INSA – Gaussian processes

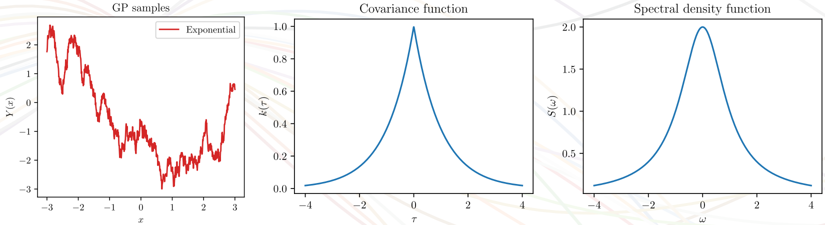
An introduction to reproducing kernel Hilbert-spaces (RKHS)

Andrés F. López-Lopera

The French Aerospace Lab ONERA, France
Information Processing and Systems Department (DTIS)
Multidisciplinary Methods, Integrated Concepts (M2CI) Research Unit

Kernel embedding

Kernel functions



- Every kernel k is the covariance function of some centred Gaussian stochastic process Y: e.g. *Ornstein-Uhlenbeck process*
- Any symmetric and p.s.d. function is a valid kernel
- Every spectral density $S(\omega)$ defines a (stationary) kernel

Kernel embedding

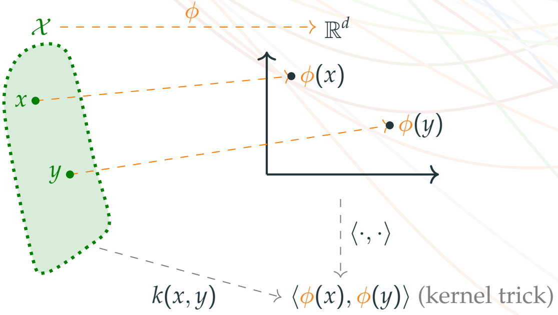
- Kernels can also be defined in more general normed vector spaces

Theorem (Kernel embedding)

*A function $k : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ is a kernel iff there exists a **Hilbert space** $\mathcal{H}$ and a map $\phi : \mathcal{X} \to \mathcal{H}$ such that*

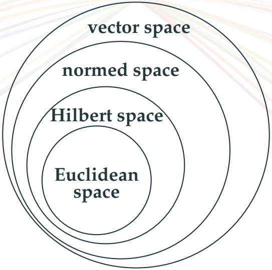
$$k(x, y) = \langle \phi(x), \phi(y) \rangle_{\mathcal{H}}, \qquad \textit{for } \; x, y \in \mathcal{X}.$$

- For instance, just think about the space $\mathbb{R}^d$



Hilbert space

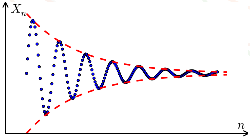
- A **Hilbert space** $\mathcal{H}$ is a natural extension of the usual Euclidean space.



- It is used to quantify distances for abstract objects: functions, probabilities, sequences, etc.
- $\mathcal{H}$ is a vector space with a scalar product

$$\langle \cdot, \cdot \rangle_{\mathcal{H}},$$

  and a norm $\|f\|^2_{\mathcal{H}} = \langle f, f \rangle_{\mathcal{H}}$.
- $\mathcal{H}$ is complete, i.e. all Cauchy sequences are convergent



Hilbert space

- Examples of finite-dimensional Hilbert spaces are:
  - The real numbers $\mathbb{R}^d$ with $\langle \boldsymbol{v}, \boldsymbol{u} \rangle$ the scalar product of $\boldsymbol{v}$ and $\boldsymbol{u}$.
  - The complex numbers $\mathbb{C}^d$ with $\langle \boldsymbol{v}, \boldsymbol{u} \rangle$ the scalar product of $\boldsymbol{v}$ and $\overline{\boldsymbol{u}}$.
- An example of an infinite-dimensional Hilbert spaces is:
  - The set of functions $f : \mathbb{R} \to \mathbb{R}$ such that $\int_{-\infty}^{\infty} f^2(x)dx < \infty$. In this case,

$$\langle f, g \rangle = \int_{-\infty}^{\infty} f(x)g(x)dx.$$

- Hilbert spaces are a powerful tool for studying linear prediction problems
- See further discussion in [Stein, 1999, Section 1.3]

Theorem (Kernel embedding (continue))

A function $k : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ is a kernel iff there exists a ***Hilbert space*** *$\mathcal{H}$ and a map $\phi : \mathcal{X} \to \mathcal{H}$ such that*

$$k(x, y) = \langle \phi(x), \phi(y) \rangle_{\mathcal{H}}, \qquad \textit{for } \; x, y \in \mathcal{X}.$$

Proof.

$\Leftarrow$ Let $\mathcal{H}$ be a Hilbert space and $\phi : \mathcal{X} \to \mathcal{H}$ be a feature map. Then, $k(x, y) = \langle \phi(x), \phi(y) \rangle_{\mathcal{H}}$ is a kernel by definition:

- $\langle \cdot, \cdot \rangle_{\mathcal{H}}$ is symmetric: $\langle \phi(x), \phi(y) \rangle_{\mathcal{H}} = \langle \phi(y), \phi(x) \rangle_{\mathcal{H}}$
- positive definiteness of scalar products:

$$\sum_{i,j=1}^{n} a_i a_j k(x_i, x_j) = \sum_{i,j=1}^{n} a_i a_j \langle \phi(x_i), \phi(x_j) \rangle = \langle \sum_{i=1}^{n} a_i \phi(x_i), \sum_{j=1}^{n} a_j \phi(x_j) \rangle \geq 0$$

ONERA
THE FRENCH AEROSPACE LAB

Proof (continue).

⇒ We have to prove that, given $\mathcal{X}$ and k, there exists a vector space $\mathcal{H}$ with a scalar product $\langle \cdot, \cdot \rangle_{\mathcal{H}}$ and a mapping $\phi : \mathcal{X} \to \mathcal{H}$ such that

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

for all $x, x' \in \mathcal{X}$.

· The ⇒ of the proof relies on reproducing kernel Hilbert spaces (RKHS)

Reproducing kernel Hilbert space (RKHS)

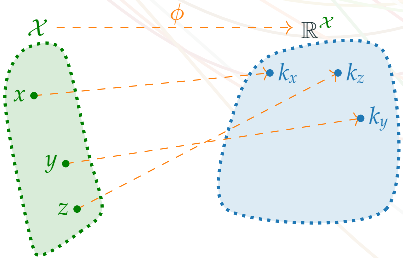
Definition of the vector space

We are going to use a space of functions:

- Consider a mapping $\phi : \mathcal{X} \to \mathbb{R}^{\mathcal{X}}$ (where $\mathbb{R}^{\mathcal{X}}$ denotes the space of all real-valued functions from $\mathcal{X}$ to $\mathbb{R}$), defined as

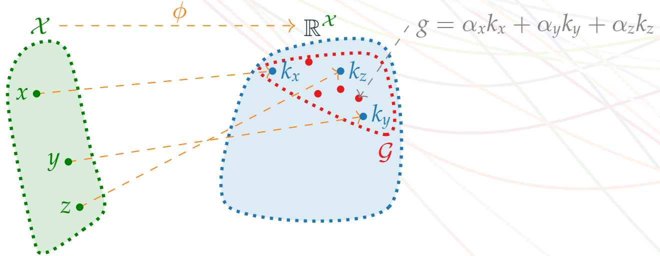
$$x \mapsto \phi(x) := k_x(\cdot) := k(x, \cdot),$$

i.e. $x \in \mathcal{X}$ is mapped to the function $k_x : \mathcal{X} \to \mathbb{R}$, $k_x(t) = k(x, t)$



- Now consider the images $\{k_x | x \in \mathcal{X}\}$ as a spanning set of a vector space, i.e. define $\mathcal{G}$ as the space containing all finite linear combinations of $k_{x_1}, \ldots, k_{x_r}$:

$$\mathcal{G} := \left\{ \sum_{i=1}^{r} \alpha_i k(x_i, \cdot) \middle| \alpha_i \in \mathbb{R}, r \in \mathbb{N}, x_i \in \mathcal{X} \right\}$$



Definition of the scalar product

- For the spanning functions we define:

$$\langle \cdot, \cdot \rangle = \langle k(x, \cdot), k(y, \cdot) \rangle := k(x, y)$$

- For general functions in $\mathcal{G}$ the scalar product is then given as follows: $g = \sum_i \alpha_i k(x_i, \cdot)$ and $f = \sum_j \beta_j k(y_j, \cdot)$ then

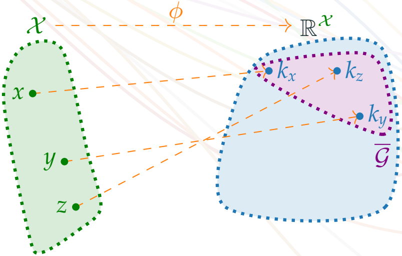
$$\begin{aligned} \langle f, g \rangle_{\mathcal{G}} &= \langle \sum_j \beta_j k(y_j, \cdot), \sum_i \alpha_i k(x_i, \cdot) \rangle_{\mathcal{G}} \\ &= \sum_{i,j} \alpha_i \beta_j \langle k(y_j, \cdot), k(x_i, \cdot) \rangle = \sum_{i,j} \alpha_i \beta_j k(y_j, x_i) \end{aligned}$$

· To check that this is really a scalar product, we need to prove (**exercise**):

- it is well-defined (not obvious because there might be several different linear combinations for the same function)
- it satisfies all properties of a scalar product (crucial ingredient is the fact that k is positive definite)

ONERA
THE FRENCH AEROSPACE LAB

· Finally, to make $\mathcal{G}$ a proper Hilbert space, we need to take its topological completion $\overline{\mathcal{G}}$ obtained by adding all limits of Cauchy sequences.



Summary.

1. We considered a mapping $\phi : \mathcal{X} \to \mathbb{R}^{\mathcal{X}}$ defined as $x \mapsto \phi(x) := k(x, \cdot)$

2. We defined $\mathcal{G}$ as the space containing all finite linear combinations of k:

$$\mathcal{G} := \left\{ \sum_{i=1}^{r} \alpha_i k(x_i, \cdot) \middle| \alpha_i \in \mathbb{R}, r \in \mathbb{N}, x_i \in \mathcal{X} \right\}$$

3. We defined a scalar product on $\mathcal{G}$: $\langle \cdot, \cdot \rangle_{\mathcal{G}} = \langle k(x, \cdot), k(y, \cdot) \rangle_{\mathcal{G}} := k(x, y)$

4. We took the completion $\overline{\mathcal{G}}$ obtained by adding all limits of Cauchy seqs

· The space $\mathcal{H} := \overline{\mathcal{G}}$ is called the reproducing kernel Hilbert space

- By construction, it has the property that $k(x, y) = \langle \phi(x), \phi(y) \rangle_{\mathcal{H}}$
- Then k is known as the *reproducing kernel*

Reproducing property

· Let $f = \sum_i \alpha_i k(x_i, \cdot)$. Then $\langle f, k(x, \cdot)\rangle_{\mathcal{H}} = f(x)$.

Proof.

$$\begin{aligned}
\langle f, k(x, \cdot)\rangle_{\mathcal{H}} &= \langle \sum_i \alpha_i k(x_i, \cdot), k(x, \cdot)\rangle_{\mathcal{H}} \\
&= \sum_i \alpha_i \langle k(x_i, \cdot), k(x, \cdot)\rangle \\
&= \sum_i \alpha_i k(x_i, x) \\
&= f(x)
\end{aligned}$$

□

Note. The word *reproducing* is used the sense that the function value is reproduced from a so-called *reproducing kernel* that does not depend on f.

Reproducing kernel Hilbert space (RKHS)

Definition (RKHS)

An RKHS $\mathcal{H}$ is a Hilbert space of real-valued functions,[a] defined on some set $\mathcal{X}$, for which all evaluation functionals

$$\begin{aligned} \delta_x : \mathcal{H} &\to \mathbb{R}, \\ f &\mapsto f(x), \end{aligned}$$

are continuous for any $x \in \mathcal{X}$.

[a] We focus here on real-valued functions, but the theory is similar for complex-valued ones

- With Riesz theorem, there exists $k_x \in \mathcal{H}$ such that for all $x \in \mathcal{X}$

$$\langle f, k(x, \cdot) \rangle_{\mathcal{H}} = f(x) \quad \text{(reproducing property)}$$

- Observe that $f(x)$ is obtained by computing the inner product between:
  - one part that is purely local, depends only on the input x
  - one part that is global and depends only on the function f

Examples

1. All Hilbert spaces are not RKHS

- It can be shown that $L^2(0,1)$, with its usual Hilbert structure $\langle f, g\rangle = \int_0^1 f(t)g(t)dt$, is not a RKHS (**exercise**)

2. Finite-dimensional spaces.

· Every finite dimensional (real) Hilbert space of functions is an RKHS

- The kernel is given by

$$k(x,y) = \sum_{i=1}^{n} e_i(x)e_i(y),$$

  where $e_1(\cdot), \ldots, e_n(\cdot)$ is an orthonormal basis
- This results from the basis expansion $f = \sum_{i=1}^{n} \langle f, e_i(\cdot)\rangle e_i$, evaluated at x:

$$f(x) = \sum_{i=1}^{n} \langle f, e_i(\cdot)\rangle e_i(x) = \langle f, \sum_{i=1}^{n} e_i(\cdot)e_i(x)\rangle$$

ONERA THE FRENCH AEROSPACE LAB

3. Sobolev space H_0^1 and Brownian motion.

· Denote $H_0^1 = \{h \in L^2(0,1), h(0) = 0, h' \in L^2(0,1)\}$, with scalar product

$$\langle h, g \rangle := \int_0^1 h'(u) g'(u) du, \qquad \text{for} \quad h, g \in H_0^1,$$

where derivatives are taken in the sense of distributions

· As an example of Sobolev space, it is well known that H_0^1 is a Hilbert space

· It is an RKHS, whose kernel can be obtained directly by definition

$$h(x) = \int_0^x h'(u) du = \int_0^1 h'(u) \mathbb{1}_{[0,x]}(u) du$$

· Then k_x must be equal to the primitive of $\mathbb{1}_{[0,x]}$ that vanishes at 0, i.e.

$$k(x, y) := k_x(y) = \min(x, y)$$

· Observe that k is the covariance function of the Brownian motion

ONERA
THE FRENCH AEROSPACE LAB

Application: dissociation between response and design

· In the context of design of experiments, we aim at choosing design points without knowing the response values at these points [Roustant, 2011]

· RKHS are also well suited to dissociate the design from the response

· The reproducing property itself:

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

shows that the value $f(x)$ is dissociated between one part depending only on the function f and another one only on its input x.

· A direct application of the Cauchy-Schwarz inequality implies that:

$$|f(x)| \leq \|f\|_{\mathcal{H}} \times \sqrt{k(x, x)},$$

which shows that the dissociation is also obtained for upper bounds.

ONERA
THE FRENCH AEROSPACE LAB

Theorem (Moore-Aronszajn theorem)

If k is a reproducing kernel, then it is symmetric and positive definite. Conversely, if k is a kernel, one can construct a unique RKHS $\mathcal{H}$ with k as a reproducing kernel.

· This means that there is an equivalence between RKHS, reproducing kernels and covariance functions.

Proof.

$\Rightarrow$ k is a kernel by definition:

- $\langle \cdot, \cdot \rangle$ is symmetric: $\langle k(x,\cdot), k(y,\cdot) \rangle = \langle k(y,\cdot), k(x,\cdot) \rangle$
- positive definiteness of scalar products:

$$\sum_{i,j=1}^{n} a_i a_j k(x_i, x_j) = \sum_{i,j=1}^{n} a_i a_j \langle k(x_i,\cdot), k(x_j,\cdot) \rangle = \langle \sum_{i=1}^{n} a_i k(x_i,\cdot), \sum_{j=1}^{n} a_j k(x_j,\cdot) \rangle \geq 0$$

$\Leftarrow$ See construction of the RKHS

ONERA THE FRENCH AEROSPACE LAB

Operations on RKHS/kernels and associated kernels/RKHS

· Remember that we can create new kernels by combining predefined ones:

$$\begin{array}{ll} \text{Sum of kernels:} & k(\boldsymbol{x},\boldsymbol{x}') = k_1(\boldsymbol{x},\boldsymbol{x}') + k_2(\boldsymbol{x},\boldsymbol{x}') \\ \text{Product of kernels:} & k(\boldsymbol{x},\boldsymbol{x}') = k_1(\boldsymbol{x},\boldsymbol{x}') \times k_2(\boldsymbol{x},\boldsymbol{x}') \end{array}$$

· One may ask what are the associated RKHS?

· Conversely, some operations on Hilbert spaces preserve the RKHS structure. What are the associated kernels?

RKHS associated to a sum of two kernels.

· The first observation is that the RKHS associated to a sum of kernels is not the usual algebraic sum of the associated RKHS

· With Moore-Aronszajn theorem, we see that if k_1 and k_2 are kernels, then:

$$\mathcal{H}_{k_1+k_2} = \overline{\mathsf{span}(k_1(x,\cdot) + k_2(x,\cdot), x \in \mathcal{X})}$$

· This is in general *strictly* included in the vector space $\mathcal{H}_{k_1} + \mathcal{H}_{k_2}$ since it does not contain all the $k_1(x,\cdot) + k_2(x,\cdot)$

ONERA
THE FRENCH AEROSPACE LAB

Operations on RKHS/kernels and associated kernels/RKHS

· The definition of the norm of $\mathcal{H}_{k_1+k_2}$ involves the couples $(h_1, h_2) \in \mathcal{H}_{k_1} \times \mathcal{H}_{k_2}$ such that $h_1 + h_2 = h$, which are not unique as soon as $\mathcal{H}_{k_1} \cap \mathcal{H}_{k_2} \neq 0$

· The norm can be obtained by solving the optimization problem:

$$\|h\|_{\mathcal{H}_{k_1+k_2}} = \min_{\substack{(h_1,h_2)\in\mathcal{H}_{k_1}\times\mathcal{H}_{k_2} \\ h_1+h_2=h}} \|(h_1, h_2)\|_{\mathcal{H}_{k_1}\times\mathcal{H}_{k_2}},$$

where $\|(h_1, h_2)\|_{\mathcal{H}_{k_1}\times\mathcal{H}_{k_2}} = \|h_1\|^2_{\mathcal{H}_{k_1}} + \|h_2\|^2_{\mathcal{H}_{k_2}}$ is the usual product norm

ONERA
THE FRENCH AEROSPACE LAB

Kernel associated to the orthogonal projection of a RKHS.

· Let $\mathcal{G} = \Pi(\mathcal{H})$ be the image of a RKHS $\mathcal{H}$ with reproducing kernel k by an orthogonal projection Π

· As a closed supspace, $\mathcal{G}$ is a Hilbert space

· Now, for all $g \in \mathcal{G}$ and $x \in \mathcal{X}$, we can use the reproducing property in $\mathcal{H}$:

$$g(x) = \langle g, k(x, \cdot) \rangle$$

· By definition of Π, $k(x, \cdot) - \Pi(k(x, \cdot))$ is orthogonal to g, and hence

$$g(x) = \langle g, \Pi(k(x, \cdot)) \rangle$$

· As $\Pi(k(x, \cdot)) \in \mathcal{G}$, this shows that G is an RKHS with reproducing kernel

$$(x, y) \to \Pi(k(x, \cdot))(y)$$

Regularity of functions in RKHS

· Real-valued functions over $\mathcal{X}$, in the RKHS $\mathcal{H}$ with reproducing kernel k, fulfil a Lipschitz-like condition, with Lipschitz constant given by $\|f\|_{\mathcal{H}}$

· By the Cauchy-Schwartz inequality, we get for all $x, y \in \mathcal{X}$

$$\begin{aligned} |f(x) - f(y)| &= |\langle f, k(x,\cdot)\rangle - \langle f, k(y,\cdot)\rangle| \\ &= |\langle f, k(x,\cdot) - k(y,\cdot)\rangle| \\ &\leq \|f\|_{\mathcal{H}} \|k(x,\cdot) - k(y,\cdot)\| \\ &\leq \|f\|_{\mathcal{H}} d(x,y), \end{aligned}$$

with the distance d over $\mathcal{X}$ defined by

$$\begin{aligned} d^2(x,y) &= \langle k(x,\cdot) - k(y,\cdot), k(x,\cdot) - k(y,\cdot)\rangle \\ &= \langle k(x,\cdot), k(x,\cdot)\rangle - 2\langle k(x,\cdot), k(y,\cdot)\rangle + \langle k(y,\cdot), k(y,\cdot)\rangle \\ &= k(x,x) - 2k(x,y) + k(y,y) \end{aligned}$$

ONERA
THE FRENCH AEROSPACE LAB

The representer theorem

The representer theorem

- In general, the RKHS is an infinite-dimensional vector space
  - a basis has to contain infinitely many vectors
- The **representer theorem** shows that in practice, we only have to deal with a finite-dimensional subspace.

- Assume we are given a kernel k. Denote the corresponding RKHS with $\mathcal{X}$, and the norm and scalar product in the space by $\|\cdot\|_{\mathcal{X}}$ and $\langle \cdot, \cdot \rangle_{\mathcal{H}}$.
- Assume that we want to learn a linear real-valued function $f : \mathcal{H} \to \mathbb{R}$ that acts on the RKHS $\mathcal{H}$ of a kernel k.
- All such functions have the form $f(x) = \langle w, x \rangle_{\mathcal{H}}$ for some $w \in \mathcal{H}$, that is we can identify the function f with the corresponding vector $w \in \mathcal{H}$.

The representer theorem

Theorem (Representer theorem)

Consider a regularised risk minimisation problem of the form

$$\min_{w \in \mathcal{H}} J_\lambda(w), \qquad \textit{with} \quad J_\lambda(w) = R_n(w) + \lambda\Omega(\|w\|_{\mathcal{H}}), \tag{1}$$

where $\mathcal{X}$ is an arbitrary input space, $\mathcal{Y}$ is the output space, $k : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ is a kernel with $\mathcal{H}$ the corresponding RKHS, and $\lambda \in \mathbb{R}^+$ a regularization parameter.

For a given training set of $(x_i, y_i)_{1 \le i \le n} \subset \mathcal{X} \times \mathcal{Y}$ and a classifier $f_w(u) = \langle w, u \rangle_{\mathcal{H}}$, let R_n be the empirical risk of the classifier w.r.t. a loss function ℓ, and $\Omega : [0, \infty) \to \mathbb{R}$ a strictly monotonically increasing function. Then, the problem in (1) *always has an optimal solution of the form*

$$w^* = \sum_{i=1}^{n} \alpha_i k(x_i, \cdot).$$

ONERA
THE FRENCH AEROSPACE LAB

Proof intuition.

· In general the problem is posed in a function space $\mathcal{H}$, which is very often *not* finite dimensional [Wahba, 1990]

· Split the space $\mathcal{H}$ into two subspaces:

- $\mathcal{G} := \mathsf{span}(k_{x_1}, \ldots, k_{x_n})$ (induced by the data)
- $\mathcal{G}^{\perp}$ the orthogonal complement

· Using $\mathcal{H} = \mathcal{G} + \mathcal{G}^{\perp}$, we can write $h = f + g$ with $f \in \mathcal{G}$ and $g \in \mathcal{G}^{\perp}$

· Applying the reproducing property leads to $g(x_i) = \langle g, k(x_i, \cdot)\rangle = 0$

· We obtain that $J_\lambda(h) = J_\lambda(f) + \lambda\Omega(\|g\|_{\mathcal{H}})$

- The optimisation can be done independently along f and g, then $g = 0$
- The optimum is obtained for $h = f \in \mathcal{G}$

$\square$

ONERA
THE FRENCH AEROSPACE LAB

- In many practical situations, we have

$$\min_{h \in \mathcal{H}} J_\lambda(h), \qquad \text{with} \quad J_\lambda(h) = \sum_{i=1}^{n} (y_i - h(x_i))^2 + \lambda \|h\|_{\mathcal{H}}^2.$$

- We know that h lies in the finite-dimensional space spanned by $k_{x_1}, \dots, k_{x_n}$:

$$h(x) = \sum_{i=1}^{n} \alpha_i k(x_i, x) = \boldsymbol{k}^\top(x)\boldsymbol{\alpha},$$

where $\boldsymbol{k}(x) = [k(x_i, x), \dots, k(x_n, x)]^\top$ and $\boldsymbol{\alpha} = [\alpha_1, \dots, \alpha_n]^\top$.

ONERA
THE FRENCH AEROSPACE LAB

Link with splines and Kriging

- Denoting $\mathbf{K} = (k(x_i, x_j))_{1 \le i,j \le n}$, then $\|h\|^2_{\mathcal{H}} = \boldsymbol{\alpha}^\top \mathbf{K}\boldsymbol{\alpha}$ and $h(x_i) = (\mathbf{K}\boldsymbol{\alpha})_i$
- Thus, the criterion $J_\lambda(h)$ can be written as second order polynomial in $\boldsymbol{\alpha}$:

$$J_\lambda = [\boldsymbol{y} - \mathbf{K}\boldsymbol{\alpha}]^\top [\boldsymbol{y} - \mathbf{K}\boldsymbol{\alpha}] + \lambda \boldsymbol{\alpha}^\top \mathbf{K}\boldsymbol{\alpha},$$

with $\boldsymbol{y} = [y_1, \dots, y_n]^\top$, and minimising w.r.t. $\boldsymbol{\alpha}$ leads to (**exercise**):

$$h(x) = \boldsymbol{k}^\top(x)[\mathbf{K} + \lambda \mathbf{I}_n]^{-1}\boldsymbol{y}.$$

Link with splines and Kriging

· The expression $h(x) = \boldsymbol{k}^\top(x)[\mathbf{K} + \lambda \mathbf{I}_n]^{-1}\boldsymbol{y}$ corresponds to the formula for smoothing splines and to the Gaussian process prediction (with noisy observations) [see, e.g., Rasmussen and Williams, 2005].

· When λ tends to zero, we get the formula for interpolation splines and for Kriging prediction:

$$h(x) = \boldsymbol{k}^\top(x)\mathbf{K}^{-1}\boldsymbol{y}$$

ONERA
THE FRENCH AEROSPACE LAB

Equivalence between RKHS and random processes

· RKHS and stochastic processes are strongly connected by the so-called *Loève representation theorem*.

· As $\mathcal{H}$ is spanned by $k(x,\cdot)$, the idea is to consider $\overline{\mathcal{L}}(Z) = \overline{\mathsf{span}(Z_x, x \in \mathcal{X})}$, for a centred second order random process $Z = (Z_x)_{x\in\mathcal{X}}$ with kernel k

· $\overline{\mathcal{L}}(Z)$ is a Hilbert space with scalar product induced by $\langle V, W\rangle = \mathbb{E}\left\{VW\right\}$.

· Furthermore, $\langle k(x,\cdot), k(y,\cdot)\rangle := k(x,y) = \langle Z_x, Z_y\rangle$, and it results that $\overline{\mathcal{L}}(Z)$ is isometric to the RKHS $\mathcal{H}$ through the map defined on the $k(x,\cdot)$ by:

$$\begin{aligned} \phi : \mathcal{H} &\to \overline{\mathcal{L}}(Z), \\ k(x,\cdot) &\mapsto Z_x, \end{aligned}$$

and extended by linearity and continuity.

· This important result serves as a dictionary to translate a functional problem into a probabilistic one, and vice-versa.

Conclusions

Conclusions

· Here, we studied the link between covariance functions (kernels) and RKHS

· An RKHS is a Hilbert space of real-valued functions for which all evaluation functionals are continuous

· If k is a kernel, there exists a unique RKHS with k as its reproducing kernel

· If k is a reproducing kernel, then it is a covariance function

· There exists an equivalence between RKHS and stochastic processes (*Loève representation theorem*)

References

Alain Berlinet and Christine Thomas-Agnan. *Reproducing kernel Hilbert spaces in probability and statistics*. Springer Science & Business Media, 2011.

Carl E. Rasmussen and Christopher K. I. Williams. *Gaussian Processes for Machine Learning (Adaptive Computation and Machine Learning)*. MIT Press, 2005.

Olivier Roustant. Statistical models and methods for computer experiments. Technical report, 2011. HDR – Université Jean Monnet, Saint-Étienne, France.

Michael L. Stein. *Interpolation of Spatial Data: Some Theory for Kriging*. Springer, 1999.

Ulrike von Luxburg. Statistical Machine Learning Part 19 - The reproducing kernel Hilbert space, 2020. University of Tübingen.

Grace Wahba. *Spline models for observational data*. SIAM, 1990.