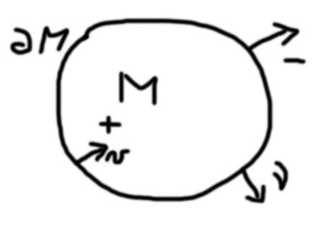
Consistent inversion of noisy non-abelian X-ray transforms

Gabriel P. Paternain

IAS Workshop on IP, Imaging and PDEs, May 23rd 2019
Joint work with François Monard and Richard Nickl

Setting

- (M, g) is a compact Riemannian surface with boundary ∂M.
- $SM = \{(x, v) \in TM : \ |v| = 1\}$ is the unit sphere bundle with boundary $\partial(SM)$.
- $\partial_{\pm}(SM) = \{(x, v) \in \partial(SM) \ : \ \pm\langle v, \nu\rangle \leq 0\}$, where ν is the the outer unit normal vector.
- We will assume ∂M is strictly convex (positive definite second fundamental form).



We let $\tau(x, v)$ be the first time when a geodesic starting at (x, v) leaves M.

Definition. We say (M, g) is non-trapping if $\tau(x, v) < \infty$ for all $(x, v) \in SM$.

We will assume that our surface is **simple**: there is non-trapping and there no conjugate points.

Examples: Strictly convex domains in the plane and small C^2 perturbations of them.

Non-abelian X-ray

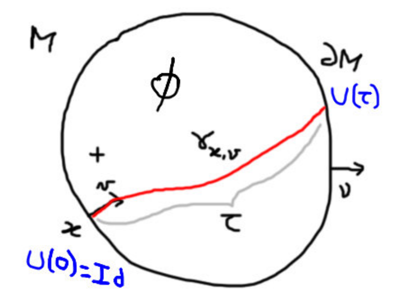
Let $\Phi \in C_c(M, \mathbb{C}^{n\times n})$ be a matrix field.

Given a unit-speed geodesic $\gamma : [0, \tau] \to M$ with endpoints $\gamma(0), \gamma(\tau) \in \partial M$, we consider the matrix ODE

$$\dot{U} + \Phi(\gamma(t))U = 0, \qquad U(0) = \mathrm{Id}.$$

We define the scattering data of Φ on γ to be $C_\Phi(\gamma) := U(\tau)$.

When Φ is scalar, we obtain $\log U(\tau) = -\int_0^\tau \Phi(\gamma(t))\, dt$, the classical X-ray/Radon transform of Φ along the curve γ.



- The collection of all such data makes up the *scattering data* or *non-Abelian X-ray transform* of Φ, viewed as a map

$$C_\Phi : \partial_+ SM \to GL(n, \mathbb{C}).$$

- **Inverse Problem:** recover Φ from C_Φ.

Injectivity

The state of the art on injectivity is:

Theorem 1 (P-Salo-Uhlmann 2012, P-Salo 2018)
Let (M, g) be a simple surface. The map $\Phi \mapsto C_\Phi$ is injective in the following cases:
(a) $\Phi : M \to \mathfrak{u}(n)$, where $\mathfrak{u}(n)$ in the set of skew-hermitian matrices (Lie algebra of $U(n)$).
(b) M has negative curvature.

Early work on this problem for Euclidean domains by Vertgeim (1992), R. Novikov (2002) and G. Eskin (2004).

Polarimetric Neutron Tomography (PNT)

The non-abelian X-ray transform arises naturally when trying to reconstruct a magnetic field from spin measurements of neutrons.

In this case

$$\Phi(x) = \begin{bmatrix} 0 & B_3 & -B_2 \\ -B_3 & 0 & B_1 \\ B_2 & -B_1 & 0 \end{bmatrix} \in \mathfrak{so}(3)$$

where $B(x) = (B_1, B_2, B_3)$ is the magnetic field.

The scatteting data takes values $C_\Phi : \partial_+ SM \to SO(3)$.

Cf. [Desai, Lionheart et al., Nature Sc. Rep. 2018] and [Hilger et al., Nature Comm. 2018].

The experiment

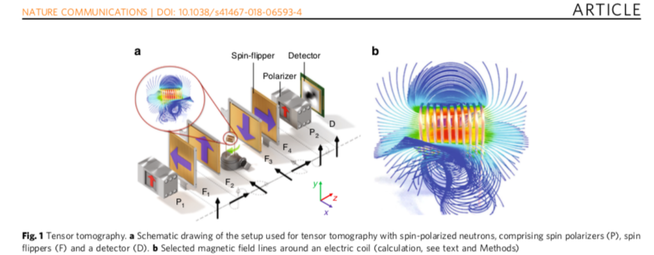


Fig. 1 Tensor tomography. **a** Schematic drawing of the setup used for tensor tomography with spin-polarized neutrons, comprising spin polarizers (P), spin flippers (F) and a detector (D). **b** Selected magnetic field lines around an electric coil (calculation, see text and Methods)

From Hilger et al., Nature Comm. 2018.

- Data produced: $C_\Phi(x, v) \in SO(3)$.
- This is done with an ingenious sequence of spin flippers and rotators placed before and after the magnetic field being measured.
- The material containing the magnetic field can also be rotated so as to produce parallel beams from different angles.

But we face the usual problems:

- No explicit reconstruction formula.
- Measurements are noisy.

Thus we have observations $(X_i, V_i) \in \partial_+ SM$ and

$$Y_i = C_\Phi(X_i, V_i) + \varepsilon_i, \quad 1 \le i \le N, \quad (\varepsilon_i)_{jk} \sim^{\text{i.i.d.}} \mathcal{N}(0, \sigma^2).$$

We will assume $(X_i, V_i) \sim^{\text{i.i.d}} \lambda$, where λ is the probability measure given by the standard area form of $\partial_+ SM$ (independent of ε_i).

We let P_Φ^N be the joint probability law of $(Y_i, (X_i, V_i))_{i=1}^N$.

Bayesian numerics magic

First a word from a magician (1988 paper):

BAYESIAN NUMERICAL ANALYSIS

PERSI DIACONIS

Department of Statistics
Stanford University
Stanford, California 94305, U.S.A.

1. INTRODUCTION

Consider a given function $f\colon [0,1] \to \mathrm{R}$ such as

$$f(x) = \exp\left\{\cosh\left(\frac{x + 2x^2 + \cos x}{3 + \sin x^3}\right)\right\}. \tag{1}$$

If you require $\int_0^1 f(x)dx$, a formula such as (1) isn't of much use and leads to questions like "What does it mean to 'know' a function?" The formula says some things (e.g. f is smooth, positive, and bounded by 20 on $[0,1]$) but there are many other facts about f that we don't know (e.g., is f monotone, unimodal, or convex?).

Once we allow that we don't know f, but do know some things, it becomes natural to take a Bayesian approach to the quadrature problem:

- Put a prior on continuous functions $C[0,1]$
- Calculate f at $x_1, x_2, \ldots, x_n$
- Compute a posterior
- Estimate $\int_0^1 f$ by the Bayes rule

Most people, even Bayesians, think this sounds crazy when they first hear about it. The following examples may help.

We adopt the same magical approach.

- We put a Gaussian process prior Π on Φ; more details on this later. The use of Gaussian process priors for inverse problems has been advocated by A. Stuart.
- Using the observations we compute the posterior $\Pi(\cdot|(Y_i, (X_i, V_i)_{i=1}^N))$ using Bayes rule;
- From the posterior we extract the mean $\bar{\Phi}_N$. This is a somewhat formidable object given by a Bochner integral

$$\bar{\Phi}_N = \int \Phi \, d\Pi(\Phi|(Y_i, (X_i, V_i)_{i=1}^N)).$$

In more detail:

- We have

$$\Pi(A|(Y_i,(X_i,V_i)_{i=1}^N)) = \frac{\int_A e^{\ell(\Phi)}\,d\Pi(\Phi)}{\int e^{\ell(\Phi)}\,d\Pi(\Phi)},$$

  where the log-likelihood is

$$\ell(\Phi) := -\frac{1}{2\sigma^2}\sum_{i=1}^{N}\|Y_i - C_\Phi(X_i,V_i)\|^2.$$

- And the posterior mean is

$$\bar{\Phi}_N = \frac{\int \Phi e^{\ell(\Phi)}\,d\Pi(\Phi)}{\int e^{\ell(\Phi)}\,d\Pi(\Phi)}.$$

The magician will tell you:

"as $N \to \infty$, $\bar{\Phi}_N$ will approach the true Φ_0 you so much desire to reconstruct; I have performed this trick many times".

Can this magic be debunked? No, this actually works.

Theorem 2 (Version I, Monard-Nickl-P 2019)

The estimator $\bar{\Phi}_N$ is consistent in the sense that in $P^N_{\Phi_0}$-probability

$$\|\bar{\Phi}_N - \Phi_0\|_{L^2} \to 0$$

as the sample size $N \to \infty$.

Assumptions on the prior:

Let $\alpha > \beta > 2$. The prior Π is a centred Gaussian Borel probability measure on the Banach space $C(M)$ that is supported in a separable linear subspace of $C^\beta(M)$, and assume its RKHS $(\mathcal{H}, \|\cdot\|_{\mathcal{H}})$ is continuously imbedded into the Sobolev space $H^\alpha(M)$.

<u>An example</u>:

Consider a Matérn kernel $k : \mathbb{R}^2 \to \mathbb{R}$ with associated (centered) Gaussian process G with covariance $E[G(x)G(y)] = k(x-y)$, $x, y \in \mathbb{R}^2$.

Explicitly

$$k(r) = \frac{2^{1-\nu}}{\Gamma(\nu)} \left(\frac{\sqrt{2\nu} r}{\ell} \right)^{\nu} K_{\nu}(\sqrt{2\nu} r / \ell),$$

where K_ν is a modified Bessel function and $r = |x - y|$. The parameter ν controls the Sobolev regularity.

Consider $M \subset \mathbb{R}^2$ and restrict the process to M to obtain a prior Π satisfying the required conditions as long as $\alpha = \nu > \beta + 1 > 3$. For this process $\mathcal{H} = H^{\alpha}(M)$.

This assumption on the prior describes a *very flexible* class.

Note: we put independent scalar valued processes on each entry of Φ.

Consistency: full version

There is one further trick that has to be performed on the prior before we can state in detail the consistency theorem.

Given Π as above, we "temper" it by introducing a new prior Π_{temp} by setting

$$\Pi_{temp}(A) := \Pi(\psi A/\sqrt{N^{1/(\alpha+1)}})$$

where A is a Borel subset of $C(M)$ and ψ is a cut-off function which equals 1 on $M_0 \subset M$ and is compactly supported in M^{int}.

Theorem 3 (Full Version, Monard-Nickl-P 2019)

With Π_{temp} as above, assume Φ_0 belongs to $\mathcal{H}$ and is supported in M_0. Then we have, for some $\eta > 0$

$$P^N_{\Phi_0}\left(\|\bar{\Phi}_N - \Phi_0\|_{L^2(M)} > N^{-\eta}\right) \to 0 \ \text{ as } N \to \infty.$$

Ingredients for the proof of consistency

- We show first that the Bayesian algorithm recovers the "regression function" C_Φ consistently in a natural statistical distance function. This uses ideas from Bayesian nonparametrics (van der Vaart and van Zanten, 2008).
- This statistical distance function is equivalent to the L^2-distance in our case, since C_Φ takes values in a compact Lie group.
- We combine this with a new quantitative version of the injectivity result of [P-Salo-Uhlmann 2012] (stability estimate).
- This blending requires a careful use of fine properties of Gaussian measures in infinite dimensions.

The new stability estimate can be stated as follows:

Theorem 4 (Monard-Nickl-P 2019)

Let (M, g) be a simple surface. Given two matrix fields Φ and Ψ in $C^1_c(M, \mathfrak{u}(n))$ we have

$$\|\Phi - \Psi\|_{L^2(M)} \leq c(\Phi, \Psi)\|C_\Phi C_\Psi^{-1} - \mathrm{Id}\|_{H^1(\partial_+ SM)},$$

where

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

and the constants C_1, C_2 only depend on (M, g).

Relation between linear and non-linear

Pseudo-linearization identity (cf. Stefanov-Uhlmann 1998 for lens rigidity) :

$$C_\Phi^{-1} C_\Psi = Id + I_{\Theta(\Phi,\Psi)}(\Psi - \Phi),$$

where $I_{\Theta(\Phi,\Psi)}$ is an attenuated X-ray transform with matrix attenuation $\Theta(\Phi,\Psi)$, an endomorphism on $\mathbb{C}^{n\times n}$ with pointwise action

$$\Theta(\Phi,\Psi)\cdot U = \Phi U - U\Psi, \qquad U \in \mathbb{C}^{n\times n}.$$

Thus the proof is reduced to a stability estimate for an attenuated X-ray transform where the weight depends on Φ and Ψ. This uses scalar holomorphic integrating factors, whose existence is guaranteed by the surjectivity of I_0^* (Pestov-Uhlmann 2005).

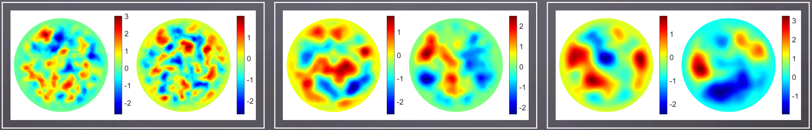
Implementation

We use MCMC averages of the pre-conditioned Crank-Nicholson algorithm to approximate the posterior mean.

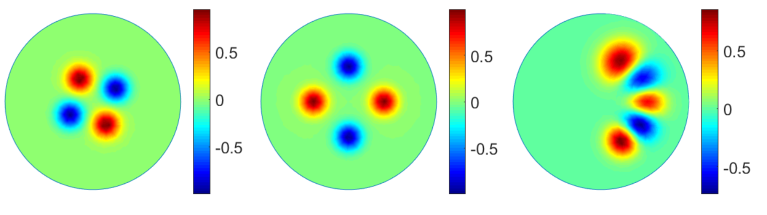
Hairer, Stuart, Vollmer (2014) proved dimension-free spectral gaps for the chain, so we have very good mixing properties towards the posterior.

We use a Matérn kernel as described before for $\nu = 3$.

Various parameters need to be fine-tuned.

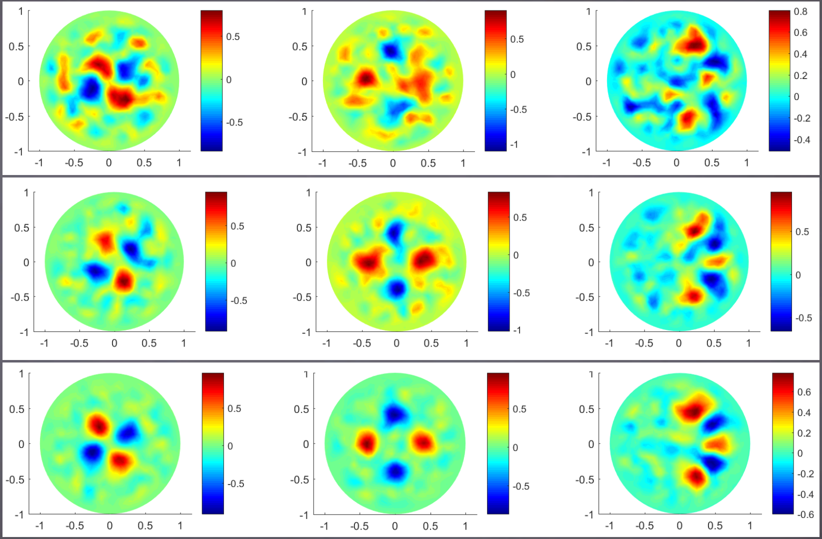


Left to right: two Matérn prior samples with $\ell = 0.1$, 0.2 and 0.3.



This is the true field Φ_0.

We generate synthetic data C_{Φ_0} from Φ_0 and then we add noise.



Top to bottom: The posterior mean field for sample sizes $N = 200, 400, 800$. The number of Monte-Carlo iterations is 100000.

Main message

- The consistency theorem is a potent tranquilizer if you suffer anxiety about Bayesian approaches to inverse problems. You can now relax and use the pCN algorithm with confidence.
- The ultimate tranquilizer is a Bernstein-von-Mises theorem which describes a dream scenario for the posterior as the sample limit $N \to \infty$.
- This is within reach for PNT. It requires a fine understanding of the inverse Fisher information operator, something of independent interest eventually leading to a complete understanding of boundary behaviour.
- There is a beautiful interplay here between problems motiviated by statistical thinking and geometric inverse problems. Lots more to be done!

pCN algorithm

We use the preconditioned Crank-Nicholson (pCN) method to sample from the posterior distribution.

Recall that the log-likelihood function given the data $(Y_i, (X_i, V_i))_{i=1}^N$ is

$$\ell(\Phi) = -\frac{1}{2\sigma^2}\sum_{i=1}^{N}\|Y_i - C_\Phi(X_i, V_i)\|^2.$$

We approximate the posterior mean by a Monte Carlo average $\widehat{\Phi} = \frac{1}{N_s}\sum_{n=0}^{N_s}\Phi_n$ of a Markov chain (Φ_n) of length N_s.

Let Π be a Gaussian prior for Φ; initialise $\Phi_n = 0$ for $n = 0$, then repeat:

1. Draw $\Psi \sim \Pi$ and for $\delta > 0$ define the proposal $p_{\Phi_n} := \sqrt{1-2\delta}\ \Phi_n + \sqrt{2\delta}\ \Psi$.
2. Set

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

The algorithm is terminated at $n = N_s$ and requires evaluation of $\ell(\Phi_n)$ and thus of the scattering data $C_{\Phi_n}(X_i, V_i)$ for every Φ_n and (X_i, V_i)

$\Phi \mapsto C_\Phi$ is non-linear and non convex so optimization methods are challenging.