Proceeding Series of the Brazilian Society of Computational and Applied Mathematics, v. 11, n. 1, 2025.

Trabalho apresentado no XLIII CNMAC, Centro de Convenções do Armação Resort - Porto de Galinhas - PE, 2024

Proceeding Series of the Brazilian Society of Computational and Applied Mathematics

Bayesian Shape-constrained Curve-fitting with Gaussian Processes: Prior Elicitation and Computation

Eduardo Adame[1], Luiz Max Carvalho[2]
School of Applied Mathematics, Getulio Vargas Foundation, Rio de Janeiro, RJ

In many applications, one is interested in reconstructing a function f when only few (potentially very noisy) evaluations $f(x)$ are available, usually due to budget restrictions. When information about the "shape" of f is available, e.g., whether it is monotonic, convex/concave, etc., it is desirable to include this information into the curve-fitting procedure. Here we build on the Gaussian process literature to propose a comprehensive framework for flexibly modeling f and its first two derivatives given evaluations of f, f', f'' at potentially irregular grids. We show how to include shape-constraints in a principled way through the prior and apply the developed methods to function emulation for noisy Markov chain Monte Carlo.

A typical strategy for evaluating a computationally intensive function f across a large collection of points S is to construct a more cost-effective estimator $\hat{f}(q) \approxeq f(q)\ \forall\ q \in Q \subseteq S$, and then use $\hat{f}(q)$ for approximation over $q \in (\mathcal{D}(f) \setminus Q)$, where $\mathcal{D}(f)$ is the domain of f, covering the remaining points in the domain. Employing Gaussian processes (see Definition 1) facilitates direct uncertainty quantification and enables the computation of probabilities such as $\mathbb{P}[\hat{f}(a) \in A]$ for each $a \in \mathcal{D}(\hat{f})$ and every $A \subseteq \operatorname{Im}(\hat{f})$, where $\operatorname{Im}(\hat{f})$ denotes the image of $\hat{f}$. Furthermore, we can incorporate shape constraints on f like monotonicity and convexity by manipulating the prior measure over its derivatives. This is done by exploiting Property 1, an idea initially explored in [2–4]. For our purposes, we can define a Gaussian process as follows:

Definition 1 (Gaussian Process)**.** *A Gaussian Process is a collection of random variables, such that any finite number of which have a joint Gaussian distribution. For a function $f : \mathcal{X} \to \mathbb{R}$, modelled as $f(\boldsymbol{X}) \mid \boldsymbol{X} \sim \mathcal{GP}(m(\boldsymbol{X}), k(\boldsymbol{X}, \boldsymbol{X}'))$, where $m(\boldsymbol{X})$ and $k(\boldsymbol{X}, \boldsymbol{X}')$ represent the mean and kernel (or covariance) functions applied to each entry $X \in \boldsymbol{X}$, respectively. The kernel function should be symmetric and hold $k(X, X) > 0$ for every (same) X.*

Considering the model $\boldsymbol{Y} \mid (\boldsymbol{f}, \boldsymbol{X}) \sim \mathcal{N}(\boldsymbol{f}, \sigma^2 I)$ with $\boldsymbol{f} \mid \boldsymbol{X} \sim \mathcal{GP}(m(\boldsymbol{X}), k(\boldsymbol{X}, \boldsymbol{X}'))$, where $\sigma^2 > 0$ is a known constant, $\boldsymbol{Y}, \boldsymbol{X} \in \mathbb{R}^n$ are random vectors, and I denotes the identity matrix in $\mathbb{R}^{n \times n}$, we can take a Bayesian approach and compute a posterior distribution $\boldsymbol{f} \mid \boldsymbol{X}, \boldsymbol{Y}$ for a sample size of n. In this case, we are able to sample from any unobserved $\boldsymbol{Y}_\star$ for a new observed $\boldsymbol{x}_\star$ by the predictive posterior $p(\boldsymbol{Y}_\star \mid \boldsymbol{y}, \boldsymbol{x}_\star, \boldsymbol{x})$, which is also Normally distributed [2]. The main result that will make us able to present the shape-constrained Gaussian processes (SCGPs) approach was also introduced by [2], which is:

Property 1 (Derivative of a GP is a GP)**.** *Let $(f(\boldsymbol{X}) \mid \boldsymbol{X}) = (\boldsymbol{f} \mid \boldsymbol{X}) \sim \mathcal{GP}(m(\boldsymbol{X}), k(\boldsymbol{X}, \boldsymbol{X}'))$, then for every $\frac{\partial f}{\partial X_d}$ we will have that:*

$$Cov\left(f_i, \frac{\partial f_j}{\partial X_{dj}}\right) = \frac{\partial}{\partial X_{dj}} k(X_i, X_j) \quad \textit{and} \quad Cov\left(\frac{\partial f_i}{\partial X_{di}}, \frac{\partial f_j}{\partial X_{ej}}\right) = \frac{\partial^2}{\partial X_{di} \partial X_{ej}} k(X_i, X_j). \tag{1}$$

[1]eduardo.salles@fgv.br
[2]luiz.fagundes@fgv.br

the equivalent follows for the expected values. Therefore:

$$\begin{bmatrix} \boldsymbol{f} \\ \boldsymbol{f}' \end{bmatrix} \mid \boldsymbol{X} \sim \mathcal{GP}\left(\begin{bmatrix} m(\boldsymbol{X}) \\ \frac{\partial}{\partial \boldsymbol{X}} m(\boldsymbol{X}) \end{bmatrix}, \begin{bmatrix} k(\boldsymbol{X}, \boldsymbol{X}') & \frac{\partial}{\partial \boldsymbol{X}'} k(\boldsymbol{X}, \boldsymbol{X}') \\ \frac{\partial}{\partial \boldsymbol{X}} k(\boldsymbol{X}, \boldsymbol{X}') & \frac{\partial^2}{\partial \boldsymbol{X} \partial \boldsymbol{X}'} k(\boldsymbol{X}, \boldsymbol{X}') \end{bmatrix} \right). \tag{2}$$

Considering the predictive posterior and equation (2), we are able to perform a shape-constrained Gaussian process regression using observations of $((x_1, y), (x_2, y'))$ which do not need to be on the same input values.

Here we showcase a function emulation application involving the approximation of a marginal likelihood. Consider a tempered posterior of the form $p_\alpha(\theta \mid \boldsymbol{z}) \propto l(\boldsymbol{z} \mid \theta)^\alpha \pi(\theta)$ for $\alpha \in [0, 1]$. Our goal is to emulate $f_{\boldsymbol{z}}(\alpha) = \int_\Omega l(\boldsymbol{z} \mid t)^\alpha \pi(t)\, \mathrm{d}\mu(t)$ with 20 evaluations, which are quite costly. In our experiments, we were able to achieve some intuition about the information that the derivative processes bring to the regression, although it might propagate uncertainty, as can be seen in Figure 1.

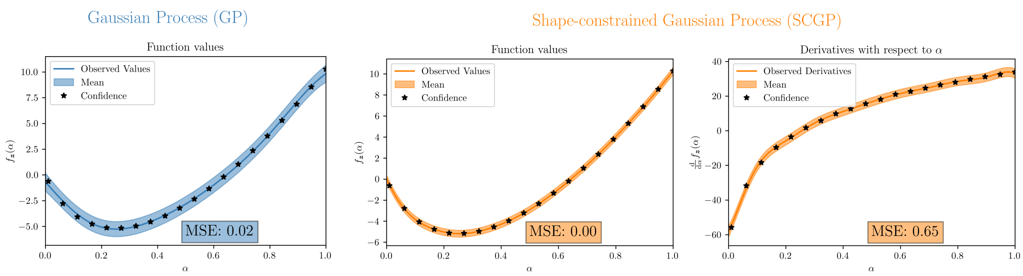

Figure 1: On the left: Gaussian process regression. Right: shape-constrained Gaussian process with first derivative observation. Both using the Radial Basis Function (RBF) as the kernel function. Source: the authors.

Acknowledgments

The authors wish to express their gratitude to CNPq and INCTMat for the financial support (process 126397/2023-6). The code for this study is based on GPyTorch [1] and is available on https://github.com/adamesalles/shape-constrained-gaussian-processes.

References

[1] J. R. Gardner, G. Pleiss, D. Bindel, K. Q. Weinberger, and A. G. Wilson. **GPyTorch: Blackbox Matrix-Matrix Gaussian Process Inference with GPU Acceleration**. 2021. arXiv: `1809.11165 [cs.LG]`.

[2] C. E. Rasmussen and C. K. I. Williams. **Gaussian Processes for Machine Learning**. Adaptive Computation and Machine Learning series. MIT Press, 2005. ISBN: 9780262182539.

[3] J. Riihimäki and A. Vehtari. "Gaussian processes with monotonicity information". In: **Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics**. Ed. by Yee Whye Teh and Mike Titterington. Vol. 9. Proceedings of Machine Learning Research. Chia Laguna Resort, Sardinia, Italy: PMLR, May 2010, pp. 645–652.

[4] X. Wang and J. O. Berger. "Estimating Shape Constrained Functions Using Gaussian Processes". In: **SIAM/ASA Journal on Uncertainty Quantification** 4.1 (2016), pp. 1–25. DOI: `10.1137/140955033`.