Kriging models under Inequality Constraints

Andrés F. López-Lopera[1], François Bachoc[2], Nicolas Durrande[1,3], and Olivier Roustant[1]

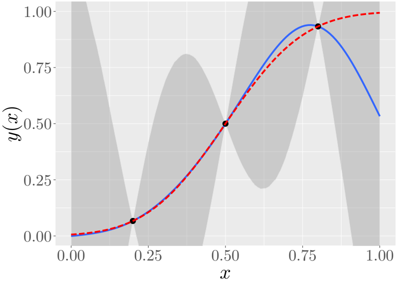
[1]École des Mines de Saint-Étienne (EMSE), France.
[2]Institut de Mathématiques de Toulouse (IMT), France.
[3]PROWLER.io, Cambridge, UK.

This work is funded by the chair of applied mathematics OQUAIDO.

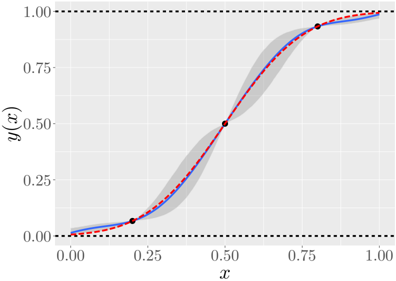
Nov 22, 2018

Gaussian process models: motivation

Target function: bounded and monotonic.



Unconstrained GP.



Constrained GP.

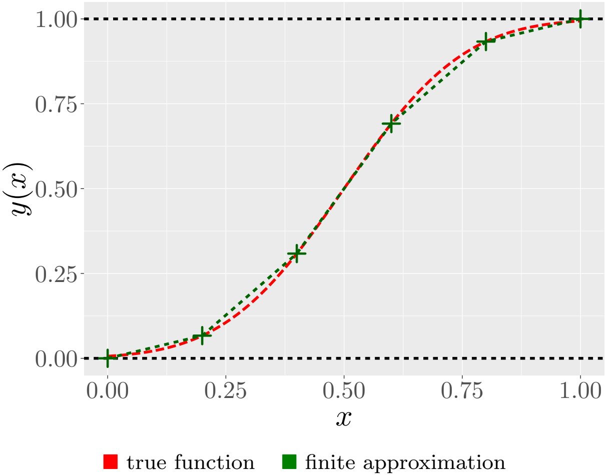
- true function
- predictive mean
- training points
- confidence intervals

Table of Contents

Table of Contents

Table of Contents

Finite representation: also bounded and monotonic.

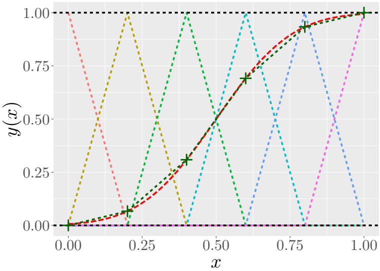


$\Rightarrow$ Imposing the inequality constraints on the knots is enough.

Let the finite-dimensional GP approximation be defined as

$$Y_m(x) = \sum_{j=1}^{m} \xi_j \phi_j(x), \text{ s.t. } \begin{cases} Y_m(x_i) = y_i & \text{(interpolation conditions)}, \\ \boldsymbol{\xi} \in \mathcal{C} & \text{(inequality conditions)}, \end{cases} \quad (1)$$

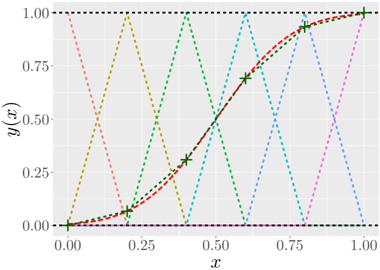
where $\boldsymbol{\xi} = \left[\xi_1, \cdots, \xi_m\right]^\top \sim \mathcal{N}(\mathbf{0}, \mathbf{\Gamma})$, with covariance matrix $\mathbf{\Gamma}$ and $\phi_j : [0,1] \to \mathbb{R}$ are hat functions (see López-Lopera et al. (2018)):



Let the finite-dimensional GP approximation be defined as

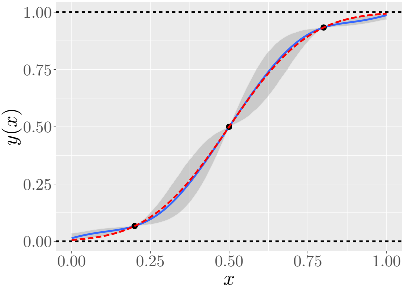
$$Y_m(x) = \sum_{j=1}^{m} \xi_j \phi_j(x), \text{ s.t. } \begin{cases} Y_m(x_i) = y_i & \text{(interpolation conditions)}, \\ \boldsymbol{\xi} \in \mathcal{C} & \text{(inequality conditions)}, \end{cases} \quad (1)$$

where $\boldsymbol{\xi} = [\xi_1, \cdots, \xi_m]^\top \sim \mathcal{N}(\mathbf{0}, \mathbf{\Gamma})$, with covariance matrix $\mathbf{\Gamma}$ and $\phi_j : [0, 1] \to \mathbb{R}$ are hat functions (see López-Lopera et al. (2018)):

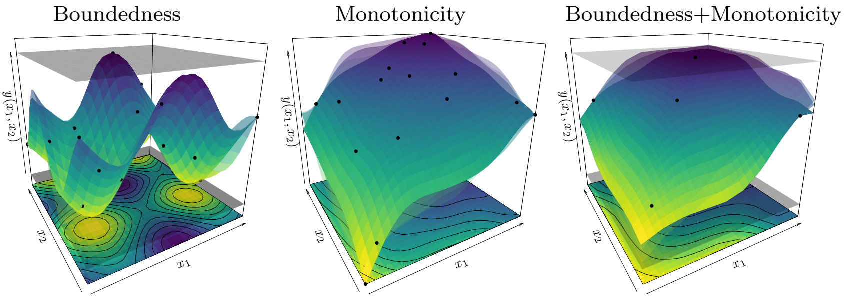


✦ Since linearity preserves Gaussian distributions, quantifying uncertainty on Y_m relies on simulating a truncated Gaussian vector $\boldsymbol{\xi} \in \mathcal{C}$ (e.g. MCMC).

1D example under boundedness and monotonicity constraints



$$\underbrace{\begin{bmatrix}0\\0\\0\\\vdots\\0\\0\\0\end{bmatrix}}_{\boldsymbol{l}} \leq \underbrace{\begin{bmatrix}1 & 0 & 0 & \cdots & 0 & 0\\ -1 & 1 & 0 & \cdots & 0 & 0\\ 0 & -1 & 1 & \cdots & 0 & 0\\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots\\ 0 & 0 & 0 & \cdots & 1 & 0\\ 0 & 0 & 0 & \cdots & -1 & 1\\ 0 & 0 & 0 & \cdots & 0 & 1\end{bmatrix}}_{\boldsymbol{\Lambda}} \underbrace{\begin{bmatrix}\xi_1\\ \xi_2\\ \xi_3\\ \vdots\\ \xi_{m-1}\\ \xi_m\end{bmatrix}}_{\boldsymbol{\xi}} \leq \underbrace{\begin{bmatrix}1\\ \infty\\ \infty\\ \vdots\\ \infty\\ \infty\\ 1\end{bmatrix}}_{\boldsymbol{u}}$$



Examples of 2D Gaussian models with different types of constraints.

Table of Contents

Partial contributions (papers, conferences, ...)

Papers

- ✦ A.F. López-Lopera, F. Bachoc, N. Durrande, and O. Roustant (2018). *Finite-dimensional Gaussian approximation with linear inequality constraints.* SIAM/ASA Journal on Uncertainty Quantification, 6(3): 1224–1255.
- F. Bachoc, A. Lagnoux, and A.F. López-Lopera (+2018). *Maximum likelihood estimation for Gaussian processes under inequality constraints* (submitted).
- A.F. López-Lopera, S. John, N. Durrande (+2018). *Gaussian process modulated Cox processes under linear inequality constraints* (in revision).

International Conferences

- A.F. López-Lopera, F. Bachoc, N. Durrande, and O. Roustant, *Finite-dimensional Gaussian approximation with linear inequality constraints*, SIAM-UQ 18, Garden Grove, California, USA, April 16-19, 2018.
- ——, *Efficiently approximating Gaussian process emulators with inequality constraints using MC/MCMC*, MCQMC 2018, Rennes, France, July 1-6, 2018.

R Packages

- A.F. López-Lopera. *LineqGPR: Gaussian process regression models with linear inequality constraints*, 2018. Freely available on CRAN: `https://cran.r-project.org/web/packages/lineqGPR/index.html`.

Proposed future works

At (Lyon, 2018), we proposed as future works:

- To scale our framework for higher dimensions (e.g. $d > 2$).
  - Additive functions
  - Triangular designs
  - Sparse grids

Proposed future works

At (Lyon, 2018), we proposed as future works:

- To scale our framework for higher dimensions (e.g. $d > 2$).
  - Additive functions ✓
  - Triangular designs ✓2D (working on $d > 2$) – see (Lyon, 2018)
  - Sparse grids ✗

  $\Rightarrow$ satisfying inequalities everywhere requires rectangular grids!
    - Alternative: sequential construction of rectangular grids.

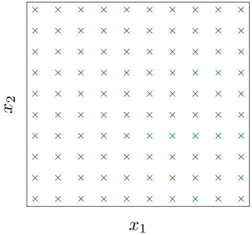
Table of Contents

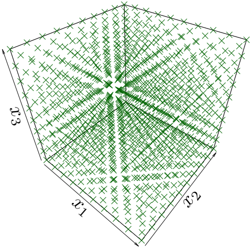
Course of dimensionality of the finite approximation

The finite-dimensional Gaussian approximation could be extended (in theory) to higher dimensions by tensorisation:

$$Y_{m_1,\cdots,m_d}(x_1,\cdots,x_d) := \sum_{j_1=1}^{m_1}\cdots\sum_{j_d=1}^{m_d}\xi_{j_1,\cdots,j_d}\prod_{k=1}^{d}\phi_{j_k}^{k}(x_k), \tag{2}$$

but...





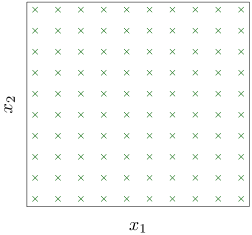
✗ Design of the knots by tensorisation (intractable in practice!!)

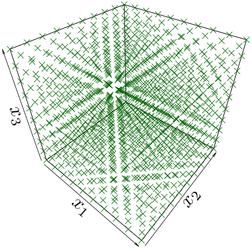
Course of dimensionality of the finite approximation

The finite-dimensional Gaussian approximation could be extended (in theory) to higher dimensions by tensorisation:

$$Y_{m_1,\cdots,m_d}(x_1,\cdots,x_d) := \sum_{j_1=1}^{m_1} \cdots \sum_{j_d=1}^{m_d} \xi_{j_1,\cdots,j_d} \prod_{k=1}^{d} \phi_{j_k}^{k}(x_k), \tag{2}$$

but...





✗ Design of the knots by tensorisation (intractable in practice!!)

⇒ Additional assumptions can be made for complexity simplification.

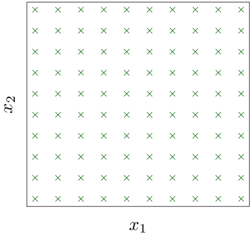
- e.g. inactive dimensions, additive functions.

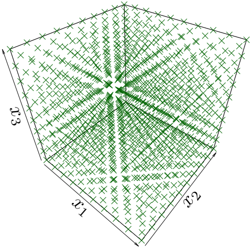
Course of dimensionality of the finite approximation

The finite-dimensional Gaussian approximation could be extended (in theory) to higher dimensions by tensorisation:

$$Y_{m_1,\cdots,m_d}(x_1,\cdots,x_d) := \sum_{j_1=1}^{m_1} \cdots \sum_{j_d=1}^{m_d} \xi_{j_1,\cdots,j_d} \prod_{k=1}^{d} \phi_{j_k}^{k}(x_k), \qquad (2)$$

but...





✗ Design of the knots by tensorisation (intractable in practice!!)

⇒ Additional assumptions can be made for complexity simplification.

- e.g. inactive dimensions, additive functions.

⇒ Other designs from finite elements methods can be explored.

- e.g. triangular designs, sparse grids.

Table of Contents

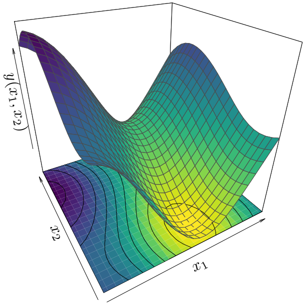
2D Illustration of Additive GPs

Let Y be a first-order additive GP on $\mathbb{R}$ with covariance function k,

$$Y(x_1, x_2) = Y_1(x_1) + Y_2(x_2),$$

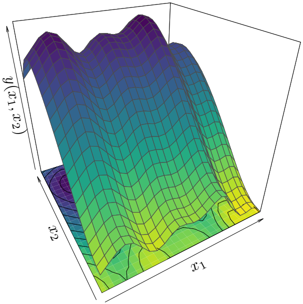
where the Y_κ's are centred GPs on $\mathbb{R}$ with covariances k_κ's for $\kappa = 1, 2$.

Simulation example 1



$(\sigma_1^2, \theta_1) = (\sigma_2^2, \theta_2) = (1, 0.3)$

Simulation example 2



$(\sigma_1^2, \theta_1) = (0.1, 0.1)$, $(\sigma_2^2, \theta_2) = (1, 0.3)$

Additive GPs

- Let Y be a first-order additive GP on $\mathbb{R}$ with covariance function k,

$$Y(x_1, \cdots, x_d) = \sum_{\kappa=1}^{d} Y_\kappa(x_\kappa), \tag{3}$$

where the Y_κ's are centred GPs on $\mathbb{R}$ with covariances k_κ's.

- Assuming that $Y_1, \cdots, Y_d$ are independent, then the covariance k of the additive GP Y is given by

$$k(x_1, \cdots, x_d; x_1{}', \cdots, x_d{}') = \sum_{\kappa=1}^{d} k_\kappa(x_\kappa, x_\kappa{}'). \tag{4}$$

- Finally, Kriging models can be established (see, e.g., Durrande et al., 2012; Duvenaud et al., 2011).

Finite-dimensional representation of additive GPs

Pros.

- Kriging in high dimension is reduced into a series of (highly parallelizable) 1D Kriging problems.

Cons.

- First-order additive GPs are very restrictive models in practice.

  - However, they can be extended into more flexible models, e.g. additivity per blocks,

$$Y(x_1, x_2, x_3, x_4) = Y_1(x_1) + Y_2(x_2) + Y_{3,4}(x_3, x_4),$$

where $Y_1, Y_2, Y_{3,4}$ are independent GPs on $\mathbb{R}$.

Finite-dimensional representation of additive GPs

- Now, we assume that the first-order additive GP Y exhibits certain inequality constraints respect to each input dimension $\kappa = 1, \cdots, d$.
- Let Y_m be the finite representation of the additive GP Y given by

$$Y_m(x_1, \cdots, x_d) = \sum_{\kappa=1}^{d} Y_\kappa^{m_\kappa}(x_\kappa), \text{ s.t. } \begin{cases} Y_m(x_1^i, \cdots, x_d^i) = y_i, \\ \boldsymbol{\xi}^\kappa \in \mathcal{C}_\kappa, \end{cases} \tag{5}$$

where $(x_1^i, \cdots, x_d^i) \in [0, 1]^d$ and $y_i \in \mathbb{R}$ for $i = 1, \cdots, n$.

Finite-dimensional representation of additive GPs

- Now, we assume that the first-order additive GP Y exhibits certain inequality constraints respect to each input dimension $\kappa = 1, \cdots, d$.
- Let Y_m be the finite representation of the additive GP Y given by

$$Y_m(x_1, \cdots, x_d) = \sum_{\kappa=1}^{d} Y_\kappa^{m_\kappa}(x_\kappa), \text{ s.t. } \begin{cases} Y_m(x_1^i, \cdots, x_d^i) = y_i, \\ \boldsymbol{\xi}^\kappa \in \mathcal{C}_\kappa, \end{cases} \tag{5}$$

where $(x_1^i, \cdots, x_d^i) \in [0, 1]^d$ and $y_i \in \mathbb{R}$ for $i = 1, \cdots, n$.

Some remarks:

- Imposing inequality constraints in high dimension can be reduced into a series of (highly parallelizable) 1D constrained problems.
- A noise variance is commonly required to "relax" the interpolation constraints, $Y_m(x_1^i, \cdots, x_d^i) = y_i$, for $i = 1, \cdots n$.

Finite-dimensional representation of additive GPs

Pros.

- Since only 1D simulations are required, MC/MCMC can be performed efficiently.
- Previous implementations in `lineqGPR` can be used into the routine of the constrained additive GPs.

Cons.

- Some constraints cannot be imposed, e.g. boundedness, due to

$$Y_\kappa^{m_\kappa} \in \mathcal{E} \ \forall \ \kappa = 1, \cdots, d \quad \not\Rightarrow \quad Y_m \in \mathcal{E}.$$

Example

$\Rightarrow$ Let $Y_m(x_1, x_2) = Y_1^{m_1}(x_1) + Y_2^{m_1}(x_2)$ with $Y_\kappa^{m_\kappa} \in [a, b]$. Assume the case $Y_1(x_1) = Y_2(x_2) = b$. Then,

$$Y_m(x_1, x_2) = 2b > b.$$

Hence, $Y \notin [a, b]$.

Additive GPs with monotonicity constraints

Additive GPs with monotonicity constraints

- Let Y_m be the finite representation of the additive GP Y.
- Consider $Y_\kappa^{m_\kappa} \in \mathcal{E}_\uparrow$ for $\kappa = 1, \cdots, d$, where $\mathcal{E}_\uparrow$ is the convex set of non-decreasing functions.
- Then,

$$Y_m \in \mathcal{E}_\uparrow \quad \Leftrightarrow \quad Y_\kappa^{m_\kappa} \in \mathcal{E}_\uparrow \ \forall \ \kappa = 1, \cdots, d.$$

Proof.

$\Rightarrow$ For simplicity, we consider the case $d = 2$.

- Consider $Y_m \in \mathcal{E}_\uparrow$, and let $a \in \mathbb{R}$. Then,

$$Y_m(x_1, a) = Y_1^{m_1}(x_1) + Y_2^{m_2}(a).$$

  Since $Y_2^{m_2}(a)$ is constant, we have

$$Y_m \in \mathcal{E}_\uparrow \Rightarrow Y_1^{m_1} \in \mathcal{E}_\uparrow.$$

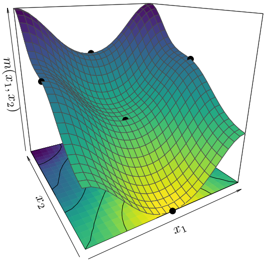
- Similarly, for $Y_m(a, x_2)$, we have $Y_m \in \mathcal{E}_\uparrow \Rightarrow Y_2^{m_2} \in \mathcal{E}_\uparrow$.

$\Leftarrow$ Since $Y_\kappa^{m_\kappa} \in \mathcal{E}_\uparrow$, for $\kappa = 1, \cdots, d$, then $Y_m \in \mathcal{E}_\uparrow$.

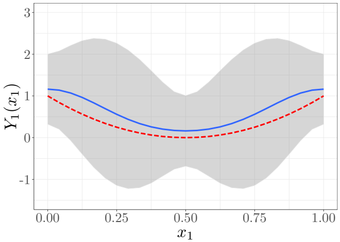
Target function: $Y(x_1, x_2) = \underbrace{4(x_1 - 0.5)^2}_{Y_1(x_1)} + \underbrace{2x_2}_{Y_2(x_2)}$

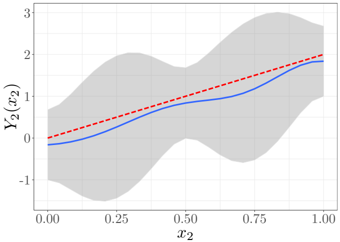
Predictive mean without constraints

$$m(x_1, x_2) = m_1(x_1) + m_2(x_2)$$



$$(\sigma_1^2, \theta_1) = (\sigma_2^2, \theta_2) = (1, 0.2)$$



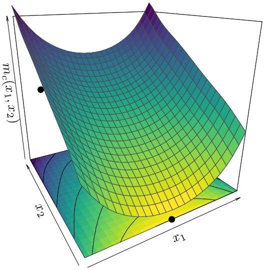


Numerical illustration in 2D

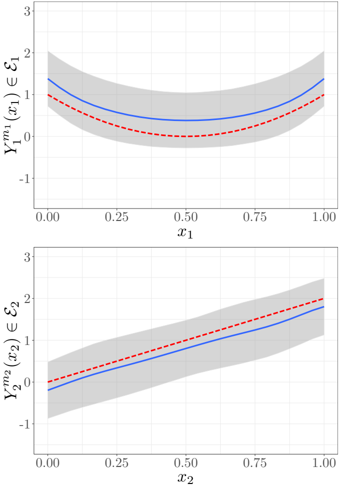
Target function: $Y(x_1, x_2) = \underbrace{4(x_1 - 0.5)^2}_{Y_1(x_1)} + \underbrace{2x_2}_{Y_2(x_2)}$

Predictive mean with constraints

$$m_c(x_1, x_2) = m_{c,1}(x_1) + m_{c,2}(x_2)$$

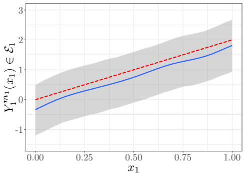


$$(\sigma_1^2, \theta_1) = (\sigma_2^2, \theta_2) = (1, 0.2)$$

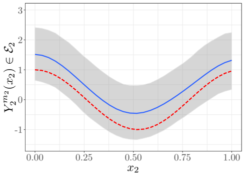


Target function:

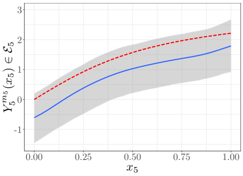
$$Y(x_1,\cdots,x_5) = \underbrace{2x_1}_{Y_1(x_1)} + \underbrace{\cos(6x_2)}_{Y_2(x_2)} + \underbrace{2x_3^2}_{Y_3(x_3)} + \underbrace{4(x_4-0.5)^2}_{Y_4(x_4)} + \underbrace{2\arctan(2x_5)}_{Y_5(x_5)}$$



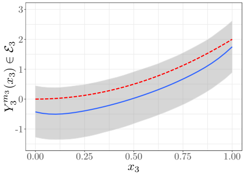
(a) Monotonicity



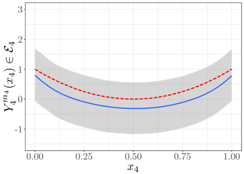
(b) No constraints



(e) Monotonicity



(c) Convexity

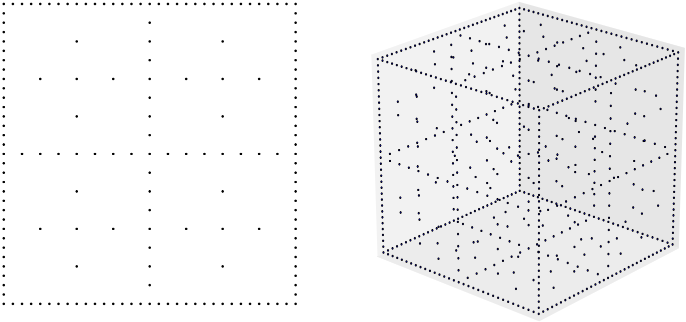


(d) Convexity

- We used 25 points from a maximin LHS.
- For $\kappa=1,\cdots,5$, we fixed $(\sigma_\kappa^2, \theta_\kappa) = (1, 0.2)$.

Table of Contents

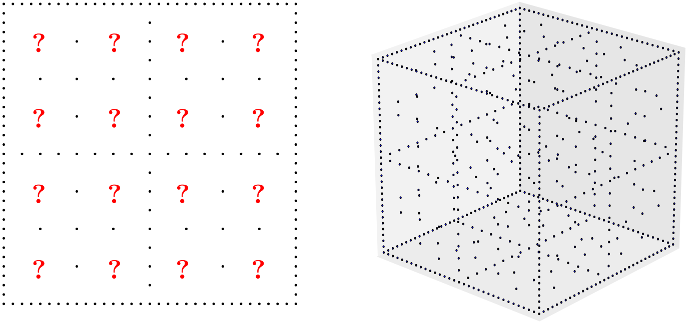
Sparse grids? (bad news)



2D and 3D examples of sparse grids by (Garcke, 2013).

⇒ Sparse grids have been widely studied in finite element methods.

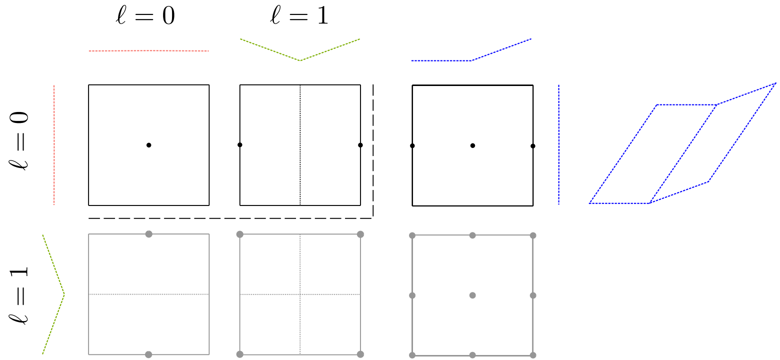
Sparse grids? (bad news)



2D and 3D examples of sparse grids by (Garcke, 2013).

⇒ Sparse grids have been widely studied in finite element methods.

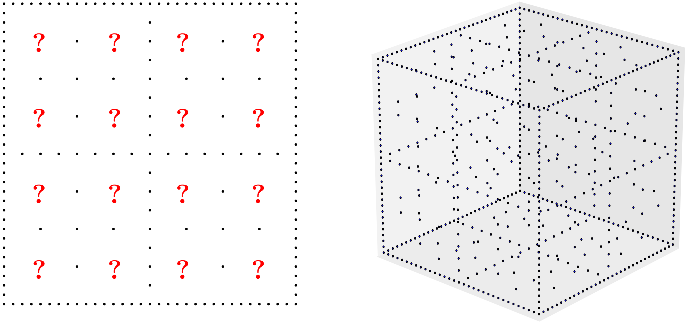
⇒ Inequalities are not necessarily satisfied everywhere :(
- rectangular designs are required.



$\Rightarrow$ Increasing constraints can be imposed w.r.t. the 1st dimension.

$\Rightarrow$ But, they cannot be imposed for the 2nd dimension!!

Sparse grids? (bad news)



2D and 3D examples of sparse grids by (Garcke, 2013).

$\Rightarrow$ Sparse grids have been widely studied in finite element methods.

$\Rightarrow$ Inequalities are not necessarily satisfied everywhere :(
- rectangular designs are required

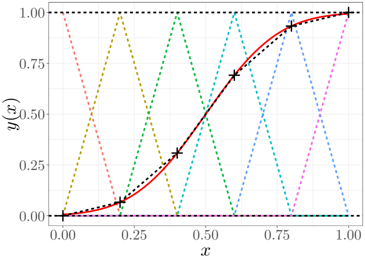
$\Rightarrow$ However, a similar sequential construction can be exploited :)

Table of Contents

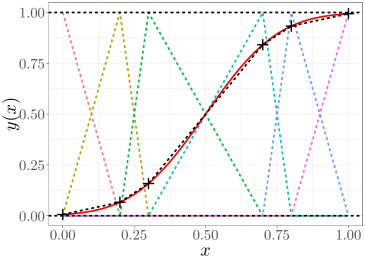
Construction of asymmetric hat basis functions

Consider the knots $t_1 < \cdots < t_m$. For $x \in [t_{j-1},\ t_{j+1}]$, the (asymmetric) hat basis function $\phi_j(x)$ is given by

$$\phi_j(x) := \begin{cases} \frac{x - t_{j-1}}{t_j - t_{j-1}} & \text{if } t_{j-1} \le x < t_j, \\ \frac{t_{j+1} - x}{t_{j+1} - t_j} & \text{if } t_j \le x < t_{j+1}, \\ 0 & \text{otherwise.} \end{cases} \tag{6}$$



Standard hat basis functions.



Asymmetric hat basis functions.

Construction of asymmetric hat basis functions

Pros.

- Since we continue using hat functions, the properties of the finite approximation of GPs are preserved (e.g. piecewise linearity).
- The asymmetric hat functions allow us to refine the grid in places requiring better resolution (e.g. regions with high variability).
- The extension for $d > 1$ are obtained by tensorisation.
- Since rectangular designs are obtained for $d > 1$, then

$$Y_{m_1,\cdots,m_d} \in \mathcal{E} \quad \Leftrightarrow \quad \boldsymbol{\xi} \in \mathcal{C},$$

where $\mathcal{E}$ is a convex set of functions defined by some ineq's.

Construction of asymmetric hat basis functions

Pros.

- Since we continue using hat functions, the properties of the finite approximation of GPs are preserved (e.g. piecewise linearity).
- The asymmetric hat functions allow us to refine the grid in places requiring better resolution (e.g. regions with high variability).
- The extension for $d > 1$ are obtained by tensorisation.
- Since rectangular designs are obtained for $d > 1$, then

$$Y_{m_1,\cdots,m_d} \in \mathcal{E} \quad \Leftrightarrow \quad \boldsymbol{\xi} \in \mathcal{C},$$

  where $\mathcal{E}$ is a convex set of functions defined by some ineq's.

$\Rightarrow$ An evolution criterion is required to add knots in "optimal" places.

Evolution criterion for adding new knots in 1D

- Consider the finite approximation with m_i initial knots given by

$$Y_{m_i}(x) = \sum_{j=1}^{m_i} \xi_j \phi_{i,j}(x), \tag{7}$$

where $\boldsymbol{\xi}$ is a Gaussian vector with covariance $\boldsymbol{\Gamma}$.

- Let $Y_{m_i}^{\text{MAP}}$ be the MAP solution obtained by maximising the pdf of (7) conditioned to both interpolation and inequality constraints,

$$Y_{m_i}^{\text{MAP}}(x) = \sum_{j=1}^{m_i} \mu_{i,j}^M \phi_{i,j}(x), \tag{8}$$

with $\boldsymbol{\mu}_i^M = [\mu_{i,1}^M, \cdots, \mu_{i,m_i}^M]^\top$,

$$\boldsymbol{\mu}_i^M = \min\{\boldsymbol{\xi}_i^\top \boldsymbol{\Gamma}_i^{-1} \boldsymbol{\xi}_i | \ \boldsymbol{\Phi}_i \boldsymbol{\xi}_i = \mathbf{y}, \boldsymbol{\xi}_i \in \mathcal{C}_i\}.$$

Evolution criterion for adding new knots in 1D

- Consider adding a knot t_* to the initial representation (7),

$$Y_{m_{i+1}}(x) = \sum_{j=1}^{m_i} \xi_j \phi_{i+1,j}(x) + \xi_* \phi_{i+1,*}(x), \tag{9}$$

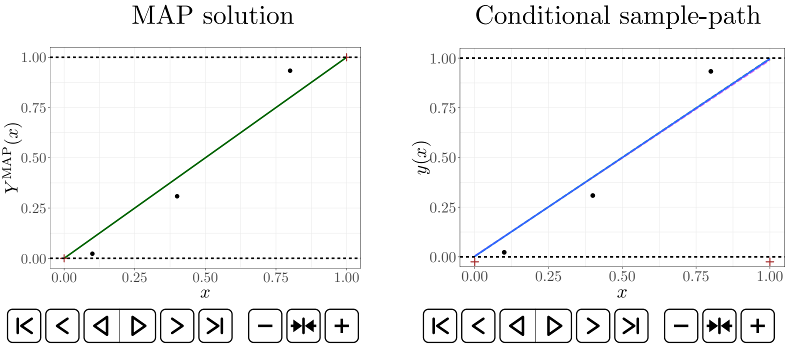
  with $\xi_* = Y(t_*)$. Let $Y_{m_{i+1}}^{\mathrm{MAP}}$ be the MAP solution of (9).
- Then,

$$t_*^{\mathrm{opt}} = \mathrm{argmax}\ {}_{t_*} \int_0^1 [Y_{m_{i+1}}^{\mathrm{MAP}}(x) - Y_{m_i}^{\mathrm{MAP}}(x)]^2 dx. \tag{10}$$

  where the inner integral can be solved matricially.

$\Rightarrow$ By maximising (10), we aim to add new knots in places yielding the highest variation between consecutive MAP solutions.

Numerical illustration in 1D

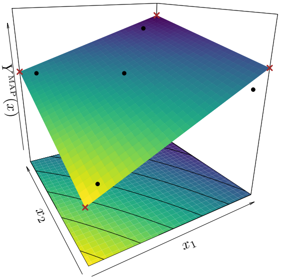


● training points + knots ■ MAP solution
■ predictive mean ■ 90% confidence intervals

Numerical illustration in 2D

For the multidimensional case, i.e. $d \geq 2$, the criterion is given by

$$t_*^{\text{opt}} = \operatorname{argmax}_{t_*} \int_{\mathbf{x}\in[0,1]^d} [Y_{m_{i+1}}^{\text{MAP}}(\mathbf{x}) - Y_{m_i}^{\text{MAP}}(\mathbf{x})]^2 d\mathbf{x}. \tag{11}$$



● training points   + knots   ■ MAP solution

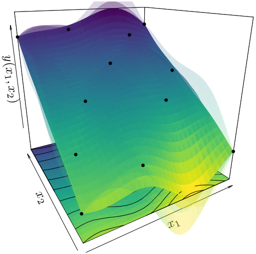
Table of Contents

Conclusions

- We introduced a constrained version of additive GPs under inequality conditions.
- We (sadly) concluded that the inequality constraints cannot be satisfied everywhere using sparse grids.
- We explored an algorithm for the automatic construction of (full) rectangular grids.

Future works

- To keep exploring alternatives in higher dimensions (e.g. using triangular designs, ~~sparse grids~~, ...)
- To scale our frameworks into higher dimensions when additional assumption are made:



  - Inactive dimensions
  - Additivity per blocks
  - ...

- To further investigate some properties of the sequential algorithm for the construction of rectangular grids.
- ✦ To (enthusiastically) start writing the thesis.

Table of Contents

References I

Adams, R. P., Murray, I., and MacKay, D. J. (2009). Tractable nonparametric Bayesian inference in Poisson processes with Gaussian process intensities. In *Proceedings of the 26th International Conference on Machine Learning*, Montreal, Canada.

Bachoc, F. (2014). Asymptotic analysis of the role of spatial sampling for covariance parameter estimation of Gaussian processes. *Journal of Multivariate Analysis*, 125(Supplement C):1 – 35.

Bachoc, F., Lagnoux, A., and López-Lopera, A. F. (2018). Maximum likelihood estimation for Gaussian processes under inequality constraints. *ArXiv e-prints*.

Bay, X., Grammont, L., and Maatouk, H. (2016). Generalization of the Kimeldorf-Wahba correspondence for constrained interpolation. *Electronic journal of statistics*, 10(1):1580–1595.

Botev, Z. I. (2017). The normal law under linear restrictions: simulation and estimation via minimax tilting. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 79(1):125–148.

Cox, D. R. (1955). Some Statistical Methods Connected with Series of Events. *Journal of the Royal Statistical Society. Series B*, 17(2):129–164.

Durrande, N., Ginsbourger, D., and Roustant, O. (2012). Additive covariance kernels for high-dimensional Gaussian process modeling. *Annales de la Faculté de Sciences de Toulouse*, 21(3):481–499.

Duvenaud, D. K., Nickisch, H., and Rasmussen, C. E. (2011). Additive Gaussian processes. In *Advances in Neural Information Processing Systems 24*, pages 226–234. Curran Associates, Inc.

Garcke, J. (2013). Sparse grids in a nutshell. In Garcke, J. and Griebel, M., editors, *Sparse Grids and Applications*, pages 57–80, Berlin, Heidelberg. Springer Berlin Heidelberg.

Genz, A. (1992). Numerical computation of multivariate normal probabilities. *Journal of Computational and Graphical Statistics*, 1:141–150.

John, S. T. and Hensman, J. (2018). Large-Scale Cox Process Inference using Variational Fourier Features. *ArXiv e-prints*.

Lloyd, C. M., Gunter, T., Osborne, M. A., and Roberts, S. J. (2015). Variational inference for Gaussian process modulated Poisson processes. In *Proceedings of the 32nd International Conference on Machine Learning, ICML 2015, Lille, France, 6-11 July 2015*, pages 1814–1822.

López-Lopera, A. F., Bachoc, F., Durrande, N., and Roustant, O. (2018). Finite-dimensional Gaussian approximation with linear inequality constraints. *SIAM/ASA Journal on Uncertainty Quantification*, 6(3):1224–1255.

Luo, Y. and Duraiswami, R. (2013). Fast near-grid Gaussian process regression. In Carvalho, C. M. and Ravikumar, P., editors, *Proceedings of the Sixteenth International Conference on Artificial Intelligence and Statistics*, volume 31 of *Proceedings of Machine Learning Research*, pages 424–432, Scottsdale, Arizona, USA. PMLR.

Maatouk, H. and Bay, X. (2017). Gaussian process emulators for computer experiments with inequality constraints. *Mathematical Geosciences*, 49(5):557–582.

Møller, J., Syversveen, A. R., and Waagepetersen, R. P. (2001). Log Gaussian Cox processes. *Scandinavian Journal of Statistics*, 25(3):451–482.

Moreira-Matias, L., Gama, J., Ferreira, M., Mendes-Moreira, J., and Damas, L. (2013). Predicting taxi-passenger demand using streaming data. *IEEE Transactions on Intelligent Transportation Systems*, 14(3):1393–1402.

Pakman, A. and Paninski, L. (2014). Exact Hamiltonian Monte Carlo for truncated multivariate Gaussians. *Journal of Computational and Graphical Statistics*, 23(2):518–542.

References III

Plumlee, M. (2014). Fast prediction of deterministic functions using sparse grid experimental designs. *Journal of the American Statistical Association*, 109(508):1581–1591.

Rasmussen, C. E. and Williams, C. K. I. (2005). *Gaussian Processes for Machine Learning (Adaptive Computation and Machine Learning)*. The MIT Press.

Stein, M. (1999). *Interpolation of Spatial Data: Some Theory for Kriging*. Springer Series in Statistics. Springer New York.