

Reproducing kernel Hilbert space

From Wikipedia, the free encyclopedia

In functional analysis (a branch of mathematics), a reproducing kernel Hilbert space (RKHS) is a Hilbert space of functions in which point evaluation is a continuous linear functional. Roughly speaking, this means that if two functions f and g in the RKHS are close in norm, i.e., $\|f-g\|$ is small, then f and g are also pointwise close, i.e., $|f(x)-g(x)|$ is small for all x .

The reverse need not be true. It is not entirely straightforward to construct a Hilbert space of functions which is not an RKHS.^[1] Note that L^2 spaces are not Hilbert spaces of functions (and hence not RKHSs), but rather Hilbert spaces of equivalence classes of functions (for example, the functions f and g defined by $f(x)=0$ and $g(x)=1_{\mathbb{Q}}$ are equivalent in L^2). However, there are RKHSs in which the norm is an L^2 -norm, such as the space of band-limited functions (see the example below). An RKHS is associated with a kernel that reproduces every function in the space in the sense that for any x in the set on which the functions are defined, "evaluation at x " can be performed by taking an inner product with a function determined by the kernel. Such a *reproducing kernel* exists if and only if every evaluation functional is continuous.

Figure illustrates related but varying approaches to viewing RKHS

The reproducing kernel was first introduced in the 1907 work of Stanisław Zaremba concerning boundary value problems for harmonic and biharmonic functions. James Mercer simultaneously examined functions which satisfy the reproducing property in the theory of integral equations. The idea of the reproducing kernel remained untouched for nearly twenty years until it appeared in the dissertations of Gábor Szegő, Stefan Bergman, and Salomon Bochner. The subject was eventually systematically developed in the early 1950s by Nachman Aronszajn and Stefan Bergman.^[2]

These spaces have wide applications, including complex analysis, harmonic analysis, and quantum mechanics. Reproducing kernel Hilbert spaces are particularly important in the field of statistical learning theory because of the celebrated representer theorem which states that every function in an RKHS that minimises an empirical risk function can be written as a linear combination of the kernel function evaluated at the training points. This is a practically useful result as it effectively simplifies the empirical risk minimization problem from an infinite dimensional to a finite dimensional optimization problem.

For ease of understanding, we provide the framework for real-valued Hilbert spaces. The theory can be easily extended to spaces of complex-valued functions and hence include the many important examples of reproducing kernel Hilbert spaces that are spaces of analytic functions.^[3]

Contents

- 1 Definition

- 2 Example
- 3 Moore–Aronszajn theorem
- 4 Integral operators and Mercer's theorem
- 5 Feature maps
- 6 Properties
- 7 Examples
- 8 Extension to vector-valued functions
- 9 See also
- 10 Notes
- 11 References

Definition

Let X be an arbitrary set and H a Hilbert space of real-valued functions on X . The evaluation functional over the Hilbert space of functions H is a linear functional that evaluates each function at a point x ,

$$L_x : f \mapsto f(x) \quad \forall f \in H.$$

We say that H is a reproducing kernel Hilbert space if L_x is continuous at any f in H or, equivalently, if for all x in X , L_x is a bounded operator on H , i.e. there exists some $M > 0$ such that

$$L_x[f] := f(x) \leq M\|f\|_H \quad \forall f \in H. \quad (1)$$

While property (1) is the weakest condition that ensures both the existence of an inner product and the evaluation of every function in H at every point in the domain, it does not lend itself to easy application in practice. A more intuitive definition of the RKHS can be obtained by observing that this property guarantees that the evaluation functional can be represented by taking the inner product of f with a function K_x in H . This function is the so-called reproducing kernel for the Hilbert space H from which the RKHS takes its name. More formally, the Riesz representation theorem implies that for all x in X there exists a unique element K_x of H with the reproducing property,

$$f(x) = L_x(f) = \langle f, K_x \rangle \quad \forall f \in H. \quad (2)$$

Since K_y is itself a function in H we have that for each y in X

$$K_y(x) = \langle K_y, K_x \rangle_H.$$

This allows us to define the reproducing kernel of H as a function $K : X \times X \rightarrow \mathbb{R}$ by

$$K(x, y) = \langle K_x, K_y \rangle_H.$$

From this definition it is easy to see that $K : X \times X \rightarrow \mathbb{R}$ is both symmetric and positive definite, i.e.

$$\sum_{i,j=1}^n c_i c_j K(x_i, x_j) \geq 0$$

for any $n \in \mathbb{N}$, $x_1, \dots, x_n \in X$, and $c_1, \dots, c_n \in \mathbb{R}$.^[4] The Moore-Aronszajn theorem (see below) is a sort of converse to this: if a function K satisfies these conditions then there is a Hilbert space of functions on X for which it is a reproducing kernel.

Example

The space of bandlimited functions H is a RKHS. Fix some $a < \infty$ and define

$$H = \{f \in L_2(\mathbb{R}) \mid \text{supp}(\phi) \subset [-a, a]\}$$

where $\phi(\omega) = \int f(x) e^{-i\omega x} dx$ is the Fourier transform of f . One can show that if $f \in H$ then

$$f(x) = \frac{1}{2\pi} \int_{-a}^a \phi(\omega) e^{ix\omega} d\omega$$

for some $\phi \in L_2[-a, a]$. It then follows by the Cauchy-Schwarz inequality and Plancherel's Theorem that

$$|f(x)| \leq \sqrt{\frac{1}{2\pi} \int_{-a}^a 1 d\omega} \sqrt{\frac{1}{2\pi} \int_{-a}^a |\phi(\omega)|^2 d\omega} = \sqrt{\frac{a}{\pi}} \|f\|.$$

As this inequality shows that the evaluation functional is bounded and H is also a Hilbert space, H is indeed a RKHS.

The kernel function K_x in this case is given by

$$K_x(y) = \frac{a}{\pi} \text{sinc}(a(y-x)) = \frac{\sin(a(y-x))}{\pi(y-x)}.$$

Note, that K_x in this case is the "bandlimited version" of the Dirac delta distribution and that K_x converges to $\delta(\cdot - x)$ in the weak sense, as explained in the entry for the sinc function.

Moore-Aronszajn theorem

We have seen how a reproducing kernel Hilbert space defines a reproducing kernel function that is both symmetric and positive definite. The Moore-Aronszajn theorem goes in the other direction; it states that every symmetric, positive definite kernel defines a unique reproducing kernel Hilbert space. The theorem first appeared in Aronszajn's *Theory of Reproducing Kernels*, although he attributes it to E. H. Moore.

Theorem. Suppose K is a symmetric, positive definite kernel on a set X . Then there is a unique Hilbert space of functions on X for which K is a reproducing kernel.

Proof. For all x in X , define $K_x = K(x, \cdot)$. Let H_0 be the linear span of $\{K_x : x \in X\}$. Define an inner product on H_0 by

$$\left\langle \sum_{j=1}^n b_j K_{y_j}, \sum_{i=1}^m a_i K_{x_i} \right\rangle = \sum_{i=1}^m \sum_{j=1}^n a_i b_j K(y_j, x_i).$$

The symmetry of this inner product follows from the symmetry of K and the non-degeneracy follows from the fact that K is positive definite.

Let H be the completion of H_0 with respect to this inner product. Then H consists of functions of the form

$$f(x) = \sum_{i=1}^{\infty} a_i K_{x_i}(x)$$

where $\sum_{i=1}^{\infty} a_i^2 K(x_i, x_i) < \infty$. The fact that the above sum converges for every x follows from the Cauchy-Schwarz inequality.

Now we can check the reproducing property (2):

$$\langle f, K_x \rangle = \left\langle \sum_{i=1}^{\infty} a_i K_{x_i}, K_x \right\rangle = \sum_{i=1}^{\infty} a_i K(x_i, x) = f(x).$$

To prove uniqueness, let G be another Hilbert space of functions for which K is a reproducing kernel. For any x and y in X , (2) implies that

$$\langle K_x, K_y \rangle_H = K(x, y) = \langle K_x, K_y \rangle_G.$$

By linearity, $\langle \cdot, \cdot \rangle_H = \langle \cdot, \cdot \rangle_G$ on the span of $\{K_x : x \in X\}$. Then $G = H$ by the uniqueness of the completion.

Integral operators and Mercer's theorem

We may characterize a symmetric positive definite kernel K via the integral operator using Mercer's theorem and obtain an additional view of the RKHS. Let X be a compact space equipped with a strictly positive finite Borel measure μ and $K : X \times X \rightarrow \mathbb{R}$ a continuous, symmetric, and positive definite function. Define the integral operator $T_K : L_2(X) \rightarrow L_2(X)$ as

$$[T_K f](\cdot) = \int_X K(\cdot, t) f(t) d\mu(t)$$

where $L_2(X)$ is the space of square integrable functions with respect to μ .

Mercer's theorem states that the spectral decomposition of the integral operator T_K of K yields a series representation of K in terms of the eigenvalues and eigenfunctions of T_K . This then implies that K is a reproducing kernel so that the corresponding RKHS can be defined in terms of these eigenvalues and eigenfunctions. We provide the details below.

Under these assumptions T_K is a compact, continuous, self-adjoint, and positive operator. The spectral theorem for self-adjoint operators implies that there is an at most countable decreasing sequence $(\sigma_i)_i \geq 0$ such that $\lim_{i \rightarrow \infty} \sigma_i = 0$ and

$T_K \phi_i(x) = \sigma_i \phi_i(x)$, where the $\{\phi_i\}$ form an orthonormal basis of $L_2(X)$. By the positivity T_K , $\sigma_i > 0 \forall i$. One can also show that T_K maps continuously into the space of continuous functions $C(X)$ and therefore we may choose continuous functions as the eigenvectors, that is, $\phi_i \in C(X) \forall i$. Then by Mercer's theorem K may be written in terms of the eigenvalues and continuous eigenfunctions as

$$K(x, y) = \sum_{j=1}^{\infty} \sigma_j \phi_j(x) \phi_j(y)$$

for all x, y in X such that $\lim_{n \rightarrow \infty} \sup_{u, v} |K(u, v) - \sum_{j=1}^n \sigma_j \phi_j(u) \phi_j(v)| = 0$. This above series representation is referred to as a Mercer kernel or Mercer representation of K .

Furthermore, it can be shown that the RKHS H of K is given by

$$H = \left\{ f \in L_2(X) \mid \sum_{i=1}^{\infty} \frac{\langle f, \phi_i \rangle^2}{\sigma_i} < \infty \right\}$$

where the inner product of \mathbf{H} given by $\langle \mathbf{f}, \mathbf{g} \rangle_{\mathbf{H}} = \sum_{i=1}^{\infty} \frac{\langle \mathbf{f}, \phi_i \rangle_{L_2} \langle \mathbf{g}, \phi_i \rangle_{L_2}}{\sigma_i}$. This

representation of the RKHS has application in probability and statistics, for example to the Karhunen-Loeve representation for stochastic processes and kernel PCA.

Feature maps

A feature map is a map $\varphi : \mathbf{X} \rightarrow \mathbf{F}$, where \mathbf{F} is a Hilbert space which we will call the feature space. The first sections presented the connection between bounded/continuous evaluation functions, positive definite functions, and integral operators and in this section we provide another representation of the RKHS in terms of feature maps.

We first note that every feature map defines a kernel via

$$K(\mathbf{x}, \mathbf{y}) = \langle \varphi(\mathbf{x}), \varphi(\mathbf{y}) \rangle. \quad (3)$$

Clearly K is symmetric and positive definiteness follows from the properties of inner product in \mathbf{F} . Conversely, every positive definite function and corresponding reproducing kernel Hilbert space has infinitely many associated feature maps such that (3) holds.

For example, we can trivially take $\mathbf{F} = \mathbf{H}$ and $\varphi(\mathbf{x}) = K_{\mathbf{x}}$ for all $\mathbf{x} \in \mathbf{X}$. Then (3) is satisfied by the reproducing property. Another classical example of a feature map relates to the previous section regarding integral operators by taking $\mathbf{F} = \ell^2$ and $\varphi(\mathbf{x}) = (\sqrt{\sigma_i} \phi_i(\mathbf{x}))_i$.

This connection between kernels and feature maps provides us with a new way to understand positive definite functions and hence reproducing kernels as inner products in \mathbf{H} . Moreover, every feature map can naturally define a RKHS by means of the definition of a positive definite function.

Lastly, feature maps allow us to construct function spaces that reveal another perspective on the RKHS. Consider the linear space

$$\mathbf{H}_{\varphi} = \{f : \mathbf{X} \rightarrow \mathbb{R} \mid \exists w \in \mathbf{F}, f(\mathbf{x}) = \langle w, \varphi(\mathbf{x}) \rangle_{\mathbf{F}}, \forall \mathbf{x} \in \mathbf{X}\}.$$

We can define a norm on \mathbf{H}_{φ} by

$$\|f\|_{\varphi} = \inf\{\|w\|_{\mathbf{F}} : w \in \mathbf{F}, f(\mathbf{x}) = \langle w, \varphi(\mathbf{x}) \rangle_{\mathbf{F}}, \forall \mathbf{x} \in \mathbf{X}\}.$$

It can be shown that \mathbf{H}_{φ} is a RKHS with kernel defined by $K(\mathbf{x}, \mathbf{y}) = \langle \varphi(\mathbf{x}), \varphi(\mathbf{y}) \rangle$.

This representation implies that the elements of the RKHS are inner products of elements in the feature space and can accordingly be seen as hyperplanes. This view of the RKHS is related to the kernel trick in machine learning.^[5]

Properties

The following properties of RKHSs may be useful to readers.

- Let $(X_i)_{i=1}^p$ be a sequence of sets and $(K_i)_{i=1}^p$ be a collection of corresponding positive definite functions on $(X_i)_{i=1}^p$. It then follows that

$$K((x_1, \dots, x_p), (y_1, \dots, y_p)) = K_1(x_1, y_1) \dots K_p(x_p, y_p)$$

is a kernel on $X = X_1 \times \dots \times X_p$.

- Let $X_0 \subset X$, then the restriction of K to $X_0 \times X_0$ is also a reproducing kernel.
- Consider a normalized kernel K such that $K(x, x) = 1$ for all $x \in X$. Define a pseudo-metric on X as

$$d_K(x, y) = \|K_x - K_y\|_H^2 = 2(1 - K(x, y)) \quad \forall x, y \in X.$$

By the Cauchy-Schwarz inequality,

$$K(x, y)^2 \leq K(x, x)K(y, y) \quad \forall x, y \in X.$$

This inequality allows us to view K as a measure of similarity between inputs. If $x, y \in X$ are similar then $K(x, y)$ will be closer to 1 while if $x, y \in X$ are dissimilar then $K(x, y)$ will be closer to 0.

- The closure of the span of $\{K_x | x \in X\}$ coincides with H .^[6]

Examples

Common examples of kernels include:

- Linear Kernel:

$$K(x, y) = \langle x, y \rangle$$

The RKHS H corresponding to this kernel is the dual space, consisting of functions

$$f(x) = \langle x, \beta \rangle \text{ with squared norm } \|f\|_H^2 = \|\beta\|^2$$

- Polynomial Kernel:

$$K(x, y) = (\alpha \langle x, y \rangle + 1)^d, \alpha \in \mathbb{R}, d \in \mathbb{N}$$

Other common examples are kernels which satisfy $K(x, y) = K(\|x - y\|)$. These are the radial basis function kernels.

- Radial Basis Function Kernels:

- Gaussian Kernel:

Sometimes referred to as the Radial basis function kernel, or squared exponential kernel

$$K(x, y) = e^{-\frac{\|x-y\|^2}{2\sigma^2}}, \sigma > 0$$

- Laplacian Kernel:

$$K(x, y) = e^{-\frac{\|x-y\|}{\sigma}}, \sigma > 0$$

The squared norm of a function f in the RKHS H with this kernel is

$$\|f\|_H^2 = \int f(x)^2 dx + \int f'(x)^2 dx.[7]$$

We also provide examples of Bergman kernels. Let X be finite and let H consist of all complex-valued functions on X . Then an element of H can be represented as an array of complex numbers. If the usual inner product is used, then K_x is the function whose value is 1 at x and 0 everywhere else, and $K(x, y)$ can be thought of as an identity matrix since $K(x, y) = 1$ when $x = y$ and $K(x, y) = 0$ otherwise. In this case, H is isomorphic to \mathbb{C}^n .

The case of $X = D$ (where D denotes the unit disc) is more sophisticated. Here the Bergman space $H^2(D)$ is the space of square-integrable holomorphic functions on D . It can be shown that the reproducing kernel for $H^2(D)$ is

$$K(x, y) = \frac{1}{\pi} \frac{1}{(1 - x\bar{y})^2}.$$

Lastly, the space of band limited functions f in $L^2(\mathbb{R})$ with bandwidth $2a$ are a RKHS with reproducing kernel

$$K(x, y) = \frac{\sin a(x - y)}{\pi(x - y)}.$$

Extension to vector-valued functions

In this section we extend the definition of the RKHS to spaces of vector-valued functions as this extension is particularly important in multi-task learning and manifold regularization. The main difference is that the reproducing kernel Γ is a symmetric function that is now a positive semi-definite *matrix* for any x, y in X . More formally, we define a vector-valued RKHS (vvRKHS) as a Hilbert space of functions $f : X \rightarrow \mathbb{R}^T$ such that for all $c \in \mathbb{R}^T$ and $x \in X$

$$\Gamma_x c(y) = \Gamma(x, y)c \in H \text{ for } y \in X$$

and

$$\langle f, \Gamma_x c \rangle_H = f(x)^\top c.$$

This second property parallels the reproducing property for the scalar-valued case. We note that this definition can also be connected to integral operators, bounded evaluation functions, and feature maps as we saw for the scalar-valued RKHS. We can equivalently define the vvrKHS as a vector-valued Hilbert space with a bounded evaluation functional and show that this implies the existence of a unique reproducing kernel by the Riesz Representation theorem. Mercer's theorem can also be extended to address the vector-valued setting and we can therefore obtain a feature map view of the vvrKHS. Lastly, it can also be shown that the closure of the span of $\{\Gamma_x c : x \in X, c \in \mathbb{R}^T\}$ coincides with H , another property similar to the scalar-valued case.

We can gain intuition for the vvrKHS by taking a component-wise perspective on these spaces. In particular, we find that every vvrKHS is isometrically isomorphic to a scalar-valued RKHS on a particular input space. Let $\Lambda = \{1, \dots, T\}$. Consider the space $X \times \Lambda$ and the corresponding reproducing kernel

$$\gamma : X \times \Lambda \times X \times \Lambda \rightarrow \mathbb{R}. \quad (4)$$

As noted above, the RKHS associated to this reproducing kernel is given by the closure of the span of $\{\gamma_{(x,t)} : x \in X, t \in \Lambda\}$ where $\gamma_{(x,t)}(y, s) = \gamma((x, t), (y, s))$ for every set of pairs $(x, t), (y, s) \in X \times \Lambda$.

The connection to the scalar-valued RKHS can then be made by the fact that every matrix-valued kernel can be identified with a kernel of the form of (4) via

$$\Gamma(x, y)_{(t,s)} = \gamma((x, t), (y, s)).$$

Moreover, every kernel with the form of (4) defines a matrix-valued kernel with the above expression. Now letting the map $D : H_\Gamma \rightarrow H_\gamma$ be defined as

$$(Df)(x, t) = \langle f(x), e_t \rangle_{\mathbb{R}^T}$$

where e_t is the t^{th} component of the canonical basis for \mathbb{R}^T , one can show that D is bijective and an isometry between H_Γ and H_γ .

While this view of the vvrKHS can be quite useful in multi-task learning, it should be noted that this isometry does not reduce the study of the vector-valued case to that of the scalar-valued case. In fact, this isometry procedure can make both the scalar-valued kernel and the input space too difficult to work with in practice as properties of the original kernels are often lost.^{[8][9][10]}

An important class of matrix-valued reproducing kernels are *separable* kernels which can factorized as the product of a scalar valued kernel and a T -dimensional symmetric positive semi-definite matrix. In light of our previous discussion these kernels are of the form

$$\gamma((x, t), (y, s)) = K(x, y)K_T(t, s)$$

for all x, y in X and t, s in T . As the scalar-valued kernel encodes dependencies between the inputs, we can observe that the matrix-valued kernel encodes dependencies among both the inputs and the outputs.

We lastly remark that the above theory can be further extended to spaces of functions with values in function spaces but obtaining kernels for these spaces is a more difficult task.^[11]

See also

- Positive definite kernel
- Mercer's theorem
- Kernel trick
- Kernel embedding of distributions
- Representer theorem

Notes

1. Alpay, D., and T. M. Mills. "A family of Hilbert spaces which are not reproducing kernel Hilbert spaces." J. Anal. Appl. 1.2 (2003): 107-111.
2. Okutmustur
3. Paulson
4. Durrett
5. Rosasco
6. Rosasco
7. Berlinet, Alain and Thomas, Christine. *Reproducing kernel Hilbert spaces in Probability and Statistics*, Kluwer Academic Publishers, 2004
8. De Vito
9. Zhang
10. Alvarez
11. Rosasco

References

- Alvarez, Mauricio, Rosasco, Lorenzo and Lawrence, Neil, "Kernels for Vector-Valued Functions: a Review," <http://arxiv.org/abs/1106.6251>, June 2011.
- Aronszajn, Nachman (1950). "Theory of Reproducing Kernels". *Transactions of the American Mathematical Society*. 68 (3): 337–404. doi:10.1090/S0002-9947-1950-0051437-7. JSTOR 1990404. MR 51437.
- Berlinet, Alain and Thomas, Christine. *Reproducing kernel Hilbert spaces in Probability and Statistics*, Kluwer Academic Publishers, 2004.
- Cucker, Felipe; Smale, Steve (2002). "On the Mathematical Foundations of Learning".

Bulletin of the American Mathematical Society. 39 (1): 1–49.

doi:10.1090/S0273-0979-01-00923-5. MR 1864085.

- De Vito, Ernest, Umanita, Veronica, and Villa, Silvia. "An extension of Mercer theorem to vector-valued measurable kernels," <http://arxiv.org/pdf/1110.4017.pdf>, June 2013.
- Durrett, Greg. 9.520 Course Notes, Massachusetts Institute of Technology, http://www.mit.edu/~9.520/scribe-notes/class03_gdurett.pdf, February 2010.
- Kimeldorf, George; Wahba, Grace (1971). "Some results on Tchebycheffian Spline Functions" (PDF). *Journal of Mathematical Analysis and Applications*. 33 (1): 82–95. doi:10.1016/0022-247X(71)90184-3. MR 290013.
- Okutmustur, Baver. "Reproducing Kernel Hilbert Spaces," Ph.D. dissertation, Bilkent University, <http://www.thesis.bilkent.edu.tr/0002953.pdf>, August 2005.
- Paulsen, Vern. "An introduction to the theory of reproducing kernel Hilbert spaces," <http://www.math.uh.edu/~vern/rkhs.pdf>.
- Steinwart, Ingo; Scovel, Clint (2012). "Mercer's theorem on general domains: On the interaction between measures, kernels, and RKHSs". *Constr. Approx.* 35 (3): 363–417. MR 2914365.
- Rosasco, Lorenzo and Poggio, Thomas. "A Regularization Tour of Machine Learning - MIT 9.520 Lecture Notes" Manuscript, Dec. 2014.
- Wahba, Grace, *Spline Models for Observational Data*, SIAM (<http://www.siam.org/books/>), 1990.
- Zhang, Haizhang, Xu, Yuesheng, and Zhang, Qinghui (2012). "Refinement of Operator-valued Reproducing Kernels." *Journal of Machine Learning Research* 13 91-136.

Retrieved from "https://en.wikipedia.org/w/index.php?title=Reproducing_kernel_Hilbert_space&oldid=733368530"

Categories: Hilbert space

- This page was last modified on 7 August 2016, at 09:34.
- Text is available under the Creative Commons Attribution-ShareAlike License; additional terms may apply. By using this site, you agree to the Terms of Use and Privacy Policy. Wikipedia® is a registered trademark of the Wikimedia Foundation, Inc., a non-profit organization.