Reproducing Kernel Hilbert Spaces - Part II

Scribes: Malcolm Davies and Ahmed Abouserie

1 Properties of Reproducing Kernel Hilbert Spaces

We have seen that there is a one-to-one correspondence between kernels and reproducing kernel Hilbert spaces (RKHS). In this section, we study properties of functions in a RKHS that are inherited from the reproducing kernel function. Conversely, properties of a reproducing kernel such as boundedness, measurability and integrability are characterized in terms of analogous properties of all functions in the associated RKHS.

1.1 Boundedness

We recall that, for a function f on a topological space Z, the uniform norm of f is given by $\|f\|_b = \sup_{z \in Z} |f(z)|$ (see [1, Chapters 4.2]). For kernel functions K on $X \times X$, we introduce the norm $\|\cdot\|_\infty$ (not to be confused with the essential supremum) that is given by:

$$\|K\|_\infty := \sup_{x \in X} \sqrt{K(x,x)}. \tag{1}$$

Note that in general, $\|K\|_b \neq \|K\|_\infty$ for any such K. However, it is natural to inquire whether there exists a correlation between the two norms in terms of characterizing the boundedness of K. The following lemma is a consequence of the reproducing property.

Published: February 21, 2020

Lemma 1. *Let $K : X \times X \to \mathbb{K}$ be a kernel on a reproducing kernel Hilbert space $\mathscr{H}$ with the feature map $\Phi : X \to \mathscr{H}$. Then K is bounded iff*

$$\|K\|_\infty := \sup_{x\in X} \sqrt{K(x,x)} < \infty.$$

Proof. Recall that the reproducing property of $\mathscr{H}$ implies that for any $x, x' \in X$,

$$K(x,x') = \left\langle \Phi(x'), \Phi(x) \right\rangle = \left\langle K(\cdot,x), K(\cdot,x') \right\rangle. \tag{2}$$

Hence, by the Cauchy-Schwarz inequality, we get that

$$\begin{aligned} |K(x,x')|^2 &= |\left\langle K(\cdot,x), K(\cdot,x') \right\rangle|^2 \\ &\leq \|K(\cdot,x)\|^2_{\mathscr{H}} \cdot \|K(\cdot,x')\|^2_{\mathscr{H}} \\ &= K(x,x) \cdot K(x',x'). \end{aligned}$$

Using properties of the supremum, along with the continuity and monotone increasing nature of the square root function on $[0,\infty)$, we get that

$$\begin{aligned} \|K\|_b &:= \sup_{(x,x')\in X\times X} |K(x,x')| \\ &\leq \sup_{x,x'\in X} \sqrt{K(x,x)} \cdot \sqrt{K(x',x')} \\ &= \sup_{x\in X} \sqrt{K(x,x)} \cdot \sup_{x'\in X} \sqrt{K(x',x')} \\ &= \sup_{x\in X} K(x,x). \end{aligned}$$

On the other hand, we have that

$$\sup_{(x,x')\in X\times X} |K(x,x')| \geq \sup_{(x,x)\in X\times X} K(x,x)$$

since $\{(x,x)|x \in X\} \subseteq \{(x,x')|x,x' \in X\}$ by choosing $x' = x$. Thus, equality holds and we get that K is bounded iff

$$\|K\|_b := \sup_{(x,x')\in X\times X} |K(x,x')| = \sup_{x\in X} K(x,x) < \infty.$$

Again, by the monotone continuity of the square root and the fact $K(x,x) \geq 0$, this is equivalent to boundedness with respect to our more useful norm definition from [2, Chapter 4.3]. Thus, K is bounded iff $\|K\|_b < \infty$ iff

$$\|K\|_\infty := \sup_{x\in X} \sqrt{K(x,x)} < \infty$$

Thus, we conclude that either of the norms, $\|.\|_b$ and $\|.\|_\infty$, can be used to determine boundedness of K. □

Next, we relate the boundedness of K to the boundedness of its *feature map* Φ.

Lemma 2. *Let $K : X \times X \to \mathbb{K}$ be a kernel on a reproducing kernel Hilbert space $\mathscr{H}$ with feature space $\mathscr{H}_0$ and feature map $\Phi : X \to \mathscr{H}_0$. Then K is bounded iff Φ is bounded.*

Proof. Since $\Phi : X \to \mathscr{H}_0$ is a feature map for K, an application of (2) gives that

$$\|\Phi(x)\|^2_{\mathscr{H}_0} = \langle \Phi(x), \Phi(x)\rangle_{\mathscr{H}_0} = \langle K(\cdot,x), K(\cdot,x)\rangle_{\mathscr{H}} = K(x,x).$$

Taking the supremum over X on both sides gives $\|\Phi\|_b^2 = \|K\|_\infty^2$. Thus, $\|\Phi\|_b^2 < \infty$ iff $\|K\|_\infty^2 < \infty$ which is our desired characterization of boundedness.

□

Now we are able to characterize the boundedness of the reproducing kernel in terms of the feature space elements $f \in \mathscr{H}$.

Lemma 3. *Let $K : X \times X \to \mathbb{K}$ be a kernel on X with a reproducing kernel Hilbert space $\mathscr{H}$. Then K is bounded iff every $f \in \mathscr{H}$ is bounded (as a function).*

Moreover, in this case the induction map $(id) : \mathscr{H} \to l^\infty(X)$ is continuous, with

$$\|(id) : \mathscr{H} \to l^\infty(X)\| = \|K\|_\infty.$$

Proof. ($\implies$): Assume K is bounded. By virtue of the properties of RKHS's for $\mathscr{H}$ and the Cauchy-Schwarz inequality, we have for all $x \in X$ and $f \in \mathscr{H}$,

$$|f(x)|^2 = |\langle f, K(\cdot,x)\rangle_{\mathscr{H}}|^2 \leq \|f\|^2_{\mathscr{H}}\, K(x,x).$$

Taking the supremum over X gives

$$\|f\|_b \leq \|f\|_{\mathscr{H}}\, \|K\|_\infty.$$

Since K is assumed bounded and $f \in \mathscr{H} \implies \|f\|_{\mathscr{H}} < \infty$, then we have $\|f\|_b < \infty$, showing boundedness.

This also shows that $(id) : \mathscr{H} \to l^\infty(X)$ is well-defined, and that

$$\|(id) : \mathscr{H} \to l^\infty(X)\| \leq \|K\|_\infty. \tag{3}$$

($\impliedby$): Conversely, if every $f \in \mathscr{H}$ is bounded, we get that the inclusion $(id) : \mathscr{H} \to l^\infty(X)$ is well-defined. We also observe that (id) is clearly a linear map since

$$(id)(\alpha f + g)(x) = (\alpha f + g)(x) = \alpha f(x) + g(x) = \alpha (id)(f)(x) + (id)(g)(x)$$

for any $\alpha \in \mathbb{K}$ and $f, g \in \mathscr{H}$.

We will use the closed graph theorem to prove that (id) is bounded. To show that (id) has a closed graph, let $(f_n)_{n=1}^\infty$ be a sequence of functions in $\mathscr{H}$ such that

$$\lim_{n\to\infty} \|f_n - f\|_{\mathscr{H}} = 0 \tag{4}$$

$$\text{and} \quad \lim_{n\to\infty} \|id(f_n) - g\|_\infty = \lim_{n\to\infty} \|f_n - g\|_\infty = 0 \tag{5}$$

for some $f \in \mathscr{H}$ and $g \in l^\infty(X)$. Then, we have that for any $x \in X$,

$$\begin{aligned}
\lim_{n\to\infty} |f_n(x) - f(x)|^2 &= \lim_{n\to\infty} |\langle f_n, K(\cdot,x)\rangle_{\mathscr{H}} - \langle f, K(\cdot,x)\rangle_{\mathscr{H}}|^2 && \text{(reproducing kernel property)} \\
&= \lim_{n\to\infty} |\langle f_n - f, K(\cdot,x)\rangle_{\mathscr{H}}|^2 && \text{(inner product property)} \\
&\le \lim_{n\to\infty} \|f_n - f\|_{\mathscr{H}}^2 \|K(\cdot,x)\|_{\mathscr{H}}^2 && \text{(Cauchy-Schwarz inequality)} \\
&= 0. && \text{(by (5))}
\end{aligned}$$

This relation holds iff f_n converges to f pointwise as $n \to \infty$. Also, since $|f_n(x) - g(x)| \le \|f_n - g\|_\infty$ for any $x \in X$, line (5) implies that $\lim_{n\to\infty} |f_n(x) - g(x)| = 0$ for every $x \in X$, which means that $f(x) = g(x)$ for all $x \in X$ iff $f = (id)(f) = g$. Therefore, we conclude that $(id) : \mathscr{H} \to l^\infty(X)$ has a closed graph and hence is bounded. Since (id) is linear, this means that it is also continuous.

Finally, for any $x \in X$ we have that

$$\begin{aligned}
|K(x,x)| \le \|K(\cdot,x)\|_\infty &\le \|(id) : \mathscr{H} \to l^\infty(X)\| \|K(\cdot,x)\|_{\mathscr{H}} \\
&= \|(id) : \mathscr{H} \to l^\infty(X)\| \sqrt{(K(x,x)}
\end{aligned}$$

which implies that $\sqrt{K(x,x)} \le \|(id) : \mathscr{H} \to l^\infty(X)\|$. Since this holds for every $x \in X$, taking the sup over X on both sides gives that

$$\|K\|_\infty \le \|(id) : \mathscr{H} \to l^\infty(X)\|. \tag{6}$$

By Lemma 1, this shows that K is bounded. In addition, inequalities (3) and (6) give that $\|K\|_\infty \le \|(id) : \mathscr{H} \to l^\infty(X)\|$. □

1.2 Measurability

Here, we focus on kernels over a measurable space, (X, μ). We start with the following characterization of the measurability of a kernel K in terms of the measurability of the functions in the associated RKHS.

Lemma 4. *Let (X, μ) be a measurable space and K be a kernel on X with reproducing kernel Hilbert space $\mathscr{H}$. Every $f \in \mathscr{H}$ is μ-measurable iff the restricted kernel function $K(\cdot, x') : X \to \mathbb{R}$ is μ-measurable for all $x' \in X$.*

If, in addition, we know that the *feature space* $\mathscr{H}$ is separable, then we are able to relate the measurability of $K(\cdot,x)$ to that of its *feature map* Φ. The next lemma gives sufficient conditions for the measurability of the Φ and the measurability of the reproducing kernel.

Lemma 5. *Let (X,μ) be a measurable space and K be a kernel on X with reproducing kernel Hilbert space $\mathscr{H}$ such that the restricted kernel function $K(\cdot,x') : X \to \mathbb{R}$ is μ-measurable for all $x' \in X$. If $\mathscr{H}$ is separable, then*

(i) the canonical feature map $\Phi : X \to \mathscr{H}$ is μ-measurable,
(ii) the full kernel $K : X \times X \to \mathbb{R}$ is $\mu \times \mu$-measurable on the product space $X \times X$.

Proof. (sketch):

1. Consider an element of the dual space of bounded linear functionals $w \in \mathscr{H}'$. By the Riesz representation theorem, there exists $f \in \mathscr{H}$ such that for any $x \in X$

$$w(\Phi(x)) = \langle w, \Phi(x) \rangle_{\mathscr{H}',\mathscr{H}} = \langle f, \Phi(x) \rangle_{\mathscr{H}} = f(x). \tag{7}$$

   Hence, $w(\Phi(\cdot))$ is measurable by the above lemma (4).

2. Recall the statement of Petti's measurability theorem [2, A.5.19]:
   Let E be a Banach space and $(\Omega,\mathscr{A})$ be a measurable space. Then $f : \Omega \to E$ is an E-valued measurable function iff the following two conditions are satisfied:

   (i) f is weakly measurable (i.e. $\langle x', f \rangle : \Omega \to \mathbb{R}$ is measurable for all $x \in E'$).
   (ii)$f(\Omega)$ is a separable subset of E

   Hence, (7) and the seperability of $\mathscr{H}$ imply that Φ is μ-measurable.

3. The second assertion now follows from $K(x,x') = \langle \Phi(x), \Phi(x') \rangle$ and the fact that the inner product is continuous (Continuity implies measurability on metric or topological spaces). □

1.3 Integrability

Finally, we investigate the integrability of measurable kernels, K, over a measurable space, (X,μ). We start by stating some useful properties of L^p integral operators of kernels.

Theorem 1. *Let (X,μ) be a measurable space, μ be a σ-finite measure on X, and $\mathscr{H}$ be a separable RKHS over X with measurable kernel $K : X \times X \to \mathbb{R}$. If there exists $p \in [1,\infty)$ such that*

$$\|K\|_{L^p} := \left(\int_X K(x,x)^{p/2} \, d\mu(x) \right)^{1/p} < \infty, \tag{8}$$

then the following holds:

(i) $\mathscr{H}$ consists of $L^p(\mu)$-integrable functions.
(ii) The inclusion map $(id): \mathscr{H} \to L^p(\mu)$ is continuous.
(iii) The adjoint of the inclusion map exists. It is the operator $S_K : L^{p'} \to \mathscr{H}$ given by

$$S_K g(x) = \int_X K(x,x')g(x')\,d\mu(x') \tag{9}$$

for $g \in L^{p'}$, $x \in X$, and conjugate exponents $\frac{1}{p} + \frac{1}{p'} = 1$.

Remark: The L^p norm here is not the standard one, rather it takes the L^p norm of the square root of the input K. This is analogous to the special ∞-norm for kernels developed above.

Proof. (i),(ii) Fix $f \in \mathscr{H}$. Then since $\|K(\cdot,x)\|_{\mathscr{H}} = \sqrt{K(x,x)}$ we have

$$\begin{aligned}
\|f\|_{L^p}^p &= \int_X |f(x)|^p\,d\mu(x) \\
&= \int_X |\langle f, K(\cdot,x)\rangle|^p\,d\mu(x) \\
&\le \|f\|_{\mathscr{H}}^p \int_X (K(x,x))^{p/2}\,d\mu(x) \\
&= \|f\|_{\mathscr{H}}^p\,\|K\|_{L^p}^p
\end{aligned}$$

This shows that $f \in L^p(\mu)$ and that $(id): \mathscr{H} \to L^p(\mu)$ is continuous with

$$\|(id): \mathscr{H} \to L^p(\mu)\| \le \|K\|_{L^p}.$$

(iii) For $g \in L^{p'}$ we have that

$$\begin{aligned}
\int_X |K(x,x')g(x')|\,d\mu(x') &\le \sqrt{K(x,x)} \int_X \sqrt{K(x',x')}\,|g(x')|\,d\mu(x') &(10)\\
&\le \sqrt{K(x,x)}\|K\|_{L^p}\|g\|_{L^{p'}} &(11)
\end{aligned}$$

Here, line (10) follows from the Schwartz inequality and the fact (from the proof of Lemma 1) that $|K(x,x')|^2 \le K(x,x)K(x',x')$. Line (11) is due to Hölder's inequality. Altogether, this shows the integrability of $K(x,x')g(x')$ and thus the existence (as a number in $\mathbb{R}$) of $S_K g(x)$ for all $x \in X$. Finally, we have

$$\begin{aligned}
S_K g(x) &= \int_X \left\langle \Phi(x'),\ \Phi(x)\right\rangle_{\mathscr{H}} g(x')\,d\mu(x') \\
&= \left\langle \int_X g(x')\Phi(x')\,d\mu(x'),\ \Phi(x) \right\rangle_{\mathscr{H}}.
\end{aligned}$$

This shows that $S_K g := \bar{g} = \int_X g(x')\Phi(x')\,d\mu(x') \in \mathscr{H}$. □

Remark: (Topological properties of $\mathscr{H}$ in relation to the adjoint S_K)
Under the conditions of Theorem 1, using the fact that a bounded linear operator has

a dense image if and only if its adjoint is injective, one can also derive the following properties for the feature space $\mathscr{H}$ in terms of the adjoint map S_K

1. $\mathscr{H}$ is dense in L^p iff the adjoint operator $S_K : L^{p'} \to \mathscr{H}$ is injective.
2. The adjoint $S_K : L^{p'} \to \mathscr{H}$ has a dense image $S_K(L^{p'})$ iff the inclusion $(id) : \mathscr{H} \to L^p$ is injective.

We conclude by stating some special properties of L^2 integral kernel operators over a measurable space (X, μ) for which there exists a σ-finite measure μ.

Theorem 2. *Let (X, μ) be a measurable space, μ be a σ-finite measure on X, and $\mathscr{H}$ be a separable RKHS over X with measurable kernel $K : X \times X \to \mathbb{R}$ such that*

$$\|K\|_{L^2} = \left(\int_X K(x,x)\, d\mu(x) \right)^{1/2} < \infty, \tag{12}$$

then

(i) the Hilbert-Schmidt (adjoint) operator $S_K : L^2 \to \mathscr{H}$ given by

$$S_K g(x) = \int_X K(x,x') g(x')\, d\mu(x')$$

exists for $g \in L^2$, $x \in X$.

(ii) the Hilbert-Schmidt (adjoint) operator has

$$\|S_K\|_{HS} = \|K\|_{L^2} \tag{13}$$

where we define $\|S\|_{HS}^2 := \sum_i \|Se_i\|_{L^2}^2$ for general operator S and orthonormal basis $\{e_i\} \subseteq \mathscr{H}$.

(iii) The integral operator $T_K = S_K^* S_K : L^2(\mu) \to L^2(\mu)$ is compact, positive, and self-adjoint.

Proof. (i) Recall from the proof of Theorem 1 that for $f \in \mathscr{H}$, $g \in L^2$, and $(id) : \mathscr{H} \to L^2$,

$$\begin{aligned} \langle g, (id) f \rangle_{L^2} &= \int_X g(x) \, \langle f, K(\cdot, x) \rangle_{\mathscr{H}} \, d\mu \\ &= \left\langle f, \int_X g(x) K(\cdot, x)\, d\mu(x) \right\rangle_{\mathscr{H}} \\ &= \langle f, S_K g \rangle_{\mathscr{H}} \end{aligned}$$

which gives that $\bar{g} := S_K g \in \mathscr{H}$.

(ii) Let $\{e_i\} \subseteq \mathscr{H}$ be an orthonormal basis. Then

$$\begin{aligned}
\|S_K\|_{HS}^2 &= \sum_{i=1}^{\infty} \|S_K^* e_i\|_{L^2}^2 \\
&= \int_X \sum_{i=1}^{\infty} |S_K^* e_i(x)|^2 \, d\mu(x) \\
&= \int_X \sum_{i=1}^{\infty} |e_i(x)|^2 \, d\mu(x) \\
&\overset{*}{=} \int_X K(x,x)^2 \, d\mu(x) \\
&= \|K\|_{L^2}^2
\end{aligned}$$

where the equality (*) follows from a previous result that $K(x,x') = \sum_j e_i(x)\overline{e_i(x')}$. Since $\|S_K^*\|_{HS}^2 = \|K\|_{L^2} < \infty$, this shows S_K^* is HS-norm bounded, and $S_K^* \in HS$. Consequently, its adjoint S_K is Hilbert-Schmidt as well.

(iii) The properties of the integral operator T_K follow directly from its definition and the properties of S_k. $\square$

Note: Here, we have that $S_K^* = (id) : \mathscr{H} \to L^2(\mu)$, and so $T_K := S_K^* S_K = (id) S_K^*$. However, it does *not* follow that $T_K = S_K$. The critical point is that $L^2(\mu)$ is not a space of functions, but rather a space of equivalence classes of functions. Hence, $S_K g(x) \in \mathscr{H}$ is defined for $x \in X$ but $T_K f(x) \in L^2(\mu)$ is not.

References

1. Gerald B. Folland. *Real Analysis: Modern Techniques and Their Methods*. Wiley, 1999.
2. Ingo Steinwart and Andreas Christmann. *Support vector machines*. Springer Science & Business Media, 2008.