Reproducing kernel Hilbert Space and C^*-Module

Jiyoung Park[1]

[1]Department of Statistics
Texas A&M University

October 10, 2025

Table of Contents

Reproducing Kernel Hilbert Space (RKHS)

- Reproducing Kernel Hilbert Space (RKHS) over $\mathbb{R}$ (generally $\mathbb{F}$).
  - $\mathcal{X}$: An arbitrary set (often called a data space).
  - $\mathcal{H} \subseteq \mathbb{R}^{\mathcal{X}} = \{f : \mathcal{X} \to \mathbb{R}\}$: A Hilbert space with a pointwise addition and multiplication.

Definition (RKHS)

1. *A positive definite kernel* $k : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ *is called a reproducing kernel of* $\mathcal{H}$ *if for all* $x \in \mathcal{X}$ $k(\cdot, x) \in \mathcal{H}$, *and for all* $f \in \mathcal{H}$ $f(x) = \langle f, k(\cdot, x)\rangle_{\mathcal{H}}$.
2. *If* $\mathcal{H}$ *has a reproducing kernel, it is called a 'reproducing kernel Hilbert space'.*
3. $\phi(x) := k(\cdot, x) \in \mathcal{H}$ *is called a 'feature map'.*

- We can also construct RKHS from the certain kernel.
  - Given k, a positive definite kernel, let $\phi(x) := k(\cdot, x) \in \mathbb{R}^{\mathcal{X}}$. Then, consider the following subset of $\mathbb{R}^{\mathcal{X}}$:

$$\mathcal{H}_k^0 := \left\{ \sum_{i=1}^{n} a_i \phi(x_i) \middle| n \in \mathbb{N},\ a_i \in \mathbb{R},\ x_i \in \mathcal{X} \right\}$$

with the inner product (well-defined due to positive definiteness of k)

$$\left\langle \sum_{i=1}^{n} a_i \phi(x_i), \sum_{j=1}^{m} b_j \phi(y_j) \right\rangle := \sum_i \sum_j a_i b_j k(x_i, y_j).$$

Then, $\mathcal{H}_k$, the completion of $\mathcal{H}_k^0$ by the given inner product structure, is RKHS with respect to the kernel k.

Theorem

The followings are equivalent:

1. $\mathcal{H}$ *is a RKHS.*
2. *For all* $x \in \mathcal{X}$, *a linear functional* $L_x : \mathcal{H} \to \mathbb{R}$ *defined by* $L_x(f) = f(x)$ *is continuous.*

- Meaning: Norm of RKHS stands for the pointwise convergence.

- Sketch of the proof:
  - $\Rightarrow$: For $f_n \xrightarrow{\mathcal{H}} f$, observe the following:

$$\begin{aligned} L_x(f) = f(x) = \langle f, k(x,\cdot)\rangle_{\mathcal{H}} &= \langle \lim f_n, k(x,\cdot)\rangle_{\mathcal{H}} \\ &= \lim \langle f_n, k(x,\cdot)\rangle_{\mathcal{H}} = \lim f_n(x) = \lim L_x(f_n). \end{aligned}$$

  - $\Leftarrow$: Consider the Riesz Representation (Hilbert space version) of L_x, denoted $h_x \in \mathcal{H}$. Define a kernel by $k(x,y) := h_x(y)$, and it is easy to check this kernel satisfies the desired properties.

- An 1-1 correspondence between a RKHS and a kernel.

Theorem

1. *For every $\mathcal{H}$ a RKHS, its reproducing kernel k is unique.*
2. *For every positive definite kernel k, $\mathcal{H}_k$ is the unique RKHS with respect to the kernel k.*

- Sketch of the proof
  1. Let k_1, k_2 be a kernels w.r.t. $\mathcal{H}$. Then, let $\phi_i(x) = k_i(\cdot, x) \in \mathcal{H}$. We observe the following:

$$\begin{aligned}\|\phi_1(x) - \phi_2(x)\|_{\mathcal{H}}^2 &= \langle \phi_1(x) - \phi_2(x), \phi_1(x) - \phi_2(x)\rangle \\ &= \langle \phi_1(X) - \phi_2(x), \phi_1(x)\rangle - \langle \phi_1(x) - \phi_2(x), \phi_2(x)\rangle \\ &= 0\end{aligned}$$

by reproducing properties. This implies $k_1(x, y) = k_2(x, y)$ for all $x, y \in \mathcal{X}$.

2. Let $\mathcal{H}_1, \mathcal{H}_2$ be RKHSs w.r.t. a kernel k.
   - Then, $\mathcal{H}_k^0$ is a linear dense subspace of both $\mathcal{H}_1$ and $\mathcal{H}_2$.
   - Therefore, for $f \in \mathcal{H}_i$, we can pick a Cauchy sequence $f_n \in \mathcal{H}_k^0$ converging to $f \in \mathcal{H}_i$ and $g \in \mathcal{H}_{-i}$.
   - Observe $f(x) = \lim f_n(x) = \lim \langle f_n, k(\cdot, x) \rangle_{\mathcal{H}_k^0} = g(x)$ for all $x \in \mathcal{X}$.
   - Therefore $\mathcal{H}_i \subseteq \mathcal{H}_{-i}$, and the vice versa.

Mercer's Theorem I

- Mercer Theorem: When $\mathcal{X}$ is compact, RKHS has a nice formulation.
  - Consider the case $\mathcal{X}$ is a compact Hausdorff. We fix k a kernel.
  - Consider a linear operator $T_k : L^2(\mathcal{X};\mu) \to L^2(\mathcal{X};\mu)$ defined by

$$T_k(f)(\cdot) = \int_{\mathcal{X}} k(\cdot, x) f(x) d\mu(x)$$

(Hilbert–Schmidt integral operator).

Theorem (Mercer)

There exists an orthonormal basis $\{e_i\} \subset L^2(\mathcal{X};\mu)$ *being eigenfunctions of* T_k *and* $\lambda_i \geq 0$ *being eigenvalues. Moreover,*

$$k(x,y) = \sum_{i=0}^{\infty} \lambda_i e_i(x) e_i(y)$$

and this series converges in $L^2(\mathcal{X};\mu)$*.*

- Takeaway: If $\mathcal{X}$ is compact, $\mathcal{H}_k$ is the eigenspace of T_k. *i.e.* $\{\phi_i = \sqrt{\lambda_i} e_i\}_i$ forms an orthonormal basis of $\mathcal{H}_k$.

TEXAS A&M UNIVERSITY

- Sketch of the proof.
  - The key is to show T_k is a self-adjoint compact operator, and then apply Spectral Theorem.
  - Self-adjointness comes from the positive definiteness of the kernel k.
  - Compactness of T_k comes from using Arzela-Ascoli on the image of T_k on the unit ball of $L^2(\mathcal{X})$ (Arzela-Ascoli part is where we use the compactness of $\mathcal{X}$).
  - Once obtain the compactness, we have Spectral Theorem for T_k, and therefore $\lambda_i e_i(x) = \int k(x,y)e_i(y)dy = \langle k(x,\cdot), e_i \rangle$ and

$$k(x,\cdot) = \sum_i \langle e_i, k(x,\cdot)\rangle e_i(\cdot) = \sum_i \lambda_i e_i(x) e_i(\cdot).$$

    Plugging-in $\cdot = y$ yields the result.
  - Convergence of the sum is guaranteed by the eigendecomposition.

- One of applications of RKHS theory is to the probability (or general measure) theory.
  - Fix a probability measure $\mathbb{P}$ in $\mathcal{X}$.
  - Fix a kernel k.
  - The, the following operation μ is called a 'kernel mean embedding'.

$$\mu_k(\mathbb{P}) = \mathbb{E}_{\mathbb{P}}(\phi(x)) = \int_{\mathcal{X}} k(\cdot, x) d\mathbb{P}(x) \in \mathcal{H}_k$$

  - This map is well-defined if $\mathbb{E}_{\mathbb{P}}\sqrt{k(x,x)} < \infty$ ($\because$ Hölder inequality).
  - $\langle \mu_k(\mathbb{P}), f \rangle = \mathbb{E}_{\mathbb{P}}(f(x))$, *i.e.* $\mu_k(\mathbb{P})$ is the Riesz representation (measure version) of $\mathbb{P}$.
  - Caution: Depending on k, $\mu_k(\cdot)$ may or may not be injective.
- Takeaway: Instead of dealing with measures (often a Banach space or metric space), we can deal with RKHS $\mathcal{H}_k$, which is easier.

Reproducing Kernel Hilbert C^*-Module (RKHM)

- Motivating example: Gaussian process
  - A stochastic process $\{X_i\}_{i\in\mathcal{I}} \subset X(\mathbb{R})$ is called a Gaussian process if for any $\mathcal{I}_d \subset \mathcal{I}$, $\{X_i\}_{\mathcal{I}_d}$ is a d-dimensional multivariate Gaussian distribution.
  - A zero-mean Gaussian process is fully determined if a covariance kernel $k(X_i, X_j) = Cov(X_i, X_j)$ is determined ($\approx$ A Gaussian distribution is fully determined with a mean and variance).

- What if $\{X_i\}_{i\in\mathcal{I}} \subset X(\mathbb{R}^d)$, a multivariate Gaussian process?
- Analogous framework would be $k(X_i, X_j)$ being a covariance matrix.
- k is no longer a $\mathbb{R}$-valued positive definite kernel, but still a symmetric positive definite operator.
- Can we extend the concept of RKHS for the case when k is an operator-valued kernel?
- RKHM is a fairly new concept to generalize RKHS to this kind of kernels ([Heo08], [HII$^+$21]).

- Reproducing Kernel Hilbert C^*-Module (RKHM)
  - $\mathcal{A}$: A C^*-algebra (not necessarily unital or commutative).

Definition (Hilbert C^*-Module)

1. *$\mathcal{M}$ is called a C^*-Module if $(\mathcal{M}, +)$ is an Abelien group and (right) $\mathcal{A}$-Module structure.*
2. *A map $\langle \cdot, \cdot \rangle_{\mathcal{M}} : \mathcal{M} \times \mathcal{M} \to \mathcal{A}$ is called $\mathcal{A}$-valued inner product if it satisfies the properties of inner product (substituting the positive definiteness condition by positiveness in $\mathcal{A}$).*
3. *A norm in $\mathcal{M}$ is defined by $\|u\|_{\mathcal{M}} := \|\langle u, u \rangle_{\mathcal{M}}^{1/2}\|_{\mathcal{A}}$.*
4. *If $(\mathcal{M}, \langle \cdot, \cdot \rangle_{\mathcal{M}})$ is complete w.r.t. $\| \cdot \|_{\mathcal{M}}$, then $\mathcal{M}$ is called a Hilbert C^*-Module.*

- RKHM can be constructed as the same way we constructed RKHS from a kernel.
  - A kernel $k : \mathcal{X} \times \mathcal{X} \to \mathcal{A}$ is called a $\mathcal{A}$-valued positive definite kernel if $k(x, y) = k(y, x)^*$ and $\sum_{i,j}^{n} a_i^* k(x_i, x_j) a_j \geq_{\mathcal{A}} 0$ for all $n \in \mathbb{N}$, $a_{i,j} \in \mathcal{A}$ and $x_{i,j} \in \mathcal{X}$.

TEXAS A&M UNIVERSITY

Definition (RKHM)

Let $\phi(x) := k(\cdot, x) : \mathcal{X} \to \mathcal{A}^{\mathcal{X}}$ *be a feature map. Consider the following subset of* $\mathcal{A}^{\mathcal{X}}$*:*

$$\mathcal{M}_k^0 := \left\{ \sum_{i=1}^{n} \phi(x_i) a_i \,\middle|\, n \in \mathbb{N}, a_i \in \mathcal{A}, x_i \in \mathcal{X} \right\}$$

equipped with the $\mathcal{A}$*-valued inner product*

$$\left\langle \sum_{i=1}^{n} \phi(x_i) a_i, \sum_{j=1}^{m} \phi(y_j) b_j \right\rangle_{\mathcal{M}_k^0} := \sum_i \sum_j a_i^* k(x_i, y_j) b_j.$$

Then, the completion of $\mathcal{M}_k^0$ *w.r.t. a norm* $\|\cdot\|_{\mathcal{M}_k^0}$ *is called a Reproducing Kernel Hilbert* C^**-Module (RKHM).*

Properties of RKHM I

- By construction, again kernel has a reproducing property, *i.e.* $\langle f, k(\cdot, x)\rangle_{\mathcal{M}_k} = f(x) \in \mathcal{A}$.
- Also, 1-1 correspondence between a kernel and RKHM also holds. The proof are exactly same with the 1-1 correspondence of RKHS.

Properties of RKHM II

- However. other properties of RKHS are not generalized well for RKHM with $\mathcal{A}$ being a general C^*-algebra.
  - e.g. Orthogonal projection Lemma, Riesz representation theorem.
- If we restrict $\mathcal{A}$ being a Von-Neumann algebra, Riesz representation theorem and Orthogonal projection Lemma are satisfied (this is true for general W^*-modules).
  - I tried to find a proof for this but could not...
- Therefore, we focus on the case when $\mathcal{A}$ is a Von-Neumann algebra. So precisely, we are now analyzing Reproducing Kernel Hilbert W^*-Module.

TEXAS A&M UNIVERSITY

- Here, we consider a set $\mathcal{D}(\mathcal{X}, \mathcal{A})$: a set of all $\mathcal{A}$-values finite Borel measures. Then, for all $\mathbb{P} \in \mathcal{D}(\mathcal{X}, \mathcal{A})$, Kernel mean embedding of $\mathbb{P}$ by $\mathcal{M}_k$ is

$$\mu_k(\mathbb{P}) := \mathbb{E}_{\mathbb{P}}(\phi(x)) = \int_{\mathcal{X}} k(\cdot, x) d\mathbb{P}(x) \in \mathcal{M}_k.$$

- The well-definedness is non-trivial in this case, unlike in RKHS.

Theorem (Well-definedness)

If $\int_{x\in\mathcal{X}} \|k(\cdot, x)\|_{\mathcal{M}_k} d|\mathbb{P}|(x) < \infty$, *then for all* $f \in \mathcal{M}_k$,

$$\langle \mu(\mathbb{P}), f \rangle_{\mathcal{M}_k} = \int_{x\in\mathcal{X}} d\mathbb{P}(x)^* f(x)$$

Kernel Mean Embedding II

- Sketch of the proof:
  1. Define $L_{\mathbb{P}} : \mathcal{M}_k \to \mathcal{A}$ by $L_{\mathbb{P}}(f) = \int_{x \in \mathcal{X}} d\mathbb{P}(x)^* f(x)$.
  2. Observe $L_{\mathbb{P}}$ is bounded. *i.e.*

$$\begin{aligned} \|L_{\mathbb{P}} f\|_{\mathcal{M}_k} &\leq \int_{x \in \mathcal{X}} \|f(x)\|_{\mathcal{A}} d|\mathbb{P}|(x) = \int_{x \in \mathcal{X}} \langle f, k(\cdot, x) \rangle_{\mathcal{A}} d|\mathbb{P}|(x) \\ &\leq \|f\|_{\mathcal{M}_k} \underbrace{\int_{x \in \mathcal{X}} \|k(\cdot, x)\|_{\mathcal{M}_k} d|\mathbb{P}|(x)}_{<\infty \text{ by the condition}}. \end{aligned}$$

  3. Apply Riesz representation theorem for Hilbert W^*-Module: There exists an $\mu(\mathbb{P}) \in \mathcal{M}_k$ such that $L_{\mathbb{P}} f = \langle f, \mu(\mathbb{P}) \rangle_{\mathcal{M}_k}$, which completes the proof.

Concluding remarks

- Reproducing Kernel property is useful in many sides (e.g. Quantum mechanics, Machine Learning, ...).
- RKHS is the most basic one, and there are many extensions (e.g. Reproducing Kernel Banach space, ...). Here we focused on RKHM.

- Jaeseong Heo, *Reproducing kernel hilbert c-modules and kernels associated with cocycles*, Journal of Mathematical Physics **49** (2008), no. 10.
- Yuka Hashimoto, Isao Ishikawa, Masahiro Ikeda, Fuyuta Komura, Takeshi Katsura, and Yoshinobu Kawahara, *Reproducing kernel hilbert c*-module and kernel mean embeddings*, 2021.