Constrained Gaussian processes: methodology, theory and applications

Hassan MAATOUK

hassan.maatouk@univ-rennes2.fr

Workshop on Gaussian Processes, November 6-7, 2017, St-Etienne (France)

UNIVERSITÉ RENNES 2

Plan

1. General introduction of GP regression and motivating example
2. Gaussian processes with inequality constraints
   - Finite-dimensional approximation of GPs
   - Simulation of truncated Gaussian vectors
3. Generalization of the Kimeldorf-Wahba correspondence
4. Real application in Insurance and Finance : estimation of discount factors and default probabilities
   - Discount factors
   - Default probabilities 'Credit Default Swaps (CDS)'
5. Noisy observations case

Plan

1. General introduction of GP regression and motivating example
2. Gaussian processes with inequality constraints
   - Finite-dimensional approximation of GPs
   - Simulation of truncated Gaussian vectors
3. Generalization of the Kimeldorf-Wahba correspondence
4. Real application in Insurance and Finance : estimation of discount factors and default probabilities
   - Discount factors
   - Default probabilities 'Credit Default Swaps (CDS)'
5. Noisy observations case

Gaussian Process Regression (GPR) or Kriging

- The following nonparametric function estimation is considered

$$y = f(\boldsymbol{x}), \qquad \boldsymbol{x} \in \mathbb{R}^d.$$

- Observations : $f(\boldsymbol{x}^{(i)}) = y_i, \ i = 1, \ldots, n$.

- In statistical framework, y is viewed as a realization of a GP Y :

$$Y(\boldsymbol{x}) := \eta(\boldsymbol{x}) + Z(\boldsymbol{x}),$$

where η is the mean and Z is a zero-mean GP with covariance function K.

- The conditional process remains a GP

$$\left\{Y(\boldsymbol{x}) \,\middle|\, Y(\boldsymbol{x}^{(1)}) = y_1, \ldots, Y(\boldsymbol{x}^{(n)}) = y_n\right\} \sim \mathcal{N}\left(\zeta(\boldsymbol{x}), \tau^2(\boldsymbol{x})\right),$$

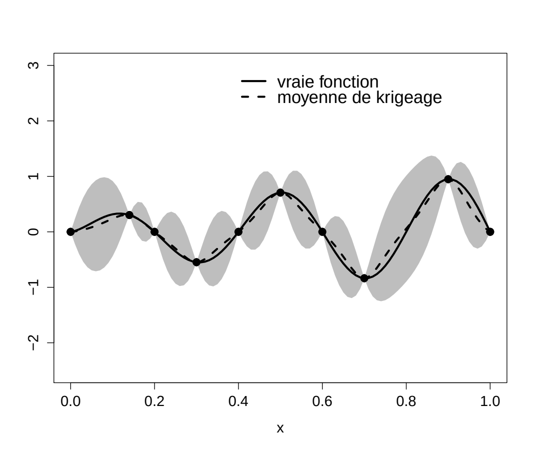
with

$$\begin{cases} \zeta(\boldsymbol{x}) = \eta(\boldsymbol{x}) + \boldsymbol{k}(\boldsymbol{x})^\top \mathbb{K}^{-1}(\boldsymbol{y} - \boldsymbol{\mu}) \\ \tau^2(\boldsymbol{x}) = K(\boldsymbol{x}, \boldsymbol{x}) - \boldsymbol{k}(\boldsymbol{x})^\top \mathbb{K}^{-1} \boldsymbol{k}(\boldsymbol{x}) \end{cases}$$

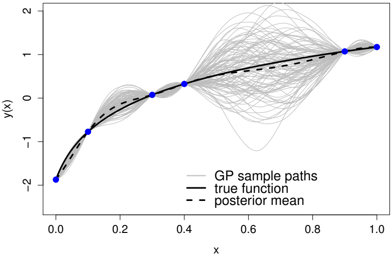
$\boldsymbol{\mu}_i = \eta(\boldsymbol{x}^{(i)})$, $\mathbb{K}_{i,j} = K(\boldsymbol{x}^{(i)}, \boldsymbol{x}^{(j)}), \ i, j = 1, \ldots, n$ and $(\boldsymbol{k}(\boldsymbol{x}))_i = \left(K(\boldsymbol{x}, \boldsymbol{x}^{(i)})\right)$.

Gaussian Process Regression

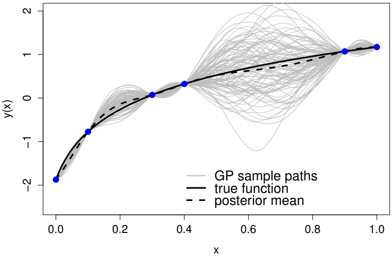
Let $f(x) = \sqrt{x} \times sin(5\pi x), \; x \in [0, 1]$ be a real function.

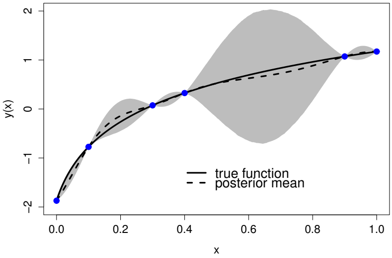


Motivating example

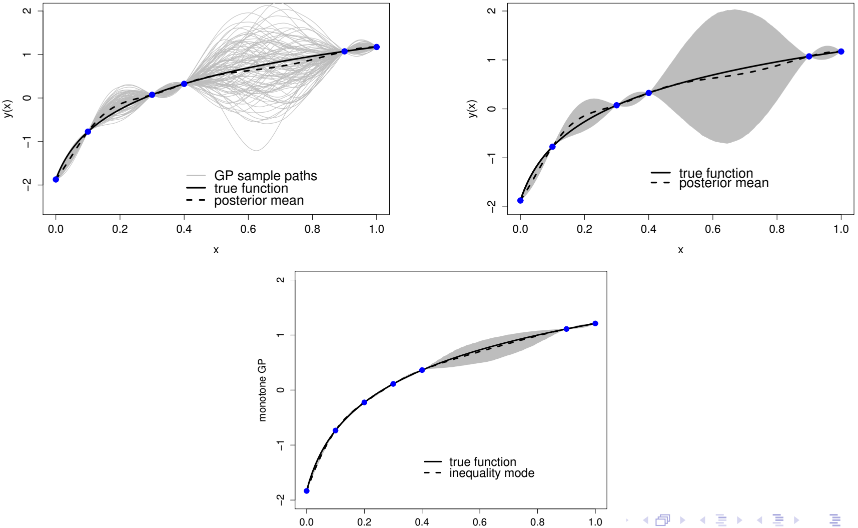


Motivating example





Motivating example



Plan

1. General introduction of GP regression and motivating example
2. Gaussian processes with inequality constraints
   - Finite-dimensional approximation of GPs
   - Simulation of truncated Gaussian vectors
3. Generalization of the Kimeldorf-Wahba correspondence
4. Real application in Insurance and Finance : estimation of discount factors and default probabilities
   - Discount factors
   - Default probabilities 'Credit Default Swaps (CDS)'
5. Noisy observations case

Gaussian processes with inequality constraints

- I is the space of interpolation conditions : $f(x^{(i)}) = y_i, \ i = 1, \ldots, n$.
- C is the space of convexity constraints (such as boundedness, monotonicity and convexity).
- $(Y(\boldsymbol{x}))_{\boldsymbol{x} \in [0,1]^d}$ is a zero-mean GP with covariance function :

$$K(\boldsymbol{x}, \boldsymbol{x}') = \mathrm{Cov}(Y(\boldsymbol{x}), Y(\boldsymbol{x}')) = \mathbb{E}(Y(\boldsymbol{x})Y(\boldsymbol{x}')).$$

Gaussian processes with inequality constraints

- I is the space of interpolation conditions : $f(x^{(i)}) = y_i, \; i = 1, \ldots, n$.
- C is the space of convexity constraints (such as boundedness, monotonicity and convexity).
- $(Y(\boldsymbol{x}))_{\boldsymbol{x} \in [0,1]^d}$ is a zero-mean GP with covariance function :

$$K(\boldsymbol{x}, \boldsymbol{x}') = \mathrm{Cov}\left(Y(\boldsymbol{x}), Y(\boldsymbol{x}')\right) = \mathbb{E}\left(Y(\boldsymbol{x})Y(\boldsymbol{x}')\right).$$

- **Formulation of the problem :** simulate the Gaussian process Y conditionally to :

$$\begin{aligned} & Y(x^{(i)}) = y_i, \quad i = 1, \ldots, n, \\ & Y \in C. \end{aligned} \qquad (I \cap C)$$

Gaussian processes with inequality constraints

- I is the space of interpolation conditions : $f(x^{(i)}) = y_i, \ i = 1, \dots, n$.
- C is the space of convexity constraints (such as boundedness, monotonicity and convexity).
- $(Y(\boldsymbol{x}))_{\boldsymbol{x} \in [0,1]^d}$ is a zero-mean GP with covariance function :

$$K(\boldsymbol{x}, \boldsymbol{x}') = \mathrm{Cov}(Y(\boldsymbol{x}), Y(\boldsymbol{x}')) = \mathbb{E}(Y(\boldsymbol{x})Y(\boldsymbol{x}')).$$

- **Formulation of the problem :** simulate the Gaussian process Y conditionally to :

$$\begin{aligned} &Y(x^{(i)}) = y_i, \quad i = 1, \dots, n, \\ &Y \in C. \end{aligned} \qquad (I \cap C)$$

Remark

This is a difficult problem because the conditional process $Y \mid Y \in I \cap C$ is not a GP in general.

Finite-dimensional approximation of GPs

- **Methodology :** we develop a finite-dimensional approximation of GPs of the form

$$Y^N(\boldsymbol{x}) := \sum_{j=0}^{N} \xi_j \phi_j(\boldsymbol{x}),$$

with $\xi = (\xi_0 \cdots \xi_N)^\top \sim \mathcal{N}(\mathbf{0}, \Gamma^N)$ and $\{\phi_j\}$ the deterministic basis functions.

Finite-dimensional approximation of GPs

- **Methodology :** we develop a finite-dimensional approximation of GPs of the form

$$Y^N(\boldsymbol{x}) := \sum_{j=0}^{N} \xi_j \phi_j(\boldsymbol{x}),$$

with $\xi = (\xi_0 \cdots \xi_N)^\top \sim \mathcal{N}(\mathbf{0}, \Gamma^N)$ and $\{\phi_j\}$ the deterministic basis functions.

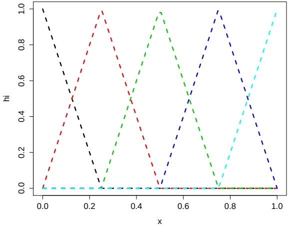
Fundamental property of the basis functions $(\phi_j)_{j=0,\cdots,N}$

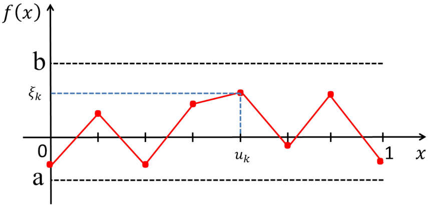
we choose the basis functions $\{\phi_j\}$ such that

- $Y^N(.)$ is positive $\Longleftrightarrow$ $\xi_j \geq 0 \; ; \; 0 \leq j \leq N$.
- $Y^N(\boldsymbol{x}) \in [a, b]$ $\Longleftrightarrow$ $a \leq \xi_j \leq b \; ; \; 0 \leq j \leq N$.
- $Y^N(.)$ is non-decreasing $\Longleftrightarrow$ $\xi_j \geq 0 \; ; \; 0 \leq j \leq N$.
- $Y^N(.)$ is convex $\Longleftrightarrow$ $\xi_j \geq 0 \; ; \; 0 \leq j \leq N$.

Boundedness constraints

1. $C = \{f \ : \ [0,1] \longrightarrow \mathbb{R} \ : \ a \leq f(x) \leq b\}$.
2. The basis functions $(h_j)_j$ are the hat functions associated to the knots $(u_j)_{j=0,\ldots,N}$ such that : $h_j(u_k) = \delta_{j,k}$.





- $Y^N(x) = \sum_{j=0}^{N} \xi_j h_j(x)$ and $Y^N(u_k) = \sum_{j=0}^{N} \xi_j h_j(u_k) = \sum_{j=0}^{N} \xi_j \delta_{j,k} = \xi_k$.
- $Y^N(x) = \sum_{j=0}^{N} Y^N(u_j) h_j(x)$ is a piecewise linear function.

$$Y^N(x) = \sum_{j=0}^{N} Y^N(u_j) h_j(x) \in [a,b], \iff \xi_j = Y^N(u_j) \in [a,b], \ j = 0, \ldots, N.$$

Covariance matrix of the random coefficients ξ_j

By the special choose of the basis functions, we have

$$Y^N(x) := \sum_{j=0}^{N} \xi_j h_j(x) = \sum_{j=0}^{N} Y^N(u_j) h_j(x).$$

To ensure the almost surly uniform convergence of Y^N to Y we suppose : $Y^N(u_j) = Y(u_j)$.
Thus,

$$Y^N(x) = \sum_{j=0}^{N} Y(u_j) h_j(x).$$

Covariance matrix of the random coefficients ξ_j

By the special choose of the basis functions, we have

$$Y^N(x) := \sum_{j=0}^{N} \xi_j h_j(x) = \sum_{j=0}^{N} Y^N(u_j) h_j(x).$$

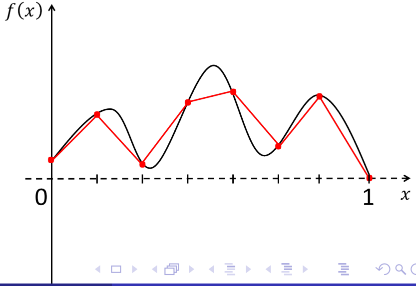
To ensure the almost surly uniform convergence of Y^N to Y we suppose : $Y^N(u_j) = Y(u_j)$. Thus,

$$Y^N(x) = \sum_{j=0}^{N} Y(u_j) h_j(x).$$

$$\begin{aligned} \Gamma^N_{i,j} &= \mathrm{Cov}(\xi_i, \xi_j) = \mathrm{Cov}(Y(u_i), Y(u_j)) \\ &= K(u_i, u_j), \end{aligned}$$

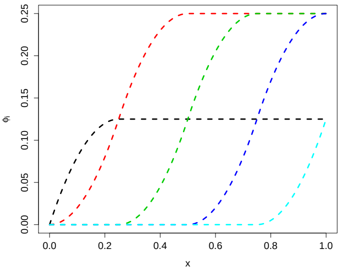
with K the covariance function of the original Gaussian process Y.

$$\mathrm{Cov}\left(Y^N(x), Y^N(x')\right) = h(x)^\top \Gamma^N h(x).$$



Monotonicity constraints

The basis functions are choose as the primitive of the hat functions :
$\phi_j(x) := \int_0^x h_j(t)dt$.



In that case, the finite-dimensional approximation of GPs can be reformulated as

$$Y^N(x) := \zeta + \sum_{j=0}^{N} \xi_j \phi_j(x) = Y(0) + \sum_{j=0}^{N} Y'(u_j)\phi_j(x).$$

Thus, $\Gamma^N_{j,k} = \mathrm{Cov}(\xi_j, \xi_k) = \mathrm{Cov}\left(Y'(u_j), Y'(u_k)\right) = \frac{\partial^2 K(u_j,u_k)}{\partial x \partial x'}$.

New formulation of the problem - case of monotonicity constraints

The simulation of Y^N conditionally to interpolation and inequality constraints is equivalent to the simulation of the Gaussian vector ξ such that

$$(A\xi)_i := \zeta + \sum_{j=0}^{N} \xi_j \phi_j\left(x^{(i)}\right) = y_i, \quad i = 1, \dots, n \qquad \text{(n interpolation conditions)} \quad I_\xi$$

$$\xi_j \geq 0, \quad j = 0, \dots, N \qquad \text{(N+1 inequality constraints)} \quad C_{\text{coef}}$$

with $A_{i,j} := \phi_j\left(x^{(i)}\right)$ and $A_{i,1} = 1, \; i = 1, \dots, n.$

New formulation of the problem - case of monotonicity constraints

The simulation of Y^N conditionally to interpolation and inequality constraints is equivalent to the simulation of the Gaussian vector ξ such that

$$(A\xi)_i := \zeta + \sum_{j=0}^{N} \xi_j \phi_j \left(x^{(i)}\right) = y_i, \quad i = 1, \dots, n \qquad \text{(n interpolation conditions)} \quad I_\xi$$

$$\xi_j \geq 0, \quad j = 0, \dots, N \qquad \text{(N+1 inequality constraints)} \quad C_{\text{coef}}$$

with $A_{i,j} := \phi_j \left(x^{(i)}\right)$ and $A_{i,1} = 1, \ i = 1, \dots, n$.

The problem is reduced to simulate a truncated Gaussian vector restricted to convex sets.

New formulation of the problem - case of monotonicity constraints

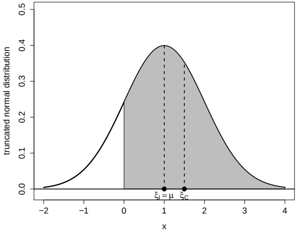
The simulation of Y^N conditionally to interpolation and inequality constraints is equivalent to the simulation of the Gaussian vector ξ such that

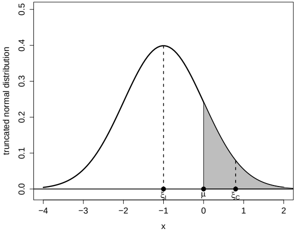
$$(A\xi)_i := \zeta + \sum_{j=0}^{N} \xi_j \phi_j\left(x^{(i)}\right) = y_i, \quad i = 1, \ldots, n \qquad \text{(n interpolation conditions)} \quad I_\xi$$

$$\xi_j \geq 0, \quad j = 0, \ldots, N \qquad \text{(N+1 inequality constraints)} \quad C_{\text{coef}}$$

with $A_{i,j} := \phi_j\left(x^{(i)}\right)$ and $A_{i,1} = 1, \ i = 1, \ldots, n.$

The problem is reduced to simulate a truncated Gaussian vector restricted to convex sets.





Mean and maximum of the posterior distribution

Definition (mean a posteriori estimate)

The mean of the posterior distribution is defined as

$$m_{\text{pos}}^N(x) := \mathbb{E}\left(Y^N(x) \;\middle|\; Y^N(x^{(i)}) = y_i,\ \xi \in C_{\text{coef}}\right) = \phi(x)^\top \xi_{\text{pos}},$$

where $\xi_{\text{pos}} = \mathbb{E}\left(\xi \;\middle|\; \xi \in I_\xi \cap C_{\text{coef}}\right)$.

Mean and maximum of the posterior distribution

Definition (mean a posteriori estimate)

The mean of the posterior distribution is defined as

$$m^N_{\text{pos}}(x) := \mathbb{E}\left(Y^N(x) \;\middle|\; Y^N(x^{(i)}) = y_i,\ \xi \in C_{\text{coef}}\right) = \phi(x)^\top \xi_{\text{pos}},$$

where $\xi_{\text{pos}} = \mathbb{E}\left(\xi \;\middle|\; \xi \in I_\xi \cap C_{\text{coef}}\right)$.

Let μ be the mode of the truncated Gaussian vector $\left\{\xi \;\middle|\; \xi \in I_\xi \cap C_{\text{coef}}\right\}$. Then

$$\mu = \arg \min_{\xi \in I_\xi \cap C_{coef}} \left(\frac{1}{2}\xi^\top (\Gamma^N)^{-1} \xi\right), \tag{1}$$

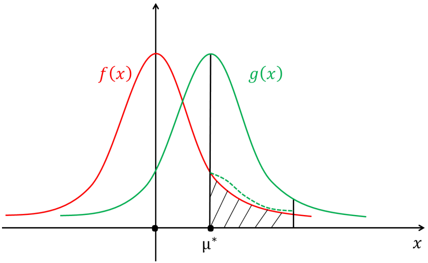
with Γ^N the covariance matrix of the Gaussian vector ξ.

Definition (maximum a posteriori estimate)

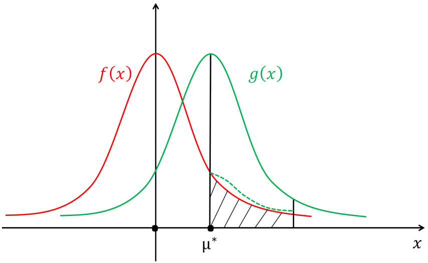
The maximum a posterior (MAP) estimate is defined as

$$M^N_{\text{pos}}(x) := \sum_{j=0}^{N} \mu_j \phi_j(x), \qquad \mu = (\mu_0, \ldots, \mu_N)^\top \text{ is computed from (1).}$$

Truncated normal variables



Truncated normal variables



Proposition

Let $\tilde{f}$ and $\tilde{g}$ be two pseudo-density functions defined as

$$\tilde{f}(x) := f(x \mid 0, \Sigma)1_{x \in C} \quad \text{and} \quad \tilde{g}(x) := g(x \mid \mu^*, \Sigma)1_{x \in C} \ .$$

The optimal constant k such that $\tilde{f}(x) \leq k\tilde{g}(x), \ x \in C$ is equal to

$$k^* = \exp\left(-\frac{1}{2}(\mu^*)^\top \Sigma^{-1} \mu^*\right).$$

Simulation of the truncated multivariate Gaussian random variables

Accept/Rejection algorithm

1. Generate X with density g while $X \notin C$.
2. Generate U uniformly on $[0, 1]$ and accept X if $U \leq e^{(\mu^*)^T \Sigma^{-1} \mu^* - X^T \Sigma^{-1} \mu^*}$. Otherwise, repeat from step 1.

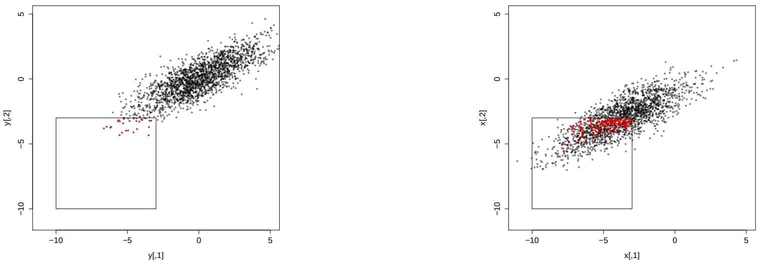
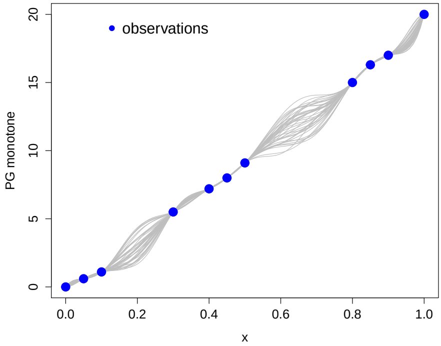
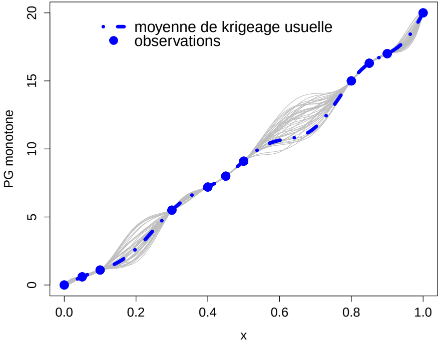


FIGURE: Crude rejection sampling with 2% acceptance rate (left) and the so-called rejection sampling from the mode (RSM) with 20% acceptance rate (right)

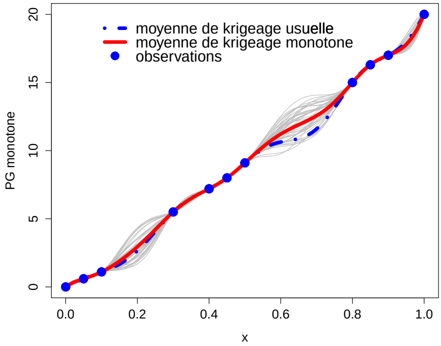
Illustrative example - monotonicity constraints cases



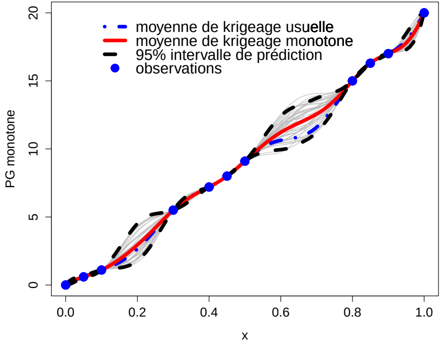
Illustrative example - monotonicity constraints cases



Illustrative example - monotonicity constraints cases



Illustrative example - monotonicity constraints cases



Illustrative example - monotonicity constraints cases

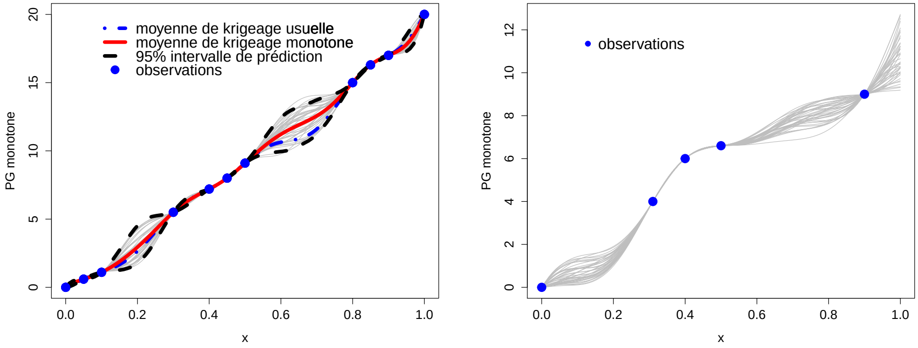


FIGURE: 100 sample paths taken from the GP approximation conditionally to interpolation condition and monotonicity constraints. The unconstrained kriging mean coincides with the MAP estimate in the left figure but not in the right one

Illustrative example - monotonicity constraints cases

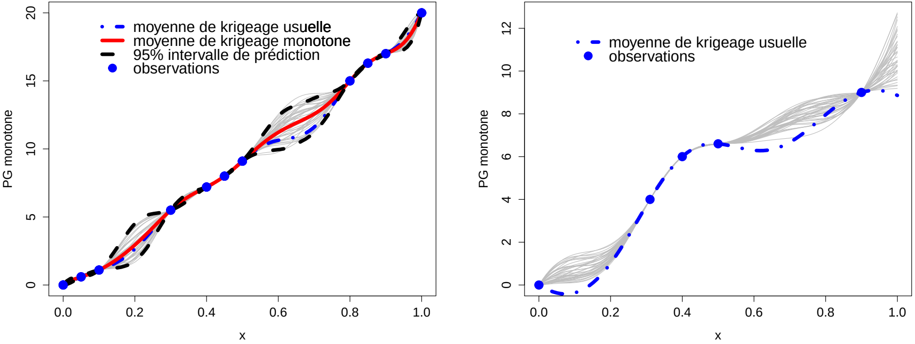


FIGURE: 100 sample paths taken from the GP approximation conditionally to interpolation condition and monotonicity constraints. The unconstrained kriging mean coincides with the MAP estimate in the left figure but not in the right one

Illustrative example - monotonicity constraints cases

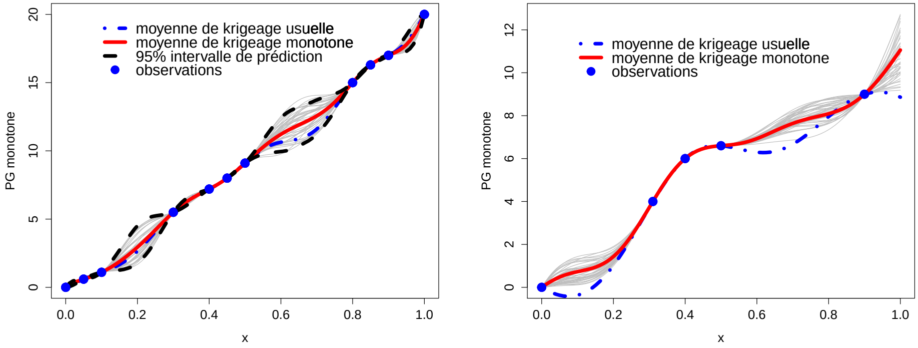


FIGURE: 100 sample paths taken from the GP approximation conditionally to interpolation condition and monotonicity constraints. The unconstrained kriging mean coincides with the MAP estimate in the left figure but not in the right one

Illustrative example - monotonicity constraints cases

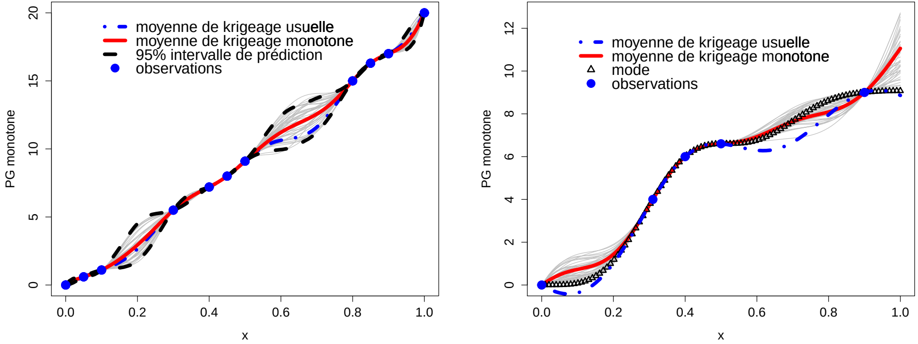


FIGURE: 100 sample paths taken from the GP approximation conditionally to interpolation condition and monotonicity constraints. The unconstrained kriging mean coincides with the MAP estimate in the left figure but not in the right one

Illustrative example - monotonicity constraints cases

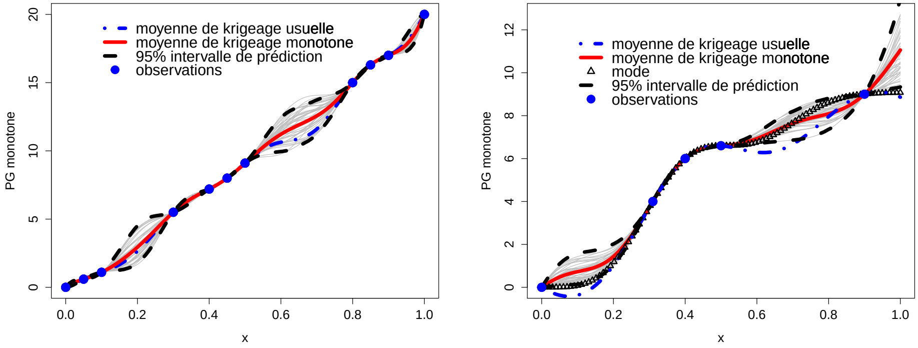


FIGURE: 100 sample paths taken from the GP approximation conditionally to interpolation condition and monotonicity constraints. The unconstrained kriging mean coincides with the MAP estimate in the left figure but not in the right one

Illustrative example - boundedness constraints case

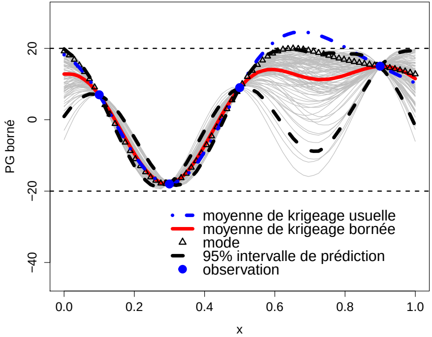
Gaussian process approximation :

$$Y^N(x) := \sum_{j=0}^{N} \xi_j h_j(x) = \sum_{j=0}^{N} Y(u_j) h_j(x). \quad (2)$$

Illustrative example - boundedness constraints case

Gaussian process approximation :

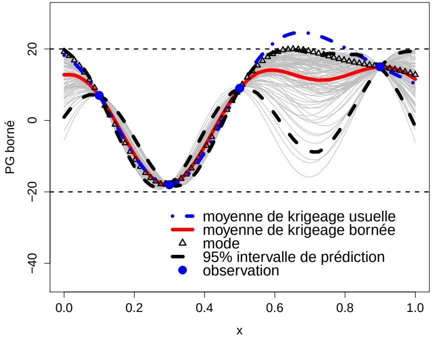
$$Y^N(x) := \sum_{j=0}^{N} \xi_j h_j(x) = \sum_{j=0}^{N} {\color{red}Y(u_j)} h_j(x). \qquad (2)$$



Gaussian process approximation :

$$Y^N(x) := \sum_{j=0}^{N} \xi_j h_j(x) = \sum_{j=0}^{N} Y(u_j) h_j(x). \qquad (2)$$

- The basis functions $h_j, \ j = 0, \ldots, N$ are the hat functions.
- $Y^N(x) \in [a, b]$ if and only if $Y(u_j) \in [a, b]$.
- The sample paths verify boundedness constraints in the entire domain.
- The mean and the maximum of the posterior distribution respect boundedness constraints contrarily to the unconstrained kriging mean.

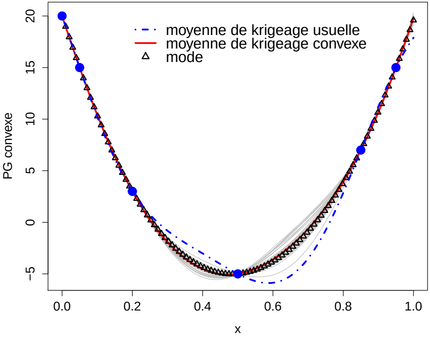


Illustrative example - convexity constraints case

The Gaussian process approximation is equal to

$$Y^N(x) := Y(0) + xY'(0) + \sum_{j=0}^{N} Y''(u_j)\varphi_j(x). \quad (3)$$

- $\varphi_j := \int_0^x \left(\int_0^t h_j(u)du \right) dt.$
- Y^N is convex if and only if $Y''(u_j) \geq 0$.
- The simulated paths respect convexity constraints in the entire domain.
- The mean and the maximum of the posterior distribution respect convexity constraints contrarily to the unconstrained kriging mean.



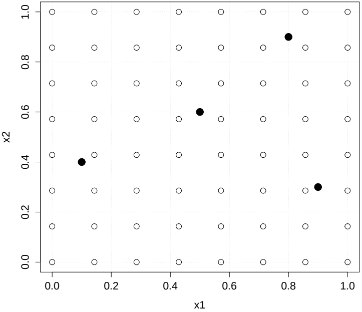
Isotonicity in two dimensions

The input $x = (x_1, x_2)$ is supposed to be in $[0,1]^2$. The monotonicity constraints with respect to the two inputs is defined as

$$\forall x_1 \in [0,1],\ x_2 \longmapsto f(x_1, x_2) \text{ is monotone}$$

and

$$\forall x_2 \in [0,1],\ x_1 \longmapsto f(x_1, x_2) \text{ is monotone.}$$



Isotonicity in two dimensions

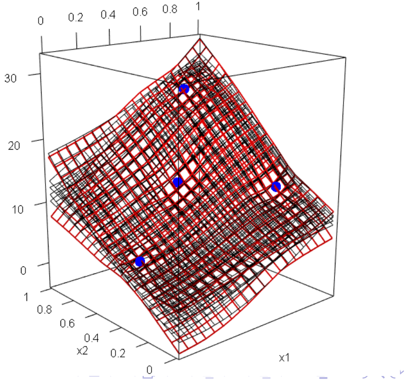
The finite-dimensional approximation of GPs is defined as

$$Y^N(x_1, x_2) := \sum_{i,j=0}^{N} Y(u_i, u_j) h_i(x_1) h_j(x_2), \tag{4}$$

where $h_j, \; j = 0, \ldots, N$ are the hat functions.

Y^N is monotone with respect to the two inputs if and only if the random coefficients $Y(u_i, u_j)$ verify

- $Y(u_{i-1}, u_j) \leq Y(u_i, u_j)$
  $Y(u_i, u_{j-1}) \leq Y(u_i, u_j)$,
  $i, j = 1, \ldots, N$.
- $Y(u_{i-1}, u_0) \leq Y(u_i, u_0), \quad i = 1, \ldots, N$.
- $Y(u_0, u_{j-1}) \leq Y(u_0, u_j), \quad j = 1, \ldots, N$.



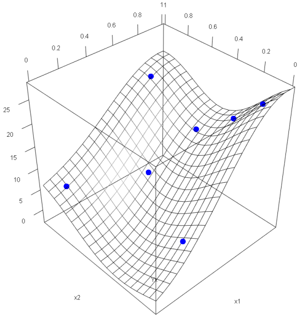
Isotonicity with respect to only one variable

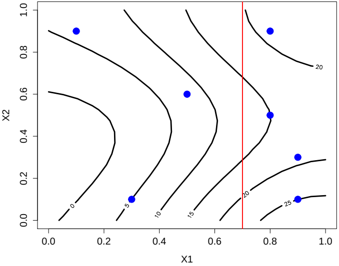
The finite-dimensional approximation of GPs

$$Y^N(x_1, x_2) = \sum_{i,j=0}^{N} {\color{red}Y(u_i, u_j)} h_i(x_1) h_j(x_2) \quad (5)$$

respects monotonicity constraints for only the first variable if and only if :

- $Y(u_{i-1}, u_j) \leq Y(u_i, u_j), \quad i = 1, \ldots, N$ and $j = 0, \ldots, N$.





Plan

1. General introduction of GP regression and motivating example
2. Gaussian processes with inequality constraints
   - Finite-dimensional approximation of GPs
   - Simulation of truncated Gaussian vectors
3. Generalization of the Kimeldorf-Wahba correspondence
4. Real application in Insurance and Finance : estimation of discount factors and default probabilities
   - Discount factors
   - Default probabilities 'Credit Default Swaps (CDS)'
5. Noisy observations case

Kimeldorf-Wahba correspondence

We consider the following optimization problem :

$$\min_{h \in H \cap I} \|h\|_H^2, \qquad (Q)$$

- H the RKHS with r.k. K which is the covariance kernel of Y,
- I the space of functions which verify the interpolation condition.

Kimeldorf-Wahba correspondence

We consider the following optimization problem :

$$\min_{h \in H \cap I} \|h\|_H^2, \qquad (Q)$$

- H the RKHS with r.k. K which is the covariance kernel of Y,
- I the space of functions which verify the interpolation condition.

Theorem (Kimeldorf and Wahba 1970)

The problem (Q) has the kriging mean as an unique solution :

$$h_{opt}(x) = \boldsymbol{k}(x)^\top \mathbb{K}^{-1} \boldsymbol{y}, \qquad (6)$$

where $\boldsymbol{y} = (y_1, \ldots, y_n)^\top$, $\boldsymbol{k}(x) = \left(K(x^{(i)}, x)\right)_i$ and $\mathbb{K}_{i,j} = \left(K(x^{(i)}, x^{(j)})\right)$.

Generalization of the Kimeldorf-Wahba correspondence

We consider the following convex optimization problem :

$$\min_{h \in H \cap I \cap C} \|h\|_H^2, \qquad (P)$$

- H the RKHS with r.k. K which is the covariance kernel of Y,
- I the space of functions which verify the interpolation condition,
- C the space of functions which verify inequality constraints (such as monotonicity, convexity, ...).

Theorem (Bay, Grammont, Maatouk, 2016. Electron. J. Statist.)

The maximum a posteriori estimate (MAP) converges to the constrained spline

$$M_{\text{pos}}^N(x) := \sum_{j=0}^{N} \mu_j \phi_j(x) \xrightarrow[N \to +\infty]{\| \, \|_\infty} h_{opt} := \arg \min_{h \in H \cap I \cap C} \|h\|_H^2.$$

Numerical illustration of the new correspondence

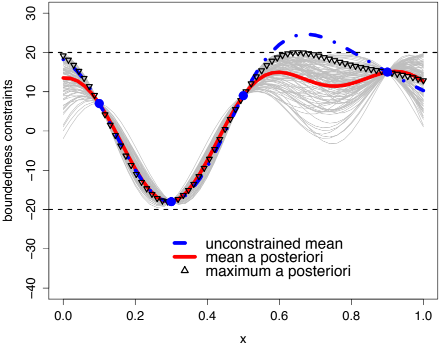


FIGURE: The unconstrained kriging mean coincides with the MAP in the right figure but not in the left one

Numerical illustration of the new correspondence

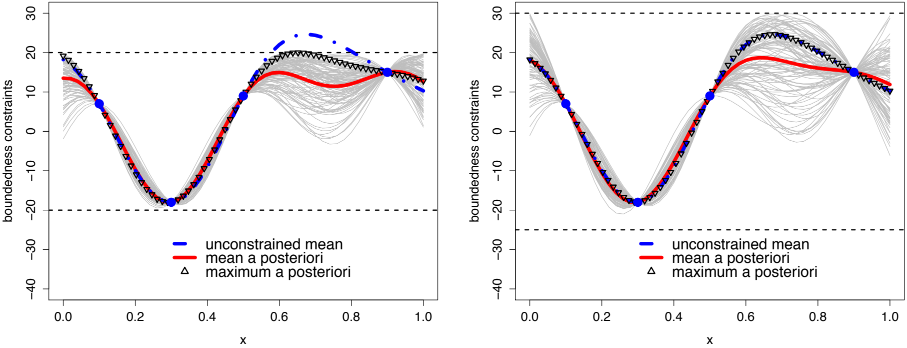


FIGURE: The unconstrained kriging mean coincides with the MAP in the right figure but not in the left one

Numerical illustration of the new correspondence

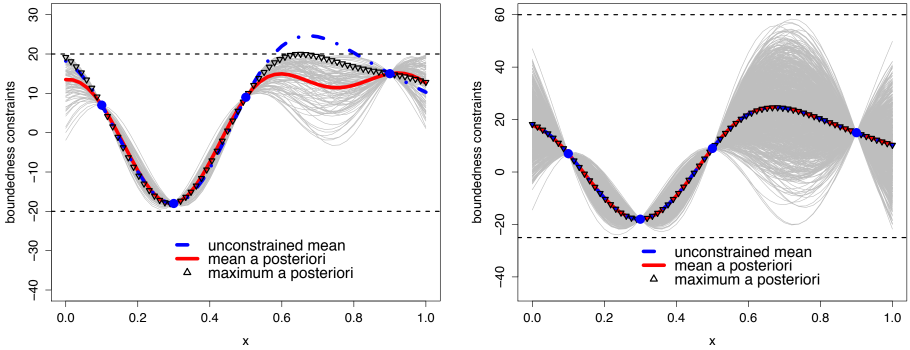


FIGURE: The unconstrained kriging mean coincides with the MAP in the right figure but not in the left one

Constrained cubic spline interpolation

- The cubic spline is the function minimises the linear enegry criterion (LE) (see, e.g., [Wolberg and Alfy, 2002]) :

$$E_L = \int_0^1 (h''(t))^2 \, dt. \tag{7}$$

Subject to $h \in I \cap C$:

$$\min \left\{ \int_0^1 (h''(t))^2 \, dt, \; h \in H^2 \cap I \cap C \right\},$$

where $H^2 = \{ h \in L^2([0,1]) \text{ tel que } h', h'' \in L^2([0,1]) \}$.

Constrained cubic spline interpolation

- The cubic spline is the function minimises the linear enegry criterion (LE) (see, e.g., [Wolberg and Alfy, 2002]) :

$$E_L = \int_0^1 \left(h''(t)\right)^2 dt. \tag{7}$$

Subject to $h \in I \cap C$:

$$\min \left\{ \int_0^1 \left(h''(t)\right)^2 dt,\ h \in H^2 \cap I \cap C \right\},$$

where $H^2 = \left\{h \in L^2([0,1]) \text{ tel que } h', h'' \in L^2([0,1])\right\}$.

- This is equivalent to

$$\min_{\substack{h \in H \\ \alpha+\beta x^{(i)}+h\left(x^{(i)}\right)=y_i \\ \alpha+\beta x+h(x)\in C}} \int_0^1 (h''(t))^2 dt = \|h\|_H^2.$$

Monotone cubic spline interpolation [Wolberg and Alfy, 2002]

TABLE: Wolberg's data used to compare different methods

x	$f(x)$
0.0	0.0
1.0	1.0
2.0	4.8
3.0	6.0
4.0	8.0
5.0	13.0
6.0	14.0
7.0	15.5
8.0	18.0
9.0	19.0
10.0	23.0
11.0	24.1

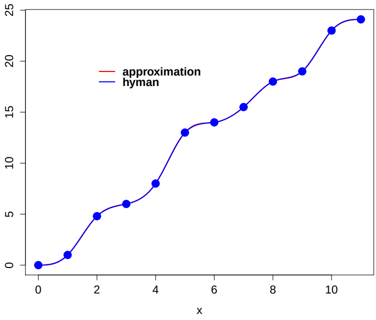


FIGURE: Monotone cubic spline interpolation using Wolberg's data : Hyman approach (blue curve) and the proposed approach $M^N_{\text{pos}}(x)$ with $N = 1000$ (red curve)

Monotone cubic spline interpolation [Wolberg and Alfy, 2002]

TABLE: Linear Energy (LE) criterion using Wolberg's data

Method	E_L
approximation	**131.68**
Hyman	133.19
CSE	132.91
FE	**131.68**
LE	**131.68**
SDDE	223.55
MDE	131.71
FB	236.30

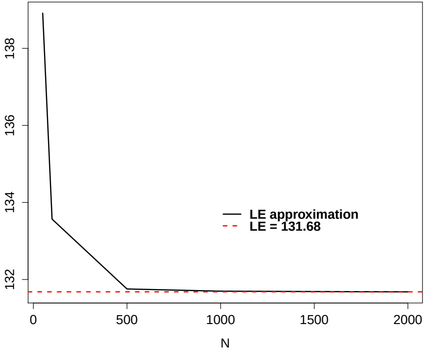


FIGURE: LE using our approach and Wolberg's data

Plan

1. General introduction of GP regression and motivating example
2. Gaussian processes with inequality constraints
   - Finite-dimensional approximation of GPs
   - Simulation of truncated Gaussian vectors
3. Generalization of the Kimeldorf-Wahba correspondence
4. Real application in Insurance and Finance : estimation of discount factors and default probabilities
   - Discount factors
   - Default probabilities 'Credit Default Swaps (CDS)'
5. Noisy observations case

Discount factors

- Discount factor is known to be a non-increasing function with respect to time-to-maturities. It verifies the following linear equality constraints :

$$A \cdot f(\boldsymbol{X}) = \boldsymbol{b}, \qquad f(\boldsymbol{X}) = \left(f(x^{(1)}), \ldots, f(x^{(n)})\right)^\top, \tag{8}$$

where A is a $m \times n$ matrix and $\boldsymbol{b} \in \mathbb{R}^m$.

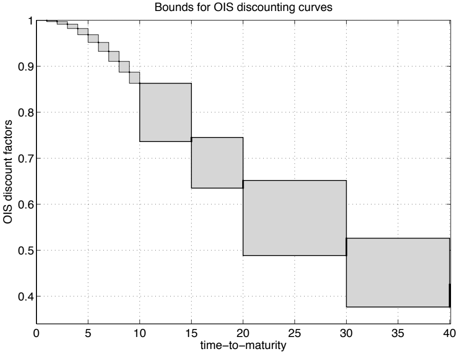
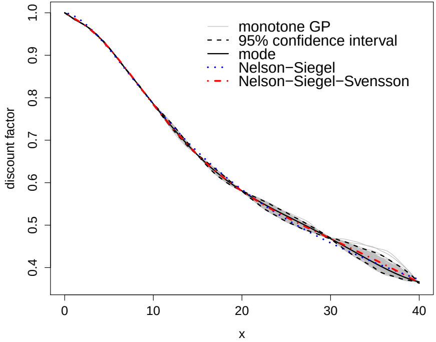


FIGURE: Cousin and Niang 2014

Discount factors using Swap vs Euribor 6M market quotes as of 02/06/2010



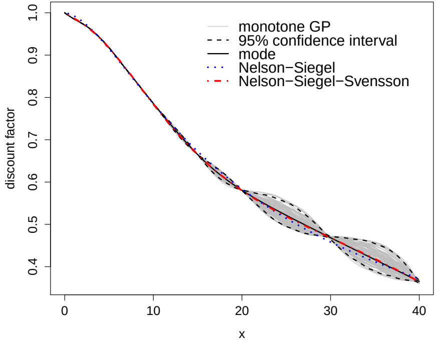


FIGURE: Simulated paths (gray lines) taken from the conditional GP with non-increasing constraints and market-fit constraints using the Gaussian covariance function with nugget equal to 10^{-5} (left) and the Matérn 5/2 covariance function without nugget (right). Swap vs. Euribor 6M market quotes as of 02/06/2010

Forward rates using Swap vs Euribor 6M market quotes as 02/06/2010

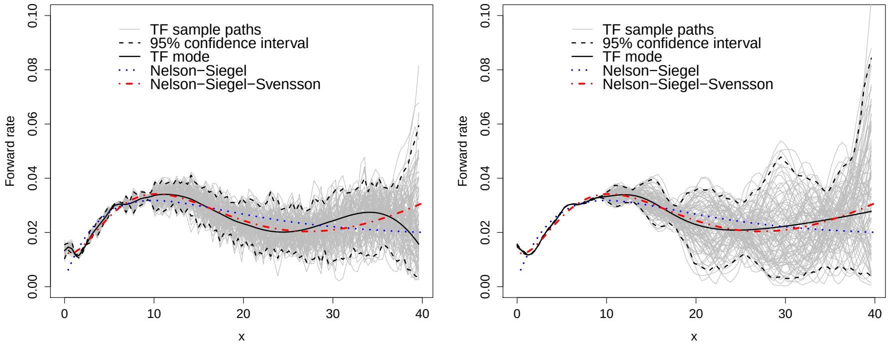


FIGURE: Forward rates obtained from sample paths of previous figures with Gaussian covariance function (left) and Matérn 5/2 covariance function (right)

Several quotation dates Swap vs Euribor discount factors

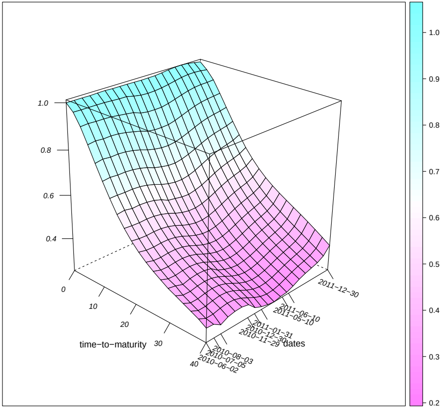


FIGURE: Swap-vs.-Euribor discount factors as a function of time-to-maturities and quotation dates

Default probabilities ‘Credit Default Swaps (CDS)’ on 06/01/2005

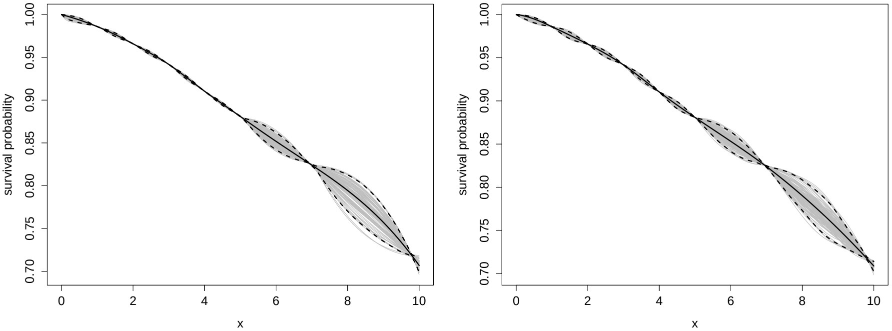


FIGURE: CDS implied survival curves (gray lines) given as simulated paths of a conditional GP with non-increasing constraints using a Gaussian covariance function (left) and a Matérn 5/2 covariance function (right)

Several quotation dates 'Credit Default Swaps (CDS)'

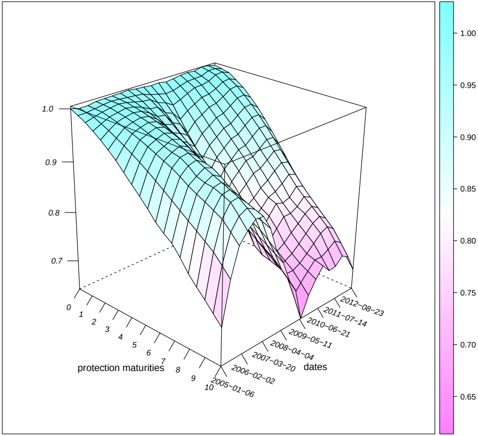


FIGURE: CDS implied survival probabilities as a function of time-to-maturities and quotation dates

Plan

1. General introduction of GP regression and motivating example
2. Gaussian processes with inequality constraints
   - Finite-dimensional approximation of GPs
   - Simulation of truncated Gaussian vectors
3. Generalization of the Kimeldorf-Wahba correspondence
4. Real application in Insurance and Finance : estimation of discount factors and default probabilities
   - Discount factors
   - Default probabilities 'Credit Default Swaps (CDS)'
5. Noisy observations case

Noisy observations

- In many application situations, an approximate response is available

$$f(x^{(i)}) = y_i + \epsilon_i = \tilde{y}_i, \quad i = 1, \ldots, n,$$

where $\epsilon \sim \mathcal{N}(\mathbf{0}, \sigma^2_{\text{noise}} \mathbf{I})$, with σ^2_{noise} the noise variance and $\mathbf{I}$ the identity matrix.

- Conditionally to noisy observations $\tilde{\boldsymbol{y}} = (\tilde{y}_1, \ldots, \tilde{y}_n)^\top$, the process remains a GP

$$Y(\boldsymbol{x}) \mid Y(\boldsymbol{X}) = \tilde{\boldsymbol{y}} \sim \mathcal{GP}\left(\zeta(\boldsymbol{x}), \tau^2(\boldsymbol{x})\right),$$

where

$$\zeta(\boldsymbol{x}) = \eta(\boldsymbol{x}) + \boldsymbol{k}(\boldsymbol{x})^\top (\mathbb{K} + \sigma^2_{\text{noise}} \mathbf{I})^{-1} (\tilde{\boldsymbol{y}} - \boldsymbol{\mu});$$
$$\tau^2(\boldsymbol{x}) = K(\boldsymbol{x}, \boldsymbol{x}) - \boldsymbol{k}(\boldsymbol{x})^\top (\mathbb{K} + \sigma^2_{\text{noise}} \mathbf{I})^{-1} \boldsymbol{k}(\boldsymbol{x}),$$

and $\boldsymbol{\mu} = \eta(\boldsymbol{X})$ is the vector of trend values at the design of experiments, $\mathbb{K}_{i,j} = K(\boldsymbol{x}^{(i)}, \boldsymbol{x}^{(j)}),\ i, j = 1, \ldots, n$ is the covariance matrix of $Y(\boldsymbol{X})$ and $\boldsymbol{k}(\boldsymbol{x}) = (K(\boldsymbol{x}, \boldsymbol{x}^{(i)}))_i$ is the vector of covariance between $Y(\boldsymbol{x})$ and $Y(\boldsymbol{X})$.

Illustrative example - boundedness constraints

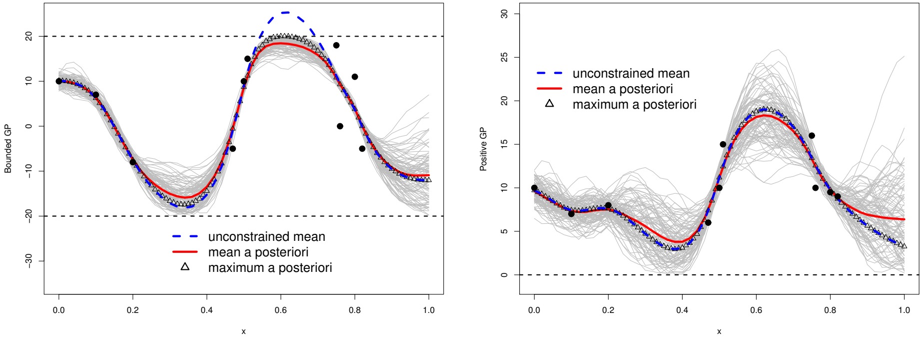


FIGURE: The GP approximation with positivity constraints (right) and boundedness constraints (left). The unconstrained mean coincides with the maximum a posteriori in the right figure but not in the left one

Illustrative example - monotonicity constraints

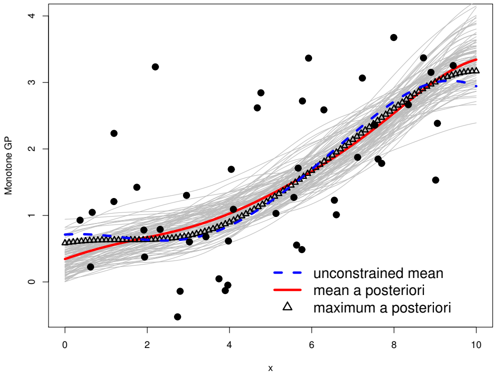


FIGURE: The GP approximation with monotonicity constraints for sinusoidal function $f(x) = 0.32(x + sin(x))$. The unconstrained mean does not coincide with the maximum a posteriori

Illustrative example - isotonicity in two dimensions

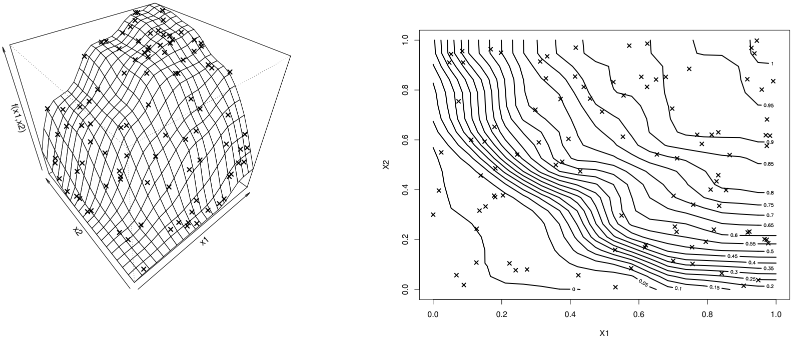


FIGURE: The maximum a posteriori estimate respecting monotonicity (non-decreasing) constraints for the two inputs, and the associated contour levels

Simulation study

- The real non-decreasing functions proposed by [Holmes and Heard, 2003, Neelon and Dunson, 2004] and used in a comparative study by [Shively et al., 2009, Lin and Dunson, 2014] are considered
  1. flat function $f_1(x) = 3, \ x \in (0, 10]$;
  2. sinusoidal function $f_2(x) = 0.32\{x + sin(x)\}, \ x \in (0, 10]$;
  3. step function $f_3(x) = 3$ if $x \in (0, 8]$ and $f_3(x) = 8$ if $x \in (8, 10]$;
  4. linear function $f_4(x) = 0.3x, \ x \in (0, 10]$;
  5. exponential function $f_5(x) = 0.15 \exp(0.6x - 3), \ x \in (0, 10]$;
  6. logistic function $f_6(x) = 3/\{1 + \exp(-2x + 10)\}, \ x \in (0, 10]$.

- The root-mean-square error (RMSE) of the estimates is computed at the one hundred x values taken uniformly (equidistant) in the interval $(0, 10]$:

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(f(x_i) - \hat{f}(x_i)\right)^2},$$

where $\hat{f}(x)$ is the estimate of $f(x)$ and x_i are the n equally-spaced x-values.

Simulation study

TABLE: Root-mean-square error ($\times$ 100) for data of size $n = 100$. The results are obtained by repeating the simulation 5000 times

	Flat	Step	Linear	Exponential	Logistic	Sinusoidal
GP	15.1	27.1	16.7	19.7	25.5	21.9
GP projection	11.3	25.3	16.3	19.1	22.4	21.1
Regression spline	9.7	28.5	24.0	21.3	19.4	22.9
GP approximation	**8.2**	41.1	**15.8**	20.8	21.0	**20.6**

- **GP projection** : Lin, L. and Dunson, D. B. (2014). Bayesian monotone regression using Gaussian process projection. *Biometrika*, 101(2) :303–317.
- **Regression spline** : Shively, T. S., Sager, T. W., and Walker, S. G. (2009). A Bayesian approach to non-parametric monotone function estimation. *J. R. Stat. Soc. B*, 71(1) :159–175.
- **GP approximation** : Maatouk, H. and Bay, X. (2017). Gaussian process emulators for computer experiments with inequality constraints. *Math.Geosci.*, 49 :557–582.

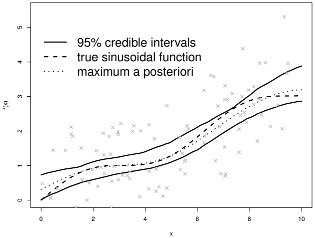


FIGURE: The 95% credible intervals of the GP approximation together with the sinusoidal function, the observations (grey crosses) and the maximum a posteriori estimate

TABLE: Empirical coverage (%) for 95% credible intervals at different x values. The simulations are repeated 1000 times

	0.5	1	1.5	2	2.5	3	3.5	4	4.5	5
GP	97.3	94.6	91.8	88.0	90.5	95.2	96.8	91.0	86.5	86.3
GP projection	94.1	95.4	92.0	89.5	93.1	94.6	96.0	90.0	89.0	86.9
GP approximation	97.0	93.0	89.6	90.1	94.1	97.1	95.5	89.5	85.4	86.7

Simulation study (methodology based on the knowledge of the derivatives of the GP at some input locations)

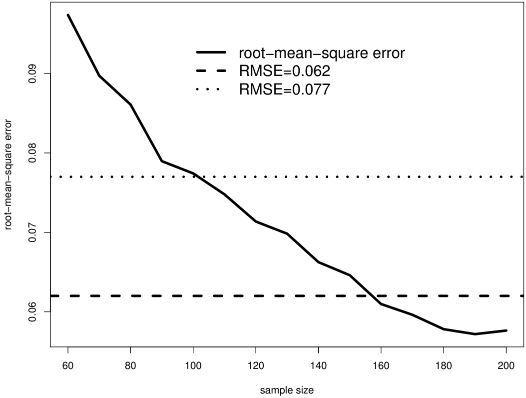


FIGURE: The root-mean-square error at different sample sizes together with the optimal values obtained in [Riihimäki and Vehtari, 2010]. The results are based on 1000 simulation replicates using the logistic function : $2/\{1 + \exp(-8x + 4)\}$, $x \in [0, 1]$

References

Holmes, C. and Heard, N. (2003).
Generalized monotonic regression using random change points.
Stat. Med., 22(4) :623–638.

Lin, L. and Dunson, D. B. (2014).
Bayesian monotone regression using Gaussian process projection.
Biometrika, 101(2) :303–317.

Maatouk, H. and Bay, X. (2016).
A new rejection sampling method for truncated multivariate Gaussian random variables restricted to convex sets.
In Cools, R. and Nuyens, D., editors, *Monte Carlo and Quasi-Monte Carlo Methods*, pages 521–530. Springer International Publishing, Cham.

Neelon, B. and Dunson, D. B. (2004).
Bayesian isotonic regression and trend analysis.
Biometrics, 60(2) :398–406.

Riihimäki, J. and Vehtari, A. (2010).
Gaussian processes with monotonicity information.
J. Mach. Learn. Res., 9 :645–652.

Shively, T. S., Sager, T. W., and Walker, S. G. (2009).
A Bayesian approach to non-parametric monotone function estimation.
J. R. Stat. Soc. B, 71(1) :159–175.

Wolberg, G. and Alfy, I. (2002).
An energy-minimization framework for monotonic cubic spline interpolation.
Journal of Computational and Applied Mathematics, 143(2) :145 – 188.

Thank you for your attention.