What is hidden behind the Sobolev kernels involved in the HSIC-ANOVA decomposition?

Gabriel Sarazin, Amandine Marrel, Sébastien da Veiga, Vincent Chabridon

Speaker: Gabriel Sarazin

What is hidden behind the Sobolev kernels involved in the HSIC-ANOVA decomposition?

Gabriel Sarazin
CEA, DES, IRESNE, DER, Cadarache F-13108 Saint-Paul-Lez-Durance,
gabriel.sarazin@cea.fr
Amandine Marrel
CEA, DES, IRESNE, DER, Cadarache F-13108 Saint-Paul-Lez-Durance,
Institut de Mathématiques de Toulouse, F-31062, Toulouse, amandine.marrel@cea.fr
Sébastien Da Veiga
Safran Tech, Modeling & Simulation Rue des Jeunes Bois, Châteaufort,
F-78114 Magny-Les-Hameaux, sebastien.daveiga@gmail.com
Vincent Chabridon
EDF R&D
6 Quai Watier, F-78401, Chatou, vincent.chabridon@edf.fr

In many industrial fields, physical phenomena are commonly modeled by numerical simulators. These simulation codes can take a high number (dozens, if not hundreds) of uncertain parameters as input variables. Since computer-based simulation experiments are time-consuming, the permissible number of runs is most often very limited. In this context, the benefits brought by a preliminary global sensitivity analysis are twofold. On the one hand, the screening step allows to select a smaller amount of influential inputs in order to reduce dimensionality or to support the construction of a surrogate model. On the other hand, the influence of each input variable is quantified and the associated rankings provide a relevant piece of information that can be used in the sequential building process of a metamodel as in [4]. In view of these two purposes, total-order Sobol indices are often regarded as the most adapted indicators. Unfortunately, unlike their first-order counterparts, their estimation requires a too large number of simulations.

To remedy this difficulty, a new class of sensitivity measures built upon the theory of reproducing kernel Hilbert spaces (RKHS) has emerged over the last decade [1]. Each input variable is assigned a continuous positive definite kernel and is equipped with the related RKHS. The same is done for the output variable. The distance between the joint input-output bivariate distribution and the bivariate distribution obtained under independence is measured through the distance of their respective embeddings in the tensorized RKHS, which leads to the Hilbert-Schmidt Independence Criterion (HSIC) [3]. Since HSIC indices can be expressed by means of kernel-based statistical moments, their estimation can be achieved at minimal computational cost and without estimating the joint input-output distribution. However, HSIC indices are often blamed for their lack of interpretability because their sum is not equal to 1. As a consequence, they do not fit into the advantageous framework that goes with analysis of variance (ANOVA). To bridge this gap, it has been recently proved that the use of orthogonal kernels enables an ANOVA-like decomposition for HSIC indices [2]. When the input variables are uniformly distributed, Sobolev kernels (parametrized by a smoothness parameter r) match all the required conditions. They allow to define higher order HSIC indices (especially total-order HSIC indices) and therefore to rigorously separate main effects and interactions.

The main objective of this work is to provide new insights into Sobolev kernels which have indeed been very few studied so far. First, we demonstrate that Sobolev kernels are characteristic. This ensures that the nullity of Sobolev-based HSIC indices is equivalent to independence. Then, another notable contribution of this work is the identification of one explicit feature map for Sobolev kernels, which may help the user understand the spectrum of dependency patterns that can be captured by this novel HSIC index. As with most kernels, the canonical feature map (resulting from Aronszajn's theorem) is uninformative. An additional pitfall comes from the fact that Sobolev kernels are not shift-invariant, which prevents the use of Bochner's theorem and the following characterization in terms of Fourier transform. Instead, another kind of feature map stems from Mercer's theorem which asserts that any continuous kernel on a compact set may be rewritten as an infinite sum that only implies the eigenvalues and eigenfunctions of the corresponding kernel integral operator. This key result discloses a feature map that sends the unit interval into the Hilbert space of real-valued square-summable sequences. The eigenvalue problem arising from this new feature map is tackled in two different ways. Firstly, we investigate the benefits of kernel feature analysis where eigenvalues and eigenfunctions are estimated from Gram matrix simulation. It allows to visualize eigenfunctions and sometimes to come up with closed-form candidate functions. Secondly, we demonstrate that the eigenvalue problem is equivalent to a Cauchy problem consisting of a linear homogeneous ordinary differential equation with constant coefficients and a sufficient number of boundary conditions. Then, two situations deserve a specific study. When $r = 1$, an exact analytical solution is available. On the contrary, when $r = 2$, the Cauchy problem cannot be completely solved. A solution in the form of a linear combination of analytical functions is obtained but the coefficients cannot be retrieved. An asymptotic approximation is then derived to accurately estimate small eigenvalues while robust numerical resolution is achieved by use of a semi-analytical adhoc calibration algorithm.

[1] Sébastien Da Veiga. Global sensitivity analysis with dependence measures. *Journal of Statistical Computation and Simulation*, 85(7):1283–1305, 2015.

[2] Sébastien Da Veiga. Kernel-based ANOVA decomposition and Shapley effects: application to global sensitivity analysis. *arXiv preprint arXiv:2101.05487*, 2021.

[3] Arthur Gretton, Olivier Bousquet, Alex Smola, and Bernhard Schölkopf. Measuring statistical dependence with Hilbert-Schmidt norms. In *International conference on algorithmic learning theory*, pages 63–77. Springer, 2005.

[4] Amandine Marrel, Bertrand Iooss, and Vincent Chabridon. Statistical identification of penalizing configurations in high-dimensional thermalhydraulic numerical experiments: the ICSCREAM methodology. *arXiv preprint arXiv:2004.04663*, 2020.

More about HSIC-ANOVA indices…

What is hidden behind Sobolev kernels?

DE LA RECHERCHE À L'INDUSTRIE

10th SAMO conference, Tallahassee, Florida, USA

G. Sarazin, A. Marrel, S. Da Veiga (Safran) & V. Chabridon (EDF R&D)

Wednesday, March 16th, 2022

IRESNE I DER I SESI I LEMS

Institut de recherche sur les systèmes nucléaires pour la production d'énergie bas carbone

Summary

Introduction

1) From HSIC indices to HSIC-ANOVA indices
- Reminders on HSIC indices
- Towards an ANOVA decomposition in the HSIC paradigm
- Focus on Sobolev kernels

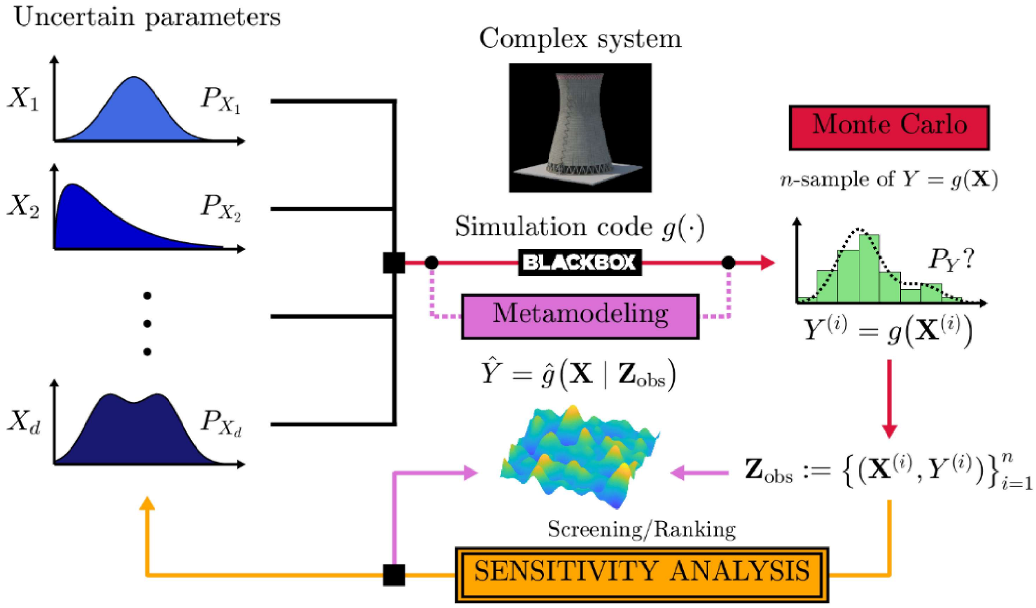
2) How to find a feature map for Sobolev kernels?
- Approach 1: kernel feature analysis
- Approach 2: transformation into a Cauchy problem
- Approach 3: series expansions of Bernoulli polynomials

3) Are Sobolev kernels characteristic?

Conclusion

Introduction

- → Let $\boldsymbol{X} := [X_1, \ldots, X_d]$ be a random vector with **independent** components ($d \approx 100$).
- → Let $Y := g(\boldsymbol{X})$ where $g: \mathbb{R}^d \to \mathbb{R}$ is a **computationally-expensive** simulation code.



- → Total-order Sobol' indices cannot be estimated from a small amount of input-output samples.
- → The HSIC indices of Gretton *et al.* (2005) have been inscreasingly used since Da Veiga (2014).

From HSIC indices to HSIC-ANOVA indices

Reminders on HSIC indices

- → Let $\mathbb{P}_{\boldsymbol{X}}$ and $\mathbb{P}_Y$ be the probability measures associated to $\boldsymbol{X}$ to Y.
- → A kernel $K_i : \boldsymbol{\mathcal{X}}_i \times \boldsymbol{\mathcal{X}}_i \to \mathbb{R}$ (with feature space $\boldsymbol{\mathcal{F}}_i$ and feature map φ_i) is assigned to X_i.
- → A kernel $K_Y : \boldsymbol{\mathcal{Y}} \times \boldsymbol{\mathcal{Y}} \to \mathbb{R}$ (with feature space $\boldsymbol{\mathcal{G}}$ and feature map φ_Y) is assigned to Y.

$$\forall 1 \leq i \leq d, \quad \boxed{K_i(x_i, x_i') = \langle \varphi_i(x_i'), \varphi_i(x_i) \rangle_{\mathcal{F}_i}} \quad \text{and} \quad \boxed{K_Y(y, y') = \langle \varphi_Y(y'), \varphi_Y(y) \rangle_{\mathcal{G}}}$$

- ➢ This allows to define the **Hilbert-Schmidt Independence Criterion (HSIC)**.
  - ✓ **Kernel-based dissimilarity measure** between $\mathbb{P}_{X_i Y}$ and $\mathbb{P}_{X_i} \otimes \mathbb{P}_Y$.

From HSIC indices to HSIC-ANOVA indices

Reminders on HSIC indices

- Let $\mathbb{P}_{\boldsymbol{X}}$ and $\mathbb{P}_Y$ be the probability measures associated to $\boldsymbol{X}$ to Y.
- A kernel $K_i : \mathcal{X}_i \times \mathcal{X}_i \to \mathbb{R}$ (with feature space $\mathcal{F}_i$ and feature map φ_i) is assigned to X_i.
- A kernel $K_Y : \mathcal{Y} \times \mathcal{Y} \to \mathbb{R}$ (with feature space $\mathcal{G}$ and feature map φ_Y) is assigned to Y.

$$\forall 1 \leq i \leq d, \quad \boxed{K_i(x_i, x_i') = \langle \varphi_i(x_i'), \varphi_i(x_i) \rangle_{\mathcal{F}_i}} \quad \text{and} \quad \boxed{K_Y(y, y') = \langle \varphi_Y(y'), \varphi_Y(y) \rangle_{\mathcal{G}}}$$

- This allows to define the **Hilbert-Schmidt Independence Criterion (HSIC)**.
  - ✓ **Kernel-based dissimilarity measure** between $\mathbb{P}_{X_i Y}$ and $\mathbb{P}_{X_i} \otimes \mathbb{P}_Y$.
  - ✓ Norm of a **covariance-like operator** between the feature maps φ_i and φ_Y.

$$\mathrm{HSIC}(X_i, Y) = \|C_{X_i Y}\|^2_{\mathrm{HS}} \quad \text{with} \quad \begin{array}{l} C_{X_i Y} : \mathcal{G} \longrightarrow \mathcal{F}_i \\ C_{X_i Y} = \mathbb{E}[\varphi_i(X_i) \otimes \varphi_Y(Y)] - \mathbb{E}[\varphi_i(X_i)] \otimes \mathbb{E}[\varphi_Y(Y)] \end{array}$$

- Explicit knowledge about the feature maps φ_i and φ_Y may help understand HSIC indices.

From HSIC indices to HSIC-ANOVA indices

Reminders on HSIC indices

- Let $\mathbb{P}_{\boldsymbol{X}}$ and $\mathbb{P}_Y$ be the probability measures associated to $\boldsymbol{X}$ to Y.
- A kernel $K_i : \mathcal{X}_i \times \mathcal{X}_i \to \mathbb{R}$ (with feature space $\mathcal{F}_i$ and feature map φ_i) is assigned to X_i.
- A kernel $K_Y : \mathcal{Y} \times \mathcal{Y} \to \mathbb{R}$ (with feature space $\mathcal{G}$ and feature map φ_Y) is assigned to Y.

$$\forall 1 \leq i \leq d, \quad \boxed{K_i(x_i, x_i') = \langle \varphi_i(x_i'), \varphi_i(x_i) \rangle_{\mathcal{F}_i}} \quad \text{and} \quad \boxed{K_Y(y, y') = \langle \varphi_Y(y'), \varphi_Y(y) \rangle_{\mathcal{G}}}$$

- This allows to define the **Hilbert-Schmidt Independence Criterion (HSIC)**.
  - ✓ **Kernel-based dissimilarity measure** between $\mathbb{P}_{X_i Y}$ and $\mathbb{P}_{X_i} \otimes \mathbb{P}_Y$.
  - ✓ Norm of a **covariance-like operator** between the feature maps φ_i and φ_Y.

- Let (X_i, Y) and (X_i', Y') be two independent copies of $\mathbb{P}_{X_i Y}$.

$$\begin{aligned} \mathrm{HSIC}(X_i, Y) = \; & \mathbb{E}_{\substack{X_i Y \\ X_i' Y'}}\left[K_i(X_i, X_i')\, K_Y(Y, Y')\right] + \mathbb{E}_{X_i X_i'}\left[K_i(X_i, X_i')\right] \mathbb{E}_{Y Y'}\left[K_Y(Y, Y')\right] \\ & -2\mathbb{E}_{X_i Y}\left[\mathbb{E}_{X_i'}\left[K_i(X_i, X_i')\right] \mathbb{E}_{Y'}\left[K_Y(Y, Y')\right]\right] \end{aligned}$$

- The estimation of $\mathrm{HSIC}(X_i, Y)$ can be considered with either a **U-statistic** or a **V-statistic**.

Towards an ANOVA decomposition in the HSIC paradigm

Keynote 4 - Sébastien DA VEIGA
A kernel-based ANOVA decompositon: extending sensitivity analysis and Shapley effects with kernels

- Da Veiga (2021) → $\mathrm{HSIC}(\boldsymbol{X}, Y)$ may be apportioned between **independent** inputs.

$$\mathrm{HSIC}(\mathbf{X}, Y) = \sum_{\mathbf{u}\subseteq\{1,\dots,d\}} \mathrm{HSIC}_{\mathbf{u}} = \sum_{\mathbf{u}\subseteq\{1,\dots,d\}} \sum_{\mathbf{v}\subseteq\mathbf{u}} (-1)^{|\mathbf{u}|-|\mathbf{v}|}\mathrm{HSIC}(\mathbf{X}_{\mathbf{v}}, Y)$$

- The **first-order and total-order HSIC-ANOVA indices** are then defined by:

$$\forall 1 \le i \le d, \quad S_i^{\mathrm{HSIC}} := \frac{\mathrm{HSIC}(X_i, Y)}{\mathrm{HSIC}(\mathbf{X}, Y)} \quad \text{and} \quad T_i^{\mathrm{HSIC}} := 1 - \frac{\mathrm{HSIC}(\mathbf{X}_{-i}, Y)}{\mathrm{HSIC}(\mathbf{X}, Y)}$$

- Each input kernel K_i must be an **ANOVA kernel**.
  - ✓ $K_i = 1 + k_i$ with k_i an **orthogonal kernel**.

$$\forall x_i \in \mathcal{X}_i, \quad \int_{\mathcal{X}_i} k_i(x_i, z)\, \mathrm{d}\mathbb{P}_{X_i}(z) = 0$$

Focus on Sobolev kernels

→ The input variables $X_1, \dots, X_d$ need to be **independent** and **uniformly distributed** on $[0,1]$.

- Sobolev kernels $K_{Sob}^r = 1 + k_{Sob}^r$ (with r a smoothing parameter) are then **orthogonal**.

$$\begin{aligned} K_{\mathrm{Sob}}^r(x, x') &= 1 + \sum_{i=1}^{r} \frac{B_j(x)B_j(x')}{(j!)^2} + \frac{(-1)^{r+1}}{(2r)!} B_{2r}(|x - x'|) \\ &= 1 + k_A^r(x, x') + k_B^r(x, x') \\ \langle \theta_{\mathrm{Sob}}^r(x'), \theta_{\mathrm{Sob}}^r(x) \rangle_{\mathcal{H}_{\mathrm{Sob}}^r} &= 1 + \langle \varphi_A^r(x'), \varphi_A^r(x) \rangle_{\mathbb{R}^r} + \langle \theta_B^r(x'), \theta_B^r(x) \rangle_{\mathcal{H}_B^r} \end{aligned}$$

trivial feature map in $\mathbb{R}^r$ — **canonical feature map**

1. The definition is based on **Bernoulli polynomials** → $\int_0^1 B_k(t)dt = 0$ for all $k \ge 1$.
2. It is easy to see that $k_A^r(x, x') = \langle \varphi_A^r(x'), \varphi_A^r(x) \rangle_{\mathbb{R}^r}$ with $\varphi_A^r(x) = \left[\frac{B_j(x)}{j!}\right]_{1 \le j \le r}$.
3. Bochner's theorem cannot be applied because $k_B^r(x, x') = t_B^r(|x - x'|)$ with $t_B^r \notin L^2(\mathbb{R})$.

❑ **How to find a unified feature map φ_{Sob}^r?**

How to find a feature map for Sobolev kernels?

Approach 1: kernel feature analysis

- Let $\mathbb{P}_X$ be a probability measure on a **compact** domain $\mathcal{X}$.
- The **kernel integral operator** is defined by:

$$\begin{aligned} T_K : L^2(\mathcal{X},\mathbb{P}_X) &\longrightarrow L^2(\mathcal{X},\mathbb{P}_X) \\ f &\longmapsto T_K f \quad \text{with } [T_K f](x) := \int_{\mathcal{X}} K(x,z) f(z)\, d\mathbb{P}_X(z) \end{aligned}$$

Mercer's theorem

There exists an **orthonormal basis** $\{\phi_i\}_{i \in I}$ of $L^2(\mathcal{X},\mathbb{P}_X)$ consisting of **eigenfunctions** of T_K.
The associated eigenvalues $\{\lambda_i\}_{i \in I}$ are **non-negative**.
K admits the following representation:

$$K(x,x') = \sum_{i \in I} \lambda_i\, \phi_i(x)\phi_i(x') = \langle \varphi(x'), \varphi(x) \rangle_{\ell^2(I,\mathbb{R})}$$

and convergence is **absolute** and **uniform**. In addition, $\varphi(x) = \left[\sqrt{\lambda_i}\phi_i(x)\right]_{i \in I}$.

- An **infinite-dimensional eigenvalue problem** has to be solved → $T_K\phi = \lambda\phi$ with $\lambda > 0$

How to find a feature map for Sobolev kernels?

Approach 1: kernel feature analysis

- Let $\mathbb{P}_X$ be a probability measure on a **compact** domain $\mathcal{X}$.
- The **kernel integral operator** is defined by:

$$\begin{aligned} T_K : L^2(\mathcal{X},\mathbb{P}_X) &\longrightarrow L^2(\mathcal{X},\mathbb{P}_X) \\ f &\longmapsto T_K f \quad \text{with } [T_K f](x) := \int_{\mathcal{X}} K(x,z) f(z)\, d\mathbb{P}_X(z) \end{aligned}$$

Kernel Feature Analysis (KFA) / Kernel PCA

- Let $\{x_j\}_{1 \le j \le n}$ be a n-sample of a random variable $X \sim \mathbb{P}_X$.
- Discretization leads to a **finite-dimensional eigenvalue problem**:

$$\mathbf{L}\mathbf{v} = (n\lambda)\mathbf{v} \quad \text{with} \quad \mathbf{L} = \left[K(x_i,x_j)\right]_{1 \le i,j \le n} \quad \text{and} \quad \mathbf{v} = \left[\phi(x_i)\right]_{1 \le i \le n}$$

- Computation of the **eigenvalues** and **eigenvectors** of the Gram matrix $\boldsymbol{L}$.
  - ✓ Estimation of the n largest **eigenvalues** and **eigenfunctions** of T_K.

❑ **What about Sobolev kernels k^r_{Sob}?**

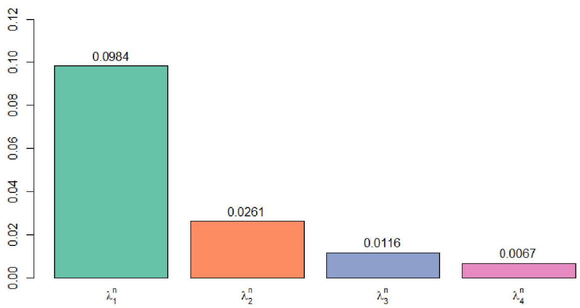
How to find a feature map for Sobolev kernels?

1 Sobolev kernel with $r = 1$

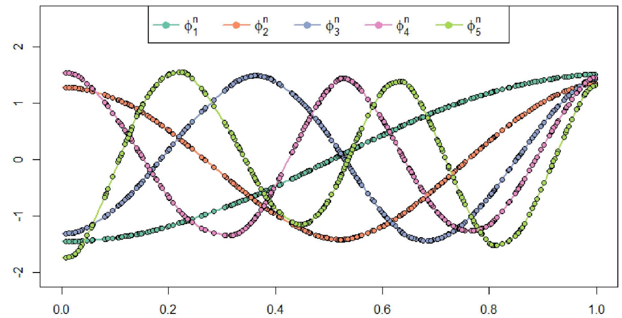
→ $\mathbb{P}_X = \mathcal{U}([0,1])$

→ $k_{Sob}^1(x, x') = B_1(x)B_1(x') + \frac{1}{2}B_2(|x - x'|)$

❑ **What do the eigenfunctions look like?**



Eigenvalues



Eigenfunctions

➢ <u>Conjecture</u>: $\forall k \geq 1, \phi_k(t) = \sqrt{2}\cos(k\pi t)$

➢ <u>Calculation by hand</u>:

$$[T_{k_{\text{Sob}}^1}\phi_k](x_0) = \int_0^1 k_{\text{Sob}}^1(x_0, z)\phi_k(z)\,\mathrm{d}z = \ldots = \overbrace{\frac{1}{(k\pi)^2}}^{\lambda_k}\phi_k(x_0)$$

➢ <u>Conclusion:</u> $K_{\text{Sob}}^1(x, x') = 1 + 2\sum_{k=1}^{\infty}\frac{1}{(k\pi)^2}\cos(k\pi x)\cos(k\pi x')$

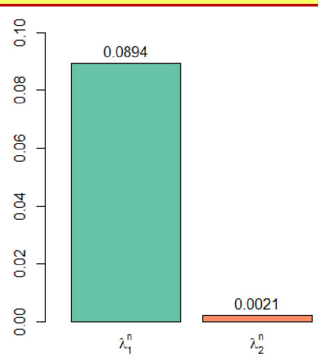
How to find a feature map for Sobolev kernels?

2 Sobolev kernel with $r = 2$

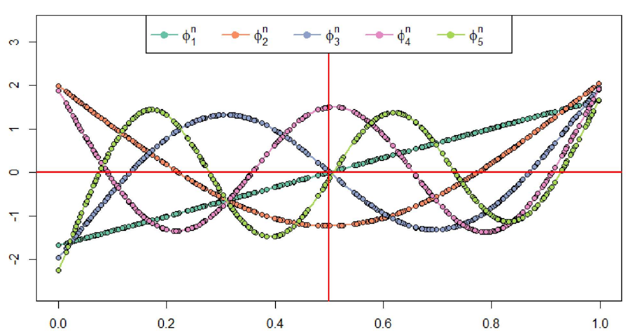
→ $\mathbb{P}_X = \mathcal{U}([0,1])$

→ $k_{Sob}^2(x, x') = B_1(x)B_1(x') + \frac{1}{4}B_2(x)B_2(x') - \frac{1}{24}B_4(|x - x'|)$

❑ **What do the eigenfunctions look like?**



Eigenvalues



Eigenfunctions

➢ <u>Conjecture</u>: $\forall k \geq 1, \phi_k(t) = \sqrt{2k+1}\, L_k(2t - 1)$ with L_k the k-th Legendre polynomial

➢ <u>Calculation by hand</u>: this is a numerical illusion!

✓ $\lambda_1 \gg \lambda_2$ and ϕ_1 pseudo-linear → $K_{Sob}^2(x, x') \approx 1 + 12\,\lambda_1\left(x - \frac{1}{2}\right)\left(x' - \frac{1}{2}\right)$

➢ <u>Conclusion:</u> K_{Sob}^2 is very close to the **dot-product kernel**.

How to find a feature map for Sobolev kernels?

Approach 2: transformation into a Cauchy problem

- Any eigenfunction ϕ of $T_{k^r_{Sob}}$ is **infinitely differentiable** on $[0,1]$. We proved that:

$$T_{k^r_{\mathrm{Sob}}}\phi = \lambda\phi$$
$$\Updownarrow$$
$$\lambda\phi^{[2r]} + (-1)^{r+1}\phi = 0 \quad \text{with} \quad \begin{cases} \phi^{[r]}(0) = \phi^{[r]}(1) = 0 \\ \forall 0 \le p \le r-2, (-1)^{r+p}\left(\phi^{[p]}(1) - \phi^{[p]}(0)\right) = \phi^{[2r-p-1]}(0) \\ \forall 0 \le p \le r-2, \phi^{[2r-p-1]}(0) = \phi^{[2r-p-1]}(1) \end{cases}$$

- For k^r_{Sob}, the eigenvalue problem is equivalent to a Cauchy problem $(\mathcal{C}_r)$ comprised of:
  - ✓ An homogeneous linear ODE of order $2r$ with constant coefficients.
  - ✓ $2r$ boundary conditions on ϕ and its following derivatives (up to order $2r-1$).

1 Sobolev kernel with $r = 1$

$$(\mathcal{C}_1):\ \lambda\phi''(t) + \phi(t) = 0 \ \text{ with } \ \phi'(0) = 0 \ \text{ and } \ \phi'(1) = 0$$

- There are solutions to $(\mathcal{C}_1)$ if and only if $\lambda \in \left(\frac{1}{(k\pi)^2}\right)_{k\ge 1}$.
- Analytical resolution yields $\phi_k(t) = \sqrt{2}\cos(k\pi t)$ → consistent with KFA!

How to find a feature map for Sobolev kernels?

Approach 2: transformation into a Cauchy problem

- Any eigenfunction ϕ of $T_{k^r_{Sob}}$ is **infinitely differentiable** on $[0,1]$. We proved that:

$$T_{k^r_{\mathrm{Sob}}}\phi = \lambda\phi$$
$$\Updownarrow$$
$$\lambda\phi^{[2r]} + (-1)^{r+1}\phi = 0 \quad \text{with} \quad \begin{cases} \phi^{[r]}(0) = \phi^{[r]}(1) = 0 \\ \forall 0 \le p \le r-2, (-1)^{r+p}\left(\phi^{[p]}(1) - \phi^{[p]}(0)\right) = \phi^{[2r-p-1]}(0) \\ \forall 0 \le p \le r-2, \phi^{[2r-p-1]}(0) = \phi^{[2r-p-1]}(1) \end{cases}$$

- For k^r_{Sob}, the eigenvalue problem is equivalent to a Cauchy problem $(\mathcal{C}_r)$ comprised of:
  - ✓ An homogeneous linear ODE of order $2r$ with constant coefficients.
  - ✓ $2r$ boundary conditions on ϕ and its following derivatives (up to order $2r-1$).

2 Sobolev kernel with $r = 2$

$$(\mathcal{C}_2):\ \lambda\phi^{[4]}(t) - \phi(t) = 0 \ \text{ with } \ \begin{cases} \phi''(0) = \phi''(1) = 0 \\ \phi(1) - \phi(0) = \phi^{[3]}(0) = \phi^{[3]}(1) \end{cases}$$

- The eigenfunctions are actually **not polynomials**.
- We cannot go further! There is **no closed-form expression** for these eigenvalues.

How to find a feature map for Sobolev kernels?

Approach 3: series expansions of Bernoulli polynomials

$$K^r_{\text{Sob}}(x,x') = 1 + k^r_A(x,x') + k^r_B(x,x') \quad \text{with} \quad k^r_B(x,x') = \frac{(-1)^{r+1}}{(2r)!} B_{2r}(|x-x'|)$$

$$\forall n \geq 2, \quad \forall 0 \leq x \leq 1, \quad B_n(x) = (-2) \times n! \sum_{k=1}^{\infty} \frac{\cos(2k\pi x - \frac{n\pi}{2})}{(2k\pi)^n}$$

➢ With this in mind, it is straighforward to see that:

$$\begin{aligned} k^r_B(x,x') &= 2\sum_{k=1}^{\infty} \frac{1}{(2k\pi)^{2r}} \left[\cos(2k\pi x)\cos(2k\pi x') + \sin(2k\pi x)\sin(2k\pi x')\right] \\ &= \sum_{k=1}^{\infty} \lambda_k c_k(x) c_k(x') + \sum_{k=1}^{\infty} \lambda_k s_k(x) s_k(x') \ \text{ with } \begin{cases} \lambda_k & := \frac{1}{(2k\pi)^{2r}} \\ c_k(t) & := \sqrt{2}\cos(2k\pi t) \\ s_k(t) & := \sqrt{2}\sin(2k\pi t) \end{cases} \end{aligned}$$

1. Explicit feature map φ^r_B of k^r_B in $\ell^2(\mathbb{N},\mathbb{R})$. It can be combined with φ^r_A which arrives in $\mathbb{R}^r$.
2. The functions $(c_k)_{k\geq 1}$ and $(s_l)_{l\geq 1}$ are orthonormal. This is the Mercer decomposition of k^r_B.
3. For $j \geq 1$, (c_j, s_j) is one orthonormal basis for the 2-dimensional eigenspace of λ_j.

How to find a feature map for Sobolev kernels?

Approach 3: series expansions of Bernoulli polynomials

$$K^r_{\text{Sob}}(x,x') = 1 + k^r_A(x,x') + k^r_B(x,x') \quad \text{with} \quad k^r_B(x,x') = \frac{(-1)^{r+1}}{(2r)!} B_{2r}(|x-x'|)$$

$$\forall n \geq 2, \quad \forall 0 \leq x \leq 1, \quad B_n(x) = (-2) \times n! \sum_{k=1}^{\infty} \frac{\cos(2k\pi x - \frac{n\pi}{2})}{(2k\pi)^n}$$

➢ **Global feature map** from $[0,1]$ in $\ell^2(\mathbb{N},\mathbb{R})$ that explicits feature functions:

$$K^r_{\text{Sob}}(x,x') = 1 + \sum_{k=1}^{r} \frac{B_k(x)B_k(x')}{(k!)^2} + 2\sum_{k=1}^{\infty} \frac{1}{(2k\pi)^{2r}} \left[\cos(2k\pi x)\cos(2k\pi x') + \sin(2k\pi x)\sin(2k\pi x')\right]$$

$$K^r_{\text{Sob}}(x,x') = \left\langle \varphi^r_{\text{Sob}}(x'), \varphi^r_{\text{Sob}}(x) \right\rangle_{\ell^2(\mathbb{N},\mathbb{R})} \quad \text{with:}$$

$$\varphi^r_{\text{Sob}}(x) = \left[1, \left(\frac{B_k(x)}{k!}\right)_{1\leq k\leq r}, \left(\frac{\sqrt{2}\cos(2k\pi x)}{(2k\pi)^r}\right)_{k\geq 1}, \left(\frac{\sqrt{2}\sin(2k\pi x)}{(2k\pi)^r}\right)_{k\geq 1}\right]$$

Are Sobolev kernels characteristic?

	constant features		polynomial features		sinusoidal features
$K^r_{\text{Sob}} =$	1	$+$	k^r_A	$+$	k^r_B
$\downarrow$	$\downarrow$		$\downarrow$		$\downarrow$
$\mathcal{H}^r_{\text{Sob}} =$	$\text{Span}(\{1\})$	$\oplus$	$\mathcal{H}^r_A$	$\oplus$	$\mathcal{H}^r_B$
	constant functions		zero-mean functions		

$$\mathcal{H}^r_{\text{Sob}} = \left\{ f \in \mathbb{R}^{[0,1]} \;\middle|\; \begin{array}{l} f(\cdot) = \gamma_0 + \sum_{i=1}^{r} \beta_i B_i(\cdot) + \sum_{k=1}^{\infty} \frac{a_k}{(2k\pi)^r} c_k(\cdot) + \sum_{k=1}^{\infty} \frac{b_k}{(2k\pi)^r} s_k(\cdot) \\ \text{with } \gamma_0 \in \mathbb{R},\ (\beta_i)_{1 \le i \le r} \in \mathbb{R}^r \text{ and } (a_k)_{k\ge1}, (b_k)_{k\ge1} \in \ell^2(\mathbb{N}^*, \mathbb{R}) \end{array} \right\}$$

Are Sobolev kernels characteristic?

	constant features		polynomial features		sinusoidal features
$K^r_{\text{Sob}} =$	1	$+$	k^r_A	$+$	k^r_B
$\downarrow$	$\downarrow$		$\downarrow$		$\downarrow$
$\mathcal{H}^r_{\text{Sob}} =$	$\text{Span}(\{1\})$	$\oplus$	$\mathcal{H}^r_A$	$\oplus$	$\mathcal{H}^r_B$
	constant functions		zero-mean functions		

$$\mathcal{H}^r_{\text{Sob}} = \left\{ f \in \mathbb{R}^{[0,1]} \;\middle|\; \begin{array}{l} f(\cdot) = \gamma_0 + \sum_{i=1}^{r} \beta_i B_i(\cdot) + \sum_{k=1}^{\infty} \frac{a_k}{(2k\pi)^r} c_k(\cdot) + \sum_{k=1}^{\infty} \frac{b_k}{(2k\pi)^r} s_k(\cdot) \\ \text{with } \gamma_0 \in \mathbb{R},\ (\beta_i)_{1 \le i \le r} \in \mathbb{R}^r \text{ and } (a_k)_{k\ge1}, (b_k)_{k\ge1} \in \ell^2(\mathbb{N}^*, \mathbb{R}) \end{array} \right\}$$

- The RKHS contains all Bernoulli polynomials, hence all polynomials:

$$B_0 = 1 \in \text{Span}(\{1\}) \quad ; \quad B_1, \dots, B_r \in \mathcal{H}^r_A \quad ; \quad (B_k)_{k \ge r+1} \in \mathcal{H}^r_B$$

- Stone-Weierstrass theorem → the polynomials are dense in $C([0,1])$.

K^r_{Sob} UNIVERSAL ⇔ K^r_{Sob} CHARACTERISTIC

Are Sobolev kernels characteristic?

		constant features		polynomial features		sinusoidal features
K^r_{Sob}	$=$	1	$+$	k^r_A	$+$	k^r_B
$\downarrow$		$\downarrow$		$\downarrow$		$\downarrow$
$\mathcal{H}^r_{\text{Sob}}$	$=$	$\text{Span}(\{1\})$	$\oplus$	$\mathcal{H}^r_A$	$\oplus$	$\mathcal{H}^r_B$
		constant functions		zero-mean functions		

$$\mathcal{H}^r_{\text{Sob}} = \left\{ f \in \mathbb{R}^{[0,1]} \,\middle|\, \begin{array}{l} f(\cdot) = \gamma_0 + \sum_{i=1}^{r} \beta_i B_i(\cdot) + \sum_{k=1}^{\infty} \frac{a_k}{(2k\pi)^r} c_k(\cdot) + \sum_{k=1}^{\infty} \frac{b_k}{(2k\pi)^r} s_k(\cdot) \\ \text{with } \gamma_0 \in \mathbb{R},\ (\beta_i)_{1 \le i \le r} \in \mathbb{R}^r \text{ and } (a_k)_{k\ge1}, (b_k)_{k\ge1} \in \ell^2(\mathbb{N}^*, \mathbb{R}) \end{array} \right\}$$

- Sobolev kernels K^r_{Sob} are **characteristic.**
  - ✓ First-order **HSIC-ANOVA indices** allow to detect independence:

$$S_i^{\text{HSIC}} = 0 \iff \text{HSIC}(X_i, Y) = 0 \iff X_i \perp Y$$

Conclusion

What must be remembered about Sobolev kernels?

$$K^r_{\text{Sob}}(x,x') = 1 + k^r_{\text{Sob}}(x,x') = 1 + k^r_A(x,x') + k^r_B(x,x')$$

- $\mathbb{P}_X = \mathcal{U}([0,1]) \rightarrow K^r_{\text{Sob}}(x,x')$
- $k^r_{\text{Sob}}(x,x')$: orthogonality constraint $\rightarrow$ HSIC-ANOVA
- $k^r_A(x,x')$: sum of separable functions $\rightarrow$ polynomial features
- $k^r_B(x,x')$: translation invariance $\rightarrow$ sinusoidal features

- They are **CHARACTERISTIC** and they therefore can be used to build **tests of independence**.

1 Sobolev kernel with $r = 1$

- The joint effect of k_A^1 and k_B^1 results in sinusoidal features.

2 Sobolev kernel with $r = 2$ → It behaves like the **dot-product kernel**.

- The eigenvalues of $T_{k_B^2}$ have a faster decay speed ($1/k^4$ instead of $1/k^2$).
- The sinusoidal features vanish. They are replaced by polynomial-like features.

MUST NOT BE USED

References

- **ARONSZAJN, N. 1950,** « Theory of reproducing kernels », *Transactions of the American mathematical society*, vol. 68, n°3, p. 337-404.
- **DA VEIGA, S. 2015,** « Global sensitivity analysis with dependence measures », *Journal of Statistical Computation and Simulation*, vol. 85, n°7, p. 1283-1305.
- **DA VEIGA, S. 2021,** « Kernel-based ANOVA decomposition and Shapley effects - Application to global sensitivity analysis », preprint.
- **DURRANDE, N., GINSBOURGER, D., ROUSTANT, O. and CARRARO, L. 2013**, « ANOVA kernels and RKHS of zero mean functions for model-based sensitivity analysis », *Journal of Multivariate Analysis*, vol. *115*, p. 57-67.
- **GINSBOURGER, D., ROUSTANT, O., SCHUHMACHER, D., DURRANDE, N. and LENZ, N. 2016**, « On ANOVA decompositions of kernels and Gaussian random field paths », *Monte Carlo and Quasi-Monte Carlo Methods*, pp. 315-330, Springer, Cham.
- **GRETTON, A., BOUSQUET, O., SMOLA, A. and SCHÖLKOPF, B. 2005,** « Measuring statistical dependence with Hilbert-Schmidt norms », *International conference on algorithmic learning theory*, pages 63-77, Springer.
- **IOOSS, B. and MARREL, A., 2019,** « Advanced methodology for uncertainty propagation in computer experiments with large number of inputs », *Nuclear Technology*.
- **MARREL, A., IOOSS, B. and CHABRIDON, V., 2020,** « Statistical identification of penalizing configurations in high-dimensional thermalhydraulic numerical experiments: The ICSCREAM methodology », preprint.
- **SOBOL, I., M., 1993,** « Sensitivity analysis for non-linear mathematical models », *Mathematical modelling and computational experiment*, vol. 1, p. 407-414.

anr agence nationale de la recherche

cea

THANKS FOR YOUR ATTENTION

Work funded and carried out within the SAMOURAI ANR research project.

Simulation **A**nalytics and **M**etamodel-based solutions for **O**ptimization, **U**ncertainty and **R**eliability **A**nalys**I**s

EDF SAFRAN IFP Energies nouvelles cea MINES Saint-Étienne CentraleSupélec GERAD

IRESNE I DER I SESI I LEMS

Institut de recherche sur les systèmes nucléaires pour la production d'énergie bas carbone