Reproducing kernels for spaces of zero mean functions. Application to sensitivity analysis

N. Durrande[a,*], D. Ginsbourger[b], O. Roustant[a], L. Carraro[c]

[a] *CROCUS - Ecole Nationale Supérieure des Mines de St Etienne, 158 cours Fauriel - 42023 St Etienne cedex 2, France*
[b] *Institute of Mathematical Statistics and Actuarial Science, University of Berne, Alpeneggstrasse 22 - 3012 Bern, Switzerland*
[c] *TELECOM St Etienne, 25 rue du Dr Rémy Annino - 42000 Saint Etienne, France*

Abstract

Given a Reproducing Kernel Hilbert Space $(\mathcal{H}, \langle ., . \rangle)$ of real-valued functions and a suitable measure μ over the source space, we decompose $\mathcal{H}$ as sum of a subspace of centered functions for μ and its orthogonal in $\mathcal{H}$. This decomposition leads to a special case of ANOVA kernels, for which the functional ANOVA representation of the minimal norm interpolator can be elegantly derived. The proposed kernels appear to be particularly convenient for analyzing the effect of each (group of) variable(s) and computing sensitivity indices without recursivity.

Keywords: Kernel Methods, Global Sensitivity Analysis, Sobol-Hoeffding Decomposition, Gaussian Process Regression, Computer Experiments

1. Introduction

Let f be a real-valued function defined over $D \subset \mathbb{R}^d$. We assume that f is costly to evaluate and that we want to study some global properties of f such as the influence of each variable on f. As the number of evaluations of f is limited, it may be unaffordable to run sensitivity analysis methods directly on f. Thus, it can be helpful to replace f by a mathematical approximation for performing

*Corresponding author. Nicolas Durrande, email: durrande@emse.fr, phone: +33 (0)4 77 42 66 38.

such studies [7]. We propose in this article a class of functional approximations that is well suited for performing global sensitivity analysis. First of all, we present some background in sensitivity analysis, interpolation in RKHS, and a class of kernels from the state of the art called ANOVA kernels. We then construct RKHS of zero mean functions and derive a new class of ANOVA kernels that is well suited for sensitivity analysis. Finally, we illustrate the use of those kernels on a classical example from the sensitivity analysis literature.

1.1. Sensitivity analysis

Let us consider $f \in L^2(D, \mu)$, where $D = D_1 \times \cdots \times D_d$ is a product space of bounded sets $D_i \subset \mathbb{R}$ and $\mu = \mu_1 \times \cdots \times \mu_d$ is a product probability measure over D. The purpose of global sensitivity analysis is to analyze the influence of all (groups of) variables on f. A common approach is to study the variance of $f(\mathbf{X})$ where $\mathbf{X}$ is a random vector with distribution μ.

If $d = 1$, any $g \in L^2(D, \mu)$ can be canonically decomposed as a sum of a constant plus a zero mean function,

$$g = \int_D g(s)\mathrm{d}\mu(s) + \left(g - \int_D g(s)\mathrm{d}\mu(s)\right)$$

so that we have a geometric decomposition of $L^2(D, \mu)$:

$$L^2(D, \mu) = L_1^2(D, \mu) \overset{\perp}{\oplus} L_0^2(D, \mu) \tag{1}$$

where $L_1^2(D, \mu)$ denote the subspace of constant functions and $L_0^2(D, \mu)$ the subspace of zero mean functions for μ.

Similarly, if $d > 1$, the space $L^2(D, \mu)$ has a tensor product structure [6]

$$L^2(D, \mu) = \bigotimes_{i=1}^{d} L^2(D_i, \mu_i). \tag{2}$$

Using Eq. 1 and the notation $L_P^2(D, \mu) = \bigotimes_{i=1}^{d} L_{P_i}^2(D_i, \mu_i)$ for $P \in \{0, 1\}^d$ we

obtain

$$L^2(D,\mu) = \bigotimes_{i=1}^{d} \left(L_1^2(D_i,\mu_i) \overset{\perp}{\oplus} L_0^2(D_i,\mu_i) \right) = \bigoplus_{P\in\{0,1\}^d}^{\perp} L_P^2(D,\mu). \tag{3}$$

A key property is that the subspaces L_P^2 and L_Q^2 are orthogonal whenever $P \neq Q$. Given an arbitrary function $f \in L^2(D,\mu)$, the orthogonal projections of f onto those subspaces leads to the functional ANOVA representation [4, 10] (or *Sobol-Hoeffding decomposition*) of f into main effects and interactions:

$$f(\mathbf{x}) = f_0 + \sum_{i=1}^{d} f_i(x_i) + \sum_{i<j} f_{i,j}(x_i,x_j) + \cdots + f_{1,\ldots,d}(\mathbf{x}). \tag{4}$$

Let us remark that f_0 can be seen as a constant function over D (i.e. an element of $L^2_{\{0\}^n}$), and each $f_{i_1,\ldots,i_k}$ $(1 \leq k \leq d,\, i_1,\ldots,i_k \in [1,d])$ can be represented as an element of $L^2_{P(I)}(D,\mu)$ (I denotes here $\{i_1,\ldots,i_k\}$) by identifying it with

$$f_{P(I)} : \mathbf{x} \in D \longrightarrow f_{i_1,\ldots,i_k}(x_{i_1},\ldots,x_{i_k}) \in \mathbb{R} \tag{5}$$

where $P(I) \in \{0,1\}^d$ with $P(I)_i = 1$ if $i \in I$ and $P(I)_i = 0$ if $i \notin I$. So the integral of $f_{P(I)}$ with respect to any of the variables indexed by $i \in I$ is zero. This representation of f gives an insight on the influence of each variable or couple of variables on f. For the constant term, the main effects, and the two-factor interactions, one gets the classical expressions

$$\begin{aligned} f_0 &= \int_D f(\mathbf{x})\mathrm{d}\mu(\mathbf{x}) \\ f_i(x_i) &= \int_{D_{-\{i\}}} f(\mathbf{x})\mathrm{d}\mu_{-i}(\mathbf{x}_{-i}) - f_0 \\ f_{i,j}(x_i,x_j) &= \int_{D_{-\{i,j\}}} f(\mathbf{x})\mathrm{d}\mu_{-\{i,j\}}(\mathbf{x}_{-\{i,j\}}) - f_i(x_i) - f_j(x_j) - f_0 \end{aligned} \tag{6}$$

where $D_{-I} := \prod_{i\notin I} D_i$ and $\mu_{-I} := \otimes_{i\notin I}\mu_i$. Similarly, the calculation of any f_I requires to have recursively calculated all the f_J's for $J \in I$, which makes it cumbersome (if not practically impossible) to get higher order interactions.

Coming back to the case of a random vector $\mathbf{X}$ with distribution μ, the variance of the random variable $f(\mathbf{X})$ can be decomposed as

$$\mathrm{var}(f(\mathbf{X})) = \sum_{i=1}^{d} \mathrm{var}(f_i(X_i)) + \sum_{i<j} \mathrm{var}(f_{i,j}(\mathbf{X}_{i,j})) + \cdots + \mathrm{var}(f_{1,\ldots,d}(\mathbf{X})) \quad (7)$$

and the global sensitivity index S_I for a subset of indices I is usually defined as

$$S_I = \mathrm{var}(f_I(\mathbf{X}_I))/\mathrm{var}(f(\mathbf{X})). \quad (8)$$

S_I represents the proportion of variance of $f(\mathbf{X})$ explained by the interaction between the variables contained in I. The knowledge of the indices S_I is very helpful for understanding the influence of the inputs, but the computation of the f_I's is cumbersome when the evaluation of f is costly since they rely on the computation of the integrals of Eq 6. Following [7], it can then be advantageous to perform the sensitivity analysis on a surrogate model m approximating f.

1.2. Optimal interpolation in RKHS

The class of functional approximation techniques considered in this work, commonly referred to as Kriging or Gaussian Process Regression in contemporary statistical learning settings, boils down to optimal interpolation in Reproducing Kernel Hilbert Spaces (RKHS). f is here assumed to be known at a set of points $\mathcal{X} = \{\mathcal{X}_1, \ldots, \mathcal{X}_n\} \in D$. Given $\mathcal{H}$ a RKHS of real-valued functions over D with kernel $K(.,.)$, the interpolator m of f at $\mathcal{X}$ that minimizes $||m||_{\mathcal{H}}$ is [8]:

$$m(\mathbf{x}) = \mathbf{k}(\mathbf{x})^t \mathbf{K}^{-1} \mathbf{F} \quad (9)$$

where $\mathbf{F} = f(\mathcal{X})$ is the column vector of observations, $\mathbf{k}(.)$ is the column vector of functions $(K(\mathcal{X}_i, .))_{1 \leq i \leq n}$ and $\mathbf{K}$ is the Gram matrix $(\mathbf{K})_{i,j} = K(\mathcal{X}_i, \mathcal{X}_j)$.

A striking fact about Eq 9 is that m can be an interpolator even if $f \notin \mathcal{H}$. K can be any symmetric positive definite kernel and it has to be chosen in practice. This choice has a great impact on the resulting model, and it is customary to

select K among family of parametric functions of positive type according to some prior knowledge about f, and to estimate the corresponding parameters based on observed data. We will focus here on a particular family of such kernels, called ANOVA, which are furthermore designed to offer good interpretability properties. The main contribution of this paper (in Section 3) deals with a special case of ANOVA kernels tailored for an improved disentanglement of multivariate effects.

1.3. ANOVA kernels and a candidate ANOVA-like decomposition of m

ANOVA kernels (See e.g. [2] section 5.4 for a historic approach) have been proposed in the literature of multivariate regression for an enhanced interpretability of splines and related models. They are constructed [5] by taking tensor products of univariate kernels $1 + k^i$, where 1 stands for a *bias term* and the k^i's are arbitrary symmetric definite positive kernels on $D_i \times D_i$ $(1 \leq i \leq d)$:

$$K_{ANOVA}(\mathbf{x}, \mathbf{y}) = \prod_{i=1}^{d}(1 + k^i(x_i, y_i)) = 1 + \sum_{I \subset \{1,\dots,d\}} \prod_{i \in I} k^i(x_i, y_i). \tag{10}$$

Denoting by $\mathbb{1}^i$ and $\mathcal{H}^i$ the RKHS of functions defined over D_i with respective reproducing kernels 1 and k^i, K_{ANOVA} is in fact the reproducing kernel of the space $\mathcal{H}_{ANOVA} = \bigotimes_{i=1}^{d}(\mathbb{1}^i + \mathcal{H}^i)$. Now, back to Eq. 10, the particular structure of K_{ANOVA} allows us to develop the $n \times 1$ vector $\mathbf{k}(\mathbf{x})$ of eq 9 as follows:

$$\mathbf{k}(.) = \mathbf{1} + \sum_{I \subset \{1,\dots,d\}} \bigodot_{i \in I} \mathbf{k}^i(.) \tag{11}$$

where $\odot$ denotes a term-wise product. Injecting this relation in eq 9, we get:

$$\begin{aligned} m(\mathbf{x}) &= \mathbf{1}^t\mathbf{K}^{-1}\mathbf{F} + \sum_{I \subset \{1,\dots,d\}} \left(\bigodot_{i \in I} \mathbf{k}^i(x_i) \right)^t \mathbf{K}^{-1}\mathbf{F} \\ &= \mathbf{1}^t\mathbf{K}^{-1}\mathbf{F} + \sum_{I \subset \{1,\dots,d\}} \prod_{i \in I} \left(\mathbf{k}^i(x_i)^t\mathbf{K}^{-1}\mathbf{F} \right) \end{aligned} \tag{12}$$

Noting $m_0 = \mathbf{1}^t\mathbf{K}^{-1}\mathbf{F}$ and $m_I(\mathbf{x}) = \prod_{i \in I} \mathbf{k}^i(x_i)^t\mathbf{K}^{-1}\mathbf{F}$ ($I \subset \{1, \dots, d\}$ and

$I \neq \varnothing$), we obtain a development of m which looks quite similar to its FANOVA representation:

$$m(\mathbf{x}) = m_0 + \sum_{i=1}^{d} m_i(x_i) + \sum_{i<j} m_{i,j}(\mathbf{x}_{i,j}) + \cdots + m_{1,\ldots,d}(\mathbf{x}_{1,\ldots,d}), \qquad (13)$$

where the m_I's have the nice feature of not requiring any recursive computation of integrals. However, the properties of the ANOVA representation are not respected since the m_I's are not necessarily zero mean functions, i.e. any two terms of the decomposition do not have to be orthogonal in L_2. For example, if k^i is an Ornstein-Uhlenbeck kernel [1], it is known that $\mathbb{1}^i \subset \mathcal{H}^i$.

Let us remark that the submodels m_0 and m_I respectively belongs to the spaces $\mathbb{1}^1 \otimes \cdots \otimes \mathbb{1}^d$ and $\bigotimes_{i\in I} \mathcal{H}^i \bigotimes_{i\notin I} \mathbb{1}^i$, but that they are not necessarily orthogonal projection onto those spaces. In order to ensure that the decomposition of Eq. 12 has the properties required in Eq. 4, we have to consider RKHS $\mathcal{H}^i$ that are L_2-orthogonal to the constant functions $\mathbb{1}^i$, ie RKHS of zero mean functions for μ_i [11]. With such a construction, we would benefit from the advantages of the two decompositions: the meaning of the decomposition given by Eq. 4 for the analysis of variance and the easiness of computation of the m_I's from Eq. 12.

2. RKHS of zero mean functions

We will show in this section how to extract a RKHS of zero mean functions from a RKHS with arbitrary symmetric definite positive kernel K (k if $d = 1$).

2.1. Decomposition of one-dimensional RKHS

Let $\mathcal{H}$ be a RKHS of functions defined over a compact set $D \subset \mathbb{R}$ and μ a finite Borel measure over D. Furthermore, we consider the couple of hypotheses:

H 1.
(i) $k: D \times D \to \mathbb{R}$ *is* $\mu \otimes \mu$*-measurable.*
(ii) $\int_D \sqrt{k(s,s)}\mathrm{d}\mu(s) < \infty$.

As D is compact, any bounded kernel satisfies the condition (ii) so this hypothesis is not very restrictive. For example, usual stationary kernels such as the Gaussian, power-exponential and Matérn kernels satisfy it.

Proposition 1. *Under H1,* $\mathcal{H}$ *can be decomposed as a sum of two orthogonal sub-RKHS,* $\mathcal{H} = \mathcal{H}_0 \overset{\perp}{\oplus} \mathcal{H}_1$ *where* $\mathcal{H}_0$ *is a RKHS of zero-mean functions for* μ*, and its orthogonal* $\mathcal{H}_1$ *is at most 1-dimensional.*

Proof. Following H1 (i), the integral operator $I : \mathcal{H} \to \mathbb{R},\ h \mapsto \int_D h(s)\mathrm{d}\mu(s)$ is bounded since for $h \in \mathcal{H}$

$$|I(h)| \leq \int_D |\langle h, k(s,.)\rangle_{\mathcal{H}}|\mathrm{d}\mu(s) \leq ||h||_{\mathcal{H}} \int_D \sqrt{k(s,s)}\mathrm{d}\mu(s) \tag{14}$$

According to the Riesz representation theorem, there exists a unique $R \in \mathcal{H}$ such that $\forall h \in \mathcal{H},\ I(h) = \langle h, R\rangle_{\mathcal{H}}$. If $R(.) = 0$, then all $f \in \mathcal{H}$ are centered functions for μ, so that $\mathcal{H}_0 = \mathcal{H}$ and $\mathcal{H}_1 = \{0\}$. If $R(.) \neq 0$, then $\mathcal{H}_1 = span(R)$ is a 1-dimensional sub-RKHS of $\mathcal{H}$, and the subspace $\mathcal{H}_0$ of centered functions for μ can be defined by $\mathcal{H}_0 = \mathcal{H}_1^{\perp}$. □

Remark 1. *For all* $x \in D$ *the value of* $R(x)$ *can be calculated explicitly. Indeed, recalling that* $k(x,.)$ *and* R *are respectively the representers in* $\mathcal{H}$ *of the evaluation functional at* x *and of the integral operator, we get:*

$$R(x) = \langle k(x,.), R\rangle_{\mathcal{H}} = I(k(x,.)) = \int_D k(x,s)\mathrm{d}\mu(s). \tag{15}$$

The reproducing kernels k_0, k_1 of $\mathcal{H}_0$ and $\mathcal{H}_1$ satisfy $k = k_0 + k_1$. Let π denote the orthogonal projection onto $\mathcal{H}_1$. Following [2] we obtain

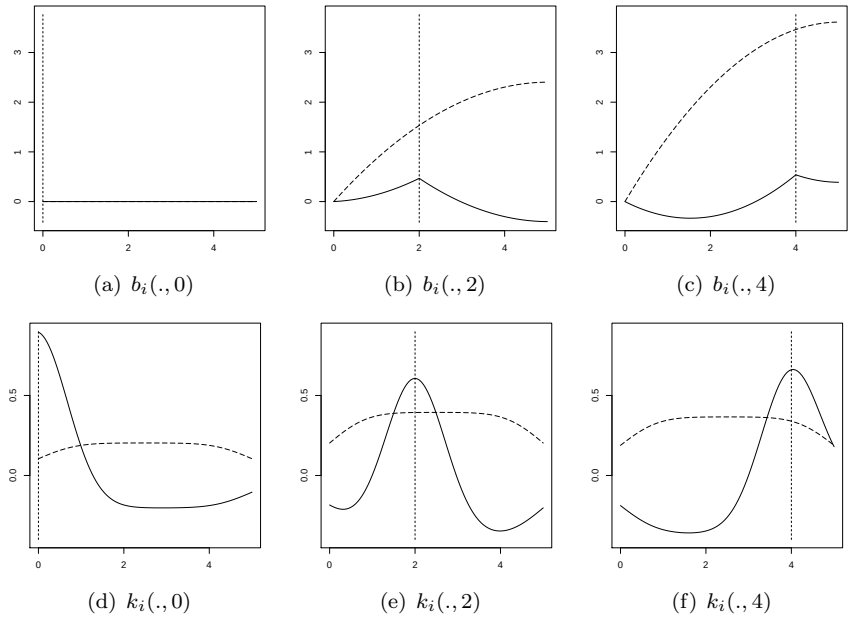
$$\begin{aligned} k_0(x,y) &= k(x,y) - \pi(k(x,.))(y) \\ &= k(x,y) - \frac{\int_D k(x,s)\mathrm{d}\mu(s) \int_D k(y,s)\mathrm{d}\mu(s)}{\iint_{D\times D} k(s,t)\mathrm{d}\mu(s)\mathrm{d}\mu(t)} \end{aligned} \tag{16}$$

2.2. Example

Let us briefly illustrate the previous results for two usual kernels:

$$b(x,y) = \min(x,y) \qquad \text{and} \qquad g(x,y) = \exp\left(-(x-y)^2\right), \tag{17}$$

known as the *Brownian* and the *Gaussian* covariance kernels, respectively. Here $D = [0, 5]$ and μ is the Lebesgue measure over D. Figure 1 represents sections of the reproducing kernels $k_i(x, y)$ and $g_i(x, y)$ $(i \in \{0, 1\})$ outcomes of the decomposition of b and g, for various values of y.



(a) $b_i(., 0)$ (b) $b_i(., 2)$ (c) $b_i(., 4)$

(d) $k_i(., 0)$ (e) $k_i(., 2)$ (f) $k_i(., 4)$

Figure 1: Representation of the sub kernels $b_i(., y)$ and $g_i(., y)$ for $y = 0, 2, 4$ and i=0,1. The dashed lines correspond to b_1, g_1 and the solid lines are for b_0 and g_0.

We observe on this figure that $b_0(., y)$ and $g_0(., y)$ take negative values and that they are zero mean functions (as elements of $\mathcal{H}_0$). Moreover, $b_0(., y)$ and $b_1(., y)$ (respectively $k_0(., y)$, $k_1(., y)$) are orthogonal for the scalar product of their RKHS but are not orthogonal for $L^2(D, \mu)$.

2.3. Generalization for multi-dimensional RKHS

The former decomposition of one-dimensional kernels leads directly to the decomposition of tensor product kernels

$$K(\mathbf{x}, \mathbf{y}) = \prod_{i=1}^{d} k^i(x_i, y_i) = \prod_{i=1}^{d} (k_0^i(x_i, y_i) + k_1^i(x_i, y_i)). \tag{18}$$

if the k^i's satisfy H1. Since $k_0^i(x_i, y_i)$ is a 1-dimensional kernel, it can be seen as a bias term so this equation highlights the similarity between the usual tensor product kernels (power-exponential, Brownian, Matérn, ...) and ANOVA kernels.

3. A new class of kernels for sensitivity analysis

We now propose a special class of ANOVA kernels,

$$K^\star_{ANOVA}(\mathbf{x}, \mathbf{y}) = \prod_{i=1}^{d}(1 + k_0^i(x_i, y_i)) = 1 + \sum_{I \subset \{1,\dots,d\}} \prod_{i \in I} k_0^i(x_i, y_i), \tag{19}$$

where the k_0^i are obtained by decomposing kernels as in the previous section.

Proposition 2. *If m is a best predictor based on $K^\star_{ANOVA}$,*

$$m_I = \prod_{i \in I} \mathbf{k}_0^i(x_i)^t \mathbf{K}^{-1} \mathbf{F} \tag{20}$$

is the term of the functional ANOVA representation of m indexed by I. Hence, the decomposition of m given by Eq. 13 coincides with its functional ANOVA representation (Eq. 4).

Proof. The kernels k_0^i are associated to RKHS $\mathcal{H}_0^i$ of zero-mean functions, so we have $\mathbb{1}^i \perp_{L_2} \mathcal{H}_0^i$. The underlying RKHS associated to $K^\star_{ANOVA}$ is

$$\mathcal{H}^\star_{ANOVA} = \prod_{i=1}^{d} \left(\mathbb{1}^i \overset{\perp}{\oplus} \mathcal{H}_0^i \right) \tag{21}$$

where $\perp$ stands for the L^2 scalar product. The result follows. □

Corollary 1. *Contrarily to usual ANOVA kernels, the class of $K^\star_{ANOVA}$ ensures that the terms m_I are mutually orthogonal in the L^2 sense.*

As the expression of the submodels is simple, the computation of the sensitivity indices can be performed analytically and efficiently.

Corollary 2. *The sensitivity indices S_I of m are given by:*

$$S_I = \frac{\mathrm{var}(m_I(\mathbf{X}_I))}{\mathrm{var}(m(\mathbf{X}))} = \frac{\mathbf{F}^t \mathbf{K}^{-1} \left(\bigodot_{i \in I} \Gamma_i \right) \mathbf{K}^{-1} \mathbf{F}}{\mathbf{F}^t \mathbf{K}^{-1} \left(\bigodot_{i=1}^{d} (1_{n \times n} + \Gamma_i) - 1_{n \times n} \right) \mathbf{K}^{-1} \mathbf{F}} \tag{22}$$

where Γ_i is the $n \times n$ matrix $\Gamma_i = \int_{D_i} \mathbf{k}_0^i(x_i) \mathbf{k}_0^i(x_i)^t \mathrm{d}\mu_i(x_i)$ and $1_{n \times n}$ is the $n \times n$ matrix of 1.

Proof. The numerator is obtained by direct calculation:

$$\begin{aligned}\mathrm{var}(m_I(\mathbf{X}_I)) &= \mathrm{var}\left(\prod_{i\in I}\mathbf{k}_0^i(x_i)^t\mathbf{K}^{-1}\mathbf{F}\right)\\ &= \mathbf{F}^t\mathbf{K}^{-1}\bigodot_{i\in I}\left(\int_{D_i}\mathbf{k}_0^i(x_i)\mathbf{k}_0^i(x_i)^t\mathrm{d}\mu_i(x_i)\right)\mathbf{K}^{-1}\mathbf{F}.\end{aligned} \tag{23}$$

For the denominator, we obtain similarly

$$\begin{aligned}\mathrm{var}(m(\mathbf{X})) = \mathbf{F}^t\mathbf{K}^{-1}\bigodot_{i\in I}\left(\int_{D_i}\left(1_{n\times 1}+\mathbf{k}_0^i(x_i)\right)\left(1_{n\times 1}+\mathbf{k}_0^i(x_i)\right)^t\mathrm{d}\mu_i(x_i)\right)\mathbf{K}^{-1}\mathbf{F}\\ -\mathbf{F}^t\mathbf{K}^{-1}1_{n\times n}\mathbf{K}^{-1}\mathbf{F}.\end{aligned} \tag{24}$$

We then use the property that $k_0^i(x,.)$ is a zero mean function so we have $\int_{D_i}\left(1_{n\times 1}+\mathbf{k}_0^i(x_i)\right)\left(1_{n\times 1}+\mathbf{k}_0^i(x_i)\right)^t\mathrm{d}\mu_i(x_i) = 1_{n\times n}+\Gamma_i$. □

Conversely to the method developed in [3], the computation of S_I does not require here to compute all S_J for $J\subset I$.

3.1. example: the g-function of Sobol

In order to illustrate the use of the kernels $K^\star_{ANOVA}$ we consider the so-called *g-function of Sobol*, defined over $[0,1]^d$ by

$$g(x_1,\ldots,x_d) = \prod_{k=1}^{d}\frac{|4x_k-2|+a_k}{1+a_k} \qquad \text{with } a_k>0. \tag{25}$$

This function is well known in the literature [9] and one particular advantage is that the Sobol sensitivity indices associated to the variables x_i, $i=1,\ldots,d$ can be obtained analytically:

$$S_i = \frac{\frac{1}{3(1+a_i)^2}}{\prod_{k=1}^{d}\left(1+\frac{1}{3(1+a_k)^2}\right)-1} \tag{26}$$

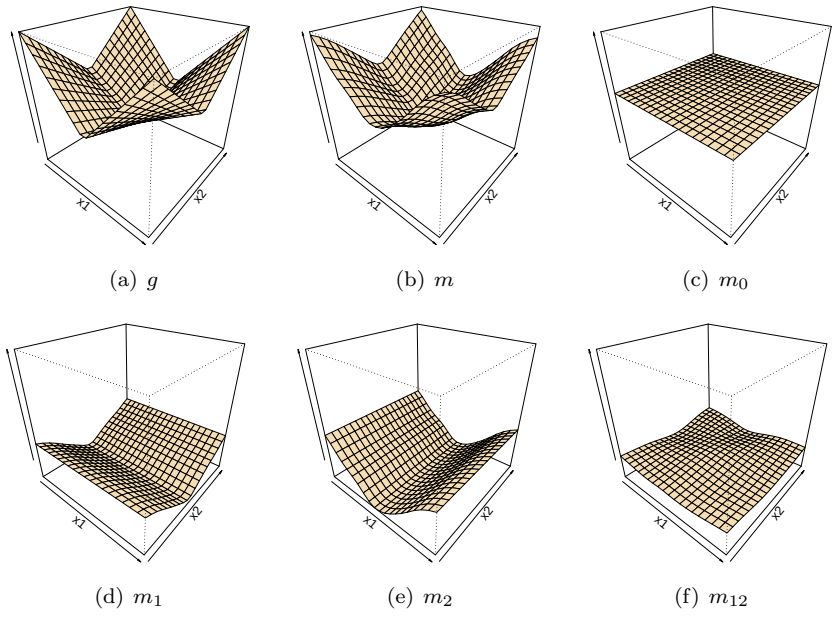
Here we limit ourself to the case $d=2$ and we choose $a_1=1,\ a_2=2$. Starting from a one-dimensional Matérn 3/2 kernel

$$k(x,y) = (1+2\,|x-y|)\exp(-2\,|x-y|), \tag{27}$$

we can derive the expression of $K^{\star}_{ANOVA}$ using Eq. 16 and 19:

$$K^{\star}_{ANOVA}(\mathbf{x},\mathbf{y}) = \prod_{i=1}^{2}\left(1 + k(x_i,y_i) - \frac{\int_0^1 k(x_i,s)\mathrm{d}\mu(s)\int_0^1 k(y_i,s)\mathrm{d}\mu(s)}{\iint_0^1 k(s,t)\mathrm{d}\mu(s)\mathrm{d}\mu(t)}\right). \tag{28}$$

We then build the optimal interpolator $m \in \mathcal{H}^{\star}_{ANOVA}$ based on the observation of g at 20 points of $[0,1]^2$ (those points steem from a LHS-maximin procedure). According to what we have seen, the function m can be split as a sum of 4 submodels m_0, m_1, m_2 and m_{12} which are represented on Fig. 3.1.



(a) g (b) m (c) m_0

(d) m_1 (e) m_2 (f) m_{12}

Figure 2: Representation of the g-function, the model m and all the submodels on $[0,1]^2$. The z scale is the same on all graphs.

We observe numerically that the mean value of m_1, m_2 and m_{12} is lower than $1e-15$ (in absolute value), corroborating that these functions are zero-mean. More generally, after numerical computations of the scalar products between any two functions of the set $\{m_0, m_1, m_2, m_{12}\}$, we observe that $|\langle m_I, m_J\rangle_{L^2}| <$

$1e-15$ for $I \neq J$.

Using Eq. 22, the sensitivity indices calculated for m are $S_1 = 0.69$, $S_2 = 0.30$ and $S_{12} = 0.02$ (the sum is slighly different from 1 because of rounding up). Those figures can be compared to the analytical values given by Eq. 26 which are $S_1 = 0.675$, $S_2 = 0.30$ and $S_{12} = 0.025$. The accuracy of the computation of Global sensitivity indices can be judged satisfactory in this example.

4. Concluding remark

We proposed a new class of kernels for which the functional ANOVA decomposition of the mean predictor can be obtained analytically, without the usual recursive integral calculations for higher order interaction terms. This new class is a special case of usual ANOVA kernels, with particular univariate kernels so that an orthogonality to constants is respected. Up to a calculation or a tabulation of the integral of univariate kernels, the replacement of usual ANOVA kernels by the ones proposed here may be done at neglectable cost in applications, with substantial benefits for the model interpretability and global sensitivity analysis studies.

The issue of the estimation of the parameters of $K^{\star}_{ANOVA}$ has not been raised yet in this article. This is however an important point for the practical use of those kernels. The use of the likelihood theory has been considered, but many points such as the links between the optimal parameters for K and the optimal parameters for the associated $K^{\star}_{ANOVA}$ needs to be studied in detail. Finally, since the pattern of the proof of Prop. 1 can be applied to any bounded operator on $\mathcal{H}$, the perspectives for future research include a focus on other operators than the integral operator I, for example for building RKHS respecting orthogonality to a family of trend basis functions.

References

[1] A. Antoniadis. Analysis of variance on function spaces. Statistics, 15:59–71, 1984.

[2] Alain Berlinet and Christine Thomas-Agnan. Reproducing kernel Hilbert spaces in probability and statistics. Kluwer Academic Publishers, 2004.

[3] W. Chen, R. Jin, and A. Sudjianto. Analytical variance-based global sensitivity analysis in simulation-based design under uncertainty. Journal of mechanical design, 127, 2005.

[4] B. Efron and C. Stein. The jackknife estimate of variance. The Annals of Statistics, 9:586–596, 1981.

[5] S.R. Gunn and J.S. Kandola. Structural modelling with sparse kernels. Machine learning, 48:137–163, 2002.

[6] P. Krée. Produits tensoriels complétés d'espaces de Hilbert. Séminaire Paul Krée, Vol 1:No. 7, 1974–1975.

[7] A. Marrel, B. Iooss, B. Laurent, and O. Roustant. Calculations of sobol indices for the gaussian process metamodel. Reliability Engineering & System Safety, 94:742–751, 2009.

[8] C.E. Rasmussen and C. Williams. Gaussian Processes for Machine Learning. MIT Press, 2006.

[9] A Saltelli, K. Chan, and E. Scott. Sensitivity analysis. Wiley Series in Probability and Statistics, 2000.

[10] I.M. Sobol. Global sensitivity indices for nonlinear mathematical models and their monte carlo estimates. Mathematics and Computers in Simulation, 55:271–280, 2001.

[11] G. Wahba, Y. Wang, C. Gu, R. Klein, and B. Klein. Smoothing spline anova for exponential families, with application to the wisconsin epidemiological study of diabetic retinopathy. The Annals of Statistics, 23:1865–1895, 1995.