Notes on Functional Analysis*

Mengxi Wu
University of Southern California

Contents

*Author Email: mengxiwu@usc.edu. The first draft of these notes was written during Math 525b (Spring 2025), taught by Professor Calum Rickard.

Metric Spaces and Normed Vector Spaces

Metric Space

A metric on a set M is a function $\rho : M \times M \to [0, \infty)$ such that for all $x, y, z \in M$:

1. $\rho(x, y) = \rho(y, x)$ (Symmetry).
2. $\rho(x, y) \leq \rho(x, z) + \rho(z, y)$ (Triangle Inequality).
3. $\rho(x, y) = 0 \iff x = y$ (Identity of Indiscernibles).

Example

Let $M = \mathbb{R}^p$. For $x = (x_1, \ldots, x_p) \in \mathbb{R}^p$ and $y = (y_1, \ldots, y_p) \in \mathbb{R}^p$, we can define different metrics:

$$\rho_2(x, y) = \sqrt{(x_1 - y_1)^2 + \cdots + (x_p - y_p)^2},$$
$$\rho_1(x, y) = |x_1 - y_1| + \cdots + |x_p - y_p|, \quad \rho_\infty(x, y) = \max_{1 \leq i \leq p} |x_i - y_i|.$$

Convergence

A sequence $x_n \in M$ converges to $x \in M$ if $\lim_{n\to\infty} \rho(x_n, x) = 0$. We write $\lim_{n\to\infty} x_n = x$ or $x_n \to x$ as $n \to \infty$.

Note. For the standard metrics on $\mathbb{R}^p$ $(\rho_1, \rho_2, \rho_\infty)$, a sequence of vectors $x^n = (x_1^n, \ldots, x_p^n)$ converges to $x = (x_1, \ldots, x_p)$ if and only if each component converges individually, i.e., $x_i^n \to x_i$ as $n \to \infty$ for all $1 \leq i \leq p$.

Example

Let $M = C([0, 1])$, the space of continuous functions on $[0, 1]$. We can define:

$$\rho_1(x, y) = \int_0^1 |x(t) - y(t)| \, \mathrm{d}t, \quad \rho_\infty(x, y) = \sup_{0 \leq t \leq 1} |x(t) - y(t)|.$$

Let $K = \rho_\infty(x, y)$. Then:

$$\rho_1(x, y) = \int_0^1 |x(t) - y(t)| \, \mathrm{d}t \leq \int_0^1 K \, \mathrm{d}t = K = \rho_\infty(x, y).$$

Note that since $\rho_1(x, y) \leq \rho_\infty(x, y)$, ρ_∞-convergence (uniform convergence) implies ρ_1-convergence (convergence in the L^1 norm), but the converse is not necessarily true.

Cauchy Sequence

A sequence $x_n \in M$ is a Cauchy sequence if $\rho(x_n, x_m) \to 0$ as $n, m \to \infty$.

Proposition 1

If $x_n \to x$ in M, then (x_n) is a Cauchy sequence.

Proof. By the triangle inequality, $\rho(x_n, x_m) \le \rho(x_n, x) + \rho(x_m, x) \to 0$ as $n, m \to \infty$. $\square$

Completeness of Metric Space

A metric space (M, ρ) is complete if every Cauchy sequence inside it converges to a limit in M. That is, if $\rho(x_n, x_m) \to 0$ as $n, m \to \infty$, then there exists some $x \in M$ such that $x_n \to x$.

Example

a. $M = \mathbb{R}$ is complete, but the rational numbers $\mathbb{Q} \subset \mathbb{R}$ (equipped with the subspace metric) are not complete.

b. $M = L^1([0,1])$ with the metric $\rho_1(x, y) = \int_0^1 |x(t) - y(t)| \, \mathrm{d}t$ is complete.

c. The space of continuous functions $C([0,1])$ viewed as a subspace of $L^1([0,1])$ under the ρ_1 metric is not complete.

Remark. For example c, recall that every $x \in L^1$ can be approximated by a sequence of continuous functions $x_n \in C([0,1])$ such that $\int_0^1 |x_n(t) - x(t)| \, \mathrm{d}t \to 0$. Consider the specific discontinuous step function:

$$x(t) = \begin{cases} 0 & \text{if } 0 \le t < \frac{1}{2} \\ 1 & \text{if } \frac{1}{2} \le t \le 1 \end{cases} \in L^1([0,1]).$$

We can approximate it using the following continuous piecewise linear functions $x_n(t) \in C([0,1])$:

$$x_n(t) = \begin{cases} 0 & \text{if } 0 \le t < \frac{1}{2} \\ n(t - \frac{1}{2}) & \text{if } \frac{1}{2} \le t < \frac{1}{2} + \frac{1}{n} \\ 1 & \text{if } \frac{1}{2} + \frac{1}{n} \le t \le 1 \end{cases}.$$

Then we can compute the distance:

$$\int_0^1 |x_n(t) - x(t)| \, \mathrm{d}t = \frac{1}{2n} \to 0 \quad \text{as } n \to \infty.$$

This shows that the sequence x_n is Cauchy with respect to ρ_1, but its limit $x \notin C([0,1])$. Thus, the space is not complete.

Dense Set

A subset $B \subseteq M$ is dense in M if every element $x \in M$ can be approximated by elements from B. That is, for any $x \in M$, there exists a sequence $x_n \in B$ such that $x_n \to x$.

Theorem 1 (Completion Theorem)

For any metric space (M, ρ), there exists a complete metric space $(\tilde{M}, \tilde{\rho})$ such that M is isometric to a dense subset $B \subseteq \tilde{M}$. Specifically, there exists a distance-preserving bijection $h : M \to B$ satisfying:

$$\rho(x, y) = \tilde{\rho}(h(x), h(y)), \quad \forall x, y \in M.$$

This space $(\tilde{M}, \tilde{\rho})$ is uniquely defined up to isometry and is called the *completion* of (M, ρ).

Remark. *Isometric means distance-preserving.* When we say that a map $h : M \to B$ is an isometry, it means that the distance between any two points in M is exactly equal to the distance between their corresponding images in B under the mapping h. In Theorem 1 above, because distances are preserved perfectly, both h and its inverse $h^{-1} : B \to M$ are automatically continuous.

Example

1. $\mathbb{R}$ is the completion of $\mathbb{Q}$ under the standard absolute value metric.
2. $L^1([0, 1])$ is the completion of $C([0, 1])$ with respect to the integral metric $\rho_1(x, y) = \int_0^1 |x(t) - y(t)| \,\mathrm{d}t$.

Equivalence Relation on Cauchy Sequences

If (x_n) and (y_n) are two Cauchy sequences in M, we say they are equivalent, denoted $(x_n) \sim (y_n)$, if:

$$\lim_{n\to\infty} \rho(x_n, y_n) = 0.$$

Continuous Functions

Let $f : (M, \rho) \to (Y, \nu)$ be a function between two metric spaces.

1. f is continuous at a point $x_0 \in M$ if whenever $x_n \to x_0$ in M, we have $f(x_n) \to f(x_0)$ in Y. Equivalently, for any $\epsilon > 0$, there exists a $\delta > 0$ such that:

   $$\nu(f(x), f(x_0)) \leq \epsilon \quad \text{whenever} \quad \rho(x_0, x) \leq \delta.$$

   We write $\lim_{x\to x_0} f(x) = f(x_0)$.
2. f is continuous if it is continuous at every point $x \in M$.

Open and Closed Balls

The open ball $B_\delta(x)$ and closed ball $B_\delta[x]$ centered at x with radius $\delta > 0$ are defined as:

$$B_\delta(x) = \{y \in M \mid \rho(x, y) < \delta\}, \quad B_\delta[x] = \{y \in M \mid \rho(x, y) \leq \delta\}.$$

Open Set

A subset $A \subseteq M$ is open if every point in A has some wiggle room around it inside A. That is, $\forall x \in A$, there exists some $\delta > 0$ such that $B_\delta(x) \subseteq A$.

Closed Set

A subset $F \subseteq M$ is closed if its complement $F^c = M \setminus F$ is an open set.

Theorem 2

Let $\mathcal{T}$ be the collection of all open subsets of M. Then:

1. $M \in \mathcal{T}$ and $\emptyset \in \mathcal{T}$.
2. $\mathcal{T}$ is closed under arbitrary (finite or infinite) unions.
3. $\mathcal{T}$ is closed under finite intersections.

Remark. Notice that the empty set $\emptyset$ and the whole space M satisfy the conditions for being both open and closed at the same time! The collection $\mathcal{T}$ is formally called the *topology* of M.

Corollary 1

Let $\mathcal{F}$ be the set of all closed subsets of M. Then $\mathcal{F}$ contains M and $\emptyset$, and it is closed under arbitrary intersections and finite unions.

Neighborhood

A subset $A \subseteq M$ is a neighborhood of a point x if $x \in A$ and there is enough room to fit an open ball around x inside A, i.e., $B_\delta(x) \subseteq A$ for some $\delta > 0$.

Interior Point (Inner Point)

A point $x \in A$ is called an interior point of A if there exists an open ball centered at x completely contained in A: $B_\delta(x) \subseteq A$ for some $\delta > 0$.

Fact. The set of all interior points of A (denoted A° or $\text{int}(A)$) is always an open set.

Closure

Given a subset $A \subseteq M$, its closure $\overline{A}$ is defined as the smallest closed set that contains A.

Theorem 3

The closure can also be described sequentially:

$$\overline{A} = \{x \in M \mid \exists \{x_n\} \subseteq A \text{ such that } \lim_{n\to\infty} x_n = x\}.$$

Corollary 2

A subset $F \subseteq M$ is closed if and only if $\overline{F} = F$. Equivalently, F is closed if and only if whenever a sequence $\{x_n\} \subseteq F$ converges to some $x \in M$, the limit x must also belong to F.

Normed Vector Spaces (nvs)

A normed vector space E is a vector space over a field (usually $\mathbb{R}$ or $\mathbb{C}$) equipped with a norm function $\|\cdot\|_E \colon E \to [0,\infty)$ satisfying:

1. $\|\alpha x\|_E = |\alpha|\|x\|_E$ (Absolute Homogeneity).
2. $\|x+y\|_E \le \|x\|_E + \|y\|_E$ (Triangle Inequality).
3. $\|x\|_E = 0 \iff x = 0$ (Definiteness).

Note. Every norm automatically induces a metric on the space by defining the distance between two vectors as the norm of their difference: $\rho(x,y) = \|x-y\|_E$.

Triangle Inequality for Induced Metric

$$\|x-y\|_E = \|(x-z)+(z-y)\|_E \le \|x-z\|_E + \|z-y\|_E.$$

Example

1. On $\mathbb{R}^d$, the standard metrics we saw earlier come directly from these norms: for a vector $x = (x_1, \dots, x_d) \in \mathbb{R}^d$,

$$\|x\|_2 = \sqrt{x_1^2 + x_2^2 + \cdots + x_d^2}, \quad \|x\|_1 = |x_1| + \cdots + |x_d|, \quad \|x\|_\infty = \max_{1\le i\le d} |x_i|.$$

2. For the space of functions $E = C([0,1])$, the corresponding norms are:

$$\|x\|_1 = \int_0^1 |x(t)|\,\mathrm{d}t, \quad \|x\|_\infty = \sup_{0\le t\le 1} |x(t)|.$$

Completeness of Normed Vector Space

If a normed vector space E is complete under its induced metric, it is called a *Banach space*.

Example

1. $\mathbb{R}^d$ paired with any of the norms ($\|x\|_1, \|x\|_2, \|x\|_\infty$) is a Banach space.
2. $C([0,1])$ is a Banach space when equipped with the supremum norm $\|x\|_\infty$, but it is *not* a Banach space under the $\|x\|_1$ norm.
3. $L^1([0,1])$ with the norm $\|x\|_1 = \int_0^1 |x(t)|\,\mathrm{d}t$ is a complete Banach space, and $C([0,1])$ forms a dense subset within it.
4. $\mathbb{Q}^d$ is not a Banach space because the scalar field $\mathbb{Q}$ contains gaps.

Linear Operators

Let E and F be two normed vector spaces, and let $T : E \to F$ be a linear mapping. The following three conditions are completely equivalent:

a. T is continuous at the origin 0.

b. T is continuous everywhere on E.

c. T is bounded: There exists a constant $C > 0$ such that $\|T(x)\|_F \leq C\|x\|_E$ for all $x \in E$.

Notation. We denote by $\mathcal{L}(E, F)$ the set of all bounded linear operators from E to F.

Proof. The implication b $\implies$ a is trivial.

For a $\implies$ b: Let $x_n \to x$ in E. By linearity, this means $x_n - x \to 0$ in E. Since T is continuous at 0, we have $T(x_n - x) = T(x_n) - T(x) \to 0$ as $n \to \infty$, which proves continuity at x. Thus, a $\implies$ b.

For a $\implies$ c: Since T is continuous at 0, choosing $\epsilon = 1$, there must exist some radius $\delta > 0$ such that:

$$\|T(x)\|_F = \|T(x) - T(0)\|_F \leq 1 \quad \text{whenever} \quad \|x\|_E \leq \delta.$$

Now, pick any non-zero vector $y \in E$. We can rescale it to have norm exactly equal to δ by considering the vector $\frac{y}{\|y\|_E} \cdot \delta$. Since this scaled vector has norm δ, applying our property gives:

$$\left\| T\left(\frac{y}{\|y\|_E} \cdot \delta\right) \right\|_F \leq 1 \implies \frac{\delta}{\|y\|_E}\|T(y)\|_F \leq 1 \implies \|T(y)\|_F \leq \frac{1}{\delta}\|y\|_E.$$

Setting $C = \frac{1}{\delta}$ satisfies condition c.

For c $\implies$ a: If $x_n \to 0$ as $n \to \infty$, then by bounded inequality:

$$\|T(x_n)\|_F \leq C\|x_n\|_E \to 0,$$

which confirms T is continuous at 0. □

Operator Norm

Given a bounded linear operator $T \in \mathcal{L}(E, F)$, its operator norm is defined as:

$$\|T\| = \sup_{\|x\|_E \leq 1} \|T(x)\|_F = \sup_{\|x\|_E = 1} \|T(x)\|_F.$$

Note.

i. $\|T\| \leq C$ for any constant C that satisfies condition c.

ii. $\|T\|$ represents the absolute *smallest* valid constant $C > 0$ for which condition c holds.

iii. *Crucial Property.* $\|\mathbf{T}(\mathbf{x})\|_{\mathbf{F}} \leq \|\mathbf{T}\|\|\mathbf{x}\|_{\mathbf{E}}$.

Dual Space

Given a normed vector space E, its *dual space* (denoted E^*) is the space of all continuous linear functionals mapping from E into its underlying scalar field:

$$E^* = \mathcal{L}(E, \mathbb{R}) \quad (\text{or } E^* = \mathcal{L}(E, \mathbb{C})).$$

Note that because the target fields $\mathbb{R}$ or $\mathbb{C}$ are complete, E^* is always a complete Banach space, even if E itself is not! For any functional $\ell \in E^*$ and vector $x \in E$, we often write the evaluation using dual pairing notation:

$$\ell(x) = \langle \ell, x \rangle_{E^*,E} = \langle \ell, x \rangle.$$

Example

1. Let $E = \mathbb{R}^d$. Any linear functional $\ell \in E^* = \mathcal{L}(\mathbb{R}^d, \mathbb{R})$ can be represented as a standard dot product:

   $$\ell(x) = a \cdot x = \langle a, x \rangle = \sum_{i=1}^{d} a_i x_i,$$

   where $a = (a_1, \ldots, a_d) \in \mathbb{R}^d$ is a fixed vector uniquely identifying the functional.

2. Let E and F be normed vector spaces. We can construct their product space $E \times F$, which is also a normed vector space under any of the following natural norms:

   $$\|(x, y)\|_2 = \sqrt{\|x\|_E^2 + \|y\|_F^2},$$

   $$\|(x, y)\|_1 = \|x\|_E + \|y\|_F,$$

   $$\|(x, y)\|_\infty = \max\{\|x\|_E, \|y\|_F\}.$$

   Furthermore, if both component spaces E and F are complete Banach spaces, then the product space $E \times F$ is guaranteed to be a Banach space as well.

Main Theorems of Normed Vector Spaces

Hahn-Banach Theorem

Let E be a normed vector space and $M \subseteq E$ be a vector subspace. Let $f \in M^* = \mathcal{L}(M, \mathbb{R})$ be a bounded linear functional on M. Then there exists an extension $\ell \in E^* = \mathcal{L}(E, \mathbb{R})$ such that $\ell|_M = f$ and:

$$\|\ell\|_{E^*} = \|f\|_{M^*}.$$

Remark. The general proof relies on Zorn's Lemma. However, Zorn's Lemma is not required if the space E is separable (i.e., possesses a countable dense subset), in which case a sequential induction argument suffices.

Corollary 1 (Separation from Closed Subspaces)

Let E be a normed vector space and $F \subset E$ be a closed vector subspace such that $F \neq E$. Then there exists a nonzero functional $\ell \in E^*$ such that $\ell|_F = 0$.

Moreover, if $x_0 \notin F$, we can choose $\ell \in E^*$ such that $\ell(x_0) = 1$ and $\ell|_F = 0$.

Proof. Let $x_0 \notin F$. We construct the expanded subspace M spanned by x_0 and F:

$$M = \mathbb{R}x_0 + F = \{tx_0 + y : t \in \mathbb{R},\, y \in F\}.$$

First, we observe that this representation is unique: if $tx_0 + y = 0$ with $t \neq 0$, then $x_0 = -\frac{y}{t}$. Since F is a vector subspace, this implies $x_0 \in F$, which directly contradicts our assumption $x_0 \notin F$. Thus, $t = 0$, which immediately implies $y = 0$.

We can therefore safely define a linear functional $f : M \to \mathbb{R}$ by:

$$f(tx_0 + y) = t.$$

By construction, this functional satisfies $f(x_0) = 1$ and $f(y) = 0$ for all $y \in F$. To formally verify it is well-defined, suppose $t_1x_0 + y_1 = t_2x_0 + y_2$. Then $(t_1 - t_2)x_0 + (y_1 - y_2) = 0$, which by our unique representation requires $t_1 = t_2$ and $y_1 = y_2$. Hence $f(t_1x_0 + y_1) = f(t_2x_0 + y_2)$. Linearity follows directly:

$$f\big(\alpha_1(t_1x_0 + y_1) + \alpha_2(t_2x_0 + y_2)\big) = \alpha_1t_1 + \alpha_2t_2 = \alpha_1 f(t_1x_0 + y_1) + \alpha_2 f(t_2x_0 + y_2).$$

To establish that f is bounded, we define the distance from x_0 to the subspace F:

$$d = \inf_{z \in F} \|x_0 - z\|_E.$$

Because F is closed and $x_0 \notin F$, it follows that $d > 0$. (If $d = 0$, there would exist a sequence $\{z_n\} \subseteq F$ converging to x_0, forcing $x_0 \in F = \overline{F}$, a contradiction).

For any arbitrary element $tx_0 + y \in M$ with $t \neq 0$, we can factor out the scalar:

$$\|tx_0 + y\|_E = |t| \left\| x_0 - \left(-\frac{y}{t}\right) \right\|_E \geq |t|d,$$

since $-\frac{y}{t} \in F$. This inequality trivially holds for $t = 0$ as well. Consequently, if $\|tx_0 + y\|_E \leq 1$, we must have $|t| \leq \frac{1}{d}$. The norm of f is therefore bounded by:

$$\|f\|_{M^*} = \sup_{\|tx_0 + y\|_E \leq 1} |f(tx_0 + y)| = \sup_{\|tx_0 + y\|_E \leq 1} |t| \leq \frac{1}{d} < \infty.$$

By the Hahn-Banach Theorem, f extends to a global linear functional $\ell \in E^*$ preserving the norm. This extended functional satisfies $\ell(x_0) = f(x_0) = 1$ and $\ell(y) = f(y) = 0$ for all $y \in F$, completing the proof. $\square$

Corollary 1 provides a standard mechanism to prove that a vector subspace $D \subseteq E$ is dense ($\overline{D} = E$) via contradiction:

1. Assume the closure is proper: $\overline{D} \neq E$. Note that $\overline{D}$ remains a valid closed vector subspace.
2. By Corollary 1, there must exist a nonzero functional $\ell \in E^*$ vanishes entirely on $\overline{D}$ (i.e., $\ell|_{\overline{D}} = 0$).
3. If independent structural arguments demonstrate that $\ell|_D = 0 \implies \ell = 0$ on all of E, we obtain a contradiction, proving density.

By continuity, if a bounded functional vanishes on D, then for any $x_0 \in \overline{D}$ approached by a sequence $\{x_n\} \subseteq D$, we have $\ell(x_0) = \lim_{n\to\infty} \ell(x_n) = 0$.

Corollary 2 (Norm Realization and Alignment)

Let E be a normed vector space.

a. For each nonzero $x \in E$, there exists a normalized functional $\ell \in E^*$ ($\|\ell\|_{E^*} = 1$) such that:

$$\ell(x) = \langle \ell, x \rangle = \|\ell\|_{E^*}\|x\|_E.$$

We say that ℓ and x are *aligned*.

b. For any $x \in E$, the norm can be calculated via dual evaluation:

$$\|x\|_E = \max_{\|\ell\|_{E^*}\leq 1} |\ell(x)|.$$

Geometric Context. By definition, the bound $|\ell(x)| \leq \|\ell\|_{E^*}\|x\|_E$ always holds. Alignment represents the case where this inequality is sharply saturated.

- If $E = \mathbb{R}^d$ under the Euclidean norm $\|\cdot\|_2$, then $E^* \cong \mathbb{R}^d$ under $\|\cdot\|_2$. Alignment means $\langle a, x \rangle = \|a\|_2\|x\|_2$, which occurs if and only if a and x are linearly dependent and point in the same direction ($\cos\alpha = 1$).
- If $E = \mathbb{R}^d$ under the ℓ^1-norm $\|\cdot\|_1$, its dual is governed by the ℓ^∞-norm $\|\cdot\|_\infty$. Alignment requires $\langle a, x \rangle = \|a\|_\infty\|x\|_1$, meaning $a_i = \max_j |a_j| \cdot \operatorname{sgn}(x_i)$ for all indices where $x_i \neq 0$.

Proof. We assume without loss of generality that the scalar field is $\mathbb{R}$.

a. Let $x_0 \in E$ with $x_0 \neq 0$. Define the one-dimensional subspace $M = \mathbb{R}x_0$. Construct the linear functional $f : M \to \mathbb{R}$ by $f(tx_0) = t\|x_0\|_E$. Its operator norm on M is:

$$\|f\|_{M^*} = \sup_{||tx_0||_E\leq 1} |f(tx_0)| = \sup_{|t|\|x_0\|_E\leq 1} |t|\|x_0\|_E = 1.$$

By construction, $f(x_0) = \|x_0\|_E = \|f\|_{M^*}\|x_0\|_E$. Applying the Hahn-Banach Theorem yields an extension $\ell \in E^*$ matching the norm $\|\ell\|_{E^*} = \|f\|_{M^*} = 1$, which satisfies $\ell(x_0) = f(x_0) = \|x_0\|_E = \|\ell\|_{E^*}\|x_0\|_E$.

b. For any functional satisfying $\|\ell\|_{E^*} \leq 1$, we have $|\ell(x_0)| \leq \|\ell\|_{E^*}\|x_0\|_E \leq \|x_0\|_E$, hence $\sup_{\|\ell\|_{E^*}\leq 1} |\ell(x_0)| \leq \|x_0\|_E$. Part a guarantees the existence of a specific functional achieving exactly $\|x_0\|_E$, turning the supremum into a maximum. $\square$

Norm of the Adjoint Operator

Let E and F be normed vector spaces, and let $T \in \mathcal{L}(E, F)$ be a bounded linear operator. The *adjoint operator* $T' : F^* \to E^*$ is defined by mapping each linear functional $\ell \in F^*$ to the functional $T'\ell \in E^*$ via:

$$(T'\ell)(x) := \ell(T(x)) \quad \forall x \in E.$$

Proposition 1

The adjoint operator satisfies $T' \in \mathcal{L}(F^*, E^*)$, and its operator norm is identical to the original operator:

$$\|T'\| = \|T\|.$$

Proof. We expand the definition of the operator norm using nested suprema over the respective unit balls:

$$\begin{aligned}\|T'\| &= \sup_{\|\ell\|_{F^*} \le 1} \|T'\ell\|_{E^*} = \sup_{\|\ell\|_{F^*} \le 1} \sup_{\|x\|_E \le 1} |(T'\ell)(x)| \\ &= \sup_{\|\ell\|_{F^*} \le 1} \sup_{\|x\|_E \le 1} |\ell(T(x))| = \sup_{\|x\|_E \le 1} \sup_{\|\ell\|_{F^*} \le 1} |\ell(T(x))|.\end{aligned}$$

Applying Corollary 2 b directly to the inner supremum over the element $T(x) \in F$:

$$\sup_{\|\ell\|_{F^*} \le 1} |\ell(T(x))| = \|T(x)\|_F.$$

Substituting this back into the norm expression completes the proof:

$$\|T'\| = \sup_{\|x\|_E \le 1} \|T(x)\|_F = \|T\|.$$

$\square$

Canonical Isometric Embedding into the Bidual

Let E be a normed vector space and $E^{**} = (E^*)^*$ denote its bidual space. Every element $x_0 \in E$ induces an evaluation functional $J_{x_0} \in E^{**}$ defined via:

$$J_{x_0}(\ell) = \ell(x_0) \quad \forall \ell \in E^*.$$

Boundedness follows immediately from standard duality: $|J_{x_0}(\ell)| = |\ell(x_0)| \le \|\ell\|_{E^*} \|x_0\|_E$.

Proposition 2

The canonical evaluation map $J : E \to E^{**}$ defined by $x_0 \mapsto J_{x_0}$ is a linear isometric embedding. Thus, E can be viewed as an isometric subspace of E^{**} (written $E \subseteq E^{**}$).

Proof. Linearity is inherited from the vector space operations on E^*. Utilizing Corollary 2b once more to evaluate the norm of the functional J_{x_0} on E^*:

$$\|J_{x_0}\|_{E^{**}} = \sup_{\|\ell\|_{E^*} \le 1} |J_{x_0}(\ell)| = \sup_{\|\ell\|_{E^*} \le 1} |\ell(x_0)| = \|x_0\|_E.$$

$\square$

Baire's Category Theorem (Closed Set Formulation)

Let (M, ρ) be a complete metric space. If M is expressed as the countable union of closed subsets $M = \bigcup_{n=1}^{\infty} M_n$, then there exists an index $n_0 \geq 1$ such that its interior is nonempty ($\mathring{M}_{n_0} \neq \emptyset$). That is, M_{n_0} contains an open ball $B_{r_0}(x_0)$.

Proof. We proceed by contradiction. Assume that $\mathring{M}_n = \emptyset$ for all $n \geq 1$. This implies that the complements:

$$O_n = M \setminus M_n$$

are open and dense in M. Consequently, every open ball in M must intersect each O_n. We aim to show that the intersection $G = \bigcap_{n=1}^{\infty} O_n$ is dense. However, by De Morgan's laws:

$$G = \bigcap_{n=1}^{\infty} O_n = \bigcap_{n=1}^{\infty} M_n^c = \left(\bigcup_{n=1}^{\infty} M_n \right)^c = M^c = \emptyset,$$

which yields a contradiction since a dense subset of a nonempty complete space cannot be empty.

To prove G is dense, it suffices to show that an arbitrary open ball $B_{r_0}(x_0) \subseteq M$ contains at least one point of G. Since O_1 is open and dense, the intersection $B_{r_0}(x_0) \cap O_1$ is open and nonempty. We can therefore choose a nested closed ball $\overline{B}_{r_1}(x_1)$ inside it:

$$\overline{B}_{r_1}(x_1) \subseteq B_{r_0}(x_0) \cap O_1 \quad \text{with } r_1 \leq \frac{r_0}{2}.$$

Next, because O_2 is dense, the open set $B_{r_1}(x_1) \cap O_2$ is nonempty. We select a smaller nested closed ball:

$$\overline{B}_{r_2}(x_2) \subseteq B_{r_1}(x_1) \cap O_2 \quad \text{with } r_2 \leq \frac{r_1}{2} \leq \frac{r_0}{2^2}.$$

Proceeding inductively, given the open ball $B_{r_n}(x_n)$, density of O_{n+1} allows us to find:

$$\overline{B}_{r_{n+1}}(x_{n+1}) \subseteq B_{r_n}(x_n) \cap O_{n+1} \quad \text{with } r_{n+1} \leq \frac{r_0}{2^{n+1}}.$$

The constructed sequence of centers $\{x_n\}_{n=1}^{\infty}$ is Cauchy. Indeed, for any indices $j, j' \geq n$, both points belong to the ball $B_{r_n}(x_n)$, so:

$$\rho(x_j, x_{j'}) \leq \rho(x_j, x_n) + \rho(x_{j'}, x_n) < 2r_n \leq \frac{2r_0}{2^n} \longrightarrow 0 \quad \text{as } n \to \infty.$$

Since M is complete, the limit $x = \lim_{n\to\infty} x_n$ exists. Because the balls are nested, for any fixed index n, the tail of the sequence lies completely inside the closed ball $\overline{B}_{r_n}(x_n)$, forcing the limit to remain inside it:

$$x \in \overline{B}_{r_n}(x_n) \subseteq B_{r_{n-1}}(x_{n-1}) \cap O_n \subseteq O_n.$$

Since this holds for every index, $x \in \bigcap_{m=1}^{\infty} O_m = G$. Thus $B_{r_0}(x_0) \cap G \neq \emptyset$, proving density and completing the contradiction. $\square$

Banach-Steinhaus Theorem (Uniform Boundedness Principle)

Let E be a Banach space and F be a normed vector space. Let $\{T_i\}_{i\in I} \subset \mathcal{L}(E, F)$ be a family of bounded linear operators. Suppose that the family is pointwise uniformly bounded, meaning:

$$\sup_{i\in I} \|T_i(x)\|_F < \infty \quad \forall x \in E.$$

Then the operators are uniformly bounded in operator norm:

$$\sup_{i\in I} \|T_i\| < \infty.$$

Proof. For each integer $n \geq 1$, we define the sets:

$$M_n = \left\{x \in E : \sup_{i\in I} \|T_i(x)\|_F \leq n\right\} = \bigcap_{i\in I} \{x \in E : \|T_i(x)\|_F \leq n\}.$$

Because each operator T_i is continuous, the set $\{x \in E : \|T_i(x)\|_F \leq n\}$ is the inverse image of a closed set, hence closed. The intersection of any family of closed sets remains closed, so each M_n is closed. The pointwise uniform boundedness assumption guarantees that every element $x \in E$ belongs to M_n for a sufficiently large n, yielding the countable decomposition $E = \bigcup_{n=1}^{\infty} M_n$.

Since E is a complete Banach space, Baire's Category Theorem dictates that there exists an index $n_0 \geq 1$ such that $\mathring{M}_{n_0} \neq \emptyset$. Thus, there exists a closed ball centered at some $x_0 \in E$ completely contained within this set:

$$\overline{B}_{r_0}(x_0) \subseteq M_{n_0} \implies \sup_{i\in I} \|T_i(x)\|_F \leq n_0 \quad \forall x \in \overline{B}_{r_0}(x_0).$$

We now estimate the operator norm by evaluating perturbations around x_0. Let $y \in E$ be an arbitrary vector satisfying $\|y\|_E \leq r_0$. The shifted vector $x_0 + y$ belongs to $\overline{B}_{r_0}(x_0)$. By linearity, we can isolate y:

$$\|T_i(y)\|_F = \|T_i(y + x_0 - x_0)\|_F \leq \|T_i(x_0 + y)\|_F + \|T_i(x_0)\|_F.$$

Since both x_0 and $x_0 + y$ reside within M_{n_0}, their images under any T_i are bounded by n_0:

$$\|T_i(y)\|_F \leq n_0 + n_0 = 2n_0 \quad \forall i \in I.$$

Taking the supremum over the ball of radius r_0 allows us to calculate the operator norm bound:

$$\|T_i\| = \frac{1}{r_0} \sup_{\|y\|_E \leq r_0} \|T_i(y)\|_F \leq \frac{2n_0}{r_0} < \infty.$$

As this bound is entirely independent of the index i, we conclude $\sup_{i\in I} \|T_i\| \leq \frac{2n_0}{r_0}$. $\square$

Finite-Dimensional Example. Let $T_i \in \mathcal{L}(\mathbb{R}^d, \mathbb{R}^d)$ be represented by matrices $(a^i_{k\ell})$ for $1 \leq k, \ell \leq d$. Pointwise boundedness $\sup_i \|T_i(x)\| < \infty$ for all $x \in \mathbb{R}^d$ implies uniform boundedness of all matrix entries: $\sup_i |a^i_{k\ell}| < \infty$. This can be checked by testing the standard basis vectors e_ℓ:

$$\sup_i |a^i_{k\ell}| = \sup_i |e_k \cdot (T_i(e_\ell))| \leq \sup_i \|e_k\|_2 \|T_i(e_\ell)\|_2 = \sup_i \|T_i(e_\ell)\|_2 < \infty.$$

The Open Mapping Theorem

Theorem 3

Let E and F be Banach spaces and let $T \in \mathcal{L}(E, F)$ be a surjective (onto) bounded linear operator. Then there exists a scaling constant $c > 0$ such that the image of the open unit ball contains an open ball centered at the origin in F:

$$B_c^F \subseteq T(B_1^E).$$

Proof. By surjectivity, the images of expanding balls cover the entire target space: $F = \bigcup_{n=1}^{\infty} T(B_n^E)$. To invoke Baire's Category Theorem, we look at their closures: $F = \bigcup_{n=1}^{\infty} \overline{T(B_n^E)}$. Complete continuity of F guarantees that there exists an index n_0 containing an open neighborhood:

$$B_{r_0}^F(x_0) \subseteq \overline{T(B_{n_0}^E)}.$$

By linearity, the image is symmetric about the origin ($y \in T(B_{n_0}^E) \iff -y \in T(B_{n_0}^E)$), which implies the closure is symmetric as well, so $B_{r_0}^F(-x_0) \subseteq \overline{T(B_{n_0}^E)}$. Let $y \in B_{r_0}^F$ be an arbitrary vector centered at the origin. We can decompose it convexly:

$$y = \frac{1}{2}(x_0 + y) + \frac{1}{2}(-x_0 + y).$$

Since $x_0 + y \in B_{r_0}^F(x_0)$ and $-x_0 + y \in B_{r_0}^F(-x_0)$, both belong to the convex set $\overline{T(B_{n_0}^E)}$. Summing them shows that the centered ball is absorbed: $B_{r_0}^F \subseteq \overline{T(B_{n_0}^E)}$. By linear scaling, dividing by n_0 yields:

$$B_\delta^F \subseteq \overline{T(B_1^E)} \quad \text{where } \delta = \frac{r_0}{n_0}.$$

We now show that this closure can be swallowed by a slightly larger open ball, specifically $\overline{T(B_1^E)} \subseteq T(B_2^E)$. Let $z \in \overline{T(B_1^E)}$. By scaling the absorption property, $B_{\delta/2}^F \subseteq \overline{T(B_{1/2}^E)}$. This allows us to approximate z closely: there exists an element $x_1 \in B_1^E$ such that:

$$z - T(x_1) \in B_{\delta/2}^F \subseteq \overline{T(B_{1/2}^E)}.$$

Iterating this approximation property at scale $1/2$, there must exist an $x_2 \in B_{1/2}^E$ satisfying:

$$z - T(x_1) - T(x_2) \in B_{\delta/4}^F \subseteq \overline{T(B_{1/4}^E)}.$$

Inductively, for step n, we can find a correction vector $x_n \in B_{1/2^{n-1}}^E$ such that:

$$z - T(x_1 + \cdots + x_n) \in B_{\delta/2^n}^F \subseteq \overline{T(B_{1/2^n}^E)},$$

which means $\|z - T(\sum_{k=1}^n x_k)\|_F \le \frac{\delta}{2^n} \longrightarrow 0$ as $n \to \infty$. The series $\sum_{k=1}^{\infty} x_k$ converges absolutely within the Banach space E because:

$$\sum_{k=1}^{\infty} \|x_k\|_E < 1 + \sum_{j=1}^{\infty} \frac{1}{2^j} = 2.$$

Completeness implies the limit $x = \sum_{k=1}^{\infty} x_k$ exists and satisfies $\|x\|_E < 2$. By continuity of T:

$$T(x) = \lim_{n\to\infty} T\left(\sum_{k=1}^{n} x_k\right) = z.$$

Since $\|x\|_E < 2$, it follows that $z \in T(B_2^E)$, which proves $\overline{T(B_1^E)} \subseteq T(B_2^E)$. Combining our results:

$$B_\delta^F \subseteq \overline{T(B_1^E)} \subseteq T(B_2^E) \implies B_{\delta/2}^F \subseteq T(B_1^E).$$

Setting $c = \delta/2$ completes the proof. □

Corollary 3 (Openness of Operator Images)

Let E and F be Banach spaces and let $T \in \mathcal{L}(E, F)$ be surjective. If $U \subseteq E$ is an open set, then its image $T(U) \subseteq F$ is open.

Proof. Let $y_0 \in T(U)$. By definition, there exists an $x_0 \in U$ such that $T(x_0) = y_0$. Since U is open, we can find a radius $\delta > 0$ such that $B_\delta^E(x_0) = x_0 + B_\delta^E \subseteq U$. By Theorem 3, there exists a constant $c > 0$ such that $B_c^F \subseteq T(B_1^E)$, which scales to $B_{c\delta}^F \subseteq T(B_\delta^E)$. Using linearity to translate the sets:

$$y_0 + B_{c\delta}^F \subseteq T(x_0) + T(B_\delta^E) = T(x_0 + B_\delta^E) \subseteq T(U).$$

Thus, $T(U)$ contains an open ball around each of its points, meaning it is open. □

Inverse Mapping Theorem

Let E and F be Banach spaces. If a bounded linear operator $T \in \mathcal{L}(E, F)$ is bijective, then its set-theoretic inverse is automatically bounded: $T^{-1} \in \mathcal{L}(F, E)$.

Proof. To establish continuity of T^{-1}, it suffices to show that the preimage of any open set $U \subseteq E$ under T^{-1} is open in F. Notice that:

$$(T^{-1})^{-1}(U) = T(U).$$

By the Open Mapping Theorem (Corollary 3), since T is surjective, the image $T(U)$ is guaranteed to be open. Thus T^{-1} is continuous, which is equivalent to being bounded. □

Closed Graph Theorem

Let E and F be Banach spaces and let $T : E \to F$ be a linear operator. If the graph of the operator:

$$\Gamma = \{(x, T(x)) \in E \times F : x \in E\}$$

is a closed subspace of the product space $E \times F$, then T is bounded ($T \in \mathcal{L}(E, F)$).

Proof. The product space $E \times F$ equipped with the norm $\|(x, y)\|_{E\times F} = \sqrt{\|x\|_E^2 + \|y\|_F^2}$ is a Banach space. Because Γ is assumed to be a closed subspace, it is a complete Banach space in its own right under this induced norm. We consider the two standard canonical coordinate projections restricted to Γ:

$$\pi_1 : \Gamma \to E, \quad \pi_1(x, T(x)) = x,$$

$$\pi_2 : \Gamma \to F, \quad \pi_2(x, T(x)) = T(x).$$

The projection π_1 is bounded because $\|\pi_1(x, T(x))\|_E = \|x\|_E \le \|(x, T(x))\|_{E\times F}$. Furthermore, π_1 is clearly a bijection between Γ and E. Applying the Inverse Mapping Theorem, the inverse operator $\pi_1^{-1} : E \to \Gamma$ is automatically bounded.

We can now express the original operator as a composition of these two projections:

$$T = \pi_2 \circ \pi_1^{-1}.$$

Since π_2 is obviously bounded ($\|T(x)\|_F \le \|(x, T(x))\|_{E\times F}$), T is the composition of two bounded linear operators, making it bounded. □

Operational Summary. The Closed Graph Theorem simplifies proofs of continuity by relaxing the convergent sequence condition:

- *Standard Proof Method.* Assume $x_n \to x$ in E. One must simultaneously prove that the limit $\lim_{n\to\infty} T(x_n) = y$ exists, and that $y = T(x)$.
- *Closed Graph Method.* Assume $x_n \to x$ in E and assume that $T(x_n) \to y$ in F. One only needs to show the structural identity $y = T(x)$.

Hilbert Spaces

Inner Product

Let H be a complex vector space. An inner product on H is a $\mathbb{C}$-valued function denoted by $\langle x, y\rangle$, for $x, y \in H$, such that:

i. $\langle x, x\rangle \ge 0 \ \forall x \in H$; $\langle x, x\rangle = 0$ if and only if $x = 0$

ii. $\forall x, y, z \in H$, and $\alpha, \beta \in \mathbb{C}$:

$$\langle \alpha x + \beta y, z\rangle = \alpha\langle x, z\rangle + \beta\langle y, z\rangle,$$

$$\langle z, \alpha x + \beta y\rangle = \overline{\alpha}\langle z, x\rangle + \overline{\beta}\langle z, y\rangle,$$

$$\langle x, y\rangle = \overline{\langle y, x\rangle}.$$

Examples

1. $H = \mathbb{C}^d$, $x = (x_1, \dots, x_d)$, $y = (y_1, \dots, y_d) \in \mathbb{C}^d$:

$$\langle x, y\rangle = \sum_{j=1}^{d} x_j \overline{y}_j.$$

Then, $\langle x, x\rangle = \sum_{j=1}^{d} |x_j|^2$.

2. ℓ^2 is the space of all sequences $(x_1, x_2, \dots)$ such that

$$\sum_{n=1}^{\infty} |x_n|^2 < \infty,$$

with the inner product $\langle x, y \rangle = \sum_{n=1}^{\infty} x_n \overline{y}_n$.

3. $\mathcal{C} = C([a, b])$ is the space of continuous complex-valued functions on $[a, b]$ with

$$\langle f, g \rangle = \int_a^b f(t)\overline{g(t)}\, dt.$$

Then, $\langle f, f \rangle = \int_a^b |f(t)|^2 dt$.

4. $L^2(X, \mu)$ is the space of complex-valued square-integrable functions (equivalence classes) on a measure space $(X, \mathcal{X}, \mu)$ with

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

In fact, Examples 1 and 2 are just specific cases of Example 4 using the counting measure.

Inner Product Space (Pre-Hilbert Space)

A complex vector space equipped with an inner product is called an inner product space or pre-Hilbert space.

Orthogonality

Let H be a pre-Hilbert space. We say $x \perp y$ in H if $\langle x, y \rangle = 0$. The induced norm is defined as:

$$\|x\|_H := \sqrt{\langle x, x \rangle}, \quad x \in H.$$

Pythagorean Theorem

If $x \perp y$, then

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

Proof. Let $\langle x, y \rangle = \langle y, x \rangle = 0$. Then,

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

Orthonormal System

A subset $\{e_j : j = 1, \ldots, N\}$ is an orthonormal system (ONS) if

$$\langle e_k, e_\ell \rangle = \begin{cases} 1 & \text{if } k = \ell \\ 0 & \text{if } k \neq \ell \end{cases}.$$

Note that $\|e_k\|_H^2 = \langle e_k, e_k \rangle = 1$.

Theorem 2

Let $\{e_j : j = 1, \ldots, N\}$ be an ONS. Then for any $x \in H$:

i. $\sum_{j=1}^N \langle x, e_j \rangle e_j$ and $x - \sum_{j=1}^N \langle x, e_j \rangle e_j$ are orthogonal.

ii. $\|x\|_H^2 = \|x - \sum_{j=1}^N \langle x, e_j \rangle e_j\|_H^2 + \| \sum_{j=1}^N \langle x, e_j \rangle e_j\|_H^2$.

iii. *Bessel's Inequality.*

$$\sum_{j=1}^N |\langle x, e_j \rangle|^2 \leq \|x\|_H^2.$$

Proof. We first prove i by showing that $x - \sum_{j=1}^N \langle x, e_j \rangle e_j \perp e_k$ for all $k = 1, \ldots, N$:

$$\langle x - \sum_{j=1}^N \langle x, e_j \rangle e_j, e_k \rangle = \langle x, e_k \rangle - \langle x, e_k \rangle \langle e_k, e_k \rangle = \langle x, e_k \rangle - \langle x, e_k \rangle = 0.$$

Next, ii follows directly from i and the Pythagorean Theorem, and iii clearly follows by dropping the non-negative first term from the right-hand side of ii.

Remark. If $\{e_j : j = 1, \ldots, N\}$ is an ONS, then $\{e_1, \ldots, e_N\}$ is linearly independent. Indeed,

$$0 = \sum_{j=1}^N \alpha_j e_j \implies \alpha_k = 0 \quad \forall k$$

which is shown by taking the inner product of both sides with each e_k.

Cauchy-Schwarz Inequality

For any $x, y \in H$,

$$|\langle x, y \rangle| \leq \|x\|_H \|y\|_H.$$

Proof. Let $y \neq 0$. Consider the single-element ONS $\{\frac{y}{\|y\|_H}\}$. By Bessel's inequality:

$$\sum_{j=1}^1 |\langle x, e_j \rangle|^2 \leq \|x\|_H^2 \implies \left| \left\langle x, \frac{y}{\|y\|_H} \right\rangle \right|^2 \leq \|x\|_H^2 \implies |\langle x, y \rangle|^2 \leq \|x\|_H^2 \|y\|_H^2.$$

Corollary 1

$\|x\|_H = \sqrt{\langle x, x\rangle}$, $x \in H$, is a well-defined norm.

Proof. We prove the triangle inequality. For any $x, y \in H$:

$$\begin{aligned}
\|x+y\|_H^2 &= \langle x+y, x+y\rangle \\
&= \langle x, x\rangle + \langle x, y\rangle + \langle y, x\rangle + \langle y, y\rangle \\
&\leq \langle x, x\rangle + |\langle x, y\rangle| + |\langle y, x\rangle| + \langle y, y\rangle \\
&\leq \|x\|_H^2 + \|y\|_H^2 + 2\|x\|_H\|y\|_H \quad \text{(by Cauchy-Schwarz)} \\
&= (\|x\|_H + \|y\|_H)^2.
\end{aligned}$$

Hilbert Space

A pre-Hilbert space H is called a *Hilbert space* if it is complete with respect to the induced norm $\|\cdot\|_H$.

Projection Theorem

Let H be a pre-Hilbert space. Given a vector subspace $M \subseteq H$, its orthogonal complement is defined as:

$$M^\perp = \{x \in H : \langle x, m\rangle = 0 \quad \forall m \in M\}.$$

We write $x \perp M$ if $x \in M^\perp$. Note that $M^\perp$ is always a closed vector subspace of H due to the continuity of the inner product.

Theorem 3

Let H be a Hilbert space and $M \subseteq H$ be a closed vector subspace.

i. $\forall x \in H$, there exists a unique $m_0 \in M$ such that

$$d := \inf_{m \in M} \|x - m\|_H = \|x - m_0\|_H$$

Moreover, $(x - m_0) \perp M$. The vector m_0 is called the *orthogonal projection* of x onto M, denoted by $P_M x = m_0$.

ii. Let $m_1 \in M$ be such that $(x - m_1) \perp M$. Then $m_1 = m_0$.

Remark. Theorem 3 implies that $\forall x \in H$, there exists a unique decomposition $x = m_0 + w$ where $m_0 \in M$ and $w \in M^\perp$. Thus, $H = M \oplus M^\perp$.

Proof.

i. Let $x \in H$. By the definition of infimum, there exists a sequence $\{m_k\} \subseteq M$ such that $\|x - m_k\|_H \to d$ as $k \to \infty$. We show that $\{m_k\}$ is a Cauchy sequence using the Parallelogram Law:

$$\|m_k - m_\ell\|_H^2 = \|(x - m_\ell) - (x - m_k)\|_H^2 = 2\|x - m_\ell\|_H^2 + 2\|x - m_k\|_H^2 - \|2x - (m_\ell + m_k)\|_H^2.$$

Since M is a vector subspace, $\frac{m_\ell+m_k}{2} \in M$, which implies $\|x - (\frac{m_\ell+m_k}{2})\|_H \geq d$. Thus,

$$\|2x - (m_\ell + m_k)\|_H^2 = 4\left\|x - \left(\frac{m_\ell + m_k}{2}\right)\right\|_H^2 \geq 4d^2.$$

Substituting this back gives:

$$\|m_k - m_\ell\|_H^2 \leq 2\|x - m_\ell\|_H^2 + 2\|x - m_k\|_H^2 - 4d^2 \xrightarrow{k,\ell\to\infty} 2d^2 + 2d^2 - 4d^2 = 0.$$

Since H is complete, $m_0 := \lim_{k\to\infty} m_k$ exists, and by closure of M, $m_0 \in M$ with $\|x - m_0\|_H = d$.

For uniqueness, if $\tilde{m}_0 \in M$ also achieves the distance d, applying the same inequality with $m_k = m_0$ and $m_\ell = \tilde{m}_0$ yields:

$$\|m_0 - \tilde{m}_0\|_H^2 \leq 2d^2 + 2d^2 - 4d^2 = 0 \implies m_0 = \tilde{m}_0.$$

To prove that $(x - m_0) \perp M$, let $m \in M$ and $t \in \mathbb{R}$. Expanding the norm using inner product properties:

$$\|(x - m_0) - tm\|_H^2 = \|x - m_0\|_H^2 - 2t\mathrm{Re}\{\langle x - m_0, m\rangle\} + t^2\|m\|_H^2.$$

Since m_0 minimizes the distance, we have $d^2 \leq \|(x - m_0) - tm\|_H^2$, which simplifies to:

$$0 \leq t^2\|m\|_H^2 - 2t\mathrm{Re}\{\langle x - m_0, m\rangle\} \quad \forall t \in \mathbb{R}.$$

For this quadratic function to remain non-negative for all t, the linear coefficient must vanish:

$$\mathrm{Re}\{\langle x - m_0, m\rangle\} = 0.$$

Replacing t by it in the original formulation yields $\mathrm{Im}\{\langle x - m_0, m\rangle\} = 0$. Since both parts are zero, $\langle x - m_0, m\rangle = 0$, proving orthogonality.

ii. Let $m_1 \in M$ satisfy $(x - m_1) \perp M$. For any $m \in M$, since $(m_1 - m) \in M$, we have $(x - m_1) \perp (m_1 - m)$. By the Pythagorean Theorem:

$$\|x - m\|_H^2 = \|(x - m_1) + (m_1 - m)\|_H^2 = \|x - m_1\|_H^2 + \|m_1 - m\|_H^2 \geq \|x - m_1\|_H^2.$$

Taking the infimum over all $m \in M$ implies $\|x - m_1\|_H^2 = d^2$. By the uniqueness proven in part i, $m_1 = m_0$.

Corollary 2

Let $\{e_1, \ldots, e_N\}$ be an ONS in a Hilbert space H, and let $M = \mathrm{span}\{e_1, \ldots, e_N\}$. Then for any $x \in H$:

$$P_M(x) = \sum_{j=1}^{N} \langle x, e_j\rangle e_j.$$

Proof. Theorem 2 established that $(x - \sum_{j=1}^N \langle x, e_j\rangle e_j) \perp e_k$ for all k, which implies it is orthogonal to the entire span M. Applying Theorem 3 ii completes the proof.

Hilbert Space Basis

Let H be a Hilbert space. An orthonormal system $S \subseteq H$ is called a *complete orthonormal system (CONS)* or a *Hilbert basis* if for any $z \in H$:

$$\langle z, e_\alpha \rangle = 0 \quad \forall e_\alpha \in S \implies z = 0.$$

Example

Let μ be the counting measure on $[0, 1]$. Consider $H = L^2([0, 1], \mu)$, the space of functions $f : [0, 1] \to \mathbb{R}$ such that

$$\sum_{t \in [0,1]} [f(t)]^2 = \int_{[0,1]} [f(t)]^2 d\mu < \infty.$$

The inner product is given by $\langle f, g \rangle = \sum_{t \in [0,1]} f(t)g(t)$. For each $\alpha \in [0, 1]$, defining the standard indicator functions $e_\alpha(t) = 1$ if $\alpha = t$ and 0 otherwise forms a non-countable CONS where $f(t) = \sum_{\alpha \in [0,1]} \langle f, e_\alpha \rangle e_\alpha(t)$.

Theorem 4

Let H be a Hilbert space with a CONS $\{e_\alpha : \alpha \in A\}$. For any $x \in H$:

i. The set $\mathcal{A} := \{\alpha \in A : \langle x, e_\alpha \rangle \neq 0\}$ is at most countable: $\mathcal{A} = \{\alpha_1, \alpha_2, \ldots\}$.

ii. $x = \sum_{\alpha \in A} \langle x, e_\alpha \rangle e_\alpha = \sum_{k=1}^{\infty} \langle x, e_{\alpha_k} \rangle e_{\alpha_k} = \lim_{n \to \infty} \sum_{k=1}^{n} \langle x, e_{\alpha_k} \rangle e_{\alpha_k}$.

iii. *Parseval's Identity.*

$$\|x\|_H^2 = \sum_{\alpha \in A} |\langle x, e_\alpha \rangle|^2 = \sum_{k=1}^{\infty} |\langle x, e_{\alpha_k} \rangle|^2.$$

Proof.

i. For any finite subset $\{\beta_1, \ldots, \beta_n\} \subseteq A$, Bessel's inequality states $\sum_{k=1}^{n} |\langle x, e_{\beta_k} \rangle|^2 \leq \|x\|_H^2 < \infty$. For each $n \geq 1$, define $A_n = \{\alpha \in A : |\langle x, e_\alpha \rangle| > \frac{1}{n}\}$. If any A_n were infinite, it would contain a countable sequence contradicting Bessel's bounded sum. Hence, each A_n is finite, making $\mathcal{A} = \bigcup_{n=1}^{\infty} A_n$ a countable union of finite sets, which is countable.

ii. Let $\{\alpha_k\}_{k=1}^{\infty}$ be an enumeration of $\mathcal{A}$ and define the partial sums $s_n = \sum_{k=1}^{n} \langle x, e_{\alpha_k} \rangle e_{\alpha_k}$. For $m > n$, by the Pythagorean theorem:

$$\|s_m - s_n\|_H^2 = \sum_{k=n+1}^{m} |\langle x, e_{\alpha_k} \rangle|^2.$$

As $n, m \to \infty$, this goes to 0 because the total series $\sum_{k=1}^{\infty} |\langle x, e_{\alpha_k} \rangle|^2$ converges. Thus $\{s_n\}$ is Cauchy, and since H is complete, it converges to some $y \in H$. To show $y = x$, we look at the inner product for any $\alpha \in A$:

$$\langle x - y, e_\alpha \rangle = \langle x, e_\alpha \rangle - \lim_{n \to \infty} \sum_{k=1}^{n} \langle x, e_{\alpha_k} \rangle \langle e_{\alpha_k}, e_\alpha \rangle.$$

If $\alpha \notin \mathcal{A}$, both terms are 0. If $\alpha \in \mathcal{A}$, then $\alpha = \alpha_n$ for some n, yielding $\langle x, e_{\alpha_n} \rangle - \langle x, e_{\alpha_n} \rangle = 0$. Since $\langle x - y, e_\alpha \rangle = 0$ for all α, by the completeness of the system, $x - y = 0 \implies x = y$.

iii. Applying Theorem 2 ii to the finite partial sums yields:

$$\|x\|_H^2 = \left\| x - \sum_{k=1}^{n} \langle x, e_{\alpha_k} \rangle e_{\alpha_k} \right\|_H^2 + \sum_{k=1}^{n} |\langle x, e_{\alpha_k} \rangle|^2.$$

Taking $n \to \infty$, the first norm term vanishes due to part ii, leaving Parseval's Identity.

Theorem 5

A Hilbert space H has a countable CONS if and only if H is separable (contains a countable dense subset).

Proof. ($\implies$) Let $\{e_n\}_{n \geq 1}$ be a countable CONS. The set of all finite linear combinations $\sum_{k=1}^{n} r_k e_k$ where $r_k \in \mathbb{Q} + i\mathbb{Q}$ forms a countable dense subset in H. ($\impliedby$) Let $\{y_1, y_2, \ldots\}$ be a countable dense subset. We can filter out linearly dependent vectors to get a linearly independent sequence $\{x_n\}$ spanning the same subspace. Applying the standard Gram-Schmidt procedure to $\{x_n\}$ constructs a countable orthonormal system $\{e_n\}$ whose span is dense in H, hence forming a CONS.

Proposition 2

Let H be a pre-Hilbert space. For a fixed $z \in H$, define $\ell_z(x) = \langle x, z \rangle$. Then $\ell_z \in H^*$ and $\|\ell_z\|_{H^*} = \|z\|_H$.

Proof. Boundedness follows from Cauchy-Schwarz: $|\ell_z(x)| = |\langle x, z \rangle| \leq \|z\|_H \|x\|_H$, showing $\|\ell_z\|_{H^*} \leq \|z\|_H$. Testing the functional on the unit vector $x = \frac{z}{\|z\|_H}$ (for $z \neq 0$) yields exactly $\|z\|_H$, establishing equality.

Riesz Representation Theorem

Let H be a Hilbert space. For every bounded linear functional $\ell \in H^*$, there exists a unique $z_\ell \in H$ such that

$$\ell(x) = \langle x, z_\ell \rangle, \quad \forall x \in H.$$

Moreover, $\|\ell\|_{H^*} = \|z_\ell\|_H$.

Proof. If $\ell = 0$, choosing $z_\ell = 0$ works. Assume $\ell \neq 0$, and let $M = \ker(\ell)$. Since ℓ is continuous, M is a closed subspace, and $M \neq H$. By the Projection Theorem, $M^\perp \neq \{0\}$. Pick a nonzero $x_0 \in M^\perp$ scaled such that $\ell(x_0) = 1$. For any $x \in H$, the vector $x - \ell(x)x_0$ belongs to M because $\ell(x - \ell(x)x_0) = \ell(x) - \ell(x) = 0$. Since $x_0 \in M^\perp$, we have:

$$\langle x - \ell(x)x_0, x_0 \rangle = 0 \implies \langle x, x_0 \rangle = \ell(x)\|x_0\|_H^2 \implies \ell(x) = \left\langle x, \frac{x_0}{\|x_0\|_H^2} \right\rangle.$$

Setting $z_\ell = \frac{x_0}{\|x_0\|_H^2}$ proves existence. Uniqueness follows since if $\langle x, z_1 \rangle = \langle x, z_2 \rangle$ for all x, then $\langle x, z_1 - z_2 \rangle = 0 \implies z_1 - z_2 = 0$.

Adjoint Operator

Let H be a Hilbert space. For each $T \in \mathcal{L}(H)$, there exists a unique $T^* \in \mathcal{L}(H)$ such that

$$\langle Tx, y \rangle = \langle x, T^*y \rangle \quad \forall x, y \in H.$$

Moreover, $\|T^*\| = \|T\|$.

Proof. For a fixed $y \in H$, $x \mapsto \langle Tx, y \rangle$ is a bounded linear functional on H. By the Riesz Representation Theorem, there exists a unique vector, which we denote as T^*y, such that $\langle Tx, y \rangle = \langle x, T^*y \rangle$. To verify that T^* is linear on a complex Hilbert space, consider $y_1, y_2 \in H$ and $\alpha, \beta \in \mathbb{C}$. For any $x \in H$:

$$\begin{aligned}
\langle x, T^*(\alpha y_1 + \beta y_2) \rangle &= \langle Tx, \alpha y_1 + \beta y_2 \rangle \\
&= \overline{\alpha} \langle Tx, y_1 \rangle + \overline{\beta} \langle Tx, y_2 \rangle \quad \text{(conjugate-linear in 2nd slot)} \\
&= \overline{\alpha} \langle x, T^*y_1 \rangle + \overline{\beta} \langle x, T^*y_2 \rangle \\
&= \langle x, \alpha T^*y_1 + \beta T^*y_2 \rangle \quad \text{(bringing scalars back inside)}
\end{aligned}$$

Since this holds for all $x \in H$, $T^*(\alpha y_1 + \beta y_2) = \alpha T^*y_1 + \beta T^*y_2$. The norm equality follows from:

$$\|T^*\| = \sup_{\|y\| \le 1} \sup_{\|x\| \le 1} |\langle x, T^*y \rangle| = \sup_{\|y\| \le 1} \sup_{\|x\| \le 1} |\langle Tx, y \rangle| = \sup_{\|x\| \le 1} \|Tx\|_H = \|T\|.$$

Self-Adjoint Operator

An operator $T \in \mathcal{L}(H)$ is called *self-adjoint* if $T^* = T$.

Lemma

Let H be a Hilbert space and $T \in \mathcal{L}(H)$. Then,

$$\|T^*T\| = \|T\|^2 = \|T^*\|^2.$$

Proof. For any $x \in H$ with $\|x\|_H \le 1$:

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

Taking the supremum over all such x yields $\|T\|^2 \le \|T^*T\|$. Combined with the standard norm inequality $\|T^*T\| \le \|T^*\|\|T\| = \|T\|^2$, we get equality.

Remark

Let H be a Hilbert space. Then:

1. $\mathcal{A} := \mathcal{L}(H)$ is a C^*-algebra.
2. $\mathcal{A}$ is a Banach algebra satisfying $\|ST\| \le \|S\|\|T\|$.
3. The adjoint operation $T \mapsto T^*$ is an involution satisfying:

- $(\alpha T + \beta S)^* = \overline{\alpha} T^* + \overline{\beta} S^*$.
- $(ST)^* = T^* S^*$.
- $\|T^* T\| = \|T\|^2$.

L^p Spaces

Let $(X, \mathcal{X}, \mu)$ be a measure space. For $0 < p \leq \infty$, we define the space $L^p(X, \mathcal{X}, \mu)$ (abbreviated as $L^p(\mu)$ or L^p) as:

$$L^p = \left\{ f : X \to \mathbb{C} \text{ is measurable} : \int_X |f|^p \, d\mu < \infty \right\}.$$

We identify functions that are equal almost everywhere (a.e.) to form equivalence classes. Additionally, we define L^∞ to be the space of equivalence classes of measurable functions $f : X \to \mathbb{C}$ that are essentially bounded:

$$|f| \leq C \text{ a.e. for some } C > 0.$$

Remark. L^p is a vector space. For instance, if $p \in (0, \infty)$ and $a, b > 0$, the inequality $(a+b)^p \leq 2^p(a^p + b^p)$ holds because:

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

For $f \in L^p$ where $p \in (0, \infty)$, we define the L^p norm (or quasi-norm if $p < 1$) as:

$$\|f\|_{L^p} = \|f\|_p = \left(\int_X |f|^p \, d\mu \right)^{1/p}.$$

For $p = \infty$, the essential supremum norm is defined as:

$$\|f\|_{L^\infty} = \|f\|_\infty = \inf\{C \geq 0 : |f| \leq C \text{ a.e.}\}.$$

Note. By definition, $|f| \leq \|f\|_\infty$ a.e.

Young's Inequality

For $a, b \geq 0$, $p \in (1, \infty)$, and $\frac{1}{p} + \frac{1}{q} = 1$,

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

Proof. Let $a, b > 0$ (the case where $a = 0$ or $b = 0$ is trivial). Since the natural logarithm function ln is concave, we can apply Jensen's inequality:

$$\ln\left(\frac{a^p}{p} + \frac{b^q}{q} \right) \geq \frac{1}{p} \ln(a^p) + \frac{1}{q} \ln(b^q) = \ln(a) + \ln(b) = \ln(ab).$$

Taking the exponential (exp) of both sides yields the desired result.

Hölder's Inequality

Let $p \in [1, \infty]$ and let $\frac{1}{p} + \frac{1}{q} = 1$ (with $\frac{1}{\infty} = 0$). If $f \in L^p$ and $g \in L^q$, then $fg \in L^1$ and:

$$\int_X |fg| \, d\mu \leq \|f\|_p \|g\|_q$$

Proof. For $p = 1$ or $p = \infty$, the proof is straightforward. Let $p \in (1, \infty)$, $f \neq 0$, and $g \neq 0$ a.e. By Young's inequality, for any parameter $t > 0$, we have:

$$\int_X |fg| \, d\mu = \int_X |(tf)(t^{-1}g)| \, d\mu \leq \frac{t^p}{p} \int_X |f|^p \, d\mu + \frac{t^{-q}}{q} \int_X |g|^q \, d\mu$$

Let $A = \int_X |f|^p \, d\mu$ and $B = \int_X |g|^q \, d\mu$. Define the auxiliary function $L(t)$ for $t > 0$:

$$L(t) = \frac{t^p}{p} A + \frac{t^{-q}}{q} B$$

To minimize $L(t)$, we find its derivative and set it to zero:

$$L'(t) = t^{p-1}A - t^{-q-1}B = 0 \implies t^{p+q} = \frac{B}{A} \implies t = \left(\frac{B}{A}\right)^{\frac{1}{pq}}$$

Evaluating $L(t)$ at this critical point gives:

$$\inf_{t>0} L(t) = \frac{1}{p}\left(\frac{B}{A}\right)^{1/q} A + \frac{1}{q}\left(\frac{B}{A}\right)^{-1/p} B = \frac{1}{p}A^{1/p}B^{1/q} + \frac{1}{q}A^{1/p}B^{1/q} = A^{1/p}B^{1/q}$$

Substituting $A^{1/p} = \|f\|_p$ and $B^{1/q} = \|g\|_q$ completes the proof.

Remark. If $f \in L^p$ and $\frac{1}{p} + \frac{1}{q} = 1$, then $|f|^{p-1} \in L^q$.

Proposition 1 (Minkowski's Inequality)

For $p \in [1, \infty]$, $\|\cdot\|_p$ satisfies the triangle inequality and forms a norm on L^p.

Proof. For $p = 1$ or $p = \infty$, the proof is left as an exercise. Let $p \in (1, \infty)$ and $f, g \in L^p$. We assume $f + g \neq 0$ dynamically. Splitting the integrand yields:

$$\int_X |f+g|^p \, d\mu = \int_X |f+g|^{p-1}|f+g| \, d\mu \leq \int_X |f+g|^{p-1}|f| \, d\mu + \int_X |f+g|^{p-1}|g| \, d\mu$$

Applying Hölder's inequality to both integrals on the right-hand side using conjugate exponents p and $q = \frac{p}{p-1}$, we get:

$$\int_X |f+g|^p \, d\mu \leq \left(\int_X |f+g|^{(p-1)q} \, d\mu\right)^{1/q} \|f\|_p + \left(\int_X |f+g|^{(p-1)q} \, d\mu\right)^{1/q} \|g\|_p$$

Since $(p-1)q = p$ and $1 - \frac{1}{q} = \frac{1}{p}$, this simplifies directly to:

$$\int_X |f+g|^p \, d\mu \leq \left(\int_X |f+g|^p \, d\mu \right)^{1-\frac{1}{p}} (\|f\|_p + \|g\|_p)$$

Dividing both sides by $\left(\int_X |f+g|^p \, d\mu\right)^{1-\frac{1}{p}}$ yields:

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

Remark. Let $f_n, f \in L^\infty$. Then $\|f_n - f\|_\infty \to 0$ if and only if there exist representatives $\tilde{f}_n, \tilde{f}$ in their respective equivalence classes such that $\sup_{x \in X} |\tilde{f}_n(x) - \tilde{f}(x)| \to 0$ as $n \to \infty$.

Chebyshev's Inequality

Let $p \in (0, \infty)$ and $f \in L^p$. Then, for any $\epsilon > 0$:

$$\mu(\{x \in X : |f(x)| > \epsilon\}) \leq \frac{\int_X |f|^p \, d\mu}{\epsilon^p}$$

Proof. Let $\epsilon > 0$. By utilizing monotonic sets, we observe:

$$\int_X |f|^p \, d\mu \geq \int_{\{|f|>\epsilon\}} |f|^p \, d\mu \geq \epsilon^p \mu(\{x \in X : |f(x)| > \epsilon\})$$

Theorem 2

L^p is a Banach space for any $p \in [1, \infty]$.

Proof. The case $p = \infty$ is left as an exercise. Let $p \in [1, \infty)$ and let (f_n) be a Cauchy sequence in L^p. By Chebyshev's Inequality, for any $\epsilon > 0$:

$$\mu(\{|f_n - f_m| > \epsilon\}) \leq \frac{\int_X |f_n - f_m|^p \, d\mu}{\epsilon^p} \xrightarrow{n,m\to\infty} 0$$

Hence, (f_n) is Cauchy in measure. This implies there exists a measurable function f and a subsequence (f_{n_k}) such that $f_{n_k} \to f$ a.e. as $k \to \infty$.

Fix $\epsilon > 0$. Choose N such that $\int_X |f_n - f_m|^p \, d\mu \leq \epsilon$ for all $n, m \geq N$. Letting $m = n_l \to \infty$, Fatou's Lemma implies:

$$\int_X |f_n - f|^p \, d\mu = \int_X \lim_{l\to\infty} |f_n - f_{n_l}|^p \, d\mu \leq \liminf_{l\to\infty} \int_X |f_n - f_{n_l}|^p \, d\mu \leq \epsilon$$

for all $n \geq N$. Thus, $f - f_n \in L^p$, which means $f = (f - f_n) + f_n \in L^p$, and $f_n \to f$ in L^p.

Comparison of L^p Spaces

In general, spaces are not nested: neither $L^p \subseteq L^q$ nor $L^q \subseteq L^p$ if $p \neq q$ on arbitrary measure spaces.

Proof. Let $X = (0, \infty)$ equipped with Lebesgue measure μ. Consider the function:

$$f(x) = x^{-\alpha}\chi_{(0,1)}(x) + x^{-\beta}\chi_{[1,\infty)}(x) \quad (\alpha, \beta > 0)$$

Calculating the L^r integral gives $\int_X |f(x)|^r \, d\mu = \int_0^1 x^{-\alpha r} \, dx + \int_1^\infty x^{-\beta r} \, dx$. This converges if and only if $\alpha < \frac{1}{r} < \beta$. Thus, if we select exponents such that $\frac{1}{q} < \alpha < \frac{1}{p} < \beta$, then $f \in L^p$ but $f \notin L^q$. Conversely, choosing $\alpha < \frac{1}{q} < \beta < \frac{1}{p}$ yields $f \in L^q$ but $f \notin L^p$.

Proposition 3

Let $\mu(X) < \infty$ and $0 < p < q \leq \infty$. Then $L^q \subseteq L^p$ and:

$$\|f\|_p \leq \|f\|_q \, \mu(X)^{\frac{1}{p} - \frac{1}{q}}$$

Proof. Applying Hölder's inequality with conjugate exponents $\frac{q}{p}$ and $\frac{q}{q-p}$ yields:

$$\int_X |f|^p \cdot 1 \, d\mu \leq \left(\int_X |f|^q \, d\mu \right)^{\frac{p}{q}} \left(\int_X 1 \, d\mu \right)^{1 - \frac{p}{q}} = \|f\|_q^p \, \mu(X)^{1 - \frac{p}{q}}$$

Taking the p-th root of both sides proves the assertion.

Proposition 4

Let $0 < p < q < r \leq \infty$. Then $L^q \subseteq L^p + L^r$.

Proof. Let $f \in L^q$. We decompose $f = f\chi_{\{|f| \leq 1\}} + f\chi_{\{|f| > 1\}} =: k + h$. Then $|k| \leq |f|$ and $|k| \leq 1$, so $|k|^r \leq |f|^q \in L^1$, implying $k \in L^r$. Similarly, $|h| \leq |f|$ and on the support of h, $|f| > 1$, meaning $|h|^p \leq |f|^q \in L^1$, implying $h \in L^p$.

Proposition 5

Let $0 < p < q < r \leq \infty$. Then $L^p \cap L^r \subseteq L^q$ and $\|f\|_q \leq \|f\|_p^t \|f\|_r^{1-t}$ where $\frac{1}{q} = \frac{t}{p} + \frac{1-t}{r}$.

Proof. Let $f \in L^p \cap L^r$.

i. If $r = \infty$, then $\int_X |f|^q \, d\mu = \int_X |f|^{q-p} |f|^p \, d\mu \leq \|f\|_\infty^{q-p} \|f\|_p^p$. Taking the q-th root gives the result with $t = p/q$.

ii. If $r < \infty$, rewrite $|f|^q = |f|^{tq} |f|^{(1-t)q}$. Apply Hölder's inequality with conjugate exponents $\frac{p}{tq}$ and $\frac{r}{(1-t)q}$. The conditions require $\frac{tq}{p} + \frac{(1-t)q}{r} = 1$, which aligns perfectly with the definition of t.

Proposition 6 (Sequence Spaces)

If $X = \mathbb{N}$ and μ is the counting measure, we denote $\ell^p = L^p(\mathbb{N}, \mu)$. If $0 < p < q \leq \infty$, then $\ell^p \subseteq \ell^q$ and $\|x\|_q \leq \|x\|_p$.

Proof. Let $x \in \ell^p$. For $q = \infty$, $\|x\|_\infty^p = \sup_n |x_n|^p \leq \sum_n |x_n|^p = \|x\|_p^p \implies \|x\|_\infty \leq \|x\|_p$. For $q < \infty$, by interpolation (Proposition 5), $\|x\|_q \leq \|x\|_p^{p/q} \|x\|_\infty^{1-p/q} \leq \|x\|_p$.

Simple Function Spaces

Define $\Sigma = \{f \in \mathcal{E} : f$ vanishes outside a set of finite measure$\}$, where $\mathcal{E}$ is the space of all complex-valued simple functions:

$$\mathcal{E} = \left\{ f = \sum_{j=1}^{n} z_j \chi_{E_j} : n \geq 1,\ z_j \in \mathbb{C},\ E_j \in \mathcal{X} \text{ are disjoint} \right\}$$

Remark. $\mathcal{E} \subseteq L^\infty$ and $\Sigma \subseteq L^p$ for any $p \in (0, \infty]$.

Theorem 2.10 [Folland]

Let $(X, \mathcal{X})$ be a measurable space. If $f : X \to \mathbb{C}$ is measurable, there exists a sequence $\{f_n\} \subseteq \mathcal{E}$ such that $0 \leq |f_1| \leq |f_2| \leq \cdots \leq |f|$, $f_n \to f$ pointwise, and $f_n \to f$ uniformly on any set where f is bounded.

Proposition 7

i. $\mathcal{E}$ is dense in L^∞.

ii. Σ is dense in L^p if $p \in [1, \infty)$.

Proof. Apply Theorem 2.10. For i, since $f \in L^\infty$ is bounded outside a null set, $f_n \to f$ uniformly a.e., meaning $\|f_n - f\|_\infty \to 0$. For ii, if $f \in L^p$, $|f_n - f|^p \leq 2^{p+1}|f|^p \in L^1$. By the Dominated Convergence Theorem, $\int_X |f_n - f|^p \, d\mu \to 0$.

Generalized Dominated Convergence Theorem

Let $\{f_n\}$ be a sequence of measurable functions such that $f_n \to f$ a.e. Suppose there exists a sequence of nonnegative integrable functions $\{g_n\}$ such that $|f_n| \leq g_n$ a.e., $g_n \to g$ a.e., and $\int_X g_n \, d\mu \to \int_X g \, d\mu < \infty$. Then $f_n \to f$ in L^1.

Proof. Consider the sequence of nonnegative functions $h_n = g_n + g - |f_n - f|$. Since $|f_n - f| \leq g_n + g$ a.e., $h_n \geq 0$ a.e. Applying Fatou's Lemma to $\{h_n\}$ yields:

$$\int_X \liminf_{n\to\infty} h_n \, d\mu \leq \liminf_{n\to\infty} \int_X h_n \, d\mu$$

Since $f_n \to f$ and $g_n \to g$ a.e., $\liminf h_n = 2g$ a.e. Thus:

$$\int_X 2g \, d\mu \leq \int_X g \, d\mu + \int_X g \, d\mu - \limsup_{n\to\infty} \int_X |f_n - f| \, d\mu$$

Subtracting $2\int_X g\,d\mu < \infty$ from both sides results in $\limsup_{n\to\infty}\int_X |f_n - f|\,d\mu \le 0$, which ensures $\lim_{n\to\infty}\|f_n - f\|_1 = 0$.

The Dual of L^p

Let $p \in (1,\infty]$ and $q \in [1,\infty)$ with $\frac{1}{p}+\frac{1}{q}=1$. For any $h \in L^q$, define the linear functional T_h on L^p by:

$$T_h(f) = \int_X fh\,d\mu, \quad f \in L^p$$

Then $T_h \in (L^p)^*$ and $\|T_h\|_{(L^p)^*} = \|h\|_q = \sup\{|\int_X fh\,d\mu| : \|f\|_p \le 1\}$.

Theorem 4 (Riesz Representation Theorem for L^p)

Let $p \in (1,\infty)$. For each bounded linear functional $T \in (L^p)^*$, there exists a unique $h \in L^q$ such that $T(f) = \int_X hf\,d\mu$ for all $f \in L^p$, and $\|T\|_{(L^p)^*} = \|h\|_q$. If μ is σ-finite, this also holds for $p = 1$ and $q = \infty$.

Weak L^p

Let $p \in (0,\infty)$. A measurable function f belongs to *weak* L^p if its distribution function $\lambda_f(\alpha) = \mu(\{x \in X : |f(x)| > \alpha\})$ satisfies:

$$[f]_p := \left(\sup_{\alpha>0}\alpha^p\lambda_f(\alpha)\right)^{1/p} < \infty$$

1. By Chebyshev's Inequality, $L^p \subseteq$ weak L^p and $[f]_p \le \|f\|_p$.
2. Using layer-cake integration by parts, if $f \in L^p$, its norm can be computed via:

$$\|f\|_p^p = \int_X |f|^p\,d\mu = p\int_0^\infty \alpha^{p-1}\lambda_f(\alpha)\,d\alpha$$

Convolution and Smoothing

Let $X = \mathbb{R}^d$, $\mathcal{X} = \mathcal{L}$ (the Lebesgue σ-algebra), and let $\mu = m$ denote the standard Lebesgue measure, so $d\mu = dm = dx$.

Lemma 1 (Commutativity of Integrability)

Let $f, g : \mathbb{R}^d \to \mathbb{C}$ be measurable functions, and fix $x_0 \in \mathbb{R}^d$. Then,

$$f(x_0 - \cdot)g(\cdot) \in L^1(\mathbb{R}^d) \iff g(x_0 - \cdot)f(\cdot) \in L^1(\mathbb{R}^d) \tag{1}$$

and whenever the compatibility condition (1) holds, we have the identity:

$$\int_{\mathbb{R}^d} f(x_0 - y)g(y)\,dy = \int_{\mathbb{R}^d} g(x_0 - y)f(y)\,dy.$$

Proof. Let $G : \mathbb{R}^d \to \mathbb{R}^d$ be a global diffeomorphism (bijective with $G, G^{-1} \in C^1(\mathbb{R}^d)$), defining the coordinate transformation $x = G(y) = (g_1(y), \dots, g_d(y))$. By the multi-dimensional change-of-variables formula:

$$\int_{\mathbb{R}^d} f(x)\,dx = \int_{\mathbb{R}^d} f(G(y))\,|\det \nabla G(y)|\,dy,$$

where the Jacobian matrix $\nabla G(y)$ is given by:

$$\nabla G(y) = \begin{bmatrix} \frac{\partial g_1(y)}{\partial y_1} & \cdots & \frac{\partial g_1(y)}{\partial y_d} \\ \vdots & \ddots & \vdots \\ \frac{\partial g_d(y)}{\partial y_1} & \cdots & \frac{\partial g_d(y)}{\partial y_d} \end{bmatrix}.$$

In our specific framework, we set the translation-reflection map $G(y) = x_0 - y = z$. The corresponding Jacobian is:

$$\nabla G(y) = \begin{bmatrix} -1 & 0 & \cdots & 0 \\ 0 & -1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & -1 \end{bmatrix} = -I_d \implies |\det \nabla G(y)| = |(-1)^d| = 1.$$

Applying this transformation preserves the integration measure ($dy = dz$) and yields:

$$\int_{\mathbb{R}^d} f(x_0 - y)g(y)\,dy = \int_{\mathbb{R}^d} f(G(y))g(x_0 - G(y))\,|\det \nabla G(y)|\,dy = \int_{\mathbb{R}^d} g(x_0 - z)f(z)\,dz.$$

□

Convolution

Let $f, g : \mathbb{R}^d \to \mathbb{C}$ be measurable functions such that the integrability condition $f(x - \cdot)g(\cdot) \in L^1(\mathbb{R}^d)$ holds for m-almost every $x \in \mathbb{R}^d$. The *convolution* of f and g, denoted by $f * g$, is the well-defined function given by:

$$(f * g)(x) = \int_{\mathbb{R}^d} f(x - y)g(y)\,dy = \int_{\mathbb{R}^d} g(x - y)f(y)\,dy = (g * f)(x).$$

We say that $(f * g)(x)$ is defined at any point $x \in \mathbb{R}^d$ where the absolute integrand converges:

$$\int_{\mathbb{R}^d} |f(x - y)||g(y)|\,dy < \infty.$$

Support

Given a function $f : \mathbb{R}^d \to \mathbb{C}$, the *support* of f, denoted $\operatorname{supp}(f)$, is the topological closure of its non-vanishing set:

$$\operatorname{supp}(f) = \overline{\{x \in \mathbb{R}^d : f(x) \neq 0\}}.$$

We denote by $C_c(\mathbb{R}^d)$ the vector space of continuous functions with compact support.

Lemma 2 (Convolution of Compactly Supported Functions)

If $f, g \in C_c(\mathbb{R}^d)$, then their convolution $f * g \in C_c(\mathbb{R}^d)$. Moreover, its support satisfies:

$$\operatorname{supp}(f * g) \subseteq \operatorname{supp}(f) + \operatorname{supp}(g),$$

where $A + B = \{x + y : x \in A, y \in B\}$ denotes the algebraic Minkowski sum.

Proof. Because $f, g \in C_c(\mathbb{R}^d)$, both functions are continuous on compact sets and are thus uniformly bounded. Consequently, the integral converges absolutely for every $x \in \mathbb{R}^d$:

$$\int_{\mathbb{R}^d} |f(x-y)||g(y)|\, dy < \infty \quad \forall x \in \mathbb{R}^d.$$

Hence, $f * g$ is pointwise well-defined across all of $\mathbb{R}^d$.

To establish continuity, let $\{x_n\}_{n=1}^{\infty}$ be a sequence in $\mathbb{R}^d$ converging to x_0. Define the auxiliary sequences $h_n(y) = [f(x_n - y) - f(x_0 - y)]g(y)$ and $h(y) = 0$. Since f is continuous, $f(x_n - y) \to f(x_0 - y)$ pointwise for each fixed y, meaning $h_n(y) \to h(y) = 0$ pointwise everywhere.

To invoke the Dominated Convergence Theorem, let $M = \|f\|_\infty < \infty$. We find a uniform L^1-dominating function ϕ:

$$|h_n(y)| = |[f(x_n - y) - f(x_0 - y)]g(y)| \leq (|f(x_n - y)| + |f(x_0 - y)|)|g(y)| \leq 2M|g(y)|.$$

Since $g \in C_c(\mathbb{R}^d)$, the function $\phi(y) = 2M|g(y)|$ is continuous with compact support and is therefore integrable ($\phi \in L^1(\mathbb{R}^d)$). Applying the DCT:

$$\lim_{n\to\infty} ((f * g)(x_n) - (f * g)(x_0)) = \lim_{n\to\infty} \int_{\mathbb{R}^d} h_n(y)\, dy = \int_{\mathbb{R}^d} \lim_{n\to\infty} h_n(y)\, dy = 0.$$

This proves $f * g$ is continuous.

Now consider the algebraic restriction on non-zero values. If $(f * g)(x) \neq 0$, there must exist a point $y \in \mathbb{R}^d$ such that $f(x - y)g(y) \neq 0$. This implies:

$$x - y \in \operatorname{supp}(f) \quad \text{and} \quad y \in \operatorname{supp}(g),$$

which yields $x = (x - y) + y \in \operatorname{supp}(f) + \operatorname{supp}(g)$, and thus:

$$\{x \in \mathbb{R}^d : (f * g)(x) \neq 0\} \subseteq \operatorname{supp}(f) + \operatorname{supp}(g).$$

Taking the topological closure on both sides gives:

$$\operatorname{supp}(f * g) = \overline{\{x \in \mathbb{R}^d : (f * g)(x) \neq 0\}} \subseteq \overline{\operatorname{supp}(f) + \operatorname{supp}(g)}.$$

Since $\operatorname{supp}(f)$ and $\operatorname{supp}(g)$ are compact subsets of $\mathbb{R}^d$, their Minkowski sum $\operatorname{supp}(f) + \operatorname{supp}(g)$ is also compact, which implies it is closed. Therefore, the closure bar can be dropped:

$$\overline{\operatorname{supp}(f) + \operatorname{supp}(g)} = \operatorname{supp}(f) + \operatorname{supp}(g),$$

which confirms that $\operatorname{supp}(f * g) \subseteq \operatorname{supp}(f) + \operatorname{supp}(g)$. Because $\operatorname{supp}(f * g)$ is a closed subset of a compact set, it is compact, meaning $f * g \in C_c(\mathbb{R}^d)$. $\square$

Remark.

1. Any continuous function with compact support is automatically bounded ($C_c(\mathbb{R}^d) \subseteq C_b(\mathbb{R}^d)$) by the Extreme Value Theorem, since it achieves finite absolute maxima on its compact support domain and vanishes identically outside it.
2. The standard Lebesgue measure space $(\mathbb{R}^d, \mathcal{L}, m)$ is σ-finite, as it can be exhaustively partitioned into a countable union of compact balls $B_n(0) = \{x \in \mathbb{R}^d : |x| \leq n\}$, each possessing finite volume:

$$m(B_n(0)) = \frac{\pi^{d/2}}{\Gamma\left(\frac{d}{2}+1\right)} n^d < \infty.$$

Proposition 1 (Convolution $L^p \times L^q$)

Let $p, q \in [1, \infty]$ be conjugate exponents satisfying $\frac{1}{p} + \frac{1}{q} = 1$. If $f \in L^p(\mathbb{R}^d)$ and $h \in L^q(\mathbb{R}^d)$, then $(f * h)(x)$ is explicitly defined for all $x \in \mathbb{R}^d$ and satisfies the uniform bound:

$$\|f * h\|_\infty = \sup_{x \in \mathbb{R}^d} |(f * h)(x)| \leq \|f\|_p \|h\|_q.$$

Furthermore, $f * h$ is uniformly continuous on $\mathbb{R}^d$, and if $p \in (1, \infty)$, it vanishes at infinity: $\lim_{|x| \to \infty} (f * h)(x) = 0$.

Proof. For any fixed $x \in \mathbb{R}^d$, the shift-invariance of the Lebesgue measure implies $\|f(x - \cdot)\|_p = \|f\|_p$. Applying Hölder's inequality:

$$\int_{\mathbb{R}^d} |f(x-y)||h(y)|\, dy \leq \|f(x - \cdot)\|_p \|h\|_q = \|f\|_p \|h\|_q < \infty.$$

Thus, the convolution integral converges absolutely for every x, and the triangle inequality yields:

$$|(f * h)(x)| \leq \int_{\mathbb{R}^d} |f(x-y)||h(y)|\, dy \leq \|f\|_p \|h\|_q \implies \sup_{x \in \mathbb{R}^d} |(f * h)(x)| \leq \|f\|_p \|h\|_q.$$

Continuity Analysis:

i. Let $p \in (1, \infty)$, which implies $q \in (1, \infty)$. By the density of smooth, compactly supported functions in L^p spaces, there exist sequences $\{f_n\}_{n=1}^\infty, \{h_n\}_{n=1}^\infty \subseteq C_c(\mathbb{R}^d)$ such that $\|f_n - f\|_p \to 0$ and $\|h_n - h\|_q \to 0$. By Lemma 2, each $f_n * h_n \in C_c(\mathbb{R}^d)$. Expanding the difference using a telescoping sum gives:

$$\begin{aligned}
\sup_{x \in \mathbb{R}^d} |(f_n * h_n)(x) - (f * h)(x)| &\leq \sup_{x \in \mathbb{R}^d} |((f_n - f) * h_n)(x)| + \sup_{x \in \mathbb{R}^d} |(f * (h_n - h))(x)| \\
&\leq \|f_n - f\|_p \|h_n\|_q + \|f\|_p \|h_n - h\|_q \\
&\leq \left(\sup_{k \in \mathbb{N}} \|h_k\|_q\right) \|f_n - f\|_p + \|f\|_p \|h_n - h\|_q \xrightarrow{n \to \infty} 0.
\end{aligned}$$

Because the sequence of continuous, compactly supported functions $f_n * h_n$ converges uniformly to $f * h$, the limit $f * h$ is uniformly continuous. Furthermore, since each $f_n * h_n$ vanishes identically outside a sufficiently large ball, uniform convergence preserves this behavior at infinity, so $\lim_{|x| \to \infty} (f * h)(x) = 0$.

ii. Let $p = 1$, implying $q = \infty$, so $f \in L^1(\mathbb{R}^d)$ and $h \in L^\infty(\mathbb{R}^d)$. Let $f_n \in C_c(\mathbb{R}^d)$ be a sequence approximating f in L^1 norm ($\|f_n - f\|_1 \to 0$). Then:

$$\sup_{x\in\mathbb{R}^d} |(f_n * h)(x) - (f * h)(x)| \leq \|f_n - f\|_1 \|h\|_\infty \xrightarrow{n\to\infty} 0.$$

Thus, $f_n * h \to f * h$ uniformly. To check the continuity of each $f_n * h$, we invoke the DCT on the shifted functions. Since $f_n \in C_c(\mathbb{R}^d)$, it is uniformly continuous, so $\lim_{x\to x_0}[f_n(x-y) - f_n(x_0-y)] = 0$ pointwise everywhere. Since the integrand is dominated by $2\|f_n\|_\infty |h(y)| \chi_{\mathrm{supp}(f_n)}(x-y)$, the DCT gives:

$$\lim_{x\to x_0} ((f_n * h)(x) - (f_n * h)(x_0)) = \int_{\mathbb{R}^d} \lim_{x\to x_0} [f_n(x-y) - f_n(x_0-y)] h(y)\, dy = 0.$$

The uniform limit of these continuous functions guarantees that $f * h$ is continuous. $\square$

Proposition 2 (Young's Inequality for Convolutions - Base Case)

Let $p \in [1, \infty)$, $f \in L^p(\mathbb{R}^d)$, and $h \in L^1(\mathbb{R}^d)$. Then $f * h$ is well-defined m-almost everywhere, $f * h \in L^p(\mathbb{R}^d)$, and satisfies:

$$\|f * h\|_p \leq \|f\|_p \|h\|_1.$$

Proof. We use Minkowski's inequality for integrals (the continuous integral version of the triangle inequality):

$$\begin{aligned}
\|f * h\|_p = \left(\int_{\mathbb{R}^d} \left| \int_{\mathbb{R}^d} f(x-y)h(y)\, dy \right|^p dx \right)^{1/p} &\leq \int_{\mathbb{R}^d} \left(\int_{\mathbb{R}^d} |f(x-y)|^p\, dx \right)^{1/p} |h(y)|\, dy \\
&= \int_{\mathbb{R}^d} \|f\|_p |h(y)|\, dy \\
&= \|f\|_p \|h\|_1 < \infty.
\end{aligned}$$

Fubini's theorem implies the absolute integral $\int_{\mathbb{R}^d} |f(x-y)||h(y)|\, dy$ is finite for m-almost every x, confirming that $f * h \in L^p(\mathbb{R}^d)$ is well-defined with $\|f * h\|_p \leq \|f\|_p \|h\|_1$. $\square$

Continuity of Shifts

For a measurable function f on $\mathbb{R}^d$ and a translation vector $z \in \mathbb{R}^d$, the translation operator τ_z is defined as:

$$\tau_z f(x) = f(x+z), \quad x \in \mathbb{R}^d.$$

By the translation invariance of Lebesgue measure, $\|\tau_z f\|_p = \|f\|_p$ for any $p \in (0, \infty]$. Note that τ_0 is the identity operator.

Lemma 3 (L^p Continuity of Translation)

Let $p \in [1, \infty)$ and $f \in L^p(\mathbb{R}^d)$. Then:

1. $\lim_{z\to 0} \|\tau_z f - f\|_p = 0$, which means $\int_{\mathbb{R}^d} |f(x+z) - f(x)|^p \, dx \xrightarrow{z\to 0} 0$.
2. $\lim_{z\to z_0} \|\tau_z f - \tau_{z_0} f\|_p = \lim_{z\to z_0} \|\tau_{z-z_0} f - f\|_p = 0$.

Proof. First, let $g \in C_c(\mathbb{R}^d)$. Fix a compact neighborhood of the translation vector, say $|z| \leq 1$. If $\operatorname{supp}(g) \subseteq B_R(0)$, then for all $|z| \leq 1$, the shifted support satisfies $\operatorname{supp}(\tau_z g) \subseteq B_{R+1}(0)$. We can bound the shifted difference using a uniform L^p envelope:

$$|g(x+z) - g(x)|^p \leq (|g(x+z)| + |g(x)|)^p \leq \left(2\|g\|_\infty \chi_{B_{R+1}(0)}(x)\right)^p \in L^1(\mathbb{R}^d).$$

Since g is continuous, $\lim_{z\to 0} |g(x+z) - g(x)|^p = 0$ pointwise everywhere. Applying the DCT directly yields $\lim_{z\to 0} \|\tau_z g - g\|_p = 0$.

Now let $f \in L^p(\mathbb{R}^d)$ be arbitrary. For any test function $g \in C_c(\mathbb{R}^d)$, we use the triangle inequality to insert g:

$$\begin{aligned}
\|\tau_z f - f\|_p &\leq \|\tau_z f - \tau_z g\|_p + \|\tau_z g - g\|_p + \|g - f\|_p \\
&= \|\tau_z (f - g)\|_p + \|\tau_z g - g\|_p + \|g - f\|_p \\
&= 2\|f - g\|_p + \|\tau_z g - g\|_p.
\end{aligned}$$

Taking the limit superior as $z \to 0$ gives:

$$\limsup_{z\to 0} \|\tau_z f - f\|_p \leq 2\|f - g\|_p + \limsup_{z\to 0} \|\tau_z g - g\|_p = 2\|f - g\|_p.$$

Since $C_c(\mathbb{R}^d)$ is dense in $L^p(\mathbb{R}^d)$ for $p < \infty$, for any $\epsilon > 0$ we can choose a function $g \in C_c(\mathbb{R}^d)$ such that $\|f - g\|_p < \frac{\epsilon}{2}$. This means:

$$\limsup_{z\to 0} \|\tau_z f - f\|_p \leq 2\left(\frac{\epsilon}{2}\right) = \epsilon.$$

Because $\epsilon > 0$ is arbitrary, we have $\limsup_{z\to 0} \|\tau_z f - f\|_p = 0$, which implies $\lim_{z\to 0} \|\tau_z f - f\|_p = 0$. $\square$

Standard Mollifiers

Let $\ell : \mathbb{R} \to \mathbb{R}$ be the standard smooth bump generator, and define the radial profile $\tilde{w}(x) := \ell(|x|_2^2)$ for $x \in \mathbb{R}^d$, where $|x|_2 = \sqrt{x_1^2 + \ldots + x_d^2}$. This construction ensures that:

$$\tilde{w} \in C_c^\infty(\mathbb{R}^d) \quad \text{and} \quad \operatorname{supp}(\tilde{w}) = \overline{B_1(0)}.$$

We define the normalized profile $w(x) = c\tilde{w}(x)$, picking the scaling constant $c > 0$ such that $\int_{\mathbb{R}^d} w(x) \, dx = 1$. For any scale parameter $\varepsilon > 0$, we define the *standard mollifier* w_ε by:

$$w_\varepsilon(x) := \varepsilon^{-d} w\left(\frac{x}{\varepsilon}\right), \quad x \in \mathbb{R}^d.$$

By construction, $w_\varepsilon \in C_c^\infty(\mathbb{R}^d)$, $\mathrm{supp}(w_\varepsilon) = \overline{B_\varepsilon(0)}$, $w_\varepsilon(x) \geq 0$, and $\int_{\mathbb{R}^d} w_\varepsilon(x)\,dx = 1$ for all $\varepsilon > 0$.

Probabilistic Interpretation. The mollifier w can be viewed as the probability density function of a random variable $\xi \in \mathbb{R}^d$ that satisfies $|\xi|_2 \leq 1$ almost surely. The rescaled function w_ε then acts as the density of the scaled variable $\varepsilon\xi$. Expectation computations align directly with scaling transformations:

$$\mathbb{E}[f(x - \varepsilon\xi)] = \int_{\mathbb{R}^d} f(x - \varepsilon y) w(y)\,dy = \int_{\mathbb{R}^d} f(x - z) w_\varepsilon(z)\,dz = (f * w_\varepsilon)(x).$$

Smoothing Theorem

Let $p \in [1, \infty)$ and $f \in L^p(\mathbb{R}^d)$. We define the mollified approximation f_ε by:

$$f_\varepsilon(x) = (f * w_\varepsilon)(x) = \int_{\mathbb{R}^d} f(x - y) w_\varepsilon(y)\,dy = \mathbb{E}[f(x - \varepsilon\xi)], \quad x \in \mathbb{R}^d.$$

Then the following properties hold:

1. $f_\varepsilon \in L^p(\mathbb{R}^d)$, with $\|f_\varepsilon\|_p \leq \|f\|_p$, and $f_\varepsilon \to f$ in L^p norm: $\lim_{\varepsilon\to 0} \|f_\varepsilon - f\|_p = 0$.
2. $f_\varepsilon \in C_b^\infty(\mathbb{R}^d)$, meaning f_ε is infinitely differentiable with bounded derivatives of all orders.
3. For any multi-index $\gamma = (\gamma_1, \ldots, \gamma_d)$ with order $|\gamma| = \gamma_1 + \ldots + \gamma_d$, the differential operator $D^\gamma = \frac{\partial^{|\gamma|}}{\partial x_1^{\gamma_1} \ldots \partial x_d^{\gamma_d}}$ distributes onto the mollifier:

$$D^\gamma f_\varepsilon(x) = \int_{\mathbb{R}^d} D^\gamma w_\varepsilon(x - y) f(y)\,dy = \left((D^\gamma w_\varepsilon) * f\right)(x), \quad x \in \mathbb{R}^d.$$

Proof.

1. Applying Proposition 2 with $\|w_\varepsilon\|_1 = 1$ gives $\|f_\varepsilon\|_p = \|f * w_\varepsilon\|_p \leq \|f\|_p \|w_\varepsilon\|_1 = \|f\|_p$. To prove norm convergence, we exploit the identity $\int_{\mathbb{R}^d} w_\varepsilon(y)\,dy = 1$ to rewrite the difference:

$$|f_\varepsilon(x) - f(x)| = \left| \int_{\mathbb{R}^d} [f(x - y) - f(x)] w_\varepsilon(y)\,dy \right| \leq \int_{\mathbb{R}^d} |\tau_{-y} f(x) - f(x)| w_\varepsilon(y)\,dy.$$

Taking the L^p norm and applying Minkowski's inequality for integrals yields:

$$\|f_\varepsilon - f\|_p \leq \int_{\mathbb{R}^d} \|\tau_{-y} f - f\|_p w_\varepsilon(y)\,dy.$$

Since $w_\varepsilon(y)$ vanishes outside $B_\varepsilon(0)$, we can restrict the integration domain and bound the integrand by its supremum over that ball:

$$\|f_\varepsilon - f\|_p \leq \int_{B_\varepsilon(0)} \left(\sup_{|y|\leq\varepsilon} \|\tau_{-y} f - f\|_p \right) w_\varepsilon(y)\,dy = \sup_{|y|\leq\varepsilon} \|\tau_{-y} f - f\|_p.$$

By Lemma 3, $\lim_{\varepsilon\to 0} \left(\sup_{|y|\leq\varepsilon} \|\tau_{-y} f - f\|_p\right) = 0$, which proves that $\lim_{\varepsilon\to 0} \|f_\varepsilon - f\|_p = 0$.

2. The infinite differentiability follows from Leibniz's rule for differentiating under the integral sign, which is justified because $w_\varepsilon \in C_c^\infty(\mathbb{R}^d)$ has bounded derivatives of all orders.

3. Let $q \in [1, \infty]$ be the conjugate exponent of p. Because $w_\varepsilon \in C_c^\infty(\mathbb{R}^d)$, taking any partial derivative preserves smoothness and compact support, so $D^\gamma w_\varepsilon \in C_c^\infty(\mathbb{R}^d) \subset L^q(\mathbb{R}^d)$. Applying Proposition 1:

$$\sup_{x \in \mathbb{R}^d} |D^\gamma f_\varepsilon(x)| = \sup_{x \in \mathbb{R}^d} |((D^\gamma w_\varepsilon) * f)(x)| \leq \|f\|_p \|D^\gamma w_\varepsilon\|_q < \infty.$$

   This confirms that every derivative is uniformly bounded, so $f_\varepsilon \in C_b^\infty(\mathbb{R}^d)$. □

Representations of Positive Linear Functionals

A linear functional $\ell : C(X) \to \mathbb{R}$ is *positive* if $\ell(f) \geq 0$ whenever $f \geq 0$.

Remark. If $f \geq g$, then $\ell(f) \geq \ell(g)$ for a positive functional ℓ. In fact, ℓ is positive if and only if this property holds.

Proposition 2 (Boundedness of Positive Linear Functionals)

If ℓ is positive, then $\ell \in [C(X)]^*$.

Proof. Let $f \in C(X)$. Then we can bound f by its uniform norm:

$$-\|f\|_u \leq f \leq \|f\|_u \implies -\|f\|_u \ell(1) \leq \ell(f) \leq \|f\|_u \ell(1) \implies |\ell(f)| \leq \ell(1)\|f\|_u.$$

This immediately implies continuity with respect to the supremum norm. □

Definition

Let μ be a Borel measure on a metric space X, and let $A \in \mathcal{B}$ (the Borel σ-algebra). We say μ is:

i. *Outer regular* on A if

$$\mu(A) = \inf\{\mu(U) : U \supseteq A,\, U \text{ is open}\}$$

ii. *Inner regular* on A if

$$\mu(A) = \sup\{\mu(K) : K \subseteq A,\, K \text{ is compact}\}$$

iii. *Regular* if μ is both inner and outer regular for all $E \in \mathcal{B}$.

Theorem 1 (Riesz Representation Theorem for Positive Linear Functionals)

Let $\ell : C(X) \to \mathbb{R}$ be a positive linear functional. Then, there exists a unique finite Borel regular measure μ on X such that:

$$\ell(f) = \int_X f \, \mathrm{d}\mu, \quad \forall f \in C(X)$$

i. For any open set $U \subseteq X$,

$$\mu(U) = \sup\{\ell(f) : f \prec U\} \leq \ell(1) = \mu(X)$$

and consequently,

$$\ell(1) = \int_X 1 \, \mathrm{d}\mu = \mu(X).$$

ii. For any compact set $K \subseteq X$,

$$\mu(K) = \inf\{\ell(f) : f \geq \chi_K, \, f \in C(X)\}.$$

Note. For an open set $U \subseteq X$, its indicator function satisfies:

$$\chi_U(x) = \sup\{f(x) : f \prec U\}, \quad \forall x \in X.$$

Comment. Statement i implies statement ii: if K is compact, then its complement K^c is open, so we can write:

$$\begin{aligned}
\mu(K) &= \mu(X) - \mu(K^c) \\
&= \ell(1) - \sup\{\ell(f) : f \prec K^c\} \\
&= \ell(1) - \sup\{\ell(f) : 0 \leq f \leq 1, \, \mathrm{supp}(f) \subseteq K^c\} \\
&= \inf\{\ell(1 - f) : 0 \leq f \leq 1, \, \mathrm{supp}(f) \subseteq K^c\} \\
&\geq \inf\{\ell(g) : g \geq \chi_K, \, g \in C(X)\}.
\end{aligned}$$

For f with $\mathrm{supp}(f) \subseteq K^c$, f is zero on K since K and $\mathrm{supp}(f)$ are disjoint. Thus, $1 - f \geq 1$ on K because $f = 0$ on K, and $1 - f \geq 0$ everywhere since $f \leq 1$. Therefore, $1 - f \geq \chi_K$, where χ_K is the indicator function of K.

Proof. *Uniqueness.* Let μ be a finite regular measure satisfying $\ell(f) = \int f \, \mathrm{d}\mu$. Let $U \subseteq X$ be open. By the note above,

$$\mu(U) = \int_U 1 \, \mathrm{d}\mu = \sup\{\ell(f) : f \prec U\}$$

Let $K \subseteq U$ be compact. By Urysohn's Lemma (Proposition 1), there exists $f \prec U$ such that $f|_K = 1$. Then,

$$\mu(K) \leq \int_X f \, \mathrm{d}\mu = \ell(f)$$

$$\implies \mu(U) = \sup\{\mu(K') : K' \subseteq U,\, K' \text{ compact}\} \le \sup\{\ell(f) : f \prec U\}$$
$$\implies \mu(U) = \sup\{\ell(f) : f \prec U\}, \text{ which is exactly i.}$$

Why does it imply uniqueness? Suppose there were two measures μ_1 and μ_2 representing ℓ:

$$\ell(f) = \int_X f \, \mathrm{d}\mu_1 = \int_X f \, \mathrm{d}\mu_2 \quad \forall f \in C(X).$$

For any open set U, both $\mu_1(U)$ and $\mu_2(U)$ must satisfy:

$$\mu_i(U) = \sup\{\ell(f) : f \prec U\}, \quad i = 1, 2.$$

Since the right-hand side depends only on ℓ, we must have:

$$\mu_1(U) = \mu_2(U).$$

Thus, μ_1 and μ_2 agree on all open sets. The measure μ is outer regular, meaning for any Borel set $E \subseteq X$,

$$\mu(E) = \inf\{\mu(U) : U \supseteq E,\, U \text{ open}\}.$$

Since μ_1 and μ_2 agree on open sets, they must assign the same measure to any Borel set E:

$$\mu_1(E) = \inf\{\mu_1(U) : U \supseteq E\} = \inf\{\mu_2(U) : U \supseteq E\} = \mu_2(E).$$

Hence, $\mu_1 = \mu_2$ on the entire Borel σ-algebra. In short, open sets generate the topology of X, and their measures are uniquely specified by ℓ. Since μ is outer regular, its values on arbitrary sets are determined by its values on open sets.

Existence. We begin by defining the measure on open sets and then constructing an outer measure:

$$\mu(U) := \sup\{\ell(f) : f \prec U\} \text{ for } U \subseteq X \text{ open}.$$
$$\mu^*(E) := \inf\{\mu(U) : U \supseteq E,\, U \text{ open}\} \text{ for } E \subseteq X.$$

Claim 1. μ^* is an outer measure.

An outer measure must satisfy:

1. $\mu^*(\emptyset) = 0$.
2. $A \subseteq B \implies \mu^*(A) \le \mu^*(B)$.
3. $\mu^*\left(\bigcup_{n=1}^\infty E_n\right) \le \sum_{n=1}^\infty \mu^*(E_n)$.

$\mu^*(\emptyset) = 0$ because $\emptyset$ is covered by itself, and $\mu(\emptyset) = 0$. If $A \subseteq B$, any open cover of B also covers A, so the infimum for A is over a larger set of covers, so $\mu^*(A) \le \mu^*(B)$. Thus, it is enough to prove that

$$\mu^*\left(\bigcup_{n=1}^\infty U_n\right) \le \sum_{n=1}^\infty \mu^*(U_n) \text{ for } U_n \text{ open}.$$

The key observation is that open sets can be used to approximate arbitrary sets when defining the outer measure. There exists $f \prec \bigcup_{n=1}^\infty U_n = U$ such that $\ell(f) > \mu\left(\bigcup_{n=1}^\infty U_n\right) - \epsilon$ for a given $\epsilon > 0$ since $\mu(U) = \sup\{\ell(f) : f \prec U\}$. Since every open cover of a compact set has a finite subcover, there exists N such that $\operatorname{supp}(f) \subseteq \bigcup_{n=1}^{N} U_n$. By a partition of unity argument (Corollary 1), there exists $f_n \prec U_n$ such that

$$\left(\sum_{n=1}^{N} f_n\right)\Bigg|_{\operatorname{supp}(f)} = 1, \quad \text{and thus } f = \sum_{n=1}^{N}(f_n f).$$

Applying the functional gives:

$$\mu\left(\bigcup_{n=1}^{\infty} U_n\right) \leq \ell(f) + \epsilon = \left[\sum_{n=1}^{N} \ell(f f_n)\right] + \epsilon \leq \left[\sum_{n=1}^{N} \mu(U_n)\right] + \epsilon.$$

Hence μ^* is a valid outer measure.

Claim 2. μ has a finite regular Borel extension.

By Carathéodory's theorem, μ^* is a measure on the σ-algebra $\mathcal{F}$ of all $A \subseteq X$ satisfying the condition:

$$\mu^*(E) \geq \mu^*(E \cap A) + \mu^*(E \setminus A) \text{ for all } E \subseteq X. \tag{2}$$

We now show $\mathcal{F}$ contains all the open sets of X, i.e., $U \in \mathcal{F}$ if $U \subseteq X$ is open. If this is the case, $\mathcal{F}$ contains $\mathcal{B}$ and μ can be extended to a measure on $\mathcal{B}$.

Let $U \subseteq X$ be open. Equation (2) holds if E itself is open. In this case, $E \cap U$ is open, so there exists $f \prec E \cap U$ such that $\ell(f) > \mu(E \cap U) - \frac{\epsilon}{2}$ where $\epsilon > 0$. Similarly, there exists $g \prec E \setminus \operatorname{supp}(f)$ such that:

$$\ell(g) > \mu(E \setminus \operatorname{supp}(f)) - \frac{\epsilon}{2}.$$

Then, $f + g \prec E$ since $\operatorname{supp}(f)$, $\operatorname{supp}(g) \subseteq E$. Therefore:

$$\begin{aligned}\mu(E) \geq \ell(f+g) &= \ell(f) + \ell(g) \\ &> \mu(E \cap U) - \frac{\epsilon}{2} + \mu(E \setminus \operatorname{supp}(f)) - \frac{\epsilon}{2} \\ &\geq \mu(E \cap U) + \mu^*(E \setminus U) - \epsilon.\end{aligned}$$

Now, we prove Equation (2) holds for an arbitrary set $E \subseteq X$. For all $\epsilon > 0$, there exists an open set $\tilde{E} \supset E$ such that:

$$\mu^*(E) \geq \mu(\tilde{E}) - \epsilon$$

which implies:

$$\mu^*(E) \geq \mu(\tilde{E}) - \epsilon \geq \mu^*(\tilde{E} \cap U) + \mu^*(\tilde{E} \setminus U) - \epsilon \geq \mu^*(E \cap U) + \mu^*(E \setminus U) - \epsilon.$$

Since $\mu(U) \leq \ell(1) < \infty$ for all open $U \subseteq X$, μ is guaranteed to be finite.

We define for any subset $E \subseteq X$ by:

$$\mu^*(E) = \inf\{\mu(U) : U \supseteq E,\ U \text{ open}\}.$$

This ensures μ^* is outer regular.

Why Outer Regularity on the Borel σ-Algebra Implies Inner Regularity (in a Compact Space)? Let $A \in \mathscr{X}$ be a Borel set. Since μ is outer regular, for the complement A^c, there exists an open set $U \supseteq A^c$ such that:

$$\mu(U) \leq \mu(A^c) + \epsilon \text{ for any } \epsilon > 0.$$

Since $U \supseteq A^c$, its complement $K = U^c$ satisfies $K \subseteq A$. In a compact space, closed subsets are compact, so K is compact. From $\mu(X) = \mu(A) + \mu(A^c)$ and $\mu(U) \leq \mu(A^c) + \epsilon$, we have:

$$\mu(K) = \mu(U^c) = \mu(X) - \mu(U) \geq \mu(X) - (\mu(A^c) + \epsilon) = \mu(A) - \epsilon.$$

Since $K \subseteq A$, this shows:

$$\mu(A) \leq \mu(K) + \epsilon \leq \sup\{\mu(K') : K' \subseteq A,\ K' \text{ compact}\} + \epsilon.$$

Taking $\epsilon \to 0$ gives:

$$\mu(A) \leq \sup\{\mu(K) : K \subseteq A,\ K \text{ compact}\}.$$

The reverse inequality ($\geq$) is trivial since $\mu(K) \leq \mu(A)$ for $K \subseteq A$. Thus,

$$\mu(A) = \sup\{\mu(K) : K \subseteq A,\ K \text{ compact}\}.$$

We conclude μ is inner regular. Since μ is both outer and inner regular, μ is regular.

Claim 3. $\int_X f \, \mathrm{d}\mu = \ell(f)$ for all $f \in C(X)$.

Let $0 \leq f \leq 1$, $f \in C(X)$. Let $N > 1$. We can partition the range of f:

$$\begin{aligned} f &= \sum_{n=1}^{N} f \chi_{\{\frac{n-1}{N} \leq f < \frac{n}{N}\}} \\ &= \sum_{n=1}^{N} \left[\left(f - \frac{n-1}{N} \right) \chi_{\{\frac{n-1}{N} \leq f < \frac{n}{N}\}} + \frac{n-1}{N} \chi_{\{\frac{n-1}{N} \leq f < \frac{n}{N}\}} \right] \end{aligned}$$

which yields:

$$f = \sum_{n=1}^{N} \left[\left(f - \frac{n-1}{N} \right) \chi_{\{\frac{n-1}{N} \leq f \leq \frac{n}{N}\}} + \frac{1}{N} \chi_{\{\frac{n}{N} \leq f\}} \right] = \sum_{n=1}^{N} f_n$$

where each $f_n \in C(X)$.

Let us unpack the simple function approximation:

$$\begin{aligned} &\sum_{n=1}^{N} \frac{n-1}{N} \left(\chi_{\{\frac{n-1}{N} \leq f\}} - \chi_{\{\frac{n}{N} \leq f\}} \right) \\ &= \frac{1}{N} \sum_{n=1}^{N} (n-1) \left(\chi_{\{\frac{n-1}{N} \leq f\}} - \chi_{\{\frac{n}{N} \leq f\}} \right) \\ &= \frac{1}{N} \left[1 \cdot \left(\chi_{\{\frac{1}{N} \leq f\}} - \chi_{\{\frac{2}{N} \leq f\}} \right) + \ldots + (N-1) \cdot \left(\chi_{\{\frac{N-1}{N} \leq f\}} - \chi_{\{\frac{N}{N} \leq f\}} \right) \right] \\ &= \frac{1}{N} \left[\chi_{\{\frac{1}{N} \leq f\}} - \chi_{\{\frac{2}{N} \leq f\}} + 2\chi_{\{\frac{2}{N} \leq f\}} - 2\chi_{\{\frac{3}{N} \leq f\}} + \ldots - (N-1)\chi_{\{\frac{N}{N} \leq f\}} \right] \\ &= \frac{1}{N} \left[\chi_{\{\frac{1}{N} \leq f\}} + \chi_{\{\frac{2}{N} \leq f\}} + \ldots + \chi_{\{\frac{N-1}{N} \leq f\}} - (N-1) \cdot \chi_{\{\frac{N}{N} \leq f\}} \right]. \end{aligned}$$

Since $(N-1)\cdot\chi_{\{\frac{N}{N}\le f\}} = 0$ is handled carefully across the sets, we have:

$$\sum_{n=1}^{N}\frac{n-1}{N}\chi_{\{\frac{n-1}{N}\le f<\frac{n}{N}\}} = \sum_{n=1}^{N}\frac{n-1}{N}\left(\chi_{\{\frac{n-1}{N}\le f\}} - \chi_{\{\frac{n}{N}\le f\}}\right) = \frac{1}{N}\sum_{n=1}^{N}\chi_{\{\frac{n}{N}\le f\}}.$$

For each n, the continuous components f_n satisfy:

$$f_n = \left(f - \frac{n-1}{N}\right)\chi_{\{\frac{n-1}{N}\le f\le\frac{n}{N}\}} + \frac{1}{N}\chi_{\{\frac{n}{N}\le f\}} \implies \frac{1}{N}\chi_{\{\frac{n}{N}\le f\}} \le f_n \le \frac{1}{N}\chi_{\{\frac{n-1}{N}\le f\}}.$$

By using statement ii, we can bound the measure of these sets:

$$\mu\left(\left\{x : \frac{n}{N} \le f(x)\right\}\right) \le N\ell(f_n) \le \mu\left(\left\{x : \frac{n-1}{N} \le f(x)\right\}\right)$$

Summing from 1 to N gives:

$$\frac{1}{N}\sum_{n=1}^{N}\mu\left(\frac{n}{N}\le f\right) \le \sum_{n=1}^{N}\ell(f_n) \le \frac{1}{N}\sum_{n=1}^{N}\mu\left(\frac{n-1}{N}\le f\right).$$

Focus on the right-hand side (RHS):

$$\ell(f) \le \sum_{n=1}^{N}\frac{1}{N}\mu\left(\frac{n-1}{N}\le f\right).$$

We can expand each term sequentially:

$$\mu\left(\frac{n-1}{N}\le f\right) = \sum_{k=n}^{N}\mu\left(\frac{k-1}{N}\le f<\frac{k}{N}\right) + \mu(f\ge 1).$$

Since $f \le 1$, the boundary term vanishes, so:

$$\mu\left(\frac{n-1}{N}\le f\right) = \sum_{k=n}^{N}\mu\left(\frac{k-1}{N}\le f<\frac{k}{N}\right).$$

Substitute this back into the RHS sum:

$$\ell(f) \le \sum_{n=1}^{N}\frac{1}{N}\sum_{k=n}^{N}\mu\left(\frac{k-1}{N}\le f<\frac{k}{N}\right).$$

Interchanging the order of summation yields:

$$\sum_{n=1}^{N}\sum_{k=n}^{N}\frac{1}{N}\mu\left(\frac{k-1}{N}\le f<\frac{k}{N}\right) = \sum_{k=1}^{N}\sum_{n=1}^{k}\frac{1}{N}\mu\left(\frac{k-1}{N}\le f<\frac{k}{N}\right)$$

$$\implies \sum_{n=1}^{k} \frac{1}{N}\mu\left(\frac{k-1}{N} \le f < \frac{k}{N}\right) = \frac{k}{N}\mu\left(\frac{k-1}{N} \le f < \frac{k}{N}\right)$$

$$\implies \ell(f) \le \sum_{k=1}^{N} \frac{k}{N}\mu\left(\frac{k-1}{N} \le f < \frac{k}{N}\right).$$

Writing $\frac{k}{N} = \frac{k-1}{N} + \frac{1}{N}$:

$$\ell(f) \le \sum_{k=1}^{N} \left(\frac{k-1}{N} + \frac{1}{N}\right)\mu\left(\frac{k-1}{N} \le f < \frac{k}{N}\right)$$

$$\implies \ell(f) \le \sum_{k=1}^{N} \frac{k-1}{N}\mu\left(\frac{k-1}{N} \le f < \frac{k}{N}\right) + \frac{1}{N}\sum_{k=1}^{N}\mu\left(\frac{k-1}{N} \le f < \frac{k}{N}\right).$$

The second sum represents the full support:

$$\frac{1}{N}\sum_{k=1}^{N}\mu\left(\frac{k-1}{N} \le f < \frac{k}{N}\right) = \frac{1}{N}\mu(f > 0)$$

because the sets $\frac{k-1}{N} \le f < \frac{k}{N}$ form a disjoint partition of $\{f > 0\}$.

Combining both bounds, we trap the functional value:

$$\sum_{n=1}^{N} \frac{n-1}{N}\mu\left(\frac{n-1}{N} \le f < \frac{n}{N}\right) \le \ell(f) \le \sum_{n=1}^{N} \frac{n-1}{N}\mu\left(\frac{n-1}{N} \le f < \frac{n}{N}\right) + \frac{1}{N}\mu(f > 0).$$

The left-hand side is the lower Riemann sum and the right-hand side is the upper Riemann sum. As we integrate via Riemann sums and let $N \to \infty$, both sides converge to the same integral:

$$\int_X f \, \mathrm{d}\mu = \ell(f) = \int_X f \, \mathrm{d}\mu.$$

Therefore, $\int_X f \, \mathrm{d}\mu = \ell(f)$ holds for all $f \in C(X)$. $\square$

Corollary 2 (Regularity of Finite Borel Measures on Compact Metric Spaces)

Let X be a compact metric space. Then, any finite Borel measure γ is regular.

Proof. Let $\ell(f) = \int_X f \, \mathrm{d}\gamma$ for $f \in C(X)$. Since γ is a standard measure, ℓ is positive. By Theorem 1, there exists a regular finite measure μ such that:

$$\ell(f) = \int_X f \, \mathrm{d}\mu = \int_X f \, \mathrm{d}\gamma \quad \forall f \in C(X).$$

Let $K \subset X$ be compact and define the closed sets:

$$F_n = \left\{x \in X : \rho(x, K) \ge \frac{1}{n}\right\} \quad \forall n \ge 1.$$

We show that the complement $F_n^c = \{x \in X : \rho(x, K) < 1/n\}$ is open: Take any $x \in F_n^c$, meaning $\rho(x, K) = d < 1/n$. By definition of infimum, there exists $y \in K$ such that $\rho(x, y) < (1/n + d)/2$. Now, consider the open ball $B(x, r)$ where $r = (1/n - d)/2 > 0$. For any $z \in B(x, r)$, by the triangle inequality:

$$\rho(z, K) \leq \rho(z, y) \leq \rho(z, x) + \rho(x, y) < r + \frac{1/n + d}{2} = \frac{1/n - d}{2} + \frac{1/n + d}{2} = 1/n.$$

Thus, $\rho(z, K) < 1/n$, so $z \in F_n^c$. Since every $x \in F_n^c$ has an open neighborhood entirely contained in F_n^c, F_n^c is open, and hence F_n is closed.

By Urysohn's lemma, there exists continuous functions $f_n : X \to [0, 1]$ such that $f_n|_K = 1$ and $f_n|_{F_n} = 0$. Note that $f_n(x) \to \chi_K(x)$ pointwise for all $x \in X$. By the Dominated Convergence Theorem:

$$\int_X f_n \,\mathrm{d}\mu = \int_X f_n \,\mathrm{d}\gamma \implies \mu(K) = \gamma(K).$$

Therefore, γ matches the regular measure μ on compact sets. For an open set U, complementation gives:

$$\gamma(U) = \gamma(X) - \gamma(U^c) = \mu(X) - \mu(U^c) = \mu(U).$$

Let $A \in \mathcal{B}$ and $\epsilon > 0$. Since μ is regular, there exists open and compact bounds $U_\epsilon \supseteq A \supseteq K_\epsilon$ such that:

$$\mu(U_\epsilon \setminus K_\epsilon) < \epsilon \implies \gamma(U_\epsilon) \leq \gamma(K_\epsilon) + \epsilon \leq \gamma(A) + \epsilon.$$

Hence, γ is outer regular on A. Similarly,

$$\mu(K_\epsilon) \geq \mu(A) - \epsilon \implies \gamma(K_\epsilon) \geq \gamma(A) - \epsilon.$$

Thus, γ is inner regular. We conclude γ is fully regular, meaning $\gamma = \mu$ identically across the entire σ-algebra. □

Lemma 2 (Jordan Decomposition of Linear Functionals)

Let $\ell \in [C(X)]^*$. Then, $\ell = \ell_1 - \ell_2$ where ℓ_1 and ℓ_2 are positive linear functionals.

Proof. Define the positive cone $P = \{f \in C(X) : f \geq 0\}$. For any $f \in P$, define:

$$\ell_1(f) := \sup\{\ell(g) : 0 \leq g \leq f,\, g \in C(X)\}.$$

Note that $\ell_1(f) \geq 0$ since we can choose $g = 0$ in the supremum. We first show that ℓ_1 is linear on P. For any $f, f' \in P$ and $c \geq 0$, we check:

1. *Additivity.* $\ell_1(f + f') = \ell_1(f) + \ell_1(f')$.

   For any $0 \leq g \leq f$ and $0 \leq g' \leq f'$, we have $0 \leq g + g' \leq f + f'$, and thus:

$$\ell(g) + \ell(g') = \ell(g + g') \leq \ell_1(f + f').$$

Taking suprema over g and g' independently gives $\ell_1(f) + \ell_1(f') \leq \ell_1(f + f')$. For the reverse direction, given any $0 \leq h \leq f + f'$, we can decompose $h = u + u'$ where $0 \leq u \leq f$ and $0 \leq u' \leq f'$ by setting $u = \min(h, f)$ and $u' = h - u$. Then,

$$\ell(h) = \ell(u) + \ell(u') \leq \ell_1(f) + \ell_1(f').$$

Taking the supremum over h gives $\ell_1(f + f') \leq \ell_1(f) + \ell_1(f')$.

2. *Homogeneity.* $\ell_1(cf) = c\ell_1(f)$.

$$\ell_1(cf) = \sup\{\ell(g) : 0 \leq g \leq cf\} = \sup\{\ell(ch) : 0 \leq h \leq f\} = c\sup\{\ell(h) : 0 \leq h \leq f\} = c\ell_1(f).$$

We now extend ℓ_1 linearly to the entire space $C(X)$. Any $f \in C(X)$ can be split into positive and negative parts: $f = f_1 - f_2$ where $f_1, f_2 \in P$. We define:

$$\ell_1(f) := \ell_1(f_1) - \ell_1(f_2).$$

If $f = f_1 - f_2 = f_1' - f_2'$, then rearranging gives $f_1 + f_2' = f_1' + f_2$. By additivity on P:

$$\ell_1(f_1) + \ell_1(f_2') = \ell_1(f_1') + \ell_1(f_2) \implies \ell_1(f_1) - \ell_1(f_2) = \ell_1(f_1') - \ell_1(f_2').$$

Thus, the extension is well-defined. Finally, define $\ell_2 := \ell_1 - \ell$. To see that ℓ_2 is positive, observe that for any $f \geq 0$:

$$\ell_2(f) = \ell_1(f) - \ell(f) = \sup\{\ell(g) : 0 \leq g \leq f\} - \ell(f) \geq \ell(f) - \ell(f) = 0.$$

Thus, ℓ_1 and ℓ_2 are both positive linear functionals satisfying $\ell = \ell_1 - \ell_2$. □

Proposition 4 (Density of Continuous Functions in L^p Spaces)

Let X be a compact metric space, μ be a finite regular Borel measure, and $p \in [1, \infty)$. Then, $C(X)$ is dense in $L^p(X, \mu)$.

Proof. Since simple functions (Σ) are dense in L^p and the measure is finite, it suffices to approximate the indicator function χ_E of an arbitrary Borel set $E \in \mathcal{B}$ by continuous functions.

Let $\epsilon > 0$. Since μ is regular, there exist a compact set $K_\epsilon \subseteq E$ and an open set $U_\epsilon \supseteq E$ such that:

$$\mu(U_\epsilon) < \mu(E) + \frac{\epsilon}{2}, \quad \mu(K_\epsilon) > \mu(E) - \frac{\epsilon}{2}.$$

Subtracting these inequalities gives:

$$\mu(U_\epsilon \setminus K_\epsilon) = \mu(U_\epsilon) - \mu(K_\epsilon) < \epsilon.$$

Since $K_\epsilon \subseteq E \subseteq U_\epsilon$, Urysohn's Lemma allows us to construct a continuous function f such that $f = 1$ on K_ϵ, $f = 0$ outside U_ϵ, and $0 \leq f \leq 1$ everywhere. The difference $|\chi_E - f|$ is nonzero only on the symmetric difference region $U_\epsilon \setminus K_\epsilon$, because:

- On K_ϵ, $\chi_E = 1$ and $f = 1$, so $|\chi_E - f| = 0$.

- Outside U_ϵ, $\chi_E = 0$ and $f = 0$, so $|\chi_E - f| = 0$.
- On $U_\epsilon \setminus K_\epsilon$, $|\chi_E - f| \leq 1$ since $0 \leq f \leq 1$ and $\chi_E \in \{0, 1\}$.

Therefore:

$$\int_X |\chi_E - f|^p \, \mathrm{d}\mu \leq \int_{U_\epsilon \setminus K_\epsilon} 1 \, \mathrm{d}\mu = \mu(U_\epsilon \setminus K_\epsilon) < \epsilon.$$

Since simple functions are dense in L^p, and we can approximate any simple function component by continuous functions, we can approximate any general L^p-function by continuous functions. Thus, $C(X)$ is dense in $L^p(X, \mu)$. □

Lusin's Theorem

Let X be a compact metric space and μ be a finite measure. Let f be a bounded measurable function on X ($\|f\|_\infty \leq C$). Then, for all $\epsilon > 0$, there exists $g \in C(X)$ such that $\mu(\{f \neq g\}) \leq \epsilon$. Moreover, $\|g\|_\infty \leq \sup_x |f(x)|$, and $0 \leq g \leq \sup_x |f(x)|$ if $f \geq 0$.

Proof. By Proposition 4, since $C(X)$ is dense in $L^1(X, \mu)$, there exists a sequence of continuous functions $\{f_n\} \subset C(X)$ such that $f_n \to f$ in L^1, which implies a subsequence converges to f almost everywhere. Since $f_n \to f$ almost everywhere and $\mu(X) < \infty$, Egoroff's Theorem guarantees that for any $\epsilon > 0$, there exists a measurable set $A \subset X$ such that:

$$\mu(X \setminus A) \leq \frac{\epsilon}{2},$$

and $f_n \to f$ uniformly on A. Since μ is regular from Corollary 2, for this set A, there exists a compact subset $K \subseteq A$ such that:

$$\mu(A \setminus K) \leq \frac{\epsilon}{2} \implies \mu(X \setminus K) = \mu(X \setminus A) + \mu(A \setminus K) \leq \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon.$$

Since $f_n \to f$ uniformly on A, and $K \subseteq A$, the restriction $f|_K$ is continuous as a uniform limit of continuous functions. By the Tietze Extension Theorem, since K is closed in X and X is a compact metric space, there exists a continuous function $g \in C(X)$ extending it over the whole space:

$$g|_K = f|_K, \quad \sup_X g = \sup_K f, \quad \inf_X g = \inf_K f.$$

By construction, $f = g$ on K and $\mu(X \setminus K) \leq \epsilon$. Therefore:

$$\{x \in X : f(x) \neq g(x)\} \subseteq X \setminus K \implies \mu(\{f \neq g\}) \leq \mu(X \setminus K) \leq \epsilon.$$

Since g is an extension of $f|_K$, and f is bounded by C, we obtain:

$$|g(x)| \leq \sup_K |f| \leq \sup_X |f| = C.$$

If $f \geq 0$, then $g \geq 0$ holds everywhere since $\inf_X g \geq \inf_K f \geq 0$. □

Proposition 5

Let X be a compact metric space, and μ be a finite signed measure. Let $\ell_\mu(f) = \int_X f \, \mathrm{d}\mu$ for $f \in C(X)$. Then, $\ell_\mu \in [C(X)]^*$ and $\|\ell_\mu\|_{[C(X)]^*} = |\mu|(X) = \|\mu\|$.

Proof. By standard integration inequalities, since μ is a finite signed measure, we know $\ell_\mu \in [C(X)]^*$ and $\|\ell_\mu\| \leq \|\mu\|$.

Let D^+, D^- be the disjoint components of the Hahn decomposition such that $\mathrm{d}|\mu| = \chi_{D^+}\mathrm{d}\mu - \chi_{D^-}\mathrm{d}\mu$ and $\mathrm{d}\mu = (\chi_{D^+} - \chi_{D^-})\mathrm{d}|\mu|$. Let $\rho = \chi_{D^+} - \chi_{D^-}$, noting that $|\rho| = \rho^2 = 1$.

Let $\epsilon > 0$. By Lusin's Theorem, there exists $g \in C(X)$ such that $|\mu|(\{\rho \neq g\}) \leq \epsilon$ and $\|g\|_\infty \leq 1$. We compute the action:

$$\begin{aligned}
\ell_\mu(g) = \int_X g \, \mathrm{d}\mu &= \int_X g\rho \, \mathrm{d}|\mu| \\
&= \int_{\{g=\rho\}} g\rho \, \mathrm{d}|\mu| + \int_{\{g\neq\rho\}} g\rho \, \mathrm{d}|\mu| \\
&= \int_X \rho^2 \, \mathrm{d}|\mu| - \int_{\{g\neq\rho\}} \rho^2 \, \mathrm{d}|\mu| + \int_{\{g\neq\rho\}} g\rho \, \mathrm{d}|\mu| \\
&\geq |\mu|(X) - \epsilon - \epsilon = |\mu|(X) - 2\epsilon.
\end{aligned}$$

This follows from the fact that:

$$0 \leq \int_{\{g\neq\rho\}} \rho^2 \, \mathrm{d}|\mu| = |\mu|(\{g \neq \rho\}) \leq \epsilon$$

and

$$\left| \int_{\{g\neq\rho\}} g\rho \, \mathrm{d}|\mu| \right| \leq \int_{\{g\neq\rho\}} |g\rho| \, \mathrm{d}|\mu| \leq \int_{\{g\neq\rho\}} \mathrm{d}|\mu| = |\mu|(\{g \neq \rho\}) \leq \epsilon \implies \int_{\{g\neq\rho\}} g\rho \, \mathrm{d}|\mu| \geq -\epsilon.$$

Therefore,

$$\|\mu\| \geq |\ell_\mu(g)| \geq \ell_\mu(g) \geq |\mu|(X) - 2\epsilon.$$

Taking $\epsilon \to 0$, we confirm $\|\ell_\mu\| = \|\mu\|$. $\square$

Notation. Let $M(X)$ be the space of finite signed Borel measures equipped with the total variation norm $\|\mu\| = |\mu|(X)$.

Theorem 3 (Riesz-Markov-Kakutani Representation Theorem)

Let X be a compact metric space. Then, for every $\ell \in [C(X)]^*$, there exists a unique finite signed measure $\mu \in M(X)$ such that:

$$\ell(f) = \int_X f \, \mathrm{d}\mu, \quad \forall f \in C(X).$$

Moreover, $\|\ell\|_{[C(X)]^*} = \|\mu\|_{M(X)}$.

Proof. Let $\ell \in [C(X)]^*$. By Lemma 2, we split the functional into positive components: $\ell = \ell_1 - \ell_2$. Applying Theorem 1 to each component gives unique finite regular positive measures μ_1, μ_2 such that for all $f \in C(X)$:

$$\ell_1(f) = \int_X f \,\mathrm{d}\mu_1, \quad \text{and } \ell_2(f) = \int_X f \,\mathrm{d}\mu_2.$$

Taking the difference:

$$\ell(f) = \int_X f \,\mathrm{d}(\mu_1 - \mu_2) = \int_X f \,\mathrm{d}\mu, \quad \forall f \in C(X)$$

where $\mu = \mu_1 - \mu_2$. The fact that μ is a finite regular signed measure and is unique follows directly by adapting the uniqueness proof from Corollary 2. Finally, Proposition 5 ensures that $\|\ell\| = \|\mu\|$. □

Schwartz Space of Functions with Rapid Decrease

From Fourier Series to Fourier Transform

1. Let $T > 0$. Then, the family $\left\{\frac{1}{\sqrt{2\pi}} e^{\frac{in\pi}{T}x} : n \in \mathbb{Z}\right\}$ is $2T$-periodic and forms a complete orthonormal system (CONS) in $L^2([-T, T])$ (for complex-valued functions).
2. If $h \in L^2([-T, T])$, then its Fourier series expansion converges in $L^2([-T, T])$:

$$h(x) = \sum_{n=-\infty}^{\infty} c_n e^{\frac{in\pi}{T}x} \quad \text{with} \quad c_n = \frac{1}{2T} \int_{-T}^{T} h(x) e^{-\frac{in\pi}{T}x} \,\mathrm{d}x.$$

Moreover, Parseval's identity holds:

$$\int_{-T}^{T} |h(x)|^2 \,\mathrm{d}x = 2T \sum_{n=-\infty}^{\infty} |c_n|^2.$$

Proof. First, we compute the L^2 norm squared of h:

$$\int_{-T}^{T} |h(x)|^2 \,\mathrm{d}x = \int_{-T}^{T} h(x)\overline{h(x)} \,\mathrm{d}x,$$

Substituting the Fourier series representations for $h(x)$ and $\overline{h(x)}$ yields:

$$\int_{-T}^{T} \left(\sum_{n=-\infty}^{\infty} c_n e^{\frac{in\pi}{T}x} \right) \left(\sum_{m=-\infty}^{\infty} \overline{c_m} e^{-\frac{im\pi}{T}x} \right) \mathrm{d}x,$$

Rearranging the terms, we obtain:

$$\int_{-T}^{T} \sum_{n=-\infty}^{\infty} \sum_{m=-\infty}^{\infty} c_n \overline{c_m} e^{\frac{i(n-m)\pi}{T}x} \,\mathrm{d}x,$$

By Fubini's theorem, we can interchange the summation and integration:

$$\sum_{n=-\infty}^{\infty} \sum_{m=-\infty}^{\infty} c_n \overline{c_m} \int_{-T}^{T} e^{\frac{i(n-m)\pi}{T} x} \, \mathrm{d}x,$$

The exponential functions form an orthogonal system. When $n \neq m$, we have:

$$\int_{-T}^{T} e^{\frac{i(n-m)\pi}{T} x} \, \mathrm{d}x = \left[\frac{T}{i\pi(n-m)} e^{\frac{i(n-m)\pi}{T} x} \right]_{-T}^{T} = 0.$$

When $n = m$, the integral becomes:

$$\int_{-T}^{T} e^{0} \, \mathrm{d}x = \int_{-T}^{T} 1 \, \mathrm{d}x = 2T.$$

Therefore, only the diagonal terms survive, and we get:

$$\sum_{n=-\infty}^{\infty} c_n \overline{c_n} \cdot 2T = 2T \sum_{n=-\infty}^{\infty} |c_n|^2.$$

$\square$

3. Let $f \in C_c^\infty(\mathbb{R})$ such that $f(x) = 0$ for $|x| > R > 0$. Then, for any $T > R + 1$ and $|x| < T$, we can view f as a $2T$-periodic function and expand it as:

$$f(x) = \sum_{n=-\infty}^{\infty} \frac{1}{2T} \int_{-T}^{T} f(y) e^{-\frac{in\pi}{T} y} \, \mathrm{d}y \, e^{\frac{in\pi}{T} x} = \sum_{n=-\infty}^{\infty} \frac{1}{2T} \int_{-\infty}^{\infty} f(y) e^{-i\frac{n}{2T} 2\pi y} \, \mathrm{d}y \, e^{i\frac{n}{2T} 2\pi x}.$$

Let $\hat{f}(\xi) = \int_{-\infty}^{\infty} f(y) e^{-i2\pi\xi y} \, \mathrm{d}y$ for $\xi \in \mathbb{R}$. Note that $\hat{f} \in C^\infty(\mathbb{R})$. Hence, we can write:

$$f(x) = \sum_{n=-\infty}^{\infty} \frac{1}{2T} \hat{f}\left(\frac{n}{2T}\right) e^{i\frac{n}{2T} 2\pi x}, \quad \text{for } |x| < T \text{ and } T > R + 1.$$

By Parseval's identity applied to the interval $[-T, T]$:

$$\int_{-\infty}^{\infty} |f(x)|^2 \, \mathrm{d}x = \int_{-T}^{T} |f(x)|^2 \, \mathrm{d}x = \frac{1}{2T} \sum_{n=-\infty}^{\infty} \left| \hat{f}\left(\frac{n}{2T}\right) \right|^2 \quad \text{for } T > R + 1.$$

For any fixed $x \in \mathbb{R}$, taking the limit as $T \to \infty$ yields:

$$f(x) = \lim_{T \to \infty} \sum_{n=-\infty}^{\infty} \frac{1}{2T} \hat{f}\left(\frac{n}{2T}\right) e^{i2\pi \frac{n}{2T} x} = \int_{-\infty}^{\infty} \hat{f}(\xi) e^{i2\pi\xi x} \, \mathrm{d}\xi.$$

Similarly, for the L^2 norm, we consider the sum for finite $T > R + 1$:

$$\int_{-T}^{T} |f(x)|^2 \, \mathrm{d}x = \frac{1}{2T} \sum_{n=-\infty}^{\infty} \left| \hat{f}\left(\frac{n}{2T}\right) \right|^2.$$

Since f has compact support within $[-R, R]$, whenever $T > R+1$, we have:

$$\int_{-\infty}^{\infty} |f(x)|^2 \, \mathrm{d}x = \int_{-T}^{T} |f(x)|^2 \, \mathrm{d}x.$$

Setting $\xi_n = \frac{n}{2T}$ and $\Delta\xi = \frac{1}{2T}$, the right-hand side can be rewritten as:

$$\sum_{n=-\infty}^{\infty} \left| \hat{f}\left(\frac{n}{2T}\right) \right|^2 \cdot \frac{1}{2T} = \sum_{n=-\infty}^{\infty} |\hat{f}(\xi_n)|^2 \cdot \Delta\xi.$$

This expression is precisely a Riemann sum for the function $|\hat{f}(\xi)|^2$ over the entire real line. As $T \to \infty$, the grid spacing $\Delta\xi = \frac{1}{2T} \to 0$, rendering the partition infinitely fine. Under these conditions (which $\hat{f}$ naturally satisfies since it is the Fourier transform of a smoothly compactly supported function $f \in C_c^\infty(\mathbb{R})$), the Riemann sum converges to the integral:

$$\lim_{T\to\infty} \sum_{n=-\infty}^{\infty} |\hat{f}(\xi_n)|^2 \cdot \Delta\xi = \int_{-\infty}^{\infty} |\hat{f}(\xi)|^2 \, \mathrm{d}\xi.$$

Thus, we establish the inversion formula and Plancherel's theorem for $C_c^\infty(\mathbb{R})$:

$$f(x) = \int_{-\infty}^{\infty} \hat{f}(\xi) e^{i2\pi\xi x} \, \mathrm{d}\xi, \quad x \in \mathbb{R} \quad \text{and} \quad \int_{-\infty}^{\infty} |f(x)|^2 \, \mathrm{d}x = \int_{-\infty}^{\infty} |\hat{f}(\xi)|^2 \, \mathrm{d}\xi.$$

Fourier Transform

The *Fourier transform* of $f \in C_c^\infty(\mathbb{R})$ is defined as:

$$\hat{f}(\xi) = (\mathcal{F}f)(\xi) = \int_{-\infty}^{\infty} f(x) e^{-i2\pi\xi x} \, \mathrm{d}x, \quad \xi \in \mathbb{R}.$$

The *inverse Fourier transform* of $g \in C_c^\infty(\mathbb{R})$ is defined as:

$$(\mathcal{F}^{-1}g)(x) = \int_{-\infty}^{\infty} g(\xi) e^{i2\pi\xi x} \, \mathrm{d}\xi, \quad x \in \mathbb{R}.$$

Consequently, we have $\mathcal{F}^{-1}\mathcal{F}f = f$. Note that $(\mathcal{F}^{-1}f)(x) = \hat{f}(-x)$ for any $x \in \mathbb{R}$, which shows that $\mathcal{F}$ is a unitary operator. It turns out that $\mathcal{F}(C_c^\infty(\mathbb{R})) \subseteq S(\mathbb{R})$, where $S(\mathbb{R})$ denotes the Schwartz space of rapidly decreasing functions, and $\mathcal{F} : S(\mathbb{R}) \to S(\mathbb{R})$ is a bijection.

Space of Functions of Rapid Decrease

On $\mathbb{R}$, the Schwartz space is given by:

$$S(\mathbb{R}) = \left\{ f \in C^\infty(\mathbb{R}) : \sup_{x\in\mathbb{R}} (1+|x|)^N |f^{(n)}(x)| < \infty \text{ for all } N, n \in \mathbb{N}_0 \right\}.$$

To generalize this to higher dimensions, let $f, g : \mathbb{R}^d \to \mathbb{C}$. For a multi-index $\alpha = (\alpha_1, \ldots, \alpha_d) \in \mathbb{N}_0^d$, we define:

- $|\alpha| = \alpha_1 + \cdots + \alpha_d$ (the total order of differentiation),
- $\partial^\alpha = \frac{\partial^{|\alpha|}}{\partial x_1^{\alpha_1} \cdots \partial x_d^{\alpha_d}}$ (the total derivative operator),
- $\alpha! = \alpha_1! \cdots \alpha_d!$ with $0! = 1$,
- $x^\alpha = x_1^{\alpha_1} \cdots x_d^{\alpha_d}$ for $x = (x_1, \dots, x_d) \in \mathbb{R}^d$.

Product Rule (Leibniz's Formula)

$$\partial^\alpha (fg) = \sum_{\beta \le \alpha} \frac{\alpha!}{\beta!(\alpha - \beta)!} \partial^\beta f \, \partial^{\alpha - \beta} g,$$

where $\beta \le \alpha$ signifies that $\beta_k \le \alpha_k$ for each $k = 1, \dots, d$.

Schwartz Space

For $\alpha \in \mathbb{N}_0^d$ and $N \in \mathbb{N}_0$, we introduce the family of seminorms:

$$|f|_{N,\alpha} = \sup_{x \in \mathbb{R}^d} (1 + |x|)^N |\partial^\alpha f(x)|.$$

The Schwartz space of rapidly decreasing functions on $\mathbb{R}^d$ is then defined as:

$$S = S(\mathbb{R}^d) = \left\{ f \in C^\infty(\mathbb{R}^d) : |f|_{N,\alpha} < \infty \text{ for all } N \in \mathbb{N}_0, \alpha \in \mathbb{N}_0^d \right\}.$$

Basic Properties

1. The condition $|f|_{N,\alpha} < \infty$ is true if and only if there exists a constant $C_{\alpha,N} > 0$ such that:
$$|\partial^\alpha f(x)| \le \frac{C_{\alpha,N}}{(1 + |x|)^N} \quad \text{for all } x \in \mathbb{R}^d.$$
2. $C_c^\infty(\mathbb{R}^d) \subseteq S(\mathbb{R}^d)$.
3. As an example, the Gaussian function $e^{-|x|^2} \in S(\mathbb{R}^d)$, but $e^{-|x|^2} \notin C_c^\infty(\mathbb{R}^d)$ because its support is not compact.

Seminorms and Convergence

1. Each $|\cdot|_{N,\alpha}$ constitutes a seminorm on S:
$$|f + g|_{N,\alpha} \le |f|_{N,\alpha} + |g|_{N,\alpha}, \quad |af|_{N,\alpha} = |a||f|_{N,\alpha}.$$
Note that $|f|_{N,\alpha} = 0$ does not individually imply $f = 0$.
2. A sequence $f_k \in S$ converges to $f \in S$ (denoted $f_k \to f$ in S) if:
$$|f_k - f|_{N,\alpha} \to 0 \quad \text{as } k \to \infty \quad \text{for all } N \in \mathbb{N}_0, \alpha \in \mathbb{N}_0^d.$$

3. For $N \in \mathbb{N}_0$, we can define a directed system of norms on S:

$$|f|_{(N)} = \max_{|\alpha| \le N} |f|_{N,\alpha} = \max_{|\alpha| \le N} \sup_{x \in \mathbb{R}^d} (1+|x|)^N |\partial^\alpha f(x)|.profile$$

   These norms are monotonically increasing in N, meaning $|f|_{(N)} \le |f|_{(N+1)}$.

4. The topology of $S(\mathbb{R}^d)$ can be fully structuralized via the translation-invariant metric:

$$\rho(f,g) = \sum_{N=1}^{\infty} 2^{-N} \min\left(|f-g|_{(N)}, 1\right).$$

Structural Properties of $S(\mathbb{R}^d)$

1. If $f \in S$, then $\partial^\alpha f \in S$ for any multi-index α.
2. If $f, g \in S$, then their pointwise product satisfies $fg \in S$.
3. If $f \in S$, then $Pf \in S$ for any polynomial function $P(x)$.
4. If $f \in S$, then $f \in L^p(\mathbb{R}^d)$ for any $p > 0$.

   Proof. For $p = \infty$, the statement follows directly from the definition of the space. For $0 < p < \infty$, we write:

$$\int_{\mathbb{R}^d} |f(x)|^p \, \mathrm{d}x = \int_{\mathbb{R}^d} \left(|f(x)|(1+|x|)^{\frac{d+1}{p}}\right)^p \frac{\mathrm{d}x}{(1+|x|)^{d+1}} \le |f|^p_{\lceil \frac{d+1}{p} \rceil, 0} \int_{\mathbb{R}^d} \frac{\mathrm{d}x}{(1+|x|)^{d+1}} < \infty.$$

   □

5. If $f, g \in S$, then their convolution $f * g \in S$.

   Proof. By the differentiating properties of convolutions, we have:

$$\partial^\alpha (f * g)(x) = \int_{\mathbb{R}^d} \partial^\alpha f(x-y) g(y) \, \mathrm{d}y.$$

   Using the submultiplicative inequality $1+|x| \le 1+|x-y|+|y| \le (1+|x-y|)(1+|y|)$, we find:

$$\begin{aligned}
(1+|x|)^N |\partial^\alpha (f*g)(x)| &\le \int_{\mathbb{R}^d} (1+|x-y|)^N |\partial^\alpha f(x-y)| (1+|y|)^N |g(y)| \, \mathrm{d}y \\
&\le |f|'_{N,\alpha} \int_{\mathbb{R}^d} (1+|y|)^{N+d+1} |g(y)| \frac{\mathrm{d}y}{(1+|y|)^{d+1}} \\
&\le |f|_{N,\alpha} |g|'_{N+d+1,0} \int_{\mathbb{R}^d} \frac{\mathrm{d}y}{(1+|y|)^{d+1}} < \infty.
\end{aligned}$$

   □

Continuous Linear Functionals on S

A linear functional $T : S(\mathbb{R}^d) \to \mathbb{C}$ is continuous with respect to the metric ρ if and only if:

> There exist constants $C > 0$ and $N \in \mathbb{N}_0$ such that $|T(f)| \leq C|f|_{(N)}$ for all $f \in S$.

Proof. ($\Rightarrow$) Assume T is continuous. Since T is linear and continuous at zero, for any $\epsilon = 1$, there exists a $\delta > 0$ such that $|T(f)| \leq 1$ whenever $\rho(f, 0) \leq 2\delta$. We observe that:

$$\rho(f, 0) = \sum_{N=1}^{\infty} 2^{-N} \min\{1, |f|_{(N)}\} \leq \sum_{N=1}^{N_0} 2^{-N}|f|_{(N_0)} + \sum_{N=N_0+1}^{\infty} 2^{-N} \leq |f|_{(N_0)} + 2^{-N_0}.$$

Choose $N_0 \in \mathbb{N}$ large enough so that $2^{-N_0} \leq \delta$. Now, pick any non-zero $g \in S$, and scale it by defining $f = \frac{\delta}{|g|_{(N_0)}} g$. Then $|f|_{(N_0)} = \delta$, which implies:

$$\rho(f, 0) \leq |f|_{(N_0)} + 2^{-N_0} \leq \delta + \delta = 2\delta \implies |T(f)| \leq 1.$$

By substituting back the relation between f and g using linearity, we conclude:

$$\left| T\left(\frac{\delta}{|g|_{(N_0)}} g \right) \right| \leq 1 \implies |T(g)| \leq \frac{1}{\delta} |g|_{(N_0)}.$$

Setting $C = \frac{1}{\delta}$ completes this direction.

($\Leftarrow$) Conversely, suppose there exist $C > 0$ and $N_0 \in \mathbb{N}_0$ such that $|T(f)| \leq C|f|_{(N_0)}$ for all $f \in S$. Let $f_n \to f$ in S. Then by definition, $|f_n - f|_{(N_0)} \to 0$ as $n \to \infty$. By linearity, we have:

$$|T(f_n) - T(f)| = |T(f_n - f)| \leq C|f_n - f|_{(N_0)} \to 0.$$

Thus, $T(f_n) \to T(f)$, proving continuity. $\square$

Corollary 0

A linear functional $T : S \to \mathbb{C}$ is continuous if and only if there exist a constant $C > 0$, finite collections of multi-indices $\alpha_1, \ldots, \alpha_m \in \mathbb{N}_0^d$, and integers $N_1, \ldots, N_m \in \mathbb{N}_0$ such that:

$$|T(f)| \leq C \sum_{k=1}^{m} |f|'_{N_k, \alpha_k} \quad \text{for all } f \in S.$$

Fourier Transform on $S(\mathbb{R}^d)$

i. Given $f \in L^1(\mathbb{R}^d)$, its *Fourier transform* $\mathcal{F}f = \hat{f}$ is defined by:

$$\mathcal{F}f(\xi) = \hat{f}(\xi) = \int_{\mathbb{R}^d} f(x) e^{-i2\pi x \cdot \xi} \, \mathrm{d}x, \quad \xi \in \mathbb{R}^d,$$

where $x \cdot \xi = x_1\xi_1 + \cdots + x_d\xi_d$.

ii. Given $g \in L^1(\mathbb{R}^d)$, its *inverse Fourier transform* $\mathcal{F}^{-1}g$ is defined by:

$$\mathcal{F}^{-1}g(x) = \int_{\mathbb{R}^d} g(\xi) e^{i2\pi x \cdot \xi} \, \mathrm{d}\xi, \quad x \in \mathbb{R}^d.$$

Basic Properties of $\mathcal{F}$ and $\mathcal{F}^{-1}$

a. $\mathcal{F}f(\xi)$ and $\mathcal{F}^{-1}g(x)$ are bounded, uniformly continuous functions.

b. $\mathcal{F}$ and $\mathcal{F}^{-1}$ are linear operators.

c. $\|\mathcal{F}f\|_{L^\infty} \leq \|f\|_{L^1}$ because $|\mathcal{F}f(\xi)| \leq \int_{\mathbb{R}^d} |f(x)|\,\mathrm{d}x$.

d. $\|\mathcal{F}^{-1}g\|_{L^\infty} \leq \|g\|_{L^1}$.

e. $\mathcal{F}^{-1}g(x) = \mathcal{F}g(-x)$ for all $x \in \mathbb{R}^d$.

f. Distributional derivatives transform as $\widehat{\partial^\alpha f}(\xi) = (i2\pi\xi)^\alpha \hat{f}(\xi)$.

Properties of Fourier Transform

Property 1

If $f \in S$, then for any multi-index $\alpha \in \mathbb{N}_0^d$:

$$\widehat{\partial^\alpha f}(\xi) = (i2\pi\xi)^\alpha \hat{f}(\xi) = (i2\pi)^{|\alpha|}\xi_1^{\alpha_1}\cdots\xi_d^{\alpha_d}\hat{f}(\xi).$$

Proof. Consider the case $d = 1$ and $\alpha = 1$. We want to show $\widehat{f^{(1)}}(\xi) = i2\pi\xi\hat{f}(\xi)$. Using integration by parts:

$$\begin{aligned}\widehat{f^{(1)}}(\xi) &= \int_{-\infty}^{\infty} f'(x)e^{-i2\pi x\xi}\,\mathrm{d}x = \lim_{n\to\infty}\int_{-n}^{n} f'(x)e^{-i2\pi x\xi}\,\mathrm{d}x \\ &= \lim_{n\to\infty}\left[f(x)e^{-i2\pi x\xi}\right]_{-n}^{n} - \int_{-n}^{n} f(x)\left(-i2\pi\xi e^{-i2\pi x\xi}\right)\,\mathrm{d}x.\end{aligned}$$

Since $f \in S$, $\lim_{n\to\infty} f(\pm n) = 0$, so the boundary terms vanish, leaving:

$$\widehat{f^{(1)}}(\xi) = 0 + i2\pi\xi\int_{-\infty}^{\infty} f(x)e^{-i2\pi x\xi}\,\mathrm{d}x = i2\pi\xi\hat{f}(\xi).$$

The general result follows by iterating this process for each partial derivative. □

Example

Let $\Delta u(x) = \sum_{k=1}^d \frac{\partial^2}{\partial x_k^2}u(x)$ be the Laplacian of $u \in S$. Find $\widehat{\Delta u}(\xi)$.

Each operator $\frac{\partial^2}{\partial x_k^2}$ matches a multi-index $\alpha = (0, \ldots, 2, \ldots, 0)$ where 2 is in the k-th position. Applying Property 1:

$$\begin{aligned}\widehat{\Delta u}(\xi) &= \sum_{k=1}^d \widehat{\frac{\partial^2 u}{\partial x_k^2}}(\xi) = \sum_{k=1}^d (i2\pi\xi_k)^2\hat{u}(\xi) \\ &= \hat{u}(\xi)\left(-4\pi^2\sum_{k=1}^d \xi_k^2\right) = -4\pi^2|\xi|^2\hat{u}(\xi).\end{aligned}$$

Property 2

1. Let $f \in S$ and $y \in \mathbb{R}^d$. Then shifts transform into modulation:

$$\widehat{f(\cdot + y)}(\xi) = e^{i2\pi\xi\cdot y}\hat{f}(\xi).$$

Proof. Using the substitution $z = x + y$ with $\mathrm{d}z = \mathrm{d}x$:

$$\begin{aligned}\widehat{f(\cdot + y)}(\xi) &= \int_{\mathbb{R}^d} f(x+y)e^{-i2\pi x\cdot\xi}\,\mathrm{d}x = \int_{\mathbb{R}^d} f(z)e^{-i2\pi(z-y)\cdot\xi}\,\mathrm{d}z \\ &= e^{i2\pi y\cdot\xi}\int_{\mathbb{R}^d} f(z)e^{-i2\pi z\cdot\xi}\,\mathrm{d}z = e^{i2\pi y\cdot\xi}\hat{f}(\xi).\end{aligned}$$

□

2. Let $f \in S$ and $\delta \in \mathbb{R}\setminus\{0\}$. Dilation transforms inversely:

$$\widehat{f(\delta\cdot)}(\xi) = |\delta|^{-d}\hat{f}\left(\frac{\xi}{\delta}\right).$$

Proof. Using the change of coordinates $z = \delta x$, we have $\mathrm{d}z = |\delta|^d\mathrm{d}x$:

$$\widehat{f(\delta\cdot)}(\xi) = \int_{\mathbb{R}^d} f(\delta x)e^{-i2\pi x\cdot\xi}\,\mathrm{d}x = \int_{\mathbb{R}^d} f(z)e^{-i2\pi\frac{z\cdot\xi}{\delta}}|\delta|^{-d}\,\mathrm{d}z = |\delta|^{-d}\hat{f}\left(\frac{\xi}{\delta}\right).$$

□

Example

Let $f \in S$ such that $\int_{\mathbb{R}^d} f(x)\,\mathrm{d}x \neq 0$. Show that:

$$\lim_{\delta\to 0}|\widehat{f(\delta\cdot)}(\xi)| = \begin{cases} 0, & \xi \neq 0, \\ +\infty, & \xi = 0. \end{cases}$$

For any $\delta \neq 0$, Property 2b gives $|\widehat{f(\delta\cdot)}(\xi)| = |\delta|^{-d}\left|\hat{f}\left(\frac{\xi}{\delta}\right)\right|$. Since $\hat{f} \in S$, it is bounded by its seminorms:

$$|\hat{f}(\xi)| \leq \frac{C}{(1+|\xi|)^{d+1}}, \quad \text{where } C = |\hat{f}|_{d+1,0}.$$

When $\xi \neq 0$, we can manipulate the expression:

$$|\widehat{f(\delta\cdot)}(\xi)| = \frac{1}{|\xi|^d}\cdot\left(\frac{|\xi|}{|\delta|}\right)^d\cdot\left|\hat{f}\left(\frac{\xi}{\delta}\right)\right| \leq \frac{1}{|\xi|^d}\cdot\left(\frac{|\xi|}{|\delta|}\right)^d\cdot\frac{|\hat{f}|_{d+1,0}}{(1+|\xi|/|\delta|)^{d+1}} \to 0 \quad \text{as } \delta \to 0.$$

When $\xi = 0$, the expression simplifies to:

$$|\widehat{f(\delta\cdot)}(0)| = |\delta|^{-d}|\hat{f}(0)| \to \infty \quad \text{as } \delta \to 0,$$

since $\hat{f}(0) = \int_{\mathbb{R}^d} f(x)\,\mathrm{d}x \neq 0$. This highlights the reciprocal relation between physical scale and frequency scale under dilation.

Property 3

If $f \in S$, then $\hat{f} \in C^\infty(\mathbb{R}^d)$ with all derivatives bounded, satisfying:

$$\partial^\beta \hat{f}(\xi) = \hat{g}(\xi) \quad \text{where} \quad g(x) = (-i2\pi x)^\beta f(x) \in S, \quad \beta \in \mathbb{N}_0^d.$$

Proof. Differentiating under the integral sign (justified by the rapid decay of f):

$$\partial^\beta \hat{f}(\xi) = \int_{\mathbb{R}^d} f(x) \frac{\partial^\beta}{\partial \xi^\beta} \left(e^{-i2\pi x\cdot\xi} \right) \mathrm{d}x = \int_{\mathbb{R}^d} f(x)(-i2\pi x)^\beta e^{-i2\pi x\cdot\xi} \, \mathrm{d}x = \hat{g}(\xi).$$

□

Proposition 1

The Fourier transform maps the Schwartz space into itself continuously, $\mathcal{F} : S \to S$.

Proof. Let $f \in S$. By Property 3, $\hat{f} \in C^\infty(\mathbb{R}^d)$. To verify $\hat{f} \in S$, we must show that $\sup_{\xi\in\mathbb{R}^d} |\xi^\alpha \partial^\beta \hat{f}(\xi)| < \infty$ for all $\alpha, \beta \in \mathbb{N}_0^d$. We write:

$$\begin{aligned} \xi^\alpha \partial^\beta \hat{f}(\xi) &= (i2\pi)^{-|\alpha|} (i2\pi\xi)^\alpha \partial^\beta \hat{f}(\xi) \\ &= (i2\pi)^{-|\alpha|} (i2\pi\xi)^\alpha \hat{g}(\xi) \quad \text{with } g(x) = (-i2\pi x)^\beta f(x) \in S. \end{aligned}$$

Applying Property 1 to the function $g \in S$, this becomes:

$$\xi^\alpha \partial^\beta \hat{f}(\xi) = (i2\pi)^{-|\alpha|} \widehat{\partial^\alpha g}(\xi),$$

Taking the supremum over all frequencies $\xi \in \mathbb{R}^d$:

$$\sup_{\xi\in\mathbb{R}^d} |\xi^\alpha \partial^\beta \hat{f}(\xi)| = (2\pi)^{-|\alpha|} \sup_{\xi\in\mathbb{R}^d} |\widehat{\partial^\alpha g}(\xi)| \le (2\pi)^{-|\alpha|} \|\partial^\alpha g\|_{L^1} < \infty,$$

where the L^1 norm is finite since $\partial^\alpha g \in S \subset L^1(\mathbb{R}^d)$. Thus, $\hat{f} \in S$. □

Proposition 2

Let $f, g \in S$. Then $\widehat{f * g}(\xi) = \hat{f}(\xi)\hat{g}(\xi)$.

Proof. By expanding the definition and applying Fubini's Theorem:

$$\begin{aligned} \widehat{f * g}(\xi) &= \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} f(x-y) g(y) \, \mathrm{d}y \, e^{-i2\pi x\cdot\xi} \, \mathrm{d}x \\ &= \int_{\mathbb{R}^d} \left(\int_{\mathbb{R}^d} f(x-y) \, e^{-i2\pi(x-y)\cdot\xi} \, \mathrm{d}x \right) g(y) \, e^{-i2\pi y\cdot\xi} \, \mathrm{d}y. \end{aligned}$$

Setting $z = x - y$ with $\mathrm{d}z = \mathrm{d}x$ inside the inner integral yields:

$$\widehat{f * g}(\xi) = \int_{\mathbb{R}^d} \left(\int_{\mathbb{R}^d} f(z) \, e^{-i2\pi z\cdot\xi} \, \mathrm{d}z \right) g(y) \, e^{-i2\pi y\cdot\xi} \, \mathrm{d}y = \hat{f}(\xi) \cdot \hat{g}(\xi).$$

Fubini's theorem is fully applicable because:

$$\int_{\mathbb{R}^d} \int_{\mathbb{R}^d} |f(x-y)||g(y)| \, \mathrm{d}y \, \mathrm{d}x = \|f\|_{L^1} \|g\|_{L^1} < \infty.$$

□

Proposition 3

Let $f, g \in S$. Then the following dual identities hold:

$$\int_{\mathbb{R}^d} \hat{f}(z)g(z)\,\mathrm{d}z = \int_{\mathbb{R}^d} f(y)\hat{g}(y)\,\mathrm{d}y,$$

$$\int_{\mathbb{R}^d} \mathcal{F}^{-1}(f)(z)g(z)\,\mathrm{d}z = \int_{\mathbb{R}^d} f(y)\mathcal{F}^{-1}(g)(y)\,\mathrm{d}y.$$

Proof. Using Fubini's Theorem to change the order of integration:

$$\begin{aligned}\int_{\mathbb{R}^d} \hat{f}(z)g(z)\,\mathrm{d}z &= \int_{\mathbb{R}^d}\left(\int_{\mathbb{R}^d} f(y)\,e^{-i2\pi y\cdot z}\,\mathrm{d}y\right)g(z)\,\mathrm{d}z\\ &= \int_{\mathbb{R}^d}\left(\int_{\mathbb{R}^d} g(z)\,e^{-i2\pi y\cdot z}\,\mathrm{d}z\right)f(y)\,\mathrm{d}y = \int_{\mathbb{R}^d}\hat{g}(y)f(y)\,\mathrm{d}y.\end{aligned}$$

This is valid since $f, g \in S \subset L^1(\mathbb{R}^d)$, meaning:

$$\int_{\mathbb{R}^d}\int_{\mathbb{R}^d} |f(y)||g(z)||e^{-i2\pi y\cdot z}|\,\mathrm{d}y\,\mathrm{d}z = \|f\|_{L^1}\|g\|_{L^1} < \infty.$$

An identical argument swapping $e^{-i2\pi y\cdot z}$ for $e^{i2\pi y\cdot z}$ establishes the $\mathcal{F}^{-1}$ identity. $\square$

Inverse of Fourier Transform

Theorem 1

Let $f \in S$. Then $\mathcal{F}^{-1}\mathcal{F}(f)(x) = \mathcal{F}\mathcal{F}^{-1}(f)(x) = f(x)$ for all $x \in \mathbb{R}^d$. In particular,

$$\mathcal{F}^{-1}(\hat{f})(x) = \int_{\mathbb{R}^d} \hat{f}(\xi)e^{i2\pi x\cdot\xi}\,\mathrm{d}\xi = f(x).$$

Consequently, both $\mathcal{F}, \mathcal{F}^{-1} : S \to S$ are bijections.

Proof. Let $f \in S$. We first prove the inversion identity at the origin $x = 0$, i.e., $f(0) = \int_{\mathbb{R}^d} \hat{f}(\xi)\,\mathrm{d}\xi$. Let $k(x) = e^{-\pi|x|^2}$ be the standard Gaussian, whose Fourier transform satisfies $\hat{k}(y) = e^{-\pi|y|^2} = k(y)$. For any $\delta > 0$, by the Lebesgue Dominated Convergence Theorem:

$$\int_{\mathbb{R}^d} f(\delta y)\hat{k}(y)\,\mathrm{d}y \longrightarrow f(0)\int_{\mathbb{R}^d}\hat{k}(y)\,\mathrm{d}y = f(0) \quad \text{as } \delta \to 0.$$

Applying Proposition 3 to the left-hand side:

$$\begin{aligned}\int_{\mathbb{R}^d} f(\delta y)\hat{k}(y)\,\mathrm{d}y &= \int_{\mathbb{R}^d}\widehat{f(\delta\cdot)}(\xi)k(\xi)\,\mathrm{d}\xi\\ &= \int_{\mathbb{R}^d}\delta^{-d}\hat{f}\left(\frac{\xi}{\delta}\right)k(\xi)\,\mathrm{d}\xi \quad \text{(by Property 2b)}\\ &= \int_{\mathbb{R}^d}\delta^{-d}\hat{f}(z)k(\delta z)\delta^d\,\mathrm{d}z \quad \left(\text{substituting } z = \frac{\xi}{\delta},\ \mathrm{d}\xi = \delta^d\mathrm{d}z\right)\\ &= \int_{\mathbb{R}^d}\hat{f}(z)k(\delta z)\,\mathrm{d}z.\end{aligned}$$

By the Dominated Convergence Theorem, as $\delta \to 0$, $k(\delta z) = e^{-\pi\delta^2|z|^2} \to k(0) = 1$ pointwise, so:

$$\lim_{\delta\to 0} \int_{\mathbb{R}^d} \hat{f}(z)k(\delta z)\,\mathrm{d}z = \int_{\mathbb{R}^d} \hat{f}(z)\cdot 1\,\mathrm{d}z = \int_{\mathbb{R}^d} \hat{f}(z)\,\mathrm{d}z.$$

Equating the two limits gives $f(0) = \int_{\mathbb{R}^d} \hat{f}(z)\,\mathrm{d}z$.

To extend this to a general point $y \in \mathbb{R}^d$, we define a shifted function $f_y(x) = f(x+y)$, which also belongs to S. Applying the origin inversion formula to f_y:

$$\begin{aligned} f(y) = f_y(0) &= \int_{\mathbb{R}^d} \mathcal{F}(f_y)(\xi)\,\mathrm{d}\xi \\ &= \int_{\mathbb{R}^d} \hat{f}(\xi)e^{i2\pi\xi\cdot y}\,\mathrm{d}\xi \quad \text{(by Property 2a)} \\ &= \mathcal{F}^{-1}(\hat{f})(y). \end{aligned}$$

This establishes $\mathcal{F}^{-1}\mathcal{F}(f) = f$.

To prove $\mathcal{F}\mathcal{F}^{-1}(f) = f$, we recall that $\mathcal{F}^{-1}g(x) = \mathcal{F}g(-x)$. Let $\tilde{f}(x) = f(-x)$. By change of variables:

$$\mathcal{F}^{-1}f(x) = \int_{\mathbb{R}^d} f(\xi)e^{i2\pi x\cdot\xi}\,\mathrm{d}\xi = \int_{\mathbb{R}^d} f(-z)e^{-i2\pi z\cdot x}\,\mathrm{d}z = \mathcal{F}\tilde{f}(x).$$

Thus, substituting this structural relation:

$$\mathcal{F}(\mathcal{F}^{-1}(f))(x) = \mathcal{F}(\mathcal{F}\tilde{f})(x) = \mathcal{F}^{-1}(\mathcal{F}\tilde{f})(-x) = \tilde{f}(-x) = f(-(-x)) = f(x).$$

Therefore, $\mathcal{F}\mathcal{F}^{-1}(f)(x) = f(x) = \mathcal{F}^{-1}\mathcal{F}(f)(x)$, completing the proof.

Analysis of the Dominated Convergence Theorem (DCT) Steps

1. *First Use of DCT in the Proof.*

$$\int_{\mathbb{R}^d} f(\delta y)\hat{k}(y)\,\mathrm{d}y \longrightarrow f(0) \quad \text{as } \delta \to 0.$$

   - *Pointwise Convergence:* Since $f \in S$, it is continuous everywhere, meaning $f(\delta y) \to f(0)$ as $\delta \to 0$ for each fixed vector y.
   - *Dominating Function:* Because $f \in S$, it is bounded globally by a constant $M = \|f\|_{L^\infty}$. Thus, $|f(\delta y)\hat{k}(y)| \le M|\hat{k}(y)|$. Since $\hat{k}(y) = e^{-\pi|y|^2}$ is a integrable Gaussian function, it serves as a valid dominator.

2. *Second Use of DCT in the Proof.*

$$\int_{\mathbb{R}^d} \hat{f}(\xi)k(\delta\xi)\,\mathrm{d}\xi \longrightarrow \int_{\mathbb{R}^d} \hat{f}(\xi)\,\mathrm{d}\xi \quad \text{as } \delta \to 0.$$

   - *Pointwise Convergence:* $k(\delta\xi) = e^{-\pi\delta^2|\xi|^2} \to e^0 = 1$ as $\delta \to 0$ for each $\xi \in \mathbb{R}^d$.

- *Dominating Function:* Since k is a Gaussian, $|k(\delta\xi)| \leq 1$ for all δ. This yields the bound $|\hat{f}(\xi)k(\delta\xi)| \leq |\hat{f}(\xi)|$. Since $f \in S$, Proposition 1 ensures that $\hat{f} \in S \subset L^1(\mathbb{R}^d)$, making $\hat{f}$ its own integrable dominating function.

3. *Significance of the Schwartz Space S.*

   - Functions in S decay rapidly: Both f and $\hat{f}$ decrease faster than any inverse polynomial, which natively guarantees absolute integrability over $\mathbb{R}^d$.
   - Standard Gaussians $k(x) = e^{-\pi|x|^2}$ act as perfect smoothing multipliers: they are smooth, self-dual under $\mathcal{F}$, and guarantee convergence margins in integral dualities. Without the rapid polynomial decay inherent to S, uniform integrable boundaries might fail.

$\square$

Plancherel's Equality

For $f \in S$, $\int_{\mathbb{R}^d} |f(x)|^2 \,\mathrm{d}x = \int_{\mathbb{R}^d} |\hat{f}(\xi)|^2 \,\mathrm{d}\xi$, so $\|f\|_{L^2} = \|\hat{f}\|_{L^2}$ where $L^2 = L^2(\mathbb{R}^d)$.

Idea of the Proof

Use the property $f(0) = \int_{\mathbb{R}^d} \hat{f}(\xi) \,\mathrm{d}\xi$ and the Fourier transform of the convolution. Let $(f * g)(x) = \int_{\mathbb{R}^d} f(x-y)g(y) \,\mathrm{d}y$ and $\mathcal{F}(f * g)(\xi) = \hat{f}(\xi)\hat{g}(\xi)$. Applying the inversion formula at $x = 0$, we obtain:

$$(f * g)(0) = \int_{\mathbb{R}^d} g(y)f(-y) \,\mathrm{d}y = \int_{\mathbb{R}^d} \mathcal{F}(f * g)(\xi) \,\mathrm{d}\xi = \int_{\mathbb{R}^d} \hat{f}(\xi)\hat{g}(\xi) \,\mathrm{d}\xi \quad (\#).$$

In (#) above, we choose an appropriate substitution for f and g to handle the reflection $f(-y)$ and introduce the complex conjugate.

Proposition 4 (Plancherel's Equality)

For any $f, g \in S$,

$$\int_{\mathbb{R}^d} g(y)\overline{f(y)} \,\mathrm{d}y = \int_{\mathbb{R}^d} \hat{g}(\xi)\overline{\hat{f}(\xi)} \,\mathrm{d}\xi.$$

Thus, $\langle g, f\rangle_{L^2} = \langle \hat{g}, \hat{f}\rangle_{L^2}$. In particular, $\|f\|_{L^2} = \|\hat{f}\|_{L^2}$ when $f = g$.

Proof. Let $f, g \in S$. Define $\tilde{f}(y) = \overline{f(-y)}$ for $y \in \mathbb{R}^d$, and note that $\tilde{f} \in S$. By identity (#), we have:

$$\int_{\mathbb{R}^d} g(y)\overline{f(y)} \,\mathrm{d}y = \int_{\mathbb{R}^d} g(y)\tilde{f}(-y) \,\mathrm{d}y = \int_{\mathbb{R}^d} \hat{g}(\xi)\mathcal{F}(\tilde{f})(\xi) \,\mathrm{d}\xi.$$

Now, we compute $\mathcal{F}(\tilde{f})(\xi)$:

$$\begin{aligned}\mathcal{F}(\tilde{f})(\xi) &= \int_{\mathbb{R}^d} \overline{f(-y)} e^{-i2\pi y\cdot\xi}\,\mathrm{d}y \\ &= \int_{\mathbb{R}^d} \overline{f(x)} e^{i2\pi x\cdot\xi}\,\mathrm{d}x \quad \text{(substituting } x=-y,\ \mathrm{d}x=\mathrm{d}y,\ |J|=1) \\ &= \overline{\int_{\mathbb{R}^d} f(x) e^{-i2\pi x\cdot\xi}\,\mathrm{d}x} = \overline{\hat{f}(\xi)}.\end{aligned}$$

Hence, $\int_{\mathbb{R}^d} g(y)\overline{f(y)}\,\mathrm{d}y = \int_{\mathbb{R}^d} \hat{g}(\xi)\overline{\hat{f}(\xi)}\,\mathrm{d}\xi$, which implies $\langle g, f\rangle_{L^2} = \langle \hat{g}, \hat{f}\rangle_{L^2}$. □

Remark. The operators $\mathcal{F}$ and $\mathcal{F}^{-1}$ can be uniquely extended to unitary maps on $L^2(\mathbb{R}^d)$. For any $f \in L^2$, there exists a sequence $f_n \in S$ such that $f_n \to f$ in L^2, since $C_c^\infty \subseteq S$ and C_c^∞ is dense in L^2. By Proposition 4, if $\|f_n - f_m\|_{L^2} \to 0$ as $m, n \to \infty$, then:

$$\|f_n - f_m\|_{L^2} = \|\hat{f}_n - \hat{f}_m\|_{L^2} = \|\mathcal{F}^{-1}(f_n) - \mathcal{F}^{-1}(f_m)\|_{L^2}.$$

This allows us to define $\mathcal{F}f = \lim_{n\to\infty} \mathcal{F}(f_n)$ and $\mathcal{F}^{-1}f = \lim_{n\to\infty} \mathcal{F}^{-1}(f_n)$ via Cauchy completeness. This definition is independent of the chosen sequence. Indeed, if $g_n \in S$ is another sequence with $g_n \to f$ in L^2, then $\|g_n - f_n\|_{L^2} \to 0$ as $n \to \infty$. By Proposition 4,

$$\|g_n - f_n\|_{L^2} = \|\hat{g}_n - \hat{f}_n\|_{L^2} = \|\mathcal{F}^{-1}g_n - \mathcal{F}^{-1}f_n\|_{L^2} \to 0,$$

which implies $\hat{g}_n \to \mathcal{F}f$ and $\mathcal{F}^{-1}g_n \to \mathcal{F}^{-1}f$ in L^2. Hence, $\mathcal{F} : L^2 \to L^2$ is a unitary operator satisfying $\mathcal{F}^* = \mathcal{F}^{-1}$.

Corollary 1

The maps $\mathcal{F}, \mathcal{F}^{-1} : S \to S$ uniquely extend to bounded linear operators $\mathcal{F}, \mathcal{F}^{-1} : L^2 \to L^2$ via:

$$\mathcal{F}f = \hat{f} = \lim_{n\to\infty} \mathcal{F}f_n, \quad \mathcal{F}^{-1}f = \lim_{n\to\infty} \mathcal{F}^{-1}f_n \quad (\text{in } L^2),$$

where $f_n \in S$ and $f_n \to f$ in L^2. Furthermore, $\mathcal{F}\mathcal{F}^{-1}f = \mathcal{F}^{-1}\mathcal{F}f = f$ and

$$\|\mathcal{F}f\|_{L^2} = \|\mathcal{F}^{-1}f\|_{L^2} = \|f\|_{L^2} \quad \text{for all } f \in L^2.$$

Proposition 6

Let $f \in L^1 \cap L^2$. Let $\hat{f} = \lim_{n\to\infty} \hat{f}_n$ and $\mathcal{F}^{-1}f = \lim_{n\to\infty} \mathcal{F}^{-1}f_n$ in L^2, where $f_n \in S$ and $f_n \to f$ in L^2. Then the standard integral representations hold pointwise almost everywhere:

$$\hat{f}(\xi) = \int_{\mathbb{R}^d} f(x)e^{-i2\pi x\cdot\xi}\,\mathrm{d}x \quad \text{and} \quad \mathcal{F}^{-1}f(x) = \int_{\mathbb{R}^d} f(\xi)e^{i2\pi x\cdot\xi}\,\mathrm{d}\xi.$$

Proof. By standard properties of the Fourier transform on S, for any test function $h \in S$, we have:

$$\int_{\mathbb{R}^d} \hat{f}_n h\,\mathrm{d}x = \int_{\mathbb{R}^d} f_n\hat{h}\,\mathrm{d}x \quad \text{and} \quad \int_{\mathbb{R}^d} \mathcal{F}^{-1}(f_n)h\,\mathrm{d}x = \int_{\mathbb{R}^d} f_n\mathcal{F}^{-1}(h)\,\mathrm{d}x.$$

Taking the limit as $n \to \infty$, L^2 convergence implies weak convergence against $h \in S$:

$$\int_{\mathbb{R}^d} \hat{f}_n h \to \int_{\mathbb{R}^d} \hat{f} h, \quad \int_{\mathbb{R}^d} f_n \hat{h} \to \int_{\mathbb{R}^d} f \hat{h},$$

yielding $\int_{\mathbb{R}^d} \hat{f} h = \int_{\mathbb{R}^d} f \hat{h}$ (and similarly for $\mathcal{F}^{-1}$). Now, for any $h \in S$:

$$\begin{aligned}
\int_{\mathbb{R}^d} \hat{f}(y) h(y)\,\mathrm{d}y &= \int_{\mathbb{R}^d} f(x) \hat{h}(x)\,\mathrm{d}x \\
&= \int_{\mathbb{R}^d} f(x) \left(\int_{\mathbb{R}^d} h(y) e^{-i2\pi x \cdot y}\,\mathrm{d}y \right) \mathrm{d}x \\
&= \int_{\mathbb{R}^d} \left(\int_{\mathbb{R}^d} f(x) e^{-i2\pi x \cdot y}\,\mathrm{d}x \right) h(y)\,\mathrm{d}y \quad \text{(by Fubini's Theorem).}
\end{aligned}$$

Since this holds for all $h \in S$, by the fundamental lemma of the calculus of variations, we conclude:

$$\hat{f}(y) = \int_{\mathbb{R}^d} f(x) e^{-i2\pi x \cdot y}\,\mathrm{d}x \quad \text{a.e.}$$

Fubini's theorem is justified since $\int_{\mathbb{R}^d} \int_{\mathbb{R}^d} |f(x)| \cdot |h(y)|\,\mathrm{d}x\,\mathrm{d}y = \|f\|_{L^1} \|h\|_{L^1} < \infty$. An analogous proof holds for $\mathcal{F}^{-1}$. □

Corollary 2

Let $f \in L^2(\mathbb{R}^d)$. Then the Fourier transform can be evaluated as a principal value limit:

$$\begin{aligned}
\mathcal{F} f(\xi) = \hat{f}(\xi) &= \lim_{n \to \infty} \int_{|x| \le n} f(x) e^{-i2\pi x \cdot \xi}\,\mathrm{d}x \quad (\text{in } L^2), \\
\mathcal{F}^{-1} f(x) &= \lim_{n \to \infty} \int_{|\xi| \le n} f(\xi) e^{i2\pi x \cdot \xi}\,\mathrm{d}\xi \quad (\text{in } L^2).
\end{aligned}$$

Proof. Let $f \in L^2(\mathbb{R}^d)$ and define $f_n = \chi_{\{|x| \le n\}} f$. By Hölder's inequality, $\int_{\mathbb{R}^d} |f_n| \le \|f\|_{L^2} \|\chi_{\{|x| \le n\}}\|_{L^2} < \infty$, so $f_n \in L^1 \cap L^2$. Furthermore, $f_n \to f$ in L^2 by the Dominated Convergence Theorem, since $|f_n| \le |f|$ and $|f|^2 \in L^1$.

By Proposition 6, since $f_n \in L^1 \cap L^2$:

$$\hat{f}_n(\xi) = \int_{\mathbb{R}^d} f_n(x) e^{-i2\pi x \cdot \xi}\,\mathrm{d}x = \int_{|x| \le n} f(x) e^{-i2\pi x \cdot \xi}\,\mathrm{d}x.$$

Since $\mathcal{F}$ is unitary on L^2, $\|\hat{f}_n - \hat{f}\|_{L^2} = \|f_n - f\|_{L^2} \to 0$, meaning $\hat{f}_n \to \hat{f}$ in L^2. Hence,

$$\hat{f}(\xi) = \lim_{n \to \infty} \int_{|x| \le n} f(x) e^{-i2\pi x \cdot \xi}\,\mathrm{d}x \quad \text{in } L^2.$$

An analogous argument using $g_n = \chi_{\{|\xi| \le n\}} f$ yields the formula for $\mathcal{F}^{-1}$. □

Some Applications of the Fourier Transform

1. Heisenberg Uncertainty Principle

Let $\psi \in S(\mathbb{R})$ be a normalized wave function, satisfying $\int_{\mathbb{R}} |\psi(x)|^2 \, dx = 1$. In quantum mechanics, $|\psi(x)|^2$ represents the probability density function (pdf) of a particle's position. By Plancherel's theorem:

$$\int_{\mathbb{R}} |\hat{\psi}(\xi)|^2 \, d\xi = \int_{\mathbb{R}} |\psi(x)|^2 \, dx = 1,$$

meaning $|\hat{\psi}(\xi)|^2$ serves as the pdf of the particle's momentum. If x_0 and ξ_0 denote the expected position and momentum respectively, their corresponding variances (mean square errors) are:

$$(\Delta x)^2 = \int_{\mathbb{R}} |x - x_0|^2 \, |\psi(x)|^2 \, dx, \quad (\Delta \xi)^2 = \int_{\mathbb{R}} |\xi - \xi_0|^2 \, |\hat{\psi}(\xi)|^2 \, d\xi.$$

Theorem 2 (Heisenberg Uncertainty Principle).

Let $\psi \in S(\mathbb{R})$ with $\int_{\mathbb{R}} |\psi(x)|^2 \, dx = 1$. Then:

$$(\Delta x)^2 (\Delta \xi)^2 \geq \frac{1}{16\pi^2}.$$

Proof. First, consider the case where $x_0 = \xi_0 = 0$, giving $(\Delta x)^2 = \int_{\mathbb{R}} x^2 |\psi(x)|^2 \, dx$ and $(\Delta \xi)^2 = \int_{\mathbb{R}} \xi^2 |\hat{\psi}(\xi)|^2 \, d\xi$. Since $\psi \in S(\mathbb{R})$, it decays faster than any polynomial, implying $x|\psi(x)|^2 \to 0$ as $|x| \to \infty$. Integration by parts yields:

$$\begin{aligned} 1 = \int_{-\infty}^{\infty} |\psi(x)|^2 \, dx &= \Big[x|\psi(x)|^2 \Big]_{-\infty}^{\infty} - \int_{-\infty}^{\infty} x \big(\psi'(x) \overline{\psi(x)} + \psi(x) \overline{\psi'(x)} \big) \, dx \\ &= -2 \, \mathrm{Re} \int_{-\infty}^{\infty} x \, \psi'(x) \, \overline{\psi(x)} \, dx. \end{aligned}$$

Applying the Cauchy–Schwarz inequality:

$$\begin{aligned} 1 &\leq 2 \int_{-\infty}^{\infty} |x \, \psi(x)| \, |\psi'(x)| \, dx \\ &\leq 2 \left(\int_{-\infty}^{\infty} x^2 |\psi(x)|^2 \, dx \right)^{1/2} \left(\int_{-\infty}^{\infty} |\psi'(x)|^2 \, dx \right)^{1/2} \\ &= 2 \, (\Delta x) \left(\int_{-\infty}^{\infty} |\psi'|^2 \, dx \right)^{1/2}. \end{aligned}$$

By Plancherel's equality and derivative properties, $\widehat{\psi'}(\xi) = i2\pi\xi \, \hat{\psi}(\xi)$, which implies:

$$\int_{-\infty}^{\infty} |\psi'(x)|^2 \, dx = \int_{-\infty}^{\infty} |\widehat{\psi'}(\xi)|^2 \, d\xi = \int_{-\infty}^{\infty} (2\pi|\xi|)^2 |\hat{\psi}(\xi)|^2 \, d\xi = 4\pi^2 (\Delta \xi)^2.$$

Combining these gives $1 \leq 2(\Delta x) \cdot 2\pi(\Delta\xi) = 4\pi(\Delta x)(\Delta\xi)$, squaring both sides results in:

$$(\Delta x)^2(\Delta\xi)^2 \geq \frac{1}{16\pi^2}.$$

For general $x_0, \xi_0 \in \mathbb{R}$, apply the above result to the shifted wave function $\varphi(x) = e^{-i2\pi\xi_0 x}\psi(x + x_0)$. It easily follows that $\int_{\mathbb{R}} |\varphi|^2 = 1$ and $|\hat{\varphi}(\xi)|^2 = |\hat{\psi}(\xi + \xi_0)|^2$. A standard change of variables confirms that the variances transform accordingly, completing the proof. $\square$

2. Helmholtz Equation in $S(\mathbb{R}^d)$

Consider the Helmholtz-type equation:

$$u(x) - \Delta u(x) = f(x), \quad x \in \mathbb{R}^d,\ f \in S(\mathbb{R}^d), \tag{H}$$

where $\Delta = \sum_{j=1}^{d} \frac{\partial^2}{\partial x_j^2}$ denotes the Laplacian operator. Taking the Fourier transform gives:

$$\widehat{\Delta u}(\xi) = -4\pi^2|\xi|^2\, \hat{u}(\xi).$$

Proposition 7

For any $f \in S(\mathbb{R}^d)$, there exists a unique solution $u \in S(\mathbb{R}^d)$ solving (H). That is, the operator $(I - \Delta) : S \to S$ is a bijection.

Proof. Let $f \in S$. Then $\hat{f} \in S$. Since the polynomial $1 + 4\pi^2|\xi|^2$ is smooth, bounded below by 1, and has bounded derivatives of all orders, its reciprocal $\frac{1}{1+4\pi^2|\xi|^2}$ serves as a multiplier on S. Thus, $\frac{\hat{f}(\xi)}{1+4\pi^2|\xi|^2} \in S$.

(1) Uniqueness. Suppose $u \in S$ satisfies $u - \Delta u = f$. Taking the Fourier transform on both sides:

$$\widehat{u - \Delta u}(\xi) = \hat{f}(\xi) \implies (1 + 4\pi^2|\xi|^2)\, \hat{u}(\xi) = \hat{f}(\xi) \implies \hat{u}(\xi) = \frac{\hat{f}(\xi)}{1 + 4\pi^2|\xi|^2}.$$

Thus $\hat{u}$ is uniquely determined. Since $\mathcal{F} : S \to S$ is bijective, u is unique.

(2) Existence. Given $f \in S$, define $g(\xi) = \frac{\hat{f}(\xi)}{1+4\pi^2|\xi|^2} \in S$. Setting $u(x) = (\mathcal{F}^{-1}g)(x) \in S$, we check that it solves (H):

$$\widehat{u - \Delta u}(\xi) = (1 + 4\pi^2|\xi|^2)\, \hat{u}(\xi) = (1 + 4\pi^2|\xi|^2)\, g(\xi) = \hat{f}(\xi).$$

By the injectivity of $\mathcal{F}$, we conclude $u - \Delta u = f$. $\square$

Tempered Distributions

Fundamental Definitions

Notation. Let $\mathcal{S}' = \mathcal{S}'(\mathbb{R}^d)$ denote the dual space containing all linear continuous maps $T : \mathcal{S}(\mathbb{R}^d) \to \mathbb{C}$. The elements of $\mathcal{S}'$ are called *tempered distributions*, and $\mathcal{S}$ represents the Schwartz space of rapidly decreasing functions.

Continuity in $\mathcal{S}$

A linear map $T : \mathcal{S} \to \mathbb{C}$ is continuous if and only if it is bounded with respect to the topology generated by the Schwartz semi-norms. Specifically, there exist constants $m \geq 1$, $C \geq 0$, multi-indices $\alpha_1, \ldots, \alpha_m \in \mathbb{N}_0^d$, and integers $N_1, \ldots, N_m \in \mathbb{N}_0$ such that:

$$|T(f)| \leq C \sum_{k=1}^{m} |f|_{N_k,\alpha_k} \quad \forall f \in \mathcal{S}.$$

Equivalently, using the directed system of semi-norms $|f|_{(N)} = \max_{|\alpha|,\beta \leq N} \sup_x |x^\beta \partial^\alpha f(x)|$, continuity is satisfied if there exist $N \in \mathbb{N}_0$ and $C \geq 0$ such that:

$$|T(f)| \leq C|f|_{(N)} \quad \forall f \in \mathcal{S}.$$

Convergence in $\mathcal{S}'$

Let $\{T_n\}_{n=1}^{\infty} \subset \mathcal{S}'$ and $T \in \mathcal{S}'$. We say that $T_n \to T$ weakly-* in $\mathcal{S}'$ if:

$$T_n(f) \to T(f) \quad \forall f \in \mathcal{S}.$$

Example: The Space of Smooth Functions of Slow Growth ($\mathcal{O}_M$)

Let $\mathcal{O}_M(\mathbb{R}^d)$ be the space of all smooth functions $g \in C^\infty(\mathbb{R}^d)$ whose derivatives grow at most polynomially. That is, for every multi-index $\alpha \in \mathbb{N}_0^d$, there exist constants $C = C(\alpha) \geq 0$ and $N = N(\alpha) \in \mathbb{N}_0$ such that:

$$|\partial^\alpha g(x)| \leq C(1 + |x|)^N \quad \forall x \in \mathbb{R}^d.$$

a. *Regular Distribution Mapping.* Every $g \in \mathcal{O}_M$ induces a well-defined regular tempered distribution $T_g \in \mathcal{S}'$ via the canonical pairing:

$$T_g(f) = \int_{\mathbb{R}^d} g(x)f(x)\,\mathrm{d}x, \quad f \in \mathcal{S}.$$

Proof. Let $\alpha = 0$. By definition, there exist $C(0) \geq 0$ and $N \in \mathbb{N}_0$ such that $|g(x)| \leq C(0)(1 + |x|)^N$. For any $f \in \mathcal{S}$, we can ensure convergence by controlling the polynomial growth with a sufficiently large algebraic weight:

$$\begin{aligned}
\int_{\mathbb{R}^d} |g(x)| \cdot |f(x)|\,\mathrm{d}x &\leq C(0) \int_{\mathbb{R}^d} (1 + |x|)^N |f(x)|\,\mathrm{d}x \\
&= C(0) \int_{\mathbb{R}^d} (1 + |x|)^{N+d+1} |f(x)| \cdot \frac{\mathrm{d}x}{(1 + |x|)^{d+1}} \\
&\leq C(0)\,|f|_{N+d+1,\,0} \int_{\mathbb{R}^d} \frac{\mathrm{d}x}{(1 + |x|)^{d+1}}.
\end{aligned}$$

Setting $C = C(0) \int_{\mathbb{R}^d} (1 + |x|)^{-(d+1)}\,\mathrm{d}x < \infty$, we obtain:

$$|T_g(f)| \leq \left| \int_{\mathbb{R}^d} g(x)f(x)\,\mathrm{d}x \right| \leq C\,|f|_{N+d+1,\,0}.$$

This verifies that T_g is bounded relative to the Schwartz semi-norms, so $T_g \in \mathcal{S}'$. $\square$

b. *Injectivity.* The linear mapping $\mathcal{O}_M \ni g \mapsto T_g \in \mathcal{S}'$ is injective.

Proof. Let $g_1, g_2 \in \mathcal{O}_M$ such that $T_{g_1}(f) = T_{g_2}(f)$ for all $f \in \mathcal{S}$. It follows that $\int_{\mathbb{R}^d}(g_1(x) - g_2(x))f(x)\,\mathrm{d}x = 0$. Since $C_c^\infty(\mathbb{R}^d) \subset \mathcal{S}$ is dense in $L^1_{\mathrm{loc}}(\mathbb{R}^d)$, and smooth functions are locally integrable, it follows by the fundamental lemma of the calculus of variations that $g_1 = g_2$ almost everywhere. By continuity, $g_1(x) = g_2(x)$ identically for all $x \in \mathbb{R}^d$. $\square$

c. *Action on Derivatives.* Let $\alpha \in \mathbb{N}_0^d$. If $g \in \mathcal{O}_M$, then $\partial^\alpha g \in \mathcal{O}_M$. Since functions $f \in \mathcal{S}$ decay rapidly at infinity, boundary terms vanish under integration by parts, yielding:

$$\langle \partial^\alpha g, f\rangle = \int_{\mathbb{R}^d} \partial^\alpha g(x) \cdot f(x)\,\mathrm{d}x = (-1)^{|\alpha|} \int_{\mathbb{R}^d} g(x) \cdot \partial^\alpha f(x)\,\mathrm{d}x = (-1)^{|\alpha|} T_g(\partial^\alpha f).$$

Summary of Embeddings. We establish the nested chain of topological vector spaces:

$$C_c^\infty(\mathbb{R}^d) \subseteq \mathcal{S}(\mathbb{R}^d) \subseteq \mathcal{O}_M(\mathbb{R}^d) \subseteq \mathcal{S}'(\mathbb{R}^d).$$

Moving forward, we simplify notation by writing $T_g \equiv g$ and $T_g(f) \equiv \langle g, f\rangle$.

Polynomial growth

A measurable function $g : \mathbb{R}^d \to \mathbb{C}$ is of *polynomial growth* if there exist constants $C > 0$ and $N \geq 1$ such that:

$$|g(x)| \leq C(1 + |x|)^N \quad \forall x \in \mathbb{R}^d.$$

Remark. Any such function defines a regular tempered distribution $T_g \in \mathcal{S}'$ with the continuous control bound given by $|T_g(f)| \leq C|f|_{N+d+1,\,0}$.

Example: Functions of Polynomial Growth & Distributional Derivatives

Using duality, for any general tempered distribution $T \in \mathcal{S}'$, we define its *distributional derivative* $\partial^\alpha T \in \mathcal{S}'$ by shifting the differential operator onto the test function:

$$(\partial^\alpha T)(f) = (-1)^{|\alpha|} T(\partial^\alpha f) \quad \forall f \in \mathcal{S}. \tag{\#}$$

Formally, as an operator composition, this reads:

$$\partial^\alpha T = (-1)^{|\alpha|} T \circ \partial^\alpha, \quad \mathcal{S} \xrightarrow{\partial^\alpha} \mathcal{S} \xrightarrow{T} \mathbb{C}.$$

The Ramp Functions. Consider the one-dimensional ramp function $g : \mathbb{R} \to \mathbb{R}$:

$$g(x) = \max(0, x) = \begin{cases} x, & x \geq 0 \\ 0, & x < 0 \end{cases}$$

a. *First Distributional Derivative (g').* For any $f \in \mathcal{S}(\mathbb{R})$, applying definition (#) yields:

$$\begin{aligned}\left\langle \frac{\mathrm{d}}{\mathrm{d}x} g, f \right\rangle &= -\int_{-\infty}^{\infty} g(x) f'(x)\,\mathrm{d}x = -\int_{0}^{\infty} x f'(x)\,\mathrm{d}x \\ &= -\Big[x f(x)\Big]_0^{\infty} + \int_0^{\infty} f(x)\,\mathrm{d}x.\end{aligned}$$

The boundary term vanishes at ∞ because $f \in \mathcal{S}$ decays faster than any polynomial. Thus:

$$\left\langle \frac{\mathrm{d}}{\mathrm{d}x} g, f \right\rangle = \int_0^{\infty} f(x)\,\mathrm{d}x = \int_{-\infty}^{\infty} H(x) f(x)\,\mathrm{d}x,$$

where $H(x)$ is the Heaviside step function. Consequently, $g' = H$ in the distributional sense:

$$H(x) = \begin{cases} 1, & x \geq 0 \\ 0, & x < 0 \end{cases}$$

b. *Second Distributional Derivative (H').* Differentiating the Heaviside distribution again:

$$\begin{aligned}\left\langle \frac{\mathrm{d}H}{\mathrm{d}x}, f \right\rangle &= -\int_{-\infty}^{\infty} H(x) f'(x)\,\mathrm{d}x = -\int_{0}^{\infty} f'(x)\,\mathrm{d}x \\ &= -\Big[f(x)\Big]_0^{\infty} = f(0).\end{aligned}$$

This maps to the evaluation at the origin, which corresponds to the Dirac delta distribution δ_0:

$$\langle \delta_0, f \rangle = f(0) \implies \frac{\mathrm{d}H}{\mathrm{d}x} = \delta_0.$$

Formally, this relation highlights the link between distributional derivatives and Lebesgue-Stieltjes integration: $\int_{-\infty}^{\infty} f(x) H'(x)\,\mathrm{d}x = \int_{-\infty}^{\infty} f(x)\,\mathrm{d}H(x)$.

Example: L^p Spaces as Tempered Distributions

Let $g \in L^p(\mathbb{R}^d)$ with $p \geq 1$. Then g induces a regular tempered distribution $T_g \in \mathcal{S}'$.

Proof. Let q be the conjugate exponent matching p ($\frac{1}{p} + \frac{1}{q} = 1$). Applying Hölder's inequality:

$$\int_{\mathbb{R}^d} |g(x)| \cdot |f(x)|\,\mathrm{d}x \leq \|g\|_{L^p} \|f\|_{L^q}.$$

To dominate the L^q norm of the test function $f \in \mathcal{S}$ using standard semi-norms, choose an integer index $m = \lceil \frac{d+1}{q} \rceil$. This selection ensures polynomial convergence:

$$\begin{aligned}\|f\|_{L^q}^q &= \int_{\mathbb{R}^d} |f(x)|^q\,\mathrm{d}x \\ &= \int_{\mathbb{R}^d} \left(|f(x)|(1+|x|)^{\frac{d+1}{q}}\right)^q \cdot \frac{\mathrm{d}x}{(1+|x|)^{d+1}} \\ &\leq |f|_{m,0}^q \int_{\mathbb{R}^d} \frac{\mathrm{d}x}{(1+|x|)^{d+1}}.\end{aligned}$$

Since $\int_{\mathbb{R}^d}(1+|x|)^{-(d+1)}\,\mathrm{d}x = C < \infty$, taking the q-th root gives $\|f\|_{L^q} \le C^{1/q}|f|_{m,0}$. Combining these results:

$$|T_g(f)| \le \|g\|_{L^p}\|f\|_{L^q} \le C^{1/q}\,\|g\|_{L^p}\,|f|_{m,0} < \infty.$$

Thus, T_g is a continuous linear functional on $\mathcal{S}$, meaning $L^p(\mathbb{R}^d) \subset \mathcal{S}'(\mathbb{R}^d)$. Injectivity follows via a standard density argument analogous to the one used in part b of the preceding example. □

We embed $L^p(\mathbb{R}^d) \subseteq \mathcal{S}'(\mathbb{R}^d)$ for $p \ge 1$ by identifying $g \in L^p$ with its regular tempered distribution $T_g \equiv g$, defined via:

$$\langle g, f\rangle = \int_{\mathbb{R}^d} g(x)f(x)\,\mathrm{d}x, \quad f \in \mathcal{S}(\mathbb{R}^d).$$

In particular, $L^2(\mathbb{R}^d) \subseteq \mathcal{S}'(\mathbb{R}^d)$. Recall that if $g \in L^2$, its Fourier transform $\mathcal{F}g = \hat{g}$ and inverse Fourier transform $\mathcal{F}^{-1}g$ also belong to L^2. By the Parseval-Plancherel identities on $\mathcal{S}$, we have:

$$\int_{\mathbb{R}^d} \hat{g}(x)f(x)\,\mathrm{d}x = \int_{\mathbb{R}^d} g(x)\hat{f}(x)\,\mathrm{d}x, \quad \int_{\mathbb{R}^d}(\mathcal{F}^{-1}g)(x)f(x)\,\mathrm{d}x = \int_{\mathbb{R}^d} g(x)(\mathcal{F}^{-1}f)(x)\,\mathrm{d}x,$$

which holds by density for all $f \in L^2$. This motivates the general definitions of basic operations on $\mathcal{S}'(\mathbb{R}^d)$ via duality.

Main Operations on $\mathcal{S}'(\mathbb{R}^d)$

1. *Differentiation.* Given $T \in \mathcal{S}'(\mathbb{R}^d)$ and a multi-index $\alpha \in \mathbb{N}_0^d$, the distributional derivative $\partial^\alpha T \in \mathcal{S}'(\mathbb{R}^d)$ is defined by:

$$\langle \partial^\alpha T, f\rangle = (-1)^{|\alpha|}\langle T, \partial^\alpha f\rangle \quad \forall f \in \mathcal{S}(\mathbb{R}^d).$$

2. *Fourier Transform.* For any $T \in \mathcal{S}'(\mathbb{R}^d)$, we define its Fourier transform $\mathcal{F}T = \hat{T} \in \mathcal{S}'(\mathbb{R}^d)$ and inverse Fourier transform $\mathcal{F}^{-1}T \in \mathcal{S}'(\mathbb{R}^d)$ via:

$$\langle \mathcal{F}T, f\rangle = \langle T, \mathcal{F}f\rangle, \quad \langle \mathcal{F}^{-1}T, f\rangle = \langle T, \mathcal{F}^{-1}f\rangle \quad \forall f \in \mathcal{S}(\mathbb{R}^d).$$

   Note. Operationally, $\mathcal{F}T = T \circ \mathcal{F}$, corresponding to the composition $\mathcal{S} \xrightarrow{\mathcal{F}} \mathcal{S} \xrightarrow{T} \mathbb{C}$.

3. *Multiplication by Smooth Functions of Polynomial Growth ($\mathcal{O}_M$).* Given $T \in \mathcal{S}'(\mathbb{R}^d)$ and $F \in \mathcal{O}_M(\mathbb{R}^d)$, the product $FT \in \mathcal{S}'(\mathbb{R}^d)$ is defined by:

$$\langle FT, f\rangle = \langle T, Ff\rangle \quad \forall f \in \mathcal{S}(\mathbb{R}^d),$$

   which is well-defined since the map $f \mapsto Ff$ is continuous from $\mathcal{S}$ into itself.

The Space of Moderate Measures ($\mathcal{M}_{\mathrm{mod}}$)

Let $\mathcal{M}_{\mathrm{mod}}(\mathbb{R}^d)$ denote the space of signed (or complex) Borel measures μ on $\mathbb{R}^d$ of *moderate growth*, meaning there exists an integer $N \geq 1$ such that:

$$\int_{\mathbb{R}^d} (1+|x|)^{-N}\,\mathrm{d}|\mu|(x) < \infty. \tag{\# mod}$$

The Lebesgue measure on $\mathbb{R}^d$ belongs to $\mathcal{M}_{\mathrm{mod}}$ since (# mod) converges for any integer $N \geq d+1$. Every $\mu \in \mathcal{M}_{\mathrm{mod}}$ induces a linear functional $T_\mu \in \mathcal{S}'$ via integration:

$$T_\mu(f) = \int_{\mathbb{R}^d} f(x)\,\mathrm{d}\mu(x), \quad f \in \mathcal{S}.$$

T_μ is a continuous linear functional on $\mathcal{S}$.

Proof. Linearity is immediate. It remains to prove continuity. Let $f \in \mathcal{S}$. Using the moderate growth bound (# mod):

$$\int_{\mathbb{R}^d} |f(x)|\,\mathrm{d}|\mu|(x) = \int_{\mathbb{R}^d} |f(x)|(1+|x|)^N \cdot (1+|x|)^{-N}\,\mathrm{d}|\mu|(x) \leq |f|_{N,0} \int_{\mathbb{R}^d} (1+|x|)^{-N}\,\mathrm{d}|\mu|(x).$$

Setting $C = \int_{\mathbb{R}^d} (1+|x|)^{-N}\,\mathrm{d}|\mu|(x) < \infty$, we obtain $|T_\mu(f)| \leq C|f|_{N,0}$. This verifies that $T_\mu \in \mathcal{S}'$. $\square$

The mapping $\mu \mapsto T_\mu$ is injective, allowing the identification $\mathcal{M}_{\mathrm{mod}} \subseteq \mathcal{S}'$. We write $\langle \mu, f\rangle = \int_{\mathbb{R}^d} f\,\mathrm{d}\mu$. Its operational extensions to derivatives and Fourier transforms follow by duality:

$$\langle \partial^\alpha \mu, f\rangle = (-1)^{|\alpha|}\langle \mu, \partial^\alpha f\rangle, \quad \langle \mathcal{F}\mu, f\rangle = \langle \mu, \mathcal{F}f\rangle, \quad \langle \mathcal{F}^{-1}\mu, f\rangle = \langle \mu, \mathcal{F}^{-1}f\rangle.$$

Now assume that μ is a finite signed Borel measure. By Fubini's theorem, we can compute its inverse Fourier transform explicitly:

$$\begin{aligned}
\langle \mathcal{F}^{-1}\mu, f\rangle = \langle \mu, \mathcal{F}^{-1}f\rangle &= \int_{\mathbb{R}^d} (\mathcal{F}^{-1}f)(x)\,\mathrm{d}\mu(x) \\
&= \int_{\mathbb{R}^d} \left(\int_{\mathbb{R}^d} f(\xi)e^{i2\pi x\cdot\xi}\,\mathrm{d}\xi\right)\mathrm{d}\mu(x) \\
&= \int_{\mathbb{R}^d} f(\xi)\left(\int_{\mathbb{R}^d} e^{i2\pi x\cdot\xi}\,\mathrm{d}\mu(x)\right)\mathrm{d}\xi = \int_{\mathbb{R}^d} f(\xi)\phi(\xi)\,\mathrm{d}\xi,
\end{aligned}$$

where $\phi(\xi) = \int_{\mathbb{R}^d} e^{i2\pi x\cdot\xi}\,\mathrm{d}\mu(x) = \mathbb{E}\big[e^{i2\pi X\cdot\xi}\big]$ is the characteristic function of the measure μ (equivalently, its Fourier–Stieltjes transform). Therefore, in the sense of $\mathcal{S}'$ we have $\mathcal{F}^{-1}\mu = \phi$, so the inverse transform is the regular distribution induced by the bounded continuous function ϕ.

Proposition 1

The operators $\mathcal{F}, \mathcal{F}^{-1} : \mathcal{S}'(\mathbb{R}^d) \to \mathcal{S}'(\mathbb{R}^d)$ are linear bijections.

Proof. By the inversion theorem on $\mathcal{S}$, we know $\mathcal{F}\mathcal{F}^{-1} = \mathcal{F}^{-1}\mathcal{F} = \mathrm{Id}_{\mathcal{S}}$. Testing the distributional composition on an arbitrary $f \in \mathcal{S}$:

$$\langle \mathcal{F}^{-1}\mathcal{F}T, f\rangle = \langle \mathcal{F}T, \mathcal{F}^{-1}f\rangle = \langle T, \mathcal{F}\mathcal{F}^{-1}f\rangle = \langle T, f\rangle.$$

Similarly, $\langle \mathcal{F}\mathcal{F}^{-1}T, f\rangle = \langle T, f\rangle$. Thus $\mathcal{F}^{-1}\mathcal{F}T = \mathcal{F}\mathcal{F}^{-1}T = T$, completing the proof. $\square$

Derivative-Polynomial Duality

Let $T \in \mathcal{S}'(\mathbb{R}^d)$ and $\alpha \in \mathbb{N}_0^d$. Prove that $\widehat{\partial^\alpha T} = (i2\pi\xi)^\alpha \hat{T}$.

Proof. Since $(i2\pi\xi)^\alpha$ has smooth coefficients growing polynomially, it belongs to $\mathcal{O}_M$. For any $f \in \mathcal{S}$:

$$\langle \widehat{\partial^\alpha T}, f\rangle = \langle \partial^\alpha T, \hat{f}\rangle = (-1)^{|\alpha|}\langle T, \partial^\alpha \hat{f}\rangle.$$

Using the standard identity $\partial^\alpha \hat{f} = \mathcal{F}\big((-i2\pi x)^\alpha f\big)$, we set $g(x) = (-i2\pi x)^\alpha f(x) \in \mathcal{S}$. Continuing the evaluation:

$$\begin{aligned}(-1)^{|\alpha|}\langle T, \hat{g}\rangle = (-1)^{|\alpha|}\langle \hat{T}, g\rangle &= (-1)^{|\alpha|}\int_{\mathbb{R}^d} \hat{T}(\xi)(-i2\pi\xi)^\alpha f(\xi)\,\mathrm{d}\xi \\ &= \int_{\mathbb{R}^d} (i2\pi\xi)^\alpha \hat{T}(\xi) f(\xi)\,\mathrm{d}\xi = \langle (i2\pi\xi)^\alpha \hat{T}, f\rangle.\end{aligned}$$

Thus, $\widehat{\partial^\alpha T} = (i2\pi\xi)^\alpha \hat{T}$ holds in the sense of distributions. $\square$

Theorem 1 (Structure Theorem for Tempered Distributions)

For every $T \in \mathcal{S}'(\mathbb{R}^d)$, there exists a continuous function g of polynomial growth and a multi-index $\alpha \in \mathbb{N}_0^d$ such that $T = \partial^\alpha g$. That is,

$$\langle T, f\rangle = (-1)^{|\alpha|}\int_{\mathbb{R}^d} g(x)\,\partial^\alpha f(x)\,\mathrm{d}x \quad \forall f \in \mathcal{S}(\mathbb{R}^d).$$

Note on Polynomial Growth. Recall that a function $g : \mathbb{R}^d \to \mathbb{C}$ is of *polynomial growth* if $|g(x)| \le C(1+|x|)^N$ for all $x \in \mathbb{R}^d$, for some $C > 0$ and integer $N \ge 1$. As an example, the function $g(x) = \max\{x, 0\}$ satisfies $g \in \mathcal{S}'(\mathbb{R})$ and its second derivative yields the Dirac delta distribution, $\frac{\mathrm{d}^2}{\mathrm{d}x^2} g = \delta_0$.

Proof. Case $d = 1$. We first establish an integration lemma.

Lemma 1

Let $h : \mathbb{R} \to \mathbb{C}$ be a measurable function of polynomial growth. Then there exists a continuous function H of polynomial growth such that:

$$\int_{-\infty}^{\infty} h(x) f^{(n)}(x)\,\mathrm{d}x = \int_{-\infty}^{\infty} H(x) f^{(n+1)}(x)\,\mathrm{d}x \quad \forall f \in \mathcal{S}(\mathbb{R}),\, n \in \mathbb{N}_0.$$

Proof. Define the primitive $G(x) = \int_0^x h(t)\,\mathrm{d}t$. Since h is of polynomial growth, there are constants such that $|h(x)| \leq C_1(1+|x|)^N$. Using $|x| \leq \frac{|x|^2+1}{2}$, we can also bound it as $|h(x)| \leq C_2(1+|x|^2)^N \leq C_3(1+|x|^{2N})$ (since $|x|^k \leq \frac{|x|^{2N}}{C_4} + C_5$). For $x \geq 0$:

$$|G(x)| \leq \int_0^x |h(t)|\,\mathrm{d}t \leq C_6 \int_0^x (1+t^{2N})\,\mathrm{d}t = C_6 x + C_6 \frac{x^{2N+1}}{2N+1}.$$

A symmetric argument holds for $x < 0$, where $|G(x)| \leq \int_x^0 |h(t)|\,\mathrm{d}t \leq C_6|x| + C_6\frac{|x|^{2N+1}}{2N+1}$; hence $|G(x)| \leq C_7(1+|x|)^{2N+1}$, meaning G is continuous and grows at most polynomially. Integrating by parts:

$$\int_{-\infty}^{\infty} G(x) f^{(n+1)}(x)\,\mathrm{d}x = \Big[G(x) f^{(n)}(x)\Big]_{-\infty}^{\infty} - \int_{-\infty}^{\infty} h(x) f^{(n)}(x)\,\mathrm{d}x.$$

Because $f \in \mathcal{S}$, $f^{(n)}(x)$ decays faster than any power of x, causing the boundary term to vanish. Setting $H = -G$ yields $\langle H, f^{(n+1)}\rangle = 0 - \langle h, f^{(n)}\rangle$, which completes the proof of the lemma. □

Proof of Theorem 1 Continued. Let $T \in \mathcal{S}'(\mathbb{R})$. There exist constants $C > 0$ and $N \in \mathbb{N}_0$ such that:

$$|\langle T, f\rangle| \leq C|f|_{(N)} \leq C \sum_{k=0}^{N} \sup_{x\in\mathbb{R}} (1+|x|)^N |f^{(k)}(x)| \leq C_1 \sum_{k=0}^{N} \sup_{x\in\mathbb{R}} (1+|x|^2)^N |f^{(k)}(x)|. \tag{1}$$

Introduce the algebraic dampening weight $\varphi(x) = (1+|x|^2)^{-N}$. Note that $\varphi \in \mathcal{O}_M$ and $|\varphi^{(\ell)}(x)| \leq C_\ell \varphi(x)$ for all $\ell \in \mathbb{N}_0$. We analyze the modulated distribution φT, defined by $\langle \varphi T, f\rangle = \langle T, \varphi f\rangle$.

Claim 1. The tempered distribution φT satisfies the bound:

$$|\langle \varphi T, f\rangle| \leq C_2 \sum_{k=0}^{N} \sup_{x\in\mathbb{R}} |f^{(k)}(x)| \quad \forall f \in \mathcal{S}.$$

Proof. Using inequality (1) on the product φf:

$$|\langle \varphi T, f\rangle| = |\langle T, \varphi f\rangle| \leq C_1 \sum_{k=0}^{N} \sup_{x\in\mathbb{R}} (1+|x|^2)^N |(\varphi f)^{(k)}(x)|.$$

By the Leibniz rule, $|(\varphi f)^{(k)}(x)| \leq \sum_{j=0}^{k} \binom{k}{j} |f^{(j)}(x)||\varphi^{(k-j)}(x)| \leq \sum_{j=0}^{k} C'_{k,j}|f^{(j)}(x)|\varphi(x)$, where $C'_{k,j} = \binom{k}{j} C_{k-j}$. Since $(1+|x|^2)^N\varphi(x) = 1$, we see:

$$|\langle \varphi T, f\rangle| \leq C_1 \sum_{k=0}^{N} \sum_{j=0}^{k} C'_{k,j} \sup_{x\in\mathbb{R}} |f^{(j)}(x)| \leq C_2 \sum_{k=0}^{N} \sup_{x\in\mathbb{R}} |f^{(k)}(x)|.$$

The weight cancels out, yielding the desired uniform derivative bound. □

Claim 2. There exist finite complex Borel measures $\mu_0, \mu_1, \ldots, \mu_N$ on $\mathbb{R}$ such that:

$$\langle \varphi T, f\rangle = \sum_{k=0}^{N} \int_{\mathbb{R}} f^{(k)}(x)\, \mathrm{d}\mu_k(x) \quad \forall f \in \mathcal{S}.$$

Proof. Let $C_0(\mathbb{R})$ be the Banach space of continuous functions vanishing at infinity, equipped with the supremum norm $\|g\|_u = \sup_x |g(x)|$. Form the product Banach space $E = \prod_{k=0}^{N} C_0(\mathbb{R})$ with the sum norm $\|(g_0, \ldots, g_N)\|_u = \sum_{k=0}^{N} \|g_k\|_u$. Define the linear subspace:

$$M = \left\{ \left(f, f', \ldots, f^{(N)}\right) : f \in \mathcal{S} \right\} \subseteq E.$$

The functional $\Lambda : M \to \mathbb{C}$ given by $\Lambda\left(f, \ldots, f^{(N)}\right) = \langle \varphi T, f\rangle$ is bounded under $\|\cdot\|_u$ by Claim 1:

$$|\Lambda(f, \ldots, f^{(N)})| = |\langle \varphi T, f\rangle| \leq C_2 \|(f, \ldots, f^{(N)})\|_u.$$

By the Hahn-Banach theorem, Λ extends to a bounded linear functional on all of E with $\|\Lambda\|_{E^*} \leq C_2$. Since $\Lambda(g_0, 0, \ldots, 0)$ defines a bounded functional on $C_0(\mathbb{R})$, the Riesz Representation Theorem implies that there exists a finite complex Borel measure μ_0 such that $\Lambda(g_0, 0, \ldots, 0) = \int_{\mathbb{R}} g_0 \,\mathrm{d}\mu_0$. Proceeding slot by slot (using linearity $\Lambda(g_0, g_1) = \Lambda(g_0, 0) + \Lambda(0, g_1)$), we get components $\mu_1, \ldots, \mu_N$ satisfying $\Lambda(g_0, \ldots, g_N) = \sum_{k=0}^{N} \int_{\mathbb{R}} g_k(x)\, \mathrm{d}\mu_k(x)$, yielding the stated integral representation for $\langle \varphi T, f\rangle$. □

Claim 3. There exist bounded measurable functions $F_0, F_1, \ldots, F_N$ such that:

$$\langle \varphi T, f\rangle = \sum_{k=0}^{N} \int_{\mathbb{R}} f^{(k+1)}(x) F_k(x)\, \mathrm{d}x \quad \forall f \in \mathcal{S}.$$

Proof. Let $G_k(x) = \mu_k((-\infty, x])$ be the cumulative distribution functions. Since the measures are finite, each G_k is bounded and of bounded variation. Integrating by parts against the smooth test function $f^{(k)}$:

$$\langle \varphi T, f\rangle = \sum_{k=0}^{N} \int_{-\infty}^{\infty} f^{(k)}(x)\, \mathrm{d}G_k(x) = \sum_{k=0}^{N} \left[f^{(k)}(x) G_k(x) \right]_{-\infty}^{\infty} - \sum_{k=0}^{N} \int_{-\infty}^{\infty} f^{(k+1)}(x) G_k(x)\, \mathrm{d}x.$$

Because $f^{(k)} \in \mathcal{S}$ and G_k is bounded, the boundary terms vanish, leaving $F_k = -G_k$ as the required bounded functions. □

Proof of Theorem 1 Continued. We reconstruct T by writing $T = \frac{1}{\varphi}(\varphi T)$. Because $\frac{1}{\varphi}(x) = (1+|x|^2)^N \in \mathcal{O}_M$, the test function $\frac{f}{\varphi}$ belongs to $\mathcal{S}$:

$$\langle T, f\rangle = \left\langle \varphi T, \frac{f}{\varphi} \right\rangle = \sum_{k=0}^{N} \int_{\mathbb{R}} F_k(x) \left(\frac{f}{\varphi}\right)^{(k+1)} (x)\, \mathrm{d}x.$$

Expanding $(f/\varphi)^{(k+1)}$ via the Leibniz rule gives $\left(\frac{1}{\varphi} f\right)^{(k+1)} = \sum_{\ell=0}^{k+1} \binom{k+1}{\ell} f^{(\ell)} \left(\frac{1}{\varphi}\right)^{(k+1-\ell)}$. This yields a finite sum of terms of the form:

$$\langle T, f\rangle = \sum_{k=0}^{N} \sum_{\ell=0}^{k+1} \int_{\mathbb{R}} f^{(\ell)}(x)\, H_{k,\ell}(x)\, \mathrm{d}x, \quad \text{where } H_{k,\ell} = \binom{k+1}{\ell} F_k \left(\tfrac{1}{\varphi}\right)^{(k+1-\ell)}.$$

Each $H_{k,\ell}$ is a bounded function multiplied by a derivative of $(1+|x|^2)^N$, which is measurable and of polynomial growth.

Applying Lemma 1 repeatedly to each term allows us to shift all derivatives up to a uniform high order $N+2$, converting each component to $\int_{\mathbb{R}} f^{(N+2)}(x) \widetilde{H}_{k,\ell}(x)\, \mathrm{d}x$, where $\widetilde{H}_{k,\ell}$ are continuous and of polynomial growth. Summing these up, we define:

$$g(x) = (-1)^{N+2} \sum_{k=0}^{N} \sum_{\ell=0}^{k+1} \widetilde{H}_{k,\ell}(x),$$

which is continuous and of polynomial growth. This leaves us with

$$\langle T, f\rangle = (-1)^{N+2} \int_{\mathbb{R}} g(x) f^{(N+2)}(x)\, \mathrm{d}x = \langle \partial^{N+2} g, f\rangle,$$

confirming $T = \partial^{N+2} g$ (with $\alpha = N+2$). □

The Helmholtz Equation in $\mathcal{S}'(\mathbb{R}^d)$

We investigate the generalized Helmholtz equation on tempered distributions:

$$u - \Delta u = g \quad \text{in } \mathcal{S}'(\mathbb{R}^d), \tag{H}$$

where $g \in \mathcal{S}'(\mathbb{R}^d)$ is given and $u \in \mathcal{S}'(\mathbb{R}^d)$ is the unknown distribution.

Proposition 2

For any $g \in \mathcal{S}'(\mathbb{R}^d)$, there exists a unique solution $u \in \mathcal{S}'(\mathbb{R}^d)$ to (H). Thus, the operator $(I - \Delta) : \mathcal{S}'(\mathbb{R}^d) \to \mathcal{S}'(\mathbb{R}^d)$ is a bijection, and the solution is given explicitly by:

$$u = \mathcal{F}^{-1} \left(\frac{1}{1 + 4\pi^2 |\xi|^2} \hat{g} \right).$$

Proof. The polynomial factor $p(\xi) = 1 + 4\pi^2|\xi|^2$ and its reciprocal $p(\xi)^{-1} = \frac{1}{1+4\pi^2|\xi|^2}$ both belong to $\mathcal{O}_M(\mathbb{R}^d)$. Applying the Fourier transform to both sides of (H) converts differentiation into algebraic multiplication (recalling $\widehat{\partial^\alpha T} = (i2\pi\xi)^\alpha \hat{T}$, which means $\widehat{\Delta u} = -4\pi^2|\xi|^2\hat{u}$):

$$u - \Delta u = g \text{ in } \mathcal{S}' \iff \widehat{u - \Delta u} = \hat{g} \text{ in } \mathcal{S}' \iff (1 + 4\pi^2|\xi|^2)\hat{u} = \hat{g} \text{ in } \mathcal{S}'.$$

Since multiplication by $p(\xi)^{-1}$ is a valid and invertible operation within $\mathcal{S}'(\mathbb{R}^d)$, we solve uniquely for $\hat{u}$:

$$\hat{u} = \frac{1}{1 + 4\pi^2|\xi|^2}\hat{g} \text{ in } \mathcal{S}'.$$

Applying $\mathcal{F}^{-1}$ to both sides yields $u = \mathcal{F}^{-1}\big(\frac{1}{1+4\pi^2|\xi|^2}\hat{g}\big)$. This uniquely determines $u \in \mathcal{S}'$, completing the bijection proof. □

The L^2 Setting and Sobolev Spaces

Let $s \in \mathbb{N}_0$. We define the classical Sobolev space $H^s(\mathbb{R}^d)$ as:

$$H^s(\mathbb{R}^d) = \left\{ u \in \mathcal{S}'(\mathbb{R}^d) : \partial^\alpha u \in L^2(\mathbb{R}^d) \text{ for all } |\alpha| \leq s \right\}.$$

This space is a Hilbert space under the norm:

$$\|u\|_{H^s} = \left(\sum_{|\alpha| \leq s} \|\partial^\alpha u\|_{L^2}^2 \right)^{1/2}.$$

Operationally, $u \in H^s(\mathbb{R}^d)$ implies that $u \in L^2(\mathbb{R}^d)$ (since $H^s \subseteq L^2$) and its distributional derivatives up to order s can be represented by L^2 functions satisfying $\int_{\mathbb{R}^d} \partial^\alpha u \, f \, \mathrm{d}x = (-1)^{|\alpha|} \int_{\mathbb{R}^d} u \, \partial^\alpha f \, \mathrm{d}x$ for all $f \in \mathcal{S}$.

Proposition 3

For every $g \in L^2(\mathbb{R}^d)$, the unique solution u to (H) belongs to $H^2(\mathbb{R}^d)$. It satisfies $u - \Delta u = g$ almost everywhere, and obeys the structural stability estimate:

$$\|u\|_{H^2} \leq C\|g\|_{L^2},$$

where C is a constant independent of g. Thus, $(I - \Delta) : H^2(\mathbb{R}^d) \to L^2(\mathbb{R}^d)$ is a bijection.

Proof. Let $g \in L^2(\mathbb{R}^d)$, which implies $\hat{g} \in L^2(\mathbb{R}^d)$. By Proposition 2, the unique distribution solution is $u = \mathcal{F}^{-1}\left(\frac{1}{1+4\pi^2|\xi|^2}\hat{g}\right)$, which belongs to $L^2(\mathbb{R}^d)$ since its transform is bounded by $|\hat{g}|$. Taking the Fourier transform representation, for any multi-index $|\alpha| \leq 2$:

$$\widehat{\partial^\alpha u}(\xi) = (i2\pi\xi)^\alpha \hat{u}(\xi) = \frac{(i2\pi\xi)^\alpha}{1 + 4\pi^2|\xi|^2}\hat{g}(\xi).$$

Because $|(i2\pi\xi)^\alpha| \le (2\pi|\xi|)^{|\alpha|} \le 1+4\pi^2|\xi|^2$ holds for any $|\alpha| \le 2$, the multiplier is uniformly bounded by 1. Consequently, $|\widehat{\partial^\alpha u}(\xi)| \le |\hat{g}(\xi)|$. By Plancherel's theorem, $\partial^\alpha u \in L^2(\mathbb{R}^d)$ and satisfies:

$$\|\partial^\alpha u\|_{L^2} = \|\widehat{\partial^\alpha u}\|_{L^2} \le \|\hat{g}\|_{L^2} = \|g\|_{L^2}.$$

Summing over all $|\alpha| \le 2$ yields the norm stability estimate $\|u\|_{H^2} \le C\|g\|_{L^2}$. Uniqueness is guaranteed by the overarching uniqueness in $\mathcal{S}'$ (Proposition 2). Thus, $(I-\Delta) : H^2(\mathbb{R}^d) \to L^2(\mathbb{R}^d)$ is a bijection. $\square$

Approximation via Smooth Limits. Alternatively, we can resolve the equation for $g \in L^2$ as a limit of smooth solutions. Choose a sequence $g_n \in \mathcal{S}$ such that $g_n \to g$ in L^2. Since $(I-\Delta) : \mathcal{S} \to \mathcal{S}$ is an isomorphism, there correspond unique smooth solutions $u_n \in \mathcal{S}$. Linearity implies $(u_n - u_m) - \Delta(u_n - u_m) = g_n - g_m$, and applying the stability estimate from Proposition 3:

$$\|\partial^\alpha u_n - \partial^\alpha u_m\|_{L^2} \le C\|g_n - g_m\|_{L^2} \to 0 \quad \text{as } n, m \to \infty,$$

for all $|\alpha| \le 2$. By the completeness of L^2, each derivative sequence $\partial^\alpha u_n$ converges to some function $u_\alpha \in L^2$ (with $u \equiv u_0$). For any test function $f \in \mathcal{S}$, the standard pairing satisfies $\int_{\mathbb{R}^d} \partial^\alpha u_n \, f \, \mathrm{d}x = (-1)^{|\alpha|} \int_{\mathbb{R}^d} u_n \, \partial^\alpha f \, \mathrm{d}x$. Passing to the limit $n \to \infty$ inside the integrals gives:

$$\int_{\mathbb{R}^d} u_\alpha \, f \, \mathrm{d}x = (-1)^{|\alpha|} \int_{\mathbb{R}^d} u \, \partial^\alpha f \, \mathrm{d}x,$$

confirming that $u_\alpha = \partial^\alpha u$ in the distributional sense. Hence $u \in H^2(\mathbb{R}^d)$ and satisfies $u - \Delta u = g$ almost everywhere.

Proposition 4

For any source term $g \in H^s(\mathbb{R}^d)$ ($s \in \mathbb{N}_0$), the unique solution u to (H) satisfies $u \in H^{s+2}(\mathbb{R}^d)$. The operator $(I-\Delta) : H^{s+2}(\mathbb{R}^d) \to H^s(\mathbb{R}^d)$ is a bijection.

Proof. Differentiating the equation (H) distributionally yields $\partial^\alpha u - \Delta(\partial^\alpha u) = \partial^\alpha g$ for any multi-index α. If $g \in H^s(\mathbb{R}^d)$, then for all $|\alpha| \le s$, the derivative term satisfies $\partial^\alpha g \in L^2(\mathbb{R}^d)$. Applying Proposition 3 to this differentiated equation, we find that $\partial^\alpha u \in H^2(\mathbb{R}^d)$, which implies that for all $|\beta| \le 2$:

$$\|\partial^{\beta+\alpha} u\|_{L^2} = \|\partial^\beta(\partial^\alpha u)\|_{L^2} \le C\|\partial^\alpha g\|_{L^2}.$$

Any multi-index γ with $|\gamma| \le s+2$ can be split into components $\gamma = \beta + \alpha$ where $|\alpha| \le s$ and $|\beta| \le 2$. Accumulating these bounds implies:

$$\|\partial^\gamma u\|_{L^2} \le C\|g\|_{H^s} \quad \text{for all } |\gamma| \le s+2.$$

Hence, $u \in H^{s+2}(\mathbb{R}^d)$ solves (H), and uniqueness follows directly from the broad uniqueness property in $\mathcal{S}'$ established in Proposition 2. Thus, $(I-\Delta) : H^{s+2}(\mathbb{R}^d) \to H^s(\mathbb{R}^d)$ is a bijection. $\square$

Weak Convergence

Motivation and Fundamental Definition

Let E be a normed vector space and $E^* = \mathcal{L}(E, \mathbb{R})$ denote its continuous topological dual space. Consider a sequence $\{x_n\}_{n=1}^{\infty} \subset E$ approximating an element $x \in E$.

By a fundamental consequence of the Hahn–Banach theorem, the standard norm topology characterizes strong convergence via uniform control over the dual unit ball:

$$\|x_n - x\|_E = \sup_{\|\ell\|_{E^*} \leq 1} |\ell(x_n - x)| \longrightarrow 0 \quad \text{as } n \to \infty.$$

Equivalently, norm convergence dictates that $\ell(x_n) \to \ell(x)$ *uniformly* over bounded subsets of functionals. *Weak convergence* relaxes this topological requirement by demanding only pointwise (element-by-element) convergence across the dual space.

Weak Convergence

We say that a sequence $\{x_n\}_{n=1}^{\infty} \subset E$ converges *weakly* to $x \in E$, denoted by $x_n \rightharpoonup x$, if:

$$\ell(x_n) \longrightarrow \ell(x) \quad \text{as } n \to \infty \quad \forall \ell \in E^*.$$

Using duality pairing notation, this is expressed as:

$$\langle \ell, x_n \rangle_{E^*, E} \longrightarrow \langle \ell, x \rangle_{E^*, E} \quad \text{as } n \to \infty \quad \forall \ell \in E^*.$$

Concrete Realizations in Classical Spaces

- *The L^p Setting.* Let $E = L^p(\Omega, \mu)$ with $p \in [1, \infty)$. Under standard σ-finite conditions, the dual space is isomorphic to $L^q(\Omega, \mu)$ where $\frac{1}{p} + \frac{1}{q} = 1$. Consequently, a sequence $f_n \rightharpoonup f$ weakly in L^p if and only if:

$$\int_\Omega f_n(x) g(x) \, \mathrm{d}\mu(x) \longrightarrow \int_\Omega f(x) g(x) \, \mathrm{d}\mu(x) \quad \forall g \in L^q(\Omega, \mu).$$

- *The Space of Continuous Functions $C(X)$.* Let X be a compact metric space and $E = C(X)$ equipped with the uniform supremum norm $\|f\|_u = \sup_{x \in X} |f(x)|$. By the Riesz–Markov–Kakutani representation theorem, the dual space $[C(X)]^*$ is isometric to $M(X)$, the space of finite signed Radon measures on X. Thus, $f_n \rightharpoonup f$ weakly in $C(X)$ if and only if:

$$\int_X f_n(x) \, \mathrm{d}\mu(x) \longrightarrow \int_X f(x) \, \mathrm{d}\mu(x) \quad \forall \mu \in M(X).$$

- *Hilbert Spaces.* Let $E = H$ be a Hilbert space. By the Riesz representation theorem, $H^* \cong H$, meaning weak convergence $x_n \rightharpoonup x$ is fully characterized by the inner product:

$$\langle x_n, z \rangle_H \longrightarrow \langle x, z \rangle_H \quad \text{as } n \to \infty \quad \forall z \in H.$$

  Connection to strong convergence: $x_n \to x$ strongly in H if and only if $x_n \rightharpoonup x$ weakly and $\limsup_{n \to \infty} \|x_n\|_H \leq \|x\|_H$.

The Finite-Dimensional Special Case

Proposition 1

Let E be a finite-dimensional normed vector space (e.g., $E = \mathbb{R}^d$). Then the weak and strong topologies coincide; that is,

$$x_n \to x \quad \iff \quad x_n \rightharpoonup x.$$

Proof.

- ($\Rightarrow$) Assume $x_n \to x$ strongly. For each continuous linear functional $\ell \in E^*$, boundedness implies:

  $$|\ell(x_n) - \ell(x)| = |\ell(x_n - x)| \leq \|\ell\|_{E^*} \|x_n - x\|_E \longrightarrow 0 \quad \text{as } n \to \infty,$$

  which proves $x_n \rightharpoonup x$ weakly. This direction holds universally in any normed space.

- ($\Leftarrow$) Suppose $\dim E = d < \infty$ and $x_n \rightharpoonup x$. Let $\{v_1, \ldots, v_d\}$ be a basis for E, and let $\{\ell_1, \ldots, \ell_d\} \subset E^*$ represent the corresponding dual basis satisfying $\ell_i(v_j) = \delta_{ij}$. Every element $y \in E$ can be decomposed as $y = \sum_{i=1}^{d} \ell_i(y) v_i$. By the definition of weak convergence, we obtain coordinate-wise convergence: $\ell_i(x_n) \to \ell_i(x)$ in $\mathbb{R}$ for each $i \in \{1, \ldots, d\}$. Because all norms on a finite-dimensional space are equivalent, we can bound the standard norm by the ℓ^1-coordinate norm:

  $$\|x_n - x\|_E \leq C \sum_{i=1}^{d} |\ell_i(x_n) - \ell_i(x)| \longrightarrow 0 \quad \text{as } n \to \infty,$$

  confirming strong norm convergence. $\square$

Example. For a sequence of vectors $\{x^{(n)}\}_{n=1}^{\infty} \subset \mathbb{R}^d$, strong convergence $x^{(n)} \to x$ is precisely equivalent to component-wise convergence: $x_i^{(n)} \to x_i$ for every coordinate index $i = 1, 2, \ldots, d$.

The Infinite-Dimensional Counterexample

In general infinite-dimensional spaces, weak convergence does *not* imply strong convergence. Consider an infinite-dimensional Hilbert space H containing a complete orthonormal system (CONS) $\{e_n\}_{n=1}^{\infty}$. For any $n \neq m$, Pythagoras' theorem dictates:

$$\|e_n - e_m\|_H^2 = \|e_n\|^2 - 2\langle e_n, e_m \rangle_H + \|e_m\|^2 = 1 - 0 + 1 = 2.$$

Because the distance between any two distinct elements is always $\sqrt{2}$, the sequence $\{e_n\}$ cannot contain any strongly convergent subsequence. However, $e_n \rightharpoonup 0$ weakly. Indeed, for every element $x \in H$, Bessel's inequality / Parseval's identity guarantees that:

$$\sum_{n=1}^{\infty} |\langle x, e_n \rangle_H|^2 = \|x\|_H^2 < \infty \implies \lim_{n \to \infty} \langle e_n, x \rangle_H = 0 = \langle 0, x \rangle_H.$$

The Canonical Embedding and Weakly Cauchy Sequences

The canonical evaluation map $J : E \to E^{**}$ into the bidual space $[E^*]^*$. Every element $x \in E$ acts as a bounded functional $T_x \in E^{**}$ via $T_x(\ell) = \ell(x)$ for all $\ell \in E^*$. The Hahn–Banach theorem guarantees this map is a linear isometry:

$$\|T_x\|_{E^{**}} = \sup_{\|\ell\|_{E^*} \leq 1} |T_x(\ell)| = \sup_{\|\ell\|_{E^*} \leq 1} |\ell(x)| = \|x\|_E.$$

Weakly Cauchy Sequence

A sequence $\{x_n\}_{n=1}^{\infty} \subset E$ is called *weakly Cauchy* if the scalar sequence $\{\ell(x_n)\}_{n=1}^{\infty}$ is Cauchy in $\mathbb{R}$ (or $\mathbb{C}$) for every continuous linear functional $\ell \in E^*$.

Note. Every weakly convergent sequence is fundamentally weakly Cauchy because any convergent sequence in $\mathbb{R}$ is a Cauchy sequence.

Proposition 2

Let E be a normed vector space. Then the following properties hold:

a. If $x_n \rightharpoonup x$, then the norm is sequentially lower semicontinuous:

$$\|x\|_E \leq \liminf_{n\to\infty} \|x_n\|_E.$$

b. If $\{x_n\}$ is weakly Cauchy, then it is bounded in norm: $\sup_n \|x_n\|_E < \infty$.

c. If E is a reflexive Banach space, then the space is weakly sequentially complete; i.e., every weakly Cauchy sequence converges weakly to an element in E.

Proof.

a. *Lower Semicontinuity.* By the Hahn–Banach theorem, choose a normalized functional $\ell \in E^*$ such that $\|\ell\|_{E^*} = 1$ and $\ell(x) = \|x\|_E$. Evaluating the limit under weak convergence:

$$\|x\|_E = \ell(x) = \lim_{n\to\infty} \ell(x_n) = \liminf_{n\to\infty} \ell(x_n).$$

Since $\ell(x_n) \leq \|\ell\|_{E^*}\|x_n\|_E = \|x_n\|_E$ holds uniformly for all n, taking the lower limit yields:

$$\|x\|_E = \liminf_{n\to\infty} \ell(x_n) \leq \liminf_{n\to\infty} \|x_n\|_E.$$

□

b. *Uniform Boundedness.* Let $\{x_n\}$ be weakly Cauchy. For each individual functional $\ell \in E^*$, the sequence $\{\ell(x_n)\}$ converges in $\mathbb{R}$, which implies it is bounded: $\sup_n |\ell(x_n)| < \infty$. By setting $T_{x_n}(\ell) = \ell(x_n)$. The sequence of operators $\{T_{x_n}\} \subset E^{**}$ satisfies a pointwise boundedness condition on the Banach space E^*: $\sup_n |T_{x_n}(\ell)| < \infty$ for all $\ell \in E^*$. Applying the Banach–Steinhaus Uniform Boundedness Principle yields a uniform bound on the operator norms:

$$\sup_n \|x_n\|_E = \sup_n \|T_{x_n}\|_{E^{**}} = \sup_n \sup_{\|\ell\|_{E^*} \leq 1} |T_{x_n}(\ell)| < \infty.$$

$\square$

c. *Reflexive Completeness.* Let $\{x_n\}$ be weakly Cauchy inside a reflexive space ($E = E^{**}$). By part b, there exists a uniform bound $M = \sup_n \|x_n\|_E < \infty$. Define a functional $T : E^* \to \mathbb{R}$ point-by-point via the pointwise limit:

$$T(\ell) = \lim_{n\to\infty} \ell(x_n).$$

Linearity of T follows directly from the linearity of the limit operation. Taking boundaries, the inequality $|\ell(x_n)| \le M\|\ell\|_{E^*}$ implies $|T(\ell)| \le M\|\ell\|_{E^*}$, hence $T \in E^{**}$. Because E is reflexive, the canonical mapping is surjective, ensuring there exists a unique element $x \in E$ such that $T(\ell) = \ell(x)$ for all $\ell \in E^*$. Substituting this back gives $\ell(x) = \lim_{n\to\infty} \ell(x_n)$, which matches the definition of $x_n \rightharpoonup x$. $\square$

Weak Compactness

Solving $Lx = g$ by Approximation

Solving a bounded linear equation

$$Lx = g, \quad L \in \mathcal{L}(E, F),$$

together with a family of *approximating* problems

$$L_n x_n = g_n, \quad \text{with the a priori bound} \quad \sup_n \|x_n\|_E < \infty.$$

Two questions govern whether the approximate solutions $\{x_n\}$ recover an exact solution of the original equation.

- *Extraction of a limit.* Does there exist a subsequence and a candidate limit $x \in E$ with

$$x_{n_k} \rightharpoonup x \quad \text{(weakly)} \quad ?$$

- *Identification of the limit.* Once such an x exists, can we pass to the limit and conclude $Lx = g$? Using the dual (adjoint) operator L' and the duality pairing $\langle \cdot, \cdot \rangle_{E^*,E}$,

$$\langle (L_n)'\ell,\, x_n\rangle = \langle \ell,\, L_n x_n\rangle = \langle \ell,\, g_n\rangle \quad \xrightarrow{n\to\infty} \quad \langle L'\ell,\, x\rangle = \langle \ell,\, g\rangle, \quad \forall \ell \in F^*.$$

  Since $\langle L'\ell, x\rangle = \langle \ell, Lx\rangle$, this yields $\langle \ell, Lx\rangle = \langle \ell, g\rangle$ for all $\ell \in F^*$, hence $Lx = g$, the range of L being a closed subspace of the Banach space F.

Lemma 1

Reflexivity passes to closed subspaces: if $E^{**} = E$, then for every closed subspace $M \subseteq E$ one has $M^{**} = M$; that is, M is itself reflexive.

Lemma 2

Separability is inherited from the dual: if E^* is separable, then E is separable.

Weak Compactness Theorem

Let E be a reflexive Banach space. Then every bounded sequence in E has a weakly convergent subsequence. Explicitly, if

$$\|x_n\|_E \le C \quad \forall n \quad \text{for some } C > 0,$$

then there exists a subsequence $\{x_{n_k}\}$ and an element $x \in E$ such that

$$x_{n_k} \rightharpoonup x \quad \text{as } k \to \infty.$$

Proof. Let $\{x_n\} \subset E$ satisfy $\|x_n\|_E \le C$ for all n. Since every weakly Cauchy sequence in a reflexive Banach space possesses a weak limit, it suffices to produce a subsequence $\{x_{n_k}\}$ that is *weakly Cauchy*, i.e., such that

$$\{\ell(x_{n_k})\}_{k=1}^{\infty} \text{ is Cauchy in } \mathbb{R} \quad \forall \ell \in E^*.$$

A separable, reflexive ambient subspace. Let M denote the closure of the linear span of the sequence,

$$M = \overline{\operatorname{span}}\{x_n : n \ge 1\}.$$

Then M is *separable*: rational (finite) linear combinations of the x_n form a countable dense subset of M. Moreover, M is a closed subspace of the reflexive space E, hence *reflexive* by Lemma 1. Consequently $(M^*)^* = M$ is separable, so applying Lemma 2 to M^* shows that M^* is separable.

A countable dense family of functionals. Choose a countable dense set

$$D = \{\ell_m : m \ge 1\} \subseteq M^*, \quad \overline{D} = M^*,$$

which exists by separability of M^*. For each fixed m, the scalar sequence $\{\ell_m(x_n)\}_n$ is bounded:

$$|\ell_m(x_n)| \le \|\ell_m\|_{M^*}\|x_n\|_E \le C\|\ell_m\|_{M^*} < \infty, \quad \|\ell_m\|_{M^*} = \sup_{\substack{x \in M \\ \|x\|_E = 1}} |\ell_m(x)|,$$

so that $\sup_n |\ell_m(x_n)| < \infty$.

Diagonal extraction. By Bolzano–Weierstrass, each bounded sequence $\{\ell_m(x_n)\}_n$ admits a convergent subsequence. Applying this successively for $m = 1, 2, 3, \dots$ and passing to a diagonal subsequence yields a single subsequence $\{x_{n_k}\}$ for which

$$\lim_{k\to\infty} \ell_m(x_{n_k}) \quad \text{exists for every } m \ge 1.$$

The same subsequence works simultaneously for all m.

Weak Cauchy property. Fix an arbitrary $\ell \in E^*$ and let $R\ell := \ell|_M \in M^*$ denote its restriction. For any $m \ge 1$, inserting and removing ℓ_m gives

$$\begin{aligned}
|\ell(x_{n_k}) - \ell(x_{n_j})| &= |R\ell(x_{n_k}) - R\ell(x_{n_j})| \\
&\le |R\ell(x_{n_k}) - \ell_m(x_{n_k})| + |\ell_m(x_{n_k}) - \ell_m(x_{n_j})| + |\ell_m(x_{n_j}) - R\ell(x_{n_j})| \\
&\le \|R\ell - \ell_m\|_{M^*}\|x_{n_k}\|_E + |\ell_m(x_{n_k}) - \ell_m(x_{n_j})| + \|R\ell - \ell_m\|_{M^*}\|x_{n_j}\|_E \\
&\le 2C\|R\ell - \ell_m\|_{M^*} + |\ell_m(x_{n_k}) - \ell_m(x_{n_j})|.
\end{aligned}$$

Taking the limit superior over $k, j \to \infty$,

$$\limsup_{k,j\to\infty} |\ell(x_{n_k}) - \ell(x_{n_j})| \le 2C\|R\ell - \ell_m\|_{M^*} + \underbrace{\limsup_{k,j\to\infty} |\ell_m(x_{n_k}) - \ell_m(x_{n_j})|}_{=0},$$

the bracketed term vanishing because $\lim_k \ell_m(x_{n_k})$ exists for every $m \ge 1$.

Density closes the argument. Since $\{\ell_m\}_{m\ge1}$ is dense in M^*, the quantity $2C\|R\ell - \ell_m\|_{M^*}$ can be made arbitrarily small by a suitable choice of m. Therefore

$$\limsup_{k,j\to\infty} |\ell(x_{n_k}) - \ell(x_{n_j})| = 0,$$

i.e., $\{\ell(x_{n_k})\}_k$ is Cauchy for every $\ell \in E^*$. Hence $\{x_{n_k}\}$ is weakly Cauchy and, by reflexivity, admits a weak limit $x \in E$ with $x_{n_k} \rightharpoonup x$. □

Weak-$*$ Convergence and Weakly-$*$ Compact Sets

Let E be a normed vector space with continuous dual $E^* = \mathcal{L}(E, \mathbb{R})$. Whereas weak convergence on E^* tests against the full bidual E^{**}, weak-$*$ convergence tests only against the original space E.

Weak-$*$ Convergence

Let $\ell_n, \ell \in E^*$. We say that ℓ_n converges *weakly*-$*$ to ℓ, denoted $\ell_n \overset{*}{\rightharpoonup} \ell$, in E^* if

$$\ell_n(x) \longrightarrow \ell(x) \quad \text{as } n \to \infty \quad \forall x \in E.$$

Note. On the dual space E^* there are three distinct modes of convergence:

1. *Norm (strong) convergence* in E^*. $\|\ell_n - \ell\|_{E^*} \to 0$.
2. *Weak convergence.* $\rightharpoonup$ tested against every functional in the bidual E^{**}.
3. *Weak-$*$ convergence.* $\overset{*}{\rightharpoonup}$ tested only against the evaluations $x \in E \subseteq E^{**}$.

Remarks.

a. *Hierarchy of the three modes.* Recall the canonical embedding $E \subseteq (E^*)^* = E^{**}$. Because weak-$*$ convergence tests against the smaller collection $E \subseteq E^{**}$, each stronger mode implies the next:

$$\|\ell_n - \ell\|_{E^*} \to 0 \quad \Longrightarrow \quad \ell_n \rightharpoonup \ell \quad \Longrightarrow \quad \ell_n \overset{*}{\rightharpoonup} \ell.$$

The first implication is the general fact that norm convergence implies weak convergence; concretely, for any $x \in E^{**}$,

$$|x(\ell_n) - x(\ell)| \le \|x\|_{E^{**}} \|\ell_n - \ell\|_{E^*} \longrightarrow 0.$$

b. *Coincidence under reflexivity.* If E is a reflexive Banach space, so that $(E^*)^* = E$, then the bidual reduces to E and weak and weak-$*$ convergence *coincide*:

$$\ell_n \rightharpoonup \ell \quad \Longleftrightarrow \quad \ell_n \overset{*}{\rightharpoonup} \ell.$$

Consequently, weak-$*$ convergence is genuinely interesting only when E is *not* reflexive.

c. *The space L^1.* Let $E = L^1(\mu)$ with μ a σ-finite measure. Then $E^* = L^\infty(\mu)$, while E is *not* reflexive since $(L^\infty)^* \neq L^1$. Here, for $f_n, f \in L^\infty$,

$$f_n \overset{*}{\rightharpoonup} f \text{ in } L^\infty \quad \Longleftrightarrow \quad \int f_n\, g \,\mathrm{d}\mu \longrightarrow \int f\, g \,\mathrm{d}\mu \quad \forall g \in L^1(\mu).$$

The dualities $(L^p)^* = L^q$; in particular σ-finiteness gives $(L^1)^* = L^\infty$, but $(L^\infty)^* \neq L^1$.

d. *The space $C(X)$.* Let X be a compact metric space and $E = C(X)$ with the uniform norm $\|f\|_u$. Then $E^* = M(X)$, the space of finite signed (real) Borel measures on X, and for $\mu_n, \mu \in M(X)$,

$$\mu_n \overset{*}{\rightharpoonup} \mu \text{ in } M(X) \quad \Longleftrightarrow \quad \int_X f \,\mathrm{d}\mu_n \longrightarrow \int_X f \,\mathrm{d}\mu \quad \forall f \in C(X).$$

Weak-$*$ Sequential Compactness

Recall that for functionals $\ell_n, \ell \in E^*$, weak-$*$ convergence $\ell_n \overset{*}{\rightharpoonup} \ell$ in E^* means

$$\ell_n(x) \longrightarrow \ell(x) \quad \text{as } n \to \infty, \quad \text{for every } x \in E.$$

We first record an example showing that weak-$*$ convergence is strictly weaker than weak convergence, and then prove the sequential form of the Banach–Alaoglu theorem.

Example: Weak-$*$ but Not Weak Convergence

Let

$$E = c_0 = \Big\{ x = (x_n)_{n\geq 1} \in \ell^\infty : \lim_{n\to\infty} x_n = 0 \Big\},$$

equipped with the ℓ^∞-norm $\|x\|_\infty = \sup_n |x_n|$. One can show that the dual is

$$E^* = \ell^1, \quad \text{and hence} \quad (E^*)^* = (\ell^1)^* = \ell^\infty \supsetneq c_0.$$

In particular c_0 is *not* reflexive: the bidual ℓ^∞ is strictly larger than c_0.

For each $n \geq 1$, let $\ell_n \in \ell^1 = E^*$ be the sequence with a single 1 in the n-th position and 0 elsewhere:

$$\ell_n = (\underbrace{0, \dots, 0}_{n-1}, 1, 0, \dots).$$

As a functional on c_0, ℓ_n acts on $x = (x_k) \in c_0$ by

$$\ell_n(x) = \sum_{k=1}^\infty (\ell_n)_k\, x_k = x_n.$$

$\ell_n \xrightarrow{*} 0$. Fix any $x = (x_k) \in c_0$. By the definition of c_0 we have $x_n \to 0$ as $n \to \infty$, so

$$\ell_n(x) = x_n \longrightarrow 0 = \mathbf{0}(x) \quad \text{for every } x \in c_0.$$

This is exactly the statement that $\ell_n \xrightarrow{*} 0$, weak-$*$ convergence to the zero functional.

Claim: ℓ_n has no weak limit in $E^* = \ell^1$. Weak convergence of $\{\ell_n\}$ in ℓ^1 would require $\Lambda(\ell_n)$ to converge for *every* $\Lambda \in (\ell^1)^* = \ell^\infty$. Take the specific element

$$z = (0, 1, 0, 1, \dots) \in \ell^\infty \quad \text{(entries alternating between 0 and 1)},$$

regarded as the functional on ℓ^1 given by $z(\ell) = \sum_{k=1}^{\infty} z_k \, \ell_k$. Then

$$z(\ell_n) = \sum_{k=1}^{\infty} z_k \, (\ell_n)_k = z_n = \begin{cases} 0, & n \text{ odd}, \\ 1, & n \text{ even}, \end{cases}$$

and $\lim_{n\to\infty} z_n$ does not exist. So $\Lambda(\ell_n)$ fails to converge for this $\Lambda = z$, and $\{\ell_n\}$ has no weak limit, even though it has the weak-$*$ limit 0. This is why weak-$*$ convergence is interesting precisely when E is not reflexive.

Banach–Alaoglu Theorem (sequential form)

Let E be a separable normed vector space. Then every bounded sequence in E^* has a weak-$*$ convergent subsequence. Explicitly: if $\{\ell_n\} \subset E^*$ satisfies

$$\|\ell_n\|_{E^*} \le C \quad \text{for all } n, \text{ for some } C > 0,$$

then there exist a subsequence $\{\ell_{n_k}\}$ and a functional $\ell \in E^*$ such that $\ell_{n_k} \xrightarrow{*} \ell$.

Proof. Let $\{\ell_n\} \subset E^*$ with $\|\ell_n\|_{E^*} \le C$ for all n. The proof has five steps.

A countable dense set in E. Since E is separable, we may choose a countable set

$$\{x_m : m \ge 1\} \subseteq E \quad \text{that is dense in } E.$$

Each scalar sequence is bounded. Fix m. For every n,

$$|\ell_n(x_m)| \le \|\ell_n\|_{E^*} \, \|x_m\|_E \le C \, \|x_m\|_E.$$

The right-hand side does not depend on n, so the real sequence $\{\ell_n(x_m)\}_{n\ge1}$ is bounded:

$$\sup_n |\ell_n(x_m)| \le C \, \|x_m\|_E < \infty.$$

Diagonal extraction of a subsequence. We extract one subsequence that makes *all* of these scalar sequences converge at once.

- For $m = 1$: $\{\ell_n(x_1)\}_n$ is a bounded real sequence, so by Bolzano–Weierstrass it has a convergent subsequence. Write its indices as $\{n_k^{(1)}\}_{k\ge1}$.

- For $m = 2$: along the previous subsequence, $\{\ell_{n_k^{(1)}}(x_2)\}_k$ is again bounded, so it has a further convergent subsequence $\{n_k^{(2)}\} \subseteq \{n_k^{(1)}\}$.
- Continuing inductively, at stage m we obtain $\{n_k^{(m)}\} \subseteq \{n_k^{(m-1)}\}$ along which $\ell_{n_k^{(m)}}(x_j)$ converges as $k \to \infty$ for each $j = 1, \dots, m$.

Now set the *diagonal* sequence $n_k := n_k^{(k)}$. For each fixed m, the tail $\{n_k : k \geq m\}$ is a subsequence of $\{n_k^{(m)}\}$, along which $\ell_{n_k}(x_m)$ converges. Therefore

$$\lim_{k\to\infty} \ell_{n_k}(x_m) \quad \text{exists for every } m \geq 1.$$

$\{\ell_{n_k}(x)\}_k$ is Cauchy for every $x \in E$. Fix $x \in E$ and $\varepsilon > 0$. Since $\{x_m\}$ is dense, choose m with $\|x - x_m\|_E < \varepsilon$. For any indices j, k, insert and subtract the values at x_m and use the triangle inequality:

$$|\ell_{n_k}(x) - \ell_{n_j}(x)| \leq \underbrace{|\ell_{n_k}(x) - \ell_{n_k}(x_m)|}_{\text{(I)}} + \underbrace{|\ell_{n_k}(x_m) - \ell_{n_j}(x_m)|}_{\text{(II)}} + \underbrace{|\ell_{n_j}(x_m) - \ell_{n_j}(x)|}_{\text{(III)}}.$$

For term (I), use linearity of ℓ_{n_k} and the bound $\|\ell_{n_k}\|_{E^*} \leq C$:

$$\text{(I)} = |\ell_{n_k}(x - x_m)| \leq \|\ell_{n_k}\|_{E^*}\, \|x - x_m\|_E \leq C\, \varepsilon.$$

By the same reasoning, (III) $\leq C\,\varepsilon$. Hence

$$|\ell_{n_k}(x) - \ell_{n_j}(x)| \leq 2C\varepsilon + |\ell_{n_k}(x_m) - \ell_{n_j}(x_m)|.$$

By diagonal extraction of a subsequence, $\{\ell_{n_k}(x_m)\}_k$ converges, so it is Cauchy; thus there is K such that

$$|\ell_{n_k}(x_m) - \ell_{n_j}(x_m)| < \varepsilon \quad \text{for all } j, k \geq K.$$

For such j, k we get $|\ell_{n_k}(x) - \ell_{n_j}(x)| \leq 2C\varepsilon + \varepsilon = (2C + 1)\,\varepsilon$. Since $\varepsilon > 0$ was arbitrary, the sequence $\{\ell_{n_k}(x)\}_k$ is Cauchy in $\mathbb{R}$.

The limit is a bounded linear functional. Since $\mathbb{R}$ is complete, for each $x \in E$ the limit

$$\ell(x) := \lim_{k\to\infty} \ell_{n_k}(x) \quad \text{exists in } \mathbb{R}.$$

Linearity: for $x, y \in E$ and scalars a, b, using linearity of each ℓ_{n_k} and of the limit,

$$\ell(ax + by) = \lim_k \ell_{n_k}(ax + by) = \lim_k \left[a\, \ell_{n_k}(x) + b\, \ell_{n_k}(y)\right] = a\, \ell(x) + b\, \ell(y).$$

Boundedness: taking the limit in $|\ell_{n_k}(x)| \leq C\|x\|_E$ gives

$$|\ell(x)| = \lim_{k\to\infty} |\ell_{n_k}(x)| \leq C\, \|x\|_E \quad \text{for every } x \in E,$$

so $\ell \in E^*$ with $\|\ell\|_{E^*} \leq C$. Finally, by the definition of ℓ we have $\ell_{n_k}(x) \to \ell(x)$ for every $x \in E$, which is exactly $\ell_{n_k} \overset{*}{\rightharpoonup} \ell$. $\square$

Strong and Weak Operator Convergence

Throughout, E and F are normed vector spaces, and $\mathcal{L}(E,F)$ is the space of bounded linear operators $T : E \to F$ with the operator norm $\|T\| = \sup_{\|x\|_E \le 1} \|T(x)\|_F$.

Strong and Weak Operator Convergence

Let $T_n, T \in \mathcal{L}(E,F)$.

a. *Strong operator convergence.* We say $T_n \to T$ *strongly*, written $T_n \xrightarrow{s} T$, if

$$\|T_n(x) - T(x)\|_F \longrightarrow 0 \quad \text{as } n \to \infty, \quad \text{for every } x \in E.$$

b. *Weak operator convergence.* We say $T_n \to T$ *weakly*, written $T_n \xrightarrow{w} T$, if $T_n(x) \rightharpoonup T(x)$ weakly in F for every $x \in E$; that is,

$$\ell\big(T_n(x)\big) \longrightarrow \ell\big(T(x)\big) \quad \text{as } n \to \infty, \quad \text{for every } \ell \in F^* \text{ and every } x \in E.$$

Note. On $\mathcal{L}(E,F)$ there are three types of convergence, ordered from strongest to weakest:

$$\|T_n - T\| \to 0 \quad \Longrightarrow \quad T_n \xrightarrow{s} T \quad \Longrightarrow \quad T_n \xrightarrow{w} T,$$

where $\|T_n - T\| \to 0$ is convergence in operator norm (uniform convergence). The first implication holds because, for each fixed x,

$$\|T_n(x) - T(x)\|_F \le \|T_n - T\| \, \|x\|_E \longrightarrow 0.$$

The second holds because norm convergence in F implies weak convergence in F: if $T_n(x) \to T(x)$ in F, then $T_n(x) \rightharpoonup T(x)$ in F. The converses fail in general, as the next example shows.

Example: Shift Operators on ℓ^2

Let $E = F = \ell^2$. For each $n \ge 1$ define the *right shift by n*, $T_n \in \mathcal{L}(\ell^2, \ell^2)$, and the *left shift by n*, $T_n^* \in \mathcal{L}(\ell^2, \ell^2)$, by

$$T_n(x) = \big(\underbrace{0, \dots, 0}_{n \text{ zeros}}, x_1, x_2, \dots\big), \quad T_n^*(x) = (x_{n+1}, x_{n+2}, \dots), \quad x = (x_k) \in \ell^2.$$

i. *T_n preserves the norm.* Right-shifting only inserts zeros, so

$$\|T_n(x)\|_{\ell^2}^2 = \sum_{k=1}^{\infty} |x_k|^2 = \|x\|_{\ell^2}^2,$$

hence $\|T_n(x)\|_{\ell^2} = \|x\|_{\ell^2}$ for all x, and in particular $\|T_n\| = 1$.

ii. *T_n^* is the adjoint of T_n.* For $x, y \in \ell^2$, the only nonzero entries of $T_n(x)$ are $(T_n x)_{n+j} = x_j$, so

$$\langle T_n(x), y\rangle = \sum_{k=1}^{\infty} (T_n x)_k \, y_k = \sum_{j=1}^{\infty} x_j \, y_{n+j} = \sum_{j=1}^{\infty} x_j \, (T_n^* y)_j = \langle x, T_n^*(y)\rangle.$$

iii. $T_n^* \xrightarrow{s} 0$ *but* $\|T_n^*\| = 1$. For a fixed $x = (x_k) \in \ell^2$,

$$\|T_n^*(x)\|_{\ell^2}^2 = \sum_{k=n+1}^{\infty} |x_k|^2 \longrightarrow 0 \quad \text{as } n \to \infty,$$

since it is the tail of the convergent series $\sum_{k=1}^{\infty} |x_k|^2 = \|x\|_{\ell^2}^2 < \infty$. So $T_n^* \xrightarrow{s} 0$ (strong operator convergence). However $\|T_n^*\| = 1$ for every n: applying T_n^* to the basis vector $e_{n+1} = (0, \dots, 0, 1, 0, \dots)$ gives $T_n^*(e_{n+1}) = e_1$, which has norm 1 while $\|e_{n+1}\| = 1$. Thus $T_n^* \not\to 0$ in operator norm. *Strong operator convergence does not imply norm convergence.*

iv. $T_n \xrightarrow{w} 0$ *but* $T_n \not\xrightarrow{s} 0$. Since $F = \ell^2$ is a Hilbert space, every functional on ℓ^2 is $\langle \cdot, y \rangle$ for some $y \in \ell^2$, so weak operator convergence $T_n \xrightarrow{w} 0$ means $\langle T_n(x), y \rangle \to 0$ for all $x, y \in \ell^2$. Using part ii and the Cauchy–Schwarz inequality,

$$|\langle T_n(x), y \rangle| = |\langle x, T_n^*(y) \rangle| \leq \|x\|_{\ell^2} \, \|T_n^*(y)\|_{\ell^2} \longrightarrow 0,$$

where the limit is 0 by part iii. Hence $T_n \xrightarrow{w} 0$. But T_n does *not* converge strongly to 0, because by part i

$$\|T_n(x) - 0\|_{\ell^2} = \|T_n(x)\|_{\ell^2} = \|x\|_{\ell^2} \not\to 0 \quad (x \neq 0).$$

Weak operator convergence does not imply strong operator convergence.

Remark. If $F = \mathbb{R}$, then $\mathcal{L}(E, \mathbb{R}) = E^*$ and the operators T_n are just functionals. In this case strong operator convergence $T_n \xrightarrow{s} T$ means $|T_n(x) - T(x)| \to 0$ for every $x \in E$, i.e. $T_n(x) \to T(x)$ for every $x \in E$, which is exactly weak-$*$ convergence $T_n \overset{*}{\rightharpoonup} T$ in E^*. So for scalar-valued operators the two notions coincide.

Proposition 3

Let $T \in \mathcal{L}(E, F)$ and suppose $x_n \rightharpoonup x$ weakly in E. Then $T(x_n) \rightharpoonup T(x)$ weakly in F.

Proof. We must show $\ell\big(T(x_n)\big) \to \ell\big(T(x)\big)$ for every $\ell \in F^*$. Fix $\ell \in F^*$. The composition $\ell \circ T : E \to \mathbb{R}$ is linear (a composition of linear maps) and bounded, since for every $x \in E$,

$$|(\ell \circ T)(x)| = |\ell(T(x))| \leq \|\ell\|_{F^*} \, \|T(x)\|_F \leq \|\ell\|_{F^*} \, \|T\| \, \|x\|_E.$$

Therefore $\ell \circ T \in E^*$. Applying the definition of weak convergence $x_n \rightharpoonup x$ to this particular functional $\ell \circ T$,

$$\ell\big(T(x_n)\big) = (\ell \circ T)(x_n) \longrightarrow (\ell \circ T)(x) = \ell\big(T(x)\big).$$

Since $\ell \in F^*$ was arbitrary, $T(x_n) \rightharpoonup T(x)$. □

Proposition 4

Let $T_n, T \in \mathcal{L}(E, F)$.

a. If $T_n \xrightarrow{s} T$, then $\|T\| \leq \liminf_{n \to \infty} \|T_n\|$.

b. Suppose $\sup_n \|T_n\| < \infty$ and $T_n(y) \to T(y)$ in F for every y in a dense subset $D \subseteq E$. Then $T_n \xrightarrow{s} T$.

Proof.

a. Fix $x \in E$. Strong convergence gives $T_n(x) \to T(x)$ in F, so by continuity of the norm $\|T_n(x)\|_F \to \|T(x)\|_F$; in particular the limit equals the lower limit:

$$\|T(x)\|_F = \lim_{n\to\infty} \|T_n(x)\|_F = \liminf_{n\to\infty} \|T_n(x)\|_F.$$

For each n we have $\|T_n(x)\|_F \le \|T_n\| \, \|x\|_E$. Taking $\liminf$ of both sides (the factor $\|x\|_E$ is a nonnegative constant),

$$\|T(x)\|_F = \liminf_{n\to\infty} \|T_n(x)\|_F \le \liminf_{n\to\infty} \left(\|T_n\| \, \|x\|_E\right) = \Big(\liminf_{n\to\infty} \|T_n\|\Big) \|x\|_E.$$

This holds for all $x \in E$. Dividing by $\|x\|_E$ for $x \neq 0$ and taking the supremum over such x,

$$\|T\| = \sup_{x\neq 0} \frac{\|T(x)\|_F}{\|x\|_E} \le \liminf_{n\to\infty} \|T_n\|. \qquad \square$$

b. Write $M := \sup_k \|T_k\| < \infty$. Fix $x \in E$ and let $\varepsilon > 0$. For any $y \in D$, the triangle inequality gives

$$\|T(x) - T_n(x)\|_F \le \|T(x) - T(y)\|_F + \|T(y) - T_n(y)\|_F + \|T_n(y) - T_n(x)\|_F.$$

Bound the first and third terms using boundedness of T and of each T_n:

$$\|T(x) - T(y)\|_F = \|T(x - y)\|_F \le \|T\| \, \|x - y\|_E,$$

$$\|T_n(y) - T_n(x)\|_F = \|T_n(y - x)\|_F \le \|T_n\| \, \|x - y\|_E \le M \, \|x - y\|_E.$$

Writing $A := \|T\| \, \|x - y\|_E$, $B := M \, \|x - y\|_E$, and $C := \|T(y) - T_n(y)\|_F$, we obtain

$$\|T(x) - T_n(x)\|_F \le \underbrace{(\|T\| + M) \, \|x - y\|_E}_{A+B} + \underbrace{\|T(y) - T_n(y)\|_F}_{C}.$$

Now choose the two pieces small. Since D is dense in E, pick $y \in D$ with

$$\|x - y\|_E \le \frac{\varepsilon}{2(\|T\| + M + 1)}, \quad \text{so that} \quad A + B = (\|T\| + M) \, \|x - y\|_E \le \frac{\varepsilon}{2}.$$

Fix this y. By hypothesis $T_n(y) \to T(y)$ in F (because $y \in D$), so there is N with

$$C = \|T(y) - T_n(y)\|_F \le \frac{\varepsilon}{2} \quad \text{for all } n \ge N.$$

Therefore, for all $n \ge N$,

$$\|T(x) - T_n(x)\|_F \le (A + B) + C \le \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon.$$

Since $\varepsilon > 0$ was arbitrary, $\|T_n(x) - T(x)\|_F \to 0$; and since $x \in E$ was arbitrary, $T_n \xrightarrow{s} T$. $\square$

Theorem 3

Let E be a Banach space and F a reflexive Banach space. Let $T_n \in \mathcal{L}(E,F)$ be such that for every $x \in E$ the sequence $\{T_n(x) : n \geq 1\}$ is weakly Cauchy in F. Then there exists $T \in \mathcal{L}(E,F)$ such that $T_n \xrightarrow{w} T$.

Proof. *Define T pointwise.* Fix $x \in E$. By assumption $\{T_n(x)\}_n$ is weakly Cauchy in F. Since F is a reflexive Banach space, every weakly Cauchy sequence in a reflexive Banach space has a weak limit, which gives an element of F, which we call $T(x)$, with

$$T_n(x) \rightharpoonup T(x) \quad \text{weakly in } F.$$

The weak limit is unique, so $T(x)$ is well defined. This defines a map $T : E \to F$.

T is linear. Let $x, x' \in E$ and let a, b be scalars. Each T_n is linear, so $T_n(ax + bx') = a\,T_n(x) + b\,T_n(x')$. Weak limits respect linear combinations: for any $\ell \in F^*$,

$$\ell\big(a\,T_n(x)+b\,T_n(x')\big) = a\,\ell(T_n(x))+b\,\ell(T_n(x')) \longrightarrow a\,\ell(T(x))+b\,\ell(T(x')) = \ell\big(aT(x)+bT(x')\big).$$

Thus $T_n(ax + bx') \rightharpoonup aT(x) + bT(x')$, and since weak limits are unique, $T(ax + bx') = aT(x) + bT(x')$.

T is bounded. First, each weakly Cauchy sequence $\{T_n(x)\}_n$ is bounded in F by Proposition 2(b):

$$\sup_n \|T_n(x)\|_F < \infty \quad \text{for every } x \in E.$$

Since E is a Banach space and each $T_n \in \mathcal{L}(E,F)$, the family $\{T_n\}$ is pointwise bounded on E. By the Uniform Boundedness Principle (Banach–Steinhaus),

$$\sup_n \|T_n\| =: C < \infty.$$

Next, applying weak lower semicontinuity of the norm to $T_n(x) \rightharpoonup T(x)$, then using $\|T_n(x)\|_F \leq \|T_n\|\,\|x\|_E$,

$$\|T(x)\|_F \leq \liminf_{n\to\infty} \|T_n(x)\|_F \leq \liminf_{n\to\infty} \big(\|T_n\|\,\|x\|_E\big) \leq \Big(\sup_k \|T_k\|\Big)\|x\|_E = C\,\|x\|_E.$$

So T is bounded with $\|T\| \leq C$, that is, $T \in \mathcal{L}(E,F)$.

From above $T_n(x) \rightharpoonup T(x)$ weakly in F for every $x \in E$, which is exactly $T_n \xrightarrow{w} T$. $\square$

Bounded Operators

The Banach Algebra of Bounded Operators

Throughout this section, E is a Banach space *over the complex field* $\mathbb{C}$. Consider the space of bounded linear operators from E to itself,

$$\mathcal{B} := \mathcal{L}(E) = \mathcal{L}(E,E).$$

This $\mathcal{B}$ is a Banach space under the operator norm $\|A\| = \sup_{\|x\|_E \le 1} \|A(x)\|_E$, and it also carries a *multiplication*, namely composition of operators,

$$AB := A \circ B, \quad \text{which satisfies} \quad \|AB\| \le \|A\|\,\|B\| \quad (A, B \in \mathcal{B}).$$

The inequality $\|AB\| \le \|A\|\,\|B\|$ is called submultiplicativity. A Banach space with such a multiplication is a *Banach algebra*. Moreover $\mathcal{B}$ has a unit, the identity map $I \in \mathcal{B}$:

$$IA = AI = A \quad \text{for every } A \in \mathcal{B}.$$

Invertible Operator

We say $T \in \mathcal{B}$ is *invertible* if there exists $S \in \mathcal{B}$ with

$$ST = TS = I.$$

Such an S is unique; we denote it $S = T^{-1}$.

Example: Neumann Series

Let $T \in \mathcal{B}$ with $\|T\| < 1$. Then $I - T$ is invertible, and

$$(I - T)^{-1} = \sum_{n=0}^{\infty} T^n \quad (\text{with the convention } T^0 := I).$$

Proof. *The series converges in $\mathcal{B}$.* By submultiplicativity, $\|T^n\| \le \|T\|^n$ for every $n \ge 0$ (and $\|T^0\| = \|I\| = 1 = \|T\|^0$). Since $\|T\| < 1$, the geometric series of norms converges:

$$\sum_{n=0}^{\infty} \|T^n\| \le \sum_{n=0}^{\infty} \|T\|^n = \frac{1}{1 - \|T\|} < \infty.$$

Let $S_n := \sum_{k=0}^{n} T^k$ be the partial sums. For $m > n$,

$$\|S_m - S_n\| = \left\| \sum_{k=n+1}^{m} T^k \right\| \le \sum_{k=n+1}^{m} \|T^k\| \le \sum_{k=n+1}^{m} \|T\|^k \xrightarrow[n,m\to\infty]{} 0,$$

because the right-hand side is a tail of a convergent series. So $\{S_n\}$ is Cauchy in $\mathcal{B}$. Since E is a Banach space, $\mathcal{B} = \mathcal{L}(E)$ is also complete, so $\{S_n\}$ converges to some

$$S := \sum_{n=0}^{\infty} T^n \in \mathcal{B}.$$

S is a two-sided inverse of $I - T$. Compute the product $S_n(I - T)$ and watch the terms telescope:

$$S_n(I - T) = \sum_{k=0}^{n} T^k (I - T) = \sum_{k=0}^{n} \left(T^k - T^{k+1}\right) = \underbrace{\sum_{k=0}^{n} T^k}_{I + T + \cdots + T^n} - \underbrace{\sum_{k=0}^{n} T^{k+1}}_{T + T^2 + \cdots + T^{n+1}} = I - T^{n+1}.$$

Now let $n \to \infty$ in the equation $S_n(I - T) = I - T^{n+1}$.

- *Right-hand side.* $\|T^{n+1}\| \le \|T\|^{n+1} \to 0$, so $T^{n+1} \to 0$ in $\mathcal{B}$, and therefore $I - T^{n+1} \to I$.
- *Left-hand side.* Multiplication is continuous: if $A_n \to A$ in $\mathcal{B}$, then for any fixed B,

$$\|A_n B - AB\| = \|(A_n - A)B\| \le \|A_n - A\| \, \|B\| \to 0,$$

  so $A_n B \to AB$. Taking $A_n = S_n$, $A = S$, and $B = I - T$ gives $S_n(I - T) \to S(I - T)$.

Both sides converge, so their limits agree: $S(I - T) = I$. The same telescoping with the factor on the left gives $(I - T)S_n = I - T^{n+1} \to I$, and by continuity $(I - T)S_n \to (I - T)S$, so $(I - T)S = I$. Therefore $S = (I - T)^{-1}$. $\square$

Remark. In this computation S_n and $I - T$ commute, because both are polynomials in T; so the left and right products are equal and a single calculation handles both sides. In general, to prove an operator is invertible one must check *both* $ST = I$ and $TS = I$ separately.

Spectrum and Resolvent

We now study, for a fixed $T \in \mathcal{B}$ and a scalar $\lambda \in \mathbb{C}$, the invertibility of

$$\lambda - T := \lambda I - T \in \mathcal{B}.$$

Finite dimensions. Let $E = \mathbb{C}^d$, so T is a $d \times d$ complex matrix. Then

$$\lambda I - T \text{ is invertible} \iff \det(\lambda I - T) \neq 0 \iff \lambda \text{ is not an eigenvalue of } T,$$

and λ is an eigenvalue exactly when $\det(\lambda I - T) = 0$. In infinite dimensions there is no determinant, so we keep invertibility of $\lambda I - T$ as the basic notion and make the following definitions.

Resolvent Set, Resolvent, Spectrum

Let $T \in \mathcal{B} = \mathcal{L}(E)$.

a. The *resolvent set* of T is

$$\rho(T) = \{\lambda \in \mathbb{C} : \lambda I - T \text{ is invertible in } \mathcal{B}\}.$$

   For $\lambda \in \rho(T)$, the operator

$$R_\lambda(T) := (\lambda I - T)^{-1} \in \mathcal{B}$$

   is called the *resolvent* of T at λ.

b. The *spectrum* of T is the complement

$$\sigma(T) = \mathbb{C} \setminus \rho(T) = \{\lambda \in \mathbb{C} : \lambda I - T \text{ is not invertible}\}.$$

Eigenvalue, Eigenvector

A *nonzero* vector $x \in E$, $x \neq 0$, is an *eigenvector* of T if $T(x) = \lambda x$ for some $\lambda \in \mathbb{C}$; this λ is then an *eigenvalue* of T. If λ is an eigenvalue, then $(\lambda I - T)(x) = 0$ with $x \neq 0$, so $\lambda I - T$ is not injective, hence not invertible, and therefore $\lambda \in \sigma(T)$. Thus every eigenvalue lies in the spectrum.

The Resolvent Identity and Commuting

Let $\lambda, \mu \in \rho(T)$. The idea is to rewrite each resolvent by inserting the other one. Since $(\mu I - T)R_\mu(T) = I$ and $R_\lambda(T)(\lambda I - T) = I$,

$$R_\lambda(T) = R_\lambda(T)\,\underbrace{(\mu I - T)R_\mu(T)}_{=I}, \quad R_\mu(T) = \underbrace{R_\lambda(T)(\lambda I - T)}_{=I}\,R_\mu(T).$$

Subtracting these two expressions and factoring out $R_\lambda(T)$ on the left and $R_\mu(T)$ on the right,

$$R_\lambda(T) - R_\mu(T) = R_\lambda(T)\big[(\mu I - T) - (\lambda I - T)\big]R_\mu(T).$$

The middle bracket simplifies: $(\mu I - T) - (\lambda I - T) = \mu I - \lambda I = (\mu - \lambda)I$. Therefore

$$R_\lambda(T) - R_\mu(T) = (\mu - \lambda)\,R_\lambda(T)R_\mu(T). \tag{1}$$

Now swap the names λ and μ in (1):

$$R_\mu(T) - R_\lambda(T) = (\lambda - \mu)\,R_\mu(T)R_\lambda(T), \quad \text{i.e.} \quad R_\lambda(T) - R_\mu(T) = (\mu - \lambda)\,R_\mu(T)R_\lambda(T). \tag{2}$$

Comparing (1) and (2) gives $(\mu - \lambda)R_\lambda(T)R_\mu(T) = (\mu - \lambda)R_\mu(T)R_\lambda(T)$. When $\lambda \neq \mu$ we cancel the nonzero scalar $(\mu - \lambda)$ to obtain

$$R_\lambda(T)R_\mu(T) = R_\mu(T)R_\lambda(T)$$

and for $\lambda = \mu$ this is trivial. Using this commutativity in (1) also yields the stated form $R_\lambda(T) - R_\mu(T) = (\mu - \lambda)R_\mu(T)R_\lambda(T)$.

Theorem 1

Let $T \in \mathcal{B} = \mathcal{L}(E)$, with E a Banach space over $\mathbb{C}$. Then:

a. The resolvent set $\rho(T) \subseteq \mathbb{C}$ is open, and the map $\lambda \mapsto R_\lambda(T)$ is a $\mathcal{B}$-valued analytic function on $\rho(T)$: for each $\lambda_0 \in \rho(T)$ there are coefficients $c_n \in \mathcal{B}$ with

$$R_\lambda(T) = \sum_{n=0}^{\infty} c_n\,(\lambda - \lambda_0)^n \quad \text{for all } \lambda \text{ sufficiently close to } \lambda_0.$$

Moreover, for all $\lambda, \mu \in \rho(T)$ the *resolvent identity* holds,

$$R_\lambda(T) - R_\mu(T) = (\mu - \lambda)\,R_\mu(T)\,R_\lambda(T),$$

and the resolvents commute: $R_\mu(T)\,R_\lambda(T) = R_\lambda(T)\,R_\mu(T)$.

b. The spectrum $\sigma(T) \subseteq \mathbb{C}$ is a nonempty compact set, and its *spectral radius* satisfies

$$r(T) := \sup\{|\lambda| : \lambda \in \sigma(T)\} \leq \|T\|.$$

Proof of Part a. *Motivation (treat everything as numbers).* Fix $\lambda_0 \in \rho(T)$. If λ, λ_0, and T were ordinary numbers, we would write, for λ near λ_0,

$$\frac{1}{\lambda - T} = \frac{1}{(\lambda_0 - T) - (\lambda_0 - \lambda)} = \frac{1}{\lambda_0 - T} \cdot \frac{1}{1 - \frac{\lambda_0 - \lambda}{\lambda_0 - T}} = \frac{1}{\lambda_0 - T} \sum_{n=0}^{\infty} \left(\frac{\lambda_0 - \lambda}{\lambda_0 - T}\right)^n,$$

the geometric series being valid when $\left|\frac{\lambda_0 - \lambda}{\lambda_0 - T}\right| < 1$. This is only a guide; we now do the same thing rigorously with operators, replacing $\frac{1}{\lambda_0 - T}$ by the operator $R_{\lambda_0}(T)$.

Build a candidate inverse. Assume

$$|\lambda - \lambda_0| < \frac{1}{\|R_{\lambda_0}(T)\|}, \quad \text{equivalently} \quad |\lambda - \lambda_0| \, \|R_{\lambda_0}(T)\| < 1.$$

$R_{\lambda_0}(T)$ is invertible, so it is not the zero operator and $\|R_{\lambda_0}(T)\| > 0$. Then the operator $(\lambda_0 - \lambda)R_{\lambda_0}(T)$ has norm

$$\|(\lambda_0 - \lambda)R_{\lambda_0}(T)\| = |\lambda_0 - \lambda| \, \|R_{\lambda_0}(T)\| < 1,$$

so by the Neumann Series Example, $I - (\lambda_0 - \lambda)R_{\lambda_0}(T)$ is invertible and

$$\left[I - (\lambda_0 - \lambda)R_{\lambda_0}(T)\right]^{-1} = \sum_{n=0}^{\infty} (\lambda_0 - \lambda)^n \left[R_{\lambda_0}(T)\right]^n. \tag{$*$}$$

Define the candidate operator

$$\widetilde{R}_\lambda(T) := R_{\lambda_0}(T) \left[I - (\lambda_0 - \lambda)R_{\lambda_0}(T)\right]^{-1} \in \mathcal{B}.$$

A commuting fact and a rewriting. By $(*)$, the operator $[I - (\lambda_0 - \lambda)R_{\lambda_0}(T)]^{-1}$ is a norm-convergent series of powers of $R_{\lambda_0}(T)$. Since $R_{\lambda_0}(T)$ commutes with each of its own powers, and multiplication is continuous, $R_{\lambda_0}(T)$ commutes with this inverse:

$$R_{\lambda_0}(T) \left[I - (\lambda_0 - \lambda)R_{\lambda_0}(T)\right]^{-1} = \left[I - (\lambda_0 - \lambda)R_{\lambda_0}(T)\right]^{-1} R_{\lambda_0}(T). \tag{$**$}$$

Also write $\lambda I - T$ in terms of λ_0:

$$\lambda I - T = (\lambda_0 I - T) - (\lambda_0 - \lambda)I, \quad \text{since } \lambda_0 I - (\lambda_0 - \lambda)I = \lambda I.$$

Check that $\widetilde{R}_\lambda(T)$ inverts $\lambda I - T$. Using the rewriting and then $(**)$ to move the inverse to the front,

$$\begin{aligned}
\widetilde{R}_\lambda(T)\,(\lambda I - T) &= R_{\lambda_0}(T) \left[I - (\lambda_0 - \lambda)R_{\lambda_0}(T)\right]^{-1} \left[(\lambda_0 I - T) - (\lambda_0 - \lambda)I\right] \\
&= \left[I - (\lambda_0 - \lambda)R_{\lambda_0}(T)\right]^{-1} R_{\lambda_0}(T) \left[(\lambda_0 I - T) - (\lambda_0 - \lambda)I\right] \\
&= \left[I - (\lambda_0 - \lambda)R_{\lambda_0}(T)\right]^{-1} \left[R_{\lambda_0}(T)(\lambda_0 I - T) - (\lambda_0 - \lambda)R_{\lambda_0}(T)\right].
\end{aligned}$$

Now $R_{\lambda_0}(T)(\lambda_0 I - T) = (\lambda_0 I - T)^{-1}(\lambda_0 I - T) = I$, so the bracket simplifies to

$$R_{\lambda_0}(T)(\lambda_0 I - T) - (\lambda_0 - \lambda)R_{\lambda_0}(T) = I - (\lambda_0 - \lambda)R_{\lambda_0}(T),$$

and therefore

$$\widetilde{R}_\lambda(T)\,(\lambda I - T) = \big[I - (\lambda_0 - \lambda)R_{\lambda_0}(T)\big]^{-1}\big[I - (\lambda_0 - \lambda)R_{\lambda_0}(T)\big] = I.$$

The identical computation with the two factors in the opposite order (again using $(**)$) gives $(\lambda I - T)\,\widetilde{R}_\lambda(T) = I$. Hence $\lambda I - T$ is invertible, with inverse $\widetilde{R}_\lambda(T)$.

Openness and the power series. We have shown: whenever $|\lambda - \lambda_0| < 1/\|R_{\lambda_0}(T)\|$, the point λ lies in $\rho(T)$, and

$$R_\lambda(T) = \widetilde{R}_\lambda(T) = R_{\lambda_0}(T)\sum_{n=0}^{\infty}(\lambda_0 - \lambda)^n[R_{\lambda_0}(T)]^n = \sum_{n=0}^{\infty}(-1)^n\,[R_{\lambda_0}(T)]^{n+1}\,(\lambda - \lambda_0)^n,$$

where we used $(\lambda_0 - \lambda)^n = (-1)^n(\lambda - \lambda_0)^n$. So every $\lambda_0 \in \rho(T)$ is the center of an open disk of radius $1/\|R_{\lambda_0}(T)\|$ contained in $\rho(T)$; this proves $\rho(T)$ is open. On that disk, $R_\lambda(T)$ is given by a norm-convergent power series in $(\lambda - \lambda_0)$ with coefficients $c_n = (-1)^n[R_{\lambda_0}(T)]^{n+1} \in \mathcal{B}$, so $R_\lambda(T)$ is analytic on $\rho(T)$.

The resolvent identity and commuting. Let $\lambda, \mu \in \rho(T)$. Start from the elementary identity

$$(\mu I - T) - (\lambda I - T) = (\mu - \lambda)I.$$

Multiply on the left by $R_\mu(T) = (\mu I - T)^{-1}$ and on the right by $R_\lambda(T) = (\lambda I - T)^{-1}$:

$$R_\mu(T)\big[(\mu I - T) - (\lambda I - T)\big]R_\lambda(T) = (\mu - \lambda)\,R_\mu(T)R_\lambda(T).$$

Expand the left-hand side into two terms and simplify each:

$$\underbrace{R_\mu(T)(\mu I - T)}_{=I}\,R_\lambda(T) - R_\mu(T)\,\underbrace{(\lambda I - T)R_\lambda(T)}_{=I} = R_\lambda(T) - R_\mu(T).$$

This gives the resolvent identity

$$R_\lambda(T) - R_\mu(T) = (\mu - \lambda)\,R_\mu(T)R_\lambda(T).$$

Swapping the roles of λ and μ in this identity gives $R_\mu(T) - R_\lambda(T) = (\lambda - \mu)R_\lambda(T)R_\mu(T)$, that is, $R_\lambda(T) - R_\mu(T) = (\mu - \lambda)R_\lambda(T)R_\mu(T)$. Comparing the two expressions for $R_\lambda(T) - R_\mu(T)$, if $\mu \neq \lambda$ we may cancel the nonzero factor $(\mu - \lambda)$ to get $R_\mu(T)R_\lambda(T) = R_\lambda(T)R_\mu(T)$; for $\mu = \lambda$ this is trivial. So the resolvents commute. □

Proof of Part b. *Large λ lie in $\rho(T)$.* As motivation, if λ and T were numbers we would write, for $|\lambda| > |T|$,

$$\frac{1}{\lambda - T} = \frac{1}{\lambda}\cdot\frac{1}{1 - T/\lambda} = \frac{1}{\lambda}\sum_{n=0}^{\infty}\frac{T^n}{\lambda^n}.$$

For operators, fix $\lambda \in \mathbb{C}$ with $|\lambda| > \|T\|$. Then $\left\|\frac{1}{\lambda}T\right\| = \|T\|/|\lambda| < 1$, so by the Neumann Series Example $I - \frac{1}{\lambda}T$ is invertible. Set

$$\widetilde{R}_\lambda(T) := \frac{1}{\lambda}\Big(I - \tfrac{1}{\lambda}T\Big)^{-1}.$$

Since $\lambda I - T = \lambda\big(I - \frac{1}{\lambda}T\big)$, a direct check gives

$$\widetilde{R}_\lambda(T)(\lambda I - T) = \tfrac{1}{\lambda}\big(I - \tfrac{1}{\lambda}T\big)^{-1}\lambda\big(I - \tfrac{1}{\lambda}T\big) = \big(I - \tfrac{1}{\lambda}T\big)^{-1}\big(I - \tfrac{1}{\lambda}T\big) = I,$$

and the same on the other side. So $\lambda \in \rho(T)$ and $R_\lambda(T) = \widetilde{R}_\lambda(T)$ for every $|\lambda| > \|T\|$.

$\sigma(T)$ is compact, and $r(T) \le \|T\|$. Previous part says every λ with $|\lambda| > \|T\|$ is in $\rho(T)$. Equivalently,

$$\sigma(T) \subseteq \{z \in \mathbb{C} : |z| \le \|T\|\},$$

so $\sigma(T)$ is bounded and $r(T) = \sup\{|\lambda| : \lambda \in \sigma(T)\} \le \|T\|$. Also $\sigma(T) = \mathbb{C} \setminus \rho(T)$ is closed, being the complement of the open set $\rho(T)$. A closed and bounded subset of $\mathbb{C}$ is compact (Heine–Borel), so $\sigma(T)$ is compact.

$\sigma(T)$ is nonempty. Suppose, for contradiction, that $\sigma(T) = \emptyset$, i.e. $\rho(T) = \mathbb{C}$. Then $\lambda \mapsto R_\lambda(T)$ is analytic on all of $\mathbb{C}$. For $|\lambda| > \|T\|$, expanding the Neumann series from Previous part,

$$R_\lambda(T) = \frac{1}{\lambda}\Big(I - \tfrac{1}{\lambda}T\Big)^{-1} = \frac{1}{\lambda}\sum_{n=0}^{\infty}\frac{T^n}{\lambda^n},$$

$$\|R_\lambda(T)\| \le \frac{1}{|\lambda|}\sum_{n=0}^{\infty}\frac{\|T\|^n}{|\lambda|^n} = \frac{1}{|\lambda|}\cdot\frac{1}{1 - \|T\|/|\lambda|} = \frac{1}{|\lambda| - \|T\|} \xrightarrow[|\lambda|\to\infty]{} 0.$$

Thus $R_\lambda(T)$ is an entire $\mathcal{B}$-valued function that is bounded on $\mathbb{C}$ (it is continuous, hence bounded on the disk $|\lambda| \le \|T\| + 1$, and small outside) and tends to 0 at infinity. By Liouville's theorem (for Banach-space-valued analytic functions) a bounded entire function is constant; since it also tends to 0, that constant is 0, so $R_\lambda(T) \equiv 0$. But $R_\lambda(T) = (\lambda I - T)^{-1}$ means $(\lambda I - T)R_\lambda(T) = I$; if $R_\lambda(T) = 0$ then $(\lambda I - T)\cdot 0 = 0 \ne I$, a contradiction. Hence $\sigma(T) \ne \emptyset$. $\square$

Spectral Radius Formula

Let $T \in \mathcal{L}(E)$. Then the limit $\lim_{n\to\infty}\|T^n\|^{1/n}$ exists and equals the spectral radius:

$$r(T) = \lim_{n\to\infty}\|T^n\|^{1/n} \quad \big(\le \|T\|\big).$$

The bound $\|T^n\|^{1/n} \le \|T\|$ holds for every n, because $\|T^n\| \le \|T\|^n$ by submultiplicativity; hence the limit is $\le \|T\|$, which refines the estimate $r(T) \le \|T\|$ from Theorem 1(b). The equality $r(T) = \lim_n \|T^n\|^{1/n}$ is proved using the analyticity of the resolvent from Theorem 1; we take it as stated here.

Example: Invertible Operators

Let $T \in \mathcal{L}(E)$ be invertible. Then $0 \notin \sigma(T)$, since $0 \cdot I - T = -T$ is invertible. Then:

a. $\sigma(T^{-1}) = \{1/\lambda : \lambda \in \sigma(T)\}$.

b. If $S \in \mathcal{L}(E)$ commutes with T, that is $ST = TS$, then S commutes with T^{-1}: $ST^{-1} = T^{-1}S$.

Proof.

a. We want to show $\lambda \in \sigma(T^{-1})$ if and only if $1/\lambda \in \sigma(T)$. Equivalently, we want to show $\lambda \in \rho(T^{-1})$ if and only if $1/\lambda \in \rho(T)$. Since T and T^{-1} are invertible, $0 \notin \sigma(T)$ and $0 \notin \sigma(T^{-1})$, so we only consider $\lambda \neq 0$. For such λ, factor
$$\lambda I - T^{-1} = -\lambda T^{-1}\Big(\tfrac{1}{\lambda} I - T\Big).$$
Indeed, expanding the right side: $-\lambda T^{-1}\big(\frac{1}{\lambda}I - T\big) = -\lambda T^{-1} \cdot \frac{1}{\lambda} I + \lambda T^{-1}T = -T^{-1} + \lambda I = \lambda I - T^{-1}$. The factor $-\lambda T^{-1}$ is invertible (a nonzero scalar times an invertible operator). Therefore $\lambda I - T^{-1}$ is invertible if and only if $\frac{1}{\lambda}I - T$ is invertible:
$$\lambda \in \rho(T^{-1}) \iff \tfrac{1}{\lambda} \in \rho(T).$$
Taking complements, $\lambda \in \sigma(T^{-1}) \iff \frac{1}{\lambda} \in \sigma(T)$, which is exactly $\sigma(T^{-1}) = \{1/\lambda : \lambda \in \sigma(T)\}$.

b. We want $ST^{-1} = T^{-1}S$. Since T is invertible, it is left-cancellable, so it is enough to show $T(ST^{-1}) = T(T^{-1}S)$. Compute both sides using $TS = ST$:
$$T(ST^{-1}) = (TS)T^{-1} = (ST)T^{-1} = S(TT^{-1}) = S, \quad T(T^{-1}S) = (TT^{-1})S = S.$$
Both equal S, so $T(ST^{-1}) = T(T^{-1}S)$; multiplying on the left by T^{-1} gives $ST^{-1} = T^{-1}S$. $\square$

We extend the polynomial calculus $P \mapsto P(T)$ to entire functions. Recall that an *entire* function $f : \mathbb{C} \to \mathbb{C}$ is given by a power series
$$f(z) = \sum_{n=0}^{\infty} c_n z^n, \quad z \in \mathbb{C},$$
with infinite radius of convergence; equivalently, $\sum_{n=0}^{\infty} |c_n|\, r^n < \infty$ for every $r > 0$.

Definition ($f(T)$)

Given an entire function $f(z) = \sum_{n=0}^{\infty} c_n z^n$ and $T \in \mathcal{L}(E)$, set
$$f(T) := \sum_{n=0}^{\infty} c_n T^n \in \mathcal{L}(E).$$
This series converges absolutely in $\mathcal{L}(E)$: taking $r = \|T\|$ in the convergence condition,
$$\sum_{n=0}^{\infty} \|c_n T^n\| \le \sum_{n=0}^{\infty} |c_n|\, \|T\|^n < \infty,$$
and since $\mathcal{L}(E)$ is complete, the partial sums converge.

Spectral Mapping Theorem

Let $T \in \mathcal{L}(E)$ and let $f : \mathbb{C} \to \mathbb{C}$ be entire. Then

$$\sigma(f(T)) = f(\sigma(T)) = \{f(\lambda) : \lambda \in \sigma(T)\}.$$

Proof. For f a polynomial, we prove the two parts. The general entire case is the same idea, with extra care about convergence of the infinite series. Let $f(z) = \sum_{n=0}^{N} c_n z^n$ be a polynomial of degree N. Then $f(T) = \sum_{n=0}^{N} c_n T^n$ is a polynomial in T. We show $\sigma(f(T)) = f(\sigma(T))$.

1. $f(\sigma(T)) \subseteq \sigma(f(T))$. Let $\alpha \in \sigma(T)$; we show $f(\alpha) \in \sigma(f(T))$. View $f(\alpha) - f(z)$ as a polynomial in z. It vanishes at $z = \alpha$, so $(z - \alpha)$ divides it; writing the linear factor as $(\alpha - z)$, there is a polynomial g with

   $$f(\alpha) - f(z) = (\alpha - z)\, g(z).$$

   Substituting the operator T for z (a constant $f(\alpha)$ becomes $f(\alpha)I$), the polynomial calculus gives

   $$f(\alpha)I - f(T) = (\alpha I - T)\, g(T).$$

   Suppose, for contradiction, that $G := f(\alpha)I - f(T)$ is invertible. Then

   $$I = G^{-1}G = G^{-1}(\alpha I - T)\, g(T).$$

   All of G^{-1}, $g(T)$, and $\alpha I - T$ are polynomials in T, so they commute with one another. Rearranging,

   $$I = (\alpha I - T)\big[g(T)G^{-1}\big] = \big[g(T)G^{-1}\big](\alpha I - T),$$

   which exhibits $g(T)G^{-1}$ as a two-sided inverse of $\alpha I - T$. So $\alpha \in \rho(T)$, contradicting $\alpha \in \sigma(T)$. Hence $f(\alpha)I - f(T)$ is not invertible, i.e. $f(\alpha) \in \sigma(f(T))$.

2. $\sigma(f(T)) \subseteq f(\sigma(T))$. Suppose, for contradiction, there is $\mu \in \sigma(f(T))$ with $\mu \notin f(\sigma(T))$. Factor the polynomial $\mu - f(z)$ (of degree $N \geq 1$, with leading coefficient $c \neq 0$) into linear factors over $\mathbb{C}$:

   $$\mu - f(z) = c \prod_{k=1}^{N} (z_k - z),$$

   where $z_1, \ldots, z_N$ are its roots, so $f(z_k) = \mu$ for each k. None of these roots lies in $\sigma(T)$: if some $z_k \in \sigma(T)$, then $\mu = f(z_k) \in f(\sigma(T))$, contradicting $\mu \notin f(\sigma(T))$. Hence every $z_k \in \rho(T)$, so each $z_k I - T$ is invertible. Substituting T for z,

   $$\mu I - f(T) = c \prod_{k=1}^{N} (z_k I - T).$$

   This is a nonzero scalar times a product of invertible operators, which all commute, being polynomials in T, so it is invertible. Therefore $\mu \in \rho(f(T))$, contradicting $\mu \in \sigma(f(T))$.

The two parts give $\sigma(f(T)) = f(\sigma(T))$. □

Example: Spectra of ST and TS

Let $S, T \in \mathcal{L}(E)$. Then $\sigma(ST)$ and $\sigma(TS)$ can differ only at the point 0:

$$\sigma(ST) \setminus \{0\} = \sigma(TS) \setminus \{0\}.$$

Proof. It suffices to show that for $\lambda \neq 0$, if $\lambda \in \rho(TS)$ then $\lambda \in \rho(ST)$; the reverse inclusion follows by swapping the roles of S and T. So fix $\lambda \neq 0$ with $\lambda \in \rho(TS)$, and put

$$A := (\lambda I - TS)^{-1}, \quad \text{so} \quad (\lambda I - TS)A = A(\lambda I - TS) = I.$$

We claim that $\frac{1}{\lambda}(I + SAT)$ is the inverse of $\lambda I - ST$. Compute

$$(\lambda I - ST)(I + SAT) = \lambda I + \lambda SAT - ST - ST\,SAT = \lambda I - ST + \big[\lambda SAT - S(TS)AT\big].$$

The bracketed terms combine, using $\lambda A - (TS)A = (\lambda I - TS)A = I$:

$$\lambda SAT - S(TS)AT = S\big[\lambda A - (TS)A\big]T = S\big[(\lambda I - TS)A\big]T = S\,I\,T = ST.$$

Therefore $(\lambda I - ST)(I + SAT) = \lambda I - ST + ST = \lambda I$. The analogous computation (using $A(\lambda I - TS) = I$) gives $(I + SAT)(\lambda I - ST) = \lambda I$. Hence

$$(\lambda I - ST) \cdot \tfrac{1}{\lambda}(I + SAT) = \tfrac{1}{\lambda}(I + SAT) \cdot (\lambda I - ST) = I,$$

so $\lambda I - ST$ is invertible and $\lambda \in \rho(ST)$. This proves $\rho(TS) \setminus \{0\} \subseteq \rho(ST)$; by symmetry the two agree off 0, and so do their complements. $\square$