Reproducing Kernel Hilbert Spaces - Part I

Scribes: Yingxue Su, Qianfan Bai

In this section, we introduce reproducing kernel Hilbert spaces (RKHSs) and describe their relation to kernels following the material from [1, Ch. 4.2].

Definition 1. Let $X \neq \emptyset$ and H be a $\mathbb{K}-Hilbert$ function space over X, i.e., a Hilbert space that consists of functions with domain in X and range into $\mathbb{K}$.

i) A function $k : X \times X \to \mathbb{K}$ is called a *reproducing kernel* of H if $k(\cdot,x) \in H$ for all $x \in X$ and it satisfies the *reproducing property*

$$f(x) = \langle f, k(\cdot,x)\rangle_H$$

for all $f \in H$ and all $x \in X$.

ii) The space H is called a *reproducing kernel Hilbert space (RKHS)* over X if for all $x \in X$ the *Dirac functional* $\delta_x : H \to \mathbb{K}$ defined by

$$\delta_x(f) = f(x), \quad f \in H,$$

is continuous.

Remark 1. We observe that a RKHS is a space of functions, hence $L^2(d\mu)$ is not a RKHS.

If H is a RKHS, then norm convergence implies pointwise convergence. To show that this is the case, let $(f_n) \in H$ be such that $\|f_n - f\|_H \to 0$ as $n \to \infty$ with $f \in H$. It follows that for any $x \in X$ there is a constant c such that

$$|\delta_x(f_n) - \delta_x(f)| \leq c\|f - f_n\|_H.$$

Hence

$$\lim_{n\to\infty} f_n(x) = \lim_{n\to\infty} \delta_x(f_n) = \delta_x(f) = f(x).$$

Lemma 1. *(Reproducing kernels are kernels). Let H be a Hilbert function space over X that has a reproducing kernel k. Then H is a RKHS and H is also a feature*

Published: January 23, 2020

space of k, where the feature map $\Phi : X \to H$ is given by

$$\Phi(x) = k(\cdot, x), \;\; x \in X.$$

We call Φ the canonical feature map of k.

Proof. First we show that δ_x is continuous so that H is a RKHS. Since k is a reproducing kernel in H, for any $f \in H$,

$$|\delta_x(f)| = |f(x)| = |\langle f, k(\cdot,x)\rangle_H| \leq \|f\|\|k(\cdot,x)\|.$$

This shows that δ_x is continuous for any $x \in X$. Therefore, H is a *RKHS*.

Next , we show that H is a feature space of k with feature map Φ. For a fixed $x' \in X$, let $f = k(\cdot, x')$. Then, for any $x \in X$,

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

Therefore, H is a feature space of k with a feature map Φ.

We have just seen that every Hilbert function space with a reproducing kernel is a RKHS. The following theorem shows that, conversely, every RKHS has a (unique) reproducing kernel over X and that this kernel can be determined by the Dirac functionals δ_x, $x \in X$.

Theorem 1. *(Every RKHS has a unique reproducing kernel). Let H be a RKHS over X and H' be the dual space of H. Then $k : X \times X \to \mathbb{K}$ defined by*

$$k(x,x') = \langle \delta_x, \delta_{x'}\rangle_{H'}, \;\; x, x' \in X,$$

is the only reproducing kernel of H. Furthermore, if $(e_i)_{i\in I}$ is an orthonormal basis of H, then for all $x, x' \in X$, we have

$$k(x,x') = \sum_{i\in I} e_i(x)\overline{e_i(x')}.$$

Proof. First, we show that k is a reproducing kernel by showing that the reproducing property holds. By Riesz representation theorem, there exists an isometric antilinear isomorphisim $I : H' \to H$ that assigns to any $g' \in H'$ a representing element in H; that is

$$g'(f) = \langle f, Ig'\rangle,$$

for all $f \in H$, $g' \in H'$. In particular, for $g' = \delta_x \in H'$, $f = I\delta_{x'} \in H$, then

$$\langle I\delta_{x'}, I\delta_x\rangle_H = \delta_x(I\delta_{x'}).$$

With this observation, for all $x, x' \in X$,

$$k(x,x') \overset{\text{def}}{=} \langle \delta_x, \delta_{x'}\rangle_{H'} \overset{\text{Riesz}}{=} \langle I\delta_{x'}, I\delta_x\rangle_H = \delta_x(I\delta_{x'}) \overset{\text{def}}{=} I\delta_{x'}(x).$$

This shows that $k(\cdot,x') = I\delta_{x'}$ for all $x' \in X$. Hence,

$$f(x') \stackrel{\text{def}}{=} \delta_{x'}(f) = \langle f, I\delta_{x'}\rangle_H = \langle f, k(\cdot,x')\rangle$$

for all $x' \in X$, and this shows that k has the reproducing property.

To show uniqueness, let $\tilde{k}$ be an arbitrary reproducing kernel on H. For any $x' \in X$, given a basis $(e_i)_{i\in I} \in H$, we have

$$\tilde{k}(\cdot,x') = \sum_{i\in I}\langle \tilde{k}(\cdot,x'), e_i\rangle e_i = \sum_{i\in I}\overline{\langle e_i, \tilde{k}(\cdot,x')\rangle} e_i.$$

By the reproducing property of $\tilde{k}$, we have

$$\overline{\langle e_i, \tilde{k}(\cdot,x')\rangle} e_i = \overline{e_i(x')} e_i.$$

Therefore,

$$\tilde{k}(\cdot,x') = \overline{e_i(x')} e_i.$$

Since $\tilde{k}$ and $(e_i)_{i\in I}$ are arbitrarily chosen, we find $\tilde{k} = k$. Therefore, k is the only reproducing kernel of H.

Remark 2. We remark that the Hilbert space H in the proof does not have to be separable. If H is not separable, $(e_i)_{i\in I}$ is an uncountable set and the proof still works for this case. Recall that every Hilbert space H has an orthonormal basis. H is separable if and only if it admits a *countable* orthonarmal basis.

Theorem 1 shows that a RKHS uniquely determines its reproducing kernel; this reproducing kernel is actually a kernel by Lemma 1. The following theorem now shows that, conversely, every kernel has a unique RKHS. Consequently, we have a one-to-one relation between a kernel and a RKHS.

Theorem 2. *Let $X \neq \emptyset$ and k be a kernel over X with feature space H_0 and feature map $\Phi_0 : X \to H_0$. Then*

$$H := \{f : X \to \mathbb{K} \mid \exists w \in H_0 \;\; with \;\; f(x) = \langle w, \Phi_0(x)\rangle_{H_0} \;\; for \;\; all \;\; x \in X\}$$

equipped with the norm

$$\|f\|_H := \inf\{\|w\|_{H_0} : \; w \in H_0 \;\; with \;\; f = \langle w, \Phi_0(\cdot)\rangle_{H_0}\}$$

is the only RKHS for which k is a reproducing kernel.

Proof. First, we show that H is a Hilbert space. We observe that H is a vector space of functions from $X \to \mathbb{K}$. Define $V : H_0 \to H$ by $Vw = \langle w, \Phi_0(\cdot)\rangle_{H_0}$ for $w \in H_0$. By this notation, for any $f \in H$, we write

$$\|f\|_H = \inf_{w\in V^{-1}(f)} \|w\|_{H_0}.$$

To show that $\|\cdot\|_H$ is a Hilbert space norm, let $(w_n) \subset \ker V = \{w \in H_0 | Vw = 0\}$ with the property that $\lim_{n\to\infty} w_n = w$. Then $\langle w, \Phi(x)\rangle_{H_0} = \lim_{n\to\infty}\langle w_n, \Phi(x)\rangle = 0$ for any $x \in X$. This shows that $w \in \ker V$ and, hence, $\ker V$ is closed. Denoting $\tilde{H} = (\ker V)^{\perp}$, we can write $H_0 = \ker V \oplus \tilde{H}$. Then, by construction, the restriction $V_{|\tilde{H}} : \tilde{H} \to H$ of V to $\tilde{H}$ is injective. We want to show that $V_{|\tilde{H}}$ is also surjective. For any $f \in H$, there exists $w \in H_0$ with $f(x) = \langle w, \Phi_0(x)\rangle_{H_0} = Vw(x)$. Rewrite $w = w_0 + \tilde{w}$ with $w_0 \in H_0$ and $\tilde{w} \in \tilde{H}$. Then $f = V(w_0 + \tilde{w}) = V\tilde{w} = V_{|\tilde{H}}\tilde{w}$. This shows that $V_{|\tilde{H}}$ is surjective and, thus, $V_{|\tilde{H}}$ is also bijective.

Let $(V_{|\tilde{H}})^{-1}$ be the inverse operator of $V_{|\tilde{H}}$. Then we have

$$\|f\|_H^2 = \inf_{w\in V^{-1}(\{f\})} \|w\|_{H_0}^2 = \inf_{w_0\in\ker V, \tilde{w}\in\tilde{H}, w_0+\tilde{w}\in V^{-1}(\{f\})} \|w_0 + \tilde{w}\|_{H_0}^2$$

$$= \inf_{w_0\in\ker V, \tilde{w}\in\tilde{H}, w_0+\tilde{w}\in V^{-1}(\{f\})} \|w_0\|_{H_0}^2 + \|\tilde{w}\|_{H_0}^2 = \|(V_{|\tilde{H}})^{-1} f\|_{\tilde{H}}^2.$$

Since $\tilde{H}$ is a Hilbert space norm, then $\|\cdot\|_H$ is also a Hilbert space norm. Hence we have shown that $V_{|\tilde{H}}$ is an isometric isomorphism from $\tilde{H}$ to H.

To show that k is a reproducing kernel of H, note that, for any $x \in X$,

$$k(\cdot, x) = \langle \Phi_0(x), \Phi_0(\cdot)\rangle = V\Phi_0(x).$$

Since $\langle w, \Phi_0(x)\rangle = Vw(x) = 0$ for any $w \in \ker V$, then $\Phi_0(x) \in (\ker V)^{\perp} = \tilde{H}$. Since $V_{|\tilde{H}} : \tilde{H} \to H$ is isometric, we obtain that

$$f(x) = \langle (V_{|\tilde{H}})^{-1} f, \Phi_0(x)\rangle_{H_0} = \langle f, V_{|\tilde{H}}\Phi_0(x)\rangle_H = \langle f, k(\cdot, x)\rangle_H$$

for all $f \in H, x \in X$, which is the reproducing property of k. Therefore, H is a RKHS by Lemma 1.

To prove the uniqueness, we first show that the set

$$H_{pre} := \{\sum_{i=1}^{n} \alpha_i k(\cdot, x_i) : n \in \mathbb{N}, \alpha_1, ..., \alpha_n \in \mathbb{K}, x_1, ..., x_n \in X\}$$

is dense in any RKHS $\hat{H}$ with k as the reproducing kernel. By the definition of reproducing kernel, we observe that $k(\cdot, x) \in \hat{H}$ for all $x \in X$. Hence, $H_{pre} \subset \hat{H}$. Now we suppose that H_{pre} is not dense in $\hat{H}$. Then, $(H_{pre})^{\perp} \neq \{0\}$. Therefore, there exists a function $g \in (H_{pre})^{\perp}$ and a $x \in X$ with $g(x) \neq 0$. Since $g \in (H_{pre})^{\perp}$ and $k(\cdot, x) \in \hat{H}$, $\langle g, k(\cdot, x)\rangle = 0$. By the reproducing property of k, $\langle g, k(\cdot, x)\rangle = g(x) \neq 0$. This is a contradiction. Therefore, H_{pre} is dense in any RHKS.

For any $f := \sum_{i=1}^{n} \alpha_i k(\cdot, x_i) \in H_{pre}$, by the reproducing property, notice that

$$\|f\|_{\hat{H}}^2 = \sum_{i=1}^{n}\sum_{j=1}^{n} \alpha_i \overline{\alpha_j} \langle k(\cdot, x_i), k(\cdot, x_j)\rangle_{\hat{H}} = \sum_{i=1}^{n}\sum_{j=1}^{n} \alpha_i \overline{\alpha_j} k(x_j, x_i).$$

Let us now prove that k has only one RKHS. Let H_1 and H_2 be two *RKHSs* of k. We just proved that H_{pre} is dense in both H_1 and H_2 and that the norms of H_1 and H_2 coincide on H_{pre}. Choose $f \in H_1$. There exists a sequence $(f_n) \subset H_{pre}$ with $\|f_n - f\|_{H_1} \to 0$. Since $H_{pre} \subset H_2$, the sequence (f_n) is also contained in H_2, and since the norms of H_1 and H_2 coincide on H_{pre}, the sequence (f_n) is a Cauchy sequence in H_2. Therefore, there exists a $g \in H_2$ with $\|f_n - g\|_{H_2} \to 0$. Since the convergence with respect to a RHKS norm implies pointwise convergence, we then find $f(x) = g(x)$ for all $x \in X$, i.e., we have shown $f \in H_2$. Furthermore, $\|f_n - f\|_{H_1} \to 0$ and $\|f_n - f\|_{H_2} \to 0$ imply

$$\|f\|_{H_1} = \lim_{n\to\infty} \|f_n\|_{H_1} = \lim_{n\to\infty} \|f_n\|_{H_{pre}} = \lim_{n\to\infty} \|f_n\|_{H_2} = \|f\|_{H_2}.$$

Therefore, H_1 is isometrically included in H_2. Similarly, we can prove that $H_2 \subset H_1$. So the reproducing kernel k has a unique RKHS.

References

1. Ingo Steinwart; Andreas Christmann. *Support Vector Machines*. Springer, 2008.