Master internship and PhD thesis offer
Additive constrained Gaussian processes in high dimension

General context. This master internship and PhD thesis are part of the ANR JCJC "GAP" project, a 5-year research project funded by the French national research agency. GAP focuses on both theoretical and practical aspects related to Gaussian processes (GPs), a class of stochastic processes widely used in machine learning and computer experiments. The guiding principle of GAP is to push the boundaries of the current mathematical models and algorithmic procedures based on GPs, with a strong emphasis on reliability. The research proposal seeks to exploit additive structures in constrained GPs for which constraints are verified everywhere. This reliability property is key for real-world applications concerned by risk assessment, especially in criticality safety.

The framework of constrained GPs. Adding physical information in models can significantly reduce the size of the training set which is necessary to reach a certain level of accuracy. We consider here the information given by inequality constraints, such as monotonicity. This research considers a class of GP models proposed by [3], for which the inequality constraints are verified everywhere in the space. This GP model has been investigated under different angles: simulation [6], inference properties [1], noisy observations [5] and Bayesian inference [8]. However, in its original form, the model is constructed by tensorization of 1-dimensional models, and thus structurally limited to low dimensional problems (typically less than 5 dimensions). To scale to higher dimensions, a promising algorithm called MaxMod has been introduced, which sequentially finds the more active input variables [2]. Recently, MaxMod has been extended to hundreds of dimensions under additivity assumptions of the GP model [7].

Research directions. Constrained additive GPs are of great interest to address high dimensional problems with inequality information. However, the additivity assumption used in [7] excludes the presence of interactions of input variables, which may be very restrictive in real-world applications. To conciliate these two objectives, we propose to consider block-additive GPs of the form,

$$\sum_{I \in \mathcal{I}} Z_I(x_I). \tag{1}$$

which are additive with respect to *groups* of input variables. In (1), $\mathcal{I}$ contains groups of input variables. While considering interactions in additive model is not new (see, e.g., [4] in machine learning, or [9] in computer experiments), there still remain difficult questions to solve, especially when inequality constraints are considered. To cite a few: how to select the groups, namely the subsets $I \in \mathcal{I}$? How to adapt the MaxMod algorithm? How to deal with non-overlapping groups? How to obtain theoretical guarantees?

Expected research outputs. The output is methodological and will lead to publications in international journals in applied mathematics and/or conferences in machine learning. Furthermore, the methodology may also be applied on real-world applications, such as the environmental risk studies considered in [2, 7]. We expect to significantly improve the efficiency of the current models and algorithms by using both the wider class of block-additive GPs and the physical information brought by inequality constraints.

Supervising team and organizational details. The master internship is funded for about 6 months and will start in Spring 2023. The PhD is funded for 3 years and will start in Fall 2023. For a student already holding a master degree, the PhD training can start as soon as possible. The student will be supervised by a leading team on constrained Gaussian processes:

- François Bachoc, Ass. Prof. at *Institut de Mathématiques de Toulouse* (IMT),
- Andrés F. Lopéz-Lopera, Ass. Prof. at *Université Polytechnique Hauts-de-France*,
- Olivier Roustant, Prof. at *INSA Toulouse & IMT*.

The student will be located at IMT, France. An international mobility during the PhD is envisaged, as a collaboration with other researchers of the ANR JCJC GAP project.

How to apply? Applications will be considered until the position is filled. The candidates should have master-level skills in mathematics / statistics / machine learning. Please send a CV (either in English or French), application letter and grade transcripts (bachelor and master level) to

- François Bachoc (francois.bachoc@math.univ-toulouse.fr),
- Andrés F. Lopéz-Lopera (andres.lopezlopera@uphf.fr), and
- Olivier Roustant (roustant@insa-toulouse.fr).

References related to the PhD subject.

[1] F. Bachoc, A. Lagnoux, and A. F. López-Lopera. Maximum likelihood estimation for Gaussian processes under inequality constraints. *Electronic Journal of Statistics*, 13(2):2921–2969, 2019.

[2] F. Bachoc, A. F. López-Lopera, and O. Roustant. Sequential construction and dimension reduction of Gaussian processes under inequality constraints. *SIAM Journal on Mathematics of Data Science*, 4(2):772–800, 2022.

[3] X. Bay, L. Grammont, and H. Maatouk. Generalization of the Kimeldorf-Wahba correspondence for constrained interpolation. *Electronic journal of statistics*, 10(1):1580–1595, 2016.

[4] D. Duvenaud, H. Nickisch, and C. Rasmussen. Additive Gaussian processes. In *NeurIPS*, 2011.

[5] A. F. López-Lopera, F. Bachoc, N. Durrande, J. Rohmer, D. Idier, and O. Roustant. Approximating Gaussian process emulators with linear inequality constraints and noisy observations via MC and MCMC. In *Monte Carlo and Quasi-Monte Carlo Methods*, 2020.

[6] A. F. López-Lopera, F. Bachoc, N. Durrande, and O. Roustant. Finite-dimensional Gaussian approximation with linear inequality constraints. *SIAM/ASA Journal on Uncertainty Quantification*, 6(3):1224–1255, 2018.

[7] A. F. López-Lopera, F. Bachoc, and O. Roustant. High-dimensional additive Gaussian processes under monotonicity constraints. *Accepted for publication in NeuRIPS 2022* `https://arxiv.org/abs/2205.08528`, 2022.

[8] A. F. Lopez-lopera, S. John, and N. Durrande. Gaussian process modulated Cox processes under linear inequality constraints. In *AISTATS*, 2019.

[9] T. Muehlenstaedt, O. Roustant, L. Carraro, and S. Kuhnt. Data-driven Kriging models based on FANOVA-decomposition. *Statistics & Computing*, 22(3):723–738, 2012.