Ecole Nationale Supérieure des Mines SAINT-ETIENNE

INSTITUT de MATHEMATIQUES de TOULOUSE

Finite-Dimensional Gaussian Approximation with Linear 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.

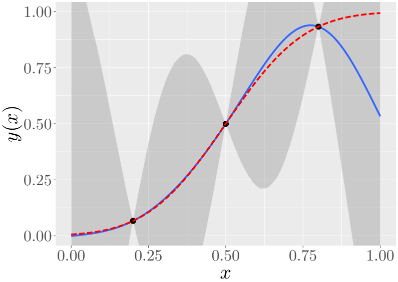
This work is funded by the chair of applied mathematics OQUAIDO.

SIAM UQ 18

April 17, 2018

Gaussian process models: motivation

Target function: is bounded and monotonic.

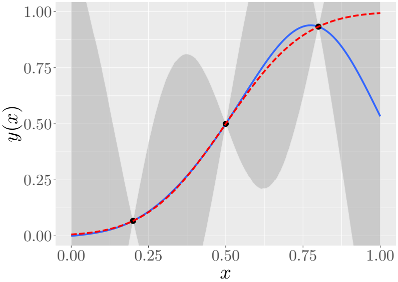


Unconstrained GP.

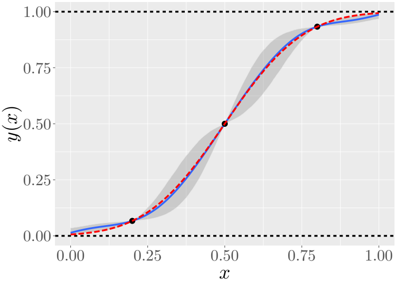
■ true function ● training points
■ predictive mean ■ confidence intervals

Gaussian process models: motivation

Target function: is bounded and monotonic.



Unconstrained GP.



Constrained GP.

true function
predictive mean
● training points
confidence intervals

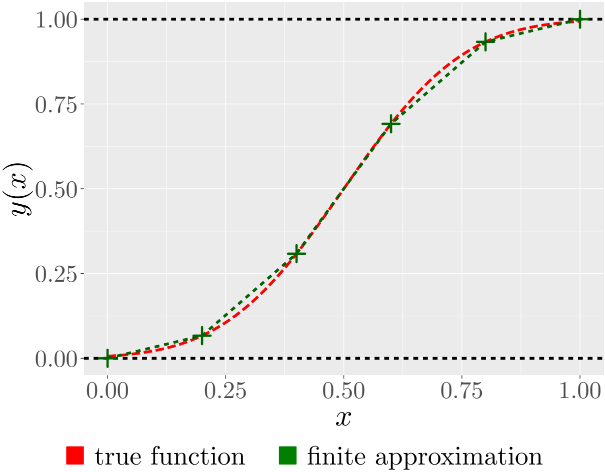
Table of Contents

Table of Contents

Table of Contents

Finite-dimensional Gaussian approximation

Finite representation: is also bounded and monotonic.



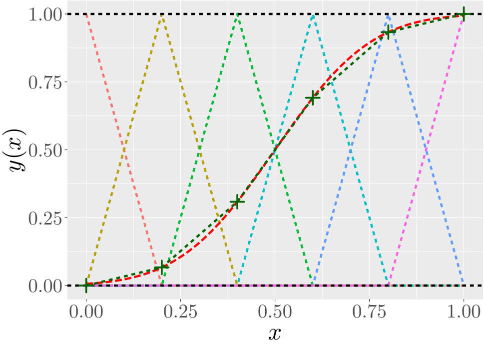
$\Rightarrow$ Imposing both inequality constraints on the knots is enough.

Finite-dimensional Gaussian approximation

Let the finite-dimensional GP approximation be defined as

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

where $\boldsymbol{\xi} = [\xi_1, \cdots, \xi_m]^\top \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Gamma})$, with covariance matrix $\boldsymbol{\Gamma}$ and $\phi_j : [0,1] \to \mathbb{R}$ are hat functions (Maatouk and Bay, 2017):



Finite-dimensional Gaussian approximation

Satisfying linear inequality constraints:

Since $Y_m \in \mathcal{E} \Leftrightarrow \boldsymbol{\xi} \in \mathcal{C}$, we consider $\mathcal{C}$ is composed by

$$\mathcal{C} = \left\{ \mathbf{c} \in \mathbb{R}^m ; \; \forall \, k = 1, \ldots, q \; : \; \ell_k \leq \sum_{j=1}^{m} \lambda_{k,j} c_j \leq u_k \right\}.$$

Therefore, we have $\boldsymbol{\Lambda\xi} | \{\boldsymbol{\Phi\xi} = \mathbf{y}\} \sim \mathcal{N}\left(\boldsymbol{\Lambda\mu}, \boldsymbol{\Lambda\Sigma\Lambda}^\top\right)$ where

$$\boldsymbol{\mu} = \boldsymbol{\Gamma\Phi}^\top [\boldsymbol{\Phi\Gamma\Phi}^\top]^{-1} \mathbf{y}, \quad \text{and} \quad \boldsymbol{\Sigma} = \boldsymbol{\Gamma} - \boldsymbol{\Gamma\Phi}^\top [\boldsymbol{\Phi\Gamma\Phi}^\top]^{-1} \boldsymbol{\Phi\Gamma}.$$

Finite-dimensional Gaussian approximation

Satisfying linear inequality constraints:

Since $Y_m \in \mathcal{E} \Leftrightarrow \boldsymbol{\xi} \in \mathcal{C}$, we consider $\mathcal{C}$ is composed by

$$\mathcal{C} = \left\{ \mathbf{c} \in \mathbb{R}^m ;\ \forall\, k = 1, \ldots, q \ :\ \ell_k \leq \sum_{j=1}^{m} \lambda_{k,j} c_j \leq u_k \right\}.$$

Therefore, we have $\boldsymbol{\Lambda\xi} | \{\boldsymbol{\Phi\xi} = \mathbf{y}\} \sim \mathcal{N}\left(\boldsymbol{\Lambda\mu}, \boldsymbol{\Lambda\Sigma\Lambda}^\top\right)$ where

$$\boldsymbol{\mu} = \boldsymbol{\Gamma\Phi}^\top [\boldsymbol{\Phi\Gamma\Phi}^\top]^{-1}\mathbf{y}, \quad \text{and} \quad \boldsymbol{\Sigma} = \boldsymbol{\Gamma} - \boldsymbol{\Gamma\Phi}^\top [\boldsymbol{\Phi\Gamma\Phi}^\top]^{-1}\boldsymbol{\Phi\Gamma}.$$

Posterior distribution (López-Lopera et al., 2017)

Then, the posterior follows a truncated multinormal distribution

$$\boldsymbol{\Lambda\xi} | \{\boldsymbol{\Phi\xi} = \mathbf{y}, \boldsymbol{l} \leq \boldsymbol{\Lambda\xi} \leq \boldsymbol{u}\} \sim \mathcal{TN}\left(\boldsymbol{\Lambda\mu},\ \boldsymbol{\Lambda\Sigma\Lambda}^\top,\ \boldsymbol{l},\ \boldsymbol{u}\right). \tag{1}$$

Finite-dimensional Gaussian approximation

Satisfying linear inequality constraints:

Since $Y_m \in \mathcal{E} \Leftrightarrow \boldsymbol{\xi} \in \mathcal{C}$, we consider $\mathcal{C}$ is composed by

$$\mathcal{C} = \left\{ \mathbf{c} \in \mathbb{R}^m;\ \forall\, k = 1, \ldots, q\ :\ \ell_k \leq \sum_{j=1}^{m} \lambda_{k,j} c_j \leq u_k \right\}.$$

Therefore, we have $\boldsymbol{\Lambda\xi} | \{\boldsymbol{\Phi\xi} = \mathbf{y}\} \sim \mathcal{N}\left(\boldsymbol{\Lambda\mu}, \boldsymbol{\Lambda\Sigma\Lambda}^\top\right)$ where

$$\boldsymbol{\mu} = \boldsymbol{\Gamma\Phi}^\top [\boldsymbol{\Phi\Gamma\Phi}^\top]^{-1} \mathbf{y}, \quad \text{and} \quad \boldsymbol{\Sigma} = \boldsymbol{\Gamma} - \boldsymbol{\Gamma\Phi}^\top [\boldsymbol{\Phi\Gamma\Phi}^\top]^{-1} \boldsymbol{\Phi\Gamma}.$$

Posterior distribution (López-Lopera et al., 2017)

Then, the posterior follows a truncated multinormal distribution

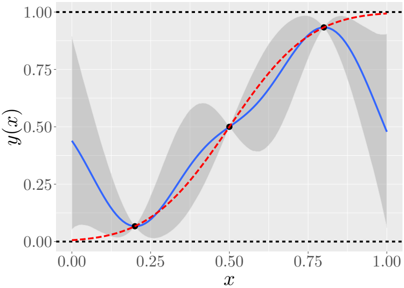
$$\boldsymbol{\Lambda\xi} | \{\boldsymbol{\Phi\xi} = \mathbf{y}, \boldsymbol{l} \leq \boldsymbol{\Lambda\xi} \leq \boldsymbol{u}\} \sim \mathcal{TN}\left(\boldsymbol{\Lambda\mu},\ \boldsymbol{\Lambda\Sigma\Lambda}^\top,\ \boldsymbol{l},\ \boldsymbol{u}\right). \tag{1}$$

$\Rightarrow$ (1) can be approximated via Hamiltonian MC (Pakman and Paninski, 2014).

$\Rightarrow$ What about $\boldsymbol{\Lambda}$, $\boldsymbol{l}$, $\boldsymbol{u}$?

Finite-dimensional Gaussian approximation

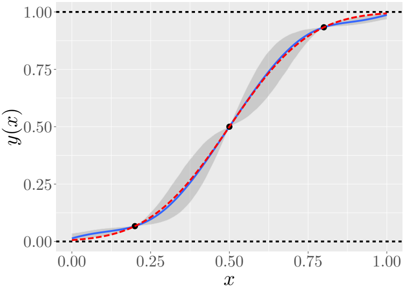
Defining boundedness constraints



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

Finite-dimensional Gaussian approximation

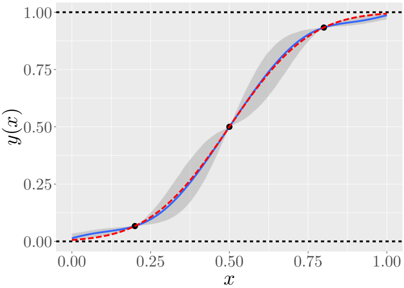
Defining boundedness and monotonicity constraints



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

Finite-dimensional Gaussian approximation

Defining boundedness and monotonicity constraints



or simply,

$$\underbrace{\begin{bmatrix} \boldsymbol{l}_b \\ \boldsymbol{l}_m \end{bmatrix}}_{\boldsymbol{l}} \leq \underbrace{\begin{bmatrix} \boldsymbol{\Lambda}_b \\ \boldsymbol{\Lambda}_m \end{bmatrix}}_{\boldsymbol{\Lambda}} \boldsymbol{\xi} \leq \underbrace{\begin{bmatrix} \boldsymbol{u}_b \\ \boldsymbol{u}_m \end{bmatrix}}_{\boldsymbol{u}}$$

Examples on 2D input spaces

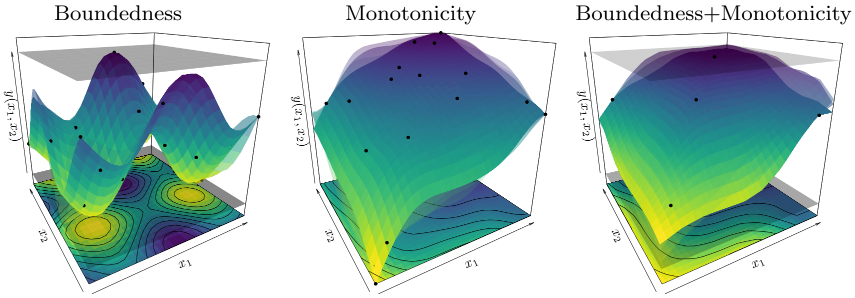


Figure: Examples of 2D Gaussian models with different types of constraints.

Table of Contents

Maximum likelihood (ML): asymptotic properties.

Let $\mathcal{E}_\kappa$ be one of the following convex set of functions

$$\mathcal{E}_\kappa = \begin{cases} f \; : \; \mathbb{X} \to \mathbb{R}, f \text{ is } C^0 \text{ and } \forall \mathbf{x} \in \mathbb{X}, \; \ell \leq f(\mathbf{x}) \leq u & \text{if } \kappa = 0, \\ f \; : \; \mathbb{X} \to \mathbb{R}, f \text{ is } C^1 \text{ and } \forall \mathbf{x} \in \mathbb{X}, \; \forall i = 1, \cdots, d, \; \frac{\partial}{\partial x_i} f(\mathbf{x}) \geq 0 & \text{if } \kappa = 1, \\ f \; : \; \mathbb{X} \to \mathbb{R}, f \text{ is } C^2 \text{ and } \forall \mathbf{x} \in \mathbb{X}, \; \frac{\partial^2}{\partial \mathbf{x}^2} f(\mathbf{x}) \text{ is a non-negative definite matrix} & \text{if } \kappa = 2, \end{cases}$$

which corresponds to boundedness, monotonicity, and convexity constraints. We will focus on the GP Y and the observation vector

$$\mathbf{Y}_n = \left[Y(x_1), \cdots, Y(x_n) \right]^\top .$$

Let the unconstrained likelihood

$$\mathcal{L}_n(\boldsymbol{\theta}) = -\frac{1}{2} \log(\det(\mathbf{R}_{\boldsymbol{\theta}})) - \frac{1}{2} \mathbf{Y}_n^\top \mathbf{R}_{\boldsymbol{\theta}}^{-1} \mathbf{Y}_n - \frac{n}{2} \log 2\pi,$$

with $\mathbf{R}_{\boldsymbol{\theta}} = (k_{\boldsymbol{\theta}}(x_i, x_j))_{1 \leq i,j \leq n}$.

Constrained maximum likelihood (CML)

Proposition: asymptotic consistency of CML

Let $P_{\boldsymbol{\theta}}$ be the distribution of Y with covariance function $k_{\boldsymbol{\theta}}$. Let

$$\mathcal{L}_{\mathcal{C},n}(\boldsymbol{\theta}) = \mathcal{L}_n(\boldsymbol{\theta}) + \log P_{\boldsymbol{\theta}}(Y \in \mathcal{E}_\kappa | \mathbf{Y}_n) - \log P_{\boldsymbol{\theta}}(Y \in \mathcal{E}_\kappa). \quad \text{(Constrained ML)}$$

Assume that $\forall \varepsilon > 0$ and $\forall M < \infty$, (Consistency of the unconditional ML)

$$P\left(\sup_{\|\boldsymbol{\theta}-\boldsymbol{\theta}^*\| \geq \varepsilon} (\mathcal{L}_n(\boldsymbol{\theta}) - \mathcal{L}_n(\boldsymbol{\theta}^*)) \geq -M \right) \xrightarrow[n\to\infty]{} 0.$$

Then, (Consistency of the conditional CML)

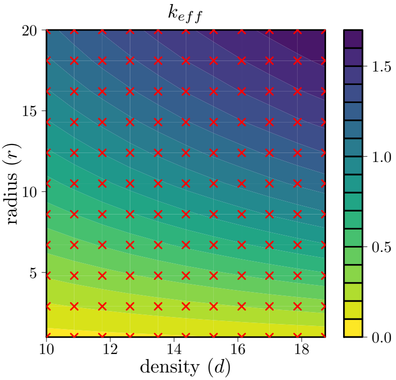
$$P\left(\sup_{\|\boldsymbol{\theta}-\boldsymbol{\theta}^*\| \geq \varepsilon} (\mathcal{L}_{\mathcal{C},n}(\boldsymbol{\theta}) - \mathcal{L}_{\mathcal{C},n}(\boldsymbol{\theta}^*)) \geq -M \,\middle|\, Y \in \mathcal{E}_\kappa \right) \xrightarrow[n\to\infty]{} 0.$$

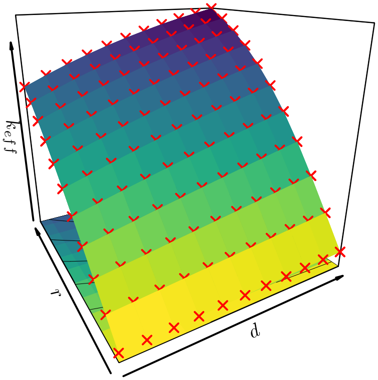
Consequently (Consistency of ML and CML estimators)

$$\operatorname{argmax}_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \ \mathcal{L}_n(\boldsymbol{\theta}) \xrightarrow[n\to\infty]{P} \boldsymbol{\theta}^*, \quad \text{and} \quad \operatorname{argmax}_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \ \mathcal{L}_{\mathcal{C},n}(\boldsymbol{\theta}) \xrightarrow[n\to\infty]{P|Y\in\mathcal{E}_\kappa} \boldsymbol{\theta}^*.$$

Table of Contents

Nuclear criticality dataset (IRSN)



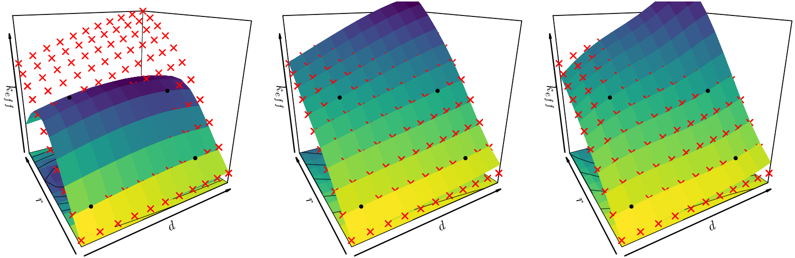


Nuclear criticality safety assessments: IRSN's dataset.

$\Rightarrow$ k_{eff} is positive and non-decreasing.

Nuclear criticality dataset (IRSN)

2D Gaussian models for interpolating the IRSN's dataset using $n = 4$.

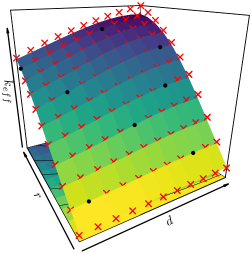


(a) Standard GP+MLE (b) Constrained GP+MLE (c) Constrained GP+cMLE

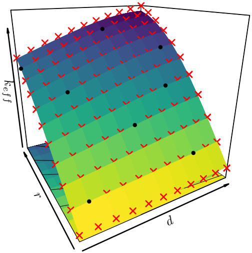
× test data • training points

Nuclear criticality dataset (IRSN)

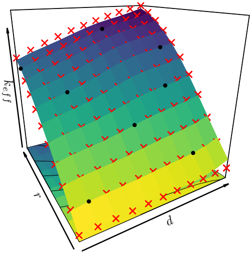
2D Gaussian models for interpolating the IRSN's dataset using $n = 8$.



(d) Standard GP+MLE



(e) Constrained GP+MLE



(f) Constrained GP+cMLE

× test data • training points

Now, we repeat the procedure for 20 random LHDs, and we compute the Q^2 and coverage accuracy (CA) criteria...

Prediction accuracy

Let n_t be the number of test points, $z_1, \cdots, z_{n_t}$ and $\widehat{z}_1, \cdots, \widehat{z}_{n_t}$ the sets of test and predicted observations (respectively), then:

Q^2 criterion:

$$Q^2 = 1 - \frac{\sum_{i=1}^{n_t} (\widehat{z}_i - z_i)^2}{\sum_{i=1}^{n_t} (\overline{z} - z_i)^2}, \tag{2}$$

where $\overline{z}$ is the mean of the test data. $\Rightarrow Q^2 \to 1$ ✓

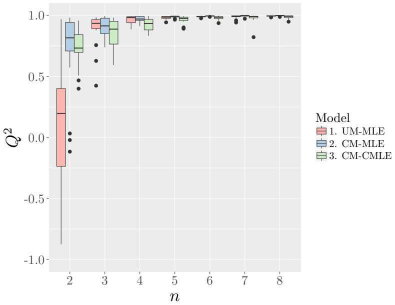
Coverage accuracy (CA) criterion:

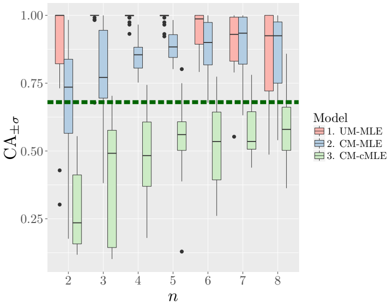
$$\mathrm{CA}_{\pm\sigma} = \frac{1}{n_t} \sum_{i=1}^{n_t} \mathbb{1}_{z_i \in [\widehat{z}_i - \widehat{\sigma}_i, \widehat{z}_i + \widehat{\sigma}_i]} \tag{3}$$

where $\widehat{\sigma}_i$ are the predictive standard deviations. $\Rightarrow \mathrm{CA}_{\pm\sigma} \to 0.68$ ✓

MINES Saint-Étienne

INSTITUT DE MATHÉMATIQUES DE TOULOUSE





Assessment of the models for interpolating the IRSN's dataset using different number of training points n and using twenty different Latin hypercube designs.

$\Rightarrow$ Unconstrained model was often outperformed by constrained ones.

$\Rightarrow$ MLE yields good tradeoff between prediction accuracy and computational cost.

$\Rightarrow$ cMLE provides more accurate confidence intervals.

Table of Contents

Conclusions and Future Works

Conclusions

- We extended the approach from (Maatouk and Bay, 2017): now it works for any linear inequality constraint in 1D or 2D.
- We suggested Hamiltonian MC to approximate the posterior.
- We proved the consistency of the constrained maximum likelihood.
- We implemented the R package: `lineqGPR` (available in June!!).

✦ A.F. López-Lopera, F. Bachoc, N. Durrande, and O. Roustant (2017). Finite-dimensional Gaussian approximation with linear inequality constraints. ArXiv e-prints.

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

Table of Contents

References I

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.

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

López-Lopera, A. F., Bachoc, F., Durrande, N., and Roustant, O. (2017). Finite-dimensional Gaussian approximation with linear inequality constraints. *ArXiv e-prints*.

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

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

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.