| CIS3333: Mathematics of Machine Learning | Fall 2025 |
|---|---|

# Lecture: Function Spaces and Hilbert Spaces

*Date: October 20th, 2025* — *Author: Surbhi Goel*

Last time we saw eigendecomposition and PCA—finding optimal bases for data. All of our vector spaces so far have been finite-dimensional: $\mathbb{R}^n$, polynomials of degree $\leq n$, etc. Today we extend everything to **infinite dimensions**. This is the key to kernel methods: we'll learn in infinite-dimensional spaces without computing infinite coordinates.

# 1 From Vectors to Functions

## 1.1 Function Spaces as Vector Spaces

- A **function space** is a vector space where elements are functions.
- **Example:** Polynomials $\mathcal{P}_n = \{p(x) = a_0 + a_1 x + \cdots + a_n x^n : a_i \in \mathbb{R}\}$
  - Vector addition: $(p + q)(x) = p(x) + q(x)$
  - Scalar multiplication: $(\lambda p)(x) = \lambda p(x)$
  - Finite basis: $\{1, x, x^2, \ldots, x^n\}$, so $\dim(\mathcal{P}_n) = n + 1$
- **Example:** Continuous functions $C[0,1] = \{f : [0,1] \to \mathbb{R} \text{ continuous}\}$
  - Still has vector addition and scalar multiplication
  - But now **infinite-dimensional**—no finite basis
- You can think of a function $f : [0,1] \to \mathbb{R}$ as an "infinite vector"—one coordinate $f(x)$ for every $x \in [0,1]$.

## 1.2 Inner Products on Function Spaces

- For functions $f, g : [a,b] \to \mathbb{R}$, a natural inner product is:

$$\langle f, g \rangle = \int_a^b f(x)g(x)dx$$

- This satisfies all inner product axioms:
  - Bilinearity: $\langle \alpha f_1 + \beta f_2, g \rangle = \alpha \langle f_1, g \rangle + \beta \langle f_2, g \rangle$
  - Symmetry: $\langle f, g \rangle = \langle g, f \rangle$
  - Positive definite: $\langle f, f \rangle = \int_a^b f(x)^2 dx \geq 0$, with equality iff $f = 0$
- This induces a norm: $\|f\|_2 = \sqrt{\langle f, f \rangle} = \sqrt{\int_a^b f(x)^2 dx}$ (the $L^2$ norm)
- Functions are orthogonal if $\langle f, g \rangle = 0$.

- An orthonormal basis for a function space satisfies norm 1 for each basis function and orthogonality between each pair of basis functions.
- Just like in $\mathbb{R}^n$, we can project functions onto subspaces. If $U$ is a subspace of a function space, any function $f$ decomposes as:

$$f = f_U + f_{U^\perp}$$

  where $f_U \in U$ is the projection and $f_{U^\perp}$ is orthogonal to all of $U$. Alternatively, $f_U$ is the closest point in $U$ to $f$:

$$f_U = \arg\min_{g \in U} \|f - g\|$$

## 2 Hilbert Spaces

Not all function spaces are well-behaved. We need **completeness** to ensure limits exist.

**Motivation: why completeness matters.** Sometimes a sequence of functions gets closer and closer together, but the "limit" isn't actually in the function space. This is problematic when we want to find best approximations or solutions to learning problems since the answer might not exist in our space!

**Example: approximating a step function.** Consider continuous functions $C[0,1]$ with the $L^2$ norm $\|f\|_2 = \sqrt{\int_0^1 f(x)^2 dx}$.

Define a sequence of continuous functions that approximate a step function:

$$f_n(x) = \begin{cases} 0 & \text{if } x \in [0, \frac{1}{2} - \frac{1}{n}] \\ n\left(x - \frac{1}{2} + \frac{1}{n}\right) & \text{if } x \in [\frac{1}{2} - \frac{1}{n}, \frac{1}{2} + \frac{1}{n}] \\ 1 & \text{if } x \in [\frac{1}{2} + \frac{1}{n}, 1] \end{cases}$$

- Each $f_n$ is continuous (smooth ramp from 0 to 1 near $x = 1/2$)
- For $f_1$: transition region has width $2/1 = 2$
- For $f_{100}$: transition region has width $2/100 = 0.02$
- As $n \to \infty$, transition region gets narrower (width $2/n \to 0$), making the ramp steeper

The functions in the sequence get closer and closer together: you can verify that $\|f_m - f_n\|_2 \to 0$ as $m, n \to \infty$. But what does the sequence converge to?

The "limit" is the step function:

$$f(x) = \begin{cases} 0 & \text{if } x < \frac{1}{2} \\ 1 & \text{if } x \geq \frac{1}{2} \end{cases}$$

This is discontinuous. It jumps from 0 to 1 at $x = 1/2$. So $f \notin C[0,1]$! The space $C[0,1]$ has a "hole" where $f$ should be. It's incomplete.

**The solution: completeness.** We need spaces where all such limits exist.

- $L^2[0,1]$ completes $C[0,1]$ by including discontinuous functions with finite $L^2$ norm
- The step function is in $L^2[0,1]$: $\int_0^1 f(x)^2 dx = 1/2 < \infty$
- Working in complete spaces ensures that best approximations and solutions always exist

## 2.1 Formal Definitions

**Cauchy sequences.** A sequence $(x_n)$ in a normed space is a *Cauchy sequence* if elements get arbitrarily close together:

$$\forall \epsilon > 0, \exists N : \text{for all } m, n > N, \|x_m - x_n\| < \epsilon$$

**Complete spaces.** A normed space is *complete* if every Cauchy sequence converges to a limit *within the space*. For example:

- $\mathbb{R}$ is complete (every Cauchy sequence of reals converges to a real)
- $\mathbb{Q}$ is not complete (e.g., decimal approximations of $\sqrt{2}$)
- $L^2[0,1]$ is complete
- $C[0,1]$ with $\|\cdot\|_2$ is not complete (step function example)

**Hilbert spaces:** A Hilbert space is a complete inner product space. This means: (1) it is a vector space, (2) it has an inner product, and (3) it is complete.

## 2.2 Examples of Hilbert Spaces

- **Example 1:** $\mathbb{R}^n$ with standard inner product $\langle x, y\rangle = x^T y$ is a Hilbert space (it's complete because every Cauchy sequence in $\mathbb{R}^n$ converges).
- **Example 2:** $L^2[a,b] = \{f : [a,b] \to \mathbb{R} : \int_a^b f(x)^2 dx < \infty\}$ with $\langle f, g\rangle = \int_a^b f(x)g(x)dx$ is a Hilbert space.
  - Contains square-integrable functions (finite $L^2$ norm)
  - Even if a sequence $(f_n)$ of continuous functions converges to a discontinuous $f$, as long as $\int f^2 < \infty$, we have $f \in L^2$
  - $L^2$ fills in the gaps left by continuous functions
- $C[0,1]$ with $\|\cdot\|_2$ is *not* a Hilbert space (as we saw above).
- $\ell^2 = \{(x_1, x_2, \dots) : \sum_{i=1}^{\infty} x_i^2 < \infty\}$ (infinite sequences) with $\langle x, y\rangle = \sum_{i=1}^{\infty} x_i y_i$ is a Hilbert space. This is like "$\mathbb{R}^\infty$" with countably many coordinates.

### 2.3 Bases for Hilbert Spaces

- An **orthonormal basis** for a Hilbert space $\mathcal{H}$ is a sequence $(\phi_1, \phi_2, \dots)$ such that:
  - Orthonormality: $\langle \phi_i, \phi_j \rangle = \delta_{ij}$ (perpendicular unit vectors)
  - Completeness: Every $f \in \mathcal{H}$ can be written as an *infinite sum* $f = \sum_{i=1}^{\infty} \alpha_i \phi_i$ where $\alpha_i = \langle f, \phi_i \rangle$ and the sum converges in the Hilbert space norm.

  Recall that in $\mathbb{R}^n$, every vector is a *finite* sum of basis vectors. In infinite-dimensional Hilbert spaces, we need convergent *infinite series*.

**Example: The Fourier Basis** The Fourier basis is a fundamental example of an orthonormal basis for a Hilbert space.

- $L^2[-\pi, \pi]$ has the orthonormal basis:

$$\left\{ \frac{1}{\sqrt{2\pi}}, \frac{1}{\sqrt{\pi}} \cos(x), \frac{1}{\sqrt{\pi}} \sin(x), \frac{1}{\sqrt{\pi}} \cos(2x), \frac{1}{\sqrt{\pi}} \sin(2x), \dots \right\}$$

- It can be verified by direct integration that all pairs of distinct basis functions are orthogonal, and each has a norm of 1.
- Any $f \in L^2[-\pi, \pi]$ can be written as an infinite sum:

$$f = \sum_{i=1}^{\infty} \langle f, \phi_i \rangle \phi_i$$

  where $\phi_i$ are the basis functions above. The sum converges in $L^2$ norm. This is the **Fourier series** of $f$.

# 3 From Evaluation Functionals to Kernels

A key operation in machine learning is evaluating a function $f$ at a specific point $x$, for instance in a loss function $\ell(f(x_i), y_i)$. Can we treat this evaluation, $f \mapsto f(x)$, as a well-behaved operation in a general Hilbert space like $L^2$?

Let's define an **evaluation functional** $\delta_x : \mathcal{H} \to \mathbb{R}$ as $\delta_x(f) = f(x)$. This is a linear map from the vector space to its scalars. However, in many infinite-dimensional spaces, this functional is not **continuous**, which creates problems.

**Example: Evaluation in $L^2[0, 1]$ is not continuous.** To see why, consider the sequence of "spike" functions centered at $x = 0.5$:

$$f_n(x) = \begin{cases} \sqrt{n} & \text{if } x \in [0.5 - \frac{1}{2n}, 0.5 + \frac{1}{2n}] \\ 0 & \text{otherwise} \end{cases}$$

The $L^2$ norm is $\|f_n\|_2 = 1$ for all $n$, but the value at $x = 0.5$ is $f_n(0.5) = \sqrt{n}$, which goes to $\infty$. Since we have a sequence of functions with a constant norm but an unbounded value at a point, the evaluation functional $\delta_{0.5}$ is discontinuous.

**The Riesz Representation Theorem.** A fundamental result for Hilbert spaces states that every *continuous* linear functional $L$ can be represented as an inner product with a unique element in that space.

$$\text{For a continuous linear functional } L, \exists! u_L \in \mathcal{H} \text{ such that } L(f) = \langle f, u_L \rangle \text{ for all } f \in \mathcal{H}$$

If the evaluation functional $\delta_x$ were continuous, this theorem would guarantee the existence of a special function, let's call it $k_x \in \mathcal{H}$, that *represents* evaluation:

$$f(x) = \delta_x(f) = \langle f, k_x \rangle$$

This property—being able to represent evaluation as an inner product—is highly desirable, but as our example shows, it doesn't hold for all Hilbert spaces. This motivates us to seek out spaces where it does.

**Deriving Kernels.** Let's assume we are in such a "nice" Hilbert space where every evaluation functional has a representer $k_x$. What properties would a function $K(x, y) = \langle k_x, k_y \rangle$ have?

The properties of the inner product would be transferred to $K$:

1. **Symmetry:** $K(x, y) = \langle k_x, k_y \rangle = \langle k_y, k_x \rangle = K(y, x)$.
2. **Positive semi-definiteness:** For any points $x_1, \dots, x_N$ and coefficients $c_1, \dots, c_N$, the squared norm of $f = \sum_{i=1}^{N} c_i k_{x_i}$ must be non-negative:

$$0 \le \|f\|^2 = \left\langle \sum_i c_i k_{x_i}, \sum_j c_j k_{x_j} \right\rangle = \sum_i \sum_j c_i c_j \langle k_{x_i}, k_{x_j} \rangle = \sum_i \sum_j c_i c_j K(x_i, x_j)$$

This leads us to a crucial definition. The objects that can be used to build these well-behaved Hilbert spaces are functions that satisfy these two properties.

**Definition: Kernel** A **kernel** is a function $K : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ that is symmetric and positive semi-definite.

**Examples of Kernels**

- **Linear kernel:** $K(x, y) = x^T y$ (standard dot product in $\mathbb{R}^d$)
- **Polynomial kernel:** $K(x, y) = (x^T y + c)^p$
- **Gaussian (RBF) kernel:** $K(x, y) = \exp\left(-\frac{\|x-y\|^2}{2\sigma^2}\right)$. This corresponds to an infinite-dimensional feature space.
- **Kernel combinations:** Kernels are highly composable: sums $(K_1 + K_2)$, positive scalings ($cK_1$ for $c \ge 0$), and products $(K_1 K_2)$ of kernels are also kernels. This allows us to build complex similarity metrics from simple ones.

It turns out that symmetry and positive semi-definiteness are the only conditions needed. For any function $K$ satisfying them, we can construct a unique Hilbert space—a **Reproducing Kernel Hilbert Space (RKHS)**—where $K$ acts as the inner product between representers of evaluation. We will see this in the next lecture.