Lecture 10: Hilbert Space
10.1 Function Space

- A function space is a set of functions \( \mathcal{F} \) that has some structure.
- Often a nonparametric density estimation or a function approximation is chosen to lie in some function space, where the assumed structure is exploited by algorithms and theoretical analysis.

Let \( V \) be a vector space over \( \mathbb{R} \). A **norm** is a mapping \( \| \cdot \| : V \to [0, \infty) \) that satisfies

1. \( \| x + y \| \leq \| x \| + \| y \| \).
2. \( \| ax \| = |a| \| x \| \) for all \( a \in \mathbb{R} \).
3. \( \| x \| = 0 \) implies \( x = 0 \).

A vector space equipped with a norm is called a **normed vector space**.

**Example.** \( V = \mathbb{R}^k \) with \( \| x \| = \sqrt{\sum_i x_i^2} \).
10.1 Function Space

- A sequence \( x_1, x_2, \ldots \) is said to converge to \( x \) if \( \|x_n - x\| \to 0 \) as \( n \to \infty \).

- A sequence \( x_1, x_2, \ldots \) in a normed space is a **Cauchy sequence** if \( \|x_m - x_n\| \to 0 \) as \( m, n \to \infty \).

- The space is **complete** if every Cauchy sequence converges to a limit.

- A complete, normed vector space is called a **Banach space**.

**Example.** \( L^p([0, 1]) \) spaces, \( 1 \leq p \leq \infty \)

\[
\{ f(x) : \int |f^p(x)|^p \, dx < \infty \}. 
\]
10.2 Hilbert Space

► An inner product is a mapping $\langle \cdot, \cdot \rangle : V \times V \rightarrow \mathbb{R}$ that satisfies, for all $x, y, z \in V$ and $a \in \mathbb{R}$:

1. $\langle x, y \rangle \geq 0$ and $\langle x, x \rangle = 0$ if and only if $x = 0$
2. $\langle x, y + z \rangle = \langle x, y \rangle + \langle x, z \rangle$
3. $\langle x, ay \rangle = a \langle x, y \rangle$
4. $\langle x, y \rangle = \langle y, x \rangle$

Example. $V = \mathbb{R}^k$ with $\langle x, y \rangle = \sum_i x_i y_i$.

Example. $V = L^2([0, 1])$ with $\langle f, g \rangle = \int f(x)g(x)dx$.

► $x$ and $y$ are orthogonal if $\langle x, y \rangle = 0$

► Cauchy-Schwartz inequality

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

Example. $\int f(x)g(x)dx \leq (\int f^2 dx)^{1/2}(\int g^2 dx)^{1/2}$
10.2 Hilbert Space

- An inner product space is a normed space with the norm \( \|x\| = \langle x, x \rangle \).
- Parallelogram property

\[
\|x + y\|^2 + \|x - y\|^2 = 2(\|x\|^2 + \|y\|^2).
\]

- A **Hilbert space**, \( H \), is a complete, inner product vector space.
- Given a set \( S \subset H \) of a normed linear space and some point \( b \) outside of \( S \), the distance between \( b \) and \( S \) is defined as

\[
d(b, S) = \inf_{x \in S} \|x - b\|.
\]

- Note that in general there is no guarantee that there exists a point \( u \in S \) such that \( d(b, S) = \|u - b\| \) (this is why we have inf instead of min).
10.2 Hilbert Space

**Theorem.** A set $S \subset H$ is called **closed** if every convergent sequence $\{x_n\}$ in $S$ converges to an element of $S$. If $S$ is a closed linear space of a Hilbert space $H$ and $b$ is an element of $H$, then there exists $u \in S$ such that $\|u - b\| = d(b, S)$.

**Idea of Proof.**

- There exists a sequence $\{u_n\} \in S$ such that $\|u_n - b\| \to d(b, S)$ as $n \to \infty$.
- This does not mean that $\{u_n\}$ has a limit in $S$ in general.
- From
  \[
  \left\| \frac{1}{2}(b - u_m) \right\|^2 + \left\| \frac{1}{2}(b - u_n) \right\|^2 = \frac{1}{2} \left\| b - \frac{1}{2}(u_n + u_m) \right\|^2 + \frac{1}{8} \left\| u_n - u_m \right\|^2,
  \]

  $\|u_n - u_m\| \to 0$ and thus $\{u_n\}$ is a Cauchy sequence.
- From the definition of the Hilbert space, there exists $u \in S$ such that $d(b, S) = \|u - b\|$.  

10.2 Hilbert Space

For a closed subspace $S$ of $H$ and $x \in S$, $\hat{x}$ such that $d(x, S) = \|\hat{x} - x\|$ is called the closest point of $x$ in $S$ or the projection of $x$ onto $S$.

**Theorem.** Let $S$ be a closed linear subspace of $H$, let $x$ be any element of $S$, $b$ any element of $V$, and $\hat{b}$ the project of $b$ onto $S$. Then

$$\langle x - \hat{b}, b - \hat{b} \rangle = 0.$$  

**Proof.** If $x = \hat{b}$, we are done. Otherwise, set

$$\theta(x - \hat{b}) - (b - \hat{b}) = \theta x + (1 - \theta)\hat{b} - b = y - b$$

where $y = \theta x + (1 - \theta)\hat{b}$.

Since $y$ is in $S$ and $\|y - b\| \geq \|\hat{b} - b\|$, we have

$$\|\theta(x - \hat{b}) - (b - \hat{b})\|^2 = \theta^2\|x - \hat{b}\|^2 - 2\theta\langle x - \hat{b}, b - \hat{b} \rangle + \|b - \hat{b}\|^2 \geq \|b - \hat{b}\|^2.$$  

Therefore, $\theta^2\|x - \hat{b}\|^2 - 2\theta\langle x - \hat{b}, b - \hat{b} \rangle \geq 0$ for all $\theta$. 
10.2 Hilbert Space

For a closed subspace $S$ of $H$ and $x \in S$, $\hat{x}$ such that $d(x, S) = \|\hat{x} - x\|$ is called the closest point of $x$ in $S$ or the projection of $x$ onto $S$.

**Theorem.** Let $S$ be a closed linear subspace of $H$, let $x$ be any element of $S$, $b$ any element of $V$, and $\hat{b}$ the project of $b$ onto $S$. Then

$$\langle x - \hat{b}, b - \hat{b} \rangle = 0.$$

**Proof.**

Therefore, $\theta^2\|x - \hat{b}\|^2 - 2\theta\langle x - \hat{b}, b - \hat{b} \rangle \geq 0$ for all $\theta$. The left-hand side attains its minimum value when $\theta = \langle x - \hat{b}, b - \hat{b} \rangle/\|x - \hat{b}\|^2$, in which case

$$-\langle x - \hat{b}, b - \hat{b} \rangle^2/\|x - \hat{b}\|^2 \geq 0.$$

This implies

$$\langle x - \hat{b}, b - \hat{b} \rangle = 0.$$
10.2 Hilbert Space

For a closed subspace $S$ of $H$ and $x \in S$, $\hat{x}$ such that $d(x, S) = \|\hat{x} - x\|$ is called the closest point of $x$ in $S$ or the projection of $x$ onto $S$.

**Theorem.** Let $S$ be a closed linear subspace of $H$, let $x$ be any element of $S$, $b$ any element of $V$, and $\hat{b}$ the project of $b$ onto $S$. Then

$$\langle x - \hat{b}, b - \hat{b} \rangle = 0.$$ 

**Proof.**

**Corollary.** $b - \hat{b}$ is orthogonal to $S$.

**Corollary.** $\hat{b}$ is unique.
10.3 Parametric Regression

**Example.** For a given data set \( \{(x_i, y_i), i = 1, 2, ..., m\} \), let we want a linear function fit to the data

\[
y_i = ax_i + b.
\]

In a matrix form, we are looking for \( a \) and \( b \) such that

\[
\begin{pmatrix}
x_1 & 1 \\
\vdots & \vdots \\
x_m & 1 \\
\end{pmatrix}
\begin{pmatrix}
a \\
b \\
\end{pmatrix} =
\begin{pmatrix}
y_1 \\
\vdots \\
y_n \\
\end{pmatrix}
\]

In a more general form, we are looking for a solution \( u \) to the following linear system

\[
Au = v
\]

where \( A \) is an \( m \times n \) and \( m > n \).
10.3 Parametric Regression

Example. (cont’d) If we assume that the column vectors of $A$ are linearly independent, what can we say about the existence of the solution $u$?

▶ If $Au = \nu$ has a solution, then one can express $\nu$ as a linear combination of $A_1, A_2, \ldots, A_n$ (the column vectors of $A$). That is, if $\nu$ is not in the column space of $A$, there is no solution.

▶ The best we can represent about $\nu$ is the projection of $\nu$ into the column space of $A$, $\hat{\nu}$.

▶ Then, does $Au = \hat{\nu}$ has a solution? From the previous slides, we know

$$
\langle A_1, \hat{\nu} - \nu \rangle = 0, \langle A_2, \hat{\nu} - \nu \rangle = 0, \ldots, \langle A_n, \hat{\nu} - \nu \rangle = 0.
$$

That is, $A^T(Au - \nu) = A^T(\hat{\nu} - \nu) = 0$, i.e., $A^T Au = A^T \nu$.

As $A^T A$ is invertible, we have

$$
u = (A^T A)^{-1} A^T \nu,$$ the regression formula!
10.4 Orthonormal Basis

**Definition.** An orthonormal basis of a Hilbert Space $H$ is a family ${e_k \in H}_{k \in B}$ if it satisfies

1. $\langle e_k, e_j \rangle = 0$ for all $k \neq j, k, j \in B$.
2. $\|e_k\| = 1$ for all $k \in B$.
3. The linear span of $\{e_k\}$ is dense in $H$.

If the index set $B$ is countable, the Hilbert space is called **separable.** That is, for any $u \in H$, $u$ can be represented as

$$u = \sum_{i \in B} \beta_i e_i$$

for an orthonormal basis $\{e_i\}$.

**Note.** We will consider only separable Hilbert spaces.

- $\beta_i = \langle u, e_i \rangle$.
- $\|u\|^2 = \sum_{i \in B} \beta_i^2$. 

▶ $\beta_i = \langle u, e_i \rangle$.
▶ $\|u\|^2 = \sum_{i \in B} \beta_i^2$. 
10.4 Orthonormal Basis

Examples.

- \( \{ e^{2\pi inx} \}_{n=-\infty}^{\infty} \) is an orthogonal basis of \( L^2[0, 1] \).
- The Legendre polynomial is another orthogonal basis for \( L^2[0, 1] \).
- The Hermite polynomial is an orthogonal basis of \( L^2(\mathbb{R}) \).