Consistency of Bayesian inference with Gaussian priors in an elliptic nonlinear inverse problem

Matteo Giordano

Joint work with Richard Nickl

Cambridge Centre for Analysis, University of Cambridge

8th European Congress of Mathematics, 21st June 2021

Statistical inverse problems

- A statistical inverse problem is the task of recovering an unknown object f from noisy indirect measurements

$$Y = \mathscr{G}f + W$$

where $\mathscr{G} : f \mapsto \mathscr{G}f$ is a 'forward' map and W is random noise.

Statistical inverse problems

- A statistical inverse problem is the task of recovering an unknown object f from noisy indirect measurements

$$Y = \mathscr{G} f + W$$

where $\mathscr{G} : f \mapsto \mathscr{G} f$ is a 'forward' map and W is random noise.

- Given a bounded smooth domain $\mathcal{O} \subset \mathbb{R}^d$, consider parameters $f \in \mathcal{F} \subset L^2(\mathcal{O})$, a measurable map $\mathscr{G} : \mathcal{F} \to L^2(\mathcal{O})$, and observations

$$Y \equiv (Y_i)_{1 \le i \le N}, \qquad Y_i = \mathscr{G} f(x_i) + W_i, \quad W_i \overset{\text{iid}}{\sim} N(0,1)$$

over a grid $\{x_1, \dots, x_N\} \subset \mathcal{O}$.

Statistical inverse problems

- A statistical inverse problem is the task of recovering an unknown object f from noisy indirect measurements

$$Y = \mathscr{G}f + W$$

where $\mathscr{G} : f \mapsto \mathscr{G}f$ is a 'forward' map and W is random noise.

- Given a bounded smooth domain $\mathcal{O} \subset \mathbb{R}^d$, consider parameters $f \in \mathcal{F} \subset L^2(\mathcal{O})$, a measurable map $\mathscr{G} : \mathcal{F} \to L^2(\mathcal{O})$, and observations

$$Y \equiv (Y_i)_{1 \leq i \leq N}, \qquad Y_i = \mathscr{G}f(x_i) + W_i, \quad W_i \overset{\text{iid}}{\sim} N(0,1)$$

over a grid $\{x_1, \dots, x_N\} \subset \mathcal{O}$.

- In many applications, f is a parameter governing a PDE and $\mathscr{G}f$ is the PDE solution. Application areas include medical imaging [Bertero and Piana, Springer 2006], geophysics [Snieder and Trampert, Springer 1999], acoustics [Collins and Kuperman, Inv. Probl. 1994], finance [Baumeister, Rec. Dev. Comp. Fin. 2013] and many others.

Non-convex estimation problems

- For observations $Y_i = \mathscr{G}f(x_i) + W_i$, the log-likelihood equals (up to an additive constant) the negative least squares criterion

$$\ell(f) = -\frac{1}{2}\sum_{i=1}^{N} |Y_i - \mathscr{G}f(x_i)|^2, \qquad f \in \mathcal{F}.$$

Non-convex estimation problems

- For observations $Y_i = \mathscr{G}f(x_i) + W_i$, the log-likelihood equals (up to an additive constant) the negative least squares criterion

$$\ell(f) = -\frac{1}{2}\sum_{i=1}^{N} |Y_i - \mathscr{G}f(x_i)|^2, \qquad f \in \mathcal{F}.$$

- In many inverse problems arising from PDEs, the solution map $\mathscr{G} : f \mapsto \mathscr{G}f$ is nonlinear.

Non-convex estimation problems

- For observations $Y_i = \mathscr{G} f(x_i) + W_i$, the log-likelihood equals (up to an additive constant) the negative least squares criterion

$$\ell(f) = -\frac{1}{2} \sum_{i=1}^{N} |Y_i - \mathscr{G} f(x_i)|^2, \qquad f \in \mathcal{F}.$$

- In many inverse problems arising from PDEs, the solution map $\mathscr{G} : f \mapsto \mathscr{G} f$ is nonlinear.
- Hence, penalised least squares objective functionals are in general non-convex, and 'off-the-shelf' optimisation methods are not guaranteed to converge to global minima.

Bayesian approach with Gaussian priors

- Following [Stuart, Acta Numer. 2010], Bayesian models with Gaussian priors Π for f are a popular framework for inference in nonlinear inverse problems.

Bayesian approach with Gaussian priors

- Following [Stuart, Acta Numer. 2010], Bayesian models with Gaussian priors Π for f are a popular framework for inference in nonlinear inverse problems.
- The posterior distribution $\Pi(\cdot|Y)$ provides a principled approach to (non optimisation-based) estimation, via the posterior mean

$$\bar{f} = E^{\Pi}[f|Y] = \int_{\mathcal{F}} f d\Pi(f|Y),$$

and to uncertainty quantification, via posterior credible sets.

Bayesian approach with Gaussian priors

- Following [Stuart, Acta Numer. 2010], Bayesian models with Gaussian priors Π for f are a popular framework for inference in nonlinear inverse problems.
- The posterior distribution $\Pi(\cdot|Y)$ provides a principled approach to (non optimisation-based) estimation, via the posterior mean

$$\bar{f} = E^{\Pi}[f|Y] = \int_{\mathcal{F}} f d\Pi(f|Y),$$

  and to uncertainty quantification, via posterior credible sets.
- Modern MCMC methods (such as the pCN algorithm) can be used to sample from $\Pi(\cdot|Y)$ and reliably compute $\bar{f}$ and credible sets [Cotter et al., Stat. Sci. 2013; Hairer, Stuart and Vollmer, Ann. Appl. Prob. 2014].

Statistical guarantees

- Only few results are available to provide statistical guarantees on the performance of Bayesian inversion in nonlinear inverse problems [Monard, Nickl and Paternain, Comm. Pure Appl. Math. 2020; Abraham and Nickl, Math. Stat. Learn. 2019].

Statistical guarantees

- Only few results are available to provide statistical guarantees on the performance of Bayesian inversion in nonlinear inverse problems [Monard, Nickl and Paternain, Comm. Pure Appl. Math. 2020; Abraham and Nickl, Math. Stat. Learn. 2019].
- Assuming that $Y_i = \mathscr{G} f_0(x_i) + W_i$ for some fixed true f_0, we investigate posterior contraction and the consistency of the posterior mean, studying if (and at what rate) as $N \to \infty$,

$$\Pi(\cdot|Y) \to \delta_{f_0}, \qquad \bar{f} \to f_0.$$

Elliptic inverse problems

- For $f \in C^1(\mathcal{O}) : \inf_{x\in\mathcal{O}} f(x) > 0$, and known $g \in C^\infty(\mathcal{O})$, consider the elliptic boundary value problem

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

By elliptic theory, there exists a unique classical solution $u_f \in C^2(\mathcal{O})$. The forward map is the solution map $\mathscr{G} : f \mapsto \mathscr{G}f \equiv u_f$.

Elliptic inverse problems

- For $f \in C^1(\mathcal{O}) : \inf_{x\in\mathcal{O}} f(x) > 0$, and known $g \in C^\infty(\mathcal{O})$, consider the elliptic boundary value problem

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

  By elliptic theory, there exists a unique classical solution $u_f \in C^2(\mathcal{O})$. The forward map is the solution map $\mathscr{G} : f \mapsto \mathscr{G}f \equiv u_f$.

- We address the problem of estimating f from observations $(Y, X) \equiv (Y_i, X_i)_{1\le i\le N} \sim P_f^N$, where

$$Y_i = \mathscr{G}f(X_i) + W_i, \qquad W_i \overset{\text{iid}}{\sim} N(0,1), \quad X_i \overset{\text{iid}}{\sim} \mathrm{Un}(\mathcal{O}).$$

Elliptic inverse problems

- For $f \in C^1(\mathcal{O}) : \inf_{x\in\mathcal{O}} f(x) > 0$, and known $g \in C^\infty(\mathcal{O})$, consider the elliptic boundary value problem

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

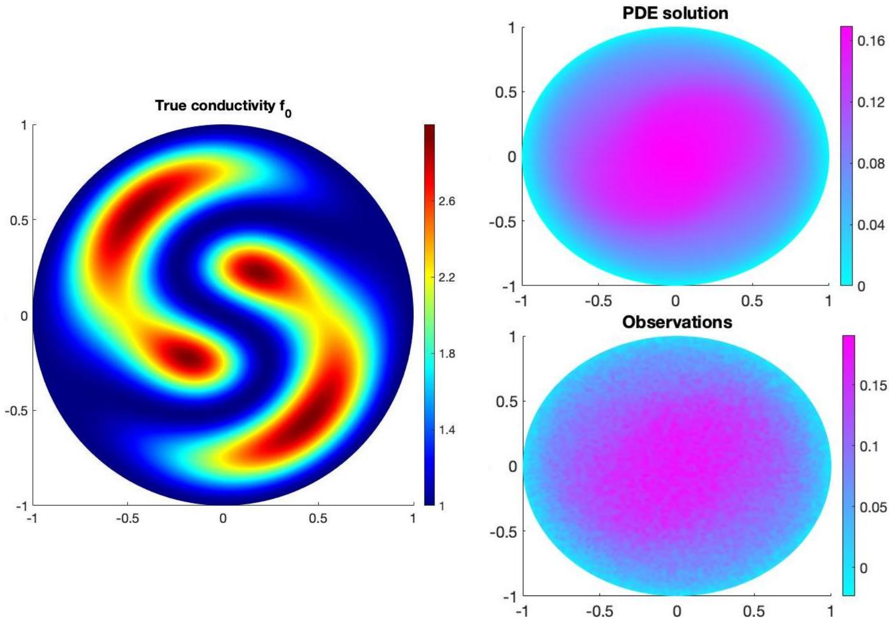
  By elliptic theory, there exists a unique classical solution $u_f \in C^2(\mathcal{O})$. The forward map is the solution map $\mathscr{G} : f \mapsto \mathscr{G} f \equiv u_f$.

- We address the problem of estimating f from observations $(Y, X) \equiv (Y_i, X_i)_{1\leq i\leq N} \sim P_f^N$, where

$$Y_i = \mathscr{G} f(X_i) + W_i, \qquad W_i \overset{\text{iid}}{\sim} N(0, 1), \quad X_i \overset{\text{iid}}{\sim} \mathrm{Un}(\mathcal{O}).$$

- Standard benchmark example for evaluating the performance of recovery algorithms in statistics [Bissantz et al., Inv. Probl. 2004; Dashti and Stuart, SIAM J. Numer. Anal. 2011; Cotter et al., Stat. Sc. 2013; . . .] and applied mathematics [Richter, SIAM J. Appl. Math. 1981; Knowles, J. Comput. Appl. Math. 2001; Bonito et al., SIAM J. Math. Anal. 2017; . . .].

Elliptic inverse problems



Scaled Gaussian priors

- We assign scaled Gaussian priors of the form

$$\Pi = \mathcal{L}(F), \qquad F = \chi \frac{F'}{\sqrt{N^{1/(2\alpha+2+2d)}}}, \quad F' \sim \Pi'$$

where Π' is a 'standard' α-regular Gaussian prior on $L^2(\mathcal{O})$ (e.g., Matérn or squared exponential), $\alpha > 0$, and $\chi \in C_c^\infty(\mathcal{O})$.

Scaled Gaussian priors

- We assign scaled Gaussian priors of the form

$$\Pi = \mathcal{L}(F), \qquad F = \chi \frac{F'}{\sqrt{N^{1/(2\alpha+2+2d)}}}, \quad F' \sim \Pi'$$

where Π' is a 'standard' α-regular Gaussian prior on $L^2(\mathcal{O})$ (e.g., Matérn or squared exponential), $\alpha > 0$, and $\chi \in C_c^\infty(\mathcal{O})$.

- The scaling enforces additional regularisation onto the posterior draws, ensuring that for $\beta < \alpha$,

$$\Pi(f : \|f\|_{H^\beta} \le M | Y, X) \overset{P_{f_0}}{\to} 1,$$

which is used crucially in the proofs to handle the nonlinearity of $\mathscr{G}$ [Monard, Nickl and Paternain, Comm. Pure Appl. Math. 2020; Abraham and Nickl, Math. Stat. Learn. 2019].

Scaled Gaussian priors

- We assign scaled Gaussian priors of the form

$$\Pi = \mathcal{L}(F), \qquad F = \chi \frac{F'}{\sqrt{N^{1/(2\alpha+2+2d)}}}, \quad F' \sim \Pi'$$

  where Π' is a 'standard' α-regular Gaussian prior on $L^2(\mathcal{O})$ (e.g., Matérn or squared exponential), $\alpha > 0$, and $\chi \in C_c^\infty(\mathcal{O})$.

- The scaling enforces additional regularisation onto the posterior draws, ensuring that for $\beta < \alpha$,

$$\Pi(f : \|f\|_{H^\beta} \le M | Y, X) \overset{P_{f_0}}{\to} 1,$$

  which is used crucially in the proofs to handle the nonlinearity of $\mathscr{G}$ [Monard, Nickl and Paternain, Comm. Pure Appl. Math. 2020; Abraham and Nickl, Math. Stat. Learn. 2019].

- The positivity constraint on f is modelled by composing $f = \Phi \circ F$, where $\Phi : \mathbb{R} \to [\delta, \infty)$, $\delta > 0$, is a suitable smooth 'link' function.

Convergence rates for the posterior mean

Theorem (G., Nickl 2020)

For $\alpha > 2 + d/2$, let $f_0 \in H^\alpha(\mathcal{O})$ satisfy $\inf_{x\in\mathcal{O}} f_0(x) > 0$ and suitable boundary conditions, and consider observations $(X, Y) \sim P^N_{f_0}$. Let Π be the above scaled Gaussian prior, and consider the posterior mean estimator $\bar{f} = \Phi \circ E^\Pi[F|Y, X]$. Then, as $N \to \infty$,

$$\|\bar{f} - f_0\|_{L^2} = O_{P_{f_0}}(N^{-\zeta}), \quad 0 < \zeta < \frac{\alpha - 1}{2\alpha + 2 + d}.$$

Convergence rates for the posterior mean

Theorem (G., Nickl 2020)

For $\alpha > 2 + d/2$, let $f_0 \in H^\alpha(\mathcal{O})$ satisfy $\inf_{x\in\mathcal{O}} f_0(x) > 0$ and suitable boundary conditions, and consider observations $(X, Y) \sim P^N_{f_0}$. Let Π be the above scaled Gaussian prior, and consider the posterior mean estimator $\bar{f} = \Phi \circ E^\Pi[F|Y, X]$. Then, as $N \to \infty$,

$$\|\bar{f} - f_0\|_{L^2} = O_{P_{f_0}}(N^{-\zeta}), \quad 0 < \zeta < \frac{\alpha - 1}{2\alpha + 2 + d}.$$

- Hence, $\bar{f}$ is a consist estimator of f_0, with an algebraic-in-N upper bound on the convergence rate, reliably computable via MCMC.

Convergence rates for the posterior mean

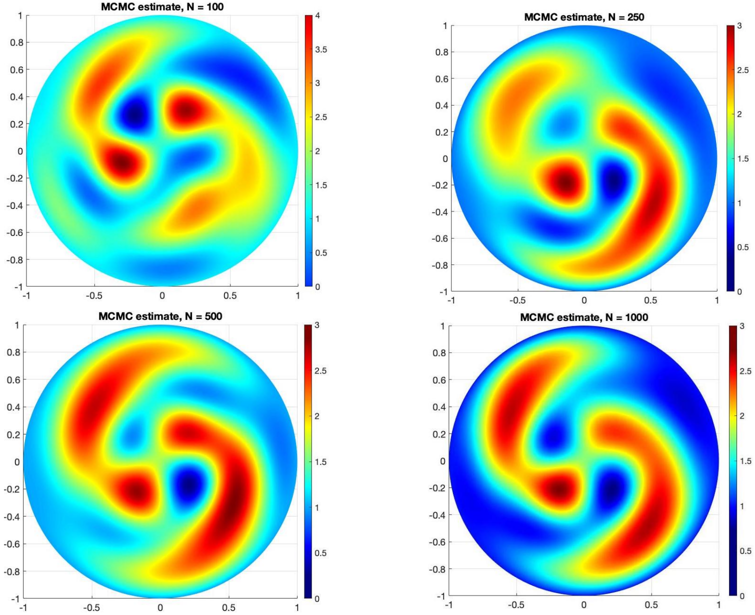
Theorem (G., Nickl 2020)

For $\alpha > 2 + d/2$, let $f_0 \in H^\alpha(\mathcal{O})$ satisfy $\inf_{x\in\mathcal{O}} f_0(x) > 0$ and suitable boundary conditions, and consider observations $(X, Y) \sim P^N_{f_0}$. Let Π be the above scaled Gaussian prior, and consider the posterior mean estimator $\bar{f} = \Phi \circ E^\Pi[F|Y, X]$. Then, as $N \to \infty$,

$$\|\bar{f} - f_0\|_{L^2} = O_{P_{f_0}}(N^{-\zeta}), \quad 0 < \zeta < \frac{\alpha - 1}{2\alpha + 2 + d}.$$

- Hence, $\bar{f}$ is a consist estimator of f_0, with an algebraic-in-N upper bound on the convergence rate, reliably computable via MCMC.
- (Minimax) optimal rates of estimation for f in the elliptic inverse problem are currently unknown. If $f_0 \in C^\infty(\mathcal{O})$, then we can tune the prior Π so that ζ is arbitrarily close to $1/2$, approaching the 'parametric' rate $N^{-1/2}$.

Convergence rates for the posterior mean



The PDE-constrained regression problem

- Building on nonparametric posterior contraction theory [van der Vaart and van Zanten, Ann. Statist. 2008], we first show that the induced posterior on the PDE solution $\Pi(\mathscr{G}f|Y,X)$ contracts around the true PDE solution $\mathscr{G}f_0$.

The PDE-constrained regression problem

- Building on nonparametric posterior contraction theory [van der Vaart and van Zanten, Ann. Statist. 2008], we first show that the induced posterior on the PDE solution $\Pi(\mathscr{G}f|Y,X)$ contracts around the true PDE solution $\mathscr{G}f_0$.

Theorem (G., Nickl 2020)

For $\alpha > 1 + d/2$, let $f_0 \in H^\alpha(\mathcal{O})$ satisfy $\inf_{x\in\mathcal{O}} f_0(x) > 0$ and suitable boundary conditions, and consider observations $(X, Y) \sim P^N_{f_0}$. Let Π be the above scaled Gaussian prior. Then, for some $L > 0$ as $N \to \infty$,

$$\Pi(f : \|\mathscr{G}f - \mathscr{G}f_0\|_{L^2} > LN^{-\frac{\alpha+1}{2\alpha+2+d}}|Y,X) \overset{P_{f_0}}{\to} 0.$$

The PDE-constrained regression problem

- Building on nonparametric posterior contraction theory [van der Vaart and van Zanten, Ann. Statist. 2008], we first show that the induced posterior on the PDE solution $\Pi(\mathscr{G}f|Y,X)$ contracts around the true PDE solution $\mathscr{G}f_0$.

Theorem (G., Nickl 2020)

For $\alpha > 1 + d/2$, let $f_0 \in H^\alpha(\mathcal{O})$ satisfy $\inf_{x\in\mathcal{O}} f_0(x) > 0$ and suitable boundary conditions, and consider observations $(X, Y) \sim P_{f_0}^N$. Let Π be the above scaled Gaussian prior. Then, for some $L > 0$ as $N \to \infty$,

$$\Pi(f : \|\mathscr{G}f - \mathscr{G}f_0\|_{L^2} > LN^{-\frac{\alpha+1}{2\alpha+2+d}}|Y,X) \overset{P_{f_0}}{\to} 0.$$

- The rate $N^{-\frac{\alpha+1}{2\alpha+2+d}}$ is optimal for estimating $\mathscr{G}f_0,\ f_0 \in H^\alpha(\mathcal{O})$ [Nickl, van der Geer and Wang, SIAM J. U. Q. 2020].

The PDE-constrained regression problem

- Building on nonparametric posterior contraction theory [van der Vaart and van Zanten, Ann. Statist. 2008], we first show that the induced posterior on the PDE solution $\Pi(\mathscr{G}f|Y,X)$ contracts around the true PDE solution $\mathscr{G}f_0$.

Theorem (G., Nickl 2020)

For $\alpha > 1 + d/2$, let $f_0 \in H^\alpha(\mathcal{O})$ satisfy $\inf_{x\in\mathcal{O}} f_0(x) > 0$ and suitable boundary conditions, and consider observations $(X,Y) \sim P^N_{f_0}$. Let Π be the above scaled Gaussian prior. Then, for some $L > 0$ as $N \to \infty$,

$$\Pi\left(f : \|\mathscr{G}f - \mathscr{G}f_0\|_{L^2} > LN^{-\frac{\alpha+1}{2\alpha+2+d}} \middle| Y,X\right) \overset{P_{f_0}}{\to} 0.$$

- The rate $N^{-\frac{\alpha+1}{2\alpha+2+d}}$ is optimal for estimating $\mathscr{G}f_0$, $f_0 \in H^\alpha(\mathcal{O})$ [Nickl, van der Geer and Wang, SIAM J. U. Q. 2020].

- Adaptation to the smoothness α is possible via a randomly truncated Gaussian series prior.

Contraction at f_0

- To show posterior contraction around f_0, the previous result is combined with the stability estimate in [Nickl, van der Geer and Wang, SIAM J. U. Q. 2020]: for $\alpha > 2 + d/2$,

$$\|f - f_0\|_{L^2} \leq c(\|f_0\|_{C^1}, \|f\|_{H^\beta})\|\mathscr{G}f - \mathscr{G}f_0\|_{L^2}^{\frac{\beta-1}{\beta+1}}, \qquad \beta < \alpha.$$

Contraction at f_0

- To show posterior contraction around f_0, the previous result is combined with the stability estimate in [Nickl, van der Geer and Wang, SIAM J. U. Q. 2020]: for $\alpha > 2 + d/2$,

$$\|f - f_0\|_{L^2} \leq c(\|f_0\|_{C^1}, \|f\|_{H^\beta})\|\mathscr{G}f - \mathscr{G}f_0\|_{L^2}^{\frac{\beta-1}{\beta+1}}, \qquad \beta < \alpha.$$

- Above, $c(\|f_0\|_{C^1}, \|f\|_{H^\beta}) \approx e^{\|f_0\|_{C^1}}\|f\|_{H^\beta}$. Hence, we need a bound on $\|f\|_{H^\beta}$ under the posterior, which motivates the use of scaled priors.

Contraction at f_0

- To show posterior contraction around f_0, the previous result is combined with the stability estimate in [Nickl, van der Geer and Wang, SIAM J. U. Q. 2020]: for $\alpha > 2 + d/2$,

$$\|f - f_0\|_{L^2} \le c(\|f_0\|_{C^1}, \|f\|_{H^\beta}) \|\mathscr{G} f - \mathscr{G} f_0\|_{L^2}^{\frac{\beta-1}{\beta+1}}, \qquad \beta < \alpha.$$

- Above, $c(\|f_0\|_{C^1}, \|f\|_{H^\beta}) \approx e^{\|f_0\|_{C^1}} \|f\|_{H^\beta}$. Hence, we need a bound on $\|f\|_{H^\beta}$ under the posterior, which motivates the use of scaled priors.

Theorem (G., Nickl 2020)

For $\alpha > 2 + d/2$, let $f_0 \in H^\alpha(\mathcal{O})$ satisfy $\inf_{x\in\mathcal{O}} f_0(x) > 0$ and suitable boundary conditions, and consider observations $(X, Y) \sim P_{f_0}^N$. Let Π be the above scaled Gaussian prior. Then, for some $L > 0$ as $N \to \infty$,

$$\Pi(f : \|f - f_0\|_{L^2} > L N^{-\zeta} | Y, X) \overset{P_{f_0}}{\to} 0, \qquad 0 < \zeta < \frac{\alpha - 1}{2\alpha + 2 + d}.$$

References

Giordano, M., and Nickl, R. (2020), *Consistency of Bayesian inference with Gaussian process priors in an elliptic inverse problem*, Inverse Problems 36, p. 1-35.

References

Giordano, M., and Nickl, R. (2020), *Consistency of Bayesian inference with Gaussian process priors in an elliptic inverse problem*, Inverse Problems 36, p. 1-35.

Thank you!