.

NOTES ON REPRODUCING KERNEL HILBERT SPACE

XIN LU TAN

Given a linear operator A on a vector space $\mathcal{H}$, we have the following relationship:

$$\mathrm{Ker}A = (\mathrm{Ran}A^*)^\perp, \qquad \mathrm{Ker}A^* = (\mathrm{Ran}A)^\perp,$$
$$(\mathrm{Ker}A)^\perp = \overline{\mathrm{Ran}A^*}, \qquad (\mathrm{Ker}A^*)^\perp = \overline{\mathrm{Ran}A}.$$

Let $A_g(f) = \langle f, g\rangle_{\mathcal{H}}$, then A_g is a linear functionals, i.e. $A_g(c_1 f_1 + c_2 f_2) = c_1 A_g(f_1) + c_2 A_g(f_2)$. Moreover, A_g is bounded, $|A_g(f)| \leq \|A_g\|\|f\|_{\mathcal{H}} = \|f\|_{\mathcal{H}}\|g\|_{\mathcal{H}}$.

Theorem 1 (Riesz Representation Theorem)**.** *In a Hilbert space $\mathcal{H}$, for every bounded linear functionals $A : \mathcal{H} \longrightarrow \mathbb{R}$, there exists a unique $g \in \mathcal{H}$ such that*

$$Af = \langle f, g\rangle_{\mathcal{H}}$$

An evaluation functional over a Hilbert space of functions $\mathcal{H}$ is a linear functional $\delta_x : \mathcal{H} \longrightarrow \mathbb{R}$ that evaluates each function in the space at the point x, or $\delta_x(f) = f(x)$. Evaluation functional is always linear: For $f_1, f_2 \in \mathcal{H}$, and $c_1, c_2 \in \mathbb{R}$, $\delta_x(c_1 f_1 + c_2 f_2) = (c_1 f_1 + c_2 f_2)(x) = c_1 f_1(x) + c_2 f_2(x) = c_1 \delta_x(f_1) + c_2 \delta_x(f_2)$.

Definition 1. *A Hilbert space $\mathcal{H}$ is a reproducing kernel Hilbert space if the evaluation functionals are bounded (equivalently, continuous), i.e. there exists a positive constant M such that*

$$|\delta_x(f)| = |f(x)| \leq M\|f\|_{\mathcal{H}} \qquad \forall f \in \mathcal{H}.$$

In a reproducing kernel Hilbert space, norm convergence implies pointwise convergence.

Theorem 2. *If $\lim_{n\to\infty} \|f_n - f\|_{\mathcal{H}} = 0$, then $\lim_{n\to\infty} f_n(x) = f(x)$, for all $x \in \mathcal{X}$.*

Proof. $|f_n(x) - f(x)| = |\delta_x(f_n) - \delta_x(f)| = |\delta_x(f_n - f)| \leq \|\delta_x\|\|f_n - f\|_{\mathcal{H}}$. □

We will discuss three distinct concepts:

(1) reproducing kernel
(2) kernel
(3) positive definite function

and then show that they are all equivalent.

Definition 2 (Reproducing Kernel)**.** *Let $\mathcal{H}$ be a Hilbert space of functions $f : \mathcal{X} \longrightarrow \mathbb{R}$ defined on a non-empty set $\mathcal{X}$. A function $k : \mathcal{X} \times \mathcal{X} \longrightarrow \mathbb{R}$ is called a reproducing kernel of $\mathcal{H}$ if it satisfies*

(1) $\forall x \in \mathcal{X}, k_x = k(x, \cdot) \in \mathcal{H}$,
(2) $\forall x \in \mathcal{X}, \forall f \in \mathcal{H}, \langle f, k_x\rangle_{\mathcal{H}} = f(x)$ *(the reproducing property).*

In particular, for any $x, y \in \mathcal{X}$, $k(x, y) = \langle k_x, k_y\rangle_{\mathcal{H}} = \langle k_y, k_x\rangle_{\mathcal{H}} = k(y, x)$.

Theorem 3. *A Hilbert space $\mathcal{H}$ is a reproducing kernel Hilbert space if and only if it has a reproducing kernel.*

X. L. Tan: The Wharton School, Department of Statistics, University of Pennsylvania, Philadelphia, PA 19104. Email address: xtan@wharton.upenn.edu.

Proof. If $\mathcal{H}$ is a reproducing kernel Hilbert space, then by definition all the evaluation functionals are bounded. By the Riesz Representation Theorem, for each $x \in \mathcal{X}$, there exists a unique representer $k_x \in \mathcal{H}$ of $\mathcal{F}_x$ such that $\mathcal{F}_x(f) = f(x) = \langle k_x, f\rangle_{\mathcal{H}}$. So for each $x, y \in \mathcal{H}$, we define $k(x,y) = k_y(x) = \langle k_x, k_y\rangle_{\mathcal{H}}$. By symmetry of the inner product $\langle \cdot,\cdot\rangle_{\mathcal{H}}$ we have $k(x,y) = k(y,x)$. We immediately see that $\forall x \in \mathcal{X}$, the function $k_x = k(x,\cdot) \in \mathcal{H}$. Also, $\forall f \in \mathcal{H}$, $\langle f, k(x,\cdot)\rangle_{\mathcal{H}} = \langle f, k_x\rangle_{\mathcal{H}} = f(x)$. So $k(x,y)$ is the reproducing kernel of $\mathcal{H}$. On the other hand, given a space of functions $\mathcal{H}$, if there exists $k : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ such that the two properties in Definition 2 hold, then the evaluation functionals are bounded. This can be seen as follows: $\forall x \in \mathcal{X}, \forall f \in \mathcal{H}$, we have $|\delta_x(f)| = |f(x)| = |\langle f, k_x\rangle_{\mathcal{H}}| \le \|k_x\|_{\mathcal{H}}\|f\|_{\mathcal{H}} = \|\delta_x\|\|f\|_{\mathcal{H}}$, where $\|\delta_x\| = \|k_x\|_{\mathcal{H}} < \infty$ since $k_x \in \mathcal{H}$. □

Theorem 4. *If it exists, reproducing kernel is unique. Equivalently, a reproducing kernel Hilbert space uniquely determines its reproducing kernel.*

Proof. If both $k(x,y)$ and $k'(x,y)$ are reproducing kernels of $\mathcal{H}$, then we know that both $k(x,y)$ and $k'(x,y)$ must be symmetric. By the reproducing property we then have $k(x,y) = \langle k(x,\cdot), k'(y,\cdot)\rangle_{\mathcal{H}} = \langle k'(y,\cdot), k(x,\cdot)\rangle_{\mathcal{H}} = k'(y,x) = k'(x,y)$. This is true for all $x, y \in \mathcal{X}$, so $k = k'$. □

Definition 3 (Kernel)**.** *A function $k : \mathcal{X} \times \mathcal{X} \longrightarrow \mathbb{R}$ is called a kernel on $\mathcal{X}$ is there exists a Hilbert space (not necessarily a reproducing kernel Hilbert space) $\mathcal{F}$ and a map $\mathbf{\Phi} : \mathcal{X} \longrightarrow \mathcal{F}$ such that $k(x,y) = \langle \mathbf{\Phi}(x), \mathbf{\Phi}(y)\rangle_{\mathcal{F}}$.*

Here $\mathcal{H}$ may not be a reproducing kernel Hilbert space. We call $\mathbf{\Phi} : \mathcal{X} \to \mathcal{F}$ a *feature map* and $\mathcal{F}$ a *feature space.*

Corollary 1. *Every reproducing kernel is a kernel.*

Proof. We simply take $\mathbf{\Phi} : x \longmapsto k(x,\cdot)$, i.e. the reproducing kernel Hilbert space $\mathcal{H}$ is a feature space, and $k(x,y) = \langle k(x,\cdot), k(y,\cdot)\rangle_{\mathcal{H}} = \langle \mathbf{\Phi}(x), \mathbf{\Phi}(y)\rangle_{\mathcal{H}}$. □

Example 1 (Non-uniqueness of feature representation)**.** *Consider $\mathcal{X} = \mathbb{R}^2$, and $k(x,y) = \langle x, y\rangle^2$. We can write $k(x,y)$ as*

$$\begin{aligned}
k(x,y) &= (x_1y_1 + x_2y_2)^2 \\
&= x_1^2y_1^2 + x_2^2y_2^2 + 2x_1x_2y_1y_2 \\
&= \begin{pmatrix} x_1^2 & x_2^2 & \sqrt{2}x_1x_2 \end{pmatrix} \begin{pmatrix} y_1^2 \\ y_2^2 \\ \sqrt{2}y_1y_2 \end{pmatrix} \\
&= \begin{pmatrix} x_1^2 & x_2^2 & x_1x_2 & x_1x_2 \end{pmatrix} \begin{pmatrix} y_1^2 \\ y_2^2 \\ y_1y_2 \\ y_1y_2 \end{pmatrix}.
\end{aligned}$$

So we can use the feature maps $\phi(x) = (x_1^2, x_2^2, \sqrt{2}x_1x_2)$ or $\tilde{\phi}(x) = (x_1^2, x_2^2, x_1x_2, x_1x_2)$, with feature spaces $\mathcal{F} = \mathbb{R}^3$ or $\tilde{\mathcal{F}} = \mathbb{R}^4$. Note that $\mathcal{F}$ and $\tilde{\mathcal{F}}$ are not reproducing kernel Hilbert space since they are not unique! There are, in fact, infinitely many feature space representations (and we can even work in one or more of them). But what remains unique is the kernel and its reproducing kernel Hilbert space!

Definition 4. *A symmetric function $k : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ is positive definite if $\forall n \in \mathbb{N}$, $\forall (a_1, \ldots, a_n) \in \mathbb{R}^n$ and $\forall (x_1, \ldots, x_n) \in \mathcal{X}^n$,*

$$\sum_{i=1}^{n} \sum_{j=1}^{n} a_i a_j k(x_i, x_j) \geq 0.$$

The function $k(\cdot, \cdot)$ is strictly positive definite if for mutually distinct x_i, the equality holds only when all the a_i are zero.

Every inner product is a positive definite function, and more generally:

Corollary 2. *Every kernel is a positive definite function.*

Proof. Let $k(x, y)$ be a kernel, then there exists a Hilbert space (not necessarily a reproducing kernel Hilbert space) $\mathcal{F}$ and a map $\phi : \mathcal{X} \to \mathcal{F}$ such that $k(x, y) = \langle \phi(x), \phi(y) \rangle_{\mathcal{F}}$. We then have

$$\begin{aligned}
\sum_{i=1}^{n} \sum_{j=1}^{n} a_i a_j k(x_i, x_j) &= \sum_{i=1}^{n} \sum_{j=1}^{n} a_i a_j \langle \phi(x_i), \phi(x_j) \rangle_{\mathcal{F}} \\
&= \left\langle \sum_{i=1}^{n} a_i \phi(x_i), \sum_{j=1}^{n} a_j \phi(x_j) \right\rangle_{\mathcal{F}} \\
&= \left\| \sum_{i=1}^{n} a_i \phi(x_i) \right\|_{\mathcal{F}}^2 \geq 0.
\end{aligned}$$

□

So far, we have shown that reproducing kernel $\implies$ kernel $\implies$ positive definite function. In Theorem 5, we will show that positive definite function $\implies$ reproducing kernel, and conclude that a reproducing kernel, a kernel and a positive definite function are equivalent. Before stating the theorem, we first spell out the following two lemmas which are immediate consequences of these equivalences.

Lemma 1 (Sum and scaling of kernels)**.** *If k, k_1, k_2 are kernels on $\mathcal{X}$, and $\alpha \geq 0$ is a scalar, then αk and $k_1 + k_2$ are kernels.*

A difference of kernels is not necessarily s kernel! This is because we cannot have $k_1(x, x) - k_2(x, x) = \langle \phi(x), \phi(x) \rangle_{\mathcal{H}} < 0$. This gives the set of all kernels the geometry of a *closed convex cone.*

Lemma 2 (Product of kernels)**.** *Let k_1 and k_2 be kernels on $\mathcal{X}$ and $\mathcal{Y}$, respectively.*

(a) Then $k = k_1 \otimes k_2$, given by

$$k((x, y), (x', y')) = k_1(x, x') k_2(y, y')$$

is a kernel on $\mathcal{X} \times \mathcal{Y}$.

(b) If $\mathcal{X} = \mathcal{Y}$, then $k = k_1 \cdot k_2$, given by

$$k(x, x') = k_1(x, x') k_2(x, x')$$

is a kernel on $\mathcal{X}$.

Also, $\mathcal{H}_{k_1 \otimes k_2} \cong \mathcal{H}_{k_1} \otimes \mathcal{H}_{k_2}$ (isomorphic?!).

We will now show that a positive definite function is always a reproducing kernel of an RKHS.

Theorem 5 (Moore-Aronszajn)**.** *Let $k : \mathcal{X} \times \mathcal{X} \longrightarrow \mathbb{R}$ be positive definite. Then there is a unique reproducing kernel Hilbert space with reproducing kernel k. In particular, the span of the reproducing kernel k is dense in the resulting reproducing kernel Hilbert space.*

Proof. Let the space $\mathcal{H}_0 = \text{span}\{k(x, \cdot) : x \in \mathcal{X}\}$ be endowed with the inner product

$$\langle f, g\rangle_{\mathcal{H}_0} = \sum_{i=1}^{n}\sum_{j=1}^{m} \alpha_i \beta_j k(x_i, y_j),$$

where $f = \sum_{i=1}^{n} \alpha_i k(x_i, \cdot)$ and $g = \sum_{j=1}^{m} \beta_j k(y_j, \cdot)$. It can be seen that the evaluation functionals are continuous on $\mathcal{H}_0$, and any Cauchy sequence f_n is $\mathcal{H}_0$ which converges pointwise to 0 also converges in $\mathcal{H}_0$-norm to 0. Define $\mathcal{H}$ to be the set of functions $f : \mathcal{X} \to \mathbb{R}$ for which there exists a Cauchy sequence $\{f_n\} \in \mathcal{H}_0$ converging pointwise to f. We define the inner product between $f, g \in \mathcal{H}$ as the limit of an inner product of the Cauchy sequences $\{f_n\}, \{g_n\}$ converging to f and g respectively. It is easy to see that the inner product is well-defined and independent of the sequences used. Also, $\langle f, f\rangle_{\mathcal{H}} = 0$ if and only if $f = 0$. It can also be shown that the evaluation functionals are still continuous on $\mathcal{H}$, and that $\mathcal{H}$ is a Hilbert space. □

Now we have shown that reproducing kernel $\Longrightarrow$ kernel $\Longrightarrow$ positive definite function $\Longrightarrow$ reproducing kernel, so our proof of equivalence is complete. It is now easy to see that Lemma 1 is trivial if we use a positive-definiteness argument.

Corollary 3. *If a vector space $\mathcal{V}$ of functions on a set $\mathcal{X}$ is the direct sum of reproducing kernel Hilbert spaces $\mathcal{H}_1$ with kernel k_1 and $\mathcal{H}_2$ with kernel k_2, that is, $\mathcal{V} = \mathcal{H}_1 + \mathcal{H}_2$ and $\mathcal{H}_1 \cap \mathcal{H}_2 = \{0\}$, then $k = k_1 + k_2$ is a reproducing kernel for all of $\mathcal{V}$, and the subspaces $\mathcal{H}_1$ and $\mathcal{H}_2$ are orthogonal in the resulting reproducing kernel Hilbert space $\mathcal{V}$.*

A more general version of the corollary says that:

Theorem 6. *If $k_i(x, y)$ is the reproducing kernel of the class $\mathcal{F}_i$ with the norm $\|\cdot\|_i$, then $k(x, y) = k_1(x, y) + k_2(x, y)$ is the reproducing kernel of the class $\mathcal{F}$ of all functions $f = f_1 + f_2$ with $f_i \in \mathcal{F}_i$, and with the norm defined by*

$$\|f\|^2 = \min[\|f_1\|_1^2 + \|f_2\|_2^2],$$

the minimum taken for all the decompositions $f = f_1 + f_2$ with $f_i \in \mathcal{F}_i$.

References

[1] N. Aronszajn. Theory of reproducing kernels. *Trans. Amer. Math. Soc.*, 68:337–404, 1950.

[2] M. Reed and B. Simon. *Methods of modern mathematical physics. I.* Academic Press Inc. [Harcourt Brace Jovanovich Publishers], New York, second edition, 1980. Functional analysis.

[3] B. Schölkopf and A. J. Smola. *Learning with kernels: Support Vector Machines, Regularization, Optimization, and Beyond.* MIT Press, Cambridge, Massachusetts; London, England, first edition, 2002.

[4] G. Wahba. *Spline models for observational data*, volume 59 of *CBMS-NSF Regional Conference Series in Applied Mathematics.* Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 1990.