Bayesian inverse problems, Gaussian processes and partial differential equations

RICHARD NICKL

University of Cambridge (UK)

8th European Congress of Mathematics, Portoroz, June 24 2021

Acknowledgements

We present work in collaboration with a number of people:

- K. Abraham (Cambridge, now Paris)
- J. Bohr (Cambridge)
- M. Giordano (Cambridge, $\rightarrow$ Oxford)
- F. Monard (UC Santa Cruz)
- G.P. Paternain (Cambridge)
- S. Wang (Cambridge, now M.I.T.)

Statistical inverse regression models

Consider statistical observations arising as random vectors $(Y_i, X_i)_{i=1}^N$, where the X_i represent a discretisation of a d-dimensional manifold $\mathcal{X}$, where

$$Y_i = \mathscr{G}_\theta(X_i) + \varepsilon_i,\ \varepsilon_i \sim^{i.i.d.} \mathcal{N}_V(0, I_V),$$

where V is a vector space ($dim V < \infty$) and a collection of 'regression' fields

$$\{\mathscr{G}_\theta : \theta \in \Theta\},\quad \mathscr{G}_\theta : \mathcal{X} \to V$$

is indexed by the infinite-dimensional parameter $\theta \in \Theta$.

Statistical inverse regression models

Consider statistical observations arising as random vectors $(Y_i, X_i)_{i=1}^N$, where the X_i represent a discretisation of a d-dimensional manifold $\mathcal{X}$, where

$$Y_i = \mathscr{G}_\theta(X_i) + \varepsilon_i, \ \varepsilon_i \sim^{i.i.d.} \mathcal{N}_V(0, I_V),$$

where V is a vector space ($dim V < \infty$) and a collection of 'regression' fields

$$\{\mathscr{G}_\theta : \theta \in \Theta\}, \ \ \mathscr{G}_\theta : \mathcal{X} \to V$$

is indexed by the infinite-dimensional parameter $\theta \in \Theta$. In inverse problems

$$\mathscr{G} : \Theta \to L^2(\mathcal{X}, V), \ \ \theta \mapsto \mathscr{G}_\theta,$$

is the possibly non-linear solution 'forward' map of a PDE.

Statistical inverse regression models

Consider statistical observations arising as random vectors $(Y_i, X_i)_{i=1}^N$, where the X_i represent a discretisation of a d-dimensional manifold $\mathcal{X}$, where

$$Y_i = \mathscr{G}_\theta(X_i) + \varepsilon_i,\ \ \varepsilon_i \sim^{i.i.d.} \mathcal{N}_V(0, I_V),$$

where V is a vector space ($dim V < \infty$) and a collection of 'regression' fields

$$\{\mathscr{G}_\theta : \theta \in \Theta\},\ \ \mathscr{G}_\theta : \mathcal{X} \to V$$

is indexed by the infinite-dimensional parameter $\theta \in \Theta$. In inverse problems

$$\mathscr{G} : \Theta \to L^2(\mathcal{X}, V),\ \ \theta \mapsto \mathscr{G}_\theta,$$

is the possibly non-linear solution 'forward' map of a PDE.

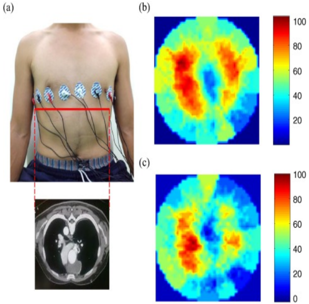
Example: Suppose $\mathscr{G}_\theta = u_\theta$ is the solution to

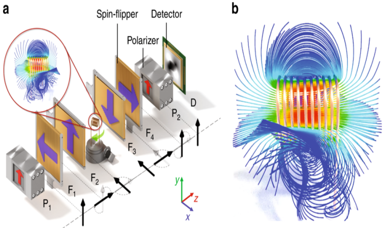
$$\begin{aligned} \mathscr{D}u + \theta u &= 0 \ \text{ on } \mathcal{X}, \\ u &= g \ \text{ on } \partial\mathcal{X}, \end{aligned}$$

with $\mathscr{D}$ a given linear partial differential op. (e.g., Laplacian, or geodesic vector field).

Examples for $\mathscr{G}$ from partial differential equations (PDEs)

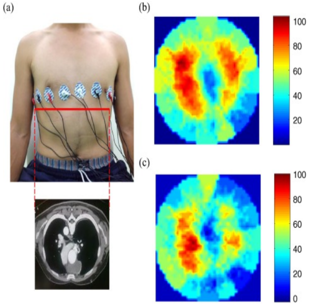
- Elliptic PDEs: Electric impedance tomography (Caldéron problem), groundwater flow & oil reservoir analysis, photoacoustics, etc..

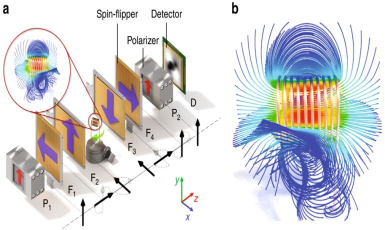




Examples for $\mathscr{G}$ from partial differential equations (PDEs)

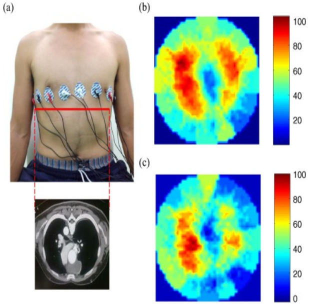
- Elliptic PDEs: Electric impedance tomography (Caldéron problem), groundwater flow & oil reservoir analysis, photoacoustics, etc..
- Transport PDEs: X-ray transforms in (non-linear) tomography

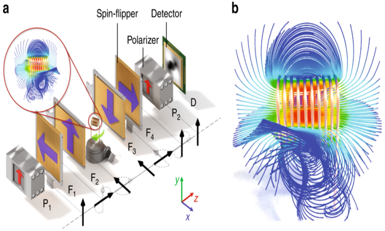




Examples for $\mathscr{G}$ from partial differential equations (PDEs)

- Elliptic PDEs: Electric impedance tomography (Caldéron problem), groundwater flow & oil reservoir analysis, photoacoustics, etc..
- Transport PDEs: X-ray transforms in (non-linear) tomography
- Time evolution equations: wave, heat, Euler and Navier-Stokes equations





Bayesian Inverse Problems (Stuart (2010))

The 'log-likelihood' or 'least squares fit' function is

$$\ell_N(\theta) = -\frac{1}{2}\sum_{i=1}^{N} \|Y_i - \mathscr{G}_\theta(X_i)\|_V^2, \quad \theta \in \Theta.$$

Note $-\ell_N$ is non-convex when $\mathscr{G}$ is non-linear.

Bayesian Inverse Problems (Stuart (2010))

The 'log-likelihood' or 'least squares fit' function is

$$\ell_N(\theta) = -\frac{1}{2}\sum_{i=1}^{N} \|Y_i - \mathscr{G}_\theta(X_i)\|_V^2, \quad \theta \in \Theta.$$

Note $-\ell_N$ is non-convex when $\mathscr{G}$ is non-linear. There is need for regularisation of the complexity of θ, which can be modelled by a Bayesian approach.

Let Π be p.m. on Θ arising from a Gaussian process prior. The posterior distribution $\Pi(\cdot|(Y_i, X_i)_{i=1}^N)$ on Θ then equals

$$d\Pi(\theta|(Y_i, X_i)_{i=1}^N) \propto e^{\ell_N(\theta)} d\Pi(\theta) \propto e^{\ell_N(\theta) - \frac{1}{2}\|\theta\|_{\mathcal{H}}^2}, \quad \theta \in \Theta.$$

where $\|\cdot\|_{\mathcal{H}}$ is the RKHS-norm associated to Π.

Bayesian Inverse Problems (Stuart (2010))

The 'log-likelihood' or 'least squares fit' function is

$$\ell_N(\theta) = -\frac{1}{2}\sum_{i=1}^{N} \|Y_i - \mathscr{G}_\theta(X_i)\|_V^2, \quad \theta \in \Theta.$$

Note $-\ell_N$ is non-convex when $\mathscr{G}$ is non-linear. There is need for regularisation of the complexity of θ, which can be modelled by a Bayesian approach.

Let Π be p.m. on Θ arising from a Gaussian process prior. The posterior distribution $\Pi(\cdot|(Y_i, X_i)_{i=1}^N)$ on Θ then equals

$$d\Pi(\theta|(Y_i, X_i)_{i=1}^N) \propto e^{\ell_N(\theta)} d\Pi(\theta) \propto e^{\ell_N(\theta) - \frac{1}{2}\|\theta\|_{\mathcal{H}}^2}, \quad \theta \in \Theta.$$

where $\|\cdot\|_{\mathcal{H}}$ is the RKHS-norm associated to Π.

- The posterior mode, also called maximum a posteriori (MAP) estimate coincides with a penalised least squares (Tikhonov) regulariser.

Bayesian Inverse Problems (Stuart (2010))

The 'log-likelihood' or 'least squares fit' function is

$$\ell_N(\theta) = -\frac{1}{2}\sum_{i=1}^{N} \|Y_i - \mathscr{G}_\theta(X_i)\|_V^2, \quad \theta \in \Theta.$$

Note $-\ell_N$ is non-convex when $\mathscr{G}$ is non-linear. There is need for regularisation of the complexity of θ, which can be modelled by a Bayesian approach.

Let Π be p.m. on Θ arising from a Gaussian process prior. The posterior distribution $\Pi(\cdot|(Y_i, X_i)_{i=1}^N)$ on Θ then equals

$$d\Pi(\theta|(Y_i, X_i)_{i=1}^N) \propto e^{\ell_N(\theta)} d\Pi(\theta) \propto e^{\ell_N(\theta) - \frac{1}{2}\|\theta\|_{\mathcal{H}}^2}, \quad \theta \in \Theta.$$

where $\|\cdot\|_{\mathcal{H}}$ is the RKHS-norm associated to Π.

- The posterior mode, also called maximum a posteriori (MAP) estimate coincides with a penalised least squares (Tikhonov) regulariser.
- Avoiding optimisation methods one can estimate θ by the posterior mean

$$E^\Pi[\theta|(Y_i, X_i)_{i=1}^N] = \int_\Theta \theta d\Pi(\theta|(Y_i, X_i)_{i=1}^N).$$

Computation: gradient based MCMC

Let us single out one popular MCMC method for illustration:

Unadjusted discretised Langevin algorithm

The log-posterior density on a high-dimensional discretisation space $\Theta \supset \mathbb{R}^D$ is

$$p_N(\theta) := \log \pi_N(\theta|(Y_i, X_i)_{i=1}^N) = \ell_N(\theta) - \frac{1}{2}\|\theta\|_{\mathcal{H}}^2, \quad \theta \in \mathbb{R}^D.$$

Then fix $\delta > 0$ and initialise at ϑ_0. For $k \geq 0$ and $\xi_k \sim^{iid} \mathcal{N}(0, I)$ in $\mathbb{R}^D$, do:

$$\vartheta_{k+1} = \vartheta_k + \delta \nabla p_N(\vartheta_k) + \sqrt{2\delta}\xi_k.$$

Computation: gradient based MCMC

Let us single out one popular MCMC method for illustration:

Unadjusted discretised Langevin algorithm

The log-posterior density on a high-dimensional discretisation space $\Theta \supset \mathbb{R}^D$ is

$$p_N(\theta) := \log \pi_N(\theta|(Y_i, X_i)_{i=1}^N) = \ell_N(\theta) - \frac{1}{2}\|\theta\|_{\mathcal{H}}^2, \quad \theta \in \mathbb{R}^D.$$

Then fix $\delta > 0$ and initialise at ϑ_0. For $k \geq 0$ and $\xi_k \sim^{iid} \mathcal{N}(0, I)$ in $\mathbb{R}^D$, do:

$$\vartheta_{k+1} = \vartheta_k + \delta \nabla p_N(\vartheta_k) + \sqrt{2\delta}\xi_k.$$

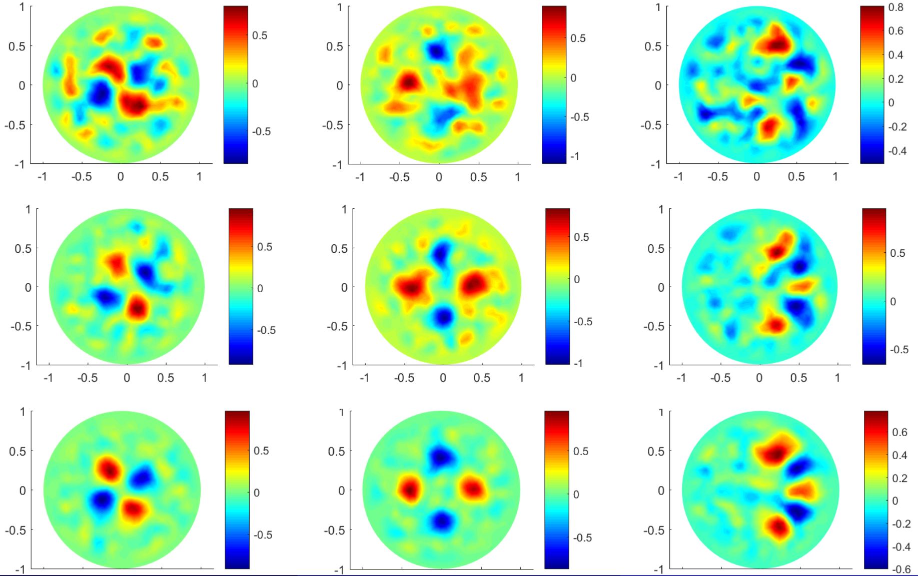
Note that (ϑ_k) is the Euler discretisation of the D-dimensional Langevin SDE

$$d\vartheta_t = \nabla \log p_N(\vartheta_t)dt + dW_t, \; 0 \leq t \leq T,$$

which has $\Pi(\cdot|(Y_i, X_i)_{i=1}^N)$ as invariant measure. [δ remains constant in k.]

Bayesian inversion with Gaussian process priors in action

Posterior mean fields for $N = 200, 400, 800$ and $N_s = 100000$ MCMC iterations



Mathematical and statistical guarantees?

Statistical consistency?

Does the posterior distribution reliably solve the inverse problem? Do we recover the ground truth θ_0 generating the data $(Y_i, X_i)_{i=1}^N \sim P_{\theta_0}^N$?

Mathematical and statistical guarantees?

Statistical consistency?

Does the posterior distribution reliably solve the inverse problem? Do we recover the ground truth θ_0 generating the data $(Y_i, X_i)_{i=1}^N \sim P_{\theta_0}^N$?

Computability?

Can the posterior measure $\Pi(\cdot|(Y_i, X_i)_{i=1}^N)$ (and its mean vector) be reliably computed specifically in high dimensions and when it is not log-concave?

Mathematical and statistical guarantees?

Statistical consistency?

Does the posterior distribution reliably solve the inverse problem? Do we recover the ground truth θ_0 generating the data $(Y_i, X_i)_{i=1}^N \sim P_{\theta_0}^N$?

Computability?

Can the posterior measure $\Pi(\cdot|(Y_i, X_i)_{i=1}^N)$ (and its mean vector) be reliably computed specifically in high dimensions and when it is not log-concave?

Uncertainty quantification?

Can we use posterior distributions for the construction of scientifically reliable confidence regions (i.e., 'error bars') for statistical inference on θ?

Mathematical and statistical guarantees?

Statistical consistency?

Does the posterior distribution reliably solve the inverse problem? Do we recover the ground truth θ_0 generating the data $(Y_i, X_i)_{i=1}^N \sim P_{\theta_0}^N$?

Computability?

Can the posterior measure $\Pi(\cdot|(Y_i, X_i)_{i=1}^N)$ (and its mean vector) be reliably computed specifically in high dimensions and when it is not log-concave?

Uncertainty quantification?

Can we use posterior distributions for the construction of scientifically reliable confidence regions (i.e., 'error bars') for statistical inference on θ?

- The answers to these questions depend on analytical properties of the PDE encoded in the non-linear map $\mathscr{G}$ and its linearisation $D\mathscr{G}_{\theta_0}$ near θ_0.

Algorithmic guarantees I: Posterior consistency

If the data arises from a ground truth θ_0, a first question one can ask concerns statistical consistency of the posterior measure: Do we have as $N \to \infty$ that

$$\Pi\big(\theta : \|\theta - \theta_0\| \geq \varepsilon_N | (Y_i, X_i)_{i=1}^N\big) \to^{P_{\theta_0}^N} 0?$$

- Here $\varepsilon_N \to 0$ is a 'posterior contraction rate' and $\|\cdot\|$ a norm on Θ. Convergence to θ_0 of the posterior mean vectors then typically follows too.

Algorithmic guarantees I: Posterior consistency

If the data arises from a ground truth θ_0, a first question one can ask concerns statistical consistency of the posterior measure: Do we have as $N \to \infty$ that

$$\Pi\big(\theta : \|\theta - \theta_0\| \geq \varepsilon_N | (Y_i, X_i)_{i=1}^N\big) \to^{P_{\theta_0}^N} 0?$$

- Here $\varepsilon_N \to 0$ is a 'posterior contraction rate' and $\|\cdot\|$ a norm on Θ. Convergence to θ_0 of the posterior mean vectors then typically follows too.
- Such theorems were recently proved in some PDE settings ($\|\cdot\| = \|\cdot\|_{L^2}$)

- Non-linear X-rays: $\varepsilon_N \approx N^{-\gamma}, \gamma > 0$ (Monard, Nickl, Paternain 2021)
- Calderón problem: $\varepsilon_N \approx (\log N)^{-\gamma}, \gamma > 0$ (Abraham, Nickl, 2019),
- 'Fruitfly' elliptic PDE: $\varepsilon_N \approx N^{-\gamma}, \gamma > 0$ (Giordano, Nickl, 2020)
- Parabolic Schrödinger eq., $\varepsilon_N \approx N^{-\gamma}, \gamma > 0$ (Kekkonen 2021)

Algorithmic guarantees I: Posterior consistency

If the data arises from a ground truth θ_0, a first question one can ask concerns statistical consistency of the posterior measure: Do we have as $N \to \infty$ that

$$\Pi\big(\theta : \|\theta - \theta_0\| \geq \varepsilon_N | (Y_i, X_i)_{i=1}^N\big) \to^{P_{\theta_0}^N} 0?$$

- Here $\varepsilon_N \to 0$ is a 'posterior contraction rate' and $\|\cdot\|$ a norm on Θ. Convergence to θ_0 of the posterior mean vectors then typically follows too.
- Such theorems were recently proved in some PDE settings ($\|\cdot\| = \|\cdot\|_{L^2}$)

- Non-linear X-rays: $\varepsilon_N \approx N^{-\gamma}, \gamma > 0$ (Monard, Nickl, Paternain 2021)
- Calderón problem: $\varepsilon_N \approx (\log N)^{-\gamma}, \gamma > 0$ (Abraham, Nickl, 2019),
- 'Fruitfly' elliptic PDE: $\varepsilon_N \approx N^{-\gamma}, \gamma > 0$ (Giordano, Nickl, 2020)
- Parabolic Schrödinger eq., $\varepsilon_N \approx N^{-\gamma}, \gamma > 0$ (Kekkonen 2021)

- See Giordano & Nickl (2020) and Bohr & Nickl (2021) for general results. The key analytical property of the PDE required is a stability estimate

$$\|\theta - \theta_0\| \lesssim \|\mathscr{G}(\theta) - \mathscr{G}(\theta_0)\|' \quad \forall \theta \in \Theta \text{ s.t. } \|\theta\|_{\mathcal{H}} \leq C$$

Algorithmic guarantees II: Computation in high-dimensions

Hairer, Stuart, Vollmer (2014) prove a D-independent spectral gap for the pCN MCMC scheme, allowing polynomial time computational guarantees when $\mathscr{G}$ is uniformly bounded and N is moderate (so that the posterior is still 'spread out').

This applies to a relevant class of PDEs but breaks down when N is large.

Algorithmic guarantees II: Computation in high-dimensions

Hairer, Stuart, Vollmer (2014) prove a D-independent spectral gap for the pCN MCMC scheme, allowing polynomial time computational guarantees when $\mathscr{G}$ is uniformly bounded and N is moderate (so that the posterior is still 'spread out').

This applies to a relevant class of PDEs but breaks down when N is large.

Nickl & Wang (2020), Bohr and Nickl (2021)

Assume (among other things) a 'gradient' stability condition

$$\|D\mathscr{G}_{\theta_0}[v]\|_{L^2}^2 \gtrsim D^{-\kappa}\|v\|^2 \quad \forall v \in \Theta, \quad \kappa \geq 0.$$

This can be verified for important classes of PDEs. Then Langevin-type MCMC can compute the posterior measure and its mean with precision $\epsilon > 0$ after

$$k_{mixing-time} = O(N^{b_1} D^{b_2} \epsilon^{-b_3}) \text{ Monte Carlo iterations.}$$

Algorithmic guarantees II: Computation in high-dimensions

Hairer, Stuart, Vollmer (2014) prove a D-independent spectral gap for the pCN MCMC scheme, allowing polynomial time computational guarantees when $\mathscr{G}$ is uniformly bounded and N is moderate (so that the posterior is still 'spread out').

This applies to a relevant class of PDEs but breaks down when N is large.

Nickl & Wang (2020), Bohr and Nickl (2021)

Assume (among other things) a 'gradient' stability condition

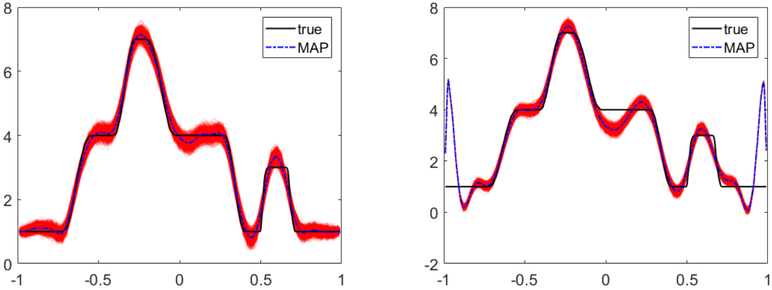
$$\|D\mathscr{G}_{\theta_0}[v]\|_{L^2}^2 \gtrsim D^{-\kappa}\|v\|^2 \quad \forall v \in \Theta, \quad \kappa \geq 0.$$

This can be verified for important classes of PDEs. Then Langevin-type MCMC can compute the posterior measure and its mean with precision $\epsilon > 0$ after

$$k_{mixing-time} = O(N^{b_1} D^{b_2} \epsilon^{-b_3}) \text{ Monte Carlo iterations.}$$

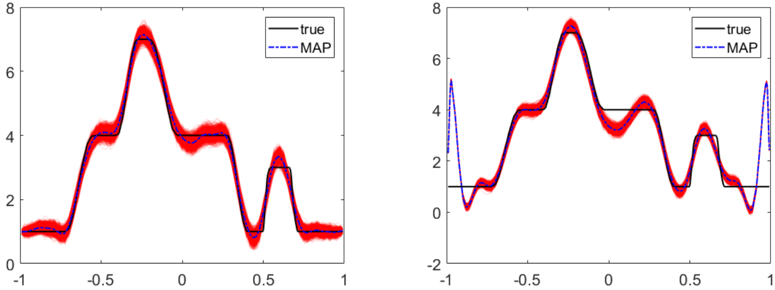
The proofs exploit i) that posterior measures are well approximated in global Wasserstein distance by log-concave measures ($N \to \infty$) and ii) fast dimension-free mixing of Langevin diffusions towards log-concave targets.

Algorithmic guarantees III: Uncertainty quantification



Do Bayesian methods provide valid frequentist error bars?

Algorithmic guarantees III: Uncertainty quantification

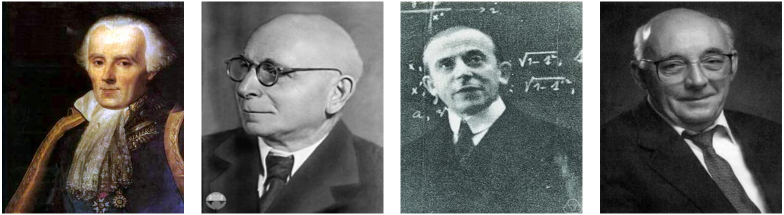


Do Bayesian methods provide valid frequentist error bars?

Confident credibility: If $R_N \equiv R_\alpha((Y_i, X_i)_{i=1}^N)$ are posterior quantiles for some norm, and $\bar{\theta}$ is the posterior mean or mode, do we have

$$P_{\theta_0}^N(\theta_0 \in \{\bar{\theta} - R_N, \bar{\theta} + R_N\}) \approx 1 - \alpha, \quad \text{as} \quad N \to \infty?$$

The Bernstein-von Mises (BvM) theorem



Discovered by Laplace (1812), expanded by Bernstein & von Mises in the early 20th century, and proved in general by Le Cam (1986), the BvM theorem states

$$\left\| \Pi(\cdot|(Y_i,X_i)_{i=1}^N) - \mathcal{N}(\bar{\theta}, I_N(\theta_0)^{-1}) \right\|_{TV} \to^{P_{\theta_0}^N}_{N\to\infty} 0, \ \theta_0 \in \Theta \subset \mathbb{R}^D,$$

when the prior Π has a positive density on Θ, if the Fisher information $I_N(\theta_0)$ is invertible, and for $D = dim(\Theta)$ fixed. See van der Vaart (1998) for proofs.

A key consequence for (typical) credible sets C_N is that

$$C_N \text{ s.t. } \Pi(C_N|(Y_i,X_i)_{i=1}^N) = 1-\alpha \Rightarrow P_{\theta_0}^N(\theta_0 \in C_N) \to 1-\alpha.$$

Bernstein-von Mises theorem for inverse problems

Following Castillo & Nickl (2013/14), such theorems do extend to $dim(\Theta) = \infty$.

Bernstein-von Mises theorem for inverse problems

Following Castillo & Nickl (2013/14), such theorems do extend to $dim(\Theta) = \infty$.

Theorem [Monard, Nickl, Paternain (2021)]

For sufficiently regular $\mathscr{G}$, smooth test functions and ground truth $\psi, \theta_0 \in C^\infty$, posterior draw $\theta \sim \Pi(\cdot|Y_i, X_i)_{i=1}^N$ and mean $\bar{\theta} = E^\Pi[\theta|(Y_i, X_i)_{i=1}^N]$, the statistics

$$\sqrt{N}\langle \theta - \bar{\theta}, \psi\rangle_{L^2(\mathcal{X})}|(Y_i, X_i)_{i=1}^N \to^d \mathcal{N}(0, \langle \psi, (I_{\theta_0}^* I_{\theta_0})^{-1}\psi\rangle_{L^2(\mathcal{X})})$$

are asymptotically Gaussian as $N \to \infty$, in $P_{\theta_0}^N$-probability.

Bernstein-von Mises theorem for inverse problems

Following Castillo & Nickl (2013/14), such theorems do extend to $dim(\Theta) = \infty$.

Theorem [Monard, Nickl, Paternain (2021)]

For sufficiently regular $\mathscr{G}$, smooth test functions and ground truth $\psi, \theta_0 \in C^\infty$, posterior draw $\theta \sim \Pi(\cdot|Y_i, X_i)_{i=1}^N$) and mean $\bar{\theta} = E^\Pi[\theta|(Y_i, X_i)_{i=1}^N]$, the statistics

$$\sqrt{N}\langle \theta - \bar{\theta}, \psi \rangle_{L^2(\mathcal{X})} | (Y_i, X_i)_{i=1}^N \to^d \mathcal{N}(0, \langle \psi, (I_{\theta_0}^* I_{\theta_0})^{-1} \psi \rangle_{L^2(\mathcal{X})})$$

are asymptotically Gaussian as $N \to \infty$, in $P_{\theta_0}^N$-probability.

- Here $I_{\theta_0}^* I_{\theta_0} = D\mathscr{G}_{\theta_0}^* D\mathscr{G}_{\theta_0}$ is the 'information operator' of the model, whose existence and surjectivity properties need to be ensured by PDE techniques.

Bernstein-von Mises theorem for inverse problems

Following Castillo & Nickl (2013/14), such theorems do extend to $dim(\Theta) = \infty$.

Theorem [Monard, Nickl, Paternain (2021)]

For sufficiently regular $\mathscr{G}$, smooth test functions and ground truth $\psi, \theta_0 \in C^\infty$, posterior draw $\theta \sim \Pi(\cdot|Y_i, X_i)_{i=1}^N$) and mean $\bar{\theta} = E^\Pi[\theta|(Y_i, X_i)_{i=1}^N]$, the statistics

$$\sqrt{N}\langle \theta - \bar{\theta}, \psi \rangle_{L^2(\mathcal{X})} | (Y_i, X_i)_{i=1}^N \to^d \mathcal{N}(0, \langle \psi, (I_{\theta_0}^* I_{\theta_0})^{-1} \psi \rangle_{L^2(\mathcal{X})})$$

are asymptotically Gaussian as $N \to \infty$, in $P_{\theta_0}^N$-probability.

- Here $I_{\theta_0}^* I_{\theta_0} = D\mathscr{G}_{\theta_0}^* D\mathscr{G}_{\theta_0}$ is the 'information operator' of the model, whose existence and surjectivity properties need to be ensured by PDE techniques.

- For the example inverse problem with $\mathscr{D} = \Delta$, the inverse $(I_{\theta_0}^* I_{\theta_0})^{-1}$ can be shown to exist on C_c^∞ using elliptic PDE techniques (cf. Nickl (2020)).

Bernstein-von Mises theorem for inverse problems

Following Castillo & Nickl (2013/14), such theorems do extend to $dim(\Theta) = \infty$.

Theorem [Monard, Nickl, Paternain (2021)]

For sufficiently regular $\mathscr{G}$, smooth test functions and ground truth $\psi, \theta_0 \in C^\infty$, posterior draw $\theta \sim \Pi(\cdot|Y_i, X_i)_{i=1}^N)$ and mean $\bar{\theta} = E^\Pi[\theta|(Y_i, X_i)_{i=1}^N]$, the statistics

$$\sqrt{N}\langle \theta - \bar{\theta}, \psi \rangle_{L^2(\mathcal{X})} |(Y_i, X_i)_{i=1}^N \to^d \mathcal{N}(0, \langle \psi, (I_{\theta_0}^* I_{\theta_0})^{-1} \psi \rangle_{L^2(\mathcal{X})})$$

are asymptotically Gaussian as $N \to \infty$, in $P_{\theta_0}^N$-probability.

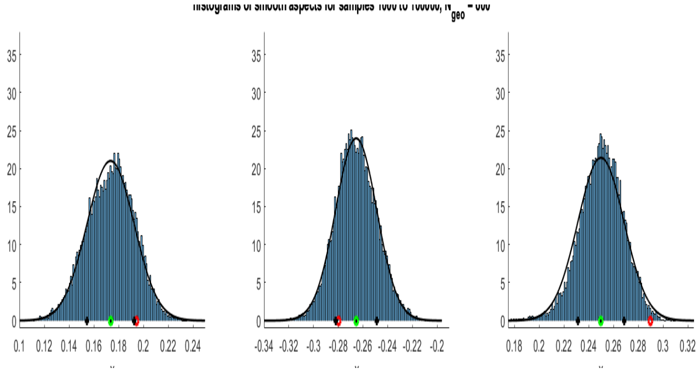
• Here $I_{\theta_0}^* I_{\theta_0} = D\mathscr{G}_{\theta_0}^* D\mathscr{G}_{\theta_0}$ is the 'information operator' of the model, whose existence and surjectivity properties need to be ensured by PDE techniques.

• For the example inverse problem with $\mathscr{D} = \Delta$, the inverse $(I_{\theta_0}^* I_{\theta_0})^{-1}$ can be shown to exist on C_c^∞ using elliptic PDE techniques (cf. Nickl (2020)).

• For X-ray transforms with $\mathscr{D}$ the geodesic vector field, $I_{\theta_0}^* I_{\theta_0}$ is a ΨDO of order -1, elliptic in the interior of $\mathcal{X}$. One uses Hörmander's generalised transmission condition or a calculus of Monard (2020) to deal with behaviour near $\partial\mathcal{X}$.

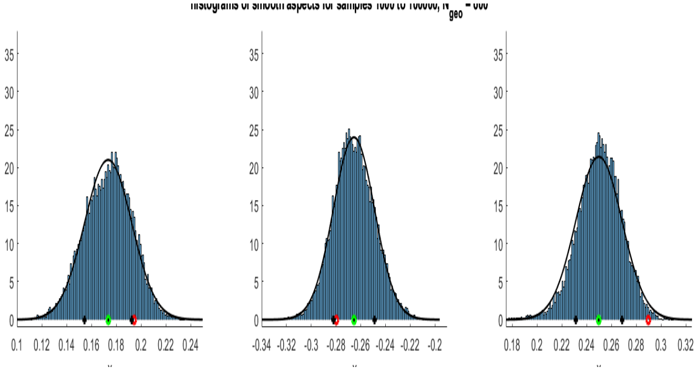
Numerical illustration

- Numerical MCMC plots of posterior draws of $\langle \theta, \psi \rangle_{L^2(\mathcal{X})} | (Y_i, X_i)_{i=1}^N$ around the posterior mean (green). The true value is marked in red, a Gaussian superimposed.



Numerical illustration

• Numerical MCMC plots of posterior draws of $\langle \theta, \psi \rangle_{L^2(\mathcal{X})} | (Y_i, X_i)_{i=1}^N$ around the posterior mean (green). The true value is marked in red, a Gaussian superimposed.



• These findings illustrate that the non-Gaussian posterior measures arising in non-linear inverse problems actually may produce Gaussian statistics for moderate $N = 600$. In particular this illustrates that uncertainty quantification based on posterior ‘credible sets’ often works in practice.

Does the inverse information $(I_{\theta_0}^* I_{\theta_0})^{-1}$ always exist?

We want to infer a smooth conductivity $\theta > 0$ from solutions $\mathscr{G}_\theta \equiv u_\theta$ of the PDE

$$\begin{aligned} \nabla \cdot (\theta \nabla u) &= f \text{ in } \mathcal{X}, \\ u &= g \text{ on } \partial\mathcal{X}. \end{aligned}$$

Here $f > 0$ is a given positive smooth source term, $\mathcal{X}$ is a bounded domain in $\mathbb{R}^d$, and g are given smooth boundary temperatures on $\partial\mathcal{X}$.

Does the inverse information $(I_{\theta_0}^* I_{\theta_0})^{-1}$ always exist?

We want to infer a smooth conductivity $\theta > 0$ from solutions $\mathscr{G}_\theta \equiv u_\theta$ of the PDE

$$\begin{aligned} \nabla \cdot (\theta \nabla u) &= f \text{ in } \mathcal{X}, \\ u &= g \text{ on } \partial \mathcal{X}. \end{aligned}$$

Here $f > 0$ is a given positive smooth source term, $\mathcal{X}$ is a bounded domain in $\mathbb{R}^d$, and g are given smooth boundary temperatures on $\partial \mathcal{X}$.

Bayesian inversion is consistent for θ thanks to global stability estimates for $\mathscr{G}_\theta$, see Giordano & Nickl (2020) (and Nickl, van de Geer, Wang (2020)).

Does the inverse information $(I_{\theta_0}^* I_{\theta_0})^{-1}$ always exist?

We want to infer a smooth conductivity $\theta > 0$ from solutions $\mathscr{G}_\theta \equiv u_\theta$ of the PDE

$$\nabla \cdot (\theta \nabla u) = f \text{ in } \mathcal{X},$$
$$u = g \text{ on } \partial\mathcal{X}.$$

Here $f > 0$ is a given positive smooth source term, $\mathcal{X}$ is a bounded domain in $\mathbb{R}^d$, and g are given smooth boundary temperatures on $\partial\mathcal{X}$.

Bayesian inversion is consistent for θ thanks to global stability estimates for $\mathscr{G}_\theta$, see Giordano & Nickl (2020) (and Nickl, van de Geer, Wang (2020)).

Nevertheless the Fisher information is not invertible in a large class of smooth directions, and hence the BvM phenomenon does not occur for this PDE.

Theorem [Nickl and Paternain (2021+)]

Suppose $\mathcal{X}$ is the unit disk and $g = 0, f = 2$. Suppose $0 \neq \psi \geq 0$ is smooth but vanishes along some straight line connecting $(0, 0)$ and the boundary. Then the Fisher information at $\theta = 1$ for inference on the functional $\Psi(\theta) = \langle \psi, \theta \rangle_{L^2}$ is zero, and no locally near $\theta = 1$ uniformly $\sqrt{N}$-consistent estimator of $\Psi(\theta)$ exists.

Conclusions

• After about 10 years of 'Bayesian inverse problems' becoming increasingly popular in applied mathematics, we have now some theoretical understanding of why such methods can be trusted in non-linear settings.

• The theory requires often fairly deep understanding of PDEs and high-dimensional statistics, and there cannot be a general 'off the shelf' theorem that says that Bayes methods always work for all purposes.

• The area certainly stimulates new questions at the interface of statistics, applied and pure mathematics. **Much more to be done!**

Conclusions

• After about 10 years of 'Bayesian inverse problems' becoming increasingly popular in applied mathematics, we have now some theoretical understanding of why such methods can be trusted in non-linear settings.

• The theory requires often fairly deep understanding of PDEs and high-dimensional statistics, and there cannot be a general 'off the shelf' theorem that says that Bayes methods always work for all purposes.

• The area certainly stimulates new questions at the interface of statistics, applied and pure mathematics. **Much more to be done!**

References

Bayesian consistency in PDE models

F. Monard, R. Nickl, G.P. Paternain, Consistent inversion of noisy non-Abelian X-ray transforms, *Comm. Pure Appl. Math.* (2021)

K. Abraham, R. Nickl, On statistical Calderón problems. *Math. Stat. Learn.* 2 (2019)

M. Giordano, R. Nickl, Consistency of Bayesian inference with Gaussian process priors in an elliptic inverse problem. *Inv. Prob.* 8 (2020)

Bernstein-von Mises theorems in PDE models

R. Nickl, Bernstein-von Mises theorems for statistical inverse problems I: Schrödinger equation, *J. Eur. Math. Soc.* 22 (2020)

F. Monard, R. Nickl, G.P. Paternain, Statistical guarantees for Bayesian uncertainty quantification in non-linear inverse problems with Gaussian process priors, *Ann. Statist.* (2021)

Computational guarantees in high-dimensional PDE models

R. Nickl, S. Wang, On polynomial-time computation of high-dimensional posterior measures by Langevin-type algorithms, *arXiv*, (2020)

J. Bohr, R. Nickl, On log-concave approximations of high-dimensional posterior measures and stability properties in non-linear inverse problems, *arXiv*, (2021)