Applied Linear Analysis

Course Notes by Xiaobing Feng
Transcribed and Annotated by Gregory $\mathbb{Z}$itelli

Measure Theory

Definition. Given a set X and a collection of subsets $\mathcal{G}$, we say that $\mathcal{G}$ is a ring if the first three conditions are satisfied, and a σ-ring if all of the following conditions are satisfied:

(i) $\varnothing \in \mathcal{G}$
(ii) $A, B \in \mathcal{G}$ implies that $A \backslash B \in \mathcal{G}$
(iii) $A, B \in \mathcal{G}$ implies that $A \cup B \in \mathcal{G}$
(iv) $A_n \in \mathcal{G}$ implies that $\bigcup_{n=1}^{\infty} A_n \in \mathcal{G}$

If in addition $X \in \mathcal{G}$, then we say that $\mathcal{G}$ is a σ-algebra.

Definition. A set function $\mu : \mathcal{G} \to \overline{\mathbb{R}}$ (or to $\underline{\mathbb{R}}$) is called a measure if

(i) $E \in \mathcal{G}$ implies that $\mu(E) \geq 0$
(ii) $\mu(\varnothing) = 0$
(iii) μ is σ-additive. For any $A_n \in \mathcal{G}$, $A_i \cap A_j = \varnothing$ for $i \neq j$, then $\mu\left(\bigcup_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} \mu(A_n)$

If a set function μ satisfies (ii) and (iii), we say it is a signed measure. If it satisfies only (i) and (ii), and property (iii) is relaxed to an inequality $\mu\left(\bigcup_{n=1}^{\infty} A_n\right) \leq \sum_{n=1}^{\infty} \mu(A_n)$, then we say it is an outer measure.

Definition. Let μ be a (signed) measure, defined on a σ-algebra $\mathcal{G}$. We say that μ is complete if for any set $E \in \mathcal{G}$ such that $\mu(E) = 0$, any subset $F \subseteq E$ is in $\mathcal{G}$.

Definition. Let X be a set, $\mathcal{G} \subseteq 2^X$ a σ-algebra on X, and μ a measure defined on $\mathcal{G}$. Then the triplet $(X, \mathcal{G}, \mu)$ is called a measure space. If $\mu(X) < \infty$ then we say that $(X, \mathcal{G}, \mu)$ is finite. If $\mu(X) = 1$ then we say that $(X, \mathcal{G}, \mu)$ is a probability space. If μ is complete then we say that $(X, \mathcal{G}, \mu)$ is complete.

Definition. A function $f : X \to Y$ between a measure space X and a topological space Y is said to be measurable if the inverse image of open sets in Y is measurable in X. If f is real valued ($\overline{\mathbb{R}}$ or $\underline{\mathbb{R}}$), this is equivalent to saying that $f^{-1}((-\infty, a)) \in \mathcal{G}$ for all $a \in \mathbb{R}$ (and $\{-\infty\} \in \mathcal{G}$ if f is $\underline{\mathbb{R}}$ valued).

Theorem. *Measurable functions are closed under the standard operations of* $+$, $-, \times, \div, \limsup, \liminf$.

Integration

Our motivation for Lebesgue integration is to partition the range of a function, rather than its domain. In a typical Riemann integral over a set $[a, b] \subseteq \mathbb{R}$, we create a partition of the domain $P = \{a = x_0 < x_1 < \ldots < x_n = b\}$, and then consider the minimum and maximum values of the function over the smaller sets $[x_i, x_{i+1}]$. Lebesgue integration approaches the situation from the opposite direction, beginning with simple functions $s(x) = \sum_{i=1}^{n} a_i \chi_{A_i}(x)$. The sets A_i are essentially the inverse image of the function in question over a small interval in the range, meaning they can end up being quite wild. The integral is defined for these simple functions, and then the integral of an actual function is the limit as it is approximated by simple function.

Definition. We define the following types of convergence for a function $f : X \to \mathbb{R}$:

(i) Pointwise convergence: $\lim_{n\to\infty} f_n(x) = f(x)$ for all $x \in X$.

(ii) Almost everywhere pointwise convergence: $\lim_{n\to\infty} f_n(x) = f(x)$ for almost every $x \in X$.

(iii) Convergence in measure: For any $\epsilon > 0$,

$$\lim_{n\to\infty} \mu\left(\{x \in X : |f_n(x) - f(x)| > \epsilon\}\right) = 0$$

(iv) Uniform convergence: For any $\epsilon > 0$, there exists some $N \in \mathbb{N}$ such that $n > N$ implies that $|f_n(x) - f(x)| < \epsilon$ for all $x \in X$.

(v) Almost uniform convergence: For any $\epsilon > 0$, there exists some $N \in \mathbb{N}$ such that $n > N$ implies that $|f_n(x) - f(x)| < \epsilon$ for almost every $x \in X$.

Theorem (Lebesgue Dominated Convergence Theorem)**.** *Let $f_n \to f$ pointwise almost everywhere or in measure, with f_n integrable for all n. Suppose that there exists some g integrable such that $|f_n| \leq g$ almost everywhere, then f is integrable and*

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

$$\lim_{n\to\infty} \int f_n d\mu = \int f d\mu$$

Theorem (Fatou's Lemma)**.** *Suppose that $f_n \geq 0$ almost everywhere is integrable, then*

$$\int \lim_{n\to\infty} f_n d\mu \leq \lim_{n\to\infty} f_n d\mu$$

Radon-Nikodyn Theorem

Definition. A function $g : (a, b) \to \mathbb{R}$ is absolutely continuous if for any $\epsilon > 0$ there exists some $\delta > 0$ such that for any sequence of intervals $\{(a_j, b_j\} \subseteq (a, b)$ with $\sum_{j=1}^{\infty}(b_j - a_j) < \delta$,

$$\sum_{j=1}^{\infty} |g(b_j) - g(a_j)| < \epsilon$$

It turns out that $g : (a, b) \to \mathbb{R}$ is absolutely continuous if and only if there exists a Lebesgue integrable $f : (a, b) \to \mathbb{R}$ such that $g(x) = g(a) + \int_a^x f(t)dt$ (where dt here represents Lebesgue measure on $\mathbb{R}$). We call this f the derivative of g, and in fact g is differentiable almost everywhere and $f = g'$ at such points.

Definition. A function $g : [a, b] \to \mathbb{R}$ is of bounded variation if there exists some $K > 0$ such that for any finite partition of $[a, b]$, $\{a = x_0 < x_1 < \ldots < x_n = b\}$,

$$\sum_{j=1}^{\infty} |g(x_j) - g(x_{j-1})| \leq K$$

In general, such a K will not be unique. We define $V(g)$ to be the infimum of such K, called the total variation of g on $[a, b]$. Functions of bounded variation are incredibly important, because they can always be expressed as the difference of two monotonically increasing functions. Since monotonically increasing functions are differentiable almost everywhere (in contrast to typical functions, which are often differentiable nowhere), it follows that functions of bounded variation are also differentiable almost everywhere. It can be shown that absolutely continuous functions are also functions of bounded variation, although the converse is not true.

The previous definition of absolute continuity applies specifically to functions from open intervals (a, b) into $\mathbb{R}$. Absolute continuity leads to a notion of the derivative, which we would like to generalize. One way to do this is to create a version of absolute continuity for measures. Our initial definition of absolute continuity for one measure with respect to another does not appear to resemble the one for functions, but we will see that a version can be constructed using ϵ's and δ's.

Definition. Let μ, ν be two signed measures on the same σ-algebra $\mathcal{G}$. The measure ν is said to be absolutely continuous with respect to μ if $\nu(E) = 0$ whenever $|\mu|(E) = 0$ for $E \in \mathcal{G}$, written $\nu << \mu$.

Theorem. *The following are equivalent:*

(i) $\nu << \mu$
(ii) $\nu^+ << \mu$
(iii) $|\nu| << \mu$

Theorem. *If μ, ν are two signed measure on the same σ-algebra $\mathcal{G}$, and $|\nu|(E) < \infty$ whenever $|\mu|(E) < \infty$ for $E \in \mathcal{G}$, then the following are equivalent:*

(i) $\nu << \mu$
(ii) For any $\epsilon > 0$ there exists $\delta > 0$ such that for any measurable $E \in \mathcal{G}$ such that $|\mu|(E) < \delta$, it follows that $|\lambda(E)| < \epsilon$.

Theorem (Radon-Nikodyn Theorem)**.** *Let $(X, \mathcal{G}, \mu)$ be a finite measure space. Suppose that ν is a finite signed measure on $\mathcal{G}$. Then $\nu << \mu$ if and only if there exists some μ-integrable function f such that $\nu(E) = \int_X f d\mu$ for any $E \in \mathcal{G}$. Furthermore, such an f is unique almost everywhere with respect to μ.*

We write $\frac{d\nu}{d\mu} = f$. The Radon-Nikodyn Theorem is something of a generalization of the fundamental theorem of calculus for set functions. Just like a real valued function has a derivative almost everywhere if it is absolutely continuous, a measure also has a derivative with respect to another measure if it is absolutely continuous with respect to that measure.

Definition. Let μ, ν be two signed measures. We say that they are mutually singular if there exist measurable sets A, B such that $A \cup B = X$, $A \cap B = \varnothing$, $|\mu|(A) = 0$ and $|\nu|(B) = 0$. We write $\mu \perp \nu$ or $\nu \perp \mu$.

Theorem (Lebesgue Decomposition)**.** *Let μ, ν be σ-finite signed measures, then there exist σ-finite signed measures ν_0, ν_1 such that $\nu = \nu_0 + \nu_1$, $\nu_0 \perp \mu$ and $\nu_1 << \mu$.*

Product Measures

Given a finite sequence of measure spaces $(X_i, \mathcal{G}_i, \mu_i)$, we wish to turn the set $X_1 \times \ldots \times X_n$ into a measure space. We typically think of X_i as being $\mathbb{R}$ equipped with Lebesgue measure, and we would like to define a measure on $\mathbb{R}^n$ which agrees with our notion of area and volume. Specifically, we would like the measures of rectangles to be the product of their side lengths.

We accomplish this by defining measurable n-rectangles, which will be sets of the form $A_1 \times A_2 \times \ldots A_n$ where each $A_i \in \mathcal{G}_i$. Letting $\mathcal{G}$ be the σ-algebra generated by these measurable n-rectangles, we then define the following outer measure on $\mathcal{G}$:

$$\mu^*(E) = \inf \left\{ \sum_{j=1}^{\infty} \mu_1 \left(A_1^{(j)} \right) \mu_2 \left(A_2^{(j)} \right) \cdots \mu_n \left(A_n^{(j)} \right) : E \subseteq \bigcup_{j=1}^{\infty} A_1^{(j)} \times \ldots \times A_n^{(j)} \right\}$$

We then use this outer measure to construct a measure. If each X_i is indeed $\mathbb{R}$ equipped with Lebesgue measure, then we say that μ is n-dimensional Lebesgue measure.

Theorem (Fubini's Theorem)**.** *If $f : X \to \overline{\mathbb{R}}$ (or $\underline{\mathbb{R}}$) is integrable on a product measure space X, then*

$$\int f d\mu = \int \cdots \int f d\mu_{j_1} d\mu_{j_2} \ldots d\mu_{j_n}$$

for any permutation j_i of $1 \ldots n$.

Fixed Point Theorem

Definition. Given a metric space X, with $X_0 \subseteq X$, a function (or a mapping) $T : X_0 \to X$ is called a contraction if there exists some $\theta \in (0,1)$ such that $\rho(Tx, Ty) \leq \theta\rho(x, y)$ for any $x, y \in X_0$.

Theorem (Fixed Point Theorem)**.** *Let (X, ρ) be a complete metric space, $T : X \to X$ a contraction. Then there exists a unique point $z \in X$ such that $T(z) = z$.*

Proof. We begin by proving that such a point exists. Take some $x_0 \in X$, and define the recursive sequence $x_{n+1} = Tx_n$ for $n \in \mathbb{N}$. We claim that x_n is a Cauchy sequence. Observe that

$$\rho(x_{n+1}, x_n) = \rho(Tx_n, Tx_{n-1}) \leq \theta\rho(x_n, x_{n-1})$$

Since this holds for any $n \geq 1$, it follows that $\rho(x_{n+1}, x_n) \leq \theta^n\rho(x_1, x_0)$

$$\begin{aligned}
\rho(x_{n+p}, x_n) &\leq \rho(x_{n+p}, x_{n+p-1}) + \rho(x_{n+p-1}, x_{n+p-2}) + \ldots + \rho(x_{n+1}, x_n) \\
&\leq \left(\theta^{n+p-1} + \theta^{n+p-2} + \ldots + \theta^n\right)\rho(x_1, x_0) \\
&\leq \frac{\theta^n}{1-\theta}\rho(x_1, x_0)
\end{aligned}$$

As $n \to \infty$, this expression approaches 0 for any fixed p. Therefore, letting $m = n + p$, the sequence is Cauchy. Since X is complete, there exists some $z \in X$ such that $\rho(x_n, z) \to 0$ as $n \to \infty$.

We now want to show that z is a fixed point of T. We have that

$$\begin{aligned}
\rho(z, Tz) &\leq \rho(z, x_n) + \rho(x_n, Tz) \\
&= \rho(z, x_n) + \rho(Tx_{n-1}, Tz) \\
&\leq \rho(z, x_n) + \theta\rho(x_{n-1}, z)
\end{aligned}$$

Since the right hand side approaches 0 as $n \to \infty$, we have that $\rho(z, Tz) = 0$, so $Tz = z$.

We now need to show that z is unique. Suppose that $z_1, z_2 \in X$ exist such that $z_1 = Tz_1$ and $z_2 = Tz_2$. Then

$$\rho(z_1, z_2) = \rho(Tz_1, Tz_2) \leq \theta\rho(z_1, z_2)$$

But this implies that $(1-\theta)\rho(z_1, z_2) \leq 0$, so $\rho(z_1, z_2) = 0$ and $z_1 = z_2$. ∎

Corollary. *Suppose $Y \subseteq X$ is a closed subset of a complete metric space X. If $T : Y \to Y$ is a contraction on Y, then there exists a unique $z \in Y$ such that $Tz = z$.*

Remark.

(i) If $X = \mathbb{R}^n$ and $Y = \overline{B_1(0)}$, then the contraction requirement can be relaxed so that $\rho(Tx, Ty) \leq \rho(x, y)$. However, the fixed point may not be unique.
(ii) If the contraction requirement is removed, the existence still holds if T is a compact mapping/operator.

Hausdorff Measures & Dimensions

Consider the sets $Y = \{y = ax + b\}$ and $Z = \{(x - x_0) \cdot n = 0\}$ in $\mathbb{R}^3$, a line and a plane. Under Lebesgue measure in $\mathbb{R}^3$, they both have measure 0. Yet the sets are considerably different, with the first having co-dimension 2 and the second having co-dimension 1. Hausdorff measure is a construction of geometric measure theory that allows us to make an appropriate distinction between such sets, and generalizes the notion of dimension.

We begin with a metric space (X, ρ), and a subset $E \subseteq X$. For $\delta > 0$, $d \geq 0$, we define the set function $H^d_\delta(E)$ as follows:

$$H^d_\delta(E) = \inf \left\{ \sum_{i=1}^{\infty} (\mathrm{diam}(A_i))^d : E \subseteq \bigcup_{i=1}^{\infty} A_i, \mathrm{diam}(A_i) < \delta \right\}$$

Recall that $\mathrm{diam}(B) = \sup\{\rho(x, y) : x, y \in B\}$. Unlike Lebesgue outer measure which relies on open sets, our sets A_i can be anything as long as their diameters are small enough. It is a fact that H^d_δ is increasing as δ decreases. Then it follows that $\limsup_{\delta\to 0} H^d_\delta(E) = \lim_{\delta\to 0} H^d_\delta(E)$ exists, although it may be ∞.

We define $H^d(E) = \sup_{\delta>0} H^d_\delta(E) = \lim_{\delta\to 0} H^d_\delta(E)$, called the Hausdorff measure. It is also a fact that H^d is an outer measure on X. If we let $\mathcal{S}(X)$ be the set of all H^d-measurable subsets of X, then H^d is a metric measure on $(X, \mathcal{S}(X))$. Furthermore, $\mathcal{S}(X)$ contains all the Borel sets of X, under the metric topology.

Definition. We define the Hausdorff dimension of a set E as

$$\dim_H(E) = \inf \left\{ d \geq 0 : H^d(E) = 0 \right\} = \sup \left\{ d \geq 0 : H^d(E) = \infty \right\}$$

The Hausdorff dimension of a set does not need to be an integer. For a line and a plane, the dimensions agrees with our intuition of 1 and 2. There are many fractal sets whose Hausdorff dimension are fractions, or even transcendental numbers.

Metric Spaces of Radon Measures

Let $(X, \mathcal{F})$ be a topological space, $\mathcal{G}(\mathcal{F})$ the σ-algebra generated by $\mathcal{F}$. Each set $\mathcal{G}(\mathcal{F})$ is called a Borel set. All open and closed sets are Borel sets, as well as countable unions and intersections of such sets.

Definition. Any measure μ on $(X, \mathcal{G}(\mathcal{F}))$ is called a Borel measure. $(X, \mathcal{G}(\mathcal{F}), \mu)$ is called a Borel measure space.

Definition. Assume $(X, \mathcal{F})$ is a topological space, μ a measure on X.

(i) μ is said to be inner regular or tight if for any Borel set E,

$$\mu(E) = \sup \{\mu(F) : F \subseteq E, F \text{ is compact}\}$$

(ii) μ is said to be outer regular if for any Borel set E,

$$\mu(E) = \inf \{\mu(B) : E \subseteq B : B \text{ is open}\}$$

(iii) μ is said to be locally finite if for any $x \in X$, there exists a neighborhood N_x of x such that $\mu(N_x) < \infty$.

(iv) μ is called a Radon measure if μ is inner regular and locally finite.

(v) We define the support of a measure as

$$\text{supp}(\mu) = \{x \in X : \forall N_x \in \mathcal{F}, \mu(N_x) > 0\}$$

We commonly write $\mathcal{M}(X)$ for the set of all signed Radon measures on X, and $M_+(X)$ for the set of all nonnegative Radon measures on X. We additionally write $\mathcal{P}(X) \subseteq \mathcal{M}_+(X)$ for the space of probability Radon measures, such that the measure of the whole space is 1. We would like to define some kind of metric on these spaces of measures.

Example. The Radon metric on $\mathcal{M}_+(X)$ is defined as

$$\rho(\mu_1, \mu_2) = \sup \left\{ \int_X f d(\mu_1 - \mu_2) : f : X \to [-1, 1], f \text{ is continuous} \right\}$$

Is the Radon metric any good? It is a fact that $\rho(\mu_n, \mu) \xrightarrow{n \to \infty} 0$ implies that μ_n converges to μ weakly, although the converse may not be true in general. Additionally, $(\mathcal{P}(X), \rho)$ is not a sequentially compact metric space. Are question then becomes, are there any attractive metrics on $\mathcal{P}(X)$? It turns out there is, called the Wasserstein metric.

Definition. Let (X, d) be a metric space, we say that $m_p(\mu) = \int_X (d(x, x_0))^p \, d\mu(x)$ is the p^{th} moment of $\mu \in \mathcal{P}(X)$. We write the space of measures with finite p^{th} moments as

$$\mathcal{P}_p(X) = \{\mu \in \mathcal{P}(X) : m_p(\mu) < \infty\}$$

Definition. We say that a measure γ on $X^2 = X \times X$ is a marginal of μ (on the first factor) and ν (on the second factor) if $\gamma(E_x) = \nu(E_x)$ and $\gamma(E_y) = \mu(E_y)$. We write $\Gamma(\mu, \nu)$ for the space of all such marginals.

Definition. The p^{th} Wasserstein metric (for $p \geq 1$) is defined on $\mathcal{P}(X)$ by

$$W_p(\mu, \nu) = \inf_{\gamma \in \Gamma(\mu,\nu)} \left(\int_{X \times X} (d(x, y)) \, d\gamma(x, y) \right)^{1/p}$$

It can be proven that $(\mathcal{P}(X), W_p)$ is a seperable and complete if X is seperable and complete, and is compact if X is compact. The metrics W_p have important application in optimal transportation problems.

Normed Linear Spaces

Our goals for this section will be to review the general structure of linear spaces, and begin to study linear operators on them. We will typically be working with linear spaces of infinite dimension, such that for any linearly independent collection there exists another element in the space that can be added to the collection without making it dependent.

Definition. A function $\varphi : X \to [0, \infty)$ is called a norm on a linear space X if it satisfies the following properties:

(i) $\varphi(x) \geq 0$ for all $x \in X$.
(ii) $\varphi(x) = 0$ if and only if $x = 0$.
(iii) $\varphi(\lambda x) = |\lambda|\varphi(x)$ for all $x \in X$, λ is any scalar.
(iv) $\varphi(x + y) \leq \varphi(x) + \varphi(y)$ for all $x, y \in X$.

A linear space equipped with a norm is called a normed linear space.

We usually write $||x|| = \varphi(x)$. The scalars used to define our linear space must come from some field. We usually consider this field to be either $\mathbb{R}$ or $\mathbb{C}$, since we need the notion of absolute values. For finite fields like $\mathbb{Z}/p\mathbb{Z}$, no such absolute value concept exists.

Norms generalize the notion of length to linear spaces, while metrics generalize distance. The two concepts are closely related, but not completely identical. We will see that every norm induces a metric, which in turn means that normed linear spaces have a natural topology given by the norm.

Definition. A metric linear space is a linear space X with a metric ρ such that the functions $f : X \times \mathbb{R} \to X$ and $g : X \times X \to X$ given by $f(x, \lambda) = \lambda x$ and $g(x, y) = x + y$ are continuous according to the measures $\rho, \rho + \rho, \rho + |\cdot|$.

Definition. A metric linear space (X, ρ) is called a Frechet space if

(i) $\rho(x, y) = \rho(x - y, 0)$ for any $x, y \in X$.
(ii) X is complete.

Theorem. *Every normed linear space is a metric linear space with metric ρ given by $\rho(x, y) = ||x - y||$ for all $x, y \in X$.*

The converse is generally not true. Not every metric linear space has a natural norm given by the metric. We are then led to ask, which kinds of metric linear spaces do have norms? It turns out that the condition is that the metric linear space is a Frechet space.

Definition. A linear space X is called a topological linear space if:

(i) There exists a topology $\mathcal{F}$ on X.
(ii) $(X, \mathcal{F})$ is Hausdorff.
(iii) The functions $f : X \times \mathbb{R} \to X$ and $g : X \times X \to X$ given by $f(x, \lambda) = \lambda x$ and $g(x, y) = x + y$ are continuous with respect to the standard topology on $\mathbb{R}$ and $\mathcal{F}$.

Topological linear spaces are simply metric linear spaces without a metric. The definition of topological linear spaces is identical to that for metric linear spaces, simply using the continuity endowed by the base topology rather than the metric.

Lemma. *Let (X, ρ) be a metric linear space, and define $||x|| = \rho(x, 0)$ for all $x \in X$. Then $||\cdot||$ is a norm if*

(i) $\rho(x, y) = \rho(x - y, 0)$
(ii) $\rho(\lambda x, \lambda y) = |\lambda| \rho(x, y)$

This lemma gives the precise conditions for a metric linear space to be a normed linear space. Additionally, the norm defined by $||x|| = \rho(x, 0)$ can be used to construct a metric $\hat{\rho}(x, y) = ||x - y||$ which will be identical to the original metric ρ. Therefore, all norms can be used to construct a metric, many metrics can be used to construct a norm, and all such linear spaces have a natural topology.

Banach Spaces

Definition. Given a normed vector space $(X, ||\cdot||)$, we say that X is a Banach space if it is complete, i.e. every Cauchy sequence has a convergent limit in X.

Definition. Two norms $||\cdot||_1$ and $||\cdot||_2$ on a linear space X are said to be equivalent if there exist constants $c_1, c_2 > 0$ such that for all $x \in X$,

$$c_1||x||_1 \leq ||x||_2 \leq c_2||x||_1$$

Theorem. *Norms $||\cdot||_1$ and $||\cdot||_2$ are equivalent if and only if either of the following:*

(i) Any bounded subset in $||\cdot||_1$ is also bounded in $||\cdot||_2$.
(ii) Any convergent sequence in $||\cdot||_1$ is also convergent in $||\cdot||_2$.

Example. Let $X = \mathbb{R}^n$, and $p \geq 1$. We can define the norms $||x||_p = \left(\sum_{i=1}^n |x_i|^p\right)^{1/p}$ and $||x||_\infty = \max_{1\leq i\leq n} |x_i|$. When $p = 2$, this is the standard Euclidean norm.

Example. Given a measure space (Ω, μ) and some $p \geq 1$, we define the p-norms on measurable functions from Ω to $\mathbb{R}$ by

$$||f||_p = \left(\int_\Omega |f|^p d\mu\right)^{1/p}$$

We consider the collection of functions whose p-norm is finite. We define $L^p(\Omega, \mu)$ as the space of equivalence classes of such function. Each equivalence class is the collection of functions that are equal almost everywhere, such that if $\tilde{f} = \{f\}$ and $g \in \tilde{f}$ then $\mu(\{x \in \Omega : f(x) \neq g(x)\}) = 0$. The p-norm on the equivalence class $\tilde{f}$ is defined by taking the norm of any representation.

We also define the ∞-norm as the essential supremum of bounded functions, which can be shown to be $||f||_\infty = \lim_{p\to\infty} ||f||_p$.

Lemma. *The spaces $(L^p(\Omega, \mu), ||\cdot||_p)$ for $1 \leq p \leq \infty$ are Banach spaces.*

Example. The continuous functions on a compact metric space X, written as $C(X)$, have a norm defined by $||f|| = \sup_{x\in\Omega} |f(x)|$.

Convex Hulls

Definition. Let X be a linear space. Given any $x, y \in X$, we define the segment determined by x and y to be the set

$$S(x,y) = \{z = \lambda x + (1-\lambda)y : \lambda \in [0,1]\}$$

We say that a subset $C \subseteq X$ is convex if $S(x,y) \subseteq C$ for any $x, y \in C$.

Definition. For any subset of a linear space $E \subseteq X$, the convex hull of E, denoted $\text{con}(E)$, is defined as the intersection of all convex sets in X containing E.

Theorem. *Let X be a linear space, $E \subseteq X$, then the convex hull of E is the collection of all linear combinations of elements in E with coefficients whose sum is 1,*

$$\text{con}(E) = \left\{ \sum_{i=1}^{n} \alpha_i x_i : \alpha_i \geq 0,\ \sum_{i=1}^{n} \alpha_i = 1,\ x_i \in E \right\}$$

Proof. Let C be the set defined above. We would like to show that $\text{con}(E) \subseteq C$ and $C \subseteq \text{con}(E)$. C is clearly convex, and since $E \subseteq C$ then $\text{con}(E) \subseteq C$ since $\text{con}(E)$ is the intersection of all convex sets containing E.

Now let V be some convex set such that $E \subseteq V$. If we can show that $C \subseteq V$ then it will follow that $C \subseteq \text{con}(E)$. We proceed by induction on the number n used in the definition of C, showing that each subset $C_n \subseteq C$ is also a subset of V. We let $C_1 = \{\alpha_1 x_1 : \alpha = 1, x_1 \in E\} = E \subseteq V$. Now suppose that $C_{n-1} \subseteq V$ for some $n \in \mathbb{N}$. Let $y = \alpha_1 x_1 + \alpha_2 x_2 + \ldots + \alpha_n x_n$ be any element of C_n, and let $\alpha = \alpha_1 + \ldots + \alpha_{n-1}$. Assuming that $\alpha_n \neq 0$, $z = \frac{1}{\alpha_n}(\alpha_1 x_1 + \ldots + \alpha_{n-1} x_{n-1}) \in C_n, V$. We also know that $x_n \in E \subseteq V$, and since V is convex $y = \alpha z + (1-\alpha)x_n \in V$. Therefore, $C_n \subseteq V$.

Since each $C_n \subseteq V$ and $C = \bigcup_{n=1}^{\infty} C_n$, it follows that $C \subseteq V$. Since this can be done for all convex sets V containing E, their intersection $\text{con}(E)$ also contains C. Combining these two inclusions, we have that $C = \text{con}(E)$. ■

Subspaces

Definition. Suppose $(X, ||\cdot||)$ is a normed linear space and $Y \subseteq X$ is a linear subspace of X, then $(Y, ||\cdot||)$ is called a normed linear subspace of $(X, ||\cdot||)$. If Y is closed in X with regard to the topology induced by the norm then we say that Y is a closed subspace of X.

Definition. Given a subspace $E \subseteq X$ of a linear space X, we define the span of E to be

$$\operatorname{span}(E) = \left\{ \sum_{i=1}^{n} \lambda_i x_i : n \in \mathbb{N}, x_i \in E, \lambda_i \in \mathbb{R} \right\}$$

The span of a set will always be a subspace. We had a similar definition for the convex hull of a set, but with the restriction that the linear combinations were convex (having coefficients λ_i nonnegative and summing to 1). Since the span is the set of all linear combinations, it will always be convex since it includes the convex linear combinations.

Theorem (Existence of a Basis). *Every linear space X contains a set $G \subseteq X$ of linearly independent elements such that* $\operatorname{span}(G) = X$.

This theorem appears trivially true for finite linear spaces since any linear space of dimension n is isomorphic to $\mathbb{R}^n$, and the canonical basis for $\mathbb{R}^n$ always exist. The fact that a basis exists for all infinite dimension spaces is much more complicated to prove, and relies on Zorn's Lemma, which additionally requires that the axiom of choice is assumed true.

Quotient Spaces

We recall that a binary relation on a set X is called an equivalence relation if it is reflexive, symmetric, and transition. These properties ensure that an equivalence relation behaves like the equality symbol. An equivalence relation is usually written by adjoining elements with the tilde symbol, where x equivalent to y would be written $x \sim y$. We define the equivalence class of an element to be $[x] = \{y \in X : x \sim y\}$.

Definition. Let X be a linear space and Y_0 be a linear subspace. We say that $x, y \in X$ are related by Y_0 if $x - y \in Y_0$.

The relation of elements in X being related by Y_0 is an equivalence relation. We define the quotient space X/Y_0 to be the collection of all equivalence classes of elements in X. If we define $[x] + [y] = [x + y]$ and $\lambda[x] = [\lambda x]$, then X/Y_0 is a linear space. If X is a normed linear space, then it would be convenient to define a norm on our new quotient space.

Lemma. *The function* $||[x]|| = \inf\{||y||_X : y \in [x]\}$ *is a norm on the linear space* X/Y_0 *provided that* Y_0 *is a closed subspace.*

Theorem. *If* X *is a Banach space and* Y_0 *is closed, then* X/Y_0 *is a Banach space.*

Remark. The space $H^k(\Omega)$ is called the Sovolev space. In applications we are frequently interested in the quotient space $H^k(\Omega)/P_n(\Omega)$, where $P_n(\Omega)$ is the space of n degree polynomials.

Finite-Dimensional Normed Linear Spaces

Theorem. *Any finite-dimensional subspace of a normed linear space is closed.*

The theorem does not require that the original normed linear space is itself finite-dimensional, but simply that the subspace under consideration is finite-dimensional. The proof relies on the fact that a finite-dimensional subspace can be shown to be isomorphic to $\mathbb{R}^n$, and any subspace of $\mathbb{R}^n$ is closed.

The following lemma will assist in our proof of the next theorem.

Lemma. *Let* Y *be a linear subspace of* X*, and suppose that* Y *is a true subset of* X *so that* $Y \neq X$*. Then for any* $0 < \epsilon < 1$ *there exists* $z_\epsilon \in X$ *such that* $||z_\epsilon - e|| \geq 1 - \epsilon$ *for all* $y \in Y$*.*

Proof. Since $Y \neq X$, there exists some $x_0 \in X \backslash Y$. Define $\delta = d(x_0, Y) > 0$ and $\eta_\epsilon = \frac{\epsilon}{1-\epsilon}\delta$. Then there exists $y_\epsilon \in Y$ such that

$$\delta \leq ||x_0 - y_\epsilon|| \leq \delta + \eta_\epsilon$$

Now let $z_\epsilon = \frac{x_0 - y_\epsilon}{||x_0 - y_\epsilon||}$ so that $||z_\epsilon|| = 1$. Then for any $y \in Y$,

$$||z_\epsilon - y|| \geq \frac{\delta}{\delta + \eta_\epsilon} = 1 - \frac{\eta_\epsilon}{\delta + \eta_\epsilon} = 1 - \epsilon$$

■

Theorem. *A normed linear space is finite-dimensional if and only if every bounded subset is relatively compact, i.e. the closure of every bounded subset is compact.*

Proof. Suppose that X is a finite-dimensional linear space with basis $\{e_1, \ldots, e_k\}$. Let $y_m = \sum_{k=1}^n \lambda_k^{(m)} e_k$ be a sequence contained in a bounded subset of X, then the sequences $\lambda_k^{(m)}$ (in m) are bounded sequences of real numbers and therefore contain convergence subsequences. Using a diagonalization argument, we can construct a sequence y_{m_j} such that $y_{m_j} \to y = \sum_{k=1}^n \lambda_k e_k$. Therefore, X is sequentially compact.

Now suppose that every bounded subset of X is relatively compact. Assume that the dimension of X is infinite. Let $x_1 \in X$ be some element such that $||x_1|| = 1$, and define $Y_1 = \text{span}\{x_1\}$. Then $Y_1 \neq X$, and by the lemma there exists some $x_2 \in X \backslash Y_1$ such that $||x_2|| = 1$ and $||x_2 - x_1|| > \frac{1}{2}$. Repeating the argument with $Y_2 = \text{span}\{x_1, x_2\}$, there is some $x_3 \in X \backslash (Y_1 \cup Y_2)$ such that $||x_3 - x_j|| > \frac{1}{2}$ for $j = 1, 2$.

We can now construct a sequence $x_k \in X$ such that $||x_k|| = 1$ such that $||x_k - x_j|| > \frac{1}{2}$ for $k > j$. But this sequence has no convergent subsequence, a contradiction. ■

Linear Operators

Definition. Let X and Y be normed linear spaces. Any function $T : D_T \to T(D_T)$, where $D_T \subseteq X$ and $T(D_T) \subseteq Y$, is called a transformation, mapping, or operator.

(i) T is said to be linear if $T(\lambda_1 x_1 + \lambda_2 x_2) = \lambda_1 T(x_1) + \lambda_2 T(x_2)$ for any $\lambda_1, \lambda_2 \in \mathbb{R}$ and $x_1, x_2 \in D_T$.

(ii) T is said to be onto (surjective) if $T(D_T) = Y$.

(iii) T is said to be one-to-one (injective) if $Tx_1 = Tx_2$ implies $x_1 = x_2$.
(iv) T is said to be continuous at $\hat{x} \in D_T$ if $x_n \to \hat{x}$ implies that $Tx_n \to T\hat{x}$.
(v) T is said to be bounded if there exists some $M > 0$ such that $\|Tx\|_Y \leq M\|x\|_X$ for any $x \in D_T$.

Theorem. *Let $T : X \to Y$ be linear, then T is continuous on all of X if and only if T is continuous at a single point $x_0 \in X$.*

Proof. If T is continuous everywhere on X then it is certainly continuous at x_0. Conversely, we suppose that T is continuous at x_0. Let $y \in X$, and suppose that $y_n \to y$. Let $x_n = y_n - y + x_0$, such that $x_n \to x_0$. By the continuity at x_0, it follows that $Ty_n - T_y + Tx_0 = Tx_n \to Tx_0$, and so $Ty_n \to Ty$. Therefore, T is continuous on all of X. ■

Theorem. *Let X and Y be normed linear spaces. Then $T : X \to Y$ is continuous if and only if T is bounded.*

Proof. First, suppose that T is bounded, and let $M > 0$ be a real number such that $\|Tx\|_Y \leq M\|x\|_X$ for all $x \in X$. By the previous theorem, we simply need to show that T is continuous at 0 in order to show it is continuous everywhere. Let $x_n \to 0$, then $\|Tx_n\|_Y \leq M\|x_n\|_X \to 0$, and so $Tx_n \to 0$.

For the other direction, suppose that T is not bounded, so for any $n \in \mathbb{N}$ there exists some $x_n \in X$ such that $\|Tx_n\|_Y > M\|x_n\|_X$. This implies that $x_n \neq 0$, since this would contradict the strict inequality. Now let $y_n = \frac{x_n}{n\|x_n\|_X}$, such that $\|y_n\|_X = \frac{1}{n} \to 0$, which implies that $y_n \to 0$. Then the following inequality holds:

$$\|Ty_n\| = \left\| \frac{Tx_n}{n\|x_n\|_X} \right\| > 1$$

However, this implies that $Ty_n \nrightarrow 0$ when $y_n \to 0$, and so T is not continuous at 0. Therefore, T is continuous if and only if it is bounded. ■

We denote the set of all linear mappings from X to Y by $\mathcal{L}(X, Y)$, and the set of all continuous (or bounded) mappings by $\mathcal{B}(X, Y)$. If $X = Y$, we simply write $\mathcal{L}(X)$ and $\mathcal{B}(X)$. In order to make these sets linear spaces themselves, we define addition and scalar multiplication of linear maps by their pointwise addition and multiplication. We introduce the norm for the space $\mathcal{B}(X, Y)$ with the following equivalent expressions:

$$\|T\| = \sup_{\substack{x \in X \\ x \neq 0}} \frac{\|Tx\|_Y}{\|x\|_X} = \sup_{\substack{x \in X \\ \|x\|_X = 1}} \|Tx\|_Y = \sup_{\substack{x \in X \\ \|x\|_X \leq 1}} \|Tx\|_Y$$

This norm is defined for elements in $\mathcal{L}(X, Y)$ which may potentially be unbounded, however for certain operators the supremum may be ∞. This norm is typically called an induced norm, since it is constructed from the norms on X and Y.

Theorem. *If Y is a Banach space, then $\mathcal{B}(X, Y)$ is a Banach space.*

Proof. Let T_n be a Cauchy sequence in $\mathcal{B}(X, Y)$, such that for any $\epsilon > 0$ the exists $N \in \mathbb{N}$ such that $m, n \geq N$ implies that

$$\|T_n - T_m\| < \epsilon$$

We claim that there is a uniform bound on all T_n. Since T_n is Cauchy, there exists some $N_1 \in \mathbb{N}$ such that $\|T_n\| \leq \|T_n - T_{N_1}\| + \|T_{N_1}\| \leq 1 + \|T_{N_1}\|$ for $n \geq N_1$, and so $\max\{\|T_1\|, \|T_2\|, \ldots, \|T_{N_1-1}\|, \|T_{N_1}\| + 1\}$ bounds the norm of every T_n.

Now since $\|T_n x - T_m x\|_Y \leq \|T_n - T_m\| \cdot \|x\|_X$ for some $x \in X$, it follows that $T_n x$ is a Cauchy sequence in Y. Then there exists $y \in Y$ such that $T_n x \to y$. We define an operator $T : X \to Y$ by $Tx = \lim_{n\to\infty} T_n x$. We now need to show that T is linear and bounded, and that $T_n \to T$ under the induced norm. The fact that T is linear is trivial. To see that T is bounded, let M be the uniform bound on the norms of T_n, then

$$\|Tx\|_Y \leq \|Tx - T_n x\|_Y + \|T_n x\|_Y \leq \|Tx - T_n x\|_Y + M\|x\|_X$$

Since we can make $\|Tx - T_n x\|_Y$ as small as we would like, it follows that $\|Tx\|_Y \leq$

$M\|x\|_X$. Therefore, $T \in \mathcal{B}(X,Y)$.

Now to show that $T_n \to T$. By T_n Cauchy, given $\epsilon > 0$ choose $N \in \mathbb{N}$ such that for $m, n \geq N$,

$$\begin{aligned}\|T_n x - Tx\|_Y &\leq \lim_{m\to\infty} \|T_n x - T_m x\|_Y \\ &\leq \lim_{m\to\infty} \|T_n - T_m\| \|x\|_X \\ &\leq \epsilon \lim_{m\to\infty} \|x\|_X = \epsilon\|x\|_X\end{aligned}$$

Therefore, $\|(T_n - T)x\|_Y \leq \epsilon\|x\|_X$ for all $x \in X$, and so $T_n \to T$ in $\mathcal{B}(X,Y)$. Since every Cauchy sequence can be shown to converge, $\mathcal{B}(X,Y)$ is Banach. ■

It is interesting to note that the previous proof is similar to the one used to show that $C(X)$ is complete if X is compact.

Definition. Let X be a Banach space. X is called a Banach algebra if there exists a binary product on the space denoted by xy such that

(i) X is a ring with addition performed normally in the linear space and multiplication given by xy.
(ii) $\|xy\| \leq \|x\|\|y\|$ for all $x, y \in X$.

The continuous functions on a compact metric space $C(X)$ form a Banach algebra, with addition, multiplication, and scalar multiplication performed as usual (note that this X is not the space referred to in the definition). $\mathcal{B}(X)$ is also a Banach algebra, with multiplication defined through composition as $(ST)x = S \circ Tx$. If $X = \mathbb{R}^n$, then the space $\mathcal{B}(X)$ is equivalent to the space of $n \times n$ matrices.

Fundamental Theorems in Banach Space Theory

Recall the statement of the Baire Category Theorem, that a nonempty complete metric space is of the second category, meaning it cannot be written as the countable union of nowhere dense subsets. Also recall that a nowhere dense subset is one whose closure has empty interior. This will assist us in our proof of the next theorem.

Theorem (Banach-Steinhaus Theorem, Principle of Uniform Boundedness)**.** *Let X be a Banach space and Y be a normed linear space. Let $\{T_\alpha\} \subseteq \mathcal{B}(X,Y)$ be a collection of bounded operators from X to Y. If the collection $\{T_\alpha x\}$ is bounded for any $x \in X$, then the family $\{T_\alpha\}$ is uniformly bounded on X, i.e. there exists $M > 0$ such that* $\sup_\alpha \|T_\alpha\| \leq M$.

Proof. Define $X_n = \{x \in X : \sup_\alpha \|T_\alpha x\| \leq n\}$. We claim that each X_n is closed, and that $X = \bigcup_{n=1}^\infty X_n$. Since X is complete, by the Baire Category Theorem we know that X is of the second category. This is implies that there exists at least one X_{n_0} that is nowhere dense. So there exists $x_0 \in X_{n_0}$ and $\overline{B_X(x_0, r_0)} \subseteq X$, which is to say that x_0 is an interior point of X_{n_0}. Now for some $z \in \overline{B_X(0,1)}$, let $x = x_0 + r_0 z$ such that $z = \frac{x - x_0}{r_0}$ and $x \in \overline{B_X(x_0, r_0)} \subseteq X_{n_0}$. It follows that

$$\|T_\alpha z\|_Y \leq \frac{1}{r_0}\Big(\|T_\alpha x\|_Y + \|T_\alpha x_0\|_Y\Big) \leq \frac{2n_0}{r_0}$$

This inequality holds for all $z \in \overline{B_X(0,1)}$, and so

$$\|T_\alpha\| = \sup_{\|z\| \leq 1} \|T_\alpha z\|_Y \leq \frac{2n_0}{r_0}$$

■

Definition. A collection of operators $\{T_n\} \subseteq \mathcal{B}(X,Y)$ is said to be strongly convergent if $\lim_{n\to\infty} T_n x$ exists for all $x \in X$. If there exists some $T \in \mathcal{B}(X,Y)$ such that $\lim_{n\to\infty} T_n x$ for all $x \in X$, then we say that T_n is strongly convergent to T.

Theorem. *Let X be a Banach space, Y a normed linear space, and suppose that the collection of operators $\{T_n\} \subseteq \mathcal{B}(X,Y)$ is strongly convergent. Then there exists some $T \in \mathcal{B}(X,Y)$ such that T_n is strongly convergent to T.*

Suppose now that we have an operator $T \in \mathcal{B}(X,Y)$ that is a bijection, then T^{-1} certainly exists. We are curious what conditions must be placed on T so that we can ensure that T^{-1} is bounded. We first establish the following lemma, which is used in the proof of the next theorem. Although the assumption that X and Y are Banach spaces is not used explicitly in the theorem, it is used in the proof of the lemma.

Lemma. *Let X and Y be Banach spaces, with $T \in \mathcal{B}(X,Y)$ being a surjective operator. Then for any $\epsilon > 0$ there exists some $\delta > 0$ such that $T(B_X(0, 2\epsilon)) \supseteq B_Y(0, \delta)$.*

Theorem (Open Mapping Theorem)**.** *Suppose X and Y are Banach spaces, and $T \in \mathcal{B}(X,Y)$ is a surjective bounded operator. Then T maps open sets onto open sets, such that if A is an open set in X then $T(A)$ is an open set in Y.*

Proof. Let $G \subseteq X$ be open. For all $y \in T(G)$ there exists some $x \in G$ such that $Tx = y$. Since G is open, there exists $\epsilon > 0$ such that $B_X(x, \epsilon) \subseteq G$. By the lemma, there exists $\delta > 0$ such that $B_Y(0, \delta) \subseteq T(B_X(0, \epsilon))$, so by linearity,

$$B_Y(y, \delta) \subseteq T(B_X(x, \epsilon)) \subseteq T(G)$$

So for any $y \in T(G)$ there exists a neighborhood $B_Y(y, \delta)$ also contained in $T(G)$, making $T(G)$ open. ■

Corollary. *If $T \in \mathcal{B}(X,Y)$ is a bijection, and $T^{-1} : Y \to X$ is the operator defined by $T^{-1}y = x$ where $Tx = y$, then $T^{-1} \in \mathcal{B}(Y, X)$.*

Proof. This follows immediately from the definition of continuity, since the preimage (through T^{-1}) of any open set is open by the open mapping theorem, due to the fact that T is a bijection. ■

Corollary. *Let $\|\cdot\|_1$ and $\|\cdot\|_2$ be two norms defined on a linear space X, such that X is a Banach space under both norms. Suppose that there exists some $M > 0$ such that $\|x\|_1 \leq \|x\|_2$ for any $x \in X$. Then the norms are equivalent.*

Proof. We write X_1 for X equipped with the $\|\cdot\|_1$ norm, and likewise for X_2. Define $T : X_1 \to X_2$ given by $Tx = x$. T is clearly continuous, so $T \in \mathcal{B}(X_1, X_2)$, and moreover T is a bijection. Therefore, $T^{-1} \in \mathcal{B}(X_2, X_1)$, which implies that there exists $K > 0$ such that

$$\|Tx\|_2 = \|x\|_2 \leq K\|x_1\|$$

Thus, the norms are equivalent. ■

During the previous week we used the following lemma to prove the Open Mapping Theorem. We provide a proof here. It should be mentioned that the fact that X and Y are both Banach spaces is used for this proof, and is therefore necessary for the eventual proof of the Open Mapping Theorem. The primary use of the completeness of Y is to establish that Y is of the second category.

Lemma. *Let X and Y be Banach spaces, with $T \in \mathcal{B}(X,Y)$ being a surjective operator. Then for any $\epsilon > 0$ there exists some $\delta > 0$ such that $T(B_X(0,2\epsilon)) \supseteq B_Y(0,\delta)$.*

Proof. Let $\epsilon > 0$ be given. Since $X = \bigcup_{n=1}^{\infty} B_X(0,n\epsilon)$ it follows that

$$Y = T(X) = \bigcup_{n=1}^{\infty} T(B_X(0,n\epsilon)) = \bigcup_{n=1}^{\infty} nT(B_X(0,\epsilon))$$

Since Y is complete, it is of the second category and therefore cannot be written as the countable union of nowhere dense subsets (sets whose closure contains no interior points). Then there must exist some $n_0 \in \mathbb{N}$ such that $n_0\overline{T(B_X(0,\epsilon))}$ has at least one interior point. This implies that there exists some $z \in Y$, $r > 0$ such that $B_Y(z,r) \subseteq n_0\overline{T(B_X(0,\epsilon))}$. Letting $y_0 = \frac{z}{n_0} \in Y$ and $\delta = \frac{r}{n_0} > 0$, we have that $B_Y(y_0,\delta) \subseteq \overline{T(B_X(0,\epsilon))}$. We now define the sets

$$P = \{y_1 - y_2 : y_1, y_2 \in B_Y(y_0,\delta)\} \qquad Q = \{x_1 - x_2 : x_1, x_2 \in B_X(0,\epsilon)\}$$

By linearity and continuity of T, $P \subseteq \overline{T(Q)}$. Since $Q \subseteq B_X(0,2\epsilon)$, it follows that $P \subseteq \overline{T(Q)} \subseteq \overline{T(B_X(0,2\epsilon))}$. Now for any $y \in B_Y(0,\delta)$, we can write $y = (y + y_0) - y_0 \in P$, and so we finally have that $B_Y(0,\delta) \subseteq \overline{T(B_X(0,2\epsilon))}$.

We now only need to show that the closure of $T(B_X(0,2\epsilon))$ is unnecessary. By what we've just shown, for any $\epsilon_0 > 0$ there exists $\delta_0 > 0$ such that $B_Y(0,\delta_0) \subseteq \overline{T(B_X(0,\epsilon_0))}$. Choose some sequence $\epsilon_n > 0$ such that $\epsilon_n > \epsilon_{n+1} \to \epsilon_0$ and that $\sum_{n=1}^{\infty} \epsilon_n < \epsilon_0$. Then there exist $\delta_n \to 0$ such that $B_Y(0,\delta_n) \subseteq \overline{T(B_X(0,\epsilon_n))}$.

Now let $y \in B_Y(0,\delta_0)$, then there exists $x_0 \in B_X(0,\epsilon_0)$ such that $\|y - Tx_0\| < \delta_1$. This further implies that there exists $x_1 \in B_X(0,\epsilon_1)$ such that $\|y - Tx_0 - Tx_1\| < \delta_2$. By repeating the process for arbitrary n,

$$\left\| y - \sum_{j=0}^{n} Tx_j \right\| < \delta_{n+1}$$

Since $\|x_n\| \leq \epsilon_n$, it follows that $\sum_{n=0}^{\infty} \|x_n\| < 2\epsilon_0$. By a homework problem, we know that since X is a Banach space and x_n is absolutely convergent, it is also convergent. Then $x = \sum_{n=0}^{\infty} x_n \in B_X(0,2\epsilon_0)$ exists, and $\|y - Tx\| = 0$, implying that $B_Y(0,\delta_0) \subseteq T(B_X(0,2\epsilon_0))$. ∎

Definition. Let $T : D_T \to Y$, $D_T \subseteq X$ be an operator. The graph of T is the set

$$G_T = \{(x, Tx) : x \in D_T\}$$

T is said to be a closed operator if G_T is a closed set in $X \times Y$.

Theorem. *Let X and Y be Banach spaces, $T : D_T \to Y$ linear. If T is closed, then $T \in \mathcal{B}(D_T, Y)$.*

Linear Functionals

Definition. Any operator in $\mathcal{L}(X, \mathbb{R})$ or $\mathcal{L}(X, \mathbb{C})$ is called a linear functional. We call $X^* = \mathcal{B}(X, \mathbb{R})$ or $\mathcal{B}(X, \mathbb{C})$ the dual space of X.

Theorem. *The dual space X^* is always a Banach space.*

Proof. The proof follows immediately from the fact that X^* maps onto either $\mathbb{R}$ or $\mathbb{C}$, both of which are Banach spaces. ■

Out next goal will be to prove the Hahn-Banach Theorem. Suppose that we begin with a normed linear space X which has some linear subspace Y. This subspace has a dual space Y^*, and elements of the dual space $y^* \in Y^*$ are linear functionals that act on elements of Y. Since y^* is linear and Y is simply a subspace of X, we might be interested in defining y^* on the more general elements of X as well. The Hahn-Banach Theorem essentially tells us that this is possible. Specifically, it says that there exists some $x^* \in X^*$ which agrees with y^* on the elements in Y, but of course is also linear on the rest of X. To lead us towards the proof, we begin with the following lemma.

Lemma (Hahn-Banach Lemma). *Let X be a real normed linear space and $p : X \to \mathbb{R}$ satisfying*

(i) $p(\lambda x) = \lambda p(x)$ for all $\lambda > 0$ and $x \in X$.
(ii) $p(x + y) \leq p(x) + p(y)$ for all $x, y \in X$.

If $Y \subseteq X$ is a linear subspace and $f \in Y^*$ such that $f(y) \leq p(y)$ for all $y \in Y$, then there exists $F \in X^*$ such that $F|_Y = f$ and $F(x) \leq p(x)$ on X.

Theorem (Hahn-Banach Theorem)**.** *Let X be a normed linear space, with a linear subspace $Y \subseteq X$. Then for any $y^* \in Y^*$, there exists some $x^* \in X^*$ such that $x^*|_Y = y^*$ and $\|x^*\| = \|y^*\|$.*

Proof. If X is a real normed linear space, let $p(x) = \|y^*\|\|x\|$ and $f(y) = y^*(y)$. By the Hahn-Banach lemma, there exists $F \in X^*$ such that $F|_Y = f$ and $F(x) \leq p(x)$ on X. Choosing $x^* = F$ provides the result.

On the other hand, if X is a complex normed linear space, we write $y^*(y) = f_1(y) + if_2(y)$ where $f_1, f_2 \in \mathcal{B}(Y, \mathbb{R})$. We define $p(x) = \|y^*\|\|x\|$. Now regarding X as a real normed linear space, by the lemma there exists $F_1 \in \mathcal{B}(X, \mathbb{R})$ such that $F_1|_Y = f_1$ and $F_1(x) \leq p(x)$ on X. Next we define $F(x) = F_1(x) - iF_1(ix)$, and set $x^* = F$. ■

Theorem. *Let X be a normed linear space, $Y \subset X$ a subspace. Let $x_0 \in X$ such that $\delta = \inf_{y \in Y} \|x_0 - y\| > 0$ (or equivalently that $x_0 \notin \overline{Y}$). Then there exists $x^* \in X^*$ such that $x^*(x_0) = 1$, $\|x^*\| = \frac{1}{\delta}$, and x^* is 0 on Y.*

Proof. Let $Y_! = \operatorname{span}\{Y, x_0\} \subseteq X$. Define $z^* \in Y_1^*$ by $z^*(x) = z^*(y + \lambda x_0) = \lambda$. Then $z^*(x_0) = 1$ and $z^*(y) = 0$ for any $y \in Y$. Now if $\lambda \neq 0$ then

$$\|x\| = \|y + \lambda x_0\| = |\lambda| \left\| \frac{y}{\lambda} + x_0 \right\| \geq |\lambda| \delta$$

and so $\|z^*(x)\| = |\lambda| \leq \frac{\|x\|}{\delta}$ on Y_1, which implies that $\|z^*\| \leq \frac{1}{\delta}$. Now choose $y_n \in Y$ such that $\delta \leq \|x_0 - y_n\| \leq \delta + \frac{1}{n}$. Then

$$1 = z^*(x_0 - y_n) \leq \|z^*\|\|x_0 - y_n\| \to \delta$$

Which implies that $\|z^*\| = \frac{1}{\delta}$. By the Hahn-Banach Theorem, there exists $x^* \in X$ such that x^* agrees with z^* on Y_1, and $\|x^*\| = \|z^*\| = \frac{1}{\delta}$. ■

Corollary. *For any $x \in X$, there exists $x^* \in X^*$ such that $\|x^*\| = 1$ and $x^*(x) = \|x\|$.*

Proof. Letting $Y = \{0\}$, and $\delta = \inf_{y \in Y} \|x - y\| = \|x\|$, by the previous theorem we have that there exists $z^* \in X^*$ such that $\|z^*\| = \frac{1}{\|x\|}$ and $z^*(x) = 1$. Then $x^* = \|x\|z^*$ is the element we want. ■

Corollary.

(i) *For any* $y, z \in X$, $y \neq z$, *there exists some* $x^* \in X^*$ *such that* $x^*(y) \neq x^*(z)$. *Essentially, there exist non-trivial elements in* X^*.

(ii)
$$\|x\| = \sup_{\substack{x^* \in X^* \\ x^* \neq 0}} \frac{|x^*(x)|}{\|x^*\|} = \sup_{\|x^*\|=1} |x^*(x)|$$

(iii) *If* $Y \subset X$ *is a subspace that is not dense in* X, *then there exists some nonzero* $x^* \in X^*$ *such that* x^* *is 0 on* Y.

(iv) *Let* $N_{x^*} = \{x \in X : x^*(x) = 0\}$ *be the null space of* x^*. *There exists some* $x_0 \in X$ *such that* $X = N_{x^*} \oplus span\{x_0\}$.

Definition. For nonzero $x^* \in X^*$, the set $H_c = \{x \in X : \mathrm{Re}x^*(x) = c\}$ is called a hyperplane in X.

Dual Spaces and Reflexive Spaces

Theorem. *Let* X *be a normed linear space. If* X^* *is separable, then* X *is separable.*

The converse to this theorem is generally not true. As an counter example, let $X = L^1(\Omega)$ and $X^* = L^\infty(\Omega)$. Although X is separable, X^* is not.

Definition. Let X and Y be normed linear space.

(i) X and Y are linearly isomorphic if there exists a bijective $T \in \mathcal{L}(XY)$. Such a T is called an isomorphism.

(ii) X and Y are isometrically isomorphic if there exists a bijective $T \in \mathcal{L}(X, Y)$ such that $\|Tx\|_Y = \|x\|_X$ for all $x \in X$.

(iii) The space X^{**} is defined as the dual space of X^*, also called the double-dual space of X.

(iv) For a fixed $x \in X$, we define $\hat{x}$ as the element of X^{**} such that $\hat{x}(z^*) = z^*(x)$ for $z^* \in X^*$.

(v) We define $\kappa : X \to X^{**}$ by $\kappa(x) = \hat{x}$. κ is a linear operator, and is called the natural embedding of X into X^{**}. We write $\hat{X} = \kappa(X)$, and $\hat{X}$ is a linear subspace of X. κ is an isometric isomorphism between X and $\hat{X}$.

Theorem (Dual of Banach-Steinhaus)**.** *Let* $\{x_\alpha\} \subseteq X$ *such that* $\sup_\alpha |x^*(x_\alpha)| < \infty$ *for any* $x^* \in X^*$. *Then* $\sup_\alpha \|x_\alpha\| < \infty$.

Proof. Let $\hat{x}_\alpha = \kappa(x_\alpha) \in X^{**}$. By the Banach-Steinhaus Theorem, $\sup_\alpha \|\hat{x}_\alpha\| < \infty$. This provides the result directly, since $\|x_\alpha\| = \|\hat{x}_\alpha\|$. ■

Definition. A normed linear space is called reflexive if $\hat{X} = \kappa(X) = X^{**}$.

The reflexivity of a normed linear space imposes a very strict structure on the dual spaces of X. Not only does the fact that $\kappa(X) = X^{**}$ imply that X is isometrically isomorphic to $X^{(2n)*}$ for any $n \in \mathbb{N}$ (all double dual spaces), but the next theorem shows that it also guarantees that all the $X^{(2n-1)*}$ are isometrically isomorphic as well.

Theorem.

(i) Any closed subspace of a reflexive normed linear space is reflexive.
(ii) X is reflexive if and only if X^* is reflexive.