Linear Analysis
Part II

1 Introduction: What is “linear analysis”?

The objects in this course are infinite dimensional vector spaces (hence the term “linear”) over $\mathbb{R}$ or $\mathbb{C}$, together with additional structure (a “norm” or “inner product”) which “respects” in some way the linear structure. This additional structure will allow us to do “analysis”. The most pedestrian way to understand the last sentence is that it will allow us to “take limits”.

In fact, the extra structure allows much more than just “taking limits”. Hidden in the notion of norm and inner product are notions of convexity, duality, and orthogonality. In the most complicated structure we will discuss, that of a *Hilbert space*, all these notions will be present simultaneously together with completeness.

The point of view of this course, like most undergraduate mathematics courses, will be axiomatic/revisionist. The objects will be defined, and the basic theorems discussed without any “motivation”. The objects in this course have been come to be viewed as so central for mathematics that it is presently inaccurate to consider them as having any single “motivation”.

This being said, the subject arose and developed under specific mathematical circumstances, and this may be useful for deeper understanding of the theory. Nice notes can be found in [3, 2]. At the very least, however, everyone starting out in this subject should know that the vector spaces that originally gave rise to this theory were spaces of *functions*, and the linear operators were *linear partial differential operators*, their inverses, and more generally so-called *linear integral operators*. So the linear algebraic question “Under what conditions can you invert an operator?”–and this represents the kind of question we will be answering in this course–in practice meant “Under what conditions can you solve a linear pde?”.

Because of this connection to spaces of functions, the subject is almost always known by the name *Functional Analysis*, and it is under this name that you will probably search the literature for more material, if you find these notes to be confusing.

2 Normed vector spaces

2.1 Vector spaces

I assume that you are familiar with the notion of *vector space*, *linear transformation*, *subspace*, *quotient space*, *image*, *kernal*, etc. In this course, vector spaces will be over $\mathbb{R}$ or $\mathbb{C}$.

Recall that the *span* of a set S in V is the smallest subspace of V containing S, or alternatively, the set of all finite linear combinations $\sum_{i=1}^{m} \alpha_i s_i$. If this set is V we say S *spans* V. A set $S \subset V$ is *linearly independent* if $\sum_{i=1}^{m} \alpha_i s_i = 0 \implies \alpha_i = 0$ for all i. A set $B \subset V$ is called a *basis* if it spans V and is linearly independent. If S_1 and S_2 are subsets of V, and λ_1, λ_2 are given scalars, we will often use the notation $\lambda_1 S_1 + \lambda_2 S_2$ to denote the set of points[1] of the form $\{\lambda_1 s_1 + \lambda_2 s_2\}_{s_1 \in S_1, s_2 \in S_2}$. Zorn's lemma implies that all vector spaces have a basis, and any two baseis have the same cardinality. We will go through this proof later, as a sort of warm up for the use of Zorn's lemma in the Hahn-Banach Theorem. If the cardinality of B is finite, we say that V is *finite dimensional*, otherwise, *infinite dimensional*.

Most of the Theorems in this course will be easy to prove in the finite dimensional case by purely algebraic methods. Thus, the main emphasis of this course will be the infinite dimensional case.

2.2 Normed vector spaces: the definition

We turn immediately to the central object of study in this course:

Definition 2.1. *A* normed vector space V *is a vector space (over* $\mathbb{R}$ *or* $\mathbb{C}$*) together with*[2] *a function* $V \to \mathbb{R}$*, the value of which at* v *will be denoted* $|v|$*, such that* $|\cdot|$ *satisfies the following*

1. $|v| \geq 0$ *for all* v*, and* $|v| = 0$ *iff* $v = 0$ *(positive definiteness)*
2. $|\lambda v| = |\lambda||v|$*, for a scalar*[3] λ*, and for* $v \in V$
3. $|v + w| \leq |v| + |w|$ *(triangle inequality)*

I will give a whole list of examples soon enough. For the time being, let us be content with $\mathbb{R}^n$, with the Euclidean norm

$$|(x_1, \ldots x_n)| \doteq \sqrt{x_1^2 + \ldots + x_n^2}. \tag{1}$$

(More generally, Let V be an n-dimensional vector space, let $\mathbf{e}_i$ be a basis and let $(x_1, \ldots x_n)$ denote the components of a vector $\mathbf{x}$ with respect to this basis. Then the expression defined by the right hand side of (1) defines a norm on V.) Thus, the notion just defined incorporates in particular one of the most familiar mathematical structures.

2.3 The relation with the topology

Let's think more carefully about the above definition to try to disect its structure. It should be clear that $|\cdot|$ defines a *metric space* structure on V, with

[1] Elements of a vector space will often be referred to as "points".

[2] I assume this use of language is familiar and unproblematic.

[3] One can distinguish from the context whether $|\cdot|$ denotes norm of a vector or absolute value of a scalar.

metric d given by $d(x, y) = |x - y|$, whose three defining properties are inherited from those of the norm.

In particular, the metric space structure allows us to speak of a *topology*, i.e. to identify *open* and *closed* sets and *continuous* mappings. We will see soon that the latter, when required also to be linear, have an alternative characterization with respect to the norm.

In addition to open, closed, and continuous, we may talk about the notion of *basis* for the topology, *product topology*, *homeomorphism*, *dense*, *separable*, *convergent* sequence. Since the topology is metrizable, we may also talk about *Cauchy sequences* and the notion of *completeness*. In the context of normed vector spaces, these notions will be applied without further comment.

To understand the sense in which the topological structure is "married" to the linear, we present the following

Proposition 2.1. *Let $V, |\cdot|$ be a normed vector space. The vector space operations are continuous maps $V \times V \to V$, $\mathbb{R} \times V \to V$.*

Proof. We will do only $+$. Let $\mathcal{U} \subset V$ be open. Want to show that the inverse image $+^{-1}(\mathcal{U})$ is open. Let $(v_1, v_2) \in +^{-1}(\mathcal{U})$. That is to say, $v_1 + v_2 = v$ for some $v \in \mathcal{U}$. Let $B(\epsilon)$ denote the open ball of radius ϵ around the origin. $v + B(\epsilon)$ is the open ball around v of radius ϵ. Clearly $v + B(\epsilon) \subset \mathcal{U}$ for some $\epsilon > 0$. By the triangle inequality, $v_1 + B(\epsilon/2) + v_2 + B(\epsilon/2) = v + B(\epsilon)$. But $(v_1 + B(\epsilon/2), v_2 + B(\epsilon/2))$ is an open neighborhood around (v_1, v_2). So $+^{-1}(\mathcal{U})$ is indeed open, as desired. □

Corollary 2.1. *Let V be as above. Translations and dilations are homeomorphisms.*

Proof. Just consider, for any $v_0 \in V$, the map $V \to V$ given by composing the continuous map $V \to V \times V$ defined by $v \mapsto (v_0, v)$ with the addition map $V \times V \to V$. By the previous Proposition, this is continuous. On the other hand, its (continuous again!) inverse is given by the analogous map defined by $-v_0$. So the map is a homeomorphism.

Similarly for dilations. Let $\lambda \neq 0$ and consider the composition of the map $V \to \mathbb{R} \times V$ defined by $v \mapsto (\lambda, v)$ with the map of the previous proposition, etc. □

2.4 A more abstract setting: topological vector spaces

To understand the message of the above proposition, let us formalise this concept in a definition.

Definition 2.2. *A* topological vector space *is a vector space V, together with a topology, so that the vector space operations are continuous, and so that points are closed sets.*

The latter assumption is technical, and is to ensure that the space is Hausdorff.

2.5 Norms and convexity

At the most abstract level, the main object of this subject would be the topological vector space. Our previous proposition shows that normed vector spaces are a special case.

To understand, at a slightly more abstract level, what is the extra structure given by a norm, let us make the following

Definition 2.3. *Let V be a vector space, and $C \subset V$ a subset. C is said to be* convex *if*

$$tC + (1-t)C \subset C$$

for all $t \in [0,1]$.

Proposition 2.2. *Let $V, |\cdot|$ be a normed vector space. The* unit ball $B(1) \subset V$ *is convex.*

Remark. Note that if C is convex, then translations of C, i.e. sets of the form $p + C$, are also convex.

Definition 2.4. *A* locally convex topological vector space V *is a topological vector space with a basis of convex sets.*

By the above, remark, it is clear that a sufficient condition for a topological vector space to be locally convex is that every open set $\mathcal{U} \subset V$ containing the origin contains a convex neighborhood of the origin. In any case, normed vector spaces are clearly examples of locally convex topological vector spaces in view of Proposition 2.2.

In the class of locally convex topological vector spaces, how special are metric spaces? Let us make the following

Definition 2.5. *Let V be a topological vector space. A subset $B \subset V$ is said to be* bounded *if for any open neighborhood $\mathcal{U} \subset V$ of 0, there exists an $s > 0$ such that $B \subset t\mathcal{U}$ for all $t > s$.*

Remark. If V is a normed vector space, then B is bounded iff $B \subset B(t)$ for some $t > 0$, where $B(t)$ denotes the open unit ball around the origin of radius t.

Proposition 2.3. *Let V be a topological vector space, and assume $C \subset V$ is a bounded convex neighborhood of 0. Then V is* normable, *that is to say, a norm $|\cdot|$ can be defined on V with the same induced topology.*

Proof. For this we need

Lemma 2.1. *$C \subset \tilde{C}$, where $\tilde{C}$ is a* balanced *bounded convex neighborhood of the origin, that is to say, one for which in addition $\lambda\tilde{C} \subset \tilde{C}$ for all $|\lambda| \leq 1$.*

The proof is straightforward and omitted. Now define the function $\mu_{\tilde{C}}$ by

$$\mu_{\tilde{C}}(v) = \inf\{t : v \in t\tilde{C}\}.$$

This function is called the *Minkowski functional* of $\tilde{C}$. One checks explicitly that $|v| \doteq \mu_{\tilde{C}}(v)$ defines the structure of a normed vector space on V. □

Definition 2.6. *A topological vector space V is said to be* locally bounded *if there exists a bounded open neighborhood $\mathcal{U}$ of the origin.*

Normed vector spaces are clearly locally bounded. Proposition 2.3 says that a topological vector space is normable if it is locally bounded and locally convex.

2.6 Banach spaces

After the above diversion, we safely return to the world of normed vector spaces. Fundamental for analysis is being able to take limits. For this, the metric space notion of completeness is extremely useful. When this requirement is added to a normed vector space, one arrives at the definition of a Banach space.

Definition 2.7. *A normed vector space V, $|\cdot|$ is called a* Banach space *if it is complete under the induced metric.*

2.7 Examples

2.7.1 Finite dimensions

We have already seen the example of $\mathbb{R}^n$, or $\mathbb{C}^n$. As we shall see later on, these are Banach spaces, as is any finite dimensional normed vector space.

2.7.2 The space $\mathcal{C}(X)$

In general, let S be a set, and let $\mathcal{F}_{\mathbb{R}}(S)$ denote the set of real valued functions on S, and let $\mathcal{F}_{\mathbb{C}}(S)$ denote the set of complex valued functions on B. These are clearly vector spaces.

Let $\mathcal{B}_{\mathbb{R}}(S)$ denote the set of bounded real functions. This is a subspace of $\mathcal{F}_{\mathbb{R}}(S)$, and thus a vector space. Define

$$|f| = \sup_{s\in S} |f(s)|. \tag{2}$$

It is easy to see that $|f|$ defines a norm, making $\mathcal{B}_{\mathbb{R}}(S)$ into a normed vector space.

On the other hand, let X be a compact Hausdorff space, and consider $\mathcal{C}_{\mathbb{R}}(X)$ the set of all real valued continuous functions. We have

$$\mathcal{C}_{\mathbb{R}}(X) \subset \mathcal{B}_{\mathbb{R}}(S)$$

by well known properties of continuous functions. It inherits thus the norm. (A subspace of a n.v.s is a n.v.s.!) It turns out that $\mathcal{C}_{\mathbb{R}}(X)$ is a Banach space. Actually, you have shown this already in your analysis classes. For you have probably encountered a theorem "Suppose a sequence of continuous f_i on X converge uniformly to f. Then f is continuous." **Exercise:** How far away is this theorem from the statement that $\mathcal{C}_{\mathbb{R}}(X)$ is a Banach space?

All the above considerations apply equally well to $\mathcal{C}_{\mathbb{C}}(X)$

Understanding the topology of $\mathcal{C}_{\mathbb{R}}(X)$, in particular, idenitifying its compact subsets, will be important later on.

2.7.3 $C^k(\bar{U})$

Let $U \subset \mathbb{R}^n$ be open and bounded, and consider now the space of functions $f : U \to \mathbb{R}$ such that $D^\alpha f$ is continuous and uniformly bounded on $\mathcal{U}$ for all multiindices $|\alpha| \leq k$.[4] Denote this space by $C^k(\bar{U})$. Consider the norm $|\cdot|_k$ defined by

$$|f|_k = \max_{|\alpha|\leq k} |D^\alpha f|$$

where $|\cdot|$ denotes the supremum norm. This makes $C^k(\bar{U})$ into a Banach space.

2.7.4 L_p

Let $X = [0,1]$ and let $p > 1$, and consider the set $\hat{L}_p([0,1])$ of all $f : [0,1] \to \mathbb{R}$ such that f is continuous. Define a norm on $\hat{L}_p$ by

$$|f| = \left(\int_0^1 |f|^p\right)^{1/p} \tag{3}$$

This is indeed a norm! The triangle inequality holds in view of Minkowski's inequality. **Exercise:** Why is f positive definite?

For better or for worse, one of the most important facts of life in mathematics is that the above space is **not** a Banach space. That is to say, it is not complete under the induced norm.

From one point of view, this is not a problem. Every incomplete metric space can be embedded into a larger one, which is complete. This larger one is called the completion. For normed vector spaces, one easily shows that every normed vector space V can be realised as a subspace of a larger one V', such that V inherits its norm from V', V' is a Banach space, and $\overline{V} = V'$. Moreover V' is unique up to isometry (we'll see what that means later). We call V' the completion of V.

The problem is that the above construction is completely abstract. Elements of the completion are equivalence classes of Cauchy sequences of V.

One of the miracles then of analysis is that the completion of $\hat{L}_p$ can be realised as equivalence classes of Lebesgue measurable functions $f : [0,1] \to \mathbb{R} \cup \pm\infty$ where $f \sim g$ iff $f = g$ a.e. Let us call this space L_p, without the hat. The norm if just (3), where this integral is interpreted in the sense of Lebesgue. L_p is thus a Banach space.

2.7.5 l_p

The previous example is one of the most fruitful for the application of linear analysis. Since we do not have the technology of measure theory at our disposal, we will have to settle with a baby example, where functions are replaced by sequences. (The latter of course are just functions on the natural numbers, and

[4] Here $\alpha = (\alpha_1, \ldots \alpha_n)$ nonnegative integers, with $|\alpha| = \sum_i \alpha_i$, and $D^\alpha f$ denotes $\frac{\partial^{|\alpha|}}{\partial x_1^{\alpha_1} \cdots \partial x_n^{\alpha_n}}$.

this example can be thought of as a special case of the previous where $[0, 1]$ is replaced by $\mathbb{N}$.) This is so-called "little" l_p.

For $0 < p < \infty$, we define

$$l_p(\mathbb{C}) = \{(x_1, x_2, x_3, \ldots), x_i \in \mathbb{C} : \sum_{i=1}^{\infty} |x_i|^p < \infty\},$$

and for $p = \infty$, we define

$$l_\infty(\mathbb{C}) = \{(x_1, x_2, x_3, \ldots), x_i \in \mathbb{C} : \sup_i |x_i| < \infty\}.$$

Alternatively, we may replace $\mathbb{C}$ with $\mathbb{R}$. We may think of these as subsets of $\mathcal{F}_{\mathbb{C}}(\mathbb{N})$ or $\mathcal{F}_{\mathbb{R}}(\mathbb{N})$. By the Minkowski inequality, l_p is a subspace, and can be made into a normed vector space with norm defined by

$$|x| = \left(\sum_{i=1}^{\infty} |x_i|^p \right)^{1/p} \tag{4}$$

for $p \geq 1$, and

$$|x| = \sup_i |x_i|, \tag{5}$$

for $p = \infty$. We will see later on that these are in fact Banach spaces. In the case $0 < p < 1$, l_p is again a subspace of $\mathcal{F}_{\mathbb{C}}(\mathbb{N})$, but the expression (4) does not define a norm.

2.7.6 $\mathcal{C}(\Omega)$, $C^\infty(\Omega)$

Let $\Omega \subset \mathbb{R}^n$ be *open*, and let $\mathcal{C}(\Omega)$ denote the set of continuous functions on Ω. Continuous functions are no longer necessarily bounded, thus the expression (2) is no longer well defined. We can still make $\mathcal{C}(\Omega)$ into a topological vector space as follows. Choose a sequence of compact K_i such that $\cup K_i = \Omega$, and $K_i \subset K_{i+1}$. Let

$$V(i, n) = \{f : |f|_{\mathcal{C}(K_i)} < 1/n\},$$

and consider the topology on $\mathcal{C}(\Omega)$ generated by this family and all its translates. This defines on $\mathcal{C}(\Omega)$ the structure of a locally convex topological vector space.

2.8 Bounded linear maps

Now we know what the objects of interest are in this course. So what are the morphisms, will ask the well-trained student in the language of category theory. These are the so-called bounded linear maps.

Definition 2.8. *Let V, W be topological vector spaces. A linear map $T : V \to W$ is said to be bounded if $E \subset V$ bounded implies $T(E) \subset W$ is bounded.*

Proposition 2.4. *Let V, W be topological vector space, and $T : V \to W$ be linear, and assume V and W are locally bounded. Then T is bounded iff T is continuous.*

Proof. First, I claim that T is continuous iff T is continuous at 0, i.e. iff for any open subset $\mathcal{U}$ in W around 0, there exists an open neighborhood of 0 in V contained in $T^{-1}(\mathcal{U})$.

To see this, suppose then that indeed for any open neighborhood $\mathcal{U}$ of the origin in W, there exists an open neighborhood $\tilde{\mathcal{U}}$ of the origin in V, with $\tilde{\mathcal{U}} \subset T^{-1}(\mathcal{U})$. Let v_0 and w_0 be arbitrary points such that $T(v_0) = w_0$. Then

$$T(v_0 + \tilde{\mathcal{U}}) = w_0 + T(\tilde{\mathcal{U}}) \subset w_0 + \mathcal{U}.$$

This shows that the the inverse image of any open subset is open. (The other direction of the implication is of course immediate.)

We now continue with the proof of the proposition. Suppose then T is bounded. Let $\mathcal{U} \subset W$ be an arbitrary open neighborhood of 0, and let $\tilde{\mathcal{U}} \subset V$ be a bounded open neighborhood of 0. The latter exists since V is locally bounded. We have that $T(\tilde{\mathcal{U}})$ is bounded, by assumption. Thus, by definition, there exits an $n > 0$ such that $T(\tilde{\mathcal{U}}) \subset n\mathcal{U}$. But then

$$n^{-1}\tilde{\mathcal{U}} \subset T^{-1}(\mathcal{U}),$$

and this proves continuity, since $\mathcal{U}$ open implies $n^{-1}\mathcal{U}$ open by Corollary 2.1.

Suppose conversely that T is continuous, and let $E \subset V$ be bounded. Let $\mathcal{U}$ be an arbitrary open neighborhood of 0 in W. Since $T^{-1}(\mathcal{U})$ is open, there exits an open neighborhood of the origin $\tilde{\mathcal{U}} \subset V$ such that $T^{-1}(\mathcal{U}) \supset \tilde{\mathcal{U}}$, i.e. such that

$$\mathcal{U} \supset T(\tilde{\mathcal{U}}).$$

On the other hand, by boundedness, it follows that

$$E \subset t\tilde{\mathcal{U}}$$

for all $t > s$, but then

$$T(E) \subset T(t\tilde{\mathcal{U}}) \subset tT(\tilde{\mathcal{U}}) = t\mathcal{U},$$

so $T(E)$ is bounded. Thus T is bounded. □

Specialising the definition to normed vector spaces we obtain

Definition 2.9. *Let V, W be normed vector spaces. A linear map $T : V \to W$ is said to be bounded if $T(B(1)) \subset B(t)$ for some $t > 0$. We define $||T||$ to be the infimum of all $t > 0$ such that the previous inclusion holds.*

It is easy to see that $||T||$ is alternatively characterized by

$$||T|| = \sup_{|v|\leq 1} |Tv| = \sup_{|v|=1} |Tv| = \sup_{|v|<1} |Tv|.$$

Let $\mathcal{L}(V, W)$ denote the set of linear maps $T : V \to W$. This is clearly a vector space. Define $\mathcal{B}(V, W)$ to be set of all bounded linear maps. We have

Proposition 2.5. *$\mathcal{B}(V,W)$ is a subspace of $\mathcal{L}(V,W)$. $||T||$ defines a norm on $\mathcal{B}(V,W)$.*

Proof. The proof of this is very easy. Clearly $0 \in \mathbb{B}(V,W)$ and $||0|| = 0$. By linearity, it is clear that if $T \in \mathcal{B}(V,W)$, then $\lambda T \in \mathcal{B}(V,W)$ with $||T|| = |\lambda|||T||$, since

$$(\lambda T)(B(1)) = T(\lambda B(1)) = T(|\lambda|B(1)) = |\lambda|(T(B(1)))$$

so (if $\lambda \neq 0$[5])

$$T(B(1)) \subset B(t)$$

iff

$$(\lambda T)(B(1)) \subset |\lambda|B(t)$$

which is equivalent to

$$(\lambda T(B(1)) \subset B(|\lambda|t).$$

Thus the set is closed under scalar multiplication, and the norm satisfies the second property. Finally, suppose $T_1, T_2 \in \mathcal{B}(V,W)$, and the definition is satisfied with t_1, t_2. We have

$$\begin{aligned}(T_1 + T_2)(B(1)) &= T_1(B(1)) + T_2(B(1)) \\ &\subset B(t_1) + B(t_2) \\ &= B(t_1 + t_2)\end{aligned}$$

Taking the infimum over T, we obtain in particular

$$||T_1 + T_2|| \leq ||T_1|| + ||T_2||.$$

□

2.9 The dual space V^*

Definition 2.10. *Let V be a normed vector space over $\mathbb{R}$ ($\mathbb{C}$, respectively). The space $\mathcal{B}(V,\mathbb{R})$ (respectively, $\mathcal{B}(V,\mathbb{C})$) is known as the* dual *of V and is denoted V^*.*

To distinguish, we will call $\mathcal{L}(V,\mathbb{R})$ (respectively, $\mathcal{L}(V,\mathbb{C})$) the *algebraic dual* of V.

Proposition 2.6. *Let W be a Banach space, and V a normed vector space. Then $\mathcal{B}(V,W)$ is a Banach space.*

Proof. Let $T_i \in \mathcal{B}(V,W)$ be Cauchy. Let $v \in V$. $T_i(v)$ is clearly Cauchy in W, because

$$|T_i(v) - T_j(v)| \leq ||T_i - T_j|| \cdot |v|.$$

[5]If $\lambda = 0$, again, there is nothing to show.

Thus $T_i(v)$ converges by the completeness of W. Define $T(v) = \lim T_i(v)$. Claim. T is linear, i.e. $T \in \mathcal{L}(V, W)$. This follows immediately from the continuity of addition in W. On the other hand, T is clearly bounded. For given any $|v| \leq 1$,

$$\begin{aligned} |T(v)| &\leq |T(v) - T_i(v)| + |T_i(v)| \\ &\leq \epsilon + ||T_i|| \cdot |v| \\ &\leq \epsilon + ||T_i||. \end{aligned}$$

The above should be interpreted as follows. For any $\epsilon > 0$, $v > 0$ there exists an i depending on ϵ and v such that the second inequality holds.

Thus we have $||T|| < \infty$, so $T \in \mathcal{B}(V, W)$. Finally we must show that $||T - T_i|| \to 0$. We compute

$$\begin{aligned} |(T - T_i)(v)| &\leq |(T - T_j)(v)| + |(T_j - T_i)(v)| \\ &\leq \epsilon + ||T_j - T_i|| \cdot |v| \\ &\leq \epsilon + ||T_j - T_i||, \end{aligned}$$

where this is to be interpreted as follows: For any $\epsilon > 0$, there exists a j depending on v such that the second inequality holds for $j > i$. Now choose i such that $||T_j - T_i|| < \epsilon$ and we're done.

Note of course that $||T|| = \lim ||T_i||$. □

In particular, since $\mathbb{R}$ (resp. $\mathbb{C}$) are Banach spaces, it follows that

Corollary 2.2. *Let V be a normed vector space. Then V^* is a Banach space.*

It turns out that a good technique for proving that a n.v.s. is in fact Banach is exhibiting as the dual space of another n.v.s.

2.10 Adjoint map

Let V, W, be normed vector spaces, $T \in \mathcal{B}(V, W)$. Let us define a map $T^* : W^* \to V^*$ as follows. For any $w^* \in W^*$, let $T^*(w^*)$ be the map $V \to \mathbb{R}$ (or $\mathbb{C}$ respectively) defined by

$$T^*(w^*)(v) = w^*(Tv).$$

$T^*(w^*)$ is clearly in the algebraic dual of V. Moreover, for $|v| \leq 1$,

$$|T^*(w^*)(v)| = |w^*(Tv)| \leq ||w^*|| \cdot |Tv| \leq ||w^*|| ||T||$$

so in fact $T^*(w^*) \in V^*$. So $T^* \in \mathcal{B}(W^*, V^*)$. On the other hand, the above computation shows that

$$||T^*(w^*)|| \leq ||w^*|| ||T||$$

and thus,

$$||T^*|| \leq ||T||.$$

In fact, $||T^*|| = ||T||$, but this will need Hahn-Banach.

2.11 V^{**}

We define the double dual V^{**} of a normed vector space to be $(V^*)^*$, i.e. the dual of the dual.

Proposition 2.7. *The map $\phi : V \to V^{**}$ defined by*

$$v \mapsto v^{**}$$

where $v^{**} \in V^{**}$ is in turn defined by $v^{**}(f) = f(v)$, for all $f \in V^*$, is a bounded linear map.

Proof. Linearity is clear. Moreover, the identity

$$|v^{**}(f)| = |f(v)| \leq ||f|||v|$$

shows that $||\phi|| \leq 1$. □

Later on in the course, the Hahn-Banach theorem will show that this map is an isometry, in particular the map is injective.

Note that, in contrast to the finite dimensional case, the map ϕ is not in general surjective.

2.12 Examples!

2.12.1 Finite dimensions

Let V, W be finite dimensional normed vector spaces. Then any linear map $T : V \to W$ is bounded. We will see this later on. **Exercise:** Let v_i, w_i be baseis for V, W, respectively. What is the relation between the matrix of T and T^*?

2.12.2 Infinite dimensions

Consider l_p, and define $T : l_p \to l_p$ by $(x_1, x_2, \ldots) \mapsto (0, x_1, x_2, \ldots)$. Note that this map is bounded, has $||T|| = 1$, and is injective but not surjective.

Consider the map

$$D : C^1[0,1] \to C^0[0,1]$$

taking $f \mapsto f'$, where $'$ denotes the derivative. This is linear by the properties of derivatives ($(f+g)' = f' + g'$, etc.). Moreover, by defiinition of the relevant norms, one sees trivially that $||D|| \leq 1$, and immediately that $||D|| = 1$.

For a linear map which is *not* bounded, consider the space X, which as an underlying vector space is $C^1[0,1]$, but is made into a normed vector space with the induced norm of $C^0[0,1]$, i.e. the sup norm. Consider the map

$$id : X \to C^1[0,1]$$

where the target is the usual C^1 defined as in Section 2.7.3, and the map is just the identity. This map is clearly linear! But it is not bounded. For one can consider a sequence f_i of C^1 functions such that $|f_i| \leq 1$, yet $|f_i'| \to \infty$.

3 Finite dimensional normed vector spaces

The point of the theory developed in this course is to treat the infinite dimensional case. In this section, we shall see why the finite dimensional case is so very different.

Let us begin with a definition.

Definition 3.1. *We say that two norms* $|\cdot|_1$ *and* $|\cdot|_2$ *on a vector space* V *are* equivalent *if there exits constants* $0 < c < C < \infty$ *such that*

$$c|v|_1 \leq |v|_2 \leq C|v|_1.$$

One easily sees that this defines an equivalence relation on the set of norms on V. Clearly the induced topology defined by two equivalent norms is the same. In particular, the notion of bounded operator with respect to the norms coincides. Moreover, the notion of a Cauchy sequence with respect to the two norms is the same. Thus, if $|\cdot|_1$ and $|\cdot|_2$ are equivalent, then v_i is Cauchy with respect to $|\cdot|_1$ iff it is Cauchy with respect to $|\cdot|_2$.

In this section we will show that all finite dimensional normed vector spaces have equivalent norms and are Banach. We will show that all linear maps are bounded. Moreover, we will show the closed unit ball in a finite dimensional n.v.s. is compact. Finally, we shall show that the latter fact is a characterization of finite dimensionality for normed vector spaces, that is to say, if the closed unit ball is compact, then the dimension is *necessarily* finite.

Given a finite n dimensional vector space V, we may think of it as $\mathbb{R}^n$ or $\mathbb{C}^n$ after choice of basis. Let us define a norm on $\mathbb{R}^n$, resp. $\mathbb{C}^n$, by

$$|x|_1 = \sum_{i=1}^{n} |x_i|.$$

We denote the induced normed vector space l_1^n.

Proposition 3.1. *Let* $|\cdot|$ *be a norm on* $\mathbb{R}^n$*. Then* $|\cdot|$ *is equivalent to* $|\cdot|_1$*.*

Note the immediate

Corollary 3.1. *All norms on* $\mathbb{R}^n$ *are equivalent.*

Proof. First, the easy direction. Let e_i denote the standard basis. We have

$$|x| \leq \sum_i |x_i||e_i| = \big(\max_{1\leq i\leq n} |e_i|\big) \sum_{i=1}^{n} |x_i| \leq C|x|_1$$

where $C = \max_{1\leq i\leq n} |e_i|$.

For the other direction, we need some Lemmas.

Lemma 3.1. *The function* $|\cdot|$ *is continuous on the unit circle* $S(1)$ *with respect to* $|\cdot|_1$*.*

Proof. This follows from the computation

$$||x|-|y|| \le |x-y| \le C|x-y|_1.$$

$\square$

Lemma 3.2. *The closed unit ball (and thus the unit circle) are compact in the topology of l_1^n.*

Proof. Suppose x^i is a sequence. Then x_p^i is a sequence in $\mathbb{R}$ or $\mathbb{C}$. Choose a subsequence $x_1^{j_k}$ that converges to $\tilde{x}_1$, then a further subsequence $x_2^{j_{k_m}}$, etc. Construct from the final choices a sequence we will again denote by x^i. Clearly $x^i \to \tilde{x} = (\tilde{x}_1, \dots, \tilde{x}_n)$. Since $|x^i - \tilde{x}|_1 = \sum_{j=1}^n |x_j^i - \tilde{x}_j^i| \to 0$. $\square$

In fact, the above proof shows that any closed ball is compact in the topology of l_1^1, and even more generally, any closed and bounded set.

From the above we know that $|\cdot|$ a continuous function on a compact set attains its infimum. The latter then must be strictly positive. Thus we have

$$0 < c \le |x|$$

for all $|x|_1 = 1$. But then, applying this to an arbitrary $\tilde{x} = \lambda x$, for $0 \neq |\lambda| = |\tilde{x}|_1$ we obtain

$$0 < |\tilde{x}|_1 c = |\lambda| c \le |\lambda||x| = |\lambda x| = |\tilde{x}|.$$

This completes the proof. $\square$

Proposition 3.2. *Let V be a finite dimensional normed vector space. Then the closed unit ball is compact.*

Proof. This follows from the statement proven about l_1^1 in Lemma 3.2 and from the previous proposition. $\square$

Proposition 3.3. *Let V be a finite dimensional normed vector space. Then V is a Banach space.*

Proof. Let v_i be Cauchy in V. It follows that v_i is in particular bounded, i.e. there exists an R such that $v_i \in \overline{B(R)}$. But $\overline{B(R)}$ is compact (so in particular complete!). So v_i converges. $\square$

Corollary 3.2. *Let $V \subset W$, where W is a normed vector space, and V is a finite dimensional subspace. Then V is closed.*

Proposition 3.4. *Let $T : V \to W$ be linear, where V and W are normed vector spaces and V is finite dimensional. Then T is bounded.*

Proof. Since ImT is clearly finite dimensional, it suffices to consider the case where W is also finite dimensional. It suffices to prove this when $V = l_1^n$, $W = l_1^m$. Consider the matrix T_{ij} associated to T. We have

$$T(x_1, \dots, x_n) = \big(\sum T_{i1} x_i, \dots, \sum T_{im} x_i \big)$$

and thus

$$\begin{aligned} |T(x_1, \dots, x_n)|_1 &= \sum_{ij} |T_{ij} x_i| \\ &= \sum_{ij} |T_{ij}||x_i| \\ &\leq \max_{ij} |T_{ij}| \sum_i |x_i| \\ &= C|(x_1, \dots, x_n)|_1, \end{aligned}$$

where $C = \max_{ij} |T_{ij}|$. This of course completes the proof. $\square$

Finally:

Proposition 3.5. *Let V be a normed vector space with the property that the closed unit ball is compact. Then V is finite dimensional.*

Proof. Consider the open cover of the closed unit ball consisiting of all open balls around arbitrary points, of radius 1/2. By compactness, there exits a finite subcover, i.e.

$$B(1) \subset \cup_{i=1}^n (x_i + B(1/2)).$$

Now let W denote the subset of spanned by x_i. This of course is finite dimensional with $\dim W \leq n$. We have

$$B(1) \subset W + B(1/2).$$

Iterating once we obtain,

$$B(1) \subset W + \frac{1}{2}(W + B(1/2)) = W + B(1/4).$$

Iterating arbitrarily many times, we obtain

$$B(1) \subset W + B(1/2^m)$$

for all $m \geq 0$, and thus

$$B(1) \subset \cap_m (W + B(1/2^m)) = \overline{W} = W$$

where for the last equality we have used Corollary 3.2. But of course, this implies that $V \subset W$, and thus $V = W$. So V is finite dimensional with $\dim V \leq n$. $\square$

The above proof can easily be generalised to a statement about *locally compact* topological vector spaces, i.e. topological vector spaces with a neighborhood of the origin with compact closure.

4 The Hahn-Banach Theorem

The Hahn-Banach Theorem is essentially a statement about the richness of V^*, i.e. it says (or rather, its corollaries show) that this space is sufficiently big. **At this point, we don't even know whether in general $V \neq 0$ implies $V^* \neq 0$!**

The idea here is that one constructs elements of V^* be defining linear functionals on subspaces of V (for instance finite dimensional ones, where this is more or less trivial) and then extends them "little by little" to the whole of V. The fact that one can extend a bounded linear functional from a codimension 1 subspace to the whole space is the content of Proposition 4.3. If

$$V = \cup V_i, \tag{6}$$

where V_i are finite dimensional subspaces with $V_i \subset V_{i+1}$ of codimension 1, then this allows one to obtain bounded linear functionals on V by induction.

The analogue of this induction procedure in the general case when (6) does not hold is known as transfinite induction. We begin with a general discussion.

4.1 Introduction to transfinite induction

Everyone has their favourite method of transfinite induction. For us, it will be via Zorn's lemma.

Definition 4.1. *A partially ordered set is a set $\mathcal{S}$ together with a binary relation $\leq$, i.e. a relation such that for all x, y in S either $x \leq y$ or $x \not\leq y$, satisfying*

$$x \leq x,$$

$$x \leq y,\ y \leq z \implies x \leq z,$$

$$x \leq y,\ y \leq x \implies x = y.$$

Definition 4.2. *Let $\mathcal{S}$ be a partially ordered set. A subset $\mathcal{T} \subset \mathcal{S}$ is called* totally ordered *if $x \not\leq y \implies y \leq x$.*

Definition 4.3. *Let $\mathcal{S}$ be a partially ordered set, and $\tilde{\mathcal{S}}$ a subset. An element $b \in \mathcal{S}$ is said to be an* upper bound *for $\tilde{\mathcal{S}}$ if $x \leq b$ for all $x \in \tilde{\mathcal{S}}$. An element $l \in \mathcal{S}$ is said to be a* least upper bound *for $\tilde{\mathcal{S}}$ it is an upper bound, and moreover $l \leq b$ for any upper bound b.*

Definition 4.4. *Let $\mathcal{S}$ be a partially ordered set. We say that $\mathcal{S}$ has the* least upper bound property *if every non-empty totally ordered subset has a least upper bound.*

Definition 4.5. *Let $\mathcal{S}$ be a partially ordered set. An element $m \in \mathcal{S}$ is called* maximal *if $m \leq x \implies m = x$.*

Lemma 4.1. *(Zorn) Let $\mathcal{S}$ be a non-empty partially ordered set with the least upper bound property. There exists a maximal element $m \in \mathcal{S}$.*

The above "lemma" is equivalent to the axiom of choice.

4.2 Application: Every vector space has a basis

Proposition 4.1. *Let $V \neq \{0\}$ be a vector space. Then there exists a basis for V.*

In fact we will prove something stronger namely

Proposition 4.2. *Let $V \neq \{0\}$ be a vector space, and let $S \subset V$ be linearly independent. Then there exists a basis B of V with $S \subset B \subset V$.*

Proof. Let $\mathcal{S}$ be the set of all linear independent subsets of V, containing B. This set is non-empty since it contains S. We may partially order $\mathcal{S}$ as follows. For $S_1, S_2 \in \mathcal{S}$, we say that $S_1 \leq S_2$ if $S_1 \subset S_2$. One checks easily that this defines an order relation. Suppose now that $\mathcal{T} \subset \mathcal{S}$ is a non-empty totally ordered subset. Define $S_b = \cup_{\tilde{S} \in \mathcal{T}} \tilde{S}$. The set $\tilde{S}$ contains S, and is linearly independent, because if $\sum_{i=1}^{m} \alpha_i x_i = 0$, then there exists by total ordering a $\tilde{S} \in \mathcal{T}$ such that $x_i \in \tilde{S}$ for all $i = 1, \ldots n$, and one applies linear independence of $\tilde{S}$ to deduce that $\alpha_i = 0$. Thus, S_b is clearly a least upper bound in $\mathcal{S}$, so $\mathcal{S}$ has the least upper bound property.

By Zorn's lemma, it follows that $\mathcal{S}$ has a maximal element. Call this B. To show B is a basis, we need only show that $\mathrm{Span}(B) = V$. Suppose this is not the case, i.e. suppose $\exists v : v \notin \mathrm{SpanB}$. Consider the set

$$\tilde{B} = \{v\} \cup B.$$

Claim. $\tilde{B}$ is linearly independent. For if $\alpha v + \sum_{i=1}^{n} \alpha_i v_i = 0$ for $v_i \in \tilde{B}$, and $\alpha \neq 0$, one obtains

$$v = -\sum \alpha^{-1} \alpha_i v_i$$

and thus $v \in \mathrm{Span}(B)$, a contradiction. But now we have $B \leq \tilde{B}$, but $B \neq \tilde{B}$, and this is contradicts the maximality of B. So B is indeed a basis. □

Note how it was important in the above that the scalars constitute a field. The above result does not hold for modules.

Exercise. Show that any two baseis have the same cardinality.

4.3 The statement of the theorem

Let us restrict to real vector spaces V. We sometimes call linear maps $f : V \to \mathbb{R}$ *linear functionals*. The Hahn-Banach theorem is a statement about the extendibility of bounded linear functionals defined on a subspace to linear functionals on the whole space with the same norm.

4.3.1 The codimension 1 case

We begin with the case of codimension 1.

Although our primary application is to bounded linear functions, the following somewhat more general setting is convenient.

Proposition 4.3. *Let V be a real vector space, and let $p : V \to \mathbb{R}$ be a function $p : V \to \mathbb{R}$, with $p(v_1 + v_2) \leq p(v_1) + p(v_2)$, and $p(\lambda v) = \lambda p(v)$ for all $\lambda > 0$. Let W be a codimension 1 subspace of V, and suppose that $f : W \to \mathbb{R}$ is a linear functional such that $f(w) \leq p(w)$ for all $w \in W$. Then there exists a linear functional $\tilde{f} : V \to \mathbb{R}$ such that $\tilde{f}|_W = f$ and $\tilde{f}(v) \leq p(v)$ for all $v \in V$.*

The assumptions on p imply in particular that is is a convex function. Note that, in the case of a normed vector space V, if $||f|| < \infty$, then the conditions of the proposition are satisfied with $p(x) = ||f|||x|$, and one has that $||\tilde{f}|| = ||f||$.

Proof. Pick $v_0 \in V \setminus W$. By codimensionality 1, it suffices to define $\tilde{f}$ so that $\tilde{f}(w + av_0) \leq p(w + av_0)$ for all $a \in \mathbb{R}$. That is to say $\tilde{f}(w) + af(v_0) \leq p(w + av_0)$. This is equivalent to

$$a\tilde{f}(v_0) \leq p(w + av_0) - f(w)$$

and thus, with, $a > 0$, we obtain

$$\tilde{f}(v_0) \leq a^{-1}p(w + av_0) - f(a^{-1}w),$$

and thus

$$\tilde{f}(v_0) \leq p(a^{-1}w + v_0) - f(a^{-1}w),$$

for all w, a condition which, by redefining w, we may write as

$$\tilde{f}(v_0) \leq p(w + v_0) - f(w) \tag{7}$$

for all w.

On the other hand, for $a < 0$, we obtain the condition

$$\tilde{f}(v_0) \geq a^{-1}p(w + av_0) - a^{-1}f(w),$$

or equivalently

$$\tilde{f}(v_0) \geq -(-a^{-1})p(w + av_0) + f(-a^{-1}w)$$

and thus, after redefining w as above, we obtain the condition

$$\tilde{f}(v_0) \geq -p(w - v_0) + f(w). \tag{8}$$

for all w.

So now, the condition that $\tilde{f}(v_0)$ can be chosen so that inequalities (7) and (8) both hold is just that, for all w, $\tilde{w}$:

$$f(\tilde{w}) + f(w) \leq p(w + v_0) + p(\tilde{w} - v_0). \tag{9}$$

But indeed, (9) holds. For we have

$$\begin{aligned} f(\tilde{w}) + f(w) &= f(\tilde{w} + w) \\ &\leq p(\tilde{w} + w) \\ &= p(w + v_0 + \tilde{w} - v_0) \\ &\leq p(w + v_0) + p(\tilde{w} - v_0). \end{aligned}$$

□

4.3.2 The general case

Theorem 4.1. *Proposition 4.3 holds without the assumption that W has codimension 1.*

Proof. Consider the set $\mathcal{S}$ of all extensions $\tilde{f} : \tilde{V} \to \mathbb{R}$ of f, where $W \subset \tilde{V} \subset V$ is a subspace, where $\tilde{f}|_V = f$, and where $\tilde{f}(x) \leq p(x)$ for all $x \in \tilde{V}$.

This set is nonempty, as it contains f itself. We may partially order the set by setting $\tilde{f}_1 \prec \tilde{f}_2$ if $\tilde{V}_1 \subset \tilde{V}_2$, where $\tilde{V}_i$ are the domains of $\tilde{f}_i$, and $\tilde{f}_2|_{\tilde{V}_1} = \tilde{f}_1$. This clearly defines a partial ordering.

Claim. Under this partial ordering, $\mathcal{S}$ satisfies the least upper bound property. For if $\mathcal{T}$ is a totally ordered subset, we can consider a $\tilde{f}$ defined on $\cup_{\tilde{V} \in \mathcal{T}} \tilde{V}$ by $\tilde{f}(x) = \tilde{f}_\alpha(x)$ for some $\tilde{f}_\alpha$ such that its domain $\tilde{V}_\alpha$ contains x. One easily sees that since $\mathcal{T}$ is totally ordered, this definition does not depend on the choice. Finally, it is clear that $\tilde{f}_\alpha \prec \tilde{f}$, for any $\tilde{f}_\alpha \in \mathcal{T}$.

We may thus apply Zorn's lemma to obtain a maximal $\tilde{f} \in \mathcal{S}$. We are left with proving that the domain of $\tilde{f}$ is V. So let us suppose that this is not the case. Let $v_0 \in V \setminus \tilde{W}$, where $\tilde{W}$ denotes the domain of $\tilde{f}$. Consider the set $\tilde{V} = \mathrm{Span}(v_0, \tilde{W})$. We have that $\tilde{W} \subset \tilde{V}$ is codimension 1. We may apply then Proposition 4.3 to extend $\tilde{f}$ to a linear $\tilde{\tilde{f}} : \tilde{V} \to \mathbb{R}$, with $\tilde{\tilde{f}}(x) \leq p(x)$. But clearly, $\tilde{f} \prec \tilde{\tilde{f}}$ and $\tilde{f} \neq \tilde{\tilde{f}}$. This contradicts maximality.

So the domain of $\tilde{f}$ is V, and the theorem is proven. □

Corollary 4.1. *Let V be a normed vector space, and $W \subset V$ a subspace. Let $f \in W^*$. Then there exits an $\tilde{f} \in V^*$ with $\tilde{f}|_W = f$, and $||\tilde{f}|| = ||f||$.*

We will refer to Theorem 4.1 or Corollary 4.1 as the *Hahn-Banach theorem.*

4.4 The dual space revisited

Armed with Hahn-Banach, we may now show that the dual space is big enough. The main tool will be the following

Proposition 4.4. *Let V be a normed vector space, and $v \in V$ an arbitrary element. There exists an $f_v \in V^*$ such that $||f_v|| = 1$, and $f_v(v) = |v|$.*

Such an f_v is called a support functional for v.

Proof. Consider the one-dimensional subspace W spanned by v, and define a $f_v \in W^*$ by $f_v(v) = |v|$. Clearly, $||f_v|| = 1$. Now apply Hahn-Banach in the form of Corollary 4.1. □

Corollary 4.2. *Let V be a normed vector space, and let $v \in V$. Then $v = 0$ iff $f(v) = 0$ for all $f \in V^*$.*

In particular,

Corollary 4.3. *Let $V \neq 0$ be a normed vector space. Then $V^* \neq 0$.*

In fact

Corollary 4.4. *Let V be a normed vector space, $v, w \in V$ with $v \neq w$. Then there exists an $f \in V^*$ such that $f(v) \neq f(w)$.*

Proof. Just take $f = f_{v-w}$. $\square$

For another manifestation of the richness of V^*, let us consider V^{**}. We have

Proposition 4.5. *The map $\phi : V \to V^{**}$ is an isometry, i.e. $||\phi(v)|| = |v|$.*

In particular, ϕ is injective.

Proof. We have already shown in Proposition 2.7 that $||\phi(v)|| \leq |v|$. For the other direction, just note that for $|v| = 1$, we can choose a support functional f_v with $||f_v|| = 1$, and this gives

$$|\phi(v)(f_v)| = |f_v(v)| = |v| = 1,$$

and thus $||\phi(v)|| \geq 1$ for all $|v| = 1$. But this gives $||\phi(v)|| \geq |v|$ for all v. $\square$

Finally, in a similar spirit, we show

Proposition 4.6. *Let V and W be normed vector spaces and let $T : V \to W$ be a bounded linear map. Then $T^* : W^* \to V^*$ satisfies $||T^*|| = ||T||$.*

Proof. Again, we have already shown that $||T^*|| \leq ||T||$. Let $v \in V$ be arbitrary with $|v| = 1$, and choose a support functional f_w for $w = Tv$. We have

$$(T^* f_w)(v) = f_w(T(v)) = |w|.$$

Thus, $||T^* f_w|| \geq |w|$. Since $||f_w|| = 1$, we have that

$$||T^*|| \geq |Tv|.$$

We thus have that

$$||T^*|| \geq \sup_{|v|=1} |Tv| = ||T||$$

and this completes the proof. $\square$

4.5 Remarks and Examples of V^*

One is often quite weary in mathematics to take seriously anything proven with the help of Zorn's lemma. Perhaps a healthy attitude to take towards the Hahn-Banach Theorem is that it is a satisfying and clarifying general statement that in practice one rarely invokes.

Here, by "in practice", one should read, when applying the theory to a particular normed vector space. For most normed vector spaces appearing in analysis, one has a good idea of the what their dual is, or at least, can easily construct a large subset of the dual. One has no need for applying the axiom of choice to construct such elements.

Example. Consider the space $C([0,1])$. If g is a continuous function then we can identify g with an element of $(C([0,1]))^*$ defined by

$$f \mapsto \int_0^1 fg.$$

One easily sees that the norm of g thought of as an element in $(C([0,1]))^*$ is $\int |g|$. Since $(C([0,1]))^*$ is a Banach space, and the set $C([0,1])$ is not complete under the L^1 norm, it is clear that $(C([0,1]))^*$ contains elements not of the above form. For instance, $f \mapsto f(x_0)$ is an element of the dual not induced as above. It turns out that the dual of $C([0,1])$ can be identified with the space of signed Borel measures. This is not a space we have the technology to work with in this course.

Example. Consider the space l_p for $\infty > p \geq 1$. Let $x \in l_q$ where q is defined by

$$\frac{1}{p} + \frac{1}{q} = 1, \tag{10}$$

with the convention that if $p = 1$, $q = \infty$. For $s \geq 1$, let $|\cdot|_s$ denote the l_s norm. We call q the *conjugate exponent* of p.

The element x defines an element of l_p^* by the action

$$y \mapsto \sum x_i y_i.$$

The convergence of this sum follows from the Hölder inequality, stating that for $x \in l_q$, $y \in l_p$ satisfiying (10)

$$|xy|_1 \leq |x|_q |y|_p.$$

It turns out (and you will show this on example sheets) that this identification yields an isometric isomorphism of l_p^* and l_q. What about l_∞^*?

5 Completeness

In the last section, we essentially exploited the structure of *convexity* provided by a norm defined on a vector space. (Our slightly more general formulation of Hahn-Banach in terms of a convex functional p should make this clear.) In this section, we will exploit *completeness*. Thus, Banach spaces will become here important.

The considerations of the current section stem from the observation that complete metric spaces are necessarily "big". This notion of "big" is captured by a concept known as *Baire category*.

5.1 Baire category

Definition 5.1. *Let X be a topological space. We say that X is of* first category *if*

$$X = \cup_i E_i$$

where E_i is nowhere dense[6]*. Otherwise, we say that X is of* second category*.*

Our convention is that $\cup_i$ always denotes a countable union. We can alternatively characterize the spaces of second category as follows: A nonempty topological space X is of second category if either of the following equivalent statements is true

1. For all $\mathcal{U}_i$ countable collection of open dense sets, $\cap_i \mathcal{U}_i$ is nonempty.
2. If $X = \cup C_i$, where C_i is closed, then there exists an i such that C_i has non-empty interior.

Theorem 5.1. *Let X be a complete metric space. Then X is of second category.*

Proof. Let $\mathcal{U}_i$ be a sequence of open dense sets in X. Choose $x_1 \in \mathcal{U}_1$. By the density of $\mathcal{U}_2$, and openness of $\mathcal{U}_1$, there exists an open ball of radius ≤ 1, such that $\overline{B_{x_1}(\epsilon_1)} \subset \mathcal{U}_1$, and $x_2 \in \overline{B_{x_1}(\epsilon_1)}$ such that $x_2 \in \mathcal{U}_2$. Given now $0 < \epsilon_j \leq j^{-1}$, and $\overline{B_{x_j}(\epsilon_j)} \subset \mathcal{U}_j$, and $\overline{B_{x_j}(\epsilon_j)} \subset \overline{B_{x_{j-1}}(\epsilon_{j-1})}$, for $j = 1, \ldots, i$, we can choose $x_{i+1} \in \mathcal{U}_{i+1} \cap B_{x_i}(\epsilon_i)$, and an ϵ_{i+1} such that $\overline{B_{x_{i+1}}(\epsilon_{i+1})} \subset \overline{B_{x_i}(\epsilon_i)} \cap \mathcal{U}_{i+1}$.

Clearly, x_i is a Cauchy sequence, and converges thus to some x. Moreover, Since $x_j \subset \overline{B_{x_i}(\epsilon_i)}$, for $j \geq i$, and $\overline{B_{x_i}(\epsilon_i)}$ is of course closed, then $x = \lim x_j$ is also contained in $\overline{B_{x_i}(\epsilon_i)}$. Thus, $x \in \mathcal{U}_i$ for all i, so $\cap \mathcal{U}_i \neq \emptyset$. □

5.2 Applications

5.2.1 The existence of irrationals

Proposition 5.1. $\mathbb{R} \setminus \mathbb{Q} \neq \emptyset$.

Proof. $\mathbb{Q}$ is of first category, as it is countable and points are closed with empty interior, yet $\mathbb{R}$ is of second category, as it is complete. □

Note that this is a completely nonconstructive proof. Compare this with the difficulty of constructing a particular number which is irrational.

5.2.2 Nowhere differentiable functions

Proposition 5.2. *There exists a function $f \in C^0([0,1])$ such that f is not differentiable anywhere.*

[6] A set is *nowhere dense* if its closure has empty interior

Proof. Consider the space $C^0([0,1])$. This is a Banach space. For all integer $n \geq 1$, and rationals $p \in \mathbb{Q} \cap [0,1]$, define

$$E_n = \{f \in C^0([0,1]) : \exists p \in [0,1] : |f'(p)| \leq n\}$$

Let F_n denote the closure of E_n. One easily shows that F_n is nowhere dense, as any open set in $C^0([0,1])$ contains functions such that the Lipschitz constant is arbitrarily large at all p, and thus cannot be approximated by members of E_n. Thus, $\cup F_n$ is of the first category. Since C^0 is complete, we must have $C^0 \neq \cup F_n$, i.e. there exists an f which is differentiable nowhere. □

Showing that F_n is nowhere dense may give you an idea for a constructive proof of the statement of the Proposition. Try it.

5.3 Completeness meets linearity: Banach-Steinhaus

The following result is often called the Banach-Steinhaus theorem:

Theorem 5.2. *Let V be a Banach spaces, W a normed vector space, and let T_α be an arbitrary collection of bounded linear maps $T_\alpha : V \to W$, such that for each $x \in V$,*

$$\sup_\alpha |T_\alpha x| < \infty. \tag{11}$$

Then

$$\sup_\alpha ||T_\alpha|| < \infty. \tag{12}$$

Proof. For positive integers n, define $F_n = \{x : \forall_\alpha, |T_\alpha x| \leq n\}$. By the continuity of T_α (Remember Proposition 2.4!), we have that F_n is closed as it is an intersection of closed sets. By our assumption (11),

$$\cup F_n = V.$$

Since V is complete by assumption, and thus of second category by Theorem 5.1, it must be that at least one of the F_n contains an open ball $x_0 + B(\epsilon)$ for an $\epsilon > 0$, i.e. for all α,

$$T_\alpha(x_0 + B(\epsilon)) \subset \overline{B(n)}.$$

But this means that for all α,

$$T_\alpha(B(\epsilon)) \subset T_\alpha x_0 + \overline{B(n)}$$

Let $n_0 = \sup_\alpha |T_\alpha x_0|$. We have then

$$T_\alpha(B(\epsilon)) \subset \overline{B(n_0)} + \overline{B(n)} = \overline{B(n_0 + n)}$$

and thus

$$T_\alpha(B(1)) \subset \epsilon^{-1}\overline{B(n_0 + n)} = \overline{B(\epsilon^{-1}(n_0 + n))}$$

and thus $||T_\alpha|| \leq \epsilon^{-1}(n_0 + n)$, i.e. we have shown (12). □

5.4 Open mapping, inverse mapping, closed graph

Theorem 5.3. *Let V and W be Banach spaces, and let T be a surjective bounded linear map $T : V \to W$. Then T is an open map, i.e. $T(\mathcal{U})$ is open if $\mathcal{U}$ is open.*

Proof. First note the following

Lemma 5.1. *T is open iff $T(B(1)) \supset B(\epsilon)$ for some $\epsilon > 0$.*

Proof. One direction is obvious. For the other, note that if $\mathcal{U}$ is an arbitrary open set, and $q = T(p)$ for $p \in \mathcal{U}$, then there exists a δ such that $p + B(\delta) \subset \mathcal{U}$. But then

$$\begin{aligned} T(\mathcal{U}) &\supset T(p + B(\delta)) \\ &= T(p) + T(B(\delta)) \\ &= q + \delta T(B(1)) \\ &\supset q + B(\delta\epsilon), \end{aligned}$$

and this shows that $T(\mathcal{U})$ contains an open set around an arbitrary element q of it. ∎

So it suffices for us to show that under the assumptions of the Proposition, $T(B(1)) \supset B(\epsilon)$ for some $\epsilon > 0$. We certainly have $V = \cup_{n=1}^{\infty} n(B(1))$. Thus, by surjectivity,

$$\begin{aligned} W &= T\big(\cup_{n=1}^{\infty} nB(1)\big) \\ &= \cup_{n=1}^{\infty} nT(B(1)). \end{aligned}$$

Thus we may certainly write

$$W = \cup_{n=1}^{\infty} n\overline{T(B(1))}.$$

Since we have exhibited W as a countable union of closed sets, and W is of second category by assumption, it follows that there exists an n such that

$$n\overline{T(B(1))} \supset y_0 + B(\delta),$$

so

$$\overline{T(B(1))} \supset n^{-1}y_0 + B(n^{-1}\delta).$$

Now since $B(1) = -B(1)$ (The unit ball is balanced!), we have

$$\overline{T(B(1))} = \overline{T(-B(1))} \supset -n^{-1}y_0 + B(n^{-1}\delta).$$

So

$$\begin{aligned} \overline{T(B(2))} &= \overline{T(B(1))} + \overline{T(B(1))} \\ &= \overline{T(-B(1))} + \overline{T(B(1))} \\ &\supset -n^{-1}y_0 + B(n^{-1}\delta) + n^{-1}y_0 + B(n^{-1}\delta) \\ &= B(2n^{-1}\delta). \end{aligned}$$

By rescaling T, the proof is clearly complete by the following

Lemma 5.2. *Let T be a bounded linear map $T : V \to W$, where V is a Banach space and W a normed vector space, such that $\overline{T(B(1))} \supset B(1)$. Then $T(B(1)) \supset B(1)$.*

Proof. Let $w \in B(1) \subset W$. We have that $w \in B(1-\delta)$ for some $\delta > 0$. By the density of $T(B(\epsilon))$ in $B(\epsilon)$ for all $\epsilon > 0$ (this follows by assumption, and linearity), and the fact that, for any set $X \subset V$, $\overline{X} \subset X + B(\tilde{\epsilon})$, for any $\tilde{\epsilon} > 0$, we have that for all $i \geq 1$,

$$
\begin{aligned}
B(2^{-i}(1-2^{-i}\delta)) &= B(2^{-i-1}(1-2^{-i}\delta)) + B(2^{-i-1}(1-2^{-i}\delta) \\
&= \overline{T(B(2^{-i-1}(1-2^{-i}\delta)))} \cap B(2^{-i-1}(1-2^{-i}\delta)) \\
&\quad + B(2^{-i-1}(1-2^{-i}\delta)) \\
&\subset T(B(2^{-i-1}(1-2^{-i}\delta))) \cap B(2^{-i-1}(1-2^{-i}\delta)) \\
&\quad + B(2^{-2i-2}\delta) + B(2^{-i-1}(1-2^{-i}\delta)) \\
&= T(B(2^{-i-1}(1-2^{-i}\delta))) \cap B(2^{-i-1}(1-2^{-i}\delta)) \\
&\quad + B(2^{-i-1}(1-2^{-i-1}\delta)).
\end{aligned}
$$

By induction, it follows that we may write w as $w = \sum_{i=1}^{\infty} w_i$, where

$$w_i \in T(B(2^{-i}(1-2^{-i+1}\delta))) \cap B(2^{-i}(1-2^{-i+1}\delta)).$$

If $v_i \in B(2^{-i}(1-2^{-i+1}\delta))$ is thus such that $Tv_i = w_i$, then $\sum v_i$ converges to a $v \in B(1)$ since V is a Banach space. By continuity and linearity, $Tv = w$. Thus $w \in T(B(1))$. □

□

Note how the completeness assumptions for V and W enter in very different ways. The completeness of W was only used to infer it is second category. In fact, one can replace the assumptions that W is Banach and T is surjective with the assumption that the image of T is of second category in W. Surjectivity then *follows* after having shown that the map is open.

The following result is known as the *Inverse Mapping Theorem*. It is important in various applications of the theory. It is essentially an immediate corollary of Theorem 5.3.

Theorem 5.4. *Let V and W be Banach spaces, and let $T : V \to W$ be an injective and surjective bounded linear map. Then T^{-1} is bounded.*

Proof. The map T^{-1} exists and is linear. Since by Theorem 5.3, we have

$$T(B(1)) \supset B(\delta),$$

for some $\delta > 0$, it follows that

$$B(1) \supset T^{-1}(B(\delta)),$$

i.e.

$$T^{-1}(B(1)) \subset B(\delta^{-1}),$$

i.e. T^{-1} is bounded with $||T^{-1}|| \leq \delta^{-1}$. □

Another celebrated Corollary of Theorem 5.3 is the so-called *Closed graph theorem.*

Theorem 5.5. *Let V and W be Banach spaces, and let $T : V \to W$ be linear. Then $T : V \to W$ is bounded iff the graph Γ of V is closed as a subset of $V \times W$.*

By the graph of V, we mean the set $\Gamma = \{v, Tv\}_{v\in V} \subset V \times W$.

Proof. Certainly, if T is bounded, then the graph is closed. So it suffices to show the other implication.

Assume then that Γ is closed. As Γ is evidently a linear subspace of the Banach space $V \to W$, it follows that Γ is itself a Banach space.

Consider the map $\phi : \Gamma \to V$ defined by

$$\phi : (v, Tv) \mapsto v.$$

The map ϕ is clearly linear. Moreover, it is bounded, as $|(v, Tv)| = \max\{|v|, |Tv|\}$, and thus $||\phi|| \leq 1$. Finally, the map is manifestly both injective and surjective. It follows from Theorem 5.4 that ϕ^{-1} is a bounded linear map. But this implies that there exists a C such that for all v,

$$|Tv| \leq \max\{|v|, |Tv|\} \leq C|v|,$$

i.e. T is bounded. □

To see the gain apparent from the previous Theorem, consider the sequential characterization of continuity. To show that T is continuous, one has to show that if $v_i \to v$ then $Tv_i \to Tv$. Armed with the above theorem, it suffices to show that if $v_i \to v$ **and if** $Tv_i \to w$, then $w = Tv$.

5.5 More applications

We have already seen some elementary applications of category arguments to show the existence of "exotic" objects in analysis. (If $\mathbb{R} \setminus \mathbb{Q}$ can be thought of as exotic...)

Now that we have a surprising set of theorems which demonstrate the power of mixing category arguments with linearity, we can show the existence of "exotic" objects characeterized by properties of linear maps.

For instance, in the example sheets you will use the theorems of this section to show that there exist (in fact many) continuous functions on $\mathbb{S}^1$ such that their Fourier series diverges at some point.

A sketch of the argument is as follows: One obtains an explicit integral representation

$$S_n f(x) = \int_{-\pi}^{\pi} f(t) \frac{\sin\left(\left(n + \frac{1}{2}\right)(x - t)\right)}{\sin\left((x - t)/2\right)} dt \tag{13}$$

for the n'th partial sum operator

$$S_n : C(\mathbb{S}^1) \to C(\mathbb{S}^1),$$

defined by

$$S_n : f \mapsto \sum_{k=-n}^{n} e^{ikt}\hat{f}(k)$$

where

$$\hat{f}(k) = \frac{1}{2\pi}\int_{-\pi}^{\pi} f(t)e^{-ikt}dt.$$

Let ϕ_n denote the composition of S_n with the map "evaluation at 0." Assuming now that $\forall f$, $\sup_n |\phi_n(f)| < \infty$, one applies Banach-Steinhaus to obtain that $\sup ||\phi_n|| < \infty$. On the other hand, this is easily contradicted by using (13), and choosing an appropriate f for each n.[7] Thus there exists a continuous f such that its Fourier series does not converge at the origin.

Another such argument, also left for the example sheets, is the following. One can show that the Fourier series of an L^1 function on $\mathbb{S}^1$ is an element of c_0, the set of complex functions on $\mathbb{Z}$ that tend to 0 as $|n| \to \infty$. Do all sequences in c_0 arise like this?

This would mean that the map $\Lambda : L^1 \to c_0$ defined by taking a function to its Fourier series is surjective. One first shows that this is an injective bounded linear map. Were it surjective, its inverse would be bounded in view of the inverse mapping theorem, Theorem 5.4. One obtains a contradiction by demonstrating a sequence of L^1 functions whose L^1 norm goes to infinity, while the sup norm of their Fourier coefficients remains bounded. Thus the answer is no.

One should make reference at this point to other results in analysis which are beyond the scope of this class. If $f \in C^1(\mathbb{S}^1)$, then $S_n f$ converges everywhere. This is a 19th century theorem. On the other hand, for an L^p function on $\mathbb{S}^1$, with $p > 1$ (in particular for continuous functions) the Fourier series converges almost everywhere (i.e. the set where it doesn't has measure 0). For $p = 2$, this is the celebrated Carleson's Theorem. This theorem is hard.

Finally, a classic result of Kolmogorov shows that there exists a function in $L^1(\mathbb{S}^1)$ such that the Fourier series diverges everywhere!

6 The topology of $C(K)$

The space of continuous functions on a compact Hausdorff space plays a fundamental role in functional analysis. In this section, we will try to understand better how big this space is, and what is its topology. In particular, we will be interested in constructing (Section 6.1) a rich class of functions in $C(K)$, identifying the compact subsets of $C(K)$ (Section 6.2), as well as useful dense subsets (Section 6.4).

An application of density and compactness results for solving an actual problem in analysis is given in Section 6.3.

[7] Why is this less work than a constructive proof? Because the f you choose in showing $\sup ||S_n|| \to \infty$ can depend on n. The convergence of these f is irrelevant.

6.1 The extension theorem

Our first task will be to show that the space $C(K)$ is sufficiently rich. Specifically, we shall show that functions on $C(K)$ can be constructed by extending functions on closed subsets. Compare with the Hahn-Banach Theorem, Theorem 4.1. In particular, $C(K)$ separates points (Corollary 6.1).

6.1.1 compact Hausdorff $\implies$ normal

Recall the following

Definition 6.1. *A topological space X is said to be* Hausdorff*, when given $p \neq q$ in X, there exist open neighborhoods $\mathcal{U}_p$ of p and $\mathcal{U}_q$ of q, such that $\mathcal{U}_p \cap \mathcal{U}_q = \emptyset$.*

On the other hand, we have

Definition 6.2. *A topological space X is said to be* normal *if given C_1, C_2 closed with $C_1 \cap C_2 = \emptyset$, then there exist open $\mathcal{U}_1 \supset C_1$, $\mathcal{U}_2 \supset C_2$ such that $\mathcal{U}_1 \cap \mathcal{U}_2 = \emptyset$.*

An alternative characterization of normality is: X is normal iff for all $C_1 \subset \mathcal{U}_2$, with C_1 closed and $\mathcal{U}_2$ open, there exist $\mathcal{U}_1$, C_2, open, closed, respectively, such that $C_1 \subset \mathcal{U}_1 \subset C_2 \subset \mathcal{U}_2$. This follows by considering the complement of $\mathcal{U}_2$, etc.

It is the normality condition that will be very useful for extending continuous functions. Thus, it's nice to know the following

Proposition 6.1. *Let K be a compact Hausdorff space. Then K is normal.*

Proof. Let C_1, C_2 be closed subsets of K. As closed subsets of a compact set, they are compact. Given any $p \in C_1$, $q \in C_2$, let $p \in \mathcal{U}_{p,q}$, $q \in \mathcal{V}_{p,q}$ be open neighborhoods such that $\mathcal{U}_{p,q} \cap \mathcal{V}_{p,q} = \emptyset$. Fix q say. Then $\{\mathcal{U}_{p.q}\}_{p \in C_1}$ forms an open cover of C_1, and thus there exists a finite subcover $\{\mathcal{U}_{p_i,q}\}_{i=1\ldots n}$. Let us define

$$\mathcal{U}_q = \cup_{i=1}^n \mathcal{U}_{p_i,q},$$
$$\mathcal{V}_q = \cap_{i=1}^n \mathcal{V}_{p_i,q}.$$

Clearly $\mathcal{U}_q \cap \mathcal{V}_q = \emptyset$, and $\mathcal{U}_q \supset C_1$.

Now repeat this construction for arbitrary q, and consider the resulting sets $\{\mathcal{U}_q\}_{q \in C_2}$ and $\{\mathcal{V}_q\}_{q \in C_2}$. The latter forms an open cover of C_2, and thus, by compactness, there exits a finite subcover $\{\mathcal{V}_{q_i}\}_{i=1\ldots m}$. Define

$$\mathcal{V} = \cup_{i=1}^m \mathcal{V}_{q_i}$$
$$\mathcal{U} = \cap_{i=1}^m \mathcal{U}_{q_i}.$$

These sets are clearly open. By definition of a cover, $\mathcal{V} \supset C_2$, while clearly $\mathcal{V} \cap \mathcal{U} = \emptyset$. On the other hand, since $\mathcal{U}_{q_i} \supset C_1$ for all i, it follows that $\mathcal{U} \supset C_1$. ◻

Note that the compactness of K was only used to assert the compactness of the C_i. It follows thus from the proof that in any Hausdorff space X, one can separate compact sets by open sets.

6.1.2 Urysohn's lemma

We begin with the following

Lemma 6.1. *Let X be a normal space, and let C_0, C_1 be disjoint closed sets. Then there exists an $f \in C(X)$ with range $[0,1]$ such that $f = 0$ on C_0 and $f = 1$ on C_1.*

Proof. Enumerate the rationals of $\mathbb{Q} \cap [0,1]$ as $\{q_i\}_{i=0}^{\infty}$, where $q_0 = 0$, $q_1 = 1$. Define inductively a collection of open and closed sets $\mathcal{U}_i \subset C_i$ as follows: Let $\mathcal{U}_0 = \emptyset$, $F_0 = C_0$, let $\mathcal{U}_1$ be $X \setminus C_1$ and $F_1 = X$. Given $\mathcal{U}_i$, F_i, for $0 \leq i \leq n$, with the property that $\mathcal{U}_i \subset F_i$, and, if $q_i < q_j$, then $F_i \subset \mathcal{U}_j$, there exists a unique interval (q_{i_1}, q_{i_2}) containing q_{n+1}, with $0 \leq i_1, i_2 \leq n$, and no other $q_j \in (q_{i_1}, q_{i_2})$ for $j = 0, \ldots n$. We have by normality that there exists $\mathcal{U}_{n+1} \subset F_{n+1}$ such that $F_{i_1} \subset \mathcal{U}_{n+1} \subset F_{n+1} \subset \mathcal{U}_{i_2}$. It now follows that this holds for all $q_i < q_j$ for $i = 0 \ldots n+1$, and thus, by induction[8], one defines such a sequence for all $i = 0 \ldots \infty$.

Now define

$$f(x) = \inf_{n=0\ldots\infty} \{q_n : x \in F_n\}.$$

Certainly, if $x \in C_0$, then $x \in F_0$, so $f(x) = 0$. On the other hand, if $x \in C_1$, then $x \notin \mathcal{U}_1$, so $x \notin F_j$ for $q_j < 1$, so $f(x) = 1$.

Claim: f is both lower and upper semicontinuous, and thus continuous. For, given α, $\{x : f(x) > \alpha\}$ is clearly open. For if y is in this set, this implies that there exists a $q_n > \alpha$ such that $y \notin F_n$. Since the complement of F_n is open, this implies that there exists a neighborhood of y which does not intersect F_n. But this means that $f \geq q_n > \alpha$ on this neighborhood. Thus $\{x : f(x) > \alpha\}$ is open.

On the other hand, $\{x : f(x) < \alpha\}$ is also open. For let y be in this set. This means that there exist $q_n < q_m < \alpha$ such that $y \in F_n$. But this means that $y \in \mathcal{U}_m$. Now since $\mathcal{U}_m$ is open there is a neighborhood of y which is contained in $\mathcal{U}_m$. But now, since $\mathcal{U}_m \subset F_m$, it follows that this neighborhood is contained in F_m, i.e. $f \leq q_m < \alpha$ on this neighhorhood. Thus $\{x : f(x) < \alpha\}$ is open. We have thus shown continuity, as desired. □

In view of Proposition 6.1, the above result applies to compact Hausdorff spaces. In particular,

Corollary 6.1. *Let K be compact Hausdorff. Then $C(K)$ separates points, i.e. given $p \neq q$ in K, there exists an $f \in C(K)$ such that $f(p) \neq f(q)$.*

Proof. It suffices to remark that points are closed sets in a Hausdorff space. □

[8]Note that to have an interval (q_{i_1}, q_{i_2}), one needs $n > 1$. Thus the importance of defining two base cases (0 and 1) for the induction.

6.1.3 The Tietze-Urysohn extension theorem

Finally,

Theorem 6.1. *Let X be normal, and $f : C \to \mathbb{C}$ a bounded continuous function on a closed subset C. Then there exists a continuous extension $\tilde{f} : X \to \mathbb{C}$, with $\tilde{f}|_C = f$, and $|\tilde{f}| = |f|$, where $|\cdot|$ denotes the sup norm.*

Note that Lemma 6.1 is a special case of Theorem 6.1 where the range of f consists of 2 points.

Proof. Clearly, by taking real and imaginary parts, translating and rescaling, it suffices to consider the case where the range of f is $[0, 1]$. Also, one need not worry about the condition $|\tilde{f}| = |f|$. For given $\hat{f} : X \to \mathbb{C}$ any continuous extension of f, we may define $\tilde{f}$ by $\tilde{f}(x) = \hat{f}(x)$, if $|\hat{f}(x)| \leq |f|$, $\tilde{f}(x) = e^{i(\arg \hat{f}(x))}|f|$ otherwise, and $\tilde{f}$ is again a continuous extension of f with $|\tilde{f}| = |f|$.

Define a sequence of closed sets and functions by induction as follows. Let $f_0 = f$, $C_0 = f^{-1}\big([0, \frac{1}{3}]\big)$, $F_0 = f^{-1}\big([\frac{2}{3}, 1]\big)$, and let $g_0 : X \to [0, \frac{1}{3}]$ be a function taking the value 0 on C_0 and $1/3$ on F_0, obtained by applying the previous Proposition, and let $f_1 = f - g_0|_C$ be a function $f_1 : C \to \mathbb{R}$. Note that $0 \leq f_1 \leq \frac{2}{3}$, i.e.

$$f_1 : C \to \big[0, \frac{2}{3}\big].$$

Now, given $f_i : C \to \big[0, \big(\frac{2}{3}\big)^i\big]$, define

$$C_i = f_i^{-1}\big([0, \frac{1}{3}\big(\frac{2}{3}\big)^i]\big),$$

and

$$F_i = f_i^{-1}\big([\frac{2}{3}\big(\frac{2}{3}\big)^i, \big(\frac{2}{3}\big)^i]\big),$$

let g_i be a function retrieving 0 on C_i and $\frac{1}{3}\big(\frac{2}{3}\big)^i$ on F_i, given by Lemma 6.1. Set

$$f_{i+1} = f_i - g_i|_C. \tag{14}$$

Clearly $0 \leq f_{i+1} \leq \big(\frac{2}{3}\big)^{i+1}$. We thus have defined inductively functions

$$g_i : X \to \big[0, \frac{1}{3}\big(\frac{2}{3}\big)^i\big].$$

$$f_{i+1} : C \to \big[0, \big(\frac{2}{3}\big)^{i+1}\big].$$

Clearly, from (14), we obtain

$$\sum_{i=0}^{\infty} g_i|_C = f_0 = f.$$

Setting $\tilde{f} = \sum_{i=0}^{\infty} g_i$, we have $\tilde{f} \in C(X)$, since

$$\sum_{i=n}^{m} |g_i| \leq (2/3)^n;$$

we're done. ◻

6.2 Compactness: Arzela-Ascoli

Now we shall find a characterization for the compact subsets of $C(K)$. The following definition will be useful

Definition 6.3. *Let X be a metric space and E a subset, and $\epsilon > 0$. A set $N \subset X$ is said to be an* ϵ-net *for E if*

$$E \subset \cup_{x \in N} B_\epsilon(x).$$

Definition 6.4. *Let X be a metric space, and E a subset. E is said to be* totally bounded *if for all $\epsilon > 0$ there exists a finite ϵ-net N for E in X.*

Note that a subset S of a totally bounded set E is again totally bounded, and an ϵ-net for E is an ϵ-net for S. If E is totally bounded than so is $\overline{E}$. Also note that totally bounded sets are certainly bounded. Finally, note that if $N \subset X$ is an ϵ-net for E, then there exists a 2ϵ net $\tilde{N} \subset E$ for E. Thus in the above definition, we can equivalently require that $N \subset E$.

We have the following

Proposition 6.2. *A set E is totally bounded iff for every sequence $y_i \in E$, there exists a subsequence which is Cauchy.*

Proof. Suppose E is totally bounded and let y_i be a sequence in E. Let N_j denote an $1/j$ net for E, which exists by assumption. For $j = 1$, there is an $x_1 \in N_1$ such that $B_1(x_1)$ that contains infinitely many of the y_i. Let i_1 be the smallest such i. Now given $y_{i_k} \in \cap_{j=1}^{k} B_{1/j}(x_j)$ for all $1 \leq k \leq n$, such that $x_j \in N_j$ and $B_{1/j}(x_j)$ contains infinitely many $\{y_i\} \cap B_{1/m}(x_m)$ for $m < j$, let x_{n+1} be such that $B_{1/(n+1)}(x_{n+1})$ contains infinitely many of the $\{y_i\} \cap \bigcap_{j=1}^{n} B_{1/j}(x_j)$, and let i_{n+1} be the first $i \geq i_k$ such that $y_{i_{n+1}} \in \cap_{j=1}^{n+1} B_{1/j}(x_j)$. For $n < m$, we have

$$\begin{aligned} d(y_{i_n}, y_{i_m}) &\leq d(y_{i_n}, x_n) + d(x_n, y_{i_m}) \\ &\leq 2/n. \end{aligned}$$

Thus y_{i_k} is Cauchy.

Now suppose E is not totally bounded. Let $\epsilon > 0$ be such that there does not exist a finite ϵ-net. Choose $y_1 \in E$ at random. Having chosen $y_1, \ldots, y_n \in E$ with the property that $d(y_i, y_j) > \epsilon$ for $1 \leq i \leq j \leq n$, since $E \not\subset \cup_{i=1}^{n} B_\epsilon(y_i)$, there exists a $y_{i+1} \in E \setminus \cup_{i=1}^{n} B_\epsilon(y_i)$. By induction one obtains a sequence $\{y_i\}$ such that $d(y_i, y_j) > \epsilon$ for all i, j. No subsequence of y_i can be Cauchy. ◻

Corollary 6.2. *Let X be a complete metric space. A set E is totally bounded iff $\overline{E}$ is compact.*

Thus, since $C(K)$ is a Banach space we are more than happy to classify totally bounded sets.

We have

Definition 6.5. *A subset $\mathcal{F} \subset C(K)$ is called* equicontinuous at $x \in K$ *if for every $\epsilon > 0$ there exists a neighborhood $\mathcal{U}$ of x such that for $y \in \mathcal{U}$, $|f(y)-f(x)| < \epsilon$ for all $f \in \mathcal{F}$. We say that $\mathcal{F}$ is* equicontinuous *if it is equicontinuous at x for all $x \in K$.*

Note than finite subsets of $C(K)$ are clearly equicontinuous.

The Arzela-Ascoli theorem states

Theorem 6.2. *Let K be compact Hausdorff. Then $\mathcal{F} \subset C(K)$ is totally bounded iff it is bounded and equicontinuous.*

Proof. Suppose $\mathcal{F}$ is totally bounded. It is certainly bounded. Given $\epsilon > 0$, let $\{f_i\}_{i=1}^n$ be an ϵ-net for $\mathcal{F}$. Given $x \in K$, then $\mathcal{U}_i$ be subsets so that $|f_i(y) - f_i(x)| < \epsilon$. Define $\mathcal{U} = \cap_i \mathcal{U}_i$. We have for $y \in \mathcal{U}$,

$$|f(y) - f(x)| \leq |f(y) - f_i(y)| + |f_i(y) - f_i(x)| + |f_i(x) - f(x)|.$$

Since $\{f_i\}$ forms an ϵ-net, there exists an i such that $|f - f_i| < \epsilon$. For this i, we have then

$$|f(y) - f(x)| \leq \epsilon + |f_i(y) - f_i(x)| + \epsilon \leq 3\epsilon.$$

We have shown that $\mathcal{F}$ is equicontinuous at x.

Conversely, suppose $\mathcal{F}$ is bounded and equicontinuous. Given $\epsilon > 0$, x, let $\mathcal{U}_x$ denote an open set around x such that $|f(y) - f(x)| < \epsilon$ for all $y \in \mathcal{U}_x$. The collection $\{\mathcal{U}_x\}$ forms an open cover for K, and thus there exists a finite subcover $\{\mathcal{U}_{x_i}\}_{i=1}^n$.

Consider $\mathcal{F}|_{\{x_i\}}$ as a subset of l_∞^n. By assumption, this is a bounded subset, and thus, since we are in l_∞^n, totally bounded. There thus exists an ϵ-net $\{f_j|_{\{x_i\}}\}_{j=1}^m \subset \mathcal{F}|_{\{x_i\}}$ for this subset. For any j, we have

$$\begin{aligned}
|f - f_j| &= \max_i \sup_{y \in \mathcal{U}_i} |f(y) - f_j(y)| \\
&\leq \max_i \sup_{y \in \mathcal{U}_i} |f(y) - f(x_i)| + |f(x_i) - f_j(x_i)| + |f_j(x_i) - f_j(y)| \\
&\leq \max_i (\epsilon + |f(x_i) - f_j(x_i)| + \epsilon) \\
&= \left|(f - f_j)|_{\{x_i\}}\right|_\infty + 2\epsilon.
\end{aligned}$$

Since $\{f_j|_{\{x_i\}}\}_{j=1}^m \subset \mathcal{F}|_{\{x_i\}}$ is an ϵ-net, there exists a j such that $|(f - f_j)|_{\{x_i\}}|_\infty < \epsilon$, thus, for this j,

$$|f - f_j| < 3\epsilon.$$

The functions f_j thus constitute a 3ϵ-net for $\mathcal{F}$ in $C(K)$. Thus $\mathcal{F}$ is totally bounded. ∎

Note that examining the above proof, it is clearly enough to assume that the family $\mathcal{F}$ is pointwise bounded, that is, for all x, $\sup_{f\in\mathcal{F}}|f(x)|<\infty$.

Example. Consider $K=\overline{\mathcal{U}}$, where $\mathcal{U}$ is a bounded open subset in $\mathbb{R}^n$. Let $B^1(R)$ denote the open ball of radius $R>0$ in $C^1(\overline{\mathcal{U}})$. Then $B^1(R)\subset C^0(\overline{\mathcal{U}})$ is equicontinuous.

6.3 Aside: The role of compactness results in analysis

What are compactness theorems in function spaces good for? We give some applications. One should think of these as at the heart of the theory of differential equations, ordinary and partial.

A pedestrian way of thinking about the role of compactness is as follows: Suppose you have a problem where the unknown is a function f, and you think you have a good way of approximating solutions to this problem, call these approximate solutions f_i. Then if these $f_i\in\mathcal{B}$, where $\mathcal{B}$ is compact, it follows that one can draw a convergent subsequence $f_{i_k}\to f$ in the relevant space. The function f is a good candidate for a solution to the problem.

We will proceed in this section to illustrate this idea in proving the existence of solutions to o.d.e.'s with low regularity.

You are probably already familiar with the fundamental existence and uniqueness theorem for o.d.e.'s, which states

Theorem 6.3. *Let $f:\mathbb{R}\to\mathbb{R}$ be locally Lipschitz. Then for any t_0, x_0, there exists a unique interval (T_-,T_+) such that $-\infty\le T_-<t_0<T_+\le\infty$ and a unique C^1 function $x:(T_-,T_+)\to\mathbb{R}$ satisfying the initial value problem*

$$x'=f(x),\qquad x(t_0)=x_0,$$

such that x is not the restriction of an $\tilde{x}$ satisfying the above on a larger interval. Moreover, let $K\subset\mathbb{R}$ be compact. Then if $T_\pm\neq\infty$ there exist $t_+<T_+$, $t_->T_-$, respectively, such that such that $x([t_+,T_+))\cap K=\emptyset$, $x((T_-,t_-])\cap K=\emptyset$, respectively.

This theorem can be proven with the so-called Banach fixed point theorem or with Picard iteration. On the other hand, what happens when we weaken the Lipschitz condition to continuity? In this section, we will prove the following

Theorem 6.4. *Let $f:\mathbb{R}\to\mathbb{R}$ be continuous. Then for any t_0, x_0, there exists an $\epsilon>0$ and a function $x:(t_0-\epsilon,t_0+\epsilon)\to\mathbb{R}$ satisfying the initial value problem*

$$x'=f(x),\qquad x(t_0)=x_0. \tag{15}$$

That is to say, we still have existence, but we have lost uniqueness.

Example. Consider the IVP: $x'=2|x|^{1/2}$, $x(0)=0$. One solution is $x=0$, and an other is $x=0$ for $t\le0$, and $x=t^2$ for $t\ge0$. What are all solutions?

Proof. We shall prove Theorem 6.4 from Theorem 6.3. We shall make fundamental use of Arzela-Ascoli.

Choose $\delta > 0$, and let $F = \sup_{|x-x_0|\le\delta} |f(x)|$. Let $\epsilon > 0$ be such that $\epsilon(F+1) \le \delta/2$. Let B denote the set $\{f + \tilde{g}\}$ where $\tilde{g}$ ranges over functions $\mathbb{R} \to \mathbb{R}$ with supremum less than or equal to 1.

Lemma 6.2. *If $x : [t_0 - \epsilon_-, t_0 + \epsilon_+] \to \mathbb{R}$ solves (15), for $0 < \epsilon_\pm \le \epsilon$, with $\tilde{f} \in B$ in place of f, then*

$$|x - x_0| < \delta, \tag{16}$$

$$|x'| \le F + 1. \tag{17}$$

The above lemma is an example of an *a priori* estimate.

Proof. For such an x, we have

$$x' = \tilde{f}(x)$$

for a $\tilde{f}$ in B. Thus, in view of the definition of B and F, (17) would follow immediately from (16).

Consider the subset $t \in [t_0, t_0 + \epsilon]$ such that (16) holds for all $\bar{t} \in [2t_0 - t, t] \cap [t_0 - \epsilon_-, t_0 + \epsilon_+]$. Call this subset $\mathcal{T}$. The set $\mathcal{T}$ is clearly open in $[t_0, t_0 + \epsilon]$, by continuity of x. Moreover, $t_0 \in \mathcal{T}$, and thus $\mathcal{T} \neq \emptyset$. On the other hand, for $t \in \mathcal{T}$, we have

$$\begin{aligned} \sup_{[2t_0-t,t]\cap[t_0-\epsilon_-,t_0+\epsilon_+]} |\tilde{f}(x(\bar{t}))| &\le \sup_{|x-x_0|\le\delta} |f(x)| + 1 \\ &\le F + 1. \end{aligned}$$

Thus, for $\bar{t} \in [2t_0 - t, t] \cap [t_0 - \epsilon_-, t_0 + \epsilon_+]$,

$$\begin{aligned} |x(\bar{t}) - x_0| &= \left| \int_{t_0}^{\bar{t}} x'(\hat{t}) d\hat{t} \right| \\ &= \left| \int_{t_0}^{\bar{t}} f(x(\hat{t})) d\hat{t} \right| \\ &\le (F+1) \left| \int_{t_0}^{\bar{t}} d\hat{t} \right| \\ &\le (F+1)\epsilon \\ &\le \delta/2. \end{aligned}$$

By continuity of x, this implies that $\mathcal{T} \subset [t_0, t_0 + \epsilon]$ is closed. By connectedness of $[t_0, t_0 + \epsilon]$, it follows that $\mathcal{T} = [t_0, t_0 + \epsilon]$. That is to say, we have shown that (16) holds for all $t \in [t_0 - \epsilon_-, t_0 + \epsilon_+]$. $\square$

Now consider, $\tilde{f} \in B \cap C^1$, and let $x : (T_-, T_+) \to \mathbb{R}$ be the solution of Theorem 6.3. In view of Lemma 6.2, we have that $T_- < t_0 - \epsilon < t_0 + \epsilon < T_+$. Thus, we may restrict x to a function $x : [t_0 - \epsilon, t_0 + \epsilon]$.

Consider the subset $S \subset C^1([t_0 - \epsilon, t_0 + \epsilon])$ which consists of solutions to (15) on the interval $[t_0 - \epsilon, t_0 + \epsilon]$, with f replaced by $\tilde{f} \in B$, restricted to $[x_0 - \delta, x_0 + \delta]$. We have just shown that this set is nonempty, and it contains a unique function corresponding to any $\tilde{f} \in B \cap C^1([x_0 - \delta, x_0 + \delta])$.

Now we show

Lemma 6.3. *$S \subset C([t_0 - \epsilon, t_0 + \epsilon])$ is totally bounded.*

Proof. In view of Arzela-Ascoli, it suffices to show the uniform boundedness and equicontinuity of S. This is immediate from (16) and (17). □

On the other hand, we have

Lemma 6.4. *$C^1([x_0 - \delta, x_0 + \delta]) \subset C([x_0 - \delta, x_0 + \delta])$ is dense.*

Proof. Omitted. This will follow in particular from the next section. In the meantime, see if you can come up with a direct proof. □

In particular, consider a sequence $f_i \to f$ with $f_i \in B \cap C^1([x_0 - \delta, x_0 + \delta])$. By density, such a sequence exists. Let $x_i \in S$ denote the corresponding solution to (15) with f_i. By Lemma 6.3, there exists a subsequence x_{i_k} such that $x_{i_k} \to x$ for some $x \in C^0([t_0 - \epsilon, t_0 + \epsilon])$. The function x is a good candidate for a solution of (15)! But we have to be careful; we're not done yet. We still must deduce that x actually solves (15). (At this point, we don't even know yet that x is differentiable!)

This doesn't turn out to be so difficult. Since $f_{i_k} \to f$ uniformly, and $x_{i_k} \to x$ uniformly, it follows that $f_{i_k}(x_{i_k})(t) \to f(x(t))$ uniformly in t. Since $x'_{i_k} = f_{i_k}(x_{i_k})$, this implies by an elementary result in real analysis that x' exists and equals $\lim x'_{i_k}$, i.e. $f(x)$. □

The importance of these ideas is even greater when one passes from ordinary differential equations to partial. Such applications would take us, however, too far afield.

6.4 Dense subsets in $C(K)$: Stone-Weierstrass

We have already seen in the last section the importance of having good dense subsets of a relevant function space. See Lemma 6.4. In this section, we shall prove a general theorem that will allow us to come up with many useful dense subsets of $C(K)$. Here two algebraic notions will play a crucial role, that of a lattice and that of an algebra.

Definition 6.6. *Let $(V, +)$ be a vector space over $\mathbb{R}$ or over $\mathbb{C}$, and let $\cdot : V \times V \to V$ be a binary operation. We say that $(V, +, \cdot)$ is an* algebra *(over $\mathbb{R}$ of $\mathbb{C}$) if $(V, +, \cdot)$ is a rng*[9]*, and $\lambda(v \cdot w) = v \cdot (\lambda w) = (\lambda v) \cdot w$, for scalars λ. If V is a normed vector space and $|v \cdot w| \leq |v||w|$, we say that V is a* normed algebra*. If V is in addition Banach, we say that V is a* Banach algebra*. If V*

[9] ring without necessarily multiplicative identity

is commutative as a ring, we say that V is a commutative algebra. *If V has a multiplicative identity, we say V is* unitial.

Examples. The space $C(K)$ is certainly a unitial commutative Banach algebra under the usual multiplication of functions. $\mathcal{L}(V, V)$ is an algebra under composition of linear transformations. If V is a normed vector space, then $\mathcal{B}(V, V) \subset \mathcal{L}(V, V)$ is a unitial subalgebra which is a normed algebra, which if V is a Banach space, is a unitial Banach algebra.

On the other hand

Definition 6.7. *A* lattice *is a partially ordered set $\mathcal{L}$ such that any two element subset has a least upper bound and a greatest lower bound.*

If p and q are two elements of $\mathcal{L}$, we may denote the least upper bound and greatest lower bound by $p \vee q$ and $p \wedge q$, respectively. We may give equivalently an algebraic characterization of the notion of lattice in terms of a set $\mathcal{L}$ and two binary operations $\vee$ and $\wedge$ satisfying a collection of axioms. What are the axioms?

The vector space $\mathcal{F}_{\mathbb{R}}(X)$, the **real** valued functions on a topological space X, is a lattice where $\mathcal{F}_{\mathbb{R}}(X)$ is given the obvious partial ordering $f \leq g \iff \forall_x f(x) \leq g(x)$. We have then

$$(f \vee g)(x) = \max\{f(x), g(x)\}, \qquad (f \wedge g)(x) = \min\{f(x), g(x)\}.$$

The space of continuous functions $C_{\mathbb{R}}(X) \subset \mathcal{F}_{\mathbb{R}}(X)$ is easily seen to be closed under the lattice operations, i.e. it is a *sublattice* of $\mathcal{F}_{\mathbb{R}}(X)$ and thus itself a lattice.

In what follows for now, we will be considering $C_{\mathbb{R}}(K)$, where K is compact Hausdorff. The main theorem of this section, known as the Stone-Weierstrass Theorem, is the following

Theorem 6.5. *If $\mathcal{A} \subset C_{\mathbb{R}}(K)$ is a subalgebra such that $\mathcal{A}$ separates points, then $\overline{\mathcal{A}} = C(K)$, or there exists an $x \in K$ such that $\overline{\mathcal{A}} = \{f \in C(K) : f(x) = 0\}$.*

Example. Let $\mathcal{A}$ denote the set of polynomials in $x_1 \ldots x_n$, thought of as functions on $\overline{\mathcal{U}}$, where $\mathcal{U}$ is a bounded open subset of $\mathbb{R}^n$. The set $\mathcal{A}$ is easily seen to be an algebra separating points, and since $1 \in \mathcal{A}$, the second possibility of the theorem is excluded. Thus $\overline{\mathcal{A}} = C(\overline{\mathcal{U}})$. This is the classical *Weierstrass theorem.* Note that since $\mathcal{A} \subset C^k(\overline{\mathcal{U}})$ for any $k \geq 0$, we have in particular $\overline{C^k(\overline{\mathcal{U}})} = C(\overline{\mathcal{U}})$.

Proof. The importance of real-valued functions is precisely that we may exploit the lattice property. We have the following

Lemma 6.5. *Suppose $\mathcal{L} \subset C_{\mathbb{R}}(K)$ is a sublattice, and suppose that $g \in C_{\mathbb{R}}(K)$ is such that for each $x, y \in K$, there exists an $f \in \mathcal{L}$ such that $|f(x) - g(x)| < \epsilon$, $|f(y) - g(y)| < \epsilon$. Then $g \in \overline{\mathcal{L}}$. In particular, if this assumption is true for all $g \in C_{\mathbb{R}}(K)$, then $\overline{\mathcal{L}} = C_{\mathbb{R}}(K)$.*

Proof. Let $g(x) \in C_{\mathbb{R}}(K)$ be as in the assumption of the lemma. Given $\epsilon > 0$, we will produce an $f \in \mathcal{L}$ such that $|f - g| < \epsilon$.

Pick an $x \in K$, and for each $y \in K$, let $f_{x,y}$ be a function with

$$|f_{x,y}(x) - g(x)| < \epsilon, \qquad |f_{x,y}(y) - g(y)| < \epsilon.$$

By continuity of $f_{x,y}$, g, the inequality $|f_{x,y} - g| < \epsilon$ must hold in open sets around x, y. Call these open sets $\mathcal{V}_{x,y}$, $\mathcal{U}_{x,y}$. We have that $\{\mathcal{U}_{x,y}\}$ is an open cover for K, and thus there exists a finite subcover $\mathcal{U}_{x,y_i}$. If we consider $\mathcal{V}_x = \cap_{i=1}^{n} \mathcal{V}_{x,y_i}$, and

$$f_x = f_{x,y_1} \wedge \cdots \wedge f_{x,y_n},$$

we have

$$f_x(y) < \epsilon + g(y) \tag{18}$$

for all y, and

$$g - \epsilon < f_x \tag{19}$$

in $\mathcal{V}_x$. By the defintion of a sublattice, $f_x \in \mathcal{L}$. Now, consider the open cover $\mathcal{V}_x$ for K. By compactness of K there exists a finite subcover $\mathcal{V}_{x_j}$. Define

$$f = f_{x_1} \vee \cdots \vee f_{x_m}.$$

Again, since $\mathcal{L}$ is closed under the lattice operations, $f \in \mathcal{L}$. We have

$$f(y) < \epsilon + g(y)$$

on account of (18), and

$$g(y) - \epsilon < f(y)$$

on account of (19), by definition of $\vee$ and the fact that $\mathcal{V}_x$ is a cover. That is to say, we have, for all y,

$$|f(y) - g(y)| < \epsilon,$$

as required. ☐

What has all this to do with subalgebras, you will say. The answer is given by the following lemma.

Lemma 6.6. *Let $\mathcal{A} \subset C_{\mathbb{R}}(K)$ be a subalgebra, closed in the topology of $C_{\mathbb{R}}(K)$. Then $\mathcal{A}$ is a sublattice of $C_{\mathbb{R}}(K)$.*

Proof. We note that $\max\{f(x), g(x)\} = \frac{1}{2}(f(x) + g(x) + |f(x) - g(x)|)$, and similarly for min. Thus, it suffices to show that if $f \in \mathcal{A}$, so is $x \mapsto |f(x)|$. Moreover, it suffices to show this for f satisfying, say, $|f| \leq 1$. (Why?)

Now, consider the function $g_\epsilon(x) = \sqrt{\epsilon^2 + x}$. For $-1 \leq x \leq 1$, we have that $|g_\epsilon(x^2) - |x|| \leq \epsilon$.

The function g_ϵ is analytic in a neighborhood of $t = \frac{1}{2}$ and equal to its power series in $[0, 1]$, which moreover, converges uniformly. Write

$$g_\epsilon(x) = c_0 + c_1 \left(x - \frac{1}{2}\right) + c_2 \left(x - \frac{1}{2}\right)^2 + \ldots$$

and let S_n denote the partial sum $\sum_{i=0}^{n} c_i(x-\frac{1}{2})^i$, and let $\tilde{S}_n = S_n - S_n(0)$. We have that $\tilde{S}_n$ is a polynomial without constant term. On the other hand, $S_n(0) \to \epsilon$, as $n \to \infty$, thus, for $n \geq N$, we have $|S_n(0)| < 2\epsilon$.

We have that $\tilde{S}_n \circ f^2 \in \mathcal{A}$, as $\mathcal{A}$ is an algebra, and $\tilde{S}_n$ has no constant term. Since $|f| \leq 1$ implies that $0 \leq f^2 \leq 1$, we have that, given $\epsilon > 0$, there exists an $\tilde{N}$ such that, for $n \geq \tilde{N}$,

$$\begin{aligned} |\tilde{S}_n \circ f^2 - |f|| &\leq |\tilde{S}_n \circ f^2 - g_\epsilon \circ f^2| + |g_\epsilon \circ f^2 - |f|| \\ &\leq |\tilde{S}_n \circ f^2 - S_n \circ f^2| + |S_n \circ f^2 - g_\epsilon \circ f^2| \\ &\quad + |g_\epsilon \circ f^2 - |f|| \\ &< 2\epsilon + \epsilon + \epsilon = 4\epsilon. \end{aligned}$$

(In the above centred formula, $|f|$ denotes the function $x \mapsto |f(x)|$, not the sup of f.) Thus, since $\mathcal{A}$ is assumed closed, and $\epsilon > 0$ is arbitrary, we are done. $\square$

Aside. One can remark that in the above lemma we do not use continuity. We could replace $C_{\mathbb{R}}(K)$ in the statement with the subspace $\mathcal{F}^b_{\mathbb{R}}(K) \subset \mathcal{F}_{\mathbb{R}}(K)$ of bounded functions, given the supremum norm. Compare with Lemma 6.5 where continuity is used in a fundamental way.

To complete the proof of Theorem 6.5, there is very little to say. Consider $\overline{\mathcal{A}}$. Note that this is again an algebra by the continuity of the multiplication, addition, etc. Thus, by Lemma 6.6, $\overline{\mathcal{A}}$ is a sublattice of $C(K)$.

Let us suppose that for all $x \in K$, there exists an $f \in \overline{\mathcal{A}}$ such that $f(x) \neq 0$. For $x \neq y$, let f_x and be a function such that $f_x(x) \neq 0$, let f_y be a function such that $f_y(y) \neq 0$, and let $f_{x,y}$ be a function such that $f_{x,y}(x) \neq f_{x,y}(y)$. By defining $\tilde{f} = f_x + \alpha f_{x,y} + \beta f_y$, for some $\alpha, \beta \in \mathbb{R}$, we obtain a function such that $\tilde{f}(x) \neq 0$, $\tilde{f}(y) \neq 0$, and $\tilde{f}(x) \neq \tilde{f}(y)$. It follows that $(\tilde{f}(x), \tilde{f}(y))$, $(\tilde{f}^2(x), \tilde{f}^2(y))$ are linearly independent, and thus the assumptions of Lemma 6.5 hold for arbitrary $g \in C(K)$.

Applying thus Lemma 6.5, we have shown thus that if for all $x \in K$ there exists an $f \in \overline{\mathcal{A}}$ such that $f(x) \neq 0$, it follows that $\overline{\mathcal{A}} = C(K)$.

On the other hand, suppose now that there exists a point such that for all $f \in \overline{\mathcal{A}}$, $f(x) = 0$. Let us consider the algebra $\mathcal{A}'$ which is spanned by $\mathcal{A}$ and the constants. This is easily seen to equal $\{\mathcal{A} + \lambda 1\}_{\lambda \in \mathbb{R}}$. We have that $\overline{\mathcal{A}'}$ satisfies the property enunciated in the previous paragraph, and in addition, separates points. Thus, we have $\overline{\mathcal{A}'} = C(K)$. But now let $g \in C(K)$ such that $g(x) = 0$, and let $\epsilon > 0$ be arbitrary. This means we may write

$$|g - (f + \lambda)| < \epsilon$$

for some $f \in \mathcal{A}$, $\lambda \in \mathbb{R}$. Evaluating at x, since $f(x) = g(x) = 0$, we obtain $|\lambda| < \epsilon$. Thus, $|g - f| < 2\epsilon$. It follows that $g \in \overline{\mathcal{A}}$. $\square$

We state the complex version of Stone-Weierstrass

Theorem 6.6. *Let $\mathcal{A} \subset C_{\mathbb{C}}(K)$ be a subalgebra over $\mathbb{C}$ separating points, and moreover, suppose that $\mathcal{A}$ is closed under complex conjugation. Then $\overline{\mathcal{A}} = C_{\mathbb{C}}(K)$, or there exists an $x \in K$ such that $\mathcal{A} = \{f \in C_{\mathbb{C}}(K) : f(x) = 0\}$.*

Proof. Recall that $\Re f = \frac{1}{2}(f + \bar{f})$, $\Im f = \frac{i}{2}(\bar{f} - f)$. Thus, by assumption, $f \in \mathcal{A} \implies \Re f \in \mathcal{A}, \Im f \in \mathcal{A}$. Consider the subalgebra $\mathcal{A}'$ (over $\mathbb{R}$) of $C_{\mathbb{R}}(K)$ generated by $\Re f$, $\Im f$, for all $f \in \mathcal{A}$. We have $\mathcal{A}' \subset \mathcal{A}$, and moreover, it is easily seen to separate points.

Suppose that for all $x \in K$, there exists an $f \in \mathcal{A}'$ such that $f(x) \neq 0$. Then by Theorem 6.5, we have that $\overline{\mathcal{A}'} = C_{\mathbb{R}}(K)$. But then for any $u \in C_{\mathbb{R}}(K)$, $v \in C_{\mathbb{R}}(K)$ there exists $f_j \in \mathcal{A}' \to u$, $g_j \in \mathcal{A}' \to v$, and thus $f_j + ig_j \to u + iv$. But $f_j + ig_j \in \mathcal{A}$. Thus, we have shown in this case $\overline{\mathcal{A}} = C_{\mathbb{C}}(K)$.

If on the other hand, there exists an x such that $f(x) = 0$ for all $x \in \mathcal{A}'$, then argue as in the last part of the proof of Theorem 6.5. $\square$

6.5 Applications to Fourier series and the discovery of orthogonality

Theorem 6.6 is useful in the context of classical Fourier series.

Let $K = \mathbb{S}^1$, parametrized, say, by $(-\pi, \pi]$, and consider the vector space $\mathcal{A}$ generated by $\{e^{inx}\}_{n\in\mathbb{Z}}$. The elements of $\mathcal{A}$ are called the *trigonometric polynomials*. $\mathcal{A}$ is clearly an algebra. Moreover, $\mathcal{A}$ separates points, and contains the constants. Finally, since $\overline{e^{inx}} = e^{-inx}$, it follows that $\mathcal{A}$ is closed under complex conjugation. It follows that any $f \in C(\mathbb{S}^1)$ can be approximated by an element of $\mathcal{A}$.

This fact is crucial in the proof of:

Proposition 6.3. *Let $f \in C(\mathbb{S}^1)$, and let $S_N(f) = \sum_{n=-N}^{N} \hat{f}(n)e^{inx}$ be the N'th partial sum of f's Fourier series, i.e. where*

$$\hat{f}(n) = \frac{1}{2\pi}\int_{-\pi}^{\pi} f(t)e^{-int}.$$

We have that

$$\lim_{N\to\infty}\int |f - S_N(f)|^2 \to 0.$$

Proof. If P is a trigonometric polynomial and $N \geq \deg P$, then it follows that $S_N(P) = P$. By Stone-Weierstrass applied to trigonometric polynomials, given ϵ, there exists a P such that $|P - f| < \epsilon$. For such N we have

$$\begin{aligned}
|f - S_N(f)|^2 &\leq |f - P|^2 + |P - S_N(P)|^2 + |S_N(P) - S_N(f)|^2 \\
&= |f - P|^2 + |P - S_N(P)|^2 + |S_N(P - f)|^2 \\
&= |f - P|^2 + |S_N(P - f)|^2
\end{aligned}$$

and thus

$$\begin{aligned}
\int_{-\pi}^{\pi} |f - S_N(f)|^2 &\leq \int_{-\pi}^{\pi} |f - P|^2 + \int_{-\pi}^{\pi} |S_N(P - f)|^2 \\
&\leq 2\pi\epsilon^2 + 2\pi\epsilon^2.
\end{aligned}$$

$\square$

Where did the last inequality come from? The claim is that for any function $g \in C(\mathbb{S}^1)$, we have

$$\int_{-\pi}^{\pi} |S_N(g)|^2 \leq \int_{-\pi}^{\pi} |g|^2. \tag{20}$$

Try showing this directly at this stage.

The geometric structure of (20) may not be immediately apparent. The e^{inx} form what is known as an orthonormal set in $C(\mathbb{S}^1)$, with respect to the "inner product"

$$f \cdot g = \frac{1}{2\pi} \int_{-\pi}^{\pi} f\bar{g}.$$

Orthonormal means $e^{inx} \cdot e^{imx} = \delta_{nm}$, where $\delta_{nm} = 1$ if $n = m$ and 0 otherwise. Modulo the convenient factor of $(2\pi)^{-1}$, this inner product is related to the L^2 norm by $|f|_2 = f \cdot f$.

In this language, we can understand S_N as an "orthogonal projection", and (20) will just follow from the general statement that orthogonal projections do not increase the norm.

We embark in the next section on a general study of normed spaces whose norm arises from an inner product in the sense described above. These are called *Euclidean spaces*. A Euclidean space which is also complete is called a *Hilbert space*. We will return to thus to (20) later…

7 Hilbert space

In this section, we shall introduce the concept of a *Hilbert space*, that is to say a Banach space whose norm arises from an inner product. We have already seen to a certain extent at the end of the previous section how the notion of orthogonality may be useful. It is hard to give a sense of just how important this notion is in analysis.

The inner product structure will also allow us to revisit the notion of duality, defined for general Banach space earlier. As we shall see, for Hilbert spaces, the dual and the adjoint map have very concrete realisations.

7.1 Defintions: inner product space, Euclidean space, Hilbert space

Definition 7.1. *Let V be a vector space over $\mathbb{R}$ or $\mathbb{C}$. Let $p : V \times V \to \mathbb{R}$, $\mathbb{C}$, respectively, be a map such that*

$$p(v, w) = \overline{p(w, v)} \tag{21}$$

$$p(\lambda_1 v_1 + \lambda_2 v_2, w) = \lambda_1 p(v_1, w) + \lambda_2 p(v_2, w) \tag{22}$$

$$p(v, v) \geq 0, \qquad p(v, v) = 0 \text{ iff } v = 0. \tag{23}$$

We call p an inner product *on V; we often denote it by $< v, w >$ instead of $p(v, w)$. A vector space V together with such an inner product $<,>$ is known as an* inner product space.

The properties (21), (22) together are sometimes called *sesquilinearity* and define the notion of a Hermitian form. In the real case, then this just means that p is a bilinear form.

If $< v, w >= 0$ we say that v and w are *orthogonal*.

The most fundamental perhaps property of inner product spaces is the so-called Schwarz inequality.

Proposition 7.1. *Let $(V, <, >)$ be an inner product space. Then*

$$| < v, w > | \leq \sqrt{< v, v >< w, w >}, \tag{24}$$

with equality iff $v = \lambda w$.

Proof. If $v = \lambda w$, we clearly have equality. By replacing w by $\tilde{w} = \alpha w$ with $|\alpha| = 1$, we can arrange so that $< v, \tilde{w} >$ is real. The validity of the equality for $\tilde{w}$ implies that for w. So in what follows we may assume $< v, w >$ is real, and, moreover, that $v \neq \lambda w$

Consider $< v + tw, v + tw >$, for $t \in \mathbb{R}$. By (23) and our assumption, this is strictly positive for all t. Expanding, we obtain

$$\begin{aligned} < v + tw, v + tw > &= < v, v > + < v, tw > + < tw, v > +t^2 < w, w > \\ &= < v, v) + 2t < v, w > +t^2 < w, w > . \end{aligned}$$

Since this polynomial is strictly positive its roots are not real, that is to say $4 < v, w >^2 -4 < v, v >< w, w >< 0$. But this is the desired ineqality. □

Aside: In general, we call a vector v *isotropic* with respect to a Hermitian form if $p(v, v) = 0$; we call a Hermitian form *positive* if $p(w, w) \geq 0$ for all w, and we call it *non-degenerate* if $p(v, w) = 0$ for all w implies $v = 0$. The above proposition has thus shown that a positive hermitian form is non-degenerate iff there are no isotropic vectors, i.e. iff it is positive definite.

Proposition 7.2. *Let $(V, <, >)$ be an inner product space. Define $| \cdot |$ on V by*

$$|v| = \sqrt{< v, v >}. \tag{25}$$

Then $| \cdot |$ defines a norm on V.

Proof. It is enough to show the triangle inequality. We have

$$\begin{aligned} |v + w|^2 &= < v + w, v + w > \\ &= < v, v > + < v, w > + < w, v > + < w, w > \\ &\leq < v, v > +2| < v, w > |+ < w, w > \\ &\leq < v, v > +2\sqrt{< v, v >< w, w >}+ < w, w > \\ &= (\sqrt{< v, v >} + \sqrt{< w, w >})^2 \end{aligned}$$

from which the triangle inequality follows immediately. □

Definition 7.2. *A normed vector space* $(E, |\cdot|)$ *where* $|\cdot|$ *is defined by* (25) *for some inner product* $<,>$ *on* E*, is called a* Euclidean space.

Proposition 7.3. *Let* $(E, |\cdot|)$ *be a Euclidean space. Then there is a unique* $<,>$ *such that* (25) *holds.*

Proof. Let $<,>$ be such that (25) holds. In the case of a real inner product, we have

$$\begin{aligned} < v, w > &= \frac{1}{2}(< v+w, v+w > - < v, v > - < w, w >) \\ &= \frac{1}{2}(|v+w|^2 - |v|^2 - |w|^2) \end{aligned}$$

whereas in the case of a complex inner product, we have

$$< v, w > + < w, v >=< v+w, v+w > - < v, v > - < w, w >,$$

while

$$-i < v, w > +i < w, v >=< v+iw, v+iw > - < v, v > + < w, w >$$

from which we obtain

$$< v, w >= \dots,$$

an expression the reader is invited to fill in. □

The above identities are sometimes known as the *polarization identities*. In what follows then, given a Euclidean space V, then $<,>$ will denote this uniquely defined inner product giving $|\cdot|$ by (25).

The following parallelogram law may be familiar.

Proposition 7.4. *Let* $(E, |\cdot|)$ *be a Euclidean space. Then*

$$|v-w|^2 + |v+w|^2 = 2|v|^2 + 2|w|^2 \tag{26}$$

Proof. We expand using the inner product.

$$\begin{aligned} |v-w|^2 + |v+w|^2 &= < v-w, v-w > + < v+w, v+w > \\ &= < v, v > - < v, w > - < w, v > + < w, v > \\ &\quad + < v, v > + < v, w > + < w, v > + < w, w > \\ &= 2 < v, v > +2 < w, w > \\ &= 2|v|^2 + 2|w|^2. \end{aligned}$$

□

Also,

Proposition 7.5. *Let* $(E, |\cdot|)$ *be a Euclidean space. Suppose* v *and* w *are orthogonal. Then*

$$|v+w|^2 = |v|^2 + |w|^2 \tag{27}$$

Proof. Expand using the inner product... $\square$

Iterating, we have that if v_i are pairwise orthogonal than

$$\left|\sum_{i=1}^{n} v_i\right|^2 = \sum_{i=1}^{n} |v_i|^2.$$

Finally, we have

Definition 7.3. *Let $(H, |\cdot|)$ be a Euclidean space which is Banach, i.e. such that $|\cdot|$ defines a complete metric. We say H is a* Hilbert space.

It turns out that any Euclidean space can be embedded into a larger Hilbert space by taking the completion. For this, let us first note the following easy proposition:

Proposition 7.6. *Let E be a Euclidean space. Then $<,>: E \times E \to \mathbb{C}$ is continuous.*

Proof. This follows from the computation

$$\begin{aligned} |<v,w> - <\tilde{v},\tilde{w}>| &\leq |<v,w> - <v,\tilde{w}>| + |<v,\tilde{w}> - <\tilde{v},\tilde{w}>| \\ &= |<v,w-\tilde{w}>| + |<v-\tilde{v},\tilde{w}>| \\ &\leq |v||w-\tilde{w}| + |v-\tilde{v}||\tilde{w}|. \end{aligned}$$

$\square$

From this follows easily

Proposition 7.7. *Let E denote a Euclidean space, and $\bar{E}$ its completion. Then the inner product extends to an inner product on $\bar{E}$, making the latter into a Hilbert space.*

Proof. This also follows easily from the computation of the previous proposition. The uniqueness follows in particular directly from the continuity of the inner product. $\square$

In view of the above Proposition, Euclidean spaces are sometimes called pre-Hilbert spaces. We shall not use such awful terminology here.[10]

7.1.1 Examples

For us, this whole story began with $C(\mathbb{S}^1)$ with

$$< f, g > \doteq \int_{-\pi}^{\pi} f\bar{g}.$$

[10] I will also resist the temptation to call Hilbert spaces post-Euclidean spaces.

We could also equally well have taken $C([a,b])$ for some $a<b$. This is clearly a Euclidean space with the "L^2 norm"

$$|f|_2 = \sqrt{< f, f >} = \sqrt{\int_a^b |f|^2}$$

Alas, this is not a Hilbert space, as we know already from our discussion of Banach spaces. Its completion is, by Proposition 7.7. It is a miraculous fact that this completion can be realised as a set of (equivalence classes of) Lebesgue measurable functions, where $< f, g >$ is defined with respect to the Lebesgue integral. This is the space L^2, which in any case, we have discussed before in the context of Banach spaces. For this, however, you will need more mathematical technology than you have.

So we have to settle for little l_2. This space is easily seen to be a Euclidean space with inner product

$$< a, b >= \sum a_i \bar{b}_i.$$

Since we know it to be a Banach space with the norm $|a|_2 = \sqrt{< a, a >}$, it follows that l_2 is a Hilbert space.

The space l_2 is clearly separable, i.e. there exists a countable dense subset. We will see later on that all infinite dimensional separable Hilbert spaces are isometrically isomorphic to l_2. So l_2 is not a bad example to have...

This being said, the whole point about setting up the theory of Hilbert spaces is applying it to solve problems in analysis. Knowing, thus, that a particular space is indeed a Hilbert space, is fundamental. This is why it really is a shame that we cannot talk about L^2.

7.2 Orthogonal complements

7.2.1 Basic definitions

Let E be Euclidean. We have already defined what it means for v and w to be orthogonal, namely $< v, w >= 0$.

Definition 7.4. *Let $S \subset E$. We define the* orthogonal space of S, *denoted $S^\perp$, to be the set $S^\perp \doteq \{v \in E : \forall_{w \in S},\ < v, w >= 0\}$.*

A trivial application of continuity and linearity of the inner product yields

Proposition 7.8. *Let $S \subset E$. Then $S^\perp$ is a closed subspace of E, and $S^\perp = (\overline{\text{Span } S})^\perp$.*

Finally, if V is itself a subspace, since we have that $V \cap V^\perp = 0$, it follows that $V + V^\perp$ in our usual notation is equal to the direct sum $V \oplus V^\perp$.

7.2.2 Existence of projection maps

Experience from finite dimensional euclidean space may lead you to believe that $V \oplus V^\perp = E$, for any E, at least if V is closed. This isn't always true in Euclidean spaces! However, we have

Theorem 7.1. *If $F \subset E$ be a subspace, where F is assumed complete and E is Euclidean, then $F \oplus F^{\perp} = E$. Moreover, for arbitrary $x \in E$, writing uniquely $x = x_1 + x_2$, where $x_1 \in F$, $x_2 \in F^{\perp}$, then x_1 is characterized uniquely by*

$$|x_2| = |x_1 - x| = \inf_{y \in F} |y - x|.$$

Note that, of course, the assumptions of the theorem are satisfied if F is finite dimensional, or alternatively, if E is Hilbert and F is closed. In particular, in a Hilbert space, they are satisfied if $F = S^{\perp}$.

Proof. The statement of the theorem should tip you off to the nature of the proof. Take a minimising sequence y_i for $|y - x|$, that is to say, a sequence $y_i \in F$ such that $\lim |y_i - x| = \inf_{y \in F} |y - x| \doteq d$.

First, I claim that y_i is Cauchy. For this, apply (26) for $v = x - y_i$, $w = x - y_j$ to obtain

$$|y_j - y_i|^2 + |2x - y_i - y_j|^2 = 2|y_i - x|^2 + 2|y_j - x|^2,$$

and thus

$$\begin{aligned} |y_j - y_i|^2 &= 2|y_i - x|^2 + 2|y_j - x|^2 - |2x - y_i - y_j|^2 \\ &= 2|y_i - x|^2 + 2|y_j - x|^2 - 4|x - (y_i - y_j)/2|^2 \\ &\leq 2(d + \epsilon) + 2(d + \epsilon) - 4d \\ &\leq 4\epsilon \end{aligned}$$

if i, j are such that $|y_i - x|^2 \leq d + \epsilon$, $|y_j - x|^2 \leq d + \epsilon$.

So y_i is indeed Cauchy, and thus, by the assumption of completeness, converges to some y. By continuity of $|\cdot|$, it follows that $|y - x| = d$.

So the claim is now that y is the unique element of F that achieves d, and that we can set $x_1 = y$, $x_2 = y - x$.

So define x_2 as above and suppose $x_2 \notin F^{\perp}$. Let $\tilde{y} \in F$ such that $< x_2, \tilde{y} > \neq 0$. By multiplying $\tilde{y}$ by a scalar, we can arrange so that $< x_2, \tilde{y} >< 0$. Now consider for $t > 0$,

$$\begin{aligned} < y + t\tilde{y} - x, y + t\tilde{y} - x > &= < y - x, y - x > + < t\tilde{y}, y - x > \\ &\quad + < y - x, t\tilde{y} > + t^2 < \tilde{y}, \tilde{y} > \\ &= d^2 + 2t < \tilde{y}, x_2 > + t^2 |\tilde{y}|^2. \end{aligned}$$

Now, for small enough t the second term is negative and dominates the last term. For such a t then we have

$$< y + t\tilde{y} - x, y + t\tilde{y} - x >< d^2$$

and this contradicts the definition of d in view of the fact that $y + t\tilde{y} \in F$. $\square$

When $F \oplus F^{\perp} = E$, we say that $F^{\perp}$ is an *orthogonal complement* of F.

We may interpret the above Theorem in terms of projection operators.

Corollary 7.1. *Let F, E be as in Theorem 7.1. There exists a unique operator $P : E \to E$ such that $P(E) = F$, $P(F^\perp) = 0$, $P^2 = P$, $(I - P)(E) = F^\perp$, $(I - P)(F) = 0$, $(I - P)^2 = (I - P)$, and $||P|| \leq 1$, $||I - P|| \leq 1$, with equality if $F \neq 0$, $F \neq E$, respectively.*

7.2.3 The Riesz representation theorem

Let E be a Euclidean space. Clearly, by the Schwarz inequality, an element v defines an element of the Banach space E^* by

$$w \mapsto < w, v > .$$

Denote this map $\phi_v \in E^*$, and let $\phi : E \to E^*$ take v to ϕ_v. We have $|\phi_v| = |v|$.

Now suppose H is Hilbert. We have

Proposition 7.9. *Let H be Hilbert, and ϕ be defined as above. Then $\phi : H \to H^*$ is an isometric anti-isomorphism. In particular, H^* is a Hilbert space.*

Proof. Let $0 \neq \xi \in H^*$. Consider $\ker \xi$. By continuity of ξ, $\ker \xi$ is a closed subspace, and thus, by Theorem 7.1, there exists an orthogonal complement. By linear algebra, this is 1-dimensional and thus spanned by some $\tilde{v}$. Define $v = a\tilde{v}$, so that

$$< v, v >= \xi(v).$$

By linearity, now $< bv, v >= \xi(bv)$. Thus ϕ_v and ξ coincide on $\ker \xi$ and on $(\ker \xi)^\perp$. We have shown surjectivity for ϕ. The anti-isomorphism part follows easily. □

Orthogonality and completeness allow us to realise the abstract dual in a very explict way!

7.2.4 Aside: Return to Fourier series

Consider $C(\mathbb{S}^1)$ with the L^2 norm and the problem considered in Section 6.5. The claim is that the map S_N is precisely the map $P : E \to F$ of the Corollary 7.1, where E is $C(\mathbb{S}^1)$, and F is the finite dimensional, and thus complete, subspace spanned by $\{e^{in\theta}\}_{|n| \leq N}$. Given this, the statement $||S_N|| \leq 1$ follows from Corollary 7.1.

So why is S_N a projection operator? We know $F \oplus F^\perp = E$. We know already that $f \in F$ implies $S_N f = f$. On the other hand, $f \in F^\perp$ means that $< f, e^{in\theta} >= 0$, for $|n| \leq N$, and thus $S_N f = 0$. But now this implies that S_N satisfies the conditions for P.

7.3 Orthonormal systems and baseis

We just saw thus in the previous section the importance of having an orthonormal set of vectors, i.e. unit vectors which are pairwise orthogonal.

We now turn to a general discussion. First the definitions, then a return to the examples.

7.3.1 Definitions and existence

Definition 7.5. *Let E be Euclidean. A set $\{e_\alpha\}$ of unit vectors is called an* orthonormal system *if $< e_\alpha, e_\beta >= 0$ for $\alpha \neq \beta$.*

Definition 7.6. *Let E be Euclidean. An orthonormal system is called maximal if it cannot be extended to a strictly larger orthonormal system.*

By Zorn's lemma, there always exists a maximal orthonormal system in any Euclidean space E.

Proposition 7.10. *Let H be Hilbert, and S a maximal orthonormal system. Then $\overline{\text{Span } S} = H$.*

Proof. Consider $S^\perp$. By Proposition 7.8, we have that $S^\perp = \overline{\text{Span } S}^\perp$. Setting $F = \overline{\text{Span } S}$, we have by Theorem 7.1 that $H = F \oplus F^\perp$. If $F^\perp = 0$, we have that $H = F = \overline{\text{Span } S}$, as desired. Otherwise, we have that $S^\perp = F^\perp \neq 0$, in which case there exists a unit $x \in S^\perp$. The system $\{x\} \cup S$ is then orthonormal and contains strictly S. This contradicts the maximality of S. □

Note also the easy converse, which holds in any Euclidean space:

Proposition 7.11. *Let E be Euclidean, and S be an orthonormal system such that $\overline{\text{Span } S} = E$. Then S is maximal.*

Definition 7.7. *Let H be Hilbert. A maximal orthonormal system is called a* Hilbert space basis.

Propostion 7.11 thus provides an alternative characterization of a Hilbert space basis.

We shall see in the examples that either characterization may be easier to check in showing that a given orthonormal system is indeed a basis.

As discussed before, Zorn's lemma is not something one should use lightly. It turns out that in the separable case we can avoid it completely in constructing Hilbert space baseis. First we note the following

Proposition 7.12. *Let $\{x_i\}_{i=1}^N$ be linearly independent for some $N \leq \infty$. Then there exist $\{e_i\}_{i=1}^N$ such that for all $n = 1 \ldots N$, we have* Span $(e_j)_{j=1\ldots n}$ = Span $(x_j)_{j=1\ldots n}$.

The above Proposition, is nothing but the celebrated Gram-Schmidt orthogonalization procedure.

Proof. By induction. Let $e_1 = x_1/|x_1|$. Having constructed $e_1 \ldots e_n$, for $n \geq 1$, define

$$e_{n+1} = \left| x_{n+1} - \sum_{i=1}^{n} < x_{n+1}, e_i > e_i \right|^{-1} \big(x_{n+1} - \sum_{i=1}^{n} < x_{n+1}, e_i > e_i \big).$$

(Note that $e_{n+1} = |P_n x_{n+1}|^{-1}(P_n x_{n+1})$, where P_n is the orthogonal projection to the subspace Span $(x_j)_{j=1\ldots n}$.) □

From this, we obtain the following

Proposition 7.13. *Let H be separable, and let $\{y_i\}$ be countable set such that* $\overline{\text{Span }(y_i)} = H$. *Then there exists a countable Hilbert space basis for H in the span of $\{y_i\}$.*

Proof. Go down the list of $\{y_i\}$, and discard y_j in the span of the previous. One arrives at a $\{y_{i_k}\}$ which are linearly independent, and such that Span $\{y_{i_k}\} =$ Span $\{y_i\}$. Now, apply Gram-Schmidt to $\{y_{i_k}\}$. □

7.3.2 Examples

Take l_2, and let $e_i = (0, \ldots, 0, 1, 0, \ldots)$ where the 1 is in the i'th place. This is clearly maximal, and thus a basis.

Take the completion H of $C(\mathbb{S}^1)$ with respect to the L^2 norm. (This space is also known as L^2.) We know by Stone-Weierstrass that the algebra of trigonometric polynomials is dense in $C(\mathbb{S}^1)$ with respect to the sup norm. This means that it is also dense with respect to the L^2 norm. (Why?) Thus $\mathcal{A}$ is dense in H as well. By orthogonality, it follows that $\mathcal{A}$ is a basis.

As abstract Hilbert spaces, the above two examples are actually the same. A countable, non-finite basis for a Hilbert space can be thought of as an isometric isomorphism with l_2. We turn to this now.

7.4 The isomorphism with l_2

For warm up, try proving for yourself the following statements: Let H be a finite dimensional Hilbert space. Then there exists a Hilbert space basis $\{e_i\}_{i=1}^n$. H is isometrically isomorphic to l_2^n via the map

$$x \mapsto (< x, e_1 >, \ldots < x, e_n >).$$

We have

$$x =< x, e_1 > e_1 + \ldots + < x, e_n > e_n,$$

$$|x|^2 = | < x, e_1 > |^2 + \cdots + | < x, e_n > |^2.$$

We now move on to the general, separable case. First we show the following

Lemma 7.1. *(Bessel's inequality) Let E be Euclidean, and $\{e_i\}_{i=1}^N$ be a countable orthonormal system, for some $1 \le N \le \infty$. For $x \in E$, define $x_i =< x, e_i >$. Then*

$$\sum_{i=1}^N |x_i|^2 \le |x|^2.$$

If $N = \infty$, $(x_i) \in l_2$.

Proof. Let F_n the span of $\{e_i\}_{i=1}^n$, note that $\sum_{i=1}^n |x_i|^2 = |P_{F_n} x|^2 \le |x|^2$. If $N = \infty$, take the limit. □

Proposition 7.14. *Let H be a separable Hilbert space, and let $\{e_i\}_{i=1}^N$ be a countable basis for some $1 \leq N \leq \infty$. Let $x, y \in H$, and define $x_i =< x, e_i >$, $y_i =< y, e_i >$. Then*

$$x = \sum_{i=1}^{N} x_i e_i, \qquad y = \sum_{i=1}^{N} y_i e_i, \tag{28}$$

$$< x, y >= \sum_{i=1}^{N} x_i \overline{y}_i, \tag{29}$$

where if $N = \infty$, the latter series converges absolutely.

Proof. Consider $s_n = \sum_{i=1}^n x_i e_i$. In the case $N = \infty$, by Bessel's inequality, s_n is Cauchy, Thus s_n converges to some s by completeness. In the case $N < \infty$, set $s = s_N$.

Consider $s - x$, and $s_n - x$. Since $< s_n - x, e_m >= 0$ for $n \geq m$, it follows (by continuity of the inner product in the case $N = \infty$, and trivially if $N < \infty$) that $< s - x, e_m >= 0$ for all $m < N + 1$. Thus, $s - x \in \overline{\text{Span } \{e_i\}}^{\perp} = 0$. So $s = x$.

To show (29), Consider $< s_n, \tilde{s}_n >$, where $\tilde{s}_n$ is defined for y replacing x. By the properties of inner products, we have

$$< s_n, \tilde{s}_n >= \sum_{i=1}^{n} x_i \overline{y}_i.$$

In the case $N < \infty$, this immediately gives (29) for $n = N$. Otherwise, by the continuity of the inner product, the left hand side converges to $< x, y >$. On the other hand, the right hand side is an absolutely convergent series since $\sum_{i=1}^n |x_i \overline{y}_i| \leq \sqrt{\sum_{i=1}^n |x_i|^2 \sum_{i=1}^n |y_i|^2} \leq |x||y|$. $\square$

Formula (29), and its specialisation to $x = y$, are classically known as *Parseval's identities.*

Essentially, we have shown that H isometrically embeds to l_2 or l_2^N by the map $x \to (x_i)$. In the finite dimensional case, by dimensionality considerations, the above immediately proves the statement claimed at the beginning of this section, that is to say, H is isometrically isomorphic to l_2^N.

In the countably infinite case, to complete the isometric isomorphism with l_2, we need the surjectivity of this map. This is provided by the so-called Riesz-Fisher theorem, which here is demoted to a

Proposition 7.15. *Let H be a separable Hilbert space with countably infinite basis $\{e_i\}$. Let $(x_i) \in l_2(\mathbb{C})$. Then there exists an $x \in H$ such that $< x, e_i >= x_i$, namely*

$$x \doteq \sum_{i=1}^{\infty} x_i e_i.$$

Proof. Consider the partial sum $s_n = \sum_{i=1}^{n} x_i e_i$. Since $(x_i) \in l_2$, it follows by the Pythagorean theorem that the sequence s_n is Cauchy. Thus by completeness $s_n \to x$. We show as before now that $< x, e_i >= x_i$ by noting that $< s_n, e_i >= x_i$ for $n \geq i$, and the continuity of the inner product. $\square$

8 Compact self-adjoint operators and Spectral Theory

8.1 The spectrum and resolvent: definition

Definition 8.1. *Let X be a Banach space, and $T \in \mathcal{B}(X, X)$. We define the* spectrum *of T, denoted $\sigma(T)$, by*

$$\sigma(T) = \{\lambda \in \mathbb{C} : (T - \lambda I)^{-1} \text{ does not exist}\}.$$

The resolvent set *$\rho(T)$ is defined to be complement of the spectrum, i.e.*

$$\rho(T) \doteq \mathbb{C} \setminus \sigma(T).$$

Note that if $(T - \lambda I)^{-1}$ exists, then it is bounded by the inverse mapping theorem. For this, the completeness of X is essential.

Definition 8.2. *The* resolvent *of T is a map $R : \rho(T) \to \mathcal{B}(X, X)$ defined by $\lambda \mapsto (T - \lambda I)^{-1}$.*

Definition 8.3. *Let X, T be as above. If $Ker(T - \lambda I) \neq 0$, we say that λ is an* eigenvalue *for T, and $Ker(T - \lambda I)$ is the* eigenspace *of λ. The set of eigenvalues is called the* point spectrum *and denoted $\sigma_p(T)$. Clearly $\sigma_p(T) \subset \sigma(T)$.*

In finite dimensions, for $(T - \lambda I)^{-1}$ not to exist, it must be by the dimension theorem that $Ker(T - \lambda I) \neq 0$, that is to say, in finite dimensions

$$\sigma_p(T) = \sigma(T).$$

In infinite dimensions, in general $\sigma_p(T) \neq \sigma(T)$.

8.2 Structure of the spectrum

In this section, we show the closedness, boundedness, and non-emptyness of the spectrum.

Theorem 8.1. *Let X, T be as above. Then $\sigma(T)$ is a closed, nonempty subset of $\{|\lambda| \leq ||T||\}$.*

Proof. To prove that $\sigma(T)$ is closed: This is equivalent to showing that the resolvent set $\rho(T)$ is open.

First we note the following. Let $\mathcal{U} \subset \mathcal{B}(X, X)$ denote the subset of invertible operators.

Lemma 8.1. *$\mathcal{U}$ is open.*

Proof. Let $S_1 \in \mathcal{U}$. Let S_2 be such that $||S_1 - S_2|| < \epsilon$. Note that

$$S_2 = S_1(I - S_1^{-1}(S_1 - S_2)).$$

Define $Q = S_1^{-1}(S_1 - S_2)$. If $||Q|| < 1$, then

$$(I - Q)^{-1} = 1 + Q + Q^2 + Q^3 + \dots \tag{30}$$

and

$$||(I - Q)^{-1}|| \leq \frac{1}{1 - ||Q||}.$$

We have thus that S_2 is invertible for $\epsilon < ||S_1^{-1}||^{-1}$, with S_2^{-1} given by $S_2^{-1} = (I - Q)^{-1}S_1^{-1}$ and moreover,

$$||S_2^{-1}|| \leq \frac{||S_1^{-1}||}{1 - ||S_1^{-1}||||S_1 - S_2||}.$$

□

Now suppose that $\lambda \in \rho(T)$. This means that $(T - \lambda I) \in \mathcal{U}$. But then $T - \mu I \in \mathcal{U}$, for μ sufficiently close to λ, since

$$||(T - \lambda I) - (T - \mu I)|| = |\lambda - \mu|.$$

We have shown thus openness of the resolvent set , and thus, closedness of the spectrum.

To remark that $\sigma(T) \subset \{|\lambda| \leq ||T||\}$, or equivalently $\rho(T) \supset \{|\lambda| > ||T||\}$, just note that for $|\lambda| > ||T||$, $T - \lambda I = \lambda(\lambda^{-1}T - I)$, and $\lambda^{-1}T - I$ is invertible by the previous, since $||\lambda^{-1}T|| < 1$. Thus $(T - \lambda I)^{-1}$ exists, i.e., $\lambda \in \rho(T)$. Note moreover, that for such λ, we have

$$||R(\lambda)|| = ||(T - \lambda I)^{-1}|| \leq |\lambda|^{-1}(1 - |\lambda^{-1}|||T||)^{-1}. \tag{31}$$

We turn to showing non-emptyness. Choose a point λ_0 in the resolvent, and let λ be sufficiently close to λ_0. The formula (30) says that

$$\begin{aligned} T - \lambda I &= (T - \lambda_0 I)(I - (T - \lambda_0 I)^{-1}(T - \lambda_0 I - (T - \lambda I))) \\ &= (T - \lambda_0 I)(I - (T - \lambda_0 I)^{-1}((\lambda - \lambda_0)I)) \end{aligned}$$

Thus, for small enough $|\lambda - \lambda_0|$,

$$R(\lambda) = (T - \lambda I)^{-1} = \sum_{i=0}^{\infty}(T - \lambda_0 I)^{-i-1}(\lambda - \lambda_0)^{-i}I.$$

That is to say, R is an operator valued holomorphic function on $\rho(T)$.

So suppose that $\sigma(T) = \emptyset$. That is to say, $\rho(T) = \mathbb{C}$. In the language of complex analysis, this means that R is entire. On the other hand, by (31), we have that R is bounded. Liouville's theorem from complex analysis says that R must be constant.

But of course, R cannot be constant, because $(T - \lambda I)^{-1} \neq (T - \mu I)^{-1}$ if $\lambda \neq \mu$. The contradiction proves $\rho(T) \neq \mathbb{C}$, and thus, $\sigma(T) \neq \emptyset$. $\square$

One might want to compare with the finite dimensional case. Specialising to that case, in view of $\sigma_p(T) = \sigma(T)$, we have shown that every linear transformation has an eigenvalue.

The usual proof of this latter fact goes through the characteristic polynomial. Any root of the characteristic polynomial is an eigenvalue. All polynomials over $\mathbb{C}$ have a root, by the fundamental theorem of algebra. Thus, any T has an eigenvalue.

The algebraic device of the characteristic polynomial is not available to us in infinite dimensions. But one must remember, that even in finite dimensions, this does not render the proof completely algebraic. For the fundamental theorem of algebra requires an analytic argument. In fact, one classic proof proceeds precisely via Liouville's theorem!

8.3 Compact operators

8.3.1 Definitions

Definition 8.4. *Let X, Y be Banach spaces. An operator $T \in \mathcal{L}(X, Y)$ is said to be* compact *if $E \subset X$ bounded implies $T(E)$ is totally bounded.*

Note that a compact operator is in particular, bounded.

Proposition 8.1. *Let X, Y be Banach spaces. Then $T \in \mathcal{L}(X, Y)$ is compact iff $T(B(1))$ is totally bounded iff $\overline{T(B(1))}$ is compact.*

The proof of the above is immediate.

8.3.2 Examples and non-examples

If X and Y are finite dimensions, then every $T \in \mathcal{L}(X, Y)$ is compact. In infinite dimensions, "most" operators are not compact. In particular we easily see that the identity map $I : X \to X$ is compact iff X is finite dimension.

The zero map is certainly compact. A non-trivial example of a compact operator is again the identity map, considered though as a map $\phi : C^1(\overline{\mathcal{U}}) \to C^0(\overline{\mathcal{U}})$. The compactness of this map is a corollary of Arzela-Ascoli.

A more interesting example of a compact operator is the map $L : C^2(\mathbb{S}^1) \to C^0(D)$, where D denotes the closed unit ball, and L takes a function f to the unique solution $\psi \in C^0(D) \cap C^2(B(1))$ of $\Box\psi = 0$ with boundary values f. The compactness of this operator follows follows from estimates you proved on example sheets.

8.3.3 The spectrum of compact operators

We shall not give a general discussion of the theory of compact operators on general Banach spaces, in particular, their spectrum. Let us just quote the following theorem:

Theorem 8.2. *Let $T : X \to X$ be compact. Then the point spectrum of T is a countable set $\{\lambda_i\}$. If X is infinite dimensional, then $\sigma(T) = \{0\} \cup \{\lambda_i\}$, and $\lambda_i \to 0$ if there are infinitely many λ_i. Moreover, the eigenspace corresponding to $\lambda_i \neq 0$ is finite dimensional.*

8.4 Self-adjoint operators on Hilbert space

We have already defined the adjoint

$$T^* : Y^* \to X^*$$

of an operator

$$T : X \to Y.$$

Suppose now that $X = Y = H$ a Hilbert space. We know that H can be identified with H^* via the Riesz Representation Theorem. Thus, we can "compare" T and T^*. Let $\phi : H \to H^*$ be the antilinear isometry.

Definition 8.5. *We say that $T : H \to H$ is* self-adjoint *if $\phi \circ T \circ \phi^{-1} = T^*$.*

We have

Proposition 8.2. *Let $T : H \to H$ be bounded, and let $T^* : H^* \to H^*$ be the adjoint, and let ϕ be the map of the Riesz Representation theorem. Then*

$$< Tx, y >=< x, \phi^{-1} \circ T^* \circ \phi(y) > . \tag{32}$$

In particular $T : H \to H$ is self adjoint iff $< Tx, y >=< x, Ty >$ for all x, y.

Proof. Formula (32) is just obtained by chasing arrows.

$$\begin{aligned} < x, \phi^{-1} \circ T^* \circ \phi(y) > &= (T^* \circ \phi(y))(x) \\ &= (T^*(\phi(y)))(x) \\ &= \phi(y)(T(x)) \\ &= < Tx, y > . \end{aligned}$$

Note that $< x, \phi^{-1} \circ T^* \circ \phi(y) >$ for all x, y, determines T^*. Thus the iff statement. □

8.4.1 Eigenvalues and eigenspaces of compact self-adjoint operators

First some notation. Let $T : H \to H$. We have defined already the eigenspace corresponding to an eigenvalue λ. Let us denote this by E_λ. And Let P_{E_λ} denote the orthogonal projection to E_λ.

Proposition 8.3. *Let $T : H \to H$ be self-adjoint. Then $\sigma_p(T) \subset \mathbb{R}$.*

Proof. Let $\lambda \in \sigma_p(T)$ with $0 \neq v \in E_\lambda$. Then

$$\lambda|v|^2 =< \lambda v, v >=< Tv, v >=< v, Tv >=< v, \lambda v >= \bar{\lambda}|v|^2$$

from which we obtain $\lambda = \bar{\lambda}$. □

Proposition 8.4. *Let $T : H \to H$ be self-adjoint, and let $\lambda, \nu \in \sigma_p(T)$ be distinct. Then $E_\lambda \perp E_\nu$.*

Proof. Let $0 \neq v \in E_\lambda$, $0 \neq w \in E_\nu$. We have

$$\lambda < v, w >=< \lambda v, w >=< Tv, w >=< v, Tw >= \nu < v, w >$$

from which we obtain, since $\lambda \neq \nu$, that $< v, w >= 0$. □

Proposition 8.5. *Let $T : H \to H$ be self-adjoint, and $\lambda_\alpha \in \sigma_p(T)$. Then*

$$T(\oplus_\alpha E_{\lambda_\alpha}) \subset \oplus_\alpha E_{\lambda_\alpha}$$

$$T(\overline{\oplus_\alpha E_{\lambda_\alpha}}) \subset \overline{\oplus_\alpha E_{\lambda_\alpha}}$$

$$T((\oplus_\alpha E_{\lambda_\alpha})^\perp) \subset (\oplus_\alpha E_{\lambda_\alpha})^\perp$$

Finally, $\sigma_p(T|_{(\oplus_\alpha E_{\lambda_\alpha})^\perp}) = \sigma_p(T) \setminus \cup_\alpha \{\lambda_\alpha\}$.

Proof. The first inclusion is clear, the second follows from the first by continuity. The third follows from the following. Let $w \in (\oplus_\alpha E_{\lambda_\alpha})^\perp$. We have that for all $v \in E_{\lambda_\alpha}$, $< w, v >= 0$. By then $< Tw, v >=< w, Tv >= \lambda < w, v >= 0$. Thus $Tw \in (\oplus_\alpha E_{\lambda_\alpha})^\perp$. □

Proposition 8.6. *Let $T : H \to H$ be compact, self-adjoint, and let $\lambda_0 > 0$. Then*

$$\dim \oplus_{\lambda_0 \leq |\lambda_\alpha|, \lambda_\alpha \in \sigma_p(T)} E_{\lambda_\alpha} < \infty$$

Proof. Suppose not. Then there exists an infinite sequence of unit vectors $\{v_i\}$ with $Tv_i \perp Tv_j$ for $i \neq j$, and $||Tv_i|| \geq \lambda_0$. It follows that

$$||Tv_i - Tv_j||^2 =< Tv_i, Tv_i > + < Tv_j, Tv_j >\geq 2\lambda_0^2,$$

thus no subsequence of Tv_i converges. Thus T is not compact. □

Corollary 8.1. *Let T be compact, self-adjoint. If $\lambda \neq 0$, then E_λ is finite dimensional.*

Corollary 8.2. *Let T be compact, self-adjoint. Given, ϵ, here are only finitely many λ_α with $|\lambda_\alpha| > \epsilon$. Thus $\sigma_p(T)$ is either finite or countably infinite with $\lambda_i \to 0$.*

8.4.2 The spectral theorem

Theorem 8.3. *Let T be a self-adjoint, compact operator $T : H \to H$. Then the point spectrum of T is a real countable set $\{\lambda_i\}_{i=1}^N$ for some $1 \leq N \leq \infty$. The eienspaces E_{λ_i} are finite dimensional for $\lambda_i \neq 0$, and $E_{\lambda_i} \perp E_{\lambda_j}$ for $i \neq j$. If their are infinitely many λ_i, then $\lambda_i \to 0$. Moreover, if H is infinite dimensional then $\sigma(T) = \sigma_p(T) \cup \{0\}$. Finally, we may write T as*

$$T = \sum_{i=1}^{N} \lambda_i P_{E_{\lambda_i}}.$$

Proof. Like other results in Hilbert space theory, this is proven by exploiting variational methods. In particular, we have a variational characterization of the eigenvalues of T. For this, the following Lemma will be useful:

Lemma 8.2. *Let $T : H \to H$ be self-adjoint. Then*

$$||T|| = \sup_{|x|=1} |< Tx, x >|. \tag{33}$$

Proof. This is not immediately obvious because $||T||$ is defined as $\sup \sqrt{< Tx, Tx >}$. First we note that

$$||T|| = \sup_{|x|=1, |y|=1} |< Tx, y >|. \tag{34}$$

Let x_i be a maximizing sequence for $|Tx_i|$, with $|x_i| = 1$. This means that

$$\begin{aligned} ||T|| &= ||T||^{-1} \lim < Tx_i, Tx_i > \\ &= ||T||^{-1} \lim < x_i, T(Tx_i) > \\ &= \lim < x_i, T(|Tx_i|^{-1} Tx_i > \end{aligned}$$

So, setting $y_i = |Tx_i|^{-1} Tx_i$, we have that $< x_i, Ty_i > \to ||T||$. On the other hand, $|< x, Ty >| \leq ||T||$ for any $|x| = 1$, $|y| = 1$.

Let λ denote the right hand side of (33). We compute that

$$\begin{aligned} |< Tx, y >| &\leq \frac{1}{4} |< T(x+y), (x+y) > - < T(x-y), (x-y) >| \\ &\leq \frac{1}{4} (\lambda ||x+y||^2 + \lambda ||x-y||^2) \\ &\leq \frac{\lambda}{4} (||x+y||^2 + ||x-y||^2) \\ &= \frac{\lambda}{4} (2||x||^2 + 2||y||^2) \\ &= \lambda, \end{aligned}$$

where we have used the parallelogram law. Thus, (34) implies (33). □

The notation λ was meant to be suggestive! Without loss of generality, let us assume that $\lambda = \sup_{|x|=1} < Tx, x >$. We will show that λ is an eigenvalue.

For this, let x_i be a maximising sequence for $< Tx, x >$, i.e., let $< Tx_i, x_i >\to \lambda$ with $|x_i| \leq 1$. By compactness of T, there exists a subsequence x_{i_k} such that $Tx_{i_k} \to y$.

Now we have

$$< Tx_i - \lambda x_i, Tx_i - \lambda x_i >= ||T(x_i)||^2 - 2\lambda < Tx_i, x_i > +\lambda^2 < x_i, x_i >^2 .$$

Thus $||Tx_i - \lambda x_i||^2 \to 0$. Since $Tx_{i_k} \to 0$, then $x_{i_k} \to 0$ and $Ty = \lambda y$. Note that $y \neq 0$. We have produced an eigenvector of eigenvalue λ.

Consider now the eigenspace E_λ. By compactness, it is necessarily finite dimensional. Writing $H = E_\lambda \oplus E_\lambda^\perp$, we have that $T(E_\lambda) = E_\lambda$, $T(E_\lambda)^\perp = E_\lambda^\perp$. The problem is thus reduced to understanding an operator on a smaller space.

Iterating the above argument with $E_\lambda^\perp$, in place of H, etc., we obtain a sequence of distinct eigenvalues $|\lambda_1| \geq |\lambda_2| \geq \cdots$. If this sequence is finite, we must have

$$H = \oplus_{i=1}^n E_{\lambda_i}$$

from which one easily deduces

$$T = \sum_{i=1}^n \lambda_i P_{E_{\lambda_i}}.$$

If the λ_i are infinite in number, then $|\lambda_i| \to 0$. For otherwise, there is an infinite dimensional subspace $H_\infty \subset H$ such that for $|x| = 1$, $< Tx, Tx >\geq \lambda_0 < x, x >\geq \lambda_0$, from which one contradicts compactness.

It follows thus that we must have that T is the 0 map on

$$\overline{(\oplus E_{\lambda_i})}^\perp = (\oplus E_{\lambda_i})^\perp,$$

for $TP_{(\oplus E_{\lambda_i})^\perp}$ cannot have a nonzero eigenvalue. That is to say

$$H = E_0 \oplus \overline{\oplus_{i=1}^\infty E_{\lambda_i}},$$

from which it is clear that $\sigma_p \subset \{0\} \cup \lambda_i$.

From $|| \sum_{i=n}^m \lambda_i P E_{\lambda_I} || \leq \max |\lambda_i|$, we obtain immediately that from the fact that $|\lambda_i| \to 0$ and previous considerations that

$$\sum_{i=1}^\infty \lambda_i P_{E_{\lambda_i}}$$

exists and equals T.

The only thing left is to show that, in the infinite dimensional case $\sigma \subset \{0\} \cap \sigma_p$, as the other inclusion is immediate from the closedness of the spectrum.

For this, let $\nu \notin \{0\} \cup \sigma_p(T)$. Let T_n denote the partial sum of the series representing T. We have

$$T_n - \nu I = \sum_{i=1}^n (\lambda_i - \nu) P_{E_{\lambda_i}} - \nu P_{(\oplus_{i=1}^n E_{\lambda_i})^\perp}$$

From which is is clear that $(T_n-\nu I)^{-1}$ exists, and $||(T_n-\nu I)^{-1}|| \leq \max\{|\nu|^{-1}, |\lambda_i-\nu|^{-1}\}$. We know that for every ϵ there exists an n such that $||T-T_n|| < \epsilon$, and thus, by the above computation

$$||(T_n-\nu I)^{-1}||||T-\nu I-(T_n-\nu I)|| < \epsilon'$$

thus, we have the invertibility of $T-\nu I$. It follows that $\nu \notin \sigma(T)$. $\square$

9 Thanks

Thanks to Susan Thomas and Paul Jefferys for comments and corrections.

References

[1] B. Bollobas *Linear Analysis: an introductory course* second edition, Cambridge University Press, 1999

[2] J. Dieudonné *History of Functional Analysis* North-Holland Mathematics Studies, 49, Amsterdam, 1981

[3] P. Lax *Functional Analysis* Wiley-Interscience, New York, 2002

[4] W. Rudin *Functional Analysis* second edition, McGraw-Hill, New York, 1991