Self-study: Reproducing Kernel Hilbert Space

Tianpei Xie

Jul. 6th., 2015

Contents

1 Hilbert Space and Functional Analysis Basis

1.1 Complete Metric Space

- **Definition** A ***metric space*** is a set M and a real-valued function $d(\cdot, \cdot) : M \times M \to \mathbb{R}$ which satisfies:

  1. (***Non-Negativity***) $d(x, y) \geq 0$
  2. (***Definiteness***) $d(x, y) = 0$ if and only if $x = y$
  3. (***Symmetric***) $d(x, y) = d(y, x)$
  4. (***Triangle Inequality***) $d(x, z) \leq d(x, y) + d(y, z)$

  The function d is called a ***metric*** on M. The metric space M equipped with metric d is denoted as (M, d).

- **Definition** (***Cauchy Sequence***)
  A sequence of elements $\{x_n\}$ of a metric space (M, d) is called a ***Cauchy sequence*** if $\forall \epsilon > 0$, there exists $N \in \mathbb{N}$, for all $n, m \geq N$, $d(x_n, x_m) < \epsilon$.

- **Proposition 1.1** *Any convergent sequence is a Cauchy sequence.*

  Note that this is the direct result of *triangle inequality property of a metric.*

- **Definition** (***Complete Metric Space***)
  A metric space in which ***all Cauchy sequences converge*** is called ***complete***.

- **Definition** (***Denseness***)
  A set B in a *metric space* M is called ***dense*** if *every* $m \in M$ *is a limit of elements in* B.

- **Definition** (***Continuity***)
  A function $f : (X, d) \to (Y, p)$ is called ***continuous*** at x if $f(x_n) \xrightarrow{p} f(x)$ whenever $x_n \xrightarrow{d} x$.

- **Definition** (***Isometry***)
  A ***bijection*** $h : (X, d) \to (Y, p)$ which ***preserves*** *the metric*, that is,

  $$p(h(x), h(y)) = d(x, y)$$

  is called an ***isometry***. It is automatically *continuous*. (X, d) and (Y, p) are said to be ***isometric*** if such an isometry exists.

- **Definition** (***Normed Linear Space***)
  A ***normed linear space*** is a vector space, V, over $\mathbb{R}$ (or $\mathbb{C}$) and a function, $\|\cdot\| : V \to \mathbb{R}$ which satisfies:

  1. (***Non-Negativity***): $\|v\| \geq 0$ for all v in V;
  2. (***Positive Definiteness***): $\|v\| = 0$ if and only if $v = 0$;
  3. (***Absolute Homogeneity***) $\|\alpha v\| = |\alpha| \, \|v\|$ for all v in V and α in $\mathbb{R}$ (or $\mathbb{C}$)
  4. (***Subadditivity / Triangle Inequality***) $\|v + w\| \leq \|v\| + \|w\|$ for all v and w in V

  We denote the normed linear space as $(V, \|\cdot\|)$.

1.2 Hilbert Space

- **Definition** An *inner product space* (*pre-Hilbert space*) X is an abstract vector space possessing the structure of an inner product that allows length and angle to be measured. An inner product on X is a mapping on $X \times X$ to a scale field E of X; that is, for every pair $\boldsymbol{x}, \boldsymbol{y} \in X$, the associated scalar in E as the inner product, denoted as $\langle \boldsymbol{x}\,,\, \boldsymbol{y} \rangle$ satisfies the following properties

  1. Addition $\langle \boldsymbol{x} + \boldsymbol{y}\,,\, \boldsymbol{z} \rangle = \langle \boldsymbol{x}\,,\, \boldsymbol{z} \rangle + \langle \boldsymbol{y}\,,\, \boldsymbol{z} \rangle$ for all $\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z} \in X$;
  2. Scalar product $\langle a\boldsymbol{x}\,,\, \boldsymbol{y} \rangle = a \langle \boldsymbol{x}\,,\, \boldsymbol{y} \rangle$, for all $\boldsymbol{x}, \boldsymbol{y} \in X$, $a \in E$;
  3. Hermitian $\langle \boldsymbol{x}\,,\, \boldsymbol{y} \rangle = \overline{\langle \boldsymbol{y}\,,\, \boldsymbol{x} \rangle}$;
  4. Nonegative $\langle \boldsymbol{x}\,,\, \boldsymbol{x} \rangle \geq 0$ with equality holds iff $\boldsymbol{x} = \boldsymbol{0}$.

- **Remark** A *norm* is induced by inner product via

$$\|\boldsymbol{x}\| = \sqrt{\langle \boldsymbol{x}\,,\, \boldsymbol{x} \rangle}$$

  and a *metric* is defined via

$$d(\boldsymbol{x}, \boldsymbol{y}) = \sqrt{\langle \boldsymbol{x} - \boldsymbol{y}\,,\, \boldsymbol{x} - \boldsymbol{y} \rangle} = \|\boldsymbol{x} - \boldsymbol{y}\|\,.$$

- **Definition** A ***complete*** *inner product space* is called ***a Hilbert space***.

  Inner product spaces are sometimes called ***pre-Hilbert spaces***.

- **Remark** Consider a *complex* Hilbert space F_c, where a function $f_1 + i\, f_2 \in F_c$ for every $f_1, f_2 \in F$, a Hilbert space of real valued functions. Note that $\|f_1 + i\, f_2\|_2^2 = \|f_1\|_2^2 + \|f_2\|_2^2$. The following property holds:

  1. If $f \in F_c$, then $\overline{f} \in F_c$;
  2. $\|f\| = \|\overline{f}\|$.

- **Definition** (***Complete Orthonormal Basis***)
  If S is *an orthonormal set* in a Hilbert space $\mathcal{H}$ and *no other orthonormal set* contains S as a proper subset, then S is called *an* ***orthonormal basis*** (or a ***complete orthonormal system***) for $\mathcal{H}$.

- **Theorem 1.2** ***(Existence of Orthonormal Basis)***
  Every Hilbert space $\mathcal{H}$ has an ***orthonormal basis****.*

- **Proposition 1.3** ***(Orthogonal Representation of Element in Hilbert Space)***
  Let $\mathcal{H}$ be a Hilbert space and $S = (x_\alpha)_{\alpha \in A}$ an ***orthonormal basis****. Then for each $y \in \mathcal{H}$,*

$$y = \sum_{\alpha \in A} \langle y\,,\, x_\alpha \rangle\, x_\alpha \tag{1}$$

  and

$$\|y\|_{\mathcal{H}} = \sum_{\alpha \in A} |\langle y\,,\, x_\alpha \rangle|^2 \tag{2}$$

The equality in (1) *means that the sum on the right-hand side converges (independent of order) to y in $\mathcal{H}$.* ***Conversely****, if $\sum_{\alpha\in A}|c_\alpha|^2 < \infty$, $c_\alpha \in \mathbb{C}$, then $\sum_{\alpha\in A} c_\alpha x_\alpha$ converges to an element of $\mathcal{H}$.*

- **Remark** ***Orthogonality*** is the central concept of Hilbert space. In the presence of closed subspaces, the orthogonality allows us to decompose the Hilbert space into the direct sum of the *subspace* and its *orthogonal complement*.

- **Definition** (***Direct Sum***)
  Suppose that $\mathcal{H}_1$ and $\mathcal{H}_2$ are Hilbert spaces. Then the set of *pairs* (x, y) with $x \in \mathcal{H}_1, y \in \mathcal{H}_2$ is a *Hilbert space* with *inner product*

  $$\langle (x_1, y_1)\, ,\, (x_2, y_2)\rangle := \langle x_1\, ,\, x_2\rangle_{\mathcal{H}_1} + \langle y_1\, ,\, y_2\rangle_{\mathcal{H}_2}$$

  This space is called ***the direct sum*** *of the spaces $\mathcal{H}_1$ and $\mathcal{H}_2$* and is denoted by $\mathcal{H}_1 \oplus \mathcal{H}_2$.

- **Definition** (***Orthogonal Complement***)
  Let $\mathcal{M} \subseteq \mathcal{H}$ is a ***closed*** *linear subspace* of Hilbert space $\mathcal{H}$ with *induced inner product* $\langle\, ,\, \rangle$ (i.e. $\langle x\, ,\, y\rangle_{\mathcal{M}} = \langle x\, ,\, y\rangle_{\mathcal{H}}$ for all $x, y \in \mathcal{M}$). $\mathcal{M}$ is also a *Hilbert space*.

  We denote by $\mathcal{M}^\perp$ the set of vectors in $\mathcal{H}$ which are *orthogonal* to $\mathcal{M}$; $\mathcal{M}^\perp$ is called ***the orthogonal complement of*** $\mathcal{M}$. It follows from the linearity of the inner product that $\mathcal{M}^\perp$ is a *linear subspace* of $\mathcal{H}$ and an elementary argument shows that $\mathcal{M}^\perp$ is *closed*. So $\mathcal{M}^\perp$ is also a *Hilbert space*.

- **Lemma 1.4** *Let $\mathcal{H}$ be a Hilbert space, $\mathcal{M}$ a closed subspace of $\mathcal{H}$, and suppose $x \in \mathcal{H}$. Then there exists in $\mathcal{M}$ a* ***unique*** *element z* ***closest*** *to x.*

- **Theorem 1.5** ***(The Projection Theorem)***
  Let $\mathcal{H}$ be a Hilbert space, $\mathcal{M}$ a closed subspace. Then every $x \in \mathcal{H}$ can be ***uniquely*** *written $x = z + w$ where $z \in \mathcal{M}$ and $w \in \mathcal{M}^\perp$.*

- **Remark** The projection theorem sets up a natural *isomorphism* $\mathcal{M} \oplus \mathcal{M}^\perp \to \mathcal{H}$ given by

  $$(z, w) \mapsto z + w$$

  We will often suppress the isomorphism and simply write $\mathcal{H} = \mathcal{M} \oplus \mathcal{M}^\perp$.

- **Definition** (***Separability***)
  A *metric space* which has a ***countable dense subset*** is said to be ***separable***.

- **Remark** Most Hilbert space we have seen is separable.

- **Proposition 1.6** ***(Canonical Hilbert Space)***
  A Hilbert space $\mathcal{H}$ is ***separable*** *if and only if it has a* ***countable orthonormal basis*** *S. If there are $N < \infty$ elements in S, then $\mathcal{H}$ is* ***isomorphic*** *to $\mathbb{C}^N$, If there are* ***countably many*** *elements in S, then $\mathcal{H}$ is* ***isomorphic*** *to ℓ^2.*

1.3 Bounded Linear Operator and Dual Space

- **Definition** (***Bounded Linear Operator***)
  A ***bounded linear transformation*** (or ***bounded operator***) is a mapping $T : (X, \|\cdot\|_X) \to (Y, \|\cdot\|_Y)$ from a normed linear space X to a normed linear space Y that satisfies

1. (***Linearity***) $T(\alpha x + \beta y) = \alpha T(x) + \beta T(y)$ for all $x, y \in X$, $\alpha, \beta \in \mathbb{R}$ or $\mathbb{C}$
2. (***Boundedness***) $\|Tx\|_Y \le C\, \|x\|_X$ for small $C \ge 0$.

The smallest such C is called *the **norm** of T*, written $\|T\|$ or $\|T\|_{X,Y}$. Thus

$$\|T\| := \sup_{\|x\|_X = 1} \|Tx\|_Y$$

- **Remark** Denote the space of ***all bounded linear operator*** between Hilbert space $\mathcal{H}_1$ and $\mathcal{H}_2$ as $\mathcal{L}(\mathcal{H}_1, \mathcal{H}_2)$. The space $\mathcal{L}(\mathcal{H}_1, \mathcal{H}_2)$ is linear space with norm

$$\|T\| := \sup_{\|x\|_{\mathcal{H}_1} = 1} \|Tx\|_{\mathcal{H}_2}, \quad \forall T \in \mathcal{L}(\mathcal{H}_1, \mathcal{H}_2).$$

It can be shown that $\mathcal{L}(\mathcal{H}_1, \mathcal{H}_2)$ is a *complete normed space (i.e. a Banach space).*

- **Definition** (***Dual Space***)
The space $\mathcal{L}(\mathcal{H}, \mathbb{C})$ is called the ***dual space*** of $\mathcal{H}$ and is denoted by $\mathcal{H}^*$. The elements of $\mathcal{H}^*$ are called ***continuous linear functionals***. That is, the dual space $\mathcal{H}^*$ is the space of *continuous linear functionals* on $\mathcal{H}$.

- **Remark** The *dual space* $\mathcal{H}^*$ is also called ***covector space*** with respect to a vector space $\mathcal{H}$ and the linear functionals are called ***covectors***. This terms are mostly used in *differential geometry* when the vector space is *the tangent space*.

- **Theorem 1.7** *(**The Riesz Representation Theorem**) [Reed and Simon, 1980, Kreyszig, 1989, Conway, 2019]*
*For each $T \in \mathcal{H}^*$, there is a **unique** $y_T \in \mathcal{H}$ such that*

$$T(x) = \langle x\, ,\, y_T \rangle$$

for all $x \in \mathcal{H}$. In addition $\|y_T\|_{\mathcal{H}} = \|T\|_{\mathcal{H}^*}$.

- **Remark** The *the Riesz representation theorem* together with *the Cauchy-Schwarz inequality* defines an ***isomorphism*** $\mathcal{H}^* \to \mathcal{H}$ between a Hilbert space $\mathcal{H}$ and its dual $\mathcal{H}^*$. In other words, the bounded linear functional on Hilbert space has a simple form.

- **Corollary 1.8** *(**The Riesz Representation for Sesquilinear Form**)*
Let $B(\cdot, \cdot)$ be a function from $\mathcal{H} \times \mathcal{H}$ to $\mathbb{C}$ which satisfies:

  1. *(**Linearity**) $B(\alpha x + \beta y, z) = \alpha B(x, z) + \beta B(y, z)$*
  2. *(**Conjugate Linearity**) $B(x, \alpha y + \beta z) = \overline{\alpha} B(x, y) + \overline{\beta} B(x, z)$*
  3. *(**Boundedness**) $|B(x, y)| \le C\, \|x\|_{\mathcal{H}}\, \|y\|_{\mathcal{H}}$*

*for all $x, y, z \in \mathcal{H}$, $\alpha, \beta \in \mathbb{C}$. Then there is a **unique bounded linear transformation** $A : \mathcal{H} \to \mathcal{H}$ so that*

$$B(x, y) = \langle Ax\, ,\, y \rangle$$

*for all $x, y \in \mathcal{H}$. The **norm** of A is the smallest constant C such that (3) holds.*

- **Remark** A bilinear function on $\mathcal{H}$ obeying (1) and (2) is called a ***sesquilinear form*** (as a generalization of ***blinear form*** in complex vector space).

In terms of this, an inner product in complex vector space is ***a complex Hermitian form*** (also called a ***symmetric sesquilinear form***).

1.4 Hilbert Adjoints and Self Adjoint Operator

- **Definition** (***Hilbert Space Adjoint***)
Let $T : \mathcal{H}_1 \to \mathcal{H}_2$ be a *bounded linear operator*, where $\mathcal{H}_1$ and $\mathcal{H}_2$ are *Hilbert spaces*. Then ***the Hilbert-adjoint operator*** T^* of T is the operator
$$T^* : \mathcal{H}_2 \to \mathcal{H}_1$$
such that for all $x \in \mathcal{H}_1$ and $y \in \mathcal{H}_2$,
$$\langle Tx\,,\, y\rangle_{\mathcal{H}_2} = \langle x\,,\, T^*y\rangle_{\mathcal{H}_1} \tag{3}$$

- **Proposition 1.9** *(**Existence of Adjoint Operator**) [Kreyszig, 1989]*
*The Hilbert-adjoint operator T^* of T **exists**, is **unique** and is a **bounded linear operator** with norm*
$$\|T^*\| = \|T\|\,.$$

- **Proposition 1.10** *(**Properties of Hilbert-adjoint operators**). [Reed and Simon, 1980, Kreyszig, 1989]*
Let $\mathcal{H}_1$, $\mathcal{H}_2$ be Hilbert spaces, $S : \mathcal{H}_1 \to \mathcal{H}_2$ and $T : \mathcal{H}_1 \to \mathcal{H}_2$ bounded linear operators and α any scalar. Then we have
  1. $\langle T^*y\,,\, x\rangle = \langle y\,,\, Tx\rangle$, *$(x \in H_1, y \in \mathcal{H}_2)$*
  2. $(S + T)^* = S^* + T^*$
  3. $(\alpha T)^* = \alpha T^*$
  4. $(T^*)^* = T$
  5. $\|T^*T\| = \|TT^*\| = \|T\|^2$
  6. $T^*T = 0 \Leftrightarrow T = 0$
  7. $(ST)^* = T^*S^*$ *(assuming $\mathcal{H}_2 = \mathcal{H}_1$)*
  8. *If T has a **bounded inverse**, T^{-1}, then T^* has a **bounded inverse** and $(T^*)^{-1} = (T^{-1})^*$.*

- **Definition** A ***bounded linear operator*** $T : \mathcal{H} \to \mathcal{H}$ on a Hilbert space $\mathcal{H}$ is said to be
  1. ***self-adjoint*** or ***Hermitian*** if
$$T^* = T \quad \Leftrightarrow \quad \langle Tx\,,\, y\rangle = \langle x\,,\, Ty\rangle$$
  2. ***unitary*** if T is *bijective* and
$$T^* = T^{-1}$$

3. ***normal*** if

$$T^*T = TT^*$$

- **Definition** (***Projection Operator***)
If $P \in \mathcal{L}(\mathcal{H})$ and $P^2 = P$, then P is called a ***projection***. If in addition $P = P^*$, then P is called an ***orthogonal projection***.

- **Remark** If T is ***self-adjoint*** and ***unitary***, then T is ***normal***.

- **Proposition 1.11** *(**Self-adjointness**). [Kreyszig, 1989]*
Let $T : \mathcal{H} \to \mathcal{H}$ be a bounded linear operator on a Hilbert space $\mathcal{H}$. Then:

  1. *If T is **self-adjoint**, $\langle Tx\,,\, x\rangle$ is **real** for all $x \in \mathcal{H}$.*
  2. *If $\mathcal{H}$ is complex and $\langle Tx\,,\, x\rangle$ is **real** for all $x \in \mathcal{H}$, the operator T is **self-adjoint***

- **Proposition 1.12** *(**Self-adjointness of product**). [Kreyszig, 1989]*
*The product of two bounded **self-adjoint** linear operators S and T on a Hilbert space $\mathcal{H}$ is **self-adjoint** if and only if the operators **commute**,*

$$ST = TS.$$

- **Proposition 1.13** *(**Sequences of self-adjoint operators**). [Kreyszig, 1989]*
*Let (T_n) be a sequence of **bounded self-adjoint** linear operators $T_n : \mathcal{H} \to \mathcal{H}$ on a Hilbert space $\mathcal{H}$. Suppose that (T_n) converges, say,*

$$T_n \to T, \quad i.e. \;\; \|T_n - T\| \to 0$$

*where $\|cdot\|$ is the norm on the space $\mathcal{L}(\mathcal{H}, \mathcal{H})$. Then the limit operator T is a **bounded self-adjoint** linear operator on H.*

- **Proposition 1.14** *(**Unitary operator**). [Kreyszig, 1989]*
*Let the operators $U : \mathcal{H} \to \mathcal{H}$ and $V : \mathcal{H} \to \mathcal{H}$ be **unitary**; here, $\mathcal{H}$ is a Hilbert space. Then:*

  1. *U is **isometric**; thus $\|Ux\| = \|x\|$ for all $x \in \mathcal{H}$;*
  2. *$\|U\| = 1$, provided $\mathcal{H} \neq \{0\}$,*
  3. *$U^{-1} = U^*$ is **unitary**,*
  4. *UV is unitary,*
  5. *U is normal.*
  6. *A bounded linear operator T on a complex Hilbert space $\mathcal{H}$ is **unitary** if and only if T is **isometric** and **surjective**.*

1.5 Regular Measure and Duality of $\mathcal{C}_0(X)$

- **Definition** (***Subspace of Continuous Functions***)
Let $\mathcal{C}(X) := \mathcal{C}(X, \mathbb{R})$ be *the space of **continuous** real-valued functions* on topological space X and $\mathcal{B}(X) := \mathcal{B}(X, \mathbb{R})$ be *the space of **bounded** real-valued functions* on X.

1. The intersection of $\mathcal{B}(X)$ and $\mathcal{C}(X)$ is the space of all ***bounded continuous*** *functions*

$$\mathcal{BC}(X) := \mathcal{BC}(X, \mathbb{R}) = \mathcal{B}(X, \mathbb{R}) \cap \mathcal{C}(X, \mathbb{R})$$

   Note that $\mathcal{BC}(X) \subseteq \mathcal{B}(X)$ is a ***closed subspace***.

2. Define the ***support*** of a function f, $\text{supp}(f)$ as *the* ***smallest closed set*** *outside of which f vanishes*. The subset $\mathcal{C}_c(X) \subseteq \mathcal{C}(X)$ is the space of all *continuous functions* with ***compact support***

$$\mathcal{C}_c(X) = \{f \in \mathcal{C}(X, \mathbb{R}) : \text{supp }(f) \textit{ is compact}\}.$$

   Note that by *Tietze Extension Theorem*, *the locally compact Hausdorff space* X has a rich supply of continuous functions that vanishes outside a compact set.

3. Recall also that $\mathcal{C}_0(X)$ is the space of *continuous functions* on X that ***vanishes at infinity***, i.e. for all $\epsilon > 0$, $|f(x)| < \epsilon$ if $x \in X \setminus C$ for some ***compact subset*** $C \subseteq X$.

$$\mathcal{C}_0(X) = \{f \in \mathcal{C}(X, \mathbb{R}) : f \textit{ vanishes at infinity}\}.$$

Note that

$$\mathcal{C}_c(X) \subseteq \mathcal{C}_0(X) \subseteq \mathcal{BC}(X) \subseteq \mathcal{C}(X)$$

- **Definition** (***Radon Measure***) [Folland, 2013]
A ***Radon measure*** μ on X is a *Borel measure* that is

  1. ***finite*** on *all* ***compact*** *sets*; i.e. for any ***compact subset*** $K \subseteq X$,

$$\mu(K) < \infty.$$

  2. ***outer regular*** on *all Borel sets*; i.e. for any *Borel set* E

$$\mu(E) = \inf\{\mu(U) : E \subseteq U, U \textit{ is open}\}.$$

  3. ***inner regular*** on all *open sets*; i.e. for any *open set* E

$$\mu(E) = \sup\{\mu(C) : C \subseteq E, C \textit{ is compact and Borel}\}.$$

- **Definition** (***Complex Radon Measure***)
A ***signed Radon measure*** is a ***signed Borel measure*** whose ***positive*** and ***negative variations*** are ***Radon***, and a ***complex Radon measure*** is a ***complex Borel measure*** whose *real and imaginary parts* are *signed Radon measures.*

- **Definition** (***Space of Complex Radon Measures***)
On *locally compact Hausdorff space* X, We denote *the space of complex Radon measures on* X by $\mathcal{M}(X)$. For $\mu \in \mathcal{M}(X)$ we define

$$\|\mu\| = |\mu|(X),$$

where $|\mu|$ is the ***total variation*** of μ.

- **Theorem 1.15** ***(The Riesz-Markov Theorem, Locally Compact Version)*** *[Reed and Simon, 1980, Folland, 2013]*
  Let X be a ***locally compact Hausdorff*** *space. For any continuous linear functional I on $\mathcal{C}_0(X)$, (the space of* continuous functions *on X that vanishes at infinity), there is a* ***unique regular countably additive complex Borel measure*** *μ on X such that*

$$I(f) = \int_X f d\mu, \quad \text{for all } f \in \mathcal{C}_0(X).$$

  The ***norm*** *of I as a linear functional is* ***the total variation*** *of μ, that is*

$$\|I\| = |\mu|(X).$$

  Finally, I is ***positive*** *if and only if the measure μ is* ***non-negative****.*

- **Remark** In other word, the map $\mu \mapsto I_\mu$, is an ***isometric isomorphism*** from $\mathcal{M}(X)$ to $(\mathcal{C}_0(X))^*$, or

$$\mathcal{M}(X) \simeq (\mathcal{C}_0(X))^*.$$

- **Corollary 1.16** *[Reed and Simon, 1980, Folland, 2013]*
  Let X be a ***compact Hausdorff*** *space. Then the* ***dual space*** *$\mathcal{C}(X)^*$ is* ***isometric isomorphism*** *to $\mathcal{M}(X)$.*

- **Definition** Given $\mathcal{M}(X) \simeq (\mathcal{C}_0(X))^*$, we define subspaces of $\mathcal{M}$:

$$\begin{aligned} \mathcal{M}_+(X) &= \{I \in \mathcal{M}(X) : I \text{ is a positive linear functional}\}, \\ \mathcal{M}_{+,1}(X) &= \{I \in \mathcal{M}(X) : \|I\| = 1\}. \end{aligned}$$

  Thus $\mathcal{M}_+(X)$ is *identified* with ***the space of all positive Randon measures on*** X.

- **Remark** (***Isometric Embedding of*** $L^1(\mu)$ ***into*** $M(X)$)
  Let μ be a fixed *positive Radon measure* on X. If $f \in L^1(\mu)$, *the complex measure*

$$d\nu_f = f d\mu$$

  is easily seen to be ***Radon***, and $\|\nu\| = \int |f| d\mu = \|f\|_1$. Thus $f \mapsto \nu_f$ is an ***isometric embedding*** of $L^1(\mu)$ into $M(X)$ whose range consists precisely of those $\nu \in \mathcal{M}(X)$ such that $\nu \ll \mu$.

- **Proposition 1.17** *($\mathcal{M}(X)$* ***is Normed Linear Space****)* *[Folland, 2013]*
  If μ is a ***complex Borel measure****, then μ is* ***Radon*** *if and only if $|\mu|$ is* ***Radon****. Moreover, $\mathcal{M}(X)$ is a vector space and $\mu \mapsto \|\mu\|$ is a* ***norm*** *on it.*

- **Remark** (***Two Perspectives of Measures***)
  For *regular Borel measure* μ or in general, *Radon measures* on ***locally compact*** space X, there are two perspectives:

  1. ***Nonegative set function on the σ-algebra $\mathscr{A}$***: as a ***measure of the volume*** of a *subset* in X;

  2. ***Positive linear functional on*** $\mathcal{C}_0(X)$: as a ***integral*** *of compactly supported continuous functions with respect to* ***given measure****.*

  In some cases, it is important to think of ***measures*** not merely as individual objects but instead as *elements of* $(\mathcal{C}_0(X))^*$, so that we can employ *geometric* ideas.

1.6 Spectrum of Bounded Linear Operator

- **Definition** (***Resolvent*** *and* ***Spectrum***)
  Let $T \in \mathcal{L}(X)$. A complex number λ is said to be in ***the resolvent set*** $\rho(T)$ *of* T if

  $$\lambda I - T$$

  is a ***bijection*** with a ***bounded inverse***.

  $$R_\lambda(T) := (\lambda I - T)^{-1}$$

  is called ***the resolvent*** *of* T *at* λ. Note that $R_\lambda(T)$ is defined on $\mathrm{Ran}\,(\lambda I - T)$.

  If $\lambda \notin \rho(T)$, then λ is said to be in the ***spectrum*** $\sigma(T)$ ***of*** T.

- **Remark** The name "***resolvent***" is appropriate, since $R_\lambda(T)$ helps to solve the equation $(\lambda I - T)\,x = y$. Thus, $x = (\lambda I - T)^{-1}\,y = R_\lambda(T)y$ provided $R_\lambda(T)$ exists.

- **Definition** (***Point Spectrum****,* ***Continuous Spectrum*** *and* ***Residual Spectrum***)
  Let $T \in \mathcal{L}(X)$

  1. ***Point Spectrum***: An $x \neq 0$ which satisfies

     $$Tx = \lambda x$$
     $$\text{or } (\lambda I - T)\,x = 0, \quad \text{for some } \lambda \in \mathbb{C}$$

     is called an ***eigenvector*** *of* T; λ is called ***the corresponding eigenvalue***.

     If λ is an *eigenvalue*, then $(\lambda I - T)$ is ***not injective*** (i.e. $\mathrm{Ker}\,(\lambda I - T) \neq \{0\}$) so λ is *in the spectrum of* T. ***The set of all eigenvalues*** is called ***the point spectrum of*** T. It is denoted as $\sigma_p(T)$.

  2. ***Continuous Spectrum***: If λ is ***not an eigenvalue*** and if $\mathrm{Ran}\,(\lambda I - T)$ is ***dense*** but the resolvent $R_\lambda(T)$ is ***unbounded***, then λ is said to be in ***the continuous spectrum***. It is denoted as $\sigma_c(T)$.

  3. ***Residual Spectrum***: If λ is ***not an eigenvalue*** and if $\mathrm{Ran}\,(\lambda I - T)$ is ***not dense***, then λ is said to be in ***the residual spectrum***. It is denoted as $\sigma_r(T)$.

- **Remark** (***Pure Point Spectrum for Finite Dimensional Case***)
  If X is ***finite dimensional*** *normed linear space*, $T \in \mathcal{L}(X)$ then $\sigma_c(T) = \sigma_r(T) = \emptyset$.

- **Remark** If X is a function space, the *eigenvectors* of *linear operator* T is called the ***eigenfunctions*** of T.

- **Definition** (***Eigenspace of Linear Operator***)
  The subspace of domain $D(T)$ consisting of $\{0\}$ and ***all eigenvectors*** of T corresponding to *an eigenvalue* λ of T is called ***the eigenspace of*** T corresponding to that eigenvalue λ.

- **Definition** (***Spectral Radius of Linear Operator***)
  Let

  $$r(T) = \sup_{\lambda \in \sigma(T)} |\lambda|$$

  $r(T)$ is called ***the spectral radius of*** T.

- **Proposition 1.18** *(**Spectral Radius Calculation**) [Reed and Simon, 1980]*
  *Let X be a **Hilbert space**, $T \in \mathcal{L}(X)$ and T is **self-adjoint**. Then*

$$r(T) = \|T\|$$

- **Theorem 1.19** *(**Spectrum and Resolvent of Adjoint**) (**Phillips**) [Reed and Simon, 1980]*
  *If X is a **Hilbert space** and $T \in \mathcal{L}(X)$, then*

$$\sigma(T) = \sigma(T^*) \;\; and \;\; R_\lambda(T^*) = (R_\lambda(T))^*.$$

- **Proposition 1.20** *(**Spectrum of Self-Adjoint Operator**) [Reed and Simon, 1980]*
  *Let be a **self-adjoint operator** on a **Hilbert space** $\mathcal{H}$. Then,*

  1. *T has **no residual spectrum**, i.e. $\sigma_r(T) = \emptyset$.*
  2. *$\sigma(T)$ is a subset of $\mathbb{R}$.*
  3. ***Eigenvectors** corresponding to **distinct eigenvalues** of T are **orthogonal**.*

- **Definition** (***Positive-Semidefinite Operator***)
  Let $\mathcal{H}$ be a ***Hilbert space***. An operator $B \in \mathcal{L}(\mathcal{H})$ is called ***positive-semidefinite*** if

$$\langle Bx\,,\,x\rangle \geq 0 \text{ for all } x \in \mathcal{H}.$$

  We write $B \succeq 0$ if is *positive-semidefinite* and $B \succeq A$ if $(B - A) \succeq 0$.

  Similarly, B is called ***positive-definite*** if

$$\langle Bx\,,\,x\rangle > 0 \text{ for all } x \neq 0 \in \mathcal{H}.$$

  The *positive semidefinite operator* is sometimes called ***positive*** *operator*.

- **Proposition 1.21** *(**Positive Semi-Definiteness $\Rightarrow$ Self-Adjoint**) [Reed and Simon, 1980]*
  *Every (bounded) **positive semidefinite** operator on a **complex Hilbert space** is **self-adjoint**.*

  Theorem 1.22 *(**Square Root Lemma**) [Reed and Simon, 1980]*
  *Let $A \in \mathcal{L}(\mathcal{H})$ and $A \succeq 0$. Then there is a **unique** $B \in \mathcal{L}(\mathcal{H})$ with $B \succeq 0$ and $B^2 = A$. Furthermore, B **commutes** with every bounded operator which commutes with A.*

- **Definition** For $A \in \mathcal{L}(\mathcal{H})$, we can define ***absolute value*** *of* A as the square root of its normal operation

$$|A| := \sqrt{A^* A}$$

1.7 Compact Operator

- **Definition** (***Compact Operator***)
  Let X and Y be *Banach spaces*. An operator $T \in \mathcal{L}(X, Y)$ is called ***compact*** (or ***completely continuous***) if T takes ***bounded sets*** in X into ***precompact sets*** in Y.

  Equivalently, T is ***compact*** if and only if for every ***bounded*** *sequence* $\{x_n\} \subseteq X$, $\{Tx_n\}$ has a ***subsequence convergent*** in Y.

- **Example** (***Finite Rank Operators***)
  Suppose that *the* ***range*** *of* T *is* ***finite dimensional***. That is, every vector in the range of T can be written

  $$Tx = \sum_{i=1}^{n} \alpha_i y_i,$$

  for some fixed family $\{y_i\}_{i=1}^{n}$ in Y. If x_n is any *bounded sequence* in X, the corresponding $\alpha_i^{(n)}$ are *bounded* since T is *bounded*. The usual subsequence trick allows one to extract a *convergent subsequence* from $\{Tx_n\}$ which proves that T is *compact*. $\blacksquare$

- An important property of the compact operator is

  Theorem 1.23 *(**Weakly Convergent** + **Compact Operator** = **Uniformly Convergent**) [Reed and Simon, 1980]*
  *A **compact** operator maps **weakly convergent** sequences into **norm convergent** sequences; i.e. if $T \in \mathcal{L}(X)$ is compact, then*

  $$x_n \overset{w}{\to} x \quad \Rightarrow \quad Tx_n \overset{norm}{\to} Tx.$$

  *The converse holds true if X is **reflective**.*

- **Theorem 1.24** *(**Compact Operator Approximated by Finite Rank Operator**)[Reed and Simon, 1980]*
  *Let $\mathcal{H}$ be a **separable Hilbert space**. Then every **compact operator** on $\mathcal{H}$ is the **norm limit** of a sequence of operators of **finite rank**.*

- **Theorem 1.25** *(**Analytic Fredholm Theorem**) [Reed and Simon, 1980]*
  *Let D be an **open connected** subset of $\mathbb{C}$. Let $f : D \to \mathcal{L}(\mathcal{H})$ be an **analytic operator-valued function** such that $f(z)$ is **compact** for each $z \in D$. Then, either*

  1. *$(I - f(z))^{-1}$ exists for **no** $z \in D$; or*
  2. *$(I - f(z))^{-1}$ exists for **all** $z \in D \setminus S$ where S is a **discrete** subset of D (i.e. S is a set which has no limit points in D.) In this case, $(I - f(z))^{-1}$ is **meromorphic** in D, **analytic** in $D \setminus S$, the **residues** at the poles are **finite rank operators**, and if $z \in S$ then*

     $$f(z)\varphi = \varphi$$

     *has a **nonzero solution** in $\mathcal{H}$*

- **Corollary 1.26** *(**The Fredholm Alternative**) [Reed and Simon, 1980]*
  *If A is a **compact** operator on $\mathcal{H}$, then **either** $(I - A)^{-1}$ exists **or** $\varphi = \varphi$ has a solution.*

- **Theorem 1.27** *(**Riesz-Schauder Theorem**) [Reed and Simon, 1980]*
  *Let A be a **compact** operator on $\mathcal{H}$, then $\sigma(A)$ is a **discrete set** having **no limit points except** perhaps $\lambda = 0$.*

  *Further, any **nonzero** $\lambda \in \sigma(A)$ is an **eigenvalue** of **finite multiplicity** (i.e. the corresponding space of eigenvectors is **finite dimensional**).*

- **Remark** (***Compact** Operator has only **Nonzero Point Spectrum** with **Finite Dimensional Eigenspace***)

Riesz-Schauder Theorem states that *the **spectrum** for **compact** operator on **Hilbert** space* consists of *only* the point spectrum besides $\lambda = 0$.

Moreover, *the **eigenspace*** corresponding to *each **nonzero eigenvalue*** is *finite dimensional.*

- **Theorem 1.28** ***(The Hilbert-Schmidt Theorem)*** *[Reed and Simon, 1980]*
  *Let A be a **self-adjoint compact operator** on $\mathcal{H}$. Then, there is a **complete orthonormal basis**, $\{\phi_n\}_{n=1}^{\infty}$, for $\mathcal{H}$ so that*

  $$A\phi_n = \lambda_n \phi_n$$

  and $\lambda_n \to 0$ as $n \to \infty$.

- **Remark** (***Eigendecomposition*** *of Hilbert Space based on* ***Self-Adjoint Compact Operator***)
  In other word, given a self-adjoint compact operator A on $\mathcal{H}$, *the HIlbert space $\mathcal{H}$ is the direct sum of eigenspaces of A.*

  $$\mathcal{H} = \bigoplus_{\lambda_n \in \sigma(A) \subset \mathbb{R}} \text{Ker}\,(\lambda_n I - A)$$

  A ***self-adjoint compact operator*** on $\mathcal{H}$ is the closest counterpart of ***Hermitian matrix / Symmetric Real matrix*** in infinite dimensional space.

- **Theorem 1.29** ***(Canonical Form for Compact Operators)*** *[Reed and Simon, 1980]*
  *Let A be a **compact** operator on $\mathcal{H}$. Then there exist **(not necessarily complete) orthonormal sets** $\{\psi_n\}_{n=1}^{N}$ and $\{\phi_n\}_{n=1}^{N}$ and **positive real numbers** $\{\lambda_n\}_{n=1}^{N}$ with $\lambda_n \to 0$ so that*

  $$A = \sum_{n=1}^{N} \lambda_n \left\langle \psi_n \,,\, \cdot \right\rangle \phi_n \tag{4}$$

  *The sum in (4), which may be finite or infinite, **converges in norm**. The numbers, $\{\lambda_n\}_{n=1}^{N}$, are called the **singular values of** A.*

1.8 Trace Class and Hilbert-Schmidt Operators

- **Definition** (***Trace of Positive Semi-Definite Operator***)
  Let $\mathcal{H}$ be a ***separable Hilbert space***, $\{\phi_n\}_{n=1}^{\infty}$ an ***orthonormal basis*** Then for any ***positive semi-definite*** operator $A \in \mathcal{L}(\mathcal{H})$, we define

  $$\text{tr}\,(A) = \sum_{n=1}^{\infty} \left\langle A\phi_n \,,\, \phi_n \right\rangle$$

  The number $\text{tr}\,(A)$ is called ***the trace of*** A.

- **Remark** (***Trace of General Linear Operator***)
  Let $A \in \mathcal{L}(\mathcal{H})$ be a bounded linear operator on separable Hilbert space. Instead of considering *the trace of A*, we consider *the trace of absolute value of A*,

  $$\text{tr}\,(|A|) = \text{tr}\left(\sqrt{A^{*}A}\right).$$

- **Definition** (***Hilbert-Schmidt Operator***)
  An operator $T \in \mathcal{L}(\mathcal{H})$ is called ***Hilbert-Schmidt*** if and only if
  $$\mathrm{tr}\,(T^*T) < \infty.$$
  The family of all Hilbert-Schmidt operators is denoted by $\mathcal{B}_2(\mathcal{H})$ or $\mathcal{B}_{HS}(\mathcal{H})$.

- **Proposition 1.30** *(**Space of Hilbert-Schmidt Operator**) [Reed and Simon, 1980]*
  1. *The space of all Hilbert-Schmidt operators $\mathcal{B}_2(\mathcal{H})$ is a **$*$-ideal** in $\mathcal{L}(\mathcal{H})$,*
  2. *(**Inner Product**): If $A, B \in \mathcal{B}_2(\mathcal{H})$, then for **any orthonormal basis** $\{\varphi_n\}_{n=1}^{\infty}$,*
     $$\sum_{n=1}^{\infty} \langle A^* B\varphi_n \, , \, \varphi_n\rangle$$
     *is **absolutely summable**, and its **limit**, denoted by $\langle A \, , \, B\rangle_{HS}$, is **independent** of the orthonormal basis chosen, i.e.*
     $$\langle A \, , \, B\rangle_{HS} = \mathrm{tr}\,(A^* B)$$
  3. *$\mathcal{B}_2(\mathcal{H})$ with inner product $\langle \cdot \, , \, \cdot\rangle_{HS}$ is a **Hilbert space**.*
  4. *(**Norm**): Let $\|\cdot\|_2$ be defined in $\mathcal{B}_2(\mathcal{H})$ by*
     $$\|A\|_2 := \sqrt{\langle A \, , \, A\rangle_{HS}} = \sqrt{\mathrm{tr}\,(A^* A)}.$$
     Then
     $$\|A\| \le \|A\|_2 \le \|A\|_1 \, , \quad \textit{and} \quad \|A\|_2 = \|A^*\|_2$$
  5. *(**Compactness**) Every $A \in \mathcal{B}_2(\mathcal{H})$ is **compact** and a **compact operator**, A, is in $\mathcal{B}_2(\mathcal{H})$ **if and only if***
     $$\sum_{n=1}^{\infty} \lambda_n^2 < \infty$$
     *where $\{\lambda_n\}$ are the **singular values** of A.*
  6. *(**Finite Rank Approximation**) The **finite rank operators** are $\|\cdot\|_2$-**dense** in $\mathcal{B}_2(\mathcal{H})$.*
  7. *$A \in \mathcal{B}_2(\mathcal{H})$ **if and only if***
     $$\{\|A\varphi_n\|\}_{n=1}^{\infty} \in \ell^2$$
     *for **some** orthonormal basis $\{\varphi_n\}_{n=1}^{\infty}$.*
  8. *$A \in \mathcal{B}_1(\mathcal{H})$ if and only if $A = BC$ with $B, C \in \mathcal{B}_2(\mathcal{H})$.*
  9. *$\mathcal{B}_2(\mathcal{H})$ is not $\|\cdot\|$-closed in $\mathcal{L}(\mathcal{H})$.*

- **Theorem 1.31** *(**Hilbert-Schmidt Operator of** L^2 **Space**) [Reed and Simon, 1980]*
*Let (M, μ) be a **measure space** and $\mathcal{H} = L^2(M, \mu)$. Then $T \in \mathcal{L}(\mathcal{H})$ is **Hilbert-Schmidt if and only if** there is a function*

$$K \in L^2(M \times M, \mu \otimes \mu)$$

with

$$(Tf)(x) = \int_M K(x, y) f(y) d\mu(y),$$

Moreover,

$$\|T\|_2^2 = \int_{M \times M} |K(x, y)|^2 \, d\mu(x) d\mu(y).$$

- **Definition** (***Kernel of Integral Operator***)
Consider the simple operator T_K, defined in $\mathcal{C}[0, 1]$ by

$$(T_K f)(x) = \int_0^1 K(x, y) f(y) dy,$$

where the function $K(x, y)$ is *continuous* on the square $0 \le x, y \le 1$. T_K is called an ***integral kernel operator*** and $K(x, y)$ is called *the **kernel** of the integral operator* T_K.

- **Remark** (***Properties of Integral Kernel Operator***)
We summary some important property of the integral kernel operator T_K:

  1. T_K is ***compact operator*** on $\mathcal{C}[0, 1]$.
  2. For $K^*(x, y) := \overline{K(y, x)}$,

  $$(T_K)^* = T_{K^*}$$

  3. Let B_M denote the functions f in $\mathcal{C}[0, 1]$ such that $\|f\|_\infty \le M$, i.e. closed $\|\|_\infty$-ball in $\mathcal{C}[0, 1]$

  $$B_M := \{f \in \mathcal{C}[0, 1] : \|f\|_\infty \le M\}$$

  The set of functions $T_K(B_M) := \{T_K f : f \in B_M\}$ is ***equicontinuous***.
  4. The *operator norm* of T_K is *bounded above* by the L^2 *norm* of kernel function K

  $$\|T_K\| \le \|K\|_{L^2}$$

  5. The eigenfunctions of T_K, $\{\varphi_n\}_{n=1}^\infty$, forms a complete orthonormal basis in $L^2(M, \mu)$. Then

  $$K(x, y) = \sum_{n=1}^{\infty} \lambda_n \varphi_n(x) \overline{\varphi_n(y)}$$

  where λ_n is the eigenvalue corresponding to eigenfunction φ_n.

- **Theorem 1.32** ***(Mercer's Theorem)***
  Suppose Ω is a ***compact domain*** *and T is a* ***positive Hilbert-Schmidt operator*** *on $L^2(\Omega)$. If the integral kernel $K(\cdot, \cdot)$ is* ***continuous*** *on $\Omega \times \Omega$, then the* ***eigenfunction*** *φ_k is* ***continuous*** *on Ω if $\lambda_k > 0$, and the expansion*

$$K(x, y) = \sum_{n=1}^{\infty} \lambda_n \varphi_n(x) \overline{\varphi_n(y)}$$

  converges ***uniformly*** *on* ***compact*** *sets.*

2 Reproducing Kernel Hilbert Space (RKHS)

2.1 Definitions

- **Definition** (***Evaluation Functional***)
  Let X be a space, $\mathcal{H}$ be the *Hilbert space* of complex-valued functions on X, a linear functional $\delta_x : \mathcal{H} \to \mathbb{C}$ is called an ***evalution functional*** if

$$\delta_x(f) = f(x), \quad \forall f \in \mathcal{H}$$

  That is, δ_x evaluates each function $f \in \mathcal{H}$ at a point x.

- **Definition** (***Reproducing Kernel Hilbert Space***)
  A *Hilbert space* $\mathcal{H}$ is a ***reproducing kernel Hilbert space*** (*RKHS*) if, ***for all*** *x* ***in*** *X, the evaluation functional δ_x* is ***bounded linear operator*** on $\mathcal{H}$, i.e. there exists some $M_x > 0$ such that

$$|\delta_x(f)| := |f(x)| \le M_x \|f\|_{\mathcal{H}} \qquad \forall f \in \mathcal{H}.$$

  Equivalently, for every $x \in X$, δ_x is ***continuous*** *at* ***every*** *f in $\mathcal{H}$*

- **Remark**

$$\mathcal{H} \text{ is a RKHS } \Leftrightarrow \delta_x \in \mathcal{H}^*, \forall x \in X$$

- **Remark** (***Unique Representation of Evaluation Functional at Each Point***)
  If $\mathcal{H}$ is a *reproducing kernel Hilbert space*, $\delta_x \in \mathcal{H}^*$, then *the Riesz Representation theorem* implies that for all $x \in X$, there exists ***a unique function*** $k_x \in \mathcal{H}$ so that

$$\delta_x = \langle \cdot \, , \, k_x \rangle$$
$$\Rightarrow f(x) = \delta_x(f) = \langle f \, , \, k_x \rangle \, , \forall f \in \mathcal{H}$$

  Note that $k_x : X \to \mathbb{C}$ is a complex-valued function on X, so

$$k_x(y) := \delta_y(k_x) = \langle k_x \, , \, k_y \rangle := K(x, y)$$

  where the complex-valued function $K : X \times X \to \mathbb{C}$ is called a ***reproducing kernel***

- **Definition** (***Reproducing Kernel***)
  Let $\mathcal{H}$ be a class of functions on X forming a Hilbert space (complex in the latter, but possibly real). A function $K : X \times X \to \mathbb{C}$ is called a ***reproducing kernel*** (*r.k.*) of $\mathcal{H}$, if

1. For every $x \in X$, the kernel $K(x, \cdot)$ as a function belongs to $\mathcal{H}$; i.e., $K(x, \cdot) := k_x \in \mathcal{H}$;
2. The ***reproducing property***: for every $x \in X$ and every $f \in \mathcal{H}$,

$$f(x) = \delta_x(f) = \langle f\,,\, k_x \rangle_{\mathcal{H}} = \langle f\,,\, K(x, \cdot)\rangle_{\mathcal{H}} \tag{5}$$

- **Remark** (***Reproducing Kernel via Inner Product in RKHS***)
We can define ***the reproducing kernel*** $K : X \times X \to \mathbb{C}$ using ***the inner product***

$$K(x, y) = \langle k_x\,,\, k_y \rangle_{\mathcal{H}}\,, \quad \forall x, y \in X$$

where $k_x, k_y \in \mathcal{H}$ correspond to evaluation functionals δ_x and δ_y in RKHS $\mathcal{H}$, respesctively.

Equivalently, we can the following equation:

$$K(x, y) = \langle K(x, \cdot)\,,\, K(y, \cdot)\rangle_{\mathcal{H}}$$

- **Remark** The following properties hold for reproducing kernels:

1. (***Existence***). The existence of reproducing kernel K is based on the definition of RKHS $\mathcal{H}$ that $\delta_x \in \mathcal{H}^*$ for all $x \in X$. Then by *the Riesz representation theorem (Riesz Lemma)*, we can find a unique k_x corresponding to δ_x so that $K(x, y) := \delta_y(k_x) = \langle k_x\,,\, k_y \rangle$.
2. (***Uniqueness***) If a reproducing kernel $K(x, y)$ *exists*, it is ***unique***. This is due to *the Riesz representation theorem (Riesz Lemma)*.
3. (***Positive Semi-Definite***) $K(x, y)$ is ***positive semidefinite*** in X; i.e.,

$$\sum_{i,j=1}^{n} K(x_i, x_j)\xi_i \bar{\xi}_j \geq 0$$

for all $x_1 \ldots, x_n \in X$ and all $\xi_1, \ldots, \xi_n \in \mathbb{C}$. It follows that

$$\begin{aligned}
\sum_{i,j=1}^{n} K(x_i, x_j)\xi_i \bar{\xi}_j &= \sum_{i,j=1}^{n} \left\langle k_{x_i}\,,\, k_{x_j} \right\rangle_{\mathcal{H}} \xi_i \bar{\xi}_j \\
&= \left\langle \sum_{i=1}^{n} \xi_i k_{x_i}\,,\, \sum_{j=1}^{n} \xi_j k_{x_j} \right\rangle_{\mathcal{H}} \\
&= \left\| \sum_{i=1}^{n} \xi_i k_{x_i} \right\|_{\mathcal{H}}^2 \geq 0 \quad \blacksquare
\end{aligned}$$

4. (***Hermitian***): $K(x, y)$ is *Hermitian* i.e.

$$K(x, y) = \overline{K(y, x)}$$

This is due to the Hermitian property of inner product.

5. (***Cauchy-Schwartz Inequality***)

$$|K(x, y)|^2 \leq K(x, x)^{1/2} K(y, y)^{1/2}.$$

2.2 Properties

- **Proposition 2.1** ***(Closed Subspace)***
  A ***closed linear subspaces*** *$\mathcal{F}$ of reproducing kernel Hilbert space $\mathcal{H}$ is a reproducing kernel Hilbert space with the reproducing kernel $K_{\mathcal{F}} = K|_{\mathcal{F}}$.*

- **Proposition 2.2** ***(Orthorgonal Complements)***
  If $\mathcal{H}'$ and $\mathcal{H}''$ are **complementary** *subspaces of $\mathcal{H}$, i.e. $\mathcal{H} = \mathcal{H}' \oplus \mathcal{H}''$, then their reproducing kernels satisfy the equation $K' + K'' = K$.*

- **Remark** (***Projection*** *via* ***Reproducing Kernel***)
  If the class $\mathcal{F}$ with the reproducing kernel K is a *subspace* of a larger Hilbert space $\mathcal{H}$, then the formula
  $$f(x) = \langle h\, ,\, K(x, \cdot)\rangle_{\mathcal{H}}\, ,$$
  gives the projection f of $h \in \mathcal{H}$ in $\mathcal{F}$.

- **Proposition 2.3** *If K is the reproducing kernel of the class F with the norm $\|\cdot\|$, and if the linear class $F_1 \subset F$ forms a Hilbert space with the norm $\|\cdot\|_1$ such that $\|f_1\|_1 \geq \|f_1\|$ for every $f_1 \in F_1$, then the class F_1 possesses a reproducing kernel K_1 satisfying $K_1 \preceq K$; i.e., $K - K_1$ is positive definite.*

2.3 Convergence Properties

- **Remark** Recall different convergence:

  1. **Definition** (***Pointwise Convergence***). [Kreyszig, 1989]
     A sequence (f_n) in a normed space $\mathcal{H}$ is said to be ***pointwise convergent*** (or ***convergent in pointwise topology***) if there is an $f \in \mathcal{H}$ such that for every $x \in X$
     $$\lim_{n\to\infty} f_n(x) = f(x).$$

  2. **Definition** (***Strong Convergence***). [Kreyszig, 1989]
     A sequence (f_n) in a normed space $\mathcal{H}$ is said to be ***strongly convergent*** (or ***convergent in the norm***) if there is an $f \in \mathcal{H}$ such that
     $$\lim_{n\to\infty} \|f_n - f\| = 0.$$
     This is written $\lim_{n\to\infty} f_n = f$ or simply $f_n \to f$ is called *the* ***strong limit*** *of* (f_n), and we say that (f_n) *converges* ***strongly*** *to* f.

  3. **Definition** (***Weak Convergence***). [Kreyszig, 1989]
     A sequence (f_n) in a normed space $\mathcal{H}$ is said to be ***weakly convergent*** if there is an $f \in \mathcal{H}$ such that *for* ***every*** $I \in \mathcal{H}^*$,
     $$\lim_{n\to\infty} I(f_n) = I(f).$$
     This is written $f_n \xrightarrow{w} f$ or $f_n \rightharpoonup f$. The element f is called ***the weak limit*** *of* (f_n), and we say that (f_n) ***converges weakly*** *to* f.

- **Proposition 2.4** *(**Convergence in Norm** leads to **Pointwise Convergence**)*
  *If the class $\mathcal{H}$ possesses a reproducing kernel $K(x, y)$, every sequence of functions $\{f_n\}$ which converges **strongly** to a function f in the Hilbert space $\mathcal{H}$, converges also **at every point** in the ordinary sense, i.e.*

$$\|f_n - f\|_{\mathcal{H}} \to 0 \;\Rightarrow\; f_n(x) \to f(x), \quad \text{for each } x \in X$$

  *This convergence becomes **uniform** in every subset of E in which $K(x, y)$ is **uniformly bounded**.*

  Proof: This follows from

$$\begin{aligned}|f(x) - f_n(x)| &= |\langle f - f_n \,,\, K(x, \cdot)\rangle_{\mathcal{H}}| \\ &\le \|f - f_n\| \; \|K(x, \cdot)\| = \|f - f_n\| \; K(x, x)^{1/2}.\end{aligned} \tag{6}$$

  Thus $\|f - f_n\| \to 0$ leads to $|f(x) - f_n(x)| \to 0$ for every $x \in X$.

  If $\{f_n\}$ converges ***weakly*** to f; i.e., $\langle f_n \,,\, K(x, \cdot)\rangle \to \langle f \,,\, K(x, \cdot)\rangle$ for every $x \in X$, we have again $f_n(x) \to f(x)$ for every x. That is, in RKHS,

$$\text{strong convergence } \Rightarrow \text{weak convergence } \Rightarrow \text{ pointwise convergence}$$

  there exists non-increasing nested sets $E_1 \subset E_2 \ldots$ in which f_n ***uniformly*** *converges* to f. Let $E = \lim_{n\to\infty} E_n = \bigcup_n E_n$ Moreover, if $x \mapsto K(x, \cdot)$ is a transformation that is *continuous* from X to a subset of $\mathcal{H}$, then in every ***compact*** $E_1 \subset E$, f_n converges ***uniformly*** to f and it transforms to a ***compact subset*** of $\mathcal{H}$.

  To see that, for every $\epsilon > 0$, $\exists(x_1, \ldots, x_n) \subset E_1$ such that for every $x \in E_1$, there exists at least one x_k such that $\|K(x, \cdot) - K(x_k, \cdot)\| \le \epsilon/4 \;\|f\| \le \epsilon/4\, M$ for $M = \sup_{\boldsymbol{x}\in E} \|f(x)\|$. Further if we choose n_0, so that $n > n_0$, $|f(x_k) - f_n(x_k)| \le \epsilon/4$, then for the selected $x \in E_1$, the following holds

$$\begin{aligned}|f(x) - f(x_k)| &\le |f(x_k) - f_n(x_k)| + |\langle f(x) - f_n(x) \,,\, K(x, y) - K(x_k, y)\rangle| \\ &\le \frac{\epsilon}{4} + \|f - f_n\| \; \|K(x, \cdot) - K(x_k, \cdot)\| \\ &\le \frac{\epsilon}{4} + 2\, M \frac{\epsilon}{4M} < \epsilon.\end{aligned}$$

  The *continuity* of the correspondence $x \mapsto K(x, \cdot)$ is equivalent to ***equicontinuity*** of all functions of $\mathcal{H}$ with $\|f(x)\| < M$ for any $M > 0$. ∎

- **Remark** In reproducing kernel Hilbert space,

$$\text{strong (norm) convergence } \Rightarrow \text{weak convergence } \Rightarrow \text{ pointwise convergence}$$

2.4 Construction from Hermitian Positive Definite Kernel

- **Definition** Let X be a nonempty set. A *Hermitian form* $K : X \times X \to \mathbb{C}$ is called a ***positive-definite (p.d.) kernel*** on X if

$$\sum_{i=1}^{n}\sum_{j=1}^{n} c_i \bar{c}_j K(x_i, x_j) \ge 0$$

  holds for any $x_1, \ldots, x_n \in X$, given $n \in \mathbb{N}$, $c_1, \ldots, c_n \in \mathbb{C}$.

- In this section, we show that a RKHS can be constructed from any positive definite kernels:

Theorem 2.5 ***(RKHS from Positive Definite Kernel) (Moore-Aronszajn)***
Suppose K is a ***symmetric, positive definite kernel*** *on a set X. Then there is a* ***unique*** *Hilbert space of functions on X for which K is a* ***reproducing kernel****.*

Proof: For all $x \in X$, define $K_x := K(x, \cdot)$. Let $\mathcal{H}_0$ be the linear span of $\{K_x : x \in X\}$, that is, it is the space of functions of the form

$$\sum_{k=1}^{n} \xi_k K_{x_k}$$

where $x_1, x_2, \ldots x_n \in X$ and $\xi_1, \xi_2, \ldots \xi_n \in \mathbb{C}$. Define an inner product on $\mathcal{H}_0$ by

$$\left\langle \sum_{k=1}^{n} \xi_k K_{x_k} \,,\, \sum_{j=1}^{m} \eta_j K_{y_j} \right\rangle_{\mathcal{H}_0} := \sum_{i=1}^{n} \sum_{j=1}^{m} K(x_i, y_j) \xi_i \overline{\eta}_j .$$

which implies $K(x, y) = \langle K_x \,,\, K_y \rangle_{\mathcal{H}_0}$. It is an inner product due to symmetric and positive definite property of kernel K.

Let $\mathcal{H}$ be the completion of $\mathcal{H}_0$ with respect to this inner product. Then $\mathcal{H}$ consists of functions of the form

$$f := \sum_{k=1}^{\infty} \xi_k K_{x_k}$$

where

$$\lim_{n \to \infty} \sup_{p \geq 0} \| \sum_{i=n}^{n+p} \xi_i K_{x_i} \|^2_{\mathcal{H}_0} = 0$$

Now we can check the reproducing property

$$\langle f \,,\, K_x \rangle_{\mathcal{H}} = \left\langle \sum_{k=1}^{\infty} \xi_k K_{x_k} \,,\, K_x \right\rangle_{\mathcal{H}} = \sum_{k=1}^{\infty} \xi_k \langle K_{x_k} \,,\, K_x \rangle_{\mathcal{H}} = \sum_{k=1}^{\infty} \xi_k K(x_k, x) = f(x)$$

To prove ***uniqueness***, let $\mathcal{G}$ be *another Hilbert space* of functions for which K is a reproducing kernel. For every x and y in X, the reproducing property implies that

$$\langle K_x, K_y \rangle_{\mathcal{H}} = K(x, y) = \langle K_x, K_y \rangle_{\mathcal{G}}.$$

By linearity, $\langle \cdot, \cdot \rangle_{\mathcal{H}} = \langle \cdot, \cdot \rangle_{\mathcal{G}}$ on the span of $\{K_x : x \in X\}$. Then $\mathcal{H} \subseteq \mathcal{G}$ because $\mathcal{G}$ is complete and contains $\mathcal{H}_0$ and hence contains its completion.

Now we need to prove that every element of $\mathcal{G}$ is in $\mathcal{H}$. Let f be an element of $\mathcal{G}$. Since $\mathcal{H}$ is a *closed subspace* of $\mathcal{G}$, we can write $f = f_{\mathcal{H}} + f_{\mathcal{H}^\perp}$ where $f_{\mathcal{H}} \in \mathcal{H}$ and $f_{\mathcal{H}^\perp} \in \mathcal{H}^\perp$. Now if $x \in X$ then, since K is a reproducing kernel of $\mathcal{G}$ and $\mathcal{H}$:

$$\begin{aligned} f(x) = \langle K_x \,,\, f \rangle_{\mathcal{G}} &= \langle K_x \,,\, f_{\mathcal{H}} \rangle_{\mathcal{G}} + \langle K_x \,,\, f_{\mathcal{H}^\perp} \rangle_{\mathcal{G}} \\ &= \langle K_x \,,\, f_{\mathcal{H}} \rangle_{\mathcal{G}} \\ &= \langle K_x \,,\, f_{\mathcal{H}} \rangle_{\mathcal{H}} = f_{\mathcal{H}}(x) \end{aligned}$$

where we have used the fact that K_x belongs to $\mathcal{H}$ so that its inner product with $f_{\mathcal{H}^\perp}$ in $\mathcal{G}$ is zero. This shows that $f = f_H$ in $\mathcal{G}$. ∎

2.5 Construction from Integral Kernel Operator on Compact Space

- **Remark** (***Integral Operator***)
  Let X be a ***compact*** space equipped with a *strictly positive* ***finite*** *Borel measure* μ and $K : X \times X \to \mathbb{R}$ a ***continuous***, ***symmetric***, and ***positive definite function***. We can define a linear operator T_K on $L^2(X, \mu)$ by

$$(T_K f)(x) := \int_X K(x, y) f(y) d\mu(y),$$

  i.e. T_K is a ***integral kernel operator*** on $L^2(X, \mu)$.

- **Remark** (***RKHS from Integral Kernel Operator***)
  We see that

  1. $T_K \in \mathcal{B}_2(L^2(X, \mu))$ is a *Hilbert-Schmidt operator*, thus
  2. T_K is a ***compact operator***.
  3. T_K is a ***self-adjoint, positive semi-definite*** operator on $L^2(X, \mu)$ since K is a symmetric and positive definite kernel.
  4. By *Hilbert-Schmidt theorem*, since T_K is *self-adjoint and compact*, the Hilbert space $L^2(X, \mu)$ has a ***complete orthonormal basis*** $\{\varphi_n\}_{n=1}^{\infty}$ where each φ_n is the ***eigen-function*** of T_K corresponding to ***eigenvalue*** $\lambda_n \ge 0$ with $\lambda_n \to 0$.
  5. T_K maps ***continuously*** into the space of *continuous functions* $\mathcal{C}(X)$.
  6. By *Mercer's Theorem*, there exists an orthonormal basis $\{\varphi_n\}_{n=1}^{\infty}$ on $L^2(X, \mu)$ where each φ_n is a ***continuous*** *eigenfunction* of T_K corresponding to the ***eigenvalue*** $\lambda_n \ge 0$ so that the kernel K has an expansion

$$K(x, y) = \sum_{n=1}^{\infty} \lambda_n \varphi_n(x) \overline{\varphi_n(y)}$$

     that converges ***uniformly*** on ***compact*** set X. This above series representation is referred to as a ***Mercer kernel*** or ***Mercer representation*** of K. Thus any function f in $L^2(X, \mu)$ can be represented as

$$f(x) = \sum_{n=1}^{\infty} \alpha_n \varphi_n(x).$$

  7. Finally, a ***reproducing kernel Hilbert space*** $\mathcal{H} \subseteq L^2(X, \mu)$ based on spectral decomposition of T_K is given by

$$\mathcal{H} = \left\{ f \in L^2(X, \mu) : \sum_{n=1}^{\infty} \frac{|\langle f \,,\, \varphi_n \rangle_{L^2}|^2}{\lambda_n} < \infty \right\}$$

     where the inner product of $\mathcal{H}$ given by

$$\langle f \,,\, g \rangle_{\mathcal{H}} = \sum_{n=1}^{\infty} \frac{\langle f \,,\, \varphi_n \rangle_{L^2} \; \langle g \,,\, \varphi_n \rangle_{L^2}}{\lambda_n}.$$

     The kernel K is the reproducing kernel of $\mathcal{H}$.

2.6 Construction from Feature Map

- **Definition** (***Feature Map***) [Scholkopf and Smola, 2001]
  A ***feature map*** is a map $\Phi : X \to \mathcal{F}$, where $\mathcal{F}$ is a *Hilbert space* such that the image of X under Φ, $\mathcal{H} := \Phi(X) \subseteq \mathcal{F}$ is a ***reproducing kernel Hilbert space*** with kernel function

$$K(x, y) := \langle \Phi(x)\,,\, \Phi(y) \rangle_{\mathcal{F}}\,.$$

- **Remark** (***Feature Map via Kernel Function***)
  We can think of Φ as a vector-valued function with possibly *infinite-dimensional* output. Moreover, given kernel function K, let $K_x := K(x, \cdot) \in \mathcal{H}$, we can define the feature map as

$$\Phi : x \to K_x = K(x, \cdot)$$

- **Remark** (***Feature Map via Eigenfunction of Integral Operator***) [Scholkopf and Smola, 2001, Rasmussen and Williams, 2005]
  Any *symmetric positive definite kernel* K induces a ***integral kernel operator*** T_K that is *self-adjoint* and *compact*. T_K has discrete real spectrum $\sigma(T_K) \subset \mathbb{R}$ with eigenfunctions $\{\varphi_n\}$ that spans the entire space $\mathcal{F}$.

  Use the *Mercer's theorem*. Given the kernel function $K : X \times X \to \mathbb{C}$, the *eigenfunction* $\varphi_n : X \to \mathbb{C}$ associated with the *eigenvalue* $\lambda_n \geq 0$ is defined by the integral equation

$$\lambda_n \varphi_n(x) = \int_X K(x, y)\varphi_n(y) d\mu(y).$$
$$\text{where } K(x, y) = \sum_{n=1}^{\infty} \lambda_n \varphi_n(x)\overline{\varphi_n(y)}$$

  And we can define the feature map Φ via

$$\Phi : x \mapsto \left(\sqrt{\lambda_n}\varphi_n(x)\right)_{n=1}^{\infty}.$$

  Note that the output dimension of Φ is determined by the Mercer representation of K. It can be finite dimensional if the kernel K is simple. In this way, we have

$$K(x, y) := \langle \Phi(x)\,,\, \Phi(y) \rangle_{\mathcal{F}}\,.$$

- **Remark** (***Equivalence of Two Representations***)
  The kernel map and the Mercer's feature map are equivalent in that there exists an ***isometric isomorphism*** between them so that the inner product is preserved. In specific, $\Phi : x \mapsto K(x, \cdot)$ maps a feature to a function in $\mathcal{F}$ and the Mercer's kernel $\Phi : X \mapsto (\sqrt{\lambda_n}\varphi_n(x))_{j=1}^{\infty}$ maps a feature vector to a ***vector representation*** of $K(x, \cdot)$ under a set of orthonormal basis $\left\{\sqrt{\lambda_n}\varphi_n(\cdot)\right\}_{n=1}^{\infty} \subset \mathcal{F}$.

  Note, however, $\{K(x, \cdot)\}_{n \in S}$ for a set of features $\{x_n\}_{n \in S}$ are not orthonormal. $\{K(x, \cdot)\}_{n \in S} \not\subset \left\{\sqrt{\lambda_n}\varphi_n(\cdot)\right\}_{n=1}^{\infty}$.

3 Equivalent Definition of Reproducing Kernel Hilbert Space

We summarize four different ways to construct a reproducing kernel Hilbert space (RKHS):

1. (***Bounded Evaluation Functional***)
   A RKHS $\mathcal{H}$ is a *Hilbert space* of functions on X such that ***the evalution functional*** $\delta_x \in \mathcal{H}^*$ is ***bounded linear functional*** for all $x \in X$.

   - This implies that
     $$f(x) := \delta_x(f) = \langle f \, , \, K_x \rangle$$
     for some unique $K_x \in \mathcal{H}$ for each $x \in X$;
   - Define *the reproducing kernel* as function $K : X \times X \to \mathbb{C}$ such that
     $$K(x, y) = \langle K_x \, , \, K_y \rangle = K_x(y).$$
     Thus $K(x, y)$ satisfies the reproducing property:
     $$f(x) = \langle f \, , \, K(x, \cdot) \rangle$$

2. (***Hermitian Positive Definite Kernel***)
   Given a ***Hermitian positive definite kernel***, $K : X \times X \to \mathbb{C}$, there exists a ***unique RKHS*** $\mathcal{H}$ that admits K as its *reproducing kernel*.

   - From the subspace $\mathcal{H}_0 = \operatorname{span}\{K_x : x \in X\}$ where $K_x := K(x, \cdot)$:
     $$f \in \mathcal{H}_0 \Rightarrow f = \sum_{k=1}^{n} \xi_k K_{x_k}, \;\; \exists n \in \mathbb{N}, \{x_i\}_{i=1}^{n} \subset X, \; \{\xi_i\} \subset \mathbb{C}$$
   - Define the inner product on $\mathcal{H}_0$ as
     $$\left\langle \sum_{k=1}^{n} \xi_k K_{x_k} \, , \, \sum_{j=1}^{m} \eta_j K_{y_j} \right\rangle_{\mathcal{H}_0} := \sum_{i=1}^{n} \sum_{j=1}^{m} K(x_i, y_j) \xi_i \overline{\eta}_j.$$
     Due to Hermitian and positive definite property of K, the inner product above is well-defined.
   - $K(x, y) = \langle K_x \, , \, K_y \rangle_{\mathcal{H}_0}$ by definition. The reproducing property holds as well.
   - Construct the RKHS $\mathcal{H}$ by the ***completion*** of $\mathcal{H}_0$.

3. (***Integral Kernel Operator***)
   Consider a measure space (X, μ) where X is a ***compact*** space and μ is a *Borel measure*. Given $K : X \times X \to \mathbb{C}$ as a ***continuous Hermitian positive definite kernel*** on X, we can define a ***integral kernel operator*** T_K on $L^2(X, \mu)$ by
   $$(T_K f)(x) := \int_X K(x, y) f(y) d\mu(y).$$
   - T_K is a ***self-adjoint, positive*** and ***compact operator*** on *separable Hilbert space*.

- The ***spectrum*** of T_K is ***discrete*** and is of ***real nonnegative value*** $\lambda_n \geq 0$ such that $\lambda_n \to 0$.
- There exists a ***complete orthonormal basis*** in $L^2(X, \mu)$ that are ***eigenfunctions*** $\{\varphi_n(x)\}$ of T_K.
- There exists a ***orthonormal basis*** formed by continuous eigenfunctions $\{\varphi_n(x)\}$ and their eigenvalues $\{\lambda_n\}$ so that the expansion

$$K(x, y) = \sum_{n=1}^{\infty} \lambda_n \varphi_n(x) \overline{\varphi_n(y)}$$

  converges *uniformly* on compact set X.
- The RKHS $\mathcal{H} \subseteq L^2(X, \mu)$ based on spectral decomposition of T_K is given by

$$\mathcal{H} = \left\{ f \in L^2(X, \mu) : \sum_{n=1}^{\infty} \frac{|\langle f \, , \, \varphi_n \rangle_{L^2}|^2}{\lambda_n} < \infty \right\}$$

  where the inner product of $\mathcal{H}$ given by

$$\langle f \, , \, g \rangle_{\mathcal{H}} = \sum_{n=1}^{\infty} \frac{\langle f \, , \, \varphi_n \rangle_{L^2} \, \langle g \, , \, \varphi_n \rangle_{L^2}}{\lambda_n}.$$

  The kernel K is the reproducing kernel of $\mathcal{H}$.

4. (***Feature Map***)
Define ***feature map*** $\Phi : X \to \mathcal{F}$ from X to a Hilbert space $\mathcal{F}$ so that $\mathcal{H} := \Phi(X)$ is a RKHS with the reproducing kernel

$$K(x, y) := \langle \Phi(x) \, , \, \Phi(y) \rangle_{\mathcal{F}} \, , \quad \forall x, y \in X$$

- We can define

$$\Phi : x \to K_x = K(x, \cdot)$$

- We can also define

$$\Phi : x \mapsto \left(\sqrt{\lambda_n} \varphi_n(x) \right)_{n=1}^{\infty}$$

  where the eigenfunctions $\{\varphi_n(x)\}$ and their eigenvalues $\{\lambda_n\}$ form expansion of kernel K

$$K(x, y) = \sum_{n=1}^{\infty} \lambda_n \varphi_n(x) \overline{\varphi_n(y)}$$

- These two definitions are equivalent based on Mercer's theorem.

4 Reproducing Kernel Hilbert Space in Machine Learning

4.1 Empirical Feature Map

- **Definition** (***Empirical Feature Map***) [Scholkopf and Smola, 2001]
  Given a set of samples $S := (z_1, \ldots, z_m) \subset X$, the ***empirical feature map*** $\Phi_m : X \to \mathbb{R}^m$ is the empirical estimate of the feature map $\Phi : x \mapsto K(x, \cdot) \in \mathcal{H}$ under S. That is

  $$\Phi_m : x \mapsto K(x, \cdot)|_{(z_1, \ldots, z_m)} \equiv (K(x, z_n))_{n=1}^m.$$

- **Remark** Note that the image of empirical feature map $\Phi_m(X) \subset \mathbb{R}^m$ does *not necessarily* form a *closed linear subspace*. Also the inner product defined in the linear span of $\{\Phi_m(\boldsymbol{z}_i), 1 \le i \le m\}$ is ***not canonical***, since $\Phi_m(\boldsymbol{x}_i)$ are ***not orthogonal*** in $\mathbb{R}^m$ in general.

- **Remark** (***Induced Inner Product on*** $\mathbb{R}^m$ *from* ***Empirical Feature Map***)
  The empirical feature map that is associated with kernel K should be defined by inducing an ***inner product*** of $\mathbb{R}^m$ into $\Phi_m(X)$ as

  $$\langle \Phi_m(x)\,,\, \Phi_m(y) \rangle_m = K(x, y),$$

  where $\langle \cdot\,,\, \cdot \rangle_m \equiv \langle M \cdot\,,\, \cdot \rangle_{\mathbb{R}^m}$ for ***positive definite matrix*** M. Enforcing $x, y \in S := (z_1, \ldots, z_m)$ be in training set, we can obtain the equation

  $$\boldsymbol{K} = \boldsymbol{K}\,\boldsymbol{M}\,\boldsymbol{K},$$
  $$\Rightarrow \boldsymbol{M} = \boldsymbol{K}^{\dagger} = \boldsymbol{K}^{-1}.$$

  where $\boldsymbol{K} = [K(z_i, z_j)]_{i,j=1}^m \in \mathbb{R}^{m \times m}$ is ***the matrix representation*** of T_K in $\mathbb{R}^m$.

- **Remark** (***Explict Form of Empirical Feature Map***) [Scholkopf and Smola, 2001]
  Therefore, we could define ***empirical feature map*** *that is* ***associated with kernel*** K as

  $$\Phi_m : x \mapsto \boldsymbol{K}^{-\frac{1}{2}} (K(x, z_n))_{n=1}^m.$$

  The above is equivalent to the ***Kernel PCA whitening***.

- **Remark** (***Empirical Feature Map as Finite Dimensional Approximation***)
  This Φ_m maps X to a m***-dimensional space*** $\mathbb{R}^m$ as opposed to the original Φ that maps to $\mathcal{H}$, *a* ***Hilbert space of functions*** with *high or infinite dimensionality.* Moreover, the induced inner product on $\mathbb{R}^m$ has representation

  $$\langle \Phi_m(x)\,,\, \Phi_m(y) \rangle_m \equiv \boldsymbol{k}_x^T \boldsymbol{K}^{-1} \boldsymbol{k}_y$$

  where $\boldsymbol{K} = [K(z_i, z_j)]_{i,j}^m$, and $\boldsymbol{k}_x = ((K(x, z_i))_{i=1}^m)^T$.

4.2 Representer Theorem

- **Definition** (***Loss Function***)
  Denote by $(x, y, f(x)) \in \mathcal{X} \times \mathcal{Y} \times \mathcal{Y}$ the triplet consisting of a ***pattern*** x, an ***observation*** y and a ***prediction*** $f(x)$. Then the map

  $$c : \mathcal{X} \times \mathcal{Y} \times \mathcal{Y} \to [0, \infty)$$

  with the property $c(x, y, y) = 0$ for all $x \in \mathcal{X}$ and $y \in \mathcal{Y}$ will be called ***a loss function***.

- **Definition** (***Expected Risk***)
Let $((\mathcal{X},\mathcal{Y}),\mathscr{F},\mathcal{P})$ be a probability space on domain $(\mathcal{X},\mathcal{Y})$ and $f:\mathcal{X}\to\mathcal{Y}$ be a measurable function on $\mathcal{X}$. ***The expected risk*** of f *with respect to* $\mathcal{P}$ *and* c is defined as

$$\mathcal{R}(f) = \mathbb{E}_{\mathcal{P}}\left[c(x,y,f(x))\right] = \int_{\mathcal{X}\times\mathcal{Y}} c(x,y,f(x))\, d\mathcal{P}(x,y)$$

- **Definition** (***Empirical Risk***)
Since $\mathcal{P}$ is unknown, given a set of samples $\mathcal{D} := \{(x_n,y_n)\}_{n=1}^{m} \subset \mathcal{X}\times\mathcal{Y}$, we replace $\mathcal{P}$ by ***the empirical probability measure***

$$\widehat{\mathcal{P}}_m = \frac{1}{m}\sum_{n=1}^{m}\delta_{(x_n,y_n)}.$$

Then we define ***the empirical risk*** of f *with respect to* $\widehat{\mathcal{P}}_m$ *and* c as

$$\mathcal{R}_{emp}(f) = \mathbb{E}_{\widehat{\mathcal{P}}_m}\left[c(x,y,f(x))\right] = \frac{1}{m}\sum_{n=1}^{m} c(x_n,y_n,f(x_n))$$

- **Remark** We assume the *empirical risk functional* $\mathcal{R}_{emp}(f)$ is ***continuous*** with respect to f.

- **Remark** (***Regularization***)
The key idea in ***regularization*** is to restrict *the class of possible minimizers* $\mathcal{F}$ (with $f\in\mathcal{F}$) of *the empirical risk functional* $\mathcal{R}_{emp}(f)$ such that $\mathcal{F}$ becomes a ***compact set***.

We do not directly specify a compact set $\mathcal{F}$, since this leads to a *constrained optimization problem*, which can be cumbersome in practice. Instead, we add a ***stabilization (regularization) term*** $\Omega(f)$ to the original objective function; the latter could be $\mathcal{R}_{emp}(f)$, for instance. This, too, leads to ***better conditioning*** of the problem. We consider *the following class of* ***regularized risk functionals***:

$$\mathcal{R}_{reg}(f) := \mathcal{R}_{emp}(f) + \lambda\Omega(f)$$

Here $\lambda > 0$ is the so-called ***regularization parameter*** which specifies the ***tradeoff*** between minimization of $\mathcal{R}_{emp}(f)$ and *the* ***smoothness*** *or* ***simplicity*** which is enforced by small $\Omega(f)$. Usually one chooses $\Omega(f)$ to be ***convex***, since this ensures that there exists *only one global minimum*, provided $\mathcal{R}_{emp}(f)$ is also *convex*

- **Definition** (***Regularized Risk in Reproducing Kernel Hilbert Space***)
Suppose that $f\in\mathcal{H}$ where $\mathcal{H}$ is a ***reproducing kernel Hilbert space*** on X. $\mathcal{R}_{emp}(f)$ is *the empirical risk functional*. The ***regularized risk functionals*** on $\mathcal{H}$ is defined as

$$\mathcal{R}_{reg}(f) := \mathcal{R}_{emp}(f) + \frac{\lambda}{2}\left\|f\right\|_{\mathcal{H}}^2$$

- **Lemma 4.1** ***(Operator Inversion Lemma)*** *[Scholkopf and Smola, 2001]*
Let X *be a* ***compact*** *set and let the map* $f: X\to Y$ *be* ***continuous***. *Then there exists an* ***inverse*** *map* $f^{-1}: f(X)\to X$ *that is also* ***continuous***.

- **Theorem 4.2** *(**Representer Theorem**) [Scholkopf and Smola, 2001]*
  *Let $\mathcal{X}$ be a set, and $c : (\mathcal{X} \times \mathbb{R} \times \mathbb{R})^m \to \mathbb{R} \cup \{\infty\}$ be an arbitrary loss function, $\mathcal{H}$ be the reproducing kernel Hilbert space associated with kernel K on X. Denote $\Omega : [0, \infty) \to \mathbb{R}$ as a **strictly monotonic increasing** function. Then each minimizer $f \in \mathcal{H}$ of the regularized risk*

  $$c\left(\left(x_1, y_1, f(x_1)\right), \ldots, \left(x_m, y_m, f(x_m)\right)\right) + \Omega\left(\|f\|_{\mathcal{H}}\right) \tag{7}$$

  *admits a **representation** of the form*

  $$f(x) = \sum_{n=1}^{m} \alpha_n K(x_n, x).$$

  Proof: For convenience we will assume that we are dealing with $\bar{\Omega}(\|f\|_{\mathcal{H}}^2) := \Omega(\|f\|_{\mathcal{H}})$ rather than $\Omega(\|f\|_{\mathcal{H}})$. This is no restriction at all, since *the quadratic function is strictly monotonic* on $[0, \infty)$, and therefore $\bar{\Omega}$ is strictly monotonic on $[0, \infty)$ if and only if Ω also satisfies this requirement.

  We may decompose any $f \in \mathcal{H}$ into a part contained $\mathcal{H}_0 = \operatorname{span}\{K(x_i, \cdot), i = 1, \ldots, m\}$ and one in the ***orthogonal complement*** $\mathcal{H}_0^{\perp}$;

  $$f(x) = f_{\mathcal{H}_0}(x) + f_{\mathcal{H}_0^{\perp}}(x) = \sum_{n=1}^{m} \alpha_n K(x_n, x) + f_{\mathcal{H}_0^{\perp}}(x)$$

  Here $\alpha_n \in \mathbb{R}$ and $f_{\mathcal{H}_0^{\perp}} \in \mathcal{H}$ with $\langle f_{\mathcal{H}_0^{\perp}}, K(x_n, \cdot)\rangle_{\mathcal{H}} = 0$ for all $n \in [m] := \{1, \ldots, m\}$. By reproducing property of K we may write $f(x_i)$ (for all $i \in [m]$) as

  $$\begin{aligned} f(x_i) &= \langle f\,,\, K(x_i, \cdot)\rangle_{\mathcal{H}} \\ &= \sum_{n=1}^{m} \alpha_n K(x_n, x_i) + \langle f_{\mathcal{H}_0^{\perp}}, K(x_i, \cdot)\rangle_{\mathcal{H}} \\ &= \sum_{n=1}^{m} \alpha_n K(x_n, x_i) \end{aligned}$$

  Second, for all $f_{\mathcal{H}_0^{\perp}}$, by *Pythagorean theorem* and the *monotonicity* of Ω,

  $$\Omega(\|f\|_{\mathcal{H}}) := \bar{\Omega}\left(\left\|\sum_{n=1}^{m} \alpha_n K(x_n, \cdot)\right\|_{\mathcal{H}}^2 + \left\|f_{\mathcal{H}_0^{\perp}}\right\|_{\mathcal{H}}^2\right) \geq \bar{\Omega}\left(\left\|\sum_{n=1}^{m} \alpha_n K(x_n, \cdot)\right\|_{\mathcal{H}}^2\right)$$

  Thus for any fixed $\alpha_n \in \mathbb{R}$ the risk functional (7) is minimized for $f_{\mathcal{H}_0^{\perp}} = 0$. Since this also has to hold for the solution, the theorem holds. $\blacksquare$

- **Remark** (***Monotonicity of Regularizer Functional*** $\Omega(\cdot)$ ***is Required***)
  Monotonicity of Ω is ***necessary*** to ensure that the theorem holds. It does not prevent the ***regularized risk functional*** from having ***multiple local minima***. To ensure a single minimum, we would need to require ***convexity***. If we discard the ***strictness of the monotonicity***, then it no longer follows that *each minimizer of the regularized risk admits an expansion*; it still follows, however, that *there is always another solution that is **as good***, and that does *admit the expansion*.

- **Remark** (***Function Space Minimizer Lies in Finite Dimensional Subspace***)
  The ***significance*** of *the Representer Theorem* is that although we might be trying to solve *an* ***optimization problem in an infinite-dimensional space*** $\mathcal{H}$, containing *linear combinations of kernels centered* on ***arbitrary*** points of X, it states that the solution lies ***in the span of*** m ***particular kernels*** – *those centered on the training points.*

  In the *Support Vector* community,

$$f(x) = \sum_{n=1}^{m} \alpha_n K(x_n, x)$$

  is called ***the Support Vector expansion***. For suitable choices of loss functions, it has empirically been found that many of the α_n often equal 0.

5 Example and Computation

- K as an operator is self-adjoint, i.e.

$$\langle f\,,\, Kg \rangle = \langle Kf\,,\, g \rangle\,.$$

- The inner product in Reproducing Kernel Hilbert Space (RKHS) $\mathcal{H}$ is given by [Ramm, 1998]:

$$\langle f\,,\, g \rangle_{\mathcal{H}} \equiv \left\langle K^{-1} f\,,\, g \right\rangle_{\boldsymbol{x}} \tag{8}$$

  where K^{-1} is inverse to the linear operator $K : \mathcal{H} \to \mathcal{H}$ given by

$$(Kf)(\boldsymbol{z}) = \int_E K(\boldsymbol{z}, \boldsymbol{x}) f(\boldsymbol{x}) d\boldsymbol{x}.$$

  Note that the reproducing property holds

$$\begin{aligned} \langle f\,,\, K(\cdot, \boldsymbol{y}) \rangle_{\mathcal{H}} &= \left\langle K^{-1} f\,,\, K(\cdot, \boldsymbol{y}) \right\rangle_{\boldsymbol{x}} = \left\langle f\,,\, K^{-1} K(\cdot, \boldsymbol{y}) \right\rangle_{\boldsymbol{x}} = \langle f\,,\, \delta_{\boldsymbol{y}} \rangle_{\boldsymbol{x}} \\ &= f(\boldsymbol{y}). \end{aligned}$$

- The *Gaussian kernel*

$$K(\boldsymbol{x}, \boldsymbol{x}') = \exp\left(-\frac{\|\boldsymbol{x} - \boldsymbol{x}'\|_2^2}{2\lambda} \right), \lambda > 0 \tag{9}$$

  The Gaussian kernel has universally bounded norm $|K(\boldsymbol{x}, \boldsymbol{x})|^{1/2} = \|\Phi(\boldsymbol{x})\| = 1$. Moreover, $K(\boldsymbol{x}, \boldsymbol{x}') > 0$ for $\boldsymbol{x} \neq \boldsymbol{x}'$; i.e., all points lies in the same orthant

$$\cos(\angle \boldsymbol{x}, \boldsymbol{x}') = \left\langle \Phi(\boldsymbol{x})\,,\, \Phi(\boldsymbol{x}') \right\rangle = K(\boldsymbol{x}, \boldsymbol{x}') > 0.$$

  This indicates that in the Gaussian case, the mapped data lie in a fairly restricted area of feature space. *However*, in another sense, they occupy a space which is as large as possible: given distinct points $(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_m) \subset E$, $\{\Phi(\boldsymbol{x}_1), \ldots, \Phi(\boldsymbol{x}_m)\}$ are linearly independent and $[K_{i,j}] = [K(\boldsymbol{x}_i, \boldsymbol{x}_j)]$ has full rank.

$\{\Phi(\boldsymbol{x}_1), \ldots, \Phi(\boldsymbol{x}_m)\}$ span an m-dimensional subspace of F. Therefore a Gaussian kernel defined on a domain of infinite cardinality, with no a priori restriction on the number of training examples, produces a feature space of *infinite*-dimension.

The eigenfunction of Gaussian kernel can be found using Fourier transformation; i.e., $K(\boldsymbol{x}, \boldsymbol{x}') = f(\|\boldsymbol{x} - \boldsymbol{x}'\|) \equiv G(\boldsymbol{x} - \boldsymbol{x}')$

$$\begin{aligned} \lambda\psi(\boldsymbol{x}) &= \int G(\boldsymbol{x} - \boldsymbol{x}')\psi(\boldsymbol{x}')d\mu(\boldsymbol{x}') = G \otimes \psi \\ \Rightarrow \lambda\mathcal{F}\{\psi\}(\boldsymbol{s}) &= \mathcal{F}\{G\}\,\mathcal{F}\{\psi\} = G(\boldsymbol{s})\,\mathcal{F}\{\psi\}(\boldsymbol{s}) \\ \psi &\in \mathcal{F}^{-1}\{N(\lambda I - G(\boldsymbol{s}))\}, \\ &\text{where } N(\lambda I - G(\boldsymbol{s})) = \{F(\boldsymbol{s}) : (\lambda I - G(\boldsymbol{s}))\,F(\boldsymbol{s}) = 0\} \end{aligned}$$

Note that $\mathcal{F}\{G\}$ of Gaussian is also Gaussian which is rescaled in mass and variance from the original one by some constants. Since the spheres centered at 0 are the sets on which the multiplier equality $\lambda = G(\boldsymbol{s})$ can hold, $\psi \equiv 0$ for $\boldsymbol{s} \in$ the complementary of a sphere centered at 0.

Thus, the eigenfunctions will be inverse Fourier transforms of *tempered distributions* [Grafakos, 2008] supported in spheres centered at the origin. There are a lot of them, for example, the most familiar ones are the *Bessel functions*, which correspond to uniform surface measure on a nondegenerate sphere.

- The *homogeneous polynomial kernel*

$$K(\boldsymbol{x}, \boldsymbol{x}') = \left\langle \boldsymbol{x}\,,\, \boldsymbol{x}' \right\rangle^d, d > 0 \tag{10}$$

and *inhomogeneous polynomial kernel*

$$K(\boldsymbol{x}, \boldsymbol{x}') = \left(\left\langle \boldsymbol{x}\,,\, \boldsymbol{x}' \right\rangle + c\right)^d, d > 0 \tag{11}$$

- The B_n*-spline kernel*

$$K(\boldsymbol{x}, \boldsymbol{x}') = B_{2p+1}\left(\left\|\boldsymbol{x} - \boldsymbol{x}'\right\|\right), p > 0 \tag{12}$$

where $B_n = \otimes_{i=1}^{n} I[-\frac{1}{2}, \frac{1}{2}]$ and $f \otimes g = \int f(t)g(\tau - t)dt$.

- All the kernel above (w/o inhomogeneous one) is invariant under the unitary transformation $\boldsymbol{U}$, i.e.

$$\begin{aligned} K(\boldsymbol{U}\boldsymbol{x}, \boldsymbol{U}\boldsymbol{x}') &= \left\langle \Phi(\boldsymbol{U}\boldsymbol{x})\,,\, \Phi(\boldsymbol{U}\boldsymbol{x}') \right\rangle \\ &= \left\langle \Phi(\boldsymbol{x})\,,\, \Phi(\boldsymbol{x}') \right\rangle = K(\boldsymbol{x}, \boldsymbol{x}') \end{aligned}$$

- The *Radial basis function* (*RBF*) kernels are kernels that can be written in the form

$$K(\boldsymbol{x}, \boldsymbol{x}') = f\left(d(\boldsymbol{x}, \boldsymbol{x}')\right) \tag{13}$$

for $d(\boldsymbol{x}, \boldsymbol{x}')$ is the metric on E.

The RBF kernels are *unitary invariant*, too. In addition, they are *translation invariant*.

By Bochner's theorem, if a kernel K can be written in terms of $\|x-y\|$, i.e. $K(x,y) = f(\|x-y\|)$ for some f, then K is a kernel iff the Fourier transform of f is non-negative.

$$K(\boldsymbol{x},\boldsymbol{x}') = \int_{\mathbb{R}^D} S(\boldsymbol{s}) \exp\left(-i\,\boldsymbol{s}^T(\boldsymbol{x}-\boldsymbol{x}')\right) d\boldsymbol{s}$$

In terms of this, for RBF kernel, the eigenfunctions can be obtained by Fourier analysis; in particular, it could be Bessel functions etc.

The RBF kernel is sometimes called a convolutional kernel, with the feature map

$$\begin{aligned}\Phi_{\boldsymbol{u}} &: E \mapsto L^2 \\ \boldsymbol{x} &\mapsto \frac{1}{(2\pi)^{n/2}} \mathcal{F}^{-1}(\sqrt{S})(\boldsymbol{x}-\boldsymbol{u})\end{aligned}$$

So that

$$K(\boldsymbol{x},\boldsymbol{x}') = \int_{\mathbb{R}^D} \Phi_{\boldsymbol{u}}(\boldsymbol{x})\Phi_{\boldsymbol{u}}(\boldsymbol{x}')d\boldsymbol{u}$$

For example, for Gaussian kernel

$$\exp\left(-\frac{\|\boldsymbol{x}-\boldsymbol{x}'\|^2}{\sigma^2}\right) = \left(\frac{4}{\sigma^2\pi}\right)^{n/2} \int_{\mathbb{R}^D} \exp\left(-\frac{2\,\|\boldsymbol{x}-\boldsymbol{u}\|^2}{\sigma^2}\right) \exp\left(-\frac{2\,\|\boldsymbol{x}'-\boldsymbol{u}\|^2}{\sigma^2}\right) d\boldsymbol{u}$$

with the convolutional feature map

$$\Phi_{\boldsymbol{u}}(\boldsymbol{x}) = \left(\frac{2}{\sigma\sqrt{\pi}}\right)^{n/2} \exp\left(-\frac{2\,\|\boldsymbol{x}-\boldsymbol{u}\|^2}{\sigma^2}\right).$$

References

Nachman Aronszajn. Theory of reproducing kernels. *Transactions of the American mathematical society*, pages 337–404, 1950.

John B Conway. *A course in functional analysis*, volume 96. Springer, 2019.

Gerald B Folland. *Real analysis: modern techniques and their applications.* John Wiley & Sons, 2013.

Loukas Grafakos. *Classical fourier analysis*, volume 2. Springer, 2008.

Thomas Hofmann, Bernhard Schölkopf, and Alexander J Smola. Kernel methods in machine learning. *The annals of statistics*, pages 1171–1220, 2008.

Erwin Kreyszig. *Introductory functional analysis with applications*, volume 81. wiley New York, 1989.

AG Ramm. On the theory of reproducing kernel hilbert spaces. *Journal of Inverse and Ill posed problems*, 6:515–520, 1998.

Carl Edward Rasmussen and Christopher K. I. Williams. *Gaussian Processes for Machine Learning (Adaptive Computation and Machine Learning).* The MIT Press, 2005. ISBN 026218253X.

Michael Reed and Barry Simon. *Methods of modern mathematical physics: Functional analysis*, volume 1. Gulf Professional Publishing, 1980.

Bernhard Scholkopf and Alexander J Smola. *Learning with kernels: support vector machines, regularization, optimization, and beyond.* MIT press, 2001.