On the Bayesian formulation of fractional inverse problems and data-driven discretization of forward maps

Nicolás García Trillos
Brown University

Fractional PDEs: Theory, Algorithms and Applications
ICERM
June 18th 2018

Outline

1. Bayesian formulation of fractional inverse problems.
2. Data driven discretization of forward maps.

This presentation mostly based on:

1. *The Bayesian Formulation and Well-Posedness of Fractional Elliptic Inverse Problems* (2017 Inverse Problems) with D. Sanz-Alonso.
2. *Data driven discretizations of forward maps in Bayesian inverse problems* (In preparation) with D. Bigoni, Y. Marzouk and D. Sanz-Alonso.

Part 1: Bayesian formulation of fractional inverse problems.

Inverse problem: learn a permeability field from partial and noisy observations of pressure field.

PDE version: Learn diffusion coefficient and **order** of a (FPDE) based on partial and noisy observations of its solution.

Inverse problem: learn a permeability field from partial and noisy observations of pressure field.

PDE version: Learn diffusion coefficient and **order** of a (FPDE) based on partial and noisy observations of its solution.

$$u = (s, A) \longrightarrow \mathcal{F}(u) \longrightarrow \mathcal{O} \circ \mathcal{F}(u)$$

$$\mathcal{G} := \mathcal{O} \circ \mathcal{F}$$

$$y = \mathcal{G}(u) + Noise$$

Inverse problem: learn a permeability field from partial and noisy observations of pressure field.

PDE version: Learn diffusion coefficient and **order** of a (FPDE) based on partial and noisy observations of its solution.

$$u = (s, A) \longrightarrow \mathcal{F}(u) \longrightarrow \mathcal{O} \circ \mathcal{F}(u)$$

$$\mathcal{G} := \mathcal{O} \circ \mathcal{F}$$

$$\phi(y; \mathcal{G}(u))$$

- **Forward map:** $p = \mathcal{F}(u)$.

$$\begin{cases} L_A^s p &= f, \quad \text{in } D, \\ \partial_A p &= 0, \quad \text{on } \partial D, \end{cases} \tag{1}$$

  where $\partial_A p := A(x)\nabla p \cdot \nu$, and ν is the exterior unit normal to ∂D.

- **Observation map:** $\mathcal{O}(p) := (p(x_1), \ldots, p(x_n))$ for some $x_i \in D$.
- **Noise model:** $\phi(y, \mathcal{G}(u)) = \exp\left(-\frac{1}{2\gamma^2}\|y - \mathcal{G}(u)\|^2\right)$.

- **Forward map:** $p = \mathcal{F}(u)$.

$$\begin{cases} L_A^s p &= f, \quad \text{in } \ D, \\ \partial_A p &= 0, \quad \text{on } \ \partial D, \end{cases} \tag{2}$$

where $\partial_A p := A(x)\nabla p \cdot \nu$, and ν is the exterior unit normal to ∂D.
Here,

$$L_A^s p = \sum_{k=1}^{\infty} \lambda_{A,k}^s p_k \psi_{A,k}.$$

- **Observation map:** $\mathcal{O}(p) := (p(z_1), \ldots, p(z_m))$ for some $z_i \in D$.
- **Noise model:** $\phi(y, \mathcal{G}(u)) = \exp\left(-\frac{1}{2\gamma^2}\|y - \mathcal{G}(u)\|^2\right)$.

- A. M. Stuart. Inverse problems: a Bayesian perspective. (2010).
- J. Kaipio and E. Somersalo. Statistical and computational inverse problems (2006).

Bayesian formulation

- Prior: $u \sim \pi_u$
- Likelihood model: $\pi_{y|u}$
- Bayes rule (informally):

$$\nu^y(u) := \pi_{u|y} \propto \pi_{y|u} \cdot \pi_u$$

Posterior distribution.

ν^y is the fundamental object in Bayesian inference.

- Estimates:

$$\mathbb{E}_{u\sim\nu^y}\left(R(u)\right)$$

- Uncertainty quantification:

$$Var_{u\sim\nu^y}\left(R(u)\right)$$

Advantages of Bayesian formulation

Well defined mathematical framework:

- Stability (Well-posedness).
- Posterior consistency (contraction rates, scalings for parameters, etc).
- Consistency of numerical methods.

Prior

- π_u is a distribution on $(0,1) \times \mathcal{H}$.
- For example,

$$\pi_u = \pi_s \otimes \pi_A$$

$$A = e^v I_d, \qquad \text{where } v \sim N(0, K)$$

- Karhunen-Loeve expansion:

$$v = \sum_{i=1}^{\infty} \lambda_{K,i} \zeta_i \Psi_i, \quad \zeta_i \sim N(0,1).$$

Well-posedness of Bayesian formulation

Theorem (NGT and D. Sanz-Alonso 17')

Suppose that $\mathcal{G}$ is continuous in supp(π_u). Then, posterior distribution ν^y is absolutely continuous with respect to prior:

$$d\nu^y(u) \propto \phi(y; \mathcal{G}(u)) d\pi_u(u),$$

Recall: $\mathcal{G} : (s, A) \to \mathbb{R}^m$.

Well-posedness of Bayesian formulation

Theorem (NGT and D. Sanz-Alonso 17')

Suppose that $\mathcal{G} \in L^2_{\pi_u}$. Then the map

$$y \mapsto \nu^y$$

is Locally Lipschitz in the Hellinger distance. That is, For $|y_1|, |y_2| \leq r$ we have

$$d_{hell}(\nu^{y_1}, \nu^{y_2}) \leq C_r \|y_1 - y_2\|.$$

Well-posedness of Bayesian formulation

The analysis reduces to studying stability through regularity of FPDEs.

- L. A. Caffarelli and P. R. Stinga. Fractional elliptic equations, Caccioppoli estimates and regularity. (2016)

Well-posedness and posterior consistency

- M. Dashti and A. M. Stuart. Uncertainty quantification and weak approximation of an elliptic inverse problem. (2011).
- S. Agapiou, S. Larsson, and A.M. Stuart. Posterior contraction rates for the Bayesian approach to linear ill-posed inverse problems. (2013) .
- S. Volmer. Posterior consistency for Bayesian inverse problems through stability and regression results. (2013).

Numerical methods: MCMC

- Need a way to approximate expectations with respect to ν^y.
- Standard procedure: MCMC.
  Generate a path of a Markov chain with invariant distribution ν^y: $u_1, \dots, u_k, \dots$ and then use

$$\frac{1}{k}\sum_{i=1}^{k} R(u_i)$$

- However, careful with:
  1. Discretization of u.
  2. Discretization of forward map.

Idealized MCMC algorithm

For the sake of simplicity assume $A = e^v \cdot I_d$ and known $s \in (0,1)$.
Metropolis Hastings with pCN proposal:
Having defined v_k, v_{k+1} is generated according to:

1. Proposal: $\tilde{v} = \sqrt{1-\beta^2} v_k + \beta \xi$, where $\xi \sim \pi_v$.
2. Acceptance probability:

$$\alpha(\tilde{u}, u_k) := \min\left\{1, \frac{\phi(y;\mathcal{G}(\tilde{u})}{\phi(y;\mathcal{G}(u_k))}\right\}$$

3. $v_{k+1} := \begin{cases} \tilde{v} & \text{with prob } \alpha(\tilde{u}, u_k) \\ v_k & \text{with prob } 1-\alpha(\tilde{u}, u_k) \end{cases}$

Idealized MCMC algorithm

For the sake of simplicity assume $A = e^v \cdot I_d$ and known $s \in (0,1)$.

Metropolis Hastings with pCN proposal:

Having defined v_k, v_{k+1} is generated according to:

1. Proposal: $\tilde{v} = \sqrt{1-\beta^2} v_k + \beta\xi$, where $\xi \sim \pi_v$.
   Compare to: $\tilde{v} = v_k + \beta\xi$
   S. L. Cotter, G. O. Roberts, A. M. Stuart, and D. White. MCMC methods for functions: modifying old algorithms to make them faster. Statistical Science.
2. Acceptance probability:

$$\alpha(\tilde{u}, u_k) := \min\left\{1, \frac{\phi(y; \mathcal{G}(\tilde{u}))}{\phi(y; \mathcal{G}(u_k))}\right\}$$

Idealized MCMC algorithm

For the sake of simplicity assume $A = e^v \cdot I_d$ and known $s \in (0,1)$.

Metropolis Hastings with pCN proposal:

Having defined v_k, v_{k+1} is generated according to:

1. Proposal: $\tilde{v} = \sqrt{1-\beta^2} v_k + \beta\xi$, where $\xi \sim \pi_v$.
   Compare to: $\tilde{v} = v_k + \beta\xi$
   S. L. Cotter, G. O. Roberts, A. M. Stuart, and D. White. MCMC methods for functions: modifying old algorithms to make them faster. Statistical Science.
   Robustness to truncation:

$$\xi = \sum_{i=1}^{L} \lambda_{K,i} \zeta_i \Psi_i, \quad \zeta_i \sim N(0,1)$$

2. Acceptance probability:

$$\alpha(\tilde{u}, u_k) := \min\left\{1, \frac{\phi(y; \mathcal{G}(\tilde{u}))}{\phi(y; \mathcal{G}(u_k))}\right\}$$

Part 2: Data driven discretization of forward maps

$$u \longrightarrow \mathcal{F}(u) \longrightarrow \mathcal{O} \circ \mathcal{F}(u)$$

$$\mathcal{G} := \mathcal{O} \circ \mathcal{F}$$

$$\phi(y; \mathcal{G}(u))$$

MCMC: Metropolis with pCN proposal

To produce u_{k+1}:

1. Proposal: $\tilde{u} = \sqrt{1-\beta^2}u_k + \beta\xi$, $\xi \sim \pi_u$.
2. Compute acceptance probability:

$$\alpha(\tilde{u}, u_k) := \min\left\{1, \frac{\phi(y;\mathcal{G}(\tilde{u}))}{\phi(y;\mathcal{G}(u_k))}\right\}$$

3. $$u_{k+1} := \begin{cases} \tilde{u} & \text{with prob } \alpha(\tilde{u}, u_k) \\ u_k & \text{with prob } 1-\alpha(\tilde{u}, u_k) \end{cases}$$

MCMC:Metropolis with pCN proposal

To produce u_{k+1}:

1. Proposal: $\tilde{u} = \sqrt{1-\beta^2}u_k + \beta\xi$, $\xi \sim \pi_u$.
2. Compute acceptance probability:

$$\alpha(\tilde{u}, u_k) := \min\left\{1, \frac{\phi(y; \mathcal{G}_X(\tilde{u}))}{\phi(y; \mathcal{G}_X(u_k))}\right\}$$

3. $v_{k+1} := \begin{cases} \tilde{v} & \text{with prob } \alpha(\tilde{u}, u_k) \\ v_k & \text{with prob } 1-\alpha(\tilde{u}, u_k) \end{cases}$

MCMC:Metropolis with pCN proposal

To produce u_{k+1}:

1. Proposal: $\tilde{u} = \sqrt{1-\beta^2}u_k + \beta\xi$, $\xi \sim \pi_u$.
2. Compute acceptance probability:

$$\alpha(\tilde{u}, u_k) := \min\left\{1, \frac{\phi(y; \mathcal{G}_X(\tilde{u}))}{\phi(y; \mathcal{G}_X(u_k))}\right\}$$

3. $u_{k+1} := \begin{cases} \tilde{u} & \text{with prob } \alpha(\tilde{u}, u_k) \\ u_k & \text{with prob } 1-\alpha(\tilde{u}, u_k) \end{cases}$

How do we choose the discretization? How fine? Inhomogeneous in space?

$$(u, X) \longrightarrow \mathcal{F}_X(u) \longrightarrow \mathcal{O} \circ \mathcal{F}(u)$$

$$\mathcal{G}_X := \mathcal{O} \circ \mathcal{F}_X$$

$$\phi(y; \mathcal{G}_X(u))$$

$$(u, X) \longrightarrow \mathcal{F}_X(u) \longrightarrow \mathcal{O} \circ \mathcal{F}(u)$$

$$\mathcal{G}_X := \mathcal{O} \circ \mathcal{F}_X$$

$$\phi(y; \mathcal{G}_X(u))$$

More specifically: $X = (N, \{x_1, \dots, x_N\})$.

True and Surrogate Bayesian inverse problems

Prior: $u \sim \pi_u$.
Likelihood: $\phi(y; \mathcal{G}(u))$
Posterior:

$$d\nu^y(u) \propto \phi(y; \mathcal{G}(u)) d\pi_u(u)$$

Prior: $(u, X) \sim \pi_{u,X}$.
Likelihood: $\phi(y; \mathcal{G}_X(u))$
Posterior:

$$d\nu^y(u, X) \propto \phi(y; \mathcal{G}_X(u)) d\pi_{u,X}(u, X)$$

Prior for surrogate problem

For simplicity our prior takes the form:

$$\pi_{u,X} = \pi_{x_1,\dots,x_N|N} \cdot \pi_N \cdot \pi_u.$$

- π_u is as for the true problem.
- $\pi_{u,X}$ treats X and u independently.
- $\pi_{x_1,\dots,x_N|N} = dx_1 \dots dx_N$ on D^N.
- π_N takes into account cost of discretization of $\mathcal{F}$ using N elements:

$$\pi_N \propto \exp(-C(N))$$

Likelihood

$$\phi(y; \mathcal{G}_X(u))$$

Recall $\mathcal{G}_X = \mathcal{O} \circ \mathcal{F}_X$.

Likelihood

$$\phi(y; \mathcal{G}_X(u))$$

Recall $\mathcal{G}_X = \mathcal{O} \circ \mathcal{F}_X$.
However, we don't triangulate directly using X. First, we regularize the points $x_1, \dots, x_n$.

Likelihood

$$\phi(y; \mathcal{G}_X(u))$$

Recall $\mathcal{G}_X = \mathcal{O} \circ \mathcal{F}_X$.
However, we don't triangulate directly using X. We regularize the points $x_1, \dots, x_n$.
$X = (\{x_1, \dots, x_n\}, N)$ induces a density estimator ρ_X.

1. ρ_X is used to choose points in a master grid.

or

2. Using ρ_X we start a flow of the points $x_, \dots, x_n$ attempting to minimize an energy of the form:

$$E(x_1, \dots, x_n) \sim \sum_{i,j} \exp(-|x_i - x_j|^2/(h_N \cdot \rho_X(x_i))^2)$$

Sampling from $\nu^y(u, X)$

Metropolis within Gibbs:
Alternate:

1. Update the unknown; discretization is fixed. $u|X, y$
2. Change distribution of elements. $x_1, \dots, x_N|N, u, y$
3. Coarsen or refine discretization. $x_1, \dots, x_N, N|u, y$

Update u

Given (u_k, X_k) produce u_{k+1} by:

1. Proposal: $\tilde{u} = \sqrt{1-\beta^2}u_k + \beta\xi$, $\xi \sim \pi_u$.
2. Compute acceptance probability:

$$\alpha(\tilde{u}, u_k) := \min\left\{1, \frac{\phi(y; \mathcal{G}_{X_k}(\tilde{u}))}{\phi(y; \mathcal{G}_{X_k}(u_k))}\right\}$$

3. $u_{k+1} := \begin{cases} \tilde{u} & \text{with prob } \alpha(\tilde{u}, u_k) \\ u_k & \text{with prob } 1 - \alpha(\tilde{u}, u_k) \end{cases}$

Change distribution of elements

Given (u_k, X_k) produce $x_1^{k+1}, \ldots, x_N^{k+1}$ by:

1. Proposal: $\tilde{x}_i = x_i + F(x_i)$, for all $i = 1, \ldots, N$.

$$F := (\Psi_1, \Psi_2) \sim N(0, K_1 \otimes K_2)$$

2. Compute acceptance probability:

$$\alpha(\tilde{X}, X_k) := \min \left\{ 1, \frac{\phi(y; \mathcal{G}_{\tilde{X}}(u_k))}{\phi(y; \mathcal{G}_{X_k}(u_k))} \cdot \frac{p(X_k|\tilde{X})}{p(\tilde{X}|X_k)} \right\}$$

3. $X_{k+1} := \begin{cases} \tilde{X} & \text{with prob } \alpha(\tilde{X}, X_k) \\ X_k & \text{with prob } 1 - \alpha(\tilde{X}, X_k) \end{cases}$

Coarsen or refine discretization

Given (u_k, X_k) we produce N_{k+1} and $x_1^{k+1}, \dots, x_{N_{k+1}}^{k+1}$ by:

1. Proposal: First construct ρ_X and generate $\tilde{N} \sim p(\cdot|N_k)$.
   - If $\tilde{N} < N$ let $\tilde{x}_i = x_i$ for $i = 1, \dots, \tilde{N}$.
   - If $\tilde{N} \geq N$ let $\tilde{x}_i = x_i$ for $i = 1, \dots, N$ and generate $\tilde{x}_{N+j} \sim \rho_X$ for $j = 1, \dots, \tilde{N} - N$.
2. Compute acceptance probability:

$$\alpha(\tilde{X}, X_k) := \min \left\{ 1, \frac{\phi(y; \mathcal{G}_{\tilde{X}}(u_k))}{\phi(y; \mathcal{G}_{X_k}(u_k))} \cdot p(X, \tilde{X}) \cdot \frac{p(\tilde{N}|N_k)}{p(N_k|\tilde{N})} \cdot \frac{\exp(-C(\tilde{N}))}{\exp(-C(N))} \right\} \tag{3}$$

3. $X_{k+1} := \begin{cases} \tilde{X} & \text{with prob } \alpha(\tilde{X}, X_k) \\ X_k & \text{with prob } 1 - \alpha(\tilde{X}, X_k) \end{cases}$

Conclusions and future work

1. Bayesian formulation of fractional inverse problems.
2. Data driven discretization of forward maps.

Thank you for your attention!