Bayesian inference in Inverse problems

Bani Mallick

bmallick@stat.tamu.edu

Department of Statistics, Texas A&M University, College Station
Inverse Problems

- Inverse problems arise from indirect observations of a quantity of interest
- Observations may be limited in numbers relative to the dimension or complexity of the model space
- Inverse problems ill posed
- Classical approaches have used regularization methods to impose well-posedness ans solved the resulting deterministic problems by optimization
Bayesian approach

- A natural mechanism for regularization in the form of prior information
- Can handle non linearity, non Gaussianity
- Focus is on uncertainties in parameters, as much as on their best (estimated) value.
- Permits use of prior knowledge, e.g., previous experiments, modeling expertise, physics constraints.
- Model-based.
- Can add data sequentially
Inverse problem

- Inverse problems whose solutions are unknown functions: Spatial or temporal fields
- Estimating fields rather than parameters typically increases the ill-posedness of the inverse problem since one is recovering an infinite dimensional object from finite amounts of data
- Obtaining physically meaningful results requires the injection of additional information on the unknown field
- A standard Bayesian approach is to employ Gaussian process or Markov Random field priors
Forward Model and Inverse problem

\[ Z = F(K) + \epsilon \]

where

- \( F \) is the forward model, simulator, computer code which is non-linear and expensive to run.
- \( K \) is a spatial field
- \( Z \) is the observed response
- \( \epsilon \) is the random error usually assumed to be Gaussian
- Want to estimate \( K \) with UQ
- This is a non-linear inverse problem
Fluid flow in porous media

- Studying flow of liquids (Ground water, oil) in aquifer (reservoir)
- Applications: Oil production, Contaminant cleanup
- Forward Model: Models the flow of liquid, output is the production data, inputs are physical characteristics like permeability, porosity
- Inverse problem: Inferring the permeability from the flow data
Permeability

- Primary parameter of interest is the permeability field.
- Permeability is a measure of how easily liquid flows through the aquifer at that point.
- This permeability values vary over space.
- Effective recovery procedures rely on good permeability estimates, as one must be able to identify high permeability channels and low permeability barriers.
Forward Model

Darcy’s law:

\[ v_j = -\frac{k_{rj}(S)}{\mu_j} k_f \nabla p, \]  

- \( v_j \) is the phase velocity
- \( k_f \) is the fine-scale permeability field
- \( k_{rj} \) is the relative permeability to phase \( j \) (\( j=\)oil or water)
- \( S \) is the water saturation (volume fraction)
- \( p \) is the pressure.
Forward Model

Combining Darcy’s law with a statement of conservation of mass allows us to express the governing equations in terms of pressure and saturation equations:

\[
\nabla \cdot (\lambda(S) k_f \nabla p) = Q_s, \tag{2}
\]

\[
\frac{\partial S}{\partial t} + v \cdot \nabla f(S) = 0, \tag{3}
\]

- $\lambda$ is the total mobility
- $Q_s$ is a source term
- $f$ is the fractional flux of water
- $v$ is the total velocity
Production (amount of oil in the produced fluid, fractional Flow or water-cut) $F_{k_f}(t)$ is given by

$$F_{k_f}(t) = \int_{\partial\Omega^{out}} v_n f(S) dl$$

where $\partial\Omega^{out}$ is outflow boundaries and $v_n$ is normal velocity field.
Permeability field

Forward Simulator

Output
Fine-scale Permeability field → Forward Simulator → Output
Bayesian way

- If $p(K)$ is the prior for the spatial field $K$: usually Gaussian processes
- $p(Z|k)$ is the likelihood depending on the distribution of $\epsilon$: Gaussian, non-Gaussian
- Then posterior distribution: $p(K|Z) \propto p(Z|K)p(K)$ is the Bayesian solution of this inverse problem
Inverse Problem

- Dimension reduction: Replacing $K$ by a finite set of parameters $\tau$
- Building enough structures through models and priors.
- Additional data: coarse-scale data
- Need to link data at different scales
- Bayesian hierarchical models have the ability to do all these things simultaneously
Multiscale Data

- $K_f$ is the fine scale field of interest (data: well logs, cores)
- Additional data: from coarse scale field $K_c$ (seismic traces)
- Some of the observed fine-scale permeability values $K_f^o$ at some spatial locations
- We want to infer $K_f$ conditioned on $Z$, $K_c$ and $K_f^o$
- The posterior distribution of interest: $p(K_f|Z, K_c, K_f^o)$
\[ \text{Coarse–grid} \quad \text{Fine–grid} \]

\[ \phi = 1 \quad \text{div}(k_f(x)\Delta \phi) = 0 \quad \phi = 0 \]

\[ (k_c(x)e_j, e_l) = \frac{1}{|K|} \int_K (k_f(x)\Delta \phi_j(x), e_l)dx \]
Dimension reduction

- We need to reduce the dimension of the spatial field $K_f$
- This is a spatial field denoted by $K_f(x, \omega)$ where $x$ is for the spatial locations and $\omega$ denotes the randomness in the process
- Assuming $K_f$ to be a real-valued random field with finite second moments we can represent it by Karhunen-Loeve (K-L) expansion
K-L expansion

\[ K_f(x, \omega) = \theta_0 + \sum_{l=1}^{\infty} \sqrt{\lambda_l} \theta_l(\omega) \phi_l(x) \]

where

- \( \lambda \): eigen values
- \( \phi(x) \): eigen functions
- \( \theta \): uncorrelated with zero mean and unit variance
- If \( K_f \) is Gaussian process then \( \theta \) will be Gaussian
If the covariance kernel is $C$ then we obtain them by solving

$$\int C(x_1, x_2) \phi_l(x_2) \, dx_2 = \lambda_l \phi_l(x_1)$$

and can express $C$ as

$$C(x_1, x_2) = \sum_{l=1}^{\infty} \lambda_l \phi_l(x_1) \phi_l(x_2)$$
Spatial covariance

We assume the correlation structure

\[ C(x, y) = \sigma^2 \exp \left( -\frac{|x_1 - y_1|^2}{2l_1^2} - \frac{|x_2 - y_2|^2}{2l_2^2} \right). \]

where, \( l_1 \) and \( l_2 \) are correlation lengths.

For an \( m \)-term KLE approximation

\[
K_f^m = \theta_0 + \sum_{i=1}^{m} \sqrt{\lambda_i} \theta_i \Phi_i, \\
= B(l_1, l_2, \sigma^2) \theta, \text{(say)}
\]

(1)
Existing methods

- The energy ratio of the approximation is given by

\[ e(m) := \frac{E\|k_f^m\|^2}{E\|k_f\|^2} = \frac{\sum_{i=1}^{m} \lambda_i}{\sum_{i=1}^{\infty} \lambda_i}. \]

- Assume correlation length \( l_1, l_2 \) and \( \sigma^2 \) are known.

- We treat all of them as model parameters, hence

\[ \tau = (\theta, \sigma^2, l_1, l_2, m). \]
Hierarchical Bayes’ model

\[ P(\theta, l_1, l_2, \sigma^2|Z, k_c, k_f^o) \propto P(z|\theta, l_1, l_2, \sigma^2)P(k_c|\theta, l_1, l_2, \sigma^2) \]
\[ P(k_f^o|\theta, l_1, l_2, \sigma^2)P(\theta)P(l_1, l_2)P(\sigma^2) \]

- \( P(z|\theta, l_1, l_2, \sigma^2) \): Likelihood
- \( P(k_c|\theta, l_1, l_2, \sigma^2) \): Upscale model linking fine and coarse scales
- \( P(k_f^o|\theta, l_1, l_2, \sigma^2) \): Observed fine scale model
- \( P(\theta)P(l_1, l_2)P(\sigma^2) \): Priors
The likelihood can be written as follows:

\[ Z = F[B(l_1, l_2, \sigma^2)\theta] + \epsilon_f \]
\[ = F_1(\theta, l_1, l_2, \sigma^2) + \epsilon_f \]

where, \( \epsilon_f \sim MVN(0, \sigma_f^2 I) \).
Likelihood calculations

\[ Z = F(\tau) + \epsilon \]

For Gaussian model the likelihood will be

\[
P(Z|\tau) = \frac{1}{\sqrt{2\pi\sigma_1}} \exp\left(-\frac{[Z - F(\tau)]^2}{2\sigma_1^2}\right)
\]

where \( \sigma_1^2 \) is the variance of \( \epsilon \).
Likelihood Calculations

- It is like a black-box likelihood which we can’t write analytically, although we do have a code $F$ that will compute it.
- We need to run $F$ to compute the likelihood which is expensive.
- Hence, no hope of having any conjugacy in the model, other than for the error variance in the likelihood.
- Need to be somewhat intelligent about the update steps during MCMC so that do not spend too much time computing likelihoods for poor candidates.
Upscale model

The Coarse-scale model can be written as follows.

\[
    k_c = L_1(k_f) + \epsilon_c \\
    = L_1(\theta, l_1, l_2, \sigma^2) + \epsilon_c
\]

where, \( \epsilon_c \sim MVN(0, \sigma^2_c I) \).

i.e \( k_c|\theta, l_1, l_2, \sigma^2, \sigma^2_c \sim MVN(L_1(\theta, l_1, l_2, \sigma^2), \sigma^2_c I) \).

- \( L_1 \) is the upsampling operator
- It could be as simple as average
- It could be more complex where you need to solve the original system on the coarse grid with boundary conditions
\[ \nabla \cdot (k_f(x) \Delta \phi) = 0 \]

\[ (k_c(x)e_j, e_l) = \frac{1}{|K|} \int_K (k_f(x) \Delta \phi_j(x), e_l) dx \]
Observed fine scale model

We assume the model \( k_f^O = k_p^O + \epsilon_k \)

where, \( \epsilon_k \sim MVN(0, \sigma_k^2) \).

\( k_p^O \) is the spatial field obtained from K-L the expansion at the observed well locations.

So here we assume, \( k_f^O|\theta, l_1, l_2, \sigma^2, \sigma_k^2 \sim MVN(k_p^O, \sigma_k^2) \),
\begin{align*}
l_1, l_2, \sigma^2 & \rightarrow \text{Covariance Matrix} \\
\lambda, \phi & \rightarrow \text{K.L.} \\
\theta & \rightarrow \text{Expansion} \\
K_f & \rightarrow \text{Upscaling} \\
F(.) & \rightarrow \text{Solve} \\
K_c & \rightarrow \text{Forward} \\
Z & \rightarrow \text{\textit{K}}^0_f \\
\end{align*}
Inverse problem

- We can show that the posterior measure is Lipschitz continuous with respect to the data in the total variation distance.
- It guaranties that this Bayesian inverse problem is well-posed.
- Say, $y$ is the total dataset, i.e., $y = \begin{pmatrix} z \\ k_c \\ k_f^0 \end{pmatrix}$
- $g(\tau, y)$ is the likelihood and $\pi_0(\tau)$ is the prior.
Theorem 0.1. $\forall r > 0, \exists C = C(r)$ such that the posterior measures $\pi_1$ and $\pi_2$ for two different data sets $y_1$ and $y_2$ with $\max (\|y_1\|_2, \|y_2\|_2) \leq r$, satisfy

$$\|\pi_1 - \pi_2\|_{TV} \leq C \|y_1 - y_2\|_2,$$
MCMC computation

- Metropolis-Hastings (M-H) Algorithm to generate the parameters.
- Reversible jump M-H algorithm when the dimension $m$ of the K-L expansion is treated as model unknown.
- Two step MCMC or Langevin can accelerate our computation.
• Two stage Metropolis
Numerical Results

- We consider the isotropic case $l_1 = l_2 = l$
- We consider a 50X50 fine scale permeability field on unit square
- We fix $l = 0.25$ and $\sigma^2 = 1$
- The observed coarse-scale permeability field is calculated in a 5X5 coarse grid
- The fine-scale permeability field is observed at 100 locations
10 percent fine-scale data observed and no coarse-scale data available
25 percent fine-scale data observed and no coarse-scale data available
25 percent fine-scale data observed and no coarse-scale data available
Numerical results with unknown K-L terms

- We generate 15 fine-scale permeability field with $l = 0.3$, $\sigma^2 = 0.2$ and the reference permeability field is taken to be the average of these 15 permeability field.
- We take the first 20 terms in the K-L expansion while generating the reference field.
- The mode of the posterior distribution of $m$ comes out to be 19.
- The posterior mean of fine-scale permeability field resembles very close to the reference permeability field.
- The posterior density of $l$ is bimodal but the highest peak is near $0.3$.
- The posterior density $\sigma^2$ are centered around $0.2$. 
Numerical Results using Reversible Jump MCMC

- **Reference fine-scale permeability**
- **Initial fine-scale permeability**
- **Observed coarse-scale permeability**
- **Posterior mean of the fine-scale permeability**
Numerical Results using Reversible Jump MCMC
Numerical Results using Reversible Jump MCMC
Discussion

- We have developed a Bayesian hierarchical model which is very flexible. We can use it for other fluid dynamics, weather forecasting problems.
- To use other dimension reduction techniques like predictive processes.
- In two stage MCMC: can we use approximate solvers (Polynomial Chaos,...) or emulators at the first stage.
- Bayes Theorem in Infinite dimension: Warwick, A. Stuart and his group.