A class of ANOVA kernels dedicated to sensitivity analysis

N. Durrande, D. Ginsbourger, O. Roustant

SAMO 2013 — Nice
Outline:

1. ANOVA kernels
2. HDMR, FANOVA, Hoeffding-Sobol, ...
3. Kernels of zero-mean functions
4. Application to sensitivity analysis
5. Examples
Let $f$ be the function of interest. We assume $f \in L^2(\mu)$ with:

- $D = D_1 \times \cdots \times D_d$ with $D_i \subset \mathbb{R}$.
- $\mu$ is a separable probability measure: $\mu = \mu_1 \times \cdots \times \mu_d$

Given $n$ observations $f(x^{(i)}) = y^{(i)}$, we are interested in:

- The ANOVA representation of $f$
- Sobol sensitivity indices
A common approach is to approximate $f$ with a mathematical model and to compute the sensitivity indices of $m$.

We focus here on Gaussian process regression:

$$m(x) = k(x)^t K^{-1} Y$$
$$c(x, y) = k(x, y) - k(x)^t K^{-1} k(y)$$

The choice of the kernel has a great impact on the model... Is there a specific kernel that gives directly $m$’s FANOVA?
A first idea is to look at ANOVA kernels [Stitson 97]:

\[
K(x, y) = \prod_{i=1}^{d} (1 + k_i(x_i, y_i))
\]

With such kernels, the decomposition of \( m \) can be obtained directly:

**Example**

In 2D we have \( K = (1 + k_1) \times (1 + k_2) = 1 + k_1 + k_2 + k_1 k_2 \).

The best predictor can be written as

\[
m(x) = (1 + k_1(x_1) + k_2(x_2) + k_1(x_1)k_2(x_2))^{tK^{-1}} F
\]

\[
= 1^{tK^{-1}} F + k_1(x_1)^{tK^{-1}} F + k_2(x_2)^{tK^{-1}} F + k_1(x_1)k_2(x_2)^{tK^{-1}} F
\]

\[
= \underbrace{m_0^{tK^{-1}} F}_{m_0} + \underbrace{m_1(x_1)^{tK^{-1}} F}_{m_1(x_1)} + \underbrace{m_2(x_2)^{tK^{-1}} F}_{m_2(x_2)} + \underbrace{m_{12}(x)}_{m_{12}(x)}
\]
A first idea is to look at ANOVA kernels [Stitson 97]:

\[ K(x, y) = \prod_{i=1}^{d} (1 + k_i(x_i, y_i)) \]

With such kernels, the decomposition of \( m \) can be obtained directly:

**Example**

In 2D we have \( K = (1 + k_1) \times (1 + k_2) = 1 + k_1 + k_2 + k_1 k_2 \).

The best predictor can be written as

\[
m(x) = (1 + k_1(x_1) + k_2(x_2) + k_1(x_1)k_2(x_2))^t K^{-1} F
\]

\[
= 1^t K^{-1} F + k_1(x_1)^t K^{-1} F + k_2(x_2)^t K^{-1} F + k_1(x_1)k_2(x_2)^t K^{-1} F
\]

\[
= m_0 + m_1(x_1) + m_2(x_2) + m_{12}(x)
\]

However, the \( m_l \) do not satisfy \( \int_{D_i} m_l(x_i) dx_i = 0 \).
By construction $L^2(\mu) = \bigotimes_{i=1}^{d} L^2(\mu_i)$. Furthermore, $L^2(\mu_i)$ can be decomposed as

$$L^2(\mu_i) = 1_{D_i} + L^2_0(\mu_i)$$

$$f_i(x_i) = \int_{D_i} f_i(s) d\mu(s) + \left(f_i(x_i) - \int_{D_i} f_i(s) d\mu(s)\right).$$

We thus obtain:

$$L^2(\mu) = \bigotimes_{i=1}^{d} \left(1_{D_i} + L^2_0(\mu_i)\right) = 1_D + \sum_{i=1}^{d} L^2_0(\mu_i) + \cdots + \prod_{i=1}^{d} L^2_0(\mu_i)$$

FANOVA is given by the projections onto these subspaces.
The best predictor $m$ belongs to the RKHS $\mathcal{H}$ associated to $K$

**Definition (RKHS)**

A RKHS $\mathcal{H}$ with kernel $K$ is a Hilbert space of function such that

- $K(x, .) \in \mathcal{H}$ for all $x \in D$
- $\langle K(x, .), h \rangle_{\mathcal{H}} = h(x)$, for all $h \in \mathcal{H}$, for all $x \in D$.

Wahba suggests to consider the following structure for the RKHS:

$$\mathcal{H} = \bigotimes_{i=1}^{d} \left( 1_{D_i} + \mathcal{H}_i^0 \right)$$

How can we build a RKHS of zero mean function?
Let $\mathcal{H}$ be a RKHS of 1-dimensional functions with kernel $k$.

**Theorem**

If the kernel satisfies

$$\int_D \sqrt{k(s, s)} \, d\mu(s) < \infty$$

then $\mathcal{H}$ can be written

$$\mathcal{H} = \mathcal{H}_0 \perp \mathcal{H}_1$$

where $\mathcal{H}_0$ is a RKHS of zero-mean functions for $\mu$

$\mathcal{H}_1$ is at most 1-dimensional.

Let $R$ be the Riesz representer of $\int \cdot \, dx$:

$$\int_D h(s) \, d\mu(s) = \langle h, R \rangle_{\mathcal{H}}.$$

$\mathcal{H}_0$ corresponds to $R^\perp$. 

\[ \mathcal{H}_1 = \text{span}(R) \]
The expression of $R(x)$ can be obtained easily

$$R(x) = \langle R, k(x, .) \rangle_{\mathcal{H}} = \int_{D} k(x, s) ds$$
Finally, we have $\mathcal{H} = \mathcal{H}^1 \perp \mathcal{H}^0$ with

- $\mathcal{H}^1 = \text{span}(R)$ a one dimensional RKHS
- $\mathcal{H}^0$ a RKHS of zero mean functions

The kernels of those spaces are:

$$k^1(x, y) = \frac{\langle k(x, .), R \rangle_{\mathcal{H}} R(y)}{||R||^2} = \frac{\int k(x, s)ds \int k(y, s)ds}{\int \int k(s, t)dsdt}$$

$$k^0(x, y) = k(x, y) - \frac{\int k(x, s)ds \int k(y, s)ds}{\int \int k(s, t)dsdt}$$
As for the ANOVA representation in $L^2$, we can build a RKHS $\mathcal{H}$

$$\mathcal{H} = \bigotimes_{i=1}^{d} (\mathbb{1}_{D_i} + \mathcal{H}_i^0)$$

$$K(x, y) = \prod_{i=1}^{d} (1 + k_i^0(x_i, y_i))$$

with this space, the ANOVA representation is obtained naturally

$$m(x) = (1 + k_1^0(x_1) + k_2^0(x_2) + k_1^0(x_1)k_2^0(x_2))^t K^{-1} F$$

$$= m_0 + m_1(x_1) + m_2(x_2) + m_{12}(x)$$

here the $m_i$ satisfy $\int_{D_i} m_i(x_i) dx_i = 0$. 
The decomposition of the kernel gives directly a decomposition of the Gaussian process $Z(x) = Z_0 + Z_1(x_1) + Z_2(x_2) + Z_{12}(x)$:

where the $Z_i$:
- satisfy the FANOVA properties
- are independent
Using $K$, the sensitivity indices $S_I$ can be computed analytically:

$$S_I = \frac{\text{var}(m_I(X_I))}{\text{var}(m(X))} = \frac{\text{var}(k_I(X_I)^t K^{-1} Y)}{\text{var}(k(X)^t K^{-1} Y)}$$

$$= \frac{Y^T K^{-1} (\bigodot_{i \in I} \Gamma_i) K^{-1} Y}{Y^T K^{-1} \left( \bigodot_{i=1}^d (1_{n \times n} + \Gamma_i) - 1_{n \times n} \right) K^{-1} Y}$$

where $\Gamma_i$ is the matrix $\Gamma_i = \int_{D_i} k_i^0(s_i) k_i^0(s_i)^T ds_i$, $1_{n \times n}$ is the matrix of 1 and where $\bigodot$ is a term wise product.

Contrarily to other methods, the computation of $S_I$ do not require to compute all $S_J$ for $J \subset I$. 
We consider a test function defined on \([-5, 5]^2\)

\[ f(x) = x_1 + x_2^2 + x_1 x_2 \]

Using the kernels described above, we obtain:

\[
m(x_1, x_2) = m_0 + m_1(x_1) + m_2(x_2) + m_{12}(x_1, x_2)
\]
The computation of the sensitivity indices on $m$ gives

<table>
<thead>
<tr>
<th></th>
<th>$S_1$</th>
<th>$S_2$</th>
<th>$S_{12}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>model</td>
<td>0.23</td>
<td>0.48</td>
<td>0.29</td>
</tr>
<tr>
<td>analytical</td>
<td>0.25</td>
<td>0.5</td>
<td>0.25</td>
</tr>
</tbody>
</table>

On this example, the model gives a very good approximation.
Let us consider the random test function $f : [0, 1]^{10} \rightarrow \mathbb{R}$:

$$x \mapsto 10 \sin(\pi x_1 x_2) + 20(x_3 - 0.5)^2 + 10x_4 + 5x_5 + \mathcal{N}(0, 1)$$

The univariate sub-models are:
Conclusion:

- kernels can be adapted to the ANOVA representation
- The Sobol sensitivity indices can be computed efficiently

Future work:

- taking the prediction variance into account
- Estimation of the kernel parameters
- RKHS orthogonal to other operators than $\int$
References: