New insights into the feature maps of Sobolev kernels: application in global sensitivity analysis

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

NEW INSIGHTS INTO THE FEATURE MAPS OF SOBOLEV KERNELS: APPLICATION IN GLOBAL SENSITIVITY ANALYSIS

Gabriel Sarazin[1], Amandine Marrel[2,3], Sebastien Da Veiga[4] and Vincent Chabridon[5]

Abstract. As part of the study of an input-output numerical simulator, performing a sensitivity analysis allows to identify the input parameters having the greatest influence on the output variability. Since the variance-based approach (also known as the ANOVA framework) is too expensive, and the kernel-based approach (leading to HSIC indices) lacks interpretability, the HSIC-ANOVA framework has recently emerged to marry the advantages of both. A major particularity of this new methodology is the need to use the unanchored Sobolev kernels. This paper investigates how sensitivity is measured according to the chosen Sobolev kernel. To achieve this, at least one explicit feature map is extracted from each Sobolev kernel and this helps identify the dependence patterns captured by HSIC-ANOVA indices. For the Sobolev kernel of order $r = 1$, three different proof techniques are proposed to disclose its Mercer feature map. For higher-order Sobolev kernels ($r \geq 2$), it is proved that the Mercer feature map does not have a closed-form expression. In response, a slightly relaxed feature map is obtained after considering a sub-kernel decomposition. This latter feature map allows to justify why the Sobolev kernels of order $r \geq 2$ should not be used to estimate HSIC-ANOVA indices.

Résumé. Lors de l'étude d'un modèle numérique, la réalisation d'une analyse de sensibilité permet d'identifier les paramètres d'entrée les plus influents sur la sortie. Comme les méthodes basées sur le partage de la variance (approche ANOVA) sont trop coûteuses, et que les méthodes à noyaux (approche HSIC) sont difficiles à interpréter, la méthode HSIC-ANOVA a récemment vu le jour pour combiner leurs avantages respectifs. Une particularité majeure de cette méthode HSIC-ANOVA est l'utilisation d'une famille très spécifique de noyaux de Sobolev. Ce papier s'intéresse à la manière dont la sensibilité est mesurée selon le choix du noyau de Sobolev. Pour cela, on extrait de chacun des noyaux de Sobolev au moins une *feature map* car la connaissance des *features* aide ensuite à comprendre les motifs de dépendance qui sont capturés par les indices HSIC-ANOVA. Pour le noyau de Sobolev d'ordre $r = 1$, on propose trois techniques de preuve pour retrouver la *feature map* de Mercer de ce noyau. Pour les noyaux de Sobolev d'ordre supérieur ($r \geq 2$), on démontre que la *feature map* de Mercer n'a pas d'expression explicite. Une *feature map* d'un autre type est alors obtenue en considérant une décomposition des noyaux de Sobolev en plusieurs sous-noyaux. Cette *feature map* révèle notamment qu'il ne faut pas estimer les indices HSIC-ANOVA avec un noyau de Sobolev d'ordre $r \geq 2$.

2020 Mathematics Subject Classification. 34A30, 46E35, 47G10, 62G05, 62G10.

Keywords and phrases: Sensitivity analysis, Hilbert-Schmidt independence criterion, cross-covariance operators, Mercer kernels, kernel feature maps, kernel integral operators, ordinary differential equations.

[1] Université Paris-Saclay, CEA, DES, ISAS, DM2S, SGLS, 91191, Gif-sur-Yvette ; e-mail: gabriel.sarazin@cea.fr

[2] CEA, DES, IRESNE, DER, SESI, Cadarache, 13108, Saint-Paul-Lez-Durance ; e-mail: amandine.marrel@cea.fr

[3] Institut de Mathématiques de Toulouse, 118 Route de Narbonne, 31062, Toulouse

[4] ENSAI, 51 rue Blaise Pascal, 35170, Bruz ; e-mail: sebastien.da-veiga@ensai.fr

[5] EDF R&D, 6 Quai Watier, 78401, Chatou ; e-mail: vincent.chabridon@edf.fr

Submitted November 29, 2023.

1. Introduction

1.1. Kernel methods

Kernels, to be understood here as symmetric and positive definite functions of two arguments, are essential tools in probability and statistics. In probability, they are rather called *covariance functions* and they are used in stochastic calculus to characterize the properties of random fields and processes [63]. In statistics, they are the backbone of numerous non-linear methods designed to carry out machine learning tasks, whether supervised (*e.g.* Gaussian process regression [116], support vector classification [25], denoising with smoothing splines [112, 113]) or unsupervised (*e.g.* spectral clustering [7], kernel k-means [34], kernel PCA [90]). Kernels have also enabled the development of advanced methods in areas such as novelty detection [22, 23], Bayesian quadrature [6, 20], independent component analysis [8], hypothesis testing (especially independence testing [53] or two-sample testing [51]), optimal experimental design [75], optimal quantization [107] and many others. The ubiquity of kernels in statistical applications is mainly explained by their ability to introduce non-linearity where standard methods are only able to handle linearity. From a mathematical viewpoint, a kernel can be described in two different but equivalent ways.

First, in virtue of the Moore-Aronszajn theorem [5], every kernel is related to a unique *reproducing kernel Hilbert space* (RKHS) which is generated from the only knowledge of the kernel and inherits most of its properties (separability, measurability, boundedness, continuity, differentiability) [12, 25]. In particular, the link between a kernel and its RKHS is summarized by a reproducing property, which often leads to speak of *reproducing kernels* (instead of *kernels*). The RKHS offers a sound theoretical framework where some challenging problems find easy solutions. A classic example is provided by kernel ridge regression [86] where a penalized least-square score has to be minimized over a high-dimensional class of candidate functions [56, 112]. If an RKHS is chosen, the representer theorem [88] indicates that any minimizer can be expressed as a weighted linear combination of functions picked from the RKHS.

Then, beyond the abstract existence of the RKHS, there is always a transformation (called *feature map*) that allows to rewrite the kernel as the inner product in a specific Hilbert space (called *feature space*) between the representatives (called *feature functions*) of the two initial arguments [25]. The role played by a kernel within an algorithm is sometimes easier to understand by adopting the feature viewpoint. A typical example is the case of binary classification with support vector machines [57]. In presence of non-separable data (*i.e.* when linear boundaries fail to separate the two classes), the idea is to use a kernel in order to transport the data into a feature space where they become linearly separable. The knowledge of the kernel feature map then allows to clearly identify the transformations applied to the data to make them separable.

Global sensitivity analysis [31, 84] is another area where kernel methods overcome the limitations of the standard methodology and where the feature maps allow to understand how kernels perform the expected task.

1.2. Global sensitivity analysis

In many industrial fields, reliability and risk assessment is based on the joint use of computer simulation and uncertainty management. For highly sophisticated systems, the physical phenomena involved in accidental scenarios are modeled by numerical simulators which compute one (or several) output(s) of interest from a large number of uncertain input parameters. The generic Monte Carlo approach for uncertainty propagation relies on input-output evaluations of the simulator. As a complement, global sensitivity analysis (GSA) ambitions to quantify the influence of input uncertainties on the output uncertainty [16, 83]. When computer-based experiments are time-consuming, the maximum feasible number of runs may be very limited. The main challenge of GSA is then to deliver accurate sensitivity estimates from the few available simulation data.

The most prevailing approach to perform GSA is to apportion the output variance between all subsets of input variables [96]. Variance allocation is notably guided by the ANOVA[1] decomposition of the multivariate function representing the simulator [61]. The resulting Sobol' indices [96, 97] are very popular because their definition (as percentages of the output variance) is very intuitive. However, the computational burden required to accurately estimate them is often prohibitive. To bypass this problem, it is often preferred to adopt a pairwise approach based on the estimation and comparison of dependence measures [15, 17, 18]. This can be achieved in many different ways, but kernel methods are certainly the most attractive option, as they notably allow to define the *Hilbert Schmidt Independence Criterion* (HSIC) [52]. Computing the HSIC between two random variables (each previously assigned an appropriate kernel) amounts to using a dissimilarity measure to compare their joint distribution and their hypothetical distribution under independence (within an RKHS built from the two selected kernels). Interestingly, and unlike the numerator of high-order Sobol' indices, the HSIC can be estimated accurately, even from a limited number of samples. Hence, to meet the needs of GSA when simulation data are expensive, the HSIC can be applied to all input-output pairs [29], and the resulting sensitivity measures are called HSIC indices. Among their many advantages, HSIC indices are often praised for their ability to characterize independence because, assuming that all kernels are characteristic, an HSIC index is equal to zero if and only if the output variable is independent from the input variable.

Despite all this, HSIC indices suffer from a lack of interpretability, and this hinders their dissemination to a wider audience. There are actually two problems. Firstly, as HSIC indices do not arise from an ANOVA-like decomposition, their sum is not equal to one, and they cannot be universally upper bounded (*i.e.* any bound can be questioned with exotic choices of kernels and/or distributions). Secondly, the nature of the information captured by HSIC indices is not directly accessible, unless adopting the cross-covariance viewpoint [9, 109]. A prerequisite to understand finely the way HSIC indices measure sensitivity is the prior knowledge of the feature maps nested in the kernels chosen for the input and output variables. Many long-established results may be found in the dedicated literature [12, 25, 66, 76], especially for Gaussian and Matérn kernels. However, for a given kernel, there is no guarantee that a fully analytical feature map can be easily found. This must be discussed on a case-by-case basis, which makes the interpretation of HSIC indices very dependent on the selected kernels.

1.3. **New generation of HSIC indices based on Sobolev kernels**

In response to the first interpretation problem, an ANOVA-like decomposition has been recently set up for HSIC indices [30]. This HSIC-ANOVA decomposition allows to define the sensitivity indices of the same name. A contribution (defined in the HSIC style) can be assigned to each subset of inputs, and the sum of all contributions is then equal to one. This is very similar to Sobol' indices, except that all HSIC-ANOVA contributions can be estimated from a small amount of input-output data.

Unfortunately, the HSIC-ANOVA decomposition is obtained at the cost of strong restrictions on the input variables (assumed independent) and the input kernels (subject to additional orthogonality constraints). The input and output kernels must also be characteristic in order to preserve the ability to detect independence. Finding kernels that satisfy all these constraints is a major obstacle to the implementation of the HSIC-ANOVA methodology. The solution promoted in [30] is to apply a preliminary transformation to the input variables so that they all follow the standard uniform distribution[2], and then to assign them some specific Sobolev kernels (namely the *unanchored* Sobolev kernels).

Sobolev kernels have already proved to be efficient for spline regression [28, 56, 112]. They are also of great theoretical use to demonstrate the (strong) tractability of some quasi-Monte Carlo techniques in very high dimension [38, 77, 94]. Still from a theoretical perspective, they may be used to construct Hilbert spaces of multivariate functions in which the ANOVA components can be computed with simplified formulas [70, 114]. However, in the context of GSA, the use of Sobolev kernels raises a dilemma. On the one hand, Sobolev kernels are necessary to compute HSIC-ANOVA indices (which may be seen as a more interpretable collection of HSIC

[1]ANOVA means *ANalysis Of VAriance*. This acronym is encountered in many different areas of statistics [4, 49, 56, 87, 104].

[2]In this work, the standard uniform distribution refers to the uniform distribution on $[0,1]$. For convenience, this distribution will sometimes be denoted by $\mathcal{U}([0,1])$ in equations.

indices). On the other hand, this gain in interpretability is counterbalanced by a loss of transparency regarding the way sensitivity is measured. Therefore, the main objective of this paper is to investigate the mathematical properties of Sobolev kernels, especially those which have immediate consequences in GSA. For this, several points deserve to be clarified.

(Q1) We need to identify at least one orthonormal basis (ONB) for each Sobolev RKHS. The knowledge of the basis functions is indeed the key to understanding the preliminary transformations applied to the input variables when HSIC-ANOVA indices are computed with Sobolev kernels.

(Q2) We need to extract at least one easily interpretable feature map for each Sobolev kernel. Of all possible feature maps, those arriving in ℓ^2-spaces are particularly attractive because they often lead to an ONB of the RKHS.

(Q3) We need to know if Sobolev kernels are characteristic. This will determine the ability of HSIC-ANOVA indices to characterize independence.

1.4. Organization of the document

This paper is divided into seven sections which can be split into two parts. The first part, consisting of Sections 2 to 4, introduces all useful concepts before explaining why the acquisition of new knowledge about Sobolev kernels gives even more credit to the HSIC-ANOVA approach.

- Section 2 provides the essential ideas of the theory of reproducing kernels. The notion of feature map is introduced and its importance in the understanding of kernel action is emphasized. In particular, Section 2.4 presents two different strategies that can be used to identify an ONB of the RKHS associated to a given kernel. The first one (based on Mercer features) is well-known whereas the second one (based on ℓ^2-linearly independent features) is a first contribution of this work.

- Section 3 says a few words about Sobolev kernels. Their links with Sobolev spaces are detailed and their most relevant properties are presented. The question (Q3) is addressed at this step because it is ultimately quite simple after introducing all the necessary concepts. In addition, the unanchored Sobolev kernels are introduced as a specific family of Sobolev kernels built upon Bernoulli polynomials and parameterized by an integer $r \geq 1$ (accounting for the order of smoothness in the associated Sobolev space). Since the unanchored Sobolev spaces are among the few kernels fulfilling all the requirements of the HSIC-ANOVA approach, the study will be restricted to them.

- Section 4 is then focused on kernel-based GSA. The advantages of HSIC indices (over Sobol' indices) are highlighted. The added value of the newly-developed HSIC-ANOVA framework is justified and the problems caused by the use of Sobolev kernels are clearly stated. In this section, the information captured by HSIC-based sensitivity measures is thoroughly examined from the perspective of kernel feature maps. It is notably explained how basic manipulations on kernels sometimes allow to rewrite HSIC indices as series of explicit covariance terms.

In the second part of this work, Sections 5 to 8 strive to answer the questions (Q1) and (Q2). These four sections are therefore the core contribution of this work.

- In Section 5, a simulation-based method is applied to Sobolev kernels in order to get a first idea of the feature map arising from Mercer's theorem.

- In Section 6, the eigenvalue problem (solved numerically in Section 5) is transformed into a boundary value problem. In particular, it is shown that this equivalent version of the original problem can be solved analytically only for $r = 1$.

- In Section 7, a sub-kernel decomposition of Sobolev kernels is used to disclose a feature map of a different type which is a little less informative than Mercer's but holds whatever is $r \geq 2$.

- Section 8 deals with the limit Sobolev kernel, obtained in the hypothetical situation where $r = \infty$.

Section 9 provides some concluding remarks. The key findings are summarized and they are reformulated in terms of the original questions. Throughout this work, the proofs which are sufficiently short are inserted in the body of the text. On the contrary, the long and more technical ones are reported to Appendix F.

2. Basic reminders on kernels

In this first section, the objective is to introduce briefly the few elements from the theory of reproducing kernels that will be used in this work. Section 2.1 is dedicated to recall some basic definitions. Then, Section 2.2 deals with the particular links between kernels and probability distributions, in particular through the notions of orthogonal kernels and kernel mean embeddings. In Section 2.3, the fundamental concept of feature map is put forward and some examples are provided. The specific case of Mercer kernels is carefully examined in Section 2.4 as it will be of great importance in the forthcoming developments.

2.1. **Kernels and their reproducing kernel Hilbert spaces**

In this work, for the sake of simplicity, $\mathcal{X}$ will always denote an interval in $\mathbb{R}$ but most definitions and results can be extended to locally compact spaces in $\mathbb{R}^d$ (with $d \in \mathbb{N}^*$) or to even more general spaces. In the following, $\mathbb{R}^{\mathcal{X}}$ denotes the space of all real-valued functions defined on $\mathcal{X}$.

Definition 2.1. A function $K : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ is said to be a *kernel* if it is symmetric and positive definite.

- Symmetry: $\forall\, x, x' \in \mathcal{X}, \quad K(x,x') = K(x',x)$,
- Positive definiteness: $\forall\, n \geq 1, \quad \forall\, x_1, \dots, x_n \in \mathcal{X}, \quad \forall\, c_1, \dots, c_n \in \mathbb{R}, \quad \sum_{i=1}^{n}\sum_{j=1}^{n} c_i\, c_j\, K(x_i, x_j) \geq 0$.

After adopting a matrix viewpoint, the positive definiteness property may be rewritten as follows:

$$\boldsymbol{c}^T\, \boldsymbol{K}_n\, \boldsymbol{c} \geq 0 \quad \text{with} \quad \boldsymbol{K}_n := \Big[K(x_i, x_j)\Big]_{1\leq i,j\leq n} \quad \text{and} \quad \boldsymbol{c} := [c_i]_{1\leq i\leq n} \quad .$$

The n-by-n matrix $\boldsymbol{K}_n$ is called the Gram matrix (or more simply the kernel matrix).

Definition 2.2. Let $\mathcal{H}$ be a Hilbert space[3] (of functions $h : \mathcal{X} \to \mathbb{R}$) with inner product $\langle \cdot, \cdot \rangle_{\mathcal{H}}$. For any $x \in \mathcal{X}$, the evaluation functional at x is the linear form defined by $L_x : h \in \mathcal{H} \mapsto h(x) \in \mathbb{R}$. Then, $\mathcal{H}$ is said to be an RKHS if all evaluation functionals are continuous:

$$\forall\, x \in \mathcal{X}, \quad \exists\, C_x > 0 \quad \text{such that} \quad \forall\, h \in \mathcal{H}, \quad |h(x)| \leq C_x\, \|h\|_{\mathcal{H}} \ .$$

Definition 2.3. Let $K : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ be a kernel and let $\mathcal{H}$ be an RKHS (of functions $h : \mathcal{X} \to \mathbb{R}$). K is said to be a *reproducing kernel* of $\mathcal{H}$ if the two following conditions are verified:

- Embedding property: $\forall\, x \in \mathcal{X}, \quad K(\cdot, x) \in \mathcal{H}$,
- Reproducing property: $\forall\, x \in \mathcal{X}, \quad \forall\, h \in \mathcal{H}, \quad h(x) = \langle h, K(\cdot, x)\rangle_{\mathcal{H}}$.

Remark 2.4. In Definition 2.3, the two points are crucial. Together, they notably lead to the fundamental property verified by K:

$$\forall\, x, x' \in \mathcal{X}, \quad K(x,x') = \langle K(\cdot, x)\,, K(\cdot, x')\rangle_{\mathcal{H}} \ . \tag{2.1}$$

[3]Strictly speaking, a Hilbert space is an ordered pair $(\mathcal{H}, \langle \cdot, \cdot \rangle_{\mathcal{H}})$ where $\mathcal{H}$ is a space and $\langle \cdot, \cdot \rangle_{\mathcal{H}}$ is an inner product on $\mathcal{H}$. However, a commonly accepted abuse of notation is to write $\mathcal{H}$ instead of $(\mathcal{H}, \langle \cdot, \cdot \rangle_{\mathcal{H}})$. Remember that an inner product space $(\mathcal{H}, \langle \cdot, \cdot \rangle_{\mathcal{H}})$ is a Hilbert space if and only if the normed space $(\mathcal{H}, \|\cdot\|_{\mathcal{H}})$ built from the induced norm $\|\cdot\|_{\mathcal{H}}$ is complete. Otherwise, the mathematical structure $(\mathcal{H}, \langle \cdot, \cdot \rangle_{\mathcal{H}})$ is merely a pre-Hilbert space.

In fact, any RKHS admits a reproducing kernel and any kernel is reproducing for an RKHS. More precisely, there is a one-to-one mapping between kernels and RKHSs. For a given RKHS, the Riesz-Fréchet representation theorem [19] (see Theorem 5.5, p. 135) applied to all evaluation functionals allows to construct a reproducing kernel and it is the only one for this RKHS. Conversely, for a given kernel K, the linear span of the functions $K(\cdot, x)$ is a pre-Hilbert which can be completed into a Hilbert space. By uniqueness of the completion, this Hilbert space is the only RKHS having K as reproducing kernel [5].

Theorem 2.5 (Moore-Aronszajn)**.** *For any kernel $K : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$, there is a unique Hilbert space for which K is a reproducing kernel.*

A detailed proof is provided in [12] (see Theorem 3, p. 19).

Remark 2.6. If K_1 and K_2 are two kernels on $\mathcal{X}$, it is straightforward to see that $K_1 + K_2$ is also a kernel on $\mathcal{X}$. The RKHS induced by $K_1 + K_2$ is not so simple to characterize [12] (see Theorem 5, p. 24). In the particular case where $\mathcal{H}_1 \cap \mathcal{H}_2 = \{0\}$, it is exactly the direct sum $\mathcal{H}_1 \oplus \mathcal{H}_2$. If K_1 and K_2 are defined on two (different) intervals $\mathcal{X}_1$ and $\mathcal{X}_2$, it can be proved[4] that the tensor product $K_1 \otimes K_2$ is a kernel on $\mathcal{X}_1 \times \mathcal{X}_2$. The related RKHS (denoted by $\mathcal{H}_1 \otimes \mathcal{H}_2$) is obtained after completing the linear span of all the functions $h_1 \otimes h_2$ where $h_1 \in \mathcal{H}_1$ and $h_2 \in \mathcal{H}_2$ [12] (see Theorem 13, p. 31).

2.2. Links between kernels and distributions

For any interval $\mathcal{X} \subseteq \mathbb{R}$, the space of all Borel[5] probability measures on $\mathcal{X}$ is denoted by $\mathcal{M}_1^+(\mathcal{X})$.

2.2.1. *Orthogonal kernels*

Definition 2.7. A kernel $k : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ is said to be *orthogonal*[6] with respect to $\nu \in \mathcal{M}_1^+(\mathcal{X})$ if one has:

$$\forall\, x \in \mathcal{X}, \quad \int_{\mathcal{X}} k(\xi, x)\, \mathrm{d}\nu(\xi) = 0 \; . \tag{2.2}$$

For fixed $\nu \in \mathcal{M}_1^+(\mathcal{X})$, let k be an orthogonal kernel and let $\mathcal{F}$ be the associated RKHS. By definition, all the functions $k(\cdot, x)$ with $x \in \mathcal{X}$ have zero mean. Due to the denseness of the pre-Hilbert space $\mathcal{F}^{\text{pre}} := \text{Span}(\{k(\cdot, x) \text{ with } x \in \mathcal{X}\})$ in $\mathcal{F}$, the zero-mean property can actually be extended to any function $f \in \mathcal{F}$. As a result, $\mathcal{F}$ does not contain any non-zero constant function, which may be written as $\mathbb{R} \cap \mathcal{F} = \{0\}$. In this context, $\mathbb{R}$ does not represent the real line but the space of all constant functions on $\mathcal{X}$. This space is an RKHS, and more precisely the RKHS induced by the constant kernel $(x, x') \mapsto 1$.

Definition 2.8. A kernel $K : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ is said to be ANOVA (with respect to ν) if $K = 1 + k$ where $k : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ is an orthogonal kernel (with respect to ν).

Let $K = 1 + k$ be an ANOVA kernel. As it is known that $\mathbb{R} \cap \mathcal{F} = \{0\}$, one can write $\mathcal{H} = \mathbb{R} \oplus \mathcal{F}$.

For most widespread parametric families of distributions, no ANOVA kernel can be found in the literature. The only exception is the standard uniform distribution. In this particular case, some Sobolev kernels naturally satisfy the orthogonality constraint. This point will be further discussed in Section 3.

---

[4]This directly follows from the Schur product theorem [105] (see Theorem 3.1, p. 221) which guarantees that the Hadamard product of two Gram matrices is indeed a symmetric positive semi-definite matrix.

[5]Let $\mathcal{B}(\mathcal{X})$ be the Borel σ-algebra of $\mathcal{X}$, *i.e.* the σ-algebra generated by the open sets of $\mathcal{X}$. Any probability measure ν defined on $\mathcal{B}(\mathcal{X})$ is called a Borel probability measure on $\mathcal{X}$.

[6]In order to establish a clear distinction with other kernels, the letters k and $\mathcal{F}$ will be used (instead of K and $\mathcal{H}$) for denoting orthogonal kernels and their RKHSs.

2.2.2. *Embedding in* L^p*-spaces*

Let $K : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ be a kernel and let $\nu \in \mathcal{M}_1^+(\mathcal{X})$. If the random variable $K(X,X)$ with $X \sim \nu$ verifies some finite-moment assumptions, it can easily be shown that $\mathcal{H} \subset \mathbb{R}^{\mathcal{X}}$ is contained in an L^p-space related to ν:

$$\forall\, p \geq 1, \quad \mathbb{E}_\nu\left[K(X,X)^{p/2}\right] < \infty \quad \Longrightarrow \quad \mathcal{H} \subset \mathrm{L}^p(\mathcal{X},\nu)\ . \tag{2.3}$$

Thus, it is enough to assume that $\mathbb{E}_\nu[K(X,X)] < \infty$ for $\mathcal{H}$ to be part of $\mathrm{L}^2(\mathcal{X},\nu)$. In particular, if K is a bounded kernel, $\mathcal{H}$ is contained in any space $\mathrm{L}^p(\mathcal{X},\nu)$ with $p \geq 1$ and $\nu \in \mathcal{M}_1^+(\mathcal{X})$.

2.2.3. *Kernel embedding of probability distributions*

Under the only assumption that $\mathbb{E}_\nu\left[\sqrt{K(X,X)}\right] < \infty$, the kernel mean embedding $\mu_\nu \in \mathcal{H}$ of the probability distribution ν can be defined as:

$$\forall\, x \in \mathcal{X}, \quad \mu_\nu(x) = \mathbb{E}_\nu\left[K(X,x)\right] = \int_{\mathcal{X}} K(\xi,x)\,\mathrm{d}\nu(\xi)\ . \tag{2.4}$$

The embedding mechanism allows to define a dissimilarity measure on $\mathcal{M}_1^+(\mathcal{X})$ which is called the *maximum mean discrepancy* (MMD) since [50]:

$$\forall\, \nu_1,\nu_2 \in \mathcal{M}_1^+(\mathcal{X}), \quad \mathrm{MMD}(\nu_1,\nu_2) := \|\mu_{\nu_1} - \mu_{\nu_2}\|_{\mathcal{H}}\ .$$

For two given probability measures ν_1 and ν_2, the MMD must be understood as the distance between their images μ_{ν_1} and μ_{ν_2} in a function space where the two images are well-defined and the distance can be easily computed (or at least approximated from sample data). After reverting to the initial definition of the norm $\|\cdot\|_{\mathcal{H}}$ in terms of the inner product $\langle\cdot,\cdot\rangle_{\mathcal{H}}$, an alternative formula of the MMD can be derived:

$$\forall\, \nu_1,\nu_2 \in \mathcal{M}_1^+(\mathcal{X}), \quad \mathrm{MMD}^2(\nu_1,\nu_2) = \mathbb{E}_{\nu_1\otimes\nu_1}[K(X,X')] + \mathbb{E}_{\nu_2\otimes\nu_2}[K(X,X')] - 2\,\mathbb{E}_{\nu_1\otimes\nu_2}[K(X,X')]\ . \tag{2.5}$$

In the first (resp. second) term, X and X' are two independent copies of ν_1 (resp. ν_2). By way of comparison, in the third term, X and X' are still independent but they are, this time, distributed according to ν_1 and ν_2.

Definition 2.9. Let $K : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ be a kernel with RKHS denoted by $\mathcal{H}$. The kernel K is said to be characteristic to $\mathcal{M}_1^+(\mathcal{X})$ if the map $\nu \in \mathcal{M}_1^+(\mathcal{X}) \mapsto \mu_\nu \in \mathcal{H}$ is injective.

Most of the time, it is quite difficult to determine whether a given kernel is characteristic or not. A necessary condition for a kernel to be characteristic is to generate an infinite-dimensional RKHS. For some specific types of kernels, there also exist sufficient conditions (no longer involving the notion of kernel mean embedding) [101–103]. A recent overview of existing results is also provided in [92].

Remark 2.10. Let $K_1 : \mathcal{X}_1\times\mathcal{X}_1 \to \mathbb{R}$ and $K_2 : \mathcal{X}_2\times\mathcal{X}_2 \to \mathbb{R}$ be two kernels which are respectively characteristic to $\mathcal{M}_1^+(\mathcal{X}_1)$ and $\mathcal{M}_1^+(\mathcal{X}_2)$. Then, the tensor product kernel $K_1 \otimes K_2$ is characteristic to $\mathcal{M}_1^+(\mathcal{X}_1 \times \mathcal{X}_2)$. Against all odds, the generalization of this result to more than two kernels is not so simple. Indeed, it is true for translation-invariant kernels [101] (see Proposition 8, p. 777) but false in general [106] (see Example 2, p. 9).

2.3. Kernels and their feature maps

Definition 2.11. For a given kernel $K : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$, let us assume that there exists a Hilbert space $\mathcal{G}$ and a map $\psi : \mathcal{X} \to \mathcal{G}$ such that:

$$\forall\, x,x' \in \mathcal{X}, \quad K(x,x') = \langle\psi(x),\psi(x')\rangle_{\mathcal{G}}\ . \tag{2.6}$$

ψ is called a *feature map*, $\mathcal{G}$ a *feature space* and any object $\psi(x)$ a *feature function* (or simply a *feature*).

Having access to a feature map $\psi : \mathcal{X} \to \mathcal{G}$ provides a different overview of the kernel K. In particular, if ψ is sufficiently explicit, the kernel behavior (in other words, the mathematical operations performed on the provided data when K is evaluated) may be understood at a deeper level. Given two points x and x' in $\mathcal{X}$, computing the value $K(x, x')$ amounts to:

- creating two feature functions $\psi(x)$ and $\psi(x')$ living in a possibly very sophisticated feature space $\mathcal{G}$,
- comparing these two feature functions with the metric existing in $\mathcal{G}$.

When $\mathcal{G}$ is infinite-dimensional (for instance when $\mathcal{G}$ is a sequence space or a function space), the metric in $\mathcal{G}$ is expected to capture a much richer and more subtle information than the initial metric in $\mathcal{X}$. On the one side, from a mathematical viewpoint, computing $K(x, x')$ amounts to applying preliminary transformations to x and x' before comparing their respective images $\psi(x)$ and $\psi(x')$ in a higher-dimensional feature space. On the other side, from a numerical viewpoint, everything remains tractable because the value $K(x, x')$ is computed with a simple analytical formula. Thus, the feature space (hidden within the kernel structure) can be leveraged without even knowing the feature map ψ and the inner product $\langle \cdot, \cdot \rangle_{\mathcal{G}}$. Kernel evaluations are indeed sufficient to virtually create pairs of feature functions and measure their similarity in the feature space. This property, known as the *kernel trick*, explains in large part the great popularity of kernel methods [62, 110].

Mathematically speaking, there always exists a feature map. Indeed, for any given kernel K, the fundamental property stated in Eq. (2.1) always holds and therefore always offers a naive way of satisfying Eq. (2.6):

$$\forall\, x, x' \in \mathcal{X}, \quad K(x, x') = \langle \theta(x), \theta(x') \rangle_{\mathcal{H}} \quad \text{with} \quad \theta(x) := K(\cdot, x) \in \mathcal{H}\ .$$

$\theta : \mathcal{X} \to \mathcal{H}$ is called the *canonical feature map*[7] because the feature space is exactly the RKHS induced by K. Unfortunately, this feature map is not very informative in most cases. In fact, it is often difficult to understand intuitively the true mathematical nature of the operations that allow to transform a point $x \in \mathcal{X}$ into the function $K(\cdot, x) = [\theta(x)](\cdot) \in \mathcal{H} \subset \mathbb{R}^{\mathcal{X}}$, especially when $\mathcal{H}$ is unknown or partly implicit.

Thankfully, the canonical feature map is not the only existing feature map. Depending on the kernel, there may exist many other feature maps and the associated feature spaces may have very different natures. Sometimes, the use of basic tools in mathematical analysis is sufficient to extract a feature map from the analytical expression of the kernel under study, as illustrated in the example below.

Example 2.12. The Gaussian kernel $K_G : \mathbb{R}^2 \to \mathbb{R}$ (with bandwidth parameter $\gamma > 0$) is defined by:

$$\forall\, x, x' \in \mathbb{R}, \quad K_G(x, x') = \exp\left[-\frac{1}{2} \left(\frac{x - x'}{\gamma} \right)^2 \right]. \tag{2.7}$$

Remember that the Taylor series expansion of $\exp(\cdot)$ holds everywhere on $\mathbb{R}$ because the associated radius of convergence is infinite. Hence, the initial expression of K_G can thus be rewritten as follows:

$$K_G(x, x') = \exp\left[-\frac{1}{2} \left(\frac{x}{\gamma} \right)^2 \right] \exp\left[-\frac{1}{2} \left(\frac{x'}{\gamma} \right)^2 \right] \exp\left(\frac{x\, x'}{\gamma^2} \right) \text{ with } \exp\left(\frac{x\, x'}{\gamma^2} \right) = \sum_{k=0}^{\infty} \frac{1}{k!} \left(\frac{x}{\gamma} \right)^k \left(\frac{x}{\gamma} \right)^k .$$

This points the way towards a feature map ψ_G from $\mathbb{R}$ to the Hilbert space $\ell^2(\mathbb{N})$ of all square-summable sequences indexed by $\mathbb{N}$:

$$\begin{aligned} K_G(x, x') = \langle \psi_G(x), \psi_G(x') \rangle_{\ell^2(\mathbb{N})} \quad & \text{where} \quad \psi_G(x) = [\, \psi_k(x) \,]_{k \geq 0} \in \ell^2(\mathbb{N}) \\ & \text{with} \quad \psi_k(x) := \frac{1}{\sqrt{k!}} \exp\left[-\frac{1}{2} \left(\frac{x}{\gamma} \right)^2 \right] \left(\frac{x}{\gamma} \right)^k . \end{aligned} \tag{2.8}$$

[7] In the rest of the paper, the Greek letter θ will be exclusively reserved for denoting the canonical feature maps.

As the feature space is now $\ell^2(\mathbb{N})$, ψ_G can be interpreted as a catalogue of reference transformations applied to the data. Here, the catalogue consists of an infinite number of damped polynomial features.

Beyond mathematical tricks, for certain types of kernels, there also exists a well-adapted framework to address feature extraction. The case of Mercer kernels is covered in the next section.

2.4. **Focus on Mercer kernels**

2.4.1. *Mercer expansions and their* L^2*-orthogonal feature maps*

From now on, the space of all continuous functions on $\mathcal{X}$ is denoted by $\mathrm{C}(\mathcal{X})$.

Definition 2.13. A kernel $K : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ is said to be a Mercer kernel if $\mathcal{X}$ is compact and K is continuous.

Generally speaking, the RKHS associated to any bounded and continuous kernel K is composed of bounded and continuous functions [25] (see Lemma 4.28, p. 128), meaning that $\mathcal{H} \subset \mathrm{C}(\mathcal{X})$.

Definition 2.14. A Mercer kernel $K : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ is said to be *universal* if $\mathcal{H}$ is uniformly dense in $\mathrm{C}(\mathcal{X})$:

$$\forall\, f \in \mathrm{C}(\mathcal{X}), \quad \forall\, \epsilon > 0, \quad \exists\, h_\epsilon \in \mathcal{H} \quad \text{such that} \quad \|f - h_\epsilon\|_\infty := \sup_{x \in \mathcal{X}} |f(x) - h_\epsilon(x)| < \epsilon \ .$$

Remark 2.15. Several variants of Definition 2.14 may be found in the literature [92, 101]. In particular, a continuous kernel K defined on $\mathbb{R}$ is said to be c_0-universal if $\mathcal{H}$ is uniformly dense in the space $\mathrm{C}_0(\mathbb{R})$ of all continuous functions vanishing at infinity.

Remark 2.16. For a Mercer kernel, being universal is a sufficient condition to be characteristic to $\mathcal{M}_1^+(\mathcal{X})$ [101] (see Theorem 13, p. 778).

Let $K : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ be a Mercer kernel and let ν be a probability measure (with support $\mathcal{X}$). K and ν are the only two elements needed to define the following integral transform:

$$\begin{array}{rccc} T_K : & \mathrm{L}^2(\mathcal{X}, \nu) & \longrightarrow & \mathrm{L}^2(\mathcal{X}, \nu) \\ & f & \longmapsto & T_K\, f \end{array} \quad \text{with} \quad \big[T_K\, f\big](x) := \mathbb{E}_\nu\big[K(x, X)\, f(X)\big] = \int_{\mathcal{X}} K(x, \xi)\, f(\xi)\, \mathrm{d}\nu(\xi) \ . \tag{2.9}$$

The linear operator T_K is called the *kernel integral operator*. Changing either K or ν gives rise to another operator with possibly very different properties. The integral operators built from Mercer kernels and Borel probability measures have many remarkable properties. Only those which are directly useful for the rest of this work are recalled hereafter.

T_K is a Hilbert-Schmidt operator (*i.e.* the Hilbert-Schmidt norm of T_K is finite). In particular, one has:

$$\|T_K\|_{\mathrm{HS}}^2 := \sum_{i \geq 1} \sum_{j \geq 1} |\langle T_K\, e_i\, , e_j \rangle_{\mathrm{L}^2}|^2 = \sum_{i \geq 1} \sum_{j \geq 1} |\langle K, e_i \otimes e_j \rangle_{\mathrm{L}^2}|^2 = \|K\|_{\mathrm{L}^2}^2 < \infty \ , \tag{2.10}$$

where $(e_i)_{i \geq 1}$ denotes any possible ONB of $\mathrm{L}^2(\mathcal{X}, \nu)$. The first equality is obtained by combining Eq. (2.9) and Fubini's theorem in $\mathrm{L}^2(\mathcal{X}^2, \nu^{\otimes 2})$. The second one follows from Parseval's identity after noting that the tensorized system $(\phi_i \otimes \phi_j)_{i,j \geq 1}$ is an ONB of $\mathrm{L}^2(\mathcal{X}^2, \nu^{\otimes 2})$. In addition, $\|K\|_{\mathrm{L}^2}$ is always a finite norm because K is a Mercer kernel (and is therefore bounded on $\mathcal{X}$).

Remark 2.17. In light of Eq. (2.9) and (2.10), it is enough to assume that $K \in \mathrm{L}^2(\mathcal{X}^2, \nu^{\otimes 2})$ for T_K to be both well-defined and Hilbert-Schmidt. However, only the case of Mercer kernels is considered here because it provides a nice framework where all relevant results can be stated without restriction.

In addition to being Hilbert-Schmidt, T_K is positive, compact and self-adjoint. These properties come from the fact that T_K is built from a Mercer kernel. At this point, the spectral theory of linear operators allows to go a step further. In fact, the spectral theorem for compact self-adjoint operators [25] (see Theorem A.5.13, p. 505) provides an eigendecomposition of T_K which is the cornerstone to demonstrate Mercer's theorem [67] (see Theorem 3.a.1, p. 145).

Theorem 2.18 (Mercer)**.** *Let $K : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ be a Mercer kernel and let ν be a probability measure with support $\mathcal{X}$. Let $T_K : \mathrm{L}^2(\mathcal{X},\nu) \to \mathrm{L}^2(\mathcal{X},\nu)$ denote the resulting kernel integral operator. Then, there exists an ONB of $\mathrm{L}^2(\mathcal{X},\nu)$ denoted by $(\phi_i)_{i\geq 1}$ which is only composed of eigenfunctions of T_K. The associated eigenvalues $(\lambda_i)_{i\geq 1}$ are all non-negative and the eigenfunctions corresponding to positive eigenvalues are continuous on $\mathcal{X}$. In addition, K can be decomposed as follows:*

$$\forall\, x, x' \in \mathcal{X}, \quad K(x,x') = \sum_{i=1}^{\infty} \lambda_i\, \phi_i(x)\, \phi_i(x') \quad \textit{where} \quad \lambda_1 \geq \lambda_2 \geq \ldots \geq 0\ , \tag{2.11}$$

and the convergence of the series is absolute and uniform.

Eq. (2.11) is called the *Mercer decomposition* (or the *Mercer expansion*, or the *Mercer representation*) of K with respect to ν. The *rank* of K is defined as the number of positive eigenvalues (counted with multiplicity) in the eigenspectrum $\boldsymbol{\lambda}(T_K) := (\lambda_i)_{i\geq 1}$. If K only has a finite number of positive eigenvalues, it is said to be a *degenerate kernel* (or a *finite-rank kernel*).

Remark 2.19. Using the ONB of eigenfunctions $(\phi_i)_{i\geq 1}$ in Eq. (2.10) leads to $\|T_K\|_{\mathrm{HS}}^2 = \sum_{i\geq 1} \lambda_i^2 < \infty$. For a Mercer kernel, a little more is known about the decay speed of the eigenvalues. In fact, T_K is actually a trace-class operator [21], which implies that $\mathrm{Tr}(T_K) = \sum_{i\geq 1} \lambda_i < \infty$. Hence, one has $(\lambda_i)_{i\geq 1} \in \ell^1(\mathbb{N}^*)$.

Remark 2.20. For a fixed probability measure ν, the Mercer decomposition established in Eq. (2.11) is not uniquely defined. To be more precise, the eigenspectrum and the eigenspaces of T_K are invariant, but there may be many different ways to pick the eigenfunctions from the eigenspaces, especially if some eigenspaces have dimension greater than or equal to 2.

Theorem 2.18 may be useful for feature extraction because it provides a feature map φ from $\mathcal{X}$ to $\ell^2(\mathbb{N}^*)$:

$$K(x,x') = \langle \varphi(x), \varphi(x') \rangle_{\ell^2(\mathbb{N}^*)} \quad \text{where} \quad \varphi(x) := \left(\sqrt{\lambda_i}\, \phi_i(x)\right)_{i\geq 1}\ . \tag{2.12}$$

φ is called the Mercer feature map[8] of K (with respect to ν). Let us insist on the fact that φ strongly depends on ν (either for the decay speed of the eigenvalues or for the shapes of the eigenfunctions). If ν is replaced by another probability measure, a completely different collection of Mercer features could emerge from T_K.

Theorem 2.18 must not be seen as a miracle solution because the features revealed in Eq. (2.12) are only defined implicitly. In fact, the ability to extract a Mercer feature map φ relies on the ability to solve the infinite-dimensional eigenvalue problem defined by:

$$T_K\, \phi = \lambda\, \phi \quad \text{with} \quad \phi \in \mathrm{L}^2(\mathcal{X},\nu) \quad \text{and} \quad \lambda > 0\ . \tag{2.13}$$

For common characteristic kernels and probability distributions, this eigenvalue problem seldom has a closed-form solution. A list of the few rares examples where this happens is given in Appendix B.

For a given Mercer kernel K, the knowledge of an explicit Mercer decomposition allows to characterize the related RKHS thanks to the eigenvalues and eigenfunctions of T_K.

Theorem 2.21. *In regard of the Mercer expansion stated in Eq.* (2.11)*, the RKHS induced by K is:*

$$\mathcal{H} = \left\{ h \in \mathbb{R}^{\mathcal{X}} \;\middle|\; h(\cdot) = \sum_{i=1}^{\infty} a_i \sqrt{\lambda_i}\, \phi_i(\cdot) \;\; \textit{with} \;\; (a_i)_{i\geq 1} \in \ell^2(\mathbb{N}^*) \right\}, \tag{2.14}$$

[8] In the rest of the paper, the Greek letter φ will be exclusively reserved for denoting the Mercer feature maps.

with inner product:

$$\begin{array}{cccccc} \langle \cdot,\cdot \rangle_{\mathcal{H}} : & \mathcal{H} & \times & \mathcal{H} & \longrightarrow & \mathbb{R} \\ & \Bigg(h_1(\cdot) = \sum_{i=1}^{\infty} a_i \sqrt{\lambda_i}\, \phi_i(\cdot) & , & h_2(\cdot) = \sum_{i=1}^{\infty} b_i \sqrt{\lambda_i}\, \phi_i(\cdot) \Bigg) & \longmapsto & \sum_{i=1}^{\infty} a_i\, b_i \ . \end{array} \tag{2.15}$$

The system $(\sqrt{\lambda_i}\,\phi_i)_{i\geq 1}$ is therefore an ONB of $\mathcal{H}$.

The reader is referred to [25] (see Theorem 4.51, pp. 150–151) for the detailed proof.

Remark 2.22. With Eq. (2.14), it can be seen that the size of the RKHS is directly linked to the decay rate of the eigenvalues $(\lambda_i)_{i\geq 1}$. The slower the decay rate, the larger the RKHS. Thus, for two kernels having the same basis of eigenfunctions $(\phi_i)_{i\geq 1}$, the RKHS induced by the kernel having the fastest decay rate is contained in the RKHS induced by the kernel having the slowest decay rate.

2.4.2. *Kernel expansions leading to non-orthogonal feature maps*

Theorem 2.21 is actually a specific case of a more general result which applies to any kernel (whether Mercer or not) for which a series expansion is known. Indeed, if a kernel can be decomposed as a series of symmetric and separable functions, a feature map can be directly identified and a feature-based characterization of the associated RKHS follows from this decomposition. This result is stated in the theorem below.

Theorem 2.23. *Let $\mathcal{X} \subseteq \mathbb{R}$ be an interval and let $K : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ be a kernel. Let us assume that there exist a countable set I and a system $(g_i)_{i\in I}$ of ℓ^2-linearly independent functions such that:*

$$\forall\, x, x' \in \mathcal{X}, \quad K(x,x') = \sum_{i\in I} g_i(x)\, g_i(x') \ . \tag{2.16}$$

Then, the RKHS induced by K is:

$$\mathcal{H} = \left\{ h \in \mathbb{R}^{\mathcal{X}} \;\middle|\; h(\cdot) = \sum_{i\in I} a_i\, g_i(\cdot) \ \textit{ with } \ (a_i)_{i\in I} \in \ell^2(I) \right\}, \tag{2.17}$$

with inner product:

$$\begin{array}{cccccc} \langle \cdot,\cdot \rangle_{\mathcal{H}} : & \mathcal{H} & \times & \mathcal{H} & \longrightarrow & \mathbb{R} \\ & \Bigg(h_1(\cdot) = \sum_{i\in I} a_i\, g_i(\cdot) & , & h_2(\cdot) = \sum_{i\in I} b_i\, g_i(\cdot) \Bigg) & \longmapsto & \sum_{i\in I} a_i\, b_i \ . \end{array} \tag{2.18}$$

The system $(g_i)_{i\in I}$ is therefore an ONB of $\mathcal{H}$.

The reader is referred to Appendix F.1 for the detailed proof. It is much inspired from what is done in [25] (see the proof of Theorem 4.51, pp. 150–151) but it is more general since no assumption is made on K.

Remark 2.24. An important point of Theorem 2.23 lies in the fact that the system $(g_i)_{i\in I}$ has to be ℓ^2-linearly independent. As pointed out in [81], there exist several non-equivalent ways of defining linear independence in an infinite-dimensional Hilbert space. Among them, ℓ^2-linear independence means that the implication:

$$\textstyle\sum_{i\in I} a_i\, g_i(\cdot) = 0 \implies \forall\, i \in I, \quad a_i = 0 \tag{2.19}$$

holds for any square-summable sequence $(a_i)_{i\in I} \in \ell^2(I)$ such that the series $\sum_{i\in I} a_i\, g_i(\cdot)$ converges everywhere on $\mathcal{X}$. By way of comparison, the system $(g_i)_{i\in I}$ is said to be ω-independent if Eq. (2.19) is verified for any sequence leading to a convergent series. For this reason, the ℓ^2-linear independence of the features $(g_i)_{i\in I}$ is the weakest assumption one can make for Eq. (2.18) to define an inner product in $\mathcal{H}$. In particular, for any function $h(\cdot) = \sum_{i\in I} a_i\, g_i(\cdot)$ belonging to $\mathcal{H}$, this assumption ensures that the sequence $(a_i)_{i\in I}$ is unique in $\ell^2(I)$. It also guarantees that the series $\sum_{i\in I} a_i\, b_i$ in Eq. (2.18) has always finite sum.

For a given kernel, Theorem 2.23 reveals that finding a decomposition in the same form as Eq. (2.16) is enough to extract a feature map, characterize the RKHS and derive an ONB. Mercer's theorem is nothing but a particular case of this general result, where the prior choice of a probability measure allows to define an integral operator (making L^2-orthogonal features emerge). Hence, the search for a feature map can be done in two different ways:

- One may try to write K as a sum (or series) of symmetric and separable functions using various expansion methods.
- One may try to find a probability measure $\nu \in \mathcal{M}_1^+(\mathcal{X})$ for which the eigenvalue problem in Eq. (2.13) can be solved analytically.

The first strategy seems simpler but requires the right intuition about how to transform the kernel formula. In comparison, the framework laid by Mercer's theorem, although not offering a miracle solution, clearly indicates a class of eigenvalue problems whose solutions disclose features.

The next section is devoted to a particular family of Mercer kernels which is derived from Sobolev spaces.

3. Sobolev kernels

In this section, the main goal is to provide a quick overview of Sobolev kernels. First, Section 3.1 describes the Hilbertian setting from which Sobolev kernels originate. In particular, it will be shown that they emerge from the Sobolev spaces $\mathrm{H}^r([0,1])$ after using well-adapted inner products. For $r \geq 2$, we will see in Section 3.2 that it not so simple to obtain a kernel with closed-form expression unless considering the unanchored Sobolev spaces. Some general properties of Sobolev kernels will eventually be put forward in Section 3.3.

Throughout this section, when nothing is mentioned, the reference probability measure is the standard uniform distribution. This remark mainly concerns orthogonality properties and Mercer expansions.

3.1. **Sobolev spaces and their reproducing kernels**

Before we start, let us take time to clarify some notations (although they are, admittedly, very standard in mathematical analysis).

- $\mathrm{C}([0\,,1])$ is the space of all continuous functions on $[0\,,1]$.
- $\mathrm{C}^k([0\,,1])$ is the space of all k-times continuously differentiable functions on $[0\,,1]$.
- $\mathrm{C}^\infty([0\,,1])$ is the space of all infinitely differentiable functions on $[0\,,1]$.
- $\mathrm{C}_0^\infty([0\,,1])$ is the space of all infinitely differentiable functions on $[0\,,1]$ such that $\phi(0) = \phi(1) = 0$.
- For any $k \geq 1$ and for any $h \in \mathrm{C}^k([0\,,1])$, $h^{[k]}$ is the k-th (classic) derivative of f on $[0\,,1]$. The notations $h' := h^{[1]}$ and $h'' := h^{[2]}$ will also be used in some situations that lend themselves well to this.

Now, let us consider $h \in \mathrm{L}^1([0\,,1])$. Remember that there exists a unique solution $g_h \in \mathrm{L}^1([0\,,1])$ to the following integral equation:

$$\int_0^1 g_h(x)\,\phi(x)\,\mathrm{d}x = -\int_0^1 h(x)\,\phi'(x)\,\mathrm{d}x \quad \text{with} \quad \phi \in \mathrm{C}_0^\infty([0\,,1]). \tag{3.1}$$

This solution is called the *weak derivative* of h and is often denoted by $D^1 h$. Thus, the use of an appropriate set of smooth test functions allows to define a generalized notion of derivative in $\mathrm{L}^1([0\,,1])$. Further technical details on the mathematical foundations of weak derivatives may be found in [2] (see Sections 1.55–1.62, pp. 19–22). In particular, if the bounded interval $[0\,,1]$ is replaced by a possibly unbounded domain $\Omega \in \mathbb{R}^n$ (for $n \geq 1$), everything said above remains valid, provided you replace $\mathrm{L}^1([0\,,1])$ by the space $\mathrm{L}^1_{\mathrm{loc}}(\Omega)$ of all locally integrable functions on Ω.

Since $D^1 h \in \mathrm{L}^1([0\,,1])$ by construction, the process can be iterated to define $D^2 h$, then $D^3 h$, and finally any weak derivative $D^k h$ of order $k \geq 1$. Knowing that $\mathrm{L}^2([0\,,1]) \subset \mathrm{L}^1([0\,,1])$, the weak differentiation mechanism can therefore be applied infinitely many times in $\mathrm{L}^2([0\,,1])$ and all the resulting weak derivatives belong to $\mathrm{L}^1([0\,,1])$. However, there is no guarantee that these functions also belong to $\mathrm{L}^2([0\,,1])$. For this reason, the subspace of $\mathrm{L}^2([0\,,1])$ where the weak derivatives up to order $r \geq 1$ remain in $\mathrm{L}^2([0\,,1])$ is a very specific subspace of $\mathrm{L}^2([0\,,1])$ called the *Sobolev space* of order r (on $[0\,,1]$ and for the L^2-norm):

$$\mathrm{H}^r([0\,,1]) := \left\{ h \in \mathbb{R}^{[0\,,1]} \;\middle|\; \forall\, 0 \leq k \leq r, \;\; D^k h \in \mathrm{L}^2([0\,,1]) \right\}. \tag{3.2}$$

The integer parameter r directly controls the level of smoothness within $\mathrm{H}^r([0\,,1])$. Indeed, even if not necessarily intuitive, the integrability conditions imposed on the weak derivatives $D^k h$ determine how many times h can be differentiated (in the classic meaning of the word). This point is precisely the subject of the Sobolev embedding theorem [2] (see Theorem 4.12, p. 85) which ensures that $\mathrm{H}^r([0\,,1])$ is continuously embedded in $\mathrm{C}^{r-1}([0\,,1])$ for any $r \geq 1$. This notably shows that the functions contained in the Sobolev space $\mathrm{H}^r([0\,,1])$ are at least continuous. Hence, there is no need to consider the quotient space of $\mathrm{H}^r([0\,,1])$ for the almost-everywhere equality relation.

Now, let us see which inner products are suitable to get the most out of Sobolev spaces. Of course, a naive attempt could be to equip $\mathrm{H}^r([0\,,1])$ with the L^2-inner product but this would unfortunately not produce an RKHS. A more specific inner product must be used instead:

$$\forall\, h_1, h_2 \in \mathrm{H}^r([0\,,1]), \quad \langle h_1, h_2 \rangle_{\mathrm{H}^r} := \sum_{k=0}^{r} \left(\int_0^1 D^k h_1(x)\, D^k h_2(x)\, \mathrm{d}x \right). \tag{3.3}$$

The inner product $\langle \cdot, \cdot \rangle_{\mathrm{H}^r}$ truly accounts for the entire specificity of the functions in $\mathrm{H}^r([0\,,1])$. It is therefore considered as the standard inner product. The resulting induced norm $\|\cdot\|_{\mathrm{H}^r}$ involves the L^2-norms of all the weak derivatives (from $h = D^0 h$ to $D^r h$). If endowed with this norm, $\mathrm{H}^r([0\,,1])$ becomes an RKHS [12] (see Theorem 121, p. 276) with reproducing kernel denoted by K^r. This RKHS (resp. its kernel) is better known as the standard Sobolev RKHS (resp. standard Sobolev kernel) of order r. For $r = 1$, it was demonstrated in [41] (see Chapter IV) that K^1 has a closed-form expression:

$$\forall\, x, x' \in [0\,,1], \quad K^1(x, x') = \frac{2e}{e^2 - 1} \cosh\left[\, \min(x, x')\right] \cosh\left[1 - \max(x, x')\right], \tag{3.4}$$

where $\cosh(\cdot)$ denotes the hyperbolic cosine function. Moreover, it was later shown in [108] (see Corollary 2, p. 27) that K^1 admits a Mercer decomposition based on sinusoidal eigenfunctions (more precisely cosine functions with increasing frequencies) and eigenvalues decaying at a polynomial rate of $1/k^2$ (see Appendix B.3.3). When $r \geq 2$, unlike what happens for $r = 1$, there is no explicit formula for the reproducing kernel K^r of the standard Sobolev RKHS. In the hope of obtaining a kernel with a closed-form expression even for $r \geq 2$, a well-known strategy is to replace $\langle \cdot, \cdot \rangle_{\mathrm{H}^r}$ by another inner product defined on $\mathrm{H}^r([0\,,1])$. Let $\langle \cdot, \cdot \rangle_{\mathcal{H}}$ be one of them. The associated induced norm $\|\cdot\|_{\mathcal{H}}$ is said to be equivalent to $\|\cdot\|_{\mathrm{H}^r}$ if:

$$\exists\, 0 < c_1 < c_2 < \infty \quad \text{such that} \quad \forall\, h \in \mathrm{H}^r([0\,,1]), \quad c_1 \|h\|_{\mathrm{H}^r} \leq \|h\|_{\mathcal{H}} \leq c_2 \|h\|_{\mathrm{H}^r}\,. \tag{3.5}$$

By Definition 2.2, if an inner product $\langle \cdot, \cdot \rangle_{\mathcal{H}}$ indeed leads to an equivalent norm, $\mathrm{H}^r([0\,,1])$ equipped with $\langle \cdot, \cdot \rangle_{\mathcal{H}}$ remains an RKHS. This new RKHS is denoted by $\mathcal{H}$ in order to avoid any confusion with the standard Sobolev RKHS. According to Theorem 2.5, the reproducing kernel $K_{\mathcal{H}}$ (associated to $\mathcal{H}$) is unique and different from the standard kernel K^r. Since K^r and $K_{\mathcal{H}}$ come out from the same Sobolev space, they can be considered as equivalent Sobolev kernels. This terminology is strictly delineated in [79].

Definition 3.1. Let $K : [0\,,1] \times [0\,,1] \to \mathbb{R}$ be a kernel with associated RKHS denoted by $\mathcal{H}$. Then, K is said to be a Sobolev kernel (of order r) if $\|\cdot\|_{\mathcal{H}}$ and $\|\cdot\|_{\mathrm{H}^r}$ are two equivalent norms on $\mathrm{H}^r([0\,,1])$.

The next section focuses on the particular case of the unanchored Sobolev RKHSs. Just as the standard Sobolev RKHSs, they are built from the Sobolev spaces $\mathrm{H}^r([0\,,1])$. The only difference lies in the choice of the inner product used to build the Hilbertian structure.

3.2. Unanchored Sobolev spaces

A simple way to define an alternative inner product on $\mathrm{H}^r([0\,,1])$ is to take:

$$\langle h_1, h_2\rangle_{\mathcal{H}^r_{\mathrm{Sob}}} := \sum_{k=0}^{r-1} \left(\int_0^1 D^k h_1(x)\,\mathrm{d}x\right)\left(\int_0^1 D^k h_2(x)\,\mathrm{d}x\right) + \int_0^1 D^r h_1(x)\, D^r h_2(x)\,\mathrm{d}x\ . \tag{3.6}$$

The induced norm $\|\cdot\|_{\mathcal{H}^r_{\mathrm{Sob}}}$ is equivalent to the standard Sobolev norm $\|\cdot\|_{\mathrm{H}^r}$. The two inequalities in Eq. (3.5) are indeed satisfied by $\|\cdot\|_{\mathcal{H}^r_{\mathrm{Sob}}}$. The right-hand side is a direct consequence of the Cauchy-Schwarz inequality (when it is applied in $\|h\|^2_{\mathcal{H}^r_{\mathrm{Sob}}}$ to majorize the mean values of the derivatives). As regards the left-hand side, it results from a recursive use of the Poincaré-Wirtinger inequality [19] (see Comment 3.A on Chapter 9, p. 312). In particular, this inequality allows to derive successive upper bounds of $\|h\|^2_{\mathrm{H}^r}$ which are gradually freed from the L^2-norms of the low-order derivatives.

Definition 3.2. The RKHS obtained when $\mathrm{H}^r([0\,,1])$ is endowed with the inner product defined in Eq. (3.6) is called the *unanchored* Sobolev space (of order r) and denoted by $\mathcal{H}^r_{\mathrm{Sob}}$.

Contrary to what was deplored for the reproducing kernel K^r of the standard Sobolev space, the reproducing kernel K^r_{Sob} of the unanchored Sobolev space has an easily computable expression at all orders.

Theorem 3.3. *For any $r \geq 1$, the reproducing kernel of the unanchored Sobolev space $\mathcal{H}^r_{\mathrm{Sob}}$ is given by:*

$$\forall\, x, x' \in [0\,,1], \quad K^r_{\mathrm{Sob}}(x,x') := 1 + k^r_{\mathrm{Sob}}(x,x') = \sum_{k=0}^{r} \frac{B_k(x)\, B_k(x')}{(k!)^2} + \frac{(-1)^{r+1}}{(2r)!}\, B_{2r}(|x-x'|)\ , \tag{3.7}$$

where $(B_k)_{k\geq 0}$ denotes the sequence of Bernoulli polynomials.

For the detailed proof, the interested reader is invited to consult [56] (see Section 2.3.3, p. 35–38). Unlike the scattered information found in older works [28, 112], this proof is self-contained and easy to read.

Remark 3.4. The kernels $(K^r_{\mathrm{Sob}})_{r\geq 1}$ are all ANOVA. This can be checked very easily with the properties of Bernoulli polynomials (recalled in Appendix A.1), especially the zero-mean property (see Appendix A.1.6) and the symmetry properties (see Appendix A.1.4).

3.3. Properties of Sobolev kernels

Several general results regarding Sobolev kernels (in the meaning of Definition 3.1) are now highlighted because they are essential to understand the rest of this work.

Proposition 3.5. *All Sobolev kernels are Mercer kernels.*

Proof. Let K be a Sobolev kernel (of order $r \geq 1$). The Sobolev embedding theorem yields $\mathrm{H}^r([0\,,1]) \subset \mathrm{C}([0\,,1])$. Then, Lemma 4.28 in [25] allows to justify that K is continuous. □

Proposition 3.6. *All Sobolev kernels are characteristic to $\mathcal{M}^+_1([0\,,1])$.*

Proof. Let K be a Sobolev kernel (of order $r \geq 1$). According to Proposition 3.5 and Remark 2.16, it is sufficient to prove that K is universal. In other words, the only point to justify is the uniform denseness of $\mathrm{H}^r([0\,,1])$ in $\mathrm{C}([0\,,1])$. With Eq. (3.2), it is obvious that $\mathrm{H}^r([0\,,1])$ contains all the polynomial functions. According to the Stone-Weierstrass theorem, the algebra of polynomials is uniformly dense in $\mathrm{C}([0\,,1])$. Hence, the same is true for $\mathrm{H}^r([0\,,1])$. Therefore, K is universal, and thus characteristic to $\mathcal{M}_1^+([0\,,1])$. □

Proposition 3.7. *A Sobolev kernel is not always ANOVA.*

Proof. A counterexample is provided by the kernel $K^1_{\text{anch}}(x,x') = 1 + k^1_{\text{anch}}(x,x')$ where $k^1_{\text{anch}}(x,x') := \min(x,x')$. K^1_{anch} is the reproducing kernel of the RKHS (denoted by $\mathcal{H}^1_{\text{anch}}$) obtained when $\mathrm{H}^1([0\,,1])$ is equipped with the inner product $\langle h_1, h_2 \rangle_{\mathcal{H}^1_{\text{anch}}} := h_1(0)\, h_2(0) + \langle\, h_1', h_2' \,\rangle_{\mathrm{L}^2}$ [12] (see Example 23, p. 322). This RKHS is sometimes called the Sobolev space (of order 1) *anchored* at 0 [55, 60, 70]. This explains why the subscript "anch" is used here to denote the RKHS and its reproducing kernel. Coming back to the proof, the kernel k^1_{anch} cannot be orthogonal since it is positive almost everywhere on $[0\,,1]^2$. Therefore, K^1_{anch} is not an ANOVA kernel. □

Theorem 3.8. *If K is a Sobolev kernel (of order $r \geq 1$), then the eigenvalues of the kernel integral operator $T_K : \mathrm{L}^2([0\,,1]) \to \mathrm{L}^2([0\,,1])$ verify $\lambda_k = \mathcal{O}(1/k^{2r})$.*

This result is mentioned in many recent papers [6, 115, 119] but the original proof dates back to much older works [13, 14] where more general conclusions on Sobolev spaces are enunciated.

Several illustrations of Theorem 3.8 can be found in Appendix B.3. Polynomial eigendecays are one major singularity of Sobolev kernels. For comparison, Gaussian kernels are characterized by exponential eigendecays and thus correspond to much smaller RKHSs.

Now that all the necessary mathematical concepts have been clearly introduced, it is time to explain what kernels are used for in sensitivity analysis.

4. Kernel-based global sensitivity analysis

The main ambition of this section is to provide a brief overview of kernel methods in global sensitivity analysis (GSA) and to explain why the feature maps of Sobolev kernels provide valuable insights into the newly developed HSIC-ANOVA framework. First, Section 4.1 deals with HSIC indices and highlights their ability to characterize independence. Then, Section 4.2 explains why the (Mercer) feature maps of the input and output kernels help identify the (most important) dependence patterns captured by HSIC indices. Finally, Section 4.3 focuses on the HSIC-ANOVA decomposition. The reasons for the use of Sobolev kernels are detailed and the questions raised by this choice are put forward.

4.1. Sensitivity measures based on the Hilbert-Schmidt independence criterion

Let us take the usual notations in the field of uncertainty quantification. In this context, an output of interest Y is computed by a numerical simulator $g : \mathcal{X}_1 \times \ldots \times \mathcal{X}_d \to \mathcal{Y}$ which is given a set of possibly dependent random variables $X_1, \ldots, X_d$ gathered in a random vector $\boldsymbol{X}$. In terms of modeling, the simulator is handled as a deterministic black-box function. Each random input X_i takes its values in $\mathcal{X}_i \subseteq \mathbb{R}$ and follows a probability distribution $\mathbb{P}_{X_i} \in \mathcal{M}_1^+(\mathcal{X}_i)$. In addition, a continuous kernel $K_i : \mathcal{X}_i \times \mathcal{X}_i \to \mathbb{R}$ (with RKHS denoted by $\mathcal{H}_i$) is assigned to X_i. Similar mathematical objects and notations are adopted for the output variable Y.

4.1.1. *Definition and estimation of HSIC indices*

As recalled in [18, 29], many existing sensitivity measures seek to quantify the discrepancy between the joint input-output distribution $\mathbb{P}_{X_i Y}$ and the product of marginal distributions $\mathbb{P}_{X_i} \otimes \mathbb{P}_Y$ (representing hypothetical independence between X_i and Y). The key step is thus the choice of a dissimilarity measure on $\mathcal{M}_1^+(\mathcal{X}_i \times \mathcal{Y})$. For example, taking the total-variation distance leads to Borgonovo's indices [15]. Despite real advantages, this method involves a delicate preliminary step in which all input-output densities need to be estimated from the available data [33]. To avoid this, a kernel strategy consists in using the MMD (see Section 2.2.3) related to the tensor product kernel $K_i \otimes K_Y$. The discrepancy between $\mathbb{P}_{X_i Y}$ and $\mathbb{P}_{X_i} \otimes \mathbb{P}_Y$ is thus measured through the distance of their respective images in the tensor product RKHS $\mathcal{H}_i \otimes \mathcal{H}_Y$. This kernel-based approach amounts to applying the *Hilbert-Schmidt independence criterion* (HSIC) to all input-output pairs:

$$\forall\, 1 \leq i \leq d, \quad \mathrm{HSIC}(X_i, Y) := \mathrm{MMD}^2(\mathbb{P}_{X_i Y}, \mathbb{P}_{X_i} \otimes \mathbb{P}_Y) = \|\mu_{\mathbb{P}_{X_i Y}} - \mu_{\mathbb{P}_{X_i} \otimes \mathbb{P}_Y}\|^2_{\mathcal{H}_i \otimes \mathcal{H}_Y} \ . \tag{4.1}$$

The resulting sensitivity measures are called HSIC indices. Although they were initially intended for variable selection in machine learning [52, 54], they have been increasingly used over the past few years in GSA [29, 74]. The alternative formula of the MMD given in Eq. (2.5) allows to rewrite each index $\mathrm{HSIC}(X_i, Y)$ as a sum of three expectations:

$$\mathrm{HSIC}(X_i, Y) = \mathbb{E}\left[K_i(X_i, X_i')\, K_Y(Y, Y')\right] + \mathbb{E}\left[K_i(X_i, X_i')\right]\, \mathbb{E}\left[K_Y(Y, Y')\right] - 2\, \mathbb{E}\left[K_i(X_i, X_i')\, K_Y(Y, Y'')\right], \tag{4.2}$$

where (X_i, Y), (X_i', Y') and (X_i'', Y'') are three random pairs following the joint input-output distribution $\mathbb{P}_{X_i Y}$ while being independent of each other. Based on Eq. (4.2), the quantity $\mathrm{HSIC}(X_i, Y)$ can be expressed as a single expectation (see Appendix C.2.1), and more precisely as the expectation of a symmetric function involving four independent copies of the input-output vector $\boldsymbol{Z} := (X_i, Y)$. This paves the way to an estimator of $\mathrm{HSIC}(X_i, Y)$ in the form of a U-statistic or a V-statistic [53, 98]. Further technical details related to the construction of $\widehat{H}_i^U$ (U-statistic estimator) and $\widehat{H}_i^V$ (V-statistic estimator) are postponed to Appendix C.2.2. In the following, when a result is valid for both $\widehat{H}_i^U$ and $\widehat{H}_i^V$, no distinction will be made between them and the HSIC estimator will be merely denoted by $\widehat{H}_i$.

From a theoretical perspective, an important point regarding the efficiency of inference is the existence of a central limit theorem (CLT) that governs the asymptotic behavior of $\widehat{H}_i$ [53] (see Theorem 1, p. 4). This notably shows that $\widehat{H}_i$ converges to the exact value $\mathrm{HSIC}(X_i, Y)$ at the rate of $1/n$. A similar convergence rate can also be achieved for most Sobol' index estimators, as highlighted by the CLTs established in [64] (see Sections 3.1 and 4.2, pp. 345–348) and more recently in [47] (see Section 4.2, p. 2351). Although the convergence speed of an HSIC estimator (taken alone) seems comparable to that of any Sobol' index estimator, it is actually much easier to estimate all HSIC indices than all Sobol' indices. Indeed, the entire collection of HSIC estimates can be directly computed from a single Monte Carlo design of size n, whereas the computation of all Sobol' estimates asks for a Pick-Freeze design composed of at least $n\,(d+1)$ samples [82]. Hence, HSIC indices are particularly attractive to tackle large-scale problems because the number of input-output evaluations required to reach sufficient accuracy (on the entire collection of sensitivity indices) does not increase with the dimensionality.

4.1.2. *Characterization of independence with HSIC indices*

From Eq. (4.1), it is clear that one has $\mathrm{HSIC}(X_i, Y) = 0$ in case of independence between X_i and Y. The converse is false in general but becomes true if both K_i and K_Y are sufficiently sophisticated.

Proposition 4.1. *Let $K_i : \mathcal{X}_i \times \mathcal{X}_i \to \mathbb{R}$ and $K_Y : \mathcal{Y} \times \mathcal{Y} \to \mathbb{R}$ be two kernels such that K_i is characteristic to $\mathcal{M}_1^+(\mathcal{X}_i)$ and K_Y is characteristic to $\mathcal{M}_1^+(\mathcal{Y})$. Then, one has:*

$$X_i \perp Y \iff \mathrm{HSIC}(X_i, Y) = 0\ .$$

Proof. It is straightforward in view of the definitions given so far. As already mentioned in Remark 2.10, the tensor product kernel $K_i \otimes K_Y$ is characteristic to $\mathcal{M}_1^+(\mathcal{X}_i \times \mathcal{Y})$. The map $\nu \in \mathcal{M}_1^+(\mathcal{X}_i \times \mathcal{Y}) \mapsto \mu_\nu \in \mathcal{H}_1 \otimes \mathcal{H}_Y$ is therefore injective and this ends the proof. $\square$

Proposition 4.1 is fundamental in the prospect of building a test of independence. Indeed, a plethora of test procedures have been developed on the basis of this property [3, 26, 32, 43, 44]. Generally speaking, a test of independence is a statistical procedure intended to make a choice between:

$$(H_0) : X_i \perp Y \qquad \text{vs.} \quad (H_1) : X_i \text{ and } Y \text{ are dependent variables.}$$

If K_i and K_Y are two characteristic kernels, taking the HSIC allows to transform the initial problem of hypothesis test into a simpler one:

$$(H_0) : \mathrm{HSIC}(X_i, Y) = 0 \qquad \text{vs.} \quad (H_1) : \mathrm{HSIC}(X_i, Y) > 0.$$

From here, any estimator of $\mathrm{HSIC}(X_i, Y)$ with known distribution under (H_0) is a suitable test statistic. In particular, the estimators $\widehat{H}_i^U$ and $\widehat{H}_i^V$ (see Appendix C.2.2) are two typical examples. Their asymptotic distributions under (H_0) were first derived in [53] (see Theorem 2, p. 4) and several refinements have been proposed since then [117, 118]. Naturally, the acceptance (resp. rejection) region of (H_0) is composed of the smallest (resp. largest) values taken by the test statistic $\widehat{H}_i$. When the sample size is small, asymptotic test procedures are no longer valid and non-parametric test procedures based on random permutations must be used instead [32, 53].

4.2. **Connections between feature maps and HSIC indices**

Beyond their original mathematical construction based on kernel mean embeddings, HSIC indices can also be rewritten as generalized covariance operators [9, 109]. In fact, it is proved in [54] that $\mathrm{HSIC}(X_i, Y) = \|C_{X_iY}\|_{\mathrm{HS}}^2$ where the cross-covariance operator $C_{X_iY} : \mathcal{H}_Y \to \mathcal{H}_i$ is defined by:

$$\forall\, h_i \in \mathcal{H}_i, \quad \forall\, h_Y \in \mathcal{H}_Y, \quad \langle C_{X_iY}\, h_Y, h_i\rangle_{\mathcal{H}_i} = \mathrm{Cov}(h_i(X_i), h_Y(Y)) \ .$$

After replacing $\|\cdot\|_{\mathrm{HS}}$ by its definition, one has:

$$\forall\, 1 \le i \le d, \quad \mathrm{HSIC}(X_i, Y) = \sum_k \sum_l \left|\mathrm{Cov}\left(u_{ik}(X_i)\,, v_l(Y)\right)\right|^2 \quad \text{with} \quad \begin{cases} (u_{ik})_k & \text{an ONB of } \mathcal{H}_i \ , \\ (v_l)_l & \text{an ONB of } \mathcal{H}_Y \ . \end{cases} \tag{4.3}$$

With the above reformulation, $\mathrm{HSIC}(X_i, Y)$ appears to be an aggregation of covariance terms obtained after scanning all basis directions in the two RKHSs. Quantifying with $\mathrm{Cov}(\cdot, \cdot)$ the linear dependence between $u_{ik}(X_i)$ and $v_l(Y)$ amounts to quantifying a part of the non-linear dependence between X_i and Y, more precisely the non-linear dependence pattern characterized by the pair of preliminary transformations (u_{ik}, v_l).

Remark 4.2. In Eq. (4.3), the indexation sets associated to k and l are deliberately not specified. This is done to emphasize the fact that indexing will depend on the selected kernels. Indeed, if a kernel as finite (resp. infinite) rank, the induced RKHS has finite (resp. infinite) dimension, and the associated ONB is composed of a finite (resp. infinite) number of basis functions.

Eq. (4.3) reveals that the identification of the dependence patterns captured by the HSIC (when built with K_i and K_Y) simply requires the knowledge of ONBs of $\mathcal{H}_i$ and $\mathcal{H}_Y$. For a given RKHS, it has already been explained in Section 2.4 that an ONB can be directly obtained from the kernel expression by finding a Mercer expansion (see Theorem 2.21) or any other series expansion based on symmetric and separable functions (see Theorem 2.23). Depending on the kernel, this may be more or less difficult.

Example 4.3. Let us imagine that $\mathrm{HSIC}(X_i, Y)$ is computed with two Gaussian kernels (with scale parameters respectively denoted by γ_i and γ_Y). This is the most common situation in practice. A series expansion of the Gaussian kernel K_G was provided in Example 2.12 to reveal the hidden existence (within K_G) of infinitely many damped polynomial features. Theorem 2.23 can then be used to justify that the features exactly form an ONB of the RKHS (induced by K_G). With this in mind, the generalized covariance defined in Eq. (4.3) becomes:

$$\mathrm{HSIC}(X_i, Y) = \sum_{k=0}^{\infty} \sum_{l=0}^{\infty} \left|\mathrm{Cov}\left(u_{ik}(X_i)\,, v_l(Y)\right)\right|^2 \quad \text{with} \quad \begin{cases} \forall\, k \ge 0, \quad u_{ik}(x_i) := \dfrac{1}{\sqrt{k!}} \exp\left[-\dfrac{1}{2}\left(\dfrac{x_i}{\gamma_i}\right)^2\right] \left(\dfrac{x_i}{\gamma_i}\right)^k, \\[2ex] \forall\, l \ge 0, \quad v_l(y) := \dfrac{1}{\sqrt{l!}} \exp\left[-\dfrac{1}{2}\left(\dfrac{y}{\gamma_Y}\right)^2\right] \left(\dfrac{y}{\gamma_Y}\right)^l. \end{cases}$$

Now, let $(\overline{u_{ik}})_k$ and $(\overline{v_l})_l$ be the functions obtained after an L^2-normalization (with respect to $\mathbb{P}_{X_i}$ and $\mathbb{P}_Y$) of the basis functions extracted from $\mathcal{H}_i$ and $\mathcal{H}_Y$. With these new notations, Eq. (4.3) becomes:

$$\mathrm{HSIC}(X_i, Y) = \sum_k \sum_l \|u_{ik}\|^2_{\mathrm{L}^2} \, \|v_l\|^2_{\mathrm{L}^2} \left|\mathrm{Cov}\left(\overline{u_{ik}}(X_i)\,, \overline{v_l}(Y)\right)\right|^2 , \tag{4.4}$$

and it is now possible to find a uniform upper bound for all covariance terms:

$$0 \leq \left|\mathrm{Cov}\left(\overline{u_{ik}}(X_i)\,, \overline{v_l}(Y)\right)\right|^2 \leq \mathbb{V}\Big(\overline{u_{ik}}(X_i)\Big)\,\mathbb{V}\Big(\overline{v_l}(Y)\Big) = \Big(1 - \mathbb{E}_{\mathbb{P}_{X_i}}\Big[\overline{u_{ik}}(X_i)\Big]^2\Big)\,\Big(1 - \mathbb{E}_{\mathbb{P}_Y}\Big[\overline{v_l}(Y)\Big]^2\Big) \leq 1 \; .$$

Hence, each covariance term in Eq. (4.4) lies in $[0\,, 1]$ and it is weighted by a coefficient equal to $\|u_{ik}\|^2_{\mathrm{L}^2}\|v_l\|^2_{\mathrm{L}^2}$. This means that the HSIC captures many different dependence patterns but they are not equally weighted and the weighting system depends on the encountered distributions $\mathbb{P}_{X_i}$ and $\mathbb{P}_Y$.

- For each input variable X_i, the L^2-norms of the basis functions $(u_{ik})_k$ can be calculated by hand or approximated by numerical integration because $\mathbb{P}_{X_i}$ is provided.
- For the output variable Y, since $\mathbb{P}_Y$ is unknown, the L^2-norms of the basis functions $(v_l)_l$ can only be estimated from the available output samples.

It can be easily proved that the two sequences $(\|u_{ik}\|_{\mathrm{L}^2})_k$ and $(\|v_l\|_{\mathrm{L}^2})_l$ are square summable and thus vanish at infinity. As a result, only a small number of dependence patterns actually count towards the final value of $\mathrm{HSIC}(X_i, Y)$. Furthermore, an enhanced (if not optimal) version of Eq. (4.4) can be achieved if K_i and K_Y have explicit Mercer decompositions (for $\mathbb{P}_{X_i}$ and $\mathbb{P}_Y$ respectively):

$$\begin{aligned} &\forall\, x_i, x_i' \in \mathcal{X}_i, \quad K_i(x_i, x_i') = \sum_k \lambda_k\, \phi_{ik}(x_i)\, \phi_{ik}(x_i') \quad \text{with} \quad \lambda_{i1} \geq \lambda_{i2} \geq \ldots > 0 \;, \\ &\forall\, y, y' \in \mathcal{Y}, \qquad K_Y(y, y') = \sum_l \mu_l\, \psi_l(y)\, \psi_l(y') \qquad \text{with} \quad \mu_1 \geq \mu_2 \geq \ldots > 0 \;. \end{aligned}$$

Theorem 2.21 allows to rewrite Eq. (4.3) with the eigenvalues and eigenfunctions of T_{K_i} and T_{K_Y}:

$$\mathrm{HSIC}(X_i, Y) = \sum_k \sum_l \lambda_{ik}\, \mu_l \left|\mathrm{Cov}\left(\phi_{ik}(X_i)\,, \psi_l(Y)\right)\right|^2 . \tag{4.5}$$

Since the eigenfunctions are already L^2-normalized, Eq. (4.5) is already of the same form as Eq. (4.4) and no renormalization effort is necessary. Therefore, each coefficient $\lambda_{ik}\,\mu_l$ must be interpreted as the weight assigned to the dependence pattern characterized by the pair (ϕ_{ik}, ψ_l). Just as the L^2-norms in Eq. (4.4), the two sequences of eigenvalues $(\lambda_{ik})_k$ and $(\mu_l)_l$ are square summable and vanish at infinity. They even decrease faster since the eigenfunctions $(\phi_{ik})_k$ and $(\psi_l)_l$ are orthogonal in the L^2-sense, contrary to the features $(u_{ik})_k$ and $(v_l)_l$ in Eq. (4.4). In short, the Mercer feature map of K_i (resp. K_Y) with respect to $\mathbb{P}_{X_i}$ (resp. $\mathbb{P}_Y$) provides the most relevant description of how X_i (resp. Y) is transformed when $\mathrm{HSIC}(X_i, Y)$ is computed. In particular, Mercer features have two major advantages over any other collection of features:

- They are naturally ranked by increasing order of influence.
- Their L^2-norms have a faster decay rate, which results in a sparser representation of $\mathrm{HSIC}(X_i, Y)$.

4.3. **An ANOVA framework for HSIC indices**

Despite their many advantages (low estimation cost, characterization of independence, almost no limiting assumption on the input and output distributions), HSIC indices also suffer from major shortcomings. Above all, as the sum of all HSIC indices is not equal to 1, they cannot be interpreted as percentages of influence. Moreover, in the absence of a universal bound for HSIC indices, it is difficult to know what is meant by small and large values. To remedy this difficulty, the ANOVA framework (until now reserved for Sobol' indices) has been freshly extended to kernel-based sensitivity measures [10, 11, 30]. More specifically, an ANOVA-like decomposition has been set up for HSIC indices in [30] and it has immediately aroused much interest in the GSA community [45, 78, 85]. However, this breakthrough was obtained at the cost of stronger assumptions (on both the input kernels and the input probability distributions).

Theorem 4.4. *It is assumed that:*

(A1) The input variables $X_1, \ldots, X_d$ are mutually independent.

(A2) There is an ANOVA kernel $K_i = 1 + k_i$ (with RKHS $\mathcal{H}_i = \mathbb{R} \oplus \mathcal{F}_i$) for each marginal distribution $\mathbb{P}_{X_i}$. A multivariate ANOVA kernel can then be constructed for each subset of input variables:

$$\forall\, \boldsymbol{u} := \{u_1, \ldots, u_s\} \subseteq \{1, \ldots, d\}, \quad K_{\boldsymbol{u}} := K_{u_1} \otimes K_{u_2} \otimes \ldots \otimes K_{u_s} \ .$$

(A3) The joint input distribution $\mathbb{P}_{\boldsymbol{X}}$ and the output distribution $\mathbb{P}_Y$ are such that:

$$\forall\, \boldsymbol{u} \subseteq \{1, \ldots, d\}, \quad \mathbb{E}_{\mathbb{P}_{\boldsymbol{X_u}}}[K_{\boldsymbol{u}}(\boldsymbol{X_u}, \boldsymbol{X_u})] < \infty \quad \textit{and} \quad \mathbb{E}_{\mathbb{P}_Y}[K_Y(Y, Y)] < \infty \ .$$

An HSIC index can then be computed for each subset of input variables:

$$\forall\, \boldsymbol{u} \subseteq \{1, \ldots, d\}, \quad \mathrm{HSIC}(\boldsymbol{X_u}, Y) := \mathrm{MMD}^2(\mathbb{P}_{\boldsymbol{X_u}Y}, \mathbb{P}_{\boldsymbol{X_u}} \otimes \mathbb{P}_Y) \ .$$

The HSIC-ANOVA decomposition is then given by:

$$\mathrm{HSIC}(\boldsymbol{X}, Y) = \sum_{\boldsymbol{u} \subseteq \{1, \ldots, d\}} \mathrm{HSIC}_{\boldsymbol{u}} = \sum_{\boldsymbol{u} \subseteq \{1, \ldots, d\}} \sum_{\boldsymbol{v} \subseteq \boldsymbol{u}} (-1)^{|\boldsymbol{u}| - |\boldsymbol{v}|} \mathrm{HSIC}(\boldsymbol{X_v}, Y) \ . \tag{4.6}$$

The proof may be found in [30] (see Appendix A.3, pp. 33–37). It notably relies on earlier works dealing with ANOVA decompositions of multivariate functions in well-adapted RKHSs [70, 114]. The assumption (A1) may seem too restrictive but it is in fact similar to what is demanded for Sobol' indices, at least in their most usual setting[9]. Theorem 4.4 provides a rigorous decomposition of the quantity $\mathrm{HSIC}(\boldsymbol{X}, Y)$ into the sum of $2^d - 1$ HSIC terms, one per each subset $\boldsymbol{X_u}$ of input variables. If taking $\boldsymbol{u} := \{i\}$ in Eq. (4.6), one has $\mathrm{HSIC}_i = \mathrm{HSIC}(X_i, Y)$. This suggests to renormalize HSIC indices in the following way:

$$\forall\, 1 \leq i \leq d, \quad S_i^{\mathrm{HSIC}} := \frac{\mathrm{HSIC}(X_i, Y)}{\mathrm{HSIC}(\boldsymbol{X}, Y)} \ . \tag{4.7}$$

The resulting collection of sensitivity indices are called the first-order HSIC-ANOVA indices [31]. Indices of higher order can also be defined in the same spirit. In particular, the total-order HSIC-ANOVA indices are studied in [85]. Unlike the R^2-HSIC indices (see Appendix C.1.2), introduced much earlier [29] and resulting from a different renormalization technique, HSIC-ANOVA indices can be regarded as percentage-like importance measures, especially because their sum is equal to 1.

Remark 4.5. Let us assume that the input and output kernels verify (A2) and (A3). In addition, if all kernels are characteristic, HSIC indices are able to characterize independence, and the same is true for the first-order HSIC-ANOVA indices S_i^{HSIC} which are proportional to them.

[9]The original definition of Sobol' indices [96, 97] is based on the Sobol'-Hoeffding decomposition of the input-output numerical simulator $g : \mathcal{X}_1 \times \ldots \times \mathcal{X}_d \to \mathcal{Y}$. Since the uniqueness of this decomposition is only guaranteed in the presence of mutually independent inputs, Sobol' indices are well defined only under this assumption. Over the past fifteen years, there have been several attempts to extend Sobol' indices to correlated inputs [24, 73, 80] but none of them has proved sufficiently satisfactory.

From a theoretical viewpoint, the HSIC-ANOVA methodology successfully combines the advantages of the HSIC and ANOVA frameworks. Now comes the question of its implementation in practice. To comply with the assumptions of Theorem 4.4 and Remark 4.5, the output kernel K_Y must be characteristic and a characteristic ANOVA kernel $K_i = 1 + k_i$ must be found for each input probability distribution $\mathbb{P}_{X_i} \in \mathcal{M}_1^+(\mathcal{X}_i)$. As already evoked in Section 2.2.1, there is no characteristic ANOVA kernel for most parametric families, except for the standard uniform distribution. A well-known trick is to replace the initial variable X_i by $U_i := F_{X_i}(X_i)$ where F_{X_i} denotes the cumulative distribution function of $\mathbb{P}_{X_i}$. Of course, the mathematical modeling of the numerical simulator is adapted accordingly and becomes $g : [0\,,1]^d \to \mathcal{Y}$. As the problem can always be reformulated in this way, $X_1, \ldots, X_d$ are now assumed to follow standard uniform distributions. Then, among all possible kernel choices, the practice promoted in [30] is to use Sobolev kernels, and more precisely the unanchored Sobolev kernels $(K^r_{\text{Sob}})_{r\geq 1}$. They are several reasons for this choice:

- They are ANOVA (see Remark 3.4) and characteristic (see Proposition 3.6).
- They have a simple analytical formula (unlike the orthogonalized kernels popularized in [42, 49]).
- No parameter tuning is necessary and the level of smoothness in $\mathcal{H}^r_{\text{Sob}}$ can be adjusted with the integer parameter $r \geq 1$.

Up to now, HSIC indices have been mainly estimated with kernels drawn from Gaussian process regression (Gaussian kernels, Laplace kernels, Matérn kernels). To the best of our knowledge, the initiative of computing HSIC indices with Sobolev kernels is one major peculiarity of the HSIC-ANOVA framework. For a complete understanding of the dependence patterns captured by the first-order HSIC-ANOVA indices, one must return to the cross-covariance viewpoint. In particular, Eq. (4.3) becomes:

$$S_i^{\text{HSIC}} \propto \text{HSIC}(X_i, Y) = \sum_k \sum_l \left|\text{Cov}\left(u_{ik}(X_i)\,, v_l(Y)\right)\right|^2 \quad \text{with} \quad \begin{cases} (u_{ik})_k & \text{an ONB of } \mathcal{H}^r_{\text{Sob}}\ , \\ (v_l)_l & \text{an ONB of } \mathcal{H}_Y\ . \end{cases} \tag{4.8}$$

If K_Y is the Gaussian kernel, an ONB of $\mathcal{H}_Y$ has already been provided in Example 4.3. The identification of an ONB of $\mathcal{H}^r_{\text{Sob}}$ seems to be the only remaining technical deadlock. According to Theorems 2.21 and 2.23, a solution may be found through the extraction of a feature map from K^r_{Sob}. This objective is pursued from Section 5 to Section 8. Of all possible feature maps, Mercer's is the most instructive because it provides the most synthetic description of the role played by K^r_{Sob} in the computation of S_i^{HSIC}. That is why Sections 5 and 6 only focus on this specific feature map. In Section 5, numerical experiments are first conducted and this surprisingly helps respond many theoretical answers.

Since the rest of this work is dedicated to the study of the unanchored Sobolev kernels, a simplified terminology is adopted. *Sobolev kernels* will only refer to the kernels $(K^r_{\text{Sob}})_{r\geq 1}$ even if Definition 3.1 is more general. Likewise, *Sobolev spaces* will only refer to the unanchored Sobolev spaces $(\mathcal{H}^r_{\text{Sob}})_{r\geq 1}$.

5. Numerical extraction of the Mercer feature maps

This section aims at using a numerical method called kernel feature analysis (KFA) to get a first idea of what the Mercer expansion of Sobolev kernels may be like. Since it is precised in Theorem 3.8 that the decay speed of the eigenvalues is always polynomial for Sobolev kernels, the only unknown here is the mathematical nature of the eigenfunctions. In Section 5.1, KFA is described in the general context of a Mercer kernel. In Section 5.2, it is applied to Sobolev kernels and valuable conclusions are drawn from this numerical study.

5.1. **Key principles of KFA**

Let $K : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ be a Mercer kernel, let ν be a probability measure with support $\mathcal{X}$ and let T_K be the integral operator built from K and ν. By Theorem 2.18, K admits an eigendecomposition based on a sequence of L^2-normalized eigenfunctions with positive eigenvalues. The general idea of KFA (originally called the *Nyström method* [99, 100]) is to estimate this eigendecomposition by solving a discretized version of the eigenvalue problem. Unlike the Galerkin method [69, 93] that relies on a deterministic discretization of the eigenvalue equation $T_K\,\phi = \lambda\,\phi$, KFA is a simulation-based approach. Only the key points of the method are outlined in what follows. For a step-by-step explanation, the reader is referred to [72] or to Appendix D.

Let $f \in \mathrm{L}^2(\mathcal{X}, \nu)$ be an eigenfunction of T_K associated to a positive eigenvalue λ. For now, no constraint is imposed on the L^2-norm of f. It will be seen later how to recover a unit-norm function (as demanded by Theorem 2.18). The starting point to estimate both f and λ is to draw a n-sample $\boldsymbol{x}_{\mathrm{sim}} := (x_i)_{1\le i\le n}$ from ν. In each pointwise equality $[T_K\,f](x_i) = \lambda\,f(x_i)$, the left-hand term is then replaced by its Monte Carlo approximation. The resulting system of n equations then includes $n+1$ unknowns (namely λ and the n values of f). The Gram matrix $\boldsymbol{K}_n$ (built from $\boldsymbol{x}_{\mathrm{sim}}$) helps rewrite the system as a matrix equation:

$$\boldsymbol{K}_n\,\boldsymbol{v} = (n\,\lambda)\,\boldsymbol{v} \quad \text{with} \quad \boldsymbol{K}_n := [K(x_i, x_j)]_{1\le i,j\le n} \in \mathbb{R}^{n\times n} \quad \text{and} \quad \boldsymbol{v} := [f(x_i)]_{1\le i\le n} \in \mathbb{R}^n\,. \tag{5.1}$$

The eigenvalue problem related to T_K is thus transformed into a matrix eigenvalue problem which can be solved numerically (as long as n is not too large). Let $(\gamma_k)_{1\le k\le n}$ denote the eigenvalues of $\boldsymbol{K}_n$ (with $\gamma_1 \ge \ldots \ge \gamma_n \ge 0$) and let $(\boldsymbol{v}_k)_{1\le k\le n}$ denote the corresponding eigenvectors. In light of Eq. (5.1), the k-th largest eigenvalue of T_K can be estimated by $\hat{\lambda}_k = \gamma_k/n$. In addition, the eigenvector $\boldsymbol{v}_k$ admits the latent structure $[f_k(x_i)]_{1\le i\le n}$ where f_k is an eigenfunction of T_K associated to λ_k (the k-th largest eigenvalue). For this reason, $\boldsymbol{v}_k$ must be understood as a discrete estimate of f_k. As standard eigenvalue algorithms are designed to return unit eigenvectors in $\mathbb{R}^n$, one has $\|f_k\|_{\mathrm{L}^2}\approx 1/\sqrt{n}$ in practice. To bypass this problem, $\boldsymbol{v}_k$ is replaced by $\boldsymbol{w}_k := \sqrt{n}\,\boldsymbol{v}_k$ which is now a discrete estimate of the L^2-normalized eigenfunction $\phi_k := f_k/\|f_k\|_{\mathrm{L}^2}$. Then, for extrapolation purposes, it can be useful to construct an estimate of ϕ_k that can be evaluated everywhere on the domain $\mathcal{X}$. An interpolation method called the *Nyström extension* [46] can be used to build $\widehat{\phi}_k : \mathcal{X} \to \mathbb{R}$ from the only knowledge of γ_k and $\boldsymbol{v}_k$ (see Appendix D).

From a practical viewpoint, plotting the histogram of the eigenvalues $(\hat{\lambda}_k)_{1\le k\le n}$ allows to visualize how fast the eigenvalues decrease. In particular, if the eigendecay is very fast, the Mercer decomposition can be truncated after a few terms and the corresponding eigenfunctions stand for the most influential features. The eigenfunction estimates $(\widehat{\phi}_k)_{1\le k\le n}$ are also worth plotting. They may help understand what the true eigenfunctions look like, or at least identify some of their properties (monotonicity, periodicity, upper and lower bounds). Sometimes the displayed curves have such distinctive shapes that one can easily guess the analytical expression of the eigenfunctions (see Figures 1, 3 and 4). The purpose of the next section is to apply KFA to Sobolev kernels in order to extract relevant information about their eigenvalues and eigenfunctions, and from there to deduce information about their Mercer feature maps.

5.2. **Application to Sobolev kernels**

The procedure detailed in Section 5.1 is now applied to Sobolev kernels. This means that ν is the uniform distribution, $\mathcal{X} = [0\,,1]$ and $K = k^r_{\mathrm{Sob}}$ (for some $r \ge 1$). The eigenvalue problem under study is then:

$$\boxed{T_{k^r_{\mathrm{Sob}}}\,\phi = \lambda\,\phi \quad \text{with} \quad \phi \in \mathrm{L}^2([0\,,1]) \quad \text{and} \quad \lambda > 0}\,. \tag{$\mathcal{S}^r_\lambda$}$$

The notation $(\mathcal{S}^r_\lambda)$ will be extensively used in all what follows, and sometimes declined to specific values of r. In particular, the letter $\mathcal{S}$ was chosen to indicate that the ambition is to extract a feature map by means of a *spectral* approach. For convenience, two additional notations are introduced:

- $\boldsymbol{\lambda}(T_{k^r_{\mathrm{Sob}}})$ is the eigenspectrum of the integral operator $T_{k^r_{\mathrm{Sob}}}$.
- For any fixed eigenvalue λ, $E_r(\lambda)$ is the eigenspace of λ, *i.e.* the subspace of $\mathrm{L}^2([0\,,1])$ composed of all the eigenfunctions of $T_{k^r_{\mathrm{Sob}}}$ associated to λ.

5.2.1. *Application to the Sobolev kernel* k^1_{Sob}

The Gram matrix $\boldsymbol{K}_n$ is built from $n = 500$ sample points and the results are represented in Figure 1. The histogram plot in Figure 1 (A) allows to visually appreciate the decay speed of the estimated eigenvalues. In Figure 1 (B), the eigenvectors are plotted together with the Nyström approximations that they allow to construct. For each pair $(\hat{\lambda}_k, \boldsymbol{w}_k)$, the eigenvector coefficients are marked by colored dots, while the resulting Nyström extension $\widehat{\phi}_k$ is represented as a solid line (of the same color) which interpolates these dots. To check if the eigendecay is truly polynomial, a simple trick is to apply a logarithmic transformation to the estimated eigenvalues. As can be observed in Figure 2 (A), the resulting estimates are then arranged along a decreasing straight line. The estimated slope $\hat{\beta}_1 \approx -2.0$ means that $\lambda_k = \mathcal{O}(1/k^2)$. This is fully consistent with Theorem 3.8, which gives significant credit to the KFA method.

Going back to Figure 1 (B), it can be observed that the eigenfunction estimates $(\widehat{\phi}_k)_{1\le k\le 6}$ look like some sinusoidal functions. Beyond the general appearance of the curves, some other particularities can be observed. Firstly, the eigenfunctions correspond to an increasing number of half-periods. Secondly, there is no phase shift and the eigenfunctions take their largest amplitude at the bounds of the interval $[0\,,1]$. In light of these elements, all the functions $t \mapsto \sqrt{2}\,\cos(k\pi t)$ with $k \ge 1$ emerge as plausible eigenfunctions. This conjecture is indeed true, as confirmed by the proposition below.

Theorem 5.1. *For any* $k \ge 1$*, the function* $c_k : t \in [0\,,1] \mapsto \sqrt{2}\,\cos(k\pi t)$ *is an* L^2*-normalized eigenfunction of* $T_{k^1_{\mathrm{Sob}}}$ *with associated eigenvalue* $\lambda_k := 1/(k\pi)^2$*. The eigenspectrum* $\boldsymbol{\lambda}(T_{k^1_{\mathrm{Sob}}})$ *consists of the null eigenvalue* 0 *and the eigenvalues* $(\lambda_k)_{k\ge1}$*. Accordingly, the Mercer expansion of* k^1_{Sob} *may be written as follows:*

$$\forall x, x' \in [0\,,1], \quad k^1_{\mathrm{Sob}}(x,x') = \sum_{k=1}^{\infty} \frac{1}{(k\pi)^2}\, c_k(x)\, c_k(x')\ . \tag{5.2}$$

Proof. For any given $x \in [0\,,1]$, $[T_{k^1_{\mathrm{Sob}}} c_k](x)$ can be expressed as an integral over $[0\,,1]$. Then, basic calculations allow to show that $[T_{k^1_{\mathrm{Sob}}} c_k](x) = c_k(x)/(k\pi)^2$. They are detailed in Appendix G.1 as supplementary material. This confirms that c_k is an eigenfunction and this also indicates that $\lambda_k := 1/(k\pi)^2$ is the corresponding eigenvalue. In addition, since k^1_{Sob} is an orthogonal kernel, $T_{k^1_{\mathrm{Sob}}}\,\mathbf{1} = 0$ with $\mathbf{1}$ denoting the constant function equal to 1. This proves that $0 \in \boldsymbol{\lambda}(T_{k^1_{\mathrm{Sob}}})$ and that $\mathbf{1} \in E_1(0)$. There is no other eigenvalue since the orthonormal system (ONS) defined by $(c_k)_{k\ge0} := \{\mathbf{1}\,; (c_k)_{k\ge1}\}$ is already an ONB of $\mathrm{L}^2([0\,,1])$ as recalled in [59] (see Theorem 4.1, p. 21). By Theorem 2.18, the Mercer decomposition of k^1_{Sob} directly follows. □

Remark 5.2. It must be acknowledged that the Mercer expansion of k^1_{Sob} was already existing in the literature. A specific remark on the eigendecomposition of $T_{k^1_{\mathrm{Sob}}}$ may be found in [35] while a detailed proof is provided in [37] (see Lemma 1, p. 9). The proof technique is rather different from ours since the idea is to expand k^1_{Sob} in an ONB of $\mathrm{L}^2([0\,,1]^2)$ obtained by tensorization of the ONB $(c_k)_{k\ge0}$. In the continuation of this work, two additional proof techniques will be proposed. In Section 6, the Mercer decomposition is recovered by solving a boundary value problem. In Section 7, the trick consists in rewriting k^1_{Sob} only in terms of B_2 before using a Fourier series expansion.

Theorem 5.1 allows to disclose the Mercer feature map of $K^1_{\mathrm{Sob}} = 1 + k^1_{\mathrm{Sob}}$. One can see that it is composed of one constant feature and an infinite number of purely sinusoidal features. The Mercer decomposition of K^1_{Sob} paves the way to a feature-based characterization of the unanchored Sobolev space $\mathcal{H}^1_{\mathrm{Sob}}$. Importantly, Theorem 2.21 ensures that the system $\{\mathbf{1}\,; (\tilde{c}_k)_{k\ge1}\}$ with $\tilde{c}_k(\cdot) := c_k(\cdot)/k\pi$ is an ONB of $\mathcal{H}^1_{\mathrm{Sob}}$.

Remark 5.3. The Mercer expansion of the unanchored Sobolev kernel K^1_{Sob} is actually very similar to that of the standard Sobolev kernel K^1 (see Appendix B.3). The eigenfunctions $\{\mathbf{1}\,; (c_k)_{k\ge1}\}$ are the same and the eigenvalues are asymptotically equivalent. At first sight, this may be surprising because the analytical expressions of K^1 and K^1_{Sob} are much different. However, from a theoretical viewpoint, this is pretty natural since the two kernels are related to the same function space $\mathrm{H}^1([0\,,1])$ and they result from the use of two equivalent metrics. Furthermore, Theorem 3.8 guarantees that the eigenvalues of T_{K^1} and $T_{K^1_{\text{Sob}}}$ have the same decay speed, namely $\mathcal{O}(1/k^2)$.

Now that k^1_{Sob} has been carefully studied, the question is whether k^2_{Sob} has similar properties.

5.2.2. *Application to the Sobolev kernel* k^2_{Sob}

The estimation method is exactly the same and the results (for $n = 500$ sample points) are represented in Figure 3. The largest estimated eigenvalue $\hat{\lambda}_1$ is predominant to such an extent that it is almost equal to the total eigensum. In comparison with the histogram plot from Figure 1 (A), the eigendecay observed in Figure 3 (A) is much faster. This is not really surprising since Theorem 3.8 indicates that the new decay order is $1/k^4$. However, this situation could not have been fully anticipated only on the basis of Theorem 3.8 because the very start of the eigendecay could as well have been much slower than the final asymptotic trend.

Just as for $r = 1$, the logarithmic transformation is applied to the eigenvalue estimates. The results are shown in Figure 2 (B) and one can see that the log-transformed eigenvalue estimates perfectly match with a straight line. After being rounded, the estimated slope $\hat{\beta}_1 = -4.44$ yields $\lambda_k = \mathcal{O}(1/k^4)$ and the decay rate stipulated in Theorem 3.8 is recovered.

As the largest eigenvalue λ_1 seems to be heavily dominating in $\boldsymbol{\lambda}(T_{k^2_{\text{Sob}}})$, the Mercer expansion of k^2_{Sob} may be restricted to its first term. Numerically, one can check that $\hat{\lambda}_1 \approx 1/12$. It can also be observed in Figure 3 (B) that $\widehat{\phi}_1$ looks like the L^2-normalized zero-mean linear function $P_1 : t \in [0\,,1] \mapsto 2\sqrt{3}\,(t - 1/2)$. Hence, k^2_{Sob} can be approximated by:

$$k^2_{\text{Sob}}(x, x') \approx \lambda_1\, \phi_1(x)\, \phi_1(x') \approx \hat{\lambda}_1\, P_1(x)\, P_1(x') = \left(x - \frac{1}{2}\right)\left(x' - \frac{1}{2}\right) = B_1(x)\, B_1(x') =: k_{\text{lin}}(x, x')\ . \tag{5.3}$$

k_{lin} is the dot-product kernel centered at the midpoint $(1/2, 1/2)$ of $[0\,,1]^2$. This kernel is not characteristic and it is therefore unsuitable for many tasks. The fact that $k^2_{\text{Sob}} \approx B_1 \otimes B_1$ also indicates that the numerical behavior of k^2_{Sob} is almost entirely driven by one single term of its definition. The strong similarity between k^2_{Sob} and $k_{\text{lin}} = B_1 \otimes B_1$ is an important conclusion whose implications will be further discussed.

The eigenfunction estimates $(\widehat{\phi}_k)_{1 \le k \le 6}$ shown in Figure 3 (B) deserve to be paid some attention. In fact, the shape of $\widehat{\phi}_k$ suggests that the underlying eigenfunction ϕ_k might be a polynomial function of degree k: $\widehat{\phi}_1$ looks like a straight line, $\widehat{\phi}_2$ like a parabola, $\widehat{\phi}_3$ like a cubic curve and so on. Remember that the shifted Legendre polynomials (see Appendix A.2) are a family of orthogonal polynomials in $\mathrm{L}^2([0,1])$. More precisely, it is the only family $(P_k)_{k\ge 0}$ such that $\deg(P_k) = k$ and $\langle P_k, P_l\rangle_{\mathrm{L}^2} = \delta_{kl}$. As a result, one can naturally wonder whether the eigenfunctions of $T_{k^2_{\text{Sob}}}$ are the shifted Legendre polynomials. Surprisingly, the conjecture is false. Therefore, the visual insights brought by KFA are misleading this time. This situation can be illustrated in the specific case of $P_1 = 2\sqrt{3}\, B_1$. In fact, when $T_{k^2_{\text{Sob}}}$ is applied to $B_1 \propto P_1$, a polynomial of degree 5 comes out:

$$\forall x\, \in [0\,,1], \quad [T_{k^2_{\text{Sob}}} B_1](x) = \frac{1}{12}\left(\frac{708}{720} x - \frac{1}{2}\right) + \left(\frac{1}{120} x^5 - \frac{1}{48} x^4 + \frac{1}{72} x^3\right). \tag{5.4}$$

The calculation details are provided in Appendix G.2 as supplementary material. Eq. (5.4) shows that $T_{k^2_{\text{Sob}}} B_1$ and B_1 cannot be proportional. Hence, P_1 cannot be an L^2-normalized eigenfunction of $T_{k^2_{\text{Sob}}}$. The same strategy could be used for P_2 but the integral calculations (which are already tedious for P_1) would become extremely tiresome. In Section 6, a much faster way to reach a more general conclusion will be presented.

The next objective is to study what happens when $r \ge 3$.

5.2.3. *Application to the Sobolev kernels* k^r_{Sob} *with* $r \ge 3$

Performing KFA for different Sobolev kernels such that $r \ge 3$ confirms that the eigenvalues of $T_{k^r_{\text{Sob}}}$ decrease with a polynomial rate of $1/k^{2r}$. In addition, two important properties are always verified. Firstly, the largest eigenvalue is strongly predominant and it can be approximated by $\hat{\lambda}_1 = 1/12$. Secondly, the eigenfunction estimates $(\widehat{\phi}_k)_{k\geq 1}$ look like the shifted Legendre polynomials $(P_k)_{k\geq 1}$. Based on these two remarks, all what was said in Section 5.2.2 can be repeated identically and the same conclusions can be drawn. As a result, the Sobolev kernels $(k^r_{\text{Sob}})_{r\geq 2}$ and the dot-product kernel k_{lin} are expected to show similar behaviors. This closeness is even accentuated because the eigendecay occurs faster (as r increases) and the truncation of the Mercer expansion in Eq. (5.3) is therefore all the more justified.

One major difference between k^2_{Sob} and k^r_{Sob} emerges as r becomes larger. This can be observed in Figure 4 where KFA is used to estimate the eigenfunctions associated to k^5_{Sob}. The eigenfunction estimates $(\widehat{\phi}_k)_{1\leq k\leq 6}$ and the shifted Legendre polynomials $(P_k)_{1\leq k\leq 6}$ seem to perfectly match. Although this is again a numerical illusion (as will be demonstrated in Section 6), this suggests that the shifted Legendre polynomials may be asymptotic eigenfunctions (in other words, the eigenfunctions of the limit integral operator obtained as $r\to\infty$). Such a thorny question is not answered here but it will be further studied in Section 8.

As a conclusion, KFA was of great help in the study of the Mercer feature maps associated to Sobolev kernels. For $r = 1$, KFA provided the guidance to explicit the Mercer decomposition of k^1_{Sob} and thereby understand that the associated feature map only consists of sinusoidal features. For $r = 2$, KFA highlighted the fact that k^r_{Sob} behaves like the dot-product kernel k_{lin} centered at $(1/2, 1/2)$ but the exact Mercer decomposition of k^2_{Sob} could not be disclosed. More generally, for $r\geq 3$, KFA showed that the eigenfunctions involved in the Mercer decomposition of k^r_{Sob} are very close to the shifted Legendre polynomials.

To go beyond these first conclusions, further investigations on Sobolev kernels are conducted in the next section. In particular, the use of a differential approach provides new answers.

6. Extraction of Mercer feature maps with a differential approach

In this section, a differential approach is proposed to gain additional insights into the Mercer feature maps of Sobolev kernels (especially in the case where $r\geq 2$). For this, the key idea is to differentiate the eigenvalue equation coming from Mercer's theorem and to transform the infinite-dimensional eigenvalue problem $(\mathcal{S}^r_\lambda)$ into a boundary value problem, *i.e.* an ordinary differential equation (ODE) subject to a system of boundary conditions. Section 6.1 explains how to switch from one formulation of the problem to the other. Section 6.2 deals with the resolution of the resulting boundary value problem. Depending on the smoothness parameter r, a closed-form solution may exist or not.

6.1. Transformation of the eigenvalue problem into a boundary value problem

Before computing derivatives, the important issue of the differentiability of the eigenfunctions involved in $(\mathcal{S}^r_\lambda)$ must be addressed. For fixed $r\geq 1$, let $\phi\in \mathrm{L}^2([0\,,1])$ be an eigenfunction of $T_{k^r_{\text{Sob}}}$. With Theorem 2.21, it is clear that ϕ belongs to the function space $\mathrm{H}^r([0\,,1])$. Then, in virtue of the basic results recalled in Section 3.1, and especially the definition of $\mathrm{H}^r([0\,,1])$ provided in Eq. (3.2), it can be said that:

(i) the weak derivative $D^k\phi$ is well defined at any order $k\geq 1$,

(ii) $D^k\phi$ remains in $\mathrm{L}^2([0\,,1])$ for all $0\leq k\leq r$,

(iii) ϕ is $r-1$ times continuously differentiable on $[0\,,1]$.

In short, ϕ is always at least continuous, with the worst case occurring for $r = 1$. It must be acknowledged that this simple conclusion could have been directly drawn from Theorem 2.18 (as k^r_{Sob} is continuous on $[0\,,1]^2$ whatever $r\geq 1$), and independently of the Sobolev embedding theorem, which is (in a sense) too strong a result. The continuity of ϕ is an essential property. It is actually the key argument to demonstrate that ϕ is infinitely differentiable on $[0\,,1]$. As a consequence, ϕ is much smoother than first expected, and appears to be a very special function within $\mathrm{H}^r([0\,,1])$. This point is clarified in the theorem below.

Proposition 6.1. *Let $\phi \in \mathrm{L}^2([0\,,1])$ be an eigenfunction of $T_{k^r_{\mathrm{Sob}}}$ with eigenvalue $\lambda > 0$. Then, $\phi \in \mathrm{C}^\infty([0\,,1])$ and it is a solution of the following ODE:*

$$\lambda\,\phi^{[2r]} + (-1)^{r+1}\,\phi = 0\ . \tag{$\mathcal{E}^r_\lambda$}$$

The reader is referred to Appendix F.2 for the detailed proof. Knowing that $\phi \in \mathrm{C}([0\,,1])$, the Leibniz integral rule allows to justify that the eigenvalue equation $\lambda\,\phi = T_{k^r_{\mathrm{Sob}}}\,\phi$ can be differentiated on both sides. This leads to an integral expression of $\lambda\,\phi^{[1]}$ and the same technique can be repeated. The derivatives $\phi^{[1]}, \ldots, \phi^{[2r]}$ are computed recursively and the properties of Bernoulli polynomials make ϕ be a solution of the ODE ($\mathcal{E}^r_\lambda$). As $\phi^{[2r]} \propto \phi$, a simple proof by induction leads to $\phi \in \mathrm{C}^\infty([0\,,1])$.

With the ODE ($\mathcal{E}^r_\lambda$) from Proposition 6.1, it can be proved much more easily that $T_{k^r_{\mathrm{Sob}}}$ does not admit any polynomial eigenfunction, as stated in Corollary 6.2 below.

Corollary 6.2. *Let $\phi \in \mathrm{L}^2([0\,,1])$ be an eigenfunction of $T_{k^r_{\mathrm{Sob}}}$ with eigenvalue $\lambda > 0$. Then, ϕ cannot be a polynomial function.*

Proof. For the sake of contradiction, let us assume that $T_{k^r_{\mathrm{Sob}}}$ admits one polynomial eigenfunction ϕ with eigenvalue $\lambda > 0$. Let $P \in \mathbb{R}[X]$ denote the polynomial expression associated to the polynomial function $\phi \in \mathbb{R}^{[0,1]}$. According to Theorem 6.1, ϕ is one solution of the ODE ($\mathcal{E}^r_\lambda$). After switching to polynomial expressions, one has $\lambda\,P^{[2r]} = (-1)^r P$. Taking the degree on both sides of the previous equality leads to a contradiction. $\square$

It is important to realize that Proposition 6.1 provides a necessary but not sufficient condition for being an eigenfunction. In particular, for any positive eigenvalue λ, the solution space associated to ($\mathcal{E}^r_\lambda$) is too big compared to what is expected for the solution space of ($\mathcal{S}^r_\lambda$). On the one side, the solution space associated to ($\mathcal{S}^r_\lambda$) is the eigenspace $E_r(\lambda)$. According to what was said in Section 5 (especially in Theorem 5.1), $E_r(\lambda)$ is a 1-dimensional linear subspace of $\mathrm{L}^2([0\,,1])$. On the other side, the solution space associated to ($\mathcal{E}^r_\lambda$) is a $2r$-dimensional linear subspace of $\mathrm{L}^2([0\,,1])$. As a consequence, additional constraints must be added to ($\mathcal{E}^r_\lambda$) in order to restrain the solution space. This is the subject of the theorem below.

Theorem 6.3. *Let $\phi \in \mathrm{L}^2([0\,,1])$ and $\lambda > 0$. Then, the two following statements are equivalent:*

(i) ϕ is an eigenfunction of the integral operator $T_{k^r_{\mathrm{Sob}}}$ with eigenvalue λ.

(ii) ϕ is a solution of the boundary value problem defined as:

$$\boxed{\lambda\,\phi^{[2r]} + (-1)^{r+1}\,\phi = 0\ \textit{with}\ \begin{cases} & \phi^{[r]}(0) = \phi^{[r]}(1) = 0 \\ \forall\, 0 \le p \le r-2, & (-1)^{r+p}(\phi^{[p]}(1) - \phi^{[p]}(0)) = \phi^{[2r-p-1]}(0) \\ \forall\, 0 \le p \le r-2, & \phi^{[2r-p-1]}(0) = \phi^{[2r-p-1]}(1) \end{cases}}\ . \tag{$\mathcal{B}^r_\lambda$}$$

The reader is referred to Appendix F.3 for the detailed proof. As Theorem 6.3 is an extended version of Proposition 6.1, the proof does not repeat what is said in Appendix F.2 and rather focuses on how to obtain the boundary conditions and the converse statement. For the boundary conditions, the derivatives of the eigenvalue equation (already calculated for the purposes of Proposition 6.1) need to be evaluated at the bounds of $[0\,,1]$ and the properties of Bernoulli polynomials eventually lead to the expected constraints. For the converse statement, one must write $T_{k^r_{\mathrm{Sob}}}\,\phi = (-1)^r\,\lambda\,T_{k^r_{\mathrm{Sob}}}\,\phi^{[2r]}$ and repeat $2r$ times a trick that consists in using an integration by parts before selecting an appropriate boundary condition to enable simplifications.

Remark 6.4. Theorem 6.3 is an important result since the equivalence between the eigenvalue problem ($\mathcal{S}^r_\lambda$) and the boundary value problem ($\mathcal{B}^r_\lambda$) is clearly established. Differentiating the eigenvalue equation $T_K\,\phi = \lambda\,\phi$ in order to obtain an equivalent boundary value problem is a well-known proof technique to disclose the Mercer expansion of a given kernel. Other examples of boundary value problems obtained in this way are provided in the Appendix B.

As many analytical techniques are available to solve linear ODEs, the reformulation of $(\mathcal{S}_\lambda^r)$ as $(\mathcal{B}_\lambda^r)$ opens the way for a possibly faster and simpler analytical resolution. This question is discussed in the next section.

6.2. **Analytical resolution of the boundary value problem**

For $r=1$, the boundary value problem $(\mathcal{B}_\lambda^r)$ becomes:

$$\lambda\,\phi'' + \phi = 0 \quad \text{with} \quad \phi'(0) = \phi'(1) = 0 \quad \text{and} \quad \lambda > 0\ . \tag{$\mathcal{B}_\lambda^1$}$$

The resolution is straightforward (see Appendix E.1) and fortunately allows to recover the Mercer expansion already established in Theorem 5.1.

For $r \geq 2$, solving the homogeneous linear ODE $(\mathcal{E}_\lambda^r)$ simply boils down to finding the complex roots of the characteristic polynomial $\chi_r(z) := \lambda\, z^{2r} + (-1)^{r+1}$. To achieve this, one can write:

$$\chi_r(z) = 0 \iff \lambda\, z^{2r} = (-1)^r \iff \left[\sqrt[2r]{\lambda}\, z\right]^{2r} = i^{2r} \iff \zeta^{2r} = 1 \ \text{ with } \ \zeta := -i\ \sqrt[2r]{\lambda}\, z\ .$$

After using the $2r$-th roots of unity in $\mathbb{C}$, a solution set containing $2r$ elements can be derived:

$$\begin{aligned} \chi_r(z) = 0 &\iff \zeta \in \left\{ e^{2i\pi\left(\frac{j}{2r}\right)} \ \text{ with } \ 0 \leq j \leq 2r-1 \right\} \\ &\iff z \in \left\{ \xi\, e^{2i\pi\left(\frac{j}{2r}+\frac{1}{4}\right)} \ \text{ with } \ \xi := \frac{1}{\sqrt[2r]{\lambda}} \ \text{ and } \ \left\lfloor \frac{r}{2} \right\rfloor \leq j \leq \left\lceil \frac{3r}{2} \right\rceil - 1 \right\}\ . \end{aligned}$$

where $\lfloor\cdot\rfloor$ and $\lceil\cdot\rceil$ are respectively the floor and ceiling functions. For the sake of convenience, the roots of χ_r are indifferently denoted by $(z_i)_{1\leq i\leq 2r}$ in what follows. Any solution of the ODE $(\mathcal{E}_\lambda^r)$ can then be expressed as:

$$\forall t \in [0\,,1], \quad \phi(t) = \sum_{i=1}^{2r} w_i\, \exp(z_i\, t) \quad \text{with} \quad \boldsymbol{w} := (w_1, \ldots, w_{2r}) \in \mathbb{C}^{2r}\ . \tag{6.1}$$

Since all coefficients in χ_r are real numbers, there also exists a real-valued variant of Eq. (6.1) but it would be tedious and useless to formalize it. Thus, the construction of the solution space associated to $(\mathcal{E}_\lambda^r)$ is not a problem, especially if a basis of complex-valued solutions is used to do so. On the contrary, it is much harder to restrain this solution space according to the boundary conditions specified in $(\mathcal{B}_\lambda^r)$. In fact, when the boundary conditions are rewritten in terms of the general solution given in Eq. (6.1), a non-linear system depending on both ξ and $w_1, \ldots, w_{2r}$ comes out and it cannot be solved analytically.

Remark 6.5. For the Laplace kernel and the uniform distribution on a symmetric interval $[-a\,,a]$, the same pitfall is encountered (see Appendix B.2). The eigenvalue problem is transformed into a boundary value problem which cannot be solved completely [48] (see Section 2.3.3, pp. 27–45). In particular, the Mercer expansion remains partly implicit because the eigenvalues rely on the unknown solutions of a transcendental equation.

Let us take a close look at what happens for $r=2$. This time, the boundary value problem is given by:

$$\lambda\,\phi^{[4]} - \phi = 0 \quad \text{with} \quad \begin{cases} \phi''(0) = \phi''(1) = 0 \\ \phi(1) - \phi(0) = \phi^{[3]}(0) = \phi^{[3]}(1) \end{cases} \quad \text{and} \quad \lambda > 0\ , \tag{$\mathcal{B}_\lambda^2$}$$

and any real-valued solution of the ODE may be written as:

$$\phi(t) = \alpha e^{\xi t} + \beta e^{-\xi t} + \gamma\cos(\xi t) + \delta\sin(\xi t) \quad \text{with} \quad \xi := \frac{1}{\sqrt[4]{\lambda}} \quad \text{and} \quad (\alpha,\beta,\gamma,\delta) \in \mathbb{R}^4. \tag{6.2}$$

With this kind of general solution, the boundary conditions may be rearranged as a system of $2r = 4$ equations:

$$\boldsymbol{M}_2(\xi)\,\boldsymbol{w} = \boldsymbol{0} \text{ where } \boldsymbol{M}_2(\xi) := \left[\begin{array}{c|c|c|c} 1 & 1 & -1 & 0 \\ e^{\xi} & e^{-\xi} & -\cos(\xi) & -\sin(\xi) \\ e^{\xi}-1 & 1-e^{-\xi} & \sin(\xi) & 1-\cos(\xi) \\ e^{\xi}-\xi^3-1 & e^{-\xi}+\xi^3-1 & \cos(\xi)-1 & \sin(\xi)+\xi^3 \end{array}\right] \text{ and } \boldsymbol{w} := \begin{bmatrix} \alpha \\ \beta \\ \gamma \\ \delta \end{bmatrix} .$$

Contrary to what such a rearrangement might suggest, the equation $\boldsymbol{M}_2(\xi)\,\boldsymbol{w} = \boldsymbol{0}$ is not a linear system because most coefficients in $\boldsymbol{M}_2(\xi)$ depend on ξ (or equivalently on λ) which is unknown. Therefore, a system of $2r = 4$ equations in $2r + 1 = 5$ unknowns must be solved. However, as ξ is the only unknown involved in the parametrization of the square matrix $\boldsymbol{M}_2(\xi)$, the specific structure adopted by the boundary conditions paves the way to a simple characterization of the eigenvalues:

- If $\boldsymbol{M}_2(\xi)$ is an invertible matrix, the linear equation $\boldsymbol{M}_2(\xi)\,\boldsymbol{w} = \boldsymbol{0}$ admits $\boldsymbol{w} = \boldsymbol{0}$ as unique solution in $\mathbb{R}^4$. In this case, the solution space of $(\mathcal{B}^2_\lambda)$ is restricted to $\phi = 0$. Hence, $\lambda = 1/\xi^4$ is not an eigenvalue.
- If $\boldsymbol{M}_2(\xi)$ is a singular matrix, $\boldsymbol{M}_2(\xi)\,\boldsymbol{w} = \boldsymbol{0}$ admits infinitely many non-zero solutions in $\mathbb{R}^4$. The functions resulting from the use of such coefficients in Eq. (6.2) are non-zero solutions of the boundary value problem $(\mathcal{B}^2_\lambda)$. They thus live in the eigenspace $E_2(\lambda)$ associated to the eigenvalue $\lambda = 1/\xi^4$.

As a consequence, non-trivial eigenfunctions only arise from situations where the matrix $\boldsymbol{M}_2(\xi)$ is singular. The eigenspectrum of $T_{k^2_{\text{Sob}}}$ can thus be found out after solving the equation:

$$\eta_2(\xi) := \det\left[\boldsymbol{M}_2(\xi)\right] = 0 .$$

Computing analytically the determinant of $\boldsymbol{M}_2(\xi)$ yields:

$$\forall \xi > 0, \quad \eta_2(\xi) = -2e^{-\xi}\left[2\,(e^{\xi}-1)^2\,\sin(\xi) + (\xi^3 + e^{2\xi}\,(\xi^3 - 2) + 2)\cos(\xi) - 2\,(1 + \xi^3 e^{\xi} - e^{2\xi})\right] . \tag{6.3}$$

The roots of η_2 on $(0\,,+\infty)$ do not have a closed-form expression. However, η_2 is asymptotically equivalent to the very simple function η_2^∞ defined below.

Proposition 6.6. *The asymptotic behavior of η_2 obeys the following equivalence relation:*

$$\eta_2(\xi) \underset{\xi\to\infty}{\sim} \eta_2^\infty(\xi) := -a_2(\xi)\,b_2(\xi) \quad \textit{with} \quad \begin{cases} a_2(\xi) := 2\,\xi^3\,e^{\xi} \\ b_2(\xi) := \cos(\xi) \end{cases} .$$

Proof. After identifying that $\xi^3\,e^{2\xi}$ is the leading-order factor in Eq. (6.3), η_2^∞ follows immediately:

$$\begin{aligned} \eta_2(\xi) &= -2\xi^3 e^{\xi}\left[\underbrace{\frac{2(e^{\xi}-1)^2}{\xi^3 e^{2\xi}}}_{\xrightarrow[\xi\to\infty]{}0}\sin(\xi) + \underbrace{\frac{\xi^3 + e^{2\xi}(\xi^3-2)+2}{\xi^3 e^{2\xi}}}_{\xrightarrow[\xi\to\infty]{}1}\cos(\xi) - \underbrace{\frac{2(1+\xi^3 e^{\xi} - e^{2\xi})}{\xi^3 e^{2\xi}}}_{\xrightarrow[\xi\to\infty]{}0}\right] \\ &\underset{\xi\to\infty}{\sim} -2\,\xi^3\,e^{\xi}\,\cos(\xi) . \end{aligned}$$

$\square$

A direct consequence of Proposition 6.6 is the following approximation formula for small eigenvalues:

$$\begin{matrix} \eta_2(\xi) = 0 \\ \text{for } \xi \to \infty \end{matrix} \iff \begin{matrix} \eta_2^\infty(\xi) = 0 \\ \text{for } \xi \to \infty \end{matrix} \iff \begin{matrix} \cos(\xi) = 0 \\ \text{for } \xi \to \infty \end{matrix} \iff \begin{matrix} \xi_k = \left(\frac{2k+1}{2}\right)\pi \\ \text{for } k \to \infty \end{matrix} \iff \boxed{\begin{matrix} \lambda_k = \dfrac{1}{\left[\left(\frac{2k+1}{2}\right)\pi\right]^4} \\ \text{for } k \to \infty \end{matrix}} \,. \tag{6.4}$$

These asymptotic derivations are consistent with Theorem 3.8 which says that the eigenvalues decrease with a rate proportional to $1/k^4$ in the case of K^2_{Sob}. In fact, Eq. (6.4) does even better than Theorem 3.8 because the eigendecay is henceforth known up to the multiplicative constant (and not only bounded).

For $r = 3$, the same method can be used to deduce the decay rate of the eigenvalues. The general solution of the ODE $(\mathcal{E}^3_\lambda)$ is easy to obtain (see Appendix E.2 for further details) but it is still a bit more complicated to leverage the extra information brought by the boundary conditions. They can be reorganized as $\boldsymbol{M}_3(\xi)\,\boldsymbol{w} = \boldsymbol{0}$ with $\boldsymbol{w}$ containing the $2r = 6$ coefficients involved in the general solution of the ODE. Therefore, the eigenvalues of $T_{k^3_{\text{Sob}}}$ correspond to the roots of the function $\eta_3 : \xi \mapsto \det[\boldsymbol{M}_3(\xi)]$ on $(0\,,+\infty)$. The intractable expression of $\boldsymbol{M}_3(\xi)$ makes pointless the calculation by hand of the determinant involved in the definition of η_3. Therefore, no counterpart of Eq. (6.3) is available when $r = 3$. However, an asymptotically equivalent function η_3^∞ can be found, as shown below.

Proposition 6.7. *The asymptotic behavior of η_3 obeys the following equivalence relation:*

$$\eta_3(\xi) \underset{\xi\to\infty}{\sim} \eta_3^\infty(\xi) := -a_3(\xi)\, b_3(\xi) \quad \textit{with} \quad \begin{cases} a_3(\xi) := \dfrac{3}{4}\,\xi^8\, e^{\sqrt{3}\,\xi} \\ b_3(\xi) := \sin(\xi) \end{cases} .$$

The reader is referred to Appendix E.2 for the entire proof. Once again, an approximation formula for small eigenvalues can be derived:

$$\begin{matrix} \eta_3(\xi) = 0 \\ \text{for } \xi \to \infty \end{matrix} \iff \begin{matrix} \eta_3^\infty(\xi) = 0 \\ \text{for } \xi \to \infty \end{matrix} \iff \begin{matrix} \sin(\xi) = 0 \\ \text{for } \xi \to \infty \end{matrix} \iff \begin{matrix} \xi_k = k\pi \\ \text{for } k \to \infty \end{matrix} \iff \boxed{\begin{matrix} \lambda_k = \dfrac{1}{(k\pi)^6} \\ \text{for } k \to \infty \end{matrix}} \,. \tag{6.5}$$

Just as for $r = 2$, the approximation formula obtained after asymptotic derivations may be seen as a refined version of Theorem 3.8 which only allowed to affirm that $\lambda_k = \mathcal{O}(1/k^6)$.

For $r \geq 4$, the general solution of the ODE $(\mathcal{E}^r_\lambda)$ can be derived in the same way as for $r \in \{2, 3\}$. However, the analytical resolution of the boundary value problem $(\mathcal{B}^r_\lambda)$ cannot be pushed any further because the expression of $\boldsymbol{M}_{2r}(\xi)$ is too cumbersome. Therefore, no generalization of Eq. (6.4) and (6.5) might be obtained, unless using a symbolic math software.

As a conclusion, Section 6 focused on the transformation of the original eigenvalue problem $(\mathcal{S}^r_\lambda)$ into the boundary value problem $(\mathcal{B}^r_\lambda)$. A general solution can always be found for the associated ODE but it is then much harder to take into account the boundary conditions. A complete analytical resolution is possible for $r = 1$ and allows to recover the Mercer expansion of k^1_{Sob}. For $r \geq 2$, the boundary conditions result in a non-linear system of equations without closed-form solution. This unfortunately means that the Mercer expansion of k^r_{Sob} remains partly implicit, and that we should not hope for better.

Beyond the expected conclusions, the differential approach also brings additional answers. On the one hand, it is now mathematically proved that the eigenfunctions cannot be of polynomial type. On the second hand, when the boundary conditions do not allow for an exact resolution, they can still be used to derive approximation formulas which confirm that the decay rates are polynomial with order of $1/k^{2r}$ (at least for $r \in \{1, 2, 3\}$).

Despite those findings, the identification of a fully explicit feature map remains an open question. In the next section, an approach based on Fourier series provides complementary insights.

7. Feature maps based on a sub-kernel decomposition

Significant efforts were made in Sections 5 and 6 with the aim of discovering the Mercer decomposition of the Sobolev kernel k^r_{Sob}. The related feature map φ^r_{Sob} is of particular interest because its features take the form of mutually orthogonal functions in $\mathrm{L}^2([0\,,1])$. Unfortunately, φ^r_{Sob} has a closed-form expression only for $r=1$ and remains partly implicit in all other cases. However, the study of Sobolev kernels must not be stopped here.

The main objective of this section is therefore to identify another feature map $\psi^r_{\mathrm{Sob}} : [0\,,1] \to \ell^2(\mathbb{N}^*)$. To this end, Section 7.1 seeks to rewrite k^r_{Sob} as the sum of two kernels in order to split the initial problem into two simpler problems. Then, it is shown in Section 7.2 that an explicit feature map can be easily extracted for each of the two identified kernels. The feature map ψ^r_{Sob} obtained after merging the two collections of features offers a different view on Sobolev kernels and notably allows to derive an ONB of $\mathcal{H}^r_{\mathrm{Sob}}$ (whatever is $r \geq 2$). Finally, in Section 7.3, the links between the non-orthogonal feature map ψ^r_{Sob} and the Mercer feature map φ^r_{Sob} are discussed. In particular, the additional insights provided by ψ^r_{Sob} are leveraged to better understand the numerical results observed in Section 5.

7.1. Decomposition of Sobolev kernels into a sum of two sub-kernels

Let us consider the following decomposition of k^r_{Sob} as the sum of two kernels:

$$k^r_{\mathrm{Sob}}(x,x') = k^r_A(x,x') + k^r_B(x,x') \quad \text{with} \quad k^r_A := \sum_{i=1}^{r} \frac{B_i(x)\, B_i(x')}{(i!)^2} \quad \text{and} \quad k^r_B(x,x') := \frac{(-1)^{r+1}}{(2r)!} B_{2r}(|x-x'|)\ .$$

It is trivial to see that k^r_A is a kernel as it is a positive linear combination of r symmetric separable functions. In the case of k^r_B, positive definiteness deserves further explanations. First, it should be noted that all Bernoulli polynomials (except B_1) have equal boundary values, *i.e.* they verify $B_n(0) = B_n(1)$ for $n \neq 1$. Hence, each polynomial B_n may be envisioned as the restriction on $[0\,,1]$ of a 1-periodic continuous function defined on $\mathbb{R}$. This suggests to compute the Fourier series expansion of B_n. The Fourier coefficients of B_n are actually easy to derive (see Appendix A.1.5) and one eventually has:

$$\forall\, n \geq 2, \quad \forall\, x \in [0\,,1], \quad B_n(x) = (-2)\, n! \sum_{k=1}^{\infty} \frac{\cos(2k\pi x - \frac{n\pi}{2})}{(2k\pi)^n}\ . \tag{7.1}$$

After replacing B_{2r} by its Fourier series expansion, k^r_B turns into:

$$\forall\, x,x' \in [0\,,1], \quad k^r_B(x,x') \quad = \quad 2\sum_{k=1}^{\infty} \frac{1}{(2k\pi)^{2r}} \left[\cos(2k\pi x)\cos(2k\pi x') + \sin(2k\pi x)\sin(2k\pi x')\right]\ . \tag{7.2}$$

Since k^r_B is expressed as a series of symmetric separable functions with positive coefficients, it is now obvious that it is a positive definite function on $[0\,,1]^2$.

Remark 7.1. $K^r_B = 1 + k^r_B$ is a well-known kernel in the literature [6, 12, 28]. It is often introduced as the reproducing kernel of the following periodic Sobolev space:

$$\mathrm{H}^r_{\mathrm{per}}([0\,,1]) := \left\{ h \in \mathbb{R}^{[0,1]} \;\middle|\; \begin{array}{ll} h^{[k]} \in \mathrm{L}^2([0\,,1]) & \forall\, 0 \leq k \leq r \\ h^{[k]}(0) = h^{[k]}(1) & \forall\, 0 \leq k \leq r-1 \end{array} \right\}, \tag{7.3}$$

when this one is equipped with the inner product:

$$\begin{array}{rccc} \langle\cdot,\cdot\rangle_{\mathrm{H}^r_{\mathrm{per}}} : & \mathrm{H}^r_{\mathrm{per}}([0\,,1])\times\mathrm{H}^r_{\mathrm{per}}([0\,,1]) & \longrightarrow & \mathbb{R} \\ & (\,h_1\,,\,h_2\,) & \longmapsto & \left(\int_0^1 h_1(x)\,\mathrm{d}x\right)\left(\int_0^1 h_2(x)\,\mathrm{d}x\right)+\int_0^1 h_1^{[r]}(x)\,h_2^{[r]}(x)\,\mathrm{d}x\ . \end{array} \tag{7.4}$$

$\mathrm{H}^r_{\mathrm{per}}([0\,,1])$ is obtained from $\mathrm{H}^r([0\,,1])$ by adding boundary conditions that force the first $r-1$ derivatives to have equal boundary values.

7.2. Identification of explicit feature maps

7.2.1. *Explicit feature map for k^r_A*

For the kernel k^r_A, a feature map $\psi^r_A:[0\,,1]\to\mathbb{R}^r$ immediately stands out:

$$k^r_A(x,x')=\langle\psi^r_A(x),\psi^r_A(x')\rangle_{\mathbb{R}^r}\quad\text{with}\quad\psi^r_A(x)=\left(\widetilde{B}_k(x)\right)_{1\le k\le r}\quad\text{where}\quad\widetilde{B}_k(x):=\frac{B_k(x)}{k!}\ . \tag{7.5}$$

The feature space is the Euclidean space $\mathbb{R}^r$ and every feature function $\psi^r_A(x)$ consists of r polynomial features.

Remark 7.2. As Bernoulli polynomials are not orthogonal in $\mathrm{L}^2([0\,,1])$, the definition of k^r_A is not a Mercer expansion of k^r_A with respect to the uniform distribution on $[0\,,1]$. To obtain one, there is no other option than solving analytically the eigenvalue problem related to $T_{k^r_A}$. Whether it is straightforward to prove that:

$$T_{k^r_A}\,\phi=\lambda\,\phi\quad\Longrightarrow\quad\phi\in\mathrm{Span}(\{B_1,\dots,B_r\})\ ,$$

it is much more complicated to go further. In fact, tedious and error-prone hand calculations are required in order to determine which linear combinations of $B_1,\dots,B_r$ correspond to the eigenfunctions of $T_{k^r_A}$. Even for $r=3$, this asks for considerable efforts. Consequently, there is no point in going further in this direction.

7.2.2. *Explicit feature map for k^r_B*

For the kernel k^r_B, a feature map $\psi^r_B:[0\,,1]\to\ell^2(\mathbb{N}^*)$ stems from the series expansion established in Eq. (7.2):

$$\begin{gathered} k^r_B(x,x')=\langle\psi^r_B(x),\psi^r_B(x')\rangle_{\ell^2}\quad\text{with}\quad\psi^r_B(x):=\left[\,(\tilde{c}^r_{2k})_{k\ge1}\ ;\ (\tilde{s}^r_{2k})_{k\ge1}\,\right] \\ \text{where}\quad\begin{cases}\tilde{c}^r_{2k}(x):=c_{2k}(x)/(2k\pi)^r\\ \tilde{s}^r_{2k}(x):=s_{2k}(x)/(2k\pi)^r\end{cases}\quad\text{and}\quad\begin{cases}c_{2k}(x):=\sqrt{2}\,\cos(2k\pi x)\\ s_{2k}(x):=\sqrt{2}\,\sin(2k\pi x)\end{cases}\ . \end{gathered} \tag{7.6}$$

Here, the feature space is the Hilbert space $\ell^2(\mathbb{N}^*)$ and every feature function $\psi^r_B(x)$ is composed of an infinite number of sinusoidal features which are organized into pairs. From one pair to the next, the frequency $f_k:=k$ increases whereas the amplitude $A^r_k:=\sqrt{2}/(2k\pi)^r$ decreases.

The ONS $\{(c_{2k})_{k\ge1},(s_{2k})_{k\ge1}\}$ is nothing but the Fourier ONB of $\mathrm{L}^2([0\,,1])$, after removing the constant function $\mathbf{1}$ from it. As a consequence, Eq. (7.2) is exactly the Mercer decomposition of k^r_B and the eigenvalues of $T_{k^r_B}$ are given by $\mu_k:=1/(2k\pi)^{2r}$ with $k\ge1$. The polynomial rate of the eigendecay is perfectly in line with Theorem 3.8 because k^r_B is the reproducing kernel of an infinite-dimensional sub-RKHS of $\mathrm{H}^r([0\,,1])$. In the rest of this work, the notation ψ^r_B is replaced by φ^r_B to stress that this feature map is of Mercer's type.

It is worth noting that the smoothness parameter r has a very targeted influence on the Mercer decomposition of k^r_B. One the one hand, the exponent of the power law describing the eigendecay is equal to $2r$. On the other hand, r does not interfere in the definition of the eigenfunctions $(c_{2k})_{k\ge1}$ and $(s_{2k})_{k\ge1}$. This means that all kernels k^r_B (with $r\ge1$) induce the same initial collection of sinusoidal features but they are not weighted in the same way from one kernel k^r_B to another. High-frequency features are always more penalized than low-frequency features, and the imbalance between them worsens as r increases.

7.2.3. *Explicit feature map for k^r_{Sob}*

Replacing k^r_B by its Mercer expansion in the initial definition of k^r_{Sob} yields:

$$k^r_{\text{Sob}}(x,x') = \sum_{k=1}^{r} \frac{B_k(x)\,B_k(x')}{(k!)^2} + \sum_{k=1}^{\infty} \frac{c_{2k}(x)\,c_{2k}(x')}{(2k\pi)^{2r}} + \sum_{k=1}^{\infty} \frac{s_{2k}(x)\,s_{2k}(x')}{(2k\pi)^{2r}} \tag{7.7}$$

$$= \sum_{k=1}^{r} \widetilde{B}_k(x)\,\widetilde{B}_k(x') + \sum_{k=1}^{\infty} \tilde{c}^r_{2k}(x)\,\tilde{c}^r_{2k}(x') + \sum_{k=1}^{\infty} \tilde{s}^r_{2k}(x)\,\tilde{s}^r_{2k}(x') \tag{7.8}$$

and a global feature map $\psi^r_{\text{Sob}} : [0\,,1] \to \ell^2(\mathbb{N}^*)$ follows:

$$k^r_{\text{Sob}}(x,x') = \langle \psi^r_{\text{Sob}}(x), \psi^r_{\text{Sob}}(x')\rangle_{\ell^2} \quad \text{with} \quad \psi^r_{\text{Sob}}(x) := [\,\psi^r_A(x)\,;\,\psi^r_B(x)\,] = \left[\,\left(\widetilde{B}_k\right)_{1\le k\le r}\,;\,(\tilde{c}^r_{2k})_{k\ge1}\,;\,(\tilde{s}^r_{2k})_{k\ge1}\,\right]\,.$$

ψ^r_{Sob} is not a Mercer feature map because the identified features are not mutually orthogonal in $\mathrm{L}^2([0\,,1])$ even though there exist orthogonality relations between some of them.

In return for its non-orthogonality, ψ^r_{Sob} has the great advantage of being totally explicit. In particular, ψ^r_{Sob} allows to clearly establish the coexistence of two different kinds of features within k^r_{Sob}. Moreover, the role played by the smoothness parameter r in the balance between the polynomial and sinusoidal features is now clear. When r is increased by one unit, the former polynomial features $(\widetilde{B}_k)_{1\le k\le r-1}$ remain unchanged, a new polynomial feature $\widetilde{B}_r$ comes into play and the weights $(\sqrt{\mu_k})_{k\ge1}$ assigned to the sinusoidal features $(c_{2k})_{k\ge1}$ and $(s_{2k})_{k\ge1}$ are all multiplied by the factor $1/(2k\pi)$. In short, the amount of polynomial features increases while the amplitude of sinusoidal features is gradually shrunk. This explains why KFA shows polynomial-like eigenfunctions for $r=2$ and beyond (see Figures 3 to 4).

The numerical results brought by KFA and the new insights brought by the analysis ψ^r_{Sob} are indeed consistent and complementary. From a theoretical viewpoint, the analytical expression of ψ^r_{Sob} indicates that there is a gradual evolution of the balance between polynomial and sinusoidal features as r becomes larger. From a numerical viewpoint, KFA reveals that this transition is rather abrupt and corresponds to the switch from $r=1$ to $r\ge2$. Given ψ^1_{Sob}, inserting $\widetilde{B}_2$ and replacing the initial weights $1/(2k\pi)$ by smaller weights $1/(2k\pi)^2$ is thus enough to trigger a definitive reversal in the balance between the polynomial and sinusoidal features. This question is further investigated in Section 7.3.

7.2.4. *Feature-based characterization of $\mathcal{H}^r_{\text{Sob}}$*

k^r_{Sob} was rewritten in Eq. (7.8) as a series of symmetric and separable functions. In view of Theorem 2.23, one may monder whether the system defined by:

$$(g^r_k)_k := \left\{\,\mathbf{1}\,;\,(\widetilde{B}_k)_{1\le k\le r}\,;\,(\tilde{c}^r_{2k})_{k\ge1}\,;\,(\tilde{s}^r_{2k})_{k\ge1}\,\right\} = \{\,\mathbf{1}\,;\,(f^r_k)_k\} \tag{7.9}$$

is an ONB of $\mathcal{H}^r_{\text{Sob}}$ or not. It can be easily proved that this system is not ω-independent (see Remark 2.24). Indeed, after taking $n=2$ in Eq. (7.1), a series expansion of $\widetilde{B}_2$ follows:

$$\widetilde{B}_2(\cdot) = \frac{1}{2}\,B_2(\cdot) = 2\sum_{k=1}^{\infty}(2k\pi)^{r-2}\,\tilde{c}^r_{2k}(\cdot)\,,$$

which means that one can find a non-zero sequence of coefficients $(\gamma_k)_k$ making the series $\sum_k \gamma_k\, g^r_k(\cdot)$ converge and be equal to zero everywhere on $[0\,,1]$. However, no square-summable sequence is able to do the same.

Proposition 7.3. *The system $(g^r_k)_k$ defined by Eq. (7.9) is ℓ^2-linearly independent.*

The reader is referred to Appendix F.4 for the detailed proof. A key step is to prove that $\mathrm{H}^r([0\,,1])$ is the direct sum of the function spaces induced by the sub-kernels k_A^r and K_B^r.

Now that Proposition 7.3 is stated, Theorem 2.23 can be applied rigorously. Therefore, it can be concluded that the system $(g_k^r)_k$ is indeed an ONB of $\mathcal{H}_{\mathrm{Sob}}^r$.

7.3. **Comparison with Mercer feature maps**

The objective is now to compare the Mercer feature map φ_{Sob}^r (known for $r = 1$ and approximated for $r \geq 2$) and the non-orthogonal feature map ψ_{Sob}^r (known whatever is $r \geq 1$). φ_{Sob}^r involves features which are either purely sinusoidal (for $r = 1$) or pseudo-polynomial (for $r \geq 2$). On the contrary, the closed-form expression of ψ_{Sob}^r shows that polynomial and sinusoidal features always coexist. To bridge the gap between the two feature maps, the L^2-norms of the non-orthogonal features $(g_k^r)_k$ must be computed. In fact, according to what was said in Section 4.2, and more precisely what was established in Eq. (4.4), the contribution of each input basis function g_k^r to the final value of the index S_i^{HSIC} defined in Eq. (4.7) can be read through its norm in $\mathrm{L}^2([0\,,1])$. Moreover, it can also be shown that:

$$\|T_{K_{\mathrm{Sob}}^r}\|_{\mathrm{HS}}^2 \quad = \quad \sum_{k \geq 1} (\lambda_k^r)^2 \quad \geq \quad \sum_k \|g_k^r\|_{\mathrm{L}^2}^4 \ , \tag{7.10}$$

where $(\lambda_k^r)_{k \geq 1}$ are the eigenvalues of $T_{K_{\mathrm{Sob}}^r}$. For each feature g_k^r, the computation of its L^2-norm allows to better understand its contribution to $\|T_{K_{\mathrm{Sob}}^r}\|_{\mathrm{HS}}^2$, and therefore its influence on the Mercer decomposition of K_{Sob}^r. For more details on how the above inequality is obtained, the reader is referred to Appendix F.5, where a proof is given in the general case of a Mercer kernel $K : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$.

7.3.1. *Calculation of* L^2*-norms*

The properties of Bernoulli polynomials and trigonometric functions allow to derive the L^2-norms of the features identified in Eq. (7.9). Of course, one immediately has $\|\tilde{c}_{2k}^r\|_{\mathrm{L}^2} = \|\tilde{s}_{2k}^r\|_{\mathrm{L}^2} = 1/(2k\pi)^r$. As regards the polynomial features $(\widetilde{B}_k)_{1 \leq k \leq r}$, their L^2-norms are expected to decrease extremely fast because of the presence of factorials in the denominators. In fact, the L^2-norms of Bernoulli polynomials are closely related to Bernoulli numbers (see Appendix A.1.7) and asymptotic results (see Appendix A.1.2) are available for them:

$$\|B_k\|_{\mathrm{L}^2}^2 = |B_{2k}(0)|\, \frac{(k!)^2}{(2k)!} \quad \text{with} \quad |B_{2k}(0)| \underset{k \to \infty}{\sim} 4\sqrt{k\pi} \left(\frac{k}{\pi\, e}\right)^{2k} .$$

Then, Stirling's approximation for factorials can be used to deduce an asymptotically-equivalent sequence:

$$k! \underset{k \to \infty}{\sim} \sqrt{2k\pi} \left(\frac{k}{e}\right)^k \quad \text{and thus} \quad \|\widetilde{B}_k\|_{\mathrm{L}^2} = \frac{\|B_k\|_{\mathrm{L}^2}}{k!} = \left(\frac{|B_{2k}(0)|}{(2k)!}\right)^{1/2} \underset{k \to \infty}{\sim} \frac{\sqrt{2}}{(2\pi)^k} \ . \tag{7.11}$$

The decay speed is therefore much lower than first thought. It can be checked numerically that the asymptotic approximation of $\|\widetilde{B}_k\|_{\mathrm{L}^2}$ is actually pretty accurate for small values of k (including $k = 1$). Hence, the less influential polynomial feature (*i.e.* $\widetilde{B}_r$) and the most influential pair of sinusoidal features (*i.e.* $\tilde{c}_2^r$ and $\tilde{s}_2^r$) are weighted the same:

$$\|\widetilde{B}_r\|_{\mathrm{L}^2} \approx \sqrt{2}/(2\pi)^r \quad \text{and} \quad \|\tilde{c}_2^r\|_{\mathrm{L}^2} = \|\tilde{s}_2^r\|_{\mathrm{L}^2} = 1/(2\pi)^r \ .$$

This yields $\|\widetilde{B}_r\|_{\mathrm{L}^2}^2 \approx \|\tilde{c}_2^r\|_{\mathrm{L}^2}^2 + \|\tilde{s}_2^r\|_{\mathrm{L}^2}^2$, which strengthens the idea of a competition between $\widetilde{B}_r$ and $(\tilde{c}_2^r, \tilde{s}_2^r)$.

7.3.2. *Application to the Sobolev kernel* k_{Sob}^1

For $r = 1$, $\widetilde{B}_1$ is the unique existing polynomial feature and it is facing the entire collection of sinusoidal features $\{(\tilde{c}_{2k}^1)_{k \geq 1}\,; (\tilde{s}_{2k}^1)_{k \geq 1}\}$. In particular, the pair $(\tilde{c}_2^1, \tilde{s}_2^1)$ is directly challenging $\widetilde{B}_1$ in terms of L^2-norm. Because of this competition, the joint action of polynomial and sinusoidal features generates a specific collection of Mercer features, namely the collection of sinusoidal features identified in Theorem 5.1.

Remark 7.4. If the Mercer expansion of k^1_{Sob} had not be provided by Theorem 5.1, it could also have been obtained only with the tools from Section 7. The trick consists in rewriting k^1_{Sob} in the following way:

$$k^1_{\text{Sob}}(x,x') = B_1(x)\,B_1(x') + \frac{1}{2}B_2(|x-x'|) = \frac{1}{2}\left[x^2+(x')^2\right] - \max(x,x') + \frac{1}{3} = B_2\left(\frac{|x-x'|}{2}\right) + B_2\left(\frac{x+x'}{2}\right).$$

Then, Eq. (7.1) applied to B_2 provides:

$$k^1_{\text{Sob}}(x,x') = 4\sum_{k=1}^{\infty}\frac{\cos\left[k\pi(x-x')\right]}{(2k\pi)^2} + 4\sum_{k=1}^{\infty}\frac{\cos\left[k\pi(x+x')\right]}{(2k\pi)^2} = \sum_{k=1}^{\infty}\frac{c_k(x)\,c_k(x')}{(k\pi)^2}\ ,$$

and this offers a third manner of accessing the Mercer decomposition of k^1_{Sob}.

7.3.3. *Application to the Sobolev kernels k^r_{Sob} with $r \geq 2$*

For $r \geq 2$, the competition between $\widetilde{B}_r$ and $(\tilde{c}^{\,r}_2, \tilde{s}^{\,r}_2)$ is still existing but it is overwhelmed by the lower-degree polynomial features $\widetilde{B}_1, \dots, \widetilde{B}_{r-1}$ which have much larger L^2-norms. Since a small number of polynomial features contribute almost exclusively to $\|T_{K^r_{\text{Sob}}}\|^2_{\text{HS}}$, it is quite natural that the Mercer decomposition of k^r_{Sob} involves eigenfunctions which have much in common with the shifted Legendre polynomials (see Appendix A.2). To go even further, it should be emphasized that the L^2-norm of $\widetilde{B}_1$ is the largest by far. This is consistent with the conclusion stated in Section 5.2 where it was pointed out that $B_1 \otimes B_1$ is the leading term in k^r_{Sob} (as soon as $r \geq 2$).

In this section, the objective was to identify a fully analytical and easily interpretable feature map. This was achieved by rewriting k^r_{Sob} as the sum of the two sub-kernels. The unified feature map ψ^r_{Sob} arising from this approach is composed of non-orthogonal features (in the L^2-sense) but has the tremendous advantage of clearly establishing the coexistence of polynomial and sinusoidal features. The balance between both types of features is ruled by the smoothness parameter r. For $r = 1$, there is a true competition and this gives birth to a purely sinusoidal Mercer feature map. For $r \geq 2$, the sinusoidal features are too heavily penalized and they become negligible in comparison with the polynomial features. Knowing this, one may think that the sinusoidal features will vanish asymptotically (*i.e.* as $r \to \infty$). This intuition will be demonstrated in Section 8.

8. A glance at the asymptotic feature maps

The objective of this section is to understand how the behavior of Sobolev kernels $(K^r_{\text{Sob}})_{r\geq 1}$ evolves when the smoothness parameter $r \geq 1$ becomes extremely large. Findings from Section 7 indicate that the influence of sinusoidal features vanishes as r increases. This trend suggests that there will be no sinusoidal feature asymptotically. This intuition has now to be formalized theoretically.

In Section 8.1, it is shown that the sequence of Sobolev kernels converges uniformly to a limit kernel. The analytical expression of this kernel allows to identify an asymptotic feature map and to derive an ONB of the limit RKHS. In Section 8.2, further attention is paid to the asymptotic Mercer feature map. Once again, the shifted Legendre polynomials are proved not to be the unknown eigenfunctions.

8.1. **Limit Sobolev kernel and associated RKHS**

Since Bernoulli polynomials are continuous functions on $[0\,,1]$, they are bounded and attain their bounds:

$$\forall\, n \geq 1, \quad -\infty < m_n := \min_{x\in[0\,,1]} B_n(x) \leq \max_{x\in[0\,,1]} B_n(x) =: M_n < +\infty\ . \tag{8.1}$$

Explicit formulas were provided by [71] in order to compute (or at least approximate) those bounds:

$$\forall\, n \geq 1, \quad \begin{cases} m_n = \dfrac{2\,\zeta(n)\,n!}{(2\pi)^n} & \text{if } n \equiv 0 \pmod 4 \\[2ex] m_n > -\dfrac{2\,n!}{(2\pi)^n} & \text{otherwise} \end{cases} \quad \text{and} \quad \begin{cases} M_n = \dfrac{2\,\zeta(n)\,n!}{(2\pi)^n} & \text{if } n \equiv 2 \pmod 4 \\[2ex] M_n < \dfrac{2\,n!}{(2\pi)^n} & \text{otherwise} \end{cases} , \tag{8.2}$$

where ζ is the Riemann zeta function:

$$\begin{array}{rccl} \zeta: & \mathbb{N}\setminus\{0,1\} & \longrightarrow & \mathbb{R} \\ & n & \longmapsto & \zeta(n) := \displaystyle\sum_{k=1}^{\infty} \frac{1}{k^n} \end{array} .$$

As ζ is decreasing on $\mathbb{N}\setminus\{0,1\}$, one has $1 \leq \zeta(n) \leq \zeta(2) = \pi^2/6 \leq 2$ and this leads to:

$$\forall\, n \geq 1, \quad \|B_n\|_\infty = \max_{x\in[0\,,1]} |B_n(x)| =: M_n^+ \leq \frac{2\,\zeta(n)\,n!}{(2\pi)^n} \leq \frac{4\,n!}{(2\pi)^n} . \tag{8.3}$$

The above inequality may be seen as a slightly relaxed summary of Eq. (8.1) and (8.2) obtained after taking a loose bound of $\zeta(n)$. The term on the right-hand side is not the tightest possible upper bound for M_n^+ but it is quite sufficient to prove the convergence results presented below.

Proposition 8.1. *For any $r \geq 1$, let k_A^r and k_B^r be the kernels introduced in Section 7.1 to decompose k_{Sob}^r.*

(a) The kernel sequence $(k_A^r)_{r\geq 1}$ converges uniformly to the continuous kernel k_A^∞ defined by:

$$\forall\, x, x' \in [0\,,1], \quad k_A^\infty(x,x') := \lim_{r\to\infty} k_A^r(x,x') = \sum_{i=1}^{\infty} \frac{B_i(x)\,B_i(x')}{(i!)^2} . \tag{8.4}$$

(b) The kernel sequence $(k_B^r)_{r\geq 1}$ converges uniformly to the zero kernel $k_B^\infty := \lim_{r\to\infty} k_B^r = 0$.

The reader is referred to Appendix F.6 for the detailed proof. This proposition allows to define the limit Sobolev kernel as $K_{\text{Sob}}^\infty := 1 + k_{\text{Sob}}^\infty = 1 + k_A^\infty$. The related RKHS is denoted by $\mathcal{H}_{\text{Sob}}^\infty = \mathbb{R} \oplus \mathcal{F}_{\text{Sob}}^\infty$ since k_{Sob}^∞ remains an orthogonal kernel.

As Bernoulli polynomials are not mutually orthogonal in $\mathrm{L}^2([0\,,1])$, the closed-form expression in Eq. (8.4) is not the Mercer expansion of k_{Sob}^∞. However, a feature map $\psi_{\text{Sob}}^\infty : [0\,,1] \to \ell^2(\mathbb{N}^*)$ can still be identified:

$$k_{\text{Sob}}^\infty(x,x') = \langle \psi_{\text{Sob}}^r(x), \psi_{\text{Sob}}^r(x') \rangle_{\mathbb{R}^r} \quad \text{with} \quad \psi_{\text{Sob}}^r(x) = \left(\widetilde{B}_k(x)\right)_{k\geq 1} . \tag{8.5}$$

It is composed of an infinite number of polynomial features having increasing degrees. ψ_{Sob}^∞ must be regarded as the infinite-dimensional generalization of the feature map ψ_A^r extracted from the finite-rank kernel k_A^r. In the spirit of what was done in Section 7.2.4 with the polynomial and sinusoidal features extracted from k_{Sob}^r, the question here is whether the polynomial features $(\widetilde{B}_k)_{k\geq 0}$ are the basis functions of the limit RKHS.

Proposition 8.2. *The system $(\widetilde{B}_k)_{k\geq 0}$ is ℓ^2-linearly independent.*

The reader is referred to Appendix F.7 for the detailed proof. Theorem 2.23 can then be applied seamlessly to K_{Sob}^∞ in order to show that the polynomial features $(\widetilde{B}_k)_{k\geq 0}$ form an ONB of $\mathcal{H}_{\text{Sob}}^\infty$.

Remark 8.3. If a sequence of kernels $(K_n)_{n\geq 1}$ converges pointwise, the limit function K_∞ is also a kernel [25] (see Corollary 4.17, p. 119). However, the characterization of the limit RKHS is in general a delicate issue [5] (see Section 9, pp. 362–368). In the case of $K_{\mathrm{Sob}}^{\infty}$, the feature-based formalism proposed in Theorem 2.23 offers a very convenient way to describe $\mathcal{H}_{\mathrm{Sob}}^{\infty}$. In particular, the function space can be expressed as:

$$\mathcal{H}_{\mathrm{Sob}}^{\infty} = \left\{ h \in \mathbb{R}^{[0,1]} \,\middle|\, h(\cdot) = \sum_{k=0}^{\infty} a_k \frac{B_k(\cdot)}{k!} \text{ with } (a_k)_{k\geq 0} \in \ell^2(\mathbb{N}) \right\}.$$

The feature-based viewpoint can also be used to write $\langle \cdot, \cdot \rangle_{\mathcal{H}_{\mathrm{Sob}}^{\infty}}$. It would have been much more difficult to obtain the same results with the theory of Sobolev spaces of infinite order [39, 40].

This first section has allowed to lay the foundations of the asymptotic framework. The limit Sobolev kernel $K_{\mathrm{Sob}}^{\infty}$ has been identified and its properties have been carefully studied. In the next section, the connections between the Mercer feature map of $K_{\mathrm{Sob}}^{\infty}$ and the shifted Legendre polynomials are investigated.

8.2. **Asymptotic Mercer feature maps**

The question is whether the Mercer decomposition of $k_{\mathrm{Sob}}^{\infty}$ has a closed-form expression relying on basic math functions. Since $k_{\mathrm{Sob}}^{\infty} = k_A^{\infty}$ is the limit kernel of the sequence $(k_A^r)_{r\geq 1}$, the Mercer expansion of these kernels could help guess the asymptotic Mercer decomposition. Unfortunately, it was explained in Section 7.2.1 that the eigenvalue problem related to $T_{k_A^r}$ is more or less difficult to solve depending on the value of r. The resolution is trivial for $r \in \{1, 2\}$ because the kernel definitions $k_A^1 = \widetilde{B}_1 \otimes \widetilde{B}_1$ and $k_A^2 = \widetilde{B}_1 \otimes \widetilde{B}_1 + \widetilde{B}_2 \otimes \widetilde{B}_2$ are almost Mercer representations (up to the normalization of $\widetilde{B}_1$ and $\widetilde{B}_2$). On the contrary, the eigenvalue equation is much more difficult to solve when $r \geq 3$.

- For $r = 3$, the Mercer decomposition of k_A^3 can be obtained after long and exhausting hand calculations. The three resulting eigenfunctions are polynomials (which is quite normal since the RKHS induced by k_A^r is the space of all zero-mean polynomials with degree at most 3) but they are not P_1, P_2 and P_3.
- For $r \geq 4$, there is no point in trying to apply a brute-force approach based on hand calculations.

Therefore, the Mercer decomposition of $k_{\mathrm{Sob}}^{\infty}$ cannot be derived from a general result on the Mercer decomposition of k_A^r. A simple method in order to obtain a definitive answer is to compute $T_{k_{\mathrm{Sob}}^{\infty}} P_1$ and to prove that it cannot be proportional to P_1. For the sake of convenience, the integral calculation is carried out with $B_1 \propto P_1$:

$$\begin{aligned}\forall x \in [0,1],\ \left[T_{k_{\mathrm{Sob}}^{\infty}} B_1\right](x) &= \int_0^1 k_{\mathrm{Sob}}^{\infty}(x,\xi)\, B_1(\xi)\, \mathrm{d}\xi = \int_0^1 \xi\, k_{\mathrm{Sob}}^{\infty}(x,\xi)\, \mathrm{d}\xi = \sum_{k=1}^{\infty} \left(\int_0^1 \xi\, \widetilde{B}_k(\xi)\, \mathrm{d}\xi \right) \widetilde{B}_k(x) \\ &= \sum_{k=1}^{\infty} (-1)^{k+1}\, \widetilde{B}_{k+1}(0)\, \widetilde{B}_k(x) = \frac{1}{12}\, \widetilde{B}_1(x) + \sum_{k=1}^{\infty} \widetilde{B}_{2k}(0)\, \widetilde{B}_{2k+1}(x)\ .\end{aligned}$$

An integration by parts is used to switch from the first line to the second line. Then, the properties of Bernoulli numbers (see Appendix A.1.2) enable simplifications leading to the final expression. According to Proposition 8.2, the polynomials $(\widetilde{B}_k)_{k\geq 1}$ are ℓ^2-linearly independent. This allows to prove that $T_{k_{\mathrm{Sob}}^{\infty}} B_1$ and B_1 cannot be proportional. As a consequence, B_1 is not an eigenfunction of $T_{k_{\mathrm{Sob}}^{\infty}}$ and neither is P_1. With the same technique but longer calculations, it could be proved that P_2 is not an eigenfunction of $T_{k_{\mathrm{Sob}}^{\infty}}$, and so on.

Once again, the shifted Legendre polynomials $(P_k)_{k\geq 1}$ appear not to be the eigenfunctions of the integral operator under study. Thus, the asymptotic framework does not explain why those polynomials provide so accurate approximations of the eigenfunctions involved in the Mercer expansions of the kernels $(k_A^r)_{r\geq 1}$ and $(k_{\mathrm{Sob}}^r)_{r\geq 2}$. Identifying the Mercer expansion of $k_{\mathrm{Sob}}^{\infty}$ thus remains an open question.

9. Conclusion

The declared objective of this work was to shed light on how the new generation of HSIC indices (called HSIC-ANOVA indices and based on the use of Sobolev kernels) measures sensitivity. More specifically, as the knowledge of the feature maps is the key to identify the non-linear transformations applied to the input and output variables, the most expected achievement was the extraction of at least one fully explicit feature map from each (unanchored) Sobolev kernel.

Three strategies have been explored: (i) KFA, (ii) the differential approach and (iii) the decomposition into sub-kernels. It was found that their conclusions overlap on many points, while remaining highly complementary in the information they provide. Regarding how the feature maps of Sobolev kernels change with r, the three approaches agree that only two cases need to be distinguished (namely $r = 1$ vs. $r \geq 2$). The Sobolev kernel K^1_{Sob} is only composed of purely sinusoidal features. The eigenfunctions are cosine functions and the decay rate of the eigenvalues is $1/k^2$. Hence, if K^1_{Sob} is assigned to all input variables, the information captured by the first-order HSIC-ANOVA indices is now transparent:

$$S_i^{\text{HSIC}} \propto \text{HSIC}(X_i, Y) = \sum_{k=1}^{\infty} \sum_{l} \frac{1}{(k\pi)^2} \left| \text{Cov}\left(c_k(X_i), v_l(Y)\right) \right|^2 \quad \text{with} \quad \begin{cases} \forall\, k \geq 1, & c_k(x_i) = \sqrt{2}\cos(k\pi x_i)\ , \\ (v_l)_l & \text{an ONB of } \mathcal{H}_Y\ . \end{cases}$$

In contrast, if $r \geq 2$, K^r_{Sob} consists of polynomial-like features. The eigenfunctions estimated with the KFA method look like the shifted Legendre polynomials and the eigenvalues decrease much faster. From a numerical standpoint, K^2_{Sob} and the following Sobolev kernels behave almost like the dot-product kernel K_{lin} because their Mercer feature maps are characterized by the predominance of one single linear feature. If K^2_{Sob} is used to implement the HSIC-ANOVA decomposition, the information captured by the first-order indices is now:

$$S_i^{\text{HSIC}} \propto \text{HSIC}(X_i, Y) \approx \sum_{l} \left| \text{Cov}\left(B_1(X_i), v_l(Y)\right) \right|^2 \quad \text{with} \quad \begin{cases} & B_1(x_i) = x_i - \frac{1}{2}\ , \\ (v_l)_l & \text{an ONB of } \mathcal{H}_Y\ . \end{cases}$$

An important lesson is therefore to avoid using a Sobolev kernel K^r_{Sob} of order $r \geq 2$ within the HSIC. Even though all Sobolev kernels are characteristic, the computation of HSIC-ANOVA indices must be restricted to K^1_{Sob}. In a more general perspective, this prompts to question the importance of using characteristic kernels in GSA. Of course, characteristic kernels offer theoretical guarantees regarding the ability of the HSIC to characterize independence. However, in practice, this is not sufficient to ensure an efficient detection of statistical dependence. Using a characteristic kernel is a good practice but investigating the macroscopic kernel behavior is just as important, if not more so.

At the end of this work, one can say that most of the questions raised by the use of Sobolev kernels received a satisfactory answer. In light of the numerical and theoretical results, K^1_{Sob} appears to be a nice kernel that should be trusted to implement the HSIC-ANOVA decomposition. As the first-order HSIC-ANOVA indices are sufficient to characterize independence (see Remark 4.5), one may wonder about the added value of the total-order indices. This question echoes the numerical study proposed in [30] where the first-order and total-order HSIC-ANOVA indices are almost equal in all presented test cases. The very nature of the additional information captured by higher-order indices will be investigated in future works.

Acknowledgments

This research work is part of the SAMOURAI[10] project funded by the French National Research Agency (ANR-20-CE46-0013). This financial support is gratefully acknowledged. The authors would also like to thank

[10] Simulation Analytics and Meta-model-based solutions for Optimization, Uncertainty and Reliability AnalysIs.

Olivier Roustant[11], Anthony Nouy[12] and Luc Pronzato[13] for sharing their expertise on the theory of reproducing kernel Hilbert spaces.

[11]INSA Toulouse, Institut Mathématiques de Toulouse, Université Toulouse III - Paul Sabatier, France.
[12]École Centrale Nantes, Nantes Université, Laboratoire de Mathématiques Jean Leray, UMR CNRS 6629, France.
[13]Laboratoire I3S, Université Côte d'Azur - CNRS, Sophia Antipolis, France.

References

[1] M. Abramowitz and I. A. Stegun. *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*, volume 55. US Government printing office, 1964.
[2] R. A. Adams and J. J. Fournier. *Sobolev spaces*. Elsevier, 2003.
[3] M. Albert, B. Laurent, A. Marrel, and A. Meynaoui. Adaptive test of independence based on HSIC measures. *The Annals of Statistics*, 50(2):858–879, 2022.
[4] G. Archer, A. Saltelli, and I. M. Sobol. Sensitivity measures, ANOVA-like techniques and the use of bootstrap. *Journal of Statistical Computation and Simulation*, 58(2):99–120, 1997.
[5] N. Aronszajn. Theory of reproducing kernels. *Transactions of the American Mathematical Society*, 68(3):337–404, 1950.
[6] F. Bach. On the equivalence between kernel quadrature rules and random feature expansions. *The Journal of Machine Learning Research*, 18(1):714–751, 2017.
[7] F. Bach and M. Jordan. Learning spectral clustering. *Advances in Neural Information Processing Systems*, 16, 2003.
[8] F. R. Bach and M. I. Jordan. Kernel independent component analysis. *Journal of Machine Learning Research*, 3(Jul):1–48, 2002.
[9] C. R. Baker. Joint measures and cross-covariance operators. *Transactions of the American Mathematical Society*, 186:273–289, 1973.
[10] J. Barr and H. Rabitz. A generalized kernel method for global sensitivity analysis. *SIAM/ASA Journal on Uncertainty Quantification*, 10(1):27–54, 2022.
[11] J. Barr and H. Rabitz. Kernel-based global sensitivity analysis obtained from a single data set. *Reliability Engineering & System Safety*, 235:109173, 2023.
[12] A. Berlinet and C. Thomas-Agnan. *Reproducing Kernel Hilbert Spaces in Probability and Statistics*. Springer Science & Business Media, 2011.
[13] M. S. Birman and M. Z. Solomyak. Piecewise-polynomial approximations of functions of the classes W_p^α. *Matematicheskii Sbornik*, 115(3):331–355, 1967.
[14] M. S. Birman and M. Z. Solomyak. Estimates of singular numbers of integral operators. *Russian Mathematical Surveys*, 32(1):15, 1977.
[15] E. Borgonovo. A new uncertainty importance measure. *Reliability Engineering & System Safety*, 92(6):771–784, 2007.
[16] E. Borgonovo. *Sensitivity Analysis: an Introduction for the Management Scientist*. International Series in Operations Research and Management Science. Springer Cham, 2017.
[17] E. Borgonovo, G. B. Hazen, and E. Plischke. A common rationale for global sensitivity measures and their estimation. *Risk Analysis*, 36(10):1871–1895, 2016.
[18] E. Borgonovo and E. Plischke. Sensitivity analysis: a review of recent advances. *European Journal of Operational Research*, 248(3):869–887, 2016.
[19] H. Brezis. *Functional Analysis, Sobolev Spaces and Partial Differential Equations*, volume 2. Springer, 2011.
[20] F.-X. Briol, C. Oates, M. Girolami, and M. A. Osborne. Frank-Wolfe Bayesian quadrature: probabilistic integration with theoretical guarantees. *Advances in Neural Information Processing Systems*, 28, 2015.
[21] C. Brislawn. Kernels of trace class operators. *Proceedings of the American Mathematical Society*, 104(4):1181–1190, 1988.
[22] C. Campbell. Kernel methods: a survey of current techniques. *Neurocomputing*, 48(1-4):63–84, 2002.
[23] C. Campbell and K. Bennett. A linear programming approach to novelty detection. *Advances in Neural Information Processing Systems*, 13, 2000.
[24] G. Chastaing, F. Gamboa, and C. Prieur. Generalized Sobol' sensitivity indices for dependent variables: Numerical methods. *Journal of statistical computation and simulation*, 85(7):1306–1333, 2015.
[25] A. Christmann and I. Steinwart. Kernels and reproducing kernel Hilbert spaces. In *Support Vector Machines*, pages 111–164. Springer, 2008.
[26] K. Chwialkowski and A. Gretton. A kernel independence test for random processes. In *International Conference on Machine Learning*, pages 1422–1430. PMLR, 2014.
[27] S. Corlay and G. Pagès. Functional quantization-based stratified sampling methods. *Monte Carlo Methods and Applications*, 21(1):1–32, 2015.
[28] P. Craven and G. Wahba. Smoothing noisy data with spline functions: estimating the correct degree of smoothing by the method of generalized cross-validation. *Numerische Mathematik*, 31(4):377–403, 1978.
[29] S. Da Veiga. Global sensitivity analysis with dependence measures. *Journal of Statistical Computation and Simulation*, 85(7):1283–1305, 2015.
[30] S. Da Veiga. Kernel-based ANOVA decomposition and Shapley effects – Application to global sensitivity analysis, 2021. Preprint available at `https://hal.archives-ouvertes.fr/hal-03108628`.
[31] S. Da Veiga, F. Gamboa, B. Iooss, and C. Prieur. *Basics and Trends in Sensitivity Analysis: Theory and Practice in R*. Society for Industrial and Applied Mathematics, 2021.
[32] M. De Lozzo and A. Marrel. New improvements in the use of dependence measures for sensitivity analysis and screening. *Journal of Statistical Computation and Simulation*, 86(15):3038–3058, 2016.

[33] P. Derennes, J. Morio, and F. Simatos. A nonparametric importance sampling estimator for moment independent importance measures. *Reliability Engineering & System Safety*, 187:3–16, 2019.

[34] I. S. Dhillon, Y. Guan, and B. Kulis. Kernel k-means: spectral clustering and normalized cuts. In *Proceedings of the 10th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 551–556, 2004.

[35] J. Dick, A. Hinrichs, and F. Pillichshammer. Proof techniques in quasi-Monte Carlo theory. *Journal of Complexity*, 31(3):327–371, 2015.

[36] J. Dick, P. Kritzer, F. Pillichshammer, and H. Woźniakowski. Approximation of analytic functions in Korobov spaces. *Journal of Complexity*, 30(2):2–28, 2014.

[37] J. Dick, D. Nuyens, and F. Pillichshammer. Lattice rules for nonperiodic smooth integrands. *Numerische Mathematik*, 126(2):259–291, 2014.

[38] J. Dick, I. H. Sloan, X. Wang, and H. Woźniakowski. Liberating the weights. *Journal of Complexity*, 20(5):593–623, 2004.

[39] J. A. Dubinskii. *Sobolev Spaces of Infinite Order and Differential Equations*, volume 3 of *Mathematics and its Applications. East European Series*. Springer Science & Business Media, 1986.

[40] Y. A. Dubinskii. Sobolev spaces of infinite order. *Russian Mathematical Surveys*, 46(6):107, 1991.

[41] M. Duc-Jacquet. *Approximation des fonctionnelles linéaires sur les espaces Hilbertiens autoreproduisants*. PhD thesis, Université Joseph-Fourier-Grenoble I, 1973.

[42] N. Durrande, D. Ginsbourger, O. Roustant, and L. Carraro. ANOVA kernels and RKHS of zero mean functions for model-based sensitivity analysis. *Journal of Multivariate Analysis*, 115:57–67, 2013.

[43] M. R. El Amri and A. Marrel. More powerful HSIC-based independence tests, extension to space-filling designs and functional data, Oct. 2021. Preprint available at `https://hal-cea.archives-ouvertes.fr/cea-03406956`.

[44] M. R. El Amri and A. Marrel. Optimized HSIC-based tests for sensitivity analysis: Application to thermalhydraulic simulation of accidental scenario on nuclear reactor. *Quality and Reliability Engineering International*, 38(3):1386–1403, 2022.

[45] N. Fellmann, C. Blanchet-Scalliet, C. Helbert, A. Spagnol, and D. Sinoquet. Kernel-based sensivity analysis for (excursion) sets, 2023. arXiv preprint arXiv:2305.09268.

[46] C. Fowlkes, S. Belongie, F. Chung, and J. Malik. Spectral grouping using the Nyström method. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 26(2):214–225, 2004.

[47] F. Gamboa, P. Gremaud, T. Klein, and A. Lagnoux. Global sensitivity analysis: a novel generation of mighty estimators based on rank statistics. *Bernoulli*, 28(4):2345–2374, 2022.

[48] R. G. Ghanem and P. D. Spanos. *Stochastic Finite Elements: a Spectral Approach*. Courier Corporation, 2003.

[49] D. Ginsbourger, O. Roustant, D. Schuhmacher, N. Durrande, and N. Lenz. On ANOVA decompositions of kernels and Gaussian random field paths. In *Monte Carlo and Quasi-Monte Carlo Methods*, pages 315–330. Springer, 2016.

[50] A. Gretton, K. Borgwardt, M. Rasch, B. Schölkopf, and A. Smola. A kernel method for the two-sample-problem. *Advances in Neural Information Processing Systems*, 19:513–520, 2006.

[51] A. Gretton, K. M. Borgwardt, M. J. Rasch, B. Schölkopf, and A. Smola. A kernel two-sample test. *The Journal of Machine Learning Research*, 13(1):723–773, 2012.

[52] A. Gretton, O. Bousquet, A. Smola, and B. Schölkopf. Measuring statistical dependence with Hilbert-Schmidt norms. In *International Conference on Algorithmic Learning Theory*, volume 20, pages 63–77. Springer, 2005.

[53] A. Gretton, K. Fukumizu, C. H. Teo, L. Song, B. Schölkopf, and A. J. Smola. A kernel statistical test of independence. In *Proceedings of the 20th International Conference on Neural Information Processing Systems*, pages 585–592, 2007.

[54] A. Gretton, R. Herbrich, A. Smola, O. Bousquet, and B. Schölkopf. Kernel methods for measuring independence. *Journal of Machine Learning Research*, 6:2075–2129, 2005.

[55] M. Griebel. Sparse grids for higher dimensional problems. In *Foundations of Computational Mathematics, Santander 2005*, London Mathematical Society Lecture Note Series, pages 106–161. Cambridge University Press, 2006.

[56] C. Gu. *Smoothing Spline ANOVA Models*, volume 297. Springer, 2013.

[57] T. Hastie, R. Tibshirani, J. H. Friedman, and J. H. Friedman. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, volume 2. Springer, 2009.

[58] D. Hawkins. Some practical problems in implementing a certain sieve estimator of the Gaussian mean function. *Communications in Statistics-Simulation and Computation*, 18(2):481–500, 1989.

[59] E. Hernández and G. Weiss. *A First Course on Wavelets*. CRC press, 1996.

[60] F. Hickernell, I. Sloan, and G. Wasilkowski. On tractability of weighted integration over bounded and unbounded regions in $\mathbb{R}^s$. *Mathematics of Computation*, 73(248):1885–1901, 2004.

[61] W. Hoeffding. A class of statistics with asymptotically normal distribution. *Breakthroughs in Statistics: Foundations and Basic Theory*, pages 308–334, 1992.

[62] T. Hofmann, B. Schölkopf, and A. J. Smola. Kernel methods in machine learning. *The Annals of Statistics*, 36(3):1171–1220, 2008.

[63] D. T. Hristopulos. *Random Fields for Spatial Data Modeling*. Springer, 2020.

[64] A. Janon, T. Klein, A. Lagnoux, M. Nodet, and C. Prieur. Asymptotic normality and efficiency of two Sobol index estimators. *ESAIM: Probability and Statistics*, 18:342–364, 2014.

[65] X. Jiang, Y. Motai, R. R. Snapp, and X. Zhu. Accelerated kernel feature analysis. In *2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'06)*, volume 1, pages 109–116. IEEE, 2006.
[66] M. Kanagawa, P. Hennig, D. Sejdinovic, and B. K. Sriperumbudur. Gaussian processes and kernel methods: a review on connections and equivalences, 2018. arXiv preprint arXiv:1807.02582.
[67] H. König. *Eigenvalue Distribution of Compact Operators*, volume 16. Birkhäuser, 2013.
[68] O. Kouba. Lecture notes: Bernoulli polynomials and applications, 2013. arXiv preprint arXiv:1309.7560.
[69] M. A. Krasnosel'skii, G. M. Vainikko, R. Zabreyko, Y. Ruticki, and V. Stet'senko. *Approximate Solution of Operator Equations*. Springer Science & Business Media, 2012.
[70] F. Kuo, I. Sloan, G. Wasilkowski, and H. Woźniakowski. On decompositions of multivariate functions. *Mathematics of Computation*, 79(270):953–966, 2010.
[71] D. H. Lehmer. On the maxima and minima of Bernoulli polynomials. *The American Mathematical Monthly*, 47(8):533–538, 1940.
[72] Z. Liang and Y. Lee. Eigen-analysis of nonlinear PCA with polynomial kernels. *Statistical Analysis and Data Mining: the ASA Data Science Journal*, 6(6):529–544, 2013.
[73] T. A. Mara and S. Tarantola. Variance-based sensitivity indices for models with dependent inputs. *Reliability Engineering & System Safety*, 107:115–121, 2012.
[74] A. Marrel, B. Iooss, and V. Chabridon. The ICSCREAM methodology: identification of penalizing configurations in computer experiments using screening and metamodel – Applications in thermal-hydraulics. *Nuclear Science and Engineering*, 196(3):301–321, 2022.
[75] M. Mutny and A. Krause. Experimental design for linear functionals in reproducing kernel Hilbert spaces. *Advances in Neural Information Processing Systems*, 35:20175–20188, 2022.
[76] E. Novak, M. Ullrich, H. Woźniakowski, and S. Zhang. Reproducing kernels of Sobolev spaces on $\mathbb{R}^d$ and applications to embedding constants and tractability. *Analysis and Applications*, 16(05):693–715, 2018.
[77] E. Novak and H. Woźniakowski. Intractability results for integration and discrepancy. *Journal of Complexity*, 17(2):388–441, 2001.
[78] P. Novello, T. Fel, and D. Vigouroux. Making sense of dependence: efficient black-box explanations using dependence measure, 2022. arXiv preprint arXiv:2206.06219.
[79] C. J. Oates. Minimum discrepancy methods in uncertainty quantification, 2021. arXiv preprint arXiv:2109.06075.
[80] A. B. Owen and C. Prieur. On Shapley value for measuring importance of dependent inputs. *SIAM/ASA Journal on Uncertainty Quantification*, 5(1):986–1002, 2017.
[81] S. Saliani. ℓ^2-linear independence for the system of integer translates of a square integrable function. *Proceedings of the American Mathematical Society*, 141(3):937–941, 2013.
[82] A. Saltelli. Making best use of model evaluations to compute sensitivity indices. *Computer Physics Communications*, 145(2):280–297, 2002.
[83] A. Saltelli, M. Ratto, T. Andres, F. Campolongo, J. Cariboni, D. Gatelli, M. Saisana, and S. Tarantola. *Sensitivity Analysis in Practice: a Guide to Assessing Scientific Models*. Wiley, 2004.
[84] A. Saltelli, M. Ratto, T. Andres, F. Campolongo, J. Cariboni, D. Gatelli, M. Saisana, and S. Tarantola. *Global Sensitivity Analysis: the Primer*. John Wiley & Sons, 2008.
[85] G. Sarazin, A. Marrel, S. Da Veiga, and V. Chabridon. Independence test based on total-order HSIC-ANOVA indices. In *53èmes Journées de Statistique de la SFdS*, 2022.
[86] C. Saunders. Ridge regression learning algorithm in dual variables. In *Proceedings of the 15th International Conference on Machine Learning*, 1998.
[87] H. Scheffe. *The Analysis of Variance*, volume 72. John Wiley & Sons, 1999.
[88] B. Schölkopf, R. Herbrich, and A. Smola. A generalized representer theorem. In *14th Annual Conference on Computational Learning Theory (COLT 2001) and 5th European Conference on Computational Learning Theory (EuroCOLT 2001)*, pages 416–426. Springer, 2001.
[89] B. Schölkopf, A. Smola, and K.-R. Müller. Nonlinear component analysis as a kernel eigenvalue problem. *Neural Computation*, 10(5):1299–1319, 1998.
[90] B. Schölkopf, A. J. Smola, and K.-R. Müller. Kernel principal component analysis. In *Advances in Kernel Methods: Support Vector Learning*, pages 327–352. Cambridge, MA: MIT Press, 1999.
[91] R. J. Serfling. *Approximation Theorems of Mathematical Statistics*. John Wiley & Sons, 2009.
[92] C.-J. Simon-Gabriel and B. Schölkopf. Kernel distribution embeddings: Universal kernels, characteristic kernels and kernel metrics on distributions. *The Journal of Machine Learning Research*, 19(1):1708–1736, 2018.
[93] I. H. Sloan. Iterated Galerkin method for eigenvalue problems. *SIAM Journal on Numerical Analysis*, 13(5):753–760, 1976.
[94] I. H. Sloan and H. Woźniakowski. When are quasi-Monte Carlo algorithms efficient for high dimensional integrals? *Journal of Complexity*, 14(1):1–33, 1998.
[95] A. J. Smola, O. L. Mangasarian, and B. Schölkopf. Sparse kernel feature analysis. In *Classification, Automation, and New Media*, pages 167–178. Springer, 2002.

[96] I. M. Sobol'. Sensitivity analysis for non-linear mathematical models. *Mathematical Modelling and Computational Experiment*, 1:407–414, 1993.

[97] I. M. Sobol'. Global sensitivity indices for nonlinear mathematical models and their Monte Carlo estimates. *Mathematics and Computers in Simulation*, 55(1-3):271–280, 2001.

[98] L. Song, A. Smola, A. Gretton, K. M. Borgwardt, and J. Bedo. Supervised feature selection via dependence estimation. In *Proceedings of the 24th International Conference on Machine Learning*, pages 823–830, 2007.

[99] A. Spence. On the convergence of the Nyström method for the integral equation eigenvalue problem. *Numerische Mathematik*, 25:57–66, 1975.

[100] A. Spence. Error bounds and estimates for eigenvalues of integral equations. *Numerische Mathematik*, 29(2):133–147, 1978.

[101] B. Sriperumbudur, K. Fukumizu, and G. Lanckriet. On the relation between universality, characteristic kernels and RKHS embedding of measures. In *Proceedings of the 13th International Conference on Artificial Intelligence and Statistics*, pages 773–780. JMLR Workshop and Conference Proceedings, 2010.

[102] B. K. Sriperumbudur, K. Fukumizu, A. Gretton, G. R. Lanckriet, and B. Schölkopf. Kernel choice and classifiability for RKHS embeddings of probability distributions. In *Proceedings of the 20th International Conference on Neural Information Processing Systems*, volume 22, pages 1750–1758, 2009.

[103] B. K. Sriperumbudur, A. Gretton, K. Fukumizu, B. Schölkopf, and G. R. Lanckriet. Hilbert space embeddings and metrics on probability measures. *The Journal of Machine Learning Research*, 11:1517–1561, 2010.

[104] M. Stitson, A. Gammerman, V. Vapnik, V. Vovk, C. Watkins, and J. Weston. Support vector regression with ANOVA decomposition kernels. *Advances in Kernel Methods: Support Vector Learning*, pages 285–292, 1999.

[105] G. P. Styan. Hadamard products and multivariate statistical analysis. *Linear Algebra and its Applications*, 6:217–240, 1973.

[106] Z. Szabó and B. K. Sriperumbudur. Characteristic and universal tensor product kernels. *Journal of Machine Learning Research*, 18:1–29, 2018.

[107] O. Teymur, J. Gorham, M. Riabiz, and C. Oates. Optimal quantisation of probability measures using maximum mean discrepancy. In *International Conference on Artificial Intelligence and Statistics*, pages 1027–1035. PMLR, 2021.

[108] C. Thomas-Agnan. Computing a family of reproducing kernels for statistical applications. *Numerical Algorithms*, 13:21–32, 1996.

[109] N. Vakhania, V. Tarieladze, and S. Chobanyan. *Probability Distributions on Banach Spaces*, volume 14. Springer Science & Business Media, 1987.

[110] J.-P. Vert, K. Tsuda, and B. Schölkopf. A primer on kernel methods. *Kernel Methods in Computational Biology*, 47:35–70, 2004.

[111] G. Wahba. Smoothing noisy data with spline functions. *Numerische mathematik*, 24(5):383–393, 1975.

[112] G. Wahba. *Spline Models for Observational Data*, volume 29 of *CBMS-NSF Regional Conference Series in Applied Mathematics*. Society for Industrial and Applied Mathematics, 1990.

[113] Y. Wang. *Smoothing Splines: Methods and Applications*. CRC press, 2011.

[114] G. W. Wasilkowski and H. Woźniakowski. Finite-order weights imply tractability of linear multivariate problems. *Journal of Approximation Theory*, 130(1):57–77, 2004.

[115] Y. Wei, F. Yang, and M. J. Wainwright. Early stopping for kernel boosting algorithms: a general analysis with localized complexities. *Advances in Neural Information Processing Systems*, 30, 2017.

[116] C. K. Williams and C. E. Rasmussen. *Gaussian Processes for Machine Learning*, volume 2 of *Adaptive Computation and Machine Learning series*. MIT Press, 2006.

[117] K. Zhang, J. Peters, D. Janzing, and B. Schölkopf. Kernel-based conditional independence test and application in causal discovery. In *Proceedings of the 27th Conference on Uncertainty in Artificial Intelligence*, pages 804–813, 2011.

[118] Q. Zhang, S. Filippi, A. Gretton, and D. Sejdinovic. Large-scale kernel methods for independence testing. *Statistics and Computing*, 28(1):113–130, 2018.

[119] Y. Zhang, J. Duchi, and M. Wainwright. Divide and conquer kernel ridge regression. In *Conference on Learning Theory*, pages 592–617. PMLR, 2013.

Appendix A. Remarkable properties of two families of polynomials

A.1. Bernoulli polynomials

This section brings together various results on Bernoulli polynomials which are notably used in the proofs of Appendix F. A similar table dedicated to Bernoulli polynomials can be found in [1] (see Section 23, pp. 804–808). For a much more comprehensive overview of existing results on Bernoulli polynomials, it is strongly recommended to have a look at [68].

A.1.1. *Explicit expressions for low degrees*

Bernoulli polynomials $(B_n)_{n\geq 0}$ are all defined on $[0\,,1]$. The analytical expressions of the first Bernoulli polynomials are detailed below:

- $B_0(x) \;=\; 1$;
- $B_1(x) \;=\; x - \dfrac{1}{2}$;
- $B_2(x) \;=\; x^2 - x + \dfrac{1}{6}$;
- $B_3(x) \;=\; x^3 - \dfrac{3}{2}\,x^2 + \dfrac{1}{2}$;
- $B_4(x) \;=\; x^4 - 2\,x^3 + x^2 - \dfrac{1}{30}$;
- $B_5(x) \;=\; x^5 - \dfrac{5}{2}\,x^4 + \dfrac{5}{3}\,x^3 - \dfrac{1}{6}\,x$;
- $B_6(x) \;=\; x^6 - 3\,x^5 + \dfrac{5}{2}\,x^4 - \dfrac{1}{2}\,x^2 + \dfrac{1}{42}$;
- $B_7(x) \;=\; x^7 - \dfrac{7}{2}\,x^6 + \dfrac{7}{2}\,x^5 - \dfrac{7}{6}\,x^3 + \dfrac{1}{6}\,x$.

A.1.2. *Bernoulli numbers*

For any $n \geq 0$, the value $B_n(0)$ is called the n-th Bernoulli number. Here is a list of the first ones:

$$B_0(0) = 1\ ; \quad B_1(0) = -\frac{1}{2}\ ; \quad B_2(0) = \frac{1}{6}\ ; \quad B_3(0) = 0\ ; \quad B_4(0) = -\frac{1}{30}\ ; \quad B_5(0) = 0\ ;$$
$$B_6(0) = \frac{1}{42}\ ; \quad B_7(0) = 0\ ; \quad B_8(0) = -\frac{1}{30}\ ; \quad B_9(0) = 0\ ; \quad B_{10}(0) = \frac{5}{66}\ ; \quad B_{11}(0) = 0\ \ldots$$

In particular, Bernoulli numbers verify the following properties:

- $\forall\, n \geq 1, \quad B_{2n+1}(0) = 0$;
- $\forall\, n \geq 1, \quad |B_{2n}(0)| = (-1)^{n+1} B_{2n}(0)$;
- $|B_{2n}(0)| \underset{n\to\infty}{\sim} 4\sqrt{n\pi}\left(\dfrac{n}{\pi\, e}\right)^{2n}$.

A.1.3. *Derivatives*

$$\forall\, n \geq 1, \quad B_n'(\cdot) = n\, B_{n-1}(\cdot)$$

A.1.4. *Symmetries and boundary values*

$$\forall\, n \geq 0, \quad \forall\, x \in [0\,,1], \quad B_n(1-x) = (-1)^n\, B_n(x)$$

This leads to distinguish three cases:

$$\begin{cases} \qquad\qquad B_1(1) = -B_1(0) = \dfrac{1}{2} \ ; \\ \forall\, n \geq 0, \quad B_{2n}(1) = B_{2n}(0) \neq 0 \ ; \\ \forall\, n \geq 1, \quad B_{2n+1}(1) = B_{2n+1}(0) = 0 \ . \end{cases}$$

For any $n \geq 2$, the polynomial function $B_n : [0\,,1] \to \mathbb{R}$ can be seen as the restriction of a continuous 1-periodic function defined on $\mathbb{R}$.

A.1.5. *Fourier series expansions*

Any Bernoulli polynomial (except B_1) is equal to its Fourier series on $[0\,,1]$. For B_1, the equality only holds on $(0\,,1)$ since B_1 cannot be extended as a continuous 1-periodic function. The complex and real versions of the Fourier series are respectively given by:

- $\forall\, 0 \leq x \leq 1, \quad B_0(x) = 1 \ ;$
- $\forall\, 0 < x < 1, \quad B_1(x) = -\dfrac{1}{2\pi i} \displaystyle\sum_{k \in \mathbb{Z}^*} \dfrac{e^{2\pi i k x}}{k} = -\sum_{k=1}^{\infty} \dfrac{\sin(2k\pi x)}{k\pi} \ ;$
- $\forall\, n \geq 2, \quad \forall\, 0 \leq x \leq 1, \quad B_n(x) = -\dfrac{n!}{(2\pi i)^n} \displaystyle\sum_{k \in \mathbb{Z}^*} \dfrac{e^{2\pi i k x}}{k^n} = (-2)\, n! \sum_{k=1}^{\infty} \dfrac{\cos(2k\pi x - \frac{n\pi}{2})}{(2k\pi)^n} \ .$

A.1.6. *Mean values*

$$\forall\, n \geq 1, \quad \int_0^1 B_n(x)\,\mathrm{d}x = 0$$

A.1.7. *Integrals*

- $\forall\,(i,j) \in \mathbb{N}^2, \quad \beta_{ij} := \langle B_i, B_j \rangle_{\mathrm{L}^2} = \displaystyle\int_0^1 B_i(x)\, B_j(x)\,\mathrm{d}x = (-1)^{i+1} \dfrac{i!\, j!}{(i+j)!}\, B_{i+j}(0) \ .$
- $\forall\, n \in \mathbb{N}, \quad \|B_n\|_{\mathrm{L}^2}^2 = \displaystyle\int_0^1 B_n(x)^2\,\mathrm{d}x = \dfrac{(n!)^2}{(2n)!}\, |B_{2n}(0)| \ .$

A.1.8. *Upper bound*

$$\forall\, n \geq 1, \quad \|B_n\|_\infty = \max_{x \in [0\,,1]} |B_n(x)| =: M_n^+ \leq \frac{4\, n!}{(2\pi)^n}$$

A.2. **Legendre polynomials**

A few elements about Legendre polynomials are introduced in this section. More details on orthogonal families of polynomials are provided in [1] (see pp. 773–780).

A.2.1. *Definition*

Definition A.1. There exists a unique family $(L_k)_{k \geq 0}$ of orthogonal polynomial functions in $\mathrm{L}^2([-1\,,1])$ such that $\deg(L_k) = k$ and $L_k(1) = 1$ for all $k \geq 0$. They are called *Legendre polynomials* and they form a complete and orthogonal system in $\mathrm{L}^2([-1\,,1])$.

Remark A.2. Legendre polynomials are not an ONB of $\mathrm{L}^2([-1\,,1])$ because they do not have unit norm:

$$\forall\, k \geq 0, \quad \|L_k\|_{\mathrm{L}^2}^2 = \int_{-1}^{1} L_k(t)^2\,\mathrm{d}t = \frac{2}{2k+1} \neq 1\ .$$

For any given interval $[a\,,b]$, an ONB of $\mathrm{L}^2([a\,,b])$ can be derived from Legendre polynomials. Such basis will be denoted by $(P_k)_{k\geq 1}$ in this work. It is simply obtained by rescaling and renormalizing Legendre polynomials:

$$\forall\, t \in [a\,,b], \quad \sigma(t) = \frac{2t-(a+b)}{b-a} \quad \text{and} \quad \forall\, k \geq 0, \quad P_k(t) := \frac{[L_k \circ \sigma](t)}{\|L_k \circ \sigma\|_{\mathrm{L}^2}} = \sqrt{\frac{2k+1}{b-a}}\, L_k\left(\frac{2t-(a+b)}{b-a}\right).$$

The shift function σ is a linear mapping from $[a\,,b]$ into $[-1\,,1]$ that helps go back to the standard definition domain of Legendre polynomials. Using σ allows to preserve orthogonality among the polynomials $(P_k)_{k\geq 1}$ while the multiplying factor $\|L_k \circ \sigma\|_{\mathrm{L}^2}^{-1} = \sqrt{(2k+1)/(b-a)}$ ensures that the polynomial functions $(P_k)_{k\geq 0}$ are normalized in $\mathrm{L}^2([a\,,b])$. In the specific case where $a=0$ and $b=1$, the polynomials $(P_k)_{k\geq 1}$ are called the *shifted Legendre polynomials*:

$$\forall t \in [0\,,1], \quad \forall k \geq 0, \quad P_k(t) := \sqrt{2k+1}\, L_k(2t-1).$$

A.2.2. *Explicit expressions for low degrees*

The very first Legendre and shifted Legendre polynomials are given below:

- $L_0(x) = 1$;
- $L_1(x) = x$;
- $L_2(x) = \frac{1}{2}\left(3\,x^2 - 1\right)$;
- $L_3(x) = \frac{1}{2}\left(5\,x^3 - 3\,x\right)$;
- $L_4(x) = \frac{1}{8}\left(35\,x^4 - 30\,x^2 + 3\right)$;

- $P_0(x) = 1$;
- $P_1(x) = 2\,x - 1$;
- $P_2(x) = 6\,x^2 - 6\,x + 1$;
- $P_3(x) = 20\,x^3 - 30\,x^2 + 12\,x - 1$;
- $P_4(x) = 70\,x^4 - 140\,x^3 + 90\,x^2 - 20\,x + 1$.

Appendix B. Examples of Mercer decompositions

B.1. Gaussian kernel

$$\boxed{\forall\, x, x' \in \mathbb{R}, \quad K_G(x,x') := \exp\left[-\frac{1}{2}\left(\frac{x-x'}{\gamma}\right)^2\right] \quad \text{with} \quad \gamma > 0} \quad \text{and} \quad \boxed{\nu \text{ is } \mathcal{N}(0,\sigma^2)}\ .$$

Eigenvalues and eigenfunctions:

$$\forall\, k \geq 0, \quad \lambda_k = \sqrt{\frac{2a}{A}}\, B^k \quad \text{and} \quad \phi_k(x) = \exp\left[-(c-a)\,x^2\right]\, H_{2k}\left(\sqrt{2c}\,x\right),$$

$$\text{with} \quad a = \frac{1}{4\,\sigma^2}\ ; \quad b = \frac{1}{2\,\gamma^2}\ ; \quad c = \sqrt{a^2 + 2\,ab}\ ; \quad A = a+b+c\ ; \quad B = \frac{b}{A}\ .$$

$(H_k)_{k\geq 1}$ are the Hermite polynomials [1] (see pages 773-780).

B.2. Laplace kernel

$$\boxed{\forall\, a > 0, \quad \forall\, x, x' \in [-a\,,a], \quad K_L(x,x') = \exp\left(-\frac{|x-x'|}{\gamma}\right) \quad \text{with} \quad \gamma > 0} \quad \text{and} \quad \boxed{\nu \text{ is } \mathcal{U}([-a\,,a])}\ .$$

Equivalent boundary value problem:

$$\text{For any } \lambda > 0, \quad T_{K_L}\,\phi = \lambda\,\phi \quad \Longleftrightarrow \quad \lambda\,\phi'' + w^2\,\phi = 0 \quad \text{with} \quad \begin{cases} \frac{1}{\gamma}\,\phi(a) + \phi'(a) = 0 \\ \frac{1}{\gamma}\,\phi(-a) - \phi'(-a) = 0 \end{cases}$$

$$\text{and} \quad w^2 = \frac{2\,\gamma - \lambda}{\lambda\,\gamma^2}\ .$$

Eigenvalues and eigenfunctions:

- $$\forall\, k \geq 1, \quad \lambda_{1k} = \frac{2\gamma}{1+\gamma^2 w_{1k}^2} \quad \text{and} \quad \phi_{1k}(x) = \frac{\cos(w_{1k}x)}{\sqrt{a + \dfrac{\sin(2w_{1k}a)}{2w_{1k}}}}\ ;$$
- $$\forall\, k \geq 1, \quad \lambda_{2k} = \frac{2\gamma}{1+\gamma^2 w_{2k}^2} \quad \text{and} \quad \phi_{2k}(x) = \frac{\sin(w_{2k}x)}{\sqrt{a - \dfrac{\sin(2w_{2k}a)}{2w_{2k}}}}\ .$$

The sequences $(w_{1k})_{k\geq 1}$ and $(w_{2k})_{k\geq 1}$ are the (unknown) solutions of the two following non-linear equations:

$$(E_1): \quad \frac{1}{\gamma} - w\,\tan(wa) = 0 \qquad \text{and} \qquad (E_2): \quad w + \frac{1}{\gamma}\tan(wa) = 0\ .$$

B.3. **Sobolev kernels**

B.3.1. *Covariance kernel of the Wiener process*

$$\boxed{\forall\, x, x' \in [0\,,1], \quad K_W(x,x') := \min(x,x')} \quad \text{and} \quad \boxed{\nu \text{ is } \mathcal{U}([0\,,1])}\ .$$

Eigenvalues and eigenfunctions:

$$\forall\, k \geq 1, \quad \lambda_k = \frac{1}{\left[\left(\frac{2k-1}{2}\right)\pi\right]^2} \quad \text{and} \quad \phi_k(x) = \sqrt{2}\,\sin\left[\left(\frac{2k-1}{2}\right)\pi x\right].$$

Equivalent boundary value problem:

$$\text{For any } \lambda > 0, \quad T_{K_W}\,\phi = \lambda\,\phi \quad \Longleftrightarrow \quad \lambda\,\phi'' + \phi = 0 \quad \text{with} \quad \begin{cases} \phi(0) = 0 \\ \phi'(1) = 0 \end{cases} .$$

B.3.2. *Covariance kernel of the Brownian bridge*

$$\boxed{\forall\, x, x' \in [0\,,1], \quad K_{\mathcal{B}}(x,x') := \min(x,x') - xx'} \quad \text{and} \quad \boxed{\nu \text{ is } \mathcal{U}([0\,,1])}\ .$$

Eigenvalues and eigenfunctions:

$$\forall\, k \geq 1, \quad \lambda_k = \frac{1}{(k\pi)^2} \quad \text{and} \quad \phi_k(x) = \sqrt{2}\,\sin(k\pi x)\ .$$

Equivalent boundary value problem:

$$\text{For any } \lambda > 0, \quad T_{K_{\mathcal{B}}}\,\phi = \lambda\,\phi \quad \Longleftrightarrow \quad \lambda\,\phi'' + \phi = 0 \quad \text{with} \quad \begin{cases} \phi(0) = 0 \\ \phi(1) = 0 \end{cases} .$$

B.3.3. *Reproducing kernel of the standard Sobolev space of order $r = 1$*

$$\boxed{\forall\, x, x' \in [0\,,1], \quad K^1(x,x') = \frac{2e}{e^2 - 1}\,\cosh\left[\,\min(x,x')\right]\,\cosh\left[1 - \max(x,x')\right]} \quad \text{and} \quad \boxed{\nu \text{ is } \mathcal{U}([0\,,1])}\ .$$

Eigenvalues and eigenfunctions:

- $\lambda_0 = 1 \quad \text{and} \quad \phi_0(x) = 1$;
- $\forall\, k \geq 1, \quad \lambda_k = \dfrac{1}{1 + (k\pi)^2} \quad \text{and} \quad \phi_k(x) = \sqrt{2}\,\cos(k\pi x)$.

Equivalent boundary value problem:

$$\text{For any } \lambda > 0, \quad T_{K^1}\,\phi = \lambda\,\phi \quad \Longleftrightarrow \quad \lambda\,\phi'' + (1-\lambda)\,\phi = 0 \quad \text{with} \quad \begin{cases} \phi'(0) = 0 \\ \phi'(1) = 0 \end{cases} .$$

B.3.4. *Reproducing kernel of the unanchored Sobolev space of order $r = 1$*

$$\boxed{\forall\, x, x' \in [0\,,1], \quad K^1_{\mathrm{Sob}}(x,x') = 1 + k^1_{\mathrm{Sob}}(x,x') = 1 + B_1(x)B_1(x') + \frac{1}{2}\, B_2\left(|x-x'|\right)} \quad \text{and} \quad \boxed{\nu \text{ is } \mathcal{U}([0\,,1])}\,.$$

Eigenvalues and eigenfunctions:

- $\lambda_0 = 1 \quad \text{and} \quad \phi_0(x) = 1 \quad ;$
- $\forall\, k \geq 1, \quad \lambda_k = \dfrac{1}{(k\pi)^2} \quad \text{and} \quad \phi_k(x) = \sqrt{2}\, \cos(k\pi x) \quad .$

Equivalent boundary value problem:

$$\text{For any } \lambda > 0, \quad T_{k^1_{\mathrm{Sob}}}\, \phi = \lambda\, \phi \quad \Longleftrightarrow \quad \lambda\, \phi'' + \phi = 0 \quad \text{with} \quad \begin{cases} \phi'(0) = 0 \\ \phi'(1) = 0 \end{cases} \quad .$$

B.3.5. *Reproducing kernel of the periodic Sobolev space of order $r \geq 1$*

$$\boxed{\forall\, x, x' \in [0\,,1], \quad K^r_{\mathrm{per}}(x,x') = 1 + k^r_{\mathrm{per}}(x,x') = 1 + \frac{(-1)^{r+1}}{(2r)!}\, B_{2r}(|x-x'|)} \quad \text{and} \quad \boxed{\nu \text{ is } \mathcal{U}([0\,,1])}\,.$$

Eigenvalues and eigenfunctions:

- $\lambda_0 = 1 \quad \text{and} \quad \phi_0(x) = 1 \quad ;$
- $\forall\, k \geq 1, \quad \lambda_{1k} = \dfrac{1}{(2k\pi)^{2r}} \quad \text{and} \quad \phi_{1k}(x) = \sqrt{2}\, \cos(2k\pi x) \quad ;$
- $\forall\, k \geq 1, \quad \lambda_{2k} = \dfrac{1}{(2k\pi)^{2r}} \quad \text{and} \quad \phi_{2k}(x) = \sqrt{2}\, \sin(2k\pi x) \quad .$

Equivalent boundary value problem:

$$\text{For any } \lambda > 0, \quad T_{k^r_{\mathrm{per}}}\, \phi = \lambda\, \phi \quad \Longleftrightarrow \quad \lambda\, \phi^{[2r]} + (-1)^{r+1}\phi = 0 \quad \text{with} \quad \begin{cases} \phi^{[k]}(0) = \phi^{[k]}(1) \\ \forall\, 0 \leq k \leq 2r-1 \end{cases} \quad .$$

B.4. **Related literature**

If additional details are sought, the reader is kindly referred to the following papers:

- For K_G: [116] (see pp. 97–98).
- For K_L: [48] (see pp. 29–32).
- For K_W: [58] (see pp. 487–492).
- For $K_{\mathcal{B}}$: [27] (see p. 7).
- For K^1: [41] (see p. IV.8) and [108] (see p. 27).
- For K^1_{Sob}: [36] (see pp. 9–10).
- For K^r_{per}: [111] (see p. 386).

Appendix C. Inference of kernel-based sensitivity measures

Let $g : \mathcal{X}_1 \times \ldots \mathcal{X}_d \to \mathcal{Y}$ be a numerical simulator. The input variables $X_1, \ldots, X_d$ and the output variable $Y := g(X_1, \ldots, X_d)$ are assumed to be scalar[14]. They are respectively assigned the kernels $K_1, \ldots, K_d$ and K_Y. In addition, for each pair $\boldsymbol{Z}_i := (X_i, Y)$ composed of one given input and the common output, the two marginal distributions and the joint distribution are respectively denoted by $\mathbb{P}_{X_i}$, $\mathbb{P}_Y$ and $\mathbb{P}_{X_iY}$.

C.1. **Kernel-based sensitivity measures**

C.1.1. *HSIC indices*

The most convenient way to write HSIC indices is to use their formula based on expectations [54]. In fact, the HSIC index for the pair (X_i, Y) may be expressed as:

$$\mathrm{HSIC}(X_i, Y) = \mathbb{E}\left[K_i(X_i, X_i')\, K_Y(Y, Y')\right] + \mathbb{E}\left[K_i(X_i, X_i')\right] \mathbb{E}\left[K_Y(Y, Y')\right] - 2\,\mathbb{E}\left[K_i(X_i, X_i')\, K_Y(Y, Y'')\right], \quad \text{(C.1)}$$

where $\boldsymbol{Z}_i = (X_i, Y)$, $\boldsymbol{Z}_i' := (X_i', Y')$ and $\boldsymbol{Z}_i'' := (X_i'', Y'')$ are three independent random pairs following the same bivariate distribution $\mathbb{P}_{X_iY}$.

C.1.2. *R^2-HSIC indices*

The normalized HSIC index (also called the R^2-HSIC index) between X_i and Y is defined in the same fashion as Pearson's correlation coefficient (with the HSIC replacing the covariance operator):

$$\mathcal{R}^2(X_i, Y) := \frac{\mathrm{HSIC}(X_i, Y)}{\sqrt{\mathrm{HSIC}(X_i, X_i)}\sqrt{\mathrm{HSIC}(Y, Y)}}\ . \quad \text{(C.2)}$$

Normalization has the desired effect since one has $0 \leq \mathcal{R}^2(X_i, Y) \leq 1$. However, the proof of this result is much more complicated than a simple application of the Cauchy-Schwarz inequality (see [31] for more details).

C.1.3. *HSIC-ANOVA indices*

Provided that the input and output kernels satisfy all the assumptions of Theorem 4.4, an ANOVA-like decomposition exists for HSIC indices [30]. In particular, the HSIC between $\boldsymbol{X} := [X_1, \ldots, X_d]$ and Y can be decomposed as follows:

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

This allows to define the first-order and total-order HSIC-ANOVA indices in a similar way as Sobol' indices:

$$S_i^{\mathrm{HSIC}} := \frac{\mathrm{HSIC}(X_i, Y)}{\mathrm{HSIC}(\boldsymbol{X}, Y)} \quad \text{and} \quad T_i^{\mathrm{HSIC}} := 1 - \frac{\mathrm{HSIC}(\boldsymbol{X}_{-i}, Y)}{\mathrm{HSIC}(\boldsymbol{X}, Y)} \quad \text{(C.3)}$$

where $\boldsymbol{X}_{-i} := [X_j]_{j \neq i}$. Moreover, after realizing that $\mathrm{HSIC}(\boldsymbol{X}, Y) - \mathrm{HSIC}(\boldsymbol{X}_{-i}, Y) = \sum_{\boldsymbol{u} \ni i} \mathrm{HSIC}(\boldsymbol{X}_{\boldsymbol{u}}, Y)$, it can be easily proved that $0 \leq S_i^{\mathrm{HSIC}} \leq T_i^{\mathrm{HSIC}} \leq 1$.

C.2. **Estimation of the HSIC**

Here, the objective is to show how to estimate the HSIC between an input variable $X \sim \mathbb{P}_X$ (equipped with the kernel K_X) and an output variable $Y \sim \mathbb{P}_Y$ (equipped with the kernel K_Y). Of course, everything can be generalized to a vector of input variables, thus allowing any of the kernel-based sensitivity measure defined in Section C.1 to be estimated.

[14] This choice is made for the sake of simplicity. However, note that everything remains true when the random objects live in (possibly different) separable spaces [52].

C.2.1. *Preliminary formalization*

After grouping all its terms within a single expectation, the HSIC becomes:

$$\begin{aligned}\mathrm{HSIC}(X,Y) &= \mathbb{E}\left[K_X(X,X)\,K_Y(Y,Y')\right] + \mathbb{E}\left[K_X(X,X')\right]\,\mathbb{E}\left[K_Y(Y,Y')\right] - 2\,\mathbb{E}\left[K_X(X,X')\,K_Y(Y,Y'')\right]\\ &= \mathbb{E}\left[K_X(X_1,X_2)\,K_Y(Y_1,Y_2) + K_X(X_1,X_2)\,K_Y(Y_3,Y_4) - 2\,K_X(X_1,X_2)\,K_Y(Y_1,Y_3)\right]\\ &= \mathbb{E}\left[\theta(\boldsymbol{Z}_1,\boldsymbol{Z}_2,\boldsymbol{Z}_3,\boldsymbol{Z}_4)\right]\end{aligned}$$

where $\boldsymbol{Z}_1 := (X_1,Y_1)$, $\boldsymbol{Z}_2 := (X_2,Y_2)$, $\boldsymbol{Z}_3 := (X_3,Y_3)$ and $\boldsymbol{Z}_4 := (X_4,Y_4)$ are four independent random pairs following the same bivariate distribution $\mathbb{P}_{XY}$, and $\theta : \mathcal{Z}^4 \to \mathbb{R}$ (with $\mathcal{Z} := \mathcal{X}\times\mathcal{Y}$) is the 4-argument function defined by:

$$\theta : (\boldsymbol{z}_1,\boldsymbol{z}_2,\boldsymbol{z}_3,\boldsymbol{z}_4) \longmapsto K_X(x_1,x_2)\,K_Y(y_1,y_2) + K_X(x_1,x_2)\,K_Y(y_3,y_4) - 2\,K_X(x_1,x_2)\,K_Y(y_1,y_3)\ .$$

Since θ is not symmetric (*i.e.* it is not invariant under all possible permutations of its arguments), the theory of U-statistics and V-statistics [91] (see Chapters 5 and 6, pp. 171–242) cannot be applied directly. To circumvent this pitfall, θ needs to be replaced by the symmetric function:

$$\widetilde{\theta}(\boldsymbol{z}_1,\boldsymbol{z}_2,\boldsymbol{z}_3,\boldsymbol{z}_4) := \frac{1}{4!}\sum_{\sigma\in\mathbb{S}_4}\theta\left(\boldsymbol{z}_{\sigma(1)},\boldsymbol{z}_{\sigma(2)},\boldsymbol{z}_{\sigma(3)},\boldsymbol{z}_{\sigma(4)}\right)$$

where $\mathbb{S}_4$ is the set of all permutations of $\{1,2,3,4\}$. Ultimately, one has $\mathrm{HSIC}(X,Y) = \mathbb{E}\left[\widetilde{\theta}(\boldsymbol{Z}_1,\boldsymbol{Z}_2,\boldsymbol{Z}_3,\boldsymbol{Z}_4)\right]$, and this expression is the key to build Monte Carlo estimators.

C.2.2. *U-statistic and V-statistic estimators*

The available data is a Monte Carlo design of experiments composed of n input-output samples:

$$\boldsymbol{Z}_{\mathrm{obs}} := \left\{\boldsymbol{Z}^{(i)}\right\}_{1\le i\le n} = \left\{\left(X^{(i)},Y^{(i)}\right)\right\}_{1\le i\le n} \sim \left(\mathbb{P}_{XY}\right)^{\otimes n}\ .$$

The U-statistic estimator of $\mathrm{HSIC}(X,Y)$ is given by:

$$\begin{aligned}\widehat{H}^U &:= \binom{n}{4}^{-1}\sum_{1\le i_1<\ldots<i_4\le n}\widetilde{\theta}\left(\boldsymbol{Z}^{(i_1)},\boldsymbol{Z}^{(i_2)},\boldsymbol{Z}^{(i_3)},\boldsymbol{Z}^{(i_4)}\right)\\ &= \frac{1}{(n)_2}\sum_{i_1\ne i_2}K_X\left(X^{(i_1)},X^{(i_2)}\right)\,K_Y\left(Y^{(i_1)},Y^{(i_2)}\right)\ldots\\ &\qquad - \frac{2}{(n)_3}\sum_{i_1\ne i_2\ne i_3}K_X\left(X^{(i_1)},X^{(i_2)}\right)\,K_Y\left(Y^{(i_1)},Y^{(i_3)}\right)\ldots\\ &\qquad\qquad + \frac{1}{(n)_4}\sum_{i_1\ne\ldots\ne i_4}K_X\left(X^{(i_1)},X^{(i_2)}\right)\,K_Y\left(Y^{(i_3)},Y^{(i_4)}\right),\end{aligned}$$

with $(n)_p := n!/(n-p)!$ for any $0\le p\le n$. By way of comparison, the V-statistic estimator of $\mathrm{HSIC}(X,Y)$ has a similar (although slightly simpler) expression:

$$\widehat{H}^V := \frac{1}{n^4}\sum_{1\le i_1,\ldots,i_4\le n}\widetilde{\theta}\left(\boldsymbol{Z}^{(i_1)},\boldsymbol{Z}^{(i_2)},\boldsymbol{Z}^{(i_3)},\boldsymbol{Z}^{(i_4)}\right)$$

$$= \frac{1}{n^2} \sum_{i_1, i_2} K_X \left(X^{(i_1)}, X^{(i_2)} \right) K_Y \left(Y^{(i_1)}, Y^{(i_2)} \right) \dots$$

$$- \frac{2}{n^3} \sum_{i_1, i_2, i_3} K_X \left(X^{(i_1)}, X^{(i_2)} \right) K_Y \left(Y^{(i_1)}, Y^{(i_3)} \right) \dots$$

$$+ \frac{1}{n^4} \sum_{i_1, i_2, i_3, i_4} K_X \left(X^{(i_1)}, X^{(i_2)} \right) K_Y \left(Y^{(i_3)}, Y^{(i_4)} \right).$$

Appendix D. Kernel feature analysis

Let $K : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ be a Mercer kernel, let ν be a probability measure with support $\mathcal{X}$ and let T_K be the integral operator built from K and ν. The objective here is to give additional details regarding the numerical procedure described in Section 5.1 to solve the eigenvalue problem:

$$T_K\,\phi = \lambda\,\phi \quad \text{with} \quad \phi \in \mathrm{L}^2(\mathcal{X},\nu) \quad \text{and} \quad \lambda > 0\ .$$

Originally, the approach described below was known as the *Nyström method* [99,100]. In more recent papers, it is also called *kernel principal component analysis* (kernel PCA) [72,89,90] or *kernel feature analysis* (KFA) [65,95]. In this work, it was decided to use the latter terminology.

D.1. Discretization of the initial eigenvalue problem

Let $f \in \mathrm{L}^2(\mathcal{X},\nu)$ be an eigenfunction of T_K associated to $\lambda > 0$. For now, the idea is to solve the eigenvalue equation $T_K\,f = \lambda\,f$ without paying attention to the fact that the eigenfunction mentioned in Theorem 2.18 are expected to have unit L^2-norm. This renormalization issue will be addressed later. The starting point is to draw a n-sample $\boldsymbol{x}_{\mathrm{sim}} := (x_i)_{1\le i\le n}$ from ν. For any $x \in \mathcal{X}$, the pointwise equality $[T_K\,f](x) = \lambda\,f(x)$ can be discretized with the Monte Carlo method:

$$\lambda\,f(x) = \int_{\mathcal{X}} K(x,\xi)\,f(\xi)\,\mathrm{d}\xi \approx \frac{1}{n}\sum_{j=1}^{n} K(x,x_j)\,f(x_j) \quad \text{with} \quad x_1,\ldots,x_n \overset{\text{i.i.d.}}{\sim} \nu\ .$$

Taking $x = x_i$ for $i \in \{1,\ldots,n\}$ yields a system of n linear equations:

$$\forall\,1 \le i \le n, \quad \lambda\,f(x_i) \approx \frac{1}{n}\sum_{j=1}^{n} K(x_i,x_j)\,f(x_j)\ .$$

After denoting by $v_i := f(x_i)$ the unknown values taken by f at the simulated points, the above system of equations may be rewritten as a matrix equation based on the simulated Gram matrix $\boldsymbol{K}_n$:

$$\boldsymbol{K}_n\,\boldsymbol{v} = (n\,\lambda)\,\boldsymbol{v} \quad \text{with} \quad \boldsymbol{K}_n := [K(x_i,x_j)]_{1\le i,j\le n} \in \mathbb{R}^{n\times n} \quad \text{and} \quad \boldsymbol{v} := [v_i]_{1\le i\le n} \in \mathbb{R}^n\ . \tag{D.1}$$

The eigenvalue problem related to T_K is thus transformed into a matrix eigenvalue problem which can be solved numerically (as long as n is not too large).

Remark D.1. Let $\psi : \mathcal{X} \to \mathcal{F}$ be one possible feature map of K. Knowing that $K(x_i,x_j) = \langle \psi(x_i), \psi(x_j)\rangle_{\mathcal{F}}$, the Gram matrix $\boldsymbol{K}_n$ can be envisioned as the covariance matrix of the n feature functions $\psi(x_i)$ in the feature space $\mathcal{F}$. Furthermore, since KFA is basically an eigenanalysis of $\boldsymbol{K}_n$, it is often called kernel PCA.

D.2. Estimation of the eigenvalues and eigenfunctions

Let $(\gamma_k)_{1\le k\le n}$ denote the eigenvalues of $\boldsymbol{K}_n$ and let $(\boldsymbol{v}_k)_{1\le k\le n}$ denote the corresponding eigenvectors. By convention, it is assumed that $\gamma_1 \ge \ldots \ge \gamma_n \ge 0$. Note that the eigenvalues are all non-negative because $\boldsymbol{K}_n$ is a symmetric positive semi-definite matrix. With Eq. (D.1), it is clear that the k-th largest eigenvalue of T_K can be estimated by $\hat{\lambda}_k = \gamma_k/n$.

Now, the question is how to estimate the eigenfunctions $(\phi_k)_{1\le k\le n}$ of T_K from the eigenvectors $(\boldsymbol{v}_k)_{1\le k\le n}$ of the matrix $\boldsymbol{K}_n$. As already said when establishing Eq. (D.1), the k-th eigenvector $\boldsymbol{v}_k := [v_{ik}]_{1\le i\le n}$ may be seen as a discrete estimate of an eigenfunction f_k associated to λ_k:

$$\forall\,1 \le i \le n, \quad v_{ik} \approx f_k(x_i)\ .$$

By default, the Euclidean norm of the eigenvector $\boldsymbol{v}_k$ returned by the eigenvalue algorithm is equal to 1. Hence, the L^2-norm of f_k is decreasing with the sample size n:

$$\|f_k\|_{\mathrm{L}^2}^2 = \int_{\mathcal{X}} |f_k(x)|^2 \,\mathrm{d}\nu(x) \approx \frac{1}{n} \sum_{i=1}^{n} |f_k(x_i)|^2 \approx \frac{1}{n} \sum_{i=1}^{n} |v_{ik}|^2 = \frac{1}{n} \|\boldsymbol{v}_k\|_{\mathbb{R}^n}^2 = \frac{1}{n} \ .$$

This means $\|f_k\|_{\mathrm{L}^2} \approx 1/\sqrt{n}$ whereas the eigenfunction ϕ_k involved in the Mercer decomposition must respect the unit-norm constraint in $\mathrm{L}^2(\mathcal{X}, \nu)$. To bypass this problem, $\boldsymbol{v}_k$ is replaced by $\boldsymbol{w}_k := \sqrt{n}\,\boldsymbol{v}_k$ and it is then possible to construct a discrete estimate of the L^2-normalized eigenfunction $\phi_k := f_k/\|f_k\|_{\mathrm{L}^2}$:

$$w_{ik} \approx \phi_k(x_i) \ .$$

Of course, one could decide to take $-\boldsymbol{w}_k$ instead of $\boldsymbol{w}_k$. A rule has to be laid down to eliminate this last degree of freedom. For this, the simulated data in $\boldsymbol{x}_{\mathrm{sim}}$ are sorted in increasing order and the eigenvector $\boldsymbol{w}_k$ is renumbered accordingly. There exists an integer m_k such that the subsequence $(w_{ik})_{m_k \leq i \leq n}$ is monotonous. The sign of $\boldsymbol{w}_k$ is then chosen so that this subsequence is increasing. Thus, among the two possible functions, ϕ_k is assumed to be the one which is increasing in the neighborhood of the upper bound. This decision rule applies in all the upcoming numerical experiments.

Remark D.2. The eigenvalue and eigenfunction estimates of T_K are simply obtained by properly renormalizing the eigendecompositon of the Gram matrix $\boldsymbol{K}_n$:

$$\boldsymbol{K}_n = \sum_{k=1}^{n} \gamma_k \,\boldsymbol{v}_k \boldsymbol{v}_k^T = \sum_{k=1}^{n} \left(\frac{\gamma_k}{n}\right) \left(\sqrt{n}\,\boldsymbol{w}_k\right) \left(\sqrt{n}\,\boldsymbol{w}_k\right)^T = \sum_{k=1}^{n} \hat{\lambda}_k \,\widehat{\phi}_k(\boldsymbol{x}_{\mathrm{sim}})\,\widehat{\phi}_k(\boldsymbol{x}_{\mathrm{sim}})^T \ .$$

D.3. **Nyström extension**

For extrapolation purposes, it would be interesting to construct a continuous estimate of the unknown eigenfunction ϕ_k. To do this, an interpolation method called the Nyström extension [46] can be used. In this approach, for any given point $x \in \mathcal{X}$, the value $\phi_k(x)$ is estimated from the only knowledge of γ_k and $\boldsymbol{w}_k$:

$$\widehat{\phi}_k^{\mathrm{Nys}}(x) := \frac{1}{\gamma_k} \sum_{i=1}^{n} w_{ik}\, K(x, x_i) \approx \frac{1}{\hat{\lambda}_k} \int_{\mathcal{X}} K(x,\xi)\,\phi_k(\xi)\,\mathrm{d}\nu(\xi) \approx \phi_k(x) \ , \tag{D.2}$$

and this provides an overall estimator $\widehat{\phi}_k^{\mathrm{Nys}}$ of the k-th eigenfunction ϕ_k on $\mathcal{X}$. Computing the value of $\widehat{\phi}_k^{\mathrm{Nys}}(x)$ for any given prediction point $x \in \mathcal{X}$ only asks for n additional kernel evaluations between x and the points in $\boldsymbol{x}_{\mathrm{sim}}$. As $\widehat{\phi}_k^{\mathrm{Nys}}$ is expressed as a linear combination of n canonical features, $\widehat{\phi}_k^{\mathrm{Nys}}$ inherits from the regularity of the kernel K. In particular, as K is a continuous kernel, $\widehat{\phi}_k^{\mathrm{Nys}}$ is in turn a continuous function.

Remark D.3. Although it was not built to this end, note that the function $\widehat{\phi}_k^{\mathrm{Nys}}$ interpolates the points $\boldsymbol{x}_{\mathrm{sim}}$ and $\boldsymbol{w}_k$. Indeed, one has:

$$\forall\, 1 \leq j \leq n, \quad \widehat{\phi}_k^{\mathrm{Nys}}(x_j) = \frac{1}{\gamma_k} \sum_{i=1}^{n} w_{ik}\, K(x_j, x_i) = \frac{1}{\gamma_k} (\boldsymbol{K}_n\, \boldsymbol{w}_k)_j = \frac{1}{\gamma_k} (\gamma_k\, \boldsymbol{w}_k)_j = w_{jk} \ .$$

Appendix E. Analytical resolution of the boundary value problem

E.1. **Exact resolution for** $r = 1$

In this case, the boundary value problem $(\mathcal{B}_\lambda^r)$ is restricted to:

$$\boxed{\lambda\,\phi'' + \phi = 0 \quad \text{with} \quad \phi'(0) = \phi'(1) = 0 \quad \text{and} \quad \lambda > 0} \;. \tag{$\mathcal{B}_\lambda^1$}$$

To begin with, any real-valued solution of this ODE may be written as:

$$\forall t \in [0\,,1], \quad \phi(t) = \alpha\cos(\xi t) + \beta\sin(\xi t) \quad \text{with} \quad \xi := \frac{1}{\sqrt{\lambda}} \quad \text{and} \quad (\alpha,\beta) \in \mathbb{R}^2 \;.$$

The boundary conditions on ϕ' yield $\beta = 0$ and $\sin(\xi) = 0$. Hence, there are countably infinitely many possible solutions for ξ and they are all given by $\xi_k = k\pi$ (with $k \geq 1$ since $\lambda > 0$). The sequence of eigenvalues $(\lambda_k)_{k\geq 1}$ directly follows with $\lambda_k = 1/\xi_k^2 = 1/(k\pi)^2$. For each eigenvalue λ_k, a pair $(\alpha_k, \beta_k) \in \mathbb{R}^2$ is used to parametrize the corresponding eigenfunction ϕ_k. One has $\beta_k = 0$ but α_k remains free of any constraint. This is natural since the solution space is expected to be an eigenspace (*i.e.* a linear subspace of dimension at least 1). More precisely, the eigenspace associated to λ_k can be expressed as:

$$E_1(\lambda_k) := \{\phi_k : t \in [0\,,1] \longmapsto \alpha_k \cos(k\pi t) \;\text{ with }\; \alpha_k \in \mathbb{R}\} \;.$$

The solution space of $(\mathcal{B}_\lambda^1)$ contains infinitely many solutions but the Mercer decomposition of k_{Sob}^1 requires L^2-normalized eigenfunctions. This leads to take $\alpha_k = \pm\sqrt{2}$ and $\phi_k : t \in [0\,,1] \mapsto \sqrt{2}\cos(k\pi t)$. Everything is therefore consistent with the Mercer expansion of k_{Sob}^1 already found in Section 5.

E.2. **Partial resolution for** $r = 3$

E.2.1. *Construction of the matrix* $\boldsymbol{M}_3(\xi)$

For $r = 3$, the boundary value problem is given by:

$$\lambda\phi^{[6]} + \phi = 0 \quad \text{with} \quad \begin{cases} \phi^{[3]}(0) = \phi^{[3]}(1) = 0 \\ \phi(0) - \phi(1) = \phi^{[5]}(0) = \phi^{[5]}(1) \\ \phi'(1) - \phi'(0) = \phi^{[4]}(0) = \phi^{[4]}(1) \end{cases} \quad \text{and} \quad \lambda > 0 \;. \tag{$\mathcal{B}_\lambda^3$}$$

Any real-valued solution of this ODE may be written as:

$$\begin{aligned} \phi(t) \;=\; & \alpha\,\cos(\xi t) + \beta\,\sin(\xi t) + \gamma\,\exp\big(\tfrac{\sqrt{3}}{2}\,\xi t\big)\,\cos\big(\tfrac{1}{2}\,\xi t\big) + \delta\,\exp\big(\tfrac{\sqrt{3}}{2}\,\xi t\big)\,\sin\big(\tfrac{1}{2}\,\xi t\big)\,\ldots \\ & + \mu\,\exp\big(-\tfrac{\sqrt{3}}{2}\,\xi t\big)\,\cos\big(\tfrac{1}{2}\,\xi t\big) + \nu\,\exp\big(-\tfrac{\sqrt{3}}{2}\,\xi t\big)\,\sin\big(\tfrac{1}{2}\,\xi t\big) \end{aligned} \tag{E.1}$$

$$\text{with} \quad \xi := \frac{1}{\sqrt[6]{\lambda}} \quad \text{and} \quad (\alpha,\beta,\gamma,\delta,\epsilon,\zeta) \in \mathbb{R}^6 \;.$$

To go further, the boundary conditions of $(\mathcal{B}_\lambda^3)$ need to be rewritten in terms of the general solution of the ODE. For the sake of convenience, Eq. (E.1) is rewritten in a more compact style:

$$\phi(t) = \boldsymbol{w}^T \boldsymbol{Q}(t) = \sum_{i=1}^{6} w_i \, Q_i(t \mid \xi)$$

$$\text{with} \quad \boldsymbol{w} := [\, w_i \,]_{1\le i\le 6} = \begin{bmatrix} \alpha \\ \beta \\ \gamma \\ \delta \\ \epsilon \\ \zeta \end{bmatrix} \quad \text{and} \quad \boldsymbol{Q}(t) := [\, Q_i(t \mid \xi) \,]_{1\le i\le 6} = \begin{bmatrix} \cos(\xi t) \\ \sin(\xi t) \\ \exp\left(\frac{\sqrt{3}}{2}\,\xi t\right)\, \cos\left(\frac{1}{2}\,\xi t\right) \\ \exp\left(\frac{\sqrt{3}}{2}\,\xi t\right)\, \sin\left(\frac{1}{2}\,\xi t\right) \\ \exp\left(-\frac{\sqrt{3}}{2}\,\xi t\right) \cos\left(\frac{1}{2}\,\xi t\right) \\ \exp\left(-\frac{\sqrt{3}}{2}\,\xi t\right)\, \sin\left(\frac{1}{2}\,\xi t\right) \end{bmatrix}.$$

Assuming that $\lambda = 1/\xi^6$ is an eigenvalue of $T_{k^3_{\mathrm{Sob}}}$, the functions $[Q_i(\cdot \mid \xi)]_{1\le i\le 6}$ form a basis of the 6-dimensional solution space of the ODE involved in $(\mathcal{B}_\lambda^3)$. After rewriting the trigonometric functions as the real parts of complex-valued functions, the derivatives of the basis functions can be easily calculated:

$$\forall k \ge 0, \quad \begin{cases} Q_1^{[k]}(t \mid \xi) = \xi^k \, \cos\left(\xi t + \frac{k\pi}{2}\right) \\ Q_2^{[k]}(t \mid \xi) = \xi^k \, \sin\left(\xi t + \frac{k\pi}{2}\right) \\ Q_3^{[k]}(t \mid \xi) = \xi^k \, \exp\left(\frac{\sqrt{3}}{2}\,\xi t\right)\, \cos\left(\frac{1}{2}\,\xi t + \frac{k\pi}{6}\right) \\ Q_4^{[k]}(t \mid \xi) = \xi^k \, \exp\left(\frac{\sqrt{3}}{2}\,\xi t\right)\, \sin\left(\frac{1}{2}\,\xi t + \frac{k\pi}{6}\right) \\ Q_5^{[k]}(t \mid \xi) = \xi^k \, \exp\left(-\frac{\sqrt{3}}{2}\,\xi t\right)\, \cos\left(\frac{1}{2}\,\xi t + \frac{5k\pi}{6}\right) \\ Q_6^{[k]}(t \mid \xi) = \xi^k \, \exp\left(-\frac{\sqrt{3}}{2}\,\xi t\right)\, \sin\left(\frac{1}{2}\,\xi t + \frac{5k\pi}{6}\right) \end{cases}.$$

These formulas allow to rearrange the boundary conditions as a system of equations which can be summarized by $\boldsymbol{M}_3(\xi)\, \boldsymbol{w} = \boldsymbol{0}$. In order to provide the analytical expression of all coefficients, $\boldsymbol{M}_3(\xi)$ is described with the help of its two half-matrices:

$$\boldsymbol{M}_3(\xi) = \left[\; \boldsymbol{M}_A(\xi) \;\middle|\; \boldsymbol{M}_B(\xi) \;\right] \in \mathbb{R}^{6\times 6} \quad \text{with} \quad \begin{cases} \boldsymbol{M}_A(\xi) \in \mathbb{R}^{6\times 3} \\ \boldsymbol{M}_B(\xi) \in \mathbb{R}^{6\times 3} \end{cases}.$$

They are respectively given by:

$$\boldsymbol{M}_A(\xi) = \begin{bmatrix} 0 & -1 & 0 \\ \sin(\xi) & -\cos(\xi) & -e^{\frac{\sqrt{3}}{2}\xi} \sin\left(\frac{1}{2}\,\xi\right) \\ \cos(\xi) - 1 & \sin(\xi) & e^{\frac{\sqrt{3}}{2}\xi} \cos\left(\frac{1}{2}\,\xi + \frac{2\pi}{3}\right) + \frac{1}{2} \\ -\sin(\xi) & \cos(\xi) - 1 & e^{\frac{\sqrt{3}}{2}\xi} \cos\left(\frac{1}{2}\,\xi + \frac{5\pi}{6}\right) + \frac{\sqrt{3}}{2} \\ \cos(\xi) - 1 & \sin(\xi) + \xi^5 & e^{\frac{\sqrt{3}}{2}\xi} \cos\left(\frac{1}{2}\,\xi\right) - 1 - \frac{\sqrt{3}}{2}\,\xi^5 \\ -(\sin(\xi) + \xi^3) & \cos(\xi) - 1 & e^{\frac{\sqrt{3}}{2}\xi} \cos\left(\frac{1}{2}\,\xi + \frac{\pi}{6}\right) - \frac{\sqrt{3}}{2} + \frac{1}{2}\,\xi^3 \end{bmatrix}$$

and:

$$\boldsymbol{M}_B(\xi) = \begin{bmatrix} 1 & 0 & 1 \\ e^{\frac{\sqrt{3}}{2}\xi}\cos\left(\frac{1}{2}\,\xi\right) & -e^{-\frac{\sqrt{3}}{2}\xi}\sin\left(\frac{1}{2}\,\xi\right) & e^{-\frac{\sqrt{3}}{2}\xi}\cos\left(\frac{1}{2}\,\xi\right) \\ e^{\frac{\sqrt{3}}{2}\xi}\sin\left(\frac{1}{2}\,\xi+\frac{2\pi}{3}\right)-\frac{\sqrt{3}}{2} & e^{-\frac{\sqrt{3}}{2}\xi}\cos\left(\frac{1}{2}\,\xi-\frac{2\pi}{3}\right)+\frac{1}{2} & e^{-\frac{\sqrt{3}}{2}\xi}\sin\left(\frac{1}{2}\,\xi-\frac{2\pi}{3}\right)+\frac{\sqrt{3}}{2} \\ e^{\frac{\sqrt{3}}{2}\xi}\sin\left(\frac{1}{2}\,\xi+\frac{5\pi}{6}\right)-\frac{1}{2} & e^{-\frac{\sqrt{3}}{2}\xi}\cos\left(\frac{1}{2}\,\xi+\frac{\pi}{6}\right)-\frac{\sqrt{3}}{2} & e^{-\frac{\sqrt{3}}{2}\xi}\sin\left(\frac{1}{2}\,\xi+\frac{\pi}{6}\right)-\frac{1}{2} \\ e^{\frac{\sqrt{3}}{2}\xi}\sin\left(\frac{1}{2}\,\xi\right)+\frac{1}{2}\xi^5 & e^{-\frac{\sqrt{3}}{2}\xi}\cos\left(\frac{1}{2}\,\xi\right)-1+\frac{\sqrt{3}}{2}\,\xi^5 & e^{-\frac{\sqrt{3}}{2}\xi}\sin\left(\frac{1}{2}\,\xi\right)+\frac{1}{2}\xi^5 \\ e^{\frac{\sqrt{3}}{2}\xi}\sin\left(\frac{1}{2}\,\xi+\frac{\pi}{6}\right)-\frac{1}{2}-\frac{\sqrt{3}}{2}\,\xi^3 & e^{-\frac{\sqrt{3}}{2}\xi}\cos\left(\frac{1}{2}\,\xi+\frac{5\pi}{6}\right)+\frac{\sqrt{3}}{2}+\frac{1}{2}\,\xi^3 & e^{-\frac{\sqrt{3}}{2}\xi}\sin\left(\frac{1}{2}\,\xi+\frac{5\pi}{6}\right)-\frac{1}{2}+\frac{\sqrt{3}}{2}\,\xi^3 \end{bmatrix}.$$

E.2.2. *Proof of Proposition 6.7*

It would be useless to seek to obtain a one-line expression of $\eta_3(\xi) = \det[\boldsymbol{M}_3(\xi)]$ with a computation by hand of the determinant. To find the function η_3^∞ which is asymptotically equivalent to η_3, the trick is to factorize each row $\boldsymbol{L}_i$ (with $1 \le i \le 6$) and each column $\boldsymbol{C}_j$ (with $1 \le j \le 6$) in $\boldsymbol{M}_3(\xi)$ by its leading-order term. In particular:

- $\boldsymbol{L}_5$ must be factorized by ξ^5.
- $\boldsymbol{L}_6$ must be factorized by ξ^3.
- $\boldsymbol{C}_3$ and $\boldsymbol{C}_4$ must be factorized by $e^{\frac{\sqrt{3}}{2}\xi}$.

Then, one has $\eta_3(\xi) = \xi^8\, e^{\sqrt{3}\xi} \det[\boldsymbol{N}_3(\xi)]$ with $\boldsymbol{N}_3(\xi)$ the matrix obtained after operating all factorizations. The idea is then to find a simple function which is asymptotically equivalent to $\det[\boldsymbol{N}_3(\xi)]$. A first step consists in taking asymptotically equivalent functions for all the coefficients of the matrix $\boldsymbol{N}_3(\xi)$:

$$\det[\boldsymbol{N}_3(\xi)] \underset{\xi\to\infty}{\sim} \begin{vmatrix} 0 & -1 & 0 & e^{-\frac{\sqrt{3}}{2}\xi} & 0 & 1 \\ \sin(\xi) & -\cos(\xi) & -\sin\left(\frac{1}{2}\,\xi\right) & \cos\left(\frac{1}{2}\,\xi\right) & e^{-\frac{\sqrt{3}}{2}\xi}\sin\left(\frac{1}{2}\,\xi\right) & e^{-\frac{\sqrt{3}}{2}\xi}\cos\left(\frac{1}{2}\,\xi\right) \\ \cos(\xi)-1 & \sin(\xi) & \boxed{\cos\left(\frac{1}{2}\,\xi+\frac{2\pi}{3}\right)} & \boxed{\sin\left(\frac{1}{2}\,\xi+\frac{2\pi}{3}\right)} & \boxed{\frac{1}{2}} & \boxed{\frac{\sqrt{3}}{2}} \\ -\sin(\xi) & \cos(\xi)-1 & \boxed{\cos\left(\frac{1}{2}\,\xi+\frac{5\pi}{6}\right)} & \boxed{\sin\left(\frac{1}{2}\,\xi+\frac{5\pi}{6}\right)} & \boxed{-\frac{\sqrt{3}}{2}} & \boxed{-\frac{1}{2}} \\ \frac{\cos(\xi)-1}{\xi^5} & 1 & \boxed{\frac{\cos\left(\frac{1}{2}\,\xi\right)}{\xi^5}} & \boxed{\frac{\sin\left(\frac{1}{2}\,\xi\right)}{\xi^5}} & \boxed{\frac{\sqrt{3}}{2}} & \boxed{\frac{1}{2}} \\ -1 & \frac{\cos(\xi)-1}{\xi^3} & \boxed{\frac{\cos\left(\frac{1}{2}\,\xi+\frac{\pi}{6}\right)}{\xi^3}} & \boxed{\frac{\sin\left(\frac{1}{2}\,\xi+\frac{\pi}{6}\right)}{\xi^3}} & \boxed{\frac{1}{2}} & \boxed{\frac{\sqrt{3}}{2}} \end{vmatrix}.$$

$$\underset{\xi\to\infty}{\sim} \det[\boldsymbol{N}_3^\infty(\xi)]$$

The framed coefficients in $\boldsymbol{N}_3^\infty(\xi)$ correspond to asymptotic approximations as opposed to all other coefficients which correspond to the original terms in $\boldsymbol{N}_3(\xi)$. It is much more convenient to handle $\boldsymbol{N}_3^\infty(\xi)$ because the coefficients have simple expressions and most of them vanish asymptotically. To go further, the determinant calculation rules (and especially the cofactor expansion) must be applied cleverly. More specifically, at each step, the calculations must be carried out so that the coefficients which are asymptotically vanishing enable simplifications. It is a fastidious job but it must be completed patiently in order to obtain:

$$\det\left[\boldsymbol{N}_3^{\infty}(\xi)\right] \underset{\xi\to\infty}{\sim} \tfrac{3}{4}\,\sin(\xi)\ .$$

The details of the step-by-step dimension reduction are not provided here for conciseness. This finally leads to the expected result:

$$\begin{aligned}\eta_3(\xi) = \det\left[\boldsymbol{M}_3(\xi)\right] = \xi^8\, e^{\sqrt{3}\xi}\,\det\left[\boldsymbol{N}_3(\xi)\right] &\underset{\xi\to\infty}{\sim} \xi^8\, e^{\sqrt{3}\xi}\,\det\left[\boldsymbol{N}_3^{\infty}(\xi)\right]\\ &\underset{\xi\to\infty}{\sim} \eta_3^{\infty}(\xi) = \tfrac{3}{4}\,\sin(\xi)\,\xi^8\, e^{\sqrt{3}\xi}\ .\end{aligned}$$

Appendix F. Proofs

F.1. **Proof of Theorem 2.23**

The idea is to prove that K is indeed the reproducing kernel of the RKHS defined in Theorem 2.23. First, it must be checked that all functions in $\mathcal{H}$ are well-defined. For any $h \in \mathcal{H}$, there exists $(a_i)_{i\in I} \in \ell^2(I)$ such that $h(\cdot) = \sum_{i\in I} a_i\, g_i(\cdot)$ and the convergence of the series must be justified at any point $x \in \mathcal{X}$. For this, simply apply the Cauchy-Schwarz inequality:

$$\forall\, x \in \mathcal{X}, \quad |h(x)| = \left|\sum_{i\in I} a_i\, g_i(x)\right| \leq \left(\sum_{i\in I} a_i^2\right)^{1/2} \left(\sum_{i\in I} g_i(x)^2\right)^{1/2} = \|(a_i)_{i\in I}\|_{\ell^2} \sqrt{K(x,x)} < \infty \ .$$

The vector space $\mathcal{H}$ from Eq. (2.17) is therefore composed of well-defined functions. The map $\langle\cdot,\cdot\rangle_{\mathcal{H}} : \mathcal{H}\times\mathcal{H} \to \mathbb{R}$ defined in Eq. (2.18) is also well-defined:

- The sequences $(a_i)_{i\in I}$ and $(b_i)_{i\in I}$ used to write h_1 and h_2 are uniquely defined since the functions $(g_i)_{i\in I}$ are assumed to be ℓ^2-linearly independent.
- The series $\sum_{i\in I} a_i b_i$ is always convergent because the sequences $(a_i)_{i\in I}$ and $(b_i)_{i\in I}$ belong to $\ell^2(I)$.

From this, it can be easily proved that $(\mathcal{H}, \langle\cdot,\cdot\rangle_{\mathcal{H}})$ is a Hilbert space. In particular, the normed space $(\mathcal{H}, \|\cdot\|_{\mathcal{H}})$ inherits from the completeness of $\ell^2(I)$. Now, it only remains to prove that K is a reproducing kernel for the constructed Hilbert space. For any $x \in \mathcal{X}$, one has:

$$K(\cdot, x) = \sum_{i\in I} g_i(x)\, g_i(\cdot) \quad \text{with} \quad \|(g_i(x))_{i\in I}\|_{\ell^2}^2 = \sum_{i\in I} g_i(x)^2 = K(x,x) < \infty \ ,$$

which allows to prove that $K(\cdot, x)$ belongs to $\mathcal{H}$. In addition, the reproducing property is verified:

$$\forall\, h \in \mathcal{H}, \quad \forall\, x \in \mathcal{X}, \quad \langle h, K(\cdot,x)\rangle_{\mathcal{H}} = \left\langle \sum_{i\in I} a_i\, g_i(\cdot)\, , \sum_{i\in I} g_i(x)\, g_i(\cdot) \right\rangle_{\mathcal{H}} = \sum_{i\in I} a_i\, g_i(x) = h(x) \ .$$

$(\mathcal{H}, \|\cdot\|_{\mathcal{H}})$ is therefore a Hilbert space for which K is reproducing. By Theorem 2.5, we know that it is the only one, which ends the proof.

F.2. **Proof of Proposition 6.1**

Let $\phi \in \mathrm{L}^2([0\,,1])$ be an eigenfunction of $T_{k^r_{\mathrm{Sob}}}$ associated to $\lambda > 0$. It must be proved that $\phi \in \mathrm{C}^\infty([0\,,1])$ and that ϕ is a solution of the ODE recalled below:

$$\boxed{\lambda\, \phi^{[2r]} + (-1)^{r+1}\, \phi = 0} \ . \qquad (\mathcal{E}^r_\lambda)$$

First, let us prove that ϕ has zero mean:

$$\begin{aligned}
\int_0^1 \phi(x)\,\mathrm{d}x = \frac{1}{\lambda}\int_0^1 [T_{k^r_{\mathrm{Sob}}}\,\phi](x)\,\mathrm{d}x &= \int_0^1 \left(\int_0^1 k^r_{\mathrm{Sob}}(x,\xi)\,\phi(\xi)\,\mathrm{d}\xi\right)\mathrm{d}x \quad \text{by definition of } T_{k^r_{\mathrm{Sob}}}, \\
&= \int_0^1 \left(\int_0^1 k^r_{\mathrm{Sob}}(x,\xi)\,\mathrm{d}x\right)\phi(\xi)\,\mathrm{d}\xi \quad \text{with Fubini's theorem,} \\
&= 0 \quad \text{as } k^r_{\mathrm{Sob}} \text{ is orthogonal.}
\end{aligned}$$

Before going further, let us recall a very classical result in differential calculus.

Theorem F.1 (Leibniz integral rule)**.** *Let $I \subseteq \mathbb{R}$ be an interval. Let $a : I \to \mathbb{R}$ and $b : I \to \mathbb{R}$ be two continuously differentiable functions. Let $h : I \times \mathbb{R} \longmapsto \mathbb{R}$ such that:*

- *h is continuous on $I \times \mathbb{R}$,*
- *$h(\cdot, \xi)$ is continuously differentiable on I for all $\xi \in \mathbb{R}$,*
- *$h(x, \cdot)$ is integrable on $\mathbb{R}$ for all $x \in I$.*

Then, the function defined by:

$$\begin{array}{rccc} H: & I & \longrightarrow & \mathbb{R} \\ & x & \longmapsto & \displaystyle\int_{a(x)}^{b(x)} h(x,\xi)\,\mathrm{d}\xi \ , \end{array}$$

is continuously differentiable on I with:

$$\boxed{\forall\, x \in I, \quad H'(x) = \int_{a(x)}^{b(x)} \frac{\partial h}{\partial x}(x,\xi)\,\mathrm{d}\xi + b'(x)\,h(x,b(x)) - a'(x)\,h(x,a(x))} \ . \tag{F.1}$$

The main idea of our proof is to use the Leibniz integral rule to justify that the eigenvalue equation $[T_{k^r_{\mathrm{Sob}}}\,\phi](\cdot) = \lambda\,\phi(\cdot)$ is (at least) $2r$ times differentiable on $[0\,,1]$. The derivatives $\lambda\,\phi^{[k]} = (T_{k^r_{\mathrm{Sob}}}\,\phi)^{[k]}$ for all $k \in \{1,\ldots,2r\}$ can then be computed with the help of Eq. (F.1). To begin, let us revert to the integral expression of $T_{k^r_{\mathrm{Sob}}}\,\phi$:

$$\begin{aligned} \lambda\,\phi(x) = \left[T_{k^r_{\mathrm{Sob}}}\,\phi\right](x) &= \int_0^1 k^r_{\mathrm{Sob}}(x,\xi)\,\phi(\xi)\,\mathrm{d}\xi \\ &= \int_0^1 \left(\sum_{k=1}^r \frac{B_k(x)B_k(\xi)}{(k!)^2} + \frac{(-1)^{r+1}}{(2r)!}B_{2r}(|x-\xi|)\right)\phi(\xi)\,\mathrm{d}\xi \\ &= \sum_{k=1}^r \left(\int_0^1 B_k(\xi)\,\phi(\xi)\,\mathrm{d}\xi\right)\frac{B_k(x)}{(k!)^2} + \frac{(-1)^{r+1}}{(2r)!}\int_0^1 B_{2r}(|x-\xi|)\,\phi(\xi)\,\mathrm{d}\xi \\ &= \sum_{k=1}^r \frac{\beta_k}{(k!)^2}\,B_k(x) + \frac{(-1)^{r+1}}{(2r)!}\left[\int_0^x B_{2r}(x-\xi)\,\phi(\xi)\,\mathrm{d}\xi + \int_x^1 B_{2r}(\xi-x)\,\phi(\xi)\,\mathrm{d}\xi\right] \\ &= G_A(x) + \frac{(-1)^{r+1}}{(2r)!}\,G_B(x) \ . \end{aligned} \tag{F.2}$$

Of course, $G_A \in \mathrm{C}^\infty([0\,,1])$. Because of the recursive formula ruling the differentiation of Bernoulli polynomials (see Appendix A.1.3), the derivatives of G_A can be expressed as follows:

$$\begin{cases} \forall\, p \in \{1,\ldots,r\}, & G_A^{[p]}(x) = \displaystyle\sum_{k=p}^r \frac{\beta_k}{k!(k-p)!}\,B_{k-p}(x) \ ; \\ \forall\, p > r, & G_A^{[p]}(x) = 0 \ . \end{cases} \tag{F.3}$$

Regarding G_B, it is a bit more complicated. It must be noted that Theorem F.1 applies to both integrals. In particular, ϕ is continuous on $[0\,,1]$ because $\lambda > 0$ (according to Theorem 2.18). As a consequence, the functions $\xi \mapsto B_{2r}(x-\xi)\,\phi(\xi)$ and $\xi \mapsto B_{2r}(\xi-x)\,\phi(\xi)$ verify all the assumptions of Theorem F.1. Therefore, G_B is continuously differentiable on $[0\,,1]$ and one has:

$$\begin{aligned} G_B'(x) &= 2r \int_0^x B_{2r-1}(x-\xi)\,\phi(\xi)\,\mathrm{d}\xi + B_{2r}(0)\,\phi(x) - 2r \int_0^x B_{2r-1}(\xi-x)\,\phi(\xi)\,\mathrm{d}\xi - B_{2r}(0)\,\phi(x) \\ &= 2r \left[\int_0^x B_{2r-1}(x-\xi)\,\phi(\xi)\,\mathrm{d}\xi - \int_x^1 B_{2r-1}(x-\xi)\,\phi(\xi)\,\mathrm{d}\xi \right] . \end{aligned}$$

This suggests that G_B' is also continuously differentiable and G_B'' can be obtained in the same way. In fact, a simple proof by induction (not given here for conciseness) allows to show that G_B is (at least) $2r-1$ times continuously differentiable on $[0\,,1]$ with:

$$\forall\, 1 \le p \le 2r-1,\; G_B^{[p]}(x) = \frac{(2r)!}{(2r-p)!} \left[\int_0^x B_{2r-p}(x-\xi)\,\phi(\xi)\,\mathrm{d}\xi + (-1)^p \int_x^1 B_{2r-p}(\xi-x)\,\phi(\xi)\,\mathrm{d}\xi \right] . \tag{F.4}$$

The key arguments are the Leibniz integral rule, the derivatives of Bernoulli polynomials (see Appendix A.1.3) and the boundary values of Bernoulli polynomials (see Appendix A.1.4). For $p = 2r-1$, Eq. (F.4) becomes:

$$G_B^{[2r-1]}(x) = (2r)! \left[\int_0^x B_1(x-\xi)\,\phi(\xi)\,\mathrm{d}\xi - \int_x^1 B_1(\xi-x)\,\phi(\xi)\,\mathrm{d}\xi \right] .$$

Hence, $G_B^{[2r-1]}$ is continuously differentiable on $[0\,,1]$ and $G_B^{[2r]}$ can be computed with Eq. (F.1):

$$\begin{aligned} G_B^{[2r]}(x) &= (2r)! \left[\int_0^x \phi(\xi)\;\mathrm{d}\xi + B_1(0)\,\phi(\xi) - \left(\int_x^1 \phi(\xi)\;\mathrm{d}\xi - B_1(0)\,\phi(x) \right) \right] \\ &= (2r)! \left[\int_0^1 \phi(\xi)\,\mathrm{d}\xi - \phi(x) \right] = -(2r)!\,\phi(\xi) \qquad \text{since } B_1(0) = -\frac{1}{2} \text{ and } \phi \text{ has zero mean.} \end{aligned}$$

Finally, the eigenvalue equation $[T_{k^r_{\mathrm{Sob}}}\,\phi](\cdot) = \lambda\,\phi(\cdot)$ can be differentiated $2r$ times and one has:

$$\lambda\,\phi^{[2r]}(x) = (T_{k^r_{\mathrm{Sob}}}\,\phi)^{[2r]}(x) = G_A^{[2r]}(x) + \frac{(-1)^{r+1}}{(2r)!} G_B^{[2r]}(x) = (-1)^r \phi(x)\ ,$$

which justifies that ϕ is indeed a solution of $(\mathcal{E}^r_\lambda)$. Besides, the equality $\lambda\,\phi^{[2r]} = (-1)^r \phi$ allows to prove that ϕ is infinitely smooth since:

$$\phi \in \mathrm{C}^{2r}([0\,,1]) \implies \phi^{[2r]} \in \mathrm{C}^{2r}([0\,,1]) \implies \phi \in \mathrm{C}^{4r}([0\,,1]) \implies \phi^{[2r]} \in \mathrm{C}^{4r}([0\,,1]) \implies \ldots \implies \phi \in \mathrm{C}^{\infty}([0\,,1])\,.$$

Remark F.2. Eq. (F.3) and (F.4) will be very useful in the proof of Theorem 6.3 since they provide an integral expression of the first $2r-1$ derivatives of ϕ. In short, one may write:

$$\boxed{\begin{aligned}\forall\, p \in \{0,\dots,2r-1\},\quad \lambda\,\phi^{[p]}(x) &= G_A^{[p]} + \frac{(-1)^{r+1}}{(2r)!}\,G_B^{[p]} \\ &= \sum_{k=p}^{r} \frac{\beta_k}{k!(k-p)!}\,B_{k-p}(x)\ \dots \\ &\qquad + \frac{(-1)^{r+1}}{(2r-p)!}\left[\int_0^x B_{2r-p}(x-\xi)\,\phi(\xi)\,\mathrm{d}\xi\ \dots\right. \\ &\qquad\qquad \left. + (-1)^p \int_x^1 B_{2r-p}(\xi-x)\,\phi(\xi)\,\mathrm{d}\xi\right]\end{aligned}}\ . \tag{F.5}$$

Note that the first term is equal to zero as soon as $p > r$.

F.3. **Proof of Theorem 6.3**

Let $\phi \in \mathrm{L}^2([0\,,1])$ and $\lambda > 0$. It must be proved that the two following statements are equivalent:

(i) ϕ is an eigenfunction of the integral operator $T_{k^r_{\text{Sob}}}$ with eigenvalue λ.

(ii) ϕ is a solution of the boundary value problem defined by:

$$\boxed{\lambda\,\phi^{[2r]} + (-1)^{r+1}\,\phi = 0 \text{ with } \begin{cases} \phi^{[r]}(0) = \phi^{[r]}(1) = 0 \\ \forall\, 0 \le p \le r-2,\quad (-1)^{r+p}\left(\phi^{[p]}(1) - \phi^{[p]}(0)\right) = \phi^{[2r-p-1]}(0) \\ \forall\, 0 \le p \le r-2,\quad \phi^{[2r-p-1]}(0) = \phi^{[2r-p-1]}(1)\end{cases}}\ . \tag{$\mathcal{B}^r_\lambda$}$$

For convenience, specific labels are used to distinguish the three different types of boundary conditions:

$(\mathcal{C}_1)$ $\qquad \phi^{[r]} = \phi^{[r]}(1) = 0$;

$(\mathcal{C}_2)$ $\quad \forall\, 0 \le p \le r-2,\quad (-1)^{r+p}(\phi^{[p]}(1) - \phi^{[p]}(0)) = \phi^{[2r-p-1]}(0)$;

$(\mathcal{C}_3)$ $\quad \forall\, 0 \le p \le r-2,\quad \phi^{[2r-p-1]}(0) = \phi^{[2r-p-1]}(1)$.

First, let us prove that $\boxed{\text{(i)} \implies \text{(ii)}}$.

As ϕ is an eigenfunction of $T_{k^r_{\text{Sob}}}$ with $\lambda > 0$, Theorem 6.1 states that it is a solution of the ODE $(\mathcal{E}^r_\lambda)$. Then, the integral formula given in Eq. (F.5) to express the first $2r-1$ derivatives of $T_{k^r_{\text{Sob}}}\,\phi$ can be used to demonstrate that ϕ verifies all the boundary conditions.

Boundary conditions $(\mathcal{C}_1)$

- $\lambda\,\phi^{[r]}(0) = \dfrac{\beta_r}{r!} + \dfrac{(-1)^{2r+1}}{r!}\displaystyle\int_0^1 B_r(\xi)\,\phi(\xi)\,\mathrm{d}\xi = \dfrac{\beta_r}{r!} - \dfrac{\beta_r}{r!} = 0$.
- $\lambda\,\phi^{[r]}(1) = \dfrac{\beta_r}{r!} + \dfrac{(-1)^{r+1}}{r!}\displaystyle\int_0^1 B_r(1-\xi)\,\phi(\xi)\,\mathrm{d}\xi = \dfrac{\beta_r}{r!} + \dfrac{(-1)^{r+1}}{r!}\displaystyle\int_0^1 (-1)^r B_r(\xi)\,\phi(\xi)\,\mathrm{d}\xi = \dfrac{\beta_r}{r!} - \dfrac{\beta_r}{r!} = 0$.

As a result, $\lambda\,\phi^{[r]}(0) = \lambda\,\phi^{[r]}(1) = 0$ and this leads to $(\mathcal{C}_1)$ since $\lambda > 0$.

Boundary conditions $(\mathcal{C}_3)$
For any $q \in \{r+1,\dots,2r-1\}$, one has:

- $\lambda\,\phi^{[q]}(0) = \dfrac{(-1)^{r+1}}{(2r-q)!}\,(-1)^q \displaystyle\int_0^1 B_{2r-q}(\xi)\,\phi(\xi)\,\mathrm{d}\xi = \dfrac{(-1)^{r+q+1}}{(2r-q)!}\,\beta_{2r-q}\ .$
- $\lambda\,\phi^{[q]}(1) = \dfrac{(-1)^{r+1}}{(2r-q)!} \displaystyle\int_0^1 B_{2r-q}(1-\xi)\,\phi(\xi)\,\mathrm{d}\xi = \dfrac{(-1)^{r+1}}{(2r-q)!} \displaystyle\int_0^1 (-1)^{2r-q} B_{2r-q}(\xi)\,\phi(\xi)\,\mathrm{d}\xi = \dfrac{(-1)^{r+q+1}}{(2r-p)!}\,\beta_{2r-q}\ .$

When $p \in \{0,\dots,r-2\}$, note that $q = 2r-p-1 \in \{r+1,\dots,2r-1\}$. After operating the associated index shift, one has $\lambda\,\phi^{[2r-p-1]}(0) = \lambda\,\phi^{[2r-p-1]}(1)$ for any $p \in \{0,\dots,r-2\}$ and this leads to $(\mathcal{C}_3)$ since $\lambda > 0$.

Boundary conditions $(\mathcal{C}_2)$
For any $p \in \{0,\dots,r-2\}$, one has:

- $\lambda\,\phi^{[p]}(0) = \displaystyle\sum_{k=p}^{r} \dfrac{\beta_k}{k!(k-p)!}\,B_{k-p}(0) + \dfrac{(-1)^{r+p+1}}{(2r-p)!}\,\beta_{2r-p}\ .$
- $\lambda\,\phi^{[p]}(1) = \displaystyle\sum_{k=p}^{r} \dfrac{\beta_k}{k!(k-p)!}\,B_{k-p}(1) + \dfrac{(-1)^{r+p+1}}{(2r-p)!}\,\beta_{2r-p}\ .$

Because of the boundary values of Bernoulli polynomials (see Appendix A.1.4), the difference of the two terms above can be expressed very simply:

$$\lambda\,(\phi^{[p]}(1) - \phi^{[p]}(0)) = \sum_{k=p}^{r} \frac{\beta_k}{k!(k-p)!}(B_{k-p}(1) - B_{k-p}(0)) = \frac{\beta_{p+1}}{(p+1)!}\ .$$

To compute $\phi^{[2p-r-1]}(0)$, the formula obtained at the previous step can be applied since $2r-p-1 \in \{r+1,\dots,2r-1\}$:

$$\phi^{[2p-r-1]}(0) = \frac{(-1)^{3r-p}}{(p+1)!}\,\beta_{p+1} = (-1)^{r+p}\,\frac{\beta_{p+1}}{(p+1)!}\ .$$

As a result, $(-1)^{r+p}\,\lambda\,(\phi^{[p]}(1) - \phi^{[p]}(0)) = \lambda\,\phi^{[2r-p-1]}(0)$ and this leads to $(\mathcal{C}_2)$ since $\lambda > 0$.

As a conclusion, ϕ verifies both the ODE and the boundary conditions, hence verifies $(\mathcal{B}^r_\lambda)$.

Now, let us prove that $\boxed{\text{(ii)} \implies \text{(i)}}$.

For this, ϕ is assumed to be a solution of $(\mathcal{B}^r_\lambda)$. This means that ϕ verifies the equality $\lambda\,\phi^{[2r]} = (-1)^r\phi$ and all the boundary conditions. Under these assumptions, it must be proved that ϕ is an eigenfunction of $T_{k^r_{\mathrm{Sob}}}$ with eigenvalue λ. To begin, it can be proved that ϕ has zero mean since:

$$\int_0^1 \phi(\xi)\,\mathrm{d}\xi = (-1)^r\,\lambda \int_0^1 \phi^{[2r]}(\xi)\,\mathrm{d}\xi = (-1)^r\,\lambda\,\left(\phi^{[2r-1]}(1) - \phi^{[2r-1]}(0)\right) = 0\ .$$

The final equality is provided by the boundary values of $\phi^{[2r-1]}$ in $(\mathcal{C}_3)$. Coming back to the main objective, the key idea of our proof is to write $[T_{k^r_{\mathrm{Sob}}}\,\phi](\cdot) = (-1)^r\,\lambda\,[T_{k^r_{\mathrm{Sob}}}\phi^{[2r]}](\cdot)$ and to transform $T_{k^r_{\mathrm{Sob}}}\,\phi^{[2r]}$ by repeating integration by parts in order to demonstrate that ϕ verifies the eigenvalue equation $T_{k^r_{\mathrm{Sob}}}\,\phi = \lambda\,\phi$. Before doing this, let us introduce the following notation:

$$\forall\, g : [0\,,1] \to \mathbb{R}, \quad g(1) - g(0) = [g(t)]_{t=0}^1\ .$$

This will be notably useful to summarize the result of an integration by parts. The integral expression of $T_{k^r_{\mathrm{Sob}}}\,\phi^{[2r]}$ is now splitted into three terms:

$$[T_{k^r_{\mathrm{Sob}}}\,\phi^{[2r]}](x) = \int_0^1 k^r_{\mathrm{Sob}}(x,\xi)\,\phi^{[2r]}(\xi)\,\mathrm{d}\xi = \int_0^1 \left(\sum_{j=1}^{r} \frac{B_j(x)B_j(x')}{(j!)^2} + \frac{(-1)^{r+1}}{(2r)!}B_{2r}(|x-\xi|)\right)\,\phi^{[2r]}(\xi)\,\mathrm{d}\xi$$

$$
\begin{aligned}
&= \sum_{j=1}^{r} \left(\int_0^1 B_j(\xi)\, \phi^{[2r]}(\xi)\, \mathrm{d}\xi \right) \frac{B_j(x)}{(j!)^2} \quad \dots \\
&\qquad + \frac{(-1)^{r+1}}{(2r)!} \left[\int_0^x B_{2r}(x-\xi)\, \phi^{[2r]}(\xi)\, \mathrm{d}\xi + \int_x^1 B_{2r}(\xi - x)\, \phi^{[2r]}(\xi)\, \mathrm{d}\xi \right] \\
&= \sum_{j=1}^{r} \frac{I_j}{(j!)^2}\, B_j(x) + \frac{(-1)^{r+1}}{(2r)!} [\, J_-(x) + J_+(x) \,] \ . \qquad \text{(F.6)}
\end{aligned}
$$

Whether for I_j, $J_-(x)$ or $J_+(x)$, the integrand is the product of $\phi^{[2r]}$ by a Bernoulli polynomial. Therefore, after repeating integration by parts as many times as necessary, these integrals can be calculated, or at least transformed into expressions where integration only concerns ϕ.

Integration by parts to transform I_j
It can be shown that:

$$
\begin{aligned}
\forall\, 1 \le j \le r, \quad \forall\, 1 \le k \le j, \qquad I_j = &\sum_{l=0}^{k-1} (-1)^l \frac{j!}{(j-l)!} \Big[B_{j-l}(t)\, \phi^{[2r-l-1]}(t) \Big]_{t=0}^{1} \quad \dots \\
&+ (-1)^k \frac{j!}{(j-k)!} \int_0^1 B_{j-k}(\xi)\, \phi^{[2r-k]}(\xi)\, \mathrm{d}\xi \ .
\end{aligned} \tag{F.7}
$$

The proof (by induction) relies on the formula of the derivatives of Bernoulli polynomials (see Appendix A.1.3). Then, taking $k = j$ in Eq. (F.7) yields:

$$
\boxed{\forall\, 1 \le j \le r, \quad I_j = \sum_{l=0}^{j} (-1)^l \frac{j!}{(j-l)!} \Big[B_{j-l}(t)\, \phi^{[2r-l-1]}(t) \Big]_{t=0}^{1}} \ . \tag{F.8}
$$

Integration by parts to transform $J_-(x)$
It can be shown that:

$$
\begin{aligned}
\forall\, 1 \le k \le 2r, \quad J_-(x) = &\sum_{l=0}^{k-1} \frac{(2r)!}{(2r-l)!} \Big[B_{2r-l}(x-t)\, \phi^{[2r-l-1]}(t) \Big]_{t=0}^{x} \quad \dots \\
&+ \frac{(2r)!}{(2r-k)!} \int_0^x B_{2r-k}(x-\xi)\, \phi^{[2r-k]}(\xi)\, \mathrm{d}\xi \ .
\end{aligned} \tag{F.9}
$$

Once again, the proof (by induction) relies on the formula of the derivatives of Bernoulli polynomials. Then, taking $k = 2r$ in Eq. (F.9) yields:

$$
\boxed{J_-(x) = \sum_{l=0}^{2r-1} \frac{(2r)!}{(2r-l)!} \Big[B_{2r-l}(x-t)\, \phi^{[2r-l-1]}(t) \Big]_{t=0}^{x} + (2r)! \int_0^x \phi(\xi)\, \mathrm{d}\xi} \ . \tag{F.10}
$$

Integration by parts to transform $J_+(x)$

It can be shown that:

$$\forall\, 1 \le k \le 2r, \quad J_+(x) = \sum_{l=0}^{k-1} (-1)^l \, \frac{(2r)!}{(2r-l)!} \left[B_{2r-l}(t-x)\, \phi^{[2r-l-1]}(t) \right]_{t=x}^{1} \quad \dots$$
$$+ (-1)^k \, \frac{(2r)!}{(2r-k)!} \int_x^1 B_{2r-k}(\xi - x)\, \phi^{[2r-k]}(\xi)\, \mathrm{d}\xi \; . \tag{F.11}$$

The proof (by induction) is similar to what was done for $J_-(x)$. Then, taking $k = 2r$ in Eq. (F.11) yields:

$$\boxed{J_+(x) = \sum_{l=0}^{2r-1} (-1)^l \, \frac{(2r)!}{(2r-l)!} \left[B_{2r-l}(t-x)\, \phi^{[2r-l-1]}(t) \right]_{t=x}^{1} + (2r)! \int_x^1 \phi(\xi)\, \mathrm{d}\xi} \; . \tag{F.12}$$

The summation of Eq. (F.10) and (F.12) leads to:

$$J_-(x) + J_+(x) = \sum_{l=0}^{2r-1} \frac{(2r)!}{(2r-l)!} \left(\left[B_{2r-l}(x-t)\, \phi^{[2r-l-1]}(t) \right]_{t=0}^{x} \quad \dots \right.$$
$$\left. + (-1)^l \left[B_{2r-l}(t-x)\, \phi^{[2r-l-1]}(t) \right]_{t=x}^{1} \right) + \int_0^1 \phi(\xi)\, \mathrm{d}\xi \tag{F.13}$$
$$= \sum_{k=1}^{2r} \frac{(2r)!}{k!} \left(\left[B_k(x-t)\, \phi^{[k-1]}(t) \right]_{t=0}^{x} + (-1)^{2r-k} \left[B_k(t-x)\, \phi^{[k-1]}(t) \right]_{t=x}^{1} \right) \tag{F.14}$$
$$= \sum_{k=1}^{2r} \frac{(2r)!}{k!} \left(- B_k(x)\, \phi^{[k-1]}(0) + (-1)^k \, B_k(1-x)\, \phi^{[k-1]}(1) \right) \quad \dots$$
$$+ \sum_{k=1}^{2r} \frac{(2r)!}{k!} \left(B_k(0)\, \phi^{[k-1]}(x) - (-1)^k \, B_k(0)\, \phi^{[k-1]}(x) \right)$$
$$= S_1(x) + S_2(x) \; . \tag{F.15}$$

To switch from Eq. (F.13) to Eq. (F.14), one needs to perform the index shift $k = 2r - l$ and to remember that ϕ has zero mean. In addition:

- With $(\mathcal{C}_1)$ and the symmetry properties of Bernoulli polynomials (see Appendix A.1.4), the sum $S_1(x)$ can be simplified:

$$\boxed{S_1(x) = \sum_{k=1}^{r} \frac{(2r)!}{k!} \left(\phi^{[k-1]}(1) - \phi^{[k-1]}(0) \right) B_k(x)} \; . \tag{F.16}$$

- With the boundary values of Bernoulli polynomials (see Appendix A.1.4), one can see that the sum $S_2(x)$ is actually composed of only one non-zero term:

$$\boxed{S_2(x) = -(2r)!\, \phi(x)} \; . \tag{F.17}$$

After bringing together Eq. (F.6), (F.8), (F.15), (F.16) and (F.17), one has:

$$\left[T_{k^r_{\mathrm{Sob}}} \, \phi^{[2r]} \right](x) = \sum_{k=1}^{r} \frac{I_k}{(k!)^2} \, B_k(x) + (-1)^{r+1} \sum_{k=1}^{r} \frac{1}{k!} \left(\phi^{[k-1]}(1) - \phi^{[k-1]}(0) \right) B_k(x) + (-1)^r \, \phi(x)$$

$$= \sum_{k=1}^{r} \frac{1}{k!} \left[\frac{I_k}{k!} + (-1)^{r+1} \left(\phi^{[k-1]}(1) - \phi^{[k-1]}(0) \right) \right] B_k(x) + (-1)^r \, \phi(x)$$
$$= \sum_{k=1}^{r} \frac{\alpha_k}{k!} \, B_k(x) + (-1)^r \, \phi(x) \ .$$

The next step is to demonstrate that $\alpha_1 = \ldots = \alpha_r = 0$.

Nullity of α_r
Given the definition of α_r, it follows that:

$$\begin{aligned}
\alpha_r &= \frac{I_r}{r!} + (-1)^{r+1} \left(\phi^{[r-1]}(1) - \phi^{[r-1]}(0) \right) \\
&= \sum_{l=0}^{r-1} \frac{(-1)^l}{(r-l)!} \left(B_{r-l}(1) \, \phi^{[2r-l-1]}(1) - B_{r-l}(0) \, \phi^{[2r-l-1]}(0) \right) \ \ldots \\
&\qquad + (-1)^r \left(\phi^{[r-1]}(1) - \phi^{[r-1]}(0) \right) + (-1)^{r+1} \left(\phi^{[r-1]}(1) - \phi^{[r-1]}(0) \right) && \text{(F.18)} \\
&= \sum_{l=0}^{r-2} \frac{(-1)^l}{(r-l)!} \left(B_{r-l}(1) \, \phi^{[2r-l-1]}(1) - B_{r-l}(0) \, \phi^{[2r-l-1]}(0) \right) \ \ldots \\
&\qquad + (-1)^{r-1} \left(B_1(1) \, \phi^{[r]}(1) - B_1(0) \, \phi^{[r]}(0) \right) && \text{(F.19)} \\
&= \sum_{\substack{0 \le l \le r-2 \\ r-l = 2l'}} \frac{(-1)^l}{(r-l)!} \left(B_{r-l}(1) \, \phi^{[2r-l-1]}(1) - B_{r-l}(0) \, \phi^{[2r-l-1]}(0) \right) \ \ldots \\
&\qquad + \sum_{\substack{0 \le l \le r-2 \\ r-l = 2l'+1}} \frac{(-1)^l}{(r-l)!} \left(B_{r-l}(1) \, \phi^{[2r-l-1]}(1) - B_{r-l}(0) \, \phi^{[2r-l-1]}(0) \right) = 0 \ . && \text{(F.20)}
\end{aligned}$$

Note that:

- Eq. (F.18) is obtained after replacing I_r by the sum given in Eq. (F.8). Then, the term indexed by $l = r$ is taken out of the sum and it is exactly the opposite of the last term.
- Eq. (F.19) is obtained after taking the term indexed by $l = r - 1$ out of the sum. In fact, this term is equal to zero because of $(\mathcal{C}_1)$.
- In Eq. (F.20), the sum indexed by even integers is equal to zero for two reasons. Firstly, all the associated Bernoulli polynomials have equal boundary values (see Appendix A.1.4). Secondly, all the functions $\phi^{[2r-l-1]}$ have equal boundary values according to $(\mathcal{C}_3)$.
- In Eq. (F.20), the sum indexed by odd integers is equal to zero because all the associated Bernoulli polynomials have zero boundary values (see Appendix A.1.4).

Nullity of α_k with $k \in \{1, \ldots, r-1\}$
Given the definition of α_k, it follows that:

$$\alpha_k = \frac{I_k}{k!} + (-1)^{r+1} \left(\phi^{[k-1]}(1) - \phi^{[k-1]}(0) \right)$$

$$= \sum_{l=0}^{k-1} \frac{(-1)^l}{(k-l)!} \left(B_{k-l}(1)\, \phi^{[2r-l-1]}(1) - B_{k-l}(0)\, \phi^{[2r-l-1]}(0) \right) \; \dots$$
$$+ (-1)^k \left(\phi^{[2r-k-1]}(1) - \phi^{[2r-k-1]}(0) \right) + (-1)^{r+1} \left(\phi^{[k-1]}(1) - \phi^{[k-1]}(0) \right) \tag{F.21}$$
$$= \sum_{l=0}^{k-2} \frac{(-1)^l}{(k-l)!} \left(B_{k-l}(1)\, \phi^{[2r-l-1]}(1) - B_{k-l}(0)\, \phi^{[2r-l-1]}(0) \right) \; \dots$$
$$+ (-1)^{k-1} \left(B_1(1)\, \phi^{[2r-k]}(1) - B_1(0)\, \phi^{[2r-k]}(0) \right) + (-1)^{r+1} \left(\phi^{[r-1]}(1) - \phi^{[r-1]}(0) \right) \tag{F.22}$$
$$= \sum_{\substack{0 \le l \le r-2 \\ k-l=2l'}} \frac{(-1)^l}{(k-l)!} \left(B_{k-l}(1)\, \phi^{[2r-l-1]}(1) - B_{k-l}(0)\, \phi^{[2r-l-1]}(0) \right) \; \dots$$
$$+ \sum_{\substack{0 \le l \le r-2 \\ k-l=2l'+1}} \frac{(-1)^l}{(k-l)!} \left(B_{k-l}(1)\, \phi^{[2r-l-1]}(1) - B_{k-l}(0)\, \phi^{[2r-l-1]}(0) \right) = 0 \; . \tag{F.23}$$

Note that:

- Eq. (F.21) is obtained after replacing I_j by the sum given in Eq. (F.8). Then, the term indexed by $l = r$ is taken out of the sum. This term is always equal to zero because of $(\mathcal{C}_1)$ when $k = r - 1$ and because of $(\mathcal{C}_3)$ when $1 \le k \le r - 2$.
- Eq. (F.22) is obtained after taking the term indexed by $l = r - 1$ out of the sum. The entire expression in the second line of Eq. (F.22) is equal to zero because of $(\mathcal{C}_2)$ and $(\mathcal{C}_3)$.
- In Eq. (F.23), the sum indexed by even integers is equal to zero for two reasons. Firstly, all the associated Bernoulli polynomials have equal boundary values (see Appendix A.1.4). Secondly, all the functions $\phi^{[2r-l-1]}$ have equal boundary values according to $(\mathcal{C}_3)$.
- In Eq. (F.23), the sum indexed by odd integers is equal to zero because all the associated Bernoulli polynomials have zero boundary values (see Appendix A.1.4).

As a conclusion, having $\alpha_1 = \ldots = \alpha_r = 0$ finally leads to $T_{k^r_{\text{Sob}}}\, \phi^{[2r]=(-1)^r\, \phi}$ and then to $T_{k^r_{\text{Sob}}}\, \phi = \lambda\, \phi$.

F.4. **Proof of Proposition 7.3**

F.4.1. *Sum of kernels*

In this proof, a fundamental theorem regarding the RKHS induced by the sum of two kernels is required.

Theorem F.3. *Let $\mathcal{X} \subseteq \mathbb{R}$ be an interval. Let $K_1 : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ and $K_2 : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ be two kernels with respective RKHSs denoted by $\mathcal{H}_1$ and $\mathcal{H}_2$. The RKHS $\mathcal{H}_{12}$ induced by the kernel $K_1 + K_2$ is:*

$$\mathcal{H}_{12} = \mathcal{H}_1 + \mathcal{H}_2 = \left\{ h \in \mathbb{R}^{[0,1]} \;\middle|\; h = h_1 + h_2 \;\; \textit{with} \;\; \begin{array}{c} h_1 \in \mathcal{H}_1 \\ h_2 \in \mathcal{H}_2 \end{array} \right\} .$$

If $\mathcal{H}_1 \cap \mathcal{H}_2 = \{0\}$, one must write $\mathcal{H}_{12} = \mathcal{H}_1 \oplus \mathcal{H}_2$ and the inner product is:

$$\begin{array}{ccccccc} \langle \cdot, \cdot \rangle_{\mathcal{H}_{12}} : & \mathcal{H}_{12} & \times & \mathcal{H}_{12} & \longrightarrow & \mathbb{R} \\ & (f = f_1 + f_2 & , & g = g_1 + g_2) & \longmapsto & \langle f_1, g_1 \rangle_{\mathcal{H}_1} + \langle f_2, g_2 \rangle_{\mathcal{H}_2} \; . \end{array}$$

F.4.2. *A few more details on the sub-kernel decomposition*

The unanchored Sobolev space (of order r) is the RKHS obtained when $\mathrm{H}^r([0\,,1])$ is equipped with the inner product $\langle\cdot,\cdot\rangle_{\mathcal{H}^r_{\mathrm{Sob}}}$ defined in Eq. (3.6). This RKHS is denoted by $\mathcal{H}^r_{\mathrm{Sob}}$ and its reproducing kernel is:

$$K^r_{\mathrm{Sob}}(x,x') = 1 + k^r_{\mathrm{Sob}}(x,x') = 1 + \sum_{k=1}^{r} \frac{B_k(x)B_k(x')}{(k!)^2} + \frac{(-1)^{r+1}}{(2r)!} B_{2r}(|x-x'|) = 1 + k^r_A(x,x') + k^r_B(x,x')\ .$$

Preliminary results on the two kernels k^r_A and k^r_B need to be stated.

Lemma F.4. *For any $r \geq 1$, the following statements are true:*

(a) k^r_A and k^r_B are orthogonal kernels.

(b) The RKHS $\mathcal{F}^r_A$ induced by k^r_A is:

$$\mathcal{F}^r_A = \left\{ f \in \mathbb{R}^{[0,1]} \;\middle|\; f(\cdot) = \sum_{k=1}^{r} a_k\, \widetilde{B}_k(\cdot) \ \ \textit{with}\ \ (a_k)_{1\leq k \leq r} \in \mathbb{R}^r \right\},$$

with inner product:

$$\begin{array}{cccccc} \langle\cdot,\cdot\rangle_{\mathcal{F}^r_A} : & \mathcal{F}^r_A & \times & \mathcal{F}^r_A & \longrightarrow & \mathbb{R} \\ & \Bigg(f_1(\cdot) = \displaystyle\sum_{k=1}^{r} a_k\, \widetilde{B}_k(\cdot) & , & f_2(\cdot) = \displaystyle\sum_{k=1}^{r} \alpha_k\, \widetilde{B}_k(\cdot) \Bigg) & \longmapsto & \displaystyle\sum_{k=1}^{r} a_k\, \alpha_k\ . \end{array}$$

(c) The RKHS $\mathcal{F}^r_B$ induced by k^r_B is:

$$\mathcal{F}^r_B = \left\{ f \in \mathbb{R}^{[0,1]} \;\middle|\; f(\cdot) = \sum_{k=1}^{\infty} p_k\, \tilde{c}^r_{2k}(\cdot) + q_k\, \tilde{s}^r_{2k}(\cdot) \ \ \textit{with}\ \ \begin{array}{c} (p_k)_{k\geq 1} \in \ell^2(\mathbb{N}^*) \\ (q_k)_{k\geq 1} \in \ell^2(\mathbb{N}^*) \end{array} \right\},$$

with inner product:

$$\begin{array}{cccccc} \langle\cdot,\cdot\rangle_{\mathcal{F}^r_B} : & \mathcal{F}^r_B & \times & \mathcal{F}^r_B & \longrightarrow & \mathbb{R} \\ & \Bigg(f_1(\cdot) = \displaystyle\sum_{k=1}^{\infty} p_k\, \tilde{c}^r_{2k}(\cdot) + q_k\, \tilde{s}^r_{2k}(\cdot) & , & f_2(\cdot) = \displaystyle\sum_{k=1}^{\infty} \pi_k\, \tilde{c}^r_{2k}(\cdot) + \rho_k\, \tilde{s}^r_{2k}(\cdot) & \longmapsto & \displaystyle\sum_{k=1}^{\infty} p_k\, \pi_k + q_k\, \rho_k\ . \end{array}$$

(d) $\mathcal{F}^r_A \cap \mathcal{F}^r_B = \{0\}$.

Proof. The statement (a) can be proved very easily by using the zero-mean property of Bernoulli polynomials (see Appendix A.1.6). As regards the statements (b) and (c), they are direct consequences of Theorems 2.23 and 2.21 respectively. Only the statement (d) deserves further clarifications. For any $f \in \mathcal{F}^r_A \cap \mathcal{F}^r_B$, it must be proved that $f = 0$. Since $f \in \mathcal{F}^r_A$, one can write:

$$f(\cdot) = \sum_{i=1}^{r} a_i\, \frac{B_i(\cdot)}{i!} \quad \text{with} \quad (a_i)_{1\leq i\leq r} \in \mathbb{R}^r\ .$$

Now, let us see why all coefficients are zero. First, let us assume that $a_1 \neq 0$ for the sake of contradiction. It is said in Appendix A.1.4 that $B_1(1) = -B_1(0) = 1/2$ and $B_k(1) = B_k(0)$ for all $k \geq 2$ and this leads to

$f(1) - f(0) = a_1 \neq 0$. Therefore, f cannot belong to $\mathcal{F}_B^r$ because it only contains 1-periodic functions. Since there is a contradiction with $f \in \mathcal{F}_B^r$, it must be that $a_1 = 0$. Then, let us write f in the formalism of $\mathcal{F}_B^r$:

$$\begin{aligned}
f(x) &= \sum_{j=2}^{r} \frac{a_j}{j!}\, B_j(x) \qquad \text{since } a_1 = 0 \;, \\
&= \sum_{j=2}^{r} \frac{a_j}{j!} \left(-2\,(j!) \sum_{k=1}^{\infty} \frac{\cos\left(2k\pi x - j\pi/2\right)}{(2k\pi)^j} \right) \qquad \text{with the Fourier series expansion of } B_j \;, \\
&= \sum_{j=2}^{r} \sum_{k=1}^{\infty} \frac{a_j}{(2k\pi)^j}\, \epsilon_k(x) \qquad \text{with } \epsilon_k := \begin{cases} -c_{2k} & \text{if } j = 0 \mod 4 \\ -s_{2k} & \text{if } j = 1 \mod 4 \\ c_{2k} & \text{if } j = 2 \mod 4 \\ s_{2k} & \text{if } j = 3 \mod 4 \end{cases} \\
&= \sum_{k=1}^{\infty} \left(\sum_{j=2}^{r} a_j\, (2k\pi)^{r-j} \right) \frac{\epsilon_k(x)}{(2k\pi)^r} = \sum_{k=1}^{\infty} \gamma_k\, \frac{\epsilon_k(x)}{(2k\pi)^r} \;.
\end{aligned}$$

The question is now to examine which conditions (on $a_2, \dots, a_r$) are required so that the sequence $(\gamma_k)_{k\geq 1}$ is square summable. A simple factorization allows to see that:

$$\gamma_k = \sum_{j=2}^{r} a_j\, (2k\pi)^{r-j} = (2k\pi)^{r-2} \sum_{l=0}^{r-2} \frac{a_{l+2}}{(2k\pi)^l} \underset{k\to\infty}{\sim} \begin{cases} a_2\, (2k\pi)^{r-2} & \text{if } a_2 \neq 0 \\ a_3\, (2k\pi)^{r-3} & \text{if } a_2 = 0 \text{ but } a_3 \neq 0 \\ \vdots & \vdots \\ a_r & \text{if } a_2 = \ldots = a_{r-1} = 0 \text{ but } a_r \neq 0 \\ 0 & \text{if } a_2 = \ldots = a_{r-1} = a_r = 0 \;. \end{cases}$$

The only way to make $(\gamma_k)_{k\geq 1}$ be square summable is thus to take $a_2 = \ldots = a_r = 0$. This leads to $f = 0$. $\square$

Knowing (a), one has $\mathbb{R} \cap \mathcal{F}_A^r = \{0\}$ and $\mathbb{R} \cap \mathcal{F}_B^r = \{0\}$. Since (d) also ensures that $\mathcal{F}_A^r \cap \mathcal{F}_B^r = \{0\}$, the sum of the three sub-RKHSs is direct. With the help of Theorem F.3, it is then straightforward to obtain:

$$\boxed{\mathrm{H}^r([0\,,1]) = \mathbb{R} \oplus \mathcal{F}_A^r \oplus \mathcal{F}_B^r} \;. \tag{F.24}$$

F.4.3. *Proof of the main result*

It must be proved that the system $(g_k^r)_k$ defined by Eq. (7.9) is truly ℓ^2-linearly independent. Let us consider a square-summable sequence $(\gamma_k)_k$ which is rearranged in the following way:

$$(\gamma_k)_k := \left\{\; (a_k)_{0\leq k\leq r} \quad ; \quad (p_k)_{k\geq 1} \quad ; \quad (q_k)_{k\geq 1} \;\right\} \;.$$

This rearrangement allows for a better matching of the coefficients in $(\gamma_k)_k$ and the functions in $(g_k^r)_k$. Assuming that $(\gamma_k)_k$ is square summable amounts to assuming that the subsequences $(p_k)_{k\geq 1}$ and $(q_k)_{k\geq 1}$ are both in $\ell^2(\mathbb{N}^*)$. Note that the series defined as:

$$S(\cdot) = \underbrace{a_0}_{\in\mathbb{R}} + \underbrace{\sum_{k=1}^{r} a_k\, \widetilde{B}_k(\cdot)}_{\in\mathcal{F}_A^r} + \underbrace{\sum_{k=1}^{\infty} p_k\, \tilde{c}_{2k}^r(\cdot) + q_k\, \tilde{s}_{2k}^r(\cdot)}_{\in\mathcal{F}_B^r} = a_0 + f_A^r(\cdot) + f_B^r(\cdot)$$

belongs to $\mathrm{H}^r([0\,,1])$ in virtue of Eq. (F.24).

Now, let us assume that the square-summable sequence $(\gamma_k)_k$ is such that $S = 0$. As the zero function belongs simultaneously to $\mathbb{R}$, $\mathcal{F}_A^r$ and $\mathcal{F}_B^r$, one can write:

$$S(\cdot) = 0 + 0 + 0 = a_0 + f_A^r(\cdot) + f_B^r(\cdot)$$

Because of the direct sum, the decomposition of S into a sum of three sub-functions belonging respectively to $\mathbb{R}$, $\mathcal{F}_A^r$ and $\mathcal{F}_B^r$ is unique. This leads to:

$$a_0 = 0 \ ,$$

$$f_A^r(\cdot) = \sum_{k=1}^{r} a_k \, \widetilde{B}_k(\cdot) = 0 \ , \tag{F.25}$$

$$f_B^r(\cdot) = \sum_{k=1}^{\infty} p_k \, \tilde{c}_{2k}^r(\cdot) + q_k \, \tilde{s}_{2k}^r(\cdot) = 0 \ . \tag{F.26}$$

Eq. (F.25) implies $a_1 = \ldots = a_r = 0$ because the polynomials $(\widetilde{B}_k)_{1 \le k \le r}$ are linearly independent. Furthermore, after turning back to the definition of the sinusoidal features, Eq. (F.26) may be rewritten in the following way:

$$\sum_{k=1}^{\infty} \frac{p_k}{(2k\pi)^r} \, c_{2k}(\cdot) + \frac{q_k}{(2k\pi)^r} \, s_{2k}(\cdot) = 0 \ . \tag{F.27}$$

Remember that the system $\{\mathbf{1} \, ; (c_{2k})_{k \ge 1} \, ; (s_{2k})_{k \ge 1}\}$ is the Fourier basis of $\mathrm{L}^2([0\,,1])$ and is therefore ℓ^2-linearly independent (because of Parseval's identity). Additionally, as the sequences with general terms $p_k/(2k\pi)^r$ and $q_k/(2k\pi)^r$ are in $\ell^2(\mathbb{N}^*)$, Eq. (F.27) cannot be verified unless $p_k = 0$ and $q_k = 0$ for all $k \ge 1$. Finally, $(\gamma_k)_k$ is only composed of zero coefficients. This proves that the system $(g_k^r)_k$ is ℓ^2-linearly independent.

F.5. **Additional details for Section 7.3**

Let $K : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ be a Mercer kernel, let ν be a probability measure with support $\mathcal{X}$ and let T_K be the integral operator built from K and ν. Just as in Theorem 2.18, $(\lambda_i)_{i \ge 1}$ and $(\phi_i)_{i \ge 1}$ respectively denote the eigenvalues of T_K and an ONB of $\mathrm{L}^2(\mathcal{X}, \nu)$ only composed of eigenfunctions of T_K. In addition, let us assume that there also exists a series expansion of K based on a system $(g_i)_{i \ge 1}$ of non-orthogonal but ℓ^2-linearly independent features. In light of these assumptions, K can be decomposed in two different ways:

$$\forall\, x, x' \in \mathcal{X}, K(x, x') = \sum_{i \ge 1} \lambda_i \, \phi_i(x) \, \phi_i(x') = \sum_{i \ge 1} g_i(x) \, g_i(x') \ . \tag{F.28}$$

Theorem 2.21 and Theorem 2.23 ensure that $(\sqrt{\lambda_i}\, \phi_i)_{i \ge 1}$ and $(g_i)_{i \ge 1}$ are two possible ONBs of $\mathcal{H}$. However, as the functions $(g_i)_{i \in I}$ are not mutually orthogonal (in the L^2-sense), they cannot be eigenfunctions of T_K. The difference between the two families of basis functions can be summarized by the following inequality:

$$\|T_K\|_{\mathrm{HS}}^2 \quad = \quad \sum_{i \ge 1} \lambda_i^2 \quad \ge \quad \sum_{i \ge 1} \|g_i\|_{\mathrm{L}^2}^4 \ ,$$

with equality holding if and only if the system $(g_i)_{i \ge 1}$ is L^2-orthogonal. Let us demonstrate this result.

For the first equality, just write the Hilbert-Schmidt norm of T_K with the ONB of eigenfunctions $(\phi_i)_{i \ge 1}$:

$$\|T_K\|_{\mathrm{HS}}^2 = \sum_{i \ge 1} \sum_{j \ge 1} |\langle T_K \, e_i \, , e_j \rangle_{\mathrm{L}^2}|^2 = \sum_{i \ge 1} \lambda_i^2 \ .$$

Then, remember that it was stated in Eq. (2.10) that $\|T_K\|_{\mathrm{HS}}^2 = \|K\|_{\mathrm{L}^2}^2$. If using the second decomposition in Eq. (F.28), $\|K\|_{\mathrm{L}^2}^2$ can be computed differently:

$$\begin{aligned}
\|K\|_{\mathrm{L}^2}^2 = \left\| \sum_{i\geq 1} g_i \otimes g_i \right\|_{\mathrm{L}^2}^2 &= \left\langle \sum_{i\geq 1} g_i \otimes g_i, \sum_{j\geq 1} g_j \otimes g_j \right\rangle_{\mathrm{L}^2} \\
&= \sum_{i\geq 1}\sum_{j\geq 1} \langle g_i \otimes g_i, g_j \otimes g_j \rangle_{\mathrm{L}^2} && \text{with the dominated convergence theorem,} \\
&= \sum_{i\geq 1}\sum_{j\geq 1} \langle g_i, g_j \rangle_{\mathrm{L}^2}^2 && \text{by definition of the inner product in } \mathrm{L}^2(\mathcal{X}^2, \nu^{\otimes 2}), \\
&\geq \sum_{i\geq 1} \|g_i\|_{\mathrm{L}^2}^4 && \text{by only taking the terms such that } i = j.
\end{aligned}$$

Before using Lebesgue's dominated convergence theorem (to switch from the first to the second line), it must be proved that the absolute value of the integrand belongs to $\mathrm{L}^1(\mathcal{X}^2, \nu^{\otimes 2})$:

$$\begin{aligned}
\int_{\mathcal{X}\times\mathcal{X}} \left(\sum_{i\geq 1} g_i(x)\, g_i(\xi) \right)^2 \mathrm{d}\nu(x)\, \mathrm{d}\nu(\xi) &\leq \int_{\mathcal{X}\times\mathcal{X}} \left(\sum_{i\geq 1} g_i(x)^2 \right) \left(\sum_{i\geq 1} g_i(\xi)^2 \right) \mathrm{d}\nu(x)\, \mathrm{d}\nu(\xi) \\
&= \int_{\mathcal{X}\times\mathcal{X}} K(x,x)\, K(\xi,\xi)\, \mathrm{d}\nu(x)\, \mathrm{d}\nu(\xi) = \left(\int_{\mathcal{X}} K(x,x)\, \mathrm{d}\nu(x) \right)^2 < \infty \ .
\end{aligned}$$

The first line is obtained by applying the Cauchy-Schwarz inequality in $\ell^2(\mathbb{N}^*)$. The two equalities in the second line stem from Eq. (F.28) and Fubini's theorem. The final quantity is indeed finite because K is a Mercer kernel. Eventually, one has $\|T_K\|_{\mathrm{HS}}^2 = \sum_{i\geq 1} \lambda_i^2 = \|K\|_{\mathrm{L}^2}^2 \geq \sum_{i\geq 1} \|g_i\|_{\mathrm{L}^2}^4$.

F.6. **Proof of Proposition 8.1**

It must be proved that:

(a) The sequence $(k_A^r)_{r\geq 1}$ converges uniformly to the continuous kernel $k_A^\infty = \sum_{k\geq 1} \widetilde{B}_k \otimes \widetilde{B}_k$.

(b) The sequence $(k_A^r)_{r\geq 1}$ converges uniformly to $k_B^\infty = 0$.

For the two convergence results, the idea of the proof is the same. The difference between the reproducing kernel of order r and the proposed asymptotic kernel must be uniformly bounded on $[0\,,1]^2$ by a constant $C(r)$ converging to zero as r tends to ∞. This may be summarized by:

$$\forall\, x, x' \in [0\,,1], \quad |k_\star^r(x,x') - k_\star^\infty(x,x')| \leq C(r) \xrightarrow[r\to\infty]{} 0 \ ,$$

where $\star$ is either A or B.

Proof of (a)

First, it must be checked that the series k_A^∞ is convergent for any point $(x,x') \in [0\,,1]$. This can be easily achieved with the help of Eq. (8.3):

$$|k_A^\infty(x,x')| = \left| \sum_{k=1}^{\infty} \frac{B_k(x)\, B_k(x')}{(k!)^2} \right| \leq \sum_{k=1}^{\infty} \left(\frac{M_k^+}{k!} \right)^2 \leq 16 \sum_{k=1}^{\infty} \left(\frac{1}{2\pi} \right)^{2k} = \frac{16}{4\pi^2 - 1} < \infty \ .$$

The function k_A^∞ is thus well-defined everywhere on $[0\,,1]^2$. In addition, the remainder of the series can be uniformly bounded:

$$|k_A^\infty(x,x') - k_A^r(x,x')| = \left|\sum_{k=r+1}^{\infty} \frac{B_k(x)\,B_k(x')}{(k!)^2}\right| \leq \sum_{k=r+1}^{\infty} \left(\frac{M_k^+}{k!}\right)^2 \leq \frac{16}{4\pi^2-1}\left(\frac{1}{2\pi}\right)^{2r} \xrightarrow[r\to\infty]{} 0\ . \tag{F.29}$$

This justifies that $(k_A^r)_{r\geq 1}$ uniformly converges to k_A^∞ on $[0\,,1]^2$. It is trivial to see that k_A^∞ is a kernel because it is expressed as a convergent series of symmetric and separable functions. More precisely, k_A^r is a continuous kernel because of the uniform limit theorem.

Proof of (b)
The same kind of reasoning must be applied but it is even simpler because there is only one single Bernoulli polynomial to bound:

$$|k_B^r(x,x')| = \frac{1}{(2r)!}\left|B_{2r}(|x-x'|)\right| \leq \frac{M_{2r}^+}{(2r)!} \leq 4\left(\frac{1}{2\pi}\right)^{2r} \xrightarrow[r\to\infty]{} 0\ .$$

This justifies that $(k_B^r)_{r\geq 1}$ uniformly converges to the zero kernel on $[0\,,1]^2$.

F.7. Proof of Proposition 8.2

First, it must be noted that the series $\sum_{k\geq 0} a_k\,\widetilde{B}_k(\cdot)$ is uniformly convergent for any sequence $(a_k)_{k\geq 0} \in \ell^2(\mathbb{N})$. Indeed, with the Cauchy-Schwarz inequality and the same upper bounding technique as in Eq. (F.29), one has:

$$\begin{aligned}\left|\sum_{k=0}^{r} a_k\,\widetilde{B}_k(x) - \sum_{k=0}^{\infty} a_k\,\widetilde{B}_k(x)\right| = \left|\sum_{k=r+1}^{\infty} a_k\,\widetilde{B}_k(x)\right| &\leq \left(\sum_{k=r+1}^{\infty} a_k^2\right)^{1/2}\left(\sum_{k=r+1}^{\infty} \widetilde{B}_k(x)^2\right)^{1/2}\\ &\leq \|(a_k)_{k\geq 0}\|_{\ell^2}\left[\sum_{k=r+1}^{\infty}\left(\frac{M_k^+}{k!}\right)^2\right]^{1/2}\\ &\leq \|(a_k)_{k\geq 0}\|_{\ell^2}\,\frac{4}{\sqrt{4\pi^2-1}}\left(\frac{1}{2\pi}\right)^{r} \xrightarrow[r\to\infty]{} 0\ .\end{aligned}$$

Now, let us assume that there exist a sequence $(a_k)_{k\geq 0} \in \ell^2(\mathbb{N})$ such that $S(\cdot) := \sum_{k\geq 0} a_k\,\widetilde{B}_k(\cdot) = 0$. Under this assumption, the mean value of the function $S : [0\,,1] \to \mathbb{R}$ is equal to zero. The zero-mean property of Bernoulli polynomials (see Appendix A.1.6) leads to:

$$0 = \int_0^1 S(x)\,\mathrm{d}x = \int_0^1 \left(\sum_{k=0}^{\infty} a_k\,\widetilde{B}_k(x)\right)\mathrm{d}x = \sum_{k=0}^{\infty} a_k \left(\int_0^1 \widetilde{B}_k(x)\,\mathrm{d}x\right) = a_0\ . \tag{F.30}$$

In the above equation, the integral over $[0\,,1]$ and the summation over $\mathbb{N}$ are freely interchanged because the series converges uniformly on $[0\,,1]$. With Eq. (F.30), one has $a_0 = 0$. Let us now explain why all remaining coefficients $(a_i)_{i\geq 1}$ are also zero. As S is a uniformly convergent series of polynomials (defined on a bounded interval), it can be easily justified that $S \in \mathrm{C}^\infty([0\,,1])$. In addition, the derivatives of S can be computed through term-by-term differentiation:

$$\forall\, i\geq 1,\quad S^{[i]}(\cdot) = \sum_{k=0}^{\infty} a_k\,\widetilde{B}_k^{[i]}(\cdot) = \sum_{k=i}^{\infty} a_k\,\widetilde{B}_{k-i}(\cdot) = \sum_{k=0}^{\infty} a_{k+i}\,\widetilde{B}_k(\cdot) = 0\ .$$

Just as for Eq. (F.30), the mean value of $S^{[i]}$ (for any $i \geq 1$) is equal to zero and this yields $a_i = 0$. Therefore, the nullity of $\sum_{k\geq 0} a_k \, \widetilde{B}_k(\cdot)$ implies the nullity of the coefficients $(a_k)_{k\geq 0}$. This proves that the system $(\widetilde{B}_k)_{k\geq 1}$ is ℓ^2-linearly independent.

Appendix G. Supplementary material

G.1. **Proof of Theorem 5.1**

It only remains to prove that $\left[T_{k^1_{\mathrm{Sob}}} c_k\right](\cdot) = c_k(\cdot)/(k\pi)^2$ for any $k \geq 1$. This can be done by simply calculating the integral $\left[T_{k^1_{\mathrm{Sob}}} c_k\right](x)$ for a given point $x \in [0\,,1]$. This is not particularly difficult but the calculations are tedious and they must be carried out patiently. The main calculation steps are provided below. First, one has:

$$\begin{aligned}\left[T_{k^1_{\mathrm{Sob}}} c_k\right](x) &= \int_0^1 k^1_{\mathrm{Sob}}(x,\xi)\, c_k(\xi)\,\mathrm{d}\xi \\ &= \int_0^1 \left(B_1(x)\,B_1(\xi) + \frac{1}{2}\,B_2(|x-\xi|)\right) c_k(\xi)\,\mathrm{d}\xi \\ &= \sqrt{2}\left(x-\frac{1}{2}\right)\int_0^1 \left(\xi - \frac{1}{2}\right)\cos(k\pi\xi)\,\mathrm{d}\xi + \frac{1}{\sqrt{2}}\int_0^1 \left[(x-\xi)^2 - |x-\xi| + \frac{1}{6}\right]\cos(k\pi\xi)\,\mathrm{d}\xi \\ &= \sqrt{2}\,\alpha\left(x-\frac{1}{2}\right) + \frac{1}{\sqrt{2}}\,\beta(x) \quad \text{after denoting by } \alpha \text{ and } \beta(x) \text{ the two integrals.}\end{aligned} \tag{G.1}$$

It can be proved that $\boxed{\alpha = \frac{1}{(k\pi)^2}\,\left[(-1)^k - 1\right]}$.

Then, the integral $\beta(x)$ may be divided into three terms:

$$\begin{aligned}\beta(x) &= \int_0^1 \left[(x-\xi)^2 - |x-\xi| + \frac{1}{6}\right]\cos(k\pi\xi)\,\mathrm{d}\xi \\ &= \int_0^1 (x-\xi)^2\,\cos(k\pi\xi)\,\mathrm{d}\xi - \int_0^1 |x-\xi|\,\cos(k\pi\xi)\,\mathrm{d}\xi + \frac{1}{6}\int_0^1 \cos(k\pi\xi)\,\mathrm{d}\xi \\ &= \beta_2(x) - \beta_1(x) + \frac{1}{6}\,\beta_0 \quad \text{after denoting by } \beta_0 \text{ and } \beta_1(x) \text{ and } \beta_2(x) \text{ the three integrals.}\end{aligned} \tag{G.2}$$

It can be proved that $\boxed{\beta_2(x) = \frac{2}{(k\pi)^2}\,\left[(-1)^k\,(1-x) + x\right]}$ and $\boxed{\beta_0 = 0}$.

To eliminate the absolute value in the integral expression of $\beta_1(x)$, one may write:

$$\begin{aligned}\beta_1(x) &= \int_0^1 |x-\xi|\,\cos(k\pi\xi)\,\mathrm{d}\xi \\ &= \int_0^x (x-\xi)\,\cos(k\pi\xi)\,\mathrm{d}\xi + \int_x^1 (\xi - x)\,\cos(k\pi\xi)\,\mathrm{d}\xi \\ &= \beta_1^-(x) + \beta_1^+(x) \quad \text{after denoting by } \beta_1^-(x) \text{ and } \beta_1^+(x) \text{ the two integrals.}\end{aligned} \tag{G.3}$$

It can be proved that $\boxed{\beta_1^-(x) = \frac{1}{(k\pi)^2}\,[1-\cos(k\pi x)]}$ and $\boxed{\beta_1^+ = \frac{1}{(k\pi)^2}\,\left[(-1)^k - \cos(k\pi x)\right]}$.

With the analytical expressions of α, β_0, $\beta_1^-(x)$, $\beta_1^+(x)$ and $\beta_2(x)$, Eq. (G.1), (G.2) and (G.3) yield:

$$\left[T_{k^1_{\mathrm{Sob}}} c_k\right](x) = \sqrt{2}\,\alpha\left(x-\frac{1}{2}\right) + \frac{1}{\sqrt{2}}\left(\beta_2(x) - \beta_1^-(x) - \beta_1^+(x) + \frac{1}{6}\,\beta_0\right) = \frac{1}{(k\pi)^2}\,c_k(x)\ .$$

G.2. **Additional details for Section 5.2.2**

It only remains to prove that $\left[T_{k^2_{\text{Sob}}} B_1\right](\cdot) = g_A(\cdot) + g_B(\cdot)$ as stated in Eq. (5.4). This can be done by simply calculating the integral $[T_{k^2_{\text{Sob}}} B_1](x)$ for a given point $x \in [0\,,1]$. This is not particularly difficult but the calculations are very tedious. To save time, most integrals were computed with the online integral calculator proposed by Wolfram|Alpha[15]. First, one has:

$$\begin{aligned}
[T_{k^2_{\text{Sob}}} B_1](x) &= \int_0^1 k^2_{\text{Sob}}(x,\xi)\, B_1(\xi)\, \mathrm{d}\xi \\
&= \int_0^1 \left(B_1(x)\, B_1(\xi) + \frac{1}{4}\, B_2(x)\, B_2(\xi) - \frac{1}{24}\, B_4(|x-\xi|) \right) B_1(\xi)\, \mathrm{d}\xi \\
&= B_1(x) \int_0^1 B_1(\xi)\, B_1(\xi)\, \mathrm{d}\xi + \frac{1}{4} B_2(x) \int_0^1 B_2(\xi)\, B_1(\xi)\, \mathrm{d}\xi - \frac{1}{24} \int_0^1 B_4(|x-\xi|)\, B_1(\xi)\, \mathrm{d}\xi \\
&= \alpha\, B_1(x) + \frac{1}{4}\, \beta\, B_2(x) - \frac{1}{24}\, \gamma(x) \quad \text{after denoting by } \alpha \text{ and } \beta \text{ and } \gamma(x) \text{ the three integrals.}
\end{aligned} \tag{G.4}$$

It can be proved that $\boxed{\alpha = \frac{1}{12}}$ and $\boxed{\beta = 0}$.

Then, the integral $\gamma(x)$ may be divided into four terms:

$$\begin{aligned}
\gamma(x) &= \int_0^1 B_4(|x-\xi|)\, B_1(\xi)\, \mathrm{d}\xi = \int_0^1 \left[(x-\xi)^4 - 2\,|x-\xi|^3 + (x-\xi)^2 - \frac{1}{30} \right] B_1(\xi)\, \mathrm{d}\xi \\
&= \int_0^1 (x-\xi)^4\, B_1(\xi)\, \mathrm{d}\xi - 2 \int_0^1 |x-\xi|^3\, B_1(\xi)\, \mathrm{d}\xi + \int_0^1 (x-\xi)^2\, B_1(\xi)\, \mathrm{d}\xi + \frac{1}{30} \int_0^1 B_1(\xi)\, \mathrm{d}\xi \\
&= \gamma_4(x) - 2\,\gamma_3(x) + \gamma_2(x) + \frac{1}{30}\,\gamma_0 \quad \text{after denoting by } \gamma_0,\ \gamma_2(x),\ \gamma_3(x) \text{ and } \gamma_4(x) \text{ the four integrals.}
\end{aligned} \tag{G.5}$$

It can be proved that $\boxed{\gamma_4(x) = \frac{1}{30}\,(-10\,x^3 + 15\,x^2 - 9\,x + 2)}$, $\boxed{\gamma_2(x) = \frac{1}{12}\,(1-2x)}$ and $\boxed{\gamma_0 = 0}$.

To eliminate the absolute value in the integral expression of $\gamma_3(x)$, one may write:

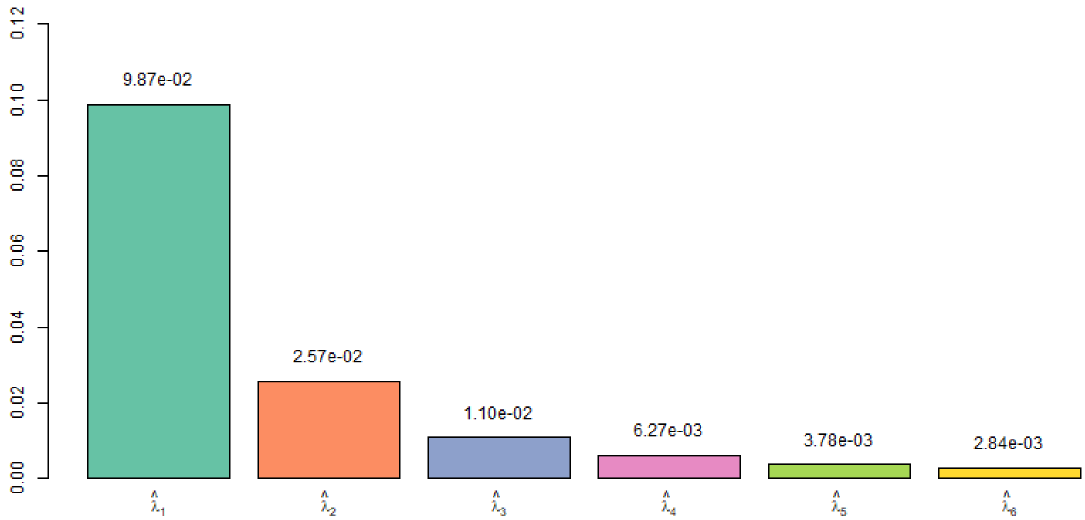
$$\begin{aligned}
\gamma_3(x) &= \int_0^1 |x-\xi|^3\, B_1(\xi)\, \mathrm{d}\xi \\
&= \int_0^x (x-\xi)^3\, B_1(\xi)\, \mathrm{d}\xi + \int_x^1 (\xi-x)^3\, B_1(\xi)\, \mathrm{d}\xi \\
&= \gamma_3^-(x) + \gamma_3^+(x) \quad \text{after denoting by } \gamma_3^-(x) \text{ and } \gamma_3^+(x) \text{ the two integrals.}
\end{aligned} \tag{G.6}$$

It can be proved that $\boxed{\gamma_3^-(x) = \frac{1}{40}\left[2\,x^5 - 5\,x^4\right]}$ and $\boxed{\gamma_3^+(x) = \frac{1}{40}\left[2\,x^5 - 5\,x^4 + 10\,x^2 - 10\,x + 3\right]}$.

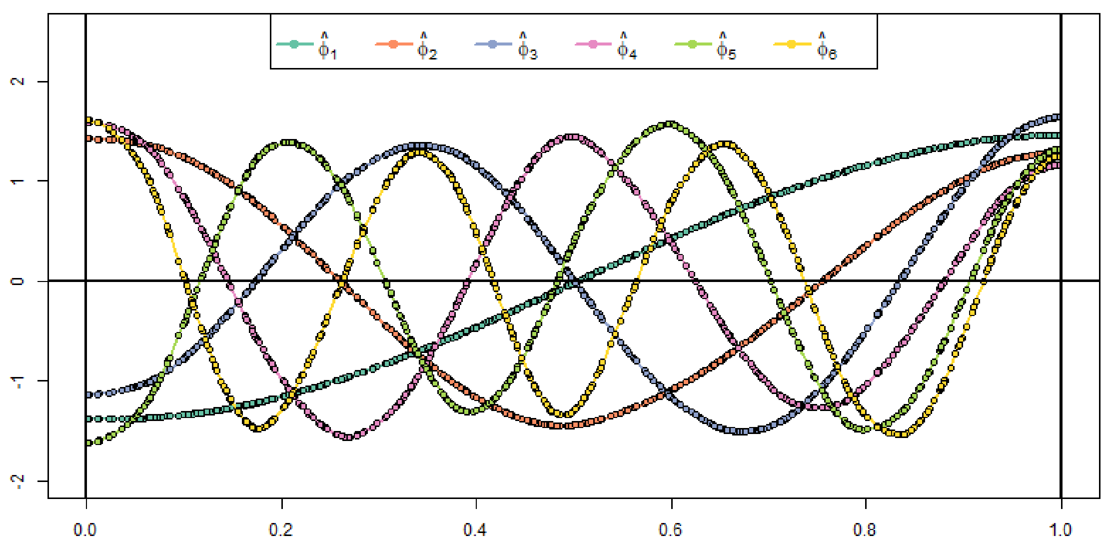
[15] https://www.wolframalpha.com

With the analytical expressions of α, β, γ_0, $\gamma_2(x)$, $\gamma_3^-(x)$, $\gamma_3^+(x)$ and $\gamma_4(x)$, Eq. (G.4), (G.5) and (G.6) yield:

$$\begin{aligned}\left[T_{k^2_{\mathrm{Sob}}} B_1\right](x) &= \alpha\, B_1(x) + \frac{1}{4}\,\beta\, B_2(x) - \frac{1}{24}\left[\gamma_4(x) - 2\left(\gamma_3^-(x) + \gamma_3^+(x)\right) + \gamma_2(x) + \frac{1}{30}\,\gamma_0\right] \\ &= g_A(x) + g_B(x)\ .\end{aligned}$$

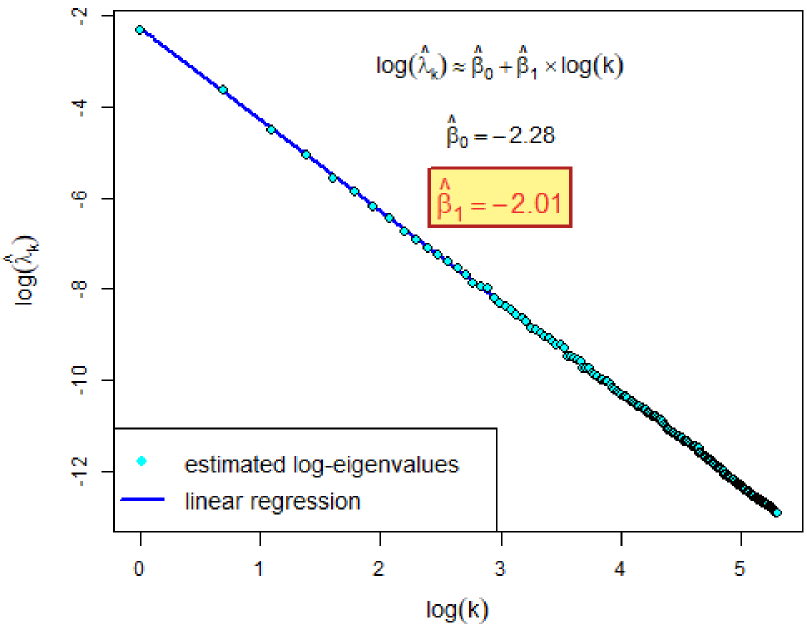


(A) Histogram of the estimated eigenvalues.

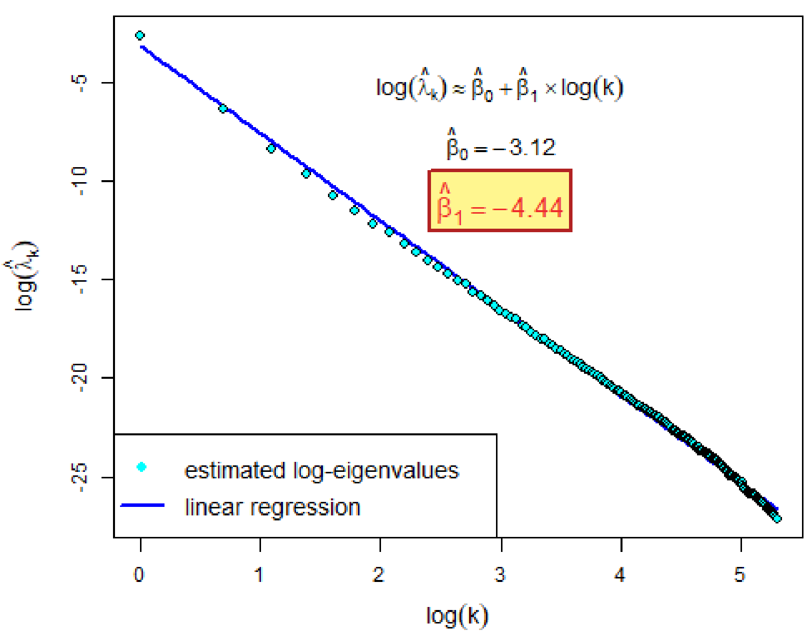


(B) Curves of the estimated eigenfunctions.

FIGURE 1. Estimation of the eigenvalues and eigenfunctions involved in the Mercer decomposition of the kernel k^1_{Sob}. The KFA method is performed with $n = 500$ sample points.

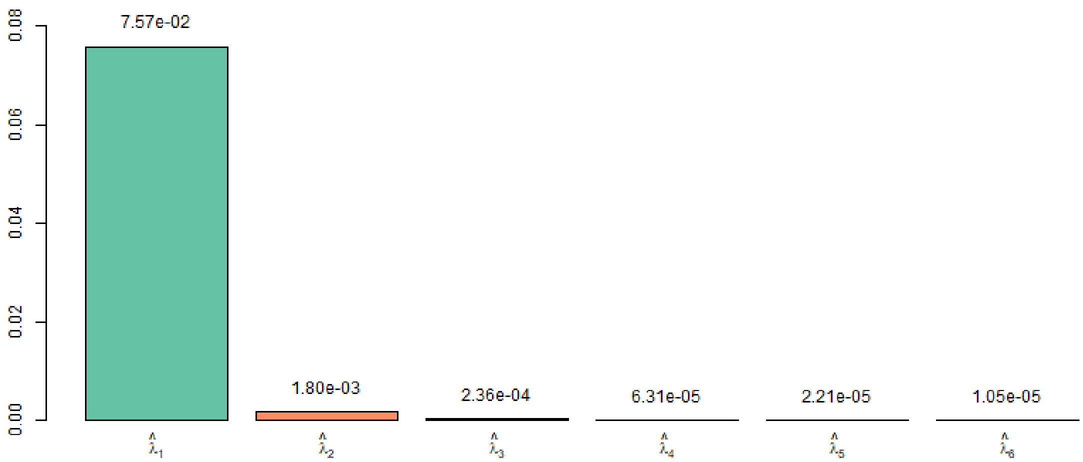


(A) Estimation of the decay rate for the kernel k^1_{Sob} after a logarithmic transformation.

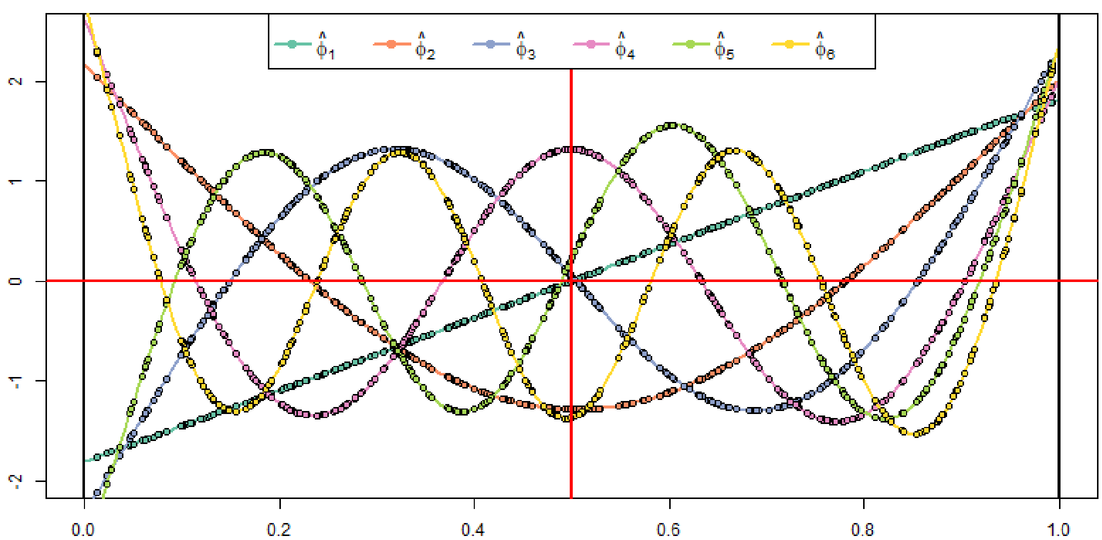


(B) Estimation of the decay rate for the kernel k^2_{Sob} after a logarithmic transformation.

FIGURE 2. Regression-based eigendecay analysis for the kernel k^r_{Sob} (with $r \in \{1, 2\}$). The eigenvalues are first estimated by KFA (with $n = 500$ points).

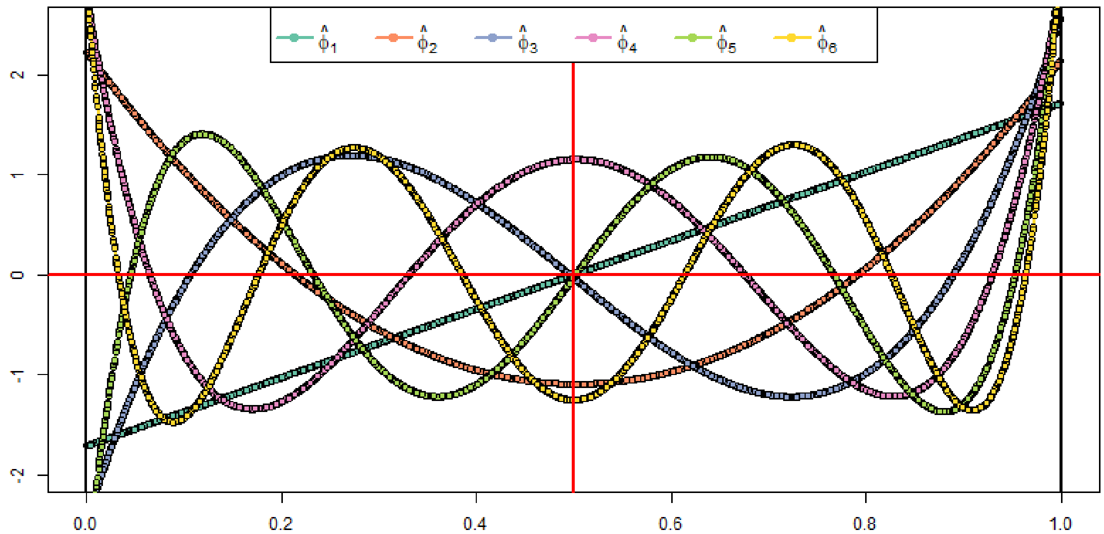


(A) Histogram of the estimated eigenvalues.

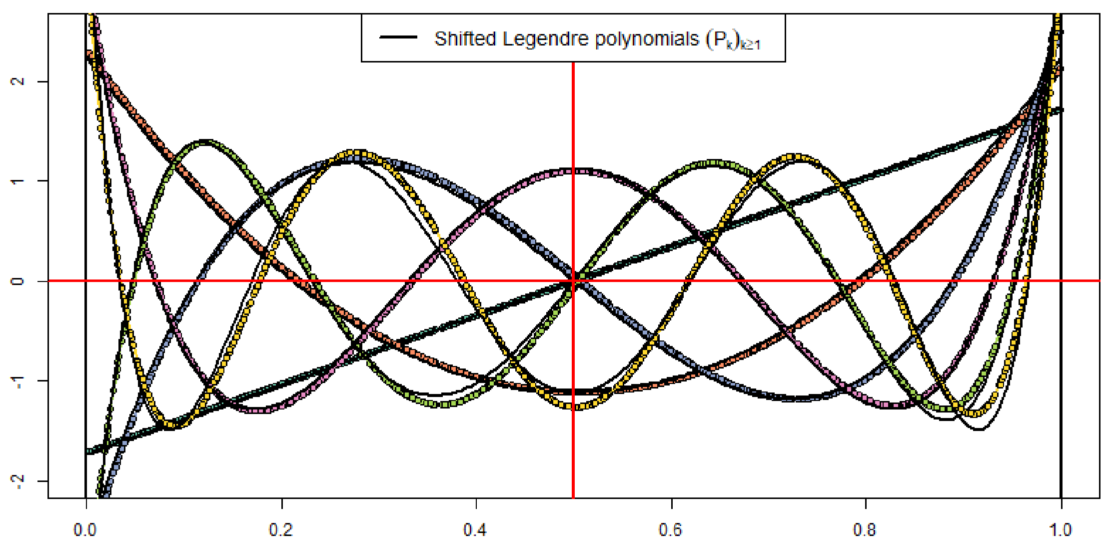


(B) Curves of the estimated eigenfunctions.

FIGURE 3. Estimation of the eigenvalues and eigenfunctions involved in the Mercer decomposition of the kernel k^2_{Sob}. The KFA method is performed with $n = 500$ sample points.



(A) Curves of the estimated eigenfunctions.



(B) Approximation of the theoretical eigenfunctions.

FIGURE 4. Brute-force estimation of the eigenfunctions involved in the Mercer decomposition of the kernel k^5_{Sob}. The KFA method is performed with $n = 3000$ sample points.