Cross-validation estimations of hyper-parameters of Gaussian processes with inequality constraints

Hassan Maatouk, Olivier Roustant, Yann Richet

Available online at www.sciencedirect.com

ScienceDirect

Procedia Environmental Sciences 00 (2015) 000–000

Procedia Environmental Sciences

www.elsevier.com/locate/procedia

Spatial Statistics 2015: Emerging Patterns

Cross-Validation Estimations of Hyper-Parameters of Gaussian Processes with Inequality Constraints

Hassan Maatouk[a,]*, Olivier Roustant[a], Yann Richet[b]

[a]*Mines de Saint-Etienne, 158 Cours Fauriel, 42023 St-Etienne, France*
[b]*Institut de Radioprotection et de Sûreté Nucléaire (IRSN, Paris), France*

Abstract

In many situations physical systems may be known to satisfy inequality constraints with respect to some or all input parameters. When building a surrogate model of this system (like in the framework of computer experiments[7]), one should integrate such expert knowledge inside the emulator structure. We proposed a new methodology to incorporate both equality conditions and inequality constraints into a Gaussian process emulator such that all conditional simulations satisfy the inequality constraints in the whole domain[6]. An estimator called mode (maximum a posteriori) is calculated and satisfies the inequality constraints.

Herein we focus on the estimation of covariance hyper-parameters and cross validation methods[1]. We prove that these methods are suited to inequality constraints. Applied to real data in two dimensions, the numerical results show that the Leave-One-Out mean square error criterion using the mode is more efficient than the usual (unconstrained) Kriging mean.

Keywords: Gaussian process emulator ; inequality constraints ; cross validation

1. Introduction

In the literature of incorporating inequality constraints into a Gaussian process (GP) emulator, some methodologies are based on the knowledge of the derivatives of the GP at some input locations[3,4,8]. The methodology presented in (Maatouk and Bay, 2014)[6] is quite different from the methods constructed so far. The

* Corresponding author. Tel.: +33(4)77-42-66-37; fax: +0-000-000-0000 .
E-mail address: hassan.maatouk@mines-stetienne.fr

inequality constraints are forced by constructions. The main idea is the approximation of the original GP by a finite-dimensional one. It is done via incorporating Gaussian random coefficients and deterministic basis functions. The basis functions[6] are chosen such that the inequality constraints of the GP are *equivalent* to constraints on the coefficients. By this special choice of the basis functions, the problem is reduced to simulate a truncated Gaussian vector (random coefficients) restricted to convex sets which is a well-known problem with existing algorithms, see e.g. the algorithm described in (Maatouk and Bay, 2014)[5].

In this paper, estimating covariance hyper-parameters is studied to inequality constraints and Cross Validation (CV) methods are used. We focus on the Leave-One-Out (LOO) mean square error criterion[2] which is closely related to traditional maximum likelihood estimation[1]. Let us mention that in the case of inequality constraints, the covariance parameters are estimated without constraints. Herein a suited cross validation technique to inequality constraints is derived. Additionally, a real application in two dimensions to investigate the performance of the proposed algorithm is included.

2. Gaussian processes with equality and inequality constraints

Let $\left(Y(\boldsymbol{x})\right)_{\boldsymbol{x}\in\mathbb{R}^d}$ be a centered Gaussian Process (GP) with continuous covariance functions:

$$K : (\mathbf{u},\mathbf{v}) \in \mathbb{R}^d \times \mathbb{R}^d \rightarrow K(\mathbf{u},\mathbf{v}) = \mathrm{cov}(Y(\mathbf{u}), Y(\mathbf{v})) \in \mathbb{R}. \tag{1}$$

In the running example in (Maatouk and Bay, 2014)[6], the Gaussian covariance function is considered:

$$K(x,x') = \sigma^2 \prod_{k=1}^{d} e^{-\frac{(x_k - x_k')^2}{2\theta}} = \sigma^2 C(x,x'), \tag{2}$$

for all $\boldsymbol{x}, \boldsymbol{x}' \in \mathbb{R}^d$, where C is the correlation function, σ^2 *and* $\boldsymbol{\theta} = (\theta_1,\dots,\theta_d)$ are parameters that will be estimated and cross validation methods well be used. Without loss of generality, the input $\boldsymbol{x}$ is in $[0, 1]^d$. Let C be the space of continuous functions verify some properties such as boundary, monotonicity or convexity constraints. The interpolation condition and the inequality constraints of Y are given respectively as follow:

$$Y(x^{(i)}) = y_i, \quad i = 1,\dots,n$$

$$Y \in C$$

where the design points $\boldsymbol{x}^{(i)} \in [0,1]^d$ are given by the row of the matrix $\mathbf{X} = (\boldsymbol{x}^{(1)}, \dots, \boldsymbol{x}^{(n)})^\top$.

2.1. Finite-dimensional Gaussian processes

In this section we recall the model presented in (Maatouk and Bay, 2014)[6]. The main idea is the approximation of the original Gaussian process Y by a finite-dimensional one of the form

$$Y^N(x) := \sum_{j=0}^{N} \xi_j \phi_j(x), \quad x \in \mathbb{R}$$

where $\xi = (\xi_0,\dots,\xi_N)^\top$ is a centered Gaussian vector with carefully chosen covariance matrix Γ^N and deterministic basis function $(\phi_j)_j$, $j=0, \dots, N$. The choice of these basis function and Γ^N depend on the type of the inequality constraints[6]. The basis functions are chosen such that the inequality constraints of Y^N are *equivalent* to constraints on the coefficients ξ_j. Hence the problem is equivalent to simulate the truncated Gaussian vector ξ restricted to

$$Y^N(x^{(i)}) = \sum_{j=0}^{N} \xi_j \phi_j(x^{(i)}) = y_i, \qquad i = 1,\ldots,n$$
$$\xi \in C_\xi$$

where $C_\xi = \left\{c \in R^d : \sum_{j=0}^{N} c_j \phi_j \in C\right\}$, $c=(c_0, \ldots, c_N)^\top$ see (Maatouk and Bay 2014)[6] for details.

3. Cross validation estimation of covariance hyper parameters

In the framework of estimating covariance hyper parameters, we find two types of methods. The first one is the Maximum Likelihood (ML) estimator and the second one is the CV on which we focus on this paper.

3.1. Cross validation without inequality constraints

Let us recall that the Leave-One-Out (LOO) mean square error criterion is defined as the following estimator of the correlation length hyper parameters $\boldsymbol{\theta}$:

$$\hat{\theta}_{CV} = \arg\min_{\theta\in\Theta} \sum_{i=1}^{n} (y_i - \hat{y}_{i,\theta}(y_{-i}))^2 \tag{3}$$

where $\hat{y}_{i,\theta}(y_{-i}) = E_\theta(y_i \mid y_1, \ldots, y_{i-1}, y_{i+1}, \ldots, y_n)$ and $\boldsymbol{\Theta}$ is a compact set of $\mathbb{R}^d$, see (Bachoc, 2013)[1]. Additionally, the variance parameter is estimated using the following criterion:

$$C_{LOO} = \frac{1}{n} \sum_{i=1}^{n} \frac{(y_i - \hat{y}_{i,\hat{\theta}_{CV}}(y_{-i}))^2}{\sigma^2 c_{i,-i}^2}, \tag{4}$$

where $c_{i,-i}^2 = \mathrm{var}(y_i \mid y_1,\ldots, y_{i-1}, y_{i+1},\ldots, y_n)$. From (Cressie, 1993)[2], the variance parameter is computed such that criterion (4) is closed to 1 and then:

$$\hat{\sigma}_{CV}^2 = \frac{1}{n} \sum_{i=1}^{n} \frac{(y_i - \hat{y}_{i,\hat{\theta}_{CV}}(y_{-i}))^2}{c_{i,-i}^2},$$

where $\hat{\theta}_{CV}$ is calculated from Equation (3).

3.2. Cross validation with inequality constraints

Let us mention that the two methods cited above (ML and CV) are not suited to inequality constraints. This is because the usual unconstrained kriging mean defined as the mean of the Gaussian process conditionally to given observation data is not guaranteed to satisfy inequality constraints in the whole domain. To this end, the idea is to use a new estimator called *mode* (maximum a posteriori) which is defined in (Maatouk and Bay, 2014)[6] and its analytical expression is equal to:

$$M_{KI}(x) = \sum_{j=0}^{N} \mu_j \phi_j(x), \tag{5}$$

where $\boldsymbol{\mu} = (\mu_0, ..., \mu_N)^\top$ is the solution of the following quadratic optimization problem

$$\mu = \arg\min_{c \in I \cap C_\xi} \left(\frac{1}{2} c^t (\Gamma^N)^{-1} c\right),$$

with Γ^N the covariance matrix of the Gaussian vector ξ. The vector $\boldsymbol{\mu}$ can be seen as the *mode* of the Gaussian vector ξ restricted to $I \cap C_\xi$, where

$$I := \left\{ \sum_{j=0}^{N} \xi_j \phi_j(x^{(i)}) = y_i, i = 1, ..., n \right\}.$$

The positive point of such estimator is that verifies equality condition and inequality constraints in the whole domain. Moreover, it does not depend on the variance parameters σ^2 since μ and the basis functions do not depend on it as well. Now, the idea in this paper is to replace the usual unconstrained kriging mean by the mode in Equation (3). In that case, the LOO mean square error criterion can be reformulated as follows:

$$\hat{\theta}_{CV} = \arg\min_{\theta \in \Theta} \sum_{i=1}^{n} (y_i - \hat{M}_{i,\theta}(y_{-i}))^2 \tag{6}$$

where $\hat{M}_{i,\theta}(y_{-i}) = M_{KI}(y_i \mid y_1, ..., y_{i-1}, y_{i+1}, ..., y_n; \theta)$ and M_{KI} is defined in (5). Moreover, the variance parameter is estimated using the following criterion:

$$C_{LOO} = \frac{1}{n} \sum_{i=1}^{n} \frac{(y_i - \hat{M}_{i,\hat{\theta}_{CV}}(y_{-i}))^2}{\sigma^2 c_{i,-i}^2}, \tag{7}$$

where $c_{i,-i}^2$ is obtained by incorporating inequality constraints into the Kriging variance which is used in Equation (4) and it is calculated from simulations. Now, the variance parameter is computed such that criterion (7) is closed to 1 and then:

$$\hat{\sigma}_{CV}^2 = \frac{1}{n} \sum_{i=1}^{n} \frac{(y_i - \hat{M}_{i,\hat{\theta}_{CV}}(y_{-i}))^2}{c_{i,-i}^2},$$

where $\hat{\theta}_{CV}$ is calculated from Equation (6).

3.3. Real Application

The aim of this section is to show the performance of the proposed estimator and to compare it with the usual kriging mean. We consider the real data given in Fig. 1 (a). These observation data (n=121) defined on $[0,20] \times [10,20]$ respect monotonicity (non-decreasing) constraints for the two input variables. The idea is to fix some design points and test the estimator at the other observation data.

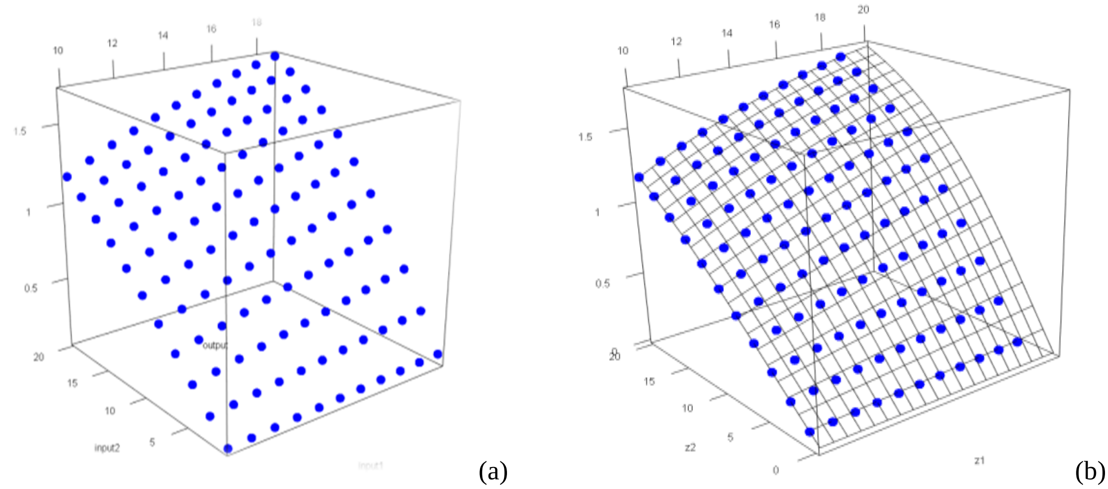


Fig. 1. (a) the observation data; (b) our estimator mode under both equality conditions and inequality constraints.

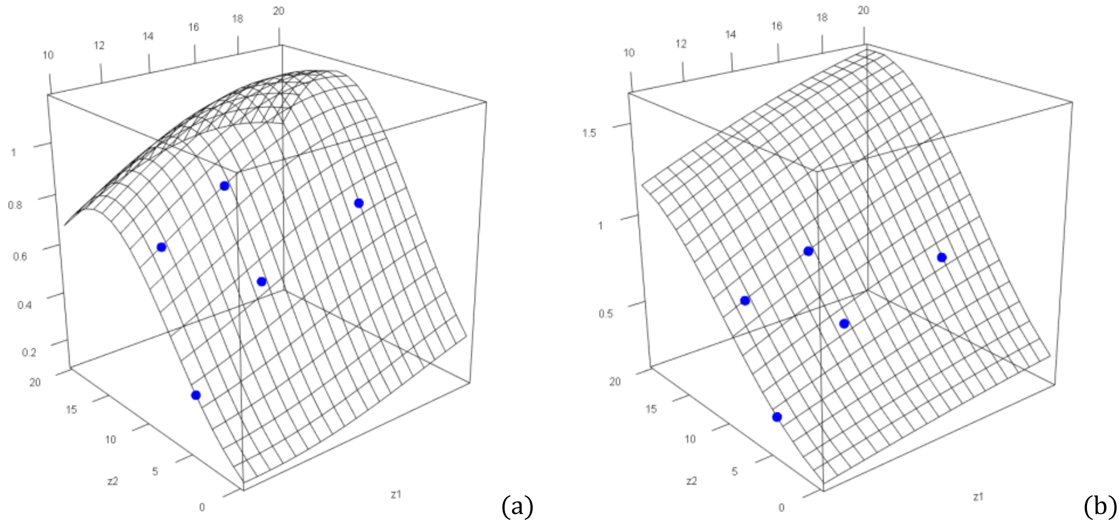
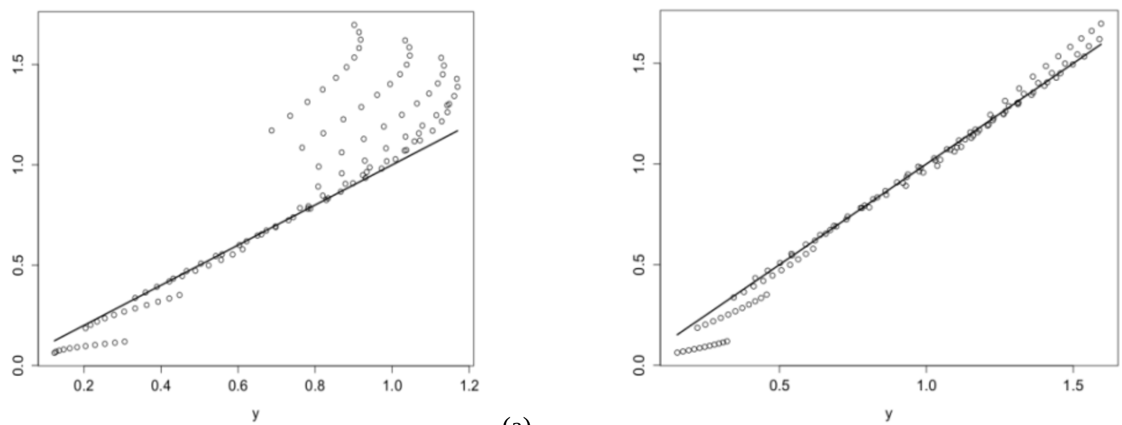


Fig. 2. (a) the usual kriging mean using the length parameter $\hat{\theta}_{\mathrm{CV}} = (9.05, 9.10)$ estimated by LOO criterion; (b) the function mode using the length parameters $\hat{\theta}_{\mathrm{CV}} = (25.17, 10.57)$ estimated by suited LOO criterion to inequality constraints.

In Fig. 2 (a), we plot the usual kriging mean with the length parameters $\hat{\theta}_{\mathrm{CV}} = (9.05, 9.10)$ estimated by LOO criterion[9]. The monotonicity (non-decreasing) constraint is not respected in the whole domain, contrarily to Fig. 2 (b), where the estimator mode is illustrated using the length parameters $\hat{\theta}_{\mathrm{CV}} = (25.17, 10.57)$ estimated by suited LOO criterion to inequality constraints described in this paper. In Fig. 3, we compare the values estimated versus the real one. The estimators used are the usual kriging mean Fig. 3 (a) and mode Fig. 3 (b). To investigate the performance of the proposed estimator mode, we calculate the criterion Q_2 defined as:

$$Q_2 = 1 - \frac{\sum_{i=1}^{n} (y_i - \hat{y}_i)^2}{\sum_{i=1}^{n} (y_i - \bar{y})^2},$$

for the two methods. It is equal to 0.98 for the method using the mode as an estimator $(\hat{y}_i)$ and equal to 0.69 for one using the usual kriging mean.



(a) (b)

Fig. 3. the values estimated versus the real values using the usual kriging mean as an estimator (a) and mode (b).

4. Conclusion

In this paper, we have proposed a new technique to estimate covariance hyper-parameters of Gaussian processes with inequality constraints. A suited cross validation algorithm to inequality constraints is derived. The performance of the proposed method is investigated by a real application in two dimensions.

Acknowledgements

This work has been conducted within the frame of the ReDice Consortium, gathering industrial (CEA, EDF, IFPEN, IRSN, Renault) and academic (Mines de Saint-Etienne, INRIA, and the University of Bern) partners around advanced methods for Computer Experiments. The authors also thank Xavier Bay (EMSE) and Laurence Grammont (ICJ, Lyon1) for constructive comments.

References

1. Bachoc F. Cross validation and maximum likelihood estimations of hyper-parameters of Gaussian processes with model misspecification. *Computational Statistics & Data Analysis* 2013*;* **60(**0**)**:55-69.
2. Cressie N. A. C.. Statistics for Spatial Data, revised edition. *John Wiley & Sons, New York* 1993.
3. Da Veiga S., Marrel A. Gaussian process modeling with inequality constraints. *Annales de la faculté des sciences de Toulouse* 2012; **21**(3):529-555, 4.
4. Golchi S., Bingham D. R., Chipman H., Campbell D. A.. Monotone Emulation of Computer Experiments. *ArXiv e-prints*, September 2013.
5. Maatouk H., Bay X. A New Rejection Sampling Method for Truncated Multivariate Gaussian Random Variables Restricted to Convex Sets. *hal-01063978*, September 2014.
6. Maatouk H., Bay X. Gaussian Process Emulators for Computer Experiments with Inequality Constraints. *hal-01096751*, December 2014.
7. Rasmussen C. E., Williams C. K. I. Gaussian Processes for Machine Learning (Adaptive Computation and Machine Learning). *The MIT Press* 2005.
8. Riihimaki J., Vehtari A. Gaussian processes with monotonicity information. *In AISTATS, volume 9 of JMLR Proceedings*. p. 645-652. JMLR.org, 2010.
9. Roustant O., Ginsbourger D., Deville Y. Dicekriging, diceoptim: Two r packages for the analysis of computer experiments by kriging-based metamodeling and optimization. *Journal of Statistical Software* 10 2012; **51**(1):1-55.