

# Variational Bayesian Approximation of Inverse Problems using Sparse Precision Matrices

## Variational Bayes as an alternative to MCMC?

**Jan Povala<sup>1</sup>** Ieva Kazlauskaitė<sup>1</sup> Eky Febrianto Fehmi Cirak  
Mark Girolami  
Imperial College London / The Alan Turing Institute

June 8, 2022

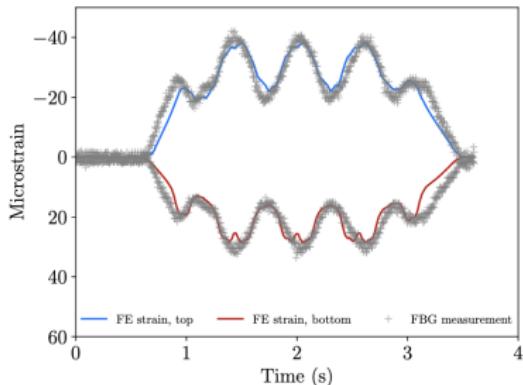
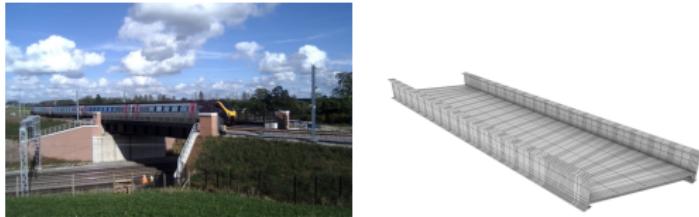
---

<sup>1</sup>equal contribution

# Motivation

Febrianto, Butler, Girolami & Cirak (2021)

A railway bridge in Staffordshire which has been instrumented with fibre optic sensors (top left), its digital twin (top right), sample sensor measurements (bottom):



## Using the Model and the Data

1. Using the model to answer ‘what if’ scenarios: How would the system behave under different conditions?
2. Using the measurements of the state of the system to infer the system’s properties.

The latter problem is termed the **inverse problem**.

## The Rest of the Talk

1. Bayesian Inverse problem formulation.
2. Variational Bayes as an alternative to Markov Chain Monte Carlo methods.
3. Leveraging problem structure for efficient inference.
4. Results on an elliptic PDE.
5. Conclusion.

## General Model (Stuart 2010)

- ▶ For a given parameter  $\kappa$ , the solution of a PDE model is given by a possibly non-linear solution operator  $\mathcal{A}$ :

$$u = \mathcal{A}(\kappa). \quad (1)$$

- ▶ We observe the solution at a finite number of points:

$$\mathbf{y} = \mathcal{P}(\mathcal{A}(\kappa)) + \boldsymbol{\eta}, \quad (2)$$

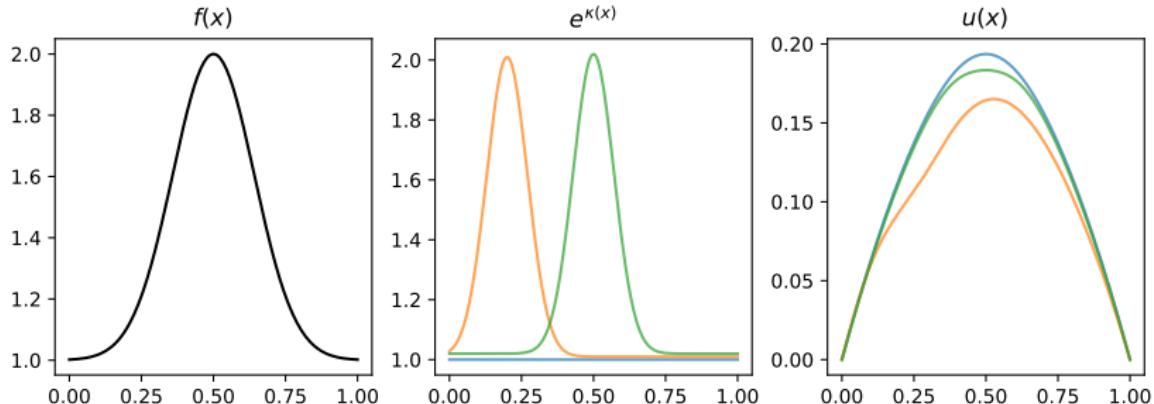
where  $\mathcal{P}$  is the projection operator and  $\boldsymbol{\eta}$  is non-correlated noise.

- ▶ Objective: want to find  $\kappa$ , the parameter to a model, given  $\mathbf{y}$ , a noisy observation of the solution of a PDE.

## Example – Poisson Equation

Given forcing  $f$ , log-diffusivity  $\kappa$ , and appropriate boundary conditions, we have:

$$-\nabla \cdot (\exp(\kappa(x)) \nabla u(x)) = f(x), \quad (3)$$



## Inverse Problem: Finding $\kappa$

- ▶ General approach: finding  $\kappa$  such that the data misfit,  $\|\mathbf{y} - \mathcal{P}(\mathcal{A}(\kappa))\|_2$ , is minimised.
- ▶ This is an ill-posed problem: there may be **no solution**, it may **not** be **unique**, and it may depend **sensitively** on  $y$ .
- ▶ Assumptions about  $\kappa$  are implemented via some form of **regularisation**. Bayesian approach:
  1. Describe our prior knowledge about  $\kappa$  in terms of a prior probability measure,  $\mu_0(\kappa)$ .
  2. Use Bayes' formula to calculate the posterior probability measure for  $\kappa$  given  $y$ ,  $\mu^y(\kappa)$ .
- ▶ To proceed with inference, we have two options:
  1. **Derive an inference scheme first, discretise afterwards**
  2. **Discretise first, derive an inference afterwards.**

## Bayesian Inference

- ▶ Discretise the problem before deriving an inference scheme.
- ▶ Many widely used methods fall under this approach:
  - **Metropolis-Hastings** (Metropolis et al. 1953, Hastings 1970)
  - **Hamiltonian Monte Carlo** (HMC) (Duane et al. 1987),
  - **Laplace approximation**,
  - **Variational Bayes** (Jordan et al. 1998, 1999).
- ▶ This leads to the familiar formulation of the Bayes' theorem:

$$p(\boldsymbol{\kappa} \mid \mathbf{y}) = \frac{p(\mathbf{y} \mid \boldsymbol{\kappa}) p(\boldsymbol{\kappa})}{p(\mathbf{y})} \propto p(\mathbf{y} \mid \boldsymbol{\kappa}) p(\boldsymbol{\kappa}), \quad (4)$$

where  $p(\boldsymbol{\kappa} \mid \mathbf{y})$  is the posterior density,  $p(\mathbf{y} \mid \boldsymbol{\kappa})$  is the likelihood of the observed data  $\mathbf{y}$  for a given discretised  $\boldsymbol{\kappa}$ .

## Likelihood – Back to the Poisson Problem

- ▶ The likelihood,  $p(\mathbf{y} | \boldsymbol{\kappa})$ , involves solving the PDE.
- ▶ We consider a specific problem – the Poisson problem:

$$-\nabla \cdot (\exp(\kappa(\mathbf{x})) \nabla u(\mathbf{x})) = f(\mathbf{x}), \quad (5)$$

- ▶ Discretise the weak form of the problem with a standard **finite element approach** to obtain a linear system:

$$\mathbf{A}(\boldsymbol{\kappa})\mathbf{u} = \mathbf{f}, \quad (6)$$

where  $\mathbf{A}(\boldsymbol{\kappa}) \in \mathbb{R}^{n_u \times n_u}$  is the stiffness matrix,  $\boldsymbol{\kappa} \in \mathbb{R}^{n_\kappa}$  is the vector of log-diffusion coefficients,  $\mathbf{f} \in \mathbb{R}^{n_u}$  is the nodal source vector.

- ▶ Finally, the likelihood is given by

$$p(\mathbf{y} | \boldsymbol{\kappa}) = p(\mathbf{y} | \mathbf{u}(\boldsymbol{\kappa})) = \mathcal{N}(\mathbf{P}\mathbf{A}(\boldsymbol{\kappa})^{-1}\mathbf{f}, \sigma_y^2 \mathbf{I}), \quad (7)$$

where the matrix  $\mathbf{P}$  represents the discretisation of the observation operator  $\mathcal{P}$ .

## Prior

- ▶ Place a zero-mean Gaussian Process prior on  $\kappa$ :

$$\kappa \sim \mathcal{GP}(0, k_{\text{SE}}(\cdot, \cdot)). \quad (8)$$

- ▶ Discretise  $\kappa$  by assuming  $\kappa(\mathbf{x})$  is constant on each element of the mesh. Centroids of the finite elements are the training points of the GP, giving:

$$p(\boldsymbol{\kappa}) = \mathcal{N}(\mathbf{0}, \mathbf{K}_\psi(\mathbf{X}, \mathbf{X})). \quad (9)$$

- ▶ Parameters  $\psi$  are covariance function parameters such as length-scale and amplitude.

## Stochastic Variational Bayes I

We posit a family of distributions  $\mathcal{D}_q$  from which we choose the minimiser of the Kullback-Leibler divergence:

$$q^*(\boldsymbol{\kappa}) = \arg \min_{q(\boldsymbol{\kappa}) \in \mathcal{D}_q} \text{KL}(q(\boldsymbol{\kappa}) \parallel p(\boldsymbol{\kappa} \mid \mathbf{y})), \quad (10)$$

where KL divergence is defined as

$$\text{KL}(q(\boldsymbol{\kappa}) \parallel p(\boldsymbol{\kappa} \mid \mathbf{y})) = \int \log \frac{q(\boldsymbol{\kappa})}{p(\boldsymbol{\kappa} \mid \mathbf{y})} q(\boldsymbol{\kappa}) d\boldsymbol{\kappa}. \quad (11)$$

The minimisation problem is equivalent to maximising the evidence lower bound (ELBO):

$$q^*(\boldsymbol{\kappa}) = \arg \max_{q(\boldsymbol{\kappa}) \in \mathcal{D}_q} \mathbb{E}_q [\log p(\mathbf{y} \mid \boldsymbol{\kappa})] - \text{KL}(q(\boldsymbol{\kappa}) \parallel p(\boldsymbol{\kappa})). \quad (12)$$

## Stochastic Variational Bayes II

Generally,  $\mathbb{E}_q [\log p(y | \kappa)]$  is not available in closed form. Instead, we use a Monte Carlo approximation with  $N_{\text{SVI}}$  samples from  $q(\kappa)$  as follows

$$\mathbb{E}_q [\log p(y | \kappa)] \approx N_{\text{SVI}}^{-1} \sum_{i=1}^{N_{\text{SVI}}} \log p(y | \kappa^{(i)}), \quad (13)$$

where  $\kappa^{(i)}$  is the  $i$ -th sample from  $q(\kappa)$ .

## Choice of Variational Family

- ▶ The choice determines how much structure of the true posterior distribution is captured.
- ▶ Trade-off between ‘richness’ and ‘convenience’.

Different parametrisations of  $q(\kappa)$  for the family of multivariate Gaussian distributions:

1. Diagonal covariance matrix, known as mean-field VB [MFVB]

$$q(\kappa) \sim \mathcal{N}(\mu, D)$$

2. Fully-specified Cholesky factor. (full-covariance VB [FCVB])

$$q(\kappa) \sim \mathcal{N}(\mu, LL^\top)$$

3. Cholesky factor of the sparse precision matrix [PMVB]

$$q(\kappa) \sim \mathcal{N}(\mu, (L_Q L_Q^\top)^{-1})$$

Other than Gaussians, other options include mixtures and hierarchical parametrisations.

## Precision Matrices and Conditional Independence

- ▶ Specify the multivariate Gaussian through the precision matrix:  
$$Q = \Sigma^{-1}$$
.
- ▶ The elements of the precision matrix reflect **conditional independence**:

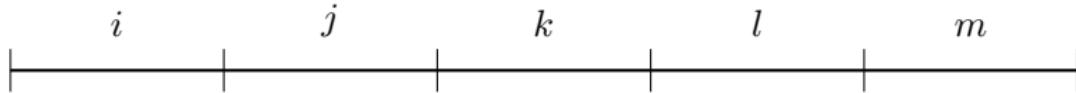
$$p(\kappa_i, \kappa_j \mid \kappa_{-\{i,j\}}) = p(\kappa_i \mid \kappa_{-\{i,j\}})p(\kappa_j \mid \kappa_{-\{i,j\}}) \Leftrightarrow Q_{ij} = 0, \quad (14)$$

whereas for **independence** (of Gaussian random variables!), we have:

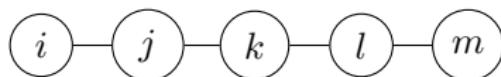
$$p(\kappa_i, \kappa_j) = p(\kappa_i)p(\kappa_j) \Leftrightarrow \Sigma_{ij} = 0. \quad (15)$$

- ▶ For more details, see Rue & Held (2005).

# Leveraging Sparsity 1D Example

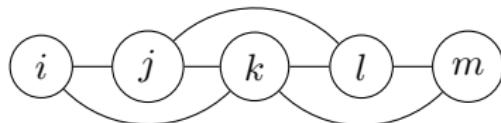


(a) Labelling of the five elements.



	$i$	$j$	$k$	$l$	$m$
$i$	×				
$j$		×			
$k$			×		
$l$				×	
$m$					×

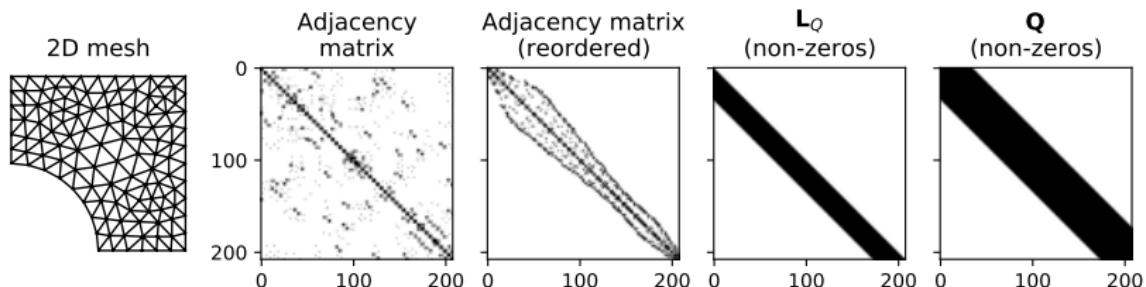
(b) 1-neighbourhood structure



	$i$	$j$	$k$	$l$	$m$
$i$	×				
$j$		×			
$k$			×		
$l$				×	
$m$					×

(c) 2-neighbourhood structure

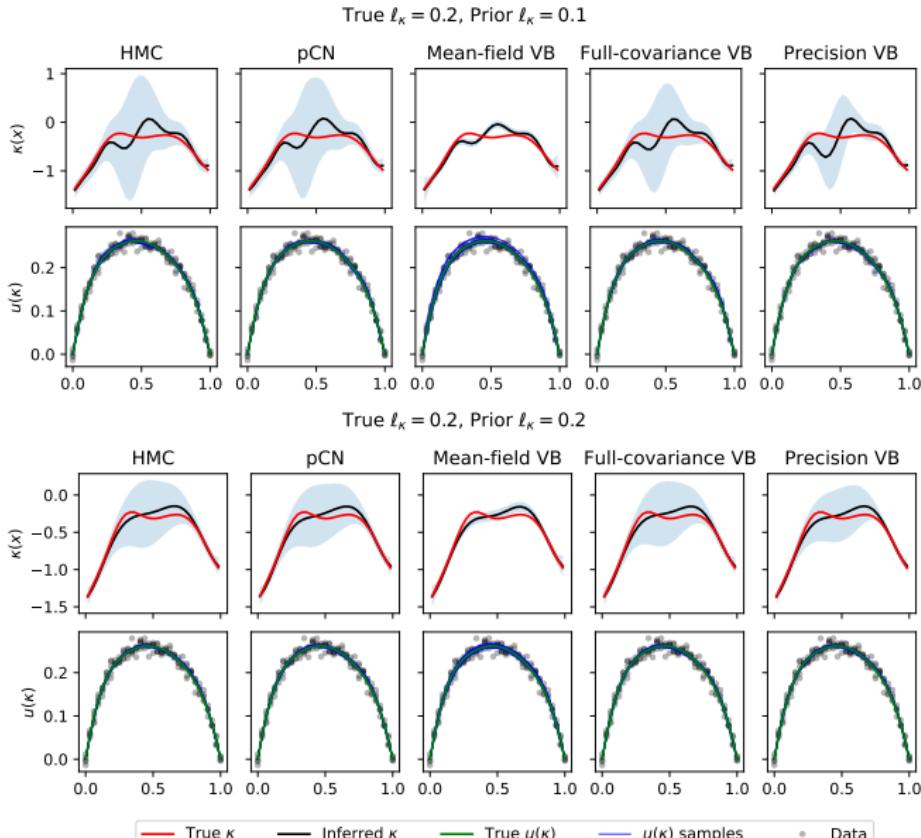
## Leveraging Sparsity 2D Example



## Experiments – 1D Poisson Problem

- ▶ Objective: infer the posterior distribution of  $\kappa$  given measurements of  $y$ .
- ▶ Sensors are placed on the discretisation nodes and the sensor noise is  $\sigma_y = 0.01$ . Five readings are taken.
- ▶ Constant forcing,  $f(x) = 1$  is assumed and we impose Dirichlet boundary conditions as  $u(0) = u(1) = 0$ .
- ▶ Data generated by assuming  $\kappa$  is a sample from a unit-variance zero-mean GP with lengthscale  $\ell_\kappa = 0.2$ .

# Results – Uncertainty Estimate

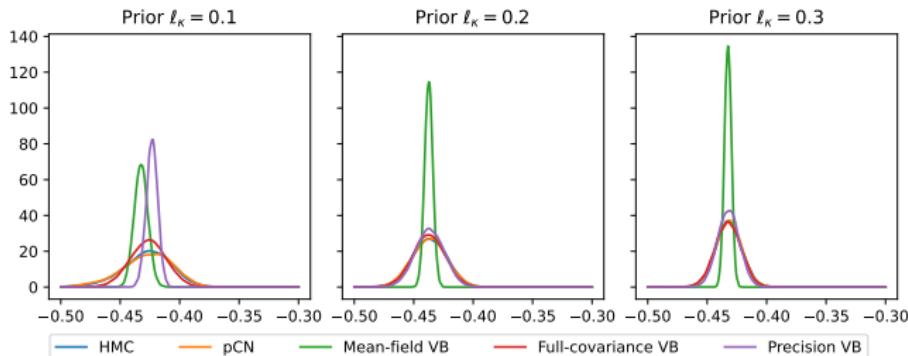


## Results – Uncertainty Propagation

We assess the propagation of uncertainty using log of total boundary flux through the left boundary:

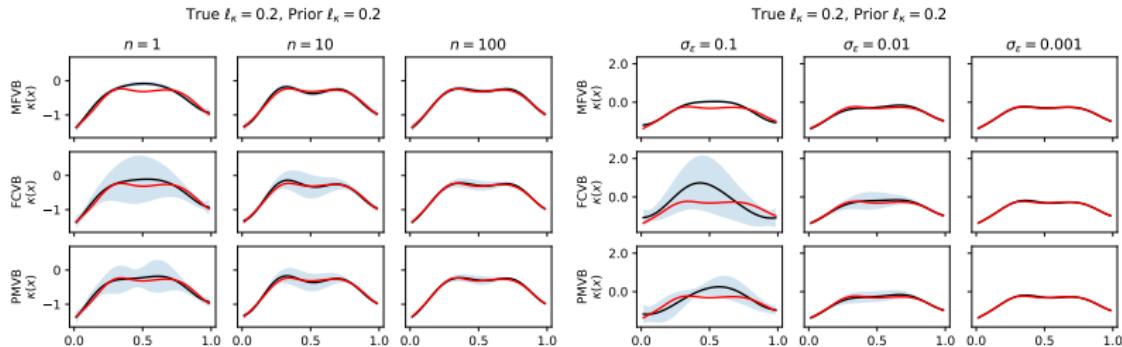
$$r(\kappa) = \log \int_{\Gamma_b} e^{\kappa(s)} \nabla u(s) \cdot \mathbf{n} \, ds, \quad (16)$$

where  $\mathbf{n}$  is a unit vector normal to the boundary.



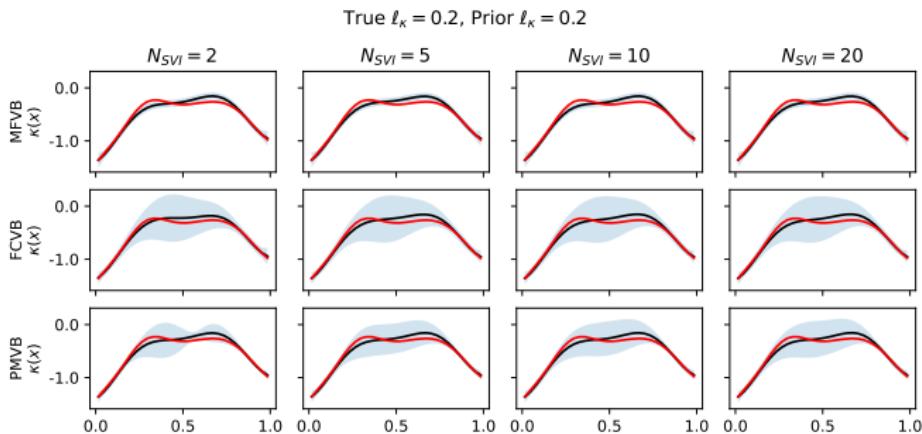
**Figure:** Log of the boundary flux at the left boundary node ( $x = 0$ ) for the 1D Poisson example. For PMVB, the precision matrix bandwidth of 10 is used.

## Results – Data Quality



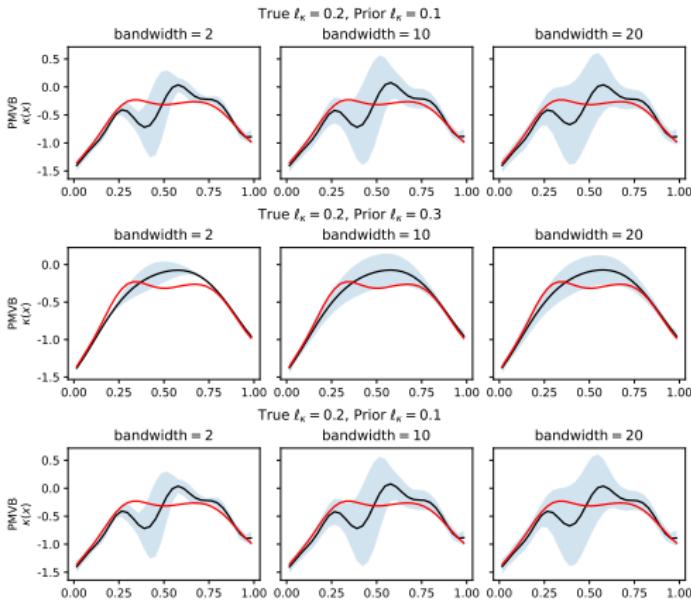
- ▶ Mean estimates for all methods get closer to the true  $\kappa$  with improved information.
- ▶ FCVB and PMVB uncertainty estimates get narrower with increasing number of observations and with decreasing observational noise.

## Results – Number of Samples for VB



- ▶ Figure shows the posterior estimates for different number of Monte Carlo samples in the estimation of ELBO,
- ▶ On a qualitative level, a small number of samples is sufficient to obtain a good estimate.

## Results – Increasing PMVB Bandwidth



- ▶ Increasing the order of neighbourhood improves the estimation of uncertainty (at the cost of more parameters).

## Conclusions

- ▶ the mean of the variational posterior provides an accurate point estimate irrespective of the choice of the parametrisation of the covariance structure,
- ▶ VB with a full-covariance or precision matrix structure adequately estimates posterior uncertainty compared to MCMC methods,
- ▶ sparse precision matrix parameterisation leverages the structure of the problem to balance computational complexity and the ability to capture dependencies in the posterior distribution,
- ▶ order of magnitude faster than MCMC methods,
- ▶ the VB estimates may be used effectively in propagating uncertainty.

## Implementation

- ▶ FEM C++ code written by CSMLab, led by Prof Fehmi Cirak.
- ▶ Our Tensorflow (Python) code performs the inference and interfaces with the FEM module.
- ▶ Available at <https://github.com/jp2011/bip-pde-vi>.
- ▶ For more details, see the paper: Povala et al. (2022)

## Bibliography I

- Duane, S., Kennedy, A. D., Pendleton, B. J. & Roweth, D. (1987), 'Hybrid Monte Carlo', *Physics Letters B* **195**(2), 216–222.
- Febrianto, E., Butler, L., Girolami, M. & Cirak, F. (2021), 'Digital twinning of self-sensing structures using the statistical finite element method'.
- Hastings, W. K. (1970), 'Monte Carlo sampling methods using Markov chains and their applications', *Biometrika* **57**(1), 97–109.
- Jordan, M. I., Ghahramani, Z., Jaakkola, T. S. & Saul, L. K. (1999), 'An introduction to variational methods for graphical models', *Machine Learning* **37**(2), 183–233.
- Jordan, R., Kinderlehrer, D. & Otto, F. (1998), 'The variational formulation of the Fokker-Planck equation', *Siam Journal on Mathematical Analysis* **29**(1), 1–17.
- Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H. & Teller, E. (1953), 'Equation of state calculations by fast computing machines', *The Journal of Chemical Physics* **21**(6), 1087–1092.

## Bibliography II

- Povala, J., Kazlauskaite, I., Febrianto, E., Cirak, F. & Girolami, M. (2022), 'Variational Bayesian approximation of inverse problems using sparse precision matrices', *Computer Methods in Applied Mechanics and Engineering* **393**, 114712.
- Rue, H. & Held, L. (2005), *Gaussian Markov Random Fields: Theory and Applications*, number 104 in 'Monographs on Statistics and Applied Probability', Chapman & Hall/CRC, Boca Raton.
- Stuart, A. M. (2010), 'Inverse problems: A Bayesian perspective', *Acta Numerica* **19**, 451–559.

## Results – Computation Time

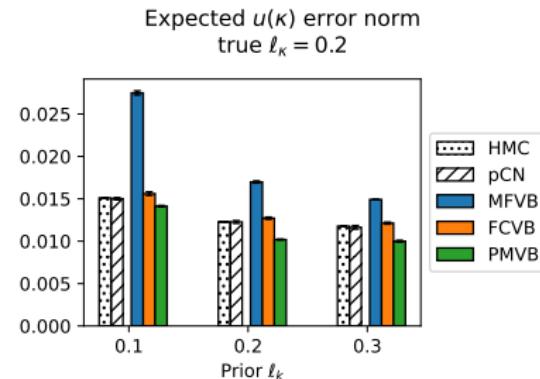
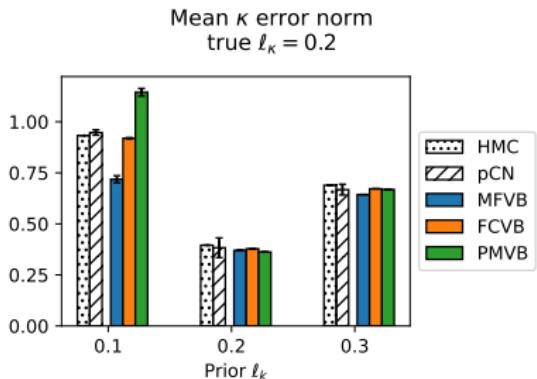
true $\ell_\kappa$	prior $\ell_\kappa$	Time (hours)			
		HMC	MFVB	FCVB	PMVB
0.1	0.1	15.2 (871–3244)	1.1	3.6	2.1
	0.2	11.1 (1043–4006)	0.7	2.7	2.1
	0.3	7.2 (1130–5408)	0.6	2.3	2.0
0.2	0.1	15.2 (1600–4700)	0.6	2.2	1.8
	0.2	10.4 (1067–3468)	0.6	2.3	2.0
	0.3	7.0 (1487–3969)	0.5	1.7	1.8

**Table:** Run-times in hours for the Poisson 1D problem. For VB methods,  $N_{\text{SVI}} = 3$ . The column for HMC includes the range of effective sample sizes (ESS) across different components of  $\kappa$ .

## Results – Error

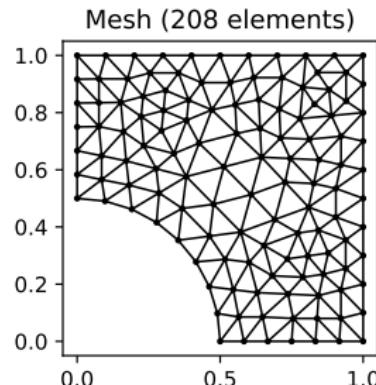
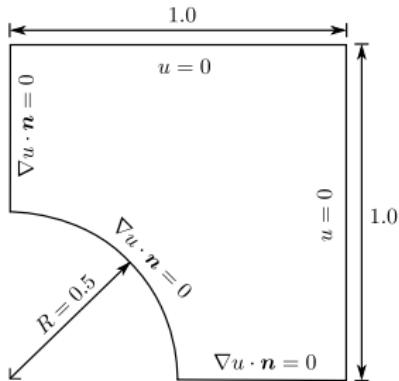
$$\left\| \frac{1}{S} \sum_{s=1}^S \boldsymbol{\kappa}^{(s)} - \boldsymbol{\kappa}_{\text{true}} \right\|_2$$

$$\frac{1}{S} \sum_{s=1}^S \left\| \mathbf{u}(\boldsymbol{\kappa}^{(s)}) - \mathbf{u}(\boldsymbol{\kappa}_{\text{true}}) \right\|_2$$

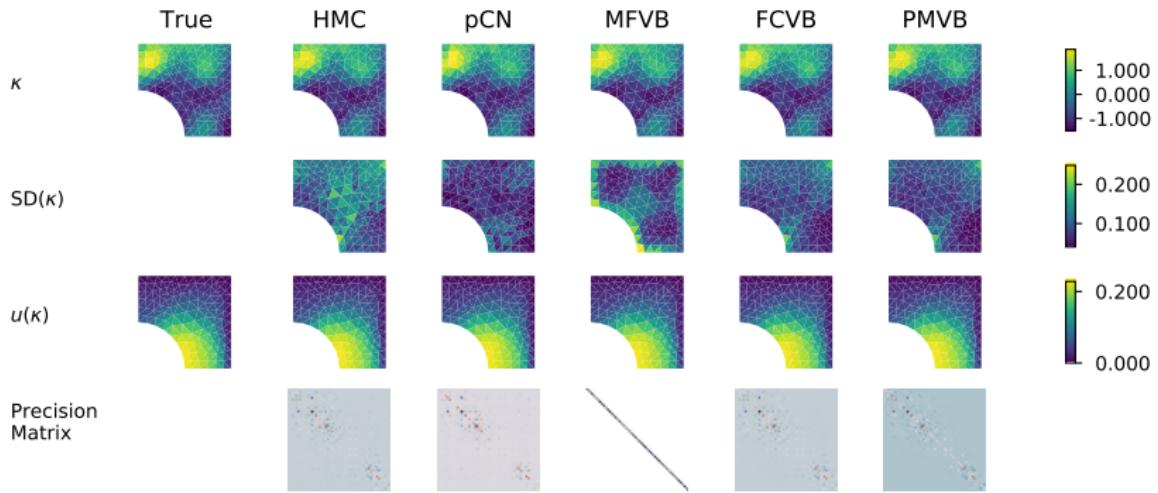


## Experiments – 2D Poisson Problem

- ▶ Same setup as before, but now with the sensor noise  $\sigma_y = 0.001$ , and different boundary conditions.
- ▶ Dirichlet boundary conditions  $u(x, y) = 0$  when  $x = 1$  or  $y = 1$ . Neumann boundary conditions on the rest of the boundary



## 2D Example – Results I



**Figure:** Posterior mean and standard deviation for  $\kappa$  and the corresponding  $u$  for 2D Poisson example with prior length-scale  $\ell_\kappa = 0.1$ . The bottom row shows the structure of the precision matrix for each inference scheme.

## 2D Example –Computation Cost

true $\ell_\kappa$	prior $\ell_\kappa$	Time (hours)				
		HMC	MFVB	FCVB	PMVB	
0.1	0.1	240.6 (930–11200)	6.4	29.6	28.1	
	0.2	295.5 (1537–11067)	6.6	32.6	28.9	
	0.3	242.0 (1057–6068)	7.3	27.3	30.6	
0.2	0.1	242.7 (1102–18235)	6.2	34.3	27.2	
	0.2	264.3 (1304–9848)	7.4	33.7	34.0	
	0.3	221.9 (1192–6356)	7.8	31.3	34.0	

**Table:** Run-times for different inference schemes in seconds. The number of Monte Carlo samples is  $N_{\text{SVI}} = 5$  for all MFVB, FCVB, and PMVB. The column for HMC includes the range of effective sample sizes (ESS) across different components of  $\kappa$ .