Gaussian process regression models under linear inequality conditions

A. F. López-Lopera
Mines Saint-Étienne (EMSE), France.

Supervisor(s): O. Roustant (EMSE, France), F. Bachoc (Institut de Mathématiques de Toulouse, France) and N. Durrande (EMSE, France; PROWLER.io, Cambridge, UK)

Ph.D. expected duration: Oct. 2016 - Sep. 2019

Adress: Mines Saint-Étienne, UMR CNRS 6158, LIMOS, F-42023 Saint-Étienne, France.

Email: andres-felipe.lopez@emse.fr

Abstract:

In the last decades, Gaussian processes (GPs) have become one of the most attractive Bayesian framework due to their ability to perform both regression and classification tasks [1]. However, due to their pure data-driven nature, they do not account for the physical properties exhibited in real-world data (e.g. positivity, monotonicity), which can lead to more realistic data interpolation and uncertainty quantifications [2, 3]. Figure 1 shows an example where the target function satisfies both boundedness (i.e. $0 \leq y(x_i) \leq 1$, for all $x_i \in [0, 1]$) and monotonicity constraints (i.e. $y(x_{i-1}) \leq y(x_i)$, if $x_{i-1} \leq x_i$). Our aim is to investigate deeper GP regression models under inequality constraints.

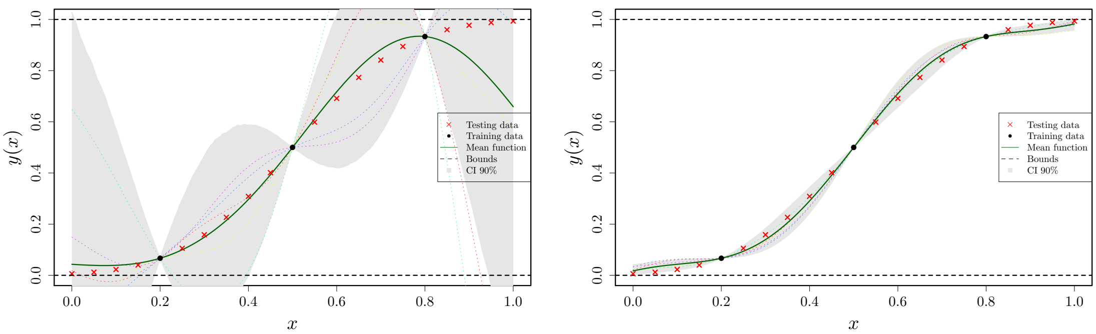

Figure 1: GP models interpolating the Gaussian CDF $x \mapsto \Phi(\frac{x-0.5}{0.2})$: (left) classical (unconstrained) GP model, (right) constrained GP model satisfying both boundedness and monotonicity constraints.

To the best of our knowledge, the only probabilistic model which satisfies specific inequalities everywhere in the input space is proposed in [2]. With this approach, the posterior converges to the one provided by splines methods [4]. Our work builds on this framework and our contributions are threefold. First, we extend their framework to deal with any linear inequality constraint (not only boundedness, monotonicity or convexity). Second, we suggest an efficient Hamiltonian Monte Carlo-based sampler to approximate the posterior distribution satisfying both interpolation and inequality constraints. Finally, we investigate theoretical and numerical properties of a constrained likelihood for covariance parameter estimation.

The model was tested under both synthetic and real-world data in 1D or 2D. According to the experimental results under different types of inequality constraints, the proposed method fits properly the observations and provides realistic confidence intervals (see Figure 1). On a

2D nuclear criticality safety dataset, it provides reliable results on both data prediction and uncertainty quantification satisfying both positivity and monotonicity conditions exhibited by the nuclear data. Table 1 shows the performances of the Gaussian models for interpolating the nuclear database using different number of training points n and using twenty different random Latin hypercube designs. The models are assessed using the Q^2 criterion.[1] One can observe that constrained GP models outperformed the unconstrained one in all the cases.

Table 1: Assessment of the Gaussian models for interpolating the nuclear database using different number of training points n and using twenty different random Latin hypercube designs. Predictive accuracy is evaluated using the mean μ and the standard deviation σ of the Q^2 results. MLE: maximum likelihood estimation. CMLE: constrained maximum likelihood estimation.

	Q^2		
n	unconstrained GP + MLE $\mu \pm \sigma$	constrained GP + MLE $\mu \pm \sigma$	constrained GP + CMLE $\mu \pm \sigma$
2	-0.128 ± 1.004	$\mathbf{0.967 \pm 0.026}$	0.952 ± 0.043
4	0.558 ± 0.260	0.981 ± 0.014	$\mathbf{0.996 \pm 0.006}$
6	0.858 ± 0.139	0.940 ± 0.059	$\mathbf{0.995 \pm 0.004}$
8	0.962 ± 0.035	$\mathbf{0.995 \pm 0.003}$	0.981 ± 0.011

We would like to present the results obtained during the 1st year of the PhD training at MascotNum Annual Conference 2018 for the poster session. We believe that an oral presentation would be more appropriate at a later edition of the conference, when the results of currently ongoing work could be presented. This ongoing work consists in the extension of our framework for higher dimension, and the investigation of the asymptotics for the constrained likelihood. We also believe that our presentation at the conference could be helpful to discuss possible research directions in the future.

References

[1] C. E. Rasmussen and C. K. I. Williams, *Gaussian Processes for Machine Learning (Adaptive Computation and Machine Learning).* The MIT Press, 2005.

[2] H. Maatouk and X. Bay, "Gaussian process emulators for computer experiments with inequality constraints," *Mathematical Geosciences*, vol. 49, no. 5, pp. 557–582, 2017.

[3] S. Da Veiga and A. Marrel, "Gaussian process modeling with inequality constraints," *Annales de la faculté des sciences de Toulouse Mathématiques*, vol. 21, no. 3, pp. 529–555, 2012.

[4] X. Bay, L. Grammont, and H. Maatouk, "Generalization of the Kimeldorf-Wahba correspondence for constrained interpolation," *Electronic journal of statistics* , vol. 10, no. 1, pp. 1580–1595, 2016.

Short biography – A. F. López-Lopera received the BEng and MEng degrees in electrical engineering from the Universidad Tecnológica de Pereira, Colombia. Currently, he is working toward the PhD degree in applied mathematics at EMSE, France. In the thesis, *Metamodelling under inequality constraints*, GP-based models are explored in two main directions: the estimation under inequality constraints, and the extension to higher dimensions. The thesis is being funded by the chair in applied mathematics OQUAIDO.

[1] Denoting by n_t the number of test points, $z_1, \cdots, z_{n_t}$ and $\widehat{z}_1, \cdots, \widehat{z}_{n_t}$ the sets of test and predicted observations (respectively), then $Q^2 = 1 - \sum_{i=1}^{n_t} (\widehat{z}_i - z_i)^2 / \sum_{i=1}^{n_t} (\overline{z} - z_i)^2$, where $\overline{z}$ is the mean of the test data. Hence, for noise-free observations, the Q^2 indicator is equal to one if the predictors $\widehat{z}_1, \cdots, \widehat{z}_{n_t}$ are exactly equal to the test data (ideal case), zero if they are equal to the constant prediction $\overline{z}$, and negative if they perform worse than $\overline{z}$.