Bayesian Inference for Latent Gaussian Models Governed by PDEs

Sonia Reilly, Georg Stadler

Courant Institute, New York University

Goal

To characterize the posterior of a hierarchical linear PDE-governed Bayesian inverse problem modeled as a latent Gaussian model by computing the marginal distribution of the hyperparameters.

Latent Gaussian Models (LGMs)

Hierarchical model with Gaussian prior:

$$\begin{aligned} \boldsymbol{\theta} &\sim \pi_{\text{hyp}}(\boldsymbol{\theta}) && \text{low-dim hyperparameter} \\ \boldsymbol{m}|\boldsymbol{\theta} &\sim \mathcal{N}(\boldsymbol{\mu}_{\text{pr}}(\boldsymbol{\theta}), \boldsymbol{Q}_{\text{pr}}^{-1}(\boldsymbol{\theta})) && \text{high-dim latent variable} \\ \boldsymbol{y}|\boldsymbol{m}, \boldsymbol{\theta} &\sim \pi_{\text{like}}(\boldsymbol{y}|\boldsymbol{m}, \boldsymbol{\theta}) && \text{observed data} \end{aligned}$$

Linear Gaussian Bayesian inverse problem as LGM:

$$\boldsymbol{y} = \boldsymbol{A}\boldsymbol{m} + \boldsymbol{\varepsilon} \quad \text{with} \quad \boldsymbol{\varepsilon} \sim \mathcal{N}(0, \boldsymbol{Q}_{\varepsilon}(\boldsymbol{\theta})),$$

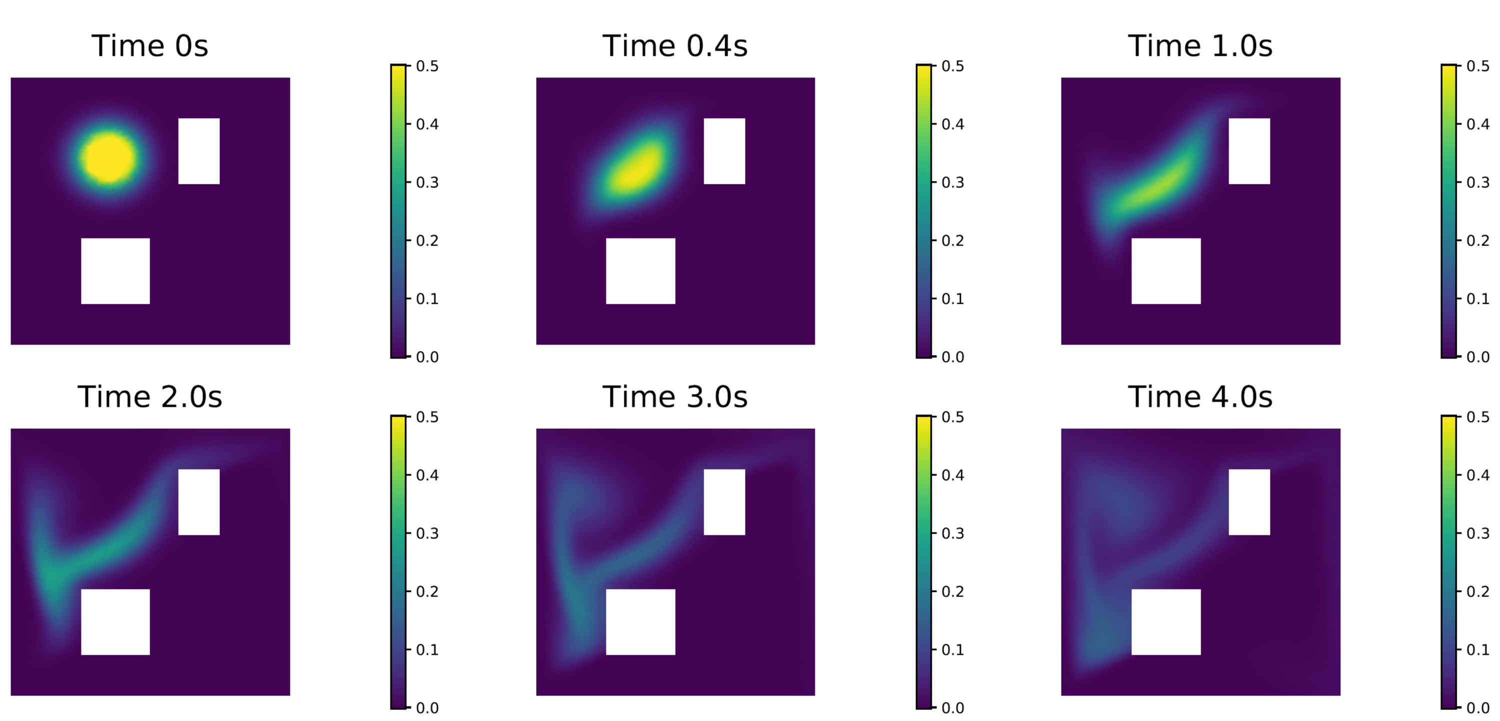
with $\boldsymbol{A}$ a discretization of a linear PDE. Want to characterize $\pi(\boldsymbol{m}|\boldsymbol{y})$.

Example: Hierarchical Initial Condition Inference

Infer distribution over initial condition of advection diffusion equation

$$\begin{cases} u_t - \kappa\Delta u + v \cdot \nabla u = 0 & \text{in} \quad \Omega \times [0, T] \\ u(0, x) = m(x) & \text{in} \quad \Omega \\ \kappa \nabla u \cdot n = 0 & \text{on} \quad \partial\Omega, \end{cases}$$

from solution evaluated at points x_i and times t_i, given Gaussian prior with unknown hyperparameters.



- Data: $y_i = u(x_i, t_i) + \varepsilon_i$
- Latent variable: discretized initial condition $\boldsymbol{m}$
- Forward map $\boldsymbol{A}$: map from discretized $\boldsymbol{m}$ to $\boldsymbol{y}$
- Prior: Matérn Gaussian random field with precision $(\delta I - \gamma\Delta)^2$
- Likelihood $\pi_{\text{like}}(\boldsymbol{y}|\boldsymbol{m}, \boldsymbol{\theta})$: $\mathcal{N}(\boldsymbol{A}\boldsymbol{m}, \boldsymbol{Q}_{\varepsilon}(\boldsymbol{\theta}))$
- Hyperparameter $\boldsymbol{\theta}$: prior parameters δ, γ
- Hyperprior $\pi_{\text{hyp}}(\boldsymbol{\theta})$: gamma distributions, deliberately uninformative

Sampling & Integration by Marginalization

Sampling: sample $\boldsymbol{m}^* \sim \pi(\boldsymbol{m}|\boldsymbol{y})$ by

1. sampling $\boldsymbol{\theta}^* \sim \pi(\boldsymbol{\theta}|\boldsymbol{y})$ (low-dimensional, so can use MCMC)
2. sampling $\boldsymbol{m}^* \sim \pi(\boldsymbol{m}|\boldsymbol{\theta}^*, \boldsymbol{y})$ (Gaussian, since it is the posterior of a linear Gaussian Bayesian inverse problem)

Integration:

$$\begin{aligned} \int f(\boldsymbol{m})\pi(\boldsymbol{m}|\boldsymbol{y})\, d\boldsymbol{m} &= \int\int f(\boldsymbol{m})\pi(\boldsymbol{m}, \boldsymbol{\theta}|\boldsymbol{y})\, d\boldsymbol{\theta} d\boldsymbol{m} \\ &= \int \left(\int f(\boldsymbol{m})\pi(\boldsymbol{m}|\boldsymbol{\theta}, \boldsymbol{y})\, d\boldsymbol{m} \right) \pi(\boldsymbol{\theta}|\boldsymbol{y})\, d\boldsymbol{\theta} \end{aligned}$$

Can integrate by sampling as above, or for the outer integral, by quadrature. For quadrature, find max of $\pi(\boldsymbol{\theta}|\boldsymbol{y})$ and set of quadrature points around the max.

Fast Computation of Marginal: Low Rank Approx

Both require many evaluations of $\pi(\boldsymbol{\theta}|\boldsymbol{y})$ with different $\boldsymbol{\theta}$:

$$\begin{aligned} \pi(\boldsymbol{\theta}|\boldsymbol{y}) &\propto \frac{\pi(\boldsymbol{m}, \boldsymbol{\theta}, \boldsymbol{y})}{\pi(\boldsymbol{m}|\boldsymbol{\theta}, \boldsymbol{y})} = \frac{\pi_{\text{like}}(\boldsymbol{y}|\boldsymbol{m}, \boldsymbol{\theta})\pi_{\text{pr}}(\boldsymbol{m}|\boldsymbol{\theta})\pi_{\text{hyp}}(\boldsymbol{\theta})}{\pi(\boldsymbol{m}|\boldsymbol{\theta}, \boldsymbol{y})} \\ &\propto \left(\frac{|\boldsymbol{Q}_{\text{pr}}||\boldsymbol{Q}_{\varepsilon}|}{|\boldsymbol{Q}_{\text{post}}|} \right)^{1/2} \exp\left(-\frac{1}{2} \left[||\boldsymbol{y}||_{Q_{\varepsilon}} + ||\boldsymbol{\mu}_{\text{pr}}||_{Q_{\text{pr}}} - ||\boldsymbol{\mu}_{\text{post}}||_{Q_{\text{post}}} \right] \right) \pi_{\text{hyp}}(\boldsymbol{\theta}) \end{aligned}$$

where

$$\boldsymbol{Q}_{\text{post}} = \boldsymbol{Q}_{\text{pr}} + \underbrace{\boldsymbol{A}^T \boldsymbol{Q}_{\varepsilon} \boldsymbol{A}}_{\text{update}}$$

and $\boldsymbol{\mu}_{\text{post}}$ are the precision and mean of the Gaussian posterior $\pi(\boldsymbol{m}|\boldsymbol{\theta}, \boldsymbol{y})$.

Each evaluation with new $\boldsymbol{\theta}$ requires solving a Bayesian inverse problem. Computational cost is dominated by forward and adjoint PDE solve in $\boldsymbol{A}^T \boldsymbol{Q}_{\varepsilon} \boldsymbol{A}$. Instead, compute rank-$r$ approximation using randomized SVD:

$$\boldsymbol{A}^T \boldsymbol{Q}_{\varepsilon} \boldsymbol{A} \approx \boldsymbol{V}_r \Lambda_r \boldsymbol{V}_r^T.$$

Costs $O(r)$ PDE solves up front, but each subsequent $\pi(\boldsymbol{\theta}|\boldsymbol{y})$ evaluation costs no PDE solves.

Decreasing the Rank: Prior Preconditioning

Problem: Large r may be needed, and unclear how to choose r.
Solution: Approximate *prior-preconditioned* update,

$$\boldsymbol{Q}_{\text{pr}}^{-1/2}(\boldsymbol{\theta}) \boldsymbol{A}^T \boldsymbol{Q}_{\varepsilon} \boldsymbol{A} \boldsymbol{Q}_{\text{pr}}^{-1/2}(\boldsymbol{\theta}) \approx \boldsymbol{V}_r \Lambda_r \boldsymbol{V}_r^T,$$

since $\boldsymbol{Q}_{\text{pr}}$ is a smoothing operator that lowers r and implies a natural bound on truncation error.

But $\boldsymbol{Q}_{\text{pr}}$ depends on $\boldsymbol{\theta}$!

Weakest Prior Preconditioning

Instead,

$$\boldsymbol{Q}_{\text{pr}}^{-1/2}(\boldsymbol{\theta}_0) \boldsymbol{A}^T \boldsymbol{Q}_{\varepsilon} \boldsymbol{A} \boldsymbol{Q}_{\text{pr}}^{-1/2}(\boldsymbol{\theta}_0) \approx \boldsymbol{V}_r \Lambda_r \boldsymbol{V}_r^T$$

for some fixed $\boldsymbol{\theta}_0$ and convert for each $\boldsymbol{\theta}$:

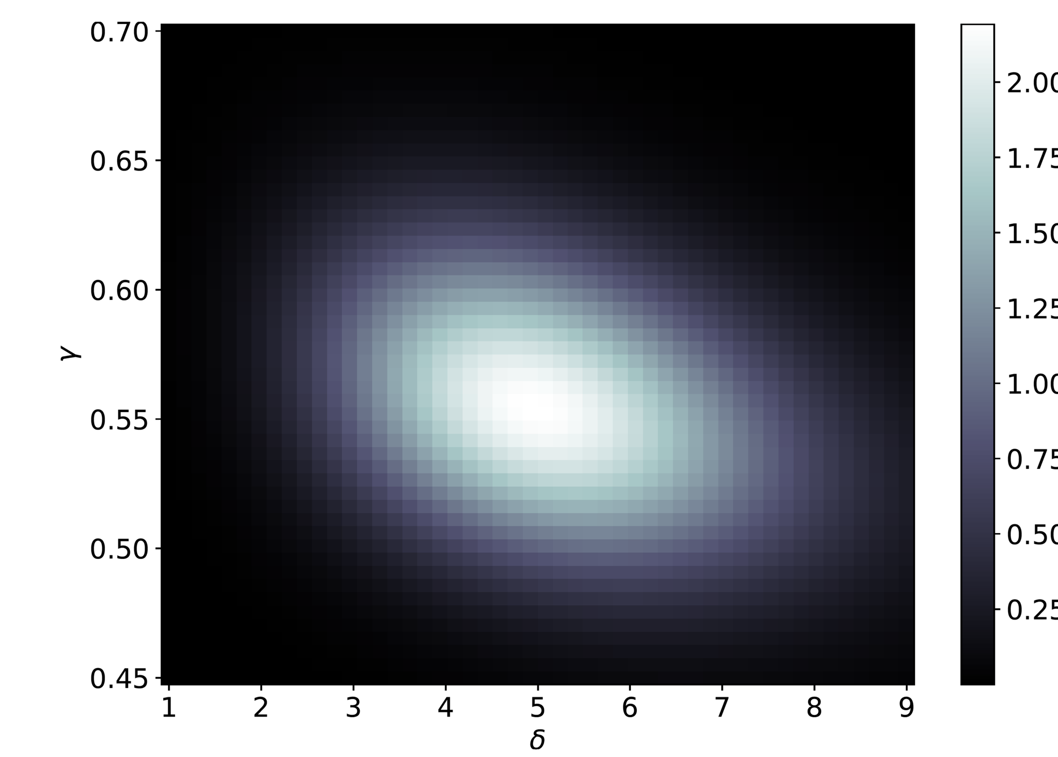
$$\boldsymbol{Q}_{\text{pr}}^{-1/2}(\boldsymbol{\theta}) \boldsymbol{A}^T \boldsymbol{Q}_{\varepsilon} \boldsymbol{A} \boldsymbol{Q}_{\text{pr}}^{-1/2}(\boldsymbol{\theta}) \approx \boldsymbol{Q}_{\text{pr}}^{-1/2}(\boldsymbol{\theta}) \boldsymbol{Q}_{\text{pr}}^{1/2}(\boldsymbol{\theta}_0) \boldsymbol{V}_r \Lambda_r \boldsymbol{V}_r^T \boldsymbol{Q}_{\text{pr}}^{1/2}(\boldsymbol{\theta}_0) \boldsymbol{Q}_{\text{pr}}^{-1/2}(\boldsymbol{\theta}).$$

Choose $\boldsymbol{\theta}_0$ to be the parameters of the least smoothing, "weakest" prior to avoid amplifying truncation error.

Initial Condition Inference $\pi(\boldsymbol{\theta}|\boldsymbol{y})$

Hyperparameter marginal $\pi(\gamma, \delta|\boldsymbol{y}) = \pi(\boldsymbol{\theta}|\boldsymbol{y})$.

- High probability γ, δ represent approximate smoothness and variance of true initial condition
- Unimodal, easy to sample and to integrate with quadrature



Next Steps

- Use various discretizations of PDE to speed up MCMC
- Find a way to adapt low rank approximation when $\boldsymbol{A}$ depends on $\boldsymbol{\theta}$
- Extend to nonlinear PDEs using ideas from Integrated Nested Laplace Approximation

References

[1] Bui-Thanh, T., Ghattas, O., Martin, J. and Stadler, G., 2013. A computational framework for infinite-dimensional Bayesian inverse problems Part I: The linearized case, with application to global seismic inversion. SIAM Journal on Scientific Computing, 35(6), pp.A2494-A2523.

[2] Norton, R.A., Christen, J.A. and Fox, C., 2018. Sampling hyperparameters in hierarchical models: improving on Gibbs for high-dimensional latent fields and large datasets. Communications in Statistics-Simulation and Computation, 47(9), pp.2639-2655.

Acknowledgements

This material is based upon work supported by the U.S. Department of Energy, Office of Science, Office of Advanced Scientific Computing Research, Department of Energy Computational Science Graduate Fellowship under Award Number DE-SC0022158.