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

Cedric Travelletti [1] David Ginsbourger [1] Niklas Linde [2]

[1]University of Bern [2]University of Lausanne

Problem Overview

Want to recover the *matter density field* inside the **Stromboli** volcano:

$$\rho : D \to \mathbb{R}$$

from observations of the (vertical component of) the induced **gravity field** at points $s_1, ..., s_n$ on the surface:

$$\mathcal{G}_{s_i}[\rho] = \int_D \rho(x) \frac{x^{(3)} - s^{(3)}}{||x - s_i||^3}$$

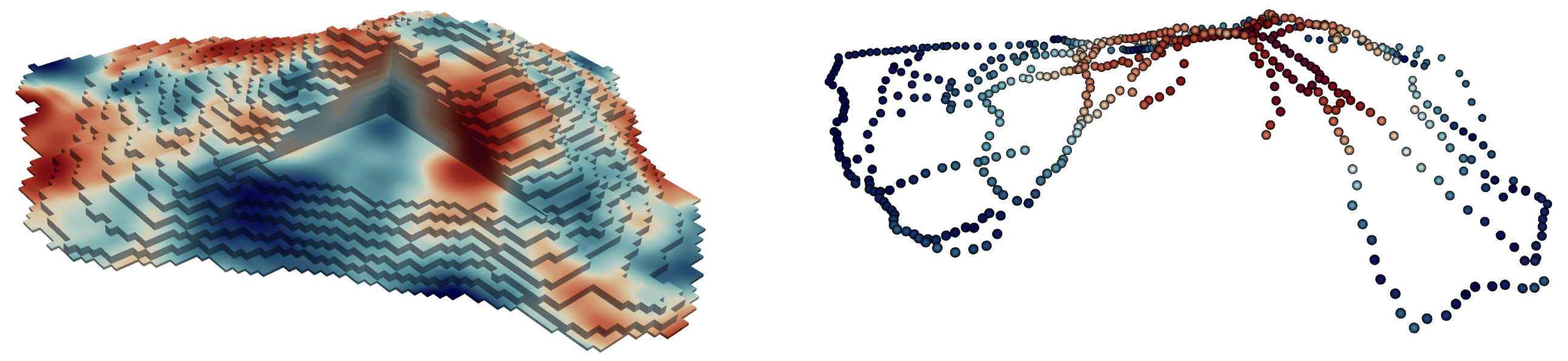

Figure 1. Problem Overview: Underground mass density (simulated), vertical intensity of the generated gravity field (arbitrary colorscale)..

We here restrict ourselves to **continuous** density fields. Available data may then be described as an linear operator between Banach spaces $\mathcal{G} : C(D) \to \mathbb{R}^n$ (the **forward operator**). Our task is then to recover the continuous scalar field ρ on D from the data

$$\boldsymbol{y} = \mathcal{G}[\rho] + \boldsymbol{\epsilon}$$

where $\boldsymbol{\epsilon} \sim \mathcal{N}(0, \tau^2 \boldsymbol{I})$ is some Gaussian-distributed noise. This is a so-called **Inverse Problem**.

Bayesian Inversion

Inversion may be performed in a Bayesian way by putting a prior on $C(D)$ and using the posterior to approximate ρ. In the following, let $Z \sim Gp(\mu, k)$ denote a Gaussian Process on D with (almost surely) continuous trajectories.

After discretization of the forward operator, posterior mean and covariance can be computed analytically. Given a set of **grid points** $\boldsymbol{W} = (w_1, ..., w_m)$ in D, discretize the forward into an $n \times m$ matrix G:

$$\mathcal{G}[\rho] \approx G(\rho(w_1), ..., \rho(w_m))^T.$$

Then, given two sets of points $\boldsymbol{X} = (x_1, ...x_k)^T$ and $\boldsymbol{X'} = (x'_1, ..., x'_l)$ in D and using $K_{\boldsymbol{X}\boldsymbol{X'}}$ to denote the $k \times l$ matrix with elements $k(x_i, x'_j)$, the posterir mean and covariance are given by:

$$\tilde{\mu}_{\boldsymbol{X}} = \mu_{\boldsymbol{X}} + K_{\boldsymbol{X}\boldsymbol{W}} G^T \left(G K_{\boldsymbol{W}\boldsymbol{W}} G^T + \tau^2 \boldsymbol{I}\right)^{-1} (\boldsymbol{y} - G\mu_{\boldsymbol{W}})$$
$$\tilde{K}_{\boldsymbol{X}\boldsymbol{X'}} = K_{\boldsymbol{X}\boldsymbol{X'}} - K_{\boldsymbol{X}\boldsymbol{W}} G^T \left(G K_{\boldsymbol{W}\boldsymbol{W}} G^T + \tau^2 \boldsymbol{I}\right)^{-1} G K^T_{\boldsymbol{X'}\boldsymbol{W}}.$$

Challenges

It turns out that the conditioning equations become **intractable** when the problem is discretized on fine resolution grids. In the following, let $\boldsymbol{X}$ denote a set of m points used to discretize the volcano, also let $\boldsymbol{W} = \boldsymbol{X}$.

- **Forward operator** is an integral operator and hence involves all points of the discretization.
- **Typical discretization size** for Stromboli volcano is into cubic cells of 50m side length. This yields rougly $m = 200k$ discretization points.
- Formulae involve $m \times m$ matrices, which translate to **hundreds of gigabytes of memory**.
- **Intermediate matrix** $K_{\boldsymbol{W}\boldsymbol{W}}$ and **posterior covariance** $\tilde{K}_{\boldsymbol{X}\boldsymbol{X}}$ become **too large for storage** .

Sequential Inversion

The problem gets even worse when data is assimilated in stages $i = 1, ..., n$, with G_i denoting forward operator at data assimilation stage i and $K^{(i)}$ denoting posterior covariance at stage i.

Implicit Covariance Representation

Fast, sequential updating of the posterior is achieved by introducing an **implicit representation** and **chunking**.

- **Chunking:** Prior covariance matrix $K^{(0)}$ may be constructed on-demand (analytical formula exists for each element). So products $K^{(0)}A$ can be computed in chunks by only constructing *bands* of $K^{(0)}$.
- **Implicit Representation:** Only maintain and update multiplication routine for the posterior covariance. All computations either brought back to multiplication with the prior (which can be performed by chunking) or reduced to multiplication with small matrices.

Multiplication routine for *thin* matrices $A \in \mathbb{R}^{m \times q}$, $q \ll m$ is maintained at every stage (thin ≈ small enough so that resulting product can fit in memory).

$$\mathit{CovMul}_n : A \mapsto K^{(n)}A.$$

Routine may then be updated iteratively thanks to:

Lemma

For any $n \in \mathbb{N}$ and any $m \times q$ matrix A:

$$K^{(n)}A = K^{(0)}A - \sum_{i=1}^{n} \bar{K}_i R_i^{-1} \bar{K}_i^T A,$$

with intermediate matrices $\bar{K}_i$ and R_i^{-1} defined as:

$$\bar{K}_i := K^{(i-1)} G_i^T,$$
$$R_i^{-1} := \left(G_i K^{(i-1)} G_i^T + \tau^2 \boldsymbol{I}\right)^{-1}.$$

So multiplication routine at each stage may be defined by **only storing low rank matrices** at each data acquisition stage.

Posterior Sampling

Our implicit representation integrates well with **residual kriging** to allow sampling from the posterior on large grids.

Excursion Set Estimation

Goal: Estimate high density regions $\Gamma^* := \{x \in D : \rho(x) \geq T\}$.

Set Estimation: Posterior gives rise to a *random set* $\Gamma := \{x \in D : \tilde{Z}_x \geq T\}$, where $\tilde{Z}$ is any GP distributed according to the posterior.

Can then use the **excursion probability** $p_\Gamma(x) := \mathbb{P}[x \in \Gamma]$ to produce estimator (**Vorob'ev expectation**):

$$\hat{\Gamma}_V := \{x \in D : p_\Gamma(x) \geq \alpha_V\}$$

where α_V chosen such that $\hat{\Gamma}_V$ has same volume as expected volume of Γ (computable).

Use Case: Sequential Design for Excursion Set Estimation

Sequential Design of Experiment: At stage n, goal is to choose next observation point s_{n+1} on surface of volcano, using data available up to stage n in order to decrease uncertainty on our estimate of Γ.

Can do this by maximizing *weighted integrated variance reduction* (wIVR) criterion:

$$\mathrm{wIVR}^n(s) = \int_D \left(K^{(n)}_{xx} - K^{(n+1)}_{xx}[G_s]\right) p_n(x) dx,$$

where $K^{(n+1)}$ denotes the conditional covariance after including the observations described by G_s (this is independent of the observed realisation of the data) and p_n denotes the *coverage function* at stage n (according to the posterior at stage n).

- Requires computation of (hypothetical) covariance matrices on full grid $\Longrightarrow$ too expansive for traditional approaches, requires **implicit representation**.
- Criterion optimized for single new observation point at each stage (myopic).
- Candidate locations chosen among finite set (nearest neighbors from last stage, ...).

Results:

Run wIVR strategy on Stromboli volcano discretized at $50m$ resolution. Compare reconstruction accuracy (using Vorob'ev expectation) for different strategies. Compare to best possible reconstruction (**infill**) which corresponds to gathering data at each point of the (discretized) surface.

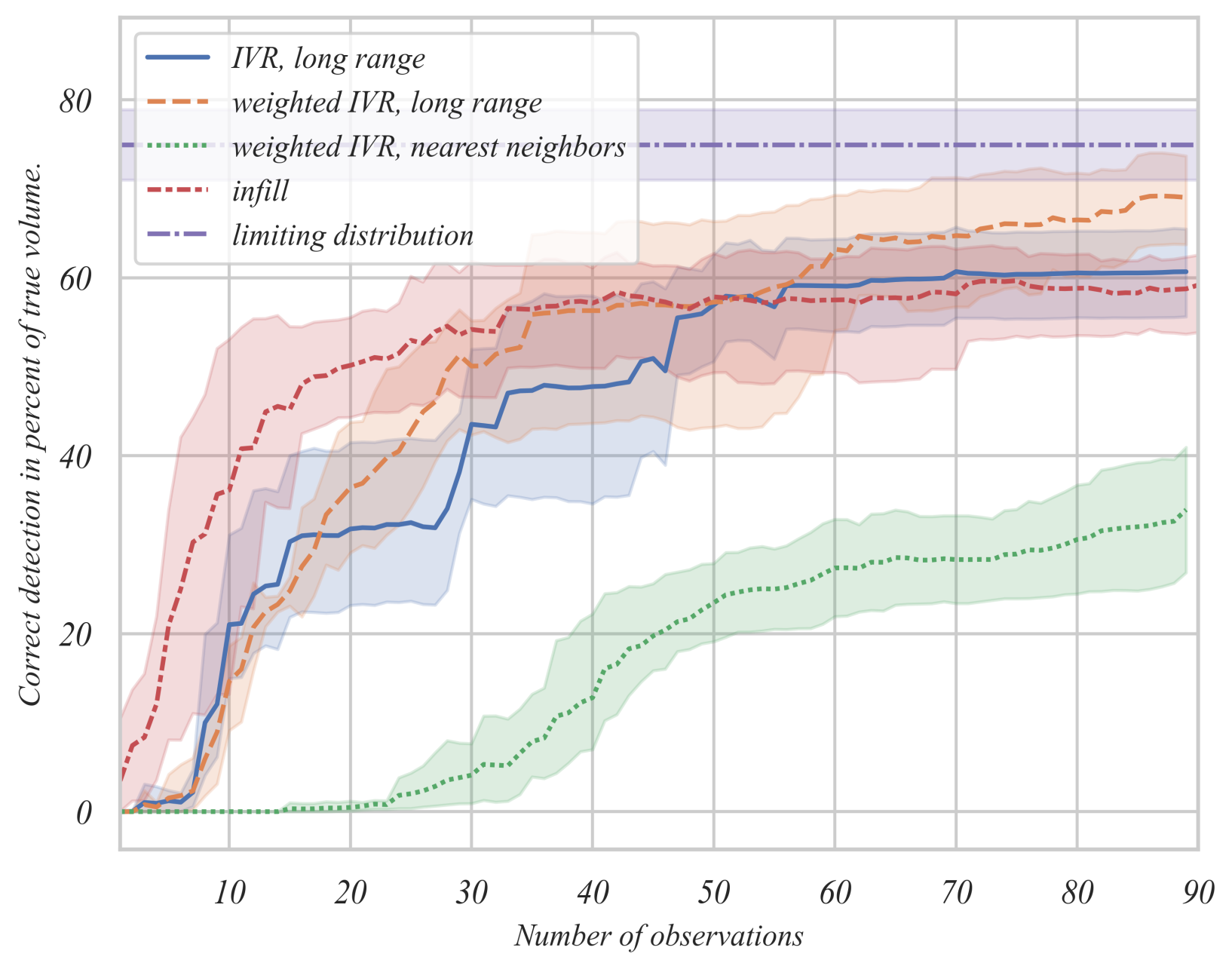

Figure 2. Comparison of excursion set reconstruction quality for different strategies.

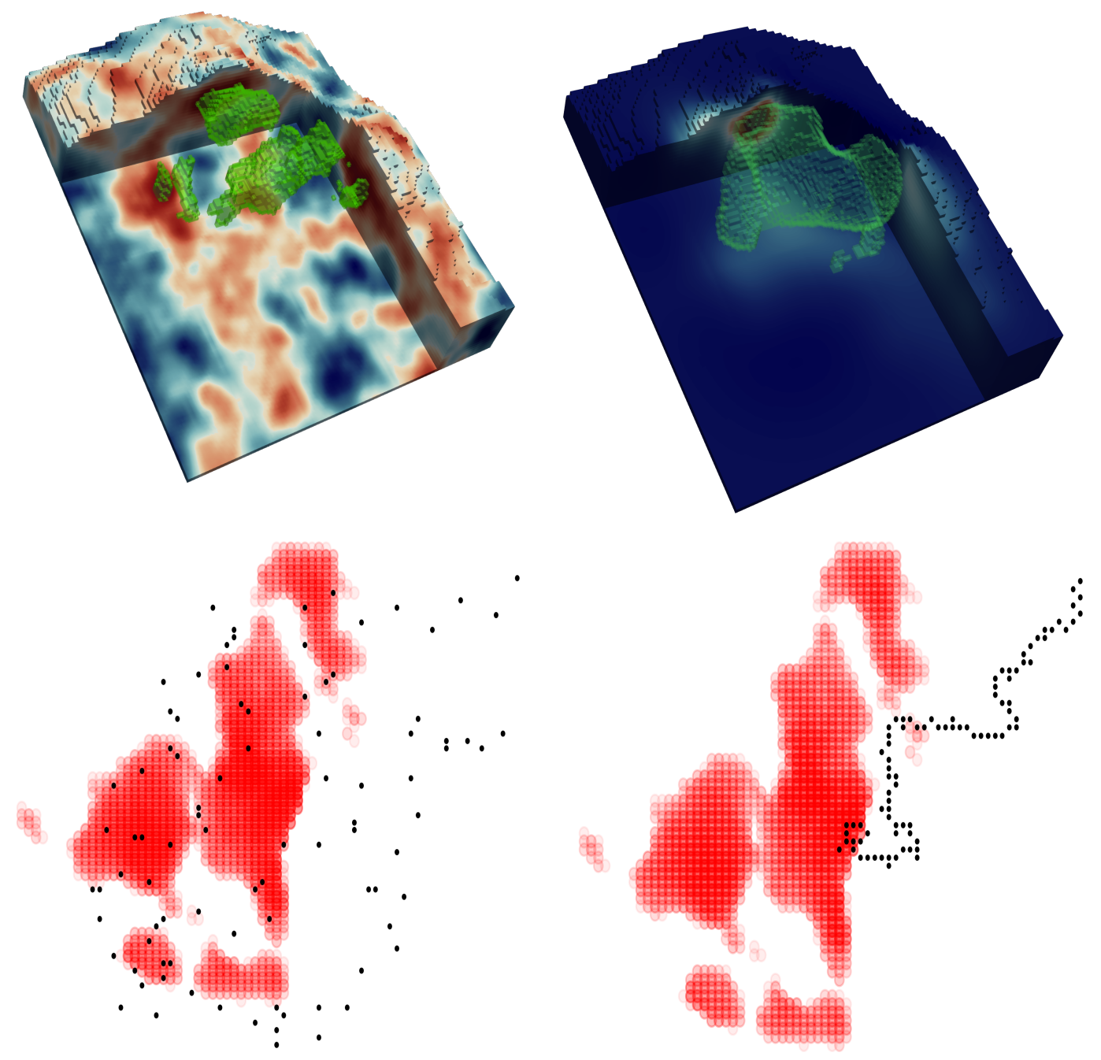

Figure 3. Ground truth (with excursion set), Vorob'ev estimate at end of wIVR strategy and locations visited by the strategy (long steps and nearest neighbors).

MASCOT NUM 2021

cedric.travelletti@stat.unibe.ch