Kernel Method: Data Analysis with Positive Definite Kernels

2. Positive Definite Kernel and Reproducing Kernel Hilbert Space

Kenji Fukumizu

The Institute of Statistical Mathematics.
Graduate University of Advanced Studies /
Tokyo Institute of Technology

Nov. 17-26, 2010
Intensive Course at Tokyo Institute of Technology

Outline

Positive definite kernel

- Definition and examples of positive definite kernel

Reproducing kernel Hilbert space

- RKHS and positive definite kernel

Some basic properties

- Properties of positive definite kernels
- Properties of RKHS

Positive definite kernel

Definition and examples of positive definite kernel

Reproducing kernel Hilbert space

RKHS and positive definite kernel

Some basic properties

Properties of positive definite kernels

Properties of RKHS

Definition of Positive Definite Kernel

Definition. Let $\mathcal{X}$ be a set. $k : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ is a positive definite kernel if $k(x, y) = k(y, x)$ and for every $x_1, \ldots, x_n \in \mathcal{X}$ and $c_1, \ldots, c_n \in \mathbb{R}$

$$\sum_{i,j=1}^{n} c_i c_j k(x_i, x_j) \geq 0,$$

i.e. the symmetric matrix

$$(k(x_i, x_j))_{i,j=1}^{n} = \begin{pmatrix} k(x_1, x_1) & \cdots & k(x_1, x_n) \\ \vdots & \ddots & \vdots \\ k(x_n, x_1) & \cdots & k(x_n, x_) \end{pmatrix}$$

is positive semidefinite.

- The symmetric matrix $(k(x_i, x_j))_{i,j=1}^{n}$ is often called a Gram matrix.

Definition: Complex-valued Case

Definition. Let $\mathcal{X}$ be a set. $k : \mathcal{X} \times \mathcal{X} \to \mathbb{C}$ is a positive definite kernel if for every $x_1, \ldots, x_n \in \mathcal{X}$ and $c_1, \ldots, c_n \in \mathbb{C}$

$$\sum_{i,j=1}^{n} c_i \overline{c_j} k(x_i, x_j) \geq 0.$$

Remark. The Hermitian property $k(y, x) = \overline{k(x, y)}$ is derived from the positive-definiteness. [Exercise]

If $k(x, y)$ is positive definite, so is $\overline{k(x, y)} = k(y, x)$.

Some Basic Properties

Facts. Assume $k : \mathcal{X} \times \mathcal{X} \to \mathbb{C}$ is positive definite. Then, for any x, y in $\mathcal{X}$,

1. $k(x,x) \geq 0.$
2. $|k(x,y)|^2 \leq k(x,x)k(y,y)$.

Proof. (1) is obvious. For (2), with the fact $k(y,x) = \overline{k(x,y)}$, the definition of positive definiteness implies that the eigenvalues of the hermitian matrix

$$\begin{pmatrix} k(x,x) & \overline{k(x,y)} \\ k(x,y) & k(y,y) \end{pmatrix}$$

is non-negative, thus, its determinant $k(x,x)k(y,y) - |k(x,y)|^2$ is non-negative. $\square$

Examples

Real valued positive definite kernels on $\mathbb{R}^n$:

- Linear kernel[1]

$$k_0(x, y) = x^T y$$

- Exponential

$$k_E(x, y) = \exp(\beta x^T y) \qquad (\beta > 0)$$

- Gaussian RBF (radial basis function) kernel

$$k_G(x, y) = \exp\Big(-\frac{1}{2\sigma^2}\|x - y\|^2\Big) \qquad (\sigma > 0)$$

- Laplacian kernel

$$k_L(x, y) = \exp\Big(-\alpha \textstyle\sum_{i=1}^n |x_i - y_i|\Big) \qquad (\alpha > 0)$$

- Polynomial kernel

$$k_P(x, y) = (x^T y + c)^d \qquad (c \geq 0, d \in \mathbb{N})$$

[1] [Exercise] prove that the linear kernel is positive definite.

Feature Map must be Positive Definite

Proposition 1

Let V be an vector space with an inner product $\langle \cdot, \cdot \rangle$. If we have a map

$$\Phi : \mathcal{X} \to V, \qquad x \mapsto \Phi(x),$$

the kernel on $\mathcal{X}$ defined by

$$k(x, y) = \langle \Phi(x), \Phi(y) \rangle$$

is positive definite.

Proof. Let $x_1, \ldots, x_n$ in $\mathcal{X}$ and $c_1, \ldots, c_n \in \mathbb{C}$.

$$\begin{aligned}
\textstyle\sum_{i,j=1}^n c_i \overline{c_j} k(x_i, x_j) &= \textstyle\sum_{i,j=1}^n c_i \overline{c_j} \langle \Phi(x_i), \Phi(x_j) \rangle \\
&= \left\langle \textstyle\sum_{i=1}^n c_i \Phi(x_i), \textstyle\sum_{j=1}^n c_j \Phi(x_j) \right\rangle \\
&= \left\| \textstyle\sum_{i=1}^n c_i \Phi(x_i) \right\|^2 \quad \geq 0.
\end{aligned}$$

Positive definite kernel

Definition and examples of positive definite kernel

Reproducing kernel Hilbert space

RKHS and positive definite kernel

Some basic properties

Properties of positive definite kernels

Properties of RKHS

Reproducing kernel Hilbert space

Definition. Let $\mathcal{X}$ be a set. A reproducing kernel Hilbert space (RKHS) (over $\mathcal{X}$) is a Hilbert space $\mathcal{H}$ consisting of functions on $\mathcal{X}$ such that for each $x \in \mathcal{X}$ there is a function $k_x \in \mathcal{H}$ with the property

$$\langle f, k_x \rangle_{\mathcal{H}} = f(x) \qquad (\forall f \in \mathcal{H}) \qquad \text{(reproducing property)}.$$

$k(\cdot, x) := k_x(\cdot)$ is called a reproducing kernel of $\mathcal{H}$.

Fact 1. A reproducing kernel is Hermitian (or symmetric).

Proof.
$$k(y, x) = \langle k(\cdot, x), k_y \rangle = \langle k_x, k_y \rangle = \overline{\langle k_y, k_x \rangle} = \overline{\langle k(\cdot, y), k_x \rangle} = \overline{k(x, y)}.$$

Fact 2. The reproducing kernel is unique, if exists. [Exercise]

Positive Definite Kernel and RKHS I

Proposition 2 (RKHS $\Rightarrow$ positive definite kernel)

The reproducing kernel of a RKHS is positive definite.

Proof. Special case of Proposition 1, because with $\Phi(x) = k(\cdot, x)$

$$\overline{k(x,y)} = k(y,x) = \langle k(\cdot,x), k(\cdot,y)\rangle = \langle \Phi(x), \Phi(y)\rangle.$$

$\square$

Positive Definite Kernel and RKHS II

Theorem 3 (positive definite kernel $\Rightarrow$ RKHS. Moore-Aronszajn)

Let $k : \mathcal{X} \times \mathcal{X} \to \mathbb{C}$ (or $\mathbb{R}$) be a positive definite kernel on a set $\mathcal{X}$. Then, there uniquely exists a RKHS $\mathcal{H}_k$ on $\mathcal{X}$ such that

1. $k(\cdot, x) \in \mathcal{H}_k$ *for every* $x \in \mathcal{X}$,
2. $\mathrm{Span}\{k(\cdot, x) \mid x \in \mathcal{X}\}$ *is dense in* $\mathcal{H}_k$,
3. k *is the reproducing kernel on* $\mathcal{H}_k$, i.e.,
$$\langle f, k(\cdot, x)_{\mathcal{H}} \rangle = f(x) \qquad (\forall x \in \mathcal{X}, \forall f \in \mathcal{H}_k).$$

Proof omitted.

Positive Definite Kernel and RKHS III

One-to-one correspondence between positive definite kernels and RKHS.

$$k \qquad \longleftrightarrow \qquad \mathcal{H}_k$$

- Proposition 2: RKHS $\mapsto$ positive definite kernel k.
- Theorem 3: $k \mapsto \mathcal{H}_k$ (injective).

RKHS as Feature Space

If we define

$$\Phi : \mathcal{X} \to \mathcal{H}_k, \quad x \mapsto k(\cdot, x),$$

then,

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

RKHS associated with a positive definite kernel k gives a desired feature space!!

In kernel methods, the above feature map and feature space are always used.

Examples of RKHS

1) Linear kernel on $\mathbb{R}^m$: $k_0(x, y) = x^T y$.

- RKHS for the linear kernel:
  $\mathcal{H} = \{f : \mathbb{R}^m \to \mathbb{R} \mid f(u) = k_0(u, a) = a^T u \ (a \in \mathbb{R}^m)\}$.
- Inner product of $\mathcal{H}$:

$$\langle a^T u, b^T u \rangle = a^T b \qquad (\langle k_0(\cdot, a), k_0(\cdot, b) \rangle = k_0(a, b))$$

- $\mathcal{H}$ is isomorphic to $\mathbb{R}^m$:

$$\mathcal{H} \cong \mathbb{R}^m, \qquad a^T u \mapsto a.$$

- Feature map $\Phi(x) = k_0(u, x) = x^T u$ is simply the above isomorphism.
- Nothing is gained from the viewpoint of data analysis.

2) Positive definite kernel on finite set $\mathcal{X} = \{1, 2, \ldots, m\}$.

- Positive definite kernel $k(i, j)$ corresponds one-to-one to a $m \times m$ positive semidefinite matrix K.
- Eigendecomposition:

$$K = U \Lambda U^T, \qquad U = (u_1, \ldots, u_m), \Lambda = \mathrm{diag}(\lambda_1, \ldots, \lambda_m).$$

- Assume K is (strictly) positive definite, *i.e.*, $\lambda_i > 0$.
- RKHS of k:
  $\mathcal{H} = \{f : \{1, \ldots, m\} \to \mathbb{R}\} = \{(f(1), \ldots, f(m))^T \in \mathbb{R}^m\}$.
- Inner product of $\mathcal{H}$: for $f = \sum_{i=1}^m a_i u_i, g = \sum_{i=1}^m b_i u_i$,

$$\langle f, g \rangle_{\mathcal{H}} = f^T K^{-1} g = \sum_{i=1}^m \frac{a_i b_i}{\lambda_i}.$$

- Reproducing property:
  Let $K = (k_1, \dots, k_m)$ ***i.e.*** $k_i = k(\cdot, i)$.
  For $f = \sum_{j=1}^m a_j k_j \in \mathcal{H}$,

$$\langle f, k(\cdot, i)\rangle = \sum_{j=1}^m a_j k_j^T K^{-1} k_i = \sum_{j=1}^m a_j k_j^T e_i = \sum_{j=1}^m a_j k(j, i) = f(i).$$

3) Gaussian kernel

- RKHS for Gaussian kernel is infinite dimensional.
- Discussed in the next week.

Positive definite kernel

Definition and examples of positive definite kernel

Reproducing kernel Hilbert space

RKHS and positive definite kernel

Some basic properties

Properties of positive definite kernels

Properties of RKHS

Operations that Preserve Positive Definiteness I

Proposition 4

If $k_i : \mathcal{X} \times \mathcal{X} \to \mathbb{C}$ $(i = 1, 2, \ldots)$ are positive definite kernels, then so are the following:

1. *(positive combination)* $ak_1 + bk_2 \quad (a, b \geq 0)$.
2. *(product)* $k_1 k_2 \quad (k_1(x, y)k_2(x, y))$.
3. *(limit)* $lim_{i\to\infty} k_i(x, y)$, *assuming the limit exists.*

Remark. From Proposition 4, the set of all positive definite kernels is a closed (w.r.t. pointwise convergence) convex cone stable under multiplication.

Operations that Preserve Positive Definiteness II

Proof.

(1): Obvious.

(3): The non-negativity in the definition holds also for the limit.

(2): It suffices to show that two Hermitian matrices A and B are positive semidefinite, so is their component-wise product. This is done by the following lemma. □

Definition. For two matrices A and B of the same size, the matrix C with $C_{ij} = A_{ij}B_{ij}$ is called the **Hadamard product** of A and B.

The Hadamard product of A and B is denoted by $A \odot B$.

Lemma 5

Let A and B be non-negative Hermitian matrices of the same size. Then, $A \odot B$ is also non-negative.

Operations that Preserve Positive Definiteness III

Proof.

Let

$$A = U \Lambda U^*$$

be the eigendecomposition of A, where

$U = (u^1, \dots, u^p)$: a unitary matrix, *i.e.*, $\quad U^* = \overline{U}^T$

Λ: diagonal matrix with non-negative entries $(\lambda_1, \dots, \lambda_p)$.

Then, for arbitrary $c_1, \dots, c_p \in \mathbb{C}$,

$$\sum_{i,j=1} c_i \bar{c}_j (A \odot B)_{ij} = \sum_{a=1}^{p} \lambda_a c_i \bar{c}_j u_i^a \bar{u}_j^a B_{ij} = \sum_{a=1}^{p} \lambda_a \xi^{aT} B \overline{\xi^a},$$

where $\xi^a = (c_1 u_1^a, \dots, c_p u_p^a)^T \in \mathbb{C}^p$.

Since $\xi^{aT} B \overline{\xi^a}$ and λ_a are non-negative for each a, so is the sum. $\square$

Modification

Proposition 6

Let $k : \mathcal{X} \times \mathcal{X} \to \mathbb{C}$ *be a positive definite kernel and* $f : \mathcal{X} \to \mathbb{C}$ *be an arbitrary function. Then,*

$$\tilde{k}(x, y) = f(x)k(x, y)\overline{f(y)}$$

is positive definite. In particular,

$$f(x)\overline{f(y)}$$

and

$$\frac{k(x, y)}{\sqrt{k(x, x)}\sqrt{k(y, y)}} \qquad (\textit{normalized kernel})$$

are positive definite.

Proof is left as an exercise.

Proofs for Positive Definiteness of Examples

- Linear kernel: Proposition 1
- Exponential:

$$\exp(\beta x^T y) = 1 + \beta x^T y + \frac{\beta^2}{2!}(x^T y)^2 + \frac{\beta^3}{3!}(x^T y)^3 + \cdots$$

  Use Proposition 4.
- Gaussian RBF kernel:

$$\exp\left(-\frac{1}{2\sigma^2}\|x-y\|^2\right) = \exp\left(-\frac{\|x\|^2}{2\sigma^2}\right)\exp\left(\frac{x^T y}{\sigma^2}\right)\exp\left(-\frac{\|y\|^2}{2\sigma^2}\right).$$

  Apply Proposition 6.
- Laplacian kernel: The proof is shown later.
- Polynomial kernel: Just sum and product.

□

Definition by Evaluation Map

Proposition 7

Let $\mathcal{H}$ be a Hilbert space consisting of functions on a set $\mathcal{X}$. Then, $\mathcal{H}$ is a RKHS if and only if the evaluation map

$$e_x : \mathcal{H} \to \mathbb{K}, \qquad e_x(f) = f(x),$$

is a continuous linear functional for each $x \in \mathcal{X}$.

Proof. Assume $\mathcal{H}$ is a RKHS. The boundedness of e_x is obvious from

$$|e_x(f)| = |\langle f, k_x \rangle| \leq \|k_x\| \|f\|.$$

Conversely, assume the evaluation map is continuous. By Riesz lemma, there is $k_x \in \mathcal{H}$ such that

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

which means $\mathcal{H}$ is a RKHS having k_x as a reproducing kernel. $\square$

Continuity

The functions in a RKHS are "nice" functions.

Proposition 8

Let k be a positive definite kernel on a topological space $\mathcal{X}$, and $\mathcal{H}_k$ be the associated RKHS. If $x \mapsto k(x,x)$ is continuous for any x and $\mathrm{Re}[k(y,x)]$ is continuous for every $x, y \in \mathcal{X}$, then all the functions in $\mathcal{H}_k$ are continuous.

Proof. Let f be an arbitrary function in $\mathcal{H}_k$.

$$|f(x) - f(y)| = |\langle f, k(\cdot,x) - k(\cdot,y)\rangle| \le \|f\|\|k(\cdot,x) - k(\cdot,y)\|.$$

The assertion is easy from

$$\|k(\cdot,x) - k(\cdot,y)\|^2 = k(x,x) + k(y,y) - 2\mathrm{Re}[k(x,y)].$$

$\square$

Remark. If $k(x,y)$ is differentiable, then all the functions in $\mathcal{H}_k$ are differentiable.

c.f. L^2 space contains non-continuous functions.

Summary of Sections 1 and 2

- We would like to use a feature map $\Phi : \mathcal{X} \to \mathcal{H}$ to incorporate nonlinearity or high order moments.
- The inner product in the feature space must be computed efficiently. Ideally,

$$\langle \Phi(x), \Phi(y) \rangle = k(x, y).$$

- To satisfy the above relation, the kernel k must be positive definite.
- A positive definite kernel k defines an associated RKHS, where k is the reproducing kernel;

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

- Use a RKHS as a feature space, and $\Phi : x \mapsto k(\cdot, x)$ as the feature map.

Vector space with inner product I

Definition. V: vector space over a field $\mathbb{K} = \mathbb{R}$ or $\mathbb{C}$.
V is called an inner product space if it has an inner product (or scalar product, dot product) $(\cdot, \cdot) : V \times V \to \mathbb{K}$ such that for every $x, y, z \in V$

1. (Strong positivity) $(x, x) \geq 0$, and $(x, x) = 0$ if and only if $x = 0$,
2. (Addition) $(x + y, z) = (x, z) + (y, z)$,
3. (Scalar multiplication) $(\alpha x, y) = \alpha(x, y)$ ($\forall \alpha \in \mathbb{K}$),
4. (Hermitian) $(y, x) = \overline{(x, y)}$.

Vector space with inner product II

$(V, (\cdot, \cdot))$: inner product space.

Norm of $x \in V$:

$$\|x\| = (x, x)^{1/2}.$$

Metric between x and y:

$$d(x, y) = \|x - y\|.$$

Theorem 9 (Cauchy-Schwarz inequality)

$$|(x, y)| \leq \|x\| \|y\|.$$

Remark: Cauchy-Schwarz inequality holds without requiring $\|x\| = 0 \Rightarrow x = 0$.

Hilbert space I

Definition. A vector space with inner product $(\mathcal{H}, (\cdot, \cdot))$ is called Hilbert space if the induced metric is complete, *i.e.* every Cauchy sequence[2] converges to an element in $\mathcal{H}$.

Remark 1:
A Hilbert space may be either finite or infinite dimensional.

Example 1.
$\mathbb{R}^n$ and $\mathbb{C}^n$ are finite dimensional Hilbert space with the ordinary inner product

$$(x, y)_{\mathbb{R}^n} = \sum_{i=1}^n x_i y_i \quad \text{or} \quad (x, y)_{\mathbb{C}^n} = \sum_{i=1}^n x_i \overline{y_i}.$$

[2] A sequence $\{x_n\}_{n=1}^{\infty}$ in a metric space (X, d) is called a Cauchy sequence if $d(x_n, x_m) \to 0$ for $n, m \to \infty$.

Hilbert space II

Example 2. $L^2(\Omega, \mu)$.
Let $(\Omega, \mathcal{B}, \mu)$ is a measure space.

$$\mathcal{L} = \left\{ f : \Omega \to \mathbb{C} \;\middle|\; \int |f|^2 d\mu < \infty \right\}.$$

The inner product on $\mathcal{L}$ is define by

$$(f, g) = \int f\overline{g} d\mu.$$

$L^2(\Omega, \mu)$ is defined by the equivalent classes identifying f and g if their values differ only on a measure-zero set.

- $L^2(\Omega, \mu)$ is complete.
- $L^2(\mathbb{R}^n, dx)$ is infinite dimensional.

Orthogonality

- Orthogonal complement.

  Let $\mathcal{H}$ be a Hilbert space and V be a closed subspace.

  $$V^{\perp} := \{x \in \mathcal{H} \mid (x, y) = 0 \text{ for all } y \in V\}$$

  is a closed subspace, and called the orthogonal complement.

- Orthogonal projection.

  Let $\mathcal{H}$ be a Hilbert space and V be a closed subspace. Every $x \in \mathcal{H}$ can be uniquely decomposed

  $$x = y + z, \qquad y \in V \quad \text{and} \quad z \in V^{\perp},$$

  that is,

  $$\mathcal{H} = V \oplus V^{\perp}.$$

Complete orthonormal system I

- ONS and CONS.

  A subset $\{u_i\}_{i \in I}$ of $\mathcal{H}$ is called an orthonormal system (ONS) if $(u_i, u_j) = \delta_{ij}$ (δ_{ij} is Kronecker's delta).

  A subset $\{u_i\}_{i \in I}$ of $\mathcal{H}$ is called a complete orthonormal system (CONS) if it is ONS and if $(x, u_i) = 0$ ($\forall i \in I$) implies $x = 0$.

  Fact: Any ONS in a Hilbert space can be extended to a CONS.

Complete orthonormal system II

- Separability
  A Hilbert space is separable if it has a countable CONS.

Assumption

In this course, a Hilbert space is always assumed to be separable.

Complete orthonormal system III

Theorem 10 (Fourier series expansion)

Let $\{u_i\}_{i=1}^{\infty}$ *be a CONS of a separable Hilbert space. For each* $x \in \mathcal{H}$*,*

$$x = \textstyle\sum_{i=1}^{\infty}(x, u_i)u_i, \qquad \textit{(Fourier expansion)}$$

$$\|x\|^2 = \textstyle\sum_{i=1}^{\infty}|(x, u_i)|^2. \qquad \textit{(Parseval's equality)}$$

Proof omitted.

Example: CONS of $L^2([0\ 2\pi], dx)$

$$u_n(t) = \tfrac{1}{\sqrt{2\pi}} e^{\sqrt{-1}nt} \qquad (n = 0, 1, 2, \ldots)$$

Then,

$$f(t) = \textstyle\sum_{n=0}^{\infty} a_n u_n(t)$$

is the (ordinary) Fourier expansion of a periodic function.

Bounded operator I

Let $\mathcal{H}_1$ and $\mathcal{H}_2$ be Hilbert spaces. A linear transform $T : \mathcal{H}_1 \to \mathcal{H}_2$ is often called operator.

Definition. A linear operator $\mathcal{H}_1$ and $\mathcal{H}_2$ is called bounded if

$$\sup_{\|x\|_{\mathcal{H}_1}=1} \|Tx\|_{\mathcal{H}_2} < \infty.$$

The operator norm of a bounded operator T is defined by

$$\|T\| = \sup_{\|x\|_{\mathcal{H}_1}=1} \|Tx\|_{\mathcal{H}_2} = \sup_{x\neq 0} \frac{\|Tx\|_{\mathcal{H}_2}}{\|x\|_{\mathcal{H}_1}}.$$

(Corresponds to the largest singular value of a matrix.)

Fact. If $T : \mathcal{H}_1 \to \mathcal{H}_2$ is bounded,

$$\|Tx\|_{\mathcal{H}_2} \leq \|T\|\|x\|_{\mathcal{H}_1}.$$

Bounded operator II

Proposition 11

A linear operator is bounded if and only if it is continuous.

Proof. Assume $T : \mathcal{H}_1 \to \mathcal{H}_2$ is bounded. Then,

$$\|Tx - Tx_0\| \leq \|T\|\|x - x_0\|$$

means continuity of T.

Assume T is continuous. For any $\varepsilon > 0$, there is $\delta > 0$ such that $\|Tx\| < \varepsilon$ for all $x \in \mathcal{H}_1$ with $\|x\| < 2\delta$.
Then,

$$\sup_{\|x\|=1} \|Tx\| = \sup_{\|x\|=\delta} \frac{1}{\delta}\|Tx\| \leq \frac{\varepsilon}{\delta}.$$

$\square$

Riesz lemma I

Definition. A linear functional is a linear transform from $\mathcal{H}$ to $\mathbb{C}$ (or $\mathbb{R}$).

The vector space of all the bounded (continuous) linear functionals called the dual space of $\mathcal{H}$, and is denoted by $\mathcal{H}^*$.

Theorem 12 (Riesz lemma)

For each $\phi \in \mathcal{H}^*$, there is a unique $y_\phi \in \mathcal{H}$ such that

$$\phi(x) = (x, y_\phi) \qquad (\forall x \in \mathcal{H}).$$

Proof.

Consider the case of $\mathbb{R}$ for simplicity.

$\Leftarrow$) Obvious by Cauchy-Schwartz.

Riesz lemma II

$\Rightarrow$) If $\phi(x) = 0$ for all x, take $y = 0$. Otherwise, let

$$V = \{x \in \mathcal{H} \mid \phi(x) = 0\}.$$

Since ϕ is a bounded linear functional, V is a closed subspace, and $V \neq \mathcal{H}$.
Take $z \in V^{\perp}$ with $\|z\| = 1$. By orthogonal decomposition, for any $x \in \mathcal{H}$,

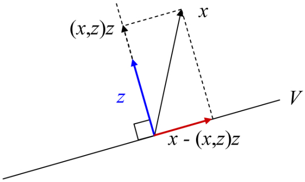
$$x - (x, z)z \in V.$$

Apply ϕ, then

$$\phi(x) - (x, z)\phi(z) = 0, \qquad \text{i.e.,} \quad \phi(x) = (x, \phi(z)z).$$

Take $y_\phi = \phi(z)z$. $\square$

Riesz lemma III



Proof of Theorem 3 I

Proof. (Described in $\mathbb{R}$ case.)

- Construction of an inner product space:

$$H_0 := \mathrm{Span}\{k(\cdot, x) \mid x \in \mathcal{X}\}.$$

Define an inner product on H_0:
for $f = \sum_{i=1}^{n} a_i k(\cdot, x_i)$ and $g = \sum_{j=1}^{m} b_j k(\cdot, y_j)$,

$$\langle f, g \rangle := \sum_{i=1}^{n} \sum_{j=1}^{m} a_i b_j k(x_i, y_j).$$

This is independent of the way of representing f and g from the expression

$$\langle f, g \rangle = \sum_{j=1}^{m} b_j f(y_j) = \sum_{i=1}^{n} a_i g(x_i).$$

Proof of Theorem 3 II

- Reproducing property on H_0:

$$\langle f, k(\cdot, x)\rangle = \sum_{i=1}^{n} a_i k(x_i, x) = f(x).$$

- Well-defined as an inner product:
  It is easy to see $\langle \cdot, \cdot \rangle$ is bilinear form, and

$$\|f\|^2 = \sum_{i,j=1}^{n} a_i a_j k(x_i, x_j) \geq 0$$

  by the positive definiteness of f.
  If $\|f\| = 0$, from Cauchy-Schwarz inequality,[3]

$$|f(x)| = |\langle f, k(\cdot, x)\rangle| \leq \|f\| \|k(\cdot, x)\| = 0$$

  for all $x \in \mathcal{X}$; thus $f = 0$.

Proof of Theorem 3 III

- Completion:
  Let $\mathcal{H}$ be the completion of H_0.
  - H_0 is dense in $\mathcal{H}$ by the completion.
  - $\mathcal{H}$ is realized by functions:
    Let $\{f_n\}$ be a Cauchy sequence in $\mathcal{H}$. For each $x \in \mathcal{X}$, $\{f_n(x)\}$ is a Cauchy sequence, because

    $$|f_n(x) - f_m(x)| = |\langle f_n - f_m, k(\cdot, x)\rangle| \leq \|f_n - f_m\| \|k(\cdot, x)\|.$$

    Define $f(x) = \lim_n f_n(x)$.
    This value is the same for equivalent sequences, because $\{f_n\} \sim \{g_n\}$ implies

    $$|f_n(x) - g_n(x)| = |\langle f_n - g_n, k(\cdot, x)\rangle| \leq \|f_n - g_n\| \|k(\cdot, x)\| \to 0.$$

    Thus, any element $[\{f_n\}]$ in $\mathcal{H}$ can be regarded as a function f on $\mathcal{X}$.

---

[3] Note that Cauchy-Schwarz inequality holds without assuming strong positivity of the inner product.