Gaussian processes with inequality constraints: theory and computation

François Bachoc[1], Nicolas Durrande[2], Agnès Lagnoux[1], Andrés Felipe Lopez Lopera[3], Olivier Roustant[1]

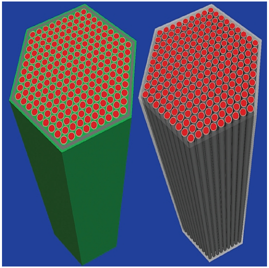
[1]Institut de Mathématiques de Toulouse
[2]Monumo (Cambridge)
[3]Université de Valenciennes

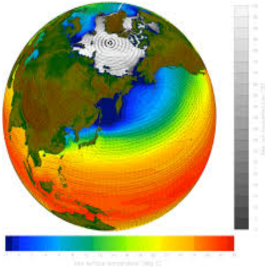
Rutgers Statistics Seminar

April 2024

Outline

Computer models have become essential in science and industry !





For clear reasons : cost reduction, possibility to explore hazardous or extreme scenarios...

Computer models as expensive functions

A computer model can be seen as a deterministic function

$$\begin{aligned} f : \mathbb{X} \subset \mathbb{R}^d &\to \mathbb{R} \\ x &\mapsto f(x). \end{aligned}$$

- x : tunable simulation parameter (e.g. geometry).
- $f(x)$: scalar quantity of interest (e.g. energetic efficiency).

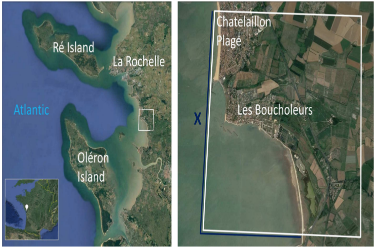
The function f is usually

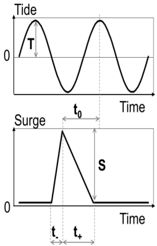
- continuous (at least)
- non-linear
- only available through evaluations $x \mapsto f(x)$.

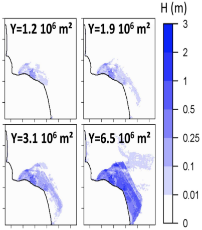
$\Longrightarrow$ Black box model.

Follow-along example : coastal flooding

Figures from [Azzimonti et al., 2019].
Hydrodynamic numerical simulations made by BRGM [Rohmer et al., 2018].





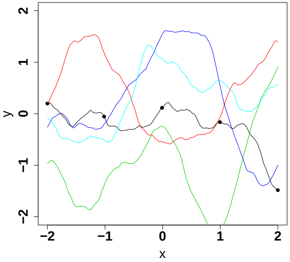


- Input x with $d = 5$.
  - : Tide (meter).
  - : Surge peak (meter).
  - : Phase difference between surge peak and high tide (hour).
  - : Time duration of raising part of surge (hour).
  - : Time duration of falling part of surge (hour).
- Output $f(x)$.
  - Maximal flooding area (m^2).
  - 1 hour simulation time.

Gaussian process

Gaussian processes (Kriging model)

Modeling the **black box function** as a **single realization** of a Gaussian process $x \to \xi(x)$ on the domain $\mathbb{X} \subset \mathbb{R}^d$.



Usefulness

Predicting the continuous realization function, from a finite number of **observation points**.

Remark : Gaussian processes are widely used in geostatistics as well.

Gaussian processes

Definition

A stochastic process $\xi : \mathbb{X} \to \mathbb{R}$ is Gaussian if for any $x_1, ..., x_n \in \mathbb{X}$, the vector $(\xi(x_1), ..., \xi(x_n))$ is a Gaussian vector.

Mean and covariance functions

The distribution of a Gaussian process is characterized by :

- Its mean function :
$$x \mapsto m(x) = \mathbb{E}(\xi(x)).$$
  - Can be any function $\mathbb{X} \to \mathbb{R}$.
  - Will be the zero function throughout this talk !
- Its covariance function :
$$(x_1, x_2) \mapsto k(x_1, x_2) = Cov(\xi(x_1), \xi(x_2)).$$
  - Must be symmetric non-negative definite (to provide “valid” covariance matrices).

Conditional distribution

Gaussian process ξ observed at $x_1, ..., x_n$, without noise.

Notation

- $y = (\xi(x_1), ..., \xi(x_n))^\top$.
- R is the $n \times n$ matrix $[k(x_i, x_j)]$.
- $r(x) = (k(x, x_1), ..., k(x, x_n))^\top$.

Conditional mean

The conditional mean is $m_n(x) = \mathbb{E}(\xi(x)|\xi(x_1), ..., \xi(x_n)) = r(x)^\top R^{-1} y$.

Conditional variance

The conditional variance is $k_n(x, x) = var(\xi(x)|\xi(x_1), ..., \xi(x_n)) =$
$\mathbb{E}\left[(\xi(x) - m_n(x))^2\right] = k(x, x) - r(x)^\top R^{-1} r(x)$.

Conditional distribution

Conditionally to $\xi(x_1), ..., \xi(x_n)$, ξ is a Gaussian process with (conditional) mean function m_n and (conditional) covariance function
$(u, v) \mapsto k_n(u, v) = k(u, v) - r(u)^\top R^{-1} r(v)$.

Illustration of conditional mean and variance

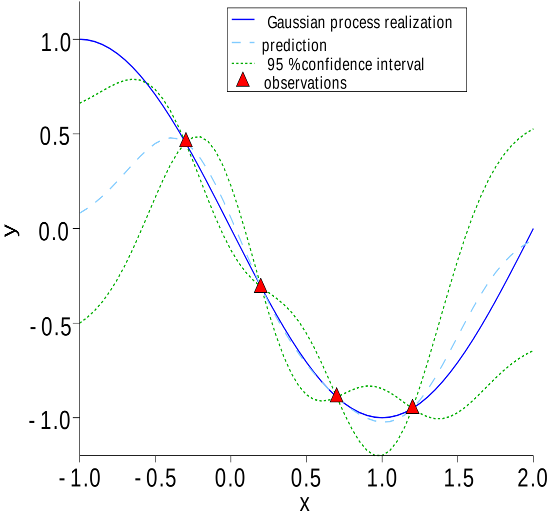
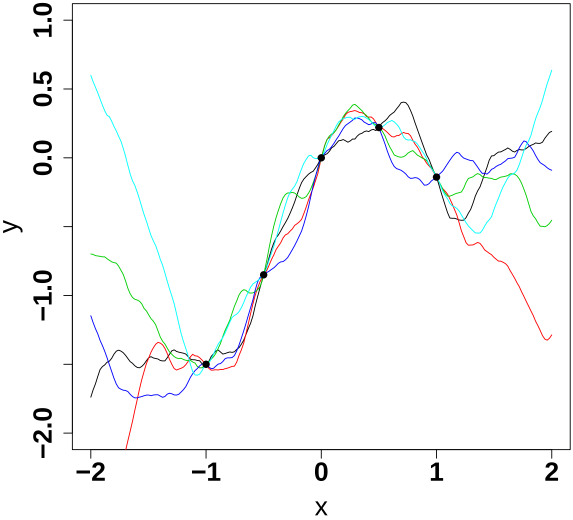


Illustration of the conditional distribution



Parametric covariance function estimation

Parameterization

Covariance function model $\{k_\theta, \theta \in \Theta\}$ for the Gaussian process ξ.

- $\Theta \subset \mathbb{R}^p$.
- θ is the multidimensional covariance parameter.
- k_θ is a covariance function.

Observations

ξ is observed at $x_1, ..., x_n \in \mathbb{X}$, yielding the Gaussian vector $y = (\xi(x_1), ..., \xi(x_n))^\top$.

Estimation

Objective : build estimator $\hat{\theta}(y)$.

Maximum likelihood (ML) for estimation

Explicit Gaussian likelihood function for the observation vector y.

Maximum likelihood

Define R_θ as the covariance matrix of $y = (\xi(x_1), ..., \xi(x_n))^\top$ with covariance function $k_\theta : R_\theta = [k_\theta(x_i, x_j)]_{i,j=1,\ldots,n}$.
The maximum likelihood estimator of θ is

$$\hat{\theta}_{ML} \in \underset{\theta \in \Theta}{\operatorname{argmax}} \, \mathcal{L}_n(\theta)$$

with

$$\mathcal{L}_n(\theta) = \log(p_\theta(y)) = \log\left(\frac{1}{(2\pi)^{n/2}|R_\theta|} e^{-\frac{1}{2} y^\top R_\theta^{-1} y}\right).$$

Outline

Inequality constraints

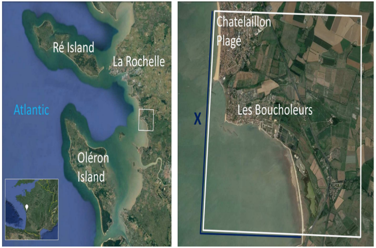
We consider a Gaussian process ξ on $\mathbb{X} = [0,1]^d$ for which we assume that additional information is available.

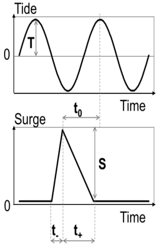
- $\xi(x)$ belongs to $[\ell, u]$ for $x \in [0,1]^d$ (**boundedness constraints**).
- $\partial\xi(x)/\partial x_i \geq 0$ for $x \in [0,1]^d$ and $i = 1, \ldots, d$ (**monotonicity constraints**).
- ξ is convex on $[0,1]^d$ (**convexity constraints**).
- Modifications and/or combinations of the above constraints.

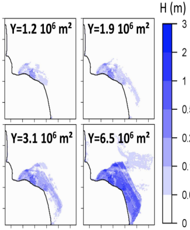
Application examples in computer experiments.

- **Boundedness :** computer model output belongs to $\mathbb{R}^+$ (energy) or $[0,1]$ (concentration, energetic efficiency).
- **Monotonicity :** inputs are known to have positive effects (more input power $\rightarrow$ more output energy).

Coastal flooding : the constraints







- Input x.
  - : Tide (meter). Output increases when tide increases !
  - : Surge peak (meter). Output increases when surge increases !
  - : Phase difference between surge peak and high tide (hours).
  - : Time duration of raising part of surge (hours).
  - : Time duration of falling part of surge (hours).
- Output $f(x)$.
  - Maximal flooding area (m^2).

Impact of the constraints

Generic form of the constraints :

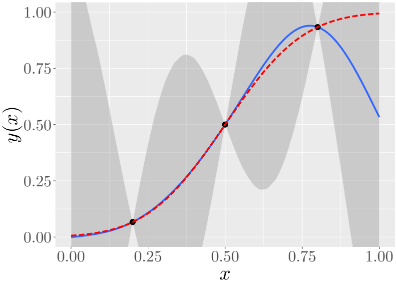
$$\xi \in \mathcal{E}$$

where $\mathcal{E}$ is a set of functions from $[0,1]^d \to \mathbb{R}$ such that $P(\xi \in \mathcal{E}) > 0$.

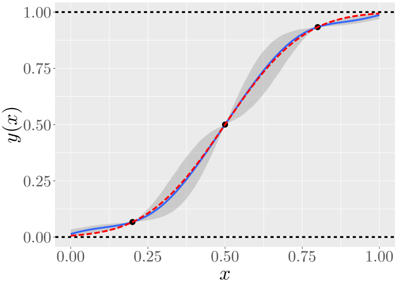
Impact.

- **New Bayesian model :** The prior on the realization function is $P(\xi \in .|\xi \in \mathcal{E})$.
- **New conditional distribution :** Conditional distribution of ξ given
  - $\xi(x_1) = y_1, \dots, \xi(x_n) = y_n$ (data interpolation),
  - $\xi \in \mathcal{E}$ (inequality constraints).
- **New estimation** of the covariance parameters θ in the covariance model $\{k_\theta; \theta \in \Theta\}$.

Target function : bounded and monotonic.



Unconstrained Gaussian process.

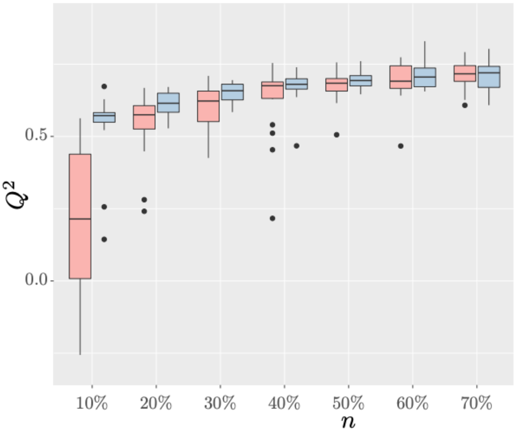


Constrained Gaussian process.

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

Gaussian process predictive score.

- Without constraints.
- With constraints.



The Q^2 (≤ 1) measures the prediction quality,

- $Q^2 = 1$: perfect prediction,
- $Q^2 = 0$: no better than constant prediction.

An application to nuclear engineering

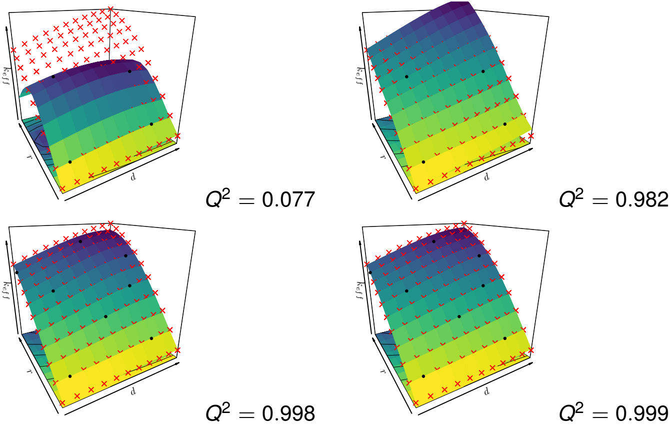


Figure – Two-dimensional nuclear engineering example.
Radius and **density** of uranium sphere $\Longrightarrow$ **criticality coefficient**.
Mononicity constraints.

- Left : unconstrained Gaussian process models.
- Right : constrained Gaussian process models.

Outline

Constrained maximum likelihood estimator

The constrained maximum likelihood estimator for θ is

$$\hat{\theta}_{cML} \in \underset{\theta \in \Theta}{\operatorname{argmax}} \mathcal{L}_{\mathcal{C},n}(\theta)$$

with

$$\begin{aligned} \mathcal{L}_{\mathcal{C},n}(\theta) &= \log(p_\theta(y|\xi \in \mathcal{E})) \\ &= \log(p_\theta(y)) - \log(\mathbb{P}_\theta(\xi \in \mathcal{E})) + \log(\mathbb{P}_\theta(\xi \in \mathcal{E}|y)). \end{aligned}$$

- The additional terms $\log(\mathbb{P}_\theta(\xi \in \mathcal{E}))$ and $\log(\mathbb{P}_\theta(\xi \in \mathcal{E}|y))$ have no explicit expressions.
- They need to be approximated by numerical integration or Monte Carlo : [Genz, 1992, Botev, 2017].
- We do not address this approximation issue in this theory section (see next computation section).

Main questions :

- $\hat{\theta}_{ML}$ ignores the constraints. Is it biased conditionally to the constraints ?
  - For instance if $\hat{\theta}_{ML}$ is the variance estimator, if the true variance is 4 and if the constraints are $\xi \in [-1, 1]$, does $\hat{\theta}_{ML}$ underestimate the variance ?
- Does $\hat{\theta}_{cML}$ improve over $\hat{\theta}_{ML}$ by taking the constraints into account ?

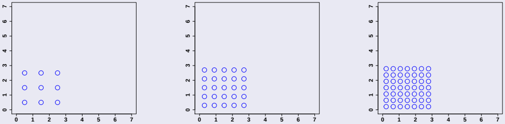
We address these questions **asymptotically**.

Fixed-domain asymptotics with fixed constraints

- Asymptotics (number of observations $n \to +\infty$) is an active area of research.
- Mostly without constraints.
- There are several asymptotic frameworks because there are several possible location patterns for the observation points.

Fixed-domain asymptotics

The observation points $x_1, \dots, x_n$ are dense in a bounded domain.



Fixed constraints

Fixed constraint set $\mathcal{E}$ with

$$\mathbb{P}(\xi \in \mathcal{E}) > 0.$$

Microergodic parameters

- Consistent estimation is impossible for some covariance parameters (identifiable in finite-sample), see e.g. [Zhang, 2004, Stein, 1999].
  - Covariance parameters that yield equivalent Gaussian measures are called non-microergodic. They cannot be estimated consistently.
  - Covariance parameters that yield orthogonal Gaussian measures are called microergodic. They can be estimated consistently.
- For instance, consider the set of covariance functions $\{k_\theta, \theta \in (0, \infty)^2\}$ on [0, 1] given by $\theta = (\sigma^2, \alpha)$ and $k_\theta(t_1, t_2) = \sigma^2 e^{-\alpha |t_1 - t_2|}$.
  - σ^2 is non-microergodic.
  - α is non-microergodic.
  - $\sigma^2 \alpha$ is microergodic.

Some initial properties

Let $\theta_0 \in \Theta$ such that $k = k_{\theta_0}$ (true covariance parameter).

- A non-microergodic parameter cannot be estimated consistently conditionally to the constraints.
  - Has a short proof using that $\mathbb{P}(\xi \in \mathcal{E}) > 0$ is fixed.
- If $\hat{\theta}_{ML} - \theta_0 = O_{\mathbb{P}}(n^{-1/2})$ then $\hat{\theta}_{ML} - \theta_0 = O_{\mathbb{P}|\xi \in \mathcal{E}}(n^{-1/2})$ which means

$$\limsup_{n\to\infty} \mathbb{P}\left(\sqrt{n} \|\hat{\theta}_{ML} - \theta_0\| \geq M \middle| \xi \in \mathcal{E} \right) \underset{M\to\infty}{\longrightarrow} 0.$$

  - Holds because

$$\begin{aligned}\mathbb{P}\left(\sqrt{n} \|\hat{\theta}_{ML} - \theta_0\| \geq M \middle| \xi \in \mathcal{E} \right) &= \frac{1}{\mathbb{P}(\xi \in \mathcal{E})} \mathbb{P}\left(\sqrt{n} \|\hat{\theta}_{ML} - \theta_0\| \geq M, \xi \in \mathcal{E} \right) \\ &\leq \frac{1}{\mathbb{P}(\xi \in \mathcal{E})} \mathbb{P}\left(\sqrt{n} \|\hat{\theta}_{ML} - \theta_0\| \geq M \right)\end{aligned}$$

  and $\mathbb{P}(\xi \in \mathcal{E}) > 0$ is fixed.

$\Longrightarrow$ Rate of convergence is preserved with constraints.
$\Longrightarrow$ What about asymptotic distribution ?

Asymptotic normality result 1 : variance estimation

Setting :

- Gaussian process ξ on $[0, 1]^d$.
- Monotonicity, boundedness or convexity constraints.
- Observation point sequence $(x_i)_{i \in \mathbb{N}}$ is dense in $[0, 1]^d$.
- $\theta = \sigma^2$ and $k_\theta(u_1, u_2) = \sigma^2 \tilde{k}(u_1, u_2)$, for some fixed $\tilde{k}$.
- True covariance function $k = \sigma_0^2 \tilde{k}$.

Asymptotic normality without constraints

- It is well-known that in this case

$$\sqrt{n} \left(\hat{\sigma}^2_{ML} - \sigma_0^2 \right) \xrightarrow[n \to \infty]{\mathcal{L}} \mathcal{N}(0, 2\sigma_0^4).$$

Asymptotic normality result 1 : variance estimation

Notation (convergence in distribution given the constraints) : we write

$$X_n \xrightarrow[n\to\infty]{\mathcal{L}|\xi\in\mathcal{E}} L$$

when for all bounded measurable function f :

$$\mathbb{E}(f(X_n)|\xi \in \mathcal{E}) \underset{n\to\infty}{\longrightarrow} \int f(x)dL(x).$$

Theorem [Bachoc et al., 2019]

Under technical conditions on k and the sequence $(x_i)_{i\in\mathbb{N}}$ (see paper), we have

$$\sqrt{n}\left(\hat{\sigma}^2_{ML} - \sigma_0^2\right) \xrightarrow[n\to\infty]{\mathcal{L}|\xi\in\mathcal{E}} \mathcal{N}(0, 2\sigma_0^4)$$

and

$$\sqrt{n}\left(\hat{\sigma}^2_{cML} - \sigma_0^2\right) \xrightarrow[n\to\infty]{\mathcal{L}|\xi\in\mathcal{E}} \mathcal{N}(0, 2\sigma_0^4).$$

- Same asymptotic distribution as the (unconstrained) maximum likelihood estimator, in the unconstrained case.
- No asymptotic impact of the constraints.

Asymptotic normality result 2 : Matérn model

Setting :

- Gaussian process ξ on $[0,1]^d$, $d = 1, 2, 3$, with covariance function k.
- Monotonicity, boundedness or convexity constraints.
- Observation point sequence $(x_i)_{i \in \mathbb{N}}$ is dense in $[0,1]^d$.
- $\theta = (\sigma^2, \rho) \in (0, \infty)^2$ and

$$k_{\theta,\nu}(x, x') = \sigma^2 K_\nu \left(\frac{\|x - x'\|}{\rho} \right) = \frac{\sigma^2}{\Gamma(\nu) 2^{\nu-1}} \left(\frac{\|x - x'\|}{\rho} \right)^\nu \kappa_\nu \left(\frac{\|x - x'\|}{\rho} \right).$$

  - Γ is the Gamma function.
  - κ_ν is the modified Bessel function of the second kind.
  - $\nu > 0$ (assumed known) is the smoothness parameter : $\nu > r \implies$ corresponding Gaussian process if r times differentiable.
- True covariance function $k = k_{\theta_0,\nu}$, $\theta_0 = (\sigma_0^2, \rho_0)$.

In this case :

- σ^2 is non-microergodic
- ρ is non-microergodic
- $\sigma^2 / \rho^{2\nu}$ is microergodic and

$$\sqrt{n} \left(\frac{\widehat{\sigma}^2_{ML}}{\widehat{\rho}^{2\nu}_{ML}} - \frac{\sigma_0^2}{\rho_0^{2\nu}} \right) \xrightarrow[n \to +\infty]{\mathcal{L}} \mathcal{N} \left(0, 2 \left(\frac{\sigma_0^2}{\rho_0^{2\nu}} \right)^2 \right).$$

This is shown in [Kaufman and Shaby, 2013] using results from [Du et al., 2009, Wang and Loh, 2011].

Asymptotic normality result 2 : Matérn model

Theorem [Bachoc et al., 2019]

Under technical conditions on ν and the sequence $(x_i)_{i\in\mathbb{N}}$ (see paper), we have

$$\sqrt{n}\left(\frac{\widehat{\sigma}^2_{ML}}{\widehat{\rho}^{2\nu}_{ML}} - \frac{\sigma_0^2}{\rho_0^{2\nu}}\right) \xrightarrow[n\to+\infty]{\mathcal{L}|\xi\in\mathcal{E}} \mathcal{N}\left(0, 2\left(\frac{\sigma_0^2}{\rho_0^{2\nu}}\right)^2\right)$$

and

$$\sqrt{n}\left(\frac{\widehat{\sigma}^2_{cML}}{\widehat{\rho}^{2\nu}_{cML}} - \frac{\sigma_0^2}{\rho_0^{2\nu}}\right) \xrightarrow[n\to+\infty]{\mathcal{L}|\xi\in\mathcal{E}} \mathcal{N}\left(0, 2\left(\frac{\sigma_0^2}{\rho_0^{2\nu}}\right)^2\right).$$

- Same conclusions as for the estimation of a variance parameter.

An illustration

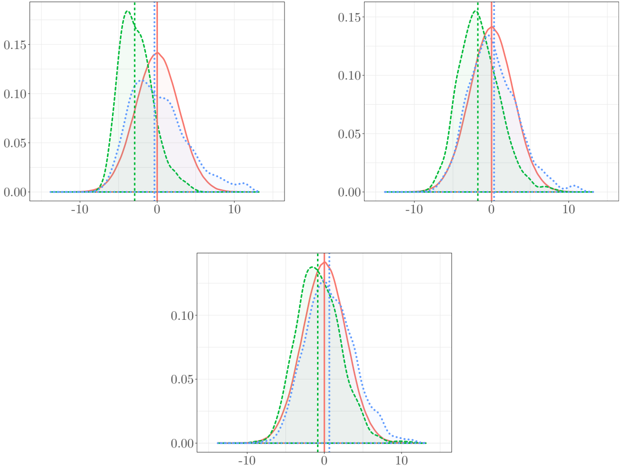


Figure – An example with the estimation of σ_0^2 with boundedness constraints.
Distribution of $n^{1/2}(\hat{\sigma}^2 - \sigma_0^2)$. $n = 20$ (top left), $n = 50$ (top right) and $n = 80$ (bottom).

- Green : ML.
- Blue : cML.
- Red : Gaussian limit.

Some proof ideas : ML for the variance

When $k_\theta = \sigma^2 \tilde{k}$ and for boundedness constraint.

- Write the variance estimator as

$$\begin{aligned}
\widehat{\sigma}^2_{ML} &= \frac{\sigma_0^2}{n} \sum_{i=1}^{n} \frac{(y_i - \mathbb{E}[y_i|y_1,\ldots,y_{i-1}])^2}{\mathrm{Var}(y_i|y_1,\ldots,y_{i-1})} \\
&= \frac{\sigma_0^2}{n} \sum_{i=1}^{m} \frac{(y_i - \mathbb{E}[y_i|y_1,\ldots,y_{i-1}])^2}{\mathrm{Var}(y_i|y_1,\ldots,y_{i-1})} + \frac{\sigma_0^2}{n} \sum_{i=m+1}^{n} \frac{(y_i - \mathbb{E}[y_i|y_1,\ldots,y_{i-1}])^2}{\mathrm{Var}(y_i|y_1,\ldots,y_{i-1})} \\
&:= A_m + B_{m,n}
\end{aligned}$$

  with fixed m and as $n \to \infty$.
- Approximate boundedness event by $\{y_i \in [\ell, u]; i = 1, \ldots, m\}$.
- A_m is negligible as $n \to \infty$.
- Conditioning by approximated boundedness does not affect $B_{m,n}$ by independence so $\sqrt{n}(B_{m,n} - \sigma_0^2) \to \mathcal{N}(0, 2\sigma_0^4)$ also conditionally.
- Conclude by letting $m = m_n \to \infty$ as $n \to \infty$ slowly enough.
- Same method for monotonicity and convexity.

Some proof ideas : ML for Matérn

- Introduce estimated variance with imposed correlation length

$$\bar{\sigma}_n^2(\rho) \in \underset{\sigma^2 \in (0,\infty)}{\text{argmax}} \; \mathcal{L}_n(\sigma^2, \rho).$$

- Then from [Kaufman and Shaby, 2013, Du et al., 2009, Wang and Loh, 2011], for $0 < \rho_l < \rho_u < \infty$,

$$\sup_{\rho_1, \rho_2 \in [\rho_l, \rho_u]} \left| \frac{\bar{\sigma}_n^2(\rho_1)}{\rho_1^{2\nu}} - \frac{\bar{\sigma}_n^2(\rho_2)}{\rho_2^{2\nu}} \right| = o_{\mathbb{P}}(1/\sqrt{n}).$$

- We conclude with the previous result for

$$\frac{\bar{\sigma}_n^2(\rho_0)}{\rho_0^{2\nu}}.$$

Some proof ideas : cML

For boundedness constraint.

- We show that the two added terms in the constrained likelihood are negligible.
- For the unconditional constraints :

$$|\log(\mathbb{P}(\sigma_1 \xi \in \mathcal{E})) - \log(\mathbb{P}(\sigma_2 \xi \in \mathcal{E}))| \leq \text{Constant} \left| \sigma_1^2 - \sigma_2^2 \right| .$$

  Using Tsirelson's theorem.
- For the conditional constraints :

$$\sup_{\theta \in \Theta} |\log(\mathbb{P}_\theta(\xi \in \mathcal{E}|y))| = o_{\mathbb{P}|\xi \in \mathcal{E}}(1).$$

  Because conditional constraint probability $\to 1$. More technical part. Using Borel-TIS inequality, and RKHS arguments for the Matérn case.

Outline

- For boundedness constraints, it is possible to consider models of the form $y_i = T(\xi(x_i))$ with T bijective from $\mathbb{R}$ to $[\ell, u]$ and ξ a Gaussian process.
  - No computational problem.
- For monotonicity and convexity constraints, the model $P(\xi \in .|\xi \in \mathcal{E})$ has become standard.
  - But the constraint $\xi \in \mathcal{E}$ needs to be approximated.
  - $\xi \in \mathcal{E}$ is replaced by a finite number of constraints on inducing points in [Da Veiga and Marrel, 2012, Golchi et al., 2015].

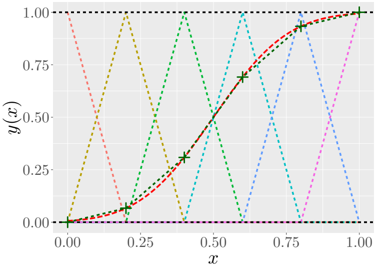
$$(\partial_i \xi)(s) \geq 0, s \in [0,1]^d \qquad \approx \qquad (\partial_i \xi)(s_j) \geq 0, j = 1, \ldots, m.$$

  - ξ is replaced by a finite-dimensional approximation ξ_m in [López-Lopera et al., 2018, Maatouk and Bay, 2017].

In dimension 1, for $x \in [0,1]$:

$$\xi_m(x) = \sum_{i=1}^{m} \xi(t_i)\phi_i(x) = \sum_{i=1}^{m} \xi_m(t_i)\phi_i(x),$$

- $0 = t_1 < \cdots < t_m = 1$: knots,
- ϕ_i : hat basis function centered at t_i.



Finite-dimensional linear inequalities for the constraints

In dimension 1

- Boundedness

$$\xi_m \text{ is bounded in } [\ell, u] \text{ on } [0,1] \iff \xi_m(t_i) \in [\ell, u] \text{ for } i = 1, \ldots, m.$$

- Monotonicity

$$\xi_m \text{ is non-decreasing on } [0,1] \iff \xi_m(t_i) \le \xi_m(t_{i+1}) \text{ for } i = 1, \ldots, m-1.$$

In dimension d

- Finite-dimensional approximation, for $u = (u_1, \ldots, u_d) \in [0,1]^d$,

$$\xi_m(u_1, \ldots, u_d) = \sum_{i_1=1}^{m_1} \cdots \sum_{i_d=1}^{m_d} \xi_m(t_{i_1}^{(1)}, \ldots, t_{i_d}^{(d)}) \phi_{i_1}^{(1)}(u_1) \cdots \phi_{i_d}^{(d)}(u_d),$$

  - $(t_{i_1}^{(1)}, \ldots, t_{i_d}^{(d)})$: multi-dimensional knot,
  - $\phi_{i_1}^{(1)}(\cdot) \cdots \phi_{i_d}^{(d)}(\cdot)$: multi-dimensional hat basis function.

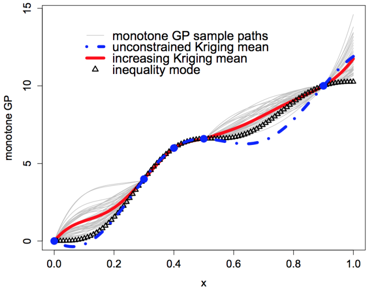
- For boundedness, monotonicity, component-wise convexity :

$$\xi_m \in \mathcal{E} \iff \text{finite number of linear inequalities on } [\xi_m(t_{i_1}^{(1)}, \ldots, t_{i_d}^{(d)})]_{i_1, \ldots, i_d}.$$

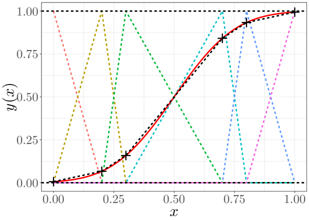
In the frame of [López-Lopera et al., 2018, Maatouk and Bay, 2017].

$\Longrightarrow$ Boils down to optimizing/sampling w.r.t. the Gaussian vector $[\xi_m(t^{(1)}_{i_1}, \ldots, t^{(d)}_{i_d})]_{i_1,\ldots,i_d}$.

- The mode is the "most likely" function for ξ_m, obtained by quadratic optimization with linear constraints.
- Conditional realizations of ξ_m can be sampled approximately, for instance by Hamiltonian Monte Carlo for truncated Gaussian vectors [Pakman and Paninski, 2014].



Introduced in [Bachoc et al., 2022].



- Let $\widehat{Y}$ be the mode function with an ordered set of knots :

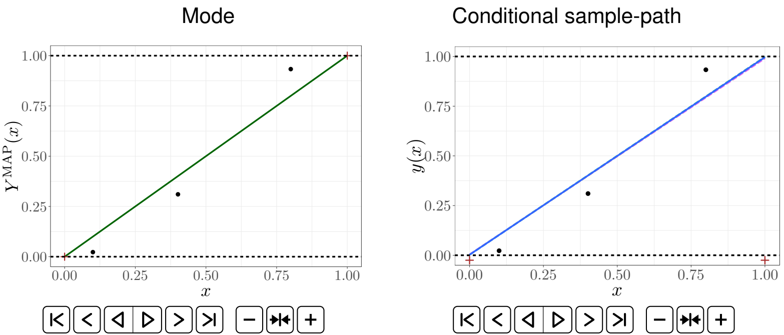
$$\{t_1, \dots, t_m\}, \quad \text{with} \quad 0 = t_1 < \cdots < t_m = 1.$$

- Here, we aim at adding a new knot t (where ?).
- To do so, we aim at *maximising the total modification of the mode* :

$$I(t) = \int_{[0,1]} \left(\widehat{Y}_{+t}(x) - \widehat{Y}(x) \right)^2 dx. \tag{1}$$

The integral in (1) has a closed-form expression.

We write the mode $\widehat{Y} = Y^{\mathrm{MAP}}$.



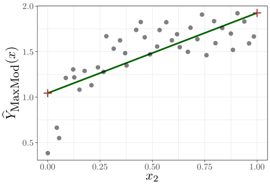
● Observation points + Knots ■ Mode
■ Predictive mean ■ 90% confidence intervals

2D example under monotonicity constraints

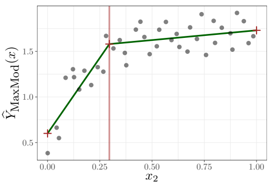
MaxMod in multiD

- Adding new active variables or adding new knots to active variables.

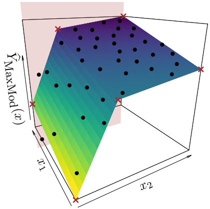
Figure – Evolution of the MaxMod algorithm using $f(x) = \frac{1}{2}x_1 + \mathrm{arctan}(10x_2)$



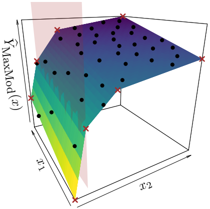
(a) iteration 0



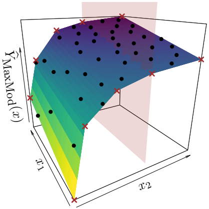
(b) iteration 1



(c) iteration 2



(d) iteration 3

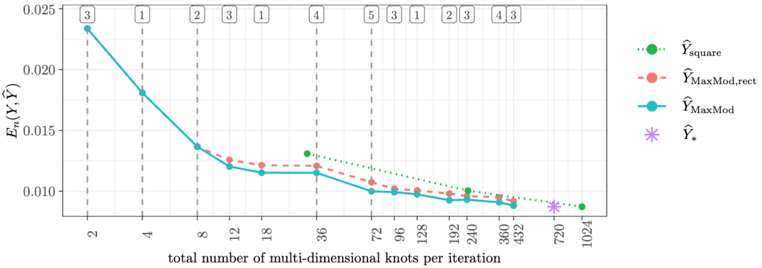


(e) iteration 4

• training points + knots ■ mode

MaxMod results on coastal example

- $E_n(Y, \hat{Y})$: relative square error.
- $\hat{Y}_{\text{square}}$: regularly spaced knots, identical number per variable.
- $\hat{Y}_{\text{MaxMod,rect}}$: regularly spaced knots, numbers per variable given by MaxMod.
- $\hat{Y}_*$: optimized by hand in a previous study.



Approach	m	$E_n(Y, \hat{Y})$ $[1 \times 10^{-3}]$	CPU time [s]		
			Training step	Computation of $\hat{Y}$	Sampling step with 100 realizations
$\hat{Y}_{\text{square}}$	1024	8.72	49.1	8.03	non converged after 1 day
$\hat{Y}_{\text{MaxMod}}$	432	8.81	949.5	0.58	108.72

Outline

When the sequence of knots is fixed and dense

Setting :

- Fixed data set from now on.
- $\mathcal{I}$: set of functions interpolating the data set.
- For variable $j \in \{1, \ldots, d\}$: sequence of one-dimensional knots $t_1^{(j)}, \ldots, t_{m_j}^{(j)}$ and $m_j \to \infty$. The sequence is dense in $[0, 1]$.
- The mode $\widehat{Y}_{m_1,\ldots,m_d} : [0, 1]^d \to \mathbb{R}$.
- Kernel k with corresponding RKHS $\mathcal{H}$ of functions from $[0, 1]^d$ to $\mathbb{R}$.
- Inequality set $\mathcal{C}$ of functions from $[0, 1]^d$ to $\mathbb{R}$.

Theorem [Bay et al., 2017, Bay et al., 2016]

Under some technical conditions

$$\widehat{Y}_{m_1,\ldots,m_d} \to Y_{\mathrm{opt}},$$

uniformly on $[0, 1]^d$, with

$$Y_{\mathrm{opt}} = \operatorname*{argmin}_{f \in \mathcal{H} \cap \mathcal{C} \cap \mathcal{I}} \|f\|_{\mathcal{H}}.$$

Multiaffine extension

Definition

Let $F_1, \dots, F_d$ be (general) closed subsets of $[0,1]$ containing 0 and 1.
Let f be a continuous function on $F = F_1 \times \cdots \times F_d$.
Then, there exists a *unique continuous extension of f on* $[0,1]^d$ *such that any 1D marginal cut functions* $u_i \mapsto f(u_i, t_{\sim i})$ *is affine* on intervals of $[0,1] \setminus F_i$.
Denoted $P_{F \to [0,1]^d}(f)$, it is obtained by sequential 1D affine interpolations.

$\Longrightarrow$ $P_{F \to [0,1]^d}(f)$ is called the multiaffine extension of f.

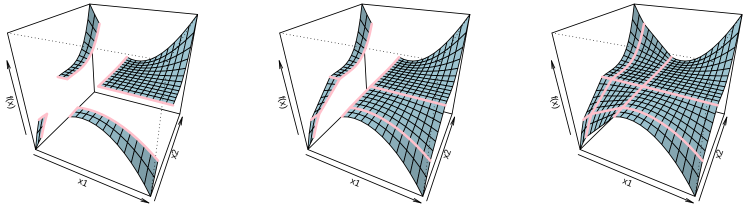


Figure – Sequential construction of the multiaffine extension (2D case).

Multiaffine extension

Properties

- The multiaffine extension is expressed with 2^d neighbours as

$$P_{F \to [0,1]^d}(f)(u_1, \ldots, u_d) = \sum_{\epsilon_1, \ldots, \epsilon_d \in \{-,+\}} \left(\prod_{j=1}^{d} \omega_{\epsilon_j}(u_j) \right) f(u_1^{\epsilon_1}, \ldots, u_d^{\epsilon_d}),$$

  where u_j^-, u_j^+ are the closest left and right neighbours of u_j in F_j,
  $\omega_+(u_j) = \frac{u_j - u_j^-}{u_j^+ - u_j^-}$ if $u_j \notin F_j$ and $\frac{1}{2}$ otherwise, and $\omega_-(u_j) = 1 - \omega_+(u_j)$.
- It preserves boundedness, monotonicity and componentwise convexity.

The multiaffine extension for a fixed sequence of knots that is not dense

Setting :

- For variable $j \in \{1, \dots, d\}$: sequence of one-dimensional knots $t_1^{(j)}, \dots, t_{m_j}^{(j)}$ and $m_j \to \infty$. The sequence has closure $F_j \subset [0, 1]$.

First approach : can we still find a limit function from $[0, 1]^d$ to $\mathbb{R}$?

$\longrightarrow$ Not successful to stay on $[0, 1]^d$ here.

Instead : Work on $F := F_1 \times \cdots \times F_d$ and define

- $\mathcal{H}_F$ RKHS of k restricted to $F \times F$.
- $\mathcal{C}_F$: set of functions from F to $\mathbb{R}$ which multi-affine extensions satisfy inequality constraints.
- $\mathcal{I}_F$: set of functions from F to $\mathbb{R}$ which multi-affine extensions interpolate the data set.

The multiaffine extension for a fixed sequence of knots that is not dense

Theorem [Bachoc et al., 2022]

Under some technical conditions

$$\widehat{Y}_{m_1,\ldots,m_d} \to Y_{\mathrm{opt,F}},$$

uniformly on F, with

$$Y_{\mathrm{opt,F}} = \underset{f \in \mathcal{H}_F \cap \mathcal{C}_F \cap \mathcal{I}_F}{\operatorname{argmin}} \|f\|_{\mathcal{H}_F}.$$

As a consequence

$$\widehat{Y}_{m_1,\ldots,m_d} \to P_{F \to [0,1]^d}\left(Y_{\mathrm{opt,F}}\right),$$

uniformly on $[0, 1]^d$.

Application to convergence of MaxMod

- Mode $\widehat{Y}_{\text{MaxMod},m}$ at iteration m of MaxMod.
- We add an exploration reward to MaxMod.

Theorem [Bachoc et al., 2022]

Under some technical conditions, as $m \to \infty$,

$$\widehat{Y}_{\text{MaxMod},m} \to Y_{\text{opt}},$$

uniformly on $[0,1]^d$, with

$$Y_{\text{opt}} = \underset{f \in \mathcal{H} \cap \mathcal{C} \cap \mathcal{I}}{\text{argmin}} \|f\|_{\mathcal{H}}.$$

Application to convergence of MaxMod

Proof arguments :

$\Longrightarrow$ let us show that sequence of knots is dense.

- As is common for algorithms maximizing acquisition functions (EGO,...), two ingredients :
  - $\rightarrow$ Show that acquisition function is small at points close to existing ones.
  - $\rightarrow$ Show that acquisition function is large at points away from existing ones.
- Here :
  - $\rightarrow$ Show that mode perturbation vanishes from $\widehat{Y}_{\text{MaxMod},m}$ to $\widehat{Y}_{\text{MaxMod},m+1}$ $\longrightarrow$ previous convergence result.
  - $\rightarrow$ Acquisition function is large at points away from existing ones $\longrightarrow$ the exploration reward.

Conclusion

Summary.

- Inequality constraints correspond to additional information (e.g. physical knowledge).
- Taking them into account can significantly improve the predictions.
- With a computational cost (explicit $\Longrightarrow$ Monte Carlo).
- Asymptotically, we do not see an impact of the constraints and ML $\approx$ cML.
- MaxMod algorithm for higher dimension.

Main open question on likelihood theory.

- How to analyse asymptotically n-dependent constraints $\xi \in \mathcal{E}_n$ with

$$\mathbb{P}(\xi \in \mathcal{E}_n) \underset{n \to \infty}{\longrightarrow} 0.$$

  - For instance boundedness with tighter and tighter bounds or monotonicity over larger and larger domains.
  - Should yield more impacts of the constraints ?
  - Previous proof techniques do not apply.

Subsequent and ongoing work on computation.

- Additive model and corresponding MaxMod : [López-Lopera et al., 2022].
- Block additive models and corresponding MaxMod : in preparation.

Main references

References.

- Constrained Gaussian processes : [López-Lopera et al., 2018].
- Constrained Maximum Likelihood : [Bachoc et al., 2019].
- MaxMod : [Bachoc et al., 2022].
- Extension of MaxMod for additive models : [López-Lopera et al., 2022].
- R package LineqGPR : `https://github.com/anfelopera/lineqGPR`.

Thank you for your attention !

Bibliography I

Azzimonti, D., Ginsbourger, D., Rohmer, J., and Idier, D. (2019).
Profile extrema for visualizing and quantifying uncertainties on excursion regions : application to coastal flooding.
Technometrics.

Bachoc, F., Lagnoux, A., and López-Lopera, A. F. (2019).
Maximum likelihood estimation for Gaussian processes under inequality constraints.
Electronic Journal of Statistics, 13(2) :2921–2969.

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

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.

Bay, X., Grammont, L., and Maatouk, H. (2017).
A new method for interpolating in a convex subset of a Hilbert space.
Computational Optimization and Applications, 68(1) :95–120.

Bibliography II

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.

Da Veiga, S. and Marrel, A. (2012).
Gaussian process modeling with inequality constraints.
Annales de la Faculté des Sciences de Toulouse Mathématiques, 21(6) :529–555.

Du, J., Zhang, H., and Mandrekar, V. (2009).
Fixed-domain asymptotic properties of tapered maximum likelihood estimators.
The Annals of Statistics, 37 :3330–3361.

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

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

Kaufman, C. and Shaby, B. (2013).
The role of the range parameter for estimation and prediction in geostatistics.
Biometrika, 100 :473–484.

Bibliography III

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

López-Lopera, A. F., Bachoc, F., and Roustant, O. (2022).
High-dimensional additive Gaussian processes under monotonicity constraints.
In *NeurIPS*.

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.

Rohmer, J., Idier, D., Paris, F., Pedreros, R., and Louisor, J. (2018).
Casting light on forcing and breaching scenarios that lead to marine inundation : Combining numerical simulations with a random-forest classification approach.
Environmental Modelling and Software, 104 :64–80.

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

Wang, D. and Loh, W.-L. (2011).
On fixed-domain asymptotics and covariance tapering in Gaussian random field models.
Electronic Journal of Statistics, 5 :238–269.

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