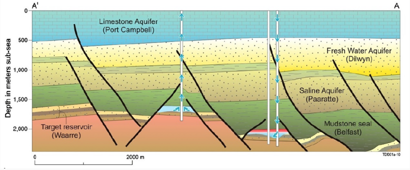
Fast methods for Bayesian inverse problems with uncertain PDE forward models

Ilona Ambartsumyan and Omar Ghattas

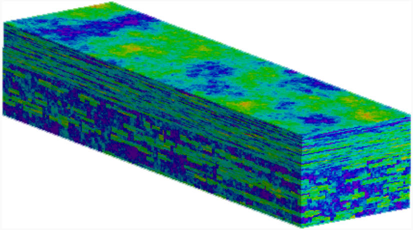
Oden Institute for Computational Engineering and Sciences
The University of Texas at Austin

Example of inverse problem with uncertain forward model

Infer fault transmissibility in a subsurface flow model:



Berkeley Lab, newscenter.lbl.gov



SPE 10, spe.org/web/csp

- PDE model: multi-phase porous medium flow (expensive to solve)
- Observations: fluid pressure and well production (typically sparse and noisy)
- Presence of random parameter field: background permeability (reconstructed from previous studies)

Derivation of the algorithm

Inverse problem governed by random PDE forward problem

Infer the inversion parameter $m \in \mathcal{M}$, given noisy observations $d \in V$ of the state variable $u \in \mathcal{U}$, governed by the random PDE

$$r(u, m, k) = 0$$

where

- Random parameter $k \in \mathcal{K}$ is normally distributed, $k \sim \mathcal{N}(\bar{k}, \Gamma_k)$
- Observations are polluted by Gaussian additive noise, $d = \mathcal{B}u(k, m) + \beta, \qquad \beta \sim \mathcal{N}(0, \Gamma_\beta)$
- Prior information about m is given by the measure μ_{pr}

Bayesian solution of inverse problem

By (infinite dimensional) Bayes' Theorem,

$$\frac{d\mu_{\text{post}}}{d\mu_{\text{pr}}} = \frac{1}{\mathcal{Z}} \pi_{\text{like}}(d|m)$$

Marginalize out the random parameter k:

$$\begin{aligned}\pi_{\text{like}}(d|m) &\propto \int_{\mathcal{K}} \pi(d|m,k)\, d\mu_k \\ &= \int_{\mathcal{K}} \exp\left[-\frac{1}{2}(d - \mathcal{B}u(m,k))^T \Gamma_\beta^{-1}(d - \mathcal{B}u(m,k))\right] d\mu_k\end{aligned}$$

For sufficiently complex forward models, this integral is prohibitive to approximate via Monte Carlo sampling (and it has to be done for each sample of m)

Approximation of the likelihood

Consider the first-order approximation of u wrt k:

$$u^L(m,k) = u(m,\bar{k}) + D_k u(m,\bar{k})[k-\bar{k}]$$

where D_k is the Fréchet derivative of u wrt k

The likelihood based on u^L becomes:

$$\pi^L_{\text{like}} \propto \exp\left(\frac{1}{2}\|d - \mathcal{B}u(m,\bar{k})\|^2_{\Gamma_\nu^{-1}} - \frac{1}{2}\log\,\det(\Gamma_\nu)\right)$$

where $\Gamma_\nu := \Gamma_\beta + (\mathcal{B}D_k u)\Gamma_k(\mathcal{B}D_k u)^T$

The term $(\mathcal{B}D_k u)\Gamma_k(\mathcal{B}D_k u)^T$ is the contribution of model uncertainty to the noise covariance, and results from pushing forward the covariance of k, Γ_k, through the Jacobian of the parameter-to-observable map, $\mathcal{B}D_k u$.

Low rank approximation of sensitivity matrix

- As m changes across optimization iterations/MCMC sampling steps, we need to evaluate (and later compute derivatives wrt m of) the term

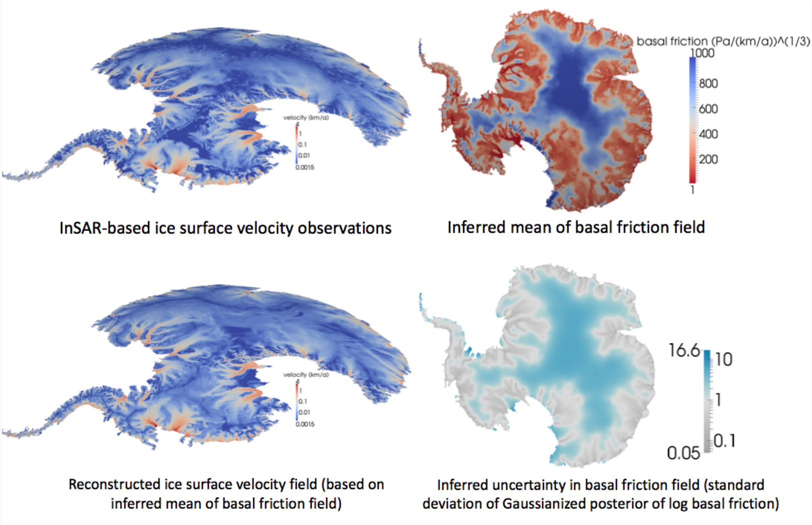
$$\Gamma_\beta + \overbrace{(\mathcal{B}D_k u)}^{J(m,\bar{k})} \Gamma_k \overbrace{(\mathcal{B}D_k u)^T}^{J(m,\bar{k})^T}$$

- $J\,\Gamma_k J^T$ resembles the Gauss-Newton Hessian (of the log-likelihood wrt k), but turned inside-out (i.e. JJ^T operating on data space instead of J^TJ operating on parameter space)
- Explicit construction of J requires as many (linearized) forward PDE solves as $\min$(data dimension, k dimension), which is prohibitive
- For many problems, singular values of $\mathcal{B}D_k u$ decay rapidly and a (randomized) low rank approximation can be made:

$$J\,\Gamma_k J^T \approx V_r \Lambda_r V_r^T$$

  where the columns of V_r contain the r eigenvectors of $J\,\Gamma_k J^T$ and the diagonal elements of Λ_r contains its eigenvalues
- Use randomized SVD which requires $O(r)$ matrix-vector products with random vectors ξ_i, amounting to $O(r)$ linearized adjoint solves $\eta_i = J^T \xi_i$ and linearized forward solves $J\Gamma_k \eta_i$

Example: Antarctic ice sheet flow inversion



InSAR-based ice surface velocity observations

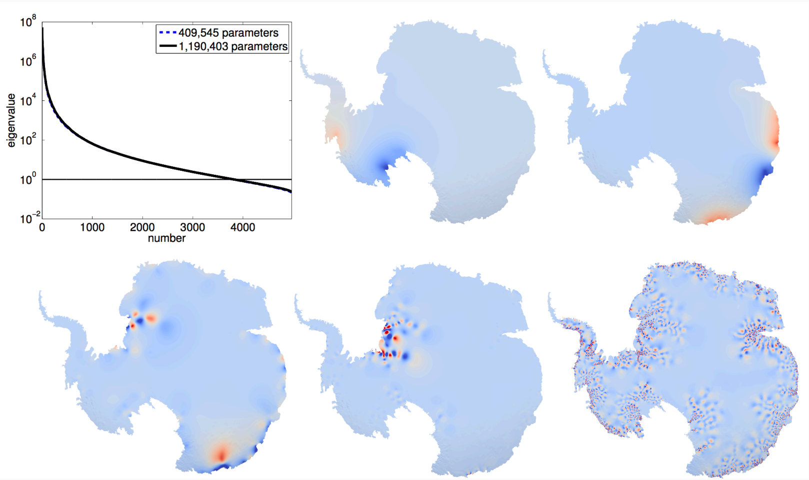
Inferred mean of basal friction field

Reconstructed ice surface velocity field (based on inferred mean of basal friction field)

Inferred uncertainty in basal friction field (standard deviation of Gaussianized posterior of log basal friction)

T. Isaac, N. Petra, G. Stadler, O. Ghattas, JCP 2015

Spectrum of Hessian for Antarctic ice flow inverse problem



$O(i^{-3})$ eigenvalue decay of prior-preconditioned data misfit Hessian (5000 out of 1.19 million) and eigenvectors 1, 7, 100, 500, 4000

Eigenproblem-constrained optimization for MAP point

With the low-rank approximation, the maximum a posteriori estimate, m_{MAP}, is found by minimizing the negative log of the posterior distribution:

$$m_{\mathsf{MAP}} = \underset{m\in\mathcal{M}}{\arg\min} \left(\frac{1}{2}\|\mathcal{B}u(m,\bar{k}) - d\|^2_{(\Gamma_\beta + V_r\Lambda_r V_r^T)^{-1}} + \frac{1}{2}\|\Gamma_{\mathsf{pr}}^{-\frac{1}{2}}(m-\bar{m})\|^2_{L^2(\Omega)} + \frac{1}{2}\log\det(\Gamma_\beta + V_r\Lambda_r V_r^T)\right)$$

where the state solution u depends on the parameters m, $\bar{k}$ through

$$(r(u,m,\bar{k}),\tilde{u}) = 0 \qquad \forall\, \tilde{u}\in\mathcal{U}$$

and the eigenvalues λ_i & eigenvectors v_i depend on u, m, $\bar{k}$ through

$$(J(\bar{k},m,u)\,\Gamma_k J(\bar{k},m,u)^T v_i, \tilde{v}_i) = \lambda_i(v_i,\tilde{v}_i) \qquad \forall\,\tilde{v}_i\in V,\ i=1,\dots,r$$

- Solving the eigenvalue problem via randomized SVD amounts to solving $O(r)$ pairs of linearized adjoint/forward PDEs
- Computing the gradient requires differentiating through the eigenvalue problem to obtain eigenvalue and eigenvector sensitivities, necessitating an solution of $O(r)$ adjoint eigenvalue problems.

Current simpler implementation based on MC estimator

- Have constructed fast algorithms for Hessian-constrained optimization in other contexts (e.g., P. Chen, U. Villa, and O. Ghattas, JCP 2019)
- But somewhat complicated implementation, so initially consider a Monte Carlo estimator of $J\Gamma_k J^T$
- Let e be a standard normal random variable; then

$$
\begin{aligned}
(\mathcal{B}D_k u(m,\bar{k}))\Gamma_k(\mathcal{B}D_k u(m,\bar{k}))^T &= (\mathcal{B}D_k u(m,\bar{k}))\Gamma_k^{1/2}\mathbb{E}[ee^T]\Gamma_k^{1/2}(\mathcal{B}D_k u(m,\bar{k}))^T \\
&= \mathbb{E}[\mathcal{B}D_k u(m,\bar{k}))\Gamma_k^{1/2}e\,(\mathcal{B}D_k u(m,\bar{k}))\Gamma_k^{1/2}e)^T]
\end{aligned}
$$

- Consider a Monte Carlo estimator with $\xi_i \sim \mathcal{N}(0,\Gamma_k)$, $i = 1,\ldots,n_s$,

$$
(\mathcal{B}D_k u(m,\bar{k}))\Gamma_k(\mathcal{B}D_k u(m,\bar{k}))^T \approx \frac{1}{n_s}\sum_{i=1}^{n_s}\left(\mathcal{B}D_k u(m,\bar{k})[\xi_i]\right)\left(\mathcal{B}D_k u(m,\bar{k})[\xi_i]\right)^T
$$

MC estimator-based MAP point computation

The maximum a posteriori estimate, m_{MAP}, is found by minimizing the negative log of the posterior distribution:

$$m_{\text{MAP}} = \underset{m \in \mathcal{M}}{\arg\min} \Big(\frac{1}{2} \|\mathcal{B}u(m, \bar{k}) - d\|^2_{(\Gamma_\beta + \frac{1}{n_s}\sum_{i=1}^{n_s} \mathcal{B}U_i \mathcal{B}U_i^T)^{-1}} + \frac{1}{2}\|\Gamma_{\text{pr}}^{-\frac{1}{2}}(m - \bar{m})\|^2_{L^2(\Omega)} + \frac{1}{2}\log\det(\Gamma_\beta + \frac{1}{n_s}\sum_{i=1}^{n_s} \mathcal{B}U_i \mathcal{B}U_i^T) \Big)$$

where the state solution u depends on the parameters m, $\bar{k}$ through

$$(r(u, m, \bar{k}), \tilde{u}) = 0 \qquad \forall\, \tilde{u} \in \mathcal{U}$$

and the sensitivity solutions, U_i, depend on u, m, $\bar{k}$ through

$$((\partial_u r)(\bar{k}, m, u)[U_i], \tilde{U}_i) = -((\partial_k r)(\bar{k}, m, u)[\xi_i], \tilde{U}_i) \qquad \forall\, \tilde{U}_i \in \mathcal{U},\ i = 1, \ldots, n_s$$

This is an $n_s + 1$ PDE constrained optimization problem, but it can be solved efficiently using gradient based optimization, since n_s is typically small and independent of the parameter dimension. This requires an efficient computation of the gradient...

Gradient computation for the MAP optimization problem

The (weak form of the) gradient is given by

$$(\mathcal{G}(m), \tilde{m}) = (m - \bar{m}, \Gamma_{pr}^{-1}\tilde{m}) + ((\partial_m r)(\bar{k}, m, u)[\tilde{m}], v)$$
$$+ \sum_{i=1}^{n_s} \left[((\partial_{um} r)(\bar{k}, m, u)[U_i, \tilde{m}], V_i) + ((\partial_{km} r)(\bar{k}, m, u)[\xi_i, \tilde{m}], V_i) \right], \forall \tilde{m} \in \mathcal{M}$$

where u and $\{U_i\}_{i=1}^{n_s}$ solve the "forward" problems

$$(r(u, m, \bar{k}), \tilde{u}) = 0 \qquad \forall\, \tilde{u} \in V$$
$$((\partial_u r)(\bar{k}, m, u)[U_i], \tilde{U}_i) = -((\partial_k r)(\bar{k}, m, u)[\xi_i], \tilde{U}_i) \qquad \forall\, \tilde{U}_i \in V,\ i = 1, \ldots, n_s$$

and v and $\{V_i\}_{i=1}^{n_s}$ solve the "adjoint" problems

$$((\partial_u r)(u, m, \bar{k})[\tilde{v}], v) = -\left(\mathcal{B}u - d, \Gamma_\nu^{-1}\tilde{v}\right)$$
$$- \sum_{i=1}^{n_s} \left[((\partial_{uu} r)(\bar{k}, m, u)[U_i, \tilde{v}], V_i) + ((\partial_{ku} r)(\bar{k}, m, u)[\xi_i, \tilde{v}], V_i) \right], \forall\, \tilde{v} \in V$$

$$((\partial_{uU_i} r)(\bar{k}, m, u)[U_i, \tilde{V}_i], V_i) + ((\partial_{kU_i} r)(\bar{k}, m, u)[\xi_i, \tilde{V}_i], V_i) = -\frac{1}{n_s}(\mathcal{B}^T\Gamma_\nu^{-1}\mathcal{B}U_i, \tilde{V}_i)$$
$$+ \frac{1}{n_s}(\mathcal{B}^T\Gamma_\nu^{-1}(\mathcal{B}u - d)(\mathcal{B}u - d)^T\Gamma_\nu^{-1}\mathcal{B}U_i, \tilde{V}_i), \forall\, \tilde{V}_i \in V,\ i = 1, \ldots, n_s$$

Numerical examples

Test case 1: Poisson problem

The forward problem: find $u \in V$ s.t.

$$
\begin{aligned}
-\nabla \cdot (e^m \nabla) u &= k, && \text{in } \Omega, \\
u &= 0, && \text{on } \Gamma_D, \\
e^m \nabla u \cdot n &= 0, && \text{on } \Gamma_N.
\end{aligned}
$$

- The domain is $[0,1] \times [0,1]$;
- Dirichlet BC on $\Gamma_D = \{0\} \times [0,1]$;
- 4225 dofs in state space,
  1089 dofs in $(m,\, k)$ parameter space
- 16 observations with random locations;
- $n_s = 30$ samples;
- 1% Gaussian noise.

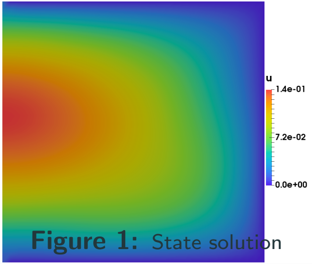


Figure 1: State solution

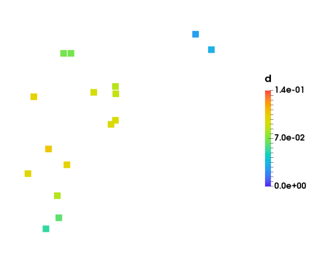


Figure 2: Observations

Test case 1: Random parameter

We assume that $\mu_k \sim \mathcal{N}(\bar{k}, \Gamma_k)$, where the covariance is given by $\Gamma_k = \mathcal{A}^{-2}$ s.t. $\mathcal{A}k = \tilde{k}$ satisfies

$$\gamma_k(\Theta_k \nabla k, \nabla v) + \delta_k(k, v) = (\tilde{k}, v) \quad \forall v \in H^1(\Omega)$$

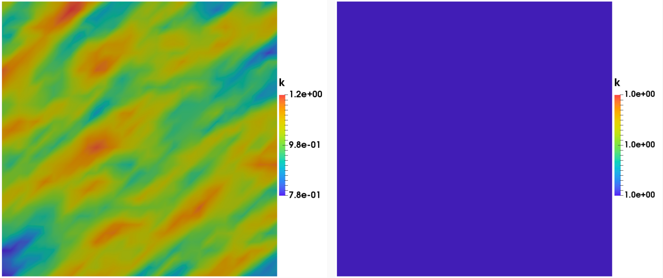


Figure 3: Left: sample from μ_k (used to synthesize data); right: mean of μ_k (used for deterministic inversion)

The prior is given by a distribution with similar structure

Test case 1: Some comments

- Random parameter-to-state map is linear, so approximation is exact
- Small number of observations, not evenly distributed over domain
- Relatively high variance of random parameter distribution, dominates observational noise

Algorithm has been implemented using FEniCS[1] and hIPPYlib[2] open source libraries

[1] https://fenicsproject.org/
[2] https://hippylib.github.io/

Test case 1: MAP point

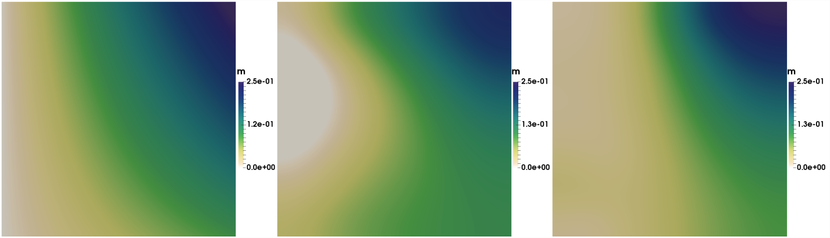


Figure 4: Left: true m; middle: MAP point m^{det}; right: MAP point m^{stoch}

We compare the true field m, the MAP estimate obtained using the deterministic algorithm m^{det}, and the MAP point obtained by the proposed algorithm m^{stoch}.

Test case 1: Eigenvalue of data misfit Hessian (wrt m)

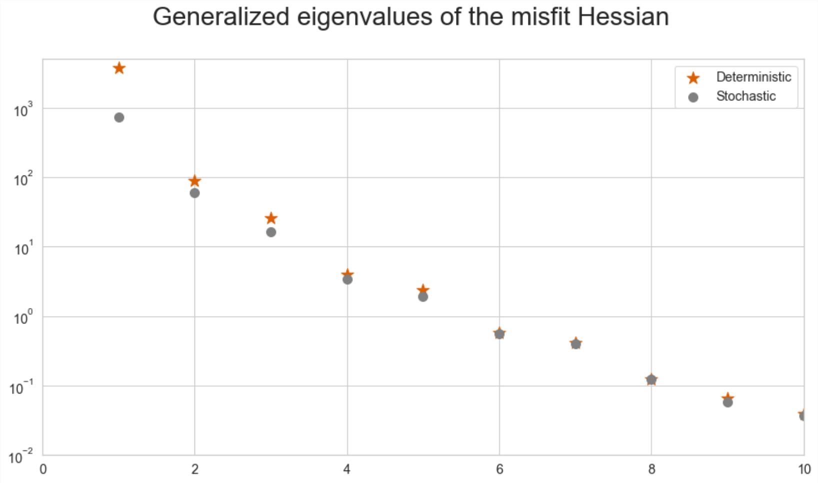


Figure 5: Generalized eigenvalues of prior preconditioned data misfit Hessian (wrt m)

Test case 2: Advection–diffusion–reaction problem

The forward problem: find $u \in V$ s.t.

$$-\nabla \cdot (\kappa \nabla u) + \nabla u \cdot k + cu = m, \quad \text{in } \Omega,$$
$$u = 0, \quad \text{on } \Gamma_D.$$

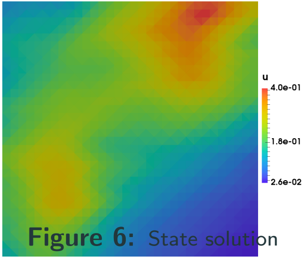


Figure 6: State solution

- The domain is $[0,1] \times [0,1]$;
- Dirichlet BC on $\Gamma_D = \{0\} \times [0,1] \cup [0,1] \times \{0\}$;
- 2048 dofs in state space, 1089 dofs in parameter space
- 50 observations with random locations;
- $n_s = 30$ samples;
- 1% Gaussian noise.

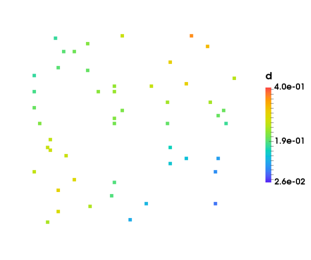


Figure 7: Observations

Test case 2: Random parameter

We assume that $\mu_k \sim \mathcal{N}(\bar{k}, \Gamma_k)$, where covariance given by $\Gamma_k = \mathcal{A}^{-2}$ s.t. $\mathcal{A}k = \tilde{k}$ satisfies

$$\gamma_k(\Theta_k \nabla k, \nabla v) + \delta_k(k, v) = (\tilde{k}, v), \quad \forall v \in H^1(\Omega).$$

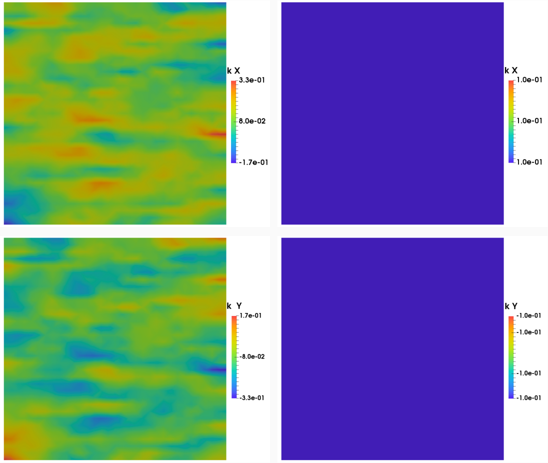


Figure 9: Left: sample from μ_k; right: mean of μ_k.

Test case 2: State solution

The state solution is sensitive to the random parameter

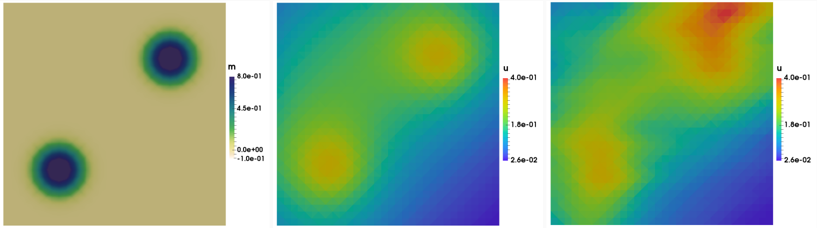


Figure 10: Left: true m; middle: $u(m, \bar{k})$; right: $u(m, k)$.

Diffusion coefficient $\kappa = 10^{-2}$, reaction coefficient $c = 0.4$

Test case 2: Some comments

- Random parameter-to-state map is not linear, so approximation is not exact
- Moderate number of observations, evenly distributed over the domain
- Large variance in random parameter distribution, dominates noise
- Sharp interfaces in the inversion parameter field (TV prior used)

Test case 2: MAP point

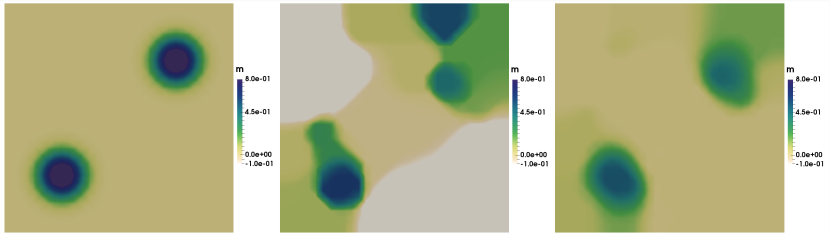


Figure 11: Left: true m; middle: MAP point m^{det}; right: MAP point m^{stoch}.

We compare the true field m, the MAP estimate obtained using traditional algorithm m^{det}, and the MAP point obtained by the proposed algorithm m^{stoch}.

Conclusions

Conclusions

We considered the Bayesian inference of an unknown parameter field m in the presence of uncertainty in another parameter field k

- We proposed a method based on a linearized approximation of the map from k to u that accounts for both the uncertainty and the sensitivity of the state wrt k
- Computation of the negative log posterior and its gradient require a number of PDE solves (n_s+1) that is small and independent of the dimensions of the unknown parameter m and the random parameter k
- Examples illustrate the effect of accounting for model uncertainty in the Bayesian solution

Ongoing and future work

- Approximation of the Jacobian J via randomized truncated SVD (greater accuracy for fewer n_s)
  - Requires addition of eigenvalue problem with forw/adjoint PDEs
- Comparison of our approach to full MCMC + inner MC solution
- Extension of the method to allow for more general distributions μ_k of the random parameter
- Use of the approximation as a control variate:

$$
\begin{aligned}
\pi_{\text{like}}(d|m) &\propto \mathbb{E}_{\mu_k}[\pi(d|m,k)] \\
&= \mathbb{E}_{\mu_k}[\pi^L(d|m,k)] + \mathbb{E}_{\mu_k}[\pi(d|m,k) - \pi^L(d|m,k)] \\
&\approx \exp\left(\frac{1}{2}\|d - \mathcal{B}u(m,\bar{k})\|^2_{\Gamma_\nu^{-1}} - \frac{1}{2}\log\det(\Gamma_\nu)\right) \\
&\quad + \frac{1}{N}\sum_{i=1}^{N}\exp\left(\frac{1}{2}\|d - \mathcal{B}u(m,k_i)\|^2_{\Gamma_\beta^{-1}}\right) \\
&\quad - \frac{1}{N}\sum_{i=1}^{N}\exp\left(\frac{1}{2}\|d - \mathcal{B}u(m,\bar{k}) - \mathcal{B}D_k u(m,\bar{k})[k_i - \bar{k}]\|^2_{\Gamma_\beta^{-1}}\right)
\end{aligned}
$$