Applied Linear Analysis

Course Notes by Xiaobing Feng
Transcribed and Annotated by Gregory Zitelli

Weak Convergence

Definition. Given $x_n, x \in X$, we say that x_n converges to x weakly if for any $x^* \in X^*$ we have that $\lim_n x^*(x_n) = x^*(x)$. The point x is called the weak limit of x_n, written $x_n \xrightarrow{w} x$ or $x_n \rightharpoonup x$.

The weak limit of a sequence is always unique, and can be used to define notions of weakly sequentially compact sets, weakly bounded sets, weakly closed sets, and weakly Cauchy sequences. As its name implies, weak convergence is implies by strong convergence (convergence in the norm), but does not necessarily imply strong convergence.

Theorem. *Suppose that $x_n \rightharpoonup x$, then*

(i) $\{\|x_n\|\}$ *is bounded.*
(ii) $\|x\| \leq \liminf_n \|x_n\|$

Theorem. *Let X be reflexive. The subset $K \subseteq X$ is weakly sequentially compact if and only if K is bounded and weakly closed.*

Proof. Suppose that K is weakly sequentially compact. Let $x_n \in K$ and $x_n \rightharpoonup x \in X$. By hypothesis, there exists $x_{n_j} \rightharpoonup y \in K$, but the uniqueness of the limit means that $x = y \in K$. Therefore, K is weakly closed.

To show that K is bounded, suppose not. Then for any $n \geq 1$ there exists $x_n \in K$ such that $\|x_n\| \geq n$. Then there exists some x_{n_k} such that $x_{n_k} \rightharpoonup z \in K$, then $\|z\| < \infty$ which contradicts the fact that $\|x_{n_k}\| \geq n_k \to \infty$.

Next, assume that K is bounded and weakly closed. Let $x_j \in K$ be a sequence bounded by some $C > 0$. Define $Z = \mathrm{span}\{x_j\} \subseteq K$, a separable closed subspace of

X. By the previous theorems, Z is also reflexive and Z^* is separable. Let $z_n^* \in Z^*$ be a dense sequence.

$$|z_n^*(x_j)| \leq \|z_n^*\| \, \|x_j\| \leq C \, \|z_n^*\|$$

Holding n fixed, the sequences in j of real numbers $z^*(x_j)$ are bounded, and so there exists convergent subsequences. Now using diagonalization we construct a subsequence $x_{j_k} = y_k$ such that $\lim_k z_n^*(y_k)$ exists. By our homework, since such a construction works for a dense subset z_n^* of Z^*, the limit $\lim_k z^*(x_n)$ exists for any $z^* \in Z^*$. Therefore, $\lim_k z^*(y_k) = \lim_k \kappa(y_k)(z^*) = \lim_k y_k^{**}(z^*)$ exists, and y_k^{**} converges pointwise on Z^*. Then by our theorems, there exists some $y^{**} \in Z^{**}$ such that $\lim_k y_k^{**}(z^*) = y^{**}(z^*)$ for all $z^* \in Z^*$. By the reflexivity of Z, there is $y \in Z$ such that $\kappa(y) = y^{**}$, and so $\lim_k y_k^{**}(z^*) = \lim_k z^*(y_k) = y^{**}(z^*) = z^*(y)$. Thus, $x_{n_k} \rightharpoonup y \in Z \subseteq K$. ■

Theorem. *A reflexive normed linear space is always weakly complete.*

Definition. A sequence of linear functionals $x_n^* \in X^*$ is said to be weakly (or weakly*) convergent to $x^* \in X^*$ if $\lim_n x^*(x) = x^*(x)$ for all $x \in X$. We say that x^* is the weak limit of x_n^*.

Again, the weak* limit of a sequence of functionals is always unique. Notice that the definition of weak and weak* convergence are mirror images of one another, with $x_n \xrightarrow{w} x$ when $x^*(x_n) \to x^*(x)$ for every $x^* \in X$, while $x_n^* \xrightarrow{w^*} x$ when $x_n^*(x) \to x^*(x)$ for every $x \in X$. The proof of the following theorem is very similar to the previous one, and involves using a similar diagonalization argument to form a subsequence of the x_n^* sequence.

Theorem. *Suppose X is separable, then every bounded sequence $x_n^* \in X^*$ has a weakly convergent subsequence.*

Both the weak and weak* convergence induce what is called the weak topology on X and X^*, respectively. This leads to notions of weak and weak* compactness. The weak topology on X is sometimes called the initial topology with respect to X^*. Specifically, it is the topology with the fewest open sets such that the functionals X^* remain continuous on X. The following theorem shows a major difference between the standard topology on X^* and the weak* topology.

Theorem (Banach-Alaoglu Theorem)**.** *If X is a Banach space, then the closed unit ball $\overline{B(0,1)} \subseteq X^*$ is compact in the weak* topology.*

Adjoint Operators

In finite-dimensional linear algebra, the adjoint of a linear operator thought of as a matrix $A \in \mathbb{C}^{n\times m}$ is the conjugate-transpose matrix $A^* \in \mathbb{C}^{m\times n}$. The matrices satisfy $\langle Ax, y\rangle = \langle x, A^*y\rangle$ for $x, y \in \mathbb{C}^m$. We wish to construct a similar adjoint for operators on arbitrary normed linear spaces.

Definition. Let $T \in \mathcal{B}(X,Y)$. Define $T^* : Y^* \to X^*$ by $T^*y^*(x) = y^*(Tx)$ for all $x \in X$ and $y^* \in Y^*$.

Theorem. *Let $T \in \mathcal{B}(X,Y)$, $S \in \mathcal{B}(Y,Z)$.*

(i) $T^* \in \mathcal{B}(Y^*, X^*)$

(ii) We define the operator $\sigma : \mathcal{B}(X,Y) \to \mathcal{B}(Y^*,X^*)$ by $\sigma(T) = T^*$. Then σ is a linear operator.

(iii) $(ST)^* = T^*S^*$

(iv) For any $T \in \mathcal{B}(X,Y)$, we define $\hat{T} \in \mathcal{B}(\hat{X},\hat{Y})$ by $\hat{T}(\hat{x}) = \hat{y}$ whenever $Tx = y$. If X is reflective, then $(T^*)^* = T^{**} = \hat{T}$.

Theorem. *Let X be a Banach space, and Y be a normed linear space. Then for any $T \in \mathcal{B}(X,Y)$, there exists an inverse T^{-1} which is bounded if and only if $(T^*)^{-1}$ exists and is bounded. In that case, $(T^*)^{-1} = (T^{-1})^*$.*

Definition.

(i) Given $A \subseteq X$, define $A^\perp = \{x^* \in X^* : \forall x \in A,\ x^*(x) = 0\} \subseteq X^*$. $A^\perp$ is called the orthogonal complement of A.

(ii) Given $\Sigma \subseteq X^*$, define $\Sigma^\perp = \{x \in X : \forall x^* \in \Sigma,\ x^*(x) = 0\} \subseteq X$. $\Sigma^\perp$ is called the orthogonal complement of Σ.

(iii) For any $T \in \mathcal{B}(X,Y)$, $T^* \in \mathcal{B}(Y^*,X^*)$, we define the null spaces of T and T^* as $N_T = \{x \in X : Tx = 0\}$ and $N_{T^*} = \{y^* \in Y^* : T^*y^* = 0\}$.

(iv) We define the range or image of T and T^* as $R_T = \{y \in Y : \exists x \in X,\ Tx = y\}$ and $R_{T^*} = \{x^* \in X^* : \exists y^* \in Y^*,\ T^*y^* = x^*\}$.

Theorem. *Let $T \in \mathcal{B}(X,Y)$, then $\overline{R_T} = N_{T^*}^\perp$.*

We mention that the similar expression $\overline{R_{T^*}} = N_T^\perp$ is generally not true. When X and Y are Banach spaces and R_T is closed, then it follows that R_{T^*} is closed and $R_{T^*} = N_T^\perp$.

Corollary. *If $N_{T^*} = \{0\}$ (so T^* is injective) and R_T is closed then $R_T = Y$ (T is surjective). Therefore, $Tx = y$ has a unique solution $x \in X$ for any $y \in Y$.*

Compact Operators

When considering operators onto a finite-dimensional space Y, we have the property that all bounded subsets of Y are relatively compact. Consequently, any bounded operator $T \in \mathcal{B}(X,Y)$ maps bounded sets in X onto relatively compact subsets of Y. This may not be true in general, however, when Y is not finite-dimensional. Consequently, we are interested in operators that retain this property, such that the image of bounded sets are relatively compact.

Definition. Let X and Y be normed linear spaces. An operator $T \in \mathcal{B}(X,Y)$ is called compact (or completely continuous) if T maps any bounded subset of X into a compact set of Y (onto a relatively compact set).

Theorem. *T maps any weakly convergent sequence into a strongly convergent sequence (a sequence that converges in the norm of X).*

Proof. Let $x_0, x_n \in X$ such that $x_n \rightharpoonup x_0$. We first claim that $Tx_n \rightharpoonup Tx_0$, which does not rely on the compactness of T. Let $y^* \in Y^*$, then

$$|y^*(Tx_n) - y^*(Tx_0)| = |(y^*T)x_n - (y^*T)x_0|$$

However, this converges to 0, since $y^*T \in X^*$. Next, we suppose that Tx_n does not converge in the norm to Tx_0. Then there exists some $\epsilon > 0$ and subsequence x_{n_k} such that $\|Tx_{n_k} - Tx_0\| \geq \epsilon$. Since $x_{n_k} \rightharpoonup x_0$, we know that $\{x_{n_k}\}$ is bounded. By the compactness of T, the set $\{Tx_{n_k}\}$ is relatively compact. Then there exists a subsequence $Tx_{n'_k}$ of Tx_{n_k} such that $Tx_{x'_k} \to \overline{y} \in Y$. By the uniqueness of the weak limit, we know that $\overline{y} = Tx_0$, however this contradicts the fact that $\|Tx_{n_k} - Tx_0\| \geq \epsilon$. ■

For the proof of the next theorem, recall that a set is called totally bounded if for any $\epsilon > 0$ there exists a finite ϵ-net for the set. A relatively compact is pre-compact, or totally bounded, if the underlying space is complete.

Theorem. *Suppose that X is a normed linear space, and Y is a Banach space. If the operators $T_n \in \mathcal{B}(X,Y)$ are compact, and $T_n \to T \in \mathcal{B}(X,Y)$, then this operator T is compact.*

Proof. Let $A \subseteq X$ be bounded. We want to prove that $\overline{T(A)}$ is compact, which is equivalent to showing that it is totally bounded since Y is complete. Now since $T_n \to T$, given $\epsilon > 0$ choose some $N_\epsilon > 1$ such that $n > N_\epsilon$ implies that $\|T_n x - T_x\| < \frac{\epsilon}{3}$ for all $x \in A$. Since T_{N_ϵ} is a compact operator, $\overline{T_{N_\epsilon}(A)}$ is compact in Y. Since Y is complete, $T_{N_\epsilon}(A)$ is pre-compact, so there exists some $\{x_j\}_{j=1}^m$ such that

$$\min_{1 \le j \le m} \|T_{N_\epsilon} x - T_{N_\epsilon} x_j\|_Y < \frac{\epsilon}{3}$$

for all $x \in A$. Therefore,

$$\begin{aligned} &\min_{1 \le j \le m} \|Tx - Tx_j\|_Y \\ &\le \min_{1 \le j \le m} \Big(\|Tx - T_{N_\epsilon} x\|_Y + \|T_{N_\epsilon} x - T_{N_\epsilon} x_j\|_Y + \|T_{N_\epsilon} x_j - T_{N_\epsilon} x_j\|_Y \Big) \\ &< \frac{\epsilon}{3} + \frac{\epsilon}{3} + \frac{\epsilon}{3} = \epsilon \end{aligned}$$

So $\{B_Y(Tx_j, \epsilon)\}_{j=1}^m$ forms an ϵ-net for $T(A)$, making it totally bounded. ∎

Theorem. *Let $S, T \in \mathcal{B}(X, Y)$. If T is compact, then ST and TS are compact.*

Theorem. *Let X be a normed linear space, and Y be a Banach space. Then $T \in \mathcal{B}(X, Y)$ is compact if and only if $T^* \in \mathcal{B}(Y^*, X^*)$ is compact.*

We provide a brief outline of the proof of the preceding theorem. We want to show that if T is compact then T^* is also compact. Since $T(X)$ is separable, we take a countable dense set A inside it. Given a bounded sequence $y_n^* \in Y^*$, we use a diagonalization procedure to construct a subsequence $y_{n_j}^*$ such that $\lim_j y_{n_j}^*(y)$ exists for all $y \in A$. We define $x_j^* = T^* y_{n_j}^*$, and it follows that $x_j^* \rightharpoonup x^*$ weakly (since it converges pointwise on the dense A). By showing that $x_j^* \to x^*$, we can demonstrate that $T^* y_n^*$ is sequentially compact, which is equivalent to it being compact by the completeness of Y.

For the other direction, we assume that T^* is compact and wish to show that T is compact. By the previous part, T^{**} is compact, and so for $x_n \in X$ bounded we take $\hat{x}_n = \kappa(x_n)$, also bounded. Then is $T^{**}\hat{x}_n$ pre-compact, and so there exists a subsequence $\hat{x}_{n_j}$ such that $\lim_j T^{**}\hat{x}_{n_j}$ exists. Since Y is a Banach space, so is $\mathcal{B}(X, Y)$ is as well. This implies that Tx_{n_j} is Cauchy, and so the limit exists and the sequential compactness again implies that T is compact.

Fredholm-Riesz-Schauder Theory

Suppose that $T \in \mathcal{B}(X, Y)$ is a compact operator, and that we wish to solve the operator equation $(\lambda I - T)x = y$ for some $x \in X$, $y \in Y$. We are curious when this equation has a solution, and when that solution is unique. We introduce the following notation to solve this kind of problem:

Definition.

(i) $T_\lambda = \lambda I - T \in \mathcal{B}(X, Y)$
(ii) $T_\lambda^* = \lambda I - T^* \in \mathcal{B}(Y^*, X^*)$
(iii) $N_\lambda = N_{T_\lambda} = \{x \in X : T_\lambda x = 0\}$
(iv) $N_\lambda^* = N_{T_\lambda^*} = \{x^* \in X^* : T_\lambda^* x^* = 0\}$
(v) $R_\lambda = R_{T_\lambda} = \{y \in X : \exists x \in X,\ y = T_\lambda x\}$
(vi) $R_\lambda^* = R_{T_\lambda^*} = \{y^* \in X^* : \exists x^* \in X,\ y^* = T_\lambda^* x^*\}$

Lemma.

(i) N_λ *and* N_λ^* *are finite-dimensional.*
(ii) R_λ *and* R_λ^* *are closed subspaces.*
(iii) $R_\lambda = (N_\lambda^*)^\perp$
(iv) $R_\lambda^* = N_\lambda^\perp$

This lemma establishes that the null space of the operators T_λ and T_λ^* are always finite-dimensional with closed range. Such an operator is called a Fredholm operator.

Corollary.

(i) $T_\lambda x = y$ *has a solution* ($y \in R_\lambda$) *if and only if* $y \in (N_\lambda^*)^\perp$.
(ii) $T_\lambda x = y$ *has a solution for any* $y \in Y$ *if and only if* $N_\lambda^* = \{0\}$.

Theorem. $R_\lambda = X$ *if and ony if* $N_\lambda = \{0\}$.

Theorem. $dim(N_\lambda) = dim(N_\lambda^*)$

The preceding two theorems and the final two elements of the lemma form what is referred to as the Fredholm-Riesz-Schauder theory for compact operators.

Theorem (The Fredholm Alternative)**.** *Let T be compact, $T \in \mathcal{B}(X)$. Then precisely one of the following properties must hold:*

(i) *For every $y \in X$ there exists a unique element $x \in X$ such that $T_\lambda x = y$.*
(ii) *There exists some nonzero $x \in X$ such that $T_\lambda x = 0$. This is equivalent to saying that λ is an eigenvalue of T.*

Definition. Let $T \in \mathcal{B}(X)$. Define $\rho(T) = \{\lambda \in \mathbb{C} : T_\lambda^{-1} \in \mathcal{B}(X)\}$ and $\sigma(T) = \mathbb{C} \backslash \rho(T)$. The set $\rho(T)$ is called the resolvent of T, and $\sigma(T)$ is called the spectrum of T.

The spectrum can be decomposed into the union of the continuous spectrum (value of λ for which R_λ is dense in X), the risidual spectrum (values of λ for which R_λ is not dense), and the eigenvalues of T (values of λ fr which T_λ^{-1} does not exist). We are interested in properties of the spectrum if the original operator T is compact.

Theorem. *If $T \in \mathcal{B}(X)$ is compact and X is an infinite-dimensional Banach space, then $\sigma(T)$ consists of 0 and either a finite number of eigenvalues, or and infinite number of eigenvalue which converge to 0.*

This reveals that the spectrum of a compact operator is precisely its eigenvalues and zero. Additionally, if a compact operator has infinitely many eigenvalues, then they converge to zero.

Schauder Fixed Point Theorem

We can now begin to develop a theory for fixed points of nonlinear operators. This is known as the Schauder Fixed Point Theorem, and generalizes our previous Brouwer's Fixed Point Theorem.

Corollary. *Suppose $dim(X) < \infty$, and let $K \subseteq X$ be a nonempty, compact, convex set. Let $T : K \to K$ be a continuous operator. Then there exists some $x \in K$ such that $Tx = x$.*

Definition. Let X and Y be normed linear spaces. For $E \subseteq X$ we say that the operator $T : E \to Y$ is compact if T is continuous $\overline{T(A)}$ is compact in Y for any bounded set $A \subseteq E$.

Theorem (Schauder Fixed Point Theorem)**.** *Let X be a normed linear space with a closed, bounded, convex subset $E \subseteq X$. If $T : E \to X$ is compact such that $T(E) \subseteq E$, then there exists some $x \in K$ such that $Tx = x$.*

The hypothesis that T is compact is often difficult to show, and so the following corollaries generalize the result by removing the condition on T and placing more restrictions on the underlying subset E.

Corollary. *Let X be a Banach space, $T : E \to X$ continuous such that $T(E) \subseteq E$. If E is a compact and convex subset of X, then there exists some $xinE$ such that $Tx = x$.*

Corollary. *Let X be a Banach space, $T : E \to X$ continuous. If E is closed an convex, and if $T(E) \subseteq E$ is pre-compact, then there exists $x \in E$ such that $Tx = x$.*

Theorem. *Let X be a Banach space, $T : X \to X$ compact. Let $S = \{x \in X : \exists \sigma \in [0,1],\ \sigma Tx = x\}$ be the set of points in X that T contracts. Then if there exists some bound $M > 0$ such that $\|x\|_X < M$ for all $x \in S$, then there exists some $x \in X$ such that $Tx = x$.*

Inner Product Spaces

Definition. An inner product is a function $\varphi : X \times X \to \mathbb{C}$ on a linear space X that satisfies the following properties for any $x, y, z \in X$ and $\lambda \in \mathbb{C}$:

(i) $\varphi(x, x) \geq 0$
(ii) $\varphi(x + y, z) = \varphi(x, z) + \varphi(y, z)$
(iii) $\varphi(\lambda x, y) = \lambda\varphi(xy)$
(iv) $\varphi(x, y) = \overline{\varphi(y, x)}$

A normed linear space with an inner product is called an inner product space. We typically write $(\cdot,\cdot)_X = \varphi(\cdot,\cdot)$ to denote the inner product on a set X. An inner product always induces a norm on X by the construction $\|\cdot\| = \sqrt{(\cdot,\cdot)}$. Hence, every inner product space is also a normed space, and therefore also a metric space.

Lemma (Schwarz Inequality)**.** *If $\|\cdot\|$ is the induces norm on an inner product space, then $|(x,y)| \leq \|x\|\|y\|$ for all elements x, y in the space.*

Lemma (Parallelogram Law)**.** *If $\|\cdot\|$ is the induces norm on an inner product space, then $\|x+y\|^2 + \|x-y\|^2 = 2\|x\|^2 + 2\|y\|^2$ for all elements x, y in the space.*

A normed linear space is called strictly convex if $\|x\| = \|y\| = 1$ and $\|x+y\| = 2$ imply that $x = y$ for elements on the space. All norms induced by an inner products are strictly convex in this sense.

Theorem. *Let X be a complex normed linear space with norm $\|\cdot\|_X$ satisfying the parallelogram law. Then there exists an inner product $(\cdot,\cdot)_X$ such that $\|\cdot\|_X = \sqrt{(\cdot,\cdot)}$.*

We have seen that norms constructed from inner products satisfy a number of nice properties. This theorem provides insight into the other direction. In order for a norm to come from an inner product, we only need it to satisfy the parallelogram law. The inner product that satisfies the above property that $\|\cdot\|_X = \sqrt{(\cdot,\cdot)}$ can be written explicitly as

$$(x,y)_X = \frac{1}{4}\Bigg(\Big(\|x+y\|_X^2 - \|x-y\|_X^2\Big) + i\Big(\|x+iy\|_X^2 - \|x-iy\|_X^2\Big)\Bigg)$$

Definition.

(i) A complete inner product space X is called a Hilbert space.
(ii) If $(x,y) = 0$ we say that the elements $x, y \in X$ are orthogonal, written $x \perp y$.
(iii) If $x \in X$ and $A \subseteq X$, we write $x \perp A$ if $x \perp y$ for every $y \in A$.
(iv) If $A, B \subseteq X$, we write $A \perp B$ if $x \perp y$ for every $x \in A$ and $y \in B$.
(v) If $A \subseteq X$, the orthogonal compliment is the set $A^\perp = \{x \in X : x \perp A\} \subseteq X$.

Projection Theorem

Let Y be a subspace of a normed linear space X. If Y is finite-dimensional, then for any $x_0 \in X$ there exists some element $y_0 \in Y$ such that

$$\|x_0 - y_0\| = \inf_{y \in Y} \|x_0 - y\|$$

If the norm $\|\cdot\|$ is strictly convex, then we've shown that the point y_0 is actually unique. Since all the norms induced by inner products are strictly convex, it follows that such a point will always be unique. However, we would like to be able to drop the condition that Y be finite-dimensional. The following lemma shows that the condition is not necessary in a Hilbert space.

Lemma. *Let H be a Hilbert space, and let $M \subseteq H$ be a closed and convex subset. Then for any $x_0 \in H$ there exists a unique element $y_0 \in M$ such that*

$$\|x_0 - y_0\| = \inf_{y \in M} \|x_0 - y\|$$

Theorem. *Let H be a Hilbert space, $M \subseteq H$ a closed subspace. Then for any $x_0 \in H$, there exist unique elements $y_0 \in M$ and $z_0 \in M^\perp$ such that $x_0 = y_0 + z_0$. This is also written $H = M \oplus M^\perp$.*

The proof of this theorem relies on the previous lemma. By choosing y_0 as the unique element, it can be shown that $z_0 = x_0 - y_0 \in M^\perp$. The following theorem confirms that all proper subspaces have at least one orthogonal element in the underlying Hilbert space.

Corollary. *Let $M \subset H$ be a proper closed subspace of a Hilbert space H. Then there exists a nonzero element $z_0 \in M$ such that $z_0 \perp H$ such that $z_0 \perp M$.*

Riesz Representation Theorem

Theorem. *Let H be a Hilbert space. For all $x^* \in H^*$, there exists a unique element $z \in H$ such that $\|x^*\| = \|z\|$ and $x^*(x) = (x, z)$ for all $x \in H$.*

Proof. Let $x^* \in H^*$ be some element. Let $N = \{x \in H : x^*(x) = 0\}$ be the kernel of x^*, a closed subspace of H. If $N = H$ then x^* is the zero function, and so we can take the element $z = 0 \in H$. Otherwise, suppose that $N \neq H$. By the previous corollary, let $z_0 \in N^\perp$ be a nonzero element of the orthogonal compliment of N, and let $\alpha = x^*(z_0) \neq 0$. Now choose some $x \in H$ and define $y = x - \frac{x^*(x)z_0}{\alpha} \in H$. Then

$$x^*(y) = x^*(x) - \frac{x^*(x)}{\alpha}x^*(z_0) = 0$$

So $y \in N$, and $y \perp z_0$. Additionally, $(x, z_0) = \frac{x^*(x)}{\alpha}(z_0, z_0) = \frac{x^*(x)}{\alpha}\|z_0\|^2$, so $(x, \frac{\overline{\alpha}z_0}{\|z_0\|^2} = x^*(x)$. This element $z = \frac{\overline{\alpha}z_0}{\|z_0\|^2}$ is independent of x, and so it works for all $x \in H$.

Now suppose there exist $z, z' \in H$ that both satisfy this property. Then $(x, z - z') = 0$ for any $x \in H$. Choose $x = z - z'$, then we have that $(z - z', z - z') = \|z - z'\|^2 = 0$, and so $z = z'$ and the element is unique.

Lastly, to prove that the norms are equal, observe that

$$\|x^*\| = \sup_{\|x\|=1} |x^*(x)| = \sup_{\|x\|=1} |(x, z)| \leq \sup_{\|x\|=1} \|x\|\|z\| = \|z\|$$

We also have that $\|z\|^2 = (z, z) = x^*(z) = |x^*(z)| \leq \|x^*\|\|z\|$, which implies that $\|z\| \leq \|x^*\|$. Combining these, we have that $\|z\| = \|x^*\|$. ■

Corollary. *The map $\sigma : H \to H^*$ given by $(\sigma x(y)) = (y, x)$ is an isometric embedding of H onto H^*, and $\sigma(\lambda x + \mu z) = \overline{\lambda}\sigma(x) + \overline{\mu}\sigma(z)$.*

Corollary. *Every Hilbert space H is reflexive and weakly complete. A subset $A \subseteq H$ is weakly compact if and only if it is bounded and weakly closed. This implies that if a sequence $\{x_n\}_{n=1}^\infty \subset H$ is bounded, then there exists a weakly convergent subsequence $x_{n_j} \rightharpoonup x \in H$.*

Proof. Let $x^*, y^* \in H^*$. By the previous corollary, there exist elements $x, y \in H$ such that $x^* = \sigma x$ and $y^* = \sigma y$. Define the inner product on H^* by $(x^*, y^*)_{H^*} = (x, y)_H$. This turns H^* into a Hilbert space itself. Define the map $\tau : H^* \to H^{**}$ by $(\tau\sigma y)(\sigma x) = (\sigma x, \sigma y)$. We now need to show that the following diagram commutes.

$$\begin{tikzcd}
H \arrow[r, "\sigma"] \arrow[dr, "\kappa"'] & H^* \arrow[d, "\tau"] \\
& H^{**}
\end{tikzcd}$$

We want to show that $\tau\sigma = \kappa$, the natural isometric embedding from H into H^{**}. To do this, let $y \in h$ and $x^* \in H^*$. Then there exists $x \in H$ such that $\sigma x = x^*$, and $\tau\sigma(y)(x^*) = (\tau\sigma y)(\sigma x) = (\sigma x, \sigma y) = (y, x) = (\sigma x)(y) = \kappa(y)(x^*)$, which implies that $\tau\sigma(y) = \kappa(y)$. ■

Lax-Milgram Theorem

Theorem. *Let $B(x, y)$ be a bilinear form on a Hilbert space H, a function on $H \times H$ that is linear in both components. Suppose that there exist some $C, c > 0$ such that $C\|x\|\|y\|$ (boundedness) and $|B(x, y)| \geq c\|x\|^2$ (connectivity) for all $x, y \in H$. Then for $x^* \in H^*$ there exists $z \in H$ such that $x^*(x) = B(x, z)$ for all $x \in H$.*

Proof. For fixed $y \in H$ let $C' = C\|y\|$, then we have that $|B(x, y)| \leq C\|y\|\|x\| = C'\|x\|$. Then the functional that takes x to $B(x, y)$ is an element of H^*, so by the Riesz Representation theorem there exists a unique $\hat{y} \in H$ such that $B(x, y) = (x, \hat{y})$. Now define $T : H \to H$ by $T(y) = \hat{y}$, such that $\|Ty\|^2 = (Ty, Ty) = (\hat{y}, \hat{y}) = |B(\hat{y}, y)| \leq C\|\hat{y}\|\|y\| = C\|Ty\|\|y\|$, which implies that $\|Ty\| \leq C\|y\|$. This shows that T is bounded, and so $T \in \mathcal{B}(H)$. We will show that this T is bijective, and use it to construct the element z that we wanted.

Now since $c\|x\|^2 \leq |B(x, x)| = |(x, Tx)| \leq \|x\|\|Tx\|$, which implies that $c\|x\| \leq \|Tx\|$ and so T is injective. Now let $u \in \overline{R_T}$ be in the range of T, such that $u_n \in R_T$ and $u_n \to u$. This further implies that u_n is Cauchy, and so if $x_n \in H$ such that $Tx_n = u_n$, it is easy to show that x_n is Cauchy. This can be used to imply that $Tx_n \to Tx = u$, and so R_T is closed. Similarly, if $R_T \neq H$ then there exists nonzero $z_0 \in H$ such that $z_0 \perp R_T$, however $0 = |(z_0, Tz_0)| = |B(z_0, z_0)| \geq c\|z_0\|^2$, a contradiction. Therefore, we also have that $R_T = H$.

Now let $x^* \in H^*$, and let $z_0 \in H$ such that $x^*(x) = (x, z_0)$ for all $x \in H$. Then let $z \in H$ such that $Tz = z_0$ since T is surjective. Then we finally have that $x^*(x) = (x, z_0) = (x, Tz) = B(x, z)$. Uniqueness follows easily. ■

Definition. Let M and N be subspaces of a Hilbert space H. We say that H is the direct sum of M and N, written $H = M \oplus N$, if for any $x \in H$ there exists unique elements $m \in M$ and $n \in N$ such that $x = m + n$.

When the dimension of a subspace M is one, we can always find some functional x^* that is not identically zero on M such that H is the direct sum of M and the kernel of x^*. We also have the decomposition $H = M \perp M^{\perp}$ for any closed subspace M.

Operators on Hilbert Spaces

Definition. Let $M \subseteq H$ be a closed subspace of a Hilbert space H. By the Projection theorem, for any $x \in H$ there exists unique elements $y \in M$ and $z \in M^\perp$ such that $x = y + z$. Define the operator $P_M : H \to M$ by $P_M x = y$, where y is as constructed in the Projection theorem. P_M is called the projection on M.

Definition. Let X be an inner product space, and let $T \in \mathcal{B}(X)$ be a bounded operator on X. The operator $T^* \in \mathcal{B}(X)$ such that $(Tx, y) = (x, T^*y)$ for all $x, y \in X$ is called the adjoint of T. If $T = T^*$ then we say that T is self-adjoint.

When dealing with bounded linear maps between Banach spaces, the definition of the adjoint was much different. If $T : X \to Y$ is a bounded linear map, then the adjoint $T^* : Y^* \to X^*$ was defined as the map such that $T^*y^*(x) = y^*(Tx)$. The connection is due to the fact that Hilbert spaces are isomorphic to their own dual space by the map σ. Therefore, if $T : X \to X$ is our initial map and $T^* : X^* \to X^*$ is its adjoint, there is a natural map $T^* : X \to X$ that mirrors the action of T^* on X^*.

Theorem. *All projection operators P on a Hilbert space H are self-adjoint and idempotent (meaning $P^2 = P$). If $P \neq 0$ then $\|P\| = 1$.*

Theorem. *Suppose that an operator P on a Hilbert space H is self-adjoint. Then P is the projection onto some closed subspace of H if $P^2 = P$.*

Theorem. *Two projections P_1 and P_2 are said to be orthogonal if $P_1P_2 = 0$ (the zero operator).*

(i) $P_1P_2 = 0$ *if and only if* $P_2P_1 = 0$.
(ii) $P_MP_N = 0$ *if and only if* $M \perp N$.
(iii) If P is a projection, then $I - P$ is also a projection.

Theorem. *Suppose that $A : H \to H$ is a self-adjoint operator. Then (Ax, x) is real valued for any $x \in H$ and $\|A\| = \sup_{\|x\|=1} |(Ax, x)|$.*

Proof. $(Ax, x) = (x, Ax) = \overline{(Ax, x)}$, and so (Ax, x) is real valued. Now suppose that A is not the zero operator, in which case the expression for $\|A\|$ is trivial. Let $\alpha = \sup_{\|x\|=1} |(Ax, x)|$, then $\alpha \leq \|Ax\|\|x\| \leq \|A\|\|x\|^2 = \|A\|$. Now choose nonzero $z \in H$ such that $Az \neq 0$. Let $\lambda = \sqrt{\frac{\|Az\|}{\|z\|}}$ and $u = \frac{Az}{\lambda}$. Then

$$\begin{aligned}
\|Az\| = \big(A(\lambda z), u\big) &= \frac{1}{4}\Big(\big(A(\lambda z + u), \lambda z + u\big) - \big(A(\lambda z - u), \lambda z - u\big)\Big) \\
&\leq \frac{\alpha}{4}\Big(\|\lambda z + u\|^2 + \|\lambda z - u\|^2\Big) = \frac{\alpha}{2}\Big(\|\lambda z\|^2 + \|u\|^2\Big) \\
&= \frac{\alpha}{2}\Big(\lambda^2\|z\|^2 + \frac{1}{\lambda^2}\|Az\|^2\Big) = \alpha\|z\|\|Az\|
\end{aligned}$$

which implies that $\|A\| \leq \alpha$. Together, this shows that $\|A\| = \alpha$. ■

Theorem. *The eigenvalues of a self-adjoint operator are real valued, and eigenvectors (elements of H) corresponding to distinct eigenvalues are orthogonal.*

In addition to the eigenvalues of a self-adjoint operator being strictly real (having no imaginary part), it is also true that all the elements in the spectrum of the operator are real valued. When H is finite dimensional, self-adjoint operators correspond to Hermitian matrices. Hermitian matrices are matrices whose transpose and conjugate coincide, so that $a_{ij} = \overline{a_{ji}}$.

Definition. The upper bound and lower bound for a self-adjoint operator are defined as $M = \sup_{\|x\|=1}(Ax, x)$ and $m = \sup_{\|x\|=1}(Ax, x)$.

As a corollary to the above theorems, it turns out that the norm of a self-adjoint operator A can be expressed in terms of its upper and lower bounds as $\|A\| = \max\{|M|, |m|\}$.

Theorem. *Suppose that the operator A is self-adjoint, and let $\sigma(A)$ be its spectrum. Then $\sigma(A) \subseteq [m, M]$, and $m, M \in \sigma(A)$.*

Definition. Let $A \in \mathcal{B}(H)$.

(i) A is semi-positive (definite) if $\mathrm{Re}(Ax, x) \geq 0$ for all $x \in H$.
(ii) A is positive (definite) if $\mathrm{Re}(Ax, x) > 0$ for all nonzero $x \in H$.
(iii) We write $A \geq B$ (or $A > B$) if $A - B$ is semi-positive or positive (definite).
(iv) If A is a self-adjoint, positive definite operator (written SPD), then $m > 0$.

Orthonormal Basis

Definition. A subset $K \subseteq X$ of an inner product space X is called orthonormal if

(i) $\|x\| = 1$ for all $x \in K$.
(ii) $x \perp y$ for all $x, y \in K$, $x \neq y$.

If $K^\perp = \{0\}$ then we say that K is complete. We say that K is an orthonormal basis for X if for all $x \in X$ we have that

$$x = \sum_{y \in K_x} (x, y)y \qquad K_x\{y \in K : (x, y) \neq 0\}$$

Theorem. *There exists an orthonormal basis for every Hilbert space.*

Theorem. *Let $K \subseteq H$ be an orthonormal subset of a Hilbert space H. The following are equivalent:*

(i) K is complete.
(ii) span$(K) = H$
(iii) K is an orthonormal basis for H.
(iv) (Parseval's Formula) For all $x \in H$,

$$\|x\|^2 = \sum_{y \in Kx} |(x, y)|^2$$

Distributions

Distributions attempt to generalize functions so that that more exotic things like the Dirac delta "function" are well behaved, and can be differentiated. This is achieved by looking at a collection of continuous linear functionals on a space of other functions, with some kind of differential operator defined on the functionals.

The underlying space is called the space of test functions, and consists of infinitely differentiable functions with compact support. Let Ω be an open subset of $\mathbb{R}^n$, then $C^\infty(\Omega)$ is the space of all functions on Ω which are infinitely differentiable on all of Ω. Recall that the support of a function $u \in C^\infty(\Omega)$ is defined as the set of all nonzero points, $\text{supp}(u) = \{x \in \Omega : f(x) \neq 0\}$. A function $u \in C^\infty(\Omega)$ is said to have compact support if there exists a compact subset $K \subseteq \Omega$ such that $\text{supp}(u) \subseteq K$. We then define the space $C_0^\infty(\Omega) \subseteq C^\infty(\Omega)$ as the space of all infinitely differentiable functions on Ω with compact support. We are now ready to introduce our space of test functions.

Definition. A test function is an element of the topological linear space $\mathcal{D}(\Omega)$, which is defined to be $C_0^\infty(\Omega)$ equipped with the following topology: $\{\varphi_k\}_{k=1}^\infty \subset \mathcal{D}(\Omega)$ is said to converge to an element $\varphi \in \mathcal{D}(\Omega)$ if

(i) there exists a compact subset $K \subseteq \Omega$ such that $\bigcup_{k=1}^\infty \operatorname{supp}(\varphi_k) \subseteq K$.

(ii) for any multi-index $\alpha = (\alpha_1, \ldots, \alpha_n)$ where $\alpha_j \in \mathbb{Z}^+$, the partial derivatives $D^\alpha \varphi_k \to D^\alpha \varphi$ uniformly on Ω. Recall that $D^\alpha \varphi(x) = \frac{\partial^{|\alpha|}\varphi(x)}{\partial^{\alpha_1}x_1 \partial^{\alpha_2}x_2 \cdots \partial^{\alpha_n}x_n}$.

Definition. A distribution is an element of the dual space $(\mathcal{D}(\Omega))^* = \mathcal{B}(\mathcal{D}(\Omega), \mathbb{R})$. The space is typically denoted $\mathcal{D}'(\Omega)$. The action of a distribution $T \in \mathcal{D}'(\Omega)$ on a function $\varphi \in \mathcal{D}(\Omega)$ is written $T(\varphi) = \langle T, \varphi \rangle$.

A notable example of a distribution is the Dirac delta function δ, which can be expressed as the distribution such that $\langle \delta, \varphi \rangle = \varphi(x_0)$ for some $x_0 \in \mathbb{R}^n$. Classically, the Dirac delta function was thought of as the function that could be integrated against to recover another function's value at a point, which is technically impossible using traditional integration.

When $\mathcal{D}'(\Omega)$ is equipped with the weak*-topology, it is a locally convex topological vector space. Because the underlying space $\mathcal{D}(\Omega)$ is not a Banach space, there is no norm associated with it. Consequently, the distributions are not bounded linear functionals, since there exists no concept of bound on the appropriate spaces. Instead, they simply must be continuous on the elements in $\mathcal{D}(\Omega)$ under the convergence criteria established on the test functions.

Theorem. *Let $\Lambda \in \mathcal{L}(\mathcal{D}(\Omega), \mathbb{R})$ be an arbitrary functional on $\mathcal{D}(\Omega)$. Then the following are equivalent:*

(i) $\Lambda \in \mathcal{D}'(\Omega)$

(ii) For all compact $K \subseteq \Omega$, there exists some nonnegative integer N and $C > 0$ such that for all $\varphi \in \mathcal{D}_K$ (infinitely differentiable functions with compact support contained in K),

$$|\langle \Lambda, \varphi \rangle| \leq C \|\varphi\|_{C^N(\Omega)}$$

The norm $\|\varphi\|_{C^N(\Omega)}$ on the space $C^N(\Omega)$ of N-differentiable functions on Ω is simply the sum of the supremums of each partial derivative of the original function φ. Since the function has compact support, these supremums are well defined.

Functions and Measures as Distributions

The definition of distributions allows a large class of functions and measures to be thought of as distributions acting on smooth functions. A measure μ, for instance, can be thought of as the distribution given by $\langle \mu, \varphi \rangle = \int \varphi d\mu$. Functions themselves can be viewed as distributions by simply integrating against them under some measure.

Definition. A function $f : \Omega \to \mathbb{R}$ is called locally Lebesgue-integrable if f is integrable on every compact subset contained in Ω, i.e. if $K \subseteq \Omega$ is compact then $\int_K |f(x)|dx = \int_\Omega \chi_K(x)|f(x)|dx < \infty$. We denote all such functions as $L^1_{\text{loc}}(\Omega)$.

Locally integrable functions are essentially the worst kinds of functions available to use to make distributions. They may not even be globally integrable, but their integrability on compact sets allows us to define the following kinds of operators.

Definition. Let $f \in L^1_{\text{loc}}(\Omega)$. Define the operator $T_f : \mathcal{D}(\Omega) \to \mathbb{R}$ by $\langle Tf, \varphi \rangle = \int_\Omega f(x)\varphi(x)dx$ for all $\varphi \in \mathcal{D}(\Omega)$. Such an integral is well defined, since φ has compact support.

Lemma. *The functional T_f constructed from $f \in L^1_{loc}(\Omega)$ is a distribution.*

Proof. Let $\text{supp}(\varphi) \subseteq K$, K compact. Then

$$|\langle T_f, \varphi \rangle| \leq \left(\int_K |f(x)|dx \right) \|\varphi\|_{C^0(\Omega)}$$

Choosing $N = 0$ and $C = \int_K |f(x)|dx$, by the previous theorem we have that $T_f \in \mathcal{D}'(\Omega)$. ∎

Definition. Let μ be a Borel measure or a positive measure, with $\mu(K) < \infty$ for any compact K. Define $\Lambda_\mu : \mathcal{D}(\Omega) \to \mathbb{R}$ by $\langle \Lambda_\mu, \varphi \rangle = \int_\Omega \varphi d\mu$.

Lemma. *The functional Λ_μ is a distribution.*

Derivatives of Distributions

Definition. Let α be a multi-index, and let $\Lambda \in \mathcal{D}'(\Omega)$ be a distribution. We define the α^{th}-order distributional derivative $D^\alpha \Lambda : \mathcal{D}(\Omega) \to \mathbb{R}$ by

$$\langle D^\alpha \Lambda, \varphi \rangle = (-1)^{|\alpha|} \langle \Lambda, D^\alpha \varphi \rangle$$

Lemma. *For any distribution Λ and any multi-index α, the distributional derivative $D^\alpha \Lambda$ is a distribution.*

Proof. Since $\Lambda \in \mathcal{D}'(\Omega)$, there exists a compact set $K \subseteq \Omega$, nonnegative integer N and $C > 0$ such that for any $\varphi \in \mathcal{D}(\Omega)$,

$$|\langle \Lambda, \varphi| \leq C \|\varphi\|_{C^N(\Omega)}$$

Then,

$$|\langle D^\alpha \Lambda, \varphi \rangle| = |\langle \Lambda, D^\alpha \varphi \rangle| \leq C \left\| D^\alpha \varphi \right\|_{C^N(\Omega)} \leq C \|\varphi\|_{C^{N+|\alpha|}(\Omega)}$$

Then by the theorem, choosing a new N as $N + |\alpha|$, we have that $D^\alpha \Lambda \in \mathcal{D}'(\Omega)$. ■

The distributional derivative behaves similar to the classical derivative, with properties like $D^\alpha D^\beta \Lambda = D^{\alpha+\beta} \Lambda = D^\beta D^\alpha \Lambda$. Interestingly, this allows us to differentiate distributions infinitely many times, despite the fact that they can be constructed from very poorly behaved functions.

Distributional Derivatives of Functions

In the previous week we have seen that locally integrable functions can be represented as distributions, and all distributions can be "differentiated" infinitely many times. The class of locally integrable functions is very large, and includes all differentiable, continuous, and Lebesgue integrable functions. Although these functions are rarely differentiable by classical standards, our theory of distributions allows us to define derivatives for them by first representing them as distributions. On the other hand, when the original function is sufficiently differentiable then we will see that these distributional derivatives coincide with its classical derivatives.

Definition. Let $f \in L^1_{\text{loc}}(\Omega)$ be a function defined on an open set $\Omega \subseteq \mathbb{R}^n$. We define its representation distribution Λ_f by $\langle \Lambda_f, \varphi \rangle = \int_\Omega f(x)\varphi(x)dx$, where dx represents Lebesgue measure on $\mathbb{R}^n$. Then the distribution $D^\alpha \Lambda_f$ is called the α^{th} distributional derivative of f, and can be expressed as follows:

$$\langle D^\alpha \Lambda_f, \varphi \rangle = (-1)^{|\alpha|} \langle \Lambda_f, D^\alpha \varphi \rangle = (-1)^{|\alpha|} \int_\Omega f(x) D^\alpha \varphi(x) dx$$

The distributional derivative $D^\alpha \Lambda_f$ is written $D^\alpha f$.

Lemma. *If $f \in C^{|\alpha|}(\Omega)$ then $D^\alpha \Lambda_f = \Lambda_{D^\alpha f}$.*

Consider the Heaviside function, given by $H(x) = 1$ for $x > 0$ and $H(x) = 0$ for $x \leq 0$. Although the function is discontinuous, we can compute its distributional derivative $H' = DH = D\Lambda_H$.

$$\langle H, \varphi \rangle = \int_{\mathbb{R}} H(x)\varphi(x)dx = \int_0^\infty \varphi(x)dx$$

$$\begin{aligned} \langle DH, \varphi \rangle &= -\langle H, \varphi' \rangle = -\int_{\mathbb{R}} H(x)\varphi'(x)dx = -\int_0^\infty \varphi'(x)dx \\ &= -\lim_{x \to \infty} \varphi(x) + \varphi(0) = \varphi(0) = \langle \delta_0, \varphi \rangle \end{aligned}$$

So the distributional derivative of the Heaviside function is δ_0, the delta function. In fact, the shift of the Heaviside function H_{x_0} has distributional derivative δ_{x_0}.

Multiplication by Smooth Functions

Definition. Let $\Lambda \in \mathcal{D}'(\Omega)$, $f \in C^\infty(\Omega)$. Define $f\Lambda : \mathcal{D}(\Omega) \to \mathbb{R}$ by

$$\langle f\Lambda, \varphi \rangle = \langle \Lambda, f\varphi \rangle$$

The function f is not necessarily one of our test functions, since it may not have compact support. The product $f\varphi$, however, will have compact support since φ is a test function. Therefore, the evaluation $\langle \Lambda, f\varphi \rangle$ is well defined for all test functions φ. When performing such a multiplication, it is customary to write $f\Lambda$ rather than Λf, since the latter is often confused for the evaluation $\Lambda(f) = \langle \Lambda, f \rangle$.

Sequences of Distributions

Recall that the weak* convergence $\Lambda_j \rightharpoonup \Lambda$ was so called pointwise convergence, where $\langle \Lambda_j, \varphi \rangle \to \langle \Lambda, \varphi \rangle$ for all φ in the underlying space. When speaking of distributions, we say that Λ_j converges to Λ in $\mathcal{D}'(\Omega)$ if its convergence is weak*.

Definition. Let $f_j, f \in L^1_{\text{loc}}(\Omega)$. If $\Lambda_{f_j} \to \Lambda_f$ in distribution, then we say that f_j converges to f in distribution.

Theorem. *Suppose that $\Lambda_j \in \mathcal{D}'(\Omega)$ and $\Lambda_j \to \Lambda$ in $\mathcal{D}'(\Omega)$. Then $\Lambda \in \mathcal{D}'(\Omega)$, and for any multi-index α we have that $D^\alpha \Lambda_j \to D^\alpha \Lambda$ in $\mathcal{D}'(\Omega)$.*

Theorem. *If $\Lambda_j \to \Lambda$ in $\mathcal{D}'(\Omega)$ and $g_j \to g$ in $C^\infty(\Omega)$, then $g_j \Lambda_j \to g\Lambda$ in $\mathcal{D}'(\Omega)$.*

Distributions with Compact Support

Definition. Let $V \subseteq \Omega \subseteq \mathbb{R}^n$, where V and Ω are open. We define $i_{V\Omega} : \mathcal{D}(V) \to \mathcal{D}(\Omega)$ as the natural (trivial) extension. Given a distribution Λ on $\mathcal{D}(V)$, the restriction distribution $\tilde{\Lambda}$ as $\langle \tilde{\Lambda}, \varphi \rangle = \langle \Lambda, i_{V\Omega}\varphi \rangle$. We also define the restriction operator $r_{V\Omega} : \mathcal{D}'(\Omega) \to \mathcal{D}'(V)$ by $r_{V\Omega}\Lambda = \tilde{\Lambda}$.

The restriction of a distribution is important in understanding where that distribution does not put any weight. If $\tilde{\Lambda}$ is the zero operator, then we say that Λ vanishes on V. This leads to the definition of the support of a distribution.

Definition. Let $\Lambda \in \mathcal{D}'(\Omega)$. We define $W = \bigcup_{r_{V\Omega}\Lambda \equiv 0} V$. Then the support of Λ is given by $\operatorname{supp}(\Lambda) = \Omega \backslash W$.

Theorem. *Suppose that Λ is a distribution with compact support. Then Λ has finite order, meaning there exists a constant $C > 0$ and some smallest natural number N such that*

$$|\langle \Lambda, \varphi \rangle| \leq C \|\varphi\|_{C^N(\Omega)}$$

Moreover, Λ extends uniquely to some $\hat{\Lambda} \in \mathcal{B}(C^\infty(\Omega), \mathbb{R})$.

Distributions as Derivatives of Continuous Functions

Possibly one of the most staggering properties of distributions is their following representation. The theorem essentially says that any distribution can be written as the sum of distributional derivatives of some distribution that arose from a continuous function. In this way, every distribution (a continuous functional on the space of test functions) is the "derivative" of some continuous function, acting as a functional under integration.

Theorem. *Let $\Lambda \in \mathcal{D}'(\Omega)$. Then there exist continuous functions g_α on Ω, one for each multi-index α, such that*

(i) each compact set $K \subseteq \Omega$ intersects the supports of only finitely many g_α.
(ii) $\Lambda = \sum_\alpha D^\alpha g_\alpha$, or similarly that

$$\langle \Lambda, \varphi \rangle = \sum_\alpha (-1)^{|\alpha|} \int_\Omega g_\alpha(x) D^\alpha \varphi(x) x$$

Since compact sets only intersect finite many supports of the functions g_α, the sum will only contain finitely many terms.

Convolution of Distributions with Test Functions

Given two functions f, g on $\mathbb{R}^n$, the convolution is defined to be the function generated by the other two from the process $f * g(x) = \int_{\mathbb{R}^n} f(y)g(x-y)dy$. We typically construct the convolution operation by first defining the operators $(\tau_x g)(y) = g(x-y)$ and $\breve{g}(y) = g(-y)$, and then writing $f * g(x) = \int_{\mathbb{R}^n} f(y)\tau_x \breve{g}(y)dy$. This in turn can be expressed as the $L^2(\mathbb{R}^n)$ inner product $(f, \tau_x \breve{g})$. We mimic this form with our definition of the convolution of a distribution with one of its test functions.

Definition. Let $\Lambda \in \mathcal{D}'(\Omega)$, $g \in \mathcal{D}(\Omega)$. We define the function $\Lambda * g : \Omega \to \mathbb{R}$ by

$$\Lambda * g(x) = \langle \Lambda, \tau_x \breve{g} \rangle$$

This real valued function is called the convolution of Λ and g.

It should be emphasized that, although we have defined various operations on distributions, the convolution of a distribution with a function on $\mathbb{R}^n$ is another function on $\mathbb{R}^n$. The following theorem shows that such convolutions are very nicely behaved, and allow the transference of differentiation to their components.

Theorem. *$\Lambda * g \in C^\infty(\Omega)$ and $D^\alpha(\Lambda * g) = D^\alpha \Lambda * g = \Lambda * D^\alpha g$.*

Convolution occurs naturally when distributions are used to solve PDEs. An abstract linear differential operator of order N is represented as $P(D) = \sum_{\alpha \leq N} c_\alpha(x) D^\alpha$. One example is the well known Laplacian, which can be expressed as $\Delta = P(D) = D^2_{x_1} + \ldots + D^2_{x_n} = \frac{\partial}{\partial x_1^2} + \ldots + \frac{\partial}{\partial x_n^2}$, which is a second order differential operator. In two dimensions, the operator given by $P(D) = \frac{\partial^4}{\partial x_1^4} + 2\frac{\partial^4}{\partial x_1^2 \partial x_2^2} + \frac{\partial^4}{\partial x_2^4}$ is typically written Δ^2. Problems in PDEs can often by expressed as finding a distribution u such that $P(D)u = v$ for some given v. It is a fact that the solution is given by $u = E * v$ (where $E * v$ represents the distribution induced by the function obtained by convolution) for some distribution E where $P(D)E = \delta_x$. The distribution E is called the fundamental solution for $P(D)$.

Tempered Distributions

In order to generalize the space of test functions, we can consider the collection of rapidly decreasing test functions. Rather than having compact support, these are functions on $\mathbb{R}^n$ that decay quickly as they approach infinity. This is defined precisely as functions whose derivatives converge to zero when multiplied by any power of x as the limit $\|x\| \to \infty$ is taken. Distributions which act on this space (a subset of our original distributions) are called tempered distributions. Tempered distributions have an advantage over regular distributions in that there are well defined Fourier transforms defined on them. They are useful in finding solutions of linear PDEs with constant coefficients of arbitrary order.

Weak Derivatives and Sobolev Spaces

Recall that the equality $D^\alpha \Lambda_f = \Lambda_{D^\alpha f}$ may not hold in general, but is guaranteed if $f \in C^{|\alpha|}(\Omega)$. Since the class of functions we are most interested is $L^1_{\text{loc}}(\Omega)$ however, it is unlikely that our function f will be many times continuously differentiable. Instead, we ask whether or not there exists some separate function $g \in L^1_{\text{loc}}(\Omega)$ such that $D^\alpha \Lambda_f = \Lambda_g$. This is equivalent to asking whether there exists g such that

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

Although it is not terribly likely that such a g exists for an arbitrary $f \in L^1_{\text{loc}}(\Omega)$, we are interested in functions with these kinds of relationships. If the above equation does hold, then we say that g is the α^{th} order weak derivative of f. A weak derivative can be thought of as a sort of Reisz representation of $(-1)^{|\alpha|}D^\alpha\Lambda_f$ under the $L^2(\Omega)$ inner product, although these functions may not actually be in $L^2(\Omega)$. If a function f has an α weak derivative g, then g is indeed unique as we would expect.

Theorem. *Let $u, v \in L^1_{loc}(\Omega)$. Then $v = D^\alpha u$ if and only if there exists a sequence $u_m \in C^\infty(\Omega)$ such that $u_m \to u$ in $L^1_{loc}(\Omega)$ and $D^\alpha u_m \to v$ in $L^1_{loc}(\Omega)$, where $D^\alpha u_m$ refers to the classical α derivative of the smooth function u_m.*

Theorem. *Suppose that $f \in C^1(\mathbb{R})$, $f' \in L^\infty(\mathbb{R})$, $u \in L^1_{loc}(\mathbb{R})$, and the weak derivative Du exists, then the weak derivative $D(f \circ u)$ also exists and*

$$D(f \circ u) = f'(u)Du$$

Review of L^p Spaces

Recall that for a measurable function $u : \Omega \to \mathbb{R}$, we define the p-norm as $\|u\|_p = \left(\int_\Omega |u(x)|^p dx\right)^{1/p}$ for $p \geq 1$ and $\|u\|_\infty = \text{ess sup}_\Omega\, |u|$. The space $L^p(\Omega)$ is the collection of measurable functions with finite p-norm. On each $L^p(\Omega)$ space, the process $\|\cdot\|_p$ acts as an actual norm, and each $(L^p(\Omega), \|\cdot\|_p)$ is a Banach space.

Remark.

(i) If $|\Omega| < \infty$ then for any $1 \leq p < q \leq \infty$, we have that $L^q(\Omega) \subseteq L^p(\Omega)$ and $\|u\|_p \leq |\Omega|^{\frac{1}{p}-\frac{1}{q}}\|u\|_q$ for all $u \in L^q(\Omega)$.

(ii) If $\frac{1}{p} + \frac{1}{p^*}$ then $\left|\int_\Omega u(x)v(x)dx\right| \leq \|u\|_p\|v\|_{p^*}$ for all $u \in L^p(\Omega)$ and $v \in L^{p^*}(\Omega)$.

(iii) If $\frac{1}{q} = \frac{\lambda}{p} + \frac{1-\lambda}{r}$ for $1 \leq p \leq q \leq r$, then for any $u \in L^r(\Omega)$

$$\|u\|_q \leq \|u\|_p^\lambda \|u\|_r^{1-\lambda}$$

(iv) If $\frac{1}{p_1} + \frac{1}{p_2} + \ldots + \frac{1}{p_n} = 1$ then for all $u_j \in L^{p_j}(\Omega)$

$$\left|\int_\Omega u_1(x)\ldots u_n(x)dx\right| \leq \|u_1\|_{p_1}\cdots\|u_n\|_{p_n}$$

(v) Let $\Phi_p(u) \left(\frac{1}{|\Omega|}\int_\Omega |u(x)|^p dx\right)^{1/p} = |\Omega|^{-1/p}\|u\|_p$ for $-\infty < p < \infty$. Then

$$\lim_{p\to\infty} \Phi_p(u) = \|u\|_\infty$$

$$\lim_{p\to-\infty} \Phi_p(u) = \text{ess inf}_\Omega\ |u|$$

$$\lim_{p\to\infty} \Phi_p(u) = e^{\frac{1}{|\Omega|}\int_\Omega \log|u(x)|dx}$$

(vi) The space $C_0^\infty(\Omega)$ is dense in $L^p(\Omega)$, and $L^p(\Omega)$ is separable for $1 \le p < \infty$. However, $L^\infty(\Omega)$ is not separable.

(vii) For $1 \le p < \infty$, $K \subseteq L^p(\Omega)$ is bounded. Then K is compact if and only if for all $\epsilon > 0$ there exists some $\delta > 0$ and compact $G \subseteq \Omega$ such that for all $u \in K$ and $|h| < \delta$, where $\tilde{u}$ denotes the trivial extention of u outside of Ω, we have the following properties:

$$\int_\Omega |\tilde{u}(x+h) - \tilde{u}(x)|^p dx < \epsilon^p \qquad \int_{\Omega\setminus G} |u(x)|^p dx < \epsilon^p$$

(viii) For $1 < p < \infty$, $L^p(\Omega)$ are strictly convex. $L^1(\Omega)$ and $L^\infty(\Omega)$ are not.

(ix) For $1 < p < \infty$, for all $T \in (L^p(\Omega))^*$ there exists $v \in L^{p^*}(\Omega)$ such that $T(u) = \int_\Omega v(x)u(x)dx$ for every $u \in L^p(\Omega)$. Furthermore, $\|T\| = \|v\|_{p^*}$. This gives an isometric isomorphism between $(L^p(\Omega))^*$ and $L^{p^*}(\Omega)$.

(x) $L^\infty(\Omega)$ is isometrically isomorphic to the dual space of $L^1(\Omega)$, but the inclusion $L^1(\Omega) \subset (L^\infty(\Omega))^*$ is strict.

(xi) For $1 < p < \infty$, $L^p(\Omega)$ is reflexive, but $L^1(\Omega)$ and $L^\infty(\Omega)$ are not.

Sobolev Spaces $W^{m,p}(\Omega)$

Definition. Let $m \in \mathbb{Z}_+$, $1 \le p \le \infty$. Let u be a measurable function, with weak derivatives $D^\alpha u$ for multi-indices α. Then we define

$$\|u\|_{m,p} = \left(\sum_{|\alpha|\le m} \|D^\alpha u\|_p^p\right)^{1/p}$$

$$\|u\|_{m,\infty} = \max_{|\alpha|\le m} \|D^\alpha u\|_\infty$$

Then we define the Sobolev spaces as the sets

$$\begin{aligned} W^{m,p}(\Omega) &= \{u \in L^p(\Omega) : \|u\|_{m,p} < \infty\} \\ &= \{u \in L^p(\Omega) : D^\alpha u \in L^p(\Omega),\ |\alpha| \le m\} \end{aligned}$$

equipped with the norms $\|\cdot\|_{m,p}$.

Theorem. *The spaces $W^{m,p}(\Omega)$ are Banach spaces for $m \in \mathbb{Z}_+$, $1 \leq p \leq \infty$.*

Now consider the following construction. The space $C^m(\Omega)$ is all functions with continuous m^{th} order derivatives. Let $H^{m,p}(\Omega) = \overline{C^m(\Omega)}$, where the completion is performed under the $\|\cdot\|_{m,p}$ norm. These are Banach spaces under the $\|\cdot\|_{m,p}$ norm, and $H^{m,p}(\Omega) \subseteq W^{m,p}(\Omega)$. In 1964 it was proved by Meyers and Serrin that $H^{m,p}(\Omega) = W^{m,p}(\Omega)$. The primary result of this equality is that $C^m(\Omega)$ is dense in $W^{m,p}(\Omega)$ for $m \in \mathbb{Z}_+$ and $1 \leq p \leq \infty$. The notation involving W and H are both used in literature.

Definition. $W_0^{m,p}(\Omega) = H_0^{m,p}(\Omega)$ is the space defined as $\overline{C_0^m(\Omega)}$, where completion is once again under the norm $\|\cdot\|_{m,p}$.

Exercise. Let Ω be an open, bounded subset of $\mathbb{R}^n$. Prove that $u(x) = \frac{1}{|x|^\alpha} \in W^{1,p}(\Omega)$ if and only if $0 \leq \alpha < \frac{n}{p} - 1$ and $u \notin W^{1,p}(\Omega)$ for $p \geq n$.

Global Approximations of Sobolev Functions

The following theorem establishes that every $W^{1,p}(\Omega)$ function can be approximated by a sequence of infinitely differentiable functions that converge under the $\|\cdot\|_{1,p}$ norm.

Theorem. *Suppose that $\Omega \subseteq \mathbb{R}^n$ is an open, bounded subset. Let $u \in W^{1,p}(\Omega)$ for $1 \leq p < \infty$. Then there exists a sequence $u_m \in C^\infty(\Omega)$ such that $u_m \to u$ in the $\|\cdot\|_{1,p}$ norm.*

The following theorem strengthens this result, showing that these approximations can be extended to the boundary of a sufficiently nice region Ω.

Theorem. *Let $\Omega \subseteq \mathbb{R}^n$ be open and bounded, and suppose that $\partial\Omega \in C^1$ (meaning that $\partial\Omega$ is locally a C^1 graph). Then for any $u \in W^{1,p}(\Omega)$ with $1 \leq p < \infty$, there exists a sequence $u_m \in C^\infty(\overline{\Omega})$ such that $u_m \to u$ in the $\|\cdot\|_{1,p}$ norm.*

Trace Theorem

As the last theorem shows, we are often concerned with the behavior of our Sobolev functions (or approximations) near the boundary of the set they are acting on. Since we are mainly concerned with L^p spaces over $\mathbb{R}^n$, we have a problem measuring the boundary, since $\mu^n(\partial\Omega) = 0$ (where μ^n is n-dimensional Lebesgue measure) for any $\Omega \subseteq \mathbb{R}^n$. Consequently, $u|_{\partial\Omega}$ has no meaning, since u is only defined up to sets of measure zero.

Theorem (Trace Theorem)**.** *Suppose $\Omega \subseteq \mathbb{R}^n$ is bounded and $\partial\Omega \in C^1$. Then there exists an operator $T \in \mathcal{L}(W^{1,p}(\Omega), L^p(\partial\Omega)$ (where $L^p(\partial\Omega)$ is understood to be equipped with Lebesgue measure on $\mathbb{R}^{n-1}$) for $1 \leq p < \infty$ such that*

(i) $Tu = u|_{\partial\Omega}$ *if* $u \in W^{1,p}(\Omega) \cap C(\overline{\Omega})$.

(ii) $\|Tu\|_{L^p(\partial\Omega)} \leq C(p, \Omega)\|u\|_{1,p}$ *for all* $u \in W^{1,p}(\Omega)$*, where* $C(p, \Omega)$ *is a constant depending only on* p *and* Ω.

The operator T is called the Trace operator.

The construction of the Trace operator involves looking at the boundary of our global, smooth approximations u_m that were described in the previous theorem. Although u is simply in $W^{1,p}(\Omega)$, and so its boundary values are not well defined, the approximations u_m are continuous.

Theorem. *Suppose that $\Omega \subseteq \mathbb{R}^n$ is bounded and $\partial\Omega \in C^1$. Let $u \in W^{1,p}(\Omega)$, then $u \in W_0^{1,p}(\Omega)$ if and only if $Tu \equiv 0$ on $\partial\Omega$.*

Sobolev Embedding

For our last topic, we are curious if the Sobolev spaces $W^{1,p}(\Omega)$ are contained in any more familiar spaces. Here we continue to write $\Omega \subseteq \mathbb{R}^n$. It turns out that they are, depending on p and the underlying dimension of the space n. Three different results hold if $1 \leq p < n$, $p = n$, or $p > n$.

Theorem (Gagliardo-Ninenberg-Sobolev)**.** *Suppose $1 \leq p < n$, $\partial\Omega \in C^1$. Then there exists $C = C(p, \Omega) > 0$ such that for all $u \in W^{1,p}(\Omega)$, $q \in [1, \hat{p}]$ where $\hat{p} = \frac{np}{n-p} > p$,*

$$\|u\|_q \leq C\|u\|_{1,p}$$

Theorem (Poincare Inequality)**.**

(i) For $1 \leq p < \infty$ *and any* $u \in W_0^{1,p}(\Omega)$,

$$\|u\|_p \leq C(n,\Omega)\|Du\|_p$$

(ii) For $u \in W^{1,p}(\Omega)$ *and* $u_G = \frac{1}{|G|}\int_G u(x)dx$ *for* $G \subseteq \Omega$ *measurable,*

$$\|u - u_G\|_p \leq C(n,\Omega)\|Du\|_p$$

Definition. Let X and Y be Banach spaces, $Y \subseteq X$. Y is said to be compactly embedded in X, written $Y \hookrightarrow X$, if

(i) $\|z\|_X \leq C\|z\|_Y$ for some constant C and all $z \in Y$.

(ii) each bounded sequence of Y is precompact in X.

An immediate example of a compact embedding is that $H^1(\Omega) = W^{1,2}(\Omega) \hookrightarrow L^2(\Omega)$. It is also true that $W^{1,\infty}(\Omega) = C^{0,1}(\overline{\Omega})$, where $C^{0,1}$ indicates the set of Lipschitz functions, and $W^{1,\infty}_{\text{loc}}(\Omega) = C^{0,1}_{\text{loc}}(\Omega)$. Additionally, all the elements of the space $C^{0,1}_{\text{loc}}(\Omega)$ are classically differentiable almost everywhere with respect to Lebesgue measure. Another commonly used space is $H^{-1}(\Omega) = W^{-1,2}(\Omega) = (H_0^1(\Omega))^*$, where the norm is the operator norm.