Lecture 4. Hilbert Space

Functional Data Analysis

Jun Song

Department of Statistics
Korea University

Table of contents

Hilbert Space

Vector Space

Definition (Real vector space)

A real vector space V is a set of elements, called, vectors, with given operations of vector addition $+\colon V \times V \to V$ and scalar multiplication $\cdot\colon \mathbb{R} \times V \to V$ such that

1. *(Commutativity)* $v + w = w + v$ for $\forall v, w \in V$
2. *(Associativity)*$(v + u) + w = v + (u + w)$ for $\forall v, w, u \in V$
3. *(Identity element)*There is a zero vector 0 such that $v + 0 = v$ for $\forall v \in V$
4. *(Inverse element)*For each $v \in V$, $\exists - v$ such that $v + (-v) = 0$
5. For $\lambda \in \mathbb{R}$ and $v, w \in V$, $\lambda \cdot (v + w) = \lambda \cdot v + \lambda \cdot w$
6. For $\lambda, m \in \mathbb{R}$ and $v \in V$, $\lambda \cdot (m \cdot v) = (\lambda m) \cdot v$
7. For $\lambda, m \in \mathbb{R}$ and $v \in V$, $(\lambda + m) \cdot v = \lambda \cdot v + m \cdot v$
8. *(Identity element)*$1 \cdot v = v$ for $\forall v \in V$

Vector Space

Linear Combination: $a_1, \ldots, a_n \in \mathbb{R}$, $v_1, \ldots, v_n \in V$

$$a_1 v_1 + \cdots + a_n v_n \in V$$

$a_1, \ldots, a_n$ are often called the *coefficients* of the linear combination.

Definition (Linearly independent vectors)

$v_1, \ldots, v_n \in V$ are linearly independent if, for $a_1, \ldots, a_n \in \mathbb{R}$, $a_1 x_1 + \ldots + a_n x_n = 0$ has a unique solution $a_1 = \ldots = a_n = 0$. Therefore, if $v_i = \sum_{j \neq i} c_j v_j$ for some c_j, then $v_1, \ldots, v_n$ are not linearly independent, that is, linearly dependent.

Vector Space

Definition (Linearly independent vectors)

$v_1, \dots, v_n \in V$ are **linearly independent** if, for $a_1, \dots, a_n \in \mathbb{R}$, $a_1 v_1 + \dots + a_n v_n = 0$ has a unique solution $a_1 = \dots = a_n = 0$. Therefore, if $v_i = \sum_{j \neq i} c_j v_j$ for some c_j, then $v_1, \dots, v_n$ are not linearly independent, that is, linearly dependent.

Example

Let $e_i \in \mathbb{R}^n$ for $i = 1, \dots, n$ are the standard basis of $\mathbb{R}^n$. i.e., $e_1 = (1, 0, \dots, 0)^T$, ..., $e_n = (0, \dots, 0, 1)^T$. Then $e_1, \dots, e_n$ are linearly independent.

Vector Space

Definition (Linear Subspace)

S is a **subspace** of a vector space V if $S \subseteq V$ and S itself is a vector space. $S \leq V$

Definition (Linear Span)

Let V be a vector space and let $A = \{v_1, \ldots, v_p\} \subseteq V$. The **span** of A is the collection of all the possible linear combinations of elements in A.

$$\text{span}(A) = \left\{ \sum_{i=1}^{p} a_i v_i : a_1, \ldots, a_p \in \mathbb{R} \right\}$$

Remark: $\text{span}(A)$ is also a vector space. Thus, $\text{span}(A) \leq V$

Metric Space

Definition (Metric space(M,d))

A vector space M is a **metric space** if M is equipped with a distance function $d : M \times M \to \mathbb{R}$ satisfying followings:

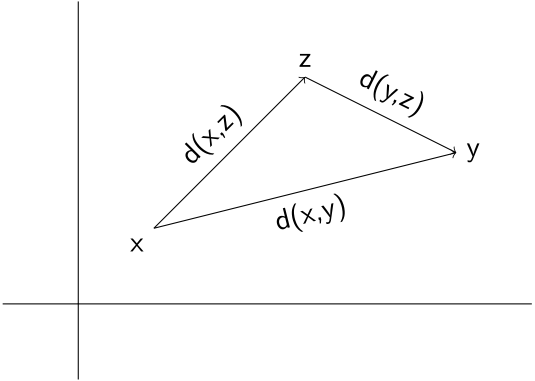
1. $d(x, y) \geq 0 \quad \forall x, y \in M$
2. $d(x, y) = 0 \iff x = y \quad \forall x, y \in M$
3. $d(x, y) = d(y, x) \quad \forall x, y \in M$
4. $d(x, y) \leq d(x, z) + d(z, y) \quad \forall x, y, z \in M$ (triangle inequality)

Metric Space

Example

$M = \mathbb{R}^2, x = (x_1, x_2)^T, y = (y_1, y_2)^T$

$$d(x, y) = \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2}$$



Normed Space

Definition (Normed space ($\mathcal{V}$,$||\cdot||$))

A vector space V is a **normed space**(or **normed vector space**) if V is equipped with a norm function $||\cdot|| : \mathcal{V} \to \mathbb{R}$ such that:

1. $||v|| \geq 0 \quad \forall v \in V$
2. $||v|| = 0 \iff v = 0$
3. $||\lambda \cdot v|| = |\lambda| \cdot ||v|| \quad \forall \lambda \in \mathbb{R},\ \forall v \in V$
4. $||v + w|| \leq ||v|| + ||w|| \quad \forall v, w \in V$

A normed space can be a metric space, by defining distance function $d(x, y) = ||x - y||$. However, the converse is not true in general.

Remark: A metric space does not have the condition for scalar multiplication. A *Banach space* is a *complete* normed vector space. We say that the normed vector space $(\|\cdot\|, \mathcal{V})$ is complete if every Cauchy sequence converges to a point in the space. Recall that a sequence $\{x_n\}$ is called Cauchy if for any $\epsilon > 0$ there exists N such that

$$\|x_n - x_m\| < \epsilon \quad \forall n, m > N.$$

Normed Space: Examples

1. Euclidean Norm (a standard norm): $x \in \mathbb{R}^n$,

$$||x||_2 = \sqrt{x_1^2 + \cdots + x_n^2}$$

2. ℓ_p-norm for $p \geq 1$

$$||x||_p = \left(\sum_{i=1}^{n} |x_i|^p \right)^{1/p}$$

   When $p = 1$, $||x||_1$ is frequently used in sparse modelings.

3. Maximum norm (infinite norm, supremum norm)

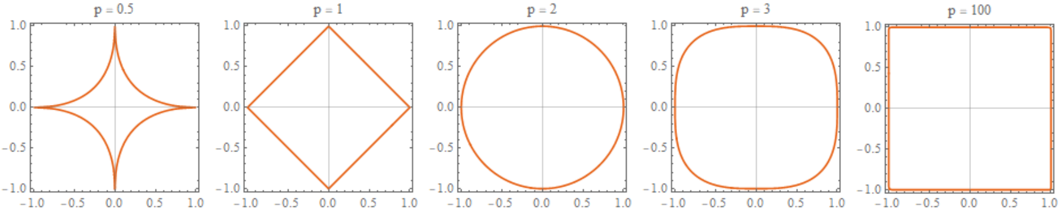
$$||x||_\infty = \max_i |x_i|$$

4. ℓ_0-norm (it's actually not a norm)

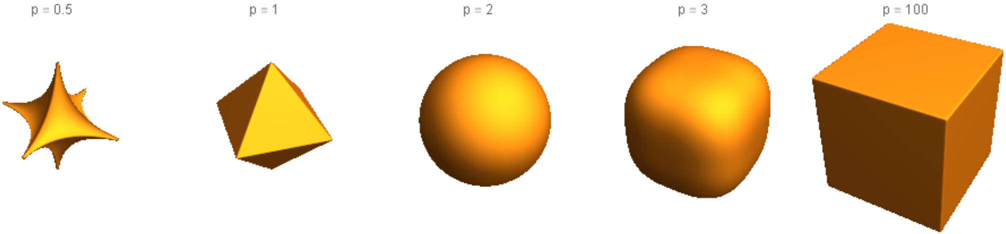
$$||x||_0 = \# \text{ of nonzero elements of x}$$

Normed Space: ℓ_p-norm

When $x \in \mathbb{R}^2$, $\{x \in \mathbb{R}^2 : \|x\|_p = 1\}$ is



When $x \in \mathbb{R}^3$, $\{x \in \mathbb{R}^2 : \|x\|_p = 1\}$ is



Inner Product Space

Definition (Inner product space)

A real vector space V is an **inner product space** if V is equipped with an inner product $\langle \cdot, \cdot \rangle : V \times V \to \mathbb{R}$ such that:

1. $\langle v, v \rangle \geq 0 \quad \forall v \in V$
2. $\langle v, v \rangle = 0 \iff v = 0$
3. $\langle \lambda v, w \rangle = \langle \lambda v, w \rangle \quad \forall \lambda \in \mathbb{R}, \; \forall v, w \in V$
4. $\langle v, w + u \rangle = \langle v, w \rangle + \langle v, u \rangle \quad \forall v, w, u \in V$
5. $\langle v . w \rangle = \langle w, v \rangle \quad \forall v, w \in V$

An inner product space V can be a normed space by defining $||v|| = \sqrt{\langle v, v \rangle}$. But the converse is not true generally.

Definition (Hilbert Space)

A **Hilbert space** is a **complete** inner product space.

Hilbert spaces might seem only slightly different than Banach spaces, but the mathematics becomes much simpler. Most properties you can think of from

Example - $\mathbb{R}^d$

Let $\mathbb{R}^d$ be the set of all d dimensional vectors with coordinates in $\mathbb{R}$ and $d < \infty$. The Euclidian norm is given by

$$|x|^2 = \sum_{i=1}^{d} x_i^2.$$

Under this norm, $\mathbb{R}^d$ is a Hilbert space. More generally, one, for $p \geq 1$, can define the ℓ_p norm as

$$|x|_p^p = \sum_{i=1}^{d} x_i^p \qquad |x|_\infty = \max_{1 \leq i \leq p} |x_i|.$$

Under any of these norms, $\mathbb{R}^d$ is only a Banach space unless $p = 2$.

Example - $\mathcal{C}[0,1]$

The space of continuous functions over $[0,1]$ is often denoted $\mathcal{C}[0,1]$. Equipped with the *sup-norm* this space is a Banach space:

$$\|x\| = \sup_{0 \le t \le 1} |x(t)|.$$

However the sup-norm is not an inner product norm (why?), and thus this is not a Hilbert space.

Example - $L^2[0,1]$

This space is the "bread and butter" of many FDA methods. The space $L^2[0,1]$ is the space of real-valued functions over $[0,1]$ which are square integrable:

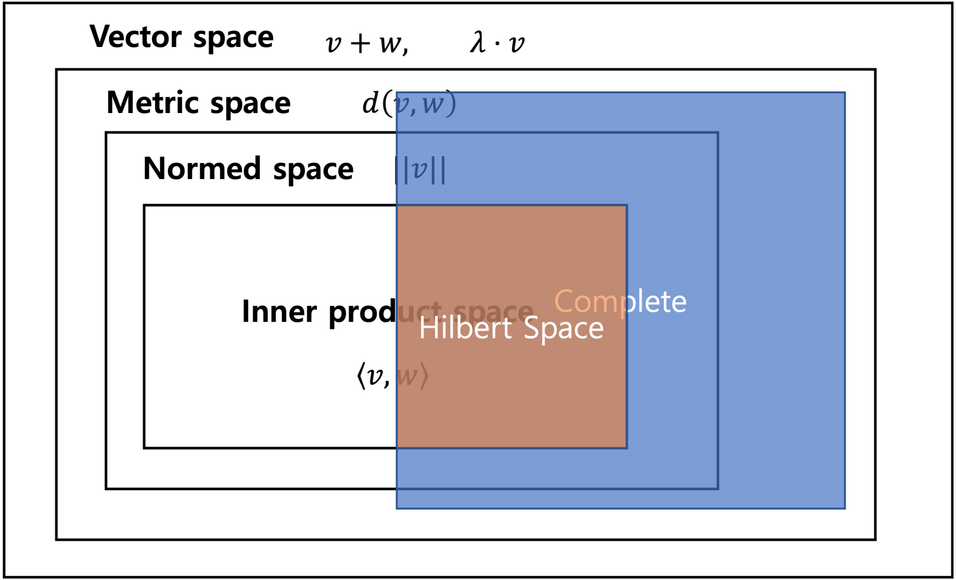
$$\|x\|^2 = \int_0^1 x(t)^2 \, dt.$$

Showing that it is an inner product space is straightforward.

Other examples

- Reproducing kernel Hilbert spaces,
- Sobolev spaces,
- Tensor product and Cartesian product spaces.

Recap & More



Vector space $v + w, \quad \lambda \cdot v$
Metric space $d(v, w)$
Normed space $||v||$
Inner product space
$\langle v, w \rangle$
Complete
Hilbert Space

Recap & More

- Inner product space $\subsetneq$ normed space $\subsetneq$ metric space $\subsetneq$ vector space
- geometrical information: Inner product $>$ norm $>$ metric
- A "complete" space is needed when we investigate asymptotic properties & approximation.

In Statistics

- When we have/collect data, we usually assume that the datasets are the **realizations of random elements** in a certain mathematical space.
- **Univariate data**
  A random variable $X : \Omega \to \mathbb{R}$. The observation(data) is $X(\omega) \in \mathbb{R}$
- **Multivariate data**
  A random vector $X : \Omega \to \mathbb{R}^p$
- **Functional data**
  A random function $X : \Omega \to L_2(\mathbb{R})$(or an arbitrary Hilbert space, $\mathcal{H}$)
- $\mathbb{R}$, $\mathbb{R}^p$, $L_2(\mathbb{R})$ are Hilbert spaces (complete inner product spaces).
- **Distributional data**; e.g., a target is a probability distribution $f(\cdot)$ - resides in a metric space.

Projections

Projections

- A major tool in FDA is functional principal component analysis, which involves projecting the data onto a lower-dimensional subspace.
- An advantage of using Hilbert spaces is that the inner product actually defines the geometry of the space.
- For example, we say that two elements are *perpendicular* if $\langle x, y \rangle = 0$.
- We will review some facts about projections.

Uniqueness of Projections

Theorem

Let $\mathcal{G}$ be a closed subspace of $\mathcal{H}$. For any $x \in \mathcal{H}$ define

$$\delta_x = \inf_{z \in \mathcal{G}} \|x - z\|.$$

Then there exists a unique point $y \in \mathcal{G}$ which achives $\|y - x\| = \delta_x$.

The point y is called the projection of x onto $\mathcal{G}$, and is denoted $P_{\mathcal{G}}(x)$ or just $P(x)$ for short.

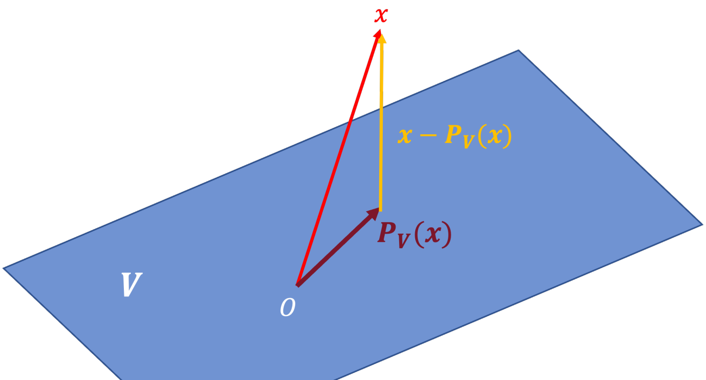
Orthogonal Projection onto a subspace

For each $x \in \mathcal{H}$, $P_V(x) \in V$ and

$$(x - P_V(x)) \perp v, \quad \text{for all } v \in V$$

Equivalently, $P_V(x)$ is the element in V that is closest to $x \in \mathcal{H}$.

$$P_V(x) = \text{argmin}_{v \in V} \|x - v\|$$



Properties

The orthogonal complement of $\mathcal{G}$, denoted $\mathcal{G}^{\perp}$ is defined as

$$\mathcal{G}^{\perp} = \{x \in \mathcal{H} : \langle x, y \rangle = 0 \text{ for some } y \in \mathcal{G}\}.$$

Note that $\mathcal{G}^{\perp}$ is also a closed linear subspace (why?). Denote $Q(\cdot)$ as the projection onto $\mathcal{G}^{\perp}$.

- Every $x \in \mathcal{H}$ can be uniquely decomposed as $x = P(x) + Q(x)$.
- $P(x)$ and $Q(x)$ are nearest points to x in $\mathcal{G}$ and $\mathcal{G}^{\perp}$.
- The mappings $P(\cdot)$ and $Q(\cdot)$ are linear, i.e.

$$P(ax + by) = aP(x) + bP(y).$$

- The norm of x satisfies $\|x\|^2 = \|P(x)\|^2 + \|Q(x)\|^2$.

Cauchy-Schwarz inequality

A very useful result for Hilbert spaces is the Cauchy-Schwarz inequality:

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

It is a fairly simple proof. Let $e = \|y\|^{-1} y$, then

$$\begin{aligned} \|x\|^2 &= \|x - \langle x, e\rangle e\|^2 + \|\langle x, e\rangle e\|^2 \\ &\geq \langle x, e\rangle^2 = \frac{\langle x, y\rangle^2}{\|y\|^2}. \end{aligned}$$

So, as desired

$$\|x\|^2 \|y\|^2 \geq \langle x, y\rangle^2.$$

(Note $P(x) = \langle e, x\rangle e$ and $Q(x) = x - P(x)$)

Riesz Representation Theorem

Theorem

If $L : \mathcal{H} \to \mathbb{R}$ is continuous and linear then there exists $y \in \mathcal{H}$ such that

$$L(x) = \langle x, y \rangle \qquad \forall x \in \mathcal{H}.$$

Clearly the function $\langle \cdot, y \rangle$ is continuous and linear, but the Riesz Representation Theorem states that every continuous linear functional is actually of this form. (Note: We will use the term "functional" to denote a mapping to the real line).

For any Hilbert (or Banach space), the set of continuous linear functionals is called the *Dual Space* and usually denoted as $\mathcal{H}^*$. The Riesz Representation Theorem states that, in some sense $\mathcal{H} = \mathcal{H}^*$.

Basis Expansions

Suppose that $e_1, \ldots, e_d$ is an orthonormal basis for a subspace $\mathcal{G} \subset \mathcal{H}$. This means that for any $x \in \mathcal{H}$ we have $P(x) \in \mathcal{G}$ and so

$$P(x) = a_1 e_1 + \cdots + a_d e_d \quad \text{and} \quad \langle e_i, e_j \rangle = 1_{i=j}.$$

Using our discussed properties we have

$$\begin{aligned}
\langle x, e_i \rangle &= \langle P(x) + Q(x), e_i \rangle \\
&= \langle P(x), e_i \rangle + \langle Q(x), e_i \rangle \\
&= \langle a_1 e_1 + \cdots + a_d e_d, e_i \rangle = a_i.
\end{aligned}$$

This means that $a_i = \langle x, e_i \rangle$ and so the projection, $P(\cdot)$, can be written as

$$P(x) = \sum_{i=1}^{d} \langle x, e_i \rangle e_i.$$

Separability

We say that the space $\mathcal{H}$ is separable if it contains a countable basis (there is a more general definition for other spaces). This means that any $x \in \mathcal{H}$ can be expressed as

$$x = \sum_{i=1}^{\infty} a_i e_i.$$

Theorem (Parceval's Theorem)

Let $\{e_i\}$ be an orthonormal basis of a real separable Hilbert space, $\mathcal{H}$. Then for any $x \in \mathcal{H}$ we have

$$x = \sum_{i=1}^{\infty} \langle x, e_i \rangle e_i \quad \text{and} \quad \|x\|^2 = \sum_{i=1}^{\infty} \langle x, e_i \rangle^2.$$

Nonseparable Banach spaces are uncommon, but not so difficult to construct. Nonseparable Hilbert spaces, however, are fairly exotic.

Linear Operators

Linear operators

Recall that an operator $L : \mathcal{H} \to \mathcal{H}$ is called linear if $L(ax + by) = aL(x) + bL(y)$. We say the linear operator is *bounded* if

$$\|L\|_{\mathcal{L}} = \sup_{\|x\| \leq 1} \|L(x)\| < \infty.$$

We denote the space of all bounded operators as $\mathcal{L}$. Under the norm $\| \cdot \|_{\mathcal{L}}$, this space is a Banach space.

Operator inequality

By definition of $\|L\|_{\mathcal{L}}$ one has the the fairly trivial *operator inequality*

$$\|L(x)\| \leq \|L\|_{\mathcal{L}}\|x\|.$$

Simple, but like Cauchy-Schwarz, it can be very useful.

Hilbert-Schmidt operators

A bounded linear operator, L, is called Hilbert-Schmidt if

$$\|L\|_{\mathcal{S}}^2 = \sum_{i=1}^{\infty} \|L(e_i)\|^2 < \infty.$$

where e_i is some orthonormal basis. The above quantity is invariant with respect to the choice of basis. We denote the space of Hilbert-Schmidt operators as $\mathcal{S}$, equipped with the norm $\|\cdot\|_{\mathcal{S}}$ it is a Hilbert space.

Example - Identity operator

The identity operator, L, is defined as $L(x) = x$. This operator is bounded since

$$\|L\|_{\mathcal{L}} = \sup_{\|x\| \leq 1} \|L(x)\| = \sup_{\|x\| \leq 1} \|x\| = 1.$$

However, it is not Hilbert-Schmidt

$$\|L\|_{\mathcal{S}}^2 = \sum_{i=1}^{\infty} \|L(e_i)\|^2 = \sum_{i=1}^{\infty} \|e_i\|^2 = \sum_{i=1}^{\infty} 1 = \infty.$$

Example - Integral operator

Suppose that $\mathcal{H} = L^2[0,1]$. Consider the bivariate function $\psi(t,s)$ which satisfies

$$\int\int \psi(t,s)^2 \; dtds < \infty.$$

This function can be used to define an *integral operator*

$$\Psi(x)(t) = \int \psi(t,s)x(s) \; ds.$$

This operator is bounded and Hilbert-Schmidt.

Example - Integral operator

Let $\|x\| \leq 1$ then by the Cauchy-Scwartz inequality

$$\begin{aligned}
\|\Psi(x)\|^2 &= \int \left(\int \psi(t,s)x(s)\ ds \right)^2 dt \\
&\leq \int \left(\int \psi(t,s)^2\ ds \right) \left(\int x(s)^2\ ds \right) dt \\
&\leq \int \int \psi(t,s)^2\ dsdt.
\end{aligned}$$

So $\Psi \in \mathcal{L}$

Example - Integral operator

Let $\psi_t(s) = \psi(t,s)$, then ψ_t is an element of $L^2[0,1]$. We have

$$\begin{aligned}
\|\Psi\|_{\mathcal{S}}^2 &= \sum \|\Psi(e_i)\|^2 \\
&= \sum \int \left(\int \psi(t,s) e_i(s) \; ds \right)^2 \; dt \\
&= \int \sum \langle \psi_t, e_i \rangle^2 \; dt \\
&= \int \|\psi_t\|^2 \; dt = \int \int \psi(t,s)^2 \; dt ds.
\end{aligned}$$

So $\Psi \in \mathcal{S}$.

Bounded vs Hilbert-Schmidt

It is important to remember that the Hilbert-Schmidt property is stronger than being bounded, i.e. $\mathcal{S} \subset \mathcal{L}$ and we have

$$\|\Psi\|_{\mathcal{L}} \leq \|\Psi\|_{\mathcal{S}} \qquad \text{for all } \Psi \in \mathcal{S}.$$

The proof involves a basis expansion:

$$\Psi(x) = \sum \langle x, e_i \rangle \Psi(e_i)$$

followed by the triangle inequality and Cauchy-Schwarz:

$$\begin{aligned} \|\Psi(x)\| &\leq \sum |\langle x, e_i \rangle| \|\Psi(e_i)\| \\ &\leq \left(\sum |\langle x, e_i \rangle|^2 \right)^{1/2} \left(\sum \|\Psi(e_i)\|^2 \right)^{1/2} = \|x\| \|\Psi\|_{\mathcal{S}}. \end{aligned}$$

Spectral Theory

Self adjoint operators

In function spaces, it is not clear what it means for an operator to be symmetric (unlike a symmetric matrix). However, for Hilbert spaces, the inner product gives a clear path for defining such operators. We say L^* is the *adjoint* of an operator L if

$$\langle L(x), y \rangle = \langle x, L^*(y) \rangle \qquad \text{for all } x, y \in \mathcal{H}.$$

We say L is *self-adjoint* (or symmetric for real Hilbert spaces) if $L = L^*$.

Spectral Theory - Matrices

Recall that for matrices, we have a number of decompositions. For symmetric matrices, the most famous is likely the eigendecomposition. If $\mathbf{A}$ is a symmetric matrix, then we can write

$$\mathbf{A} = \mathbf{Q}\mathbf{\Lambda}\mathbf{Q}^\top = \sum_{i=1}^{d} \lambda_i \mathbf{q}_i \mathbf{q}_i^\top,$$

where $\mathbf{q}_i$ are the eigenvectors, λ_i are the eigenvlaues, $\mathbf{Q} = [\mathbf{q}_1 \ldots \mathbf{q}_d]$, and $\mathbf{\Lambda}$ is a diagonal matrix of eigenvalues.

Spectral Theory - Linear operators

We have an analogous result for symmetric Hilbert-Schmidt operators. In particular, if $\Psi \in \mathcal{S}$ and is symmetric then for any $x \in \mathcal{H}$ we have

$$\Psi(x) = \sum_{i=1}^{\infty} \lambda_i \langle x, v_i \rangle v_i,$$

where λ_i and v_i are the eigenvalues and eigenfunctions of Ψ in the sense that

$$\Psi(v_i) = \lambda_i v_i.$$

The eigenfunctions v_i are orthonormal.

Spectral Theory - Some facts

Below are some basic facts for symmetric Hilbert-Schmidt operators.

- The v_i form an orthnormal basis for $\mathcal{H}$.
- We say Ψ is positive definite if $\langle \Psi x, x \rangle \geq 0$ for all $x \in \mathcal{H}$. In this case $\lambda_i \geq 0$.
- If Ψ is symmetric and positive definite then $\sup \langle \Psi x, x \rangle$, such that $\|x\| = 1$ and $\langle x, v_i \rangle = 0$ for $i \leq I$, occurs at v_{I+1} and the sup is λ_{I+1}.
- $\|\Psi\|_{\mathcal{L}} = \max\{|\lambda_i|\}$
- $\|\Psi\|_{\mathcal{S}}^2 = \sum \lambda_i^2$

Tensors

Tensors

Tensors are an incredibly useful concept for working in general Hilbert spaces. We have already encountered two examples of tensors (though they weren't called that).

- In the spectral theorem for matrices, we wrote $\mathbf{q}_i\mathbf{q}_i^\top$, this is a type of tensor product: $\mathbf{q}_i \otimes \mathbf{q}_i$.
- The other example was when we talked about bivariate functional objects (covariance functions). There these objects were reprsented as $\sum c_{ij}B_i(t)B_j(s)$, the terms $B_i(t)B_j(s)$ are tensor products: $B_i \otimes B_j$.

Tensors - Bilinear forms

There are a number of ways to view tensors (1) operators (2) bilinear functionals (3) tensor objects. The easiest way (at least as an intro) is to view them as bilinear functionals. In particular, if $x_1 \in \mathcal{H}_1$ and $x_2 \in \mathcal{H}_2$, then $x_1 \otimes x_2$ is a bilinear functional given by:

$$(x_1 \otimes x_2)(y_1, y_2) = \langle x_1, y_1 \rangle \langle x_2, y_2 \rangle.$$

So $x_1 \otimes x_2 : \mathcal{H}_1 \times \mathcal{H}_2 \to \mathbb{R}$.

Tensor space

Once one has a tensor product $\otimes$ we can actually define a vector space of tensors:

$$\mathcal{A} = \left\{ \sum_{j=1}^{J} x_{1j} \otimes x_{2j} : x_{1j} \in \mathcal{H}_1, x_{2j} \in \mathcal{H}_2 \right\}$$

where

$$\left(\sum_{j=1}^{J} x_{1j} \otimes x_{2j} \right) (y_1, y_2) := \sum_{j=1}^{J} (x_{1j} \otimes x_{2j})(y_1, y_2).$$

(Can you verify that this is actually a vector space?)

Tensor space

The vector space $\mathcal{A}$ can be equipped with the following inner product:

$$\left\langle \sum_{j=1}^{J} x_{1j} \otimes x_{2j}, \sum_{k=1}^{K} x_{3k} \otimes x_{4k} \right\rangle = \sum_{j=1}^{J} \sum_{k=1}^{K} \langle x_{1j}, x_{3k} \rangle_{\mathcal{H}_1} \langle x_{2j}, x_{4k} \rangle_{\mathcal{H}_2},$$

(Can you verify that this is actually an inner product?). Using this inner product, we can define a norm over $\mathcal{A}$.

Relaxing our view of tensors

While we defined the object $x_1 \otimes x_2$ as a bilinear form, it can also be viewed as an operator:

$$(x_1 \otimes x_2)(y) = \langle x_2, y \rangle x_1$$

which is a bit more common in practice. Often we will just use the notation $x_1 \otimes x_2$ which can then be interpreted as needed.

$$\|x_1 \otimes x_2\|_{\mathcal{H}_1 \otimes \mathcal{H}_2} = \|x_1 \otimes x_2\|_{\mathcal{S}}.$$

Example - $\mathbb{R}$

Suppose that $\mathcal{H}_1 = \mathbb{R}^n$ and $\mathcal{H}_2 = \mathbb{R}^m$, then we have

$$(x_1 \otimes x_2)(y_1, y_2) = \langle x_1, y_1 \rangle_{\mathbb{R}^n} \langle x_2, y_2 \rangle_{\mathbb{R}^m} = y_1^\top x_1 x_2^\top y_2.$$

So $x_1 \otimes x_2$ is equivalent to $x_1 x_2^\top$, which is an $m \times n$ matrix. Thus $\mathcal{H}_1 \otimes \mathcal{H}_2$ is equivalent to the space of $n \times m$ matrices, $\mathbb{R}^{n \times m}$.

Example - $L^2[0,1]$

Suppose that $\mathcal{H}_1 = \mathcal{H}_2 = L^2[0,1]$, then we have that

$$(x_1 \otimes x_2)(y_1, y_2) = \langle x_1, y_1 \rangle \langle x_2, y_2 \rangle = \int x_1(t) y_1(t) \; dt \int x_2(s) y_2(s) \; ds,$$

so $x_1 \otimes x_2$ is equivalent to the bivariate function $x_1(t)x_2(s)$ (each of which is square integrable). The tensor space $\mathcal{H}_1 \otimes \mathcal{H}_2$ is therefore equivalent to the space of square-integrable bivariate functions, i.e. $L^2([0,1]^2)$.

Example - Mixing

Suppose that $\mathcal{H}_1 = \mathbb{R}^d$ and $\mathcal{H}_2 = L^2[0,1]$, then we have that

$$(x_1 \otimes x_2)(y_1, y_2) = \langle x_1, y_1 \rangle \langle x_2, y_2 \rangle = x_1^\top y_1 \int x_2(t) y_2(t)\ dt,$$

so $x_1 \otimes x_2$ is equivalent to the vector of functions $x_2(t)x_1$. The tensor space $\mathcal{H}_1 \otimes \mathcal{H}_2$ is therefore equivalent to product space $(L^2[0,1])^d$, i.e. the space of d dimensional vectors of functions, with each coordinate being square integrable.

Tensor - Basis

Lastly, we mention that bases for tensor spaces can be constructed fairly easily. Namely, if $\{e_i\}$ is a basis for $\mathcal{H}_1$ and $\{f_j\}$ a basis for $\mathcal{H}_2$, then $\{e_i \otimes f_j\}$ is a basis for $\mathcal{H}_1 \otimes \mathcal{H}_2$. In addition, if the original bases were orthonormal then so is $\{e_i \otimes f_j\}$. We often call this a *tensor basis*.

It is a convenient construction as it works for any Hilbert space, however, it might not be optimal from a computational perspective as one often ends up with lots of basis functions in this way.