What is an RKHS?

Dino Sejdinovic, Arthur Gretton

March 11, 2014

1 Outline

- Normed and inner product spaces. Cauchy sequences and completeness. Banach and Hilbert spaces.
- Linearity, continuity and boundedness of operators. Riesz representation of functionals.
- Definition of an RKHS and reproducing kernels.
- Relationship with positive definite functions. Moore-Aronszajn theorem.

2 Some functional analysis

We start by reviewing some elementary Banach and Hilbert space theory. Two key results here will prove useful in studying the properties of reproducing kernel Hilbert spaces: (a) that a linear operator on a Banach space is continuous if and only if it is bounded, and (b) that all continuous linear functionals on a Hilbert space arise from the inner product. The latter is often termed *Riesz representation theorem*.

2.1 Definitions of Banach and Hilbert spaces

We will focus on *real* Banach and Hilbert spaces, which are, first of all, vector spaces[1] over the field $\mathbb{R}$ of real numbers. We remark that the theory remains valid in the context of *complex* Banach and Hilbert spaces, defined over the field $\mathbb{C}$ of complex numbers, modulo appropriately placed complex conjugates. In particular, the complex inner product satisfies conjugate symmetry instead of symmetry.

Definition 1 (Norm)**.** Let $\mathcal{F}$ be a vector space over $\mathbb{R}$. A function $\|\cdot\|_{\mathcal{F}} : \mathcal{F} \to [0, \infty)$ is said to be *a norm* on $\mathcal{F}$ if

1. $\|f\|_{\mathcal{F}} = 0$ if and only if $f = \mathbf{0}$ (*norm separates points*),

[1] A vector space can also be known as a linear space Kreyszig (1989, Definition 2.1-1).

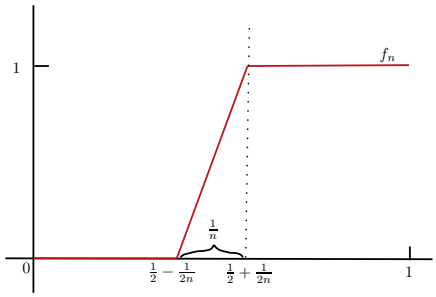

Figure 2.1: An example of a Cauchy sequence of continuous functions with no continuous limit, w.r.t. L_2-norm

2. $\|\lambda f\|_{\mathcal{F}} = |\lambda| \, \|f\|_{\mathcal{F}}$, $\forall \lambda \in \mathbb{R}$, $\forall f \in \mathcal{F}$ (*positive homogeneity*),

3. $\|f + g\|_{\mathcal{F}} \leq \|f\|_{\mathcal{F}} + \|g\|_{\mathcal{F}}$, $\forall f, g \in \mathcal{F}$ (*triangle inequality*).

Note that all elements in a normed space must have finite norm - if an element has infinite norm, it is not in the space. The norm $\|\cdot\|_{\mathcal{F}}$ induces a metric, i.e., a notion of distance on $\mathcal{F}$: $d(f, g) = \|f - g\|_{\mathcal{F}}$. This means that $\mathcal{F}$ is endowed with a certain topological structure, allowing us to study notions like continuity and convergence. In particular, we can consider when a sequence of elements of $\mathcal{F}$ converges with respect to induced distance. This gives rise to the definition of a convergent and of a Cauchy sequence:

Definition 2 (Convergent sequence)**.** A sequence $\{f_n\}_{n=1}^{\infty}$ of elements of a normed vector space $(\mathcal{F}, \|\cdot\|_{\mathcal{F}})$ is said to *converge* to $f \in \mathcal{F}$ if for every $\epsilon > 0$, there exists $N = N(\varepsilon) \in \mathbb{N}$, such that for all $n \geq N$, $\|f_n - f\|_{\mathcal{F}} < \epsilon$.

Definition 3 (Cauchy sequence)**.** A sequence $\{f_n\}_{n=1}^{\infty}$ of elements of a normed vector space $(\mathcal{F}, \|\cdot\|_{\mathcal{F}})$ is said to be *a Cauchy (fundamental) sequence* if for every $\epsilon > 0$, there exists $N = N(\varepsilon) \in \mathbb{N}$, such that for all $n, m \geq N$, $\|f_n - f_m\|_{\mathcal{F}} < \epsilon$.

From the triangle inequality $\|f_n - f_m\|_{\mathcal{F}} \leq \|f_n - f\|_{\mathcal{F}} + \|f - f_m\|_{\mathcal{F}}$, it is clear that every convergent sequence is Cauchy. However, not every Cauchy sequence in every normed space converges!

Example 4. The field of rational numbers $\mathbb{Q}$ with absolute value $|\cdot|$ as a norm is a normed vector space over itself. The sequence 1, 1.4, 1.41, 1.414, 1.4142, ... is a Cauchy sequence in $\mathbb{Q}$ which does not converge - because $\sqrt{2} \notin \mathbb{Q}$.

Example 5. In the space $C[0, 1]$ of bounded continuous functions on segment $[0, 1]$ endowed with the norm $\|f\| = \left(\int_0^1 |f(x)|^2 \, dx\right)^{1/2}$, a sequence $\{f_n\}$ of functions shown in Fig. 2.1, that take value 0 on $[0, \frac{1}{2} - \frac{1}{2n}]$ and value 1 on $[\frac{1}{2} + \frac{1}{2n}, 1]$ is Cauchy, but has no continuous limit.

Cauchy sequences are always bounded (Kreyszig, 1989, Exercise 4 p. 32), i.e., there exists $M < \infty$, s.t., $\|f_n\|_{\mathcal{F}} \leq M$, $\forall n \in \mathbb{N}$.

Next we define a complete space (Kreyszig, 1989, Definition 1.4-3):

Definition 6 (Complete space)**.** A space $\mathcal{X}$ is complete if every Cauchy sequence in $\mathcal{X}$ converges: it has a limit, and this limit is in $\mathcal{X}$.

Definition 7 (Banach space)**.** Banach space is a complete normed space, i.e., it contains the limits of all its Cauchy sequences.

Example 8. For an index set A and $p \geq 1$, the space $\ell^p(A)$ of sequences $\{x_\alpha\}_{\alpha \in A}$ of real numbers, satisfying $\sum_{\alpha \in A} |x_\alpha|^p < \infty$ is a Banach space with norm $\left\|\{x_\alpha\}_{\alpha \in A}\right\|_{\ell^p(A)} = \left(\sum_{\alpha \in A} |x_\alpha|^p\right)^{1/p}$.

Example 9. If μ is a positive measure on $\mathcal{X} \subset \mathbb{R}^d$ and $p \geq 1$, then the space

$$L_p(\mathcal{X}; \mu) := \left\{ f : X \to \mathbb{R} \text{ measurable} \;\middle|\; \int_{\mathcal{X}} |f(x)|^p d\mu < \infty \right\} \tag{2.1}$$

is a Banach space with norm $\|f\|_p = \left(\int_{\mathcal{X}} |f(x)|^p d\mu\right)^{1/p}$.

In order to study useful geometrical notions analogous to those of Euclidean spaces $\mathbb{R}^d$, e.g., orthogonality, one requires additional structure on a Banach space, that is provided by a notion of inner product:

Definition 10 (Inner product)**.** Let $\mathcal{F}$ be a vector space over $\mathbb{R}$. A function $\langle \cdot, \cdot \rangle_{\mathcal{F}} : \mathcal{F} \times \mathcal{F} \to \mathbb{R}$ is said to be *an inner product* on $\mathcal{F}$ if

1. $\langle \alpha_1 f_1 + \alpha_2 f_2, g \rangle_{\mathcal{F}} = \alpha_1 \langle f_1, g \rangle_{\mathcal{F}} + \alpha_2 \langle f_2, g \rangle_{\mathcal{F}}$
2. $\langle f, g \rangle_{\mathcal{F}} = \langle g, f \rangle_{\mathcal{F}}$
3. $\langle f, f \rangle_{\mathcal{F}} \geq 0$ and $\langle f, f \rangle_{\mathcal{F}} = 0$ if and only if $f = 0$.

Vector space with an inner product is said to be an inner product (or unitary) space. Some immediate consequences of Definition 10 are that:

- $\langle f, g \rangle_{\mathcal{F}} = 0, \ \forall f \in \mathcal{F}$ if and only if $g = 0$.
- $\langle f, \alpha_1 g_1 + \alpha_2 g_2 \rangle_{\mathcal{F}} = \alpha_1 \langle f, g_1 \rangle_{\mathcal{F}} + \alpha_2 \langle f, g_2 \rangle_{\mathcal{F}}$.

One can always define *a norm* induced by the inner product:

$$\|f\|_{\mathcal{F}} = \langle f, f \rangle_{\mathcal{F}}^{1/2},$$

and the following useful relations between the norm and the inner product hold:

- $|\langle f, g \rangle| \leq \|f\| \cdot \|g\|$ (*Cauchy-Schwarz inequality*)
- $\|f + g\|^2 + \|f - g\|^2 = 2\|f\|^2 + 2\|g\|^2$ (*the parallelogram law*)

- $4\langle f, g\rangle = \|f+g\|^2 - \|f-g\|^2$ (*the polarization identity*[2])

Definition 11 (Hilbert space)**.** Hilbert space is a complete inner product space, i.e., it is a Banach space with an inner product.

Two key examples of Hilbert spaces are given below.

Example 12. For an index set A, the space $\ell^2(A)$ is a Hilbert space with inner product

$$\langle \{x_\alpha\}, \{y_\alpha\} \rangle_{\ell^2(A)} = \sum_{\alpha \in A} x_\alpha y_\alpha.$$

Example 13. If μ is a positive measure on $\mathcal{X} \subset \mathbb{R}^d$, then the space $L_2(\mathcal{X}; \mu)$ is a Hilbert space with inner product

$$\langle f, g \rangle = \int_X f(x) g(x) d\mu.$$

Strictly speaking, $L_2(\mathcal{X}; \mu)$ is the space of equivalence classes of functions that differ by at most a set of μ-measure zero[3]. If μ is the Lebesgue measure, it is customary to write $L_2(\mathcal{X})$ as a shorthand[4].

More Hilbert space examples can be found in Kreyszig (1989, p. 132 and 133).

2.2 Bounded/Continuous linear operators

In the following, we take $\mathcal{F}$ and $\mathcal{G}$ to be normed vector spaces over $\mathbb{R}$ (for instance, they could both be the Banach spaces of functions mapping from $\mathcal{X} \subset \mathbb{R}$ to $\mathbb{R}$, with L_p-norm)

Definition 14 (Linear operator)**.** A function $A : \mathcal{F} \to \mathcal{G}$, where $\mathcal{F}$ and $\mathcal{G}$ are both normed linear spaces over $\mathbb{R}$, is called a **linear operator** if and only if it satisfies the following properties:

- **Homogeneity:** $A(\alpha f) = \alpha (Af) \quad \forall \alpha \in \mathbb{R},\, f \in \mathcal{F}$,
- **Additivity:** $A(f+g) = Af + Ag \quad \forall f, g \in \mathcal{F}$.

Example 15. Let $\mathcal{F}$ be an inner product space. For $g \in \mathcal{F}$, operator $A_g : \mathcal{F} \to \mathbb{R}$, defined with $A_g(f) := \langle f, g \rangle_{\mathcal{F}}$ is a linear operator. Note that the codomain of A_g is the underlying field $\mathbb{R}$, which is trivially a normed linear space over itself[5]. Such scalar-valued operators are called *functionals* on $\mathcal{F}$.

[2] the polarization identity is different in complex Hilbert spaces and reads: $4\langle f, g\rangle = \|f+g\|^2 - \|f-g\|^2 + i\|f+ig\|^2 - i\|f-ig\|^2$

[3] Norm defined in (**??**) does not separate functions f and g which differ only on some set $A \subset \mathcal{X}$, for which $\mu(A) = 0$, since $f - g \neq 0$ and $\|f-g\|_2^2 = \int_{\mathcal{X}} \left(|f(x) - g(x)|^2\right) d\mu = 0$. Thus, we consider all such functions as a single element in the space $L_2(\mathcal{X}; \mu)$.

[4] In fact, $\ell^2(A)$ is just $L_2(\mathcal{X}; \mu)$ where μ is the *counting measure*, where the "size" of a subset is taken to be the number of elements in the subset

[5] with norm $|\cdot|$

Definition 16 (Continuity)**.** A function $A : \mathcal{F} \to \mathcal{G}$ is said to be **continuous** at $f_0 \in \mathcal{F}$, if for every $\epsilon > 0$, there exists a $\delta = \delta(\epsilon, f_0) > 0$, s.t.

$$\|f - f_0\|_{\mathcal{F}} < \delta \qquad \text{implies} \qquad \|Af - Af_0\|_{\mathcal{G}} < \epsilon. \tag{2.2}$$

A is **continuous** on $\mathcal{F}$, if it is continuous at every point of $\mathcal{F}$. If, in addition, δ depends on ϵ only, i.e., $\forall \epsilon > 0$, $\exists \delta = \delta(\epsilon) > 0$, s.t. (2.2) holds for every $f_0 \in \mathcal{F}$, A is said to be **uniformly continuous**.

In other words, continuity means that a convergent sequence in $\mathcal{F}$ is mapped to a convergent sequence in $\mathcal{G}$. An even stronger form of continuity than uniform continuity is Lipschitz continuity:

Definition 17 (Lipschitz continuity)**.** A function $A : \mathcal{F} \to \mathcal{G}$ is said to be **Lipschitz continuous** if $\exists C > 0$, s.t. $\forall f_1, f_2 \in \mathcal{F}$, $\|Af_1 - Af_2\|_{\mathcal{G}} \leq C \|f_1 - f_2\|_{\mathcal{F}}$.

It is clear that Lipschitz continuous function is uniformly continuous since one can choose $\delta = \epsilon / C$.

Example 18. For $g \in \mathcal{F}$, $A_g : \mathcal{F} \to \mathbb{R}$, defined with $A_g(f) := \langle f, g \rangle_{\mathcal{F}}$ is continuous on $\mathcal{F}$:

$$|A_g(f_1) - A_g(f_2)| = |\langle f_1 - f_2, g \rangle_{\mathcal{F}}| \leq \|g\|_{\mathcal{F}} \|f_1 - f_2\|_{\mathcal{F}} .$$

Definition 19 (Operator norm)**.** The operator norm of a linear operator $A : \mathcal{F} \to \mathcal{G}$ is defined as

$$\|A\| = \sup_{f \in \mathcal{F}} \frac{\|Af\|_{\mathcal{G}}}{\|f\|_{\mathcal{F}}}$$

Definition 20 (Bounded operator)**.** The linear operator $A : \mathcal{F} \to \mathcal{G}$ is said to be a bounded operator if $\|A\| < \infty$.

It can readily be shown (Kreyszig, 1989) that operator norm satisfies all the requirements of a norm (triangle inequality, zero iff the operator maps only to the zero function, $\|\lambda A\| = |\lambda| \|A\|$ for $c \in \mathbb{R}$), and that the set of bounded linear operators $A : \mathcal{F} \to \mathcal{G}$ (for which the norm is defined) is therefore itself a normed vector space. Another way to write the above is to say that, for $f \in \mathcal{F}$ (possibly) not attaining the supremum, we have

$$\|Af\|_{\mathcal{G}} \quad \leq \quad \|A\| \|f\|_{\mathcal{F}} ,$$

so there exists a non-negative real number λ for which $\|Af\|_{\mathcal{G}} \leq \lambda \|f\|_{\mathcal{F}}$, for all $f \in \mathcal{F}$, and the **smallest** such λ is precisely the operator norm. In other words, a bounded subset in $\mathcal{F}$ is mapped to a bounded subset in $\mathcal{G}$.

WARNING: In calculus, a bounded function is a function whose range is a bounded set. This definition is *not* the same as the above, which simply states that the effect of A on f is bounded by some scaling of the norm of f. There is a useful geometric interpretation of the operator norm: A maps the closed unit ball in $\mathcal{F}$, into a subset of the closed ball in $\mathcal{G}$ centered at $0 \in \mathcal{G}$ and with radius $\|A\|$. Note also the result in Kreyszig (1989, p. 96): every linear operator on a normed, finite dimensional space is bounded.

Theorem 21. *Let $(\mathcal{F}, \|\cdot\|_{\mathcal{F}})$ and $(\mathcal{G}, \|\cdot\|_{\mathcal{G}})$ be normed linear spaces. If L* is a linear operator, then the following three conditions are equivalent:

1. *L is a bounded operator.*
2. *L is continuous on $\mathcal{F}$.*
3. *L is continuous at one point of $\mathcal{F}$.*

Proof. (1)⇒(2), since $\|L(f_1 - f_2)\|_{\mathcal{G}} \leq \|L\| \, \|f_1 - f_2\|_{\mathcal{F}}$, L is Lipschitz continuous with a Lipschitz constant $\|L\|$, and (2)⇒(3) trivially. Now assume that L is continuous at one point $f_0 \in \mathcal{F}$. Then, there is a $\delta > 0$, s.t. $\|L\Delta\|_{\mathcal{G}} = \|L(f_0 + \Delta) - Lf_0)\|_{\mathcal{G}} \leq 1$, whenever $\|\Delta\|_{\mathcal{F}} \leq \delta$. But then, $\forall f \in \mathcal{F}\backslash\{0\}$, since $\left\|\delta \frac{f}{\|f\|}\right\|_{\mathcal{F}} = \delta$,

$$\begin{aligned} \|Lf\|_{\mathcal{G}} &= \delta^{-1} \|f\|_{\mathcal{F}} \left\| L\left(\delta \frac{f}{\|f\|}\right) \right\|_{\mathcal{G}} \\ &\leq \delta^{-1} \|f\|_{\mathcal{F}}, \end{aligned}$$

so $\|L\| \leq \delta^{-1}$, and (3)⇒(1), q.e.d. □

Definition 22 (Topological dual)**.** If $\mathcal{F}$ is a normed space, then the space $\mathcal{F}'$ of *continuous linear* functionals $A : \mathcal{F} \to \mathbb{R}$ is called the topological dual space of $\mathcal{F}$.

Note that there is an alternative notion of a dual space: that of an *algebraic dual*, i.e., the space of all *linear* functionals $A : \mathcal{F} \to \mathbb{R}$ (which need not be continuous). In finite-dimensional space, the two notions of dual spaces coincide but this is not the case in infinite dimensions. Unless otherwise specified, we will always refer to the topological dual when discussing the dual of $\mathcal{F}$.

We have seen in Examples 15, 18 that the functionals of the form $\langle \cdot, g \rangle_{\mathcal{F}}$ on an inner product space $\mathcal{F}$ are both linear and continuous, i.e., they lie in the topological dual $\mathcal{F}'$ of $\mathcal{F}$. It turns out that if $\mathcal{F}$ is a Hilbert space, all elements of $\mathcal{F}'$ take this form[6].

Theorem 23. (Riesz representation) *In a Hilbert space $\mathcal{F}$, all continous linear functionals are of the form $\langle \cdot, g \rangle_{\mathcal{F}}$, for some $g \in \mathcal{F}$.*

Two Hilbert spaces may have elements of different nature, e.g., functions vs. sequences, but still have exactly the same geometric structure. This is the notion of *isometric isomorphism* of two Hilbert spaces. It combines notions of *vector space isomorphism* (a linear bijection) and of *isometry* (transformation that preserves distances). Once the isometric isomorphism of two spaces is established, it is customary to use whichever of the spaces is more convenient for the problem.

[6] An approachable proof of Riesz representation theorem is in Rudin (1987, Theorem 4.12).

Definition 24 (Hilbert space isomorphism)**.** Two Hilbert spaces $\mathcal{H}$ and $\mathcal{F}$ are said to be *isometrically isomorphic* if there is a linear bijective map $U : \mathcal{H} \to \mathcal{F}$, which preserves the inner product, i.e., $\langle h_1, h_2 \rangle_{\mathcal{H}} = \langle Uh_1, Uh_2 \rangle_{\mathcal{F}}$.

Note that Riesz representation theorem gives us a natural isometric isomorphism[7] $\psi : g \mapsto \langle \cdot, g \rangle_{\mathcal{F}}$ between $\mathcal{F}$ and $\mathcal{F}'$, whereby $\|\psi(g)\|_{\mathcal{F}'} = \|g\|_{\mathcal{F}}$. This property will be used below when defining a kernel on RKHSs. In particular, note that the dual space of a Hilbert space is also a Hilbert space.

3 Reproducing kernel Hilbert space

3.1 Definition of an RKHS

We begin by describing in general terms the reproducing kernel Hilbert space, and its associated kernel. Let $\mathcal{H}$ be a Hilbert space[8] of functions mapping from some non-empty set $\mathcal{X}$ to $\mathbb{R}$ (we write this: $\mathcal{H} \subset \mathbb{R}^{\mathcal{X}}$). A very interesting property of an RKHS is that if two functions $f \in \mathcal{H}$ and $g \in \mathcal{H}$ are close in the norm of $\mathcal{H}$, then $f(x)$ and $g(x)$ are close for all $x \in \mathcal{X}$. We write the inner product on $\mathcal{H}$ as $\langle f, g \rangle_{\mathcal{H}}$, and the associated norm $\|f\|_{\mathcal{H}}^2 = \langle f, f \rangle_{\mathcal{H}}$. We may alternatively write the function f as $f(\cdot)$, to indicate it takes an argument in $\mathcal{X}$.

Note that since $\mathcal{H}$ is now a space of functions on $\mathcal{X}$, there is for every $x \in \mathcal{X}$ a very special functional on $\mathcal{H}$: the one that assigns to each $f \in \mathcal{H}$, its value at x:

Definition 25 (Evaluation functional)**.** Let $\mathcal{H}$ be a Hilbert space of functions $f : \mathcal{X} \to \mathbb{R}$, defined on a non-empty set $\mathcal{X}$. For a fixed $x \in \mathcal{X}$, map $\delta_x : \mathcal{H} \to \mathbb{R}$, $\delta_x : f \mapsto f(x)$ is called the (Dirac) evaluation functional at x.

It is clear that evaluation functionals are always linear: For $f, g \in \mathcal{H}$ and $\alpha, \beta \in \mathbb{R}$, $\delta_x(\alpha f + \beta g) = (\alpha f + \beta g)(x) = \alpha f(x) + \beta g(x) = \alpha \delta_x(f) + \beta \delta_x(g)$. So the natural question is whether they are also continuous (recall that this is the same as bounded). This is exactly how reproducing kernel Hilbert spaces are defined (Steinwart & Christmann, 2008, Definition 4.18(ii)):

Definition 26 (Reproducing kernel Hilbert space)**.** A Hilbert space $\mathcal{H}$ of functions $f : \mathcal{X} \to \mathbb{R}$, defined on a non-empty set $\mathcal{X}$ is said to be a Reproducing Kernel Hilbert Space (RKHS) if δ_x is continuous $\forall x \in \mathcal{X}$.

A useful consequence is that RKHSs are particularly well behaved, relative to other Hilbert spaces.

Corollary 27. *(Norm convergence in $\mathcal{H}$ implies pointwise convergence)(Berlinet & Thomas-Agnan, 2004, Corollary 1) If two functions converge in RKHS norm, then they converge at every point, i.e., if* $\lim_{n\to\infty} \|f_n - f\|_{\mathcal{H}} = 0$*, then* $\lim_{n\to\infty} f_n(x) = f(x)$, $\forall x \in \mathcal{X}$.

[7] in complex Hilbert spaces, due to conjugate symmetry of inner product, this map is antilinear, i.e., $\psi(\alpha g) = \bar{\alpha}(\psi g)$

[8] This is a complete linear space with a dot product - see earlier.

Proof. For any $x \in \mathcal{X}$,

$$\begin{aligned} |f_n(x) - f(x)| &= |\delta_x f_n - \delta_x f| \\ &\leq \|\delta_x\| \, \|f_n - f\|_{\mathcal{H}} , \end{aligned}$$

where $\|\delta_x\|$ is the norm of the evaluation operator (which is bounded by definition on the RHKS). □

Example 28. (Berlinet & Thomas-Agnan, 2004, p. 2) If we are *not* in an RKHS, then norm convergence does not necessarily imply pointwise convergence. Let $\mathcal{H}$ be the space of polynomials over $[0, 1]$ endowed with the L_2-metric:

$$\|f_1 - f_2\|_{L_2([0,1])} = \left(\int_0^1 |f_1(x) - f_2(x)|^2 \, dx \right)^{1/2},$$

and consider the sequence of functions $\{q_n\}_{n=1}^{\infty}$, where $q_n = x^n$. Then

$$\begin{aligned} \lim_{n\to\infty} \|q_n - 0\|_{L_2([0,1])} &= \lim_{n\to\infty} \left(\int_0^1 x^{2n} dx \right)^{1/2} \\ &= \lim_{n\to\infty} \frac{1}{\sqrt{2n+1}} \\ &= 0, \end{aligned}$$

and yet $q_n(1) = 1$ for all n, i.e., $q_n \to 0 \in \mathcal{H}$, but $q_n(1) \nrightarrow 0$. In other words, the evaluation of functions at point 1 is not continuous.

3.2 Reproducing kernels

The reader will note that there is no mention of a kernel in the definition of an RKHS! We next define what is meant by a kernel, and then show how it fits in with the above definition.

Definition 29. (Reproducing kernel (Berlinet & Thomas-Agnan, 2004, p. 7))

Let $\mathcal{H}$ be a Hilbert space of $\mathbb{R}$-valued functions defined on a non-empty set $\mathcal{X}$. A function $k : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ is called *a reproducing kernel* of $\mathcal{H}$ if it satisfies

- $\forall x \in \mathcal{X}, \;\; k(\cdot, x) \in \mathcal{H}$,
- $\forall x \in \mathcal{X}, \forall f \in \mathcal{H}, \;\; \langle f, k(\cdot, x) \rangle_{\mathcal{H}} = f(x)$ (the reproducing property).

In particular, for any $x, y \in \mathcal{X}$,

$$k(x, y) = \langle k(\cdot, x), k(\cdot, y) \rangle_{\mathcal{H}}. \tag{3.1}$$

The definition above raises a number of questions. What does the kernel have to do with the definition of the RKHS? Does this kernel exist? Is it unique? What properties does it have? We first consider uniqueness, which is immediate from the definition of the reproducing kernel.

Proposition 30. (Uniqueness of the reproducing kernel) *If it exists, reproducing kernel is unique.*

Proof. Assume that $\mathcal{H}$ has two reproducing kernels k_1 and k_2. Then,

$$\langle f, k_1(\cdot, x) - k_2(\cdot, x)\rangle_{\mathcal{H}} = f(x) - f(x) = 0,\ \forall f \in \mathcal{H}, \forall x \in \mathcal{X}.$$

In particular, if we take $f = k_1(\cdot, x) - k_2(\cdot, x)$, we obtain $\|k_1(\cdot, x) - k_2(\cdot, x)\|^2_{\mathcal{H}} = 0$, $\forall x \in \mathcal{X}$, i.e., $k_1 = k_2$. □

To establish existence of a reproducing kernel in an RKHS, we will make use of the Riesz representation theorem - which tells us that in an RKHS, evaluation itself can be represented as an inner product!

Theorem 31. (Existence of the reproducing kernel) *$\mathcal{H}$ is a reproducing kernel Hilbert space (i.e., its evaluation functionals δ_x are continuous), if and only if $\mathcal{H}$ has a reproducing kernel.*

Proof. Given that a Hilbert space $\mathcal{H}$ has a reproducing kernel k with the reproducing property $\langle f, k(\cdot, x)\rangle_{\mathcal{H}} = f(x)$, then

$$\begin{aligned} |\delta_x f| &= |f(x)| \\ &= |\langle f, k(\cdot, x)\rangle_{\mathcal{H}}| \\ &\leq \|k(\cdot, x)\|_{\mathcal{H}} \|f\|_{\mathcal{H}} \\ &= \langle k(\cdot, x), k(\cdot, x)\rangle^{1/2}_{\mathcal{H}} \|f\|_{\mathcal{H}} \\ &= k(x, x)^{1/2} \|f\|_{\mathcal{H}} \end{aligned}$$

where the third line uses the Cauchy-Schwarz inequality. Consequently, $\delta_x : \mathcal{F} \to \mathbb{R}$ is a bounded linear operator.

To prove the other direction, assume that $\delta_x \in \mathcal{H}'$, i.e. $\delta_x : \mathcal{F} \to \mathbb{R}$ is a bounded linear functional. The Riesz representation theorem (Theorem 23) states that there exists an element $f_{\delta_x} \in \mathcal{H}$ such that

$$\delta_x f = \langle f, f_{\delta_x}\rangle_{\mathcal{H}},\ \forall f \in \mathcal{H}.$$

Define $k(x', x) = f_{\delta_x}(x')$, $\forall x, x' \in \mathcal{X}$. Then, clearly $k(\cdot, x) = f_{\delta_x} \in \mathcal{H}$, and $\langle f, k(\cdot, x)\rangle_{\mathcal{H}} = \delta_x f = f(x)$. Thus, k is the reproducing kernel. □

From the above, we see $k(\cdot, x)$ is in fact the *representer of evaluation* at x. We now turn to one of the most important properties of the kernel function: specifically, that it is positive definite (Berlinet & Thomas-Agnan, 2004, Definition 2), (Steinwart & Christmann, 2008, Definition 4.15).

Definition 32 (Positive definite functions)**.** A symmetric[9] function $h : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ is positive definite if $\forall n \geq 1$, $\forall (a_1, \ldots a_n) \in \mathbb{R}^n$, $\forall (x_1, \ldots, x_n) \in \mathcal{X}^n$,

$$\sum_{i=1}^{n} \sum_{j=1}^{n} a_i a_j h(x_i, x_j) \geq 0. \tag{3.2}$$

[9]Note that we require symmetry of h in addition to (3.2). In the complex case, $\sum_{i=1}^{n} \sum_{j=1}^{n} a_i \bar{a_j} h(x_i, x_j) \geq 0$, satisfied for all complex scalars $(a_1, \ldots a_n) \in \mathbb{C}^n$ will itself imply conjugate symmetry of h. Can you construct a non-symmetric h that satisfies (3.2)?

The function $h(\cdot,\cdot)$ is *strictly* positive definite if for mutually distinct x_i, the equality holds only when all the a_i are zero.[10]

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

Lemma 33. *Let $\mathcal{F}$ be any Hilbert space (not necessarily an RKHS), $\mathcal{X}$ a non-empty set and $\phi : \mathcal{X} \to \mathcal{F}$. Then $h(x,y) := \langle \phi(x), \phi(y) \rangle_{\mathcal{F}}$ is a positive definite function.*

Proof.

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

□

Corollary 34. *Reproducing kernels are positive definite.*

Proof. For a reproducing kernel k in an RKHS $\mathcal{H}$, one has $k(x,y) = \langle k(\cdot,x), k(\cdot,y)\rangle_{\mathcal{H}}$, so it is sufficient to take $\phi : x \mapsto k(\cdot,x)$. □

The following Lemma goes in the converse direction and shows that *all* positive definite functions are in fact intimately related to inner products in feature spaces, because a Cauchy-Schwarz type inequality holds. We will later see (in the Moore-Aronszajn theorem) that they are actually equivalent concepts.

Lemma 35. *If h is positive definite, then $|h(x_1,x_2)|^2 \leq h(x_1,x_1)h(x_2,x_2)$.*

Proof. If $h(x_1,x_2) = 0$, inequality is clear. Otherwise, take $a_1 = a$, $a_2 = h(x_1,x_2)$. We have that $q(a) = a^2 h(x_1,x_1) + 2a\,|h(x_1,x_2)|^2 + |h(x_1,x_2)|^2\, h(x_2,x_2) \geq 0$. Since q is a quadratic function of a and inequality holds $\forall a$, we have $|h(x_1,x_2)|^4 \leq |h(x_1,x_2)|^2\, h(x_1,x_1)h(x_2,x_2)$, which proves the claim. □

3.3 Feature space, and other kernel properties

This section summarizes the relevant parts of Steinwart & Christmann (2008, Section 4.1).

Following Lemma 33, one can define a *kernel* (note that we drop qualification *reproducing* here - later we will see that these two notions are the same), as a function which can be represented via inner product, and this is the approach taken in Steinwart & Christmann (2008, Section 4.1):

[10] Note that Wendland (2005, Definition 6.1 p. 65) uses the terminology "positive semi-definite" vs "positive definite". This is probably more logical, since it then coincides with the terminology used in linear algebra. However, we proceed with the terminology prevalent with machine learning literature.

Definition 36 (*Kernel*)**.** Let $\mathcal{X}$ be a non-empty set. The function $k : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ is said to be a *kernel* if there exists a real Hilbert space $\mathcal{H}$ and a map $\phi : \mathcal{X} \to \mathcal{H}$ such that $\forall x, y \in \mathcal{H}$,

$$k(x, y) = \langle \phi(x), \phi(y) \rangle_{\mathcal{H}} .$$

Such map $\phi : \mathcal{X} \to \mathcal{H}$ is referred to as the feature map, and space $\mathcal{H}$ as the feature space. For a given kernel, there may be more than one feature map, as demonstrated by the following example.

Example 37. Consider $\mathcal{X} = \mathbb{R}$, and

$$k(x, y) = xy = \begin{bmatrix} \frac{x}{\sqrt{2}} & \frac{x}{\sqrt{2}} \end{bmatrix} \begin{bmatrix} \frac{y}{\sqrt{2}} \\ \frac{y}{\sqrt{2}} \end{bmatrix},$$

where we defined the feature maps $\phi(x) = x$ and $\tilde{\phi}(x) = \begin{bmatrix} \frac{x}{\sqrt{2}} & \frac{x}{\sqrt{2}} \end{bmatrix}$, and where the feature spaces are respectively, $\mathcal{H} = \mathbb{R}$, and $\tilde{\mathcal{H}} = \mathbb{R}^2$.

Lemma 38. [ℓ_2 convergent sequences are kernel feature maps] *For every $x \in \mathcal{X}$, assume the sequence $\{f_n(x)\} \in \ell_2$ for $n \in \mathbb{N}$, where $f_n : \mathcal{X} \to \mathbb{R}$. Then*

$$k(x_1, x_2) := \sum_{n=1}^{\infty} f_n(x_1) f_n(x_2) \tag{3.3}$$

is a kernel.

Proof. Hölder's inequality states

$$\sum_{n=1}^{\infty} |f_n(x_1) f_n(x_2)| \leq \|f_n(x_1)\|_{\ell_2} \|f_n(x_2)\|_{\ell_2} .$$

so the series (3.3) converges absolutely. Defininig $\mathcal{H} := \ell_2$ and $\phi(x) = \{f_n(x)\}$ completes the proof. □

4 Construction of an RKHS from a kernel: Moore-Aronsajn

We have seen previously that *given* a reproducing kernel Hilbert space $\mathcal{H}$, we may define a unique reproducing kernel associated with $\mathcal{H}$, which is a positive definite function. Then we considered kernels, i.e., functions that can be written as an inner product in a feature space. All reproducing kernels are kernels. In Example (37), we have seen that the representation of a kernel as an inner product in a feature space may not be unique. However, neither of the feature spaces in that example is an RKHS, as they are not spaces of functions on $\mathcal{X} = \mathbb{R}$.

Our goal now is to show that for every positive definite function $k(x, y)$, there corresponds a *unique RKHS* $\mathcal{H}$, for which k is a reproducing kernel. The proof is rather tricky, but also very revealing of the properties of RKHSs, so it is worth understanding (it also occurs in very incomplete form in a number of books and tutorials, so it is worth seeing what a complete proof looks like).

Starting with the positive definite function, we will construct a pre-RKHS $\mathcal{H}_0$, from which we will form the RKHS $\mathcal{H}$. The pre-RKHS $\mathcal{H}_0$ must satisfy two properties:

1. the evaluation functionals δ_x are continuous on $\mathcal{H}_0$,
2. Any Cauchy sequence f_n in $\mathcal{H}_0$ which converges pointwise to 0 also converges in $\mathcal{H}_0$-norm to 0.

The last result has an important implication: Any Cauchy sequence $\{f_n\}$ in $\mathcal{H}_0$ that converges pointwise to $f \in \mathcal{H}_0$, also converges to f in $\|\cdot\|_{\mathcal{H}_0}$, since in that case $\{f_n - f\}$ converges pointwise to 0, and thus $\|f_n - f\|_{\mathcal{H}_0} \to 0$. **PREVIEW**: we can already say what the pre-RKHS $\mathcal{H}_0$ will look like: it is the set of functions

$$f(x) = \sum_{i=1}^{n} \alpha_i k(x_i, x). \tag{4.1}$$

After the proof, we'll show in Section (4.5) that these functions satisfy conditions (1) and (2) of the pre-Hilbert space.

Next, **define** $\mathcal{H}$ to be the set of functions $f \in \mathbb{R}^{\mathcal{X}}$ for which there exists an $\mathcal{H}_0$-Cauchy sequence $\{f_n\} \in \mathcal{H}_0$ converging pointwise to f: note that $\mathcal{H}_0 \subset \mathcal{H}$, since the limits of these Cauchy sequences might not be in $\mathcal{H}_0$. Our goal is to prove that $\mathcal{H}$ is an RKHS. The two properties above hold if and only if

- $\mathcal{H}_0 \subset \mathcal{H} \subset \mathbb{R}^{\mathcal{X}}$ and the topology induced by $\langle \cdot, \cdot \rangle_{\mathcal{H}_0}$ on $\mathcal{H}_0$ coincides with the topology induced on $\mathcal{H}_0$ by $\mathcal{H}$.
- $\mathcal{H}$ has reproducing kernel $k(x, y)$.

We concern ourselves with proving that (1), (2) imply the above bullet points, since the reverse direction is easy to prove. This takes four steps:

1. 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. Is the inner product well defined, and independent of the sequences used? This is proved in Section 4.1.
2. Recall that an inner product space must satisfy $\langle f, f \rangle_{\mathcal{H}} = 0$ iff $f = 0$. Is this true when we define the inner product on $\mathcal{H}$ as above? (Note that we can also check that the remaining requirements for an inner product on $\mathcal{H}$ hold, but these are straightforward)
3. Are the evaluation functionals still continuous on $\mathcal{H}$?
4. Is $\mathcal{H}$ complete?

Finally, we'll see that the functions (4.1) define a valid pre-RKHS $\mathcal{H}_0$. We will also show that the kernel $k(\cdot, x)$ has the reproducing property on the RKHS $\mathcal{H}$.

4.1 Is the inner product well defined in $\mathcal{H}$?

In this section we prove that if we define the inner product in $\mathcal{H}$ of all limits of Cauchy sequences as (4.2) below, then this limit is *well defined*: (1) it converges, and (2) it depends only on the *limits* of the Cauchy sequences, and not the particular sequences themselves.

This result is from Berlinet & Thomas-Agnan (2004, Lemma 5).

Lemma 39. *For $f, g \in \mathcal{H}$ and Cauchy sequences (wrt the $\mathcal{H}_0$ norm) $\{f_n\}$, $\{g_n\}$ converging pointwise to f and g, define $\alpha_n = \langle f_n, g_n \rangle_{\mathcal{H}_0}$. Then, $\{\alpha_n\}$ is convergent and its limit depends only on f and g. We thus define*

$$\langle f, g \rangle_{\mathcal{H}} := \lim_{n \to \infty} \langle f_n, g_n \rangle_{\mathcal{H}_0} \tag{4.2}$$

Proof that $\alpha_n = \langle f_n, g_n \rangle_{\mathcal{H}_0}$ is convergent: For $n, m \in \mathbb{N}$,

$$\begin{aligned}
|\alpha_n - \alpha_m| &= \left|\langle f_n, g_n \rangle_{\mathcal{H}_0} - \langle f_m, g_m \rangle_{H_0}\right| \\
&= \left|\langle f_n, g_n \rangle_{H_0} - \langle f_m, g_n \rangle_{H_0} + \langle f_m, g_n \rangle_{H_0} - \langle f_m, g_m \rangle_{H_0}\right| \\
&= \left|\langle f_n - f_m, g_n \rangle_{H_0} + \langle f_m, g_n - g_m \rangle_{H_0}\right| \\
&\leq \left|\langle f_n - f_m, g_n \rangle_{H_0}\right| + \left|\langle f_m, g_n - g_m \rangle_{H_0}\right| \\
&\leq \|g_n\|_{H_0} \|f_n - f_m\|_{H_0} + \|f_m\|_{H_0} \|g_n - g_m\|_{H_0} .
\end{aligned}$$

Take $\epsilon > 0$. Every Cauchy sequence is bounded, so $\exists A, B \in \mathbb{R}$, $\|f_m\|_{H_0} \leq A$, $\|g_n\|_{H_0} \leq B$, $\forall n, m \in \mathbb{N}$.

By taking $N_1 \in \mathbb{N}$ s.t. $\|f_n - f_m\|_{H_0} < \frac{\epsilon}{2B}$, for $n, m \geq N_1$, and $N_2 \in \mathbb{N}$ s.t. $\|g_n - g_m\|_{H_0} < \frac{\epsilon}{2A}$, for $n, m \geq N_1$, we have that $|\alpha_n - \alpha_m| < \epsilon$, for $n, m \geq \max(N_1, N_2)$, which means that $\{\alpha_n\}$ is a Cauchy sequence in $\mathbb{R}$, which is complete, and the sequence is convergent in $\mathbb{R}$.

Proof that limit is independent of Cauchy sequence chosen:

If some $\mathcal{H}_0$-Cauchy sequences $\{f'_n\}$, $\{g'_n\}$ also converge pointwise to f and g, and $\alpha'_n = \langle f'_n, g'_n \rangle_{H_0}$, one similarly shows that

$$|\alpha_n - \alpha'_n| \leq \|g_n\|_{H_0} \|f_n - f'_n\|_{H_0} + \|f'_n\|_{H_0} \|g_n - g'_n\|_{H_0} .$$

Now, since $\{f_n\}$ and $\{f'_n\}$ both converge pointwise to f, $\{f_n - f'_n\}$ converges pointwise to 0, and so does $\{g_n - g'_n\}$. But then they also converge to 0 in $\|\cdot\|_{H_0}$ by the pre-RKHS axiom 2, and therefore $\{\alpha_n\}$ and $\{\alpha'_n\}$ must have the same limit.

4.2 Does it hold that $\langle f, f \rangle_{\mathcal{H}} = 0$ iff $f = 0$?

In this section, we verify that all the expected properties of an inner product from Definition (10) hold for $\mathcal{H}$. It turns out that the only challenging property to show is the third one - the others follow from the inner product definition on the pre-RKHS. This result is from Berlinet & Thomas-Agnan (2004, Lemma 6).

Lemma 40. *Let $\{f_n\}$ be Cauchy sequence in $\mathcal{H}_0$ converging pointwise to $f \in \mathcal{H}$. If $\lim_{n\to\infty} \|f_n\|^2_{\mathcal{H}_0} = 0$, then $f(x) = 0$ pointwise for all x (we assumed pointwise convergence implies norm convergence - we now want to prove the other direction, bearing in mind that the inner product in $\mathcal{H}$ is defined as the limit of inner products in $\mathcal{H}_0$ by (4.2)).*

Proof: We have$\forall x \in \mathcal{X}$,

$$f(x) = \lim_{n\to\infty} f_n(x) = \lim_{n\to\infty} \delta_x(f_n) \underset{\text{(a)}}{\leq} \lim_{n\to\infty} \|\delta_x\| \, \|f_n\|_{\mathcal{H}_0} \underset{\text{(b)}}{=} 0,$$

where in (a) we used that the evaluation functional δ_x is continuous on $\mathcal{H}_0$, by the pre-RKHS axiom 1 (hence bounded, with a well defined operator norm $\|\delta_x\|$); and in (b) we used the assumption in the lemma that f_n converges to 0 in $\|\cdot\|_{\mathcal{H}_0}$.

4.3 Are the evaluation functionals continuous on $\mathcal{H}$?

Here we need to establish a preliminary lemma, before we can continue.

Lemma 41. *$\mathcal{H}_0$ is dense in $\mathcal{H}$ (Berlinet & Thomas-Agnan, 2004, Lemma 7, Corollary 2).*

Proof. It suffices to show that given any $f \in \mathcal{H}$ and its associated Cauchy sequence $\{f_n\}$ wrt $\mathcal{H}_0$ converging pointwise to f (which exists by definition), $\{f_n\}$ also converges to f in $\|\cdot\|_{\mathcal{H}}$ (note: this is the *new* norm which we defined above in terms of limits of Cauchy sequences in $\mathcal{H}_0$).

Since$\{f_n\}$ is Cauchy in $\mathcal{H}_0$-norm, for all $\epsilon > 0$, there is $N \in \mathbb{N}$, s.t. $\|f_m - f_n\|_{\mathcal{H}_0} < \epsilon$, $\forall m, n \geq N$. Fix $n^* \geq N$. The sequence $\{f_m - f_{n^*}\}_{m=1}^{\infty}$ converges pointwise to $f - f_{n^*}$. We now simply use the definition of the inner product in $\mathcal{H}$ from (4.2),

$$\|f - f_{n^*}\|^2_{\mathcal{H}} \quad = \quad \lim_{m\to\infty} \|f_m - f_{n^*}\|^2_{\mathcal{H}_0} \leq \epsilon^2,$$

whereby $\{f_n\}_{n=1}^{\infty}$ converges to f in $\|\cdot\|_{\mathcal{H}}$. □

Lemma 42. *The evaluation functionals are continuous on $\mathcal{H}$ (Berlinet & Thomas-Agnan, 2004, Lemma 8).*

Proof. We show that δ_x is continuous at $f = 0$, since this implies by linearity that it is continuous everywhere. Let $x \in \mathcal{X}$, and $\epsilon > 0$. By pre-RKHS axiom 1, δ_x is continuous on $\mathcal{H}_0$. Thus, $\exists \eta$, s.t.

$$\|g - 0\|_{\mathcal{H}_0} = \|g\|_{\mathcal{H}_0} < \eta \Rightarrow |\delta_x(g)| = |g(x)| < \epsilon/2. \tag{4.3}$$

To complete the proof, we just need to show that there is a $g \in \mathcal{H}_0$ close (in $\mathcal{H}$-norm) to some $f \in \mathcal{H}$ with small norm, and that this function is also close at each point.

We take $f \in \mathcal{H}$ with $\|f\|_{\mathcal{H}} < \eta/2$. By Lemma 41 there is a Cauchy sequence $\{f_n\}$ in $\mathcal{H}_0$ converging both pointwise to f and in $\|\cdot\|_{\mathcal{H}}$ to f, so one can find $N \in \mathbb{N}$, s.t.

$$\begin{aligned} |f(x) - f_N(x)| &< \epsilon/2, \\ \|f - f_N\|_{\mathcal{H}} &< \eta/2. \end{aligned}$$

We have from these definitions that

$$\|f_N\|_{\mathcal{H}_0} = \|f_N\|_{\mathcal{H}} \leq \|f\|_{\mathcal{H}} + \|f - f_N\|_{\mathcal{H}} < \eta.$$

Thus $\|f\|_{\mathcal{H}} < \eta/2$ implies $\|f_N\|_{\mathcal{H}_0} < \eta$. Using (4.3) and setting $g := f_N$, we have that $\|f_N\|_{\mathcal{H}_0} < \eta$ *implies* $|f_N(x)| < \epsilon/2$, and thus $|f(x)| \leq |f(x) - f_N(x)| + |f_N(x)| < \epsilon$. In other words, $\|f\|_{\mathcal{H}} < \eta/2$ is shown to imply $|f(x)| < \epsilon$. This means that δ_x is continuous at 0 in the $\|\cdot\|_{\mathcal{H}}$ sense, and thus by linearity on all $\mathcal{H}$. □

4.4 Is $\mathcal{H}$ complete (a Hilbert space)?

The idea here is to show that every Cauchy sequence wrt the $\mathcal{H}$-norm converges to a function in $\mathcal{H}$.

Lemma 43. *$\mathcal{H}$ is complete.*

Let $\{f_n\}$ be any Cauchy sequence in $\mathcal{H}$. Since evaluation functionals are linear continuous on $\mathcal{H}$ by 42, then for any $x \in \mathcal{X}$, $\{f_n(x)\}$ is convergent in $\mathbb{R}$ to some $f(x) \in \mathbb{R}$ (since $\mathbb{R}$ is complete, it contains this limit). The question is thus whether the function $f(x)$ defined pointwise in this way is still in $\mathcal{H}$ (recall that $\mathcal{H}$ is defined as containing the limit of $\mathcal{H}_0$-Cauchy sequences that converge pointwise).

The proof strategy is to define a sequence of functions$\{g_n\}$, where $g_n \in \mathcal{H}_0$, which is "close" to the $\mathcal{H}$-Cauchy sequence $\{f_n\}$. These functions will then be shown **(1)** to converge pointwise to f, and **(2)** to be Cauchy in $\mathcal{H}_0$. Hence by our original construction of $\mathcal{H}$, we have $f \in \mathcal{H}$. Finally, we show $f_n \to f$ in $\mathcal{H}$-norm.

Define $f(x) := \lim_{n\to\infty} f_n(x)$. For $n \in \mathbb{N}$, choose $g_n \in \mathcal{H}_0$ such that $\|g_n - f_n\|_{\mathcal{H}} < \frac{1}{n}$. This can be done since $\mathcal{H}_0$ is dense in $\mathcal{H}$. From

$$\begin{aligned} |g_n(x) - f(x)| &\leq |g_n(x) - f_n(x)| + |f_n(x) - f(x)| \\ &\leq |\delta_x(g_n - f_n)| + |f_n(x) - f(x)|. \end{aligned}$$

The first term in this sum goes to zero due to the continuity of δ_x on $\mathcal{H}$ (Lemma 42), and thus $\{g_n(x)\}$ converges to $f(x)$, satisfying criterion (1). For criterion (2), we have

$$\begin{aligned} \|g_m - g_n\|_{\mathcal{H}_0} &= \|g_m - g_n\|_{\mathcal{H}} \\ &\leq \|g_m - f_m\|_{\mathcal{H}} + \|f_m - f_n\|_{\mathcal{H}} + \|f_n - g_n\|_{\mathcal{H}} \\ &\leq \frac{1}{m} + \frac{1}{n} + \|f_m - f_n\|_{\mathcal{H}}, \end{aligned}$$

hence $\{g_n\}$ is Cauchy in $\mathcal{H}_0$.

Finally, is this limiting f a limit with respect to the $\mathcal{H}$-norm? Yes, since by Lemma 41 (denseness of $\mathcal{H}_0$ in $\mathcal{H}$: see the first lines of the proof), g_n tends to f in the $\mathcal{H}$-norm sense, and thus f_n converges to f in $\mathcal{H}$-norm,

$$\begin{aligned} \|f_n - f\|_{\mathcal{H}} &\leq \|f_n - g_n\|_{\mathcal{H}} + \|g_n - f\|_{\mathcal{H}} \\ &\leq \frac{1}{n} + \|g_n - f\|_{\mathcal{H}}. \end{aligned}$$

Thus $\mathcal{H}$ is complete.

4.5 How to build a valid pre-RKHS $\mathcal{H}_0$

Here we show how to build a valid pre-RKHS. Importantly, in doing this, we prove that for every positive definite function, there corresponds a unique RKHS $\mathcal{H}$.

Theorem 44. (Moore-Aronszajn)

Let $k : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ be positive definite. There is a unique RKHS $\mathcal{H} \subset \mathbb{R}^{\mathcal{X}}$ with reproducing kernel k. Moreover, if space $\mathcal{H}_0 = span\left[\{k(\cdot, x)\}_{x \in \mathcal{X}}\right]$ is endowed with the inner product

$$\langle f, g \rangle_{\mathcal{H}_0} = \sum_{i=1}^{n} \sum_{j=1}^{m} \alpha_i \beta_j k(x_i, y_j), \tag{4.4}$$

where $f = \sum_{i=1}^{n} \alpha_i k(\cdot, x_i)$ and $g = \sum_{j=1}^{m} \beta_j k(\cdot, y_j)$, then $\mathcal{H}_0$ is a valid pre-RKHS.

We first need to show that (4.4) is a **valid inner product**. First, is it independent of the particular α_i and β_i used to define f, g? Yes, since

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

As a useful consequence of this result we get the **reproducing property** on $\mathcal{H}_0$, by setting $g = k(\cdot, x)$,

$$\langle f, g \rangle_{\mathcal{H}_0} = \sum_{i=1}^{n} \alpha_i g(x_i) = \sum_{i=1}^{n} \alpha_i k(x_i, x) = f(x).$$

Next, we check that the form (4.4) is indeed a valid inner product on $\mathcal{H}_0$. The only nontrivial axiom to be verified is

$$\langle f, f \rangle_{\mathcal{H}_0} = 0 \implies f = 0.$$

The proof of this is a straightforward modification of the proof of 35. Let $f = \sum_{i=1}^{n} \alpha_i k(\cdot, x_i)$, fix $x \in \mathcal{X}$ and put $a_i = a\alpha_i$, $i = 1, \ldots, n$, $a_{n+1} = f(x)$,

$x_{n+1} = x$. By positive definiteness of k:

$$\begin{aligned} 0 &\leq \sum_{i=1}^{n+1}\sum_{j=1}^{n+1} a_i a_j k(x_i, x_j) \\ &= a^2 \langle f, f\rangle_{\mathcal{H}_0} + 2a\,|f(x)|^2 + |f(x)|^2 k(x,x), \end{aligned}$$

whereby $|f(x)|^4 \leq |f(x)|^2 k(x,x) \langle f, f\rangle_{\mathcal{H}_0}$, so it is clear that if $\langle f, f\rangle_{\mathcal{H}_0} = 0$, f must vanish everywhere.

We now proceed to the main proof.

Proof. (that $\mathcal{H}_0$ satisfies the pre-RKHS axioms). Let $x \in \mathcal{X}$. Note that for $f = \sum_{i=1}^n \alpha_i k(\cdot, x_i)$

$$\langle f, k(\cdot, x)\rangle_{\mathcal{H}_0} = \sum_{i=1}^n \alpha_i k(x, x_i) = f(x), \tag{4.5}$$

and thus for $f, g \in \mathcal{H}_0$,

$$\begin{aligned} |\delta_x(f) - \delta_x(g)| &= |\langle f - g, k(\cdot, x)\rangle_{H_0}| \\ &\leq k^{1/2}(x,x)\,\|f - g\|_{\mathcal{H}_0}, \end{aligned}$$

by Cauchy-Schwarz, implying that δ_x is continuous on $\mathcal{H}_0$, and the **first** pre-RKHS requirement is satisfied.

Now, take $\epsilon > 0$ and define a Cauchy $\{f_n\}$ in $\mathcal{H}_0$ that converges pointwise to 0. Since Cauchy sequences are bounded, we may define $A > 0$, s.t. $\|f_n\|_{\mathcal{H}_0} < A$, $\forall n \in \mathbb{N}$. One can find $N_1 \in \mathbb{N}$, s.t. $\|f_n - f_m\|_{H_0} < \epsilon/2A$, for $n, m \geq N_1$. Write $f_{N_1} = \sum_{i=1}^r \alpha_i k(\cdot, x_i)$. Take $N_2 \in \mathbb{N}$, s.t. $n \geq N_2$ implies $|f_n(x_i)| < \frac{\epsilon}{2r|\alpha_i|}$ for $i = 1, \ldots, r$. Now, for $n \geq \max(N_1, N_2)$

$$\begin{aligned} \|f_n\|^2_{\mathcal{H}_0} &\leq |\langle f_n - f_{N_1}, f_n\rangle_{\mathcal{H}_0}| + |\langle f_{N_1}, f_n\rangle_{\mathcal{H}_0}| \\ &\leq \|f_n - f_{N_1}\|_{\mathcal{H}_0}\,\|f_n\|_{\mathcal{H}_0} + \sum_{i=1}^r |\alpha_i f_n(x_i)| \\ &< \epsilon, \end{aligned}$$

so f_n converges to 0 in $\|\cdot\|_{\mathcal{H}_0}$. Thus, all the pre-RKHS axoims are satisfied, and $\mathcal{H}$ is an RKHS.

To see that the **reproducing kernel** on $\mathcal{H}$ is k, simply note that if $f \in \mathcal{H}$, and $\{f_n\}$ in $\mathcal{H}_0$ converges to f pointwise,

$$\begin{aligned} \langle f, k(\cdot, x)\rangle_{\mathcal{H}} &\underset{\text{(a)}}{=} \lim_{n\to\infty} \langle f_n, k(\cdot, x)\rangle_{\mathcal{H}_0} \\ &= \lim_{n\to\infty} f_n(x) \\ &= f(x). \end{aligned}$$

where in (a) we use the definition of inner product on $\mathcal{H}$ in (4.2). Since $\mathcal{H}_0$ is dense in $\mathcal{H}$, $\mathcal{H}$ is the unique RKHS that contains $\mathcal{H}_0$. But since $k(\cdot, x) \in \mathcal{H}$, $\forall x \in \mathcal{X}$, it is clear that any RKHS with reproducing kernel k must contain $\mathcal{H}_0$. □

4.6 Summary

Moore-Aronszajn theorem tells us that every positive definite function is a reproducing kernel. We have previously seen that every reproducing kernel is a kernel and that every kernel is a positive definite function. Therefore, all three notions are exactly the same! In addition, we have established a bijection between the set of all positive definite functions on $\mathcal{X} \times \mathcal{X}$, denoted by $\mathbb{R}_+^{\mathcal{X} \times \mathcal{X}}$, and the set of all reproducing kernel Hilbert spaces, denoted by $Hilb(\mathbb{R}^{\mathcal{X}})$, which consists of subspaces of $\mathbb{R}^{\mathcal{X}}$. It turns out that this bijection also preserves the geometric structure of these sets, which are in both cases *closed convex cones*, and we will give some intuition on this in the next Section.

5 Operations with kernels

Since kernels are just positive definite functions, the following Lemma is immediate:

Lemma 45. (Sum and scaling of kernels)
If k, k_1, and k_2 are kernels on $\mathcal{X}$, and $\alpha \geq 0$ is a scalar, then αk, $k_1 + k_2$ are kernels.

Note that a difference of kernels is not necessarily a kernel! This is because we cannot have $k_1(x, x) - k_2(x, x) < 0$, since we would then have a feature map $\phi : \mathcal{X} \to \mathcal{H}$ for which $\langle \phi(x), \phi(x) \rangle_{\mathcal{H}} < 0$. Mathematically speaking, these properties give the set of all kernels the structure of a convex cone (*not* a linear space). Now, consider the following: since we know that $k = k_1 + k_2$ also has an RKHS $\mathcal{H}_k$, what is the relationship between $\mathcal{H}_k$ and the RKHSs $\mathcal{H}_{k_1}$ and $\mathcal{H}_{k_2}$ of k_1 and k_2? The following theorem gives an answer.

Theorem 46. (Sum of RKHSs)
Let $k_1, k_2 \in \mathbb{R}_+^{\mathcal{X} \times \mathcal{X}}$, and $k = k_1 + k_2$. Then,

$$\mathcal{H}_k = \mathcal{H}_{k_1} + \mathcal{H}_{k_2} = \{ f_1 + f_2 \,:\, f_1 \in \mathcal{H}_{k_1}, f_2 \in \mathcal{H}_{k_2} \}, \tag{5.1}$$

and $\forall f \in \mathcal{H}_k$,

$$\|f\|_{\mathcal{H}_k}^2 = \min_{f_1 + f_2 = f} \left\{ \|f_1\|_{\mathcal{H}_{k_1}}^2 + \|f_2\|_{\mathcal{H}_{k_2}}^2 \right\}. \tag{5.2}$$

The product of kernels is also a kernel. Note that this contains as a consequence a familiar fact from linear algebra: that the Hadamard product of two positive definite matrices is positive definite.

Theorem 47. (Product of kernels)
Let k_1 and k_2 be kernels on $\mathcal{X}$ and $\mathcal{Y}$, respectively. Then

$$k\left((x,y),(x',y')\right) \quad := \quad k_1(x,x')k_2(y,y')$$

is a kernel on $\mathcal{X}\times\mathcal{Y}$. In addition, there is an isometric isomorphism between $\mathcal{H}_k$ and the Hilbert space tensor product $\mathcal{H}_{k_\mathcal{X}}\otimes\mathcal{H}_{k_\mathcal{Y}}$. In addition, if $\mathcal{X}=\mathcal{Y}$,

$$k\left(x,x'\right) \quad := \quad k_1(x,x')k_2(x,x')$$

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

The above results enable us to construct many interesting kernels by multiplication, addition and scaling by non-negative scalars. To illustrate this assume that $\mathcal{X}=\mathbb{R}$. The trivial (linear) kernel on $\mathbb{R}^d$ is $k_{lin}(x,x')=\langle x,x'\rangle$. Then for any polynomial $p(t)=a_mt^m+\cdots+a_1t+a_0$ with non-negative coefficients a_i, $p(\langle x,x'\rangle)$ defines a valid kernel on $\mathbb{R}^d$. This gives rise to the **polynomial kernel** $k_{poly}(x,x')=(\langle x,x'\rangle+c)^m$, for $c\geq 0$. One can extend the same argument to all functions which have the Taylor series with non-negative coefficients (Steinwart & Christmann, 2008, Lemma 4.8). This leads us to the **exponential kernel** $k_{exp}(x,x')=\exp(2\sigma\langle x,x'\rangle)$, for $\sigma>0$. Furthermore, let $\phi:\mathbb{R}^d\to\mathbb{R}$, $\phi(x)=\exp(-\sigma\|x\|^2)$. Then, since it is representable as an inner product in $\mathbb{R}$ (i.e., ordinary product), $\tilde{k}(x,x')=\phi(x)\phi(x')=\exp(-\sigma\|x\|^2)\exp(-\sigma\|x'\|^2)$ is a kernel on $\mathbb{R}^d$. Therefore, by Theorem 47, so is:

$$\begin{aligned} k_{gauss}(x,x') &= \tilde{k}(x,x')k_{exp}(x,x') \\ &= \exp\left(-\sigma\left[\|x\|^2+\|x'\|^2-2\langle x,x'\rangle\right]\right) \\ &= \exp\left(-\sigma\|x-x'\|^2\right), \end{aligned}$$

which is the **gaussian kernel** on $\mathbb{R}^d$.

6 Mercer representation of RKHS

Moore-Aronszajn theorem gives a construction of an RKHS without imposing any additional assumptions on $\mathcal{X}$ (apart from it being a non-empty set) nor on kernel k (apart from it being a positive definite function). In this Section, we will consider $\mathcal{X}$ to be a compact metric space (with metric $d_\mathcal{X}$) and that $k:\mathcal{X}\times\mathcal{X}\to\mathbb{R}$ is a *continuous* positive definite function. These assumptions will allow us to give an alternative construction and interpretation of RKHS.

6.1 Integral operator of a kernel

Definition 48 (Integral operator)**.** Let k be a continuous kernel on compact metric space $\mathcal{X}$, and let ν be a finite Borel measure on $\mathcal{X}$. Let S_k be the linear

map:

$$\begin{aligned} S_k : L_2(\mathcal{X};\nu) &\rightarrow \mathcal{C}(\mathcal{X}), \\ (S_k f)(x) &= \int k(x,y) f(y) d\nu(y), \quad f \in L_2(\mathcal{X};\nu), \end{aligned}$$

and $T_k = I_k \circ S_k$ its composition with the inclusion $I_k : \mathcal{C}(\mathcal{X}) \hookrightarrow L_2(\mathcal{X};\nu)$. T_k is said to be the *integral operator* of kernel k.

Let us first show that the operator S_k is well-defined, i.e., that $S_k f$ is a continuous function $\forall f \in L_2(\mathcal{X};\nu)$. Indeed, $\forall x, y \in \mathcal{X}$, we have that:

$$\begin{aligned} |(S_k f)(x) - (S_k f)(y)| &= \left| \int (k(x,z) - k(y,z)) f(z) d\nu(z) \right| \\ &= \left| \langle k(x,\cdot) - k(y,\cdot), f \rangle_{L_2} \right| \\ &\leq \|k(x,\cdot) - k(y,\cdot)\|_2 \|f\|_2 \\ &\leq \left[\int (k(x,z) - k(y,z))^2 d\nu(z) \right]^{1/2} \|f\|_2 \\ &\leq \sqrt{\nu(\mathcal{X})} \max_{z \in \mathcal{X}} |k(x,z) - k(y,z)| \, \|f\|_2 . \end{aligned}$$

At this point, we use the fact that k is uniformly continuous on $\mathcal{X} \times \mathcal{X}$ (as it is a continuous function on a compact domain. Namely, $\forall \epsilon > 0$, $\exists \delta = \delta(\epsilon)$, s.t. $d_{\mathcal{X}}(x,y) < \delta$ implies $|k(x,z) - k(y,z)| < \frac{\epsilon}{\sqrt{\nu(\mathcal{X})}\|f\|_2}$, $\forall x, y, z \in \mathcal{X}$. From here,

$$d_{\mathcal{X}}(x,y) < \delta \quad \Rightarrow \quad |(S_k f)(x) - (S_k f)(y)| < \epsilon, \qquad \forall x, y \in \mathcal{X},$$

i.e., $S_k f$ is a continuous function on $\mathcal{X}$.

Note that the operator $T_k : L_2(\mathcal{X};\nu) \rightarrow L_2(\mathcal{X};\nu)$ is distinct from S_k. In particular, while $S_k f$ is a continuous function, $T_k f$ is an equivalence class, so $(S_k f)(x)$ is defined, while $(T_k f)(x)$ is *not*.

The integral operator inherits various properties of the kernel function. In particular, it is readily shown that symmetry of k implies that T_k is a *self-adjoint* operator, i.e., that $\langle f, T_k g \rangle = \langle T_k f, g \rangle$, $\forall f, g \in L_2(\mathcal{X};\nu)$, and that positive definiteness of k implies that T_k is a positive operator, i.e., that $\langle f, T_k f \rangle \geq 0$ $\forall f \in L_2(\mathcal{X};\nu)$. Furthermore, continuity of k implies that T_k is also a *compact operator* - the proof of this requires the use of *Arzela-Ascoli theorem*, which can be found in Rudin (1987, Theorem 11.28, p.245). Thus, one can apply an important result of functional analysis, the spectral theorem, to the operator T_k, which states that any compact, self-adjoint operator can be diagonalized in an appropriate orthonormal basis.

Theorem 49. (Spectral theorem)

Let $\mathcal{F}$ be a Hilbert space, and $T : \mathcal{F} \rightarrow \mathcal{F}$ a compact, self-adjoint operator. There is an at most countable orthonormal set (ONS) $\{e_j\}_{j \in J}$ of $\mathcal{F}$ and $\{\lambda_j\}_{j \in J}$ with $|\lambda_1| \geq |\lambda_2| \geq \cdots > 0$ converging to zero, such that

$$Tf = \sum_{j \in J} \lambda_j \langle f, e_j \rangle_{\mathcal{F}} e_j, \qquad f \in \mathcal{F}.$$

6.2 Mercer's theorem

Let us know fix a finite measure ν on $\mathcal{X}$ with $\text{supp}[\nu] = \mathcal{X}$. Recall that the integral operator T_k is compact, positive and self-adjoint on $L_2(\mathcal{X}; \nu)$, so, by the spectral theorem, there exists ONS $\{\tilde{e}_j\}_{j\in J}$ and the set of eigenvalues $\{\lambda_j\}_{j\in J}$, where J is at most countable set of indices, corresponding to the *strictly positive eigenvalues* of T_k. Note that each $\tilde{e}_j$ is an equivalence class in the ONS of $L_2(\mathcal{X}; \nu)$, but to each equivalence class we can also assign a continuous function $e_j = \lambda_j^{-1} S_k \tilde{e}_j \in \mathcal{C}(\mathcal{X})$. To show that e_j is in the class $\tilde{e}_j$, note that:

$$I_k e_j = \lambda_j^{-1} T_k \tilde{e}_j = \lambda_j^{-1} \lambda_j \tilde{e}_j = \tilde{e}_j.$$

With this notation, the following Theorem holds:

Theorem 50. (Mercer's theorem)
Let k be a continuous kernel on compact metric space $\mathcal{X}$, and let ν be a finite Borel measure on $\mathcal{X}$ with supp$[\nu] = \mathcal{X}$. Then $\forall x, y \in \mathcal{X}$

$$k(x, y) = \sum_{j\in J} \lambda_j e_j(x) e_j(y),$$

and the convergence of the sum is uniform on $\mathcal{X} \times \mathcal{X}$, and absolute for each pair $(x, y) \in \mathcal{X} \times \mathcal{X}$.

Note that Mercer's theorem gives us another feature map for the kernel k, since:

$$\begin{aligned} k(x, y) &= \sum_{j\in J} \lambda_j e_j(x) e_j(y) \\ &= \left\langle \sqrt{\lambda_j} e_j(x), \sqrt{\lambda_j} e_j(y) \right\rangle_{\ell^2(J)}, \end{aligned}$$

so we can take $\ell^2(J)$ as a feature space, and the corresponding feature map is:

$$\begin{aligned} \phi : \mathcal{X} &\to \ell^2(J) \\ \phi : x &\mapsto \left\{ \sqrt{\lambda_j} e_j(x) \right\}_{j\in J}. \end{aligned}$$

This map is well defined as $\sum_{j\in J} \left| \sqrt{\lambda_j} e_j(x) \right|^2 = k(x, x) < \infty$.

Apart from the representation of the kernel function, Mercer theorem also leads to a construction of RKHS using the eigenfunctions of the integral operator T_k. In particular, first note that sum $\sum_{j\in J} a_j e_j(x)$ converges absolutely $\forall x \in \mathcal{X}$ whenever sequence $\left\{ \frac{a_j}{\sqrt{\lambda_j}} \right\} \in \ell^2(J)$. Namely, from the Cauchy-Schwartz

inequality in $\ell^2(J)$, we have that:

$$\begin{aligned}\sum_{j\in J}|a_j e_j(x)| &\leq \left[\sum_{j\in J}\left|\frac{a_j}{\sqrt{\lambda_j}}\right|^2\right]^{1/2}\cdot\left[\sum_{j\in J}\left|\sqrt{\lambda_j}e_j(x)\right|^2\right]^{1/2}\\ &= \left\|\left\{\frac{a_j}{\sqrt{\lambda_j}}\right\}\right\|_{\ell^2}\sqrt{k(x,x)}.\end{aligned}$$

In that case, $\sum_{j\in J}a_j e_j$ is a well defined function on $\mathcal{X}$. The following theorem tells us that the RKHS of k is exactly the space of functions of this form.

Theorem 51. *Let $\mathcal{X}$ be a compact metric space and $k:\mathcal{X}\times\mathcal{X}\to\mathbb{R}$ a continuous kernel. Define:*

$$\mathcal{H} = \left\{f=\sum_{j\in J}a_j e_j \;:\; \left\{\frac{a_j}{\sqrt{\lambda_j}}\right\}\in\ell^2(J)\right\},$$

with inner product:

$$\left\langle\sum_{j\in J}a_j e_j,\sum_{j\in J}b_j e_j\right\rangle_{\mathcal{H}} = \sum_{j\in J}\frac{a_j b_j}{\lambda_j}. \tag{6.1}$$

Then $\mathcal{H}=\mathcal{H}_k$ (they are the same spaces of functions with the same inner product).

Proof. Routine work shows that (6.1) defines an inner product and that $\mathcal{H}$ is a Hilbert space. By Mercer's theorem, $k(\cdot,x)=\sum_{j\in J}(\lambda_j e_j(x))\,e_j$, and:

$$\begin{aligned}\sum_{j\in J}\left|\frac{\lambda_j e_j(x)}{\sqrt{\lambda_j}}\right|^2 &= \sum_{j\in J}\lambda_j e_j^2(x)\\ &= k(x,x)<\infty,\end{aligned}$$

so $k(\cdot,x)\in\mathcal{H}$, $\forall x\in\mathcal{X}$. Furthermore, let $f=\sum_{j\in J}a_j e_j\in\mathcal{H}$ with $\left\{\frac{a_j}{\sqrt{\lambda_j}}\right\}\in\ell^2(J)$. Then,

$$\begin{aligned}\langle f,k(\cdot,x)\rangle_{\mathcal{H}} &= \left\langle\sum_{j\in J}a_j e_j,\sum_{j\in J}(\lambda_j e_j(x))\,e_j\right\rangle_{\mathcal{H}}\\ &= \sum_{j\in J}\frac{a_j\lambda_j e_j(x)}{\lambda_j}\\ &= f(x).\end{aligned}$$

Thus, $\mathcal{H}$ is a Hilbert space of functions with a reproducing kernel k, so it must be equal to $\mathcal{H}_k$ by the uniqueness of RKHS. □

A consequence of the above theorem is that although space $\mathcal{H}$ is defined using the integral operator T_k and its associated eigenfunctions $\{e_j\}_{j\in J}$ which depend on the underlying measure ν, it coincides exactly with the RKHS $\mathcal{H}_k$ of k, so by uniqueness of the RKHS shown in the Moore-Aronszajn theorem, $\mathcal{H}$ actually does not depend on the choice of ν at all.

6.3 Relation between $\mathcal{H}_k$ and $L_2(\mathcal{X};\nu)$

Assume now that $\{\tilde{e}_j\}_{j\in J}$ is an orthonormal basis of $L_2(\mathcal{X};\nu)$, i.e., that all eigenvalues of T_k are strictly positive. Write $\hat{f}(j) = \langle f, \tilde{e}_j\rangle_{L_2}$ for Fourier coefficients of $f \in L_2(\mathcal{X};\nu)$, w.r.t. the basis $\{\tilde{e}_j\}_{j\in J}$. Then,

$$T_k f = \sum_{j\in J} \lambda_j \hat{f}(j)\tilde{e}_j, \qquad f \in L_2(\mathcal{X};\nu),$$

so in the expansion w.r.t. orthonormal basis $\{\tilde{e}_j\}_{j\in J}$, T_k simply scales the Fourier coefficients with respective eigenvalues. The operator $T_k^{1/2}$ for which $T_k = T_k^{1/2} \circ T_k^{1/2}$ is given by:

$$T_k^{1/2} f = \sum_{j\in J} \sqrt{\lambda_j}\hat{f}(j)\tilde{e}_j, \qquad f \in L_2(\mathcal{X};\nu).$$

Note that if we replace classes $\tilde{e}_j$ with their representers $e_j = \lambda_j^{-1} S_k \tilde{e}_j$, we obtain a function in the RKHS, i.e.,

$$\sum_{j\in J} \left|\hat{f}(j)\right|^2 = \|f\|_2^2 < \infty \Rightarrow \left\{\hat{f}(j)\right\} \in \ell^2(J) \quad \Rightarrow \quad \sum_{j\in J} \sqrt{\lambda_j}\hat{f}(j)e_j \in \mathcal{H}_k$$

Thus, $T_k^{1/2}$ induces an isometric isomorphism between $L_2(\mathcal{X};\nu)$ and $\mathcal{H}_k$ (and both are isometrically isomorphic to $\ell^2(J)$). In the case where not all eigenvalues of T_k are strictly positive, $\{\tilde{e}_j\}_{j\in J}$ does not span all of $L_2(\mathcal{X};\nu)$, but $\mathcal{H}_k$ is still isometrically isomorphic to its subspace $span\,\{\tilde{e}_j \ : j \in J\} \subseteq L_2(\mathcal{X};\nu)$.

7 Further results

- Separable RKHS: Steinwart & Christmann (2008, Lemma 4.33)
- Measurability of canonical feature map: Steinwart & Christmann (2008, Lemma 4.25)
- Relation between RKHS and $L_2(\mu)$: Steinwart & Christmann (2008, Theorem 4.26, Theorem 4.27). Note in particular Steinwart & Christmann (2008, Theorem 4.47): the mapping from L_2 to $\mathcal{H}$ for the Gaussian RKHS is injective.

- Expansion of kernel in terms of basis functions: Berlinet & Thomas-Agnan (2004, Theorem 14 p. 32)
- Mercer's theorem: Steinwart & Christmann (2008, p. 150).

8 What functions are in an RKHS?

- Gaussian RKHSs do not contain constants: Steinwart & Christmann (2008, Corollary 4.44).
- Universal RKHSs are dense in the space of bounded continuous functions: Steinwart & Christmann (2008, Section 4.6)
- The bandwidth of the kernel limits the bandwidth of the functions in the RKHS: (c.f., e.g., Appendix of Christian Walder PhD thesis, University of Queensland, 2008).

9 Acknowledgements

Thanks to Sivaraman Balakrishnan and Zoltán Szabó for careful proofreading.

References

Berlinet, A. and Thomas-Agnan, C. *Reproducing Kernel Hilbert Spaces in Probability and Statistics*. Kluwer, 2004.

Kreyszig, E. *Introductory Functional Analysis with Applications*. Wiley, 1989.

Rudin, W. *Real and Complex Analysis*. McGraw-Hill, 3rd edition, 1987.

Steinwart, Ingo and Christmann, Andreas. *Support Vector Machines*. Information Science and Statistics. Springer, 2008.

Wendland, H. *Scattered Data Approximation*. Cambridge University Press, Cambridge, UK, 2005.