Fast, large scale Gaussian Process-based Bayesian inversion for set estimation in Geophysics

C. Travelletti
Idiap Research Institute and University of Bern

Supervisor(s): Prof. D. Ginsbourger (Idiap Research Institute and University of Bern)

Ph.D. expected duration: Nov. 2018 - Nov. 2022

Adress: Rue Marconi 19, 1920 Martigny, Switzerland

Email: cedric.travelletti@idiap.ch

Abstract:

In natural sciences and engineering, one is often faced with the problem of reconstructing some unknown function u_0 from indirect data that has been generated by a known physical process. Here indirect means that we do not have access to the actual value of the function at some selected points, but only to (for example) integrals, or other linear forms of the function. Such problems are broadly known as *inverse problems*.

Inverse problems can be solved in a Bayesian way by putting a prior on the unknown u_0 (usually gaussian process priors are used) and then using the conditional distribution to approximate the unknown function. There is a rich litterature dedicated to such approaches [3].

Nevertheless, if one is not interested in a full reconstruction of the unknown function itself, but only of implicit regions defined through it, the picture gets more complex. One might for example be interested in regions where the unknown u_0 is above some threshold (excursion sets), or regions where it varies sharply.

In this regard, this work aims at answering the following questions

- How can we estimate implicit sets that are defined through a function that is a solution to an inverse problem.
- How can we quantify the uncertainty on those estimates (set UQ).

These two questions serve a longer term goal which is to develop algorithms to guide the data collection process in order to optimally improve the reconstruction of the implicit region of interest. Towards this end, we want to adapt the GP-based adaptive set estimation techniques first pioneered in [1] to the inverse problem setting.

It turns out that transposing GP methods to the inverse problem world is not straightforward. Indeed, when using GP priors to solve inverse problems, the size of the involved matrices grows quadratically with the number of cells used to discretize the inversion region and one quickly reaches a memory bottleneck even for problems of moderate sizes.

Hence, as a first step towards our longer term goals, we present Gaussian Process-based Bayesian inversion techniques for linear inverse problems which are suited for fast inversion on large scale grids (several hundreds of thousands cells). Those techniques bypass the memory bottleneck by distributing the computations to multiple graphical processing units (GPUs) and by only accessing *bigger-than-memory* matrices through matrix-matrix products, in a fashion similar to [5]. The result of this work is a full-fledged library for solving bayesian inverse problems on GPUs that is generic and can be used on any linear inverse problem.

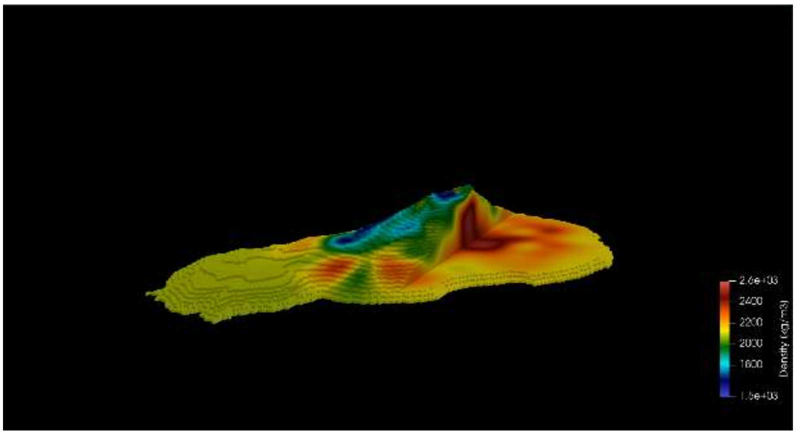

Figure 1: GP-based inversion of Stromboli data (posterior mean).

We will demonstrate those techniques on a test case from volcano geophysics that aims at reconstructing the mass density field inside the Stromboli island from gravimetric data [2].

Our whole Bayesian inversion machinery is distributed as an open source Python package [4] which can be used on any linear inverse problem, the user only having to provide the forward operator and the geometry of the inversion grid.

This presentation is based on joint work with David Ginsbourger and Niklas Linde.

References

[1] Dario Azzimonti, David Ginsbourger, Clment Chevalier, Julien Bect, and Yann Richet. Adaptive design of experiments for conservative estimation of excursion sets. *Technometrics*, 0(0):1–14, 2019.

[2] Niklas Linde, Ludovic Baron, Tullio Ricci, Anthony Finizola, André Revil, Filippo Muccini, Luca Cocchi, and Cosmo Carmisciano. 3-d density structure and geological evolution of stromboli volcano (aeolian islands, italy) inferred from land-based and sea-surface gravity data. *Journal of volcanology and geothermal research*, 273:58–69, 2014.

[3] Albert. Tarantola. *Inverse Problem Theory and Methods for Model Parameter Estimation*. Society for Industrial and Applied Mathematics, 2005.

[4] Cédric Travelletti. Volcapy. `https://github.com/CedricTravelletti/Volcano`, 2020.

[5] Ke Wang, Geoff Pleiss, Jacob Gardner, Stephen Tyree, Kilian Q Weinberger, and Andrew Gordon Wilson. Exact gaussian processes on a million data points. In H. Wallach, H. Larochelle, A. Beygelzimer, F. d' Alché-Buc, E. Fox, and R. Garnett, editors, *Advances in Neural Information Processing Systems 32*, pages 14622–14632. Curran Associates, Inc., 2019.

Short biography – Cédric Travelletti obtained his MSc in Physics from ETH Zürich in 2016. He then worked in the insurance and banking industry before joining the UQOD group at Idiap Research Institute and enrolling as a Ph.D. student at University of Bern in November 2018. His thesis focuses on stochastic approaches to estimate implicit sets under indirect measurements. This thesis in funded by the Swiss National Science Fundation under project nr. 178858.