Reproducing kernel Hilbert C^*-module for data analysis

Yuka Hashimoto

NTT Network Service Systems Laboratories /RIKEN AIP

May 2nd, 2023

- Y. Hashimoto, I. Ishikawa, M. Ikeda, F. Komura, T. Katsura, and Y. Kawahara, JMLR, 22(267):1–56 (updated version : arXiv:2101.11410v2)
- Y. Hashimoto, F. Komura, and M. Ikeda, Matrix and Operator Equations, to appear
- Y. Hashimoto, M. Ikeda, and H. Kadri, AISTATS 2023

Introduction

Yuka Hashimoto

NTT Network Service Systems Laboratories / RIKEN AIP

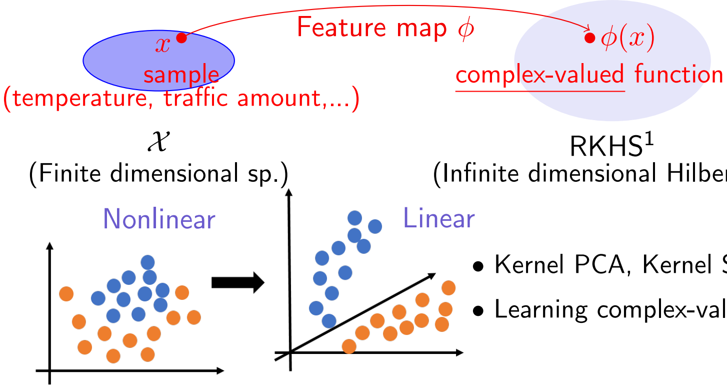
- 2018 Received Master's degree from Keio University
- 2018- NTT Network Service Systems Laboratories
- 2022 Received Ph.D. from Keio University
- 2022- Visiting researcher at RIKEN AIP

Backgrounds / Interests

- Operator theoretic data analysis
- Kernel methods
- Numerical linear algebra

Contents

Kernel methods



x

sample

(temperature, traffic amount,...)

Feature map ϕ

$\phi(x)$

complex-valued function

$\mathcal{X}$

(Finite dimensional sp.)

RKHS[1]

(Infinite dimensional Hilbert sp.)

Nonlinear

Linear

- Kernel PCA, Kernel SVM
- Learning complex-valued functions

Advantages of RKHS

- Nonlinearity in the original space is transformed into a linear one.
- We can compute inner products in RKHS exactly by computers.

---

[1] Schölkopf and Smola, MIT Press, Cambridge, 2001

Reproducing kernel Hilbert space (RKHS)

Let $\mathcal{X}$ be a set. A map $k : \mathcal{X} \times \mathcal{X} \to \mathbb{C}$ is called a positive definite kernel if it satisfies:

1. $k(x, y) = \overline{k(y, x)}$ for $x, y \in \mathcal{X}$ and
2. $\sum_{t,s=1}^{n} \overline{c_t} k(x_t, x_s) c_s \geq 0$ for $n \in \mathbb{N}$, $c_1, \ldots, c_n \in \mathbb{C}$, $x_1, \ldots, x_n \in \mathcal{X}$.

$\phi(x) := k(\cdot, x)$ ($\phi : \mathcal{X} \to \mathbb{C}^{\mathcal{X}}$: feature map associated with k),

$$\mathcal{H}_{k,0} := \Big\{ \sum_{t=1}^{n} \phi(x_t) c_t \Big| \ n \in \mathbb{N}, \ c_t \in \mathbb{C}, \ x_t \in \mathcal{X} \Big\}. \tag{1}$$

We can define an inner product $\langle \cdot, \cdot \rangle_k : \mathcal{H}_{k,0} \times \mathcal{H}_{k,0} \to \mathbb{C}$ as

$$\Big\langle \sum_{s=1}^{n} \phi(x_s) c_s, \sum_{t=1}^{l} \phi(y_t) d_t \Big\rangle_k := \sum_{s=1}^{n} \sum_{t=1}^{l} \overline{c_s} k(x_s, y_t) d_t. \tag{2}$$

RKHS $\mathcal{H}_k$: completion of $\mathcal{H}_{k,0}$

Representer theorem in RKHSs

The representer theorem guarantees that solutions of a minimization problem are represented only with given samples[2].

$\mathcal{H}_k$: RKHS
$\mathbb{R}_+ := \{a \in \mathbb{R} \mid a \geq 0\}$

Theorem 1 Representer theorem in RKHSs

Let $x_1, \ldots, x_n \in \mathcal{X}$ and $a_1, \ldots, a_n \in \mathbb{C}$. Let $h : \mathcal{X} \times \mathbb{C}^2 \to \mathbb{R}_+$ be an error function and $g : \mathbb{R}_+ \to \mathbb{R}_+$ satisfy $g(c) < g(d)$ for $c < d$. Then, any $u \in \mathcal{H}_k$ minimizing $\sum_{i=1}^n h(x_i, a_i, u(x_i)) + g(\|u\|_k)$ admits a representation of the form $\sum_{i=1}^n \phi(x_i)c_i$ for some $c_1, \ldots, c_n \in \mathbb{C}$.

The result can be applied to supervised problems.

---

[2] Schölkopf et al., COLT 2001.

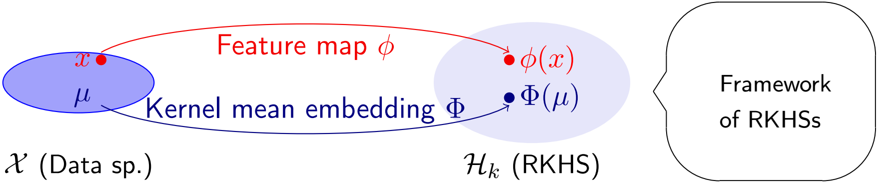
Kernel mean embedding in RKHSs

The kernel mean embedding enables us to generalize the framework of RKHSs to analyzing measures.

$k : \mathcal{X} \times \mathcal{X} \to \mathbb{C}$: positive definite kernel, $\mathcal{H}_k$: RKHS

$\mathcal{D}(\mathcal{X})$: set of all complex-valued finite regular Borel measures
Kernel mean embedding in $\mathcal{H}_k$[3]:

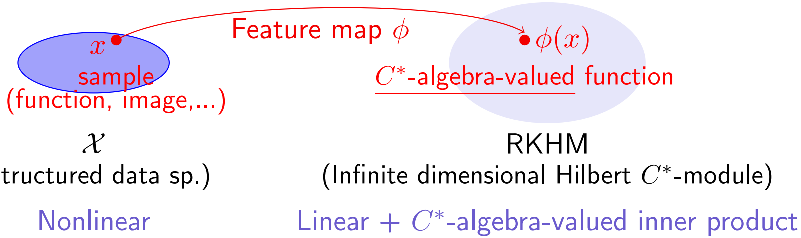
$$\Phi : \mathcal{D}(\mathcal{X}) \to \mathcal{H}_k, \ \langle \Phi(\mu), v \rangle_k = \int_{x \in \mathcal{X}} v(x) d\mu(x)$$



---

[3] Muandet et al., Kernel Mean Embedding of Distributions: A Review and Beyond, 2017.

Goal: Generalization of data analysis in RKHS to RKHM



Advantages of RKHM:

- C^*-algebra-valued inner products extract information of structures.

We constructed a framework of data analysis with RKHM.

- We can reconstruct existing RKHSs by using RKHMs.
- We have shown fundamental properties for data analysis in RKHMs (e.g. representer theorem, kernel mean embedding).

C^*-algebra and von Neumann-algebra

C^*-algebra : Banach space equipped with a product & an involution * + α

e.g.

- $C(\mathcal{Z})$ for a compact space $\mathcal{Z}$
  Norm : sup norm, **Product** : pointwise product,
  Involution : pointwise complex conjugate
- $\mathcal{K}(\mathcal{H}) = \{$compact operators on a Hilbert space $\mathcal{H}\}$
  Norm : operator norm, **Product** : composition, **Involution** : adjoint

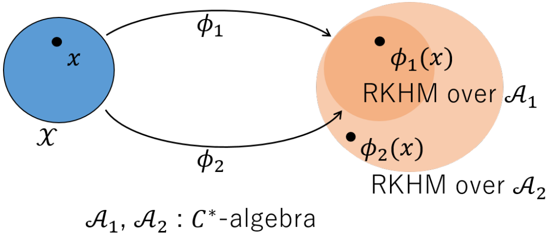
Von Neumann-algebra : C^*-algebra that is closed in the strong operator topology

e.g.

- $L^\infty(\mathcal{Z})$ for a measure space $\mathcal{Z}$
- $\mathcal{B}(\mathcal{H}) = \{$bounded linear operators on a Hilbert space $\mathcal{H}\}$

Advantages of RKHM

- Enlarge representation spaces using C^*-algebras (e.g. use the C^*-algebra of continuous functions for functional data).



- Construct positive definite kernels from the perspective of C^*-algebra. (Make use of the product structure.)
  e.g. polynomial kernel $k(x, y) = x^*y + x^*x^*yy$ ($x, y \in \mathcal{A}_1$ or $\mathcal{A}_2$)

Hilbert C^*-module

$\mathcal{A}$: C^*-algebra
$\mathcal{M}$: right $\mathcal{A}$-module ($u \in \mathcal{M}$, $c \in \mathcal{A} \to uc \in \mathcal{M}$)

Definition 1 $\mathcal{A}$-valued inner product

A map $\langle \cdot, \cdot \rangle : \mathcal{M} \times \mathcal{M} \to \mathcal{A}$ is called an $\mathcal{A}$-valued inner product if it satisfies the following properties for $u, v, w \in \mathcal{M}$ and $c, d \in \mathcal{A}$:

1. $\langle u, vc + wd \rangle = \langle u, v \rangle c + \langle u, w \rangle d$,
2. $\langle v, u \rangle = \langle u, v \rangle^*$,
3. $\langle u, u \rangle \geq 0$ and if $\langle u, u \rangle = 0$ then $u = 0$.

→ *$\mathcal{A}$-valued absolute value* $|u| := \langle u, u \rangle^{1/2}$ → Norm $\|u\| := \| \langle u, u \rangle \|_{\mathcal{A}}^{1/2}$

Hilbert C^*-module $\mathcal{M}$[4]: complete $\mathcal{A}$-module equipped with an $\mathcal{A}$-valued inner-product

---

[4] Lance, Cambridge University Press, 1995.

Review of reproducing kernel Hilbert C^*-module

$\mathcal{A}$: C^*-algebra

RKHS ($\mathcal{H}_k$):

- $\mathbb{C}$-valued positive definite kernel k
- $\mathbb{C}$-valued functions
- $\mathbb{C}$-valued inner product

RKHM over $\mathcal{A}$ ($\mathcal{M}_k$):

- $\mathcal{A}$-valued positive definite kernel k
- $\mathcal{A}$-valued functions
- $\mathcal{A}$-valued inner product

Reproducing kernel Hilbert C^*-module (RKHM)

Let $\mathcal{X}$ be a set. A map $k : \mathcal{X} \times \mathcal{X} \to \mathcal{A}$ is called an $\mathcal{A}$-valued positive definite kernel if it satisfies:

1. $k(x, y) = k(y, x)^*$ for $x, y \in \mathcal{X}$ and
2. $\sum_{t,s=1}^{n} c_t^* k(x_t, x_s) c_s \geq 0$ for $n \in \mathbb{N}$, $c_1, \ldots, c_n \in \mathcal{A}$, $x_1, \ldots, x_n \in \mathcal{X}$.

$\phi(x) := k(\cdot, x)$ ($\phi : \mathcal{X} \to \mathcal{A}^{\mathcal{X}}$: feature map associated with k),

$$\mathcal{M}_{k,0} := \left\{ \sum_{t=1}^{n} \phi(x_t) c_t \,\middle|\, n \in \mathbb{N},\ c_t \in \mathcal{A},\ x_t \in \mathcal{X} \right\}. \tag{3}$$

We can define an $\mathcal{A}$-valued inner product $\langle \cdot, \cdot \rangle_k : \mathcal{M}_{k,0} \times \mathcal{M}_{k,0} \to \mathcal{A}$ as

$$\left\langle \sum_{s=1}^{n} \phi(x_s) c_s, \sum_{t=1}^{l} \phi(y_t) d_t \right\rangle_k := \sum_{s=1}^{n} \sum_{t=1}^{l} c_s^* k(x_s, y_t) d_t. \tag{4}$$

RKHM $\mathcal{M}_k$: completion of $\mathcal{M}_{k,0}$

Representer theorem in RKHMs

To generalize complex-valued supervised problems to $\mathcal{A}$-valued ones, we show a representer theorem.

$\mathcal{M}_k$: RKHM over $\mathcal{A}$, $|\cdot|_k$: absolute value in $\mathcal{M}_k$
$\mathcal{A}_+ := \{a \in \mathcal{A} \mid {}^\exists b \in \mathcal{A}$ such that $a = b^* b\}$

Theorem 2 Representer theorem in RKHMs

Let $\mathcal{A}$ be a unital C^*-algebra, $x_1, \ldots, x_n \in \mathcal{X}$ and $a_1, \ldots, a_n \in \mathcal{A}$. Let $h : \mathcal{X} \times \mathcal{A}^2 \to \mathcal{A}_+$ be an error function and $g : \mathcal{A}_+ \to \mathcal{A}_+$ satisfy $g(c) < g(d)$ for $c < d$. If $\mathrm{Span}_{\mathcal{A}}\{\phi(x_i)\}_{i=1}^n$ is closed, any $u \in \mathcal{M}_k$ minimizing $\sum_{i=1}^n h(x_i, a_i, u(x_i)) + g(|u|_k)$ admits a representation of the form $\sum_{i=1}^n \phi(x_i) c_i$ for some $c_1, \ldots, c_n \in \mathcal{A}$.

Key point of the proof:
For a Hilbert C^*-module $\mathcal{M}$ over a unital C^*-algebra $\mathcal{A}$ and any finitely generated closed submodule $\mathcal{V}$ of $\mathcal{M}$, $u \in \mathcal{M}$ is decomposed into $u = u_1 + u_2$ where $u_1 \in \mathcal{V}$ and $u_2 \in \mathcal{V}^\perp$.

Approximate representer theorem in RKHMs

If $\mathcal{A}$ is a von Neumann algebra, we can show an approximate representer theorem under mild conditions.

Theorem 3 Approximate representer theorem in RKHMs

Let $\mathcal{A}$ be a von Neumann-algebra, $x_1, \ldots, x_n \in \mathcal{X}$ and $a_1, \ldots, a_n \in \mathcal{A}$. Let $h : \mathcal{X} \times \mathcal{A}^2 \to \mathcal{A}_+$ be a Lipschitz continuous error function with Lipschitz constant L and $g : \mathcal{A}_+ \to \mathcal{A}_+$ satisfy $g(c) < g(d)$ for $c < d$. Assume $f(u) := \sum_{i=1}^n h(x_i, a_i, u(x_i)) + g(|u|_k)$ has a minimizer u. Then, for any $\epsilon > 0$, there exists $v \in \mathcal{M}_k$ of the form $\sum_{i=1}^n \phi(x_i) c_i$ such that $\|f(v) - f(u)\| \leq Ln\epsilon\|u\|$.

Key point of the proof:

If $\mathcal{A}$ is a von Neumann-algebra, we can apply the spectral decomposition and construct an "orthonormalization" to the module generated by $\{\phi(x_i)\}_{i=1}^n$.

Kernel mean embedding in RKHMs

$\mathcal{A} \subseteq \mathcal{B}(\mathcal{H})$, $\mathcal{M}$: Hilbert $\mathcal{A}$-module.

Definition 2 Internal tensor

The completion of $\mathcal{M} \otimes \mathcal{H}$ w.r.t. $\langle w_1 \otimes u_1, w_2 \otimes u_2 \rangle = \langle u_1, \langle w_2, w_2 \rangle_{\mathcal{M}} u_2 \rangle_{\mathcal{H}}$ is denoted as $\mathcal{M} \otimes_{\mathcal{A}} \mathcal{H}$.

Definition 3 Von Neumann module

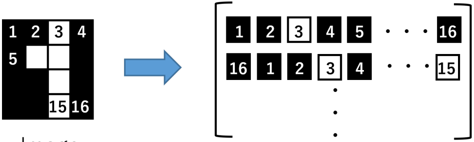
$\mathcal{W} = \mathcal{M} \otimes_{\mathcal{A}} \mathcal{H}$. Identify $w \in \mathcal{M}$ with the map $u \mapsto w \otimes_{\mathcal{A}} u$ and regard $\mathcal{M} \subseteq \mathcal{B}(\mathcal{H}, \mathcal{W})$. If $\mathcal{M} \subseteq \mathcal{B}(\mathcal{H}, \mathcal{W})$ is strongly closed, $\mathcal{M}$ is called a von Neumann $\mathcal{A}$-module.

Assume k is bounded and $\phi(x) = k(\cdot, x) \in C_0(\mathcal{X}, \mathcal{A})$ for any $x \in \mathcal{X}$. Assume $\mathcal{M}_k$ is a von Neumann-module (Riesz representation theorem is available). Kernel mean embedding $\Phi : \mathcal{D}(\mathcal{X}, \mathcal{A}) \to \mathcal{M}_k$ is defined as

$$\Phi(\mu) = \int_{x \in \mathcal{X}} \phi(x) d\mu(x). \tag{5}$$

Supervised learning in RKHM

$x_1, \ldots, x_n \in \mathcal{X}$: input training samples (e.g. $Circ(p)$)
$a_1, \ldots, a_n \in \mathcal{A}$: output training samples (e.g. $\mathbb{C}^{p \times p}$)



Image

Circulant matrix

Product structure:
Convolution

Minimization problem ($\lambda \geq 0$: regularization parameter)

$$\min_{f \in \mathcal{M}_k} \left(\sum_{i=1}^{n} |f(x_i) - a_i|_{\mathcal{A}}^2 + \lambda |f|_k^2 \right). \tag{6}$$

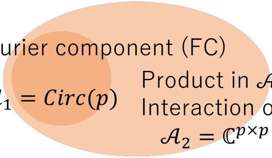
→ Apply representer theorem in RKHM.
→ Solve the minimization problem using the Gram matrix.

Supervised learning in RKHM

Example of kernels (q-th degree polynomial kernel)

$\mathcal{A}_1 := Circ(p) \subset \mathbb{C}^{p\times p} =: \mathcal{A}_2$, $\mathcal{X} \subseteq \mathcal{A}_1^d$ (e.g. images),
$a_{i,j} \in \mathcal{A}_2$ $(i = 1, \ldots d, j = 1, \ldots q+1)$, $x_1 = [x_{1,1}, \ldots, x_{1,n}]$

$$k(x_1, x_2) = \sum_{i=1}^{d} \bigg(\prod_{j=1}^{q} a_{i,j}^* x_{1,i}^* \bigg) a_{i,q+1}^* a_{i,q+1} \bigg(\prod_{j=1}^{q} x_{2,i} a_{i,q+1-j} \bigg) \in \mathcal{A}_2. \quad (7)$$



Product in $\mathcal{A}_1$ (convolution):
Pointwise product of each Fourier component (FC)

$\mathcal{A}_1 = Circ(p)$

Product in $\mathcal{A}_2$:
Interaction of different FCs

$\mathcal{A}_2 = \mathbb{C}^{p\times p}$

By setting $a_{i,j} \in \mathcal{A}_2$, we beyond the convolution in existing methods.

Connection with CNN

$\mathcal{A}_1 := Circ(p)$, $a_1, \ldots, a_L, b_1, \ldots, b_L \in \mathcal{A}_1$, $\sigma_1, \ldots, \sigma_L : \mathcal{A}_1 \to \mathcal{A}_1$

$$\begin{aligned}\hat{k}(x, y) := \sigma_L(b_L^* b_L + \sigma_{L-1}(b_{L-1}^* b_{L-1} + \cdots \\ + \sigma_2(b_2^* b_2 + \sigma_1(b_1^* b_1 + x^* a_1^* a_1 y) a_2^* a_2) \cdots a_{L-1}^* a_{L-1}) a_L^* a_L).\end{aligned} \tag{8}$$

Then, $\hat{k}$ is an $\mathcal{A}_1$-valued positive definite kernel.

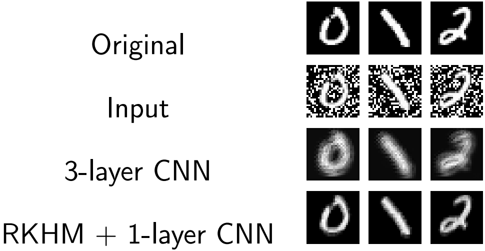
The solution f of the supervised problem is

$$\begin{aligned}f(x) = \sum_{i=1}^{n} \sigma_L(b_L^* b_L + \sigma_{L-1}(b_{L-1}^* b_{L-1} + \cdots \\ + \sigma_2(b_2^* b_2 + \underset{\text{Activation}}{\sigma_1}(\overset{\text{Bias}}{b_1^* b_1} + \underset{\text{Convolution}}{x^* a_1^* a_1 x_i}) a_2^* a_2) \cdots a_{L-1}^* a_{L-1}) a_L^* a_L) c_i,\end{aligned} \tag{9}$$

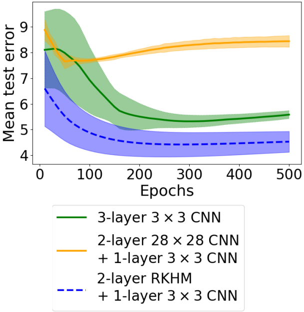
By setting $a_1 \ldots, a_L, b_1, \ldots, b_L \in \mathcal{A}_2 \supset \mathcal{A}_1$, we beyond CNNs.

Numerical results

Noise reduction for MNIST (number of samples : 20)



Input and output images



Mean test error versus the number of epochs

A CNN with an RKHM outperformed a CNN without an RKHM.

Conclusion

- RKHM is a natural generalization of RKHS.
- We showed a representer theorem and an approximate representer theorem in RKHMs and defined a kernel mean embedding in RKHMs.
- RKHMs are useful for analyzing image data.