

Bayesian Inference for Nonlinear Inverse Problems

RICHARD NICKL

Department of Pure Mathematics and Mathematical Statistics
University of Cambridge (UK)

Aad's 60th birthday conference
Leiden, June 17 2019



UNIVERSITY OF
CAMBRIDGE



European Research Council
Established by the European Commission

Aad's football career



Aad's football career



Aad's football career



Social Media & Dutch National Football

OVERVIEW

VIDEOS

STATS

TOP STORIES

MORE PEOPLE



Twitter



Facebook

Top stories

[MORE](#)



Football Oranje

[Van der Vaart announces retirement](#)

10 hours ago



football.london

I. STATISTICAL NONLINEAR INVERSE PROBLEMS

Statistical inverse regression models

- Inverse problems are often abstractly modelled by a **forward map**

$$f \mapsto \mathcal{G}(f), \quad \mathcal{G} : \mathbb{H}_1 \rightarrow \mathbb{H}_2, \quad \mathbb{H}_i \text{ Hilbert spaces.}$$

Statistical inverse regression models

- Inverse problems are often abstractly modelled by a **forward map**

$$f \mapsto \mathcal{G}(f), \quad \mathcal{G} : \mathbb{H}_1 \rightarrow \mathbb{H}_2, \quad \mathbb{H}_i \text{ Hilbert spaces.}$$

- A **statistical inverse problem** arises when **random noise** is added to $\mathcal{G}(f)$.

Statistical inverse regression models

- Inverse problems are often abstractly modelled by a **forward map**

$$f \mapsto \mathcal{G}(f), \quad \mathcal{G} : \mathbb{H}_1 \rightarrow \mathbb{H}_2, \quad \mathbb{H}_i \text{ Hilbert spaces.}$$

- A **statistical inverse problem** arises when **random noise** is added to $\mathcal{G}(f)$.
- Real world measurements are of discrete form $(Y_i, x_i)_{i=1}^N$, where the (x_i) are a uniform discretisation of the domain of $\mathcal{G}(f)$ and where

$$Y_i = \mathcal{G}(f)(x_i) + \varepsilon_i, \quad \varepsilon_i \sim N_V(0, \sigma^2 I_V), \quad \sigma > 0.$$

Statistical inverse regression models

- Inverse problems are often abstractly modelled by a **forward map**

$$f \mapsto \mathcal{G}(f), \quad \mathcal{G} : \mathbb{H}_1 \rightarrow \mathbb{H}_2, \quad \mathbb{H}_i \text{ Hilbert spaces.}$$

- A **statistical inverse problem** arises when **random noise** is added to $\mathcal{G}(f)$.
- Real world measurements are of discrete form $(Y_i, x_i)_{i=1}^N$, where the (x_i) are a uniform discretisation of the domain of $\mathcal{G}(f)$ and where

$$Y_i = \mathcal{G}(f)(x_i) + \varepsilon_i, \quad \varepsilon_i \sim N_V(0, \sigma^2 I_V), \quad \sigma > 0.$$

- Sometimes a continuous limit of this equation is considered: for $\mathbb{W}$ a centred Gaussian white noise indexed by $\mathbb{H}_2$, one postulates functional observations

$$Y = \mathcal{G}(f) + \varepsilon \mathbb{W}, \quad \text{at noise level } \varepsilon \approx \frac{\sigma}{\sqrt{N}}.$$

Statistical inverse regression models

- Inverse problems are often abstractly modelled by a **forward map**

$$f \mapsto \mathcal{G}(f), \quad \mathcal{G} : \mathbb{H}_1 \rightarrow \mathbb{H}_2, \quad \mathbb{H}_i \text{ Hilbert spaces.}$$

- A **statistical inverse problem** arises when **random noise** is added to $\mathcal{G}(f)$.
- Real world measurements are of discrete form $(Y_i, x_i)_{i=1}^N$, where the (x_i) are a uniform discretisation of the domain of $\mathcal{G}(f)$ and where

$$Y_i = \mathcal{G}(f)(x_i) + \varepsilon_i, \quad \varepsilon_i \sim N_V(0, \sigma^2 I_V), \quad \sigma > 0.$$

- Sometimes a continuous limit of this equation is considered: for $\mathbb{W}$ a centred Gaussian white noise indexed by $\mathbb{H}_2$, one postulates functional observations

$$Y = \mathcal{G}(f) + \varepsilon \mathbb{W}, \quad \text{at noise level } \varepsilon \approx \frac{\sigma}{\sqrt{N}}.$$

- There is a rich literature on how to solve such problems when $\mathcal{G}$ is linear, but much less is known when $\mathcal{G}$ is **non-linear**.

Penalised Least Squares / Tikhonov Regularisation

- A natural approach is to attempt to minimise a penalised least squares fit

$$\sum_{i=1}^n \|Y_i - \mathcal{G}(f)(x_i)\|_{V_2}^2 + \lambda \cdot \text{pen}(f), \quad \lambda > 0$$

over a suitable space $\mathcal{H}$ of parameters f .

Penalised Least Squares / Tikhonov Regularisation

- A natural approach is to attempt to minimise a penalised least squares fit

$$\sum_{i=1}^n \|Y_i - \mathcal{G}(f)(x_i)\|_{V_2}^2 + \lambda \cdot \text{pen}(f), \quad \lambda > 0$$

over a suitable space $\mathcal{H}$ of parameters f .

- Classical choices are **Sobolev norms**

$$\text{pen}(f) = \|f\|_{\mathcal{H}}^2 = \|f\|_{H^\alpha(\mathcal{O}_1)}^2 \simeq \max_{|j|=\alpha} \|D^j f\|_{L^2(\mathcal{O}_1)}^2, \quad \alpha \in \mathbb{N},$$

and the resulting estimators are sometimes called ‘**Tikhonov regularisers**’.

Penalised Least Squares / Tikhonov Regularisation

- A natural approach is to attempt to minimise a penalised least squares fit

$$\sum_{i=1}^n \|Y_i - \mathcal{G}(f)(x_i)\|_{V_2}^2 + \lambda \cdot \text{pen}(f), \quad \lambda > 0$$

over a suitable space $\mathcal{H}$ of parameters f .

- Classical choices are **Sobolev norms**

$$\text{pen}(f) = \|f\|_{\mathcal{H}}^2 = \|f\|_{H^\alpha(\mathcal{O}_1)}^2 \simeq \max_{|j|=\alpha} \|D^j f\|_{L^2(\mathcal{O}_1)}^2, \quad \alpha \in \mathbb{N},$$

and the resulting estimators are sometimes called ‘**Tikhonov regularisers**’.

- When $\mathcal{G}$ is non-linear, the associated least squares criterion is non-convex.

Penalised Least Squares / Tikhonov Regularisation

- A natural approach is to attempt to minimise a penalised least squares fit

$$\sum_{i=1}^n \|Y_i - \mathcal{G}(f)(x_i)\|_{V_2}^2 + \lambda \cdot \text{pen}(f), \quad \lambda > 0$$

over a suitable space $\mathcal{H}$ of parameters f .

- Classical choices are **Sobolev norms**

$$\text{pen}(f) = \|f\|_{\mathcal{H}}^2 = \|f\|_{H^\alpha(\mathcal{O}_1)}^2 \simeq \max_{|j|=\alpha} \|D^j f\|_{L^2(\mathcal{O}_1)}^2, \quad \alpha \in \mathbb{N},$$

and the resulting estimators are sometimes called ‘**Tikhonov regularisers**’.

- When $\mathcal{G}$ is non-linear, the associated least squares criterion is non-convex.
- Global minima of such optimisation problems can then not obviously be found, and alternatives to optimisation based methods are of interest.

Gaussian process priors and Inverse Problems

- Bayesian models for f based on Gaussian process priors Π are frequently used in such inverse problems following work by A. Stuart (2010, *Acta Numerica*).

Gaussian process priors and Inverse Problems

- Bayesian models for f based on Gaussian process priors Π are frequently used in such inverse problems following work by A. Stuart (2010, Acta Numerica).
- If Π is a Gaussian prior probability measure on the function space $\mathbb{H}_1$ with reproducing kernel Hilbert space (RKHS) $\mathcal{H}$, and if $\ell(f)$ is the log-likelihood function, then the posterior distribution formally equals

$$d\Pi(f|(Y_i, x_i)_{i=1}^N) \propto e^{\ell(f)} d\Pi(f) \propto e^{\ell(f) - \frac{1}{2}\|f\|_{\mathcal{H}}^2}, \quad f \in \mathbb{H}_1.$$

Gaussian process priors and Inverse Problems

- Bayesian models for f based on Gaussian process priors Π are frequently used in such inverse problems following work by A. Stuart (2010, Acta Numerica).
- If Π is a Gaussian prior probability measure on the function space $\mathbb{H}_1$ with reproducing kernel Hilbert space (RKHS) $\mathcal{H}$, and if $\ell(f)$ is the log-likelihood function, then the posterior distribution formally equals

$$d\Pi(f|(Y_i, x_i)_{i=1}^N) \propto e^{\ell(f)} d\Pi(f) \propto e^{\ell(f) - \frac{1}{2}\|f\|_{\mathcal{H}}^2}, \quad f \in \mathbb{H}_1.$$

- The posterior mode, also called maximum a posteriori (MAP) estimate $\hat{f}$, coincides with a Tikhonov regulariser since ($\sigma^2 = 1$)

$$-\ell(f) + \frac{1}{2}\|f\|_{\mathcal{H}}^2 \propto \frac{1}{n} \sum_{i=1}^n \|Y_i - \mathcal{G}(f)(x_i)\|_{V_2}^2 + \frac{1}{2N}\|f\|_{\mathcal{H}}^2.$$

Gaussian process priors and Inverse Problems

- Bayesian models for f based on Gaussian process priors Π are frequently used in such inverse problems following work by A. Stuart (2010, Acta Numerica).
- If Π is a Gaussian prior probability measure on the function space $\mathbb{H}_1$ with reproducing kernel Hilbert space (RKHS) $\mathcal{H}$, and if $\ell(f)$ is the log-likelihood function, then the posterior distribution formally equals

$$d\Pi(f|(Y_i, x_i)_{i=1}^N) \propto e^{\ell(f)} d\Pi(f) \propto e^{\ell(f) - \frac{1}{2}\|f\|_{\mathcal{H}}^2}, \quad f \in \mathbb{H}_1.$$

- The posterior mode, also called maximum a posteriori (MAP) estimate $\hat{f}$, coincides with a Tikhonov regulariser since ($\sigma^2 = 1$)

$$-\ell(f) + \frac{1}{2}\|f\|_{\mathcal{H}}^2 \propto \frac{1}{n} \sum_{i=1}^n \|Y_i - \mathcal{G}(f)(x_i)\|_{V_2}^2 + \frac{1}{2N}\|f\|_{\mathcal{H}}^2.$$

- Avoiding optimisation methods one can estimate f by the posterior mean

$$\bar{f} = \int_{\mathbb{H}_1} f d\Pi(f|(Y_i, x_i)_{i=1}^N).$$

Posterior sampling via MCMC

- For Gaussian priors (Cotter, Stuart, Roberts & White (Stat. Sci. 2013)):

pre-conditioned Crank-Nicholson (pCN) algorithm

Let Π be a Gaussian prior for f . Fix $\delta > 0$ and initialise f_0 . For $n \geq 0$ do:

1. Draw $\xi \sim \Pi$ and set $p_{f_n} = \sqrt{1 - 2\delta}f_n + \sqrt{2\delta}\xi$.
2. With $\ell(f) = -\sum_{i=1}^n \|Y_i - \mathcal{G}(f)(x_i)\|^2$ the log-likelihood function, set

$$f_{n+1} = \begin{cases} p_{f_n}, & \text{with probability } 1 \wedge \exp\{\ell(p_{f_n}) - \ell(f_n)\} \\ f_n, & \text{elsewhere.} \end{cases}$$

Posterior sampling via MCMC

- For Gaussian priors (Cotter, Stuart, Roberts & White (Stat. Sci. 2013)):

pre-conditioned Crank-Nicholson (pCN) algorithm

Let Π be a Gaussian prior for f . Fix $\delta > 0$ and initialise f_0 . For $n \geq 0$ do:

1. Draw $\xi \sim \Pi$ and set $p_{f_n} = \sqrt{1 - 2\delta}f_n + \sqrt{2\delta}\xi$.
2. With $\ell(f) = -\sum_{i=1}^n \|Y_i - \mathcal{G}(f)(x_i)\|^2$ the log-likelihood function, set

$$f_{n+1} = \begin{cases} p_{f_n}, & \text{with probability } 1 \wedge \exp\{\ell(p_{f_n}) - \ell(f_n)\} \\ f_n, & \text{elsewhere.} \end{cases}$$

- One can show that $\{f_n : n \geq 0\}$ forms a Markov chain with unique invariant measure equal to the posterior distribution $\Pi(\cdot | (Y_i, x_i)_{i=1}^N)$.

Posterior sampling via MCMC

- For Gaussian priors (Cotter, Stuart, Roberts & White (Stat. Sci. 2013)):

pre-conditioned Crank-Nicholson (pCN) algorithm

Let Π be a Gaussian prior for f . Fix $\delta > 0$ and initialise f_0 . For $n \geq 0$ do:

1. Draw $\xi \sim \Pi$ and set $p_{f_n} = \sqrt{1 - 2\delta}f_n + \sqrt{2\delta}\xi$.
2. With $\ell(f) = -\sum_{i=1}^n \|Y_i - \mathcal{G}(f)(x_i)\|^2$ the log-likelihood function, set

$$f_{n+1} = \begin{cases} p_{f_n}, & \text{with probability } 1 \wedge \exp\{\ell(p_{f_n}) - \ell(f_n)\} \\ f_n, & \text{elsewhere.} \end{cases}$$

- One can show that $\{f_n : n \geq 0\}$ forms a Markov chain with unique invariant measure equal to the posterior distribution $\Pi(\cdot | (Y_i, x_i)_{i=1}^N)$.
- Hairer, Stuart, Vollmer (2014, AoAP) prove non-asymptotic mixing and fast convergence of $\text{Law}(f_n) \rightarrow \Pi(\cdot | Y)$, under conditions on $\ell(f)$ that **do not require convexity**. The posterior mean $\bar{f}$ can then be reliably computed.

Posterior sampling via MCMC

- For Gaussian priors (Cotter, Stuart, Roberts & White (Stat. Sci. 2013)):

pre-conditioned Crank-Nicholson (pCN) algorithm

Let Π be a Gaussian prior for f . Fix $\delta > 0$ and initialise f_0 . For $n \geq 0$ do:

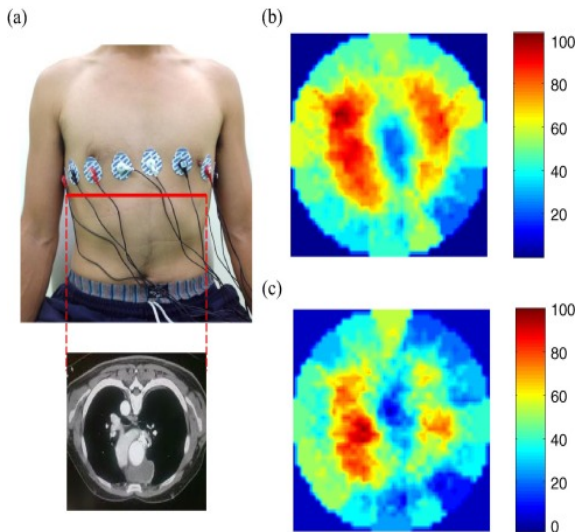
1. Draw $\xi \sim \Pi$ and set $p_{f_n} = \sqrt{1 - 2\delta}f_n + \sqrt{2\delta}\xi$.
2. With $\ell(f) = -\sum_{i=1}^n \|Y_i - \mathcal{G}(f)(x_i)\|^2$ the log-likelihood function, set

$$f_{n+1} = \begin{cases} p_{f_n}, & \text{with probability } 1 \wedge \exp\{\ell(p_{f_n}) - \ell(f_n)\} \\ f_n, & \text{elsewhere.} \end{cases}$$

- One can show that $\{f_n : n \geq 0\}$ forms a Markov chain with unique invariant measure equal to the posterior distribution $\Pi(\cdot | (Y_i, x_i)_{i=1}^N)$.
- Hairer, Stuart, Vollmer (2014, AoAP) prove non-asymptotic mixing and fast convergence of $\text{Law}(f_n) \rightarrow \Pi(\cdot | Y)$, under conditions on $\ell(f)$ that **do not require convexity**. The posterior mean $\bar{f}$ can then be reliably computed.
- Can we give statistical guarantees, for the convergence towards the ground truth f_0 , of the posterior mean $\bar{f}$ viewed purely as a frequentist algorithm?

I. TWO EXAMPLES OF NONLINEAR INVERSE PROBLEMS IN PDEs

Example 1: Electric Impedance Tomography (EIT)



The Calderón Problem

- For $\mathcal{O}$ some domain in $\mathbb{R}^d$, consider solutions $u = u_{f,g}$ of the PDE

$$\begin{aligned}\nabla \cdot (f \nabla u) &= 0 \text{ on } \mathcal{O} \\ \text{s.t. } u &= g \text{ on } \partial\mathcal{O}.\end{aligned}$$

The Calderón Problem

- For $\mathcal{O}$ some domain in $\mathbb{R}^d$, consider solutions $u = u_{f,g}$ of the PDE

$$\begin{aligned}\nabla \cdot (f \nabla u) &= 0 \text{ on } \mathcal{O} \\ \text{s.t. } u &= g \text{ on } \partial\mathcal{O}.\end{aligned}$$

- In EIT one measures the resulting current flux on the patient's skin $\partial\mathcal{O}$ arising from boundary voltage applications described by g . Mathematically one is given 'Neumann-data'

$$\left\{ \Lambda_f(g) = \frac{\partial u_{f,g}}{\partial \nu} \Big|_{\partial\mathcal{O}} : g \in H^{1/2}(\partial\mathcal{O}) \right\}.$$

Note that the map $f \mapsto \Lambda_f(\cdot)$ is nonlinear. Does Λ_f identify f ?

The Calderón Problem

- For $\mathcal{O}$ some domain in $\mathbb{R}^d$, consider solutions $u = u_{f,g}$ of the PDE

$$\begin{aligned}\nabla \cdot (f \nabla u) &= 0 \text{ on } \mathcal{O} \\ \text{s.t. } u &= g \text{ on } \partial\mathcal{O}.\end{aligned}$$

- In EIT one measures the resulting current flux on the patient's skin $\partial\mathcal{O}$ arising from boundary voltage applications described by g . Mathematically one is given 'Neumann-data'

$$\left\{ \Lambda_f(g) = \frac{\partial u_{f,g}}{\partial \nu} \Big|_{\partial\mathcal{O}} : g \in H^{1/2}(\partial\mathcal{O}) \right\}.$$

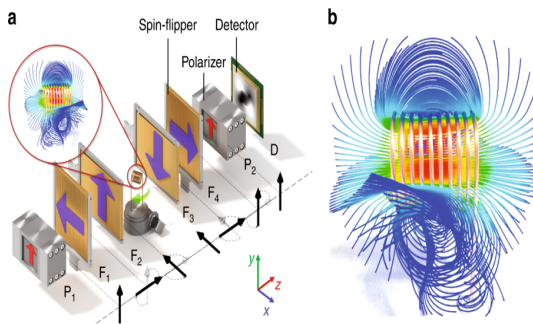
Note that the map $f \mapsto \Lambda_f(\cdot)$ is nonlinear. Does Λ_f identify f ?

Theorem [Sylvester & Uhlmann, 1987 Ann. Math.]

Suppose $d \geq 3$. If $\Lambda_{f_1} = \Lambda_{f_2} \Rightarrow f_1 = f_2$.

- See Uhlmann (2009, Inv. Prob.) for a survey.

Example 2: Polarised Neutron Tomography (PNT)

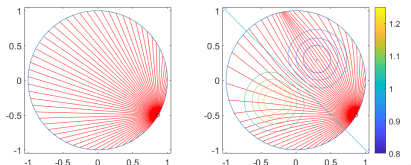


In Hilger et al. (Nat. Comm. 2018), Lionheart et al. (Nat. Rep. 2018)], physical neutron spin experiments are performed to reconstruct a 'magnetic field'

$$\Phi(x) = \begin{bmatrix} 0 & B_3(x) & -B_2(x) \\ -B_3(x) & 0 & B_1(x) \\ B_2(x) & -B_1(x) & 0 \end{bmatrix}, \quad x \in M.$$

from certain Lie group valued scattering data. Here $M = \{x \in \mathbb{R}^2 : \|x\| \leq 1\}$.

Matrix ODEs and X-ray transforms

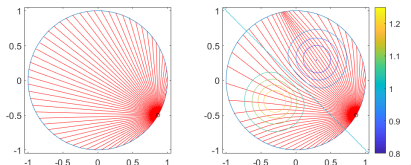


- Let M be the unit disk and consider geodesics

$$\gamma = \gamma(t) = \gamma_{(x,v)}(t), \quad 0 \leq t \leq \tau_\gamma$$

from $x \in \partial M$ into M in v -direction with 'exit time' τ_γ .

Matrix ODEs and X-ray transforms



- Let M be the unit disk and consider geodesics

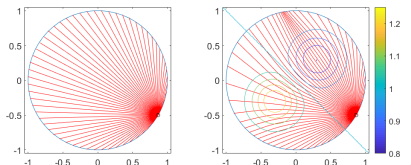
$$\gamma = \gamma(t) = \gamma_{(x,v)}(t), \quad 0 \leq t \leq \tau_\gamma$$

from $x \in \partial M$ into M in v -direction with 'exit time' τ_γ .

- Given a **matrix field** $\Phi : M \rightarrow \mathbb{C}^{n \times n}$, consider the following linear matrix ODE along each geodesic $\gamma \in \partial_+ SM$:

$$\frac{d}{dt}S + \Phi(\gamma(t))S = 0, \quad S(0) = \text{id}, \quad 0 \leq t \leq \tau_\gamma.$$

Matrix ODEs and X-ray transforms



- Let M be the unit disk and consider geodesics

$$\gamma = \gamma(t) = \gamma_{(x,v)}(t), \quad 0 \leq t \leq \tau_\gamma$$

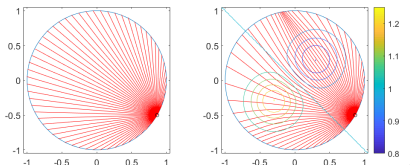
from $x \in \partial M$ into M in v -direction with 'exit time' τ_γ .

- Given a **matrix field** $\Phi : M \rightarrow \mathbb{C}^{n \times n}$, consider the following linear matrix ODE along each geodesic $\gamma \in \partial_+ SM$:

$$\frac{d}{dt}S + \Phi(\gamma(t))S = 0, \quad S(0) = \text{id}, \quad 0 \leq t \leq \tau_\gamma.$$

- The solution $S(t)$ of this ODE generates scattering data $\mathbf{C}_\Phi(\gamma) := S(\tau_\gamma)$.

Matrix ODEs and X-ray transforms



- Let M be the unit disk and consider geodesics

$$\gamma = \gamma(t) = \gamma_{(x,v)}(t), \quad 0 \leq t \leq \tau_\gamma$$

from $x \in \partial M$ into M in v -direction with 'exit time' τ_γ .

- Given a **matrix field** $\Phi : M \rightarrow \mathbb{C}^{n \times n}$, consider the following linear matrix ODE along each geodesic $\gamma \in \partial_+ SM$:

$$\frac{d}{dt} S + \Phi(\gamma(t)) S = 0, \quad S(0) = \text{id}, \quad 0 \leq t \leq \tau_\gamma.$$

- The solution $S(t)$ of this ODE generates scattering data $C_\Phi(\gamma) := S(\tau_\gamma)$.
- For $n = 1$ we obtain the usual Radon transform as then

$$C_\Phi(\gamma) = e^{-\int_0^{\tau_\gamma} \Phi(\gamma(t)) dt}.$$

Non-Abelian X-ray transforms

- If Φ takes values in the Lie algebra $\mathfrak{so}(n)$, i.e.,

$$\Phi(x) = \begin{bmatrix} 0 & B_3(x) & -B_2(x) \\ -B_3(x) & 0 & B_1(x) \\ B_2(x) & -B_1(x) & 0 \end{bmatrix}, \quad x \in M, \text{ when } n = 3,$$

then the scattering data $C_\Phi(\gamma)$ takes values in the compact Lie group

$$SO(n) = \{C \in \mathbb{R}^{n \times n} : C^T C = id\}.$$

Non-Abelian X-ray transforms

- If Φ takes values in the Lie algebra $\mathfrak{so}(n)$, i.e.,

$$\Phi(x) = \begin{bmatrix} 0 & B_3(x) & -B_2(x) \\ -B_3(x) & 0 & B_1(x) \\ B_2(x) & -B_1(x) & 0 \end{bmatrix}, \quad x \in M, \text{ when } n = 3,$$

then the **scattering data** $C_\Phi(\gamma)$ takes values in the compact Lie group

$$SO(n) = \{C \in \mathbb{R}^{n \times n} : C^T C = id\}.$$

- The **inverse problem** is to recover the matrix (magnetic) field

$$\Phi : M \rightarrow \mathfrak{so}(n) \text{ from } C_\Phi : \partial_+ SM \rightarrow SO(n).$$

Non-Abelian X-ray transforms

- If Φ takes values in the Lie algebra $\mathfrak{so}(n)$, i.e.,

$$\Phi(x) = \begin{bmatrix} 0 & B_3(x) & -B_2(x) \\ -B_3(x) & 0 & B_1(x) \\ B_2(x) & -B_1(x) & 0 \end{bmatrix}, \quad x \in M, \text{ when } n = 3,$$

then the **scattering data** $C_\Phi(\gamma)$ takes values in the compact Lie group

$$SO(n) = \{C \in \mathbb{R}^{n \times n} : C^T C = id\}.$$

- The **inverse problem** is to recover the matrix (magnetic) field

$$\Phi : M \rightarrow \mathfrak{so}(n) \text{ from } C_\Phi : \partial_+ SM \rightarrow SO(n).$$

Theorem [Paternain, Salo, Uhlmann (Geom. Funct. Anal. 2012)]

Let Φ, Ψ be maps defined on M that take values in the algebra $\mathfrak{so}(n)$ of $n \times n$ skew-symmetric matrices, $n \geq 2$. If their non-Abelian Radon transforms $C_\Phi = C_\Psi$ coincide as maps into $SO(n)$, then $\Phi = \Psi$.

- For $n \geq 2$ the map $\Phi \mapsto C_\Phi$ involves a non-linear matrix exponential.

Statistical observations in PNT

- Consider the random design regression model

$$Y_{i,j,k} = (C_\Phi(X_i, V_i))_{j,k} + \varepsilon_{i,j,k}, \quad \varepsilon_{i,j,k} \sim^{i.i.d.} N(0, \sigma^2), \quad \sigma^2 > 0,$$

with $1 \leq i \leq N, 1 \leq j, k \leq 3$, representing the physical PNT experiments.

Statistical observations in PNT

- Consider the random design regression model

$$Y_{i,j,k} = (C_\Phi(X_i, V_i))_{j,k} + \varepsilon_{i,j,k}, \quad \varepsilon_{i,j,k} \sim^{i.i.d.} N(0, \sigma^2), \quad \sigma^2 > 0,$$

with $1 \leq i \leq N, 1 \leq j, k \leq 3$, representing the physical PNT experiments.

- Here the (X_i, V_i) are sampled i.i.d. from the uniform distribution on geodesic space $\partial_+ SM$, and where C_Φ are the scattering data in $SO(n)$.

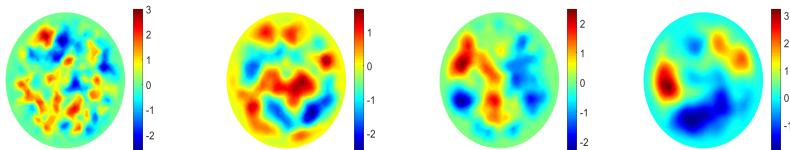
Statistical observations in PNT

- Consider the random design regression model

$$Y_{i,j,k} = (C_\Phi(X_i, V_i))_{j,k} + \varepsilon_{i,j,k}, \quad \varepsilon_{i,j,k} \sim^{i.i.d.} N(0, \sigma^2), \quad \sigma^2 > 0,$$

with $1 \leq i \leq N, 1 \leq j, k \leq 3$, representing the physical PNT experiments.

- Here the (X_i, V_i) are sampled i.i.d. from the **uniform distribution on geodesic space $\partial_+ SM$** , and where C_Φ are the scattering data in $SO(n)$.
- The vector field entries B_1, B_2, B_3 of Φ are assigned independent **Matérn process priors** on M .



Scattering Data (in $\mathfrak{su}(2)$) for neutron tomography

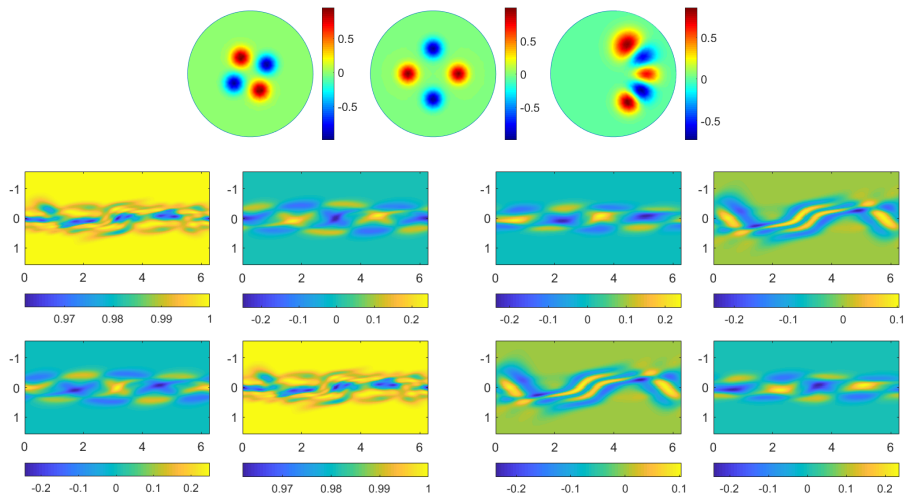


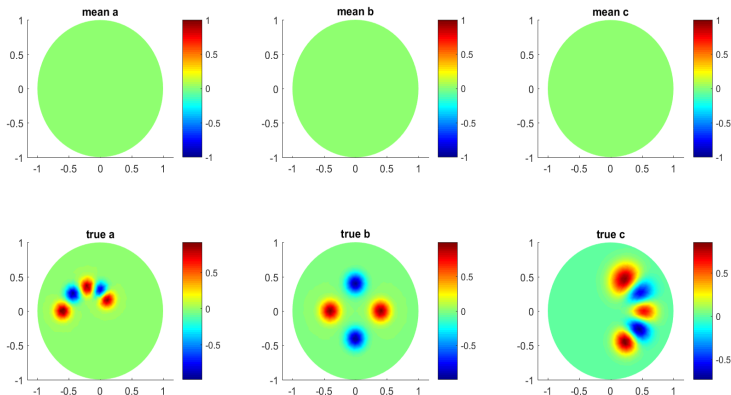
Figure: Top: the components (B_1, B_2, B_3) of the magnetic field Φ . Bottom: real (left 2×2 block) and imaginary (right 2×2 block) parts of the scattering data C_Φ in $\mathfrak{su}(2)$.

Implementation for neutron spin tomography

We use MCMC averages of the pCN algorithm to approximate the posterior mean

$$(\bar{B}_1, \bar{B}_2, \bar{B}_3) \approx \bar{\Phi} = \int_{\mathbb{H}_1} \Phi d\Pi(\Phi | (Y_i, x_i)_{i=1}^N).$$

parameters: 886 unknowns, $n=5000$, $\delta = 0.00025$, $\sigma = 1$, $\nu = 3$, $\text{ell} = 0.15$, $\epsilon = 0.03$

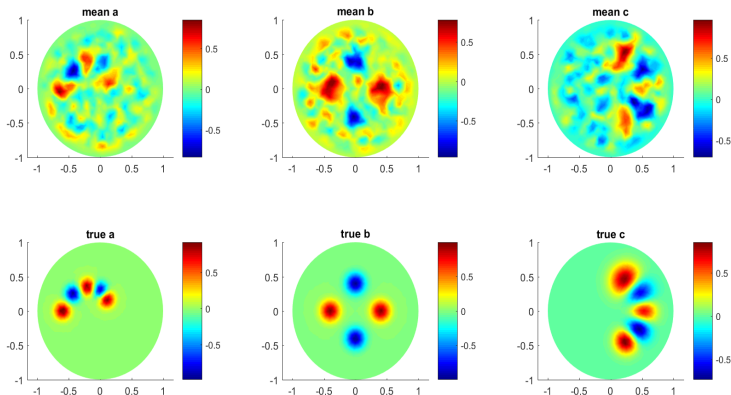


Implementation for neutron spin tomography

We use MCMC averages of the pCN algorithm to approximate the posterior mean

$$(\bar{B}_1, \bar{B}_2, \bar{B}_3) \approx \bar{\Phi} = \int_{\mathbb{H}_1} \Phi d\Pi(\Phi | (Y_i, x_i)_{i=1}^N).$$

iterations : 20000 -- acc. rate : 0.1882

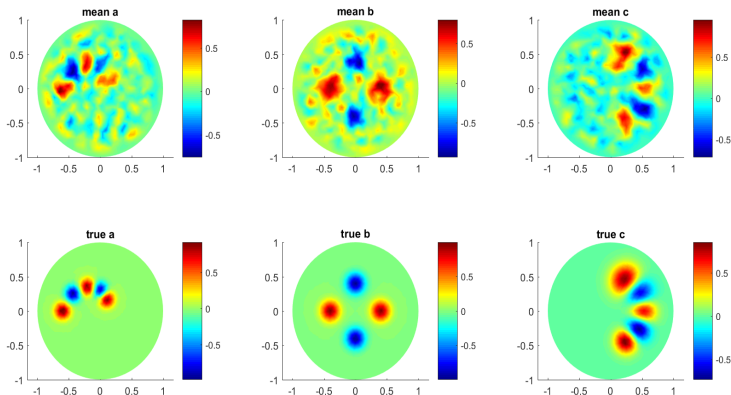


Implementation for neutron spin tomography

We use MCMC averages of the pCN algorithm to approximate the posterior mean

$$(\bar{B}_1, \bar{B}_2, \bar{B}_3) \approx \bar{\Phi} = \int_{\mathbb{H}_1} \Phi d\Pi(\Phi | (Y_i, x_i)_{i=1}^N).$$

iterations : 50000 -- acc. rate : 0.18686



Implementation for neutron spin tomography

We use MCMC averages of the pCN algorithm to approximate the posterior mean

$$(\bar{B}_1, \bar{B}_2, \bar{B}_3) \approx \bar{\Phi} = \int_{\mathbb{H}_1} \Phi d\Pi(\Phi | (Y_i, x_i)_{i=1}^N).$$

iterations : 99500 -- acc. rate : 0.1894

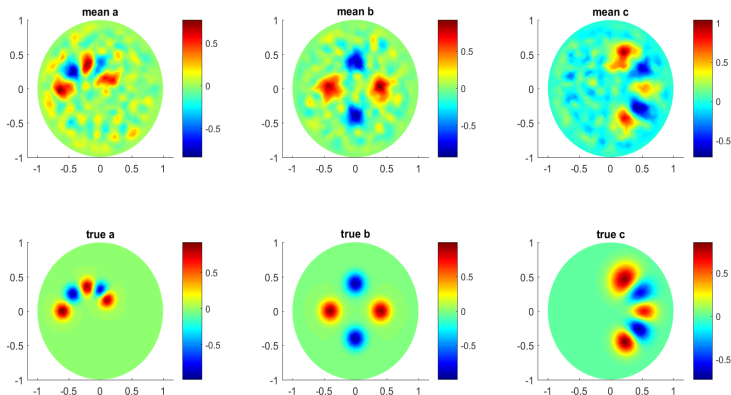
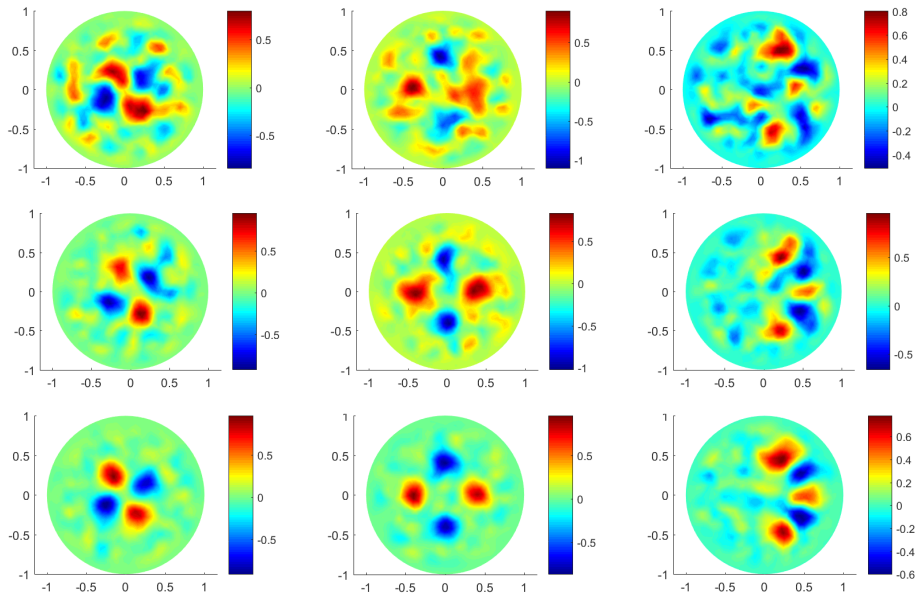


Illustration of consistency in PNT: $N = 200, 400, 800$



III. STATISTICAL THEORY FOR BAYESIAN INVERSION

Consistency of Bayesian Polarised Neutron Tomography

- Let a prior Π for Φ arise from i.i.d. copies B_1, B_2, B_3 drawn from the law of

$$B = \frac{\zeta f'}{\sqrt{N^{1/(\alpha+1)}}}, \quad f' \sim \Pi'_{\alpha,\beta},$$

where $\zeta \in C_c^\infty(M)$ is s.t. $\zeta = 1$ on a compact subset M_0 of the interior of M , and where $\Pi'_{\alpha,\beta}$ is a Matérn prior.

Consistency of Bayesian Polarised Neutron Tomography

- Let a prior Π for Φ arise from i.i.d. copies B_1, B_2, B_3 drawn from the law of

$$B = \frac{\zeta f'}{\sqrt{N^{1/(\alpha+1)}}}, \quad f' \sim \Pi'_{\alpha,\beta},$$

where $\zeta \in C_c^\infty(M)$ is s.t. $\zeta = 1$ on a compact subset M_0 of the interior of M , and where $\Pi'_{\alpha,\beta}$ is a Matérn prior.

Theorem [Monard, N and Paternain (2019)]

Let $\Pi(\cdot | (Y_i, (X_i, V_i))_{i=1}^N)$ be the posterior distribution for Φ arising from the above prior Π with $\alpha > \beta > 2$ and observations $(Y_i, (X_i, V_i))_{i=1}^N \sim P_\Phi^N$. Denote by

$$\bar{\Phi} = E[\Phi | (Y_i, (X_i, V_i))_{i=1}^N]$$

the corresponding posterior mean. If $\Phi_0 \in C^\alpha(M)$ is supported in M_0 then

$$\|\bar{\Phi} - \Phi_0\|_{L^2(M)} = O_{P_{\Phi_0}^N}(N^{-\eta}) \text{ as } N \rightarrow \infty.$$

- As $\alpha, \beta \rightarrow \infty$ we can take $\eta \rightarrow 1/2$.

Statistical solution of the Caldéron problem

- Next let $\mathcal{O}$ be a bounded domain in $\mathbb{R}^d$, $d \geq 3$. Consider ‘Neumann’ data

$$Y = \left\{ \frac{\partial u_{f_1}(g)}{\partial \nu} \Big|_{\partial \mathcal{O}} : g \in H^r(\partial \mathcal{O}) \right\} + \frac{1}{\sqrt{N}} \mathbb{W}, \quad Y \sim P_f^Y,$$

of the elliptic PDE from earlier, where $\mathbb{W}$ is a Gaussian white noise in an appropriate space of Hilbert-Schmidt operators, $r \in \mathbb{R}$.

Statistical solution of the Caldéron problem

- Next let $\mathcal{O}$ be a bounded domain in $\mathbb{R}^d$, $d \geq 3$. Consider ‘Neumann’ data

$$Y = \left\{ \frac{\partial u_{f_1}(g)}{\partial \nu} \Big|_{\partial \mathcal{O}} : g \in H^r(\partial \mathcal{O}) \right\} + \frac{1}{\sqrt{N}} \mathbb{W}, \quad Y \sim P_f^Y,$$

of the elliptic PDE from earlier, where $\mathbb{W}$ is a Gaussian white noise in an appropriate space of Hilbert-Schmidt operators, $r \in \mathbb{R}$.

Theorem [Abraham and N (2019)]

For suitable Gaussian (e.g., Matérn) priors Π , let $\bar{f} = E^\Pi[f|Y]$ be the posterior mean arising from observations $Y \sim P_f^Y$. Assume $f_0 \in H^\alpha(\mathcal{O})$, $\alpha > 3 + d$, satisfies $f_0 > f_{\min} > 0$ and that $f_0 = 1$ in a neighbourhood of $\partial \mathcal{O}$. Then as $N \rightarrow \infty$,

$$\|\bar{f} - f_0\|_{L^\infty(\mathcal{O})} = O_{P_{f_0}^Y}((\log N)^{-\delta}), \quad \delta > 0.$$

- A statistical minimax lower bound with a different exponent δ can also be proved, showing that the Caldéron problem is information-theoretically ‘harder’ than PNT (and other inverse problems).

The Annals of Statistics

2008, Vol. 36, No. 3, 1435–1463

DOI: 10.1214/009053607000000613

© Institute of Mathematical Statistics, 2008

RATES OF CONTRACTION OF POSTERIOR DISTRIBUTIONS BASED ON GAUSSIAN PROCESS PRIORS

BY A. W. VAN DER VAART AND J. H. VAN ZANTEN¹

Vrije Universiteit Amsterdam

We derive rates of contraction of posterior distributions on nonparametric or semiparametric models based on Gaussian processes. The rate of contraction is shown to depend on the position of the true parameter relative to the reproducing kernel Hilbert space of the Gaussian process and the small ball probabilities of the Gaussian process. We determine these quantities for a range of examples of Gaussian priors and in several statistical settings. For instance, we consider the rate of contraction of the posterior distribution based on sampling from a smooth density model when the prior models the log density as a (fractionally integrated) Brownian motion. We also consider regression with Gaussian errors and smooth classification under a logistic or probit link function combined with various priors.

Proof ideas for PNT: Hellinger stability

- The proof proceeds by first showing a contraction rate

$$\Pi(\Phi : h(C_\Phi, C_{\Phi_0}) \geq \delta_N | Y) \xrightarrow{P_{\Phi_0}^Y} 0$$

in the [Hellinger distance \$h\$ induced by the statistical model](#), adapting ideas from Bayesian Nonparametrics (van der Vaart & van Zanten (2008)).

Proof ideas for PNT: Hellinger stability

- The proof proceeds by first showing a contraction rate

$$\Pi(\Phi : h(C_\Phi, C_{\Phi_0}) \geq \delta_N | Y) \xrightarrow{P_{\Phi_0}^Y} 0$$

in the **Hellinger distance h induced by the statistical model**, adapting ideas from Bayesian Nonparametrics (van der Vaart & van Zanten (2008)).

- For our measurement model $h(C_\Phi, C_\Psi) \simeq \|C_\Phi - C_\Psi\|_{L^2(\partial_+ SM)}$ can be shown.

Proof ideas for PNT: Hellinger stability

- The proof proceeds by first showing a contraction rate

$$\mathbb{P}(\Phi : h(C_\Phi, C_{\Phi_0}) \geq \delta_N | Y) \xrightarrow{P_{\Phi_0}^Y} 0$$

in the **Hellinger distance h induced by the statistical model**, adapting ideas from Bayesian Nonparametrics (van der Vaart & van Zanten (2008)).

- For our measurement model $h(C_\Phi, C_\Psi) \simeq \|C_\Phi - C_\Psi\|_{L^2(\partial_+ SM)}$ can be shown.
- Then we require a quantitative version of the injectivity result:

Theorem [Monard, N and Paternain (2019)]

For M a simple surface, let $\Phi, \Psi \in L^2(M, \mathfrak{so}(n))$ have non-Abelian Radon transforms C_Φ, C_Ψ in $L^2(\partial_+ SM, SO(n))$. Then

$$\|\Phi - \Psi\|_{L^2} \leq c_{\Phi, \Psi} \|C_\Phi - C_\Psi\|_{H^1}$$

where, for C_1 a numerical constant,

$$c_{\Phi, \Psi} = C_1(1 + (|\Phi|_{C^1} \vee |\Psi|_{C^1}))^{1/2} e^{C_2(|\Phi|_{C^1} \vee |\Psi|_{C^1})} (1 + \|\Psi\|_{C^1}).$$

Continuation of the proof

- A forward estimate

$$\|C_\Phi - C_\Psi\|_{H^\beta} \leq (1 + \|\Phi\|_{C^\beta}^\beta + \|\Psi\|_{C^\beta}^\beta) \|\Phi - \Psi\|_{H^\beta}, \quad \beta \in \mathbb{N},$$

can also be shown.

Continuation of the proof

- A forward estimate

$$\|C_\Phi - C_\Psi\|_{H^\beta} \leq (1 + \|\Phi\|_{C^\beta}^\beta + \|\Psi\|_{C^\beta}^\beta) \|\Phi - \Psi\|_{H^\beta}, \quad \beta \in \mathbb{N},$$

can also be shown.

- Then, we can use an interpolation inequality

$$\begin{aligned} \|\Phi - \Phi_0\|_{L^2} &\leq c_{\Phi, \Phi_0} \|C_\Phi - C_{\Phi_0}\|_{H^1} \\ &\lesssim c_{\Phi, \Phi_0} \|C_\Phi - C_{\Phi_0}\|_{L^2}^{(\beta-1)/\beta} \|C_\Phi - C_{\Phi_0}\|_{H^\beta}^{1/\beta}, \quad 0 \leq 1 \leq \beta, \end{aligned}$$

and a bound for the Sobolev regularity of Φ , C_Φ when $\Phi \sim \Pi(\cdot|Y)$. Key techniques involve a careful use of fine properties of [Gaussian measures in infinite dimensions](#).

Continuation of the proof

- A forward estimate

$$\|C_\Phi - C_\Psi\|_{H^\beta} \leq (1 + \|\Phi\|_{C^\beta}^\beta + \|\Psi\|_{C^\beta}^\beta) \|\Phi - \Psi\|_{H^\beta}, \quad \beta \in \mathbb{N},$$

can also be shown.

- Then, we can use an interpolation inequality

$$\begin{aligned} \|\Phi - \Phi_0\|_{L^2} &\leq c_{\Phi, \Phi_0} \|C_\Phi - C_{\Phi_0}\|_{H^1} \\ &\lesssim c_{\Phi, \Phi_0} \|C_\Phi - C_{\Phi_0}\|_{L^2}^{(\beta-1)/\beta} \|C_\Phi - C_{\Phi_0}\|_{H^\beta}^{1/\beta}, \quad 0 \leq 1 \leq \beta, \end{aligned}$$

and a bound for the Sobolev regularity of Φ , C_Φ when $\Phi \sim \Pi(\cdot|Y)$. Key techniques involve a careful use of fine properties of [Gaussian measures in infinite dimensions](#).

- From what precedes we can derive a posterior contraction rate

$$\Pi(f : \|\Phi - \Phi_0\|_{L^2} \geq \delta_N^{(\beta-1)/\beta} | Y) \xrightarrow{P_{f_0}^Y} 0.$$

Combining [tight exponential bounds](#) $\exp\{-CN\varepsilon_N^2\}$ for the last limit with a [quantitative uniform integrability argument](#) we deduce $\|\bar{f} - f_0\|_{L^2} = O_P(\varepsilon_N)$.

Consistency of Gaussian priors in non-linear problems:

F. Monard, R. Nickl, G.P. Paternain, Consistent inversion of noisy non-Abelian X -ray transforms, arXiv:1905.00860

K. Abraham, R. Nickl, On statistical Caldéron problems. arXiv:1906.03486