Hierarchical Ensemble Kalman Methods with Sparsity-Promoting Gamma Hyperpriors

Hwanwoo Kim

University of Chicago

Joint work with D.Sanz-Alonso, A.Strang

2023 SIAM CSE

February 27, 2023

Problem Setup

Given data y of the form

$$y = \mathcal{G}(u) + \eta, \quad \eta \sim \mathcal{N}(0, \Gamma),$$

where $\mathcal{G} : \mathbb{R}^d \to \mathbb{R}^n$ and $\Gamma \in \mathbb{R}^{n\times n}$ are known,

- Point estimation of u under sparsity assumption?
- Uncertainty quantification of u?

Outline

1. Hierarchical Bayesian Model
2. Iterative Alternating Scheme (IAS)
3. ℓ_1-regularized Iterative Ensemble Kalman Filter (ℓ_1-IEKF)
4. Numerical Examples

Hierarchical Bayesian Model

- Data generating model (Likelihood) $\implies y|u \sim \mathcal{N}(\mathcal{G}(u), \Gamma)$
- Bayesian framework $\implies u|\theta \sim \mathcal{N}(0, D_\theta),\ D_\theta = \text{diag}(\theta)$
- Hierarchical setup $\implies \theta_i \sim \text{Gamma}(\alpha, \beta),\ 1 \leq i \leq d$
- Ultimate objective: Posterior distribution $p(u, \theta|y)$

Posterior Distribution

$$p(u,\theta|y) = \frac{p(y|u,\theta)p(u|\theta)p(\theta)}{p(y)} \propto \exp\bigl(-J(u,\theta)\bigr),$$

where

$$J(u,\theta) := \overbrace{\frac{1}{2}\|y-\mathcal{G}(u)\|_{\Gamma}^{2} + \rlap{$\underbrace{\phantom{\frac{1}{2}\|u\|_{D_\theta}^{2} + \sum_{i=1}^{d}\left[\frac{\theta_i}{\alpha_i} - \left(\beta-\frac{3}{2}\right)\log\frac{\theta_i}{\alpha_i}\right]}}_{(b)}$}\frac{1}{2}\|u\|_{D_\theta}^{2}}^{(a)} + \sum_{i=1}^{d}\left[\frac{\theta_i}{\alpha_i} - \left(\beta-\frac{3}{2}\right)\log\frac{\theta_i}{\alpha_i}\right].$$

Iterative Alternating Scheme (IAS)

When $\mathcal{G}(u)$ is linear, i.e., $\mathcal{G}(u) = Au$ for some matrix A,

1. **Initialize** θ^0, $k = 0$.
2. **Iterate until convergence:**
   (i) Main parameter update:

$$\begin{aligned} u^{k+1} &= \arg\min_u J(u, \theta^k) \\ &= \arg\min_u \frac{1}{2}\|y - Au\|_\Gamma^2 + \frac{1}{2}\|u\|_{D_\theta^k}^2 \\ &= (A^\top \Gamma^{-1} A + D_{\theta^k}^{-1})^{-1} A^\top \Gamma^{-1} y. \end{aligned}$$

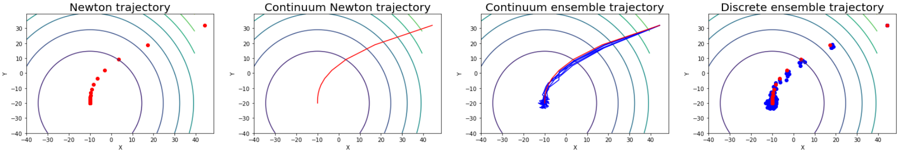
   (ii) Variance parameter(regularization) update:

$$\begin{aligned} \theta^{k+1} &= \arg\min_\theta J(u^{k+1}, \theta) \\ &= \alpha \left(\frac{\tilde{\beta}}{2} + \sqrt{\frac{\tilde{\beta}^2}{4} + \frac{(u_i^{k+1})^2}{2\alpha_i}} \right), \quad \tilde{\beta} = \beta - 3/2. \end{aligned}$$

Extensions of IAS

- More details in [Calvetti et al., 2019].
- Variational approximation of the posterior allows one to build approximate credible intervals. $\implies$ VIAS [Agrawal et al., 2022]
- Nonlinear extension $\implies$ ℓ_1-regularized IEKF [Kim et al., 2023]

Iterative Ensemble Kalman Filter (IEKF)



- IEKF is a sequential nonlinear optimization method.
- Introduce an initial ensemble (a set of particles).
- Particles are roughly updated according to a trajectory of Gauss-Newton iterates.
- After sufficient number of iterations, particles will be centered around the optimum.
- The sample mean of these particles will be our solution.

ℓ_1-regularized IEKF

- In the main parameter update step:

$$\arg\min_u \frac{1}{2}\|y - \mathcal{G}(u)\|_\Gamma^2 + \frac{1}{2}\|u\|_{D_\theta^k}^2,$$

  employ IEKF, an ensemble-based nonlinear optimization method.
- Combined with the variance parameter update, one can promote sparse structure in the solution: ℓ_1-regularized version of IEKF.
- Stronger $\ell_{p<1}$-regularization is possible with a generalized gamma hyperprior.
- Utilize particles to build approximate credible intervals.
- More details in [Kim et al., 2023].

Example 1: First order PDE inversion I

Consider the following partial differential equation

$$\begin{aligned} \partial_{x_1} v - \partial_{x_2} v - u(x_1)v = 0, \qquad & (x_1, x_2) \in (0,1) \times (0,1), \\ v(x_1, 0) = \phi(x_1), \qquad & x_1 \in [0,1]. \end{aligned}$$

If u is continuous and ϕ is continuously differentiable, then it admits the solution

$$v(x_1, x_2) = \phi(x_1 + x_2) \exp\left(\int_{x_1+x_2}^{x_1} u(z) dz \right).$$

With $\phi(x) = \cos(x)$, the data y is obtained according to

$$y(x_1, x_2) = v(x_1, x_2) + \epsilon, \quad \epsilon \sim N(0, 0.1^2).$$

Example 1: first order PDE inversion II

Our goal is to recover the function u given the data y. We further assume that u admits a representation

$$u(x) = \sum_{j=1}^{30} u_j \sin(j\pi x) + \sum_{j=1}^{30} \tilde{u}_j \cos(j\pi x), \ x \in [0,1],$$

with only three components of $\{u_j\}_{j=1}^{30}$ and $\{\tilde{u}_j\}_{j=1}^{30}$ are nonzero.

Example 1: first order PDE inversion III

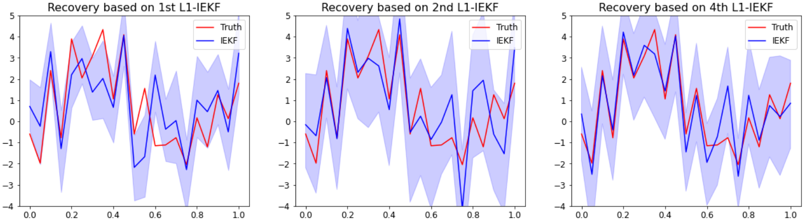


Figure: Red: target function to recover. Blue: ℓ_1-recovery. Left column: vanilla version. Middle column: ℓ_1-recovery after one outer iteration. Right column: ℓ_1-recovery after three outer iterations. Shaded: 2.5/97.5 percentile of the recovery.

Example 2: Subsurface flow inversion I

Consider the elliptic PDE, given by

$$-\mathsf{div}\big(e^{u(x)}\nabla v(x)\big) = f(x),\ x = (x_1, x_2) \in [0,1] \times [0,1]$$

with boundary conditions

$$v(x_1, 0) = 100,\ \frac{\partial v}{\partial x_1}(1, x_2) = 0,\ -e^{u(x)}\frac{\partial v}{\partial x_1}(0, x_2) = 500,\ \frac{\partial v}{\partial x_2}(x_1, 1) = 0,$$

and source term

$$f(x) = f(x_1, x_2) = \begin{cases} 0 & 0 \le x_2 \le \frac{4}{6}, \\ 137 & \frac{4}{6} < x_2 \le \frac{5}{6}, \\ 274 & \frac{5}{6} < x_2 \le 1. \end{cases}$$

The domain is discretized in a uniform 15×15 grid in $[0,1] \times [0,1]$.

Example 2: Subsurface flow inversion II

- Assume the log diffusion coefficient to be expressed as

$$u(x_1, x_2) = \sum_{i=0}^{19} \sum_{j=0}^{19} u_{ij} \phi_{ij}(x_1, x_2),$$

  where $\phi_{ij}(x_1, x_2) = \cos(i\pi x_1) \cos(j\pi x_2)$.
- Only six components of $\{u_{i,j}\}_{i,j=0}^{19}$ were chosen to be nonzero.
- We aim to recover $\{u_{i,j}\}_{i,j=0}^{19} \in \mathbb{R}^{400}$ from the data

$$y(x_1, x_2) = \mathcal{G}\big(u(x_1, x_2)\big) + \eta = v(x_1, x_2) + \eta,$$

  evaluated along the grid with $\eta \sim N(0, 0.1^2)$.

Example 2: Subsurface flow inversion III

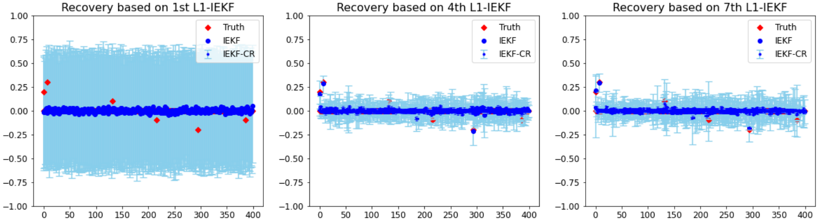


Figure: Parameter recovery for 2D-elliptic inverse problem based on ℓ_1-IEKF. Red: Truth. Blue: ℓ_1-IEKF estimate. Left column: vanilla IEKF. Middle column: ℓ_1-IEKF after three outer iterations. Right column: ℓ_1-IEKF after six outer iterations. Shaded: elementwise 2.5/97.5 percentile for parameter estimate.

Summary

- Hierarchical Bayesian model with gamma hyperprior to induce sparse structure in parameter of interest.
- Under the linear model setting, one can use coordinate descent or a variational inference to approximate posterior distribution of the target parameter.
- Under the nonlinear model setting, one can use Iterative Ensemble Kalman Filter(IEKF) to solve for parameters and build approximate credible intervals using particles.
- Hierarchical Bayesian framework provides flexible regularizations tools in IEKF.
- Codes available at: https://github.com/hwkim12

References

Agrawal, S., Kim, H., Sanz-Alonso, D., and Strang, A. (2022).
A variational inference approach to inverse problems with gamma hyperpriors.
SIAM/ASA Journal of Uncertainty Quantification, 10(4):1533–1559.

Calvetti, D., Somersalo, E., and Strang, A. (2019).
Hierachical bayesian models and sparsity: ℓ_2-magic.
Inverse Problems, 35(3):035003.

Kim, H., Sanz-Alonso, D., and Strang, A. (2023).
Hierarchical ensemble kalman methods with sparsity-promoting generalized gamma hyperpriors.
arXiv preprint arXiv:2205.09322, to appear in Foundations of Data Science.