Functional Analysis
Lecture notes for 18.102, Spring 2020

Richard Melrose

Department of Mathematics, MIT

Version 0.9A; Revised: 29-9-2018; Run: May 16, 2020 .

Contents

Preface

These are notes for the course 'Introduction to Functional Analysis' – or in the MIT style, 18.102, from various years culminating in Spring 2020. There are many people who I should like to thank for comments on and corrections to the notes over the years, but for the moment I would simply like to thank, as a collective, the MIT undergraduates who have made this course a joy to teach, as a result of their interest and enthusiasm.

Introduction

This course is intended for 'well-prepared undergraduates' meaning specifically that they have a rigourous background in analysis at roughly the level of the first half of Rudin's book [**4**] – at MIT this is 18.100B. In particular the basic theory of metric spaces is used freely. Some familiarity with linear algebra is also assumed, but not at a very sophisticated level.

The main aim of the course in a mathematical sense is the presentation of the standard constructions of linear functional analysis, centred on Hilbert space and its most significant analytic realization as the Lebesgue space $L^2(\mathbb{R})$ and leading up to the spectral theory of ordinary differential operators. In a one-semester course at MIT it is only just possible to get this far. Beyond the core material I have included other topics that I believe may prove useful both in showing how to use the 'elementary' results in various directions.

Dirichlet problem. The treatment of the eigenvalue problem with potential perturbation on an interval is one of the aims of this course, so let me describe it briefly here for orientation.

Let $V : [0,1] \longrightarrow \mathbb{R}$ be a real-valued continuous function. We are interested in 'oscillating modes' on the interval; something like this arises in quantum mechanics for instance. Namely we want to know about functions $u(x)$ – twice continuously differentiable on $[0,1]$ so that things make sense – which satisfy the differential equation

$$\begin{gathered} -\frac{d^2u}{dx^2}(x) + V(x)u(x) = \lambda u(x) \\ \text{and the boundary conditions} \\ u(0) = u(1) = 0. \end{gathered} \tag{1}$$

Here the eigenvalue, λ is an 'unknown' constant. More precisely we wish to know which such λ's can occur. In fact all λ's can occur with $u \equiv 0$ but this is the 'trivial solution' which will always be there for such an equation. What other solutions are there? The main result is that there is an infinite sequence of λ's for which there is a non-trivial solution of (1) $\lambda_j \in \mathbb{R}$ – they are all real, no non-real complex λ's can occur. For each of these λ_j there is at least exactly a one-dimensional space of solutions, u_j, to (1). We can say a lot more about everything here but one main aim of this course is to get at least to this point. From a Physical point of view, (1) represents a linearized oscillating string with fixed ends.

The journey to a discussion of the Dirichlet problem is rather extended and apparently wayward. The relevance of Hilbert space and the Lebesgue integral is not immediately apparent – and indeed one can proved the results as stated above without Hilbert space methods – but I hope this will become clear as we proceed. It is the *completeness* of the eigenfunctions which uses Hilbert space.

The basic idea of functional analysis is that we consider a space of all 'putative' solutions to the problem at hand. In this case one might take the space of all twice continuously differentiable functions on $[0,1]$ – we will consider such spaces below. One of the weaknesses of this choice of space is that it is not closely connected with the 'energy' invariant of a solution, which is the integral

$$\int_0^1 (|\frac{du}{dx}|^2 + V(x)|u(x)|^2)dx. \tag{2}$$

It is the importance of such integrals which brings in the Lebesgue integral and leads to a Hilbert space structure.

In any case one of the significant properties of the equation (1) is that it is 'linear'. So we start with a brief discussion of linear (I usually say vector) spaces. What we are dealing with here can be thought of as the eigenvalue problem for an 'infinite matrix'. This in fact is not a very good way of thinking about operators on infinite-dimensional spaces, they are not really like infinite matrices, but in this case it is justified by the appearance of *compact operators* which *are* rather more like infinite matrices. There was a matrix approach to quantum mechanics in the early days but it was replaced by the sort of 'operator' theory on Hilbert space that we will discuss below. One of the crucial distinctions between the treatment of finite dimensional matrices and an infinite dimensional setting is that in the latter *topology* is encountered. This is enshrined in the notion of a *normed linear space* which is the first important topic we shall meet.

After a brief treatment of normed and Banach spaces, the course proceeds to the construction of the Lebesgue integral and the associated spaces of 'Lebesgue integrable functions' (as you will see this is by way of a universally accepted falsehood, but a useful one). To some extent I follow here the idea of Jan Mikusiński that one can simply define integrable functions as the almost everywhere limits of absolutely summable series of step functions and more significantly the basic properties can be deduced this way. While still using this basic approach I have dropped the step functions almost completely and instead emphasize the completion of the space of continuous functions to get the Lebesgue space. Even so, Mikusiński's approach still underlies the explicit identification of elements of the completion with Lebesgue 'functions'. This approach is followed in the book of Debnaith and Mikusiński [**1**].

After about a two-week stint of integration and then a little measure theory the course proceeds to the more gentle ground of Hilbert spaces. Here I have been most guided by the (old now) book of Simmons [**5**] which is still very much worth reading. We proceed to a short discussion of operators and the spectral theorem for compact self-adjoint operators. I have also included in the notes (but generally not in the lectures) various things that a young mathematician should know(!) such as Kuiper's Theorem. Then in the last third or so of the semester this theory is applied to the treatment of the Dirichlet eigenvalue problem, followed by a short discussion of the Fourier transform and the harmonic oscillator. Finally various loose ends are brought together, or at least that is my hope.

CHAPTER 1

Normed and Banach spaces

In this chapter we introduce the basic setting of functional analysis, in the form of normed spaces and bounded linear operators. We are particularly interested in complete, i.e. Banach, spaces and the process of completion of a normed space to a Banach space. In lectures I proceed to the next chapter, on Lebesgue integration after Section 7 and then return to the later sections of this chapter at appropriate points in the course.

There are many good references for this material and it is always a good idea to get at least a couple of different views. The treatment here, whilst quite brief, does cover what is needed later.

1. Vector spaces

You should have some familiarity with linear, or I will usually say 'vector', spaces. Should I break out the axioms? Not here I think, but they are included in Section 14 at the end of the chapter. In short it is a space V in which we can add elements and multiply by scalars with rules quite familiar to you from the the basic examples of $\mathbb{R}^n$ or $\mathbb{C}^n$. Whilst these special cases are (very) important below, this is not what we are interested in studying here. What we want to come to grips with are spaces of *functions* hence the name of the course.

Note that for us the 'scalars' are either the real numbers or the complex numbers – usually the latter. To be neutral we denote by $\mathbb{K}$ either $\mathbb{R}$ or $\mathbb{C}$, but of course consistently. Then our set V – the set of vectors with which we will deal, comes with two 'laws'. These are maps

$$+ : V \times V \longrightarrow V,\ \cdot : \mathbb{K} \times V \longrightarrow V. \tag{1.1}$$

which we denote not by $+(v, w)$ and $\cdot(s, v)$ but by $v + w$ and sv. Then we impose the *axioms of a vector space* – see Section 14 below! These are commutative group axioms for $+$, axioms for the action of $\mathbb{K}$ and the distributive law linking the two.

The basic examples:

- The field $\mathbb{K}$ which is either $\mathbb{R}$ or $\mathbb{C}$ is a vector space over itself.
- The vector spaces $\mathbb{K}^n$ consisting of ordered n-tuples of elements of $\mathbb{K}$. Addition is by components and the action of $\mathbb{K}$ is by multiplication on all components. You should be reasonably familiar with these spaces and other finite dimensional vector spaces.
- Seriously non-trivial examples such as $\mathrm{C}([0, 1])$ the space of continuous functions on $[0, 1]$ (say with complex values).

In these and many other examples we will encounter below, the 'component addition' corresponds to the addition of functions.

Lemma 1.1. *If X is a set then the spaces of all functions*

$$\mathcal{F}(X;\mathbb{R}) = \{u : X \longrightarrow \mathbb{R}\},\ \mathcal{F}(X;\mathbb{C}) = \{u : X \longrightarrow \mathbb{C}\} \tag{1.2}$$

are vector spaces over $\mathbb{R}$ and $\mathbb{C}$ respectively.

Non-Proof. Since I have not written out the axioms of a vector space it is hard to check this – and I leave it to you as the first of many important exercises. In fact, better do it more generally as in Problem 1.2 – then you can sound sophisticated by saying 'if V is a linear space then $\mathcal{F}(X;V)$ inherits a linear structure'. The main point to make sure you understand is precisely this; because we *do* know how to add and multiply in either $\mathbb{R}$ and $\mathbb{C}$, we can add functions and multiply them by constants (we can multiply functions by each other but that is not part of the definition of a vector space so we ignore it for the moment since many of the spaces of functions we consider below are *not* multiplicative in this sense):-

$$(c_1f_1 + c_2f_2)(x) = c_1f_1(x) + c_2f_2(x) \tag{1.3}$$

defines the function $c_1f_1 + c_2f_2$ if $c_1,\ c_2 \in \mathbb{K}$ and $f_1,\ f_2 \in \mathcal{F}(X;\mathbb{K})$. □

You should also be familiar with the notions of linear subspace and quotient space. These are discussed a little below and most of the linear spaces we will meet are either subspaces of these function-type spaces, or quotients of such subspaces – see Problems 1.3 and 1.5.

Although you are probably most comfortable with finite-dimensional vector spaces it is the infinite-dimensional case that is most important here. The notion of dimension is based on the concept of the linear independence of a subset of a vector space. Thus a subset $E \subset V$ is said to be *linearly independent* if for any finite collection of distinct elements $v_i \in E$, $i = 1, \dots, N$, and any collection of 'constants' $a_i \in \mathbb{K}$, $i = 1, \dots, N$ we have the following implication

$$\sum_{i=1}^{N} a_iv_i = 0 \Longrightarrow a_i = 0\ \forall\ i. \tag{1.4}$$

That is, it is a set in which there are 'no non-trivial finite linear dependence relations between the elements'. A vector space is finite-dimensional if every linearly independent subset is finite. It follows in this case that there is a finite and maximal linearly independent subset – a basis – where maximal means that if any new element is added to the set E then it is no longer linearly independent. A basic result is that any two such 'bases' in a finite dimensional vector space have the same number of elements – an outline of the finite-dimensional theory can be found in Problem 1.1.

Still it is time to leave this secure domain behind, since we are most interested in the other case, namely infinite-dimensional vector spaces. As usual with such mysterious-sounding terms as 'infinite-dimensional' it is defined by negation.

Definition 1.1. A vector space is infinite-dimensional if it is not finite dimensional, i.e. for any $N \in \mathbb{N}$ there exist N elements with no, non-trivial, linear dependence relation between them.

Thus the infinite-dimensional vector spaces, which you may be quite keen to understand, appear just as the non-existence of something. That is, it is the 'residual' case, where there is no finite basis. This means that it is 'big'.

So, finite-dimensional vector spaces have finite bases, infinite-dimensional vector spaces do not. Make sure that you see the gap between these two cases, i.e. either a vector space has a finite-dimensional basis or else it has an infinite linearly independent set. In particular if there is a linearly independent set with N elements for any N then there is an infinite one, there is a point here – if an independent finite set has the property that there is no element of the space which can be added to it so that it remains independent then it already is a basis and any other independent set has the same or fewer elements.

The notion of a basis in an infinite-dimensional vector spaces needs to be modified to be useful analytically. Convince yourself that the vector space in Lemma 1.1 is infinite dimensional if and only if X is infinite. [1]

2. Normed spaces

We need to deal effectively with infinite-dimensional vector spaces. To do so we need the control given by a metric (or even more generally a non-metric topology, but we will only get to that much later in this course; first things first). A norm on a vector space leads to a metric which is 'compatible' with the linear structure.

Definition 1.2. A norm on a vector space is a function, traditionally denoted

$$\|\cdot\| : V \longrightarrow [0,\infty), \tag{1.5}$$

with the following properties

(*Definiteness*)

$$v \in V,\ \|v\| = 0 \Longrightarrow v = 0. \tag{1.6}$$

(*Absolute homogeneity*) For any $\lambda \in \mathbb{K}$ and $v \in V$,

$$\|\lambda v\| = |\lambda|\|v\|. \tag{1.7}$$

(*Triangle Inequality*) For any two elements v, $w \in V$

$$\|v + w\| \le \|v\| + \|w\|. \tag{1.8}$$

Note that (1.7) implies that $\|0\| = 0$. Thus (1.6) means that $\|v\| = 0$ is equivalent to $v = 0$. This definition is based on the same properties holding for the standard norm(s), $|z|$, on $\mathbb{R}$ and $\mathbb{C}$. You should make sure you understand that

$$\begin{gathered}\mathbb{R} \ni x \longrightarrow |x| = \begin{cases} x & \text{if } x \ge 0 \\ -x & \text{if } x \le 0 \end{cases} \in [0,\infty) \text{ is a norm as is} \\ \mathbb{C} \ni z = x + iy \longrightarrow |z| = (x^2 + y^2)^{\frac{1}{2}}. \end{gathered} \tag{1.9}$$

Situations do arise in which we do not have (1.6):-

Definition 1.3. A function (1.5) which satisfes (1.7) and (1.8) but possibly not (1.6) is called a *seminorm.*

[1] Hint: For each point $y \in X$ consider the function $f : X \longrightarrow \mathbb{C}$ which takes the value 1 at y and 0 at every other point. Show that if X is finite then any function $X \longrightarrow \mathbb{C}$ is a finite linear combination of these, and if X is infinite then this is an infinite set with no finite linear relations between the elements.

A metric, or distance function, on a set is a map

$$d : X \times X \longrightarrow [0, \infty) \tag{1.10}$$

satisfying three standard conditions

$$d(x, y) = 0 \Longleftrightarrow x = y, \tag{1.11}$$

$$d(x, y) = d(y, x) \ \forall \ x, y \in X \text{ and} \tag{1.12}$$

$$d(x, y) \leq d(x, z) + d(z, y) \ \forall \ x, y, z \in X. \tag{1.13}$$

As you are no doubt aware, a set equipped with such a metric function is called a metric space.

If you do not know about metric spaces, then you are in trouble. I suggest that you take the appropriate course now and come back next year. You could read the first few chapters of Rudin's book [**4**] before trying to proceed much further but it will be a struggle to say the least. The point is

Proposition 1.1. *If $\|\cdot\|$ is a norm on V then*

$$d(v, w) = \|v - w\| \tag{1.14}$$

is a distance on V turning it into a metric space.

Proof. Clearly (1.11) corresponds to (1.6), (1.12) arises from the special case $\lambda = -1$ of (1.7) and (1.13) arises from (1.8). □

We will not use any special notation for the metric, nor usually mention it explicitly – we just subsume all of metric space theory from now on. So $\|v - w\|$ is the distance between two points in a normed space.

Now, we need to talk about a few examples; there are more in Section 7. The most basic ones are the usual finite-dimensional spaces $\mathbb{R}^n$ and $\mathbb{C}^n$ with their Euclidean norms

$$|x| = \left(\sum_i |x_i|^2\right)^{\frac{1}{2}} \tag{1.15}$$

where it is at first confusing that we just use single bars for the norm, just as for $\mathbb{R}$ and $\mathbb{C}$, but you just need to get used to that.

There are other norms on $\mathbb{C}^n$ (I will mostly talk about the complex case, but the real case is essentially the same). The two most obvious ones are

$$\begin{gathered} |x|_\infty = \max |x_i|, \ x = (x_1, \dots, x_n) \in \mathbb{C}^n, \\ |x|_1 = \sum_i |x_i| \end{gathered} \tag{1.16}$$

but as you will see (if you do the problems) there are also the norms

$$|x|_p = (\sum_i |x_i|^p)^{\frac{1}{p}}, \ 1 \leq p < \infty. \tag{1.17}$$

In fact, for $p = 1$, (1.17) reduces to the second norm in (1.16) and in a certain sense the case $p = \infty$ is consistent with the first norm there.

In lectures I usually do not discuss the notion of equivalence of norms straight away. However, two norms on the one vector space – which we can denote $\|\cdot\|_{(1)}$ and $\|\cdot\|_{(2)}$ are *equivalent* if there exist constants C_1 and C_2 such that

$$\|v\|_{(1)} \leq C_1 \|v\|_{(2)}, \ \|v\|_{(2)} \leq C_2 \|v\|_{(1)} \ \forall \ v \in V. \tag{1.18}$$

The equivalence of the norms implies that the metrics define the same open sets – the topologies induced are the same. You might like to check that the reverse is also true, if two norms induced the same topologies (just meaning the same collection of open sets) through their associated metrics, then they are equivalent in the sense of (1.18) (there are more efficient ways of doing this if you wait a little).

Look at Problem 1.6 to see why we are not so interested in norms in the finite-dimensional case – namely any two norms on a finite-dimensional vector space are equivalent and so in that case a choice of norm does not tell us much, although it certainly has its uses.

One important class of normed spaces consists of the spaces of bounded continuous functions on a metric space X :

$$\mathcal{C}_\infty(X) = \mathcal{C}_\infty(X;\mathbb{C}) = \{u : X \longrightarrow \mathbb{C},\ \text{continuous and bounded}\}\,. \tag{1.19}$$

That this is a linear space follows from the (pretty obvious) result that a linear combination of bounded functions is bounded and the (less obvious) result that a linear combination of continuous functions is continuous; this we are supposed to know. The norm is the best bound

$$\|u\|_\infty = \sup_{x\in X} |u(x)|. \tag{1.20}$$

That this *is* a norm is straightforward to check. Absolute homogeneity is clear, $\|\lambda u\|_\infty = |\lambda|\|u\|_\infty$ and $\|u\|_\infty = 0$ means that $u(x) = 0$ for all $x \in X$ which is exactly what it means for a function to vanish. The triangle inequality 'is inherited from $\mathbb{C}$' since for any two functions and any point,

$$|(u+v)(x)| \le |u(x)| + |v(x)| \le \|u\|_\infty + \|v\|_\infty \tag{1.21}$$

by the definition of the norms, and taking the supremum of the left gives

$$\|u+v\|_\infty \le \|u\|_\infty + \|v\|_\infty.$$

Of course the norm (1.20) is defined even for bounded, not necessarily continuous functions on X. Note that convergence of a sequence $u_n \in \mathcal{C}_\infty(X)$ (remember this means with respect to the distance induced by the norm) is precisely *uniform convergence*

$$\|u_n - v\|_\infty \to 0 \Longleftrightarrow u_n(x) \to v(x) \text{ uniformly on } X. \tag{1.22}$$

Other examples of infinite-dimensional normed spaces are the spaces l^p, $1 \le p \le \infty$ discussed in the problems below. Of these l^2 is the most important for us. It is in fact one form of Hilbert space, with which we are primarily concerned:-

$$l^2 = \{a : \mathbb{N} \longrightarrow \mathbb{C}; \sum_j |a(j)|^2 < \infty\}. \tag{1.23}$$

It is not immediately obvious that this is a linear space, nor that

$$\|a\|_2 = \left(\sum_j |a(j)|^2\right)^{\frac{1}{2}} \tag{1.24}$$

is a norm. It is. From now on we will generally use sequential notation and think of a map from $\mathbb{N}$ to $\mathbb{C}$ as a sequence, so setting $a(j) = a_j$. Thus the 'Hilbert space' l^2 consists of the square summable sequences.

3. Banach spaces

You are supposed to remember from metric space theory that there are three crucial properties, completeness, compactness and connectedness. It turns out that normed spaces are always connected, so that is not very interesting, and they are never compact (unless you consider the trivial case $V = \{0\}$) so that is not very interesting either – in fact we will ultimately be very interested in compact subsets. So that leaves completeness. This is so important that we give it a special name in honour of Stefan Banach who first emphasized this property.

Definition 1.4. A normed space which is complete with respect to the induced metric is a *Banach* space.

Lemma 1.2. *The space $\mathcal{C}_\infty(X)$, defined in* (1.19) *for any metric space X, is a Banach space.*

Proof. This is a standard result from metric space theory – basically that the uniform limit of a sequence of (bounded) continuous functions on a metric space is continuous. However, it is worth recalling how one proves completeness at least in outline. Suppose u_n is a Cauchy sequence in $\mathcal{C}_\infty(X)$. This means that given $\delta > 0$ there exists N such that

$$n, m > N \Longrightarrow \|u_n - u_m\|_\infty = \sup_{x\in X} |u_n(x) - u_m(x)| < \delta. \tag{1.25}$$

Fixing $x \in X$ this implies that the sequence $u_n(x)$ is Cauchy in $\mathbb{C}$. We know that this space is complete, so each sequence $u_n(x)$ must converge (we say the sequence of functions converges pointwise). Since the limit of $u_n(x)$ can only depend on x, we may define $u(x) = \lim_n u_n(x)$ in $\mathbb{C}$ for each $x \in X$ and so define a function $u : X \longrightarrow \mathbb{C}$. Now, we need to show that this is bounded and continuous and is the limit of u_n with respect to the norm. Any Cauchy sequence is bounded in norm – take $\delta = 1$ in (1.25) and it follows from the triangle inequality that

$$\|u_m\|_\infty \le \|u_{N+1}\|_\infty + 1,\ m > N \tag{1.26}$$

and the finite set $\|u_n\|_\infty$ for $n \le N$ is certainly bounded. Thus $\|u_n\|_\infty \le C$, but this means $|u_n(x)| \le C$ for all $x \in X$ and hence $|u(x)| \le C$ by properties of convergence in $\mathbb{C}$ and thus $\|u\|_\infty \le C$, so the limit is bounded.

The uniform convergence of u_n to u now follows from (1.25) since we may pass to the limit in the inequality to find

$$\begin{aligned} n > N \Longrightarrow |u_n(x) - u(x)| &= \lim_{m\to\infty} |u_n(x) - u_m(x)| \le \delta \\ &\Longrightarrow \|u_n - u\|_\infty \le \delta. \end{aligned} \tag{1.27}$$

The continuity of u at $x \in X$ follows from the triangle inequality in the form

$$\begin{aligned} |u(y) - u(x)| \le |u(y) - u_n(y)| + |u_n(y) - u_n(x)| + |u_n(x) - u(x)| \\ \le 2\|u - u_n\|_\infty + |u_n(x) - u_n(y)|. \end{aligned}$$

Given $\delta > 0$ the first term on the far right can be make less than $\delta/2$ by choosing n large using (1.27) and then, having chosen n, the second term can be made less than $\delta/2$ by choosing $d(x, y)$ small enough, using the continuity of u_n. □

I have written out this proof (succinctly) because this general structure arises often below – first find a candidate for the limit and then show it has the properties that are required.

There is a space of sequences which is really an example of this Lemma. Consider the space c_0 consisting of all the sequences $\{a_j\}$ (valued in $\mathbb{C}$) such that $\lim_{j\to\infty} a_j = 0$. As remarked above, sequences are just functions $\mathbb{N} \longrightarrow \mathbb{C}$. If we make $\{a_j\}$ into a function $\alpha : D = \{1, 1/2, 1/3, \dots\} \longrightarrow \mathbb{C}$ by setting $\alpha(1/j) = a_j$ then we get a function on the metric space D. Add 0 to D to get $\overline{D} = D \cup \{0\} \subset [0, 1] \subset \mathbb{R}$; clearly 0 is a limit point of D and $\overline{D}$ is, as the notation dangerously indicates, the closure of D in $\mathbb{R}$. Now, you will easily check (it is really the definition) that $\alpha : D \longrightarrow \mathbb{C}$ corresponding to a sequence, extends to a *continuous* function on $\overline{D}$ vanishing at 0 if and only if $\lim_{j\to\infty} a_j = 0$, which is to say, $\{a_j\} \in c_0$. Thus it follows, with a little thought which you should give it, that c_0 is a Banach space with the norm

$$\|a\|_\infty = \sup_j \|a_j\|. \tag{1.28}$$

What is an example of a non-complete normed space, a normed space which is *not* a Banach space? These are legion of course. The simplest way to get one is to 'put the wrong norm' on a space, one which does not correspond to the definition. Consider for instance the linear space $\mathcal{T}$ of sequences $\mathbb{N} \longrightarrow \mathbb{C}$ which 'terminate', i.e. each element $\{a_j\} \in \mathcal{T}$ has $a_j = 0$ for $j > J$, where of course the J may depend on the particular sequence. Then $\mathcal{T} \subset c_0$, the norm on c_0 defines a norm on $\mathcal{T}$ but it cannot be complete, since the closure of $\mathcal{T}$ is easily seen to be all of c_0 – so there are Cauchy sequences in $\mathcal{T}$ without limit in $\mathcal{T}$. Make sure you are not lost here – you need to get used to the fact that we often need to discuss the 'convergence of sequences of convergent sequences' as here.

One result we will exploit below, and I give it now just as preparation, concerns *absolutely summable series.* Recall that a series is just a sequence where we 'think' about adding the terms. Thus if v_n is a sequence in some vector space V then there is the corresponding sequence of partial sums $w_N = \sum\limits_{i=1}^{N} v_i$. I will say that $\{v_n\}$ is a series if I am thinking about summing it.

Definition 1.5. A series $\{v_n\}$ with partial sums $\{w_N\}$ is said to be *absolutely summable* if

$$\sum_n \|v_n\|_V < \infty, \text{ i.e. } \sum_{N>1} \|w_N - w_{N-1}\|_V < \infty. \tag{1.29}$$

Proposition 1.2. *The sequence of partial sums of any absolutely summable series in a normed space is Cauchy and a normed space is complete if and only if every absolutely summable series in it converges, meaning that the sequence of partial sums converges.*

Proof. The sequence of partial sums is

$$w_n = \sum_{j=1}^{n} v_j. \tag{1.30}$$

Thus, if $m > n$ then

$$w_m - w_n = \sum_{j=n+1}^{m} v_j. \tag{1.31}$$

It follows from the triangle inequality that

$$\|w_n - w_m\|_V \leq \sum_{j=n+1}^{m} \|v_j\|_V. \tag{1.32}$$

So if the series is absolutely summable then

$$\sum_{j=1}^{\infty} \|v_j\|_V < \infty \text{ and } \lim_{n\to\infty} \sum_{j=n+1}^{\infty} \|v_j\|_V = 0.$$

Thus $\{w_n\}$ is Cauchy if $\{v_j\}$ is absolutely summable. Hence if V is complete then every absolutely summable series is summable, i.e. the sequence of partial sums converges.

Conversely, suppose that every absolutely summable series converges in this sense. Then we need to show that every Cauchy sequence in V converges. Let u_n be a Cauchy sequence. It suffices to show that this has a subsequence which converges, since a Cauchy sequence with a convergent subsequence is convergent. To do so we just proceed inductively. Using the Cauchy condition we can for every k find an integer N_k such that

$$n, m > N_k \Longrightarrow \|u_n - u_m\| < 2^{-k}. \tag{1.33}$$

Now choose an increasing sequence n_k where $n_k > N_k$ and $n_k > n_{k-1}$ to make it increasing. It follows that

$$\|u_{n_k} - u_{n_{k-1}}\| \leq 2^{-k+1}. \tag{1.34}$$

Denoting this subsequence as $u'_k = u_{n_k}$ it follows from (1.34) and the triangle inequality that

$$\sum_{n=1}^{\infty} \|u'_n - u'_{n-1}\| \leq 4 \tag{1.35}$$

so the sequence $v_1 = u'_1$, $v_k = u'_k - u'_{k-1}$, $k > 1$, is absolutely summable. Its sequence of partial sums is $w_j = u'_j$ so the assumption is that this converges, hence the original Cauchy sequence converges and V is complete. □

Notice the idea here, of 'speeding up the convergence' of the Cauchy sequence by dropping a lot of terms. We will use this idea of absolutely summable series heavily in the discussion of Lebesgue integration.

4. Operators and functionals

The vector spaces we are most interested in are, as already remarked, spaces of functions (or something a little more general). The elements of these are the objects of primary interest but we are especially interested in the way they are related by linear maps. The sorts of maps we have in mind here are differential and integral operators. For example the indefinite Riemann integral of a continuous function $f : [0, 1] \longrightarrow \mathbb{C}$ is also a continuous function of the upper limit:

$$I(f)(x) = \int_0^x f(s)ds. \tag{1.36}$$

So, $I : \mathcal{C}([0, 1]) \longrightarrow \mathcal{C}([0, 1])$ it is an 'operator' which turns one continuous function into another. You might want to bear such an example in mind as you go through this section.

A map between two vector spaces (over the same field, for us either $\mathbb{R}$ or $\mathbb{C}$) is linear if it takes linear combinations to linear combinations:-

$$T : V \longrightarrow W,\ T(a_1v_1+a_2v_2) = a_1T(v_1)+a_2T(v_2),\ \forall\ v_1,\ v_2 \in V,\ a_1, a_2 \in \mathbb{K}. \tag{1.37}$$

In the finite-dimensional case linearity is enough to allow maps to be studied. However in the case of infinite-dimensional normed spaces we will require continuity, which is automatic in finite dimensions. It makes perfectly good sense to say, demand or conclude, that a map as in (1.37) is continuous if V and W are normed spaces since they are then metric spaces. Recall that for metric spaces there are several different equivalent conditions that ensure a map, $T : V \longrightarrow W$, is continuous:

$$v_n \to v \text{ in } V \Longrightarrow Tv_n \to Tv \text{ in } W \tag{1.38}$$

$$O \subset W \text{ open } \Longrightarrow T^{-1}(O) \subset V \text{ open} \tag{1.39}$$

$$C \subset W \text{ closed } \Longrightarrow T^{-1}(C) \subset V \text{ closed.} \tag{1.40}$$

For a linear map between normed spaces there is a direct characterization of continuity in terms of the norm.

Proposition 1.3. *A linear map* (1.37) *between normed spaces is continuous if and only if it is* bounded *in the sense that there exists a constant C such that*

$$\|Tv\|_W \le C\|v\|_V\ \forall\ v \in V. \tag{1.41}$$

Of course bounded for a function on a metric space already has a meaning and this is not it! The usual sense would be $\|Tv\| \le C$ but this would imply $\|T(av)\| = |a|\|Tv\| \le C$ so $Tv = 0$. Hence it is not so dangerous to use the term 'bounded' for (1.41) – it is really 'relatively bounded', i.e. takes bounded sets into bounded sets. From now on, bounded for a linear map means (1.41).

Proof. If (1.41) holds then if $v_n \to v$ in V it follows that $\|Tv - Tv_n\| = \|T(v - v_n)\| \le C\|v - v_n\| \to 0$ as $n \to \infty$ so $Tv_n \to Tv$ and continuity follows.

For the reverse implication we use the second characterization of continuity above. Denote the ball around $v \in V$ of radius $\epsilon > 0$ by

$$B_V(v, \epsilon) = \{w \in V; \|v - w\| < \epsilon\}.$$

Thus if T is continuous then the inverse image of the the unit ball around the origin, $T^{-1}(B_W(0,1)) = \{v \in V; \|Tv\|_W < 1\}$, contains the origin in V and so, being open, must contain some $B_V(0, \epsilon)$. This means that

$$T(B_V(0, \epsilon)) \subset B_W(0, 1) \text{ so } \|v\|_V < \epsilon \Longrightarrow \|Tv\|_W \le 1. \tag{1.42}$$

Now proceed by scaling. If $0 \ne v \in V$ then $\|v'\| < \epsilon$ where $v' = \epsilon v/2\|v\|$. So (1.42) shows that $\|Tv'\| \le 1$ but this implies (1.41) with $C = 2/\epsilon$ – it is trivially true if $v = 0$. □

Note that a bounded linear map is in fact *uniformly continuous* – given $\delta > 0$ there exists $\epsilon > 0$ such that

$$\|v - w\|_V = d_V(v, w) < \epsilon \Longrightarrow \|Tv - Tw\|_W = d_W(Tv, TW) < \delta \tag{1.43}$$

namely $\epsilon = \delta/C$. One consequence of this is that a linear map $T : U \longrightarrow W$ into a Banach space, defined and continuous on a linear subspace, $U \subset V$. (with respect to the restriction of the norm from V to U) extends uniquely to a continuous map $T : \overline{U} \longrightarrow W$ on the closure of U.

As a general rule we drop the distinguishing subscript for norms, since which norm we are using can be determined by what it is being applied to.

So, if $T : V \longrightarrow W$ is continous and linear between normed spaces, or from now on 'bounded', then

$$\|T\| = \sup_{\|v\|=1} \|Tv\| < \infty. \tag{1.44}$$

Lemma 1.3. *The bounded linear maps between normed spaces V and W form a linear space $\mathcal{B}(V, W)$ on which $\|T\|$ defined by* (1.44) *or equivalently*

$$\|T\| = \inf\{C;\ (1.41)\ \textit{holds}\} \tag{1.45}$$

is a norm.

Proof. First check that (1.44) is equivalent to (1.45). Define $\|T\|$ by (1.44). Then for any $v \in V$, $v \neq 0$,

$$\|T\| \ge \|T(\frac{v}{\|v\|})\| = \frac{\|Tv\|}{\|v\|} \Longrightarrow \|Tv\| \le \|T\|\|v\| \tag{1.46}$$

since as always this is trivially true for $v = 0$. Thus $C = \|T\|$ is a constant for which (1.41) holds.

Conversely, from the definition of $\|T\|$, if $\epsilon > 0$ then there exists $v \in V$ with $\|v\| = 1$ such that $\|T\| - \epsilon < \|Tv\| \le C$ for any C for which (1.41) holds. Since $\epsilon > 0$ is arbitrary, $\|T\| \le C$ and hence $\|T\|$ is given by (1.45).

From the definition of $\|T\|$, $\|T\| = 0$ implies $Tv = 0$ for all $v \in V$ and for $\lambda \neq 0$,

$$\|\lambda T\| = \sup_{\|v\|=1} \|\lambda Tv\| = |\lambda|\|T\| \tag{1.47}$$

and this is also obvious for $\lambda = 0$. This only leaves the triangle inequality to check and for any T, $S \in \mathcal{B}(V, W)$, and $v \in V$ with $\|v\| = 1$

$$\|(T + S)v\|_W = \|Tv + Sv\|_W \le \|Tv\|_W + \|Sv\|_W \le \|T\| + \|S\| \tag{1.48}$$

so taking the supremum, $\|T + S\| \le \|T\| + \|S\|$. □

Thus we see the very satisfying fact that the space of bounded linear maps between two normed spaces is itself a normed space, with the norm being the best constant in the estimate (1.41). Make sure you absorb this! Such bounded linear maps between normed spaces are often called 'operators' because we are thinking of the normed spaces as being like function spaces.

You might like to check boundedness for the example, I, of a linear operator in (1.36), namely that in terms of the supremum norm on $\mathcal{C}([0, 1])$, $\|T\| \le 1$.

One particularly important case is when $W = \mathbb{K}$ is the field, for us usually $\mathbb{C}$. Then a simpler notation is handy and one sets $V' = \mathcal{B}(V, \mathbb{C})$ – this is called the *dual space* of V (also sometimes denoted V^*).

Proposition 1.4. *If W is a Banach space then $\mathcal{B}(V, W)$, with the norm* (1.44), *is a Banach space.*

Proof. We simply need to show that if W is a Banach space then every Cauchy sequence in $\mathcal{B}(V, W)$ is convergent. The first thing to do is to find the limit. To say that $T_n \in \mathcal{B}(V, W)$ is Cauchy, is just to say that given $\epsilon > 0$ there exists N such that n, $m > N$ implies $\|T_n - T_m\| < \epsilon$. By the definition of the norm, if $v \in V$

then $\|T_nv - T_mv\|_W \le \|T_n - T_m\|\|v\|_V$ so T_nv is Cauchy in W for each $v \in V$. By assumption, W is complete, so

$$T_nv \longrightarrow w \text{ in } W. \tag{1.49}$$

However, the limit can only depend on v so we can define a map $T : V \longrightarrow W$ by $Tv = w = \lim_{n\to\infty} T_nv$ as in (1.49).

This map defined from the limits is linear, since $T_n(\lambda v) = \lambda T_nv \longrightarrow \lambda Tv$ and $T_n(v_1 + v_2) = T_nv_1 + T_nv_2 \longrightarrow Tv_2 + Tv_2 = T(v_1 + v_2)$. Moreover, $|\|T_n\| - \|T_m\|| \le \|T_n - T_m\|$ so $\|T_n\|$ is Cauchy in $[0, \infty)$ and hence converges, with limit S, and

$$\|Tv\| = \lim_{n\to\infty} \|T_nv\| \le S\|v\| \tag{1.50}$$

so $\|T\| \le S$ shows that T is bounded.

Returning to the Cauchy condition above and passing to the limit in $\|T_nv - T_mv\| \le \epsilon\|v\|$ as $m \to \infty$ shows that $\|T_n - T\| \le \epsilon$ if $n > M$ and hence $T_n \to T$ in $\mathcal{B}(V, W)$ which is therefore complete. □

Note that this proof is structurally the same as that of Lemma 1.2.

One simple consequence of this is:-

Corollary 1.1. *The dual space of a normed space is always a Banach space.*

However you should be a little suspicious here since we have not shown that the dual space V' is non-trivial, meaning we have not eliminated the possibility that $V' = \{0\}$ even when $V \neq \{0\}$. The Hahn-Banach Theorem, discussed below, takes care of this.

One game you can play is 'what is the dual of that space'. Of course the dual is the dual, but you may well be able to identify the dual space of V with some other Banach space by finding a linear bijection between V' and the other space, W, which identifies the norms as well. We will play this game a bit later.

5. Subspaces and quotients

The notion of a linear subspace of a vector space is natural enough, and you are likely quite familiar with it. Namely $W \subset V$ where V is a vector space is a (linear) subspace if any linear combinations $\lambda_1w_1 + \lambda_2w_2 \in W$ if λ_1, $\lambda_2 \in \mathbb{K}$ and w_1, $w_2 \in W$. Thus W 'inherits' its linear structure from V. Since we also have a topology from the metric we will be especially interested in closed subspaces. Check that you understand the (elementary) proof of

Lemma 1.4. *A subspace of a Banach space is a Banach space in terms of the restriction of the norm if and only if it is closed.*

There is a second very important way to construct new linear spaces from old. Namely we want to make a linear space out of 'the rest' of V, given that W is a linear subspace. In finite dimensions one way to do this is to give V an inner product and then take the subspace orthogonal to W. One problem with this is that the result depends, although not in an essential way, on the inner product. Instead we adopt the usual 'myopia' approach and take an equivalence relation on V which identifies points which differ by an element of W. The equivalence classes are then 'planes parallel to W'. I am going through this construction quickly here under the assumption that it is familiar to most of you, if not you should think about it carefully since we need to do it several times later.

So, if $W \subset V$ is a linear subspace of V we define a relation on V – remember this is just a subset of $V \times V$ with certain properties – by

$$v \sim_W v' \Longleftrightarrow v - v' \in W \Longleftrightarrow \ \exists\, w \in W \text{ s.t. } v = v' + w. \tag{1.51}$$

This satisfies the three conditions for an equivalence relation:

(1) $v \sim_W v$
(2) $v \sim_W v' \Longleftrightarrow v' \sim_W v$
(3) $v \sim_W v',\ v' \sim_W v'' \Longrightarrow v \sim_W v''$

which means that we can regard it as a 'coarser notion of equality.'

Then V/W is the set of equivalence classes with respect to $\sim_W$. You can think of the elements of V/W as being of the form $v + W$ – a particular element of V plus an arbitrary element of W. Then of course $v' \in v+W$ if and only if $v' - v \in W$ meaning $v \sim_W v'$.

The crucial point here is that

$$V/W \text{ is a vector space.} \tag{1.52}$$

You should check the details – see Problem 1.5. Note that the 'is' in (1.52) should really be expanded to 'is in a natural way' since as usual the linear structure is inherited from V :

$$\lambda(v + W) = \lambda v + W,\ (v_1 + W) + (v_2 + W) = (v_1 + v_2) + W. \tag{1.53}$$

The subspace W appears as the origin in V/W.

Now, two cases of this are of special interest to us.

Proposition 1.5. *If $\|\cdot\|$ is a seminorm on V then*

$$E = \{v \in V; \|v\| = 0\} \subset V \tag{1.54}$$

is a linear subspace and

$$\|v + E\|_{V/E} = \|v\| \tag{1.55}$$

defines a norm on V/E.

Proof. That E is linear follows from the properties of a seminorm, since $\|\lambda v\| = |\lambda|\|v\|$ shows that $\lambda v \in E$ if $v \in E$ and $\lambda \in \mathbb{K}$. Similarly the triangle inequality shows that $v_1 + v_2 \in E$ if $v_1,\, v_2 \in E$.

To check that (1.55) defines a norm, first we need to check that it makes sense as a function $\|\cdot\|_{V/E} \longrightarrow [0,\infty)$. This amounts to the statement that $\|v'\|$ is the same for all elements $v' = v + e \in v + E$ for a fixed v. This however follows from the triangle inequality applied twice:

$$\|v'\| \le \|v\| + \|e\| = \|v\| \le \|v'\| + \|-e\| = \|v'\|. \tag{1.56}$$

Now, I leave you the exercise of checking that $\|\cdot\|_{V/E}$ is a norm, see Problem **??**. □

The second application is more serious, but in fact we will not use it for some time so I usually do not do this in lectures at this stage.

Proposition 1.6. *If $W \subset V$ is a closed subspace of a normed space then*

$$\|v + W\|_{V/W} = \inf_{w \in W} \|v + w\|_V \tag{1.57}$$

defines a norm on V/W; if V is a Banach space then so is V/W.

For the proof see Problems **??** and **??**.

6. Completion

A normed space not being complete, not being a Banach space, is considered to be a defect which we might, indeed will, wish to rectify.

Let V be a normed space with norm $\|\cdot\|_V$. A *completion* of V is a Banach space B with the following properties:-

(1) There is an injective (i.e. 1-1) linear map $I : V \longrightarrow B$
(2) The norms satisfy

$$\|I(v)\|_B = \|v\|_V \ \forall\ v \in V. \tag{1.58}$$

(3) The range $I(V) \subset B$ is dense in B.

Notice that if V is itself a Banach space then we can take $B = V$ with I the identity map.

So, the main result is:

THEOREM 1.1. *Each normed space has a completion.*

There are several ways to prove this, we will come across a more sophisticated one (using the Hahn-Banach Theorem) later. In the meantime I will describe two proofs. In the first the fact that any metric space has a completion in a similar sense is recalled and then it is shown that the linear structure extends to the completion. A second, 'hands-on', proof is also outlined with the idea of motivating the construction of the Lebesgue integral – which is in our near future.

PROOF 1. One of the neater proofs that any metric space has a completion is to use Lemma 1.2. Pick a point in the metric space of interest, $p \in M$, and then define a map

$$M \ni q \longmapsto f_q \in \mathcal{C}_\infty(M),\ f_q(x) = d(x,q) - d(x,p)\ \forall\ x \in M. \tag{1.59}$$

That $f_q \in \mathcal{C}_\infty(M)$ is straightforward to check. It is bounded (because of the second term) by the reverse triangle inequality

$$|f_q(x)| = |d(x,q) - d(x,p)| \le d(p,q)$$

and is continuous, as the difference of two continuous functions. Moreover the distance between two functions in the image is

$$\sup_{x\in M} |f_q(x) - f_{q'}(x)| = \sup_{x\in M} |d(x,q) - d(x,q')| = d(q,q') \tag{1.60}$$

using the reverse triangle inequality (and evaluating at $x = q$). Thus the map (1.59) is well-defined, injective and even distance-preserving. Since $\mathcal{C}_\infty(M)$ is complete, the closure of the image of (1.59) is a complete metric space, X, in which M can be identified as a dense subset.

Now, in case that $M = V$ is a normed space this all goes through. The disconcerting thing is that the map $q \longrightarrow f_q$ is *not* linear. Nevertheless, we can give X a linear structure so that it becomes a Banach space in which V is a dense linear subspace. Namely for any two elements $f_i \in X$, $i = 1, 2$, define

$$\lambda_1 f_1 + \lambda_2 f_2 = \lim_{n\to\infty} f_{\lambda_1 p_n + \lambda_2 q_n} \tag{1.61}$$

where p_n and q_n are sequences in V such that $f_{p_n} \to f_1$ and $f_{q_n} \to f_2$. Such sequences exist by the construction of X and the result does not depend on the choice of sequence – since if p'_n is another choice in place of p_n then $f_{p'_n} - f_{p_n} \to 0$ in X (and similarly for q_n). So the element of the left in (1.61) is well-defined. All of the properties of a linear space and normed space now follow by continuity from $V \subset X$ and it also follows that X is a Banach space (since a closed subset of a complete space is complete). Unfortunately there are quite a few annoying details to check! □

'Proof 2' (the last bit is left to you). Let V be a normed space. First we introduce the rather large space

$$\widetilde{V} = \left\{ \{u_k\}_{k=1}^{\infty}; u_k \in V \text{ and } \sum_{k=1}^{\infty} \|u_k\| < \infty \right\} \tag{1.62}$$

the elements of which, if you recall, are said to be absolutely summable. Notice that the elements of $\widetilde{V}$ are *sequences*, valued in V so two sequences are equal, are the same, only when each entry in one is equal to the corresponding entry in the other – no shifting around or anything is permitted as far as equality is concerned. We think of these as series (remember this means nothing except changing the name, a series is a sequence and a sequence is a series), the only difference is that we 'think' of taking the limit of a sequence but we 'think' of summing the elements of a series, whether we can do so or not being a different matter.

Now, each element of $\widetilde{V}$ is a Cauchy series – meaning the corresponding sequence of partial sums $v_N = \sum\limits_{k=1}^{N} u_k$ is Cauchy if $\{u_k\}$ is absolutely summable. As noted earlier, this is simply because if $M \geq N$ then

$$\|v_M - v_N\| = \|\sum_{j=N+1}^{M} u_j\| \leq \sum_{j=N+1}^{M} \|u_j\| \leq \sum_{j \geq N+1} \|u_j\| \tag{1.63}$$

gets small with N by the assumption that $\sum\limits_{j} \|u_j\| < \infty$.

Moreover, $\widetilde{V}$ is a linear space, where we add sequences, and multiply by constants, by doing the operations on each component:-

$$t_1\{u_k\} + t_2\{u'_k\} = \{t_1 u_k + t_2 u'_k\}. \tag{1.64}$$

This always gives an absolutely summable series by the triangle inequality:

$$\sum_k \|t_1 u_k + t_2 u'_k\| \leq |t_1| \sum_k \|u_k\| + |t_2| \sum_k \|u'_k\|. \tag{1.65}$$

Within $\widetilde{V}$ consider the linear subspace

$$S = \left\{ \{u_k\}; \sum_k \|u_k\| < \infty, \ \sum_k u_k = 0 \right\} \tag{1.66}$$

of those which sum to 0. As discussed in Section 5 above, we can form the quotient

$$B = \widetilde{V}/S \tag{1.67}$$

the elements of which are the 'cosets' of the form $\{u_k\} + S \subset \widetilde{V}$ where $\{u_k\} \in \widetilde{V}$. This is our completion, we proceed to check the following properties of this B.

(1) A norm on B (via a seminorm on $\tilde{V}$) is defined by

$$\|b\|_B = \lim_{n \to \infty} \|\sum_{k=1}^{n} u_k\|, \ b = \{u_k\} + S \in B. \tag{1.68}$$

(2) The original space V is imbedded in B by

$$V \ni v \longmapsto I(v) = \{u_k\} + S,\ u_1 = v,\ u_k = 0\ \forall\ k > 1 \tag{1.69}$$

and the norm satisfies (1.58).

(3) $I(V) \subset B$ is dense.

(4) B is a Banach space with the norm (1.68).

So, first that (1.68) is a norm. The limit on the right does exist since the limit of the norm of a Cauchy sequence always exists – namely the sequence of norms is itself Cauchy but now in $\mathbb{R}$. Moreover, adding an element of S to $\{u_k\}$ does not change the norm of the sequence of partial sums, since the additional term tends to zero in norm. Thus $\|b\|_B$ is well-defined for each element $b \in B$ and $\|b\|_B = 0$ means exactly that the sequence $\{u_k\}$ used to define it tends to 0 in norm, hence is in S hence $b = 0$ in B. The other two properties of norm are reasonably clear, since if $b,\ b' \in B$ are represented by $\{u_k\}$, $\{u'_k\}$ in $\widetilde{V}$ then tb and $b + b'$ are represented by $\{tu_k\}$ and $\{u_k + u'_k\}$ and

(1.70)

$$\lim_{n\to\infty}\|\sum_{k=1}^{n} tu_k\| = |t|\lim_{n\to\infty}\|\sum_{k=1}^{n} u_k\|, \Longrightarrow \|tb\| = |t|\|b\|$$

$$\lim_{n\to\infty}\|\sum_{k=1}^{n}(u_k + u'_k)\| = A \Longrightarrow$$

$$\text{for } \epsilon > 0\ \exists\ N \text{ s.t. }\ \forall\ n \geq N,\ A - \epsilon \leq \|\sum_{k=1}^{n}(u_k + u'_k)\| \Longrightarrow$$

$$A - \epsilon \leq \|\sum_{k=1}^{n} u_k\| + \|\sum_{k=1}^{n} u'_k)\|\ \forall\ n \geq N \Longrightarrow A - \epsilon \leq \|b\|_B + \|b'\|_B\ \forall\ \epsilon > 0 \Longrightarrow$$

$$\|b + b'\|_B \leq \|b\|_B + \|b'\|_B.$$

Now the norm of the element $I(v) = v, 0, 0, \cdots$, is the limit of the norms of the sequence of partial sums and hence is $\|v\|_V$ so $\|I(v)\|_B = \|v\|_V$ and $I(v) = 0$ therefore implies $v = 0$ and hence I is also injective.

We need to check that B is complete, and also that $I(V)$ is dense. Here is an extended discussion of the difficulty – of course maybe you can see it directly yourself (or have a better scheme). Note that I suggest that you to write out your own version of it carefully in Problem **??**.

Okay, what does it mean for B to be a Banach space, as discussed above it means that every absolutely summable series in B is convergent. Such a series $\{b_n\}$ is given by $b_n = \{u_k^{(n)}\} + S$ where $\{u_k^{(n)}\} \in \widetilde{V}$ and the summability condition is that

$$\infty > \sum_n \|b_n\|_B = \sum_n \lim_{N\to\infty}\|\sum_{k=1}^{N} u_k^{(n)}\|_V. \tag{1.71}$$

So, we want to show that $\sum\limits_n b_n = b$ converges, and to do so we need to find the limit b. It is supposed to be given by an absolutely summable series. The 'problem' is that this series should look like $\sum\limits_n \sum\limits_k u_k^{(n)}$ in some sense – because it is supposed to represent the sum of the b_n's. Now, it would be very nice if we had the estimate

$$\sum_n \sum_k \|u_k^{(n)}\|_V < \infty \tag{1.72}$$

since this should allow us to break up the double sum in some nice way so as to get an absolutely summable series out of the whole thing. The trouble is that (1.72) need not hold. We know that *each* of the sums over k – for given n – converges, but not the sum of the sums. All we know here is that the sum of the 'limits of the norms' in (1.71) converges.

So, that is the problem! One way to see the solution is to note that we do not have to choose the original $\{u_k^{(n)}\}$ to 'represent' b_n – we can add to it any element of S. One idea is to rearrange the $u_k^{(n)}$ – I am thinking here of fixed n – so that it 'converges even faster.' I will not go through this in full detail but rather do it later when we need the argument for the completeness of the space of Lebesgue integrable functions. Given $\epsilon > 0$ we can choose p_1 so that for all $p \ge p_1$,

$$|\|\sum_{k\le p} u_k^{(n)}\|_V - \|b_n\|_B| \le \epsilon,\ \sum_{k\ge p} \|u_k^{(n)}\|_V \le \epsilon. \tag{1.73}$$

Then in fact we can choose successive $p_j > p_{j-1}$ (remember that little n is fixed here) so that

$$|\|\sum_{k\le p_j} u_k^{(n)}\|_V - \|b_n\|_B| \le 2^{-j}\epsilon,\ \sum_{k\ge p_j} \|u_k^{(n)}\|_V \le 2^{-j}\epsilon\ \forall\ j. \tag{1.74}$$

Now, 'resum the series' defining instead $v_1^{(n)} = \sum\limits_{k=1}^{p_1} u_k^{(n)}$, $v_j^{(n)} = \sum\limits_{k=p_{j-1}+1}^{p_j} u_k^{(n)}$ and do this setting $\epsilon = 2^{-n}$ for the nth series. Check that now

$$\sum_n \sum_k \|v_k^{(n)}\|_V < \infty. \tag{1.75}$$

Of course, you should also check that $b_n = \{v_k^{(n)}\} + S$ so that these new summable series work just as well as the old ones.

After this fiddling you can now try to find a limit for the sequence as

$$b = \{w_k\} + S,\ w_k = \sum_{l+p=k+1} v_l^{(p)} \in V. \tag{1.76}$$

So, you need to check that this $\{w_k\}$ is absolutely summable in V and that $b_n \to b$ as $n \to \infty$.

Finally then there is the question of showing that $I(V)$ is dense in B. You can do this using the same idea as above – in fact it might be better to do it first. Given an element $b \in B$ we need to find elements in V, v_k such that $\|I(v_k) - b\|_B \to 0$ as $k \to \infty$. Take an absolutely summable series u_k representing b and take $v_j = \sum\limits_{k=1}^{N_j} u_k$ where the p_j's are constructed as above and check that $I(v_j) \to b$ by computing

$$\|I(v_j) - b\|_B = \lim_{\to\infty} \|\sum_{k>p_j} u_k\|_V \le \sum_{k>p_j} \|u_k\|_V. \tag{1.77}$$

□

7. More examples

Let me collect some examples of normed and Banach spaces. Those mentioned above and in the problems include:

- c_0 the space of convergent sequences in $\mathbb{C}$ with supremum norm, a Banach space.
- l^p one space for each real number $1 \le p < \infty$; the space of p-summable series with corresponding norm; all Banach spaces. The most important of these for us is the case $p = 2$, which is (a) Hilbert space.
- l^∞ the space of bounded sequences with supremum norm, a Banach space with $c_0 \subset l^\infty$ as a closed subspace with the same norm.
- $\mathcal{C}([a, b])$ or more generally $\mathcal{C}(M)$ for any compact metric space M – the Banach space of continuous functions with supremum norm.
- $\mathcal{C}_\infty(\mathbb{R})$, or more generally $\mathcal{C}_\infty(M)$ for any metric space M – the Banach space of bounded continuous functions with supremum norm.
- $\mathcal{C}_0(\mathbb{R})$, or more generally $\mathcal{C}_0(M)$ for any metric space M – the Banach space of continuous functions which 'vanish at infinity' (see Problem **??**) with supremum norm. A closed subspace, with the same norm, in $\mathcal{C}_\infty(M)$.
- $\mathcal{C}^k([a, b])$ the space of k times continuously differentiable (so $k \in \mathbb{N}$) functions on $[a, b]$ with norm the sum of the supremum norms on the function and its derivatives. Each is a Banach space – see Problem **??**.
- The space $\mathcal{C}([0, 1])$ with norm

$$\|u\|_{L^1} = \int_0^1 |u| dx \tag{1.78}$$

  given by the Riemann integral of the absolute value. A normed space, but not a Banach space. We will construct the concrete completion, $L^1([0, 1])$ of Lebesgue integrable 'functions'.
- The space $\mathcal{R}([a, b])$ of Riemann integrable functions on $[a, b]$ with $\|u\|$ defined by (1.78). This is only a seminorm, since there are Riemann integrable functions (note that u Riemann integrable does imply that $|u|$ is Riemann integrable) with $|u|$ having vanishing Riemann integral but which are not identically zero. This cannot happen for continuous functions. So the quotient is a normed space, but it is not complete.
- The same spaces – either of continuous or of Riemann integrable functions but with the (semi- in the second case) norm

$$\|u\|_{L^p} = \left(\int_a^b |u|^p \right)^{\frac{1}{p}}. \tag{1.79}$$

  Not complete in either case even after passing to the quotient to get a norm for Riemann integrable functions. We can, and indeed will, define $L^p(a, b)$ as the completion of $\mathcal{C}([a, b])$ with respect to the L^p norm. However we will get a concrete realization of it soon.
- Suppose $0 < \alpha < 1$ and consider the subspace of $\mathcal{C}([a, b])$ consisting of the 'Hölder continuous functions' with exponent α, that is those $u : [a, b] \longrightarrow \mathbb{C}$ which satisfy

$$|u(x) - u(y)| \le C|x - y|^\alpha \text{ for some } C \ge 0. \tag{1.80}$$

Note that this already implies the continuity of u. As norm one can take the sum of the supremum norm and the 'best constant' which is the same as

$$\|u\|_{\mathcal{C}^\alpha} = \sup_{x\in[a,b]|} |u(x)| + \sup_{x\neq y\in[a,b]} \frac{|u(x)-u(y)|}{|x-y|^\alpha}; \tag{1.81}$$

it is a Banach space usually denoted $\mathcal{C}^\alpha([a,b])$.

- Note the previous example works for $\alpha = 1$ as well, then it is not denoted $\mathcal{C}^1([a,b])$, since that is the space of once continuously differentiable functions; this is the space of Lipschitz functions $\Lambda([a,b])$ – again it is a Banach space.
- We will also talk about Sobolev spaces later. These are functions with 'Lebesgue integrable derivatives'. It is perhaps not easy to see how to define these, but if one takes the norm on $\mathcal{C}^1([a,b])$

$$\|u\|_{H^1} = \left(\|u\|^2_{L^2} + \|\frac{du}{dx}\|^2_{L^2}\right)^{\frac{1}{2}} \tag{1.82}$$

and completes it, one gets the Sobolev space $H^1([a,b])$ – it is a Banach space (and a Hilbert space). In fact it is a subspace of $\mathcal{C}([a,b]) = \mathcal{C}([a,b])$.

Here is an example to see that the space of continuous functions on $[0,1]$ with norm (1.78) is not complete; things are even worse than this example indicates! It is a bit harder to show that the quotient of the Riemann integrable functions is not complete, feel free to give it a try.

Take a simple non-negative continuous function on $\mathbb{R}$ for instance

$$f(x) = \begin{cases} 1-|x| & \text{if } |x| \le 1 \\ 0 & \text{if } |x| > 1. \end{cases} \tag{1.83}$$

Then $\int_{-1}^1 f(x) = 1$. Now scale it up and in by setting

$$f_N(x) = Nf(N^3x) = 0 \text{ if } |x| > N^{-3}. \tag{1.84}$$

So it vanishes outside $[-N^{-3}, N^{-3}]$ and has $\int_{-1}^1 f_N(x)dx = N^{-2}$. It follows that the sequence $\{f_N\}$ is absolutely summable with respect to the integral norm in (1.78) on $[-1,1]$. The pointwise series $\sum\limits_N f_N(x)$ converges everywhere except at $x = 0$ – since at each point $x \neq 0$, $f_N(x) = 0$ if $N^3|x| > 1$. The resulting function, even if we ignore the problem at $x = 0$, is not Riemann integrable because it is not bounded.

You might respond that the sum of the series is 'improperly Riemann integrable'. This is true but does not help much.

It is at this point that I start doing Lebesgue integration in the lectures. The following material is from later in the course but fits here quite reasonably.

8. Baire's theorem

At least once I wrote a version of the following material on the blackboard during the first mid-term test, in an an attempt to distract people. It did not work very well – its seems that MIT students have already been toughened up by this stage. Baire's theorem will be used later (it is also known as 'Baire category theory' although it has nothing to do with categories in the modern sense).

This is a theorem about complete metric spaces – it could be included in the earlier course 'Real Analysis' but the main applications are in Functional Analysis.

THEOREM 1.2 (Baire). *If M is a non-empty complete metric space and $C_n \subset M$, $n \in \mathbb{N}$, are closed subsets such that*

$$M = \bigcup_n C_n \tag{1.85}$$

then at least one of the C_n's has an interior point, i.e. contains a non-empty ball in M.

PROOF. We will assume that each of the C_n's has empty interior, hoping to arrive at a contradiction to (1.85) using the other properties. Thus if $p \in M$ and $\epsilon > 0$ the open ball $B(p, \epsilon)$ is not contained in any one of the C_n.

We start by choosing $p_1 \in M \setminus C_1$ which must exist since M is not empty and otherwise $C_1 = M$. Now, there must exist $\epsilon_1 > 0$ such that $B(p_1, \epsilon_1) \cap C_1 = \emptyset$, since C_1 is closed. No open ball around p_1 can be contained in C_2 so there exists $p_2 \in B(p_1, \epsilon_1/3)$ which is not in C_2. Again since C_2 is closed there exists $\epsilon_2 > 0$, $\epsilon_2 < \epsilon_1/3$ such that $B(p_2, \epsilon_2) \cap C_2 = \emptyset$.

Proceeding inductively we suppose there is are k points p_i, $i = 1, \dots, k$ and positive numbers

$$0 < \epsilon_k < \epsilon_{k-1}/3 < \epsilon_{k-2}/3^2 < \cdots < \epsilon_1/3^{k-1} \tag{1.86}$$

such that

$$p_j \in B(p_{j-1}, \epsilon_{j-1}/3),\ B(p_j, \epsilon_j) \cap C_j = \emptyset. \tag{1.87}$$

Then we can add another p_{k+1} by using the properties of C_{k+1} – it has empty interior so there is some point in $B(p_k, \epsilon_k/3)$ which is not in C_{k+1} and then $B(p_{k+1}, \epsilon_{k+1}) \cap C_{k+1} = \emptyset$ where $\epsilon_{k+1} > 0$ but $\epsilon_{k+1} < \epsilon_k/3$. Thus, we have a sequence $\{p_k\}$ in M satisfying (1.86) and (1.87) for all k.

Since $d(p_{k+1}, p_k) < \epsilon_k/3$ this is a Cauchy sequence, in fact

$$d(p_k, p_{k+l}) < \epsilon_k/3 + \cdots + \epsilon_{k+l-1}/3 < 2\epsilon_k. \tag{1.88}$$

Since M is assumed to be complete this sequence converges to a limit, $q \in M$. Notice however that $p_l \in B(p_k, 2\epsilon_k/3)$ for all $k > l$ so $d(p_k, q) \le 2\epsilon_k/3$ which implies that $q \notin C_k$ for any k. This is the desired contradiction to (1.85).

Thus, at least one of the C_n must have non-empty interior. □

In applications one might get a complete metric space written as a countable union of subsets

$$M = \bigcup_n E_n,\ E_n \subset M \tag{1.89}$$

where the E_n are not necessarily closed. We can still apply Baire's theorem however, just take $C_n = \overline{E_n}$ to be the closures – then of course (1.85) holds since $E_n \subset C_n$. The conclusion from (1.89) for a complete M is

(1.90) For at least one n the closure of E_n has non-empty interior.

9. Uniform boundedness

One application of Baire's theorem is often called the *uniform boundedness principle* or Banach-Steinhaus Theorem.

Theorem 1.3 (Uniform boundedness). *Let B be a Banach space and suppose that T_n is a sequence of bounded (i.e. continuous) linear operators $T_n : B \longrightarrow V$ where V is a normed space. Suppose that for each $b \in B$ the set $\{T_n(b)\} \subset V$ is bounded (in norm of course) then* $\sup_n \|T_n\| < \infty$.

Proof. This follows from a pretty direct application of Baire's theorem to B. Consider the sets

$$S_p = \{b \in B,\ \|b\| \le 1,\ \|T_n b\|_V \le p\ \forall\ n\},\ p \in \mathbb{N}. \tag{1.91}$$

Each S_p is closed because T_n is continuous, so if $b_k \to b$ is a convergent sequence in S_p then $\|b\| \le 1$ and $\|T_n(b)\| \le p$. The union of the S_p is the whole of the closed ball of radius one around the origin in B :

$$\{b \in B; d(b,0) \le 1\} = \bigcup_p S_p \tag{1.92}$$

because of the assumption of 'pointwise boundedness' – each b with $\|b\| \le 1$ must be in one of the S_p's.

So, by Baire's theorem one of the sets S_p has non-empty interior, it therefore contains a closed ball of positive radius around some point. Thus for some p, some $v \in S_p$, and some $\delta > 0$,

$$w \in B,\ \|w\|_B \le \delta \Longrightarrow \|T_n(v+w)\|_V \le p\ \forall\ n. \tag{1.93}$$

Since $v \in S_p$ is fixed it follows that $\|T_n w\| \le \|T_n(v+w)\| + \|T_n v\| \le 2p$ for all w with $\|w\| \le \delta$. This however implies that the norms are uniformly bounded:

$$\|T_n\| \le 2p/\delta \tag{1.94}$$

as claimed. □

10. Open mapping theorem

The second major application of Baire's theorem is to

Theorem 1.4 (Open Mapping). *If $T : B_1 \longrightarrow B_2$ is a bounded and surjective linear map between two Banach spaces then T is open:*

$$T(O) \subset B_2 \textit{ is open if } O \subset B_1 \textit{ is open.} \tag{1.95}$$

This is 'wrong way continuity' and as such can be used to prove the continuity of inverse maps as we shall see. The proof uses Baire's theorem pretty directly, but then another similar sort of argument is needed to complete the proof. Note however that the proof is considerably simplified if we assume that B_1 is a Hilbert space. There are more direct but more computational proofs, see Problem **??**. I prefer this one because I have a reasonable chance of remembering the steps.

Proof. What we will try to show is that the image under T of the unit open ball around the origin, $B(0,1) \subset B_1$ contains an open ball around the origin in B_2. The first part, of the proof, using Baire's theorem shows that the *closure* of the image, so in B_2, has 0 as an interior point – i.e. it contains an open ball around the origin in B_2 :

$$\overline{T(B(0,1)} \supset B(0,\delta),\ \delta > 0. \tag{1.96}$$

To see this we apply Baire's theorem to the sets

$$C_p = \mathrm{cl}_{B_2}\, T(B(0,p)) \tag{1.97}$$

the closure of the image of the ball in B_1 of radius p. We know that

$$B_2 = \bigcup_p T(B(0,p)) \tag{1.98}$$

since that is what surjectivity means – every point is the image of something. Thus one of the closed sets C_p has an interior point, v. Since T is surjective, $v = Tu$ for some $u \in B_1$. The sets C_p increase with p so we can take a larger p and v is still an interior point, from which it follows that $0 = v - Tu$ is an interior point as well. Thus indeed

$$C_p \supset B(0,\delta) \tag{1.99}$$

for some $\delta > 0$. Rescaling by p, using the linearity of T, it follows that with δ replaced by δ/p, we get (1.96).

If we assume that B_1 is a Hilbert space (and you are reading this after we have studied Hilbert spaces) then (1.96) shows that if $v \in B_2$, $\|v\| < \delta$ there is a sequence u_n with $\|u_n\| \le 1$ and $Tu_n \to v$. As a bounded sequence u_n has a weakly convergent subsequence, $u_{n_j} \rightharpoonup u$, where we know this implies $\|u\| \le 1$ and $Au_{n_j} \rightharpoonup Au = v$ since $Au_n \to v$. This strengthens (1.96) to

$$T(B(0,1) \supset B(0,\delta/2)$$

and proves that T is an open map.

If B_1 is a Banach space but not a Hilbert space (or you don't yet know about Hilbert spaces) we need to work a little harder. Having applied Baire's thereom, consider now what (1.96) means. It follows that each $v \in B_2$, with $\|v\| = \delta$, is the limit of a sequence Tu_n where $\|u_n\| \le 1$. What we want to find is such a sequence, u_n, which converges. To do so we need to choose the sequence more carefully. Certainly we can stop somewhere along the way and see that

$$v \in B_2,\ \|v\| = \delta \Longrightarrow\ \exists\ u \in B_1,\ \|u\| \le 1,\ \|v - Tu\| \le \frac{\delta}{2} = \frac{1}{2}\|v\| \tag{1.100}$$

where of course we could replace $\frac{\delta}{2}$ by any positive constant but the point is the last inequality is now relative to the norm of v. Scaling again, if we take any $v \ne 0$ in B_2 and apply (1.100) to $v/\|v\|$ we conclude that (for $C = p/\delta$ a fixed constant)

$$v \in B_2 \Longrightarrow\ \exists\ u \in B_1,\ \|u\| \le C\|v\|,\ \|v - Tu\| \le \frac{1}{2}\|v\| \tag{1.101}$$

where the size of u only depends on the size of v; of course this is also true for $v = 0$ by taking $u = 0$.

Using this we construct the desired *better* approximating sequence. Given $w \in B_1$, choose $u_1 = u$ according to (1.101) for $v = w = w_1$. Thus $\|u_1\| \le C$, and $w_2 = w_1 - Tu_1$ satisfies $\|w_2\| \le \frac{1}{2}\|w\|$. Now proceed by induction, supposing that we have constructed a sequence u_j, $j < n$, in B_1 with $\|u_j\| \le C2^{-j+1}\|w\|$ and $\|w_j\| \le 2^{-j+1}\|w\|$ for $j \le n$, where $w_j = w_{j-1} - Tu_{j-1}$ – which we have for $n = 1$. Then we can choose u_n, using (1.101), so $\|u_n\| \le C\|w_n\| \le C2^{-n+1}\|w\|$

and such that $w_{n+1} = w_n - Tu_n$ has $\|w_{n+1}\| \le \frac{1}{2}\|w_n\| \le 2^{-n}\|w\|$ to extend the induction. Thus we get a sequence u_n which is absolutely summable in B_1, since $\sum\limits_n \|u_n\| \le 2C\|w\|$, and hence converges by the assumed completeness of B_1 this time. Moreover

$$w - T(\sum_{j=1}^{n} u_j) = w_1 - \sum_{j=1}^{n}(w_j - w_{j+1}) = w_{n+1} \tag{1.102}$$

so $Tu = w$ and $\|u\| \le 2C\|w\|$.

Thus finally we have shown that each $w \in B(0,1)$ in B_2 is the image of some $u \in B_1$ with $\|u\| \le 2C$. Thus $T(B(0,3C)) \supset B(0,1)$. By scaling it follows that the image of any open ball around the origin contains an open ball around the origin.

Now, the linearity of T shows that the image $T(O)$ of any open set is open, since if $w \in T(O)$ then $w = Tu$ for some $u \in O$ and hence $u + B(0,\epsilon) \subset O$ for $\epsilon > 0$ and then $w + B(0,\delta) \subset T(O)$ for $\delta > 0$ sufficiently small. □

One important corollary of this is something that seems like it should be obvious, but definitely needs completeness to be true.

Corollary 1.2. *If $T : B_1 \longrightarrow B_2$ is a bounded linear map between Banach spaces which is 1-1 and onto, i.e. is a bijection, then it is a homeomorphism – meaning its inverse, which is necessarily linear, is also bounded.*

Proof. The only confusing thing is the notation. Note that T^{-1} is generally used both for the inverse, when it exists, and also to denote the inverse map on sets even when there is no true inverse. The inverse of T, let's call it $S : B_2 \longrightarrow B_1$, is certainly linear. If $O \subset B_1$ is open then $S^{-1}(O) = T(O)$, since to say $v \in S^{-1}(O)$ means $S(v) \in O$ which is just $v \in T(O)$, is open by the Open Mapping theorem, so S is continuous. □

11. Closed graph theorem

For the next application you should check, it is one of the problems, that the product of two Banach spaces, $B_1 \times B_2$, – which is just the linear space of all pairs (u, v), $u \in B_1$ and $v \in B_2$, – is a Banach space with respect to the sum of the norms

$$\|(u, v)\| = \|u\|_1 + \|v\|_2. \tag{1.103}$$

Theorem 1.5 (Closed Graph). *If $T : B_1 \longrightarrow B_2$ is a linear map between Banach spaces then it is bounded if and only if its graph*

$$\mathrm{Gr}(T) = \{(u, v) \in B_1 \times B_2; v = Tu\} \tag{1.104}$$

is a closed subset of the Banach space $B_1 \times B_2$.

Proof. Suppose first that T is bounded, i.e. continuous. A sequence $(u_n, v_n) \in B_1 \times B_2$ is in $\mathrm{Gr}(T)$ if and only if $v_n = Tu_n$. So, if it converges, then $u_n \to u$ and $v_n = Tu_n \to Tv$ by the continuity of T, so the limit is in $\mathrm{Gr}(T)$ which is therefore closed.

Conversely, suppose the graph is closed. This means that viewed as a normed space in its own right it is complete. Given the graph we can reconstruct the map it comes from (whether linear or not) in a little diagram. From $B_1 \times B_2$ consider the two projections, $\pi_1(u, v) = u$ and $\pi_2(u, v) = v$. Both of them are continuous since the norm of either u or v is less than the norm in (1.103). Restricting them to $\mathrm{Gr}(T) \subset B_1 \times B_2$ gives

$$\begin{tikzcd} & \mathrm{Gr}(T) \arrow[dl, shift left, "\pi_1"] \arrow[dr, "\pi_2"] & \\ B_1 \arrow[ur, shift left, "S"] \arrow[rr, "T"] & & B_2. \end{tikzcd} \tag{1.105}$$

This little diagram commutes. Indeed there are two ways to map a point $(u, v) \in \mathrm{Gr}(T)$ to B_2, either directly, sending it to v or first sending it to $u \in B_1$ and then to Tu. Since $v = Tu$ these are the same.

Now, as already noted, $\mathrm{Gr}(T) \subset B_1 \times B_2$ is a closed subspace, so it too is a Banach space and π_1 and π_2 remain continuous when restricted to it. The map π_1 is 1-1 and onto, because each u occurs as the first element of precisely one pair, namely $(u, Tu) \in \mathrm{Gr}(T)$. Thus the Corollary above applies to π_1 to show that its inverse, S is continuous. But then $T = \pi_2 \circ S$, from the commutativity, is also continuous proving the theorem. □

You might wish to entertain yourself by showing that conversely the Open Mapping Theorem is a consequence of the Closed Graph Theorem.

The characterization of continuous linear maps through the fact that their graphs are closed has led to significant extensions. For instance consider a linear map but only defined on a subspace (often required to be dense in which case it is said to be 'densely defined') $D \subset B$, where B is a Banach space,

$$A : D \longrightarrow B \text{ linear.} \tag{1.106}$$

Such a map is said to be *closed* if its graph

$$\mathrm{Gr}(A) = \{(u, Au); u \in D\} \subset B \times B \text{ is closed.}$$

Check for example that if $H^1(\mathbb{R}) \subset L^2(\mathbb{R})$ (I'm assuming that you are reading this near the end of the course ...) is the space defined in Chapter 4, as consisting of the elements with a strong derivative in $L^2(\mathbb{R})$ then

$$\frac{d}{dx} : D = H^1(\mathbb{R}) \longrightarrow L^2(\mathbb{R}) \text{ is closed.} \tag{1.107}$$

This follows for instance from the 'weak implies strong' result for differentiation. If $u_n \in H^1(\mathbb{R})$ is a sequence such that $u_n \to u$ in $L^2(\mathbb{R})$ and $du_n/dx \longrightarrow v$ in L^2 (which is convergence in $L^2(\mathbb{R}) \times L^2(\mathbb{R})$) then $u \in H^1(\mathbb{R})$ and $v = du/dx$ in the same strong sense.

Such a closed operator, A, can be turned into a bounded operator by changing the norm on the domain D to the 'graph norm'

$$\|u\|_{\mathrm{Gr}} = \|u\| + \|Au\|. \tag{1.108}$$

12. Hahn-Banach theorem

Now, there is always a little pressure to state and prove the Hahn-Banach Theorem. This is about extension of functionals. Stately starkly, the basic question is: Does a normed space have *any* non-trivial continuous linear functionals on it? That is, is the dual space always non-trivial (of course there is always the zero linear functional but that is not very amusing). We do not really encounter this problem since for a Hilbert space, or even a pre-Hilbert space, there is always the space itself, giving continuous linear functionals through the pairing – Riesz' Theorem says that in the case of a Hilbert space that is all there is. If you are following the course then at this point you should also see that the only continuous linear functionals on a pre-Hilbert space correspond to points in the completion. I could have used the Hahn-Banach Theorem to show that any normed space has a completion, but I gave a more direct argument for this, which was in any case much more relevant for the cases of $L^1(\mathbb{R})$ and $L^2(\mathbb{R})$ for which we wanted *concrete* completions.

Theorem 1.6 (Hahn-Banach). *If $M \subset V$ is a linear subspace of a normed space and $u : M \longrightarrow \mathbb{C}$ is a linear map such that*

$$|u(t)| \le C\|t\|_V \ \forall\ t \in M \tag{1.109}$$

then there exists a bounded linear functional $U : V \longrightarrow \mathbb{C}$ with $\|U\| \le C$ and $U\big|_M = u$.

First, by computation, we show that we can extend any continuous linear functional 'a little bit' without increasing the norm.

Lemma 1.5. *Suppose $M \subset V$ is a subspace of a normed linear space, $x \notin M$ and $u : M \longrightarrow \mathbb{C}$ is a bounded linear functional as in* (1.109) *then there exists $u' : M' \longrightarrow \mathbb{C}$, where $M' = \{t' \in V; t' = t + ax,\ a \in \mathbb{C}\}$, such that*

$$u'\big|_M = u,\ |u'(t+ax)| \le C\|t+ax\|_V,\ \forall\ t \in M,\ a \in \mathbb{C}. \tag{1.110}$$

Proof. Note that the decomposition $t' = t + ax$ of a point in M' is unique, since $t + ax = \tilde{t} + \tilde{a}x$ implies $(a - \tilde{a})x \in M$ so $a = \tilde{a}$, since $x \notin M$ and hence $t = \tilde{t}$ as well. Thus

$$u'(t+ax) = u'(t) + au(x) = u(t) + \lambda a,\ \lambda = u'(x) \tag{1.111}$$

and all we have at our disposal is the choice of λ. Any choice will give a linear functional extending u, the problem of course is to arrange the continuity estimate without increasing the constant C. In fact if $C = 0$ then $u = 0$ and we can take the zero extension. So we might as well assume that $C = 1$ since dividing u by C arranges this and if u' extends u/C then Cu' extends u and the norm estimate in (1.110) follows. So we now assume that

$$|u(t)| \le \|t\|_V \ \forall\ t \in M. \tag{1.112}$$

We want to choose λ so that

$$|u(t) + a\lambda| \le \|t + ax\|_V \ \forall\ t \in M,\ a \in \mathbb{C}. \tag{1.113}$$

Certainly when $a = 0$ this represents no restriction on λ. For $a \ne 0$ we can divide through by $-a$ and (1.113) becomes

$$|a||u(-\frac{t}{a}) - \lambda| = |u(t) + a\lambda| \le \|t + ax\|_V = |a|\| - \frac{t}{a} - x\|_V \tag{1.114}$$

and since $-t/a \in M$ we only need to arrange that

$$|u(t) - \lambda| \le \|t - x\|_V \ \forall\ t \in M \tag{1.115}$$

and the general case will follow by reversing the scaling.

A complex linear functional such as u can be recovered from its real part, as we see below, so set

$$w(t) = \operatorname{Re}(u(t)),\ |w(t)| \le \|t\|_V \ \forall\ t \in M. \tag{1.116}$$

We proceed to show the real version of the Lemma, that w can be extended to a linear functional $w' : M + \mathbb{R}x \longrightarrow \mathbb{R}$ if $x \notin M$ without increasing the norm. The same argument as above shows that the only freedom is the choice of $\lambda = w'(x)$ and we need to choose $\lambda \in \mathbb{R}$ so that

$$|w(t) - \lambda| \le \|t - x\|_V \ \forall \ t \in M. \tag{1.117}$$

The norm estimate on w shows that

$$|w(t_1) - w(t_2)| \le |u(t_1) - u(t_2)| \le \|t_1 - t_2\| \le \|t_1 - x\|_V + \|t_2 - x\|_V. \tag{1.118}$$

Writing this out using the reality we find

$$\begin{gathered} w(t_1) - w(t_2) \le \|t_1 - x\|_V + \|t_2 - x\|_V \Longrightarrow \\ w(t_1) - \|t_1 - x\| \le w(t_2) + \|t_2 - x\|_V \ \forall \ t_1, \ t_2 \in M. \end{gathered} \tag{1.119}$$

We can then take the supremum on the left and the infimum on the right and choose λ in between – namely we have shown that there exists $\lambda \in \mathbb{R}$ with

$$\begin{aligned} w(t) - \|t - x\|_V \le \sup_{t_2 \in M} (w(t_1) - \|t_1 - x\|) \le \lambda \\ \le \inf_{t_2 \in M} (w(t_1) + \|t_1 - x\|) \le w(t) + \|t - x\|_V \ \forall \ t \in M. \end{aligned} \tag{1.120}$$

This in turn implies that

$$-\|t - x\|_V \le -w(t) + \lambda \le \|t - x\|_V \Longrightarrow |w(t) - \lambda| \le \|t - x\|_V \ \forall \ t \in M. \tag{1.121}$$

So we have an extension of w to a real functional $w' : M + \mathbb{R}x \longrightarrow \mathbb{R}$ with $|w'(t + ax)| \le \|t + ax\|_V$ for all $a \in \mathbb{R}$. We can repeat this argument to obtain a further extension $w'' : M + \mathbb{C}x = M + \mathbb{R}x + \mathbb{R}(ix) \longrightarrow \mathbb{R}$ without increasing the norm.

Now we find the desired extension of u by setting

$$u'(t + cx) = w''(t + ax + b(ix)) - iw''(it - bx + a(ix)) : M + \mathbb{C}x \longrightarrow \mathbb{C} \tag{1.122}$$

where $c = a + ib$. This is certainly linear over the reals and linearity over complex coefficients follows since

$$\begin{aligned} u'(it + icx) = w''(it - bx + a(ix))) - iw''(-t - ax - b(ix)) \\ = i(w''(t + ax + b(ix) - iw''(it - bx + a(ix)))) = iu'(t + cx). \end{aligned} \tag{1.123}$$

The uniqueness of a complex linear functional with given real part also shows that $u'\big|_M = u$.

Finally, to estimate the norm of u' notice that for each $t \in M$ and $c \in \mathbb{C}$ there is a unique $\theta \in [0, 2\pi)$ such that

$$|u'(t + cx)| = \operatorname{Re} e^{i\theta} u'(t + cx) = w''(e^{i\theta} t + e^{i\theta} cx) \le \|e^{i\theta} t + e^{i\theta} cx\|_V = \|t + cx\|_V. \tag{1.124}$$

This completes the proof of the Lemma. □

Proof of Hahn-Banach. This is an application of Zorn's Lemma. I am not going to get into the derivation of Zorn's Lemma from the Axiom of Choice, but if you believe the latter – and you are advised to do so, at least before lunchtime – you should believe the former. See also the discussion in Section 2.9

Zorn's Lemma is a statement about partially ordered sets. A partial order on a set E is a subset of $E \times E$, so a relation, where the condition that (e, f) be in the relation is written $e \prec f$ and it must satisfy the three conditions

$$\begin{gathered} e \prec e \;\forall\; e \in E \\ e \prec f \text{ and } f \prec e \Longrightarrow e = f \\ e \prec f \text{ and } f \prec g \Longrightarrow e \prec g. \end{gathered} \tag{1.125}$$

So, the missing ingredient between this and an order is that two elements need not be related at all, either way.

A subset of a partially ordered set inherits the partial order and such a subset is said to be a *chain* if each pair of its elements *is* related one way or the other. An *upper bound* on a subset $D \subset E$ is an element $e \in E$ such that $d \prec e$ for all $d \in D$. A *maximal* element of E is one which is not majorized, that is $e \prec f$, $f \in E$, implies $e = f$.

Lemma 1.6 (Zorn). *If every chain in a (non-empty) partially ordered set has an upper bound then the set contains at least one maximal element.*

So, we are just accepting this Lemma as axiomatic. However, make sure that you appreciate that it is true for countable sets. Namely if C is countable and has no maximal element then it must contain a chain which has no upper bound. To see this, write C as a sequence $\{c_i\}_{i\in\mathbb{N}}$. Then $x_1 = c_1$ is not maximal so there exists some c_k, $k > 1$, with $c_1 \prec c_k$ in terms of the order in C. From the properties of $\mathbb{N}$ it follows that there is a *smallest* $k = k_2$ such that c_k has this property, but $k_2 > 1$. Let this be $x_2 = c_{k_2}$ and proceed in the same way – $x_3 = c_{k_3}$ where $k_3 > k_2$ is the smallest such integer for which $c_{k_2} \prec c_{k_3}$. Assuming C is infinite in the first place this grinds out an infinite chain x_i. Now you can check that this cannot have an upper bound because every element of C is either one of these, and so cannot be an upper bound, or else it is c_j with $k_l < j < k_{l+1}$ for some l and then it is not greater than x_l.

The point of Zorn's Lemma is precisely that it applies to uncountable sets.

One consequence of Zorn's Lemma is the existence of 'Hamel bases' for infinite dimensional vector spaces. This is pretty much irrelevant for us but I include it since you can use this to show the existence of non-continuous linear functionals on a Banach space; the proof is also an easier version of the proof of the Hahn-Banach theorem.

Definition 1.6. A *Hamel basis* $B \subset V$ of a vector space is a linearly independent subspace which spans, i.e. every element of V is a finite linear combination of elements of B.

Hamel bases have a strong tendency to be big. Notice that for a finite dimensional space this is just the usual notion of a basis.

Proof. Look at the collection $\mathcal{X}$ of all linearly independent subsets of V. This is non-empty – assuming $V \neq \{0\}$ we can take $\{x\} \in \mathcal{X}$ for any $x \neq 0$ in V. Inclusion is a partial order on $\mathcal{X}$. Suppose $\mathcal{Y} \subset \mathcal{X}$ is a chain with respect to this partial order – so for any two elements one is contained in the other. Let $L = \cup\mathcal{Y}$ be the union of all the elements of $\mathcal{Y}$. Each element of $\mathcal{C}$ is contained in L so L is an upper bounde for $\mathcal{Y}$. Thus by Zorn's lemma $\mathcal{X}$ just contain a maximal element, say B. Then B is a Hamel basis, since if some element $w \in V$ were not a finite linear combination of elements of B then $B \cup \{w\}$ would also be linearly independent and hence an element of $\mathcal{X}$ which contains B which contradicts its maximality. $\square$

So, back to Hahn-Banach.

We are given a functional $u : M \longrightarrow \mathbb{C}$ defined on some linear subspace $M \subset V$ of a normed space where u is bounded with respect to the induced norm on M. We will apply Zorn's Lemma to the set E consisting of *all* extensions (v, N) of u with the same norm; it is generally non-countable. That is,

$$V \supset N \supset M,\ v\big|_M = u \text{ and } \|v\|_N = \|u\|_M.$$

This is certainly non-empty since it contains (u, M) and has the natural partial order that $(v_1, N_1) \prec (v_2, N_2)$ if $N_1 \subset N_2$ and $v_2\big|_{N_1} = v_1$. You should check that this is a partial order.

Let C be a chain in this set of extensions. Thus for any two elements $(v_i, N_i) \in C$, $i = 1$, 2, either $(v_1, N_1) \prec (v_2, N_2)$ or the other way around. This means that

$$\tilde{N} = \bigcup\{N; (v, N) \in C \text{ for some } v\} \subset V \tag{1.126}$$

is a linear space. Note that this union need not be countable, or anything like that, but any two elements of $\tilde{N}$ are each in one of the N's and one of these must be contained in the other by the chain condition. Thus each pair of elements of $\tilde{N}$ is actually in a common N and hence so is their linear span. Similarly we can define an extension

$$\tilde{v} : \tilde{N} \longrightarrow \mathbb{C},\ \tilde{v}(x) = v(x) \text{ if } x \in N,\ (v, N) \in C. \tag{1.127}$$

There may be many pairs $(v, N) \in C$ satisfying $x \in N$ for a given x but the chain condition implies that $v(x)$ is the same for all of them. Thus $\tilde{v}$ is well defined, and is clearly also linear, extends u and satisfies the norm condition $|\tilde{v}(x)| \le \|u\|_M \|x\|_V$. Thus $(\tilde{v}, \tilde{N})$ is an upper bound for the chain C.

So, the set of all extensions, E, with the norm condition, satisfies the hypothesis of Zorn's Lemma, so must – at least in the mornings – have a maximal element $(\tilde{u}, \tilde{M})$. If $\tilde{M} = V$ then we are done. However, in the contary case there exists $x \in V \setminus \tilde{M}$. This means we can apply our little lemma and construct an extension $(u', \tilde{M}')$ of $(\tilde{u}, \tilde{M})$ which is therefore also an element of E and satisfies $(\tilde{u}, \tilde{M}) \prec (u', \tilde{M}')$. This however contradicts the condition that $(\tilde{u}, \tilde{M})$ be maximal, so is forbidden by Zorn. $\square$

There are many applications of the Hahn-Banach Theorem. As remarked earlier, one significant one is that the dual space of a non-trivial normed space is itself non-trivial.

Proposition 1.7. *For any normed space V and element $0 \neq v \in V$ there is a continuous linear functional $f : V \longrightarrow \mathbb{C}$ with $f(v) = 1$ and $\|f\| = 1/\|v\|_V$.*

Proof. Start with the one-dimensional space, M, spanned by v and define $u(zv) = z$. This has norm $1/\|v\|_V$. Extend it using the Hahn-Banach Theorem and you will get a continuous functional f as desired. $\square$

13. Double dual

Let me give another application of the Hahn-Banach theorem, although I have generally not covered this in lectures. If V is a normed space, we know its dual space, V', to be a Banach space. Let $V'' = (V')'$ be the dual of the dual.

Proposition 1.8. *If $v \in V$ then the linear map on V' :*

$$T_v : V' \longrightarrow \mathbb{C},\ T_v(v') = v'(v) \tag{1.128}$$

is continuous and this defines an isometric linear injection $V \hookrightarrow V''$, $\|T_v\| = \|v\|$.

Proof. The definition of T_v is 'tautologous', meaning it is almost the definition of V'. First check T_v in (1.128) is linear. Indeed, if v'_1, $v'_2 \in V'$ and λ_1, $\lambda_2 \in \mathbb{C}$ then $T_v(\lambda_1 v'_1 + \lambda_2 v'_2) = (\lambda_1 v'_1 + \lambda_2 v'_2)(v) = \lambda_1 v'_1(v) + \lambda_2 v'_2(v) = \lambda_1 T_v(v'_1) + \lambda_2 T_v(v'_2)$. That $T_v \in V''$, i.e. is bounded, follows too since $|T_v(v')| = |v'(v)| \leq \|v'\|_{V'}\|v\|_V$; this also shows that $\|T_v\|_{V''} \leq \|v\|$. On the other hand, by Proposition 1.7 above, if $\|v\| = 1$ then there exists $v' \in V'$ such that $v'(v) = 1$ and $\|v'\|_{V'} = 1$. Then $T_v(v') = v'(v) = 1$ shows that $\|T_v\| = 1$ so in general $\|T_v\| = \|v\|$. It also needs to be checked that $V \ni v \longmapsto T_v \in V''$ is a linear map – this is clear from the definition. It is necessarily 1-1 since $\|T_v\| = \|v\|$. □

Now, it is definitely not the case in general that $V'' = V$ in the sense that this injection is also a surjection. Since V'' is always a Banach space, one necessary condition is that V itself should be a Banach space. In fact the closure of the image of V in V'' is a completion of V. If the map to V'' is a bijection then V is said to be *reflexive*. It is pretty easy to find examples of non-reflexive Banach spaces, the most familiar is c_0 – the space of infinite sequences converging to 0. Its dual can be identified with l^1, the space of summable sequences. Its dual in turn, the bidual of c_0, is the space l^∞ of bounded sequences, into which the embedding is the obvious one, so c_0 is not reflexive. In fact l^1 is not reflexive either. There are useful characterizations of reflexive Banach spaces. You may be interested enough to look up James' Theorem:- A Banach space is reflexive if and only if every continuous linear functional on it attains its supremum on the unit ball.

14. Axioms of a vector space

In case you missed out on one of the basic linear algebra courses, or have a poor memory, here are the axioms of a vector space over a field $\mathbb{K}$ (either $\mathbb{R}$ or $\mathbb{C}$ for us).

A *vector space structure* on a set V is a pair of maps

$$+ : V \times V \longrightarrow V,\ \cdot : \mathbb{K} \times V \longrightarrow V \tag{1.129}$$

satisfying the conditions listed below. These maps are written $+(v_1, v_2) = v_1 + v_2$ and $\cdot(\lambda, v) = \lambda v$, $\lambda \in \mathbb{K}$, V, v_1, $v_2 \in V$.

additive commutativity $v_1 + v_2 = v_2 + v_1$ for all v_1, $v_2 \in V$.
additive associativity $v_1 + (v_2 + v_3) = (v_1 + v_2) + v_3$ for all v_1, v_2, $v_3 \in V$.
existence of zero There is an element $0 \in V$ such that $v + 0 = v$ for all $v \in V$.
additive invertibility For each $v \in V$ there exists $w \in V$ such that $v + w = 0$.
distributivity of scalar additivity $(\lambda_1 + \lambda_2)v = \lambda_1 v + \lambda_2 v$ for all λ_1, $\lambda_2 \in \mathbb{K}$ and $v \in V$.
multiplicativity $\lambda_1(\lambda_2 v) = (\lambda_1\lambda_2)v$ for all λ_1, $\lambda_2 \in \mathbb{K}$ and $v \in V$.
action of multiplicative identity $1v = v$ for all $v \in V$.
distributivity of space additivity $\lambda(v_1 + v_2) = \lambda v_1 + \lambda v_2$ for all $\lambda \in \mathbb{K}$ v_1, $v_2 \in V$.

CHAPTER 2

The Lebesgue integral

In this second part of the course the basic theory of the Lebesgue integral is presented. Here I follow an idea of Jan Mikusiński, of completing the space of step functions on the line under the L^1 norm but in such a way that the limiting objects are seen directly as functions (defined almost everywhere). There are other places you can find this, for instance the book of Debnaith and Mikusiński [**1**]. Here I start from the Riemann integral, since this is a prerequisite of the course; this streamlines things a little. The objective is to arrive at a working knowledge of Lebesgue integration as quickly as seems acceptable, to pass on to the discussion of Hilbert space and then to more analytic questions.

So, the treatment of the Lebesgue integral here is intentionally compressed, while emphasizing the completeness of the spaces L^1 and L^2. In lectures everything is done for the real line but in such a way that the extension to higher dimensions – carried out partly in the text but mostly in the problems – is not much harder.

1. Integrable functions

Recall that the Riemann integral is defined for a certain class of bounded functions $u : [a, b] \longrightarrow \mathbb{C}$ (namely the Riemann integrable functions) which includes all continuous functions. It depends on the compactness of the interval and the boundedness of the function, but can be extended to an 'improper integral' on the whole real line for which however some of the good properties fail. This is NOT what we will do. Rather we consider the space of continuous functions 'with compact support':

$$\mathcal{C}_{\mathrm{c}}(\mathbb{R}) = \{u : \mathbb{R} \longrightarrow \mathbb{C}; u \text{ is continuous and } \exists\ R \text{ such that } u(x) = 0 \text{ if } |x| > R\}. \tag{2.1}$$

Thus each element $u \in \mathcal{C}_{\mathrm{c}}(\mathbb{R})$ vanishes outside an interval $[-R, R]$ where the R depends on the u. Note that the *support* of a continuous function is defined to be the complement of the largest open set on which it vanishes (or as the closure of the set of points at which it is non-zero – make sure you see why these are the same). Thus (2.1) says that the support, which is necessarily closed, is contained in some interval $[-R, R]$, which is equivalent to saying it is compact.

LEMMA 2.1. *The Riemann integral defines a continuous linear functional on $\mathcal{C}_c(\mathbb{R})$ equipped with the L^1 norm*

$$\begin{gathered}\int_{\mathbb{R}} u = \lim_{R\to\infty} \int_{[-R,R]} u(x)dx,\\ \|u\|_{L^1} = \lim_{R\to\infty} \int_{[-R,R]} |u(x)|dx, \\ |\int_{\mathbb{R}} u| \le \|u\|_{L^1}.\end{gathered} \tag{2.2}$$

The limits here are trivial in the sense that the functions involved are constant for large R.

PROOF. These are basic properties of the Riemann integral see Rudin [**4**]. □

Note that $\mathcal{C}_{\rm c}(\mathbb{R})$ is a normed space with respect to $\|u\|_{L^1}$ as defined above; that it is not complete is one of the main reasons for passing to the Lebesgue integral. With this small preamble we can directly define the 'space' of Lebesgue integrable functions on $\mathbb{R}$.

DEFINITION 2.1. A function $f : \mathbb{R} \longrightarrow \mathbb{C}$ is *Lebesgue integrable*, written $f \in \mathcal{L}^1(\mathbb{R})$, if there exists a series with partial sums $f_n = \sum\limits_{j=1}^{n} w_j$, $w_j \in \mathcal{C}_{\rm c}(\mathbb{R})$ which is absolutely summable,

$$\sum_j \int |w_j| < \infty \tag{2.3}$$

and such that

$$\sum_j |w_j(x)| < \infty \Longrightarrow \lim_{n\to\infty} f_n(x) = \sum_j w_j(x) = f(x). \tag{2.4}$$

This is a somewhat convoluted definition which you should think about a bit. Its virtue is that it is all there. The problem is that it takes a bit of unravelling. Before we go any further note that the sequence w_j obviously determines the sequence of partial sums f_n, both in $\mathcal{C}_{\rm c}(\mathbb{R})$ but the converse is also true since

$$\begin{gathered}w_1 = f_1,\ w_k = f_k - f_{k-1},\ k > 1,\\ \sum_j \int |w_j| < \infty \Longleftrightarrow \sum_{k>1} \int |f_k - f_{k-1}| < \infty.\end{gathered} \tag{2.5}$$

You might also notice that can we do some finite manipulation, for instance replace the sequence w_j by

$$W_1 = \sum_{j\le N} w_j,\ W_k = w_{N+k-1},\ k > 1 \tag{2.6}$$

and nothing much changes, since the convergence conditions in (2.3) and (2.4) are properties only of the *tail* of the sequences and the sum in (2.4) for $w_j(x)$ converges if and only if the corresponding sum for $W_k(x)$ converges and then converges to the same limit.

Before massaging the definition a little, let me give a simple example and check that this definition does include continuous functions defined on an interval and extended to be zero outside – the theory we develop will include the usual Riemann integral although I will not quite prove this in full, but only because it is not particularly interesting.

LEMMA 2.2. *If $f \in \mathcal{C}([a,b])$ then*

$$\tilde{f}(x) = \begin{cases} f(x) & \text{if } x \in [a,b] \\ 0 & \text{otherwise} \end{cases} \tag{2.7}$$

is an integrable function.

PROOF. Just 'add legs' to $\tilde{f}$ by considering the sequence

$$f_n(x) = \begin{cases} 0 & \text{if } x < a - 1/n \text{ or } x > b + 1/n, \\ (1 + n(x-a))f(a) & \text{if } a - 1/n \le x < a, \\ (1 - n(x-b))f(b) & \text{if } b < x \le b + 1/n, \\ f(x) & \text{if } x \in [a,b]. \end{cases} \tag{2.8}$$

This is a continuous function on each of the open subintervals in the description with common limits at the endpoints, so $f_n \in \mathcal{C}_{\text{c}}(\mathbb{R})$. By construction, $f_n(x) \to \tilde{f}(x)$ for each $x \in \mathbb{R}$. Define the sequence w_j which has partial sums the f_n, as in (2.5) above. Then $w_j = 0$ in $[a,b]$ for $j > 1$ and it can be written in terms of the 'legs'

$$l_n = \begin{cases} 0 & \text{if } x < a - 1/n,\ x \ge a \\ (1 + n(x-a)) & \text{if } a - 1/n \le x < a, \end{cases}$$

$$r_n = \begin{cases} 0 & \text{if } x \le b,\ x > b + 1/n \\ (1 - n(x-b)) & \text{if } b \le x \le b + 1/n, \end{cases}$$

as

$$|w_n(x)| = (l_n - l_{n-1})|f(a)| + (r_n - r_{n-1})|f(b)|,\ n > 1. \tag{2.9}$$

It follows that

$$\int |w_n(x)| = \frac{(|f(a)| + |f(b)|)}{n(n-1)}$$

so $\{w_n\}$ is an absolutely summable sequence showing that $\tilde{f} \in \mathcal{L}^1(\mathbb{R})$. □

Returning to the definition, notice that we only say 'there exists' an absolutely summable sequence and that it is required to converge to the function *only* at points at which the pointwise sequence is absolutely summable. At other points anything is permitted. So it is not immediately clear that there are any functions *not* satisfying this condition. Indeed if there was a sequence like w_j above with $\sum\limits_j |w_j(x)| = \infty$ always, then (2.4) would represent no restriction at all. So the point of the definition is that absolute summability – a condition on the integrals in (2.3) – does imply something about (absolute) convergence of the pointwise series. Let us reenforce this idea with another definition:-

DEFINITION 2.2. A set $E \subset \mathbb{R}$ is said to be *of measure zero* in the sense of Lebesgue (which is pretty much always the meaning here) if there is a series $g_n = \sum\limits_{j=1}^{n} v_j$, $v_j \in \mathcal{C}_{\text{c}}(\mathbb{R})$ which is absolutely summable, $\sum\limits_j \int |v_j| < \infty$, and such that

$$\sum_j |v_j(x)| = \infty \ \forall\ x \in E. \tag{2.10}$$

Notice that we do not require E to be precisely the set of points at which the series in (2.10) diverges, only that it does so at all points of E, so E is just a subset of the set on which some absolutely summable series of functions in $\mathcal{C}_{\mathrm{c}}(\mathbb{R})$ does not converge absolutely. So any subset of a set of measure zero is automatically of measure zero. To introduce the little trickery we use to unwind the definition above, consider first the following (important) result.

Lemma 2.3. *Any finite union of sets of measure zero is a set of measure zero.*

Proof. Since we can proceed in steps, it suffices to show that the union of two sets of measure zero has measure zero. So, let the two sets be E and F and two corresponding absolutely summable sequences, as in Definition 2.2, be v_j and w_j. Consider the alternating sequence

$$u_k = \begin{cases} v_j & \text{if } k = 2j-1 \text{ is odd} \\ w_j & \text{if } k = 2j \text{ is even.} \end{cases} \tag{2.11}$$

Thus $\{u_k\}$ simply interlaces the two sequences. It follows that u_k is absolutely summable, since

$$\sum_k \|u_k\|_{L^1} = \sum_j \|v_j\|_{L^1} + \sum_j \|w_j\|_{L^1}. \tag{2.12}$$

Moreover, the pointwise series $\sum\limits_k |u_k(x)|$ diverges precisely where one or other of the two series $\sum\limits_j |v_j(x)|$ or $\sum\limits_j |w_j(x)|$ diverges. In particular it must diverge on $E \cup F$ which is therefore, from the definition, a set of measure zero. □

The definition of $f \in \mathcal{L}^1(\mathbb{R})$ above certainly requires that the equality on the right in (2.4) should hold outside a set of measure zero, but in fact a specific one, the one on which the series on the left diverges. Using the same idea as in the lemma above we can get rid of this restriction.

Proposition 2.1. *If $f : \mathbb{R} \longrightarrow \mathbb{C}$ and there exists a series $f_n = \sum\limits_{j=1}^{n} w_j$ with $w_j \in \mathcal{C}_c(\mathbb{R})$ which is absolutely summable, so $\sum\limits_j \|w_j\|_{L^1} < \infty$, and a set $E \subset \mathbb{R}$ of measure zero such that*

$$x \in \mathbb{R} \setminus E \Longrightarrow f(x) = \lim_{n\to\infty} f_n(x) = \sum_{j=1}^{\infty} w_j(x) \tag{2.13}$$

then $f \in \mathcal{L}^1(\mathbb{R})$.

Recall that when one writes down an equality such as on the right in (2.13) one is implicitly saying that $\sum\limits_{j=1}^{\infty} w_j(x)$ converges *and* the equality holds for the limit. We will call a sequence as the w_j above an ‘approximating series’ for $f \in \mathcal{L}^1(\mathbb{R})$.
This is indeed a refinement of the definition since all $f \in \mathcal{L}^1(\mathbb{R})$ arise this way, taking E to be the set where $\sum\limits_j |w_j(x)| = \infty$ for a series as in the defintion.

Proof. By definition of a set of measure zero there is some series v_j as in (2.10). Now, consider the series obtained by alternating the terms between w_j, v_j

and $-v_j$. Explicitly, set

$$u_j = \begin{cases} w_k & \text{if } j = 3k-2 \\ v_k & \text{if } j = 3k-1 \\ -v_k(x) & \text{if } j = 3k. \end{cases} \tag{2.14}$$

This defines a series in $\mathcal{C}_{\mathrm{c}}(\mathbb{R})$ which is absolutely summable, with

$$\sum_j \|u_j(x)\|_{L^1} = \sum_k \|w_k\|_{L^1} + 2\sum_k \|v_k\|_{L^1}. \tag{2.15}$$

The same sort of identity is true for the pointwise series which shows that

$$\sum_j |u_j(x)| < \infty \text{ iff } \sum_k |w_k(x)| < \infty \text{ and } \sum_k |v_k(x)| < \infty. \tag{2.16}$$

So if the pointwise series on the left converges absolutely, then $x \notin E$, by definition and hence, using (2.13), we find that

$$\sum_j |u_j(x)| < \infty \Longrightarrow f(x) = \sum_j u_j(x) \tag{2.17}$$

since the sequence of partial sums of the u_j cycles through f_n, $f_n(x) + v_n(x)$, then $f_n(x)$ and then to $f_{n+1}(x)$. Since $\sum\limits_k |v_k(x)| < \infty$ the sequence $|v_n(x)| \to 0$ so (2.17) indeed follows from (2.13). □

This is the trick at the heart of the definition of integrability above. Namely we can manipulate the series involved in this sort of way to prove things about the elements of $\mathcal{L}^1(\mathbb{R})$. One point to note is that if w_j is an absolutely summable series in $\mathcal{C}_{\mathrm{c}}(\mathbb{R})$ then

$$F(x) = \begin{cases} \sum\limits_j |w_j(x)| & \text{when this is finite} \\ 0 & \text{otherwise} \end{cases} \Longrightarrow F \in \mathcal{L}^1(\mathbb{R}). \tag{2.18}$$

The sort of property (2.13), where some condition holds on the complement of a set of measure zero is so commonly encountered in integration theory that we give it a simpler name.

Definition 2.3. A condition that holds on $\mathbb{R} \setminus E$ for some set of measure zero, E, is said to hold *almost everywhere.* In particular we write

$$f = g \textit{ a.e.} \text{ if } f(x) = g(x) \ \forall \ x \in \mathbb{R} \setminus E,\ E \text{ of measure zero.} \tag{2.19}$$

Of course as yet we are living dangerously because we have done nothing to show that sets of measure zero are 'small' let alone 'ignorable' as this definition seems to imply. Beware of the trap of 'proof by declaration'!

Now Proposition 2.1 can be paraphrased as 'A function $f : \mathbb{R} \longrightarrow \mathbb{C}$ is Lebesgue integrable if and only if it is the pointwise sum a.e. of an absolutely summable series in $\mathcal{C}_{\mathrm{c}}(\mathbb{R})$.'

2. Linearity of $\mathcal{L}^1$

The word 'space' is quoted in the definition of $\mathcal{L}^1(\mathbb{R})$ above, because it is not immediately obvious that $\mathcal{L}^1(\mathbb{R})$ is a linear space, even more importantly it is far from obvious that the integral of a function in $\mathcal{L}^1(\mathbb{R})$ is well defined (which is the point of the exercise after all). In fact we wish to define the integral to be

$$\int_{\mathbb{R}} f = \sum_n \int w_n \tag{2.20}$$

where $w_n \in \mathcal{C}_{\mathrm{c}}(\mathbb{R})$ is any 'approximating series' meaning now as the w_j in Proposition 2.1. This is fine in so far as the series on the right (of complex numbers) does converge – since we demanded that $\sum\limits_n \int |w_n| < \infty$ so this series converges absolutely – but not fine in so far as the answer might well depend on *which* series we choose which 'approximates f' in the sense of the definition or Proposition 2.1.

So, the immediate aim is to prove these two things. First we will do a little more than prove the linearity of $\mathcal{L}^1(\mathbb{R})$. Recall that a function is 'positive' if it takes only non-negative values.

Proposition 2.2. *The space $\mathcal{L}^1(\mathbb{R})$ is linear (over $\mathbb{C}$) and if $f \in \mathcal{L}^1(\mathbb{R})$ the real and imaginary parts,* $\operatorname{Re} f$, $\operatorname{Im} f$ *are Lebesgue integrable as are their positive parts and as is also the absolute value, $|f|$. For a real Lebesgue integrable function there is an approximating sequence as in Proposition 2.1 which is real and if $f \geq 0$ the sequence of partial sums can be arranged to be non-negative.*

Proof. We first consider the real part of a function $f \in \mathcal{L}^1(\mathbb{R})$. Suppose $w_n \in \mathcal{C}_{\mathrm{c}}(\mathbb{R})$ is an approximating series as in Proposition 2.1. Then consider $v_n = \operatorname{Re} w_n$. This is absolutely summable, since $\int |v_n| \leq \int |w_n|$ and

$$\sum_n w_n(x) = f(x) \Longrightarrow \sum_n v_n(x) = \operatorname{Re} f(x). \tag{2.21}$$

Since the left identity holds a.e., so does the right and hence $\operatorname{Re} f \in \mathcal{L}^1(\mathbb{R})$ by Proposition 2.1. The same argument with the imaginary parts shows that $\operatorname{Im} f \in \mathcal{L}^1(\mathbb{R})$. This also shows that a real element has a real approximating sequence.

The fact that the sum of two integrable functions is integrable really is a simple consequence of Proposition 2.1 and Lemma 2.3. Indeed, if $f,\ g \in \mathcal{L}^1(\mathbb{R})$ have approximating series w_n and v_n as in Proposition 2.1 then $u_n = w_n + v_n$ is absolutely summable,

$$\sum_n \int |u_n| \leq \sum_n \int |w_n| + \sum_n \int |v_n| \tag{2.22}$$

and

$$\sum_n w_n(x) = f(x),\ \sum_n v_n(x) = g(x) \Longrightarrow \sum_n u_n(x) = f(x) + g(x).$$

The first two conditions hold outside (probably different) sets of measure zero, E and F, so the conclusion holds outside $E \cup F$ which is of measure zero. Thus $f + g \in \mathcal{L}^1(\mathbb{R})$. The case of cf for $c \in \mathbb{C}$ is more obvious.

The proof that $|f| \in \mathcal{L}^1(\mathbb{R})$ if $f \in \mathcal{L}^1(\mathbb{R})$ is similar but perhaps a little trickier. Again, let $\{w_n\}$ be an approximating series as in the definition showing that $f \in$

$\mathcal{L}^1(\mathbb{R})$. To make a series for $|f|$ we can try the 'obvious' thing. Namely we know that

$$\sum_{j=1}^{n} w_j(x) \to f(x) \text{ if } \sum_j |w_j(x)| < \infty \tag{2.23}$$

so certainly it follows that

$$|\sum_{j=1}^{n} w_j(x)| \to |f(x)| \text{ if } \sum_j |w_j(x)| < \infty.$$

So, set

$$v_1(x) = |w_1(x)|,\ v_k(x) = |\sum_{j=1}^{k} w_j(x)| - |\sum_{j=1}^{k-1} w_j(x)|\ \forall\ x \in \mathbb{R}. \tag{2.24}$$

Then, for sure,

$$\sum_{k=1}^{N} v_k(x) = |\sum_{j=1}^{N} w_j(x)| \to |f(x)| \text{ if } \sum_j |w_j(x)| < \infty. \tag{2.25}$$

So equality holds off a set of measure zero and we only need to check that $\{v_j\}$ is an absolutely summable series.

The triangle inequality in the 'reverse' form $||v| - |w|| \le |v - w|$ shows that, for $k > 1$,

$$|v_k(x)| = ||\sum_{j=1}^{k} w_j(x)| - |\sum_{j=1}^{k-1} w_j(x)|| \le |w_k(x)|. \tag{2.26}$$

Thus

$$\sum_k \int |v_k| \le \sum_k \int |w_k| < \infty \tag{2.27}$$

so the v_k's do indeed form an absolutely summable series and (2.25) holds almost everywhere, so $|f| \in \mathcal{L}^1(\mathbb{R})$.

For a positive function this last argument yields a real approximating sequence with positive partial sums. □

By combining these results we can see again that if f, $g \in \mathcal{L}^1(\mathbb{R})$ are both real valued then

$$f_+ = \max(f, 0),\ \max(f, g),\ \min(f, g) \in \mathcal{L}^1(\mathbb{R}). \tag{2.28}$$

Indeed, the positive part, $f_+ = \frac{1}{2}(|f| + f)$, $\max(f, g) = g + (f - g)_+$, $\min(f, g) = -\max(-f, -g)$.

3. The integral on $\mathcal{L}^1$

Next we want to show that the integral is well defined via (2.20) for any approximating series. From Propostion 2.2 it is enough to consider only real functions. For this, recall a result concerning a case where uniform convergence of continuous functions follows from pointwise convergence, namely when the convergence is monotone, the limit is continuous, and the space is compact. It works on a general compact metric space but we can concentrate on the case at hand.

Lemma 2.4. *If $u_n \in \mathcal{C}_c(\mathbb{R})$ is a decreasing sequence of non-negative functions such that $\lim_{n\to\infty} u_n(x) = 0$ for each $x \in \mathbb{R}$ then $u_n \to 0$ uniformly on $\mathbb{R}$ and*

$$\lim_{n\to\infty} \int u_n = 0. \tag{2.29}$$

Proof. Since all the $u_n(x) \ge 0$ and they are decreasing (which really means not increasing of course) if $u_1(x)$ vanishes at x then all the other $u_n(x)$ vanish there too. Thus there is one $R > 0$ such that $u_n(x) = 0$ if $|x| > R$ for all n, namely one that works for u_1. So we only need consider what happens on $[-R, R]$ which is compact. For any $\epsilon > 0$ look at the sets

$$S_n = \{x \in [-R, R]; u_n(x) \ge \epsilon\}.$$

This can also be written $S_n = u_n^{-1}([\epsilon, \infty)) \cap [-R, R]$ and since u_n is continuous it follows that S_n is closed and hence compact. Moreover the fact that the $u_n(x)$ are decreasing means that $S_{n+1} \subset S_n$ for all n. Finally,

$$\bigcap_n S_n = \emptyset$$

since, by assumption, $u_n(x) \to 0$ for each x. Now the property of compact sets in a metric space that we use is that if such a sequence of decreasing compact sets has empty intersection then the sets themselves are empty from some n onwards. This means that there exists N such that $\sup_x u_n(x) < \epsilon$ for all $n > N$. Since $\epsilon > 0$ was arbitrary, $u_n \to 0$ uniformly.

One of the basic properties of the Riemann integral is that the integral of the limit of a uniformly convergent sequence (even of Riemann integrable functions but here continuous) is the limit of the sequence of integrals, which is (2.29) in this case. □

We can easily extend this in a useful way – the direction of monotonicity is reversed really just to mentally distinquish this from the preceding lemma.

Lemma 2.5. *If $v_n \in \mathcal{C}_c(\mathbb{R})$ is any increasing sequence such that $\lim_{n\to\infty} v_n(x) \ge 0$ for each $x \in \mathbb{R}$ (where the possibility $v_n(x) \to \infty$ is included) then*

$$\lim_{n\to\infty} \int v_n dx \ge 0 \text{ including possibly } +\infty. \tag{2.30}$$

Proof. This is really a corollary of the preceding lemma. Consider the sequence of functions

$$w_n(x) = \begin{cases} 0 & \text{if } v_n(x) \ge 0 \\ -v_n(x) & \text{if } v_n(x) < 0. \end{cases} \tag{2.31}$$

Since this is the maximum of two continuous functions, namely $-v_n$ and 0, it is continuous and it vanishes for large x, so $w_n \in \mathcal{C}_{\rm c}(\mathbb{R})$. Since $v_n(x)$ is increasing, w_n is decreasing and it follows that $\lim w_n(x) = 0$ for all x – either it gets there for some finite n and then stays 0 or the limit of $v_n(x)$ is zero. Thus Lemma 2.4 applies to w_n so

$$\lim_{n\to\infty} \int_{\mathbb{R}} w_n(x) dx = 0.$$

Now, $v_n(x) \ge -w_n(x)$ for all x, so for each n, $\int v_n \ge -\int w_n$. From properties of the Riemann integral, $v_{n+1} \ge v_n$ implies that $\int v_n dx$ is an increasing sequence and it is bounded below by one that converges to 0, so (2.30) is the only possibility. □

From this result applied carefully we see that the integral behaves sensibly for absolutely summable series.

Lemma 2.6. *Suppose $u_n \in \mathcal{C}_c(\mathbb{R})$ is an absolutely summable series of real-valued functions, so $\sum\limits_n \int |u_n| dx < \infty$, and also suppose that*

$$\sum_n u_n(x) = 0 \;\; a.e. \tag{2.32}$$

then

$$\sum_n \int u_n dx = 0. \tag{2.33}$$

Proof. As already noted, the series (2.33) does converge, since the inequality $|\int u_n dx| \le \int |u_n| dx$ shows that it is absolutely convergent (hence Cauchy, hence convergent).

If E is a set of measure zero such that (2.32) holds on the complement then we can modify u_n as in (2.14) by adding and subtracting a non-negative absolutely summable sequence v_k which diverges absolutely on E. For the new sequence u_n (2.32) is strengthened to

$$\sum_n |u_n(x)| < \infty \Longrightarrow \sum_n u_n(x) = 0 \tag{2.34}$$

and the conclusion (2.33) holds for the new sequence if and only if it holds for the old one.

Now, we need to get ourselves into a position to apply Lemma 2.5. To do this, just choose some integer N (large but it doesn't matter yet) and consider the sequence of functions – it depends on N but I will suppress this dependence –

$$U_1(x) = \sum_{n=1}^{N+1} u_n(x), \;\; U_j(x) = |u_{N+j}(x)|, \;\; j \ge 2. \tag{2.35}$$

This is a sequence in $\mathcal{C}_c(\mathbb{R})$ and it is absolutely summable – the convergence of $\sum\limits_j \int |U_j| dx$ only depends on the 'tail' which is the same as for u_n. For the same reason,

$$\sum_j |U_j(x)| < \infty \Longleftrightarrow \sum_n |u_n(x)| < \infty. \tag{2.36}$$

Now the sequence of partial sums

$$g_p(x) = \sum_{j=1}^{p} U_j(x) = \sum_{n=1}^{N+1} u_n(x) + \sum_{j=2}^{p} |u_{N+j}| \tag{2.37}$$

is increasing with p – since we are adding non-negative functions. If the two equivalent conditions in (2.36) hold then

$$\sum_n u_n(x) = 0 \Longrightarrow \sum_{n=1}^{N+1} u_n(x) + \sum_{j=2}^{\infty} |u_{N+j}(x)| \ge 0 \Longrightarrow \lim_{p\to\infty} g_p(x) \ge 0, \tag{2.38}$$

since we are only increasing each term. On the other hand if these conditions do not hold then the tail, any tail, sums to infinity so

$$\lim_{p\to\infty} g_p(x) = \infty. \tag{2.39}$$

Thus the conditions of Lemma 2.5 hold for g_p and hence

$$\sum_{n=1}^{N+1}\int u_n + \sum_{j\ge N+2}\int |u_j(x)|dx \ge 0. \tag{2.40}$$

Using the same inequality as before this implies that

$$\sum_{n=1}^{\infty}\int u_n \ge -2\sum_{j\ge N+2}\int |u_j(x)|dx. \tag{2.41}$$

This is true for any N and as $N\to\infty$, $\lim_{N\to\infty}\sum\limits_{j\ge N+2}\int|u_j(x)|dx = 0$. So the fixed number on the left in (2.41), which is what we are interested in, must be non-negative.

In fact the signs in the argument can be reversed, considering instead

$$h_1(x) = -\sum_{n=1}^{N+1} u_n(x),\ h_p(x) = |u_{N+p}(x)|,\ p\ge 2 \tag{2.42}$$

and the final conclusion is the opposite inequality in (2.41). That is, we conclude what we wanted to show, that

$$\sum_{n=1}^{\infty}\int u_n = 0. \tag{2.43}$$

□

Finally then we are in a position to show that the integral of an element of $\mathcal{L}^1(\mathbb{R})$ is well-defined.

Proposition 2.3. *If $f\in\mathcal{L}^1(\mathbb{R})$ then*

$$\int f = \lim_{n\to\infty}\sum_n\int u_n \tag{2.44}$$

is independent of the approximating sequence, u_n, used to define it. Moreover,

$$\begin{gathered}\int|f| = \lim_{N\to\infty}\int|\sum_{k=1}^N u_k|,\\ |\int f|\le\int|f| \text{ and}\\ \lim_{n\to\infty}\int|f-\sum_{j=1}^n u_j| = 0.\end{gathered} \tag{2.45}$$

So in some sense the definition of the Lebesgue integral 'involves no cancellations'. There are various extensions of the integral which do exploit cancellations – I invite you to look into the definition of the Henstock integral (and its relatives).

Proof. The uniqueness of $\int f$ follows from Lemma 2.6. Namely, if u_n and u'_n are two series approximating f as in Proposition 2.1 then the real and imaginary parts of the difference $u'_n - u_n$ satisfy the hypothesis of Lemma 2.6 so it follows that

$$\sum_n\int u_n = \sum_n\int u'_n.$$

Then the first part of (2.45) follows from this definition of the integral applied to $|f|$ and the approximating series for $|f|$ devised in the proof of Proposition 2.2. The inequality

$$|\sum_n \int u_n| \le \sum_n \int |u_n|, \tag{2.46}$$

which follows from the finite inequalities for the Riemann integrals

$$|\sum_{n\le N} \int u_n| \le \sum_{n\le N} \int |u_n| \le \sum_n \int |u_n|$$

gives the second part.

The final part follows by applying the same arguments to the series $\{u_k\}_{k>n}$, as an absolutely summable series approximating $f - \sum\limits_{j=1}^{n} u_j$ and observing that the integral is bounded by

$$\int |f - \sum_{k=1}^{n} u_k| \le \sum_{k=n+1}^{\infty} \int |u_k| \to 0 \text{ as } n \to \infty. \tag{2.47}$$

□

4. Summable series in $\mathcal{L}^1(\mathbb{R})$

The next thing we want to know is when the 'norm', which is in fact only a seminorm, on $\mathcal{L}^1(\mathbb{R})$, vanishes. That is, when does $\int |f| = 0$? One way is fairly easy. The full result we are after is:-

Proposition 2.4. *For an integrable function $f \in \mathcal{L}^1(\mathbb{R})$, the vanishing of $\int |f|$ implies that f is a null function in the sense that*

$$f(x) = 0 \;\forall\; x \in \mathbb{R} \setminus E \text{ where } E \text{ is of measure zero.} \tag{2.48}$$

Conversely, if (2.48) *holds then $f \in \mathcal{L}^1(\mathbb{R})$ and $\int |f| = 0$.*

Proof. The main part of this is the first part, that the vanishing of $\int |f|$ implies that f is null. The converse is the easier direction in the sense that we have already done it.

Namely, if f is null in the sense of (2.48) then $|f|$ is the limit a.e. of the absolutely summable series with all terms 0. It follows from the definition of the integral above that $|f| \in \mathcal{L}^1(\mathbb{R})$ and $\int |f| = 0$. □

For the forward argument we will use the following more technical result, which is also closely related to the completeness of $L^1(\mathbb{R})$ (note the small notational difference, L^1 is the Banach space which is the quotient by the null functions, see below).

Proposition 2.5. *If $f_n \in \mathcal{L}^1(\mathbb{R})$ is an absolutely summable series, meaning that $\sum\limits_n \int |f_n| < \infty$, then*

$$E = \{x \in \mathbb{R}; \sum_n |f_n(x)| = \infty\} \text{ has measure zero.} \tag{2.49}$$

If $f : \mathbb{R} \longrightarrow \mathbb{C}$ satisfies

$$f(x) = \sum_n f_n(x) \text{ a.e.} \tag{2.50}$$

then $f \in \mathcal{L}^1(\mathbb{R})$,

$$\int f = \sum_n \int f_n,$$
$$|\int f| \le \int |f| = \lim_{n\to\infty} \int |\sum_{j=1}^n f_j| \le \sum_j \int |f_j| \textit{ and} \tag{2.51}$$
$$\lim_{n\to\infty} \int |f - \sum_{j=1}^n f_j| = 0.$$

This basically says we can replace 'continuous function of compact support' by 'Lebesgue integrable function' in the definition and get the same result. Of course this makes no sense without the original definition, so what we are showing is that iterating it makes no difference – we do not get a bigger space.

Proof. The proof is very like the proof of completeness via absolutely summable series for a normed space outlined in the preceding chapter.

By assumption each $f_n \in \mathcal{L}^1(\mathbb{R})$, so there exists a sequence $u_{n,j} \in \mathcal{C}_{\mathrm{c}}(\mathbb{R})$ with $\sum\limits_j \int |u_{n,j}| < \infty$ and

$$\sum_j |u_{n,j}(x)| < \infty \Longrightarrow f_n(x) = \sum_j u_{n,j}(x). \tag{2.52}$$

We might hope that $f(x)$ is given by the sum of the $u_{n,j}(x)$ over both n and j, but in general, this double series is not absolutely summable. However we can replace it by one that is. For each n choose N_n so that

$$\sum_{j>N_n} \int |u_{n,j}| < 2^{-n}. \tag{2.53}$$

This is possible by the assumed absolute summability – the tail of the series therefore being small. Having done this, we replace the series $u_{n,j}$ by

$$u'_{n,1} = \sum_{j \le N_n} u_{n,j}(x),\ u'_{n,j}(x) = u_{n,N_n+j-1}(x)\ \forall\ j \ge 2, \tag{2.54}$$

summing the first N_n terms. This still sums to f_n on the same set as in (2.52). So in fact we can simply replace $u_{n,j}$ by $u'_{n,j}$ and we have in addition the estimate

$$\sum_j \int |u'_{n,j}| \le \int |f_n| + 2^{-n+1}\ \forall\ n. \tag{2.55}$$

This follows from the triangle inequality since, using (2.53),

$$\int |u'_{n,1} + \sum_{j=2}^N u'_{n,j}| \ge \int |u'_{n,1}| - \sum_{j\ge 2} \int |u'_{n,j}| \ge \int |u'_{n,1}| - 2^{-n} \tag{2.56}$$

and the left side converges to $\int |f_n|$ by (2.45) as $N \to \infty$. Using (2.53) again gives (2.55).

Dropping the primes from the notation and denoting the new series again as $u_{n,j}$ we can let v_k be some enumeration of the $u_{n,j}$ – using the standard diagonalization procedure for instance. This gives a new series of continuous functions of compact support which is absolutely summable since

$$\sum_{k=1}^{N}\int|v_k|\le\sum_{n,j}\int|u_{n,j}|\le\sum_{n}(\int|f_n|+2^{-n+1})<\infty. \tag{2.57}$$

Using the freedom to rearrange absolutely convergent series we see that

$$\sum_{n,j}|u_{n,j}(x)|<\infty\Longrightarrow f(x)=\sum_{k}v_k(x)=\sum_{n}\sum_{j}u_{n,j}(x)=\sum_{n}f_n(x). \tag{2.58}$$

The set where (2.58) fails is a set of measure zero, by definition. Thus $f\in\mathcal{L}^1(\mathbb{R})$ and (2.49) also follows. To get the final result (2.51), rearrange the double series for the integral (which is also absolutely convergent). □

For the moment we only need the weakest part, (2.49), of this. To paraphrase this, for any absolutely summable series of integrable functions the absolute pointwise series converges off a set of measure zero – it can only diverge on a set of measure zero. It is rather shocking but this allows us to prove the rest of Proposition 2.4! Namely, suppose $f\in\mathcal{L}^1(\mathbb{R})$ and $\int|f|=0$. Then Proposition 2.5 applies to the series with each term being $|f|$. This is absolutely summable since all the integrals are zero. So it must converge pointwise except on a set of measure zero. Clearly it diverges whenever $f(x)\neq0$,

$$\int|f|=0\Longrightarrow\{x;f(x)\neq0\}\text{ has measure zero} \tag{2.59}$$

which is what we wanted to show to finally complete the proof of Proposition 2.4.

5. The space $L^1(\mathbb{R})$

At this point we are able to define the standard Lebesgue space

$$L^1(\mathbb{R})=\mathcal{L}^1(\mathbb{R})/\mathcal{N},\ \mathcal{N}=\{\text{null functions}\} \tag{2.60}$$

and to check that it is a Banach space with the norm (arising from, to be pedantic) $\int|f|$.

Theorem 2.1. *The quotient space $L^1(\mathbb{R})$ defined by* (2.60) *is a Banach space in which the continuous functions of compact support form a dense subspace.*

The elements of $L^1(\mathbb{R})$ are *equivalence classes of functions*

$$[f]=f+\mathcal{N},\ f\in\mathcal{L}^1(\mathbb{R}). \tag{2.61}$$

That is, we 'identify' two elements of $\mathcal{L}^1(\mathbb{R})$ if (and only if) their difference is null, which is to say they are equal off a set of measure zero. Note that the set which is ignored here is not fixed, but can depend on the functions.

Proof. For an element of $L^1(\mathbb{R})$ the integral of the absolute value is well-defined by Propositions 2.2 and 2.4

$$\|[f]\|_{L^1}=\int|f|,\ f\in[f] \tag{2.62}$$

and gives a *semi-norm* on $\mathcal{L}^1(\mathbb{R})$. It follows from Proposition 1.5 that on the quotient, $\|[f]\|$ is indeed a norm.

The completeness of $L^1(\mathbb{R})$ is a direct consequence of Proposition 2.5. Namely, to show a normed space is complete it is enough to check that any absolutely summable series converges. If $[f_j]$ is an absolutely summable series in $L^1(\mathbb{R})$ then f_j is absolutely summable in $\mathcal{L}^1(\mathbb{R})$ and by Proposition 2.5 the sum of the series exists so we can use (2.50) to define f off the set E and take it to be zero on E. Then, $f \in \mathcal{L}^1(\mathbb{R})$ and the last part of (2.51) means precisely that

$$\lim_{n\to\infty} \|[f] - \sum_{j<n}[f_j]\|_{L^1} = \lim_{n\to\infty} \int |f - \sum_{j<n} f_j| = 0 \tag{2.63}$$

showing the desired completeness. □

Note that despite the fact that it is technically incorrect, everyone says '$L^1(\mathbb{R})$ is the space of Lebesgue integrable functions' even though it is really the space of equivalence classes of these functions modulo equality almost everywhere. Not much harm can come from this mild abuse of language.

Another consequence of Proposition 2.5 and the proof above is an extension of Lemma 2.3.

Proposition 2.6. *Any countable union of sets of measure zero is a set of measure zero.*

Proof. If E is a set of measure zero then any function f which is defined on $\mathbb{R}$ and vanishes outside E is a null function – is in $\mathcal{L}^1(\mathbb{R})$ and has $\int |f| = 0$. Conversely if the characteristic function of E, the function equal to 1 on E and zero in $\mathbb{R} \setminus E$ is integrable and has integral zero then E has measure zero. This is the characterization of null functions above. Now, if E_j is a sequence of sets of measure zero and χ_k is the characteristic function of

$$\bigcup_{j\le k} E_j \tag{2.64}$$

then $\int |\chi_k| = 0$ so this is an absolutely summable series with sum, the characteristic function of the union, integrable and of integral zero. □

6. The three integration theorems

Even though we now 'know' which functions are Lebesgue integrable, it is often quite tricky to use the definitions to actually show that a particular function has this property. There are three standard results on convergence of sequences of integrable functions which are powerful enough to cover most situations that arise in practice – a Monotonicity Lemma, Fatou's Lemma and Lebesgue's Dominated Convergence theorem.

Lemma 2.7 (Montonicity). *If $f_j \in \mathcal{L}^1(\mathbb{R})$ is a monotone sequence, either $f_j(x) \ge f_{j+1}(x)$ for all $x \in \mathbb{R}$ and all j or $f_j(x) \le f_{j+1}(x)$ for all $x \in \mathbb{R}$ and all j, and $\int f_j$ is bounded then*

$$\{x \in \mathbb{R}; \lim_{j\to\infty} f_j(x) \textit{ is finite}\} = \mathbb{R} \setminus E \tag{2.65}$$

where E has measure zero and

$$\begin{gathered} f = \lim_{j\to\infty} f_j(x) \textit{ a.e. is an element of } \mathcal{L}^1(\mathbb{R}) \\ \textit{with} \int f = \lim_{j\to\infty} \int f_j \textit{ and } \lim_{j\to\infty} \int |f - f_j| = 0. \end{gathered} \tag{2.66}$$

In the usual approach through measure one has the concept of a measureable, non-negative, function for which the integral 'exists but is infinite' – we do not have this (but we could easily do it, or rather you could). Using this one can drop the assumption about the finiteness of the integral but the result is not significantly stronger.

PROOF. Since we can change the sign of the f_i it suffices to assume that the f_i are monotonically increasing. The sequence of integrals is therefore also montonic increasing and, being bounded, converges. Turning the sequence into a series, by setting $g_1 = f_1$ and $g_j = f_j - f_{j-1}$ for $j \geq 2$ the g_j are non-negative for $j \geq 1$ and

$$\sum_{j\geq 2} \int |g_j| = \sum_{j\geq 2} \int g_j = \lim_{n\to\infty} \int f_n - \int f_1 \tag{2.67}$$

converges. So this is indeed an absolutely summable series. We therefore know from Proposition 2.5 that it converges absolutely a.e., that the limit, f, is integrable and that

$$\int f = \sum_j \int g_j = \lim_{n\to\infty} \int f_j. \tag{2.68}$$

The second part, corresponding to convergence for the equivalence classes in $L^1(\mathbb{R})$ follows from the fact established earlier about $|f|$ but here it also follows from the monotonicity since $f(x) \geq f_j(x)$ a.e. so

$$\int |f - f_j| = \int f - \int f_j \to 0 \text{ as } j \to \infty. \tag{2.69}$$

□

Now, to Fatou's Lemma. This really just takes the monotonicity result and applies it to a sequence of integrable functions with bounded integral. You should recall that the max and min of two real-valued integrable functions is integrable and that

$$\int \min(f, g) \leq \min(\int f, \int g). \tag{2.70}$$

This follows from the identities

$$2\max(f, g) = |f - g| + f + g,\ 2\min(f, g) = -|f - g| + f + g. \tag{2.71}$$

LEMMA 2.8 (Fatou). *Let $f_j \in \mathcal{L}^1(\mathbb{R})$ be a sequence of real-valued non-negative integrable functions such that $\int f_j$ is bounded then*

$$\begin{gathered} f(x) = \liminf_{n\to\infty} f_n(x) \text{ exists a.e., } f \in \mathcal{L}^1(\mathbb{R}) \text{ and} \\ \int \liminf f_n = \int f \leq \liminf \int f_n. \end{gathered} \tag{2.72}$$

PROOF. You should remind yourself of the properties of lim inf as necessary! Fix k and consider

$$F_{k,n} = \min_{k\leq p\leq k+n} f_p(x) \in \mathcal{L}^1(\mathbb{R}). \tag{2.73}$$

As discussed above this is integrable. Moreover, this is a decreasing sequence, as n increases, because the minimum is over an increasing set of functions. The $F_{k,n}$ are non-negative so Lemma 2.7 applies and shows that

$$g_k(x) = \inf_{p\ge k} f_p(x) \in \mathcal{L}^1(\mathbb{R}),\ \int g_k \le \int f_n\ \forall\ n \ge k. \tag{2.74}$$

Note that for a decreasing sequence of non-negative numbers the limit exists and is indeed the infimum. Thus in fact,

$$\int g_k \le \liminf \int f_n\ \forall\ k. \tag{2.75}$$

Now, let k vary. Then, the infimum in (2.74) is over a set which decreases as k increases. Thus the $g_k(x)$ are increasing. The integrals of this sequence are bounded above in view of (2.75) since we assumed a bound on the $\int f_n$'s. So, we can apply the monotonicity result again to see that

$$\begin{gathered} f(x) = \lim_{k\to\infty} g_k(x) \text{ exists a.e and } f \in \mathcal{L}^1(\mathbb{R}) \text{ has} \\ \int f \le \liminf \int f_n. \end{gathered} \tag{2.76}$$

Since $f(x) = \liminf f_n(x)$, by definition of the latter, we have proved the Lemma. □

Now, we apply Fatou's Lemma to prove what we are really after:-

Theorem 2.2 (Dominated convergence). *Suppose $f_j \in \mathcal{L}^1(\mathbb{R})$ is a sequence of integrable functions such that*

$$\begin{gathered} \exists\ h \in \mathcal{L}^1(\mathbb{R}) \text{ with } |f_j(x)| \le h(x) \text{ a.e. and} \\ f(x) = \lim_{j\to\infty} f_j(x) \text{ exists a.e.} \end{gathered} \tag{2.77}$$

then $f \in \mathcal{L}^1(\mathbb{R})$ and $[f_j] \to [f]$ in $L^1(\mathbb{R})$, so $\int f = \lim_{n\to\infty} \int f_n$ (including the assertion that this limit exists).

Proof. First, we can assume that the f_j are real since the hypotheses hold for the real and imaginary parts of the sequence and together give the desired result. Moreover, we can change all the f_j's to make them zero on the set on which the initial estimate in (2.77) does not hold. Then this bound on the f_j's becomes

$$-h(x) \le f_j(x) \le h(x)\ \forall\ x \in \mathbb{R}. \tag{2.78}$$

In particular this means that $g_j = h - f_j$ is a non-negative sequence of integrable functions and the sequence of integrals is also bounded, since (2.77) also implies that $\int |f_j| \le \int h$, so $\int g_j \le 2\int h$. Thus Fatou's Lemma applies to the g_j. Since we have *assumed* that the sequence $g_j(x)$ converges a.e. to $h - f$ we know that

$$\begin{gathered} h - f(x) = \liminf g_j(x) \text{ a.e. and} \\ \int h - \int f \le \liminf \int (h - f_j) = \int h - \limsup \int f_j. \end{gathered} \tag{2.79}$$

Notice the change on the right from liminf to limsup because of the sign.

Now we can apply the same argument to $g_j'(x) = h(x) + f_j(x)$ since this is also non-negative and has integrals bounded above. This converges a.e. to $h(x) + f(x)$ so this time we conclude that

$$\int h + \int f \le \liminf \int (h + f_j) = \int h + \liminf \int f_j. \tag{2.80}$$

In both inequalities (2.79) and (2.80) we can cancel an $\int h$ and combining them we find

$$\limsup \int f_j \le \int f \le \liminf \int f_j. \tag{2.81}$$

In particular the limsup on the left is smaller than, or equal to, the liminf on the right, for the same real sequence. This however implies that they are equal and that the sequence $\int f_j$ converges. Thus indeed

$$\int f = \lim_{n\to\infty} \int f_n. \tag{2.82}$$

Convergence of f_n to f in $L^1(\mathbb{R})$ follows by applying the results proved so far to $|f - f_n|$, converging almost everywhere to 0. In this case (2.82) becomes

$$\lim_{n\to\infty} \int |f - f_n| = 0.$$

□

Generally in applications it is Lebesgue's dominated convergence which is used to prove that some function is integrable. Of course, since we deduced it from Fatou's lemma, and the latter from the Monotonicity lemma, you might say that Lebesgue's theorem is the weakest of the three! However, it is very handy and often a combination does the trick. For instance

LEMMA 2.9. *A continuous function $u \in \mathcal{C}(\mathbb{R})$ is Lebesgue integrable if and only if the 'improper Riemann integral'*

$$\lim_{R\to\infty} \int_{-R}^{R} |u(x)| dx < \infty. \tag{2.83}$$

Note that the 'improper integral' without the absolute value can converge without u being Lebesgue integrable.

PROOF. If (2.83) holds then consider the sequence of functions $v_N = \chi_{[-N,N]}|u|$, which we know to be in $L^1(\mathbb{R})$ by Lemma 2.2. This is monotonic increasing with limit $|u|$, so the Monotonicity Lemma shows that $|u| \in L^1(\mathbb{R})$. Then consider $w_N = \chi_{[-N,N]}u$ which we also know to be in $L^1(\mathbb{R})$. Since it is bounded by $|u|$ and converges pointwise to u, it follows from Dominated Convergence that $u \in L^1(\mathbb{R})$. Conversely, if $u \in L^1(\mathbb{R})$ then $|u| \in L^1(\mathbb{R})$ and $\chi_{[-N,N]}|u| \in L^1(\mathbb{R})$ converges to $|u|$ so by Dominated Convergence (2.83) must hold. □

So (2.83) holds for any $u \in L^1(\mathbb{R})$.

7. Notions of convergence

We have been dealing with two basic notions of convergence, but really there are more. Let us pause to clarify the relationships between these different concepts.

(1) Convergence of a sequence in $L^1(\mathbb{R})$ (or by slight abuse of language in $\mathcal{L}^1(\mathbb{R})$) – f and $f_n \in\ L^1(\mathbb{R})$ and

$$\|f - f_n\|_{L^1} \to 0 \text{ as } n \to \infty. \tag{2.84}$$

(2) Convergence almost everywhere:- For some sequence of functions f_n and function f,

$$f_n(x) \to f(x) \text{ as } n \to \infty \text{ for } x \in \mathbb{R} \setminus E \tag{2.85}$$

where $E \subset \mathbb{R}$ is of measure zero.

(3) Dominated convergence:- For $f_j \in L^1(\mathbb{R})$ (or representatives in $\mathcal{L}^1(\mathbb{R})$) such that $|f_j| \le F$ (a.e.) for some $F \in L^1(\mathbb{R})$ and (2.85) holds.

(4) What we might call 'absolutely summable convergence'. Thus $f_n \in L^1(\mathbb{R})$ are such that $f_n = \sum\limits_{j=1}^{n} g_j$ where $g_j \in L^1(\mathbb{R})$ and $\sum\limits_j \int |g_j| < \infty$. Then (2.85) holds for some f.

(5) Monotone convergence. For $f_j \in \mathcal{L}^1(\mathbb{R})$, real valued and montonic, we require that $\int f_j$ is bounded and it then follows that $f_j \to f$ almost everywhere, with $f \in \mathcal{L}^1(\mathbb{R})$ and that the convergence is $\mathcal{L}^1$ and also that $\int f = \lim_j \int f_j$.

So, one important point to know is that 1 does not imply 2. Nor conversely does 2 imply 1 even if we assume that all the f_j and f are in $L^1(\mathbb{R})$.

However, montone convergence implies dominated convergence. Namely if f is the limit then $|f_j| \le |f| + 2|f_1|$ and $f_j \to f$ almost everywhere. Also, monotone convergence implies convergence with absolute summability simply by taking the sequence to have first term f_1 and subsequence terms $f_j - f_{j-1}$ (assuming that f_j is monotonically increasing) one gets an absolutely summable series with sequence of finite sums converging to f. Similarly absolutely summable convergence implies dominated convergence for the sequence of partial sums; by montone convergence the series $\sum\limits_n |f_n(x)|$ converges a.e. and in L^1 to some function F which dominates the partial sums which in turn converge pointwise. I suggest that you make a diagram with these implications in it so that you are clear about the relationships between them.

8. The space $L^2(\mathbb{R})$

So far we have discussed the Banach space $L^1(\mathbb{R})$. The real aim is to get a good hold on the (Hilbert) space $L^2(\mathbb{R})$. This can be approached in several ways. We could start off as for $L^1(\mathbb{R})$ and define $L^2(\mathbb{R})$ as the completion of $\mathcal{C}_{\rm c}(\mathbb{R})$ with respect to the norm

$$\|f\|_{L^2} = \left(\int |f|^2\right)^{\frac{1}{2}}. \tag{2.86}$$

This would be rather repetitious; instead we adopt an approach based on Dominated Convergence. You might think, by the way, that it is enough just to ask that $|f|^2 \in \mathcal{L}^1(\mathbb{R})$. This does not work, since even if real the sign of f could jump around and make it non-integrable (provided you believe in the axiom of choice).

Nor would this approach work for $L^1(\mathbb{R})$ since $|f| \in L^1(\mathbb{R})$ does not imply that $f \in L^1(\mathbb{R})$.

Definition 2.4. A function $f : \mathbb{R} \longrightarrow \mathbb{C}$ is said to be 'Lebesgue square integrable', written $f \in \mathcal{L}^2(\mathbb{R})$, if there exists a sequence $u_n \in \mathcal{C}_{\mathrm{c}}(\mathbb{R})$ such that

$$u_n(x) \to f(x) \text{ a.e. and } |u_n(x)|^2 \le F(x) \text{ a.e. for some } F \in \mathcal{L}^1(\mathbb{R}). \tag{2.87}$$

Proposition 2.7. *The space $\mathcal{L}^2(\mathbb{R})$ is linear, $f \in \mathcal{L}^2(\mathbb{R})$ implies $|f|^2 \in \mathcal{L}^1(\mathbb{R})$ and* (2.86) *defines a seminorm on $\mathcal{L}^2(\mathbb{R})$ which vanishes precisely on the null functions $\mathcal{N} \subset \mathcal{L}^2(\mathbb{R})$.*

Definition 2.5. We define $L^2(\mathbb{R}) = \mathcal{L}(\mathbb{R})/\mathcal{N}$.

So we know that $L^2(\mathbb{R})$ is a normed space. It is in fact complete and much more!

Proof. First to see the linearity of $\mathcal{L}^2(\mathbb{R})$ note that if $f \in \mathcal{L}^2(\mathbb{R})$ and $c \in \mathbb{C}$ then $cf \in \mathcal{L}^2(\mathbb{R})$ since if u_n is a sequence as in the definition for f then cu_n is such a sequence for cf.

Similarly if $f,\ g \in \mathcal{L}^2(\mathbb{R})$ with sequences u_n and v_n then $w_n = u_n + v_n$ has the first property – since we know that the union of two sets of measure zero is a set of measure zero and the second follows from the estimate

$$|w_n(x)|^2 = |u_n(x) + v_n(x)|^2 \le 2|u_n(x)|^2 + 2|v_n(x)|^2 \le 2(F+G)(x) \tag{2.88}$$

where $|u_n(x)|^2 \le F(x)$ and $|v_n(x)|^2 \le G(x)$ with $F,\ G \in \mathcal{L}^1(\mathbb{R})$.

Moreover, if $f \in \mathcal{L}^2(\mathbb{R})$ then the sequence $|u_n(x)|^2$ converges pointwise almost everywhere to $|f(x)|^2$ so by Lebesgue's Dominated Convergence, $|f|^2 \in \mathcal{L}^1(\mathbb{R})$. Thus $\|f\|_{L^2}$ is well-defined. It vanishes if and only if $|f|^2 \in \mathcal{N}$ but this is equivalent to $f \in \mathcal{N}$ – conversely $\mathcal{N} \subset \mathcal{L}^2(\mathbb{R})$ since the zero sequence works in the definition above.

So we only need to check the triangle inquality, absolute homogeneity being clear, to deduce that $L^2 = \mathcal{L}^2/\mathcal{N}$ is at least a normed space. In fact we checked this earlier on $\mathcal{C}_{\mathrm{c}}(\mathbb{R})$ and the general case follows by continuity:-

$$\begin{aligned} \|u_n + v_n\|_{L^2} \le \|u_n\|_{L^2} + \|v_n\|_{L^2}\ \forall\ n \Longrightarrow \\ \|f+g\|_{L^2} = \lim_{n\to\infty} \|u_n + v_n\|_{L^2} \le \|f\|_{L^2} + \|g\|_{L^2}. \end{aligned} \tag{2.89}$$

□

We will get a direct proof of the triangle inequality as soon as we start talking about (pre-Hilbert) spaces.

So it only remains to check the completeness of $L^2(\mathbb{R})$, which is really the whole point of the discussion of Lebesgue integration.

Theorem 2.3. *The space $L^2(\mathbb{R})$ is complete with respect to $\|\cdot\|_{L^2}$ and is a completion of $\mathcal{C}_{\mathrm{c}}(\mathbb{R})$ with respect to this norm.*

Proof. That $\mathcal{C}_{\mathrm{c}}(\mathbb{R}) \subset L^2(\mathbb{R})$ follows directly from the definition and the fact that a continuous null function must vanish. This is a dense subset since, if $f \in \mathcal{L}^2(\mathbb{R})$ a sequence $u_n \in \mathcal{C}_{\mathrm{c}}(\mathbb{R})$ as in Definition 2.4 satisfies

$$|u_n(x) - u_m(x)|^2 \le 4F(x)\ \forall\ n,\ m, \tag{2.90}$$

and converges almost everwhere to $|f(x)-u_m(x)|^2$ as $n \to \infty$. Thus, by Dominated Convergence, $|f(x)-u_m(x)|^2 \in \mathcal{L}^1(\mathbb{R})$. As $m \to \infty$, $|f(x)-u_m(x)|^2 \to 0$ almost everywhere and $|f(x)-u_m(x)|^2 \le 4F(x)$ so again by dominated convergence

$$\|f-u_m\|_{L^2} = \left(\|(|f-u_m|^2)\|_{L^1})\right)^{\frac{1}{2}} \to 0. \tag{2.91}$$

This shows the density of $\mathcal{C}_{\mathrm{c}}(\mathbb{R})$ in $L^2(\mathbb{R})$, the quotient by the null functions.

To prove completeness, we only need show that any absolutely L^2-summable sequence in $\mathcal{C}_{\mathrm{c}}(\mathbb{R})$ converges in L^2 and the general case follows by density. So, suppose $\phi_n \in \mathcal{C}_{\mathrm{c}}(\mathbb{R})$ is such a sequence:

$$\sum_n \|\phi_n\|_{L^2} < \infty.$$

Consider $F_k(x) = \left(\sum\limits_{n\le k} |\phi_k(x)|\right)^2$. This is an increasing sequence in $\mathcal{C}_{\mathrm{c}}(\mathbb{R})$ and its L^1 norm is bounded:

$$\|F_k\|_{L^1} = \|\sum_{n\le k} |\phi_n|\|^2_{L^2} \le \left(\sum_{n\le k} \|\phi_n\|_{L^2}\right)^2 \le C^2 < \infty \tag{2.92}$$

using the triangle inequality and absolutely L^2 summability. Thus, by Monotone Convergence, $F_k(x) \to F(x)$ a.e., $F_k \to F \in \mathcal{L}^1(\mathbb{R})$ and $F_k(x) \le F(x)$ a.e., where we define $F(x)$ to be the limit when this exists and zero otherwise.

Thus the sequence of partial sums $u_k(x) = \sum\limits_{n\le k} \phi_n(x)$ converges almost everywhere – since it converges (absoliutely) on the set where F_k is bounded. Let $f(x)$ be the limit. We want to show that $f \in \mathcal{L}^2(\mathbb{R})$ but this follows from the definition since

$$|u_k(x)|^2 \le (\sum_{n\le k} |\phi_n(x)|)^2 = F_k(x) \le F(x) \text{ a.e.} \tag{2.93}$$

As in (2.91) it follows that

$$\int |u_k(x) - f(x)|^2 \to 0. \tag{2.94}$$

As for the case of $L^1(\mathbb{R})$ it now follows that $L^2(\mathbb{R})$ is complete. □

We want to check that $L^2(\mathbb{R})$ is a Hilbert space (which I will define very soon, even though it is in the next Chapter); to do so observe that if $f,\ g \in \mathcal{L}^2(\mathbb{R})$ have approximating sequences u_n, v_n as in Definition 2.4, so $|u_n(x)|^2 \le F(x)$ and $|v_n(x)|^2 \le G(x)$ with $F,\ G \in \mathcal{L}^1(\mathbb{R})$ then

$$u_n(x)v_n(x) \to f(x)g(x) \text{ a.e. and } |u_n(x)v_n(x)| \le \frac{1}{2}(F(x)+G(x)) \tag{2.95}$$

shows that $fg \in \mathcal{L}^1(\mathbb{R})$ by Dominated Convergence. This leads to the basic property of the norm on a (pre)-Hilbert space – that it comes from an inner product. In this case

$$\langle f, g\rangle_{L^2} = \int f(x)\overline{g(x)},\ \|f\|_{L^2} = \langle f, f\rangle^{\frac{1}{2}}. \tag{2.96}$$

At this point I normally move on to the next chapter on Hilbert spaces with $L^2(\mathbb{R})$ as one motivating example.

9. Measurable and non-measurable sets

The σ-algebra of Lebesgue measurable sets on the line is discussed below but we can directly consider the notion of a set of finite Lebesgue measure. Namely such a set $A \subset \mathbb{R}$ is defined by the condition that the chactacteristic function

$$\chi_A(x) = \begin{cases} 1 & \text{if } x \in A \\ 0 & \text{if } x \notin A \end{cases} \tag{2.97}$$

is Lebesgue integrable, $\chi_A \in \mathcal{L}^1(\mathbb{R})$. The measure of the set (think 'length') is then $\mu(A) = \int_{\mathbb{R}} \chi_A$ the properties of which are discussed below. Certainly if $A \subset [-R, R]$ has finite measure then $\mu(A) \le 2R$ from the properties of the integral. Similalry if $A_i \subset [-R, R]$ are a sequence of sets of finite measure which are disjoint, $A_i \cap A_j = \emptyset$, $i \ne j$, then

$$A = \bigsqcup_i A_i \text{ has finite measure and } \mu(A) = \sum_i \mu(A_i) \tag{2.98}$$

using Monotone Convergence.

Now the question arises, enquiring minds want to know after all:- Are there bounded sets which are not of finite measure? Similarly, are there functions of bounded support which are not integrable? It turns out this question gets us into somewhat deep water, but it is important to understand some of the limitiations that the insistence on precision in Mathematics places on its practitioners!

Let me present a standard construction of a non-(Lebesgue-)measurable subset of $[0, 1]$ and then comment on the issues that it raises. We start with the quotient space and quotient map

$$q : \mathbb{R} \longrightarrow \mathbb{R}/\mathbb{Q},\ q(x) = \{y \in \mathbb{R}; y = x + r,\ r \in \mathbb{Q}\}. \tag{2.99}$$

This partitions $\mathbb{R}$ into disjoint subsets

$$\mathbb{R} = \bigsqcup_{\tau \in \mathbb{R}/\mathbb{Q}} q^{-1}(\tau). \tag{2.100}$$

Two of these sets intersect if and only if they have elements differing by a rational, and then they are the same.

Now, each of these sets $q^{-1}(\tau)$ intersects $[0, 1]$. This follows from the density of the rationals in the reals, since if $x \in q^{-1}(\tau)$ there exists $r \in \mathbb{Q}$ such that $|x - r| < \frac{1}{2}$ and then $x' = x + (-r + \frac{1}{2}) \in q^{-1}(\tau) \cap [0, 1]$. So we can 'localize' (2.100) to

$$[0, 1] = \bigsqcup_{\tau \in \mathbb{R}/\mathbb{Q}} L(\tau),\ L(\tau) = q^{-1}(\tau) \cap [0, 1] \tag{2.101}$$

where all the sets $L(\tau)$ are non-empty.

Definition 2.6. A *Vitali set,* $V \subset [0, 1]$, is a set which contains precisely one element from each of the $L(\tau)$.

Take such a set V and consider the translates of it by rationals in $[-1, 1]$,

$$V_r = \{y \in [-1, 2]; y = x + r,\ x \in V\},\ r \in \mathbb{Q},\ |r| \le 1. \tag{2.102}$$

For different r these are disjoint – since by construction no two distinct elements of V differ by a rational. The union of these sets however satisfies

$$[0, 1] \subset \bigsqcup_{r \in \mathbb{Q}, |r| \le 1} V_r \subset [-1, 2]. \tag{2.103}$$

Now, we can simply order the sets V_r into a sequence A_i by ordering the rationals in $[-1,1]$.

Suppose V is of finite Lebesgue measure. Then we know that all the V_r are of finite measure and $\mu(V_r) = \mu(V) = \mu(A_i)$ for all i, from the properties of the Lebesgue integral. This means that (2.98) applies, so we have the inequalities

$$\mu([0,1]) = 1 \le \sum_{i=1}^{\infty} \mu(V) \le 3 = \mu([-1,2]). \tag{2.104}$$

Clearly we have a problem! The only way the right-hand inequality can hold is if $\mu(V) = 0$, but then the left-hand inequality fails.

Our conclusion then is that V cannot be Lebesgue measurable! Or is it? Since we are careful people we trace back through the discussion and see (it took people a long, long, time to recognize this) more precisely:-

Proposition 2.8. *If a Vitali set, $V \subset [0,1]$ exists, containing precisely one element of each of the sets $L(\tau)$, then it is bounded and not of finite Lebesgue measure; its characteristic function is a non-negative function of bounded support which is not Lebesgue integrable.*

Okay, so what is the 'issue' here. It is that the existence of such a Vitali set requires the *Axiom of Choice*. There are lots of sets $L(\tau)$ so from the standard (Zermelo-Fraenkel) axions of set theory it does not follow that you can 'choose an element from each' to form a new set. That is a (slightly informal) version of the additional axiom. Now, it has been shown (namely by Gödel and Cohen) that the Axiom of Choice is independent of the Zermelo-Fraenkel Axioms. This does not mean consistency, it means conditional consistency. The Zermelo-Fraenkel axioms together with the Axiom of Choice are inconsistent if and only if the Zermelo-Fraenkel axioms on their own are inconsistent.

Conclusion: As a working Mathematician you are free to choose to believe in the Axiom of Choice or not. It will make your life easier if you do, but it is up to you. Note that if you do not admit the Axiom of Choice, it does not mean that all bounded real sets are measurable, in the sense that you can prove it. Rather it means that it is consistent to believe this (as shown by Solovay).

See also the discussion of the Hahn-Banach Theorem in Section 1.12.

10. Measurable functions

From our original definition of $\mathcal{L}^1(\mathbb{R})$ we know that $\mathcal{C}_{\mathrm{c}}(\mathbb{R})$ is dense in $L^1(\mathbb{R})$. We also know that elements of $\mathcal{C}_{\mathrm{c}}(\mathbb{R})$ can be approximated uniformly, and hence in $L^1(\mathbb{R})$ by step functions – finite linear combinations of the characteristic functions of intervals. It is usual in measure theory to consider a somewhat larger class of functions which contains the step functions:

Definition 2.7. A *simple* function on $\mathbb{R}$ is a finite linear combination (generally with complex coefficients) of characteristic functions of subsets of finite measure:

$$f = \sum_{j=1}^{N} c_j \chi(B_j),\ \chi(B_j) \in \mathcal{L}^1(\mathbb{R}),\ c_j \in \mathbb{C}. \tag{2.105}$$

The real and imaginary parts of a simple function are simple and the positive and negative parts of a real simple function are simple. Since step functions are simple, we know that simple functions are dense in $\mathcal{L}^1(\mathbb{R})$ and that if $0 \le F \in \mathcal{L}^1(\mathbb{R})$ then there exists a sequence of simple functions (take them to be a summable sequence of step functions) $f_n \ge 0$ such that $f_n \to F$ almost everywhere and $f_n \le G$ for some other $G \in \mathcal{L}^1(\mathbb{R})$.

We elevate a special case of the second notion of convergence above to a definition.

Definition 2.8. A function $f : \mathbb{R} \longrightarrow \mathbb{C}$ is *(Lebesgue) measurable* if it is the pointwise limit almost everywhere of a sequence of simple functions.

Lemma 2.10. *A function is Lebesgue measurable if and only if it is the pointwise limit, almost everywhere, of a sequence of continuous functions of compact support.*

Proof. Continuous functions of compact support are the uniform limits of step functions, so this condition certainly implies measurability in the sense of Definition 2.8. Conversely, suppose a function f is the limit almost everywhere of a squence u_n of simple functions. Each of these functions is integrable, so we can find $\phi_n \in \mathcal{C}_{\mathrm{c}}(\mathbb{R})$ such that $\|u_n - \phi_n\|_{L^1} < 2^{-n}$. Then the telescoped sequence $v_1 = u_1 - \phi_1$, $v_k = (u_k - \phi_k) - (u_{k-1} - \phi_{k-1})$, $k > 1$, is absolutely summable so $u_n - \phi_n \to 0$ almost everywhere, and hence $\phi_n \to f$ off a set of measure zero. □

So replacing 'simple functions' by continuous functions in Definition 2.8 makes no difference – and the same for approximation by elements of $\mathcal{L}^1(\mathbb{R})$.

The measurable functions form a linear space since if f and g are measurable and f_n, g_n are sequences of simple functions as required by the definition then $c_1 f_n(x) + c_2 f_2(x) \to c_1 f(x) + c_2 g(x)$ on the intersection of the sets where $f_n(x) \to f(x)$ and $g_n(x) \to g(x)$ which is the complement of a set of measure zero.

Now, from the discussion above, we know that each element of $\mathcal{L}^1(\mathbb{R})$ is measurable. Conversely:

Lemma 2.11. *A function $f : \mathbb{R} \longrightarrow \mathbb{C}$ is an element of $\mathcal{L}^1(\mathbb{R})$ if and only if it is measurable and there exists $F \in \mathcal{L}^1(\mathbb{R})$ such that $|f| \le F$ almost everywhere.*

Proof. If f is measurable there exists a sequence of simple functions f_n such that $f_n \to f$ almost everywhere. The real part, $\operatorname{Re} f$, is also measurable as the limit almost everywhere of $\operatorname{Re} f_n$ and from the hypothesis $|\operatorname{Re} f| \le F$. We know that there exists a sequence of simple functions g_n, $g_n \to F$ almost everywhere and $0 \le g_n \le G$ for another element $G \in \mathcal{L}^1(\mathbb{R})$. Then set

$$u_n(x) = \begin{cases} g_n(x) & \text{if } \operatorname{Re} f_n(x) > g_n(x) \\ \operatorname{Re} f_n(x) & \text{if } -g_n(x) \le \operatorname{Re} f_n(x) \le g_n(x) \\ -g_n(x) & \text{if } \operatorname{Re} f_n(x) < -g_n(x). \end{cases} \tag{2.106}$$

Thus $u_n = \max(v_n, -g_n)$ where $v_n = \min(\operatorname{Re} f_n, g_n)$ so u_n is simple and $u_n \to f$ almost everywhere. Since $|u_n| \le G$ it follows from Lebesgue Dominated Convergence that $\operatorname{Re} f \in \mathcal{L}^1(\mathbb{R})$. The same argument shows $\operatorname{Im} f = -\operatorname{Re}(if) \in \mathcal{L}^1(\mathbb{R})$ so $f \in \mathcal{L}^1(\mathbb{R})$ as claimed. □

11. The spaces $L^p(\mathbb{R})$

We use Lemma 2.11 as a model:

Definition 2.9. For $1 \leq p < \infty$ we set

$$\mathcal{L}^p(\mathbb{R}) = \{f : \mathbb{R} \longrightarrow \mathbb{C}; f \text{ is measurable and } |f|^p \in \mathcal{L}^1(\mathbb{R})\}. \tag{2.107}$$

For $p = \infty$ we set

$$\mathcal{L}^\infty(\mathbb{R}) = \{f : \mathbb{R} \longrightarrow \mathbb{C}; f \text{ measurable and } \exists\ C \text{ s.t. } |f(x)| \leq C \text{ a.e}\} \tag{2.108}$$

Observe that, in view of Lemma 2.10, the case $p = 2$ gives the same space as Definition 2.4.

Proposition 2.9. *For each* $1 \leq p < \infty$,

$$\|u\|_{L^p} = \left(\int |u|^p\right)^{\frac{1}{p}} \tag{2.109}$$

is a seminorm on the linear space $\mathcal{L}^p(\mathbb{R})$ *vanishing only on the null functions and making the quotient* $L^p(\mathbb{R}) = \mathcal{L}^p(\mathbb{R})/\mathcal{N}$ *into a Banach space.*

Proof. The real part of an element of $\mathcal{L}^p(\mathbb{R})$ is in $\mathcal{L}^p(\mathbb{R})$ since it is measurable and $|\operatorname{Re} f|^p \leq |f|^p$ so $|\operatorname{Re} f|^p \in \mathcal{L}^1(\mathbb{R})$. Similarly, $\mathcal{L}^p(\mathbb{R})$ is linear; it is clear that $cf \in \mathcal{L}^p(\mathbb{R})$ if $f \in \mathcal{L}^p(\mathbb{R})$ and $c \in \mathbb{C}$ and the sum of two elements, f, g, is measurable and satisfies $|f + g|^p \leq 2^p(|f|^p + |g|^p)$ so $|f + g|^p \in \mathcal{L}^1(\mathbb{R})$.

We next strengthen (2.107) to the approximation condition that there exists a sequence of simple functions v_n such that

$$v_n \to f \text{ a.e. and } |v_n|^p \leq F \in \mathcal{L}^1(\mathbb{R}) \text{ a.e.} \tag{2.110}$$

which certainly implies (2.107). As in the proof of Lemma 2.11, suppose $f \in \mathcal{L}^p(\mathbb{R})$ is real and choose f_n real-valued simple functions and converging to f almost everywhere. Since $|f|^p \in \mathcal{L}^1(\mathbb{R})$ there is a sequence of simple functions $0 \leq h_n$ such that $|h_n| \leq F$ for some $F \in \mathcal{L}^1(\mathbb{R})$ and $h_n \to |f|^p$ almost everywhere. Then set $g_n = h_n^{\frac{1}{p}}$ which is also a sequence of simple functions and define v_n by (2.106). It follows that (2.110) holds for the real part of f but combining sequences for real and imaginary parts such a sequence exists in general.

The advantage of the approximation condition (2.110) is that it allows us to conclude that the triangle inequality holds for $\|u\|_{L^p}$ defined by (2.109) since we know it for simple functions and from (2.110) it follows that $|v_n|^p \to |f|^p$ in $\mathcal{L}^1(\mathbb{R})$ so $\|v_n\|_{L^p} \to \|f\|_{L^p}$. Then if w_n is a similar sequence for $g \in \mathcal{L}^p(\mathbb{R})$

$$\|f+g\|_{L^p} \leq \limsup_n \|v_n+w_n\|_{L^p} \leq \limsup_n \|v_n\|_{L^p}+\limsup_n \|w_n\|_{L^p} = \|f\|_{L^p}+\|g\|_{L^p}. \tag{2.111}$$

The other two conditions being clear it follows that $\|u\|_{L^p}$ is a seminorm on $\mathcal{L}^p(\mathbb{R})$.

The vanishing of $\|u\|_{L^p}$ implies that $|u|^p$ and hence $u \in \mathcal{N}$ and the converse follows immediately. Thus $L^p(\mathbb{R}) = \mathcal{L}^p(\mathbb{R})/\mathcal{N}$ is a normed space and it only remains to check completeness.

We know that completeness is equivalent to the convergence of any absolutely summable series. So, we can suppose $f_n \in \mathcal{L}^p(\mathbb{R})$ have

$$\sum_n \left(\int |f_n|^p\right)^{\frac{1}{p}} < \infty. \tag{2.112}$$

Consider the sequence $g_n = f_n\chi_{[-R,R]}$ for some fixed $R > 0$. This is in $\mathcal{L}^1(\mathbb{R})$ and

$$\|g_n\|_{L^1} \leq (2R)^{\frac{1}{q}}\|f_n\|_{L^p} \tag{2.113}$$

by the integral form of Hölder's inequality

$$f \in \mathcal{L}^p(\mathbb{R}),\ g \in \mathcal{L}^q(\mathbb{R}),\ \frac{1}{p} + \frac{1}{q} = 1 \Longrightarrow fg \in \mathcal{L}^1(\mathbb{R}) \text{ and } |\int fg| \le \|f\|_{L^p}\|g\|_{L^q} \tag{2.114}$$

which can be proved by the same approximation argument as above, see Problem **??**. Thus the series g_n is absolutely summable in L^1 and so converges absolutely almost everywhere. It follows that the series $\sum_n f_n(x)$ converges absolutely almost everywhere – since it is just $\sum_n g_n(x)$ on $[-R, R]$, to a function, f.

So, we only need show that $f \in \mathcal{L}^p(\mathbb{R})$ and that $\int |f - F_n|^p \to 0$ as $n \to \infty$ where $F_n = \sum\limits_{k=1}^n f_k$. By Minkowski's inequality we know that $h_n = (\sum\limits_{k=1}^n |f_k|)^p$ has bounded L^1 norm, since

$$\||h_n|\|_{L^1}^{\frac{1}{p}} = \|\sum_{k=1}^n |f_k|\|_{L^p}. \le \sum_k \|f_k\|_{L^p}. \tag{2.115}$$

Thus, h_n is an increasing sequence of functions in $\mathcal{L}^1(\mathbb{R})$ with bounded integral, so by the Monotonicity Lemma it converges a.e. to a function $h \in \mathcal{L}^1(\mathbb{R})$. Since $|F_n|^p \le h$ and $|F_n|^p \to |f|^p$ a.e. it follows by Dominated convergence that

$$|f|^p \in \mathcal{L}^1(\mathbb{R}),\ \||f|^p\|_{L^1}^{\frac{1}{p}} \le \sum_n \|f_n\|_{L^p} \tag{2.116}$$

and hence $f \in \mathcal{L}^p(\mathbb{R})$. Applying the same reasoning to $f - F_n$ which is the sum of the series starting at term $n + 1$ gives the norm convergence:

$$\|f - F_n\|_{L^p} \le \sum_{k>n} \|f_k\|_{L^p} \to 0 \text{ as } \ n \to \infty. \tag{2.117}$$

□

A function $f : \mathbb{R} \longrightarrow \mathbb{C}$ is *locally integrable* if

$$F_{[-N,N]} = \begin{cases} f(x) & x \in [-N, N] \\ 0 & x \text{ if } |x| > N \end{cases} \Longrightarrow F_{[-N,N]} \in \mathcal{L}^1(\mathbb{R})\ \forall\ N. \tag{2.118}$$

So any continuous function on $\mathbb{R}$ is locally integrable as is any element of $\mathcal{L}^1(\mathbb{R})$.

Lemma 2.12. *The locally integrable functions form a linear space,* $\mathcal{L}^1_{\text{loc}}(\mathbb{R})$ *and*

$$\begin{gathered} \mathcal{L}^p(\mathbb{R}) = \{f \in \mathcal{L}^1_{\text{loc}}(\mathbb{R}); |f|^p \in \mathcal{L}^1(\mathbb{R})\}\ 1 \le p < \infty \\ \mathcal{L}^\infty(\mathbb{R}) = \{f \in \mathcal{L}^1_{\text{loc}}(\mathbb{R}); \sup_{\mathbb{R}\setminus E} |f(x)| < \infty \textit{ for some } E \textit{ of measure zero.}\} \end{gathered} \tag{2.119}$$

The proof is left as an exercise.

12. Lebesgue measure

In case anyone is interested in how to define Lebesgue measure from where we are now we can just use the integral.

Definition 2.10. A set $A \subset \mathbb{R}$ is *measurable* if its characteristic function χ_A is locally integrable. A measurable set A has finite measure if $\chi_A \in \mathcal{L}^1(\mathbb{R})$ and then

$$\mu(A) = \int \chi_A \tag{2.120}$$

is the Lebesgue measure of A. If A is measurable but not of finite measure then $\mu(A) = \infty$ by definition.

We know immediately that any interval (a, b) is measurable (indeed whether open, semi-open or closed) and has finite measure if and only if it is bounded – then the measure is $b - a$.

Proposition 2.10. *The complement of a measurable set is measurable and any* countable *union or countable intersection of measurable sets is measurable.*

Proof. The first part follows from the fact that the constant function 1 is locally integrable and hence $\chi_{\mathbb{R}\setminus A} = 1 - \chi_A$ is locally integrable if and only if χ_A is locally integrable.

Notice the relationship between a characteristic function and the set it defines:-

$$\chi_{A\cup B} = \max(\chi_A, \chi_B),\ \chi_{A\cap B} = \min(\chi_A, \chi_B). \tag{2.121}$$

If we have a sequence of sets A_n then $B_n = \bigcup_{k\le n} A_k$ is clearly an increasing sequence of sets and

$$\chi_{B_n} \to \chi_B,\ B = \sum_n A_n \tag{2.122}$$

is an increasing sequence which converges pointwise (at each point it jumps to 1 somewhere and then stays or else stays at 0.) Now, if we multiply by $\chi_{[-N,N]}$ then

$$f_n = \chi_{[-N,N]}\chi_{B_n} \to \chi_{B\cap[-N,N]} \tag{2.123}$$

is an increasing sequence of integrable functions – assuming that is that the A_k's are measurable – with integral bounded above, by $2N$. Thus by the monotonicity lemma the limit is integrable so χ_B is locally integrable and hence $\bigcup_n A_n$ is measurable.

For countable intersections the argument is similar, with the sequence of characteristic functions decreasing. □

Corollary 2.1. *The (Lebesgue) measurable subsets of* $\mathbb{R}$ *form a collection,* $\mathcal{M}$, *of the power set of* $\mathbb{R}$, *including* $\emptyset$ *and* $\mathbb{R}$ *which is closed under complements, countable unions and countable intersections.*

Such a collection of subsets of a set X is called a 'σ-algebra' – so a σ-algebra Σ in a set X is a collection of subsets of X containing X, $\emptyset$, the complement of any element and countable unions and intersections of any element. A (positive) measure is usually defined as a map $\mu : \Sigma \longrightarrow [0, \infty]$ with $\mu(\emptyset) = 0$ and such that

$$\mu(\bigcup_n E_n) = \sum_n \mu(E_n) \tag{2.124}$$

for any sequence $\{E_m\}$ of sets in Σ which are disjoint (in pairs).

As for Lebesgue measure a set $A \in \Sigma$ is 'measurable' and if $\mu(A)$ is not of finite measure it is said to have infinite measure – for instance $\mathbb{R}$ is of infinite measure in this sense. Since the measure of a set is always non-negative (or undefined if it isn't measurable) this does not cause any problems and in fact Lebesgue measure is countably additive as in (2.124) provided we allow ∞ as a value of the measure. It is a good exercise to prove this!

13. Higher dimensions

I have never actually covered this in lectures – there is simply not enough time. Still it is worth knowing that the Lebesgue integral in higher dimensional Euclidean spaces can be obtained following the same line of reasoning. So, we want – with the advantage of a little more experience – to go back to the beginning and define $\mathcal{L}^1(\mathbb{R}^n)$, $L^1(\mathbb{R}^n)$, $\mathcal{L}^2(\mathbb{R}^n)$ and $L^2(\mathbb{R}^n)$. In fact relatively little changes but there are some things that one needs to check a little carefully.

The first hurdle is that I am not assuming that you have covered the Riemann integral in higher dimensions; it is in my view a rather pointless thing to do anyway. Fortunately we do not really need that since we can just iterate the one-dimensional Riemann integral for continuous functions. So, define

$$\mathcal{C}_{\mathrm{c}}(\mathbb{R}^n) = \{u : \mathbb{R}^n \longrightarrow \mathbb{C};\ \text{continuous and such that } u(x) = 0 \text{ for } |x| > R\} \tag{2.125}$$

where of course the R can depend on the element. Now, if we hold say the last $n-1$ variables fixed, we get a continuous function of one variable which vanishes when $|x| > R$:

$$u(\cdot, x_2, \dots, x_n) \in \mathcal{C}_{\mathrm{c}}(\mathbb{R}) \text{ for each } (x_2, \dots, x_n) \in \mathbb{R}^{n-1}. \tag{2.126}$$

So we can integrate it and get a function

$$I_1(x_2, \dots, x_n) = \int_{\mathbb{R}} u(x, x_1, \dots, x_n),\ I_1 : \mathbb{R}^{n-1} \longrightarrow \mathbb{C}. \tag{2.127}$$

Lemma 2.13. *For each $u \in \mathcal{C}_c(\mathbb{R}^n)$, $I_1 \in \mathcal{C}_c(\mathbb{R}^{n-1})$.*

Proof. Certainly if $|(x_2, \dots, x_n)| > R$ then $u(\cdot, x_2, \dots, x_n) \equiv 0$ as a function of the first variable and hence $I_1 = 0$ in $|(x_2, \dots, x_n)| > R$. The continuity follows from the uniform continuity of a function on the compact set $|x| \le R$, $|(x_2, \dots, x_n) \le R$ of $\mathbb{R}^n$. Thus given $\epsilon > 0$ there exists $\delta > 0$ such that

$$|x - x'| < \delta,\ |y - y'|_{\mathbb{R}^{n-1}} < \delta \Longrightarrow |u(x, y) - u(x', y')| < \epsilon. \tag{2.128}$$

From the standard estimate for the Riemann integral,

$$|I_1(y) - I_1(y')| \le \int_{-R}^{R} |u(x, y) - u(x, y')| dx \le 2R\epsilon \tag{2.129}$$

if $|y - y'| < \delta$. This implies the (uniform) continuity of I_1. Thus $I_1 \in \mathcal{C}_{\mathrm{c}}(\mathbb{R}^{n-2})$ □

The upshot of this lemma is that we can integrate again, and hence a total of n times and so define the (iterated) Riemann integral as

$$\int_{\mathbb{R}^n} u(z)dz = \int_{-R}^{R} \int_{-R}^{R} \cdots \int_{-R}^{R} u(x_1, x_2, x_3, \dots, x_n) dx_1 dx_2 \dots dx_n \in \mathbb{C}. \tag{2.130}$$

Lemma 2.14. *The interated Riemann integral is a well-defined linear map*

$$\mathcal{C}_c(\mathbb{R}^n) \longrightarrow \mathbb{C} \tag{2.131}$$

which satisfies

$$|\int u| \le \int |u| \le (2R)^n \sup |u| \ \textit{if } u \in \mathcal{C}_c(\mathbb{R}^n) \textit{ and } u(z) = 0 \textit{ in } |z| > R. \tag{2.132}$$

Proof. This follows from the standard estimate in one dimension. □

Now, one slightly annoying thing is that we would really want to know that the integral is independent of the order of integration. In fact it is not hard – see Problem XX. Again using properties of the one-dimensional Riemann integral we find:-

Lemma 2.15. *The iterated integral*

$$\|u\|_{L^1} = \int_{\mathbb{R}^n} |u| \tag{2.133}$$

is a norm on $\mathcal{C}_c(\mathbb{R}^n)$.

Definition 2.11. The space $\mathcal{L}^1(\mathbb{R}^n)$ is defined to consist of those functions $f : \mathbb{R}^n \longrightarrow \mathbb{C}$ such that there exists a sequence $\{f_n\}$ which is absolutely summable with respect to the L^1 norm and such that

$$\sum_n |f_n(x)| < \infty \Longrightarrow \sum_n f_n(x) = f(x). \tag{2.134}$$

Now you can go through the whole discusion above in this higher dimensional case, and the only changes are really notational!

Things get a littlem more complicated in the discussion of change of variable. This is covered in the problems. There are also a few other theorems it is good to know!

CHAPTER 3

Hilbert spaces

There are really three 'types' of Hilbert spaces (over $\mathbb{C}$). The finite dimensional ones, essentially just $\mathbb{C}^n$, for different integer values of n, with which you are pretty familiar, and two infinite dimensional types corresponding to being separable (having a countable dense subset) or not. As we shall see, there is really only one separable infinite-dimensional Hilbert space (no doubt you realize that the $\mathbb{C}^n$ are separable) and that is what we are mostly interested in. Nevertheless we try to state results in general and then give proofs (usually they are the nicest ones) which work in the non-separable cases too.

I will first discuss the definition of pre-Hilbert and Hilbert spaces and prove Cauchy's inequality and the parallelogram law. This material can be found in many other places, so the discussion here will be kept succinct. One nice source is the book of G.F. Simmons, "Introduction to topology and modern analysis" [**5**]. I like it – but I think it is long out of print.

RBM:Add description of contents when complete and mention problems

1. pre-Hilbert spaces

A pre-Hilbert space, H, is a vector space (usually over the complex numbers but there is a real version as well) with a Hermitian inner product

$$\begin{gathered}\langle,\rangle : H \times H \longrightarrow \mathbb{C},\\ \langle \lambda_1 v_1 + \lambda_2 v_2, w\rangle = \lambda_1\langle v_1, w\rangle + \lambda_2\langle v_2, w\rangle,\\ \langle w, v\rangle = \overline{\langle v, w\rangle}\end{gathered} \tag{3.1}$$

for any v_1, v_2, v and $w \in H$ and λ_1, $\lambda_2 \in \mathbb{C}$ which is positive-definite

$$\langle v, v\rangle \geq 0,\ \langle v, v\rangle = 0 \Longrightarrow v = 0. \tag{3.2}$$

Note that the reality of $\langle v, v\rangle$ follows from the 'Hermitian symmetry' condition in (3.1), the positivity is an additional assumption as is the positive-definiteness.

The combination of the two conditions in (3.1) implies 'anti-linearity' in the second variable

$$\langle v, \lambda_1 w_1 + \lambda_2 w_2\rangle = \overline{\lambda_1}\langle v, w_1\rangle + \overline{\lambda_2}\langle v, w_2\rangle \tag{3.3}$$

which is used without comment below.

The notion of 'definiteness' for such an Hermitian inner product exists without the need for positivity – it just means

$$\langle u, v\rangle = 0\ \forall\ v \in H \Longrightarrow u = 0. \tag{3.4}$$

Lemma 3.1. *If H is a pre-Hilbert space with Hermitian inner product $\langle,\rangle$ then*

$$\|u\| = \langle u, u\rangle^{\frac{1}{2}} \tag{3.5}$$

is a norm on H.

Proof. The first condition on a norm follows from (3.2). Absolute homogeneity follows from (3.1) since

$$\|\lambda u\|^2 = \langle \lambda u, \lambda u\rangle = |\lambda|^2\|u\|^2. \tag{3.6}$$

So, it is only the triangle inequality we need. This follows from the next lemma, which is the Cauchy-Schwarz inequality in this setting – (3.8). Indeed, using the 'sesqui-linearity' to expand out the norm

$$\begin{aligned}\|u+v\|^2 &= \langle u+v, u+v\rangle \\ &= \|u\|^2 + \langle u,v\rangle + \langle v,u\rangle + \|v\|^2 \le \|u\|^2 + 2\|u\|\|v\| + \|v\|^2 \\ &= (\|u\|+\|v\|)^2.\end{aligned} \tag{3.7}$$

□

Lemma 3.2. *The Cauchy-Schwarz inequality,*

$$|\langle u,v\rangle| \le \|u\|\|v\| \ \forall\ u,v \in H \tag{3.8}$$

holds in any pre-Hilbert space.

Proof. This inequality is trivial if either u or v vanishes. For any non-zero u, $v \in H$ and $s \in \mathbb{R}$ positivity of the norm shows that

$$0 \le \|u+sv\|^2 = \|u\|^2 + 2s\operatorname{Re}\langle u,v\rangle + s^2\|v\|^2. \tag{3.9}$$

This quadratic polynomial in s is non-zero for s large so can have only a single minimum at which point the derivative vanishes, i.e. it is where

$$2s\|v\|^2 + 2\operatorname{Re}\langle u,v\rangle = 0. \tag{3.10}$$

Substituting this into (3.9) gives

$$\|u\|^2 - (\operatorname{Re}\langle u,v\rangle)^2/\|v\|^2 \ge 0 \Longrightarrow |\operatorname{Re}\langle u,v\rangle| \le \|u\|\|v\| \tag{3.11}$$

which is what we want except that it is only the real part. However, we know that, for some $z \in \mathbb{C}$ with $|z| = 1$, $\operatorname{Re}\langle zu,v\rangle = \operatorname{Re} z\langle u,v\rangle = |\langle u,v\rangle|$ and applying (3.11) with u replaced by zu gives (3.8). □

Corollary 3.1. *The inner product is continuous on the metric space (i.e. with respect to the norm) $H \times H$.*

Proof. Corollaries really aren't supposed to require proof! If $(u_j, v_j) \to (u,v)$ then, by definition $\|u-u_j\| \to 0$ and $\|v-v_j\| \to 0$ so from

$$\begin{aligned}|\langle u,v\rangle - \langle u_j,v_j\rangle| &\le |\langle u,v\rangle - \langle u,v_j\rangle| + |\langle u,v_j\rangle - \langle u_j,v_j\rangle| \\ &\le \|u\|\|v-v_j\| + \|u-u_j\|\|v_j\|\end{aligned} \tag{3.12}$$

continuity follows. □

Corollary 3.2. *The Cauchy-Scwharz inequality is optimal in the sense that*

$$\|u\| = \sup_{v\in H; \|v\|\le 1} |\langle u,v\rangle|. \tag{3.13}$$

I really will leave this one to you.

2. Hilbert spaces

Definition 3.1. A Hilbert space H is a pre-Hilbert space which is complete with respect to the norm induced by the inner product.

As examples we know that $\mathbb{C}^n$ with the usual inner product

$$\langle z, z'\rangle = \sum_{j=1}^{n} z_j \overline{z'_j} \tag{3.14}$$

is a Hilbert space – since any finite dimensional normed space is complete. The example we had from the beginning of the course is l^2 with the extension of (3.14)

$$\langle a, b\rangle = \sum_{j=1}^{\infty} a_j \overline{b_j},\ a, b \in l^2. \tag{3.15}$$

Completeness was shown earlier.

The whole outing into Lebesgue integration was so that we could have the 'standard example' at our disposal, namely

$$L^2(\mathbb{R}) = \{u \in \mathcal{L}^1_{\text{loc}}(\mathbb{R}); |u|^2 \in \mathcal{L}^1(\mathbb{R})\}/\mathcal{N} \tag{3.16}$$

where $\mathcal{N}$ is the space of null functions. The inner product is

$$\langle u, v\rangle = \int u\overline{v}. \tag{3.17}$$

Note that we showed that if $u,\ v \in \mathcal{L}^2(\mathbb{R})$ then $uv \in \mathcal{L}^1(\mathbb{R})$. We also showed that if $\int |u|^2 = 0$ then $u = 0$ almost everywhere, i.e. $u \in \mathcal{N}$, which is the definiteness of the inner product (3.17). It is fair to say that we went to some trouble to prove the completeness of this norm, so $L^2(\mathbb{R})$ is indeed a Hilbert space.

3. Orthonormal sequences

Two elements of a pre-Hilbert space H are said to be orthogonal if

$$\langle u, v\rangle = 0 \text{ which can be written } u \perp v. \tag{3.18}$$

A sequence of elements $e_i \in H$, (finite or infinite) is said to be *orthonormal* if $\|e_i\| = 1$ for all i and $\langle e_i, e_j\rangle = 0$ for all $i \neq j$.

Proposition 3.1 (Bessel's inequality). *If e_i, $i \in \mathbb{N}$, is an orthonormal sequence in a pre-Hilbert space H, then*

$$\sum_i |\langle u, e_i\rangle|^2 \leq \|u\|^2\ \forall\ u \in H. \tag{3.19}$$

Proof. Start with the finite case, $i = 1, \dots, N$. Then, for any $u \in H$ set

$$v = \sum_{i=1}^{N} \langle u, e_i\rangle e_i. \tag{3.20}$$

This is supposed to be 'the projection of u onto the span of the e_i'. Anyway, computing away we see that

$$\langle v, e_j\rangle = \sum_{i=1}^{N} \langle u, e_i\rangle\langle e_i, e_j\rangle = \langle u, e_j\rangle \tag{3.21}$$

using orthonormality. Thus, $u - v \perp e_j$ for all j so $u - v \perp v$ and hence

$$0 = \langle u - v, v\rangle = \langle u, v\rangle - \|v\|^2. \tag{3.22}$$

Thus $\|v\|^2 = |\langle u, v\rangle|$ and applying the Cauchy-Schwarz inequality we conclude that $\|v\|^2 \le \|v\|\|u\|$ so either $v = 0$ or $\|v\| \le \|u\|$. Expanding out the norm (and observing that all cross-terms vanish)

$$\|v\|^2 = \sum_{i=1}^{N} |\langle u, e_i\rangle|^2 \le \|u\|^2$$

which is (3.19).

In case the sequence is infinite this argument applies to any finite subsequence, e_i, $i = 1, \dots, N$ since it just uses orthonormality, so (3.19) follows by taking the supremum over N. □

4. Gram-Schmidt procedure

Definition 3.2. An orthonormal sequence, $\{e_i\}$, (finite or infinite) in a pre-Hilbert space is said to be *maximal* if

$$u \in H,\ \langle u, e_i\rangle = 0\ \forall\ i \Longrightarrow u = 0. \tag{3.23}$$

Theorem 3.1. *Every separable pre-Hilbert space contains a maximal orthonormal sequence.*

Proof. Take a countable dense subset – which can be arranged as a sequence $\{v_j\}$ and the existence of which is the definition of separability – and orthonormalize it. First if $v_1 \neq 0$ set $e_i = v_1/\|v_1\|$. Proceeding by induction we can suppose we have found, for a given integer n, elements e_i, $i = 1, \dots, m$, where $m \le n$, which are orthonormal and such that the linear span

$$\mathrm{sp}(e_1, \dots, e_m) = \mathrm{sp}(v_1, \dots, v_n). \tag{3.24}$$

We certainly have this for $n = 1$. To show the inductive step observe that if v_{n+1} is in the span(s) in (3.24) then the same e_i's work for $n + 1$. So we may as well assume that the next element, v_{n+1} is not in the span in (3.24). It follows that

$$w = v_{n+1} - \sum_{j=1}^{n} \langle v_{n+1}, e_j\rangle e_j \neq 0 \text{ so } e_{m+1} = \frac{w}{\|w\|} \tag{3.25}$$

makes sense. By construction it is orthogonal to all the earlier e_i's so adding e_{m+1} gives the equality of the spans for $n + 1$.

Thus we may continue indefinitely, since in fact the only way the dense set could be finite is if we were dealing with the space with one element, 0, in the first place. There are only two possibilities, either we get a finite set of e_i's or an infinite sequence. In either case this must be a maximal orthonormal sequence. That is, we claim

$$H \ni u \perp e_j\ \forall\ j \Longrightarrow u = 0. \tag{3.26}$$

This uses the density of the v_j's. There must exist a sequence w_k where each w_k is a v_j, such that $w_k \to u$ in H, assumed to satisfy (3.26). Now, each v_j, and hence each w_k, is a finite linear combination of e_l's so, by Bessel's inequality

$$\|w_k\|^2 = \sum_l |\langle w_k, e_l\rangle|^2 = \sum_l |\langle u - w_k, e_l\rangle|^2 \le \|u - w_k\|^2 \tag{3.27}$$

where $\langle u, e_l\rangle = 0$ for all l has been used. Thus $\|w_k\| \to 0$ and hence $u = 0$. □

Although a non-complete but separable pre-Hilbert space has maximal orthonormal sets, these are not much use without completeness.

5. Orthonormal bases

Definition 3.3. In view of the following result, a maximal orthonormal sequence in a separable Hilbert space will be called an orthonormal basis; it is often called a 'complete orthonormal basis' but the 'complete' is really redundant.

This notion of basis is not quite the same as in the finite dimensional case (although it is a legitimate extension of it). There are other, quite different, notions of a basis in infinite dimensions. See for instance 'Hamel basis' which arises in some settings – it is discussed briefly in §1.12 and can be used to show the existence of a non-continuous functional on a Banach space.

Theorem 3.2. *If $\{e_i\}$ is an orthonormal basis (a maximal orthonormal sequence) in a Hilbert space then for any element $u \in H$ the 'Fourier-Bessel series' converges to u :*

$$u = \sum_{i=1}^{\infty} \langle u, e_i\rangle e_i. \tag{3.28}$$

In particular a Hilbert space with an orthonormal basis is separable!

Proof. The sequence of partial sums of the Fourier-Bessel series

$$u_N = \sum_{i=1}^{N} \langle u, e_i\rangle e_i \tag{3.29}$$

is Cauchy. Indeed, if $m < m'$ then

$$\|u_{m'} - u_m\|^2 = \sum_{i=m+1}^{m'} |\langle u, e_i\rangle|^2 \le \sum_{i>m} |\langle u, e_i\rangle|^2 \tag{3.30}$$

which is small for large m by Bessel's inequality. Since we are now assuming completeness, $u_m \to w$ in H. However, $\langle u_m, e_i\rangle = \langle u, e_i\rangle$ as soon as $m > i$ and $|\langle w - u_n, e_i\rangle| \le \|w - u_n\|$ so in fact

$$\langle w, e_i\rangle = \lim_{m\to\infty} \langle u_m, e_i\rangle = \langle u, e_i\rangle \tag{3.31}$$

for each i. Thus $u - w$ is orthogonal to all the e_i so by the assumed completeness of the orthonormal basis must vanish. Thus indeed (3.28) holds. □

6. Isomorphism to l^2

A finite dimensional Hilbert space is isomorphic to $\mathbb{C}^n$ with its standard inner product. Similarly from the result above

Proposition 3.2. *Any infinite-dimensional separable Hilbert space (over the complex numbers) is isomorphic to l^2, that is there exists a linear map*

$$T : H \longrightarrow l^2 \tag{3.32}$$

which is 1-1, onto and satisfies $\langle Tu, Tv\rangle_{l^2} = \langle u, v\rangle_H$ and $\|Tu\|_{l^2} = \|u\|_H$ for all u, $v \in H$.

Proof. Choose an orthonormal basis – which exists by the discussion above – and set

$$Tu = \{\langle u, e_j\rangle\}_{j=1}^{\infty}. \tag{3.33}$$

This maps H into l^2 by Bessel's inequality. Moreover, it is linear since the entries in the sequence are linear in u. It is 1-1 since $Tu = 0$ implies $\langle u, e_j\rangle = 0$ for all j implies $u = 0$ by the assumed completeness of the orthonormal basis. It is surjective since if $\{c_j\}_{j=1}^{\infty} \in l^2$ then

$$u = \sum_{j=1}^{\infty} c_j e_j \tag{3.34}$$

converges in H. This is the same argument as above – the sequence of partial sums is Cauchy since if $n > m$,

$$\| \sum_{j=m+1}^{n} c_j e_j \|_H^2 = \sum_{j=m+1}^{n} |c_j|^2. \tag{3.35}$$

Again by continuity of the inner product, $Tu = \{c_j\}$ so T is surjective.

The equality of the norms follows from equality of the inner products and the latter follows by computation for finite linear combinations of the e_j and then in general by continuity. □

7. Parallelogram law

What exactly is the difference between a general Banach space and a Hilbert space? It is of course the existence of the inner product defining the norm. In fact it is possible to formulate this condition intrinsically in terms of the norm itself.

Proposition 3.3. *In any pre-Hilbert space the parallelogram law holds –*

$$\|v + w\|^2 + \|v - w\|^2 = 2\|v\|^2 + 2\|w\|^2, \ \forall \ v, w \in H. \tag{3.36}$$

Proof. Just expand out using the inner product

$$\|v + w\|^2 = \|v\|^2 + \langle v, w\rangle + \langle w, v\rangle + \|w\|^2 \tag{3.37}$$

and the same for $\|v - w\|^2$ and see the cancellation. □

Proposition 3.4. *Any normed space where the norm satisfies the parallelogram law,* (3.36)*, is a pre-Hilbert space in the sense that*

$$\langle v, w\rangle = \frac{1}{4}\left(\|v + w\|^2 - \|v - w\|^2 + i\|v + iw\|^2 - i\|v - iw\|^2\right) \tag{3.38}$$

is a positive-definite Hermitian inner product which reproduces the norm.

Proof. A problem below. □

So, when we use the parallelogram law and completeness we are using the essence of the Hilbert space.

8. Convex sets and length minimizer

The following result does not need the hypothesis of separability of the Hilbert space and allows us to prove the subsequent results – especially Riesz' theorem – in full generality.

PROPOSITION 3.5. *If $C \subset H$ is a subset of a Hilbert space which is*

(1) *Non-empty*
(2) *Closed*
(3) *Convex, in the sense that $v_1,\ v_2 \in C$ implies $\frac{1}{2}(v_1+v_2) \in C$*

then there exists a unique element $v \in C$ closest to the origin, i.e. such that

$$\|v\| = \inf_{u \in C} \|u\|. \tag{3.39}$$

PROOF. By definition of the infimum of a non-empty set of real numbers which is bounded below (in this case by 0) there must exist a sequence $\{v_n\}$ in C such that $\|v_n\| \to d = \inf_{u\in C}\|u\|$. We show that v_n converges and that the limit is the point we want. The parallelogram law can be written

$$\|v_n - v_m\|^2 = 2\|v_n\|^2 + 2\|v_m\|^2 - 4\|(v_n+v_m)/2\|^2. \tag{3.40}$$

Since $\|v_n\| \to d$, given $\epsilon > 0$ if N is large enough then $n > N$ implies $2\|v_n\|^2 < 2d^2 + \epsilon^2/2$. By convexity, $(v_n+v_m)/2 \in C$ so $\|(v_n+v_m)/2\|^2 \ge d^2$. Combining these estimates gives

$$n, m > N \Longrightarrow \|v_n - v_m\|^2 \le 4d^2 + \epsilon^2 - 4d^2 = \epsilon^2 \tag{3.41}$$

so $\{v_n\}$ is Cauchy. Since H is complete, $v_n \to v \in C$, since C is closed. Moreover, the distance is continuous so $\|v\| = \lim_{n\to\infty}\|v_n\| = d$.

Thus v exists and uniqueness follows again from the parallelogram law. If v and v' are two points in C with $\|v\| = \|v'\| = d$ then $(v+v')/2 \in C$ so

$$\|v - v'\|^2 = 2\|v\|^2 + 2\|v'\|^2 - 4\|(v+v')/2\|^2 \le 0 \Longrightarrow v = v'. \tag{3.42}$$

Alternatively you can just observe that we have actually shown above that *any* sequence in C such that $\|v_n\| \to d$ converges, so this is true for the alternating sequence of v and v'. □

9. Orthocomplements and projections

PROPOSITION 3.6. *If $W \subset H$ is a linear subspace of a Hilbert space then*

$$W^\perp = \{u \in H; \langle u, w\rangle = 0\ \forall\ w \in W\} \tag{3.43}$$

is a closed linear subspace and $W \cap W^\perp = \{0\}$. If W is also closed then

$$H = W \oplus W^\perp \tag{3.44}$$

meaning that any $u \in H$ has a unique decomposition $u = w + w^\perp$ where $w \in W$ and $w^\perp \in W^\perp$.

PROOF. That $W^\perp$ defined by (3.43) is a linear subspace follows from the linearity of the condition defining it. If $u \in W^\perp$ and $u \in W$ then $u \perp u$ by the definition so $\langle u, u\rangle = \|u\|^2 = 0$ and $u = 0$; thus $W \cap W^\perp = \{0\}$. Since the map $H \ni u \longrightarrow \langle u, w\rangle \in \mathbb{C}$ is continuous for each $w \in H$ its null space, the inverse image of 0, is closed. Thus

$$W^\perp = \bigcap_{w\in W} \{u \in H; \langle u, w\rangle = 0\} \tag{3.45}$$

is closed.

Now, suppose W is closed. If $W = H$ then $W^\perp = \{0\}$ and there is nothing to show. So consider $u \in H$, $u \notin W$ and set

$$C = u + W = \{u' \in H; u' = u + w,\ w \in W\}. \tag{3.46}$$

Then C is closed, since a sequence in it is of the form $u'_n = u + w_n$ where w_n is a sequence in W and u'_n converges if and only if w_n converges. Also, C is non-empty, since $u \in C$ and it is convex since $u' = u + w'$ and $u'' = u + w''$ in C implies $(u' + u'')/2 = u + (w' + w'')/2 \in C$.

Thus the length minimization result above applies and there exists a unique $v \in C$ such that $\|v\| = \inf_{u' \in C} \|u'\|$. The claim is that this v is orthogonal to W – draw a picture in two real dimensions! To see this consider an aritrary point $w \in W$ and $\lambda \in \mathbb{C}$ then $v + \lambda w \in C$ and

$$\|v + \lambda w\|^2 = \|v\|^2 + 2\operatorname{Re}(\lambda\langle v, w\rangle) + |\lambda|^2\|w\|^2. \tag{3.47}$$

Choose $\lambda = te^{i\theta}$ where t is real and the phase is chosen so that $e^{i\theta}\langle v, w\rangle = |\langle v, w\rangle| \geq 0$. Then the fact that $\|v\|$ is minimal means that

$$\begin{gathered} \|v\|^2 + 2t|\langle v, w\rangle)| + t^2\|w\|^2 \geq \|v\|^2 \Longrightarrow \\ t(2|\langle v, w\rangle| + t\|w\|^2) \geq 0\ \forall\ t \in \mathbb{R} \Longrightarrow |\langle v, w\rangle| = 0 \end{gathered} \tag{3.48}$$

which is what we wanted to show.

Thus indeed, given $u \in H \setminus W$ we have constructed $v \in W^\perp$ such that $u = v + w$, $w \in W$. This is (3.44) with the uniqueness of the decomposition already shown since it reduces to 0 having only the decomposition $0 + 0$ and this in turn is $W \cap W^\perp = \{0\}$. □

Since the construction in the preceding proof associates a unique element in W, a closed linear subspace, to each $u \in H$, it defines a map

$$\Pi_W : H \longrightarrow W. \tag{3.49}$$

This map is linear, by the uniqueness since if $u_i = v_i + w_i$, $w_i \in W$, $\langle v_i, w_i\rangle = 0$ are the decompositions of two elements then

$$\lambda_1 u_1 + \lambda_2 u_2 = (\lambda_1 v_1 + \lambda_2 v_2) + (\lambda_1 w_1 + \lambda_2 w_2) \tag{3.50}$$

must be the corresponding decomposition. Moreover $\Pi_W w = w$ for any $w \in W$ and $\|u\|^2 = \|v\|^2 + \|w\|^2$, Pythagoras' Theorem, shows that

$$\Pi_W^2 = \Pi_W,\ \|\Pi_W u\| \leq \|u\| \Longrightarrow \|\Pi_W\| \leq 1. \tag{3.51}$$

Thus, projection onto W is an operator of norm 1 (unless $W = \{0\}$) equal to its own square. Such an operator is called a projection or sometimes an idempotent (which sounds fancier).

Finite-dimensional subspaces are always closed by the Heine-Borel theorem.

Lemma 3.3. *If $\{e_j\}$ is any finite or countable orthonormal set in a Hilbert space then the orthogonal projection onto the closure of the span of these elements is*

$$Pu = \sum \langle u, e_k\rangle e_k. \tag{3.52}$$

Proof. We know that the series in (3.52) converges and defines a bounded linear operator of norm at most one by Bessel's inequality. Clearly $P^2 = P$ by the same argument. If W is the closure of the span then $(u - Pu) \perp W$ since $(u - Pu) \perp$

e_k for each k and the inner product is continuous. Thus $u = (u - Pu) + Pu$ is the orthogonal decomposition with respect to W. □

LEMMA 3.4. *If $W \subset H$ is a linear subspace of a Hilbert space which contains the orthocomplement of a finite dimensional space then W is closed and $W^\perp$ is finite-dimensional.*

PROOF. If $U \subset W$ is a closed subspace with finite-dimensional orthocomplement then each of the N elements, v_i, of a basis of $(\operatorname{Id} - \Pi_U)W$ is the image of some $w_i \in W$. Since U is the null space of $\operatorname{Id} - \Pi_U$ it follows that any element of W can be written uniquely in the form

$$w = u + \sum_{i=1}^{N} c_i v_i,\ u = \Pi_U w \in U,\ c_i = \langle w, v_i \rangle. \tag{3.53}$$

Then if ϕ_n is a sequence in W which converges in H it follows that $\Pi_U \phi_n$ converges in U and $\langle \phi_n, v_i \rangle$ converges and hence the limit is in W. □

Note that the existence of a non-continuous linear functional $H \longrightarrow \mathbb{C}$ is equivalent to the existence of a non-closed subspace of H with a one-dimensional complement. Namely the null space of a non-continuous linear functional cannot be closed, since from this continuity follows, but it does have a one-dimensional complement (not orthocomplement!)

QUESTION 1. Does there exist a non-continuous linear functional on an infinite-dimensional Hilbert space? [1]

10. Riesz' theorem

The most important application of the convexity result above is to prove Riesz' representation theorem (for Hilbert space, there is another one to do with measures).

THEOREM 3.3. *If H is a Hilbert space then for any continuous linear functional $T : H \longrightarrow \mathbb{C}$ there exists a unique element $\phi \in H$ such that*

$$T(u) = \langle u, \phi \rangle\ \forall\ u \in H. \tag{3.54}$$

PROOF. If T is the zero functional then $\phi = 0$ gives (3.54). Otherwise there exists some $u' \in H$ such that $T(u') \neq 0$ and then there is some $u \in H$, namely $u = u'/T(u')$ will work, such that $T(u) = 1$. Thus

$$C = \{u \in H; T(u) = 1\} = T^{-1}(\{1\}) \neq \emptyset. \tag{3.55}$$

The continuity of T implies that C is closed, as the inverse image of a closed set under a continuous map. Moreover C is convex since

$$T((u + u')/2) = (T(u) + T(u'))/2. \tag{3.56}$$

Thus, by Proposition 3.5, there exists an element $v \in C$ of minimal length.

Notice that $C = \{v + w; w \in N\}$ where $N = T^{-1}(\{0\})$ is the null space of T. Thus, as in Proposition 3.6 above, v is orthogonal to N. In this case it is the unique element orthogonal to N with $T(v) = 1$.

[1] The existence of such a functional requires some form of the Axiom of Choice (maybe a little weaker in the separable case). You are free to believe that all linear functionals are continuous but you will make your life difficult this way.

Now, for any $u \in H$,

$$u - T(u)v \text{ satisfies } T(u - T(u)v) = T(u) - T(u)T(v) = 0 \Longrightarrow u = w + T(u)v,\ w \in N. \tag{3.57}$$

Then, $\langle u, v\rangle = T(u)\|v\|^2$ since $\langle w, v\rangle = 0$. Thus if $\phi = v/\|v\|^2$ then

$$u = w + \langle u, \phi\rangle v \Longrightarrow T(u) = \langle u, \phi\rangle T(v) = \langle u, \phi\rangle. \tag{3.58}$$

□

11. Adjoints of bounded operators

As an application of Riesz' Theorem we can see that to any bounded linear operator on a Hilbert space

$$A : H \longrightarrow H,\ \|Au\| \le C\|u\| \ \forall\ u \in H \tag{3.59}$$

there corresponds a unique adjoint operator. This has profound consequences for the theory of operators on a Hilbert space, as we shall see.

Proposition 3.7. *For any bounded linear operator $A : H \longrightarrow H$ on a Hilbert space there is a unique bounded linear operator $A^* : H \longrightarrow H$ such that*

$$\langle Au, v\rangle_H = \langle u, A^*v\rangle_H \ \forall\ u, v \in H \text{ and } \|A\| = \|A^*\|. \tag{3.60}$$

Proof. To see the existence of A^*v we need to work out what $A^*v \in H$ should be for each fixed $v \in H$. So, fix v in the desired identity (3.60), which is to say consider

$$H \ni u \longrightarrow \langle Au, v\rangle \in \mathbb{C}. \tag{3.61}$$

This is a linear map and it is clearly bounded, since

$$|\langle Au, v\rangle| \le \|Au\|\|v\| \le (\|A\|\|v\|)\|u\|. \tag{3.62}$$

Thus it is a continuous linear functional on H which depends on v. In fact it is just the composite of two continuous linear maps

$$H \overset{u\longmapsto Au}{\longrightarrow} H \overset{w\longmapsto\langle w,v\rangle}{\longrightarrow} \mathbb{C}. \tag{3.63}$$

By Riesz' theorem there is a unique element in H, which we can denote A^*v (since it only depends on v) such that

$$\langle Au, v\rangle = \langle u, A^*v\rangle \ \forall\ u \in H. \tag{3.64}$$

This defines the map $A^* : H \longrightarrow H$ but we need to check that it is linear and continuous. Linearity follows from the uniqueness part of Riesz' theorem. Thus if v_1, $v_2 \in H$ and c_1, $c_2 \in \mathbb{C}$ then

$$\begin{aligned}\langle Au, c_1v_1 + c_2v_2\rangle &= \overline{c_1}\langle Au, v_1\rangle + \overline{c_2}\langle Au, v_2\rangle \\ &= \overline{c_1}\langle u, A^*v_1\rangle + \overline{c_2}\langle u, A^*v_2\rangle = \langle u, c_1A^*v_2 + c_2A^*v_2\rangle\end{aligned} \tag{3.65}$$

where we have used the definitions of A^*v_1 and A^*v_2 – by uniqueness we must have $A^*(c_1v_1 + c_2v_2) = c_1A^*v_1 + c_2A^*v_2$.

Using the optimality of Cauchy's inequality

$$\|A^*v\| = \sup_{\|u\|=1} |\langle u, A^*v\rangle| = \sup_{\|u\|=1} |\langle Au, v\rangle| \le \|A\|\|v\|. \tag{3.66}$$

This shows that A^* is bounded and that

$$\|A^*\| \le \|A\|. \tag{3.67}$$

The defining identity (3.60) also shows that $(A^*)^* = A$ so the reverse equality in (3.67) also holds and therefore

$$\|A^*\| = \|A\|. \tag{3.68}$$

□

One useful property of the adjoint operator is that

$$\operatorname{Nul}(A^*) = (\operatorname{Ran}(A))^\perp. \tag{3.69}$$

Indeed $w \in (\operatorname{Ran}(A))^\perp$ means precisely that $\langle w, Av\rangle = 0$ for all $v \in \mathcal{H}$ which translates to

$$w \in (\operatorname{Ran}(A))^\perp \Longleftrightarrow \langle A^*w, v\rangle = 0 \Longleftrightarrow A^*w = 0. \tag{3.70}$$

Note that in the finite dimensional case (3.69) is equivalent to $\operatorname{Ran}(A) = (\operatorname{Nul}(A^*))^\perp$ but in the infinite dimensional case $\operatorname{Ran}(A)$ is often not closed in which case this cannot be true and you can only be sure that

$$\overline{\operatorname{Ran}(A)} = (\operatorname{Nul}(A^*))^\perp. \tag{3.71}$$

12. Compactness and equi-small tails

A compact subset in a general metric space is one with the property that any sequence in it has a convergent subsequence, with its limit in the set. You will recall, with pleasure no doubt, the equivalence of this condition to the (more general since it makes good sense in an arbitrary topological space) covering condition, that *any* open cover of the set has a finite subcover. So, in a Hilbert space the notion of a compact set is already fixed. We want to characterize it, actually in two closely related ways.

In any metric space a compact set is both closed and bounded, so this must be true in a Hilbert space. The Heine-Borel theorem gives a converse to this, for $\mathbb{R}^n$ or $\mathbb{C}^n$ (and hence in any finite-dimensional normed space) any closed and bounded set is compact. Also recall that the convergence of a sequence in $\mathbb{C}^n$ is equivalent to the convergence of the n sequences given by its components and this is what is used to pass first from $\mathbb{R}$ to $\mathbb{C}$ and then to $\mathbb{C}^n$. All of this fails in infinite dimensions and we need some condition in addition to being bounded and closed for a set to be compact.

To see where this might come from, observe that

LEMMA 3.5. *In any metric space the set, S, consisting of the points of a convergent sequence, together with its limit, is compact.*

PROOF. We show that S is compact by checking that any sequence in S has a convergent subsequence. To see this, observe that a sequence $\{t_j\}$ in S either has a subsequence converging to the limit, s, of the original sequence or it does not. So we only need consider the latter case, but this means that, for some $\epsilon > 0$, $d(t_j, s) > \epsilon$; but then t_j takes values in a finite set, since $S \setminus B(s, \epsilon)$ is finite – hence some value is repeated infinitely often and there is a convergent subsequence. □

Lemma 3.6. *The image of a convergent sequence in a Hilbert space is a set with equi-small tails with respect to any orthonormal sequence, i.e. if e_k is an othonormal sequence and $u_n \to u$ is a convergent sequence then given $\epsilon > 0$ there exists N such that*

$$\sum_{k>N} |\langle u_n, e_k\rangle|^2 < \epsilon^2 \ \forall\ n. \tag{3.72}$$

Proof. Bessel's inequality shows that for any $u \in \mathcal{H}$,

$$\sum_{k} |\langle u, e_k\rangle|^2 \le \|u\|^2. \tag{3.73}$$

The convergence of this series means that (3.72) can be arranged for any single element u_n or the limit u by choosing N large enough, thus given $\epsilon > 0$ we can choose N' so that

$$\sum_{k>N'} |\langle u, e_k\rangle|^2 < \epsilon^2/2. \tag{3.74}$$

Consider the closure of the subspace spanned by the e_k with $k > N$. The orthogonal projection onto this space (see Lemma 3.3) is

$$P_N u = \sum_{k>N} \langle u, e_k\rangle e_k. \tag{3.75}$$

Then the convergence $u_n \to u$ implies the convergence in norm $\|P_N u_n\| \to \|P_N u\|$, so

$$\|P_N u_n\|^2 = \sum_{k>N} |\langle u_n, e_k\rangle|^2 < \epsilon^2,\ n > n'. \tag{3.76}$$

So, we have arranged (3.72) for $n > n'$ for some N. This estimate remains valid if N is increased – since the tails get smaller – and we may arrange it for $n \le n'$ by choosing N large enough. Thus indeed (3.72) holds for all n if N is chosen large enough. □

This suggests one useful characterization of compact sets in a separable Hilbert space since the equi-smallness of the tails, as in (3.72), for all $u \in K$ just means that the Fourier-Bessel series converges uniformly.

Proposition 3.8. *A set $K \subset \mathcal{H}$ in a separable Hilbert space is compact if and only if it is bounded, closed and the Fourier-Bessel sequence with respect to any (one) complete orthonormal basis converges* uniformly *on it.*

Proof. We already know that a compact set in a metric space is closed and bounded. Suppose the equi-smallness of tails condition fails with respect to some orthonormal basis e_k. This means that for some $\epsilon > 0$ and all p there is an element $u_p \in K$, such that

$$\sum_{k>p} |\langle u_p, e_k\rangle|^2 \ge \epsilon^2. \tag{3.77}$$

Consider the subsequence $\{u_p\}$ generated this way. No subsequence of it can have equi-small tails (recalling that the tail decreases with p). Thus, by Lemma 3.6, it cannot have a convergent subsequence, so K cannot be compact if the equi-smallness condition fails.

Thus we have proved the equi-smallness of tails condition to be necessary for the compactness of a closed, bounded set. It remains to show that it is sufficient.

So, suppose K is closed, bounded and satisfies the equi-small tails condition with respect to an orthonormal basis e_k and $\{u_n\}$ is a sequence in K. We only need show that $\{u_n\}$ has a Cauchy subsequence, since this will converge ($\mathcal{H}$ being complete) and the limit will be in K (since it is closed). Consider each of the sequences of coefficients $\langle u_n, e_k\rangle$ in $\mathbb{C}$. Here k is fixed. This sequence is bounded:

$$|\langle u_n, e_k\rangle| \le \|u_n\| \le C \tag{3.78}$$

by the boundedness of K. So, by the Heine-Borel theorem, there is a subsequence u_{n_l} such that $\langle u_{n_l}, e_k\rangle$ converges as $l \to \infty$.

We can apply this argument for each $k = 1, 2, \dots$. First extract a subsequence $\{u_{n,1}\}$ of $\{u_n\}$ so that the sequence $\langle u_{n,1}, e_1\rangle$ converges. Then extract a subsequence $u_{n,2}$ of $u_{n,1}$ so that $\langle u_{n,2}, e_2\rangle$ *also* converges. Then continue inductively. Now pass to the 'diagonal' subsequence v_n of $\{u_n\}$ which has kth entry the kth term, $u_{k,k}$ in the kth subsequence. It is 'eventually' a subsequence of each of the subsequences previously constructed – meaning it coincides with a subsequence from some point onward (namely the kth term onward for the kth subsquence). Thus, for this subsequence *each* of the $\langle v_n, e_k\rangle$ converges.

Consider the identity (the orthonormal set e_k is complete by assumption) for the difference

$$\begin{aligned}\|v_n - v_{n+l}\|^2 &= \sum_{k\le N} |\langle v_n - v_{n+l}, e_k\rangle|^2 + \sum_{k>N} |\langle v_n - v_{n+l}, e_k\rangle|^2 \\ &\le \sum_{k\le N} |\langle v_n - v_{n+l}, e_k\rangle|^2 + 2\sum_{k>N} |\langle v_n, e_k\rangle|^2 + 2\sum_{k>N} |\langle v_{n+l}, e_k\rangle|^2\end{aligned} \tag{3.79}$$

where the parallelogram law on $\mathbb{C}$ has been used. To make this sum less than ϵ^2 we may choose N so large that the last two terms are less than $\epsilon^2/2$ and this may be done for all n and l by the equi-smallness of the tails. Now, choose n so large that each of the terms in the first sum is less than $\epsilon^2/2N$, for all $l > 0$ using the Cauchy condition on each of the finite number of sequence $\langle v_n, e_k\rangle$. Thus, $\{v_n\}$ is a Cauchy subsequence of $\{u_n\}$ and hence as already noted convergent in K. Thus K is indeed compact. □

This criterion for compactness is useful but is too closely tied to the existence of an orthonormal basis to be easily applicable. However the condition can be restated in a way that holds even in the non-separable case (and of course in the finite-dimensional case, where it is trivial).

Proposition 3.9. *A subset $K \subset H$ of a Hilbert space is compact if and only if it is closed and bounded and for every $\epsilon > 0$ there is a finite-dimensional subspace $W \subset H$ such that*

$$\sup_{u\in K} \inf_{w\in W} \|u - w\| < \epsilon. \tag{3.80}$$

So we see that the extra condition needed is 'finite-dimensional approximability'.

Proof. Before proceeding to the proof consider (3.80). Since W is finite-dimensional we know it is closed and hence the discussion in §9 above applies. In particular $u = w + w^\perp$ with $w \in W$ and $w^\perp \perp W$ where

$$\inf_{w \in W} \|u - w\| = \|w^\perp\|. \tag{3.81}$$

This can be restated in the form

$$\sup_{u \in K} \|(\mathrm{Id} - \Pi_W)u\| < \epsilon \tag{3.82}$$

where Π_W is the orthogonal projection onto W (so $\mathrm{Id} - \Pi_W$ is the orthogonal projection onto $W^\perp$).

Now, let us first assume that H is separable, so we already have a condition for compactness in Proposition 3.8. Then if K is compact we can consider an orthonormal basis of H and the finite-dimensional spaces W_N spanned by the first N elements in the basis with Π_N the orthogonal projection onto it. Then $\|(\mathrm{Id} - \Pi_N)u\|$ is precisely the length of the 'tail' of u with respect to the basis. So indeed, by Proposition 3.8, given $\epsilon > 0$ there is an N such that $\|(\mathrm{Id} - \Pi_N)u\| < \epsilon/2$ for all $u \in K$ and hence (3.82) holds for $W = W_N$.

Now suppose that $K \subset H$ and for each $\epsilon > 0$ we can find a finite dimensional subspace W such that (3.82) holds. Take a sequence $\{u_n\}$ in K. The sequence $\Pi_W u_n \in W$ is bounded in a finite-dimensional space so has a convergent subsequence. Now, for each $j \in \mathbb{N}$ there is a finite-dimensional subspace W_j (not necessarily corresponding to an orthonormal basis) so that (3.82) holds for $\epsilon = 1/j$. Proceeding as above, we can find successive subsequence of u_n such that the image under Π_j in W_j converges for each j. Passing to the diagonal subsequence u_{n_l} it follows that $\Pi_j u_{n_i}$ converges for each j since it is eventually a subsequence of the jth choice of subsequence above. Now, the triangle inequality shows that

$$\|u_{n_i} - u_{n_k}\| \le \|\Pi_j(u_{n_i} - u_{n_k})\|_{W_j} + \|(\mathrm{Id} - \Pi_j)u_{n_i}\| + \|(\mathrm{Id} - \Pi_j)u_{n_k}\|. \tag{3.83}$$

Given $\epsilon > 0$ first choose j so large that the last two terms are each less than $1/j < \epsilon/3$ using the choice of W_j. Then if $i, k > N$ is large enough the first term on the right in (3.83) is also less than $\epsilon/3$ by the convergence of $\Pi_j u_{n_i}$. Thus u_{n_i} is Cauchy in H and hence converges and it follows that K is compact.

This converse argument does not require the separability of H so to complete the proof we only need to show the necessity of (3.81) in the non-separable case. Thus suppose K is compact. Then K itself is separable – has a countable dense subset – using the finite covering property (for each $p > 0$ there are finitely many balls of radius $1/p$ which cover K so take the set consisting of all the centers for all p). It follows that the closure of the span of K, the finite linear combinations of elements of K, is a separable Hilbert subspace of H which contains K. Thus any compact subset of a non-separable Hilbert space is contained in a separable Hilbert subspace and hence (3.80) holds. □

13. Finite rank operators

Now, we need to starting thinking a little more seriously about operators on a Hilbert space, remember that an operator is just a continuous linear map $T : \mathcal{H} \longrightarrow \mathcal{H}$ and the space of them (a Banach space) is denoted $\mathcal{B}(\mathcal{H})$ (rather than the more cumbersome $\mathcal{B}(\mathcal{H}, \mathcal{H})$ which is needed when the domain and target spaces are different).

DEFINITION 3.4. An operator $T \in \mathcal{B}(\mathcal{H})$ is of *finite rank* if its range has finite dimension (and that dimension is called the rank of T); the set of finite rank operators will be denoted $\mathcal{R}(\mathcal{H})$.

Why not $\mathcal{F}(\mathcal{H})$? Because we want to use this for the *Fredholm operators.*

Clearly the sum of two operators of finite rank has finite rank, since the range is contained in the sum of the ranges (but is often smaller):

$$(T_1 + T_2)u \in \operatorname{Ran}(T_1) + \operatorname{Ran}(T_2)\ \forall\ u \in \mathcal{H}. \tag{3.84}$$

Since the range of a constant multiple of T is contained in the range of T it follows that the finite rank operators form a linear subspace of $\mathcal{B}(\mathcal{H})$.

What does a finite rank operator look like? It really looks like a matrix.

LEMMA 3.7. *If $T : H \longrightarrow H$ has finite rank then there is a finite orthonormal set $\{e_k\}_{k=1}^L$ in H and constants $c_{ij} \in \mathbb{C}$ such that*

$$Tu = \sum_{i,j=1}^{L} c_{ij}\langle u, e_j\rangle e_i. \tag{3.85}$$

PROOF. By definition, the range of T, $R = T(H)$ is a finite dimensional subspace. So, it has a basis which we can diagonalize in H to get an orthonormal basis, e_i, $i = 1, \dots, p$. Now, since this is a basis of the range, Tu can be expanded relative to it for any $u \in H$:

$$Tu = \sum_{i=1}^{p} \langle Tu, e_i\rangle e_i. \tag{3.86}$$

On the other hand, the map $u \longrightarrow \langle Tu, e_i\rangle$ is a continuous linear functional on H, so $\langle Tu, e_i\rangle = \langle u, v_i\rangle$ for some $v_i \in H$; notice in fact that $v_i = T^*e_i$. This means the formula (3.86) becomes

$$Tu = \sum_{i=1}^{p} \langle u, v_i\rangle e_i. \tag{3.87}$$

Now, the Gram-Schmidt procedure can be applied to orthonormalize the sequence $e_1, \dots, e_p, v_1 \dots, v_p$ resulting in $e_1, \dots, e_L$. This means that each v_i is a linear combination which we can write as

$$v_i = \sum_{j=1}^{L} \overline{c_{ij}} e_j. \tag{3.88}$$

Inserting this into (3.87) gives (3.85) (where the constants for $i > p$ are zero). □

It is clear that

$$B \in \mathcal{B}(\mathcal{H}) \text{ and } T \in \mathcal{R}(\mathcal{H}) \text{ then } BT \in \mathcal{R}(\mathcal{H}). \tag{3.89}$$

Indeed, the range of BT is the range of B restricted to the range of T and this is certainly finite dimensional since it is spanned by the image of a basis of $\operatorname{Ran}(T)$. Similalry $TB \in \mathcal{R}(\mathcal{H})$ since the range of TB is contained in the range of T. Thus we have in fact proved most of

PROPOSITION 3.10. *The finite rank operators form a $*$-closed two-sided ideal in $\mathcal{B}(\mathcal{H})$, which is to say a linear subspace such that*

$$B_1,\ B_2 \in \mathcal{B}(\mathcal{H}),\ T \in \mathcal{R}(\mathcal{H}) \Longrightarrow B_1TB_2,\ T^* \in \mathcal{R}(\mathcal{H}). \tag{3.90}$$

Proof. It is only left to show that T^* is of finite rank if T is, but this is an immediate consequence of Lemma 3.7 since if T is given by (3.85) then

$$T^* u = \sum_{i,j=1}^{N} \overline{c_{ij}} \langle u, e_i \rangle e_j \tag{3.91}$$

is also of finite rank. □

Lemma 3.8 (Row rank=Colum rank). *For any finite rank operator on a Hilbert space, the dimension of the range of T is equal to the dimension of the range of T^*.*

Proof. From the formula (3.87) for a finite rank operator, it follows that the v_i, $i = 1, \dots, p$ must be linearly independent – since the e_i form a basis for the range and a linear relation between the v_i would show the range had dimension less than p. Thus in fact the null space of T is precisely the orthocomplement of the span of the v_i – the space of vectors orthogonal to each v_i. Since

$$\begin{gathered}\langle Tu, w \rangle = \sum_{i=1}^{p} \langle u, v_i \rangle \langle e_i, w \rangle \Longrightarrow \\ \langle w, Tu \rangle = \sum_{i=1}^{p} \langle v_i, u \rangle \langle w, e_i \rangle \Longrightarrow \\ T^* w = \sum_{i=1}^{p} \langle w, e_i \rangle v_i\end{gathered} \tag{3.92}$$

the range of T^* is the span of the v_i, so is also of dimension p. □

14. Compact operators

Definition 3.5. An element $K \in \mathcal{B}(\mathcal{H})$, the bounded operators on a separable Hilbert space, is said to be *compact* (the old terminology was 'totally bounded' or 'completely continuous') if the image of the unit ball is precompact, i.e. has compact closure – that is if the closure of $K\{u \in \mathcal{H}; \|u\|_{\mathcal{H}} \le 1\}$ is compact in $\mathcal{H}$.

Notice that in a metric space, to say that a set has compact closure is the same as saying it is contained in a compact set; such a set is said to be *precompact.*

Proposition 3.11. *An operator $K \in \mathcal{B}(\mathcal{H})$, bounded on a separable Hilbert space, is compact if and only if it is the limit of a norm-convergent sequence of finite rank operators.*

Proof. So, we need to show that a compact operator is the limit of a convergent sequence of finite rank operators. To do this we use the characterizations of compact subsets of a separable Hilbert space discussed earlier. Namely, if $\{e_i\}$ is an orthonormal basis of $\mathcal{H}$ then a subset $I \subset \mathcal{H}$ is compact if and only if it is closed and bounded and has equi-small tails with respect to $\{e_i\}$, meaning given $\epsilon > 0$ there exits N such that

$$\sum_{i>N} |\langle v, e_i \rangle|^2 < \epsilon^2 \ \forall \ v \in I. \tag{3.93}$$

Now we shall apply this to the set $K(B(0,1))$ where we assume that K is compact (as an operator, don't be confused by the double usage, in the end it turns out to be constructive) – so this set is *contained* in a compact set. Hence (3.93) applies to it. Namely this means that for any $\epsilon > 0$ there exists n such that

$$\sum_{i>n} |\langle Ku, e_i\rangle|^2 < \epsilon^2 \ \forall \ u \in \mathcal{H},\ \|u\|_{\mathcal{H}} \le 1. \tag{3.94}$$

For each n consider the first part of these sequences and define

$$K_n u = \sum_{k\le n} \langle Ku, e_i\rangle e_i. \tag{3.95}$$

This is clearly a linear operator and has finite rank – since its range is contained in the span of the first n elements of $\{e_i\}$. Since this is an orthonormal basis,

$$\|Ku - K_n u\|^2_{\mathcal{H}} = \sum_{i>n} |\langle Ku, e_i\rangle|^2 \tag{3.96}$$

Thus (3.94) shows that $\|Ku - K_n u\|_{\mathcal{H}} \le \epsilon$. Now, increasing n makes $\|Ku - K_n u\|$ smaller, so given $\epsilon > 0$ there exists n such that for all $N \ge n$,

$$\|K - K_N\|_{\mathcal{B}} = \sup_{\|u\|\le 1} \|Ku - K_n u\|_{\mathcal{H}} \le \epsilon. \tag{3.97}$$

Thus indeed, $K_n \to K$ in norm and we have shown that the compact operators are contained in the norm closure of the finite rank operators.

For the converse we assume that $T_n \to K$ is a norm convergent sequence in $\mathcal{B}(\mathcal{H})$ where each of the T_n is of finite rank – of course we know nothing about the rank except that it is finite. We want to conclude that K is compact, so we need to show that $K(B(0,1))$ is precompact. It is certainly bounded, by the norm of K. Let $W_n = T_n\mathcal{H}$ be the range of T_n. By definition it is a finite dimensional subspace and hence closed. Let Π_n be the orthogonal projection onto W_n, so $\operatorname{Id} - \Pi_n$ is projection onto $W_n^\perp$. Thus the composite $(\operatorname{Id} - \Pi_n)T_n = 0$ and hence

$$(\operatorname{Id} - \Pi_n)K = (\operatorname{Id} - \Pi_n)(K - T_n) \Longrightarrow \|(\operatorname{Id} - \Pi_n)K\| \to 0 \text{ as } n \to \infty. \tag{3.98}$$

So, for any $\epsilon > 0$ there exists n such that

$$\sup_{u\in B(0,1)} \inf_{w\in W_n} \|Ku - w\| \le \sup_{\|u\|\le 1} \|(\operatorname{Id} - \Pi_n)Ku\| < \epsilon \tag{3.99}$$

and it follows from Proposition 3.9 that $K(B(0,1))$ is precompact and hence K is compact. □

Proposition 3.12. *For any separable Hilbert space, the compact operators form a closed and $*$-closed two-sided ideal in $\mathcal{B}(H)$.*

Proof. In any metric space (applied to $\mathcal{B}(H)$) the closure of a set is closed, so the compact operators are closed being the closure of the finite rank operators. Similarly the fact that it is closed under passage to adjoints follows from the same fact for finite rank operators. The ideal properties also follow from the corresponding properties for the finite rank operators, or we can prove them directly anyway. Namely if B is bounded and T is compact then for some $c > 0$ (namely $1/\|B\|$ unless it is zero) cB maps $B(0,1)$ into itself. Thus $cTB = TcB$ is compact since the image of the unit ball under it is contained in the image of the unit ball under T; hence TB is also compact. Similarly BT is compact since B is continuous and then

$$BT(B(0,1)) \subset B(\overline{T(B(0,1))}) \text{ is compact} \tag{3.100}$$

since it is the image under a continuous map of a compact set. □

15. Weak convergence

It is convenient to formalize the idea that a sequence be bounded and that each of the $\langle u_n, e_k\rangle$, the sequence of coefficients of some particular Fourier-Bessel series, should converge.

DEFINITION 3.6. A sequence, $\{u_n\}$, in a Hilbert space, $\mathcal{H}$, is said to *converge weakly* to an element $u \in \mathcal{H}$ if it is bounded in norm and $\langle u_j, v\rangle \to \langle u, v\rangle$ converges in $\mathbb{C}$ for each $v \in \mathcal{H}$. This relationship is written

$$u_n \rightharpoonup u. \tag{3.101}$$

In fact as we shall see below, the assumption that $\|u_n\|$ is bounded and that u exists are both unnecessary. That is, a sequence converges weakly if and only if $\langle u_n, v\rangle$ converges in $\mathbb{C}$ for each $v \in \mathcal{H}$. Conversely, there is no harm in assuming it is bounded and that the 'weak limit' $u \in \mathcal{H}$ exists. Note that the weak limit is unique since if u and u' both have this property then $\langle u - u', v\rangle = \lim_{n\to\infty}\langle u_n, v\rangle - \lim_{n\to\infty}\langle u_n, v\rangle = 0$ for all $v \in \mathcal{H}$ and setting $v = u - u'$ it follows that $u = u'$.

LEMMA 3.9. *A (strongly) convergent sequence is weakly convergent with the same limit.*

PROOF. This is the continuity of the inner product. If $u_n \to u$ then

$$|\langle u_n, v\rangle - \langle u, v\rangle| \le \|u_n - u\|\|v\| \to 0 \tag{3.102}$$

for each $v \in \mathcal{H}$ shows weak convergence. □

LEMMA 3.10. *For a bounded sequence in a separable Hilbert space, weak convergence is equivalent to component convergence with respect to an orthonormal basis.*

PROOF. Let e_k be an orthonormal basis. Then if u_n is weakly convergent it follows immediately that $\langle u_n, e_k\rangle \to \langle u, e_k\rangle$ converges for each k. Conversely, suppose this is true for a bounded sequence, just that $\langle u_n, e_k\rangle \to c_k$ in $\mathbb{C}$ for each k. The norm boundedness and Bessel's inequality show that

$$\sum_{k\le p} |c_k|^2 = \lim_{n\to\infty} \sum_{k\le p} |\langle u_n, e_k\rangle|^2 \le \sup_n \|u_n\|^2 \tag{3.103}$$

for all p. Thus in fact $\{c_k\} \in l^2$ and hence

$$u = \sum_k c_k e_k \in \mathcal{H} \tag{3.104}$$

by the completeness of $\mathcal{H}$. Clearly $\langle u_n, e_k\rangle \to \langle u, e_k\rangle$ for each k. It remains to show that $\langle u_n, v\rangle \to \langle u, v\rangle$ for all $v \in \mathcal{H}$. This is certainly true for any finite linear combination of the e_k and for a general v we can write

$$\begin{aligned} \langle u_n, v\rangle - \langle u, v\rangle = \langle u_n, v_p\rangle - \langle u, v_p\rangle + \langle u_n, v - v_p\rangle - \langle u, v - v_p\rangle \Longrightarrow \\ |\langle u_n, v\rangle - \langle u, v\rangle| \le |\langle u_n, v_p\rangle - \langle u, v_p\rangle| + 2C\|v - v_p\| \end{aligned} \tag{3.105}$$

where $v_p = \sum\limits_{k\le p} \langle v, e_k\rangle e_k$ is a finite part of the Fourier-Bessel series for v and C is a bound for $\|u_n\|$. Now the convergence $v_p \to v$ implies that the last term in (3.105) can be made small by choosing p large, independent of n. Then the second last term can be made small by choosing n large since v_p is a finite linear combination of the e_k. Thus indeed, $\langle u_n, v\rangle \to \langle u, v\rangle$ for all $v \in \mathcal{H}$ and it follows that u_n converges weakly to u. □

Proposition 3.13. *Any bounded sequence $\{u_n\}$ in a separable Hilbert space has a weakly convergent subsequence.*

This can be thought of as different extension to infinite dimensions of the Heine-Borel theorem. As opposed to the characterization of compact sets above, which involves adding the extra condition of finite-dimensional approximability, here we weaken the notion of convergence.

Proof. Choose an orthonormal basis $\{e_k\}$ and apply the procedure in the proof of Proposition 3.8 to it. Thus, we may extract successive subsequence along the kth of which $\langle u_{n_p}, e_k\rangle \to c_k \in \mathbb{C}$. Passing to the diagonal subsequence, v_n, which is *eventually* a subsequence of each of these ensures that $\langle v_n, e_k\rangle \to c_k$ for each k. Now apply the preceeding Lemma to conclude that this subsequence converges weakly. □

Lemma 3.11. *For a weakly convergent sequence $u_n \rightharpoonup u$*

$$\|u\| \le \liminf \|u_n\| \tag{3.106}$$

and a weakly convergent sequence converges strongly if and only if the weak limit satisfies $\|u\| = \lim_{n\to\infty} \|u_n\|$.

Proof. Choose an orthonormal basis e_k and observe that

$$\sum_{k\le p} |\langle u, e_k\rangle|^2 = \lim_{n\to\infty} \sum_{k\le p} |\langle u_n, e_k\rangle|^2. \tag{3.107}$$

The sum on the right is bounded by $\|u_n\|^2$ independently of p so

$$\sum_{k\le p} |\langle u, e_k\rangle|^2 \le \liminf_n \|u_n\|^2 \tag{3.108}$$

by the definition of $\liminf$. Then let $p \to \infty$ to conclude that

$$\|u\|^2 \le \liminf_n \|u_n\|^2 \tag{3.109}$$

from which (3.106) follows.

Now, suppose $u_n \rightharpoonup u$ then

$$\|u - u_n\|^2 = \|u\|^2 - 2\operatorname{Re}\langle u, u_n\rangle + \|u_n\|^2. \tag{3.110}$$

Weak convergence implies that the middle term converges to $-2\|u\|^2$ so if the last term converges to $\|u\|^2$ then $u \to u_n$. □

Observe that for any $A \in \mathcal{B}(H)$, if $u_n \rightharpoonup u$ then $Au_n \rightharpoonup Au$ using the existence of the adjoint:-

$$\langle Au_n, v\rangle = \langle u_n, A^*v\rangle \to \langle u, A^*v\rangle = \langle Au, v\rangle \ \forall\ v \in \mathcal{H}. \tag{3.111}$$

Lemma 3.12. *An operator $K \in \mathcal{B}(\mathcal{H})$ is compact if and only if the image Ku_n of any weakly convergent sequence $\{u_n\}$ in $\mathcal{H}$ is strongly, i.e. norm, convergent.*

This is the origin of the old name ‘completely continuous’ for compact operators, since they turn even weakly convergent into strongly convergent sequences.

PROOF. First suppose that $u_n \rightharpoonup u$ is a weakly convergent sequence in $\mathcal{H}$ and that K is compact. We know that $\|u_n\| < C$ is bounded so the sequence Ku_n is contained in $CK(B(0,1))$ and hence in a compact set (clearly if D is compact then so is cD for any constant c.) Thus, any subsequence of Ku_n has a convergent subsequence and the limit is necessarily Ku since $Ku_n \rightharpoonup Ku$. But the condition on a sequence in a metric space that every subsequence of it has a subsequence which converges to a fixed limit implies convergence. (If you don't remember this, reconstruct the proof: To say a sequence v_n *does not* converge to v is to say that for some $\epsilon > 0$ there is a subsequence along which $d(v_{n_k}, v) \geq \epsilon$. This is impossible given the subsequence of subsequence condition converging to the fixed limit v.)

Conversely, suppose that K has this property of turning weakly convergent into strongly convergent sequences. We want to show that $K(B(0,1))$ has compact closure. This just means that any sequence in $K(B(0,1))$ has a (strongly) convergent subsequence – where we do not have to worry about whether the limit is in the set or not. Such a sequence is of the form Ku_n where u_n is a sequence in $B(0,1)$. However we know that we can pass to a subsequence which converges weakly, $u_{n_j} \rightharpoonup u$. Then, by the assumption of the Lemma, $Ku_{n_j} \to Ku$ converges strongly. Thus u_n does indeed have a convergent subsequence and hence $K(B(0,1))$ must have compact closure. □

As noted above, it is not really necessary to assume that a weakly convergent sequence in a Hilbert space is bounded, provided one has the Uniform Boundedness Principle, Theorem 1.3, at the ready.

PROPOSITION 3.14. *If $u_n \in H$ is a sequence in a Hilbert space and for all $v \in H$*

$$\langle u_n, v\rangle \to F(v) \text{ converges in } \mathbb{C} \tag{3.112}$$

then $\|u_n\|_H$ is bounded and there exists $w \in H$ such that $u_n \rightharpoonup w$.

PROOF. Apply the Uniform Boundedness Theorem to the continuous functionals

$$T_n(u) = \langle u, u_n\rangle,\ T_n : H \longrightarrow \mathbb{C} \tag{3.113}$$

where we reverse the order to make them linear rather than anti-linear. Thus, each set $|T_n(u)|$ is bounded in $\mathbb{C}$ since it is convergent. It follows from the Uniform Boundedness Principle that there is a bound

$$\|T_n\| \leq C. \tag{3.114}$$

However, this norm as a functional is just $\|T_n\| = \|u_n\|_H$ so the original sequence must be bounded in H. Define $T : H \longrightarrow \mathbb{C}$ as the limit for each u :

$$T(u) = \lim_{n\to\infty} T_n(u) = \lim_{n\to\infty} \langle u, u_n\rangle. \tag{3.115}$$

This exists for each u by hypothesis. It is a linear map and from (3.114) it is bounded, $\|T\| \leq C$. Thus by the Riesz Representation theorem, there exists $w \in H$ such that

$$T(u) = \langle u, w\rangle\ \forall\ u \in H. \tag{3.116}$$

Thus $\langle u_n, u\rangle \to \langle w, u\rangle$ for all $u \in H$ so $u_n \rightharpoonup w$ as claimed. □

16. The algebra $\mathcal{B}(H)$

Recall the basic properties of the Banach space, and algebra, of bounded operators $\mathcal{B}(\mathcal{H})$ on a separable Hilbert space $\mathcal{H}$. In particular that it is a Banach space with respect to the norm

$$\|A\| = \sup_{\|u\|_{\mathcal{H}}=1} \|Au\|_{\mathcal{H}} \tag{3.117}$$

and that the norm satisfies

$$\|AB\| \le \|A\|\|B\| \tag{3.118}$$

as follows from the fact that

$$\|ABu\| \le \|A\|\|Bu\| \le \|A\|\|B\|\|u\|.$$

Consider the set of invertible elements:

$$\mathrm{GL}(\mathcal{H}) = \{A \in \mathcal{B}(\mathcal{H}); \exists\ B \in \mathcal{B}(\mathcal{H}),\ BA = AB = \mathrm{Id}\}. \tag{3.119}$$

Note that this is equivalent to saying A is 1-1 and onto in view of the Open Mapping Theorem, Theorem 1.4.

This set is open, to see this consider a neighbourhood of the identity.

Lemma 3.13. *If $A \in \mathcal{B}(\mathcal{H})$ and $\|A\| < 1$ then*

$$\mathrm{Id} - A \in \mathrm{GL}(\mathcal{H}). \tag{3.120}$$

Proof. This follows from the convergence of the Neumann series. If $\|A\| < 1$ then $\|A^j\| \le \|A\|^j$, from (3.118), and it follows that

$$B = \sum_{j=0}^{\infty} A^j \tag{3.121}$$

(where $A^0 = \mathrm{Id}$ by definition) is absolutely summable in $\mathcal{B}(\mathcal{H})$ since $\sum\limits_{j=0}^{\infty} \|A^j\|$ converges. Since $\mathcal{B}(H)$ is a Banach space, the sum converges. Moreover by the continuity of the product with respect to the norm

$$AB = A \lim_{n\to\infty} \sum_{j=0}^{n} A^j = \lim_{n\to\infty} \sum_{j=1}^{n+1} A^j = B - \mathrm{Id} \tag{3.122}$$

and similalry $BA = B - \mathrm{Id}$. Thus $(\mathrm{Id} - A)B = B(\mathrm{Id} - A) = \mathrm{Id}$ shows that B is a (and hence *the*) 2-sided inverse of $\mathrm{Id} - A$. □

Proposition 3.15. *The invertible elements form an open subset* $\mathrm{GL}(\mathcal{H}) \subset \mathcal{B}(\mathcal{H})$.

Proof. Suppose $G \in \mathrm{GL}(\mathcal{H})$, meaning it has a two-sided (and unique) inverse $G^{-1} \in \mathcal{B}(\mathcal{H})$:

$$G^{-1}G = GG^{-1} = \mathrm{Id}\,. \tag{3.123}$$

Then we wish to show that $B(G; \epsilon) \subset \mathrm{GL}(\mathcal{H})$ for some $\epsilon > 0$. In fact we shall see that we can take $\epsilon = \|G^{-1}\|^{-1}$. To show that $G + B$ is invertible set

$$E = -G^{-1}B \Longrightarrow G + B = G(\mathrm{Id} + G^{-1}B) = G(\mathrm{Id} - E) \tag{3.124}$$

From Lemma 3.13 we know that

$$\|B\| < 1/\|G^{-1}\| \Longrightarrow \|G^{-1}B\| < 1 \Longrightarrow \mathrm{Id} - E \text{ is invertible.} \tag{3.125}$$

Then $(\mathrm{Id}-E)^{-1}G^{-1}$ satisfies

$$(\mathrm{Id}-E)^{-1}G^{-1}(G+B)=(\mathrm{Id}-E)^{-1}(\mathrm{Id}-E)=\mathrm{Id}\,. \tag{3.126}$$

Moreover $E'=-BG^{-1}$ also satisfies $\|E'\|\le\|B\|\|G^{-1}\|<1$ and

$$(G+B)G^{-1}(\mathrm{Id}-E')^{-1}=(\mathrm{Id}-E')(\mathrm{Id}-E')^{-1}=\mathrm{Id}\,. \tag{3.127}$$

Thus $G+B$ has both a 'left' and a 'right' inverse. The associativity of the operator product (that $A(BC)=(AB)C$) then shows that

$$G^{-1}(\mathrm{Id}-E')^{-1}=(\mathrm{Id}-E)^{-1}G^{-1}(G{+}B)G^{-1}(\mathrm{Id}-E')^{-1}=(\mathrm{Id}-E)^{-1}G^{-1} \tag{3.128}$$

so the left and right inverses are equal and hence $G+B$ is invertible. □

Thus $\mathrm{GL}(\mathcal{H})\subset\mathcal{B}(\mathcal{H})$, the set of invertible elements, is open. It is also a group – since the inverse of G_1G_2 if $G_1,\ G_2\in\mathrm{GL}(\mathcal{H})$ is $G_2^{-1}G_1^{-1}$.

This group of invertible elements has a smaller subgroup, $\mathrm{U}(\mathcal{H})$, the unitary group, defined by

$$\mathrm{U}(\mathcal{H})=\{U\in\mathrm{GL}(\mathcal{H});U^{-1}=U^*\}. \tag{3.129}$$

The unitary group consists of the linear isometric isomorphisms of $\mathcal{H}$ onto itself – thus

$$\langle Uu,Uv\rangle=\langle u,v\rangle,\ \|Uu\|=\|u\|\ \forall\ u,v\in\mathcal{H},\ U\in\mathrm{U}(\mathcal{H}). \tag{3.130}$$

This is an important object and we will use it a little bit later on.

The groups $\mathrm{GL}(H)$ and $\mathrm{U}(H)$ for a separable Hilbert space may seem very similar to the familiar groups of invertible and unitary $n\times n$ matrices, $\mathrm{GL}(n)$ and $\mathrm{U}(n)$, but this is somewhat deceptive. For one thing they are much bigger. In fact there are other important qualitative differences. One important fact that you should know, and there is a proof towards the end of this chapter, is that both $\mathrm{GL}(H)$ and $\mathrm{U}(\mathcal{H})$ are contractible as metric spaces – they have no significant topology. This is to be constrasted with the $\mathrm{GL}(n)$ and $\mathrm{U}(n)$ which have a lot of topology, and are not at all simple spaces – especially for large n. One upshot of this is that $\mathrm{U}(\mathcal{H})$ does not look much like the limit of the $\mathrm{U}(n)$ as $n\to\infty$. In fact there is another group which *is* essentially the large n limit of the $\mathrm{U}(n)$, namely

$$\mathrm{U}^{-\infty}(H)=\{\mathrm{Id}+K\in\mathrm{U}(H);K\in\mathcal{K}(H)\}. \tag{3.131}$$

It does have lots of interesting (and useful) topology.

Another important fact that we will discuss below is that $\mathrm{GL}(H)$ is *not* dense in $\mathcal{B}(H)$, in contrast to the finite dimensional case. In other words there are operators which are not invertible and cannot be made invertible by small perturbations.

17. Spectrum of an operator

Another direct application of Lemma 3.13, the convergence of the Neumann series, is that if $A\in\mathcal{B}(H)$ and $\lambda\in\mathbb{C}$ has $|\lambda|>\|A\|$ then $\|\lambda^{-1}A\|<1$ so $(\mathrm{Id}-\lambda^{-1}A)^{-1}$ exists and satisfies

$$(\lambda\,\mathrm{Id}-A)\lambda^{-1}(\mathrm{Id}-\lambda^{-1}A)^{-1}=\mathrm{Id}=\lambda^{-1}(\mathrm{Id}-\lambda^{-1}A)^{-1}(\lambda-A). \tag{3.132}$$

Thus, $\lambda\,\mathrm{Id}-A\in\mathrm{GL}(H)$, which we usually abbreviate to $\lambda-A$, has inverse $(\lambda-A)^{-1}=\lambda^{-1}(\mathrm{Id}-\lambda^{-1}A)^{-1}$. The set of λ for which this operator is invertible is called the *resolvent set* and we have shown

$$\begin{aligned} \operatorname{Res}(A) = \{\lambda \in \mathbb{C}; (\lambda \operatorname{Id} - A) \in \operatorname{GL}(H)\} \subset \mathbb{C} \\ \{|\lambda| > \|A\|\} \subset \operatorname{Res}(A). \end{aligned} \tag{3.133}$$

From the discussion above, it is an open, and non-empty, set on which $(A-\lambda)^{-1}$, called the resolvent of A, is defined. The complement of the resolvent set is called the spectrum of A

$$\operatorname{Spec}(A) = \{\lambda \in \mathbb{C}; \lambda \operatorname{Id} - A \notin \operatorname{GL}(H)\} \subset \{\lambda \in \mathbb{C}; |\lambda| \le \|A\|\}. \tag{3.134}$$

As follows from the discussion above it is a compact set – in fact it cannot be empty. This is quite easy to see if you know a little complex analysis since it follows from Liouville's Theorem. One way to show that $\lambda \in \operatorname{Spec}(A)$ is to check that $\lambda - A$ is not injective, since then it cannot be invertible. This means precisely that λ is an *eigenvalue* of A :

$$\exists\ 0 \ne u \in \mathcal{H} \text{ s.t. } Au = \lambda u. \tag{3.135}$$

However, you should strongly resist the temptation to think that the spectrum is the set of eigenvalues of A, this is sometimes but by no means always true. The other way to show that $\lambda \in \operatorname{Spec}(A)$ is to prove that $\lambda - A$ is not *surjective.* Note that by the Open Mapping Theorem if $\lambda - A$ is both surjective and injective then $\lambda \in \operatorname{Res}(A)$.

For a finite rank operator the spectrum does consist of the set of eigenvalues.

For a bounded self-adjoint operator we can say more quite a bit more.

Proposition 3.16. *If $A : H \longrightarrow H$ is a bounded operator on a Hilbert space and $A^* = A$ then $A - \lambda \operatorname{Id}$ is invertible for all $\lambda \in \mathbb{C} \setminus [-\|A\|, \|A\|]$ and conversely at least one of $A - \|A\| \operatorname{Id}$ and $A + \|A\| \operatorname{Id}$ is not invertible.*

The proof of this depends on a different characterization of the norm in the self-adjoint case.

Lemma 3.14. *If $A^* = A \in \mathcal{B}(H)$ then*

$$\|A\| = \sup_{\|u\|=1} |\langle Au, u\rangle|. \tag{3.136}$$

Proof. Certainly, $|\langle Au, u\rangle| \le \|A\|\|u\|^2$ so the right side can only be smaller than or equal to the left. Set

$$a = \sup_{\|u\|=1} |\langle Au, u\rangle| \le \|A\|.$$

Then for any $u, v \in H$, $|\langle Au, v\rangle| = \langle Ae^{i\theta}u, v\rangle$ for some $\theta \in [0, 2\pi)$, so we can arrange that $\langle Au, v\rangle = |\langle Au', v\rangle|$ is non-negative and $\|u'\| = 1 = \|u\| = \|v\|$. Dropping the primes and computing using the polarization identity

$$\begin{aligned} &4\langle Au, v\rangle \\ &= \langle A(u+v), u+v\rangle - \langle A(u-v), u-v\rangle + i\langle A(u+iv), u+iv\rangle - i\langle A(u-iv), u-iv\rangle. \end{aligned} \tag{3.137}$$

By the reality of the left side we can drop the last two terms and use the bound $|\langle Aw, w\rangle| \le a\|w\|^w$ on the first two to see that

$$4\langle Au, v\rangle \le a(\|u+v\|^2 + \|u-v\|^2) = 2a(\|u\|^2 + \|v\|^2) = 4a \tag{3.138}$$

Thus, $\|A\| = \sup_{\|u\|=\|v\|=1} |\langle Au, v\rangle| \le a$ and hence $\|A\| = a$. □

This suggests an improvement on the last part of the statement of Proposition 3.16, namely

$$\begin{gathered}\text{If } A^* = A \in \mathcal{B}(H) \text{ then } a_-, a_+ \in \operatorname{Spec}(A) \text{ and } \operatorname{Spec}(A) \subset [a_-, a_+] \\ \text{where } a_- = \inf_{\|u\|=1} \langle Au, u\rangle,\ a_+ = \sup_{\|u\|=1} \langle Au, u\rangle.\end{gathered} \tag{3.139}$$

Observe that Lemma 3.14 shows that $\|A\| = \max(a_+, -a_-)$.

Proof of Proposition 3.16. First we show that if $A^* = A$ then $\operatorname{Spec}(A) \subset \mathbb{R}$. Thus we need to show that if $\lambda = s + it$ where $t \neq 0$ then $A - \lambda$ is invertible. Now $A - \lambda = (A - s) - it$ and $A - s$ is bounded and selfadjoint, so it is enough to consider the special case that $\lambda = it$. Then for any $u \in \mathcal{H}$,

$$\operatorname{Im}\langle (A - it)u, u\rangle = -t\|u\|^2. \tag{3.140}$$

So, certainly $A - it$ is injective, since $(A - it)u = 0$ implies $u = 0$ if $t \neq 0$. The adjoint of $A - it$ is $A + it$ so the adjoint is injective too. It follows from (3.71) that the range of $A - it$ is dense in $\mathcal{H}$. By this density of the range, if $w \in H$ there exists a sequence $u_n \in \mathcal{H}$ with $w_n = (A - it)u_n \to w$. So again we find that

$$\begin{gathered} |\operatorname{Im}\langle (A - it)\langle u_n - u_m), (u_n - u_m)\rangle| = |t|\|u_n - u_m\|^2 \\ = |\operatorname{Im}\langle (w_n - w_m), (u_n - u_m)\rangle| \le \|w_n - w_m\|\|u_n - u_m\| \\ \Longrightarrow \|u_n - u_m\| \le \frac{1}{|t|}\|w_n - w_m\|. \end{gathered} \tag{3.141}$$

Since $w_n \to w$ it is a Cauchy sequence and hence u_n is Cauchy so by completeness, $u_n \to u$ and hence $(A - it)u = w$. Thus $A - it$ is 1-1 and onto and $\|A^{-1}\| \le 1/|t|$. So we have shown that $\operatorname{Spec}(A) \subset \mathbb{R}$.

We already know that $\operatorname{Spec}(A) \subset \{z \in \mathbb{C}; |z| \le \|A\|\}$ so finally then we need to show that one of $A \pm \|A\| \operatorname{Id}$ is NOT invertible. This follows from (3.136). Indeed, by the definition of sup there is a sequence $u_n \in H$ with $\|u_n\| = 1$ such that either $\langle Au_n, u_n\rangle \to \|A\|$ or $\langle Au_n, u_n\rangle \to -\|A\|$. Assume we are in the first case, so $\langle Au_n, u_n\rangle \to \|A\|$. Then

$$\begin{gathered} \|(A - \|A\|)u_n\|^2 = \|Au_n\|^2 - 2\|A\|\langle Au_n, u_n\rangle + \|A\|^2\|u_n\|^2 \\ \le 2\|A\|^2 - 2\|A\|\langle Au_n, u_n\rangle \to 0. \end{gathered} \tag{3.142}$$

Since the sequence is positive it follows that $\|(A - \|A\|)u_n\| \to 0$. This means that $A - \|A\| \operatorname{Id}$ is not invertible, since if it had a bounded inverse B then $1 = \|u_n\| \le \|B\|\|(A - \|A\|)u_n\|$ which is impossible. In the other case it follows similarly that $A + \|A\|$ is not invertible, or one can replace A by $-A$ and use the same argument. So one of $A \pm \|A\|$ is not invertible. □

Only slight modifications of this proof are needed to give (3.139) which we restate in a slightly different form.

Lemma 3.15. *If $A = A^* \in \mathcal{B}(H)$ then*

$$\operatorname{Spec}(A) \subset [\alpha-, \alpha_+] \Longleftrightarrow \alpha_- \le \langle Au, u\rangle \le \alpha_+ \ \forall\ u \in H,\ \|u\| = 1. \tag{3.143}$$

Proof. Take $a_\pm$ to be defined as in (3.139) then set $b = (a_+ - a_-)/2$ and consider $B = A - b \operatorname{Id}$ which is self-adjoint and clearly satisfies

$$\sup_{\|u\|=1} |\langle Bu, u\rangle| = b \tag{3.144}$$

Thus $\|B\| = b$ and $\operatorname{Spec}(B) \subset [-b, b]$ and the argument in the proof above shows that both end-points are in the spectrum. It follows that

$$\{a_-\} \cup \{a_+\} \subset \operatorname{Spec}(A) \subset [a_-, a_+] \tag{3.145}$$

from which the statement follows. □

In particular if $A = A^*$ then

$$\operatorname{Spec}(A) \subset [0, \infty) \Longleftrightarrow \langle Au, u \rangle \geq 0. \tag{3.146}$$

18. Spectral theorem for compact self-adjoint operators

One of the important differences between a general bounded self-adjoint operator and a compact self-adjoint operator is that the latter has eigenvalues and eigenvectors – lots of them.

Theorem 3.4. *If $A \in \mathcal{K}(\mathcal{H})$ is a self-adjoint, compact operator on a separable Hilbert space, so $A^* = A$, then H has an orthonormal basis consisting of eigenvectors of A, u_j such that*

$$Au_j = \lambda_j u_j,\ \lambda_j \in \mathbb{R} \setminus \{0\}, \tag{3.147}$$

combining an orthonormal basis for the possibly infinite-dimensional (closed) null space and eigenvectors with non-zero eigenvalues which can be arranged into a sequence such that $|\lambda_j|$ is non-increasing and $\lambda_j \to 0$ as $j \to \infty$ (in case $\operatorname{Nul}(A)^\perp$ *is finite dimensional, this sequence is finite).*

The operator A maps $\operatorname{Nul}(A)^\perp$ into itself so it may be clearer to first split off the null space and then look at the operator acting on $\operatorname{Nul}(A)^\perp$ which has an orthonormal basis of eigenvectors with non-vanishing eigenvalues.

Before going to the proof, let's notice some useful conclusions. One is that we have 'Fredholm's alternative' in this case.

Corollary 3.3. *If $A \in \mathcal{K}(\mathcal{H})$ is a compact self-adjoint operator on a separable Hilbert space then the equation*

$$u - Au = f \tag{3.148}$$

either has a unique solution for each $f \in \mathcal{H}$ or else there is a non-trivial finite dimensional space of solutions to

$$u - Au = 0 \tag{3.149}$$

and then (3.148) *has a solution if and only if f is orthogonal to all these solutions.*

Proof. This is just saying that the null space of $\operatorname{Id} - A$ is a complement to the range – which is closed. So, either $\operatorname{Id} - A$ is invertible or if not then the range is precisely the orthocomplement of $\operatorname{Nul}(\operatorname{Id} - A)$. You might say there is not much alternative from this point of view, since it just says the range is *always* the orthocomplement of the null space. □

Let me separate off the heart of the argument from the bookkeeping.

Lemma 3.16. *If $A \in \mathcal{K}(\mathcal{H})$ is a self-adjoint compact operator on a separable (possibly finite-dimensional) Hilbert space then*

$$F(u) = \langle Au, u \rangle,\ F : \{u \in \mathcal{H}; \|u\| = 1\} \longrightarrow \mathbb{R} \tag{3.150}$$

is a continuous function on the unit sphere which attains its supremum and infimum where

$$\sup_{\|u\|=1} |F(u)| = \|A\|. \tag{3.151}$$

Furthermore, if the maximum or minimum of $F(u)$ is non-zero it is attained at an eivenvector of A with this extremal value as eigenvalue.

Proof. Since $|F(u)|$ is the function considered in (3.136), (3.151) is a direct consequence of Lemma 3.14. Moreover, continuity of F follows from continuity of A and of the inner product so

$$|F(u)-F(u')| \le |\langle Au,u\rangle - \langle Au,u'\rangle| + |\langle Au,u'\rangle - \langle Au',u'\rangle| \le 2\|A\|\|u-u'\| \tag{3.152}$$

since both u and u' have norm one.

If we were in finite dimensions this almost finishes the proof, since the sphere is then compact and a continuous function on a compact set attains its supremum and infimum. In the general case we need to use the compactness of A. Certainly F is bounded,

$$|F(u)| \le \sup_{\|u\|=1} |\langle Au,u\rangle| \le \|A\|. \tag{3.153}$$

Thus, there is a sequence u_n^+ such that $F(u_n^+) \to \sup F$ and another u_n^- such that $F(u_n^-) \to \inf F$. The properties of weak convergence mean that we can pass to a weakly convergent subsequence in each case, and so assume that $u_n^\pm \rightharpoonup u^\pm$ converges weakly; then $\|u^\pm\| \le 1$ by the properties of weak convergence. The compactness of A means that $Au_n^\pm \to Au^\pm$ converges strongly, i.e. in norm. But then we can write

$$\begin{aligned} |F(u_n^\pm) - F(u^\pm)| &\le |\langle A(u_n^\pm - u^\pm), u_n^\pm\rangle| + |\langle Au^\pm, u_n^\pm - u^\pm\rangle| \\ &= |\langle A(u_n^\pm - u^\pm), u_n^\pm\rangle| + |\langle u^\pm, A(u_n^\pm - u^\pm)\rangle| \le 2\|Au_n^\pm - Au^\pm\| \end{aligned} \tag{3.154}$$

to deduce that $F(u^\pm) = \lim F(u_n^\pm)$ are respectively the supremum and infimum of F. Thus indeed, as in the finite dimensional case, the supremum and infimum are attained, and hence are the max and min. Note that this is *not* typically true if A is not compact as well as self-adjoint.

Now, suppose that $\Lambda^+ = \sup F > 0$. Then for any $v \in \mathcal{H}$ with $v \perp u^+$ and $\|v\| = 1$, the curve

$$L_v : (-\pi,\pi) \ni \theta \longmapsto \cos\theta u^+ + \sin\theta v \tag{3.155}$$

lies in the unit sphere. Expanding out

$$\begin{aligned} &F(L_v(\theta)) = \\ &\quad \langle AL_v(\theta), L_v(\theta)\rangle = \cos^2\theta F(u^+) + \sin(2\theta)\operatorname{Re}(Au^+,v) + \sin^2(\theta)F(v) \end{aligned} \tag{3.156}$$

we know that this function must take its maximum at $\theta = 0$. The derivative there (it is certainly continuously differentiable on $(-\pi,\pi)$) is $2\operatorname{Re}\langle Au^+,v\rangle$ which must therefore vanish. The same is true for iv in place of v so in fact

$$\langle Au^+, v\rangle = 0 \ \forall\ v \perp u^+,\ \|v\| = 1. \tag{3.157}$$

Taking the span of these v's it follows that $\langle Au^+, v\rangle = 0$ for all $v \perp u^+$ so Au^+ must be a multiple of u^+ itself. Inserting this into the definition of F it follows that $Au^+ = \Lambda^+ u^+$ is an eigenvector with eigenvalue $\Lambda^+ = \sup F$.

The same argument applies to $\inf F$ if it is negative, for instance by replacing A by $-A$. This completes the proof of the Lemma. $\square$

Proof of Theorem 3.4. First consider the Hilbert space $\mathcal{H}_0 = \text{Nul}(A)^\perp \subset \mathcal{H}$. Then, as noted above, A maps $\mathcal{H}_0$ into itself, since

$$\langle Au, v\rangle = \langle u, Av\rangle = 0 \ \forall \ u \in \mathcal{H}_0,\ v \in \text{Nul}(A) \Longrightarrow Au \in \mathcal{H}_0. \tag{3.158}$$

Moreover, A_0, which is A restricted to $\mathcal{H}_0$, is again a compact self-adjoint operator – where the compactness follows from the fact that $A(B(0,1))$ for $B(0,1) \subset \mathcal{H}_0$ is smaller than (actually of course equal to) the whole image of the unit ball.

Thus we can apply the Lemma above to A_0, with quadratic form F_0, and find an eigenvector. Let's agree to take the one associated to $\sup F_0$ unless $\sup F_0 < -\inf F_0$ in which case we take one associated to the inf . Now, what can go wrong here? Nothing except if $F_0 \equiv 0$. However in that case we know from Lemma 3.14 that $\|A\| = 0$ so $A = 0$.

So, we now know that we can find an eigenvector with non-zero eigenvalue unless $A \equiv 0$ which would implies $\text{Nul}(A) = \mathcal{H}$. Now we proceed by induction. Suppose we have found N mutually orthogonal eigenvectors e_j for A all with norm 1 and eigenvectors λ_j – an orthonormal set of eigenvectors and all in $\mathcal{H}_0$. Then we consider

$$\mathcal{H}_N = \{u \in \mathcal{H}_0 = \text{Nul}(A)^\perp; \langle u, e_j\rangle = 0,\ j = 1, \dots, N\}. \tag{3.159}$$

From the argument above, A maps $\mathcal{H}_N$ into itself, since

$$\langle Au, e_j\rangle = \langle u, Ae_j\rangle = \lambda_j \langle u, e_j\rangle = 0 \text{ if } u \in \mathcal{H}_N \Longrightarrow Au \in \mathcal{H}_N. \tag{3.160}$$

Moreover this restricted operator is self-adjoint and compact on $\mathcal{H}_N$ as before so we can again find an eigenvector, with eigenvalue either the max of min of the new F for $\mathcal{H}_N$. This process will not stop uness $F \equiv 0$ at some stage, but then $A \equiv 0$ on $\mathcal{H}_N$ and since $\mathcal{H}_N \perp \text{Nul}(A)$ which implies $\mathcal{H}_N = \{0\}$ so $\mathcal{H}_0$ must have been finite dimensional.

Thus, either $\mathcal{H}_0$ is finite dimensional or we can grind out an infinite orthonormal sequence e_i of eigenvectors of A in $\mathcal{H}_0$ with the corresponding sequence of eigenvalues such that $|\lambda_i|$ is non-increasing – since the successive F_N's are restrictions of the previous ones the max and min are getting closer to (or at least no further from) 0.

So we need to rule out the possibility that there is an infinite orthonormal sequence of eigenfunctions e_j with corresponding eigenvalues λ_j where $\inf_j |\lambda_j| = a > 0$. Such a sequence cannot exist since $e_j \rightharpoonup 0$ so by the compactness of A, $Ae_j \to 0$ (in norm) but $\|Ae_j\| \geq a$ which is a contradiction. Thus if $\text{null}(A)^\perp$ is not finite dimensional then the sequence of eigenvalues constructed above must converge to 0.

Finally then, we need to check that this orthonormal sequence of eigenvectors constitutes an orthonormal basis of $\mathcal{H}_0$. If not, then we can form the closure of the span of the e_i we have constructed, $\mathcal{H}'$, and its orthocomplement in $\mathcal{H}_0$ – which would have to be non-trivial. However, as before F restricts to this space to be F' for the restriction of A' to it, which is again a compact self-adjoint operator. So, if F' is not identically zero we can again construct an eigenfunction, with non-zero eigenvalue, which contradicts the fact the we are always choosing a largest eigenvalue, in absolute value at least. Thus in fact $F' \equiv 0$ so $A' \equiv 0$ and the eigenvectors form and orthonormal basis of $\operatorname{Nul}(A)^\perp$. This completes the proof of the theorem. □

19. Functional Calculus

As we have seen, the non-zero eigenvalues of a compact self-adjoint operator A form the image of a sequence in $[-\|A\|,\|A\|]$ either converging to zero or finite. If e_j is an orthonormal sequence of eigenfunctions which spans $\operatorname{Nul}(A)^\perp$ with associated eigenvalues λ_i then

$$A=\sum_i \lambda_i P_i,\ P_i u=\langle u,e_i\rangle e_i \tag{3.161}$$

being the projection onto the span $\mathbb{C}e_i$. Since $P_iP_j=0$ if $i\neq j$ and $P_i^2=P_i$ it follows inductively that the positive powers of A are given by similar sums converging in $\mathcal{B}(H)$:

$$A^k=\sum_i \lambda_i^k P_i,\ P_i u=\langle u,e_i\rangle e_i,\ k\in\mathbb{N}. \tag{3.162}$$

There is a similar formula for the identity of course, except we need to remember that the null space of A then appears (and the series does *not* usually converge in the norm topology on $\mathcal{B}(H)$) :

$$\operatorname{Id}=\sum_i P_i+P_N,\ N=\operatorname{Nul}(A). \tag{3.163}$$

The sum (3.163) can be interpreted in terms of a *strong limit* of operators, meaning that the result converges when applied term by term to an element of H, so

$$u=\sum_i P_i u+P_N u,\ \forall\ u\in H \tag{3.164}$$

which is a form of the Fourier-Bessel series. Combining these formulæ we see that for any polynomial $p(z)$

$$p(A)=\sum_i p(\lambda_i)P_i+p(0)P_N \tag{3.165}$$

converges strongly, and in norm provided $p(0)=0$.

In fact we can do this more generally, by choosing $f\in\mathcal{C}([-\|A\|,\|A\|)$ and defining an operator by

$$f(A)\in\mathcal{B}(H),\ f(A)u=\sum_i f(\lambda_i)(u,e_i)e_i \tag{3.166}$$

This series converges in the norm topology provided $f(0)=0$ so to a compact operator and if f is real it is self-adjoint. You can easily check that, always for $A=A^*$ compact here, this formula defines a bounded linear map

$$\mathcal{C}([-\|A\|,\|A\|]\longrightarrow\mathcal{B}(H) \tag{3.167}$$

which has nice properties. Most importantly

$$(fg)(A)=f(A)g(A),\ (f(A))^*=\bar{f}(A) \tag{3.168}$$

so it takes the product of two continuous functions to the product of the operators.

We will proceed to show that such a map exists for any bounded self-adjoint operator. Even though it may not have eigenfunctions – or even if it does, it might not have an orthonormal basis of eigenvectors. Even so, it is still possible to define $f(A)$ for a continous function defined on $[a_-, a_+]$ if $\operatorname{Spec}(A) \subset [a_-, a_+]$. (In fact it only has to be defined on the compact set $\operatorname{Spec}(A)$ which might be quite a lot smaller). This is an effective extension of the spectral theorem to the case of non-compact self-adjoint operators.

How does one define $f(A)$? Well, it is easy enough in case f is a polynomial, since then we can simply substitute A^n in place of z^n. If we factorize the polynomial this is the same as setting

$$f(z) = c(z-z_1)(z-z_2)\dots(z-z_N) \Longrightarrow f(A) = c(A-z_1)(A-z_2)\dots(A-z_N) \tag{3.169}$$

and this is equivalent to (3.166) in case A is also compact.

Notice that the result does not depend on the order of the factors or anything like that. To pass to the case of a general continuous function we need to estimate the norm in the polynomial case.

Proposition 3.17. *If $A = A^* \in \mathcal{B}(H)$ is a bounded self-adjoint operator on a Hilbert space then for any polynomial with real coefficients*

$$\|f(A)\| \le \sup_{z\in[a_-,a_+]} |f(z)|,\ \operatorname{Spec}(A) \subset [a_-, a_+]. \tag{3.170}$$

Proof. For a polynomial we have defined $f(A)$ by (3.169). We can drop the constant c since it will just contribute a factor of $|c|$ to both sides of (3.170). Now, recall from Lemma 3.14 that for a self-adjoint operator the norm can be realized as

$$\|f(A)\| = \sup\{|t|; t \in \operatorname{Spec}(f(A))\}. \tag{3.171}$$

That is, we need to think about when $f(A) - t$ is invertible. However, $f(z) - t$ is another polynomial (with leading term z^N because we normalized the leading coefficient to be 1). Thus it can also be factorized:

$$\begin{aligned} f(z) - t &= \prod_{j=1}^{N}(z - \zeta_j(t)), \\ f(A) - t &= \prod_{j=1}^{N}(A - \zeta_j(t)) \end{aligned} \tag{3.172}$$

where the $\zeta_j \in \mathbb{C}$ are the roots (which might be complex even though the polynomial is real). Written in this way we can see that

$$(f(A) - t)^{-1} = \prod_{j=1}^{N}(A - \zeta_j(t))^{-1} \text{ if } \zeta_j(t) \notin \operatorname{Spec}(A)\ \forall\ j. \tag{3.173}$$

Indeed the converse is also true, i.e. the inverse exists if and only if all the $A - \zeta_j(t)$ are invertible, but in any case we see that

$$\operatorname{Spec}(f(A)) \subset \{t \in \mathbb{C}; \zeta_j(t) \in \operatorname{Spec}(A),\ \text{for some } j = 1, \dots, N\} \tag{3.174}$$

since if t is not in the right side then $f(A) - t$ is invertible.

Now this can be restated as

$$\operatorname{Spec}(f(A)) \subset f(\operatorname{Spec}(A)) \tag{3.175}$$

since $t \notin f(\operatorname{Spec}(A))$ means $f(z) \neq t$ for $z \in \operatorname{Spec}(A)$ which means that there is no root of $f(z) = t$ in $\operatorname{Spec}(A)$ and hence (3.174) shows that $t \notin \operatorname{Spec}(f(A))$. In fact it is easy to see that there is equality in (3.175).

Then (3.170) follows from (3.171), the norm is the sup of $|z|$, for $z \in \operatorname{Spec}(f(A))$ so

$$\|f(A)\| \le \sup_{t \in \operatorname{Spec}(A)} |f(t)|.$$

□

This allows one to pass by continuity to f in the uniform closure of the polynomials, which by the Stone-Weierstrass theorem is the whole of $\mathcal{C}([a_-, a_+])$.

THEOREM 3.5. *If $A = A^* \in \mathcal{B}(H)$ for a Hilbert space H then the map defined on polynomials, through* (3.169) *extends by continuity to a bounded linear map*

$$\mathcal{C}([a_-, a_+]) \longrightarrow \mathcal{B}(H) \text{ if } \operatorname{Spec}(A) \subset [a_-, a_+],\ \operatorname{Spec}(f(A)) \subset f([a_-, a_+]). \tag{3.176}$$

PROOF. By the Stone-Weierstrass theorem polynomials are dense in continous functions on any compact interval, in the supremum norm. □

REMARK 3.1. You should check the properties of this map, which also follow by continuity, especially that (3.168) holds in this more general context. In particular, $f(A)$ is self-adjoint if $f \in \mathcal{C}([a_-, a_+])$ is real-valued and is non-negative if $f \ge 0$ on $\operatorname{Spec}(A)$.

20. Spectral projection

I have not discussed this in lectures but it is natural at this point to push a little further towards the full spectral theorem for bounded self-adjoint operators. If $A \in \mathcal{B}(H)$ is self-adjoint, and $[a_-, a_+] \supset \operatorname{Spec}(A)$, we have defined $f(A) \in \mathcal{B}(H)$ for $A \in \mathcal{C}([a_-, a_+])$ real-valued and hence, for each $u \in H$,

$$\mathcal{C}([a_-, a_+]) \ni f \longmapsto \langle f(A)u, u \rangle \in \mathbb{R}. \tag{3.177}$$

Thinking back to the treatment of the Lebesgue integral, you can think of this as a replacement for the Riemann integral and ask whether it can be extended further, to functions which are not necessarily continuous.

In fact (3.177) is essentially given by a Riemann-Stieltjes integral and this suggests finding the increasing function which defines it. Of course we have the rather large issue that this depends on a vector in Hilbert space as well – clearly we want to allow this to vary too.

One direct approach is to try to define the 'integral' of the characteristic function $(-\infty, a]$ for fixed $a \in \mathbb{R}$. To do this is consider

$$Q_a(u) = \inf\{\langle f(A)u, u \rangle; f \in \mathcal{C}([a_-, a_+]),\ f(t) \ge 0,\ f(t) \ge 1 \text{ on } [a_-, a]\}. \tag{3.178}$$

Since $f \ge 0$ we know that $\langle f(A)u, u \rangle \ge 0$ so the infimum exists and is non-negative. In fact there must exist a sequence f_n such that

$$Q_a(u) = \lim \langle f_n(A)u, u \rangle,\ f_n \in \mathcal{C}([a_-, a_+]),\ f_n \ge 0,\ f_n(t) \ge 1,\ a_- \le t \le a, \tag{3.179}$$

where the sequence f_n could depend on u. Consider an obvious choice for f_n given what we did earlier, namely

$$f_n(t) = \begin{cases} 1 & a_- \le t \le a \\ 1 - (t-a)/n & a \le t \le a + 1/n \\ 0 & t > a + 1/n. \end{cases} \tag{3.180}$$

Certainly

$$Q_a(u) \le \lim \langle f_n(A)u, u \rangle \tag{3.181}$$

where the limit exists since the sequence is decreasing.

Lemma 3.17. *For any $a \in [a_-, a_+]$,*

$$Q_a(u) = \lim \langle f_n(A)u, u\rangle. \tag{3.182}$$

Proof. For any given f as in (3.178), and any $\epsilon > 0$ there exists n such that $f(t) \geq 1/(1+\epsilon)$ in $a \leq t \leq a + 1/n$, by continuity. This means that $(1+\epsilon)f \geq g_n$ and hence $\langle f(A)u, u\rangle \geq (1+\epsilon)^{-1}\langle f_n(A)u, u\rangle$ from which (3.182) follows, given (3.181). □

Thus in fact one sequence gives the infimum for all u. Now, use the polarization identity to define

$$Q_a(u, v) = \frac{1}{4}\left(Q_a(u+v) - Q_a(u-v) + iQ_a(u+iv) - iQ_a(u-iv)\right). \tag{3.183}$$

The corresponding identity holds for $\langle f_n(A)u, v\rangle$ so in fact

$$Q_a(u, v) = \lim_{n\to\infty} \langle f_n(A)u, v\rangle. \tag{3.184}$$

It follows that $Q_a(u, v)$ is a sesquilinear form, linear in the first variable and antilinear in the second. Moreover the $f_n(A)$ are uniformly bounded in $\mathcal{B}(H)$ (with norm 1 in fact) so

$$|Q_a(u, v)| \leq C\|u\|\|v\|. \tag{3.185}$$

Now, using the linearity in v of $\overline{Q_a(u, v)}$ and the Riesz Representation theorem it follows that for each $u \in H$ there exists a unique $Q_a u \in H$ such that

$$Q_a(u, v) = \langle Q_a u, v\rangle,\ \forall\ v \in H,\ \|Q_a u\| \leq \|u\|. \tag{3.186}$$

From the uniqueness, $H \ni u \longmapsto Q_a u$ is linear so (3.186) shows that it is a bounded linear operator. Thus we have proved most of

Proposition 3.18. *For each $a \in [a_-, a_+] \supset \operatorname{Spec}(A)$ there is a uniquely defined operator $Q_a \in \mathcal{B}(H)$ such that*

$$Q_a(u) = \langle Q_a u, u\rangle \tag{3.187}$$

recovers (3.182) *and $Q_a^* = Q_a = Q_a^2$ is a projection satisfying*

$$Q_a Q_b = Q_b Q_a = Q_b \text{ if } b \leq a,\ [Q_a, f(A)] = 0\ \forall\ f \in \mathcal{C}([a_-, a_+]). \tag{3.188}$$

This operator, or really the whole family Q_a, is called the spectral projection of A.

Proof. We have already shown the existence of $Q_a \in \mathcal{B}(H)$ with the property (3.187) and since we defined it directly from $Q_a(u)$ it is unique. Self-adjointness follows from the reality of $Q_a(u) \geq 0$ since $\overline{\langle Q_a u, v\rangle} = \langle u, Q_a v\rangle$ then follows from (3.186).

From (3.184) it follows that

$$\begin{gathered}\langle Q_a u, v\rangle = \lim_{n\to\infty} \langle f_n(A)u, v\rangle \Longrightarrow \\ \langle Q_a u, f(A)v\rangle = \lim_{n\to\infty} \langle f_n(A)u, f(A)v\rangle = \langle Q_a f(A)u, v\rangle\end{gathered} \tag{3.189}$$

since $f(A)$ commutes with $f_n(A)$ for any continuous f. This proves the statement in (3.188). Since $f_n f_m \leq f_n$ is admissible in the definition of Q_a in (3.178)

$$\langle Q_a u, v\rangle = \lim_{n\to\infty} \langle (f_n f_m)(A)u, v\rangle = \lim_{n\to\infty} \langle f_n(A)u, f_m(A)v\rangle = \langle Q_a(A)u, f_m(A)v\rangle \tag{3.190}$$

and now letting $m \to \infty$ shows that $Q_a^2 = Q_a$. A similar argument shows the first identity in (3.188). □

Returning to the original thought that (3.177) represents a Riemann-Stieltjes integral for each u we see that collectively what we have is a map

$$[a_-, a_+] \ni a \longmapsto Q_a \in \mathcal{B}(H) \tag{3.191}$$

taking values in the self-adjoint projections and *increasing* in the sense of (3.188). A little more application allows one to recover the functional calculus as an integral which can be written

$$f(A) = \int_{[a_-,a_+]} f(t) dQ_t \tag{3.192}$$

which does indeed reduce to a Riemann-Stieltjes integral for each u :

$$\langle f(A)u, u\rangle = \int_{[a_-,a_+]} f(t) d\langle Q_t u, u\rangle. \tag{3.193}$$

This, meaning (3.192), is the spectral resolution of the self-adjoint operator A, replacing (and reducing to) the decomposition as a sum in the compact case

$$f(A) = \sum_n f(\lambda_j) P_j \tag{3.194}$$

where the P_j are the orthogonal projections onto the eigenspaces for λ_j.

21. Polar Decomposition

One nice application of the functional calculus for self-adjoint operators is to get the polar decomposition of a general bounded operator.

Lemma 3.18. *If $A \in \mathcal{B}(H)$ then $E = (A^*A)^{\frac{1}{2}}$, defined by the functional calculus, is a non-negative self-adjoint operator.*

Proof. That E exists as a self-adjoint operator satisfying $E^2 = A^*A$ follows directly from Theorem 3.5 and positivity follows as in Remark 3.1. □

Proposition 3.19. *Any bounded operator A can be written as a product*

$$A = U(A^*A)^{\frac{1}{2}},\ U \in \mathcal{B}(H),\ U^*U = \mathrm{Id} - \Pi_{\mathrm{Nul}(A)},\ UU^* = \Pi_{\overline{\mathrm{Ran}(A)}}. \tag{3.195}$$

Proof. Set $E = (A^*A)^{\frac{1}{2}}$. We want to define U and we can see from the first condition, $A = UE$, that

$$U(w) = Av, \quad \text{if } w = Ev. \tag{3.196}$$

This makes sense since $Ev = 0$ implies $\langle Ev, Ev\rangle = 0$ and hence $\langle A^*Av, v\rangle = 0$ so $\|Av\| = 0$ and $Av = 0$. So let us define

$$U(w) = \begin{cases} Av & \text{if } w \in \mathrm{Ran}(E),\ w = Ev \\ 0 & \text{if } w \in (\mathrm{Ran}(E))^\perp. \end{cases} \tag{3.197}$$

So U is defined on a dense subspace of H, $\operatorname{Ran}(E)\oplus(\operatorname{Ran}(E))^{\perp}$ which may not be closed if $\operatorname{Ran}(E)$ is not closed. It follows that

$$\begin{gathered}U(w_1+w_2)=U(w_1)=Av_1\Longrightarrow\\ \|U(w_1+w_2)\|^2=|\langle Av_1,Av_1\rangle|^2=\langle E^2v_1,v_1\rangle=\|Ev_1\|^2=\|w_1\|^2\le\|w_1+w_2\|^2\\ \text{if } w_1=Ev_,\ w_2\in(\operatorname{Ran}E)^{\perp}.\end{gathered}\tag{3.198}$$

Thus U is bounded on the dense subspace on which it is defined, so has a unique continuous extension to a bounded operator $U\in\mathcal{B}(H)$. From the definition of U the first, factorization, condition in (3.195) holds.

From the definition U vanishes on $\operatorname{Ran}(E)^{\perp}$. We can now check that the continuous extension is a bijection

$$U:\overline{\operatorname{Ran}(E)}\longrightarrow\overline{\operatorname{Ran}(A)}.\tag{3.199}$$

Indeed, if $w\in\overline{\operatorname{Ran}(E)}$ then $\|w\|=\|Uw\|$ from (3.198) so (3.199) is injective. The same identity shows that the range of U in (3.199) is closed since if Uw_n converges, $\|w_n-w_m\|=\|U(w_n-w_m)\|$ shows that the sequence w_n is Cauchy and hence converges; the range is therefore $\overline{\operatorname{Ran}(A)}$. This same identity, $\|Uw\|=\|w\|$, for $w\in\operatorname{Ran}(E)$, implies that

$$\langle Uw,Uw'\rangle=\langle w,w'\rangle,\ w,w'\in Ran(E).\tag{3.200}$$

This follows from the polarization identity

$$\begin{aligned}4\langle Uw,Uw'\rangle&=\|U(w+w')\|^2-\|U(w-w')\|^2+i\|U(w+iw')\|^2-i\|U(w-iw')\|^2\\&=\|w+w'\|^2-\|w-w'\|^2+i\|w+iw'\|^2-i\|w-iw'\|^2=4\langle w,w'\rangle\end{aligned}\tag{3.201}$$

The adjoint U^* of U has range contained in the orthocomplement of the null space of U, so in $\overline{\operatorname{Ran}(E)}$, and null space precisely $\operatorname{Ran}(A)^{\perp}$ so defines a linear map from $\operatorname{Ran}(A)$ to $\overline{\operatorname{Ran}(E)}$. As such it follows from (3.201) that

$$U^*U=\operatorname{Id}\ \text{ on }\ \operatorname{Ran}(E)\Longrightarrow U^*=U^{-1}\ \text{ on }\ \operatorname{Ran}(A)\tag{3.202}$$

since U is a bijection it follows that U^* is the two-sided inverse of U as a map in (3.199). The remainder of (3.195) follows from this, so completing the proof of the Proposition. □

A bounded linear operator with the properties of U above, that there are two decompositions of $H=H_1\oplus H_2=H_3\oplus H_4$ into orthogonal closed subspaces, such that $U=0$ on H_2 and $U:H_1\longrightarrow H_3$ is a bijection with $\|Uw\|=\|w\|$ for all $w\in H_1$ is called a *partial isometry.* So the polar decomposition writes a general bounded operator as product $A=UE$ where U is a partial isometry from $\overline{\operatorname{Ran}(E)}$ onto $\overline{\operatorname{Ran}(A)}$ and $E=(A^*A)^{\frac12}$. If A is injective then U is actually unitary.

EXERCISE 1. Show that in the same sense, $A=FV$ where $F=(AA^*)^{\frac12}$ and V is a partial isometry from $\overline{\operatorname{Ran}(A^*)}$ to $\overline{\operatorname{Ran}F}$.

22. Compact perturbations of the identity

I have generally not had a chance to discuss most of the material in this section, or the next, in the lectures.

Compact operators are, as we know, 'small' in the sense that they are norm limits of finite rank operators. Accepting this, then you will want to say that an operator such as

$$\operatorname{Id} - K,\ K \in \mathcal{K}(\mathcal{H}) \tag{3.203}$$

is 'big'. We are quite interested in this operator because of spectral theory. To say that $\lambda \in \mathbb{C}$ is an eigenvalue of K is to say that there is a non-trivial solution of

$$Ku - \lambda u = 0 \tag{3.204}$$

where non-trivial means other than than the solution $u = 0$ which always exists. If λ is an eigenvalue of K then certainly $\lambda \in \operatorname{Spec}(K)$, since $\lambda - K$ cannot be invertible. For general operators the converse is not correct, but for compact operators it is.

Lemma 3.19. *If $K \in \mathcal{B}(H)$ is a compact operator then $\lambda \in \mathbb{C} \setminus \{0\}$ is an eigenvalue of K if and only if $\lambda \in \operatorname{Spec}(K)$.*

Proof. Since we can divide by λ we may replace K by $\lambda^{-1}K$ and consider the special case $\lambda = 1$. Now, if K is actually finite rank the result is straightforward. By Lemma 3.7 we can choose a basis so that (3.85) holds. Let the span of the e_i be W – since it is finite dimensional it is closed. Then $\operatorname{Id} - K$ acts rather simply – decomposing $H = W \oplus W^\perp$, $u = w + w'$

$$(\operatorname{Id} - K)(w + w') = w + (\operatorname{Id}_W - K')w',\ K' : W \longrightarrow W \tag{3.205}$$

being a matrix with respect to the basis. It follows that 1 is an eigenvalue of K if and only if 1 is an eigenvalue of K' as an operator on the finite-dimensional space W. A matrix, such as $\operatorname{Id}_W - K'$, is invertible if and only if it is injective, or equivalently surjective. So, the same is true for $\operatorname{Id} - K$.

In the general case we use the approximability of K by finite rank operators. Thus, we can choose a finite rank operator F such that $\|K - F\| < 1/2$. Thus, $(\operatorname{Id} - K + F)^{-1} = \operatorname{Id} - B$ is invertible. Then we can write

$$\operatorname{Id} - K = \operatorname{Id} - (K - F) - F = (\operatorname{Id} - (K - F))(\operatorname{Id} - L),\ L = (\operatorname{Id} - B)F. \tag{3.206}$$

Thus, $\operatorname{Id} - K$ is invertible if and only if $\operatorname{Id} - L$ is invertible. Thus, if $\operatorname{Id} - K$ is *not* invertible then $\operatorname{Id} - L$ is not invertible and hence has null space and from (3.206) it follows that $\operatorname{Id} - K$ has non-trivial null space, i.e. K has 1 as an eigenvalue. □

A little more generally:-

Proposition 3.20. *If $K \in \mathcal{K}(\mathcal{H})$ is a compact operator on a separable Hilbert space then*

$$\begin{gathered}\operatorname{null}(\operatorname{Id} - K) = \{u \in \mathcal{H}; (\operatorname{Id}_K)u = 0\} \textit{ is finite dimensional}\\ \operatorname{Ran}(\operatorname{Id} - K) = \{v \in \mathcal{H}; \exists u \in \mathcal{H},\ v = (\operatorname{Id} - K)u\} \textit{ is closed and}\\ \operatorname{Ran}(\operatorname{Id} - K)^\perp = \{w \in \mathcal{H}; (w, Ku) = 0\ \forall\ u \in \mathcal{H}\} \textit{ is finite dimensional}\end{gathered} \tag{3.207}$$

and moreover

$$\dim\left(\operatorname{null}(\operatorname{Id} - K)\right) = \dim\left(\operatorname{Ran}(\operatorname{Id} - K)^\perp\right). \tag{3.208}$$

Proof of Proposition 3.20. First let's check this in the case of a finite rank operator $K = T$. Then

$$\operatorname{Nul}(\operatorname{Id} - T) = \{u \in \mathcal{H}; u = Tu\} \subset \operatorname{Ran}(T). \tag{3.209}$$

A subspace of a finite dimensional space is certainly finite dimensional, so this proves the first condition in the finite rank case.

Similarly, still assuming that T is finite rank consider the range

$$\operatorname{Ran}(\operatorname{Id} - T) = \{v \in \mathcal{H}; v = (\operatorname{Id} - T)u \text{ for some } u \in \mathcal{H}\}. \tag{3.210}$$

Consider the subspace $\{u \in \mathcal{H}; Tu = 0\}$. We know that this this is closed, since T is certainly continuous. On the other hand from (3.210),

$$\operatorname{Ran}(\operatorname{Id} - T) \supset \operatorname{Nul}(T). \tag{3.211}$$

Now, $\operatorname{Nul}(T)$ is closed and has finite *codimension* – it's orthocomplement is spanned by a finite set which maps to span the image. As shown in Lemma 3.4 it follows from this that $\operatorname{Ran}(\operatorname{Id} - T)$ itself is closed with finite dimensional complement.

This takes care of the case that $K = T$ has finite rank! What about the general case where K is compact? If K is compact then there exists $B \in \mathcal{B}(\mathcal{H})$ and T of finite rank such that

$$K = B + T, \ \|B\| < \frac{1}{2}. \tag{3.212}$$

Now, consider the null space of $\operatorname{Id} - K$ and use (3.212) to write

$$\operatorname{Id} - K = (\operatorname{Id} - B) - T = (\operatorname{Id} - B)(\operatorname{Id} - T'), \ T' = (\operatorname{Id} - B)^{-1}T. \tag{3.213}$$

Here we have used the convergence of the Neumann series, so $(\operatorname{Id} - B)^{-1}$ does exist. Now, T' is of finite rank, by the ideal property, so

$$\operatorname{Nul}(\operatorname{Id} - K) = \operatorname{Nul}(\operatorname{Id} - T') \text{ is finite dimensional.} \tag{3.214}$$

Here of course we use the fact that $(\operatorname{Id} - K)u = 0$ is equivalent to $(\operatorname{Id} - T')u = 0$ since $\operatorname{Id} - B$ is invertible. So, this is the first condition in (3.207).

Similarly, to examine the second we do the same thing but the other way around and write

$$\operatorname{Id} - K = (\operatorname{Id} - B) - T = (\operatorname{Id} - T'')(\operatorname{Id} - B), \ T'' = T(\operatorname{Id} - B)^{-1}. \tag{3.215}$$

Now, T'' is again of finite rank and

$$\operatorname{Ran}(\operatorname{Id} - K) = \operatorname{Ran}(\operatorname{Id} - T'') \text{ is closed and of finite codimension.} \tag{3.216}$$

What about (3.208)? This time let's first check first that it is enough to consider the finite rank case. For a compact operator we have written

$$(\operatorname{Id} - K) = G(\operatorname{Id} - T) \tag{3.217}$$

where $G = \operatorname{Id} - B$ with $\|B\| < \frac{1}{2}$ is invertible and T is of finite rank. So what we want to see is that

$$\dim \operatorname{Nul}(\operatorname{Id} - K) = \dim \operatorname{Nul}(\operatorname{Id} - T) = \dim \operatorname{Nul}(\operatorname{Id} - K^*). \tag{3.218}$$

However, $\operatorname{Id} - K^* = (\operatorname{Id} - T^*)G^*$ and G^* is also invertible, so

$$\dim \operatorname{Nul}(\operatorname{Id} - K^*) = \dim \operatorname{Nul}(\operatorname{Id} - T^*) \tag{3.219}$$

and hence it is enough to check that $\dim \operatorname{Nul}(\operatorname{Id} - T) = \dim \operatorname{Nul}(\operatorname{Id} - T^*)$ – which is to say the same thing for finite rank operators.

Now, for a finite rank operator, written out as (3.85), we can look at the vector space W spanned by all the f_i's and all the e_i's together – note that there is nothing to stop there being dependence relations among the combination although separately they are independent. Now, $T : W \longrightarrow W$ as is immediately clear and

$$T^*v = \sum_{i=1}^{N}(v, f_i)e_i \tag{3.220}$$

so $T : W \longrightarrow W$ too. In fact $Tw' = 0$ and $T^*w' = 0$ if $w' \in W^\perp$ since then $(w', e_i) = 0$ and $(w', f_i) = 0$ for all i. It follows that if we write $R : W \longleftrightarrow W$ for the linear map on this finite dimensional space which is equal to $\operatorname{Id} - T$ acting on it, then R^* is given by $\operatorname{Id} - T^*$ acting on W and we use the Hilbert space structure on W induced as a subspace of $\mathcal{H}$. So, what we have just shown is that

$$(\operatorname{Id} - T)u = 0 \Longleftrightarrow u \in W \text{ and } Ru = 0,\ (\operatorname{Id} - T^*)u = 0 \Longleftrightarrow u \in W \text{ and } R^*u = 0. \tag{3.221}$$

Thus we really are reduced to the finite-dimensional theorem

$$\dim \operatorname{Nul}(R) = \dim \operatorname{Nul}(R^*) \text{ on } W. \tag{3.222}$$

You no doubt know this result. It follows by observing that in this case, everything is now in W, $\operatorname{Ran}(W) = \operatorname{Nul}(R^*)^\perp$ and in finite dimensions

$$\dim \operatorname{Nul}(R) + \dim \operatorname{Ran}(R) = \dim W = \dim \operatorname{Ran}(W) + \dim \operatorname{Nul}(R^*). \tag{3.223}$$

□

23. Hilbert-Schmidt, Trace and Schatten ideals

As well as the finite rank and compact operators there are other important ideals. Since these results are not exploited in the subsequence sections, the many proofs are relegated to exercises.

First consider the Hilbert-Schmidt operators. The definition is based on

Lemma 3.20. *For a separable Hilbert space, H, if $A \in \mathcal{B}(H)$ then once the sum for any one orthonormal basis $\{e_i\}$*

$$\|A\|^2_{\mathrm{HS}} = \sum_i \|Ae_i\|^2 \tag{3.224}$$

is finite it is finite for any other orthonormal basis and is independent of the choice of basis.

It is straightforward to show that the operators of finite rank satisfy (3.224); this is basically Bessel's inequality.

Proof. This is Problem XX. Starting from (3.224) for some orthonormal basis e_i, consider any other orthonormal basis f_j. Using the completeness, expand using Bessel's identity

$$\|Ae_i\|^2 = \sum_j |\langle Ae_i, f_j\rangle|^2 = \sum_j |\langle e_i, A^*f_j\rangle|^2. \tag{3.225}$$

This converges absolutely, so the convergence of (3.224) implies the convergence of the double sum, which can then be rearranged to give

$$\sum_i \|Ae_i\|^2 = \sum_i \sum_j |\langle e_i, A^*f_j\rangle|^2 = \sum_j \sum_i |\langle e_i, A^*f_j\rangle|^2 = \sum_j \|A^*f_j\|^2 \tag{3.226}$$

where Bessel's identity is used again. Thus the sum for A^* with respect to the new basis is finite. Applying this argument again shows that the sum is independent of the basis, and the same for the adjoint. □

Proposition 3.21. *The operators for which* (3.224) *is finite form a 2-sided $*$ -ideal* $\mathrm{HS}(H) \subset \mathcal{B}(H)$, *contained in the ideal of compact operators, it is a Hilbert space and the norm satisfies*

$$\|A\|_{\mathcal{B}} \le \|A\|_{\mathrm{HS}} = \left(\sum_i \|Ae_i\|^2\right)^{\frac{1}{2}}, \tag{3.227}$$

$$\|AD\|_{\mathrm{HS}} \le \|A\|_{\mathrm{HS}}\|D\|_{\mathcal{B}},\ A \in \mathrm{HS}(H),\ D \in \mathcal{B}(H).$$

The inner product is

$$\langle A, B\rangle_{\mathrm{HS}} = \sum_i \langle Ae_i, Be_i\rangle_H,\ A,\ B \in \mathrm{HS}(H).$$

For a compact operator the polar decomposition can be given a more explicit form and we can use this to give another characterization of the Hilbert-Schmidt operators.

Proposition 3.22. *If* $A \in \mathcal{K}(H)$ *then there exist orthonormal bases* e_i *of* $\mathrm{Nul}(A)^\perp$ *and* f_j *of* $\mathrm{Nul}(A^*)^\perp$ *such that*

$$Au = \sum_i s_i \langle u, e_i\rangle f_i$$

where the s_i are the non-zero eigenvalues of $(A^*A)^{\frac{1}{2}}$ *repeated with multiplicity.*

The s_i are called the *characteristic values* of A.

Proof. First take a basis e_i of eigenvectors of A^*A restricted to $\mathrm{Nul}(A)^\perp = \mathrm{Nul}(A^*A)^\perp$ with eigenvalues $s_i^2 > 0$, so the s_i are the non-zero eigenvalues of $|A| = (A^*A)^{\frac{1}{2}}$. Then $A = U|A|$ with U a unitary operator from $\overline{\mathrm{Ran}(|A|)} = \mathrm{Nul}(A)^\perp$ to $\overline{\mathrm{Ran}(A)}$ so (3.22) follows by taking $f_i = Ue_i$. □

Extending the e_i to an orthonormal basis of H it follows that

$$\|A\|_{\mathrm{HS}} = \left(\sum_i s_i^2\right)^{\frac{1}{2}} = \|s_*\|_{l^2}.$$

So to say that A is Hilbert-Schmidt is to say that the sequence of its characteristic values is in l^2 (with the caveat that the sequence might be finite).

One reason that the Hilbert-Schmidt operators are of interest is their relation to the ideal of operators 'of trace class', $\mathcal{T}(H)$.

Definition 3.7. The space $\mathcal{T}(H) \subset \mathcal{B}(H)$ for a separable Hilbert space consists of those operators A for which

$$\|A\|_{\mathrm{Tr}} = \sup \sum_i |\langle Ae_i, f_i\rangle| < \infty \tag{3.228}$$

where the supremum is over pairs of orthonormal sequences $\{e_i\}$ and $\{f_i\}$.

Proposition 3.23. *The trace class operators form an ideal, $\mathcal{T}(H) \subset \mathrm{HS}(H)$, which is a Banach space with respect to the norm (3.228) which satisfies*

$$\|A\|_{\mathcal{B}} \le \|A\|_{\mathrm{Tr}},\ \|A\|_{\mathrm{HS}} \le \|A\|_{\mathcal{B}}^{\frac{1}{2}}\|A\|_{\mathrm{Tr}}^{\frac{1}{2}}; \tag{3.229}$$

the following two conditions are equivalent to $A \in \mathcal{T}(H)$:

(1) *The operator defined by the functional calculus,*

$$|A|^{\frac{1}{2}} = (A^*A)^{\frac{1}{4}} \in \mathrm{HS}(H). \tag{3.230}$$

(2) *There are operators $B_i,\ B_i' \in \mathrm{HS}(H)$ such that*

$$A = \sum_{i=1}^{N} B_i' B_i. \tag{3.231}$$

Proof. Note first that $\mathcal{T}(H)$ is a linear space and that $\|\cdot\|_{\mathrm{Tr}}$ is a norm on it. Now suppose $A \in \mathcal{T}(H)$ and consider its polar decomposition $A = U(A^*A)^{\frac{1}{2}}$. Here U is a partial isometry mapping $\overline{\mathrm{Ran}(A^*A)^{\frac{1}{2}}}$ to $\overline{\mathrm{Ran}(A)}$ and vanishing on $\overline{\mathrm{Ran}(A^*A)^{\frac{1}{2}}}^{\perp}$. Consider an orthonormal basis $\{e_i\}$ of $\overline{\mathrm{Ran}(A^*A)^{\frac{1}{2}}}$. This is an orthonormal sequence in H as is $f_i = Ue_i$. Inserting these into (3.228) shows that

$$\sum_i |\langle U(A^*A)^{\frac{1}{2}} e_i, f_i\rangle| = \sum_i |\langle (A^*A)^{\frac{1}{4}} e_i, (A^*A)^{\frac{1}{4}} e_i\rangle| < \infty \tag{3.232}$$

where we use the fact that $U^* f_i = U^*Ue_i = e_i$. Since the closure of the range of $(A^*A)^{\frac{1}{4}}$ is the same as the closure of the range of $(A^*A)^{\frac{1}{2}}$ it follows from (3.232) that (3.230) holds (since adding an orthonormal basis of $\mathrm{Ran}((A^*A)^{\frac{1}{4}})^{\perp}$ does not increase the sum).

Next assume that (3.230) holds for $A \in \mathcal{B}(H)$. Then the polar decomposition can be written $A = (U(A^*A)^{\frac{1}{4}})(A^*A)^{\frac{1}{4}}$ showing that A is the product of two Hilbert-Schmidt operators, so in particular of the form (3.231).

Now assume that A is of the form (3.231), so is a sum of products of Hilbert-Schmidt operators. The linearity of $\mathcal{T}(H)$ means it suffices to assume that $A = BB'$ where $B,\ B' \in \mathrm{HS}(H)$. Then,

$$|\langle Ae, f_i\rangle| = |\langle B'e_i, B^* f_i\rangle| \le \|B'e_i\|_H \|B^* f_i\|_H. \tag{3.233}$$

Taking a finite sum and applying Cauchy-Schwartz inequality

$$\sum_{i=1}^{N} |\langle Ae, f_i\rangle| \le (\sum_{i=1}^{N} \|B'e_i\|^2)^{\frac{1}{2}} (\sum_{i=1}^{N} \|B^* f_i\|^2)^{\frac{1}{2}}. \tag{3.234}$$

If the sequences are orthonormal the right side is bounded by the product of the Hilbert-Schmidt norms so

$$\|BB'\|_{\mathrm{Tr}} \le \|B\|_{\mathrm{HS}} \|B'\|_{\mathrm{HS}} \tag{3.235}$$

and $A = BB' \in \mathcal{T}(H)$.

The first inequality in (3.229) follows the choice of single unit vectors u and v as orthonormal sequences, so

$$|\langle Au, v\rangle| \le \|A\|_{\mathrm{Tr}} \Longrightarrow \|A\| \le \|A\|_{\mathrm{Tr}}. \tag{3.236}$$

The completeness of $\mathcal{T}(H)$ with respect to the trace norm follows standard arguments which can be summarized as follows

(1) If A_n is Cauchy in $\mathcal{T}(H)$ then by the equality just established, it is Cauchy in $\mathcal{B}(H)$ and so converges in norm to $A \in \mathcal{B}(H)$.

(2) A Cauchy sequence is bounded, so there is a constant $C = \sup_n \|A_m\|_{\mathrm{Tr}}$ such that for any N, any orthonormal sequences e_i, f_i,

$$\sum_{i=1}^{N} |\langle A_n e_i, f_i\rangle| \le C. \tag{3.237}$$

Passing to the limit $A_n \to A$ in the finite sum gives the same bound with A_n replaced by A and then allowing $N \to \infty$ shows that $A \in \mathcal{T}(H)$. Similarly the Cauchy condition means that for $\epsilon > 0$ there exists M such that for all N, and any orthonormal sequences e_i, f_i

$$m.n > M \Longrightarrow \sum_{i=1}^{N} |\langle (A_n - A_m)e_i, f_i\rangle| \le \epsilon. \tag{3.238}$$

Passing first to the limit $m \to \infty$ in the finite sum and then $N \to \infty$ shows that

$$n > M \Longrightarrow \|A_n - A\|_{\mathrm{Tr}} \le \epsilon$$

and so $A_n \to A$ in the trace norm.

□

Proposition 3.24. *The trace functional*

$$\mathcal{T}(H) \ni A \longmapsto \mathrm{Tr}(A) = \sum_i \langle Ae_i, e_i\rangle \tag{3.239}$$

is a continuous linear functional (with respect to the trace norm) which is independent of the choice of orthonormal basis $\{e_i\}$ and which satsifies

$$\mathrm{Tr}(AB - BA) = 0 \textit{ if } A \in \mathcal{T}(H),\ B \in \mathcal{B}(H) \textit{ or } A,\ B \in \mathrm{HS}(H). \tag{3.240}$$

Proof. The complex number $\mathrm{Tr}(AB - BA)$ depends linearly on A and, separately, on B. The ideals are $*$-closed so decomposing $A = (A + A^*)/2 + i(A - A^*)/2i$ and similarly for B shows that it suffices to assume that A and B are self-adjoint. If $A \in \mathcal{T}(H)$ we can choose use an orthonormal basis of eigenvectors for it to evaluate the trace. Then if $Ae_i = \lambda_i e_i$

$$\begin{aligned}\mathrm{Tr}(AB - BA) &= \sum_i \left(\langle Be_i, Ae_i\rangle - \langle Ae_i, Be_i\rangle\right) \\ &= \sum_i \left(\lambda_i\langle Be_i, e_i\rangle - \lambda_i\langle e_i, Be_i\rangle\right) = 0.\end{aligned} \tag{3.241}$$

The case that A, $B \in \mathrm{HS}(H)$ is similar. □

This is the fundamental property of the trace functional, that it vanishes on commutators where one of the elements is of trace class and the other is bounded. Two other important properties are that

Lemma 3.21. (1) *If A, $B \in \mathrm{HS}(H)$ then*

$$\langle A, B\rangle_{\mathrm{HS}} = \mathrm{Tr}(A^*B) \tag{3.242}$$

(2) *If $T = T^* \in \mathcal{K}(H)$ then $T \in \mathcal{T}(H)$ if and only if the sequence of non-zero eigenvalues λ_j of T (repeated with multiplicity) is in l^1 and*

$$\operatorname{Tr}(T) = \sum_j \lambda_j,\ \|T\|_{\operatorname{Tr}} = \sum_j |\lambda_j|. \tag{3.243}$$

In fact the second result extends to *Lidskii's theorem:* If $T \in \operatorname{Tr}(H)$ then the spectrum outside 0 is discrete, so countable, and each point is an eigenvalue λ_i of finite algebraic multiplicity k_i and then $\operatorname{Tr}(T) = \sum\limits_i k_i\lambda_i$ converges in l^1. The algebraic multiplicity is the limit as $k \to \infty$ of the dimension of the null space of $(T - \lambda_i)^k$. The standard proof of this is not elementary.

Next we turn to the more general Schatten classes.

Definition 3.8. An operator $A \in \mathcal{K}(H)$ is 'of Schatten class,' $A \in \operatorname{Sc}_p(H)$, $p \in [1, \infty)$ if and only if $|A|^p \in \mathcal{T}(H)$, i.e.

$$\|T\|_{\operatorname{Sc}_p} = \left(\sum_i s_i^p\right)^{\frac{1}{p}} < \infty \tag{3.244}$$

where s_i are the non-zero characteristic values of A repeated with multiplicity.

So $\mathcal{T}(H) = \operatorname{Sc}_1(H)$, $\operatorname{HS}(H) = \operatorname{Sc}_2(H)$.

Of course the notation is suggestive, but we need to be a bit careful in proving the results which are implied by the notation!

Proposition 3.25. *Each of the Schatten classes is a two-sided $*$-ideal in $\mathcal{B}(H)$ which is a Banach space with respect to the norm (3.244); the norm is also given by*

$$\|T\|^p_{\operatorname{Sc}_p} = \sup \sum_i |\langle Te_i, f_i\rangle|^p \tag{3.245}$$

with the supremum over orthonormal sequences, with finiteness implying that $T \in \operatorname{Sc}_p(H)$. If q is the conjugate index to $p \in (0, \infty)$ then

$$A \in \operatorname{Sc}_q(H),\ B \in \operatorname{Sc}_p(H) \Longrightarrow AB \in \mathcal{T}(H),\ \|AB\|_{\operatorname{Tr}} \le \|A\|_{\operatorname{Sc}_q}\|B\|_{\operatorname{Sc}_p} \tag{3.246}$$

and conversely, if $A \in \mathcal{B}(H)$ then $A \in \operatorname{Sc}_p(H)$ if and only if $AB \in \mathcal{T}(H)$ for all $B \in \operatorname{Sc}_q(H)$ and

$$\|A\|_{\operatorname{Sc}_p} = \sup_{\|B\|_{\operatorname{Sc}_q}=1} \|AB\|_{\operatorname{Tr}}.$$

Proof. The alternate realization of the Schatten norm in (3.245) is particularly useful since whilst it is clear from the definition that $cT \in \operatorname{Sc}_p(H)$ if $T \in \operatorname{Sc}_p(H)$ and $c \in \mathbb{C}$, it is not otherwise immediately clear that the space is linear (or that the triangle inequality holds).

From the definition (3.244), that if T is self-adjoint then $T \in \operatorname{Sc}_p(H)$ if and only if

$$\sup \sum_i |\langle Tf_i, f_i\rangle|^p = \|T\|^p_{\operatorname{Sc}_p} < \infty \tag{3.247}$$

with the supremum over orthonormal sequences. To see this let e_j be an orthonomal basis of eigenvectors for T. Then expanding in the Fourier-Bessel series

$$\begin{aligned}\langle Tf_i, f_i\rangle &= \sum_j \lambda_j|\langle f_i, e_j\rangle|^2 \le \sum_j |\lambda_j||\langle f_i, e_j\rangle|^{2/p}|\langle f_i, e_j\rangle|^{2/q} \\ &\le (\sum_j |\lambda_j|^p\langle f_i, e_j\rangle|^2)^{\frac{1}{p}}(\sum_j |\langle f_i, e_j\rangle|^2)^{\frac{1}{q}} = (\sum_j |\lambda_j|^p\langle f_i, e_j\rangle|^2)^{\frac{1}{p}}\end{aligned} \tag{3.248}$$

by Hölder's inequality, so

$$\sum_i |\langle Tf_i, f_i\rangle|^p \le \sum_j |\lambda_j|^p \sum_i \langle f_i, e_j\rangle|^2 = \sum_j |\lambda_j|^p = \|T\|^p_{\mathrm{Sc}_p}. \tag{3.249}$$

This proves (3.247) when $T = T^* \in \mathcal{K}(H)$.

Now consider (3.245). Let P_N be the orthogonal projection onto the span of the eigenspaces corresponding to the the N largest eigenvalues of $|T|$. Then we replace T by $T_N = TP_N$; certainly $TP_N \to T$ in norm. Since T_N has finite rank, both $\mathrm{Nul}(T_N)$ and $\mathrm{Nul}(T_N^*)$ are infinite dimensional so we can write the polar decomposition as

$$T_N = U_N A_N,\ A_N = P_N|T|P_N$$

and take U_N to be unitary (rather than a partial isometry) by extending it by an isometric isomorphism $\mathrm{Nul}(T_N)^\perp \longrightarrow \mathrm{Nul}(T_N^*)^\perp$. Then using Cauchy-Schwartz inequality and then (3.247) for A_N,

$$\begin{aligned}\sum_i |\langle T_N e_i, f_i\rangle|^p &= \sum_i |\langle A_N^{\frac{1}{2}} e_i, A_N^{\frac{1}{2}} U_N^* f_i\rangle|^p \\ &\le (\sum_i \|A_N^{\frac{1}{2}} e_i\|^{2p})^{\frac{1}{2}} (\sum_i \|A_N^{\frac{1}{2}} U_N^* f_i\|^{2p})^{\frac{1}{2}} \\ &\le (\sum_i |\langle A_N e_i, e_i\rangle|^p)^{\frac{1}{2}} |\sum_i |\langle A_N U_N^* f_i, U^* f_i\rangle|^p)^{\frac{1}{2}} \\ &\le \|A_N\|^p_{\mathrm{Sc}_p} = \|T_N\|^p_{\mathrm{Sc}_p} \le \|T\|^p_{\mathrm{Sc}_p}.\end{aligned} \tag{3.250}$$

As usual dropping to a finite sum on the left we can pass to the limit as $N \to \infty$ and obtain a uniform bound on any finite sum for T from which (3.245) follows.

At this point we know that if $A \in \mathrm{Sc}_p(H)$ and U_1, U_2 are unitary then

$$U_1AU_2 \in \mathrm{Sc}_p(H) \text{ and } \|U_1AU_2\|_{\mathrm{Sc}_p} = \|A\|_{\mathrm{Sc}_p}.$$

From (3.245) it follows directly that $\mathrm{Sc}_p(H)$ is linear, that the triangle inequality holds, so that $\|\cdot\|_{\mathrm{Sc}_p}$ is a norm, and $\mathrm{Sc}_p(H)$ is complete and that it is $*$-closed.

Now, if $A \in \mathrm{Sc}_q(H)$ and $B \in \mathrm{Sc}_p(H)$ for conjugate indices p, $q \in (1, \infty)$ choose a finite rank orthogonal projection P and consider ABP which is of finite rank, and hence of trace class. We can compute its trace with respect to any orthonormal basis. Choose an orthonormal basis e_i of the range of PAP and f_i so that the polar decomposition of PAP becomes

$$PAPf_i = s_ie_i \Longrightarrow PA^*Pe_i = s_if_i$$

where the s_i are the characteristic values of PA. With finite sums

$$
\begin{aligned}
(3.251)\quad |\operatorname{Tr}(PAPBP)| &= |\sum\langle PAP^2BPe_i,e_i\rangle| \\
&= |\sum\langle PBPe_i,PA^*Pe_i\rangle| \le \sum_{i=1}^{N} s_i|\langle PBPe_i,f_i\rangle| \\
&\le (\sum s_j^q)^{\frac{1}{q}}(\sum|\langle PBPe_i,f_i\rangle|^p)^{\frac{1}{p}} \le \|PAP\|_{\mathrm{Sc}_q}\|PBP\|_{\mathrm{Sc}_p}
\end{aligned}
$$

by Hölder's inequality. Now $|PBP| = P|B|P$ (and similarly for A) and from minimax arguments discussed earlier it follows that, $s_j(|PBP|) \le s_j(|B|)$ for all j. So we see that

$$
|\operatorname{Tr}(PAPBP)| \le \|A\|_{\mathrm{Sc}_q}\|B\|_{\mathrm{Sc}_p}. \tag{3.252}
$$

Fixing B this is true for any A, so A can be replaced by UA with U unitary, in such a way that $APB = |APB|$. We also know that $\|UA\|_{\mathrm{Sc}_q} = \|A\|_{\mathrm{Sc}_q}$ and since $P|APB|P$ is positive and

$$
\operatorname{Tr}(PAPBP) = \operatorname{Tr}(P|APB|P) = \|P|APB|P\|_{\mathrm{Tr}} \le \|A\|_{\mathrm{Sc}_q}\|B\|_{\mathrm{Sc}_p}. \tag{3.253}
$$

Taking an increasing sequence of projections P_N, it follows that $P_N|AP_NB|P_N \to |AB|$ in trace norm and that (3.246) holds.

The proof of optimality in this 'non-commutative Hölder inequality' is left as an exercise. That $\mathrm{Sc}_p(H)$ is an ideal then follows from the fact that $\mathcal{T}(H)$ is an ideal. □

24. Fredholm operators

Definition 3.9. A bounded operator $F \in \mathcal{B}(\mathcal{H})$ on a Hilbert space is said to be *Fredholm*, written $F \in \mathcal{F}(H)$, if it has the three properties in (3.207) – its null space is finite dimensional, its range is closed and the orthocomplement of its range is finite dimensional.

In view of Proposition 3.20, if $K \in \mathcal{K}(H)$ then $\operatorname{Id} - K \in \mathcal{F}(H)$. For general Fredholm operators the row-rank=colum-rank result (3.208) does not hold. Indeed the difference of these two integers, called the index of the operator,

$$
\operatorname{ind}(F) = \dim(\operatorname{null}(F)) - \dim\left(\operatorname{Ran}(F)^\perp\right) \tag{3.254}
$$

is a very important number with lots of interesting properties and uses.

Notice that the last two conditions in (3.207) are really independent since the orthocomplement of a subspace is the same as the orthocomplement of its closure. There is for instance a bounded operator on a separable Hilbert space with trivial null space and dense range which is not closed. How could this be? Think for instance of the operator on $L^2(0,1)$ which is multiplication by the function x. This is assuredly bounded and an element of the null space would have to satisfy $xu(x) = 0$ almost everywhere, and hence vanish almost everywhere. Moreover the density of the L^2 functions vanishing in $x < \epsilon$ for some (non-fixed) $\epsilon > 0$ shows that the range is dense. However this operator is not invertible and not Fredholm.

On the other hand we do know that a subspace with finite codimension is closed, so we can replace the last two conditions in Definition 3.9 by saying that the range of the operator has finite codimension. I have not done this directly since it is a little too easy to fall into the trap of thinking that it is enough to check that the *closure* of the range has finite codimension; it isn't!

Before looking at general Fredholm operators let's check that, in the case of operators of the form $\mathrm{Id} - K$, with K compact the third conclusion in (3.207) really follows from the first. This is a general fact which I mentioned, at least, earlier but let me pause to prove it.

Proposition 3.26. *If $B \in \mathcal{B}(\mathcal{H})$ is a bounded operator on a Hilbert space and B^* is its adjoint then*

$$\operatorname{Ran}(B)^\perp = (\overline{\operatorname{Ran}}(B))^\perp = \{v \in \mathcal{H}; (v, w) = 0 \ \forall \ w \in \operatorname{Ran}(B)\} = \operatorname{Nul}(B^*). \tag{3.255}$$

Proof. The definition of the orthocomplement of $\operatorname{Ran}(B)$ shows immediately that

$$\begin{aligned} v \in (\operatorname{Ran}(B))^\perp \Longleftrightarrow (v, w) = 0 \ \forall \ w \in \operatorname{Ran}(B) \Longleftrightarrow (v, Bu) = 0 \ \forall \ u \in \mathcal{H} \\ \Longleftrightarrow (B^* v, u) = 0 \ \forall \ u \in \mathcal{H} \Longleftrightarrow B^* v = 0 \Longleftrightarrow v \in \operatorname{Nul}(B^*). \end{aligned} \tag{3.256}$$

On the other hand we have already observed that $V^\perp = (\overline{V})^\perp$ for any subspace – since the right side is certainly contained in the left and $(u, v) = 0$ for all $v \in V$ implies that $(u, w) = 0$ for all $w \in \overline{V}$ by using the continuity of the inner product to pass to the limit of a sequence $v_n \to w$. □

There is a more 'analytic' way of characterizing Fredholm operators, rather than Definition 3.9.

Lemma 3.22. *An operator $F \in \mathcal{B}(H)$ is Fredholm, $F \in \mathcal{F}(H)$, if and only if it has a generalized inverse P satisfying*

$$\begin{aligned} PF &= \mathrm{Id} - \Pi_{\operatorname{Nul}(F)} \\ FP &= \mathrm{Id} - \Pi_{\operatorname{Ran}(F)^\perp} \end{aligned} \tag{3.257}$$

with the two projections of finite rank.

Proof. If (3.257) holds then F must be Fredholm, since its null space is finite dimensional, from the second identity the range of F must contain the range of $\mathrm{Id} - \Pi_{\operatorname{Nul}(F)^\perp}$ and hence it must be closed and of finite codimension.

Conversely, suppose that $F \in \mathcal{F}(H)$. We can divide H into two pieces in two ways as $H = \operatorname{Nul}(F) \oplus \operatorname{Nul}(F)^\perp$ and $H = \operatorname{Ran}(F)^\perp \oplus \operatorname{Ran}(F)$ where in each case the first summand is finite-dimensional. Then F defines four maps, from each of the two first summands to each of the two second ones but only one of these is non-zero and so F corresponds to a bounded linear map $\tilde{F} : \operatorname{Nul}(F)^\perp \longrightarrow \operatorname{Ran}(F)$. These are two Hilbert spaces with a bounded linear bijection between them, so the inverse map, $\tilde{P} : \operatorname{Ran}(F) \longrightarrow \operatorname{Nul}(F)^\perp$ is bounded by the Open Mapping Theorem and we can define

$$P = \tilde{P} \circ \Pi_{\operatorname{Nul}(F)^\perp}. \tag{3.258}$$

Then (3.257) follows directly. □

What we want to show is that the Fredholm operators form an open set in $\mathcal{B}(H)$ and that the index is locally constant. To do this we show that a weaker version of (3.257) also implies that F is Fredholm.

Lemma 3.23. *An operator $F \in \mathcal{F}(H)$ is Fredholm if and only if it has a* parametrix *$Q \in \mathcal{B}(H)$ in the sense that*

$$\begin{aligned} QF &= \mathrm{Id} - E_L \\ FQ &= \mathrm{Id} - E_R \end{aligned} \tag{3.259}$$

with E_L and E_R of finite rank. Moreover any two such parametrices differ by a finite rank operator.

The term 'parametrix' refers to an inverse modulo an ideal. Here we are looking at the ideal of finite rank operators. In fact this is equivalent to the existence of an inverse modulo compact operators. One direction is obvious – since finite rank operators are compact – the other is covered by one of the problems. Notice that the parametrix Q is itself Fredholm, since reversing the two equations shows that F is a parameterix for Q. Similarly it follows that if F is Fredholm then so is F^* and that the product of two Fredholm operators is Fredholm.

Proof. If F is Fredholm then $Q = P$ certainly is a parameterix in this sense. Conversely suppose that Q as in (3.259) exists. Then $\operatorname{Nul}(\operatorname{Id} - E_L)$ is finite dimensional – from (3.207) for instance. However, from the first identity $\operatorname{Nul}(F) \subset \operatorname{Nul}(QF) = \operatorname{Nul}(\operatorname{Id} - E_L)$ so $\operatorname{Nul}(F)$ is finite dimensional too. Similarly, the second identity shows that $\operatorname{Ran}(F) \supset \operatorname{Ran}(FQ) = \operatorname{Ran}(\operatorname{Id} - E_R)$ and the last space is closed and of finite codimension, hence so is the first. Thus the existence of such a parameterix Q implies that F is Fredholm.

Now if Q and Q' both satisfy (3.259) with finite rank error terms E'_R and E'_L for Q' then

$$(Q' - Q)F = E_L - E'_L \tag{3.260}$$

is of finite rank. Applying the generalized inverse, P, of F on the right shows that the difference

$$(Q' - Q) = (E_L - E'_L)P + (Q' - Q)\Pi_{\operatorname{Nul}(F)} \tag{3.261}$$

is indeed of finite rank. □

Observe that (3.260) can be reversed. If F is Fredholm, so has a parametrix Q then all the operators $Q + E$ where E is of finite rank are also parametrices. It is also the case that if F is Fredholm and K is compact then $F + K$ is Fredholm. Indeed, if you go through the proof above replacing 'finite rank' by 'compact' you can check this. Thus an operator is Fredholm if and only if it has invertible image in the Calkin algebra, $\mathcal{B}(H)/\mathcal{K}(H)$.

Now recall that finite-rank operators are of trace class, that the trace is well-defined and that the trace of a commutator where one factor is bounded and the other trace class vanishes. Using this we show

Lemma 3.24. *If Q and F satisfy* (3.259) *then*

$$\operatorname{ind}(F) = \operatorname{Tr}(E_L) - \operatorname{Tr}(E_R). \tag{3.262}$$

Proof. We certainly know that (3.262) holds in the special case that $Q = P$ is the generalized inverse of F, since then $E_L = \Pi_{\operatorname{Nul}(F)}$ and $E_R = \Pi_{\operatorname{Ran}(F)^\perp}$ and the traces are the dimensions of these spaces.

Now, if Q is a parameterix as in (3.259) consider the straight line of operators $Q_t = (1 - t)P + tQ$. Using the two sets of identities for the generalized inverse and paramaterix

$$\begin{aligned} Q_tF &= (1-t)PF + tQF = \operatorname{Id} - (1-t)\Pi_{\operatorname{Nul}(F)} - tE_L, \\ FQ_t &= (1-t)FP + tFQ = \operatorname{Id} - (1-t)\Pi_{\operatorname{Ran}(F)^\perp} - tE_R. \end{aligned} \tag{3.263}$$

Thus Q_t is a curve of parameterices and what we need to show is that

$$J(t) = \operatorname{Tr}((1-t)\Pi_{\operatorname{Nul}(F)} + tE_L) - \operatorname{Tr}((1-t)\Pi_{\operatorname{Ran}(F)^\perp} + tE_R) \tag{3.264}$$

is constant. This is a linear function of t as is Q_t. We can differentiate (3.263) with respect to t and see that

$$\frac{d}{dt}((1-t)\Pi_{\operatorname{Nul}(F)} + tE_L) - \frac{d}{dt}((1-t)\Pi_{\operatorname{Ran}(F)^\perp} + tE_R) = [Q-P, F] \\ \Longrightarrow J'(t) = 0 \tag{3.265}$$

since it is the trace of the commutator of a bounded and a finite rank operator (using the last part of Lemma 3.23). □

Proposition 3.27. *The Fredholm operators form an open set in $\mathcal{B}(H)$ on which the index is locally constant.*

Proof. We need to show that if F is Fredholm then there exists $\epsilon > 0$ such that $F + B$ is Fredholm if $\|B\| < \epsilon$. Set $B' = \Pi_{\operatorname{Ran}(F)} B \Pi_{\operatorname{Nul}(F)^\perp}$ then $\|B'\| \le \|B\|$ and $B - B'$ is finite rank. If $\tilde{F}$ is the operator constructed in the proof of Lemma 3.22 then $\tilde{F} + B'$ is invertible as an operator from $\operatorname{Nul}(F)^\perp$ to $\operatorname{Ran}(F)$ if $\epsilon > 0$ is small. The inverse, P'_B, extended as 0 to $\operatorname{Nul}(F)$ as P is defined in that proof, satisfies

$$\begin{aligned} P'_B(F+B) &= \operatorname{Id} - \Pi_{\operatorname{Nul}(F)} + P'_B(B - B'), \\ (F+B)P'_B &= \operatorname{Id} - \Pi_{\operatorname{Ran}(F)^\perp} + (B-B')P'_B \end{aligned} \tag{3.266}$$

and so is a parametrix for $F + B$. Thus the set of Fredholm operators is open.

The index of $F+B$ is given by the difference of the trace of the finite rank error terms in the second and first lines here. It depends continuously on B in $\|B\| < \epsilon$ so, being integer-valued, is constant. □

This shows in particular there is an open subset of $\mathcal{B}(H)$ which contains no invertible operators, in strong contrast to the finite dimensional case. In fact even the Fredholm operators do not form a dense subset of $\mathcal{B}(H)$. One such open subset consists of the *semi-Fredholm* operators – those with closed range and with *either* null space or complement of range finite-dimensional.

Why is the index important? For one thing it actually labels the components of the Fredholm operators – two Fredholm operators can be connected by a curve of Fredholms if and only if they have the same index. One of the main applications of the index is quite trivial to see – if the index of a Fredholm operator is positive then the operator must have non-trivial null space. This is a remarkably powerful method for showing that certain sorts ('elliptic' for one) of equations have non-trivial solutions.

25. Kuiper's theorem

For finite dimensional spaces, such as $\mathbb{C}^N$, the group of invertible operators – in this case matrices and denoted typically $\operatorname{GL}(N)$ – is a particularly important example of a Lie group. One reason it is important is that it carries a good deal of 'topological' structure. In particular – if you have done a little topology – its fundamental group is not trivial, in fact it is isomorphic to $\mathbb{Z}$. This corresponds to the fact that a continuous closed curve $c : \mathbb{S} \longrightarrow \operatorname{GL}(N)$ is *contractible* if and only if its winding number is zero – the effective number of times that the determinant goes around the origin in $\mathbb{C}$. There is a lot more topology than this and it is actually quite complicated.

Perhaps surprisingly, the corresponding group of the invertible bounded operators on a separable (complex) infinite-dimensional Hilbert space is contractible. This is Kuiper's theorem, and means that this group, $\mathrm{GL}(H)$, has no 'topology' at all, no holes in any dimension and for topological purposes it is like a big open ball. The proof is not really hard, but it is not exactly obvious either. It depends on an earlier idea, 'Eilenberg's swindle' - it is an unusual name for a theorem - which shows how the infinite-dimensionality is exploited. As you can guess, this is sort of amusing (if you have the right attitude . . .). The proof I give here is due to B. Mityagin, [**3**].

Let's denote by $\mathrm{GL}(H)$ this group. In view of the open mapping theorem we know that

$$\mathrm{GL}(H) = \{A \in \mathcal{B}(H); A \text{ is injective and surjective}\}. \tag{3.267}$$

Contractibility means precisely that there is a continuous map

$$\begin{gathered}\gamma : [0,1] \times \mathrm{GL}(H) \longrightarrow \mathrm{GL}(H) \text{ s.t.}\\ \gamma(0,A) = A,\ \gamma(1,A) = \mathrm{Id},\ \forall\ A \in \mathrm{GL}(H).\end{gathered} \tag{3.268}$$

Continuity here means for the metric space $[0,1] \times \mathrm{GL}(H)$ where the metric comes from the norms on $\mathbb{R}$ and $\mathcal{B}(H)$.

I will only show 'weak contractibility' of $\mathrm{GL}(H)$. This has nothing to do with weak convergence, rather just means that we only look for an homotopy over compact sets.

As a warm-up exercise, let us show that the group $\mathrm{GL}(H)$ is contractible to the unitary subgroup

$$\mathrm{U}(H) = \{U \in \mathrm{GL}(H); U^{-1} = U^*\}. \tag{3.269}$$

These are the isometric isomorphisms.

Proposition 3.28. *There is a continuous map*

$$\Gamma : [0,1] \times \mathrm{GL}(H) \longrightarrow \mathrm{GL}(H) \textit{ s.t. } \Gamma(0,A) = A,\ \Gamma(1,A) \in \mathrm{U}(H)\ \forall\ A \in \mathrm{GL}(H). \tag{3.270}$$

Proof. This is a consequence of the functional calculus, giving the 'polar decomposition' of invertible (and more generally bounded) operators. Namely, if $A \in \mathrm{GL}(H)$ then $AA^* \in \mathrm{GL}(H)$ is self-adjoint. Its spectrum is then contained in an interval $[a,b]$, where $0 < a \le b = \|A\|^2$. It follows from what we showed earlier that $R = (AA^*)^{\frac12}$ is a well-defined bounded self-adjoint operator and $R^2 = AA^*$. Moreover, R is invertible and the operator $U_A = R^{-1}A \in \mathrm{U}(H)$. Certainly it is bounded and $U_A^* = A^*R^{-1}$ so $U_A^*U_A = A^*R^{-2}A = \mathrm{Id}$ since $R^{-2} = (AA^*)^{-1} = (A^*)^{-1}A^{-1}$. Thus U_A^* is a right inverse of U_A, and (since U_A is a bijection) is the unique inverse so $U_A \in \mathrm{U}(H)$. So we have shown $A = RU_A$ then

$$\Gamma(s,A) = (s\,\mathrm{Id} + (1-s)R)U_A,\ s \in [0,1] \tag{3.271}$$

satisfies (3.270).

There is however the issue of continuity of this map. Continuity in s is clear enough but we also need to show that the map

$$\mathrm{GL}(H) \ni A \longmapsto (A^*A)^{\frac12} \in \mathrm{GL}(H), \tag{3.272}$$

defining R, is continuous.

Certainly the map $A \longmapsto A^*A$ is (norm) continuous. Suppose $A_n \to A$ in $\mathrm{GL}(H)$ then given ϵ there is a polynomial p such that

$$\|(B^*B)^{\frac{1}{2}} - p(B^*B)\| \le \epsilon/3 \text{ for } B = A,\ A_n\ \forall\ n. \tag{3.273}$$

On the other hand $(A_n^*A_n)^k \to (A^*A)^k$ for any k so

$$\|(A^*A)^{\frac{1}{2}} - (A_n^*A_n)^{\frac{1}{2}}\| \le \\ \|(A^*A)^{\frac{1}{2}} - p(A^*A)\| + \|p(A^*A) - p(A_n^*A_n)\| + \|(A_n^*A_n)^{\frac{1}{2}} - p(A_n^*A_n)\| \to 0. \tag{3.274}$$

□

So, for any compact subset $X \subset \mathrm{GL}(H)$ we seek a continuous map

$$\begin{gathered}\gamma : [0,1] \times X \longrightarrow \mathrm{GL}(H) \text{ s.t.} \\ \gamma(0,A) = A,\ \gamma(1,A) = \mathrm{Id},\ \forall\ A \in X,\end{gathered} \tag{3.275}$$

note that this is not contractibility *of* X, but *of* X in $\mathrm{GL}(H)$.

In fact, to carry out the construction without having to worry about too many things at once, just consider (path) connectedness of $\mathrm{GL}(H)$ meaning that there is a continuous map as in (3.275) where $X = \{A\}$ just consists of one point – so the map is just $\gamma : [0,1] \longrightarrow \mathrm{GL}(H)$ such that $\gamma(0) = A$, $\gamma(1) = \mathrm{Id}$.

The construction of γ is in three stages

1. Creating a gap
2. Rotating to a trivial factor
3. Eilenberg's swindle.

Lemma 3.25 (Creating a gap). *If $A \in \mathcal{B}(H)$ and $\epsilon > 0$ is given there is a decomposition $H = H_K \oplus H_L \oplus H_O$ into three closed mutually orthogonal infinite-dimensional subspaces such that if Q_I is the orthogonal projections onto H_I for $I = K,\ L,\ O$ then*

$$\|Q_L B Q_K\| < \epsilon. \tag{3.276}$$

Proof. Choose an orthonormal basis e_j, $j \in \mathbb{N}$, of H. The subspaces H_i will be determined by a corresponding decomposition

$$\mathbb{N} = K \cup L \cup O,\ K \cap L = K \cap O = L \cap O = \emptyset. \tag{3.277}$$

Thus H_I has orthonormal basis e_k, $k \in I$, $I = K,\ L,\ O$. To ensure (3.276) we choose the decomposition (3.277) so that all three sets are infinite and so that

$$|(e_l, Be_k)| < 2^{-l-1}\epsilon\ \forall\ l \in L,\ k \in K. \tag{3.278}$$

Once we have this, then for $u \in H$, $Q_K u \in H_K$ can be expanded to $\sum\limits_{k \in K} (Q_k u, e_k) e_k$ and expanding in H_L similarly,

$$\begin{gathered} Q_L B Q_K u = \sum_{l \in L} (BQ_K u, e_l) e_l = \sum_{k \in L} \sum_{k \in K} (Be_k, e_l)(Q_K u, e_k) e_l \\ \Longrightarrow \|Q_L B Q_K u\|^2 \le \sum_{k \in K} \left(|(Q_k u, e_k)|^2 \sum_{l \in L} |(Be_k, e_l)|^2 \right) \\ \le \frac{1}{2}\epsilon^2 \sum_{k \in K} |(Q_k u, e_k)|^2 \le \frac{1}{2}\epsilon^2 \|u\|^2 \end{gathered} \tag{3.279}$$

giving (3.276). The absolute convergence of the series following from (3.278).

Thus, it remains to find a decomposition (3.277) for which (3.278) holds. This follows from Bessel's inequality. First choose $1 \in K$ then $(Be_1, e_l) \to 0$ as $l \to \infty$ so $|(Be_1, e_{l_1})| < \epsilon/4$ for l_1 large enough and we will take $l_1 > 2k_1$. Then we use induction on N, choosing $K(N)$, $L(N)$ and $O(N)$ with

$$\begin{gathered} K(N) = \{k_1 = 1 < k_2 < \dots, k_N\}, \\ L(N) = \{l_1 < l_2 < \dots < l_N\},\ l_r > 2k_r,\ k_r > l_{r-1} \text{ for } 1 < r \le N \text{ and} \\ O(N) = \{1, \dots, l_N\} \setminus (K(N) \cup L(N)). \end{gathered}$$

Now, choose $k_{N+1} > l_N$ such that $|(e_l, Be_{k_{N+1}})| < 2^{-l-N}\epsilon$, for all $l \in L(N)$, and then $l_{N+1} > 2k_{N+1}$ such that $|(e_{l_{N+1}}, B_k)| < e^{-N-1-k}\epsilon$ for $k \in K(N+1) = K(N) \cup \{k_{N+1}\}$ and the inductive hypothesis follows with $L(N+1) = N(N) \cup \{l_{N+1}\}$. □

Given a fixed operator $A \in \mathrm{GL}(H)$ Lemma 3.25 can be applied with $\epsilon = \|A^{-1}\|^{-1}$. It then follows, from the convergence of the Neumann series, that the curve

$$A(s) = A - sQ_LAQ_K,\ s \in [0,1] \tag{3.280}$$

lies in $\mathrm{GL}(H)$ and has endpoint satisfying

$$Q_LBQ_K = 0,\ B = A(1),\ Q_LQ_K = 0 = Q_KQ_L,\ Q_K = Q_K^2,\ Q_L = Q_L^2 \tag{3.281}$$

where all three projections, Q_L, Q_K and $\mathrm{Id} - Q_K - Q_L$ have infinite rank.

These three projections given an identification of $H = H \oplus H \oplus H$ and so replace the bounded operators by 3×3 matrices with entries which are bounded operators on H. The condition (3.281) means that

$$B = \begin{pmatrix} B_{11} & B_{12} & B_{13} \\ 0 & B_{22} & B_{23} \\ B_{31} & B_{32} & B_{33} \end{pmatrix},\ Q_K = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix},\ Q_L = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix}. \tag{3.282}$$

So, now we have a 'little hole'. Under the conditions (3.281) consider

$$P = BQ_KB^{-1}(\mathrm{Id} - Q_L). \tag{3.283}$$

The condition $Q_LBQ_K = 0$ and the definition show that $Q_LP = 0 = PQ_L$. Moreover,

$$P^2 = BQ_KB^{-1}(\mathrm{Id} - Q_L)BQ_KB^{-1}(\mathrm{Id} - Q_L) = BQ_KB^{-1}BQ_KB^{-1}(\mathrm{Id} - Q_L) = P.$$

So, P is a projection which acts on the range of $\mathrm{Id} - Q_L$; from its definition, the range of P is contained in the range of BQ_K. Since

$$PBQ_K = BQ_KB^{-1}(\mathrm{Id} - Q_L)BQ_K = BQ_K$$

it follows that P is a projection *onto* the range of BQ_K.

The next part of the proof can be thought of as a result on 3×3 matrices but applied to a decomposition of Hilbert space. First, observe a little result on rotations.

Lemma 3.26. *If P and Q are projections on a Hilbert space with $PQ = QP = 0$ and $M = MP = QM$ restricts to an isomorphism from the range of P to the range of Q with 'inverse' $M' = M'Q = PM'$ (so $M'M = P$ and $MM' = Q$)*

$$[-\pi/2, \pi/2] \ni \theta \longmapsto R(\theta) = \cos\theta P + \sin\theta M - \sin\theta M' + \cos\theta Q + (\mathrm{Id} - P - Q) \tag{3.284}$$

is a path in the space of invertible operators such that

$$R(0)P = P,\ R(\pi/2)P = M'P. \tag{3.285}$$

Proof. Computing directly, $R(\theta)R(-\theta) = \mathrm{Id}$ from which the invertibility follows as does (3.285). □

We have shown above that the projection P has range equal to the range of BQ_K; apply Lemma 3.26 with $M = S(BQ_K)^{-1}P$ where S is a fixed isomorphism of the range of Q_K to the range of Q_L. Then

$$L_1(\theta) = R(\theta)B \text{ has } L_1(0) = B,\ L(\pi/2) = B' \text{ with } B'Q_K = Q_LSQ_K \tag{3.286}$$

an isomorphism onto the range of Q.

Next apply Lemma 3.26 again but for the projections Q_K and Q_L with the isomorphism S, giving

$$R'(\theta) = \cos\theta Q_K + \sin\theta S - \sin\theta S' + \cos\theta Q_L + Q_O. \tag{3.287}$$

Then the curve of invertibles

$$L_2(\theta) = R'(\theta - \theta')B' \text{ has } L(0) = B',\ L(\pi/2) = B'',\ B''Q_K = Q_K.$$

So, we have succeed by succesive homotopies through invertible elements in arriving at an operator

$$B'' = \begin{pmatrix} \mathrm{Id} & E \\ 0 & F \end{pmatrix} \tag{3.288}$$

where we are looking at the decomposition of $H = H \oplus H$ according to the projections Q_K and $\mathrm{Id} - Q_K$. The invertibility of this is equivalent to the invertibility of F and the homotopy

$$B''(s) = \begin{pmatrix} \mathrm{Id} & (1-s)E \\ 0 & F \end{pmatrix} \tag{3.289}$$

connects it to

$$L = \begin{pmatrix} \mathrm{Id} & 0 \\ 0 & F \end{pmatrix},\ (B''(s))^{-1} = \begin{pmatrix} \mathrm{Id} & -(1-s)EF^{-1} \\ 0 & F^{-1} \end{pmatrix} \tag{3.290}$$

through invertibles.

The final step is 'Eilenberg's swindle'. In (3.290) we arrived at a family of operators on $H \oplus H$. Reversing the factors we can consider

$$\begin{pmatrix} F & 0 \\ 0 & \mathrm{Id} \end{pmatrix}. \tag{3.291}$$

Eilenberg's idea is to connect this to the identity by an explicit curve which 'only uses F' and so is uniform in parameters. So for the moment just take F to be a fixed unitary operator on H.

We use several isomorphism involving H and $l^2(H)$ which are isomorphic of course, as separable Hilbert spaces. First consider the two simple 'rotations' on $H \oplus H$

$$\begin{pmatrix} \mathrm{Id}\cos t & F\sin t \\ -F^{-1}\sin t & \mathrm{Id}\cos t \end{pmatrix},\ \begin{pmatrix} F\cos t & F\sin t \\ -F^{-1}\sin t & \cos tF^{-1} \end{pmatrix}. \tag{3.292}$$

These are both unitary norm-continuous curves, the first starts at the identity and is off-diagonal at $t = \pi/2$ and equal to the second at that point. So by reversing the second and going back we connect

$$\mathrm{Id} = \begin{pmatrix} \mathrm{Id} & 0 \\ 0 & \mathrm{Id} \end{pmatrix} \text{ to } \begin{pmatrix} F & 0 \\ 0 & F^{-1} \end{pmatrix}. \tag{3.293}$$

Now, we can also identify $H \oplus H$ with $l^2(H \oplus H)$. So an element of this space is an l^2 sequence with values in $H \oplus H$ and the identity just acts as the identity on each 2×2 block. We can perform the to-and-fro rotation in (3.292) in each block. That this is actually a norm-continuous curve acting on $l^2(H \oplus H)$ is a consequence of the fact that it is 'the same' in each block and so it is actually a sequence of operators, each on $H \oplus H$, which are continuous, and uniformly so with respect to the index i corresponding to a sequence in l^2; so the whole operator is continuous.

This connects Id to the second matrix (3.293) acting in each block of $l^2(H \oplus H)$. So, here is one part of the swindle, we can reorder the space so it becomes $l^2(H)$ where now the operator is diagonal but with alternating entries

$$\mathrm{Diag}(F^{-1}, F, F^{-1}, F, \dots) \text{ on } l^2(H). \tag{3.294}$$

This requires just a unitary isomorphism corresponding to relabelling the basis elements.

Now, go back to the operator (3.291) and look at the lower left identity element acting on H. We can identify the H in this spot with $l^2(H)$ and then we have a curve linking this entry to (3.294). For the whole operator this gives a norm-continuous curve connecting

$$\begin{pmatrix} F & 0 \\ 0 & \mathrm{Id} \end{pmatrix} \text{ to } \mathrm{Diag}(F, F^{-1}, F, F^{-1}, F, \dots) \text{ on } H \oplus l^2(H) \tag{3.295}$$

just adding the first entry. But now we reverse the procedure using F^{-1} in place of F so the end-point in (3.295) is connected to the identity on $l^2(H) = H$!

The fact that this construction only uses F itself and 2×2 matrices means that it works uniformly when F depends continously on parameters in a compact set. So we have constructed a curve as desired in (3.275) and hence we have proved:-

Theorem 3.6 (Kuiper). *For any compact subset $X \subset \mathrm{GL}(H)$ there is a retraction γ as in* (3.275).

Note that it follows from a result of Milnor (on CW complexes) that in this case contractibility follows from weak contractibility. If you are topologically inclined you might like to look up some applications of Kuiper's Theorem - for instance that the projective unitary group is a classifying space for two dimensional integral cohomology, an Eilenberg-MacLane space.

CHAPTER 4

Differential and Integral operators

In the last part of the course some more concrete analytic questions are considered. First the completeness of the Fourier basis is shown, this is one of the settings from which the notion of a Hilbert space originates. The index formula for Toeplitz operators on Hardy space is then derived. Next operator methods are used to demonstrate the uniqueness of the solutions to the Cauchy problem. The completeness of the eigenbasis for 'Sturm-Liouville' theory is then deduced from the spectral theorem. The Fourier transform is examined and used to prove the completeness of the Hermite basis for $L^2(\mathbb{R})$. Once one has all this, one can do a lot more, but there is no time left. Such is life.

1. Fourier series

Let us now try applying our knowledge of Hilbert space to a concrete Hilbert space such as $L^2(a,b)$ for a finite interval $(a,b) \subset \mathbb{R}$. Any such interval with $b > a$ can be mapped by a linear transformation onto $(0,2\pi)$ and so we work with this special interval. You showed that $L^2(a,b)$ is indeed a Hilbert space. One of the reasons for developing Hilbert space techniques originally was precisely the following result.

Theorem 4.1. *If $u \in L^2(0,2\pi)$ then the Fourier series of u,*

$$\frac{1}{2\pi}\sum_{k\in\mathbb{Z}} c_k e^{ikx},\ c_k = \int_{(0,2\pi)} u(x)e^{-ikx}dx \tag{4.1}$$

converges in $L^2(0,2\pi)$ to u.

Notice that this does not say the series converges pointwise, or pointwise almost everywhere. In fact it is true that the Fourier series of a function in $L^2(0,2\pi)$ converges almost everywhere to u, but it is hard to prove! In fact it is an important result of L. Carleson. Here we are just claiming that

$$\lim_{n\to\infty}\int |u(x) - \frac{1}{2\pi}\sum_{|k|\le n} c_k e^{ikx}|^2 = 0 \tag{4.2}$$

for any $u \in L^2(0,2\pi)$.

Our abstract Hilbert space theory has put us quite close to proving this. First observe that if $e'_k(x) = \exp(ikx)$ then these elements of $L^2(0,2\pi)$ satisfy

$$\int e'_k\overline{e'_j} = \int_0^{2\pi} \exp(i(k-j)x) = \begin{cases} 0 & \text{if } k \ne j \\ 2\pi & \text{if } k = j. \end{cases} \tag{4.3}$$

Thus the functions

$$e_k = \frac{e'_k}{\|e'_k\|} = \frac{1}{\sqrt{2\pi}}e^{ikx} \tag{4.4}$$

form an orthonormal set in $L^2(0,2\pi)$. It follows that (4.1) is just the Fourier-Bessel series for u with respect to this orthonormal set:-

$$c_k = \sqrt{2\pi}(u, e_k) \Longrightarrow \frac{1}{2\pi} c_k e^{ikx} = (u, e_k)e_k. \tag{4.5}$$

So, we already know that this series converges in $L^2(0,2\pi)$ thanks to Bessel's inequality. So 'all' we need to show is

Proposition 4.1. *The e_k, $k \in \mathbb{Z}$, form an orthonormal basis of $L^2(0,2\pi)$, i.e. are complete:*

$$\int ue^{ikx} = 0 \;\forall\; k \Longrightarrow u = 0 \;\textit{in}\; L^2(0,2\pi). \tag{4.6}$$

This however, is not so trivial to prove. An equivalent statement is that the finite linear span of the e_k is *dense* in $L^2(0,2\pi)$. I will prove this using Fejér's method. In this approach, we check that any continuous function on $[0,2\pi]$ satisfying the additional condition that $u(0) = u(2\pi)$ is the uniform limit on $[0,2\pi]$ of a sequence in the finite span of the e_k. Since uniform convergence of continuous functions certainly implies convergence in $L^2(0,2\pi)$ and we already know that the continuous functions which vanish near 0 and 2π are dense in $L^2(0,2\pi)$ this is enough to prove Proposition 4.1. However the proof is a serious piece of analysis, at least it seems so to me! There are other approaches, for instance we could use the Stone-Weierstrass Theorem; rather than do this we will deduce the Stone-Weierstrass Theorem from Proposition 4.1. Another good reason to proceed directly is that Fejér's approach is clever and generalizes in various ways as we will see.

So, the problem is to find the sequence in the span of the e_k which converges to a given continuous function and the trick is to use the Fourier expansion that we want to check! The idea of Cesàro is close to one we have seen before, namely to make this Fourier expansion 'converge faster', or maybe better. For the moment we can work with a general function $u \in L^2(0,2\pi)$ – or think of it as continuous if you prefer. The truncated Fourier series of u is a finite linear combination of the e_k :

$$U_n(x) = \frac{1}{2\pi} \sum_{|k| \le n} (\int_{(0,2\pi)} u(t) e^{-ikt} dt) e^{ikx} \tag{4.7}$$

where I have just inserted the definition of the c_k's into the sum. Since this is a finite sum we can treat x as a parameter and use the linearity of the integral to write it as

$$U_n(x) = \int_{(0,2\pi)} D_n(x-t)u(t),\; D_n(s) = \frac{1}{2\pi} \sum_{|k| \le n} e^{iks}. \tag{4.8}$$

Now this sum can be written as an explicit quotient, since, by telescoping,

$$2\pi D_n(s)(e^{is/2} - e^{-is/2}) = e^{i(n+\frac{1}{2})s} - e^{-i(n+\frac{1}{2})s}. \tag{4.9}$$

So in fact, at least where $s \neq 0$,

$$D_n(s) = \frac{e^{i(n+\frac{1}{2})s} - e^{-i(n+\frac{1}{2})s}}{2\pi(e^{is/2} - e^{-is/2})} \tag{4.10}$$

and the limit as $s \to 0$ exists just fine.

As I said, Cesàro's idea is to speed up the convergence by replacing U_n by its average

$$V_n(x) = \frac{1}{n+1}\sum_{l=0}^{n} U_l. \tag{4.11}$$

Again plugging in the definitions of the U_l's and using the linearity of the integral we see that

$$V_n(x) = \int_{(0,2\pi)} S_n(x-t)u(t),\ S_n(s) = \frac{1}{n+1}\sum_{l=0}^{n} D_l(s). \tag{4.12}$$

So again we want to compute a more useful form for $S_n(s)$ – which is the Fejér kernel. Since the denominators in (4.10) are all the same,

$$2\pi(n+1)(e^{is/2} - e^{-is/2})S_n(s) = \sum_{l=0}^{n} e^{i(l+\frac{1}{2})s} - \sum_{l=0}^{n} e^{-i(l+\frac{1}{2})s}. \tag{4.13}$$

Using the same trick again,

$$(e^{is/2} - e^{-is/2})\sum_{l=0}^{n} e^{i(l+\frac{1}{2})s} = e^{i(n+1)s} - 1 \tag{4.14}$$

so

$$\begin{gathered} 2\pi(n+1)(e^{is/2} - e^{-is/2})^2 S_n(s) = e^{i(n+1)s} + e^{-i(n+1)s} - 2 \\ \Longrightarrow S_n(s) = \frac{1}{n+1}\frac{\sin^2(\frac{(n+1)}{2}s)}{2\pi\sin^2(\frac{s}{2})}. \end{gathered} \tag{4.15}$$

Now, what can we say about this function? One thing we know immediately is that if we plug $u = 1$ into the discussion above, we get $U_n = 1$ for $n \ge 0$ and hence $V_n = 1$ as well. Thus in fact

$$\int_{(0,2\pi)} S_n(x-\cdot) = 1,\ \forall\ x \in (0,2\pi). \tag{4.16}$$

Looking directly at (4.15) the first thing to notice is that $S_n(s) \ge 0$. Also, we can see that the denominator only vanishes when $s = 0$ or $s = 2\pi$ in $[0,2\pi]$. Thus if we stay away from there, say $s \in (\delta, 2\pi-\delta)$ for some $\delta > 0$ then, $\sin(t)$ being a bounded function,

$$|S_n(s)| \le (n+1)^{-1}C_\delta \text{ on } (\delta, 2\pi-\delta). \tag{4.17}$$

We are interested in how close $V_n(x)$ is to the given $u(x)$ in supremum norm, where now we will take u to be continuous. Because of (4.16) we can write

$$u(x) = \int_{(0,2\pi)} S_n(x-t)u(x) \tag{4.18}$$

where t denotes the variable of integration (and x is fixed in $[0,2\pi]$). This 'trick' means that the difference is

$$V_n(x) - u(x) = \int_{(0,2\pi)} S_n(x-t)(u(t)-u(x)). \tag{4.19}$$

For each x we split this integral into two parts, the set $\Gamma(x)$ where $x-t \in [0,\delta]$ or $x-t \in [2\pi-\delta, 2\pi]$ and the remainder. So

$$|V_n(x)-u(x)| \le \int_{\Gamma(x)} S_n(x-t)|u(t)-u(x)| + \int_{(0,2\pi)\setminus\Gamma(x)} S_n(x-t)|u(t)-u(x)|. \tag{4.20}$$

Now on $\Gamma(x)$ either $|t-x| \le \delta$ – the points are close together – or t is close to 0 and x to 2π so $2\pi - x + t \le \delta$ or conversely, x is close to 0 and t to 2π so $2\pi - t + x \le \delta$. In any case, by assuming that $u(0) = u(2\pi)$ and using the uniform continuity of a continuous function on $[0,2\pi]$, given $\epsilon > 0$ we can choose δ so small that

$$|u(x)-u(t)| \le \epsilon/2 \text{ on } \Gamma(x). \tag{4.21}$$

On the complement of $\Gamma(x)$ we have (4.17) and since u is bounded we get the estimate

$$|V_n(x)-u(x)| \le \epsilon/2 \int_{\Gamma(x)} S_n(x-t) + (n+1)^{-1}q(\delta) \le \epsilon/2 + (n+1)^{-1}q(\delta) \tag{4.22}$$

where $q(\delta) = 2\sin(\delta/2)^{-2}\sup|u|$ is a positive constant depending on δ (and u). Here the fact that S_n is non-negative and has integral one has been used again to bound the integral of $S_n(x-t)$ over $\Gamma(x)$ by 1. Having chosen δ to make the first term small, we can choose n large to make the second term small and it follows that

$$V_n(x) \to u(x) \text{ uniformly on } [0,2\pi] \text{ as } n \to \infty \tag{4.23}$$

under the assumption that $u \in \mathcal{C}([0,2\pi])$ satisfies $u(0) = u(2\pi)$.

So this proves Proposition 4.1 subject to the density in $L^2(0,2\pi)$ of the continuous functions which vanish near (but not of course in a fixed neighbourhood of) the ends. In fact we know that the L^2 functions which vanish near the ends are dense since we can chop off and use the fact that

$$\lim_{\delta\to 0}\left(\int_{(0,\delta)}|f|^2 + \int_{(2\pi-\delta,2\pi)}|f|^2\right) = 0. \tag{4.24}$$

This proves Theorem 4.1.

Notice that from what we have shown it follows that the finite linear combinations of the $\exp(ikx)$ are dense in the subspace of $\mathcal{C}([0,2\pi])$ consisting of the functions with equal values at the ends. Taking a general element $u \in \mathcal{C}([0,2\pi]$ we can choose constants so that

$$v = u - c - dx \in \mathcal{C}([0,2\pi]) \text{ satisfies } v(0) = v(2\pi) = 0. \tag{4.25}$$

Indeed we just need to take $c = u(0)$, $d = u(1) - c$. Then we know that v is the uniform limit of a sequence of finite sums of the $\exp(ikx)$. However, the Taylor series

$$e^{ikx} = \sum_l \frac{(ik)^l}{l!}x^l \tag{4.26}$$

converges uniformly to e^{ikx} in any (complex) disk. So it follows in turn that the polynomials are dense

Theorem 4.2 (Stone-Weierstrass). *The polynomials are dense in $\mathcal{C}([a,b])$ for any $a < b$, in the uniform topology.*

Make sure you understand the change of variable argument to get to a general (finite) interval.

2. Toeplitz operators

Although the convergence of Fourier series was stated above for functions on an interval $(0, 2\pi)$ it can be immediately reinterpreted in terms of periodic functions on the line, or equivalently functions on the circle $\mathbb{S}$. Namely a 2π-periodic function

$$u : \mathbb{R} \longrightarrow \mathbb{C},\ u(x + 2\pi) = u(x)\ \forall\ x \in \mathbb{R} \tag{4.27}$$

is uniquely determined by its restriction to $[0, 2\pi)$ by just iterating to see that

$$u(x + 2\pi k) = u(x),\ x \in [0, 2\pi),\ k \in \mathbb{Z}. \tag{4.28}$$

Conversely a function on $[0, 2\pi)$ determines a 2π-periodic function this way. Thus a function on the circle

$$\mathbb{S} = \{z \in \mathbb{C} : |z| = 1\} \tag{4.29}$$

is the same as a periodic function on the line in terms of the standard angular variable

$$\mathbb{S} \ni z = e^{2\pi i\theta},\ \theta \in [0, 2\pi). \tag{4.30}$$

In particular we can identify $L^2(\mathbb{S})$ with $L^2(0, 2\pi)$ in this way – since the missing end-point corresponds to a set of measure zero. Equivalently this identifies $L^2(\mathbb{S})$ as the *locally square integrable* functions on $\mathbb{R}$ which are 2π-periodic.

Since $\mathbb{S}$ is a compact Lie group (what is that you say? Look it up!) this brings us into the realm of harmonic analysis. Just restating the results above for any $u \in L^2(\mathbb{S})$ the Fourier series (thinking of each $\exp(ik\theta)$ as a 2π-periodic function on the line) converges in $L^2(I)$ for any bounded interval

$$u(x) = \sum_{k\in\mathbb{Z}} a_k e^{ikx},\ a_k = \int_{(0,2\pi)} u(x)e^{-ikx}dx. \tag{4.31}$$

After this adjustment of attitude, we follow G.H. Hardy (you might enjoy ”A Mathematician’s Apology”) in thinking about:

Definition 4.1. Hardy space is

$$H = \{u \in L^2(\mathbb{S}); a_k = 0\ \forall\ k < 0\}. \tag{4.32}$$

There are lots of reasons to be interested in $H \subset L^2(\mathbb{S})$ but for the moment note that it is a closed subspace – since it is the intersection of the null spaces of the continuous linear functionals $H \longmapsto a_k,\ k < 0$. Thus there is a unique orthogonal projection

$$\pi_H : L^2(\mathbb{S}) \longrightarrow H \tag{4.33}$$

with range H.

If we go back to the definition of $L^2(\mathbb{S})$ we can see that a continuous function $\alpha \in \mathcal{C}(\mathbb{S})$ defines a bounded linear operator on $L^2(\mathbb{S})$ by multiplication. It is invertible if and only if $\alpha(\theta) \neq 0$ for all $\theta \in [0, 2\pi)$ which is the same as saying that α is a continuous map

$$\alpha : \mathbb{S} \longrightarrow \mathbb{C}^* = \mathbb{C} \setminus \{0\}. \tag{4.34}$$

For such a map there is a well-defined ‘winding number’ giving the number of times that the curve in the plane defined by α goes around the origin. This is easy to define using the properties of the logarithm. Suppose that α is once continuously differentiable and consider

$$\frac{1}{2\pi i}\int_{[0,2\pi]}\alpha^{-1}\frac{d\alpha}{d\theta}d\theta = \operatorname{wn}(\alpha). \tag{4.35}$$

If we can write

$$\alpha = \exp(2\pi i f(\theta)) \tag{4.36}$$

with $f : [0, 2\pi] \longrightarrow \mathbb{C}$ continuous then necessarily f is differentiable and

$$\operatorname{wn}(\alpha) = \int_0^{2\pi}\frac{df}{d\theta}d\theta = f(2\pi) - f(0) \in \mathbb{Z} \tag{4.37}$$

since $\exp(2\pi i(f(0) - f(2\pi)))) = 1$. In fact, even for a general $\alpha \in \mathcal{C}(\mathbb{S};\mathbb{C}^*)$, it is always possible to find a continuous f satisfying (4.36), using the standard properties of the logarithm as a local inverse to exp, but ill-determined up to addition of integral multiples of $2\pi i$. Then the winding number is given by the last expression in (4.37) and is independent of the choice of f.

Definition 4.2. A Toeplitz operator on H is an operator of the form

$$T_\alpha = \pi_H\alpha\pi_H : H \longrightarrow H,\ \alpha \in \mathcal{C}(\mathbb{S}). \tag{4.38}$$

The result I want is one of the first 'geometric index theorems' – it is a very simple case of the celebrated Atiyah-Singer index theorem (which it much predates).

Theorem 4.3 (Toeplitz). *If $\alpha \in \mathcal{C}(\mathbb{S};\mathbb{C}^*)$ then the Toeplitz operator (4.38) is Fredholm (on the Hardy space H) with index*

$$\operatorname{ind}(T_\alpha) = -\operatorname{wn}(\alpha) \tag{4.39}$$

given in terms of the winding number of α.

Proof. First we need to show that T_α is indeed a Fredholm operator. To do this we decompose the original, multiplication, operator into four pieces

$$\alpha = T_\alpha + \pi_H\alpha(\operatorname{Id}-\pi_H) + (\operatorname{Id}-\pi_H)\alpha\pi_H + (\operatorname{Id}-\pi_H)\alpha(\operatorname{Id}-\pi_H) \tag{4.40}$$

which you can think of as a 2×2 matrix corresponding to writing

$$\begin{gathered} L^2(\mathbb{S}) = H \oplus H_-,\ H_- = (\operatorname{Id}-\pi_H)L^2(\mathbb{S}) = H^\perp, \\ \alpha = \begin{pmatrix} T_\alpha & \pi_H\alpha(\operatorname{Id}-\pi_H) \\ (\operatorname{Id}-\pi_H)\alpha\pi_H & (\operatorname{Id}-\pi_H)\alpha(\operatorname{Id}-\pi_H) \end{pmatrix}. \end{gathered} \tag{4.41}$$

Now, we will show that the two 'off-diagonal' terms are compact operators (on $L^2(\mathbb{S})$). Consider first $(\operatorname{Id}-\pi_H)\alpha\pi_H$. It was shown above, as a form of the Stone-Weierstrass Theorem, that the finite Fourier sums are dense in $\mathcal{C}(\mathbb{S})$ in the supremum norm. This is not the convergence of the Fourier series but there is a sequence $\alpha_k \to \alpha$ in supremum norm, where each

$$\alpha_k = \sum_{j=-N_k}^{N_k} a_{kj}e^{ij\theta}. \tag{4.42}$$

It follows that

$$\|(\operatorname{Id}-\pi_H)\alpha_k\pi_H - (\operatorname{Id}-\pi_H)\alpha\pi_H\|_{\mathcal{B}(L^2(\mathbb{S}))} \to 0. \tag{4.43}$$

Now by (4.42) each $(\mathrm{Id}-\pi_H)\alpha_k\pi_H$ is a finite linear combination of terms

$$(\mathrm{Id}-\pi_H)e^{ij\theta}\pi_H,\ |j|\le N_k. \tag{4.44}$$

However, each of these operators is of finite rank. They actually vanish if $j\ge 0$ and for $j<0$ the rank is exactly $-j$. So each $(\mathrm{Id}-\pi_H)\alpha_k\pi_H$ is of finite rank and hence $(\mathrm{Id}-\pi_H)\alpha\pi_H$ is compact. A very similar argument works for $H\alpha(\mathrm{Id}-H)$ (or you can use adjoints).

Now, again assume that $\alpha\ne 0$ does not vanish anywhere. Then the whole multiplication operator in (4.40) is invertible. If we remove the two compact terms we see that

$$T_\alpha+(\mathrm{Id}-\pi_H)\alpha(\mathrm{Id}-\pi_H) \text{ is Fredholm} \tag{4.45}$$

since the Fredholm operators are stable under addition of compact operators. Here the first part maps H to H and the second maps H_- to H_-. It follows that the null space and range of T_α are the projections of the null space and range of the sum (4.45) – so it must have finite dimensional null space and closed range with a finite-dimensional complement as a map from H to itself:-

$$\alpha\in\mathcal{C}(\mathbb{S};\mathbb{C}^*)\Longrightarrow T_\alpha \text{ is Fredholm in } \mathcal{B}(H). \tag{4.46}$$

So it remains to compute its index. Note that the index of the sum (4.45) acting on $L^2(\mathbb{S})$ vanishes, so that does not really help! The key here is the stability of both the index and the winding number.

Lemma 4.1. *If $\alpha\in\mathcal{C}(\mathbb{S};\mathbb{C}^*)$ has winding number $p\in\mathbb{Z}$ then there is a curve*

$$\alpha_t:[0,1]\longrightarrow\mathcal{C}(\mathbb{S};\mathbb{C}^*),\ \alpha_1=\alpha,\ \alpha_0=e^{ip\theta}. \tag{4.47}$$

Proof. If you take a continuous function $f:[0,2\pi]\longrightarrow\mathbb{C}$ then

$$\alpha=\exp(2\pi i f)\in\mathcal{C}(\mathbb{S};\mathbb{C}^*) \text{ iff } f(2\pi)=f(0)+p,\ p\in\mathbb{Z} \tag{4.48}$$

(so that $\alpha(2\pi)=\alpha(0)$) where p is precisely the winding number of α. So to construct a continuous family as in (4.47) we can deform f instead provided we keep the difference between the end values constant. Clearly

$$\alpha_t=\exp(2\pi i f_t),\ f_t(\theta)=p\frac{\theta}{2\pi}(1-t)+f(\theta)t,\ t\in[0,1] \tag{4.49}$$

does this since $f_t(0)=f(0)t,\ f_t(2\pi)=p(1-t)+f(2\pi)t=f(0)t+p,\ f_0=p\frac{\theta}{2\pi},$ $f_1(\theta)=f(\theta)$. □

It was shown above that the index of a Fredholm operator is constant on the components – so along any norm continuous curve such as T_{α_t} where α_t is as in (4.47). Thus the index of T_α, where α has winding number p is the same as the index of the Toeplitz operator defined by $\exp(ip\theta)$, which has the same winding number (note that the winding number is also constant under deformations of α). So we are left to compute the index of the operator $\pi_He^{ip\theta}\pi_H$ acting on H. This is just a p-fold 'shift up'. If $p\le 0$ it is actually surjective and has null space spanned by the $\exp(ij\theta)$ with $0\le j<-p$ – since these are mapped to $\exp(i(j+p)\theta)$ and hence killed by π_H. Thus indeed the index of T_α for $\alpha=\exp(ip\theta)$ is $-p$ in this case. For $p>0$ we can take the adjoint so we have proved Theorem 4.3. □

Why is this important? Suppose you have a function $\alpha \in \mathcal{C}(\mathbb{S};\mathbb{C}^*)$ and you know it has winding number $-k$ for $k \in \mathbb{N}$. Then you know that the operator T_α must have null space *at least* of dimension k. It could be bigger but this is an existence theorem hence useful. The index is generally *relatively* easy to compute and from that one can tell quite a lot about a Fredholm operator.

3. Cauchy problem

Most, if not all, of you will have had a course on ordinary differential equations so the results here are probably familiar to you at least in outline. I am not going to try to push things very far but I will use the Cauchy problem to introduce 'weak solutions' of differential equations.

So, here is a form of the Cauchy problem. Let me stick to the standard interval we have been using but as usual it does not matter. So we are interested in solutions u of the equation, for some positive integer k

$$\begin{gathered} Pu(x) = \frac{d^k u}{dx^k}(x) + \sum_{j=0}^{k-1} a_k(x) \frac{d^j u}{dx^j}(x) = f(x) \text{ on } [0, 2\pi] \\ \frac{d^j u}{dx^j}(0) = 0,\ j = 0, \dots, k-1 \\ a_j \in \mathcal{C}^j([0, 2\pi]),\ j = 0, \dots, k-1. \end{gathered} \tag{4.50}$$

So, the a_j are fixed (corresponding if you like to some physical system), u is the 'unknown' function and f is also given. Recall that $\mathcal{C}^j([0,2\pi])$ is the space (complex valued here) of functions on $[0, 2\pi]$ which have j continuous derivatives. The middle line consists of the 'homogeneous' Cauchy conditions – also called initial conditions – where homogeneous just means zero. The general case of non-zero initial conditions follows from this one.

If we want the equation to make 'classical sense' we need to assume for instance that u has continuous derivatives up to order k and f is continuous. I have written out the first term, involving the highest order of differentiation, in (4.50) separately to suggest the following observation. Suppose u is just k times differentiable, but without assuming the kth derivative is continous. The equation still makes sense but if we assume that f is continuous then it actually follows that u is k times continuously differentiable. In fact each of the terms in the sum is continuous, since this only invovles derivatives up to order $k-1$ multiplied by continuous functions. We can (mentally if you like) move these to the right side of the equation, so together with f this becomes a continuous function. But then the equation itself implies that $\frac{d^k u}{dx^k}$ is continuous and so u is actually k times continuously differentiable. This is a rather trivial example of 'elliptic' regularity which we will push much further.

So, the problem is to prove

Theorem 4.4. *For each $f \in \mathcal{C}([0, 2\pi])$ there is a unique k times continuously differentiable solution, u, to* (4.50).

Note that in general there is no way of 'writing the solution down'. We can show it exists, and is unique, and we can say a lot about it but there is no formula – although we will see that it is the sum of a reasonable series.

How to proceed? There are many ways but to adopt the one I want to use I need to manipulate the equation in (4.50). There is a certain discriminatory property of the way I have written the equation. Although it seems rather natural, writing the 'coefficients' a_k on the left involves an element of 'handism' if that is a legitimate concept. Instead we could try for the 'rigthist' approach and look at the similar equation

$$\begin{gathered}\frac{d^k u}{dx^k}(x)+\sum_{j=0}^{k-1}\frac{d^j(b_j(x)u)}{dx^j}(x)=f(x) \text{ on } [0,2\pi]\\ \frac{d^j u}{dx^j}(0)=0,\ j=0,\dots,k-1\\ b_j\in\mathcal{C}^j([0,2\pi]),\ j=0,\dots,k-1.\end{gathered}\tag{4.51}$$

As already written in (4.50) we think of P as an operator, sending u to this sum.

LEMMA 4.2. *For any functions $a_j\in\mathcal{C}^j([0,2\pi])$ there are unique functions $b_j\in\mathcal{C}^j([0,2\pi])$ so that* (4.51) *gives the same operator as* (4.50).

PROOF. Here we can simply write down a formula for the b_j in terms of the a_j. Namely the product rule for derivatives means that

$$\frac{d^j(b_j(x)u)}{dx^j}=\sum_{p=0}^{j}\binom{j}{p}\frac{d^{j-p}b_j}{dx^{j-p}}\cdot\frac{d^p u}{dx^p}.\tag{4.52}$$

If you are not quite confident that you know this, you do know it for $j=1$ which is just the usual product rule. So proceed by induction over j and observe that the formula for $j+1$ follows from the formula for j using the properties of the binomial coefficients.

Pulling out the coefficients of a fixed derivative of u show that we need b_j to satisfy

$$a_p=b_p+\sum_{j=p+1}^{k-1}\binom{j}{p}\frac{d^{j-p}b_j}{dx^{j-p}}.\tag{4.53}$$

This shows the uniqness since we can recover the a_j from the b_j. On the other hand we can solve (4.53) for the b_j too. The 'top' equation says $a_{k-1}=b_{k-1}$ and then successive equations determine b_p in terms of a_p and the b_j with $j>p$ which we already know iteratively.

Note that the $b_j\in\mathcal{C}^j([0,2\pi])$. □

So, what has been achieved by 'writing the coefficients on the right'? The important idea is that we can solve (4.50) in one particular case, namely when all the a_j (or equivalently b_j) vanish. Then we would just integrate k times. Let us denote Riemann integration by

$$I:\mathcal{C}([0,2\pi])\longrightarrow\mathcal{C}([0,2\pi]),\ If(x)=\int_0^x f(s)ds.\tag{4.54}$$

Of course we can also think of this as Lebesgue integration and then we know for instance that

$$I:L^2(0,2\pi)\longrightarrow\mathcal{C}([0,2\pi])\tag{4.55}$$

is a bounded linear operator. Note also that

$$(If)(0)=0\tag{4.56}$$

satisfies the first of the Cauchy conditions.

Now, we can apply the operator I to (4.51) and repeat k times. By the fundamental theorem of calculus

$$u \in \mathcal{C}^j([0,2\pi]),\ \frac{d^p u}{dx^p}(0) = 0,\ p = 0, \dots, j \Longrightarrow I^j(\frac{d^j u}{dx^j}) = u. \tag{4.57}$$

Thus (4.51) becomes

$$(\mathrm{Id} + B)u = u + \sum_{j=0}^{k-1} I^{k-j}(b_j u) = I^k f. \tag{4.58}$$

Notice that this argument is reversible. Namely if $u \in \mathcal{C}^k([0,2\pi])$ satisfies (4.58) for $f \in \mathcal{C}([0,2\pi])$ then $u \in \mathcal{C}^k([0,2\pi])$ does indeed satisfy (4.58). In fact even more is true

Proposition 4.2. *The operator* $\mathrm{Id} + B$ *is invertible on* $L^2(0,2\pi)$ *and if* $f \in \mathcal{C}([0,2\pi])$ *then* $u = (\mathrm{Id} + B)^{-1} I^k f \in \mathcal{C}^k([0,2\pi])$ *is the unique solution of* (4.51).

Proof. From (4.58) we see that B is given as a sum of operators of the form $I^p \circ b$ where b is multiplcation by a continuous function also denoted $b \in \mathcal{C}([0,2\pi])$ and $p \geq 1$. Writing out I^p as an iterated (Riemann) integral

$$I^p v(x) = \int_0^x \int_0^{y_1} \cdots \int_0^{y_{p-1}} v(y_p) dy_p \cdots dy_1. \tag{4.59}$$

For the case of $p = 1$ we can write

$$(I \cdot b_{k-1})v(x) = \int_0^{2\pi} \beta_{k-1}(x,t)v(t)dt,\ \beta_{k-2}(x,t) = H(x-t)b_{k-1}(x) \tag{4.60}$$

where the Heaviside function restricts the integrand to $t \leq x$. Similarly in the next case by reversing the order of integration

$$\begin{aligned}(I^2 \cdot b_{k-2})v(x) &= \int_0^x \int_0^s b(t)v(t)dtds \\ &= \int_0^x \int^x b_{k-2}(t)v(t)dsdt = \int_0^{2\pi} \beta_{k-2}(x,t)v(t)dt, \\ &\qquad \beta_{k-2} = (x-t)_+ b_{k-2}(x).\end{aligned} \tag{4.61}$$

In general

$$(I^p \cdot b_{k-p})v(x) = \int_0^{2\pi} \beta_{k-p}(x,t)v(t)dt,\ \beta_{k-p} = \frac{1}{(p-1)!}(x-t)_+^{p-1} b_{k-p}(x). \tag{4.62}$$

The explicit formula here is not that important, but (throwing away a lot of information) all the $\beta_*(t,x)$ have the property that they are of the form

$$\beta(x,t) = H(x-t)e(x,t),\ e \in \mathcal{C}([0,2\pi]^2). \tag{4.63}$$

This is a *Volterra operator*

$$Bv(x) = \int_0^{2\pi} \beta(x,t)v(t) \tag{4.64}$$

with β as in (4.63).

So now the point is that for any Volterra operator B, $\operatorname{Id}+B$ is invertible on $L^2(0,2\pi)$. In fact we can make a stronger statement that

$$B \text{ Volterra} \Longrightarrow \sum_j (-1)^j B^j \text{ converges in } \mathcal{B}(L^2(0,2\pi)). \tag{4.65}$$

This is just the Neumann series, but notice we are *not* claiming that $\|B\| < 1$ which would give the convergence as we know from earlier. Rather the key is that the powers of B behave very much like the operators I^k computed above.

Lemma 4.3. *For a Volterra operator in the sense of* (4.63) *and* (4.64)

$$B^jv(x) = \int_0^{2\pi} H(x-t)e_j(x,t)v(t),\ e_j \in \mathcal{C}([0,2\pi]^2),\ e_j \le \frac{C^j}{(j-1)!}(x-t)+^{j-1},\ j>1. \tag{4.66}$$

Proof. Proceeding inductively we can assume (4.66) holds for a given j. Then $B^{j+1} = B \circ B^j$ is of the form in (4.66) with

$$\begin{aligned} e_{j+1}(x,t) &= \int_0^{2\pi} H(x-s)e(x,s)H(s-t)e_j(s-x)ds \\ &= \int_t^x e(x,s)e_j(s-t)ds \le \sup|e|\frac{C^j}{(j-1)!}\int_t^x (s-t)_+^{j-1}ds \le \frac{C^{j+1}}{j!}(x-t)+^j \end{aligned} \tag{4.67}$$

provided $C \ge \sup|e|$. □

The estimate (4.67) means that, for a different constant

$$\|B^j\|_{L^2} \le \frac{C^j}{j-1}, j > 1 \tag{4.68}$$

which is summable, so the Neumann series (4.58) does converge.

To see the regularity of $u = (\operatorname{Id}+B)^{-1}I^kf$ when $f \in \mathcal{C}([0,2\pi])$ consider (4.58). Each of the terms in the sum maps $L^2(0,2\pi)$ into $\mathcal{C}([0,2\pi])$ so $u \in \mathcal{C}([0.2\pi])$. Proceeding iteratively, for each $p = 0, \dots, k-1$, each of these terms, $I^{k-j}(b_ju)$ maps $\mathcal{C}^p([0,e\pi])$ into $\mathcal{C}^{p+1}([0,2\pi])$ so $u \in \mathcal{C}^k([0,2\pi])$. Similarly for the Cauchy conditions. Differentiating (4.58) recovers (4.51). □

As indicated above, the case of non-vanishing Cauchy data follows from Theorem 4.4. Let

$$\Sigma : \mathcal{C}^k([0,2\pi]) \longrightarrow \mathcal{C}^k \tag{4.69}$$

denote the Cauchy data map – evaluating the function and its first $k-1$ derivatives at 0.

Proposition 4.3. *The combined map*

$$(\Sigma, P) : \mathcal{C}^k([0,2\pi]) \longrightarrow \mathbb{C}^k \oplus \mathcal{C}([0,2\pi]) \tag{4.70}$$

is an isomorphism.

Proof. The map Σ in (4.69) is certainly surjective, since it is surjective even on polynomials of degree $k-1$. Thus given $z \in \mathbb{C}^k$ there exists $v \in \mathcal{C}^k([0,2\pi])$ with $\Sigma v = z$. Now, given $f \in \mathcal{C}([0,2\pi])$ Theorem 4.4 allows us to find $w \in \mathcal{C}^k([0,2\pi])$ with $Pw = f - Pv$ and $\Sigma w = 0$. So $u = v + w$ satisfies $(\Sigma, P)u = (z, f)$ and we have shown the surjectivity of (4.70). The injectivity again follows from Theorem 4.4 so (Σ, P) is a bijection and hence and isomorphism using the Open Mapping Theorem (or directly). □

Let me finish this discussion of the Cauch problem by introducing the notion of a *weak solution.* let $\Sigma_{2\pi} : \mathcal{C}^k([0, 2\pi]) \longrightarrow \mathcal{C}^k$ be the evaluation of the Cauchy data at the top end of the interval. Then if $u \in \mathcal{C}^k([0, 2\pi])$ satisfies $\Sigma u = 0$ and $v \in \mathcal{C}([90, 2\pi])$ satisfies $\Sigma_{2\pi} v = 0$ there are no boundary terms in integration by parts for derivatives up to order k and it follows that

$$\int_{(0,2\pi)} Pu\overline{v} = \int_{(0,2\pi)} u\overline{Qv},\ Qv = (-1)^k \frac{d^k v}{dx^k} + \sum_{j=0}^{k-1} \frac{d^j \overline{a_j} v}{dx^j} \tag{4.71}$$

Thus Q is another operator just like P called the 'formal adjoint' of P.

If $Pu = f$ then (4.71) is just

$$\langle u, Qv\rangle_{L^2} = \langle f, v\rangle_{L^2}\ \forall\ v \in \mathcal{C}^k([0, 2\pi]) \text{ with } \Sigma_{2\pi} v = 0. \tag{4.72}$$

The significant point here is that (4.72) makes sense even if $u,\ f \in L^2([0, 2\pi])$.

Definition 4.3. If $u,\ f \in L^2([0, 2\pi])$ satisfy (4.72) then u is said to be a weak solution of (4.51).

Exercise 2. Prove that 'weak=strong' meaning that if $f \in \mathcal{C}([0, 2\pi])$ and $u \in L^2(0, 2\pi)$ is a weak solution of (4.72) then in fact $u \in \mathcal{C}^k([0, 2\pi])$ satisifes (4.51) 'in the classical sense'.

4. Dirichlet problem on an interval

I want to do a couple more 'serious' applications of what we have done so far. There are many to choose from, and I will mention some more, but let me first consider the Diriclet problem on an interval. I will choose the interval $[0, 2\pi]$ because we looked at it before but of course we could work on a general bounded interval instead. So, we are supposed to be trying to solve

$$-\frac{d^2u(x)}{dx^2} + V(x)u(x) = f(x) \text{ on } (0, 2\pi),\ u(0) = u(2\pi) = 0 \tag{4.73}$$

where the last part are the Dirichlet boundary conditions. I will assume that the 'potential'

$$V : [0, 2\pi] \longrightarrow \mathbb{R} \text{ is continuous and real-valued.} \tag{4.74}$$

Now, it certainly makes sense to try to solve the equation (4.73) for say a given $f \in \mathcal{C}([0, 2\pi])$, looking for a solution which is twice continuously differentiable on the interval. It may not exist, depending on V but one thing we can shoot for, which has the virtue of being explicit, is the following:

Proposition 4.4. *If $V \geq 0$ as in* (4.74) *then for each $f \in \mathcal{C}([0, 2\pi])$ there exists a unique twice continuously differentiable solution, u, to* (4.73).

There are in fact various approaches to this but we want to go through L^2 theory – not surprisingly of course. How to start?

Well, we do know how to solve (4.73) if $V \equiv 0$ since we can use (Riemann) integration. Thus, ignoring the boundary conditions for the moment, we can find a solution to $-d^2v/dx^2 = f$ on the interval by integrating twice:

$$v(x) = -\int_0^x \int_0^y f(t)dtdy \text{ satifies } -d^2v/dx^2 = f \text{ on } (0, 2\pi). \tag{4.75}$$

Moroever v really is twice continuously differentiable if f is continuous. So, what has this got to do with operators? Well, we can change the order of integration in (4.75) to write v as

$$v(x) = -\int_0^x \int_t^x f(t)dydt = \int_0^{2\pi} a(x,t)f(t)dt,\ a(x,t) = (t-x)H(x-t) \tag{4.76}$$

where the Heaviside function $H(y)$ is 1 when $y \geq 0$ and 0 when $y < 0$. Thus $a(x,t)$ is actually continuous on $[0,2\pi]\times[0,2\pi]$ since the $t-x$ factor vanishes at the jump in $H(t-x)$. So (4.76) shows that v is given by applying an integral operator, with continuous kernel on the square, to f.

Before thinking more seriously about this, recall that there is also the matter of the boundary conditions. Clearly, $v(0) = 0$ since we integrated from there. On the other hand, there is no particular reason why

$$v(2\pi) = \int_0^{2\pi} (t-2\pi)f(t)dt \tag{4.77}$$

should vanish. However, we can always add to v any linear function and still satisfy the differential equation. Since we do not want to spoil the vanishing at $x = 0$ we can only afford to add cx but if we choose the constant c correctly this will work. Namely consider

$$c = \frac{1}{2\pi}\int_0^{2\pi} (2\pi - t)f(t)dt,\ \text{then } (v + cx)(2\pi) = 0. \tag{4.78}$$

So, finally the solution we want is

$$w(x) = \int_0^{2\pi} b(x,t)f(t)dt,\ b(x,t) = \min(t,x) - \frac{tx}{2\pi} \in \mathcal{C}([0,2\pi]^2) \tag{4.79}$$

with the formula for b following by simple manipulation from

$$b(x,t) = a(x,t) + x - \frac{tx}{2\pi} \tag{4.80}$$

Thus there is a unique, twice continuously differentiable, solution of $-d^2w/dx^2 = f$ in $(0,2\pi)$ which vanishes at both end points and it is given by the *integral operator* (4.79).

Lemma 4.4. *The integral operator* (4.79) *extends by continuity from* $\mathcal{C}([0,2\pi])$ *to a compact, self-adjoint operator on* $L^2(0,2\pi)$.

Proof. Since w is given by an integral operator with a continuous real-valued kernel which is even in the sense that (check it)

$$b(t,x) = b(x,t) \tag{4.81}$$

we might as well give a more general result. □

Proposition 4.5. *If* $b \in \mathcal{C}([0,2\pi]^2)$ *then*

$$Bf(x) = \int_0^{2\pi} b(x,t)f(t)dt \tag{4.82}$$

defines a compact operator on $L^2(0,2\pi)$ *if in addition* b *satisfies*

$$\overline{b(t,x)} = b(x,t) \tag{4.83}$$

then B *is self-adjoint.*

Proof. If $f \in L^2((0,2\pi))$ and $v \in \mathcal{C}([0,2\pi])$ then the product $vf \in L^2((0,2\pi))$ and $\|vf\|_{L^2} \le \|v\|_\infty \|f\|_{L^2}$. This can be seen for instance by taking an absolutely summable approximation to f, which gives a sequence of continuous functions converging a.e. to f and bounded by a fixed L^2 function and observing that $vf_n \to vf$ a.e. with bound a constant multiple, $\sup |v|$, of that function. It follows that for $b \in \mathcal{C}([0,2\pi]^2)$ the product

$$b(x,y)f(y) \in L^2(0,2\pi) \tag{4.84}$$

for each $x \in [0,2\pi]$. Thus $Bf(x)$ is well-defined by (4.83) since $L^2((0,2\pi) \subset L^1((0,2\pi))$.

Not only that, but $Bf \in \mathcal{C}([0,2\pi])$ as can be seen from the Cauchy-Schwarz inequality,

(4.85)
$$|Bf(x') - Bf(x)| = |\int (b(x',y) - b(x,y))f(y)| \le \sup_y |b(x',y - b(x,y)|(2\pi)^{\frac{1}{2}} \|f\|_{L^2}.$$

Essentially the same estimate shows that

$$\sup_x |Bf(x)| \le (2\pi)^{\frac{1}{2}} \sup_{(x,y)} |b| \|f\|_{L^2} \tag{4.86}$$

so indeed, $B : L^2(0,2\pi) \longrightarrow \mathcal{C}([0,2\pi])$ is a bounded linear operator.

When b satisfies (4.83) and f and g are continuous

$$\int Bf(x)\overline{g(x)} = \int f(x)\overline{Bg(x)} \tag{4.87}$$

and the general case follows by approximation in L^2 by continuous functions.

So, we need to see the compactness. If we fix x then $b(x,y) \in \mathcal{C}([0,2\pi])$ and then if we let x vary,

$$[0,2\pi] \ni x \longmapsto b(x,\cdot) \in \mathcal{C}([0,2\pi]) \tag{4.88}$$

is continuous as a map into this Banach space. Again this is the uniform continuity of a continuous function on a compact set, which shows that

$$\sup_y |b(x',y) - b(x,y)| \to 0 \text{ as } x' \to x. \tag{4.89}$$

Since the inclusion map $\mathcal{C}([0,2\pi]) \longrightarrow L^2((0,2\pi))$ is bounded, i.e continuous, it follows that the map (I have reversed the variables)

$$[0,2\pi] \ni y \longmapsto b(\cdot,y) \in L^2((0,2\pi)) \tag{4.90}$$

is continuous and so has a compact range.

Take the Fourier basis e_k for $[0,2\pi]$ and expand b in the first variable. Given $\epsilon > 0$ the compactness of the image of (4.90) implies that the Fourier Bessel series converges uniformly (has uniformly small tails), so for some N

$$\sum_{|k|>N} |(b(x,y), e_k(x))|^2 < \epsilon \;\forall\; y \in [0,2\pi]. \tag{4.91}$$

The finite part of the Fourier series is continuous as a function of both arguments

$$b_N(x,y) = \sum_{|k|\le N} e_k(x)c_k(y),\; c_k(y) = (b(x,y), e_k(x)) \tag{4.92}$$

and so defines another bounded linear operator B_N as before. This operator can be written out as

$$B_N f(x) = \sum_{|k|\le N} e_k(x) \int c_k(y) f(y) dy \tag{4.93}$$

and so is of finite rank – it always takes values in the span of the first $2N+1$ trigonometric functions. On the other hand the remainder is given by a similar operator with corresponding to $q_N = b - b_N$ and this satisfies

$$\sup_y \|q_N(\cdot, y)\|_{L^2((0,2\pi))} \to 0 \text{ as } N \to \infty. \tag{4.94}$$

Thus, q_N has small norm as a bounded operator on $L^2((0,2\pi))$ so B is compact – it is the norm limit of finite rank operators. □

Now, recall from Problem# that $u_k = \pi^{-\frac{1}{2}} \sin(kx/2)$, $k \in \mathbb{N}$, is also an orthonormal basis for $L^2(0,2\pi)$ (it is not the Fourier basis!) Moreover, differentiating we find straight away that

$$-\frac{d^2 u_k}{dx^2} = \frac{k^2}{4} u_k. \tag{4.95}$$

Since of course $u_k(0) = 0 = u_k(2\pi)$ as well, from the uniqueness above we conclude that

$$B u_k = \frac{4}{k^2} u_k \ \forall\ k. \tag{4.96}$$

Thus, in this case we know the orthonormal basis of eigenfunctions for B. They are the u_k, each eigenspace is 1 dimensional and the eigenvalues are $4k^{-2}$.

Remark 4.1. As noted by Pavel Etingof it is worthwhile to go back to the discussion of trace class operators to see that B is indeed of trace class. Its trace can be computed in two ways. As the sum of its eigenvalues and as the integral of its kernel on the diagonal. This gives the well-known formula

$$\frac{\pi^2}{6} = \sum_{k \in \mathbb{N}} \frac{1}{k^2}. \tag{4.97}$$

This is a simple example of a 'trace formula'; you might like to look up some others!

So, this happenstance allows us to decompose B as the square of another operator defined directly on the othornormal basis. Namely

$$A u_k = \frac{2}{k} u_k \Longrightarrow B = A^2. \tag{4.98}$$

Here again it is immediate that A is a compact self-adjoint operator on $L^2(0,2\pi)$ since its eigenvalues tend to 0. In fact we can see quite a lot more than this.

Lemma 4.5. *The operator A maps $L^2(0,2\pi)$ into $\mathcal{C}([0,2\pi])$ and $Af(0) = Af(2\pi) = 0$ for all $f \in L^2(0,2\pi)$.*

Proof. If $f \in L^2(0,2\pi)$ we may expand it in Fourier-Bessel series in terms of the u_k and find

$$f = \sum_k c_k u_k, \ \{c_k\} \in l^2. \tag{4.99}$$

Then of course, by definition,

$$Af = \sum_k \frac{2c_k}{k} u_k. \tag{4.100}$$

Here each u_k is a bounded continuous function, with the bound $|u_k| \le C$ being independent of k. So in fact (4.100) converges uniformly and absolutely since it is uniformly Cauchy, for any $q > p$,

$$|\sum_{k=p}^{q} \frac{2c_k}{k} u_k| \le 2C \sum_{k=p}^{q} |c_k| k^{-1} \le 2C \left(\sum_{k=p}^{q} k^{-2} \right)^{\frac{1}{2}} \|f\|_{L^2} \tag{4.101}$$

where Cauchy-Schwarz has been used. This proves that

$$A : L^2(0, 2\pi) \longrightarrow \mathcal{C}([0, 2\pi])$$

is bounded and by the uniform convergence $u_k(0) = u_k(2\pi) = 0$ for all k implies that $Af(0) = Af(2\pi) = 0$. □

So, going back to our original problem we try to solve (4.73) by moving the Vu term to the right side of the equation (don't worry about regularity yet) and hope to use the observation that

$$u = -A^2(Vu) + A^2 f \tag{4.102}$$

should satisfy the equation and boundary conditions. In fact, let's anticipate that $u = Av$, which has to be true if (4.102) holds with $v = -AVu + Af$, and look instead for

$$v = -AVAv + Af \Longrightarrow (\mathrm{Id} + AVA)v = Af. \tag{4.103}$$

So, we know that multiplication by V, which is real and continuous, is a bounded self-adjoint operator on $L^2(0, 2\pi)$. Thus AVA is a self-adjoint compact operator so we can apply our spectral theory to it and so examine the invertibility of $\mathrm{Id} + AVA$. Working in terms of a complete orthonormal basis of eigenfunctions e_i of AVA we see that $\mathrm{Id} + AVA$ is invertible if and only if it has trivial null space, i.e. if -1 is *not* an eigenvalue of AVA. Indeed, an element of the null space would have to satisfy $u = -AVAu$. On the other hand we know that AVA is *positive* since

$$(AVAw, w) = (VAv, Av) = \int_{(0,2\pi)} V(x)|Av|^2 \ge 0 \Longrightarrow \int_{(0,2\pi)} |u|^2 = 0, \tag{4.104}$$

using the assumed non-negativity of V. So, there can be no null space – all the eigenvalues of AVA are at least non-negative and the inverse is the bounded operator given by its action on the basis

$$(\mathrm{Id} + AVA)^{-1} e_i = (1 + \tau_i)^{-1} e_i, \;\; AVAe_i = \tau_i e_i. \tag{4.105}$$

Thus $\mathrm{Id} + AVA$ is invertible on $L^2(0, 2\pi)$ with inverse of the form $\mathrm{Id} + Q$, Q again compact and self-adjoint since $(1 + \tau_i)^{-1} - 1 \to 0$. Now, to solve (4.103) we just need to take

$$v = (\mathrm{Id} + Q)Af \Longleftrightarrow v + AVAv = Af \text{ in } L^2(0, 2\pi). \tag{4.106}$$

Then indeed

$$u = Av \text{ satisfies } u + A^2Vu = A^2 f. \tag{4.107}$$

In fact since $v \in L^2(0,2\pi)$ from (4.106) we already know that $u \in \mathcal{C}([0,2\pi])$ vanishes at the end points.

Moreover if $f \in \mathcal{C}([0,2\pi])$ we know that $Bf = A^2 f$ is twice continuously differentiable, since it is given by two integrations – that is where B came from. Now, we know that u in L^2 satisfies $u = -A^2(Vu) + A^2 f$. Since $Vu \in L^2((0,2\pi) \Longrightarrow A(Vu) \in L^2(0,2\pi)$ and then, as seen above, $A(A(Vu)$ is continuous. So combining this with the result about $A^2 f$ we see that u itself is continuous and hence so is Vu. But then, going through the routine again

$$u = -A^2(Vu) + A^2 f \tag{4.108}$$

is the sum of two twice continuously differentiable functions. Thus it is so itself. In fact from the properties of $B = A^2$ it satisifes

$$-\frac{d^2u}{dx^2} = -Vu + f \tag{4.109}$$

which is what the result claims. So, we have proved the existence part of Proposition 4.4.

The uniqueness follows pretty much the same way. If there were two twice continuously differentiable solutions then the difference w would satisfy

$$-\frac{d^2w}{dx^2} + Vw = 0,\ w(0) = w(2\pi) = 0 \Longrightarrow w = -Bw = -A^2Vw. \tag{4.110}$$

Thus $w = A\phi$, $\phi = -AVw \in L^2(0,2\pi)$. Thus ϕ in turn satisfies $\phi = AVA\phi$ and hence is a solution of $(\operatorname{Id} + AVA)\phi = 0$ which we know has none (assuming $V \ge 0$). Since $\phi = 0$, $w = 0$.

This completes the proof of Proposition 4.4. To summarize, what we have shown is that $\operatorname{Id} + AVA$ is an invertible bounded operator on $L^2(0,2\pi)$ (if $V \ge 0$) and then the solution to (4.73) is precisely

$$u = A(\operatorname{Id} + AVA)^{-1}Af \tag{4.111}$$

which is twice continuously differentiable and satisfies the Dirichlet conditions for each $f \in \mathcal{C}([0,2\pi])$.

This may seem a 'round-about' approach but it is rather typical of methods from Functional Analysis. What we have done is to separate the two problems of 'existence' and 'regularity'. We first get existence of what is often called a 'weak solution' of the problem, in this case given by (4.111), which is in $L^2(0,2\pi)$ for $f \in L^2(0,2\pi)$ and then show, given regularity of the right hand side f, that this is actually a 'classical solution'.

Even if we do not assume that $V \ge 0$ we can see fairly directly what is happening.

Theorem 4.5. *For any $V \in \mathcal{C}([0,2\pi])$ real-valued, there is an orthonormal basis w_k of $L^2(0,2\pi)$ consisting of twice-continuously differentiable functions on $[0,2\pi]$, vanishing at the end-points and satisfying $-\frac{d^2w_k}{dx^2} + Vw_k = T_k w_k$ where $T_k \to \infty$ as $k \to \infty$. The equation (4.73) has a (twice continuously differentiable) solution for given $f \in \mathcal{C}([0,2\pi])$ if and only if*

$$T_k = 0 \Longrightarrow \int_{(0,2\pi)} f w_k = 0. \tag{4.112}$$

Proof. For a real-valued V we can choose a constant c such that $V + c \geq 0$. Then the eigenvalue equation we are trying to solve can be rewritten

$$-\frac{d^2w}{dx^2} + Vw = Tw \Longleftrightarrow -\frac{d^2w}{dx^2} + (V + c)w = (T + c)w. \tag{4.113}$$

Proposition 4.4 shows that there is indeed an orthonormal basis of solutions of the second equation, w_k as in the statement above with positive eigenvalues $T_k+c \to \infty$ with k.

So, only the solvability of (4.73) remains to be checked. What we have shown above is that if $f \in \mathcal{C}([0, 2\pi])$ then a twice continuously differentiable solution to

$$-\frac{d^2w}{dx^2} + Vw = f,\ w(0) = w(2\pi) = 0 \tag{4.114}$$

is precisely of the form $w = Av$ where

$$v \in L^2(0, 2\pi),\ (\mathrm{Id} + AVA)v = Af. \tag{4.115}$$

Going from (4.114) to (4.115) involves the properties of $B = A^2$ since (4.114) implies that

$$w + A^2Vw = A^2f \Longrightarrow w = Av \text{ with } v \text{ as in (4.115)}$$

Conversely if v satisfies (4.115) then $w = Av$ satisfies $w + A^2Vw = A^2f$ which implies that w has the correct regularity and satisfies (4.114).

Applying this to the case $f = 0$ shows that for twice continuously differentiable functions on $[0, 2\pi]$,

$$-\frac{d^2w}{dx^2} + Vw = 0,\ w(0) = w(2\pi) = 0 \Longleftrightarrow \\ w \in A\{v \in L^2(0, 2\pi); (\mathrm{Id} + AVA)v = 0\}. \tag{4.116}$$

Since AVA is compact and self-adjoint we see that

$$(\mathrm{Id} + AVA)v = Af \text{ has a solution in } L^2(0, 2\pi) \Longrightarrow \\ Af \perp \{v' \in L^2(0, 2\pi); (\mathrm{Id} + AVA)v' = 0\}. \tag{4.117}$$

However this last condition is equivalent to $f \perp A\{v \in L^2(0, 2\pi); (\mathrm{Id} + AVA)v = 0\}$ which is by the equivalence of (4.114) and (4.115) reduces precisely to (4.112). □

So, ultimately the solution of the differential equation (4.73) is just like the solution of a finite dimensional problem for a self-adjoint matrix. There is a solution if and only if the right side is orthogonal to the null space; it just requires a bit more work. Lots of 'elliptic' problems turn out to be like this.

We can also say (a great deal) more about the eigenvalues T_k and eigenfunctions w_k. For instance, the derivative

$$w_k'(0) \neq 0. \tag{4.118}$$

Indeed, were this to vanish w_k would be a solution of the Cauchy problem (4.50) for the second-order operator $P = \frac{d^2}{dx^2} - q + T_k$ with 'forcing term' $f = 0$ and hence, by Theorem 4.4 must itself vanish on the interval, which is a contradiction.

From this in turn it follows that the (non-trivial) eigenspaces

$$E_k(q) = \{w \in \mathcal{C}^2([0, 2\pi]); -\frac{d^2w}{dx^2} + qw = T_kw\} \tag{4.119}$$

are exactly one-dimensional. Indeed if w_k is one non-zero element, so satisfing (4.118) and w is another, then $w - w'(0)w_k/w'_k(0) \in E_k(q)$ again must satisfy the Cauchy conditions at 0 so $w = w'(0)w_k/w'_k(0)$.

EXERCISE 3. Show that the eigenfunctions functions normalized to have $w'_k(0) = 1$ are all real and w_k has exactly $k-1$ zeros in the interior of the interval.

You could try your hand at proving Borg's Theorem – if $q \in \mathcal{C}([0,2\pi])$ and the eigenvalues $T_k = \frac{k^2}{4}$ are the same as those for $q = 0$ then $q = 0$! This is the beginning of a large theory of the inverse problem – to what extent can one recover q from the knowledge of the T_k? In brief the answer is that one cannot do so in general. However q *is* determined if one knows the T_k and the 'norming constants' $w'_k(2\pi)/w'_k(0)$.

5. Harmonic oscillator

As a second 'serious' application of our Hilbert space theory I want to discuss the harmonic oscillator, the corresponding Hermite basis for $L^2(\mathbb{R})$. Note that so far we have not found an explicit orthonormal basis on the whole real line, even though we know $L^2(\mathbb{R})$ to be separable, so we certainly know that such a basis exists. How to construct one explicitly and with some handy properties? One way is to simply orthonormalize – using Gram-Schmidt – some countable set with dense span. For instance consider the basic Gaussian function

$$\exp(-\frac{x^2}{2}) \in L^2(\mathbb{R}). \tag{4.120}$$

This is so rapidly decreasing at infinity that the product with any polynomial is also square integrable:

$$x^k \exp(-\frac{x^2}{2}) \in L^2(\mathbb{R}) \ \forall\ k \in \mathbb{N}_0 = \{0, 1, 2, \dots\}. \tag{4.121}$$

Orthonormalizing this sequence gives an orthonormal basis, where completeness can be shown by an appropriate approximation technique but as usual is not so simple. This is in fact the Hermite basis as we will eventually show.

Rather than proceed directly we will work up to this by discussing the eigenfunctions of the harmonic oscillator

$$P = -\frac{d^2}{dx^2} + x^2 \tag{4.122}$$

which we want to think of as an operator – although for the moment I will leave vague the question of what it operates *on*.

As you probably already know, and we will show later, it is straightforward to show that P has a lot of eigenvectors using the 'creation' and 'annihilation' operators. We want to know a bit more than this and in particular I want to apply the abstract discussion above to this case but first let me go through the 'formal' theory. There is nothing wrong here, just that we cannot easily conclude the completeness of the eigenfunctions.

The first thing to observe is that the Gaussian is an eigenfunction of H

$$\begin{aligned} Pe^{-x^2/2} &= -\frac{d}{dx}(-xe^{-x^2/2} + x^2e^{-x^2/2}) \\ &= -(x^2-1)e^{-x^2/2} + x^2e^{-x^2/2} = e^{-x^2/2} \end{aligned} \tag{4.123}$$

with eigenvalue 1. It is an eigenfunction but not, for the moment, of a bounded operator on any Hilbert space – in this sense it is only a formal eigenfunction.

In this special case there is an essentially algebraic way to generate a whole sequence of eigenfunctions from the Gaussian. To do this, write

$$Pu=(-\frac{d}{dx}+x)(\frac{d}{dx}+x)u+u=(\operatorname{Cr}\operatorname{An}+1)u,\qquad \operatorname{Cr}=(-\frac{d}{dx}+x),\ \operatorname{An}=(\frac{d}{dx}+x) \tag{4.124}$$

again formally as operators. Then note that

$$\operatorname{An}e^{-x^2/2}=0 \tag{4.125}$$

which again proves (4.123). The two operators in (4.124) are the 'creation' operator and the 'annihilation' operator. They almost commute in the sense that

$$[\operatorname{An},\operatorname{Cr}]u=(\operatorname{An}\operatorname{Cr}-\operatorname{Cr}\operatorname{An})u=2u \tag{4.126}$$

for say any twice continuously differentiable function u.

Now, set $u_0=e^{-x^2/2}$ which is the 'ground state' and consider $u_1=\operatorname{Cr}u_0$. From (4.126), (4.125) and (4.124),

$$Pu_1=(\operatorname{Cr}\operatorname{An}\operatorname{Cr}+\operatorname{Cr})u_0=\operatorname{Cr}^2\operatorname{An}u_0+3\operatorname{Cr}u_0=3u_1. \tag{4.127}$$

Thus, u_1 is an eigenfunction with eigenvalue 3.

Lemma 4.6. *For $j\in\mathbb{N}_0=\{0,1,2,\dots\}$ the function $u_j=\operatorname{Cr}^j u_0$ satisfies $Pu_j=(2j+1)u_j$.*

Proof. This follows by induction on j, where we know the result for $j=0$ and $j=1$. Then

$$P\operatorname{Cr}u_j=(\operatorname{Cr}\operatorname{An}+1)\operatorname{Cr}u_j=\operatorname{Cr}(P-1)u_j+3\operatorname{Cr}u_j=(2j+3)u_j. \tag{4.128}$$

□

Again by induction we can check that $u_j=(2^jx^j+q_j(x))e^{-x^2/2}$ where q_j is a polynomial of degree at most $j-2$. Indeed this is true for $j=0$ and $j=1$ (where $q_0=q_1\equiv 0$) and then

$$\operatorname{Cr}u_j=(2^{j+1}x^{j+1}+\operatorname{Cr}q_j)e^{-x^2/2}. \tag{4.129}$$

and $q_{j+1}=\operatorname{Cr}q_j$ is a polynomial of degree at most $j-1$ – one degree higher than q_j.

From this it follows in fact that the finite span of the u_j consists of all the products $p(x)e^{-x^2/2}$ where $p(x)$ is any polynomial.

Now, all these functions are in $L^2(\mathbb{R})$ and we want to compute their norms. First, a standard integral computation[1] shows that

$$\int_{\mathbb{R}}(e^{-x^2/2})^2=\int_{\mathbb{R}}e^{-x^2}=\sqrt{\pi} \tag{4.130}$$

[1] To compute the Gaussian integral, square it and write as a double integral then introduce polar coordinates

$$(\int_{\mathbb{R}}e^{-x^2}dx)^2=\int_{\mathbb{R}^2}e^{-x^2-y^2}dxdy=\int_0^\infty\int_0^{2\pi}e^{-r^2}rdrd\theta=\pi\big[-e^{-r^2}\big]_0^\infty=\pi.$$

For $j > 0$, integration by parts (easily justified by taking the integral over $[-R, R]$ and then letting $R \to \infty$) gives

$$\int_{\mathbb{R}} (\operatorname{Cr}^j u_0)^2 = \int_{\mathbb{R}} \operatorname{Cr}^j u_0(x) \operatorname{Cr}^j u_0(x) dx = \int_{\mathbb{R}} u_0 \operatorname{An}^j \operatorname{Cr}^j u_0. \tag{4.131}$$

Now, from (4.126), we can move one factor of An through the j factors of Cr until it emerges and 'kills' u_0

$$\begin{aligned} \operatorname{An} \operatorname{Cr}^j u_0 &= 2 \operatorname{Cr}^{j-1} u_0 + \operatorname{Cr} \operatorname{An} \operatorname{Cr}^{j-1} u_0 \\ &= 2 \operatorname{Cr}^{j-1} u_0 + \operatorname{Cr}^2 \operatorname{An} \operatorname{Cr}^{j-2} u_0 = 2j \operatorname{Cr}^{j-1} u_0. \end{aligned} \tag{4.132}$$

So in fact,

$$\int_{\mathbb{R}} (\operatorname{Cr}^j u_0)^2 = 2j \int_{\mathbb{R}} (\operatorname{Cr}^{j-1} u_0)^2 = 2^j j! \sqrt{\pi}. \tag{4.133}$$

A similar argument shows that

$$\int_{\mathbb{R}} u_k u_j = \int_{\mathbb{R}} u_0 \operatorname{An}^k \operatorname{Cr}^j u_0 = 0 \text{ if } k \neq j. \tag{4.134}$$

Thus the functions

$$e_j = 2^{-j/2} (j!)^{-\frac{1}{2}} \pi^{-\frac{1}{4}} C^j e^{-x^2/2} \tag{4.135}$$

form an orthonormal sequence in $L^2(\mathbb{R})$.

We would like to show this orthonormal sequence is complete. Rather than argue through approximation, we can guess that in some sense the operator

$$\operatorname{An} \operatorname{Cr} = (\frac{d}{dx} + x)(-\frac{d}{dx} + x) = -\frac{d^2}{dx^2} + x^2 + 1 \tag{4.136}$$

should be invertible, so one approach is to use the ideas above of Friedrichs' extension to construct its 'inverse' and show this really exists as a compact, self-adjoint operator on $L^2(\mathbb{R})$ and that its only eigenfunctions are the e_i in (4.135). Another, more indirect approach is described below.

6. Fourier transform

The Fourier transform for functions on $\mathbb{R}$ is in a certain sense the limit of the definition of the coefficients of the Fourier series on an expanding interval, although that is not generally a good way to approach it. We know that if $u \in L^1(\mathbb{R})$ and $v \in \mathcal{C}_\infty(\mathbb{R})$ is a bounded continuous function then $vu \in L^1(\mathbb{R})$ – this follows from our original definition by approximation. So if $u \in L^1(\mathbb{R})$ the integral

$$\hat{u}(\xi) = \int e^{-ix\xi} u(x) dx, \ \xi \in \mathbb{R} \tag{4.137}$$

exists for each $\xi \in \mathbb{R}$ as a Lebesgue integral. Note that there are many different normalizations of the Fourier transform in use. This is the standard 'analyst's' normalization.

Proposition 4.6. *The Fourier tranform,* (4.137)*, defines a bounded linear map*

$$\mathcal{F} : L^1(\mathbb{R}) \ni u \longmapsto \hat{u} \in \mathcal{C}_0(\mathbb{R}) \tag{4.138}$$

into the closed subspace $\mathcal{C}_0(\mathbb{R}) \subset \mathcal{C}_\infty(\mathbb{R})$ of continuous functions which vanish at infinity (with respect to the supremum norm).

Proof. We know that the integral exists for each ξ and from the basic properties of the Lebesgue integal

$$|\hat{u}(\xi)| \le \|u\|_{L^1}, \text{ since } |e^{-ix\xi}u(x)| = |u(x)|. \tag{4.139}$$

To investigate its properties we restrict to $u \in \mathcal{C}_{\mathrm{c}}(\mathbb{R})$, a compactly-supported continuous function, say with support in $-R, R]$. Then the integral becomes a Riemann integral and the integrand is a continuous function of both variables. It follows that the Fourier transform is uniformly continuous since

$$|\hat{u}(\xi) - \hat{u}(\xi')| \le \int_{|x|\le R} |e^{-ix\xi} - e^{-ix\xi'}||u(x)|dx \le 2R \sup|u| \sup_{|x|\le R} |e^{-ix\xi} - e^{-ix\xi'}| \tag{4.140}$$

with the right side small by the uniform continuity of continuous functions on compact sets. From (4.139), if $u_n \to u$ in $L^1(\mathbb{R})$ with $u_n \in \mathcal{C}_{\mathrm{c}}(\mathbb{R})$ it follows that $\hat{u}_n \to \hat{u}$ uniformly on $\mathbb{R}$. Thus, as the uniform limit of uniformly continuous functions, the Fourier transform is uniformly continuous on $\mathbb{R}$ for any $u \in L^1(\mathbb{R})$ (you can also see this from the continuity-in-the-mean of L^1 functions).

Now, we know that even the compactly-supported once continuously differentiable functions, forming $\mathcal{C}^1_{\mathrm{c}}(\mathbb{R})$ are dense in $L^1(\mathbb{R})$ so we can also consider (4.137) where $u \in \mathcal{C}^1_{\mathrm{c}}(\mathbb{R})$. Then the integration by parts as follows is justified

$$\xi\hat{u}(\xi) = i\int (\frac{de^{-ix\xi}}{dx})u(x)dx = -i\int e^{-ix\xi}\frac{du(x)}{dx}dx. \tag{4.141}$$

Since $du/dx \in \mathcal{C}_{\mathrm{c}}(\mathbb{R})$ (by assumption) the estimate (4.139) now gives

$$\sup_{\xi\in\mathbb{R}} |\xi\hat{u}(\xi)| \le 2R \sup_{x\in\mathbb{R}} |\frac{du}{dx}|. \tag{4.142}$$

This certainly implies the weaker statement that

$$\lim_{|\xi|\to\infty} |\hat{u}(\xi)| = 0 \tag{4.143}$$

which is 'vanishing at infinity'. Now we again use the density, this time of $\mathcal{C}^1_{\mathrm{c}}(\mathbb{R})$, in $L^1(\mathbb{R})$ and the uniform estimate (4.139), plus the fact that if a sequence of continuous functions on $\mathbb{R}$ converges uniformly on $\mathbb{R}$ and each element vanishes at infinity then the limit vanishes at infinity to complete the proof of the Proposition. □

7. Fourier inversion

We could use the completeness of the orthonormal sequence of eigenfunctions for the harmonic oscillator discussed above to show that the Fourier tranform extends by continuity from $\mathcal{C}_{\mathrm{c}}(\mathbb{R})$ to define an isomorphism

$$\mathcal{F} : L^2(\mathbb{R}) \longrightarrow L^2(\mathbb{R}) \tag{4.144}$$

with inverse given by the corresponding continuous extension of

$$\mathcal{G}v(x) = (2\pi)^{-1}\int e^{ix\xi}v(\xi). \tag{4.145}$$

Instead, we will give a direct proof of the Fourier inversion formula, via Schwartz space and an elegant argument due to Hörmander. Then we will use this to prove the completeness of the eigenfunctions we have found.

We have shown above that the Fourier transform is defined as an integral if $u \in L^1(\mathbb{R})$. Suppose that in addition we know that $xu \in L^1(\mathbb{R})$. We can summarize the combined information as

$$(1+|x|)u \in L^1(\mathbb{R}). \tag{4.146}$$

LEMMA 4.7. *If u satisfies* (4.146) *then $\hat{u}$ is continuously differentiable and $d\hat{u}/d\xi = \mathcal{F}(-ixu)$ is bounded.*

PROOF. Consider the difference quotient for the Fourier transform:

$$\frac{\hat{u}(\xi+s)-\hat{u}(\xi)}{s} = \int D(x,s)e^{-ix\xi}u(x),\ D(x,s) = \frac{e^{-ixs}-1}{s}. \tag{4.147}$$

We can use the standard proof of Taylor's formula to write the difference quotient inside the integral as

$$D(x,s) = -ix\int_0^1 e^{-itxs}dt \Longrightarrow |D(x,s)| \le |x|. \tag{4.148}$$

It follows that as $s \to 0$ (along a sequence if you prefer) $D(x,s)e^{-ix\xi}u(x)$ is bounded by the $L^1(\mathbb{R})$ function $|x||u(x)|$ and converges pointwise to $-ie^{-ix\xi}xu(x)$. Dominated convergence therefore shows that the integral converges showing that the derivative exists and that

$$\frac{d\hat{u}(\xi)}{d\xi} = \mathcal{F}(-ixu). \tag{4.149}$$

From the earlier results it follows that the derivative is continuous and bounded, proving the lemma. □

Now, we can iterate this result and so conclude:

$$\begin{gathered}(1+|x|)^k u \in L^1(\mathbb{R})\ \forall\ k \Longrightarrow \\ \hat{u} \text{ is infinitely differentiable with bounded derivatives and} \\ \frac{d^k\hat{u}}{d\xi^k} = \mathcal{F}((-ix)^k u).\end{gathered} \tag{4.150}$$

This result shows that from 'decay' of u we deduce smoothness of $\hat{u}$. We can go the other way too. One way to ensure the assumption in (4.150) is to make the stronger assumption that

$$x^k u \text{ is bounded and continuous } \forall\ k. \tag{4.151}$$

Indeed, Dominated Convergence shows that if u is continuous and satisfies the bound

$$|u(x)| \le (1+|x|)^{-r},\ r > 1$$

then $u \in L^1(\mathbb{R})$. So the integrability of $x^j u$ follows from the bounds in (4.151) for $k \le j+2$. This is throwing away information but simplifies things below.

In the opposite direction, suppose that u is continuously differentiable and satisfies the estimates for some $r > 1$

$$|u(x)| + |\frac{du(x)}{dx}| \le C(1+|x|)^{-r}.$$

Then consider

$$\xi\hat{u} = i\int \frac{de^{-ix\xi}}{dx}u(x) = \lim_{R\to\infty} i\int_{-R}^{R}\frac{de^{-ix\xi}}{dx}u(x). \tag{4.152}$$

We may integrate by parts in this integral to get

$$\xi\hat{u} = \lim_{R\to\infty}\left(i\big[e^{-ix\xi}u(x)\big]_{-R}^{R} - i\int_{-R}^{R} e^{-ix\xi}\frac{du}{dx}\right). \tag{4.153}$$

The decay of u shows that the first term vanishes in the limit and the second integral converges so

$$\xi\hat{u} = \mathcal{F}(-i\frac{du}{dx}). \tag{4.154}$$

Iterating this in turn we see that if u has continuous derivatives of all orders and for all j

$$|\frac{d^ju}{dx^j}| \le C_j(1+|x|)^{-r},\ r>1 \text{ then the } \xi^j\hat{u} = \mathcal{F}((-i)^j\frac{d^ju}{dx^j}) \tag{4.155}$$

are all bounded.

Laurent Schwartz defined a space which handily encapsulates these results.

Definition 4.4. Schwartz space, $\mathcal{S}(\mathbb{R})$, consists of all the infinitely differentiable functions $u : \mathbb{R} \longrightarrow \mathbb{C}$ such that

$$\|u\|_{j,k} = \sup|x^j\frac{d^ku}{dx^k}| < \infty\ \forall\ j,\ k\ge 0. \tag{4.156}$$

This is clearly a linear space. In fact it is a complete metric space in a natural way. All the $\|\cdot\|_{j,k}$ in (4.156) are norms on $\mathcal{S}(\mathbb{R})$, but none of them is stronger than the others. So there is no natural norm on $\mathcal{S}(\mathbb{R})$ with respect to which it is complete. In the problems below you can find some discussion of the fact that

$$d(u,v) = \sum_{j,k\ge0} 2^{-j-k}\frac{\|u-v\|_{j,k}}{1+\|u-v\|_{j,k}} \tag{4.157}$$

is a complete metric. We will not use this here but it is the right way to understand what is going on.

Notice that there is some prejudice on the order of multiplication by x and differentiation in (4.156). This is only apparent, since these estimates (taken together) are equivalent to

$$\sup|\frac{d^k(x^ju)}{dx^k}| < \infty\ \forall\ j,\ k\ge0. \tag{4.158}$$

To see the equivalence we can use induction over N where the inductive statement is the equivalence of (4.156) and (4.158) for $j+k\le N$. Certainly this is true for $N=0$ and to carry out the inductive step just differentiate out the product to see that

$$\frac{d^k(x^ju)}{dx^k} = x^j\frac{d^ku}{dx^k} + \sum_{l+m<k+j} c_{l,m,k,j}x^m\frac{d^lu}{dx^l}$$

where one can be much more precise about the extra terms, but the important thing is that they all are lower order (in fact both degrees go down). If you want to be careful, you can of course prove this identity by induction too! The equivalence of (4.156) and (4.158) for $N+1$ now follows from that for N.

Theorem 4.6. *The Fourier transform restricts to a bijection on $\mathcal{S}(\mathbb{R})$ with inverse*

$$\mathcal{G}(v)(x) = \frac{1}{2\pi}\int e^{ix\xi}v(\xi). \tag{4.159}$$

Proof. The proof (due to Hörmander as I said above) will take a little while because we need to do some computation, but I hope you will see that it is quite clear and elementary.

First we need to check that $\mathcal{F} : \mathcal{S}(\mathbb{R}) \longrightarrow \mathcal{S}(\mathbb{R})$, but this is what I just did the preparation for. Namely the estimates (4.156) imply that (4.155) applies to all the $\frac{d^k(x^j u)}{dx^k}$ and so

$$\xi^k \frac{d^j \hat{u}}{d\xi^j} \text{ is continuous and bounded } \forall\ k,\ j \Longrightarrow \hat{u} \in \mathcal{S}(\mathbb{R}). \tag{4.160}$$

This indeed is why Schwartz introduced this space.

So, what we want to show is that with $\mathcal{G}$ defined by (4.159), $u = \mathcal{G}(\hat{u})$ for all $u \in \mathcal{S}(\mathbb{R})$. Notice that there is only a sign change and a constant factor to get from $\mathcal{F}$ to $\mathcal{G}$ so certainly $\mathcal{G} : \mathcal{S}(\mathbb{R}) \longrightarrow \mathcal{S}(\mathbb{R})$. We start off with what looks like a small part of this. Namely we want to show that

$$I(\hat{u}) = \int \hat{u} = 2\pi u(0). \tag{4.161}$$

Here, $I : \mathcal{S}(\mathbb{R}) \longrightarrow \mathbb{C}$ is just integration, so it is certainly well-defined. To prove (4.161) we need to use a version of Taylor's formula and then do a little computation.

Lemma 4.8. *If $u \in \mathcal{S}(\mathbb{R})$ then*

$$u(x) = u(0)\exp(-\frac{x^2}{2}) + xv(x),\ \ v \in \mathcal{S}(\mathbb{R}). \tag{4.162}$$

Proof. Here I will leave it to you (look in the problems) to show that the Gaussian

$$\exp(-\frac{x^2}{2}) \in \mathcal{S}(\mathbb{R}). \tag{4.163}$$

Observe then that the difference

$$w(x) = u(x) - u(0)\exp(-\frac{x^2}{2}) \in \mathcal{S}(\mathbb{R}) \text{ and } w(0) = 0.$$

This is clearly a necessary condition to see that $w = xv$ with $v \in \mathcal{S}(\mathbb{R})$ and we can then see from the Fundamental Theorem of Calculus that

$$w(x) = \int_0^x w'(y)dy = x\int_0^1 w'(tx)dt \Longrightarrow v(x) = \int_0^1 w'(tx) = \frac{w(x)}{x}. \tag{4.164}$$

From the first formula for v it follows that it is infinitely differentiable and from the second formula the derivatives decay rapidly since each derivative can be written in the form of a finite sum of terms $p(x)\frac{d^l w}{dx^l}/x^N$ where the p's are polynomials. The rapid decay of the derivatives of w therefore implies the rapid decay of the derivatives of v. So indeed we have proved Lemma 4.8. □

Let me set $\gamma(x) = \exp(-\frac{x^2}{2})$ to simplify the notation. Taking the Fourier transform of each of the terms in (4.162) gives

$$\hat{u} = u(0)\hat{\gamma} + i\frac{d\hat{v}}{d\xi}. \tag{4.165}$$

Since $\hat{v} \in \mathcal{S}(\mathbb{R})$,

$$\int \frac{d\hat{v}}{d\xi} = \lim_{R\to\infty} \int_{-R}^{R} \frac{d\hat{v}}{d\xi} = \lim_{R\to\infty} \left[\hat{v}(\xi)\right]_{-R}^{R} = 0. \tag{4.166}$$

So now we see that

$$\int \hat{u} = cu(0),\ c = \int \hat{\gamma}$$

being a constant that we still need to work out!

LEMMA 4.9. *For the Gaussian,* $\gamma(x) = \exp(-\frac{x^2}{2})$,

$$\hat{\gamma}(\xi) = \sqrt{2\pi}\gamma(\xi). \tag{4.167}$$

PROOF. Certainly, $\hat{\gamma} \in \mathcal{S}(\mathbb{R})$ and from the identities for derivatives above

$$\frac{d\hat{\gamma}}{d\xi} = -i\mathcal{F}(x\gamma),\ \xi\hat{\gamma} = \mathcal{F}(-i\frac{d\gamma}{dx}). \tag{4.168}$$

Thus, $\hat{\gamma}$ satisfies the same differential equation as γ :

$$\frac{d\hat{\gamma}}{d\xi} + \xi\hat{\gamma} = -i\mathcal{F}(\frac{d\gamma}{dx} + x\gamma) = 0.$$

This equation we can solve and so we conclude that $\hat{\gamma} = c'\gamma$ where c' is also a constant that we need to compute. To do this observe that

$$c' = \hat{\gamma}(0) = \int \gamma = \sqrt{2\pi} \tag{4.169}$$

which gives (4.167). The computation of the integral in (4.169) is a standard clever argument which you probably know. Namely take the square and work in polar coordinates in two variables:

$$\begin{aligned}(\int \gamma)^2 &= \int_0^\infty \int_0^\infty e^{-(x^2+y^2)} dxdy \\ &= \int_0^{2\pi} \int_0^\infty e^{-r^2/2} rdrd\theta = 2\pi\big[-e^{-r^2/2}\big]_0^\infty = 2\pi.\end{aligned} \tag{4.170}$$

□

So, finally we need to get from (4.161) to the inversion formula. Changing variable in the Fourier transform we can see that for any $y \in \mathbb{R}$, setting $u_y(x) = u(x+y)$, which is in $\mathcal{S}(\mathbb{R})$ if $u \in \mathcal{S}(\mathbb{R})$,

$$\mathcal{F}(u_y) = \int e^{-ix\xi} u_y(x)dx = \int e^{-i(s-y)\xi} u(s)ds = e^{iy\xi}\hat{u}. \tag{4.171}$$

Now, plugging u_y into (4.161) we see that

$$\int \hat{u}_y(0) = 2\pi u_y(0) = 2\pi u(y) = \int e^{iy\xi}\hat{u}(\xi)d\xi \Longrightarrow u(y) = \mathcal{G}u, \tag{4.172}$$

the Fourier inversion formula. So we have proved the Theorem. □

8. Convolution

There is a discussion of convolution later in the notes, I have inserted a new (but not very different) treatment here to cover the density of $\mathcal{S}(\mathbb{R})$ in $L^2(\mathbb{R})$ needed in the next section.

Consider two continuous functions of compact support $u,\ v \in \mathcal{C}_{\text{c}}(\mathbb{R})$. Their convolution is

$$u * v(x) = \int u(x-y)v(y)dy = \int u(y)v(x-y)dy. \tag{4.173}$$

The first integral is the definition, clearly it is a well-defined Riemann integral since the integrand is continuous as a function of y and vanishes whenever $v(y)$ vanishes – so has compact support. In fact if both u and v vanish outside $[-R, R]$ then $u * v = 0$ outside $[-2R, 2R]$.

From standard properties of the Riemann integral (or Dominated convergence if you prefer!) it follows easily that $u*v$ is continuous. What we need to understand is what happens if (at least) one of u or v is smoother. In fact we will want to take a very smooth function, so I pause here to point out

Lemma 4.10. *There exists a ('bump') function $\psi : \mathbb{R} \longrightarrow \mathbb{R}$ which is infinitely differentiable, i.e. has continuous derivatives of all orders, vanishes outside $[-1, 1]$, is strictly positive on $(-1, 1)$ and has integral* 1.

Proof. We start with an explicit function,

$$\phi(x) = \begin{cases} e^{-1/x} & x > 0 \\ 0 & x \le 0. \end{cases} \tag{4.174}$$

The exponential function grows faster than any polynomial at $+\infty$, since

$$\exp(x) > \frac{x^k}{k!} \text{ in } x > 0 \ \forall \ k. \tag{4.175}$$

This can be seen directly from the Taylor series which converges on the whole line (indeed in the whole complex plane)

$$\exp(x) = \sum_{k \ge 0} \frac{x^k}{k!}.$$

From (4.175) we deduce that

$$\lim_{x \downarrow 0} \frac{e^{-1/x}}{x^k} = \lim_{R \to \infty} \frac{R^k}{e^R} = 0 \ \forall \ k \tag{4.176}$$

where we substitute $R = 1/x$ and use the properties of exp. In particular ϕ in (4.174) is continuous across the origin, and so everywhere. We can compute the derivatives in $x > 0$ and these are of the form

$$\frac{d^l \phi}{dx^l} = \frac{p_l(x)}{x^{2l}} e^{-1/x}, \ x > 0, \ p_l \text{ a polynomial.} \tag{4.177}$$

As usual, do this by induction since it is true for $l = 0$ and differetiating the formula for a given l one finds

$$\frac{d^{l+1}\phi}{dx^{l+1}} = \left(\frac{p_l(x)}{x^{2l+2}} - 2l \frac{p_l(x)}{x^{2l+1}} + \frac{p_l'(x)}{x^{2l}} \right) e^{-1/x} \tag{4.178}$$

where the coefficient function is of the desired form p_{l+1}/x^{2l+2}.

Once we know (4.177) then we see from (4.176) that all these functions are continuous down to 0 where they vanish. From this it follows that ϕ in (4.174) is infinitely differentiable. For ϕ itself we can use the Fundamental Theorem of Calculus to write

$$\phi(x) = \int_\epsilon^x U(t)dt + \phi(\epsilon), \ x > \epsilon > 0. \tag{4.179}$$

Here U is the derivative in $x > 0$. Taking the limit as $\epsilon \downarrow 0$ both sides converge, and then we see that

$$\phi(x) = \int_0^x U(t)dt.$$

From this it follows that ϕ is continuously differentiable across 0 and it derivative is U, the continuous extension of the derivative from $x > 0$. The same argument applies to succssive derivatives, so indeed ϕ is infinitely differentiable.

From ϕ we can construct a function closer to the desired bump function. Namely

$$\Phi(x) = \phi(x+1)\phi(1-x).$$

The first factor vanishes when $x \le -1$ and is otherwise positive while the second vanishes when $x \ge 1$ but is otherwise positive, so the product is infinitely differentiable on $\mathbb{R}$ and positive on $(-1, 1)$ but otherwise 0. Then we can normalize the integral to 1 by taking

$$\psi(x) = \Phi(x)/\int \Phi. \tag{4.180}$$

□

In particular from Lemma 4.10 we conclude that the space $\mathcal{C}_c^\infty(\mathbb{R})$, of infinitely differentiable functions of compact support, is not empty. Going back to convolution in (4.173) suppose now that v is smooth. Then

$$u \in \mathcal{C}_{\mathrm{c}}(\mathbb{R}),\ v \in \mathcal{C}_c^\infty(\mathbb{R}) \Longrightarrow u * v \in \mathcal{C}_c^\infty(\mathbb{R}). \tag{4.181}$$

As usual this follows from properties of the Riemann integral or by looking directly at the difference quotient

$$\frac{u * v(x+t) - u * v(x)}{t} = \int u(y)\frac{v(x+t-y) - v(x-y)}{t}dt.$$

As $t \to 0$, the difference quotient for v converges uniformly (in y) to the derivative and hence the integral converges and the derivative of the convolution exists,

$$\frac{d}{dx}u * v(x) = u * (\frac{dv}{dx}). \tag{4.182}$$

This result allows immediate iteration, showing that the convolution is smooth and we know that it has compact support

Proposition 4.7. *For any $u \in \mathcal{C}_c(\mathbb{R})$ there exists $u_n \to u$ uniformly on $\mathbb{R}$ where $u_n \in \mathcal{C}_c^\infty(\mathbb{R})$ with supports in a fixed compact set.*

Proof. For each $\epsilon > 0$ consider the rescaled bump function

$$\psi_\epsilon = \epsilon^{-1}\psi(\frac{x}{\epsilon}) \in \mathcal{C}_c^\infty(\mathbb{R}). \tag{4.183}$$

In fact, ψ_ϵ vanishes outside the interval (ϵ, ϵ), is positive within this interval and has integral 1 – which is what the factor of ϵ^{-1} does. Now set

$$u_\epsilon = u * \psi_\epsilon \in \mathcal{C}_c^\infty(\mathbb{R}),\ \epsilon > 0, \tag{4.184}$$

from what we have just seen. From the supports of these functions, u_ϵ vanishes outside $[-R-\epsilon, R+\epsilon]$ if u vanishes outside $[-R, R]$. So only the convergence remains. To get this we use the fact that the integral of ψ_ϵ is equal to 1 to write

$$u_\epsilon(x) - u(x) = \int (u(x-y)\psi_\epsilon(y) - u(x)\psi_\epsilon(y))dy. \tag{4.185}$$

Estimating the integral using the positivity of the bump function

$$|u_\epsilon(x) - u(x)| = \int_{-\epsilon}^{\epsilon} |u(x-y) - u(x)|\psi_\epsilon(y)dy. \tag{4.186}$$

By the uniformity of a continuous function on a compact set, given $\delta > 0$ there exists $\epsilon > 0$ such that

$$\sup_{[-\epsilon,\epsilon]} |u(x-y) - y(x)| < \delta \ \forall \ x \in \mathbb{R}.$$

So the uniform convergence follows:-

$$\sup |u_\epsilon(x) - u(x)| \le \delta \int \phi_\epsilon = \delta \tag{4.187}$$

Pass to a sequence $\epsilon_n \to 0$ if you wish, □

Corollary 4.1. *The spaces $\mathcal{C}_c^\infty(\mathbb{R})$ and $\mathcal{S}(\mathbb{R})$ are dense in $L^2(\mathbb{R})$.*

Uniform convegence of continuous functions with support in a fixed subset is stronger than L^2 convergence, so the result follows from the Proposition above for $\mathcal{C}_c^\infty(\mathbb{R}) \subset \mathcal{S}(\mathbb{R})$.

9. Plancherel and Parseval

But which is which?

We proceed to show that $\mathcal{F}$ and $\mathcal{G}$, defined in (4.137) and (4.145), both extend to isomorphisms of $L^2(\mathbb{R})$ which are inverses of each other. The main step is to show that

$$\int u(x)\hat{v}(x)dx = \int \hat{u}(\xi)v(\xi)d\xi,\ u,\ v \in \mathcal{S}(\mathbb{R}). \tag{4.188}$$

Since the integrals are rapidly convergent at infinity we may substitute the definite of the Fourier transform into (4.188), write the result out as a double integral and change the order of integration

$$\begin{aligned}\int u(x)\hat{v}(x)dx &= \int u(x) \int e^{-ix\xi} v(\xi)d\xi dx \\ &= \int v(\xi) \int e^{-ix\xi} u(x)dxd\xi = \int \hat{u}(\xi)v(\xi)d\xi.\end{aligned} \tag{4.189}$$

Now, if $w \in \mathcal{S}(\mathbb{R})$ we may replace $v(\xi)$ by $\overline{\hat{w}}(\xi)$, since it is another element of $\mathcal{S}(\mathbb{R})$. By the Fourier Inversion formula,

$$w(x) = (2\pi)^{-1} \int e^{-ix\xi}\hat{w}(\xi) \Longrightarrow \overline{w(x)} = (2\pi)^{-1} \int e^{ix\xi}\overline{\hat{w}(\xi)} = (2\pi)^{-1}\hat{v}. \tag{4.190}$$

Substituting these into (4.188) gives Parseval's formula

$$\int u\overline{w} = \frac{1}{2\pi} \int \hat{u}\overline{\hat{w}},\ u,\ w \in \mathcal{S}(\mathbb{R}). \tag{4.191}$$

Proposition 4.8. *The Fourier transform $\mathcal{F}$ extends from $\mathcal{S}(\mathbb{R})$ to an isomorphism on $L^2(\mathbb{R})$ with $\frac{1}{\sqrt{2\pi}}\mathcal{F}$ an isometric isomorphism with adjoint, and inverse, $\sqrt{2\pi}\mathcal{G}$.*

Proof. Setting $u = w$ in (4.191) shows that

$$\|\mathcal{F}(u)\|_{L^2} = \sqrt{2\pi}\|u\|_{L^2} \tag{4.192}$$

for all $u \in \mathcal{S}(\mathbb{R})$. The density of $\mathcal{S}(\mathbb{R})$, established above, then implies that $\mathcal{F}$ extends by continuity to the whole of $L^2(\mathbb{R})$ as indicated. □

This isomorphism of $L^2(\mathbb{R})$ has many implications. For instance, we would like to define the Sobolev space $H^1(\mathbb{R})$ by the conditions that $u \in L^2(\mathbb{R})$ and $\frac{du}{dx} \in L^2(\mathbb{R})$ but to do this we would need to make sense of the derivative. However, we can 'guess' that if it exists, the Fourier transform of du/dx should be $i\xi\hat{u}(\xi)$. For a function in L^2, such as $\hat{u}$ given that $u \in L^2$, we *do* know what it means to require $\xi\hat{u}(\xi) \in L^2(\mathbb{R})$. We can then define the Sobolev spaces of any positive, even non-integral, order by

$$H^r(\mathbb{R}) = \{u \in L^2(\mathbb{R}); |\xi|^r\hat{u} \in L^2(\mathbb{R})\}. \tag{4.193}$$

Of course it would take us some time to investigate the properties of these spaces!

10. Weak and strong derivatives

In approaching the issue of the completeness of the eigenbasis for harmonic oscillator more directly, rather than by the kernel method discussed above, we run into the issue of weak and strong solutions of differential equations. Suppose that $u \in L^2(\mathbb{R})$, what does it *mean* to say that $\frac{du}{dx} \in L^2(\mathbb{R})$. For instance, we will want to understand what the 'possible solutions of'

$$\operatorname{An} u = f,\ u,\ f \in L^2(\mathbb{R}),\ \operatorname{An} = \frac{d}{dx} + x \tag{4.194}$$

are. Of course, if we assume that u is continuously differentiable then we know what this means, but we need to consider the possibilities of giving a meaning to (4.194) under more general conditions – without assuming too much regularity on u (or any at all).

Notice that there is a difference between the two terms in $\operatorname{An} u = \frac{du}{dx} + xu$. If $u \in L^2(\mathbb{R})$ we can assign a meaning to the second term, xu, since we know that $xu \in L^2_{\text{loc}}(\mathbb{R})$. This is not a normed space, but it is a perfectly good vector space, in which $L^2(\mathbb{R})$ 'sits' – if you want to be pedantic it naturally injects into it. The point however, is that we *do know* what the statement $xu \in L^2(\mathbb{R})$ means, given that $u \in L^2(\mathbb{R})$, it means that there exists $v \in L^2(\mathbb{R})$ so that $xu = v$ in $L^2_{\text{loc}}(\mathbb{R})$ (or $L^2_{\text{loc}}(\mathbb{R})$). The derivative can actually be handled in a similar fashion using the Fourier transform but I will not do that here.

Rather, consider the following three 'L^2-based notions' of derivative.

Definition 4.5.

(1) We say that $u \in L^2(\mathbb{R})$ has a *Sobolev derivative* if there exists a sequence $\phi_n \in \mathcal{C}^1_{\text{c}}(\mathbb{R})$ such that $\phi_n \to u$ in $L^2(\mathbb{R})$ and $\phi_n' \to v$ in $L^2(\mathbb{R})$, $\phi_n' = \frac{d\phi_n}{dx}$ in the usual sense of course.

(2) We say that $u \in L^2(\mathbb{R})$ has a *strong derivative* (in the L^2 sense) if the limit

$$\lim_{0\neq s\to 0} \frac{u(x+s) - u(x)}{s} = \tilde{v} \text{ exists in } L^2(\mathbb{R}). \tag{4.195}$$

(3) Thirdly, we say that $u \in L^2(\mathbb{R})$ has a *weak derivative* in L^2 if there exists $w \in L^2(\mathbb{R})$ such that

$$(u, -\frac{df}{dx})_{L^2} = (w, f)_{L^2} \ \forall \ f \in \mathcal{C}^1_{\mathrm{c}}(\mathbb{R}). \tag{4.196}$$

In all cases, we will see that it is justified to write $v = \tilde{v} = w = \frac{du}{dx}$ because these defintions turn out to be equivalent. Of course *if* $u \in \mathcal{C}^1_{\mathrm{c}}(\mathbb{R})$ then u is differentiable in each sense and the derivative is always du/dx – note that the integration by parts used to prove (4.196) is justified in that case. In fact we are most interested in the first and third of these definitions, the first two are both called 'strong derivatives.'

It is easy to see that the existence of a Sobolev derivative implies that this is also a weak derivative. Indeed, since ϕ_n, the approximating sequence whose existence is the definition of the Sobolev derivative, is in $\mathcal{C}^1_{\mathrm{c}}(\mathbb{R})$ the integration by parts implicit in (4.196) is valid and so for all $f \in \mathcal{C}^1_{\mathrm{c}}(\mathbb{R})$,

$$(\phi_n, -\frac{df}{dx})_{L^2} = (\phi_n', f)_{L^2}. \tag{4.197}$$

Since $\phi_n \to u$ in L^2 and $\phi_n' \to v$ in L^2 both sides of (4.197) converge to give the identity (4.196).

Before proceeding to the rest of the equivalence of these definitions we need to do some preparation. First let us investigate a little the consequence of the existence of a Sobolev derivative.

Lemma 4.11. *If $u \in L^2(\mathbb{R})$ has a Sobolev derivative then $u \in \mathcal{C}(\mathbb{R})$ and there exists a uniquely defined element $w \in L^2(\mathbb{R})$ such that*

$$u(x) - u(y) = \int_y^x w(s)ds \ \forall \ y \leq x \in \mathbb{R}. \tag{4.198}$$

Proof. Suppose u has a Sobolev derivative, determined by some approximating sequence ϕ_n. Consider a general element $\psi \in \mathcal{C}^1_{\mathrm{c}}(\mathbb{R})$. Then $\tilde{\phi}_n = \psi\phi_n$ is a sequence in $\mathcal{C}^1_{\mathrm{c}}(\mathbb{R})$ and $\tilde{\phi}_n \to \psi u$ in L^2. Moreover, by the product rule for standard derivatives

$$\frac{d}{dx}\tilde{\phi}_n = \psi'\phi_n + \psi\phi_n' \to \psi' u + \psi w \text{ in } L^2(\mathbb{R}). \tag{4.199}$$

Thus in fact ψu also has a Sobolev derivative, namely $\phi' u + \psi w$ if w is the Sobolev derivative for u given by ϕ_n – which is to say that the product rule for derivatives holds under these conditions.

Now, the formula (4.198) comes from the Fundamental Theorem of Calculus which in this case really does apply to $\tilde{\phi}_n$ and shows that

$$\psi(x)\phi_n(x) - \psi(y)\phi_n(y) = \int_y^x \frac{d\tilde{\phi}_n}{ds}(s)ds. \tag{4.200}$$

For any given $x = \bar{x}$ we can choose ψ so that $\psi(\bar{x}) = 1$ and then we can take y below the lower limit of the support of ψ so $\psi(y) = 0$. It follows that for this choice of ψ,

$$\phi_n(\bar{x}) = \int_y^{\bar{x}} (\psi'\phi_n(s) + \psi\phi_n'(s))ds. \tag{4.201}$$

Now, we can pass to the limit as $n \to \infty$ and the left side converges for each fixed $\bar{x}$ (with ψ fixed) since the integrand converges in L^2 and hence in L^1 on this compact interval. This actually shows that the limit $\phi_n(\bar{x})$ must exist for each fixed $\bar{x}$. In fact we can always choose ψ to be constant near a particular point and apply this argument to see that

$$\phi_n(x) \to u(x) \text{ locally uniformly on } \mathbb{R}. \tag{4.202}$$

That is, the limit exists locally uniformly, hence represents a continuous function but that continuous function must be equal to the original u almost everywhere (since $\psi\phi_n \to \psi u$ in L^2).

Thus in fact we conclude that '$u \in \mathcal{C}(\mathbb{R})$' (which really means that u *has a representative* which is continuous). Not only that but we get (4.198) from passing to the limit on both sides of

$$u(x) - u(y) = \lim_{n\to\infty}(\phi_n(x) - \phi_n(y)) = \lim_{n\to\infty}\int_y^s (\phi'(s))ds = \int_y^s w(s)ds. \tag{4.203}$$

□

One immediate consequence of this is

$$\text{The Sobolev derivative is unique if it exists.} \tag{4.204}$$

Indeed, if w_1 and w_2 are both Sobolev derivatives then (4.198) holds for both of them, which means that $w_2 - w_1$ has vanishing integral on any finite interval and we know that this implies that $w_2 = w_1$ a.e.

So at least for Sobolev derivatives we are now justified in writing

$$w = \frac{du}{dx} \tag{4.205}$$

since w is unique and behaves like a derivative in the integral sense that (4.198) holds.

Lemma 4.12. *If u has a Sobolev derivative then u has a strong derivative and if u has a strong derivative then this is also a weak derivative.*

Proof. If u has a Sobolev derivative then (3.17) holds. We can use this to write the difference quotient as

$$\frac{u(x+s) - u(x)}{s} - w(x) = \frac{1}{s}\int_0^s (w(x+t) - w(x))dt \tag{4.206}$$

since the integral in the second term can be carried out. Using this formula twice the square of the L^2 norm, which is finite, is

$$\begin{aligned} \|\frac{u(x+s) - u(x)}{s} - w(x)\|^2_{L^2} \\ = \frac{1}{s^2}\int\int_0^s\int_0^s (w(x+t) - w(x)\overline{(w(x+t') - w(x))}dtdt'dx. \end{aligned} \tag{4.207}$$

There is a small issue of manupulating the integrals, but we can always 'back off a little' and replace u by the approximating sequence ϕ_n and then everything is fine – and we only have to check what happens at the end. Now, we can apply the Cauchy-Schwarz inequality as a triple integral. The two factors turn out to be the same so we find that

$$\begin{aligned}(4.208)\quad \|\frac{u(x+s)-u(x)}{s}-w(x)\|^2_{L^2} &\le \frac{1}{s^2}\int\int_0^s\int_0^s|w(x+t)-w(x)|^2dtdt'dx\\ &=\frac{1}{s}\int_0^s\int|w(x+t)-w(x)|^2dxdt\end{aligned}$$

since the integrand does not depend on t'.

Now, something we checked long ago was that L^2 functions are 'continuous in the mean' in the sense that

$$(4.209)\qquad \lim_{0\neq t\to 0}\int|w(x+t)-w(x)|^2dx=0.$$

Applying this to (4.208) and then estimating the t integral shows that

$$(4.210)\qquad \frac{u(x+s)-u(x)}{s}-w(x)\to 0 \text{ in } L^2(\mathbb{R}) \text{ as } s\to 0.$$

By definition this means that u has w as a strong derivative. I leave it up to you to make sure that the manipulation of integrals is okay.

So, now suppose that u has a strong derivative, $\tilde{v}$. Observe that if $f\in\mathcal{C}^1_{\mathrm{c}}(\mathbb{R})$ then the limit defining the derivative

$$(4.211)\qquad \lim_{0\neq s\to 0}\frac{f(x+s)-f(x)}{s}=f'(x)$$

is *uniform.* In fact this follows by writing down the Fundamental Theorem of Calculus, as in (4.198), again using the properties of Riemann integrals. Now, consider

$$\begin{aligned}(4.212)\quad (u(x),\frac{f(x+s)-f(x)}{s})_{L^2} &=\frac{1}{s}\int u(x)\overline{f(x+s)}dx-\frac{1}{s}\int u(x)\overline{f(x)}dx\\ &=(\frac{u(x-s)-u(x)}{s},f(x))_{L^2}\end{aligned}$$

where we just need to change the variable of integration in the first integral from x to $x+s$. However, letting $s\to 0$ the left side converges because of the uniform convergence of the difference quotient and the right side converges because of the assumed strong differentiability and as a result (noting that the parameter on the right is really $-s$)

$$(4.213)\qquad (u,\frac{df}{dx})_{L^2}=-(w,f)_{L^2}\ \forall\ f\in\mathcal{C}^1_{\mathrm{c}}(\mathbb{R})$$

which is weak differentiability with derivative $\tilde{v}$. □

So, at this point we know that Sobolev differentiabilty implies strong differentiability and either of the stong ones implies the weak. So it remains only to show that weak differentiability implies Sobolev differentiability and we can forget about the difference!

Before doing that, note again that a weak derivative, if it exists, is unique – since the difference of two would have to pair to zero in L^2 with all of $\mathcal{C}^1_{\mathrm{c}}(\mathbb{R})$ which is dense. Similarly, if u has a weak derivative then so does ψu for any $\psi\in\mathcal{C}^1_{\mathrm{c}}(\mathbb{R})$

since we can just move ψ around in the integrals and see that

$$\begin{aligned}(\psi u, -\frac{df}{dx}) &= (u, -\overline{\psi}\frac{df}{dx})\\ &= (u, -\frac{d\overline{\psi}f}{dx}) + (u, \overline{\psi'}f)\\ &= (w, \overline{\psi}f + (\psi' u, f) = (\psi w + \psi' u, f)\end{aligned} \tag{4.214}$$

which also proves that the product formula holds for weak derivatives.

So, let us consider $u \in L^2_{\mathrm{c}}(\mathbb{R})$ which does have a weak derivative. To show that it has a Sobolev derivative we need to construct a sequence ϕ_n. We will do this by convolution.

Lemma 4.13. *If $\mu \in \mathcal{C}_c(\mathbb{R})$ then for any $u \in L^2_{\mathrm{c}}(\mathbb{R})$,*

$$\mu * u(x) = \int \mu(x-s)u(s)ds \in \mathcal{C}_c(\mathbb{R}) \tag{4.215}$$

and if $\mu \in \mathcal{C}^1_c(\mathbb{R})$ then

$$\mu * u(x) \in \mathcal{C}^1_c(\mathbb{R}),\ \frac{d\mu * u}{dx} = \mu' * u(x). \tag{4.216}$$

It folows that if μ has more continuous derivatives, then so does $\mu * u$.

Proof. Since u has compact support and is in L^2 it in L^1 so the integral in (4.215) exists for each $x \in \mathbb{R}$ and also vanishes if $|x|$ is large enough, since the integrand vanishes when the supports become separate – for some R, $\mu(x-s)$ is supported in $|s-x| \le R$ and $u(s)$ in $|s| < R$ which are disjoint for $|x| > 2R$. It is also clear that $\mu * u$ is continuous using the estimate (from uniform continuity of μ)

$$|\mu * u(x') - \mu * u(x)| \le \sup |\mu(x-s) - \mu(x'-s)| \|u\|_{L^1}. \tag{4.217}$$

Similarly the difference quotient can be written

$$\frac{\mu * u(x') - \mu * u(x)}{t} = \int \frac{\mu(x'-s) - \mu(x-s)}{s} u(s)ds \tag{4.218}$$

and the uniform convergence of the difference quotient shows that

$$\frac{d\mu * u}{dx} = \mu' * u. \tag{4.219}$$

□

One of the key properties of thes convolution integrals is that we can examine what happens when we 'concentrate' μ. Replace the one μ by the family

$$\mu_\epsilon(x) = \epsilon^{-1}\mu(\frac{x}{\epsilon}),\ \epsilon > 0. \tag{4.220}$$

The singular factor here is introduced so that $\int \mu_\epsilon$ is independent of $\epsilon > 0$,

$$\int \mu_\epsilon = \int \mu\ \forall\ \epsilon > 0. \tag{4.221}$$

Note that since μ has compact support, the support of μ_ϵ is concentrated in $|x| \le \epsilon R$ for some fixed R.

Lemma 4.14. *If $u \in L^2_c(\mathbb{R})$ and $0 \le \mu \in \mathcal{C}^1_c(\mathbb{R})$ then*

$$\lim_{0 \ne \epsilon \to 0} \mu_\epsilon * u = (\int \mu) u \ \textit{in}\ L^2(\mathbb{R}). \tag{4.222}$$

In fact there is no need to assume that u has compact support for this to work.

Proof. First we can change the variable of integration in the definition of the convolution and write it intead as

$$\mu * u(x) = \int \mu(s) u(x-s) ds. \tag{4.223}$$

Now, the rest is similar to one of the arguments above. First write out the difference we want to examine as

$$\mu_\epsilon * u(x) - (\int \mu)(x) = \int_{|s| \le \epsilon R} \mu_\epsilon(s)(u(x-s) - u(x)) ds. \tag{4.224}$$

Write out the square of the absolute value using the formula twice and we find that

$$\begin{aligned} &\int |\mu_\epsilon * u(x) - (\int \mu)(x)|^2 dx \\ &= \int \int_{|s| \le \epsilon R} \int_{|t| \le \epsilon R} \mu_\epsilon(s) \mu_\epsilon(t) (u(x-s) - u(x)) \overline{(u(x-s) - u(x))} ds dt dx \end{aligned} \tag{4.225}$$

Now we can write the integrand as the product of two similar factors, one being

$$\mu_\epsilon(s)^{\frac{1}{2}} \mu_\epsilon(t)^{\frac{1}{2}} (u(x-s) - u(x)) \tag{4.226}$$

using the non-negativity of μ. Applying the Cauchy-Schwarz inequality to this we get two factors, which are again the same after relabelling variables, so

$$\int |\mu_\epsilon * u(x) - (\int \mu)(x)|^2 dx \le \int \int_{|s| \le \epsilon R} \int_{|t| \le \epsilon R} \mu_\epsilon(s) \mu_\epsilon(t) |u(x-s) - u(x)|^2. \tag{4.227}$$

The integral in x can be carried out first, then using continuity-in-the mean bounded by $J(s) \to 0$ as $\epsilon \to 0$ since $|s| < \epsilon R$. This leaves

$$\begin{aligned} &\int |\mu_\epsilon * u(x) - (\int \mu) u(x)|^2 dx \\ &\le \sup_{|s| \le \epsilon R} J(s) \int_{|s| \le \epsilon R} \int_{|t| \le \epsilon R} \mu_\epsilon(s) \mu_\epsilon(t) = (\int \psi)^2 Y \sup_{|s| \le \epsilon R} \to 0. \end{aligned} \tag{4.228}$$

□

After all this preliminary work we are in a position to to prove the remaining part of 'weak=strong'.

Lemma 4.15. *If $u \in L^2(\mathbb{R})$ has w as a weak L^2-derivative then w is also the Sobolev derivative of u.*

Proof. Let's assume first that u has compact support, so we can use the discussion above. Then set $\phi_n = \mu_{1/n} * u$ where $\mu \in \mathcal{C}^1_{\mathrm{c}}(\mathbb{R})$ is chosen to be non-negative and have integral $\int \mu = 0$; μ_ϵ is defined in (4.220). Now from Lemma 4.14 it follows that $\phi_n \to u$ in $L^2(\mathbb{R})$. Also, from Lemma 4.13, $\phi_n \in \mathcal{C}^1_{\mathrm{c}}(\mathbb{R})$ has derivative given by (4.216). This formula can be written as a pairing in L^2 :

$$(\mu_{1/n})' * u(x) = (u(s), -\frac{d\mu_{1/n}(x-s)}{ds})^2_L = (w(s), \frac{d\mu_{1/n}(x-s)}{ds})_{L^2} \tag{4.229}$$

using the definition of the weak derivative of u. It therefore follows from Lemma 4.14 applied again that

$$\phi'_n = \mu_{/m1/n} * w \to w \text{ in } L^2(\mathbb{R}). \tag{4.230}$$

Thus indeed, ϕ_n is an approximating sequence showing that w is the Sobolev derivative of u.

In the general case that $u \in L^2(\mathbb{R})$ has a weak derivative but is not necessarily compactly supported, consider a function $\gamma \in \mathcal{C}^1_{\text{c}}(\mathbb{R})$ with $\gamma(0) = 1$ and consider the sequence $v_m = \gamma(x)u(x)$ in $L^2(\mathbb{R})$ each element of which has compact support. Moreover, $\gamma(x/m) \to 1$ for each x so by Lebesgue dominated convergence, $v_m \to u$ in $L^2(\mathbb{R})$ as $m \to \infty$. As shown above, v_m has as weak derivative

$$\frac{d\gamma(x/m)}{dx}u + \gamma(x/m)w = \frac{1}{m}\gamma'(x/m)u + \gamma(x/m)w \to w$$

as $m \to \infty$ by the same argument applied to the second term and the fact that the first converges to 0 in $L^2(\mathbb{R})$. Now, use the approximating sequence $\mu_{1/n} * v_m$ discussed converges to v_m with its derivative converging to the weak derivative of v_m. Taking $n = N(m)$ sufficiently large for each m ensures that $\phi_m = \mu_{1/N(m)} * v_m$ converges to u and its sequence of derivatives converges to w in L^2. Thus the weak derivative is again a Sobolev derivative. □

Finally then we see that the three definitions are equivalent and we will freely denote the Sobolev/strong/weak derivative as du/dx or u'.

11. Sobolev spaces

Now there are lots of applications of the Fourier transform which we do not have the time to get into. However, let me just indicate the definitions of Sobolev spaces and Schwartz space and how they are related to the Fourier transform.

First Sobolev spaces. We now see that $\mathcal{F}$ maps $L^2(\mathbb{R})$ isomorphically onto $L^2(\mathbb{R})$ and we can see from (**??**) for instance that it 'turns differentiations by x into multiplication by ξ'. Of course we do not know how to differentiate L^2 functions so we have some problems making sense of this. One way, the usual mathematicians trick, is to turn what we want into a definition.

Definition 4.6. The Sobolev spaces of order s, for any $s \in (0, \infty)$, are defined as subspaces of $L^2(\mathbb{R})$:

$$H^s(\mathbb{R}) = \{u \in L^2(\mathbb{R}); (1 + |\xi|^2)^s \hat{u} \in L^2(\mathbb{R})\}. \tag{4.231}$$

It is natural to identify $H^0(\mathbb{R}) = L^2(\mathbb{R})$.

These Sobolev spaces, for each positive order s, are Hilbert spaces with the inner product and norm

$$(u, v)_{H^s} = \int (1 + |\xi|^2)^s \hat{u}(\xi)\overline{\hat{v}(\xi)},\ \|u\|_s = \|(1 + |\xi|^2)^{\frac{s}{2}}\hat{u}\|_{L^2}. \tag{4.232}$$

That they are pre-Hilbert spaces is clear enough. Completeness is also easy, given that we know the completeness of $L^2(\mathbb{R})$. Namely, if u_n is Cauchy in $H^s(\mathbb{R})$ then it follows from the fact that

$$\|v\|_{L^2} \le C\|v\|_s \ \forall \ v \in H^s(\mathbb{R}) \tag{4.233}$$

that u_n is Cauchy in L^2 and also that $(1+|\xi|^2)^{\frac{s}{2}}\hat{u}_n(\xi)$ is Cauchy in L^2. Both therefore converge to a limit u in L^2 and the continuity of the Fourier transform shows that $u \in H^s(\mathbb{R})$ and that $u_n \to u$ in H^s.

These spaces are examples of what is discussed above where we have a dense inclusion of one Hilbert space in another, $H^s(\mathbb{R}) \longrightarrow L^2(\mathbb{R})$. In this case the inclusion in *not* compact but it does give rise to a bounded self-adjoint operator on $L^2(\mathbb{R})$, $E_s : L^2(\mathbb{R}) \longrightarrow H^s(\mathbb{R}) \subset L^2(\mathbb{R})$ such that

$$(u,v)_{L^2} = (E_s u, E_s v)_{H^s}. \tag{4.234}$$

It is reasonable to denote this as $E_s = (1+|D_x|^2)^{-\frac{s}{2}}$ since

$$u \in L^2(\mathbb{R}^n) \Longrightarrow \widehat{E_s u}(\xi) = (1+|\xi|^2)^{-\frac{s}{2}}\hat{u}(\xi). \tag{4.235}$$

It is a form of 'fractional integration' which turns any $u \in L^2(\mathbb{R})$ into $E_s u \in H^s(\mathbb{R})$.

Having defined these spaces, which get smaller as s increases it can be shown for instance that if $n \ge s$ is an integer then the set of n times continuously differentiable functions on $\mathbb{R}$ which vanish outside a compact set are dense in H^s. This allows us to justify, by continuity, the following statement:-

Proposition 4.9. *The bounded linear map*

$$\frac{d}{dx} : H^s(\mathbb{R}) \longrightarrow H^{s-1}(\mathbb{R}),\ s \ge 1,\ v(x) = \frac{du}{dx} \Longleftrightarrow \hat{v}(\xi) = i\xi\hat{u}(\xi) \tag{4.236}$$

is consistent with differentiation on n times continuously differentiable functions of compact support, for any integer $n \ge s$.

In fact one can even get a 'strong form' of differentiation. The condition that $u \in H^1(\mathbb{R})$, that $u \in L^2$ 'has one derivative in L^2' is actually equivalent, for $u \in L^2(\mathbb{R})$ to the existence of the limit

$$\lim_{t \to 0} \frac{u(x+t)u(x)}{t} = v,\ \text{in } L^2(\mathbb{R}) \tag{4.237}$$

and then $\hat{v} = i\xi\hat{u}$. Another way of looking at this is

$$\begin{gathered} u \in H^1(\mathbb{R}) \Longrightarrow u : \mathbb{R} \longrightarrow \mathbb{C} \text{ is continuous and} \\ u(x) - u(y) = \int_y^x v(t)dt,\ v \in L^2. \end{gathered} \tag{4.238}$$

If such a $v \in L^2(\mathbb{R})$ exists then it is unique – since the difference of two such functions would have to have integral zero over any finite interval and we know (from one of the exercises) that this implies that the function vansishes a.e.

One of the more important results about Sobolev spaces – of which there are many – is the relationship between these 'L^2 derivatives' and 'true derivatives'.

Theorem 4.7 (Sobolev embedding). *If n is an integer and $s > n + \frac{1}{2}$ then*

$$H^s(\mathbb{R}) \subset \mathcal{C}^n_\infty(\mathbb{R}) \tag{4.239}$$

consists of n times continuosly differentiable functions with bounded derivatives to order n (which also vanish at infinity).

This is actually not so hard to prove, there are some hints in the exercises below.

These are not the only sort of spaces with 'more regularity' one can define and use. For instance one can try to treat x and ξ more symmetrically and define smaller spaces than the H^s above by setting

$$H^s_{\text{iso}}(\mathbb{R}) = \{u \in L^2(\mathbb{R}); (1+|\xi|^2)^{\frac{s}{2}}\hat{u} \in L^2(\mathbb{R}),\ (1+|x|^2)^{\frac{s}{2}}u \in L^2(\mathbb{R})\}. \tag{4.240}$$

The 'obvious' inner product with respect to which these 'isotropic' Sobolev spaces $H^s_{\text{iso}}(\mathbb{R})$ are indeed Hilbert spaces is

$$(u,v)_{s,\text{iso}} = \int_{\mathbb{R}} u\overline{v} + \int_{\mathbb{R}} |x|^{2s}u\overline{v} + \int_{\mathbb{R}} |\xi|^{2s}\hat{u}\overline{\hat{v}} \tag{4.241}$$

which makes them look rather symmetric between u and $\hat{u}$ and indeed

$$\mathcal{F} : H^s_{\text{iso}}(\mathbb{R}) \longrightarrow H^s_{\text{iso}}(\mathbb{R}) \text{ is an isomorphism } \forall\ s \geq 0. \tag{4.242}$$

At this point, by dint of a little, only moderately hard, work, it is possible to show that the harmonic oscillator extends by continuity to an isomorphism

$$H : H^{s+2}_{\text{iso}}(\mathbb{R}) \longrightarrow H^s_{\text{iso}}(\mathbb{R})\ \forall\ s \geq 2. \tag{4.243}$$

Finally in this general vein, I wanted to point out that Hilbert, and even Banach, spaces are not the end of the road! One very important space in relation to a direct treatment of the Fourier transform, is the Schwartz space. The definition is reasonably simple. Namely we denote Schwartz space by $\mathcal{S}(\mathbb{R})$ and say

$$\begin{gathered} u \in \mathcal{S}(\mathbb{R}) \Longleftrightarrow u : \mathbb{R} \longrightarrow \mathbb{C} \\ \text{is continuously differentiable of all orders and for every } n, \\ \|u\|_n = \sum_{k+p\leq n} \sup_{x\in\mathbb{R}} (1+|x|)^k |\frac{d^p u}{dx^p}| < \infty. \end{gathered} \tag{4.244}$$

All these inequalities just mean that all the derivatives of u are 'rapidly decreasing at ∞' in the sense that they stay bounded when multiplied by any polynomial.

So in fact we know already that $\mathcal{S}(\mathbb{R})$ is not empty since the elements of the Hermite basis, $e_j \in \mathcal{S}(\mathbb{R})$ for all j. In fact it follows immediately from this that

$$\mathcal{S}(\mathbb{R}) \longrightarrow L^2(\mathbb{R}) \text{ is dense.} \tag{4.245}$$

If you want to try your hand at something a little challenging, see if you can check that

$$\mathcal{S}(\mathbb{R}) = \bigcap_{s>0} H^s_{\text{iso}}(\mathbb{R}) \tag{4.246}$$

which uses the Sobolev embedding theorem above.

As you can see from the definition in (4.244), $\mathcal{S}(\mathbb{R})$ is not likely to be a Banach space. Each of the $\|\cdot\|_n$ is a norm. However, $\mathcal{S}(\mathbb{R})$ is pretty clearly not going to be complete with respect to any one of these. However it is complete with respect to all, countably many, norms. What does this mean? In fact $\mathcal{S}(\mathbb{R})$ is a *metric space* with the metric

$$d(u,v) = \sum_n 2^{-n} \frac{\|u-v\|_n}{1+\|u-v\|_n} \tag{4.247}$$

as you can check. So the claim is that $\mathcal{S}(\mathbb{R})$ *is* complete as a metric space – such a thing is called a Fréchet space.

What has this got to do with the Fourier transform? The point is that

(4.248)
$$\mathcal{F} : \mathcal{S}(\mathbb{R}) \longrightarrow \mathcal{S}(\mathbb{R}) \text{ is an isomorphism and } \mathcal{F}(\frac{du}{dx}) = i\xi\mathcal{F}(u),\ \mathcal{F}(xu) = -i\frac{d\mathcal{F}(u)}{d\xi}$$

where this now makes sense. The dual space of $\mathcal{S}(\mathbb{R})$ – the space of continuous linear functionals on it, is the space, denoted $\mathcal{S}'(\mathbb{R})$, of tempered distributions on $\mathbb{R}$.

12. Schwartz distributions

We do not have time in this course to really discuss distributions. Still, it is a good idea for you to know what they are and why they are useful. Of course to really appreciate their utility you need to read a bit more than I have here. First think a little about the Schwartz space $\mathcal{S}(\mathbb{R})$ introduced above. The metric in (4.247) might seem rather mysterious but it has the important property that *each* of the norms $\|\cdot\|_n$ defines a continuous function $\mathcal{S}(\mathbb{R}) \longrightarrow \mathbb{R}$ with respect to this metric topology. In fact a linear map

(4.249)
$$T : \mathcal{S}(\mathbb{R}) \longrightarrow \mathbb{C} \text{ linear is continuous iff } \exists\ N, C \text{ s.t. } \|T\phi\| \le C\|\phi\|_N\ \forall\ \phi \in \mathcal{S}(\mathbb{R}).$$

So, the continuous linear functionals on $\mathcal{S}(\mathbb{R})$ are just those which are continous with respect to one of the norms.

These functionals are exactly the space of *tempered distributions*

$$\mathcal{S}'(\mathbb{R}) = \{T : \mathcal{S}(\mathbb{R}) \longrightarrow \mathbb{C} \text{ linear and continuous}\}. \tag{4.250}$$

The relationship to functions is that each $f \in L^2(\mathbb{R})$ (or more generally such that $(1+|x|)^{-N} \in L^1(\mathbb{R})$ for some N) defines an element of $\mathcal{S}'(\mathbb{R})$ by integration:

$$T_f : \mathcal{S}(\mathbb{R}) \ni \phi \longmapsto \int f(x)\phi(x) \in \mathbb{C} \Longrightarrow T_f \in \mathcal{S}'(\mathbb{R}). \tag{4.251}$$

Indeed, this amounts to showing that $\|\phi\|_{L^2}$ is a continuous norm on $\mathcal{S}(\mathbb{R})$ (so it must be bounded by a multiple of one of the $\|\phi\|_N$, which one?)

It is relatively straightforward to show that $L^2(\mathbb{R}) \ni f \longmapsto T_f \in \mathcal{S}'(\mathbb{R})$ is injective – nothing is 'lost'. So after a little more experience with distributions one comes to identify f and T_f. Notice that this is just an extension of the behaviour of $L^2(\mathbb{R})$ where (because we can drop the complex conjugate in the inner product) by Riesz' Theorem we can identify (linearly) $L^2(\mathbb{R})$ with it dual, exactly by the map $f \longmapsto T_f$.

Other elements of $\mathcal{S}'(\mathbb{R})$ include the delta 'function' at the origin and even its 'derivatives' for each j

$$\delta^j : \mathcal{S}(\mathbb{R}) \ni \phi \longmapsto (-1)^j\frac{d^j\phi}{dx^j}(0) \in \mathbb{C}. \tag{4.252}$$

In fact one of the main points about the space $\mathcal{S}'(\mathbb{R})$ is that differentiation and multiplication by polynomials is well defined

$$\frac{d}{dx} : \mathcal{S}'(\mathbb{R}) \longrightarrow \mathcal{S}'(\mathbb{R}),\ \times x : \mathcal{S}'(\mathbb{R}) \longrightarrow \mathcal{S}'(\mathbb{R}) \tag{4.253}$$

in a way that is consistent with their actions under the identification $\mathcal{S}(\mathbb{R}) : \phi \longmapsto T_\phi \in \mathcal{S}'(\mathbb{R})$. This property is enjoyed by other spaces of distributions but the fundamental fact that the Fourier transform extends to

$$\mathcal{F} : \mathcal{S}'(\mathbb{R}) \longrightarrow \mathcal{S}'(\mathbb{R}) \text{ as an isomorphism} \tag{4.254}$$

is more characteristic of $\mathcal{S}'(\mathbb{R})$.

13. Poisson summation formula

We have talked both about Fourier series and the Fourier transform. It is natural to ask: What is the connection between these? The Fourier series of a function in $L^2(0, 2\pi)$ we thought of as given by the Fourier-Bessel series with respect to the orthonormal basis

$$\frac{\exp(ikx)}{\sqrt{2\pi}},\ k \in \mathbb{Z}. \tag{4.255}$$

The interval here is just a particular choice – if the upper limit is changed to T then the corresponding orthonormal basis of $L^2(0, T)$ is

$$\frac{\exp(i2\pi kx/T)}{\sqrt{T}},\ k \in \mathbb{Z}. \tag{4.256}$$

Sometimes the Fourier transform is thought of as the limit of the Fourier series expansion when $T \to \infty$. This is actually not such a nice limit, so unless you *have* (or *want*) to do this I recommend against it!

A more fundamental relationship between the two comes about as follows. We can think of $L^2(0, 2\pi)$ as 'really' being the 2π-periodic functions restricted to this interval. Since the values at the end-points don't matter this does give a bijection – between 2π-periodic, locally square-integrable functions on the line and $L^2(0, 2\pi)$. On the other hand we can also think of the periodic functions as being defined on the circle, $|z| = 1$ in $\mathbb{C}$ or identified with the values of $\theta \in \mathbb{R}$ modulo repeats:

$$\mathbb{T} = \mathbb{R}/2\pi\mathbb{Z} \ni \theta \longmapsto e^{i\theta} \in \mathbb{C}. \tag{4.257}$$

Let us denote by $\mathcal{C}^\infty(\mathbb{T})$ the space of infinitely differentiable, 2π-periodic functions on the line; this is also the space of smooth functions on the circle, thought of as a manifold.

How can one construct such functions. There are plenty of examples, for instance the $\exp(ikx)$. Another way to construct examples is to sum over translations:-

Lemma 4.16. *The map*

$$A : \mathcal{S}(\mathbb{R}) \ni f \longrightarrow \sum_{k\in\mathbb{Z}} f(\cdot - 2\pi k) \in \mathcal{C}^\infty(\mathbb{T}) \tag{4.258}$$

is surjective.

Proof. That the series in (4.258) converges uniformly on $[0, 2\pi]$ (or any bounded interval) is easy enought to see, since the rapid decay of elements of $\mathcal{S}(\mathbb{R})$ shows that

$$|f(x)| \le C(1+|x|)^{-2},\ x \in \mathbb{R} \Longrightarrow |f(x-2\pi k)| \le C'(1+|k|)^{-2},\ x \in [0, 2\pi] \tag{4.259}$$

since if $k > 2$ $|x-2\pi k| \ge k$ if $x \in [0, 2\pi]$. Clearly (4.259) implies uniform convergence of the series. Since the derivatives of f are also in $\mathcal{S}(\mathbb{R})$ the series obtained by term-by-term differentiation also converges uniformly and by standard arguments the limit Ag is therefore infinitely differentiable, with

$$\frac{d^j Af}{dx^j} = A\frac{d^j f}{dx^j}. \tag{4.260}$$

This shows that the map A, clearly linear, is well-defined. Now, how to see that it is surjective? Let's first prove a special case. Indeed, look for a function $\psi \in \mathcal{C}_c^\infty(\mathbb{R}) \subset \mathcal{S}(\mathbb{R})$ which is non-negative and such that $A\psi = 1$. We know that we can find $\phi \in \mathcal{C}_c^\infty(\mathbb{R})$, $\phi \geq 0$ with $\phi > 0$ on $[0, 2\pi]$. Then consider $A\phi \in \mathcal{C}^\infty(\mathbb{T})$. It must be stricly positive, $A\phi \geq \epsilon > 0$ since it is larger that ϕ. So consider instead the function

$$\psi = \frac{\phi}{A\phi} \in \mathcal{C}_c^\infty(\mathbb{R}) \tag{4.261}$$

where we think of $A\phi$ as 2π-periodic on $\mathbb{R}$. In fact using this periodicity we see that

$$A\psi \equiv 1. \tag{4.262}$$

So this shows that the constant function 1 is in the range of A. In general, just take $g \in \mathcal{C}^\infty(\mathbb{T})$, thought of as 2π-periodic on the line, and it follows that

$$f = Bg = \psi g \in \mathcal{C}_c^\infty(\mathbb{R}) \subset \mathcal{S}(\mathbb{R}) \text{ satsifies } Af = g. \tag{4.263}$$

Indeed,

$$Ag = \sum_k \psi(x - 2\pi k) g(x - 2\pi k) = g(x) \sum_k \psi(x - 2\pi k) = g \tag{4.264}$$

using the periodicity of g. In fact B is a right inverse for A,

$$AB = \text{Id on } \mathcal{C}^\infty(\mathbb{T}). \tag{4.265}$$

□

Question 2. What is the null space of A?

Since $f \in \mathcal{S}(\mathbb{R})$ and $Af \in \mathcal{C}^\infty(\mathbb{T}) \subset L^2(0, 2\pi)$ with our identifications above, the question arises as to the relationship between the Fourier transform of f and the Fourier series of Af.

Proposition 4.10 (Poisson summation formula). *If $g = Af$, $g \in \mathcal{C}^\infty(\mathbb{T})$ and $f \in \mathcal{S}(\mathbb{R})$ then the Fourier coefficients of g are*

$$c_k = \int_{[0,2\pi]} g e^{-ikx} = \hat{f}(k). \tag{4.266}$$

Proof. Just substitute in the formula for g and, using uniform convergenc, check that the sum of the integrals gives after translation the Fourier transform of f. □

If we think of recovering g from its Fourier series,

$$g(x) = \frac{1}{2\pi} \sum_{k \in \mathbb{Z}} c_k e^{ikx} = \frac{1}{2\pi} \sum_{k \in \mathbb{Z}} \hat{f}(k) e^{ikx} \tag{4.267}$$

then in terms of the Fourier transform on $\mathcal{S}'(\mathbb{R})$ alluded to above, this takes the rather elegant form

$$\frac{1}{2\pi} \mathcal{F}\left(\sum_{k \in \mathbb{Z}} \delta(\cdot - k)\right)(x) = \frac{1}{2\pi} \sum_{k \in \mathbb{Z}} e^{ikx} = \sum_{k \in \mathbb{Z}} \delta(x - 2\pi k). \tag{4.268}$$

The sums of translated Dirac deltas and oscillating exponentials all make sense in $\mathcal{S}'(\mathbb{R})$.

APPENDIX A

Problems for Chapter 1

1. For §1

Problem 1.1. In case you are a bit shaky on it, go through the basic theory of finite-dimensional vector spaces. Define a vector space V to be finite-dimensional if there is an integer N such that any N elements of V are linearly dependent – if $v_i \in V$ for $i = 1, \dots N$, then there exist $a_i \in \mathbb{K}$, not *all* zero, such that

$$\sum_{i=1}^{N} a_i v_i = 0 \text{ in } V. \tag{A.1}$$

If N is the smallest such integer define dimension of V to be $\dim V = N-1$ and show that a finite dimensional vector space always has a basis, $e_i \in V$, $i = 1, \dots, \dim V$ such that any element of V can be written uniquely as a linear combination

$$v = \sum_{i=1}^{\dim V} b_i e_i,\ b_i \in \mathbb{K}. \tag{A.2}$$

Problem 1.2. Show from first principles that if V is a vector space (over $\mathbb{R}$ or $\mathbb{C}$) then for any set X the space of all maps

$$\mathcal{F}(X; V) = \{u : X \longrightarrow V\} \tag{A.3}$$

is a vector space over the same field, with 'pointwise operations' (which you should write down carefully).

Problem 1.3. Show that if V is a vector space and $S \subset V$ is a subset which is closed under addition and scalar multiplication:

$$v_1, v_2 \in S,\ \lambda \in \mathbb{K} \Longrightarrow v_1 + v_2 \in S \text{ and } \lambda v_1 \in S \tag{A.4}$$

then S is a vector space as well with operations 'inherited from V' (and called, of course, a subspace of V).

Problem 1.4. Recall that a map between vector spaces $L : V \longrightarrow W$ is linear if $L(v_1 + v_2) = Lv_1 + Lv_2$ and $L\lambda v = \lambda Lv$ for all elements v_1, v_2, $v \in V$ and all scalars λ. Show that given two finite dimensional vector spaces V and W over the same field

1. If $\dim V \leq \dim W$ then there is an injective linear map $L : V \longrightarrow W$.
2. If $\dim V \geq W$ then there is a surjective linear map $L : V \longrightarrow W$.
3. if $\dim V = \dim W$ then there is a linear isomorphism $L : V \longrightarrow W$, i.e. an injective and surjective linear map.

Problem 1.5. If $S \subset V$ is a linear subspace of a vector space show that the relation on V

$$v_1 \sim v_2 \Longleftrightarrow v_1 - v_2 \in S \tag{A.5}$$

is an equivalence relation and that the set of equivalence classes

$$[v] = \{w \in V; w - v \in S\},$$

denoted usually V/S, is a linear space in a natural way and that the projection map $\pi : V \longrightarrow V/S$, taking each v to $[v]$ is linear.

Problem 1.6. Show that any two norms on a finite dimensional vector space are equivalent.

Problem 1.7. Show that two norms on a vector space are equivalent if and only if the topologies induced are the same – the sets open with respect to the distance from one are open with respect to the distance coming from the other.

Problem 1.8. Write out a proof for each p with $1 \le p < \infty$ that

$$l^p = \{a : \mathbb{N} \longrightarrow \mathbb{C}; \sum_{j=1}^{\infty} |a_j|^p < \infty,\ a_j = a(j)\}$$

is a normed space with the norm

$$\|a\|_p = \left(\sum_{j=1}^{\infty} |a_j|^p\right)^{\frac{1}{p}}.$$

This means writing out the proof that this is a linear space and that the three conditions required of a norm hold.

Problem 1.9. Prove directly that each l^p as defined in Problem 1.8 is complete, i.e. it is a Banach space.

Problem 1.10. The space l^∞ consists of the bounded sequences

$$l^\infty = \{a : \mathbb{N} \longrightarrow \mathbb{C}; \sup_n |a_n| < \infty\},\ \|a\|_\infty = \sup_n |a_n|. \tag{A.6}$$

Show that this is a *non-separable* Banach space.

Problem 1.11. Another closely related space consists of the sequences converging to 0 :

$$c_0 = \{a : \mathbb{N} \longrightarrow \mathbb{C}; \lim_{n\to\infty} a_n = 0\},\ \|a\|_\infty = \sup_n |a_n|. \tag{A.7}$$

Check that this is a separable Banach space and that it is a closed subspace of l^∞ (perhaps do it in the opposite order).

Problem 1.12. Consider the 'unit sphere' in l^p. This is the set of vectors of length 1 :

$$S = \{a \in l^p; \|a\|_p = 1\}.$$

(1) Show that S is closed.
(2) Recall the sequential (so not the open covering definition) characterization of compactness of a set in a metric space (e.g. by checking in Rudin's book).
(3) Show that S is not compact by considering the sequence in l^p with kth element the sequence which is all zeros except for a 1 in the kth slot. Note that the main problem is not to get yourself confused about sequences of sequences!

Problem 1.13. Show that the norm on any normed space is continuous.

Problem 1.14. Finish the proof of the completeness of the space B constructed in the second proof of Theorem 1.1.

APPENDIX B

Problems for Chapter 4

1. Hill's equation

As an extended exercise I suggest you follow the ideas of §4.4 but now for 'Hill's equation' which is the same problem as (4.73) but with *periodic boundary conditions*:-

$$-\frac{d^2u}{dx^2}+Vu=f \text{ on } (0,2\pi),\ u(2\pi)=u(0),\ \frac{du}{dx}(2\pi)=\frac{du}{dx}(0). \tag{B.1}$$

There are several ways to do this, but you cannot proceed in *precisely* the same way since for $V=0$ the constants are solutions of (B.1) – so even if the system has a solution (which for some f it does not) this solution is not unique.

One way to proceed is to start from $V=1$ say and solve the problem explicitly. However the formulæ are not as simple as for the Dirichlet case.

So instead I will outline an approach starting from the solution of the Dirichlet problem. This is allows you to see some important concepts – for instance the Maximum Principle. You should proceed to prove this sequence of claims!

(1) If $V\ge 0$ (always real-valued in $\mathcal{C}([0,2\pi])$) then we know that the Dirichlet problem, (4.73), has a unique solution given in (4.111):

$$u=S_Vf,\ S_V=A(\operatorname{Id}+AVA)^{-1}A. \tag{B.2}$$

Recall that the eigenfunctions of this operator are twice continuously differentiable eigenfunctions for the Dirichlet problem with eigenvalues $T_k=\lambda_k^{-1}$ where the λ_k are the eigenvalues of S_V.

(2) Prove the maximal principle in this case, that if $V>0$, $f\ge 0$ then $u=S_Vf\ge 0$. Hint:- If this were not true then there would be an interior minimum at which $u(p)<0$ but at this point $-\frac{d^2u}{dx^2}(p)\le 0$ and $V(p)u(p)<0$ which contradicts (4.73) since $f(p)\ge 0$.

(3) Now, suppose u is a 'classical' (twice continuously differentiable) solution to (B.1) (with $V>0$). Then set $u_0=u(0)=u(2\pi)$ and $u'=u-u_0$ and observe that

$$-\frac{d^2u'}{dx^2}+Vu'=f-u_0V\Longrightarrow u'=S_Vf-u_0S_VV. \tag{B.3}$$

(4) Using the assumption that $V>0$ show that

$$\frac{d}{dx}S_VV(0)>0,\ \frac{d}{dx}S_VV(2\pi)<0. \tag{B.4}$$

Hint:- From the equation for S_VV observe that $\frac{d^2}{dx^2}S_VV(0)<0$ so if $\frac{d}{dx}S_VV(0)\le 0$ then $S_VV(x)<0$ for small $x>0$ violating the Maximum Principle and similarly at 2π.

(5) Conclude from (B.3) that for $V > 0$ there is a unique solution to (B.1) which is of the form

$$u = T_V f = S_V f + u_0 - u_0 S_V, \quad au_0 = \frac{d}{dx} T_V V(2\pi) - \frac{d}{dx} T_V V(0), \quad a = \frac{d}{dx} S_V V(2\pi) - \frac{d}{dx} S_V V(0) > 0. \tag{B.5}$$

(6) Show that T_V is an injective compact self-adjoint operator and that its eigenfunctions are twice continuously differentiable eigenfunctions for the periodic boundary problem. Hint:- Boundedness follows from the properties of S_V, as does compactness with a bit more effort. For self-adjointness integrate the equation by parts.

(7) Conclude the analogue of Theorem 4.5 for periodic boundary conditions, i.e. Hill's equation.

2. Mehler's formula and completeness

Starting from the ground state for the harmonic oscillator

$$P = -\frac{d^2}{dx^2} + x^2,\ Hu_0 = u_0,\ u_0 = e^{-x^2/2} \tag{B.6}$$

and using the creation and annihilation operators

$$\mathrm{An} = \frac{d}{dx} + x,\ \mathrm{Cr} = -\frac{d}{dx} + x,\ \mathrm{An}\,\mathrm{Cr} - \mathrm{Cr}\,\mathrm{An} = 2\,\mathrm{Id},\ H = \mathrm{Cr}\,\mathrm{An} + \mathrm{Id} \tag{B.7}$$

we have constructed the higher eigenfunctions:

$$u_j = \mathrm{Cr}^j u_0 = p_j(x) u_0(c),\ p(x) = 2^j x^j + \text{l.o.ts},\ Hu_j = (2j+1)u_j \tag{B.8}$$

and shown that these are orthogonal, $u_j \perp u_k$, $j \neq k$, and so when normalized give an orthonormal system in $L^2(\mathbb{R})$:

$$e_j = \frac{u_j}{2^{j/2}(j!)^{\frac{1}{2}}\pi^{\frac{1}{4}}}. \tag{B.9}$$

Now, what we want to see, is that these e_j form an orthonormal basis of $L^2(\mathbb{R})$, meaning they are complete as an orthonormal sequence. There are various proofs of this, but the only 'simple' ones I know involve the Fourier inversion formula and I want to use the completeness to *prove* the Fourier inversion formula, so that will not do. Instead I want to use a version of Mehler's formula.

To show the completeness of the e_j's it is enough to find a compact self-adjoint operator with these as eigenfunctions and no null space. It is the last part which is tricky. The first part is easy. Remembering that all the e_j are real, we can find an operator with the e_j;s as eigenfunctions with corresponding eigenvalues $\lambda_j > 0$ (say) by just defining

$$Au(x) = \sum_{j=0}^{\infty} \lambda_j (u, e_j) e_j(x) = \sum_{j=0}^{\infty} \lambda_j e_j(x) \int e_j(y) u(y). \tag{B.10}$$

For this to be a compact operator we need $\lambda_j \to 0$ as $j \to \infty$, although for boundedness we just need the λ_j to be bounded. So, the problem with this is to show that A has no null space – which of course is just the completeness of the e'_j since (assuming all the λ_j are positive)

$$Au = 0 \Longleftrightarrow u \perp e_j \ \forall \ j. \tag{B.11}$$

Nevertheless, this is essentially what we will do. The idea is to write A as an *integral operator* and then work with that. I will take the $\lambda_j = w^j$ where $w \in (0, 1)$. The point is that we can find an explicit formula for

$$A_w(x, y) = \sum_{j=0}^{\infty} w^j e_j(x)e_j(y) = A(w, x, y). \tag{B.12}$$

To find $A(w, x, y)$ we will need to compute the Fourier transforms of the e_j. Recall that

$$\begin{gathered}\mathcal{F} : L^1(\mathbb{R}) \longrightarrow \mathcal{C}_\infty(\mathbb{R}),\ \mathcal{F}(u) = \hat{u},\\ \hat{u}(\xi) = \int e^{-ix\xi}u(x),\ \sup|\hat{u}| \le \|u\|_{L^1}.\end{gathered} \tag{B.13}$$

Lemma B.1. *The Fourier transform of u_0 is*

$$(\mathcal{F}u_0)(\xi) = \sqrt{2\pi}u_0(\xi). \tag{B.14}$$

Proof. Since u_0 is both continuous and Lebesgue integrable, the Fourier transform is the limit of a Riemann integral

$$\hat{u}_0(\xi) = \lim_{R\to\infty} \int_{-R}^{R} e^{i\xi x}u_0(x). \tag{B.15}$$

Now, for the Riemann integral we can differentiate under the integral sign with respect to the parameter ξ – since the integrand is continuously differentiable – and see that

$$\begin{aligned}\frac{d}{d\xi}\hat{u}_0(\xi) &= \lim_{R\to\infty}\int_{-R}^{R} ixe^{i\xi x}u_0(x)\\ &= \lim_{R\to\infty} i\int_{-R}^{R} e^{i\xi x}(-\frac{d}{dx}u_0(x)\\ &= \lim_{R\to\infty} -i\int_{-R}^{R}\frac{d}{dx}\left(e^{i\xi x}u_0(x)\right) - \xi\lim_{R\to\infty}\int_{-R}^{R} e^{i\xi x}u_0(x)\\ &= -\xi\hat{u}_0(\xi).\end{aligned} \tag{B.16}$$

Here I have used the fact that $\operatorname{An} u_0 = 0$ and the fact that the boundary terms in the integration by parts tend to zero rapidly with R. So this means that $\hat{u}_0$ is annihilated by An :

$$(\frac{d}{d\xi} + \xi)\hat{u}_0(\xi) = 0. \tag{B.17}$$

Thus, it follows that $\hat{u}_0(\xi) = c\exp(-\xi^2/2)$ since these are the only functions in annihilated by An . The constant is easy to compute, since

$$\hat{u}_0(0) = \int e^{-x^2/2}dx = \sqrt{2\pi} \tag{B.18}$$

proving (B.14). □

We can use this formula, of if you prefer the argument to prove it, to show that

$$v = e^{-x^2/4} \Longrightarrow \hat{v} = \sqrt{\pi} e^{-\xi^2}. \tag{B.19}$$

Changing the names of the variables this just says

$$e^{-x^2} = \frac{1}{2\sqrt{\pi}} \int_{\mathbb{R}} e^{ixs - s^2/4} ds. \tag{B.20}$$

The definition of the u_j's can be rewritten

$$u_j(x) = (-\frac{d}{dx} + x)^j e^{-x^2/2} = e^{x^2/2}(-\frac{d}{dx})^j e^{-x^2} \tag{B.21}$$

as is easy to see inductively – the point being that $e^{x^2/2}$ is an integrating factor for the creation operator. Plugging this into (B.20) and carrying out the derivatives – which is legitimate since the integral is so strongly convergent – gives

$$u_j(x) = \frac{e^{x^2/2}}{2\sqrt{\pi}} \int_{\mathbb{R}} (-is)^j e^{ixs - s^2/4} ds. \tag{B.22}$$

Now we can use this formula twice on the sum on the left in (B.12) and insert the normalizations in (B.9) to find that

$$\sum_{j=0}^{\infty} w^j e_j(x) e_j(y) = \sum_{j=0}^{\infty} \frac{e^{x^2/2+y^2/2}}{4\pi^{3/2}} \int_{\mathbb{R}^2} \frac{(-1)^j w^j s^j t^j}{2^j j!} e^{isx+ity-s^2/4-t^2/4} ds dt. \tag{B.23}$$

The crucial thing here is that we can sum the series to get an exponential, this allows us to finally conclude:

Lemma B.2. *The identity* (B.12) *holds with*

$$A(w, x, y) = \frac{1}{\sqrt{\pi}\sqrt{1-w^2}} \exp\left(-\frac{1-w}{4(1+w)}(x+y)^2 - \frac{1+w}{4(1-w)}(x-y)^2\right) \tag{B.24}$$

Proof. Summing the series in (B.23) we find that

$$A(w, x, y) = \frac{e^{x^2/2+y^2/2}}{4\pi^{3/2}} \int_{\mathbb{R}^2} \exp(-\frac{1}{2}wst + isx + ity - \frac{1}{4}s^2 - \frac{1}{4}t^2) ds dt. \tag{B.25}$$

Now, we can use the same formula as before for the Fourier transform of u_0 to evaluate these integrals explicitly. One way to do this is to make a change of variables by setting

$$s = (S+T)/\sqrt{2},\ t = (S-T)/\sqrt{2} \Longrightarrow ds dt = dS dT, \tag{B.26}$$

$$-\frac{1}{2}wst + isx + ity - \frac{1}{4}s^2 - \frac{1}{4}t^2 = iS\frac{x+y}{\sqrt{2}} - \frac{1}{4}(1+w)S^2 + iT\frac{x-y}{\sqrt{2}} - \frac{1}{4}(1-w)T^2.$$

Note that the integrals in (B.25) are 'improper' (but rapidly convergent) Riemann integrals, so there is no problem with the change of variable formula. The formula for the Fourier transform of $\exp(-x^2)$ can be used to conclude that

$$\begin{aligned} \int_{\mathbb{R}} \exp(iS\frac{x+y}{\sqrt{2}} - \frac{1}{4}(1+w)S^2) dS &= \frac{2\sqrt{\pi}}{\sqrt{(1+w)}} \exp(-\frac{(x+y)^2}{2(1+w)}) \\ \int_{\mathbb{R}} \exp(iT\frac{x-y}{\sqrt{2}} - \frac{1}{4}(1-w)T^2) dT &= \frac{2\sqrt{\pi}}{\sqrt{(1-w)}} \exp(-\frac{(x-y)^2}{2(1-w)}). \end{aligned} \tag{B.27}$$

Inserting these formulæ back into (B.25) gives

$$A(w,x,y)=\frac{1}{\sqrt{\pi}\sqrt{1-w^2}}\exp\left(-\frac{(x+y)^2}{2(1+w)}-\frac{(x-y)^2}{2(1-w)}+\frac{x^2}{2}+\frac{y^2}{2}\right) \tag{B.28}$$

which after a little adjustment gives (B.24). □

Now, this explicit representation of A_w as an integral operator allows us to show

Proposition B.1. *For all real-valued $f\in L^2(\mathbb{R})$,*

$$\sum_{j=1}^{\infty}|(u,e_j)|^2=\|f\|^2_{L^2}. \tag{B.29}$$

Proof. By definition of A_w

$$\sum_{j=1}^{\infty}|(u,e_j)|^2=\lim_{w\uparrow 1}(f,A_wf) \tag{B.30}$$

so (B.29) reduces to

$$\lim_{w\uparrow 1}(f,A_wf)=\|f\|^2_{L^2}. \tag{B.31}$$

To prove (B.31) we will make our work on the integral operators rather simpler by assuming first that $f\in\mathcal{C}(\mathbb{R})$ is continuous and vanishes outside some bounded interval, $f(x)=0$ in $|x|>R$. Then we can write out the L^2 inner product as a double integral, which is a genuine (iterated) Riemann integral:

$$(f,A_wf)=\int\int A(w,x,y)f(x)f(y)dydx. \tag{B.32}$$

Here I have used the fact that f and A are real-valued.

Look at the formula for A in (B.24). The first thing to notice is the factor $(1-w^2)^{-\frac{1}{2}}$ which blows up as $w\to 1$. On the other hand, the argument of the exponential has two terms, the first tends to 0 as $w\to 1$ and the becomes very large and negative, at least when $x-y\neq 0$. Given the signs, we see that

$$\begin{gathered}\text{if } \epsilon>0,\ X=\{(x,y);|x|\le R,|y|\le R,|x-y|\ge\epsilon\}\text{ then}\\ \sup_X|A(w,x,y)|\to 0\text{ as }w\to 1.\end{gathered} \tag{B.33}$$

So, the part of the integral in (B.32) over $|x-y|\ge\epsilon$ tends to zero as $w\to 1$.

So, look at the other part, where $|x-y|\le\epsilon$. By the (uniform) continuity of f, given $\delta>0$ there exits $\epsilon>0$ such that

$$|x-y|\le\epsilon\Longrightarrow|f(x)-f(y)|\le\delta. \tag{B.34}$$

Now we can divide (B.32) up into three pieces:-

$$\begin{aligned}(f,A_wf)=&\int_{S\cap\{|x-y|\ge\epsilon\}}A(w,x,y)f(x)f(y)dydx\\&+\int_{S\cap\{|x-y|\le\epsilon\}}A(w,x,y)(f(x)-f(y))f(y)dydx\\&\qquad+\int_{S\cap\{|x-y|\le\epsilon\}}A(w,x,y)f(y)^2dydx\end{aligned} \tag{B.35}$$

where $S = [-R, R]^2$.

Look now at the third integral in (B.35) since it is the important one. We can change variable of integration from x to $t = \sqrt{\frac{1+w}{1-w}}(x-y)$. Since $|x-y| \le \epsilon$, the new t variable runs over $|t| \le \epsilon\sqrt{\frac{1+w}{1-w}}$ and then the integral becomes

$$\int_{S\cap\{|t|\le\epsilon\sqrt{\frac{1+w}{1-w}}\}} A(w, y+t\sqrt{\frac{1-w}{1+w}}, y)f(y)^2dydt, \text{ where}$$

$$\begin{aligned}(\text{B.36})\qquad A(w, y+&t\sqrt{\frac{1-w}{1+w}}, y)\\ &= \frac{1}{\sqrt{\pi}(1+w)}\exp\left(-\frac{1-w}{4(1+w)}(2y+t\sqrt{1-w})^2\right)\exp\left(-\frac{t^2}{4}\right).\end{aligned}$$

Here y is bounded; the first exponential factor tends to 1 and the t domain extends to $(-\infty, \infty)$ as $w \to 1$, so it follows that for any $\epsilon > 0$ the third term in (B.35) tends to

$$(\text{B.37})\qquad \|f\|^2_{L^2} \text{ as } w \to 1 \text{ since } \int e^{-t^2/4} = 2\sqrt{\pi}.$$

Noting that $A \ge 0$ the same argument shows that the second term is bounded by a constant multiple of δ. Now, we have already shown that the first term in (B.35) tends to zero as $\epsilon \to 0$, so this proves (B.31) – given some $\gamma > 0$ first choose $\epsilon > 0$ so small that the first two terms are each less than $\frac{1}{2}\gamma$ and then let $w \uparrow 0$ to see that the lim sup and lim inf as $w \uparrow 0$ must lie in the range $[\|f\|^2 - \gamma, \|f\|^2 + \gamma]$. Since this is true for all $\gamma > 0$ the limit exists and (B.29) follows under the assumption that f is continuous and vanishes outside some interval $[-R, R]$.

This actually suffices to prove the completeness of the Hermite basis. In any case, the general case follows by continuity since such continuous functions vanishing outside compact sets are dense in $L^2(\mathbb{R})$ and both sides of (B.29) are continuous in $f \in L^2(\mathbb{R})$. □

Now, (B.31) certainly implies that the e_j form an orthonormal basis, which is what we wanted to show – but hard work! It is done here in part to remind you of how we did the Fourier series computation of the same sort and to suggest that you might like to compare the two arguments.

3. Friedrichs' extension

Next I will discuss an abstract Hilbert space set-up which covers the treatment of the Dirichlet problem above and several other applications to differential equations and indeed to other problems. I am attributing this method to Friedrichs and he certainly had a hand in it.

Instead of just one Hilbert space we will consider two at the same time. First is a 'background' space, H, a separable infinite-dimensional Hilbert space which you can think of as being something like $L^2(I)$ for some interval I. The inner product on this I will denote $(\cdot,\cdot)_H$ or maybe sometimes leave off the 'H' since this is the basic space. Let me denote a second, separable infinite-dimensional, Hilbert space as D, which maybe stands for 'domain' of some operator. So D comes with its own inner product $(\cdot,\cdot)_D$ where I will try to remember not to leave off the subscript.

The relationship between these two Hilbert spaces is given by a linear map

$$i : D \longrightarrow H. \tag{B.38}$$

This is denoted 'i' because it is supposed to be an 'inclusion'. In particular I will always require that

$$i \text{ is injective.} \tag{B.39}$$

Since we will not want to have parts of H which are inaccessible, I will also assume that

$$i \text{ has dense range } i(D) \subset H. \tag{B.40}$$

In fact because of these two conditions it is quite safe to identify D with $i(D)$ and think of each element of D as really being an element of H. The subspace '$i(D) = D$' will not be closed, which is what we are used to thinking about (since it is dense) but rather has its own inner product $(\cdot,\cdot)_D$. Naturally we will also suppose that i is continuous and to avoid too many constants showing up I will suppose that i has norm at most 1 so that

$$\|i(u)\|_H \le \|u\|_D. \tag{B.41}$$

If you are comfortable identifying $i(D)$ with D this just means that the 'D-norm' on D is *bigger* than the H norm restricted to D. A bit later I will assume one more thing about i.

What can we do with this setup? Well, consider an arbitrary element $f \in H$. Then consider the linear map

$$T_f : D \ni u \longrightarrow (i(u), f)_H \in \mathbb{C}. \tag{B.42}$$

where I have put in the identification i but will leave it out from now on, so just write $T_f(u) = (u, f)_H$. This is in fact a continuous linear functional on D since by Cauchy-Schwarz and then (B.41),

$$|T_f(u)| = |(u, f)_H| \le \|u\|_H \|f\|_H \le \|f\|_H \|u\|_D. \tag{B.43}$$

So, by the Riesz' representation – so using the assumed completeness of D (with respect to the D-norm of course) there exists a unique element $v \in D$ such that

$$(u, f)_H = (u, v)_D \ \forall \ u \in D. \tag{B.44}$$

Thus, v only depends on f and always exists, so this defines a map

$$B : H \longrightarrow D,\ Bf = v \text{ iff } (f, u)_H = (v, u)_D \ \forall \ u \in D \tag{B.45}$$

where I have taken complex conjugates of both sides of (B.44).

Lemma B.3. *The map B is a continuous linear map $H \longrightarrow D$ and restricted to D is self-adjoint:*

$$(Bw, u)_D = (w, Bu)_D \ \forall \ u, w \in D. \tag{B.46}$$

The assumption that $D \subset H$ is dense implies that $B : H \longrightarrow D$ is injective.

Proof. The linearity follows from the uniqueness and the definition. Thus if $f_i \in H$ and $c_i \in \mathbb{C}$ for $i = 1, 2$ then

$$\begin{aligned} (c_1 f_1 + c_2 f_2, u)_H &= c_1(f_1, u)_H + c_2(f_2, u)_H \\ = c_1(Bf_1, u)_D + c_2(Bf_2, u)_D &= (c_1 Bf_1 + c_2 Bf_2, u)_D \ \forall \ u \in D \end{aligned} \tag{B.47}$$

shows that $B(c_1f_1 + c_2f_2) = c_1Bf_1 + c_2Bf_2$. Moreover from the estimate (B.43),

$$|(Bf, u)_D| \le \|f\|_H\|u\|_D \tag{B.48}$$

and setting $u = Bf$ it follows that $\|Bf\|_D \le \|f\|_H$ which is the desired continuity.

To see the self-adjointness suppose that $u,\ w \in D$, and hence of course since we are erasing i, $u,\ w \in H$. Then, from the definitions

$$(Bu, w)_D = (u, w)_H = \overline{(w, u)_H} = \overline{(Bw, u)_D} = (u, Bw)_D \tag{B.49}$$

so B is self-adjoint.

Finally observe that $Bf = 0$ implies that $(Bf, u)_D = 0$ for all $u \in D$ and hence that $(f, u)_H = 0$, but since D is dense, this implies $f = 0$ so B is injective. □

To go a little further we will assume that the inclusion i is *compact.* Explicitly this means

$$u_n \rightharpoonup_D u \Longrightarrow u_n(= i(u_n)) \to_H u \tag{B.50}$$

where the subscript denotes which space the convergence is in. Thus compactness means that a weakly convergent sequence in D is, or is mapped to, a strongly convergent sequence in H.

Lemma B.4. *Under the assumptions* (B.38), (B.39), (B.40), (B.41) *and* (B.50) *on the inclusion of one Hilbert space into another, the operator B in* (B.45) *is compact as a self-adjoint operator on D and has only positive eigenvalues.*

Proof. Suppose $u_n \rightharpoonup u$ is weakly convergent in D. Then, by assumption it is strongly convergent in H. But B is continuous as a map from H to D so $Bu_n \to Bu$ in D and it follows that B is compact as an operator on D.

So, we know that D has an orthonormal basis of eigenvectors of B. None of the eigenvalues λ_j can be zero since B is injective. Moreover, from the definition if $Bu_j = \lambda_j u_j$ then

$$\|u_j\|_H^2 = (u_j, u_j)_H = (Bu_j, u_j)_D = \lambda_j\|u_j\|_D^2 \tag{B.51}$$

showing that $\lambda_j > 0$. □

Now, in view of this we can define another compact operator on D by

$$Au_j = \lambda_j^{\frac{1}{2}} u_j \tag{B.52}$$

taking the positive square-roots. So of course $A^2 = B$. In fact $A : H \longrightarrow D$ is also a bounded operator.

Lemma B.5. *If u_j is an orthonormal basis of D of eigenvectors of B then $f_j = \lambda^{-\frac{1}{2}} u_j$ is an orthonormal basis of H and $A : D \longrightarrow D$ extends by continuity to an isometric isomorphism $A : H \longrightarrow D$.*

Proof. The identity (B.51) extends to pairs of eigenvectors

$$(u_j, u_k)_H = (Bu_j, u_k)_D = \lambda_j\delta_{jk} \tag{B.53}$$

which shows that the f_j form an orthonormal sequence in H. The span is dense in D (in the H norm) and hence is dense in H so this set is complete. Thus A maps an orthonormal basis of H to an orthonormal basis of D, so it is an isometric isomorphism. □

If you think about this a bit you will see that this is an abstract version of the treatment of the 'trivial' Dirichlet problem above, except that I did not describe the Hilbert space D concretely in that case.

There are various ways this can be extended. One thing to note is that the failure of injectivity, i.e. the loss of (B.39) is not so crucial. If i is not injective, then its null space is a closed subspace and we can take its orthocomplement in place of D. The result is the same except that the operator D is only defined on this orthocomplement.

An additional thing to observe is that the completeness of D, although used crucially above in the application of Riesz' Representation theorem, is not really such a big issue either

PROPOSITION B.2. *Suppose that $\tilde{D}$ is a* pre-Hilbert *space with inner product $(\cdot,\cdot)_D$ and $i : \tilde{A} \longrightarrow H$ is a linear map into a Hilbert space. If this map is injective, has dense range and satisfies* (B.41) *in the sense that*

$$\|i(u)\|_H \le \|u\|_D \;\forall\; u \in \tilde{D} \tag{B.54}$$

then it extends by continuity to a map of the completion, D, of $\tilde{D}$, satisfying (B.39), (B.40) *and* (B.41) *and if bounded sets in $\tilde{D}$ are mapped by i into precompact sets in H then* (B.50) *also holds.*

PROOF. We know that a completion exists, $\tilde{D} \subset D$, with inner product restricting to the given one and every element of D is then the limit of a Cauchy sequence in $\tilde{D}$. So we denote without ambiguity the inner product on D again as $(\cdot,\cdot)_D$. Since i is continuous with respect to the norm on D (and on H of course) it extends by continuity to the closure of $\tilde{D}$, namely D as $i(u) = \lim_n i(u_n)$ if u_n is Cauchy in $\tilde{D}$ and hence converges in D; this uses the completeness of H since $i(u_n)$ is Cauchy in H. The value of $i(u)$ does not depend on the choice of approximating sequence, since if $v_n \to 0$, $i(v_n) \to 0$ by continuity. So, it follows that $i : D \longrightarrow H$ exists, is linear and continuous and its norm is no larger than before so (B.38) holds. □

The map extended map may not be injective, i.e. it might happen that $i(u_n) \to 0$ even though $u_n \to u \ne 0$.

The general discussion of the set up of Lemmas B.4 and B.5 can be continued further. Namely, having defined the operators B and A we can define a new positive-definite Hermitian form on H by

$$(u, v)_E = (Au, Av)_H,\; u,\; v \in H \tag{B.55}$$

with the same relationship as between $(\cdot,\cdot)_H$ and $(\cdot,\cdot)_D$. Now, it follows directly that

$$\|u\|_H \le \|u\|_E \tag{B.56}$$

so if we let E be the completion of H with respect to this new norm, then $i : H \longrightarrow E$ is an injection with dense range and A extends to an isometric isomorphism $A : E \longrightarrow H$. Then if u_j is an orthonormal basis of H of eigenfunctions of A with eigenvalues $\tau_j > 0$ it follows that $u_j \in D$ and that the $\tau_j^{-1}u_j$ form an orthonormal basis for D while the $\tau_j u_j$ form an orthonormal basis for E.

Lemma B.6. *With E defined as above as the completion of H with respect to the inner product* (B.55)*, B extends by continuity to an isomoetric isomorphism $B : E \longrightarrow D$.*

Proof. Since $B = A^2$ on H this follows from the properties of the eigenbases above. □

The typical way that Friedrichs' extension arises is that we are actually given an explicit 'operator', a linear map $P : \tilde{D} \longrightarrow H$ such that $(u, v)_D = (u, Pv)_H$ satisfies the conditions of Proposition B.2. Then P extends by continuity to an isomorphism $P : D \longrightarrow E$ which is precisely the inverse of B as in Lemma B.6. We shall see examples of this below.

4. Dirichlet problem revisited

So, does the setup of the preceding section work for the Dirichlet problem? We take $H = L^2((0, 2\pi))$. Then, and this really is Friedrichs' extension, we take as a subspace $\tilde{D} \subset H$ the space of functions which are once continuously differentiable and vanish outside a compact subset of $(0, 2\pi)$. This just means that there is some smaller interval, depending on the function, $[\delta, 2\pi - \delta]$, $\delta > 0$, on which we have a continuously differentiable function f with $f(\delta) = f'(\delta) = f(2\pi-\delta) = f'(2\pi-\delta) = 0$ and then we take it to be zero on $(0, \delta)$ and $(2\pi - \delta, 2\pi)$. There are lots of these, let's call the space $\tilde{D}$ as above

$$\begin{aligned}\tilde{D} = \{u \in \mathcal{C}[0, 2\pi]; u \text{ continuously differentiable on } [0, 2\pi],\\ u(x) = 0 \text{ in } [0, \delta] \cup [2\pi - \delta, 2\pi] \text{ for some } \delta > 0\}.\end{aligned} \tag{B.57}$$

Then our first claim is that

$$\tilde{D} \text{ is dense in } L^2(0, 2\pi) \tag{B.58}$$

with respect to the norm on L^2 of course.

What inner product should we take on $\tilde{D}$? Well, we can just integrate formally by parts and set

$$(u, v)_D = \frac{1}{2\pi} \int_{[0,2\pi]} \frac{du}{dx} \overline{\frac{dv}{dx}} dx. \tag{B.59}$$

This is a pre-Hilbert inner product. To check all this note first that $(u, u)_D = 0$ implies $du/dx = 0$ by Riemann integration (since $|du/dx|^2$ is continuous) and since $u(x) = 0$ in $x < \delta$ for some $\delta > 0$ it follows that $u = 0$. Thus $(u, v)_D$ makes $\tilde{D}$ into a pre-Hilbert space, since it is a positive definite sesquilinear form. So, what about the completion? Observe that, the elements of $\tilde{D}$ being continuously differentiable, we can always integrate from $x = 0$ and see that

$$u(x) = \int_0^x \frac{du}{dx} dx \tag{B.60}$$

as $u(0) = 0$. Now, to say that $u_n \in \tilde{D}$ is Cauchy is to say that the continuous functions $v_n = du_n/dx$ are Cauchy in $L^2(0, 2\pi)$. Thus, from the completeness of L^2 we know that $v_n \to v \in L^2(0, 2\pi)$. On the other hand (B.60) applies to each u_n so

$$|u_n(x) - u_m(x)| = |\int_0^x (v_n(s) - v_m(s))ds| \le \sqrt{2\pi} \|v_n - v_m\|_{L^2} \tag{B.61}$$

by applying Cauchy-Schwarz. Thus in fact the sequence u_n is uniformly Cauchy in $C([0,2\pi])$ if u_n is Cauchy in $\tilde{D}$. From the completeness of the Banach space of continuous functions it follows that $u_n \to u$ in $C([0,2\pi])$ so each element of the completion, $\tilde{D}$, 'defines' (read 'is') a continuous function:

$$u_n \to u \in D \Longrightarrow u \in \mathcal{C}([0,2\pi]),\ u(0) = u(2\pi) = 0 \tag{B.62}$$

where the Dirichlet condition follows by continuity from (B.61).

Thus we do indeed get an injection

$$D \ni u \longrightarrow u \in L^2(0,2\pi) \tag{B.63}$$

where the injectivity follows from (B.60) that if $v = \lim du_n/dx$ vanishes in L^2 then $u = 0$.

Now, you can go ahead and check that with these definitions, B and A are the same operators as we constructed in the discussion of the Dirichlet problem.

5. Isotropic space

There are some functions which should be in the domain of P, namely the twice continuously differentiable functions on $\mathbb{R}$ with compact support, those which vanish outside a finite interval. Recall that there are actually a lot of these, they are dense in $L^2(\mathbb{R})$. Following what we did above for the Dirichlet problem set

$$\begin{aligned}\tilde{D} = \{u : \mathbb{R} \longmapsto \mathbb{C}; \exists\ R \text{ s.t. } u = 0 \text{ in } |x| > R,\\ u \text{ is twice continuously differentiable on } \mathbb{R}\}.\end{aligned} \tag{B.64}$$

Now for such functions integration by parts on a large enough interval (depending on the functions) produces no boundary terms so

$$(Pu, v)_{L^2} = \int_{\mathbb{R}} (Pu)\overline{v} = \int_{\mathbb{R}} \left(\frac{du}{dx}\overline{\frac{dv}{dx}} + x^2 u\overline{v} \right) = (u,v)_{\text{iso}} \tag{B.65}$$

is a positive definite hermitian form on $\tilde{D}$. Indeed the vanishing of $\|u\|_S$ implies that $\|xu\|_{L^2} = 0$ and so $u = 0$ since $u \in \tilde{D}$ is continuous. The suffix 'iso' here stands for 'isotropic' and refers to the fact that xu and du/dx are essentially on the same footing here. Thus

$$(u,v)_{\text{iso}} = (\frac{du}{dx}, \frac{dv}{dx})_{L^2} + (xu, xv)_{L^2}. \tag{B.66}$$

This may become a bit clearer later when we get to the Fourier transform.

Definition B.1. Let $H^1_{\text{iso}}(\mathbb{R})$ be the completion of $\tilde{D}$ in (B.64) with respect to the inner product $(\cdot,\cdot)_{\text{iso}}$.

Proposition B.3. *The inclusion map $i : \tilde{D} \longrightarrow L^2(\mathbb{R})$ extends by continuity to $i : H^1_{\text{iso}} \longrightarrow L^2(\mathbb{R})$ which satisfies* (B.38), (B.39), (B.40), (B.41) *and* (B.50) *with $D = H^1_{\text{iso}}$ and $H = L^2(\mathbb{R})$ and the derivative and multiplication maps define an injection*

$$H^1_{\text{iso}} \longrightarrow L^2(\mathbb{R}) \times L^2(\mathbb{R}). \tag{B.67}$$

Proof. Let us start with the last part, (B.67). The map here is supposed to be the continuous extension of the map

$$\tilde{D} \ni u \longmapsto (\frac{du}{dx}, xu) \in L^2(\mathbb{R}) \times L^2(\mathbb{R}) \tag{B.68}$$

where du/dx and xu are both compactly supported continuous functions in this case. By definition of the inner product $(\cdot,\cdot)_{\text{iso}}$ the norm is precisely

$$\|u\|_{\text{iso}}^2 = \|\frac{du}{dx}\|_{L^2}^2 + \|xu\|_{L^2}^2 \tag{B.69}$$

so if u_n is Cauchy in $\tilde{D}$ with respect to $\|\cdot\|_{\text{iso}}$ then the sequences du_n/dx and xu_n are Cauchy in $L^2(\mathbb{R})$. By the completeness of L^2 they converge defining an element in $L^2(\mathbb{R}) \times L^2(\mathbb{R})$ as in (B.67). Moreover the elements so defined only depend on the element of the completion that the Cauchy sequence defines. The resulting map (B.67) is clearly continuous.

Now, we need to show that the inclusion i extends to H^1_{iso} from $\tilde{D}$. This follows from another integration identity. Namely, for $u \in \tilde{D}$ the Fundamental theorem of calculus applied to

$$\frac{d}{dx}(ux\overline{u}) = |u|^2 + \frac{du}{dx}x\overline{u} + ux\overline{\frac{du}{dx}}$$

gives

$$\|u\|_{L^2}^2 \le \int_{\mathbb{R}} |\frac{du}{dx}x\overline{u}| + \int |ux\overline{\frac{du}{dx}}| \le \|u\|_{\text{iso}}^2. \tag{B.70}$$

Thus the inequality (B.41) holds for $u \in \tilde{D}$.

It follows that the inclusion map $i : \tilde{D} \longrightarrow L^2(\mathbb{R})$ extends by continuity to H^1_{iso} since if $u_n \in \tilde{D}$ is Cauchy with respect in H^1_{iso} it is Cauchy in $L^2(\mathbb{R})$. It remains to check that i is injective and compact, since the range is already dense on $\tilde{D}$.

If $u \in H^1_{\text{iso}}$ then to say $i(u) = 0$ (in $L^2(\mathbb{R})$) is to say that for any $u_n \to u$ in H^1_{iso}, with $u_n \in \tilde{D}$, $u_n \to 0$ in $L^2(\mathbb{R})$ and we need to show that this means $u_n \to 0$ in H^1_{iso} to conclude that $u = 0$. To do so we use the map (B.67). If $u_n\tilde{D}$ converges in H^1_{iso} then it follows that the sequence $(\frac{du}{dx}, xu)$ converges in $L^2(\mathbb{R}) \times L^2(\mathbb{R})$. If v is a continuous function of compact support then $(xu_n, v)_{L^2} = (u_n, xv) \to (u, xv)_{L^2}$, for if $u = 0$ it follows that $xu_n \to 0$ as well. Similarly, using integration by parts the limit U of $\frac{du_n}{dx}$ in $L^2(\mathbb{R})$ satisfies

$$(U, v)_{L^2} = \lim_n \int \frac{du_n}{dx}\overline{v} = -\lim_n \int u_n \overline{\frac{dv}{dx}} = -(u, \frac{dv}{dx})_{L^2} = 0 \tag{B.71}$$

if $u = 0$. It therefore follows that $U = 0$ so in fact $u_n \to 0$ in H^1_{iso} and the injectivity of i follows. □

We can see a little more about the metric on H^1_{iso}.

Lemma B.7. *Elements of H^1_{iso} are continuous functions and convergence with respect to $\|\cdot\|_{\text{iso}}$ implies uniform convergence on bounded intervals.*

Proof. For elements of the dense subspace $\tilde{D}$, (twice) continuously differentiable and vanishing outside a bounded interval the Fundamental Theorem of Calculus shows that

$$\begin{gathered} u(x) = e^{x^2/2} \int_{-\infty}^{x} (\frac{d}{dt}(e^{-t^2/2}u) = e^{x^2/2} \int_{-\infty}^{x} (e^{-t^2/2}(-tu + \frac{du}{dt})) \Longrightarrow \\ |u(x)| \le e^{x^2/2}(\int_{-\infty}^{x} e^{-t^2})^{\frac{1}{2}} \|u\|_{\text{iso}} \end{gathered} \tag{B.72}$$

where the estimate comes from the Cauchy-Schwarz applied to the integral. It follows that if $u_n \to u$ with respect to the isotropic norm then the sequence converges uniformly on bounded intervals with

$$\sup_{[-R,R]} |u(x)| \le C(R)\|u\|_{\text{iso}}. \tag{B.73}$$

□

Now, to proceed further we either need to apply some 'regularity theory' or do a computation. I choose to do the latter here, although the former method (outlined below) is much more general. The idea is to show that

Lemma B.8. *The linear map $(P+1) : \mathcal{C}^2_{\text{c}}(\mathbb{R}) \longrightarrow \mathcal{C}_c(\mathbb{R})$ is injective with range dense in $L^2(\mathbb{R})$ and if $f \in L^2(\mathbb{R}) \cap \mathcal{C}(\mathbb{R})$ there is a sequence $u_n \in \mathcal{C}^2_{\text{c}}(\mathbb{R})$ such that $u_n \to u$ in H^1_{iso}, $u_n \to u$ locally uniformly with its first two derivatives and $(P+1)u_n \to f$ in $L^2(\mathbb{R})$ and locally uniformly.*

Proof. Why $P+1$ and not P? The result is actually true for P but not so easy to show directly. The advantage of $P+1$ is that it factorizes

$$(P+1) = \operatorname{An}\operatorname{Cr} \text{ on } \mathcal{C}^2_{\text{c}}(\mathbb{R}).$$

so we proceed to solve the equation $(P+1)u = f$ in two steps.

First, if $f \in \text{c}(\mathbb{R})$ then using the natural integrating factor

$$v(x) = e^{x^2/2}\int_{-\infty}^{x} e^{t^2/2} f(t)dt + ae^{-x^2/2} \text{ satisfies } \operatorname{An} v = f. \tag{B.74}$$

The integral here is not in general finite if f does not vanish in $x < -R$, which by assumption it does. Note that $\operatorname{An} e^{-x^2/2} = 0$. This solution is of the form

$$v \in \mathcal{C}^1(\mathbb{R}),\ v(x) = a_\pm e^{-x^2/2} \text{ in } \pm x > R \tag{B.75}$$

where R depends on f and the constants can be different.

In the second step we need to solve away such terms – in general one cannot. However, we can always choose a in (B.74) so that

$$\int_{\mathbb{R}} e^{-x^2/2} v(x) = 0. \tag{B.76}$$

Now consider

$$u(x) = e^{x^2/2}\int_{-\infty}^{x} e^{-t^2/2} v(t)dt. \tag{B.77}$$

Here the integral does make sense because of the decay in v from (B.75) and $u \in \mathcal{C}^2(\mathbb{R})$. We need to understand how it behaves as $x \to \pm\infty$. From the second part of (B.75),

$$u(x) = a_- \operatorname{erf}_-(x),\ x < -R,\ \operatorname{erf}_-(x) = \int_{(-\infty,x]} e^{x^2/2 - t^2} \tag{B.78}$$

is an incomplete error function. It's derivative is e^{-x^2} but it actually satisfies

$$|x \operatorname{erf}_-(x)| \le Ce^{x^2},\ x < -R. \tag{B.79}$$

In any case it is easy to get an estimate such as Ce^{-bx^2} as $x \to -\infty$ for any $0 < b < 1$ by Cauchy-Schwarz.

As $x \to \infty$ we would generally expect the solution to be rapidly increasing, but precisely because of (B.76). Indeed the vanishing of this integral means we can rewrite (B.77) as an integral from $+\infty$:

$$u(x) = -e^{x^2/2}\int_{[x,\infty)} e^{-t^2/2}v(t)dt \tag{B.80}$$

and then the same estimates analysis yields

$$u(x) = -a_+ \operatorname{erf}_+(x),\ x < -R,\ \operatorname{erf}_+(x) = \int_{[x,\infty)} e^{x^2/2-t^2} \tag{B.81}$$

So for any $f \in \mathcal{C}_{\rm c}(\mathbb{R})$ we have found a solution of $(P+1)u = f$ with u satisfying the rapid decay conditions (B.78) and (B.81). These are such that if $\chi \in \mathcal{C}^2_{\rm c}(\mathbb{R})$ has $\chi(t) = 1$ in $|t| < 1$ then the sequence

$$u_n = \chi(\frac{x}{n})u(x) \to u,\ u_n' \to u',\ u_n'' \to u'' \tag{B.82}$$

in all cases with convergence in $L^2(\mathbb{R})$ and uniformly and even such that $x^2u_n \to xu$ uniformly and in $L^2(\mathbb{R})$.

This yields the first part of the Lemma, since if $f \in \mathcal{C}_{\rm c}(\mathbb{R})$ and u is the solution just constructed to $(P+1)u = f$ then $(P+1)u_n \to f$ in L^2. So the closure $L^2(\mathbb{R})$ in range of $(P+1)$ on $\mathcal{C}^2_{\rm c}(\mathbb{R})$ includes $\mathcal{C}_{\rm c}(\mathbb{R})$ so is certainly dense in $L^2(\mathbb{R})$.

The second part also follows from this construction. If $f \in L^2(\mathbb{R}) \cap \mathcal{C}(\mathbb{R})$ then the sequence

$$f_n = \chi(\frac{x}{n})f(x) \in \mathcal{C}_{\rm c}(\mathbb{R}) \tag{B.83}$$

converges to f both in $L^2(\mathbb{R})$ and locally uniformly. Consider the solution, u_n to $(P+1)u_n = f_n$ constructed above. We want to show that $u_n \to u$ in L^2 and locally uniformly with its first two derivatives. The decay in u_n is enough to allow integration by parts to see that

$$\int_{\mathbb{R}} (P+1)u_n\overline{u_n} = \|u_n\|^2_{\rm iso} + \|u\|^2_{L^2} = |(f_n, u_n)| \le \|f_n\|_{l^2}\|u_n\|_{L^2}. \tag{B.84}$$

This shows that the sequence is bounded in $H^1_{\rm iso}$ and applying the same estimate to $u_n - u_m$ that it is Cauchy and hence convergent in $H^1_{\rm iso}$. This implies $u_n \to u$ in $H^1_{\rm iso}$ and so both in $L^2(\mathbb{R})$ and locally uniformly. The differential equation can be written

$$(u_n)'' = x^2u_n - u_n - f_n \tag{B.85}$$

where the right side converges locally uniformly. It follows from a standard result on uniform convergence of sequences of derivatives that in fact the uniform limit u is twice continuously differentiable and that $(u_n)'' \to u''$ locally uniformly. So in fact $(P+1)u = f$ and the last part of the Lemma is also proved. □

Bibliography

[3] B. S. Mitjagin, *The homotopy structure of a linear group of a Banach space*, Uspehi Mat. Nauk **25** (1970), no. 5(155), 63–106. MR 0341523 (49 #6274a)

[4] W. Rudin, *Principles of mathematical analysis*, 3rd ed., McGraw Hill, 1976.

[5] George F. Simmons, *Introduction to topology and modern analysis*, Robert E. Krieger Publishing Co. Inc., Melbourne, Fla., 1983, Reprint of the 1963 original. MR 84b:54002

18.102 / 18.1021 Introduction to Functional Analysis
Spring 2021