Sparse Inverse Problems with Gamma Hyperpriors

Hwanwoo Kim

University of Chicago

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

2022 SIAM Mathematics of Data Science

September 29, 2022

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. Variational Iterative Alternating Scheme (VIAS)
4. ℓ_1-regularized Iterative Ensemble Kalman Filter (ℓ_1-IEKF)
5. Numerical Examples

Hierarchical Bayesian Model

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

Posterior Distribution

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

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:**
   1. Main parameter update:

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

   2. Variance parameter(regularization) update:

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

More details in [Calvetti et al., 2019].

Mean-Field Variation Inference

- Can we approximate $p(u, \theta|y)$ with other probability distributions?
  $\implies$ Mean-Field Variation Inference [Blei et al., 2017]
- Mean-Field Family:

$$\mathcal{D} := \Big\{ q(z) : q(u,\theta) = q(u)q(\theta), \quad q(\theta) = \prod_{i=1}^{d} q(\theta_i) \Big\}.$$

- Find a distribution $q^* \in \mathcal{D}$ closest to $p(u, \theta|y)$.
  $\implies$ Kullback-Leibler Divergence

$$\begin{aligned} \mathsf{KL}(q(\cdot)||p(\cdot|y)) &:= \int q(z) \log \frac{q(z)}{p(z|y)} dz \\ &= \log p(y) - \underbrace{\int q(z) \log \frac{p(z,y)}{q(z)} dz}_{=:\mathsf{ELBO}(q)} \end{aligned}$$

Variational Iterative Alternating Scheme (VIAS)

1. **Initialize** $q^0(u)$, $k = 0$.
2. **Iterate:**
   1. Update
   $$q^{k+1}(\theta) = \arg\max_{q(\theta)} \mathsf{ELBO}\big(q^k(u)q(\theta)\big)$$
   2. Update
   $$q^{k+1}(u) = \arg\max_{q(u)} \mathsf{ELBO}\big(q(u)q^{k+1}(\theta)\big)$$
   3. $k \to k+1$

VIAS: $q(\theta)$ update

With $q^{k+1}(\theta) = \prod_{i=1}^{d} q^{k+1}(\theta_i)$, one can derive

$$q^{k+1}(\theta_i) = \mathsf{GIG}(b, r_i^k, s),$$
$$b = 2\beta,\ s = \alpha - 0.5,\ r_i^k = (m_i^k)^2 + C_{ii}^k,$$

where

$$m_i^k = \mathbb{E}_{q^k(u)}[u_i], \quad C_{ii}^k = \mathsf{Var}_{q^k(u)}[u_i].$$

VIAS: $q(u)$ update

One can obtain

$$q^{k+1}(u) = \mathcal{N}(m^{k+1}, C^{k+1})$$

with

$$\begin{aligned} m^{k+1} &= (A^\top \Gamma^{-1} A + L)^{-1} A^\top \Gamma^{-1} y, \\ C^{k+1} &= (A^\top \Gamma^{-1} A + L)^{-1}, \end{aligned}$$

where $L = \text{diag}\left(\mathbb{E}_{q^{k+1}(\theta)}\left[\frac{1}{\theta}\right]\right)$ whose ith diagonal component is given by

$$L_{ii} = \frac{K_{s-1}\left(\sqrt{r_i^{k+1} b}\right)}{K_s\left(\sqrt{r_i^{k+1} b}\right)} \cdot \sqrt{\frac{b}{r_i^{k+1}}}$$

VIAS: Convergence and Initialization

- ELBO(q) $=$ ELBO(m, C, r).
- ELBO(m, C, r) is not convex.
- In general, VIAS can only guarantee convergence to a local maximum.
- Due to multiple local maxima, choosing C^0 with small condition number is recommended.
- Hyperparameters α, β can be chosen to maximize ELBO
  : Only need ballpark values.
- More details in [Agrawal et al., 2022].

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., 2022].

Learning parameters for Lorenz 63 model I

Consider the Lorenz-63 system:

$$\begin{aligned}
\frac{dx}{dt} &= \sigma(y - x), \\
\frac{dy}{dt} &= x(\rho - z) - y, \\
\frac{dz}{dt} &= xy - \zeta z,
\end{aligned}$$

with the classical parameter values $\sigma = 10, \rho = 28$, $\zeta = 8/3$.

Learning parameters for Lorenz 63 model II

$$A = \begin{bmatrix} | & | & | & | & | & | & | & | & | & & | & \\ \mathbf{x} & \mathbf{y} & \mathbf{z} & \mathbf{x}^2 & \mathbf{y}^2 & \mathbf{z}^2 & \mathbf{xy} & \mathbf{xz} & \mathbf{yz} & \dots & \mathbf{x}^5 & \dots \\ | & | & | & | & | & | & | & | & | & & | & \end{bmatrix} \in \mathbb{R}^{2000 \times 55}.$$

Learning parameters for Lorenz 63 model III

Obtain synthetic data $\dot{\mathbf{x}}$, $\dot{\mathbf{y}}$, $\dot{\mathbf{z}}$ from

$$\begin{aligned}
\dot{\mathbf{x}} &= A\Phi_1 + \eta_1, \quad \Phi_1 = [-10, 10, 0, 0, 0, 0, 0, \ldots, 0]^\top, \\
\dot{\mathbf{y}} &= A\Phi_2 + \eta_2, \quad \Phi_2 = [28, -1, 0, 0, -1, 0, 0, \ldots, 0]^\top, \\
\dot{\mathbf{z}} &= A\Phi_3 + \eta_3, \quad \Phi_3 = [0, 0, -8/3, 0, 0, 0, 1, \ldots, 0]^\top,
\end{aligned}$$

where $\eta_i \sim \mathcal{N}(0, 0.3I_{2000})$ are independent.

- Goal: recover Φ_1, Φ_2, Φ_3 based on A, $\dot{\mathbf{x}}$, $\dot{\mathbf{y}}$, $\dot{\mathbf{z}}$.
- More detailed explanation on simulation setup: [Brunton et al., 2016]

Parameter Estimation: VIAS

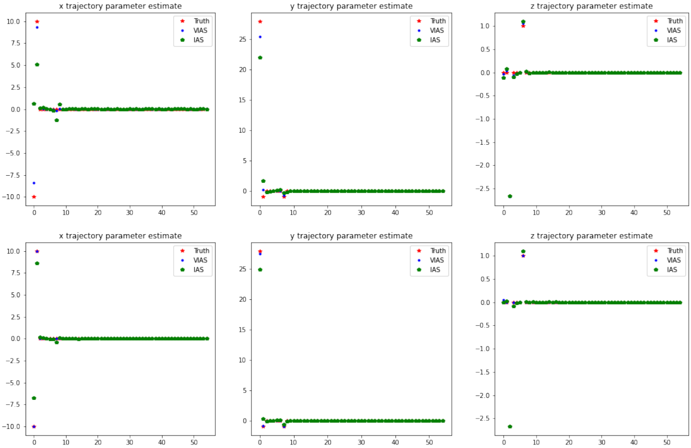

Figure: Recovery of dictionary coefficients for x-trajectory (first column), y-trajectory (second column), and z-trajectory (third column) using IAS and VIAS. Top: two iterations. Bottom: five iterations.

Uncertainty Quantification of Trajectory: VIAS

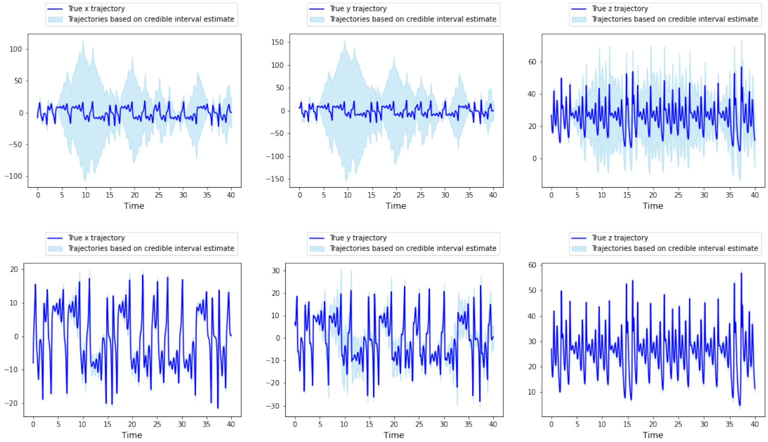

Figure: True Lorenz-63 trajectory and VIAS estimation. Top: two VIAS iterations. Bottom: five VIAS iterations. Blue line is the true dynamics. Shaded regions are constructed from 2.5 and 97.5 credible levels of coefficients.

Nonlinear Example: 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

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

Nonlinear Example: 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.

Nonlinear Example: first order PDE inversion III

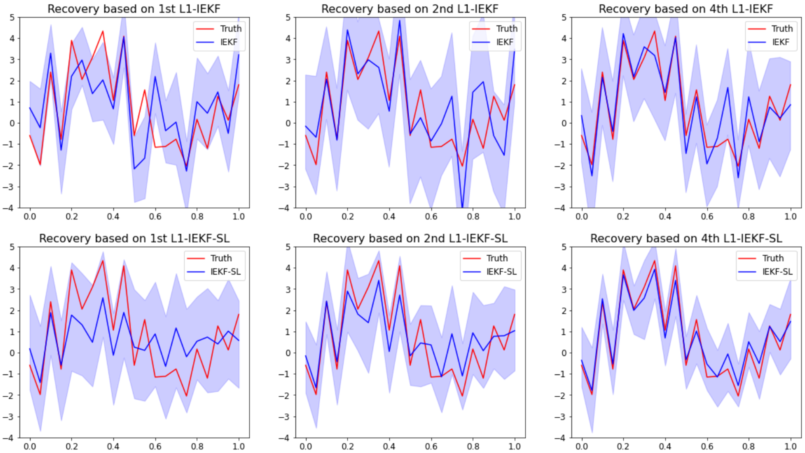

Figure: Red: target function to recover. Blue: ℓ_1-IEKF/IEKF-SL recovery. Left column: vanilla IEKF/IEKF-SL. Middle column: ℓ_1-IEKF/IEKF-SL after one outer iteration. Right column: ℓ_1-IEKF/IEKF-SL after three outer iterations. Shaded: 2.5/97.5 percentile of the recovery.

Summary

- Hierarchical Bayesian model with gamma hyperprior to induce sparse structure in parameter of interest.
- Under the linear model setting, we employed a variational inference technique to approximate posterior distribution of the target parameter.
- An coordinate ascent variational inference led to simple parameter update rules.
- Under the nonlinear model setting, we used iterative ensemble Kalman filter to solve for main parameter and build approximate credible intervals using particles.
- Reproducible 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.
arXiv preprint arXiv:2111.13329,
to appear in SIAM/ASA Journal of Uncertainty Quantification.

Blei, D. M., Kucukelbir, A., and McAuliffe, J. D. (2017).
Variational inference: A review for statisticians.
Journal of the American statistical Association, 112(518):859–877.

Brunton, S. L., Proctor, J. L., and Kutz, J. N. (2016).
Discovering governing equations from data by sparse identification of nonlinear dynamical systems.
Proceedings of the national academy of sciences, 113(15):3932–3937.

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. (2022).
Hierarchical ensemble kalman methods with sparsity-promoting generalized gamma hyperpriors.
arXiv preprint arXiv:2205.09322.