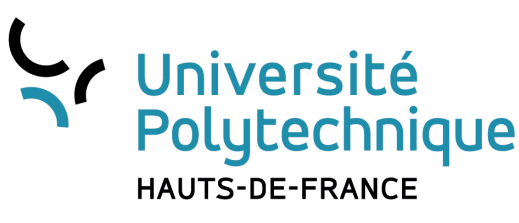


High-dimensional additive Gaussian processes under monotonicity constraints

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

October 4, 2022

ANR GAP project meeting

Joint work...

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.

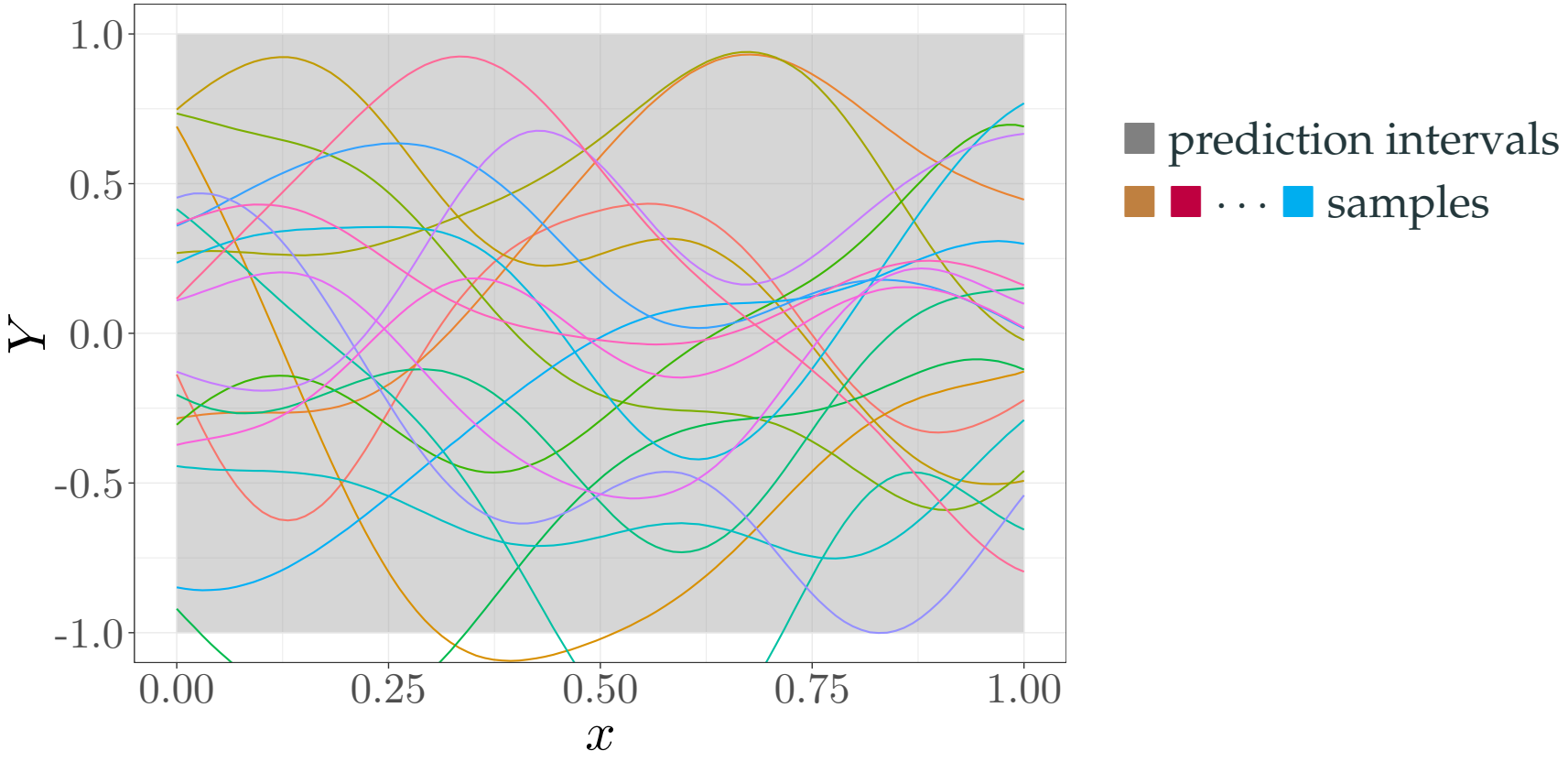
Source: https://arxiv.org/abs/2205.08528

Table of contents

Motivation: constrained GPs

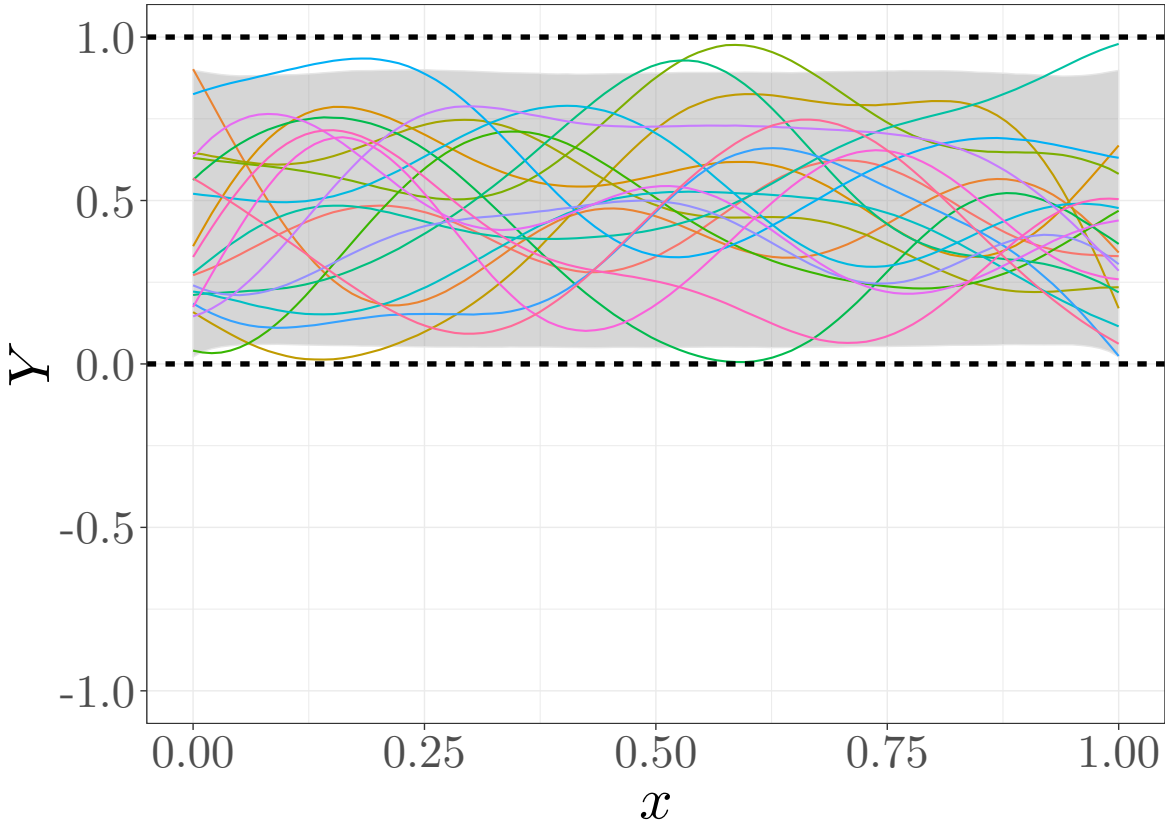
Motivation: constrained GPs

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



Motivation: constrained GPs

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



prediction intervals

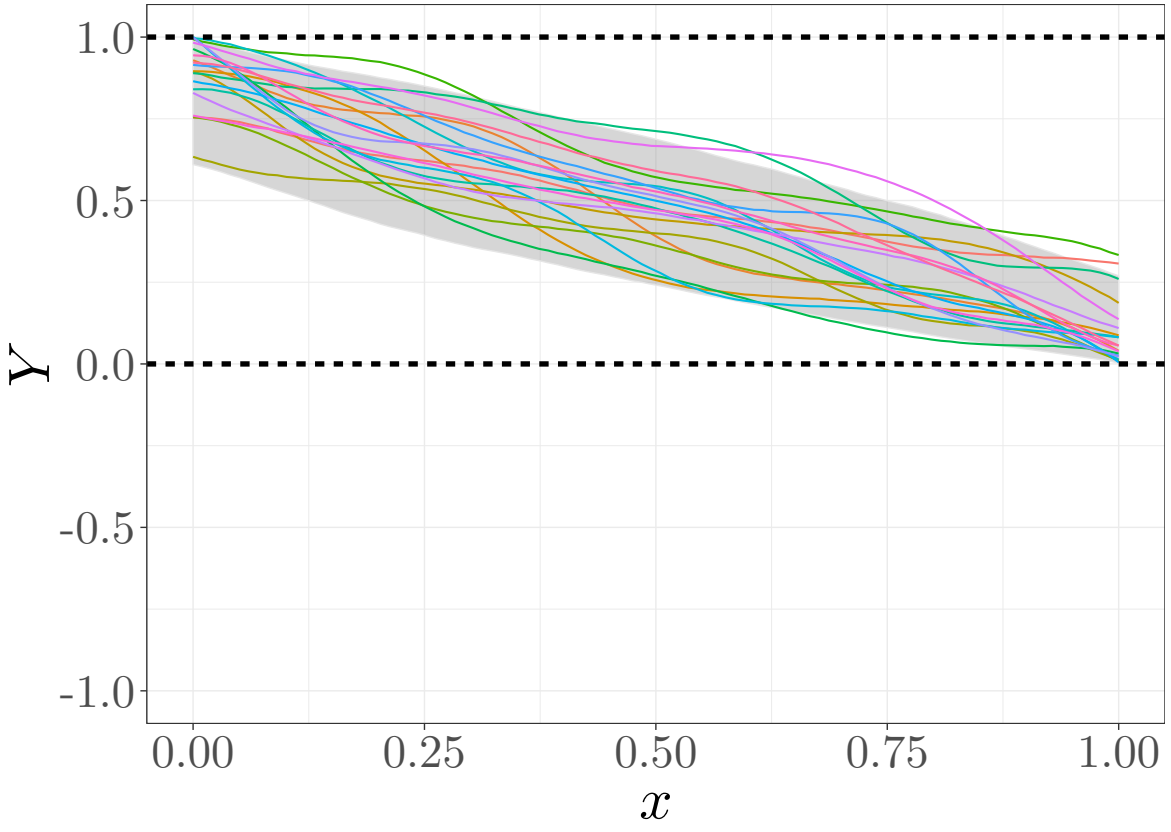
samples

Boundedness condition

$Y \in [0, 1]$

Motivation: constrained GPs

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



prediction intervals

· · · samples

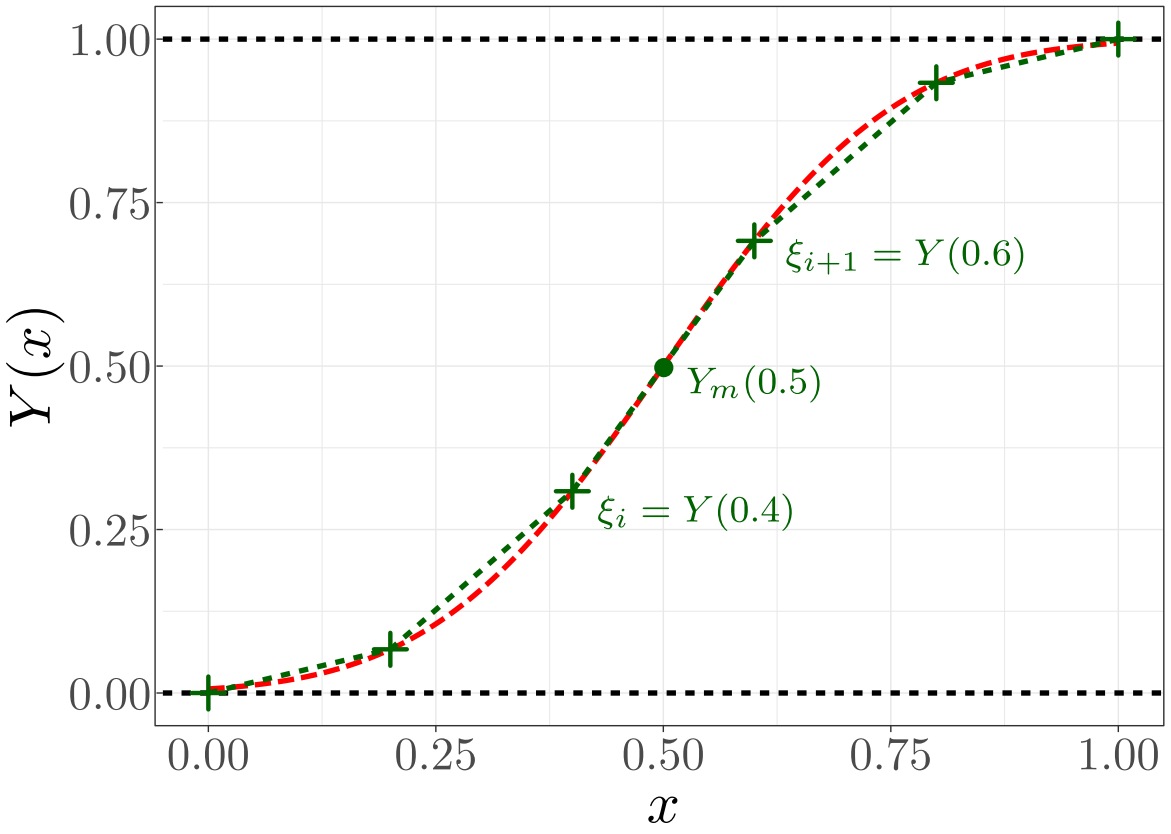
- - Boundedness condition

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

↓ Monotonicity condition

$$Y(x') \leq Y(x), \quad \forall\, x' \geq x.$$

Finite-dimensional approximation of GPs



smooth function

finite approximation

Note that:

· If $\xi_i, \xi_{i+1} \in [0, 1]$, then

$$Y_m(0.5) \in [0, 1].$$

· Or if $\xi_i < \xi_{i+1}$, then

$$\xi_i < Y_m(0.5) < \xi_{i+1}.$$

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

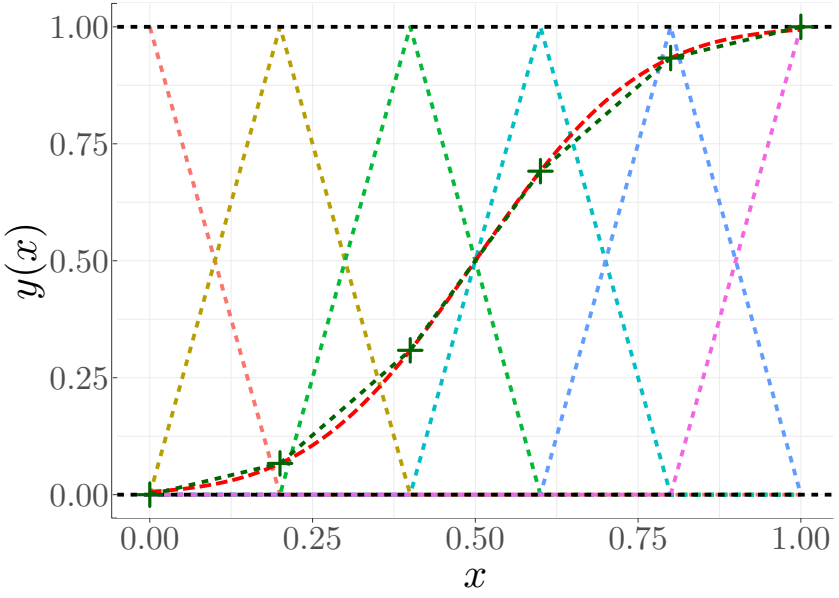
Finite-dimensional approximation of GPs

· Let the (constrained) finite-dimensional GP Y_m be defined as

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

where $x_i \in [0,1]$, $y_i \in \mathbb{R}$ for $i = 1, \ldots, n$, $\xi_j = Y(t_j)$ for $j = 1, \ldots, m$; and

- $\boldsymbol{\xi} = [\xi_1, \ldots, \xi_m]^\top \sim \mathcal{N}(0, \boldsymbol{\Sigma}_\theta)$ with covariance matrix $\boldsymbol{\Sigma}_\theta$,
- $(\boldsymbol{\Lambda}, \boldsymbol{l}, \boldsymbol{u})$ define the inequality conditions,
- $\varepsilon_i \sim \mathcal{N}(0, \tau^2)$ with noise variance τ^2, and
- $\phi_j : [0,1] \mapsto \mathbb{R}$ are (symmetric) hat basis functions:



Finite-dimensional approximation of GPs

· Then, *uncertainty quantification* relies on simulating the **truncated vector $\boldsymbol{\xi}$** [López-Lopera et al., 2018]:

$$\boldsymbol{\Lambda\xi}|\{\boldsymbol{\Phi\xi}+\boldsymbol{\varepsilon}=\boldsymbol{y},\boldsymbol{l}\leq\boldsymbol{\Lambda\xi}\leq\boldsymbol{u}\}\sim\mathcal{TN}(\boldsymbol{\Lambda}\boldsymbol{\mu}_c,\boldsymbol{\Lambda}\boldsymbol{\Sigma}_c\boldsymbol{\Lambda}^\top,\boldsymbol{l},\boldsymbol{u}), \tag{2}$$

with conditional parameters $\boldsymbol{\mu}_c$ and $\boldsymbol{\Sigma}_c$ given by

$$\mathbf{K}=\boldsymbol{\Phi\Sigma\Phi}^\top+\tau^2\boldsymbol{I},\quad \boldsymbol{\mu}_c=\boldsymbol{\Sigma\Phi}^\top\mathbf{K}^{-1}\boldsymbol{y},\quad \boldsymbol{\Sigma}_c=\boldsymbol{\Sigma}-\boldsymbol{\Sigma\Phi}^\top\mathbf{K}^{-1}\boldsymbol{\Phi\Sigma}. \tag{3}$$

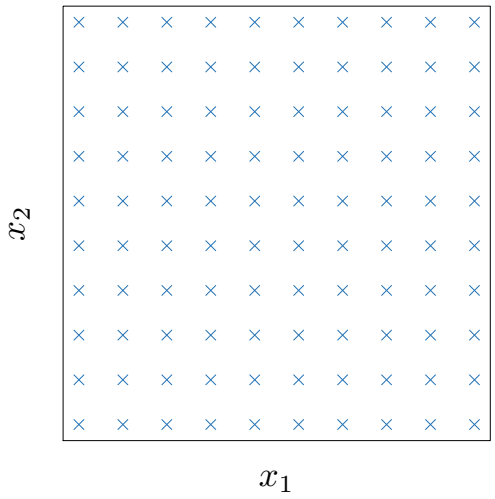
∗ Eq. (2) is computed via *Monte Carlo* (MC) or *Markov Chain MC* (MCMC):

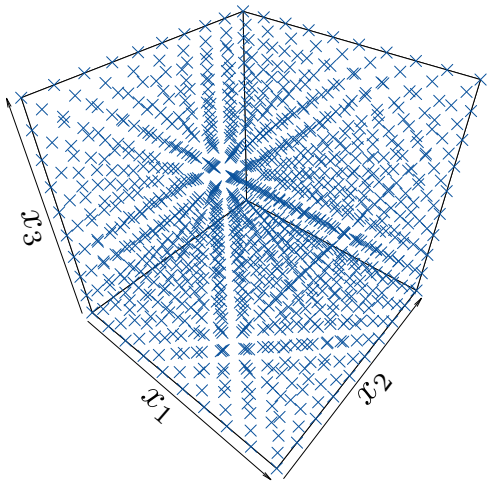
- e.g. *Hamiltonian Monte Carlo* (HMC) [Pakman and Paninski, 2014]

Finite-dimensional approximation of GPs

· **Con:** the cost of Y_m increases as d increases.

$$Y_m(\boldsymbol{x}) = \sum_{j_1,\ldots,j_d=1}^{m_1,\ldots,m_d} \left[\prod_{p=1,\ldots,d} \phi_{j_p}^{(p)}(x_p) \right] \xi_{j_1,\ldots,j_d}, \quad \text{s.t.} \quad \begin{cases} Y_m(x_i) + \varepsilon_i = y_i, \\ \boldsymbol{\xi} \in \mathcal{C}. \end{cases} \tag{4}$$





· This drawback can be mitigated by considering:

- a *"smarter" construction of rectangular grids* of knots
- and/or *further assumptions for complexity simplification*

  $\rightarrow$ e.g. *inactive variables*

The MaxMod algorithm

The maximum a posteriori (mode) function in 1D

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

$$\widehat{\boldsymbol{\xi}} = \underset{\boldsymbol{\xi} \text{ s.t. } \boldsymbol{l}\leq\boldsymbol{\Lambda}\boldsymbol{\xi}\leq\boldsymbol{u}}{\arg\max} \{-[\boldsymbol{\xi}-\boldsymbol{\mu}_c]^\top \boldsymbol{\Sigma}_c{}^{-1}[\boldsymbol{\xi}-\boldsymbol{\mu}_c]\}, \tag{5}$$

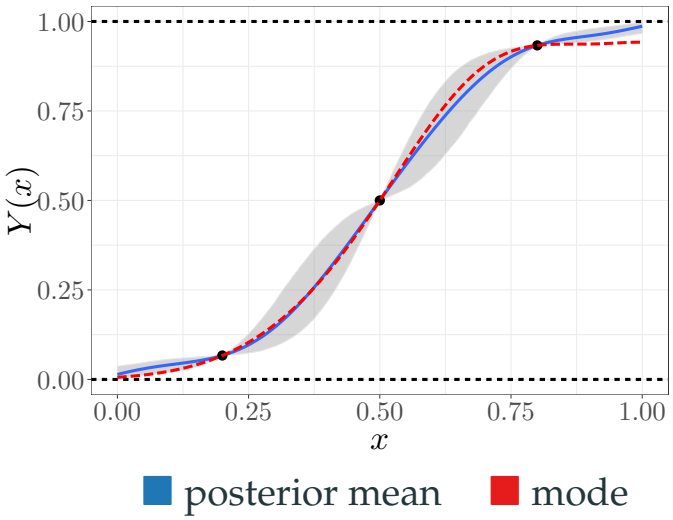
with $\widehat{\boldsymbol{\xi}} = [\widehat{\xi}_1, \ldots, \widehat{\xi}_m]^\top$.

· The *MAP estimate* of Y_m is given by

$$\widehat{Y}_m(x) = \sum_{j=1}^{m} \widehat{\xi}_j \phi_j(x). \tag{6}$$

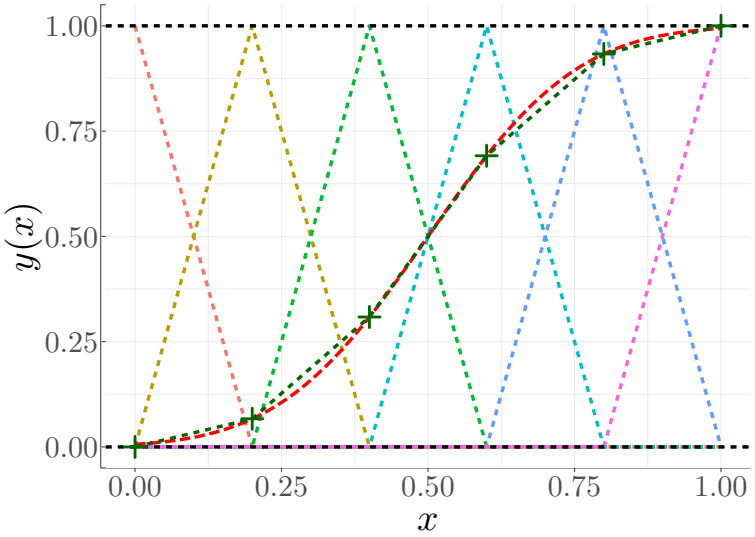
Pro:

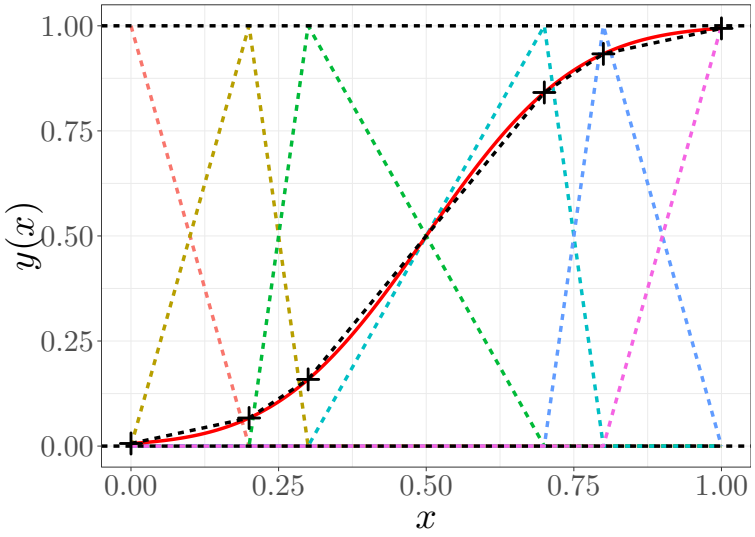
- $\widehat{Y}_m$ can be used as a point estimate
- Easy and fast calculations
- Convergence to the spline solution as $m \to \infty$ [Bay et al., 2016]
- Starting point for MCMC



Asymmetric hat basis functions

· In practice, we modify the construction of the hat basis functions ϕ:





Pros:

- This construction allows the free location of the knots
- Constrained GP model's properties are preserved [see Bachoc et al., 2022, López-Lopera, 2018]

The MaxMod algorithm in 1D

· Let $\widehat{\Upsilon}_S$ be the MAP function with an ordered set of knots:

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

· 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*:

$$I_S(t) = \int_{[0,1]} \left(\widehat{\Upsilon}_{S \cup t}(x) - \widehat{\Upsilon}_S(x)\right)^2 dx. \tag{7}$$

· The integral in (7) has a closed-form expression.

The MaxMod algorithm in 1D

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

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

· 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*:

$$I_S(t) = \int_{[0,1]} \left(\widehat{Y}_{S \cup t}(x) - \widehat{Y}_S(x)\right)^2 dx. \tag{7}$$

· The integral in (7) 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 $t^\star_{\kappa+1} \in [0,1]$ such that

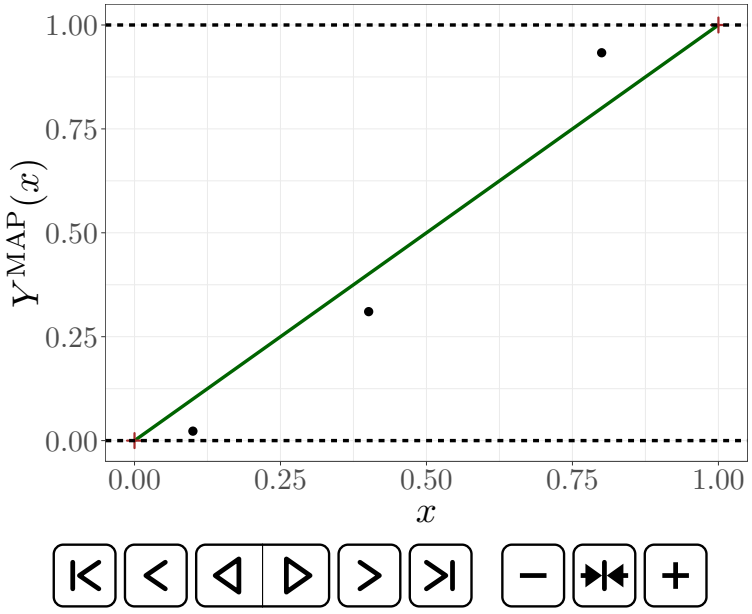
$$I_{S^{(\kappa)}}(t^\star_{\kappa+1}) \geq \sup_{t \in [0,1]} I_{S^{(\kappa)}}(t)$$

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

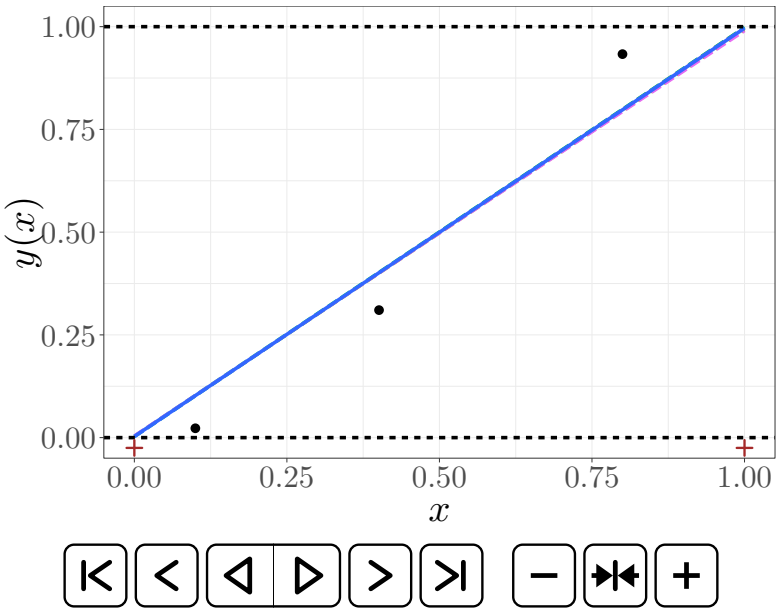
The MaxMod algorithm in 1D

1D example under boundedness and monotonicity constraints

MAP estimate



conditional sample-path



● training points + knots ■ MAP estimate

■ predictive mean ■ 90% confidence intervals

The MaxMod algorithm in higher dimensions

· Let $\widehat{Y}_{\mathcal{J},S}$ be the MAP function with $|\mathcal{J}|$ active variables and ordered sets of knots $S_{\mathcal{J}}$ for $\mathcal{J} \subseteq \{1,\dots,D\}$.

· Then, the criterion to maximise is given by

$$I_{\mathcal{J},S}(i,t) = \begin{cases} \frac{1}{N_{S,\mathcal{J},i}} \displaystyle\int_{[0,1]^d} \left(\widehat{Y}_{\mathcal{J},\, S \cup_i t}(\boldsymbol{x}) - \widehat{Y}_{\mathcal{J},\,S}(\boldsymbol{x})\right)^2 d\boldsymbol{x} & \text{if } i \in \mathcal{J}, \\ \frac{1}{N_{S,\mathcal{J},i}} \displaystyle\int_{[0,1]^{d+1}} \left(\widehat{Y}_{\mathcal{J}\cup\{i\},\, S+i}(\boldsymbol{x}) - \widehat{Y}_{\mathcal{J},\,S}(\boldsymbol{x})\right)^2 d\boldsymbol{x} & \text{if } i \notin \mathcal{J}, \end{cases} \tag{8}$$

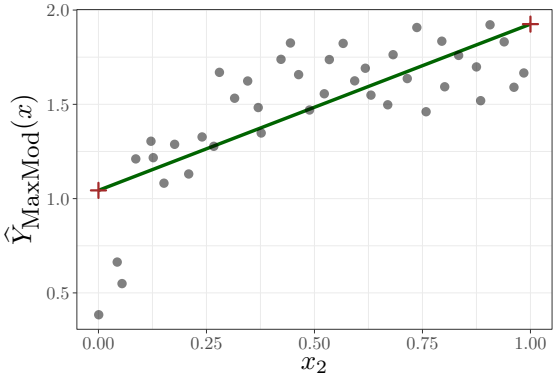
where $N_{S,\mathcal{J},i}$ is the increase of the number of basis functions.

- 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.

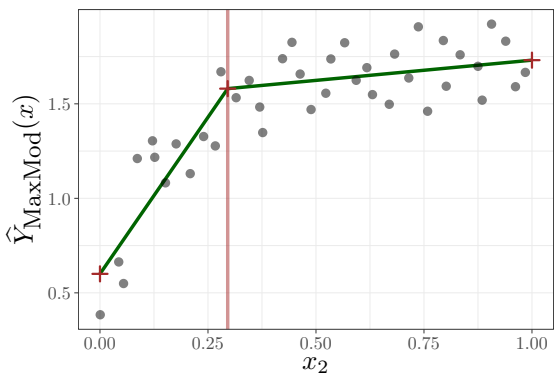
The MaxMod algorithm in higher dimensions

2D example under monotonicity constraints

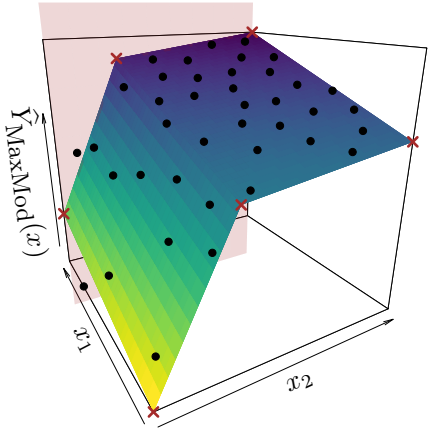
Evolution of the MaxMod algorithm using $f(x) = \frac{1}{2}x_1 + \arctan(10x_2)$



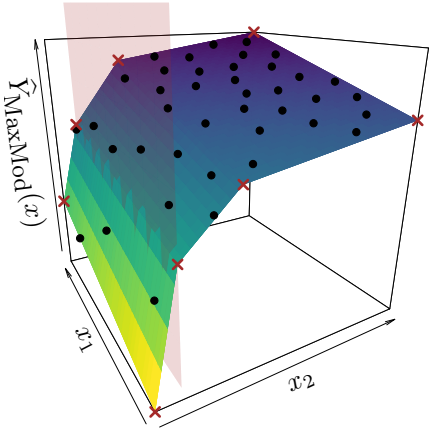
(a) iteration 0



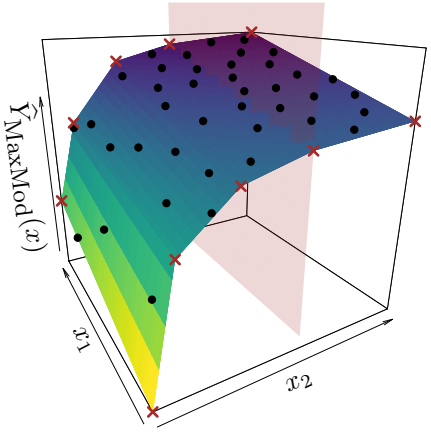
(b) iteration 1



(c) iteration 2



(d) iteration 3



(e) iteration 4

CERAMATHS • training points + knots ■ MAP estimate

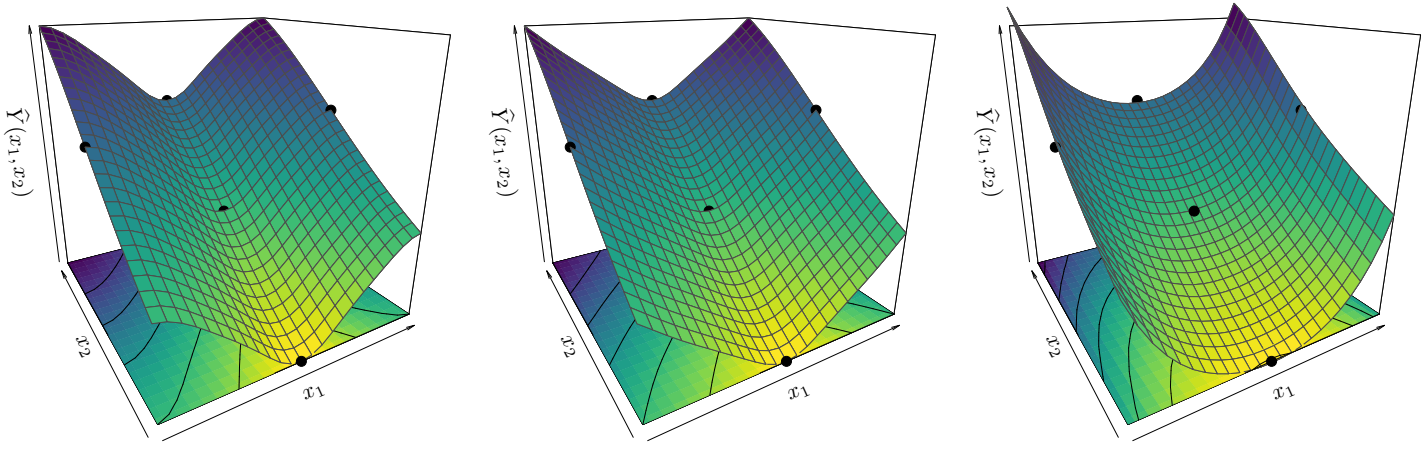
The MaxMod algorithm in higher dimensions

· The constrained GP is tractable depending on $|\mathcal{J}|$ (nb of active variable).

· According to numerical tests, we are limited to problems with $|\mathcal{J}| \leq 5$.

· 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

Additive GPs

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

$$y(\boldsymbol{x}) = y_1(x_1) + \cdots + y_d(x_d). \tag{9}$$

· Then GP priors can be placed over $y_1, \ldots, y_d$ [Durrande et al., 2012]

$$Y_i \sim \mathcal{GP}(0, k_i),$$

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

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

is also a GP and its kernel is given by

$$k(\boldsymbol{x}, \boldsymbol{x}') = k_1(x_1, x_1') + \cdots + k_d(x_d, x_d'). \tag{10}$$

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

$$Y_S(\boldsymbol{x}) = \sum_{i=1}^{d} Y_{i,S_i}(x_i) = \sum_{i=1}^{d}\sum_{j=1}^{m_i} \xi_{i,j}\phi_{i,j}(x_i), \tag{11}$$

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 $m = m_1 + \cdots + m_d$.

Finite-dimensional approximation of additive GPs

· Observe from (11) that, since $\xi_{i,j}$, for $i = 1, \ldots, d$ and $j = 1, \ldots, m_i$, are Gaussian distributed, then Y_{i,S_i} is a GP with kernel given by

$$\widetilde{k}_i(x_i, x_i') = \sum_{j=1}^{m_i} \sum_{\kappa=1}^{m_i} \phi_{i,j}(x_i)\phi_{i,\kappa}(x_i')k_i(t_{(j)}^{(S_i)}, t_{(\kappa)}^{(S_i)}). \tag{12}$$

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

· We let $\boldsymbol{\Sigma}_i = k_i(S_i, S_i)$ be the $m_i \times m_i$ covariance matrix of $\boldsymbol{\xi}_i$.

Finite-dimensional approximation of additive GPs

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

$$Y_{i,S_i} \in \mathcal{E}_i \;\Leftrightarrow\; \boldsymbol{\xi}_i \in \mathcal{C}_i \tag{13}$$

where $\boldsymbol{\xi}_i = [\xi_{i,1}, \cdots, \xi_{i,m_i}]^\top$ and $\mathcal{C}_i = \{\boldsymbol{c} \in \mathbb{R}^{m_i} : \boldsymbol{l}_i \leq \boldsymbol{\Lambda}_i \boldsymbol{c} \leq \boldsymbol{u}_i\}$.

· Examples of constraints are monotonicity and componentwise convexity.

· Given the observations and the constraints, the MAP estimate is given by

$$\widehat{Y}_S(\boldsymbol{x}) = \sum_{i=1}^{d} \sum_{j=1}^{m_i} \widehat{\xi}_{i,j} \phi_{i,j}(x_i). \tag{14}$$

Finite-dimensional approximation of additive GPs

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

$$\widehat{\boldsymbol{\xi}} = \underset{\substack{\boldsymbol{\xi}=(\boldsymbol{\xi}_1^\top,\ldots,\boldsymbol{\xi}_d^\top)^\top \\ \boldsymbol{l}_i \leq \boldsymbol{\Lambda}_i \boldsymbol{\xi}_i \leq \boldsymbol{u}_i, i=1,\ldots,d}}{\operatorname{argmin}} (\boldsymbol{\xi} - \boldsymbol{\mu}_c)^\top \boldsymbol{\Sigma}_c^{-1} (\boldsymbol{\xi} - \boldsymbol{\mu}_c), \tag{15}$$

where $\boldsymbol{\mu}_c = [\boldsymbol{\mu}_{c,1}^\top, \ldots, \boldsymbol{\mu}_{c,d}^\top]^\top$ is the $m \times 1$ vector with block i given by

$$\boldsymbol{\mu}_{c,i} = \boldsymbol{\Sigma}_i \boldsymbol{\Phi}_i^\top \left[\left(\sum_{p=1}^{d} \boldsymbol{\Phi}_p \boldsymbol{\Sigma}_p \boldsymbol{\Phi}_p^\top \right) + \tau^2 \boldsymbol{I}_n \right]^{-1} \boldsymbol{y}_n, \tag{16}$$

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

$$\boldsymbol{\Sigma}_{c,i,j} = \mathbf{1}_{i=j} \boldsymbol{\Sigma}_i - \boldsymbol{\Sigma}_i \boldsymbol{\Phi}_i^\top \left[\left(\sum_{p=1}^{d} \boldsymbol{\Phi}_p \boldsymbol{\Sigma}_p \boldsymbol{\Phi}_p^\top \right) + \tau^2 \boldsymbol{I}_n \right]^{-1} \boldsymbol{\Phi}_j \boldsymbol{\Sigma}_j. \tag{17}$$

Remarks:

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

Additive MaxMod algorithm

· Consider an additive cGP model that uses only a subset $\mathcal{J} \subseteq \{1, \ldots, d\}$ of active variables.

· Its mode function $\widehat{Y}_S$, from $\mathbb{R}^{|\mathcal{J}|}$ to $\mathbb{R}$, by, for $\boldsymbol{x} = (x_i; i \in \mathcal{J})$,

$$\widehat{Y}_S(\boldsymbol{x}) = \sum_{i \in \mathcal{J}} \sum_{j=1}^{m_i} \widehat{\xi}_{i,j} \phi_{i,j}(x_i). \tag{18}$$

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

$$I_{S,i^\star} = \int_{[0,1]^{|\mathcal{J}|+1}} \left(\widehat{Y}_S(\boldsymbol{x}) - \widehat{Y}_{S,i^\star}(\boldsymbol{x}) \right)^2 d\boldsymbol{x} \quad \text{for } i^\star \notin \mathcal{J}, \tag{19}$$

$$I_{S,i^\star,t} = \int_{[0,1]^{|\mathcal{J}|}} \left(\widehat{Y}_S(\boldsymbol{x}) - \widehat{Y}_{S,i^\star,t}(\boldsymbol{x}) \right)^2 d\boldsymbol{x} \quad \text{for } i^\star \in \mathcal{J}. \tag{20}$$

· Both (19) and (20) have analytic expression assuming $x_i \sim \mathsf{Uniform}(0, 1)$ for $i = 1, \ldots, d$ (see López-Lopera et al. [2022]), where the computational cost is linear with respect to $m = \sum_{i \in \mathcal{J}} m_i$.

INSTITUT DE MATHÉMATIQUES DE TOULOUSE

CERAMATHS

Numerical experiments

Numerical experiments: Monotonicity in hundreds of dimensions

· We consider the target function:

$$y(\boldsymbol{x}) = \sum_{i=1}^{d} \arctan\left(5\left[1 - \frac{i}{d+1}\right]x_i\right). \tag{21}$$

with $\boldsymbol{x} \in [0,1]^d$. 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]		Q^2 [%]		
		cGP mode	cGP mean	GP mean	cGP mode	cGP mean
10	50	0.1 ± 0.1	0.1 ± 0.1	82.3 ± 6.2	83.8 ± 4.2	$\mathbf{88.1 \pm 1.7}$
100	500	0.4 ± 0.1	5.2 ± 0.5	89.8 ± 1.6	90.7 ± 1.4	$\mathbf{91.5 \pm 1.3}$
250	1250	4.2 ± 0.7	132.3 ± 26.3	91.7 ± 0.8	92.9 ± 0.6	$\mathbf{93.4 \pm 0.6}$
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: Dimension reduction illustration

· We test the capability of MaxMod to account for dimension reduction considering the function in (21).

· In addition to $(x_1, \ldots, x_d)$, we include $D - d$ virtual variables, indexed as $(x_{d+1}, \ldots, x_D)$, 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**

Numerical experiments: Flood study of the Vienne river

· 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

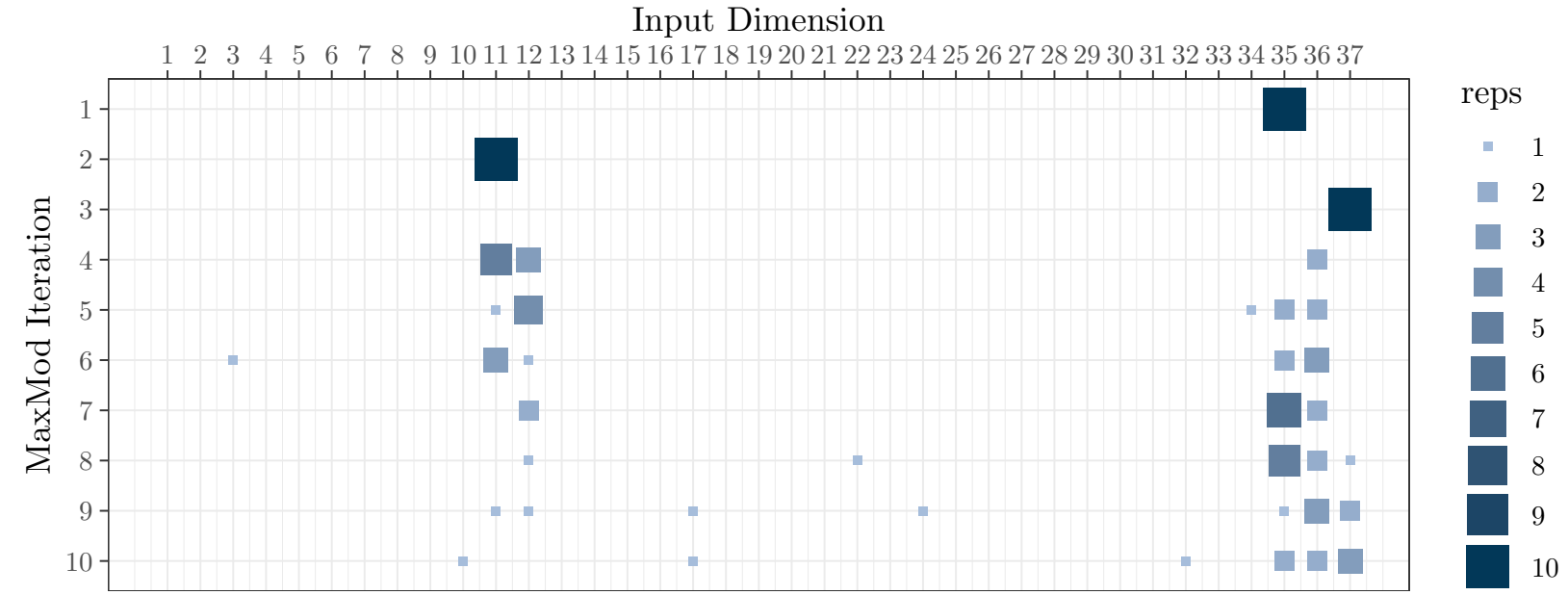
  · 37 inputs depending on: a value of flow upstream, data on the geometry of the bed, and Strickler friction coefficients

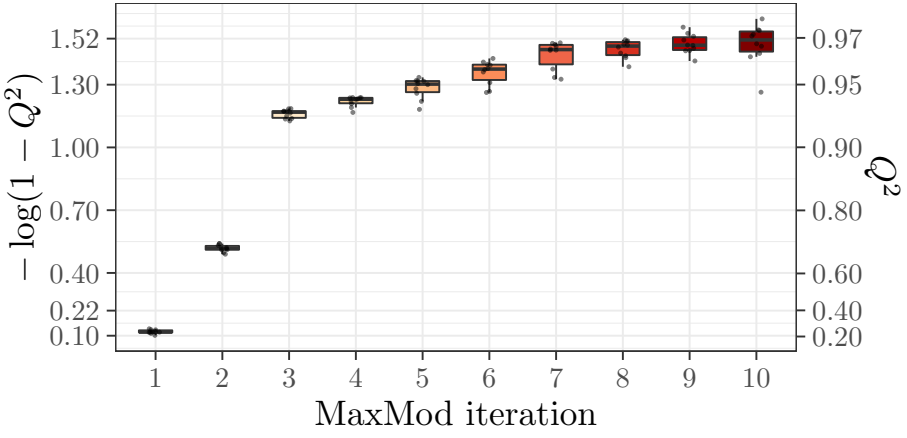
· It is possible to identify that H is decreasing along the first 24 input dimensions and increasing along dimension 37.

· 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





(top) the choice made by MaxMod per iteration and (left) Q^2 boxplots. Results are computed over 10 replicates. For the first panel, a bigger and darker square implies a more repeated choice.

Conclusions

· 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.

· We provide open-source R codes for our full framework.

References

F. Bachoc, A. F. López-Lopera, and O. Roustant. Sequential construction and dimension reduction of Gaussian processes under inequality constraints. *SIAM Journal on Mathematics of Data Science*, 4(2):772–800, 2022.

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. *Gaussian Process Modelling under Inequality Constraints*. PhD thesis, Mines Saint-Étienne, 2018.

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.

INSTITUT DE MATHÉMATIQUES DE TOULOUSE

CERAMATHS