Finite-dimensional approximation of Gaussian processes with inequality constraints

Hassan Maatouk

Finite-dimensional approximation of Gaussian processes with inequality constraints

Hassan Maatouk

IINRIA Centre de Recherche Rennes - Bretagne Atlantique, Campus de Beaulieu, 35042 Rennes, France

E-mail: hassan.maatouk@inria.fr

Summary. 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); Da Veiga and Marrel (2012), 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), the generalization of monotonic regression to multiple dimensions is studied.

The methodology developed in the present paper is quite different. It is based on a finite-dimensional approximation of GPs (or a GP approximation) 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. In Section 5, the performance of the proposed model in terms of predictive accuracy and uncertainty quantification is investigated.

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 $\boldsymbol{x}^{(i)} \in \mathbb{R}^d$, $1 \leq i \leq n$. 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 $\epsilon \sim \mathcal{N}(\mathbf{0}, \sigma^2_{\text{noise}}\mathbf{I})$ 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 $Z \sim \mathcal{GP}(\eta(\boldsymbol{x}), K(\boldsymbol{x}, \boldsymbol{x}'))$. Conditionally to noisy observations $\tilde{\boldsymbol{y}} = (\tilde{y}_1, \ldots, \tilde{y}_n)^\top$, the process remains a GP

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

where

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

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 $\boldsymbol{k}(\boldsymbol{x}) = K(\boldsymbol{x}, \boldsymbol{x}^{(i)})$ is the vector of covariance between $Z(\boldsymbol{x})$ and $Z(\boldsymbol{X})$. 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 $Z(\boldsymbol{X}) = \tilde{\boldsymbol{y}}$ (Rasmussen and Williams, 2006).

2.1. The choice of covariance function

The choice of the covariance function K has crucial consequences specially in controlling the smoothness of the kriging metamodel. It must be chosen in the set of definite and positive kernels. Some popular covariance functions used in kriging methods are given in Table 1. 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,

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	$\sigma^2 \exp\left(-\frac{\lvert x-x'\rvert}{\theta}\right)$	$\mathcal{C}^0$

1967). Since differentiation is a linear operator, the order partial derivatives of a GP remain GPs (Cramer and Leadbetter, 1967; Parzen, 1962) and

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

3. Finite-dimensional approximation of GPs

Let $Y \sim \mathcal{GP}(0, K(\boldsymbol{x}, \boldsymbol{x}'))$ be a zero-mean GP with covariance function K. In this paper, the finite-dimensional approximation of Gaussian processes developed in Maatouk and Bay (2017) and applied to finance and insurance in Cousin et al. (2016) is considered

$$Y^N(\boldsymbol{x}) = \sum_{j=0}^{N} \xi_j \phi_j(\boldsymbol{x}), \quad \boldsymbol{x} \in \mathcal{D} \subset \mathbb{R}^d, \tag{2}$$

where $\xi = (\xi_0, \ldots, \xi_N)^\top$ is a zero-mean Gaussian vector with covariance matrix Γ^N and $\phi = (\phi_0, \ldots, \phi_N)^\top$ is a vector of deterministic basis functions. In next subsections, the covariance matrix Γ^N is computed explicitly which depends on the covariance function K of the original GP Y and the choice of the basis functions is studied. The covariance function $K_N(\boldsymbol{x}, \boldsymbol{x}')$ of the Gaussian process approximation Y^N is equal to

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

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 ξ restricted to

$$\begin{aligned}&\sum_{j=0}^{N} \xi_j \phi_j(\boldsymbol{x}^{(i)}) = y_i + \epsilon_i = \tilde{y}_i, \quad i = 1, \ldots, n,\\ &\xi \in C_{\text{coef}},\end{aligned}$$

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 deduced from the choice of the basis functions. Next, we show how the

basis functions can be chosen and how the covariance matrix of the coefficients can be computed.

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 ξ_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. One-dimensional cases

The input $x \in \mathbb{R}$ and without loss of generality in the unit interval $\mathcal{D} = [0, 1]$.

3.1.1. Boundedness constraints

In this subsection, the real function is supposed continuous and belong to the convex set

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

The input set $\mathcal{D}$ is discretized uniformly to $(N+1)$ nodes $0 = t_{N,0}, \ldots, t_{N,N} = 1$ but the methodology can be adapted to non-uniform subdivision easily. The finite-dimensional approximation of GPs is defined as follow

$$Y^N(x) := \sum_{j=0}^{N} Y(t_{N,j}) h_j(x) = \sum_{j=0}^{N} \xi_j h_j(x), \quad x \in \mathcal{D}, \tag{3}$$

where $\xi_j = Y(t_{N,j})$ and $(h_j)_j$ are the hat functions associated to the nodes $t_{N,j}$: $h_j(x) = h((x - t_{N,j})/\Delta_N)$, where $\Delta_N = 1/N$ and $h(x) = (1 - |x|)\, \mathbb{1}_{(|x| \leq 1)}$, $x \in \mathbb{R}$. The value of any basis function at any node is equal to Kronecker's Delta function ($h_j(t_{N,k}) = \delta_{j,k}$, $j, k = 0, \ldots, N$), where $\delta_{j,k}$ is equal to one if $j = k$ and zero otherwise.

PROPOSITION 1. *If the realizations of the original GP Y are continuous, then the finite-dimensional approximation of GPs defined in* (3) *verifies the following properties:*

- *Y^N is a finite-dimensional GP† with covariance function $K_N(x, x') = h(x)^\top \Gamma^N h(x')$, where $h(x) = (h_0(x), \ldots, h_N(x))^\top$, $\Gamma^N_{i,j} = K(t_{N,i}, t_{N,j})$ and K is the covariance function of the original GP Y.*
- *Y^N is almost surely converge uniformly to Y when N tends to infinity.*
- *Y^N is in C if and only if the $(N+1)$ coefficients $Y(t_{N,j})$ are contained in $[a, b]$.*

From this proposition, the advantage of the proposed model is shown. In fact, the infinite number of inequality constraints of Y^N are reduced to a finite number of linear inequality constraints on the random coefficients $(Y(t_{N,j}))_j$. Thus, simulate Y^N with boundedness constraints and noisy observations is equivalent to simulate the truncated Gaussian vector $\xi = (Y(t_{N,0}), \ldots, Y(t_{N,N}))^\top$ restricted to a convex set formed by

$$Y^N(x^{(i)}) = \sum_{j=0}^{N} \xi_j h_j(\boldsymbol{x}^{(i)}) = \tilde{y}_i, \quad i = 1, \ldots, n,$$

$$\xi \in C_{\text{coef}} = \{ \xi \in \mathbb{R}^{N+1} \ : \ a \leq \xi_j \leq b, \ j = 0, \ldots, N \}.$$

†A GP with paths lying in a finite-dimensional space is called a finite-dimensional GP.

PROOF (PROOF OF PROPOSITION 1). Since Y^N is a linear combination of $(N+1)$ Gaussian variables $Y(t_{N,j}),\ j=0,\dots,N$, then it is a GP with dimension equal to $N+1$ and covariance between two input variables

$$\begin{aligned} K_N(x,x') = \mathrm{Cov}\left(Y^N(x), Y^N(x')\right) &= \sum_{i,j=0}^{N} \mathrm{Cov}\left(Y(t_{N,i}), Y(t_{N,j})\right) h_i(x) h_j(x') \\ &= \sum_{i,j=0}^{N} K(t_{N,i}, t_{N,j}) h_i(x) h_j(x'). \end{aligned}$$

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

$$Y^N(x;\omega) = \sum_{j=0}^{N} Y(t_{N,j};\omega) h_j(x).$$

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

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

Thus, on 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 $\mathcal{D}$. Now, if the $(N+1)$ coefficients $(Y(t_{N,j}))_{0\leq j\leq N}$ are in the interval $[a,b]$, then the piecewise linear approximation Y^N is in C. Conversely, suppose that Y^N is in C, then, for $i=0,\dots,N$

$$Y^N(t_{N,i}) = \sum_{j=0}^{N} Y(t_{N,j}) h_j(t_{N,i}) = \sum_{j=0}^{N} Y(t_{N,j}) \delta_{ij} = Y(t_{N,i}) \in [a,b],$$

which completes the proof of the last property, and hence the proof of the proposition.

3.1.2. Monotonicity constraints

The real function f is supposed at least differentiable. Let C be the space of functions verify monotonicity constraints

$$C = \left\{ f : \mathcal{C}^1(\mathcal{D}) \longrightarrow \mathbb{R} \ :\ f'(x) \geq 0,\ x \in \mathcal{D} \right\}.$$

In that case, the finite-dimensional approximation of GPs is defined as

$$Y^N(x) := Y(0) + \sum_{j=0}^{N} Y'(t_{N,j})\phi_j(x) = \zeta + \sum_{j=0}^{N} \xi_j \phi_j(x), \tag{4}$$

where $\zeta = Y(0)$, $\xi_j = Y'(t_{N,j})$ and $(\phi_j)_j$ are the basis functions defined as the primitive functions of the hat functions h_j

$$\phi_j(x) := \int_0^x h_j(t)dt, \quad x \in \mathcal{D}.$$

Similar to the hat functions, the derivative of the basis functions ϕ_j at any node $t_{N,i},\ i = 0, \ldots, N$ is equal to the Kronecker's delta ($\phi_j'(t_{N,i}) = \delta_{ij}$).

Proposition 2. *Suppose that the realizations of the original GP Y are almost surely continuously differentiable. The finite-dimensional approximation of GPs $\left(Y^N(x)\right)_{x\in\mathcal{D}}$ defined in* (4) *verifies the following properties:*

- *Y^N is a finite-dimensional GP with covariance function*

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

where $\phi(x) = (\phi_0(x), \ldots, \phi_N(x))^\top$ and $\tilde{\Gamma}^N$ is the covariance matrix of the Gaussian vector $(\zeta, \xi) = (Y(0), Y'(t_{N,0}), \ldots, Y'(t_{N,N}))^\top$

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

with $\Gamma^N_{i,j} = \frac{\partial^2 K}{\partial x \partial x'}(t_{N,i}, t_{N,j})$ and K the covariance function of the original GP Y.

- *Y^N is almost surely converge uniformly to Y when N tends to infinity.*
- *Y^N is non-decreasing (resp. non-increasing) if and only if the coefficients $Y'(t_{N,j})$ are all nonnegative (resp. nonpositive).*

Similar to boundedness constraints, the infinite number of inequality constraints of Y^N are reduced to a finite number of linear inequality constraints on the random coefficients $(Y'(t_{N,j}))_{0\le j\le N}$. From the last property in Proposition 2, the simulation of Y^N with monotonicity constraints and noisy observations is equivalent to the simulation of the truncated Gaussian vector (ζ, ξ) restricted to a convex set formed by

$$\begin{aligned} Y^N(x^{(i)}) &= \zeta + \sum_{j=0}^{N} \xi_j \phi_j(x^{(i)}) = \tilde{y}_i, \quad i = 1, \ldots, n, \\ (\zeta, \xi) &\in C_{\text{coef}} = \left\{(\zeta, \xi) \in \mathbb{R}^{N+2} \ :\ \xi_j \ge 0,\ j = 0, \ldots, N\right\}. \end{aligned}$$

Proof (Proof of Proposition 2). The proof of this proposition is similar to the boundedness constraints case. The first property is a consequence of the fact that the derivative of a GP remains GP. For all $x, x' \in \mathcal{D}$

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

Let us write that for any $\omega \in \Omega$

$$Y^N(x;\omega) = Y(0;\omega) + \int_0^x \left(\sum_{j=0}^{N} Y'(t_{N,j};\omega) h_j(t) \right) dt.$$

The almost sure uniform convergence of Y^N to Y can be deduced from Proposition 1. In fact, $\sum_{j=0}^{N} Y'(t_{N,j};\omega)h_j(x)$ converges uniformly pathwise to $Y'(x;\omega)$ since the realizations of the process are almost surely continuously differentiable.

Remark 1. *The monotonicity constraints can be obtained with model* (3)*. The GP approximation* $(Y^N(x))_{x\in\mathcal{D}}$ *is non-decreasing if and only if the sequence of coefficients* $(Y(t_{N,j}))_j$ *is non-decreasing (i.e.,* $Y(t_{N,j-1}) \leq Y(t_{N,j}),\ j = 1,\ldots,N$*). In that case, we have* $C_{\mathrm{coef}} = \{\xi \in \mathbb{R}^{N+1} \ :\ \xi_{j-1} \leq \xi_j,\ j = 1,\ldots,N\}$.

3.1.3. Convexity Constraints

The realizations of the original GP Y are assumed to be at least twice differentiable. The finite-dimensional approximation of GPs is defined as

$$Y^N(x) := Y(0) + Y'(0)x + \sum_{j=0}^{N} Y''(t_{N,j})\varphi_j(x) = \zeta + \kappa x + \sum_{j=0}^{N} \xi_j \varphi_j(x), \tag{5}$$

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

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

In that case, Y^N is convex *if and only if* the random coefficient $Y''(t_{N,j})$ are all non-negative. Additionally, the covariance function of the finite-dimensional approximation of GPs 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

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

and

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

REMARK 2. *The convexity constraints can be obtained with model* (3) *which is a piecewise-linear function. In that case,* $(Y^N(x))_{x \in \mathcal{D}}$ *is convex if and only if the sequence of coefficients* $(Y(t_{N,j}))_j$ *verifies*

$$\frac{Y(t_{N,j}) - Y(t_{N,j-1})}{t_{N,j} - t_{N,j-1}} \le \frac{Y(t_{N,j+1}) - Y(t_{N,j})}{t_{N,j+1} - t_{N,j}}, \quad j = 1, \ldots, N-1.$$

This is equivalent to $Y(t_{N,j}) - Y(t_{N,j-1}) \le Y(t_{N,j+1}) - Y(t_{N,j})$*, due to the uniform subdivision of the input set* $\mathcal{D}$ *used in this paper.*

The problem dimension $d \ge 2$ is considered. For boundedness and convexity constraints, the proposed model can be easily extended to multidimensional cases. In the following, isotonicity constraints are developed.

3.2. Isotonicity in two dimensions

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

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

As in the one-dimensional case, the basis functions are constructed such that monotonicity constraints are equivalent to constraints on the coefficients. The finite-dimensional approximation of GPs $\left(Y^N(\boldsymbol{x})\right)_{\boldsymbol{x} \in \mathcal{D}^2}$ is defined as

$$Y^N(x_1, x_2) := \sum_{i,j=0}^{N} Y(t_{N,i}, t_{N,j}) h_i(x_1) h_j(x_2) = \sum_{i,j=0}^{N} \xi_{i,j} h_i(x_1) h_j(x_2), \tag{6}$$

where $\xi_{i,j} = Y(t_{N,i}, t_{N,j})$ and $(h_j)_j$ 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 $\xi_{i,j}$ verify the following linear constraints:

(a) $\xi_{i-1,j} \le \xi_{i,j}$ and $\xi_{i,j-1} \le \xi_{i,j}$, $i,j = 1, \ldots, N$;
(b) $\xi_{i-1,0} \le \xi_{i,0}$, $i = 1, \ldots, N$;
(c) $\xi_{0,j-1} \le \xi_{0,j}$, $j = 1, \ldots, N$.

REMARK 3 (ISOTONICITY WITH RESPECT TO ONE VARIABLE). *If the function is non-decreasing with respect to the first variable only, then model* (6) *is non-decreasing with respect to* x_1 *if and only if the random coefficients* $\xi_{i-1,j} \le \xi_{i,j}$, $i = 1, \ldots, N$ *and* $j = 0, \ldots, N$.

3.3. Isotonicity in multidimensional cases

The input $\boldsymbol{x} = (x_1, \ldots, x_d) \in \mathbb{R}^d$ and without loss of generality in $\mathcal{D} = [0,1]^d$. The finite-dimensional approximation of GPs Y^N can be seen as a simple extension of two-dimensions

$$Y^N(\boldsymbol{x}) := \sum_{i_1,\ldots,i_d=0}^{N} Y(t_{N,i_1}, \ldots, t_{N,i_d}) \prod_{\sigma\in\{1,\ldots,d\}} h_{i_\sigma}(x_\sigma),$$

where h_{i_σ} are the hat functions defined in Subsection 3.1.1.

3.4. Simulated paths

This subsection is devoted to the sampling scheme of the proposed model conditionally to inequality constraints and noisy observations. For the sake of simplicity, the finite-dimensional approximation of GPs is supposed as in (2). In this paper, the case where the GP is observed with error is considered. The space of observations is defined as

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

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

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

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*

$$m^N(\boldsymbol{x}) := \mathbb{E}\left(Y^N(\boldsymbol{x}) \,\middle|\, Y^N(\boldsymbol{x}^{(i)}) = \tilde{y}_i, \ i = 1, \ldots, n\right) = \xi_I^\top \phi(\boldsymbol{x}),$$

where $\xi_{\text{I}} := \mathbb{E}\left(\xi \mid \xi \in I_\xi\right) = \Gamma^N A^\top \left(A\Gamma^N A^\top + \sigma^2_{\text{noise}}\mathbf{I}\right)^{-1} \tilde{\boldsymbol{y}}$.

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

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

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

REMARK 4. *The unconstrained mean $m^N(\boldsymbol{x})$ respects inequality constraints in the entire domain if and only if the conditional Gaussian vector to only noisy observations ξ_{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*

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

where $\xi_{\mathrm{pos}} := \mathbb{E}\left(\xi \mid \xi \in I_\xi \cap C_{\mathrm{coef}}\right)$ is the mean of the truncated Gaussian vector which computed from simulations.

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

$$\mu := \arg \min_{c \in I_\xi \cap C_{\mathrm{coef}}} \left(\frac{1}{2} c^\top \left(\Gamma^N\right)^{-1} c\right), \tag{7}$$

where Γ^N is the covariance matrix of the Gaussian vector ξ. The quadratic optimization problem (7) is equivalent to

$$\mu = \arg \min_{c \in C_{\mathrm{coef}}} \left(\frac{1}{2} c^\top \left(\Gamma^N_{\mathrm{cond}}\right)^{-1} c + \xi_{\mathrm{I}}^\top c\right), \tag{8}$$

where Γ^N_{cond} is the covariance matrix of the conditional Gaussian vector $\xi \mid A\xi = \tilde{\boldsymbol{y}}$. In fact, μ representes the maximum of the pdf of the Gaussian vector ξ restricted to $I_\xi \cap C_{\mathrm{coef}}$ 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 (7)-(8).

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

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

where $\mu = (\mu_0, \ldots, \mu_N)^\top$ is computed by (8).

REMARK 5. *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 6. *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*

Algorithm 1: Sampling scheme

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

interpolation function defined as the solution of the following convex optimization problem

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

where H is a reproducing kernel Hilbert space (RKHS) associated to the positive type kernel K (Aronszajn, 1950), 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 truncated Gaussian vectors ξ restricted to the convex set C_{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. Second, to show the two different cases in the simulation.

- 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.

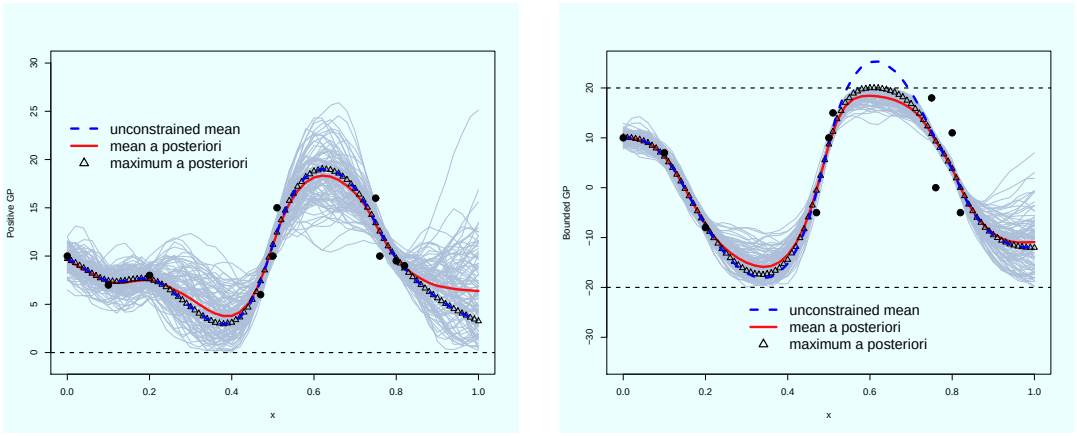
The Matérn 3/2 and squared exponential (or Gaussian) covariance functions are used (Table 1).

4.1. Boundedness constraints

The real function is supposed to respect boundedness constraints

$$C = \left\{ f \in \mathcal{C}^0([0,1]) \ : \ -\infty \leq a \leq f(x) \leq b \leq +\infty, \ x \in [0,1] \right\}. \tag{9}$$

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 (3) with $N = 50$ conditionally to positivity constraints (i.e., $a = 0$ and $b = +\infty$ in (9)). 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: it corresponds to the situation where the conditional Gaussian vector ξ_{I} lies inside the acceptance region C_{coef} (Remark 4). In Fig. 1b, the boundedness constraint is considered (i.e., $a = -20$ and $b = 20$ in (9)). 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 ξ_{I} lies outside the acceptance region C_{coef} (Remark 4).



(a) (b)

Fig. 1. The GP approximation (3) with positivity constraints (a) and boundedness constraints (b). The unconstrained mean coincides with the maximum a posteriori in (a) but not in (b)

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 $\sigma_{\text{noise}} = 1$. 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 ξ_{I} lies outside the acceptance region C_{coef} (Remark 4).

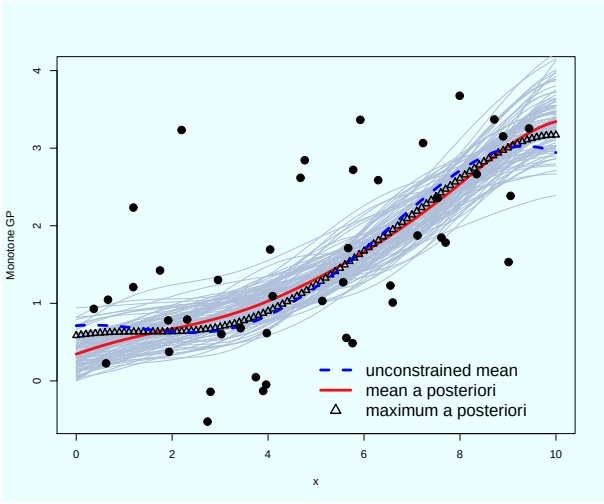


Fig. 2. The GP approximation (4) with monotonicity constraints for sinusoidal function $f(x) = 0.32(x + sin(x))$. The unconstrained mean does not coincide with the maximum a posteriori

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 $\sigma_{\mathrm{noise}} = 0.1$. In Fig. 3, the two-dimensional squared exponential covariance function is used

$$K(\boldsymbol{x}, \boldsymbol{x}') = \sigma^2 \exp\left(-\frac{(x_1 - x_1')^2}{2\theta_1^2}\right) \times \exp\left(-\frac{(x_2 - x_2')^2}{2\theta_2^2}\right), \tag{10}$$

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 (6) with $N = 10$ and the associated contour levels. It respects monotonicity (non-decreasing) constraints with respect to the two inputs.

Remark 7. *For monotonicity with respect to only one variable, model* (6) *(with noise-free data) has been used in Cousin et al. (2016) to estimate the 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.*

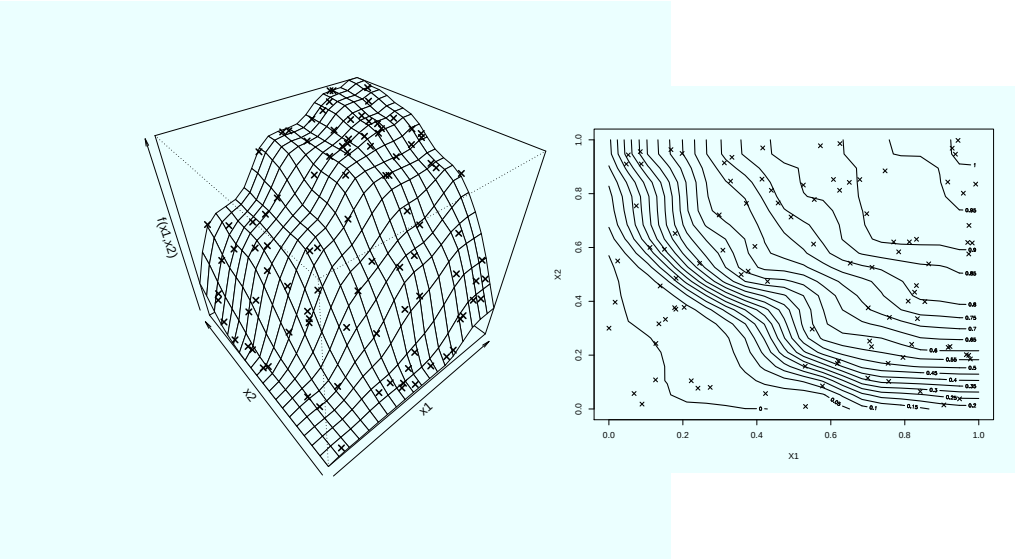


Fig. 3. The maximum of the posterior distribution drawn from model (6) respecting monotonicity (non-decreasing) constraints for the two inputs, and the associated contour levels

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 $f_6(x) = 3/\{1 + \exp(-2x + 10)\},\ x \in (0, 10]$.

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

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

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)

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

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 8. *Let us recall that the finite-dimensional approximation of GPs developed in the present paper is supposed centered (i.e., mean-zero). To be coherent, the results presented in Table 3 should be computed when the output values are normalized*

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

where $\bar{y}_i$ is the normalized value of the $i^{\boldsymbol{th}}$ output observation and $\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i$ 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

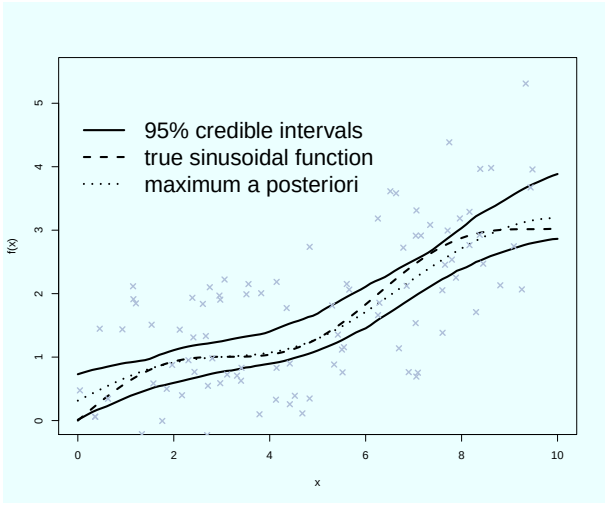


Fig. 4. The 95% credible intervals of the Gaussian process approximation together with the sinusoidal function, the observations (grey crosses) and the maximum a posteriori estimate

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

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 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. 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 used in Lin

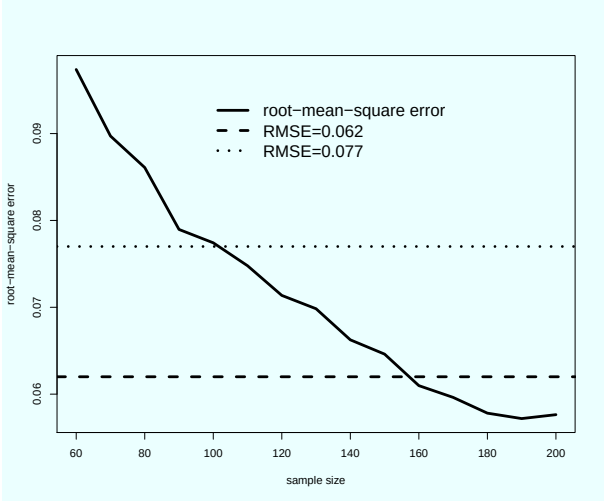


Fig. 5. The root-mean-square error at different sample sizes together with the optimal values obtained in Riihimäki and Vehtari (2010). The results are based on 1000 simulation replicates

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)

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 (10) 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

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

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. 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 predictions and uncertainty quantification, a comparison with several recently models deal 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 thanks Xavier Bay (EMSE), Olivier Roustant (EMSE), Laurence Grammont (ICJ) and Yann Richet (IRSN) for helpful discussions.

References

Abrahamsen, P. and F. E. Benth (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., L. Grammont, and H. Maatouk (2016). Generalization of the Kimeldorf-Wahba correspondence for constrained interpolation. *Electron. J. Statist. 10*(1), 1580–1595.

Bay, X., L. Grammont, and H. Maatouk (2017). A new method for interpolating in a convex subset of a Hilbert space. *Comput. Optim. Appl.*. doi:10.1007/s10589-017-9906-9.

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Maatouk, H. and X. Bay (2017). Gaussian process emulators for computer experiments with inequality constraints. *Math. Geosc.*. doi:10.1007/s11004-017-9673-2.

Maatouk, H., O. Roustant, and Y. Richet (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 D. B. Dunson (2004). Bayesian isotonic regression and trend analysis. *Biometrics 60*(2), 398–406.

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

Philippe, A. and C. P. Robert (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 C. K. Williams (2006). *Gaussian processes for machine learning*. MIT Press, Cambridge.

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

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

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

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

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