What is an RKHS?

Dino Sejdinovic, Arthur Gretton

February 14, 2012

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 Banach space arise from the inner product. The latter is often termed *Riesz representation theorem*.

2.1 Definitions of Banach and Hilbert spaces

Definition 1 (Norm)**.** Let $\mathcal{F}$ be a vector space over the field $\mathbb{K}$. A function $\|\cdot\|_{\mathcal{F}} : \mathcal{F} \to \mathbb{K}$ 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*),
2. $\|\lambda f\|_{\mathcal{F}} = |\lambda| \, \|f\|_{\mathcal{F}}$, $\forall \lambda \in \mathbb{K}$, $\forall f \in \mathcal{F}$ (*positive homogeneity*),
3. $\|f + g\|_{\mathcal{F}} \leq \|f\|_{\mathcal{F}} + \|g\|_{\mathcal{F}}$, $\forall f, g \in \mathcal{F}$ (*triangle inequality*).

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 Cauchy sequence:

Definition 2 (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 $n, m \geq N$, $\|f_n - f_m\|_{\mathcal{F}} < \epsilon$.

Cauchy sequences are always bounded [2, Exercise 4 p. 32], i.e., there exists $M < \infty$, s.t., $\|f_n\|_{\mathcal{F}} \leq M$, $\forall n \in \mathbb{N}$. However, note that not every Cauchy sequence converges: 1, 1.4, 1.41, 1.414, 1.4142, ... is a Cauchy sequence in $\mathbb{Q}$ (a normed vector space over itself) which does not converge - because $\sqrt{2} \notin \mathbb{Q}$.

Next we define a complete space [2, Definition 1.4-3]:

Definition 3 (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 4 (Banach space)**.** Banach space is a complete normed space, i.e., it contains the limits of all its Cauchy sequences.

Note that all elements in a Banach space must have finite norm - if an element has infinite norm, it is not in the space.

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

Definition 5 (Inner product)**.** Let $\mathcal{F}$ be a vector space over the field $\mathbb{K}$ ($\mathbb{R}$ or $\mathbb{C}$). A function $\langle \cdot, \cdot \rangle_{\mathcal{F}} : \mathcal{F} \times \mathcal{F} \to \mathbb{K}$ 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. $\langle f, g \rangle_{\mathcal{F}} = \overline{\langle g, f \rangle}_{\mathcal{F}}$ (complex conjugate)
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 5 are that:

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

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*, $\mathbb{K} = \mathbb{R}$)
- $4\langle f, g \rangle = \|f + g\|^2 - \|f - g\|^2 + i\|f + ig\|^2 - i\|f - ig\|^2$ (*the polarization identity*, $\mathbb{K} = \mathbb{C}$)
- $4\langle f, g \rangle = \|f + g\|^2 - \|f - g\|^2$ (*the polarization identity* if $\mathbb{K} = \mathbb{R}$)

Definition 6 (Hilbert space)**.** Hilbert space is a complete inner product space, i.e., it is a Banach space with an inner product.

As in the Banach space case, all elements in a Hilbert space must have finite norm.

Example 7. For an index set A, the space $\ell^2(A)$ of complex sequences $\{x_\alpha\}_{\alpha\in A}$, satisfying $\sum_{\alpha\in A}|x_\alpha|^2 < \infty$, endowed with the inner product

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

is a Hilbert space.

Example 8. If μ is a positive measure on $\mathcal{X}$, then the space

$$\mathcal{L}^2_\mu(\mathcal{X}) := \left\{ f : X \to \mathbb{C} \text{ measurable} \;\middle|\; \|f\|_2 = \left(\int_{\mathcal{X}} |f(x)|^2 d\mu\right)^{1/2} < \infty \right\} \tag{2.1}$$

is a Hilbert space with inner product

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

Strictly speaking, $\mathcal{L}^2_\mu(\mathcal{X})$ is the space of equivalence classes of functions that differ by at most a set of μ-measure zero.

More Hilbert space examples: [2, p. 132 and 133].

2.2 Bounded/Continuous linear operators

In the following, we take $\mathcal{F}$ and $\mathcal{G}$ to be normed vector[1] spaces over $\mathbb{K}$ (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 9 (Linear operator)**.** A function $A : \mathcal{F} \to \mathcal{G}$, where $\mathcal{F}$ and $\mathcal{G}$ are both normed linear spaces over $\mathbb{K}$, is called a **linear** operator if and only if it satisfies the following properties:

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

Example 10. Let $\mathcal{F}$ be an inner product space. For $g \in \mathcal{F}$, operator $A_g : \mathcal{F} \to \mathbb{K}$, defined with $A_g(f) := \langle f, g\rangle_{\mathcal{F}}$ is a linear operator. Note that the image space of A_g is the underlying field $\mathbb{K}$, which is trivially a normed linear space over itself. Such scalar-valued operators are called *functionals* on $\mathcal{F}$.

[1] A vector space can also be known as a linear space [2, Definition 2.1-1].

Definition 11 (Continuity)**.** A function $A : \mathcal{F} \to \mathcal{G}$ is said to be **continuous** at $f_0 \subset \mathcal{F}$, if for every $\epsilon > 0$, there exists a $\delta = \delta(\epsilon, f_0) > 0$, such that

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

A is **continuous** on $\mathcal{F}$, if it is continuous at every point of $\mathcal{F}$.

In other words, a convergent sequence in $\mathcal{F}$ is mapped to a convergent sequence in $\mathcal{G}$.

Example 12. For $g \in \mathcal{F}$, $A_g : \mathcal{F} \to \mathbb{K}$, defined[2] 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 13 (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 14 (Bounded operator)**.** The linear operator $A : \mathcal{F} \to \mathcal{G}$ is said to be a bounded operator if $\|A\| < \infty$.

It can be shown that A is bounded if and only if 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 that the **smallest** such λ is precisely the operator norm.

It can readily be shown [2] that this satisfies all the requirements of a norm (triangle inequality, zero iff the operator maps only to the zero function, $\|cA\| = |c| \|A\|$ for $c \in \mathbb{K}$), 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

$$\begin{aligned} \frac{\|Af\|_{\mathcal{G}}}{\|f\|_{\mathcal{F}}} &\leq \|A\| \\ \|Af\|_{\mathcal{G}} &\leq \|A\| \|f\|_{\mathcal{F}}. \end{aligned}$$

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 [2, p. 96]: every linear operator on a normed, finite dimensional space is bounded.

Theorem 15. *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:

[2]Here $|\cdot|$ is the norm on $\mathbb{K}$.

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 16 (Algebraic dual)**.** If $\mathcal{F}$ is a normed space, then the space $\mathcal{F}'$ of *linear* functionals $A : \mathcal{F} \to \mathbb{K}$ is called the algebraic dual space of $\mathcal{F}$.

Definition 17 (Topological dual)**.** If $\mathcal{F}$ is a normed space, then the space $\mathcal{F}'$ of *continuous linear* functionals $A : \mathcal{F} \to \mathbb{K}$ is called the topological dual space of $\mathcal{F}$.

In finite-dimensional space, the two notions of dual spaces coincide. However, this is not the case in infinite dimensions. Unless otherwise specified, we refer to the topological dual when discussing the dual of $\mathcal{F}$.

We have seen in Examples 10, 12 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.

Theorem 18. *[Riesz representation]In a Hilbert space $\mathcal{F}$, all continous linear functionals are of the form $\langle \cdot, g \rangle_{\mathcal{F}}$, for some $g \in \mathcal{F}$.*

Note that there is a natural isomorphism $\psi : g \mapsto \langle \cdot, g \rangle_{\mathcal{F}}$ 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.[3]

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[4] of functions mapping from

[3] **AG:** Define isometric isomorphism.

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

some non-empty set $\mathcal{X}$ to the field of complex numbers $\mathbb{C}$. 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 $\|f\|^2_{\mathcal{H}} = \langle f, f\rangle_{\mathcal{H}}$. 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 19 (Evaluation functional)**.** Let $\mathcal{H}$ be a Hilbert space of functions $f : \mathcal{X} \to \mathbb{K}$, defined on a non-empty set $\mathcal{X}$. For a fixed $x \in \mathcal{X}$, map $\delta_x : \mathcal{H} \to \mathbb{K}$, $\delta_x : f \mapsto f(x)$ is called the (Dirac) evaluation functional at x.

It is clear that evaluation functionals are always linear: For $f, g \in \mathcal{H}$ and $\alpha, \beta \in \mathbb{K}$, $\delta_x(\alpha f + \beta g) = (\alpha f + \beta g)(x) = \alpha f(x) + \beta g(x) = \alpha\delta_x(f) + \beta\delta_x(g)$. 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 space are defined [3, Definition 4.18(ii)]:

Definition 20 (Reproducing kernel Hilbert space)**.** A Hilbert space $\mathcal{H}$ of functions $f : \mathcal{X} \to \mathbb{K}$, 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 21. *(Norm convergence in $\mathcal{H}$ implies pointwise convergence)[1, 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* $\lim_{n\to\infty} f_n(x) = f(x)$, $\forall x \in \mathcal{X}$.

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 22. [1, p. 2] If we are *not* in an RKHS, then norm convergence does not necessarily imply pointwise convergence. Let[5] $\mathcal{H} = L_2([0, 1])$, endowed with the metric

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

[5]Note that $L_2([0, 1])$ is a Hilbert space [2].

and consider the sequence of functions $\{q_n\}$, 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. In other words, the evaluation of functions at point 1 is not continuous on the set $\{q_n\}$.

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. Recall that we use $\mathbb{K}$ to denote either $\mathbb{R}$ or $\mathbb{C}$, depending on the case considered.

Definition 23. (Reproducing kernel [1, p. 7])

Let $\mathcal{H}$ be a Hilbert space of $\mathbb{K}$-valued functions defined on a non-empty set $\mathcal{X}$. A function $k : \mathcal{X} \times \mathcal{X} \to \mathbb{K}$ 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? What properties does it have? To answer the first two questions, we will make use of the Riesz representation theorem 18.[6]

Theorem 24 (Existence of the reproducing kernel)**.** *$\mathcal{H}$ is a reproducing kernel Hilbert space (i.e., its evaluation functionals δ_x are continuous linear operators), 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

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

[6] The proof may be found in [**?**, p 346]. **AG**: fix footnote.

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

To prove the other direction, we use [3, Theorem 4.20]. Define $\mathcal{H}'$ as the dual space on $\mathcal{H}$ (Definition 17), and assume $\delta_x \in \mathcal{H}'$, i.e. $\delta_x : \mathcal{F} \to \mathbb{K}$ is a bounded linear functional. The Riesz representation theorem (Theorem 18) states that there exists an element $f_{\delta_x} \in \mathcal{H}$ such that

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

and that there is an isometric anti-linear isomorphism $I : \mathcal{H}' \to \mathcal{H}$ which maps $\delta_x \mapsto f_{\delta_x}$. We define the reproducing kernel of $\mathcal{H}$ as

$$k(x, x') = \langle \delta_x, \delta_{x'} \rangle_{\mathcal{H}'} .$$

This gives us the canonical feature map $k(\cdot, x') = I\delta_{x'}$, since

$$k(x, x') = \langle \delta_x, \delta_{x'} \rangle_{\mathcal{H}'} \underset{\text{(a)}}{=} \langle I\delta_{x'}, I\delta_x \rangle_{\mathcal{H}} \underset{\text{(b)}}{=} \delta_x(I\delta_{x'}) = I\delta_{x'}(x).$$

where in (a) we used the anti-linear isometry, and in (b) we use that $I\delta_x = f_{\delta_x}$. The canonical feature map satisfies the reproducing property,

$$f(x') = \delta_{x'} f = \langle f, I\delta_{x'} \rangle_{\mathcal{H}} = \langle f, k(\cdot, x) \rangle_{\mathcal{H}} ,$$

and 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 [1, Definition 2], [3, Definition 4.15].

Definition 25 (Positive definite functions)**.** A function $h : \mathcal{X} \times \mathcal{X} \to \mathbb{C}$ is positive definite if $\forall n \geq 1$, $\forall (a_1, \dots a_n) \in \mathbb{C}^n$, $\forall (x_1, \dots, x_n) \in \mathcal{X}^n$ for pairwise distinct x_i, x_j,

$$\sum_{i=1}^{n} \sum_{j=1}^{n} a_i \bar{a}_j h(x_i, x_j) \geq 0.$$

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.[7]

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

Lemma 26. *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.*

[7] Note that [4, 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.

Proof.

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

□

Corollary 27. *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 $\phi : x \mapsto k(\cdot,x)$. □

3.3 Feature space, and other kernel properties

This section summarizes the relevant parts of [3, Section 4.1]. Recall that we use $\mathbb{K}$ to denote either $\mathbb{R}$ or $\mathbb{C}$, depending on the case considered.

Following Lemma 26, one can define a *positive definite kernel* (or just kernel), as a function which can be represented via inner product, and that is the approach taken in [3, Section 4.1]:

Definition 28 (*Positive definite kernel*)**.** Let $\mathcal{X}$ be a non-empty set. The function $k : \mathcal{X}\times\mathcal{X} \to \mathbb{K}$ is a positive definite kernel if there exists a $\mathbb{K}$-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 a simple example, 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 29 (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 $f_n : \mathcal{X} \to \mathbb{K}$. Then*

$$k(x_1,x_2) := \sum_{n=1}^{\infty} f_n(x_1)\overline{f_n(x_2)} \tag{3.2}$$

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 (3.2) converges absolutely. Definining $\mathcal{H} := \ell_2$ and $\phi(x) = \left\{\overline{f_n(x)}\right\}$ completes the proof. □

Lemma 30 (Sums of kernels are kernels)**.** *If k, k_1, and k_2 are kernels on $\mathcal{X}$, and $\alpha \geq 0$ is a scalar, then αk, $k_1 + k_2$, and $k_1 \cdot k_2$ 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 for which $\langle \phi(x), \phi(x) \rangle < 0$. Mathematically speaking, these properties give the set of all kernels the structure of a convex cone (not a linear space).

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.

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 kernel, 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 $f \in \mathbb{C}^{\mathcal{X}}$ 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

- $\mathcal{H}_0 \subset \mathcal{H} \subset \mathbb{C}^{\mathcal{X}}$ 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? I.e., is it a Hilbert space?

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 [**?**, Lemma 5].

Lemma 31. *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 $\mathbb{C}$, which is complete, and the sequence is convergent in $\mathbb{C}$.

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

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

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 (5) 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 is [?, Lemma 6].

Lemma 32. *Let $\{f_n\}$ be Cauchy sequence in $\mathcal{H}_0$ converging pointwise to $f \in \mathcal{H}$. If $\lim_{n\to\infty} \langle f_n, f_n \rangle_{\mathcal{H}_0} = \|f_n\|^2_{\mathcal{H}_0} = 0$, 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: $\forall x \in \mathcal{X}$, $f(x) = \lim_{n\to\infty} f_n(x) = \lim_{n\to\infty} \delta_x(f_n) \underset{(a)}{\leq} \lim_{n\to\infty} \|\delta_x\| \|f_n\|_{\mathcal{H}_0} \underset{(b)}{=} 0$, 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 33. *$\mathcal{H}_0$ is dense in $\mathcal{H}$ [?, 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),

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

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

Lemma 34. *The evaluation functionals are continuous on $\mathcal{H}$ [?, 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 (33) 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 35. *$\mathcal{H}$ is complete.*

Let $\{f_n\}$ be any Cauchy sequence in $\mathcal{H}$. Since evaluation functionals are linear continuous on $\mathcal{H}$ by (34), then for any $t \in E$, $\{f_n(t)\}$ is convergent in $\mathbb{C}$ to some $f(t) \in \mathbb{C}$ (since $\mathbb{C}$ is complete, it contains this limit). The question is thus whether the function $f(t)$ 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 (34)), 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 (33) (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. Imporantly, in doing this, we prove that for every positive definite kernel, there corresponds a unique RKHS $\mathcal{H}$.

Theorem 36. (Moore-Aronszajn)

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

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

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

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

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

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

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

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

This is true since

$$\forall x \in \mathcal{X},\ f(x) = \langle f(\cdot), k(x, \cdot)\rangle \underset{(a)}{\leq} \|f\|_{\mathcal{H}_0} k^{1/2}(x, x) = 0,$$

where in (a) we use Cauchy-Schwarz. We now proceed to the main proof.

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

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

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

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

meaning 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 $f_{N_1} = \sum_{i=1}^{k} \alpha_i k(\cdot, x_i)$. Take $N_2 \in \mathbb{N}$, s.t. $|f_n(x_i)| < \frac{\epsilon}{2k|\alpha_i|}$, for $i = 1, \ldots, k$. Now, for $n \geq \max(N_1, N_2)$

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

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,

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

where in (a) we use the defnition of an 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 $\mathcal{H}_0$. □

5 Further results

- Separable RKHS: [3, Lemma 4.33]
- Measurability of canonical feature map: [3, Lemma 4.25]
- Relation between RKHS and $L_2(\mu)$: [3, Theorem 4.26, Theorem 4.27]. Note in particular [3, 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: [1, Theorem 14 p. 32]
- Mercer's theorem: [3, p. 150].

6 What functions are in an RKHS?

- Gaussian RKHSs do not contain constants: [3, Corollary 4.44].
- Universal RKHSs are dense in the space of bounded continuous functions: [3, Section 4.6]
- The bandwidth of the kernel limits the bandwidth of the functions in the RKHS: Walder thesis appendix.

7 Acknowledgements

Thanks to Sivaraman Balakrishnan for careful proofreading.

References

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

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

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

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