Using an SGFEM Surrogate to Accelerate Bayesian Inverse Uncertainty Quantification

James Rynn
with Simon Cotter, Catherine Powell (University of Manchester)
& Louise Wright (National Physical Laboratory)

MRSC, University of Manchester, UK

james.rynn@manchester.ac.uk

28/09/18

National Physical Laboratory

Overview

Bayesian Inverse Problems

Find the unknown $\boldsymbol{\theta}$ given n_z observations $\mathbf{z}$, satisfying

$$\mathbf{z} = \mathcal{G}(\boldsymbol{\theta}) + \boldsymbol{\eta}, \quad \boldsymbol{\eta} \sim \mathcal{N}(\mathbf{0}, \Sigma),$$

where

- $\mathbf{z} \in \mathbb{R}^{n_z}$ is a given vector of **observations**,
- $\mathcal{G} : \Theta \to \mathbb{R}^{n_z}$ is the **observation operator**,
- $\boldsymbol{\theta} \in \Theta$ is the **unknown**,
- $\boldsymbol{\eta} \in \mathbb{R}^{n_z}$ is a vector of **observational noise**.

Goal: Efficiently estimate the posterior density $\pi(\boldsymbol{\theta}|\mathbf{z})$ for the unknowns $\boldsymbol{\theta}$ given the data $\mathbf{z}$.

Bayes' Theorem

In the **finite-dimensional** case, from Bayes' Theorem we have

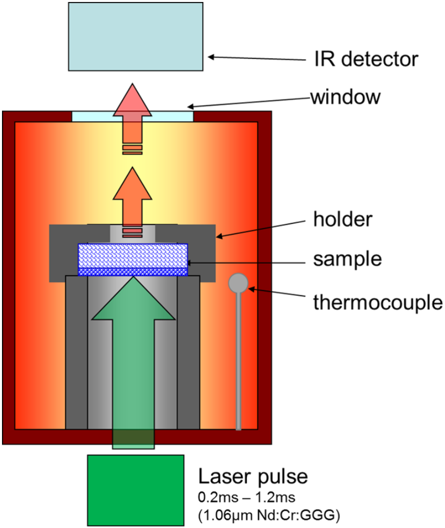
$$
\begin{aligned}
\pi(\boldsymbol{\theta}|\mathbf{z}) &\propto L(\mathbf{z}|\boldsymbol{\theta})\ \pi_0(\boldsymbol{\theta}) \\
&\propto \exp\left(-\frac{1}{2}\|\mathbf{z}-\mathcal{G}(\boldsymbol{\theta})\|_\Sigma^2\right)\ \pi_0(\boldsymbol{\theta}).
\end{aligned}
$$

Markov Chain Monte Carlo (MCMC) Methods

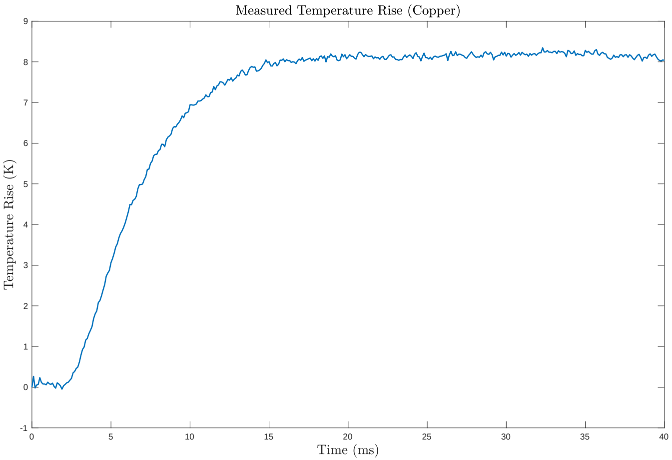
- We know $\pi(\boldsymbol{\theta}|\mathbf{z})$ up to a constant of proportionality.
- Use MCMC algorithm to generates samples $\boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \ldots, \boldsymbol{\theta}_M$ from the posterior distribution.
- Use these samples to construct Monte Carlo estimates of quantities of interest (means, variances and/or probabilities),
- e.g.

$$\mathbb{E}_\pi[\phi] = \int_\Theta \phi(\boldsymbol{\theta})\pi(\boldsymbol{\theta}|\mathbf{z})\mathrm{d}\boldsymbol{\theta} \approx \frac{1}{M}\sum_{i=1}^{M}\phi(\boldsymbol{\theta}_i).$$

Industrial Example



Industrial Example



Industrial Example

Possible unknowns:

- λ — thermal conductivity,
- I — laser intensity,
- κ — heat transfer coefficient,
- σ — standard deviation of measurement noise.

PDE Forward Problem

Solve

$$\varrho c_{\mathrm{p}} \frac{\partial u}{\partial t}(\mathbf{r}, t) - \nabla \cdot (\lambda \nabla u(\mathbf{r}, t)) = Q(\mathbf{r}, t),$$

where

$$Q(\mathbf{r}, t) = I \cdot \chi(r) \cdot 1_{\{[0, z_{\mathrm{f}}] \times [0, t_{\mathrm{f}}]\}}(z, t),$$

with appropriate initial and boundary conditions.

Goal: Find posterior density $\pi(\boldsymbol{\theta}|\mathbf{z})$ for the unknowns $\boldsymbol{\theta} := (\lambda, I)$, given observations $\mathbf{z}$ of the average (top) surface temperature $\bar{u}$ at the measurement times $t_1, t_2, \ldots, t_{n_z}$.

Observation Operator

Observation operator $\mathcal{G}$ is

$$\mathcal{G}(\boldsymbol{\theta}) = \left(\bar{u}(t_1;\boldsymbol{\theta}), \bar{u}(t_2;\boldsymbol{\theta}), \ldots, \bar{u}(t_{n_z};\boldsymbol{\theta})\right)^\top,$$

where

$$\bar{u}(t;\boldsymbol{\theta}) := \frac{1}{|D_L|}\int_{D_L} u(\mathbf{r}, t;\boldsymbol{\theta})\mathrm{d}S_z(\mathbf{r}),$$

is the average temperature over the surface D_L at time t.

Approximate $\mathcal{G}$ using

$$\mathcal{G}_{h\tau}(\boldsymbol{\theta}) = \left(\bar{u}_{h\tau}(t_1;\boldsymbol{\theta}), \bar{u}_{h\tau}(t_2;\boldsymbol{\theta}), \ldots, \bar{u}_{h\tau}(t_{n_z};\boldsymbol{\theta})\right)^\top.$$

Note: For each value of $\boldsymbol{\theta}$, to evaluate $\mathcal{G}_{h\tau}$ we are required to compute a FEM solve of a time-dependant PDE.

Random Walk Metropolis Hastings Algorithm (FEM)

Algorithm 1 RWMH Algorithm

set initial state $X^{(0)} = \boldsymbol{\theta}_0$
for $m = 1, 2, \ldots, M$ **do**
  draw proposal
  evaluate likelihood by **computing** $\mathcal{G}_{h\tau}$ **(expensive!)**
  compute acceptance probability α
  accept proposal with probability α
end for
output chain $X = (\boldsymbol{\theta}_0, \boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_M)$

Here $M \gg 10^5$.

(Results)

Unfortunately we cannot compute these as producing the samples takes far too long!

37 seconds per (time-dependant) PDE solve
$\implies$ 100m samples takes 3.7×10^9 seconds $\approx$ 117 years (1 CPU)

Parametric Forward Problem

Assume now that both λ and l may be expressed in terms of uniform random variables of mean zero and unit variance. That is,

$$\lambda = \mu_\lambda + \nu_\lambda \xi_1, \qquad l = \mu_l + \nu_l \xi_2,$$

for some given $\mu_\lambda, \mu_l, \nu_\lambda, \nu_l \in \mathbb{R}^+$ with

$$\xi_1, \xi_2 \sim \mathcal{U}(-\sqrt{3}, \sqrt{3}), \qquad \rho(\xi_i) = \frac{1}{2\sqrt{3}},$$

$$\mathbf{y} := (\xi_1(\omega), \xi_2(\omega))^\top \in \Gamma := (-\sqrt{3}, \sqrt{3})^2.$$

Parametric PDE

Solve

$$\varrho c_{\mathrm{p}} \frac{\partial u}{\partial t}(\mathbf{r}, t, \mathbf{y}) - \nabla \cdot (\lambda(\mathbf{y}) \nabla u(\mathbf{r}, t, \mathbf{y})) = Q(\mathbf{r}, t, \mathbf{y}),$$

where

$$Q(\mathbf{r}, t, \mathbf{y}) = I(\mathbf{y}) \cdot \chi(r) \cdot 1_{\{[0.z_{\mathrm{f}}] \times [0, t_{\mathrm{f}}]\}}(z, t),$$

again with appropriate initial and boundary conditions.

Solution: $u(\mathbf{r}, t, \mathbf{y})$.

PDE Forward Problem

Solve

$$\varrho c_{\mathrm{p}} \frac{\partial u}{\partial t}(\mathbf{r}, t) - \nabla \cdot (\lambda \nabla u(\mathbf{r}, t)) = Q(\mathbf{r}, t),$$

where

$$Q(\mathbf{r}, t) = I \cdot \chi(r) \cdot 1_{\{[0, z_{\mathrm{f}}] \times [0, t_{\mathrm{f}}]\}}(z, t),$$

with appropriate initial and boundary conditions.

Solution: $u(\mathbf{r}, t)$ for given $\theta = (\lambda, I)$.

SGFEM Solution

Compute a finite-dimensional approximation at time steps $\tau_1, \tau_2, \ldots, \tau_{n_t}$ such that for each $n = 1, 2, \ldots, n_t$,

$$u_{hk\tau}(\mathbf{r}, \tau_n, \mathbf{y}) = \sum_{i=1}^{n_h} \sum_{j=1}^{n_k} u_{ij}^n \phi_i(\mathbf{r}) \Psi_j(\mathbf{y}).$$

Approximate observation operator $\mathcal{G}_{hk\tau}$

$$\mathcal{G}_{hk\tau}(\mathbf{y}) = \left(\bar{u}_{hk\tau}(t_1; \mathbf{y}), \bar{u}_{hk\tau}(t_2; \mathbf{y}), \ldots, \bar{u}_{hk\tau}(t_{n_z}; \mathbf{y})\right)^\top .$$

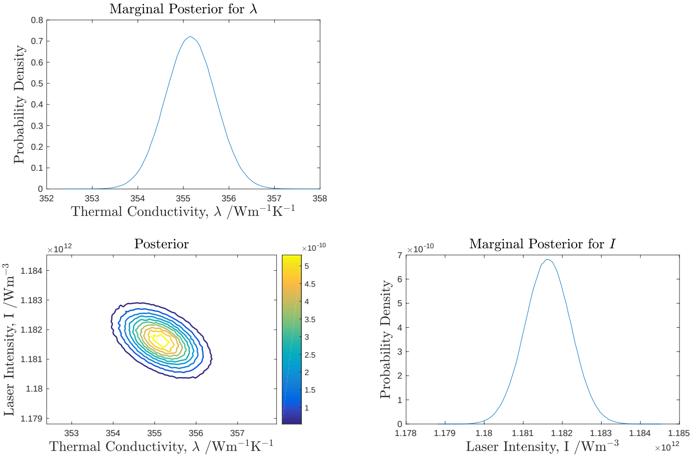
Random Walk Metropolis Hastings Algorithm (SGFEM)

Algorithm 2 RWMH Algorithm with SGFEM Surrogate

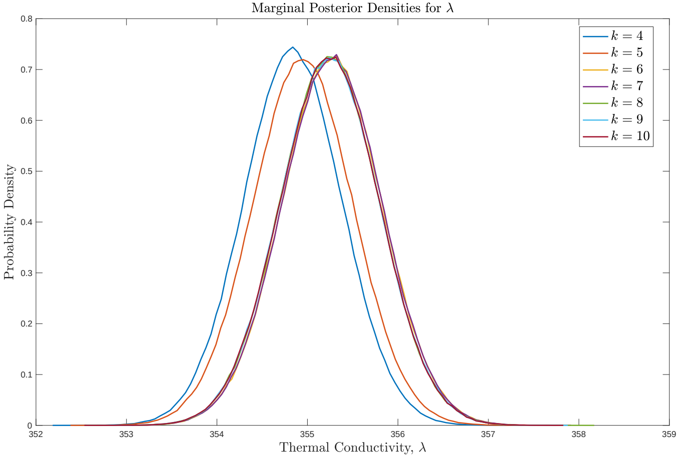
compute SGFEM solution $u_{hk\mathcal{T}}$
set initial state $X^{(0)} = \boldsymbol{\theta}_0$
for $m = 1, 2, \ldots, M$ **do**
  draw proposal
  evaluate likelihood by **evaluating** $\mathcal{G}_{hk\mathcal{T}}$ **(cheap!)**
  compute acceptance probability α
  accept proposal with probability α
end for
output chain $X = (\boldsymbol{\theta}_0, \boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_M)$

Here $M \gg 10^5$.

Posterior Density, $\pi(\theta|\mathbf{z})$



Posterior Convergence in k (Polynomial Degree)



Computational Time

Offline: Compute SGFEM solution with around 270 million DOF ($n_h \times n_t \times n_k = 12206 \times 800 \times 28$): 6 minutes

Online: Generate 100 million samples using SGFEM-RWMH: 4.28 hours

Total: 4.39 hours (1.58×10^{-4} seconds per sample).

Summary

- Achieved significant time saving (1 century $\rightarrow$ 1 "afternoon")
- Paper (in production): *"Surrogate accelerated Bayesian inversion for the determination of the thermal diffusivity of a material"*
- Extending work to a more realistic/complex forward problem:
  - spatial varying random variables
  - multi-layered material
  - express boundary heat loss parameter as random variable

V. Hoang, C. Schwab, and A. Stuart, *Complexity Analysis of Accelerated MCMC Methods for Bayesian Inversion*, Inverse Problems, 29 (2013), p. 085010.

Y. Marzouk and H. Najm, *Dimensionality Reduction and Polynomial Chaos Acceleration of Bayesian Inference in Inverse Problems*, Journal of Computational Physics, 228 (2009), pp. 1862–1902.

Y. Marzouk, H. Najm, and L. Rahn, *Stochastic Spectral Methods for Efficient Bayesian Solution of Inverse Problems*, Journal of Computational Physics, 224 (2007), pp. 560–586.

F. Nobile and R. Tempone, *Analysis and Implementation Issues for the Numerical Approximation of Parabolic Equations with Random Coefficients*, International Journal for Numerical Methods in Engineering, 80 (2009), pp. 979–1006.

MAP Estimate

Posterior:

$$\pi^{\mathbf{z}}(\boldsymbol{\theta}|\mathbf{z}) = \frac{1}{Z(\mathbf{z})} \exp(-\Phi(\boldsymbol{\theta}; \mathbf{z})) \cdot \pi_0(\boldsymbol{\theta}).$$

Maximum a posteriori (MAP) estimate $\boldsymbol{\theta}^{\mathrm{MAP}}$ satisfies

$$\boldsymbol{\theta}^{\mathrm{MAP}} := \mathrm{argmin}_{\boldsymbol{\theta} \in \mathbb{R}^2} \mathcal{J}(\boldsymbol{\theta}; \mathbf{z}),$$

where

$$\mathcal{J}(\boldsymbol{\theta}; \mathbf{z}) := -\log(\pi^{z}(\theta|\mathbf{z})),$$

is the negative log-posterior.

Gaussian Approximation

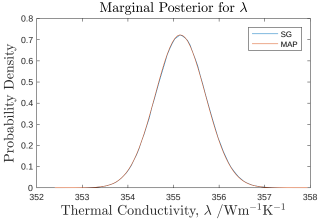
Assume

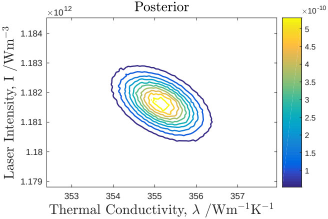
$$\boldsymbol{\theta}|\mathbf{z} \sim \mathcal{N}(\boldsymbol{\theta}^{\mathsf{MAP}}, \mathcal{C}).$$

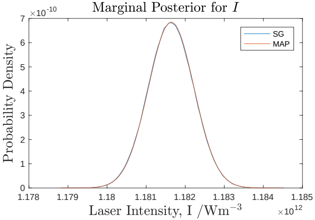
Let $\mathcal{H} \in \mathbb{R}^{2\times 2}$, $\mathcal{H}_{ij} := \frac{\partial^2 \mathcal{J}}{\partial\theta_i \partial\theta_j}$, be the Hessian of $\mathcal{J}$.

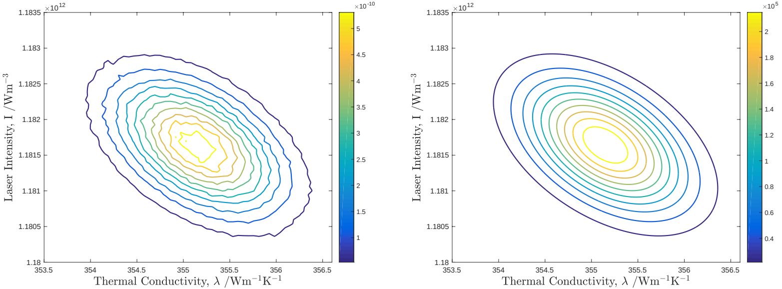
Approximating $\mathcal{H}$ at the point $\boldsymbol{\theta}^{\mathsf{MAP}}$ using finite differences we can compute an approximation to the covariance matrix

$$\mathcal{C} = \left(\mathcal{H}(\boldsymbol{\theta}^{\mathsf{MAP}})\right)^{-1}.$$

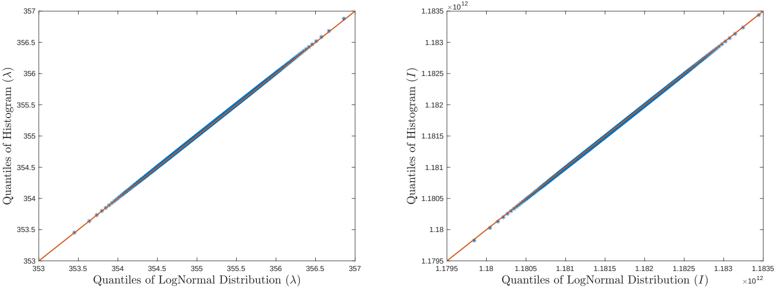

Marginal Posterior for λ
Probability Density
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
352
353
354
355
356
357
358
Thermal Conductivity, λ /Wm^{-1}K^{-1}
SG
MAP


Posterior
$\times 10^{12}$
Laser Intensity, I /Wm^{-3}
1.184
1.183
1.182
1.181
1.18
1.179
353
354
355
356
357
Thermal Conductivity, λ /Wm^{-1}K^{-1}
$\times 10^{-10}$
5
4.5
4
3.5
3
2.5
2
1.5
1


Marginal Posterior for I
$\times 10^{-10}$
Probability Density
7
6
5
4
3
2
1
0
1.178
1.179
1.18
1.181
1.182
1.183
1.184
1.185
Laser Intensity, I /Wm^{-3}
$\times 10^{12}$
SG
MAP



Relative error in the mean vectors is 8.45×10^{-6}.
Relative error in the covariance matrices is 4.00×10^{-3}.



Same distribution says Kolmogorov–Smirnov test at 1% level.