Hilbert Spaces
Applications to decomposition and sensitivity analysis
INP-ENSEEIHT
2018-2019

Serge Gratton

Chapter 1

Hilbert spaces

1.1 Introduction

Definition 1 (Inner Product)**.** *Let X be vector space with scalar field $\mathbb{K}$. $\langle \cdot, \cdot \rangle : X \times X \to \mathbb{K}$ is called an inner product if*

1. $\langle x + y, z \rangle = \langle x, z \rangle + \langle y, z \rangle$
2. $\langle \alpha x, y \rangle = \alpha \langle x, y \rangle$
3. $\langle x, y \rangle = \overline{\langle y, x \rangle}$
4. $\langle x, x \rangle \geq 0,\ \langle x, x \rangle = 0$ *iff* $x = 0$

Definition 2 (Hilbert Space)**.** *A linear space equipped with an inner product is an inner product space.*

Remark 1.

1. *If X is real then $\langle x, y \rangle = \langle y, x \rangle$.*
2. *From the definition of inner product, we obtain*

$$\begin{aligned} \langle \alpha x + \beta y, z \rangle &= \alpha \langle x, z \rangle + \beta \langle y, z \rangle \\ \langle x, \alpha y \rangle &= \overline{\alpha} \langle x, y \rangle \\ \langle x, \alpha y + \beta z \rangle &= \overline{\alpha} \langle x, y \rangle + \overline{\beta} \langle x, z \rangle \end{aligned}$$

So the inner product is linear in the first argument and semi-linear in the second argument.

Proposition 1 (Cauchy-Schwarz)**.** *In an inner product space*

$$\forall x,\ \forall y,\ |\langle x,y\rangle| \leq \sqrt{\langle x,x\rangle}\sqrt{\langle y,y\rangle}$$

Definition 3 (Norm)**.** *Let X be vector space with scalar field $\mathbb{K}$. $\|\cdot\| : X \to \mathbb{R}$ is called a norm if*

1. $\|x\| \geq 0$ *and* $[\|x\| = 0 \Longrightarrow x = 0]$
2. $\|\alpha x\| = \overline{\alpha}\|x\|$
3. $\|x+y\| \leq \|x\| + \|y\|$

Proposition 2 (Norm in an Hilbert space)**.** *In an inner product space, the function $x \to \sqrt{\langle x,x\rangle}$ is a norm.*

Proof. It follows from the definition of the inner product that $\|x\| \geq 0$ and $\|x\| = 0$ iff $x = 0$. Also, we have using the definiton of the inner product that

$$\|\alpha x\| = (\langle \alpha x, \alpha x\rangle)^{\frac{1}{2}} = (\alpha\bar{\alpha}\langle x,x\rangle)^{\frac{1}{2}} = (|\alpha|^2\langle x,x\rangle)^{\frac{1}{2}} = |\alpha|\|x\|.$$

Concerning the triangle inequality we get

$$\|x+y\|^2 = \langle x+y, x+y\rangle = \langle x,x\rangle + \langle x,y\rangle + \langle y,x\rangle + \langle y,y\rangle = \|x\|^2 + 2Re\langle x,y\rangle + \|y\|^2$$

$$\leq \|x\|^2 + 2\|x\|\|y\| + \|y\|^2 = (\|x\| + \|y\|)^2,$$

from which the result follows. □

Proposition 3 (Parallelogram equality)**.**

$$\|x+y\|^2 + \|x-y\|^2 = 2\left(\|x\|^2 + \|y\|^2\right) \tag{1.1}$$

Definition 4 (Cauchy sequence)**.** *In a normed vector space (x_n) is a Cauchy sequence iff*

$$\forall \epsilon > 0, \exists N(\epsilon), n \text{ and } p > N(\epsilon) \Longrightarrow \|x_n - x_p\| \leq \epsilon$$

Definition 5 (Hilbert Space)**.** *A complete normed space X is a space in which every Cauchy sequence of $X^{\mathbb{N}}$ is convergent. A complete subset A of a normed space X is a subset of X in which every Cauchy sequence of $A^{\mathbb{N}}$ has a limit* **in A**.
A Hilbert space is a complete *inner product space.*

Proposition 4 (Properties of Cauchy sequences)**.**

1. *A Cauchy sequence is bounded.*
2. *A convergent sequence is a Cauchy sequence. The converse is false in general.*

Example 1.

1. $\mathbb{R}^n$ *with inner product* $\langle x, y\rangle = x^T y$.
2. $\mathbb{C}^n$ *with inner product* $\langle x, y\rangle = x^T \overline{y}$.
3. *Real* $L^2[a, b]$ *with*
$$\langle x, y\rangle = \int_a^b x(t)y(t)\mu(dt).$$
4. ℓ^2 *with inner product*
$$\langle x, y\rangle = \sum_{i=1}^{\infty} x_i \overline{y_i}.$$

Problem 1.

1. *Show* ℓ^p *with* $p \neq 2$ *is not a Hilbert space.*
2. *Show* $C[a, b]$ *is not a Hilbert space.*

Lemma 1 (Continuity of Inner Product)**.** *If in an inner product space* $y_n \to y$ *and* $x_n \to x$ *then*
$$\langle x_n, y_n\rangle \to \langle x, y\rangle .$$

Problem 2. *Let* $(H_i, \langle\cdot,\cdot\rangle_i)_{i\in\mathbb{N}}$ *be a sequence of Hilbert spaces. Let* $H = \{x, x = (x_i)_{i\in\mathbb{N}}, x_i \in H_i\}$ *and* $\sum_{i=0}^{+\infty} \langle x_i, x_i\rangle_i < +\infty\}$*. Show that* H *is a Hilbert space for the inner-product* $\langle (x_i)_{i\in\mathbb{N}}, (y_i)_{i\in\mathbb{N}}\rangle = \sum_{i=0}^{+\infty} \langle x_i, y_i\rangle_i$.

1.2 Orthogonality in an Hilbert space

Definition 6. *Orthogonality.* $x, y \in X$ *are said to be* **orthogonal** *if* $\langle x, y\rangle = 0$ *and we write*
$$x \perp y.$$
For subsets $A, B \subset X$ *we say* $A \perp B$ *if*
$$a \perp b \ \forall a \in A, b \in B.$$

Theorem 1 (Topological results in normed spaces)**.** *Let $Y \subset H$ with H a Hilbert space. Then*

1. *Y is complete iff Y is closed in H.*
2. *If Y is finite dimensional then Y is complete. This is a consequence of the facts that $\mathbb{R}$ is complete, and that a finite product of complete subsets is complete.*

Definition 7 (Convex Subset)**.** *$M \subset X$ is said to be convex if*

$$\alpha x + (1-\alpha)y \in M, \ \forall \alpha \in [0,1], a, b \in M.$$

Theorem 2 (Projection)**.** *Let X be a Hilbert space and $M \neq \emptyset$ a convex subset which is closed. Then $\forall x \in X, \quad \exists! y \in M$ such that*

$$\delta = \inf_{\hat{y} \in M} \|x - \hat{y}\| = \|x - y\|.$$

The solution y of the infimum problem is called the orthogonal projection of x onto M and is denoted by $y = \Pi_M(x)$.

Proof.

1. Existence. By definition of inf,

$$\exists (y_n) \text{such that } \delta_n \to \delta, \text{where } \delta_n = \|x - y_n\|.$$

We will show (y_n) is Cauchy. Write $v_n = y_n - x$. Then $\|v_n\| = \delta_n$ and

$$\begin{aligned} \|v_n + v_m\| &= \|y_n + y_m - 2x\| \\ &= 2\|\frac{1}{2}(y_n + y_m) - x\| \geq 2\delta \end{aligned}$$

Furthermore, $y_n - y_m = v_n - v_m, \Longrightarrow$

$$\begin{aligned} \|y_n - y_m\|^2 &= \|v_n - v_m\|^2 \\ &= -\|v_n + v_m\|^2 + 2(\|v_n\|^2 + \|v_m\|) \\ &\leq -(2\delta)^2 + 2(\delta_n^2 + \delta_m^2). \end{aligned}$$

Since M is complete, $y_n \to y \in M$. Then $\|x - y\| \geq \delta$. Also, we have

$$\begin{aligned} \|x - y\| &\leq \|x - y_n\| + \|y_n - y\| \\ &= \delta_n + \|y_n - y\| \\ &\to \delta + 0 \end{aligned}$$

therefore $\|x - y\| = \delta$.

2. Uniqueness. Suppose that two such elements exist, $y_1, y_2 \in M$ and from above,

$$\|x - y_1\| = \delta = \|x - y_2\|.$$

By (1.1) we have

$$\begin{aligned} \|y_1 - y_2\|^2 &= \|(y_1 - x) + (x - y_2)\|^2 \\ &= 2\|y_1 - x\|^2 + 2\|y_2 - x\|^2 - \|(y_1 - x) + (y_2 - x)\|^2 \\ &= 2\delta^2 + 2\delta^2 - 2^2\|\frac{1}{2}(y_1 + y_2) - x\|^2 \\ &\leq 0 \end{aligned}$$

since

$$2^2\|\frac{1}{2}(y_1 + y_2) - x\|^2 \geq 4\delta^2.$$

Therefore $y_1 = y_2$.

□

Proposition 5 (characterization of the orthogonal projection on a closed nonempty convex set)**.** *Let X be a Hilbert space and $M \neq \emptyset$ a convex subset which is closed. The projection of x onto M is caracterized by*

$$\forall \hat{y} \in M, \Re \langle x - \Pi_M(x), \hat{y} - \Pi_M(x) \rangle \leq 0 \tag{1.2}$$

Proof. Let $\hat{y} \in M$ and $y_\theta = \Pi_M(x) + \theta(\hat{y} - \Pi_M(x))$ be a convex combination of $\Pi_M(x)$ and $\hat{y}$ obtain with $\theta \in [0, 1]$. We know that

$$\|x - y_\theta\|^2 = \|x - \Pi_M(x)\|^2 - 2\,\theta\,\Re \langle x - \Pi_M(x), \hat{y} - \Pi_M(x) \rangle + \theta^2\|\hat{y} - \Pi_M(x)\|^2. \tag{1.3}$$

If (1.2) holds we have, choosing $\theta = 1$, $\|x - \Pi_M(x)\| \leq \|x - \bar{y}\|$.
Vice versa, if $\Pi_M(x)$ is the solution of the projection theorem, we have that $\|x - \Pi_M(x)\|^2 \leq \|x - y_\theta\|^2$ for any $\theta \in [0, 1]$. Using (1.3) we get that $-2\,\theta\,\Re \langle x - \Pi_M(x), \hat{y} - \Pi_M(x) \rangle + \theta^2\|\hat{y} - \Pi_M(x)\|^2 \geq 0$. Dividing by $\theta > 0$ and letting $\theta \to 0^+$ yields (1.2). □

Proposition 6 (orthogonality)**.** *In the previous theorem, if M is a closed* linear subpsace *of the Hilbert space X, then $z = x - \Pi_M(x)$ is orthogonal to M, i.e.*

$$\forall \hat{y} \in M, \langle x - \Pi_M(x), \hat{y} \rangle = 0.$$

Proof. A possible proof can be obtained by using directly the characterization (1.2).
Another more elementary proof is as follows. Suppose $z \perp M$ were false, then $\exists w \in M$such that

$$\langle z, w \rangle = \beta \neq 0.$$

and so $w \neq 0$. For any α,

$$\begin{aligned} \|z - \alpha w\|^2 &= \langle z - \alpha w, z - \alpha w \rangle \\ &= \langle z, z \rangle - \overline{\alpha} \langle z, w \rangle - \alpha \left[\langle w, z \rangle - \overline{\alpha} \langle w, w \rangle \right] \\ &= \|z\|^2 - \overline{\alpha}\beta - \alpha \left[\beta - \overline{\alpha} \langle w, w \rangle \right]. \end{aligned}$$

For

$$\overline{\alpha} = \frac{\overline{\beta}}{\langle w, w \rangle},$$

we get,

$$\begin{aligned} \|z - \alpha w\|^2 &= \delta^2 - \frac{|\beta|^2}{\langle w, w \rangle} - \alpha [0] \\ &< \delta^2 \end{aligned}$$

contradicting the minimality of y, δ. Therefore we must have $z \perp M$. □

Definition 8 (Direct Sum)**.** *A vector space X is said to be* **direct sum** *of subspaces Y and Z, written*

$$X = Y \oplus Z,$$

if $\forall x \in \exists! y \in Y, \exists! z \in Z$*such that*

$$x = y + z.$$

Definition 9 (orthogonal complement)**.** *Let Y be a closed subspace of X. Then the orthogonal complement of Y in X is*

$$Y^{\perp} = \{ z \in X \text{such that } z \perp Y \}.$$

Theorem 3 (Direct Sum)**.** *Let Y be a closed subspace of H. Then*

$$H = Y \oplus Y^{\perp}.$$

Proof.

1. Existence. H is complete and Y is closed implies Y is complete. Further, we know Y is convex. Therefore $\forall x \in H, \quad \exists y \in Y$such that

$$x = y + z, \text{where } z \in Y^{\perp}.$$

2. Uniqueness. Assume $x = y + z = y_1 + z_1$. Then

$$y - y_1 \in Y$$

$$z - z_1 \in Y^{\perp}.$$

But $Y \cap Y^{\perp} = \{0\} \Longrightarrow y - y_1 = z - z_1 = 0.$

□

Definition 10 (orthogonal projection)**.** *Let Y be a closed subspace of H. From $H = Y \oplus Y^{\perp}$, we have $\forall x \in H$,*

$$x = y + z \text{ with } y \in Y \text{ and } z \in Y^{\perp}$$

The orthogonal projection onto Y is given by

$$P : H \to Y, x \mapsto Px = y.$$

Clearly P is continuous, and $P^2 = P$ (P is idempotent).

Definition 11 (orthogonal of a set)**.** *Let X be an inner product space and let M be a nonempty subset of X. Then the orthogonal of M is*

$$M^{\perp} = \{x \in X \text{such that } x \perp M\} = \{x \in X \text{such that } \langle x, v\rangle = 0 \ \forall v \in M\}.$$

Proposition 7. *If $A \subset B \subset X$, X Hulbert space, then $B^{\perp} \subset A^{\perp}$.*

Proof. Write $f_x : X \to \mathbb{C}, \ y \mapsto \langle y, x\rangle$ and use $A^{\perp} = \cap_{x \in A} f_x^{-1}(\{0\})$ □

‘

Proposition 8. *Let X be a Hilbert space. Let A be any subset of X. Let M, X be as the definition above and denote $(M^{\perp})^{\perp} = M^{\perp\perp}$. Then*

1. *$A^{\perp}$ is a closed vector space.*

2. $A^\perp = \overline{(vect(A))}^\perp$
3. $A \subset A^{\perp\perp}$.

Proof.

1. $A^\perp$ is a vector subspace of X. If $y_1, y_2 \in A^\perp$, $\alpha \in \mathbb{C}$, $\forall x \in A, \langle x, y_1 + \alpha\, y_2\rangle = \langle x, y_1\rangle + \bar{\alpha}\,\langle x, y_2\rangle = 0$.
   $A^\perp$ is closed. Let $f_x : X \to \mathbb{C}$, $y \mapsto \langle y, x\rangle$. The form f_x is linear and continous, therefore, $A^\perp = \cap_{x\in A} f_x^{-1}(\{0\})$ is a closed vector subspace of X.
2. From $A \subset \overline{(\text{vect}(A))}$ we get using Proposition 7 that $\overline{(\text{vect}(A))}^\perp \subset A^\perp$. Now let $x \in \overline{(\text{vect}(A))}$. There exists $(x_n) \in (\text{vect}(A))^{\mathbb{N}}$ such that $(x_n) \to x$. We therefore have $\langle y, x_n\rangle = 0$, and taking the limit yields $\langle y, x\rangle = 0$
3. Let $x \in A$. Then $\forall y \in A^\perp$, $\langle x, y\rangle = 0$, which shows that $x \in A^{\perp\perp}$.

□

Theorem 4. *If M is a closed subspace of a Hilbert space X then*

$$M^{\perp\perp} = M.$$

Proof. We know that $M \subset M^{\perp\perp}$ from above. Let $x \in M^{\perp\perp}$. We denot $\bar{x} = \Pi_M(x)$. We show that $x = \bar{x}$, which wields the result, using the rthogonal decompositon $X = M \oplus M^\perp$. We have $\|x - \bar{x}\|^2 = \langle x - \bar{x}, x - \bar{x}\rangle = \langle x, x - \bar{x}\rangle - \langle \bar{x}, x - \bar{x}\rangle$. Since $x \in M^{\perp\perp}$ and $x - \bar{x} \in M^\perp$, $\langle x, x - \bar{x}\rangle = 0$. Since $\bar{x} \in M$ and $x - \bar{x} \in M^\perp$, $\langle \bar{x}, x - \bar{x}\rangle = 0$. □

Theorem 5. *If A is any subset of a Hilbert space H then*

$$A^{\perp\perp} = \overline{(vect(A))}.$$

Proof. This is a consequence of $A^{\perp\perp} = \left(\overline{(\text{vect}(A))}^\perp\right)^\perp$ and $\overline{(\text{vect}(A))}$ is a closed subspace of X. □

Proposition 9 (orthogonal in the general case)**.** *Let $M \neq \emptyset$ be a subset of a Hilbert space H. Then*

$$\overline{vect(A)} = H \text{ iff } A^\perp = \{0\}.$$

Proof. Take the orthogonal in the preceeding theorem. □

Problem 3. *Let $(F_i)_{i\in I}$ by a family of linear subspaces of H. Show that*

- $(\cup_{i\in I} F_i)^\perp = \cap_{i\in I} F_i^\perp$
- *if all F_i's are closed,* $(\cap_{i\in I} F_i)^\perp = \overline{vec\ \left(\cup_{i\in I} F_i^\perp\right)}$

Definition 12 (Orthonormality)**.** *M is said to be an orthogonal set if $\forall x, y \in M$*

$$x \neq y \Longrightarrow \langle x, y\rangle = 0.$$

M is said to be orthonormal if $\forall x, y \in M$

$$\langle x, y\rangle = \delta_{xy} = \begin{cases} 0 & x \neq y \\ 1 & x = y \end{cases}.$$

If M is countable and orthogonal (resp. orthonormal) then we can write $M = (x_n)$ and we say (x_n) is an orthogonal (resp. orthonormal) sequence.

Remark 2 (Pythagorean theorem, linear independence)**.** *Let $M = \{x_1, \ldots, x_n\}$ be an orthogonal set. Then*

1. $$\|\sum_{i=1}^{n} x_i\|^2 = \sum_{i=1}^{n} \|x_i\|^2.$$
2. *M is linearly independent.*

Example 2 (Orthonormal sequences)**.**

1. $\{(1, 0, \ldots, 0), (0, 1, 0, \ldots, 0), \ldots, (0, \ldots, 0, 1, 0), (0, \ldots, 0, 1)\} \subset \mathbb{R}$.
2. $(e_n) \subset \ell^2$ *where* $e_n = \delta_n$.
3. *Let $X = C[0, 2\pi]$ with $\langle x, y\rangle = \int_0^{2\pi} x(t)y(t)dt$. Then the functions*

   $$(\sin nt), n = 1, 2, \ldots, (\cos nt), n = 0, 1, 2, \ldots$$

   form an orthogonal sequence.

Remark 3 (Unique representation with orthonormal sequences)**.** *Let X be an inner product space and let (e_k) be an orthonormal sequence in X, and suppose $x \in vect(\{e_1, \ldots, e_n\})$ where n is fixed. Then we can represent*

$$\begin{aligned} x = \sum_{k=1}^{n} \alpha_k e_k \quad &\implies \quad \langle x, e_i \rangle = \left\langle \left[\sum_{k=1}^{n} \alpha_k e_k \right], e_i \right\rangle = \alpha_i \\ &\implies \quad x = \sum_{k=1}^{n} \langle x, e_k \rangle e_k. \end{aligned}$$

Now let $x \in X$ be arbitrary and take $y \in vect(\{e_1, \ldots, e_n\})$, where

$$y = \sum_{k=1}^{n} \langle x, e_k \rangle e_k.$$

Define z by setting $x = y + z \implies z \perp y$ because

$$\begin{aligned} \langle z, y \rangle &= \langle x - y, y \rangle \\ &= \langle x, y \rangle - \langle y, y \rangle \\ &= \left\langle x, \left[\sum \langle x, e_k \rangle e_k \right] \right\rangle - \|y\|^2 \\ &= \sum \langle x, \langle x, e_k \rangle e_k \rangle - \|y\|^2 \\ &= \sum \overline{\langle x, e_k \rangle} \langle x, e_k \rangle - \sum |\langle x, e_k \rangle|^2 \\ &= 0. \end{aligned}$$

Then

$$\begin{aligned} \implies \quad \|x\|^2 \quad &= \|y\|^2 + \|z\|^2 \\ \implies \quad \|z\|^2 \quad &= \|x\|^2 - \sum |\langle x, e_k \rangle|^2 \geq 0 \\ \implies \quad \textstyle\sum_{k=1}^{n} |\langle x, e_k \rangle|^2 \quad &\leq \|x\|^2 \\ \implies \quad \textstyle\sum_{k=1}^{\infty} |\langle x, e_k \rangle|^2 \quad &\leq \|x\|^2 \end{aligned}$$

Proposition 10 (Bessel's inequality)**.** *Let X be an inner product space and let (e_k) be an orthonormal sequence in X. The* **Bessel's inequality** *holds:*

$$\sum_{k=1}^{\infty} |\langle x, e_k \rangle|^2 \leq \|x\|^2.$$

Definition 13 (Fourier Coefficients)**.** *The sequence* $(\langle x, e_k\rangle)$ *is called the Fourier coefficients of* x *w.r.t.* (e_k)*.*

Problem 4 (Fourier Coefficients are Minimizers)**.** *Let* $\{e_n, \dots, e_n\}$ *be an orthonormal set in an inner product space* X *(*n *is fixed). Let* $x \in X$ *be an arbitrary, fixed element and let* $y = \sum_{k=1}^{n} \beta_k e_k$*. Then* $\|x - y\|$ *depends on* $\beta_1, \dots, \beta_n$*. Show by direct calculation that* $\|x - y\|$ *is minimized iff* $\beta_i = \langle x, e_i\rangle\,,\ \forall i = 1, \dots, n$*.*

Solution. Let $\gamma_i = \langle x, e_i\rangle$, and $y = \sum \beta_i e_i$. Then

$$\|x-y\|^2 = \left\langle x - \sum \beta_i e_i, x - \sum \beta_i e_i \right\rangle = \|x\|^2 - \sum \bar{\beta}_i \gamma_i - \sum \beta_i \bar{\gamma}_i + \sum |\beta_i|^2$$

$$= \|x\|^2 - \sum |\gamma_i|^2 + \sum |\beta_i - \gamma_i|^2$$

and this is minumum for given x and e_is iff $\beta_i = \gamma_i$. □

Problem 5 (Gram-Schmidt)**.** *Orthonormalize the first three terms of the sequence* $(1, t, t^2, t^3, \dots)$ *on the interval* $[-1, 1]$ *where*

$$\langle x, y\rangle = \int_{-1}^{1} x(t)y(t)dt.$$

Solution. Let $f_1(t) = 1$, then $e_1(t) = \frac{f_1(t)}{\|f_1(t)\|} = \frac{1}{\sqrt{2}}$. Let $f_2(t) = t$. We have that $\langle f_2(\cdot), e_1(\cdot)\rangle = 0$, so we just need to normalize $f_2(t)$ to get $e_2(t) = \sqrt{\frac{3}{2}}t$. Let $f_3(t) = t^2$. Easy calculation shows that $\langle f_3(\cdot), e_1(\cdot)\rangle = \frac{\sqrt{2}}{3}$ and $\langle f_3(\cdot), e_2(\cdot)\rangle = 0$. Then $f_3(t) = \langle f_3(\cdot), e_2(\cdot)\rangle\, e_2(t) = t^2 - \frac{1}{3}$. Normalizing this quantity we get $e_3(t) = \sqrt{\frac{5}{8}}(3t^2 - 1)$. □

Theorem 6 (Convergence)**.** *Let* H *be a Hilbert space and let* $(e_n) \subset H$ *be an orthonormal sequence. Consider*

$$\sum_{k=1}^{\infty} \alpha_k e_k. \tag{1.4}$$

1. *The series* (1.4) *converges in the induced norm of* H *iff*

$$\sum_{k=1}^{\infty} |\alpha_k|^2 < \infty.$$

2. *If* (1.4) *converges to* x*, then* $\alpha_k = \langle x, e_k \rangle$ *and*

$$x = \sum_{k=1}^{\infty} \langle x, e_k \rangle e_k.$$

3. *For any* $x \in H$ *the series* (1.4) *with* $\alpha_k = \langle x, e_k \rangle$ *converges (in the norm of* H*).*

Proof. Let $s_n = \sum_{k=1}^{n} \alpha_k e_k$ and $\sigma_n = \sum_{k=1}^{n} |\alpha_k|^2$.

1. For $n > m$

$$\begin{aligned} \|s_n - s_m\|^2 &= \| \sum_{k=m+1}^{n} \alpha_k e_k \|^2 \\ &= \sum_{k=m+1}^{n} |\alpha_k|^2 \\ &= \sigma_n - \sigma_m. \end{aligned}$$

   Hence (s_n) is Cauchy in H iff (σ_n) is Cauchy in $\mathbb{R}$.

2. Note $\langle s_n, e_i \rangle = \alpha_i$, $\forall i = 1, \ldots, k \leq n$. By assumption $s_n \to x$, so $\alpha_i = \langle s_n, e_i \rangle \to \langle x, e_i \rangle$, for $i \leq k$ and $\alpha_i = \langle x, e_i \rangle$ $\forall i = 1, 2, \ldots$

3. This follows from Bessel's inequality and (1).

□

Definition 14 (Total Set)**.** *Let* X *be an inner product space and let* $M \subset X$*.*

1. *If* $\overline{vect(M)} = X$*,* M *is called a* **total set***.*

2. *If in addition to be a total set,* M *is an orthonormal set then* M *is a* **total orthonormal set***.*

Remark 4.

1. *A total orthonormal family in* X *is sometimes called an orthonormal basis for* X*. Note this is* **not equivalent to an algebraic basis** *unless* X *is finite dimensional.*

2. *In every nontrivial Hilbert space $H \neq \{0\}$ there is a total orthonormal set. The proof requires the axiom of choice or Zorns lemma.*

Definition 15 (Hilbert dimension)**.** *The Hilbert dimension of H is the cardinality of the smallest total orthonormal set, i.e., if $\Lambda = \left\{ M \text{such that } \overline{vect(M)} = H \right\}$ then the Hilbert dimension is*

$$\inf_{M \in \Lambda} |M| .$$

Proposition 11. *Let H be a Hilbert space and let $M \subset H$. Then*

1. *If M is total in H then*
$$M^{\perp} = \{0\} \tag{1.5}$$
2. *(1.5) $\Longrightarrow$ M is total in H.*

Proof. 1. We have $M^{\perp} = \overline{\text{vect}(M)}^{\perp}$ (see Proposition 8). Therefore from the fact that M is total, $\overline{\text{vect}(M)}^{\perp} = H^{\perp} = \{0\}$.

2. If x is a Hilbert space and $M^{\perp} = \{0\}$, then $M^{\perp\perp} = \overline{\text{vect}(M)} = \{0\}^{\perp} = H$ implies that M is total in H.

□

Proposition 12 (Parseval's equality)**.** *An orthonormal set M in a Hilbert space H is total iff*

$$\sum_k |\langle x, e_k \rangle|^2 = \|x\|^2, \ \forall x \in H.$$

Proof. If M is not total, by Problem 21 there is a nonzero $x \perp M$ in H. Since $x \perp M$ we have 0 on the left-hand side of Parseval's Equality which is not equal to $\|x\|$. Hence if Parceval's Equality holds for all $x \in H$, then M must be total in H.

Conversely, assume M to be total in H. Consider any $x \in H$ and its nonzero Fourier coefficients arranged in a sequence, i.e., $\langle x, e_1 \rangle, \langle x, e_2 \rangle, ...,$. Define $y = \sum \langle x, e_k \rangle e_k$. It follows that $x - y \in M^{\perp}$. Since M is total, then $x = y$. □

Proposition 13 (Fourier series)**.** *An orthonormal set M in a Hilbert space H is total iff*

$$\sum_k \langle x, e_k \rangle e_k = x, \ \forall x \in H.$$

Proof. Let $x \in H$. From Parseval we know that $\sum_k |\langle x, e_k\rangle|^2$ converges, which show from Theorem 6 that $\sum_k \langle x, e_k\rangle e_k$ converges to some $y \in H$. From $\forall k \ \langle x - y, e_k\rangle = 0$, we deduce that $x - y \in M^\perp$ and the conclusion follows from (1.5).

Conversely, if the Fourier series converges for all x. Let $x \in M^\perp$. Then $\langle x, \sum_{k=1}^n \langle x, e_k\rangle e_k\rangle = 0$. Taking the limit $n \to +\infty$ shows that $x = 0$. The conlusion follows from Part 2. of Proposition 11. □

1.3 Representation of functionals in Hilbert spaces

Theorem 7 (Riesz)**.** *Let H be a Hilbert space. Every bounded, linear functional on H can be represented in terms of the inner product on H, i.e.,*

$$f(x) = \langle x, z\rangle$$

where z is uniquely determined by f and

$$\|z\| = \|f\|.$$

Proof.

1. Existence of z. If $f = 0$ take $z = 0$. Otherwise assume $f \neq 0$. Then $(\mathrm{Ker} f) \neq H$ and $(\mathrm{Ker} f) \neq H \implies (\mathrm{Ker} f)^\perp \neq \{0\}$. Let $w \in (\mathrm{Ker} f)^\perp$such that $w \neq 0$ and set

$$v = f(x)w - f(w)x, \ x \in H.$$

Then

$$\begin{aligned} \implies & \ f(v) = f(x)f(w) - f(w)f(x) = 0 \\ \implies & \ v \in (\mathrm{Ker} f). \end{aligned}$$

Since $w \perp (\mathrm{Ker} f)$ we have

$$\begin{aligned} 0 &= \langle v, w\rangle \\ &= \langle f(x)w - f(w)x, w\rangle \\ &= f(x)\langle w, w\rangle - f(w)\langle x, w\rangle \\ &= f(x)\|w\|^2 - f(w)\langle x, w\rangle . \end{aligned}$$

Then

$$\begin{aligned} \implies f(x) &= \frac{f(w)}{\|w\|^2}\langle x, w\rangle \\ \implies f(x) &= \left\langle x, \overline{\left(\frac{f(w)}{\|w\|^2}\right)} w \right\rangle. \end{aligned}$$

Then

$$z = \overline{\left(\frac{f(w)}{\|w\|^2}\right)} w.$$

2. Uniqueness of z. Suppose $f(x) = \langle x, z_1\rangle = \langle x, z_2\rangle$. Then

$$\implies \langle x, z_1 - z_2\rangle = 0, \ \forall x \in H.$$

Choose $x = z_1 - z_2$. Then

$$\begin{aligned} &\implies \langle z_1 - z_2, z_1 - z_2\rangle = 0 \\ &\implies z_1 = z_2. \end{aligned}$$

3. $\|f\| = \|z\|$. If $f = 0$ then $z = 0$ and $\|f\| = \|z\| = 0$. For $f \neq 0$, $z \neq 0$. Note

$$\begin{aligned} f(z) &= \langle z, z\rangle \\ &= \|z\|^2, \text{and} \\ \|z\|^2 &\leq \|f\|\|z\| \\ \implies \|z\| &\leq \|f\|. \end{aligned}$$

Also

$$\begin{aligned} |f(x)| &= |\langle x, z\rangle| \\ &\leq \|x\|\|z\| \\ \implies \|f\| &\leq \|z\|. \end{aligned}$$

This yields

$$\|f\| = \|z\|.$$

□

Lemma 2 (Equality)**.** *Let X be an inner product space. Then*

$$\langle x, w\rangle = \langle y, w\rangle \ \forall w \in X \Longrightarrow x = y.$$

In particular,

$$\langle x, w\rangle = 0 \ \forall w \in X \Longrightarrow x = 0.$$

Definition 16 (Sesquilinear Form)**.** *Let X, Y be vector spaces over the same scalar field $\mathbb{K}$ ($\mathbb{R}$ or $\mathbb{C}$). A* **sesquilinear form** *(or* **sesquilinear functional***) h on $X \times Y$ is a mapping*

$$h : X \times Y \to \mathbb{K}$$

such that

$$\forall x, x_1, x_2 \in X \, and \, y, y_1, y_2 \in Y \, and \, \alpha, \beta \in \mathbb{K}$$

1. $h(x_1 + x_2, y) = h(x_1, y) + h(x_2, y)$
2. $h(x, y_1 + y_2) = h(x, y_1) + h(x, y_2)$
3. $h(\alpha x, y) = \alpha h(x, y)$
4. $h(x, \beta y) = \overline{\beta} h(x, y)$

Remark 5.

1. *If X, Y are real ($\mathbb{K} = \mathbb{R}$), then*

   $$h(x, \beta y) = \beta h(x, y)$$

   and h is said to be **bilinear**.

2. *If X, Y are vector spaces and $\exists c \in \mathbb{R}$ such that*

   $$|h(x, y)| \leq c\|x\|\|y\|, \ \forall x, y$$

   then h is bounded and

$$\begin{aligned} \|h\| &= \sup_{\substack{x \neq 0 \\ y \neq 0}} \frac{|h(x, y)|}{\|x\|\|y\|} \\ &= \sup_{\substack{\|x\| = 1 \\ \|y\| = 1}} |h(x, y)| \\ \Longrightarrow |h(x, y)| &\leq \|h\|\|x\|\|y\|. \end{aligned}$$

Theorem 8 (Riesz Representation)**.** *Let H_1, H_2 be Hilbert spaces and*

$$h : H_1 \times H_2 \to \mathbb{K}$$

a bounded sesquilinear form. Then h has a representation

$$h(x, y) = \langle Sx, y \rangle$$

where $S : H_1 \to H_2$ is a bounded, linear operator. S is uniquely determined by h and

$$\|S\| = \|h\|.$$

Proof.

1. Existence of S. Consider $\overline{h(x, y)}$ which is linear in y. If we fix an x, then $\overline{h(x, y)}$ is a bounded, linear functional and we can use the Riesz theorem to find z and represent

   $$\overline{h(x, y)} = \langle y, z \rangle ,$$

   hence

   $$h(x, y) = \langle z, y \rangle . \tag{1.6}$$

   Note that z is unique, but it depends on $x \in H_1$. This means that (1.6) with variable x defines an operator $S : H_1 \to H_2$ given by

   $$z = Sx,$$

   and we write

   $$h(x, y) = \langle z, y \rangle = \langle Sx, y \rangle .$$

   We must show S is linear. Observe

   $$\begin{aligned} \langle S(\alpha x_1 + \beta x_2), y \rangle &= h(\alpha x_1 + \beta x_2, y) \\ &= \alpha h(x_1, y) + \beta h(x_2, y) \\ &= \alpha \langle Sx_1, y \rangle + \beta \langle Sx_2, y \rangle , \ \forall y \in H_2 \\ \implies S(\alpha x_1 + \beta x_2) &= \alpha Sx_1 + \beta Sx_2. \end{aligned}$$

2. Uniqueness of S. If $h(x, y) = \langle Sx, y \rangle = \langle Tx, y \rangle$, then $S = T$ by the equality lemma.

3. Boundedness of S. Note first

$$\begin{aligned} \|h\| &= \sup_{\substack{x\neq 0\\ y\neq 0}} \frac{|\langle Sx, y\rangle|}{\|x\|\|y\|} \\ &\geq \sup_{\substack{x\neq 0\\ Sx\neq 0}} \frac{|\langle Sx, Sx\rangle|}{\|x\|\|Sx\|} \\ &= \sup_{x\neq 0} \frac{\|Sx\|}{\|x\|} \\ &= \|S\| \\ \Longrightarrow \|h\| &\geq \|S\|. \end{aligned}$$

So S is bounded. Also

$$\begin{aligned} \|h\| &= \sup_{\substack{x\neq 0\\ y\neq 0}} \frac{|\langle Sx, y\rangle|}{\|x\|\|y\|} \\ &\leq \sup_{\substack{x\neq 0\\ y\neq 0}} \frac{\|Sx\|\|y\|}{\|x\|\|y\|} \\ &\leq \sup_{x\neq 0} \frac{\|Sx\|}{\|x\|} \\ &= \|S\| \\ \Longrightarrow \|h\| &\leq \|S\|. \end{aligned}$$

This yields

$$\|S\| = \|h\|.$$

□

Problem 6. *Let H be a Hilbert space. Show that H' (the dual of H) is a Hilbert space with inner product $\langle \cdot, \cdot \rangle_1$ defined by*

$$\langle f_z, f_v \rangle_1 = \overline{\langle z, v\rangle} = \langle v, z\rangle ,$$

where

$$\begin{aligned} f_z(x) &= \langle x, z\rangle \\ f_v(x) &= \langle x, v\rangle \end{aligned}$$

Solution. It is easy to verify that $\langle \cdot, \cdot \rangle_1$ is an inner product on $H^{'}$. Since $\|f_z\| = \|z\| = (\langle z, z \rangle)^{\frac{1}{2}} = (\langle f_z, f_z \rangle_1)^{\frac{1}{2}}$ the norm on $H^{'}$ is induced by the inner product $\langle \cdot, \cdot \rangle_1$. We know that the normed dual is always a Banach space. It follows that $H^{'}$ is complete, and therefore a Hilbert space. □

1.4 Hilbert adjoint and pseudo-inverse

1.4.1 Hilbert adjoint

Definition 17 (Hilbert adjoint T^*)**.** *Let H_1, H_2 be Hilbert spaces and let $T : H_1 \to H_2$ be a bounded, linear operator. Then the adjoint is $T^* : H_2 \to H_1$such that*

$$\langle Tx, y \rangle = \langle x, T^* y \rangle \,, \ \forall x \in H_1 \text{and } x \in H_2.$$

Theorem 9 (Existence of the Hilbert Adjoint)**.**

1. *T^* exists.*
2. *T^* is unique.*
3. *$\|T^*\| = \|T\|$.*

Proof. Observe that $h(y, x) = \langle y, Tx \rangle$ is sesquilinear form on $H_2 \times H_1$.

$$\begin{aligned} \implies |h(x,y)| &\leq \|y\|\|Tx\| \\ &\leq \|T\|\|x\|\|y\| \\ \implies \|h\| &\leq \|T\|. \end{aligned}$$

Also

$$\begin{aligned} \|h\| &= \sup_{\substack{x \neq 0 \\ y \neq 0}} \frac{|\langle y, Tx \rangle|}{\|y\|\|x\|} \\ &\geq \sup_{\substack{x \neq 0 \\ Tx \neq 0}} \frac{|\langle Tx, Tx \rangle|}{\|Tx\|\|x\|} \\ &= \|T\| \\ \implies \|h\| &= \|T\|. \end{aligned}$$

Therefore h is a bounded sesquilinear form. Using a Riesz representation, $h(x,y) = \langle T^*y, x\rangle$ where T^* exists, is unique with norm

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

Also

$$\begin{aligned} \implies \langle y, Tx\rangle &= \langle T^*y, x\rangle \\ \implies \langle Tx, y\rangle &= \langle x, T^*y\rangle . \end{aligned}$$

□

Lemma 3 (Zero operator)**.** *Let X, Y be inner product spaces and $Q : X \to Y$ be bounded, linear operators. Then*

1. *$Q = 0$ iff $\langle Qx, y\rangle = 0\ \forall x \in X$ and $y \in Y$*
2. *Let X be* **complex** *and $Q : X \to X$. Then*

$$\langle Qx, x\rangle = 0\ \forall x \in X \implies Q = 0.$$

Proof.

1. $\implies$.

$$Q = 0 \implies Qx = 0\ \forall x \in X \implies \langle Qx, x\rangle = 0\ \forall x.$$

$\impliedby$.

$$\begin{aligned} \langle Qx, y\rangle = 0\ \forall x, y &\implies Qx = 0\ \forall x, y \\ &\implies Q = 0. \end{aligned}$$

2. Let $x, y \in X$, then $v = \alpha x + y \in X$. Then

$$\begin{aligned} \langle Qv, v\rangle &= \langle Qx, x\rangle = \langle Qy, y\rangle = 0 \\ &\implies \\ 0 &= \langle Q(\alpha x + y), \alpha x + y\rangle \\ &= |\alpha|^2 \langle Qx, x\rangle + \langle Qy, y\rangle + \alpha \langle Qx, y\rangle + \overline{\alpha} \langle Qy, x\rangle \\ &= \alpha \langle Qx, y\rangle + \overline{\alpha} \langle Qy, x\rangle . \end{aligned}$$

Now, take $\alpha = 1$ then $\alpha = i$ to obtain the two relations

$$\begin{aligned} \langle Qx, y\rangle + \langle Qy, x\rangle &= 0 \\ \langle Qx, y\rangle - \langle Qy, x\rangle &= 0. \end{aligned}$$

Then $\langle Qx, y\rangle = 0 \implies Q = 0$ by 1.

□

Remark 6. *In the previous lemma, 2 is not necessarily true if X is real.*

Theorem 10 (Properties of the Hilbert adjoint)**.** *Let H_1, H_2 be Hilbert spaces. Let $S, T : H_1 \to H_2$ be bounded, linear operators and let α be a scalar. Then*

1. $\langle T^*y, x\rangle = \langle y, Tx\rangle$
2. $(S+T)^* = S^* + T^*$
3. $(\alpha T)^* = \overline{\alpha} T^*$
4. $(T^*)^* = T$
5. $\|T^*T\| = \|TT^*\| = \|T\|^2$
6. $T^*T = 0$ *iff* $T = 0$
7. $(ST)^* = T^*S^*$, *assuming* $H_1 = H_2$.

Definition 18 (Self-Adjoint, Unitary, Normal)**.** *Let H be a Hilbert space and $T : H \to H$.*

1. *T is self-adjoint (or Hermitian) if $T^* = T$*
2. *T is unitary if T is bijection and $T^* = T^{-1}$*
3. *T is normal if $TT^* = T^*T$*

Remark 7.

1. *If T is self-adjoint then $\langle Tx, y\rangle = \langle x, Ty\rangle$.*
2. *If T is self-adjoint or normal then T is normal.*

Example 3. *Consider $\mathbb{C}^n$ with $\langle x, y\rangle = x^T\overline{y}$. Let $T : \mathbb{C}^n \to \mathbb{C}^n$. If we specify a basis for $\mathbb{C}^n$, we can represent T, T^* by matrices A, B. Then*

$$\begin{aligned} \langle Tx, y\rangle &= (Ax)^T\overline{y} \\ &= x^T A^T \overline{y} \\ \langle x, T^*y\rangle &= x^T\overline{B}\overline{y}. \end{aligned}$$

Therefore

$$\begin{aligned} \implies A^T &= \overline{B} \\ \implies B &= \overline{A}^T. \end{aligned}$$

If T is self-adjoint then $A = \overline{A}^T$.

Theorem 11 (Self-Adjointness)**.** *Let H be a Hilbert space and let $T : H \to H$ be a bounded, linear operator. Then*

1. *If T is self-adjoint then $\langle Tx, x\rangle$ is real for all $x \in H$.*

2. *If H is complex and $\langle Tx, X\rangle$ is real for all $x \in H$ then T is self-adjoint.*

Proof.

1. If T is self-adjoint, then for all $x \in X$ we have

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

2. If $\langle Tx, x\rangle$ is real for all $x \in X$, then

$$\langle Tx, x\rangle = \overline{\langle Tx, x\rangle} = \overline{\langle x, T^*x\rangle} = \langle T^*x, x\rangle .$$

   Hence

$$\begin{aligned} 0 &= \langle Tx, x\rangle - \langle T^*x, x\rangle \\ &= \langle (T - T^*)x, x\rangle \\ \implies T &= T^* \end{aligned}$$

   by the zero operator lemma, since H is complex.

□

Theorem 12. *Let H be a Hilbert space and let (T_n) be a sequence of bounded, linear, self-adjoint operators $T_n : H \to H$. Suppose that $T_n \to T$ in norm, that is $\|T_n - T\| \to 0$, where $\|\cdot\|$ is the norm on the space $B(H, H)$. Then T is a bounded, linear, self-adjoint operator on H.*

Proof. We show $T^* = T$.

$$\begin{aligned} \|T - T^*\| &\leq \|T - T_n\| + \|T_n - T_n^*\| + \|T_n^* - T^*\| \\ &= \|T - T_n\| + 0 + \|T_n^* - T^*\| \\ &= 2\|T - T_n\| \\ &\to 0. \end{aligned}$$

□

Theorem 13 (Unitary Operators)**.** *Let H be a Hilbert space and let $U, V : H \to H$ be unitary. Then*

1. *U is isometric, i.e., $\|Ux\| = \|x\|$ $\forall x \in H$*
2. *$\|U\| = 1$ provided $H \neq \{0\}$*
3. *$U^{-1} = U^*$ is unitary*
4. *UV is unitary*
5. *U is normal.*
6. *If T is a bounded, linear operator on H and H is complex, then T is unitary iff T is isometric and surjective.*

Proof.

1. $$\begin{aligned} \|Ux\|^2 &= \langle Ux, Ux\rangle \\ &= \langle x, U^*Ux\rangle \\ &= \langle x, Ix\rangle \\ &= \|x\|^2 \end{aligned}$$

2. Follows from 1.

3. Since U is bijective, so is U^{-1} and

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

4. UV is bijective and
$$(UV)^* = V^*U^* = V^{-1}U^{-1} = (UV)^{-1}.$$

5. $U^{-1} = U^*$ and $UU^{-1} = U^{-1}U = I$.

6. $\Longrightarrow$. Suppose that T is isometric and surjective. Isometry implies injectivity, so T is bijective. Need to show $T^* = T^{-1}$. By isometry,
$$\begin{aligned} \langle T^*Tx, x\rangle &= \langle Tx, Tx\rangle \\ &= \langle x, x\rangle \\ &= \langle Ix, x\rangle \\ \Longrightarrow \langle (T^*T - I)x, x\rangle &= 0 \\ \Longrightarrow T^*T &= I. \end{aligned}$$
Also,
$$\begin{aligned} TT^* &= TT^*(TT^{-1}) \\ &= T(T^*T)T^{-1} \\ &= I. \end{aligned}$$
Therefore $T^* = T^{-1}$.

$\Longleftarrow$. Conversely, T is isometric by 1 and surjective by definition.

□

1.4.2 Pseudo-inverse

We now show how to define the pseudo-inverse in a Hilbert space H.

In this section let A be an operator that is

1. linear,
2. continuous,
3. for which $\mathrm{Im}(A)$ is bounded in H.

We associate to A the operator A_0 defined by
$$\begin{aligned} A_0 : \ker(A)^\perp &\to \mathrm{Im}(A) \\ x &\mapsto A(x) \end{aligned}$$

1.4.3 Problems

Problem 7. *Let E be a normed linear space over $\mathbb{C}$. Show that its norm $\| \bullet \|$ is induced by a scalar product iff it enjoys the parallelogram identity. [Hint: express the scalar product as using $\|x \pm i\, y\|^2$. Prove that $\langle x+y, z\rangle = 2\langle x, z/2\rangle + 2\langle y, z/2\rangle$. Then show that $\langle x, z/2\rangle = \frac{1}{2}\langle x, z\rangle$]*

Problem 8. *Let E be a Hilbert space.*

1. *For $a \in E$, show that $d(x, \{a\}^\perp) = \frac{|\langle x,a\rangle|}{\|a\|}$*
2. *Let F be the vector subspace of $E = L^2(0,1)$ defined by*

$$F = \{f \in E \text{ such that } \int_0^1 f(x)\dot{x} = 0\}.$$

Determine $F^\perp$. Compute the distance to F of the function $f \in E$ defined by $f(x) = e^x$.

Problem 9 (Lax-Milgram theorem)**.** *Let E ube a real Hilbert space. Let consider a bilinear form a sur E, supposed to be continuous : $\exists C > 0$, $\forall x, y \in E$, $|a(x,y)| \leq C\|x\|\|y\|$; and coercive : $\exists \alpha > 0$, $\forall x \in E$, $a(x,x) \geq \alpha\|x\|^2$.*

1. *(a) Prove that there exists a conitnuous linear operator T defined on E such that $\forall x, y \in E$, $a(x,y) = \langle Tx, y\rangle$.*
   (b) Show that $T(E)$ is dense in E.
   (c) Prove that $\forall x \in E$, $\|Tx\| \geq \alpha\|x\|$. Conclude that T is injective and that $T(E)$ is closed.
   (d) Deduce that T is an isomorphism from E to itself.
2. *Let L be a continuous linear application defined from E to E.*
   (a) Use the previous questions to show that $\exists! u \in E$, $\forall y \in E$, $a(u,y) = L(y)$.
   (b) Suppose now that in addtion a is symmetric. We define for $x \in E$,

$$\Phi(x) = \frac{1}{2}a(x,x) - L(x).$$

Show that u is characterized by:

$$\Phi(u) = \min_{x \in E} \Phi(x).$$

Problem 10 (Lions-Stampacchia theorem)**.** *Let C be a closed non empty convex set of a real Hulbert space E, a be a coercive, continuous symmetric bilinear form coercive sur E and L be a continuous linear form on E. Let J be defined on E by*

$$J(u) = \frac{1}{2}a(u,u) - L(u).$$

Show that $\exists! c \in C,\ \forall v \in C,\ J(c) \leq J(v)$ and that c is characterized by the following condition: $\forall v \in C,$

$$a(c, v-c) \geq L(v-c).$$

Hint: From Lax-Milgram, $\exists! u \in E$ such that $\forall v \in E,\ a(u,v) = L(v)$. Check that $J(v) = a(v-u, v-u) - a(u,u)$ and then consider the Hilbert space (E,a).

Problem 11. *Let $(\lambda)_n$ be a sequence of positive members such that $\lim_{n\to+\infty} \lambda_n = +\infty$. Let $V = \left\{(u_n)_{n\in\mathbb{N}} \in \mathbb{R}^{\mathbb{N}}\right\}$*

Problem 12 (Weak convergence)**.** *Let H be a Hilbert space, $x_n, x \in H$. We say that $x_n \to x$ weakly $\left[x = w\text{-}\lim_{n\to\infty} x_n\right]$ if for any $h \in H$ we have $(x_n, h) \to (x, h)$.*

1. *Show that If $\|x_n - x\| \to 0$ then $x = w\text{-}\lim_{n\to\infty} x_n$.*
2. *Let $\{u_n\}$ be an orthonormal system. Show that $w - \lim_{n\to\infty} u_n = 0$.*
3. *If $y_n \to y$ and $w\text{-}\lim_{n\to\infty} x_n = x$, show that $(x_n, y_n) \to (x, y)$.*

Problem 13. *Let E be the Hilbert space $E \overset{n}{=} (\ell^2_{\mathbb{R}}, \langle\cdot,\cdot\rangle)$ where*

$$\ell^2_{\mathbb{R}} \overset{def}{=} \left\{ X = (x_n)_{n\geq 0} \in \mathbb{R}^{\mathbb{N}} / \sum_{n=0}^{\infty} x_n^2 < \infty \right\}$$

et $\langle X, Y\rangle \overset{def}{=} \sum_{n=0}^{\infty} x_n\, y_n$.

We consider

$$C \overset{def}{=} \left\{ X = (x_n)_{n\geq 0} \in \ell^2_{\mathbb{R}} \,/\, \forall n \in \mathbb{N} : x_{2n+1} = \frac{1}{(2n+1)^2} \right\}.$$

1. *Show C is a closed nonempty convex set of E.*

2. *Let $A \overset{def}{=} (a_n)_{n\geq 0} \in E$ with $\forall n \in \mathbb{N} : a_n \overset{def}{=} \dfrac{1}{n+1}$. Prove that A has a unique projection $\bar{A}$ on C and determine it.*

Problem 14. *Let $E \overset{n}{=} (\ell^2_{\mathbb{R}}, \|\cdot\|_2)$ where $\ell^2_{\mathbb{R}} \overset{d}{=} \{X \overset{n}{=} (x_n)_n \in \mathbb{R}^{\mathbb{N}} / \quad \sum_{n=0}^{\infty} x_n^2 < \infty\}$ and $(\forall X \in \ell^2_{\mathbb{R}}) : \|X\|_2 \overset{d}{=} (\sum_{n=0}^{\infty} x_n^2)^{1/2}$.*

We consider the right shift, operator defined by

$$\begin{array}{rccc} A: & \mathbb{R}^{\mathbb{N}} & \to & \mathbb{R}^{\mathbb{N}} \\ & X & \mapsto & AX \overset{n}{=} Y \end{array}$$

where $Y = (y_n)_{n\geq 0}$ with

$$\left\{ \begin{array}{rcll} y_0 & = & 0, & \\ y_n & = & x_{n-1} & \forall n \geq 1. \end{array} \right.$$

1. *Show that $A \in \mathcal{L}(E)$. Evaluate $\sup_{x\neq 0} \frac{(AX)}{(X)}$; is A an isometry ?*

2. *Determine the adjoint A^T of A.*

3.

   (a) *Show that the image of A is closed.*

   (b) *Compute the pseudo-inverse operator A^+ of A.*

Chapter 2

Reproducible Kernel Hilbert Spaces

2.1 The kernel trick

The ridge regression can solution can be written

$$XX^T(XX^T + \lambda I)^{-1}Y = X\hat{\beta}_\lambda^R \text{ or } X(X^TX + \lambda I)^{-1}X^TY = X\hat{\beta}_\lambda^R.$$

Since X^TX is a $p \times p$ matrix, computing it costs $O(np^2)$ operations. Since, XX^T is an $n \times n$ matrix, it $O(n^2p)$ time to compute. If $p \gg n$, then this way of computing the fitted values would be much quicker. Importantly, the result of Ridge regression only depends on $K = XX^T$ (and Y).

This can be used as follows. Suppose we believe we have a quadratic signal

$$Y_i = x_i^T\beta + \sum_{k,\ell} x_{ik}x_{i\ell}\theta_{k\ell} + \varepsilon_i.$$

We can use the products $x_{ik}x_{i\ell}$ leading to $O(p^2)$ many predictors. Even if we use the XX^T way, this has a cost of $O(n^2p^2)$, and this naive method would require $O(np^4)$ operations.

We can do better than this. The idea is that we might be able to compute K directly. Consider

$$(1 + x_i^Tx_j)^2 = 1 + 2x_i^Tx_j + \sum_{k,\ell} x_{ik}x_{i\ell}x_{jk}x_{j\ell}.$$

This is equal to the inner product between vectors of the form

$$(1, \sqrt{2}x_{i1}, \ldots, \sqrt{2}x_{ip}, x_{i1}x_{i1}, \ldots, x_{i1}x_{ip}, x_{i2}x_{i1}, \ldots, x_{ip}x_{ip}) \qquad (*)$$

If we set $K_{ij} = (1 + x_i^T x_j)^2$ and form $K(K + \lambda I)^{-1}Y$, we get the ridge regression solution with $(*)$ as our predictors. This is interesting, because computing this is only $O(n^2 p)$, and we got rid of a factor p in this computation. The key idea is that the solution depends on K, and not on the values of x_{ij} themselves.

Consider the general scenario where we try to predict the value of Y given a predictor $x \in \mathcal{X}$. In general, we don't even assume $\mathcal{X}$ has some nice linear structure where we can do linear regression.

If we want to do Ridge regression, then one thing we can do is that we can try to construct some map $\phi : \mathcal{X} \to \mathbb{R}^D$ for some D, and then run Ridge regression using $\{\phi(x_i)\}$ as our predictors. Following the idea above, we compute

$$K_{ij} = k(x_i, x_j) = \langle \phi(x_i), \phi(x_j) \rangle .$$

If we can do so, then we can simply use the above formula to obtain the fitted values.

Since it is only the function k that matters we can provide a suitable function $k : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$. If we are given such a function k, when would it correspond to a map $\phi : \mathcal{X} \to \mathbb{R}^D$?

If we want k to come from a feature map ϕ, then an immediate necessary condition is that k has to be symmetric. We will also need a condition that corresponds to the positive-definiteness of the inner product.

In this chapter, $\mathcal{H}$ is a Hilbert space and $\mathcal{X}$ is a subset of $\mathcal{H}$.

Proposition 14. *Given $\phi : \mathcal{X} \times \mathcal{X} \to \mathcal{H}$, define $k : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ by*

$$k(x, x') = \langle \phi(x), \phi(x') \rangle .$$

Then for any $x_1, \ldots, x_n \in \mathcal{X}$, the matrix $K \in \mathbb{R}^n \times \mathbb{R}^n$ with entries

$$K_{ij} = k(x_i, x_j)$$

is positive semi-definite.

Proof. Let $x_1, \ldots, x_n \in \mathcal{X}$, and $\alpha \in \mathbb{R}^n$. Then

$$\begin{aligned} \sum_{i,j} \alpha_i k(x_i, x_j) \alpha_j &= \sum_{i,j} \alpha_i \langle \phi(x_i), \phi(x_j) \rangle && (2.1) \\ &= \left\langle \sum_i \alpha_i \phi(x_i), \sum_j \alpha_j \phi(x_j) \right\rangle && (2.2) \\ &\geq 0 && (2.3) \end{aligned}$$

since the inner product is positive definite. □

Definition 19 (Positive-definite kernel)**.** *A* positive-definite kernel *(or simply* kernel*) is a symmetric map $k : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ such that for all $n \in \mathbb{N}$ and $x_1, \ldots, x_n \in \mathcal{X}$, the matrix $K \in \mathbb{R}^n \times \mathbb{R}^n$ with entries*

$$K_{ij} = k(x_i, x_j)$$

is positive semi-definite.

Example 4. *Suppose $(k_i)_{i \in \mathbb{N}}$ are kernels. Then*

- *If $\alpha_1, \alpha_2 \geq 0$, then $\alpha_1 k_1 + \alpha_2 k_2$ is a kernel. Moreover, if*

  $$k(x, x') = \lim_{m \to \infty} k_m(x, x')$$

  exists, then k is a kernel.

- *The pointwise product $k_1 k_2$ is a kernel, where*

  $$(k_1 k_2)(x, x') = k_1(x, x') k_2(x, x').$$

Example 5. *The* linear kernel is $k(x, x') = x^T x'$. It corresponds to $\phi = I$ and taking the standard inner product on $\mathbb{R}^p$.

Example 6. *The* polynomial kernel is $k(x, x') = (1 + x^T x')^d$ for all $d \in \mathbb{N}$. It is a kernel since 1 and $x^T x'$ are kernels, and sums and products preserve kernels.

Example 7. *The* Gaussian kernel is

$$k(x, x') = \exp\left(-\frac{\|x - x'\|_2^2}{2\sigma^2} \right).$$

The quantity σ is known as the *bandwidth.*

To show that it is a kernel, we decompose

$$\|x - x'\|_2^2 = \|x\|_2^2 + \|x'\|_2^2 - 2x^T x'.$$

We define

$$k_1(x, x') = \exp\left(-\frac{\|x\|_2^2}{2\sigma^2}\right) \exp\left(-\frac{\|x'\|_2^2}{2\sigma^2}\right).$$

This is a kernel by taking $\phi(\cdot) = \exp\left(-\frac{\|\cdot\|_2^2}{2\sigma^2}\right)$.

Next, we can define

$$k_2(x, x') = \exp\left(\frac{x^T x'}{\sigma^2}\right) = \sum_{r=0}^{\infty} \frac{1}{r!} \left(\frac{x^T x'}{\sigma^2}\right)^r.$$

This is the infinite linear combination of powers of kernels, hence it is a kernel. Therefore $k = k_1 k_2$ *is a kernel.*

Theorem 14 (Moore–Aronszajn theorem)**.** *For every kernel* $k : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$*, there exists an inner product space* $\mathcal{H}$ *and a feature map* $\phi : \mathcal{X} \to \mathcal{H}$ *such that*

$$k(x, x') = \langle \phi(x), \phi(x') \rangle.$$

Proof. Let $\mathcal{H}$ be generated by the family of vectors $k(\cdot, x)_{x \in \mathcal{X}}$. In other words, $\mathcal{H}$ is the vector space of functions $f : \mathcal{X} \to \mathbb{R}$ of the form

$$f(\cdot) = \sum_{i=1}^{n} \alpha_i k(\cdot, x_i) \tag{2.4}$$

for some $n \in \mathbb{N}$, $\alpha_1, \ldots, \alpha_n \in \mathbb{R}$ and $x_1, \ldots, x_n \in \mathcal{X}$. If

$$g(\cdot) = \sum_{i=1}^{m} \beta_j k(\cdot, x'_j) \in \mathcal{H},$$

then we define the inner product of f and g by

$$\langle f, g \rangle = \sum_{i=1}^{n} \sum_{j=1}^{m} \alpha_i \beta_j k(x_i, x'_j).$$

We first show that it does not depend on the particular representation chosen for f and g. To do so, we observe that

$$\sum_{i=1}^{n}\sum_{j=1}^{m}\alpha_i\beta_j k(x_i,x'_j)=\sum_{i=1}^{n}\alpha_i g(x_i)=\sum_{j=1}^{m}\beta_j f(x'_j). \tag{2.5}$$

The first equality shows that the definition of the inner product does not depend on the representation of g, while the second equality shows that it doesn't depend on the representation of f.

Note the *reproducing property. We have (take* $m=1$ *and* $x'_j=x$ *in* (2.5)*),*

$$\langle k(\cdot,x),f\rangle=f(x).$$

Our form $\langle\cdot,\cdot\rangle$*is clearly symmetric and bilinear. We show it is positive definite. We have that*

$$\langle f,f\rangle=\sum_{i=1}^{n}\sum_{j=1}^{m}\alpha_i k(x_i,x_j)\alpha_j\geq 0$$

since the kernel is positive semi-definite. Suppose now that $\langle f,f\rangle=0$.

Using the Cauchy–Schwarz inequality (which does not require positive definiteness, just positive semi-definiteness), we get

$$f(x)^2=\langle k(\cdot,x),f\rangle^2\leq\langle k(\cdot,x),k(\cdot,x)\rangle\langle f,f\rangle=0.$$

It follows that $f\equiv 0$*. Thus, we know that* $\mathcal{H}$ *is an inner product space.*

For the feature map, we define $\phi:\mathcal{X}\to\mathcal{H}$ *by*

$$\phi(x)=k(\cdot,x).$$

Then we have $\langle\phi(x),\phi(x')\rangle=\langle k(\cdot,x),k(\cdot,x')\rangle=k(x,x')$*, as desired.*

Suppose that (f_m) *is Cauchy in the inner product space* $\mathcal{H}$*, then*

$$|f_m(x)-f_n(x)|\leq k^{1/2}(x,x)\|f_m-f_n\|_{\mathcal{H}}\to 0$$

as $m,n\to\infty$*. Since every Cauchy sequence in* $\mathbb{R}$ *converges (i.e.* $\mathbb{R}$ *is complete), we set*

$$f(x)=\lim_{n\to\infty}f_n(x).$$

If we complete $\mathcal{H}$ with the limit of all Cauchy sequences, and it can be shown that the resulting space is a Hilbert space. The inner product is

$$\langle f, g \rangle = lim_{n \to +\infty} \langle f_n, g_n \rangle ,$$

where (f_n) ang (g_n) are Cauchy sequences converging respectively to f and g. □

Definition 20 (Reproducing kernel Hilbert space (RKHS))**.** *A Hilbert space $\mathcal{B}$ of functions $f : \mathcal{X} \to \mathbb{R}$ is a reproducing kernel Hilbert space if for each $x \in \mathcal{X}$, there exists a $k_x \in \mathcal{B}$ such that*

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

for all $x \in \mathcal{B}$.

The function $k : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ given by

$$k(x, x') = \langle k_x, k'_x \rangle = k_x(x') = k_{x'}(x)$$

is called the reproducing kernel associated with $\mathcal{B}$.

By the Riesz representation theorem, this condition is equivalent to saying that pointwise evaluation is continuous.

Proposition 15. *Let $\mathcal{B}$ by a RKHS. Its associated reproducing kernel is a positive definite kernel.*

Proof. The symmetry is obtained from the symmetry of the inner product in $\mathcal{B}$: $k(x, x') = \langle k_x, k'_x \rangle = \langle k'_x, k_x \rangle = k(x', x)$. For the positive definitness, let $x_1, \ldots, x_n \in \mathcal{X}$, and $\alpha \in \mathbb{R}^n$ and $f = \sum_{i=1}^n \alpha_i k(\cdot, x_i)$. Then

$$\begin{aligned} \langle f, f \rangle &= \left\langle \sum_i \alpha_i k(\cdot, x_i), \sum_j \alpha_j k(\cdot, x_j) \right\rangle \\ &= \sum_{i,j} \alpha_i \alpha_j k(x_i, x_j) \geq 0 \end{aligned}$$

and the positive definitness of $[k(x_i, x_j)]$ is a consequence of the the positive definitness of the inner product in $\mathcal{B}$. □

Example 8. *Take the linear kernel*

$$k(x, x') = x^T x'.$$

By definition, we have

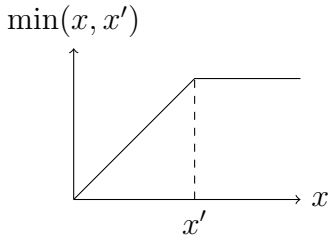
$$\mathcal{H} = \left\{ f : \mathbb{R}^p \to \mathbb{R} \mid f(x) = \sum_{i=1}^{n} \alpha_i x_i^T x \right\}$$

for some $n \in \mathbb{N}$, $x_1, \ldots, x^n \in \mathbb{R}^p$. *We then see that this is equal to*

$$\mathcal{H} = \{ f : \mathbb{R}^p \to \mathbb{R} \mid f(x) = \beta^T x \text{ for some } \beta \in \mathbb{R}^p \},$$

and if $f(x) = \beta^T x$, *then* $\|f\|_{\mathcal{H}}^2 = k(\beta, \beta) = \|\beta\|_2^2$.

Example 9. *Take the Sobolev kernel, where* $\mathcal{X} = [0, 1]$ *and* $k(x, x') = \min(x, x')$. *Then* $\mathcal{H}$ *includes all linear combinations of functions of the form* $x \mapsto \min(x, x')$, *where* $x' \in [0, 1]$, *and their pointwise limits. These functions are*



Since we allow arbitrary linear combinations and pointwise limits, this gives rise to a large class of functions. In particular, this includes all Lipschitz functions that are 0 *at the origin.*

Theorem 15 (Representer theorem)**.** *Let* $\mathcal{H}$ *be an RKHS with reproducing kernel* k. *Let* c *be an arbitrary loss function and* $J : [0, \infty) \to \mathbb{R}$ *any strictly increasing function. Then the minimizer* $\hat{f} \in \mathcal{H}$ *of*

$$Q(f) = c(Y, x_1, \ldots, x_n, f(x_1), \ldots, f(x_n)) + J(\|f\|_{\mathcal{H}}^2)$$

lies in the linear span of $\{k(\cdot, x_i)\}_{i=1}^n$.

Proof. We have for any $f \in \mathcal{H}$,

$$f = u + v$$

where $u \in V = \text{vect}(\, k(\cdot, x_i) : i = 1, \ldots, n)$ (V is finite dimensional, therefore nonempty and closed) and $v \in V^{\perp}$. Then

$$f(x_i) = \langle f, k(\cdot, x_i)\rangle = \langle u + v, k(\cdot, x_i)\rangle = \langle u, k(\cdot, x_i)\rangle = u(x_i).$$

So we know that

$$c(Y, x_1, \ldots, x_n, f(x_1), \ldots, f(x_n)) = c(Y, x_1, \ldots, x_n, u(x_1), \ldots, u(x_n)).$$

Moreover,

$$\|f\|_{\mathcal{H}}^2 = \|u + v\|_{\mathcal{H}}^2 = \|u\|_{\mathcal{H}}^2 + \|v\|_{\mathcal{H}}^2,$$

using the fact that u and v are orthogonal. So we know

$$J(\|f\|_{\mathcal{H}}^2) \geq J(\|u\|_{\mathcal{H}}^2)$$

with equality iff $v = 0$. Hence $Q(f) \geq Q(u)$ with equality iff $v = 0$, and so we must have $v = 0$ by optimality. Thus, we know that the optimizer in fact lies in V. □

2.2 Ridge regression

Ridge regression can be written

$$\hat{f} = argmin_{f \in \mathcal{H}} \left\{ \sum_{i=1}^{n} (Y_i - f(x_i))^2 + \lambda \|f\|_{\mathcal{H}}^2 \right\}, \tag{2.6}$$

where $\mathcal{H}$ is the RKHS of the linear kernel.

From the representer theorem, we know that the solution can be written in the form

$$\hat{f}(\cdot) = \sum_{i=1}^{n} \hat{\alpha}_i k(\cdot, x_i),$$

for *any* RKHS, even for an infinite dimension one. and thus we can rewrite our optimization problem (2.6) as looking for the $\hat{\alpha} \in \mathbb{R}^n$ that minimizes

$$Q(\alpha) = c(Y, x_1, \ldots, x_n, K\alpha) + J(\alpha^T K \alpha),$$

over $\alpha \in \mathbb{R}^n$ (with $K_{ij} = k(x_i, x_j)$).

For the Ridge regression, (2.6) is equivalent to minimizing

$$\|Y - K\alpha\|_2^2 + \lambda \alpha^T K \alpha,$$

which amounts to the expression for the fitted values:

$$K\hat{\alpha} = K(K + \lambda I)^{-1} Y.$$

Consider a model

$$Y_i = f^0(x_i) + \varepsilon_i$$

for $i = 1, \ldots, n$, and assume $\mathbb{E}\varepsilon = 0$, $\mathbb{V}(\varepsilon) = \sigma^2 I$.

We assume f^0 is scaled such that $\|f^0\|_{\mathcal{H}}^2 \leq 1$. Let K be the kernel matrix $K_{ij} = k(x_i, x_j)$ with eigenvalues $d_1 \geq d_2 \geq \cdots \geq d_n \geq 0$. Define

$$\hat{f}_\lambda = \text{argmin}_{f \in \mathcal{H}} \left\{ \sum_{i=1}^n (Y_i - f(x_i))^2 + \lambda \|f\|_{\mathcal{H}}^2 \right\}.$$

Theorem 16. *We have*

$$\frac{1}{n} \sum_{i=1}^n \mathbb{E}(f^0(x_i) - \hat{f}_\lambda(x_i))^2 \leq \frac{\sigma^2}{n} \sum_{i=1}^n \frac{d_i^2}{(d_i + \lambda)^2} + \frac{\lambda}{4n} \tag{2.7}$$

$$\leq \frac{\sigma^2}{n} \frac{1}{\lambda} \sum_{i=1}^n \min\left(\frac{d_i}{4}, \lambda\right) + \frac{\lambda}{4n}. \tag{2.8}$$

Proof. We have that

$$(\hat{f}_\lambda(x_i), \ldots, \hat{f}_\lambda(x_n))^T = K(K + \lambda I)^{-1} Y.$$

Also, there is some $\alpha \in \mathbb{R}^n$ such that

$$(f^0(x_1), \ldots, f^0(x_n))^T = K\alpha.$$

We can write

$$\mathbb{E} \sum_{i=1}^n (f^0(x_i) - \hat{f}_\lambda(x_i))^2 = \mathbb{E}\|K\alpha - K(K + \lambda I)^{-1}(K\alpha + \varepsilon)\|_2^2$$

Noting that $K\alpha = (K + \lambda I)(K + \lambda I)^{-1}K\alpha$, we obtain

$$\begin{aligned} &= \mathbb{E}\|\lambda(K + \lambda I)^{-1}K\alpha - K(K + \lambda I)^{-1}\varepsilon\|_2^2 \\ &= \underbrace{\lambda^2\|(K + \lambda I)^{-1}K\alpha\|_2^2}_{\text{(A)}} + \underbrace{\mathbb{E}\|K(K + \lambda I)^{-1}\varepsilon\|_2^2}_{\text{(B)}}. \end{aligned}$$

Consider the eigen-decomposition $K = UDU^T$, where U is orthogonal, $D_{ii} = d_i$ and $D_{ij} = 0$ for $i \neq j$. We get

$$\begin{aligned} \text{(A)} &= \lambda^2\|U(D + \lambda I)^{-1}\underbrace{DU^T\alpha}_{\theta}\|_2^2 \\ &= \sum_{i=1}^{n} \theta_i^2 \frac{\lambda^2}{(d_i + \lambda)^2} \end{aligned}$$

Now we have

$$\alpha^T K\alpha = \alpha^T UDU^T\alpha = \alpha^T UDD^+DU^T = \sum_{d_i>0} \frac{\theta_i^2}{d_i},$$

where D^+ ithe pseudo-inverse of the diagonal matrix D. Using the scaling of f^0, we get

$$\alpha^T K\alpha = \|f^0\|_{\mathcal{H}}^2 \leq 1,$$

which implies that $\frac{\theta_i^2}{d_i} \leq 1$ when $d_i > 0$.
Since by definition of θ_i, we see that if $d_i = 0$, then $\theta_i = 0$. So have

$$\text{(A)} = \sum_{i:d_i\geq 0} \frac{\theta_i^2}{d_i} \frac{d_i\lambda^2}{(d_i + \lambda)^2} \leq \lambda \max_{i=1,\ldots,n} \frac{d_i\lambda}{(d_i + \lambda)^2} \leq \frac{\lambda}{4}.$$

Concerning (B), we have

$$\begin{aligned} \text{(B)} &= \mathbb{E}\left[\varepsilon^T(K + \lambda I)^{-1}K^2(K + \lambda I)^{-1}\varepsilon\right] \\ &= \mathbb{E}\left[\text{trace}\left(K(K + \lambda I)^{-1}\varepsilon\varepsilon^T(K + \lambda I)^{-1}K\right)\right] \\ &= \text{trace}\left(K(K + \lambda I)^{-1}\ \mathbb{E}\left[\varepsilon\varepsilon^T\right]\ (K + \lambda I)^{-1}K\right) \\ &= \sigma^2\ \text{trace}\left(K^2(K + \lambda I)^{-2}\right) \\ &= \sigma^2 \sum_{i=1}^{n} \frac{d_i^2}{(d_i + \lambda)^2}. \end{aligned}$$

The second inequality comes from $\frac{d_i^2}{(d_i+\lambda)^2} \leq \min\left(1, \frac{d_i}{4\lambda}\right)$.

□

2.3 Support vector machines

Consider for instance the problem of predicting whether an email is a spam.

One way is to find a hyperplane that separates the two sets $\{x_i\}_{y_i=1}$ and $\{x_i\}_{i:y_i=-1}$. To start with, supose the two classes are separated by a hyperplane through the origin, defined by a normal vector $\beta \in \mathbb{R}^p$ in the following way : $y_i\ x_i^T\beta > 0$ for all i. To maximize the chance of detecting each class, can maximize the minimum distance between any of the points (given by $\frac{y_i x_i^T \beta}{\|\beta\|_2}$) and the hyperplane defined by the normal vector β. We can formulate our problem as

$$\text{maximize } M \text{ among } \beta \in \mathbb{R}^p, M \geq 0 \text{ subject to } \frac{y_i x_i^T \beta}{\|\beta\|_2} \geq M.$$

This optimization problem gives the hyperplane that maximizes the *margin* M between the two classes.

We define $t_+ = t\mathbf{1}_{t\geq 0} = \max(t, 0)$. If we cannot separate the points with an hyperplane, we can penalyze points that are not on the right side, leading for instance to a penalty term

$$\lambda \sum_{i=1}^{n} \left(1 - \frac{y_i x_i^T \beta}{M\|\beta\|_2}\right)_+,$$

and to an optimization problem like

$$\max_{M\geq 0, \beta\in\mathbb{R}^p} \left(M - \lambda \sum_{i=1}^{n} \left(1 - \frac{y_i x_i^T \beta}{M\|\beta\|_2}\right)_+\right).$$

If we scale β, such that $\|\beta\|_2 = \frac{1}{M}$, the problem reads

$$\frac{1}{\|\beta\|_2} - \lambda \sum_{i=1}^{n} (1 - y_i x_i^T \beta)_+.$$

or, replacing $\max \frac{1}{\|\beta\|_2}$ with minimizing $\|\beta\|_2^2$ and modifying consequently the penalty part,

$$\min_{\beta\in\mathbb{R}^p} \left(\|\beta\|_2^2 + \lambda \sum_{i=1}^{n} (1 - y_i x_i^T \beta)_+\right).$$

Finally, replacing the scalar λ with $\frac{1}{\lambda}$ we get the optimization problem

$$\min_{\beta\in\mathbb{R}^p} \left(\lambda\|\beta\|_2^2 + \sum_{i=1}^{n} (1 - y_i x_i^T \beta)_+\right).$$

Now we consider hyperplane to passing through the origin, and get (by translating x_i's by a fixed vector $\delta \in \mathbb{R}^p$)

$$\min_{\beta\in\mathbb{R}^p,\delta\in\mathbb{R}^p} \left(\lambda\|\beta\|_2^2 + \sum_{i=1}^n \left(1 - y_i(x_i - \delta)^T\beta\right)_+ \right)$$

which in turn yields, by replacing $\delta^T\beta$ with a constant μ,

$$(\hat{\mu}, \hat{\beta}) = \text{argmin}_{(\mu,\beta)\in\mathbb{R}\times\mathbb{R}^p} \sum_{i=1}^n \left(1 - y_i(x_i^T\beta + \mu)\right)_+ + \lambda\|\beta\|_2^2. \tag{2.9}$$

This is the *support vector classifier.*

Let H be the RKHS corresponding to the linear kernel. We can then write (2.9) as

$$(\hat{\mu}_\lambda, \hat{f}_\lambda) = \textit{argmin}_{(\mu,f)\in\mathbb{R}\times\mathcal{H}} \sum_{i=1}^n (1 - y_i(f(x_i) + \mu))_+ + \lambda\|f\|_{\mathcal{H}}^2.$$

The representer theorem tells us that the above optimization problem is equivalent to the support vector machine

$$(\hat{\mu}_\lambda, \hat{\alpha}_\lambda) = \text{argmin}_{(\mu,\alpha)\in\mathbb{R}\times\mathbb{R}^n} \sum_{i=1}^n (1 - y_i(K_i^T\alpha + \mu))_+ + \lambda\alpha^T K\alpha$$

where $K_{ij} = k(x_i, x_j)$ and k is the reproducing kernel of H. Predictions at a new x are then given by

$$\text{sign}\left(\hat{\mu}_\lambda + \sum_{i=1}^n \hat{\alpha}_{\lambda,i} k(x, x_i)\right).$$

Another estimator, using $\log(1 + \exp(-u))$ instead of $\max(0, 1 - u)$ can be obtained through the problem

$$\text{argmin}_{\ b\in\mathbb{R}^p} \sum_{i=1}^n \log(1 + \exp(-y_i x_i^T\beta)).$$

We can e introduce an error term of $\lambda\|\beta\|_2^2$. The representer theorem can be used to yield the kernelized version

$$\text{argmin}_{f\in\mathcal{H}} \left(\sum_{i=1}^n \log(1 + \exp(-y_i f(x_i))) + \lambda\|f\|_{\mathcal{H}}^2 \right).$$

We can then solve this using the representer theorem.

Chapter 3

Uncertainty quantification and Sobol indices

3.1 Notations

Let $(\Omega, \mathcal{A}, P)$ be a probability space and let Y be the output of a deterministic functon η of a random vector $\mathbf{X} = (X_1, \cdots, X_p) \in \mathbb{R}^p$, $p \geq 1$ and that P_X is the pushforward measure of P by $\mathbf{X}$,

$$Y : \begin{array}{ccccccc} (\Omega, \mathcal{A}, P) & \rightarrow & (\mathbb{R}^p, \mathcal{B}(\mathbb{R}^p), P_X) & \rightarrow & (\mathbb{R}, \mathcal{B}(\mathbb{R})) \\ \omega & \mapsto & \mathbf{X}(\omega) & \mapsto & \eta(\mathbf{X}(\omega)) \end{array}$$

Let ν be a σ–finite measure on $(\mathbb{R}^p, \mathcal{B}(\mathbb{R}^p))$. Assume that $P_X << \nu$ and let p_X be the density of P_X with respect to ν, that is $p_X = \dfrac{dP_X}{d\nu}$.
Also, assume that $\eta \in L^2_{\mathbb{R}}(\mathbb{R}^p, \mathcal{B}(\mathbb{R}^p), P_X)$. The associated innner product of this Hilbert space is:

$$\langle h_1, h_2 \rangle = \int h_1(\mathbf{x}) h_2(\mathbf{x}) p_X d\nu(\mathbf{x}) = \mathbb{E}(h_1(\mathbf{X}) h_2(\mathbf{X}))$$

Here $\mathbb{E}(\cdot)$ denotes the expectation. The corresponding norm will be classically denoted by $\| \cdot \|$.
Further, $V(\cdot) = \mathbb{E}[(\cdot - \mathbb{E}(\cdot))^2]$ denotes the variance, and $\mathrm{Cov}(\cdot, *) = \mathbb{E}[(\cdot - \mathbb{E}(\cdot))(* - \mathbb{E}(*))]$ the covariance.

Let $\mathcal{P}_p := \{1, \cdots, p\}$ and S be the collection of all subsets of $\mathcal{P}_p$.
Define $S^- := S \setminus \mathcal{P}_p$ as the collection of all subsets of $\mathcal{P}_p$ except $\mathcal{P}_p$ itself.

Further, let $X_u := (X_l)_{l \in u}$, $u \in S \setminus \{\emptyset\}$. We introduce the subspaces of $L^2_{\mathbb{R}}(\mathbb{R}^p, \mathcal{B}(\mathbb{R}^p), P_X)$ $(H_u)_{u \in S}$, $(H^0_u)_{u \in S}$ and H^0 . H_u is the set of all measurable and square integrable functions depending only on X_u. $H_\emptyset$ is the set of constants and is identical to $(H^0_\emptyset)_{u \in S}$. H^0_u, $u \in S \setminus \emptyset$, and H^0 are defined as follows:

$$H^0_u = \left\{ h_u(X_u) \in H_u,\ \langle h_u, h_v \rangle = 0, \forall\ v \subset u, \forall\ h_v \in H^0_v \right\}$$

$$H^0 = \left\{ h(X) = \sum_{u \in S} h_u(X_u), h_u \in H^0_u \right\}$$

3.2 Sobol sensitivity indices

In this section, we explain the Hoeffding-Sobol decomposition.

Let $\mathbf{x} = (x_1, \cdots, x_p) \in \mathbb{R}^p$ and assume that $\eta \in \mathbb{L}^2(\mathbb{R}^p, P_X)$. The decomposition consists in writting $\eta(\mathbf{x}) = \eta(x_1, \cdots, x_p)$ as the sum of increasing dimension functions:

$$\begin{aligned} \eta(\mathbf{x}) &= \eta_0 + \sum_{i=1}^{p} \eta_i(x_i) + \sum_{1 \le i < j \le p} \eta_{i,j}(x_i, x_j) + \cdots + \eta_{1,\cdots,p}(\mathbf{x}) \\ &= \sum_{u \subseteq \{1 \cdots p\}} \eta_u(x_u) \end{aligned} \tag{3.1}$$

The expansion (3.1) exists and is unique under one of the following assumption

$$\int \eta_u(x_u)\eta_v(x_v) dP_X = 0 \qquad \forall\ u, v \subseteq \{1 \cdots p\}, \quad u \neq v$$

It is possible to show that

$$\eta_0 = \mathbb{E}(X), \quad \eta_i = \mathbb{E}(Y|X_i) - \mathbb{E}(Y),\ i = 1, \cdots, p, \quad \eta_u = \mathbb{E}(Y|X_u) - \sum_{v \subset u} \eta_v,\ |u| \ge 2 \tag{3.2}$$

The independence of the inputs and the orthogonality properties ensure that $V(Y) = \sum_{u \in S} V(\eta_u(X_u))$.

The Sobol indices expressions are defined by:

$$S_u = \frac{V(\eta_u)}{V(Y)} = \frac{V[\mathbb{E}(Y|X_u)] - \sum_{v \subset u} V[\mathbb{E}(Y|X_v)]}{V(Y)}, \quad u \subseteq \mathcal{P}_p \tag{3.3}$$

Furthermore,

$$\sum_{u \in S} S_u = 1$$

We want therefore to estimate the quantity

$$S_u = \frac{V(\eta_u)}{V(Y)}$$

using samples from the output Y. The process involve sampling $y = \eta(x)$: $y^j = \eta(x^j)$. For this purpose, it was proposed by Janon, Klein, Lagnoux, Nodet and Prieur to use the mathematical relation

$$V[\mathbb{E}(Y|X_i, i \in u)]] = \mathrm{Cov}[Y, Y^u] \tag{3.4}$$

Exploiting this relation yields the so called pick-and-freeze approach we describe now. Let x and z be independent vector distributed as the input. For instance, we can assume that there uniformly disctributed in the square $[0,1]^n$. A complementary evaluation is performed for the random variable $y_i = \eta(x_i, z_{i^c})$, where

$$(x_i, z_{i^c})_\ell = \begin{cases} x_i \text{ if } \ell = i \\ z_l \text{ if } \ell \neq i \end{cases}$$

Perfoming the following sampling: for $j \in \{1, \dots n\}$

$$\begin{cases} y^j = \eta(x^j) \\ y_i^j = \eta((x_i, z_{i^c})^j) \end{cases}$$

We then obtain

$$\hat{S}_{i,n} = \frac{\frac{1}{n}\sum_{j=1}^n y^j y_i^j - \left(\frac{1}{n}\sum_{j=1}^n y^j y_i^j\right)\left(\frac{1}{n}\sum_{j=1}^n y^j y_i^j\right)}{\left(\frac{1}{n}\sum_{j=1}^n (y^j)^2\right) - \left(\frac{1}{n}\sum_{j=1}^n y^j\right)^2}.$$

Another pick and freeze estimator of the Sobol index i is defined by (see Gamboa, Janon, Klein, Lagnoux, Prieur)

$$\hat{S}_{i,n} = \frac{\frac{1}{n}\sum_{j=1}^{n} y^j y_i^j - \left(\frac{1}{2n}\sum_{j=1}^{n}\left(y^j + y_i^j\right)\right)^2}{\left(\frac{1}{2n}\sum_{j=1}^{n}(y^j)^2 + (y_i^j)^2\right) - \left(\frac{1}{2n}\sum_{j=1}^{n} y^j + y_i^j\right)^2}.$$

3.3 Example

Let consider the model

$$y = \prod_{i=1}^{p} f_i(x_i) \text{ with } f_i(x_i) = \frac{|4x_i - 2| + a_i}{1 + a_i}$$

We assume that x is distributed uniformly in $[0,1]^p$. Evaluate the Sobol indices in the following cases:

1. $d = 3$ and $a = (0, 1, 9)$
2. $d = 12$ and $a = (0, 0, 0, 0, 1, 1, 1, 1, 9, 9, 9, 9)$
3. $d = 24$ and $a = \left(\underbrace{0, \ldots 0}_{8 \text{ times}}, \underbrace{1, \ldots 1}_{8 \text{ times}}, \underbrace{9, \ldots 9}_{8 \text{ times}}\right)$,

and give an explanation for the values you obtain.