To-do list

Contents

Haziq Jamil
Department of Statistics
London School of Economics and Political Science
February 9, 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 it 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 field[1]. 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.

[1]In 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 $\langle\cdot,\cdot\rangle_{\mathcal{F}}$: $\mathcal{F}\times\mathcal{F}\to\mathbb{R}$ 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 $||\cdot||_{\mathcal{F}}:\mathcal{F}\times\mathcal{F}\to[0,\infty)$ 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 $f : \mathcal{X} \to \mathbb{R}$, defined on a non-empty set $\mathcal{X}$. For a fixed $x \in \mathcal{X}$, the function $\delta_x : \mathcal{F} \to \mathbb{R}$ 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 $A : \mathcal{F} \to \mathcal{G}$, 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 (Bounder operator)**.** The linear operator $A : \mathcal{F} \to \mathcal{G}$ 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 $f : \mathcal{X} \to \mathbb{R}$ 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, continuous[2]), i.e. $\exists \lambda_x \geq 0$ such that $\forall f \in \mathcal{F}$,

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

[2]For 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 $f : \mathcal{X} \to \mathbb{R}$ 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 $h : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ 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 $f_\beta : \mathcal{X} \to \mathbb{R}$, $f_\beta : x \mapsto \langle x, \beta \rangle_\mathcal{X}$ 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 $h : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ 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 $f : \mathcal{X} \to \mathbb{R}$ 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 $f_\beta \in \mathcal{X}'$,

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

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

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 $\phi : \mathcal{X} \to \mathcal{V}$ 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. $\mathrm{E}\langle x, x' \rangle_{\mathcal{X}} = \langle \mu, x' \rangle_{\mathcal{X}}$ 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 $\operatorname{E} f_\beta(x) = 0$. The reproducing kernel of $(\mathcal{X} - \mu)'$ is

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

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

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

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

Remark 1. In practice, the probability measure 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

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 $t_j \in [0, T]$, $\big(B_\gamma(t_1), \dots, B_\gamma(t_k)\big)$ is a multivariate normal random variable. $B_\gamma(t)$ is said to be a *fractional Brownian motion* (fBm) if $\operatorname{E} B_\gamma(t) = 0$ for all $t \in [0, T]$ and

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

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 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 $h : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ identical to the fBm covariance kernel yields the so-called *fractional Brownian motion RKHS*.

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

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

which depends on the Hurst coefficient $\gamma \in (0,1)$.

It is clear that the fBm kernel is positive definite by construction, and thus induces 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).

2.4.4 The squared exponential RKHS

Are SE smoother than fBm? Mainly used because of the "universal approximation" property 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.

Mandelbrot, B. B. and J. W. Van Ness (1968). “Fractional Brownian motions, fractional noises and applications”. In: *SIAM review* 10.4, pp. 422–437.

Steinwart, I. and A. Christmann (2008). *Support vector machines*. Springer Science & Business Media.

List of Figures

List of Tables

List of Theorems

List of Definitions

List of Abbreviations

RKHS Reproducing kernel Hilbert space.

List of Symbols

$\mathrm{N}_p(\mu, \Sigma)$	p-dimensional multivariate normal distribution with mean vector μ and covariance Σ.
$\sim$	Is distributed as.
$\otimes$	The tensor product.

Index