Methods for out-of-memory Bayesian Inversion with a View towards Optimal Design of Experiments

Cedric Travelletti

University of Bern

August 17, 2020

Joint work D. Ginsbourger (UniBE) and N. Linde (Unil)

Section 1

Motivating Example

Motivating Example: Stromboli Volcano

Want to learn the interior structure of the Stromboli volcano.

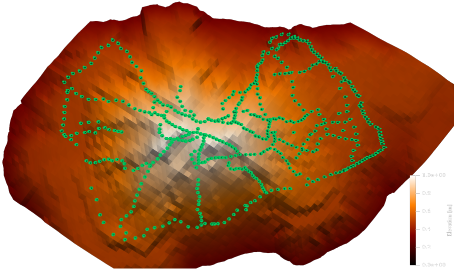
- Only allowed to measure gravitational field on the surface

Question

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

This is a linear inverse problem

- unknown density field $\rho : D \to \mathbb{R}$
- measurement site $s_1, \ldots, s_n \in S$ on the surface
- recover ρ from the data $\{G_{s_i}[\rho]\}_{i=1,\ldots,n}$.



Measurement (forward) operator

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

Bayesian Inversion

- Define a GP prior $Z \sim Gp(\mu, k)$ on D.
- Assume the unknown ρ is a realization of Z.

Data Model

$$\mathbf{y} = (y_i)_{i=1,\dots,n}, \; y_i = G_{s_i}[Z] + \epsilon$$

- Given data $\mathbf{y}$, compute conditional law of Z, conditional on the data.
- Use conditional law to approximate ρ.

Bayesian Inversion: Concrete Implementation and Challenges

- Discretization $\mathfrak{D} = \{x_1, ..., x_m\}$ into m cells.
- GP turns into gaussian vector.
- 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

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

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

Get posterior by updating mean vector and covariance matrix.

Challenges and Limitations

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

- Covariance matrix is $m \times m$.
- Forward operator is *dense*: each datapoint influenced by **all** cells in discretization.
- $\implies$ No Sparsity.

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

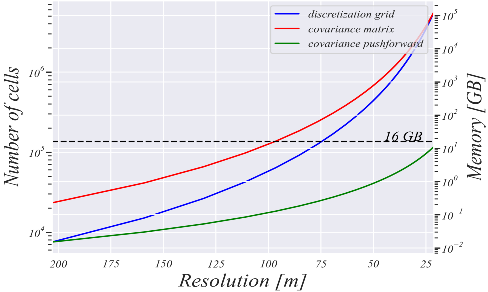


Figure: Grid and matrices size vs resolution on Stromboli example.

- From now on, only consider *large scale setting*.
- Large := discretization fine enough ($m \gtrsim 100k$ cells) so that covariance matrices do not fit in memory on a laptop.
- Big number of datapoints has already been considered [WPG+19], but big number of model points (discretization) not treated in the litterature.

Section 2

Implicit quasi-matrix free Representation of Covariance Matrices

Preliminary: Computing the Posterior Mean

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

- Only involves K through $K^{\#} := KG^T$.
- Each element of K is defined *implicitly* by a formula $K_{ij} = k(x_i, x_j)$.
- Can build elements of K on the fly.
- Matrix-Matrix products easy to parallelize (line by line).

Algorithm

- *Distribute chunks among computational units.*
  - *Each unit builds corresponding lines of K (and all columns).*
  - *Each unit computes corresponding lines of the product KG^t.*
- *Gather results and assemble complete product on main computational unit.*

Implicit Representation of Covariance Matrix

Observation

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

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

Implicit Representation of Covariance Matrix

Observation

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

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

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.

$$CovMul_n : A \mapsto K_n A$$

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

Update Algorithms

Algorithm 1 Covariance Right Multiplication Procedure after n conditioning

Require: Precomputed matrices $K_1^{\#}, ..., K_n^{\#}$ and $R_1^{-1}, ..., R_n^{-1}$.
Prior covariance right-multiplication routine $CovMul_0$.
Input matrix A.
Ensure: $K_n A$

procedure $CovMul_n$(A)
Compute $K_0 A$ using prior right-multiplication routine.
Return $K_0 A - \sum_{i=1}^{n} K_i^{\#} R_i^{-1} K_i^{\# T} A$.

Algorithm 2 Updating intermediate quantities at conditioning step n

Require: Last step covariance right-multiplication routine $CovMul_{n-1}$.
Measurement matrix G_n, measurement noise covariance Δ_n.
Ensure: Step n intermediate matrices $K_n^{\#}$ and R_n^{-1}

procedure $CovUpdate_n$
Compute $K_n^{\#} = K_{n-1} G_n^T$ using last step right-multiplication routine.
Compute $R_n^{-1} = \left(G_n K_n^{\#} + \Delta_n\right)^{-1}$.

Complexity

Proposition

Say there are N datapoints, which we group in n chunks of fixed equal size $d = \frac{N}{n}$. Then:

- *the storage requirement to define $CovMul_n$ is $\mathcal{O}(N(d+m))$,*
- *the amount of computation required to define $CovMul_n$ is $\mathcal{O}(N^2 m^2 d)$.*

Also, if $A \in \mathbb{R}^{m \times p}$, $p \ll m$, then the number of operations required to compute $K_n A$ is $\mathcal{O}(Nm^2 p)$.

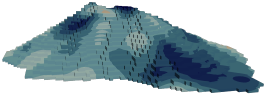
Conclusion

- It is possible to extract information from the posterior covariance matrix, even when its size is way larger than the available memory.
- Sequential data assimilation in this context is also feasible.

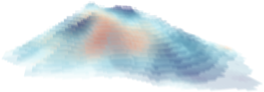
Section 3

Optimal Design

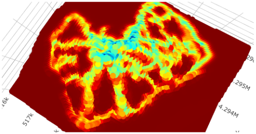
Use posterior covariance information to guide future data acquisition.



(a) simulated surface gravity



(b) reconstructed internal mass density



(c) residual variance

Implicit update formula allow extraction of posterior covariances for large-scale problems.

Criterion for Optimal Reconstruction: IVR

Question

Say we have a data collection plan: observation locations $s_1, ..., s_n$, corresponding measurement operator G.

- Can we assess (a priori) how much information this plan will provide?
- Can we compare different data collection plans?
- Can we compute an *optimal* plan?

Possible criterion: Integrated Variance Reduction:

$$\mathrm{IVR}(G) := \int_D \left(\mathrm{Var}\left[Z_x\right] - \mathrm{Var}\left[Z_x \mid G\right]\right) dx \cong \sum_{i=1}^m \left(KG^T\left(GKG^T + \Delta\right)^{-1} GK\right)_{ii}$$

- Independent of observed data.
- For large data collection plans, can be computed by chunking + update algorithm.
- Can be computed after inclusion of a prior dataset using update algorithm.

IVR on Stomboli

Want to plan field campaign on volcano:

Challenges

- Most locations hard to access $\Longrightarrow$ observations costly.
- Easier to follow existing trails.
- Some locations are inacessible.

Setup

- Collect first batch of observations at sea level on first day.
- Update model.
- Choose where to go next.

IVR on Stomboli

Want to plan field campaign on volcano:

Challenges

- Most locations hard to access $\Longrightarrow$ observations costly.
- Easier to follow existing trails.
- Some locations are inacessible.

Setup

- Collect first batch of observations at sea level on first day.
- Update model.
- Choose where to go next.

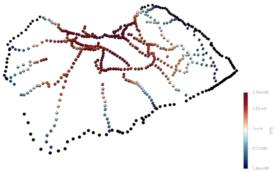


Figure: IVR for observations at different locations on the surface. Already observed data in black.

Figure: IVR for different paths. Already observed data in black.

Next Steps

Long term objectives will leverage the *sequential* aspect of update formulae

- Integrate cost and graph structure.
- Data-dependent criterions.
- Extend to dynamic planning.

But we are currently targetting our efforts towards

Set Estimation

Section 4

(Supplemental) Set Estimation and Uncertainty Quantification on Sets

Set Estimation

We want to identify high density regions (excursion sets)

$$\Gamma^* = \{x \in D : \rho(x) \geq t_0\}$$

A simple plug-in estimate can be obtained using the posterior mean

$$\Gamma_{plug-in} = \{x \in D : \tilde{\mu}(x) \geq t_0\}.$$

Better estimates can be obtained by considering the full posterior distribution.

Random Closed Sets (RACS)

The posterior distribution of the conditional field gives rise to a random closed set (RACS) Γ

$$\Gamma = \{x \in D : \tilde{Z}_x \geq t_0\}$$

Where $\tilde{Z}$ is any Gaussian Process whose law corresponds to the conditional law.
Can consider the pointwise probability to belong to the excursion set

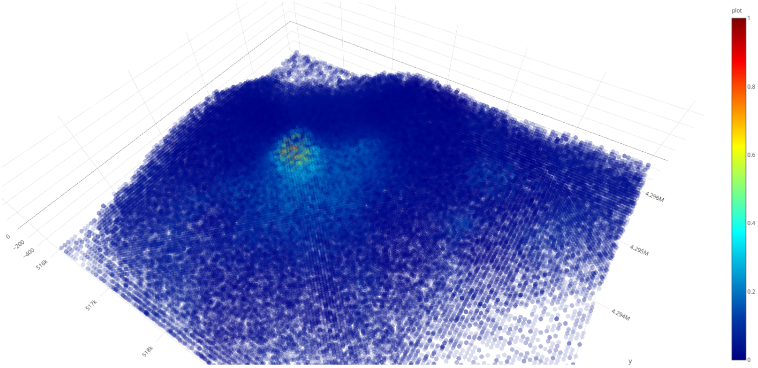
Coverage Function

$$p_\Gamma : D \to [0, 1]$$

$$p_\Gamma(x) := \mathbb{P}[x \in \Gamma]$$

Coverage function

Pointwise probability to belong the the excursion set above 2500 kg/m3.



Random Closed Sets Theory

The coverage function allows us to define a parametric family of set estimates for Γ

Vorob'ev Quantiles

$$Q_\alpha := \{x \in D : p_\Gamma \geq \alpha\}$$

The family of quantiles Q_α gives us a way to estimate Γ by controlling the (pointwise) probability α that the members of our estimate lie in Γ.

- Threshold α controls probability that points in our estimate lie in Γ.
- Can pick it such that the volume of the resulting set is equal to the expected volume of the excursion set

Vorob'ev Expectation

The Vorob'ev expectation is the quantile Q_{α_V} with threshold α_V chosen such that

$$\mu(Q_{\alpha_V}) = \mathbb{E}[\mu(\Gamma)]$$

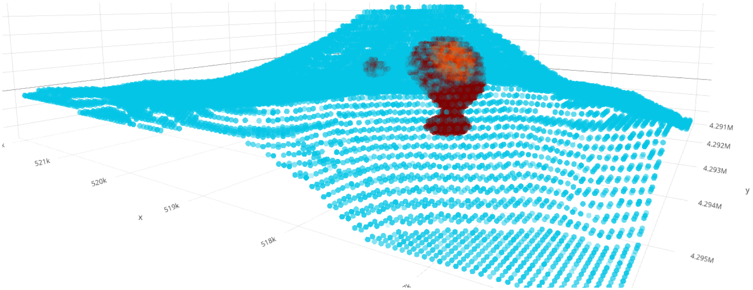
The expected volume of the excursion set can be computed using the coverage function

Robbins Theorem

$$\bar{V}_\Gamma := \mathbb{E}[\mu(\Gamma)] = \int_D p_\Gamma(x) dx$$

Vorob'ev Expectation

Plugin estimate and Vorob'ev expectation for excursion set above 2500.0 kg/m3.



Vorob'ev expectation: $\alpha = 0.22$, expected excursion measure $\mathbb{E}[\mu(\Gamma)] = 6678.16$ cells. Vorob'ev deviation: 7290.031 cells.

Vorob'ev Deviation

Can quantify uncertainty on an estimate Q for Γ by its Vorob'ev deviation

$$\mathcal{D}(Q) := \mathbb{E}[\mu(\Gamma \Delta Q)]$$

Theorem

$$\mathcal{D}(Q) = \int_Q \Big(1 - p_\Gamma(x)\Big) dx + \int_{Q^c} p_\Gamma(x) dx$$

This is the next criterion that we want to investigate.

- Compute designs that optimally recover high-density regions.
- Can leverage closed form formulae for expected excursion probability.

Vorob'ev expectationt achieves the minimum deviation among all sets that have measure equal to the expected measure of Γ.

Theorem

The Vorob'ev expectation minimizes the deviation among closed set with volume $\bar{V}_\Gamma$.

$$Q_{\alpha_V} \in \arg\min\{\mathcal{D}(Q) | Q \subset X \text{ closed}, \ \mu(Q) = \bar{V}_\Gamma\}$$

Thank You

Sources

https://www.itij.com/story/115685/tourists-flee-stromboli-volcano-eruption

Ke Alexander Wang, Geoff Pleiss, Jacob R. Gardner, Stephen Tyree, Kilian Q. Weinberger, and Andrew Gordon Wilson, *Exact gaussian processes on a million data points*, 2019.