Bayesian Inversion: Applications and Mathematical Foundations

Andrew M Stuart[1]

[1]Mathematics Institute and
Centre for Scientific Computing
University of Warwick

Woudshoten Lectures 2013
October 2^{nd} 2013

Funded by EPSRC, ERC and ONR

http://homepages.warwick.ac.uk/∼masdr/

- A.M. Stuart. *Inverse Problems: a Bayesian Perspective.* Acta Numerica **19**(2010).
  `∼masdr/BOOKCHAPTERS/stuart15c.pdf`
- M. Hairer, A.M. Stuart, J. Voss, *Signal Processing Problems on Function Space: Bayesian formulation . . . .* The Oxford Handbook of Nonlinear Filtering, editors D. Crisan and B. Rozovsky, OxfordUniversity Press, 2011.
  `∼masdr/BOOKCHAPTERS/stuart16c.pdf`
- S.L.Cotter, M. Dashti, A.M.Stuart. *Approximation of Bayesian Inverse Problems.* SINUM **48**(2010), 322–345.
  `∼masdr/JOURNALPUBS/stuart81.pdf`
- Y. Marzouk, D. Xiu. *A Stochastic Collocation Approach to Bayesian Inference in Inverse Problems.* Comm. Comp. Phys. **6**(2009), 836–847.

Outline

Outline

Motivation

- Aim: to solve $y = \mathcal{G}(u) + \mathrm{noise}$ for u given y.
- LSQ: minimize $\Phi(u; y) := \frac{1}{2}\|y - \mathcal{G}(u)\|^2$.
- Bayesian: u and $y|u$ random variables then find $u|y$.

Lagrangian Data Assimilation

- Find $u \in H = L^2_{\text{div}}(\Omega, \mathbb{R}^2)$ initial condition for Navier-Stokes:

$$\frac{dv}{dt} + \nu A v + B(v, v) = f, \; v(0) = u$$

- Given noisy Lagrangian data $y = \{z_j(t_k) + \eta_{j,k}\}$:

$$\frac{dz_j}{dt} = v(z_j, t), \; z_j(0) = z_{j,0}$$

- Abstractly: for $\mathcal{G} : X \subseteq H \mapsto Y = \mathbb{R}^{JK}$ find u given

$$y = \mathcal{G}(u) + \eta, \text{ noise.}$$

Eulerian Data Assimilation

- Find $u \in H = L^2_{\mathrm{div}}(\Omega, \mathbb{R}^2)$ initial condition for Navier-Stokes:

$$\frac{dv}{dt} + \nu A v + B(v, v) = f, \ v(0) = u$$

- Given noisy Eulerian data $y = \{v(x_j, t_k) + \eta_{j,k}\}$:

$$y_{j,k} = v(x_j, t_k) + \eta_{j,k}$$

- Abstractly: for $\mathcal{G} : X \subset H \mapsto Y = \mathbb{R}^{JK}$ find u given

$$y = \mathcal{G}(u) + \eta, \ \text{noise.}$$

Groundwater Flow Inversion

- Let $f \in H^{-1}(D)$. Find log permeability $u \in X = L^\infty(D)$:

$$\begin{aligned} -\nabla \cdot \left(e^u \nabla p\right) &= f, \quad x \in D \\ p &= 0, \quad x \in \partial D. \end{aligned}$$

- Given, for $j = 1, \dots, J$,

$$y_j = \ell_j(p) + \eta_j, \;\; \ell_j \in H^{-1}(D), \;\; \eta_j \, \text{noise}.$$

- Abstractly: for $\mathcal{G} : X \mapsto Y = \mathbb{R}^J$ find u given

$$y = \mathcal{G}(u) + \eta, \text{ noise}.$$

Outline

Motivation

- Aim: to find $u \in L^2([0,T];\mathbb{R}^p)$ solving

$$du = f(u)dt + \Sigma dB, \quad u(0) = u^-.$$

- Conditioned on:

$$u(T) = u^+.$$

- Or conditioned on $y \in L^2([0,T];\mathbb{R}^q)$ solving

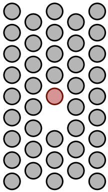
$$dy = h(u)dt + \Gamma dW, \quad y(0) = 0.$$

Molecular Dynamics

- **Find** $u \in H = L^2([0,1], \mathbb{R}^{Nd})$:

$$du = -\nabla V(u)dt + \sqrt{\frac{2}{\beta}}dB.$$

- **Conditioned on** red atom moving into vacancy:



Signal Processing

- **Find** $u \in H = L^2([0,1],\mathbb{R})$:

$$du = u - u^3 + dB, u(0) = -1$$

- **Conditioned on:** $y \in L^2([0,1],\mathbb{R})$ where

$$y(t) = \int_0^t u(s)ds + \Gamma W(t).$$

Outline

Common Probabilistic Framework

- There is a simple probability measure μ_0 on H.
- The complicated measure of interest is μ^y, also on H.
- μ^y is related to μ_0 by

$$\frac{d\mu^y}{d\mu_0}(u) = \frac{1}{Z}\exp\Big(-\Phi(u)\Big).$$

- Since $\mu^y(du) = Z^{-1}\exp\big(-\Phi(u)\big)\mu_0(du)$ we have

$$\mathbb{E}^{\mu^y} f(u) = \frac{1}{Z}\mathbb{E}^{\mu_0}\Big(\exp\big(-\Phi(u)\big)f(u)\Big).$$

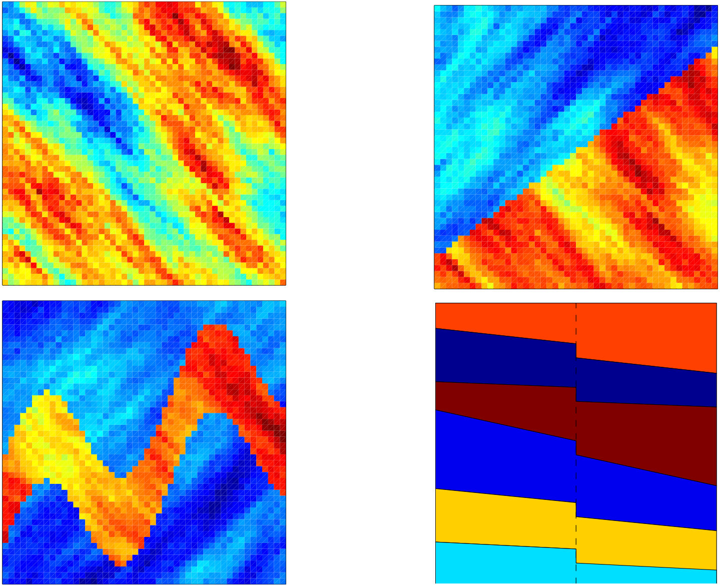
- Here Z is the normalization:

$$Z = \mathbb{E}^{\mu_0}\Big(\exp\big(-\Phi(u)\big)\Big).$$

PDE Inverse Problems

- Unknown $u \in H$.
- Data $y \in \mathbb{R}^J$.
- Prior $\mathbb{P}(u)$: $u \sim \mu_0(du)$.
- Likelihood $\mathbb{P}(y|u)$: $y|u \sim N(\mathcal{G}(u), \Gamma)$.
- Bayes' Theorem: $\mathbb{P}(u|y) \propto \mathbb{P}(y|u) \times \mathbb{P}(u)$.
- Posterior: $\mu^y(du) \propto \exp\Big(-\Phi(u; y)\Big)\mu_0(du)$.
- Potential: $\Phi(u; y) := \frac{1}{2}\Big\|\Gamma^{-\frac{1}{2}}\big(y - \mathcal{G}(u)\big)\Big\|^2$.

Priors for Permeability



Hellinger Distance

- For μ and ν which have density with respect to μ_0 we define the *Hellinger distance*

$$d_{Hell}(\mu,\nu) = \sqrt{\Big(\frac{1}{2}\int_X \Big[\Big(\frac{d\mu}{d\mu_0}\Big)^{\frac{1}{2}} - \Big(\frac{d\nu}{d\mu_0}\Big)^{\frac{1}{2}}\Big]^2 d\mu_0(u)\Big)}.$$

- Distance good because, for $f \in L^2_\mu(H;S), L^2_\nu(H;S)$:

$$\|\mathbb{E}^\mu f(u) - \mathbb{E}^\nu f(u)\|_S \le 2\Big(\mathbb{E}^\mu\|f(u)\|_S^2 + \mathbb{E}^\nu\|f(u)\|_S^2\Big)^{\frac{1}{2}} d_{Hell}(\mu,\nu).$$

The Well-Posedness Theorem

POTENTIAL CONDITIONS: $\Phi : X \times Y \to \mathbb{R}$ locally Lipschitz with appropriate growth conditions on Lipschitz constant.

Theorem

(S, Acta Numerica, 2010.) *Assume that*

- POTENTIAL CONDITIONS; $X \subseteq H$.
- $\mu_0(X) = 1$ *plus* INTEGRABILITY CONDITIONS.

Then μ^y well-defined and there is $C = C(|y_1|, |y_2|) > 0$:

$$d_{Hell}(\mu^{y_1}, \mu^{y_2}) \le C|y_1 - y_2|.$$

Implications

- **Mean**:

$$\left\|\mathbb{E}^{\mu^{y_1}} u - \mathbb{E}^{\mu^{y_2}} u\right\|_H \le C|y_1 - y_2|.$$

- **Covariance**

$$\left\|\mathbb{E}^{\mu^{y_1}} u \otimes u - \mathbb{E}^{\mu^{y_2}} u \otimes u\right\|_{H \to H} \le C|y_1 - y_2|.$$

Forward Error Gives Inverse Error

Write $\mu = \mu^y$; let μ^N denote approximation in finite dimensions.

Theorem

(Cotter, Dashti and S, SINUM, 2010).
Assume that $X \subseteq H$

- Φ *and* Φ^N *satisfy* POTENTIAL CONDITIONS, *uniformly in N;*
- *forward approximation error satisfies*

$$|\Phi(u) - \Phi^N(u)| \le M\big(\|u\|_X\big)\psi(N)$$

where $\psi(N) \to 0$ as $N \to \infty$;

- $\mu_0(X) = 1$ *plus* INTEGRABILITY CONDITIONS.

Then there is a constant C, independent of N, and such that the inverse approximation error is

$$d_{Hell}(\mu, \mu^N) \le C\psi(N).$$

Implications

- Mean:

$$\left\|\mathbb{E}^{\mu} u-\mathbb{E}^{\mu^{N}} u\right\|_{H} \leq C \psi(N).$$

- Covariance

$$\left\|\mathbb{E}^{\mu} u \otimes u-\mathbb{E}^{\mu^{N}} u \otimes u\right\|_{H \rightarrow H} \leq C \psi(N)$$

(Marzouk/Xiu used Kullback-Leibler divergence instead of Hellinger distance to prove a similar result.)

Outline

What We Have Shown

We have shown that:

- **Applications:** Many inverse problems in partial and stochastic differential equations can be formulated in the framework of Bayesian statistics on function space.
- **Common Structure:** These problems share a common mathematical structure leading to *well-posed* inverse problems for measures.
- **Approximation:** This well-posedness leads to a transfer of approximation properties from the forward problem to the inverse problem, in the Hellinger metric.
- **Algorithms:** Good algorithms follow from this infinite dimensional perspective on Bayes' Theorem – next lecture.

http://homepages.warwick.ac.uk/∼masdr/

- A.M. Stuart. *Inverse Problems: a Bayesian Perspective.* Acta Numerica **19**(2010).
  `∼masdr/BOOKCHAPTERS/stuart15c.pdf`
- M. Hairer, A.M. Stuart, J. Voss, *Signal Processing Problems on Function Space: Bayesian formulation . . . .* The Oxford Handbook of Nonlinear Filtering, editors D. Crisan and B. Rozovsky, OxfordUniversity Press, 2011.
  `∼masdr/BOOKCHAPTERS/stuart16c.pdf`
- S.L.Cotter, M. Dashti, A.M.Stuart. *Approximation of Bayesian Inverse Problems.* SINUM **48**(2010), 322–345.
  `∼masdr/JOURNALPUBS/stuart81.pdf`
- Y. Marzouk, D. Xiu. *A Stochastic Collocation Approach to Bayesian Inference in Inverse Problems.* Comm. Comp. Phys. **6**(2009), 836–847.