

Functional Analysis

Muchang Bahng

November 6, 2022

1 Metric and Normed Linear Spaces

1.1 Metric Spaces

A metric space is a space V with a metric function $d : V \times V \rightarrow \mathbb{R}$ satisfying:

1. $d(x, y) \geq 0$ and equality holds iff $x = y$
2. $d(x, y) = d(y, x)$
3. $d(x, z) \leq d(x, y) + d(y, z)$

We are only used to the Euclidean metric

$$d(x, y) = \sqrt{\sum_i (x_i - y_i)^2}$$

but we introduce some other ones. In addition to the L2 metric, the L_1 metric and L_∞ metric are defined

$$d_1(x, y) = \sum_i |x_i - y_i| \text{ and } d_\infty(x, y) = \max_i \{|x_i - y_i|\}$$

Note that

$$d_\infty(p, q) \leq d_2(p, q) \leq d_1(p, q) \leq 2d_\infty(p, q)$$

Let M be the unit circle in \mathbb{R}^2 , i.e. the set of all pairs of real numbers (x, y) s.t. $x^2 + y^2 = 1$. Then, we can set the metric as

1. $d_1[(x_1, y_1), (x_2, y_2)] := \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2}$, the metric induced by the \mathbb{R}^2 that M is embedded in.
2. Another possible metric is $d_2[p, p']$, the arc length between the points p and p' .

Given $C([0, 1])$, the set of continuous real-valued functions on $[0, 1]$, we can equip it with either of the metrics

$$d_1(f, g) := \max_{x \in [0, 1]} |f(x) - g(x)| \text{ or } d_2(f, g) := \int_0^1 |f(x) - g(x)| dx$$

Another metric is the discrete metric, defined

$$d(p, q) = \begin{cases} 1 & \text{if } p \neq q \\ 0 & \text{if } p = q \end{cases}$$

With this notion of a metric, we can now define convergence: A sequence of elements $\{x_n\}_{n=1}^\infty$ of a metric space (M, d) is said to *converge* to an element $x \in M$ if $d(x, x_n) \rightarrow 0$ as $n \rightarrow \infty$. In the unit circle $M \subset \mathbb{R}^2$, the fact that

$$d_1(p, q) \leq d_2(p, q) \leq \pi d_1(p, q)$$

means that a sequence in M converges under d_1 if and only if it converges under d_2 . However, neither $d_1 < d_2$ nor $d_2 > d_1$ within $C^1([0, 1])$.

1.2 Complete Metric Space

Definition 1.1 (Cauchy Sequence). A sequence of elements $\{x_n\}$ of a metric space (M, d) is called a *Cauchy sequence* if for all $\epsilon > 0$ there exists a $N \in \mathbb{N}$ such that $m, n > N$ implies $d(x_m, x_n) < \epsilon$.

It is clear that any convergent sequence is Cauchy. The fact that a sequence is convergent in (M, d) means that its limit is in M . Let us have a sequence $\{x_n\}$ converging to x and some $\epsilon > 0$. Choose $\epsilon/2$. Then, we know that for any $\epsilon/2 > 0$, there exists a $N \in \mathbb{N}$ such that

$$d(x, x_n) < \frac{\epsilon}{2} \text{ for all } n > N$$

and

$$d(x, x_m) < \frac{\epsilon}{2} \text{ for all } m > N$$

By triangle inequality, there exists a $N \in \mathbb{N}$ such that

$$d(x_n, x_m) \leq d(x, x_n) + d(x, x_m) < \epsilon$$

for all $n, m > N$, thus proving Cauchy. However, the converse is not true. A sequence in $(0, 1)$ may be Cauchy, but it may converge to $1 \notin (0, 1)$ and therefore not convergent.

An equivalent definition of convergence is that given a metric space (M, d) , $x \xrightarrow{d} x$ if and only if for each neighborhood N of x , there exists an M such that $m \geq M$ implies $x_m \in N$. The topology is induced by the metric d .

Definition 1.2. A function f from a metric space X to a topological space Y is continuous if and only if for all open sets $O \subset Y$, $f^{-1}(O)$ is open.

Now given a metric space (M, d) , we can talk about convergence, and this allows us to talk about the completeness of this metric space. Note that we must have a metric in order to define completeness.

Definition 1.3 (Complete Metric Space). A metric space (M, d) is called *complete* if every Cauchy sequence of points in M converges to a limit that is also in M .

Intuitively, a set of complete is there are "no points missing" from it, both inside or at the boundary. However, it is always possible to "fill all the holes", leading to the *completion* of a given space.

1. \mathbb{Q} is not complete, since $\sqrt{2}$ is a limit point of a sequence of rational numbers but is not in \mathbb{Q} .
2. \mathbb{R} is complete, since every Cauchy sequence (a sequence in which successive terms must get arbitrarily close together).
3. Any open set in \mathbb{R} is not complete. Take $(0, 1)$. We can construct a Cauchy sequence that converges to $1 \notin (0, 1)$.
4. Any open set in \mathbb{R}^n is not complete, since we can construct a Cauchy sequence that tends to a boundary point that is not in the open set. Take the unit open ball O , and let $e_1 = (1, 0, \dots, 0)$. Then, we can construct the sequence $\{\frac{n-1}{n}e_1\}$ which converges to $e_1 \notin O$.
5. Let (X, d) be

1.3 Normed Linear Spaces

Definition 1.4 (Normed Linear Space). A *normed linear space* is a vector space $(V, \|\cdot\|)$ over \mathbb{C} with a norm from V to \mathbb{R} satisfying

1. $\|v\| \geq 0$ and equality holds iff $v = 0$
2. $\|\lambda v\| = |\lambda| \|v\|$ for all $\lambda \in \mathbb{C}$
3. $\|v + w\| \leq \|v\| + \|w\|$

A norm induces a metric. A normed linear space is *complete* if it is complete as a metric space with the induced metric.

Definition 1.5 (Bounded Linear Map). A *bounded linear map* from a normed space $(V_1, \|\cdot\|_1)$ to a normed space $(V_2, \|\cdot\|_2)$ is a linear map $T : V_1 \rightarrow V_2$ satisfying that for some $C \geq 0$

$$\|Tv\|_2 \leq C\|v\|_1$$

The smallest such C is called the *norm of T* , written $\|T\|$. WLOG, we can let $\|v\|_1 = 1$. Then, $\|Tv\|_2$ would be bounded by C , and naturally,

$$\|T\| = \sup_{\|v\|_1=1} \|Tv\|_2$$

Some examples of normed vector spaces

1. \mathbb{C}^n with the norm

$$\|(x_1, \dots, x_n)\| := \sqrt{|x_1|^2 + \dots + |x_n|^2}$$

2. $C([0, 1])$ with either the norm

$$\|f\|_\infty = \sup_{x \in [0, 1]} |f(x)| \text{ or } \|f\|_1 = \int_0^1 |f(x)| dx$$

Theorem 1.1. *The following are equivalent for a linear transformation T between two normed vector spaces.*

1. T is continuous at one point.
2. T is continuous at all points.
3. T is bounded.

2 Measure and Lp Spaces

Given a set S , we know that the power set $2^S = \mathcal{P}(S)$ is the collection of all subsets $A \subset S$. But this is not too interesting to work with, since it is too big. There are two restricted notions of this collection of subsets:

1. The first is the topology, which is a collection of sets $\tau_S \subset 2^S$ such that $\emptyset \in \tau_S$, $S \in \tau_S$, it is stable under countable unions, and it is stable under finite intersections.
2. An *algebra* of sets $\mathcal{A} \subset 2^S$ contains \emptyset and S itself, and it is stable under finite union and complementation.
3. A σ -algebra is an algebra, but it is stable under not a finite union, but a countable union. Stability under countable union implies stability under finite union, and so a σ -algebra is an algebra by definition.

Remember that stability under countable unions is a stronger condition than stability under finite unions. For example, the countable union of sets

$$\bigcup_{k=2}^{\infty} \left[\frac{1}{k}, 1 - \frac{1}{k}\right] = (0, 1)$$

but a finite union of them does not equal $(0, 1)$.

Given a set S with its σ -algebra \mathcal{A} , a function $\mu : \mathcal{A} \rightarrow \mathbb{R}_0^+ \cup \{+\infty\}$ is called a *measure*. It satisfies non-negativity: $\mu(E) \geq 0$ for all $E \in \mathcal{A}$, null empty set: $\mu(\emptyset) = 0$, and countable additivity: for all countable collections $\{E_k\}_{k=1}^{\infty}$ of pairwise disjoint sets in \mathcal{A} ,

$$\mu\left(\bigcup_{k=1}^{\infty} E_k\right) = \sum_{k=1}^{\infty} \mu(E_k)$$

In measure theory, a set S with its σ -algebra \mathcal{A} , along with a measure μ , is called a measure space. A probability space $(\Omega, \mathcal{F}, \mathbb{P})$ is a measure space with the additional condition that $\mathbb{P}(\Omega) = 1$, i.e. the

measure of the entire space is 1. An element of Ω is called an outcome, an element $A \in \mathcal{A}$ is called an event, and the measure of that event $\mathbb{P}(A)$ is the probability of that event. Given a measure space (S, \mathcal{A}, μ) where the measure of the entire space S is finite, say 10, we can simply normalize it by defining a new measure $\mathbb{P} = \frac{1}{10}\mu$ to define a probability space. Of course, we cannot do this with the Lebesgue measure of \mathbb{R} since $\mu_{\text{leb}}(\mathbb{R})$ is infinite, but we can do this with the space $(0, 1)$ (this is the uniform distribution).

Now a filtration is a sequence of σ -algebras $\mathcal{F}_i \supset \mathcal{F}_{i+1}$ such that the next one is a subcollection of the next one. The fact that we work with σ -algebras rather than the big power set is an advantage for probability theory since with σ -algebras we can model our knowledge, which is done with filtrations. As events unfold, we have more and more knowledge on what is happening, and so we can construct a filtration that models conditional probabilities. For example, when we flip 3 coins, the outcome space is

$$\Omega_1 = \{000, 001, 010, 011, 100, 101, 110, 111\}$$

and its σ -algebra is, in this case, the power set 2^{Ω_1} . We start off with $\mathcal{F}_1 = 2^{\Omega_1}$. Say that we get 0 on the first coin flip. Then, our outcome space becomes restricted to

$$\Omega_2 = \{000, 001, 010, 011\}$$

and our σ -algebra becomes restricted to $2^{\Omega_2} \subset 2^{\Omega_1}$. This goes on for the second and third flip, say we get 1 and 0, respectively. Then,

$$\Omega_3 = \{010, 011\}, \quad \Omega_4 = \{010\}$$

and so we construct the filtration

$$2^{\Omega_1} \supset 2^{\Omega_2} \supset 2^{\Omega_3} \supset 2^{\Omega_4}$$

where $2^{\Omega_4} = \{\emptyset, \{010\}\}$ with the measure defined to be $\mathbb{P}(\emptyset | \Omega_4) = 0$ and $\mathbb{P}(\{010\} | \Omega_4) = 1$.

Now an L^p space is a normed vector space (with norm denoted $\|\cdot\|_{L^p(X, Y, d\mu)}$) of functions from some measure space X to a Banach space Y

$$L^p(X, Y, d\mu) := \left\{ f : X \longrightarrow Y, \|f\|_{L^p(X, Y, d\mu)} := \left(\int_X \|f\|_Y^p d\mu \right)^{1/p} < \infty \right\}$$

For $p = \infty$, the L^∞ space is defined

$$L^\infty(X, Y, d\mu) := \left\{ f : X \longrightarrow Y, \|f\|_{L^\infty(X, Y, d\mu)} := \inf\{c \text{ s.t. } \|f\|_Y \leq c \text{ almost everywhere in } X\} < \infty \right\}$$

where *almost everywhere* in X means that everywhere except for a set of measure 0. The power of p doesn't affect whether the norm is finite or not, so we don't need to pay attention to it. Like the norms of \mathbb{R}^n , $\|\cdot\|_p$ converges to $\|\cdot\|_\infty$ as $p \rightarrow \infty$.

L^p measures are used to measure the integrability (i.e. blowup) of functions. For example, the integral of $f(x) = \frac{1}{x}$ over both $(0, 1)$ and $(1, \infty)$ both go to infinity since

$$\int_0^1 \frac{1}{x} dx = \log(x) \Big|_0^1 = \infty \text{ and } \int_1^\infty \frac{1}{x} dx = \log(x) \Big|_1^\infty = \infty$$

so $f \notin L^1(0, 1)$ and $f \notin L^1(1, \infty) \implies L^1(0, \infty)$. But note that $f \in L^2(1, \infty)$ since $\int_1^\infty |\frac{1}{x}|^2 dx = (-\frac{1}{x}) \Big|_1^\infty = 1$. For $g(x) = \frac{1}{x^2}$, it is even more infinite over the interval $(0, 1)$ since now, rather than $\log(x) \Big|_0^1$, we are dealing with

$$\int_0^1 \frac{1}{x^2} dx = \left(-\frac{1}{x} \right) \Big|_0^1 = \infty$$

which is an even bigger infinity. On the other hand, $\|g\|_{L^1(1, \infty)} = 1$ is finite. Since $\frac{1}{x} \in L^2(1, \infty)$ but $\frac{1}{x} \notin L^2(0, 1)$, we can reflect it and equally claim that $\frac{1}{x} \in L^{1/2}(0, 1)$ but $\frac{1}{x} \notin L^{1/2}(1, \infty)$. In summary, the concept of L^p spaces gives us a quantitative way of measuring how bad the blowup is for a function over a given interval. That is, $\frac{1}{x}$ blows up at 0, but $\frac{1}{x^2}$ blow up even worse at 0, and this can be studied by what L^p spaces they live in.

Try looking at which spaces some functions live in by doing x^{-m} .

2.1 Inequalities

Theorem 2.1 (Holder's Inequality). *Let p and q be dual indices, i.e. $1/q + 1/p = 1$, with $1 \leq p \leq \infty$. Let $f \in L^p(\Omega)$ and $g \in L^q(\Omega)$. Then the pointwise product, $(fg)(x) = f(x)g(x)$, is in $L^1(\Omega)$ and*

$$\left| \int_{\Omega} fg \, d\mu \right| \leq \int_{\Omega} |f||g| \, d\mu \leq \|f\|_p \|g\|_q$$

3 Linear Algebra Review

In functional analysis, there are two big theorems that we care about, and those are duality theorems. Let's review some linear algebra. First, the *Riesz Representation Theorem*. Let X be a Hilbert space over \mathbb{C} . Then, for any linear functional $l \in X^*$, there exists a unique vector $v \in X$ such that

$$l(x) = \langle x, v \rangle_X$$

for all $x \in X$. That is, any linear functional $l(\cdot)$ can be represented as an inner product of the *first argument* $\langle \cdot, v \rangle$. Remember that over \mathbb{C} , the inner product has first argument linearity and second argument conjugate linearity. This map $l \mapsto v$ is a *natural isomorphism*, but only under the condition that X is a Hilbert space. This does not work when X does not have an inner product.

Let $\mathcal{L}(X, Y)$ be the vector space of linear maps from vector space X to vector space Y . Then, the adjoint map

$$* : \mathcal{L}(X, Y) \longrightarrow \mathcal{L}(Y^*, X^*)$$

is an isometric isomorphism. We construct this isomorphism as such: Given a fixed linear map $T \in \mathcal{L}(X, Y)$, choose some vector $x \in X$. The dual spaces of X, Y automatically exist without invoking anything else. Now choose any vector $l \in Y^*$. Then, there exists a unique $T^*l \in \mathcal{L}(Y^*, X^*)$ such that

$$l(Tx) = (T^*l)(x)$$

for all $x \in X, l \in Y^*$. We can visualize it with the following diagram.

$$x \in X \xrightarrow{T} Y \ni Tx$$

$$T^*l \in X^* \xleftarrow{T^*} Y^* \ni l$$

This is called the *Banach adjoint*. If we assume even further that X and Y are Hilbert spaces, then we can invoke Riesz representation theorem to define an isomorphism $i_Y : Y \rightarrow Y^*$ and $i_X : X \rightarrow X^*$ such that for any given $x \in X$ and $y \in Y$,

$$\begin{aligned} i_X(x) &= l_x \text{ s.t. } l_x(\cdot) = \langle \cdot, x \rangle_X \\ i_Y(y) &= l_y \text{ s.t. } l_y(\cdot) = \langle \cdot, y \rangle_Y \end{aligned}$$

and define the *Hilbert adjoint* to be, $T^H = i_X^{-1} \circ T^* \circ i_Y : Y \rightarrow X$, which satisfies

$$\langle y, Tx \rangle_Y = \langle T^*y, x \rangle_X$$

We differentiate between these two definitions because some texts define the adjoint to be the map from Y to X rather than their duals, but with the Riesz isometry, the Banach and Hilbert adjoints are precisely the same. The big idea is that $*$ is an isometric isomorphism. We will prove only linearity. Assume $A, B \in \mathcal{L}(X, Y)$. Then, we wish to show that $*(A+B) = *(A) + *(B)$. Given $A+B \in \mathcal{L}(X, Y)$, $*(A+B)$ is defined such that for all $x \in X, l \in Y^*$,

$$\begin{aligned} [(A+B)^*(l)](x) &= (l)[(A+B)(x)] \\ &= (l)[Ax + Bx] && \text{(by definition of } A+B) \\ &= l(Ax) + l(Bx) && \text{(by linearity of } l) \\ &= (A^*l)(x) + (B^*l)(x) && \text{(definition of adjoint)} \end{aligned}$$

We can also see that the map $T' = i_Y \circ T \circ i_X^{-1} : X^* \rightarrow Y^*$ is also well-defined. This is not necessarily the inverse of T^* .

So if we have a Hilbert space \mathcal{H} , the adjoint operator is an isometric isomorphism from $\mathcal{L}(\mathcal{H}, \mathcal{H})$ to itself (since by Riesz rep theorem $\mathcal{H}^* \simeq \mathcal{H}$).

$$* : \mathcal{L}(\mathcal{H}, \mathcal{H}) \longrightarrow \mathcal{L}(\mathcal{H}, \mathcal{H})$$

So therefore $\mathcal{L}(\mathcal{H}, \mathcal{H})$ has the following properties. Let $A : \mathcal{H} \longrightarrow \mathcal{H}$ and $B : \mathcal{H} \longrightarrow \mathcal{H}$ be elements of $\mathcal{L}(\mathcal{H}, \mathcal{H})$.

1. As we know, $\mathcal{L}(\mathcal{H}, \mathcal{H})$ is a vector space since $A + B, \lambda A \in \mathcal{L}(\mathcal{H}, \mathcal{H})$ for any $\lambda \in \mathbb{C}$.
2. We can define multiplication on $\mathcal{L}(\mathcal{H}, \mathcal{H})$ to be simply the composition of operators, so $AB \in \mathcal{L}(\mathcal{H}, \mathcal{H})$.
3. We can define the adjoint operator on $\mathcal{L}(\mathcal{H}, \mathcal{H})$ from above, which satisfies $A^{**} = A$ and $(AB)^* = B^*A^*$.

The first two conditions simply make $\mathcal{L}(\mathcal{H}, \mathcal{H})$ an algebra (in the algebraic sense), but the third condition makes it a C^* -algebra.

4 Duality

Now given the L^2 space (of complex-valued functions over a measure space X), we can upgrade it from Banach to Hilbert by defining the inner product

$$\langle f, g \rangle := \int_X \overline{f(x)}g(x) d\mu$$

Now, given p, q satisfying $\frac{1}{p} + \frac{1}{q} = 1$, let $f \in L^p$ and $g \in L^q$. Then, f defines a linear map

$$T_f : g \mapsto \int_X fg d\mu$$

Sometimes, this map T_f may be written $\int_X f$, but this does not denote an integral; it is purely notation. $\int_X f$ does not even have to be finite since f is not L^1 . They are only integrals when you feed them the appropriate terms.

Some definitions.

1. $c_0 = \{\{a_j\} \mid a_j \rightarrow 0\}$ is the set of convergent sequences. It can be seen that $c_0 \subset \ell_\infty$ since every convergent sequence is bounded.
2. $\ell_1 = \{\{a_j\} \mid \sum_j |a_j| < \infty\}$ is the set of sequences such that the sum of all (the absolute value of) its terms is finite. But upon looking at this, we can see that ℓ_1 is precisely the L^1 space of functions over the measure space $(\mathbb{N}, 2^{\mathbb{N}}, \mu)$, where μ is the counting measure (where the measure of a subset is determined by the number of elements in that set). Note that $\ell_1 \subset c_0$.
3. $\ell_2 = \{\{a_j\} \mid \sum a_j^2 < \infty\}$. That is, this is the $L^2(\mathbb{N})$ space under the counting measure.
4. ℓ_p can be defined as the same space under the counting measure, but equipped with the p -norm.
5. $\ell_\infty = \{\{a_j\} \mid |a_j| \leq c \text{ for all } j\}$ for some constant c . Clearly, this is the L^∞ -norm, and it must be for all j since the counting measure on any nonempty subset of \mathbb{N} is at least 1.

In summary,

$$\ell_1 \subset \ell_2 \subset \ell_3 \subset \dots \subset c_0 \subset \ell_\infty$$

To prove this, let's assume that some sequence $f \in \ell_p$, and let $1 \leq p < q \leq \infty$. Then, $f \in c_0$ and therefore is convergent to 0. There must be some $K \in \mathbb{N}$ such that $f(k) < 1$ for $k > K$. So, for all $k > K$, we have $|f(k)|^q < |f(k)|^p < 1$, and so

$$\sum_{k>K} |f(k)|^q < \sum_{k>K} |f(k)|^p$$

The first k terms would not affect finiteness of the norm so

$$\|f\|_p = \sum_{k \in \mathbb{N}} |f(k)|^p < \infty \implies \|f\|_q = \sum_{k \in \mathbb{N}} |f(k)|^q < \infty$$

and so $\ell_p \subset \ell_q$. \square

Let's talk about duality. Unlike that of a finite-dimensional vector space, the dual of an infinite dimensional space is not isomorphic to itself. This dual space is just some abstract space of linear functionals, but we can informally think of it as sort of an "infinite dot product." That is, given a sequence $\{f(n)\}_{n \in \mathbb{N}}$, an element ϕ in its dual space would act on f as

$$\phi(f) := \sum_{n \in \mathbb{N}} f(n) \phi(n) \in \mathbb{C}$$

For example, the dual of c_0 is not itself because $f = (\frac{1}{\sqrt{n}})_n \in c_0$, but the functional $\phi = (\frac{1}{\sqrt{n}})$ acting on f would be infinite.

$$\phi(f) = \sum_{n \in \mathbb{N}} \frac{1}{\sqrt{n}} \frac{1}{\sqrt{n}} = \sum_{n \in \mathbb{N}} \frac{1}{n} = \infty \notin \mathbb{C}$$

So what is the dual space of c_0 ? It turns out that $c_0^* \simeq \ell_1$. In order to prove this, we must prove that the map $\phi : \ell_1 \rightarrow c_0^*$ defined

$$\phi(g)(f) := \sum_{n \in \mathbb{N}} g(n) f(n)$$

is an isomorphism. Linearity is easy since

$$\begin{aligned} \phi(\lambda g_1 + g_2)(f) &= \sum_{n \in \mathbb{N}} (\lambda g_1 + g_2)(n) f(n) \\ &= \sum_{n \in \mathbb{N}} (\lambda g_1(n) + g_2(n)) f(n) \\ &= \sum_{n \in \mathbb{N}} \lambda g_1(n) f(n) + \sum_{n \in \mathbb{N}} g_2(n) f(n) \\ &= \lambda \sum_{n \in \mathbb{N}} g_1(n) f(n) + \sum_{n \in \mathbb{N}} g_2(n) f(n) \\ &= \lambda \phi(g_1)(f) + \phi(g_2)(f) \end{aligned}$$

Now, we want to prove boundedness (i.e. $\phi(g)$ is bounded for all g so doesn't explode to ∞). By Holder's Inequality, we have $f \in c_0 \subset \ell_\infty$ and $g \in \ell_1$, so

$$\sum_{n \in \mathbb{N}} |f(n) g(n)| \leq \|f\|_\infty \|g\|_1 < \infty$$

since $\phi(g)(f)$ converges absolutely, it also converges. To prove injectivity, we can just prove that the nullspace of ϕ is trivial. Assume that a nontrivial $l \in \ell_1$ is in the nullspace. Then, there exists a k th nonzero term: $l(k) \neq 0$. But we have claimed that $\phi(l)(f) = 0$ for all $f \in c_0$, but setting $f(k) = 1$, this makes a contradiction. To prove surjectivity, let $\varphi \in c_0^*$ be a linear functional. Then, we can construct a sequence g element-wise by assigning

$$g_n = \varphi(e_n)$$

where e_n is the n th basis element of c_0 . Then, $g = (g_n)_{n \in \mathbb{N}}$ is a sequence, and we want to show that it is in ℓ_1 .

Furthermore, given that p, q are duals, i.e. $\frac{1}{p} + \frac{1}{q} = 1$, we have

1. $c_0^* = \ell_1$
2. $\ell_1^* = \ell_\infty$
3. $\ell_2^* = \ell_2$. Since the dual space of ℓ_2 is itself, using Riesz Representation theorem we can construct an inner product on ℓ_2 to make it a Hilbert space. ℓ_2 is the only space that can be constructed into a Hilbert space. The rest are Banach spaces.

5 Measurability

Know what this means. Borel Cantelli

$$\bigcup_{n=1}^{\infty} \bigcap_{m>n} A_m$$

and know what

$$\limsup A_n$$

means. And know what infinitely often means. As time goes on, what is an event that happens infinitely often.

6 Fourier Transform

The *Fourier transform* \mathcal{F} is an operator (an isometric isomorphism) mapping one function to another. There are three spaces where it is naturally defined, the *Schwartz class* \mathcal{S} , the L^2 space, and the Banach dual of \mathcal{S} .

$$\begin{aligned}\mathcal{F} : \mathcal{S} &\longrightarrow \mathcal{S} \\ \mathcal{F} : L^2 &\longrightarrow L^2 \\ \mathcal{F} : \mathcal{S}' &\longrightarrow \mathcal{S}'\end{aligned}$$

and $\mathcal{S} \subset L^p \subset \mathcal{S}'$. \mathcal{S} is small, \mathcal{S}' is big, and L^p spaces live in between them. The *Schwartz class* is the function space of all functions $f : \mathbb{R}^n \rightarrow \mathbb{C}$ satisfying

1. $f \in C^\infty(\mathbb{R}^n, \mathbb{C})$
2. f decays rapidly in the sense that for any scalar α , f multiplied by any monomial x^α is bounded everywhere:

$$\sup_{x \in \mathbb{R}^n} |x^\alpha f(x)| < \infty, \text{ or a similar condition } \|x^\alpha f\|_\infty < \infty$$

Furthermore, with vector $\beta = (\beta_1, \dots, \beta_n) \in \mathbb{N}^n$ and multi-indexing $D^\beta = D_1^{\beta_1} D_2^{\beta_2} \dots D_n^{\beta_n} f$, all of its derivatives are also bounded for all β

$$\sup_{x \in \mathbb{R}^n} |x^\alpha D^\beta f(x)| < \infty, \text{ or } \|x^\alpha D^\beta f\|_\infty < \infty$$

Now Schwartz functions have both the properties of smoothness and decay. The purpose of the Fourier transform is to convert between the properties of smoothness and decay. That is, if f decays (is smooth) to some extent, then $\mathcal{F}f$ is smooth (decays) to that extent. $\mathcal{F} : \mathcal{S} \rightarrow \mathcal{S}$ is defined (note that x and ξ are just dummy variables)

$$f(x) \mapsto \mathcal{F}f(\xi) = \hat{f}(\xi) = \int_{\mathbb{R}^n} f(x) e^{-ix\xi} dx$$

This transform is a simple map from a larger class called the *Fourier integral operators*, which are maps of the form

$$f(x) \mapsto \mathcal{F}f(\xi) = \int_{\mathbb{R}^n} f(x) e^{i\varphi(x,\xi)} a(\xi, x) dx$$

Now since \mathcal{F} is an isometric automorphism of \mathcal{S} , we can take its inverse transform, defined

$$\mathcal{F}^{-1}g(x) = \int_{\mathbb{R}^n} g(\xi) e^{ix\xi} d\xi$$

The Fourier transform allows us to connect multiplication and differentiation. Let us calculate the α -fold derivative along ξ of the Fourier transform of f .

$$\begin{aligned}\partial_\xi^\alpha \hat{f}(\xi) &= \partial_\xi^\alpha \int_{\mathbb{R}^n} e^{-ix\xi} f(x) dx \\ &= \int_{\mathbb{R}^n} \partial_\xi^\alpha e^{-ix\xi} f(x) dx \\ &= \int_{\mathbb{R}^n} (-ix)^\alpha e^{-ix\xi} f(x) dx \\ &= \mathcal{F}((-ix)^\alpha f(x))\end{aligned}$$

So, a α -fold differentiation after Fourier transforming a function is the same as multiplying x^α first and then Fourier transforming the same function. Now, if we multiply a the Fourier transform of a function by a monomial this sort of looks like the Fourier transform of the derivative of the function.

$$\begin{aligned}\xi^\alpha \hat{f}(\xi) &= \int_{\mathbb{R}^n} \xi^\alpha e^{-ix\xi} f(x) dx \\ &= \int_{\mathbb{R}^n} \partial_x^\alpha (e^{-ix\xi}) f(x) dx \\ &= \int_{\mathbb{R}^n} (e^{-ix\xi}) \partial_x^\alpha f(x) dx \quad (\text{Integration by Parts})\end{aligned}$$

7 Inequalities

The Cauchy Schwartz inequality

$$ab \leq \frac{a^2}{2} + \frac{b^2}{2}$$

is a special case of Young's inequality

$$ab \leq \frac{a^p}{p} + \frac{b^q}{q}$$

where p and q are Holder conjugates ($\frac{1}{p} + \frac{1}{q} = 1$). Young's inequality is actually a specific case of a more general inequality. Let $f : \mathbb{R} \rightarrow \mathbb{R}$ be strictly monotonic satisfying $f(0) = 0$. Then, there exists an inverse $f^{-1} = g : \mathbb{R} \rightarrow \mathbb{R}$. For any point α, β within the respective domain and range,

$$ab \leq \int_0^a f(x) dx + \int_0^b g(x) dx$$

and Young's inequality sets $f(x) = x^{p-1}$ and $g(x) = x^{q-1} = x^{\frac{p}{p-1}-1} = x^{\frac{1}{p-1}} = f^{-1}(x)$. Here $1 < p < \infty$.

Minkowski Inequality says that if you have an $L^p(\Omega; \mathbb{R})$ space with norm $\|\cdot\|_{L^p} := (\int |f|^p)^{1/p}$, then the norm satisfies the triangle inequality. Using induction gives

$$\left\| \sum_j f_j(x) \right\|_{L^p(dx)} \leq \sum_j \|f_j(x)\|_{L^p(dx)}$$

That is, the norm of the sum of the functions is less than the sum of the function norms. But if we index $f_j(x)$ as $f(x, y)$, then since integration is really the same thing as summation, we can write the above as

$$\left\| \int_{\mathbb{N}} f(x, y) dy \right\|_{L^p(dx)} \leq \int_{\mathbb{N}} \|f(x, y)\|_{L^p(dx)} dy$$

expanding

$$\left(\int_{\Omega} \left| \int_{\mathbb{N}} f(x, y) dy \right|^p dx \right)^{\frac{1}{p}} \leq \int_{\mathbb{N}} \left(\int_{\Omega} |f(x, y)|^p dx \right)^{\frac{1}{p}} dy$$

So summing $\int dy$, then power to p , then norming $\int dx$, then power to $\frac{1}{p}$ is less than powering to p first, then norming $\int dx$, then power to $\frac{1}{p}$, and finally summing $\int dy$. Minkowski inequality is especially useful for functions of the form

$$f(x, y) = \phi_{\epsilon}(x - y) u(y)$$

where ϕ_{ϵ} is some sort of bump function ($\phi_{\epsilon}(x) \rightarrow \delta(x)$ as $\epsilon \rightarrow 0$). Then, defining the convolution operator

$$(J_{\epsilon}u)(x) = \int \phi_{\epsilon}(x - y)u(y) dy$$

Now we can approximate the delta function $\delta(x - y)$ with C^{∞} bump functions by "squeezing them" as $\phi_{\epsilon}(t) = \phi(t/\epsilon)\epsilon^{-n}$ (we power to n to normalize over $t \in \mathbb{R}^n$, an n -dimensional input). Therefore, we can approximate u with arbitrary accuracy with a C^{∞} function.

$$(J_{\epsilon}u)(x) = \underbrace{\int \phi_{\epsilon}(x - y)u(y) dy}_{\text{smooth}} \approx \int \delta(x - y)u(y) dy = u(x)$$

Assuming that some smooth and compactly supported $\varphi \in C_c^{\infty}(\mathbb{R}^n)$ exists (we can construct this), we can construct a family of approximate identities/deltas (since deltas are identities for convolutions), called *mollifiers* (which "mollify" or smooth out an arbitrary function)

$$\{\varphi_{\epsilon}\}_{\epsilon} \text{ s.t. } \varphi_{\epsilon}(t) = \epsilon^{-n} \varphi(t/\epsilon)$$

We have

$$(J_{\epsilon}u)(x) := (\varphi_{\epsilon} * u)(x) = \int \varphi_{\epsilon}(x - y)u(y) dy \rightarrow u(x) \text{ in } L^p$$

So, $\|(J_{\epsilon}u)(x) - u(x)\|_{L^p} \rightarrow 0$ using Minkowski inequality. This property is very useful because if we're getting an arbitrary signal, we can interpret it as a function, assume the white noise is Gaussian, and then the convolution would be a smooth signal. So in reality, we don't have to worry about smoothness.

Theorem 7.1. *If there exists some function u such that*

$$\int u\varphi = 0 \text{ for all } \varphi \in C_C^\infty$$

Then $u = 0$ almost everywhere.

8 Sobolev Spaces

8.1 Weak Derivative