Kernel-based Sensitivity Analysis on Excursion Sets

Christophette Blanchet-Scalliet[1], Noé Fellmann[1,2], Céline Helbert[1], Delphine Sinoquet[2], Adrien Spagnol[2]

[1]Institut Camille Jordan, École Centrale de Lyon
[2]IFP Énergies Nouvelles

CIROQUO — Consortium Industrie Recherche pour l'Optimisation et la QUantification d'incertitude pour les données Onéreuses

Introduction

Goal: Quantifying the influence of uncertain inputs on feasible sets associated to a constrained robust optimization problem.

$$\boldsymbol{x}^* = \arg\min_{\boldsymbol{x}\in\mathcal{X}} \mathbb{E}_{\boldsymbol{U}}[f(\boldsymbol{x},\boldsymbol{U})] \text{ s.t. } \mathbb{P}_{\boldsymbol{U}}[g(\boldsymbol{x},\boldsymbol{U})\leq 0] \geq \alpha.$$

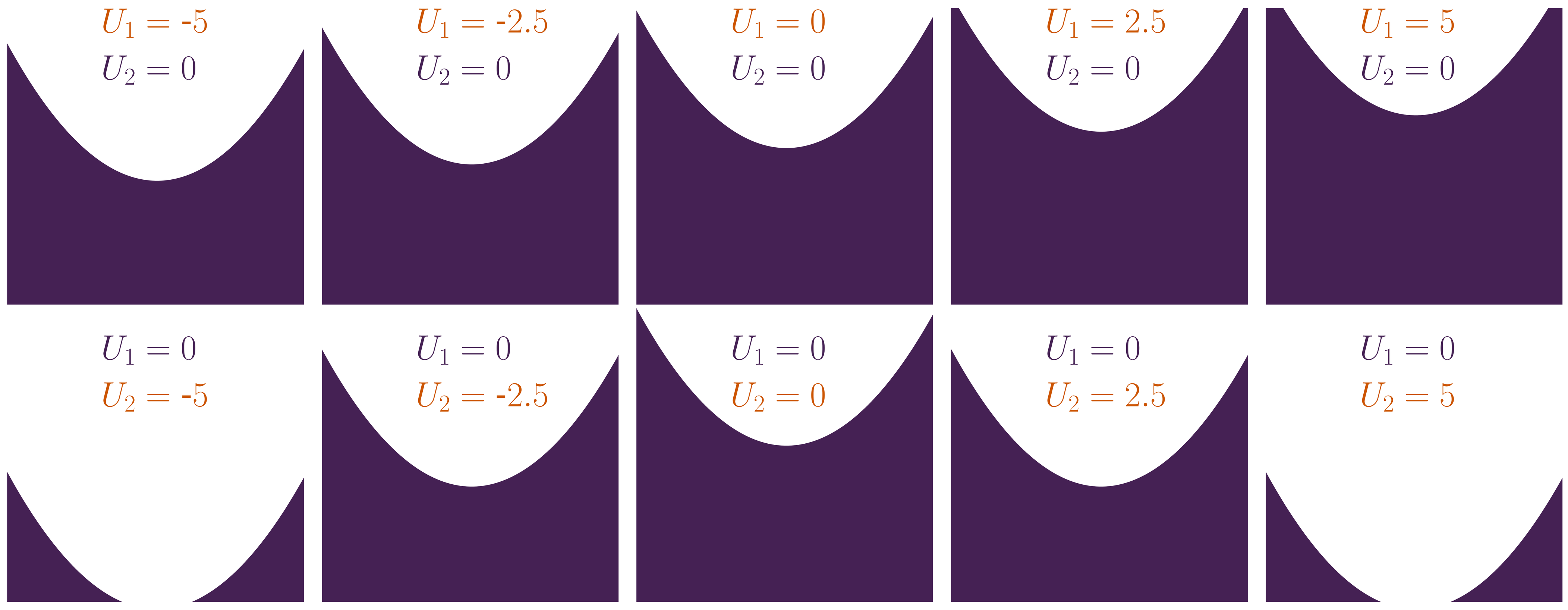
Answer: Performing Sensitivity analysis on excursion sets Γ_U using kernel-based methods. We propose a kernel k_{set} with which we compute HSIC-ANOVA indices.

$$(U_1,\dots,U_p)\mapsto \Gamma = \{x\in\mathcal{X},\ g(x,U)\leq 0\}$$

Example:

$$\Gamma = \{x\in[-5,5]^2,\ x_1^2+5x_2-U_1+U_2^2-1\leq 0\}$$

$\longrightarrow$ U_2 seems more influential than U_1 on the feasible sets.



Kernel-based Sensitivity Analysis

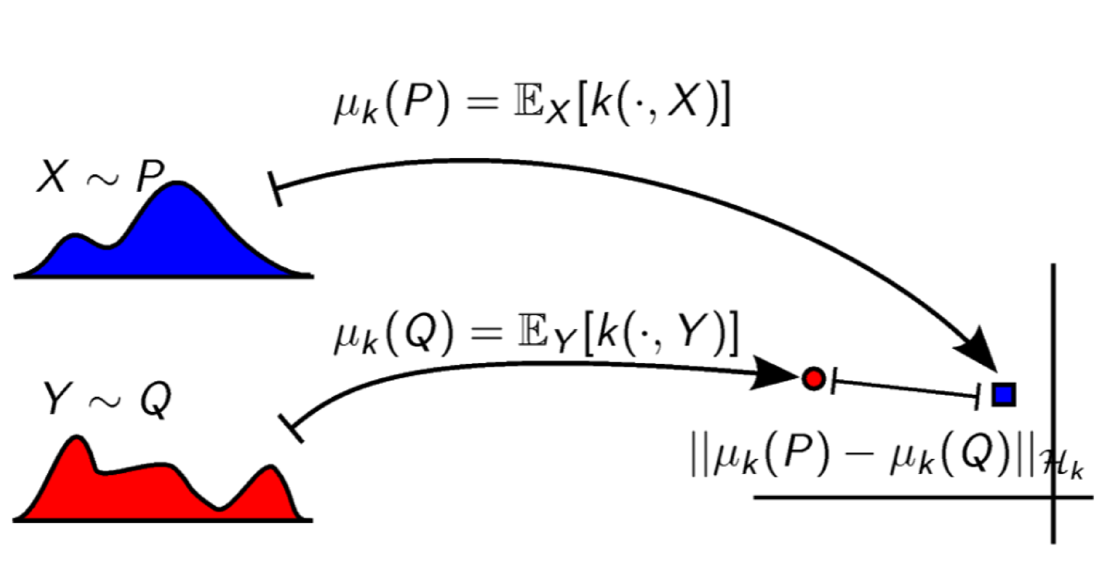


Fig. 1: Kernel mean embedding [6]

With $K = k_{\mathcal{X}}\otimes k_{\mathcal{Y}}$, the Hilbert Schmidt Independence Criterion (HSIC) is given by:

$$\begin{aligned}\mathrm{HSIC}_{k_\mathcal{X},k_\mathcal{Y}}(X,Y) &= ||\mu_K(X,Y)-\mu_{k_\mathcal{X}}(X)\otimes\mu_{k_\mathcal{Y}}(Y)||^2_{\mathcal{H}_K}\\ &= \mathbb{E}[k_\mathcal{X}(X,X')k_\mathcal{Y}(Y,Y')]\\ &\quad+\mathbb{E}[k_\mathcal{X}(X,X')]\mathbb{E}[k_\mathcal{Y}(Y,Y')]\\ &\quad-2\mathbb{E}[\mathbb{E}[k_\mathcal{X}(X,X')|X]\mathbb{E}[k_\mathcal{Y}(Y,Y')|Y]].\end{aligned}$$

When K is characteristic (injectivity of the mean embedding),

$$\mathrm{HSIC}_{k_\mathcal{X},k_\mathcal{Y}}(X,Y)) = 0 \text{ iif } X\perp Y.$$

$\longrightarrow$ screening

Assuming that the inputs are independent and that the input kernels are ANOVA,

$$\mathrm{HSIC}(\boldsymbol{X},Y) = \sum_{A\subseteq\{1,\dots,d\}}\sum_{B\subseteq A}(-1)^{|A|-|B|}\,\mathrm{HSIC}\left(\boldsymbol{X}_B,Y\right).$$

HSIC-ANOVA indices [3] are then defined as:

$$S_i^{\mathrm{HSIC}} := \frac{\mathrm{HSIC}(X_i,Y)}{\mathrm{HSIC}(\boldsymbol{X},Y)},$$

$$S_{T_i}^{\mathrm{HSIC}} := 1-\frac{\mathrm{HSIC}(\boldsymbol{X}_{-i},Y)}{\mathrm{HSIC}(\boldsymbol{X},Y)}$$

and are suited for ranking (and screening [5]).

HSIC for set-valued outputs

With $A\Delta B = A\cup B - B\cap A$ and μ the Lebesgue measure, we define a kernel on closed sets by:

$$\forall\Gamma_1,\Gamma_2\in\mathcal{F}(\mathcal{X}),\ k_{set}(\Gamma_1,\Gamma_2) = \exp\left(-\frac{\mu(\Gamma_1\Delta\Gamma_2)}{2\sigma^2}\right).$$

Proposition 1

k_{set} is a kernel [1] and is characteristic.

Given an input U_l and an associated ANOVA kernel k_l, the HSIC on sets that we call H_{set} is defined by:

$$\begin{aligned}\mathrm{H}_{set}(U_l,\Gamma) :&= \mathrm{HSIC}_{k_l,k_{set}}(U_l,\Gamma)\\ &= \mathbb{E}\left[(k_l(U_l,U_l')-1)k_{set}(\Gamma,\Gamma')\right].\end{aligned}$$

It can be estimated with its U-statistic:

$$\widehat{\mathrm{H}_{set}}\left(U_l,\Gamma\right) = \frac{2}{n(n-1)}\sum_{i<j}^{n}\left(k_l\left(U_l^{(i)},U_l^{(j)}\right)-1\right)k_{set}\left(\Gamma^{(i)},\Gamma^{(j)}\right).$$

k_{set} is rarely exactly known and requires to be estimated. Given $X^{(1)},\dots,X^{(m)}$ an iid sample of $X\sim\mathcal{U}(\mathcal{X})$, we propose the estimator:

$$\widehat{k_{set}}(\Gamma^{(i)},\Gamma^{(j)}) = e^{-\frac{\mu(\mathcal{X})}{2\sigma^2}\frac{1}{m}\sum_{k=1}^m \mathbb{1}_{\Gamma^{(i)}\Delta\Gamma^{(j)}}(X^{(k)})},$$

leading finally to the nested H_{set} estimator:

$$\widehat{\widehat{\mathrm{H}_{set}}}\left(U_l,\Gamma\right) = \frac{2}{n(n-1)}\sum_{i<j}^{n}\left(k_l\left(U_l^{(i)},U_l^{(j)}\right)-1\right)e^{-\frac{\mu(\mathcal{X})}{2\sigma^2}\frac{1}{m}\sum_{k=1}^m \mathbb{1}_{\Gamma^{(i)}\Delta\Gamma^{(j)}}(X^{(k)})}.$$

Proposition 2

The quadratic risk of the nested estimator $\widehat{\widehat{\mathrm{H}_{set}}}$ verifies:

$$\mathbb{E}\left(\widehat{\widehat{\mathrm{H}_{set}}}\left(U_l,\Gamma\right)-\mathrm{H}_{set}(U_l,\Gamma)\right)^2 \leq 2\left(\frac{2\sigma_1^2}{n(n-1)}+\frac{4(n-2)\sigma_2^2}{n(n-1)}+\frac{K^2\sigma_3^2}{m}\right).$$

HSIC-ANOVA indices on sets are then denoted $S_i^{\widehat{\widehat{\mathrm{H}_{set}}}}$ and $S_{T_i}^{\widehat{\widehat{\mathrm{H}_{set}}}}$.

Toy case 1 [4]

$$\forall x,u\in[-5,5]^2\times[-5,5]^3\quad g_1(x_1,x_2,u_1,u_2,u_3) = -x_1^2+5x_2-u_1+u_2^2-1$$

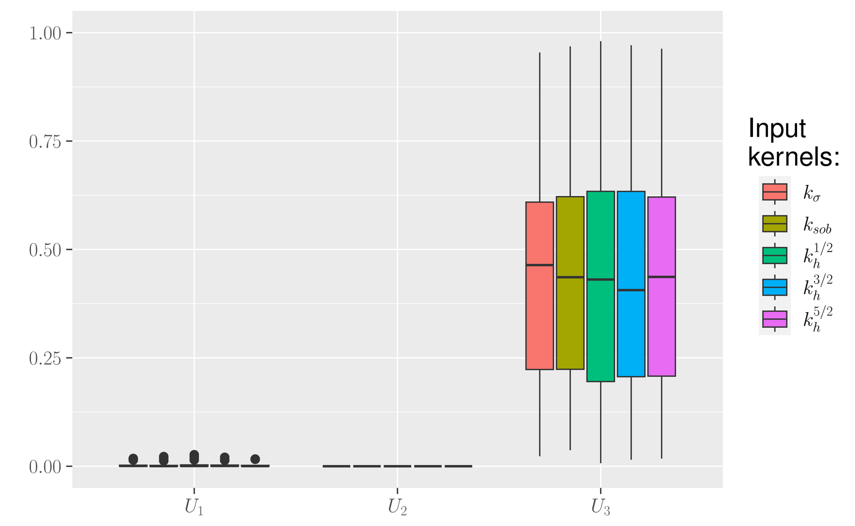


Fig. 2: p-value of $\widehat{\widehat{\mathrm{H}_{set}}}$ for $n=m=100$

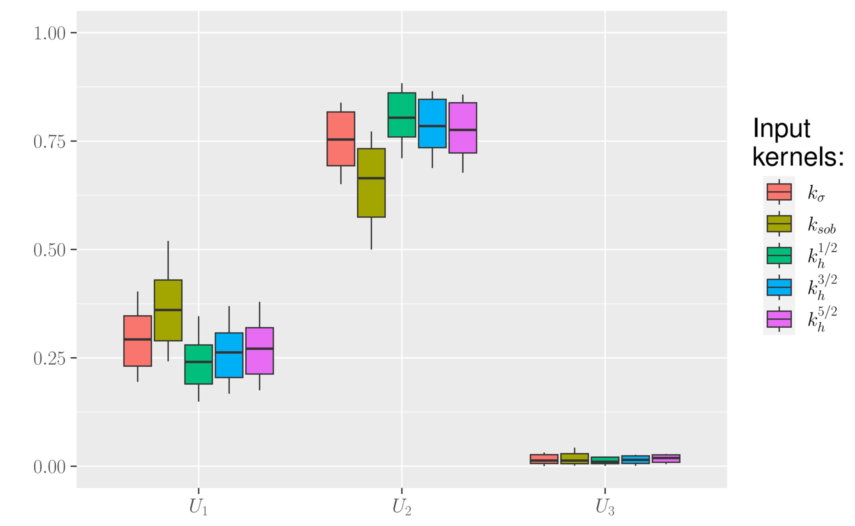


Fig. 3: $S_{T_i}^{\widehat{\widehat{\mathrm{H}_{set}}}}$ for $n=m=100$

Toy case 2 [2]

$$g_2(x_1,x_2,u_1,u_2,u_p,u_{r_1},u_{r_2}) = u_{r_2}-\max_{t\in[0,T]}\mathcal{Y}''(x_1+u_1,x_2+u_2,u_p;t),$$

with $\mathcal{Y}(x_1+u_1,x_2+u_2,u_p;t)$ the solution of the harmonic oscillator defined by:

$$(x_1+u_1)\mathcal{Y}''(t)+u_p\mathcal{Y}'(t)+(x_2+u_2)\mathcal{Y}(t) = \eta(t).$$

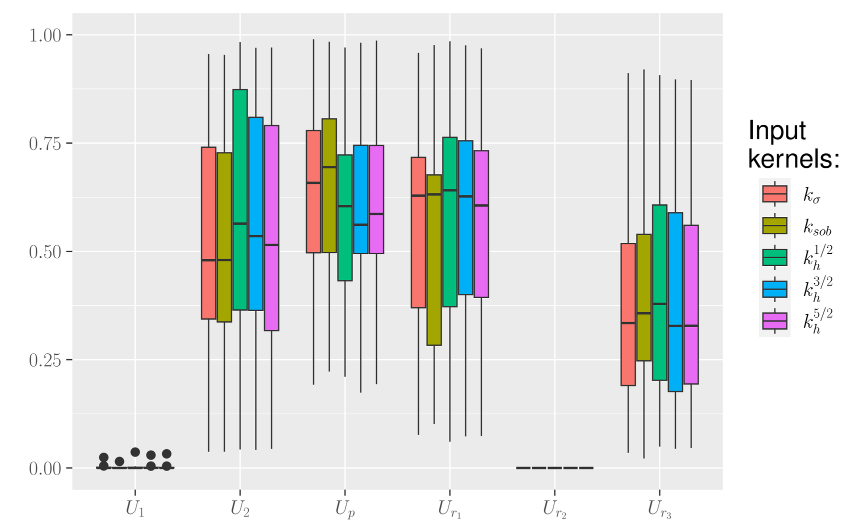


Fig. 4: p-value of $\widehat{\widehat{\mathrm{H}_{set}}}$ for $n=m=100$

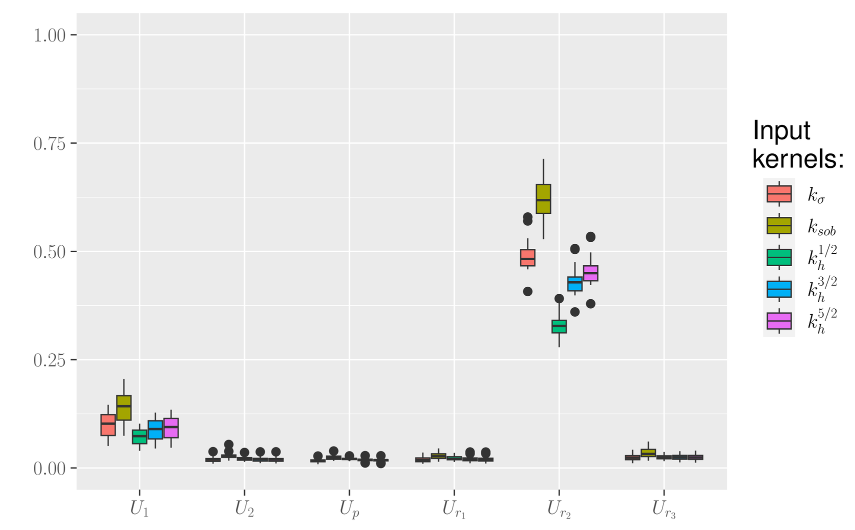


Fig. 5: $S_i^{\widehat{\widehat{\mathrm{H}_{set}}}}$ for $n=m=100$

Outlook

- Testing the indices on industrial applications (as viability kernels, or air pollutant concentration maps) and compared to other indices defined for set-valued output models.
- Incorporating this method inside a robust optimization methodology: reducing the uncertain input space dimension to get cheaper meta-models.

References

[1] P. Balança, E. Herbin. A set-indexed Ornstein-Uhlenbeck process. 2012.
[2] A. Cousin et al. A two-step procedure for time-dependent reliability-based design optimization involving piece-wise stationary Gaussian processes. 2022.
[3] S. da Veiga. Kernel-based ANOVA decomposition and Shapley effects - Application to global sensitivity analysis. 2021.
[4] R. El Amri et al. A sampling criterion for constrained Bayesian optimization with uncertainties. 2021
[5] G. Sarazin et al. Test d'indépendance basé sur les indices HSIC-ANOVA d'ordre total. 2022.
[6] D. Sejdinovic. Learning with Approximate Kernel Embeddings. RegML Workshop. 2017.