



INSTITUT
de MATHEMATIQUES
de TOULOUSE



CERAMATHS



Université
Polytechnique
HAUTS-DE-FRANCE

Gaussian processes under inequality constraints: Model selection and extension to high dimensions

A. F. López-Lopera, F. Bachoc and O. Roustant

November 30, 2023

Séminaire ULCO



A.F. López-Lopera
CERAMATHS, France



F. Bachoc
IMT, France



O. Roustant
INSA-IMT, France

High-dimensional Additive Gaussian Processes under Monotonicity Constraints

Andrés F. López-Lopera
CERAMATHS, UPHF
59313 Valenciennes, France
andres.lopezlopera@uphf.fr

François Bachoc
IMT, UMR5219 CNRS
Université Paul Sabatier
31062 Toulouse, France

Olivier Roustant
IMT, UMR5219 CNRS
INSA Toulouse
31077 Toulouse, France

Abstract

We introduce an additive Gaussian process framework accounting for monotonicity constraints and scalable to high dimensions. Our contributions are threefold. First, we show that our framework enables to satisfy the constraints everywhere in the input space. We also show that more general componentwise linear inequality constraints can be handled similarly, such as componentwise convexity. Second, we propose the additive MaxMod algorithm for sequential dimension reduction. By sequentially maximizing a squared-norm criterion, MaxMod identifies the active input dimensions and refines the most important ones. This criterion can be computed explicitly at a linear cost. Finally, we provide open-source codes for our full framework. We demonstrate the performance and scalability of the methodology in several synthetic examples with hundreds of dimensions under monotonicity constraints as well as on a real-world flood application.

Source: <https://proceedings.neurips.cc/>

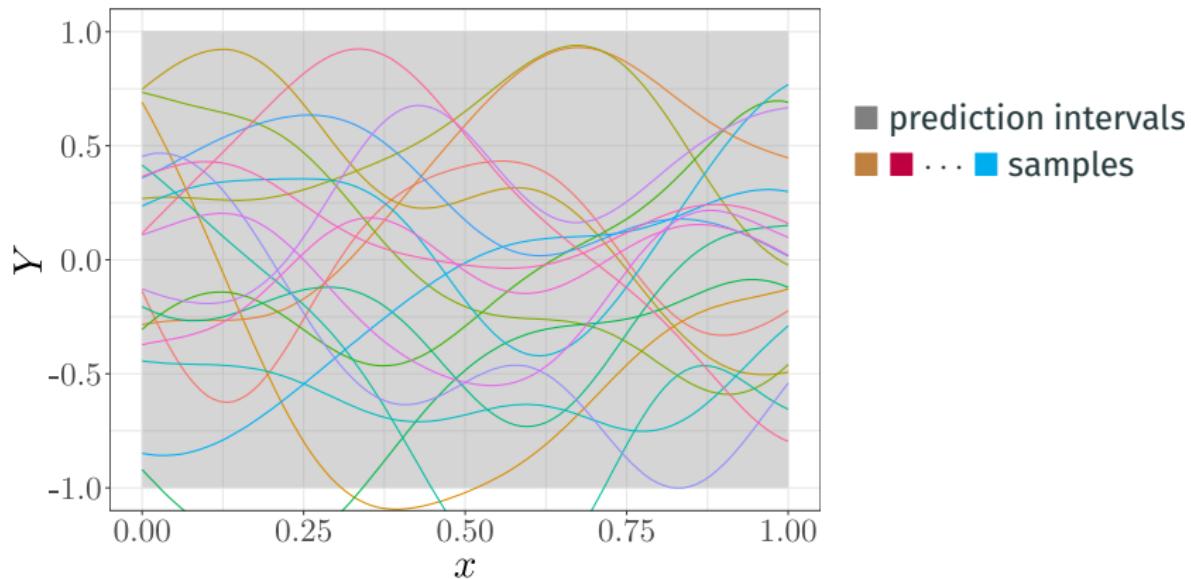
Table of contents

1. Constrained Gaussian processes
2. The MaxMod algorithm
3. Extension to additive functions
4. Numerical experiments

Constrained Gaussian processes

Gaussian processes (GPs)

GPs form a flexible **prior over functions** [Rasmussen and Williams, 2005]:



Gaussian processes (GPs)

- Let $\{Y(x); x \in \mathcal{D}\}$ be a stochastic process defined on a compact input space $\mathcal{D} \subseteq \mathbb{R}$ (e.g. $\mathcal{D} = [0, 1]$).
- Y is GP-distributed if, for all $x_1, \dots, x_n \in \mathcal{D}$,

$$\mathbf{Y}_n := \left[Y(x_1), \dots, Y(x_n) \right]^\top \sim \mathcal{N}(\boldsymbol{\mu}, \mathbf{K}).$$

with mean vector $\boldsymbol{\mu} \in \mathbb{R}^n$ and covariance matrix $\mathbf{K} \in \mathbb{R}^{n \times n}$.

- By convention, we denote the GP Y as

$$Y \sim \mathcal{GP}(\boldsymbol{\mu}, k),$$

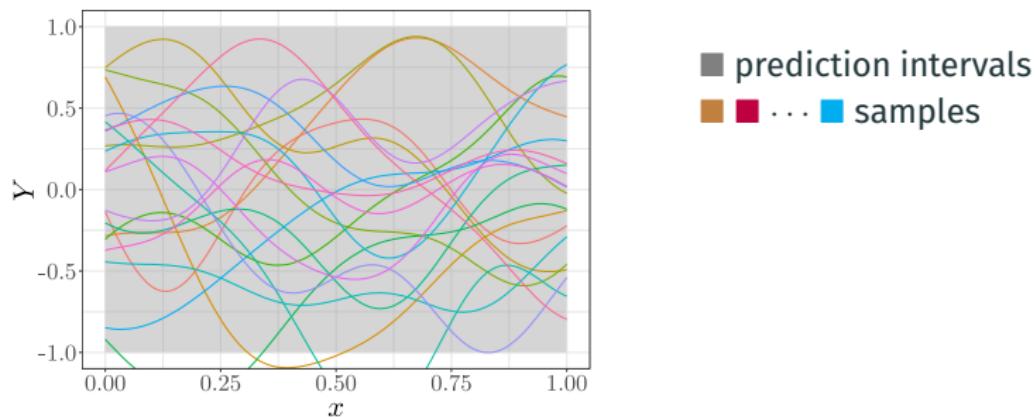
with mean function $\mu(x) = \mathbb{E}(Y(x))$ and covariance function (or kernel) $k(x, x') = \text{cov}(Y(x), Y(x')) = \mathbb{E}([Y(x) - \mu(x)][Y(x') - \mu(x')])$, for $x, x' \in \mathcal{D}$.

Gaussian processes (GPs)

- In practice, centered GPs priors Y are considered (i.e. $\mu(\cdot) = 0$). Then, Y is fully defined by its kernel function k .

Squared Exponential kernel:
$$k(x, x') = \sigma^2 \exp\left(-\frac{(x - x')^2}{2\ell}\right),$$

where σ^2 and ℓ are the variance and length-scale parameters, respectively.



[link]

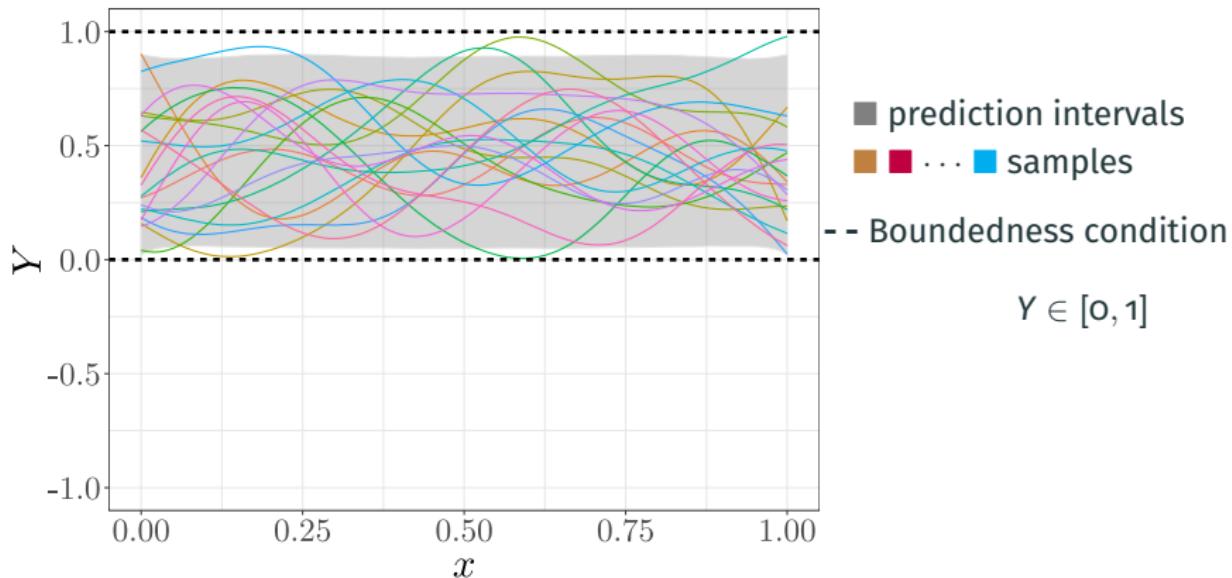
- Let $(\mathbf{X}, \mathbf{y}) = (x_i, y_i)_{1 \leq i \leq n}$ a training dataset.
- Then, the conditional distribution $Y(x^*) | (\mathbf{Y}_n = \mathbf{y})$ is Gaussian with mean and variance parameters:

$$\mu(x^*) = k(x^*, \mathbf{X})k(\mathbf{X}, \mathbf{X})^{-1}\mathbf{y},$$

$$v(x^*) = k(x^*, x^*) - k(x^*, \mathbf{X})k(\mathbf{X}, \mathbf{X})^{-1}k(\mathbf{X}, x^*),$$

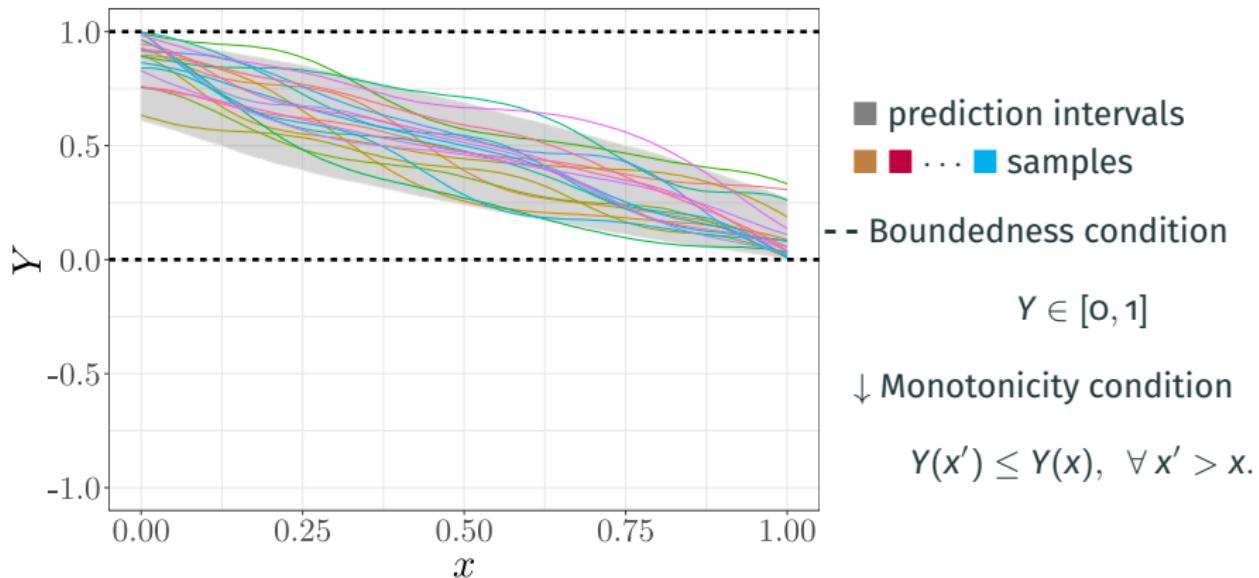
where $k(\mathbf{X}, \mathbf{X}) \in \mathbb{R}^{n \times n}$ and $k(x^*, \mathbf{X}) = (k(\mathbf{X}, x^*))^\top \in \mathbb{R}^n$.

Our interest: GP-based priors satisfying some inequality constraints...

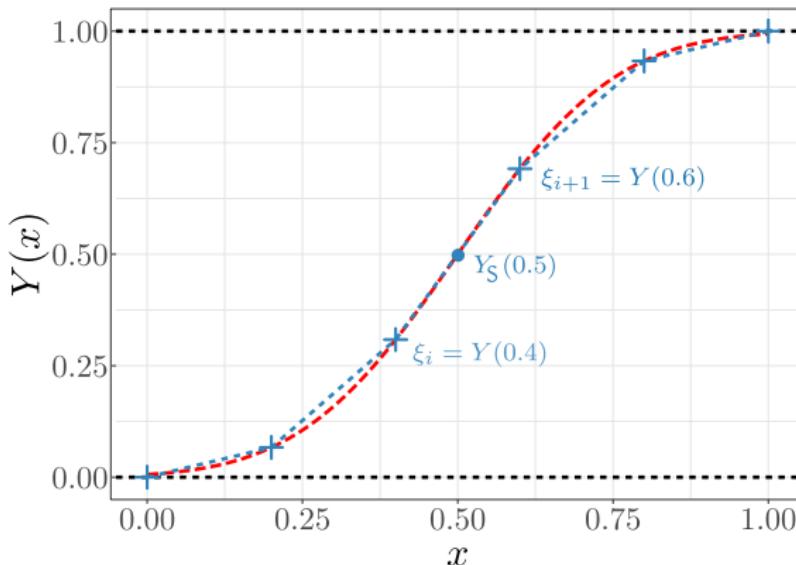


$$Y \in [0, 1]$$

Our interest: GP-based priors satisfying some inequality constraints...



Finite-dimensional approximation of GPs



- smooth function Y
- piecewise approximation Y_S

Note that:

- If $\xi_j \in [0, 1]$ for $j = 1, \dots, m$,
 $Y_S(0.5) \in [0, 1]$.
- Or if $\xi_j < \xi_{j+1}$ for $j = 1, \dots, m - 1$,
 $\xi_j < Y_S(0.5) < \xi_{j+1}$.

Pro: imposing constraints over knots is enough [Maatouk and Bay, 2017]:

$$Y_S \in \mathcal{E} \Leftrightarrow \xi \in \mathcal{C}.$$

Finite-dimensional approximation of GPs

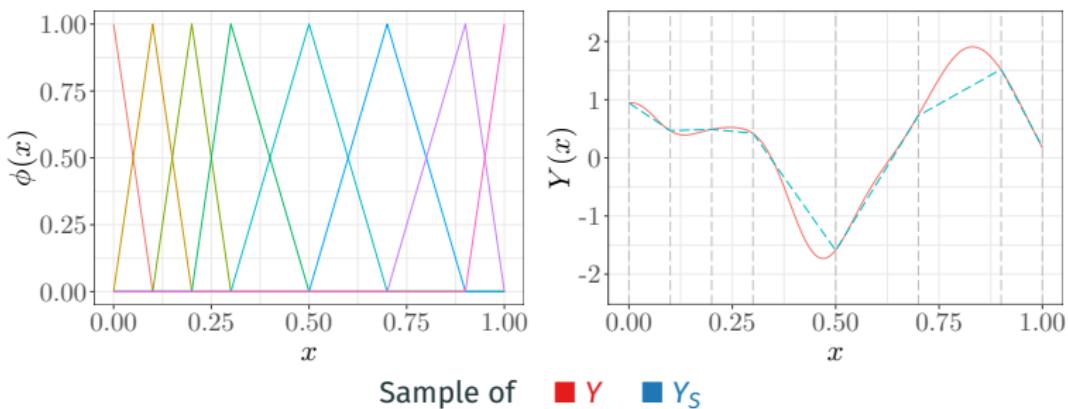
- Let \mathbf{Y}_S be the finite-dimensional GP with an ordered set of knots:

$$\mathbf{S} = \{t_0, \dots, t_m\}, \quad \text{with } 0 = t_0 < \dots < t_m = 1,$$

such that

$$\mathbf{Y}_S(x) = \sum_{j=1}^m \mathbf{Y}(t_j) \phi_j(x), \quad (1)$$

where $x \in [0, 1]$, $\mathbf{Y} \sim \mathcal{GP}(0, k_\theta)$, and $\phi_j : [0, 1] \mapsto \mathbb{R}$ are (asymmetric) hat basis functions.



Finite-dimensional approximation of GPs

- Then, for regression tasks under inequality constraints, we have

$$Y_S(x) = \sum_{j=1}^m \xi_j \phi_j(x), \text{ s.t. } \begin{cases} Y_S(x_i) + \varepsilon_i = y_i & (\text{regression conditions}), \\ \mathbf{l} \leq \boldsymbol{\Lambda} \boldsymbol{\xi} \leq \mathbf{u} & (\text{linear inequality conditions}), \end{cases} \quad (2)$$

where $x_i \in [0, 1]$, $y_i \in \mathbb{R}$ for $i = 1, \dots, n$, and

- $\xi_j := Y(t_j)$ for $j = 1, \dots, m$, i.e. $\boldsymbol{\xi} = [\xi_1, \dots, \xi_m]^\top \sim \mathcal{N}(\mathbf{o}, \boldsymbol{\Sigma}_\theta)$ with covariance matrix $\boldsymbol{\Sigma}_\theta = (\mathbf{k}_\theta(t_j, t_{j'}))_{1 \leq j, j' \leq m}$
- $\varepsilon_i \sim \mathcal{N}(\mathbf{o}, \tau^2)$, for $i = 1, \dots, n$, with noise variance τ^2
- $(\boldsymbol{\Lambda}, \mathbf{l}, \mathbf{u})$ define the ineq. constraints. For instance, for the case of monotonicity, we have

$$\underbrace{\begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}}_{\mathbf{l}} \leq \underbrace{\begin{bmatrix} -1 & 1 & 0 & \cdots & 0 & 0 \\ 0 & -1 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & -1 & 1 \end{bmatrix}}_{\boldsymbol{\Lambda}} \underbrace{\begin{bmatrix} \xi_1 \\ \xi_2 \\ \vdots \\ \xi_m \end{bmatrix}}_{\boldsymbol{\xi}} \leq \underbrace{\begin{bmatrix} \infty \\ \infty \\ \vdots \\ \infty \end{bmatrix}}_{\mathbf{u}}$$

- Since $Y \in \mathcal{E} \Leftrightarrow \xi \in \mathcal{C}$, then *uncertainty quantification* relies on simulating the **truncated vector** ξ [López-Lopera et al., 2018]:

$$\Lambda \xi | \{\Phi \xi + \varepsilon = \mathbf{y}, \mathbf{l} \leq \Lambda \xi \leq \mathbf{u}\} \sim \mathcal{T}\mathcal{N}(\Lambda \mu_c, \Lambda \Sigma_c \Lambda^\top, \mathbf{l}, \mathbf{u}), \quad (3)$$

with conditional parameters μ_c and Σ_c given by

$$\mathbf{K} = \Phi \Sigma \Phi^\top + \tau^2 \mathbf{I}, \quad \mu_c = \Sigma \Phi^\top \mathbf{K}^{-1} \mathbf{y}, \quad \Sigma_c = \Sigma - \Sigma \Phi^\top \mathbf{K}^{-1} \Phi \Sigma. \quad (4)$$

- * Eq. (3) is computed via *Monte Carlo* (MC) or *Markov Chain MC* (MCMC):
 - e.g. *Hamiltonian Monte Carlo* (HMC) [Pakman and Paninski, 2014]
 - A. López-Lopera, N. Durrande, F. Bachoc and O. Roustant, Finite-dimensional Gaussian approximation with linear inequality constraints, SIAM/ASA J. on Uncertainty Quantification, 2018.

The maximum a posteriori (mode) function in 1D

- Let $\hat{\xi}$ be the mode that maximises the pdf of $\xi | \{\Phi \xi + \varepsilon = \mathbf{y}, \mathbf{l} \leq \Lambda \xi \leq \mathbf{u}\}$:

$$\hat{\xi} = \underset{\xi \text{ s.t. } \mathbf{l} \leq \Lambda \xi \leq \mathbf{u}}{\arg \max} \{-[\xi - \mu_c]^\top \Sigma_c^{-1} [\xi - \mu_c]\}, \quad (5)$$

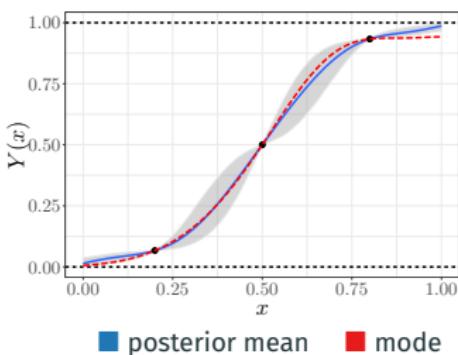
with $\hat{\xi} = [\hat{\xi}_1, \dots, \hat{\xi}_m]^\top$.

- The MAP estimate of Y_S is given by

$$\hat{Y}_S(x) = \sum_{j=1}^m \hat{\xi}_j \phi_j(x). \quad (6)$$

Pro:

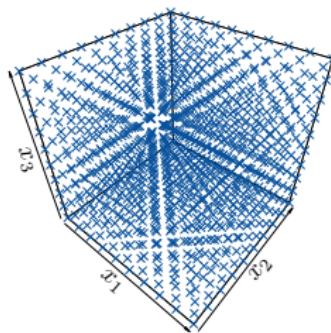
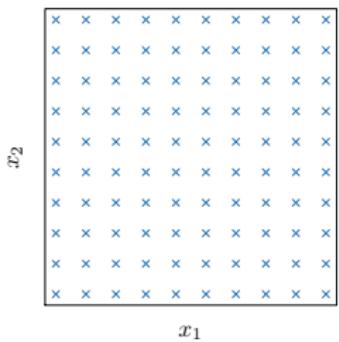
- \hat{Y}_S can be used as a point estimate
- Easy and fast calculations
- Convergence to the spline solution as $m \rightarrow \infty$ [Bay et al., 2016]
- Starting point for MCMC



The maximum a posteriori (mode) function in 1D

- **Con:** the cost of Y_S increases as d increases.

$$Y_S(\mathbf{x}) = \sum_{j_1, \dots, j_d=1}^{m_1, \dots, m_d} \left[\prod_{p=1, \dots, d} \phi_{j_p}^{(p)}(x_p) \right] \xi_{j_1, \dots, j_d}, \quad \text{s.t.} \quad \begin{cases} Y_S(x_i) + \varepsilon_i = y_i, \\ \xi \in \mathcal{C}. \end{cases} \quad (7)$$



- This drawback can be mitigated by considering:
 - a “smarter” construction of rectangular grids of knots thanks to the asymmetric construction of the hat basis functions
 - and/or further assumptions for complexity simplification
 - e.g. inactive variables

The MaxMod algorithm

The MaxMod algorithm in 1D

- Let $\widehat{Y}_{\mathbf{S}}$ be the MAP function with an ordered set of knots:

$$\mathbf{S} = \{t_0, \dots, t_m\}, \quad \text{with } 0 = t_0 < \dots < t_m = 1.$$

- Here, we aim at adding a new knot t in \mathbf{S} (where?)
- To do so, we aim at *maximising the total modification of the MAP*:

$$I_{\mathbf{S}}(t) = \int_{[0,1]} \left(\widehat{Y}_{\mathbf{S} \cup \{t\}}(x) - \widehat{Y}_{\mathbf{S}}(x) \right)^2 dx. \quad (8)$$

- The integral in (8) has a closed-form expression.

The MaxMod algorithm in 1D

- Let \widehat{Y}_S be the MAP function with an ordered set of knots:

$$S = \{t_0, \dots, t_m\}, \quad \text{with } 0 = t_0 < \dots < t_m = 1.$$

- Here, we aim at adding a new knot t in S (where?)
- To do so, we aim at *maximising the total modification of the MAP*:

$$I_S(t) = \int_{[0,1]} \left(\widehat{Y}_{S \cup t}(x) - \widehat{Y}_S(x) \right)^2 dx. \quad (8)$$

- The integral in (8) has a closed-form expression.

Algorithm MaxMod (maximum modification of the MAP) in 1D

Input parameters: the initial subdivision $S^{(0)} \in \mathcal{S}$.

Sequential procedure: for $\kappa \in \mathbb{N}$, do:

1: Set $t_{\kappa+1}^* \in [0, 1]$ such that

$$I_{S^{(\kappa)}}(t_{\kappa+1}^*) \geq \sup_{t \in [0,1]} I_{S^{(\kappa)}}(t)$$

2: $S^{(\kappa+1)} = S^{(\kappa)} \cup t_{\kappa+1}^*$.

1D example under boundedness and monotonicity constraints

MAP estimate

conditional sample-path

- training points
- predictive mean
- + knots
- MAP estimate
- 90% confidence intervals

The MaxMod algorithm in higher dimensions

- Let $\widehat{Y}_{\mathcal{J}, \mathbf{S}}$ be the MAP function with $|\mathcal{J}|$ active variables and ordered sets of knots $\mathbf{S}_{\mathcal{J}}$ for $\mathcal{J} \subseteq \{1, \dots, D\}$.
- Then, the criterion to maximise is given by

$$I_{\mathcal{J}, \mathbf{S}}(i, t) = \begin{cases} \frac{1}{N_{\mathbf{S}, \mathcal{J}, i}} \int_{[0,1]^d} (\widehat{Y}_{\mathcal{J}, \mathbf{S} \cup_i t}(\mathbf{x}) - \widehat{Y}_{\mathcal{J}, \mathbf{S}}(\mathbf{x}))^2 d\mathbf{x} & \text{if } i \in \mathcal{J}, \\ \frac{1}{N_{\mathbf{S}, \mathcal{J}, i}} \int_{[0,1]^{d+1}} (\widehat{Y}_{\mathcal{J} \cup \{i\}, \mathbf{S} + i}(\mathbf{x}) - \widehat{Y}_{\mathcal{J}, \mathbf{S}}(\mathbf{x}))^2 d\mathbf{x} & \text{if } i \notin \mathcal{J}, \end{cases} \quad (9)$$

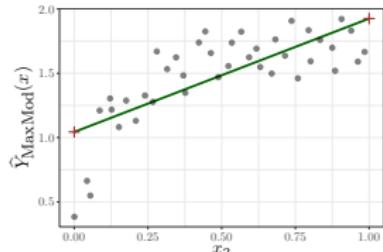
where $N_{\mathbf{S}, \mathcal{J}, i}$ is the increase of the number of basis functions.

- F. Bachoc, A. López-Lopera, and O. Roustant. Sequential construction and dimension reduction of GPs under inequality constraints. SIAM J. on Maths. of Data Science, 2022.

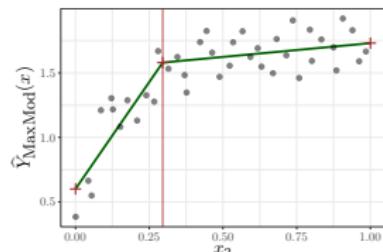
The MaxMod algorithm in higher dimensions

2D example under monotonicity constraints

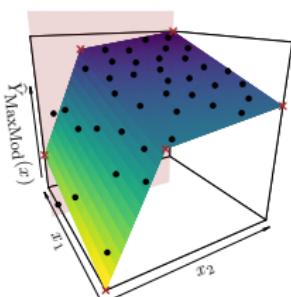
Evolution of the MaxMod algorithm using $f(x) = \frac{1}{2}x_1 + \arctan(10x_2)$



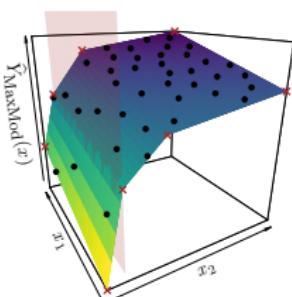
(a) iteration 0



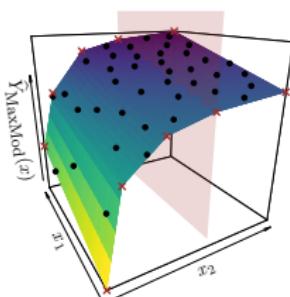
(b) iteration 1



(c) iteration 2



(d) iteration 3

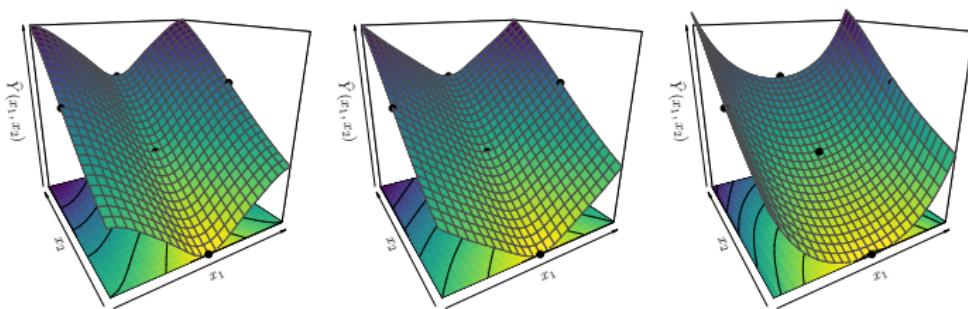


(e) iteration 4

- training points
- + knots
- MAP estimate

The MaxMod algorithm in higher dimensions

- The constrained GP is tractable depending on $|\mathcal{J}|$ (nb of active variable).
- According to numerical tests, our framework is limited to $|\mathcal{J}| \leq 5$.
- Therefore, further assumptions are required to scale the model:
 - e.g. **additive structures**



Additive GP predictions using (left) the unconstrained GP mean, (center) the cGP mode and (right) the cGP mean via HMC. The constrained model accounts for both componentwise convexity and monotonicity conditions along x_1 and x_2 , respectively.

Extension to additive functions

- In high dimension, many statistical regression models are based on additive structures of the form:

$$y(\mathbf{x}) = y_1(x_1) + \cdots + y_d(x_d). \quad (10)$$

- Then GP priors can be placed over y_1, \dots, y_d [Durrande et al., 2012]

$$Y_i \sim \mathcal{GP}(\mathbf{o}, k_i),$$

for $i = 1, \dots, d$. Taking Y_1, \dots, Y_d as independent GPs, the process

$$Y(\mathbf{x}) = Y_1(x_1) + \cdots + Y_d(x_d)$$

is also a GP and its kernel is given by

$$k(\mathbf{x}, \mathbf{x}') = k_1(x_1, x'_1) + \cdots + k_d(x_d, x'_d). \quad (11)$$

Finite-dimensional approximation of additive GPs

- For the constrained case, we can approximate Y_i by a finite-dimensional GP:

$$Y_{i,S_i}(x_i) = \sum_{j=1}^{m_i} \xi_{i,j} \phi_{i,j}(x_i),$$

with one-dimensional subdivision S_i , and m_i knots.

- We let $S = (S_1, \dots, S_d)$. The finite-dimensional GP is written,

$$Y_S(\mathbf{x}) = \sum_{i=1}^d Y_{i,S_i}(x_i) = \sum_{i=1}^d \sum_{j=1}^{m_i} \xi_{i,j} \phi_{i,j}(x_i), \quad (12)$$

where $\xi_{i,j} = Y_i(t_{(j)}^{(S_i)})$ and $\phi_{i,j} : [0, 1] \mapsto \mathbb{R}$ are asymmetric hat basis functions.

- One can note that the total number of knots is given by $m = m_1 + \dots + m_d$.

- Observe from (12) that, since $\xi_{i,j}$, for $i = 1, \dots, d$ and $j = 1, \dots, m_i$, are Gaussian distributed, then Y_{i,S_i} is a GP with kernel given by

$$\tilde{k}_i(x_i, x'_i) = \sum_{j=1}^{m_i} \sum_{\kappa=1}^{m_i} \phi_{i,j}(x_i) \phi_{i,\kappa}(x'_i) k_i(t_{(j)}^{(S_i)}, t_{(\kappa)}^{(S_i)}). \quad (13)$$

Moreover, Y_S is a GP with kernel $\tilde{k}(\mathbf{x}, \mathbf{x}') = \sum_{i=1}^d \tilde{k}_i(x_i, x'_i)$.

- We let $\Sigma_i = k_i(S_i, S_i)$ be the $m_i \times m_i$ covariance matrix of ξ_i .

- We consider the componentwise constraints $Y_{i,S_i} \in \mathcal{E}_i$, $i = 1, \dots, d$ such that

$$Y_{i,S_i} \in \mathcal{E}_i \Leftrightarrow \xi_i \in \mathcal{C}_i \quad (14)$$

where $\xi_i = [\xi_{i,1}, \dots, \xi_{i,m_i}]^\top$ and $\mathcal{C}_i = \{\mathbf{c} \in \mathbb{R}^{m_i} : \mathbf{l}_i \leq \Lambda_i \mathbf{c} \leq \mathbf{u}_i\}$.

- Examples of constraints are monotonicity and componentwise convexity.
- Given the observations and the constraints, the MAP estimate is given by

$$\hat{Y}_S(\mathbf{x}) = \sum_{i=1}^d \sum_{j=1}^{m_i} \hat{\xi}_{i,j} \phi_{i,j}(x_i). \quad (15)$$

- As in (5), the vector $\hat{\xi} = [\hat{\xi}_1^\top, \dots, \hat{\xi}_d^\top]^\top$ with $\hat{\xi}_i = [\hat{\xi}_{i,1}, \dots, \hat{\xi}_{i,m_i}]^\top$ is given by

$$\hat{\xi} = \underset{\substack{\xi = (\xi_1^\top, \dots, \xi_d^\top)^\top \\ l_i \leq \Lambda_i \xi_i \leq u_i, i=1, \dots, d}}{\operatorname{argmin}} (\xi - \mu_c)^\top \Sigma_c^{-1} (\xi - \mu_c), \quad (16)$$

where $\mu_c = [\mu_{c,1}^\top, \dots, \mu_{c,d}^\top]^\top$ is the $m \times 1$ vector with block i given by

$$\mu_{c,i} = \Sigma_i \Phi_i^\top \left[\left(\sum_{p=1}^d \Phi_p \Sigma_p \Phi_p^\top \right) + \tau^2 I_n \right]^{-1} \mathbf{y}_n, \quad (17)$$

and $(\Sigma_{c,i,j})_{i,j}$ is the $m \times m$ matrix with block (i,j) given by

$$\Sigma_{c,i,j} = \mathbf{1}_{i=j} \Sigma_i - \Sigma_i \Phi_i^\top \left[\left(\sum_{p=1}^d \Phi_p \Sigma_p \Phi_p^\top \right) + \tau^2 I_n \right]^{-1} \Phi_j \Sigma_j. \quad (18)$$

Remarks:

- $\Sigma_{c,i,j}$ involves contributions of the cross-covariances.
- The inversion is computed efficiently for $m \ll n$ (matrix inv. lemma).

Additive MaxMod algorithm

- Consider an additive cGP model that uses only a subset $\mathcal{J} \subseteq \{1, \dots, d\}$ of active variables.
- Its mode function \widehat{Y}_S , from $\mathbb{R}^{|\mathcal{J}|}$ to \mathbb{R} , by, for $\mathbf{x} = (x_i; i \in \mathcal{J})$,

$$\widehat{Y}_S(\mathbf{x}) = \sum_{i \in \mathcal{J}} \sum_{j=1}^{m_i} \widehat{\xi}_{i,j} \phi_{i,j}(x_i). \quad (19)$$

- We measure this benefit by the squared-norm modification of the cGP mode

$$I_{S,i^*} = \int_{[0,1]^{|\mathcal{J}|+1}} \left(\widehat{Y}_{S,i^*}(\mathbf{x}) - \widehat{Y}_S(\mathbf{x}) \right)^2 d\mathbf{x} \text{ for } i^* \notin \mathcal{J}, \quad (20)$$

$$I_{S,i^*,t} = \int_{[0,1]^{|\mathcal{J}|}} \left(\widehat{Y}_{S,i^*,t}(\mathbf{x}) - \widehat{Y}_S(\mathbf{x}) \right)^2 d\mathbf{x} \text{ for } i^* \in \mathcal{J}. \quad (21)$$

- Both (20) and (21) have analytic expression assuming $x_i \sim \text{Uniform}(0, 1)$ for $i = 1, \dots, d$ (see López-Lopera et al. [2022]), where the computational cost is linear with respect to $m = \sum_{i \in \mathcal{J}} m_i$.

Extension to additive functions

- For a new variable $i^* \notin \mathcal{J}$, the new mode function is

$$\widehat{Y}_{S,i^*}(\mathbf{x}) = \sum_{i \in \mathcal{J}} \sum_{j=1}^{m_i} \widetilde{\xi}_{i,j} \phi_{i,j}(x_i) + \sum_{j=1}^2 \widetilde{\xi}_{i^*,j} \phi_{i^*,j}(x_{i^*})$$

- We let $\phi_{i^*,1}(u) = 1 - u$ and $\phi_{i^*,2}(u) = u$ for $u \in [0, 1]$.

Proposition (Computation of I_{S,i^*})

We have

$$I_{S,i^*} = \sum_{i \in \mathcal{J}} \sum_{\substack{j, j'=1 \\ |j-j'| \leq 1}}^{m_i} \eta_{i,j} \eta_{i,j'} E_{j,j'}^{(S_i)} - \sum_{i \in \mathcal{J}} \left(\sum_{j=1}^{m_i} \eta_{i,j} E_j^{(S_i)} \right)^2 + \frac{\eta_{i^*}^2}{12} + \left(\sum_{i \in \mathcal{J}} \sum_{j=1}^{m_i} \eta_{i,j} E_j^{(S_i)} - \frac{\zeta_{i^*}}{2} \right)^2,$$

where $\eta_{i,j} = \widehat{\xi}_{i,j} - \widetilde{\xi}_{i,j}$, $\eta_{i^*} = \widetilde{\xi}_{i^*,2} - \widetilde{\xi}_{i^*,1}$, $\zeta_{i^*} = \widetilde{\xi}_{i^*,1} + \widetilde{\xi}_{i^*,2}$, $E_j^{(S_i)} := \int_0^1 \phi_{i,j}(t) dt$ and $E_{j,j'}^{(S_i)} := \int_0^1 \phi_{i,j}(t) \phi_{i,j'}(t) dt$ with explicit expressions in Lemma 1

[López-Lopera et al., 2022, Appendix A.3]. The matrices $(E_{j,j'}^{(S_i)})_{1 \leq j,j' \leq m_i}$ are 1-band and the computational cost is linear w.r.t. $m = \sum_{i \in \mathcal{J}} m_i$.

Additive MaxMod algorithm

- For a new t added to S_{i^*} with $i^* \in \mathcal{J}$, the new mode function is

$$\widehat{Y}_{S_{i^*}, t}(\mathbf{x}) = \sum_{i \in \mathcal{J}} \sum_{j=1}^{\tilde{m}_i} \widetilde{\xi}_{i,j} \widetilde{\phi}_{i,j}(x_i),$$

where $\tilde{m}_i = m_i$ for $i \neq i^*$, $\tilde{m}_{i^*} = m_{i^*} + 1$, $\widetilde{\phi}_{i,j} = \phi_{i,j}$ for $i \neq i^*$, and $\widetilde{\phi}_{i^*,j}$ is obtained from $S_{i^*} \cup \{t\}$ as in Proposition 1.

Proposition (Computation of $I_{S_{i^*}, t}$)

For $i \in \mathcal{J} \setminus \{i^*\}$, let $\tilde{S}_i = S_i$. Let $\tilde{S}_{i^*} = S_{i^*} \cup \{t\}$. Recall that the knots in S_{i^*} are written $0 = t_{(1)}^{(S_{i^*})} < \dots < t_{(m_{i^*})}^{(S_{i^*})} = 1$. Let $\nu \in \{1, \dots, m_{i^*} - 1\}$ be such that

$t_{(\nu)}^{(S_{i^*})} < t < t_{(\nu+1)}^{(S_{i^*})}$. Then, with a linear cost w.r.t. $\tilde{m} = \sum_{i \in \mathcal{J}} \tilde{m}_i$, we have

$$I_{S_{i^*}, t} = \sum_{i \in \mathcal{J}} \sum_{\substack{j, j'=1 \\ |j-j'| \leq 1}} \bar{\eta}_{i,j} \bar{\eta}_{i,j'} E_{j,j'}^{(\tilde{S}_i)} - \sum_{i \in \mathcal{J}} \left(\sum_{j=1}^{\tilde{m}_i} \bar{\eta}_{i,j} E_j^{(\tilde{S}_i)} \right)^2 + \left(\sum_{i \in \mathcal{J}} \sum_{j=1}^{\tilde{m}_i} \bar{\eta}_{i,j} E_j^{(\tilde{S}_i)} \right)^2,$$

where $\bar{\eta}_{i,j} = \bar{\xi}_{i,j} - \widetilde{\xi}_{i,j}$, $\bar{\xi}_{i,j} = \widehat{\xi}_{i,j}$ for $i \neq i^*$, $\bar{\xi}_{i^*,j} = \widehat{\xi}_{i^*,j}$ for $j \leq \nu$, $\bar{\xi}_{i^*,j} = \widehat{\xi}_{i^*,j-1}$ for $j \geq \nu + 2$, and

$$\bar{\xi}_{i^*,\nu+1} = \widehat{\xi}_{i^*,\nu} \frac{t_{(\nu+1)}^{(S_{i^*})} - t}{t_{(\nu+1)}^{(S_{i^*})} - t_{(\nu)}^{(S_{i^*})}} + \widehat{\xi}_{i^*,\nu+1} \frac{t - t_{(\nu)}^{(S_{i^*})}}{t_{(\nu+1)}^{(S_{i^*})} - t_{(\nu)}^{(S_{i^*})}}.$$

Numerical experiments

Numerical experiments: Monotonicity in hundreds of dimensions

- We consider the target function:

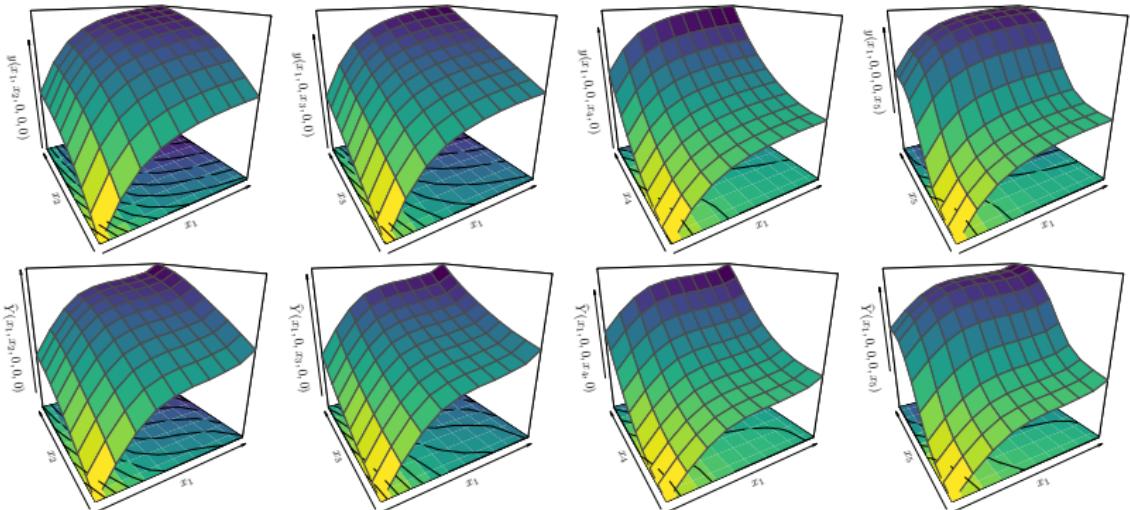
$$y(\mathbf{x}) = \sum_{i=1}^d \arctan \left(5 \left[1 - \frac{i}{d+1} \right] x_i \right). \quad (22)$$

with $\mathbf{x} \in [0, 1]^d$. y exhibits decreasing growth rates as the index i increases.

Results (mean \pm one standard deviation over 10 replicates) with $n = 2d$. For the computation of the cGP mean, 10^3 (${}^\dagger 50$) HMC samples are used.

d	m	CPU Time [s]		GP mean	Q^2 [%]	
		cGP mode	cGP mean		cGP mode	cGP mean
10	50	0.1 ± 0.1	0.1 ± 0.1	82.3 ± 6.2	83.8 ± 4.2	88.1 ± 1.7
100	500	0.4 ± 0.1	5.2 ± 0.5	89.8 ± 1.6	90.7 ± 1.4	91.5 ± 1.3
250	1250	4.2 ± 0.7	132.3 ± 26.3	91.7 ± 0.8	92.9 ± 0.6	93.4 ± 0.6
500	2500	37.0 ± 11.4	${}^\dagger 156.9 \pm 40.5$	92.5 ± 0.6	93.8 ± 0.5	${}^\dagger \mathbf{94.3 \pm 0.5}$
1000	5000	262.4 ± 35.8	${}^\dagger 10454.3 \pm 3399.3$	92.6 ± 0.3	94.6 ± 0.2	${}^\dagger \mathbf{95.1 \pm 0.2}$

Numerical experiments: Monotonicity in hundreds of dimensions



2D projections of the true profiles (top) and the constrained GP predictions (bottom)

Numerical experiments: Dimension reduction illustration

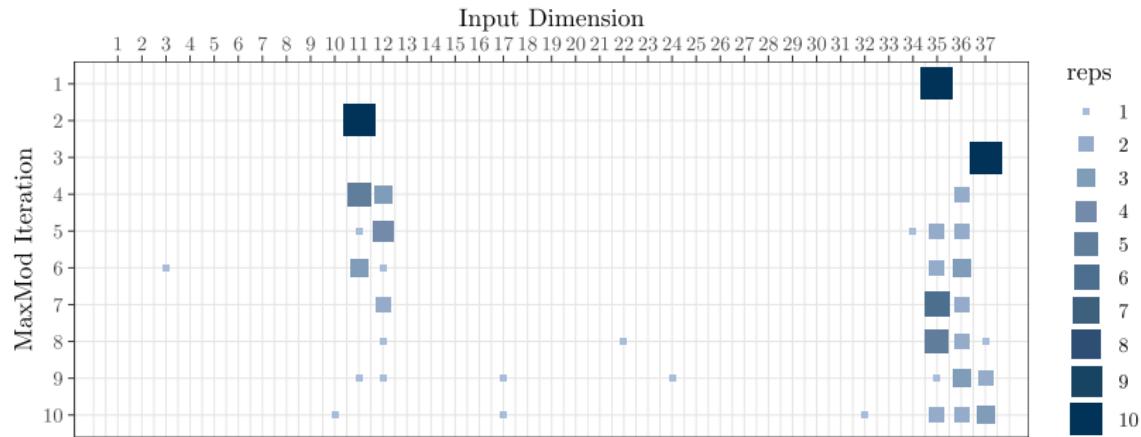
- We test the capability of MaxMod to account for dimension reduction considering the function in (22).
- In addition to (x_1, \dots, x_d) , we include $D - d$ virtual variables, indexed as (x_{d+1}, \dots, x_D) , which will compose the subset of inactive dimensions.
 - $\widehat{Y}_{\text{MaxMod}}$: the mode of the additive cGP and MaxMod.
 - $\widetilde{Y}_{\text{MaxMod}}$: the mode of the non-additive cGP and MaxMod.

Q^2 Performance of the MaxMod algorithm with $n = 10D$.

D	d	active dimensions	knots per dimension	$Q^2(\widetilde{Y}_{\text{MaxMod}}) [\%]$	$Q^2(\widehat{Y}_{\text{MaxMod}}) [\%]$
10	2	(1, 2)	(4, 3)	99.5	99.8
	3	(1, 2, 3)	(5, 5, 3)	97.8	99.8
	5	(1, 2, 3, 4, 5)	(4, 4, 4, 3, 2)	91.4	99.8
20	2	(1, 2)	(5, 3)	99.7	99.8
	3	(1, 2, 3)	(4, 4, 3)	99.0	99.9
	5	(1, 2, 3, 4, 5)	(5, 4, 3, 3, 2)	96.0	99.7

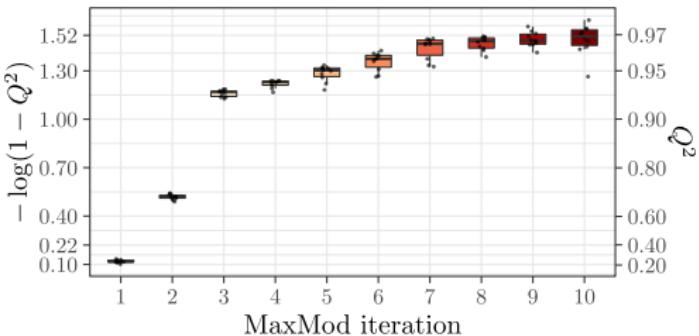
- The database contains a flood study conducted by the French multinational electric utility company EDF in the Vienne river [Petit et al., 2016].
- It is composed of $N = 2 \times 10^4$ simulations.
 - 1 output: water level H
 - 37 inputs depending on: a value of flow upstream, data on the geometry of the bed, and Strickler friction coefficients
- It is possible to identify that H is decreasing along the first 24 input dimensions and increasing along dimension 37.
- Petit et al. [2016] have shown that the additive assumption is realistic here, and that inputs 11, 35 and 37 explain most of the variance.
- We consider (approximated) LHD of size $n = 2d$ for training the cGP.

Numerical experiments: Flood study of the Vienne river



The choice made by MaxMod per iteration. Results are computed over 10 replicates. For the first panel, a bigger and darker square implies a more repeated choice.

Numerical experiments: Flood study of the Vienne river



Q^2 boxplots. Results are computed over 10 replicates. For the first panel, a bigger and darker square implies a more repeated choice.

- We combine the additive and constrained frameworks to propose an additive constrained GP prior and MaxMod algorithm.
- The corresponding mode predictor can be computed and posterior realizations can be sampled, both in a scalable way to high dimension.
- We demonstrate the performance and scalability of the framework with examples with $d \leq 1000$ and in a real-world application with $d = 37$.
- MaxMod identifies the most important input variables, with data size as low as $n = 2d$ in dimension d .
- Open-source R codes are available:

<https://github.com/anfelopera/lineqGPR>

- ★ The extension to block-additivity is being studied by Mathis Deronzier (PhD student at the IMT), for instance for disjoint blocks:

$$Y(x_1, x_2, x_3) = Y(x_1, x_2) + Y(x_3).$$

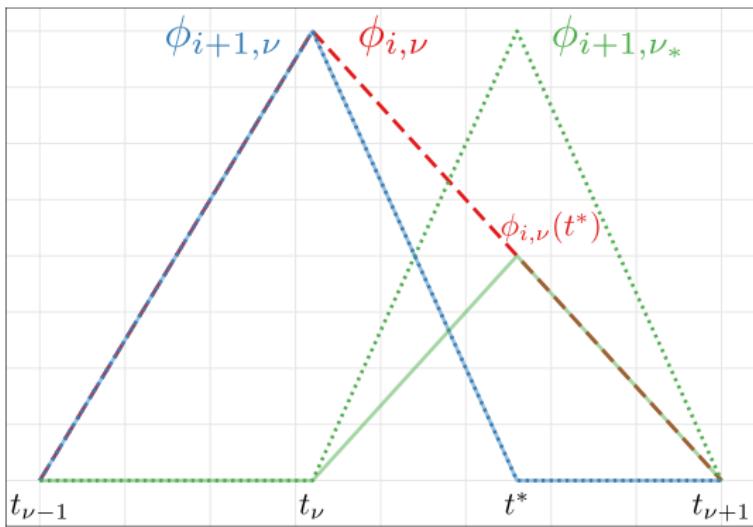
We seek to study:

- Variable selection (structure of the blocks).
 - Theoretical convergence of the (additive and block additive) MaxMod algorithm.
 - Further real-world applications.
- ★ The extension to Student-t processes.

References

- X. Bay, L. Grammont, and H. Maatouk. Generalization of the Kimeldorf-Wahba correspondence for constrained interpolation. *Electronic Journal of Statistics*, 2016.
- N. Durrande, D. Ginsbourger, and O. Roustant. Additive covariance kernels for high-dimensional Gaussian process modeling. *Annales de la Faculté de Sciences de Toulouse*, 21(3):481–499, 2012.
- A. F. López-Lopera, F. Bachoc, N. Durrande, and O. Roustant. Finite-dimensional Gaussian approximation with linear inequality constraints. *SIAM/ASA Journal on Uncertainty Quantification*, 2018.
- A. F. López-Lopera, F. Bachoc, and O. Roustant. High-dimensional additive Gaussian processes under monotonicity constraints. In *NeurIPS*. 2022.
- H. Maatouk and X. Bay. Gaussian process emulators for computer experiments with inequality constraints. *Mathematical Geosciences*, 2017.
- A. Pakman and L. Paninski. Exact Hamiltonian Monte Carlo for truncated multivariate Gaussians. *Journal of Computational and Graphical Statistics*, 2014.
- S. Petit, F. Zaoui, A.-L Popelin, C. Goeury, and N. Goutal. Couplage entre indices à base de dérivées et mode adjoint pour l'analyse de sensibilité globale. Application sur le code Mascaret. HAL e-prints, September 2016.
- C. E. Rasmussen and C. K. I. Williams. *Gaussian Processes for Machine Learning*. The MIT Press, Cambridge, MA, 2005.

Projection of the hat function into the new basis space



Projection of the hat function into the new basis space.