
To-do list

1. I think the scale parameter λ would just be absorbed by the norm, which is a single value of interest and that is what is “observed”, and the decomposition $\lambda \cdot c_f$ is not so interesting.	5
2. This is the same for any RKHS?	9
3. Are SE smoother than fBm? Lipschitz continuous. Compact convergence. May be smoother than functions in an fBm RKHS?	10

Contents

2 Reproducing kernel Krein spaces	1
2.1 Preliminaries	1
2.2 Reproducing kernel Hilbert spaces	4
2.3 Reproducing kernel Krein spaces	4
2.4 RKHS building blocks	4
2.4.1 The RKHS of constant functions	5
2.4.2 The canonical (linear) RKHS	5
2.4.3 The fractional Brownian motion RKHS	7
2.4.4 The squared exponential RKHS	10
2.4.5 The Pearson RKHS	11
2.5 Constructing RKKS from existing RKHS	11
2.5.1 Scale of an RKHS	11
2.5.2 The polynomial RKHS	11
2.5.3 The ANOVA RKKS	11
2.6 The Sobolev-Hilbert inner product	11
2.7 Discussion	11

Bibliography	12
List of Figures	13
List of Tables	14
List of Theorems	15
List of Definitions	16
List of Abbreviations	17
List of Symbols	18

Haziq Jamil

Department of Statistics

London School of Economics and Political Science

February 10, 2018

Chapter 2

Reproducing kernel Krein spaces

This chapter provides a concise review of functional analysis, especially on topic of reproducing kernel Hilbert and Krein spaces. In addition, this chapter also describes several reproducing kernel Hilbert space (RKHSs) of interest for the purpose of I-prior modelling. Choosing the appropriate RKHS allows us to fit various models of interest. In I-prior modelling, the kernel defining the RKHS turn out to be negative. In such a case, it is necessary to consider *Krein spaces*, in order to give us the required mathematical platform for I-prior modelling. Krein spaces are simply a generalisation of Hilbert spaces for which the kernels allowed to be non-positive definite in its reproducing kernel space. It is emphasised that a deep knowledge of functional analysis is not necessary for I-prior modelling; the advanced reader may wish to skip Sections 2.1–2.3. Section 2.4 describes the RKHSs and RKKSs of interest.

2.1 Preliminaries

The core study of functional analysis revolves around the treatment of functions as objects in vector spaces over a field¹. Vector spaces, or linear spaces as it is known, are sets for which its elements adhere to a set of rules (axioms) relating to additivity and multiplication by a constant. Additionally, vector spaces are endowed with some kind of structure so as to allow ideas such as closeness and limits to be conceived. Of particular interest to us is the structure brought about by *inner products*, which allow the rigorous mathematical study of various geometrical concepts such as lengths, directions, and orthogonality, among other things. We begin with the definition of an inner product.

¹In this thesis, this will be \mathbb{R} exclusively.

Definition 2.1 (Inner products). Let \mathcal{F} be a vector space over \mathbb{R} . A function $\langle \cdot, \cdot \rangle_{\mathcal{F}} : \mathcal{F} \times \mathcal{F} \rightarrow \mathbb{R}$ is said to be an inner product on \mathcal{F} if all of the following are satisfied:

- **Symmetry:** $\langle f, g \rangle_{\mathcal{F}} = \langle g, f \rangle_{\mathcal{F}}, \forall f, g \in \mathcal{F}$
- **Linearity:** $\langle af_1 + bf_2, g \rangle_{\mathcal{F}} = a\langle f_1, g \rangle_{\mathcal{F}} + b\langle f_2, g \rangle_{\mathcal{F}}, \forall f_1, f_2, g \in \mathcal{F}$ and $\forall a, b \in \mathbb{R}$
- **Non-degeneracy:** $\langle f, f \rangle_{\mathcal{F}} = 0 \Leftrightarrow f = 0$
- **Positive-definiteness:** $\langle f, f \rangle_{\mathcal{F}} \geq 0, \forall f \in \mathcal{F}$

We can always define a *norm* on \mathcal{F} using the inner product as $\|f\|_{\mathcal{F}} = \sqrt{\langle f, f \rangle_{\mathcal{F}}}$. Norms are another form of structure that specifically describes the notion of length. This is defined below.

Definition 2.2 (Norms). Let \mathcal{F} be a vector space over \mathbb{R} . A non-negative function $\|\cdot\|_{\mathcal{F}} : \mathcal{F} \times \mathcal{F} \rightarrow [0, \infty)$ is said to be a norm on \mathcal{F} if all of the following are satisfied:

- **Absolute homogeneity:** $\|\lambda f\|_{\mathcal{F}} = |\lambda| \|f\|_{\mathcal{F}}, \forall \lambda \in \mathbb{R}, \forall f \in \mathcal{F}$
- **Subadditivity:** $\|f + g\|_{\mathcal{F}} \leq \|f\|_{\mathcal{F}} + \|g\|_{\mathcal{F}}, \forall f, g \in \mathcal{F}$
- **Point separating:** $\|f\|_{\mathcal{F}} = 0 \Leftrightarrow f = 0$

The norm $\|\cdot\|_{\mathcal{F}}$ induces a metric (a notion of distance) on \mathcal{F} : $d(f, g) = \|f - g\|_{\mathcal{F}}$. The subadditivity property is also known as the *triangle inequality*. Also note that since $\|-f\|_{\mathcal{F}} = \|f\|_{\mathcal{F}}$, and by the triangle inequality and point separating property we have that $\|f\|_{\mathcal{F}} + \|-f\|_{\mathcal{F}} \geq \|f - f\|_{\mathcal{F}} = \|0\|_{\mathcal{F}} = 0$, which implies non-negativity of norms.

A vector space endowed with an inner product (c.f. norm) is called an inner product space (c.f. normed vector space). As a remark, inner product spaces can always be equipped with a norm, but not always the other way around. With these notions of distances we can then define *Cauchy sequences*. A sequence is said to be Cauchy if the elements of the sequence become arbitrarily close to one another as the sequence progresses.

Definition 2.3 (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 sequence if for every $\epsilon > 0$, $\exists N = N(\epsilon) \in \mathbb{N}$, such that $\forall n, m > N$, $\|f_n - f_m\|_{\mathcal{F}} < \epsilon$.

If the limit of the Cauchy sequence exists within the vector space, then the sequence converges to it. If the vector space contains the limits of all Cauchy sequences (or in other words, if every Cauchy sequence converges), then it is said to be *complete*.

A vector space equipped with a (positive definite) inner product that is also complete is known as a *Hilbert space*. Out of interest, an incomplete inner product space is known as a *pre-Hilbert space*, since its completion with respect to the norm induced by the inner product is a Hilbert space. A complete normed space is called a *Banach space*.

The next few definitions are introduced as a necessary precursor to defining a reproducing kernel Hilbert space. Firstly,

For a space of functions \mathcal{F} on \mathcal{X} , we define the evaluation functional that assigns a value to $f \in \mathcal{F}$ for each $x \in \mathcal{X}$.

Definition 2.4 (Evaluation functional). Let \mathcal{F} be a vector space of functions $f : \mathcal{X} \rightarrow \mathbb{R}$, defined on a non-empty set \mathcal{X} . For a fixed $x \in \mathcal{X}$, the function $\delta_x : \mathcal{F} \rightarrow \mathbb{R}$ as defined by $\delta_x(f) = f(x)$ is called the (Dirac) evaluation functional at x . Evaluation functionals are always linear.

There are two more concepts that we need to cover before defining a reproducing kernel Hilbert/Krein space.

Definition 2.5 (Linear operator). A function $A : \mathcal{F} \rightarrow \mathcal{G}$, where \mathcal{F} and \mathcal{G} are both normed vector spaces over \mathbb{R} , is called a linear operator if and only if it satisfies the following properties:

- **Homogeneity:** $A(af) = aA(f), \forall a \in \mathbb{R}, \forall f \in \mathcal{F}$
- **Additivity:** $A(f + g) = A(f) + A(g), \forall f \in \mathcal{F}, g \in \mathcal{G}$.

Definition 2.6 (Bounded operator). The linear operator $A : \mathcal{F} \rightarrow \mathcal{G}$ between two normed spaces $(\mathcal{F}, \|\cdot\|_{\mathcal{F}})$ and $(\mathcal{G}, \|\cdot\|_{\mathcal{G}})$ is said to be a bounded operator if $\exists \lambda \in [0, \infty)$ such that

$$\|A(f)\|_{\mathcal{G}} < \lambda \|f\|_{\mathcal{F}}.$$

Now we define a reproducing kernel Hilbert space.

Definition 2.7 (Reproducing kernel Hilbert space). A Hilbert space of real-valued functions $f : \mathcal{X} \rightarrow \mathbb{R}$ on a non-empty set \mathcal{X} is called a reproducing kernel Hilbert space if the evaluation functional $\delta_x : f \mapsto f(x)$ is bounded (equivalently, continuous²), i.e. $\exists \lambda_x \geq 0$ such that $\forall f \in \mathcal{F}$,

$$|f(x)| = |\delta_x(f)| \leq \lambda_x \|f\|_{\mathcal{F}}.$$

²For any two function $f, g \in \mathcal{F}$, $|f(x) - g(x)| = |\delta_x(f) - \delta_x(g)| = |\delta_x(f - g)| \leq \lambda_x \|f - g\|_{\mathcal{F}}$ for some

Theorem 2.1 (Representation theorem). *Every continuous linear functional f on a Hilbert space \mathcal{H} has the form*

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

with a unique $y \in \mathcal{M}$ and $\|f\| = \|y\|_{\mathcal{H}}$.

Theorem 2.2 (Orthogonal decomposition). *Let \mathcal{H} be a Hilbert space and $\mathcal{M} \subset \mathcal{H}$ be a closed subspace. For every $x \in \mathcal{H}$, we can write*

$$x = y + z$$

where $y \in \mathcal{M}$ and $z \in \mathcal{M}^{\perp}$, and y and z are uniquely determined by x .

Corollary 2.2.1. *Let \mathcal{M} be a subspace of a Hilbert space \mathcal{H} . Then, $\mathcal{M}^{\perp} = \{0\}$ if and only if \mathcal{M} is dense in \mathcal{H} .*

https://en.wikibooks.org/wiki/Functional_Analysis/Hilbert_spaces

In infinite-dimensional Hilbert spaces, some subspaces are not closed, but all orthogonal complements are closed. In such spaces, the orthogonal complement of the orthogonal complement of \mathcal{H} is the closure of \mathcal{H} , i.e. $(\mathcal{H}^{\perp})^{\perp} = \overline{\mathcal{H}}$. If \mathcal{M} is a closed linear subspace of \mathcal{H} , then $\mathcal{H} = \mathcal{M} \oplus \mathcal{M}^{\perp}$.

2.2 Reproducing kernel Hilbert spaces

2.3 Reproducing kernel Krein spaces

2.4 RKHS building blocks

In what follows, each of the kernel functions will have its associated scale parameter denoted by λ . Further, to make the distinction between centred and non-centred versions of the kernels, we use the notation h to denote the uncentred version, and \bar{h} to denote the centred version.

$\lambda_x \geq 0$, thus is said to be Lipschitz continuous, which implies uniform continuity. This property implies pointwise convergence from norm convergence in \mathcal{F} .

2.4.1 The RKHS of constant functions

The vector space of constant functions \mathcal{F} over a set \mathcal{X} contains the functions $f : \mathcal{X} \rightarrow \mathbb{R}$ such that $f(x) = c_f \in \mathbb{R}, \forall x \in \mathcal{X}$. These functions would be useful to model an overall average, i.e. an “intercept effect”. The space \mathcal{F} can be equipped with a norm to form an RKHS, as shown in the following lemma.

Proposition 2.3 (RKHS of constant functions). *The space \mathcal{F} as described above endowed with the norm $\|f\|_{\mathcal{F}} = |c_f|$ forms an RKHS with the reproducing kernel $h : \mathcal{X} \times \mathcal{X} \rightarrow \mathbb{R}$ as defined, rather simply by,*

$$h(x, x') = 1,$$

known as the constant kernel.

Proof. If \mathcal{F} is an RKHS with kernel h as described, then \mathcal{F} is spanned by the functions $h(\cdot, x) = 1$, so it is clear that \mathcal{F} consists of constant functions over \mathcal{X} . On the other hand, if the space \mathcal{F} is equipped with the inner product $\langle f, f' \rangle_{\mathcal{F}} = c_f c_{f'}$, then the reproducing property follows, since $\langle f, h(\cdot, x) \rangle_{\mathcal{F}} = c_f = f(x)$. Hence, $\|f\|_{\mathcal{F}} = \sqrt{\langle f, f \rangle_{\mathcal{F}}} = |c_f|$. \square

In I-prior modelling, one need not consider any scale parameter on reproducing kernel, as the scale parameter would not be identified otherwise. See later chapter for details.

I think the scale parameter λ would just be absorbed by the norm, which is a single value of interest and that is what is “observed”, and the decomposition $\lambda \cdot c_f$ is not so interesting.

2.4.2 The canonical (linear) RKHS

Consider a function space \mathcal{F} over \mathcal{X} which consists of functions of the form $f_{\beta} : \mathcal{X} \rightarrow \mathbb{R}$, $f_{\beta} : x \mapsto \langle x, \beta \rangle_{\mathcal{X}}$ for some $\beta \in \mathbb{R}$. Suppose that $\mathcal{X} \equiv \mathbb{R}^p$, then \mathcal{F} consists of the linear functions $f_{\beta}(x) = x^{\top} \beta$. More generally, if \mathcal{X} is a Hilbert space, then its continuous dual consists of elements of the form $f_{\beta} = \langle \cdot, \beta \rangle_{\mathcal{X}}$. We can show that the continuous dual space of \mathcal{X} is a RKHS which consists of these linear functions.

Proposition 2.4 (The canonical RKHS). *The continuous dual space a Hilbert space \mathcal{X} , denoted by \mathcal{X}' , is a RKHS of linear functions over \mathcal{X} of the form $\langle \cdot, \beta \rangle_{\mathcal{X}}, \beta \in \mathcal{X}$. Its*

reproducing kernel $h : \mathcal{X} \times \mathcal{X} \rightarrow \mathbb{R}$ is defined by

$$h(x, x') = \langle x, x' \rangle_{\mathcal{X}}.$$

Proof. Define $f_\beta := \langle \cdot, \beta \rangle_{\mathcal{X}}$ for some $\beta \in \mathcal{X}$. Clearly this is linear and continuous, so $f_\beta \in \mathcal{X}'$, and so \mathcal{X}' is a Hilbert space containing functions $f : \mathcal{X} \rightarrow \mathbb{R}$ of the form $f_\beta(x) = \langle x, \beta \rangle_{\mathcal{X}}$. By the Riesz representation theorem, every element of \mathcal{X}' has the form f_β . It also gives us a natural isometric isomorphism such that the following is true:

$$\langle \beta, \beta' \rangle_{\mathcal{X}} = \langle f_\beta, f_{\beta'} \rangle_{\mathcal{X}'}$$

Hence, for any $f_\beta \in \mathcal{X}'$,

$$\begin{aligned} f_\beta(x) &= \langle x, \beta \rangle_{\mathcal{X}} \\ &= \langle f_x, f_\beta \rangle_{\mathcal{X}'} \\ &= \langle \langle \cdot, x \rangle_{\mathcal{X}}, f_\beta \rangle_{\mathcal{X}'}. \end{aligned}$$

Thus, $h : \mathcal{X} \times \mathcal{X} \rightarrow \mathbb{R}$ as defined by $h(x, x') = \langle x, x' \rangle_{\mathcal{X}}$ is the reproducing kernel of \mathcal{X}' . \square

In many other literature, the kernel $h(x, x') = \langle x, x' \rangle_{\mathcal{X}}$ is also known as the *linear kernel*. The use of the term ‘canonical’ is fitting not just due to the relation between a Hilbert space and its continuous dual space. Let $\phi : \mathcal{X} \rightarrow \mathcal{V}$ be the feature map from the space of covariates (inputs) to some feature space \mathcal{V} . Suppose both \mathcal{X} and \mathcal{V} is a Hilbert space, then a kernel is defined as

$$h(x, x') = \langle \phi(x), \phi(x') \rangle_{\mathcal{V}}.$$

Taking the feature map to be $\phi(x) = \langle \cdot, x \rangle_{\mathcal{X}}$, we can prove the reproducing property to obtain $h(x, x') = \langle x, x' \rangle_{\mathcal{X}}$, which implies $\phi(x) = h(\cdot, x)$, and thus ϕ is the *canonical feature map* (Steinwart and Christmann, 2008, Lemma 4.19).

The origin of a Hilbert space may be arbitrary, in which case a centring may be appropriate. We define the centred canonical RKHS as follows.

Definition 2.8 (Centred canonical RKHS). Let \mathcal{X} be a Hilbert space, P be a probability measure over \mathcal{X} , and $\mu \in \mathcal{X}$ be the mean (i.e. $E\langle x, x' \rangle_{\mathcal{X}} = \langle \mu, x' \rangle_{\mathcal{X}}$ for all $x' \in \mathcal{X}$) with respect to this probability measure. Define $(\mathcal{X} - \mu)'$, the continuous dual space of $\mathcal{X} - \mu$,

to be the *centred canonical RKHS*. $(\mathcal{X} - \mu)'$ consists of the centred linear functions $f_\beta(x) = \langle x - \mu, \beta \rangle_{\mathcal{X}}$, for $\beta \in \mathcal{X}$, such that $\mathbb{E} f_\beta(x) = 0$. The reproducing kernel of $(\mathcal{X} - \mu)'$ is

$$h(x, x') = \langle x - \mu, x' - \mu \rangle_{\mathcal{X}}.$$

Proof. Proof of the claim $\mathbb{E} f_\beta(x) = 0$:

$$\begin{aligned} \mathbb{E} f_\beta(x) &= \mathbb{E} \langle x - \mu, \beta \rangle_{\mathcal{X}} \\ &= \mathbb{E} \langle x, \beta \rangle_{\mathcal{X}} - \langle \mu, \beta \rangle_{\mathcal{X}}, \end{aligned}$$

and since $\mathbb{E} \langle x, \beta \rangle_{\mathcal{X}} = \langle \mu, \beta \rangle_{\mathcal{X}}$ for any $\beta \in \mathcal{X}$, the results follows. \square

Remark 1. In practice, the probability measure \mathbb{P} over \mathcal{X} is unknown, so we may use the empirical distribution over \mathcal{X} , so that \mathcal{X} is centred by the sample mean $\hat{\mu} = \frac{1}{n} \sum_{i=1}^n x_i$.

2.4.3 The fractional Brownian motion RKHS

Brownian motion (also known as the Wiener process) has been an inquisitive subject in the mathematical sciences, and here, we describe a function space influenced by a generalised version of Brownian motion paths.

Suppose $B_\gamma(t)$ is a continuous-time Gaussian process on $[0, T]$, i.e. for any finite set of indices t_1, \dots, t_k , where each $t_j \in [0, T]$, $(B_\gamma(t_1), \dots, B_\gamma(t_k))$ is a multivariate normal random variable. $B_\gamma(t)$ is said to be a *fractional Brownian motion* (fBm) if $\mathbb{E} B_\gamma(t) = 0$ for all $t \in [0, T]$ and

$$\text{Cov}(B_\gamma(t), B_\gamma(s)) = \frac{1}{2}(|t|^{2\gamma} + |s|^{2\gamma} - |t - s|^{2\gamma}) \quad \forall t, s \in [0, T],$$

where $\gamma \in (0, 1)$ is called the Hurst index or Hurst parameter. Introduced by [Mandelbrot and Van Ness \(1968\)](#), fBms are a generalisation of Brownian motion. The Hurst parameter plays two roles: 1) It describes the raggedness of the resultant motion, with higher values leading to smoother motion; and 2) it determines the type of process the fBm is, as past increments of $B_\gamma(t)$ are weighted by $(t - s)^{\gamma-1/2}$. When $\gamma = 1/2$ exactly, then the fBm is a standard Brownian motion and its increments are independent; when $\gamma > 1/2$ ($\gamma < 1/2$) its increments are positively (negatively) correlated.

Let \mathcal{X} be a Hilbert space. Defining a kernel function $h : \mathcal{X} \times \mathcal{X} \rightarrow \mathbb{R}$ identical to the fBm covariance kernel yields the so-called *fractional Brownian motion RKHS*.

Definition 2.9 (Fractional Brownian motion RKHS). The fractional Brownian motion (fBm) RKHS \mathcal{F} is the space of functions on the Hilbert space \mathcal{X} possessing the reproducing kernel $h : \mathcal{X} \times \mathcal{X} \rightarrow \mathbb{R}$ defined by

$$h(x, x') = \frac{1}{2} (\|x\|_{\mathcal{X}}^{2\gamma} + \|x'\|_{\mathcal{X}}^{2\gamma} - \|x - x'\|_{\mathcal{X}}^{2\gamma}),$$

which depends on the Hurst coefficient $\gamma \in (0, 1)$. We shall reference this space as the fBm- γ RKHS.

Remark 2. When $\gamma = 1$, by the polarisation identity we get $h(x, x') = \langle x, x' \rangle_{\mathcal{X}}$, which is the (reproducing) kernel of the canonical RKHS.

From its construction, it is clear that the fBm kernel is positive definite, and thus defines an RKHS. That the fBm RKHS describes a space of functions is proved in [Cohen \(2002\)](#), who studied this space in depth. It is also noted in the collection of examples of [Berlinet and Thomas-Agnan \(2011, pp.71 & 319\)](#).

The Hurst coefficient γ controls the “smoothness” of the functions in the RKHS. We can talk about smoothness in the context of Hölder continuity of functions.

Definition 2.10 (Hölder condition). A function f over a set $(\mathcal{X}, \|\cdot\|_{\mathcal{X}})$ is said to be *Hölder continuous* of order $0 < \gamma \leq 1$ if there exists a $C > 0$ such that $\forall x, x' \in \mathcal{X}$,

$$|f(x) - f(x')| \leq C \|x - x'\|^{\gamma}.$$

Functions in the Hölder space $C^{k,\gamma}(\mathcal{X})$, where $k \geq 0$ is an integer, consists of those functions over \mathcal{X} having continuous derivatives up to order k and such that the k th partial derivatives are Hölder continuous of order γ . Unlike realisations of actual fBm paths with Hurst index γ , which are well-known to be almost surely Hölder continuous of order less than γ ([Embrechts and Maejima, 2002, Theorem 4.1.1](#)), functions in its namesake RKHS are strictly smoother.

Claim 2.5. *The fBm- γ RKHS \mathcal{F} of functions over $(\mathcal{X}, \|\cdot\|_{\mathcal{X}})$ are Hölder continuous of order γ .*

Proof. For some $f \in \mathcal{F}$ we have $f(x) = \langle f, h(\cdot, x) \rangle_{\mathcal{F}}$ by the reproducing property of the

kernel h of \mathcal{F} . It follows from the Cauchy-Schwarz inequality that for any $x, x' \in \mathcal{X}$,

$$\begin{aligned} |f(x) - f(x')| &= |\langle f, h(\cdot, x) - h(\cdot, x') \rangle_{\mathcal{F}}| \\ &\leq \|f\|_{\mathcal{F}} \cdot \|h(\cdot, x) - h(\cdot, x')\|_{\mathcal{F}} \\ &= \|f\|_{\mathcal{F}} \cdot \|x - x'\|_{\mathcal{X}}^{\gamma}, \end{aligned}$$

since

$$\begin{aligned} \|h(\cdot, x) - h(\cdot, x')\|_{\mathcal{F}}^2 &= \|h(\cdot, x)\|_{\mathcal{F}}^2 + \|h(\cdot, x')\|_{\mathcal{F}}^2 - 2\langle h(\cdot, x), h(\cdot, x') \rangle_{\mathcal{F}} \\ &= h(x, x) + h(x', x') - 2h(x, x') \\ &= \|x - x'\|_{\mathcal{X}}^{2\gamma}, \end{aligned}$$

and thus proving the claim. \square

The fBm- γ RKHS is spanned by the functions $h(\cdot, x)$, which means that $f(0) = 0$ for all $f \in \mathcal{F}$, which may be undesirable. We define the centred fBm RKHS as follows.

Definition 2.11 (Centred fBm RKHS). Let \mathcal{X} be a Hilbert space, P be a probability measure over \mathcal{X} , and $\mu \in \mathcal{X}$ be the mean (i.e. $E\langle x, x' \rangle_{\mathcal{X}} = \langle \mu, x' \rangle_{\mathcal{X}}$ for all $x' \in \mathcal{X}$) with respect to this probability measure. The kernel $\bar{h} : \mathcal{X} \times \mathcal{X} \rightarrow \mathbb{R}$ defined by

$$\bar{h}(x, x') = \frac{1}{2} E \left[\|x - X\|_{\mathcal{X}}^{2\gamma} + \|x' - X'\|_{\mathcal{X}}^{2\gamma} - \|x - x'\|_{\mathcal{X}}^{2\gamma} - \|X - X'\|_{\mathcal{X}}^{2\gamma} \right]$$

is the reproducing kernel of the *centred* fBm- γ RKHS, which consists of functions f in the fBm- γ RKHS such that $E f(X) = 0$. In the above definition, $X, X' \sim P$ are two independent copies of a random vector $X \in \mathcal{X}$.

Remark 3. Again, when $\gamma = 1$, we get the reduction

$$\begin{aligned} \bar{h}(x, x') &= \frac{1}{2} E \left[\|x - X\|_{\mathcal{X}}^2 + \|x' - X'\|_{\mathcal{X}}^2 - \|x - x'\|_{\mathcal{X}}^2 - \|X - X'\|_{\mathcal{X}}^2 \right] \\ &= \frac{1}{2} E \left[\langle X, X \rangle_{\mathcal{X}} + \langle X', X' \rangle_{\mathcal{X}} + 2\langle x, x' \rangle_{\mathcal{X}} - 2\langle x, X \rangle_{\mathcal{X}} - 2\langle x', X' \rangle_{\mathcal{X}} \right] \\ &= \langle \mu, \mu \rangle_{\mathcal{X}} + \langle x, x' \rangle_{\mathcal{X}} - \langle x, \mu \rangle_{\mathcal{X}} - \langle \mu, x' \rangle_{\mathcal{X}} \\ &= \langle x - \mu, x' - \mu \rangle_{\mathcal{X}}, \end{aligned}$$

which is the (reproducing) kernel of the centred canonical RKHS.

2. This is the same for any RKHS?

2.4.4 The squared exponential RKHS

The squared exponential (SE) kernel function is indeed known to be the default kernel used for Gaussian process regression in machine learning. It is a positive definite function, and hence defines an RKHS. The definition of the SE RKHS is as follows.

Definition 2.12 (Squared exponential RKHS). The squared exponential (SE) RKHS \mathcal{F} of functions over some set $\mathcal{X} \subset \mathbb{R}^p$ equipped with the 2-norm $\|\cdot\|_2$ is defined by the positive definite kernel $h : \mathcal{X} \times \mathcal{X} \rightarrow \mathbb{R}$

$$h(x, x') = \exp\left(-\frac{\|x - x'\|_2^2}{2l^2}\right).$$

The parameter l is called the *lengthscale* parameter, and is a smoothing parameter for the functions in the RKHS.

It is known by many other names, including the Gaussian kernel, due to its semblance to the kernel of the Gaussian pdf. Especially in the machine learning literature, the term Gaussian radial basis functions (RBF) is used, and commonly the simpler parameterisation $\gamma = 1/2l^2$ is utilised. [Duvenaud \(2014\)](#) remarks that “exponentiated quadratic” is a more fitting descriptive name for this kernel.

Despite being used extensively for learning algorithms using kernels, an explicit study of the RKHS defined by the SE kernel was not done until recently by [Steinwart, Hush, et al. \(2006\)](#). In that work, the authors describe the nature of real-valued functions in the SE RKHS by considering a real restriction on the SE RKHS of functions over complex values. Their derivation of an orthonormal basis of such an RKHS proved the SE kernel to be the reproducing kernel for the SE RKHS.

Are SE smoother than fBm? Lipschitz continuous. Compact convergence. May be smoother than functions in an fBm RKHS?

SE kernels are known to be “universal”. That is, it satisfied the following definition of universal kernels due to [Micchelli et al. \(2006\)](#).

Definition 2.13 (Universal kernel). Let $C(\mathcal{X})$ is the space of all continuous, complex-valued functions $f : \mathcal{X} \rightarrow \mathbb{C}$ equipped with the maximum norm $\|\cdot\|_\infty$, and denote $\mathcal{K}(\mathcal{X})$ as the space of *kernel sections* $\overline{\text{span}}\{h(\cdot, x) | x \in \mathcal{X}\}$, where here, h is a complex-valued kernel function. A kernel h is said to be *universal* if given any compact subset $\mathcal{Z} \subset \mathcal{X}$, any positive number ϵ and any function $f \in C(\mathcal{Z})$, there is a function $g \in \mathcal{K}(\mathcal{Z})$ such

that $\|f - g\|_{\mathcal{Z}} \leq \epsilon$.

The consequence of this universal property vis-à-vis regression modelling is that any (continuous) regression function f may be approximated very well by a function \hat{f} from the SE RKHS, and these two functions can get arbitrarily close to each other in the max norm sense. This, together with some very convenient computational advantages that the SE kernel brings (more on this in a later chapter), is a testament to the popularity of SE kernels.

2.4.5 The Pearson RKHS

2.5 Constructing RKKS from existing RKHS

2.5.1 Scale of an RKHS

2.5.2 The polynomial RKHS

2.5.3 The ANOVA RKKS

2.6 The Sobolev-Hilbert inner product

2.7 Discussion

Bibliography

- Berlinet, A. and C. Thomas-Agnan (2011). *Reproducing Kernel Hilbert Spaces in Probability and Statistics*. Springer-Verlag. DOI: 10.1007/978-1-4419-9096-9.
- Cohen, S. (2002). “Champs localement auto-similaires”. In: *Lois d’échelle, fractales et ondelettes*. Ed. by P. Abry, P. Gonçalves, and J. L. Véhel. Vol. 1. Hermès Sciences Publications.
- Duvenaud, D. (2014). “Automatic model construction with Gaussian processes”. PhD thesis. University of Cambridge.
- Embrechts, P. and M. Maejima (2002). *Selfsimilar Processes. Princeton series in applied mathematics*. Princeton University Press, Princeton, NJ.
- Mandelbrot, B. B. and J. W. Van Ness (1968). “Fractional Brownian motions, fractional noises and applications”. In: *SIAM review* 10.4, pp. 422–437.
- Micchelli, C. A., Y. Xu, and H. Zhang (2006). “Universal kernels”. In: *Journal of Machine Learning Research* 7.Dec, pp. 2651–2667.
- Steinwart, I. and A. Christmann (2008). *Support vector machines*. Springer Science & Business Media.
- Steinwart, I., D. Hush, and C. Scovel (2006). “An explicit description of the reproducing kernel Hilbert spaces of Gaussian RBF kernels”. In: *IEEE Transactions on Information Theory* 52.10, pp. 4635–4643.

List of Figures

List of Tables

List of Theorems

2.1	Theorem (Representation theorem)	4
2.2	Theorem (Orthogonal decomposition)	4

List of Definitions

2.1	Definition (Inner products)	2
2.2	Definition (Norms)	2
2.3	Definition (Cauchy sequence)	2
2.4	Definition (Evaluation functional)	3
2.5	Definition (Linear operator)	3
2.6	Definition (Bounded operator)	3
2.7	Definition (Reproducing kernel Hilbert space)	3
2.8	Definition (Centred canonical RKHS)	6
2.9	Definition (Fractional Brownian motion RKHS)	8
2.10	Definition (Hölder condition)	8
2.11	Definition (Centred fBm RKHS)	9
2.12	Definition (Squared exponential RKHS)	10
2.13	Definition (Universal kernel)	10

List of Abbreviations

RKHS	Reproducing kernel Hilbert space.
SE	Squared exponential (kernel).

List of Symbols

$N_p(\mu, \Sigma)$ p -dimensional multivariate normal distribution with mean vector μ and covariance Σ .

\sim Is distributed as.

\otimes The tensor product.

Index

fractional Brownian motion, *see* fBm

reproducing kernel Hilbert space, *see*
RKHS

RKHS, 1, **17**

SE, 10, **17**

squared exponential, *see* SE