Ecole Nationale Supérieure des Mines SAINT-ETIENNE

INSTITUT de MATHEMATIQUES de TOULOUSE

Gaussian process regression models under linear inequality conditions

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.

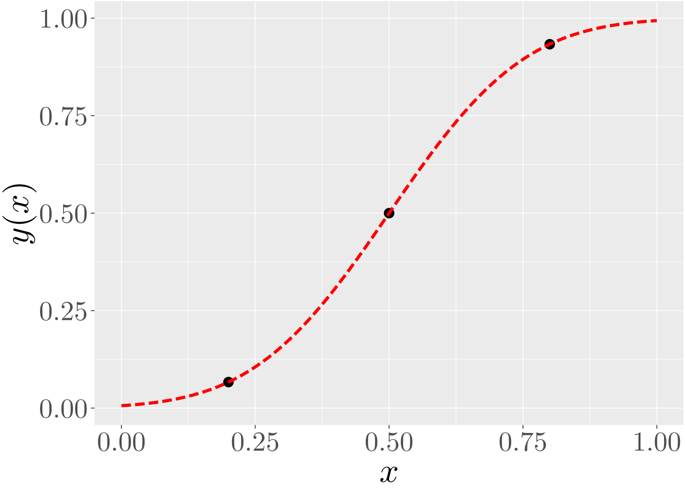
Jan 23, 2018

Table of Contents

Table of Contents

Regression models: a stochastic point of view

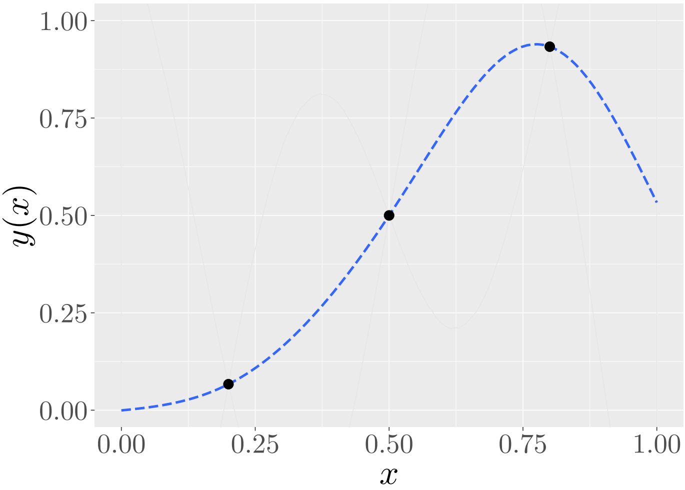
Target function: $x \mapsto \Phi(\frac{x-0.5}{0.2})$ (Gaussian CDF)



● Interpolation data

Regression models: a stochastic point of view

Target function: $x \mapsto \Phi(\frac{x-0.5}{0.2})$ (Gaussian CDF)

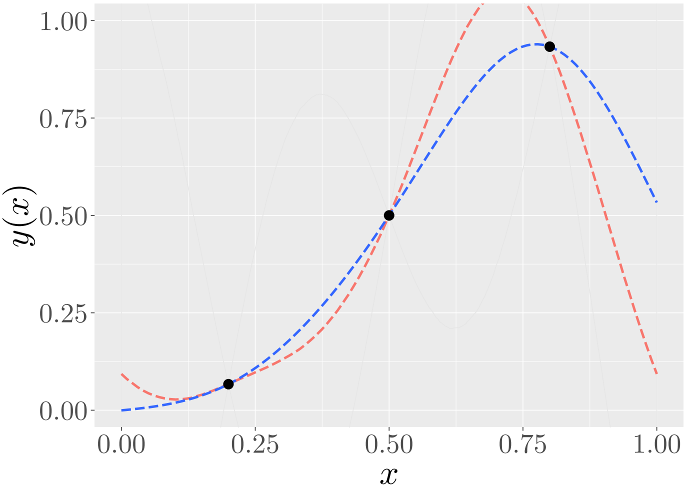


● Interpolation data

■ solution 1

Regression models: a stochastic point of view

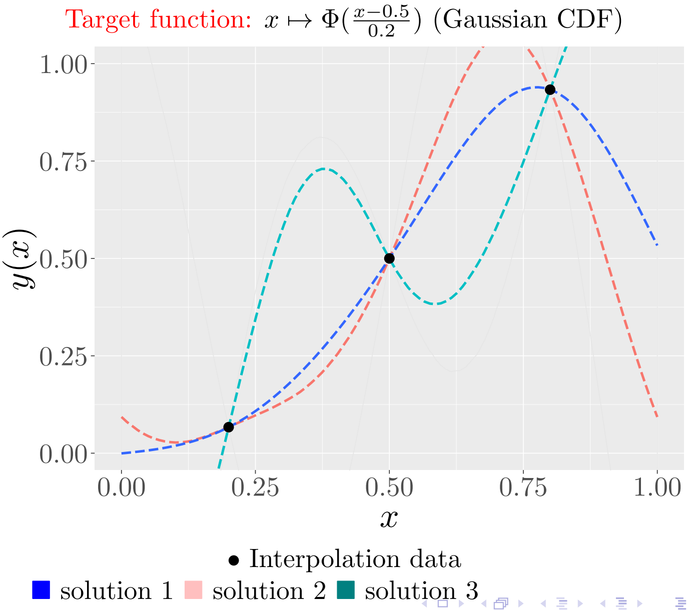
Target function: $x \mapsto \Phi(\frac{x-0.5}{0.2})$ (Gaussian CDF)



● Interpolation data

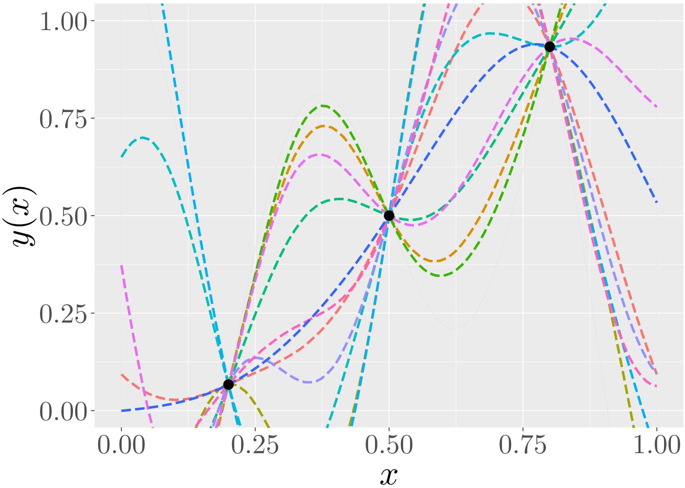
solution 1 solution 2

Regression models: a stochastic point of view



Regression models: a stochastic point of view

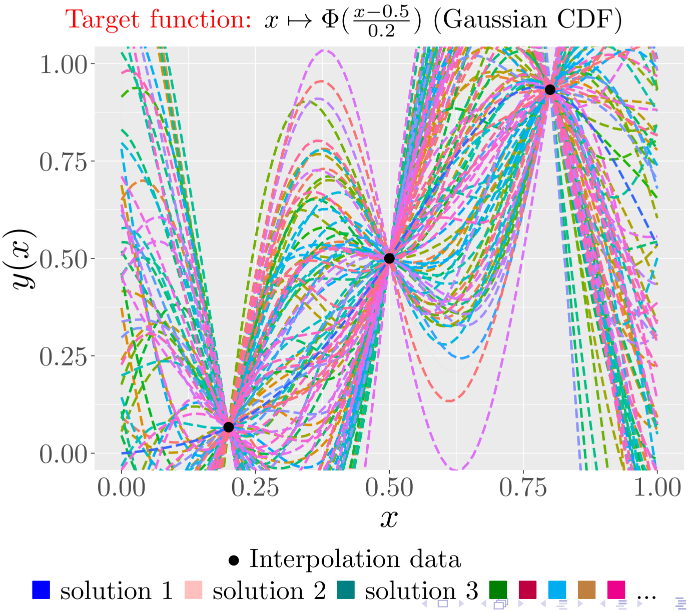
Target function: $x \mapsto \Phi(\frac{x-0.5}{0.2})$ (Gaussian CDF)



● Interpolation data

solution 1 solution 2 solution 3 ...

Regression models: a stochastic point of view



Gaussian process regression

Gaussian process (GP) (Rasmussen and Williams, 2005)

A GP is a collection of random variables, any finite number of which have a joint Gaussian distribution.

Gaussian process regression

Gaussian process (GP) (Rasmussen and Williams, 2005)

A GP is a collection of random variables, any finite number of which have a joint Gaussian distribution.

Notation (Rasmussen and Williams, 2005)

Let Y be a GP. Then, $Y(x)$ is completely defined by its mean function $m(x)$ and covariance function $k(x, x')$

$$Y(x) \sim \mathcal{GP}(m(x), k(x, x')), \tag{1}$$

with

$$\begin{aligned} m(x) &= \mathbb{E}\left\{Y(x)\right\}, \\ k(x, x') &= \mathbb{E}\left\{[Y(x) - m(x)][Y(x') - m(x')]\right\}. \end{aligned}$$

where the operator $\mathbb{E}\left\{\cdot\right\}$ denotes the expectation of a random variable.

Gaussian process regression

GP regression model (Rasmussen and Williams, 2005)

Let Y be a zero-mean GP with covariance function k. Let $\{x_i, y_i\}$ for all $i = 1, \cdots, n$ be a set of observations. Then, the joint distribution of the vector of observations $\mathbf{y}$, and the vector of predictions $\mathbf{y}^*$ is Gaussian

$$\begin{bmatrix} \mathbf{y} \\ \mathbf{y}^* \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} \mathbf{0} \\ \mathbf{0} \end{bmatrix}, \begin{bmatrix} \mathbf{K}(\mathbf{x}, \mathbf{x}) & \mathbf{K}(\mathbf{x}, \mathbf{x}^*) \\ \mathbf{K}(\mathbf{x}^*, \mathbf{x}) & \mathbf{K}(\mathbf{x}^*, \mathbf{x}^*) \end{bmatrix} \right), \tag{2}$$

and the conditional distribution $\mathbf{y}^*|\mathbf{y}$ is also Gaussian

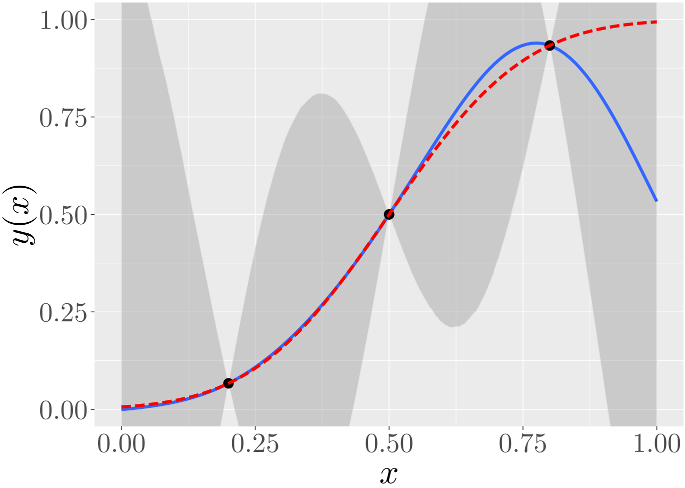
$$\mathbf{y}^*|\mathbf{y} \sim \mathcal{N}\left(\mathbf{m}_{\mathbf{y}^*|\mathbf{y}}, \mathbf{K}_{\mathbf{y}^*|\mathbf{y}} \right), \tag{3}$$

where

$$\begin{aligned} \mathbf{m}_{\mathbf{y}^*|\mathbf{y}} &= \mathbf{K}(\mathbf{x}^*, \mathbf{x})\mathbf{K}(\mathbf{x}, \mathbf{x})^{-1}\mathbf{y}, \\ \mathbf{K}_{\mathbf{y}^*|\mathbf{y}} &= \mathbf{K}(\mathbf{x}^*, \mathbf{x}^*) - \mathbf{K}(\mathbf{x}^*, \mathbf{x})\mathbf{K}(\mathbf{x}, \mathbf{x})^{-1}\mathbf{K}(\mathbf{x}, \mathbf{x}^*). \end{aligned}$$

Gaussian process regression

Target function: $x \mapsto \Phi(\frac{x-0.5}{0.2})$ (Gaussian CDF)



$$\mathbf{m}_{\mathbf{y}^*|\mathbf{y}} = \mathbf{K}(\mathbf{x}^*, \mathbf{x})\mathbf{K}(\mathbf{x}, \mathbf{x})^{-1}\mathbf{y}$$

$$\mathbf{K}_{\mathbf{y}^*|\mathbf{y}} = \mathbf{K}(\mathbf{x}^*, \mathbf{x}^*) - \mathbf{K}(\mathbf{x}^*, \mathbf{x})\mathbf{K}(\mathbf{x}, \mathbf{x})^{-1}\mathbf{K}(\mathbf{x}, \mathbf{x}^*)$$

Gaussian process regression

The GP model performed pretty well although only three training points were considered. But, how can the results be improved?

Gaussian process regression

The GP model performed pretty well although only three training points were considered. But, how can the results be improved?

Possible answers:

1. Adding more training points... but they are not always available.
2. Encoding extra information in the model about the target function.
3. Playing with the kernel function k.

Gaussian process regression

The GP model performed pretty well although only three training points were considered. But, how can the results be improved?

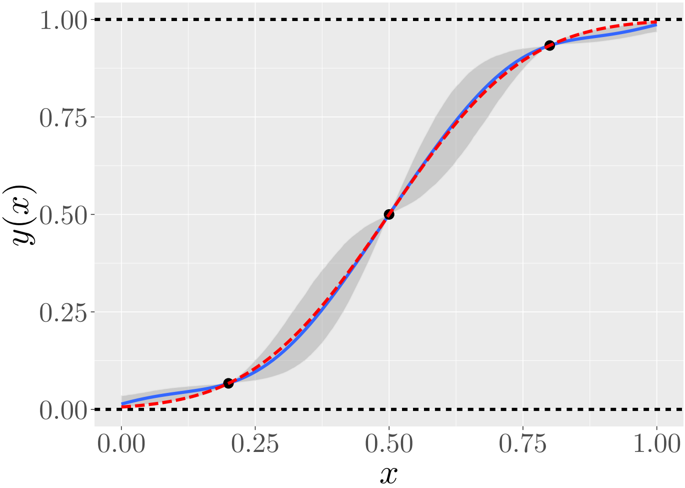
Possible answers:

1. Adding more training points... but they are not always available.
2. Encoding extra information in the model about the target function.
3. Playing with the kernel function k.

What about if we encode a priori both monotonicity and boundedness trends from the Gaussian CDF into the GP model...

Gaussian process regression

Target function: $x \mapsto \Phi(\frac{x-0.5}{0.2})$ (Gaussian CDF)



$\mathbf{m}_{\mathbf{y}^*|\mathbf{y}} = \mathbf{K}(\mathbf{x}^*, \mathbf{x})\mathbf{K}(\mathbf{x}, \mathbf{x})^{-1}\mathbf{y}$?

$\mathbf{K}_{\mathbf{y}^*|\mathbf{y}} = \mathbf{K}(\mathbf{x}^*, \mathbf{x}^*) - \mathbf{K}(\mathbf{x}^*, \mathbf{x})\mathbf{K}(\mathbf{x}, \mathbf{x})^{-1}\mathbf{K}(\mathbf{x}, \mathbf{x}^*)$?

Gaussian process regression

Advantages of using GP models

1. GP models place prior distributions over function spaces.
2. Prior assumptions (e.g. smoothness, stationarity, sparsity) can be encoded in covariance functions.
3. They can be also used in classification task.
4. They have been used in a great variety of real-world problems:
    - Computer science,
    - Physics,
    - Molecular biology and genetics,
    - ...

Gaussian process regression

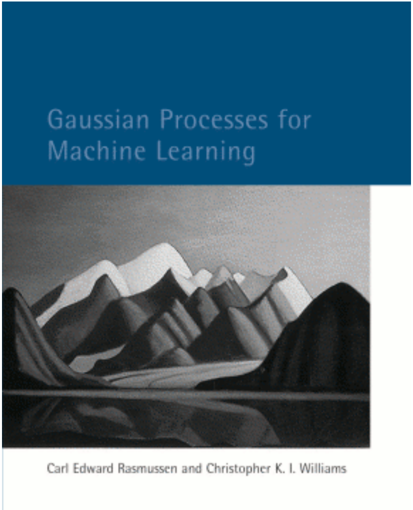
Advantages of using GP models

1. GP models place prior distributions over function spaces.
2. Prior assumptions (e.g. smoothness, stationarity, sparsity) can be encoded in covariance functions.
3. They can be also used in classification task.
4. They have been used in a great variety of real-world problems:
    - Computer science,
    - Physics,
    - Molecular biology and genetics,
    - ...

Disadvantages

1. They are purely data-driven approaches.
2. Curse of dimensionality: $\mathcal{O}(n^3)$.

Rasmussen and Williams (2005)



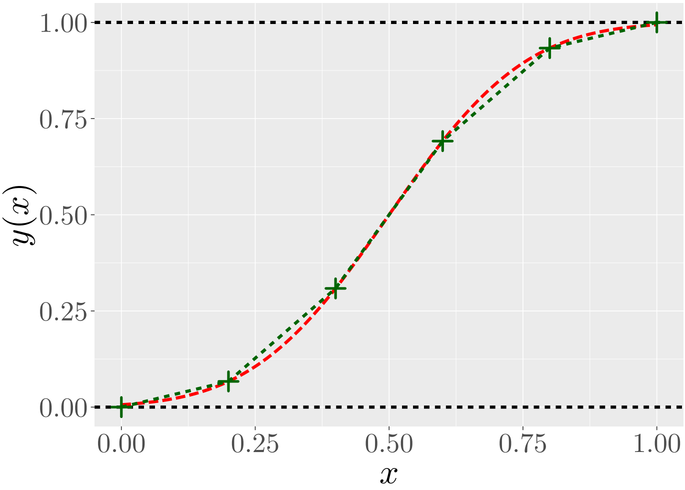
Link: `http://www.gaussianprocess.org/gpml/`

Table of Contents

Table of Contents

Finite-dimensional Gaussian approximation

Target function: $x \mapsto \Phi(\frac{x-0.5}{0.2})$ (Gaussian CDF)



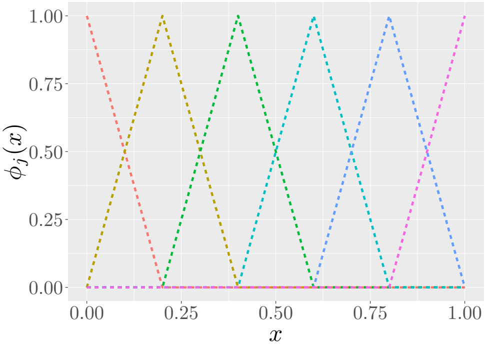
$\Rightarrow$ The finite approx. is also bounded and monotonic!!

Finite-dimensional Gaussian approximation

Let the finite-dimensional GP approximation be defined as

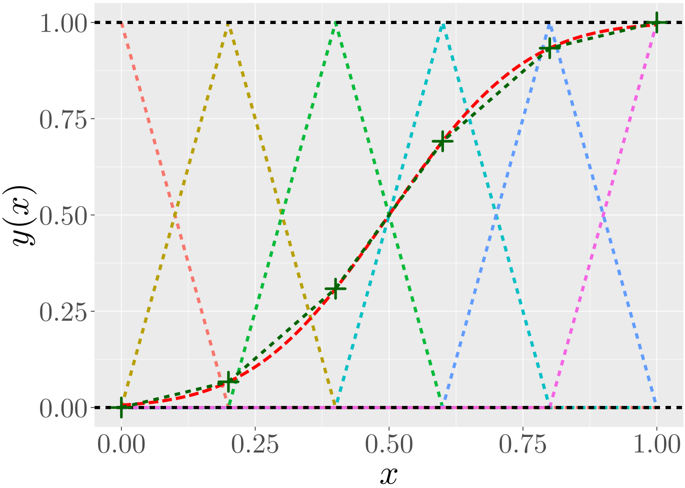
$$Y_m(x) = \sum_{j=1}^{m} \xi_j \phi_j(x), \quad \text{s.t.} \quad Y_m(x_i) = y_i \quad \text{(interpolation conditions)}$$

where $\boldsymbol{\xi} = [\xi_1, \cdots, \xi_m]^\top \sim \mathcal{N}(\mathbf{0}, \mathbf{\Gamma})$, with covariance matrix $\mathbf{\Gamma} = (k(t_i, t_j))1 \leq i, j \leq m$, and $\phi_j : [0, 1] \to \mathbb{R}$ are hat functions.



Finite-dimensional Gaussian approximation

Target function: $x \mapsto \Phi(\frac{x-0.5}{0.2})$ (Gaussian CDF)



Finite-dimensional representation.

Finite-dimensional Gaussian approximation

Since $\boldsymbol{\xi} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Gamma})$, then $\mathbf{Y}_m = [Y_m(x_1), \cdots, Y_m(x_n)]^\top = \boldsymbol{\Phi}\boldsymbol{\xi}$, is also Gaussian

$$\mathbf{Y}_m \sim \mathcal{N}\left(\mathbf{0}, \boldsymbol{\Phi}\boldsymbol{\Gamma}\boldsymbol{\Phi}^\top\right), \tag{4}$$

where $\boldsymbol{\Phi}$ is an $n \times m$ matrix defined by $\boldsymbol{\Phi}_{i,j} = \phi_j(x_i)$. Furthermore, the posterior distribution $\boldsymbol{\xi}|\{\mathbf{Y}_m = \mathbf{y}\}$ follows

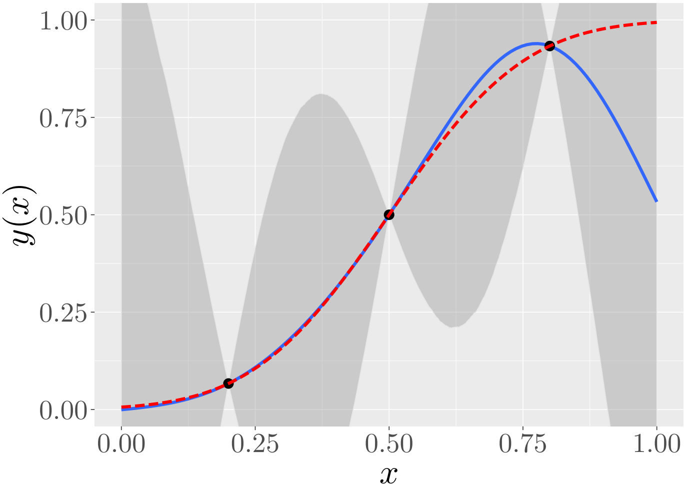
$$\boldsymbol{\xi}|\{\mathbf{Y}_m = \mathbf{y}\} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma}), \tag{5}$$

with

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

Finite-dimensional Gaussian approximation

Target function: $x \mapsto \Phi(\frac{x-0.5}{0.2})$ (Gaussian CDF)



Finite-dimensional Gaussian approximation with $m = 100$.

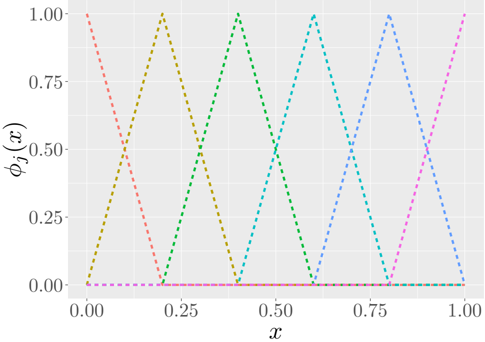
Table of Contents

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



Constrained finite-dimensional Gaussian approximation

First, since $Y_m \in \mathcal{E} \Leftrightarrow \boldsymbol{\xi} \in \mathcal{C}$, we consider $\mathcal{C}$ is composed by a set of q linear inequalities of the form

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

Constrained finite-dimensional Gaussian approximation

First, since $Y_m \in \mathcal{E} \Leftrightarrow \boldsymbol{\xi} \in \mathcal{C}$, we consider $\mathcal{C}$ is composed by a set of q linear inequalities of the form

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

Therefore, we have $\boldsymbol{\Lambda\xi} | \{\boldsymbol{\Phi\xi} = \mathbf{y}\} \sim \mathcal{N}\left(\boldsymbol{\Lambda\mu}, \boldsymbol{\Lambda\Sigma\Lambda}^\top\right)$, and the posterior follows a truncated multinormal (López-Lopera et al., 2017)

$$\boldsymbol{\Lambda\xi} | \{\boldsymbol{\Phi\xi} = \mathbf{y}, \boldsymbol{l} \leq \boldsymbol{\Lambda\xi} \leq \boldsymbol{u}\} \sim \mathcal{TN}\left(\boldsymbol{\Lambda\mu},\ \boldsymbol{\Lambda\Sigma\Lambda}^\top,\ \boldsymbol{l},\ \boldsymbol{u}\right), \tag{6}$$

where $\boldsymbol{\Lambda} = (\lambda_{k,j})_{1\leq k\leq q, 1\leq j\leq m}$, $\boldsymbol{l} = (\ell_k)_{1\leq k\leq q}$, $\boldsymbol{u} = (u_k)_{1\leq k\leq q}$, and

$$\boldsymbol{\mu} = \boldsymbol{\Gamma\Phi}^\top[\boldsymbol{\Phi\Gamma\Phi}^\top]^{-1}\mathbf{y}, \quad \text{and} \quad \boldsymbol{\Sigma} = \boldsymbol{\Gamma} - \boldsymbol{\Gamma\Phi}^\top[\boldsymbol{\Phi\Gamma\Phi}^\top]^{-1}\boldsymbol{\Phi\Gamma}. \tag{7}$$

Constrained finite-dimensional Gaussian approximation

First, since $Y_m \in \mathcal{E} \Leftrightarrow \boldsymbol{\xi} \in \mathcal{C}$, we consider $\mathcal{C}$ is composed by a set of q linear inequalities of the form

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

Therefore, we have $\boldsymbol{\Lambda\xi} | \{\boldsymbol{\Phi\xi} = \mathbf{y}\} \sim \mathcal{N}\left(\boldsymbol{\Lambda\mu}, \boldsymbol{\Lambda\Sigma\Lambda}^\top\right)$, and the posterior follows a truncated multinormal (López-Lopera et al., 2017)

$$\boldsymbol{\Lambda\xi} | \{\boldsymbol{\Phi\xi} = \mathbf{y}, \boldsymbol{l} \leq \boldsymbol{\Lambda\xi} \leq \boldsymbol{u}\} \sim \mathcal{TN}\left(\boldsymbol{\Lambda\mu},\ \boldsymbol{\Lambda\Sigma\Lambda}^\top,\ \boldsymbol{l},\ \boldsymbol{u}\right), \tag{6}$$

where $\boldsymbol{\Lambda} = (\lambda_{k,j})_{1 \leq k \leq q, 1 \leq j \leq m}$, $\boldsymbol{l} = (\ell_k)_{1 \leq k \leq q}$, $\boldsymbol{u} = (u_k)_{1 \leq k \leq q}$, and

$$\boldsymbol{\mu} = \boldsymbol{\Gamma\Phi}^\top [\boldsymbol{\Phi\Gamma\Phi}^\top]^{-1} \mathbf{y}, \quad \text{and} \quad \boldsymbol{\Sigma} = \boldsymbol{\Gamma} - \boldsymbol{\Gamma\Phi}^\top [\boldsymbol{\Phi\Gamma\Phi}^\top]^{-1} \boldsymbol{\Phi\Gamma}. \tag{7}$$

$\Rightarrow$ The distribution of Equation (6) can be approximated using MCMC.

Constrained finite-dimensional Gaussian approximation

First, since $Y_m \in \mathcal{E} \Leftrightarrow \boldsymbol{\xi} \in \mathcal{C}$, we consider $\mathcal{C}$ is composed by a set of q linear inequalities of the form

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

Therefore, we have $\boldsymbol{\Lambda}\boldsymbol{\xi} | \{\boldsymbol{\Phi}\boldsymbol{\xi} = \mathbf{y}\} \sim \mathcal{N}\left(\boldsymbol{\Lambda}\boldsymbol{\mu}, \boldsymbol{\Lambda}\boldsymbol{\Sigma}\boldsymbol{\Lambda}^\top\right)$, and the posterior follows a truncated multinormal (López-Lopera et al., 2017)

$$\boldsymbol{\Lambda}\boldsymbol{\xi} | \{\boldsymbol{\Phi}\boldsymbol{\xi} = \mathbf{y}, \boldsymbol{l} \leq \boldsymbol{\Lambda}\boldsymbol{\xi} \leq \boldsymbol{u}\} \sim \mathcal{TN}\left(\boldsymbol{\Lambda}\boldsymbol{\mu},\ \boldsymbol{\Lambda}\boldsymbol{\Sigma}\boldsymbol{\Lambda}^\top,\ \boldsymbol{l},\ \boldsymbol{u}\right), \tag{6}$$

where $\boldsymbol{\Lambda} = (\lambda_{k,j})_{1 \leq k \leq q, 1 \leq j \leq m}$, $\boldsymbol{l} = (\ell_k)_{1 \leq k \leq q}$, $\boldsymbol{u} = (u_k)_{1 \leq k \leq q}$, and

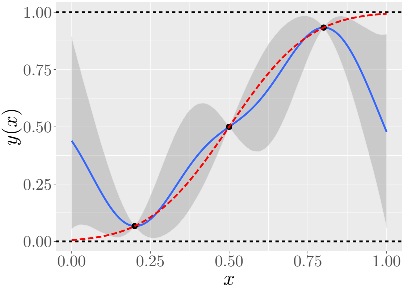
$$\boldsymbol{\mu} = \boldsymbol{\Gamma}\boldsymbol{\Phi}^\top[\boldsymbol{\Phi}\boldsymbol{\Gamma}\boldsymbol{\Phi}^\top]^{-1}\mathbf{y}, \quad \text{and} \quad \boldsymbol{\Sigma} = \boldsymbol{\Gamma} - \boldsymbol{\Gamma}\boldsymbol{\Phi}^\top[\boldsymbol{\Phi}\boldsymbol{\Gamma}\boldsymbol{\Phi}^\top]^{-1}\boldsymbol{\Phi}\boldsymbol{\Gamma}. \tag{7}$$

$\Rightarrow$ The distribution of Equation (6) can be approximated using MCMC.

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

Constrained finite-dimensional Gaussian approximation

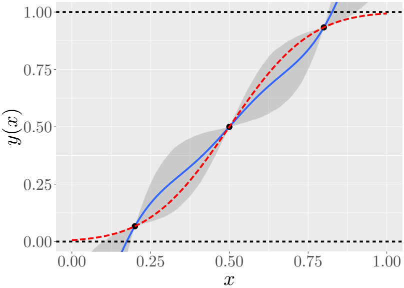
Defining boundedness constraints



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

Constrained finite-dimensional Gaussian approximation

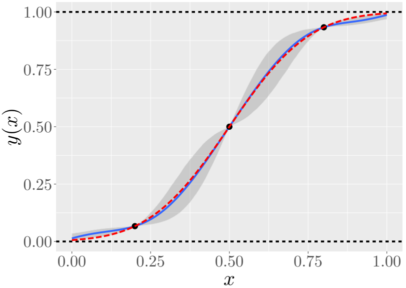
Defining monotonicity constraints



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

Constrained finite-dimensional Gaussian approximation

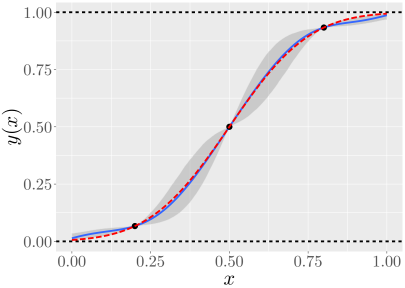
Defining boundedness and monotonicity constraints



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

Constrained finite-dimensional Gaussian approximation

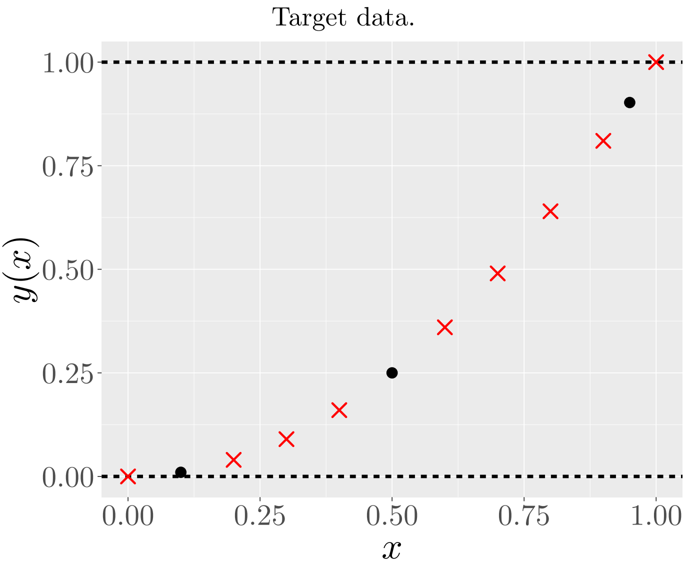
Defining boundedness and monotonicity constraints



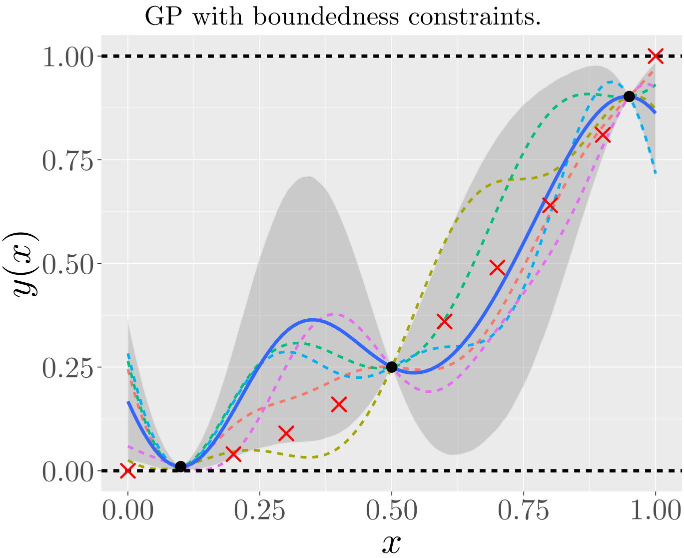
or simply,

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

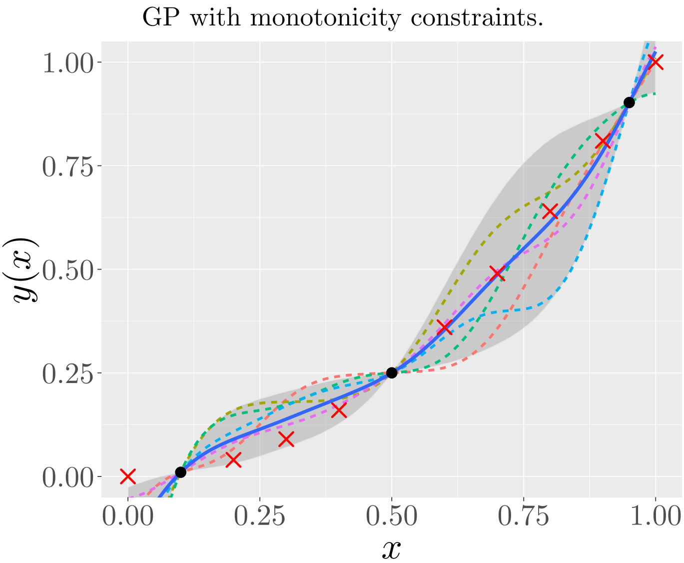
Toy example 1: $x \mapsto x^2$



Toy example 1: $x \mapsto x^2$

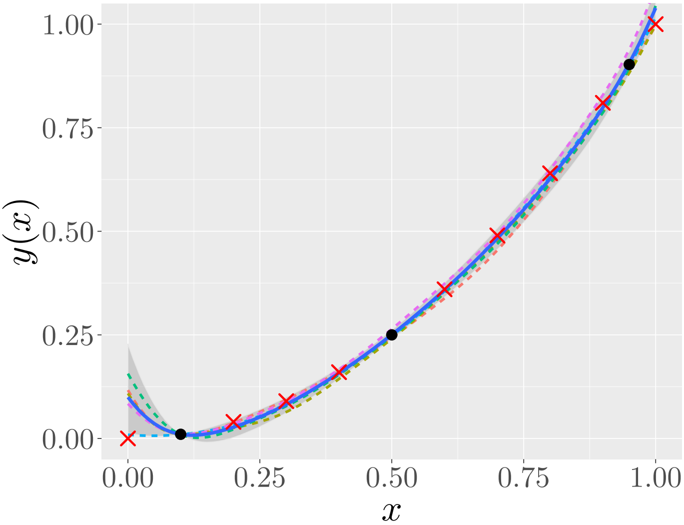


Toy example 1: $x \mapsto x^2$



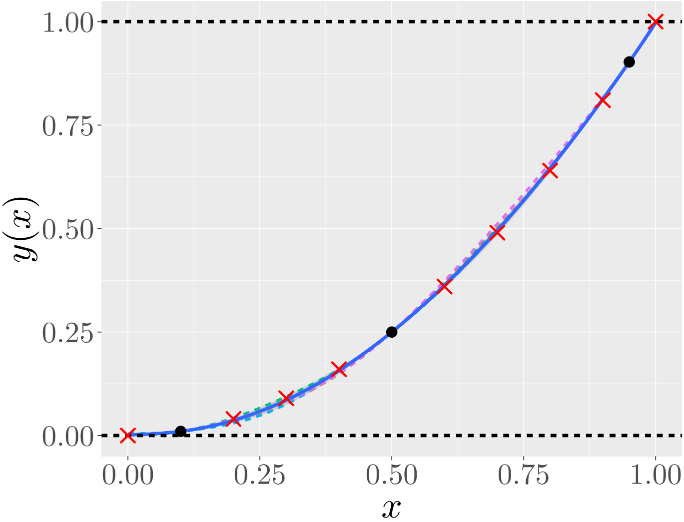
Toy example 1: $x \mapsto x^2$

GP with convexity constraints.

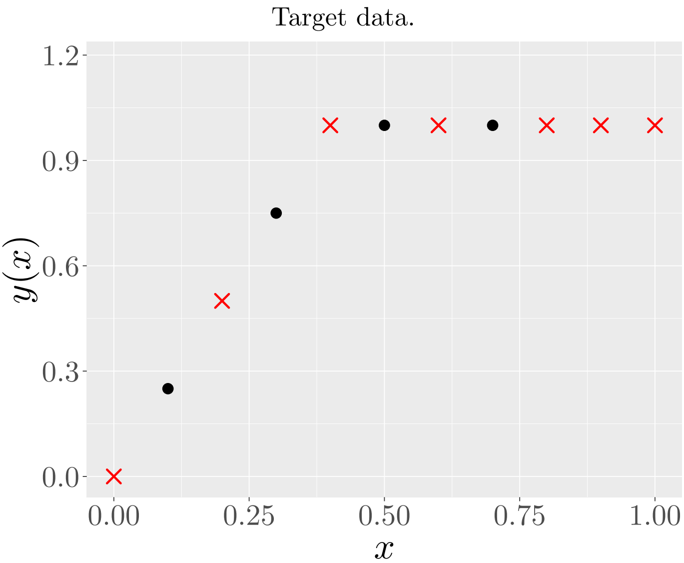


Toy example 1: $x \mapsto x^2$

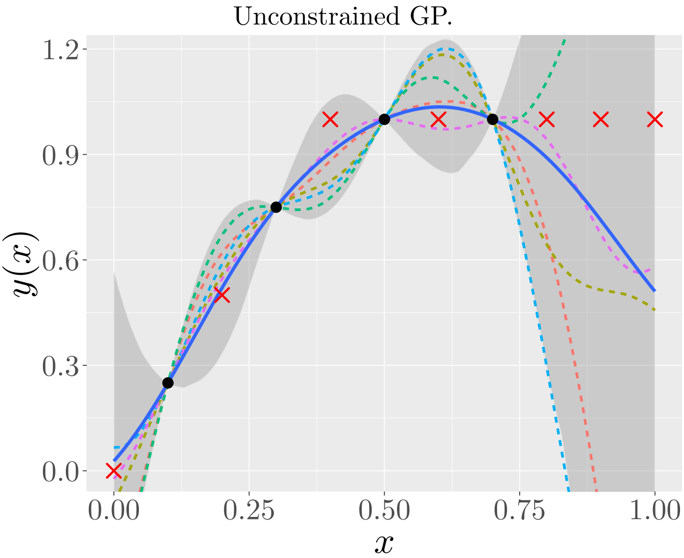
GP with boundedness+monotonicity+convexity constraints.



Toy example 2: sequential constraints

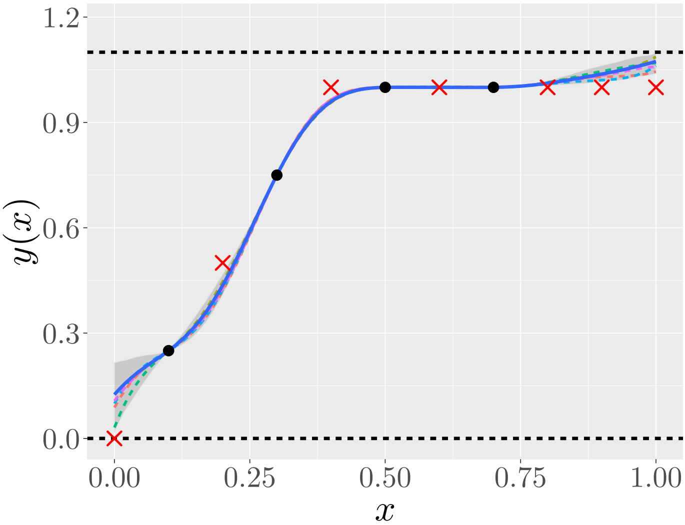


Toy example 2: sequential constraints



Toy example 2: sequential constraints

GP with boundedness+monotonicity constraints.



Toy example 2: sequential constraints

GP with sequential constraints.

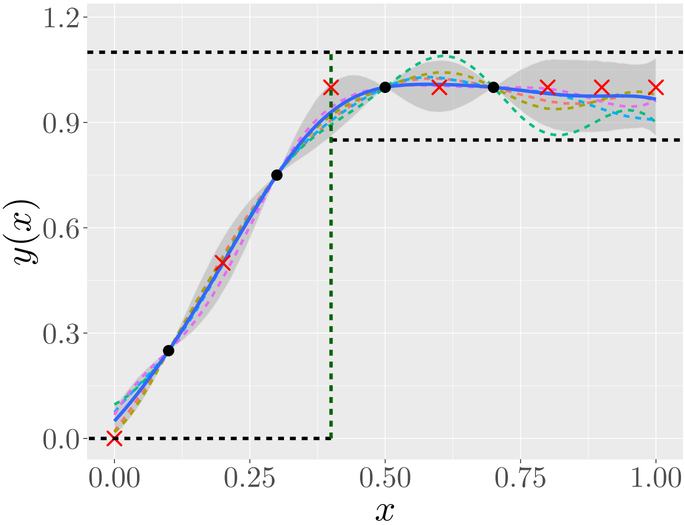


Table of Contents

Extension for 2D input spaces

The approximation can be extended to two dimensional input spaces by tensorisation:

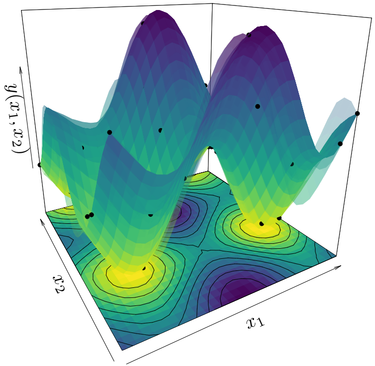
$$Y_{m_1,m_2}(x_1,x_2) := \sum_{j_1=1}^{m_1}\sum_{j_2=1}^{m_2} \xi_{j_2,j_1}\phi^1_{j_1}(x_1)\phi^2_{j_2}(x_2), \text{ s.t. } \begin{cases} Y_{m_1,m_2}\left(x_1^i, x_2^i\right) = y_i, \\ \xi_{j_2,j_1} \in \mathcal{C}, \end{cases}$$

where $\xi_{j_2,j_1} = Y(t_{j_1}, t_{j_2})$ and $(x_1^1, x_2^1), \cdots, (x_1^n, x_2^n)$ constitute a DoE, and $\phi^1_{j_1}, \phi^2_{j_2} : [0,1] \to \mathbb{R}$ are hat functions.

$\Rightarrow$ We can also assume that $\boldsymbol{\xi} = \left[\xi_{1,1}, \cdots, \xi_{1,m1}, \cdots, \xi_{m2,1}, \cdots \xi_{m2,m1}\right]^\top$ is a zero-mean Gaussian vector with covariance matrix $\boldsymbol{\Gamma}$.

Extension for 2D input spaces

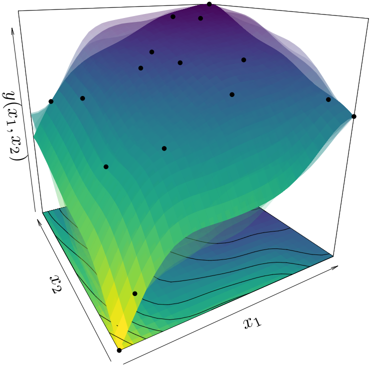
Boundedness in 2D.



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

Extension for 2D input spaces

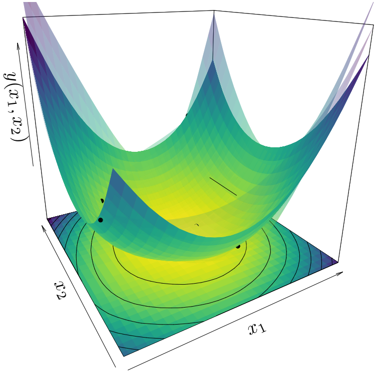
Monotonicity in 2D.



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

Extension for 2D input spaces

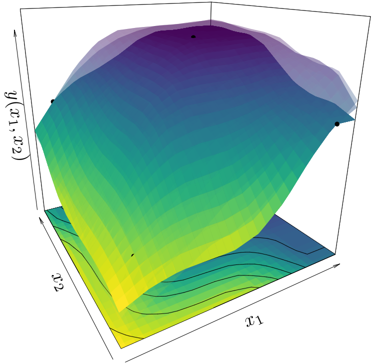
Convexity in 2D (a weak version).



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

Extension for 2D input spaces

Boundedness and monotonicity in 2D.



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

Table of Contents

Table of Contents

Maximum likelihood estimation (MLE)

Maximum likelihood (ML) estimation

Consider that the zero-mean GP Y has covariance function $k_{\boldsymbol{\theta}^*}$ for an unknown $\boldsymbol{\theta}^* \in \boldsymbol{\Theta}$. Let $\mathcal{L}_m(\boldsymbol{\theta})$ be the log likelihood of $\boldsymbol{\theta}$

$$\mathcal{L}_m(\boldsymbol{\theta}) = \log p_{\boldsymbol{\theta}}(\mathbf{Y}_m) = -\frac{1}{2}\log(\det(\mathbf{K}_{\boldsymbol{\theta}})) - \frac{1}{2}\mathbf{Y}_m^\top \mathbf{K}_{\boldsymbol{\theta}}^{-1}\mathbf{Y}_m - \frac{n}{2}\log 2\pi, \tag{8}$$

with $\mathbf{K}_{\boldsymbol{\theta}} = \boldsymbol{\Phi}\boldsymbol{\Gamma}_{\boldsymbol{\theta}}\boldsymbol{\Phi}^\top$ and $\boldsymbol{\Gamma}_{\boldsymbol{\theta}} = (k_{\boldsymbol{\theta}}(t_i, t_j))_{1 \leq i,j \leq m}$. Then, $\boldsymbol{\theta}^*$ can be estimated by maximising Equation (8) w.r.t. $\boldsymbol{\theta} \in \boldsymbol{\Theta}$

$$\widehat{\boldsymbol{\theta}}_{\mathrm{MLE}} = \arg\max_{\boldsymbol{\theta} \in \boldsymbol{\Theta}} \ \mathcal{L}_m(\boldsymbol{\theta}). \tag{9}$$

Maximum likelihood estimation (MLE)

Maximum likelihood (ML) estimation

Consider that the zero-mean GP Y has covariance function $k_{\boldsymbol{\theta}^*}$ for an unknown $\boldsymbol{\theta}^* \in \boldsymbol{\Theta}$. Let $\mathcal{L}_m(\boldsymbol{\theta})$ be the log likelihood of $\boldsymbol{\theta}$

$$\mathcal{L}_m(\boldsymbol{\theta}) = \log p_{\boldsymbol{\theta}}(\mathbf{Y}_m) = -\frac{1}{2}\log(\det(\mathbf{K}_{\boldsymbol{\theta}})) - \frac{1}{2}\mathbf{Y}_m^\top \mathbf{K}_{\boldsymbol{\theta}}^{-1}\mathbf{Y}_m - \frac{n}{2}\log 2\pi, \tag{8}$$

with $\mathbf{K}_{\boldsymbol{\theta}} = \boldsymbol{\Phi}\boldsymbol{\Gamma}_{\boldsymbol{\theta}}\boldsymbol{\Phi}^\top$ and $\boldsymbol{\Gamma}_{\boldsymbol{\theta}} = (k_{\boldsymbol{\theta}}(t_i, t_j))_{1 \leq i,j \leq m}$. Then, $\boldsymbol{\theta}^*$ can be estimated by maximising Equation (8) w.r.t. $\boldsymbol{\theta} \in \boldsymbol{\Theta}$

$$\widehat{\boldsymbol{\theta}}_{\mathrm{MLE}} = \arg\max_{\boldsymbol{\theta} \in \boldsymbol{\Theta}} \ \mathcal{L}_m(\boldsymbol{\theta}). \tag{9}$$

$\Rightarrow$ Equation (9) does not take into account the constraints $\boldsymbol{\xi} \in \mathcal{C}$.

Constrained maximum likelihood estimation (CMLE)

Constrained maximum likelihood (CML) estimation

Let $p_{\boldsymbol{\theta}}(\mathbf{Y}_m|\boldsymbol{\xi} \in \mathcal{C})$ be the conditional pdf of $\mathbf{Y}_m$ given $\boldsymbol{\xi} \in \mathcal{C}$, when Y has covariance function $k_{\boldsymbol{\theta}}$. By using Bayes' theorem, the constrained log likelihood $\mathcal{L}_{\mathcal{C},m}(\boldsymbol{\theta}) = \log p_{\boldsymbol{\theta}}(\mathbf{Y}_m|\boldsymbol{\xi} \in \mathcal{C})$ is

$$\begin{aligned}\mathcal{L}_{\mathcal{C},m}(\boldsymbol{\theta}) &= \log \frac{p_{\boldsymbol{\theta}}(\mathbf{Y}_m) P_{\boldsymbol{\theta}}(\boldsymbol{\xi} \in \mathcal{C}|\boldsymbol{\Phi}\boldsymbol{\xi} = \mathbf{Y}_m)}{P_{\boldsymbol{\theta}}(\boldsymbol{\xi} \in \mathcal{C})} \\ &= \log p_{\boldsymbol{\theta}}(\mathbf{Y}_m) + \log P_{\boldsymbol{\theta}}(\boldsymbol{\xi}\in\mathcal{C}|\boldsymbol{\Phi}\boldsymbol{\xi}=\mathbf{Y}_m) - \log P_{\boldsymbol{\theta}}(\boldsymbol{\xi}\in\mathcal{C}),\end{aligned} \tag{10}$$

Then, the constrained CML estimator is given by

$$\widehat{\boldsymbol{\theta}}_{\text{CMLE}} = \arg\max_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \mathcal{L}_{\mathcal{C},m}(\boldsymbol{\theta}). \tag{11}$$

Constrained maximum likelihood estimation (CMLE)

Constrained maximum likelihood (CML) estimation

Let $p_{\boldsymbol{\theta}}(\mathbf{Y}_m | \boldsymbol{\xi} \in \mathcal{C})$ be the conditional pdf of $\mathbf{Y}_m$ given $\boldsymbol{\xi} \in \mathcal{C}$, when Y has covariance function $k_{\boldsymbol{\theta}}$. By using Bayes' theorem, the constrained log likelihood $\mathcal{L}_{\mathcal{C},m}(\boldsymbol{\theta}) = \log p_{\boldsymbol{\theta}}(\mathbf{Y}_m | \boldsymbol{\xi} \in \mathcal{C})$ is

$$\begin{aligned} \mathcal{L}_{\mathcal{C},m}(\boldsymbol{\theta}) &= \log \frac{p_{\boldsymbol{\theta}}(\mathbf{Y}_m) P_{\boldsymbol{\theta}}(\boldsymbol{\xi} \in \mathcal{C} | \boldsymbol{\Phi}\boldsymbol{\xi} = \mathbf{Y}_m)}{P_{\boldsymbol{\theta}}(\boldsymbol{\xi} \in \mathcal{C})} \\ &= \log p_{\boldsymbol{\theta}}(\mathbf{Y}_m) + \log P_{\boldsymbol{\theta}}(\boldsymbol{\xi} \in \mathcal{C} | \boldsymbol{\Phi}\boldsymbol{\xi} = \mathbf{Y}_m) - \log P_{\boldsymbol{\theta}}(\boldsymbol{\xi} \in \mathcal{C}), \end{aligned} \tag{10}$$

Then, the constrained CML estimator is given by

$$\widehat{\boldsymbol{\theta}}_{\mathrm{CMLE}} = \arg\max_{\boldsymbol{\theta} \in \boldsymbol{\Theta}} \mathcal{L}_{\mathcal{C},m}(\boldsymbol{\theta}). \tag{11}$$

- Unconstrained log likelihood.
- Gaussian orthant probabilities (Botev, 2017; Genz, 1992).

Table of Contents

Asymptotic analysis

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

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

Maximum likelihood (ML)

Proposition 1: asymptotic consistency of ML

Let Y be a centred GP on $\mathbb{X} \subset \mathbb{R}^d$ with covariance k satisfying Condition A.1 from (López-Lopera et al., 2017). Let $\mathbf{Y}_n = \left[\, Y(x_1), \cdots, Y(x_n) \,\right]^\top$. Let

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

with $\mathbf{R}_{\boldsymbol{\theta}} = (k_{\boldsymbol{\theta}}(x_i, x_j))_{1 \leq i,j \leq n}$. Let $\widehat{\boldsymbol{\theta}} \in \arg\max_{\boldsymbol{\theta} \in \boldsymbol{\Theta}} \mathcal{L}_n(\boldsymbol{\theta})$. Assume $\forall \varepsilon > 0$,

$$P(\|\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}^*\| \geq \varepsilon) \xrightarrow[n \to \infty]{} 0. \quad \text{(Consistency of the unconditional ML)}$$

Let $\kappa \in \{0, 1, 2\}$. Let $\mathcal{E}_\kappa$. Then, we have $P(Y \in \mathcal{E}_\kappa) > 0$ from Lemmas A.3, A.4 and A.5 of (López-Lopera et al., 2017), and thus

$$P(\|\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}^*\| \geq \varepsilon \mid Y \in \mathcal{E}_\kappa) \xrightarrow[n \to \infty]{} 0. \quad \text{(Consistency of the conditional ML)}$$

Maximum likelihood (ML)

Proof 1: asymptotic consistency of ML

We have

$$P(\|\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}^*\| \geq \varepsilon |\ Y \in \mathcal{E}_\kappa) = \frac{P(\|\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}^*\| \geq \varepsilon,\ Y \in \mathcal{E}_\kappa)}{P(Y \in \mathcal{E}_\kappa)} \leq \frac{P(\|\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}^*\| \geq \varepsilon)}{P(Y \in \mathcal{E}_\kappa)}.$$

Since $P(Y \in \mathcal{E}_\kappa) > 0$ is fixed, and $P(\|\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}^*\| \geq \varepsilon) \xrightarrow[n \to \infty]{} 0$, the result follows.

Constrained maximum likelihood (CML)

Proposition 2: asymptotic consistency of CML

We use the same notations and assumptions as in Proposition 1. 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)

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

Constrained maximum likelihood (CML)

Proof 2: asymptotic consistency of CML

We have from Lemmas A.1 and A.2 from (López-Lopera et al., 2017) that $\forall \varepsilon > 0$

$$P(\log(P_{\boldsymbol{\theta}^*}(Y \in \mathcal{E}_\kappa \mid \mathbf{Y}_n)) \geq \log(1-\varepsilon) \mid Y \in \mathcal{E}_\kappa) \xrightarrow[n\to\infty]{} 1.$$

Hence $\forall \delta > 0$

$$P\left(\sup_{\|\boldsymbol{\theta}-\boldsymbol{\theta}^*\| \geq \varepsilon} \log(P_{\boldsymbol{\theta}}(Y \in \mathcal{E}_\kappa \mid \mathbf{Y}_n)) - \log(P_{\boldsymbol{\theta}^*}(Y \in \mathcal{E}_\kappa \mid \mathbf{Y}_n)) \geq \delta \,\middle|\, Y \in \mathcal{E}_\kappa \right) \xrightarrow[n\to\infty]{} 0.$$

Also, from Lemma A.6, there exists $\Delta > 0$ so that we have

$$\inf_{\|\boldsymbol{\theta}-\boldsymbol{\theta}^*\| \geq \varepsilon} P_{\boldsymbol{\theta}}(Y \in \mathcal{E}_\kappa) \geq \Delta > 0,$$

so that

$$\sup_{\|\boldsymbol{\theta}-\boldsymbol{\theta}^*\| \geq \varepsilon} -\log(P_{\boldsymbol{\theta}}(Y \in \mathcal{E}_\kappa)) + \log(P_{\boldsymbol{\theta}^*}(Y \in \mathcal{E}_\kappa)) \leq -\log(\Delta) < \infty.$$

Hence, the proposition follows.

Table of Contents

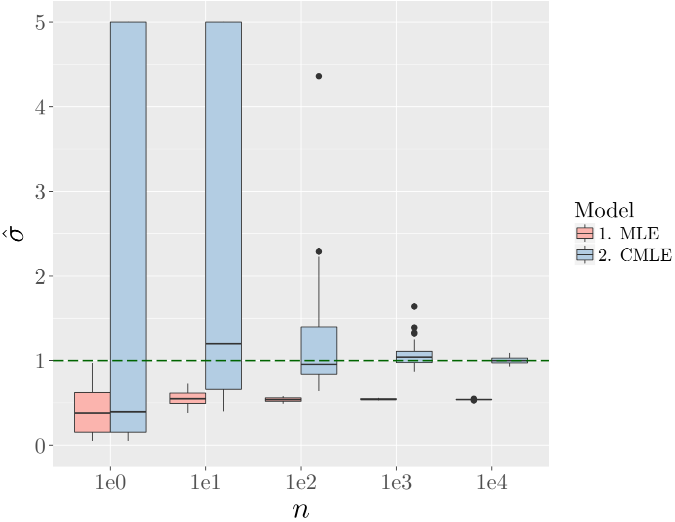
Results: univariate truncated normal

We estimate $(\sigma^*)^2 = 1$ using different nb of observations.

- We simulate from the truncated normal $Y_i \sim \mathcal{TN}\left(0, (\sigma^*)^2, -1, 1\right)$ for $i = 1, \cdots n$.
- We estimate $\hat{\sigma}$ via MLE and CMLE.
- We repeat the experiment 100 times.

Results: univariate truncated normal

Estimating $\hat{\sigma}$ given different nb of samples.



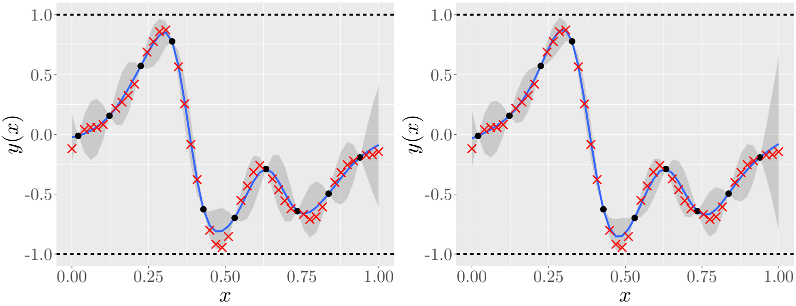
Results: simulation study

We estimate $\boldsymbol{\theta}^* = (\sigma^2 = 1, \theta = 0.2)$ using different realizations.

- We simulate from a zero-mean constrained GP Y using a Matérn 5/2 covariance function such that $Y \in [-1, 1]$.
- Nb of realizations: 100.
- We estimate $\hat{\boldsymbol{\theta}}$ via MLE and CMLE.
- We evaluate the consistently estimable ratio $\rho = \sigma^2/\theta^5$.

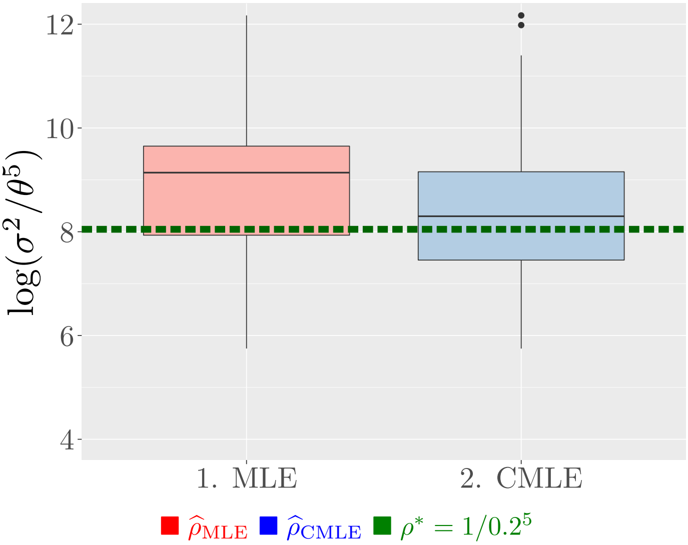
Results: simulation study

Prediction results using (left) MLE and (right) CMLE.



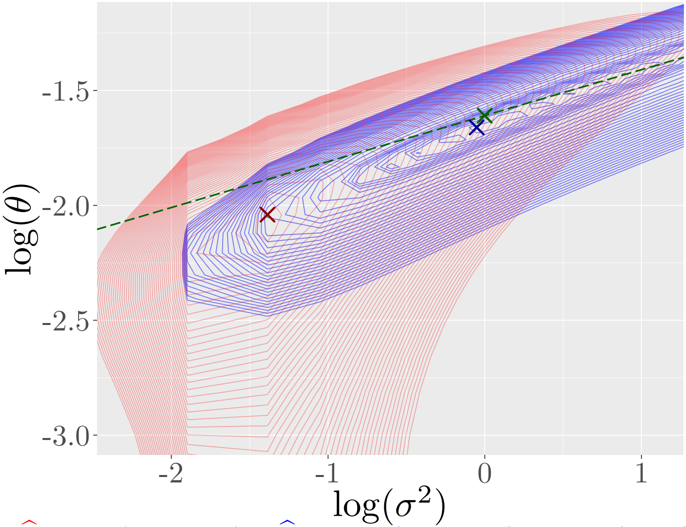
Results: simulation study

Estimation results obtained for each realization.



Results: simulation study

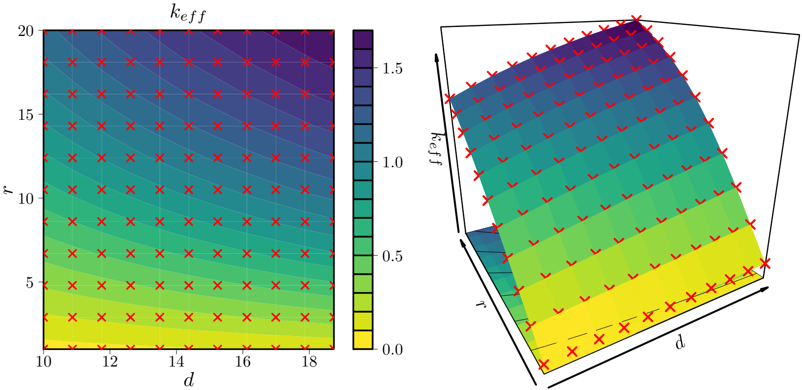
Estimation results obtained using all the realizations.



$\times\widehat{\boldsymbol{\theta}}_{\mathrm{MLE}} = (0.25, 0.13)$ $\times\widehat{\boldsymbol{\theta}}_{\mathrm{CMLE}} = (0.95, 0.19)$ $\times\boldsymbol{\theta}^* = (1, 0.2)$

Table of Contents

2D example (IRSN)



Nuclear criticality safety assessments: IRSN's dataset.

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

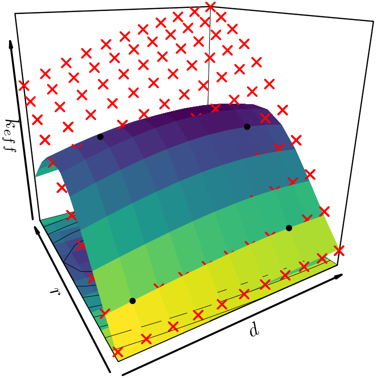
2D example (IRSN)

Procedure:

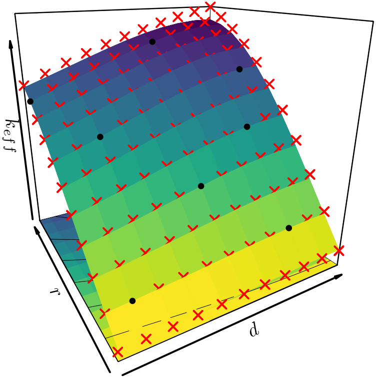
1. We used a Latin hypercube design (LHD) with different number of training points n.
2. We trained unconstrained and constrained models using either MLE or CMLE.
3. For the constrained models, we imposed both positivity and monotonicity constraints.
4. We evaluated their performances over the test points (i.e. $121 - n$).

2D example (IRSN)

Unconstrained model + MLE



(a)

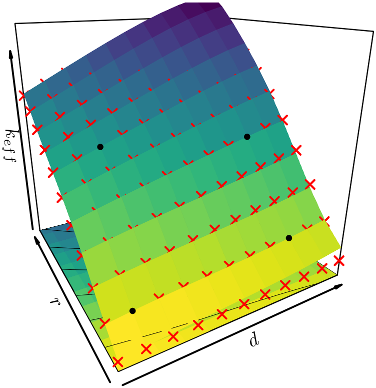


(b)

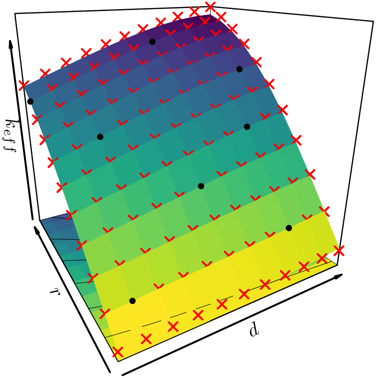
2D Gaussian models for interpolating the IRSN's dataset.

2D example (IRSN)

Constrained model + MLE



(c)

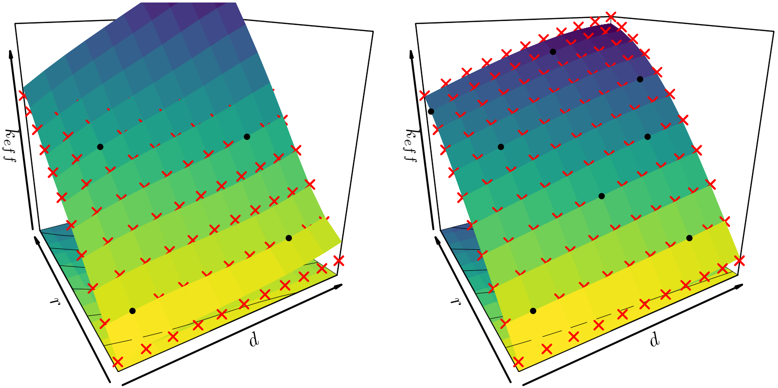


(d)

2D Gaussian models for interpolating the IRSN's dataset.

2D example (IRSN)

Constrained model + CMLE



(e) (f)

2D Gaussian models for interpolating the IRSN's dataset.

2D example (IRSN)

Now, we repeat the procedure for 20 random LHDs, and we compute the Q^2 and predictive variance adequation (PVA) criteria...

2D example (IRSN)

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:

$$Q^2 = 1 - \frac{\sum_{i=1}^{n_t} (\widehat{z}_i - z_i)^2}{\sum_{i=1}^{n_t} (\overline{z} - z_i)^2}, \tag{12}$$

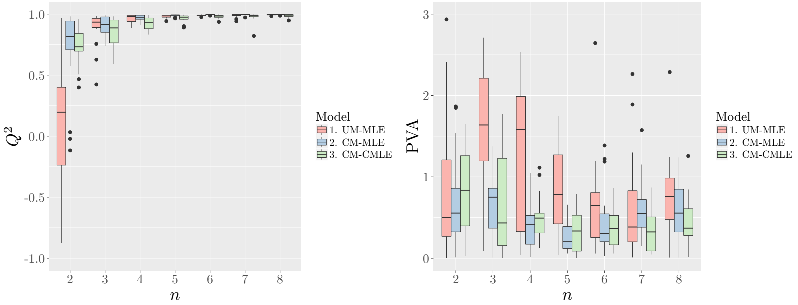
where $\overline{z}$ is the mean of the test data. $\Rightarrow Q^2 \to 1$ ✓

Predictive variance adequation (PVA) criterion:

$$\mathrm{PVA} = \left| \log \left(\frac{1}{n_t} \sum_{i=1}^{n_t} \frac{(\widehat{z}_i - z_i)^2}{\widehat{\sigma}_i^2} \right) \right|, \tag{13}$$

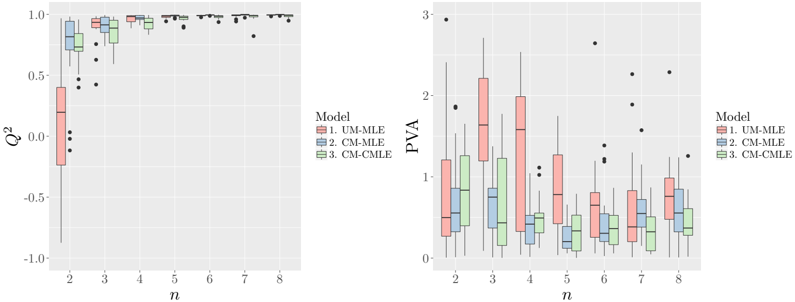
where $\widehat{\sigma}_i^2$ are the predictive variances. $\Rightarrow \mathrm{PVA} \to 0$ ✓

2D example (IRSN)



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

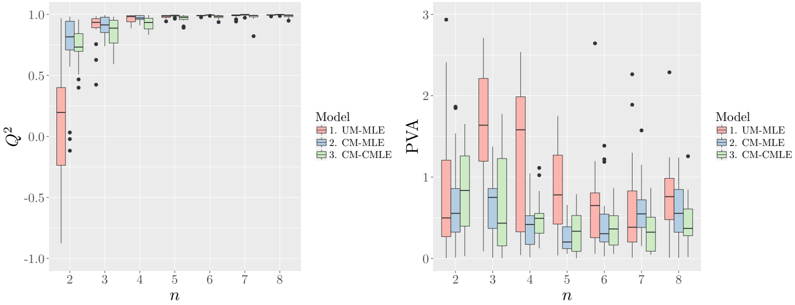
2D example (IRSN)



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.

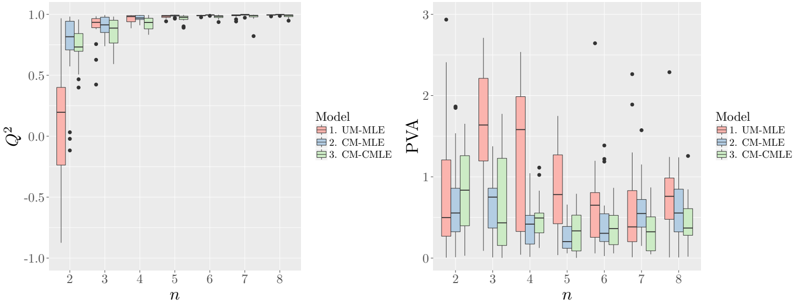
2D example (IRSN)



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 achieves a good tradeoff between prediction accuracy and computational cost.

2D example (IRSN)



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 achieves a good tradeoff between prediction accuracy and computational cost.
$\Rightarrow$ CMLE is unstable due to numerical approximations.

Table of Contents

MINES Saint-Étienne

INSTITUT DE MATHÉMATIQUES DE TOULOUSE

Conclusions and Future Works

Conclusions

- We further investigated the approach proposed in (Maatouk and Bay, 2017): now it works for any linear set of inequality constraints in 1D or 2D.
- We proved the consistency of the constrained likelihood for covariance parameter estimation.
- We implemented the R codes: `lineqGP` package.

Conclusions and Future Works

Conclusions

- We further investigated the approach proposed in (Maatouk and Bay, 2017): now it works for any linear set of inequality constraints in 1D or 2D.
- We proved the consistency of the constrained likelihood for covariance parameter estimation.
- We implemented the R codes: `lineqGP` package.

✦ Working paper:

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

MINES Saint-Étienne

Conclusions and Future Works

Future works

- To find an efficient and more reliable estimator of orthant multinormal distributions

$$\mathcal{L}_{\mathcal{C},m}(\boldsymbol{\theta}) = \log p_{\boldsymbol{\theta}}(\mathbf{Y}_m) + \log P_{\boldsymbol{\theta}}(\boldsymbol{\xi} \in \mathcal{C} | \boldsymbol{\Phi}\boldsymbol{\xi} = \mathbf{Y}_m) - \log P_{\boldsymbol{\theta}}(\boldsymbol{\xi} \in \mathcal{C}).$$

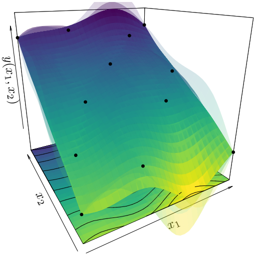
Conclusions and Future Works

Future works

- To find an efficient and more reliable estimator of orthant multinormal distributions

$$\mathcal{L}_{\mathcal{C},m}(\boldsymbol{\theta}) = \log p_{\boldsymbol{\theta}}(\mathbf{Y}_m) + \log P_{\boldsymbol{\theta}}(\boldsymbol{\xi} \in \mathcal{C} | \boldsymbol{\Phi}\boldsymbol{\xi} = \mathbf{Y}_m) - \log P_{\boldsymbol{\theta}}(\boldsymbol{\xi} \in \mathcal{C}).$$

- To work in the full framework for higher dimensions...



For example, in multidimensional problems with specific constrained dimensions.

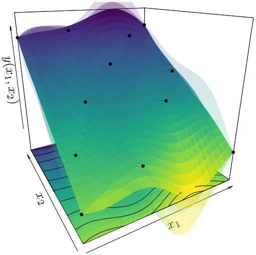
Conclusions and Future Works

Future works

- To find an efficient and more reliable estimator of orthant multinormal distributions

$$\mathcal{L}_{\mathcal{C},m}(\boldsymbol{\theta}) = \log p_{\boldsymbol{\theta}}(\mathbf{Y}_m) + \log P_{\boldsymbol{\theta}}(\boldsymbol{\xi} \in \mathcal{C} | \boldsymbol{\Phi}\boldsymbol{\xi} = \mathbf{Y}_m) - \log P_{\boldsymbol{\theta}}(\boldsymbol{\xi} \in \mathcal{C}).$$

- To work in the full framework for higher dimensions...



For example, in multidimensional problems with specific constrained dimensions.

- To study more asymptotic properties of the proposed framework.

Acknowledgement

We thank Yann Richet (IRSN) for providing the nuclear criticality safety data. This work is funded by the chair of applied mathematics OQUAIDO (Optimisation et QUAntification d'Incertitude pour les Données Onéreuses).

Table of Contents

Table of Contents

References I

Altmann, Y., McLaughlin, S., and Dobigeon, N. (2014). Sampling from a multivariate Gaussian distribution truncated on a simplex: A review. In *2014 IEEE Workshop on Statistical Signal Processing (SSP)*, pages 113–116.

Anderes, E. (2010). On the consistent separation of scale and variance for Gaussian random fields. *Ann. Statist.*, 38(2):870–893.

Azaïs, J.-M. and Wschebor, M. (2009). *Level sets and extrema of random processes and fields*. John Wiley & Sons.

Azzimonti, D. and Ginsbourger, D. (2017). Estimating orthant probabilities of high dimensional Gaussian vectors with an application to set estimation (to appear). *Journal of Computational and Graphical Statistics*.

Bachoc, F. (2013). Cross validation and maximum likelihood estimations of hyper-parameters of Gaussian processes with model misspecification. *Computational Statistics & Data Analysis*, 66:55 – 69.

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.

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.

Bect, J., Bachoc, F., and Ginsbourger, D. (2016). A supermartingale approach to Gaussian process based sequential design of experiments. *ArXiv e-prints*.

Betancourt, M. (2017). A Conceptual Introduction to Hamiltonian Monte Carlo. *ArXiv e-prints*.

References II

Bishop, C. M. (2007). *tern Recognition And Machine Learning (Information Science And Statistics)*. Springer.

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.

Brogan, W. L. (1991). *Modern Control Theory (3rd Ed.)*. Prentice-Hall, Inc., Upper Saddle River, NJ, USA.

Brooks, S., Gelman, A., Jones, G., and Meng, X. (2011). *Handbook of Markov Chain Monte Carlo*. Chapman & Hall/CRC Handbooks of Modern Statistical Methods. CRC Press.

Cousin, A., Maatouk, H., and Rullière, D. (2016). Kriging of financial term-structures. *European Journal of Operational Research*, 255(2):631–648.

Cramér, H. and Leadbetter, M. (1967). *Stationary and related stochastic processes: sample function properties and their applications*. Wiley series in probability and mathematical statistics. Tracts on probability and statistics. Wiley.

Da Veiga, S. and Marrel, A. (2012). Gaussian process modeling with inequality constraints. *Annales de la faculté des sciences de Toulouse Mathématiques*, 21(3):529–555.

Du, J., Zhang, H., and Mandrekar, V. S. (2009). Fixed-domain asymptotic properties of tapered maximum likelihood estimators. *Ann. Statist.*, 37(6A):3330–3361.

Duane, S., Kennedy, A. D., Pendleton, B. J., and Roweth, D. (1987). Hybrid Monte Carlo. *Physics letters B*, 195(2):216–222.

References III

Fernex, F., Heulers, L., Jacquet, O., Miss, J., and Richet, Y. (2005). The MORET 4B Monte Carlo code - New features to treat complex criticality systems. In *M&C international conference on mathematics and computation supercomputing, reactor physics and nuclear and biological application, Avignon, France*, volume 59.

Genton, M. G. (2001). Classes of kernels for machine learning: A statistics perspective. *Journal of Machine Learning Research*, 2:299–312.

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

Genz, A., Bretz, F., Miwa, T., and Mi, X. (2017). mvtnorm: Multivariate normal and t distributions. `https://cran.r-project.org/web/packages/mvtnorm/mvtnorm.pdf`. [Online; accessed 04-April-2017].

Golchi, S., Bingham, D. R., Chipman, H., and Campbell, D. A. (2015). Monotone emulation of computer experiments. *SIAM/ASA Journal on Uncertainty Quantification*, 3(1):370–392.

Gong, L. and Flegal, J. M. (2016). A practical sequential stopping rule for high-dimensional Markov Chain Monte Carlo. *Journal of Computational and Graphical Statistics*, 25(3):684–700.

Gramacy, R. B. and Apley, D. W. (2015). Local Gaussian process approximation for large computer experiments. *Journal of Computational and Graphical Statistics*, 24(2):561–578.

Hensman, J., Matthews, A. G. d. G., Filippone, M., and Ghahramani, Z. (2015). MCMC for variationally sparse Gaussian processes. In *NIPS 2015, 29th Annual Conference on Neural Information Processing Systems, December 7-12, 2015, Montreal, Quebec, Canada / Also on ArXiv*, Montreal, CANADA.

References IV

Holsclaw, T., Sansó, B., Lee, H. K. H., Heitmann, K., Habib, S., Higdon, D., and Alam, U. (2013). Gaussian process modeling of derivative curves. *Technometrics*, 55(1):57–67.

IAEA (2014). *Criticality Safety in the Handling of Fissile Material: Specific Safety Guide*. International Atomic Energy Agency (IAEA) Safety Standards Series No. SSG-27.

Kallenberg, O. (2002). *Foundations of Modern Probability*. Probability and its Applications. Springer New York.

Kim, Y., Kim, B., and Jang, W. (2010). Asymptotic properties of the maximum likelihood estimator for the proportional hazards model with doubly censored data. *Journal of Multivariate Analysis*, 101(6):1339 – 1351.

Kocijan, J. (2016). *Modelling and Control of Dynamic Systems Using Gaussian Process Models*. Advances in Industrial Control. Springer International Publishing, 1 edition.

Lan, S. and Shahbaba, B. (2016). *Sampling Constrained Probability Distributions Using Spherical Augmentation*, pages 25–71. Springer International Publishing, Cham.

L'Ecuyer, P., Blanchet, J. H., Tuffin, B., and Glynn, P. W. (2010). Asymptotic robustness of estimators in rare-event simulation. *ACM Trans. Model. Comput. Simul.*, 20(1):6:1–6:41.

Li, W. V. and Linde, W. (1999). Approximation, metric entropy and small ball estimates for Gaussian measures. *Ann. Probab.*, 27(3):1556–1578.

Liu, J. S. (2008). *Monte Carlo Strategies in Scientific Computing*. Springer Publishing Company, Incorporated.

Loh, W. L. (2005). Fixed-domain asymptotics for a subclass of Matérn-type Gaussian random fields. *Ann. Statist.*, 33(5):2344–2394.

References V

Loh, W. L. and Lam, T. K. (2000). Estimating structured correlation matrices in smooth Gaussian random field models. *Ann. Statist.*, 28(3):880–904.

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. (2016). *A New Rejection Sampling Method for Truncated Multivariate Gaussian Random Variables Restricted to Convex Sets*, pages 521–530. Springer International Publishing, Cham.

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

Maatouk, H., Bay, X., and Grammont, L. (2014). Simulation of Gaussian processes with interpolation and inequality constraints - A correspondence with optimal smoothing splines. In *GRF - Sim workshop : Simulation of Gaussian and related Random Fields*, Bern, Switzerland.

Maatouk, H., Roustant, O., and Richet, Y. (2015). Cross-validation estimations of hyper-parameters of Gaussian processes with inequality constraints. *Procedia Environmental Sciences*, 27:38–44.

Meyer, C. D. (2000). *Matrix Analysis and Applied Linear Algebra*. Society for Industrial and Applied Mathematics, Philadelphia, PA, USA.

Murphy, K. P. (2012). *Machine Learning: A Probabilistic Perspective (Adaptive Computation And Machine Learning Series)*. The MIT Press.

Neal, R. M. (1996). *Bayesian Learning for Neural Networks*. Springer-Verlag New York, Inc., Secaucus, NJ, USA.

Neal, R. M. (2012). MCMC using Hamiltonian dynamics. *ArXiv e-prints*.

References VI

Nickisch, H. and Rasmussen, C. (2008). Approximations for binary Gaussian process classification. *Journal of Machine Learning Research*, 9:2035–2078.

Paciorek, C. J. and Schervish, M. J. (2004). Nonstationary covariance functions for Gaussian process regression. In *In Proc. of the Conf. on Neural Information Processing Systems*. MIT Press.

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.

Riihimäki, J. and Vehtari, A. (2010). Gaussian processes with monotonicity information. In *Journal of Machine Learning Research: Workshop and Conference Proceedings*, volume 9, pages 645–652.

Ripley, B. D. (1987). *Stochastic Simulation*. John Wiley & Sons, Inc., New York, NY, USA.

Robbins, H. and Monro, S. (1951). A stochastic approximation method. *Ann. Math. Statist.*, 22(3):400–407.

Rogozhnikov, A. (2016). Brilliantly wrong thoughts on science and programming: Hamiltonian Monte Carlo explained. `http://arogozhnikov.github.io/2016/12/19/markov_chain_monte_carlo.html`. [Online; accessed 04-April-2017].

Snelson, E. and Ghahramani, Z. (2006). Sparse Gaussian processes using pseudo-inputs. In Weiss, Y., Schölkopf, P. B., and Platt, J. C., editors, *Advances in Neural Information Processing Systems 18*, pages 1257–1264. MIT Press.

References VII

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

Taylor, J. and Benjamini, Y. (2017). RestrictedMVN: multivariate normal restricted by affine constraints. `https://cran.r-project.org/web/packages/restrictedMVN/index.html`. [Online; 02-Feb-2017].

Tolvanen, V. (2014). Gaussian processes with monotonicity constraint for Big Data. Master's thesis, School of Electrical Engineering, Aalto University.

Vats, D., Flegal, J. M., and Jones, G. L. (2017). Multivariate output analysis for Markov Chain Monte Carlo. *ArXiv e-prints*.

Yadrenko, M. I. (1983). *Spectral theory of random fields*. Translation series in mathematics and engineering. Optimization Software, New York, NY. Transl. from the Russian.

Ying, Z. (1993). Maximum likelihood estimation of parameters under a spatial sampling scheme. *Ann. Statist.*, 21(3):1567–1590.

Yoo, E.-H. and Kyriakidis, P. C. (2006). Area-to-point kriging with inequality-type data. *Journal of Geographical Systems*, 8(4):357–390.

Zhang, H. (2004). Inconsistent estimation and asymptotically equal interpolations in model-based geostatistics. *Journal of the American Statistical Association*, 99(465):250–261.