Uncertainty Quantification and Experimental Design for Large-Scale Inverse Problems, with Applications to Geophysics.

Cedric Travelletti

April 13, 2022

Joint work with D. Ginsbourger (UniBe) and N. Linde (UniL)

Motivating Example: Stromboli Volcano

Want to learn the interior structure of the Stromboli volcano.

- Only allowed to measure gravitational field on the surface

Question

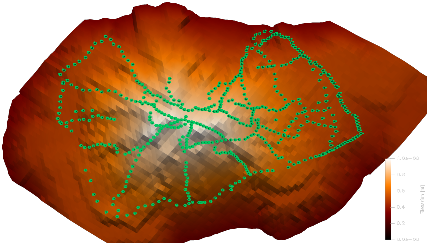
Where should I collect data to get the best possible reconstruction?

Section 1

Problem Setup

Problem Setup

- **Want to recover:** unknown density field $\rho : D \to \mathbb{R}$
- measurement sites $s_1, ..., s_n \in S$ on the surface
- **Available data:** surface gravity field $\{G_{s_i}[\rho]\}_{i=1,...,n}$.



Measurement (forward) operator

$$G_{s_i}[\rho] = \int_D \rho(x) \frac{x^{(3)} - s_i^{(3)}}{\|x - s_i\|^3} dx$$

Solve problem in a Bayesian way by putting a GP prior on subsurface density field.

- Available observations (gravity) are linear forms of the field.
- Use conditional distribution to approximate unknown field ρ.

GP posterior in Inverse Problems

GPs can handle linear operator data: $y_i = G_i(f) = \int_D f(x) d\lambda_i(x)$, where $G_i : C(D) \to \mathbb{R}$ is a linear form.

$$m_n(x) = m(x) + K_{x\boldsymbol{G}_n} K^{-1}_{\boldsymbol{G}_n\boldsymbol{G}_n} (\boldsymbol{y}_n - \boldsymbol{G}_n m.)$$

$$k_n(x, x') = k(x, x') - K_{x\boldsymbol{G}_n} K^{-1}_{\boldsymbol{G}_n\boldsymbol{G}_n} K^{T}_{x'\boldsymbol{G}_n}$$

$$K_{x\boldsymbol{G}_n} = \left(\int_D k(x, y)\, d\lambda_i(y)\right)_{i=1,\dots,n}$$

$$K_{\boldsymbol{G}_n\boldsymbol{G}_n} = \left(\int_D \int_D k(y, z)\, d\lambda_i(y)\, d\lambda_j(z)\right)_{i,j=1,\dots,n}$$

(Discrete) Bayesian Inversion

Discretize inversion domain into cubic cells of fixed side length.

- discretize GP prior on m cells: $\mathfrak{D} = \{x_1, ..., x_m\}$.
- Prior mean $\mu_0 = (\mu(x_i))_{i=1,...,m}$, covariance matrix $K_{ij} = k(x_i, x_j)$.

Posterior is gaussian with mean vector and covariance matrix

$$\tilde{\mu} = \mu_0 + KG^T \left(GKG^T + \tau^2 I \right)^{-1} (\mathbf{y} - G\mu_0)$$

$$\tilde{K} = K - KG^T \left(GKG^T + \tau^2 I \right)^{-1} GK$$

Get posterior by updating mean vector and covariance matrix.

Challenges and Limitations

Conditioning equations involve matrices that are impossible to store for (moderately) fine-grained inversion.

$$\tilde{\mu} = \mu_0 + KG^T\left(GKG^T + \tau^2 I\right)^{-1}(\mathbf{y} - G\mu_0)$$
$$\tilde{K} = K - KG^T\left(GKG^T + \tau^2 I\right)^{-1}GK$$

- Covariance matrix quadratic in number of inversion cells ($\mathcal{O}(m^2)$ storage).
- Forward operator is *dense*: each datapoint influenced by **all** cells in discretization $\implies$ No Sparsity.
- SPECIFIC TO INVERSE PROBLEMS.

Impossible to store covariance matrices for *real-world sized* problems.

Challenges and Limitations: Stromboli Memory Needs

$$\tilde{\mu} = \mu_0 + KG^T \left(GKG^T + \tau^2 I\right)^{-1} (\mathbf{y} - G\mu_0)$$

$$\tilde{K} = K - KG^T \left(GKG^T + \tau^2 I\right)^{-1} GK$$

Matrix	# Elements	Storage
K	$2.9 * 10^{10}$	115 GB
KG, G	$9.2 * 10^{7}$	369 MB
$(\cdots)^{-1}$	$2.9 * 10^{5}$	1.2 MB
μ	$1.7 * 10^{5}$	0.7 MB

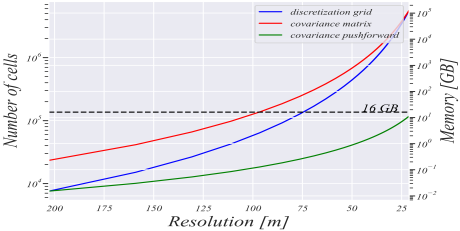


Figure: Grid and matrices size vs resolution on Stromboli example.

Compared to GP-regression, different dimensions at play.
(specific to inversion).

Summary of Challenges for large GP-based Inversion

- Larger-than-memory covariance matrices
- Difficult to exploit sparsity when observation operators are dense (inversion-specific).
- Sequential updates impossible (cannot store intermediate covariances).

Section 2

Solving the Memory Bottleneck

Solving the Memory Bottleneck

- Covariance matrix too big? $\implies$ Don't store it.

Observation

Posterior covariance information may be extracted via products with *tall and thin* matrices:

$$\tilde{K}A, \ A \in \mathbb{R}^{m \times p}, \ p \ll m$$

$\implies$ Only need to maintain a multiplication routine.

Introduce an (almost-) matrix-free implicit representation of the posterior covariance matrix.

Implicit Representation: Sequential Setting

Consider sequential data assimilation setting.

- Measurements $G_1, ..., G_n$.
- Covariance after inclusion of first n batches: $K^{(n)}$.
- Do not compute $K^{(n)}$, only maintain a right-multiplication routine.

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

- Update this *implicit* representation at every new data inclusion.

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

$$\bar{K}_i := K^{(i-1)} G_i^T,$$

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

Implicit Representation: Sequential Setting (contd.)

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

$$\bar{K}_i := K^{(i-1)} G_i^T,$$

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

- Only need to store low rank intermediate matrices:
  - $m \times d_n$ matrix $K^{(i-1)} G_i^T$
  - $d_n \times d_n$ matrix $\left(G_i K^{(i-1)} G_i^T + \tau^2 \boldsymbol{I} \right)^{-1}$
- Only involves right-multiplication with previous stage covariance.
- Need to be able to compute product with prior $K^{(0)}$ (easy if comes from kernel).

Implicit Representation: Advantages

- Drastically reduced memory footprint.

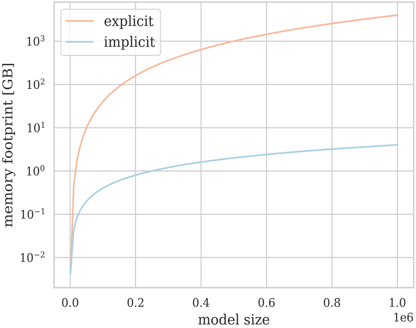


Figure: Memory footprint of posterior covariance vs grid size.

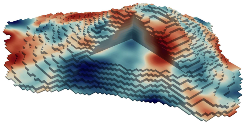
- Fast inclusion of new data.
- Update done in small chunks $\Longrightarrow$ can send to GPU.

Section 3

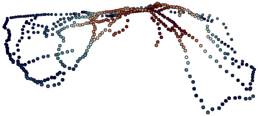
Optimal Design

Example Application: Excursion Set Estimation

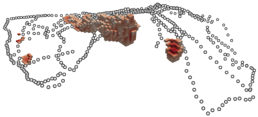
- Want to recover high-density regions.
- Correspond to geological features of interest.
- Propose *adaptive* data collection plan.



(a) density



(b) observation sites (static design)



(c) excursion set

Stepwise Uncertainty Reduction and Random Sets

Want to recover (true) excursion set $\Gamma_{\text{true}} := \{x \in D : \ \rho(x) > T\}$.

Sequential Uncertainty Reduction and Random Sets

- Collect data $G_1, ..., G_n$.
- Compute posterior.
- Approximate through random set $\Gamma^{(n)} := \{x \in D : \ Z_x^{(n)} > T\}$, where $Z^{(n)}$ distributed according to posterior.

Coverage Function

$$p_n : D \rightarrow [0, 1]$$
$$p_n(x) := \mathbb{P}\left[x \in \Gamma^{(n)}\right].$$

Weighted IVR Criterion

Select observations by sequentially optimizing an uncertainty reduction criterion (SUR framework).

(weighted) IVR criterion

$$\mathrm{IVR}(s) := \int_D \left(\mathrm{Var}_n\left[Z_x^{(n)}\right] - \mathrm{Var}_n\left[Z_x^{(n)} \mid G_{(s)}\right]\right) p_n(x) dx$$

Where n data ingestion steps have already been performed and variances and excursion probabilities are computed under the current (stage n) conditional law of the field.

- Does not depend on next observed data.

IVR and Implicit Update Framework

$$\mathrm{IVR}(G) \cong \sum_{i=1}^{m} \left(K G^T \left(G K G^T + \tau^2 I \right)^{-1} G K \right)_{ii}$$

- Computation of IVR without update formulae requires inverting concatenated dataset.
- Update formulae allow to compute only contribution of new observation G.
- Cost scales cubically in the size of the new observations (comparted to cubic in dataset size for direct approach).

Experiments

- Train GP on field campaign data.
- Sample 'realistic' volcanoes from model.
- Assess criterion in-silico using artificially sampled ground truths.

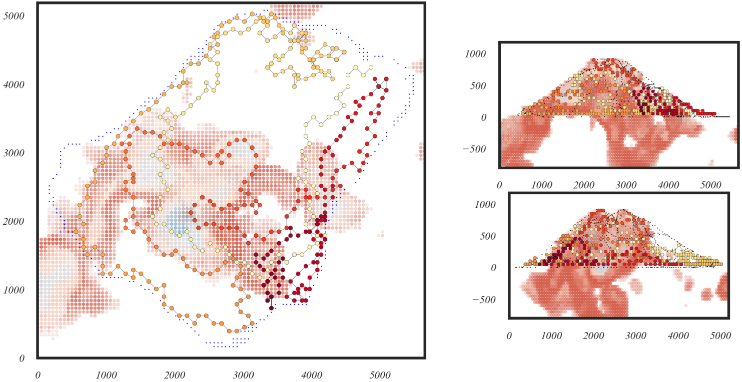


Figure: Excursion set and proposed data collection plan (tIMSE).

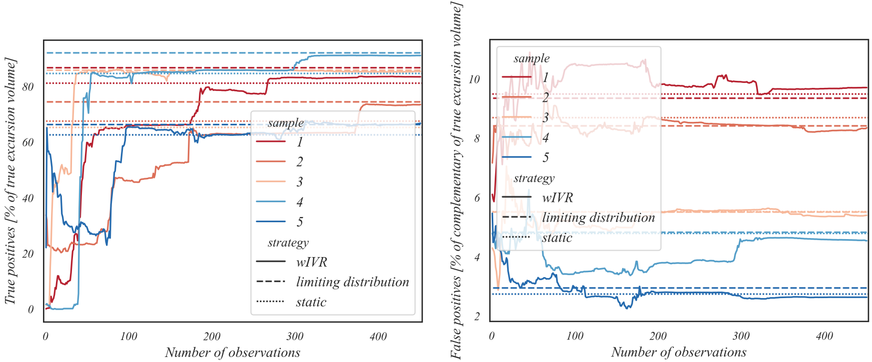
Diagnostics: Limiting Distribution

Since implicit representation allows fast inclusion of new datapoints, can study the *limiting distribution*.

Limiting Distribution

- Posterior distribution after data collected at all accessible locations.
- Gives sense of *minimal residual uncertainty* (inherent to this type of data).

Results



(a) True positives

(b) False positives

Figure: Evolution of true and false positives for the *small* scenario as a function of the number of observations.

UQ on Excursion Volume

Sample from posterior volume distribution on 5 different ground truths.

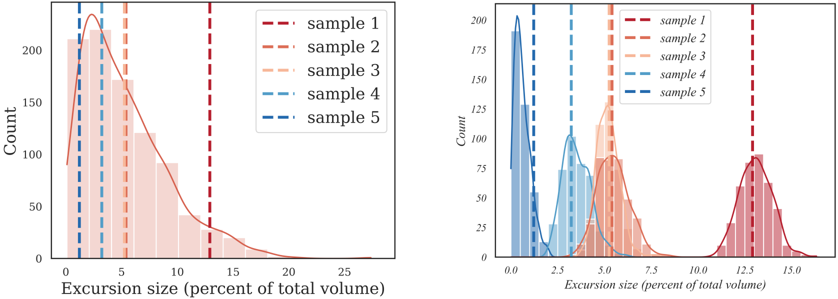
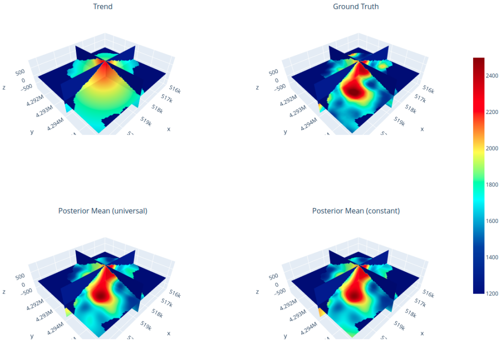


Figure: Prior (left) and empirical posterior (right) distribution (after 450 observations) of the excursion volume for each ground truth. True volumes are denoted by vertical lines.

Demonstrate uncertainty reduction and peaking around true volume.

Future Directions

- Infer trends (extend to universal kriging type inversion).



- Conservative uncertainty reduction strategies.
- Cost-aware path planning.

The End

Travelletti, C., Ginsbourger, D., Linde, N. (2021). *Uncertainty quantification and experimental design for large-scale linear inverse problems under gaussian process priors.* arXiv preprint arXiv:2109.03457.