Reproducing Kernel Hilbert Space (Part I)

Ergan Shang

USTC

November 3, 2022

Overview

Overview

1. Priliminary
2. Reproducing kernel Hilbert Space
    - Constructing from a kernel
    - Abstract Version
3. Mercer's theorem and consequences

Theorem 12.5 (Riesz representation theorem)

let L be a bounded linear functional on a Hilbert space. Then there exists a unique $g \in \mathcal{H}$ such that $L(f) = \langle f, g \rangle_{\mathcal{H}}$ for all $f \in \mathcal{H}$. And we refer to g as the representer of the functional L.

Also recall the definitions of inner product and Hilbert space.

Overview

1. Priliminary
2. Reproducing kernel Hilbert Space
   - Constructing from a kernel
   - Abstract Version
3. Mercer's theorem and consequences

PSD kernel function

Definition 12.6

A symmetric bivariate function $\mathcal{K} : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ is positive semidefinite (PSD) if for all integers $n \geq 1$ and elements $\{\boldsymbol{x}_i\}_{i=1}^n \subset \mathcal{X}$, the $n \times n$ matrix withh elements $K_{ij} = \mathcal{K}(\boldsymbol{x}_i, \boldsymbol{x}_j)$ is positive semidefinite.

12.7 Let $\mathcal{X} = \mathbb{R}^d$, we define $\mathcal{K}(\boldsymbol{x}, \boldsymbol{x}') = \langle \boldsymbol{x}, \boldsymbol{x}' \rangle$. Then with $\{\boldsymbol{x}_i\}_{i=1}^n \subset \mathbb{R}^d$, we have

$$\boldsymbol{\alpha}^\top K \boldsymbol{\alpha} = \sum_{i,j=1}^n \alpha_i \alpha_j \langle \boldsymbol{x}_i, \boldsymbol{x}_j \rangle = \left\| \sum_{i=1}^n \alpha_i \boldsymbol{x}_i \right\|_2^2 \geq 0.$$

Polynomials kernels and Feature maps

12.8 We let $\mathcal{K}(\boldsymbol{x}, \boldsymbol{z}) = (\langle \boldsymbol{x}, \boldsymbol{z} \rangle)^2$, so that

$$\mathcal{K}(\boldsymbol{x}, \boldsymbol{z}) = \sum_{j=1}^{d} x_j^2 z_j^2 + 2 \sum_{i<j} x_i x_j z_i z_j.$$

Setting $D = d + \binom{d}{2}$, and define **feature mapping** $\phi : \mathbb{R}^d \to \mathbb{R}^D$ with entries

$$\phi(\boldsymbol{x}) = \begin{bmatrix} x_j^2, & \text{for } j = 1, 2, \cdots, d \\ \sqrt{2} x_i x_j, & \text{for } i < j \end{bmatrix}.$$

As a result, we can write $\mathcal{K}(\boldsymbol{x}, \boldsymbol{z}) = \langle \phi(\boldsymbol{x}), \phi(\boldsymbol{z}) \rangle_{\mathbb{R}^D}$, which is PSD followed by the last example.

Feature maps

12.10 Consider the Fourier basis $\phi_j(x) = \sin\left(\frac{(2j-1)\pi x}{2}\right)$ and $\langle \phi_j, \phi_k \rangle = \int_0^1 \phi_j(x)\phi_k(x)dx = \delta_{jk}$. Given some sequence $\{\mu_j\}_{j=1}^{\infty}$ with $\sum_{j=1}^{\infty} \mu_j < \infty$, we let the feature map as

$$\Phi(x) := \left(\sqrt{\mu_1}\phi_1(x), \sqrt{\mu_2}\phi_2(x), \cdots\right).$$

By constrcution

$$\|\Phi(x)\|^2 = \sum_{j=1}^{\infty} \mu_j \phi_j^2(x) \leq \sum_{j=1}^{\infty} \mu_j < \infty \Rightarrow \Phi(x) \in \ell^2(\mathbb{N}).$$

Therefore, we define $\mathcal{K}(x,z) = \langle \Phi(x), \Phi(z) \rangle_{\ell^2(\mathbb{N})} = \sum_{j=1}^{\infty} \mu_j \phi_j(x)\phi_j(z)$ is a PSD kernel.

Gaussian kernel

12.9 Choose $\mathcal{X} \subset \mathbb{R}^d$ and consider Gaussian kernel $\mathcal{K}(\boldsymbol{x}, \boldsymbol{z}) = exp\left(-\frac{\|\boldsymbol{x}-\boldsymbol{z}\|_2^2}{2\sigma^2}\right)$. In order to prove it is a PSD, we can first expand exp into polynomials and by first proving product of PSD is also a PSD, and then limit of summation of PSD is also a PSD. To be specific,

(1) Let $\mathcal{K}(x, z)$ is PSD, then the polynomials $P(\mathcal{K}(x, z))$ is also a PSD, where $P(x) = \sum_{i=0}^{n} a_i x^i$.

(2) We consider $\mathcal{K}(\boldsymbol{x}, \boldsymbol{z}) = \exp\left(-\frac{\|\boldsymbol{z}\|_2^2}{2\sigma^2}\right) \cdot \exp\left(\langle \boldsymbol{x}, \boldsymbol{z}\rangle/\sigma^2\right) \cdot \exp\left(-\frac{\|\boldsymbol{z}\|_2^2}{2\sigma^2}\right)$.

We only need to prove $\mathcal{K}(x, y) = \mathcal{K}_1(x, y)\mathcal{K}_2(x, y)$ is a PSD whenever $\mathcal{K}_1$ and $\mathcal{K}_2$ are PSDs. By Linear Algebra, we let $C = A^\top A$ and $D = B^\top B$ where $c_{ij} = \mathcal{K}_1(x_i, x_j)$ and $d_{ij} = \mathcal{K}_2(x_i, x_j)$. By defining $A = (a_1, \cdots, a_n)$ and $B = (b_1, \cdots, b_n)$, we have $c_{ij} = \sum_k a_{ik}a_{jk}$ and $d_{ij} = \sum_\ell b_{i\ell}b_{j\ell}$. By denoting $e_{ij} = \mathcal{K}_1(x_i, x_j)\mathcal{K}_2(x_i, x_j)$, we calculate

$$
\begin{aligned}
\boldsymbol{u}^\top E\boldsymbol{u} &= \sum_{i,j} u_i u_j e_{ij} = \sum_{i,j}\sum_{k,\ell} u_i u_j a_{ik} a_{jk} b_{i\ell} b_{j\ell} \\
&= \sum_{k,\ell}\left(\sum_i u_i a_{ik} b_{i\ell}\right)\left(\sum_j u_j a_{jk} b_{j\ell}\right) = \sum_{k,\ell}\left(\sum_i u_i a_{ik} b_{i\ell}\right)^2 \geq 0.
\end{aligned}
$$

Constructing from a PSD

We propose the RKHS constructed from a PSD has the following property

$$\langle f, \mathcal{K}(\cdot, \boldsymbol{x}) \rangle_{\mathcal{H}} = f(\boldsymbol{x}) \quad \forall f \in \mathcal{H},$$

which is known as the kernel reproducing property. Given functions of form $f(\cdot) = \sum_{j=1}^{n} \alpha_j \mathcal{K}(\cdot, x_j)$ and $\bar{f} = \sum_{k=1}^{\bar{n}} \bar{\alpha}_k \mathcal{K}(\cdot, \bar{x}_k)$, by the linearity of inner product, we have

$$\langle f, \bar{f} \rangle = \sum_{j=1}^{n} \sum_{k=1}^{\bar{n}} \alpha_j \bar{\alpha}_k \mathcal{K}(x_j, \bar{x}_k),$$

and moreover, this kind of inner product satisfies the reproducing property by

$$\langle f, \mathcal{K}(\cdot, x) \rangle = \sum_{j=1}^{n} \alpha_j \mathcal{K}(x_j, x) = f(x).$$

Explicit Version

Theorem 12.11

Given any PSD $\mathcal{K}$, there is a unique Hilbert space $\mathcal{H}$ in which the kernel satisfies the reproducing property. It is known as the reproducing kernel Hilbert space associated with $\mathcal{K}$.

Proof:

About the sensibility of inner product, we only need to prove $\|f\|_{\mathcal{H}}^2 = 0$ iff $f = 0$. We suppose that $\langle f, f\rangle_{\tilde{\mathcal{H}}} = \sum_{i,j=1}^n \alpha_i\alpha_j\mathcal{K}(x_i, x_j) = 0$, then by arbitrarily choosing $a \in \mathbb{R}$, we have

$$0 \leq \|a\mathcal{K}(\cdot, x) + \sum_{i=1}^n \alpha_i\mathcal{K}(\cdot, x_i)\|^2 = a^2\mathcal{K}(x, x) + 2a\sum_{i=1}^n \alpha_i\mathcal{K}(x, x_i).$$

Since $\mathcal{K}(x, x) \geq 0$ and $a \in \mathbb{R}$ is arbitrary, we have $f(x) = \sum_{i=1}^n \alpha_i\mathcal{K}(x, x_i) = 0$.

Proof

Then we need to make a complete inner product space, i.e. the Hilbert space. Assuming $\{f_n\}_{n=1}^{\infty}$ is a Cauchy sequence, then $\{f_n(x)\}_{i=1}^{\infty} \subset \mathbb{R}$ is a Cauchy sequence, so that we define $f(x) = \lim_{n\to\infty} f_n(x)$ and also $\|f\|_{\mathcal{H}} := \lim_{n\to\infty} \|f_n\|_{\tilde{\mathcal{H}}}$. To verify it is well-defined, we have to prove that when the Cauchy sequence $\{g_n\}_{n=1}^{\infty}$ in $\tilde{\mathcal{H}}$ such that $\lim_{n\to\infty} g_n(x) = 0 \quad \forall x \in \mathcal{X}$, we also have $\lim_{n\to\infty} \|g_n\| = 0$. Otherwise, there is a subsequence such that $\lim_{n\to\infty} \|g_n\|^2 = 2\epsilon > 0$, so that for m, n large enoug, we have $\|g_n\|^2 \geq \epsilon$ and $\|g_m\|^2 \geq \epsilon$ and also $\|g_n - g_m\| \leq \epsilon/2$. Then we write $g_m(\cdot) = \sum_{i=1}^{N_m} \alpha_i \mathcal{K}(\cdot, x_i)$. By reproducing rpoperty,

$$\langle g_m, g_n \rangle = \sum_{i=1}^{N_m} \alpha_i \langle \mathcal{K}(\cdot, x_i), g_n \rangle = \sum_{i=1}^{N_m} \alpha_i g_n(xi) \to 0.$$

By

$$\|g_n - g_m\|^2 = \|g_n\|^2 + \|g_m\|^2 - 2\langle g_n, g_m \rangle,$$

we get contradiction.

Proof

Finally, we prove the uniqueness. Suppose that $\mathbb{G}$ is another Hilbert space with $\mathcal{K}$ being its kernel. Since $\mathbb{G}$ is complete and closed under linear operations, we have $\mathcal{H} \subset \mathbb{G}$, so that $\mathbb{G} = \mathcal{H} \bigoplus \mathcal{H}^{\perp}$. Let $g \in \mathcal{H}^{\perp}$ and noting $\mathcal{K}(\cdot, x) \in \mathcal{H}$, then $g(x) = \langle \mathcal{K}(\cdot, x), g \rangle_{\mathbb{G}} = 0$. We conclude that $\mathcal{H}^{\perp} = \{0\}$, thus $\mathcal{H} = \mathbb{G}$.

Abstract Version

Observing that $f(x) = \langle f, \mathcal{K}(\cdot, x)\rangle$, we can view $\mathcal{K}(\cdot, x)$ as the evaluation function $L_x : \mathcal{H} \to \mathbb{R}$ that performs $f \mapsto f(x)$. By Riesz representation theorem, it means all evaluation functions in RKHS are bounded. A direct question is that how large the class of Hilbert space where the evaluation functions are bounded is?

Definition 12.12

A reproducing kernel Hilbert space $\mathcal{H}$ is a Hilbert space of real-valued functions on $\mathcal{X}$ such that for each $x \in \mathcal{X}$, the evaluation functional $L_x : \mathcal{H} \to \mathbb{R}$ is bounded, i.e., there exists some $M < \infty$ such that $|L_x(f) \leq M\|f\|$ for all $f \in \mathcal{H}$.

RKHS Construction

Theorem 12.13

Given and Hilbert spaec $\mathcal{H}$ in which the evaluation functionals are all bounded, there is a unique PSD kernel $\mathcal{K}$ that satisfies the reproducing property.

Proof:

By Riesz representation theorem, there exists some element $R_x \in \mathcal{H}$ such that $f(x) = L_x(f) = \langle f, R_x \rangle$ for all $f \in \mathcal{H}$. We define $\mathcal{K}$ via $\mathcal{K}(x, z) = \langle R_x, R_z \rangle$. We only need to show it is positive semidefinite. In fact

$$\boldsymbol{\alpha}^\top K \boldsymbol{\alpha} = \sum_{j,k=1}^{n} \alpha_j \alpha_k \mathcal{K}(x_j, x_k) = \left\langle \sum_{j=1}^{n} \alpha_j R_{x_j}, \sum_{j=1}^{n} \alpha_j R_{x_j} \right\rangle = \left\| \sum_{j=1}^{n} \alpha_j R_{x_j} \right\|_{\mathcal{H}}^2 \geq 0.$$

Proof

It remains to prove the reproducing property: by

$$\mathcal{K}(y, x) = \langle R_y, R_x \rangle = R_x(y)$$

, we can see that $\mathcal{K}(\cdot, x) = R_x(\cdot)$, then by definition

$$f(x) = \langle f, R_x \rangle = \langle f, \mathcal{K}(\cdot, x) \rangle,$$

which is the reproducing property.
Finally, if there exists another kernel $\tilde{\mathcal{K}}$ satisfying the properties above, we can see

$$\begin{aligned} \mathcal{K}(x, x') &= \langle \mathcal{K}(\cdot, x), \mathcal{K}(\cdot, x') \rangle = \langle \mathcal{K}(\cdot, x), \tilde{\mathcal{K}}(\cdot, x') \rangle \\ &= \langle \tilde{\mathcal{K}}(\cdot, x), \tilde{\mathcal{K}}(\cdot, x') \rangle = \tilde{\mathcal{K}}(x, x') \quad \forall x, x' \in \mathcal{X}. \end{aligned}$$

Examples

12.14 We let $\mathcal{K}(x,z) = \langle x, z\rangle$ and $f(x) = \sum_j \mathcal{K}(x, x_j)$, thus

$$f(x) = \langle f, \mathcal{K}(\cdot, x)\rangle = \left\langle \sum_j \alpha_j \mathcal{K}(\cdot, x_j), \mathcal{K}(\cdot, x) \right\rangle = \sum_j \alpha_j \mathcal{K}(x, x_j)$$

$$= \sum_j \alpha_j \langle x, x_j\rangle = \left\langle x, \sum_j \alpha_j x_j \right\rangle .$$

It means the evaluation functional has the form $z \mapsto \langle z, \sum_{i=1}^n \alpha_i x_i\rangle$, i.e., $f_\beta(\cdot) = \langle \cdot, \beta\rangle$ and the inner product in RKHS is formed by $\langle f_\beta, f_{\beta'}\rangle_{\mathcal{H}} = \langle \beta, \beta'\rangle$.

12.16 (A simple Sobolev space) Consider the functions

$$\mathbb{H}^1[0,1] := \{f\colon [0,1] \to \mathbb{R} | f(0) = 0, \text{and } f \text{ is absolutely continuous with } f' \in L^2[0,1]\}.$$

The inner product is defined as $\langle f, g \rangle := \int_0^2 f'(z)g'(z)dz$. In order to prve it is aN RKHS, we claim its representer of evaluation: $R_x(z) = \min\{x, z\} \Rightarrow R_x'(z) = \mathbb{I}_{[0,x]}(z)$. We can calculate

$$\langle f, R_x \rangle = \int_0^1 f'(z)R_x'(z)dz = \int_0^x f'(z)dz = f(z)$$

by absolutely continuous.
Also by the process of the proof of theorem 12.13, we know that the PSD $\mathcal{K}(x, z) = \langle R_x, R_z \rangle = \int_0^2 \mathbb{I}_{[0,x]}(z)\mathbb{I}_{[0,z]}(z)dz = \langle \mathbb{I}_{[0,x]}, \mathbb{I}_{[0,z]} \rangle_{L^2[0,1]}$, therefore providing a Gram representation, leading to positive semidefinite. We conclude that $\mathcal{K}(x, z) = \min\{x, z\}$ is the unique PSD kernel.

RMK

RMK: RKHS ensures that convergence of a sequence of functions in RKHS implies pointwise convergence, which means: if we let $f_n \to f^*$ in $\mathcal{H}$ norm, we have

$$|f_n(x) - f^*(x)| = |\langle f_n - f^*, R(\cdot, x)\rangle| = |L_x(f_n - f^*)| \leq \|L_x\| \|f_n - f^*\| \to 0.$$

12.15 ($L^2[0,1]$ is not an RKHS) Consider the sequence of functions $f_n(x) = x^n$ and since $\int_0^2 f_n^2(x)dx = \frac{1}{2n+1} \to 0$. Therefore, $\|f_n\|_{\mathcal{H}} \to 0$. However, $f_n(1) \equiv 1$, thus not pointwise convergent.

Overview

12.18 (PSD matrices) Let $\mathcal{X} = [d]$ be equipped with Hamming metric, i.e. $P(\{j\}) = 1$ be the counting measure on this discrete space. Define a PSD kernel $\mathcal{K} : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ with the matrix $K = (K_{ij}) = (\mathcal{K}(i,j))_{i,j=1}^{n}$. Define the integral operator as

$$T_{\mathcal{K}}(f)(x) = \int_{\mathcal{X}} \mathcal{K}(x,z) f(z) dP(z) = \sum_{z=1}^{d} \mathcal{K}(x,z) f(z).$$

By Linear Algebra Theory, we have

$$K = \sum_{j=1}^{d} \mu_j \mathbf{v}_j \mathbf{v}_j^\top.$$

Notations

Let $\mathbb{P}$ be a non-negative measure over a compact metric space $\mathcal{X}$, and consider the function class $L^2(\mathcal{X}; \mathbb{P})$ with the usual squared norm

$$\|f\|^2_{L(\mathcal{X};\mathbb{P})} = \int_{\mathcal{X}} f^2(x) d\mathbb{P}(x).$$

Given a PSD kernel, we define a linear operator

$$T_{\mathcal{K}}(f)(x) := \int_{\mathcal{X}} \mathcal{K}(x,z) f(z) d\mathbb{P}(z).$$

We assume that

$$\int_{\mathcal{X}\times\mathcal{X}} \mathcal{K}^2(x,z) d\mathbb{P}(x) d\mathbb{P}(z) < \infty, \tag{*}$$

which is squared integral, then we have

$$\|T_{\mathcal{K}}(f)\|^2_{L^2(\mathcal{X})} = \int_{\mathcal{X}} \left(\int_{\mathcal{X}} \mathcal{K}(x,y) d\mathbb{P}(x) \right)^2 d\mathbb{P}(y) \leq \|f\|^2_{L(\mathcal{X})} \int_{\mathcal{X}\times\mathcal{X}} \mathcal{K}^2(x,y) d\mathbb{P}(x) d\mathbb{P}(y),$$

which implies the operator $T_{\mathcal{K}}$ is a bounded operator on $L^2(\mathcal{X})$.

Mercer's Theorem

Theorem 12.20

Suppose that $\mathcal{X}$ is compact, the kernel function $\mathcal{K}$ is continuous and positive semidefinite, and satisfies the Hilbert-Schmidt condition*. Then there exists a sequence of eigenfunctions $(\phi_j)_{j=1}^{\infty}$ that form an orthonormal basis of $L^2(\mathcal{X}, ; \mathbb{P})$, and non-negative eigenvalues $(\mu_j)_{j=1}^{\infty}$ such that

$$T_{\mathcal{K}}(\phi_j) = \mu_j \phi_j.$$

Moreover, the kernel function has the expansion

$$\mathcal{K}(x, z) = \sum_{j=1}^{\infty} \mu_j \phi_j(x) \phi_j(z),$$

where the convergence of the infinite series holds absolutely and uniformly.

The original Mercer's theorem is related to the spectral of compact operators in advanced Functional Analysis.

Examples

We define a mapping $\Phi : \mathcal{X} \to \ell^2(\mathbb{N})$ via

$$x \mapsto \Phi(x) := \left(\sqrt{\mu_1}\phi_1(x), \sqrt{\mu_2}\phi_2(x), \cdots\right).$$

By construction, we have $\|\Phi(x)\|^2_{\ell^2(\mathbb{N})} = \sum_{j=1}^{\infty} \mu_j \phi_j^2(x) = \mathcal{K}(x, x) < \infty$, which indeeed $\Phi \in \ell^2(\mathbb{N})$. Moreover,

$$\langle \Phi(x), \Phi(z) \rangle = \sum_{j=1}^{\infty} \mu_j \phi_j(x) \phi_j(z) = \mathcal{K}(x, z) < \infty,$$

thus providing a PSD kernel.

12.22 $\mathcal{K}(x, z) = (1 + xz)^2$ over $[-1, 1]^2$, which is equipped with Lebesgue measure. Given a function $f : [-1, 1] \to \mathbb{R}$, we have

$$\int_{-1}^{1} \mathcal{K}(x, z) f(z) dz = \left(\int_{-1}^{1} f(z) dz\right) + \left(2 \int_{-1}^{1} z f(z) dz\right) x + \left(\int_{-1}^{1} z^2 f(z) dz\right) x^2.$$

So that we let the eigenfunctions be $f(x) = a_0 + a_1 x + a_2 x^2$.

Examples

We only need to solve the linear system

$$\begin{bmatrix} 2 & 0 & 2/3 \\ 0 & 4/3 & 0 \\ 2/3 & 0 & 2/5 \end{bmatrix} \begin{bmatrix} a_0 \\ a_1 \\ a_2 \end{bmatrix} = \mu \begin{bmatrix} a_0 \\ a_1 \\ a_2 \end{bmatrix}.$$

12.23 (Eigenfunctions for a first-order Sobolev space) The PSD takes the form $\mathcal{K}(x,z) = \min\{x,z\}$. We calculate $T_{\mathcal{K}}(\phi) = \mu\phi$, which means

$$\int_0^x z\phi(z)dz + \int_x^1 x\phi(z)dz = \mu\phi(x) \quad \forall x \in [0,1].$$

Then we take derivatives twice ontaining $\mu\phi''(x) + \phi(x) = 0$. By the definition of $\mathbb{H}^1[0,1]$, we know $\phi(0) = 0$, so that $\phi(x) = \sin(x/\sqrt{\mu})$. Taking $x = 1$ in the equation above to get $\int_0^1 z\phi(z)dz = \mu\phi(1)$, we deduce that

$$\phi_j(t) = \sin\frac{(2j-1)\pi t}{2} \quad \mu_j = \left(\frac{2}{(2j-1)\pi}\right)^2 \quad j = 1, 2, \cdots.$$

Corollary

Corollary 12.26

Consider a kernel satisfying the conditions of Mercer's Theorem with associated eigenfunctions $(\phi_j)_{j=1}^{\infty}$ and non-negative eigenvalues $(\mu_j)_{j=1}^{\infty}$. It induces the reproducing kernel Hilbert space

$$\mathcal{H} = \{f = \sum_{j=1}^{\infty} \beta_j \phi_j | \text{for some } (\beta_j)_{j=1}^{\infty} \in \ell^2(\mathbb{N}) \text{ with } \sum_{j=1}^{\infty} \frac{\beta_j^2}{\mu_j} < \infty\},$$

along with inner product

$$\langle f, g \rangle_{\mathcal{H}} := \sum_{j=1}^{\infty} \frac{\langle f, \phi_j \rangle \langle g, \phi_j \rangle}{\mu_j},$$

where $\langle \cdot, \cdot \rangle$ denotes the inner product in $L^2(\mathcal{X}; \mathbb{P})$.

Proof

RMK: This Cor shows that the RKHS associated with a Mercer kernel is isomorphic to an infinite-dimensional ellipsoid contained with $\ell^2(\mathbb{N})$ - namely

$$\mathcal{E} = \left\{ (\beta_j)_{j=1}^{\infty} \in \ell^2(\mathbb{N}) | \sum_{j=1}^{\infty} \frac{\beta_j^2}{\mu_j} \leq 1 \right\}.$$

Proof:

We only need to verify that $\mathcal{H}$ has the reproducing property with respect to the given kernel. By Mercer's theorem, we have $\mathcal{K}(\cdot, x) = \sum_{j=1}^{\infty} \mu_j \phi_j(\cdot)\phi_j(x)$, so that $\beta_j = \mu_j \phi_j(x)$. Moreover,

$$\sum_{j=1}^{\infty} \frac{\beta_j^2}{\mu_j} = \sum_{j=1}^{\infty} \mu_j \phi_j^2(x) = \mathcal{K}(x, x) < \infty,$$

so that $\mathcal{K}(\cdot, x) \in \mathcal{H}$.

Proof

Let us now verify the reproducing property. By the orthonormaility of ϕ_j, we have $\langle \mathcal{K}(\cdot, x), \phi_j \rangle = \mu_j \phi_j(x)$. Thus, for any $f \in \mathcal{H}$, we have

$$\langle f, \mathcal{K}(\cdot, x) \rangle = \sum_{j=1}^{\infty} \frac{\langle f, \phi_j \rangle \langle \mathcal{K}(\cdot, x), \phi_j \rangle}{\mu_j} = \sum_{j=1}^{\infty} \langle f, \phi_j \rangle \phi_j(x) = f(x),$$

where the last equality is by the orthonormality of $(\phi_j)_{j=1}^{\infty}$.

Thank you !