Linear methods for non-linear inverse problems

Botond Szabó (Bocconi University)

Non-Linear and High Dimensional Inference

Institut Henri Poincaré
Paris, France, 05. 10. 2022.

Co-authors

Geerten Koers
(Delft)

Aad van der Vaart
(Delft)

Outline

- Brief introduction to inverse problems and frequentist Bayes
- Literature review
  - Linear inverse problems (recovery and UQ)
  - Non-linear inverse problems (recovery and UQ)
- New approach: Linear methods for non-linear inverse problems
  - General description
  - Demonstration: Schrödinger equation (recovery and UQ)
  - Theory: contraction rates and uncertainty quantification
  - Numerical Analysis
- Extensions and conclusions

Inverse problems

- **Model:** Indirect, noise observation of f_0

$$Y = \mathcal{A}f_0 + \epsilon,$$

with $\mathcal{A}$ a known, injective operator and ϵ noise.
- **Goal:** Recover f_0 and quantify the uncertainty of the procedure.

Inverse problems

- **Model:** Indirect, noise observation of f_0

$$Y = \mathcal{A}f_0 + \epsilon,$$

  with $\mathcal{A}$ a known, injective operator and ϵ noise.
- **Goal:** Recover f_0 and quantify the uncertainty of the procedure.
- **Problem:** If $\mathcal{A}^{-1}$ is not continuous, simply inverting Y amplifies noise.
- **Solution:** Bayesian approach.
- **Literature:** Dashti and Stuart (2016), Cotter et al (2013), Helin and Burger (2015), Lassas et al (2009),...etc

Bayes vs frequentist

Schools:	**Frequentist**	**Bayes**
Model:	$Y \sim P_{\theta_0},\ \theta_0 \in \Theta$	$\theta \sim \Pi$ (prior), $Y \mid \theta \sim P_\theta$
Goal:	Try to recover θ_0: Estimator, hypothesis testing	Update our belief about θ: Posterior: $\theta \mid Y$
UQ:	Confidence sets	Credible sets

Bayes vs frequentist

Schools:	**Frequentist**	**Bayes**
Model:	$Y \sim P_{\theta_0}$, $\theta_0 \in \Theta$	$\theta \sim \Pi$ (prior), $Y \mid \theta \sim P_\theta$
Goal:	Try to recover θ_0: Estimator, hypothesis testing	Update our belief about θ: Posterior: $\theta \mid Y$
UQ:	Confidence sets	Credible sets

Frequentist Bayes

Investigate Bayesian techniques from frequentist perspective, i.e. assume that there exists a true θ_0 and investigate the behaviour of the posterior $\theta \mid Y$.

Theorem (Bernstein-von Mises): Under some regularity conditions on the likelihood and prior the posterior is asymptotically normal centered around an efficient estimator and with covariance matrix equal to the inverse Fischer information, i.e.

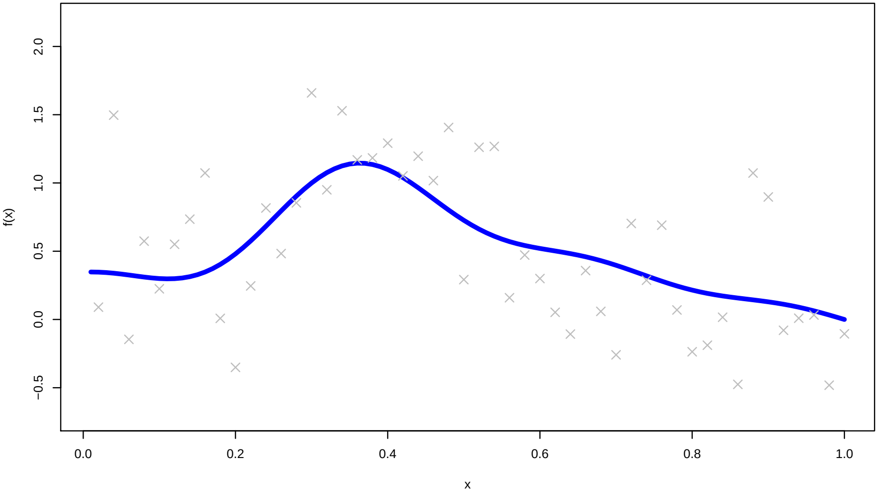
$$\|\Pi(\vartheta \in \cdot \,|\, Y_1, Y_2, \ldots) - N_d(\hat{\theta}_n, \frac{1}{n} i_{\theta_0}^{-1})\|_{TV} \stackrel{P_{\theta_0}}{\to} 0,$$

where $\hat{\theta}_n$ is the MLE and i_θ the Fischer information matrix.

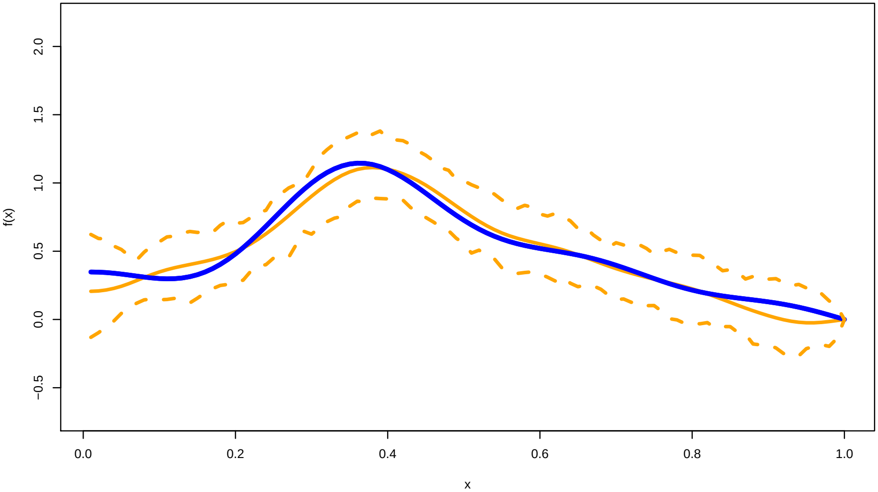
Consequences:

- Posterior mean, median are efficient estimators (asymptotically UMVU).
- Credible sets are confidence sets asymptotically.

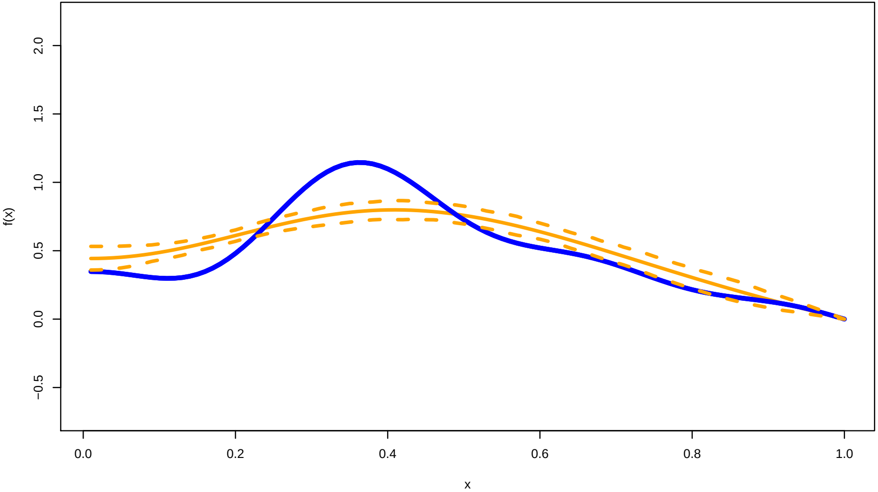
Regression



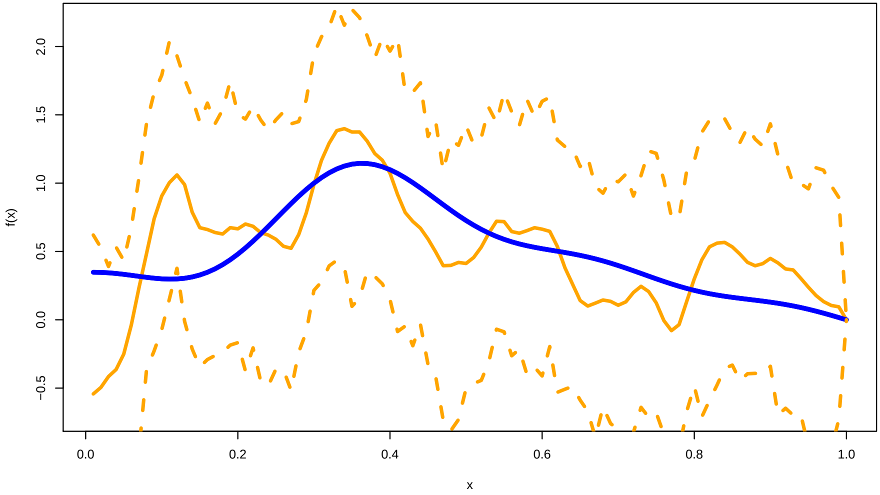
Posterior I



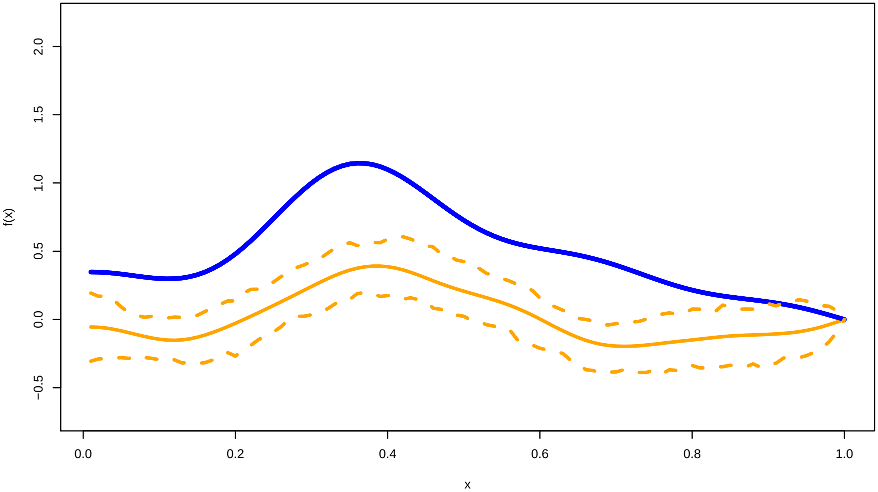
Posterior II



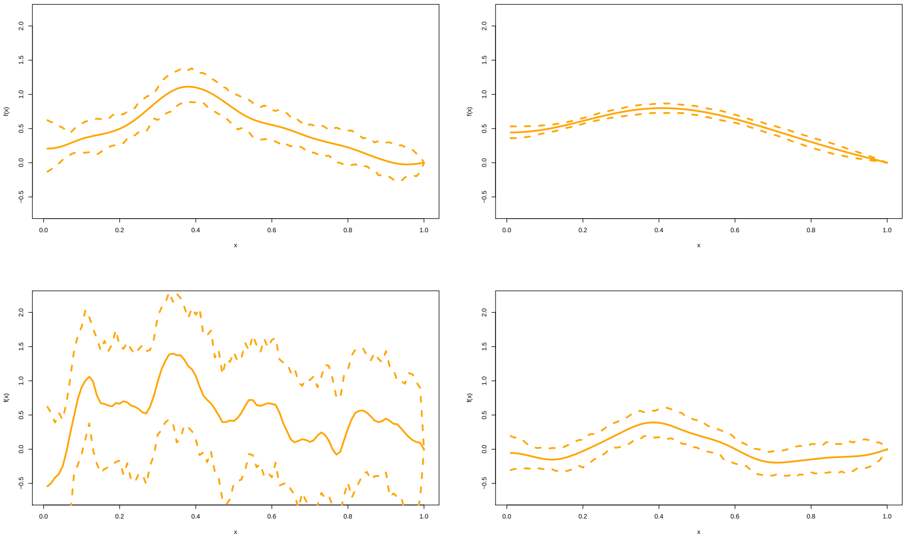
Posterior III



Posterior IV



Which one is correct?



Frequentist Bayes: Main questions

Can we achieve optimal recovery (for some pseudo metric d):

$$\inf_{\theta_0 \in \Theta_0} E_{\theta_0} \Pi\left(\theta : d(\theta, \theta_0) \leq M r_n | Y^{(n)}\right) \to 1?$$

with $M > 0$ and

$$r_n = \inf_{\hat{\theta}} \sup_{\theta_0 \in \Theta_0} E_{\theta_0} d(\theta, \theta_0).$$

Frequentist Bayes: Main questions

Can we achieve optimal recovery (for some pseudo metric d):

$$\inf_{\theta_0 \in \Theta_0} E_{\theta_0} \Pi\big(\theta : d(\theta, \theta_0) \leq M r_n | Y^{(n)}\big) \to 1?$$

with $M > 0$ and

$$r_n = \inf_{\hat{\theta}} \sup_{\theta_0 \in \Theta_0} E_{\theta_0} d(\theta, \theta_0).$$

Can we get reliable uncertainty quantification, i.e. does it hold for $\hat{C} = \{\theta : d(\theta, \hat{\theta}) \leq \rho_n\}$ with $\Pi(\hat{C} | Y^{(n)}) = 0.95$ that

$$\inf_{\theta_0 \in \Theta_0} P_{\theta_0}(\theta_0 \in \hat{C}) \geq 0.95?$$

Literature: linear inverse problems

Direct problems: $\mathcal{A}$ is the identity operator

- Optimal contraction rates in general setting: Ghoshal et al (2000), Ghoshal and vd Vaart (2007,2016), Shen and Wasserman (2001),....
- Reliable uncertainty quantification (under some possible additional constraints): Knapik et al (2011), Castillo and Nickl (2014), Sz. et al (2015), Rousseau and Sz. (2020),... etc

Direct problems: $\mathcal{A}$ is the identity operator

- Optimal contraction rates in general setting: Ghoshal et al (2000), Ghoshal and vd Vaart (2007,2016), Shen and Wasserman (2001),....
- Reliable uncertainty quantification (under some possible additional constraints): Knapik et al (2011), Castillo and Nickl (2014), Sz. et al (2015), Rousseau and Sz. (2020),... etc

Linear inverse problems: assume that $\mathcal{A}$ is a known, continuous, injective, compact, linear operator, see for instance Bissantz et al. (2007).

Benchmark model: inverse Gaussian white noise

$$Y^{(n)} = \mathcal{A}f_0 + \frac{1}{\sqrt{n}}Z,$$

where $X^{(n)}$ is the observation polluted with the Gaussian white noise Z.

Equivalent sequence formulation

Idea: For compact $\mathcal{A}$ the spectral decomposition of the self-adjoint operator $\mathcal{A}^T\mathcal{A}$ provides a convenient basis.

Equivalent Model: We observe the noisy sequence:

$$Y_i = \kappa_i f_{0,i} + \frac{1}{\sqrt{n}} Z_i, \qquad i = 1, 2, \dots$$

where κ_i^2 are the eigenvalues of $\mathcal{A}^T\mathcal{A}$, $f_{0,i}$ are the coefficients of f_0 with respect to the eigenbasis of $\mathcal{A}^T\mathcal{A}$ and Z_i are iid $N(0,1)$.

Assumptions:

- Mildly ill-posed inverse problem: $C^{-2} i^{-2p} \leq \kappa_i^2 \leq C^2 i^{-2p}$.
- For $f_0 \in S^\beta(R)$ the minimax rate: $n^{-\beta/(1+2\beta+2p)}$, see Mair and Ruymgaart (1996).

Prior: Put a Gaussian prior on the coefficients):

$$f_i \overset{ind}{\sim} N(0, i^{-1-2\alpha}), \qquad i = 1, 2, \ldots$$

Recovery: posterior contraction (for $M > 0$ large enough)

$$\sup_{f_0 \in S^\beta(R)} E_{f_0} \Pi(f : \|f - f_0\|_{L_2} \leq M n^{-\frac{\alpha \wedge \beta}{1+2\alpha+2p}} | Y^{(n)}) \to 1$$

reaches the minimax rate for $\alpha = \beta$. For $\alpha \neq \beta$ suboptimal, but rescaling can help.

Prior: Put a Gaussian prior on the coefficients):

$$f_i \overset{ind}{\sim} N(0, i^{-1-2\alpha}), \qquad i = 1, 2, \ldots$$

Recovery: posterior contraction (for $M > 0$ large enough)

$$\sup_{f_0 \in S^\beta(R)} E_{f_0} \Pi(f : \|f - f_0\|_{L_2} \leq M n^{-\frac{\alpha \wedge \beta}{1+2\alpha+2p}} | Y^{(n)}) \to 1$$

reaches the minimax rate for $\alpha = \beta$. For $\alpha \neq \beta$ suboptimal, but rescaling can help.

Uncertainty quantification: reliable for $\alpha \approx \beta$ (under slightly under-smoothing), i.e. the $1 - \gamma$ credible ball $\hat{C} = \{f : \|f - \hat{f}_\alpha\|_{L_2} \leq \rho_n\}$ satisfies

$$\inf_{f_0 \in S^\beta(R)} P_{f_0}(f_0 \in \hat{C}) \to 1.$$

Literature: Cox (1993), Knapik et al. (2011), Florens and Simoni (2012), Agapiou et al. (2013), Ray (2013), Yan et al. (2021).

Problem: β is typically unknown $\Rightarrow$ Data driven choice.

- **Empirical Bayes:** estimate α from the data $\hat{\alpha}$ (e.g. maximum marginal likelihood estimator).
- **Hierarchical Bayes:** endow α with a prior λ. Two level, hierarchical prior $\Pi = \int \Pi_\alpha \lambda(\alpha) d\alpha$.

Adaptation

Problem: β is typically unknown $\Rightarrow$ Data driven choice.

- **Empirical Bayes:** estimate α from the data $\hat{\alpha}$ (e.g. maximum marginal likelihood estimator).
- **Hierarchical Bayes:** endow α with a prior λ. Two level, hierarchical prior $\Pi = \int \Pi_\alpha \lambda(\alpha) d\alpha$.

Theorem [Adaptive contraction] Both the hierarchical (with $\lambda(\alpha) \asymp \alpha^{-c_1} e^{-c_2\alpha}$) and the empirical Bayes posterior contracts around the truth with optimal (adaptive minimax) rate, i.e. for $0 < \beta_1 < \beta_2 < \infty$

$$\inf_{\beta_1 \leq \beta \leq \beta_2} \inf_{f_0 \in S^\beta(R)} \Pi_{\hat{\alpha}}(f : \|f - f_0\|_{L_2} \leq L_n n^{-\frac{\beta}{1+2\beta+2p}} | Y^{(n)}) \xrightarrow{P_{f_0}} 1.$$

Literature: Knapik et al. (2016), vd Vaart and v Zanten (2009), Rousseau and Sz. (2017), Belitser and Ghoshal (2003), Shen et al. (2013), Castillo et al. (2013), Pati et al. (2014).

Literature: Bayesian non-linear inverse problems

Problem: if $\mathcal{A}$ is non-linear $\Rightarrow$ SVD does not work.

Literature: Bayesian non-linear inverse problems

Problem: if $\mathcal{A}$ is non-linear $\Rightarrow$ SVD does not work.

Literature: Using multi-scale analysis deriving weak Bernstein-von Mises theorem over negative Sobolev classes. This is achieved by deriving BvM results for smooth linear functionals of the functional parameter f_0.

- Optimal contraction rates for rescaled priors in non-adaptive setting.
- Reliable uncertainty quantification.

Literature: Bayesian non-linear inverse problems

Problem: if $\mathcal{A}$ is non-linear $\Rightarrow$ SVD does not work.

Literature: Using multi-scale analysis deriving weak Bernstein-von Mises theorem over negative Sobolev classes. This is achieved by deriving BvM results for smooth linear functionals of the functional parameter f_0.

- Optimal contraction rates for rescaled priors in non-adaptive setting.
- Reliable uncertainty quantification.

Models:

- Time-independent Schrodinger equation Nickl et al (2020, Nickl(2020), Monard et al (2021)
- Non-abelian x-ray transform Monard et al (2021), Monard et al (2021),
- Elliptic PDE in divergence form Giordano and Nickl (2020), Nickl et al (2020),
- Heat equation with an additional absorption Kekkonen (2022)
- Calderón problems Abraham and Nickl (2022)

Literature: Bayesian non-linear inverse problems II

Computation:

- No explicit formula for the posterior $\Rightarrow$ approximate sampling methods (standard MCMC - slow)
- Langevin (type) diffusion has polynomial computational time Nickl and Wang (2022+)
- BUT still not practical for medium to large sample size.
- Main problem: solving the forward map at each sampling step.

Consider the PDE (for $u_f = \mathcal{A}f$)

$$\begin{cases} \mathcal{L}u_f = c(f, u_f) & \text{on } \mathcal{O}, \\ u_f = g & \text{on } \Gamma \subseteq \partial\mathcal{O}. \end{cases},$$

where $\mathcal{L}$ is a linear differential operator, c and $g : \partial\mathcal{O} \to \mathbb{R}$ are known functions. Assume that $f = e(\mathcal{L}u_f, u_f)$ for some known map e.

Consider the PDE (for $u_f = \mathcal{A}f$)

$$\begin{cases} \mathcal{L}u_f = c(f, u_f) & \text{on } \mathcal{O}, \\ u_f = g & \text{on } \Gamma \subseteq \partial\mathcal{O}. \end{cases},$$

where $\mathcal{L}$ is a linear differential operator, c and $g : \partial\mathcal{O} \to \mathbb{R}$ are known functions.
Assume that $f = e(\mathcal{L}u_f, u_f)$ for some known map e.

Observational models:

- Gaussian white noise: $Y^{(n)} = u_f + \frac{1}{\sqrt{n}}Z$,
  where Z denotes the Gaussian white noise process. More formally

$$Y^{(n)}(h) = \langle h, u_f \rangle_{L_2} + Z(h),$$

  where $\{Z(h) : h \in L_2(\mathcal{O})\}$ is mean zero Gaussian process with covariance function $\mathrm{Cov}(h_1, h_2) = \langle h_1, h_2 \rangle_{L_2}$.
- Regression: $Y_i = u_f(X_i) + Z_i$, $Z_i \sim^{iid} N(0, 1)$, $i = 1, ..., n$.

Examples

- Time-independent Schrödinger equation: $\mathcal{L} = \Delta$, $c(u, f) = 2uf$, $e(v, u) = v/(2u)$.
- Heat equation with an additional absorption: $\mathcal{L} = -\Delta_x/2 + \partial_t$, $c(u, f) = uf$, $e(v, u) = v/u$.
- Exponentiating the Volterra operator: $u_f(t) = e^{\int_0^t f(s)ds}$, $\mathcal{L} = \partial_x$, $c(u, f) = uf$, $e(v, u) = v/u$.
- Darcy's problem in $d = 1$:
  - $\frac{d}{dx}(f \frac{d}{dx} u_f)(x) = g(x), \quad x \in (0, 1], \quad u_f(0) = 0$
  - $f(x) = \frac{Kg(x)+f(0)u_f'(0)}{u_f'(x)}$, where K is the Voltera operator.

Approach

Idea: Remove non-linearity imposed by boundary condition and recover the inverse of $\mathcal{L}$.

- Let K be the solution operator $u \mapsto Ku$ of

$$\begin{cases} \mathcal{L}Ku = u & \text{on } \mathcal{O}, \\ Ku = 0 & \text{on } \partial\mathcal{O}. \end{cases} \tag{1}$$

- For the given boundary function g, let $\tilde{g} : \mathcal{O} \to \mathbb{R}$ satisfy

$$\begin{cases} \mathcal{L}\tilde{g} = 0 & \text{on } \mathcal{O}, \\ \tilde{g} = g & \text{on } \partial\mathcal{O}. \end{cases} \tag{2}$$

Approach (cont)

Note: $K\mathcal{L}u_f + \tilde{g}$ is a solution of the forward problem

- $\mathcal{L}(K\mathcal{L}u_f + \tilde{g}) = \mathcal{L}K(\mathcal{L}u_f) + 0 = \mathcal{L}u_f$, on $\mathcal{O}$.
- $K\mathcal{L}u_f + \tilde{g} = 0 + g$ on $\partial\mathcal{O}$,

Thus we can rewrite the GWN observational model as a linear inverse problem

$$\tilde{Y}^{(n)} := Y^{(n)} - \tilde{g} = K(\mathcal{L}u_f) + \frac{1}{\sqrt{n}}Z.$$

Note: $K\mathcal{L}u_f + \tilde{g}$ is a solution of the forward problem

- $\mathcal{L}(K\mathcal{L}u_f + \tilde{g}) = \mathcal{L}K(\mathcal{L}u_f) + 0 = \mathcal{L}u_f$, on $\mathcal{O}$.
- $K\mathcal{L}u_f + \tilde{g} = 0 + g$ on $\partial\mathcal{O}$,

Thus we can rewrite the GWN observational model as a linear inverse problem

$$\tilde{Y}^{(n)} := Y^{(n)} - \tilde{g} = K(\mathcal{L}u_f) + \frac{1}{\sqrt{n}}Z.$$

Bayesian solution:

- Step 1: Solve the linear inverse problem by putting a prior on $\mathcal{L}u_f$
- Step 2: u_f is determined by $\mathcal{L}u_f$ and g.
- Step 3: f is determined by $\mathcal{L}u_f$ and u_f

Model (Schrödinger equation):

$$\begin{cases} \frac{1}{2}\Delta u_f = f u_f & \text{on } \mathcal{O}, \\ u_f = g & \text{on } \partial\mathcal{O}. \end{cases}$$

Model (Schrödinger equation):

$$\begin{cases} \frac{1}{2}\Delta u_f = f u_f & \text{on } \mathcal{O}, \\ u_f = g & \text{on } \partial\mathcal{O}. \end{cases}$$

Lemma There exists a continuous linear operator K on $L^2(\mathcal{O})$ such that

$$\begin{cases} \Delta K h = h & \text{on } \mathcal{O}, \\ K h = 0 & \text{on } \partial\mathcal{O}. \end{cases}$$

If $\mathcal{O}$ is sufficiently regular, and $g : \partial\mathcal{O} \to \mathbb{R}$ is continuous, then there exists a function $\tilde{g} : \mathcal{O} \to \mathbb{R}$ satisfying

$$\begin{cases} \Delta \tilde{g} = 0 & \text{on } \mathcal{O}, \\ \tilde{g} = g & \text{on } \partial\mathcal{O}. \end{cases}$$

In particular, this is true for $\mathcal{O} = [0,1]^d$.

Demonstration: Schrödinger equation (cont)

Hence $u = Kv + \tilde{g}$ solves the problem

$$\begin{cases} \Delta u = v & \text{on } [0,1]^d, \\ u = g & \text{on } \partial[0,1]^d. \end{cases}$$

Demonstration: Schrödinger equation (cont)

Hence $u = Kv + \tilde{g}$ solves the problem

$$\begin{cases} \Delta u = v & \text{on } [0,1]^d, \\ u = g & \text{on } \partial[0,1]^d. \end{cases}$$

Method:

- Step 1: We estimate $v = \Delta u_f$ from the observation model with $\tilde{Y}^{(n)} = Y^{(n)} - \tilde{g}$

$$\tilde{Y}^{(n)} = Kv + n^{-1/2} Z$$

- Step 2: Recover $u_f = Kv + \tilde{g}$.
- Step 3: Recover f from $u_f = Kv + \tilde{g}$ and $\Delta u_f = v$

$$f_v := \begin{cases} \frac{v}{2(Kv+\tilde{g})} & \text{if} \quad Kv + \tilde{g} > 0, \\ 0 & \text{otherwise.} \end{cases}$$

- Using the spectral decomposition of $K^\intercal K$ we can write the model in the sequence form

$$Y_i = \kappa_i v_{0,i} + \frac{1}{\sqrt{n}} Z_i, \qquad i = 1, 2, ...,$$

where we assume that $\kappa_i^2 \asymp i^{-2p/d}$.

- Take prior $v_i \overset{ind}{\sim} N(0, i^{-1-2\alpha/d})$.
- Using the posterior for v and the fixed g compute the posterior for $u = Kv + \tilde{g}$.
- Using the posterior for v and u compute the posterior for $f = e(v, u)$.

Bayesian approach

- Using the spectral decomposition of $K^\intercal K$ we can write the model in the sequence form

$$Y_i = \kappa_i v_{0,i} + \frac{1}{\sqrt{n}} Z_i, \qquad i = 1, 2, \ldots,$$

where we assume that $\kappa_i^2 \asymp i^{-2p/d}$.

- Take prior $v_i \overset{ind}{\sim} N(0, i^{-1-2\alpha/d})$.
- Using the posterior for v and the fixed g compute the posterior for $u = Kv + \tilde{g}$.
- Using the posterior for v and u compute the posterior for $f = e(v, u)$.

Adaptation:

- **Empirical Bayes:** plug in the maximum marginal likelihood estimator (MMLE) $\hat{\alpha}$ into the posterior
- **Hierarchical Bayes:** endow α with hyper-prior

$$c_4^{-1} \alpha^{-c_3} \exp^{-c_2 \alpha} \leq \lambda(\alpha) \leq c_4 \alpha^{-c_3} \exp^{-c_2 \alpha}$$

Notations:

- $(H^\beta, \langle .,. \rangle_{H^\beta})$: Hilbert scale defined by the SVD basis

Theorem (Contraction SVD) Let $f_0 \in L^2([0,1]^d)$, with $\mathcal{L}u_{f_0} \in H^\beta$, for $\beta > d/2$. Assume that $\|Kv\|_{L_2} \asymp \|v\|_{H^{-p}}$ and take $c(u,f) = uf$. Then for some large enough $M > 0$,

$$\Pi\left(f : \|f - f_0\|_{L^2} \geq M\epsilon_n | Y^{(n)}\right) \xrightarrow{P} 0, \qquad (3)$$

where $\epsilon_n = n^{-\alpha\wedge\beta/(d+2\alpha+2p)}$ if the regularity of the prior is chosen deterministically and $\epsilon_n = L_n n^{-\beta/(d+2\beta+2p)}$ for a slowly varying sequence L_n, if α is chosen by the empirical Bayes or hierarchical Bayes methods.

Proof idea

- By algebraic manipulations:

$$\|f - f_0\|_{L_2} \lesssim \left(\inf u_{f_0} - \|u_f - u_{f_0}\|_{L_\infty}\right)^{-1} \|v - \mathcal{L}u_{f_0}\|_{L_2}$$

- L_2-contraction rates for v: explicit computations in the sequence model (using the SVD basis), see Knapik et al. (2011)
- Supremum-norm consistency for u_f, required for Lipschitz of e, by direct computations.

- **Adaptation:** For EB show that $\hat{\alpha} \in [\underline{\alpha}_n, \overline{\alpha}_n]$ then uniform minimax contraction over $[\underline{\alpha}_n, \overline{\alpha}_n]$ using chaining arguments, see Knapik et al. (2016), Rousseau & Sz. (2018). HB follows from EB.

Idea: Use the credible sets of the posterior for $v = \mathcal{L}u_f$ to construct credible sets for f.

Credible sets:

- Let us consider credible balls

$$\tilde{C}_\gamma = \{v : \|v - \hat{v}\|_2 \leq \rho_n, \ \|v - \hat{v}\|_{L_\infty} \leq \varepsilon\},$$

  where $\hat{v}$ denotes the posterior mean and the radius ρ_n is chosen to be the smallest values such that $\Pi(v : \ v \in \tilde{C}_\gamma | X^{(n)}) \geq 1 - \gamma$.
- This defines a credible set on the function space

$$C_\gamma = \{f = e(v, u) : \ v \in \tilde{C}_\gamma\}$$

Credible set: coverage

Theorem (coverage) Assume that the conditions of Theorem (Contraction SVD) hold and take $\alpha = \beta - C/\log n$, for some large enough $C > 0$. Then

$$P_{f_0}(f_0 \in C_\gamma) \to 1,$$

and there exist a constant $M > 0$ such that

$$\|C_\gamma\|_{L_2} = \sup_{f,g \in C_\gamma} \|f - g\|_{L_2} \leq O_p(n^{-\beta/(2\beta+2p+d)}).$$

Wavelet priors

Problem: SVD often hard to compute, complex form and might vanishes at boundary (e.g. Schrödinger case).

Problem: SVD often hard to compute, complex form and might vanishes at boundary (e.g. Schrödinger case).

Wavelet priors:

- $\psi_{j,k}$, $j \in \mathbb{N}$, $k \in \{1, \ldots, 2^j\}$ an orthogonal wavelet basis of $[0,1]^d$.
- Gaussian weights: $G_{j,k} \sim N(0, 2^{-j(2\alpha/d+1)})$ be independent.
- Wavelet prior: $G = \sum_{j=0}^{\infty} \sum_{k=1}^{2^{jd}} G_{j,k}\psi_{j,k}$.

Problem: SVD often hard to compute, complex form and might vanishes at boundary (e.g. Schrödinger case).

Wavelet priors:

- $\psi_{j,k}$, $j \in \mathbb{N}$, $k \in \{1, \ldots, 2^j\}$ an orthogonal wavelet basis of $[0,1]^d$.
- Gaussian weights: $G_{j,k} \sim N(0, 2^{-j(2\alpha/d+1)})$ be independent.
- Wavelet prior: $G = \sum_{j=0}^{\infty} \sum_{k=1}^{2^{jd}} G_{j,k}\psi_{j,k}$.

Theorem (Contraction wavelets): Under the conditions Theorem (Contraction SVD) for the wavelet prior given above with $\alpha > d/2$ there exists $M > 0$ large enough, such that for $\varepsilon_n = n^{-\alpha\wedge\beta/(2\alpha+2p+d)}$

$$\Pi\left(f : \|f - f_0\|_{L_2} > M\varepsilon_n | Y^{(n)}\right) \overset{P_{f_0}}{\to} 0.$$

Problem: The basis of the prior and the operator K does not match $\Rightarrow$ can not use analytic computations in sequence model.

Galerkin projection:

$$\begin{array}{ccc} H \ni f & \xrightarrow{K} & Kf \in L_2(\mathcal{O}) \\ & & \Big\downarrow Q_j \\ H \supset V_j \ni f^{(j,n)} & \xleftarrow{K^{-1}} & Q_j Kf \in W_j \subset L_2(\mathcal{O}) \end{array},$$

where $V_j \subset H$ j-dimensional, $W_j = KV_j$, $Q_j : L_2(\mathcal{O}) \mapsto W_{j_n}$.

Remark: $K^{-1}Q_jK$ operator is close to P_j, if the operator is K is p-smoothing wrt H^β defined by $\psi_{\ell,k}$.

Proof idea II

- **Linear inverse problem:** ε_n contraction rates using Galerkin approach combined with standard testing arguments, see Yan et al (2021)

$$\Pi\left(v : \|v - \mathcal{L}u_{f_0}\|_{L_2} \geq M\epsilon_n \mid Y^{(n)}\right) \to 0$$

Proof idea II

- **Linear inverse problem:** ε_n contraction rates using Galerkin approach combined with standard testing arguments, see Yan et al (2021)

$$\Pi\left(v : \|v - \mathcal{L}u_{f_0}\|_{L_2} \geq M\epsilon_n \mid Y^{(n)}\right) \to 0$$

- For $c(u, f) = uf$ the operator e is Lipschitz:
    - $\|f - f_0\|_{L_2} \lesssim (\inf u_{f_0} - \|u_f - u_{f_0}\|_{L_\infty})^{-1} \|v - \mathcal{L}u_{f_0}\|_{L_2}$
    - By Sobolev embedding (assuming that H^γ is equivalent with S^γ, for some $d/2 < \gamma < \beta + p$):

$$\|u_f - u_{f_0}\|_\infty = \|Kv - Kv_0\|_\infty \lesssim \|Kv - Kv_0\|_{S^\gamma} \asymp \|Kv - Kv_0\|_{H^\gamma} \asymp \|v - v_0\|_{H^{\gamma-p}}.$$

    - Using Galerkin approach to derive smoothness norm contraction for v.

Obervational model: $Y_i = u_f(x_i) + Z_i, \quad Z_i \overset{iid}{\sim} N(0,1), \quad i = 1,\dots,n.$

Discrete observations

Obervational model: $Y_i = u_f(x_i) + Z_i, \quad Z_i \overset{iid}{\sim} N(0,1), \quad i = 1, \ldots, n.$

Linearization: $\tilde{Y}_i := Y_i - \tilde{g}(x_i) = K(\mathcal{L}u_f)(x_i) + Z_i, \qquad i = 1, ..., n$

Discrete observations

Obervational model: $Y_i = u_f(x_i) + Z_i, \quad Z_i \overset{iid}{\sim} N(0,1), \quad i = 1, \ldots, n.$

Linearization: $\tilde{Y}_i := Y_i - \tilde{g}(x_i) = K(\mathcal{L}u_f)(x_i) + Z_i, \qquad i = 1, ..., n$

Problem: K acts on continuous functions, but we are in discrete observational scheme.

Solution: use interpolation techniques by mapping discrete signals to continuous domain. Let $\tilde{W}_n \subset L_2$ satisfy

- $\|w\|_{L_2} \asymp \|w\|_{\mathbb{L}_n}$, for $w \in \tilde{W}_n$
- For all $f \in H^s$ there exists $\mathcal{I}_n f \in \tilde{W}_n$, such that

$$\|f - \mathcal{I}_n f\|_{L_2} \leq \delta_{n,s}\|f\|_{H^s}.$$

Theorem (Contraction regression) Under the conditions of Theorem (Contraction SVD) and the above interpolation assumptions with $\delta_{n,\beta} = cn^{-\frac{\beta}{2\alpha+2p+d}}$, there exists large enough $M > 0$, such that for the Gaussian prior with hyper-parameter $\alpha > d/2$,

$$\Pi_n\left(f : \|f - f_0\|_{L_2} > M\epsilon_n | \tilde{Y}_1, \ldots, \tilde{Y}_n\right) \overset{P_{f_0}^{(n)}}{\to} 0, \tag{4}$$

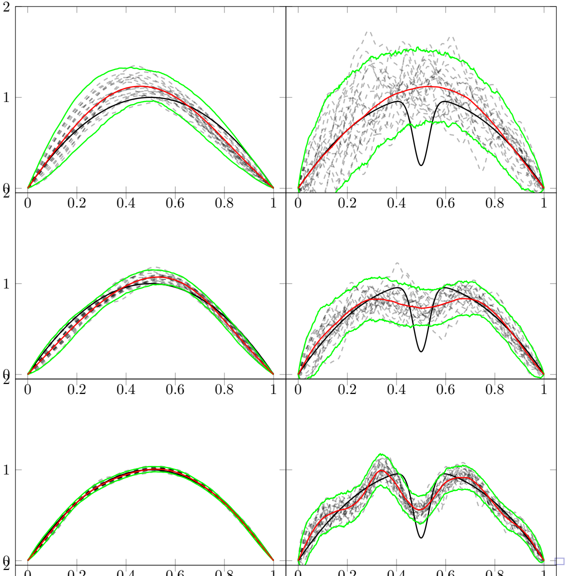
with $\epsilon_n = n^{-(\alpha\wedge\beta)/(d+2\alpha+2p)}$.

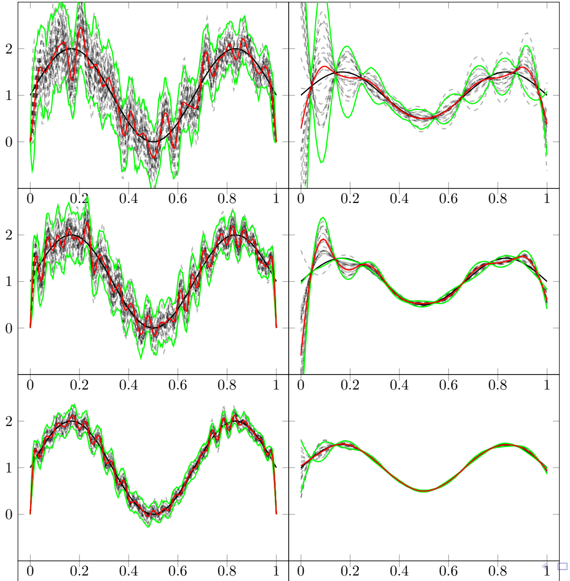
Corollary: Let us consider the time-independent Schrödinger equation corrupted with Gaussian white noise observations and assume that $f_0 \in S^\beta$, for $\beta > d/2$. Let us endow Δu_f with the SVD Gaussian prior, for $\alpha > d/2$. Then there exists a large enough constant $M > 0$

$$\Pi\left(f : \|f - f_0\|_{L_2} \geq M\epsilon_n | Y^{(n)}\right) \xrightarrow{P} 0,$$

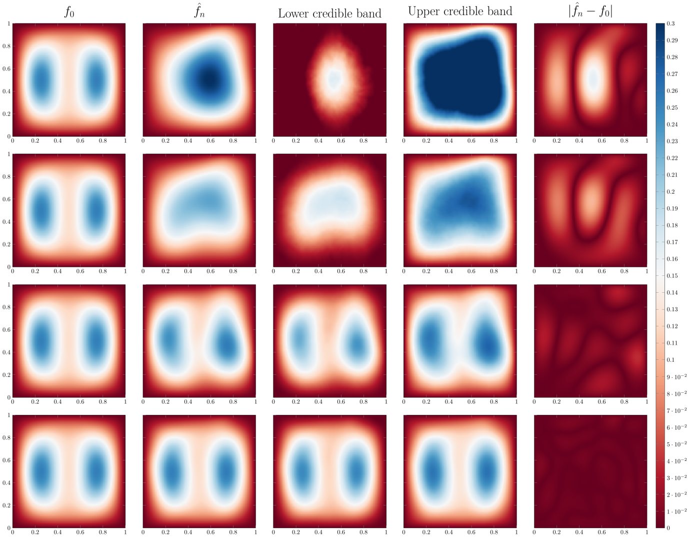
where $\epsilon_n = n^{-\alpha\wedge\beta/(d+2\alpha+4)}$ if the regularity of the prior is chosen deterministic and equal to $\alpha > d/2$ and $\epsilon_n = L_n n^{-\beta/(d+2\beta+4)}$ for a slowly varying sequence L_n, if α is chosen by the empirical Bayes or hierarchical Bayes methods. Furthermore, the credible set for $\alpha \approx \beta$ has frequentist coverage tending to one and minimax optimal radius. The same contraction rate results for fixed $\alpha > d/2$ hold in the discrete observational scheme or considering the wavelet prior.

Numerical Analysis I




2
1
0
0 0.2 0.4 0.6 0.8 1

Numerical Analysis III



Ongoing/Future work

- Extend to other non-linear inverse problems, e.g. Darcy's problem for $d \geq 2$, non Abelian X-ray transform,...
- BvM for linear functionals.
- Scale up further the algorithm, e.g. variational or distributed algorithms.

Conclusion

- Bayesian methods are becoming increasingly popular for inverse problems.
- Theoretical underpinning start to emerge even for non-linear inverse problems.
- Although polynomial time algorithms are possible, they still do not scale well.
- Proposed a new approach by first solving a linear inverse problem and then transforming the posterior.
- Theoretical guarantees and relatively fast algorithm.
- Number of models are limited, investigate extensions.