Kriging under Inequality Constraints

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

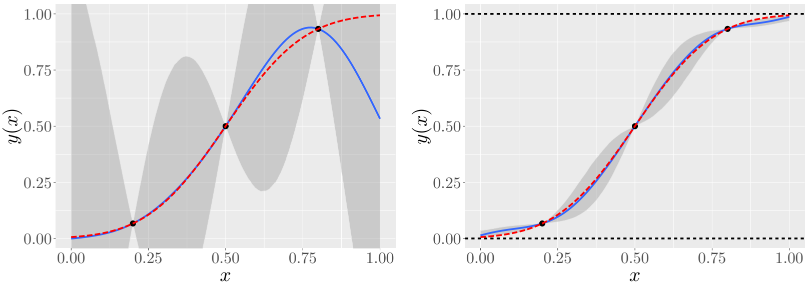
[1]École des Mines de Saint-Étienne (EMSE), France.
[2]Institut de Mathématiques de Toulouse (IMT), France.
[3]PROWLER.io, Cambridge, UK.

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

May 23, 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

MINES Saint-Étienne

Table of Contents

MINES Saint-Étienne

INSTITUT DE MATHÉMATIQUES DE TOULOUSE

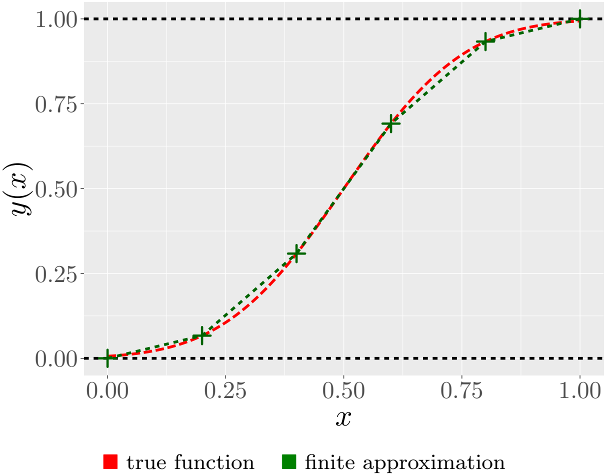
Table of Contents

MINES Saint-Étienne

INSTITUT DE MATHÉMATIQUES DE TOULOUSE

Finite-dimensional Gaussian approximation

Finite representation: also bounded and monotonic.



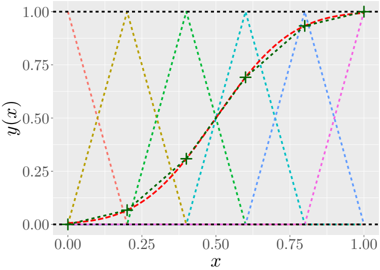
$\Rightarrow$ Imposing the inequality constraints on the knots is enough.

Finite-dimensional Gaussian approximation

Let the finite-dimensional GP approximation be defined as

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

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 (see López-Lopera et al. (2017)):

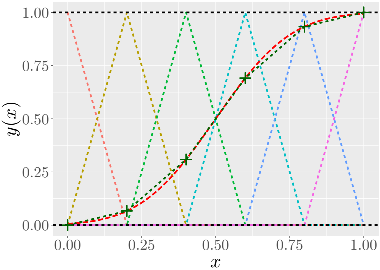


Finite-dimensional Gaussian approximation

Let the finite-dimensional GP approximation be defined as

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

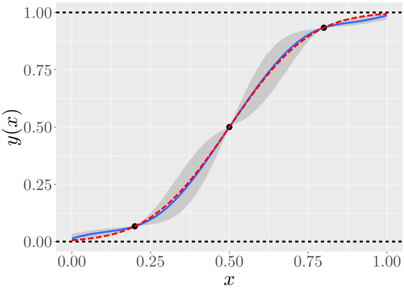
where $\boldsymbol{\xi} = [\xi_1, \; \cdots, \; \xi_m]^\top \sim \mathcal{N}(\mathbf{0}, \mathbf{\Gamma})$, with covariance matrix $\mathbf{\Gamma}$ and $\phi_j : [0,1] \to \mathbb{R}$ are hat functions (see López-Lopera et al. (2017)):



✦ 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).

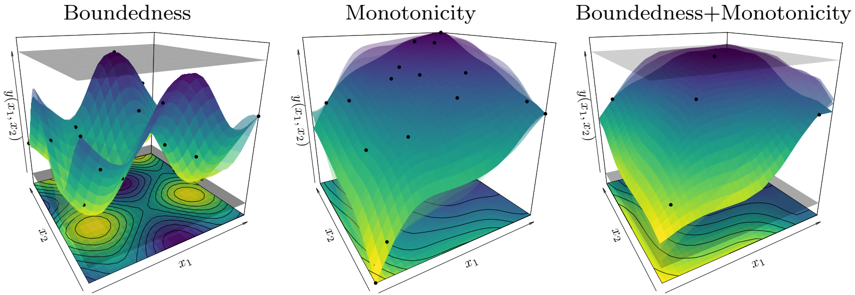
Finite-dimensional Gaussian approximation

1D example under boundedness and monotonicity constraints



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

Finite-dimensional Gaussian approximation



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

Table of Contents

- 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 $\mathcal{E}_\kappa$ 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 consider $\boldsymbol{\theta}_0$ as the true unknown covariance parameters.

Constrained maximum likelihood (CML)

- Let $\mathcal{L}_n(\boldsymbol{\theta})$ be the unconstrained log-likelihood given by

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

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

- Then, the constrained log-likelihood $\mathcal{L}_{n,c}(\boldsymbol{\theta})$ is defined by

$$\mathcal{L}_{n,c}(\boldsymbol{\theta}) = \mathcal{L}_n(\boldsymbol{\theta}) + \log P_{\boldsymbol{\theta}}(Y \in \mathcal{E}_\kappa | \mathbf{Y}_n) - \log P_{\boldsymbol{\theta}}(Y \in \mathcal{E}_\kappa),$$

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

Constrained maximum likelihood (CML)

- Let $\mathcal{L}_n(\boldsymbol{\theta})$ be the unconstrained log-likelihood given by

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

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

- Then, the constrained log-likelihood $\mathcal{L}_{n,c}(\boldsymbol{\theta})$ is defined by

$$\mathcal{L}_{n,c}(\boldsymbol{\theta}) = \mathcal{L}_n(\boldsymbol{\theta}) + \log P_{\boldsymbol{\theta}}(Y \in \mathcal{E}_\kappa | \mathbf{Y}_n) - \log P_{\boldsymbol{\theta}}(Y \in \mathcal{E}_\kappa),$$

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

✦ We showed that, loosely speaking, any consistency for ML, is preserved when adding boundedness, monotonicity and convexity constraints (see López-Lopera et al. (2017)).

MINES Saint-Étienne

INSTITUT DE MATHÉMATIQUES DE TOULOUSE

Partial contributions (papers, conferences, ...)

ArXiv preprints

- A.F. López-Lopera, F. Bachoc, N. Durrande, and O. Roustant (2017). *Finite-dimensional Gaussian approximation with linear inequality constraints* (in revision for SIAM JUQ).
- F. Bachoc, A. Lagnoux, and A.F. López-Lopera (2018). *Maximum likelihood estimation for Gaussian processes under inequality constraints* (submitted).

Conferences and talks

- A.F. López-Lopera, F. Bachoc, N. Durrande, and O. Roustant, *Gaussian process regression models under linear inequality conditions*, Mascot-Num 2018, Nantes, France, March 21-23, 2018.
- ——, *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 (free available in June/July!!).

Proposed future works

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

- To study more asymptotic properties of the cMLE.
- To scale our framework to higher dimensions.

Proposed future works

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

- To study more asymptotic properties of the cMLE. ✓
  - 4 visits to the IMT, Toulouse, France.
- To scale our framework to higher dimensions. **(in progress...)**
  - 1 visit to Prowler.io, Cambridge, UK.

Table of Contents

Table of Contents

MINES Saint-Étienne

INSTITUT DE MATHÉMATIQUES DE TOULOUSE

Notation.

- We use the same notation as before.
- For instance, we will focus on the estimation of a single variance parameter σ^2, i.e.

$$k_{\sigma^2}(x, x') = \sigma^2 k_1(x, x'),$$

  with k_1 a fixed known correlation function.
- We consider σ_0^2 as the true unknown variance parameter.
- We denote $\widehat{\sigma}_n^2$ and $\widehat{\sigma}_{n,c}^2$ the MLE and cMLE, respectively.

Theorem (Asymptotic normality of MLE, Bachoc et al. (2018))

Assume mild conditions (see Bachoc et al. (2018)). The MLE $\widehat{\sigma}_n^2$ of σ_0^2 conditioned on $\{Y \in \mathcal{E}_\kappa\}$ is asymptotically Gaussian distributed. More precisely, for $\kappa = 0, 1, 2$,

$$\sqrt{n}\left(\widehat{\sigma}_n^2 - \sigma_0^2\right) \xrightarrow[n \to +\infty]{\mathcal{L}|Y \in \mathcal{E}_\kappa} \mathcal{N}(0, 2\sigma_0^4).$$

Theorem (Asymptotic normality of cMLE, Bachoc et al. (2018))

The cMLE $\widehat{\sigma}_{n,c}^2$ of σ_0^2 is asymptotically Gaussian distributed. More precisely, for $\kappa = 0, 1, 2$,

$$\sqrt{n}\left(\widehat{\sigma}_{n,c}^2 - \sigma_0^2\right) \xrightarrow[n \to +\infty]{\mathcal{L}|Y \in \mathcal{E}_\kappa} \mathcal{N}(0, 2\sigma_0^4).$$

Theorem (Asymptotic normality of MLE, Bachoc et al. (2018))

Assume mild conditions (see Bachoc et al. (2018)). The MLE $\widehat{\sigma}_n^2$ of σ_0^2 conditioned on $\{Y \in \mathcal{E}_\kappa\}$ is asymptotically Gaussian distributed. More precisely, for $\kappa = 0, 1, 2$,

$$\sqrt{n}\left(\widehat{\sigma}_n^2 - \sigma_0^2\right) \xrightarrow[n \to +\infty]{\mathcal{L}|Y \in \mathcal{E}_\kappa} \mathcal{N}(0, 2\sigma_0^4).$$

Theorem (Asymptotic normality of cMLE, Bachoc et al. (2018))

The cMLE $\widehat{\sigma}_{n,c}^2$ of σ_0^2 is asymptotically Gaussian distributed. More precisely, for $\kappa = 0, 1, 2$,

$$\sqrt{n}\left(\widehat{\sigma}_{n,c}^2 - \sigma_0^2\right) \xrightarrow[n \to +\infty]{\mathcal{L}|Y \in \mathcal{E}_\kappa} \mathcal{N}(0, 2\sigma_0^4).$$

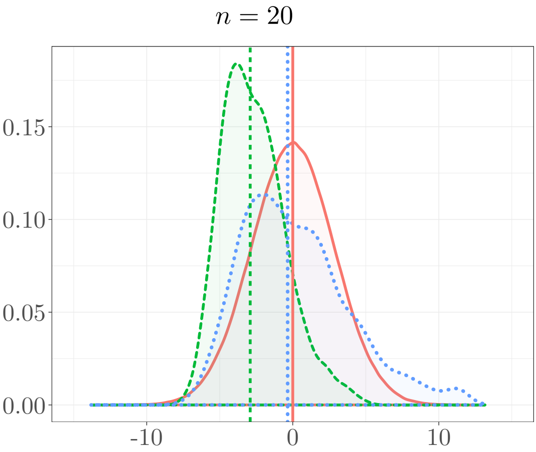
$\Rightarrow$ These results can be extended for the isotropic Matérn model (see Bachoc et al. (2018)).

Numerical illustration.

Numerical settings

- We use an isotropic Matérn 5/2 model with fixed correlation length $\rho = 0.2$ and true variance $\sigma_0^2 = 0.2$.
- We simulate $N = 1.000$ trajectories from the Gaussian approximation $Y_m \in [0,1]$ (boundedness constr.) with $m = 300$.
- For each realization, we estimate $\widehat{\sigma}_n^2$ and $\widehat{\sigma}_{n,c}^2$.

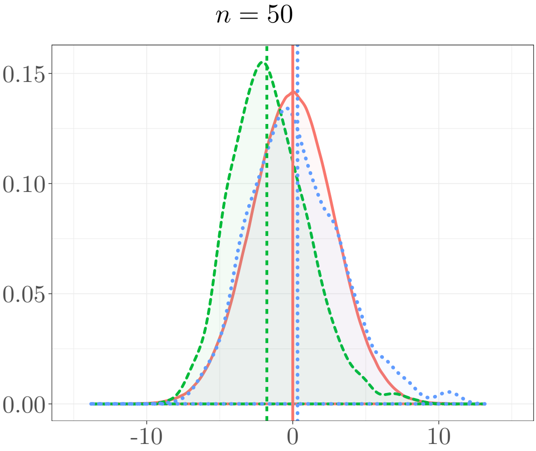
Numerical illustration.



■ limit dist. $\mathcal{N}\left(0, 2\sigma_0^4\right)$ ■ MLE: $\sqrt{n}\left(\widehat{\sigma}_n^2 - \sigma_0^2\right)$ ■ cMLE: $\sqrt{n}\left(\widehat{\sigma}_{n,c}^2 - \sigma_0^2\right)$

$\Rightarrow$ For small values of n, cMLE seems to be more accurate.

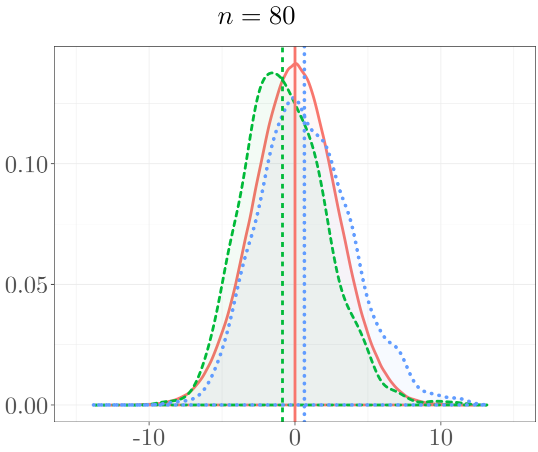
Numerical illustration.



■ limit dist. $\mathcal{N}\left(0, 2\sigma_0^4\right)$ ■ MLE: $\sqrt{n}\left(\widehat{\sigma}_n^2 - \sigma_0^2\right)$ ■ cMLE: $\sqrt{n}\left(\widehat{\sigma}_{n,c}^2 - \sigma_0^2\right)$

$\Rightarrow$ cMLE converges faster to the limit Gaussian distribution.

Numerical illustration.



limit dist. $\mathcal{N}\left(0, 2\sigma_0^4\right)$ MLE: $\sqrt{n}\left(\widehat{\sigma}_n^2 - \sigma_0^2\right)$ cMLE: $\sqrt{n}\left(\widehat{\sigma}_{n,c}^2 - \sigma_0^2\right)$

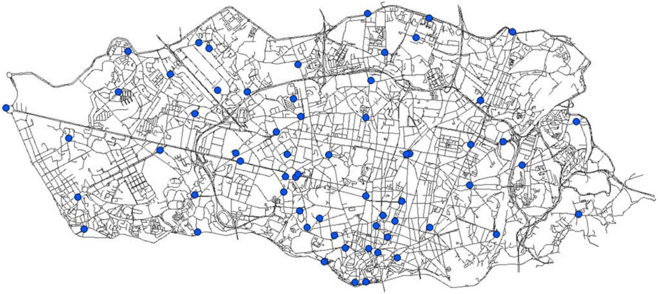
$\Rightarrow$ For large values of n , MLE and cMLE provide similar performances.

MINES Saint-Étienne

INSTITUT DE MATHÉMATIQUES DE TOULOUSE

Table of Contents

Motivation: Porto taxi pickups



Taxi-stand spatial distribution in Porto, Portugal (Moreira-Matias et al., 2013)

⇒ Aim: to model the taxi pickup rates.

⇒ Proposal: representing the pickup locations using Poisson processes.

Define an inhomogeneous Poisson process with events $\mathcal{D} = \{t_1, \cdots, t_N\}$ in the region τ to be given by

$$P(\mathcal{D}) = \exp\left\{ - \int_\tau \lambda(t) dt \right\} \prod_{n=1}^{N} \frac{\lambda(t_n)^{m_n}}{m_n!},$$

with strictly non-negative rate function λ, and where m_n denotes the multiplicity of events at location t_n.

1D Cox process models

A Cox process (CP) is an inhomogeneous Poisson process where λ is a non-negative random process (Cox, 1955). Assuming distinct events,

$$P(\mathcal{D}|\lambda) = \exp\left\{ -\int_{\mathcal{T}} \lambda(t)dt \right\} \prod_{n=1}^{N} \lambda(t_n).$$

1D Cox process models

A Cox process (CP) is an inhomogeneous Poisson process where λ is a non-negative random process (Cox, 1955). Assuming distinct events,

$$P(\mathcal{D}|\lambda) = \exp\left\{ -\int_{\mathcal{T}} \lambda(t)dt \right\} \prod_{n=1}^{N} \lambda(t_n).$$

The rate $\lambda \in \mathbb{R}^+$ can be modelled using GP models:

- log-Gaussian CPs (Møller et al., 2001),
- sigmoidal-Gaussian CPs (Adams et al., 2009),
- ...

But, these methods yield non-feasible inference of λ due to the intractable computation of the posterior distribution:

$$P(\lambda|\mathcal{D}) = \frac{P(\mathcal{D}|\lambda)P(\lambda)}{\int P(\mathcal{D}|\lambda)P(\lambda)d\lambda}$$

Following (López-Lopera et al., 2017), we will denote λ_m as the finite representation of λ at a set of given knots $u_1, \cdots, u_m$:

$$\lambda_m(t) = \sum_{j=1}^{m} \phi_j(t)\xi_j, \quad \text{s.t.} \quad \xi(u_j) \geq 0 \quad \text{(positiveness constraints)},$$

where $t \in \mathcal{T}$, $\boldsymbol{\xi} = [\xi_1, \ \cdots, \ \xi_m]^\top \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Gamma})$, with covariance matrix $\boldsymbol{\Gamma}$ and $\phi_j : [0, 1] \to \mathbb{R}$ are hat functions. Then, the likelihood is given by

$$p(\mathcal{D}|\lambda) \approx p(\mathcal{D}|\lambda_m) = \exp\left(- \sum_{j=1}^{m} c_j \xi(t_j) \right) \prod_{n} \sum_{j=1}^{m} \phi_j(t_n)\xi(u_j),$$

where $c_1 = c_m = \frac{\Delta_m}{2}$ and $c_j = \Delta_m$ for $j = 2, \cdots, m-1$.

Following (López-Lopera et al., 2017), we will denote λ_m as the finite representation of λ at a set of given knots $u_1, \cdots, u_m$:

$$\lambda_m(t) = \sum_{j=1}^{m} \phi_j(t)\xi_j, \quad \text{s.t.} \quad \xi(u_j) \geq 0 \quad \text{(positiveness constraints)},$$

where $t \in \mathcal{T}$, $\boldsymbol{\xi} = \left[\xi_1, \ \cdots, \ \xi_m\right]^\top \sim \mathcal{N}\left(\mathbf{0}, \boldsymbol{\Gamma}\right)$, with covariance matrix $\boldsymbol{\Gamma}$ and $\phi_j : [0,1] \to \mathbb{R}$ are hat functions. Then, the likelihood is given by

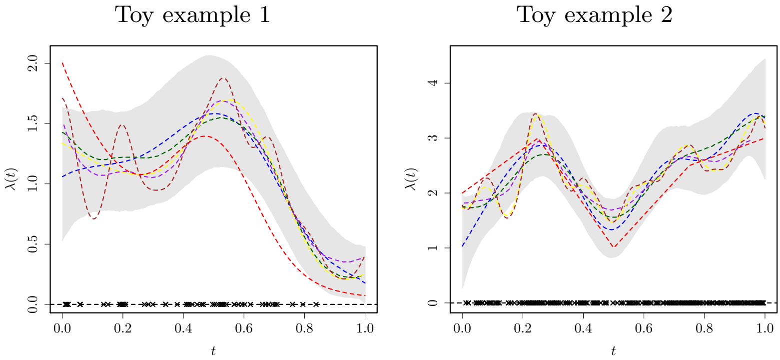
$$p(\mathcal{D}|\lambda) \approx p(\mathcal{D}|\lambda_m) = \exp\left(- \sum_{j=1}^{m} c_j \xi(t_j) \right) \prod_{n} \sum_{j=1}^{m} \phi_j(t_n)\xi(u_j),$$

where $c_1 = c_m = \frac{\Delta_m}{2}$ and $c_j = \Delta_m$ for $j = 2, \cdots, m-1$.

$\Rightarrow$ The likelihood can be simulated by sampling $\boldsymbol{\xi} \in \mathbb{R}^+$.

$\Rightarrow$ Furthermore, $P(\lambda_m|\mathcal{D})$ can be approximated via MCMC (e.g. Metropolis-Hastings algorithm) $\Rightarrow$ joint work in progress!

Synthetic examples from (Adams et al., 2009)



■ true rate ✗ events
■ log-Gaussian CP ■ sigmoidal-Gaussian CP
■ ■ variational Fourier features (John and Hensman, 2018)
■ (resp. ■) our approach (resp. confidence intervals)

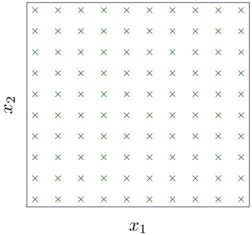
⇒ Our finite approximation is competitive with respect to the others!!

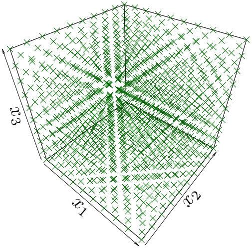
Table of Contents

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

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

but...





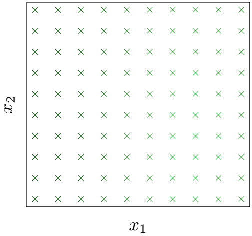
✗ Design of the knots by tensorisation (intractable in practice!!)

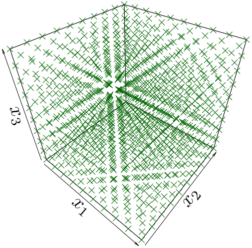
Course of dimensionality of the finite approximation

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

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

but...



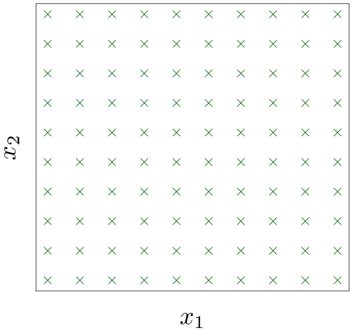


✗ Design of the knots by tensorisation (intractable in practice!!)

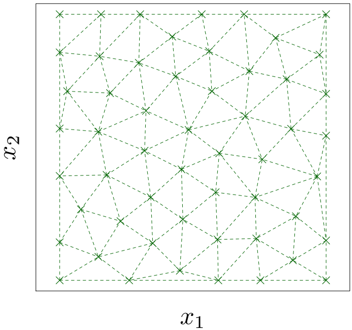
⇒ What about different designs?

- Triangular designs.
- Sparse Grids.

Table of Contents



$\Rightarrow$

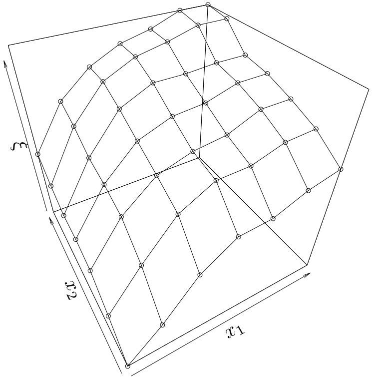


Design by tensorisation.

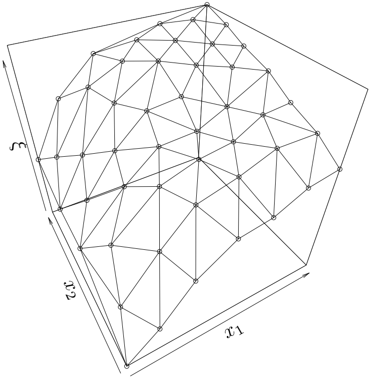
Design by Delaunay triangulation.

$\Rightarrow$ The triangulation allows a free location of knots: more suitable designs could be obtained by optimisation.

2D example under boundedness constraints



Design by tensorisation.



Design by Delaunay triangulation.

2D Gaussian approximation using triangular designs

Let 2D finite-dimensional Gaussian approximation given by:

$$Y_m(x_1, x_2) := \sum_{j=1}^{m} \xi_j \phi_j(x_1, x_2), \text{ s.t. } \begin{cases} Y_m\left(x_1^i, x_2^i\right) = y_i, \\ \xi_j \in \mathcal{C}, \end{cases}$$

where $\boldsymbol{\xi} = \left[\xi_1, \ \cdots, \ \xi_m\right]^\top \sim \mathcal{N}\left(\mathbf{0}, \boldsymbol{\Gamma}\right)$ with entries $\xi_j = Y(t_1^j, t_2^j)$, $(x_1^1, x_2^1), \cdots, (x_1^n, x_2^n)$ constitute a DoE, and ϕ_j are given by the barycentric coordinates of interpolation conditions $Y_m\left(x_1^i, x_2^i\right) = y_i$.

2D Gaussian approximation using triangular designs

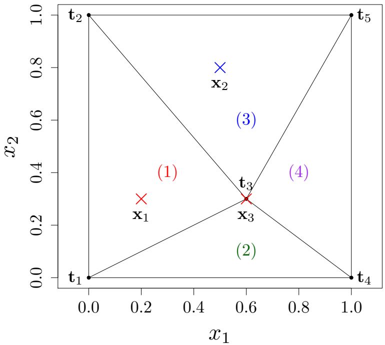
Let 2D finite-dimensional Gaussian approximation given by:

$$Y_m(x_1, x_2) := \sum_{j=1}^{m} \xi_j \phi_j(x_1, x_2), \text{ s.t. } \begin{cases} Y_m\left(x_1^i, x_2^i\right) = y_i, \\ \xi_j \in \mathcal{C}, \end{cases}$$

where $\boldsymbol{\xi} = \left[\xi_1, \ \cdots, \ \xi_m\right]^\top \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Gamma})$ with entries $\xi_j = Y(t_1^j, t_2^j)$, $(x_1^1, x_2^1), \cdots, (x_1^n, x_2^n)$ constitute a DoE, and ϕ_j are given by the barycentric coordinates of interpolation conditions $Y_m\left(x_1^i, x_2^i\right) = y_i$.

$+$ This model could scale better than the one obtained by tensorisation thanks to the triangular design (work in progress!).

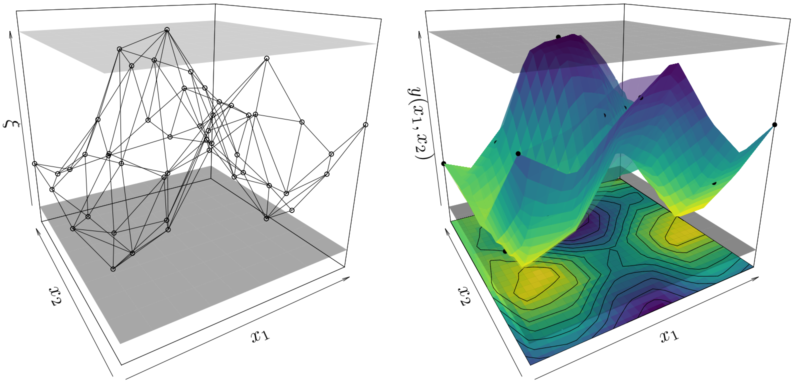
$-$ The representation of the linear inequality constraints is not always straightforward... but still possible!!



$$\begin{bmatrix} Y_m(x_1^1, x_2^1) \\ Y_m(x_1^2, x_2^2) \\ Y_m(x_1^3, x_2^3) \end{bmatrix} = \begin{bmatrix} \beta_1^{1,1} & \beta_2^{1,2} & \beta_3^{1,3} & 0 & 0 \\ 0 & \beta_1^{2,2} & \beta_2^{2,3} & 0 & \beta_3^{2,5} \\ 0 & 0 & 1 & 0 & 0 \end{bmatrix} \begin{bmatrix} \xi_1 \\ \xi_2 \\ \xi_3 \\ \xi_4 \\ \xi_5 \end{bmatrix},$$

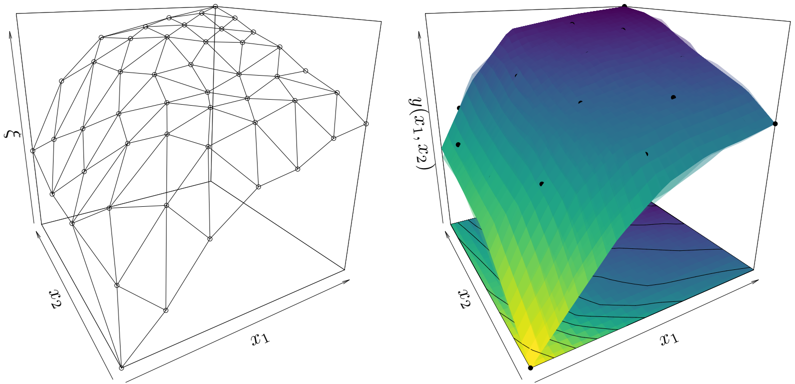
where $\beta_k^{i,1}, \beta_k^{i,2}, \beta_k^{i,3} \in [0,1]$ are the barycentric coordinates of the observation y_i w.r.t. the triangle k.

2D examples using triangular designs



2D example under boundedness constraints using Delaunay triangulation.

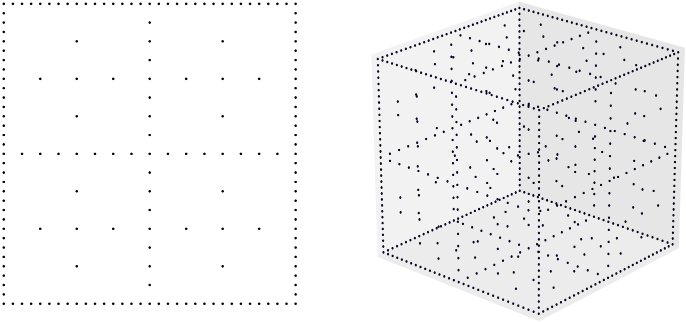
2D examples using triangular designs



2D example under monotonicity constraints using Delaunay triangulation.

Table of Contents

Sparse grids?

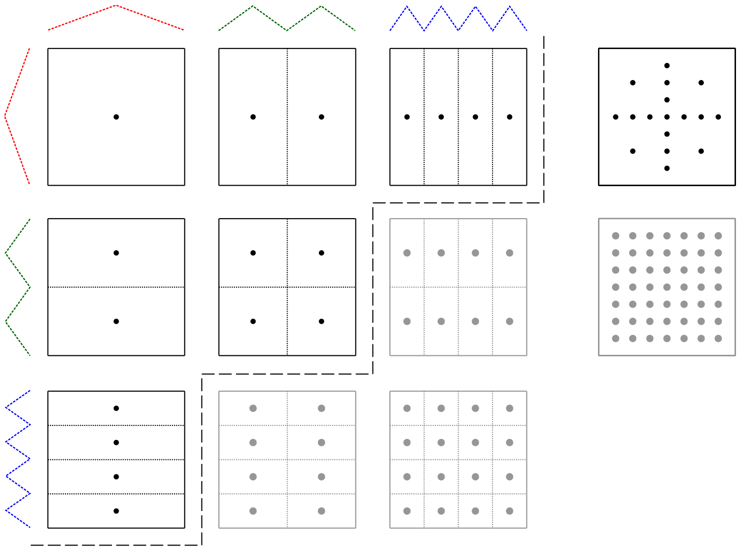


2D and 3D examples of sparse grids by (Garcke, 2013).

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

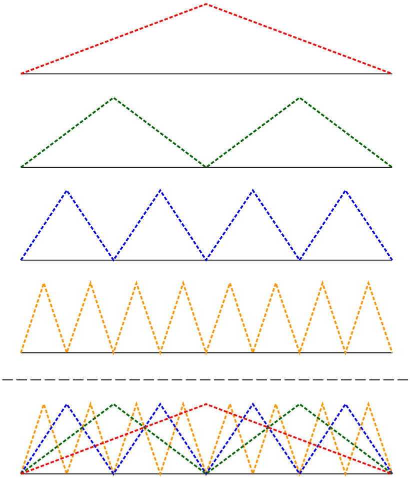
⇒ Numerical implementations are already available (e.g. SG++).

Sparse grids?



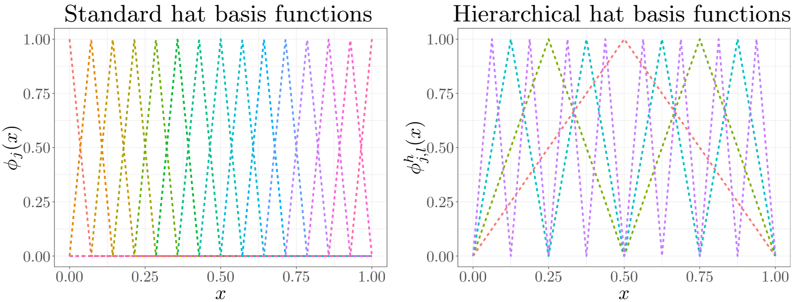
Construction of a sparse grid with level $L = 3$.

Sparse grids?



Construction of hierarchical hat basis functions with level $L = 4$.

Sparse grids?



+ Since we continue using hat functions, the properties of the finite-dimensional Gaussian approximation are preserved.

+ There exist algorithms for the automatic construction of the hierarchical hat basis functions (e.g. number of levels L).

Sparse grids?

Advantages

- Several tools have been proposed for sparse grids (e.g. SG++).
- There are works related to Gaussian processes regression models and sparse grids (Luo and Duraiswami, 2013; Plumlee, 2014).
- Sparse grids could help us to scale our framework to higher dimensions (e.g. 5 or 10 input spaces).

Table of Contents

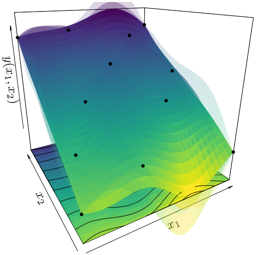
Conclusions and Future Works

Conclusions

- We proved the asymptotic normality of cMLE for:
  - variance parameter estimation,
  - parameter estimation for the isotropic Matérn model.
- We proposed a finite Gaussian approximation of 1D Cox processes.
- We investigated an alternative Gaussian approximation under inequality constraints using Delaunay triangular designs.
- Sparse grids?

Future works

- To scale the proposed framework to higher dimensions (e.g. using triangular designs, sparse grids, ...)



For instance, in multidimensional problems with specific constrained dimensions.

- To explore more asymptotic properties of cMLE?
- Open to suggestions :)

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.

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*.

References II

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. (2017). Finite-dimensional Gaussian approximation with linear inequality constraints. *ArXiv e-prints*.

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.

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

References III

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.