Functional Analysis

Richard Bass

These notes are based mostly on the book by P. Lax, *Functional Analysis*.

Functional analysis can best be characterized as infinite dimensional linear algebra. We will use some real analysis, complex analysis, and algebra, but functional analysis is not really an extension of any one of these.

1 Linear spaces

A linear space is the same as a vector space. We have an operation "+" under which the space is an Abelian group: addition is commutative, associative, there exists an identity (which we call 0) and every element has an inverse (the inverse of x will be denoted $-x$).

We define $x - y = x + (-y)$.

We also have a field F, which for us will always be the reals or the complex field. Elements of F will be called scalars.

A linear space also has an operation called scalar multiplication, which is associative ($a(bx) = (ab)x$) and distributive ($a(x + y) = ax + ay$ and $(a + b)x = ax + bx$), and we have $1x = x$, where 1 is the identity in F.

By the same proofs as in the finite dimensional case, we have $0x = 0$ because $0x = (0 + 0)x = 0x + 0x$ and $(-1)x = -x$ because

$$0 = 0x = (1)x + (-1)x = x + (-1)x.$$

Examples of linear spaces are

1. $\mathbb{R}^n$

2. The collection of all infinite sequences.

3. $B(S)$, the bounded functions on a set S.

4. $C(S)$, the continuous functions on S, where S is a topological space.

5. $C^k(\mathbb{R})$

6. $L^p(X, m)$

7. $\{f : f \text{ is analytic in } D\}$

If X is a linear space, $Y \subset X$, then Y is a linear subspace of X if $y \in Y$ implies $ay \in Y$ for all $a \in F$ and $x, y \in Y$ implies $x + y \in Y$.

Let S be a subset of X. Consider the collection

$$\{Y_\alpha : Y_\alpha \text{ is a linear subspace of } X, S \subset Y_\alpha\}.$$

It is easy to check that $\cap_\alpha Y_\alpha$ is a subspace of X, and it is called the linear span of S.

Proposition 1.1 *The linear span of S is equal to*

$$\Big\{\sum_{i=1}^n a_i x_i : a_i \in F, x_i \in S, n \in \mathbb{N}\Big\}.$$

Proof. Let M be the above set of sums. M is clearly a linear subspace of X containing S, therefore the span of S is contained in M. If Y_α is any linear subspace containing S, then Y_α must contain M, therefore $\cap_\alpha Y_\alpha$ contains M. $\square$

Let X and U be linear spaces over F. $M : X \to U$ is linear if $M(x+y) = M(x) + M(y)$ and $M(kx) = kM(x)$.

Examples of linear operators:

1) $X = L^1(m)$, g is bounded and measurable, and

$$Tf = \int f(x)g(x)\, m(dx).$$

2) Let our space be $B(S)$, fix a point $x_0 \in S$ or points $x_0, \ldots, x_n \in S$, and let $Tf = f(x_0)$ or $Tf = (f(x_0), \ldots, f(x_n))$. Here T maps $B(S)$ to $\mathbb{R}$ or $\mathbb{R}^{n+1}$.

3) Let X be the space of n-tuples, and define the ith coordinate of Mx to be $\sum_{j=1}^n a_{ij} x_j$. This is just matrix multiplication, and all linear maps in finite dimensions can be viewed in this way.

4) Let X be the set of bounded sequences, suppose that $\sup_i \sum_{j=1}^{\infty} |a_{ij}| < \infty$, and define the ith coordinate of Mx to be $\sum_{j=1}^{\infty} a_{ij} x_j$.

5) Let X be the set of bounded measurable functions on some measure space with finite measure m, and suppose $K(x, y)$ is jointly measurable and bounded. Define Mf by $Mf(x) = \int K(x, y)\, m(dy)$.

Two linear spaces are isomorphic if there exists a one-to-one linear mapping from one space onto the other.

If M is a 1-1 linear mapping from X onto U, then M^{-1} is also linear. To see this, suppose $u_1 = Mx_1$ and $u_2 = Mx_2$ are elements of U with $x_1, x_2 \in X$. Then $u_1 + u_2 = Mx_1 + Mx_2 = M(x_1 + x_2)$. Hence $M^{-1}(u_1 + u_2) = x_1 + x_2 = M^{-1}u_1 + M^{-1}u_2$. Similarly $M^{-1}(ku) = kM^{-1}u$.

A set $K \subset X$ is convex if $\lambda x + (1 - \lambda)y \in K$ whenever $x, y \in K$ and $\lambda \in [0, 1]$.

A convex combination of $x_1, \ldots, x_m$ is a sum of the form $\sum_{i=1}^{n} a_i x_i$, where $\sum_{i=1}^{n} a_i = 1$, n is a positive integer, and all the a_i are non-negative.

Lemma 1.2 *Linear subspaces are convex. Intersections of convex sets are convex. If $M : X \to U$ is linear and $K \subset X$ is convex, then $\{M(x) : x \in K\}$ is convex.*

If $S \subset X$, the convex hull of S is the intersection of all the convex sets containing S.

Proposition 1.3 *The convex hull of S is equal to the set of all convex combinations of points of x.*

If K is convex, and $E \subset K$, then E is an extreme subset of K if

(1) E is convex and non-empty

(2) if $x \in E$ and $x = \frac{y+z}{2}$ with $y, z \in K$, then $y, z \in E$.

If E is a single point, then the point is called an extreme point of K.

For an example, consider the case where K is a polygon (plus the interior) in $\mathbb{R}^2$. Each edge of K is an extreme subset. Each vertex is an extreme point.

2 Linear maps

2.1 Definitions

Definition 2.1 *If M and N are linear maps from X into U and k is a scalar, we define*

$$(M+N)(x) = M(x) + N(x), \qquad (kM)(x) = kM(x).$$

So the set of linear maps from X into U is a linear space, and we denote it $\mathcal{L}(X, U)$.

If $M : X \to U$ and $N : U \to W$, we define $(NM)(x) = N(M(x))$.

An exercise is to show this is associative but not necessarily commutative. (Multiplication by matrices is an example to show commutativity need not hold.) It is distributive:

$$M(N+K) = MN + MK, \qquad (M+K)N = MN + KN.$$

We usually write Mx for $M(x)$.

Define the identity $I : X \to X$ by $Ix = x$. We will also write I_X when we want to emphasize the space.

We say $M : X \to U$ is invertible if there exists $M^{-1} : U \to X$ such that $M^{-1}M = I_X$, $MM^{-1} = I_U$.

Definition 2.2 *The null space or kernel of M is $N_M = \{x \in X : Mx = 0\}$ and the range of M is $R_M = \{Mx : x \in X\}$.*

Observe that $N_M \subset X$ and $R_M \subset U$.

Some easily checked facts: N_M and R_M are linear subspaces, and if K, M are invertible, then $(KM)^{-1} = M^{-1}K^{-1}$.

Let Y be a subspace of X. Y is called an invariant subspace for $M : X \to X$ if M maps Y into Y.

One example of an invariant subspace is to let y be a fixed element of X and set

$$Y = \{p(M)y : p \text{ a polynomial}\}.$$

To check this, if $z \in Y$, then $z = p(M)y$ for some polynomial p. Then $Mz = Mp(M)y$, and $Mp(M) = \overline{p}(M)$, where $\overline{p}(x) = xp(x)$ is another polynomial.

A second example: suppose $T : X \to X$ and $MT = TM$. Then N_T is invariant. To see this, if $z \in N_T$, then $Tz = 0$ and then $T(Mz) = TMz = MTz = M0 = 0$, or $Mz \in N_T$.

If N and Y are subspaces of X, we write $X = N \oplus Y$ if for each $x \in X$, there exist $n \in N$ and $y \in Y$ such that $x = n + y$, and the decomposition is unique, i.e., there is only one n and one y that works for any particular x. Of course, n and y depend on x.

As an example, let $X = \mathbb{R}^3$, $N = \{(y, 0, 0\})$. There are lots of possibilities for Y, in fact, any plane in $\mathbb{R}^3$ that passes through the origin and does not contain the real axis. Given any choice of Y, though, there is only one way to write a given x as $n + y$.

We will frequently use Zorn's lemma, which is equivalent to the axiom of choice.

Suppose we have a partially ordered set S, which means that there is an order relation such that $a \leq a$ for all $a \in S$, and if $a \leq b$ and $b \leq c$, then $a \leq c$. A subset is totally ordered if for every pair x, y in the subset, either $x \leq y$ or $y \leq x$. An element u of a partially ordered set is an upper bound for a subset of S if $x \leq u$ for every x in the subset. An element x of a partially ordered set is maximal if $y \geq x$ implies $y = x$. (Lax stated this incorrectly.)

Lemma 2.3 *(Zorn's lemma) Let X be a partially ordered set. If every totally ordered subset of X has an upper bound in X, then X has a maximal element.*

Lemma 2.4 *Suppose N is a subspace of a linear space X. Then there exists a linear subspace Y such that $X = N \oplus Y$.*

Proof. Look at $\{Y : Y \text{ a subspace of } X, Y \cap N = \{0\}\}$. We partially order this collection by inclusion, If $\{Y_\alpha\}$ is a totally ordered subcollection, then $\cup_\alpha Y_\alpha$ is an upper bound. Let Y_0 be the maximal element guaranteed by Zorn's lemma.

Suppose there is a point $x \in X$ that is not in $N \oplus Y_0$. We adjoin x to Y_0 to form Y_1, that is, $Y_1 = \{ax + y : y \in Y_0, a \in \mathbb{R}\}$. Y_1 is a subspace of X that is strictly bigger than Y_0. We argue that $Y_1 \cap N = \{0\}$, a contradiction to the fact that Y_0 is maximal.

x is not in the direct sum of N and Y_0, so $x \notin N$, or else we could write $x = x + 0$. If $z \neq 0$ and $z \in Y_1 \cap N$, then there exist $a \in \mathbb{R}$ and $y \in Y_0$ such that $z = ax + y$. One possibility is that $a = 0$; but then $z = y \in Y_0 \cap N$, which isn't possible since z is nonzero. The other possibility is that $a \neq 0$. But $z \in N$, so

$$x = \frac{z}{a} + \frac{-y}{a} \in N \oplus Y_0,$$

also a contradiction. □

2.2 Index of a map

We say that $\dim X < \infty$, that is X is finite dimensional, if there exist finitely many points $x_1, \ldots, x_n \in X$ such that X is equal to the span of $\{x_1, \ldots, x_n\}$. The smallest such n is the dimension of X.

Let X be a linear space and Y a subspace. We say $x_1 \equiv x_2 (\text{mod } Y)$ or that x_1 is equivalent to x_2 if $x_1 - x_2 \in Y$. This is an equivalence relation. Let $\overline{x}$ denote the equivalence class containing x. The collection of all such equivalence classes is denoted X/Y and called the quotient space of X with respect to Y,

Let's make X/Y into a linear space. If $\overline{x}_1, \overline{x}_2$ are in X/Y, define $\overline{x}_1 + \overline{x}_2$ to be $\overline{x_1 + x_2}$. To see that this is well defined, if z_1, z_2 are elements of $\overline{x}_1, \overline{x}_2$, resp., then $(x_1 + x_2) - (z_1 + z_2) = (x_1 - z_1) + (x_2 - z_2)$, the sum of two elements of Y, hence an element of Y. We similarly define $k\overline{x} = \overline{kx}$. It is routine to verify that X/Y is now a linear space.

We define the codimension of Y by

$$\text{codim } Y = \dim X/Y.$$

Let's look at an example. Suppose $X = \mathbb{R}^5$ and $Y = \{(x, y, 0, 0, 0)\}$. $x_1 \equiv x_2$ if and only if the 3^{rd} through 5^{th} coordinates of x_1 and x_2 agree. Therefore X/Y is (essentially - at least it is isomorphic to) the 3^{rd} through 5^{th} coordinates of points in $\mathbb{R}^5$, hence isomorphic to $\mathbb{R}^3$. We see codim $Y = \dim X/Y = 3$, while $\dim Y = 2$.

Lemma 2.5 *If $X = N \oplus Y$, then X/N is isomorphic to Y.*

Proof. If $\overline{x} \in X/N$, then we can write $x = y + n$. Define $M\overline{x} = y$. We will show that M is an isomorphism.

First we need to show M is well defined. If x' is another element of $\overline{x}$, we can write $x' = y' + n'$. Then $x - x' = (y - y') + (n - n')$ is in N. Since we can write $x - x' = 0 + (x - x')$, we must have $y - y' = 0$, or $y = y'$.

Next we show M is linear. If $x_1 + x_2 \in \overline{x_1 + x_2}$, then $x_1 = y_1 + n_1$, $x_2 = y_2 + n_2$, and then $x_1 + x_2 = (y_1 + y_2) + (n_1 + n_2)$. So $M(\overline{x_1 + x_2}) = y_1 + y_2 = M\overline{x}_1 + M\overline{x}_2$. The linearity with respect to scalar multiplication is similar.

We show M is 1-1. If $M\overline{x} = M\overline{x}'$, and we write $x = y + n$, $x' = y' + n'$, then $y = M\overline{x} = M\overline{x}' = y'$. Hence

$$x - x' = (y - y') + (n - n') = n - n' \in N,$$

so $\overline{x} = \overline{x}'$.

Finally, M is onto, because if $y \in Y$, then $y = y + 0 \in X$ and $M\overline{y} = y$. □

Let $M : X \to U$. We will need the fact that

Proposition 2.6 *X/N_M is isomorphic to R_M.*

Proof. If $\overline{x} \in X/N_M$, we define $\widetilde{M}\overline{x}$ to be Mx for any $x \in \overline{x}$. If x' is any other element of $\overline{x}$, then $x - x' \in N_M$, or $M(x - x') = 0$, or $Mx = Mx'$. So the map $\widetilde{M}$ is well defined. It is routine to check that $\widetilde{M}$ is linear.

To show $\widetilde{M}$ is 1-1, if $\widetilde{M}\overline{x} = \widetilde{M}\overline{y}$, then $Mx = My$, or $M(x - y) = 0$, or $x - y \in N_M$, so $\overline{x} = \overline{y}$. To show $\widetilde{M}$ is onto, if $y \in R_M$, then $y = Mx$ for some $x \in X$. Then $\widetilde{M}\overline{x} = Mx = y$. □

We say a linear map G is degenerate if $\dim R_G < \infty$. Maps $M : X \to U$ and $L : U \to X$ are pseudo-inverses if there exist $G : X \to X$ and $H : U \to U$ that are degenerate and

$$LM = I_X + G, \qquad ML = I_U + H.$$

This concept is not interesting in finite dimensions. For example, let $X = U = \mathbb{R}^5$, let A be any 5×5 matrix, and define $Mx = Ax$, where we view x as a 5×1 matrix. Let B be any other 5×5 matrix and define $Lx = Bx$. Then $LMx = BAx = (I + G)x$, where $G = BA - I$. Obviously R_G is finite dimensional. We deal with ML similarly. So M and L are pseudo-inverses.

For a non-trivial example of pseudo-inverse, define the right and left shifts on X, the linear space of infinite sequences, by

$$R(a_1, a_2, \ldots) = (0, a_1, a_2, \ldots),$$

$$L(a_1, a_2, \ldots) = (a_2, a_3, \ldots).$$

Then

$$RL(a_1, \ldots) = (0, a_2, a_3, \ldots) = (I - G)(a_1, a_2, \ldots),$$

where

$$G(a_1, a_2, \ldots) = -(a_1, 0, 0, \ldots).$$

Clearly R_G is finite dimensional. We see that LR is the identity, so L and R are pseudo-inverses.

We will prove that $M : X \to U$ has a pseudo-inverse if and only if $\dim N_M < \infty$ and codim $R_M < \infty$.

We start by proving

Proposition 2.7 *If M has a pseudo-inverse, then*

$$\dim N_M < \infty \qquad \textit{and} \qquad \text{codim } R_M < \infty.$$

Proof. First suppose G is degenerate. We show $\dim N_{I+G} < \infty$. To see this, if $x \in N_{I+G}$, $x + Gx = 0$, so $x = -Gx \in R_G$. Therefore $N_{I+G} \subset R_G$, which implies $\dim N_{I+G} \leq \dim R_G < \infty$.

Next, $LM = I+G$, so if $x \in N_M$, then $Mx = 0$, so $LMx = 0$, or $(I+G)x = 0$, or $x \in N_{I+G}$. Therefore $N_M \subset N_{I+G}$, hence $\dim N_M \leq \dim N_{I+G} < \infty$.

Recall the X/N_G is isomorphic to R_G. Therefore codim $N_G = \dim R_G$.

We claim codim $R_{I+G} < \infty$. To prove the claim, if $x \in N_G$, then $(I + G)x = x + Gx = x$, so $N_G \subset R_{I+G}$. Then codim $R_{I+G} \leq$ codim $N_G = \dim R_G < \infty$.

Finally, since $ML = I + G$, then $R_{I+G} \subset R_M$. To see this, if $y \in R_{I+G}$, then $y = (I+G)x$, or $y = M(Lx)$. We then write

$$\text{codim } R_M \leq \text{codim } R_{I+G} < \infty.$$

□

We now prove

Proposition 2.8 *If* $\dim N_M < \infty$ *and* $\text{codim } R_M < \infty$*, then* M *has a pseudo-inverse.*

Proof. Suppose $M : X \to U$ and write $X = N_M \oplus Y$, $U = R_M \oplus V$. We claim that M restricted to Y is a 1-1 map onto R_M. First we show onto. If $z \in R_M$, then $z = Mx$ for some $x \in X$. For some $n \in N$ and $y \in Y$, $x = n + y$. So $z = My$.

Next we show 1-1. If $My_1 = My_2$, then $M(y_1 - y_2) = 0$, or $y_1 - y_2 \in N$. Let $n = y_1 - y_2$. Then $y_2 + n = y_1 + 0$. Since $X = N_M \oplus Y$, the decomposition is unique, and $n = 0$ and $y_1 = y_2$.

Therefore $M : Y \to R_M$ is invertible. Define

$$K = \begin{cases} M^{-1} & \text{on } R_M \\ 0 & \text{on } V. \end{cases}$$

Extend K to $U = R_M \oplus V$ by linearity as follows. If $u = z + v$ with $z \in R_M$ and $v \in V$, let $Ku = M^{-1}z$.

If $y \in Y$, then $My \in R_M$ and so $KMy = y$. If $n \in N_M$, then $KMn = 0$. So

$$KM = \begin{cases} 1 & \text{on } Y \\ 0 & \text{on } N_M. \end{cases}$$

If $z \in R_M$, then $z = My$ for some $y \in Y$, so $MKz = MKMy = My = z$. If $z \in V$, $Kz = 0$. So

$$MK = \begin{cases} 1 & \text{on } R_M \\ 0 & \text{on } V. \end{cases}$$

Let P be the projection onto N_M: if $x \in X$ and $x = y + n$, define $Px = n$. Then $KM = I - P$. Since $\dim R_P = \dim N_M < \infty$, P is degenerate.

Similarly $MK = I - Q$, where Q is the projection onto V. Since V is isomorphic to U/R_M by a previous lemma, $\dim V = \text{codim } R_M < \infty$, and Q is degenerate. □

A sequence of spaces and maps

$$V_0 \xrightarrow{T_0} V_1 \xrightarrow{T_1} \cdots V_{n-1} \xrightarrow{T_{n-1}} V_n$$

is an exact sequence if $R_{T_j} = N_{T_{j+1}}$.

Lemma 2.9 *Suppose we have an exact sequence, each V_j is finite dimensional, and $\dim V_0 = 0 = \dim V_n$. Then*

$$\sum_{j=0}^{n} (-1)^j \dim V_j = 0.$$

Note since $\dim V_0 = \dim V_n$, we can write the sum from $j = 0$ to n, or omit or include $j = 0$ or $j = n$ as we choose.

Proof. Write N_j for N_{T_j}, $R_j = R_{T_j}$, and write $V_j = N_j \oplus Y_j$. Because we have an exact sequence, $R_{j-1} = N_j$.

Now V_j/N_j is isomorphic to $R_j = N_{j+1}$. On the other hand Y_j is isomorphic to V_j/N_j by a lemma. Hence Y_j is isomorphic to N_{j+1}. Hence

$$\dim V_j = \dim N_j + \dim N_{j+1}.$$

Note $N_0 \subset V_0$, so $\dim N_0 = 0$. Also, because $R_{n-1} \subset V_n$, $R_{n-1} = \{0\}$. Since V_{n-1}/N_{n-1} is isomorphic to R_{n-1}, $\dim V_{n-1} = \dim N_{n-1}$.

Then

$$\begin{aligned}
\sum (-1)^j \dim V_j &= \sum (-1)^j [\dim N_j + \dim N_{j+1}] \\
&= \dim N_0 + \dim N_1 - \dim N_1 - \dim N_2 + \cdots \\
&\qquad \pm (\dim N_{n-2} + \dim N_{n-1} - \dim N_{n-1}) \\
&= \dim N_0 \\
&= 0.
\end{aligned}$$

□

Suppose M has a pseudo-inverse. Define

$$\text{ind}\,(M) = \dim N_M - \text{codim}\ R_M.$$

Theorem 2.10 *If $M : X \to U$ and $L : U \to W$ both have pseudo-inverses, then LM has a pseudo-inverse and*

$$\text{ind}\,(LM) = \text{ind}\,(L) + \text{ind}\,(M).$$

Proof. We use an exact sequence. $V_0 = 0$, $V_1 = N_M$, $V_2 = N_{LM}$, $V_3 = N_L$, $V_4 = U/R_M$, $V_5 = W/R_{LM}$, $V_6 = W/R_L$, and $V_7 = 0$.

Since $N_M \subset N_{LM}$, we let T_1 be the inclusion map. $T_2 = M$. T_3 is the map taking U to U/R_M. T_4 is the map taking U/R_M to W/R_{LM}, while T_5 is the map from W/R_{LM} to W/R_L. It is an exercise to check that this is an exact sequence. Using the lemma,

$$-\dim N_M + \dim N_{LM} - \dim N_L + \text{codim}\ R_M - \text{codim}\ R_{LM} + \text{codim}\ R_L = 0.$$

This does it. □

3 Hahn-Banach theorem

3.1 The theorem

A linear functional ℓ is a linear map from X to F. Let's take F to be the reals for now.

We will be working with a sublinear functional $p(x)$ in the statement of the theorem. An example to keep in mind is $p(x) = \|x\|$, although we have not defined what a norm is yet.

Theorem 3.1 *Suppose $p : X \to \mathbb{R}$ satisfies $p(ax) = ap(x)$ if $a > 0$ and $p(x + y) \le p(x) + p(y)$ if $x, y \in X$. Suppose Y is a linear subspace, ℓ is a linear functional on Y, and $\ell(y) \le p(y)$ for all $y \in Y$. Then ℓ can be extended to a linear functional on X satisfying $\ell(x) \le p(x)$ for all $x \in X$.*

Proof. If Y is not all of X, pick $z \in X \setminus Y$. Look at $Y_1 = \{y + az : y \in Y, a \in \mathbb{R}\}$. We want to define $\ell(z)$ to be some real number with the property that if we set

$$\ell(y + az) = \ell(y) + a\ell(z),$$

we would have $\ell(y) + a\ell(z) \le p(y + az)$ for all $y \in Y$ and $a \in \mathbb{R}$. This would give us an extension of ℓ from Y to Y_1.

For all $y, y' \in Y$,

$$\begin{aligned} \ell(y') + \ell(y) = \ell(y' + y) \le p(y' + y) &= p((y + z) + (y' - z)) \\ &\le p(y + z) + p(y' - z). \end{aligned}$$

So

$$\ell(y') - p(y' - z) \le p(y + z) - \ell(y).$$

This is true for all $y, y' \in Y$. So choose $\ell(z)$ to be a number between $\sup_{y'}[\ell(y') - p(y' - z)]$ and $\inf_y[p(y + z) - \ell(y)]$. Therefore

$$\ell(y') - p(y' - z) \le \ell(z) \le p(y + z) - \ell(y),$$

or

$$\ell(y) + \ell(z) \le p(y + z), \qquad \ell(y') - \ell(z) \le p(y' - z).$$

If $a > 0$,

$$\ell(y + az) = a\ell(\tfrac{y}{a} + z) \le ap(\tfrac{y}{a} + z) = p(y + az).$$

Similarly $\ell(y' - az) \le p(y' - az)$ if $a > 0$.

So we have extended ℓ from Y to Y_1, a larger space. Let $\{(Y_\alpha, \ell_\alpha)\}$ be the collection of all extensions of (Y, ℓ). This is partially ordered by inclusion. If $\{(Y_\beta, \ell_\beta)\}$ is a totally ordered subset, define ℓ on $\cup_\beta Y_\beta$ by setting $\ell(z) = \ell_\beta(z)$ if $z \in Y_\beta$. By Zorn's lemma, there is a maximal extension. This maximal extension must be all of X, or else by the above we could extend it. □

The Hahn-Banach theorem one learns in a real analysis course has the same proof, but one takes $p(x) = c|x|$. Saying $\ell(x) \le p(x)$ then translates to saying $|\ell| \le c$, and we extend ℓ so that the norm of the extension is the same.

3.2 Separating hyperplanes

If ℓ is a linear functional, $\{x : \ell(x) = c\}$ is a hyperplane. This splits X into two parts, those x for which $\ell(x) > c$ and those for which $\ell(x) < c$.

A point $x_0 \in S \subset X$ is interior to S if for all $y \in X$, there exists ε (depending on y) such that $x_0 + ty \in S$ if $-\varepsilon < t < \varepsilon$.

Let K be a convex set with an interior point. Without loss of generality, we may assume $x_0 = 0$. p_K, the gauge, is defined by

$$p_K(x) = \inf\left\{a > 0 : \frac{x}{a} \in K\right\}.$$

Proposition 3.2 *p_K is homogeneous and subadditive.*

Saying p_K is homogeneous means that $p_K(ax) = ap_K(x)$ if $a > 0$. Subadditive means $p_K(x + y) \le p_K(x) + p_K(y)$.

Proof. Homogeneity is obvious. We look at subadditivity. Let $x, y \in X$. If $p_K(x)$ or $p_K(y)$ is infinite, there is nothing to prove. So suppose both are finite and let $\varepsilon > 0$. Choose $p_K(x) < a < p_K(x) + \varepsilon$ and $p_K(y) < b < p_K(y) + \varepsilon$. Then $\frac{x}{a}$ and $\frac{y}{b}$ are in K. Letting $\lambda = a/(a + b)$,

$$\lambda\frac{x}{a} + (1 - \lambda)\frac{y}{b} = \frac{x + y}{a + b}$$

is in K. So

$$p_K(x + y) \le a + b \le p_K(x) + p_K(y) + 2\varepsilon.$$

Since ε is arbitrary, we are done. □

The following proposition's proof is left to the reader.

Proposition 3.3 *(a) If K is convex and $x \in K$, then $p_K(x) \le 1$. If K is convex and x is interior to K, then $p_K(x) < 1$.*

(b) Let p be positive, homogeneous, and subadditive. Then $\{x : p(x) < 1\}$ is convex and 0 is an interior point. Also $\{x : p(x) \le 1\}$ is convex.

We now prove the hyperplane separation theorem.

Theorem 3.4 *Suppose K is a nonempty convex subset of X and all points of K are interior. If $y \notin K$, then there exist ℓ and c such that $\ell(x) < c$ for all $x \in K$ and $\ell(y) = c$.*

Proof. Without loss of generality, assume $0 \in K$. Note $p_K(x) < 1$ for all $x \in K$. Set $\ell(y) = 1$ and $\ell(ay) = a$. If $a \le 0$, $\ell(ay) \le 0 \le p_K(ay)$. If $a > 0$, then since $y \notin K$, $p_K(y) \ge 1$, and so $p_K(ay) \ge a = \ell(ay)$.

We let $Y = \{ay\}$ and use Hahn-Banach to extend ℓ to all of X. We have $\ell(x) \le p_K(x) < 1$ if $x \in K$ and $l(y) = 1$. We take $c = 1$. □

Corollary 3.5 *If K is convex with at least one interior point and $y \notin K$, there exists $\ell \neq 0$ such that $\ell(x) \le \ell(y)$ for all $x \in K$.*

$A + B$ is defined to be $\{a + b : a \in A, b \in B\}$.

Corollary 3.6 *Let H and M be disjoint convex sets, with at least one having an interior point. Then there exist ℓ and c such that*

$$\ell(u) \le c \le \ell(v), \qquad u \in H, v \in M.$$

Proof. $-M$ is convex, so $K = H + (-M)$ is convex. K must have an interior point. $H \cap M = \emptyset$, so $0 \notin K$. Let $y = 0$. There exists ℓ such that

$$\ell(x) \le \ell(0) = 0 \qquad x \in K.$$

If $u, v \in H, M$, resp., then $x = u - v \in K$, so $\ell(x) \le 0$, and hence $\ell(u) \le \ell(v)$. □

3.3 Complex linear functionals

Theorem 3.7 *Let X be a linear space over $\mathbb{C}$. Suppose $p \ge 0$ satisfies $p(ax) = |a|p(x)$ for all $x \in X, a \in \mathbb{C}$, and $p(x + y) \le p(x) + p(y)$. If Y is a subspace of X, ℓ is a linear functional on Y, and $|\ell(y)| \le p(y)$ for all $y \in Y$, then ℓ can be extended to a linear functional on X with $|\ell(x)| \le p(x)$ for all x.*

Again, we think of $p(x)$ as $\|x\|$, once that is defined.

Proof. Write ℓ as $\ell(y) = \ell_1(y) + i\ell_2(y)$, the real and imaginary parts of ℓ. Since ℓ is linear,

$$i\ell(y) = \ell_1(iy) + i\ell_2(iy).$$

On the other hand

$$i\ell(y) = i\ell_1(y) - \ell_2(y)$$

by substituting in for $\ell(y)$ and multiplying by i. Equating the real parts, $\ell_1(iy) = -\ell_2(y)$.

One can work this in reverse to see that if ℓ_1 is a linear functional over the reals, and we define $\ell(x) = \ell_1(x) - i\ell_1(ix)$, we get a linear functional over the complexes.

To extend ℓ, we have

$$\ell_1(y) \le |\ell(y)| \le p(y).$$

Use Hahn-Banach to extend ℓ_1 to all of X and set $\ell(x) = \ell_1(x) - i\ell_1(ix)$.

We need to show that $|\ell(x)| \le p(x)$ for all x. Fix x and write $\ell(x) = ar$, where r is real and $|a| = 1$. Then

$$|\ell(x)| = r = a^{-1}\ell(x) = \ell(a^{-1}x).$$

Since $\ell(a^{-1}x) = |\ell(x)|$, it is real with no imaginary part, and therefore equals

$$\ell_1(a^{-1}x) \le p(a^{-1}x) = |a^{-1}|p(x) = p(x).$$

□

4 An application of the Hahn-Banach theorem

Let S be an arbitrary set and let B be the collection of real-valued bounded functions on S. We say $x \le y$ if $x(s) \le y(s)$ for all $s \in S$. (We'll use $x \ge y$ if $y \le x$.) A function x is non-negative if $0 \le x$. Let Y be a linear subspace of B. ℓ is a positive linear functional on Y if $\ell(y) \geq 0$ whenever $y \geq 0$. Note that if $x \leq y$, then $0 \leq \ell(y-x) = \ell(y) - \ell(x)$, so ℓ is monotone.

One example is to take $\ell(y) = y(s_0)$ for some point s_0 in S. Or we could take a linear combination $\sum c_i y(s_i)$ provided all the $c_i \geq 0$. Another example is if S is a measure space and we let $\ell(y) = \int y(s)\, m(dx)$.

Proposition 4.1 *Let Y be a linear subspace and suppose there exists $y_0 \in Y$ such that $y_0(s) \geq 1$ for all s. Let ℓ be a positive linear functional on Y. Then ℓ can be extended to a positive linear functional on B.*

Proof. Define
$$p(x) = \inf\{\ell(y) : y \in Y, y \geq x\}.$$
Since $-cy_0 \leq x \leq cy_0$ if x is bounded by c, we are not taking the infinum of an empty set. Since $x \leq cy_0$, then $p(x) \leq c\ell(y_0) < \infty$. If $y \geq x$ and $y \in Y$, then $-cy_0 \leq x \leq y$, so $\ell(y) \geq \ell(-cy_0) = -c\ell(y_0) > -\infty$.

It is clear that p is homogeneous. To show that p is subadditive, suppose $x_1, x_2 \in B$ and $y_1, y_2 \in Y$ with $x_1 \leq y_1, x_2 \leq y_2$. Then

$$\begin{aligned} p(x_1+x_2) &= \inf_{x_1+x_2 \leq y} \ell(y) \leq \inf_{x_1 \leq y_1, x_2 \leq y_2} \ell(y_1+y_2) \\ &= \inf_{x_1 \leq y_1} \ell(y_1) + \inf_{x_2 \leq y_2} \ell(y_2) = p(x_1) + p(x_2). \end{aligned}$$

If $y \in Y$ and $y' \geq y$ is any other element in Y, then $\ell(y) \leq \ell(y')$, so $p(y) \geq \ell(y)$. If $x \leq 0$, then since $0 \in Y$, $p(x) \leq \ell(0) = 0$.

We now use Hahn-Banach to extend ℓ to all of B. If $x \leq 0$, then $\ell(x) \leq p(x) \leq \ell(0) = 0$, which proves that ℓ is positive on B. □

5 Normed linear spaces

A norm is a map from $X \to \mathbb{R}$, denoted $|x|$, such that $|0| = 0$, $|x| > 0$ if $x \neq 0$, $|x+y| \leq |x| + |y|$, and $|ax| = |a|\,|x|$. A linear space together with its norm is called a normed linear space.

If we define $d(x,y) = |x-y|$, then d is a metric, and we can use all the terminology of topology.

Two norms $|x|_1$ and $|x|_2$ are equivalent if there exists a constant c such that

$$c|x|_1 \le |x|_2 \le c^{-1}|x|_1, \qquad x \in X.$$

Equivalent norms give rise to the same topology.

A subspace of a normed linear space is again a normed linear space. If Z and U are normed linear spaces, we can make $Z \oplus U$ into a normed linear space by defining

$$|(z,u)| = |z| + |u|,$$

For many purposes it is important to know whether a subspace is closed or not, closed being in the topological sense. Here is an example of a subspace that is not closed. Let $X = \ell^2$, the set of all sequences $\{x = (x_1, x_2, \ldots)\}$ with $|x| = (\sum_{j=1}^{\infty} |x_j|^2)^{1/2} < \infty$. Let M be the collection of points in X such that all but finitely many coordinates are zero. Clearly M is a linear subspace. Let $y_1 = (1, 0, \ldots)$, $y_2 = (1, \frac{1}{2}, 0, \ldots)$, $y_3 = (1, \frac{1}{2}, \frac{1}{4}, 0, \ldots)$ and so on. Each $y_k \in M$. But it is easy to see that $|y_k - y| \to 0$ as $k \to \infty$, where $y = (1, \frac{1}{2}, \frac{1}{4}, \frac{1}{8}, \ldots)$ and $y \notin M$. Thus M is not closed.

Proposition 5.1 *If X is a normed linear space and Y is a closed subspace, then X/Y is a normed linear space with $|\overline{x}| = \inf_{y \in \overline{x}} |y|$.*

Proof. Homogeneity is easy. To prove subadditivity, let $\varepsilon > 0$. Given $\overline{x}, \overline{z}$, there exist x, z such that $|x| < |\overline{x}| + \varepsilon$, $|z| < |\overline{z}| + \varepsilon$. then

$$|\overline{x} + \overline{z}| \le |x + z| \le |x| + |z| \le |\overline{x}| + |\overline{z}| + 2\varepsilon.$$

But ε is arbitrary.

It remains to prove positivity. Suppose $|\overline{x}| = 0$. So there exists a sequence $x_n \in \overline{x}$ such that $|x_n| \to 0$. Since each $x_n \in \overline{x}$, then $x_n - x_1 \in Y$. Let $y_n = x_1 - x_n$. So $\lim |x_1 - y_n| = \lim |x_n| = 0$. Therefore $y_n \to x_1$. Since Y is closed, $x_1 \in Y$, which implies $\overline{x} = \overline{x}_1 = 0$. □

If X is a normed linear space and Y is a subspace of X, then the closure of Y is also a linear subspace of X.

A Banach space is a complete normed linear space.

Recall that any metric space can be embedded in a complete metric space.

Examples:

1) ℓ^∞ is the collection of infinite sequences $\{a_1, a_2, \ldots\}$ with each $a_i \in \mathbb{C}$ and $\sup_i |a_i| < \infty$. We define $|x|_\infty = \sup_j |a_j|$. This is a complete space.

2) If $1 \le p < \infty$, ℓ^p is the collection of infinite sequences for which

$$|x|_p = \Big(\sum_j |a_j|^p \Big)^{1/p}$$

is finite. This is a complete space.

3) If S is a set, the collection of bounded functions on S with $|f|_\infty = \sup_s |f(s)|$ is a complete normed linear space.

4) If S is a topological space, then the collection of continuous bounded functions with $|f| = \sup_s |f(s)|$ is a complete normed linear space.

5) The L^p spaces are Banach spaces.

6) (Sobolev spaces) Let D be a domain in $\mathbb{R}^n$ and consider the C^∞ functions on D with

$$\int_D |\partial^\alpha f(x)|^p \, dx$$

finite for all $|\alpha| \le k$. Here $\partial^\alpha = \frac{\partial^{\alpha_1}}{\partial x_1^{\alpha_1}} \cdots \frac{\partial^{\alpha_n}}{\partial x_n^{\alpha_n}}$ and $|\alpha| = \alpha_1 + \cdots + \alpha_n$. For a norm, we take

$$|f|_{k,p} = \Big(\sum_{|\alpha| \le k} \int |\partial^\alpha f(x)|^p \, dx \Big)^{1/p}.$$

This is not a complete space, but its completion is denoted $W^{k,p}$ and is called a Sobolev space.

A normed linear space is separable if it contains a countable dense subset. Most of the examples are separable, but $B(S)$ is not if S is uncountable. Another example of one that is not is the collection of finite signed measures on $[0,1]$, with $|m| = \int_0^1 |dm|$. If we let δ_y be point mass at y, then $|\delta_y - \delta_z| = 2$ unless $y = z$.

5.1 The unit ball is not compact in infinite dimensions

In finite dimensions, the closed unit ball is always compact, but this is not the case in infinite dimensions. As an example, consider ℓ^2. If e_i is the sequence which has a one in the ith place and 0 everywhere else, then $|e_i - e_j| = \sqrt{2}$ if $i \neq j$. But then $\{e_i\}$ is a sequence contained in the unit ball that has no convergent subsequence, hence the unit ball is not compact.

In fact, the closed unit ball $B = \{x : |x| \le 1\}$ is never compact in infinite dimensions.

Theorem 5.2 *Let X be an infinite dimensional normed linear space. Then the closed unit ball is not compact.*

Proof. Choose y_1 such that $|y_1| = 1$. Given $y_1, \dots, y_{n-1}$, let Y_n be the linear span. Since Y_n is finite dimensional, it is closed. We will show in a moment how to find y_n such that $|y_n| = 1$ and $\inf_{y \in Y_n} |y - y_n| \ge 1/2$. We continue by induction and find a sequence $\{y_n\}$ contained in the closed unit ball such that $|y_j - y_n| \ge 1/2$ if $j < n$, hence which has no convergent subsequence.

Since Y_n is finite dimensional, it is not all of X and there exists $x \in X \setminus Y_n$. Since Y_n is closed, $d = \inf_{y \in Y_n} |y - x| > 0$. Choose $y' \in Y_n$ such that $|x - y'| < 2d$. Let $z' = x - y'$ so $|z'| < 2d$. If $y \in Y_n$, then $y + y' \in Y_n$ and

$$|z' - y| = |x - (y + y')| \ge d.$$

If we let $y_n = z'/|z'|$, then for all $y \in Y_n$

$$|y_n - y| = \left| \frac{z'}{|z'|} - y \right| = \frac{1}{|z'|} \big| z' - |z'| y \big| \ge \frac{d}{2d} = \frac{1}{2}$$

since $|z'| y \in Y_n$. □

5.2 Isometries

A (not necessarily linear) map M from X onto X is an isometry if

$$|Mx - My| = |x - y|$$

for all $x, y \in X$.

As an example, $Mx = x + u$, where u is a fixed element of X is an isometry.

The collection of isometries forms a group.

Theorem 5.3 *Let X, X' be two normed linear spaces over the reals, M an isometric map from X onto X' such that $M(0) = 0$. Then M is linear.*

Proof. Fix x, y and let $z = \frac{x+y}{2}$. Then

$$|x - z| = |y - z| = \frac{|x - y|}{2}.$$

Let

$$A = \Big\{ u : |x - u| = |y - u| = \frac{|x - y|}{2} \Big\}.$$

We show A is symmetric with respect to z. That is, if $u \in A$, we must show $v = 2z - u \in A$. To see this, $2z = x + y$, so $v - x = y - u$, $v - y = x - u$, so $|v - x| = |y - u| = \frac{|x-y|}{2}$, and similarly for $|v - y|$.

The diameter of A, d_A, is defined by $d_A = \sup_{u,w \in A} |u - w|$. Since A is symmetric with respect to z,

$$d_A \geq |u - v| = |u - (2z - u)| = |2u - 2z| = 2|u - z|,$$

so $|u - z| \leq d_A/2$.

Let

$$A_1 = \Big\{ p \in A : |u - p| \leq \tfrac{1}{2} d_A \text{ for all } u \in A \Big\}.$$

Note $z \in A_1$. A_1 is symmetric with respect to z, because if $p \in A_1$ and $q = 2z - p$, then $q - u = 2z - u - p = v - p$, so $|q - u| = |v - p| \leq \frac{1}{2} d_A$.

We have $d_{A_1} \leq \frac{1}{2} d_A$.

We define

$$A_2 = \Big\{ p \in A_1 : |u - p| \leq \tfrac{1}{2} d_{A_1} \text{ for all } u \in A_1 \Big\},$$

and so on. z is in each of these, the A_n decrease, and since $d_{A_n} \to 0$, then the intersection is the single point z.

If we let x', y' be the images of x, y under M, we define $A', A'_1, \ldots$ analogously. Since the sets are defined only in terms of distances, M maps A_n into A'_n. M^{-1} is also an isometry, and so M^{-1} maps A'_n into A_n. Therefore M maps $\cap_n A_n$ onto $\cap_n A'_n$, or

$$M\Big(\frac{x+y}{2}\Big) = \frac{x'+y'}{2}.$$

When $y = 0$, $y' = M(0) = 0$, and this becomes $M(x/2) = x'/2$, and similarly $M(y/2) = y'/2$. So

$$M\Big(\frac{x+y}{2}\Big) = \frac{x'+y'}{2}.$$

Replacing x by $2x$ and y by $2y$, we get the first property of linearity for M.

Since $M(2x) = M(x+x) = M(x) + M(x) = 2M(x)$ and $M(3x) = M(x+2x) = M(x) + M(2x) = M(x) + 2M(x) = 3M(x)$ and so on, then $M(kx) = kM(x)$. Since

$$M(x) = M(k(x/k)) = kM(x/k),$$

we have $M(x/k) = (1/k)M(x)$. Therefore $M(rx) = rM(x)$ for all r rational. Since M is an isometry, it is continuous, and therefore $M(rx) = rM(x)$ for all real r. □

To give some insight into the above proof, let $X = \ell^\infty$, the set of bounded sequences, let $x = (1, 0, \ldots)$, and $y = (0, 1, 0, 0, \ldots)$. Then

$$A = \{(\tfrac{1}{2}, \tfrac{1}{2}, a_3, a_4, \ldots) : |a_i| \le \tfrac{1}{2} \text{ for } i \ge 3\}.$$

Then

$$A_1 = \{(\tfrac{1}{2}, \tfrac{1}{2}, a_3, a_4, \ldots) : |a_i| \le \tfrac{1}{4} \text{ for } i \ge 3\},$$

and so on. As we define A_i, we close in on $(1/2, 1/2, 0, \ldots)$.

6 Hilbert space

6.1 Scalar product

We first look at the case $F = \mathbb{R}$. We have a scalar product (a, b) mapping $X \times X$ into R which is bilinear: $(x_1 + cx_2, y) = (x_1, y) + c(x_2, y)$, symmetric: $(x, y) = (y, x)$, and positive: $(x, x) > 0$ if $x \neq 0$.

When $F = \mathbb{C}$, this changes to $(ax, y) = a(x, y)$, $(x_1 + x_2, y) = (x_1, y) + (x_2, y)$, and $(y, x) = \overline{(x, y)}$, and still positivity.

Note $(x, ay) = \overline{(ay, x)} = \overline{a}\overline{(y, x)} = \overline{a}(x, y)$.

Define $\|x\| = (x, x)^{1/2}$.

We have the Cauchy-Schwarz inequality:

Proposition 6.1 $|(x, y)| \le \|x\| \, \|y\|$, *and equality holds if* $y = 0$ *or* $x = ay$.

Proof. Let t be real, $y \neq 0$.We have

$$0 \le \|x + ty\|^2 = \|x\|^2 + 2t\mathrm{Re}\,(x, y) + t^2\|y\|^2. \tag{6.1}$$

(Equality holds only if $x = ty$.) Set $t = -\mathrm{Re}\,(x, y)/\|y\|^2$. Multiplying by $\|y\|^2$, we have

$$(\mathrm{Re}\,(x, y))^2 \le \|x\|^2\|y\|^2.$$

If $\mathrm{Re}\,(x, y) = re^{i\theta}$, let $a = e^{-i\theta}$ so that $|a| = 1$ and $a(x, y)$ is real. Replace x by ax in the above. □

A corollary is that $\|x\| = \max_{\|y\| \le 1} |(x, y)|$. To see this, by Cauchy-Schwarz, the right hand side is less than or equal to the left hand side. For the other direction, take $y = x/\|x\|$.

Corollary 6.2 $\| \cdot \|$ *is a norm.*

Proof. It is only subadditivity that is not trivial, Start with

$$\|x + ty\|^2 = \|x\|^2 + 2t\mathrm{Re}\,(x, y) + t^2\|y\|^2,$$

take $t = 1$, and use Cauchy-Schwarz. □

If in (6.1) we take first $t = 1$ and then $t = -1$ and add, we get the parallelogram identity:

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

We say x and y are orthogonal if $(x, y) = 0$.

A linear space with a scalar product that is complete with respect to $\|\cdot\|$ is a Hilbert space.

One example is ℓ^2 with $(x, y) = \sum_j a_j \overline{b_j}$. Another example is L^2 with $(x, y) = \int x(t)\overline{y}(t)\, m(dt)$.

6.2 Convex subsets

Theorem 6.3 *Let K be a convex, nonempty, and closed subset of a Hilbert space H and let $x \in H$. There exists a unique $y \in K$ that is closer to x than any other point of K.*

Proof. Let $d = \inf_{z \in K} \|x - z\|$. Choose $y_n \in K$ such that $d_n \to d$, where $d_n = \|x - y_n\|$. By the parallelogram identity with $\frac{x-y_n}{2}$ and $\frac{x-y_m}{2}$, we have

$$\Big\|x - \frac{y_n + y_m}{2}\Big\|^2 + \tfrac{1}{4}\|y_n - y_m\|^2 = \tfrac{1}{2}d_n^2 + \tfrac{1}{2}d_m^2.$$

The right hand side tends to d^2. Since K is convex, $\frac{y_n+y_m}{2} \in K$, so the first term on the left is greater than or equal to d^2. Therefore $\frac{1}{4}\|y_n - y_m\|^2 \to 0$. Thus $\{y_n\}$ is a Cauchy sequence. Since H is complete, $y_n \to y$ for some $y \in H$. Since K is closed, $y \in K$. We have

$$\|x - y\| = \lim \|x - y_n\| = \lim d_n = d.$$

If y' is another point with $\|x - y\| = d$, applying the parallelogram identity with $x - y$ and $x - y'$ for the last equality in what follows,

$$\begin{aligned} 4d^2 + \|y - y'\|^2 &\le 4\Big\|x - \frac{y + y'}{2}\Big\|^2 + \|y - y'\|^2 \\ &= \|2x - (y + y')\|^2 + \|y - y\|^2 = 4d^2. \end{aligned}$$

So we must have equality, and then $\|y - y'\|^2 = 0$. □

If Y is a linear subspace of H, the orthogonal complement of Y is

$$Y^\perp = \{v \in H : (v, y) = 0 \text{ for all } y \in Y\},$$

read "Y perp."

Proposition 6.4 *Suppose Y is a closed subspace of H. Then*

1) $Y^\perp$ is a closed linear subspace of H.

2) $H = Y \oplus Y^\perp$.

3) $(Y^\perp)^\perp = Y$.

Proof. 1) (We don't need Y closed for this first part.) That $Y^\perp$ is a linear subspace is easy. If $v_n \in Y^\perp$ and $v_n \to v$, then for any $y \in H$,

$$|(v, y) - (v_n, y)| = |(v_n - v, y)| \leq \|v - v_n\| \, \|y\| \to 0,$$

so $(v, y) = \lim(v_n, y) = 0$, and hence $v \in Y^\perp$, Therefore $Y^\perp$ is closed.

2) If $x \in H$, choose $y \in Y$ closest to x and set $v = x - y$. Since y is closest, for all $z \in Y$,

$$\|v\|^2 \leq \|v + tz\|^2 = \|v\|^2 + 2t\mathrm{Re}\,(v, z) + t^2\|z\|^2.$$

Taking t very small and both positive and negative, this is only possible if $\mathrm{Re}\,(v, z) = 0$. By replacing z by az so that (v, az) is real, we see $(v, z) = 0$. Therefore $v \in Y^\perp$ and $x = y + v$.

If $x = y' + v'$ as well, then $y - y' = v' - v \in Y \cap Y^\perp$. So $y - y'$ is orthogonal to itself, and by positivity, $y - y' = 0$. Hence $y = y'$ and $v = v'$.

3) If $y \in Y$, then for any $v \in Y^\perp$ we have $(y, v) = 0$, and hence $y \in (Y^\perp)^\perp$. We thus need to show $(Y^\perp)^\perp \subset Y$.

By 2), $H = Y \oplus Y^\perp$. If $y \in (Y^\perp)^\perp$, we can write $y = v + z$ with $z \in Y \subset (Y^\perp)^\perp$ and $v \in Y^\perp$. Then $v = y - z \in (Y^\perp)^\perp$. Since v is also in $Y^\perp$, we see $v = 0$, or $y = z \in Y$. $\square$

6.3 Linear functionals

For each y, $\ell(x) = (x, y)$ is a linear functional, and $|\ell(x)| \leq c\|x\|$.

The converse also holds. Note R_ℓ is one-dimensional and R_ℓ is isomorphic to H/N_ℓ, so the codimension of N_ℓ is 1. We have $H = N_\ell \oplus N_\ell^\perp$. So the obvious candidate to take is cy, where $y \in N_\ell^\perp$.

A linear functional ℓ is bounded if there exists c such that $|\ell(x)| \leq c\|x\|$ for all x.

Theorem 6.5 *(Riesz-Frechét representation) Let ℓ be a bounded linear functional on a Hilbert space. Then there exists a unique $y \in H$ such that $\ell(x) = (x, y)$ for all x.*

Proof.

If $\ell \neq 0$, then N_ℓ is a closed subspace Y of H. (To show it is closed is easy.) We can write $H = Y \oplus Y^\perp$. Take $p \in Y^\perp$ with $\|p\| = 1$.

Let $z = \ell(x)p - \ell(p)x$. Then $\ell(z) = 0$, so $z \in Y$, and hence $(p, z) = 0$. This says

$$\ell(x)(p, p) - \ell(p)(x, p) = 0,$$

or $\ell(x) = \ell(p)(x, p)$. We take $y = \overline{\ell(p)}p$.

To show uniqueness, if $(x, y') = \ell(x) = (x, y)$, then $(x, y - y') = 0$ for all x, in particular when $x = y - y'$. Now use positivity. □

Theorem 6.6 *(Lax-Milgram lemma) Let H be a Hilbert space and suppose*

1. *for each y, $B(x, y)$ is linear in x;*
2. *for each x, $B(x, y_1 + y_2) = B(x, y_1) + B(x, y_2)$ and $B(x, cy) = \bar{c}B(x, y)$;*
3. *there exists c such that $|B(x, y| \leq c\|x\|\,\|y\|$*
4. *there exists b such that $|B(y, y)| \geq b\|y\|^2$ for all y.*

(We do not assume $B(x, y) = \overline{B(y, x)}$.) Then every bounded linear functional ℓ is of the form $\ell(x) = B(x, y)$ for some unique y.

Proof. For each y, $B(x, y)$ is a bounded linear functional of x, so there exists $z = z(y)$ such that $B(x, y) = (x, z)$ for all x, and z is unique.

If $Z = \{z : z = z(y) \text{ for some } y \in H\}$, then Z is a linear space.

Z is closed: setting $x = y$, and letting $z = z(y)$,

$$b\|y\|^2 \leq B(y, y) = (y, z) \leq c\|y\|\,\|z\|,$$

or

$$b\|y\| \leq \|z\|.$$

If $z_n \in Z$ and $z_n \to z$, let y_n be a point such that $z_n = z(y_n)$. Then $B(x, y_n) = (x, z_n)$. So $B(x, y_n - y_m) = (x, z_n - z_m)$, hence $b\|y_n - y_m\| \leq \|z_n - z_m\|$, and therefore y_n is a Cauchy sequence. H is complete; let y be the limit. Since $B(x, y_n) \to B(x, y)$ and $(x, z_n) \to (x, z)$, we have $B(x, y) = (x, z)$, and hence $z \in Z$.

$Z = H$: For each y, there exists $z(y)$ such that $B(x, y) = (x, z)$ for all y. If $Z \neq H$, there exists $x \in Z^\perp$. Applying the above with $y = x$, there exists $z(x)$ such that $B(x, x) = (x, z(x))$. Since $x \in Z^\perp$ and $z(x) \in Z$, $b\|x\|^2 \leq B(x, x) = (x, z(x)) = 0$. So $x = 0$.

Existence: given ℓ, there exists y such that $\ell(x) = (x, y)$ for all x. Then $\ell(x) = B(x, z(y))$.

Uniqueness: if there are two such z, then $B(x, z - z') = B(x, z) - B(x, z') = \ell(x) - \ell(x) = 0$. Now set $x = z - z'$. □

6.4 Linear spans

The closed linear span of S is the intersection of all closed linear subspaces containing S.

One can check that the closed linear span of S equals the closure of the linear span of S.

Proposition 6.7 *y is in the closed linear span Y of $\{y_j\}$ if and only if $(y_j, z) = 0$ for all j implies $(y, z) = 0$.*

Proof. Suppose y is in the closed linear span. If $(y_j, z) = 0$ for all j, then $(\sum a_j y_j, z) = 0$, so $(y, z) = 0$ by continuity of the inner product. So let's look at the other direction.

Let $Z = \{z : (z, y_j) = 0 \text{ for all } j\}$. We claim $Z = Y^\perp$.

If z is orthogonal to all the y_j, then z is orthogonal to all linear combinations, and hence to the limits of linear combinations. Thus $z \in Y^\perp$.

If $z \in Y^\perp$, then z is orthogonal to all the y_j, hence is in Z.

Therefore $Z = Y^\perp$, and so $Y = (Y^\perp)^\perp = Z^\perp$.

To conclude the proof, suppose $z \in Z$. Then $(z, y_j) = 0$ for all j. By hypothesis, $(z, y) = 0$. Then $y \in Z^\perp = Y$. □

We say a set $\{x_j\}$ is orthonormal if (x_j, x_k) equals 0 when $j \neq k$ and equals 1 when $j = k$. $\{x_j\}$ forms an orthonormal basis if the elements are orthonormal and the closed linear span is all of H.

If $\sum |a_j|^2 < \infty$, we can define $\sum a_j x_j$ as the limit in the Hilbert space norm of finite sums.

We need Bessel's inequality:

Proposition 6.8 *Suppose $y \in H$ and $a_j = (y, x_j)$. Then $\sum |a_j|^2 \le \|y\|^2$.*

Proof. If G is a finite collection of the x_j's, then

$$\begin{aligned} 0 \le \Big\| y - \sum_G a_j x_j \Big\|^2 &= \|y\|^2 - \sum_G \overline{a}_j (y, x_j) - \sum_G a_j (x_j, y) + \sum_G |a_j|^2 \\ &= \|y\|^2 - \sum_G |a_j|^2. \end{aligned}$$

Therefore

$$\sum_G |a_j|^2 \le \|y\|^2,$$

and we take the supremum over all finite collections. □

Lemma 6.9 *If $\{x_j\}$ is an orthonormal set, then the closed linear span is equal to*

$$\Big\{ \sum a_j x_j : \sum |a_j|^2 < \infty \Big\}.$$

Note in this case $\|x\|^2 = \sum |a_j|^2$ and $a_j = (x, x_j)$. To see this, because the $\{x_j\}$ are orthonormal, for any finite sum

$$\Big\| \sum a_j x_j \Big\|^2 = \sum \|a_j x_j\|^2$$

as the cross terms are zero. We then take limits. To get the formula for a_j,

$$(x, x_k) = \sum a_j(x_j, x_k) = a_k.$$

Theorem 6.10 *Every Hilbert space has an orthonormal basis.*

If we consider all orthonormal sets, ordered by inclusion, then we use Zorn's lemma to get a maximal element. We want to show the closed linear span Y of a maximal element is all of H.

Suppose not. Suppose there exists $y \notin Y$. Let $a_j = (y, x_j)$.

Define $x = \sum a_j x_j$, and note $x \in X$. We have

$$(y - x, x_j) = (y, x_j) - (x, x_j) = a_j - a_j = 0.$$

So $y - x$ is orthogonal to all the x_j. Since $y - x \neq 0$ (because $y \notin Y$), $\frac{y-x}{\|y-x\|}$ could be added to the collection $\{x_j\}$ to make a strictly larger orthonormal set, a contradiction. □

If H is separable, then any orthonormal basis will be countable, and we wouldn't need Zorn's lemma to obtain the orthonormal basis.

Proposition 6.11 *Suppose H is a Hilbert space, and $\{x_j\}$, $\{y_j\}$ are two orthonormal bases for H. Given $x \in H$, we have $x = \sum a_j x_j$ with $a_j = (x, x_j)$. The map $x \to y = \sum a_j y_j$ is an isometry.*

7 Applications of Hilbert spaces

7.1 The Radon-NIkodym theorem

We can use Hilbert space techniques to give an alternate proof of the Radon-Nikodym theorem.

Suppose μ and ν are finite measures on a space S and we have the condition $\nu(A) \leq \mu(A)$ for all measurable A. For $f \in L^2(\mu)$, define

$$\ell(f) = \int f \, d\nu.$$

Our condition implies $\int h\,d\nu \le \int h\,d\mu$ if $h \ge 0$. We use this with $h = |f|$ and use Cauchy-Schwarz and have

$$|\ell(f)| = \Big|\int f\,d\nu\Big| \le \int |f|\,d\nu \le \int |f|\,d\mu$$
$$\le \Big(\mu(S)\Big)^{1/2}\Big(\int f^2\,d\mu\Big)^{1/2} \le c\|f\|_{L^2(\mu)}.$$

There exists g such that $\ell(f) = (f, g)$, which translates to

$$\int f\,d\nu = \int fg\,d\mu.$$

Letting $f = \chi_A$, we get $\nu(A) = \int_A g\,d\mu$.

If ν is absolutely continuous with respect to μ, we let $\rho = \mu + \nu$ and apply the above to ν and ρ and also to μ and ρ. The absolute continuity implies that $d\mu/d\rho > 0$ a.e., and we use

$$\frac{d\nu}{d\mu} = \frac{d\nu}{d\rho}\frac{d\mu}{d\rho}.$$

7.2 The Dirichlet problem

Let D be a bounded domain in $\mathbb{R}^n$, contained in $B(0, K)$, say, where this is the ball of radius K about 0. Let (f, g) be the usual L^2 scalar product for real valued functions. It is easy to see that if $C_0^\infty(D)$ is the set of C^∞ functions that vanish on the boundary of D, then the completion of $C^\infty(D)$ with respect to the L^2 norm is simply $L^2(D)$. Define

$$\mathcal{E}(f, g) = \int_D (\nabla f(x), \nabla g(x))\,dx.$$

Clearly $\mathcal{E}$ is bilinear and symmetric.

If we start with

$$f(x_1, \ldots, x_n) = \int_{-K}^{x_1} \frac{\partial f}{\partial x_1}(y, x_2, \ldots, x_n)\,dy$$

and apply Cauchy-Schwarz, we have

$$|f(x_1, \ldots, x_n)|^2 \le \int_{-K}^{K} 1\,dy \int_{-K}^{K} |\nabla f(y, x_2, \ldots, x_n)|^2\,dy.$$

Integrating over $(x_2, \dots, x_n) \in [-K, K]^{n-1}$ we obtain

$$\int_D |f(x)|^2 \, dx \le c \int_D |\nabla f(x)|^2 \, dx,$$

or in other words,

$$(f, f) \le c\mathcal{E}(f, f).$$

If $\mathcal{E}(f, f) = 0$, then $(f, f) = 0$, and so $f = 0$ (a.e., of course). This proves that $\mathcal{E}$ is positive. We let H_0^1 be the completion of $C_0^\infty(D)$ with respect to the norm induced by $\mathcal{E}$. The superscript 1 refers to the fact we are working with first derivatives, the subscript 0 to the fact that our functions vanish on the boundary. $\mathcal{E}$ is an example of a Dirichlet form, something we will probably look at more much later.

Recall the divergence theorem:

$$\int_{\partial D} (F, n) \, d\sigma = \int_D \operatorname{div} F \, dx,$$

where D is a reasonably smooth domain, ∂D is the boundary of D, n is the outward pointing unit normal, and σ is surface measure. In three dimensions, this is also known as Gauss' theorem, and along with Green's theorem and Stokes' theorem are consequences of the fundamental theorem of calculus.

If we apply the divergence theorem to $F = u \operatorname{div} v$, then

$$\frac{\partial}{\partial x_1} F_1 = \frac{\partial u}{\partial x_1} \frac{\partial v}{\partial x_1} + u \frac{\partial^2 v}{\partial x_1^2},$$

and so

$$\operatorname{div} F = (\nabla u, \nabla v) + u \Delta v,$$

where Δv is the Laplacian. Also,

$$(\operatorname{div} F, n) = u \frac{\partial v}{\partial n},$$

where $\frac{\partial v}{\partial n}$ is the normal derivative of v. We then get Green's first identity:

$$\int_D u \Delta v + \int_D (\nabla u, \nabla v) = \int_{\partial D} u \frac{\partial v}{\partial n}.$$

Our goal is to solve the equation $\Delta v = g$ in D with $v = 0$ on the boundary of D. This is Poisson's equation, while the Dirichlet problem more properly refers to the equation $\Delta v = 0$ in D with v equal to some pre-specified function f on the boundary of D.

If we have a solution v and $u \in C_0^\infty(D)$, then by Green's identity we get

$$\int_D u(x)g(x)\,dx = -\int_D (\nabla u(x), \nabla v(x))\,dx.$$

So one way of formulating a (weak) solution to Poisson's equation is: given $g \in L^2(D)$, find $v \in H_0^1$ such that

$$\mathcal{E}(u,v) = -\int ug$$

for all $u \in C_0^\infty(D)$.

After all this, it is easy to find a weak solution to the Poisson equation. Suppose $g \in H_0^1$. Define $\ell(u) = -(u, g)$. Then

$$|\ell(u)| \le \|g\|\,\|u\| \le c\|g\|\mathcal{E}(u,u)^{1/2}.$$

By the Riesz-Frechét theorem, there exists $v \in H_0^1$ such that $\ell(u) = \mathcal{E}(u,v)$ for all u. So

$$\mathcal{E}(u,v) = \ell(u) = -(u,g),$$

and v is the desired solution.

8 Duals of normed linear spaces

8.1 Bounded linear functionals

If X is a normed linear space, a linear functional ℓ is a linear map from X to F, the field of scalars. ℓ is continuous if $|x_n - x| \to 0$ implies $\ell(x_n) \to \ell(x)$. ℓ is bounded if there exists c such that $|\ell(x)| \le c|x|$ for all x.

Theorem 8.1 *A linear functional ℓ is continuous if and only if it is bounded.*

Proof. If f is bounded,

$$|\ell(x_n) - \ell(x)| = |\ell(x_n - x)| \leq c|x_n - x| \to 0,$$

and so it is continuous.

Suppose ℓ is continuous but not bounded. So there exist x_n such that $\ell(x_n) > n|x_n|$. If

$$y_n = \frac{1}{\sqrt{n}} \frac{x_n}{|x_n|},$$

then $|y_n - 0| = |y_n| \to 0$, but

$$\ell(y_n) > \frac{1}{\sqrt{n}} n \frac{|x_n|}{|x_n|} = \sqrt{n},$$

which does not tend to $0 = \ell(0)$, a contradiction. □

The collection of all continuous linear functionals of X is called the dual of X, written X' or X^*.

Note $N_\ell = \ell^{-1}(\{0\})$ is closed, since ℓ is continuous.

Define

$$|\ell| = \sup_{|x| \neq 0} \frac{|\ell(x)|}{|x|}.$$

By linearity, this is the same as $\sup_{|x|=1} |\ell(x)|$.

Proposition 8.2 *X^* is a Banach space.*

Proof.

Subadditivity:

$$\begin{aligned} |\ell + m| &= \sup_{|x|=1} |(\ell + m)(x)| \leq \sup_{|x|=1} (|\ell(x)| + |m(x)|) \\ &\leq \sup_{|x|=1} |\ell(x)| + \sup_{|x|=1} |m(x)| = |\ell| + |m|. \end{aligned}$$

Completeness: Suppose $|\ell_n - \ell_m| \to 0$ as $n, m \to \infty$. For each x,

$$|\ell_n(x) - \ell_m(x)| \leq |\ell_n - \ell_m| \, |x| \to 0.$$

Since $\mathbb{R}$ and $\mathbb{C}$ are complete, $\lim \ell_n(x)$ exists for each x; call the limit $\ell(x)$.

Given ε, choose N such that $|\ell_n - \ell_m| < \varepsilon$ if $n, m \geq N$. So $|\ell_n(x) - \ell_m(x)| \leq \varepsilon|x|$. Let $m \to \infty$, so $|\ell_n(x) - \ell(x)| \leq \varepsilon|x|$ if $n \geq N$. This means $|\ell_n - \ell| \leq \varepsilon$ if $n \geq N$, or $\ell_n \to \ell$. □

8.2 Extensions of bounded linear functionals

Proposition 8.3 *Let X be a normed linear space, Y a subspace, ℓ a linear functional on Y with $|\ell(y)| \leq c|y|$ for all $y \in Y$. Then ℓ can be extended to a bounded linear functional on X with the same bound on X as on Y.*

Proof. This is the Hahn-Banach theorem with $p(x) = c|x|$. □

$y_1, \ldots, y_N$ are said to be linearly independent if $\sum_{i=1}^N c_i y_i = 0$ implies all the c_i are zero.

Theorem 8.4 *Suppose $y_1, \ldots, y_N$ are linearly independent and $a_1, \ldots, a_N$ are scalars. Then there exists a bounded linear functional ℓ such that $\ell(y_j) = a_j$.*

Proof. Let Y be the span of $y_1, \ldots, y_N$. If $y \in Y$, then y can be written as $\sum b_j y_j$ in only one way, for if $\sum b'_j y_j$ is another way, then

$$\sum (b_j - b'_j) y_j = y - y = 0,$$

and so $b_j = b'_j$ for all j. Define

$$\ell\Big(\sum b_j y_j\Big) = \sum a_j b_j.$$

Now use the preceding theorem to extend ℓ to all of X. □

Theorem 8.5 *If X is a normed linear space, then*

$$|y| = \max_{|\ell|=1} |\ell(y)|.$$

Proof. $|\ell(y)| \leq |\ell|\,|y|$, so the maximum on the right hand side is less than or equal to y.

If $y \in X$, let $Y = \{ay\}$ and define $\ell(ay) = a|y|$. Then the norm of ℓ on Y is 1. Now extend ℓ to all of X so as to have norm 1. □

Theorem 8.6 *Let X be a normed linear space over $\mathbb{C}$ and Y a linear subspace. Let*

$$m(z) = \inf_{y \in Y} |z - y|, \qquad M(z) = \max_{|\ell| \leq 1, \ell = 0 \text{ on } Y} |\ell(z)|.$$

Then $m(z) = M(z)$.

Proof. If $y \in Y$, $\ell = 0$ on Y, and $|\ell| \leq 1$, then

$$|\ell(z)| = |\ell(z - y)| \leq |z - y|.$$

So

$$|\ell(z)| \leq \inf_{y \in Y} |z - y| = m(z),$$

and hence $M(z) \leq m(z)$.

Let $Y_0 = \{y + az : y \in Y, a \in \mathbb{C}\}$. Define $\ell_0(z)$ on Y_0 by $\ell_0(y + az) = am(z)$. By the definition of $m(z)$, ℓ_0 is bounded on Y_0 by 1. (To see this, we write

$$|\ell_0(y + az)| = |a| m(z) \leq |a| \left| z - \frac{-y}{a} \right| = |az + y|.$$

Extend to all of X with the same bound.

$$\ell_0(z) = \ell(0 + 1 \cdot z) = m(z),$$

so $m(z) \leq M(z)$. □

We write $Y^{\perp} = \{\ell : \ell = 0 \text{ on } Y\}$ for the annihilator of Y.

Theorem 8.7 *(Spanning criterion) Let Y be the closed linear span of $\{y_j\}$. Then $z \in Y$ if and only if $\ell(y_j) = 0$ for a bounded linear functional ℓ for all j implies $\ell(z) = 0$.*

Proof. If $\ell(y_j) = 0$ for all j, then $\ell(y)$ for all y of the form $\sum a_j y_j$, and by continuity of ℓ, for all $y \in Y$.

Suppose $z \notin Y$. Then

$$\inf_{y \in Y} |z - y| = d > 0.$$

Let $Z = \{y + az : y \in Y\}$. Define ℓ_0 on Z by $\ell_0(y + az) = a$. So

$$|y + az| = |a| \left| - \frac{-y}{a} + z \right| \geq d|a|.$$

Therefore on Z, ℓ_0 is bounded by d^{-1}. Extend ℓ_0 to all of X. But then $\ell_0(y_j) = 0$ while $\ell_0(z) = 1$. □

8.3 Reflexive spaces

If $x \in X$, define the linear functional L_x on X^* by

$$L_x(\ell) = \ell(x).$$

The norm of L_x is $|x|$. So we can isomorphically embed X into X^{**}.

A Banach space is reflexive if $X^{**} = X$.

Theorem 8.8 *Hilbert spaces are reflexive.*

Proof. Recall $X^* = X$, and the result follows from this. To see $X^* = X$, if ℓ is a linear functional, there exists y such that $\ell(x) = (x, y)$ for all x. If we show $|\ell| = \|y\|$, this gives an isometry between X and X^*. By Cauchy-Schwarz, $|\ell(x)| = |(x, y)| \leq \|x\| \, \|y\|$, so $|\ell| \leq \|y\|$. Taking $x = y$, $\ell(y) = \|y\|^2$, hence $|\ell| \geq \|y\|$. □

If $1 < p, q < \infty$ and $\frac{1}{p} + \frac{1}{q} = 1$, then the dual of L^p is isomorphic to L^q. Hence the L^p spaces are reflexive.

To prove the dual of L^p is L^q, two facts from real analysis are needed. See Folland, p.190 and preceding.

1) Hölder's inequality:

$$\int |fg| \le \|f\|_p \|g\|_q.$$

2)

$$\|g\|_q = \sup\left\{ \left| \int fg \right| : \|f\|_p = 1, f \text{ simple} \right\}.$$

Theorem 8.9 $(L^p)^* = L^q$.

Proof. (Sketch) If we define $\ell(f) = \int fg$, then Hölder's inequality shows that ℓ is a bounded linear functional.

Now suppose ℓ is a bounded linear functional on L^p. The heart of the matter is when μ, the measure, is finite, so let's suppose that. Define $\nu(E) = \ell(\chi_E)$. The linearity of ℓ shows ν is finitely additive. The fact that ℓ is a continuous linear functional allows one to prove countable additivity. If $\mu(E) = 0$, then $\chi_E = 0$ (a.e.), so $\nu(E) = \ell(0) = 0$. Therefore ν is absolutely continuous with respect to μ. Note ν is either a signed measure or is complex valued.

By the Radon-Nikodym theorem there exists $g \in L^1$ such that

$$\ell(\chi_E) = \nu(E) = \int_E g.$$

By linearity, if f is simple,

$$\ell(f) = \int fg.$$

Since $\ell(f) \le c\|f\|_p$, because ℓ is a bounded linear functional on L^p, taking the supremum over simple f's with L^p norm 1, we get

$$\|g\|_q \le |\ell|.$$

We can then use the continuity of ℓ and Hölder's inequality to obtain

$$\ell(f) = \int fg$$

for all $f \in L^p$. □

Proposition 8.10 *If X is a normed linear space over $\mathbb{C}$ and X^* is separable, then X is separable.*

Proof. Since X^* is separable, there is a countable dense subset $\{\ell_n\}$. Recall $|\ell_n| = \sup_{|x|=1} |\ell_n(x)|$. So for each n there exists $x_n \in X$ such that $|x_n| = 1$ and $\ell_n(x_n) > \frac{1}{2}|\ell_n|$.

We claim $\{x_n\}$ is dense in X. To prove this, we start by showing that if ℓ is a linear functional on X that vanishes on $\{x_n\}$, then ℓ vanishes identically.

Suppose not and that there exists ℓ such that $\ell(x_n) = 0$ for all x_n but $\ell \neq 0$. We can normalize so that $|\ell| = 1$. Since the ℓ_n are dense in X^*, there exists ℓ_n such that $|\ell - \ell_n| < 1/3$. Therefore $|\ell_n| > 2/3$.Then

$$\tfrac{1}{3} > |(\ell - \ell_n)(x_n)| = |\ell_n(x_n)| > \tfrac{1}{2}|\ell_n| > \tfrac{1}{2} \cdot \tfrac{2}{3},$$

a contradiction.

Therefore by the spanning criterion, the closed linear span of the x_n is all of X. Then the set of finite linear combinations of the x_n where all the coefficients have rational coordinates, is also dense in X, and is countable. □

By the Riesz representation theorem from real analysis, the dual of $X = C([0,1])$ is the set of finite signed measures on $[0,1]$. X is separable, but X^* is not, since $|\delta_x - \delta_y| = 2$ whenever $x \neq y$. It follows that X is not reflexive: if it were, we would have X^{**} separable, but X^* not, contradicting the previous proposition.

Proposition 8.11 *Suppose X is a reflexive normed linear space and Y is a subspace. Then Y is reflexive.*

Proof. If ℓ is a linear functional on X, let ℓ_0 be the restriction of ℓ to Y. By the Hahn-Banach theorem every bounded linear functional on Y can be extended to X. So the restriction map $R : \ell \to \ell_0$ maps X^* onto Y^*.

If $\varphi \in Y^{**}$, define $\overline{R}\varphi$ by

$$\overline{R}\varphi(\ell) = \varphi(\ell_0), \qquad \ell \in X^*.$$

Here $\overline{R}$ maps Y^{**} into X^{**}.

Since X is reflexive, there exists $z \in X$ such that $\overline{R}\varphi(\ell) = \ell(z)$ for all linear functionals ℓ. Therefore $\ell(z) = \varphi(\ell_0)$.

We argue that $z \in Y$. If $\ell \in Y^{\perp}$, then $\ell_0 = 0$. So $\ell(z) = \varphi(\ell_0) = \varphi(0) = 0$. By the spanning criterion, z is in the closure of Y, and since Y is closed, it is in Y.

Therefore

$$\ell_0(z) = \varphi(\ell_0).$$

Every functional in Y^* is an ℓ_0 for some ℓ, so every $\varphi \in Y^{**}$ can be identified with some $z \in Y$. □

8.4 The support function

If M is a subset of a linear space, define $\breve{M}$ to be the closure of the convex hull of M.

If M is a bounded subset of a normed linear space X over $\mathbb{R}$, define the map S_M from X^* into $\mathbb{R}$ by

$$S_M(\ell) = \sup_{y \in M} \ell(y).$$

S_M is called the support function.

If $M = \{x_0\}$, then $S_M(\ell) = \ell(x_0)$. If $M = B_R(0)$, then $S_M(\ell) = R|\ell|$. Using the fact that $S_{M+N}(\ell) = S_M(\ell) + S_N(\ell)$, if $M = B_R(x_0)$, then $S_M(\ell) = \ell(x_0) + R|\ell|$.

Proposition 8.12 *If M is a bounded subset of X, then $z \in \breve{M}$ if and only if $\ell(z) \le S_M(\ell)$ for all $\ell \in X^*$.*

Proof. One way is easy. If $\ell \in X^*$ and $z \in \breve{M}$, then $\ell(z) \le S_{\breve{M}}(\ell)$. It is easy to check that $S_{\breve{M}}(\ell) = S_M(\ell)$.

Now suppose $z \notin \breve{M}$, but $\ell(z) \le S_M(\ell)$ for all ℓ. $\breve{M}$ is closed, so there exists R such that $B_R(z) \cap \breve{M} = \emptyset$. By the hyperplane separation theorem, there exists $\ell_0 \neq 0$ and c such that

$$\ell_0(u) \le c \le \ell_0(v), \qquad u \in \breve{M}, v \in B_R(z).$$

ℓ_0 is a bounded linear functional.

If $v \in B_R(z)$, then $v = z + Rx$ with $|x| < 1$, so

$$c \le \ell_0(z) + R\ell_0(x).$$

Also $\inf_{|x|<1} \ell_0(x) = -|\ell_0|$. So $c \le \ell_0(z) - R|\ell_0|$. We have $S_M(\ell_0) \le c$, so

$$S_M(\ell_0) + R|\ell_0| \le \ell_0(z),$$

a contradiction to $\ell(z) \le S_M(\ell)$. □

Proposition 8.13 *Suppose K is a closed and convex subset of X and $z \notin K$. Then*

$$\inf_{u \in K} |z - u| = \sup_{|\ell|=1} [\ell(z) - S_K(\ell)].$$

Proof. $S_K(\ell) \ge \ell(u)$ for all $\ell \in X^*$ and all $u \in K$. If $|\ell| = 1$, then

$$S_K(\ell) \ge \ell(u) = \ell(u) + \ell(u - z) \ge \ell(z) - |u - z|,$$

or $|u - z| \ge \ell(z) - S_K(\ell)$. So the left hand side is larger than the right hand side.

Let $0 < R < \inf_{u \in K} |z - u|$. $K + B_R$ has positive distance from z. So

$$S_{K+B_R}(\ell_0) < \ell_0(z)$$

for some $\ell_0 \in X^*$ and we can take $|\ell_0| = 1$. $S_{K+B_R}(\ell_0) = S_K(\ell_0) + R|\ell_0|$, or

$$R < \ell_0(z) - S_K(\ell_0).$$

Therefore the right hand side is larger than R. This is true for any such R, so the right side is larger than the left side. □

9 Weak convergence

Let X be a normed linear space. We say x_n converges to x weakly, written $w-\lim x_n = x$ or $x_n \xrightarrow{w} x$ if $\ell(x_n) \to \ell(x)$ for all $\ell \in X^*$.

x_n converges to x strongly, written $s-\lim x_n = x$ or $x_n \xrightarrow{s} x$ if $|x_n - x| \to 0$.

Strong convergence implies weak convergence because

$$|\ell(x_n) - \ell(x)| = |\ell(x_n - x)| \leq |\ell|\,|x_n - x| \to 0.$$

As an example where we have weak convergence but not strong convergence, let $X = \ell^2$ and let e_n be the element whose n^{th} coordinate is 1 and all other coordinate coordinates are 0. Since $|e_n| = 1$, then e_n does not converge strongly to 0. But it does converge weakly to 0. To see this, if ℓ is any bounded linear functional on X, then ℓ is of the form $\ell(x) = (x, y)$ for some $y \in X$, which means $y = (b_1, b_2, \ldots)$ with $\sum_j |b_j|^2 < \infty$. In particular, $b_j \to 0$. Then $\ell(e_n) = \overline{b}_n \to 0 = \ell(0)$.

This example stretches to any Hilbert space. If $\{x_n\}$ is an orthonormal sequence in the space, $\ell(x_n) = (x_n, y)$ for some y. By Bessel's inequality, $\sum |(x_n, y)|^2 \leq \|y\|^2$, so $(x_n, y) \to 0$,

9.1 Uniform boundedness

Theorem 9.1 *Let X be a Banach space and $\{\ell_\nu\}$ a collection of bounded functionals such that $|\ell_\nu(x)| \leq M(x)$ for all ν and each x. Then there exists c such that $|\ell_\nu| \leq c$.*

In other words, if the ℓ_ν are bounded pointwise, they are bounded uniformly.

The proof relies on the Baire category theorem, which we now recall. If A is a set, we use $\overline{A}$ for the closure of A and A^o for the interior of A. A set A is dense in X if $\overline{A} = X$ and A is nowhere dense if $(\overline{A})^o = \emptyset$.

The Baire category theorem is the following.

Theorem 9.2 *Let X be a complete metric space.*

(a) If G_n are open sets with $\overline{G_n} = X$, then $\cap_n G_n$ is dense in X.

(b) X cannot be written as the countable union of nowhere dense sets.

We now prove the uniform boundedness theorem. (A generalization to bounded linear maps is called the Banach-Steinhaus theorem).

Proof. Let $M(x) = \sup_\nu |\ell_\nu(x)|$. Let $G_n = \{x : M(x) > n\}$. Since $x \in G_n$ if and only if for some $\nu \in A$ we have $|\ell_\nu(x)| > n|x|$ and $x \to |\ell_\nu(x)|$ is a continuous function by the triangle inequality, we conclude

$$G_n = \cup_\nu \{x : |\ell_\nu(x)| > n|x|\}$$

is the union of open sets, so is open.

Suppose G_N is not dense in X. So there exists x_0 and r such that $\overline{B(x_0, r)} \cap G_N = \emptyset$, or if $|x| \le r$, then $\ell(x_0 + x) \le N$ for all ν. Since $x = (x_0 + x) - x_0$,

$$|\ell_\nu(x)| \le |\ell_\nu((x_0 + x)| + |\ell_\nu(x_0)| \le 2N$$

if $|x| \le r$, and we then have $\sup_\nu |\ell_\nu| \le c$ with $c = 2N/r$.

The other possibility, by Baire's theorem, is that every G_n is dense in X and $\cap_n G_n$ is a dense subset of X. But $M(x) = \infty$ for every $x \in \cap_n G_n$. □

Corollary 9.3 *Let X be a normed linear space, $\{x_\nu\}$ a subset such that for all $\ell \in X^*$ we have*

$$|\ell(x_\nu)| \le M(\ell) \qquad \textit{for all } x_\nu.$$

Then there exists c such that $|x_\nu| \le c$ for all x_ν.

Proof. Write $L_\nu(\ell) = \ell(x_\nu)$. So each x_ν acts as a bounded linear functional on X^*. □

Corollary 9.4 *Let X be a normed linear space and suppose x_n converges weakly to x. Then $|x| \le \liminf |x_n|$.*

Proof. There exists ℓ such that $|\ell| = 1$ and $|\ell(x)| = |x|$. Then $|\ell(x)| = \lim |\ell(x_n)|$ and $|\ell(x_n)| \le |\ell|\,|x_n| = |x_n|$. □

9.2 Weak sequential compactness

The topology that goes along with weak convergence is not necessarily derived from a metric. Thus the topology cannot be characterized by sequential convergence.

We say a subset C of a normed linear space X is weakly sequentially compact if any sequence of points in C has a subsequence converging weakly to a point of C.

Theorem 9.5 *Let X be a reflexive Banach space. Then the closed unit ball is weakly sequentially compact.*

Proof. Take $\{y_n\}$ with $|y_n| \le 1$. Let Y be the closed linear span of the y_n's. By Theorem 15 of Chapter 8, Y is reflexive also. Since $Y^{**} = Y$ is separable, then Y^* is separable. Let $\{m_j\}$ be a countable dense subset of Y^*. By a diagonalization procedure, there exists a subsequence z_n of the y_n such that $m_j(z_n)$ converges for all j. Since the m_j are dense in Y^*, we have $m(z_n)$ converges for all $m \in Y^*$. Call the limit $L(m)$.

We can check that $L(m)$ is a linear functional on Y^*. We have

$$|L(m)| \le \limsup |m(z_m)| \le |m|\,|z_n| \le |m|,$$

so the linear functional L on Y^* has norm bounded by 1. $L \in Y^{**}$. So there exists $y \in Y$ such that $L(m) = m(y)$. For all $m \in Y^*$, $m(z_n) \to m(y)$, and hence z_n converges weakly to y. Since $Y \subset X$, then $X^* \subset Y^*$, hence we have the convergence for all m in X^*. □

9.3 Weak* convergence

We say $u_n \in X^*$ is weak* convergent to u if $\lim u_n(x) = u(x)$ for all $x \in X$.

If X is reflexive, then weak* convergence is the same as weak convergence.

Weak convergence in probability theory can be identified as weak* convergence in functional analysis.

As an example, if S is a compact Hausdorff space and $X = C(S)$, then X^* is the collection of finite signed measures. Saying ia sequence of measures ν_n converges in the weak* sense means that $\int f\, d\nu_n$ converges for each continuous function f.

A set is weak* sequentially compact if every sequence in the set has a subsequence which converges in the weak* sense to an element of the set.

Theorem 9.6 *If X is a separable Banach space, then the closed unit ball in X^* is weak* sequentially compact.*

Proof. Let $u_n \in X^*$ with norms bounded by 1. Let $\{x_k\}$ be a countable dense subset of X. By diagonalization there exists a subsequence v_n of the u_n such that $v_n(x_k)$ converges for each k. Since $|v_n| \leq 1$ for all n, then $v_n(x)$ converges for all x. Call the limit $v(x)$. This is a linear functional with norm bounded by 1. □

10 Applications of weak convergence

10.1 Approximating the δ function

Let k_n be a sequence of integrable functions on the interval $[-1, 1]$. They approximate the δ function (or are an approximation to the identity) if

$$\int_{-1}^{1} f(t)k_n(t)\, dt \to f(0) \tag{10.1}$$

as $n \to \infty$ for all f continuous on $[-1, 1]$.

Theorem 10.1 *k_n approximates the δ function on $[-1, 1]$ if and only if the following three properties hold.*

(1) $\int_{-1}^{1} k_n(t)\, dt \to 1$.

(2) If g is C^∞ and 0 in a neighborhood of 0, then

$$\int_{-1}^{1} g(t)k_n(t)\, dt \to 0$$

as $n \to \infty$.

(3) There exists c such that $\int_{-1}^{1} |k_n(t)|\, dt \le c$ for all n.

Proof. If (1)–(3) hold, write $f = (f - f(0)) + f(0)$, and we may suppose without loss of generality that $f(0) = 0$. Choose $g \in C^\infty$ such that g is 0 in a neighborhood of 0 and $|g - f| < \varepsilon$. We have

$$\Big| \int_{-1}^{1} (f-g)k_n \Big| \le \varepsilon \int |k_n| \le c\varepsilon$$

and

$$\int g k_n \to 0.$$

So

$$\limsup \Big| \int f k_n \Big| \le c\varepsilon.$$

Since ε is arbitrary, this shows (10.1).

If (10.1) holds, then (1) holds by taking f identically 1 and (2) holds by taking f equal to g. So we must show (3). If X is the set C of continuous functions on $[-1, 1]$, then X^* is the collection of finite signed measures. Let $m_n(dt) = k_n(t)\, dt$ and $m_0(dt) = \delta_0(dt)$. Then (10.1) says that $m_n(f) \to m_0(f)$ for all $f \in C$, or m_n converges to m_0 in the sense of weak-* convergence. $\limsup |m_n(x)| < \infty$, so $|m_n(x)| \le M(x)$ for all x, and by the uniform boundedness principle, $|m_n| \le c$. Note $|m_n|$ is the total mass of m_n, which is $\int_{-1}^{1} |k_n(t)|\, dt$. □

10.2 Divergence of Fourier series

From the approximation of the δ-function, we can show that there exists a continuous function f whose Fourier series diverges at 0.

We look at the set of continuous functions on S^1, the unit circle. We say $f(\theta)$ has Fourier series $\sum_{-\infty}^{\infty} a_n e^{in\theta}$ with

$$a_n = \frac{1}{2\pi} \int_{-\pi}^{\pi} f(\theta) e^{-in\theta}\, d\theta.$$

The Fourier series converges at 0 if

$$\lim_{N\to\infty}\sum_{-N}^{N} a_n = f(0).$$

Now

$$\sum_{-N}^{N} a_n = \frac{1}{2\pi}\int_{-\pi}^{\pi} f(\theta)k_N(\theta)\,d\theta,$$

where

$$k_N(\theta) = \sum_{-N}^{N} e^{-in\theta} = \frac{\sin(N+\frac{1}{2})\theta}{\sin\theta/2}.$$

So the convergence of the Fourier series at 0 is equivalent to k_N being an approximation to the δ function. And if (3) fails, then $\sum_{-N}^{N} a_n$ does not converge for some f.

Since $|\sin x| \le |x|$, then

$$\Big|\frac{1}{\sin x/2}\Big| \ge \frac{2}{|x|},$$

and therefore

$$\begin{aligned}\int_{-\pi}^{\pi} |k_N(\theta)|\,d\theta &\ge 2\int_{-\pi}^{\pi} |\sin(N+\tfrac{1}{2})\theta|\frac{d\theta}{|\theta|}\\ &= 2\int_0^{(N+\frac{1}{2})x} |\sin x|\frac{dx}{x}\\ &\ge c\log N\end{aligned}$$

10.3 The Herglotz theorem

Theorem 10.2 *(Herglotz) Let f be analytic in $\{|z|<1\}$ and $h = \operatorname{Re} f \ge 0$. Then*

$$f(z) = \int_0^{2\pi} \frac{1+ze^{-i\theta}}{1-ze^{-i\theta}}\, m(d\theta) + ic$$

for some positive measure m and some constant c. If f satisfies the above equation, f is analytic with positive real part. The measure m is uniquely determined by f.

Proof. For $R < 1$ we have the Poisson integral formula:

$$f(z) = \frac{1}{2\pi} \int_0^{2\pi} \frac{R + ze^{-i\theta}}{R - ze^{-i\theta}} h(Re^{i\theta})\, d\theta + ic.$$

When $z = 0$,

$$h(0) = \frac{1}{2\pi} \int_0^{2\pi} h(Re^{i\theta})\, d\theta.$$

Take a sequence $R_n \uparrow 1$. Define ℓ_n on $C(S^1)$ by

$$\ell_n(u) = \frac{1}{2\pi} \int_0^{2\pi} h(R_n e^{i\theta}) u(\theta)\, d\theta.$$

Since $h \geq 0$,

$$|\ell_n(u)| \leq \frac{1}{\pi} |u| \int_0^{2\pi} h(R_n e^{i\theta})\, d\theta = |u| h(0).$$

$C(S^1)$ is separable, so by Helly's theorem, there exists a subsequence (also called ℓ_n) which is weak* convergent, say to ℓ. If $|u_n - u| \to 0$ and $\ell_n(u) \to \ell(u)$, then

$$|\ell_n(u_n) - \ell_n(u)| \leq |\ell_n|\, |u_n - u| \to 0,$$

so $\ell_n(u_n) \to \ell(u)$.

Let

$$u_n = \frac{R_n + ze^{-i\theta}}{R_n - ze^{-i\theta}}, \qquad u = \frac{1 + ze^{-i\theta}}{1 - ze^{-i\theta}}.$$

Fix z. Then $|u_n - u| \to 0$, and therefore $f(z) = \ell_n(u_n) \to \ell(u)$. The ℓ_n are positive linear functionals so ℓ is too. By Riesz representation, there exists a measure m such that

$$\ell(u) = \int_0^{2\pi} u(\theta)\, m(d\theta).$$

It is routine to check that if $f(z)$ has the representation in the theorem, then f is analytic.

If we take real parts,

$$h(z) = \frac{1}{2\pi} \int_0^{2\pi} \frac{1 - r^2}{1 - 2r\cos(\phi - \theta) + r^2}\, m(d\theta),$$

where $z = re^{i\phi}$. Multiply by $u(\phi)$ and integrate to get

$$\int_0^{2\pi} h(re^{i\phi})u(\phi)\,d\phi = \int_0^{2\pi} \Big[\frac{1}{2\pi}\int \frac{1-r^2}{1-2r\cos(\phi-\theta)+r^2} u(\phi)\,d\phi\Big]\, m(d\theta).$$

Call the expression inside the brackets $u_r(\theta)$. Letting $r \to 1$, $u_r \to u$ strongly. so

$$\lim_{r\to 1} \int h(re^{i\phi})u(\phi)\,d\phi = \int_0^{2\pi} u(\theta)\,dm.$$

The left hand side does not depend on m. If m_1 and m_2 are two such measures,

$$\int_0^{2\pi} u(\theta)\,m_1(d\theta) = \int_0^{2\pi} u(\theta)\,m_2(d\theta)$$

for all continuous u. Therefore $m_1 = m_2$. □

11 Weak and weak* topologies

The weak topology is the coarsest topology (i.e., fewest sets) in which all bounded linear functionals are continuous.

Bounded linear functionals are continuous in the usual norm topology (also called the strong topology), so the weak topology is coarser than the strong topology.

We argue that every open set in the weak topology is the union of finite intersections of sets of the form $\{x : a < \ell(x) < b\}$. Let

$$\mathcal{S} = \Big\{\{x : a_i < \ell_i(x) < b_i, i = 1, \ldots, n\} : n \in \mathbb{N}, a_i < b_i, \ell_i \in X^*\Big\}.$$

We claim $\mathcal{S}$ is a basis for the weak topology, that is, every open set is a union of elements of $\mathcal{S}$. Note that each element of $\mathcal{S}$ is in the weak topology because $\ell_i^{-1}((a_i, b_i))$ must be in the weak topology, so $\cap_{i=1}^n \ell_i^{-1}((a_i, b_i))$ is an element of the weak topology. Note also that unions of elements of $\mathcal{S}$ form a topology and each ℓ is continuous with respect to this topology. Therefore $\mathcal{S}$ is a basis for the weak topology.

Finite intersections of sets in $\mathcal{S}$ are unbounded if X is infinite dimensional, so every open set in the weak topology when X is infinite dimensional is unbounded. The open unit ball ($\{x : |x| < 1\}$) is a set that is open in the strong topology but not the weak topology.

Given a set S, the weak sequential closure of S is

$$\{x : \exists x_n \in S, x_n \xrightarrow{w} x\}.$$

If the weak sequential closure of S is equal to S, then S is said to be weakly sequentially closed.

Proposition 11.1 *(1) The weak sequential closure of S is contained in the closure of S with respect to the weak topology.*

(2) If X is infinite dimensional, there are sets that are are weakly sequentially closed, but are not closed in the weak topology.

Proof. 1) If x is not in the weak closure of S, there exists an open set A in the weak topology such that $x \in A$ and $A \cap S = \emptyset$. We can choose A to be the finite intersection of sets of the form $\{x : a < \ell(x) < b\}$, so

$$A = \cap_{i=1}^n \{x : a_i < \ell_i(x) < b_i\}.$$

If x is in the weak sequential closure of S, there exist $x_k \in S$ such that $x_k \xrightarrow{w} x$. For each i, $\ell_i(x_k) \to \ell_i(x)$, so for n large enough, $\ell_i(x_k) \in (a_i, b_i)$. Taking k large enough, $\ell_i(x_k) \in (a_i, b_i)$ for all $i \le n$, hence $x_k \in A$, contradicting $A \cap S = \emptyset$.

2) Let X_k be finite dimensional subsets of X with $\dim X_k = k$. Let $S_k = \{x_{kj}\}$ be a $1/k$-net for $\{|x| = k\}$ in X_k, and let $S = S_2 \cup S_3 \cup \cdots$.

0 is in the closure of S in the weak topology: If A is open and contains 0, then A contains a subset of the form $\widehat{A} = \{x : |\ell_i(x)| < \varepsilon, i = 1, \ldots, n\}$ with $|\ell_i| = 1$. If $k > n$, then the dimension of X_k is greater than the number of linear functionals, and so there exists x_k such that $|x_k| = k$ and $\ell_i(x_k) = 0$. There exists $x_{kj} \in S_k$ within $1/k$ of x_k, so

$$|\ell_i(x_{kj})| = |\ell_i(x_{kj} - x_k)| \le |x_{kj} - x_k| < 1/k.$$

So for $k > 1/\varepsilon$, $x_{kj} \in \widehat{A}$.

S contains no nontrivial weakly convergent sequences: In any ball of radius R, S has only finitely many points. Now use the uniform bounded principle: if $x_n \xrightarrow{w} x$, then $|x_n|$ is bounded; see Corollary 9.3. □

Proposition 11.2 *Suppose K is convex and contained in a Banach space X. If K is closed in the strong topology, then K is closed in the weak topology.*

Proof. Suppose $z \notin K$. We show z is not in the weak closure of K. K is closed in the strong topology, so there exists $B_R(z)$ that is disjoint from K. By the hyperplane separation theorem, there exists a nonzero linear functional ℓ and c such that

$$\ell(u) \leq c \leq \ell(v), \qquad u \in K, v \in B_R(z).$$

If $w \in B_R(0)$, $v = -w + z$, then

$$\ell(-w) = \ell(v - z) = \ell(v) - \ell(z) \geq c - \ell(z).$$

So $\ell(w) \leq \ell(z) - c$. If follows that $|\ell(w)|$ is bounded above for $w \in B_R(0)$, which implies ℓ is bounded.

If $v \in B_R(z)$, then $v = z + x$ with $|x| < R$. Then

$$c \leq \inf_{v \in B_R(z)} \ell(v) = \ell(z) + \inf_{|x|<R} \ell(x) = \ell(z) - R|\ell|.$$

Therefore $\ell(z) > c$. (Before we only had $\ell(z) \geq c$.) So $A = \{x : \ell(x) > c\}$ contains z but no point of K. Therefore A is open in the weak topology, and z is not in the weak closure of K. □

11.1 The Alaoglu theorem

Consider X^*, where X is a Banach space. For $x \in X$, define $L_x : X^* \to \mathbb{R}$ by $L_x(\ell) = \ell(x)$. The weak* topology is the coarsest topology on X^* with respect to which all the L_x with $x \in X$ are continuous.

Theorem 11.3 *(Alaoglu) The closed unit ball B in X^* is compact in the weak* topology.*

There is a connection with the Prohorov theorem of probability theory. Let $X = C(S)$ where S is a compact Hausdorff space. If μ_n is a sequence of probabiity measures, then the μ_n are elements of the closed unit ball in X^*.

The Alaoglu theorem implies there must be a subsequence which converges in the weak* sense.

Proof. If $u \in B$, then $|u(x)| \le |x|$. Let

$$P = \prod_{x \in X} I_x,$$

where $I_x = [-|x|, |x|]$. Map B into P by setting $\varphi(u) = \{u(x)\}$, the function whose x^{th} coordinate is $u(x)$. The topology on P that we use is the coarsest one where all the projections $u \to u(x)$ are continuous. As in the proof of that $\mathcal{S}$ is a basis for the weak topology, we can check that finite intersections of sets of the form $\{u : a < u(x) < b\}$ are a basis for this topology, which is the product topology. The same argument shows that it is a basis for the weak* topology. By Tychonov's theorem, P is compact. So it suffices to show that $\varphi(B)$ is closed.

Let p be in the closure of $\varphi(B)$. We will show $p = \varphi(u)$ for some $u \in B$. If p_x denotes the x^{th} component of p, we want to show $p_x = u(x)$.

Define u on X by $u(x) = p_x$. If $q \in \varphi(B)$, then $q_x + q_y = q_{x+y}$ and $q_{ax} = aq_x$ since $q = \varphi(v)$ for some $v \in X^*$. p is in the weak* closure of $\varphi(B)$, so there exist $q_n \in \varphi(B)$ such that $(q_n)_x$ converges to p_x, $(q_n)_y$ converges to p_y, $(q_n)_{x+y}$ converges to p_{x+y}, and $(q_n)_{kx}$ converges to p_{kx}. Passing to the limit, $(q_n)_x \to p_x$, etc., and we conclude p is linear. Since $p_x \in I_x$, we see the norm of p is bounded by 1. Therefore $p \in \varphi(B)$. □

Corollary 11.4 *Suppose $S \subset X^*$ is closed in the weak* topology. Then S is weak* compact if and only if it is bounded in norm.*

Proof. If S is bounded in norm, then S is contained in a closed ball about the origin. The closed ball is weak* compact, and S is a weak* closed subset, hence compact.

Now suppose S is weak* compact. For each x, the image of S under the map $u \to u(x)$ is compact, since continuous functions map compact sets to compact sets. Therefore, for each x, $\{u(x)\}$ is a bounded set. By the uniform boundedness principle, the collection $\{u\} \subset S$ is bounded in norm. □

12 Locally convex topological spaces

We look at topologies other than those defined in terms of linear functionals.

A locally convex topological (LCT) linear space is a linear space over the reals with a Hausdorff topology satisfying

1) $(x, y) \to x + y$ is a continuous mapping from $X \times X \to \mathbb{R}$.

2) $(k, x) \to kx$ is a continuous mapping from $F \times X \to X$.

3) Every open set containing the origin contains a convex open set containing the origin.

It is an exercise to show that the weak and weak* topologies are locally convex (i.e., satisfy 3)).

Proposition 12.1 *In a LCT linear space,*

(1) if T is open, so are $T - x, kT$, and $-T$.

(2) Every point of an open set T is interior to T.

Proof. (1) $T - x$ is the inverse image of T under the map $y \to x + y$. kT is similar.

(2) Suppose $0 \in T$. Fix $x \in X$. $k \to kx$ is continuous, so $\{k : kx \in T\}$ is open. Since $0 \in T$, then 0 is in this set, and therefore there exists an interval about 0 such that $kx \in T$ if k is in this interval. This is true for all x, and therefore 0 is an interior point. Use translation if the point we are interested in is other than 0. □

12.1 Separation of points

In a LCT space, we can talk about continuous linear functionals, but not bounded linear functionals.

Proposition 12.2 *Continuous linear functionals in a LCT linear space X separate points: if $y \neq z$, there exists ℓ such that $\ell(y) \neq \ell(z)$.*

Proof. Without loss of generality assume $y = 0$. There exists an open set T that contains 0 but not z, since the topology is Hausdorff. We can take T to be convex. $0 \in T$ is interior, so the gauge function p_T is finite. Recall $p_T(u) < 1$ if $u \in T$.

By the hyperplane separation theorem, there exists ℓ such that $\ell(z) = 1$ and $\ell(x) \le p_T(x)$ for all x. Since $\ell(y) = \ell(0)$, then ℓ separates.

It remains to prove that ℓ is continuous.

$H = \{w : \ell(w) < c\}$ is open: if $w \in H$ and $u \in T$, let $r = c - \ell(w)$. Then

$$\ell(w + ru) = \ell(w) + r\ell(u) \le \ell(w) + rp_T(u) < \ell(w) + c - \ell(w) = c,$$

so $w + ru \in H$. Hence H is open.

Making T symmetric about the origin by replacing $T \cap (-T)$, a similar argument shows $\{w : \ell(w) > d\}$ is open. □

Using the extended hyperplane separation theorem, we have

Corollary 12.3 *Let K be a closed convex set in a LCT space, $z \notin K$. There exists a continuous linear functional ℓ such that $\ell(y) \le c$ for $y \in K$ and $\ell(z) > c$.*

12.2 Krein-Milman theorem

Theorem 12.4 *Let K be a nonempty, compact, convex subset of a LCT linear space X. Then*

1) K has at least one extreme point.

2) K is the closure of the convex hull of its extreme points.

Proof. 1) Let $\{E_j\}$ be the collection of all nonempty closed extreme subsets of K. It is nonempty because it contains K. We partially order by inclusion. We show that if we have a totally ordered subcollection, $\cap_j E_j$ is a lower bound, and hence by Zorn's lemma a minimal element. (To be able to apply Zorn's lemma without change, say $E_1 \le E_2$ if $E_1^c \subset E_2^c$ and translate what a maximal element and upper bound means.)

The intersection of any finite totally ordered subcollection $\{E_j\}$ is just the smallest one. Since K is compact, by the finite intersection property, the intersection of any totally ordered subcollection is nonempty. (If $\cap E_j = \emptyset$, then $\{E_j^c\}$ forms an open cover of K, so there is a finite subcover, and then the intersection of those finitely many E_j is empty, a contradiction.) The intersection of closed sets is closed, and it is easy to check that the intersection of extreme sets is extreme.

By Zorn's lemma, there is a minimal element E. We claim E is a single point. If not, there exists a continuous linear functional ℓ that separates 2 of the points of E. Let μ be the maximum value of ℓ on E. Since E is compact, this maximum value is attained. Let $M = \{x \in E : \ell(x) = \mu\}$. $M \neq E$ since ℓ is not constant. ℓ is continuous and E is closed, so M is closed. $\ell^{-1}(\{\mu\})$ is extreme, so M is extreme in E, and since E is extreme in K, M is extreme in K. But this contradicts the fact that E was a minimal extreme subset.

2) Let K_e be the extreme points of K. We'll show that if z is not in the closure of the convex hull, then $z \notin K$. There exists a continuous linear functional ℓ such that $\ell(y) \leq c$ for $y \in \overline{K}_e$ and $\ell(z) > c$. K is compact and ℓ is continuous, so ℓ achieves it maximum on a closed subset E of K. E is extreme, and E must contain an extreme point p. Since $p \in E \subset K_e$, then $\ell(p) \leq c$. Since $\ell(p) = \max_K \ell(x)$, then $\ell(x) \leq \ell(p) \leq c$ for all $x \in K$. Since $\ell(z) > c$, then $z \notin K$. □

12.3 Choquet's theorem

Here is a theorem of Carathéodory.

Theorem 12.5 *Suppose K is a nonempty compact convex subset of a LCT linear space X. Let K_e be the set of extreme points. If $u \in K$, there exists a measure m_u of total mass 1 on $\overline{K}_e$ such that*

$$u = \int_{\overline{K}_e} e \, m_u(de)$$

in the weak sense.

A measure with total mass one is a probability measure, but this theorem has nothing to do with probability.

The equation holding in the weak sense means

$$\ell(u) = \int_{\overline{K}_e} \ell(e)\, m_u(de)$$

for all continuous linear functionals ℓ.

Proof. Let m, M be the minimum and maximum of ℓ on K. K is compact, so these values are achieved. Then $\{x \in K : \ell(x) = m\}$ is an extreme subset of K and similarly with m replaced by M. They each contain extreme points. So if $u \in K$,

$$\min_{p \in K_e} \ell(p) \le \ell(u) \le \max_{p \in K_e} \ell(p). \tag{12.1}$$

If ℓ_1 and ℓ_2 are equal on K_e, then applying the above to $\ell_1 - \ell_2$ shows they are equal on K.

Let L be the class of continuous functions on $\overline{K}_e$ that are the restriction of a continuous linear functional. Fix u. Define r on L by setting

$$r(\ell) = \ell(u).$$

If L contains the constant function 1, then by (12.1) we have $r(\ell) = \ell(u) = 1$. If L does not contain the constant functions, adjoin the constant function $f_0 = 1$ to L and set $r(f_0) = 1$. The set L is a linear subspace of $C(\overline{K}_e)$. Check that r is a positive linear functional on L.

Now use Hahn-Banach to extend r from L to $C(\overline{K}_e)$.

$\overline{K}_e$ is a closed subset of K, hence compact. r is a positive linear functional on $C(\overline{K}_e)$. By the Riesz representation theorem from measure theory, there exists a measure m such that

$$r(f) = \int_{\overline{K}_e} f\, dm.$$

Since $r(f_0) = 1$, then $m(\overline{K}_e) = 1$. □

An example: in $\mathbb{R}^3$, let K be the unit circle in the (x, y) plane together with $\{(1, 0, z) : |z| \le 1\}$. Then $(1, 0, 0) \notin K_e$, so the collection of extreme points is not closed.

Choquet's theorem is an important extension of Carathéodory's theorem in that we can take the integral to be over K_e rather than its closure, provided K is metrizable.

Theorem 12.6 *Let K be a nonempty compact convex subset of a LCT space X that is metrizable. Then if $u \in K$,*

$$u = \int_{K_e} e\, m_u(de)$$

in the weak sense.

13 Examples of convex sets and extreme points

Let Q be a compact Hausdorff space and $X = C(Q)$. Let P be the positive linear functionals on $X = C(Q)$ such that $\ell(1) = 1$. If $r \in Q$, define $e_r \in P$ by $e_r(f) = f(r)$.

Proposition 13.1 *P is convex and the set of extreme points is $\{e_r\}$.*

Proof. Convexity is easy. Suppose e_r were not extreme. Then $e_r = am + (1-a)\ell$, where $m, \ell \in P$ and $a \in (0,1)$. Choose $f \in C(Q)$ such that $f \geq 0$ and $f(r) = 0$. Then

$$e_r(f) = f(r) = 0 = am(f) + (1-a)\ell(f).$$

Since $m(f), \ell(f) \geq 0$ and $a \in (0,1)$, then $\ell(f) = m(f) = 0$.

Now take $f \in C(Q)$ with $f(r) = 0$. Then writing $f = f_+ - f_-$, we have $f_+(r) = 0$ and by the above $\ell(f_+) = 0$ and the same for m. Similarly $f_-(r) = 0$ so the same is true for ℓ and m. Combining, $\ell(f) = 0$ and the same for m. Thus $N_{e_r} \subset N_m, N_\ell$. A non-zero linear functional has codimension 1. Since codim $N_\ell =$ codim N_m, ℓ, m are constant multiples of e_r. Since $\ell(1) = m(1) = 1$, $\ell = e_r$ and the same for m. Therefore e_r is extreme.

Now let ℓ be an extreme element of P. By Riesz representation there exists a unique measure μ such that $\ell(f) = \int f\, d\mu$. Since $\ell(1) = 1$, then $\mu(Q) = 1$. If the support of μ is not a single point, there exist μ_1, μ_2 that are non-zero, not equal, and $\mu = a\mu_1 + (1-a)\mu_2$. Let $\ell_j = \int f\, d\mu_j$. Then $\ell = a\ell_1 + (1-a)\ell_2$; since ℓ is extreme, $\ell_1 = \ell_2$, which implies $\mu_1 = \mu_2$, a contradiction. Hence the support of μ is a single point, and since $\mu(Q)=1$, $\mu(dx)=\delta_r(dx)$ for some r. Therefore $\ell(f)=\int f(x)\,\delta_r(dx)=f(r)$, or $\ell=e_r$. □

Choquet's representation here is

$$\ell = \int e_r\, m(dr).$$

13.1 Convex functions

Let C be the set of convex functions on $[0,1]$ such that $f(0)=0$, $f(1)=1$ and $f\geq 0$. Clearly C is a convex set. Let $e_r(x)$ be the function that is 0 on $[0,r]$, 1 at 1, and linear on $[r,1]$, and let $e_1(x)$ be 0 except 1 at 1.

Proposition 13.2 *$\{e_r\}$ are the extreme points for C.*

Proof. Suppose $e_r = af+(1-a)g$ with $f,g\in C$. Since $f,g\geq 0$, they must both be 0 on $[0,r]$. If they were not equal to e_r on $[r,1]$, one of them, say f, must be larger than e_r at some point $y\in(r,1)$. But $f(r)=0$, $f(1)=1$, and by convexity f must be less than the line connecting $(r,0)$ and $(1,1)$, contradicting $f(y)>e_r(y)$. Therefore $f=g=e_r$ and e_r is extreme.

Now suppose $f\in C$ is extreme. We will show in a minute that we can write

$$f(x) = \int_0^1 e_r(x)\, m(dr) \tag{13.1}$$

for some measure m on $[0,1]$ with total mass 1. As in the proof above for P, the measure m must be supported on a single point, and hence $f=e_r$ for some r.

To prove (13.1), a convex function is differentiable almost everywhere, and its derivative f' is increasing. Let $\overline{m}$ be the Lebesgue-Stieltjes measure associated with the right continuous version of f'. Set $m(dr)=(1-r)\,\overline{m}(dr)$.

If $r > x$, then $e_r(x) = 0$. Then

$$\begin{aligned} f(x) = f(x) - f(0) &= \int_0^x f'(r)\,dr = \int_0^x \int_0^r \overline{m}(ds)\,dr \\ &= \int_0^x \int_x^s dr\,\overline{m}(ds) = \int_0^x (x-s)\,\overline{m}(ds) \\ &= \int_0^x \frac{x-r}{1-r}(1-r)\,\overline{m}(dr) = \int_0^1 e_r(x)\,m(dr). \end{aligned}$$

Since $f(1) = 1$ and $e_r(1) = 1$, we conclude $m([0,1]) = 1$. □

$f = \int_0^1 e_r\,dm$ is the Choquet representation for f.

14 Bounded linear maps

14.1 Boundedness and continuity

A linear map M, which is the same as a linear operator or linear transformation, from a Banach space X to a Banach space U is continuous if $x_n \to x$ implies $Mx_n \to Mx$. M is bounded if there exists a constant c such that $|Mx| \le c|x|$.

The following is proved just as for linear functionals.

Proposition 14.1 *M is continuous if and only if it is bounded.*

If X and U are only normed linear spaces and M is bounded, then M can be extended by continuity to a map from the completion of X to the completion of U.

Define

$$|M| = \sup_{x \neq 0} \frac{|Mx|}{|x|},$$

or what is the same,

$$|M| = \sup_{|x|=1} |Mx|.$$

Proposition 14.2 *$|aM| = |a|\,|M|$, $|M| \geq 0$ and equals 0 if and only if $M = 0$, and $|M + N| \leq |M| + |N|$.*

The proofs are easy.

We wrote $\mathcal{L}(X, U)$ for the set of linear maps from a Banach space X to a Banach space U.

Proposition 14.3 *$\mathcal{L}$ is itself a Banach space.*

The proof is the same as for linear functionals. There we used the completeness of $\mathbb{R}$ or $\mathbb{C}$; here we use the completeness of U.

Proposition 14.4 *N_M is closed.*

Proof. $\{0\}$ is closed, M is continuous, so $N_M = M^{-1}(\{0\})$ is closed. □

Suppose $M : X \to U$. We define the transpose M' as follows. $M' : U^* \to X^*$. If $\ell \in U^*$, $\ell(Mx)$ is a linear functional on X, and we call this linear functional $M'\ell$.

We sometimes use the notation $\ell(u) = (u, \ell)$ and $\xi(x) = (x, \xi)$. With this notation,

$$(Mx, \ell) = \ell(Mx) = M'\ell(x) = (x, M'\ell),$$

which justifies the name adjoint or transpose.

If R is a subspace of U, then $R^\perp$ is the set of bounded linear functionals that vanish on R, called the annihilator of R. $R^\perp \subset U^*$.

If $S \subset X^*$, then $S^\perp$ consists of those vectors that are annihilated by every linear functional in S. $S^\perp \subset X$.

Proposition 14.5 *(1) M' is bounded and $|M'| = |M|$.*

(2) $N_{M'} = R_M^\perp$.

(3) $N_M = R_{M'}^\perp$.

(4) $(M + N)' = M' + N'$.

Proof. (1) $|M'| = \sup_{|\ell|=1} |M'\ell|$ (recall $M'\ell \in X^*$) and

$$|M'\ell| = \sup_{|x|=1} |M'\ell(x)| = \sup_{|x|=1} |\ell(Mx)|.$$

So

$$|M'| = \sup_{|\ell|=1, |x|=1} |\ell(Mx)| = \sup_{|x|=1} |Mx| = |M|.$$

(2) Suppose $\ell \in N_{M'}$. Then $M'\ell = 0$. If $x \in X$, $0 = M'\ell(x) = \ell(Mx)$, or $\ell \in R_M^\perp$. On the other hand if $\ell \in R_M^\perp$, then $(M'\ell)(x) = \ell(Mx) = 0$, hence $M'\ell = 0$, or $\ell \in N_{M'}$.

(3) is similar and (4) is easy. □

14.2 Strong and weak topologies

We define some topologies on $\mathcal{L}(X, U)$.

1) Uniform: this is defined in terms of the norm M.

2) Strong: for each $x \in X$, let $R_x : \mathcal{L} \to U$ be defined by $R_x M = Mx$. The strong topology is the coarsest one where all the R_x are continuous functions.

3) Weak topology: for each $x \in X$ and $\ell \in U^*$, let $S_{x\ell} : \mathcal{L} \to \mathbb{R}$ be defined by $S_{x\ell} M = (Mx, \ell) = \ell(Mx)$. The weak topology is the coarsest one such that all the $S_{x\ell}$ are continuous.

We say M_n is strongly convergent to M if $|M_n x - Mx| \to 0$ for all x. M_n is weakly convergent to M if $M_n x \xrightarrow{w} Mx$ for all x, that is, for all $\ell \in U^*$, $\ell(M_n x) \to \ell(Mx)$.

Proposition 14.6 *Suppose X and U are Banach spaces, $M_n \in \mathcal{L}$, and $|M_n| \le c$ for all n. Suppose $s - \lim M_n x$ exists for a dense subset of x's. Then M_n converges strongly for all x.*

Proof. Let $\varepsilon > 0$, and given x choose x_k in the dense subset such that $|x - x_k| < \varepsilon$. Then $|M_n x - M_n x_k| \le c|x - x_k| \le c\varepsilon$, $M_n x_k$ converges strongly. We conclude $M_n x$ converges strongly. □

14.3 Uniform boundedness principle

Proposition 14.7 *If M_ν is a collection of elements of $\mathcal{L}(X, U)$, where X, U are Banach spaces, and $\sup_\nu |M_\nu x|$ is finite for all x, then there exists c such that $|M_\nu| \leq c$ for all c.*

The proof is nearly identical to the linear functional case.

Theorem 14.8 *Suppose X and U are Banach spaces such that for all $x \in X$ and all $\ell \in U^*$, $\sup_\nu |(M_\nu x, \ell)| < \infty$. Then there exists c such that $|M_\nu| \leq c$ for all ν.*

Proof. Fix x. Let $u_\nu = M_\nu x$. Then $\sup_\nu |\ell(u_\nu)| < \infty$. By the corollary to the uniform boundedness principle for linear functionals,

$$\sup_\nu |M_\nu x| = \sup_\nu |u_\nu| < \infty.$$

Now apply the preceding proposition. □

14.4 Composition of maps

Proposition 14.9 *Suppose X, U, W are Banach spaces, M is a linear map from X to U, and N is a linear map from U to W. Then*

1) $|NM| \leq |N|\,|M|$.

(2) $(NM)' = M'N'$.

Proof. (1)

$$|NMx| \leq |N|\,|Mx| \leq |N|\,|M|\,|x|.$$

(2)

$$(NMx, m) = (Mx, N'm) = (x, M'N'm).$$

□

14.5 Open mapping principle

Theorem 14.10 *Suppose M is a bounded linear map from a Banach space X onto a Banach space U. Then there exists d such that $B_d(0) \subset M(B_1(0))$.*

The "onto" condition is important.

Proof. M is onto, so $\cup_n M(B_n(0)) = U$. By the Baire category theorem, at least one of $M(B_n(0))$ is dense in an open set in U. If $V = M(B_n(0))$ and the open set contains a ball centered at y_0, then $V - y_0$ is dense in some open ball around the origin. Since M is onto, there exists x_0 such that $Mx_0 = y_0$. So $M(B_n(0) - x_0)$ is dense in an open ball about 0. Since $B_n(0) - x_0 \subset B_{n+|x_0|}(0)$, then $M(B_{n+|x_0|}(0))$ is dense in a ball about 0. By homogeneity, $M(B_1(0))$ is dense in $B_r(0)$ for some r, and then $M(B_c(0))$ is dense in $B_{cr}(0)$ for all $c > 0$.

We show $M(B_2(0))$ contains $B_r(0)$. Let $u \in B_r(0)$.

There exists x_1 such that $|u - Mx_1| < r/2$ and $|x_1| < 1$. Because $u - Mx_1 \in B_{r/2}(0)$, there exists x_2 such that

$$|(u - Mx_1) - Mx_2| < r/4, \qquad |x_2| < 1/2.$$

We continue with this construction.

Since $\sum_{i=m}^n |x_i| \le \sum_{i=m}^n 2^{-i+1}$, then

$$\Big|\sum_{i=m}^n x_i\Big| \to 0$$

as $m, n \to \infty$. Since X is complete, the sequence $\sum_{i=1}^n x_i$ converges, say to x. We have $|x| \le \sum |x_i| < 2$. M is bounded, so $M(\sum_{i=1}^n x_i) \to Mx$. Because $|u - M\sum_{i=1}^n x_i| \le r/2^n \to 0$, we have $Mx = u$. □

Corollary 14.11 *M maps open sets onto open sets.*

Corollary 14.12 *If M is one-to-one and onto and bounded, then M^{-1} is bounded.*

Proof. If $u \in U$ with $|u| = d/2$, there exists x with $|x| < 1$ and $Mx = u$. By homogeneity, if $u \in U$, there exists $x \in X$ with $Mx = u$ and $|x| < 2|u|/|d|$. So $x = M^{-1}u$ and $|M^{-1}| < 2/d$. □

A map $M : X \to U$ is closed if whenever $x_n \to x$ and $MX_n \to u$, then $Mx = u$. This is equivalent to the graph $\{x, Mx\}$ being a closed set.

If M is continuous, it is closed If D is the differentiation operator on the set of differentiable functions on $[0, 1]$, then D is closed, but not continuous.

Theorem 14.13 *(Closed graph theorem) If X and U are Banach spaces and M a closed linear map, then M is continuous.*

Proof. Let $G = \{g = (x, Mx)\}$ with norm $|g| = |x| + |Mx|$. It is easy to see that $|g|$ is a norm, and G is complete. Define $P : G \to X$ by $P(x, Mx) = x$, so that P is a projection onto the first coordinate.

$|Pg| = |x| \leq |x| + |Mx| = |g|$, so P is bounded with norm less than or equal to 1. P is linear and 1-1, and onto, so P^{-1} is bounded, i.e., there exists c such that $c|Pg| \geq |g|$. So $(c-1)|x| \geq |Mx|$, which proves M is bounded. □

Corollary 14.14 *Suppose X has two norms such that if $|x_n - x|_1 \to 0$ and $|x_n - y|_2 \to 0$, then $x = y$. Suppose X is complete with respect to both norms. Then the norms are equivalent.*

Proof. Let $X_1 = (X, |\cdot|_1)$ and similarly X_2. Let $I : X_1 \to X_2$. The hypothesis is equivalent to I being closed. Therefore I and I^{-1} are bounded. □

Corollary 14.15 *Suppose X is a Banach space and $X = Y \oplus Z$, where Y and Z are closed linear subspaces. If $x = y + z$, define $P_Y x = y$ and $P_Z x = z$. then*

(1) P_Y, P_Z are linear operators, $P_Y^2 = P_Y$, $P_Z^2 = P_Z$, $P_Y P_Z = 0$.

(2) P_Y, P_Z are continuous.

P_Y and P_Z are projections.

Proof. (1) is easy. (2) Since Y, Z are closed and the decomposition is unique, then the graphs of P_Y and P_Z are closed. To see this, if $x_n = y_n + z_n$, $x_n \to x$, $y_n \to y'$,and $z_n \to z'$, then $y' \in Y, z' \in Z$, and $x = y' + z'$. The decomposition is unique, so $y' = P_Y x, z' = P_Z x$.

By the closed graph theorem, P_Y, P_Z are continuous. $\square$

15 Distributions

(This is from Appendix B of Lax.)

15.1 Definitions and examples

Let C_0^∞ be the C^∞ functions on $\mathbb{R}^n$ with compact support. We will work with complex-valued functions.

Let $D_i u = \frac{\partial u}{\partial x_i}$. For $\alpha = (\alpha_1, \dots, \alpha_n)$, let $|\alpha| = \alpha_1 + \cdots + \alpha_n$, and write D^α for $D_1^{\alpha_1} \cdots D_n^{\alpha_n}$.

We say $u_k \to u$ if there exists a compact subset K such that $\operatorname{supp}(u_k) \subset K$ for all k (here $\operatorname{supp}(u)$ is the support of u) and for each $\alpha \in \mathbb{N}^n$, $D^\alpha u_k$ converges uniformly to $D^\alpha u$.

A distribution is an element of the dual of C_0^∞, that is, ℓ is a complex-valued linear functional such that $\ell(u_k) \to \ell(u)$ whenever $u_k \to u$.

For an integer N, let

$$|u|_N = \max_{|\alpha| \le N} |D^\alpha u|.$$

Proposition 15.1 *Let ℓ be a distribution, K a compact set. There exist N and c depending on K such that if $u \in C_0^\infty$ has support in K, then $|\ell(u)| \le c|u|_N$.*

Proof. If not, then for each n there exists u_n with support in K such that $\ell(u_n) = 1$ and $|u|_n \le 1/n$. Therefore $u_n \to 0$. But $\ell(u_n) = 1$ while $\ell(0) = 0$. $\square$

Let $\mathcal{D}$ be the set of distributions. We embed C_0^∞ in $\mathcal{D}$ as follows: if $v \in C_0^\infty$, define $\ell_v(u) = \int uv\,dx = (u, v)$.

We will use the notation $\ell(u) = (u, \ell)$.

Here are some examples of distributions.

1) If v is a continuous function, define $\ell(u) = \int uv\,dx$.

2) Dirac delta function: $\delta(u) = u(0)$.

3) If v is integrable and α is fixed, define $\ell(u) = \int (D^\alpha u)v\,dx$.

4) Define

$$\ell(u) = PV \int \frac{u(x)}{x}\,dx = \lim_{\varepsilon > 0} \int_{|x|>\varepsilon} \frac{u(x)}{x}\,dx.$$

If D is open, let C_D^∞ be the C^∞ functions with support contained in D. We can talk about distributions being the duals of elements of C_D^∞.

Suppose U is a continuous linear map from C_0^∞ into C_0^∞. Define $T\ell$ by

$$(v, T\ell) = (Uv, \ell).$$

It is an exercise to show $T\ell$ is a distribution. It is easy to check that $T\ell_v = \ell_{Uv}$. Therefore T acts an extension of U'.

Here are some examples.

1) Let U be multiplication by a C^∞ function t. Here T is an extension of U.

2) Let $U = -D_i$, differentiation. T is an extension of $-U$.

3) Translation: $(U_a u)(x) = u(x + a)$. T_a is an extension of U_{-a}.

4) Reflection: $Ru(x) = u(-x)$, and T is an extension of R.

5) Convolution: if t is continuous with compact support, define

$$Uu = (t * u)(x) = \int t(y)u(x - y)\,dy.$$

T is convolution with respect to Rt.

6) Let $\phi : \mathbb{R}^n \to \mathbb{R}^n$ be C^∞ and map compact sets to compact sets. Suppose $\psi = \phi^{-1}$ exists and has the same properties. Define $Uu(x) = u(\phi(x))$. Here U' is given by $U'v(y) = v(\psi(y))J(y)$.

Note that one cannot, in general, define the product of two distributions, or quantities like $\delta(x^2)$.

15.2 Local properties

Let G be open. ℓ is zero on G if $\ell(u) = 0$ for all C_0^∞ functions u whose support is contained in G.

Lemma 15.2 *If ℓ is zero on G_1 and G_2, then ℓ is zero on $G_1 \cup G_2$.*

Proof. This is just the usual partition of unity proof. Given u with support in $G_1 \cup G_2$, we will write $u = u_1 + u_2$ with $\operatorname{supp}(u_1) \subset G_1$ and $\operatorname{supp}(u_2) \subset G_2$. Since $\ell(u) = \ell(u_1) + \ell(u_2) = 0$, that will do it.

Fix $x \in \operatorname{supp}(u)$. Since G_1, G_2 are open, we can find $h = h_x$ such that h is non-negative, $h(x) > 0$, h is C^∞, and the support of h is contained either in G_1 or in G_2. The set $B_x = \{y : h_x(y) > 0\}$ is open and contains x. By compactness we can cover $\operatorname{supp} u$ by finitely many of such balls. Look at the corresponding finite collection of h's.

Let h_1 be the sum of those h's in the finite collection whose support is in G_1 and define h_2 similarly. Then let

$$u_1 = \frac{h_1}{h_1 + h_2} u, \qquad u_2 = \frac{h_2}{h_1 + h_2} u.$$

□

If we have an arbitrary collection of open sets and ℓ is zero on each one, and $\operatorname{supp} u$ is contained in their union, there by compactness there exist finitely many of them that cover $\operatorname{supp} u$, and by the above $\ell(u) = 0$.

The union of all open sets on which ℓ is zero is an open set on which ℓ is zero. Its complement is called the support of ℓ.

Example: for the Dirac delta function, the support is $\{0\}$. Note that the support of $D^\alpha \delta$ is also $\{0\}$.

Lemma 15.3 *If ℓ is a distribution and* $\operatorname{supp} \ell = \{0\}$*, then there exists N such that $D^\alpha u(0) = 0$ for $|\alpha| \le N$ implies $\ell(u) = 0$.*

Proof. Let f be 0 on $|x| < 1$ and 1 on $|x| > 2$, and in C^∞. Let $v = (1-f)u$. Since $fu = 0$ for $|x| < 1$, then $\ell(fu) = 0$ and

$$\ell(v) = \ell(u) - \ell(fu) = \ell(u).$$

Thus is suffices to consider u such that $\operatorname{supp} u \subset B_3(0)$.

There must exist N such that $|\ell(u)| \leq c|u|_N$. Define $f_k(x) = f(kx)$, $u_k = f_k u$. Then $u_k = u$ if $|x| > 2/k$.

If $|x| < 2/k$, since u and $D^\beta u$ are 0 at 0, Taylor's theorem gives

$$|D^\beta u(x)| \leq c|x|^{N+1-|\beta|} i \leq c|k|^{|\beta|-1-N}$$

if $|\beta| \leq N$. Calculus (product rule) shows that

$$|D^\alpha u_k(x)| \leq ck^{|\alpha|-N-1}$$

if $|x| < 2/k$.

Since $u_k = u$ if $|x| > 2/k$, then $u_k \to u$ in C^N norm, so $\ell(u_k) \to \ell(u)$. u_k is zero in a ball about the origin, so $\ell(u_k) = 0$, which implies $\ell(u) = 0$. □

Theorem 15.4 *If* $\operatorname{supp} \ell = \{0\}$, *then*

$$\ell = \sum_{|\alpha| \leq N} c_\alpha D^\alpha \delta.$$

Proof. If u_1, u_2 and all the derivatives up to order N agree at 0, then $\ell(u_1 - u_2) = 0$, or $\ell(u_1) = \ell(u_2)$. So $\ell(u)$ depends only on the values of u and its derivatives up to order N. So

$$\ell(u) = \sum_{|\alpha| \leq N} c_\alpha D^\alpha u \Big|_{x=0}. \tag{15.1}$$

□

To elaborate on the last line of the proof, consider the case where the dimension is 1. We take functions p_i that are equal to x^i near 0 but are in

C_0^∞. Choose the c_α so that (15.1) holds when u is equal to p_i. Then it will hold for all u.

We will need the Sobolev inequality. First suppose $1 \le p < n$ and $p^* = \frac{np}{n-p}$ so that

$$\frac{1}{p^*} = \frac{1}{p} - \frac{1}{n}$$

and $p^* > p$.

Theorem 15.5 *Suppose $1 \le p < n$. There exists C such that*

$$\|u\|_{L^{p^*}} \le c\|Du\|_{L^p}$$

for all $u \in C^1$ with compact support.

Letting $u \equiv 1$ shows why compact support is necessary, but C does not depend on the size of the support.

Proof. 1) Suppose $p = 1$. Write

$$u(x) = \int_{-\infty}^{x_i} u_i(x_1, \ldots, x_{i-1}, y_i, x_{i+1}, \ldots, x_n)\, dy_1.$$

Then

$$|u(x)| \le \int_{-\infty}^{\infty} |Du(x_1, \ldots, y_i, \ldots, x_n)|\, dy_i.$$

From here on, all integrals are over $\mathbb{R}$. Taking a product, we have

$$|u(x)|^{\frac{n}{n-1}} \le \prod_{i=1}^n \Big(\int |Du(\ldots, y_i, \ldots)|\, dy_i \Big)^{1/(n-1)}.$$

Then

$$\begin{aligned}
\int |u(x)|^{n/(n-1)}\, dx_1 &\le \int \prod_{i=1}^n \Big(\int |Du|\, dy_i \Big)^{1/(n-1)} dx_1 \\
&= \Big(\int |Du|\, dy_1 \Big)^{1/(n-1)} \int \prod_{i=2}^n \Big(\int |Du|\, dy_i \Big)^{1/(n-1)} dx_1 \\
&\le \Big(\int |Du|\, dy_1 \Big)^{1/(n-1)} \Big(\prod_{i=2}^n \int \int |Du|\, dx_1\, dy_i \Big)^{1/(n-1)},
\end{aligned}$$

where the last inequality is from the generalized Hölder inequality.

Integrating with respect to x_2,

$$\int\int |u|^{n/(n-1)}\,dx_1\,dx_2 \le \Big(\int\int |Du|\,dx_1\,dy_2\Big)^{1/(n-1)} \int \prod_{i\neq 2} I_i^{1/(n-1)}\,dx_i,$$

where

$$I_1 = \int |Du|\,dy_1$$

and

$$I_i = \int\int |Du|\,dx_1\,dy_i, \qquad i = 3, \dots, n.$$

Then

$$\begin{aligned}\int\int &|u|^{n/(n-1)}\,dx_1\,dx_2\\ &\le \Big(\int\int |Du|\,dx_1\,dy_2\Big)^{1/(n-1)}\Big(\int\int |Du|\,dy_1\,dx_2\Big)^{1/(n-1)}\\ &\qquad \times \prod_{i=3}^{n}\Big(\int\int\int |Du|\,dx_1\,dx_2\,dy_i\Big)^{1/(n-1)}.\end{aligned}$$

We continue, and after integrating with respect to dx_n, we obtain the desired inequality.

2) If $p > 1$, apply the above case to $v = |u|^\gamma$, where

$$\gamma = \frac{p(n-1)}{n-p} > 1.$$

We obtain

$$\begin{aligned}\Big(\int |u|^{\gamma n/(n-1)}\Big)^{(n-1)/n} &\le \int |D|u|^\gamma|\,dx = \gamma \int |u|^{\gamma-1}|Du|\,dx\\ &\le c\Big(\int |u|^{(\gamma-1)p/(p-1)}\Big)^{(p-1)/p}\Big(\int |Du|^p\Big)^{1/p}\end{aligned}$$

by Hölder's inequality. Dividing gives the general case. □

Theorem 15.6 *If $k < n/p$ and*

$$\frac{1}{q} = \frac{1}{p} - \frac{k}{n},$$

then

$$\|u\|_{L^q} \le c\|u\|_{W^{k,p}}.$$

Proof. Use the above theorem and induction. □

Proposition 15.7 *If ℓ has compact support, there exist L and continuous functions g_α such that*

$$\ell = \sum_{|\alpha| \le L} D^\alpha g_\alpha.$$

An example: The delta function is the derivative of h, where h is 0 for $x < 0$ and 1 for $x \ge 0$. h is the derivative of g, where g is 0 for $x < 0$ and $g(x) = x$ for $x \ge 0$. Therefore $\delta = D^2 g$.

Proof. Let $h \in C_0^\infty$ and equal to 1 on the support of ℓ. Then $\ell((1-h)u) = 0$, or $\ell(u) = \ell(hu)$. Therefore there exists N and c such that

$$|\ell(hu)| \le c|hu|_N \le c|u|_N.$$

Hence

$$|\ell(u)| \le c|u|_N.$$

Let

$$\|u\|_M = \Big(\sum_{|\alpha| \le M} \int |D^\alpha u|^2 \, dx \Big)^{1/2},$$

and let H_M be the completion of C_0^∞ with respect to this norm. This is a Hilbert space.

By the Sobolev inequality,

$$|u|_N \le c\|u\|_M, \qquad N < M - \frac{n}{2}.$$

So any Cauchy sequence in H_M is also a Cauchy sequence with respect to the C^N norm. This shows H_M can be embedded in C^N.

Since $|\ell(u)| \le c|u|_N \le c\|u\|_M$, then ℓ is a bounded linear functional in H_M. By the Riesz-Frechet representation theorem, there exists $g \in H_M$ such that

$$\ell(u) = (u, g)_M = \sum_{|\alpha| \le M} (D^\alpha u, D^\alpha g).$$

This is equal to

$$\sum (-1)^{|\alpha|} (u, D^{2\alpha} g),$$

using the example of transformations where T is differentiation. So

$$\ell = \sum (-1)^{|\alpha|} D^{2\alpha} g.$$

Since $g \in H_M$, then $g \in C^N$. Now take $L = 2M - N$. □

Lemma 15.8 *Suppose b is a C_0^∞ function with support contained in $B_1(0)$ and $\int b(x)\,dx = 1$. Let $b_k(x) = k^n b(kx)$. Then $\ell_k = b_k * \ell$ converges to ℓ in the sense of distributions.*

Proof. We have

$$\ell_k(u) = (b_k * \ell)(u) = (b_k * \ell, u) = (\ell, Rb_k * u).$$

Rb_k is an approximation to the δ function, so

$$D^\alpha (Rb_k * u) = Rb_k * D^\alpha u \to D^\alpha u.$$

Hence $Rb_k * u$ converges to u in the C_0^∞ sense, and this implies $\ell_k(u) = (\ell, Rb_k * u) \to (\ell, u) = \ell(u)$. □

Proposition 15.9 *Suppose g is continuous and $D_j g$ is continuous, where $D_j g$ is the partial derivative of g in the distribution sense. Then $D_j g$ is also the partial derivative of g in the classical sense.*

Proof. Let $g_k = b_k * g$. Then $g_k \to g$ in the sense of distributions. $D_j g_k = b_k * D_j g$, so $g_k \to g$ and $D_j g_k \to D_j g$ uniformly on compact subsets, since $g, D_j g$ are continuous. We have, if $b - a$ is a multiple of the unit vector in the x_j direction,

$$g_k(b) - g_k(a) = \int_a^b D_j g_k \, dx_j.$$

Passing to the limit,

$$g(b) - g(a) = \int_a^b D_j g \, dx_j.$$

□

A distribution ℓ is positive if $\ell(u) \geq 0$ whenever u is non-negative.

Examples: ℓ is given by a non-negative measure, $\ell(u) = \int u \, dm$.

Let $|u| = \sup |u(x)|$.

Lemma 15.10 *Suppose ℓ is a positive distribution and K is compact. There exists c depending on K such that $|\ell(u)| \leq c|u|$ for all u supported in K.*

Proof. Let $p \in C_0^\infty$ with $p \geq 0$ and $p = 1$ on K. Then

$$|u(x)| \leq |u| p(x).$$

So $\ell(u) \leq |u| \ell(p)$. Similarly, $-u \leq |u| p(x)$, so $-\ell(u) \leq |u| \ell(p)$. Now let $c = \ell(p)$. □

Proposition 15.11 *Every positive distribution is a measure.*

Proof. Extend ℓ to all continuous functions with compact support. Then use Riesz representation. □

15.3 Fourier transforms

Let $\mathcal{S}$ be the class of complex-valued C^∞ functions u such that $|x^\beta D^\alpha u(x)| \to 0$ as $|x| \to \infty$ for all multi-indexes α and all positive integers β. $\mathcal{S}$ is called the Schwartz class. An example of an element of the Schwartz class is $e^{-|x|^2}$.

Define $|u|_{\beta,\alpha} = \sup_x |x|^\beta |D^\alpha u(x)|$.

We say $u_n \in \mathcal{S}$ converges to u if $|u_n - u|_{\beta,\alpha} \to 0$ for all β, α.

We can find a metric for $\mathcal{S}$: define

$$d(u,v) = \sum_{\alpha,\beta} \frac{1}{2^{\beta+|\alpha|}} \frac{|u-v|_{\beta,\alpha}}{1+|u-v|_{\beta,\alpha}}.$$

The dual of $\mathcal{S}$ is the set of continuous linear functionals on $\mathcal{S}$. Here continuity means that if $u_n \to u$ in the sense of the Schwartz class, then $\ell(u_n) \to \ell(u)$.

Since $C_0^\infty \subset \mathcal{S}$, then $\mathcal{S}^* \subset \mathcal{D}$. This means that every element of $\mathcal{S}^*$ is a distribution. Elements of $\mathcal{S}^*$ are called tempered distributions.

Any distribution of compact support is a tempered distribution. So is $\ell(u) = \int vu\,dx$ if v grows slower than some power of $|x|$ as $|x| \to \infty$.

For $u \in \mathcal{S}$, define the Fourier transform $Fu = \widehat{u}$ by

$$\widehat{u}(\xi) = (2\pi)^{-n/2} \int u(x) e^{ix\cdot\xi}\,dx.$$

Theorem 15.12 *F maps $\mathcal{S}$ into $\mathcal{S}$ continuously.*

Proof. For elements of $\mathcal{S}$, $D^\alpha Fu = F((ix)^\alpha)u)$. If $u \in \mathcal{S}$, $|x^\alpha u| \to 0$ faster than any power of $|x|$, so $x^\alpha u \in L^1$. This implies $D^\alpha F$ is a continuous function. Therefore $Fu \in C^\infty$.

By integration by parts, $\xi^\beta D^\alpha Fu = i^{\alpha+\beta} F(D^\beta(x^\alpha u))$. By the product rule, $D^\beta(x^\alpha u)$ is in L^1. So $\xi^\beta D^\alpha Fu$ is continuous and bounded. Therefore $F \in \mathcal{S}$. □

Proposition 15.13 *For $u \in \mathcal{S}$:*
(1) If $T_a u(x) = u(x-a)$, then $FT_a u = e^{ia\cdot\xi} Fu$ and $T_a Fu = F(e^{-ia\cdot x}u)$.

(2) $F(iD_j u) = \xi_j Fu$ *and* $D_j Fu = F(ix_j u)$.

(3) If A *is a rotation about the origin and* R *is reflection about the origin,* $FR = RF$ *and* $FA = AF$.

(4) $F(u * v) = (2\pi)^{n/2}(Fu)(Fv)$.

Because

$$Fu(\xi) = c\int e^{ix\cdot\xi}u(x)\,dx,$$

$$\begin{aligned}(Fu, v) &= c\int\int e^{ix\cdot\xi}u(x)v(\xi)\,dx\,d\xi\\ &= \int (Fu)(x)v(x)\,dx\\ &= (u, Fv).\end{aligned}$$

So $F^* = F$.

If ℓ is a tempered distribution, define $F\ell$ by

$$(v, F\ell) = (Fv, \ell)$$

for all $v \in \mathcal{S}$.

It then follows that the proposition above holds for $\ell \in sS^*$.

Proposition 15.14 *Let* $\ell \equiv 1$. *Then* $F\ell = (2\pi)^{n/2}\delta$.

$\ell \equiv 1$ means $\ell = \ell_1$, or $\ell(u) = \int u \cdot 1\,dx = \int u(x)\,dx$.

Proof. Write d for $F\ell$.

$$x_j d = x_j F1 = F(iD_j 1) = F(0) = 0.$$

Since $x_j d = 0$, the support of d is contained in $\{x_j = 0\}$. This is true for each j, so the support of d is $\{0\}$. By a previous proposition

$$d = \sum_{|\alpha|\le N} c_\alpha D^\alpha \delta$$

for some N and c.

For any multi-index γ with $|\gamma| > 0$, $x^\gamma d = 0$. But $x^\gamma d = \sum_{|\alpha|\le N} c_\alpha x^\gamma D^\alpha \delta$. We claim

$$x^\gamma D^\alpha \delta = \begin{cases} 0 & \text{if } |\alpha| < |\gamma| \\ 0 & \text{if } |\alpha| = |\gamma|, \alpha \ne \gamma \\ (-1)^{|\alpha|}\alpha!\delta & \text{if } \alpha = \gamma \end{cases}$$

To show this, if $u \in C_0^\infty$,

$$(u, x^\gamma D^\alpha \delta) = (x^\gamma u, D^\alpha \delta) = (-1)^{|\alpha|}(D^\alpha(x^\gamma u), \delta) = (-1)^{|\alpha|} D^\alpha (x^\gamma u)(0).$$

Checking the three cases proves the claim.

Suppose there exists α with $|\alpha| = N > 0$ and $c_\alpha \ne 0$. But

$$0 = x^\alpha d = c_\alpha x^\alpha D^\alpha \delta = c_\alpha (-1)^{|\alpha|}\alpha!\delta,$$

or $c_\alpha = 0$. Therefore $d = c_0\delta$.

To calculate c_0, let $\ell \equiv 1$ and $v = e^{-|x|^2/2}$ in $(Fv, \ell) = (v, F\ell)$. So $F\ell = d = c_0\delta$ and $Fv = e^{-|\xi|^2/2}$. Then

$$(2\pi)^{n/2} = (e^{-|\xi|^2/2}, 1) = (e^{-x^2/2}, c_0\delta) = c_0.$$

□

Theorem 15.15 *(1) F is an invertible map from $\mathcal{S}$ into $\mathcal{S}$ and*

$$u(x) = (2\pi)^{-n/2} \int \widehat{u}(\xi) e^{-i\xi\cdot x}\, d\xi.$$

(2) F is an invertible map from $\mathcal{S}^*$ into $\mathcal{S}^*$ and $F^{-1} = FR$.

Proof. (1) $Fe^{-ia\cdot\xi} = T_a F$. So $F(e^{-ia\cdot\xi}\ell) = T_a F\ell$. Take $\ell \equiv 1$ to get

$$Fe^{-ia\cdot\xi} = (2\pi)^{n/2}\delta(x-a).$$

Now let $\ell = e^{-ia\cdot\xi}$ and use $(Fv, \ell) = (v, F\ell)$ to get

$$\int \widehat{v}(\xi) e^{-ia\cdot\xi}\, d\xi = (\widehat{v}, e^{-ia\cdot\xi}) = (2\pi)^{n/2}(v, \delta(x-a)) = (2\pi)^{n/2} v(a).$$

(2) From (1),

$$u(x) = (2\pi)^{-n/2} \int \widehat{u}(-\xi) e^{ix\cdot\xi}\, d\xi,$$

so $F^{-1} = FR$, and hence $FRF = I$. Then

$$(v, \ell) = (FRFv, \ell) = (RFv, F\ell) = (Fv, RF\ell) = (v, FRF\ell).$$

thus $FRF\ell = \ell$. The inverse of F exists because RF is a right inverse for F, FR is a left inverse, and F and R commute. Therefore $RF\ell = F^{-1}\ell$. □

Theorem 15.16 *If $u \in L^2$, then $\widehat{u} \in L^2$ and $\|\widehat{u}\|_{L^2} = \|u\|_{L^2}$.*

Proof. We will prove this for $u \in \mathcal{S}$ and then approximate functions in L^2 by functions in $\mathcal{S}$. Now

$$\overline{Fu} = (2\pi)^{-n/2} \int \overline{u} e^{-i\xi\cdot x}\, dx = RF\overline{u}.$$

Letting $v = \overline{Fu}$ in $(Fu, v) = (u, Fv)$ yields

$$(Fu, \overline{Fu}) = (u, F(RF\overline{u})) = (u, \overline{u}),$$

using $FRF = F$. □

16 Banach algebras

16.1 Normed algebras

An algebra is a linear space over $+$ and a ring over $\cdot$. We assume there is an identity for the multiplication, which we call I. Our algebras will be over the scalar field $\mathbb{C}$; the reasons will be very apparent shortly.

An algebra is a normed algebra if the linear space is normed and $|NM| \leq |N|\,|M|$ and $|I| = 1$. If the normed algebra is complete, it is called a Banach algebra.

One example is to let $\mathcal{L} = \mathcal{L}(X, X)$, the set of linear maps from X into X. Another is to let $\mathcal{L}$ be the collection of bounded continuous functions on some set.

An element M of $\mathcal{L}$ is invertible if there exists $N \in \mathcal{L}$ such that $NM = MN = I$.

M has a left inverse A if $AM = I$ and a right inverse B if $MB = I$. If it has both, then $B = AMB = A$, and so M is invertible.

Proposition 16.1 *(1) If M and K are invertible, then*

$$(MK)^{-1} = K^{-1}M^{-1}.$$

(2) If M and K commute and MK is invertible, then M and K are invertible.

Proof. (1) is easy. For (2), let $N = (MK)^{-1}$. Then $MKN = I$, so KN is a right inverse for M. Also, $I = NMK = NKM$, so NK is a left inverse for M. Since M has a left and right inverse, it is invertible. The argument for K is similar. □

Proposition 16.2 *If K is invertible, then so is $L = K - A$ provided $|A| < 1/|K^{-1}|$.*

Proof. First we suppose $K = I$. If $|B| < 1$, then

$$\Big|\sum_m^n B^i\Big| \le \sum_m^n |B^i| \le \sum_m^n |B|^i$$

is a Cauchy sequence, so $S = \sum_i B^i$ converges.We see $BS = \sum_{i=1}^\infty B^i = S-I$, so $(I - B)S = I$. Similarly $S(I - B) = I$.

For the general case, write $K - A = K(I - K^{-1}A)$, and let $B = K^{-1}A$. Then $|B| \le |K^{-1}|\,|A| < 1$, and

$$(K - A)^{-1} = (I - K^{-1}A)^{-1}K^{-1}.$$

□

The resolvent set of M, $\rho(M)$, is the set of $\lambda \in \mathbb{C}$ such that $\lambda I - M$ is invertible. The spectrum of M, $\sigma(M)$, is the set of λ for which $\lambda I - M$ is not invertible. We sometimes write $\lambda - M$ for $\lambda I - M$.

Let $f : G \to X$, where $G \subset \mathbb{C}$. $f(z)$ is strongly analytic if

$$\lim_{h\to 0} \frac{f(z+h) - f(z)}{h}$$

exists in the norm topology for all $z \in G$. One can check that most of complex analysis can be extended to strongly analytic functions.

Proposition 16.3 *(1) $\rho(M)$ is open in $\mathbb{C}$.*

(2) $(z - M)^{-1}$ is an analytic function of z in $\rho(M)$.

Proof. If $\lambda \in \rho(M)$, letting $K = \lambda I - M$ and $A = -hI$, $K - A = (\lambda + h) - M$ is invertible if h is small. So $\lambda + h \in \rho(M)$.

For (2),

$$((\lambda + h) - M)^{-1} = \sum (\lambda - M)^{n-1} h^n$$

for h small. So the resolvent can be expanded in a power series in h which is valid if $|h| < |(\lambda - M)^{-1}|^{-1}$. □

We write

$$|\sigma(M)| = \sup_{\lambda \in \sigma(M)} |\lambda|,$$

and call this the spectral radius of M.

Theorem 16.4 *(1) $\sigma(M)$ is closed, bounded, and nonempty.*

(2) $|\sigma(M)| = \lim_{k\to\infty} |M^k|^{1/k}$.

Proof. (1) $\rho(M)$ is open, so $\sigma(M)$ is closed.

$$(zI - M)^{-1} = z^{-1}(I - Mz^{-1})^{-1} = \sum_{n=0}^{\infty} M^n z^{-n-1}$$

converges if $|z^{-1}M| < 1$, or equivalently, $|z| > |M|$. Therefore, if $|z| > |M|$, then $z \in \rho(M)$. Hence the spectrum is contained in $B_{|M|}(0)$.

$$(z - M)^{-1} = \sum_{n=0}^{\infty} M^n z^{-n-1}$$

is a Laurent series. If $\sigma(M) = \emptyset$, then $(z - M)^{-1}$ would be everywhere analytic. So if $c > |M|$ and C is the circle $|z| = c$,

$$\frac{1}{2\pi i} \int_C (z - M)^{-1}\, dz = 0.$$

But integrating $\sum M^n z^{-n-1}$ term by term over the curve C, all the terms are zero except the $n = 0$ term, where we get

$$\frac{1}{2\pi i} \int_C \frac{1}{z}\, dz = I,$$

a contradiction. Therefore $\sigma(M)$ is nonempty.

(2) Fix k for the moment. If we write $n = kq + r$,

$$\Big|\sum_{n=0}^{\infty} \frac{M^n}{z^{n+1}}\Big| \le \sum \frac{|M^n|}{|z|^{n+1}} \le \sum_{n=0}^{k-1} \frac{|M|^r}{|z|^{r+1}} \sum_q \Big(\frac{|M^k|}{|z|^k}\Big)^q.$$

So $\sum M^n |z|^{-n-1}$ converges absolutely if $|M^k|/|z|^k < 1$, or if $|z| > |M^k|^{1/k}$.

If $|z| > |M^k|^{1/k}$, then $z \in \rho(M)$. Hence if $\lambda \in \sigma(M)$, then $|\lambda| \le |M^k|^{1/k}$. This is true for all k, so $|\sigma(M)| \le \liminf_{k\to\infty} |M^k|^{1/k}$.

Let C be a circle enclosing $\sigma(M)$, namely the one about the origin with radius $|\sigma(M)| + \delta$. Using $(z - M)^{-1} = \sum M^n z^{-n-1}$,

$$\frac{1}{2\pi i} \int_C (z - M)^{-1} z^n\, dz = M^n.$$

So

$$\begin{aligned} |M^n| &\le \frac{1}{2\pi} \int_C |(z - M)^{-1}|\, |z|^n\, |dz| \\ &\le c(|\sigma(M)| + \delta)^{n+1}, \end{aligned}$$

where $c = \sup_{z\in C} |(z - M)^{-1}|$.

Thus

$$|M^n|^{1/n} \le c^{1/n}(|\sigma(M)| + \delta)^{1+\frac{1}{n}}.$$

Hence

$$\limsup_{n\to\infty} |M^n|^{1/n} \le |\sigma(M)| + \delta.$$

Since this is true for all δ, that does it. □

Note not that every element of $\sigma(M)$ is an eigenvalue of M. For example, if $M : \ell^2 \to \ell^2$ is defined by

$$M(x_1, x_2, \ldots) = (x_1, x_2/2, x_3/4, \ldots),$$

then $1, 1/2, 1/4, \ldots$ are eigenvalues. Since the spectrum is closed, then $0 \in \sigma(M)$, but 0 is not an eigenvalue for M.

16.2 Functional calculus

We can define $p(M) = \sum_{i=1}^n a_i M^i$ for any polynomial p, and since

$$\limsup |M^k|^{1/k} = |\sigma(M)|,$$

also for any analytic function whose power series' radius of convergence is larger than the spectral radius (by the root test).

Let G be a domain containing $\sigma(M)$, f analytic in G, C a closed curve in $G \cap \rho(M)$ whose winding number is 1 about each point in $\sigma(M)$ and 0 about each point of G^c. Define

$$f(M) = \frac{1}{2\pi i} \int_C (z - M)^{-1} f(z)\, dz.$$

By Cauchy's theorem, this is independent of the contour chosen.

Theorem 16.5 *(1) If f is a polynomial, the two definitions agree.*

(2) $f(M)g(M) = (fg)(M)$.

(3) (Spectral mapping theorem)

$$\sigma(f(M)) = f(\sigma(M)).$$

Proof. (1) follows from

$$\frac{1}{2\pi i}\int_C (z-M)^{-1} z^n\, dz = M^n.$$

(2)

$$(zI - M) - (wI - M) = (z-w)I.$$

Multiplying this by

$$\frac{(z-M)^{-1}(w-M)^{-1}}{z-w},$$

we get

$$\frac{1}{z-w}[(w-M)^{-1} - (z-M)^{-1}] = (z-M)^{-1}(w-M)^{-1},$$

which is known as the resolvent identity.

$f \to f(M)$ is linear. Let C, D be closed curves as above with D strictly inside C. Then

$$\begin{aligned} f(M)g(M) &= \Big(\frac{1}{2\pi i}\Big)^2 \int_C\int_D (z-M)^{-1}(w-M)^{-1} f(z)g(w)\, dz\, dw \\ &= \Big(\frac{1}{2\pi i}\Big)^2 \int_C\int_D \frac{(w-M)^{-1}}{z-w} f(z)g(w)\, dz\, dw \\ &\quad - \int_C\int_D \frac{(z-M)^{-1}}{z-w} f(z)g(w)\, dz\, dw. \end{aligned}$$

Since

$$\frac{1}{2\pi i}\int_C \frac{f(z)}{z-w}\, dz = f(w),$$

the first integral is

$$\frac{1}{2\pi i}\int_D (w-M)^{-1} f(w)g(w)\, dw = (fg)(M).$$

Note C winds once around each point of D.

Because D does not wind around any point of C,

$$\frac{1}{2\pi i}\int_D \frac{g(w)}{z-w}\, dw = 0,$$

and so the second integral is 0.

(3) Suppose $\mu \notin f(\sigma(M))$; we show $\mu \notin \sigma(f(M))$. If $\mu \neq f(\lambda)$ for some $\lambda \in \sigma(M)$, then $f(z) - \mu$ does not vanish on $\sigma(M)$. So $g(z) = (f(z) - \mu)^{-1}$ is analytic in an open set containing $\sigma(M)$ and we can define $g(M)$ there.

Let

$$h = (f(z) - \mu)g(z) = 1.$$

So by (2)

$$[f(M) - \mu I]g(M) = h(M) = I.$$

Therefore $g(M)$ is the inverse of $f(M) - \mu I$, and hence $\mu \notin \sigma(f(M))$.

Suppose $\lambda \in \sigma(M)$. We show $f(\lambda) \in \sigma(f(M))$. Let

$$k(z) = \frac{f(z) - f(\lambda)}{z - \lambda}.$$

$k(z)$ is analytic in an open set containing $\sigma(M)$, so we can define $k(M)$ there. $(z - \lambda)k(z) = f(z) - f(\lambda)$, so

$$(M - \lambda I)k(M) = f(M) - f(\lambda)I.$$

Since $\lambda \in \sigma(M)$, we have that $M - \lambda I$ is not invertible. If the product were, then we saw that each factor would be as well. Therefore $f(M) - f(\lambda)I$ is not invertible. □

Corollary 16.6 *Suppose $\sigma(M) = \sigma_1 \cup \cdots \cup \sigma_N$, where the σ_i are closed and disjoint. Let C_j be a closed curve that winds once around each point of σ_j but not around any point of any other σ_k. Let*

$$P_j = \frac{1}{2\pi i} \int_{C_j} (z - M)^{-1}\, dz.$$

Then

(1) $P_j^2 = P_j$, $P_j PK = 0$.

(2) $\sum P_j = I$.

(3) $P_M \neq 0$ if $\sigma_m \neq \emptyset$.

The proof uses the observation that $C = \sum C_j$ winds once around each point of $\sigma(M)$.

17 Commutative Banach algebra

We look at commutative Banach algebras with a unit. Commutative means $MN = NM$ for all $M, N \in \mathcal{L}$.

p is a multiplicative functional on $\mathcal{L}$ if p is a homomorphism from $\mathcal{L}$ into $\mathbb{C}$.

Proposition 17.1 *Every homomorphism is a contraction.*

Proof. $M = IM$, so $p(M) = p(IM) = p(I)p(M)$, or $p(I) = 1$. If K is invertible,

$$p(K)p(K^{-1}) = p(KK^{-1}) = p(I) = 1,$$

so $p(K) \neq 0$. Suppose $|p(M)| > |M|$ for some M. Then if $B = M/p(M)$, we have $|B| < 1$, so $K = I - B$ is invertible. But

$$p(K) = p(I) - p(M/p(M)) = 1 - 1 = 0,$$

a contradiction.

We will show that if $p(K) \neq 0$ for all homomorphisms, then K is invertible.

$\mathcal{I} \subset \mathcal{L}$ is an ideal if $\mathcal{I}$ is a linear subspace, $\mathcal{I} \neq \{0\}$, $\mathcal{I} \neq \mathcal{L}$, and if $M \in \mathcal{L}$ and $J \in \mathcal{I}$, then $MJ \in \mathcal{I}$.

As an example, let $\mathcal{L} = C(S)$, let $r \in S$, and let $\mathcal{I} = \{f : f(r) = 0\}$.

Sometimes the requirement that $\mathcal{I} \neq \mathcal{L}$ is not part of the definition, and we talk about proper ideals to be the ones that are properly contained in $\mathcal{L}$.

If $I \in \mathcal{I}$, then $\mathcal{I} = \mathcal{L}$. If $\mathcal{I}$ contains an invertible element, then $\mathcal{I}$ contains the identity, and hence equals $\mathcal{L}$.

Lemma 17.2 *Let q be a homomorphism from $\mathcal{L}$ onto $\mathcal{A}$, but where q is not an isomorphism and $q(\mathcal{L}) \neq 0$. Then*

(1) $\{K \in \mathcal{L} : q(K) = 0\}$ is an ideal. (This set is called the kernel of q.)

(2) If $\mathcal{I}$ is an ideal, then $\mathcal{I}$ is the kernel of some non-trivial homomorphism.

Proof. (1) is easy. For (2), let $\mathcal{A} = \mathcal{L}/\mathcal{I}$. Let q map M into the equivalence class containing M. Then the kernel of q is $\mathcal{I}$. □

Proposition 17.3 *If $K \in \mathcal{L}$, $K \neq 0$, and K not invertible, then K lies in some ideal.*

Proof. Look at $K\mathcal{L} = \{KM : M \in \mathcal{L}\}$. Note $K\mathcal{L}$ does not contain the identity. □

Lemma 17.4 *Every ideal is contained in some maximal ideal.*

Proof. Order by inclusion. The union of a totally ordered subcollection will be an upper bound. (Note that if $I \notin \mathcal{I}_\alpha$ for all α, then $I \notin \cup_\alpha \mathcal{I}_\alpha$.) Then use Zorn's lemma. □

A division algebra is one where every nonzero element is invertible.

Proposition 17.5 *If $\mathcal{M}$ is a maximal ideal of $\mathcal{L}$, then $\mathcal{A} = \mathcal{L}/\mathcal{M}$ is a division algebra.*

Proof. Suppose $C \in \mathcal{A}$ and C is not invertible. Then $C\mathcal{A}$ is an ideal in $\mathcal{A}$. Let $q : \mathcal{L} \to \mathcal{L}/\mathcal{M} = \mathcal{A}$ be the usual map. $C = CI \in C\mathcal{A}$. $q^{-1}(C)$ is in the ideal $C\mathcal{A}$ but properly contains $\mathcal{M}$ because $q(M) = 0$ if $M \in \mathcal{M}$. This contradicts $\mathcal{M}$ being maximal. □

Lemma 17.6 *The closure of an ideal is an ideal.*

Proof. The only thing to prove is that $I \notin \overline{\mathcal{I}}$. We know $I \notin \mathcal{I}$, and so if $N \in B_1(I)$, then N is invertible, and hence not in $\mathcal{I}$. So $B_1(I)$ is an open set about I that is disjoint from $\mathcal{I}$. Therefore $I \notin \overline{\mathcal{I}}$. □

Lemma 17.7 *If $\mathcal{M}$ is a maximal ideal, then $\mathcal{M}$ is closed.*

Proof. If not, $\overline{\mathcal{M}}$ is an ideal strictly larger than $\mathcal{M}$. □

Lemma 17.8 *If $\mathcal{I}$ is a closed ideal in $\mathcal{L}$, then $\mathcal{L}/\mathcal{I}$ is a Banach algebra.*

Proposition 17.9 *If $\mathcal{A}$ is a Banach algebra with unit that is a division algebra, then $\mathcal{A}$ is isomorphic to $\mathbb{C}$.*

Proof. If $K \in \mathcal{A}$, there exists $\kappa \in \sigma(K)$. So $\kappa I - K$ is not invertible. Therefore $\kappa I - K = 0$, or $K = \kappa I$. The map $K \to \kappa$ is the desired isomorphism. □

Theorem 17.10 *$K \in \mathcal{L}$ is invertible if and only if $p(K) \neq 0$ for all homomorphisms p of $\mathcal{L}$ into $\mathbb{C}$.*

Proof. Suppose K is not invertible. K is in some maximal ideal $\mathcal{M}$. Then $\mathcal{M}$ is closed, $\mathcal{L}/\mathcal{M}$ is a division algebra, and is isomorphic to $\mathbb{C}$.

$$p : \mathcal{L} \to \mathcal{L}/\mathcal{M} \to \mathbb{C}$$

is a homomorphism onto $\mathbb{C}$, and its null space is $\mathcal{M}$. Since $K \in \mathcal{M}$, then $p(K) = 0$. □

18 Applications of Banach algebras

18.1 Absolutely convergent Fourier series

Let $\mathcal{L}$ be the set of continuous functions from the unit circle S^1 to the complex functions such that $f(\theta) = \sum c_n e^{in\theta}$ with $\sum |c_n| < \infty$. We let the norm of f be $\sum |c_n|$. We check that $\mathcal{L}$ is a Banach algebra. To do that, we use the fact that the Fourier coefficients for fg are the convolution of those for f and those for g. But the convolution of two ℓ^1 functions is in ℓ^1, so $fg \in \mathcal{L}$.

If $w \in S^1$, set $p_w(f) = f(w)$. p_w is a homomorphism from $\mathcal{L}$ to $\mathbb{C}$.

Proposition 18.1 *If p is a homomorphism from $\mathcal{L}$ to $\mathbb{C}$, then there exists w such that $p(f) = f(w)$ for all $f \in \mathcal{L}$.*

Proof. $p(I) = 1$ and $|p(M)| \le |M|$, so p has norm 1. Then

$$|p(e^{i\theta})| \le 1, \qquad |p(e^{-i\theta})| \le 1,$$

and

$$1 = p(1) = p(e^{i\theta})p(e^{-i\theta}).$$

We must have $|p(e^{i\theta})| = 1$, or we would have inequality in the above.

Therefore there exists w such that $p(e^{i\theta}) = e^{iw}$. Since p is a homomorphism, by induction $p(e^{in\theta}) = e^{inw}$. By linearity,

$$p\Big(\sum_{n=-N}^{N} c_n e^{in\theta}\Big) = \sum_{n=-N}^{N} c_n e^{inw}.$$

If $f \in \mathcal{L}$, since p is continuous and $\sum |c_n| < \infty$, we have $p(f) = f(w)$. $\square$

Theorem 18.2 *Suppose f has an absolutely convergent Fourier series and f is never 0 on S^1. Then $1/f$ also has an absolutely convergent Fourier series.*

Proof. If p is a homomorphism on $\mathcal{L}$, then $p(f) = f(w)$ for some w. Since f is never 0, $p(f) \neq 0$ for all non-trivial homomorphisms p. This implies f is invertible in $\mathcal{L}$. $\square$

18.2 The corona problem

Let H^∞ be the set of functions that are analytic and bounded in the unit disk D. If we set $|f| = \sup_{z\in D} |f(z)|$, this makes H^∞ into a commutative Banach algebra.

$M_z = \{f \in H^\infty : f(z) = 0\}$ is a maximal ideal. But if z_n is a sequence converging to the boundary, then $M = \{f \in H^\infty : \lim f(z_n) = 0\}$ is an ideal, so is contained in a maximal ideal that is not any of the M_z. Therefore

$\{M_z : z \in D\}$ is not equal to the set of all maximal ideals $\mathcal{A}$. The corona theorem says that it is dense in $\mathcal{A}$ if we provide $\mathcal{A}$ with the natural topology.

If $f \in H^\infty$, we define $\widehat{f} : \mathcal{A} \to \mathbb{C}$, the Gel'fand transform of f, as follows. If $M \in \mathcal{A}$, then H^∞/M is isomorphic to $\mathbb{C}$. Let I_M be the isomorphism. Define $\widehat{f}(M) = I_M(\overline{f})$, where $\overline{f}$ is the equivalence class of H^∞/M containing f.

Note that $\widehat{f}(M) = 0$ implies $I_M(\overline{f}) = 0$, so $\overline{f} = \overline{0}$, which implies $f \in M$.

Let $z \in D$. If $\overline{f} \in H^\infty/M_z$, $\overline{f} = \{g : g - f \in M_z\} = \{g : g(z) = f(z)\}$. So the isomorphism mapping H^∞/M_z to $\mathbb{C}$ is just $I_{M_z}(\overline{f}) = f(z)$, and thus $\widehat{f}(M_z) = f(z)$.

We define a basic neighborhood of $N \in \mathcal{A}$ to be a set of the form

$$V\{M \in \mathcal{A} : |\widehat{f}_j(M) - \widehat{f}_j(N)| < \varepsilon, j = 1, \dots, n\}$$

for some δ and some $f_1, \dots . f_n \in H^\infty$.

Theorem 18.3 *Suppose $\delta > 0$, $n > 1$, and $f_1, \dots, f_n \in H^\infty$ with*

$$\max_j |f_j(z)| \geq \delta$$

for each $z \in D$. Then there exist $g_1, \dots, g_n \in H^\infty$ with

$$f_1(z)g_1(z) + \cdots + f_n(z)g_n(z) = 1$$

for each $z \in D$.

The f_j are called corona data, the g_j corona solutions.

The corona theorem is a corollary to the above theorem.

Theorem 18.4 *The closure of $\{M_z : z \in D\}$ is equal to $\mathcal{A}$.*

Proof. Suppose not. Then there exists $N \in \mathcal{A}$ and a neighborhood V of N of the form

$$V = \{M \in \mathcal{A} : |\widehat{h}_j(M) - \widehat{h}_j(N)| < \delta, j = 1, \dots, n\}$$

that contains no M_z.

Let $f_j = h_j - \widehat{h}_j(N)$, i.e., we normalize the h_j by subtracting a constant. Then

$$V = \{M \in \mathcal{A} : |\widehat{f}_j(M)| < \delta, j = 1, \ldots, n\}.$$

By the normalization, $\widehat{f}_j(N) = 0$, so $f_j \in N$ for each j.

If $z \in D$, then $M_z \notin V$, so $|\widehat{f}_j(M_z)| \geq \delta$ for some j, or equivalently, $|f_j(z)| \geq \delta$ for some j. By the previous theorem, there exist $g_1, \ldots, g_n$ such that $f_1 g_1 + \cdots + f_n g_n = 1$. But since each $f_j \in N$ and N is an ideal, $1 \in N$, a contradiction. □

19 Operators and their spectra

19.1 Invertible maps

Proposition 19.1 *Suppose X is a Banach space, and $K : X \to X$ is bounded and onto. Then there exists ε such that if $|A| < \varepsilon$, then $K - A$ is onto.*

Note we do not assume K is 1-1. Compare this with the result that K invertible implies $K - A$ is invertible if the norm of A is small enough.

Proof. By the open mapping theorem, there exists k such that if $Kx = z$, then $|x| \leq k|z|$. Suppose $|A| < 1/k$. We show $K - A$ is onto.

Fix u, let $x_0 = 0$, and define recursively x_n by

$$Kx_{n+1} = Ax_n + u.$$

$|x_1| \leq k|u|$, and since

$$K(x_{n+1} - x_n) = A(x_n - x_{n-1}),$$

$$|x_{n+1} - x_n| \leq k|A|\,|x_n - x_{n-1}|.$$

Therefore x_n converges. If x is the limit, taking the limit in the definition of x_{n+1}, $Kx = Ax + u$, which is what we wanted. □

Note

$$|x| \leq \sum |x_{n+1} - x_n| \leq (k|A|)^n |x_1| < \frac{k}{1 - k|A|}|u|.$$

Proposition 19.2 *Suppose M is a bounded linear map from a Banach space X into itself. Then $\sigma(M') = \sigma(M)$.*

Proof. Let $K = \lambda - M$. If K is invertible, there exists L such that $KL = LK = I$. Therefore $L'K' = K'L' = I' = I$, or K' has an inverse.

If X were reflexive, we could say K' invertible implies $K = K''$ is invertible. In general, we have $K''L'' = L''K'' = I''$. Check that the restriction of K'' to X is K, the restriction of I'' to X is I, and therefore the null space of K is trivial, and hence K is 1-1.

The restriction of L'' to X is the inverse to K. Since L'' is continuous, R_K is closed. If $R_K \neq X$, we can use Hahn-Banach to find $\ell \neq 0$ annihilating R_K. Then $\ell \in N_{K'}$. But K' is invertible. □

19.2 Shifts

Let R and L be the right and left shifts, resp., on ℓ^2. We have $LR = I$ but $RL \neq I$. Check that $R' = L$ and $R = L'$.

Proposition 19.3 $\sigma(R) = \sigma(L) = \{|\lambda| \leq 1\}$.

Proof. L is a contraction, so $|L| \leq 1$. Similarly $|L^n| \leq 1$. Then $|\sigma(L)| = \lim |L^n|^{1/n} \leq 1$.

If $|\lambda| < 1$ and we set $a_n = \lambda^n a_0$, then $a \in \ell^2$ and $Lx = \lambda x$. Therefore $\sigma(L)$ contains the open unit ball. Since $\sigma(L)$ is closed, this does it. □

19.3 Volterra integral operators

Let $X = C[0, 1]$ and

$$(Vx)(s) = \int_0^s x(r)\, dr.$$

Proposition 19.4 $\sigma(V) = \{0\}$,

Proof. By induction and integration by parts,

$$V^n x(s) = \frac{1}{(n-1)!} \int_0^s (s-r)^{n-1} x(r)\, dr.$$

So

$$|V^n x(s)| \le \frac{1}{(n-1)!} \int_0^s (s-r)^n |x|\, dr \le \frac{|x|}{n!}.$$

Then $|V^n| \le 1/n!$, so $|\sigma(V)| = \lim |V^n|^{1/n} = 0$. □

19.4 Fourier transform

$F^2 = R$, reflection, and $R^2 = I$, so $F^4 = I$. Therefore $\sigma(F) \subset \{\pm 1, \pm i\}$ by the spectral mapping theorem.

20 Compact maps

20.1 Basic properties

A subset S is precompact if $\overline{S}$ is compact. Recall that if A is a subset of a metric space, A is precompact if and only if every sequence in A has a subsequence which converges in $\overline{A}$. Also, A is compact if and only if A is complete and totally bounded.

A map K from a Banach space X to a Banach space U is compact if $K(B_1^X(0))$ is precompact in U.

One example is if K is degenerate, so that R_K is finite dimensional. The identity on ℓ^2 is not compact.

Here is a more complicated example. Let $X = U = \ell^2$ and define

$$K(a_1, a_2, \ldots,) = (a_1/2, a_2/2^2, a_3/2^3, \ldots).$$

Take a sequence x_n in $K(B_1(0))$. The j^{th} coordinates of x_n are bounded by 2^{-j}. So there exists a subsequence such that x_n^j converges. Use the diagonalization procedure to get a subsequence along which x_n^j converges for every j. The limit will have the j^{th} coordinate bounded by 2^{-j}, so Kx_n converges along the subsequence to an element of $K(\overline{B}_1(0))$,

The following facts are easy:

(1) If C_1, C_2 are precompact subsets of a Banach space, then $C_1 + C_2$ is precompact.

(2) If C is precompact, so is the convex hull of C.

(3) If $M : X \to U$ and C is precompact in X, then $M(C)$ is precompact in U.

Proposition 20.1 *(1) If K_1 and K_2 are compact maps, so is $kK_1 + K_2$.*

(2) If $X \xrightarrow{L} U \xrightarrow{M}$, where M is bounded and L is compact, then ML is compact.

(3) In the same situation as (2), if L is bounded and M is compact, then ML is compact.

(4) If K_n are compact maps and $\lim |K_n - K| = 0$, then K is compact.

We can use (4) to give another proof that our example K above is compact. It is the limit in norm of K_n, where

$$K_n(a_1, a_2, \ldots) = (a_1/2, a_2/2^2, \ldots, a_n/2^n, 0, \ldots).$$

Proof. (1) For the sum, $(K_1 + K_2)(B) \subset K_1(B) + K_2(B)$, and the multiplication by k is similar.

(2) $ML(B)$ will be compact because $L(B)$ is compact and M is continuous.

(3) $L(B)$ will be contained in some ball, so $ML(B)$ is precompact.

(4) Let $\varepsilon > 0$. Choose n such that $|K_n - K| < \varepsilon$. $K_n(B)$ can be covered by finitely many balls of radius ε, so $K(B)$ is covered by the set of balls with the same centers and radius 2ε. Therefore $K(B)$ is totally bounded. □

One way of rephrasing (2)- (4) is that the set of compact maps are a closed 2-sided ideal in $\mathcal{L}(X)$.

Proposition 20.2 *If X and U are Banach spaces and $K : X \to U$ is compact and Y is a closed subspace of X, then the map $K\,|_Y$ is compact.*

Proposition 20.3 *If $K : X \to X$ is compact and Y is a closed invariant subspace of the Banach space X, then $K : X/Y \to X/Y$ is compact.*

Recall the following fact that we proved in Chapter 5. If X is a normed linear space and Y a proper closed linear subspace, then there exists $x \in X$ with $|x| = 1$ and $d(x, Y) \geq 1/2$.

Theorem 20.4 *Suppose $K : X \to X$ is compact and $T = I - K$.*

(1) $\dim(N_T) < \infty$.

(2) If $N_j = N_{T^j}$, there exists i such that $N_k = N_i$ for $k > i$.

(3) R_T is closed.

Proof. (1) $y \in N_T$ implies $y = Ky$. So the unit ball in N_T is precompact. But saying the unit ball is compact implies the space is finite dimensional.

(2) If not, N_{i-1} is a proper subset of N_i for all i. There exists $y_i \in N_i$ such that $|y_i| = 1$ and $d(y_i, N_{i-1}) > 1/2$. If $m < n$,

$$Ky_n - Ky_m = y_n - Ty_n - y_m + Ty_m.$$

Now $-Ty_n - y_m + Ty_m \in N_{n-1}$. So $|Ky_n - Ky_m| > 1/2$. therefore no subsequence of Ky_n converges, a contradiction.

(3) Suppose $y_k \to y$, where $y_k = Tx_k$. We need to show $y \in R_T$. Let $d_k = d(x_k, N_T)$.

Step 1. d_k is bounded: choose $z_k \in N_T$ such that $w_k = x_k - z_k$ satisfies $|w_k| = |x_k - z_k| < 2d_k$. $Tz_k = 0$, so $Tw_k = Tx_k - Tz_k = y_k$. $y_k \to y$, so $\{y_k\}$ is bounded. If d_k is unbounded,

$$T\Big(\frac{w_k}{d_k}\Big) = \frac{y_k}{d_k} \to 0.$$

Let $u_k = w_k/d_k$, so $|u_k| < 2$.

$$0 \leftarrow Tu_k = u_k - Ku_k.$$

Since Ku_k has a convergent subsequence, u_k does too. Suppose $u_k \to u$ (along the subsequence). Then since T is bounded, $Tu = \lim Tu_k = 0$, or $u \in N_T$. We have $|x_k - z| \geq d_k$ if $z \in N_T$. So $|w_k - z| = |x_k - (z + z_k)| \geq d_k$, hence $|w_k - d_k z| \geq d_k$, hence $|u_k - z| \geq 1$, a contradiction.

Step 2. Since $|w_k| < 2d_k$, then w_k is a bounded sequence. $w_k - Kw_k = y_k \to y$. Kw_k has a convergent subsequence, so w_k does too, say with limit w. Then $w - Kw = Tw = y$, and R_T is closed. □

Proposition 20.5 *Suppose $K : X \to X$ is compact. If $N_T = \{0\}$, then $R_T = X$.*

Proof. Since $\dim N_T = 0$, T is 1-1. Assume $X_1 = R_T$ is a proper subset of X. Then $X_2 = TX_1$ is a proper subset of X_1. To see this, suppose $u \in X$ and $u \notin X_1$. Then $Tu \in TX = X_1$. But if $Tu \in X_2$, then $Tu = Tv$ for some $v \in X_1$, and then $u = v \in X_1$ since T is 1-1, a contradiction. Continue to define X_3 and so on.

We have $X_k = R_{T^k}$, and

$$T^k = (I - K)^k = I + \sum_{j=1}^{k} (-1)^j \binom{k}{j} K^j,$$

or T^k is equal to I plus a compact operator. Therefore X_k is closed.

Pick $x_k \in X_k$ such that $|x_k| = 1$ and $\operatorname{dist}(x_k, X_{k+1}) > \frac{1}{2}$.

$$Kx_m - Kx_n = x_m - Tx_m - x_n + Tx_n.$$

The last three terms are in X_{m+1}, so $|Kx_n - Kx_m| \geq \frac{1}{2}$, hence no subsequence converges, a contradiction to K being compact. □

20.2 Spectral theory

Theorem 20.6 *Suppose X is a Banach space and $K : X \to X$.*

(1) $\sigma(K)$ consists of countably many complex numbers λ_n, whose only accumulation point is $\{0\}$. If $\dim X = \infty$, then $0 \in \sigma(K)$.

(2) Each $\lambda_j \neq 0$ is an eigenvalue:

(a) The null space of $K - \lambda_j$ is finite dimensional.

(b) There exists i such that the null space of $(K-\lambda_j)^k$ is the same as the null space of $(K-\lambda_j)^i$ if $k > i$.

(b) says that the multiplicity of each nonzero eigenvalue is finite.

Proof. We prove (2) first. Suppose $\lambda_j \in \sigma(K)$ and $\lambda_j \neq 0$. Let $T = I - \lambda_j^{-1} K$. $\lambda_j T = \lambda_j - K$ is not invertible, so its null space, which is the same as that of T, if finite dimensional. If $N_T = \{0\}$, then $R_T = X$, and then T is invertible. So N_T is larger than $\{0\}$. If $z \in N_T$, $(\lambda_j - K)z = 0$, or z is an eigenvector. By the previous theorem, the multiplicity is finite.

Now we look at (1). Suppose λ_n is a sequence of distinct non-zero eigenvalues. with corresponding eigenvectors x_n. Let Y_n be the linear space spanned by $\{x_1, \ldots, x_n\}$. We claim the x_n are linearly independent. If not, we can write $x_n = \sum_{j=1}^{n-1} c_j x_j$, and then

$$\lambda_n x_n = K x_n = \sum_{j=1}^{n} c_k K x_j = \sum_{j=1}^{n-1} c_j \lambda_j x_j,$$

or

$$\sum_{j=1}^{n-1} c_j x_j = \sum_{j=1}^{n-1} c_j \frac{\lambda_j}{\lambda_n} x_j,$$

which says $\{x_1, \ldots, x_{n-1}\}$ are linearly dependent. We then use induction. Therefore Y_{n-1} is a proper subset of Y_n. Choose $y_n \in Y_n$ such that $|y_n| = 1$ and $\operatorname{dist}(y_n, Y_{n-1}) > 1/2$. We can write

$$y_n = \sum_{j=1}^{n} a_j x_j.$$

So

$$K y_n - \lambda_n y_n = \sum_{j=1}^{n} (\lambda_j - \lambda_n) a_j x_j \in Y_{n-1}.$$

Therefore if $n > m$,

$$K y_n - K y_m = (K y_n - \lambda_n y_n) - (K y_m - \lambda_m y_m) - \lambda_m y_m,$$

which equals $\lambda_n y_n$ plus an element of Y_{n-1}, and so

$$|Ky_n - Ky_m| \geq |\lambda_n|/2.$$

Since K is compact, there can only be finitely many λ_n with $|\lambda_n| > \delta$.

If $\dim X = \infty$, define the y_n as before. $0 \notin \sigma(K)$ implies that K is invertible. Since K is compact, there is a subsequence Ky_{n_j} which converges. But then $y_{n_j} = K^{-1}Ky_{n_j}$ converges, a contradiction. □

Theorem 20.7 *K is compact if and only if K' is compact.*

Proof. We want to show that if $\{\ell_n\} \subset U^*$ with $|\ell_n| \leq 1$, then $\{K'\ell_n\}$ has a Cauchy subsequence. Let $J = \overline{K(B)}$. The quantities $|(\ell_n, u)|$ are uniformly bounded, and also equicontinuous on J:

$$|(\ell_n, u) - (\ell_n, v)| = |(\ell_n, u - v)| \leq |u - v|.$$

J is compact, so by Ascoli-Arzelà there exists a uniformly convergent subsequence. So given ε, there exists N such that $|(\ell_n, u) - (\ell_m, u)| < \varepsilon$ if $n, m \geq N$ and $u \in J$. So

$$\varepsilon > |(\ell_n - \ell_m, Kx)| = |(K'\ell_n - K'\ell_m, x)|$$

for all $x \in B$. Therefore $|K\ell_n - K\ell_m| < \varepsilon$.

Conversely, if K' is compact, then K'' is compact. But K is the restriction of K'' to X, so is also compact. □

21 Positive compact operators

We'll do the Krein-Rutman theorem, which is a generalization of the Perron-Frobenius theorem.

Theorem 21.1 *Suppose Q is compact and Hausdorff and $X = C(Q)$, the complex-valued continuous functions on Q. Suppose $K : C(Q) \to C(Q)$*

and K is compact. Suppose further than K maps real-valued functions to real-valued functions. Finally, suppose that whenever $p \geq 0$ and p is not identically zero, then Kp is strictly positive. Then K has a positive eigenvalue σ of multiplicity one, the associated eigenfunction is positive, and all the other eigenvalues of K are strictly smaller in absolute value than σ.

Examples include matrices with all positive entries, the semigroup P_t when $t = 1$ for reflecting Brownian motion on a bounded interval, and

$$Kf(x) = \int K(x, y)f(y)\,\mu(dy),$$

where K is jointly continuous, positivie, and μ is a finite measure. It is an exercise to show that the operator K is compact.

Proof. If $x \leq y$ and $x \not\equiv y$, then $y - x \geq 0$, so $K(y - x) > 0$, or $Kx < Ky$.

Step 1. Suppose $\kappa > 0$ is such that there exists $x \in C(Q)$ with $x \geq 0$, and $\kappa x \leq Kx$ at all points of Q.

Step 2. There exists such a $\kappa > 0$: Let $x \equiv 1$. So Kx is strictly positive, and we can let $\kappa = \min Kx$.

Now

$$\kappa Kx = K(\kappa x) \leq K(Kx) = K^2x.$$

So

$$\kappa^2 x \leq \kappa Kx \leq K^2 x,$$

and in general,

$$\kappa^n x \leq K^n x.$$

Since $x \geq 0$,

$$\kappa^n |x| \leq |K^n x| \leq |K^n|\,|x|,$$

so

$$|\sigma(K)| = \lim |K^n|^{1/n} \geq \kappa.$$

Therefore $|\sigma(K)|$ is strictly positive. Since K is compact, the set of eigenvalues of K is nonempty. We have shown that there exists a non-zero eigenvalue for K.

Step 3. K is compact, so there exists an eigenvalue λ and an eigenfunction z such that $Kz = \lambda z$, $|\lambda| = |\sigma(K)|$. Let λ and z be any pair with $|\lambda| = |\sigma(K)|$.

(a) We claim: if $y = |z|$ and $\sigma = |\sigma(K)|$, then $\sigma y \leq Ky$.

Proof: Let $q \in Q$. Multiply z by $\alpha \in \mathbb{C}$ such that $|\alpha| = 1$ and $\alpha\lambda z(q)$ is real and non-negative. Of course α depends on q. Write $z = u + iv$. Then

$$Ku(q) + iKv(q) = Kz(q) = \lambda z(q).$$

Looking at the real part,

$$\lambda z(q) = (Ku)(q).$$

Next, $u \leq |z| = y$, and

$$|\lambda| y(q) = |\lambda z(q)| = Ku(q) \leq (Ky)(q). \tag{21.1}$$

Then

$$\sigma y(q) \leq Ky(q). \tag{21.2}$$

Although z depends on α, which depends on q, neither σ nor y depend on q. Since q was arbitrary, the inequality (21.2) holds for all q.

(b) We claim

$$\sigma y = Ky.$$

Proof: If not, there exists q such that $\sigma y(q) < Ky(q)$. By continuity, there exists a neighborhood N about q such that

$$\sigma y(s) + \delta \leq Ky(s), \qquad s \in N.$$

Let $p > 0$ in N, 0 outside of N, and so $Kp > 0$.

We will find $c, \varepsilon > 0$ and set $x = y + \varepsilon p$, $\kappa = \sigma + c\varepsilon$, and get $\kappa x \leq Kx$. This will be a contradiction since the largest κ possible is less than or equal to $|\sigma(K)| = \sigma$.

Now $Kp > 0$, so there exists $c \leq 1$ such that $cy \leq Kp$. If $s \in N$,

$$\begin{aligned} Kx(s) &= Ky(s) + \varepsilon Kp(s) \geq Ky(s) + \varepsilon cy(s) \\ &\geq \sigma y(s) + \delta + \varepsilon cy(s). \end{aligned}$$

Now

$$\begin{aligned} \kappa x(s) &= (\sigma + c\varepsilon)(y + \varepsilon p)(s) = \sigma y(s) + \varepsilon cy(s) + \sigma\varepsilon p(s) \\ &\quad + c\varepsilon^2 p(s) \\ &\leq Kx(s) - \delta + \sigma\varepsilon p(s) + c\varepsilon^2 p(s). \end{aligned}$$

Since p is bounded above, we can take ε small enough so that the last line is less than or equal to $Kx(s)$.

If $s \notin N$, then $p(s) = 0$ and

$$\begin{aligned}\kappa x(s) &= \kappa y(s) = (\sigma + c\varepsilon)y(s) = \sigma y(s) + \varepsilon c y(s)\\ &\le Ky(s) + \varepsilon K p(s) = Kx(s).\end{aligned}$$

Step 4. We next show that any other eigenvalue that has absolute value σ is in fact equal to σ. Let z be any eigenfunction corresponding to λ with $|\lambda| = \sigma$. Fix $q \in Q$. As before, we may assume $\lambda z(q) \ge 0$. As before, write $z = u + iv$ and then $\lambda z(q) = Ku(q)$. We have $u \le |z| = y$.

Suppose $u < y$ at some point $q' \in Q$. Then $u \le y$ and $u < y$ at one point means that we have $Ku < Ky$ at every point, and so

$$|\lambda| y(q) = |\lambda z(q)| = \lambda z(q) = Ku(q) < Ky(q).$$

So $\sigma y(q)) < Ky(q)$. But we showed $\sigma y = Ky$. Therefore u is identically equal to y. This implies that z is real and positive, and then it follows that λ is real and positive. Since $z = \sigma^{-1} K z$, z is strictly positive.

Step 5. Finally, we show σ has multiplicity 1. If not, there exist real eigenfunctions y_1, y_2. But some linear combination w of y_1, y_2 will be real and take both positive and negative values. As before $|w|$ will be an eigenfunction that is non-negative, and must also take the value 0. Moreover the corresponding eigenvalue is σ. But then $0 < K|w| = \sigma|w|$, a contradiction to $|w|$ taking the value 0. □

22 Invariant subspaces

22.1 Compact maps

There exist operators without any eigenfunctions, but they still have non-trivial invariant subspaces. Y is a non-trivial invariant subspace for K if $\overline{Y}$ is a proper subset of X and $KY \subset Y$. Note that if K has an eigenvector y, then the linear subspace spanned by $\{y\}$ is a non-trivial invariant subspace. What if the only element of $\sigma(K)$ is the point 0?

Theorem 22.1 *Let X be a Banach space over $\mathbb{C}$ of dimension greater than 1 and let $K : X \to X$ be compact. Then K has a non-trivial invariant subspace.*

Proof. Suppose $K \neq 0$ and normalize so that $|K| = 1$. Choose $x_0 \in X$ such that $|x_0| > 1$, $|Kx_0| > 1$. Let $B = \overline{B_1(x_0)}$, and note $0 \notin B$. Let $D = \overline{KB}$. D is compact. Since $|x_0| > 1$ and $|K| = 1$, then $0 \notin D$.

Suppose K has no non-trivial invariant subspaces. Given $y \neq 0$, $\{p(K)y\}$ where p is a polynomial, is invariant under K. Its closure is a closed invariant subspace, so must be all of X. Therefore $\{p(K)y\}$ is dense in X.

$0 \notin D$, so if $y \in D$, there exists a polynomial p such that $|p(K)y - x_0| < 1$. The set of z satisfying $|p(K)z - x_0| < 1$ is an open set containing y. D is compact, so can be covered by finitely many of them. Therefore there exist $p_1, \ldots, p_N$ such that if $y \in D$, $|p_i(K)y - x_0| < 1$ for some p_i.

Let $K_i = p_i(K)$. If $y \in D$, $|K_i y - x_0| < 1$ for at least one i.

$x_0 \in B$ and $Kx_0 \in D$. If $y = Kx_0$, there exists i_1 such that $|K_{i_1}K(x_0) - x_0| < 1$. So $K_{i_1}Kx_0 \in B$, hence $KK_{i_1}Kx_0 \in D$. Let $y = KK_{i_1}Kx_0$ and there exists i_2 such that

$$|K_{i_2}KK_{i_1}Kx_0 - x_0| < 1.$$

Continue, so

$$\Big|\prod_{k=1}^{n}(K_{i_k}K)x_0 - x_0\Big| < 1.$$

Hence

$$\Big|\prod_{k=1}^{n}(K_{i_k}K)x_0\Big| > |x_0| - 1 > 0.$$

The K_i's and K commute, so

$$\Big|(\prod_{k=1}^{n}K_{i_k})K^n x_0\Big| > |x_0| - 1.$$

Let $c = \sup_i |K_i|$. Therefore

$$c^n|K^n|\,|x_0| > |x_0| - 1.$$

then

$$c|K^n|^{1/n}|x_0|^{1/n} > (|x_0| - 1)^{1/n}.$$

Therefore $|\sigma(K)| \geq 1/c > 0$. By spectral theory, $\sigma(K)$ contains points other than 0, and there are eigenvalues. The corresponding eigenspace is invariant under K, a contradiction. □

23 Compact symmetric operators

Let H be a complex Hilbert space. $A : H \to H$ is Hermitian (symmetric) if $A = A^*$, that is, $(Ax, y) = (x, Ay)$.

Proposition 23.1 *(1) (Ax, x) is real.*

(2) (Ax, x) is not identically 0 unless $A = 0$.

Proof. (1)

$$(Ax, x) = (x, Ax) = \overline{(Ax, x)}.$$

(2) If $(Ax, x) = 0$ for all x, then

$$\begin{aligned} 0 &= (A(x+y), x+y) = (Ax, x) + (Ay, y) + (Ax, y) + (Ay, x) \\ &= (Ax, y) + (y, Ax) = (Ax, y) + \overline{(Ax, y)}. \end{aligned}$$

So $\mathrm{Re}\,(Ax, y) = 0$. Replacing x by ix, $\mathrm{Re}\,(i(Ax, y)) = 0$. □

If $(Ax, x) \geq 0$ for all x, we say A is positive, and write $A \geq 0$. Writing $A \leq B$ means $B - A \geq 0$. For matrices, one uses positive definite.

Now suppose A is compact.

Proposition 23.2 *If $x_n \xrightarrow{w}$, then $Ax_n \xrightarrow{s}$.*

Proof. If $x_n \xrightarrow{w} x$, then $Ax_n \xrightarrow{w} Ax$, since $(Ax_n, y) = (x_n, Ay) \to (x, Ay) = (Ax, y)$. If x_n converges weakly, then $\|x_n\|$ is bounded so Ax_n lies in a precompact set.

Any subsequence of Ax_n has a further subsequence which converges strongly. The limit must be Ax. $\square$

Theorem 23.3 *(Spectral theorem) Suppose H is a complex Hilbert space, A is compact and symmetric. There exist $z_n \in H$ such that $\{z_n\}$ is an orthonormal basis for H, each z_n is an eigenvector, the eigenvalues are real, and their only point of accumulation is 0.*

Proof. If $A = 0$, any orthonormal basis will do. So suppose $A \neq 0$. Let

$$M = \sup_{\|x\|=1} (Ax, x).$$

We may assume $M > 0$ since there exists x such that $(Ax, x) \neq 0$. (Look at $-A$ if necessary.) We have $(Ax, x) \leq \|Ax\| \, \|x\|$, so $M \leq \|A\|$.

We claim the maximum is attained. Choose x_n with $\|x_n\| = 1$ such that $(Ax_n, x_n) \to M$. There exists a subsequence, also denoted x_n, which converges weakly, say to z. Since A is compact, $Ax_n \xrightarrow{s} Az$. So $(Az, z) = \lim(Ax_n, x_n) = M$. $\|x_n\| \leq 1$, so $\|z\| \leq 1$. $M > 0$ implies $z \neq 0$. Let $y = z/\|z\|$. Then $(Ay, y) = M/\|z\|^2$. If $\|z\| < 1$, then $(Ay, y) > M$, a contradiction. Therefore $\|z\| = 1$.

Note z maximizes

$$R_A(x) = \frac{(Ax, x)}{\|x\|^2}$$

over all $x \neq 0$.

Let $w \in H$, $t \in \mathbb{R}$. $R_A(z + tw) \leq R_A(z)$. So $\frac{\partial}{\partial t} R_A(z + tw) \mid_{t=0} = 0$. Doing the calculation,

$$\frac{(Aw, z) + (Az, w)}{\|z\|^2} - (Az, z)\frac{(w, z) + (z, w)}{\|z\|^4} = 0.$$

This implies

$$\mathrm{Re}\,(Az - Mz, w) = 0.$$

This is true for all w, so $Az - Mz = 0$, or z is an eigenvector.

Suppose we have found eigenvalues $z_1, z_2, \ldots, z_n$. Let Y be the orthogonal complement of the linear subspace spanned by $\{z_1, \ldots, z_n\}$. If $x \in Y$, then

$$(Ax, z_k) = (x, Az_k) = \overline{\lambda}_k (x, z_k) = 0,$$

or $Ax \in Y$. We then look at $A\,|_Y$, which will still be compact and symmetric, and find a new eigenvector z_{n+1}.

We next prove that the set of eigenvectors forms a basis. Suppose y is orthogonal to every eigenvector. Then

$$(Ay, z_i) = (y, Az_i) = (y, \alpha z_i) = 0$$

if z_i is an eigenvector, and Ay is also orthogonal to every eigenvector. If Y is the orthogonal complement to $\{z_i\}$ and $Y \neq \{0\}$, then $A : Y \to Y$, $A\,|_Y$ is symmetric, so there exists an eigenvector for $A\,|_Y$, a contradiction since Y is orthogonal to every eigenvector.

It remains to show that 0 is the only accumulation point. If $\alpha_n \neq \alpha_m$,

$$\alpha_n (z_n, z_m) = (Az_n, z_m) = (z_n, Az_m) = \alpha_m (z_n, z_m),$$

or $(z_n, z_m) = 0$. We can take the z_n to have norm 1.

Now if $\alpha_n \to \alpha \neq 0$, using the compactness of A, there is a subsequence, also called z_n, such that Az_n converges strongly, say, to w.

$$z_n = \frac{1}{\alpha_n} Az_n \to \frac{1}{\alpha} w.$$

But $\|z_n - z_m\| = 2$ if $n \neq m$, a contradiction to a_n converging. □

If $\alpha_1 \geq \alpha_2 \geq \cdots > 0$ and $Az_n = \alpha_n z_n$, then our construction shows that

$$\alpha_N = \max_{x \perp z_1, \cdots, z_{N-1}} \frac{(Ax, x)}{\|x\|^2}.$$

This is known as the Rayleigh principle.

Let

$$R_A(x) = \frac{(Ax, x)}{\|x\|^2}.$$

Proposition 23.4 *Let A be compact, symmetric, α_k the eigenvalues with $\alpha_1 \geq \alpha_2 \geq \cdots$. Then*

(1) (Fisher's principle)

$$\alpha_n = \max_{S_N} \min_{x \in S_N} R_A(x),$$

where the max is over linear subspaces S_N of dimension N.

(2) (Courant's principle)

$$\alpha_N = \min_{S_{N-1}} \max_{x \perp S_{N-1}} R_A(x).$$

Proof. Let $z_1, \ldots, z_N$ be eigenvectors with corresponding eigenvalues $\alpha_1 \geq \alpha_2 \geq \cdots \geq \alpha_N$. Let T_N be the linear subspace spanned by $\{z_1, \ldots, z_N\}$. If $y \in T_N$, we have $y = \sum_{j=1}^N c_j z_j$ for some complex numbers c_j and then

$$\begin{aligned}(Ay, y) &= \sum_{i=1}^N \sum_{j=1}^N c_i \overline{c_j} (Az_i, z_j) = \sum_i \sum_j c_i \overline{c_j} \alpha_i (z_i, z_j) \\ &= \sum_i |c_i|^2 \alpha_i \geq \sum_i |c_i|^2 \alpha_N \\ &= (y, y)\end{aligned}$$

using the fact that the z_i are orthogonal by our construction.

(1) Let z_k be the eigenvectors. Let S_N be a subspace of dimension N. There exists $y \in S_N$ such that $(y, z_k) = 0$ for $k = 1, \ldots, N-1$. Since

$$\alpha_N = \max_{x \perp z_1, \ldots, z_{N-1}} R_A(x),$$

then y is one of the vectors over which the max is being taken, so $R_A(y) \leq \alpha_N$ for this y. So $\min_{x \in S_N} R_A(x) \leq \alpha_N$. This is true for all spaces of dimension N. So the RHS is less than or equal to α_N.

But if S_N is the linear span of $\{z_1, \ldots, z_N\}$, the minimum of R_A on S_N is achieved at z_N.

(2) If $\dim S_{N-1} = N-1$, T_N, the span of $\{z_1, \ldots, z_N\}$, contains a vector y perpendicular to S_{N-1}. If $y \in T_N$, $R_A(y) \geq \alpha_N$, so $\max_{x \perp S_{N-1}} R_A(x) \geq \alpha_N$.

But taking $S_{N-1} = T_{N-1}$, we get equality. □

Proposition 23.5 *Suppose $A \le B$ with eigenvalues α_k, β_k, resp., ordered to be decreasing. Then $\alpha_k \le \beta_k$ for all k.*

Proof. $A \le B$ implies $(Ax, x) \le (Bx, x)$, so $R_A(x) \le R_B(x)$. Now use either Fisher's or Courant's principle. □

Theorem 23.6 *Suppose A is compact and symmetric. Suppose f defined on $\sigma(A)$ is bounded and complex valued. We can define $f(A)$ such that*

(1) If $f_1(\sigma) \equiv 1$, then $f_1(A) = I$.

(2) If $f_2(\sigma) \equiv \sigma$, then $f_2(A) = A$.

(3) $f \to f(A)$ is an isomorphism of the ring of bounded functions on σ into the algebra of bounded maps of H into H.

(4) $\|f(A)\| = \sup_{\sigma\in\sigma(A)} |f(\sigma)|$.

(5) If f is real, $f(A)$ is symmetric,

(6) $f \ge 0$ on $\sigma(A)$ implies $f(A)$ is positive.

Proof. If $\{z_n\}$ is an orthonormal basis of eigenvectors, and $x = \sum c_n z_n$, set $f(A)x = \sum f(\alpha_n) c_n z_n$. □

If A is positive, then $\sigma(A) \subset [0, \infty)$. We can define $\sqrt{A}$ by taking $f(\lambda) = \sqrt{\lambda}$.

24 Examples of compact symmetric operators

Let

$$Lu(x) = -u''(x) + q(x)u(x)$$

for $u \in C^2[0, 2\pi]$ with $u(0) = u(2\pi) = 0$. Suppose q is continuous and $q \ge 1$ for all x.

Fact from PDE: $Lu = f$ with $u(0) = 0$ and $u(2\pi) = 0$ has a unique solution if f is continuous.

Write $u = Af$.

Lemma 24.1

$$C = \Big\{ u : \int_0^{2\pi} (u')^2 \le 1, \int_0^{2\pi} u^2 \le 1 \Big\}$$

is precompact in L^2*.*

Proof. If $0 \le x \le y \le 2\pi$, then

$$|u(y) - u(x)| = \Big| \int_x^y u'(z)\, dz \Big| \le |y - x|^{1/2} \Big(\int_0^{2\pi} |u'(z)|^2\, dz \Big)^{1/2}$$

using Cauchy-Schwarz. Thus the functions in C are equicontinuous. If $|u(y)| \ge 100$ for some $u \in C$ and some y, then

$$|u(x)| \ge |u(y)| - |u(x) - u(y)| \ge 100 - (2\pi)^{1/2}$$

for all x, and then $\int u^2 > 1$, a contradiction to u being in C. Therefore the functions in C are uniformly bounded. By the Ascoli-Arzelà theorem, every sequence in C has a subsequence which converges uniformly on $[0, 2\pi]$, and hence in L^2. Therefore C is precompact. □

Proposition 24.2 *A is bounded, compact, symmetric, and positive.*

Proof. If $u \in C^2$ and $Lu = f$, then by integration by parts:

$$\int_0^{2\pi} uf\, dx = \int_0^{2\pi} u\, Lu\, dx = \int ((u')^2 + qu^2).$$

Now

$$\int uf \le \Big(\int u^2 \Big)^{1/2} \Big(\int f^2 \Big)^{1/2} \le \tfrac{1}{2} \int u^2 + \tfrac{1}{2} \int f^2,$$

using the inequality $xy \le \frac{1}{2}(x^2 + y^2)$. Since $q \ge 1$,

$$\int (u')^2 + \tfrac{1}{2} \int u^2 \le \tfrac{1}{2} \int f^2.$$

This implies A is bounded.

(Since

$$\int uf = \int (au)(f/a) \le \tfrac{1}{2}a^2 \int u^2 + \tfrac{1}{2a^2} \int f^2,$$

all we really need is that q is bounded below by a positive constant.)

To show A is symmetric, $(Lu, v) = (u, Lv)$. Letting $f = Lu$ and $g = Lv$, $(f, Ag) = (Af, g)$.

To see that A is positive, note

$$(Af, f) = \int uf = \int (u')^2 + qu^2 \ge 0.$$

It remains to prove A is compact.

Since $\int (u')^2 + \frac{1}{2} \int u^2 \le \frac{1}{2} \int f^2$, if f is in the unit ball in L^2, then Af has norm bounded by 1 and $(Af)'$ has norm bounded by 1/2. By the lemma, the image of the unit ball under A is precompact. □

We can extend A to all of L^2. Let e_n be eigenfunctions, α_n the eigenvalues, so $Ae_n = \alpha_n e_n$. $\|e_n'\| \le c\|Ae_n\| < \infty$. So e_n is continuous. $Le_n = \alpha_n^{-1} e_n$. The $\alpha_n \to 0$, so $\alpha_n^{-1} \to \infty$.

Remark 24.3 The above also applies to $Lu = -\Delta u + qu$ in D with $u = 0$ on the boundary.

For an example, take $q(x) \equiv c$. Solving

$$-f''(x) + cf(x) = \lambda f(x),$$

we see that the solutions are of the form $e_n = c_n \sin(nx/2)$. From our theory we know the e_n are orthogonal and form a complete orthonormal basis.

25 Trace class operators

25.1 Polar decomposition

Proposition 25.1 *Suppose $T : H \to H$ is compact. Then $T = UA$, where A is a positive symmetric operator and $U^*U = I$ on the range of A, $U = 0$ on $R_A^\perp$.*

This is called the polar decomposition and is the analog of writing a complex number as $re^{i\theta}$.

Proof. T^*T is a non-negative symmetric compact operator as $(T^*Tu, u) = (Tu, Tu) \geq 0$. So there exists a square root: $A = (T^*T)^{1/2}$.

$$\|Tu\|^2 = (Tu, Tu) = (u, T^*Tu) = (u, A^2u) = (Au, Au) = \|Au\|^2. \quad (1)$$

So if $Au = Av$, then $A(u - v) = 0$, so $T(u - v) = 0$, so $Tu = Tv$. Define $U : R_A \to H$ by $U(Au) = Tu$.

By (1), U is an isometry on R_A. Define $Un = 0$ if $n \in R_A^\perp$. For such n, $(Un, v) = (n, U^*v) = 0$ for all v, so $U^* : H \to (R_A^\perp)^\perp \subset \overline{R}_A$.

Write R for R_A. We claim $U^*Uw = w$ for $w \in \overline{R}$. If $z, w \in \overline{R}$,

$$(z, w) = (Uz, Uw) = (z, U^*Uw)$$

since U is an isometry. So $(z, U^*Uw - w) = 0$, or $(U^*Uw - w) \perp R$. Since $U^* : H \to \overline{R}$, then $U^*Uw - w$ is in $\overline{R}$ and is orthogonal to $\overline{R}$, hence is 0. Therefore $U^*Uw = w$. □

The only place we used the compactness was to define $(U^*U)^{1/2}$. We'll see later that we can define the positive square root of U^*U for any bounded linear operator U.

T compact implies that T^*T is compact, so A is compact. Let s_j be the eigenvalues of A. These are called the singular values of T.

25.2 Trace class

Let $T : H \to H$ be compact. T is of trace class if

$$\|T\|_{tr} = \sum s_j(T) < \infty.$$

Proposition 25.2 *Let T be of trace class, B bounded.*

(1) $\|T\|_{tr} = \|T^*\|_{tr}$.

(2) $\|BT\|_{tr} \leq \|B\| \, \|T\|_{tr}$.

(3) $\|TB\|_{tr} \leq \|B\| \, \|T\|_{tr}$.

(4) $\|S + T\|_{tr} \leq \|S\|_{tr} + \|T\|_{tr}$.

Proof. (1) We will show $s_j(T) = s_j(T^*)$. Since $s_j(T^*)$ is an eigenvalue of $(T^{**}T^*)^{1/2} = (TT^*)^{1/2}$, it suffices to show that TT^* and T^*T have the same eigenvalues.

Let z, λ be an eigenvector, eigenvalue pair for T^*T with $\lambda \neq 0$, so that $T^*Tz = \lambda z$. then $TT^*Tz = \lambda Tz$,and λ is an eigenvalue for TT^*. ($Tz \neq 0$ since $\lambda \neq 0$.)

(2) We will show $s_j(BT) \leq \|B\| s_j(T)$.

$$(T^*B^*BTu, u) = \|BTu\|^2 \leq \|B\|^2 \|Tu\|^2 = \|B\|^2 (T^*Tu, u).$$

So

$$(BT)^*BT \leq \|B\|^2 T^*T,$$

hence

$$s_j^2(BT) \leq \|B\|^2 s_j^2(T).$$

(3)

$$s_j(TB) = s_j(B^*T^*) \leq \|B^*\| s_j(T^*) = \|B\| s_j(T).$$

(4) We will show

$$\|T\|_{tr} = \sup \sum_n |(Tf_n, e_n)|,$$

where the supremum is over all orthonormal bases $\{f_n\}$, $\{e_n\}$.

Let z_j be normalized eigenvectors for A: $Az_j = s_j z_j$, $\|z_j\| = 1$. Then

$$f = \sum (f, z_j) z_j, \qquad Af = \sum s_j (f, z_j) z_j.$$

Then

$$Tf = UAf = \sum s_j (f, z_j) w_j,$$

where $w_j = Uz_j$. w_j is an orthonormal basis for $R = R_A$. then

$$(Tf, e) = \sum s_j (f, z_j)(w_j, e)$$

and so

$$\sum (Tf_n, e_n) = \sum \sum s_j (f_n, z_j)(w_j, e_n).$$

The right hand side is less than or equal to

$$\sum_j s_j \Big(\sum_n |(f_n, z_j)|^2\Big)^{1/2} \sum_n |(w_j, e_n)|^2\Big)^{1/2}$$
$$= \sum_j s_j \Big(\|z_j\|^2 \|w_j\|^2\Big)^{1/2} = \sum s_j = \|T\|_{tr}.$$

Therefore $\sup \sum |(Tf_n, e_n)| \le \|T\|_{tr}$.

Now let $f_n = z_n$, $e_n = w_n$ (supplemented arbitrarily on $(R_A)^\perp$. Then $(Tf_n, e_n) = s_n$ and the sup is attained.

In particular,

$$\|S+T\|_{tr} = \sup \sum_n |((S+T)f_n, e_n)|,$$

and use subadditivity on the right hand side. □

If f_n is an orthonormal basis, then we define $tr\, T = \sum (Tf_n, f_n)$. This is comparable with the trace of a matrix, which is the sum of the diagonal elements.

Proposition 25.3 *The definition of trace is independent of the choice of orthonormal basis. If T is of trace class, then the sum converges absolutely.*

Proof. We have

$$\sum (Tf_n, e_n) = \sum\sum s_j (f_n, z_j)(w_j, e_n).$$

Take $e_n = f_n$ to get

$$\sum (Tf_n, f_n) = \sum\sum s_j (f_n, z_j)(w_j, f_n).$$

We have already shown that the right hand side converges, and the sum is bounded by $\|T\|_{tr}$. By Parseval,

$$\sum_n (f_n, z_j)(f_n, w_j) = (w_j, z_j),$$

so

$$\sum (Tf_n, f_n) = \sum s_j(w_j, z_j),$$

which is independent of the sequence f_n. Here z_j are the eigenvectors of A and $w_j = Uz_j$. □

Proposition 25.4 *(1)* $|tr\, T| \le \|T\|_{tr}$.

(2) $tr\, T$ *is a linear function of* T.

(3) $tr\, T^* = tr\, T$.

(4) If B *is bounded,* $tr\, BT = tr\, TB$.

Proof. (1) has been done already. (2) and (3) follow from

$$tr\, T = \sum (Tf_n, f_n).$$

For (4),

$$Tf = UAf = U\Big(\sum s_j(f, z_j)z_j\Big) = \sum_j s_j(f, z_j)w_j.$$

Then $BTf_n = \sum s_j(f_n, z_j)Bw_j$, so using Parseval,

$$\sum (BTf_n, f_n) = \sum \sum s_j(f_n, z_j)(Bw_j, f_n) = \sum s_j(z_j, Bw_j).$$

From $Tf = \sum s_j(f, z_j)w_j$,

$$TBf = \sum s_j(Bf, z_j)w_j = \sum s_j(f, B^*z_j)w_j.$$

So

$$\begin{aligned} tr\, TB &= \sum (TBf_n, f_n) = \sum s_j(w_j, B^*z_j) \\ &= \sum \sum s_j(f_n, B^*z_j)(w_j, f_n) \\ &= \sum s_j(Bw_j, z_j). \end{aligned}$$

□

Theorem 25.5 *$tr\, T = \sum \lambda_j$ if T is of trace class, compact, and symmetric.*

Proof. Let f_n be an orthonormal basis consisting of eigenvectors of T. Then

$$tr\, T = \sum (Tf_n, f_n) = \sum \lambda_n (f_n, f_n) = \sum \lambda_n.$$

□

This theorem is true under much weaker assumptions on T.

Define $K : L^2[0,1] \to [0,1]$ by

$$Ku(s) = \int_0^1 K(s,t)u(t)\, dt.$$

K^* has kernel $\overline{K}(t,s)$.

Suppose K is continuous, symmetric, and real-valued. Then K is compact. This is an exercise, but the idea is to use equicontinuity of Ku:

$$\begin{aligned} |Ku(s) - Ku(v)| &= \Big| \int_0^1 [K(s,t) - K(v,t)]u(t)\, dt \\ &\le \Big(\int_0^1 |K(s,t) - K(v,t)|^2\, dt \Big)^{1/2} \Big(\int_0^1 u(t)^2\, dt \Big)^{1/2}. \end{aligned}$$

So $K(B)$ is a compact subset of $C[0,1] \subset L^2[0,1]$.

Therefore there exists a complete orthonormal system (κ_j, e_j). $K : L^2 \to C[0,1]$, so $e_j = \kappa_j^{-1} e_j$ is continuous if $\kappa_j \neq 0$.

Theorem 25.6 *(Mercer) Suppose K is real-valued, symmetric, and continuous. Suppose K is positive: $(Ku, u) \ge 0$ for all $u \in H$. Then*

$$K(s,t) = \sum_j \kappa_j e_j(s) e_j(t),$$

and the series is uniformly absolutely continuous.

An example is to let $K = P_t$, the transition density of absorbing or reflecting Brownian motion.

Proof. $K \geq 0$ on the diagonal: If not, if $K(r,r) < 0$, then $K(s,t) < 0$ if $|s-r|, |t-r| < \delta$ for some δ. Take $u = 1_{[r-\delta/2, r+\delta/2]}$. Then

$$(Ku, u) = \int\int K(s,t)u(t)u(s)\,ds\,dt < 0,$$

a contradiction.

Let $K_N(s,t) = \sum_{j=1}^N \kappa_j e_j(s)e_j(t)$. $K - K_N$ is a positive operator, for its eigenvectors are the e_j and its eigenvalues $\kappa_j > 0$, $j > N$. So $K - K_N$ is non-negative on the diagonal:

$$0 \leq K(s,s) - \sum_{j=1}^N \kappa_j e_j(s)^2.$$

Each term is non-negative, so the sum converges for each s. By Dini's theorem, $\sum_{j=1}^N \kappa_j e_j(s)^2$ converges uniformly and absolutely in s. By Cauchy-Schwarz, $K_N(s,t)$ converges uniformly in s and t.

Let the limit be K_∞. We need to show $K_\infty = K$. K and K_∞ have the same eigenfunctions and the same eigenvalues. They both map all functions that are orthogonal to all the e_j's to 0. Therefore $Ku = K_\infty u$ for all u. So they have the same kernel. □

Set $s = t$:

$$K(s,s) = \sum \kappa_j e_j(s)^2.$$

Integrate over s:

$$\int_0^1 K(s,s)\,ds = \sum \kappa_j.$$

Therefore, since the left hand side is finite, K is of trace class. This is true more generally: K need not be symmetric.

$K : H \to H$ is a Hilbert-Schmidt operator if there exists an orthonormal basis e_j and $\sum \|Ke_j\|^2 < \infty$. We will not prove this, but K is Hilbert-Schmidt if and only if $\sum s_j(K)^2 < \infty$.

26 Spectral theory of symmetric operators

We let M be a symmetric operator over a complex-valued Hilbert space, so that $(Mx, y) = (x, My)$.

If A is compact, we can write $x = \sum a_n e_n$ and $Ax = \sum \lambda_n a_n e_n$. Let E_n be the projection onto the eigenspace with eigenvector λ_n, so $x = \sum E_n x$ and $Ax = \sum \lambda_n E_n(x)$.

If we define a projection-valued measure $E(S)$ by

$$E(S) = \sum_{\lambda_n \in S} E_n$$

for S a Borel subset of $\mathbb{R}$, then $x = \int E(d\lambda)x$ and $Ax = \int \lambda E(d\lambda)x$.

Here E is a pure point measure. In general, we get the same result, but E might not be pure point.

Proposition 26.1 *(1) If B is bounded and symmetric, then (Bx, y) is bounded, linear in x, and skew linear in y (that is, $(Bx, cy) = \overline{c}(Bx, y)$).*

(2) If $b(x, y)$ is skew-symmetric ($b(y, x) = \overline{b(x, y)}$), linear in x, and $|b(x, y)| \le c\|x\|\,\|y\|$, then $b(x, y) = (x, By)$, where B is bounded, symmetric, and $\|B\| \le c$.

Proof. (1) is easy.

(2) Fix y, let $\ell(x) = b(x, y)$, and then $|\ell(x)| \le c\|x\|\,\|y\|$. So there exists w, depending on y, such that $b(x, y) = (x, w)$. If $x = w$,

$$\|w\|^2 = b(w, y) \le c\|w\|\,\|y\|$$

or $\|w\| \le c\|y\|$. Define B by $w = By$.

$$(x, By) = b(x, y) = \overline{b(y, x)} = \overline{(y, Bx)} = (Bx, y).$$

□

26.1 Spectrum of symmetric operators

Proposition 26.2 *If M is bounded and symmetric, then $\sigma(M) \subset \mathbb{R}$.*

Proof. Let $\lambda = \alpha + i\beta$, $\beta \neq 0$. We want to show that λ is in the resolvent. Define $B(x, y) = (x, (M - \lambda)y)$. B is linear in x and skew linear in y.

$$|B(x, y)| \leq \|x\| \, \|M - \lambda)y\| \leq \|x\| \, \|y\| \, (\|M\| + |\lambda|).$$

Now

$$B(y, y) = (y, (M - \lambda)y) = (y, My) - \alpha(y, y) - i\beta(y, y).$$

Since $(x, Mx) = (Mx, x) = (\overline{x, Mx})$, then (x, Mx) is real. Therefore

$$|B(y, y)| \geq |\beta(y, y)| = |\beta| \, \|y\|^2.$$

By the Lax-Milgram lemma, if $z \in H$ and $\ell(x) = (x, z)$, there exists y such that $B(x, y) = \ell(x)$ for all x. So

$$(x, (M - \lambda)y) = B(x, y) = (x, z).$$

This is true for all x, so $z = (M - \lambda)y$, which proves $M - \lambda$ is invertible. □

Proposition 26.3 $|\sigma(M)| = \|M\|$.

Proof.

$$\|Mx\|^2 = (Mx, Mx) = (x, M^2 x) \leq \|x\| \, \|M^2 x\| \leq \|x\|^2 \|M^2\|.$$

So $\|M\|^2 \leq \|M^2\|$. Similarly, $\|M\|^n \leq \|M^n\|$ if $n = 2^k$. The other direction follows from $\|AB\| \leq \|A\| \, \|B\|$. Taking the n^{th} root,

$$|\sigma(M)| = \lim \|M^n\|^{1/n} = \|M\|.$$

□

Proposition 26.4 *Let* $a = \inf_{\|x\|=1}(x, Mx)$ *and* b *the supremum. Then* $\sigma(M) \subset [a, b]$ *and* $a, b \in \sigma(M)$.

Proof. Let $\lambda < a$.

$$(x, (M-\lambda)x) = (x, Mx) - \lambda(x,x) \geq (a-\lambda)\|x\|^2.$$

If we define $B(x,y) = (x, (M-\lambda)y)$, then the hypotheses of the Lax-Milgram lemma hold, and as in proof of the proposition that M bounded and symmetric implies $\sigma(M) \subset \mathbb{R}$. we see that $(M-\lambda)$ is invertible. Thus $\lambda \notin \sigma(M)$ and similarly for $\lambda > b$.

We have $|(x, Mx)| \leq \|x\| \, \|Mx\| \leq \|M\|$ if $\|x\| = 1$. So $|a|, |b| \leq \|M\|$, and since $\|M\| = |\sigma(M)|$, we have $|\sigma(M)| \leq |a| \vee |b|$. So if $b > |a|$, then $b \in \sigma(M)$ and if $|a| > b$, then $a \in \sigma(M)$.

Replacing M by $M + cI$ shifts the spectrum by c. We conclude both $a, b \in \sigma(M)$. □

Proposition 26.5 *Suppose* M, N *is symmetric. Define* $\operatorname{dist}(\sigma(M), \sigma(M))$ *to be the larger of*

$$\max_{\nu \in \sigma(N)} \min_{\mu \in \sigma(M)} |\nu - \mu|, \qquad \max_{\mu \in \sigma(M)} \min_{\nu \in \sigma(N)} |\nu - \mu|.$$

Then $\operatorname{dist}(\sigma(M), \sigma(N)) \leq \|M - N\|$.

Proof. Let $d = \|M - N\|$. Suppose the first quantity is larger than d. So for some $\nu \in \sigma(N)$, $\min_{\mu \in \sigma(M)} |\mu - \nu| > d$.

$\nu \in \rho(M)$, so $M - \nu$ is invertible. Then $|\sigma((M-\nu)^{-1})| = |(\sigma(M) - \nu)^{-1}|$, and so $|\sigma(M-\nu)|^{-1} < d^{-1}$. Therefore

$$d^{-1} > |\sigma((M-\nu)^{-1})| = \|(M-\nu)^{-1}\|.$$

$$N - \nu I = M - \nu I + N - M = (M - \nu I)(I + K),$$

where

$$K = (M-\nu)^{-1}(N - M).$$

Note

$$\|K\| = \|(M-\nu)^{-1}\| \, \|N - M\| < d \cdot d^{-1} = 1.$$

Therefore $I + K$ is invertible, and so $N - \nu I$ is invertible, a contradiction to $\nu \in \sigma(N)$. □

26.2 Functional calculus for symmetric operators

Let q be a polynomial with real coefficients. If M is symmetric, then $q(M)$ is symmetric. Also, $\sigma(q(M)) = q(\sigma(M))$. So

$$\|q(M)\| = \sup_{\lambda\in\sigma(M)} |q(\lambda)|.$$

Let f be continuous on $\sigma(M)$. Extend f continuously to an interval I containing $\sigma(M)$. By the Weierstrass approximation theorem, there are polynomials q_n converging uniformly to f on I. So q_n is a Cauchy sequence. Given ε, there exists N such that

$$\sup_{\lambda\in I} |q_n(\lambda) - q_m(\lambda)| < \varepsilon$$

if $n, m \geq N$. Then

$$\|q_n(M) - q_m(M)\| < \varepsilon.$$

So $\lim_{n\to\infty} q_n(M)$ exists, and we call the limit $f(M)$.

Proposition 26.6 *(1) We have* $(f+g)(M) = f(M)+g(M)$ *and* $(fg)(M) = f(M)g(M)$.

(2) $\|f(M)\| = \sup_{\lambda\in\sigma(M)} |f(\lambda)|$.

(3) $f(M)$ *is symmetric and* $\sigma(f(M)) = f(\sigma(M))$.

Proof. (1) This holds for polynomials, so take limits.

(2)
$$\|f(M)\| = \lim \|q_n(M)\| = \lim \sup_{\lambda\in\sigma(M)} |q_n(\lambda)| = \sup |f(\lambda)|.$$

(3) $q_n(M)$ is symmetric, and then

$$(x, f(M)y) = \lim(x, q_n(M)y) = \lim(q_n(M)x, y) = (f(M)x, y).$$

Finally, since

$$\text{dist}\,(\sigma(q_n(M)), \sigma(f(M)) \leq \|q_n(M) - f(M)\| \to 0,$$

$\sigma(f(M))$ is the limit of $\sigma(q_n(M)) = q_n(\sigma(M))$, and this limit is $f(\sigma(M))$. □

M is positive if $(Mx, x) \geq 0$ for all x.

Proposition 26.7 *Let M be bounded and symmetric. M is positive if and only if $\sigma(M) \geq 0$.*

Proof. If $\sigma(M) \geq 0$, then $f(\lambda) = \sqrt{\lambda}$ is a continuous function for $\lambda \geq 0$, and so $N = \sqrt{M}$ exists. Then

$$(Mx, x) = (N^2 x, x) = (Nx, Nx) \geq 0.$$

If M is positive, $\sigma(M) \subset [a, \infty)$, where $a = \inf_{\|x\|=1}(x, Mx) \geq 0$. □

Corollary 26.8 *Every positive symmetric operator has a positive symmetric square root.*

As mentioned before, this allows us to write $T = UA$, where A is positive and symmetric and U is an isometry on R_A and 0 on $R_A^\perp$.

26.3 Spectral resolution

Fix $x, y \in H$ and define $\ell_{x,y}(f) = (f(M)x, y)$ for $f \in C(\sigma(M))$. Recall $\sigma(M)$ is bounded and closed, hence compact. ℓ is a linear functional and by Riesz representation, there exists a complex valued measure $m_{x,y}$ such that

$$(f(M)x, y) = \int_{\sigma_M} f(\lambda) m_{x,y}(d\lambda).$$

Proposition 26.9 *(1) $m_{x,y}$ is linear in x and skew linear in y.*

(2) $m_{y,x} = \overline{m_{x,y}}$.

(3) The total variation of $m_{x,y}$ is less than or equal to $\|x\|\,\|y\|$.

(4) The measure $m_{x,x}$ is real and non-negative.

Proof. (1) $\ell_{x,y}$ is linear in x and skew linear in y. $m_{x+z,y}$ and $m_{x,y}+m_{z,y}$ both represent $\ell_{x+z,y}$, and by the uniqueness of the Riesz representation, $m_{x,y}+m_{z,y}=m_{x+z,y}$.

(2) Since M is symmetric, $(f(M)x,y)$ is skew symmetric.

(3) The total variation is the same as the norm of $\ell_{x,y}$. But

$$|\ell_{x,y}(f)| = |(f(M)x,y)| \le \|f(M)\|\,\|x\|\,\|y\| \le (\sup_\lambda |f|)\|x\|\,\|y\|.$$

(4) f real implies $\sigma(f(M)) = f(\sigma(M)) \in [0,\infty)$. So $f(M)$ is a positive operator. Then $\ell_{x,x}(f) = (f(M)x,x) \ge 0$. □

If $S \subset \sigma(M)$, then $m_{x,y}(S)$ is a bounded symmetric functional of x and y. So there exists a bounded symmetric operator $E(S)$ such that

$$m_{x,y}(S) = (E(S)x,y).$$

Proposition 26.10 *(1) $E^*(S) = E(S)$.*

(2) $\|E(S)\| \le 1$.

(3) $E(\emptyset) = 0, E(\sigma(M)) = I$.

(4) If S,T are disjoint, $E(S\cup T) = E(S)+E(T)$.

(5) $E(S)$ and M commute.

(6) $E(S\cap T) = E(S)E(T)$.

(7) $E(S)$ is an orthogonal projection. If S,T are disjoint, then $E(S), E(T)$ are orthogonal.

(8) $E(S)$ and $E(T)$ commute.

Proof. (1) $E(S)$ is symmetric.

(2) From the fact that the total variation of $m_{x,y}$ is bounded by $\|x\|\,\|y\|$.

(3) $m_{x,y}(\emptyset) = 0$, so $E(\emptyset) = 0$. If $f \equiv 1$, then $f(M) = I$, and

$$(x,y) = \int_{\sigma(M)} m_{x,y}(d\lambda) = (E(\sigma(M))x,y).$$

This is true for all y, so $x = E(\sigma(M))x$ holds for all x.

(4) Because $m_{x,y}$ is additive.

(5)

$$m_{Mx,y}(f) = (f(M)Mx, y) = (Mf(M)x, y) = (f(M)x, My) = m_{x,My}(f),$$

where $m_{x,y}(f)$ denotes the integral of f with respect to $m_{x,y}$.

(6) (Lax never really does this one clearly.) Since $m_{x,y}$ is a finite measure, we can approximate $(E(S)x, y) = m_{x,y}(S)$ by $m_{x,y}(f)$. So it suffices to show $E(f)E(g) = E(fg)$ for f, g continuous and use approximations.

Now

$$\begin{aligned}(E(f)E(g)x, y) &= \int f(\lambda)\, m_{E(g)x,y}(d\lambda) = (f(M)E(g)x, y) \\ &= (E(g)x, f(M)y) = \int g(\lambda)\, m_{x,f(M)y}(d\lambda) \\ &= (g(M)x, f(M)y) = (f(M)g(M)x, y) \\ &= ((fg)(M)x, y) = \int (fg)(\lambda)\, m_{x,y}(d\lambda) \\ &= (E(fg)x, y).\end{aligned}$$

This is true for all y, so $E(f)E(g)x = E(fg)x$.

(7) Setting $S = T$ in (6) shows $E(S) = E^2(S)$, so $E(S)$ is a projection. IF $S \cap T = \emptyset$, then $E(S)E(T) = E(\emptyset) = 0$, so they are orthogonal.

(8)

$$E(S)E(T) = E(S \cap T) = E(T \cap S) = E(T)E(S).$$

□

We have proved most of the following:

Theorem 26.11 *Let H be a complex-valued Hilbert space and M a bounded symmetric operator. There exists a projection valued measure E such that $E(S \cap T) = E(S)E(T)$,*

$$f(M) = \int_{\sigma(M)} f(\lambda)\, E(d\lambda),$$

and the measure E is unique.

A few remarks.

(1) Uniqueness follows from the uniqueness of $m_{x,y}$.

(2) Suppose f is bounded and measurable. If f is simple, i.e., $f = \sum c_i \chi_{A_i}$, where the A_i are disjoint, we define $f(M) = \sum c_i E(A_i)$. In the next proposition, we show

$$\|f(M)\| = \sup_{\lambda \in \sigma(M)} |f(\lambda)|$$

if f is simple. If f is bounded and measurable, we can take f_n simple converging to f uniformly. Then

$$\|f_n(M) - f_m(M)\| = \sup_{\lambda \in \sigma(M)} |f_n - f_m|,$$

since $f_n - f_m$ is simple, and therefore $f_n(M)$ is a Cauchy sequence. We define $f(M)$ to be the limit of $f_n(M)$.

(3) We showed that the above equality holds in the weak sense:

$$(f(M)x, y) = \cdots .$$

But in fact the equality holds in the norm topology. One needs to show that the integral exists, and to do that one needs to approximate by Riemann-Stieltjes integrals. The key estimate is the next proposition.

(4) If we write $m_{x,y} = m^P_{x,y} + m^S_{x,y} + m^C_{x,y}$, where this is the decomposition of the measure into pure point part, singular part, and absolutely continuous part, we get corresponding operators E^P, E^S, E^C and we can write

$$H = R_{E^P} \oplus R_{E^S} \oplus R_{E^C}.$$

Here is the promised proposition.

Proposition 26.12 *Suppose A_i are disjoint. Then*

$$\|\sum c_i E(A_i)\| = \max |c_i|.$$

Proof. Let $r = \max_i |c_i|$. Given x, let $x_i = E(A_i)x$. Then

$$(x_i, x_j) = (E(A_i)x, E(A_j)x) = (x, E(A_i)E(A_j)x) = 0$$

if $i \neq j$. Then

$$\Big| \sum_i c_i E(A_i) x \Big|^2 = \Big(\sum_i c_i E(A_i) x, \sum_j c_j E(A_j) x \Big) = \Big(\sum_i c_i x_i, \sum_j c_j x_j \Big)$$
$$= \sum_i |c_i|^2 \|x_i\|^2 \le r^2 \sum_i \|x_i\|^2 \le r^2 \|x\|^2.$$

Therefore the operator norm is less than or equal to r. If j is such that $|c_j| = r$, then take x in the range of $E(A_j)$, and then $\sum_i c_i E(A_i)x = c_j x$, which implies that the norm is equal to r. □

Proposition 26.13

$$\|f(M)x\|^2 = \int |f(\lambda)|^2 \, m_{x,x}(d\lambda).$$

Proof. If f is real

$$\|f(M)x\|^2 = (f(M)x, f(M)x) = ((f(M))^2 x, x) = \int f^2(\lambda) \, m_{x,x}(d\lambda),$$

and the proof for complex f is similar. □

26.4 Normal operators

Let $\mathcal{F}$ be a Banach algebra with unit. Then if $Q \in \mathcal{F}$,

$$\sigma(Q) = \{p(Q) : p \text{ a homomorphism of } \mathcal{F} \text{ into } \mathbb{C}\}.$$

The reason for this is that $\lambda \in \sigma(Q)$ if and only if $\lambda I - Q$ is not invertible, which happens if and only if $p(\lambda I - Q) = 0$ for some homomorphism p. $p(I) = 1$, so this happens if and only if $\lambda = p(Q)$ for some p.

Proposition 26.14 *If p is a homomorphism, then $p(T^*) = \overline{p(T)}$.*

Proof. Let $A = (T + T^*)/2$ and $B = (T - T^*)/2$. Then $A^* = A$, $B^* = -B$, $T = A + B$, $T^* = A - B$, and so $p(T) = p(A) + p(B)$ and similarly with T replaced by T^*.

It will suffice to show $p(A)$ is real and $p(B)$ is imaginary. Write $p(A) = a + ib$ and let $U = A + itI$, so that $U^* = A - itI$. Then

$$U^*U = A^2 + t^2 I.$$

We have $p(U) = a + i(b+t)$, so $|p(U)|^2 = a^2 + (b+t)^2$. We have $|p(U)| \le \|U\|$, and hence

$$a^2 + (b+t)^2 \le \|A\|^2 + t^2$$

for all t, which can only happen if $b = 0$. (If $b > 0$, take t large positive, and $t$large negative if $b < 0$.) The operator iB is self-adjoint, so apply the above to iB. □

Proposition 26.15 *If T and T^* commute, then $\|T\| = |\sigma(T)|$.*

Proof. We already know this if T is self-adjoint. For general T,

$$p(T^*T) = p(T^*)p(T) = \overline{p(T)}p(T) = |p(T)|^2.$$

Every point in $\sigma(T)$ is of the form $p(T)$ for some homomorphism p, so

$$|\sigma(T^*T)| = |\sigma(T)|^2.$$

We also know that $\|T^*T\| = |\sigma(T^*T)|$ since T^*T is symmetric. But since

$$\|Tx\|^2 = |(x, T^*Tx)| \le \|x\|^2 \|T^*T\|,$$

then $\|T\|^2 \le \|T^*T\|$. We have $\|T^*T\| \le \|T\|^2$, and so we obtain $\|T\|^2 = |\sigma(T)|^2$. □

An operator N is normal if $N^*N = NN^*$.

Theorem 26.16 *Let N be normal. There exists an orthogonal projection valued measure E on $\sigma(N)$ such that $I = \int_{\sigma(N)} dE$ and $N = \int_{\sigma(N)} \lambda\, E(d\lambda)$.*

Proof. Let $q(x,y)$ be a polynomial in x and y. If we let $w = x + yi \in \mathbb{C}$, we can let $x = (w + \overline{w})/2$, $y = (w - \overline{w})/2$, and write $q(x,y) = R(w,\overline{w})$ for some polynomial R. Set $Q = R(N, N^*)$. Since N and N^* commute, they each commute with Q, and so Q and Q^* commute. In the lemma below we will show $\sigma(Q) = R(\lambda, \overline{\lambda})$ for $\lambda \in \sigma(N)$. We have $\|Q\| = |\sigma(Q)|$, since Q and Q^* commute. Therefore

$$\|Q\| = \sup_{\lambda \in \sigma(N)} |R(\lambda, \overline{\lambda})|.$$

Now we can define $f(N)$ as the limit of polynomials, and the rest of the proof is as before.

Lemma 26.17

$$\sigma(Q) = \{R(\lambda, \overline{\lambda}) : \lambda \in \sigma(N)\}.$$

Proof. Operators of the form $R(N, N^*)$ are a commutative algebra with unit. Let $\mathcal{F}$ be the closure in the operator norm.

Now $p(Q) = R(p(N), p(N^*)) = R(p(N), \overline{p(N)})$. Then $\sigma(Q)$ is equal to the set of points $R(p(N), \overline{p(N)})$ where p is a homomorphism, which is the same as the set of $R(\lambda, \overline{\lambda})$ where $\lambda \in \sigma(N)$. □

26.5 Unitary operators

U is unitary if it is linear, isometric, 1-1, and onto. (Cf. rotations) So $\|Ux\| = \|x\|$, or $(Ux, Ux) = (x, x)$. By polarization, $(Ux, Uy) = (x, y)$, so $(x, U^*Uy) = (x, y)$, which implies $U^*U = I$. U is invertible, since it is 1-1 and onto, and thus $U^{-1} = U^*$.

It is an exercise to show that $\sigma(U) \subset \{z : |z| = 1\}$.

$U^*U = I = UU^*$, so unitary operators are also normal operators.

27 Spectral theory of unbounded operators

An example: look at $\{f : f \in C^2, f(0) = f(1) = 0\}$. By integration by parts,

$$\int_0^1 f''g = -\int_0^1 f'g' = \int_0^1 fg'',$$

or $(Lf, g) = (f, Lg)$. Then L, the second derivative operator, is symmetric, but is an unbounded operator. Writing $Lf = \lambda f$, there are solutions satisfying the boundary condition only if $\lambda < 0$ and $\lambda = -n^2\pi^2$. The corresponding eigenfunctions are $\sin n\pi x$, although these are unnormalized. $\{-n^2\pi^2\}$ is unbounded, so L is not a bounded operator on L^2.

Proposition 27.1 *H a Hilbert space over $\mathbb{C}$, M self-adjoint. If M is defined everywhere on H, then M is bounded.*

Proof. M is closed: if $x_n \to x$ and $Mx_n \to u$, then

$$(Mx_n, y) = (x_n, My) \to (x, My) = (Mx, y).$$

Also $(Mx_n, y) \to (u, y)$. True for all y, so $Mx = u$.

By the closed graph theorem, M is bounded. □

If H is a Hilbert space over $\mathbb{C}$ and D a dense subspace with A defined on D, then D^* is the set of $v \in H$ for which there exists a vector $A^*v \in H$ such that $(Au, v) = (u, A^*v)$ for all $u \in D$. Since D is dense, $A * v$ is uniquely defined. A is self-adjoint if $D = D^*$ and $A^* = A$. (So $(Au, v) = (u, Av)$ for all $u, v \in D(A)$, and the domain cannot be enlarged.)

Our goal is to prove

Theorem 27.2 *Let A be self-adjoint, D, H be as above. There exists a projection valued measure E such that*

(1) $E(\emptyset) = 0$, $\mathbb{E}(\mathbb{R}) = I$.

(2) $E(S \cap T) = E(S)E(T)$.

(4) $E^*(S) = E(S)$.

(5) E commutes with A.

(6) $D = \{u : \int \lambda^2 (E(d\lambda)u, u) < \infty\}$ and $Au = \int \lambda E(d\lambda)u$.

We say z is in the resolvent set if $A - zI$ maps D one-to-one onto H.

Proposition 27.3 *If z is not real, then z is in the resolvent set.*

Proof. (1) $R = \text{Range}\,(A - zI)$ is a closed subspace.

R is equal to the set of all vectors u of the form $Av - zv = u$ for some $v \in D$. Then$(Av, v) - z(v, v) = (u, v)$. A is self-adjoint, so (Av, v) is real. Looking at the imaginary parts,

$$-\text{Im}\,(z, \|v\|^2) = \text{Im}\,(u, v),$$

so $|\text{Im}\, z| \|v^2\| \leq \|u\|\,\|v\|$, or

$$\|v\| \leq \frac{1}{|\text{Im}\, z|} \|u\|.$$

If $u_n \in R$ and $u_n \to u$, then $\|v_n - v_m\| \leq (1/|\text{Im}\, z|)\|u_n - u_m\|$, so v_n is a Cauchy sequence, and hence converges to some point v.

Since $Av_n - zv_n = u_n \to u$ and zv_n converges to zv, then Av_n converges, say to r, and then $r - zv = u$. Since $(Av_n, w) = (v_n, Aw)$ for $w \in D$, then $(r, w) = (v, Aw)$, which implies $r \in D$ and $Av = r$.

(2) $R = H$. If not, there exists $k \neq 0$ such that k is orthogonal to R, and then

$$(Av - zv, k) = (Av, k) - (v, \overline{z}k) = 0$$

for all $v \in D$. Then $(Av, k) = (v, \overline{z}k)$, so $k \in D$ and $Ak = \overline{z}k$. But then $(k, Ak) = z(k, k)$ is not real, a contradiction.

(3) $A - zI$ is one-to-one. If not, there exists $k \in D$ such that $(A - zI)k = 0$. But then $\|k\| \leq (1/|\text{Im}\, z|)\|0\| = 0$, or $k = 0$. □

If we set $R(z) = (A - zI)^{-1}$ the resolvent, we have

$$\|R(z)\| \leq \frac{1}{|\text{Im}\, z|}.$$

If $u, w \in H$ and $v = R(z)u$, then $(A - z)v = u$, and

$$(u, R(\overline{z})w) = ((A - z)v, R(\overline{z})w) = (v, ((A - \overline{z})R(\overline{z})w) = (v, w) = (R(z)u, w).$$

So the adjoint of $R(z)$ is $R(\overline{Z})$.

27.1 Cayley transform

Define

$$U = (A - i)(A + i)^{-1}.$$

This is the image of A under the function

$$F(z) = \frac{z - i}{z + i},$$

which maps the real line to $\partial B_1(0) \setminus \{1\}$.

Proposition 27.4 *U is a unitary operator.*

Proof. $A + i, A - i$ map $D(A)$ one-to-one onto H, so U maps H onto itself.

U is norm preserving: Let $u \in H$, $v = (A+i)^{-1}u$, $w = Uu$. So $(A+i)v = u$, $(A - i)v = w$. Then

$$\begin{aligned} \|u\|^2 = ((A + i)v, (A + i)v) &= \|Av\|^2 + \|v\|^2 + i[(v, Av) - (Av, v)] \\ &= \|Av\|^2 + \|v\|^2, \end{aligned}$$

and similarly

$$\|w\| = ((A - i)v, (A - i)v) = \|Av\|^2 + \|v\|^2.$$

□

Proposition 27.5 *If U is unitary, then $\sigma(U) \subset \{|z| = 1\}$.*

Proof. $(\lambda I - U) = \lambda(I - U/\lambda)$. Since U is an isometry, then $\|U\| = 1$. Then $I - \frac{1}{\lambda}U$ is invertible if $\frac{1}{|\lambda|}\|U\| < 1$, or if $|\lambda| > 1$.

$(\lambda I - U) = U(\lambda U^{-1} - I) = U(\lambda U^* - I)$. Since $\|\lambda U^*\| = |\lambda| < 1$, then $I - \lambda U^*$ is invertible. □

Proposition 27.6 *Given A and U as above and E the spectral resolution for U, $E(\{1\}) = 0$.*

Proof. Write E_1 for $E(\{1\})$. If $E_1 \neq 0$, there exists $z \neq 0$ in the range of E_1, so $z = E_1 w$. Then

$$Uz = \int_{\sigma(U)} \lambda\, E(d\lambda)z = \int_{\sigma(U)} \lambda\,(E - E_1)(d\lambda)z + \int_{\{1\}} \lambda\, E_1(d\lambda)z.$$

The first integral is zero since $(E - E_1)(A)$ and E_1 are orthogonal for all A. The second integral is equal to

$$E_1 z = E_1 E_1 w = E_1 w = z$$

since E_1 is a projection.

We conclude z is an eigenvector for U with eigenvalue 1. So $(A - iI)(A + iI)^{-1}z = z$. Let $v = (A + iI)^{-1}z$, or $z = (A + iI)v$. Then

$$z = (A - iI)(A + iI)^{-1}z = (A - iI)v,$$

and then $iv = -iv$, so $v = 0$, and hence $z = 0$, a contradiction. □

Let M be a bounded symmetric operator. Let f be bounded and measurable. Define

$$m_{x,y}(f) = \int_{\sigma(M)} f(\lambda)\, m_{x,y}(d\lambda).$$

This is a bounded symmetric functional of x and y, since it is true for $m_{x,y}(S)$ and we can take limits. So there exists an operator, called $f(M)$, such that

$$m_{x,y}(f) = (f(M)x, y), \qquad x, y \in H.$$

We have

$$|(f(M)x, y)| = |m_{x,y}(f)| \leq |f|\, |m_{x,y}(\sigma_M)| \leq |f|\, \|x\|\, \|y\|.$$

Therefore $\|f(M)\| \leq |f|$. We thus can extend our construction of $f(M)$ from continuous f to bounded and measurable f. We now want to define $f(M)$ for some unbounded functions f.

(The rest of this chapter is from Rudin's *Functional Analysis*.

Proposition 27.7 *Let*

$$D_f = \Big\{x : \int_{\sigma(M)} |f(\lambda)|^2\, m_{x,x}(d\lambda) < \infty\Big\}.$$

Then

(1) D_f is a dense subspace of H.

(2) If $x, y \in H$,

$$\int_{\sigma(M)} |f(\lambda)|\, |m_{x,y}|(d\lambda) \leq \|y\| \Big(\int_{\sigma(M)} |f(\lambda)|^2\, m_{x,x}(d\lambda) \Big)^{1/2}.$$

(3) If f is bounded and $v = f(M)z$, then

$$m_{x,v}(d\lambda) = \overline{f}(\lambda)\, m_{x,z}(d\lambda), \qquad x, z \in H.$$

Proof. (1) Let $S \subset \sigma(M)$ and $z = x + y$.

$$\|E(S)z\|^2 \leq (\|E(S)x\| + \|E(S)y\|)^2 \leq 2\|E(S)x\|^2 + 2\|E(S)y\|^2.$$

So

$$m_{z,z}(S) \leq 2m_{x,x}(S) + 2m_{y,y}(S).$$

This is true for all S, so

$$m_{z,z}(d\lambda) \leq 2m_{x,x}(d\lambda) + 2m_{y,y}(d\lambda).$$

Let $S_n = \{$
$min \sigma(M) : |f(\lambda)| < n\}$. Then if $x = E(S_n)z$, $E(S)x = E(S \cap S_n)x$, so $m_{x,,x}(S) = m_{x,x}(S \cap S_n)$. Then

$$\int_{\sigma(M)} |f(\lambda)|^2\, m_{x,x}(d\lambda) = \int_{S_n} |f(\lambda)|^2\, m_{x,x}(d\lambda) \leq n^2 \|x\|^2 < \infty.$$

Therefore the range of $E(S_n) \subset D(f)$. $\sigma(M) = \cup_n S_n$, so $y = \lim E(S_n)y$, hence y is in the closure of D_f.

(2) If $x, y \in H$, f bounded,

$$f(\lambda)\, m_{x,y}(d\lambda) \ll |f(\lambda)|\, |m_{x,y}|(d\lambda),$$

so there exists u with $|u| = 1$ such that

$$u(\lambda) f(\lambda)\, m_{x,y}(d\lambda) = |f(\lambda)|\, |m_{x,y}|(d\lambda).$$

So
$$\int_{\sigma(M)} |f(\lambda)|\,|m_{x,y}|(d\lambda) = (uf(M)x, y) \le \|uf(M)x\|\,\|y\|.$$
But
$$\|uf(M)x\|^2 = \int |uf|^2\,dm_{x,x} = \int |f|^2\,dm_{x,x}.$$
So (2) holds for bounded f. Now take a limit.

(3) Let g be continuous.
$$\begin{aligned}\int_{\sigma(M)} g\,dm_{x,v} &= (g(M)x, v) = (g(M)x, f(M)z)\\ &= ((\overline{f}g)(M)x, z) = \int g\overline{f}\,dm_{x,z}.\end{aligned}$$
this is true for all g continuous, so $dm_{x,x} = \overline{f}\,dm_{x,z}$. □

Theorem 27.8 *Let E be a resolution of the identity.*

(a) Suppose $f : \sigma(M) \to \mathbb{C}$ is measurable. There exists a densely defined operator $f(M)$ with domain D_f and
$$(f(M)x, y) = \int_{\sigma(M)} f(\lambda)\,m_{x,y}(d\lambda) \tag{1}$$
$$\|f(M)x\| = \int_{\sigma(M)} |f(\lambda)|^2\,m_{x,x}(d\lambda). \tag{2}$$

(b) If $D_{fg} \subset D_g$, then $f(M)g(M) = (fg)(M)$.

(c) $f(M)^* = \overline{f}(M)$ and $f(M)f(M)^* = f(M)^*f(M) = |f|^2(M)$.

Proof. (a) If $x \in D_f$, then $y \to \int_{\sigma(M)} f\,dm_{x,y}$ is a bounded linear functional with norm at most $(\int |f|^2\,dm_{x,x})^{1/2}$. Choose $f(M)x \in H$ to satisfy (1) for all y.

Let $f_n = f1_{(|f|\le n)}$. Then $D_{f-f_n} = D_f$. By dominated convergence theorem,
$$\|f(M)x - f_n(M)x\|^2 \le \int_{\sigma(M)} |f - f_n|^2\,dm_{x,x} \to 0.$$

Since f_n is bounded, (2) holds with f_n. Now let $n \to \infty$.

(b) and (c): Prove for f_n and let $n \to \infty$. □

Theorem 27.9 *(Change of measure principle) Let E be a resolution of the identity on A, $\Phi : A \to B$ one-to-one and bimeasurable. Let $E'(S') = E(\Phi^{-1}(S'))$. Then E' is a resolution of the identity on B, and*

$$\int_B f\, dm'_{x,y} = \int_B (f \circ \Phi)\, dm_{x,y}.$$

Proof. Prove for f the indicator of a set, use linearity, and take limits. □

Proof of the spectral theorem. Start with the unbounded operator A. Let $U = (A - iI)(A + iI)^{-1}$. Then U is unitary with a spectrum on $\partial B_1(0) \setminus \{1\}$. Let the resolution of the identity for U be given by E'.

Define Φ to be the map taking $\partial B_1(0) \setminus \{1\}$ to $\mathbb{R}$. Apply the change of measure principle. If we let $E(S) = E'(\Phi^{-1}(S))$, it is just a matter of checking that E is a resolution of the identity for A. □

28 Examples of self-adjoint operators

28.1 Extensions

M is symmetric if $(x, My) = (Mx, y)$. Saying M is self-adjoint includes conditions on the domains. The difference only applies for unbounded operators.

C is an extension of B if $D(B) \subset D(C)$ and $Cu = Bu$ for $u \in D(B)$.

Given B symmetric, so that $(Bu, v) = (u, Bv)$ for all $u, v \in D(B)$, can we extend B to a self-adjoint operator?

If $u_n \in D(B)$, $u_n \to u$, and $Bu_n \to w$, then

$$(Bu_n, v) = (u_n, Bv) \to (u, Bv)$$

for all $v \in D(B)$. Also $(Bu_n, v) \to (w, v)$. Since $D(B)$ is dense in H, w is uniquely determined by u. Define $\overline{B}$ by $\overline{B}u = w$ for all u, w such that $(w, v) - (u, Bv)$ for all $v \in D(B)$.

Proposition 28.1 *Let B be a densely defined operator and $\overline{B}$ its closure.*

(1) $\overline{B}$ is a closed operator.

(2) $\overline{B}$ is symmetric.

(3) If $z \notin \mathbb{R}$, then $\overline{B} - z$ maps $D(\overline{B})$ one-to-one onto a closed subspace of H.

Proof. (1) is easy.

(2) If $v \in D(\overline{B})$, choose $v_n \in D(B)$ such that $v_n \to v$ and $Bv_n \to \overline{B}v$. $(\overline{B}u, v_n) = (u, Bv_n)$. Let $n \to \infty$, and $(\overline{B}u, v) = (u, \overline{B}v)$.

(3) If $u \in D(\overline{B})$, let $f = (\overline{B} - z)u$. So

$$(\overline{B}u, u) - z(u, u) = (f, u).$$

Since $\overline{B}$ is symmetric, the first term is real. Taking imaginary parts,

$$|\mathrm{Im}\, z|\, \|u\|^2 = |\mathrm{Im}\,(f, u)| \le \|f\|\, \|u\|,$$

so

$$\|u\| \le \frac{1}{|\mathrm{Im}\, z|} \|f\|.$$

So $\overline{B} - z$ is one-to-one.

If f_n is in the range of $\overline{B} - z$ and $f_n \to f$, then write $(\overline{B} - z)u_n = f_n$. By the above inequality, $\{u_n\}$ is a Cauchy sequence, hence converges, say, to u. So u_n converges, f_n converges, hence $\overline{B}u_n$ converges. Since $\overline{B}$ is closed, $u \in D(B)$ and $\overline{B}u = f + zu$. □

Corollary 28.2 *If A is self-adjoint, then A is closed.*

Proof. $\overline{A}$ is symmetric, so $D(\overline{A}) \subset D(A^*) = D(A)$. □

Theorem 28.3 *Let A be a symmetric operator. A is self-adjoint if and only if $\mathbb{C} \setminus \mathbb{R} \subset \rho(A)$.*

Proof. That A self-adjoint implies that all non-real z are in the resolvent set has already been proved.

Suppose z is non-real, so $z, \overline{z} \in \rho(A)$.

We claim $(A-z)^{-1}$ is the adjoint of $(A-\overline{z})^{-1}$. We need to show that if $f, g \in H$, then

$$((A-z)^{-1}f, g) = (f, (A-\overline{z})^{-1}g).$$

Let $x = (A-z)^{-1}f$ and $y = (A-\overline{z})^{-1}g$. So we need to show

$$(x, (A-\overline{z})y) = ((A-zx, y).$$

Since A is symmetric, this is true for all $x, y \in D(A)$. Since $A - z$ and $A - \overline{z}$ map $D(A)$ one-to-one onto H, this is true for all $f, g \in H$.

A is self-adjoint: we have to show that if $v \in D(A^*)$, then $v \in D(A)$ and $A^*v = Av$. Suppose $v \in D(A^*)$ with $w = A^*v$. Then $(Ax, v) = (x, w)$ for all $x \in D(A)$, or $((A-z)x, v) = (x, w - \overline{z}v)$. Let $x = (A-z)^{-1}f$, $g = x - \overline{z}v$ so

$$(f, v) = ((A-z)^{-1}f, w - \overline{z}v) = (f, (A-\overline{z})^{-1}(w - \overline{z}v)).$$

This must hold for all $f \in H$, so $v = (A-\overline{z})^{-1}(w - \overline{z}v)$. Since the range of $(A-\overline{z})^{-1}$ is $D(A)$, then $v \in D(A)$. Hit both sides with $A - \overline{z}$, and then $Av = w$. □

28.2 Examples

1. Let $H = L^2(\mathbb{R})$, $B = i(d/dx)$, and $D(B) = C_0^1$, the C^1 functions with compact support.

Proposition 28.4 *B is symmetric and its closure is self-adjoint.*

Proof. Symmetry follows by integration by parts.

Let $z \in \mathbb{C}$. The range of $B - z$ is $\{f : iu' - zu = f, u \in C_0^1\} = C$.

$$\frac{d}{dx} i(e^{izx}u) = e^{izx} f,$$

or

$$0 = \int_{-\infty}^{\infty} e^{izx} f(x)\, dx,$$

using that u has compact support.

Conversely, if the equality holds, define

$$u(x) = -i \int_{-\infty}^{x} e^{iz(y-x)} f(y)\, dy.$$

Then $u \in C^1$ and if f has compact support, then u has the same support. C is dense in L^2. If $z \notin \mathbb{R}$, range of $\overline{B} - z$ is closed, so range is all of H. Since $\overline{B} - z$ is 1-1, $z \in \rho(\overline{B})$, and therefore $\overline{B}$ is self-adjoint. □

2. $H = L^2[0, \infty)$, B as before, $D(B)$ as before, but now C_0^1 means support in $(0, \infty)$.

Proposition 28.5 *B is symmetric, but $\overline{B}$ is not self-adjoint. B has no self-adjoint extension.*

Proof. $f \in D(B)$ implies $f = 0$ in a neighborhood of 0, so we can use integration by parts as before to show the symmetry.

As before, if f is in the range of $B - z$, $0 = \int_0^\infty e^{izx} f(x)\, dx$. If $\operatorname{Im} z < 0$, as before the range of $\overline{B} - z$ is dense in z.

If $\operatorname{Im} z > 0$, e^{izx} is square integrable, and therefore the range of $\overline{B} - z$ is the set of $f \in H$ such that f is orthogonal to e^{izx}. So $\overline{B}$ is not self-adjoint.

If A is a self-adjoint extension, A would be an extension of $\overline{B}$. Let $v \in D(A) \setminus D(\overline{B})$. Let $\operatorname{Im} z < 0$. $\overline{B} - z$ maps $D(\overline{B})$ onto H, so there exists $u \in D(\overline{B})$ such that

$$(\overline{B} - z)u = (A - z)v.$$

A is an extension, so $(A - z)(v - u) = 0$. A is symmetric, so eigenvectors only exist if $z \in \mathbb{R}$, so $v - u = 0$, a contradiction. □

29 Semigroups

Let X be a Banach space over the complex numbers, $Z(t) = Z_t$ linear bounded operators for $t \geq 0$. Z is a semigroup if $Z(t+s) = Z(t)Z(s)$, $Z(0) = I$.

Proposition 29.1 *(1) Let $G : X \to X$ be bounded. Then $Z(t) = e^{tG}$ (defined as $e^{tG} = \sum t^n G^n/n!$) is a semigroup that is continuous in the norm topology.*

(2) If $Z(t)$ is a semigroup and $\lim(Z(t) - I) = 0$, *then $Z = e^{tG}$ for some G.*

Proof. (1) This is functional calculus for operators.

(2) Define

$$\log Z = \log(I + Z - I) = (Z - I) - \frac{(Z - I)^2}{2} + \cdots .$$

So at least for t small we can define $\log Z$.

Choose a such that $|Z(t) - I| < \frac{1}{3}$ for $t < a$, and define $L(t) = \log Z(t)$. $Z(t)$ and $Z(s)$ commute so $L(t+s) = L(t) + L(s)$. So for t rational with $t < a$, $\frac{1}{t}L(t)$ does not depend on t. Let $G = \frac{1}{t}L(t)$, so $L(t) = tG$.

$$Z(t+h) - Z(t) = Z(t)[Z(h) - I],$$

so Z is continuous in t. So $L(t) = tG$ for all $t < a$. Then $e^{tG} = Z(t)$ for $t < a$. By the semigroup property, this holds for all t. □

29.1 Strongly continuous semigroups

To use this in PDE, we need weaker conditions. $Z(t)$ is strongly continuous at $t = 0$ if $|Z(t)x - x| \to 0$ as $t \to 0$ for all $x \in X$.

Proposition 29.2 *Suppose $Z(t)$ is a strongly continuous semigroup at 0.*

(1) There exists b and k such that $|Z(t)| \leq be^{kt}$.

(2) $Z(t)x$ is strongly continuous in t for all $x \in X$.

Proof. We claim $|Z(t)|$ is bounded near 0. If not, there exists $t_j \to 0$ such that $|Z(t_j)| \to \infty$. By the uniform boundedness principle, $Z(t_j)x$ cannot converge to x for all x, a contradiction to strong continuity. So there exists a, b such that $|Z(t)| \leq b$ for $t \leq a$.

Write $t = na + r$. $Z(t) = Z(a)^n Z(r)$, so

$$|Z(t)| \leq |Z(a)|^n |Z(r)| \leq b^{n+1} \leq be^{kt}$$

with $k = \frac{1}{a} \log b$.

(2) $Z(t)x - Z(s)x = Z(s)[Z(t-s)x - x]$, so

$$|Z(t)x - Z(s)x| \leq |Z(s)|\,|Z(t-s)x - x| \to 0.$$

□

Suppose D is dense in X and $G : D \to X$ is closed. $z \in \rho(G)$, the resolvent set, if $z - G$ maps $D = D(G)$ 1-1 onto X. Write $R(z) = R_z = (zI - G)^{-1}$. Since G is closed, then $R(z)$ is closed. $R(z)$ is defined on all of X, so by the closed graph theorem, $R(z)$ is a bounded operator.

We define G' by $(Gx, \ell) = (x, G'\ell)$, where $D(G')$ is the set of ℓ such that (Gx, ℓ) is a bounded linear functional of $x \in D(G)$.

Let Z be a strongly continuous one parameter semigroup. The infinitesimal generator of G is defined by

$$Gx = s - \lim_{h\to 0} \frac{Z(h)x - x}{h},$$

with the domain of G being those x for which the strong limit exists.

Proposition 29.3 *(1) G commutes with $Z(t)$ in the sense that if $x \in D(G)$, then $Z(t)x \in D(G)$ and $GZ(t)x = Z(t)Gx$.*

(2) $D(G)$ is dense in X.

(3) $D(G^n)$ is dense.

(4) G is closed.

(5) If $|Z(t)| \leq be^{kt}$ and $\operatorname{Re} z > k$, then $z \in \rho(G)$. The resolvent of G is the Laplace transform of $Z(t)$.

Proof. (1)

$$\frac{Z(t+h)-Z(t)}{h}x = Z(t)\frac{Z(h)-I}{h}x = \frac{Z(h)-I}{h}Z(t)x.$$

If $x \in D(G)$, the middle term converges to $Z(t)Gx$. So the limit exists in the third term, and $Z(t)x \in D(G)$. Moreover $\frac{d}{dt}Z(t)x = Z(t)Gx = GZ(t)x$.

(2) We claim

$$Z(t)x - x = G\int_0^t Z(s)x\,ds.$$

To see this, $Z(s)x$ is a continuous function of s. By a Riemann sum approximation,

$$\begin{aligned}
\frac{Z(h)-I}{h}\int_0^t Z(s)x\,ds &= \frac{1}{h}\int_0^t [Z(s+h)x - Z(s)x]\,ds \\
&= \frac{1}{h}\int_t^{t+h} Z(s)x\,ds - \frac{1}{h}\int_0^h Z(s)x\,ds \\
&\to Z(t)x - x.
\end{aligned}$$

So $\int_0^t Z(s)x\,ds \in D(G)$. But $\frac{1}{t}\int_0^t Z(s)x\,ds \to x$.

(3) Let ϕ be C^∞ and supported in $[0,1]$. Let

$$x_\phi = \int \phi(s)Z(s)x\,ds.$$

Then $Gx_\phi = -\int \phi'(s)Z(s)x\,ds$. Repeating, $x_\phi \in D(G^n)$. Take ϕ_j approximating the identity.

(4) $Z(t)x - x = \int_0^t Z(s)Gx\,ds$: To see this, both are 0 at 0. The derivative on the left is $Z(t)Gx$, which is the same as the derivative on the right. Let $x_n \in D(G)$, $x_n \to x$, $Gx_n \to y$. Then

$$Z(t)x_n - x_n = \int_0^t Z(s)Gx_n\,ds \to \int_0^t Z(s)y\,ds.$$

The left hand term converges to $Z(t)x - x$. Divide by t and let $t \to 0$. The right hand side converges to y. Therefore $x \in D(G)$ and $Gx = y$.

(5) Let

$$L(z)x = \int_0^\infty e^{-zs}Z(s)x\,ds.$$

The Riemann integral converges when $\operatorname{Re} z > k$.

$$|L(z)x| \leq \int_0^\infty be^{(k-\operatorname{Re} z)s}|x|\,ds \leq \frac{1}{\operatorname{Re} z - k}|x|.$$

We claim $L(z) = R(z)$. Check that $e^{-zt}Z(t)$ is also a semigroup with infinitesimal generator $G - zI$.

$$e^{-zt}Z(t) - x = (G - zI)\int_0^t e^{-zs}Z(s)x\,ds.$$

As $t \to \infty$, the left hand side tends to $-x$ and the right hand side tends to $(G - zI)L(z)x$. Since G is closed, $x = (zI - G)L(z)x$. So $L(z)$ is the right inverse of $(zI - G)$. Similarly, we see that it is also the left inverse. □

29.2 Generation of semigroups

Proposition 29.4 *A strongly continuous semigroup of operators is uniquely defined by its infinitesimal generator.*

Proof. If W, Z have the same generator, let $x \in D(G)$ and

$$\frac{d}{dt}W(t)Z(s-t)x = W(t)GZ(s-t)x - W(t)GZ(s-t)x = 0.$$

Therefore

$$0 = \int_0^s \frac{d}{dr}W(r)Z(s-r)x\,dr = W(s)Z(0)x - W(0)Z(s),$$

or $W(s)x = Z(s)x$. Now use the fact that $D(G)$ is dense. □

$Z(t)$ is a contraction if $|Z(t)| \leq 1$ for all t.

Theorem 29.5 *(1) The infinitesimal generator of a strongly continuous semigroup of contractions has* $(0, \infty) \subset \rho(G)$ *and*

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

(2) (Hille-Yosida theorem) Let G be a densely defined unbounded operator such that $(0, \infty) \subset \rho(G)$ and (1) is satisfied. Then G is the infinitesimal generator of a strongly continuous semigroup of contractions.

Proof. (1) We already did; this is the case $b = 1, k = 0$.

2) Note $nR(n) - I = R(n)G$ since $R(n)(nI - G) = I$. Let $G_n = nGR(n)$. Then $G_n = n^2R(n) - nI$, so G_n is a bounded operator. Define $Z_n(t) = e^{tG_n}$.

Claim: $nR(n)x \to x$ for all x.

To prove this,

$$|nR(n)x - x| = |R(n)G(x)| \leq \frac{1}{n}|Gx|,$$

so the claim is true for $x \in D(G)$. Since $|nR(n)| \leq 1$ and $D(G)$ is dense in X, this proves the claim.

If $x \in D(G)$, then $G_n(x) \to G(x)$:

$$G_nx = nGR(n)x = nR(n)Gx \to Gx.$$

We have

$$Z_n(t) = e^{tG_n} = e^{-nt}e^{n^2R(n)t} = e^{-nt}\sum \frac{(n^2t)^m}{m!}R^m(n),$$

so $|Z_n(t)| \leq e^{nt}e^{nt} = 1$.

G_n and G_m commute with Z_n and Z_m.

$$\frac{d}{dt}Z_n(s-t)Z_m(t)x = Z_n(s-t)Z_m(t)[G_m - G_n]x.$$

The norm of the right hand side is bounded by $|G_nx - G_mx|$. So

$$|Z_n(s)x - Z_m(s)x| \leq s|G_nx - G_mx| \to 0$$

as $n, m \to \infty$. Therefore $Z_n(s)x$ converges, say, to $Z(s)x$, uniformly in s. $D(G)$ is dense so this holds for all x.

$Z_n(s)$ is a strongly continuous semigroup of contractions, so the same holds for $Z(s)$.

It remains to show that G is the infinitesimal generator of Z. We have

$$Z_n(t)x - x = \int_0^t Z_n(s)G_n x\, ds.$$

If $x \in D(G)$, we can let $n \to \infty$ to get

$$Z(t)x - x = \int_0^t Z(s)Gx\, ds.$$

If H is the generator of Z, dividing by t and letting $t \to 0$, we get $D(G) \subset D(H)$ and $H = G$ on $D(G)$. So H is an extension of G. If $\lambda > 0$, then $\lambda \in \rho(G), \rho(H)$, which implies H cannot be a proper extension.

This last assertion is proved as follows. Suppose $x \in D(H) \setminus D(G)$. $(\lambda - H)x \in X$, so

$$(\lambda - G)^{-1}(\lambda - H)x \in D(G) \subset D(H).$$

So

$$(\lambda - H)(\lambda - G)^{-1}(\lambda - H)x = (\lambda - G)(\lambda - G)^{-1}(\lambda - H)x = (\lambda - H)x.$$

Hit both sides with $(\lambda - H)^{-1}$ to obtain $(\lambda - G)^{-1}(\lambda - H)x = x$. So $x \in D(G)$, a contradiction. □

29.3 Perturbation of semigroups

Lemma 29.6 *(Lumer-Phillips) Let G be densely defined in a Hilbert space H and suppose $(0, \infty) \subset \rho(G)$. Then $|R(\lambda)| \leq 1/\lambda$ if and only if $\mathrm{Re}\,(x, Gx) \leq 0$ for all $x \in D(G)$.*

If the last property holds, we say G is dissipative.

Proof.

$$\|(\lambda I - G)^{-1}u\|^2 \leq \frac{1}{\lambda^2}\|u\|^2.$$

Let $x = (\lambda I - G)^{-1}u$. So

$$(x, x) \leq \frac{1}{\lambda^2}(\lambda x - Gx, \lambda x - Gx).$$

This becomes

$$(x, Gx) + (Gx, x) \leq \frac{1}{\lambda}\|Gx\|^2.$$

This is true for all λ.

The converse is left as an exercise. □

Theorem 29.7 *(Trotter) Suppose G is the infinitesimal generator of a semi-group of contractions in a Hilbert space. Let H be a densely defined dissipative operator such that $D(G) \subset D(H)$ and there exist $b > 0$ and $a \in (0,1)$ such that*

$$\|Hx\| \leq a\|Gx\| + b\|x\|, \qquad x \in D(G).$$

Then $G + H$ (defined on $D(G)$) is the generator of a contraction semigroup.

Proof. First, $G + H$ is closed: Let $x_n \to x$ and $y_n = (G + H)x_n \to y$. So

$$G(x_n - x_m) = y_n - y_m - H(y_n - y_m),$$

and

$$\|G(x_n - x_m)\| \leq \|y_n - y_m\| + a\|G(x_n - x_m)\| + b\|x_n - x_m\|.$$

Since $a < 1$, then Gx_n converges. Therefore Hx_n converges. G is closed, so $Gx_n \to Gx$. If $x \in D(G) \subset D(H)$,

$$\|H_n x - Hx\| \leq a\|G_n x - Gx\| + b\|x_n - x\| \to 0.$$

Next, if λ is sufficiently large, then $\lambda \in \rho(G + H)$: By the Lumer-Phillips lemma, G is dissipative. H is also. So $G + H$ is dissipative. By Lumer-Phillips,

$$\|x\| \leq \frac{1}{\lambda}\|(\lambda I - (G + H))x\|.$$

Therefore the range of $(G + H) - \lambda I$ is closed.

The range is H; if not, there exists $v \neq 0$ perpendicular to the range. $G - \lambda I$ is invertible, so there exists $x \in D(G)$ such that $(G - \lambda I)x = v$. Then $v + Hx$ is in the range, or $(v + Hx, v) = 0$. So $\|v\|^2 + (Hx, v) = 0$, or

$$\|v\|^2 \leq \|Hx\|\,\|v\|,$$

and so $\|v\| \leq \|Hx\|$. Then

$$\|Gx - \lambda x\| \leq \|Hx\| \leq a\|Gx\| + b\|x\|.$$

Squaring and use the fact that G is dissipative,

$$\|Gx\|^2 + \lambda^2\|x\|^2 \leq a^2\|Gx\|^2 + 2ab\|Gx\|\,\|x\| + b^2\|x\|^\cdot$$

$a < 1$, so for λ large enough, $\|x\| = 0$. So $x = 0$ and the range is the whole space.

Now use Lumer-Phillips (not quite, since this only true for large λ.) □

Examples: Non-divergence operators and Levy processes.

30 Groups of unitary operators

We prove Stone's theorem.

Theorem 30.1 *(1) Suppose A is self-adjoint and H is a Hilbert space. There exists a strongly continuous group $U(t)$ of unitary operators with infinitesimal generator iA.*

(2) Given a strongly continuous group of unitary operators, the generator is of the form iA where A is self-adjoint.

Proof. (1) We saw $\|R(z)\| \leq 1/|\operatorname{Im} z|$. The resolvent set of iA contains the positive reals. So iA and $-iA$ satisfy the Hille-Yosida theorem. Let $U(t), V(t)$ be the respective semigroups.

V and U are inverses:

$$\frac{d}{dt}U(t)V(t) = U(t)iAV(t)x - U(t)iAV(t)x = 0.$$

So $U(t)V(t)x$ is independent of t. When $t = 0$, we get x. So $U(t)V(t)x = x$ if $x \in D(A)$. But $D(A)$ is dense.

Both U and V are contractions. Since $U(t)V(t) = I$, they must be norm preserving. This is because

$$\|x\| = \|U(t)V(t)x\| \leq \|V(t)x\| \leq \|x\|,$$

so $\|x\| = \|V(t)x\|$ and similarly with U. Since they are invertible, they are unitary. Define $U(t) = V(-t)$ for $t < 0$.

(2) Let $V(t) = U(-t)$. Then $U(t)$ and $V(t)$ are strongly continuous semigroups of contractions, and the infinitesimal generators are additive inverses. So the generators are $G, -G$.

Since both $G, -G$ are infinitesimal generators, all real numbers except 0 are in the resolvent set of G. Take $x \in D(G)$.

$$\|U(t)x\|^2 = (U(t)x, U(t)x) = \|x\|^2.$$

Take the derivative with respect to t:

$$(Gx, x) + (x, Gx) = 0.$$

Replacing x by $x + y$, we get

$$(Gx, y) = (x, Gy).$$

Replacing y by iy, we see G is antisymmetric. So G^* is an extension of $-G$. $\rho(G^*) = \overline{\rho(G)}$. If $z \neq 0$ and $z \in \mathbb{R}$, then $z \in \rho(G)$, so $z \in \rho(G^*)$. Also $z \in \rho(-G)$. So G^* cannot be a proper extension of $-G$, hence $G^* = -G$. □