Efficient nonparametric statistical inference for a non-linear inverse problem with the Schrödinger equation

Richard Nickl

Statistical Laboratory
University of Cambridge (UK)

'Meeting in Mathematical Statistics'
Luminy, Dec 2017

Inverse regression problems

- We consider a regression-type measurement model

$$Y_i = u_f(x_i) + g_i, \quad i = 1, \dots, n; \quad g_i \sim^{i.i.d.} N(0,1)$$

where the x_i's are a regular grid in the underlying domain $\mathcal{O} \subset \mathbb{R}^d$.

Inverse regression problems

- We consider a regression-type measurement model

$$Y_i = u_f(x_i) + g_i, \quad i = 1, \dots, n; \quad g_i \sim^{i.i.d.} N(0,1)$$

where the x_i's are a regular grid in the underlying domain $\mathcal{O} \subset \mathbb{R}^d$.

- The appropriate translation of this model to the continuous limit is

$$Y = u_f + \varepsilon \mathbb{W}, \quad \text{with noise level } \varepsilon = \frac{1}{\sqrt{n}},$$

where $\mathbb{W}$ is a *Gaussian white noise process* in $L^2(\mathcal{O})$.

Inverse regression problems

- We consider a regression-type measurement model

$$Y_i = u_f(x_i) + g_i, \quad i = 1, \dots, n; \quad g_i \sim^{i.i.d.} N(0,1)$$

where the x_i's are a regular grid in the underlying domain $\mathcal{O} \subset \mathbb{R}^d$.

- The appropriate translation of this model to the continuous limit is

$$Y = u_f + \varepsilon \mathbb{W}, \quad \text{with noise level } \varepsilon = \frac{1}{\sqrt{n}},$$

where $\mathbb{W}$ is a *Gaussian white noise process* in $L^2(\mathcal{O})$.

- The regression function $u_f = K(f)$ represents the forward data of some inverse problem. The goal is to infer f from the observation $Y \sim P_{u_f}$.

Partial Differential Equation (PDE) Models

• In many important applications (Stuart (2010)), the forward map $u_f = u(f)$ arises as the solution of a PDE with unknown coefficients f.

• For example $u = u_f$ is the solution of an elliptic partial differential equation

$$f_1 \Delta u + f_2 \cdot \nabla u - f_3 u = 0 \text{ on } \mathcal{O}, \quad f = (f_1, f_2, f_3),$$

on some domain $\mathcal{O}$ in $\mathbb{R}^d$, subject to suitable boundary conditions. Here

$$\Delta = \sum_{i=1}^{d} \frac{\partial^2}{\partial x_i^2}$$

is the standard Laplacian and ∇ is the gradient operator.

• Some non-standard features:

a) $f \mapsto u_f$ is **non-linear**, and

b) unless $d = 1$ the solutions u_f do not have closed-form expressions.

A special case: Schrödinger equation

• Suppose $\mathcal{O}$ is a bounded open domain in $\mathbb{R}^d$ with a smooth boundary. Let

$$f : \mathcal{O} \to [0, \infty)$$

be an unknown potential and consider solutions to the Schrödinger equation

$$\frac{\Delta}{2} u - fu = 0 \text{ on } \mathcal{O}, \quad s.t. \ u = g \text{ on } \partial\mathcal{O},$$

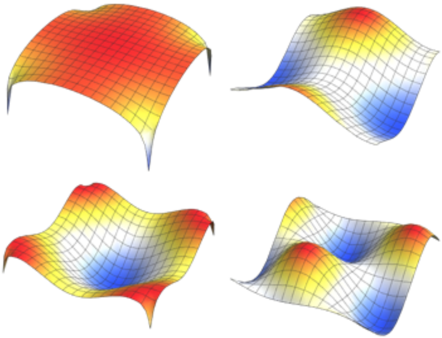
where $g : \partial\mathcal{O} \to (0, \infty)$ is a given function describing Dirichlet boundary conditions.

• The unknown function f models an attenuation of the solution of the standard Laplace equation.

Inverse medium and scattering problems

• Recovering f from observed solutions u_f is important in many physical applications, including photo-acoustics (Bal and Uhlmann, 2010) or electromagnetic waves (Bao and Li, 2005).

• Physically f models a 'cooling' or 'absorption' effect in the medium $\mathcal{O}$, where the local amount of absorption or cooling is described by the values $f(x)$.



Understanding the non-linearity

- From the famous Feynman-Kac representation we know

$$u_f(x) = E^x\left[g(X_{\tau_{\mathcal{O}}})e^{-\int_0^{\tau_{\mathcal{O}}} f(W_s)ds}\right],\ x \in \mathcal{O},$$

where $(W_s : s \geq 0)$ is a d-dimensional Brownian motion started at $x \in \mathcal{O}$, with almost surely finite exit time $\tau_{\mathcal{O}}$ from $\mathcal{O}$.

Understanding the non-linearity

- From the famous Feynman-Kac representation we know

$$u_f(x) = E^x \left[g(X_{\tau_{\mathcal{O}}}) e^{-\int_0^{\tau_{\mathcal{O}}} f(W_s) ds} \right], \ x \in \mathcal{O},$$

  where $(W_s : s \geq 0)$ is a d-dimensional Brownian motion started at $x \in \mathcal{O}$, with almost surely finite exit time $\tau_{\mathcal{O}}$ from $\mathcal{O}$.

- From the above representation, and since $g > 0$, one shows $u_f \geq c > 0$, so

$$f = \frac{\Delta u_f}{2u_f} \ \text{ on } \mathcal{O}.$$

  One also shows, using standard regularity theory for elliptic PDEs, that u_f is two degrees smoother than f.

Some first approaches

- Things **not** to do are:

  a) Estimate f by $\frac{\Delta Y}{2Y}$, as it makes no pointwise sense.

  b) Estimate f by a plug-in regression smoother $\frac{\Delta \hat{u}_f}{2\hat{u}_f}$ since nonparametric regression estimators $\hat{u}_f$ are **not** solutions of the Schrödinger equation.

Some first approaches

- Things **not** to do are:

  a) Estimate f by $\frac{\Delta Y}{2Y}$, as it makes no pointwise sense.

  b) Estimate f by a plug-in regression smoother $\frac{\Delta \hat{u}_f}{2\hat{u}_f}$ since nonparametric regression estimators $\hat{u}_f$ are **not** solutions of the Schrödinger equation.

- One may be tempted to compute a nonparametric maximum likelihood estimator to incorporate the non-linear constraints on the parameter space.

Some first approaches

- Things **not** to do are:

  a) Estimate f by $\frac{\Delta Y}{2Y}$, as it makes no pointwise sense.

  b) Estimate f by a plug-in regression smoother $\frac{\Delta \hat{u}_f}{2\hat{u}_f}$ since nonparametric regression estimators $\hat{u}_f$ are **not** solutions of the Schrödinger equation.

- One may be tempted to compute a nonparametric maximum likelihood estimator to incorporate the non-linear constraints on the parameter space.

- This would be closely related to the Tikhonov-regulariser:

$$\min_f \left[-2\langle Y, u_f\rangle_{L^2} + \|u_f\|_{L^2}^2 + \lambda\|f\|_{H^s}^2\right],$$

  which however is **not convex** in f due to the non-linearity of $f \mapsto u_f$.

Some first approaches

- Things **not** to do are:

  a) Estimate f by $\frac{\Delta Y}{2Y}$, as it makes no pointwise sense.

  b) Estimate f by a plug-in regression smoother $\frac{\Delta \hat{u}_f}{2\hat{u}_f}$ since nonparametric regression estimators $\hat{u}_f$ are **not** solutions of the Schrödinger equation.

- One may be tempted to compute a nonparametric maximum likelihood estimator to incorporate the non-linear constraints on the parameter space.

- This would be closely related to the Tikhonov-regulariser:

$$\min_f \left[-2\langle Y, u_f\rangle_{L^2} + \|u_f\|_{L^2}^2 + \lambda\|f\|_{H^s}^2\right],$$

  which however is **not convex** in f due to the non-linearity of $f \mapsto u_f$.

- As is well known, such a 'penalised least squares' approach is closely related to Bayesian inference with Gaussian priors (Kimeldorf and Wahba (1970)).

Bayes solutions of noisy inverse problems

- The Bayesian approach to statistical inverse problems has recently attracted substantial attention in applied mathematics thanks to the work of Andrew Stuart, see the surveys by Stuart (2010) and Dashti and Stuart (2016).

Bayes solutions of noisy inverse problems

- The Bayesian approach to statistical inverse problems has recently attracted substantial attention in applied mathematics thanks to the work of Andrew Stuart, see the surveys by Stuart (2010) and Dashti and Stuart (2016).
- Formally we model the function f by some prior distribution Π on function space and use Bayes' rule to compute the conditional posterior distribution

$$f \sim \Pi, \ \ Y|f \sim P_{u_f} \ \ \Rightarrow f|Y \sim \frac{dP_{u_f}(Y)d\Pi(f)}{\int dP_{u_f}(Y)d\Pi(f)}$$

where $dP_f \equiv dP_{u_f}$ is the law of a Gaussian white noise shifted by u_f.

Bayes solutions of noisy inverse problems

- The Bayesian approach to statistical inverse problems has recently attracted substantial attention in applied mathematics thanks to the work of Andrew Stuart, see the surveys by Stuart (2010) and Dashti and Stuart (2016).

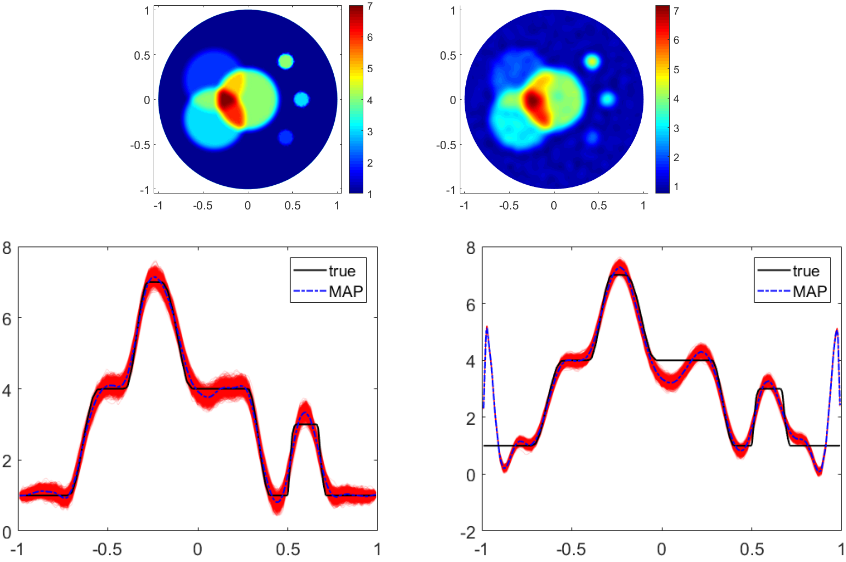
- Formally we model the function f by some prior distribution Π on function space and use Bayes' rule to compute the conditional posterior distribution

$$f \sim \Pi,\ \ Y|f \sim P_{u_f} \ \Rightarrow f|Y \sim \frac{dP_{u_f}(Y)d\Pi(f)}{\int dP_{u_f}(Y)d\Pi(f)}$$

  where $dP_f \equiv dP_{u_f}$ is the law of a Gaussian white noise shifted by u_f.

- Numerically efficient MCMC algorithms for computing posterior distributions have been put forward in the past decade; they allow to compute point estimates (MAP) and 'credible sets' for the unknown parameter f from the posterior distribution $\Pi(\cdot|Y)$.

Implementation for smooth images



Minimax rates for the inverse Schrödinger problem

• Recall that we are observing $Y = u_f + \varepsilon \mathbb{W}$ where u_f is the solution of

$$\frac{\Delta}{2} u - fu = 0 \text{ on } \mathcal{O}, \quad s.t.\ u = g \text{ on } \partial\mathcal{O}$$

in a smooth domain $\mathcal{O}$ and we consider the inverse problem of finding the positive potential f from the data as the noise level $\varepsilon \to 0$.

• Let $C^s(\mathcal{O})$ denote the usual Hölder space over the domain $\mathcal{O}$, and for later use also define the subspace $C^s_c(\mathcal{O})$ of functions compactly supported within $\mathcal{O}$.

Theorem 1

Suppose $g \in C^{s+2}(\partial\mathcal{O})$ satisfies $g \geq g_{min} > 0$. We have as $\varepsilon \to 0$ that

$$\inf_{\tilde f = \tilde f(Y,g)} \sup_{f : \|f\|_{C^s(\mathcal{O})} \leq B, f \geq f_{min} > 0} \varepsilon^{-2s/(2s+4+d)} E_f^Y \|\tilde f - f\|_{L^2(\mathcal{O})} \simeq c(s, B, g) > 0.$$

• This corresponds to a 2-ill posed problem but also incorporates the 2-smoothing property of the ‘elliptic forward map’ $f \mapsto u_f$.

A wavelet series prior and contraction theorem

• We model f by a prior Π induced by the random function

$$\log f = \sum_{l \le J, r} b_{l,r} \Phi^{\mathcal{O}}_{l,r}, \quad b_{l,r} \sim^{i.i.d.} U(-B2^{-l(s+d/2)}, 2^{-l(s+d/2)}B),$$

where the $(\Phi^{\mathcal{O}}_{l,r})$ form a (boundary corrected) wavelet basis of $L^2(\mathcal{O})$.

• We do not consider hyper-priors for B, s, J here, but regard s, B as given and choose J such that $2^J \approx \varepsilon^{-2s/(2s+4+d)}$ ('non-adaptive' case).

Theorem 2

Suppose $f_0 > 0$ satisfies $\|\log f_0\|_{C^s_c(\mathcal{O})} \le B$ for some $s > 2 + d/2$. If $\Pi(\cdot|Y)$ is the posterior distribution arising from the above prior Π, then, as $\varepsilon \to 0$ we have

$$\Pi\left(f : \|f - f_0\|_{L^2} \ge M\varepsilon^{-2s/(2s+4+d)} \log^\gamma(1/\varepsilon) | Y\right) \to 0 \text{ in } P^Y_{f_0}\text{-probability.}$$

• The posterior mean $\bar{f}(Y)$ attains the same convergence rate in $L^2(\mathcal{O})$.

Towards Bernstein -von Mises theorems: LAN expansion

• We want more precise 'efficiency' results, and to justify Bayesian 'uncertainty quantification'.

• To do this we first need to understand the 'information geometry' of the LAN expansion. Using perturbation arguments for PDEs one can prove

Theorem 3

Under mild conditions on $h \in C(\mathcal{O}), f_0 > 0, g > 0$, if $Y = u_{f_0} + \varepsilon \mathbb{W}$ then as $\varepsilon \to 0$

$$\log \frac{dP_{u_{f_0+\varepsilon h}}}{dP_{u_{f_0}}}(Y) = \langle Du_{f_0}[h], \mathbb{W} \rangle_{L^2} - \frac{1}{2}\|Du_{f_0}[h]\|_{L^2}^2 + o_{P_{f_0}^Y}(1),$$

where the score operator $Du_{f_0}[\cdot]$ maps h into the solution $v = v_h$ of the inhomogeneous Schrödinger equation

$$\frac{\Delta v}{2} - fv = h u_{f_0} \text{ on } \mathcal{O} \text{ s.t. } v = 0 \text{ on } \partial\mathcal{O}.$$

• The forward operator has the strong quadratic approximation

$$\|u_{f_0+h} - u_{f_0} - Du_{f_0}[h]\|_{L^2} = O(\|h\|_{L^2}\|h\|_{H^{-2}}), \quad h \in C(\mathcal{O}).$$

Semi-parametric Cramer-Rao bound

- To find the Cramér-Rao information bound for inference on linear forms $\langle f, \psi \rangle_{L^2}$ we need to find $\tilde{\psi}$ s.t. for all $h \in C(\mathcal{O})$,

$$\langle h, \psi \rangle_{L^2(\mathcal{O})} = \langle Du_{f_0}(h), D_{u_{f_0}}(\tilde{\psi}) \rangle_{L^2(\mathcal{O})} = \langle V_{f_0}(hu_{f_0}), V_{f_0}(\tilde{\psi} u_{f_0}) \rangle_{L^2(\mathcal{O})}$$

where V_f is the 'Green operator' of the inhomogeneous Schrödinger eq.

Semi-parametric Cramer-Rao bound

- To find the Cramér-Rao information bound for inference on linear forms $\langle f, \psi \rangle_{L^2}$ we need to find $\tilde{\psi}$ s.t. for all $h \in C(\mathcal{O})$,

$$\langle h, \psi \rangle_{L^2(\mathcal{O})} = \langle Du_{f_0}(h), D_{u_{f_0}}(\tilde{\psi}) \rangle_{L^2(\mathcal{O})} = \langle V_{f_0}(hu_{f_0}), V_{f_0}(\tilde{\psi} u_{f_0}) \rangle_{L^2(\mathcal{O})}$$

  where V_f is the 'Green operator' of the inhomogeneous Schrödinger eq.

- Using properties of the Feynman-Kac semigroup one shows that V_f has a symmetric integral kernel and inverse given by the Schrödinger operator,

$$S_f(u) = \frac{\Delta}{2} u - fu, \quad u \in C^2(\mathcal{O}).$$

Semi-parametric Cramer-Rao bound

- To find the Cramér-Rao information bound for inference on linear forms $\langle f, \psi \rangle_{L^2}$ we need to find $\tilde{\psi}$ s.t. for all $h \in C(\mathcal{O})$,

$$\langle h, \psi \rangle_{L^2(\mathcal{O})} = \langle Du_{f_0}(h), D_{u_{f_0}}(\tilde{\psi}) \rangle_{L^2(\mathcal{O})} = \langle V_{f_0}(hu_{f_0}), V_{f_0}(\tilde{\psi} u_{f_0}) \rangle_{L^2(\mathcal{O})}$$

  where V_f is the 'Green operator' of the inhomogeneous Schrödinger eq.

- Using properties of the Feynman-Kac semigroup one shows that V_f has a symmetric integral kernel and inverse given by the Schrödinger operator,

$$S_f(u) = \frac{\Delta}{2} u - fu, \quad u \in C^2(\mathcal{O}).$$

Theorem 4

For sufficiently regular, compactly supported functions ψ, the information lower bound for estimating $\langle f, \psi \rangle_{L^2(\mathcal{O})}$ at $f = f_0$ from observations $Y = u_{f_0} + \varepsilon \mathbb{W}$ is

$$I_{f_0}(\psi) = \|Du_{f_0}[\tilde{\psi}]\|^2_{L^2(\mathcal{O})} = \|S_{f_0}(\psi / u_{f_0})\|^2_{L^2(\mathcal{O})}.$$

- PDE techniques are key to derive an explicit Cramer-Rao lower bound.

Formulation of the Bernstein-von Mises theorem

- For a fixed function ψ we can now ask whether the 'semi-parametric' Bernstein von Mises theorem holds true:

Let $f \sim \Pi(\cdot|Y)$ and $\bar{f}$ the posterior mean. As $\varepsilon \to 0$ do we have

$$\varepsilon^{-1} \int_{\mathcal{O}} (f - \bar{f})\psi | Y \to^d N(0, I_{f_0}(\psi))$$

in $P_{f_0}^Y$-probability?

Formulation of the Bernstein-von Mises theorem

- For a fixed function ψ we can now ask whether the 'semi-parametric' Bernstein von Mises theorem holds true:

Let $f \sim \Pi(\cdot|Y)$ and $\bar{f}$ the posterior mean. As $\varepsilon \to 0$ do we have

$$\varepsilon^{-1} \int_{\mathcal{O}} (f - \bar{f})\psi | Y \to^d N(0, I_{f_0}(\psi))$$

in $P^Y_{f_0}$-probability?

- In fact we want more: following a multi-scale approach, we wish to prove simultaneous convergence of the stochastic processes

$$\left(\varepsilon^{-1} \int_{\mathcal{O}} (f - \bar{f})\psi : \psi \in \Psi_a | Y\right) \to^d (X(\psi) : \psi \in \Psi_a),$$

where Ψ_a is a maximal class of test functions and $X(\psi) \sim N(0, I_{f_0}(\psi))$.

A canonical limiting Gaussian distribution

- The Gaussian process

$$\mathbb{X} = (X(\psi)), \quad \mathbb{E}X(\psi)X(\psi') = \langle S_{f_0}(\psi/u_{f_0}), S_{f_0}(\psi'/u_{f_0})\rangle_{L^2(\mathcal{O})}, \ \psi, \psi' \in C_c^\alpha(\mathcal{O}),$$

is the image of a standard Gaussian white noise $\mathbb{W}$ under the Schrödinger-type operator $\psi \mapsto S_{f_0}(\psi/u_{f_0})$.

- If a sequence of stochastic processes $(X_n(\psi))$ is to converge uniformly in $\psi \in \Psi$ towards $\mathbb{X}$, then the law $\mathcal{N}_{f_0}$ of $\mathbb{X}$ needs to be **tight** for the supremum norm on Ψ.

- If $\Psi = \Psi_a$ consists of the unit ball of an α-Hölder space $C_c^\alpha(\mathcal{O})$, then the maximal spaces where this is possible are characterised in the following

Theorem 5

The Gaussian measure $\mathcal{N}_{f_0}$ induces a tight Gaussian probability measure on the topological dual space $(C_c^\alpha(\mathcal{O}))^*$ when $\alpha > 2 + d/2$ but **not** when $\alpha \leq 2 + d/2$.

The 'non-parametric' Bernstein-von Mises theorem

• The last theorem identifies the 'minimal' spaces $(C_c^\alpha(\mathcal{O}))^*, \alpha > 2 + d/2$, in which we can converge weakly towards $\mathcal{N}_{f_0}$.

Theorem 6

Let $s > \max(2 + d/2, d)$ and $\alpha > 2 + d/2$. Assume $\|\log f_0\|_{C_c^s(\mathcal{O})} < B$. Let $f \sim \Pi(\cdot|Y)$ with posterior mean $\bar{f}$, and denote by β any metric for weak convergence of probability distributions on $(C_c^\alpha(\mathcal{O}))^*$. Then

$$\beta\left(\mathcal{L}(\varepsilon^{-1}(f - \bar{f})|Y), \mathcal{N}_{f_0}\right) \to 0$$

as $\varepsilon \to 0$ in $P_{f_0}^Y$-probability. Moreover $\bar{f}$ is an efficient estimator of f.

• The proof is quite involved, based on uniform perturbation expansions of the Laplace transform of the posterior distribution.

Optimal credible sets I

- As a first application, posterior inference for linear functionals $\langle f, \psi \rangle_{L^2}, \psi \in C_c^\alpha$ is asymptotically valid and optimal, without requiring estimation of the inverse Fisher information.

- For a fixed 'credibility level' $1-\beta, \beta > 0$ and test function ψ, let

$$C_\varepsilon = \{x \in \mathbb{R} : |x - \langle \bar{f}, \psi \rangle_{L^2}| \le R_\varepsilon\}$$

with posterior quantile constants R_ε chosen such that $\Pi(C_\varepsilon | Y) = 1 - \beta$.

Theorem 7

Let $\psi \in C_c^\alpha(\mathcal{O})$ with $\alpha > 2 + d/2$. Then as $\varepsilon \to 0$ we have

$$P_{f_0}^Y(\langle f_0, \psi \rangle_{L^2} \in C_\varepsilon) \to 1 - \beta,$$

as $\varepsilon \to 0$ and the diameter R_ε of C_ε satisfies

$$\varepsilon^{-1} R_\varepsilon \to^{P_{f_0}^Y} \Phi^{-1}(1-\beta)$$

where Φ^{-1} is the inverse of the map $t \mapsto N(0, \|S_{f_0}(\psi/u_{f_0})\|_{L^2(\mathcal{O})}^2)([-t, t])$.

Optimal credible sets II

• Let $C_\varepsilon \subset supp(\Pi(\cdot|Y))$ be the smallest $(C_c^\alpha(\mathcal{O}))^*$-ball centred at the posterior mean $\bar{f}$ for which $\Pi(C_\varepsilon|Y) = 1-\beta$, with fixed credibility $1-\beta, \beta > 0$.

Theorem 8

Under the conditions of Theorem 6 the above credible set satisfies

$$P_{f_0}^Y(f_0 \in C_\varepsilon) \to 1-\beta,$$

as $\varepsilon \to 0$. Its diameter in $L^1(K)$-norm, for K any compact subset of $\mathcal{O}$, is of (near) minimax-optimal order: for any $\kappa > 0$,

$$|C_\varepsilon|_{L^1(K)} = O_{P_{f_0}^Y}\left(\varepsilon^{2s/(2s+4+d)}\varepsilon^{-\kappa}\right).$$

• See Ray (2017, AoS) for more discussion of 'geometric aspects' of such 'multiscale' credible sets in infinite dimensions.

References

Bayesian Algorithms for Inverse problems

A. Stuart, Inverse problems: A Bayesian perspective, *Acta Numerica*, 2010.

M. Dashti, A. Stuart, The Bayesian approach to inverse problems, arxiv 2016.

This talk:

R. Nickl, Bernstein-von Mises theorems for non-linear inverse problems I: Schrödinger equation, arxiv 1707.01764

Acknowledgements:
EUROPEAN RESEARCH COUNCIL (ERC)