Finite-dimensional approximation of Gaussian processes with inequality constraints and errors measurements

Hassan Maatouk

Finite-dimensional approximation of Gaussian processes with inequality constraints

November 20, 2017

H. Maatouk

INRIA Centre de Recherche Rennes - Bretagne Atlantique
Campus de Beaulieu, 35042 Rennes, France
hassan.maatouk@inria.fr

Abstract Due to their flexibility, Gaussian processes (GPs) have been widely used in nonparametric function estimation. A prior information about the underlying function is often available. For instance, the physical system (computer model output) may be known to satisfy inequality constraints with respect to some or all inputs. We develop a finite-dimensional approximation of GPs capable of incorporating inequality constraints and noisy observations for computer model emulators. It is based on a linear combination between Gaussian random coefficients and deterministic basis functions. By this methodology, the inequality constraints are respected in the entire domain. The mean and the maximum of the posterior distribution are well defined. A simulation study to show the efficiency and the performance of the proposed model in term of predictive accuracy and uncertainty quantification is included.

Keywords Gaussian processes · inequality constraints · finite-dimensional approximation · uncertainty quantification · truncated Gaussian vector

1 Introduction and related work

In the estimation of nonparametric function, Gaussian processes (GPs) are the most popular choices. This is because of their flexibility and other nice properties. For instance, the conditional GP with linear equality constraints is still a GP (Cramer and Leadbetter, 1967). Additionally, some inequality constraints (such as monotonicity and convexity) of output computer responses are related to partial derivatives. The partial derivatives of the GP remain GPs (Cramer and Leadbetter, 1967; Parzen, 1962). Incorporating an infinite number of linear inequality constraints (such as boundedness, monotonicity and convexity) into a GP model is a difficult problem. This is because the resulting conditional process is not a GP in general.

Constrained GPs (or kriging) has been studied in the domain of geostatistics (Freulon and de Fouquet, 1993; Kleijnen and Van Beers, 2013). In the literature, there are a variety of ways for incorporating linear inequality constraints into a GP emulator. In Abrahamsen and Benth (2001), the idea is based on a discrete location approximation. In that case, the inequality constraints are satisfied in a finite number of input locations. For monotonicity and isotonicity constraints, some methodologies are based on the knowledge of the derivatives of the GP at some input locations (Golchi et al., 2015; Riihimäki and Vehtari, 2010; Wang and Berger, 2016). As mentioned in Wang and Berger (2016), 'only a modest number of virtual derivative points seems to be needed to effectively impose the desired shape constraint'. In Lin and Dunson (2014), Gaussian process projection is studied. A comparison with spline-based models is included. Recently, a new methodology based on a modification of the covariance function in Gaussian processes to correctly account for known linear constraints is developed in Jidling et al. (2017).

For monotone function estimations, using B-splines was firstly introduced by Ramsay (1988, 1998). The idea is based on the integration of B-splines defined on a properly set of knots with positive coefficients to ensure monotonicity constraints. Xuming and Peide (1996) take the same approach and suggest the calculation of the coefficients by solving a finite linear minimization problem. In Delecroix et al. (1996), nonparametric function estimation in a general cone is studied. Their method is based on a projection into a discretized version of the cone, using the theory of reproducing kernel Hilbert spaces. In Shively et al. (2009), a Bayesian approach to estimate nonparametric monotone functions using restricted splines is developed. In Saarela and Arjas (2011); Tutz and Leitenstorfer (2007), the generalization of monotonic regression to multiple dimensions are studied.

The methodology developed in the present paper is quite different. It is based on a finite-dimensional approximation of GPs (or a GP approxima-

tion) that converges uniformly pathwise. It can be seen as a linear combination between deterministic basis functions and Gaussian random coefficients, where the coefficients are not independent. The main idea is to choose the basis functions such that the infinite number of inequality constraints on the GP approximation are equivalent to a finite number of constraints on the coefficients. Therefore, the simulation of the conditional GP approximation is reduced to the simulation of a Gaussian vector (random coefficients) restricted to convex sets which is a well-known problem with existing algorithms (Botts, 2013; Chopin, 2011; Maatouk and Bay, 2016; Philippe and Robert, 2003; Robert, 1995).

The article is structured as follows. In Section 2, Gaussian processes for computer experiments, their derivative processes and the choice of covariance functions are briefly reviewed. In Section 3, a finite-dimensional approximation of GPs capable of incorporating inequality constraints and noisy observations is developed. Section 4 shows some simulated examples of the finite-dimensional approximation of GPs conditionally to inequality constraints (such as boundedness and monotonicity) and noisy observations in one and two dimensions. Section 5 investigates the performance of the proposed model in terms of predictive accuracy and uncertainty quantification.

2 Gaussian processes for computer experiments

The following model is considered

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

where the simulator response y is assumed to be a deterministic real-valued function of the d-dimensional variable $\boldsymbol{x} = (x_1, \ldots, x_d) \in \mathbb{R}^d$. The true function is supposed to be continuous and evaluated at data of size n (design of experiments) given by the rows of the $n \times d$ matrix $\boldsymbol{X} = \left(\boldsymbol{x}^{(1)}, \ldots, \boldsymbol{x}^{(n)}\right)^\top$, where $\boldsymbol{x}^{(i)} \in \mathbb{R}^d,\ i = 1 \ldots, n$. In many practical situations, it is not possible to get exact evaluations of y at the design of experiments, but rather pointwise noisy measurements. In such case, an approximate response $y(\boldsymbol{X}) + \epsilon$ is available, where $\epsilon \sim \mathcal{N}(\boldsymbol{0}, \sigma^2_{\text{noise}}\mathbf{I})$ with σ^2_{noise} the noise variance and $\mathbf{I}$ the identity matrix. To simplify notations, we denote $\tilde{y}_i = y(\boldsymbol{x}^{(i)}) + \epsilon_i$, $i = 1, \ldots, n$. In the statistical framework, y is viewed as a realization of a continuous GP

$$Z(\boldsymbol{x}) = \eta(\boldsymbol{x}) + Y(\boldsymbol{x}), \quad \boldsymbol{x} \in \mathcal{D} \subset \mathbb{R}^d,$$

where $\mathcal{D}$ is a compact subset of $\mathbb{R}^d$ and the deterministic continuous function $\eta : \boldsymbol{x} \in \mathbb{R}^d \longrightarrow \eta(\boldsymbol{x}) \in \mathbb{R}$ is the mean and Y is a zero-mean GP with

continuous covariance function

$$K \ : \ (\boldsymbol{x}, \boldsymbol{x}') \in \mathcal{D} \times \mathcal{D} \ \longrightarrow \ K(\boldsymbol{x}, \boldsymbol{x}') \in \mathbb{R}.$$

In that case, the GP can be written as $Z \sim \mathcal{GP}\left(\eta(\boldsymbol{x}), K(\boldsymbol{x}, \boldsymbol{x}')\right)$. Conditionally to noisy observations $\tilde{\boldsymbol{y}} = (\tilde{y}_1, \ldots, \tilde{y}_n)^\top$, the process remains a GP

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

where

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

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 $Z(\boldsymbol{X})$ and $\boldsymbol{k}(\boldsymbol{x}) = K(\boldsymbol{x}, \boldsymbol{X})$ is the vector of covariance between $Z\left(\boldsymbol{x}\right)$ and $Z\left(\boldsymbol{X}\right)$. Additionally, the covariance function between any two inputs is equal to

$$C(\boldsymbol{x}, \boldsymbol{x}') = \text{Cov}(Z(\boldsymbol{x}), Z(\boldsymbol{x}') \mid Z(\boldsymbol{X}) = \tilde{\boldsymbol{y}}) = K(\boldsymbol{x}, \boldsymbol{x}') - \boldsymbol{k}(\boldsymbol{x})^\top (\mathbb{K} + \sigma^2_{\text{noise}}\mathbf{I})^{-1} \boldsymbol{k}(\boldsymbol{x}'),$$

where C is the covariance function of the conditional GP. The mean $\zeta(\boldsymbol{x})$ is called kriging mean prediction of $Z(\boldsymbol{x})$ based on the computer model outputs $Z\left(\boldsymbol{X}\right) = \tilde{\boldsymbol{y}}$ (Rasmussen and Williams, 2006).

2.1 The choice of covariance function

The covariance function K controls the smoothness of the kriging metamodel. It must be chosen in the set of definite and positive kernels. In Table 1, some popular covariance functions used in kriging methods are given. Notice that these covariance functions are placed in decreasing order of smoothness, the squared exponential covariance function corresponding to $\mathcal{C}^\infty$ function (i.e., the space of functions that admit derivatives of all orders) and the exponential covariance function to continuous one (Rasmussen and Williams, 2006).

2.2 Derivatives of Gaussian processes

In this subsection, the paths of the GP $(Z(\boldsymbol{x}))_{\boldsymbol{x} \in \mathbb{R}^d}$ are assumed to be of class $\mathcal{C}^p$ (i.e., the space of functions that admit derivatives up to order p). This can be guaranteed if K is smooth enough, and in particular if K is of class $\mathcal{C}^\infty$ (Cramer and Leadbetter, 1967). Since differentiation is a linear operator,

Table 1: Some popular covariance functions used in kriging methods.

Name	Expression	Class
Squared exponential	$\sigma^2 \exp\left(-\frac{(x-x')^2}{2\theta^2}\right)$	$\mathcal{C}^\infty$
Matérn 5/2	$\sigma^2 \left(1+\frac{\sqrt{5}\lvert x-x'\rvert}{\theta}+\frac{5(x-x')^2}{3\theta^2}\right)\exp\left(-\frac{\sqrt{5}\lvert x-x'\rvert}{\theta}\right)$	$\mathcal{C}^2$
Matérn 3/2	$\sigma^2 \left(1+\frac{\sqrt{3}\lvert x-x'\rvert}{\theta}\right)\exp\left(-\frac{\sqrt{3}\lvert x-x'\rvert}{\theta}\right)$	$\mathcal{C}^1$
Exponential	$\sigma^2 \exp\left(-\frac{\lvert x-x'\rvert}{\theta}\right)$	$\mathcal{C}^0$

the order partial derivatives of a GP remain GPs (Cramer and Leadbetter, 1967; Parzen, 1962) and

$$\begin{aligned} E\left(\partial^p_{x_k} Z(\boldsymbol{x})\right) &= \frac{\partial^p}{\partial x_k^p}\eta(\boldsymbol{x}), \\ \mathrm{Cov}\left(\partial^p_{x_k} Z(\boldsymbol{x}^{(i)}), \partial^q_{x_\ell} Z(\boldsymbol{x}^{(j)})\right) &= \frac{\partial^{p+q}}{\partial x_k^p \partial (x'_\ell)^q} K(\boldsymbol{x}^{(i)}, \boldsymbol{x}^{(j)}). \end{aligned}$$

3 Finite-dimensional approximation of GPs

Without loss of generality the input set is supposed the unit hypercube $\mathcal{D} = [0,1]^d$. The input set $\mathcal{D}$ is discretized uniformly to $(N+1)^d$ knots. For example, in one dimension where $\mathcal{D} = [0,1]$, the discretization can be summarized as follow: $0 = t_{N,0}, \ldots, t_{N,N} = 1$. Let $Y \sim \mathcal{GP}(0, K(\boldsymbol{x}, \boldsymbol{x}'))$ be a zero-mean GP with covariance function K. The finite-dimensional approximation of Gaussian processes is defined as

$$\begin{aligned} Y^N(\boldsymbol{x}) &= \sum_{i_1,\ldots,i_d=0}^{N} Y(t_{N,i_1}, \ldots, t_{N,i_d}) \prod_{m\in\{1,\ldots,d\}} \phi_{i_m}(x_m) \qquad (2) \\ &= \zeta_{i_1,\ldots,i_d} \prod_{m\in\{1,\ldots,d\}} \phi_{i_m}(x_m), \quad \boldsymbol{x} \in \mathcal{D} \subset \mathbb{R}^d, \end{aligned}$$

where $(\zeta_{i_1,\ldots,i_d}) = (Y(t_{N,i_1}, \ldots, t_{N,i_d}))^\top$ is a zero-mean Gaussian vector with covariance matrix Γ^N and ϕ_{i_m} is the hat function associated to the knots t_{N,i_m}: $\phi_{i_m}(x) = \phi((x - t_{N,i_m})/\Delta_N)$, where $\Delta_N = 1/N$ and $\phi(x) = (1-|x|)\,\mathbb{1}_{(|x|\le 1)}$, $x \in \mathbb{R}$. The value of any basis function at any knot is equal to Kronecker's Delta function ($\phi_{i_m}(t_{N,i_{m'}}) = \delta_{i_m,i_{m'}}$, $i_m, i_{m'} = 0, \ldots, N$), where $\delta_{i_m,i_{m'}}$ is equal to one if $i_m = i_{m'}$ and zero otherwise. The covariance function $K_N(\boldsymbol{x}, \boldsymbol{x}')$ of the Gaussian process approximation Y^N is equal to

$$K_N(\boldsymbol{x}, \boldsymbol{x}') = \Phi(\boldsymbol{x})^\top \Gamma^N \Phi(\boldsymbol{x}'),$$

where $\Phi(\boldsymbol{x}) = (\prod_{m\in\{1,\dots,d\}} \phi_{i_m}(x_m))_{i_m}$. This type of covariance functions are very similar to ones used in Cressie and Johannesson (2008), where Γ^N is a square positive definite matrix estimated from the data, which it is not the case in the present paper. By this approach (2), simulate the GP approximation is equivalent to simulate the Gaussian vector $(\zeta_{i_1,\dots,i_d})_{i_1,\dots,i_d}$ restricted to

$$\begin{aligned} &Y^N(\boldsymbol{x}^{(i)}) = y_i + \epsilon_i = \tilde{y}_i, \quad i = 1,\dots,n, \\ &(\zeta_{i_1,\dots,i_d})_{i_1,\dots,i_d} \in C_{\text{coef}}, \end{aligned}$$

where $\epsilon_i \overset{i.i.d.}{\sim} \mathcal{N}(0, \sigma^2_{\text{noise}})$ and C_{coef} is the space of coefficients which verify some linear constraints. Next, we show how C_{coef} can be computed in each inequality constraints case. In this paper, boundedness, monotonicity and convexity constraints are considered but the methodology can be easily adapted to any convex sets.

Notice that, model (2) does not correspond to a truncated Karhunen-Loève expansion $Y(x) = \sum_{j=0}^{+\infty} Z_j e_j(x)$; see, for example, Rasmussen and Williams (2006); Trecate et al. (1999) since the coefficients ζ_j are not independent (unlike the coefficients Z_j) and the basis functions ϕ_j are not the eigenfunctions e_j of the Mercer kernel $K(x, x')$.

3.1 Boundedness constraints

The one dimension is a particular case of the two dimensional one. The input $\boldsymbol{x} = (x_1, x_2) \in \mathbb{R}^2$ and without loss of generality in the unit square $\mathcal{D} = [0,1]^2$. The real function is supposed continuous and belong to the convex set

$$C = \left\{ f \in \mathcal{C}^0(\mathcal{D}) \ : \ -\infty \leq a \leq f(\boldsymbol{x}) \leq b \leq +\infty, \ \boldsymbol{x} \in \mathcal{D} \right\}.$$

The finite-dimensional approximation of GPs $\left(Y^N(\boldsymbol{x})\right)_{\boldsymbol{x}\in\mathcal{D}}$ is defined as

$$Y^N(x_1, x_2) = \sum_{i,j=0}^{N} Y(t_{N,i}, t_{N,j}) \phi_i(x_1) \phi_j(x_2) = \sum_{i,j=0}^{N} \zeta_{i,j} \phi_i(x_1) \phi_j(x_2),$$

where $\zeta_{i,j} = Y(t_{N,i}, t_{N,j})$ and $(\phi_j)_j$ are the hat functions. Then, Y^N is bounded between a and b with respect to the two inputs *if and only if* the $(N+1)^2$ random coefficients $\zeta_{i,j} \in [a, b]$. This is because Y^N is a piecewise-linear function. In that case, the space of inequality constraints on the coefficients is equal to

$$C_{\text{coef}} = \left\{ (\zeta_{i,j})_{i,j} \in \mathbb{R}^{(N+1)^2} \ : \ \zeta_{i,j} \in [a, b], \ i, j = 0, \dots, N \right\}.$$

Remark 1. *The multidimensional case is an easy extension of two dimensional one. The input* $\boldsymbol{x} \in \mathcal{D} \subset [0,1]^d$. *The model defined in* (2) *is bounded between* a *and* b *(i.e.,* $Y^N(\boldsymbol{x}) \in [a,b]$*) if and only if the random coefficients* $(\zeta_{i_1,\ldots,i_d}) = (Y(t_{N,i_1},\ldots,t_{N,i_d})) \in [a,b]$.

Proposition 1. *If the realizations of the original GP* Y *are continuous, then the finite-dimensional approximation of GPs* Y^N *is almost surely converge uniformly to* Y *when* N *tends to infinity.*

Proof. To prove the almost sure uniform convergence of the approximating random process Y^N to the limiting process Y, write more explicitly, for any $\omega \in \Omega$

$$Y^N(x;\omega) = \sum_{j=0}^{N} Y(t_{N,j};\omega)\phi_j(x), \quad x \in \mathcal{D} = [0,1].$$

Using the fact that $\phi_j(x) \geq 0$ and $\sum_{j=0}^{N} \phi_j(x) = 1$, for all $x \in \mathcal{D}$, we get

$$\begin{aligned} \left|Y^N(x;\omega) - Y(x;\omega)\right| &= \left|\sum_{j=0}^{N} (Y(t_{N,j};\omega) - Y(x;\omega))\phi_j(x)\right| \\ &\leq \sum_{j=0}^{N} \sup_{|x-x'|\leq\Delta_N} \left|Y(x';\omega) - Y(x;\omega)\right| \phi_j(x) \\ &= \sup_{|x-x'|\leq\Delta_N} \left|Y(x';\omega) - Y(x;\omega)\right|. \end{aligned}$$

Thus, one can deduce that

$$\sup_{x\in\mathcal{D}} \left|Y^N(x;\omega) - Y(x;\omega)\right| \underset{N\to+\infty}{\longrightarrow} 0$$

with probability 1, since the sample paths of the process Y are uniformly continuous on the compact interval $\mathcal{D}$. □

3.2 Isotonicity constraints

The isotonicity constraints in two dimensions are considered. The input $\boldsymbol{x} = (x_1, x_2) \in \mathbb{R}^2$ and without loss of generality in the unit square $\mathcal{D} = [0,1]^2$. The real function f is supposed to be monotone (non-decreasing) with respect to the two inputs

$$x_1 \leq x_1' \quad \text{and} \quad x_2 \leq x_2' \quad \Rightarrow \quad f(x_1,x_2) \leq f(x_1',x_2').$$

The finite-dimensional approximation of GPs $\left(Y^N(\boldsymbol{x})\right)_{\boldsymbol{x}\in\mathcal{D}^2}$ is defined as

$$Y^N(x_1,x_2)=\sum_{i,j=0}^{N} Y(t_{N,i},t_{N,j})\phi_i(x_1)\phi_j(x_2)=\sum_{i,j=0}^{N}\zeta_{i,j}\phi_i(x_1)\phi_j(x_2), \qquad (3)$$

where $\zeta_{i,j}=Y(t_{N,i},t_{N,j})$ and $(\phi_j)_j$ are the hat functions. Then, Y^N is non-decreasing with respect to the two inputs *if and only if* the $(N+1)^2$ random coefficients $\zeta_{i,j}$ verify the following linear constraints:

1. $\zeta_{i-1,j}\leq\zeta_{i,j}$ and $\zeta_{i,j-1}\leq\zeta_{i,j},\ i,j=1,\ldots,N$;
2. $\zeta_{i-1,0}\leq\zeta_{i,0},\ i=1,\ldots,N$;
3. $\zeta_{0,j-1}\leq\zeta_{0,j},\ j=1,\ldots,N$.

Remark 2 (Isotonicity with respect to one variable)**.** *If the function is non-decreasing with respect to the first variable only, then model* (3) *is non-decreasing with respect to* x_1 *if and only if the random coefficients verify:* $\zeta_{i-1,j}\leq\zeta_{i,j},\ i=1,\ldots,N$ *and* $j=0,\ldots,N$.

Remark 3 (Monotonicity constraints)**.** *For monotonicity constraints, the finite-dimensional approximation of GPs can be written as*

$$Y^N(x)=Y(0)+\sum_{j=0}^{N}Y'(t_{N,j})I_j(x)=\gamma+\sum_{j=0}^{N}\zeta_j I_j(x),\quad x\in[0,1], \qquad (4)$$

where $\gamma=Y(0)$, $\zeta_j=Y'(t_{N,j})$ *and* $I_j(x)=\int_0^x\phi_j(t)dt$. *In that case,* Y^N *is monotone if and only if the random coefficients* ζ_j *are all nonnegative. In fact, since* (I_j) *are non-decreasing functions and* (ζ_j) *are nonnegative, then* Y^N *is non-decreasing. Conversely, if* Y^N *is non-decreasing, then* $\zeta_j=(Y^N)'(t_{N,j})\geq 0$. *Thus, the space of inequality constraints on the coefficients is*

$$C_{\text{coef}}=\left\{(\gamma,\zeta)\in\mathbb{R}^{N+2}\ :\ \zeta_j\geq 0,\ j=0,\ldots,N\right\},$$

where $\zeta=(\zeta_0,\ldots,\zeta_N)^\top$. *For all* $x,x'\in\mathcal{D}=[0,1]$, *the covariance function of* Y^N *is equal to*

$$\begin{aligned}K_N(x,x') &= \operatorname{Cov}\left(Y^N(x),Y^N(x')\right)=K(0,0)+\sum_{i=0}^{N}\frac{\partial K}{\partial x}(t_{N,i},0)I_i(x)\\ &+ \sum_{j=0}^{N}\frac{\partial K}{\partial x'}(0,t_{N,j})I_j(x)+\sum_{i,j=0}^{N}\frac{\partial^2 K}{\partial x\partial x'}(t_{N,i},t_{N,j})I_i(x)I_j(x').\end{aligned}$$

3.3 Convexity constraints

For convexity in two dimensions, the finite-dimensional approximation of GPs defined as

$$Y^N(x_1, x_2) = \sum_{i,j=0}^{N} Y(t_{N,i}, t_{N,j})\phi_i(x_1)\phi_j(x_2) = \sum_{i,j=0}^{N} \zeta_{i,j}\phi_i(x_1)\phi_j(x_2),$$

is convex with respect to the two inputs *if and only if* the random coefficients verify

1. $\frac{\zeta_{i,j}-\zeta_{i-1,j}}{t_{N,i}-t_{N,i-1}} \leq \frac{\zeta_{i+1,j}-\zeta_{i,j}}{t_{N,i+1}-t_{N,i}}$ and $\frac{\zeta_{i,j}-\zeta_{i,j-1}}{t_{N,j}-t_{N,j-1}} \leq \frac{\zeta_{i,j+1}-\zeta_{i,j}}{t_{N,j+1}-t_{N,j}},\ i,j = 1,\ldots,N-1;$
2. $\frac{\zeta_{i,0}-\zeta_{i-1,0}}{t_{N,i}-t_{N,i-1}} \leq \frac{\zeta_{i+1,0}-\zeta_{i,0}}{t_{N,i+1}-t_{N,i}},\ i = 1,\ldots,N-1;$
3. $\frac{\zeta_{0,j}-\zeta_{0,j-1}}{t_{N,j}-t_{N,j-1}} \leq \frac{\zeta_{0,j+1}-\zeta_{0,j}}{t_{N,j+1}-t_{N,j}},\ j = 1,\ldots,N-1.$

Remark 4 (Convexity in one dimension)**.** *If the realizations of the original GP Y are assumed to be at least twice differentiable. Then, the finite-dimensional approximation of GPs can be defined as*

$$Y^N(x) = Y(0) + Y'(0)x + \sum_{j=0}^{N} Y''(t_{N,j})\varphi_j(x) = \gamma + \kappa x + \sum_{j=0}^{N} \zeta_j\varphi_j(x),$$

where $\gamma = Y(0)$, $\kappa = Y'(0)$ and $\zeta_j = Y''(t_{N,j})$. The basis functions $(\varphi_j)_j$ are the two times primitive functions of ϕ_j

$$\varphi_j(x) = \int_0^x \left(\int_0^t \phi_j(u)du\right) dt, \quad x \in \mathcal{D}.$$

In that case, Y^N is convex if and only if the random coefficient $Y''(t_{N,j})$ are all nonnegative. Thus, the space of inequality constraints on the coefficients C_{coef} is equal to

$$C_{\text{coef}} = \left\{(\gamma, \kappa, \zeta) \in \mathbb{R}^{N+3} \ :\ \zeta_j \geq 0,\ j = 0,\ldots,N\right\},$$

where $\zeta = (\zeta_0,\ldots,\zeta_N)^\top$. The covariance function of the GP approximation is equal to

$$K_N(x, x') = \left(1, x, \varphi(x)^\top\right)\tilde{\Gamma}^N\left(1, x', \varphi(x')^\top\right)^\top,$$

where $\varphi(x) = (\varphi_0(x),\ldots,\varphi_N(x))^\top$ and

$$\tilde{\Gamma}^N = \begin{bmatrix} K(0,0) & \frac{\partial K}{\partial x'}(0,0) & \frac{\partial^2 K}{\partial (x')^2}(0, t_{N,j}) \\ \frac{\partial K}{\partial x}(0,0) & \frac{\partial^2 K}{\partial x \partial x'}(0,0) & \frac{\partial^3 K}{\partial x \partial (x')^2}(0, t_{N,j}) \\ \frac{\partial^2 K}{\partial x^2}(t_{N,i}, 0) & \frac{\partial^3 K}{\partial x^2 \partial x'}(t_{N,i}, 0) & \Gamma^N_{i,j} \end{bmatrix}_{0 \leq i,j \leq N},$$

and

$$\Gamma^N_{i,j} = \mathrm{Cov}(Y''(t_{N,i}), Y''(t_{N,j})) = \frac{\partial^4 K}{\partial x^2 \partial (x')^2}(t_{N,i}, t_{N,j}), \quad i,j = 0,\ldots,N.$$

3.4 Simulated paths

This subsection is devoted to the sampling scheme of the proposed model conditionally to inequality constraints and noisy observations. To simplify notations, the finite-dimensional approximation of GPs in one dimension is considered

$$Y^N(x) = \sum_{j=0}^{N} Y(t_{N,j})\phi_j(x) = \sum_{j=0}^{N} \zeta_j \phi_j(x), \quad x \in \mathcal{D}.$$

In this paper, the GP is observed with error. The space of noisy observations is defined as

$$\begin{aligned} I_{\mathrm{coef}} &= \left\{ \zeta \in \mathbb{R}^{N+1} \ : \sum_{j=0}^{N} \zeta_j \phi_j(x^{(i)}) = \tilde{y}_i, \ i = 1,\ldots,n \right\} \\ &= \left\{ \zeta \in \mathbb{R}^{N+1} \ : A\zeta = \tilde{\boldsymbol{y}} \right\}, \end{aligned}$$

where $\tilde{y}_i = y_i + \epsilon_i, \ i = 1,\ldots,n, \ \epsilon_i \overset{i.i.d.}{\sim} \mathcal{N}(0, \sigma^2_{\mathrm{noise}})$ and $A_{i,j} = \phi_j(x^{(i)})$. The set of inequality constraints on the coefficients C_{coef} is a convex set (for instance, the nonnegative quadrant $\zeta_j \geq 0, \ j = 0,\ldots,N$ for non-decreasing constraints in one dimension). The sampling scheme can be summarized in two steps: first, the conditional Gaussian vector ζ with only noisy observations is simulated

$$\zeta \mid A\zeta = \tilde{\boldsymbol{y}} \sim \mathcal{N}\left((A\Gamma^N)^\top (A\Gamma^N A^\top + \sigma^2_{\mathrm{noise}}\mathbf{I})^{-1} \tilde{\boldsymbol{y}}, \Gamma^N - (A\Gamma^N)^\top (A\Gamma^N A^\top + \sigma^2_{\mathrm{noise}}\mathbf{I})^{-1} A\Gamma^N \right).$$

Second, by an improved rejection sampling (Maatouk and Bay, 2016), only the random coefficients in the convex set C_{coef} are selected. Now, the three estimates used in the illustrative examples (Section 4) are defined.

Definition 1. *The so-called unconstrained mean is defined as*

$$m^N(x) = E\left(Y^N(x) \mid Y^N(x^{(i)}) = \tilde{y}_i, \ i = 1,\ldots,n \right) = \phi(x)^\top \zeta_{\mathrm{I}},$$

where $\zeta_{\mathrm{I}} = E\left(\zeta \mid \zeta \in I_{\mathrm{coef}}\right) = \Gamma^N A^\top \left(A\Gamma^N A^\top + \sigma^2_{\mathrm{noise}}\mathbf{I} \right)^{-1} \tilde{\boldsymbol{y}}$.

Similarly to the kriging mean of the original GP Y (Eq. (1), $Z = Y$ when η is the null function), the kriging mean m^N of the finite-dimensional approximation of GPs Y^N can be written as

$$m^N(x) = \boldsymbol{k}_N(x)^\top \left(\mathbb{K}_N + \sigma^2_{\text{noise}}\mathbf{I}\right)^{-1} \tilde{\boldsymbol{y}},$$

where $\boldsymbol{k}_N(x) = K_N(x, \boldsymbol{X}) = \left(A\Gamma^N\phi(x)\right)$ is the vector of covariance between $Y^N(x)$ and $Y^N(\boldsymbol{X})$ and $(\mathbb{K}_N)_{i,j} = K_N(x^{(i)}, x^{(j)}) = \left(A\Gamma^N A^\top\right)_{i,j}$, $i, j = 1, \ldots, n$ is the covariance matrix of $Y^N(\boldsymbol{X})$.

Remark 5. *The unconstrained mean $m^N(x)$ respects inequality constraints in the entire domain if and only if the conditional Gaussian vector to only noisy observations ζ_{I} lies inside the convex set C_{coef}.*

Definition 2. *The mean of the posterior distribution of Y^N conditionally to inequality constraints and noisy observations is defined as*

$$m^N_{\text{pos}}(x) = E\left(Y^N(x) \;\middle|\; Y^N(x^{(i)}) = \tilde{y}_i,\ \zeta \in C_{\text{coef}}\right) = \phi(x)^\top \zeta_{\text{pos}},$$

where $\zeta_{\text{pos}} = E\left(\zeta \mid \zeta \in I_{\text{coef}} \cap C_{\text{coef}}\right)$ is the mean of the truncated Gaussian vector which is computed from simulations.

Finally, let μ be the maximum of the probability density function (pdf) of ζ restricted to $I_{\text{coef}} \cap C_{\text{coef}}$. It is the solution of the following convex optimization problem

$$\mu = \arg \min_{x \in I_{\text{coef}} \cap C_{\text{coef}}} \left(\frac{1}{2} x^\top \left(\Gamma^N\right)^{-1} x\right), \tag{5}$$

where Γ^N is the covariance matrix of the Gaussian vector ζ. The quadratic optimization problem (5) is equivalent to

$$\mu = \arg \min_{x \in C_{\text{coef}}} \left(\frac{1}{2} x^\top \left(\Gamma^N_{\text{cond}}\right)^{-1} x + \zeta_{\text{I}}^\top x\right), \tag{6}$$

where Γ^N_{cond} is the covariance matrix of the conditional Gaussian vector $\zeta \mid A\zeta = \tilde{\boldsymbol{y}}$. In fact, μ representes the maximum of the pdf of the Gaussian vector ζ restricted to $I_{\text{coef}} \cap C_{\text{coef}}$ and its numerical calculation is a standard problem in the minimization of positive quadratic forms subject to convex constraints (Boyd and Vandenberghe, 2004; Goldfarb and Idnani, 1983). Let us mention that in all simulated examples illustrated in this paper, the R-package 'solve.QP' described in Goldfarb and Idnani (1983) is used to solve the quadratic convex optimization problems (5)-(6).

Definition 3. *The maximum of the posterior distribution of Y^N conditionally to inequality constraints and noisy observations is defined as*

$$M^N_{\text{pos}}(x) = \sum_{j=0}^{N} \mu_j \phi_j(x), \quad x \in \mathbb{R}^d,$$

where $\mu = (\mu_0, \ldots, \mu_N)^\top$ is computed by (6).

Remark 6. *The maximum a posteriori estimate M^N_{pos} does not depend on the variance hyper-parameter σ of the covariance function K as well as on the simulations but depends on the length hyper-parameters of the covariance function $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_d)$.*

Remark 7. *In the case where the GP is observed without error (i.e., with noise-free data), the maximum a posteriori estimate M^N_{pos} converges uniformly to the constrained interpolation function solution of the following convex optimization problem*

$$\arg \min_{h \in H \cap I \cap C} \|h\|^2_H,$$

where H is the reproducing kernel Hilbert space (RKHS) associated to the positive type kernel K (Aronszajn, 1950; Berlinet and Thomas-Agnan, 2004), I is the set of functions verify interpolation conditions and the convex set C is the space of functions verify inequality constraints (Bay et al., 2016, 2017).

This generalizes to the case of interpolation conditions and inequality constraints the well known correspondence established by Kimeldorf and Wahba (1970) between Bayesian estimation on stochastic process and smoothing by splines.

In Algorithm 1, the sampling scheme of the proposed model is described. It is based on the rejection sampling from the Mode (RSM) algorithm to simulate the Gaussian vector ζ restricted to the convex set $I_{\text{coef}} \cap C_{\text{coef}}$ (see, Maatouk and Bay (2016) for more details).

4 Illustrative examples

The goal of this section is twofold: first, to illustrate the condition simulation of the GP approximation developed in the present paper with certain constraints such as boundedness, positivity and monotonicity in one and two dimensions and noisy observations. Second, to describe the two different cases in the simulation.

Algorithm 1: Sampling scheme

Initialization:
$\zeta \notin C_{\text{coef}}$; $\zeta \leftarrow \zeta_{\text{current}}$
$1 \leftarrow$ unif; $0 \leftarrow t$
while unif $> t$ **do**
  $\zeta \leftarrow \zeta_{\text{current}}$
  while $\zeta_{\text{current}} \notin C_{\text{coef}}$ **do**
    $\mathcal{N}(\mu, \Gamma^N_{\text{cond}}) \leftarrow \zeta_{\text{current}}$
  end
  $\exp\left(\mu^\top (\Gamma^N_{\text{cond}})^{-1}(\mu - \zeta_{\text{I}} - \zeta_{\text{current}}) + \zeta^\top_{\text{current}}(\Gamma^N_{\text{cond}})^{-1}\zeta_{\text{I}}\right) \leftarrow t$
  $\mathcal{U}(0,1) \leftarrow$ unif
end

- The unconstrained mean respects the constraints and then coincides with the maximum of the posterior distribution.
- The unconstrained mean does not respect the constraints, then the unconstrained mean and the maximum of the posterior distribution are different.

The Matérn 3/2 and squared exponential (or Gaussian) covariance functions are used (Table 1).

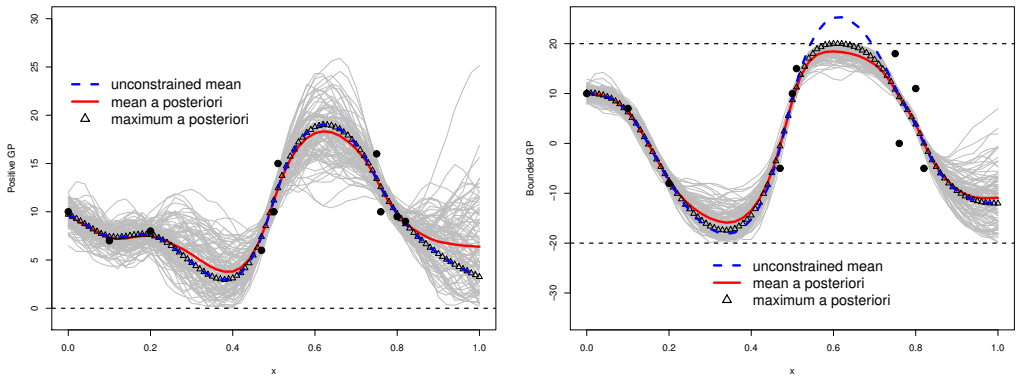
4.1 Boundedness constraints

The real function is supposed to respect boundedness constraints

$$C = \left\{ f \in \mathcal{C}^0([0,1]) \ : \ -\infty \leq a \leq f(x) \leq b \leq +\infty, \ x \in [0,1] \right\}. \quad (7)$$

The constrained data of size $n = 10$ (black points in Fig. 1) are not taken from constrained functions. The noise variance is fixed to $\sigma^2_{\text{noise}} = 1.1^2$. Additionally, the Matérn 3/2 covariance function is used with the hyper-parameters fixed to $(\theta, \sigma) = (0.3, 10)$.

In Fig. 1a, we generate one hundred sample paths taken from model (2) with $d = 1$ and $N = 50$ conditionally to positivity constraints (i.e., $a = 0$ and $b = +\infty$ in (7)). The simulated trajectories (gray lines) respect positivity constraints in the entire domain as well as the mean of the posterior distribution. The unconstrained mean and the maximum of the posterior distribution coincide and respect positivity constraints in the entire domain as well: it corresponds to the situation where the conditional Gaussian vector ζ_{I} lies inside the acceptance region C_{coef} (Remark 5). In Fig. 1b, the



(a) (b)

Figure 1: The GP approximation with positivity constraints (a) and boundedness constraints (b). The unconstrained mean coincides with the maximum *a posteriori* in (a) but not in (b).

boundedness constraint is considered (i.e., $a = -20$ and $b = 20$ in (7)). The simulated trajectories (gray lines) respect boundedness constraints in the entire domain as well as the mean and the maximum of the posterior distribution, contrarily to the unconstrained mean. This is the case where ζ_{I} lies outside the acceptance region C_{coef} (Remark 5). This numerical result can be seen as a generalization of the Kimeldorf-Wahba correspondence Kimeldorf and Wahba (1970) in the case of inequality constraints and errors measurements between Bayesian estimation on stochastic process and smoothing by splines.

4.2 Monotonicity constraints

The monotone (non-decreasing) function $f(x) = 0.32(x + sin(x)),\ x \in [0, 10]$ used in the literature to compare different models is considered. It is evaluated at data of size $n = 50$ chosen randomly on $[0, 10]$ (black points in Fig. 2) with standard deviation $\sigma_{\mathrm{noise}} = 1$.

In Fig. 2, we generate one hundred sample paths taken from model (4) with $N = 50$ conditionally to monotonicity (non-decreasing) constraints. The squared exponential covariance function is used with hyper-parameters $(\theta, \sigma) = (2.5, 1)$. Notice that, the simulated trajectories (gray lines) are non-decreasing in the entire domain as well as the mean and the maximum of the posterior distribution, contrarily to the unconstrained mean. It corresponds to the case where the conditional Gaussian vector ζ_{I} lies outside the

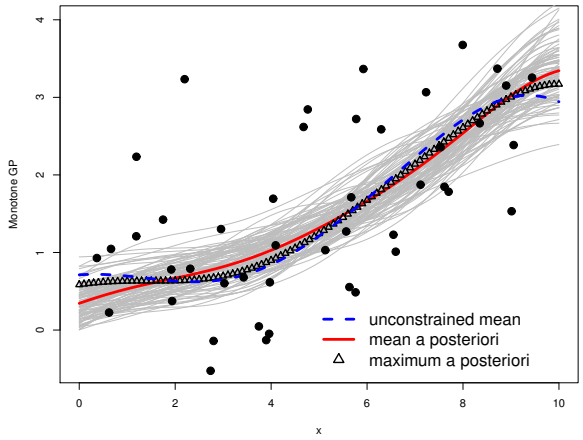


Figure 2: The GP approximation (4) with monotonicity constraints for sinusoidal function $f(x) = 0.32(x + sin(x))$. The unconstrained mean does not coincide with the maximum *a posteriori*.

acceptance region C_{coef} (Remark 5).

4.3 Isotonicity in two dimensions

In two dimensions, the monotone (non-decreasing) function with respect to the two inputs used in Saarela and Arjas (2011); Shively et al. (2009)

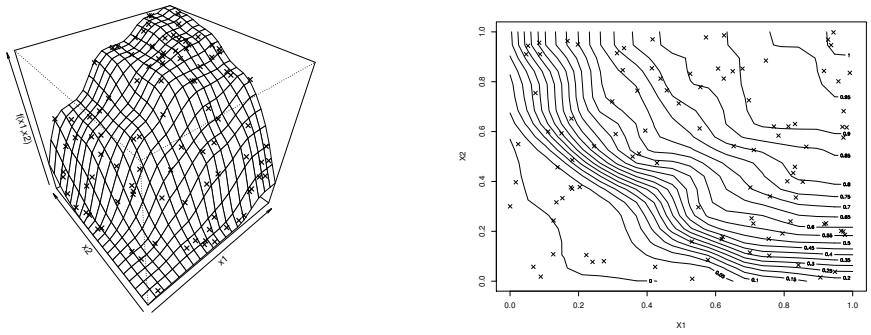
$$f(x_1, x_2) = \mathbb{1}_{\{(x-1-1)^2+(x_2-1)^2<1\}}\{1-(x_1-1)^2-(x_2-1)^2\}^{1/2}, \quad (x_1, x_2) \in [0,1]^2$$

is considered. It is evaluated at data of size $n = 100$ chosen randomly on $[0,1]^2$ with standard deviation $\sigma_{\text{noise}} = 0.1$. In Fig. 3, the two-dimensional squared exponential covariance function is used

$$K(\boldsymbol{x}, \boldsymbol{x}') = \sigma^2 \exp\left(-\frac{(x_1 - x_1')^2}{2\theta_1^2}\right) \times \exp\left(-\frac{(x_2 - x_2')^2}{2\theta_2^2}\right), \tag{8}$$

where the variance hyper-parameter $\sigma = 1$ and the length hyper-parameters $(\theta_1, \theta_2) = (0.02, 0.17)$ are estimated using cross-validation methods (Maatouk et al., 2015). Figure 3 shows the maximum of the posterior distribution using model (3) with $N = 10$ and the associated contour levels. It respects monotonicity (non-decreasing) constraints with respect to the two inputs.

Remark 8. *For monotonicity with respect to only one variable, model* (3) *(with noise-free data) has been used in Cousin et al. (2016) to estimate the*



(a) (b)

Figure 3: The maximum of the posterior distribution drawn from model (3) respecting monotonicity (non-decreasing) constraints for the two inputs, and the associated contour levels.

discount factor surface as a function of time-to-maturities and quotation dates. It is a monotone (non-increasing) function with respect to time-to-maturities at each quotation date.

5 Simulation study

In this section, a comparison between the finite-dimensional approximation of GPs developed in the present paper and models deal with monotonicity and isotonicity constraints is shown. 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

- flat function $f_1(x) = 3,\ x \in (0, 10]$;
- sinusoidal function $f_2(x) = 0.32\{x + sin(x)\},\ x \in (0, 10]$;
- step function $f_3(x) = 3$ if $x \in (0, 8]$ and $f_3(x) = 8$ if $x \in (8, 10]$;
- linear function $f_4(x) = 0.3x,\ x \in (0, 10]$;
- exponential function $f_5(x) = 0.15 \exp(0.6x - 3),\ x \in (0, 10]$;
- logistic function $f_6(x) = 3/\{1 + \exp(-2x + 10)\},\ x \in (0, 10]$.

Table 2: Length hyper-parameter estimates using a suited cross-validation method.

	Flat	Step	Linear	Exponential	Logistic	Sinusoidal
$\hat{\theta}$	100.0	0.8	8.6	1.0	2.0	2.5

Table 3: 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
Gaussian process	15.1	27.1	16.7	19.7	25.5	21.9
Gaussian process 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
Gaussian process approximation	8.2	41.1	15.8	20.8	21.0	20.6

These functions are supposed to be evaluated at data of size $n = 100$ with standard deviation $\sigma_{\text{noise}} = 1$. The root-mean-square error (RMSE) of the estimates is computed at the one hundred x values taken uniformly (equidistant) in the interval $(0, 10]$:

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

where $\hat{f}(x)$ is the estimate of $f(x)$ and x_i are the n equally-spaced x-values. For the GP approximation developed in this paper, the maximum *a posteriori* estimate (Definition 3) is used as an estimate of $f(x)$, where N is fixed to fifty. Let us recall that this estimate depends only on the length hyper-parameter θ. The squared exponential covariance function (Table 1) is used in the simulation, with σ fixed to 1 and θ estimated using the suited cross-validation method (Maatouk et al., 2015; Cousin et al., 2016). Table 2 shows the values of the parameter estimation $\hat{\theta}$.

In Table 3, the RMSE of the estimates is calculated for the finite-dimensional approximation of GPs, and it is compared with results of Gaussian process with and without projection given in Lin and Dunson (2014) and results of the regression spline method given in Shively et al. (2009). To ensure stability of results, the simulations have been repeated 5000 times. Table 3 shows that the finite-dimensional approximation of GPs outperforms regression splines (resp. Gaussian process with and without projection) except in the step and logistic cases (resp. in the step and exponential cases).

Remark 9. *Let us recall that the finite-dimensional approximation of GPs developed in the present paper is supposed centered (i.e., mean-zero). To be*

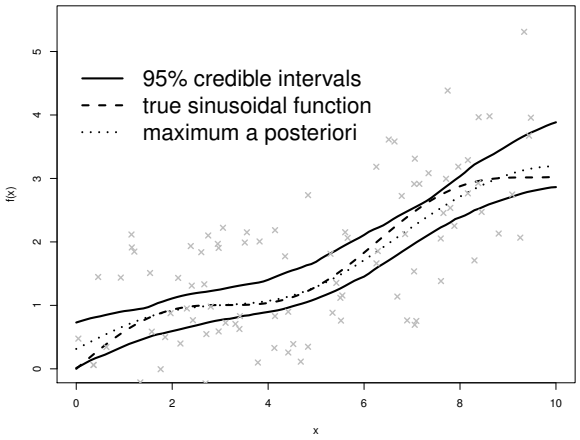


Figure 4: The 95% credible intervals of the Gaussian process approximation together with the sinusoidal function, the observations (grey crosses) and the maximum *a posteriori* estimate.

coherent, the results presented in Table 3 should be computed when the output values are normalized

$$\bar{y}_i = y_i - \bar{y}, \quad i = 1, \ldots, n,$$

where $\bar{y}_i$ *is the normalized value of the* i*th output observation and* $\bar{y} = 1/n \sum_{i=1}^{n} y_i$ *is the mean of the output observations. In that case, the finite-dimensional approximation of GPs outperforms regression spline and Gaussian process with and without projection except in the step case.*

Now, the uncertainty quantification is investigated. The monotone (non-decreasing) function $f(x) = 0.32(x + sin(x))$, $x \in (0, 10]$ (sinusoidal function) is considered (dashed lines in Fig. 4). It is evaluated at data of size $n = 100$ distributed randomly on $(0, 10]$ (grey crosses in Fig. 4), with standard deviation $\sigma_{\text{noise}} = 1$.

In Table 4, the percentage of the empirical coverage of 95% pointwise credible intervals of GP approximation is computed by repeating the simulation 1000 times. The coverage for Gaussian process approximation is closer to the nominal 95% than is that of the Gaussian process at most of input locations chosen by Lin and Dunson (2014). Additionally, the finite-dimensional approximation of GPs outperforms Gaussian process with projection at some input locations and slightly bad at the other locations.

To compare the proposed approach with the methodology based on the knowledge of the derivatives of the GP at some input locations, the logistic

Table 4: 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
Gaussian process	97.3	94.6	91.8	88.0	90.5	95.2	96.8	91.0	86.5	86.3
Gaussian process projection	94.1	95.4	92.0	89.5	93.1	94.6	96.0	90.0	89.0	86.9
Gaussian process approximation	97.0	93.0	89.6	90.1	94.1	97.1	95.5	89.5	85.4	86.7

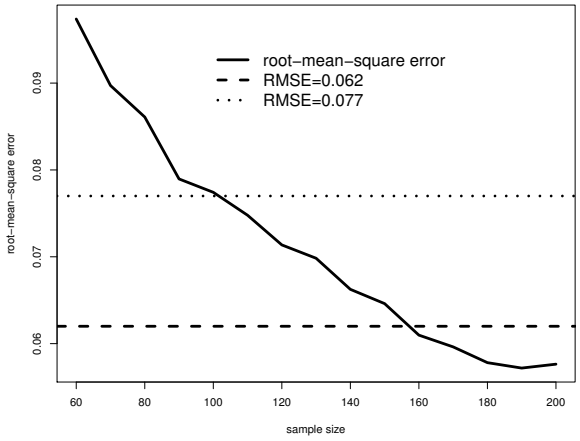


Figure 5: The root-mean-square error at different sample sizes together with the optimal values obtained in Riihimäki and Vehtari (2010).

artificial function $f(x) = 2/(1+\exp(-8x+4)),\ x \in [0,1]$ defined in Riihimäki and Vehtari (2010) is considered. This function is supposed to be evaluated at data of size n with standard deviation $\sigma_{\text{noise}} = 0.5$. The squared exponential covariance function is used.

In Riihimäki and Vehtari (2010), the RMSE is equal to 0.077 (resp. 0.062) for $n = 100$ (resp. $n = 200$). In Fig. 5, the root-mean-square error using the GP approximation is illustrated at different sample sizes together with the optimal values obtained by Riihimäki and Vehtari (2010). Notice that, we just need data of size $n = 160$ to reach the optimal value 0.062 obtained by Riihimäki and Vehtari (2010). The results are based on 1000 simulation replicates.

The isotonicity (non-decreasing) functions with respect to the two inputs

Table 5: The summary of length hyper-parameter estimates in two dimensions using cross-validation methods.

	f_1	f_2	f_3	f_4	f_5	f_6
$(\hat{\theta}_1, \hat{\theta}_2)$	(0.17,0.38)	(0.46,1.32)	(0.18,0.22)	(0.38,0.01)	(0.08,0.09)	(0.02,0.17)

Table 6: Mean square error ($\times$ 100) for data of size $n = 1024$ with standard deviation $\sigma_{\text{noise}} = 0.1$. The results are based on 100 simulation replicates.

	f_1	f_2	f_3	f_4	f_5	f_6
Gaussian process projection	0.04	0.02	0.05	0.20	0.19	0.10
Gaussian process approximation	2.86e-3	4.40e-4	7.09e-3	0.55	0.34	0.04

used in Lin and Dunson (2014); Saarela and Arjas (2011) are considered

$$
\begin{aligned}
&f_1(x_1, x_2) = \sqrt{x_1},\ (x_1, x_2) \in [0,1]^2;\\
&f_2(x_1, x_2) = 0.5x_1 + 0.5x_2,\ (x_1, x_2) \in [0,1]^2;\\
&f_3(x_1, x_2) = \min(x_1, x_2),\ (x_1, x_2) \in [0,1]^2;\\
&f_4(x_1, x_2) = 0.25x_1 + 0.25x_2 + 0.5 \times \mathbb{1}_{\{x_1+x_2>1\}},\ (x_1, x_2) \in [0,1]^2;\\
&f_5(x_1, x_2) = 0.25x_1 + 0.25x_2 + 0.5 \times \mathbb{1}_{\{\min(x_1,x_2)>5\}},\ (x_1, x_2) \in [0,1]^2;\\
&f_6(x_1, x_2) = \mathbb{1}_{\{(x_1-1)^2+(x_2-1)^2<1\}} \sqrt{1-(x_1-1)^2-(x_2-1)^2},\ (x_1, x_2) \in [0,1]^2.
\end{aligned}
$$

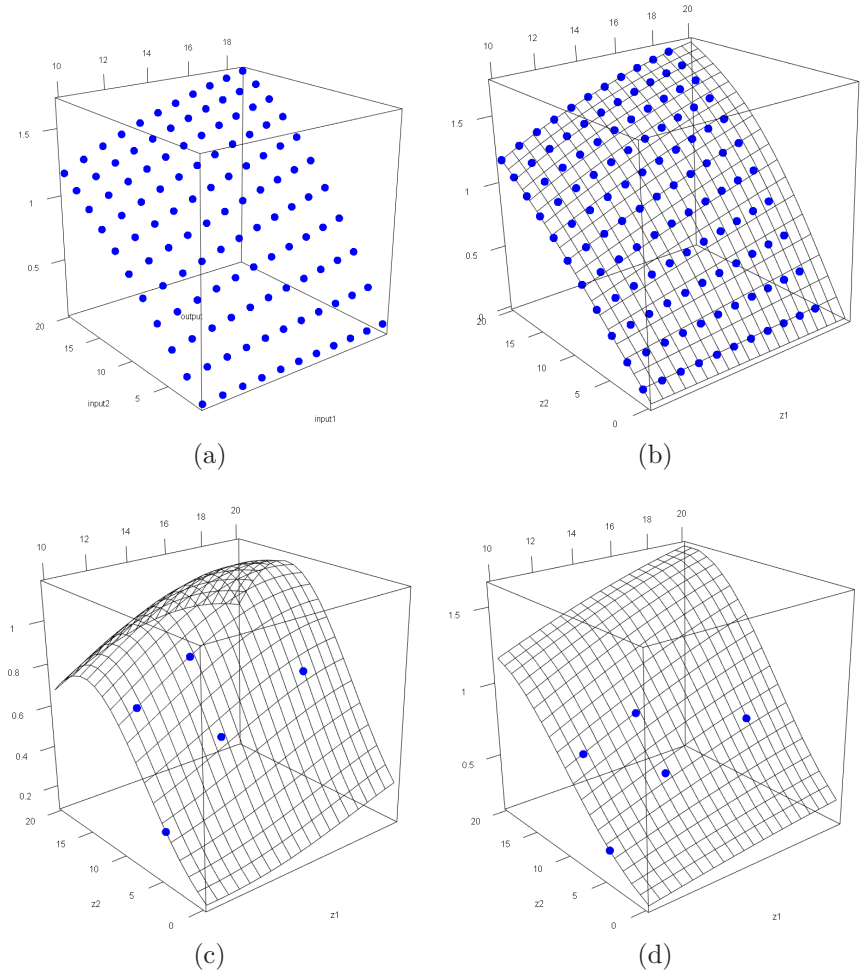
The two-dimensional squared exponential covariance function (8) is used, with σ fixed to 1 and (θ_1, θ_2) estimated using the suited cross-validation method (Maatouk et al., 2015; Cousin et al., 2016). Table 5 shows the values of the parameter estimation $(\hat{\theta}_1, \hat{\theta}_2)$.

In Table 6, the mean square error (MSE) of the estimates is calculated for the finite-dimensional approximation of GPs, and it is compared with results of Gaussian process projections given in Lin and Dunson (2014). Table 6 shows that the finite-dimensional approximation of GPs outperforms Gaussian process projections except in f_4 and f_5 cases. This is very similar to the one-dimensional case results, because of the similarity of f_4 and f_5 functions to the step case.

6 Real application: nuclear safety

In this section, the performance of the proposed model has been investigated using the real-word data provided by the 'Institut of Radioprotection and Nuclear Safety' (IRSN), France. The nuclear reactor of the uranium sphere called 'Lady Godiva device' situated at Los Alamos National Laboratory

(LANL), New Mexico, U.S. has been studied. The nuclear reactor of the sphere is increasing with respect to the two considered input parameters: its radius (between 0 and 20 cm) and density (between 10 and 20 g/cm^3). The one hundred and tweenty one observations defined on $[0, 20] \times [10 \times 20]$



(a) (b)

(c) (d)

Figure 6: (a) 3D visualization of Godiva's data. (b) the maximum of the posterior distribution with 121 observations. (c) the unconstrained GP model with fives observations. (d) the constrained GP model with fives observations.

(see Figure 6a) have been used to show the efficiency of the proposed model in term of prediction and to compare it with the unconstrained Gaussian process model. The idea is to fix some of these observations and to test the quality of prediction of the proposed estimator at the other ones. The squared exponential covariance function has been used. In Figure 6c, we

fix five observations using maximin Latin hypercube and we plot the unconstrained mean with the hyper parameters estimated by cross-validation methods. Notice that the unconstrained mean does not respect monotonicity constraints in the entire domain. In Figure 6d, the same observations have been used to plot the maximum of the posterior distribution. The hyper parameters (θ_1, θ_2) have been estimated by the suited cross-validation method Maatouk et al. (2015). We remark that with few observations, the proposed estimator verifies monotonicity constraints in the entire domain. Finally, the Q^2 criteria has been used to evaluate the quality of predictions

$$Q^2 = 1 - \frac{\sum_{i=1}^{n_t}(f(x_i) - \hat{f}(x_i))^2}{\sum_{i=1}^{n_t}(f(x_i) - \overline{y})^2},$$

where $\hat{f}$ is the proposed estimator, $\overline{y}$ is the mean of the observations and n_t is the number of tested data. The constrained model outperforms the unconstrained one with Q^2 equal to 0.98 versus 0.69.

7 Conclusion

In this paper, a finite-dimensional approximation of Gaussian processes to incorporate infinite number of inequality constraints (such as boundedness, monotonicity and convexity) and noisy observations is developed. It is based on a linear combination between Gaussian random coefficients and deterministic basis functions. The basis functions are chosen such that the infinite number of inequality constraints on the Gaussian process approximation are equivalent to a finite number of constraints on the coefficients. Consequently, simulate the conditional approximating process is equivalent to simulate a truncated Gaussian vector restricted to convex sets. By this methodology, the mean and the maximum of the posterior distribution are well defined. To show the performance of the proposed model in term of predictive accuracy and uncertainty quantification, a comparison with several recently models dealing with the same constraints is shown.

Acknowledgements

Part of this work has been conducted within the frame of the ReDice Consortium, gathering industrial (CEA, EDF, IFPEN, IRSN, Renault) and academic (École des Mines de Saint-Étienne, INRIA, and the University of Bern) partners around advanced methods for Computer Experiments. The author would like to thank Yann Richet (IRSN, Paris) for providing the nuclear safety data.

References

Abrahamsen, P. and Benth, F. E. (2001). Kriging with inequality constraints. *Math. Geo.*, 33(6):719–744.

Aronszajn, N. (1950). Theory of reproducing kernels. *Trans. Am. Math. Soc.*, 68:337–404.

Bay, X., Grammont, L., and Maatouk, H. (2016). Generalization of the Kimeldorf-Wahba correspondence for constrained interpolation. *Electron. J. Statist.*, 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. *Comput. Optim. Appl.*, 68(1):95–120. doi:10.1007/s10589-017-9906-9.

Berlinet, A. and Thomas-Agnan, C. (2004). *Reproducing kernel Hilbert spaces in probability and statistics*. Kluwer Academic Publishers.

Botts, C. (2013). An accept-reject algorithm for the positive multivariate normal distribution. *Comput. Statist.*, 28(4):1749–1773.

Boyd, S. and Vandenberghe, L. (2004). *Convex optimization*. Cambridge University Press.

Chopin, N. (2011). Fast simulation of truncated Gaussian distributions. *Stat. Comput.*, 21(2):275–288.

Cousin, A., Maatouk, H., and Rullière, D. (2016). Kriging of financial term-structures. *Eur. J. Oper. Res.*, 255(2):631 – 648.

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

Cressie, N. and Johannesson, G. (2008). Fixed rank kriging for very large spatial data sets. *J. R. Stat. Soc. B*, 70(1):209–226.

Delecroix, M., Simioni, M., and Thomas-Agnan, C. (1996). Functional estimation under shape constraints. *J. Nonparametr. Stat.*, 6(1):69–89.

Freulon, X. and de Fouquet, C. (1993). Conditioning a Gaussian model with inequalities. In Soares, A., editor, *Geostatistics Tróia '92: Volume 1*, pages 201–212. Springer Netherlands, Dordrecht.

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.

Goldfarb, D. and Idnani, A. (1983). A numerically stable dual method for solving strictly convex quadratic programs. *Math. Program.*, 27(1):1–33.

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

Jidling, C., Wahlström, N., Wills, A., and Schön, T. B. (2017). Linearly constrained Gaussian processes. *arXiv preprint arXiv:1703.00787*.

Kimeldorf, G. S. and Wahba, G. (1970). A correspondence between Bayesian estimation on stochastic processes and smoothing by splines. *Ann. Math. Stat* ., 41(2):495–502.

Kleijnen, J. P. and Van Beers, W. C. (2013). Monotonicity-preserving bootstrapped kriging metamodels for expensive simulations. *J. Oper. Res. Soc.*, 64(5):708–717.

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.

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

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

Parzen, E. (1962). *Stochastic processes*. Holden-Day series in probability and statistics. Holden-Day, San Francisco, London, Amsterdam.

Philippe, A. and Robert, C. P. (2003). Perfect simulation of positive Gaussian distributions. *Stat. Comput.*, 13(2):179–186.

Ramsay, J. O. (1988). Monotone regression splines in action. *Statist. Sci.*, 3(4):425–441.

Ramsay, J. O. (1998). Estimating smooth monotone functions. *J. R. Stat. Soc. B*, 60(2):365–375.

Rasmussen, C. E. and Williams, C. K. (2006). *Gaussian processes for machine learning*. MIT Press, Cambridge.

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

Robert, C. P. (1995). Simulation of truncated normal variables. *Stat. Comput.*, 5(2):121–125.

Saarela, O. and Arjas, E. (2011). A method for Bayesian monotonic multiple regression. *Scand. J. Statist.*, 38(3):499–513.

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.

Trecate, G. F., Williams, C. K., and Opper, M. (1999). Finite-dimensional approximation of Gaussian processes. In *Proceedings of the 1998 conference on Advances in neural information processing systems II*, pages 218–224. MIT Press.

Tutz, G. and Leitenstorfer, F. (2007). Generalized smooth monotonic regression in additive modeling. *J. Comp. Graph. Stat.*, 16(1):165–188.

Wang, X. and Berger, J. O. (2016). Estimating shape constrained functions using Gaussian processes. *SIAM/ASA Journal on Uncertainty Quantification*, 4(1):1–25.

Xuming, H. and Peide, S. (1996). Monotone B-spline smoothing. *J. Amer. Statist. Assoc.*, 93:643–650.