Spectral truncation を用いた関数値カーネルの構成

橋本悠香

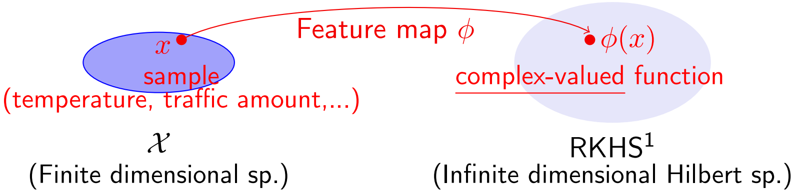
NTT / RIKEN AIP

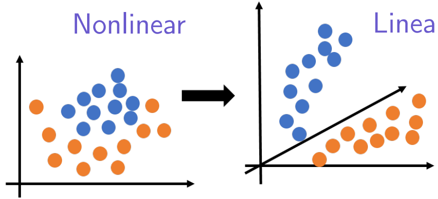
2025/6/13

- Y. Hashimoto, F. Komura, and M. Ikeda, Matrix and Operator Equations, pp. 1–27.
- Y. Hashimoto, A. Hafid, M. Ikeda, and H. Kadri, arXiv: 2405.17823

Contents

Background: Kernel methods





- 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

Background: 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),

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

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

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

Reproducing property: $\langle \phi(x), v \rangle_k = v(x)$ for $v \in \mathcal{H}_k$ and $x \in \mathcal{X}$

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

Background: 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
$\mathbb{R}_+ := \{a \in \mathbb{R} \mid a \geq 0\}$

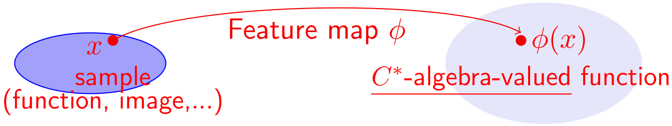
Theorem 1 Representer theorem in RKHSs

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.

Goal: Generalization of data analysis in RKHS to RKHM



$\mathcal{X}$
(Structured data sp.)

RKHM
(Infinite dimensional Hilbert C^*-module)

Nonlinear

Linear + C^*-algebra-valued inner product

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

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

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

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

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

Reproducing property: $\langle \phi(x), v \rangle_k = v(x)$ for $v \in \mathcal{M}_k$ and $x \in \mathcal{X}$

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 $w \in \mathcal{M}_k$ minimizing $f(w) := \sum_{i=1}^n h(x_i, a_i, w(x_i)) + g(|w|_k)$ 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 $\mathcal{M}$, $w \in \mathcal{M}$ is decomposed into $w = w_1 + w_2$ where $w_1 \in \mathcal{V}$ and $w_2 \in \mathcal{V}^\perp$.

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 $f(w) := \sum_{i=1}^n h(x_i, a_i, w(x_i)) + g(|w|_k)$ has a minimizer w. 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 $\|f(v) - f(w)\|_{\mathcal{A}} \leq Ln\epsilon \|w\|_{\mathcal{A}}$.

Key point of the proof:
If $\mathcal{A}$ is a von Neumann-algebra, we can apply the Gram–Schmidt orthonormalization to construct a module approximating the module generated by $\{\phi(x_i)\}_{i=1}^n$.

Background: Challenge of applying RKHM

A challenge of kernel methods: choice of kernels

We focus on the case of $\mathcal{X} = C(\mathbb{T})^d$ and $\mathcal{A} = C(\mathbb{T})$.
Typical examples of $C(\mathbb{T})$-valued kernels for functional inputs:

1. Commutative kernels
   - $k(x, y)(z) = \tilde{k}(x(z), y(z))$ for $x, y \in C(\mathbb{T})^d$,

where $\tilde{k} : \mathbb{C}^d \times \mathbb{C}^d \to \mathbb{C}$ is a complex-valued positive definite kernel.

$k(x, y)(z)$ is determined only with $x(z)$ and $y(z)$.

We can only extract local dependencies of the output function on the input function.

Background: Challenge of applying RKHM

A challenge of kernel methods: choice of kernels

Typical examples of $C(\mathbb{T})$-valued kernels for functional inputs:

2. Separable kernels
   - $k(x,y) = \tilde{k}(x,y)a$ for $x, y \in C(\mathbb{T})^d$,

where $\tilde{k} : C(\mathbb{T})^d \times C(\mathbb{T})^d \to \mathbb{C}$ is a complex-valued positive definite kernel and $a \in C(\mathbb{T})$ satisfies $a \geq 0$.

The dependency of the output function on the input function is determined only with the fixed function a.

We can only extract global dependencies of the output function on the input function.

Goal: Construct kernels that fill a gap between commutative and separable kernels.

Background: Spectral truncation

$e_j(z) = \mathrm{e}^{\mathrm{i}jz}$ $(j \in \mathbb{Z})$: Fourier function.
$x \in \mathcal{A} = C(\mathbb{T})$, M_x: multiplication operator w.r.t. x on $L^2(\mathbb{T})$
P_n: projection onto $\mathrm{Span}\{e_1, \ldots, e_n\}$

Approximate M_x by $P_n M_x P_n$. The representtion matrix is

$$R_n(x)_{j,l} := \langle e_j, M_x e_l \rangle_{L^2(\mathbb{T})} = \int_{\mathbb{T}} x(t)\mathrm{e}^{-\mathrm{i}(j-l)t}\mathrm{d}t. \tag{5}$$

- $R_n(x)$ is a Toeplitz matrix ((j,l)-element depends on $j-l$).
- $R_n(x)$ corresponds to the coefficients of $e_{-(n-1)}, \ldots, e_{n-1}$.

Define $S_n : \mathbb{C}^{n\times n} \to \mathcal{A}$ by

$$S_n(A)(z) := \frac{1}{n}\sum_{j,l=0}^{n-1} A_{j,l}\mathrm{e}^{\mathrm{i}(j-l)z}. \tag{6}$$

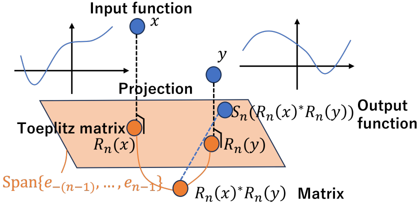
Kernels with Toeplitz matrices

New class of kernels based on spectral truncation:

- Polynomial kernel: $k_n^{\mathrm{poly},q}(x,y) = S_n\left(\sum_{i=1}^{d} \alpha_i (R_n(x_i)^*)^q R_n(y_i)^q\right)$,
- Product kernel:

$$k_n^{\mathrm{prod},q}(x,y) = S_n\left(\prod_{j=1}^{q} R_n(\tilde{k}_{1,j}(x,y))^* \prod_{j=1}^{q} R_n(\tilde{k}_{2,j}(x,y))\right),$$

where $\tilde{k}_{1,j}$, $\tilde{k}_{2,j}$, are complex-valued positive definite kernels, $x = [x_1, \ldots, x_d]$, and $\alpha_i \geq 0$.



The case of $n = \infty$: commutative kernel

$$S_n\bigg(\prod_{j=1}^{q} R_n(x_j)^* \prod_{j=1}^{q} R_n(y)\bigg)(z)$$
$$= \int_{\mathbb{T}^{2q}} \overline{x_1(t_1)} \cdots \overline{x_q(t_q)} y_1(t_{q+1}) \cdots y_q(t_{2q}) F_n^{2q,P}(z\mathbf{1} - t)\mathrm{d}t, \tag{7}$$

where $F_n^{q,P}(t) = 1/n \sum_{j=1}^{n-1} \sum_{r \in jP \bigcap \mathbb{Z}^q} \mathrm{e}^{\mathrm{i}r\cdot t}$ is the generalized Fejér kernel and $P = \{[r_1, \ldots, r_{2q}] \in \mathbb{R}^{2q} \mid |\sum_{i=l}^{m} r_i| \leq 1,\ l \leq m\}$.

Using this formula, we can show the proposed kernels are commutative if $n = \infty$ (local dependencies).

Theorem 4

For $x, y \in \mathcal{A}^d$ and $z \in \mathbb{T}$, $k_n^{\mathrm{poly},q}(x,y)(z) \to k^{\mathrm{poly},q}(x,y)(z)$, $k_n^{\mathrm{prod},q}(x,y)(z) \to k^{\mathrm{prod},q}(x,y)(z)$ as $n \to \infty$, where $k^{\mathrm{poly},q}(x,y)(z) = \sum_{i=1}^{d} \alpha_i(\overline{x_i(z)} y_i(z))^q$ and $k^{\mathrm{prod},q}(x,y)(z) = \prod_{j=1}^{q} \overline{\tilde{k}_{1,j}(x(z), y(z))} \tilde{k}_{2,j}(x(z), y(z))$.

The case of $n = 1$: separable kernel

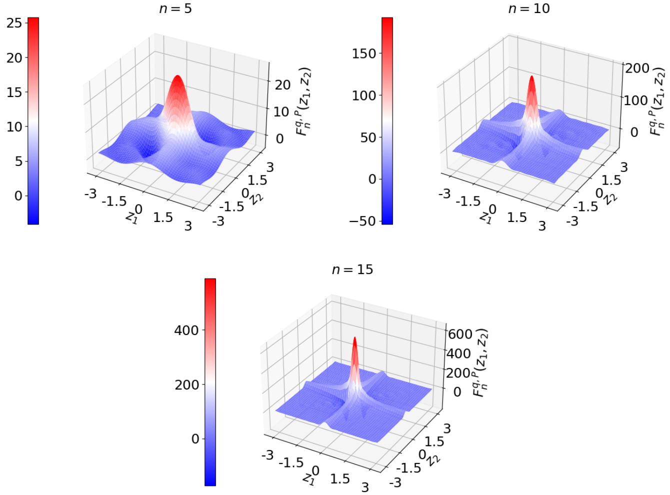
Since $R_1(x) = \int_{\mathbb{T}} x(t)\mathrm{d}t$, we have

- $k_1^{\mathrm{poly},q}(x,y) = S_n\big(\sum_{i=1}^d \alpha_i (R_n(x_i)^*)^q R_n(y_i)^q\big)$
  $= \sum_{i=1}^d \alpha_i \int_{\mathbb{T}} \overline{x(t)}\mathrm{d}t^q \int_{\mathbb{T}} y(t)\mathrm{d}t^q$,
- $k_1^{\mathrm{prod},q}(x,y) = S_n\big(\prod_{j=1}^q R_n(\tilde{k}_{1,j}(x,y))^* \prod_{j=1}^q R_n(\tilde{k}_{2,j}(x,y))\big)$
  $= \prod_{j=1}^q \int_{\mathbb{T}} \overline{\tilde{k}_{1,j}(x(t),y(t))}\mathrm{d}t \int_{\mathbb{T}} \tilde{k}_{2,j}(x(t),y(t))\mathrm{d}t$,

The proposed kernels are separable if $n = 1$ (global dependencies)

$\rightarrow$ If $1 < n < \infty$, the proposed kernels can extract both local and global dependencies.

Fejér kernel



Féjer kernel $F_n^{2,P}$ for $n = 5, 10, 15$

Positive definiteness

Proposition 1

The kernel $k_n^{\mathrm{poly},q}$ is positive definite.

As for $k_n^{\mathrm{prod},q}$, we cannot separate x and y as products. Thus, we modify the kernel to guarantee the positive definiteness.

Proposition 2

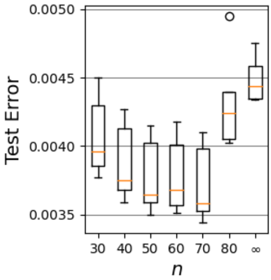
Let $\beta_n \geq -\min_{t\in\mathbb{T}^q} F_n^{2q,P}(t)$. Let

$$\hat{k}_n^{\mathrm{prod},q}(x,y) = k_n^{\mathrm{prod},q}(x,y) + \beta_n \int_{\mathbb{T}^{2q}} \prod_{j=1}^{q} \overline{\tilde{k}_{1,j}(x(t_j), y(t_j))}\tilde{k}_{2,j}(x(t_{q+j}), y(t_{q+j}))\mathrm{d}t.$$

Then, $\hat{k}_n^{\mathrm{prod},q}$ is positive definite.

Numerical results

Image reconstruction task with MNIST.
Input (x_i): Masked images, Output (a_i): Recovered images



$n = 30$ $n = 40$ $n = 50$ $n = 60$ $n = 70$ $n = 80$ $n = \infty$

Test error and the reconstructed images for different values of n.

Conclusion

- RKHM is a natural generalization of RKHS.
- RKHMs are useful for analyzing image data and functional data.
- We constructed new class of kernels based on spectral truncation.
- The proposed kernels fill a gap between existing commutative and separable kernels.