoir d'ennui avec la taille des boules, occupons-nous seulement pour l'instant de la variante de $M_\mu^{(R)}$ où l'on ne considère que des boules de rayon inférieur à R, où l'on s'est donné R (très grand) à l'avance.

Alors tous les rayons des B_j sont inférieurs à R, et le lemme de 5-recouvrement donne une famille B_j, $j \in J$, de boules disjointes, mais telle que les $5B_j$ recouvre O'_λ. Et

$$
\begin{aligned}
\mu(O'_\lambda) &\leq \mu(\cup_{j\in J} 5B_j) \leq \sum_{j\in J} \mu(5B_j) \leq C^3 \sum_{j\in J} \mu(B_j) \\
&\leq C^3\lambda^{-1} \sum_{j\in J} \nu(B_j) \leq C^3\lambda^{-1}\nu(\cup_j B_j) \leq C^3\lambda^{-1}\nu(O'_\lambda),
\end{aligned}
\tag{4.2}
$$

où la dernière inégalité vient de ce que $M'_\mu(\nu(x)) > \lambda$ pour tout $x \in B_j$, puisque $\nu(B_j) > \lambda\mu(B_j)$ et par définition de M'_μ.

Dans le cas général, on note que l'ensemble O'_λ associé à $M'_\mu(\nu)$ est l'union croissante dénombrable des ensembles associés à $M_\mu^{(R)}(\nu)$, et on passe à la limite dans (4.2).

Theorem 4.5 (Hardy-Littlewood)**.** *Pour $1 < p \leq +\infty$, l'opérateur maximal $f \to M_\mu f$ est borné sur L^p. plus préciément, il existe $C = C(p,n)$ tel que $||M_\mu f||_{L^p(d\mu)} \leq C||f||_{L^p(d\mu)}$. Même énoncé pour la fonction maximale non centrée quand μ est doublante.*

Quand $p = +\infty$, c'est clair (et même $C = 1$). Le cas général s'en déduit, par interpolation avec la continuité $L^1 \to L^1_{faible}$; we won't do that here; maybe see the french notes, or maybe Pascal Auscher will do it. □

Maintenant un théorème (important) qui s'en déduit facilement. Pour changer, on se place dans $\mathbb{R}^n$, avec la mesure de Lebesgue.

Theorem 4.6 (Théorème de différentiation de Lebesgue.)**.** *Pour toute fonction f localement intégrable pour dx, on a*

$$
\lim_{r\to 0} \frac{1}{|B(x,r)|} \int_{B(x,r)} |f(x) - f(y)|dy = 0 \tag{1}
$$

pour presque-tout $x \in \mathbb{R}^n$.

Corollaire souvent utile, même quand $d\mu = dx$: le cas de la fonction caractéristique de B borélien. Essentially the same statement; the density of E (respectively, $\mathbb{R}^n \setminus E$) is 1 (respectively, 0) almost-everywhere on E.

Notons aussi qu'ici on se donne un représentant (quelconque) de f avant d'écrire (1); bien sûr, si on modifie f sur un ensemble de mesure nulle, on modifie la liste des x tels que (1) a lieu, mais pas le théorème.

On a noté $|B(x,r)|$ la mesure de Lebesgue de $B(x,r)$. Une conséquence facile, à cause de l'inégalité triangulaire, est que

$$
f(x) = \lim_{r\to 0} \frac{1}{|B(x,r)|} \int_{B(x,r)} f(y)dy \tag{2}
$$

pour presque-tout $x \in \mathbb{R}^n$.

Démonstration. On peut se contenter du cas où $f \in L^1$, puisque pour regarder ce qui se passe dans $B(0,R)$, on peut remplacer f par $f\,\mathbf{1}_{B(0,2R)} \in L^1$.

Pour f continue, le théorème est trivial. Le cas général sera obtenu, de manière en fait très standard, par un argument de densité et de fonction maximale. Posons, pour $x \in \mathbb{R}^n$ et $r > 0$,

$$\omega_f(x,r) = \frac{1}{|B(x,r)|}\int_{B(x,r)} |f(x) - f(y)|dy, \tag{3}$$

puis

$$\omega_f(x) = \limsup_{r\to 0} \omega_f(x,r). \tag{4}$$

On veut prouver que $\omega_f(x) = 0$ presque-partout. La chose est vraie sur une classe dense, car

$$\omega_f(x) = 0 \text{ partout quand } f \text{ est continue.} \tag{5}$$

Par ailleurs, l'inégalité triangulaire dit que

$$\omega_{f+g}(x,r) \leq \omega_f(x,r) + \omega_g(x,r) \tag{6}$$

pour tout choix de f, g, x, et r, ce qui donne aussitôt

$$\begin{aligned}\omega_{f+g}(x) = \limsup_{r\to 0} \omega_{f+g}(x,r) &\leq \limsup_{r\to 0} [\omega_f(x,r) + \omega_g(x,r)] \\ &\leq \limsup_{r\to 0} \omega_f(x,r) + \limsup_{r\to 0} \omega_g(x,r) \leq \omega_f(x) + \omega_g(x).\end{aligned} \tag{4.3}$$

Enfin, on va utiliser le fait, assez clair, que

$$\omega_f(x) \leq |f(x)| + M_{dx}f(x), \tag{8}$$

où $M_{dx}f$ (qu'on noterait plutôt f^*), est la fonction maximale (centrée) de Hardy-Littlewood. Soient $f \in L^1$ et $\varepsilon > 0$. Choisissons g continue telle que $||f - g||_{L^1} \leq \varepsilon$. Alors

$$\omega_f(x) \leq \omega_{f-g}(x) + \omega_g(x) = \omega_{f-g}(x) \leq |f-g|(x) + M_{dx}(f-g)(x), \tag{9}$$

par (5), donc (9) donne

$$\begin{aligned}|\{x\,;\,|\omega_f(x)| > \lambda\}| &\leq |\{x\,;\,|(f-g)(x)| > \lambda/2\}| + |\{x\,;\,|M_{dx}(f-g)(x)| > \lambda/2\}| \\ &\leq 2||f-g||_{L^1}/\lambda + 2||M_{dx}(f-g)||_{L^1_{faible}}/\lambda \leq 2\varepsilon\lambda + 2C\varepsilon\lambda\end{aligned} \tag{4.4}$$

par le théorème maximal. On fixe λ et on applique ça avec ε aussi petit qu'on veut, et on trouve $|\{x\,;\,|\omega_f(x)| > \lambda\}| = 0$. Finalement, $\omega_f(x) = 0$ presque-partout. □

4.4 Sobolev spaces

Here and below, Ω is an open subset of $\mathbb{R}^n$, which we will often take bounded (I'll try to remember to say when we use this) and connected (otherwise, we can work independently on each connected component of Ω.

We put on Ω the (restriction of the) Lebesgue measure, which I'll denote by dx unless this creates confusion. And $|E|$ will denote the Lebesgue measure of the Borel set E (we won't need non-Borel sets).

Of course $\Delta f = \sum_{j=1}^{n} \frac{\partial^2 f}{\partial x_j{}^2}$, but we want to define a natural space where this is defined. The following brutal definition will be completed after the statement.

Definition 4.7. *Let $1 \leq p \leq +\infty$ The Sobolev space $W^{1,p}(\Omega)$ is the space of functions $f \in L^p(\Omega)$ such that the partial derivatives $\frac{\partial f}{\partial x_j}$ lie in $L^p(\Omega)$. Then $W^{2,p}(\Omega)$ is the space of functions $f \in W^{1,p}(\Omega)$ such that all the derivatives $\frac{\partial^2 f}{\partial x_j \partial x_j}$ lie in $L^p(\Omega)$.*

You already guessed the definition of $W^{k,p}$, $k \geq 0$ integer.

When I mean derivative, I mean in the sense of distribution. Thus, for $f \in L^p$, we say that $\frac{\partial f}{\partial x_j} \in L^p(\Omega)$ when there is a function $g_j \in L^p(\Omega)$ such that

$$\int_\Omega f(x) \frac{\partial \varphi(x)}{\partial x_j} dx = - \int_\Omega g_j(x) \varphi(x) dx \text{ for every } \varphi \in C_c^\infty(\Omega). \tag{4.5}$$

Where $C_c^\infty(\Omega)$ is the set of C^∞ functions with compact support in Ω. That is, we always compute the derivatives in the middle of Ω, not on the boundary. This makes sense because we would have (4.5) if f were C^1 and $g_j = \frac{\partial f}{\partial x_j}$; this would be a consequence of an integration by parts (a simple form of Green's theorem here: we just want to know that for a compactly supported C^1 function φf in Ω, the integral of $\frac{\partial f}{\partial x_j}(\varphi f)$ vanishes. And we can proceed line by line).

There is a natural norm on $W^{1,p}(\Omega)$, which is

$$||f||_{W^{1,p}} = ||f||_{L^p(\Omega)} + ||\nabla f||_{L^p(\Omega)}. \tag{4.6}$$

Here I abuse notation, by calling $\nabla f(x)$ the vector in $\mathbb{R}^n$ whose coordinates are the $g_j(x)$, where $g_j = \frac{\partial f}{\partial x_j}$ (the function that satisfies (4.5)). The way that we group the different derivatives $|g_j|$ together to make a number $|\nabla f|$ does not matter so much, but it is cleaner that way. And I'll not bother for $W^{2,p}$, where I will just define

$$||f||_{W^{2,p}} = ||f||_{L^p(\Omega)} + ||\nabla f||_{L^p(\Omega)} + \sum_{1 \leq i \leq j \leq n} ||\frac{\partial^2 f}{\partial x_j \partial x_j}||_{L^p(\Omega)}. \tag{4.7}$$

We skip the natural things to check: $W^{k,p}$, with the norm above, is a Banach space. And also, locally in the middle of Ω, smooth functions are dense in $W^{k,p}$.

When $\Omega = \mathbb{R}^n$, for instance, we don't need all these derivatives, and (due in particular to the fact that the Riesz transforms are bounded on $L^p(\mathbb{R}^n)$, so it is not a free theorem, we get that

$$||f||_{W^{1,p}} \leq C||f||_{L^p(\mathbb{R}^n)} + C||\Delta f||_{L^p(\mathbb{R}^n)}. \tag{4.8}$$

That is, f and Δf control the other derivatives of order ≤ 2.

For our spectral stories, $p = 2$ will be enough. In this case (4.8) can be made much simpler, because we can use the description in terms of the Fourier trtansform $\widehat{f}$. But on general domains Ω (and you can even think about $\Omega = B(0,1)$), the story is not as simple.

Example. Take $\Omega = I = (a, b)$, a bounded interval in $\mathbb{R}$. Take $f \in L^1_{loc}(I)$ (the minimum if we want to see f as a distribution). We claim that $f \in W^{1,p}(I)$ if and only if there exists $g \in L^p(I)$ such that

$$f(y) - f(x) = \int_x^y g(t)dt \;\; \text{for } a < x < y < b. \tag{4.9}$$

And then of course $f' = g$ in the sense of distributions.

Indeed if (4.9) holds, observe that $g \in L^1(I)$ by Hölder, so we can also define $f(a)$ and $f(b)$, which will make things simpler. Write, for $\varphi \in C^\infty_c(I)$,

$$\int_I f(x)\frac{\partial\varphi(x)}{\partial x}dx = f(a)\int_I \frac{\partial\varphi(x)}{\partial x}dx + \int\int_{x,y\in I; y\leq x} g(y)\frac{\partial\varphi(x)}{\partial x}dxdy = \int\int_{x,y\in I; y\leq x} g(y)\frac{\partial\varphi(x)}{\partial x}dxdy$$

while

$$\int_I g(y)\varphi(y)dy = -\int\int_{x,y\in I; y\leq x} g(y)\frac{\partial\varphi(x)}{dx}dxdy$$

because $g(b) = 0$, so the desired result follows from Fubini. You can take this as a poor man's integration by parts (because we don't know that f is C^1). Conversely, if $g = \frac{\partial f}{\partial x} \in L^p$, then we can use the above to see that when we remove $\int_a^x g(t)dt$ from f, we get a function $\widetilde{f}$ whose distributional derivative is 0. With a little more work, we get that $\widetilde{f}$ is constant and the result follows. The little more work is not as simple as one would think: you have to check that if $\int f\varphi' = 0$ for every compactly supported smooth function φ, then f is constant. But this is classical and I'll skip.

This case is nice and simple because there is a simple way to recover f when you know f'; this is not always so simple in $\Omega \subset \mathbb{R}^n$. Even in this simple case, it is good to know that the usual derivative (I'll call it Df) exists almost everywhere [this is a consequence (exercise) of the Lebesgue density theorem].

In fact there is a theorem of Calderón, which generalizes a (celebrated) result of Rademacher (who did it only for Lipschitz functions), that says that if $f \in W^{1,p}(\Omega)$ for some $p \geq 1$, then for almost every $x \in \Omega$, f has a differential $Df(x)$ at x. That is, we can write the expansion $f(y) = f(x) + Df(x)(y - x) + o(||y - x||)$ near x.

Let us also mention that even in dimension 1, the existence of a differential almost everywhere does not imply much on the distribution derivative of f; for the function whose graph is called de devil's staircase, there is a differential almost everywhere, which is 0, but the function is not constant.

Below, up to the next subsection about Poincaré and Sobolev, I copy more detailed french notes, but with more or less the same content as what we just said.

On va commencer par définir les espaces $W^{k,p}(U)$ lorsque k est un entier positif (ou nul) et $p \in [1, +\infty]$.

Pour $k = 0$, prendre $W^{0,p}(U) = L^p(U)$, avec la même norme.

Pour $k = 1$ (le plus important pour nous), on dira que $f \in W^{1,p}(U)$ si $f \in L^p$ et si de plus ses dérivée premières (les $\frac{\partial f}{\partial x_i}$), prises au sens des distributions, sont dans $L^p(U)$. [Voir la traduction ci-dessous.] Et on prendra la norme $||f||_{W^{1,p}(U)} = ||f||_p + \sum_i ||\frac{\partial f}{\partial x_i}||_p$.

Pour k entier plus grand, on dit que $f \in W^{k,p}(U)$ si f, ainsi que toutes ses dérivées jusqu'à l'ordre k (toujours prises au sens des distributions) sont dans L^p, et on peut prendre pour norme la somme des normes L^p.

Traduction. Soit f localement intégrable sur U. Alors, elle définit une distribution sur U, en posant $\langle f, \varphi \rangle = \int_U f(x)\varphi(x)dx$ pour $\varphi \in C_c^\infty(U)$ (effet de la distribution sur la fonction test).

Ensuite, on définit les distributions $\frac{\partial f}{\partial x_i}$ par $\langle \frac{\partial f}{\partial x_i}, \varphi \rangle = -\langle f, \frac{\partial \varphi}{\partial x_i} \rangle$ pour $\varphi \in C_c^\infty(U)$.

Donc on dira que la dérivée partielle $\frac{\partial f}{\partial x_i}$ est dans L^p s'il existe $g_i \in L^p$ telle que

$$-\langle f, \frac{\partial \varphi}{\partial x_i} \rangle = \langle g_i, \varphi \rangle \text{ pour toute fonction test } \varphi \in C_c^\infty(U). \tag{1}$$

Assez facile à vérifier: s'il existe $g \in L^p$ telle que (1) ait lieu, elle est unique. Ou, de manière à peu près équivalente, si la fonction $g \in L^1_{loc}$ définit une distribution nulle, elle est nulle presque-partout. Si c'était faux, la théorie des distributions en souffrirait beaucoup. On verra plus bas une démonstration un peu particulière basée sur le théorème de densité de Lebesgue, en dimension 1.

Parfois, le fait que $\frac{\partial f}{\partial x_i}$ est dans L^p se décide plus facilement par dualité. Commençons par discuter dans le cas où $1 < p \leq +\infty$. Soit q l'exposant conjugué de p.

Si $\frac{\partial f}{\partial x_i}$ est dans L^p, alors, en utilisant (1) on trouve que

$$|\langle f, \frac{\partial \varphi}{\partial x_i} \rangle| \leq C||\varphi||_q \text{ pour tout } \varphi \in C_c^\infty(U), \tag{2}$$

avec $C = ||g_i||_p$. Par un petit passage à la limite, ceci implique d'ailleurs que

$$|\langle f, \frac{\partial \varphi}{\partial x_i} \rangle| \leq C||\varphi||_q \text{ pour tout } \varphi \in C_c^1(U) \tag{3}$$

(avec la même constante C). Et réciproquement, si on a (2) ou (3) (et si $p > 1$), ceci implique que l'application $\varphi \to \langle f, \frac{\partial \varphi}{\partial x_i} \rangle$ s'étend en une forme linéaire continue sur L^q, donc (par Riesz), est donnée par une fonction de L^p. Bref, quand $p > 1$, (2) ou (3) implique que $\frac{\partial f}{\partial x_i}$ est dans L^p.

Quand $p = 1$, il y a un autre espace un peu plus grand que $W^{1,1}(U)$, l'espace BV des fonctions $f \in L^1$ (ou alors, $f \in L^1_{loc}$, cela pourrait dépendre des auteurs) telles que pour tout i, $\frac{\partial f}{\partial x_i}$ est une mesure (signée ou complexe) finie. Il se trouve que le dual de la fermeture, pour la norme sup, de l'ensemble des fonctions bornées à support compact, est justement l'espace vectoriel des mesures finies sur U, muni de la norme $||\nu||$ = variation totale de ν. Donc, quand $p = 1$ et $q = +\infty$, (2) et (3) caractérisent le fait que $\frac{\partial f}{\partial x_i}$ est une mesure (signée) finie. On en reparlera au chapitre sur BV.

Quand $p = 2$ et $U = \mathbb{R}^n$, les choses s'expriment très bien en termes de transformée de Fourier. L'espace $W^{k,2}$ est alors le plus souvent noté H^k, et il est assez facile de voir que $f \in H^k(\mathbb{R}^n)$ si et seulement si

$$(4) \qquad ||f||^2_{H^k} = \int_{\mathbb{R}^n} (1 + |\xi|^2)^k |\widehat{f}(\xi)|^2 d\xi < +\infty.$$

Noter au passage que quand $f \in L^2$, $\widehat{f}(\xi)$ est définie presque-partout, donc le membre de droite a un sens.

Du coup, il est facile de définir H^s pour tout $s \geq 0$, en utilisant (2) (en fait, même pour pour tout $s \in \mathbb{R}$). On voyage entre les H^s en appliquant diverses puissances réelles de l'opérateur positif $I - \Delta$, dont l'effet en transformée de Fourier est juste de multiplier $\widehat{f}(\xi)$ par $1 + |\xi|^2$.

Et on peut faire pareil chez les $W^{k,p}(\mathbb{R}^n)$ (au moins pour $1 < p < +\infty$, SVP vérifiez dans les autres cas avant de dire que c'est ma faute); il y a quelques petites vérifications à faire, pour savoir par exemple que $(I - \Delta)^{l/2}$ est un isomorphisme de $W^{k+l,p}$ dans $W^{k,p}$, ou pour vérifier que pour $k = 1$ nos deux définitions de $W^{1,p}$ coincident, ou encore pour définir $W^{k,p}$ pour k non entier.

Deux mots des espaces homogènes aussi. On notera $\dot{W}^{k,p}(U)$ l'espace des fonctions localement intégrables sur U, dont les dérivées d'ordre k sont dans L^p, et muni de la norme "homogène" obtenue en sommant les normes L^p des dérivées d'ordre k. [On ne met rien pour contrôler le caractère L^1_{loc}, mais celui-ci découle des estimations sur les dérivées, comme on le verra de manière implicite plus bas.] Sur $\mathbb{R}^n$ et pour les k non entiers, on utiliserait les puissances de $-\Delta$ au lieu de $I - \Delta$, en faisant un peu plus attention.

Remarque (produit et localisation). Soient $f \in W^{1,p}$ et g de classe C^1 sur U, disons bornée et avec une dérivée bornée. Alors $fg \in W^{1,p}$, avec $\frac{\partial (fg)}{\partial x_i} = \frac{\partial f}{\partial x_i} g + \frac{\partial g}{\partial x_i} f$.

En effet, le membre de droite est bien dans L^p, donc il s'agit seulement de montrer que pour $\varphi \in C^\infty_c$, $-\int fg \frac{\partial \varphi}{\partial x_i} = -\int f \frac{\partial (g\varphi)}{\partial x_i} + \int \frac{\partial g}{\partial x_i} f \varphi$. Toutes les intégrales portent sur un compact (le support de φ), et il s'agit seulement d'intégrer l'identité $\frac{\partial (g\varphi)}{\partial x_i} = g \frac{\partial \varphi}{\partial x_i} + \frac{\partial g}{\partial x_i} \varphi$ contre $f \in L^1_{loc}$.

L'intérêt de cette remarque est aussi qu'elle permet de définir l'espace $W^{1,p}_{loc}(U)$ comme étant l'espace des fonctions telles que $fg \in W^{1,p}(U)$ pour $g \in C^1_c(U)$. La dérivée $\frac{\partial f}{\partial x_i}$ est alors la fonction de $L^p_{loc}(U)$ telle que si $K \subset U$ est compact, et $g \in C^1_c(U)$ est égale à 1 sur un voisinage de K, alors $\frac{\partial f}{\partial x_i} = \frac{\partial (fg)}{\partial x_i}$ sur K.

Exemple. Plaçons-nous sur $U =]a, b[\subset \mathbb{R}$. Vérifions d'abord que pour obtenir $f \in W^{1,p}(U)$, il suffit de prendre $g \in L^p(U)$ et $\lambda \in \mathbb{R}$, et de poser

$$f(x) = \lambda + \int_a^x g(t)dt \quad \text{pour } x \in U. \tag{5}$$

Et alors $f' = g$ (au sens des distributions).

En effet, il suffit de voir que $\int_a^b f(x)\varphi'(x)dx = -\int_a^b g(t)\varphi(t)dt$ pour pour $\varphi \in C_c^\infty(U)$. On note d'abord que λ ne contribue ni au membre de gauche (trivialement), ni au membre de droite (parce que $\int \varphi'(x)dx = 0$ puisque φ est à support compact). On applique ensuite le théorème de Fubini pour calculer $\displaystyle\int_U \int_U g(t)\, \varphi'(x) \mathbf{1}_{x>t}\, dxdt$. Quand on intègre d'abord en x, on trouve le membre de droite. Quand on intègre d'abord en t, on trouve le membre de gauche.

Exercice: généraliser ce qui précède à une mesure finie (disons, signée). Il faut quand même faire un peu attention à la contribution des bornes quand on écrit Fubini (mais pas dans la définition de f).

Réciproquement, si $f \in W^{1,p}$, alors il existe λ telle qu'on ait la formule plus haut. En effet, on pose $g = f'$ et $F(x) = \int_a^x g(t)dt$, et on note que $F \in W^{1,p}$, avec $F' = g = f'$. Donc la dérivée de $F - f$ au sens des distributions est nulle, i.e., $\int_U (f - F)\varphi' = 0$ pour toute fonction test φ. Il reste à vérifier que $G = f - F$ est (égale presque-partout à une) constante, dès qu'elle est localement intégrable et que sa dérivées est nulle. Il y a divers moyens de le vérifier, mais le plus amusant (compte tenu de ce qu'on a fait) sera d'utiliser le théorème de densité de Lebesgue.

Soient x et y deux points de densité de Lebesgue de $G = f - F$, avec $x \neq y$, et vérifions que $G(x) = G(y)$. On en déduira le résultat, en fixant x et en notant que presque-tout point $y \in U$ est un point de densité de Lebesgue, donc tel que $G(y) = G(x)$. On se donne une fonction bosse ψ de classe C^∞ à support dans $[0, 1]$ et d'intégrale 1, on pose $\psi_k(t) = 2^k \psi(2^k t)$ (toujours d'intégrale 1), et ensuite φ_k à support compact, dont la dérivée est $\varphi_k'(t) = \psi_k(t - x) - \psi_k(t - y)$. Le support de φ_k est contenu dans U pour k assez grand, donc on sait que $\int_U G(t)[\psi_k(t - x) - \psi_k(t - y)]dt = 0$. Mais

$$\begin{aligned} \big|G(x) - \int_U G(t)\psi_k(t - x)dt\big| &= \Big| \int_{x-t}^{x+t} [G(x) - G(t)]\psi_k(t - x)dt \Big| \\ &\leq \int_{x-t}^{x+t} \big|G(x) - G(t)\big| |\psi_k(t - x)|dt \leq ||g||_\infty \frac{1}{t} \int_{x-t}^{x+t} \big|G(x) - G(t)\big|dt, \end{aligned} \tag{4.10}$$

qui tend vers 0 par la version un peu forte du théorème de densité de Lebesgue (voir plus haut).

Pareillement, $\int_U G(t)\psi_k(t - x)dt$ tend vers $G(y)$; on trouve $G(x) = G(y)$, et ceci termine notre description par (5) des fonctions de $W^{1,p}$ sur un intervalle borné. Bien sûr, on a noté que pour la discussion ci-dessus, on aurait pu se contenter de regarder le cas où $p = 1$.

Signalons encore que puisqu'on a (1), on peut prouver que f est différentiable presque-partout, avec la dérivée $g = f'$. En fait, si f est un point de Lebesgue pour g, on a

$$\lim_{r \to 0} \frac{f(x+r) - f(x)}{r} = \lim_{r \to 0} \frac{1}{r} \int_x^{x+r} g(t)dt = g(x),$$

et de même $\lim_{r \to 0} \dfrac{f(x-r) - f(x)}{-r} = g(x)$.

Mais par contre il existe des fonctions non constantes dont la dérivée existe et est nulle presque-partout. Par exemple, la fonction de Lebesgue, dont le graphe est aussi surnomé escalier du diable. En fait, quand on intègre une mesure finie $\nu = \nu_1 + \nu_2$, où $\nu_1 = gdx$ est absolument continue par rapport à la mesure de Lebesgue et ν_2 est une mesure singulière, on trouve une dérivée égale à g Lebesgue-presque-partout, et par contre (si ν_2 est une mesure positive), la dérivée de f est $+\infty$ en ν_2-presque tout point. Si ceci vous rappelle le résultat sur la différentiation des mesures, c'est normal, c'est bien de cela qu'il s'agit.

Exercice: vérifier qu'en effet la primitive d'une mesure singulière (par rapport à Lebesgue) a bien une dérivée nulle Lebesgue-p.p..

4.5 Poincaré and Sobolev

When we know ∇f (and I mean the distribution derivative) on Ω, is there a way to recover f and prove nice estimates on f? We have seen yes on the interval.

Suppose f is smooth (say, C^1). For $x, y \in \Omega$, we can always recover $f(y) - f(x)$ as

$$f(y) - f(x) = \int_I \nabla f(\gamma(t)) \cdot \gamma'(t)dt \tag{4.11}$$

for any C^1 path $\gamma : I = [a, b] \to \Omega$ such that $\gamma(a) = x$ and $\gamma(b) = y$. We shall practice this, but notice that if we only know that $f \in W^{1,p}(\Omega)$, and even assuming that (4.11) is true, we could be unlucky and integrate on a path γ such that $|\nabla f(x)| = +\infty$ on γ. Then (4.11) won't help much.

Before I return to this, let me mention that when $\Omega = \mathbb{R}^n$, for instance, there are ways to avoid (4.11), because you may be able to say that the Fourier transform $\partial_j f = \frac{\partial f}{\partial x_j}$ is the product of $\widehat{f}$ by $i\xi_j$, say, compute $\widehat{f}$ from the $i\xi_j\widehat{f}$, write f as an operator applied to those, and study that operator. On Ω we can't do that so easily, even though there are ways to localize some Fourier transform estimates.

Return to the control of $f(y) - f(x)$ in terms of $\nabla f \in L^p$.

And return to the french notes at the same time

On aura aussi besoin d'un lemme d'approximation des fonctions de $W^{1,p}$ par des fonctions régulières. We'll almost certainly take this for granted in the course.

Lemma 4.8. *On se donne $U \subset \mathbb{R}^n$, $1 \leq p < +\infty$, et $f \in W^{1,p}_{loc}(U)$. Pour tout compact $K \subset U$, il existe une suite $\{f_k\}$ dans $C_c^\infty(U)$ telle que*

$$\lim_{k \to +\infty} f_k = f \ \textit{dans}\ L^p(K) \quad \textit{et, pour}\ 1 \leq i \leq n, \quad \lim_{k \to +\infty} \frac{\partial f_k}{\partial x_i} = \frac{\partial f}{\partial x_i} \ \textit{dans}\ L^p(K).$$

On pourrait d'ailleurs (par extraction standard) trouver une même suite qui marche pour tous les compacts. Démonstration sans trop de détails: on essaie $f_k = f * \psi_k$, où $\{\psi_k\}$ est une approximation de l'identité, avec support$(\psi_k) \subset B(0, 2^{-k})$. Pour k assez grand pour que $K + B(0, 2^{-k}) \subset\subset U$, f_k est bien définie sur K. Le fait que $\{f_k\}$ converge vers f dans $L^p(K)$ est très classique. Pour la seconde partie, vérifions d'abord que $\frac{\partial f_k}{\partial x_i} = \frac{\partial f}{\partial x_i} * \psi_k$ dans un voisinage de K (et pour k assez grand). On vérifie ceci sur une fonction test φ (de support contenu dans ce voisinage de K, et pour peu que celui-ci reste à distance $> 2^{-k}$ du bord). Le membre de droite donne

$$\begin{aligned}
&\int_U \Big\{ \int_{B(0,2^{-k})} \frac{\partial f}{\partial x_i}(x-u)\psi_k(u)du \Big\} \varphi(x)dx \\
&\qquad = \int_{B(0,2^{-k})} \psi_k(u) \Big\{ \int_U \frac{\partial f}{\partial x_i}(x-u)\varphi(x)dx \Big\} du \\
&\qquad = \int_{B(0,2^{-k})} \psi_k(u) \Big\{ \int_U \frac{\partial f}{\partial x_i}(y)\varphi(y+u)dx \Big\} du \\
&\qquad = -\int_{B(0,2^{-k})} \psi_k(u) \int_U f(y) \frac{\partial \varphi}{\partial x_i}(y+u)dydu \\
&\qquad = -\int_U \int_{B(0,2^{-k})} f(y)\psi_k(u) \frac{\partial \varphi}{\partial x_i}(y+u)dudy
\end{aligned} \tag{4.12}$$

(on a utilisé la définition de $\frac{\partial f}{\partial x_i}$ après avoir vérifié que pour k assez grand, le support de $\varphi(x+u)$ est encore contenu dans U). Le membre de gauche donne

$$\begin{aligned}
-\int_U f_k \frac{\partial \varphi}{\partial x_i} &= -\int_U \int_{B(0,2^{-k})} f(x-u)\psi_k(u) \frac{\partial \varphi}{\partial x_i}(x) \\
&= -\int_U \int_{B(0,2^{-k})} f(y)\psi_k(u) \frac{\partial \varphi}{\partial x_i}(y+u)\, dxdu,
\end{aligned} \tag{4.13}$$

et on constate que c'est pareil.

Cette vérification faite, on vérifie que $\frac{\partial f_k}{\partial x_i} = \frac{\partial f}{\partial x_i} * \psi_k$ converge bien vers $\frac{\partial f}{\partial x_i}$ dans $L^p(K)$ comme plus haut. Enfin, nos fonctions ne sont pas à support compact, donc on les multiplie par une fonction C^∞ qui vaut 1 dans un voisinage de K mais est à support compact dans U. □

This one, called Lemme 2 in the french notes, is easy and will be used later.

Lemma 4.9. *On se donne $p \geq 1$, $f \in L^p(V)$, et une suite $\{f_k\}$ dans $W^{1,p}(V)$. On suppose que $f = \lim_{k\to\infty} f_k$ dans $L^p(V)$ et que les $\frac{\partial f_k}{\partial x_i}$ convergent dans $L^p(V)$ vers des fonctions g_i. Alors $f \in W^{1,p}(V)$ et $\frac{\partial f}{\partial x_i} = g_i$.*

Soit $\varphi \in C_c^\infty(V)$. Alors

$$\begin{aligned}
\Big\langle \frac{\partial f}{\partial x_i}, \varphi \Big\rangle &= -\Big\langle f, \frac{\partial \varphi}{\partial x_i} \Big\rangle = -\lim_{k\to+\infty} \int_V f_k(x) \frac{\partial \varphi}{\partial x_i}(x)dx \\
&= \lim_{k\to+\infty} \int_V \frac{\partial f_k}{\partial x_i}(x)\varphi(x)\, dx = \int_V g_i(x)\varphi(x)\, dx = \langle g_i, \varphi \rangle.
\end{aligned} \tag{4.14}$$

Il n'y a pas de problème de convergence, puisque $\frac{\partial \varphi}{\partial x_i}$ et φ sont bornées et à supports compacts dans V.

En fait, le lemme marche encore en supposant seulement que f et les g_i sont dans L^p, que $f = \lim_{k\to\infty} f_k$ faiblement, et que $g_i = \lim_{k\to\infty} \frac{\partial f_k}{\partial x_i}$ faiblement. □

Now we replace (4.11) with a more general and flexible identity.

On est à peu près prêts pour essayer de majorer $|f(x) - f(y)|$ en fonction du gradient de f près de x et y. On verra après comment en déduire des inégalités fonctionnelles sur les $W^{1,p}$.

Pour l'instant, supposons que x et y sont donnés et que f est de classe C^1 près de x et y (on aura besoin ce ça dans un petit double cône $C_{x,y}$ autour de x et y).

Quelques notations. On se donne $\alpha \in]0,1]$ pour faire général, mais $\alpha = 1$ devrait suffire. On note

$$D(x,y) = D_\alpha(x,y) = \Big\{z \in \mathbb{R}^n : |z-x| = |z-y| \text{ et } \Big|z - \frac{x+y}{2}\Big| \leq \alpha|x-y|\Big\}, \tag{4}$$

un bout d'hyperdisque à mi-chemin, et

$$C(x,y) = C_\alpha(x,y) = \operatorname{conv}\big(\{x\} \cup \{y\} \cup D(x,y)\big) \tag{5}$$

(l'enveloppe convexe), qui est une sorte de double cône (une carotte de bureau de tabac) avec ses pointes en x et y.

On paramètre aussi des chemins γ_ℓ, $\ell \in D(x,y)$: pour $\ell \in D(x,y)$, γ_ℓ est le chemin de x à y obtenu en joignant x à ℓ puis à y par les segments $[x,\ell]$ et $[\ell,y]$. On paramètre γ_ℓ par sa projection sur l'intervalle $[x,y]$, et on note aussi $\gamma_\ell : [x,y] \to C(x,y)$ le paramétrage.

Pour la suite des calculs, on suppose que $C(x,y) \subset U$ et que la fonction f est de classe C^1 sur $C(x,y)$. Les accroissements finis, appliqués à $f \circ \gamma_\ell$, donnent

$$|f(x) - f(y)| \leq \int_{[x,y]} |\nabla f(\gamma_\ell(t))||\gamma_\ell'(t)|dt = \int_{\Gamma_\ell} |\nabla f| d\mathcal{H}^1 \tag{6}$$

où la dernière égalité ne sert qu'à essayer de rendre la chose plus géométrique (on ne l'utilisera pas) et $\Gamma_\ell = [x,\ell] \cup [\ell,y]$. On moyennise cela par rapport à $\ell \in D(x,y)$, et on trouve

$$|f(x) - f(y)| \leq \frac{1}{\mathcal{H}^{n-1}(D(x,y))} \int_{D(x,y)} \int_{[x,y]} |\nabla f(\gamma_\ell(u))||\gamma_\ell'(u)|du\,d\ell \tag{7}$$

où l'on a noté indifféremment $d\ell$ et $d\mathcal{H}^{n-1}(\ell)$ la mesure de surface. On fait le changement de variable $(\ell,u) \to \gamma_\ell(u)$ et on trouve

$$|f(x) - f(y)| \leq \int_{C(x,y)} |\nabla f(z)|\theta(z)dz, \tag{8}$$

où dz est la mesure de Lebesgue et $\theta(z)$ vient du déterminant Jacobien. Le calcul donne

$$\theta(z) \leq C(n,\alpha) w_{x,y}(z), \tag{9}$$

avec

$$w_{x,y}(z) = \mathbf{1}_{C(x,y)}(z)\big\{|z-x|^{1-n} + |z-y|^{1-n}\big\}. \tag{10}$$

Autrement dit, la mesure image de $\mathcal{H}^{n-1}(D(x,y))^{-1}du\,d\ell$ par $(\ell,u) \to \gamma_\ell(u)$ est inférieure à $C(n,\alpha)\,w_{x,y}(z)$.

Remarque. Les calculs suivants donnent des majorations (parce que c'est ce qui nous intéresse ici), mais on aurait pu calculer de manière plus précise, en écrivant vraiment la formule des accroissements finis. On aurait obtenu une formule du type $f(y) - f(x) = \int_{C(x,y)} K_{x,y}(z)\nabla f(z)\,dz$, où $K_{x,y}(z)$ est une matrice qui dépend de z (et avec certaines propriétés d'invariance par rapport à $x-y$). Noter que si on fait tendre y vers l'infini (ici, en partant dans une direction donnée), on obtient une formule qui permet de calculer f à partir de ∇f, par convolution avec un noyau matriciel (et en supposant que f est C^1 à support compact, au moins pour commencer). A cause de la manière dont on s'y est pris (avec la forme de notre cône), la formule n'est pas invariante par rotations, mais en intégrant sur toutes les droites partant de x, on aurait une formule invariante par rotation. A savoir, $f(x) = -c_n \displaystyle\int_{\mathbb{R}^n} \nabla f \cdot \frac{y-x}{|y-x|^n}dy$, et où c_n est la mesure de la sphère unité. [Ca ne peut être que ça, par homogénéité, invariance par dilatations, et en calculant la constante sur des fonctions particulières, par exemple tendant vers la fonction caractéristique d'une sphère.]

Et ce serait la même formule qu'on peut obtenir en partant de ∇f, et en lui appliquant l'opérateur $\Delta^{-1}\nabla$ (à définir plus facilement en transformée de Fourier), qui est bien un opérateur de convolution. Bref, les calculs faits plus hauts sont moins jolis et homogènes, mais ils sont robustes et marchent tout seul dans des domaines.

Il va falloir commencer à distinguer des cas. Si on veut une vraie estimation ponctuelle de $|f(x) - f(y)|$, valable pour $f \in W^{1,p}$ quelconque, il faudra s'assurer que $|\nabla f(z)|$ est intégrable contre $w_{x,y}(z)$, et ça ne marchera bien tel quel que si $p > n$. On va commencer par ce cas-là. Au fait, le cas $n = 1$ est trop facile; voir le cas traité plus haut.

Notons $q = \frac{p}{p-1}$ l'exposant conjugué de p. Si $p > n$, alors $q < \frac{n}{n-1}$, le poids $|z|^{q(1-n)}$ est localement integrable, et

$$||w_{x,y}||_q \leq 2\Big\{\int_{|z-x|\leq|x-y|} |z-x|^{q(1-n)}\Big\}^{1/q} \leq C|x-y|^{(1-n)+\frac{n}{q}} = C|x-y|^{\frac{p-n}{p}} \tag{11}$$

(ce qui sera utile pour la normalisation). Finalement Hölder et (6) donnent

$$|f(x) - f(y)| \leq C\,||\nabla f||_{L^p(C(x,y))}||w_{x,y}||_q \leq C|x-y|^{\frac{p-n}{p}}\Big\{\int_{C(x,y)} |\nabla f|^p\Big\}^{1/p}. \tag{12}$$

[Statement for Hölder continuity]

Proposition 4.10. *Soient U un ouvert de $\mathbb{R}^n$, $p \in]n, +\infty]$, et $f \in W^{1,p}_{loc}(U)$. Soient $x, y \in U$, tels que de double cône $C(x,y)$ soit contenu dans U. Alors*

$$|f(x) - f(y)| \leq C\,|x-y|^{\frac{p-n}{p}}\Big\{\int_{C(x,y)} |\nabla f|^p\Big\}^{1/p}. \tag{13}$$

[The rest of the proof is a limiting argument. I don't dare to remove but you'll pass.] Ici C ne dépend que de n, p, et α (l'ouverture). En fait, il faudrait dire que f est égale presque-partout à une fonction continue, qui vérifie (13).

Pour démontrer ceci, on se donne f, et on l'approxime par des fonctions régulières f_k comme au lemme 1 (Lemma 4.8). A cause de la façon dont on a écrit le lemme, où la suite $\{f_k\}$ dépend du compact, on va s'y reprendre à deux fois.

D'abord, on se donne une boule fermée $B = \overline{B}(x_0, r)$ contenue dans U, et on construit les f_k comme au lemme 1 (Lemma 4.8). Quitte à extraire une sous-suite, on peut supposer que la suite $\{f_k\}$, qui converge déja dans $L^1(B)$ vers f, converge aussi presque-partout. Noter que pour $x, y \in B_1 = B(x_0, r/10)$, le double cône $C(x,y)$ est contenu dans $B \subset U$, donc on peut appliquer (12) aux f_k et il vient

$$(14) \qquad |f_k(x) - f_k(y)| \leq C\,|x-y|^{\frac{p-n}{p}} \Big\{ \int_B |\nabla f_k|^p \Big\}^{1/p} \leq A|x-y|^{\frac{p-n}{p}}$$

pour un nombre $A < +\infty$ (car les ∇f_k convergent dans $L^p(B)$, donc leur norme est bornée). Si $f(x)$ est la limite des $f_k(x)$ (ce qui arrive presque-partout), et pareil pour $f(y)$, on en déduit que $|f(x) - f(y)| \leq A|x-y|^{\frac{p-n}{p}}$. Ceci vaut pour x, y dans une partie de mesure pleine de B_1. On en déduit la continuité de f, après changement sur une partie de mesure nulle.

Maintenant on peut supposer f continue, on fixe x et y comme dans l'énoncé, et on applique le lemme 1 (Lemma 4.8) avec le compact $C(x,y)$. La démonstration du lemme donne aussi la convergence uniforme des f_k vers f, et (13) se déduit de (12), et du fait que les ∇f_k convergent vers ∇f dans $L^p(C(x,y))$. □

Passons très brièvement au cas où $p = n$. Prenons la fonction $f(x) = \ln(\ln(\frac{1}{|x|}))$, au voisinage de 0. Alors $|\nabla f(x)| = \frac{1}{|x|\ln(\frac{1}{|x|})}$, qui est bien dans L^n près de 0 (car $n \geq 2$). Donc $f \in W^{1,n}_{loc}$, et pourtant elle n'est ni continue ni bornée. Mais f est quand même exponentiellement intégrable quand $f \in W^{1,n}$; voir le chapitre sur BMO, la seconde remarque juste après le théorème de John et Nirenberg.

Next we go for Poincaré (the version with an average on a ball). For us $p = q$ will be enough.

Quand $p < n$, f n'est pas continue, mais par contre elle est localement dans des L^q meilleurs que L^p. Le plus pratique pour écrire ceci semble être en termes d'inégalités de Poincaré, qui sont une manière de dire comment $f \in L^q$ localement. Si on voulait des propriétés globales sur tout $\mathbb{R}^n$, on serait facilement canulé par des questions d'homogénéité. [Sauf peut-être justement pour $p = n$.]

On va utiliser les notations suivantes pour simplifier: $|B|$ est la mesure de Lebesgue de B et, si $f \in L^1(B)$, $m_B f = \frac{1}{|B|}\int_B f(x)dx$ est la moyenne de f sur B.

Proposition 4.11. *Soient $p \in [1, n]$ et $r \in [1, \frac{np}{n-p}[$. Il existe une constante $C(n, p, r)$ telle que pour toute boule ouverte B et tout $f \in W^{1,p}(B)$,*

$$\left\{ \frac{1}{|B|} \int_B |f(x) - m_B f|^r \right\}^{1/r} \leq C(n, p, r)\, |B|^{1/n} \left\{ \frac{1}{|B|} \int_B |\nabla f(x)|^p \right\}^{1/p}. \tag{15}$$

Quelques commentaires avant de commencer la démonstration.

Déjà, $r = p$ est assez intéressant, et en plus se simplifie un peu, puisqu'on doit faire moins attention à l'homogénéité, et écrire juste

$$\int_B |f(x) - m_B f|^p \leq C(n, p) \operatorname{diam}(B)^p \int_B |\nabla f(x)|^p. \tag{15'}$$

Quand $p = n$, tout $r < +\infty$ est autorisé. Et même, f est localement exponentiellement intégrable, comme on le verra pour les fonctions de BMO. Quand $p > n$ tout $r < +\infty$ est autorisé aussi, mais la Proposition 1 (Proposition 4.10) donne directement une borne sur $|f(x) - m_B f|$ qui est meilleure que (15). Donc on on s'est carrément mis dans le cas où $p \leq n$, et on n'a rien perdu.

J'ai mis les puissances de manière à simplifier la vérification d'homogénéité. Il est logique que chaque membre soit homogène de degré 1 en f. Pour ce qui est de l'homogénéité en fonction du rayon R de B, penser que quand $|\nabla f(x)|$ est de l'ordre de 1, les deux membres sont en principe de l'ordre de R ou $|B|^{1/n}$.

Noter que par Hölder, l'inégalité pour r entraîne celle pour $r' < r$. C'est pourquoi on pourra se contenter du cas où $r > p$ (noter que $\frac{np}{n-p} > p$).

On va commencer par le cas où f est de classe C^1, il faudra faire un petit passage à la limite à la fin.

Comme l'homogénéité est correcte, on pourrait se ramener facilement au cas où $B = B(0, 1)$, mais on va essayer de faire directement le cas général où $B = B(X, R)$.

Posons $B_1 = B(X, R/10)$, et notons que quand $x \in B$ et $y \in B_1$, alors le bi-cône $C(x, y)$ est contenu dans B. Alors (8) et (9) disent que

$$|f(x) - f(y)| \leq C \int_{C(x,y)} |\nabla f(z)|\, w_{x,y}(z) dz \tag{16}$$

On fait la moyenne par rapport à $y \in B_1$ et on trouve

$$|f(x) - m_{B_1} f| \leq C R^{-n} \int_{y \in B_1} \int_{z \in C(x,y)} |\nabla f(z)|\, \omega_{x,y}(z)\, dz dy, \tag{17}$$

On se souvient que $w_{x,y}(z) = \mathbf{1}_{C(x,y)}(z) \big\{ |z - x|^{1-n} + |z - y|^{1-n} \big\}$ (par (10)), et quand on intègre ceci par rapport à $y \in B_1$, on trouve moins que $C R^n (R^{1-n} + |z - x|^{1-n})$. Donc

$$|f(x) - m_{B_1} f| \leq C \int_{z \in B} |\nabla f(z)|\, (R^{1-n} + |z - x|^{1-n}) dz. \tag{18}$$

On pose $h(u) = (R^{1-n} + |u|^{1-n})\mathbf{1}_{B(0,2R)}(u)$, et on garde (18) sous la forme synthétique

$$|f(x) - m_{B_1}f| \leq C\,[(\mathbf{1}_B\,|\nabla f|) * h](x). \tag{19}$$

Puis on se souvient que la convolution envoie $L^p \times L^q$ dans L^r, où $\frac{1}{p} + \frac{1}{q} = \frac{1}{r} + 1$. Ici, on a déjà p (puisque $\nabla f \in L^p(B)$) et r (avec $p < r < \frac{np}{n-p}$). Si on prenait $r = \frac{np}{n-p}$, on trouverait $\frac{1}{q} = \frac{n-p}{np} + 1 - \frac{1}{p} = \frac{n-1}{n}$ et $q = \frac{n}{n-1}$. Ici on prend r plus petit, donc q plus petit. Bref, $q < \frac{n}{n-1}$, ce qui tombe bien, parce que cela implique que h est localement dans L^q. Noter encore que $r = p$ donne $q = 1$, donc on est dans des valeurs acceptables de q.

Retournons à (19). On trouve

$$||f(x) - m_{B_1}f||_r \leq C\,||h||_q\,||\nabla f||_{L^p(B)} \tag{20}$$

et en plus

$$||h||_q \leq C\Big\{R^{(1-n)q}R^n\Big\}^{1/q} = R^{1-n+\frac{n}{q}}. \tag{21}$$

C'est ce qu'il fallait pour (15), modulo deux choses. D'abord, la vérification de l'homogénéité en R. La puissance de R est $1 - n + \frac{n}{q} = 1 - n + n(1 + \frac{1}{r} - \frac{1}{p}) = 1 + \frac{n}{r} - \frac{n}{p}$. C'est exactement ce qu'il faut puisque $|B|$ est de l'ordre de R^n.

La seconde chose est qu'on a remplacé $m_B f$ par $m_{B_1}f$. Mais

$$\begin{aligned} |m_B f - m_{B_1}f| &\leq \frac{1}{|B|}\int_B |f(x) - m_{B_1}f|\,dx \leq \left\{\frac{1}{|B|}\int_B |f(x) - m_B f|^r\right\}^{1/r} \\ &\leq C(n,p,r)\,|B|^{1/n}\left\{\frac{1}{|B|}\int_B |\nabla f(x)|^p\right\}^{1/p}. \end{aligned} \tag{4.15}$$

par inégalité triangulaire, puis Hölder, puis ce qu'on vient de montrer; le remplacement de $m_B f$ par $m_{B_1}f$ ne coûte donc pas plus que le second membre.

Notons encore qu'en fait, dans les calculs ci-dessus, on peut remplacer $B_1 = B(X, R/10)$ par n'importe quelle boule $B(X, R_1)$, avec $R/11 \leq R_1 \leq R/9$. Soit en notant que la démonstration marche encore, soit en faisant comme pour (22).

On a presque fini. On a vérifié (15), mais seulement pour les fonctions f de classe C^1. Dans le cas général, on se donne une boule $B' = B(X, R') \subset B$, avec R' juste un peu plus petit que R, et on commence par utiliser le lemme 1 pour trouver une suite de fonctions f_k de classe C^1, qui convergent vers f dans $L^p(B')$, et telles que les ∇f_k convergent vers ∇f dans $L^p(B')$. Quitte à extraire une sous-suite on peut même s'arranger pour que $f_k(x)$ tende vers $f(x)$ pour presque tout $x \in B'$. On sait que

$$\left\{\int_{B'} |f_k(x) - m_{B_1'}f_k|^r\right\}^{1/r} \leq C\left\{\int_{B'} |\nabla f_k|^p\right\}^{1/p}, \tag{23}$$

où en principe on devrait prendre $B_1' = B(X, R'/10)$, mais où, à cause de la remarque précédente sur $R/11 \leq R_1 \leq R/9$, on peut carrément garder $B_1' = B_1$ (ce qui simplifie un peu le passage à la limite ci-dessous).

On fait tendre k vers $+\infty$, et on trouve que

$$\left\{ \int_{B'} |f(x) - m_{B_1'} f|^r \right\}^{1/r} \leq C \left\{ \int_{B'} |\nabla f|^p \right\}^{1/p} \leq C \left\{ \int_B |\nabla f|^p \right\}^{1/p}, \tag{24}$$

ou l'on a utilisé Fatou pour l'intégrale de gauche. On fait tendre R' vers R et on trouve l'analogue de (15) avec $m_{B_1} f$. Et ensuite on remplace $m_{B_1} f$ par $m_B f$ comme avant, avec (22). La proposition s'en déduit. □

Exercice. Montrer que, sous les hypothèses de la proposition 4.11, on a aussi

$$\left\{ \frac{1}{|B|^2} \int_B \int_B |f(x) - f(y)|^r \right\}^{1/r} \leq C(n,p,r)\, |B|^{1/n} \left\{ \frac{1}{|B|} \int_B |\nabla f(x)|^p \right\}^{1/p}. \tag{25}$$

We add this other version of Poincaré's inequality (compact support).

Corollary 4.12. *Soient $p \in [1,n]$ et $r \in [1, \frac{np}{n-p}[$. Il existe une constante $C(n,p,r)$ telle que pour toute boule ouverte B et tout $f \in W_0^{1,p}(B)$,*

$$\left\{ \frac{1}{|B|} \int_B |f(x)|^r \right\}^{1/r} \leq C(n,p,r)\, |B|^{1/n} \left\{ \frac{1}{|B|} \int_B |\nabla f(x)|^p \right\}^{1/p}. \tag{4.16}$$

Pareil en supposant seulement que $f \in W_0^{1,p}(\Omega)$ pour un ouvert $\Omega \subset B$.

Rappelons que $W_0^{1,p}(B)$ est l'adhérence de $C_c^\infty(B)$ dans $W_0^{1,p}(B)$. La différence avec le résultat précédent est qu'on suppose en plus que f est dans l'adhérence dans $W^{1,p}(B)$ des fonctions-test à support compact, et que par contre on n'a pas à retirer la moyenne de f sur une boule.

Démonstration rapide. The proof uses the following: when you take $f \in W^{1,p}(\Omega)$ and extend it by taking $f = 0$ outside of Ω, you still get a function $\widetilde{f} \in W^{1,p}(\mathbb{R}^n)$ with the same norm as f (in $W^{1,p}(\Omega)$).

One proves this from the definition: there is a sequence of test functions f_k that tend to f in $W^{1,k}(\Omega)$; we can of course extend them. Then the sequence converges in $W^{1,k}(\mathbb{R}^n)$ (by the Cauchy criterion), and the limit is our extension.

Ensuite, on utilise la proposition 4.11 sur $2B$ (et en particulier en intégrant sur B ou sur $2B \setminus B$) pour évaluer la taille de $m_B f$. Le résultat en découle.

4.6 Rellich-Kondrachov

Here is the form that we'll use in these notes.

Theorem 4.13. *Let $\Omega \subset \mathbb{R}^n$ be bounded, and $1 \leq p < +\infty$. Then the unit ball of $W_0^{1,p}(\Omega)$ is a relatively compact subset of $L^p(\Omega)$ (for the L^p norm). And also of $L^q(\Omega)$ for $q < \frac{np}{n-p}$.*

The case when $p > n$ is not so interesting because then the Hölder property gives compactness. Similarly, taking q smaller is less interesting (because of Hölder), so we may assume $q \geq p$.

The argument seems to be the same as below (for BV); at a point you need to estimate $\left\{ \int_Q |\nabla f|^b \right\}^{q/p}$ and then sum the estimates. And you just throw a power $(q-p)/p$ away (by saying it is bounded) and get away with it.

Here the vanishing condition at the boundary is useful to start the argument: you want to say that $\int |f|^q < +\infty$ before you start the covering argument.

The same result for $W^{1,p}(\Omega)$, even if Ω is connected and you restrict to functions whose average on a given ball is 0 (you don't want your space to contain a copy of $\mathbb{R}$) fails when the domain Ω is sufficiently ugly.

The main ingredient for this is Corollary 4.12, naturally together with the fact that completely bounded (précompact in french) plus complete implies compact. That is, at the end of the argument we prove that for all $\varepsilon > 0$, we can cover our ball with a finite (but very large) number of balls of radius ε (in L^p).

Here I will just put the french notes on the special case of $p = 1$ (but with the enlarged BV space which is a little larger than $W^{1,1}$ and different boundary conditions. The idea is the same: get a better local control.

Theorem 4.14. *Soient Q un cube ou une boule (ouvert) de $\mathbb{R}^n$, $M > 0$, et notons A l'ensemble des fonctions $f \in BV(Q)$ telles que $\int_Q |Df| \leq M$ et $|\int_Q f| \leq M$. Alors A est une partie compacte de $L^1(Q)$.*

Replace BV with $W^{1,2}$ if you want. Notons que la restriction sur l'intégrale de f (ou quelque chose de semblable) est nécessaire, à cause des fonctions constantes et puisque $\mathbb{R}$ n'est pas compact.

Rappelons que notre ensemble est une partie complète de L^1, parceque L^1 est complet et A est fermé: la condition d'intégrale passe à la limite, et pour la condition sur la variation, on utilise le théorème de semi-continuité inférieure. Il s'agit donc de montrer que A (muni de la distance $||f-g||_1$) est précompact. On se donne donc $\varepsilon > 0$ et on doit vérifier que A peut être recouvert par un nombre fini de boules de rayon ε.

On va se contenter du cas où Q est un cube; le cas d'une boule s'y ramène par changement de variable bilipschitzien, ou en modifiant un peu la construction avec des ensembles moins propres que des petits cubes. On se donne $\eta > 0$ très petit (à calculer à la fin en fonction de ε, et on recouvre Q par des cubes R_j d'intérieurs disjoints, de diamètre $\leq \eta$. Il en faut un nombre fini.

Noter qu'à cause de l'inégalité de Poincaré, $\int_Q |f| \leq |\int_Q f| + \int_Q |f - m_Q(f)| \leq M + C_Q \int_Q |Df| \leq C'_Q M$ pour $f \in A$. Alors

$$\frac{1}{|R_j|}\int_{R_j} |f| \leq \frac{1}{|R_j|}\int_Q |f| \leq \frac{C'_Q M}{|R_j|} \quad \text{pour } f \in A. \tag{13 = 8}$$

On a besoin d'un ensemble fini. Prenons l'ensemble des fonctions g de la forme

$$g = \sum_j \alpha_j \mathbf{1}_{R_j}, \tag{9}$$

où chaque α est un multiple entier de η tel que $|\alpha| \leq \frac{C'_Q M}{|R_j|}$. Ca fait beaucoup de fonctions g, mais un nombre fini quand même. On va d'ailleurs voir que le fait que la norme de g dans BV puisse être bien plus grande que M n'a aucune importance.

Soit donc $f \in A$ donnée. On choisit g comme en (9), de manière que $\left|\alpha_j - m_j\right| \leq \eta$, en posant $m_j = \frac{1}{|R_j|} \int_{R_j} f$. C'est possible à cause de (8). Et

$$\begin{aligned} ||f-g||_1 &= \sum_j \int_{R_j} |f-g| \leq \sum_j \int_{R_j} |f-m_j| + \sum_j \int_{R_j} |m_j - \alpha_j| \\ &\leq C \sum_j \int_{R_j} \operatorname{diam}(R_j)|Df| + \eta \sum_j |R_j| \\ &\leq C\eta \sum_j \int_{R_j} |Df| + \eta|Q| \leq C\eta M + \eta|Q| \end{aligned} \tag{4.17}$$

par Poincaré sur chaque R_j; l'homogénéité joue pour nous, et c'est normal puisque f est un peu régulière. Bref, quand on prend η assez petit, le membre de droite est plus petit que ε; on en déduit le théorème. $\square$

Remarque: il y a plein de variantes de ce genre d'argument; je ne suis pas sûr d'avoir choisi la plus simple. L'argument marche encore pour un domaine connexe assez général, en travaillant un peu plus la géométrie. Mais il ne faut pas permettre un domaine qui soit trop pincé par endroits, pour pouvoir contrôler, y compris localement, la norme L^1 à partir de $\int |Df|$ et de l'intégrale de f.

Variante du théorème: prenons Ω un ouvert borné, et soit A l'ensemble des fonctions caractéristiques d'ensembles F dont le périmètre dans Ω est au plus M. A nouveau A est compact dans $L^1(\Omega)$.

On peut même prendre les fonctions $f \in BV(\Omega)$ telles que $|f(x)| \leq M$ presque-partout sur Ω, et $\int_\Omega |Df| \leq M$, avec la même démonstration.

Démonstration. D'abord, A est encore complet: si f est la limite dans L^1 d'une suite $\{f_k\}$ dans A, on peut extraire une sous-suite qui converge presque partout vers f, et donc $f(x) \in \{0,1\}$ presque-partout. Le fait que $f \in BV(\Omega)$ est encore dû au théorème de semicontinuité.

Ensuite on fait comme plus haut. Soit $\varepsilon > 0$. On se donne $\eta > 0$ (à choisir bientôt), et on essaie de recouvrir Ω par des petits cubes de diamètre au plus η presque disjoints. On en met autant qu'on peut, mais il reste une partie V de Ω. Malgré tout, $|\Omega \setminus V|$ tend vers 0 quand η tend vers 0 (par convergence dominée, pour être un peu paresseux). A part ça, on utilise la même formule (9), mais on peut se contenter des cubes choisis, et (pour ce qui est des fonctions caractéristiques) de coefficients $\alpha_j \in [0,1]$. Au lieu de (10), on a

$$||f-g||_1 \leq \sum_j \int_{R_j} |f-g| + |\Omega \setminus V| \leq C\eta M + \eta|Q| + |\Omega \setminus V| \tag{4.18}$$

par le même calcul que plus haut, et on conclut pareil.

5 The laplacian on an open set of $\mathbb{R}^n$

This will be our main example. You'll be immediately amused that this is not a bounded operator, so we'll have to cheat if we want to apply the results above (we still want to study its spectrum and things like that). This will also be a good excuse to having done a bit of analysis and function spaces.

5.1 The Laplacian on a line segment

As you may guess, this is a simple variant of the next example, the Laplacian on a domain Ω. We want to diagonalize the Laplacian, and we'll find a trick: find an inverse, which is a bounded, even compact self-adjoint operator, and use the diagonalization of the inverse.

Consider $I = (a,b) \subset \mathbb{R}$, and consider $-\Delta : f \to -f''$. This is a well-defined operator from $W^{2,2}(I)$ to $L^2(I)$, for instance, and we can try to compute an inverse. Of course it sounds a little silly to write Δf for f'', but this won't hurt.

Example 1: the Dirichlet condition. Well, this would be the natural thing to do, but $-\Delta$ has a kernel (the affine functions) and we shall try to get rid of the problem by considering the subspace

$$W_0^{2,2} : W_0^{2,2}(I) = \big\{ f \in W^{2,2}(I)\,;\, f(a) = f(b) = 0 \big\}. \tag{5.1}$$

Normally this is not the definition of $W_0^{2,2}$, which instead should be the closure of $C_c^\infty(I)$ in $W^{2,2}$. But it is not hard to see that (5.1) is correct, essentially because $f \to f(a)$ is a bounded linear functional on $W^{2,2}$.

We want to find an inverse to the operator $f \to f''$, from $W_0^{2,2}$ to $L^2 = L^2(I)$, and the simplest seems to take antiderivatives twice. That is, given $f \in L^2$, consider Pf defined by

$$Pf(x) = \int_a^x f(t)dt. \tag{5.2}$$

It is easy to see that $Pf \in W^{1,2} = W^{1,2}(I)$ and in addition $Pf(a) = 0$ and $Pf(b) = \int_I f(x)dx$. Then of course we consider $-P^2$ and this could be a nice inverse for $-\Delta$. But $P^2(f)$ does not lie in $W_0^{2,2}$, because probably $P^2 f(b) \neq 0$. So we fix this, and replace $-P^2(f)$ with Lf, where

$$Lf(x) = -P^2 f(x) + c(x-a), \text{ where } c = c(f) = \frac{1}{b-a}\int_I Pf(x)dx. \tag{5.3}$$

Our L is still a bounded operator on L^2, and now $Lf \in W_0^{2,2}$ because it has the right second derivative (as before), and

$$Lf(b) = -P^2 f(b) + c(b-a) = -\int_I Pf(x)dx + c(b-a) = 0$$

because $P^2 f$ is the integral of its derivative Pf, and $Pf(a) = 0$. It is easy to check that $L : L^2 \to W_0^{2,2}$ is the inverse of $-\Delta : W_0^{2,2} \to L^2$.

Next we claim that $\widetilde{L} : L^2 \to L^2$ (obtained by composing L with the injection from $W_0^{2,2}$ to L^2, is a compact operator. This follows from the fact that L sends the unit ball of L^2 to B, where B is a ball of $W^{2,2}$ of radius $C(a,b)$. Then we use the Rellich-Kondrachov theorem.

Finally, we want to check that $\widetilde{L} : L^2 \to L^2$ is self-adjoint. This is where we have to be careful (see below for trouble in a similar-looking situation). For $f, g \in L^2$, we want to compute

$$\langle Lf, g\rangle = \int_I Lf(x)\overline{g}(x)dx.$$

We integrate by parts (using the primitive Pg of g, and the poor man's integration by parts):

$$\langle Lf, g\rangle = [L(f)\overline{Pg}]_a^b - \int_I (Lf)'(x)\overline{Pg}(x)dx = \int [Pf(x) - c(f)]\overline{Pg}(x)dx$$

because $L(f)$ vanishes at a and b, and its derivative is $-Pf + c(f)$, with $c = c(f)$ as in (5.3). Thus

$$\langle Lf, g\rangle = \int Pf(x)\overline{Pg}(x)dx - c(f)\int \overline{Pg}(x)dx = \int Pf(x)\overline{Pg}(x)dx - (b-a)c(f)\overline{c(g)}.$$

The same computation with $\langle f, Lg\rangle$ shows that $\langle Lf, g\rangle = \langle f, Lg\rangle$ (or if you prefer, the formula above clearly defines a sesquilinear form).

So L is self-adjoint. But it is also positive, i.e., $\langle Lf, f\rangle > 0$ for $f \in L^2$ (unless $f = 0$). Indeed,

$$\langle Lf, f\rangle = \int |Pf(x)|^2 - (b-a)c(f)^2 = \int |Pf(x) - c(f)|^2 = ||Pf - c(f)||_2^2$$

by Pythagorus (or because $Pf(x) - c(f)$ has integral 0, hence is orthogonal to constants). The point is that for L^2-norms, the most efficient way to take a primitive was not P. Of course this positivity is the reason why we prefer $-\Delta$.

So $\widetilde{L} : L^2 \to L^2$ is compact self-adjoint. In addition, its "eigenspace" E_0, i.e., the set of $f \in L^2$ such that $Lf = 0$, is reduced to 0. All this means that we can find an orthonormal basis of L^2 which is composed of eigenvectors for L. The situation is simple enough for us to compute eigenvalues and eigenvectors.

Note that since $\lambda \neq 0$ (and in fact $\lambda > 0$ because the operator is positive, but we'll pretend not to know that), every eigenvector for L is an eigenvector for $-\Delta$, and conversely (because $-\Delta$ has no kernel, at least in the present Dirichlet configuration where we restricted to $W_0^{2,2}$).

Let us rather study eigenvectors for $-\Delta$; this is more natural (and what we wanted). Let f be an eigenvector for $-\Delta$, with eigenvalue λ. Then $f \in W_0^{2,2}$ (by definition) and $-f'' = \lambda f$. This is in the sense of distribution, but then the equation implies that f' (the integral of $f'' \in L^2$) is continuous, then the equation says that $f \in C^3$, and so on, so f is just a regular solution, and $f(x) = \alpha exp(i\sqrt{\lambda}x) + \beta exp(-i\sqrt{\lambda}x)$ on I. If we pretend not to know that $\lambda > 0$, we just take a complex square root.

Now $f \in W_0^{2,2}$, so $f(a) = f(b) = 0$. That is, $\alpha exp(i\sqrt{\lambda}x) = -\beta exp(-i\sqrt{\lambda}x)$ for $x = a, b$. If $\alpha = 0$, we get $\beta = 0$ and f is not an eigenfunction. So $\alpha \neq 0$, and similarly $\beta \neq 0$. Now we multiply these two equations, simplify, and get that $exp(2i\sqrt{\lambda}(b-a)) = 1$. Or in other words, that $\sqrt{\lambda}(b-a)$ is an integer multiple of π. That is, the only (possible) eigenvalues of $-\Delta$ on $W_0^{2,2}$ are the numbers

$$\lambda = k^2\pi^2(b-a)^{-2} \text{ for some integer } k \geq 1. \tag{5.4}$$

In the mean time we confirmed what we knew: λ must be real because L is self-adjoint, and even positive because L is positive.

Conversely, if k is of this form (5.4), the solutions are combinations of $exp(\pm i\sqrt{\lambda}x) = exp(\pm i\frac{k\pi x}{b-a})$, our choice of k makes these functions periodic, and in order to make f vanish at both a and b, we have to choose an appropriate combination of exponentials, namely

$$f(t) = C\sin(\frac{k\pi(x-a)}{b-a}). \tag{5.5}$$

We easily check that this is a solution, that vanishes both at a and b, so we feel good and we have the full spectral decomposition of $-\Delta$ on $W^{2,2}$. Incidentally, the squares in (5.4) are not shocking, given the eigenvectors in (5.5), because $-\Delta$ has two derivatives.

Remark (trouble with d?) Why did we not do directly the same thing with the operator d, defined on $W_0^{1,2} = W_0^{1,2}(I)$ by $df = f'$?

We have a good candidate, it seems, to invert d, which is the operator P above (see (5.2)). But if we start from $f \in L^2$, then Pf usually does not lie in $W_0^{1,2}$, because probably $Pf(b) \neq 0$. So we should restrict our attention to the closed subspace

$$H = \Big\{g \in L^2\,;\, \int_I g = 0\Big\} \tag{5.6}$$

This seems fine because now $Pf \in W_0^{1,2}$ (i.e., $Pf(b) = Pf(a) = 0$) when $f \in H$, and we don't seem to lose much because $f' \in H$ when $f \in W_0^{1,2}$. So we have a beautiful inverse P to $d : W_0^{1,2} \to H$. It is not hard to check that P, seen from H to L^2, is still compact. Now, is is self-adjoint? We compute with the expected integration by parts: for $f, g \in H$,

$$\langle Pf, g\rangle = \int_I Pf(x)\overline{g}(x)dx = [Pf(x)\overline{Pg}(x)]_a^b - \int_I f(x)\overline{Pg}(x)dx = -\langle f, Pg\rangle \tag{5.7}$$

where the integrated term vanishes because $Pf(a) = Pf(b) = 0$. This is still true for $g \in L^2$ if you want (I am beginning to attract your attention to something).

OK, the sign is wrong. This is not shocking, because recall that in Fourier transform, d is associated to the multiplication by $i\xi$ (or $-i\xi$ if I did my calculation wrong), so maybe it is more reasonable to expect that $-id$, or equivalently iP, is self-adjoint and maybe positive? That would not disturb much. And indeed the computation above yields $\langle iPf, g\rangle = -i\int_I f'(x)\overline{Pg}(x)dx = \langle f, iLg\rangle$, so it looks that we could diagonalize iL after all.

Yet, when we try to find eigenvectors, we get the equation $f' = -i\lambda f$, which has no solution in $W_0^{1,0}(I)$ (except 0) because the boundary constraint is too tough. What happened? Well, unfortunately iP maps H to $W_0^{1,0} \subset L^2$, but $W_0^{1,0}$ is not contained in H, so we cannot say that $iP : H \to H$ and hence we cannot apply the spectral theorem to iP. So we missed by a tiny bit, but this is enough to create trouble.

The moral of this is that with unbounded operators, we always have to be careful about what is the domain (here it would be $W_0^{1,0}$) and the image, before we do duality. Or said in other terms, you can look self-adjoint, but in fact not be.

Example 2: the Neumann condition.

What if we don't like the Dirichlet condition $f(a) = f(b) = 0$ that defines $W_0^{2,2}(I)$, and prefer to work with the whole $W^{2,2}$ or another boundary condition?

Since many people like the so-called Neumann condition, we'll start with this. We now work on $W_n^{2,2}(I) = \{f \in W^{1,2}(I)\,;\, f'(a) = f'(b) = 0\}$. We can guess its image in L^2: this will be the same set H as above ($f'(a) = f'(b)$ implies that $\int f'' = 0$).

Now we cannot directly look for the inverse of $\Delta : W_n^{2,2} \to H$, because $-\Delta$ has a kernel. This is the space of dimension 1 of constant functions (normally we would have affine functions, but the Neumann condition kills the linear part), so it is better to work on the orthogonal of the kernel, i.e., on $W_0^{2,2}(I) \cap H$.

Now what is the inverse L of $-\Delta$? To $f \in H$ we associate its primitive Pf that vanishes on a and hence on b too. Then we take another primitive, but $P^2 f$ may not do the job because although $P^2 f$ satisfies the Neumann condition (and any other primitive of Pf would do, just because $Pf \in W_0^{1,2}$), it may not be orthogonal to the kernel. So we replace $P^2 f$ with the primitive of Pf that has integral 0, i.e., we take

$$Lf = -P^2 f + e, \quad \text{with } e = e(f) = \frac{1}{b-a}\int_I P^2 f(x)dx. \tag{5.8}$$

This way $\int Lf = 0$ and Lf is orthogonal to the kernel.

We let the reader check that $L : H \to W_n^{2,2} \cap H$ is the inverse of $-\Delta : W_n^{2,2} \cap H \to H$, and that the image of the unit ball of H is relatively compact in H (i.e., that $L : H \to H$ is compact). We are worried about self-adjoint, so we check. For $f, g \in H$,

$$\langle Lg, g\rangle = \int [e(f) - P^2]\overline{g} = [(e(f) - P^2)\overline{Pg}]_a^b + \int Pf\overline{Pg} = \int Pf\overline{Pg} \tag{5.9}$$

because $Pg = 0$ on the boundary. So again L is self-adjoint, and even positive, and we can diagonalize. We have one more eigenvalue for $-\Delta$ than before, since $-\Delta$ has a kernel of dimension 1 on $W_n^{1,2}$. For the other eigenvalues, let us no longer pretend that we don't know that they are positive. So let $f \in W^{1,2}_n$ be an eigenvector for $-\Delta$, associated to to the eigenvalue $\lambda > 0$. As before, $-f'' = \lambda f$, so $f(x) = \alpha exp(i\sqrt{\lambda}x) + \beta exp(-i\sqrt{\lambda}x)$ on I. This time the boundary constraints are that $f'(a) = f'(b) = 0$, so $0 = f'(a) = i\sqrt{\lambda}\big[\alpha exp(i\sqrt{\lambda}a) - \beta exp(-i\sqrt{\lambda}a)\big]$, and similarly for b. These are the same equations as before, except for a minus sign (see 5 lines above (5.4)) and the same proof yields that $exp(2i\sqrt{\lambda}(b-a)) = 1$, and then that λ is given by (5.4).

So the eigenvalues are the same as before, but the eigenfunctions are different. This time the computation yields

$$f(t) = C\cos\Big(\frac{k\pi(x-a)}{b-a}\Big) \tag{5.10}$$

(compare with (5.5)).

Example 3: : the periodic condition. This is where we see $[a,b]$ as the torus $\mathbb{R}/(b-a)\mathbb{Z}$. We can rename $L^2(I)$ as L^2_{per} (the values at the boundary don't make sense in L^2), and call $W^{1,2}_{per}$ (respectively $W^{2,2}_{per}$) (the restrictions to I of) functions in $W^{1,2}_{loc}(\mathbb{R})$ (respectively $W^{2,2}_{loc}(\mathbb{R})$ that are $(b-a)$-periodic.

Can we diagonalize $-\Delta$ on $W^{1,2}_{per}$? As earlier, we have a kernel (the constant functions), and also first and second derivatives of periodic functions have integral 0. So we should define the inverse L of $-\Delta$ on the same set H as before (or the obvious equivalent H_{per}). We don't need to change P: it goes from H to $W^{1,2}_0(I) \subset W^{1,2}_{per}$, and then we apply P again, we run into $W^{2,2}_0(I) \subset W^{2,2}_{per}$. As before, $P^2(f)$ is not always orthogonal the kernel N of $-\Delta$ (and if we want a nice orthogonal diagonalization, we prefer our inverse L to map into $N^\perp$, so we project on $N^\perp$ by removing the integral of $P^2 f$. That is, we take the same formula for Lf as in (5.8).

As before, this defines an inverse $L : H \to W^{2,2} \cap H$ of $-\Delta : W^{2,2} \cap H = W^{2,2} \cap N^\perp$, and a simple form of Rellich Kondrachev says that L is compact. In addition, L is self-adjoint and positive by the proof in (5.9).

What are the eigenvalues and eigenvectors for $-\Delta$ this time? We still have the eigenvalue 0, with the space E_0 of constant functions, and otherwise $\lambda > 0$, the eigenvalues still satisfy $f'' = -\lambda f$ on I, hence $f(x) = \alpha exp(i\sqrt{\lambda}x) + \beta exp(-i\sqrt{\lambda}x)$. Now we have the periodicity constraint, which we can write down as $f(b) = f(a)$ and $f'(b) = f'(a)$ (you could imagine other conditions maybe, but this will be enough). The first equation is $\alpha exp(i\sqrt{\lambda}b) + \beta exp(-i\sqrt{\lambda}b) = X$, where X is the same thing at the point a. The second equation is the same except everything is multiplied by $\sqrt{\lambda}$, and also β is replaced by $-\beta$. We add and subtract and find an equation for α, namely $\alpha exp(i\sqrt{\lambda}b) = \alpha exp(i\sqrt{\lambda}a)$, and the same for β. So we get the necessary condition that $exp(i\sqrt{\lambda}(b-a)) = 1$, i.e. $\lambda(b-a) \in 2\pi\mathbb{Z}$. So now the eigenvalues are given by

$$\lambda = k^2(2\pi)^2(b-a)^{-2} \ \text{ for some integer } k \geq 1. \tag{5.11}$$

not exactly as in (5.4), and each of its number is an eigenvalue, because both $exp(i\sqrt{\lambda}x)$ and $exp(-i\sqrt{\lambda}x)$ work. In the simple case of $I = [0, 2\pi]$, the eigenvalues are the $\lambda = k^2$ and the basis of eigenvalues that we get is the usual trigonometric system e^{ikx}, $k \in bZ$. Not so surprising after the fact.

Exercise. Consider $D : f \to -if'$, defined on $W^{1,2}_{per}(I)$. This time, can you find an inverse to $D : W^{2,2}_{per} \cap H \to H$, which is self-adjoint and diagonalizable? Does the answer fit with the previous exercise? Observe that our chances are much better than in the remark above because the $e^{ik\frac{x}{b-a}}$ are a basis of eigenvalues!

Exercise. Finally I leave the most obvious example, the diagonalization of Δ on the large space $W^{1,2}$ as an exercise. I'll suggest two ways here, but the reader should be warned that I did not do the exercise entirely. Let us take $I = [0, \pi]$ to simplify the computations.

First we try brutally. The image of $W^{2,2}$ by $-\Delta$ is the same as in the previous two examples, i.e., H, because $W^{2,2}$ is the direct sum of $W^{2,2}_0$, for instance, with the set of affine functions that are in the kernel anyway. Then we compute the kernel N of $-\Delta$ on $W^{2,2}$, and we find that N is the span of $e_0 = 1/(2\pi)$ and $e_1 = \widetilde{e}_1/||\widetilde{e}_1||$, where $\widetilde{e}_1(x) = x - \pi$ (chosen to be orthogonal to e_0). Then we can try to devise an inverse. A first try is L as above but it only sends H to $W^{2,2} \cap H$ (even on $W^{2,2}_n \cap H$, but this won't help). So we compose it with π, the orthonormal projection on $e_1^\perp$. This does not change $\Delta \circ L$, because we add a multiple of e_1, but now the image is in $W^{2,2} \cap N$.

Is this one self-adjoint? Normally yes, essentially because π is self-adjoint on L^2, but I let you check. Compactness is ass usual, and so (maybe) we have the desired inverse $\widetilde{L} = pi \circ L$ of $-\Delta : W^{1,2} \cap N$ to H. We are left with the task of diagonalizing $\widetilde{L}$; as usual the eigenfunctions are of the form $f(x) = \alpha exp(i\sqrt{\lambda}x) + \beta exp(-i\sqrt{\lambda}x)$, but now the boundary constraints are different (and less pleasant): f should just have vanishing integral and first moment. I let you try.

Then there is a slightly less brutal way. We know that $W^{2,2}_n$ is a subspace of $W^{2,2}$, and we can find a basis of the orthogonal complement of $W^{2,2}_n$ in $W^{2,2}$: we take the two functions $\sin(2x)$ and $\cos(2x)$, correctly normalized (adding a multiple of one of them fixes the value of f' at one endpoint without changing the other). Now we are lucky: both functions are eigenvectors of $-\Delta$, with eigenvalue 4. We can complete with a basis of $W^{2,2}_n$.

5.2 Solutions of $-\Delta u = f$ on a bounded domain Ω

We guessed that the case of the Laplacian on an open domain $\Omega \subset \mathbb{R}^n$ should be more interesting (but we won't be able to compute everything!). Recall that

$$\Delta f = \sum_{j=1}^{n} \frac{\partial^2 f}{\partial x_j^2} \tag{5.12}$$

(for f smooth, or else we consider $f \in L^1_{loc}(\Omega)$ and take this in the sense of distributions).

We'll try to follow the same program as above: find an orthonormal basis of $L^2(\Omega)$ which is entirely composed of eigenvectors for Δ.

As before, we will not try to do the spectral theory of the unbounded operator $-\Delta$, even though this would be possible because, with the right definitions, $-\Delta$ is self-adjoint and even positive. Instead we take a more analytic road and try to invert $-\Delta$ in some appropriate space (here L^2). Our life will be easier if

$$\Omega \text{ is bounded,} \tag{5.13}$$

so even though this is not really necessary, we assume it. Also we assume that Ω is connected (otherwise, $-\Delta$ acts independently on each component of Ω and we are just complicating our lives uselessly).

Strangely the inverse mapping that we will construct will go from L^2 to the space $W_0^{1,2}(\Omega)$, which is the closure of C_c^∞ in $W_0^{1,2}(\Omega)$. The reader may have expected the more natural $W_0^{2,2}(\Omega)$, where at least the Laplacian is well defined, but it turns out we don't want to have the unpleasant task of estimating second derivatives other than the Laplacian, so we'll manage with $W_0^{1,2}$. Our main tool here will thus be the following theorem.

Theorem 5.1. *For each $f \in L^2(\Omega)$, there is a unique function $u \in W_0^{1,2}(\Omega)$ such that*

$$\Delta u = -f \quad \textit{(as a distribution).} \tag{5.14}$$

In addition $||u||_{W^{1,2}(\Omega)} \leq C_\Omega ||f||_{L^2}$.

We'll prove this first, and you probably have an idea of how we will try to proceed then.

Here I created trouble for myself (a little bit on purpose). I should have said u is a weak solution to $\Delta u = -f$. See the discussion below; at least the advantage here is that (5.14) is well defined from what we said so far.

First of all, let us relax a little with the complex numbers. We will prove this for real-valued g (and with real-valued solutions f). Then for complex-valued g, we can easily find a solution (solve two equations). The uniqueness is not a bother either: if there is no nontrivial real solution for $\Delta u = 0$, there is no complex solution either (because the real and imaginary parts would be solutions). For other operators than Δ (typically, operators with complex coefficients) we would need to be more prudent, but here no, so we can happily forget the complex numbers.

Next $\Delta u = -f$ is a very simple example of what one may call a variational equation: there is a functional (we shall call it Q below) such that the minimizers of this functional satisfy that equation as an Euler-Lagrange equation. This makes it simpler to find solutions, because **minimizing functionals is always fun**.

So here we go: we want to solve $\Delta u = -f$, and we introduce a functional $Q = Q_f$, defined by

$$Q(u) = \int_\Omega |\nabla u|^2 - 2fu \tag{5.15}$$

[for f with complex coefficients we would take $\int Re(fu)$ here, but we don't need to]. The natural domain of definition of Q that we will take is precisely $W_0^{1,2} = W_0^{1,2}(\Omega)$.

Notice that the linear part $u \to -2\int fu$ is continuous on $W_0^{1,2}(\Omega)$. because $|\int_\Omega fu| \leq \int_\Omega |fu| \leq ||f||_{L^2(\Omega)}||u||_{L^2(\Omega)} \leq ||f||_{L^2(\Omega)}||u||_{W^{1,2}(\Omega)}$. The other part $||\nabla u||^2$ is part of the square of the norm, so it is (defined and) continuous too. So Q is well defined on $W_0^{1,2}$.

Now we want to minimize Q in $W_0^{1,2}$. And we claim it is strictly convex, which will help a lot. That is, for $u, v \in W_0^{1,2}$,

$$Q\Big(\frac{u+v}{2}\Big)-\frac{1}{2}[Q(u)+Q(v)] = \int_\Omega |\nabla\Big(\frac{u+v}{2}\Big)|^2-\frac{1}{2}\big(|\nabla u|^2+|\nabla v|^2\big) = -\frac{1}{4}\int_\Omega |\nabla(u-v)|^2 \tag{5.16}$$

(naturally the linear part goes away). So the average does significantly better than u or v. Then the existence of a unique minimizer, i.e.,

$$u_0 \in W_0^{1,2} \text{ such that } Q(u_0) = m := \inf_{u\in W_0^{1,2}} Q(u) \tag{5.17}$$

follows in a standard way that we explain now.

Before we start, observe that $m \leq 0$ (try $u = 0$). Next notice that if $Q(u) \leq 0$, then $\int_\Omega |\nabla u|^2 \leq 2\int |fu| \leq 2||f||||u|| \leq C||f||\big\{\int_\Omega |\nabla u|^2\big\}^{1/2}$, because the version of Poincaré's inequality for compactly supported functions that we proved above yieds

$$||u||_2 \leq C_\Omega ||\nabla u||_2 \text{ for } u \in W_0^{1,2} : \tag{5.18}$$

see the version of Poincaré's inequality for compactly supported functions. We simplify and find that $\big\{\int_\Omega |\nabla u|^2\big\}^{1/2} \leq C||f||_2$ when $Q(u) \leq 0$.

Because of this, m is the same if we restrict our attention to the $u \in W^{1,2}$ such that $\big\{\int_\Omega |\nabla u|^2\big\}^{1/2} \leq C||f||_2$. For these it is obvious that $2\int |fu| \leq 2||f||||u|| \leq C||f||_2$, so in particular m is finite.

Let return to the existence and uniqueness. Let $\{u_k\}$ be a minimizing sequence, and notice that if k and l large enough, then $Q(u_k) \leq m+\varepsilon$ and $Q(u_l) \leq m+\varepsilon$, so by (5.16), not only $u_{k,l} = \frac{u_k+u_l}{2}$ does better than u_k and u_l on average, but even

$$Q(u_{k,l}) \leq \frac{1}{2}[Q(u_k)+Q(u_l)] - \frac{1}{4}\int_\Omega |\nabla(u_k-u_l)|^2 \leq m+\varepsilon-\frac{1}{4}\int_\Omega |\nabla(u_k-u_l)|^2,$$

which forces $\int_\Omega |\nabla(u_k-u_l)|^2 \leq 4\varepsilon$. That is,

$$\{\nabla u_k\} \text{ is a Cauchy sequence in } L^2(\Omega). \tag{5.19}$$

This immediately forces $\{u_k\}$ to be a Cauchy sequence in $L^2(\Omega)$. Call the limit u. We want to show that u is a minimizer of Q (as in (5.17)).

Since $\{\nabla u_k\}$ is also a Cauchy sequence in $L^2(\Omega)$, it has a limit V (a vector-valued function). Now there is a lemma that says that in any ball B, if $\{u_k\}$ is a sequence in $W^{1,k}(B)$ that converges in $L^k(B)$ to a limit u, and such that $\{\nabla u_k\}$ converges to $V \in L^k(B)$, then $u \in W^{1,k}$ and its distribution derivative is given by V. Here we are interested in $k = 2$, but any $k \geq 1$ would work, and we could also weaken our convergence assumptions. The easy proof consists in observing that for every test function φ, the $\int u_k \frac{\partial\varphi}{\partial x_j} = -\int \frac{\partial u_k}{\partial x_j}\varphi$ converge to $\int u\frac{\partial\varphi}{\partial x_j} = -\int V_j\varphi$.

Thus $u \in W^{1,2}$ and $\{u_k\}$ converges to u in $W^{1,2}(\Omega)$ (by (5.19) and (5.18)). In addition, $u \in W_0^{1,2}$ because the u_k lie in $W_0^{1,2}$ (which is closed in $W^{1,2}$ by definition). Finally it is easy to see that $Q(u) = \lim_{k\to+\infty} Q(u_k) = m$ (the functional is continuous on $W^{1,2}$, for the same reasons as we said it is defined). So u is a minimizer.

Incidentally, the minimizer of Q is unique, by the convexity property (5.16). That's the advantage of (strictly) convex functionals, and otherwise any minimizing sequence would not converge readily without at least extracting a subsequence.

We also promised that $||u||_{W^{1,2}} \leq C||f||_2$. This comes from the fact that the minimum m of the functional is less than 0 (try $u = 0$), then by (5.15), the fact that $2|\int fu| \leq 2||f||_2||u||_2$, and (5.18).

Comment. Here we wanted to go fast, so se used convexity and the special form of Δ. But there is a (quite classical) similar argument that works also, at least for the operators L of the form (called divergence form) $L = -\mathrm{div}A\nabla$, where $A = A(x)$ is a matrix with bounded measurable coefficients that satisfy a suitable ellipticity condition. Then our convexity trick can be replaced by a simple but powerful tool, **the Lax-Milgran theorem**. It is good to know that to $L = -\mathrm{div}A\nabla$ we can associate a sesquilinear form on $W^{1,2}(\Omega)$, given by

$$\mathcal{F}(u,v) = \int_\Omega \langle A\nabla u, \nabla v\rangle, \tag{5.20}$$

which is just $\int_\Omega \langle \nabla u, \nabla v\rangle$ when $L = -\Delta$, and which has some good positivity properties.

Return to $-\Delta$ and $Q = Q_f$. Now we have our minimizer for Q and we need to show that it is the unique solution of $-\Delta u = f$ on $W_0^{1,2}$. We first show that u is a weak solution of $\Delta u = -f$ in Ω. This means that for every test function $\varphi \in C_c^\infty(\Omega)$,

$$\langle \nabla u, \nabla\varphi\rangle_{L^2} = \langle f, \varphi\rangle_{L^2} \tag{5.21}$$

or if you prefer, more explicitely

$$\int_\Omega \nabla u \cdot \nabla\varphi = \int_\Omega f\varphi \tag{5.22}$$

(we decided to stay real, so we don't need to put bars in the φ-terms). This is what you would get from an integration by parts if you knew that u is smooth and is a "strong" solution, since we could write that

$$0 = \int_\Omega (\Delta u + f)\varphi = \int_V f\varphi + \int_V (\Delta u)\varphi = \int_V f\varphi - \int_V \nabla u \cdot \nabla\varphi, \tag{5.23}$$

where as a precaution I decided that I could reduce the first integral to a nice smooth domain $V \subset \Omega$ that is large enough to contain a neighborhood of the support of λ, so that when we apply Green's formula to get the last identity, the integrated term on ∂V vanishes because of φ.

Now we show that u is a weak solution. We use the fact that for all $t \in \mathbb{R}$, $u + t\varphi$ lies in $W_0^{1,2}$, so $Q(u + t\varphi) \geq Q(u)$ (by minimization). We use (5.15) and expand:

$$Q(u + t\varphi) = \int_\Omega |\nabla(u + t\varphi)|^2 - 2f(u + t\varphi) = Q(u) + 2t \int_\Omega \nabla u \cdot \nabla\varphi - f\varphi + t^2 A, \tag{5.24}$$

where we don't care about A, because we just need to know that $Q(u + t\varphi)$ has a derivative at $t = 0$, and say that this derivative vanishes because u is a minimizer. We get that $\int_\Omega \nabla u \cdot \nabla\varphi - f\varphi = 0$, as required for (5.22).

We promised a distribution solution and now we obtained a weak solution; what is the difference? We need to explain more what happens with distribution derivatives. At this point we know that $u \in W_0^{1,2}$, which means that the first derivatives $\partial_j u = \frac{\partial u}{\partial x_j}$ are given by L^2 functions (call them v_j). By (5.14) we mean that for every test function φ,

$$0 = \langle \Delta u, \varphi \rangle + \langle f, \varphi \rangle = \int_\Omega f\varphi - \sum_j \langle \partial_j u, \partial_j \varphi \rangle = \int_\Omega f\varphi + \sum_j \int_\Omega \langle u, \partial_j \partial_j \varphi \rangle \tag{5.25}$$

where the first brackets describe the duality of a distribution with a measure, and then we use the definitions. The second part would give a direct definition of a distribution solution, but the first part is good enough for us, because $\sum_j \langle \partial_j u, \partial_j \varphi \rangle = \sum_j \langle v_j, \partial_j \varphi \rangle = \int_\Omega \nabla u \cdot \nabla\varphi$ and now we recognize (5.22).

Here is a related comment. A natural space that we could have tried instead of $W_0^{1,2}$ would have been $W_0^{2,2}(\Omega)$. Then we would immediately have had a definition of Δu as an L^2 function, which sounds fair. But we would also have needed to prove that all the other second derivatives of u also lie in L^2. Now Δ is special, and for instance in $\Omega = \mathbb{R}^n$ a simple Fourier transform computation shows that $u \in W^{2,2}(\mathbb{R}^n)$ as soon as u and Δu lie in L^2 (or as soon as $u - \Delta u \in L^2$); that is, Δu controls the other derivatives of order 2. The same thing actually holds in nice domains, but I have big doubts about ugly bounded open sets. But for what we want to do here, we don't need to prove that $u \in W_0^{2,2}(\Omega)$ and so $W_0^{1,2}(\Omega)$ is better.

Good, u is also a distribution solution. Now is it unique? That is, if $u \in W_0^{1,2}$ is a distribution (or weak) solution to $\Delta u = 0$, can we say that $u = 0$? Again, this looks all right: $-\Delta$ is the best known example of an elliptic operator, for which any weak solution is not only a strong solution, but C^∞. An in addition $u = 0$ (morally) on the boundary, so we expect u to vanish on Ω. But we don't want to prove all this, and we don't want to use any regularity for Ω. So we return to the functional.

If u is a weak solution of $\Delta u = 0$, then by (5.22) with $f = 0$, $\int_\Omega \nabla u \cdot \nabla\varphi = 0$ for all test functions φ. Take a sequence of test functions φ_k that converges to u in $W^{1,2}$; this is possible precisely because $u \in W_0^{1,2}$. And by taking limits above, $\int_\Omega |\nabla u|^2 = 0$. Now $u = 0$ in $W_0^{1,2}$, by (5.18). Too easy.

This completes the proof of Theorem 5.1. □

5.3 The spectrum of $-\Delta$ on a bounded domain Ω

Again bounded is not vital, but we assume that Ω is a bounded domain in $\mathbb{R}^n$.

We want to use Theorem 5.1 to say some things about the spectrum and eigenvalues of $-\Delta$ on Ω.

We want to repeat what we did in dimension 1, with less precise computations. Call G (for Green) the operator that to $f \in L^2(\Omega)$ associates the unique solution u of $-\Delta u = f$ obtained in Theorem 5.1. This is a bounded linear operator on $L^2 = L^2(\Omega)$; the linearity follows from the uniqueness in the theorem.

Now see G as an operator on L^2 (compose with the canonical injection). Then G is compact, because the image of the unit ball of L^2 is, by the theorem, contained in a ball of $W_0^{1,2}$, which is relatively compact in L^2 by the Rellich-Kondrachov theorem.

Next we claim that $G : L^2 \to L^2$ is self-adjoint and positive. Let $f, g \in L^2$ be given, and set $u = Gf$ and $v = Gg$. We have seen in (5.22) (the weak equation) that $\int_\Omega \nabla u \cdot \nabla \varphi = \int_\Omega f\varphi$ for every test function φ. This stays true, by density of C_c^∞ in $W_0^{1,2}$, when φ is replaced by v. Thus $\int_\Omega \nabla u \cdot \nabla v = \int_\Omega fv$. By symmetry, $\int_\Omega \nabla u \cdot \nabla v = \int_\Omega ug$. Thus $\int_\Omega fv = \int_\Omega ug$. This is also true with g replaced by $\overline{g}$, so we get (with easy to guess notation)

$$\langle Gf, g\rangle = \int_\Omega Gf\overline{g} = \int_\Omega u\overline{g} = \int_\Omega f\overline{v} = \langle f, Gg\rangle. \tag{5.26}$$

As for the positivity, we take $g = f$ and get that

$$\langle Gf, f\rangle = \int_\Omega u\overline{f} = \int_\Omega \nabla u \cdot \nabla \overline{u} = ||\nabla u||^2. \tag{5.27}$$

At this point, we can apply the spectral theorem. Notice that the kernel $Ker(G)$ is $\{0\}$, because if $u = 0$ then $f = -\Delta u = 0$. So we can find an orthonormal basis of eigenvectors for G. Let us summarize this, but by calling λ_j^{-1} the eigenvalues for G, so that the corresponding eigenfunctions satisfy $-\Delta f = \lambda_j f$.

Corollary 5.2. *There is an orthonormal basis $\{e_i\}$, $i \in \mathbb{N}$, of $L^2(\Omega)$, and a nondecreasing sequence $\{\lambda_i\}$, $i \in \mathbb{N}$ of positive numbers, with $\lim_{i\to+\infty} \lambda_i = +\infty$, such that*

$$-\Delta e_i = \lambda_i e_i \quad \textit{for } i \in \mathbb{N}. \tag{5.28}$$

Notice also that since $Ge_i = \lambda_i^{-1} e_i$, each e_i lies in $W_0^{1,2}(\Omega)$. Inside, due to the ellipticity of $-\Delta$, it is even smooth.

Do not dream of computing all the eigenvalues and eigenvectors of $-\Delta$ on $\Omega \subset \mathbb{R}^n$. In a very simple domain like the ball, maybe (I don't recall), but in general, no way.

But the first eigenvalue for $-\Delta$ (precisely the largest eigenvalue for G, because we did not compute $G(L^2)$) has a geometric sense, so let us discuss it. Here is a statement

Corollary 5.3. *Let μ_0 be the largest eigenvalue of $G : L^2 \to L^2$, and set $\lambda_0 = \mu_0^{-1}$. Then $\lambda_0 = C^2$, where C is the best constant in the Poincaré inequality*

$$||u||_{L^2(\Omega)} \leq C||\nabla u||_{L^2(\Omega)} \quad \textit{for } u \in W_0^{1,2}(\Omega). \tag{5.29}$$

And the best constant C in (5.29) is precisely obtained for $u \in E_{\mu_0} = \big\{u \in L^2\,;\, G(f) = \mu_0 f\big\}$

Thus the optimal functions for Poincaré are the span of the e_i that correspond to μ_0.

Don't get confused with too many inverses. The unit ball is actually the bounded open set with the worse (the largest) constant C, say, given its volume, and it fits with the fact that it has the largest spectral gap λ_0. Said otherwise, if Ω is very thin and small, then C can be taken very small. This looks good, but it prevents the existence of an eigenvector e_0 with $||e_0||$ not much smaller than the norm of its Laplacian.

Let us try to compute norms. Let $u = \sum_i \alpha_i e_i$ be any finite linear combination of the e_i, and set $f = \sum_i \lambda_i \alpha_i e_i$. Notice that $u = G(f)$ because $G(e_i) = \mu_i e_i$. Also, $||u||_2 = \sum_i |\alpha_i|^2$. Recall from (5.27) that

$$||\nabla u||_2^2 = \langle Gf, f\rangle = \Big\langle \sum \alpha_i e_i, \sum \lambda_i \alpha_i e_i \Big\rangle = \sum_i \lambda_i |\alpha_i|^2. \tag{5.30}$$

First suppose that f lies in the kernel of $G - \mu_0 I$. Then all the coefficients λ_i are equal to λ_0, and so $||\nabla u||_2^2 = \lambda_0 \sum_i |\alpha_i|^2 = \lambda_0 ||u||_2^2$. In this case (5.29) with $C = \sqrt{\lambda_0}$ is an identity for u.

It remains to show that all the other functions of $W_0^{1,2}$ do even better (so that then (5.29) holds with $C = \sqrt{\lambda_0}$). For finite linear combinations of the e_j, this follows from (5.30), because $||\nabla u||_2^2 \geq \lambda_0 \sum_i |\alpha_i|^2$. For the rest of $W_0^{1,0}$, we claim (but the reader should check the argument below with attention) that $G(L^2)$ is in fact dense in $W_0^{1,2}$. Since both sides of (5.29) are continuous functions of $u \in W_0^{1,2}$, we get the result by continuity.

Now we try to prove that $G(L^2)$ is dense in $W = W_0^{1,2}$. Let us put the norm $||u||_W^2 = ||u||_2^2 + ||\nabla u||_2^2$ on W. Then let $\xi \in W$ be orthogonal to $G(L^2)$ in W. This means in particular that for all j,

$$0 = \langle \xi, G(e_j)\rangle_W = \langle \xi, G(e_j)\rangle_{L^2} + \langle \nabla\xi, \nabla G(e_j)\rangle_{L^2} = \mu_i \langle \xi, e_j\rangle_{L^2} + \langle \nabla \xi, \nabla u_j\rangle_{L^2}$$

where we set $u_j = G(e_j)$. We apply (5.22) to $\varphi = \overline{u}_j$ (so, after a small limiting argument and maybe some conjugations) and get that $\langle \nabla\xi, \nabla u_j\rangle_{L^2} = \langle \xi, e_j\rangle_{L^2}$. Altogether, $0 = (\mu_i + 1)\langle \xi, e_j\rangle_{L^2}$ for all j, so $\xi = 0$ because the e_j are an orthonormal basis of L^2, and the proof of Corollary 5.3 is finished. □

6 Harmonic polynomials, homogeneous harmonic functions

Won't have time, did not prepare, but this sounds cool.

Look at harmonic functions on the unit disk.

We have a subspace E_n of such functions, which are homogeneous of degree n. For each one, the restriction to the circle $\mathbb{S}$ is an eigenfunction of the Laplacian (with eigenvalue $-k^2$). These eigenfunctions span the whole $L^2(\mathbb{S})$. And finally the corresponding spaces of homogeneous harmonic functions span the whole set of harmonic functions that lie in L^2, say.

Same story for harmonic polynomials on the unit ball of $\mathbb{R}^n$, eigenfunctions of the Laplacian on the sphere, and harmonic functions on a ball.

Same story for harmonic functions on a cone.

And there are some very nice relations between eigenvalues of the Laplacian on (a sector of) the sphere, and monotonicity formulas (Alt-Caffarelli-Friedman, Almgren) for the averages of a harmonic function in small balls. A nice (but more elaborate) subject, among others.