

Functional Analysis

Jacob Denson

December 27, 2019

Table Of Contents

Part I

Vector Spaces with a Metric

Functional analysis is the interlace of algebra and analysis, in which algebraic structures are endowed with topological structure. The approach's utility counts for the rapid growth of applications over the past century, be it in quantum mechanics, statistics, or computing science. Rarely are we concerned with a single object, like a function, a random variable, or a measure, but instead consider large classes of such objects. One technique for handling these classes is to add algebraic structure which elaborates on the natural relation between these objects. Functional analysis adds in analytical structure to the class, so we may combine the algebraic and analytical structure to obtain powerful results about the objects involved.

Example. *We rarely analyze a measurable function f in isolation. Instead, we prove theorems about a class of measurable functions defined on the same measurable space. If f and g are measurable, then we may consider their addition $f + g$, their multiplication fg , and scaling λf (for $\lambda \in \mathbf{R}$), which are all measurable. Thus the space of measurable functions on a set is a vector space. Similarly, the sum and product of two continuous functions on a topological space is continuous, so, we may consider $C[0, 1]$, the space of all continuous, real-valued functions on the unit interval $[0, 1]$, as a vector space.*

As is expected in the name, functional analysis mostly applies to functions between topological and algebraic structures, because there is a natural notion of scaling and addition, and we may obtain topological structure based on the various methods of convergence a function can have. The power of functional analysis rests on how strongly the algebraic structure of a space connects to the topology of the space.

The most common analytic operations are scaling and rotation. As a result, the most common vector spaces which occur in functional analysis are over the real and complex numbers. More specifically, real vector spaces X are those equipped with natural scaling operations $x \mapsto \lambda x$ for $\lambda > 0$, combined with 'reflection' operators $x \mapsto -x$. If X is a real vector space with a specified 'counterclockwise twisting', i.e. a linear map $J : X \rightarrow X$ with $J^2 = -1$, then there is a natural way to turn X into a complex vector space by defining $(a + ib)x = ax + b(Jx)$. The geometry of the complex numbers then enables us to define a counterclockwise twist by any angle θ , by the operation $x \mapsto e^{i\theta}x$. Thus we have a natural operation of the complex numbers over the space.

Moreover, if we work with vector spaces over a field of characteristic zero equipped with a 'complete' topological structure, then we should ex-

pect the underlying field of the vector space to be of characteristic zero as well. In particular, working over the real numbers *at least* isn't too particular a criterion. In these notes, we let \mathbf{K} denote either the real or complex numbers.

Chapter 1

Banach Spaces

The most natural way to associate a space with topological structure is by associating a metric on the space. There is an even more natural vector space operation on a vector space. We can think of elements of a vector space as ‘arrows’ extending from the origin. Thus a natural way to add a topology is to give these arrows a ‘length’. More precisely, a **pre-norm** on a vector space X is a map $\|\cdot\| : X \rightarrow [0, \infty)$ which is *homogenous*, in the sense that for each $\alpha \in \mathbf{K}$ and $x \in X$, $\|\alpha x\| = |\alpha| \|x\|$, satisfies the *triangle inequality* $\|x + y\| \leq \|x\| + \|y\|$ for each $x, y \in X$. If in addition, we have $\|x\| = 0$ if and only if $x = 0$, then we say $\|\cdot\|$ is a *norm*. A vector space X equipped with a norm is called a *norm space*.

On finite dimensional spaces, there is essentially a unique way to define a norm on the space. We say two norms $\|\cdot\|_1$ and $\|\cdot\|_2$ on a vector space X are *comparable* if for each $x \in X$, $\|x\|_1 \sim \|x\|_2$. It is a simple consequence of the Heine-Borel theorem that all norms on a finite dimensional space are normable. Thus there is only a single notion of ‘size’ on such spaces. On the other hand, there are many ways to measure the size of a function, or more general vectors in infinite dimensional norm spaces. For instance, we could consider a function f big if there exists x such that $f(x)$ has large magnitude, or if f is large ‘on average’. This leads to the notions of the L^p and L^∞ norms.

Example. Let X be a measure space. For each $0 < p < \infty$, let $\mathcal{L}^p(X)$ denote the space of all measurable functions $f : X \rightarrow \mathbf{K}$ such that

$$\|f\|_p = \left(\int |f(x)|^p dx \right)^{1/p} < \infty.$$

Then Minkowski's inequality tells us that the quantity $\|\cdot\|_p$ is a prenorm on $\mathcal{L}^p(X)$. We notice that if u is a function with $u(x) = 0$ for almost every x , then $\|f\|_p = \|f + u\|_p$. In particular, this means that $\|\cdot\|_p$ is a well defined operation on the quotient space of $\mathcal{L}^p(X)$ where functions are identified almost everywhere. Since $\|f\|_p = 0$ if and only if $f(x) = 0$ for almost every x , we obtain a norm space, denoted by $L^p(X)$. This is the space where most theorems of functional analysis prove useful.

Example. Given a measurable function f on a measure space X , we can define the L^∞ norm

$$\|f\|_\infty = \inf\{t : |f(x)| \leq t \text{ for almost every } x\}.$$

The space of functions with finite L^∞ prenorm is denoted by $\mathcal{L}^\infty(X)$. Considering the quotient space of functions identified almost everywhere, we obtain a norm space $L^\infty(X)$.

Example. On a compact topological space K , we can consider the family of all bounded continuous functions from K to \mathbf{K} , denoted $C(K)$. The quantity

$$\|f\|_\infty = \sup\{|f(x)| : x \in K\},$$

gives $C(K)$ the structure of a norm space.

For the purpose of the general theory, one need not know too much measure theory to get by. The most important examples of the L^p spaces for intuition and forming counterexamples are the spaces l^p , which is the collection of all sequences $\{a_n\}$ for which, if $p < \infty$,

$$\sum_{n=1}^{\infty} a_n^p < \infty,$$

and for $p = \infty$, l_∞ is the collection of all bounded sequences.

Remark. We will focus on norms in these notes, but prenorms are really no more general. If X is a space with a prenorm $\|\cdot\|$, then the space $X_0 = \{x \in X : \|x\| = 0\}$ forms a subspace of X . We have $\|x\| = \|x + x_0\|$ for each $x_0 \in X_0$, so if we form the quotient space $X' = X/X_0$, then the quantity $\|x + X_0\|$ is well defined, and this is a *norm* on X' . So results about norms automatically extend to results about prenorms where we identify vectors which differ by an element of X_0 . For instance, if we

For each norm on a space X , we can associate a metric given by $d(x, y) = \|x - y\|$. The most powerful theorems of functional analysis deal with the study of norm spaces where the induced metric gives the space X a *complete* metric space structure. These spaces are known as *Banach spaces*. It is a basic fact of measure theory that the spaces $L^p(X)$ are complete, as is the space $C(K)$.

Theorem 1.1. $L^p(X)$ is a Banach space.

Proof. We leave the $p = \infty$ case to the reader, since the argument is elementary. Let f_1, f_2, \dots be a Cauchy sequence in $L^p(X)$. Then we can find a subsequence f_{k_i} such that for each $m \geq 1$,

$$\|f_{k_{i+m}} - f_{k_i}\|_p \leq 1/4^k.$$

In particular, this means that the infinite sum

$$\sum_{i=1}^{\infty} |f_{k_{i+1}}(x) - f_{k_i}(x)|$$

converges absolutely for almost all x . This is a consequence of the Borel-Cantelli lemma, since if

$$E_k = \{x : |f_{k_{i+1}}(x) - f_{k_i}(x)| \geq 1/2^{pk}\},$$

then Markov's inequality implies $|E_i| \leq 1/2^{ip}$, and so $\sum_{i=1}^{\infty} |E_i| < \infty$. Thus we can define a measurable function f defined almost everywhere by the infinite series

$$f(x) = f_{k_1}(x) + \sum_{i=1}^{\infty} f_{k_{i+1}}(x) - f_{k_i}(x) = \lim_{n \rightarrow \infty} f_{k_n}(x).$$

Fatou's lemma and Minkowski's inequality imply that

$$\begin{aligned} \|f - f_{k_i}\|_p &= \left(\int \left| \lim_{n \rightarrow \infty} f_{k_n}(x) - f_{k_i}(x) \right|^p dx \right)^{1/p} \\ &\leq \liminf_{n \rightarrow \infty} \left(\int |f_{k_n}(x) - f_{k_i}(x)|^p dx \right)^{1/p} \\ &\leq \liminf_{n \rightarrow \infty} \|f_{k_n} - f_{k_i}\|_p \leq 1/4^k. \end{aligned}$$

Thus $f_{k_i} \rightarrow f$ in the L^p norm. Since $\{f_k\}$ is a Cauchy sequence, it is easy to justify that $f_k \rightarrow f$ in the L^p norm. Thus Cauchy sequences converge. \square

Remark. Again, studying Banach spaces is not too much more general than studying norm spaces. Given any norm space X , we can always find a Banach space X' containing X , such that X is dense in X' . One can view the elements of $X' - X$ as ‘artificial’ elements of X that nonetheless are key to understanding the geometric structure of X . Provided we are willing to deal with these artificial elements, it suffices without loss of generality to assume in the theory that we are working in a Banach space.

1.1 Subspaces

If Y is a subspace of a norm space X , then Y obviously has the natural structure of a norm space. We note that if X is a Banach space, then Y is a Banach space under this induced norm if and only if Y is a *closed* subspace of X . Here are some examples.

Example. Let $c_0 \subset l_\infty$ be the family of all sequences $\{x_n\}$ such that $\lim_{n \rightarrow \infty} x_n = 0$. Then c_0 is a closed subspace of l_∞ , and is thus a Banach space.

Example. Let \mathbf{T} be the interval $[-\pi, \pi)$ equipped with the Lebesgue measure. For each $f \in L^p(\mathbf{T})$, and $n \in \mathbf{Z}$, we define

$$\hat{f}(n) = \frac{1}{2\pi} \int_{-\pi}^{\pi} f(x) e^{-nix} dx.$$

If

Chapter 2

Hilbert Spaces

2.1 Real Inner Product Spaces

Multivariate calculus tells us how to perform analysis on \mathbf{R}^n . Hilbert spaces are those spaces closest to \mathbf{R}^n . It is here that we may measure angles and distance, which give rise to the topological properties of the space. Recall that in \mathbf{R}^2 , the angle between two vectors v and w in \mathbf{R}^2 is connected by the scalar product identity

$$v_1 w_1 + v_2 w_2 = |v||w|\cos\theta$$

Where $|v|^2 = v_1^2 + v_2^2$ and $|w|^2 = w_1^2 + w_2^2$ can be measured by Pythagoras' identity. One way to see why this equation is true is to note that the projection

$$P_w(v) = \frac{v_1 w_1 + v_2 w_2}{|w|^2} w$$

which is the projection of v onto the line spanned by w , forms a right angled triangle with v as the hypotenuse, and the equation is just given by the trigonometric identities since the length of $P_w(v)$ is $|v|\cos\theta$. Because of how useful this operation is, in higher dimensions, we define the *scalar product* in \mathbf{R}^n by the equation $v \cdot w = v_1 w_1 + \dots + v_n w_n$. The length of a vector v is then defined by $|v|^2 = v \cdot v$, and the angle θ between the vectors by the equation $v \cdot w = |v||w|\cos\theta$, where the equation makes sense because of the Cauchy Schwarz inequality, which says $|v \cdot w| \leq |v||w|$.

Inner product spaces generalize the scalar product on \mathbf{R}^n to more general vector spaces. An inner product associates to any two vectors v and

w a real number $\langle v, w \rangle$, which is *bilinear*, in the sense that it is linear in each variable once the other vector is fixed, *symmetric*, in the sense that $\langle v, w \rangle = \langle w, v \rangle$, and is *positive definite*, in the sense that $\langle v, v \rangle > 0$ for $v \neq 0$.

Example. The spaces \mathbf{R}^n are inner product spaces under the scalar product

$$v \cdot w = v_1 w_1 + \cdots + v_n w_n$$

Example. The space $L^2(\mathbf{R})$ of square summable real-valued functions on the real line is an inner product space under the inner product

$$\langle f, g \rangle_{L^2(\mathbf{R})} = \int_{-\infty}^{\infty} f(x)g(x) dx$$

The Hölder inequality guarantees that this definition is finite for all $f, g \in L^2(\mathbf{R})$. More generally, the space $L^2(X)$ of square summable functions on any measure space X is an inner product space under the inner product

$$\langle f, g \rangle_{L^2(X)} = \int_X f(x)g(x) dx$$

The most useful applications are where we let X denote a subregion of \mathbf{R}^n .

Example. The space $l^2(\mathbf{Z})$ of sequences $f : \mathbf{Z} \rightarrow \mathbf{R}$ such that

$$\sum_{n \in \mathbf{Z}} f(n)^2 < \infty$$

forms an inner product space under the inner product

$$\langle f, g \rangle_{l^2(\mathbf{Z})} = \sum_{n \in \mathbf{Z}} f(n)g(n)$$

where the discrete version of the Hölder inequality guarantees that the sum converges absolutely.

In any inner product space, we can measure the length of a vector v , generalizing the length in Euclidean space, by the equation $\|v\|^2 = \langle v, v \rangle$ (we note this definition is linear in the sense that $\|\lambda v\| = |\lambda| \|v\|$). This gives the inner product space a metric space structure, if we define the distance between two vectors v and w to be the length of the vector $w - v$, once we prove that the triangle inequality holds. First, we must prove the most

fundamental theorem in analysis, the Cauchy Schwartz inequality. This is just a generalization of the Pythagorean theorem, which we now state in the appropriate terms of inner product space theory. We say two vectors v and w are **orthogonal**, or **perpendicular**, if $\langle v, w \rangle = 0$. Note that in the case of two dimensions, if v and w are both nonzero, this means $\cos \theta = 0$, so θ is a quarter angle clockwise or anticlockwise.

Theorem 2.1 (Pythagorean Theorem). *If v and w are orthogonal,*

$$\|v + w\|^2 = \|v\|^2 + \|w\|^2$$

Proof. We just calculate that

$$\|v + w\|^2 = \langle v + w, v + w \rangle = \|v\|^2 + 2\langle v, w \rangle + \|w\|^2$$

and we then just note that $\langle v, w \rangle = 0$. □

Corollary 2.2. *The projection*

$$P_w(v) = \frac{\langle v, w \rangle}{|w|^2} w$$

is the closest vector in the line spanned by w to the vector v .

Proof. The vector $v - P_w(v)$ is orthogonal to any scalar multiple of w , and so the Pythagorean theorem guarantees that for any $\lambda \in \mathbf{R}$,

$$\|v - \lambda w\|^2 = \|v - P_w(v) + P_w(v) - \lambda w\|^2 = \|v - P_w(v)\|^2 + \|P_w(v) - \lambda w\|^2 \geq \|v - P_w(v)\|^2$$

with equality if and only if $P_w(v) - \lambda w = 0$. □

Theorem 2.3 (Cauchy Schwartz Inequality). *For any vectors v and w in an inner product space,*

$$|\langle v, w \rangle| \leq |v||w|$$

Proof. The idea of this proof, in two dimensions, is just a trivial consequence of the Pythagorean theorem: the hypotenuse is the longest side of a right angled triangle. We just rephrase this in terms of an inner product space. In two dimensions, we saw that the quantity

$$P_w(v) = \frac{\langle v, w \rangle}{|w|^2} w$$

gives the projection of v onto w . In two dimensions, we saw this formed a right angled triangle with v as the hypotenuse, and similarly, in a general inner product space, we find that w is orthogonal to the vector $v - P_w(v)$. The Pythagorean theorem guarantees that

$$\|v\|^2 = \|P_w(v)\|^2 + \|v - P_w(v)\|^2 \geq \|P_w(v)\|^2 = \frac{\langle v, w \rangle^2}{\|w\|^2}$$

Rearranging the equation and taking square roots gives the inequality. \square

Note that this theorem justifies the fact that we can define the angle between two vectors by the equation

$$\langle v, w \rangle = \|v\| \|w\| \cos \theta$$

The triangle inequality, giving the inner product space a metric space structure, is an easy consequence of the Cauchy Schwartz inequality.

Theorem 2.4 (Triangle Inequality). *For any vectors v and w ,*

$$\|v + w\| \leq \|v\| + \|w\|$$

Proof. We just calculate

$$\|v + w\|^2 = \|v\|^2 + 2\langle v, w \rangle + \|w\|^2 \leq \|v\|^2 + 2\|v\|\|w\| + \|w\|^2 = (\|v\| + \|w\|)^2$$

and then we take square roots. \square

We can now think of an inner product space as a kind of ‘infinite dimensional’ Euclidean space, with a topology given by the length function, and induced by the triangle inequality. If the metric structure is *complete*, then we say that the inner product space is a **Hilbert space**.

2.2 Complex Inner Product Spaces

To generalize inner product spaces to the complex domain, we turn to our most basic example, \mathbf{R}^2 , and reinterpret it as the complex plane \mathbf{C} . Here, complex conjugation plays a role in defining the distance function, removing the rotation properties of a complex number. The length of a

particular complex number z is given by the equation $|z|^2 = z\bar{z}$. Over \mathbf{C}^n , we therefore introduce the scalar *hermitian product*

$$v \cdot w = v_1 \overline{w_1} + \cdots + v_n \overline{w_n}$$

this equation agrees with the standard scalar product when v and w have real coordinates, and also means that the length defined by the equation $|v|^2 = v \cdot v$ agrees with the length given in \mathbf{R}^{2n} . Unlike the scalar product over the real numbers, this equation is not bilinear, but instead *sesquilinear* (a word meaning ‘one and a half’ in Greek), which is linear in the first argument, but *antilinear* in the second, so that $(v \cdot (\lambda w + \gamma u) = \bar{\lambda}(v \cdot w) + \bar{\gamma}(v \cdot u)$. In general, we define a **Hermitian inner product** (\cdot, \cdot) over a complex vector space to be a sesquilinear symmetric map which is positive definite. A Hermitian inner product space is just a space with a Hermitian inner product. A complex Hilbert space is just a Hermitian inner product space whose norm gives a complete metric space structure.

Example. If we allow for the existence of complex valued functions and sequences, then $L^2(X)$ and $l^2(\mathbf{Z})$ are Hermitian inner product spaces under the inner product

$$(f, g) = \int_X f(x) \overline{g(x)} dx$$

$$(f, g) = \sum_{n \in \mathbf{Z}} f(n) \overline{g(n)}$$

The conjugation is all that is needed to generalize these spaces to complex functions.

Example. We define the **Hardy space** $H^2(\mathbf{D}^\circ)$ to be all holomorphic functions $f : \mathbf{D}^\circ \rightarrow \mathbf{C}$ with

$$\sup_{0 \leq r < 1} \left(\frac{1}{2\pi} \int_{-\pi}^{\pi} |f(re^{it})|^2 dt \right)^{1/2} < \infty$$

We define this to be the norm on the space, induced from the inner product

$$(f, g) = \sup_{0 \leq r < 1} \frac{1}{2\pi} \int_{-\pi}^{\pi} f(re^{it}) \overline{g(re^{it})} dt$$

which is finite by the Cauchy Schwartz inequality on $L^2[-\pi, \pi]$, and easily verified to be sesquilinear and positive definite. In this definition, it is difficult

to write down what a real version of this space would look like. All these spaces are Hilbert spaces.

If W is a closed subspace of a Hilbert space H , in the sense that it is a subspace, and it is a closed set under the topology induced by the metric structure of the hermitian product, then W is also a Hilbert space, because a closed subset of a complete metric space is also complete. Conversely, a complete subspace is also closed.

Example. If D is an open subset of the complex plane, the space $L_2^a(D)$ of holomorphic square-integrable functions is a Hilbert space. If $B_r(x)$ is a ball of radius r centered at x whose closure is contained in D , and f is analytic in D , then the mean value property guarantees that

$$f(x) = \oint_{B_r(x)} f(y) \, dy$$

It follows from Cauchy Schwarz that if $0 < r < \text{dist}(x, \partial D)$, then

$$|f(x)| \leq \frac{1}{r\sqrt{\pi}} \|f\|_2$$

In particular, this implies that if a sequence of analytic functions in $L_2^a(D)$ converges in L^2 norm to some function, then they must also converge locally uniformly, so the resultant function is analytic. It follows that $L_2^a(D)$ is closed in $L^2(D)$, and so $L_2^a(D)$ is a Hilbert space.

The definition of a Hermitian inner product appears to avoid geometric intuition, but I will now attempt to explain why this definition is geometrically intuitive. As we have seen in real spaces, the normalized inner product

$$\frac{\langle v, w \rangle}{|w|^2}$$

is the number we must scale w by to obtain the projection of v onto the line spanned by the vector w . In complex vector spaces, the space spanned by a single vector is a *complex line*, or a one dimensional plane. If we set

$$\frac{\langle v, w \rangle}{|w|^2}$$

to be the value we must scale w by to obtain the projection of v onto the plane spanned by w , then we might have to rotate w , hence $\langle v, w \rangle$ might have a rotational part. Indeed, this is exactly the correct geometrical definition of the inner product, if we already have a well defined projection operation. The Pythagorean theorem, Cauchy Schwartz inequality, and triangle inequality goes through for Hermitian product spaces almost unperturbed, except that

$$\|v + w\|^2 = \|v\|^2 + 2\Re(\langle v, w \rangle) + \|w\|^2$$

The reason that everything for real inner product spaces ‘seems’ to work for Hermitian products, except for more advanced theorems which require we restrict ourselves only to Hermitian products, is one of the main reasons we push through the initial confusion.

Another approach to understanding Hermitian inner products is to use the fact that we can construct a Hermitian inner product from a real inner product. Recall that a complex vector space V is just a real vector space with a fixed twisting linear map $J : V \rightarrow V$ with $J^2 = -1$. If we already have a *real* inner product $\langle \cdot, \cdot \rangle$ on V , then our hope is to extend the real inner product to a complex inner product (\cdot, \cdot) , where the real part of the Hermitian inner product is our original real product, so that

$$\Re((v, w)) = \langle v, w \rangle$$

By geometric intuition, it is reasonable to assume that J is a twist by a right angle, so that $\langle v, Jv \rangle = 0$ for all vectors v , and that J is an isometry with respect to the inner product. It then follows that

$$\langle v, Jw \rangle = \langle Jv, J^2w \rangle = \langle -Jv, w \rangle$$

so we cannot hope for the extension to be complex linear, only sesquilinear. If (\cdot, \cdot) is a sesquilinear form, then we calculate

$$\Im(v, w) = \Re(-i(v, w)) = \Re(v, iw) = \langle v, iw \rangle$$

so the *only* way to define (\cdot, \cdot) properly is as

$$(v, w) = \langle v, w \rangle + i\langle v, iw \rangle$$

The fact that (v, w) measures the projection of v onto the complex line containing w appears again, because over the real numbers, the complex

line is spanned by w and Jw , and the value above, once normalized, gives the real projection onto this two dimensional plane.

If V is any real inner product space, then we can always complexify it by embedding V in $V \oplus V$ by the trivial map $v \mapsto (v, 0)$, and defining a real inner product $\langle (v_1, v_2), (w_1, w_2) \rangle = \langle v_1, w_1 \rangle + \langle v_2, w_2 \rangle$. We can then add an artificial twisting map $J(v, w) = (-w, v)$ gives $V \oplus V$ a complex structure, and since it satisfies the properties of the discussion above, the real inner product extends to give a complex inner product. The space with all these gadgets attached is called the **complexification** of V , denoted $V^{\mathbb{C}}$. If V is a Hilbert space, then so $V^{\mathbb{C}}$ will also be a Hilbert space. Because of this fact, we can prove all the theorems in the sequel for complex inner spaces, and then use the embedding to obtain the result for real inner product spaces. Unless otherwise mentioned, all inner product spaces in the sequel will be over the complex numbers.

2.3 Orthogonality

Because we can measure angles in a Hilbert space, we can also measure right angles. Two vectors v and w are **orthogonal**, denoted $v \perp w$, if $(v, w) = 0$. This is a symmetric relation, since if $(v, w) = 0$,

$$(w, v) = \overline{(v, w)} = \overline{0} = 0$$

The theory of right triangles essentially manifests in the theory of equations involving two to three vectors, as we have already seen in the case of the Pythagorean identity.

Theorem 2.5 (Parallelogram Law). *If x, y lie in a Hilbert space, then*

$$\|x + y\|^2 + \|x - y\|^2 = 2\|x\|^2 + 2\|y\|^2$$

which is obvious viewing $0, x, y$, and $x + y$ as vertices of a parallelogram.

Proof. We simply calculate

$$\begin{aligned} \|x + y\|^2 + \|x - y\|^2 &= \langle x + y, x + y \rangle + \langle x - y, x - y \rangle \\ &= 2\|x\|^2 + 2\|y\|^2 + \langle x, y \rangle + \langle y, x \rangle - \langle x, y \rangle - \langle y, x \rangle \\ &= 2\|x\|^2 + 2\|y\|^2 \end{aligned}$$

and this shows the theorem is true. □

A simple extension of this identity implies that if e_1, \dots, e_n are an *orthonormal* family of vectors in an inner product space, in the sense that they are pairwise orthogonal, and $\|e_i\| = 1$ for all i , then

$$\left\| \sum a_i e_i \right\|^2 = \sum |a_i|^2$$

Using slightly more analysis, and using the fact that we are working in a Hilbert space, we can extend this to an infinite sequence of vectors e_1, e_2, \dots , provided that $\sum |a_i|^2 < \infty$, because then the partial sums of $\sum_{k=1}^{\infty} a_i e_i$ are Cauchy, hence the partial sums converge in L^2 to some vector, and it is easy to see the norm must be the infinite sum $\sum |a_i|^2$. In particular, for any vector v , $v - \sum (v, e_i) e_i$ is orthogonal to the e_i , and so

$$\|v\|^2 = \left\| v - \sum (v, e_i) e_i \right\|^2 + \sum |(v, e_i)|^2 \geq \sum |(v, e_i)|^2$$

This is known as *Bessel's inequality*. We may also conclude that

$$\left\| \sum_{i=1}^{\infty} (v, e_i) e_i \right\|^2 = \sum_{i=1}^{\infty} |(v, e_i)|^2$$

If the left sum converges to v in the L^2 norm, then we obtain

$$\|v\|^2 = \sum_{i=1}^{\infty} |(v, e_i)|^2$$

this equation is known as *Parseval's identity*. Regardless, we can always take $n \rightarrow \infty$ in Bessel's inequality to obtain an infinite dimensional version of the inequality. We now give a series of conditions guaranteeing that the left sum does actually converge to v in the L^2 norm.

Theorem 2.6. *The following properties of an orthonormal sequence e_1, e_2, \dots are equivalent:*

- (a) *For any vector $v \in H$, $v = \sum_{i=1}^{\infty} (v, e_i) e_i$.*
- (b) *The finite linear combinations of the e_i are dense in H .*
- (c) *If $(v, e_i) = 0$ for all i , then $v = 0$.*
- (d) *Parseval's identity holds.*

and then we say the sequence is an **Orthonormal basis** for the space.

Proof. (a) clearly implies (b). Assuming (b), note that if $(v, e_i) = 0$ for all i , then in particular, $(v, w) = 0$ for all w which can be expressed as finite linear combinations of the e_i . But we may choose w_n with $w_n \rightarrow v$ in L^2 , and then since $|(v, w_n - v)| \leq \|v\| \|w_n - v\| \rightarrow 0$, we conclude that $0 = (v, w_n) = (v, v) + (v, w_n - v)$ converges to (v, v) , so $(v, v) = 0$, and hence $v = 0$. If v is given, then Bessel's inequality ensures that

$$\sum_{i=1}^{\infty} |(v, e_i)|^2 \leq \|v\|^2 < \infty$$

and so the partial sums of the sequence $\sum_{i=1}^{\infty} |(v, e_i)|^2$ are Cauchy, and because we are working in a Hilbert space, converge to some vector w . Now $v - w$ is orthogonal to e_i for each i , since

$$(w, e_i) = \lim_{j \rightarrow \infty} \sum_{j=1}^n (v, e_j)(e_j, e_i) = (v, e_i)$$

and therefore $v = w$, proving (a). (a) implies (d) is trivial, and if (d) holds, then Pythagoras' identity implies that

$$\|v - \sum_{i=1}^n (v, e_i) e_i\|^2 = \sum_{i=n+1}^{\infty} |(v, e_i)|^2 \rightarrow 0$$

so (a) holds. □

A topological space is *separable* if it has a countable dense subset. If H is a separable Hilbert space, then by performing the Gram Schmidt process to a countable dense subset, one finds an orthogonal sequence, which is easily verified to be an orthonormal basis for the Hilbert space. Thus we have proved that *every separable Hilbert space has a orthonormal basis*. With some more work on the geometry of Hilbert spaces, and the axiom of choice, we can show that every separable Hilbert space has an orthonormal basis, if we allow the bases to be uncountable. Before we do this, however, we consider some applications.

2.4 Fourier Series in L^2

First, let's consider the theory of Fourier series, which is the scenario where the abstract theory of Hilbert spaces found its place. Since $[-\pi, \pi]$ is a finite measure space, the space $L^2[-\pi, \pi]$ is contained within $L^1[-\pi, \pi]$, and so if $f \in L^2[-\pi, \pi]$, we can consider its Fourier series

$$\hat{f}(n) = \oint_{-\pi}^{\pi} f(x) e^{-nix} dx$$

Note that if we define $e_n(x) = e^{nix}$, then $\hat{f}(n) = (f, e_n)$, and

$$(e_n, e_m) = \oint_{-\pi}^{\pi} e^{(n-m)ix} = \begin{cases} 1 & n = m \\ 0 & n \neq m \end{cases}$$

It follows that e_n is an orthogonal sequence in $L^2[-\pi, \pi]$. Using the basic fact from Fourier analysis that for any continuous function $f \in L^2[-\pi, \pi]$ with $f(-\pi) = f(\pi)$, the sum

$$f(x) = \sum \hat{f}(n) e^{nix}$$

converges uniformly, and the fact that continuous functions are dense in $L^2[-\pi, \pi]$, we conclude that the exponentials e_n are an orthonormal basis for $L^2[-\pi, \pi]$. As a result, we obtain the classic Parseval's inequality

$$\oint_{-\pi}^{\pi} |f(x)|^2 dx = \sum |\hat{f}(n)|^2$$

and the L^2 convergence of the partial sums $\sum_{k=-N}^N \hat{f}(k) e^{kix}$ to $f(x)$.

2.5 Optimizing over Convex Sets

One of the biggest advantages of Hilbert spaces is that we have powerful theorems about the structure of convex sets in the space. A **convex set** in a vector space is a set C such that the line between any two points is contained in C . That is, for any $x, y \in C$, and $\lambda \in [0, 1]$, $\lambda x + (1 - \lambda)y \in C$.

Theorem 2.7. *If C is any closed convex subset of a Hilbert space, and $x \in H$ is arbitrary, then there is a unique point $y \in C$ which minimizes $\|x - y\|$, over all choices of y in the convex set.*

Proof. Since $C - x$ is a convex set, we may assume without loss of generality that $x = 0$, and so we are trying to minimize the norm over C . If $x, y \in C$ minimize this norm, with $\|x\| = \|y\| = d$, then

$$d \leq \left\| \frac{x+y}{2} \right\| \leq \frac{1}{2}(d+d) = \frac{d}{2}$$

Hence the parallelogram law implies

$$d^2 = \left\| \frac{x+y}{2} \right\|^2 = d^2 - \left\| \frac{x-y}{2} \right\|^2$$

which implies $x = y$. We know that there must a sequence $x_i \in C$ with $\|x_i\| \rightarrow d$ monotonically, and then the parallelogram law implies that if $\|x_n\|, \|x_m\| \leq d + \varepsilon$, then since $(x_n + x_m)/2 \in C$, then

$$\left\| \frac{x_n - x_m}{2} \right\|^2 = \frac{1}{2}\|x_n\|^2 + \frac{1}{2}\|x_m\|^2 - \left\| \frac{x_n + x_m}{2} \right\|^2 \leq 2d\varepsilon + \varepsilon^2 \rightarrow 0$$

hence $\{x_n\}$ is a Cauchy sequence, and therefore converges to a point x , which satisfies $\|x\| = d$. \square

Linear subspaces of a vector spaces are trivially verified to be convex sets, and we find that

Theorem 2.8. *If V is a closed linear subspace of a Hilbert space H , and $v \in V$ is the closest point to a fixed point x , then $v - x \perp V$, in the sense that $v - x \perp w$ for any $w \in V$. Conversely, if $v - x \perp V$, then v is the closest point to x .*

Proof. If $\langle v - x, w \rangle \neq 0$, then by multiplying w by a scalar value, we may assume that $\langle v - x, w \rangle < 0$, and $\|w\| = 1$, and then for $\alpha > 0$,

$$\|\alpha w + v - x\|^2 = \|v - x\|^2 + 2\alpha \Re \langle w, v - x \rangle + |\alpha|^2 \|w\|^2$$

and so as $\alpha \rightarrow 0$, we find that eventually $\|\alpha w + v - x\|^2 < \|v - x\|^2$. Conversely, if $\langle v - x, w \rangle = 0$ for all w , then for any w ,

$$\|w - x\|^2 = \|(w - v) + (v - x)\|^2 = \|w - v\|^2 + \|v - x\|^2 \geq \|v - x\|^2$$

so v minimizes the distance to x among all vectors in V . \square

For any subset X of an inner product space V , define X^\perp to be the set of vectors v such that $\langle x, v \rangle = 0$ for all $x \in X$. It is clearly a closed linear subspace of V . If V is a closed linear subspace of a Hilbert space H , then for any $x \in H$, there is a unique $y \in V$ with $y - x \perp V$, and we may define a map $P : H \rightarrow V$ such that $Px = y$, which we call the **orthogonal projection** of H onto V .

Theorem 2.9. *If $P : H \rightarrow V$ is defined as above, then*

1. P is a linear map.
2. $\|Px\| \leq \|x\|$.
3. $P^2 = P$.
4. $\ker P = V^\perp$, and $\operatorname{im} P = V$.

Proof. If $x_0 - y_0 \perp V$ and $x_1 - y_1 \perp V$, then $(\alpha x_0 - \alpha y_0) + (\beta x_1 - \beta y_1) \perp V$, hence $P(\alpha x_0 + \beta x_1) = \alpha y_0 + \beta y_1$, and so P is a linear map. It is obvious that P is a contraction, because $\|x\|^2 = \|Px\|^2 + \|x - Px\|^2$, and $\|x - Px\|^2 \geq 0$. For any $y \in V$, $y - y \perp V$, hence $P y = y$, and so $P^2 = P$. Finally, $P y = 0$ if and only if $y \perp V$, in which case we find $y \in V^\perp$. \square

Corollary 2.10. *For any closed subspace V of a Hilbert space H , $(V^\perp)^\perp = V$. If X is an arbitrary subset of a Hilbert space H , then $(X^\perp)^\perp = \overline{\operatorname{span}}(X)$ is the closure of the linear span of the elements of X .*

Proof. If P is the orthogonal projection of H onto V , and Q is the orthogonal projection of H onto V^\perp , then $Px + Qx = x$ for all $x \in H$, and so the kernel of Q is exactly the set of vectors x for which $Px = x$, from which we conclude that $x \in V$. If X is arbitrary, then $\overline{\operatorname{span}}(X)$ is contained in $(X^\perp)^\perp$, because $(X^\perp)^\perp$ is a closed subspace of H containing all elements in X . Conversely, we conclude that $X^\perp = \overline{\operatorname{span}}(X)^\perp = X^\perp$, because if $\langle y, x \rangle = 0$ for all $x \in X$, then surely $\langle y, \sum \alpha_i x_i \rangle = \sum \alpha_i \langle y, x_i \rangle = 0$, so $y \perp \operatorname{span}(X)$, and if $v_i \rightarrow v$, with $v_i \in \operatorname{span}(X)$, then by continuity of the inner product we find that $0 = \langle y, v_i \rangle \rightarrow \langle y, v \rangle$, hence $\langle y, v \rangle = 0$, and so $y \perp \overline{\operatorname{span}}(X)$. \square

Corollary 2.11. *A subspace V is dense in H if and only if $V^\perp = 0$.*

Proof. For then the closure of V is $(V^\perp)^\perp = 0^\perp = H$. \square

Given two Hilbert spaces H_0 and H_1 , we can define a canonical Hilbert space structure on $H_0 \oplus H_1$ by letting $\langle x_0 + x_1, y_0 + y_1 \rangle = \langle x_0, y_0 \rangle + \langle x_1, y_1 \rangle$. This is exactly the product topology, because $x_i + y_i \rightarrow x + y$ if and only if $x_i \rightarrow x$, and $y_i \rightarrow y$. The orthogonal projection of H onto V is incredibly important to the theory of Hilbert spaces, because it gives us a decomposition of H into the direct sum of V and V^\perp . The Pythagorean theorem tells us that for any $v \in V$ and $w \in V^\perp$,

$$\|v + w\|^2 = \|v\|^2 + \|w\|^2$$

and so the map $x \mapsto Px + (1 - P)x$ is actually an **isometry** between vector spaces (it preserves the inner product). This is the basic notion of equivalence between various Hilbert spaces, and so we really can think of H as being composed of the two closed subspaces V and V^\perp .

2.6 The Riesz Representation Theorem

The Riesz representation theorem shows that the isometry class of Hilbert spaces is uniquely identified by the dimension of a **Schauder basis** of the underlying Hilbert space. It is incredibly important, because it shows that to analyze a Hilbert space H with a Schauder basis of cardinality \mathfrak{a} , we need only analyze the Hilbert space $l_2(\mathfrak{a})$.

Many problems in functional analysis can be reduced to the analysis of certain linear functions between vector spaces. The easiest linear functions to analyze are the **linear functionals**, which are linear maps $f : V \rightarrow K$, where K is the base field of the vector space V . We will say a linear function $f : V \rightarrow K$ is **bounded** if there exists $M > 0$ such that $|f(x)| \leq M\|x\|$ for all $x \in V$. The bounded linear maps are exactly the continuous linear maps (and to verify continuity, we need only verify continuity at zero, so bounded linear maps are essentially ‘uniformly continuous’). If f and g are continuous, then $\lambda f + \gamma g$ is continuous, and so the set V^* of bounded linear functionals on V is a vector space, called the **dual space** of V . On the dual space of a norm space, we may define a dual norm by letting $\|f\|$ be the smallest number M such that $|f(x)| \leq M\|x\|$ holds for all x . Since $|f(x) + g(x)| \leq |f(x)| + |g(x)|$, we find $\|f + g\| \leq \|f\| + \|g\|$, and it is easy to verify that $\|\lambda f\| = |\lambda|\|f\|$.

Given a Hilbert space H , we can use the inner product structure on H to transport vectors in H to functionals in H^* . That is, for a given $x \in$

H , we define $x^* \in H^*$ to be the functional $x^*(y) = \langle x, y \rangle$. The Cauchy-Schwarz inequality tells us that $\|x^*\| = \|x\|$, so the dual map is an isometry. The Riesz representation theorem tells us that this isometry is actually invertible.

Theorem 2.12 (Riesz). *For any bounded linear function $f \in H^*$, there exists a unique vector $x \in H$ such that $f(y) = \langle x, y \rangle$ for all $y \in H$.*

Proof. Assume without loss of generality that $f \neq 0$. Let V be the kernel of f . Then V is a closed linear subspace of H , and V^\perp is one dimensional by the first isomorphism theorem. Consider the orthogonal projection P onto V^\perp . If x is a nonzero vector in V^\perp , with $\|x\| = 1$ and $f(x) = \gamma$, then for each y there is $\lambda(y)$ for which $f(y) = f(Py) = \lambda(y)$, and what's more, $\langle y, x \rangle = \langle Py, x \rangle = \lambda(y)$, and so we have found a unique y as specified, and y is unique because the dual map is an isometry. \square

Example. *For any measure space X , $L_2(X)$ is a Hilbert space, and so for any bounded linear function F on $L_2(X)$, there is a unique $g \in X$ such*

$$F(f) = \int f \bar{g}$$

so studying bounded linear functionals reduces to integration theory.

2.7 Constructing an Orthonormal Basis

In the theory of infinite dimension vector spaces, the standard definition of a basis does not lead to a natural theory, but we can obtain a much more elegant decomposition of a space by instead talking about the vector space generated by infinite linear combinations of elements of some infinite set, rather than linear combinations of a finite subset of the set. A **Schauder basis** for a norm space X is a set Y such that any $x \in X$ has a unique expansion as $x = \sum_{y \in Y} \alpha_y y$, where the sum is allowed to be infinite, and converges to x in the limit (if Y is uncountable, this sum is defined to be the limit of the net of all finite sums, which is essentially the integral with respect to the counting measure. If the sum exists, its support must be countable, and therefore this kind of sum isn't really that much more general than absolute convergence). The finite dimensional notion of a

basis will be called a **Hamel basis**, and when we refer to ‘a basis’ in a norm space, we will be referring to a Schauder basis.

Given an inner product space V , an **orthonormal set** in V is a set X such that $\langle x, y \rangle = \mathbf{I}(x = y)$ for all $x, y \in X$. An **orthonormal basis** for an inner product space V is a Schauder basis for X which forms an orthonormal set. It is clear that an orthonormal set Y is automatically linearly independent, for if $\sum a_y y = 0$, then $0 = \langle \sum a_y y, \sum a_y y \rangle = \sum |a_y|^2$, so $a_y = 0$ for all $y \in Y$.

Lemma 2.13 (Bessel’s Inequality). *If Y is an orthonormal set in an inner product space V , then for any $x \in V$,*

$$\sum |\langle x, y \rangle|^2 \leq \|x\|^2$$

hence in particular, $\sum |\langle x, y \rangle|^2 < \infty$.

Proof. For any $x \in V$, and any finite subset X of Y , $x - \sum_{y \in X} \langle x, y \rangle y$ is perpendicular to the span of the X , because for any $z \in X$,

$$\langle x - \sum_{y \in X} \langle x, y \rangle y, z \rangle = \langle x, z \rangle - \sum_{y \in X} \langle x, y \rangle \langle y, z \rangle = \langle x, z \rangle - \langle x, z \rangle = 0$$

Hence we may apply Pythagoras’ theorem to conclude that

$$\|x\|^2 = \left\| \sum_{y \in X} \langle x, y \rangle y \right\|^2 + \left\| x - \sum_{y \in X} \langle x, y \rangle y \right\|^2 \geq \sum_{y \in X} |\langle x, y \rangle|^2$$

and the general inequality holds by taking limits of the finite sums. \square

Bessel’s inequality is an equality only where $x = \sum \langle x, y \rangle y$, a fact that will soon become important. In particular, we shall find that Bessel’s inequality is always an equality when Y is an orthonormal basis, in which case we call the equation **Parseval’s equality**.

Lemma 2.14. *If Y is an orthonormal set in a Hilbert space H , then for any x , $\sum \langle x, y \rangle y$ converges in H .*

Proof. First, note that for any x , since $\sum |\langle x, y \rangle|^2 < \infty$, $\langle x, y \rangle \neq 0$ for only countably many $y \in Y$, so we may assume that Y is countable from the getgo. If we fix some ordering $Y = \{y_0, y_1, \dots\}$, then the sum

$$\sum_{k=0}^{\infty} \langle x, y_k \rangle y_k$$

converges absolutely, in the sense that

$$\sum_{k=0}^{\infty} \|\langle x, y_k \rangle y_k\|^2 = \sum_{k=0}^{\infty} |\langle x, y_k \rangle|^2 < \infty$$

and it therefore follows from basic estimates analogous to the theory of absolute convergence over the real numbers, that the entire sum converges, to the same value irrespective to the ordering we placed on Y . \square

The set of all ‘possibly infinite’ linear combinations of an orthonormal set Y forms a closed subspace V of an inner product space (it is the closure of the set of finite sums). For any $x \in H$, we find that the projection of x onto V is $Px = \sum \langle x, y \rangle y$, because for any $z \in Y$,

$$\langle x - \sum \langle x, y \rangle y, z \rangle = \langle x, z \rangle - \sum \langle x, y \rangle \langle y, z \rangle = \langle x, z \rangle - \langle x, z \rangle = 0$$

Now suppose that Y is an orthonormal basis for H , the closed span of Y is H itself, and therefore the projection onto the closed span is just the identity map, and so we obtain Parseval’s equality as we stated before, because then $x = Px = \sum \langle x, y \rangle y$.

Lemma 2.15. *An orthonormal set of linearly independent elements of a Hilbert space H is an orthonormal basis if and only if the set cannot be expanded to a larger set of linearly independent orthonormal elements.*

Proof. First, assume that Y is an orthonormal basis. If $x \in H$ satisfies $\langle x, y \rangle = 0$ for all $y \in Y$, and $x = \sum a_y y$, then for each $y \in Y$, $0 = \langle x, y \rangle = a_y$, hence $x = 0$. Conversely, suppose that Y is an orthonormal set, and $x \in H$ is linearly independent to Y . Then $x - \sum \langle x, y \rangle y \perp Y$, is linearly independent to Y , and if we normalize so that $\|x - \sum \langle x, y \rangle y\|^2 = 1$, we find that $Y \cup \{x - \sum \langle x, y \rangle y\}$ is an orthonormal set larger than Y . \square

On a finite dimensional Hilbert space H , we may repeat this process to construct an orthonormal basis on H , and the constructive process above is called **Gram-Schmidt orthogonalization**. If H is not finite dimensional, we have to rely on non-constructive measures. If $\{Y_\alpha\}$ is a chain of orthonormal sets, then $\bigcup Y_\alpha$ is orthonormal, so we may apply Zorn's lemma to conclude that a maximal orthonormal set exists, and this must be an orthonormal basis. Thus orthonormal bases exist on any Hilbert space, and we can always extend any orthonormal set to a basis.

Theorem 2.16. *Any two orthonormal bases of a Hilbert space H have the same cardinality.*

Proof. Let X and Y be two bases of a Hilbert space H . If H is finite dimensional, this is just standard linear algebra. If X has infinite cardinality \mathfrak{a} , and Y has cardinality \mathfrak{b} , then for each $x \in X$, the set Y_x of $y \in Y$ with $\langle x, y \rangle \neq 0$ is at most countable. But then $Y' = \bigcup Y_x$ has cardinality at most \mathfrak{a} . If $y \in Y' - Y$, then $y = \sum \langle y, x \rangle x = 0$, so $Y' = Y$, and therefore $\mathfrak{a} \geq \mathfrak{b}$. By symmetry, we conclude that $\mathfrak{a} = \mathfrak{b}$. \square

We may therefore define the **dimension** of a Hilbert space to be the cardinality of any orthonormal basis. This actually uniquely defines the Hilbert space, for if H_0 and H_1 are Hilbert spaces with orthonormal bases X and Y , then a bijection $f : X \rightarrow Y$ induces a map $T : H_0 \rightarrow H_1$ defined by $T(\sum a_x x) = \sum a_x f(x)$. This is well defined, for the sum $\sum a_x x$ converges if and only if $\sum |a_x|^2 < \infty$, and the same criterion shows $\sum a_x f(x)$ converges. By linear independence, and the invertibility of f , we find T is injective and surjective, hence an isomorphism. It is actually an isometry, because

$$\left\| \sum a_x x \right\|^2 = \sum |a_x|^2 = \left\| \sum a_x f(x) \right\|^2$$

hence T is an isometry, and so the Hilbert spaces are isomorphic.

Example. On $l_2(X)$, we have an orthonormal basis $e_x(y) = \mathbf{I}(x = y)$, because certainly $\langle e_x, e_y \rangle = \mathbf{I}(x = y)$, and for any $f \in l_2(X)$, $f(x) = 0$ for all but countably many $x \in X$, because $\sum \|f(x)\|^2 < \infty$, and so we may assume that X is countable from the getgo, writing $X = \{x_0, x_1, \dots\}$, thinking of f as a sequence of numbers as in l_2 , and then

$$f - \sum_{k=0}^{\infty} \langle f, e_{x_k} \rangle e_{x_k} = \lim_{n \rightarrow \infty} (0, 0, \dots, x_n, x_{n+1}, \dots)$$

and since $\sum \|f(x)\|^2 < \infty$, $\lim_{n \rightarrow \infty} \sum_{k=n}^{\infty} \|f(x_k)\|^2 \rightarrow 0$. Thus the family $\{e_x\}$ really is an orthonormal basis for $l_2(X)$. If X has cardinality \mathfrak{a} , then $l_2(X)$ has dimension \mathfrak{a} , and so we find from the previous discussion that every Hilbert space is isometric to $l_2(X)$ for some set X .

Example. The space $L_2(\mathbf{T})$ of complex-valued square summable functions with respect to the Haar measure on \mathbf{T} has an orthonormal basis consisting of the functions $z \mapsto z^n$, for $n \in \mathbf{Z}$. This follows because

$$\langle z^n, z^m \rangle = \int z^{n-m} = \frac{1}{2\pi} \int_0^{2\pi} e^{i(n-m)t} = \frac{1}{2\pi} \int_0^{2\pi} \cos((n-m)t) + i \sin((n-m)t)$$

and so $\langle z^n, z^m \rangle = \delta_{nm}$. It requires a little bit of work to prove that this is actually an orthonormal basis for $L_2(\mathbf{T})$ (see my notes on elementary Harmonic analysis), so that every square summable function $f : \mathbf{T} \rightarrow \mathbf{C}$ has a unique Fourier expansion

$$f(z) = \sum_{n=0}^{\infty} \hat{f}(n) z^n$$

where $\hat{f}(n)$ is the n 'th Fourier coefficient of f , and the convergence is in $L_2(\mathbf{T})$. The fact that $\sum |\hat{f}(n)|^2 < \infty$ is obvious from our Hilbert space theory, and this shows that the Riemann Lebesgue lemma holds for square summable functions. That is,

$$\lim_{n \rightarrow \infty} \int f(z) z^{-n} \rightarrow 0$$

The map $f \mapsto \hat{f}$ is an isometry between $L_2(\mathbf{T})$ and l_2 . In the case that $L_2(\mathbf{T})$ is the space of real-valued functions, which is isometric to $L_2[0, 1]$ under the association $t \mapsto e^{it}$, and has an orthonormal basis consisting of the maps $\sin(nt)$ and $\cos(mt)$, for $n, m \in \mathbf{Z}$.

2.8 Convexity

A function $f : (a, b) \rightarrow \mathbf{R}$ is convex when the line segment between $(a, f(a))$ and $(b, f(b))$ lies above the graph of f . The line segment connecting these two points is described by

$$\{(\lambda a + (1 - \lambda)b, \lambda f(a) + (1 - \lambda)f(b)) : 0 \leq \lambda \leq 1\}$$

and we require that $(\lambda a + (1 - \lambda)b, f(\lambda a + (1 - \lambda)b))$ lies below the corresponding point on the line defined above. This is satisfied exactly when we have a certain inequality:

A function $f : U \rightarrow \mathbf{R}$ is **convex** on (a, b) if, for any $a \leq x < y \leq b$, and $0 \leq \lambda \leq 1$, we have

$$f(\lambda x + (1 - \lambda)y) \leq \lambda f(x) + (1 - \lambda)f(y) \quad (2.1)$$

By rewording the definition, convexity is also satisfied when, for $a \leq x < y < z \leq b$,

$$\frac{f(y) - f(x)}{y - x} \leq \frac{f(z) - f(y)}{z - y} \quad (2.2)$$

Geometrically, this equation says that the slope of the tangent line from $(x, f(x))$ to $(y, f(y))$ is smaller than the tangent line from $(y, f(y))$ to $(z, f(z))$.

Lemma 2.17. *A C^1 convex function's derivative is non-decreasing on (a, b) .*

Proof. Suppose f is a C^1 function, and f' is non-decreasing, then consider any $a \leq x < y < z \leq b$. Then $f'(x) \leq f'(y) \leq f'(z)$. Applying the mean value theorem, we conclude there is $t \in (x, y)$, and $u \in (y, z)$, for which

$$f'(t) = \frac{f(y) - f(x)}{y - x} \quad f'(u) = \frac{f(z) - f(y)}{z - y} \quad (2.3)$$

And since $t < u$, (2.3) implies $f'(t) \leq f'(u)$, i.e. (2.2) is satisfied.

If f is C^1 and convex, then surely f' is non-decreasing. Fix $a \leq x < y \leq b$. In (2.2), letting y converge to x , we obtain

$$f'(x) = \lim_{y \rightarrow x^+} \frac{f(y) - f(x)}{y - x} \leq \lim_{y \rightarrow x} \frac{f(z) - f(y)}{z - y} = \frac{f(z) - f(x)}{z - x} \quad (2.4)$$

Conversely, letting $y \rightarrow z$, we obtain

$$f'(y) = \lim_{y \rightarrow z^-} \frac{f(z) - f(y)}{z - y} \geq \lim_{y \rightarrow z} \frac{f(y) - f(x)}{y - x} = \frac{f(z) - f(x)}{z - x} \quad (2.5)$$

And in tandem, (2.4), (2.5) and (2.2) imply that $f'(x) \leq f'(y)$. \square

Example. $\exp : \mathbf{R} \rightarrow \mathbf{R}$ is convex on $(-\infty, \infty)$, since $\exp'' = \exp > 0$.

Lemma 2.18. *A function is continuous on the open segments where it is convex.*

The most important inequality in analysis is the triangle inequality, undershadowed by the Schwarz inequality. Almost as important is Jensen's inequality. Despite its importance, the proof is fairly simple and intuitive. Consider the center of mass of an object.

Theorem 2.19 (Jensen's Inequality). *Let (Ω, \mathbf{P}) be a probability space. If $X \in L^1(\mathbf{P})$, where $a < X < b$, and if f is a real function, convex on (a, b) , then*

$$f(\mathbf{E}[X]) \leq \mathbf{E}[f \circ X] \quad (2.6)$$

Proof. Since $a < X < b$, $a < \mathbf{E}[X] < b$. Let

$$\beta = \sup_{a < x < \mathbf{E}[X]} \frac{g(\mathbf{E}[X]) - g(x)}{\mathbf{E}[X] - x}$$

For any $a < x < y$, $g(x) + \beta(\mathbf{E}[X] - x) \geq g(y)$. But also, by the right side of (2.2), for any $z > y$, $g(z) - \beta(z - \mathbf{E}[X]) \geq g(\mathbf{E}[X])$. For any $\omega \in \Omega$, we may restate these equations as as

$$g(f(\omega)) - \beta(f(\omega) - y) \geq g(y)$$

But then taking expectations, we obtain (2.6). \square

Jensen's inequality is incredibly useful. To see this, consider some examples.

Example. *We have seen the exponential function is convex. Hence for any $L_1(\mathbf{P})$, where \mathbf{P} is a probability measure, we have*

$$\exp\left(\int f d\mathbf{P}\right) \leq \int e^f d\mathbf{P} \quad (2.7)$$

Let \mathbf{P} be the uniform measure over a finite set $\{x_1, x_2, \dots, x_n\}$. Then (2.7) tells us that

$$e^{f(x_1)/n} e^{f(x_2)/n} \dots e^{f(x_n)/n} \leq \frac{\sum e^{f(x_i)}}{n}$$

If we let x_i be real numbers, and $f(x_i) = \log(x_i)$, we conclude

$$(x_1 x_2 \dots x_n)^{1/n} \leq \frac{\sum x_i}{n}$$

In other words, the geometric mean is always smaller than the arithmetic mean.

Jensen's inequality implies so many other inequalities in analysis.

Let $1 \leq p \leq \infty$. We say $1 \leq q \leq \infty$ is the **conjugate** of p if $p^{-1} + q^{-1} = 1$, and write $q = p'$.

Theorem 2.20 (Hölder). *If $p' = q$, (Ω, μ) is a measure space, and f, g are positive measurable functions on Ω , then*

$$\|fg\|_1 \leq \|f\|_p \|g\|_q \quad (2.8)$$

Proof. Let A, B be the values on the right hand side of (2.8)). If $A = 0$, then $f = 0$ almost everywhere, so that the theorem is trivial. Symmetry shows that the same is true if $B = 0$, so assume $A, B \neq 0$. Define $F = f/A$, and $G = g/B$. Then

$$\left(\int F^p d\mu \right)^{1/p} = A^{-1} \left(\int f^p d\mu \right)^{1/p} = 1 \quad \left(\int G^q d\mu \right)^{1/q} = B^{-1} \left(\int g^q d\mu \right)^{1/q} = 1$$

For any x , there is a number s such that $e^{s/p} = F(x)$, and t such that $e^{t/q} = G(x)$. By the convexity of the exponential function, $e^{s/p+t/q} \leq e^s p^{-1} + e^t q^{-1}$. Thus $FG \leq p^{-1} F^p + q^{-1} G^q$, and thus by integrating,

$$\int FG d\mu \leq p^{-1} + q^{-1} = 1$$

□

Corollary 2.21 (Minkowski). *For*

$$\|f + g\|_p \leq \|f\|_p + \|g\|_p$$

2.9 The Dual of a Hilbert Space

Lemma 2.22. *Fix $x, z \in H$. If $\langle x, y \rangle = \langle z, y \rangle$ for all $y \in H$, then $x = z$.*

Proof. For then $\langle x - z, y \rangle = 0$ for all y , and in particular, $\langle x - z, x - z \rangle = 0$. □

Chapter 3

Operator Theory

3.1 Operators on a Hilbert Space

If H is a Hilbert space, then H^* is a dull object to study, for it is essentially H . When one says they study Hilbert spaces, they really study the bounded operators from one space to another. This is the task we begin in this section. Given two Hilbert spaces, we let $B(H, K)$ denote the continuous linear maps from H to K .

Lemma 3.1. *If (\cdot, \cdot) is a bounded¹, sesquilinear form on a Hilbert space, then there is a unique $M \in B(H)$ such that $(x, y) = \langle x, My \rangle$.*

Proof. For each x , (\cdot, x) is a bounded linear functional, so there is a unique $Mx \in H$ such that $(y, x) = \langle y, Mx \rangle$. Since

$$(z, \lambda x + \gamma y) = \overline{\lambda}(z, x) + \overline{\gamma}(z, y) = \langle z, \lambda Mx + \gamma My \rangle$$

M is linear, because $M(\lambda x + \gamma y) = \lambda Mx + \gamma My$. M is certainly uniquely defined, and

$$\langle x, My \rangle = (x, y) \leq \|(\cdot, \cdot)\| \|x\| \|y\|$$

So $\|M\| \leq \|(\cdot, \cdot)\|$, but also

$$(x, y) = \langle x, My \rangle \leq \|M\| \|x\| \|y\|$$

so $\|(\cdot, \cdot)\| \leq \|M\|$, and we have equality of norms. □

¹In the sense that there is a ‘sesquilinear norm’ $\|(\cdot, \cdot)\|$, which is the smallest number such that $(x, y) \leq \|x\| \|y\| \|(\cdot, \cdot)\|$.

Theorem 3.2. Let H and K be Hilbert spaces. If $M \in B(H, K)$, there is a unique map $M^* \in B(K, H)$ such that

$$\langle Mx, y \rangle = \langle x, M^*y \rangle$$

M^* is known as the **adjoint** of M . We also have $\|M^*\| = \|M\|$.

Proof. The map $(x, y) = \langle Mx, y \rangle$ is a sesquilinear form, so there is a unique M^* such that $(x, y) = \langle x, M^*y \rangle$. \square

Example. If we are working in \mathbf{R}^n , then every operator $B(\mathbf{R}^n, \mathbf{R}^m)$ can be identified with a unique matrix in $M_{n,m}(\mathbf{R})$. If (a_{ij}) is that operator, then the adjoint is the transpose $(a_{ji}) \in M_{m,n}(\mathbf{R})$. In \mathbf{C}^n , the adjoint is the conjugate transpose $(\overline{a_{ji}})$.

Example. Given $\phi \in L^\infty(X)$, where X is a σ -finite measure space, we have an operator $M_\phi \in B(L^2(X))$ defined by $M_\phi(f) = \phi f$. Then

$$\langle M_\phi f, g \rangle = \int \phi f \overline{g} = \int f \overline{g \overline{\phi}} = \langle f, \overline{\phi} g \rangle$$

Thus $M_\phi^* = M_{\overline{\phi}}$.

The adjoint operation $*$ is an antilinear operator from $B(H, K)$ to $B(K, H)$, such that $(ST)^* = T^*S^*$. This is verified by calculation, since

$$\langle STx, y \rangle = \langle Tx, S^*y \rangle = \langle x, T^*S^*y \rangle$$

We also have $\|T^*T\| = \|T\|^2$, since

$$\langle T^*Tx, x \rangle = \langle Tx, Tx \rangle = \|Tx\|^2$$

so $\|T^*T\| \leq \|T\|^2$, and this inequality is attained in the suprema, since if $x_i \in B_H$ is such that $\|Tx_i\| \rightarrow \|T\|$, then

$$\langle T^*Tx_i, x_i \rangle = \|Tx_i\|^2 \rightarrow \|T\|^2$$

Thus $*$ is what is known as an involution.

We can describe some operators from a Hilbert space to itself by their action under the involution. T is a **self adjoint operator** if $T^* = T$. This is something like T being symmetric, but in the infinite dimensional case. T is unitary if $T^* = T^{-1}$. These are analogous to rotations, but in infinite

dimensions. T is **normal** if $T^*T = TT^*$. Any operator M can be expressed as $M = T + iS$, where T and S are self adjoint operators, by defining

$$T = \frac{1}{2}(M + M^*) \quad S = \frac{1}{2i}(M - M^*)$$

We shall show this is a unique expression. This is essentially like beginning with \mathbf{C} , and identifying \mathbf{R} as those elements invariant under complex conjugation.

Theorem 3.3. *If T is a self-adjoint operator on a complex Hilbert space, then*

$$\|T\| = \sup\{\langle Tx, x \rangle : \|x\| \leq 1\}$$

Proof. Let M be the supremum. Clearly $M \leq \|T\|$. If $\|x\|, \|y\| \leq 1$,

$$\begin{aligned} \langle T(x \pm y), x \pm y \rangle &= \langle Tx, x \rangle \pm \langle Ty, x \rangle \pm \langle Tx, y \rangle + \langle Ty, y \rangle \\ &= \langle Tx, x \rangle \pm \overline{\langle Tx, y \rangle} \pm \langle Tx, y \rangle + \langle Ty, y \rangle \\ &= \langle Tx, x \rangle \pm 2\Re\langle Tx, y \rangle + \langle Ty, y \rangle \end{aligned}$$

Then, subtracting, we find

$$4\Re\langle Tx, y \rangle = \langle T(x + y), x + y \rangle - \langle T(x - y), x - y \rangle$$

Thus

$$4\Re\langle Tx, y \rangle \leq M(\|x + y\|^2 + \|x - y\|^2) = 2M(\|x\|^2 + \|y\|^2) \leq 4M$$

Let λ be such that $\langle Tx, y \rangle = \lambda|\langle Tx, y \rangle|$. If we replace x by λx , we find

$$|\langle Tx, y \rangle| = \overline{\lambda}\langle Tx, y \rangle = \langle T(\overline{\lambda}x), y \rangle \leq M$$

Hence

$$\|Tx\| = \sup\{|\langle Tx, y \rangle| : y \in B_H\} \leq M$$

and therefore $\|T\| \leq M$. □

Corollary 3.4. $\langle Tx, x \rangle = 0$ if and only if $T = 0$.

Corollary 3.5. If $\langle Tx, x \rangle = \langle Sx, x \rangle$ for each $x \in H$, then $T = S$.

This theorem is false for real spaces, for instance, if we take a rotation operator on the plane.

Theorem 3.6. *T is normal if and only if $\|Tx\| = \|T^*x\|$ for each x .*

Proof. For then, for each x ,

$$\|Tx\|^2 = \langle Tx, Tx \rangle = \langle T^*Tx, x \rangle$$

$$\|T^*x\|^2 = \langle T^*x, T^*x \rangle = \langle TT^*x, x \rangle$$

which implies $T^*T = TT^*$. The converse repeats backwards through the proof, since the norm of $\|T\|$ is specified by its action through the inner product. \square

Theorem 3.7. *Any normal operator T has the following properties.*

- (a) $\ker(T) = \ker(T^*)$.
- (b) $T(H)$ is dense in H if and only if T is injective.
- (c) T is a surjective isomorphism if and only if $\|Tx\| \geq C\|x\|$ for some C .
- (d) If $Tx = \lambda x$, then $T^*x = \bar{\lambda}x$.
- (e) If λ and γ are distinct eigenvalues, then the eigenspaces are orthogonal to one another.

Proof. To prove (a) notice that $\|Tx\| = 0$ if and only if $\|T^*x\| = 0$. (b) follows since

$$T(H)^\perp = \ker(T^*) = \ker(T)$$

If $\|Tx\| \geq C\|x\|$, then $T(H)$ is closed, and dense, therefore $T(H) = H$. The converse follows by definition. If T is normal, then $\lambda - T$ is also normal, since $(\lambda - T)^* = \bar{\lambda} - T^*$, so $\ker(\lambda - T) = \ker(\lambda - \bar{\lambda})$. Finally, if $Tx = \lambda x$, and $Ty = \gamma y$, then

$$\lambda \langle x, y \rangle = \langle Tx, y \rangle = \langle x, T^*y \rangle = \langle x, \bar{\gamma}y \rangle = \gamma \langle x, y \rangle$$

so if $\lambda \neq \gamma$, then $\langle x, y \rangle = 0$. \square

Theorem 3.8. *If $U \in B(H)$, then the following are equivalent.*

- (a) U is unitary.
- (b) $GL(H) = H$ and $\langle Ux, Uy \rangle = \langle x, y \rangle$.

(c) $GL(H) = H$ and $\|Ux\| = \|x\|$, so U is an isometry from H to itself.

Proof. If U is unitary, then U is invertible, so $GL(H) = H$, and

$$\langle Ux, Uy \rangle = \langle U^* Ux, y \rangle = \langle x, y \rangle$$

If (b) holds, then

$$\|Ux\|^2 = \langle Ux, Ux \rangle = \langle x, x \rangle = \|x\|^2$$

If (c) holds, then

$$\langle U^* Ux, x \rangle = \langle Ux, Ux \rangle = \|Ux\|^2 = \|x\|^2 = \langle x, x \rangle$$

Thus $U^* U$ is the identity. Since U is injective and surjective, the open mapping theorem tells us U is invertible, so $U^* = U^{-1}$. \square

This shows that unitary operators are the natural automorphisms in the category of Hilbert spaces.

Lemma 3.9. *An operator $T \in B(H)$ in a complex Hilbert space is self adjoint if and only if $\langle Tx, x \rangle \in \mathbf{R}$ for each $x \in H$.*

Proof. If T is self adjoint, then

$$\langle Tx, x \rangle = \langle x, Tx \rangle = \overline{\langle Tx, x \rangle}$$

Conversely, if $\langle Tx, x \rangle \in \mathbf{R}$ for each $x \in H$, then

$$\langle Tx, x \rangle = \overline{\langle x, Tx \rangle} = \langle x, Tx \rangle = \langle T^* x, x \rangle$$

for each x , from which we conclude $T = T^*$. \square

Theorem 3.10. *If P is a projection, then the following are equivalent:*

(a) P is self-adjoint.

(b) P is normal.

(c) $P(H) = \ker(P)^\perp$.

(d) $\langle Px, x \rangle = \|Px\|^2$.

Proof. If P is self-adjoint, then P is trivially normal. If P is normal, then

$$P(H) = \ker(P^*)^\perp = \ker(P)^\perp$$

If (c) holds, then every $x = y + Px$, where $Py = 0$. Then

$$\langle P(y + Px), y + Px \rangle = \langle Px, y + Px \rangle = \langle Px, Px \rangle = \|Px\|^2$$

If (d) holds, then $\langle Px, x \rangle \in \mathbf{R}$ for each x , so P is self-adjoint. If (c) holds, then $\langle Px, x \rangle$ \square

3.2 Compact Operators

We already know operators on an infinite dimensional Banach space are more tricky than their finite dimensional counterparts. Thus we restrict our attention to operators which do not ‘spread themselves out’ too much. An operator is **compact** if the image of any bounded set is precompact. It suffices to verify that the image of the unit ball is precompact. Given two Banach spaces X and Y , we shall let $K(X, Y)$ denote the space of all compact operators from X to Y .

Theorem 3.11. *$K(X, Y)$ is a closed subspace of $B(X, Y)$, such that for any $M \in K(X, Y)$, $N \in B(Y, Z)$, and $L \in B(Z', X)$, $NML \in K(Z', Z)$.*

Proof. Suppose M and N are compact. Let U be a bounded subset of X . Then

$$(\lambda M + \gamma N)(U) = \lambda M(U) + \gamma N(U)$$

And the sum of two precompact sets is precompact. Thus $\lambda M + \gamma N$ is precompact. If U is a bounded subset of Z' , then $L(U)$ is a bounded subset of X , so $ML(U)$ is a precompact subset of Y , implying $NML(U)$ is a precompact subset of Z , for if every sequence y_i has a convergent subsequence, so must $N(y_i)$.

To prove $K(X, Y)$ is closed we rely on the fact that a subset of a complete metric space is precompact if and only if, for any $\varepsilon > 0$, the subset can be covered by finitely many balls of radius ε . Now suppose we have $M_1, M_2, \dots \in K(X, Y)$, and $M_i \rightarrow N$. We claim N is compact. Let U be a bounded subset of X , such that $\|x\| < K$ for each $x \in U$. Then for each $\varepsilon > 0$, there are finitely many balls of radius ε which cover $M_i(U)$, for each i . If $\|M_i - N\| < \varepsilon/K$, then $\|(M_i - N)(x)\| < \varepsilon$ for each $x \in U$, so that if $M_i x$ is contained in some ball with center y of radius ε , then Nx is contained in the ball with center y and radius 2ε . Thus $N(U)$ can be covered with finitely many balls of radius 2ε , and is therefore precompact. \square

Corollary 3.12. *$K(X)$ is a closed, two-sided ideal of $B(X)$.*

The best examples of compact operators were found in function spaces, and the best way to verify compactness in these spaces is to use the Arzela-Ascoli criterion. Say a family of functions $\{g_\alpha : X \rightarrow \mathbf{C}\}$ is **equicontinuous** at $x \in X$ if, for any $\varepsilon > 0$, there is an open set U containing x such that $\|g(a) - g(b)\| < \varepsilon$ for $a, b \in U$.

Theorem 3.13 (Arzela-Ascoli). *The precompact subsets of $l^\infty(X)$ are precisely those subsets which are equicontinuous, when X is compact.*

Proof. Suppose C is a precompact subset of $l^\infty(X)$. □

Example. *The best examples of compact operators, and the operators for which most of the theory was developed, are integral equations. Given two Hausdorff spaces A and B , equipped with respective measures μ and ν , consider some measurable function $K : A \times B \rightarrow \mathbf{C}$, and consider the integral operator \mathbf{K} mapping a measurable function f to a function on A defined by*

$$\mathbf{K}f(x) = \int_A K(x, y) f(y) d\mu(y)$$

*K is known as the **kernel** of \mathbf{K} . We assume that K has been chosen such that $\mathbf{K}f$ is measurable for each measurable f . If A and B are compact, and K is continuous, then \mathbf{K} is a compact operator from $L^1(A)$ to $L^1(B)$.*

It is a theorem of Calkin that $K(X)$ is the only closed, two-sided ideal of $B(X)$, when X is a Hilbert space. Thus we are led to consider the **Calkin Algebra** $B(X)/K(X)$, which is simple, and quite useful in K -theory.

A **finite-rank** operator is an operator whose range is finite-dimensional. It is clear that every finite-rank operator is compact. We shall eventually show that the set $F(X)$ of all finite-rank operators is dense in $K(X)$, when X is a Hilbert space.

In Linear Algebra, we know we may diagonalize any symmetric matrix. Here we apply our knowledge of Hilbert spaces to extend this to infinite dimensional spaces, for ‘symmetric’ operators which are ‘almost finite dimensional’.

Theorem 3.14. *If X is a Banach space, and Y is a closed subspace, M is a compact operator on X , and $\lambda \neq 0$ is such that*

$$\inf\{\|(\lambda - M)y : y \in B_Y\|\} = 0$$

then $Y \cap \ker(\lambda - M) \neq 0$.

Proof. Let y_i be a sequence in B_Y such that $(\lambda - M)y_i \rightarrow 0$. Since M is a compact operator, we may assume that My_i converges to some z . But then, by continuity,

$$\lim_i \lambda y_i = \lim_i \lambda y_i - My_i + My_i = \lim_i y_i - My_i + \lim_i My_i = 0 + z$$

Since $\lambda \neq 0$, $y_i \rightarrow \lambda^{-1}z$. But then $z \in Y$, and $z = Mz$. □

Corollary 3.15. *Let M be a compact operator. If $\lambda \in \sigma(M)$ is non-zero, then $\dim(\ker(\lambda - M)) < \infty$, $(\lambda - M)X$ is closed and $\text{codim}(\lambda - M) < \infty$, and there is n such that*

$$\begin{aligned} \ker[(\lambda - M)^n] &= \ker[(\lambda - M)^{n+1}] \\ (\lambda - M)^n X &= (\lambda - M)^{n+1} X \end{aligned}$$

Proof. First note that $\lambda = M$ on $\ker(\lambda - M)$, implying the identity is compact on $\ker(\lambda - M)$, and therefore that $\dim(\ker(\lambda - M)) < \infty$. It follows that $\ker(\lambda - M)$ is closed, and thus there is a subspace Y of X such that $X = \ker(\lambda - M) \oplus Y$. M restricted to Y is injective, and the last theorem implies, since $Y \cap \ker(\lambda - M) = (0)$, that

$$C = \inf\{\|(\lambda - M)y\| : y \in B_Y\} > 0$$

Hence $\|(\lambda - M)y\| \geq C\|y\|$ on Y , and so $\lambda - M$ is an isomorphism from Y to a subspace of X , which is therefore complete and thus closed. Note that

$$[(\lambda - M)X]^\perp = \{\phi \in X^* : \phi \circ (\lambda - M) = 0\} = \ker(\lambda - M^*)$$

Since T is compact, so is T^* , and because λ must also be an eigenvalue of T^* , $\dim[(\lambda - M)X]^\perp < \infty$, so that $(\lambda - M)X$ has finite codimension. □

Chapter 4

Vector Valued Integration

Ubiquitous to analysis is the integral, defined to be $\int f(x) dx$, defined to be an average of the infinitesimal values of the function f across a certain domain. In this context, f is almost always real or complex valued. But there are many spaces in which an average can be taken. Indeed, in order to define the average of a set of values a_1, \dots, a_N , given by

$$\frac{a_1 + \dots + a_N}{N}$$

we need only that we can add the a_n , and multiply them by a scalar. Thus the natural place to generalize the integration of functions is to those which take their values in a topological vector space. In this chapter we analyze such a theory.

The weakest definition of the integral exploits the fact that the integral is linear. Because integration is the limit of finite sums, and linear functions are interchangeable with summation, for any continuous linear operator $\Lambda : X \rightarrow Y$, and function $f : \Omega \rightarrow X$, we ought to have

$$\Lambda \left(\int_{\Omega} f \right) = \int_{\Omega} (\Lambda \circ f)$$

In particular, if Λ is a linear functional, the right hand side is just a scalar-valued function, and we know how to integrate those. In fact, we can *define* the **weak integral** $\int_{\Omega} f$ such that this equation holds for any continuous linear functional Λ . In particular, if X^* separates X , then such an integral is unique, and the only question of applicability is when such an

integral exists. The conditions of the next theorem, which show that such an integral exists, are automatically satisfied if X is a Fréchet space.

Theorem 4.1. *If X is a topological vector space such that X^* separates points, and μ is a Borel probability measure on a compact Hausdorff space Ω , and $f : \Omega \rightarrow X$ is a continuous function such that the closed convex hull of $f(\Omega)$ is compact, then the weak integral exists.*

Proof. Let H be the closed convex hull of $f(\Omega)$. Given a finite family of linear functionals $L = \{\Lambda_1, \dots, \Lambda_N\}$, we consider the family E_L of all points $x \in H$ such that

$$\Lambda_n(x) = \int_{\Omega} (\Lambda_n \circ f)(x) d\mu$$

If we can prove that each E_L is non-empty, then by the compactness of H , the intersection of all such E_L will be nonempty, and this will be the required weak integral. If we define a vector y by

$$y_n = \int_{\Omega} (\Lambda_n \circ f) d\mu$$

then it suffices to show that the image of H under the map $L(x) = (\Lambda_1 x, \dots, \Lambda_N x)$ contains y . Because L is a linear map into \mathbf{C}^n , $L(H)$ is compact and convex. Thus if y was not an element of this set, we can separate y from H by a linear functional. In other words, there are constants c_1, \dots, c_N such that for any $x \in H$,

$$\sum c_n y_n < \sum c_n \Lambda_n x$$

In particular,

$$\sum c_n y_n < \sum c_n (\Lambda_n \circ f)$$

Now integrating the right hand side gives

$$\sum c_n y_n < \sum c_n y_n$$

which contradicts the fact that y was not in the convex hull. \square

Thus the integral of a continuous function is defined in a great many circumstances. TODO: ESTABLISH PROPERTIES ALA RUDIN.

4.1 Convolution

Convolution of functions can be realized as a vector valued integral. The normal definition

$$(f * g)(x) = \int f(x - y)g(y) dy$$

By removing the value x , can be realized as the function-valued integral

$$f * g = \int f_y g(y) dy$$

If g is integrable, then we can consider this as integration of the function f_y against the finite measure $g dy$, so provided that the map $y \mapsto f_y$ is continuous, then the weak integral exists. The weak convolution is the normal convolution, because if h is a bounded function, then since $|f| * |g|$ is integrable, $(|f| * |g|)h$ is also integrable, so

$$\int |(f * g)(x)| |h(x)| dx \leq \int (|f| * |g|)(x) h(x) dx < \infty$$

and so we may apply Fubini's theorem to conclude that

$$\int \left(\int f(x - y)g(y) dy \right) h(x) dx = \int \left(\int f(x - y)h(x) dx \right) g(y) dy$$

Part II

Linear Analysis

Chapter 5

Topological Vector Spaces

Originally, functional analysis dealt with norm spaces, providing vectors with an explicit way of measuring their size, and the distance between vectors. But soon, vector spaces of functions appeared in problems equipped with strange, non-normable topologies. For instance, the space of holomorphic functions under the topology of locally uniform convergence, or the class of Schwartz functions on Euclidean space under the topology of rapid uniform convergence of the function and all of its derivatives. Even if we only care about normable spaces, the study of the weak and weak * topologies on a Banach space and its dual forces us to care about non normable spaces, so studying topological vector spaces in their complete generality is still useful.

In order to be completely general, we begin by making a framework to study function spaces with a topology under the weakest assumptions. A **topological vector space** is a vector space, together with a topology making addition and scalar multiplication continuous. In terms of nets, this continuity means exactly that if $x_\alpha \rightarrow x$, $y_\beta \rightarrow y$, and $\lambda_\gamma \rightarrow \lambda$, then $\lambda_\gamma(x_\alpha + y_\beta) \rightarrow \lambda(x + y)$. In terms of neighbourhoods, the continuity says that if V is an open subset containing $\lambda(x + y)$, then there exists $\varepsilon > 0$, and neighbourhoods W of x and U of y such that if $|\lambda - \gamma| < \varepsilon$, then $\gamma(W + U) \subset V$. The maps $T_y(x) = x + y$ and $M_\lambda(x) = \lambda x$ are homeomorphisms of X , as a consequence of the continuity of addition and multiplication. As a consequence, every neighbourhood of a point x can be written as $x + W$, where W is a neighbourhood of the origin. Because of this, most of the topological structure of a space can be understood by looking at the structure around the origin. Another special case of the continuity of ad-

dition and multiplication is that, since $1 \cdot (0 + 0) = 0$, we can find a neighbourhood W of the origin with $W + W \subset V$, and, by induction, we can repeat this process to find a neighbourhood W with $W + W + \cdots + W \subset V$ for any fixed, finite number of additions.

Remark. A topological vector space is, of course, an abelian topological group. However, unlike in the case of topological groups, all but the finite dimensional topological vector spaces fail to be locally compact. This is why these two theories proceed down very different paths.

Example. If X is a set, consider the space \mathbf{F}^X of functions from X to \mathbf{F} , under the product topology. Then \mathbf{F}^X is a vector space, which we can make topological if we equip the space with the topology of pointwise convergence. The continuity of the operations follows because if $f_\alpha \rightarrow f$, $g_\beta \rightarrow g$, and $\lambda_\gamma \rightarrow \lambda$, then for each $x \in X$,

$$\lambda_\alpha(f_\alpha(x) + g_\beta(x)) \rightarrow \lambda(f(x) + g(x))$$

because addition and multiplication is continuous on \mathbf{F} , and this shows that the functions $\lambda_\gamma(f_\alpha + g_\beta)$ converges pointwise to $\lambda(f + g)$. The natural neighbourhood base around the origin are the sets

$$\varepsilon B_{x_1, \dots, x_n} = \{f \in \mathbf{F}^X : |f(x_1)|, \dots, |f(x_n)| < \varepsilon\}$$

for $x_1, \dots, x_n \in X$, and $\varepsilon > 0$. This space is only first countable if X is countable.

Example. Consider the space $L_0(X)$ of all measurable functions on a measure space X with the topology induced by convergence of measure. That is, a net f_α converges to f if, for any $\varepsilon > 0$, $\mu(|f_\alpha - f| \geq \varepsilon) \rightarrow 0$. Under a probabilistic interpretation, the probability that the two functions differ by more than ε converges to zero for every ε . The neighbourhoods of the origin with respect to this topology are of the form

$$\varepsilon U_\delta = \{f \in L_0(X) : \mu(|f| < \varepsilon) \leq \delta\}$$

The space is therefore first countable.

Example. We would expect the space $L_0(X)$ to have an additional topological structure given by pointwise convergence almost everywhere. That is, we would like to have a topology such that $f_\alpha \rightarrow f$ if the set of points x where $f_\alpha(x)$ does not converge to $f(x)$ has zero measure. Tempting a notion as it is, there is no topological structure giving this notion of convergence on $L_0(X)$. To see

why, consider the space $L_0[0,1]$, induced by the Lebesgue measure on $[0,1]$. Consider the typewrite sequence of characteristic functions χ_n , defined in terms of the sets $[H_n, H_{n+1}]$ modulo 1, where H_n is the n th Harmonic number. Since $H_n \rightarrow \infty$, every point intersects infinitely many of the intervals, so that χ_n does not converge anywhere to zero. But χ_n does converge in measure to 0, so that for every subsequence of χ_n , there is a further subsequence consisting of functions which converge pointwise almost everywhere to 0. But this contradicts the topological fact that a net x_α converges to a point x if and only if every subnet of x_α has a further subnet which converges to x . Thus there can be no topology whose convergence agrees with almost everywhere convergence.

Example. Consider the space $\mathbf{F}[[X]]$ of formal power series in a single variable. Define a topology on $\mathbf{F}[[X]]$ by letting $f_\alpha \rightarrow f$ if, eventually, the lower order coefficients in the power series expansion of $f_\alpha - f$ vanish. Thus a neighbourhood basis about the origin consists of the decreasing family $\mathfrak{m}, \mathfrak{m}^2$, and so on, where \mathfrak{m} is the ideal consisting of power series with vanishing constant coefficient. It is easy to see that addition is continuous. Multiplication of two power series is also continuous, for if $f_\alpha \rightarrow f$ and $g_\beta \rightarrow g$, then

$$f_\alpha g_\beta - fg = f_\alpha(g_\beta - g) + (f_\alpha - f)g$$

and because each \mathfrak{m}^n is an ideal, if $g_\beta - g$ and $f_\alpha - f$ are in \mathfrak{m}^n , then $f_\alpha g_\beta$ is also in \mathfrak{m}^n . However, scalar multiplication is not continuous, for the convergence properties of this space are ‘too discrete’. Indeed, if we consider $\lambda_n \neq 0$ with $\lambda_n \neq 0$, then $\lambda_n X$ does not converge to 0, because $\lambda_n X \notin \mathfrak{m}^2$. Thus $\mathbf{F}[[X]]$ is a topological ring (over any field), but is not a topological vector space. It is a useful space in algebra, which favours the discrete properties of the topological structure, but it is not so useful in analysis. We can weaken the topology on $\mathbf{C}[[X]]$ by taking the topology of pointwise convergence of the coefficients, and this gives $\mathbf{C}[[X]]$ a vector space structure, but it no longer reflects the algebraic structure of the space as strongly.

5.1 Basic Theorems

We now introduce some terminology from analytical linear algebra, which will become even more prevalent in the theory of topological vector spaces.

- A **convex** subset C of a vector space X is a set such that if $x, y \in C$ imply $tx + (1 - t)y \in C$ for $0 \leq t \leq 1$, so C contains line segments

between each of its points, or in set notation, $tC + (1-t)C \subset C$ for all $0 \leq t \leq 1$. For any $t, u \in \mathbf{R}$, we have $tC + uC \subset (t+u)C$ which reduces to the ordinary convex equation by dividing both sides by $t+u$.

- A subset B is **balanced** if $\lambda B \subset B$ for $|\lambda| \leq 1$. For any scalar λ , λB is balanced, and if $|\lambda| = 1$, then $\lambda B = B$, since $B = \lambda^{-1}(\lambda B) \subset \lambda B \subset B$ which implies $B = \lambda B$.
- A subset A is **absorbing** if

$$X = \bigcup_{t>0} tA$$

so every $x \in X$ is equal to ty for some scalar $t > 0$ and $y \in A$.

- A set K is **(Von Neumann) bounded** if, for every neighbourhood U of the origin, there is $u > 0$ such that if $t > u$, $tU \supset K$.

One of the cornerstones of the basis theory of topological vector spaces is that analytic properties of a space can be established from the existence of open sets with the geometric properties above.

Theorem 5.1. *Every neighbourhood of the origin is absorbing.*

Proof. Given a topological vector space X , and a neighbourhood U of the origin. For any point x , by continuity,

$$\lim_{\lambda \downarrow 0} \lambda x = \left(\lim_{\lambda \downarrow 0} \lambda \right) x = 0 \cdot x = 0$$

Thus, eventually, $\lambda x \in U$, which implies $x \in \lambda^{-1}U$. □

Theorem 5.2. *Every topological vector space has a basis of balanced sets.*

Proof. Given a neighbourhood U of the origin in a topological vector space X , note that by continuity of multiplication there is $\varepsilon > 0$ and $V \subset U$ such that if $|\lambda| < \varepsilon$, $\lambda V \subset U$. Then $\bigcup_{|\lambda| < \varepsilon} \lambda V \subset U$ is a balanced open neighbourhood of the origin contained in U . But this implies that the balanced open neighbourhoods form a basis of the entire topology at the origin. □

Note that now if $0 < t_1 < t_2 < \dots \rightarrow \infty$, and V is a balanced neighbourhood, then $t_n V$ is balanced, hence contains uV for all $|\lambda| < t_n$ and therefore, since V is absorbing,

$$\bigcup t_n V = \bigcup_{u>0} uV = X$$

The previous theorem says that all balanced neighbourhoods contain absorbing neighbourhoods, so this statements remains true for any neighbourhood V , not just the balanced ones.

Proposition 5.3. *If a space has a bounded open set, it is first countable.*

Proof. Assume the bounded open set V lies at the origin. Then if t_n are a sequence of scalars tending to zero, then $t_n V$ is a neighbourhood basis of the origin, because if W is an arbitrary neighbourhood, then $V \subset t_n^{-1} W$ for some t_n , so $t_n V \subset W$. Thus the $t_n V$ provide a countable neighbourhood base around the origin. \square

Lemma 5.4. *If U is a neighbourhood of the origin such that $x + U$ does not contain y , then there is a neighbourhood V of the origin such that $x + V$ does not intersect $y + V$.*

Proof. Find a balanced subset U_0 of U such that $U_0 + U_0 \subset U$. If there is $u, w \in U_0$ such that $x + u = y + w$, then

$$y = x + u - w \in x + U_0 - U_0 = x + U_0 + U_0 \subset x + U$$

hence $x + U_0$ is disjoint from $y + U_0$. \square

Corollary 5.5. *T_1 topological spaces are Hausdorff.*

From now on, we will assume all the vector spaces we consider are Hausdorff. To see that a particular vector space is Hausdorff, it suffices to show that the intersection of a neighbourhood basis of the origin consists of only the origin, so that $\bar{0} = 0$. In most cases, any topological vector space can be made Hausdorff by considering an appropriate quotient space, ala identifying functions almost everywhere in the theory of L^p spaces.

Corollary 5.6. *If K is a compact set disjoint from C , then there is a neighbourhood W such that $K + W$ is disjoint from $C + W$.*

Proof. For each $x \in K$, we may find a balanced set U_x such that $x + U_x + U_x$ is disjoint from $C + U_x$. By compactness, finitely many U_{x_1}, \dots, U_{x_n} cover K , and then if $W = U_{x_1} \cap \dots \cap U_{x_n}$, we find $K + W$ is disjoint from $C + W$. \square

A corollary of this is that Hasudorff topological spaces are automatically regular. Since $C + W$ is open, we find that, in fact, $\overline{K + W}$ does not intersect $C + W$. If we choose $K = \{0\}$, and $C = V^c$ for V a neighbourhood of the origin, we find a neighbourhood W with $\overline{W} \subset V$. Thus for any neighborhood V we can find a subneighbourhood whose closure is contained in V .

Theorem 5.7. *Let X be a topological vector space.*

- (i) *If V is a neighbourhood of the origin, and A is any set, then $\overline{A} \subset A + V$, and in fact $\overline{A} = \bigcap_V A + V$.*
- (ii) *If $A, B \subset X$, then $\overline{A} + \overline{B} \subset \overline{A + B}$.*
- (iii) *The closure of a subspace is a subspace.*
- (iv) *If C is a convex, then C° and \overline{C} are convex subsets.*
- (v) *If B is balanced, then so is \overline{B} , and if $0 \in B^\circ$, then B° is balanced.*
- (vi) *If E is a bounded subset of X , then so is \overline{E} .*
- (vii) *If A and B are bounded, then so is $A + B$.*
- (viii) *If A and B are compact, so is $A + B$.*
- (ix) *If A is compact, and B is closed, $A + B$ is closed.*

Proof. $x \in \overline{A}$ if and only if $x + V$ intersects A for every neighbourhood V of the origin, hence $x \in A - V$, and $-V$ is a neighbourhood of the origin if and only if V is. This proves (i). The continuity of addition essentially proves (ii), for if $x \in \overline{A}$, and $y \in \overline{B}$, then we find $x_\alpha \rightarrow x$, $y_\alpha \rightarrow y$ with $x_\alpha, y_\alpha \in X$, then $x_\alpha + y_\alpha \rightarrow x + y$, and $x_\alpha + y_\alpha \in A + B$. If Y is a subspace, then (ii) implies $\overline{Y} + \overline{Y} \subset \overline{Y + Y} = \overline{Y}$, and $\lambda \overline{Y} = \overline{\lambda Y} = \overline{Y}$, so \overline{Y} is closed under addition and multiplication, and is therefore a subspace. Similarly, if C is convex, then

$$\lambda \overline{C} + (1 - \lambda) \overline{C} = \overline{\lambda C + (1 - \lambda) C} \subset \overline{C}$$

and if $x, y \in C^\circ$, we can find a neighbourhood U such that $x + U, y + U \subset C$, and then for any λ , by convexity

$$(\lambda x + (1 - \lambda)y) + (\lambda U + (1 - \lambda)U) = \lambda(x + U) + (1 - \lambda)(y + U) \subset C$$

which implies $\lambda x + (1 - \lambda)y \in C^\circ$, hence (iv) is shown. If $\lambda B \subset B$ for $|\lambda| < 1$, then $\lambda \bar{B} = \overline{\lambda B} \subset \bar{B}$, so \bar{B} is balanced. Conversely, λB° is open for each $\lambda \neq 0$, so for nonzero $|\lambda| < 1$, $\lambda B^\circ \subset B^\circ$, and provided $0 \in B^\circ$, we find that $\lambda B^\circ \subset B^\circ$ for all $|\lambda| < 1$. If E is a bounded subset, and V is some neighbourhood, then there is some W such that $\bar{W} \subset V$, and if we find t such that $E \subset uW$ for all $u > t$, then $\bar{E} \subset u\bar{W} \subset uV$, hence \bar{E} is bounded. If A and B are bounded, then $A + B$ is bounded, because if W is fixed, we can find U with $U + U \subset W$, and if $A, B \subset \lambda U$, then

$$A + B \subset \lambda(U + U) \subset \lambda W$$

If A and B are compact, then $A + B$ is compact, as it is the image of $A \times B$ under the continuous addition map. If A is compact, and B is closed, then $A + B$ is closed, because if $x_\alpha + y_\alpha \rightarrow z$ in $A + B$, then x_α has a subnet converging to some $x \in A$, and then $y_\alpha \rightarrow z - x$, hence $z - x \in B$, and therefore $z = x + (z - x) \in A + B$. \square

5.2 Convexity and Metrizability

We say a space is **locally convex** if it has a locally convex base.

Theorem 5.8. *Every convex neighbourhood contains a subneighbourhood which is convex and balanced.*

Proof. If U is a convex neighbourhood, consider the set $A = \bigcap_{|\lambda|=1} \lambda U$, which is, convex, and balanced because if $|\gamma| < 1$,

$$\gamma A = \gamma \bigcap_{|\lambda|=1} \lambda U = \bigcap_{|\lambda|=1} \gamma \lambda U = \bigcap_{|\lambda| < |\gamma|} \lambda U \subset A$$

If $V \subset U$ is balanced, then $V \subset A$, so $0 \in A^\circ$, and in particular, A° is a balanced, convex neighbourhood of the origin. \square

Given a set A , we let $\text{Conv}(A)$ denote the **convex hull** of A , the smallest convex set containing A . It can be written as

$$\text{Conv}(A) = \bigcup_{n=1}^{\infty} \bigcup_{t_1+\dots+t_n=1} (t_1A + \dots + t_nA)$$

It follows that the convex hull of an open set is open.

Theorem 5.9. *In a locally convex space, the convex hull of a bounded set is bounded.*

Proof. If W is a convex neighbourhood, and A is bounded, then $A \subset tW$ for some t , and so $\text{Conv}(A) \subset tW$. \square

We now move on to a very important point of the basis theory of topological vector spaces: When can we metrize the topology of a vector space. The most useful metrics on a vector space are the **translation invariant** ones, which satisfy $d(a+x, a+y) = d(x, y)$ for all points a, x, y in the vector space. It turns out that every first countable vector space has such a metrization.

Theorem 5.10. *If X is a first countable vector space with a countable base, then X is metrizable by an invariant metric d whose open balls are balanced. If X is locally convex, the metric can be chosen so the open balls are convex.*

Proof. We consider a balanced neighbourhood base of the origin V_1, V_2, \dots with $V_{n+1} + V_{n+1} + V_{n+1} \subset V_n$. If X is locally convex, choose the V_n to be convex as well. With each Dyadic rational number $r = \sum c_n 2^{-n}$ in $[0, 1]$, we can associate the interval $V_r = \sum c_n V_n$, with $V_1 = X$. These neighbourhoods are monotonic because of the inclusion properties of these neighbourhoods, and satisfy $V_r + V_s \subset V_{r+s}$ (this is most easily proved when $r = s$, in which case the general case is proved). Thus we can define

$$f(x) = \inf\{r \in [0, 1] : x \in V_r\}$$

and $d(x, y) = f(x - y)$. It is clear d is invariant, and satisfies the triangle inequality, because if $x - y \in V_r$ and $y - z \in V_s$, then $x - z \in V_r + V_s \subset V_{r+s}$. Since the V_r is balanced if the V_n are, the open balls of d are balanced, and if the V_r are convex, the open balls of d are convex. \square

Remark. The metric d obtained in the proof above is always bounded, i.e. $0 \leq d \leq 1$, and so we cannot use the same techniques to show that certain topological vector spaces are normable. We will describe the technique to obtain a norm very shortly.

The notion of Cauchy sequences can be used to determine if the metric d is complete. But even if we don't have a metric, we can still define the notion of a Cauchy sequence. We say x_α is a Cauchy net in a topological vector space if, for every neighbourhood U of the origin, eventually we have $x_\alpha - x_\beta \in U$. If x_α converges, then it is obviously Cauchy. Of course, if d is an invariant metric for the topological vector space, these notions agree exactly, which gives the surprising fact that if a single metric for a metrizable topological vector space is complete, then *any* metric for the vector space is complete.

Theorem 5.11. *Cauchy sequences are bounded.*

Proof. Let x_1, x_2, \dots be Cauchy. If U is a balanced neighbourhood of the origin, then eventually for $n \geq N$, $x_n - x_N \in U$. We know that for some λ , $x_1, \dots, x_N \in \lambda U$, hence all of the x_n are in $(\lambda + 1)U$. \square

Theorem 5.12. *A complete metrizable subspace of a topological vector space is closed.*

Proof. If Y is complete and metrizable, and $y_\alpha \rightarrow x$, then y_α is Cauchy, hence converges in Y , and since X is Hausdorff, Y is closed. \square

A **seminorm** on a space X is a positive function $\|\cdot\|$ on X such that $\|x + y\| \leq \|x\| + \|y\|$, and $\|\alpha x\| \leq |\alpha| \|x\|$. These seminorms would give X a metric by setting $d(x, y) = \|x - y\|$, except that we might have $d(x, y) = 0$ with $x \neq y$. One way to fix this is to consider the set Y of all x with $\|x\| = 0$. Then Y forms a subspace of X , and if $y \in Y$, then $\|x + y\| \leq \|x\|$, and also $\|x\| = \|x + y - y\| \leq \|x + y\|$, hence $\|x + Y\| = \|x\|$, and so the seminorm descends to X/Y , and is now a norm. This is the perspective taken in the L^p spaces, where X is the space of all p integrable functions, and Y is the space of p integrable functions equal to zero almost everywhere.

We shall now describe another way to give X a Hausdorff topological structure if, instead of a single seminorm, we have a *family* of seminorms. Given such a family, we can give X a topological structure by letting $x_\alpha \rightarrow 0$ if $\|x_\alpha\| \rightarrow 0$ for every seminorm $\|\cdot\|$ in the family. A family of seminorms

is called **separating** if for every $x \neq 0$, there is a seminorm $\|\cdot\|$ with $\|x\| \neq 0$. This implies that the topological structure is Hausdorff, for if x_α converges to two points x and x' , then for any seminorm $\|\cdot\|$

$$\|x - x'\| \leq \|x - x_\alpha\| + \|x_\alpha - x'\| \rightarrow 0$$

hence $\|x - x'\| = 0$, and since this is true for every seminorm, $x = x'$. Every space of this form is locally convex, because a neighbourhood basis is given by the family of convex sets

$$\varepsilon B_{\rho_1 \dots \rho_n} = \{x : \rho_1(x), \dots, \rho_n(x) < \varepsilon\} = \varepsilon(B_{\rho_1} \cap \dots \cap B_{\rho_n})$$

and each of these neighbourhoods is convex, since they are the intersections of convex balls with respect to an individual seminorm. In other words, the family of balls $\varepsilon B_\rho = \{x : \rho(x) < \varepsilon\}$ form a subbase for a topology. If the family of seminorms defining the topology is closed under taking the maximum operation, then these balls actually form a base for the topology.

If $\|\cdot\|$ is a seminorm, then the ‘unit ball’ with respect to this seminorm is convex, balanced, and absorbing. Conversely, suppose A is a convex subset of a topological vector space containing the origin. Then we can define the **Minkowski functional**

$$\mu_A(x) = \inf\{t \geq 0 : x \in tA\}$$

If $x \in \alpha A$ and $y \in \beta A$, then by convexity,

$$\lambda x + \gamma y \in \lambda \alpha A + \gamma \beta A \subset (\lambda \alpha + \gamma \beta) A$$

Taking infima, we find $\mu_A(\lambda x + \gamma y) \leq \lambda \mu_A(x) + \gamma \mu_A(y)$, and so, as a special case, we find $\mu_A(x+y) \leq \mu_A(x) + \mu_A(y)$. If A is an absorbing set, μ_A is always finite, and if, in addition, A is balanced, then $\mu_A(\lambda x) = |\lambda| \mu_A(x)$, so μ_A is a seminorm. If $A_0 = \mu_A^{-1}([0, 1))$ and $A_1 = \mu_A^{-1}([0, 1])$, then $A_0 \subset A \subset A_1$, and $\mu_{A_0} = \mu_A = \mu_{A_1}$. Thus every *continuous* Minkowski functional can be given in terms of an open or a closed set.

The connection between convex sets and Minkowski functionals gives a close correspondence between locally convex spaces and spaces with metrics given by seminorms. If $\|\cdot\|$ is any continuous seminorm on X , we let A consist of the points x with $\|x\| \leq 1$. Then A is convex, because if $\|x\|, \|y\| \leq 1$, then

$$\|\lambda x + (1 - \lambda)y\| \leq \lambda \|x\| + (1 - \lambda)\|y\| \leq \lambda + (1 - \lambda) = 1$$

It is absorbing, since $x \in \|x\|A$, and balanced, since $\|\lambda x\| = |\lambda|\|x\|$. We find

$$\mu_A(x) = \inf\{\lambda \geq 0 : x \in \lambda A\} = \inf\{\lambda \geq 0 : \|x\| \leq \lambda\} = \|x\|$$

so every seminorm is given by a Minkowski functional.

The previous paragraphs hint that the topology of *every* locally convex space is given by a family of seminorms. Indeed, suppose that a space has a convex balanced neighbourhood basis U_α . If $\rho_\alpha = \mu_{U_\alpha}$, then each ρ_α is continuous, because if $x_\beta \rightarrow x$, then eventually $x_\beta - x \in \varepsilon U_\alpha$ for every $\varepsilon > 0$, hence eventually $\rho_\alpha(x_\beta - x) < \varepsilon$. This implies

$$\rho_\alpha(x_\beta) \leq \rho_\alpha(x_\beta - x) + \rho_\alpha(x) \leq \rho_\alpha(x) + \varepsilon$$

$$\rho_\alpha(x_\beta) \geq \rho_\alpha(x) - \rho_\alpha(x - x_\beta) \geq \rho_\alpha(x) - \varepsilon$$

So $\rho_\alpha(x_\beta) \rightarrow \rho_\alpha(x)$. Conversely, if $\rho_\alpha(x_\beta) \rightarrow 0$ for every fixed α , then eventually $\rho_\alpha(x_\beta) \leq 1$, so $x_\beta \in U_\alpha$, so x_β is eventually in U_α for every α , so $x_\beta \rightarrow 0$. This shows that the topology of the vector space is exactly the same as the one formed from the family of seminorms ρ_α .

If the topology of a space X is specified by a countable family of seminorms, then the space is first countable, and therefore metrizable. However, in this situation we can come up with a natural metric to use, rather than appealing to the theorem we just came up with. If ρ_1, ρ_2, \dots is an ordering of the seminorms, then the functions

$$d_n(x, y) = \frac{\rho_n(x - y)}{1 + \rho_n(x - y)}$$

are bounded and satisfy the triangle inequality, and the metric for the space can be given by

$$d(x, y) = \sum_{n=1}^{\infty} \frac{d_n(x, y)}{2^n} = \sum_{n=1}^{\infty} \frac{\rho_n(x - y)}{2^n(1 + \rho_n(x - y))}$$

Since *any* invariant metric on a complete, metrizable space is complete, there is normally not a need to switch to a different metric when doing general calculations on metric spaces, unless it is possible to switch to a single norm for the entire space.

Theorem 5.13. *Let X be a locally convex, specified by a family of seminorms ρ_α . Then a set E is bounded if and only if ρ_α is bounded on E for each α .*

Proof. Suppose E is bounded. If U_α is the unit ball with respect to ρ_α , then for some $t > 0$, $E \in tU_\alpha$, so $\rho_\alpha(E) \leq t$. Conversely, suppose that $\rho_\alpha(E) \leq t_\alpha$ for each α . Then we consider a neighbourhood basis of the elements $\varepsilon U_{\rho_1 \dots \rho_N}$, as introduced above, and for each such element we find

$$E \in \max(t_1, \dots, t_N) \varepsilon U_{\rho_1 \dots \rho_N}$$

so E is bounded. □

To finish our understanding of locally convex spaces, we ask when the topology of a locally convex space can be given by a single norm. If $\|\cdot\|$ is a norm for the space, then the unit ball with respect to this norm is bounded by the theorem above, and convex. It turns out this is sufficient for a locally convex space to be normable.

Theorem 5.14. *A space is normable if and only if it has a bounded, convex neighbourhood.*

Proof. If U is a bounded, convex neighbourhood about the origin, then the εU form a neighbourhood basis around the origin. It follows that the topology of the space is specified by the single Minkowski functional μ_U . The fact that μ_U is actually a norm, rather than a seminorm, follows from the fact that the space is Hausdorff. □

Example. Consider the class $C(X)$ of continuous functions on a locally compact space X , under the topology of locally uniform convergence (uniform convergence on compact sets). Then $C(X)$ is a topological vector space. For each compact set K , we have the seminorms $\|\cdot\|_{L^\infty(K)}$, which define a seminorms which show $C(X)$ and $H(X)$ have the topology of a locally convex space. If X is σ compact, then the topology can be given by a countable family of seminorms, and so $C(X)$ is metrizable. If we write $X = \lim_{n \rightarrow \infty} K_n$, and f_n is Cauchy, then the f_n converge uniformly to some continuous function g_n on K_n , and g_{n+1} agrees with g_n on K_n , hence we can consider the function g defined everywhere by letting $g(x) = g_n(x)$ for all $x \in K_n$, and then $f_n \rightarrow g$. Thus $C(X)$ is Fréchet. However, $C(X)$ is not normable if X is not compact. We know a set $E \subset C(X)$ is bounded if and only if there is M_K such that $\|f\|_{L^\infty(K)} \leq M_K$ for each $f \in E$. But we have a neighbourhood consisting of $U_{K_1 \dots K_N \varepsilon_1 \dots \varepsilon_N}$. And since X is not compact, there is some $x \notin K_1 \cup \dots \cup K_N$, and since X is completely regular, we can find a continuous function f with $f(x) = 1$, but f vanishing on $K_1 \cup \dots \cup K_N$. It follows that $nf \in U_{K_1 \dots K_N \varepsilon_1 \dots \varepsilon_N}$ for all integers n , and the nf are not bounded. Thus $C(X)$ is not normable unless X is compact.

Example. If X is an open subset of \mathbf{C} , then the space $H(X)$ of holomorphic functions on X is also a topological vector space under locally uniform convergence. Since the local uniform limit of a holomorphic function is a holomorphic function, $H(X)$ is a closed subspace of $C(X)$, and is therefore a Fréchet space. Montel's theorem in the theory of complex variables says that $H(X)$ has the Heine-Borel property: Every closed, bounded subset of $H(X)$ is compact. But since $H(X)$ is never finite dimensional, $H(X)$ cannot be locally compact, hence $H(X)$ cannot be locally bounded. It follows that $H(X)$ is also not normable.

Example. Let X be an open subset of \mathbf{R}^n . Then we can give $C^\infty(X)$ the structure of a topological vector space by letting $f_\alpha \rightarrow f$ if all the partial derivatives of f_α converge to the partial derivatives of f locally uniformly. Arguing essentially the same as with $C(X)$ and $H(X)$, this shows $C^\infty(X)$ is a metrizable locally convex space. If f_α is a Cauchy sequence in this space, then f_α is Cauchy on each compact set K , hence f_α and all its partial derivatives converges locally uniformly on K to some C^∞ function g_K . Putting these things together gives that $C^\infty(X)$ is a Fréchet space. Surprisingly $C^\infty(X)$ also satisfies the Heine-Borel property. This is easily proven using the Arzela-Ascoli theorem. Since $C^\infty(X)$ is never finite dimensional, this space cannot be locally bounded, hence $C^\infty(X)$ isn't normable.

5.3 Linear Maps

As in the functional analysis of Banach spaces, linear maps play an important role in the theory. In particular, we are interested in those linear maps between topological vector spaces which are continuous. Continuous maps are exactly those which are uniformly continuous, because if $T : X \rightarrow Y$ is continuous, then for every neighbourhood V of the origin in Y , there is a neighbourhood U of the origin in X such that if $x - y \in U$, then $Tx - Ty \in V$.

Lemma 5.15. *If T is continuous at 0, then T is continuous everywhere.*

Proof. Suppose T is continuous at the origin. If $x_\alpha \rightarrow x$, then $x_\alpha - x \rightarrow 0$, hence $T(x_\alpha - x) = T(x_\alpha) - T(x) \rightarrow 0$, so $T(x_\alpha) \rightarrow T(x)$. Thus T is continuous everywhere. \square

The theory of linear functions is difficult to classify in the general case, but linear *functionals* are easier to understand. Given X , we let $X^\#$ denote

the class of linear functionals, and X^* the class of *continuous* linear functionals.

Theorem 5.16. *If $f \in X^\#$ is a non-zero linear functional, then the following are equivalent:*

1. *f is continuous.*
2. *The nullspace of f is closed.*
3. *The nullspace of f is not dense in X .*
4. *f is bounded in some neighbourhood of the origin.*

Proof. (1) implies (2) implies (3) is obvious. If (3) holds, so the nullspace of f is not dense in X , we can find $x \in X$ and a balanced set U such that $0 \notin f(x + U) = f(x) + f(U)$, hence $f(u) \neq -f(x)$ for any $u \in U$. Since U is balanced, this must imply that $f(u) \neq \lambda f(x)$ for any $|\lambda| \geq 1$, so $|f(u)| \leq |f(x)|$ for all $u \in U$, so f is bounded on U , implying (4). If f is bounded in some neighbourhood of the origin, then $|f(u)| \leq M$ for all $u \in U$, and if $x_\alpha \rightarrow 0$, then eventually for all $\varepsilon > 0$, $x_\alpha \in \varepsilon U$, hence $|f(x_\alpha)| \leq \varepsilon M$, implying $f(x_\alpha) \rightarrow 0$. This means f is continuous at the origin. The general statement now follows from the fact that a linear function is continuous if and only if it is continuous at the origin. \square

Example. Suppose X is a locally convex space, with the topology specified by a family of seminorms closed under the maximum operation. Then the balls εU_ρ form a basis for the space. It follows that if $f \in X^*$ is a continuous linear functional, then f is bounded on a neighbourhood on the origin, so there exists ρ, ε , and M such that if $\rho(x) \leq \varepsilon$, then $|f(x)| \leq M$. It then follows that for any $x \in X$,

$$|f(x)| = \frac{\rho(x)}{\varepsilon} \left| f\left(\frac{\varepsilon x}{\rho(x)}\right) \right| \leq (M/\varepsilon)\rho(x)$$

so the functional is directly bounded above by a seminorm.

For finite dimensional topological vector spaces, the continuous linear maps are particularly easy to consider.

Lemma 5.17. *Every linear map $T : \mathbb{F}^n \rightarrow X$ is continuous.*

Proof. If we write $x_i = Te_i$, where e_i is the i 'th element of canonical basis of \mathbf{F}^n , then $T(a) = \sum a_i x_i$, and the continuity follows because addition and scalar multiplication is continuous. \square

Corollary 5.18. *Every finite dimensional vector space has a unique topology making it into a topological vector space.*

Proof. if τ_0 and τ_1 are two topologies on \mathbf{F}^n turning it into a topological vector space, then the identity map on \mathbf{F}^n is a continuous map from (\mathbf{F}^n, τ_0) to (\mathbf{F}^n, τ_1) , and from (\mathbf{F}^n, τ_1) to (\mathbf{F}^n, τ_0) , hence we conclude the identity map is a homeomorphism. \square

Corollary 5.19. *Every finite dimensional subspace of a Hausdorff topological vector space is closed.*

Proof. If $Y \subset X$ is a subspace of X , let $T : \mathbf{F}^n \rightarrow Y$ be an isomorphism. Then we know $S = T(S_{\mathbf{F}^n})$ is compact in X , and $B = T(B_{\mathbf{F}^n})$ is open. By compactness, there is an open neighbourhood V of the origin in X which does not intersect S . Then $W = T^{-1}(V \cap Y)$ is an open neighbourhood of the origin in \mathbf{F}^n disjoint from $S_{\mathbf{F}^n}$, so that it only contains elements in B , and so \overline{W} is compact, hence $V \cap Y \subset T(\overline{W}) \subset Y$, and so the closure of each $V \cap Y$ lies in Y . Now if $z \in \overline{Y}$ is arbitrary, then $z = tx$ for $x \in V$, and since \overline{Y} is a subspace we find x lies in \overline{Y} . But if $y_\alpha \rightarrow x$, with $y_\alpha \in Y$, then y_α is eventually in $Y \cap V$, which is a closed subset of Y , hence we find $x \in Y$, hence $z \in Y$. \square

Theorem 5.20. *Every locally compact vector space X has finite dimension.*

Proof. If a vector space has a precompact neighbourhood V , then that neighbourhood is bounded, so $V/2^n$ form a local base for X . Now there are $x_1, \dots, x_n \in V$ such that

$$\overline{V} \subset (x_1 + V/2) \cup \dots \cup (x_n + V/2)$$

Then the span of the x_i is a closed subspace Y of X , and $\overline{V} \subset Y + V/2$. Then, of course, $V \subset Y + V/2$, hence, dividing by two, we find $V/2 \subset Y + V/4$, and so $\overline{V} \subset Y + V/4$. Continuing this process, we find $\overline{V} \subset \bigcap (Y + V/2^n)$, and since $V/2^n$ form a local base, we find $\overline{V} \subset \overline{Y} = Y$. We conclude that Y contains an absorbing set for X , hence $Y = X$. \square

Since we have a definition of boundedness in any topological vector space X , we can ask if a version of the Heine Borel theorem holds in X . That is, are closed, bounded subsets of X compact.

Corollary 5.21. *No infinite dimensional space can be locally bounded and satisfy the Heine-Borel theorem.*

Proof. If V is a bounded neighbourhood, then it is precompact, hence the space is locally compact, so finite dimensional. \square

In the context of norm spaces, we found a linear map was continuous if and only if it was bounded. An analogous definition of a bounded operator $T : X \rightarrow Y$ between topological vector spaces is an operator that maps bounded sets to bounded sets. It is easy to prove that a continuous map is bounded, but the converse is not always true. However, in a metrizable space, it is true.

Lemma 5.22. *if X is a metrizable space, and $x_n \rightarrow 0$, then there are constants $\lambda_n \rightarrow \infty$ with $\lambda_n x_n \rightarrow 0$.*

Proof. If d is translation invariant, it is easy to prove that for any x , $d(nx, 0) \leq nd(x, 0)$. If $d(x_n, 0) \rightarrow 0$, there are an increasing sequence of positive integers N_m with $d(x_n, 0) \leq 1/m^2$ for $n \geq N_m$. Then $d(mx_{n_m}, 0) \leq 1/m$, so if λ_n is the largest m with $n_m \leq n$, then $\lambda_n x_{n_m} \rightarrow 0$. \square

Example. *If X is non-metrizable, this need not be true. Consider the space $\mathbf{C}^{[0,1]}$ of all complex functions on $[0, 1]$, under the topology of pointwise convergence. Since $[0, 1]$ has the same cardinality as the space $\mathbf{C}^{\mathbf{N}}$ of all complex-valued sequences converging to zero pointwise, we can define a map $F : [0, 1] \rightarrow \mathbf{C}^{\mathbf{N}}$, and then define a sequence f_n of functions converging pointwise to zero by letting $f_n(x) = F(x)_n$. For any sequence of constants $\lambda_1, \lambda_2, \dots$ converging to ∞ , the sequence $1/\lambda_1, 1/\lambda_2, \dots$ converges to zero, hence there is some $x \in [0, 1]$ with $F(x)_n = 1/\lambda_n$. Then $\lambda_n f_n(x) = \lambda_n / \lambda_n = 1$, so $\lambda_n f_n$ does not converge pointwise to zero.*

Theorem 5.23. *If X is a metrizable space, every bounded operator with X as a domain is continuous.*

Proof. Since X is first countable, it suffices to prove that if x_n is a sequence converging to zero, and $T : X \rightarrow Y$ is bounded, then T is continuous. Certainly x_n is bounded, so Tx_n is bounded. If we pick λ_n as in the last

lemma, so $\lambda_n x_n \rightarrow 0$, then $\lambda_n x_n$ is bounded, so $\lambda_n T x_n$ is bounded, and since $\lambda_n^{-1} \rightarrow 0$, $T x_n = \lambda_n^{-1}(\lambda_n T x_n) \rightarrow 0$. \square

Example. A quasinorm space is a vector space X equipped with a non-negative real-valued function $\|\cdot\|$ such that $\|x\| = 0$ if and only if $x = 0$, $\|\lambda x\| = |\lambda| \|x\|$, and $\|x + y\| \lesssim \|x\| + \|y\|$, rather than the stronger inequality $\|x + y\| \leq \|x\| + \|y\|$. The most precient examples of quasinorm spaces are the $L^p(X)$ spaces over some measure μ , where $p < 1$. Since the unit ball in these spaces are not convex, we know that $\|\cdot\|_p$ is not a norm. However, the norms corresponding to these spaces are quasinorms. To see this, we use the fairly trivial inequality $(a+b)^p \leq 2^p(a^p + b^p)$, for $a, b, p > 0$, which follows because

$$(a + b)^p \leq (2 \max(a, b))^p = 2^p \max(a, b)^p \leq 2^p(a^p + b^p)$$

so that $\|f + g\|_p^p \leq 2^p(\|f\|_p^p + \|g\|_p^p)$, and therefore

$$\|f + g\|_p \leq 2(\|f\|_p^p + \|g\|_p^p)^{1/p} \leq 2^{1+1/p}(\|f\|_p + \|g\|_p)$$

So the ' L^p norm' is a quasinorm. This quasinorm gives $L^p(X)$ the structure of a complete metric space, the proof of completeness being exactly the same as in the case where $p \geq 1$. The fact that only a single quasinorm is defined in the definition of the topology shows the space is locally bounded. However, the space is not locally convex, hence not a Fréchet space. In fact, as an example, $L^p[0, 1]$ contains no convex open sets other than \emptyset and $L^p[0, 1]$. To see this, let W be a non-empty, open, convex set, and without loss of generality, let $0 \in W$, and suppose W contains all functions f with $\|f\|_p < \varepsilon$. Given any $f \in L^p[0, 1]$, and any integer N , we can divide $[0, 1]$ into disjoint intervals I_1, I_2, \dots, I_N with

$$\|f\|_{L^p(I_n)} = \frac{\|f\|_{L^p[0,1]}}{N^{1/p}}$$

If $f_n = N \chi_{I_n} f$, then $\|f_n\|_{L^p[0,1]} = N^{1-1/p} \|f\|_{L^p[0,1]}$. Since $p < 1$, we can choose N large enough that $\|f_n\|_{L^p[0,1]} < \varepsilon$, hence $f_n \in W$, and then by convexity,

$$f = \frac{f_1 + \dots + f_N}{N} \in W$$

Thus $W = L^p[0, 1]$. Now a strange result of this is that if $T : L^p[0, 1] \rightarrow X$ is a continuous linear transformation into a locally convex space X , then $T = 0$. This follows because if W is an open, convex neighbourhood, then $T^{-1}(W)$

is open and convex, hence $T^{-1}(W) = L^p[0, 1]$. But then taking W to be an arbitrary convex neighbourhood of the origin,

$$T^{-1}(0) = T^{-1}\left(\bigcap W\right) = \bigcap T^{-1}(W) = L^p[0, 1]$$

Thus there are no nonzero continuous linear functions on $L^p[0, 1]$.

For non-metrizable locally convex spaces, we still have a nice analytical criterion for continuity in terms of seminorms.

Theorem 5.24. *Let $T : X \rightarrow Y$ be a map between convex spaces. Then T is continuous if and only if for any continuous seminorm $\|\cdot\|_Y$ on Y , there is a continuous seminorm $\|\cdot\|_X$ on X such that $\|Tx\|_Y \leq \|x\|_X$.*

Proof. Suppose T is continuous. Then the unit ball B in the Y norm is open, so $T^{-1}(B)$ is an open neighbourhood of the origin in X , and there therefore contains a convex, balanced subset about the origin which is the unit ball with respect to some continuous seminorm $\|\cdot\|_X$. Thus $\|Tx\|_Y \leq 1$ if $\|x\|_X \leq 1$, which implies that

$$\|Tx\|_Y = \|x\|_X \|T(x/\|x\|_X)\|_Y \leq \|x\|_X$$

Conversely, if the other condition is satisfied and U is an open convex set about the origin in Y , then it is the unit ball with respect to some norm $\|\cdot\|_Y$, and so there is $\|\cdot\|_X$ such that $\|Tx\|_Y \leq \|x\|_X$, hence the unit ball with respect to the X norm is an open neighbourhood which maps into U , so T is continuous at the origin, hence continuous everywhere. \square

If X and Y are generated by a family of seminorms closed under maxima, we can restrict the conditions of the theorem to only the seminorms contained in this group, if we weaken the condition $\|Tx\|_Y \leq \|x\|_Y$ to $\|Tx\|_Y \lesssim \|x\|_X$.

5.4 Quotient Spaces

If Y is a subspace of a topological vector space X , then X/Y can be given the quotient topology, and one easily verifies that addition and multiplication is continuous, so the quotient of a topological vector space is a topological vector space. Since the neighbourhood of $0 + Y$ in X/Y is an open

set containing Y ,

$$\overline{0 + Y} = \bigcap_{U+Y} (0 + Y) + (U + Y) = \bigcap_U (U + Y) = \overline{Y} + Y$$

So X/Y is Hausdorff if and only if Y is closed.

We can now argue why assuming a space is Hausdorff isn't too strong a statement. Let X be a (not necessarily Hausdorff) topological vector space. Then 0 is a subspace of X , and so the closure of 0 is also a closed subspace. But now if we consider the space $X/\overline{0}$, we obtain a Hausdorff topological space. Thus the techniques of Hausdorff topological spaces can normally be used to obtain results about non Hausdorff spaces by means of quotienting.

Theorem 5.25. *The quotient map $\pi : X \rightarrow X/Y$ is open, and a neighbourhood base for X/Y is given by taking $\pi(U)$, where U ranges over a neighbourhood base.*

Proof. The fact that $\pi(U)$ is open when U is open follows because

$$\pi^{-1}(\pi(U)) = U + Y$$

is obviously a union of open neighbourhoods. If V is an open neighbourhood of the origin in X/Y , then $\pi^{-1}(V)$ is a neighbourhood of the origin, hence contains some U in a neighbourhood base, and then $\pi(U) \subset V$. \square

Theorem 5.26. *The quotient of a locally convex / locally bounded / metrizable / normable / complete metrizable space also has these properties.*

Proof. The image of a convex / balanced / absorbing / bounded set under a quotient map is also convex / balanced / absorbing / bounded. This shows that the quotient of a locally convex / bounded space is locally convex / bounded. Since the quotient of a first countable space is first countable, this shows that the quotient of a metrizable space is metrizable. The fact that the quotient of a normable space is normable follows because a normable is precisely a locally convex and locally bounded space. The only remaining fact is to show that the quotient of a complete metrizable space is complete. If d is a complete invariant metric for the space, with $f(x) = d(0, x)$, then the metric for the quotient is given by

$$d(x_0 + Y, x_1 + Y) = \inf\{f((x_1 - x_0) + y) : y \in Y\}$$

The triangle inequality follows because if $f((x_1 - x_0) + y_0) \leq \alpha$ and $f((x_2 - x_1) + y_1) \leq \beta$, then $d(x_1 + y_0, x_0) \leq \alpha$, and $d(x_2 + y_1 + y_0, x_1 + y_0) \leq \beta$, and so the triangle inequality on X gives

$$d(x_0, x_2 + y_1 + y_0) \leq \alpha + \beta$$

so $d(x_0 + Y, x_2 + Y) \leq \alpha + \beta$, and taking infima gives the triangle inequality. If $x_\alpha + Y \rightarrow 0$ in X/Y , then we can pick y_α with $x_\alpha + y_\alpha \rightarrow 0$ in X , hence $d(0, x_\alpha + y_\alpha) \rightarrow 0$, so $d(0, x_\alpha + Y) \rightarrow 0$. Conversely, if $d(0, x_\alpha + Y) \rightarrow 0$, then there are y_α with $d(0, x_\alpha + y_\alpha) \rightarrow 0$, so $x_\alpha + y_\alpha \rightarrow 0$, hence $x_\alpha + Y \rightarrow 0$. Thus d is a metric for the topological structure of X/Y . Now if d is complete on X , it remains to show d is complete on X/Y . Suppose $x_n + Y$ is Cauchy. Then we can inductively, for each n , pick k_n and y_n such that $d(x_{k_{n+1}} + y_{n+1}, x_{k_n} + y_n) \leq 1/2^n$. This means $x_{k_n} + y_n$ is Cauchy, hence converges, hence $x_{k_n} + Y$ converges, and because of the Cauchy property it follows that $x_n + Y$ converges to the same value. \square

Corollary 5.27. *If Y is a closed subspace of X , and Z is finite dimensional, then $Y + Z$ is closed.*

Proof. Z/Y is finite dimension in X/Y , hence is closed. But $Y + Z$ is the inverse image of Z/Y under the quotient map, which is continuous, so $Y + Z$ is closed. \square

Chapter 6

Weak Topologies

Example. In Banach theory, the most important examples of topological spaces are those obtained by placing other topologies on a norm space than the original norm topology. Given a norm space X , we have a space X^* consisting of continuous linear functionals on X . We say that a net x_α converges weakly to x if $f(x_\alpha) \rightarrow f(x)$ for every $f \in X^*$. This defines the **weak topology** on X . Similarly, on X^* we may define the **weak * topology**, by letting $f_\alpha \rightarrow f$ if $f_\alpha(x) \rightarrow f(x)$ for every $x \in X$. Note that we may also put a weak topology on X^* , since it is also a norm space, but the weak and weak * topologies on X^* rarely correspond (the weak topology has more stringent notions of continuity). These topologies are effectively topologies induced by pointwise convergence, because the weak topology is obtained by embedding X in \mathbb{F}^{X^*} , and the weak * topology is obtained by considering X^* as a subspace of \mathbb{F}^X . The weak topology on X is rarely equal to the original topology on X , except in degenerate circumstances. For instance, if we consider $e_n \in l_p$, for $1 < p < \infty$, then $\|e_n - e_m\| = 1$, so e_n does not converge, yet if we consider any element $f \in l_q$ of the dual space, then $\langle f, e_n \rangle = f(n)$, which converges to zero as $n \rightarrow \infty$. Thus e_n converges to zero in the weak topology. In l_1 , e_n does not converge at all, because the function $f(n) = (-1)^n$ is in l_∞ , and $\langle f, e_n \rangle = (-1)^n$ does not converge. It is important to note that the weak * topology depends on the predual we are using. If we consider $l_1 = c_0^*$, then the e_n converges to zero in the weak * topology, yet if we consider l_1 as c^* , then the e_n do not converge in the induced weak * topology.

The weak topology on a space is rarely first countable. If $A = \{\sqrt{n}e_n : n \in \mathbb{N}\}$ is viewed as a subset of l_2 , then 0 is contained in the weak closure of A , because if $\varepsilon > 0$ and we are given a finite set of sequences $\{a_n^k\}_{n \in \mathbb{N}}$, for $1 \leq k \leq K$

and, for any n , there is k_n such that $|a_n^{k_n}| \sqrt{n} > \varepsilon$, and if we consider the sum $\sum_k |a_n^k| \in l^2$, then

$$\sum_n \left| \sum_k |a_n^k| \right|^2 \geq \sum_n |a_n^{k_n}|^2 \geq \varepsilon \sum_n \frac{1}{n} = \infty$$

so there must be an element of A in each $U_{\varepsilon, \{a_n^1\}, \dots, \{a_n^{k_n}\}}$, and hence 0 is in the weak closure of A . However, 0 is not the weak limit of any sequence of elements in A , for

$$\sum_{n=0}^{\infty} \frac{e_n}{n} \in l_2$$

and if we consider any sequence $\sqrt{n_i} e_{n_i}$ that converges weakly to zero, then we find that $(n_i)^{-1/2}$ converges to zero, hence $n_i \rightarrow \infty$. But then the sequence is not norm bounded, and hence cannot converge!

Chapter 7

The Theory of Distributions

The theory of distributions is a jewel which emerges from the theory of topological vector spaces; it enables us to justify formal manipulations without the numerous technical issues which emerge from the analytical side. For instance, the ordinary integral formulation of the Fourier transform is only defined on L^1 functions, whereas the theory of distributions enables us to define the Fourier transform on essentially every function we would ever want to take the Fourier transform. Similarly, classically we can only differentiate a very particular class of functions, which does not behave so nicely under the rules of functional analysis. The theory of distributions enables us to define an object which behaves as the derivative of any continuous function. For these reasons, and the fact that distributions behave well from the perspective of functional analysis that has made them the cornerstone of modern harmonic analysis and the analysis of partial differential equations.

The power of measure theory is that it enables us to study a very general class of functions, ranging from the classes $L^p(\mathbf{R}^n)$ of integrable functions and the class $C_c^\infty(\mathbf{R}^n)$ of smooth functions with compact support. The problem is that as we allow more general classes of functions, the operations we can perform on these functions becomes more and more restricted. Nonetheless, $C_c^\infty(\mathbf{R}^n)$ is dense in almost every function space we consider. It is here that all the fundamental analytical operations can be performed, but generality is lost because a single discontinuity discounts a function from being contained in this space. Often more general results are able to be obtained from applying a density argument in the larger spaces. The theory of distributions provides an alternate viewpoint.

The general idea emerges from duality. Because the dual space of $L^p(\mathbf{R}^n)$, except in exceptional circumstances, is $L^q(\mathbf{R}^n)$, where q is the conjugate to p , functions $f \in L^p(\mathbf{R}^n)$ can be identified uniquely as ‘integrands’

$$g \mapsto \int f(x)g(x) dx$$

Similarly, the dual space of $C(K)$, where K is compact, is the space $M(K)$ of finite Borel measures on K . Thus we can think of measures as a suitable ‘generalized functions’, operating as integrands against linear functionals. From the perspective of physics this is entirely natural. Points in space are idealizations - one can never measure the exact value of some quantity of a function at a point, but rather only understand the function by looking at its averages over a small region around that point. Thus a ‘function’ can be understood by the averages with respect to a family of integrands, known as test functions, since they test the value of the function over a region. As we make the family of integrands smaller and smaller, a ‘function’ can behave more and more erratically. A distribution is an ‘integrand’ with respect to the space $C_c^\infty(\mathbf{R}^n)$ of infinitely differentiable functions with compact support. Since these functions are incredibly analytically nice, distributions are allowed to behave incredibly erratically, but we can still extend the operations of differentiation and integration to them.

7.1 The Space of Test Functions

We fix an open subset Ω of \mathbf{R}^n , and let $C_c^\infty(\Omega)$ denote the family of all smooth functions on Ω with compact support. We could equip $C_c^\infty(\Omega)$ with a locally convex, metrizable topology with respect to the seminorms

$$\|f\|_{C^n(\Omega)} = \max_{|\alpha| \leq n} \|D^\alpha f\|_{L^\infty(\Omega)}$$

However, this does not give $C_c^\infty(\Omega)$ a complete metric space structure.

Example. Let $\Omega = \mathbf{R}$, pick a bump function $\phi \in C_c^\infty(\mathbf{R})$ supported on $[0, 1]$ with $\phi > 0$ on $(0, 1)$, and define

$$\psi_m(x) = \phi(x-1) + \frac{\phi(x-2)}{2} + \cdots + \frac{\phi(x-m)}{m}$$

Then ψ_m is compactly supported on $[1, m]$, and Cauchy, since for $k \geq l$,

$$\|D^m(\psi_k - \psi_l)\|_{C^n(\mathbf{R})} = \frac{\max_{r \leq n} \|D^r \phi\|_\infty}{l}$$

but with respect to these seminorms, the ψ_m converge to the function

$$\psi(x) = \sum_{n=1}^{\infty} \psi(x - n)$$

an element of $C^\infty(\mathbf{R})$ which is not compactly supported. In other words, we have proven that $C_c^\infty(\mathbf{R})$ is not a closed subspace of $C^\infty(\mathbf{R})$.

To make $C_c^\infty(\Omega)$ complete, we give it a stronger locally convex topology. As a subspace, the topology on the space $C_c^\infty(K)$ of smooth functions compactly supported on K will have the same topology, but in the new topology, a sequence of functions ϕ_1, ϕ_2, \dots converge *only* if the union of their supports is precompact. We define a neighbourhood basis for $C_c^\infty(\Omega)$ to consist of all convex balanced sets W such that $W \cap C_c^\infty(K)$ is open in $C_c^\infty(K)$ for each compact set $K \subset \Omega$.

Theorem 7.1. *This gives a basis of a Hausdorff topology on $C_c^\infty(\Omega)$.*

Proof. If $\phi_1 + V_1$ and $\phi_2 + V_2$ both contain ϕ , then $\phi - \phi_1 \in V_1$ and $\phi - \phi_2 \in V_2$. The function ϕ, ϕ_1 , and ϕ_2 are all supported on some compact set K . By continuity of multiplication on $C_c^\infty(K)$, and the fact that $V_n \cap C_c^\infty(K)$ is open, there is a small constant δ such that $\phi - \phi_n \in (1 - \delta)V_n$. The convexity of the V_n implies that $\phi - \phi_n + \delta V_n \subset V_n$. But then $\phi + \delta V_n \subset \phi_n + V_n$, and so $\phi + \delta(V_1 \cap V_2) \subset V_1 \cap V_2$. The space is Hausdorff, because if ϕ is in every open neighbourhood of the origin, then in particular, for every compact set K , ϕ is in the convex, balanced neighbourhood consisting of functions ψ such that $\|\psi\|_{L^\infty(K)} < \varepsilon$ for each $\varepsilon > 0$. Letting $\varepsilon \rightarrow 0$ gives that ϕ vanishes on K , and since Ω is locally compact, ϕ vanishes everywhere. \square

Remark. This technique can be formulated more abstractly to give a locally convex topological structure to the direct limit of locally convex spaces. From this perspective, we also see why our metrization doesn't work; if $X = \lim X_n$, with each X_n a locally convex metrizable space, then we cannot give X a complete metrizable topology such that each X_n is an embedding and has empty interior in X , because this would contradict the

Baire category theorem. In particular, this means that the topology we have given to $C_c(\Omega)$ cannot be metrizable, and therefore the space cannot be first countable. Later we will see a more explicit proof of this.

Theorem 7.2. $C_c^\infty(\Omega)$ is a locally convex space.

Proof. Fix ϕ and ψ , and consider any neighbourhood W of the origin. By convexity, we have $(\phi + W/2) + (\psi + W/2) \subset (\phi + \psi) + W$. This shows addition is continuous. To show multiplication is continuous, fix λ , ϕ , and a neighbourhood W of the origin. Then ϕ is supported on some compact set K , and $W \cap C_c^\infty(K)$ is open, in particular absorbing, so there is $\varepsilon > 0$ such that if $|\alpha| < \varepsilon$, $\alpha\phi \in W/2$. Then if $|\gamma - \lambda| < \varepsilon$, then because W is balanced and convex,

$$\begin{aligned} \gamma \left(\phi + \frac{W}{2(|\lambda| + \varepsilon)} \right) &= \lambda\phi + (\gamma - \lambda)\phi + \frac{\gamma}{2(|\lambda| + \varepsilon)}W \\ &\subset \lambda\phi + W/2 + W/2 \subset \lambda\phi + W \end{aligned}$$

so multiplication is continuous. \square

Theorem 7.3. The map $C_c^\infty(K) \rightarrow C_c^\infty(\Omega)$ is an embedding.

Proof. We shall prove a convex, balanced neighbourhood V is open in $C_c^\infty(\Omega)$ if and only if $C_c^\infty(K) \cap V$ is open in $C_c^\infty(K)$ for each K . Since V is open, V is the union of convex, balanced sets W_α with $W_\alpha \cap C_c^\infty(K)$ open in $C_c^\infty(K)$ for each K . But then $V \cap C_c^\infty(K) = (\bigcup W_\alpha) \cap C_c^\infty(K)$ is open in $C_c^\infty(K)$. The converse is true by definition of the topology. But this statement means exactly that the map $C_c^\infty(K) \rightarrow C_c^\infty(\Omega)$ is an embedding, because it is certainly continuous, and if W is a convex neighbourhood of the origin equal to the set of ϕ supported on K with $\|\phi\|_{C^n(K)} \leq \varepsilon$ for some n , then the image is the intersection of $C_c^\infty(K)$ with the set of all ϕ supported on Ω satisfying the inequality, which is open. This shows that the map is open onto its image, hence an embedding. \square

Because of the definition in terms of open sets, it is difficult to see why the topology is that much stronger than the topology previously given. We can see this more numerically by introducing the topology in terms of seminorms. The topology we have given $C_c^\infty(\Omega)$ is the same as the locally convex topology introduced by all norms $\|\cdot\|$ on the space which are continuous when restricted to each $C_c^\infty(K)$. As an example, if we choose

an increasing family U_1, U_2, \dots of precompact open sets whose closure is contained in Ω , then any compact set K is contained in some U_N for large enough N , and for any increasing sequence $\alpha_1, \alpha_2, \dots$ of positive constants and increasing sequence k_1, k_2, \dots of positive integers the norm

$$\|f\| = \min_{\text{supp}(f) \subset U_n} \alpha_n \|f\|_{C^{k_n}(U_n)}$$

is well defined on $C_c^\infty(\Omega)$ and continuous. But this means that if $f_1, f_2, \dots \rightarrow 0$, then $\|f\| \rightarrow 0$ for any choice of constants α_n and k_n , so asymptotically as we approach the boundary of Ω (or ∞ , if Ω is unbounded), the L^∞ norms of the f_n and all of their derivatives must converge arbitrarily fast outside certain compact sets. The next theorem shows that this implies that the union of the domains f_n must actually be precompact. It is this ‘uniform compactness’ that gives us completeness.

Theorem 7.4. *E is a bounded subset of $C_c^\infty(\Omega)$ if and only if E is contained in $C_c^\infty(K)$ for some compact set K , and there are constants M_n such that $\|\phi\|_{C^n(\Omega)} \leq M_n$ for all $\phi \in E$.*

Proof. If E is not contained in some $C_c^\infty(K)$ for some K , then we can find functions $\phi_n \in E$ and a set of points $x_n \in X$ with no limit point and $\phi_n(x_n) \neq 0$. The set

$$V = \left\{ \phi : |\phi(x_n)| < \frac{|\phi_n(x_n)|}{n} \right\}$$

is an open neighbourhood of the origin, since it is convex and balanced, and for any compact set K , K can only contain finitely many of the x_n , and so $V \cap C_c^\infty(K)$ is specified by finitely many seminorms. Now E cannot be bounded with respect to V , because $\phi_n \notin nV$ for all n . Thus if E is bounded, then E is contained in some $C_c^\infty(K)$, and the existence of the constants M_n is implied by the fact that the topology induced $C_c^\infty(K)$ is just the usual topology. The converse follows by the embedding property. \square

Corollary 7.5. *$C_c^\infty(\Omega)$ has the Heine Borel property.*

Proof. This follows because if E is bounded and closed, it is a closed and bounded subset of some $C_c^\infty(K)$ for some K , hence E is compact since $C_c^\infty(K)$ satisfies the Heine-Borel property. \square

Corollary 7.6. *$C_c^\infty(\Omega)$ is quasicomplete.*

Proof. If ϕ_1, ϕ_2, \dots is a Cauchy sequence in $C_c^\infty(\Omega)$, then the sequence is bounded, hence contained in some common $C_c^\infty(K)$. Since the sequence is Cauchy, they converge in $C_c^\infty(K)$ to some ϕ , and thus the ϕ_n converge to ϕ in $C_c^\infty(\Omega)$. \square

It is often useful to use the fact that we can perform a ‘separation of variables’ to a smooth function. This is done formally in the following manner. Say $f \in C_c^\infty(\mathbf{R}^n)$ is a *tensor function* if there are $f_1, \dots, f_n \in C_c^\infty(\mathbf{R})$ such that $f(x) = f_1(x_1) \dots f_n(x_n)$. We write $f = f_1 \otimes \dots \otimes f_n$. Since the product of two tensor functions is a tensor function, the family of all finite sums of tensor functions forms an algebra.

Theorem 7.7. *Finite sums of tensor functions are dense in $C_c^\infty(\mathbf{R}^n)$.*

Proof. If $f \in C_c^\infty(\mathbf{R}^n)$ is 2π periodic, then there are coefficients a_m for each $m \in \mathbf{Z}^n$ such that

$$f(x) = \sum_m a_m e^{im \cdot x}$$

where the convergence of the sum, and all of its derivatives, is uniform, and the convergence is faster than any polynomial rate. Note that each term in the sum is a tensor function. Given any ϕ compactly supported on $[-N, N]$, we let f be the 2π periodic function which is equal to ϕ on $[-\pi N, \pi N]^n$. We have a series of trigonometric polynomials f_1, f_2, \dots with the convergence properties above. If $\psi : \mathbf{R} \rightarrow \mathbf{R}$ is a compactly supported bump function equal to one on $[-N, N]$, and vanishing outside of $[-2N, 2N]$, then $\psi^{\otimes n} f_1, \psi^{\otimes n} f_2, \dots$ converges to ϕ in $C_c^\infty(\mathbf{R}^n)$, and each term is a sum of tensor functions. \square

The power of the topology on $C_c^\infty(\Omega)$ forces uniform convergence, but it means we have too many open sets to have metrizability. However, because $C_c^\infty(\Omega)$ is the limit of metrizable spaces, its linear operators still have many of the same properties as metrizable spaces because it is the limit of metrizable spaces.

Theorem 7.8. *If $T : C_c^\infty(X) \rightarrow Y$ is a map from $C_c^\infty(X)$ to some locally convex space Y , then the following are equivalent:*

- (1) *T is continuous.*
- (2) *T is bounded.*

(3) If $\phi_1, \dots \rightarrow 0$, then $T\phi_1, \dots \rightarrow 0$.

(4) T is continuous restricted to each $C_c^\infty(K)$.

Proof. We already know that (1) implies (2). If T is bounded, and we have a sequence ϕ_1, \dots converge to zero, then this sequence is bounded, hence contained in some $C_c^\infty(K)$, and T is bounded as a map from $C_c^\infty(K)$ to Y , hence $T\phi_1, \dots \rightarrow 0$. (3) implies (4) holds because each $C_c^\infty(K)$ is metrizable, and any convergent sequence is contained in some common $C_c^\infty(K)$. To prove that (4) implies (1), we let V be a convex, balanced set. Then $T^{-1}(V) \cap C_c^\infty(K)$ is open for each K , and $T^{-1}(V)$ is convex and balanced, so $T^{-1}(V)$ is balanced. \square

Because convergence is so strict in $C_c^\infty(\Omega)$, almost every operation we want to perform on smooth functions is continuous in this space.

- Since $f \mapsto D^\alpha f$ is a continuous operator from $C_c^\infty(K)$ to itself, it is therefore continuous on the entire space $C_c^\infty(\Omega)$. More generally, any linear differential operator with coefficients in $C_c^\infty(\Omega)$ is a continuous operator.
- The inclusion $C_c^\infty(\Omega) \rightarrow L^p(\Omega)$ is continuous. To prove this, it suffices to prove for each compact K , the inclusion $C_c^\infty(K) \rightarrow L^p(\Omega)$ is continuous, and this follows because $\|f\|_p \leq |K| \|f\|_\infty$, which is easy to verify.
- If $f \in L^1(\mathbf{R}^n)$ and is compactly supported, then for any $g \in C_c^\infty(\mathbf{R}^n)$, $f * g \in C_c^\infty(\mathbf{R}^n)$. This is because $f * g$ is continuous since $g \in L^\infty(\mathbf{R}^n)$, and its support is contained in the sum of the support of f and g , as well as the identity $D^\alpha(f * g) = f * (D^\alpha g)$. In fact, the map $g \mapsto f * g$ is a continuous operator on $C_c^\infty(\mathbf{R}^n)$. This is because if we restrict our attention to $C_c^\infty(K)$, and f has support on K' , then our convolution operator maps into the compact set $K + K'$, and since

$$\|D^\alpha(g * f)\|_\infty = \|D^\alpha g * f\|_\infty \leq \|D^\alpha g\|_\infty \|f\|_1$$

we conclude

$$\|g * f\|_{C^n(K+K')} \leq \|g\|_{C^n} \|f\|_1$$

which gives continuity of the operator as a map from $C_c^\infty(K)$ to $C_c^\infty(K + K')$. Since the latter space embeds in $C_c^\infty(\mathbf{R}^n)$, we obtain continuity of the operator on $C_c^\infty(\mathbf{R}^n)$.

Theorem 7.9. *If a map $T : C_c^\infty(K) \rightarrow C_c^\infty(\mathbf{R}^n)$ is continuous, then the image of $C_c^\infty(K)$ is actually $C_c^\infty(K')$ for some compact set K' .*

Proof. Suppose there is a sequence x_1, x_2, \dots with no limit point and smooth functions f_1, f_2, \dots compactly supported on $C_c^\infty(K)$ such that $(Tf_n)(x_n) \neq 0$. But then for any sequence α_n whatsoever, we cannot have $\alpha_n Tf_n$ converging to zero, hence $\alpha_n f_n$ cannot converge to zero. But this is clearly not true, for if we let

$$\alpha_n = \frac{1}{2^n \|f_n\|_{C^n}}$$

Then for any fixed m , $\|f_n\|_{C^m}$ is eventually bounded above by $1/2^n$ and therefore converges to zero. Thus such a sequence x_n cannot exist, and therefore the image of T is supported on some compact set K' . \square

Thus the topology on the space $C_c^\infty(\mathbf{R}^n)$ is as strict as can be. As a consequence, we shall see that the weak * topology on $C_c^\infty(\mathbf{R}^n)^*$ is essentially the weakest notion of convergence available in analysis, which makes it surprising that we still be able to recover the continuity of many operators on the dual space.

7.2 The Space of Distributions

We now have the tools to explain the idea of a distribution. If f is an integrable function defined on some set Ω , then the map

$$\Lambda_f(\phi) = \int f(x)\phi(x) dx$$

is a *continuous* linear functional defined on $C_c^\infty(\Omega)$. The functional Λ_f determines f up to a set of measure zero, and so we identify Λ_f with f . But even if f satisfies weaker properties like being locally integrable, the definition above defines a continuous linear functional. The idea of the theory of distributions is to treat any continuous linear functional Λ on $C_c^\infty(\Omega)$ as if it were given by integration against a function as nice as possible. Using the properties of integration for these integration, we can usually cheat out a definition of operations usually only applicable to functions that works for all distributions. Thus the operations of analysis generalize

to an incredibly large family of objects. As an example, if f was continuously differentiable, then we would find

$$\int f'(x)\phi(x) dx = - \int f(x)\phi'(x) dx$$

Since the right hand side is defined independantly of how nice the function $f(x)$ is, we could define the *derivative* of a continuous linear functional Λ as

$$\Lambda'(\phi) = -\Lambda(\phi')$$

and more generally, for a linear functional on n dimensional space, we could define $(D^\alpha \Lambda)(\phi) = (-1)^{|\alpha|} \Lambda(D^\alpha \phi)$.

Example. Let $H(x) = \mathbf{I}(x > 0)$ denote the Heaviside step function. Then H is locally integrable, and so for any test function ϕ , we calculate

$$\int H'(x)\phi(x) dx = - \int H(x)\phi'(x) = - \int_0^\infty \phi'(x) = \phi(0)$$

Thus the ‘derivative’ of the Heaviside step function is the Dirac delta. It is not a function, but if we were to think of it as a function, it would be zero everywhere except at the origin, where it is infinitely peaked.

Example. Consider the Dirac delta function at the origin, which is the distribution $\delta(f) = f(0)$. Then

$$\delta'(f) = -\delta(f') = -f'(0)$$

which is a distribution which doesn’t arise from integration with respect to a locally integrable function nor a Radon measure. Thus the ‘derivative’ of an infinitely peaked function at the origin is the negation of a derivative.

In general, we define a **distribution** to be a continuous linear functional on the space of test functions $C_c^\infty(\Omega)$. In the last section, our exploration of continuous linear transformations on $C_c^\infty(\Omega)$ guarantees that a linear functional Λ on $C_c^\infty(\Omega)$ is continuous if and only if for every compact $K \subset X$ there is an integer N such that $|\Lambda f| \lesssim \|f\|_{C^N}$ for f supported on K . If one integer N works for all K , we define this to be the **order** of the distribution. If such an N doesn’t exist, we say the distribution has infinite order.

Example. If μ is a locally finite Borel measure, or a finite complex valued measure, then we can define

$$\Lambda_\mu(\phi) = \int \phi(x) d\mu(x)$$

Thus Λ_μ is a distribution, since if ϕ is supported on K , then

$$|\Lambda_\mu(\phi)| \leq \mu(K) \|\phi\|_{L^\infty(K)}$$

Thus Λ_μ is a distribution of order zero. When we say μ is a distribution, we are really thinking of the distribution Λ_μ .

Example. Not all distributions arise from functions or measures. For instance, if f is smooth and has compact support on the real line, then

$$p.v. \int \frac{f(x)}{x} = \lim_{\varepsilon \rightarrow 0} \int_{|x| > \varepsilon} \frac{f(x)}{x} dx$$

exists, independantly of the value of f at the origin. This is because the values of $1/x$ on either side of the real axis cancel each other out in the integral. More rigorously, if the support of f is contained in $[-N, N]$, then

$$\left| \int_{|x| > \varepsilon} \frac{f(x)}{x} \right| = \left| \int_{\varepsilon < |x| < N} \frac{f(x) - f(0)}{x} \right|$$

By the mean value theorem, $f(x) - f(0) = xf'(y)$ for some y between zero and x , so

$$\left| \int_{|x| \leq \varepsilon} \frac{f(x) - f(0)}{x} \right| \leq 2\varepsilon \|f'\|_{C^1}$$

From this, we can conclude that the values converge, and that

$$\left| p.v. \int \frac{f(x)}{x} \right| \lesssim \|f'\|_{C^1}$$

Thus this linear functional is actually continuous, hence a distribution, and it is not induced by any function or measure. This distribution is the derivative of the distribution induced by the locally integrable function $\log|x|$, since an integration by parts gives

$$\begin{aligned} - \int \log|x| f'(x) &= \lim_{\varepsilon \rightarrow 0} \int_{|x| \geq \varepsilon} \log|x| f'(x) \\ &= \log \varepsilon (f(x) - f(-x)) + \int_{|x| \geq \varepsilon} \frac{f(x)}{x} = p.v. \int \frac{f(x)}{x} dx \end{aligned}$$

As we stated before, given any distribution Λ , we can define its *derivative* $D^\alpha \Lambda$ to be the distribution

$$D^\alpha \Lambda(\phi) = (-1)^{|\alpha|} \Lambda(D^\alpha \phi)$$

which is continuous since the derivative operation is continuous on $C_c^\infty(\Omega)$. Just as the partial derivatives commutes on $C_c^\infty(\Omega)$, the partial differentiation operation commutes on the space of distributions, i.e. $D^\alpha D^\beta \Lambda = D^\beta D^\alpha \Lambda$, and we take the common value to be $D^{\alpha+\beta} \Lambda$. If $D^\alpha f$ is continuous, then we already know an integration by parts gives $D^\alpha \Lambda_f = \Lambda_{D^\alpha f}$, so we can think of the distribution derivative as a true generalization of the usual derivative. On the other hand, in general the distribution derivative may disagree with the usual derivative if the function is less well behaved. If P is a polynomial, we have

$$P(D)(\Lambda)(\phi) = \Lambda(P(-D)(\phi))$$

if we understand the polynomial applications of derivatives linearly.

Example. Let f be a left continuous function on the real line with bounded variation. Then f' exists almost everywhere, and $f' \in L^1(\mathbf{R})$. Now by Fubini's theorem, if we let μ be the measure defined by $\mu([a, b]) = f(b) - f(a)$, then for any $\phi \in C_c^\infty(\mathbf{R})$,

$$\begin{aligned} \int_{-\infty}^{\infty} \phi(x) d\mu(x) &= - \int_{-\infty}^{\infty} \int_{(x, \infty)} \phi'(y) dy d\mu(x) = - \int_{-\infty}^{\infty} \phi'(y) \int_{(-\infty, y)} d\mu(x) dy \\ &= - \int_{-\infty}^{\infty} \phi'(y) [f(y) - f(-\infty)] dy \end{aligned}$$

and we know $f(-\infty) = 0$, hence we find $\Lambda'_f = \Lambda_\mu$. In particular, we only have $\Lambda'_f = \Lambda_\mu$ if $f dx = \mu$, which only holds if f is absolutely continuous.

If $f \in L^1_{\text{loc}}(\mathbf{R}^n)$, and g is C^∞ , then fg is locally integrable. The identity

$$\int (f(x)g(x))\phi(x) dx = \int f(x)(g(x)\phi(x)) dx$$

enables us to define the product of a $C^\infty(\Omega)$ function with a distribution. Given any distribution Λ , we define $(f\Lambda)(\phi) = \Lambda(f\phi)$. To see why $f\Lambda$ is

a distribution, fix a compact set K , and pick A and N such that for any $\phi \in C_c^\infty(K)$, $|\Lambda(f)| \leq A\|f\|_{N,K}$. The Leibnitz rule tells us that

$$D^\alpha(f\phi) = \sum_{\lambda+\gamma=\alpha} C_{\lambda\gamma} D^\lambda f D^\gamma \phi$$

and so

$$|\Lambda(f\phi)| \leq A\|f\phi\|_{N,K} \leq NA \left(\max |C_{\lambda\gamma}| \|D^\lambda f\|_{L^\infty(K)} \right) \|\phi\|_{C^N(K)}$$

so $f\Lambda$ is a distribution with the same order as Λ . It is important that f is a smooth function, so that $f\phi$ is smooth for each function ϕ .

Since $C_c^\infty(X)^*$ is the dual space of a topological vector space, we can give it a natural topology, the weak * topology. Thus a net of distributions Λ_α converges to Λ if and only if $\Lambda_\alpha(\phi) \rightarrow \Lambda(\phi)$ for all test functions ϕ . This gives a further topology on the space of measures and functions, and we often write $f_\alpha \rightarrow f$ ‘in the distribution sense’ if we have a convergence $\Lambda_{f_\alpha} \rightarrow \Lambda_f$ for the corresponding distributions. Since the convergence in $C_c^\infty(\Omega)$ is incredibly strict, convergence of distributions is incredibly weak. The following is quite a surprising result which hints at the generality of the theory of distributions.

Theorem 7.10. *Suppose that $\Lambda_1, \Lambda_2, \dots$ are a sequence of distributions such that for a fixed test function ϕ ,*

$$\Lambda\phi = \lim \Lambda_n\phi$$

exists. Then Λ is a distribution, and $D^\alpha \Lambda_n \rightarrow D^\alpha \Lambda$ for each α .

Proof. Fix a compact set K . Then the Banach Steinhaus theorem guarantees that Λ restricted to $C_c^\infty(K)$ is a continuous functional, and we know this implies Λ is continuous in general. The fact that $D^\alpha \Lambda_n \rightarrow D^\alpha \Lambda$ is trivial, because for a fixed ϕ , $D^\alpha \phi \in C_c^\infty(X)$, so

$$D^\alpha \Lambda(\phi) = (-1)^{|\alpha|} \Lambda(D^\alpha \phi) = (-1)^{|\alpha|} \lim \Lambda_n(D^\alpha \phi) = \lim D^\alpha \Lambda_n(\phi)$$

Thus the sequence weakly converges to $D^\alpha \Lambda$. □

A similar application of the Banach Steinhaus theorem guarantees that if $g_n \rightarrow g$ in $C^\infty(\mathbf{R}^n)$, and $\Lambda_n \rightarrow \Lambda$ in the distributional sense, then $g_n \Lambda_n$ converges to $g\Lambda$ in the distributional sense. Thus the space of distributions is a topological $C^\infty(\mathbf{R}^n)$ module.

7.3 Localization of Distributions

Just as we can consider the local behaviour of functions around a point, we can consider the local behaviour of a distribution around points, and this local behaviour contains most of the information of the distribution. For instance, given an open subset U of X , we say two distributions Λ and Ψ are equal on U if $\Lambda\phi = \Psi\phi$ for every test function ϕ compactly supported in U . We recall the notion of a partition of unity, which, for each open cover U_α of Euclidean space, gives a family of C^∞ functions ψ_α which are positive, *locally finite*, in the sense that only finitely many functions are positive on each compact set, and satisfy $\sum \psi_\alpha = 1$ on the union of the U_α .

Theorem 7.11. *If X is covered by a family of open sets U_α , and Λ and Ψ are locally equal on each U_α , then $\Lambda = \Psi$. If we have a family of distributions Λ_α which agree with one another on $U_\alpha \cap U_\beta$, then there is a unique distribution Λ locally equal to each Λ_α .*

Proof. Since we can find a C^∞ partition of unity ψ_α compactly supported on the U_α , upon which we find if ϕ is supported on K , then finitely many of the ψ_α are non-zero on K , and so

$$\Lambda(\phi) = \sum \Lambda(\psi_\alpha \phi) = \sum \Psi(\psi_\alpha \phi) = \Psi(\phi)$$

Thus $\Lambda = \Psi$. Conversely, if we have a family of distributions Λ_α like in the hypothesis, then we can find a partition of unity $\psi_{\alpha\beta}$ subordinate to $U_\alpha \cap U_\beta$, and we can define

$$\Lambda(\phi) = \sum \Lambda_\alpha(\psi_{\alpha\beta} \phi) = \sum \Lambda_\beta(\psi_{\alpha\beta} \phi)$$

The continuity is verified by fixing a compact K , from which there are only finitely many nonzero $\psi_{\alpha\beta}$ on K , and the fact that this definition is independant of the partition of unity follows from the first part of the theorem. \square

In the language of modern commutative algebra, the association of $C_c^\infty(U)^*$ to each open subset U of Ω gives a sheaf structure to Ω . Given a distribution Λ , we might have $\Lambda(\phi) = 0$ for every ϕ supported on some open set U . The complement of the largest open set U for which this is true is called the **support** of Λ . If f vanishes on a neighbourhood of the

support of Λ , then by definition of the support, $\Lambda f = 0$. The neighbourhood condition is important – δ' is supported on $\{0\}$, since δ is, but it certainly doesn't vanish on f if $f(0) = 0$.

Theorem 7.12. *If a distribution has precompact support, the distribution has finite order, and extends uniquely to a continuous linear functional on $C^\infty(X)$.*

Proof. Let Λ be a distribution supported on a compact set. If ψ is a function with compact support with $\psi(x) = 1$ on the support of Λ , then $\psi\Lambda = \Lambda$, because for any ϕ , $\phi - \phi\psi$ is supported on a set disjoint from the support of Λ . But if ψ is supported on K , then there is N such that for any $\phi \in C_c^\infty(K)$,

$$|\Lambda(\phi)| \lesssim \|\phi\|_{N,K}$$

and so for any other compact set K ,

$$|\Lambda(\phi)| = |\Lambda(\phi\psi)| \lesssim \|\phi\psi\|_{N,K} \lesssim \|\psi\|_{C^N(K)} \|\phi\|_{C^N(K)}$$

which shows Λ has order N . We have shown that Λ is continuous with respect to the seminorm $\|\cdot\|_{C^N(K)}$ on $C^\infty(X)$, and so by the Hahn Banach theorem, Λ extends uniquely to a continuous functional on $C^\infty(X)$. \square

Example. If $\Lambda(\phi) = \sum_{|\alpha| \leq N} \lambda_\alpha D^\alpha \phi(x)$, then Λ is supported on x . Conversely, every distribution Λ supported on x is of this form. We know Λ must have finite order N , and consider ϕ with $D^\alpha \phi(x) = 0$ for all $|\alpha| \leq N$. We claim $\Lambda(\phi) = 0$. Fix $\varepsilon > 0$, and choose a compact neighbourhood K of the origin with $|D^\alpha \phi(x)| < \varepsilon$ on K for all $|\alpha| = N$. Then for $|\alpha| < N$, the mean value theorem implies that, by induction,

$$|D^\alpha \phi(x)| \leq \varepsilon n^{N-|\alpha|} |x|^{N-|\alpha|}$$

Find A such that for functions ϕ supported on K ,

$$|\Lambda(\phi)| \leq A \|\phi\|_{C^N(K)}$$

Fix a bump function ψ with support on the ball of radius one and $\psi(x) = 1$ in a neighbourhood of the origin, and define $\psi_\delta(x) = \psi(x/\delta)$. If δ is small enough,

then ψ is supported on K , and because Λ is supported on x ,

$$\begin{aligned}
|\Lambda(\phi)| &= |\Lambda(\phi\psi_\delta)| \leq A \|\phi\psi_\delta\|_{C^N(K)} \\
&\leq A \sum_{|\alpha+\beta|=N} |c_{\alpha\beta}| \|D^\alpha \phi\|_\infty \|D^\beta \psi_\delta\| \\
&\leq A \|\psi\|_{C^N} \sum_{|\alpha+\beta|=N} |c_{\alpha\beta}| \delta^{|\beta|-|\alpha|} \|D^\beta \phi\|_{L^\infty(K)} \\
&\leq \varepsilon A \left(\sum_{|\alpha+\beta|=N} |c_{\alpha\beta}| n^{N-|\beta|} \right)
\end{aligned}$$

We can then let $\varepsilon \rightarrow 0$ to conclude $\Lambda(\phi) = 0$. But this means that $\Lambda(\phi)$ is a linear function of the partial derivatives of ϕ with order $\leq N$, completing the proof.

Example. If δ is the Dirac delta distribution, then $f\delta = f(0)\delta$ for any smooth function f . Thus, in particular, $x\delta = 0$. Conversely, if Λ is any distribution with $x\Lambda = 0$, then Λ is a multiple of the Dirac delta distribution. To see this, we note that this would imply $\Lambda(f) = 0$ for all functions f such that f/x is also smooth and compactly supported. In particular, this is true if the support of f does not contain the origin. Thus Λ is supported on the origin, hence there are constants a_n such that

$$\Lambda f = \sum_{n=0}^N a_n f^{(n)}(0)$$

But $(xf)^{(n)}(0) = nf^{(n-1)}(0)$ only vanishes for all f when $n = 0$, so Λ is a multiple of the Dirac delta distribution. A more simple way to see this is that if f is compactly supported on $[-N, N]$, the function

$$g(x) = \frac{f(x) - f(0)}{x} = \int_0^1 f'(tx) dt$$

is smooth, and $f = f(0) + xg$. Since Λ and $x\Lambda$ have bounded support, they extend uniquely to $C^\infty(\Omega)$, and so $\Lambda f = f(0)\Lambda 1 + \Lambda(xg) = f(0)\Lambda 1$.

In many other ways, distributions act like functions. For instance, any distribution Λ can be uniquely written as $\Lambda_1 + i\Lambda_2$ for two distributions

Λ_1, Λ_2 that are real valued for any real-valued smooth continuous function. However, we cannot write a real-valued distribution as the difference of two positive distributions, i.e. those which are non-negative when evaluated at any non-negative functional. Given a non-negative functional Λ , we define Λf for a compactly supported continuous function $f \geq 0$ as

$$\Lambda f = \sup\{\Lambda g : g \in C_c^\infty(\mathbf{R}^n), g \leq f\}$$

and then in general define $\Lambda(f^+ - f^-) = \Lambda f^+ - \Lambda f^-$. Then Λ is obviously a positive extension of Λ to all continuous functions, and is linear. But then the Riesz representation theorem implies that there is a positive Radon measure such that $\Lambda = \Lambda_\mu$, completing the proof.

7.4 Derivatives of Continuous Functions

One of the main reasons to consider the theory of distributions is so that we can take the derivative of any function we want. We now show that, at least locally, every distribution is the derivative of some continuous function, which means the theory of distributions is essentially the minimal such class of objects which enable us to take derivatives of continuous functions.

Theorem 7.13. *If Λ is a distribution on Ω , and K is a compact set, then there is a continuous function f and α such that for every ϕ ,*

$$\Lambda \phi = (-1)^{|\alpha|} \int_{\Omega} f(x) (D^\alpha \phi)(x) dx$$

Proof. TODO □

Theorem 7.14. *If K is compact, contained in some open subset V , which in turn is a subset of Ω , and Λ has order N , then there exists finitely many continuous functions $f_\beta \in C(\Omega)$ supported on V , for each $|\beta| \leq N + 2$, with supports on V , and with $\Lambda = \sum D^\beta f_\beta$.*

Theorem 7.15. *If Λ is a distribution on Ω , then there exists continuous functions g_α on Ω such that each compact set K intersects the supports of finitely many of the g_α , and $\Lambda = \sum D^\alpha g_\alpha$. If Λ has finite order, then only finitely many of the g_α are nonzero.*

7.5 Convolutions of Distributions

Using the convolution of two functions as inspiration, we will not define the convolution of a distribution Λ with a test function ϕ , and under certain conditions, the convolution of two distributions. Recall that if $f, g \in L^1(\mathbf{R}^n)$, then their convolution is the function in $L^1(\mathbf{R}^n)$ defined by

$$(f * g)(x) = \int f(y)g(x-y) dy$$

If we define the translation operators $(T_y g)(x) = g(x-y)$, then $(f * g)(x) = \int f(y)(T_x g^*)(y) dy$, where g^* is the function defined by $g^*(x) = g(-x)$. Thus, if Λ is any distribution on \mathbf{R}^n , and ϕ is a test function on \mathbf{R}^n , we can define a function $\Lambda * \phi$ by setting $(\Lambda * \phi)(x) = \Lambda(T_x \phi^*)$. Notice that since

$$\begin{aligned} \int (T_x f)(y)g(y) dy &= \int f(y-x)g(y) dy = \int f(y)g(x+y) dy \\ &= \int f(y)(T_{-x}g)(y) dy \end{aligned}$$

Notice that we can also define the translation operators on distributions by setting $(T_x \Lambda)(\phi) = \Lambda(T_{-x} \phi)$. One mechanically verifies that convolution commutes with translations, i.e. $T_x(\Lambda * \phi) = (T_x \Lambda) * \phi = \Lambda * (T_x \phi)$.

Theorem 7.16. $\Lambda * \phi$ is C^∞ , and $D^\alpha(\Lambda * \phi) = (D^\alpha \Lambda) * \phi = \Lambda * (D^\alpha \phi)$.

Proof. It is easy to calculate that

$$\begin{aligned} (D^\alpha \Lambda * \phi)(x) &= (D^\alpha \Lambda)(\phi_x^*) = (-1)^{|\alpha|} \Lambda(D^\alpha(T_x \phi^*)) \\ &= \Lambda(T_x(D^\alpha \phi)^*) = (\Lambda * D^\alpha \phi)(x) \end{aligned}$$

If e is a unit vector, and we set $\Delta_h = h^{-1}(1 - T_{he})$, then $\Delta_h \phi$ converges to $D_e \phi$ in $C_c^\infty(\mathbf{R}^n)$, and as such,

$$\begin{aligned} \Delta_h(\Lambda * \phi)(x) &= \frac{(\Lambda * \phi)(x) - (\Lambda * \phi)(x - he)}{h} \\ &= \frac{\Lambda(T_x \phi^* - T_{-he}(T_x \phi^*))}{h} = \Lambda(T_x(\Delta_h \phi^*)) \end{aligned}$$

and this converges to $\Lambda(D_e \phi^*) = (\Lambda * D_e \phi)(x)$ as $h \rightarrow 0$. Iteration of this fact gives the general result. \square

Theorem 7.17. *If $\phi, \psi \in C_c^\infty(\mathbf{R}^n)$, then $\Lambda * (\phi * \psi) = (\Lambda * \phi) * \psi$.*

Proof. Let ϕ and ψ be supported on K . We calculate

$$(\phi * \psi)^*(x) = \int \phi^*(x + y)\psi(y) dy = \int (T_y \phi)^*(x)\psi(y) dy$$

since the map $y \mapsto (T_y \phi)^*\psi(y)$ is continuous, and vanishes out of the compact set K , so that we have a $C_c^\infty(K)$ valued integral

$$(\phi * \psi)^* = \int_K \psi^*(y) T_y \phi^* ds$$

This means precisely that

$$\begin{aligned} (\Lambda * (\phi * \psi))(0) &= \Lambda((\phi * \psi)^*) = \int_K \psi^*(y) \Lambda(T_y \phi^*) dy \\ &= \int_K \psi^*(y) (\Lambda * \phi)(y) dy = ((\Lambda * \phi) * \psi)(0) \end{aligned}$$

The commutativity in general results from applying the commutativity of the translation operators. \square

A net ϕ_α is known as an *approximate identity* in the space of distributions if $\Lambda * \phi_\alpha \rightarrow \Lambda$ as distributions for every distribution ϕ , and an approximate identity in the space of test functions if $\psi * \phi_\alpha \rightarrow \psi$ in $C_c^\infty(\mathbf{R}^n)$.

Theorem 7.18. *If ϕ_α is a family of non-negative functions in $C_c^\infty(\mathbf{R}^n)$ which are eventually supported on every neighbourhood of the origin, and integrate to one, then ϕ_α is an approximation to the identity in the space of test functions and in the space of distributions.*

Proof. It is easy to verify that if f is a continuous function, then $f * \phi_\delta$ converges locally uniformly to f as $\delta \rightarrow 0$. But now we calculate that if $f \in C_c^\infty(\mathbf{R}^n)$, then $D^\alpha(f * \phi_\delta) = (D^\alpha f) * \phi_\delta$ converges locally uniformly to $D^\alpha f$, which gives that $f * \phi$ converges to f in $C_c^\infty(\mathbf{R}^n)$. Now if Λ is a distribution, and ψ is a test function, then continuity gives

$$\begin{aligned} \Lambda(\psi^*) &= \lim_{\delta \rightarrow 0} \Lambda(\phi_\delta * \psi) = \lim_{\delta \rightarrow 0} (\Lambda * (\phi_\delta * \psi))(0) \\ &= \lim_{\delta \rightarrow 0} ((\Lambda * \phi_\delta) * \psi)(0) = \lim_{\delta \rightarrow 0} (\Lambda * \phi_\delta)(\psi^*) \end{aligned}$$

and ψ was arbitrary. \square

If Λ is a distribution on \mathbf{R}^n , then the map $\phi \mapsto \Lambda * \phi$ is a linear transformation from $C_c^\infty(\mathbf{R}^n)$ into $C^\infty(\mathbf{R}^n)$, which commutes with translations. It is also continuous. To see this, we consider a fixed compact K , and consider the map from $C_c^\infty(K)$ to $C^\infty(\mathbf{R}^n)$. We can apply the closed graph theorem to prove continuity, so we assume the existence of ϕ_1, ϕ_2, \dots converging to ϕ in $C_c^\infty(K)$ and $\Lambda * \phi_1, \Lambda * \phi_2, \dots$ converges to f . It suffices to show $f = \Lambda * \phi$. But we calculate

$$f(x) = \lim(\Lambda * \phi_n)(x) = \lim \Lambda(T_x \phi_n^*) = \Lambda(\lim T_x \phi_n^*) = \Lambda(T_x \phi^*) = (\Lambda * \phi)(x)$$

where we have used the fact that $T_x \phi_n^*$ converges to $T_x \phi^*$ in $C_c^\infty(\mathbf{R}^n)$. Surprisingly, the converse is true.

Theorem 7.19. *If $L : C_c^\infty(\mathbf{R}^n) \rightarrow C^\infty(\mathbf{R}^n)$ and commutes with translations, then there is a distribution Λ such that $L(\phi) = \Lambda * \phi$.*

Proof. If $L(\phi) = \Lambda * \phi$, then we would have

$$\Lambda(\phi) = (\Lambda * \phi^*)(0) = L(\phi^*)(0)$$

and we take this as the definition of Λ . Λ is continuous because all the operations here are continuous, and because L commutes with translations, we conclude

$$(\Lambda * \phi)(x) = \Lambda(T_x \phi^*) = L(T_{-x} \phi)(0) = L(\phi)(x)$$

which gives the theorem. □

We now move onto the case where a distribution Λ has compact support. Then Λ extends to a continuous functional on $C^\infty(\mathbf{R}^n)$, and we can define the convolution $\Lambda * \phi$ if $\phi \in C^\infty(\mathbf{R}^n)$. The same techniques as before verify that translations and derivatives are carried into the convolution.

Theorem 7.20. *If ϕ and Λ have compact support, then $\Lambda * \phi$ has compact support.*

Proof. Let ϕ and Λ be supported on K . Then $(\Lambda * \phi)(x) = \Lambda(T_x \phi^*)$. Since $T_x \phi^*$ is supported on $x - K$, for x large enough $x - K$ is disjoint from K , and so $\Lambda * \phi$ vanishes outside of $K + K$. □

Theorem 7.21. *If Λ and ψ have compact support, and $\phi \in C^\infty(\mathbf{R}^n)$, then*

$$\Lambda * (\phi * \psi) = (\Lambda * \phi) * \psi = (\Lambda * \psi) * \phi$$

Proof. Let Λ and ψ be supported on some balanced compact set K . Let V be a bounded, balanced open set containing K . If ϕ_0 is a function with compact support equal to ϕ on $V + K$, then for $x \in V$,

$$(\phi * \psi)(x) = \int \phi(x - y)\psi(y) dy = \int \phi_0(x - y)\psi(y) dy = (\phi_0 * \psi)(x)$$

Thus

$$(\Lambda * (\phi * \psi))(0) = (\Lambda * (\phi_0 * \psi))(0) = ((\Lambda * \psi) * \phi_0)(0)$$

But $\Lambda * \psi$ is supported on $K + K$, so $((\Lambda * \psi) * \phi_0)(0) = ((\Lambda * \psi) * \phi)(0)$. Now we also calculate

$$(\Lambda * (\phi * \psi))(0) = ((\Lambda * \phi_0) * \psi)(0) = ((\Lambda * \phi) * \psi)(0) \int (\Lambda * \phi_0)(-y)\psi(y)$$

where the last fact follows because $\Lambda * \phi_0$ agrees with $\Lambda * \phi$ on K . The general fact follows by applying the translation operators. \square

Now we come to the grand finale, defining the convolution of two distributions. Given two distributions Λ and Ψ , one of which has compact support, we define the linear operator

$$L(\phi) = \Lambda * (\Psi * \phi)$$

Then L commutes with translations, and is continuous, because if we have ϕ_1, ϕ_2, \dots converging to ϕ in $C_c^\infty(K)$, then $\Psi * \phi_n$ converges to $\Psi * \phi$ in $C^\infty(\mathbf{R}^n)$. If Ψ is supported on a compact support C , then the $\Psi * \phi_n$ have common compact support $C + K$, and actually converge in $C_c^\infty(C + K)$, hence $\Lambda * (\Psi * \phi_n)$ converges to $\Lambda * (\Psi * \phi)$. Conversely, if Λ has compact support, then $\Psi * \phi_n$ converges in $C^\infty(\mathbf{R}^n)$, which implies $\Lambda * (\Psi * \phi_n)$ converges to $\Lambda * (\Psi * \phi)$ in $C^\infty(\mathbf{R}^n)$. Thus L corresponds to a distribution, and we define this distribution to be $\Lambda * \Psi$.

Theorem 7.22. *If Λ and Ψ are distributions, one of which has compact support, then $\Lambda * \Psi = \Psi * \Lambda$. Let S_Λ and S_Ψ , and $S_{\Lambda * \Psi}$ denote the supports of Λ , Ψ , and $\Lambda * \Psi$. Then $\Lambda * \Psi = \Psi * \Lambda$, and $S_{\Lambda * \Psi} \subset S_\Lambda + S_\Psi$.*

Proof. We calculate that for any two test functions ϕ and ψ ,

$$(\Lambda * \Psi) * (\phi * \psi) = \Lambda * (\Psi * (\phi * \psi)) = \Lambda * ((\Psi * \phi) * \psi)$$

If Λ has compact support, then

$$\Lambda * ((\Psi * \phi) * \psi) = (\Lambda * \psi) * (\Psi * \phi)$$

Conversely, if Ψ has compact support, then

$$\Lambda * ((\Psi * \phi) * \psi) = \Lambda * (\psi * (\Psi * \phi)) = (\Lambda * \psi) * (\Psi * \phi)$$

We also calculate

$$\begin{aligned} \Psi * ((\Lambda * \phi) * \psi) &= \Psi * (\Lambda * (\phi * \psi)) = \Psi * (\Lambda * (\psi * \phi)) \\ &= \Psi * ((\Lambda * \psi) * \phi) = (\Psi * \phi) * (\Lambda * \psi) \end{aligned}$$

But since convolution is commutative, we have

$$((\Lambda * (\Psi * \phi)) * \psi) = \Lambda * ((\Psi * \phi) * \psi) = \Psi * ((\Lambda * \phi) * \psi) = (\Psi * (\Lambda * \phi)) * \psi$$

Since ψ was arbitrary, we conclude

$$(\Lambda * \Psi) * \phi = \Lambda * (\Psi * \phi) = \Psi * (\Lambda * \phi) = (\Psi * \Lambda) * \phi$$

and now since ϕ was arbitrary, we conclude $\Lambda * \Psi = \Psi * \Lambda$. Now we know convolution is commutative, we may assume S_Ψ is compact. The support of $\Psi * \phi^*$ lies in $S_\Psi - S_\phi$. But this means that if $S_\phi - S_\Psi$ is disjoint from S_Λ , which means exactly that S_ϕ is disjoint from $S_\Lambda + S_\Psi$, then

$$(\Lambda * \Psi)(\phi) = (\Lambda * (\Psi * \phi))(0) = 0$$

and this gives the support of $\Lambda * \Psi$. □

This means that the convolution of two distributions with compact support also has compact support. This means that if we have three distributions Λ, Ψ , and Φ , two of which have compact support, then the distributions $\Lambda * (\Psi * \Phi)$ and $(\Lambda * \Psi) * \Phi$ are well defined, so convolution is associative and commutative. We calculate that for any test function ϕ ,

$$(\Lambda * (\Psi * \Phi)) * \phi = \Lambda * (\Psi * (\Phi * \phi))$$

$$((\Lambda * \Psi) * \Phi) * \phi = (\Lambda * \Psi) * (\Phi * \phi)$$

If Φ has compact support, then $\Phi * \phi$ has compact support, and so we can move $(\Lambda * \Psi)$ into the equation to prove equality. If Φ does not have compact support, then Λ and Ψ have compact support, and

$$\Lambda * (\Psi * \Phi) = \Lambda * (\Phi * \Psi)$$

and we can apply the previous case to obtain that this is equal to $(\Lambda * \Phi) * \Psi$. Repeatedly applying the previous case brings this to what we want.

Theorem 7.23. *If Λ and Ψ are distributions, then*

$$D^\alpha(\Lambda * \Psi) = (D^\alpha \Lambda) * \Psi = \Lambda * (D^\alpha \Psi)$$

Proof. The Dirac delta function δ satisfies

$$(\delta * \phi)(x) = \int \phi(y) \delta(x - y) dy = \phi(x)$$

so $\delta * \phi = \phi$. Now $D^\alpha \delta$ is also supported at x , since

$$(D^\alpha \delta)(\phi) = (-1)^{|\alpha|} \int \delta(x) (D^\alpha \phi)(x) dx = (-1)^{|\alpha|} (D^\alpha \phi)(0)$$

which means that for any distribution Λ , then $(D^\alpha \delta) * \Lambda$ has compact support,

$$(((D^\alpha \delta) * \Lambda) * \phi)(0) = (D^\alpha \delta)((\Lambda * \phi)^*) = (-1)^{|\alpha|} D^\alpha(\Lambda * \phi)^* = ((D^\alpha \Lambda) * \phi)(0)$$

which verifies that $(D^\alpha \delta) * \Lambda = \delta * (D^\alpha \Lambda)$. But now we find

$$D^\alpha(\Lambda * \Psi) = (D^\alpha \delta) * \Lambda * \Psi = ((D^\alpha \delta) * \Lambda) * \Psi = D^\alpha \Lambda * \Psi$$

$$D^\alpha(\Lambda * \Psi) = D^\alpha(\Psi * \Lambda) = (D^\alpha \Psi) * \Lambda = \Lambda * (D^\alpha \Psi)$$

which verifies the theorem in general. □

Part III

Operator Theory

Chapter 8

Banach Algebras

In this chapter, we develop the machinery to study the various classes of bounded operators which occur when classifying transformations on a space. To do this, we employ tools from abstract algebra. Recall that an algebra over a field F is a (not necessarily unital) ring A together with a fixed embedding of F into A , which gives A a vector space structure. A consistent norm attached to an algebra is a nice situation to study operators over a Banach space. What's more, the general theory gives light to many other circumstance which would not have been apparent were we just analyzing concrete operators. We shall use capital letters, like M and N , to denote abstract elements of these algebras, and we denote algebras by capital letters near the beginning of the alphabet, such as A or B . A **Banach Algebra** is a Banach space which is also an algebra, and satisfies

$$\|MN\| \leq \|M\|\|N\| \quad \|1\| = 1 \quad (8.1)$$

Thus multiplication is a continuous operation. We shall only consider Banach algebras over the complex numbers, because here we can apply the unique properties of holomorphicity and conjugation. Why this is so important will become clear over time.

Example. Let K be a compact space. Then the space $C(K)$ of complex-valued continuous functions is a Banach algebra under the uniform convergence norm $\|\cdot\|_\infty$ and with pointwise multiplication. If K consists of n points, then $C(K) \cong \mathbb{C}^n$, which is perhaps the most basic commutative Banach algebra.

Example. The space $L_\infty(X)$ of essentially bounded functions on a measure space X is a Banach algebra. More generally, the space $C_b(X)$ of continuous,

bounded complex-valued functions on any topological space X is a Banach algebra. If μ is the measure of X , then provided $\mu(X) > 0$, $L_\infty(X)$ has a unit.

Example. If X is locally compact, then the space $C_0(X)$ of functions which vanish at infinity form a Banach algebra, except that the space does not always contain an identity. To add an identity, we enlarge the space to $C_c(X)$, the space of eventually constant continuous functions f – those functions for which there is $\lambda \in \mathbf{C}$ such that $f - \lambda$ vanishes at infinity.

Example. If K is a compact neighbourhood in \mathbf{C} , then we define $A(K)$ to be the set of continuous functions on K which are analytic in K° . Since uniform convergence preserves holomorphicity, $A(K)$ is a closed subalgebra of $C(K)$, and is therefore a Banach algebra. $A(\mathbf{D})$ is known as the disk algebra.

Example. Let G be locally compact, with Haar measure μ , and consider the space of functions $L^1(G)$, where multiplication is convolution,

$$(f * g)(x) = \int f(y)g(y^{-1}x)d\mu(y)$$

The space is a Banach algebra, since applying Tonelli's theorem (on each component of G , which is σ -compact) and the translation invariance of the Haar measure, we find $\|f * g\|_1 \leq \|f\|_1 \|g\|_1$. This algebra is commutative when G is a commutative. It does not always possess a unit, but we can enlarge the space so it does. Let $M(G)$ be the space of complex-valued Radon measures on G , and define convolution on $M(G)$ by letting

$$\int f d(\eta * \nu) = \int \int f(xy) d\eta(x) d\nu(y)$$

If we identify each f with the measure $f\mu$, such that $\int g d(f\mu) = \int gf d\mu$. In other words, $f\lambda$ is defined by the density equation

$$\frac{d(f\mu)}{d\mu} = f$$

The dirac delta function δ at the identity is a convolution identity on $M(G)$, so the set of all measures of the form $f\mu + \gamma\delta$ is a one dimensional extension of $L^1(G)$ which now has an identity.

Example. A concrete example of $L^1(G)$ is $L^1(\mathbf{Z})$, which can be taken to be the space of integer valued sequences $c = \{c_n\}$ (where $\sum |c_n| < \infty$), with convolution

$$(a * b)_n = \sum_{k \in \mathbf{Z}} a_k b_{n-k}$$

The dirac delta function here is just the sequence δ such that

$$\delta_k = \begin{cases} 1 & k = 0 \\ 0 & \text{elsewise} \end{cases}$$

If we define δ^1 to be the sequence valued at one, and zero elsewhere, then we see that δ^1 and its inverse, for any finitely non-zero sequence can be written as the sum of convolutions of δ^1 and $(\delta^1)^{-1}$, so that $L^1(\mathbf{Z})$ is generated by δ^1 and its inverse (a Banach algebra A is generated by M_1, \dots, M_n if $\mathbf{C}[M_1, \dots, M_n]$ is dense in A).

Almost all examples above are abelian algebras. One of the prime reasons to study Banach algebras is to study operators on a Banach space, which are almost always non-commutative. In fact, some folks call the study of Banach algebras ‘non-commutative analysis’.

Example. Let E be a Banach space. The space $B(E)$ of all bounded linear operators from E to itself is a Banach algebra with respect to the operator norm. It is a unital algebra, since it possesses the identity operator. The subset $K(E)$ of compact linear operators is a closed (double-sided) ideal of $B(E)$, and so is also a Banach algebra. $K(E)$ is unital if and only if E is finite dimensional.

Really, all that distinguishes a Banach space from a Banach algebra is a continuous multiplication structure, modulo the norm we use to define the topology of the space.

Proposition 8.1. Let X be a Banach space upon with a continuous multiplication structure. Then there is an equivalent norm on X which makes the space into a Banach algebra.

Proof. Embed X in $B(X)$ by defining $\Lambda_M(N) = MN$ (since multiplication on the right is continuous, Λ_M truly is in $B(X)$ rather than just being a linear map). It is trivial to verify this is an algebra morphism, and we have the inequality

$$\|M\| = \|\Lambda_M(1)\| \leq \|\Lambda_M\| \|1\| \quad (8.2)$$

which implies the map is injective, and its inverse continuous. The closed graph theorem implies the embedding is continuous, provided we can show that if $M_i \rightarrow M$, and $\Lambda_{M_i} \rightarrow \Lambda$, then $\Lambda = \Lambda_M$. Since multiplication is continuous,

$$\Lambda(N) = \lim \Lambda_{M_i}(N) = \lim M_i N = MN$$

So $\Lambda = \Lambda_M$, and we have continuity. The proof is complete, since in $B(X)$, the inequalities (8.1) are known to hold, except in one special case: if $X = (0)$, then $\|\text{id}_X\| = 0$, but in this case, it is impossible to define an algebra structure on X . \square

8.1 Invertibility in Banach Algebras

We shall begin Banach algebra theory by analyzing criteria for invertibility, which coincides with the spectral theory of these algebras. For obvious reasons, we restrict our attention to unital algebras. The set of all units in a unital algebra A will be denoted $GL(A)$, and called the general linear group of A . The **spectrum** and **resolvent** of an element M of a Banach algebra A are defined respectively as

$$\sigma_A(M) = \{\lambda \in \mathbf{C} : \lambda - M \notin GL(A)\}$$

$$\rho_A(M) = \{\lambda \in \mathbf{C} : \lambda - M \in GL(A)\}$$

The resolvent is the complement of the spectrum. When the underlying algebra is canonical, we just use $\sigma(M)$ and $\rho(M)$ to denote the spectrum. One way to view the addition of a complex number to an operator as an ‘infinitesimal shift’ in the effects of the operator. The spectrum tells us in which directions the operator goes bad under infinitesimal changes, and we shall find that knowledge of the spectrum will allow us to perturb operators more finely. A large number of theorems are based on relating these perturbations to the original operator.

Example. Let X be a space, and consider $f \in C_b(X)$. Then $\sigma(f) = \overline{f(X)}$. If $\lambda \in \rho(f)$, then

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

This function is bounded if and only if $\lambda \notin \overline{f(X)}$. If X is compact, and $f \in C(X)$, $\sigma(f) = f(X)$.

Example. Consider a Banach space E . The inverse mapping theorem tells us that $M \in B(E)$ is invertible if and only if it is bijective. If $\dim(E) < \infty$, this is simply the set of injective operators; the spectrum is then exactly the set of eigenvalues of the operator. One can consider eigenvalues in the infinite dimensional case, yet they are almost always a proper subset of the spectra. The collection of eigenvalues is known as the **point spectra**, denoted $\sigma_p(M)$.

The next lemma is incredibly important, and is an extension of the power series formula

$$\frac{1}{1-z} = \sum_{k=0}^{\infty} z^k$$

whose conclusion really relies on no properties of the complex numbers, aside from the products and sum all Banach algebras possess.

Lemma 8.2 (Neumann Series). *If $\|M\| \leq 1$, then $1 - M \in GL(A)$, and*

$$(1 - M)^{-1} = \sum_{k=0}^{\infty} M^k$$

in the sense that the right hand side also converges to a well-defined value.

Proof. The right side converges absolutely by the comparison test, since $\|M^k\| \leq \|M\|^k$. Because A is Banach, absolute convergence implies convergences, and so we are justified in the manipulation

$$(1 - M) \sum_{k=0}^{\infty} M^k = \sum_{k=0}^{\infty} (1 - M)M^k = \sum_{k=0}^{\infty} M^k - M^{k+1} = \lim_{n \rightarrow \infty} 1 - M^{n+1}$$

As $n \rightarrow \infty$, $M^{n+1} \rightarrow 0$, so the limit above tends to one. To obtain a right side inverse, note $(1 - M)$ commutes with $\sum_{k=0}^n M^k$ for each k , so that, by continuity, $(1 - M)$ commutes with $\sum_{k=0}^{\infty} M^k$, and thus $\sum_{k=0}^{\infty} M^k$ is also a right sided inverse. \square

Corollary 8.3. *If $\|1 - M\| < 1$, then $M \in GL(A)$, and*

$$M^{-1} = \sum_{k=0}^{\infty} (1 - M)^k$$

Corollary 8.4. *$GL(A)$ is an open subset of A .*

Proof. If $M \in GL(A)$, and if $\|M - N\| < 1/\|M^{-1}\|$, then

$$\|1 - M^{-1}N\| \leq \|M^{-1}\|\|M - N\| < 1$$

so $M^{-1}N \in GL(A)$, and thus $N \in GL(A)$. \square

Corollary 8.5. $\sigma(M)$ is a closed and bounded subset of \mathbf{C} , and $\rho(M)$ is open.

Proof. The map $f : \lambda \mapsto \lambda - M$ is a continuous operation, for

$$\|(\lambda - M) - (\mu - M)\| = \|\lambda - \mu\| = |\lambda - \mu|$$

Since $GL(A)$ is open, $f^{-1}(GL(A)) = \rho(M)$ is open, hence $\sigma(M)$ is closed. If $|\lambda| > \|M\|$, then $\|M/\lambda\| < 1$, so $(1 - M/\lambda) \in GL(A)$, which means $\lambda - M$ is also invertible. Thus $\sigma(A)$ is closed and bounded, hence compact. \square

We shall see that the spectra of complex algebras are never empty. This is why we mainly study complex algebras, rather than real algebras; there are even finite dimensional real operators with empty spectra.

Example. Consider the matrix

$$M = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}$$

as an element of the Banach algebra $M_2(\mathbf{R})$. The characteristic polynomial is calculated to be $\lambda^2 + 1$, so the matrix has no eigenvalues in the real numbers, and correspondingly, $\sigma(M) = \emptyset$. Over $M_2(\mathbf{C})$, we find $\sigma(M) = \{i, -i\}$, so the spectrum is non-empty in the complex extension of the Banach algebra.

Lemma 8.6. Inversion in an operator algebra A is continuous.

Proof. Let (M_n) be a sequence in $GL(A)$ converging to an invertible element M . Then, by continuity, $M_n M^{-1} \rightarrow 1$. If $\|1 - M_n M^{-1}\| < 1/2$, then

$$M_n^{-1}M = (M^{-1}M_n)^{-1} = \sum_{k=0}^{\infty} (1 - M^{-1}M_n)^k$$

It follows that

$$\begin{aligned} \|M_n^{-1}\| &\leq \|M^{-1}\| \|M_n^{-1}M\| \leq \|M^{-1}\| \sum_{k=0}^{\infty} \|(1 - M^{-1}M_n)^k\| \\ &\leq \|M^{-1}\| \sum_{k=0}^{\infty} 2^{-k} = 2\|M^{-1}\| \end{aligned}$$

Finally, we obtain convergence of inverses,

$$\|M_n^{-1} - M^{-1}\| = \|M_n^{-1}(M - M_n)M^{-1}\| \leq \|M_n^{-1}\| \|M - M_n\| \|M^{-1}\|$$

which tends to zero, since the first and third values are bounded, and the second converges to zero. Thus inversion is continuous. \square

The invertible elements $GL(A)$ cannot form a closed set of A , for A is path-connected, so there is a sequence of invertible elements x_i which converge to a non-invertible element. Fortunately, Banach algebras have a way of telling us when things are about to go wrong; the norm of the operator blows up.

Lemma 8.7. *If $x_i \rightarrow x$, each $x_i \in GL(A)$, and $x \notin GL(A)$, then $\|x_i^{-1}\| \rightarrow \infty$.*

Proof. If $\|x_n^{-1}\| \leq M$ for all n , then

$$\|1 - x_n^{-1}x\| \leq \|x_n^{-1}\| \|x_n - x\| \leq M \|x_n - x\|$$

If we choose n large enough that $\|x_n - x\| < 1/M$, then $x_n^{-1}x$ is invertible, implying x is invertible. \square

Theorem 8.8. *If B is a closed subalgebra of A , and $M \in GL(B)$, then the entire connected component of M in $B \cap GL(A)$ is contained in $GL(B)$.*

Proof. $GL(A)$ is open, so $B \cap GL(A)$ is open in B . If $M_n \rightarrow M$, with $M_n \in GL(B)$, and $M \in GL(A)$, then $\|M_n^{-1}\| \rightarrow \|M^{-1}\| < \infty$, so by the last lemma, M must be invertible in B . \square

Corollary 8.9. *If B is a closed subalgebra of A , then $\sigma_B(M)$ is obtained from $\sigma_A(M)$ by adding certain components of $\rho_A(M)$.*

Proof. $\sigma_B(M)$ is homeomorphic to the set $GL(B) \cap \{\lambda - M : \lambda \in \mathbf{C}\}$, and $\sigma_A(M)$ homeomorphic to $GL(A) \cap \{\lambda - M : \lambda \in \mathbf{C}\}$. Any component in $GL(B) \cap \{\lambda - M : \lambda \in \mathbf{C}\}$ can be enlarged to a component of $GL(B)$, so if $\lambda \in \sigma_A(M)$, consider the component of $\lambda - M$ in $GL(B)$. \square

Corollary 8.10. *If $A \subset B$ are algebras, and $\sigma_B(M) \subset \mathbf{R}$, then $\sigma_A(M) = \sigma_B(M)$.*

Proof. $\sigma_B(M)$ is a bounded subset of \mathbf{R} , so $\rho_B(M)$ is connected in \mathbf{C} . Hence $\sigma_A(M) = \sigma_B(M)$, since $\sigma_A(M) \neq \mathbf{C}$, and if we added any more points to the spectrum we would have to add the entire component, which would have to be all of $\rho_B(M)$. \square

8.1.1 Holomorphicity and Resolvent Formalism.

The fundamental theorem of spectral theory relies on heavy complex analysis, hence its restricted application to Banach algebras over the complex numbers. Define the resolvent of M , defined on $\rho(M)$ by

$$R(z; M) = (M - z)^{-1}$$

Fixing M , we obtain an analytic function into A .

Lemma 8.11. *R is analytic, in the sense that*

$$\lim_{w \rightarrow 0} \frac{R(z + w, M) - R(z, M)}{w}$$

converges to a well defined value. for all $z \in \rho(M)$

Proof. The proof is a pure computation.

$$\begin{aligned} & \lim_{w \rightarrow 0} \frac{R(z + w, M) - R(z, M)}{w} \\ &= \lim_{w \rightarrow 0} \frac{(z + w - M)^{-1} - (z - M)^{-1}}{w} \\ &= \lim_{w \rightarrow 0} (z + w - M)^{-1} \frac{(z - M) - (z + w - M)}{w} (z - M)^{-1} \\ &= \lim_{w \rightarrow 0} -(z + w - M)^{-1} (z - M)^{-1} \\ &= -(z - M)^{-2} \end{aligned}$$

We apply the continuity of multiplication and inversion. □

In Banach theory, we call such a mapping **strongly analytic**. A **weakly analytic function** is $f : \mathbf{C} \rightarrow A$ for which $\langle \phi, f \rangle$ is analytic for any choice of $\phi \in A^*$. It is clear that all strongly analytic functions are weakly analytic, which follows because

$$\lim_{w \rightarrow 0} \phi \left(\frac{f(z + w) - f(z)}{w} \right) = \phi \left(\lim_{w \rightarrow 0} \frac{f(z + w) - f(z)}{w} \right)$$

For fun, we shall also prove that all weakly analytic functions are strongly analytic.

Theorem 8.12. *Every weakly analytic function $f : D \rightarrow X$ is strongly analytic.*

Proof. Fix $\phi \in X^*$. Consider a particular contour winding counterclockwise around a point w in the domain, which is at least a unit distance away from w at any point on the contour. If $h, k \in \mathbb{C}$ are small enough that $w + h$ and $w + k$ are contained within the contour, then by the Cauchy integral theorem,

$$\begin{aligned} & \phi \left(\frac{1}{h-k} \left[\frac{f(w+h) - f(w)}{h} - \frac{f(w+k) - f(w)}{k} \right] \right) \\ &= \phi \left(\frac{f(w)}{hk} + \frac{f(w+h)}{(h-k)h} - \frac{f(w+k)}{(h-k)k} \right) \\ &= \frac{1}{2\pi i} \int_C \frac{\phi[f(z)] dz}{[z - (w+h)][z - (w+k)][z - w]} \end{aligned}$$

Find δ such that if $\|h\| < \delta$, the distance between any point on C and $w + h$ is greater than $1/2$. Then, if M is the length of C , and K is the supremum of f on C , then

$$\left| \frac{1}{2\pi i} \int_C \frac{\phi[f(z)] dz}{[z - (w+h)][z - (w+k)][z - w]} \right| \leq \frac{4MK}{2\pi} \|\phi\| = \frac{2MK}{\pi} \|\phi\|$$

Applying the Banach Steinhaus theorem (on X^* , viewing elements of X as elements of X^{**}), we conclude that for all h, k sufficiently small, there exists K such that

$$\left| \frac{f(w+h) - f(w)}{h} - \frac{f(w+k) - f(w)}{k} \right| \leq K|h - k|$$

By the completeness of X , the quotients of h and k converge to a well defined quantity as $h - k$ converges to zero. \square

The non-emptiness of the spectrum relies on Liouville's theorem, in Banach space form. The reason for this is that the resolvent is an entire function if the spectrum exists.

Theorem 8.13. *A bounded weakly analytic function $f : \mathbb{C} \rightarrow X$ is constant.*

Proof. Let $\phi \in X^*$. Then $\phi \circ f$ is an analytic function, and

$$|(\phi \circ f)(z)| \leq \|\phi\| \|f\|_\infty$$

Louville tells us there is $w \in \mathbf{C}$ such that $\phi \circ f = w$. Fix $x, y \in \mathbf{C}$. Then $\phi[f(x) - f(y)] = 0$ for all $\phi \in X^*$. By the Hahn Banach theorem, since ϕ was arbitrary, we must have $f(x) - f(y) = 0$, so $f(x) = f(y)$. \square

There is a deep relationship between complex analysis and Banach algebras. We shall return to ‘holomorphic functional analysis’ later.

Theorem 8.14. *The spectrum of a complex Banach algebra is non-empty.*

Proof. Assume $\sigma(M)$ is empty. Then $R(\cdot, M)$ is an analytic, entire function, and tends to zero at infinity, since if $|\lambda| > N\|M\|$,

$$\begin{aligned} \|R(\lambda, M)\| &= \|(\lambda - M)^{-1}\| = \frac{1}{|\lambda|} \left\| \sum_{k=0}^{\infty} \frac{M^k}{\lambda^k} \right\| \leq \sum_{k=0}^{\infty} \frac{\|M\|^k}{|\lambda|^{k+1}} \\ &\leq \frac{1}{|\lambda|} \sum_{k=0}^{\infty} \frac{1}{N^{k+1}} = \frac{1}{N(N-1)\|M\|} \end{aligned}$$

This implies $\|R(\cdot, M)\|$ is therefore constant. Since it tends to zero at infinity, $R(z, M) = 0$ for all $z \in \mathbf{C}$. But then

$$R(z, M) = (z - M)^{-1} = 0$$

and this is clearly impossible for any particular λ . \square

A cute little theorem arises from this property of Banach algebras that will surprisingly find great use in the analysis of abelian Banach algebras.

Corollary 8.15. *Every complex Banach division algebra is isometric to \mathbf{C} .*

Proof. In any unital Banach algebra A , $\mathbf{C} \cdot 1$ is isometric to \mathbf{C} , for

$$\|\lambda \cdot 1\| = |\lambda| \|1\| = |\lambda|$$

Let A be a complex division algebra, and fix $M \in A$. Pick some $\lambda \in \sigma(A)$. Then $\lambda - M \notin GL(A)$, hence $\lambda - M = 0$, i.e. $M = \lambda$. Thus $A = \mathbf{C} \cdot 1 \cong \mathbf{C}$. \square

The real case is much more complicated. There are three real division algebras: \mathbf{R} , \mathbf{C} , and \mathbf{Q} (the quaternions), and it is much more difficult to show that these are the only three.

Theorem 8.16. *If A is a Banach algebra for which M exists such that*

$$\|x\|\|y\| \leq M\|xy\|$$

for all $x, y \in \mathbf{C}$, then A is isometric to \mathbf{C} .

Proof. Let $y \in \partial GL(A)$. Then there are $y_i \rightarrow y$ with $y_i \in GL(A)$. Thus $\|y_i^{-1}\| \rightarrow \infty$. But since

$$\|y_i\|\|y_i^{-1}\| \leq M$$

we must have $\|y_i\| \rightarrow 0$, so $y = 0$. Suppose $z \notin GL(A)$. Consider the line between z and 1 . Consider

$$\lambda = \inf\{\gamma \in [0, 1] : \gamma + (1 - \gamma)z \in GL(A)\}$$

then $\gamma + (1 - \gamma)z \in \partial GL(A)$, so $\gamma + (1 - \gamma)z = 0$, and since $\gamma \neq 1$ (since $GL(A)$ is open), we conclude $z = \gamma/(1 - \gamma)$, and since z is not invertible, we must have $z = 0$. Thus A is a division algebra, so Gelfand implies that A is isometric to \mathbf{C} . \square

Theorem 8.17. *σ is ‘continuous’, in the sense that for any open set $U \subset \mathbf{C}$ containing $\sigma(M)$, there is an open neighbourhood V of M such that if $N \in V$, $\sigma(N) \subset U$.*

Proof. The map $\lambda \mapsto \|(\lambda - M)^{-1}\|$ is a bounded continuous function in U^c , with some bound K . If $\|N\| < 1/K$, and $\lambda \in U^c$, then

$$\lambda - (M + N) = (\lambda - M)(1 - (\lambda - M)^{-1}N)$$

is invertible, since $(\lambda - M)$ is invertible, since $\|(\lambda - M)^{-1}N\| < 1$. It follows that $\sigma(M + N) \subset U$. \square

8.1.2 Spectral Radii

The **spectral radius** of an element $M \in A$ is defined to be

$$r(M) = \sup\{|\lambda| : \lambda \in \sigma(M)\}$$

It bounds where to look for the spectrum of M occurs. What is amazing is that we can define the spectral radius without any algebraic reference; this is crazy, since if we enlarge our Banach algebra, more elements become invertible, and thus the spectrum shrinks in size. The radius formula says that there still exists points on the boundary of the spectrum.

Lemma 8.18. *Let $M \in A$, $n \in \mathbf{N}$. Then, if $\lambda \in \sigma(M)$, $\lambda^n \in \sigma(M^n)$.*

Proof. Suppose $\lambda \in \sigma(M)$. Then

$$\lambda^n - M^n = (\lambda - M) \left(\sum \lambda^{n-1-k} M^k \right) = \left(\sum \lambda^{n-1-k} M^k \right) (\lambda - M)$$

If $\lambda - M$ was invertible, then $\lambda^n - M^n$ would also be invertible. \square

Theorem 8.19 (Spectral Radius Theorem).

$$r(M) = \lim_{n \rightarrow \infty} \|M^n\|^{1/n}$$

Proof. If $\lambda \in \sigma(M)$, then $\lambda^n \in \sigma(M^n)$, and therefore by Neumann's lemma, $|\lambda|^n \leq \|M^n\|$. We conclude

$$|\lambda| \leq \|M^n\|^{1/n}$$

taking extrema,

$$r(M) = \sup_{\lambda \in \sigma(M)} |\lambda| \leq \liminf_{n \rightarrow \infty} \|M^n\|^{1/n}$$

Set $R = r(M)^{-1}$ (which can be ∞ , if $r(M) = 0$), and $r = \|M\|^{-1}$. Let λ be a complex number with modulus less than R . Then $1/|\lambda| > r(M)$, so $1 - \lambda M \in GL(A)$. If $\phi \in A^*$, define

$$f : \lambda \mapsto \langle \phi, (1 - \lambda M)^{-1} \rangle$$

Then f is holomorphic in the disk of radius R . If $|\lambda| < r$, then $\|\lambda M\| < 1$, $1 - \lambda M \in GL(A)$, and

$$\phi((1 - \lambda M)^{-1}) = \sum_{k=0}^{\infty} \phi(\lambda^k M^k)$$

power series expansions are unique, hence this expansion should work in the whole disk of radius R . But ϕ was arbitrary, so the sequence $\lambda^k M^k$ must be bounded, by Banach Steinhaus. If λ is fixed, then there is C such that

$$|\lambda^n| \|M^n\| \leq C$$

for all n , so

$$\|M^n\|^{1/n} \leq \frac{C^{1/n}}{|\lambda|}$$

Hence

$$\limsup_{n \rightarrow \infty} \|M^n\|^{1/n} \leq \lim_{n \rightarrow \infty} \frac{C^{1/n}}{\lambda} = \frac{1}{\lambda}$$

Letting $\lambda \rightarrow R$, we obtain that $\limsup \|M^k\|^{1/k} \leq r(a)$. We have shown

$$\liminf \|M^k\|^{1/k} \geq r(M) \geq \limsup \|M^k\|^{1/k}$$

from which the theorem follows. \square

Corollary 8.20. *The spectral radius of a Banach algebra element is invariant of which Banach algebra the element is in. If B is a Banach subalgebra of A , with $M \in B$, then $r_A(M) = r_B(a)$.*

Example. *We can isometrically embed $A(\mathbf{D})$ in $C(\mathbf{T})$ via the map $f \mapsto f|_{\mathbf{T}}$, so that we may view $A(\mathbf{D})$ as a closed subspace of $C(\mathbf{T})$. These properties follows simply from the maximum modulus principle. Let $z : \mathbf{T} \rightarrow \mathbf{C}$ be the identity map. Then*

$$\sigma_{A(\mathbf{D})}(z) = \mathbf{D} \not\subseteq \mathbf{T} = \sigma_{C_0(\mathbf{T})}(z|_{\mathbf{T}})$$

while the spectrum are different, the spectral radius is the same.

In our last example, enlarging the spectrum of f from \mathbf{T} to \mathbf{D} was the only thing that could occur from enlarging the algebra, for we cannot add any more points outside of the unit disk, and adding a single point in the unit disk will add the entire component to the spectrum.

8.1.3 Non-Unital Spectrum & Non-Commutative Topology

We would like to extend these tools to non-unital algebras, but there is no way to talk about invertibility since there is no unit to invert to! We still need to analyze these algebras, since they naturally occur in analysis. If X is not compact, then $C_0(X)$ does not contain an identity. Though we have noted that there is usually a natural trick for adding an identity, this modifies the structure of the algebra (for instance, the ideals change).

Nonetheless, it is psychologically relieving to find that all non-unital Banach algebras can be isometrically embedded into Banach algebras with unit. Given a non-unital algebra A , consider

$$A \ltimes \mathbf{C} = \{(a, \lambda) : a \in A, \lambda \in \mathbf{C}\}$$

with a multiplicative structure

$$(M + \lambda)(N + \gamma) = MN + \lambda N + \gamma M + \lambda \gamma$$

We have an identity $0 + 1$, and the space considered still satisfies the properties that make it a Banach algebra if we give it the norm

$$\|M + \lambda\| = \|M\| + |\lambda|$$

If $M_i + \lambda_i$ is a cauchy sequence, then M_i and λ_i are both separately cauchy sequences, and therefore converge to a well defined quantity, which also converges in the abstract norm. We may then define $\sigma_A(M) = \sigma_{A \ltimes \mathbf{C}}(M)$. The fundamental theorem of spectral theory still applies, as does the spectral radius formula.

Given X , we may embed $C_0(X)$ in $C(X^*)$, the continuous functions on the one-point compactification of X . We then find that $\mathbf{C} \ltimes C_0(X)$ is isometric to $C(X^*)$, so unitization on an arbitrary Banach algebra corresponds to the one-point compactification of the ‘underlying space’, even if there is no underlying space. This is the premise behind the school of non-commutative topology. We shall soon see that all commutative Banach algebras can be seen as subalgebras of continuous functions on some compact space. The school of non-commutative topology contends that the intuition gained from the algebra of continuous functions gives us intuitions about arbitrary Banach algebras, which are seen as spaces of continuous functions over ‘noncommutative space’.

8.1.4 Bounded Approximate Identities

We may not have an identity in a non-unital Banach algebra, but we may be able to ‘approximate’ an identity in some sense. An **approximate identity** (or **AI**) for an algebra A is a net $\{M_\alpha\}$ such that for any $N \in A$,

$$\lim M_\alpha N = N \quad \lim N M_\alpha = N$$

Nets for which one of these equations hold are known as **left** or **right** approximate identities. A **bounded approximate identity** (BLAI for short) is an approximate identity for which $\sup \|M_\alpha\| < \infty$.

Example. Let X be locally compact and Hausdorff. The set of all compact subsets is a directed, exhausting set. Using Urysohn's lemma, find f_K , for each compact K , such that $f_K|_K = 1$, and $\|f_K\|_\infty \leq 1$. Fix $g \in C_0(X)$. Pick K such that $|g| \leq \varepsilon$ outside of K . Then, in K , $|f_K g - g| = 0$, and outside of K , $|f_K g - g| < 2\varepsilon$. Thus $\|f_K g - g\|_\infty < 2\varepsilon$. $\{f_K\}$ is not Cauchy, for if $|f_K| < \varepsilon$ outside of K' , then $\|f_K - f_{K'}\|_\infty > 1 - \varepsilon$.

Example. Consider the net $P_r(z) = \sum_{\mathbf{z}} r^{|\mathbf{n}|} z^n$ in the group algebra $L^1(S^1)$, for $0 < r < 1$. Given a continuous function g , it is well known that $(P_r * g) \rightarrow g$ uniformly. Given $g \in L^1(S^1)$, pick $h \in C(S^1)$ with $\|g - h\|_1 < \varepsilon$. Then if $\|h * P_r - h\|_\infty < \varepsilon$ as well, then

$$\begin{aligned} \|g * P_r - g\|_1 &\leq \|(g - h) * P_r\|_1 + \|h * P_r - h\|_1 + \|h - g\|_1 \\ &\leq \|g - h\|_1 \|P_r\|_1 + \|h * P_r - h\|_\infty + \|h - g\|_1 \\ &\leq 3\varepsilon \end{aligned}$$

And we see that $g * P_r \rightarrow g$ in L^1 .

Example. If H is a Hilbert space, $B(H)$ always has the bounded approximation property, as do the classical sequence spaces c_0 and l_p . Finding a space without the approximation property was an open problem for more than 40 years. Proposed as a challenge by Stanislaw Mazur in 1936, the solver of this problem was promised a live goose. In 1972, the goose was granted to the Swedish mathematician Per Enflo, who solved the problem.

Advanced Banach space theory makes large uses of bounded left approximate identities. A Banach space E has the **approximation property** if there is a net $\{T_\alpha\}$ of finite rank operators which tend to id_E uniformly on compact subsets. The space has the **bounded approximation property** if $\|T_\alpha\|$ is bounded. Uniform convergence on compact sets is equivalent to pointwise convergence for bounded operators. Choose some compact K , fix $\varepsilon > 0$, and pick x_1, \dots, x_n for which $\{B_\varepsilon(x_i)\}$ is a cover of K . Then, if $x \in K$, there is x_j with $\|x - x_j\| < \varepsilon$, and then

$$\begin{aligned} \|T_\alpha x - x\| &\leq \|T_\alpha x - T_\alpha x_j\| + \|T_\alpha x_j - x_j\| + \|x_j - x\| \\ &\leq \varepsilon \sup \|T_\alpha\| + \|T_\alpha x_j - x_j\| + \varepsilon \end{aligned}$$

If we choose α big enough, then this is guaranteed to be less than 3ε . For any Banach space X , we denote by $A(X)$ the closure of all finite rank operators in $B(X)$. It is clear that $A(X) \subset K(X)$. The problem of whether $A(X) = K(X)$ is much more subtle.

Example. Let X be a Banach space with the bounded approximation property. Let $\{T_\alpha\}$ be a bounded net. If S is a compact operator, then $S(\overline{B_X})$ is precompact, and

$$\|T_\alpha S - S\| = \sup\{\|T_\alpha Sx - Sx\| : x \in B_X\} \leq \sup\{\|T_\alpha y - y\| : y \in \overline{S(B_X)}\}$$

The right side is the supremum over a compact set, and since $\{T_\alpha\}$ uniformly tends to the identity on compact sets, $\|T_\alpha S - S\| \rightarrow 0$. This shows $\{T_\alpha\}$ is a bounded left approximate identity, and that $A(X) = K(X)$.

Now suppose $A(X)$ has a bounded left approximate identity $\{T_\alpha\}$. Without loss of generality, we may assume each T_α is of finite rank, because if we choose, for each α , a bounded net $W_{\alpha,\beta}$ of finite rank operators such that $\lim_\beta W_{\alpha,\beta} = T_\alpha$, then $\{W_{\alpha,\beta}\}$ is a bounded left approximate identity. For $x \in X$, $\phi \in X^*$, let $x \otimes \phi : X \rightarrow X$ be defined by $(x \otimes \phi)(y) = \phi(y)x$. Then $x \otimes \phi$ obviously has finite rank. Pick y for which $\langle \phi, y \rangle = 1$. Then

$$\|T_\alpha x - x\| = \|T_\alpha(x \otimes \phi)(y) - (x \otimes \phi)(y)\| \leq \|y\| \|T_\alpha(x \otimes \phi) - (x \otimes \phi)\| \rightarrow 0$$

So the nets converge pointwise, which implies convergence on compact sets. To summarize, X has the bounded approximation property if and only if $A(X)$ has a bounded left approximate identity if and only if $K(X)$ has a bounded left approximate identity belonging to $A(X)$.

Lemma 8.21. If a space has the bounded left and right approximation properties then it has the two sided approximation property.

Proof. Let $\{M_\alpha\}$ be a left approximation identity, and $\{N_\beta\}$ a right approximation identity. We contend $\{M_\alpha + N_\beta - M_\alpha N_\beta\}$ is a two sided approximator. The limits below certainly converge, and the iterated limits must therefore equal the convergent factor, which is

$$\lim_\alpha \lim_\beta L(M_\alpha + N_\beta - M_\alpha N_\beta) = \lim_\alpha LM_\alpha + L - LM_\alpha = L$$

$$\lim_\beta \lim_\alpha (M_\alpha + N_\beta - M_\alpha N_\beta)L = \lim_\beta L + N_\beta L - N_\beta L = L$$

So we have a two sided approximator. □

Before the logician Paul Cohen got into logic, he was a functional analyst who contributed to the theory of approximation identities. We shall prove the theorem he contributed to the field, generalized to work over arbitrary modules. Let A be a Banach algebra. A **left banach A -module** is a Banach space X , which is a module over A , for which for any $M \in A$, $x \in X$,

$$\|Mx\| \leq \|M\|\|x\|$$

This makes the module operations continuous. As with Banach algebras, if module operations are continuous, then a Banach space can be made into a Banach A -module.

Now if A is a unital Banach algebra, then every $M \in A$ can be trivially written as NL , with $NL \in A$ (just let $N = 1$). If we have an approximate identity, then we may write M approximately as NL . Surprisingly, Cohen found that we need not approximate this factorization.

Theorem 8.22 (Cohen's Factorization Theorem). *Let A be a Banach algebra with BLAI $\{E_\alpha\}$, bounded by K . If X is a Banach A -module, let $x \in \overline{AX}$, and let $\varepsilon > 0$. Then there is $M \in A$ with $\|M\| \leq K$, $y \in \overline{AE}$ with $\|y - x\| < \varepsilon$, for which $x = My$.*

Proof. If A has an identity, then the proof is trivial. Since \overline{AX} is a closed subset of X , we might as well assume $\overline{AX} = X$. We may extend X to be an module over $A \rtimes \mathbb{C}$, by defining

$$(M + \lambda)x = Mx + \lambda x$$

Pick $\lambda > 0$ small enough such that

$$0 < \frac{\lambda}{1 - \lambda}K < 1$$

Define a net $\{L_\alpha\}$ in $A \rtimes \mathbb{C}$ by letting

$$L_\alpha = \lambda E_\alpha + (1 - \lambda) = (1 - \lambda) \left(1 + \frac{\lambda}{1 - \lambda} E_\alpha \right)$$

Then $L_\alpha N \rightarrow N$ for all $N \in A$, so $\{L_\alpha\}$ is a BLAI, and each L_α is invertible in $A \rtimes \mathbb{C}$ by the choice of λ , so

$$L_\alpha^{-1} = \frac{1}{1 - \lambda} \sum_{k=0}^{\infty} \left(-\frac{\lambda}{1 - \lambda} L_\alpha \right)^k$$

which implies

$$\begin{aligned}\|L_\alpha^{-1}N - N\| &= \|L_\alpha^{-1}N - L_\alpha^{-1}L_\alpha N\| \leq \|L_\alpha^{-1}\| \|N - L_\alpha N\| \\ &\leq \left(\frac{1}{1-\lambda} \sum_{k=0}^{\infty} \left(\frac{\lambda K}{1-\lambda} \right)^k \right) \|N - L_\alpha N\| \rightarrow 0\end{aligned}$$

Therefore L_α^{-1} is also a BLAI. Fix $x \in X$, and let

$$\delta < 1, \frac{\varepsilon}{2 + \|x\|}$$

We will inductively construct a sequence α_n of indices such that the sequence $\{L_{\alpha_n} \dots L_{\alpha_1}\}$ and $\{L_{\alpha_1}^{-1} \dots L_{\alpha_n}^{-1}\}$ are convergent subnets. Choose α_1 such that

$$\|L_{\alpha_1}^{-1}x - x\| < \delta/2$$

If $\alpha_1, \dots, \alpha_n$ has been chosen. There is a unique element $T_n \in A$ such that

$$L_{\alpha_n} \dots L_{\alpha_1} = (1 - \lambda)^n + T_n$$

Pick α_{n+1} such that

$$\|L_{\alpha_{n+1}}x - x\| < \frac{\delta}{\|L_{\alpha_1}^{-1} \dots L_{\alpha_n}^{-1}\| 2^{n+2}} \quad \|L_{\alpha_{n+1}}T_n - T_n\| < 1/2^n$$

Then

$$\|L_{\alpha_1}^{-1} \dots L_{\alpha_n}^{-1} L_{\alpha_{n+1}}^{-1}x - L_{\alpha_1}^{-1} \dots L_{\alpha_n}^{-1}x\| \leq \|L_{\alpha_1}^{-1} \dots L_{\alpha_n}^{-1}\| \|x - L_{\alpha_{n+1}}^{-1}x\| < \frac{\delta}{2^{n+2}}$$

This implies $\lim_{n \rightarrow \infty} L_{\alpha_1}^{-1} \dots L_{\alpha_n}^{-1}x = y$ exists, and

$$\begin{aligned}\|y - x\| &= \left\| \sum_{n=0}^{\infty} (L_{\alpha_1}^{-1} \dots L_{\alpha_{n+1}}^{-1}x - L_{\alpha_1}^{-1} \dots L_{\alpha_n}^{-1}x) \right\| \\ &\leq \sum_{n=0}^{\infty} \|L_{\alpha_1}^{-1} \dots L_{\alpha_{n+1}}^{-1}x - L_{\alpha_1}^{-1} \dots L_{\alpha_n}^{-1}x\| < \sum_{n=0}^{\infty} \frac{\delta}{2^{n+2}} < \delta/2\end{aligned}$$

By construction,

$$L_{\alpha_{n+1}}(L_{\alpha_n} \dots L_{\alpha_1}) = (1 - \lambda)^n L_{\alpha_{n+1}} + L_{\alpha_{n+1}}T_n$$

Thus

$$\begin{aligned}\|L_{\alpha_{n+1}}L_{\alpha_n}\dots L_{\alpha_1} - L_{\alpha_n}\dots L_{\alpha_1}\| &= \|(1-\lambda)^n L_{\alpha_{n+1}} + L_{\alpha_{n+1}}T_n - T_n - (1-\lambda)^n\| \\ &\leq (1-\lambda)^n \|L_{\alpha_{n+1}} - 1\| + \frac{1}{2^n}\end{aligned}$$

Thus $L_{\alpha_n}\dots L_{\alpha_1}$ converge to some element M . Since A is closed in $A \rtimes \mathbb{C}$, $M \in A$, and moreover,

$$x = \lim_{n \rightarrow \infty} (L_{\alpha_1}\dots L_{\alpha_n})(L_{\alpha_1}^{-1}\dots L_{\alpha_n}^{-1}x) = My$$

Also, observe that

$$\begin{aligned}T_{n+1} &= L_{\alpha_{n+1}}\dots L_{\alpha_1} - (1-\lambda)^{n+1} \\ &= L_{\alpha_{n+1}}T_n + L_{\alpha_{n+1}}(1-\lambda)^n - (1-\lambda)^{n+1} \\ &= L_{\alpha_{n+1}}T_n + (1-\lambda)^n \lambda e_{\alpha_{n+1}}\end{aligned}$$

Thus

$$\|T_{n+1} - T_n\| = \|(L_{\alpha_{n+1}}T_n - T_n + (1-\lambda)^n \lambda e_{\alpha_{n+1}}\| \leq 1/2^n + (1-\lambda)^n \lambda K$$

Since $M = \lim T_n$,

$$\|M\| \leq \|T_1\| + \left(\sum_{n=1}^{\infty} (1-\lambda)^n \lambda K \right) + \delta \leq \lambda K + (1-\lambda)K + \delta = K + \delta$$

Let $M' = \frac{K}{K+\delta}M$, $y' = \frac{K+\delta}{K}y$. Then $M'y' = My = x$, and by the choice of δ (kept hidden all this time),

$$\|x - y'\| \leq \frac{\|Kx - Ky\| + \|\delta y\|}{K} < \delta + \delta\|x\| + \delta^2 \leq \delta(2 + \|x\|) = \varepsilon$$

And we have verified what was needed. \square

Corollary 8.23. *If A is a non-unital Banach algebra with a BLAI, then each $M \in A$ may be written $M = NL$, with $N, L \in A$.*

Corollary 8.24. *Let A be a Banach algebra with a BLAI, and let $\{M_n\}$ be a sequence in A converging to 0. Then there is $M \in A$ and $\{N_m\}$ in A tending to zero such that $M_k = MN_k$.*

Proof. Let X be the set of all sequences in A that converge to zero, with the $\|\cdot\|_{\infty}$ norm. Then X is a left A module, and $X = \overline{AX}$, since A has a BLAI. Applying Cohen's theorem, we find that we may write

$$(M_1, M_2, \dots) = M(N_1, N_2, \dots)$$

hence $M_i = MN_i$ for each i . \square

8.2 Gelfand Theory

Gelfand realized that all commutative Banach algebras were really just manifestations of continuous functions over a compact set. The way to find a space representing a given algebra. This has incredibly important repercussions for all of Banach algebra theory, since all algebras contain commutative subalgebras. The ingenious trick to Gelfand theory is how to find a compact space on which to represent this algebra. We shall first analyze $C(K)$, for compact K , and find that points in K can be uniquely defined with maximal ideals in $C(K)$. This means that, in general, if a given algebra is to be represented by continuous functions, the only space we could represent it on would be homeomorphic to the space of maximal ideals, with some topology. In any algebra, we may consider this space, but it is only for abelian algebras where we obtain a full representation.

8.2.1 Algebra Homomorphisms and Ideals

Algebraically, if A is an algebra, the ideals are the objects we may quotient A by. Since A is both a ring and a vector space, an ideal therefore must be a ring ideal and a subspace. This continues to hold in the non-unital cases. We can consider left, right, and two-sided ideals, and we'll denote them by gaudy letters such as \mathfrak{a} and \mathfrak{b} . It is an easy consequence of ring theory that a proper left ideal cannot contain any left invertible elements, and a right ideal cannot contain any right invertible elements. It is also easy to verify that the closure of any ideal is an ideal. In the commutative case, the quotient of an algebra by a maximal ideal is a field. Furthermore, every left, right, or double sided ideal can be extended to a maximal ideal of the same type.

For completeness, we include a discussion of quotients of algebras by ideals, which will come very handy in the near future. Given an ideal \mathfrak{a} in an algebra A , we can surely form a subalgebra A/\mathfrak{a} , which consists of equivalence classes of the form

$$[M] = \{N : M - N \in \mathfrak{a}\}$$

where multiplication is obtained by performing the operations in the equivalence classes. This also has a natural Banach space structure, under the quotient norm.

$$\|X\| = \inf\{\|N\| : N \in X\}$$

It remains to be seen that this is a Banach algebra. By definition, for each M and $\varepsilon > 0$, there is $N \in X$ such that

$$\|N\| \leq \|X\| + \varepsilon$$

Similarly, for another equivalence class Y , there is L such that

$$\|L\| \leq \|Y\| + \varepsilon$$

From which it follows that

$$\|XY\| \leq \|NL\| \leq \|N\|\|L\| \leq \|X\|\|Y\| + [\|X\| + \|Y\|]\varepsilon + \varepsilon^2$$

Letting ε tend to zero, we find

$$\|XY\| \leq \|X\|\|Y\|$$

so the quotient structure really is a Banach algebra.

Ideals are naturally connected with homomorphisms $f : A \rightarrow B$ between algebras, maps which are both ring homomorphisms and linear maps. Every ideal is the kernel of some homomorphism, and the kernel of every homomorphism is an ideal. Correspondingly, *closed ideals* in a Banach algebra correspond to continuous homomorphisms. The most tractable homomorphisms to study are the **characters**, homomorphisms from a ring A to \mathbb{C} . In this section, we will deduce their structure.

Lemma 8.25. *If $\|M\| \leq 1$, then $|\phi(M)| \leq 1$ for a character ϕ . Thus $\|\phi\| \leq 1$.*

Proof. The kernel of every homomorphism cannot be all of A , and it is certainly an ideal of A . Therefore the kernel cannot contain any invertible elements. If $f(M) = \lambda$, then $\lambda - M \in \ker(M)$, so that $\lambda - M$ is not invertible. This implies, from our discussion of $\sigma(M)$, that $|\lambda| \leq 1$. \square

Corollary 8.26. *Every character is a continuous linear map.*

The next lemma follows because a proper ideal of an algebra cannot contain invertible elements.

Lemma 8.27. *Every maximal ideal of a Banach algebra is closed.*

Proof. Let \mathfrak{a} be a maximal ideal of an algebra A . It is easy to show the closure of any ideal is an ideal. It follows that either $\bar{\mathfrak{a}} = \mathfrak{a}$ (so that \mathfrak{a} is closed), or \mathfrak{a} is dense in A . Suppose the second option holds. Let $M \in GL(A)$ be chosen. Then there is $M_i \in \mathfrak{a}$ converging to M . But then the M_i are eventually invertible, since $GL(A)$ is open, from which we conclude $\mathfrak{a} = A$, a contradiction. \square

Corollary 8.28. *If \mathfrak{a} is a maximal ideal in an abelian Banach algebra, then A/\mathfrak{a} is isometric to \mathbb{C} .*

Proof. For then A/\mathfrak{a} is a Banach division algebra. \square

Lemma 8.29. *The kernel of every character is a maximal ideal. In an abelian algebra, every maximal ideal corresponds to the kernel of some character.*

Proof. We apply the first isomorphism theorem. Let ϕ be a character. Then ϕ is surjective, for $\phi(\mathbb{C} \cdot 1) = \mathbb{C}$. Thus if $\mathfrak{a} = \ker(\phi)$, $A/\mathfrak{a} \cong \mathbb{C}$, implying \mathfrak{a} is maximal. If A is commutative, and \mathfrak{a} is a maximal ideal, then A/\mathfrak{a} is a division ring, so we may project A down to A/\mathfrak{a} , and then apply an isometry to \mathbb{C} . \square

Lemma 8.30. *If the kernels of two characters correspond, then the two characters are equal.*

Proof. Suppose $\ker(\phi) = \ker(\psi)$. Fix $x \in A$. Then

$$\phi(x - \phi(x)) = \phi(x) - \phi(x) = 0$$

so

$$\psi(x - \phi(x)) = \psi(x) - \phi(x) = 0$$

Thus $\psi(x) = \phi(x)$. \square

Corollary 8.31. *In a commutative algebra, every maximal ideal of A is the kernel of a unique complex homomorphism, so there is a one-to-one correspondence between maximal ideals and characters.*

We denote the set of characters of a Banach algebra by A by Φ_A , known as the **character space** of A . Gelfand's theory is the study of Φ_A , and its relation to A .

Example. Let K be a compact space, and consider the algebra $C(K)$ of continuous functions. Given a closed subset C of K , the set

$$\{f \in C(K) : f|_C = 0\}$$

is an ideal of $C(K)$. We contend these are all such closed ideals. Let \mathfrak{a} be an arbitrary closed ideal of $C(K)$. It is easy to see that the set

$$C = \{x \in K : (\forall f \in \mathfrak{a} : f(x) = 0)\}$$

is closed. Let $x_i \rightarrow x$. If $f(x_i) = 0$ for all x_i , then $f(x_i) \rightarrow f(x)$, so $f(x) = 0$. Let g be an arbitrary continuous function with $g|_C = 0$. Fix $\varepsilon > 0$. Let $V = g^{-1}(B_\varepsilon(0))$. For each $y \in K - C$, there is a function $f_y \in \mathfrak{a}$ with $f_y(y) \neq 0$. We may assume $f_y(y) = g(y)$, since \mathfrak{a} is closed under multiplication by a constant. Let $U_y = (f_y - g)^{-1}(B_\varepsilon(0))$. Then $\{V\} \cup \{U_y\}$ is an open cover of K , and thus has a finite subcover $V, U_{y_1}, \dots, U_{y_n}$. Let $h_V, h_{y_1}, \dots, h_{y_n}$ be a partition of unity subordinate to the cover. Consider the function $g' = \sum h_{y_i} f_{y_i} \in \mathfrak{a}$. For $x \in K - V$,

$$|g'(x) - g(x)| = \left| \sum h_{y_i}(x)(f_{y_i}(x) - f_{y_i}(x)) \right| \leq \sum h_{y_i}(x) |f_{y_i}(x) - f_{y_i}(x)| < \varepsilon$$

For $x \in V$,

$$|g'(x) - g(x)| = \left| \sum h_{y_i}(x)(f_{y_i}(x) - f_{y_i}(x)) - h_V(x)f(x) \right| < \varepsilon + h_V(x)f(x) \leq 2\varepsilon$$

Thus $\|g' - g\|_\infty \leq 2\varepsilon$. It follows that g can be approximated to arbitrary precision by elements of \mathfrak{a} , so $g \in \mathfrak{a}$, since \mathfrak{a} is closed. We conclude \mathfrak{a} consists of all functions which vanish on C . Thus closed ideals of $C(K)$ are in one-to-one correspondence with closed sets of K . From this, it is fairly easy to identify the maximal ideals. If $C \subset D$, then the ideal corresponding to D is a subset of the ideal corresponding to C . The closed ideal corresponding to \emptyset is $C(K)$, so the maximal ideals correspond to the smallest, non-empty compact sets; these are just the points. Thus the character space of $C(K)$ is in one-to-one correspondence with K .

Lemma 8.32. If $M \in GL(A)$ then $M \notin \ker(\phi)$ for all $\phi \in \Phi_A$. The converse holds in commutative algebras.

Proof. If $M \in GL(A)$, we know $\phi(M) \neq 0$ for all $\phi \in \Phi_A$, for otherwise $\ker(\phi)$ would contain an invertible element, implying $\ker(\phi) = A$. Conversely, in a commutative ring, if $M \notin GL(A)$, then $(M) \neq A$, so (M) extends to a maximal ideal \mathfrak{a} , and the projection $\phi : A \rightarrow A/\mathfrak{a}$ contains M in its kernel, so we have an algebra isomorphism from A/\mathfrak{a} to \mathbb{C} by Gelfand-Mazur, which does not enlarge the kernel. \square

Corollary 8.33. *If $\phi(M) = \lambda$ for some character ϕ , then $\lambda \in \sigma(M)$. The converse holds for commutative algebras.*

Example. *Things can go very wrong for noncommutative algebras. Let us classify the ideals of $M_n(\mathbb{C})$. Let \mathfrak{a} be a left ideal of $M_n(\mathbb{C})$, and suppose that \mathfrak{a} contains a non-zero element M . It follows from the elementary theory of matrices that there are matrices N and L such that $NML \in \mathfrak{a}$ is a diagonal matrix non-zero somewhere. By multiplying on the left by a basis matrix e_{ij} , and normalizing, we find that \mathfrak{a} contains an element with a single non-zero entry on the diagonal. Multiplying by permutation matrices, we find that \mathfrak{a} contains all matrices with a one in a certain position, and zeroes everywhere else. Finally, by summing up, we find \mathfrak{a} contains all matrices in $M_n(\mathbb{C})$. Thus $M_n(\mathbb{C})$ is a simple algebra, it contains no non-trivial ideals. Consequently, we find every algebra homomorphism from $M_n(\mathbb{C})$ to another algebra must be injective, for the kernel is a proper ideal. Then, if $n > 1$, the character space of $M_n(\mathbb{C})$ is \emptyset ; there is no injective homomorphism from $M_n(\mathbb{C})$ to \mathbb{C} .*

There is a fairly natural topology on Φ_A . We assign to each $M \in A$ the **Gelfand transform** $\widehat{M} : \Phi_A \rightarrow \mathbb{C}$, which is defined by $\widehat{M}(\phi) = \phi(M)$. The **Gelfand topology** is the weakest topology on Φ_A which makes each \widehat{M} continuous. We shall denote the Gelfand transform itself as the map $\Gamma : A \rightarrow C(\Phi_A)$.

Lemma 8.34. *The map $M \mapsto \widehat{M}$ is a contractive algebra homomorphism into $C(\Phi_A)$, whose kernel is the Jacobson radical, the intersection of all maximal ideals. Thus the map is injective if and only if the algebra is semisimple. We have $\widehat{M}(\Phi_A) \subset \sigma(M)$, with equality in the commutative case.*

Proof. By direct calculation,

$$\widehat{MN}(\phi) = \phi(MN) = \phi(M)\phi(N) = (\widehat{M}\widehat{N})(\phi)$$

If $\widehat{M} = 0$, then $\phi(M) = 0$ for all ϕ . Thus M is contained in every maximal ideal. The converse is equally trivial. That $\widehat{M}(\Phi_A) \subset \sigma(M)$ follows from Corollary 8.33. In the commutative case, if $\lambda \in \sigma(M)$, then $\lambda - M$ is contained in a unique maximal ideal in A , which corresponds to a unique homomorphism $\phi \in \Phi_A$ for which $\phi(\lambda - M) = 0$. \square

It was Gelfand's idea to surplant the character spaces with a topological structure. This is the Gelfand topology of the character space. We shall now deduce its structure.

Theorem 8.35. Φ_A into a compact, Hausdorff space.

Proof. Let us first verify the Hausdorff condition. Given $\phi, \psi \in \Phi_A$, there is $M \in A$ such that $\phi(M) \neq \psi(M)$ (for otherwise they are equal as functions). If we pick a neighbourhood U of $\phi(x)$ and a neighbourhood V of $\psi(x)$ disjoint from each other, then $\widehat{M}^{-1}(U)$ and $\widehat{M}^{-1}(V)$ are disjoint neighbourhoods in Φ_A which separate ϕ and ψ .

Compactness is a little trickier. We may view Φ_A as a subset of A^* , and Φ_A has a weaker topology than induced under the relative topology of A^* . For each $\phi \in \Phi_A$, $\|\phi\| \leq 1$, so the Gelfand space is contained in the unit ball of A^* , which is compact, by the Banach-Alaoglu theorem. Thus we need only verify that Φ_A is weak * closed in A^* . Let Λ be in the weak * closure of Φ_A . Fix $M, N \in A$, $\varepsilon > 0$. Consider

$$W = \{\psi \in A^* : |\psi - \Lambda|(z) < \varepsilon \text{ for } z \in \{1, M, N, MN\}\}$$

Then W is a weak * neighbourhood of Λ , and contains some $\phi \in \Phi_A$. Thus

$$|\phi(1) - \Lambda(1)| = |1 - \Lambda(1)| < \varepsilon$$

so that $\Lambda(1) = 1$, since ε was arbitrary. Furthermore,

$$\begin{aligned} \Lambda(MN) - \Lambda(M)\Lambda(N) &= [\Lambda(MN) - \phi(MN)] + [\phi(M)\phi(N) - \Lambda(M)\Lambda(N)] \\ &= [\Lambda(MN) - \phi(MN)] + [\phi(N) - \Lambda(N)]\phi(M) \\ &\quad + [\phi(M) - \Lambda(M)]\Lambda(N) \end{aligned}$$

Hence, since ϕ is a complex homomorphism,

$$|\Lambda(MN) - \Lambda(M)\Lambda(N)| < (1 + |\phi(M)| + |\Lambda(N)|)\varepsilon \leq (1 + \|M\| + \|\Lambda\|\|N\|)\varepsilon$$

Letting $\varepsilon \rightarrow 0$, we find $\Lambda(MN) = \Lambda(M)\Lambda(N)$, so $\Lambda \in \Phi_A$, and thus Φ_A is compact. \square

Before we elaborate more on the Gelfand space, let us discuss an important property of semisimple algebras.

Proposition 8.36. *Every homomorphism $T : A \rightarrow B$ into a semisimple commutative Banach algebra B is continuous.*

Proof. Suppose $M_i \rightarrow M$, $TM_i \rightarrow N$. By the closed graph theorem, we now need only verify that $TM = N$. If $\phi : B \rightarrow \mathbf{C}$ is a character of B , then $\phi \circ T$ is continuous, and so $\phi(TM) = \phi(N)$. But this implies that $TM - N$ is in the kernel of every character, implying (by the semisimplicity) that $TM = N$. This proves continuity. \square

Thus every isomorphism between semisimple Banach algebras is a homeomorphism. The topology of a semisimple algebra is therefore completely determined by its algebraic structure.

Theorem 8.37. *If A is a unital commutative Banach algebra generated by the identity and a single element M , then Φ_A is homeomorphic to $\sigma(M)$.*

Proof. If $\phi : A \rightarrow \mathbf{C}$ is a character, then $\phi(M) \in \sigma(M)$, which gives us a map from Φ_A to $\sigma(M)$. If $\phi(M) = \psi(M)$, then $\phi = \psi$, since they must agree on all polynomials in M , and hence on A by continuity. Thus the map from Φ_A to $\sigma(M)$ is bijective. The map is continuous, for it is just $\Gamma(M)$. It is then a homeomorphism since Φ_A is compact. \square

Since $\Gamma(M)$ induces the ‘identity’ map on the Gelfand space representation $\sigma(M)$, $\Gamma(\sum a_i M^i) = \sum a_i z^i$. This implies that Φ_A contains all complex polynomials, and thus $\Gamma(N)$ is the uniform limit of polynomials in M , and therefore holomorphic (if Φ_A is open). Similarly results can be obtained if A is generated by n elements M_1, \dots, M_n , in which case the Gelfand space can be represented as a subset of \mathbf{C}^n , and $\Gamma(N)$ is the uniform limit of the images of polynomials in the M_i .

Example. Consider $\delta^1 \in L^1(\mathbf{Z})$. Then $\sigma(\delta^1) = \mathbf{T}$. We have $\sigma(\delta^1) \subset \mathbf{D}$ because $\|\delta^1\| = 1$. δ^1 has an inverse δ^{-1} , and since $\|\delta^{-1}\| = 1$, $\sigma(\delta^1)^{-1} = \sigma(\delta^{-1}) \subset \mathbf{D}$ (That the inverse of the spectrum is the spectrum of the inverse can be proved easily as an algebraic exercise, but we will soon see it is a special case of a much more general method of mapping spectra), so we conclude $\sigma(\delta^1) \subset \mathbf{T}$. In fact, we have equality. To see this, let us try and invert $\lambda\delta - \delta^1$, for some $|\lambda| = 1$. If $c * (\lambda\delta - \delta^1) = \delta$, then

$$\lambda c_n - c_{n-1} = \begin{cases} 1 & n = 0 \\ 0 & n \neq 0 \end{cases}$$

Solving these equations recursively, we find $c_n = \lambda^{-n}c_0$ for $n \geq 0$, and $c_{-n} = \lambda^{n-1}c_{-1}$. We must have $c \in L^1(\mathbf{Z})$, so that

$$|c_0| \sum_{k=0}^{\infty} 1 + |c_{-1}| \sum_{k=1}^{\infty} 1 < \infty$$

Hence $c_0 = c_{-1} = 0$, which is impossible if we desire invertibility to hold. Since δ^1 generates $L^1(\mathbf{Z})$, we find that the Gelfand space of $L^1(\mathbf{Z})$ is homeomorphic to \mathbf{T} .

It is interesting to note that if $c \in L^1(\mathbf{Z})$, then $\Gamma(c) = \sum c_n z^n$, so the Gelfand transform is essentially a representation of the Fourier series transform between \mathbf{Z} and \mathbf{T} . This follows because $\Gamma(\delta^1)$ induces the identity on \mathbf{T} , and the rest follows from linearity.

The key to Gelfand theory is noticing that characters of Banach algebras naturally reflect the structure of the Banach algebra in question. Understanding the character space leads to a natural understanding of the invertibility of the elements of the algebra.

Example. Let

$$f(z) = \sum a_n z^n$$

where $\sum |a_n| < \infty$. Then f is the Gelfand transform of the sequence $a = (a_k)$ in $L^1(\mathbf{Z})$. If f never vanishes, then $\sigma(a) = f(\mathbf{T})$ does not contain zero, so a is invertible, and therefore there is a sequence $b \in L^1(\mathbf{Z})$ such that $a * b = \delta$. Then $\Gamma(b) = 1/f$, and so we have shown that

$$(1/f)(z) \sim \sum b_n z^n$$

where $\sum |b_n| < \infty$. Before Gelfand theory, this was an incredibly difficult theorem of Fourier analysis. Similar results can be applied to non-continuous functions by approximation, and to higher dimensional trigonometric series, in which case the Gelfand space are the higher dimensional torii \mathbf{T}^n .

Wiener had a long, technical and messy proof of the lemma long before Banach algebras were invented. The Gelfand theory reduces the theorem to a few statements. Here's another example from complex analysis.

Example. Consider the commutative algebra $A(\mathbf{D})$. We also contend that $\Phi_{A(\mathbf{D})}$ are just the evaluations of functions at a point in the unit disk. The

identity function z has norm 1 in this space, implying that if ϕ is any homomorphism, then $\phi(z) = w \in \mathbf{D}$. Then

$$\phi\left(\sum_{k=0}^N a_k z^k\right) = \sum_{k=0}^N a_k w^k$$

and these polynomials are dense in $A(\mathbf{D})$, so $\phi = \phi_w$.

Now let $f_1, \dots, f_n \in A(\mathbf{D})$ be such that for each $z \in \mathbf{D}$, some $f_i(z)$ is non-zero. Then we claim there are $g_i \in A(\mathbf{D})$ for which $\sum f_i g_i = 1$. It is clear that this is equivalent to the ideal relation

$$(f_1, \dots, f_n) = A(\mathbf{D})$$

If this were not true, then (f_1, \dots, f_n) would be contained in a closed, maximal ideal, and would therefore be annihilated by some $\phi \in \Phi_{A(\mathbf{D})}$. But ϕ corresponds to evaluation at some $w \in \mathbf{D}$, so $f_i(w) = 0$ for all i , a contradiction which shows $(f_1, \dots, f_n) = A(\mathbf{D})$. One can apply Runge's theorem to classify the character spaces of arbitrary $A(C)$, but we leave this to the reader.

8.2.2 The Non-Unital Gelfand Space

Let us also address the non-unital case of Gelfand theory. In the Gelfand theory, we have a little trouble defining ideals. We call an ideal \mathfrak{a} (an additive subgroup closed under multiplication) for a non-unital algebra A **modular** if there is $N \in A$ such that $MN - M, NM - M \in \mathfrak{a}$ for all $M \in A$. If \mathfrak{a} is modular, it follows that A/\mathfrak{a} contains an identity. Maximal modular ideals exist with Zorn's help. We need to edit the proof which shows a maximal ideal is closed, which relies on an algebraic trick.

Lemma 8.38. *A maximal modular ideal in an algebra A is closed.*

Proof. If a maximal ideal \mathfrak{a} was not closed, then we would have $\bar{\mathfrak{a}} = A$. Let N be a right modular identity for \mathfrak{a} . Then there is $M \in \mathfrak{a}$ with $\|M - N\| < 1$,

so

$$\begin{aligned}
N &= (N - M) + M = \sum_{k=1}^{\infty} (N - M)^k - \sum_{k=2}^{\infty} (N - M)^k + M \\
&= \sum_{k=1}^{\infty} (N - M)^k - \left[\sum_{k=1}^{\infty} (N - M)^k \right] (N - M) + M \\
&= \underbrace{\left[\left(\sum_{k=1}^{\infty} (N - M)^k \right) M + M \right]}_{\in \mathfrak{a}} - \underbrace{\left[\left(\sum_{k=1}^{\infty} (N - M)^k \right) N - \left(\sum_{k=1}^{\infty} (N - M)^k \right) \right]}_{\in \mathfrak{a}}
\end{aligned}$$

Given any $L \in A$, $LM - M \in \mathfrak{a}$, implying $M \in \mathfrak{a}$. This implies $\mathfrak{a} = A$, an impossibility. \square

Lemma 8.39. *The kernel of any nonzero algebra homomorphism from A to \mathbf{C} is a maximal modular ideal, and in a commutative non-unital algebra, any maximal modular ideal is the kernel of some non-zero algebra homomorphism.*

Proof. If $\phi : A \rightarrow \mathbf{C}$ is a non-zero algebra homomorphism, then ϕ is surjective, for if $\phi(M) \neq 0$, then $\phi(\mathbf{C} \cdot x) = \mathbf{C}$. We claim that if $\mathfrak{a} = \ker(\phi)$, and if $\phi(M) = 1$, then M is a modular identify for \mathfrak{a} , because

$$\phi(MN - N) = \phi(M)\phi(N) - \phi(N) = \phi(N) - \phi(N) = 0$$

Implying $MN - N \in \mathfrak{a}$ for each N , so \mathfrak{a} is a modular ideal. The first isomorphism theorem gives us an isomorphism from A/\mathfrak{a} to \mathbf{C} , so \mathfrak{a} is maximal, since A/\mathfrak{a} is a field.

If A is commutative, and \mathfrak{a} is a maximal modular ideal, then A/\mathfrak{a} is an algebra with identity, whose only ideals consists of (0) and A/\mathfrak{a} , since \mathfrak{a} is maximal, so A/\mathfrak{a} is a field, and is therefore isometric to \mathbf{C} , giving us an algebra homomorphism. \square

We may therefore consider the Gelfand transform of a non-unital algebra, into the character space of all nonzero non-unital algebra homomorphisms.

Lemma 8.40. *If A is an algebra without identity, and ϕ is a character, then there is a unique non-zero algebra functional $\tilde{\phi}$ defined on $A \times \mathbf{C}$ which extends ϕ .*

Proof. Define $\tilde{\phi}(M + \lambda) = \phi(M) + \lambda$. Then $\tilde{\phi}$ is linear, $\tilde{\phi}(1) = 1$, and

$$\begin{aligned}\tilde{\phi}(M + \lambda)\tilde{\phi}(N + \gamma) &= [\phi(M) + \lambda][\phi(N) + \gamma] \\ &= \phi(MN) + \phi(\lambda N) + \phi(\gamma M) + \lambda\gamma \\ &= \tilde{\phi}((M + \lambda)(N + \gamma))\end{aligned}$$

The uniqueness of the extension follows from the fact any maximal modular ideal of A can be uniquely extended to a maximal ideal on $A \rtimes \mathbf{C}$, for the projection $M + \lambda \mapsto M$ maps ideals to ideals, and the only ideals of \mathbf{C} are (0) and (1) . \square

Corollary 8.41. *The Gelfand space extension $\Phi_{A \rtimes \mathbf{C}} = \Phi_A \cup \{\phi_\infty\}$, where*

$$\phi_\infty(M + \lambda) = \lambda$$

and $\sigma(M) \subset \widehat{M}(\Phi_A) \cup \{0\}$, with equality in the commutative case.

Example. *The unitization of $C_0(X)$, where X is locally compact, is equivalent to an embedding in $C(X^*)$. The maximal ideals of $C_0(X)$ correspond to the points in X , as well as the ‘infinity point’, which is the one-point added in the compactification.*

We may still apply a topology to Φ_A , but it is no longer compact.

Theorem 8.42. *The Gelfand topology on Φ_A is locally compact when A does not possess an identity, and the Gelfand transform maps A into $C_0(\Phi_A)$.*

Proof. Let $\psi \in \Phi_A$ be arbitrary, and fix $M \in A$ such that $\psi(M) = 1$. We shall verify that

$$K = \{\psi \in \Phi_A : |\psi(M)| \geq 1/2\}$$

is compact. We need only verify if it is closed, for if $\psi \in \Phi_A$, and if $\tilde{\psi} \in \Phi_{A \rtimes \mathbf{C}}$ is the unique extension, then $\|\tilde{\psi}\| \leq 1$, so $\|\psi\| \leq 1$. Thus K is contained in the closed unit ball of A^* . Since $K = \Lambda^{-1}([1/2, \infty))$, where $\Lambda : A^* \rightarrow \mathbf{C}$ is the map $\phi \mapsto |\phi(M)|$, then Λ is weak * continuous, so K is closed, hence compact. K contains an open neighbourhood

$$U = \{\psi \in \Phi_A : |\psi(M)| > 1/2\}$$

so Φ_A is locally compact. If $M \in A$, $\widehat{M} \in C_0(\Phi_A)$, since

$$\{\phi \in \Phi_A : |\phi(M)| \geq \varepsilon\}$$

is compact, and outside of this set $\widehat{M} \leq \varepsilon$. \square

8.3 The Riesz Calculus

8.3.1 Vector-valued integration

We would like to define the integral of arbitrary measurable functions $f : \Omega \rightarrow X$, where (Ω, μ) is a measure space, and X is some Banach space, to be some element $\int f d\mu \in X$. This is not so easy as in the real case, as we have no canonical order through which we can carry summations.

We would like our definition to have properties that ordinary integrals have, namely

$$\Lambda \left(\int f d\mu \right) = \int (\Lambda f) d\mu$$

for every $\Lambda \in X^*$. This is certainly true of other measures, since it is true of finite sums, and the integral can be fort of as some limit of a net of sums.

We shall define a function $f : \Omega \rightarrow X$ to be **measurable** if Λf is measurable for each $\Lambda \in X^*$, where $\Lambda f = \Lambda \circ f$. This is equivalent to the fact that $(\Lambda \circ f)^{-1}(U)$ is measurable for each measurable subset U of the complex plane, and therefore that f is Borel measurable when X has the weak topology. We shall say that a measurable f is **Pettis Integrable** if there is $x \in X$ for which

$$\Lambda x = \int (\Lambda f) d\mu$$

for all $\Lambda \in X^*$. We say x is the integral of f , and denote x by $\int f d\mu$. It is clear that x is unique, for if y also has the property, then $x - y$ is annihilated by all linear maps, hence $x - y = 0$.

Theorem 8.43. *Suppose (Ω, μ) is a compact measure space with a positive Borel measure, $f : \Omega \rightarrow X$ is a continuous map into a Banach space, and $\overline{\text{co}}(f(\Omega))$ is compact in X , then f is integrable, and $\int f d\mu \in \overline{\text{co}}(f(\Omega))$.*

Proof. Assume, without loss of generality, that μ is a probability measure. Let $H = \text{co}(f(Q))$. For a finite set L is a finite set of functionals, let

$$E_L = \left\{ x \in \overline{H} : \Lambda_i(x) = \int (\Lambda_i f) d\mu \right\}$$

Then $E_L \cap E_{L'} = E_{L \cup L'}$, and each E_L is compact, and therefore has the finite intersection property. If we can show that E_L is non-empty for each L , then $\bigcap E_L$ is non-empty, and $\int f d\mu$ exists. Given L , fix an ordering $(\Lambda_1, \dots, \Lambda_n)$,

which defines a function from X to \mathbf{R}^n , and consider $m = (m_1, \dots, m_n)$, where

$$m_i = \int (\Lambda_i f) d\mu$$

We shall show that m is in the convex hull of $(L \circ f)(\Omega)$. If m were not in the hull, then, since $\{m\}$ is a convex set, there is a linear functional which separates m and the convex hull of $(L \circ f)(\Omega)$, and since we know the structure of linear functionals on \mathbf{R}^n , there would be $c_1, \dots, c_n \in \mathbf{R}$ for which

$$\sum c_i v_i < \sum c_i m_i$$

whenever $v = (v_1, \dots, v_n)$ is in the hull. But then

$$\sum c_i [(\Lambda_i f)(\omega)] < \sum c_i m_i$$

and integrating,

$$\sum c_i m_i = \sum c_i \int (\Lambda_i f) < \sum c_i m_i$$

Hence m is in the convex hull. Since L is linear, $m = Ly$ for some $y \in H$. At this y , we have

$$\Lambda_i y = m_i = \int (\Lambda_i f) d\mu$$

so $y \in E_L$. Thus $\int f d\mu$ exists, and it lies in $\overline{\text{co}}(f(\Omega))$. \square

This theorem may be generalized to arbitrary complex valued measures by Jordan decomposition.

Theorem 8.44. *If X is a Banach space, $K \subset X$ is compact, and $\overline{H} = \overline{\text{co}}(K)$, then for each $x \in \overline{H}$ there is a regular Borel probability measure μ for which $x = \int \text{id}_K d\mu$.*

Proof. The Riesz representation theorem identifies $C(K)^*$ with the space R of all Borel probability measures on K . Define $\phi : C(K)^* \rightarrow X$ by $\phi(\mu) = \int \text{id}_K d\mu$. The theorem says that $\phi(R) = \overline{H}$. Certainly $K \subset \phi(R)$, for if δ_x is the unit measure at $x \in K$, then $\phi(\delta_x) = x$, for $\Lambda x = \int \Lambda d\delta_x$. Since R is convex, $\phi(R)$ is convex, and we know from the last theorem that it is a subset of \overline{H} . Thus all we need verify is that $\phi(R)$ is closed. This is a consequence of the fact that R is weak * compact in $C(Q)^*$, and ϕ is weak

* continuous into the weak topology on X , so $\phi(R)$ is weakly closed, and all weakly closed sets are strongly closed.

To see that R is weak * compact, notice that each probability measure has unit variation, and thus lies within the unit ball of $C(K)^*$, which is weak * compact. For $f \in C(K)^*$, let $E_f = \{\mu : \int f d\mu \geq 0\}$. Then each E_f is weak * closed, since the map $f \mapsto \int f d\mu$ is weak * continuous. Hence $\bigcap E_f$ is weak * closed. Similarly, the set $E = \{\mu : \int 1 d\mu = 1\}$ is weak * closed, and $E \cap \bigcap E_f = R$ is therefore weak * closed.

It is sufficient to show ϕ is weak * continuous at the origin. X has a weak subbasis of neighbourhoods at the origin of the form

$$W = \{x \in X : |\Lambda(x)| < r\}$$

for $r > 0$, $\Lambda \in X^*$. Thus we must verify that the set of μ for which $|\int \Lambda d\mu| < r$ is weak * open. But $\Lambda|_K \in C(K)$, so this set is open almost by definition. Thus ϕ is weak * continuous. \square

Theorem 8.45. *If Ω is a compact Hausdorff space with a positive Borel measure μ , X is a Banach space, and $f : \Omega \rightarrow X$ is continuous, then*

$$\left\| \int f d\mu \right\| \leq \int \|f\| d\mu$$

Proof. For each $\Lambda \in X^*$,

$$\left| \Lambda \left(\int f d\mu \right) \right| = \left| \int \Lambda f d\mu \right| \leq \int |\Lambda f| d\mu \leq \|\Lambda\| \int |f| d\mu$$

If Λ is chosen (by the Hahn Banach theorem) such that $\Lambda(\int f d\mu) = \|\int f d\mu\|$, and such that $\|\Lambda\| = 1$, then we obtain the needed inequality. \square

If A is a Banach algebra, then $\int f d\mu$ has an additional property, namely, for each $M \in A$,

$$M \int f d\mu = \int M f d\mu \quad \left(\int f d\mu \right) M = \int f M d\mu$$

This follows since multiplication on the left and right by x is a bounded linear function. If we let $L : N \mapsto MN$ be the left multiplication operator, and fix $\Lambda \in X^*$, then

$$\Lambda \left(M \left(\int f d\mu \right) \right) = (\Lambda \circ L) \left(\int f d\mu \right) = \int (\Lambda \circ L) f d\mu = \int \Lambda(Mf) d\mu$$

the right case is similar.

8.3.2 Holomorphic Functional Calculus

Given a polynomial $P \in \mathbf{C}[X]$, the definition of $P(M)$, for $M \in A$, is obvious – we just overload the operator. We don't need anything topological here. In this section, we use the topological structure of Banach algebra to evaluate holomorphic functions. For instance, in matrix theory it is useful to define the exponential of $M \in A$ by

$$e^M = \sum_{k=0}^{\infty} \frac{M^k}{k!}$$

We may define \sin and \cos similarly. More generally, for any power series $f = \sum_{k=0}^{\infty} c_k(X - \alpha)^k$ with a radius of convergence R , we may define

$$f(M) = \sum_{k=0}^{\infty} c_k(M - \alpha)^k$$

And this function is defined for any M satisfying $\|M - \alpha\| < R$. It is non-trivial to 'overload' arbitrary analytic functions, but it is very useful.

We shall use the formulation of complex analysis in terms of chains, arbitrary finite 'integer combinations' of differentiable curves $\gamma = \sum n_i \gamma_i$, where we let $-\gamma_i = \gamma_i^{-1}$ in the path homotopic sense (what we are really doing is considering the free abelian group $\mathbf{Z}\langle C^\infty(\mathbf{R}, \mathbf{C}) \rangle$, modulo the relation $-\gamma = \gamma^{-1}$). The trace of a curve will be its image,

$$\text{tr}(\sum n_i \gamma_i) = \bigcup_{n_i \neq 0} \text{Im}(\gamma_i)$$

If f is a continuous function defined on $\text{tr}(\gamma)$ valued in a Banach space, and γ_i is defined on $[a_i, b_i]$ we define the integral

$$\int_{\gamma} f(z) dz = \sum n_i \int_{\gamma_i} f(z) dz = \sum_i n_i \int_{a_i}^{b_i} (f \circ \gamma_i) \gamma_i'$$

where the integrals on the right side are Pettis integrals, if they exist.

If γ is a chain in \mathbf{C} , and $w \in \mathbf{C}$ is given, then we define the **index** of γ with respect to w , denoted $\text{Ind}_w \gamma$, to be the value

$$\frac{1}{2\pi i} \int_{\gamma} \frac{dz}{z - w}$$

A chain γ is **positively oriented** if $\text{Ind}_w \gamma \in \{0, 1\}$ for each $w \in \mathbf{C}$. We then let

$$\begin{aligned}\text{Int}(\gamma) &= \{w \in \mathbf{C} : \text{Ind}_w \gamma = 1\} \\ \text{Ext}(\gamma) &= \{w \in \mathbf{C} : \text{Ind}_w \gamma = 0\}\end{aligned}$$

Each is the union of connected components in $\mathbf{C} - \text{tr}(\gamma)$.

Lemma 8.46. *If D is open, $f : D \rightarrow X$ is a holomorphic function into a Banach space, and if γ satisfies $\text{Ind}_z \gamma = 0$ for all $z \notin D$, then $\int_\gamma f(z) dz = 0$.*

Proof. Take $\Lambda \in X^*$. Then $\Lambda f : D \rightarrow \mathbf{C}$ is holomorphic, and so

$$\int \Lambda f dz = 0$$

But this implies that $\Lambda(\int f dz) = 0$ for all Λ , hence $\int f dz = 0$. \square

Lemma 8.47. *Let D be an open set, and K a compact subset. There is a positively oriented chain whose interior contains K , and whose exterior contains D^c .*

Proof. Split D into its connected components. It suffices to prove the theorem assuming D is a component. Let $\varepsilon = d(K, D^c)$. Then $\varepsilon > 0$, since K is disjoint from ε . Let G be a set of positively oriented squares (a chain consisting of the sum of four clockwise lines), whose interior lies in D , and whose vertices are elements of the product $\mathbf{Z}/n \times \mathbf{Z}/n$, where $n > \varepsilon/2$. If we let $\gamma = \sum_{g \in G} g$, then by construction, $\text{Ind}_z \gamma = 0$ for all $z \in D^c$, and since the interiors of the squares are disjoint, the winding number around each point in K is 1. By construction, we have found a required chain. \square

We shall temporarily say that such a chain has the (K, D) separation property. We can now extend holomorphic functions to arbitrary algebras, by artificially introducing the cauchy integral theorem into the algebra. Let A be a Banach algebra, consider an element $M \in A$, D an open set containing all of $\sigma(M)$, and $f : D \rightarrow A$ a holomorphic function. Let γ be a positively oriented curve satisfying the $(\sigma(M), D)$ separation property, with $\sigma(M) \subset \text{Int}(\gamma)$. Define

$$f(M) = \int_\gamma f(z)(z - M)^{-1} dz$$

Then the extension of f is a well-defined holomorphic function. Of course, we must verify the map is well-defined.

Lemma 8.48. *Let D be an open set. If $f : D \rightarrow \mathbf{C}$ is holomorphic, and γ, λ be two chains in D satisfying the $(\sigma(M), D)$ separation property. When $f : D \rightarrow A$ is holomorphic, then*

$$\int_{\gamma} f(z)(z - M)^{-1} dz = \int_{\lambda} f(z)(z - M)^{-1} dz$$

Proof. Then $\gamma - \lambda$ has winding number zero around any point in $\sigma(M)$ and D^c , and we have already verified this case in lemma 8.46. \square

We have laid the foundations for the holomorphic functional calculus.

Theorem 8.49. *The map $f \mapsto f(M)$ is an algebra homomorphism from $\mathcal{O}(D)$ into A , which is really just evaluation for polynomials, and is continuous in the locally uniform topology.*

Proof. It can easily be verified (plagiarized from the proof in the real case) that uniform convergence preserves convergence of integration. The evaluation map is obviously linear, so we need only show the polynomial evaluation property for monomials $P = X^n$. Select a $(\sigma(M), D)$ chain γ far away enough from the origin that we may express inverses in Neumann series. Then

$$\begin{aligned} P(M) &= \frac{1}{2\pi i} \int_{\gamma} z^n (z - M)^{-1} dz = \frac{1}{2\pi i} \int_{\gamma} z^{n-1} \sum_{k=0}^{\infty} \frac{M^k}{z^k} dz \\ &= \frac{1}{2\pi i} \sum_{k=0}^{\infty} M^k \int_{\gamma} z^{n-1-k} dz = M^n \end{aligned}$$

since $\int_{\gamma} z^{n-1-k} dz \neq 0$ only when $n - 1 - k = -1$ ($n = k$).

Now we need to show the multiplicity of the evaluation map. This is done by brute calculation. Let $f, g \in \mathcal{O}(D)$ be given. Pick γ be as required,

and consider another $\tilde{\gamma}$ satisfying $\text{Int } \tilde{\gamma} \supset \overline{\text{Int } \gamma}$, $\mathbf{C} - D \subset \text{ext } \tilde{\gamma}$. Then

$$\begin{aligned}
f(M)g(M) &= \frac{-1}{4\pi^2} \left(\int_{\gamma} f(z)(z-M)^{-1} dz \right) \left(\int_{\tilde{\gamma}} g(w)(w-M)^{-1} \right) \\
&= \frac{-1}{4\pi^2} \int_{\gamma} \int_{\tilde{\gamma}} f(z)g(w)(z-M)^{-1}(w-M)^{-1} dw dz \\
&= \frac{-1}{4\pi^2} \int_{\gamma} \int_{\tilde{\gamma}} f(z)g(w) \left(\frac{1}{w-z} ((z-M)^{-1} - (w-M)^{-1}) \right) dw dz \\
&= \frac{-1}{4\pi^2} \int_{\gamma} f(z) \left(\int_{\tilde{\gamma}} \frac{g(w)}{w-z} dw \right) (z-M)^{-1} dz \\
&\quad + \frac{1}{4\pi^2} \int_{\tilde{\gamma}} g(w) \left(\int_{\gamma} \frac{f(z)}{w-z} dz \right) (w-M)^{-1} dw \\
&= \frac{1}{2\pi i} \int_{\gamma} f(z) \left(\frac{1}{2\pi i} \int_{\tilde{\gamma}} \frac{g(w)}{w-z} dw \right) (z-M)^{-1} dz \\
&= \frac{1}{2\pi i} \int_{\gamma} f(z)g(z)(z-M)^{-1} dz \\
&= (fg)(M)
\end{aligned}$$

and thus we have an algebra homomorphism. \square

In fact, we will show that this homomorphism is the unique one satisfying the properties. This would be trivial if $D = \mathbf{C}$, since we could expand all functions as limits of polynomials via their power series. To prove the theorem in general, we need an advanced theorem in complex analysis, Runge's theorem, which states that any function holomorphic in an open set D is locally uniformly approximated by rational functions whose poles lay outside of D . Now if r_i is rational, and Λ is a homomorphism such that $\Lambda P = P(M)$, then

$$\Lambda(P/Q) = \Lambda(P)\Lambda(1/Q) = \Lambda(P)\Lambda(Q)^{-1} = P(M)Q(M)^{-1} = (P/Q)(M)$$

The homomorphism is then uniquely defined, since the r_i are dense in the set of all holomorphic functions.

Example. Let K be compact, and let $f \in C(K)$. Let D be an open neighbourhood of $\sigma(f) = f(K)$. Consider the map from $\mathcal{O}(D) \rightarrow C(K)$ mapping g to

$g \circ f$. This map satisfies the properties of the holomorphic functional calculus, so it really is the holomorphic calculus in disguise.

Theorem 8.50. *Let $f : A \rightarrow B$ be a homomorphism between two Banach algebras, $M \in A$. Then $\sigma(f(M)) \subset \sigma(M)$, and if g is holomorphic in a neighbourhood of $\sigma(M)$, then $(g \circ f)(M) = (f \circ g)(M)$.*

Proof. That $\sigma(f(M)) \subset \sigma(M)$ is trivial, for if $\alpha - M$ is invertible, then $f(\alpha - M) = \alpha - f(M)$ is invertible. If g is a polynomial, then the theorem follows from the multiplicative property. But then the theorem is true for limits of polynomials, and hence for all g . \square

Corollary 8.51. *If D is a neighbourhood of $\sigma(M)$, and $f \in \mathcal{O}(D)$, then*

$$\widehat{f(M)} = f \circ \widehat{M}$$

The ultimatum of the holomorphic functional calculus is the proof of the spectral mapping theorem, that asserts that the ‘operation’ of the spectrum commutes with holomorphic functions. To extend this theorem to non-commutative algebras, we need an algebraic trick. We introduce the center $Z(A)$ of an algebra, which is defined exactly how it is defined in all other algebraic structures. It is a closed subalgebra of A , for if $M_i \rightarrow M$, and each $M_i \in A$, then

$$MN = \lim M_i N = \lim N M_i = NM$$

If S is a subset of A , then we may consider

$$Z(S) = \{M \in A : (\forall N \in S : MN = NM)\}$$

which is a commutative subalgebra of A . For any subset S , the double centralizer $Z(Z(S))$ also contains S . $Z(Z(S))$ is commutative if $S \subset Z(S)$, for then $Z(Z(S)) \subset Z(S)$, and thus commutes with itself by definition.

Now consider the commutative subalgebra $B = Z(Z(\{M\}))$ of an algebra A . If $\lambda - M$ is invertible in A , there is $N = (\lambda - M)^{-1}$. But then if $K \in Z(\{M\})$, then K commutes with $\lambda - M$, and

$$KN = N(\lambda - M)KN = NK(\lambda - M)N = NK$$

which implies that $\lambda - M$ is invertible in $Z(Z(\{M\}))$, so $\sigma_A(M) = \sigma_B(M)$.

Theorem 8.52 (Spectral Mapping Theorem). *If $\sigma(M) \subset D$, with D open, and $f \in \mathcal{O}(D)$, then $\sigma(f(M)) = f(\sigma(M))$.*

Proof. First, suppose A is commutative. By Gelfand theory,

$$\sigma(f(M)) = \widehat{f(M)}(\Phi_A) = f(\widehat{M}(\Phi_A)) = f(\sigma(M))$$

If A is not commutative, consider the commutative subalgebra $B = Z(Z(\{M\}))$. If f is analytic in a neighbourhood of $\sigma_A(M)$, then $f(M) \in B$, for there is surely a homomorphism from analytic functions in a neighbourhood of $\sigma_A(M)$ into B , and the embedding of B in A produces the unique homomorphism required. Thus we have justified the computation

$$\sigma_A(f(M)) = \sigma_B(f(M)) = f(\sigma_B(M)) = f(\sigma_A(M))$$

and the spectral mapping theorem is proved in general. \square

The holomorphic functional calculus is incredibly useful for it allows us to apply calculations on the complex plane to arbitrary algebras. Here are some immediate uses.

Theorem 8.53. *Suppose A is a Banach algebra, $M \in GL(A)$, and $\sigma(M)$ does not separate the origin from ∞ , then for each m , there is N such that $M = N^m$, and there is N such that $M = e^N$. If $\sigma(M) \subset \mathbf{R}^+$, then we may choose the N such that $\sigma(N) \subset \mathbf{R}^+$.*

Proof. Choose a branch of $\sqrt[m]{z}$ holomorphic on a neighbourhood of $\sigma(M)$. If $\sigma(M) \subset \mathbf{R}^+$, choose $\sqrt[m]{z}$ to be the principal branch. Then $\sqrt[m]{M}$ is well defined, and the function z^m is holomorphic everywhere, so $(\sqrt[m]{M})^m$ is defined, and

$$\left(\sqrt[m]{M}\right)^m = (z^m \circ \sqrt[m]{z})(M) = z(M) = M$$

The spectral mapping theorem implies $\sigma(\sqrt[m]{M}) = \sqrt[m]{\sigma(M)}$, so by the choice of the branch, if $\sigma(M) \subset \mathbf{R}^+$, $\sigma(\sqrt[m]{M}) \subset \mathbf{R}^+$. We may also define $\log(M)$, and $e^{\log(M)} = M$, by similar calculation. \square

Even for finite dimensional matrix algebras, this theorem has nontrivial repercussions.

Corollary 8.54. *If $M \in GL_n(\mathbf{C})$, then for each m , there is $N \in GL_n(\mathbf{C})$ such that $M = N^m$, and there is N such that $M = e^N$.*

Example. Now consider the exponential function $\exp : A \rightarrow A$. If M and N commute, it is easy to see from the power series representation that $\exp(M + N) = \exp(M)\exp(N)$. In particular, this implies that $\exp(M)$ is invertible, and $\exp(-M) = \exp(M)^{-1}$. Thus for each M , the image of the map

$$t \mapsto \exp(tM)$$

is a one-parameter subgroup of $G(A)$, which gives us a path from the identity $\exp(0)$ to $\exp(M)$. The connected component of $G(A)$ containing the identity is the **principal component**, denoted $G_0(A)$. Since $N \mapsto MNM^{-1}$ is a continuous map leaving the identity fixed, it maps $G_0(A)$ to itself, so $G_0(A)$ is a normal subgroup. The quotient group $G(A)/G_0(A)$ is sometimes called the **index group**.

Theorem 8.55. If $P(X) = (X - \alpha_1)^{m_1} \dots (X - \alpha_s)^{m_s}$, and $P(M) = 0$, then $\sigma(M) \subset \{\alpha_1, \dots, \alpha_n\}$, and if Ω is an open subset of the plane containing $\alpha_1, \dots, \alpha_n$, then for any $f \in H(\Omega)$, there is a polynomial Q of degree less than $m_1 + \dots + m_s$, such that $f - Q = gP$, for some $g \in H(\Omega)$, so $f(M) = Q(M)$.

Proof. Applying the spectral theorem, we find $\{0\} = \sigma(P(M)) = P(\sigma(M))$, from which the spectrum of M is restricted to lie in the roots of P . The Laurent series about each α_i gives constants c_{it} such that

$$g(z) = f(z)/P(z) - \sum_{i=1}^s \sum_{k=1}^{m_i} \frac{c_{it}}{(z - \alpha_i)^k}$$

is holomorphic in Ω , and

$$gP = f - \sum_{i=1}^s \sum_{k=1}^{m_i} c_{it} P(z - \alpha_i)^{-k}$$

constructing Q with degree less than n . □

In non-unital algebras, we cannot evaluate any polynomial of the form $\sum a_i z^i$, since $a_0 \in \mathbf{C}$ does not necessarily have an interpretation in the algebra. If $a_0 = 0$, we can evaluate $\sum a_i z^i$, and thus any holomorphic function f , provided $f(0) = 0$. To verify this, one need only use the simple trick of switching from a non-unital algebra A to a unital algebra $\mathbf{C} \times A$.

Chapter 9

Involutive Banach Algebras

An important property of the complex numbers is the existence of the conjugation operation $z \mapsto \bar{z}$, which allows us to ‘reverse’ the orientation of the space, and can therefore be viewed as a function which enables us to study non-holomorphic functions on the space. We shall find that Banach algebras that have a ‘conjugation operation’ have a much richer theory, analogous to ‘infinite dimensional’ complex theory.

Define an **involution** on a Banach algebra A to be an antilinear map $M \mapsto M^*$, which satisfies

$$M^{**} = M \quad (MN)^* = N^*M^*$$

A Banach algebra with a fixed involution possesses a canonical way to reverse the rotational structure provided by the complex structure on the algebra. A **Banach * Algebra** is a Banach algebra with a fixed involution. Note that an arbitrary involution could be highly discontinuous, so we normally require a least some structure on our * algebras. A **C* algebra** is a Banach * algebra A which satisfies

$$\|M^*M\| = \|M\|^2$$

for all $M \in A$.

Example. Let H be a Hilbert space. Then the map $T \mapsto T^*$ gives an involution in $B(H)$, where T^* is defined by the equation $\langle T^*x, y \rangle = \langle x, Ty \rangle$. Now if $\|x\|, \|y\| \leq 1$, then

$$|\langle T^*Tx, y \rangle| = |\langle Tx, Ty \rangle| \leq \|T\|^2$$

and this is attained, for if there are $\|x_i\| \leq 1$ such that $\|Tx_i\| \rightarrow \|T\|$, then

$$\langle T^*Tx_i, x_i \rangle = \|Tx_i\|^2 \rightarrow \|T\|^2$$

so $\|T^*T\| = \|T\|^2$. If we are working over \mathbf{C}^n , then each element of $B(\mathbf{C}^n)$ can be expressed as a matrix in $M_n(\mathbf{C})$. If (a_{ij}) is such a matrix, then one verifies that $(a_{ij})^* = (\overline{a_{ji}})$, so involution here is analogous to the complex transpose on sets of matrices.

Example. The map $f \mapsto \overline{f}$ is an involution in $C_0(X)$, where X is a locally compact Hausdorff space, and

$$\|f\overline{f}\|_\infty = \| |f|^2 \|_\infty = \|f\|_\infty^2$$

The central result of commutative involution theory shows that all commutative $*$ algebras are isometric to a subspace of $C_0(X)$.

Example. For $f \in A(\mathbf{D})$, define $f^*(z) = \overline{f(\overline{z})}$, then $*$ is an isometric involution. If $f(z) = e^{iz}$, then

$$(ff^*)(z) = e^{iz}\overline{e^{i\overline{z}}} = e^{iz}e^{-iz} = 1$$

Thus $\|ff^*\|_\infty = 1$, yet $f(1) = f^*(-1) = e > 1$, so

$$\|f\|_\infty \|f^*\|_\infty > \|ff^*\|_\infty$$

which implies $A(\mathbf{D})$ is not a C^* algebra.

Example. $L^1(G)$ has the involution $\overline{f}(g) = \overline{f(g^{-1})}$. It also has the involution $\overline{f}(g) = \overline{f(g)}$, but it turns out the first is much more important to the theory of groups. Neither give $L^1(G)$ the structure of a C^* algebra.

Proposition 9.1. In a C^* algebra, $\|M\| = \|M^*\|$.

Proof. Note that

$$\|M\|^2 = \|MM^*\| \leq \|M\| \|M^*\|$$

so $\|M\| \leq \|M^*\|$. But then, by symmetry,

$$\|M^*\| \leq \|M^{**}\| = \|M\|$$

so we have equality. □

For many Banach algebras, involutions will always be continuous, but C^* algebras are still much nicer objects to study.

Proposition 9.2. *If a Banach algebra A is commutative and semisimple, then every involution on A is continuous.*

Proof. If $T : A \rightarrow A$ is an involution, and ϕ is a character on A , then $\overline{\phi \circ T}$ is a character, and therefore continuous. This implies that if $a_i \rightarrow a$, and $Ta_i \rightarrow b$, then $\overline{\phi(Ta)} = \overline{\phi(b)}$, hence $\phi(Ta) = \phi(b)$. Since this is true for every character ϕ , $Ta = b$. \square

Proposition 9.3. *For any M in a C^* algebra,*

$$\|M\| = \sup\{\|MN\| : \|N\| \leq 1\} = \sup\{\|NM\| : \|N\| \leq 1\}$$

Proof. If $N = M^* \|M\|^{-1}$, $\|N\| = 1$, and

$$\|MN\| = \|MM^* \|M\|^{-1} = \|M\|$$

Thus the suprema on the right is greater than or equal to $\|M\|$. But if $\|N\| \leq 1$, then

$$\|MN\| \leq \|M\| \|N\| \leq \|M\|$$

so the suprema is also less than or equal to $\|M\|$. \square

It follows that we may isometrically embed any C^* -algebra A into $B(A)$, by either considering the left or right multiplications. This is known as the **left regular representation**. Thus if A is a C^* algebra without identity, then we may embed A in $B(A)$, and then consider the smallest closed C^* -subalgebra of $B(A)$ which contains A and the identity. The representation theory of C^* algebras is quite rich, and we will soon learn about the Gelfand Naimark Segal theorem, which gives rise to a universal representation of A as a subalgebra of operators on a Hilbert space.

Given a (nonunital) C^* algebra A , there is a natural C^* structure on $A^\#$, given by

$$(\lambda + M)^* = \overline{\lambda} + M^*$$

given that we define a new norm

$$\|\lambda + M\| = \sup\{\|MN + \lambda\| : N \in A, \|N\| \leq 1\}$$

The above proposition verifies that the embedding of A in $A^\#$ is an isometry.

C^* algebras are very closely related to operator theory. In fact, one can show that all C^* algebras are isometric to some C^* -subalgebra of $B(H)$, where H is a Hilbert space. Thus it makes sense to extend terminology from Hilbert space theory to the C^* domain. We say $M \in A$ is **self-adjoint** or **hermitian** if $M^* = M$. The set of all such elements is denoted A_{sa} , which forms a real subspace of A . Hermitian elements are analogous to the real numbers, which form a subset of A .

Proposition 9.4. *In a Banach $*$ algebra, any $M \in A$ can be uniquely written as $M = T + iS$, where T and S are self-adjoint elements.*

Proof. Write

$$T = \frac{M + M^*}{2} \quad S = \frac{M - M^*}{2i}$$

Then $T + iS = M$, and they are trivially verified to be self-adjoint. Now suppose $M = N + iL$, where N and L are self-adjoint. Then

$$0 = (N - T) + i(L - S)$$

is normal. Thus

$$(N - T) + i(L - S) = (N - T) - i(L - S)$$

And $i(L - S) = -i(L - S)$, which can only occur when $L - S = 0$. But then $N - T = 0$ also. \square

An element M is **normal** if

$$M^*M = MM^*$$

These elements are important because the smallest C^* subalgebra containing M will be commutative, in which case we can apply Gelfand theory to its full extent. **Unitary** elements satisfy

$$MM^* = M^*M = 1$$

the set of all such elements forming a closed subgroup $U(A)$ of $GL(A)$. A **projection** is a hermitian element M such that $M^2 = M$. Thus the terminology from operator theory carries across to arbitrary $*$ algebras, along with some of the interesting properties.

Proposition 9.5. *If A is a C^* algebra, and $M \in A$ is normal, then $\|M\| = r(M)$.*

Proof. A calculation reveals

$$\|M^2\|^2 = \|(M^2)^* M^2\| = \|(M^* M)(M^* M)^*\| = \|M^* M\|^2 = \|M\|^4$$

Thus $\|M^2\|^{1/2} = \|M\|$, and more generally by induction,

$$\|M^{2^k}\|^{1/2^k} = \left(\|(M^{2^{k-1}})^2\|^{1/2}\right)^{1/2^{k-1}} = \|M^{2^{k-1}}\|^{1/2^{k-1}} = \|M\|$$

which implies

$$r(M) = \lim_{k \rightarrow \infty} \|M^k\|^{1/k} = \lim_{k \rightarrow \infty} \|M^{2^k}\|^{1/2^k} = \|M\|$$

since $\|M^{2^k}\|^{1/2^k}$ is a subsequence of $\|M^k\|^{1/k}$. □

Corollary 9.6. *Every C^* algebra A has a unique norm making it a C^* algebra.*

Proof. The spectral radius $r(M)$ is invariant of any particular $M \in A$. Thus if $\|\cdot\|$ is any C^* norm, then

$$\|M\|^2 = \|M^* M\| = r(M^* M)$$

and thus the norm is uniquely defined. □

In a commutative C^* algebra, every element is trivially normal, so

Corollary 9.7. *The Gelfand transform from a commutative C^* algebra A to $C(\Phi_A)$ is an isometry.*

Proof. $\widehat{M}(\Phi_A) = \sigma(M)$, so $\|\widehat{M}\|_\infty = r(M) = \|M\|$. □

Gelfand's main result is that the isometry is actually surjective, which means all commutative C^* algebras are just algebras of continuous functions in disguise. The surjectivity is the tricky part, which requires some work.

Proposition 9.8. *In a C^* algebra, if M is unitary, then $\sigma(M) \subset \mathbb{T}$.*

Proof. Since

$$\|M\|^2 = \|M^*M\| = \|1\| = 1$$

and because M is normal, $r(M) = \|M\| = 1$. But $M^{-1} = M^*$ is also unitary, so $r(M^{-1}) = 1$, and by the spectral mapping theorem,

$$\sigma(M^{-1}) = \sigma(M)^{-1}$$

which implies that both spectra lie on \mathbf{T} . □

Proposition 9.9. *If M is self-adjoint, then $\sigma(M) \subset \mathbf{R}$.*

Proof. The operator e^{iM} is unitary, for

$$e^{-iM} = \sum_{k=0}^{\infty} \frac{(-i)^k M^k}{k!} = \sum_{k=0}^{\infty} \left(\frac{(iM)^k}{k!} \right)^* = (e^{iM})^*$$

The spectral mapping theorem implies that $\sigma(e^{iM}) = e^{i\sigma(M)} \subset \mathbf{T}$, which implies $\sigma(M) \subset \mathbf{R}$. □

Proposition 9.10. *If A is a C^* algebra, and ϕ a character on A , then*

- (a) $\phi(M) \in \mathbf{R}$ if M is self-adjoint.
- (b) $\phi(M^*) = \phi(M)^*$.
- (c) $\phi(M^*M) \geq 0$ for all $M \in A$.
- (d) If M is unitary, then $|\phi(M)| = 1$.

Proof. (a) and (d) follow from Gelfand theory, since we may unitize our algebra, adding at most one extra character which is uniformly zero on A . Now for any M , write $M = T + iS$, and then

$$\phi(M^*) = \phi(T - iS) = \phi(T) - i\phi(S) = \overline{\phi(T) + i\phi(S)} = \overline{\phi(M)}$$

(c) follows immediately. □

The next result fully characterizes commutative C^* algebras.

Theorem 9.11 (Gelfand-Naimark). *Let A be a commutative C^* algebra. Then the Gelfand transform is a $*$ -isomorphism between A and $C(\Phi_A)$.*

Proof. We shall show \hat{A} is dense in $C(\Phi_A)$. It is surely a subalgebra. Given $M \in A$, write $M = N + iL$, with N and L self adjoint. Now

$$\widehat{M^*}(\phi) = \overline{\widehat{M}(\phi)}$$

Thus \hat{A} is a subalgebra of $C(A)$ closed under conjugation, and which separates points. The complex Stone Weierstraß tells us \hat{A} is dense in $C(\Phi_A)$, and since it is closed, we obtain that $\hat{A} = C(\Phi_A)$. \square

The Gelfand Naimark theorem has two important applications. Recall that space X is **completely regular** if it is $T1$, and if for each closed C and $x \in X - C$, there is $f \in C_b(X)$ for which $f|_C = 0$, and $f(x) = 1$. Urysohn's lemma tells us all normal spaces (for instance, those which are compact and Hausdorff, or metrizable) are completely regular.

Theorem 9.12. *Let X be a completely regular space. Then there is a compact Hausdorff space βX with an embedding $i : X \rightarrow \beta X$ onto a dense subset of βX satisfying the following universal property; $j : X \rightarrow K$ is a continuous map into a compact Hausdorff space, then there is a unique $f : \beta X \rightarrow K$ such that*

$$\begin{array}{ccc} X & \xrightarrow{i} & \beta X \\ \downarrow j & \swarrow f & \\ K & & \end{array}$$

The universal property means βX is unique up to homeomorphism.

Proof. The map $f \mapsto \bar{f}$ is an isometric involution on $C_b(X)$. Then $C_b(X)$ is a commutative C^* algebra with identity, so $\Phi_{C_b(X)}$ is compact and Hausdorff. The map

$$i : X \rightarrow \Phi_{C_b(X)} \quad x \mapsto \phi_x$$

is continuous and injective. Fix $\phi \in \beta X - \overline{i(X)}$, if it exists. Pick $f \in C(\Phi_A)$ such that $f|_{\overline{i(X)}} = 0$, and $f(\phi) = 1$. But since $\hat{\cdot}$ is an isomorphism, $f = \hat{g}$ for some $g \in C_b(X)$, so $g(x) = 0$ for all $x \in X$, so $\hat{g} = f = 0$, a contradiction, since $f \notin i(X)$. Thus $i(X)$ is dense in βX

Now we show i is an embedding. Let $\{x_\alpha\}$ be a net in X , and suppose $i(x_\alpha) \rightarrow i(x)$. Assume x_α does not tend to x . Then there is a neighbourhood U of x and a subnet $\{x_\beta\}$ of $\{x_\alpha\}$ such that $x_\beta \notin U$ for all β . Since X is

completely regular, choose $f \in C_b(X)$ such that $f|_{X-U} = 0$, and $f(x) = 1$. Then $\hat{f}(\phi_{x_\beta})$ does not tend to $\hat{f}(\phi_x)$ even pointwise, and, consequently by the weak topology on $\Phi_{C_b(X)}$, $i(x_\beta)$ does not tend to $i(x)$, a contradiction.

Finally, let K be any compact Hausdorff space, and let $j : X \rightarrow K$ be continuous. Define

$$j^* : C(K) \rightarrow C_b(X) \quad f \mapsto f \circ j \quad \hat{j} : \Phi_{C_b(X)} \rightarrow \Phi_{C(K)} \cong K \quad \phi \mapsto \phi \circ j^*$$

then \hat{j} is continuous, and for any $f \in C(K)$,

$$[(\hat{j} \circ i)(x)](f) = [\phi_x \circ j^*](f) = f(j(x))$$

Thus if Γ is the Gelfand transform from $\Phi_{C(K)}$ to K , then

$$\Gamma^{-1}((\hat{j} \circ i)(x)) = j(x)$$

so $g \circ \hat{j} \circ i = j$, thus $g \circ \hat{j}$ is a map making the diagram commute, and it must be unique, since $i(X)$ is dense in $\Phi_{C_b(X)}$. \square

Theorem 9.13. *If B is a C^* -subalgebra of A , then $GL(A) = GL(B)$.*

Proof. Any involution satisfies $1^* = 1$, since the involution is linear, and satisfies, for any $w \in \mathbf{C}$,

$$w^* = (w1)^* = w^*1^*$$

If $N \in GL(A)$, then $N^* \in GL(A)$, for

$$N^*(N^{-1})^* = (N^{-1}N)^* = 1^* = 1 \quad (N^{-1})^*N^* = (NN^{-1})^* = 1^* = 1$$

Thus NN^* is an invertible, self adjoint operator, so $\sigma_A(NN^*) \subset \mathbf{R}$, implying $\sigma_A(NN^*) = \sigma_B(NN^*)$, which implies $NN^* \in GL(B)$. But then N has a right inverse in B . Computing with N^*N instead we find N has a left inverse, so $N \in GL(B)$. \square

Corollary 9.14. *If A is a C^* algebra, and B is a C^* -subalgebra, then*

$$\sigma_A(M) = \sigma_B(M)$$

so spectral theory over C^ -algebras is invariant of a particular C^* algebra.*

For normal operators, we can extend the holomorphic functional calculus to all continuous functions. Let A be a C^* algebra, and let $M \in A$. We let $C^*(M)$ be the smallest unital C^* algebra containing M . It can be described as the smallest closed subalgebra which contains 1, M , and M^* . Certainly, the latter is contained in the former. Consider a monomial, consider products of M and M^* . Then the involution of a monomial is a monomial, and by linearity, the span is closed under involution, hence the closed span is also closed under involution.

Theorem 9.15 (Continuous Functional Calculus). *Let M be a normal element of a C^* algebra A . Then there is a unique isometric C^* -isomorphism between $C(\sigma(M))$ and $C^*(M)$ such that*

$$1 \mapsto 1 \quad id_{\sigma(M)} \mapsto A \quad \overline{id_{\sigma(M)}} \mapsto A^*$$

Proof. The uniqueness assumption is clear, since $C(\sigma(M))$ is generated by the identity function by the Stone Weierstraß theorem. Since M is normal, $C^*(M)$ is a commutative C^* algebra, so the Gelfand transform gives us an isometric isomorphism from $C^*(M)$ to $C(\Phi_{C^*(M)})$. But the map

$$\widehat{M} : \Phi_{C^*(M)} \rightarrow \Sigma(M)$$

is a continuous bijection (for $C^*(M)$ is commutative), which therefore must be a homeomorphism. Then,

$$f \mapsto \widehat{\cdot}^{-1}(f \circ \widehat{M})$$

is what we desire, where $\widehat{\cdot}^{-1}$ is the inverse Gelfand transform. We write the image of f under this map as $f(M)$. \square

Corollary 9.16. *If A is a C^* algebra, and $M \in A$ is normal, and let $f \in C(\sigma(M))$. Then*

$$f(\sigma(M)) = \sigma(f(M))$$

Proof. We just apply Gelfand theory.

$$\sigma(f(M)) = \sigma(\Gamma^{-1}(f \circ \widehat{M})) = (f \circ \widehat{M})(\Phi_{C^*(M)}) = f(\sigma(M))$$

Thus the spectrum of normal operators is nicely defined under spectral mappings. \square

For Banach $*$ algebras, Gelfand theory is not nearly as successful. We need not even have $\phi(M^*) = \overline{\phi(M)}$, for instance, in $A(\mathbf{D})$ where $f^*(z) = \overline{f(\bar{z})}$, or in $L^1(\mathbf{Z})$, where $(a^*)_n = \bar{a}_n$. We call an algebra with this property a **Symmetric Algebra**. If we define $(a^*)_n = \overline{a_{-n}}$, then $L^1(\mathbf{Z})$ becomes a symmetric algebra, because every character is of the form $\phi(a) = \sum a_k z^k$, for $z \in \mathbf{T}$, and

$$\overline{\phi(a)} = \sum \bar{a}_k z^{-k} = \sum \overline{a_{-k}} z^k = \phi(a^*)$$

which is one of the reasons why this involution is so much more useful than pointwise conjugation.

9.1 Positivity

9.1.1 Positive Elements of Algebras

A self-adjoint element M is **positive** if $\sigma(M) \subset [0, \infty)$. In shorthand, we write $M \geq 0$. The set of all positive elements in an algebra A is denoted A_+ . We write $M \leq N$ if $N - M \geq 0$. The identification of positive elements gives us a useful ordering on A_{sa} , analogous to the ordering of the real numbers.

Example. If $f \in C_b(X)$, then f is normal if and only if f is real-valued. Similarly, f is positive if and only if $f \geq 0$, in the usual sense, for this occurs if and only if $\overline{f(X)} \subset [0, \infty)$.

Proposition 9.17. If $M \in A_+$, there is a unique $N \in A_+$ for which $M = N^2$.

Proof. Consider the map $f(z) = \sqrt{z}$. This is well defined, since $\sigma(M) \subset [0, \infty)$, and $f \in C(\sigma(M))$, so we may consider $f(M)$. Now

$$f(M)^2 = (f^2)(M) = (\text{id}_{\sigma(M)})(M) = M$$

If $N^2 = M$, then

$$NM = N^3 = (N^2)N = MN$$

similarly, $(N^*)^2 = (N^2)^* = M^*$, and since N is normal,

$$NM^* = NN^*N^* = N^*N^*N = M^*N$$

so, since $f(M)$ is a limit of certain sums of monomials of M and M^* , under which N commutes,

$$Nf(M) = f(M)N$$

Consider $B = C^*(N, f(M))$. This is a commutative sub C^* -algebra of A . Applying the Gelfand transform, for any $\phi \in \Phi_B$, $\hat{N}(\phi) \geq 0$, since N is a positive operator. Similarly, $\widehat{f(A)}(\phi) \geq 0$. Since

$$\hat{N}(\phi)^2 = \widehat{M}(\phi) = \widehat{f(A)}(\phi)^2$$

we must have $\hat{N}(\phi) = \widehat{f(A)}(\phi)$, hence $\hat{N} = \widehat{f(A)}$, so $N = f(A)$, since the Gelfand transform is injective for commutative algebras. \square

We denote the unique positive square root of $M \in A_+$ by \sqrt{M} . If self-adjoint operators behave like real numbers, positive operators behave like positive real numbers.

Proposition 9.18. *Let A be a C^* algebra, and let $M \in A_{sa}$. Then there are unique $M_+, M_- \in A_+$ for which $M = M_+ - M_-$, and $M_+M_- = 0$.*

Proof. Let $f_+(t) = \max(t, 0)$, and $f_-(t) = -\min(t, 0)$. Then f_+ and f_- are continuous maps, and we may consider $f_+(M), f_-(M)$. Since

$$f_+ - f_- = \text{id}_{\sigma(M)}$$

$f_+(M) - f_-(M) = M$. By the spectral mapping theorem,

$$\sigma(f_+(M)) = f_+(\sigma(M)) \subset [0, \infty) \quad \sigma(f_-(M)) = f_-(\sigma(M)) \subset [0, \infty)$$

so $f_+(M), f_-(M) \geq 0$. Since $f_+f_- = 0$, $f_+(M)f_-(M) = 0$.

To prove uniqueness, suppose $L - U = M$, and $LU = 0$. Let $B = C^*(L, U)$. Then $M \in B$, and thus $M_+, M_- \in B$. For any $\phi \in \Phi_B$, either $\hat{L}(\phi) = 0$, or $\hat{U}(\phi) = 0$. If $\widehat{M}(\phi) \geq 0$, then $\hat{U}(\phi) = 0$, so $\hat{L}(\phi) = \widehat{M}(\phi)$, and if $\widehat{M}(\phi) \leq 0$, then $\hat{L}(\phi) = 0$, so $\hat{U}(\phi) = -\widehat{M}(\phi)$. But then $\hat{L} = \widehat{f_+(M)}$ and $\hat{U} = \widehat{f_-(M)}$. Then equality is obtained, since B is commutative, because L and U commute. \square

Example. Suppose M is a positive element of $B(H)$. Then, for any $x \in H$,

$$\langle Mx, x \rangle = \langle \sqrt{M}\sqrt{M}x, x \rangle = \langle \sqrt{M}x, \sqrt{M}x \rangle \geq 0$$

Conversely, let $M \in B(H)$ be an operator such that, for any $x \in H$,

$$\langle Mx, x \rangle \geq 0$$

Write $M = N + iL$, with N and L self adjoint. Then

$$\langle Mx, x \rangle = \langle Nx, x \rangle + i\langle Lx, x \rangle \geq 0$$

implying $\langle Lx, x \rangle = 0$ for all x , so $L = 0$, and $M = N$ is self-adjoint. The last proposition implies that we may write $M = M_+ - M_-$. Then

$$0 \leq \langle MM_-x, M_-x \rangle = -\langle M_-^2x, M_-x \rangle \leq 0$$

Since $H = \overline{M_-H} \oplus \ker M_-$, we find that

$$\langle M_-x, x \rangle = 0$$

for all x , implying $M_- = 0$, so $M = M_+$ is positive. Thus positive operators have a nice characterization on Hilbert spaces.

Lemma 9.19. *If M is self-adjoint, and $\|M - t\| \leq t$, then M is positive. If M is positive, and $\|M\| \leq t$, then $\|M - t\| \leq t$.*

Proof. Suppose $\|M - t\| \leq t$. Then $\sigma(M - t) \in [-t, t]$, so

$$\sigma(M) = \sigma(M - t) + t \in [0, 2t]$$

Conversely, if M is positive and $\|M\| \leq t$, then $\sigma(M) \subset [0, t]$, and

$$\sigma(M - t) = \sigma(M) - t \subset [-t, 0]$$

which means $\|M - t\| = r(M - t) \leq t$. □

Proposition 9.20. *If A is a C^* algebra, then*

- (a) A_+ is closed.
- (b) If $M, N \in A_+$, then $M + N \in A_+$.
- (c) If $M \in A_+$, and $t \geq 0$, then $tM \in A_+$.
- (d) $A_+ \cap -A_+ = (0)$.

Proof. Suppose that $N \in \overline{A_+}$, choose $K > \|N\|$. Fix $\varepsilon > 0$ and pick $M \in A_+$ with $\|N - M\| < \varepsilon$. Then if ε is chosen small enough, then $\|M\| < K$, and

$$\|N - K\| = \|N - M\| + \|M - K\| \leq \varepsilon + K$$

Letting $\varepsilon \rightarrow 0$, we find $N \in A_+$. To prove (b), we apply the inequality in the lemma to conclude

$$\|M + N - (\|M\| + \|N\|)\| \leq \|M - \|M\|\| + \|N - \|N\|\| \leq \|M\| + \|N\|$$

So $M + N \in A_+$. Since $\sigma(\lambda M) = \lambda\sigma(M)$, we obtain (c). To prove (d), suppose $M \in A_+ \cap -A_+$. Then $\sigma(M) = \{0\}$, and since M is normal, $\|M\| = r(M) = 0$. \square

The next corollary says that ' $|x| \leq 0$ if and only if $x = 0$ ', but in a C^* -algebra setting.

Lemma 9.21. *If $-M^*M \in A_+$, then $M = 0$.*

Proof. Since

$$\sigma(MM^*) \cup \{0\} = \sigma(M^*M) \cup \{0\}$$

which can be proved by some algebraic tricks, $-MM^* \in A_+$ as well. If we write $M = T + iS$, where T and S are normal, then

$$M^*M + MM^* = (T - iS)(T + iS) + (T + iS)(T - iS) = 2(T^2 + S^2) \in A_+$$

Thus

$$M^*M = 2(T^2 + S^2) - MM^* \in A_+$$

which implies $M^*M = 0$. \square

Proposition 9.22. *The following are equivalent*

- (a) $M \in A_+$.
- (b) There is $N \in A_+$ such that $M = N^2$.
- (c) There is $N \in A$ such that $N^*N = M$.

Proof. We have already show (a) implies (b). The proof of (c) from (b) is trivial. To prove (a) from (c), note that if $M = N^*N$, then M is certainly self-adjoint, so we may write $M = M_+ - M_-$. Let $L = NM_-$. Then

$$-L^*L = -M_-N^*NM_- = -M_-MM_- = M_-^3 \in A_+$$

Implying $NM_- = 0$, hence

$$0 = N^*L = N^*NM_- = -M_-^2 \in A_+$$

This implies $M_- = 0$, so $M = M_+ \in A_+$. □

Thus we are inclined to define $|M| = \sqrt{MM^*}$, for any $M \in A$, which is a positive element.

Proposition 9.23. *If $M \leq N$, then*

$$(a) \quad M + L \leq N + L.$$

$$(b) \quad L^*ML \leq L^*NL.$$

If in addition, $0 \leq M \leq N$, then

$$(c) \quad \|M\| \leq \|N\|.$$

$$(d) \quad \sqrt{M} \leq \sqrt{N}.$$

and if $M, N \in GL(A)$, then

$$(e) \quad 0 \leq N^{-1} \leq M^{-1}$$

Proof. (a) is trivial. To prove (b), note that

$$L^*(N - M)L = L^*\sqrt{N - M}\sqrt{N - M}L = (\sqrt{N - ML})^*(\sqrt{N - ML}) \in A_+$$

Let us prove (c). If N is positive, then $N \leq \|N\|$, which follows by Gelfand theory. Thus if M and N are positive, then

$$0 \leq M \leq N \leq \|N\|$$

But then $\phi(M) \leq \|N\|$ for each ϕ , so $\|M\| \leq \|N\|$.

Fix $\varepsilon > 0$. Write

$$(\varepsilon + \sqrt{N} + \sqrt{M})(\varepsilon + \sqrt{N} - \sqrt{M}) = T$$

where $T, U \in A_{\text{sa}}$. By calculation

$$T = \varepsilon^2 + 2\varepsilon\sqrt{N} + N - M \geq \varepsilon^2$$

Thus T is positive and invertible, so $\varepsilon + \sqrt{N} - \sqrt{M}$ is left invertible and therefore invertible. Thus $\varepsilon \notin \sigma(\sqrt{M} - \sqrt{N})$, and thus

$$\sigma(\sqrt{M} - \sqrt{N}) \subset (-\infty, 0)$$

But this implies $\sqrt{N} - \sqrt{M}$ is positive.

If $M \geq \varepsilon$, then $M^{-1} \leq 1/\varepsilon$, which follows by Gelfand theory, since $\phi(M^{-1}) = \phi(M)^{-1}$. Since

$$1 = \sqrt{M}^{-1} M \sqrt{M}^{-1} \leq \sqrt{M}^{-1} N \sqrt{M}^{-1}$$

this yields

$$\sqrt{M} N^{-1} \sqrt{M} \leq 1$$

and by conjugation again,

$$N^{-1} = \sqrt{M}^{-1} \sqrt{M} N^{-1} \sqrt{M} \sqrt{M}^{-1} \leq M^{-1}$$

□

9.1.2 Positive Approximate Units

we show that all C^* algebras have **approximate units**, which are bounded approximate identities which are also increasing nets, in the sense of the positive ordering on the C^* algebra.

Lemma 9.24. *If $0 \leq M \leq N$, then $M(1 + M)^{-1} \leq N(1 + N)^{-1}$.*

Proof. Note that

$$M(1 + M)^{-1} = 1 - (1 + M)^{-1} \quad N(1 + N)^{-1} = 1 - (1 + N)^{-1}$$

We know $1 + M \leq 1 + N$, so $(1 + N)^{-1} \leq (1 + M)^{-1}$, so

$$N(1 + N)^{-1} = 1 - (1 + N)^{-1} \leq 1 - (1 + M)^{-1} = M(1 + M)^{-1}$$

which is exactly the inequality we needed. □

Proposition 9.25. *For a C^* algebra A , the net*

$$\{M \in A_+ : \|M\| < 1\}$$

ordered by the positivity of the algebra, is an approximate unit.

Proof. The set is a net, because any $M, N \in A_+$ are less than or equal to $\max(\|M\|, \|N\|)$. Fix $M \in A_{sa}$, and let $\varepsilon > 0$. Let B be the (non-unital) C^* algebra generated by M . Then B is commutative, and $\widetilde{M} \in C_0(\Phi_B)$. Let $g \in C_0(\Phi_B)$ be a function such that $0 \leq g < 1$, such that $\|\widetilde{M} - g\widetilde{M}\| < \varepsilon^2$, and set N to be the unique element of B such that $\widetilde{N} = g$. Then $\|N\| \leq 1$, N is positive, and $\|M - MN\| < \varepsilon^2$. Let $N \leq L < 1$. Then, since $\sigma(\sqrt{1-L}) \in [0, 1]$, $\|\sqrt{1-L}\| < 1$, and

$$\begin{aligned} \|M - LM\|^2 &= \|\sqrt{1-L}^2 M\|^2 \leq \|\sqrt{1-L} M\|^2 = \|M(1-L)M\| \\ &\leq \|M(1-N)M\| \leq \|M\|(1-N)\|M\| < \|M\|\varepsilon^2 \end{aligned}$$

Decreasing ε , we find $\lim LM = M$. Similar results show $\lim ML = M$. \square

If A is an abelian C^* algebra, then $A = A_{sa}$ forms a real Riesz lattice, which can be easily seen from an isometric representation $A \cong C(X)$. Thus if $0 \leq M \leq N + L$, where $N, L \in A_+$, then there are N', L' such that $N' \leq N$, $L' \leq L$ such that $M = N' + L'$. This theorem does not hold in general non-abelian C^* algebras, but there is a weaker decomposition theorem which does hold.

Proposition 9.26. *If A is a C^* algebra, and*

$$\sum_{i=0}^m M_i^* M_i = \sum_{j=0}^n N_j^* N_j$$

then there are $L_{i,j}$ such that

$$M_i^* M_i = \sum_j L_{i,j}^* L_{i,j} \quad N_j^* N_j = \sum_i L_{i,j}^* L_{i,j}$$

Proof. We may assume A is unital. Let $X = \sum M_i^* M_i = \sum N_j^* N_j$. Write

$$L_{i,j,t} = s$$

and so on and so forth FINISH THIS LATER. \square

9.2 Ideals and Quotients of C Star Algebras

9.2.1 C Star Operations on Quotients

If we quotient a C^* algebra by a closed ideal, we certainly get a Banach algebra back. But does this algebra still have an involution, or do we need to apply additional structure to our ideals? Such discussions naturally lead to the Gelfand Naimark construction, establishing that every C^* algebra is isometric to some subalgebra of bounded operators on a Hilbert space. Thus such an endeavor is very fruitful to discuss.

Proposition 9.27. *If \mathfrak{a} is a closed left/right ideal in a C^* algebra A containing a self-adjoint element M , and if $f \in C(\sigma(M))$ satisfies $f(0) = 0$, then $f(M) \in \mathfrak{a}$.*

Proof. Choose a sequence P_i of polynomials which uniformly approximate f , with $P_i(0) = 0$ for all i . If we write

$$P_i = \sum a_{ij} X^i \bar{X}^j$$

then $P_i(M) = \sum a_{ij} X^{i+j} \in \mathfrak{a}$, and since \mathfrak{a} is closed, we find $f(M) \in \mathfrak{a}$. \square

Corollary 9.28. *If $M \in A_{sa}$ is in \mathfrak{a} , then M_+ , M_- , $|M|$, and \sqrt{M} are all in \mathfrak{a} .*

If an ideal is closed under involution, then it certainly must be a two sided ideal, for involution reverses multiplication. We shall show that all two-sided ideals are closed under involution.

Proposition 9.29. *A closed ideal in a C^* -algebra is closed under involution.*

Proof. Given a closed ideal \mathfrak{a} of a C^* algebra A , \mathfrak{a} is a (non-unital) C^* sub-algebra of A . Let $\{E_\alpha\}$ be a two-sided approximate identity for \mathfrak{a} . Then, if $M \in \mathfrak{a}$ is arbitrary,

$$\lim M^* E_\alpha = \lim (E_\alpha M)^* = (\lim E_\alpha M)^* = M^*$$

so M^* is the limit of elements of \mathfrak{a} , and is thus in \mathfrak{a} . \square

Thus if \mathfrak{a} is a closed ideal in a C^* -algebra A , then A/\mathfrak{a} is verified to be a Banach $*$ algebra. It is a little bit more technical to verify that A/\mathfrak{a} is a C^* algebra, but we shall now carry out the details.

Lemma 9.30. *If \mathfrak{a} is a closed ideal of A , and $\{E_\alpha\}$ is the positive BAI for \mathfrak{a} , then A/\mathfrak{a} , then for any $M \in A$,*

$$\|M + \mathfrak{a}\| = \lim \|M - E_\alpha M\| = \lim \|M - ME_\alpha\|$$

Proof. Fix $M \in A$, and let $\{E_\alpha\}$ be the positive BAI for \mathfrak{a} . Choose $\varepsilon > 0$, and let $N \in \mathfrak{a}$ satisfy $\|M + N\| \leq \|M + \mathfrak{a}\| + \varepsilon$. Then

$$\begin{aligned} \limsup_\alpha \|M - E_\alpha M\| &\leq \limsup_\alpha \|(1 - E_\alpha)(M + N)\| + \|E_\alpha N - N\| \\ &\leq \limsup_\alpha \|(1 - E_\alpha)(M + N)\| + \limsup_\alpha \|N - E_\alpha N\| \\ &= \limsup_\alpha \|(1 - E_\alpha)(M + N)\| \\ &\leq \|M + N\| \leq \|M + \mathfrak{a}\| + \varepsilon \end{aligned}$$

and ε was arbitrary, so

$$\limsup_\alpha \|M - E_\alpha M\| \leq \|M + \mathfrak{a}\|$$

But $E_\alpha M \in \mathfrak{a}$, so for each α ,

$$\|M - E_\alpha M\| \geq \|M + \mathfrak{a}\|$$

and since we have bounded the limsup above and below by the same value, it is in fact a limit, and has value $\|M + \mathfrak{a}\|$. Similar results hold for $\|M - ME_\alpha\|$. \square

The lemma makes sense, for E_α approximates elements of \mathfrak{a} as best as possible, so subtracting $E_\alpha M$ subtracts the best approximation of M in \mathfrak{a} , thus giving us the quotient norm.

Theorem 9.31. *If \mathfrak{a} is a closed ideal of a C^* algebra A , then A/\mathfrak{a} is a C^* algebra.*

Proof. Letting $\{E_\alpha\}$ be the positive approximate identity for \mathfrak{a} , we find

$$\begin{aligned} \|M + \mathfrak{a}\|^2 &= \lim_\alpha \|M(1 - E_\alpha)\|^2 \\ &= \lim_\alpha \|(1 - E_\alpha)M^*M(1 - E_\alpha)\| \\ &\leq \lim_\alpha \|(1 - E_\alpha)M^*M\| \\ &= \|M^*M + \mathfrak{a}\| \\ &\leq \|M^* + \mathfrak{a}\| \|M + \mathfrak{a}\| \end{aligned}$$

And it is easy to see that $\|M^* + \mathfrak{a}\| = \|M + \mathfrak{a}\|$, for \mathfrak{a} is closed under involution, so we find

$$\|M + \mathfrak{a}\|^2 = \|M^*M + \mathfrak{a}\|$$

and this is exactly the C^* identity for the quotient algebra. \square

It is often much easier to calculate the norm on A/\mathfrak{a} by applying the theory we've created. If $\pi : A \rightarrow B$ is a C^* homomorphism with kernel \mathfrak{a} , then we obtain an induced diagram of C^* morphisms

$$\begin{array}{ccc} A & \xrightarrow{\pi} & B \\ \downarrow & \nearrow \tilde{\pi} & \\ A/\mathfrak{a} & & \end{array}$$

and we know that $\tilde{\pi}$ is an isometry, since it is injective. Thus

$$\|M + \mathfrak{a}\| = \|\pi(M)\|$$

and this is normally easier to calculate than either of the two equations we can use to calculate the quotient norm.

Example. We know the closed ideals of $C(X)$ are of the form

$$\mathfrak{a}_K = \{f \in C(X) : (\forall x \in K : f(x) = 0)\}$$

for some closed set K . Then $C(X)/\mathfrak{a}_K$ is isometric to $C(K)$, because we have the C^* homomorphism $f \mapsto f|_K$, and \mathfrak{a}_K is the kernel.

Example. Let H be a Hilbert space, and \mathfrak{a} a closed ideal of $B(H)$. If $\mathfrak{a} \neq (0)$, we claim $K(H) \subset \mathfrak{a}$. For any $x, y \in H$, let $x \otimes y \in B(H)$ be the map

$$z \mapsto \langle z, y \rangle x$$

It is clear that $K(H)$ is the closed linear hull of $\{x \otimes y : x, y \in H\}$. If $T \in \mathfrak{a}$ and $T \neq 0$, then there are v, w such that $\langle Tv, w \rangle = 1$. Then

$$(x \otimes y) = (x \otimes w) \circ (Tv \otimes y) = (x \otimes w) \circ T \circ (v \otimes y) \in \mathfrak{a}$$

from which we obtain, since $x \otimes y$ was arbitrary, that $K(H) \subset \mathfrak{a}$.

Example. In an infinite dimensional, separable Hilbert space H the only closed ideals of $B(H)$ are (0) , $K(H)$, and $B(H)$. Let \mathfrak{a} be an ideal properly containing $K(H)$. Let $T \in \mathfrak{a} - K(H)$, which is without loss of generality positive. **FINISH THIS LATER WHEN BETTER AT SPECTRAL MEASURES.**

Thus in an infinite dimensional, separable Hilbert space H , the **Calkin Algebra** $C(H) = B(H)/K(H)$ has no nontrivial closed ideals.

9.2.2 * Homomorphisms

The natural morphism between C^* algebras is a $*$ homomorphism, an algebra homomorphism T which satisfies $T(M^*) = T(M)^*$.

Proposition 9.32. *If $T : A \rightarrow B$ is a $*$ morphism, then $\|T(M)\| \leq \|M\|$ for all M .*

Proof. Assume first that A is unital. For any M , $\sigma(T(M)) \subset \sigma(M)$. But then if M is normal, then $T(M)$ is normal, and

$$\|T(M)\| = r(T(M)) \leq r(M) = \|M\|$$

In general, for any M ,

$$\|T(M)\|^2 = \|T(M)T(M)^*\| = \|T(MM^*)\| \leq \|MM^*\| = \|M\|^2$$

so $\|T\| \leq 1$. In the non-unital case, T induces a $*$ -extension between $A^\#$ and $B^\#$, in which we can apply the previous case. \square

Thus every $*$ -morphism is continuous. We shall find that injective $*$ -morphisms are isometries. To prove this, we must first verify a corresponding theorem about continuous functions on a locally compact space, which says that the uniform norm on $C(X)$ is the sharpest algebra norm we can have.

Lemma 9.33. *If $\pi : A \rightarrow B$ is an injective $*$ -morphism between abelian C^* algebras, then the dual map π^* is surjective.*

Proof. Let

$$K = \{\phi|_A : \phi \in \Phi_B\}$$

If $K \neq \Phi_A$, then using the continuous functional calculus, we may find non-zero f and g which vanish at K , for which $f = fg$. Then $1 \in \sigma(g)$, so there is $\phi \in \Phi_B$ such that $\phi(g) = 1$. But this is clearly impossible. \square

Lemma 9.34. *Let $\|\cdot\|$ be a submultiplicative norm on $C(X)$. Then $\|\cdot\|_\infty \leq \|\cdot\|$.*

Proof. Let A be the completion of $C(X)$ with respect to the $\|\cdot\|$ norm. Then A is a Banach algebra, and we obtain

$$\|f\|_\infty = r(f) \leq \|f\|$$

The fact that algebraic facts can hide within them topological facts is essential to the proof. \square

Since arbitrary C^* algebras contain abelian subalgebra on mass, we can almost effortlessly harvest facts about $*$ morphisms on arbitrary algebras from the above lemma.

Proposition 9.35. *An injective, continuous morphism $\pi : A \rightarrow B$ from a C^* algebra to a Banach algebra satisfies*

$$\|M\| \leq \|\pi\| \|\pi(M)\|$$

Proof. Given M , let $A_0 = C^*(M^*M)$. Define a norm on A_0 by letting

$$\|N\| = \|\pi(N)\|$$

Then $\|\cdot\| \leq \|\cdot\|$, and

$$\|M\|^2 = \|M^*M\| \leq \|M^*M\| = \|\pi(M^*M)\| \leq \|\pi\| \|M\| \|\pi(M)\|$$

which yields the claim. \square

Corollary 9.36. *Let $\pi : A \rightarrow B$ be a $*$ homomorphism between C^* algebras. Then $\pi(A)$ is a closed C^* algebra, and π is an isometry if it is injective.*

Proof. If π is injective, it certainly has closed range by the last proposition. But in general, letting $\mathfrak{a} = \ker(\pi)$, π induces an injective map

$$\begin{array}{ccc} A & \xrightarrow{\pi} & B \\ \downarrow & \nearrow \tilde{\pi} & \\ A/\mathfrak{a} & & \end{array}$$

and $\pi(A) = \tilde{\pi}(A)$ is closed. Thus if π is injective we obtain an inverse $*$ homomorphism $\pi^{-1} : \pi(A) \rightarrow A$, and both must be contractible, hence isometries. \square

Proposition 9.37. *If B is a C^* subalgebra of A , and \mathfrak{a} a closed ideal of A , then $B + \mathfrak{a}$ is a C^* subalgebra of A .*

Proof. Completeness is a three space property, and since \mathfrak{a} is complete, we need only show that $(B + \mathfrak{a})/\mathfrak{a}$ is complete, and from this we will obtain completeness. Let $\pi : A \rightarrow A/\mathfrak{a}$ be the quotient map. But the composed $*$ morphism ψ defined by

$$B \rightarrow A \rightarrow A/\mathfrak{a}$$

tells us that the image of B is closed, and

$$\psi(B) = (B + \mathfrak{a})/\mathfrak{a}$$

so $(B + \mathfrak{a})/\mathfrak{a}$ is complete. \square

9.2.3 Positive Operators

Every C^* algebra is essentially a ring of bounded operators on a Hilbert space. The challenge is to construct a canonical Hilbert space from a C^* algebra. The study of a certain subclass of operators will be essential. A **positive** operator $\phi : A \rightarrow \mathbf{C}$ on a C^* algebra maps A_+ into $[0, \infty)$. Certainly the set of positive maps form a real subspace of A^* .

Example. If $\phi \in \Phi_A$, then ϕ is positive, for

$$\phi(M^*M) = |\phi(M)|^2$$

and every positive element can be written in the form M^*M .

Example. If X is locally compact and Hausdorff, and μ is a complex Borel measure on X , then the functional

$$f \mapsto \int f d\mu$$

from $C_0(X)$ to \mathbf{C} is positive if and only if μ is a positive measure.

Example. Consider the trace map $tr : M_n(\mathbf{C}) \rightarrow \mathbf{C}$. Any positive matrix M can be, by a change of basis, put into the form

$$\begin{pmatrix} \lambda_1 & 0 & \dots & 0 \\ 0 & \lambda_2 & \dots & 0 \\ 0 & 0 & \ddots & 0 \\ 0 & 0 & \dots & \lambda_n \end{pmatrix}$$

where $\lambda_1, \dots, \lambda_n \geq 0$. The trace is invariant of a change in basis, and thus

$$tr(M) = \sum \lambda_i \geq 0$$

Thus the trace is positive.

Example. For a fixed $x \in H$, the map $T \mapsto \langle Tx, x \rangle$ is positive.

One need not verify continuity for positive operators, for we obtain this automatically.

Proposition 9.38. A positive operator $\phi : A \rightarrow \mathbf{C}$ is continuous.

Proof. We claim the supremum

$$K = \sup\{\phi(M) : M \in A_+, \|M\| \leq 1\}$$

is finite. Otherwise, we may pick a sequence M_1, M_2, \dots with $\|M_i\| \leq 1$ and $\phi(M_i) \geq 4^n$. Let

$$M = \sum \frac{M_i}{2^i}$$

Then for each n , $2^n M \geq M_n$, so

$$\phi(M) \geq \frac{\phi(M_n)}{2^n} \geq \frac{4^n}{2^n} = 2^n$$

which yields an immediate contradiction.

Now suppose that M is an arbitrary operator, choose N and L in A_{sa} such that

$$M = N + iL = N_+ - N_- + iL_+ - iL_-$$

Then if $\|M\| \leq 1$,

$$|\phi(M)| = |\phi(N_+) - \phi(N_-) + i\phi(L_+) - i\phi(L_-)| \leq 4K$$

which shows that ϕ is continuous. □

Theorem 9.39. *The following are equivalent*

1. ϕ is positive operator.
2. For every BAI E_α contained within the positive identity on A ,

$$\|\phi\| = \lim_\alpha \phi(E_\alpha)$$

3. There is a BAI E_α in the positive identity on A for which

$$\|\phi\| = \lim_\alpha \phi(E_\alpha)$$

Proof. Without loss of generality, assume $\|\phi\| = 1$. Suppose ϕ is positive. Fix an approximate identity E_α . Then

$$\limsup_\alpha \phi(E_\alpha) \leq 1$$

On the other hand, when $\|M\| \leq 1$,

$$|\phi(M)|^2 = \lim |\phi(E_\alpha M)|^2 \leq \liminf_\alpha \phi(E_\alpha^2) \phi(M^* M)$$

□

Chapter 10

Von Neumann Algebras

10.1 Spectral Theorem for Normal Operators

In finite dimensional theory, a normal linear transformation $T : \mathbb{C}^n \rightarrow \mathbb{C}^n$ can be written

$$T = \sum \lambda_i P_i$$

where P_i is projection onto the eigenspace corresponding to the eigenvalue λ_i . In this section we apply the theory of C^* algebras to extend this theorem to arbitrary normal bounded transformations from a Hilbert space to itself. The trick is to ‘integrate’ rather than sum up the subspaces.

A **spectral measure** on a measure space (Ω, \mathcal{F}) is a function $E : \mathcal{F} \rightarrow B(H)$, for which

$$E(\Omega) = \text{id}_H$$

such that $E(S)$ is a projection for each S , and such that for a disjoint family of sets $\{S_1, S_2, \dots\}$ in \mathcal{F} , then

$$E\left(\bigcup S_i\right) = \sum E(S_i)$$

with convergence pointwise (the strong operator topology), rather than in the operator norm. It follows that $E(\emptyset) = 0$.

Lemma 10.1. *If $S \subset W$, $E(S) \leq E(W)$.*

Proof. Since $E(W - S)$ is a projection, $E(W - S) \geq 0$, so

$$E(W) = E(W - S) + E(S) \geq E(S)$$

Hence a spectral measure is monotone on projections. □

Lemma 10.2. *If S and W are disjoint sets, then $E(S)E(W) = 0$.*

Proof. $E(S \cup W) = E(S) + E(W)$, so

$$E(S \cup W) = E(S \cup W)E(S \cup W) = E(S) + E(W)E(S) + E(S)E(W) + E(W)$$

which implies

$$E(S) + E(W) = E(S) + E(W)E(S) + E(S)E(W) + E(W)$$

hence $E(W)E(S) + E(S)E(W) = 0$. □

Lemma 10.3. $E(S \cap W) = E(S) \circ E(W)$.

Proof. We have

$$E(S \cap W) + E(S - W) = E(S) \quad E(S \cap W) + E(W - S) = E(W)$$

and

$$E(S \cup W) = E(S \cap W) + E(S - W) + E(W - S)$$

From which it follows that

$$\begin{aligned} E(S) + E(W) &= 2E(S \cap W) + E(S - W) + E(W - S) \\ &= E(S \cap W) + E(S \cup W) \end{aligned}$$

Now multiply both sides of the equation on the right by $E(W)$ gives

$$E(S)E(W) + E(W)^2 = E(S \cap W)E(W) + E(S \cup W)E(W)$$

Since $S \subset S \cup W$, $E(S) \leq E(S \cup W)$, and

$$E(S \cup W)E(S) = E(S)$$

$$\text{Now } E(S \cup W)E(S) = E(S)^2 + E(W - S)E(S) = E(S) \quad \square$$

Example. *If $H = \mathbf{C}$, then $B(H) \cong \mathbf{C}$, and any $\{0, 1\}$ valued measure μ is a spectral measure, provided $\mu(\Omega) = 1$. If $\mu(S_1) = \mu(S_2) = 1$, then*

$$1 = \mu(S_1) = \mu(S_1 \cap S_2) + \mu(S_1 - S_2)$$

$$1 = \mu(S_2) = \mu(S_1 \cap S_2) + \mu(S_2 - S_1)$$

If $\mu(S_1 \cap S_2) = 0$, then $\mu(S_1 - S_2) = \mu(S_2 - S_1) = 1$, which implies

$$\mu((S_1 - S_2) \cup (S_2 - S_1)) = \mu(S_1 - S_2) + \mu(S_2 - S_1) = 2$$

an impossibility. Thus $\mu(S_1 \cap S_2) = 1$. If $\mu(S_1) = 0$, then $S_1 \cap S_2 \subset S_1$, so $\mu(S_1 \cap S_2) \leq \mu(S_1) = 0$. The same holds if $\mu(S_2) = 0$. Thus we have verified that $\mu(S_1 \cap S_2) = \mu(S_1) \circ \mu(S_2)$ on a case by case basis. The countable summation property holds by the property of the measure itself.

Example. If H is infinite dimensional, and $N \in K(H)$, then

$$\sigma(N) = \{\lambda_0, \lambda_1, \dots\}$$

where $\lambda_0 = 0$, and each λ_i is an eigenvalue. For each $n > 0$, let P_n be orthogonal projection onto $\ker(\lambda_n - N)$, and let P_0 project onto the orthogonal complement of the closed linear space of the images of P_1, P_2, \dots . For $S \subset \sigma(N)$, let

$$E(S) = \sum_{\lambda_n \in S} P_n$$

In the sense that the sum on the right is interpreted pointwise. Then E is a spectral measure on $(\sigma(N), \mathcal{P}(\sigma(N)))$. Surely

$$E(\sigma(N)) = \sum_{n=0}^{\infty} P_n = id_H$$

And since $P_i \circ P_j = 0$ if $i \neq j$,

$$\begin{aligned} E(S_1) \circ E(S_2) &= \left(\sum_{\lambda_n \in S_1} P_n \right) \circ \left(\sum_{\lambda_m \in S_2} P_m \right) \\ &= \sum_{\lambda_n \in S_1, \lambda_m \in S_2} P_n P_m \\ &= \sum_{\lambda_n \in S_1 \cap S_2} P_n = E(S_1 \cap S_2) \end{aligned}$$

And the countable additivity follows by the pointwise definition of the sum of operators.

Example. If Ω is a σ -finite measure space with measure μ , and $H = L^2(\Omega)$, define

$$E(S)\xi = \chi_S \xi$$

Then E is a spectral measure. Surely $E(\Omega) = id_H$, $E(\emptyset) = 0$, because $\chi_\Omega = 1$, $\chi_\emptyset = 0$. Now $E(S_1 \cap S_2) = E(S_1) \circ E(S_2)$ follows because $\chi_{S_1 \cap S_2} = \chi_{S_1} \chi_{S_2}$. Similarly, $\chi_{\bigcup S_i} = \sum \chi_{S_i}$ pointwise if the S_i are disjoint.