Linear Methods for Nonlinear Inverse Problems

Aad van der Vaart

TU Delft, Netherlands

based on joint work with

Geerten Koers and Botond Szabo

April 2024

Columbia University

INVERSE PROBLEMS

$Y_n = u_f + \frac{1}{\sqrt{n}}\dot{\mathbb{W}}$, $\dot{\mathbb{W}}$ white noise (or regression)

u_f solution to a PDE that depends on parameter f

"inverse" $u_f \mapsto f$ not differentiable

"nonlinear" $f \mapsto u_f$ nonlinear (e.g. $\frac{1}{2}\Delta u_f = f u_f$)

INVERSE PROBLEMS

$Y_n = u_f + \frac{1}{\sqrt{n}}\dot{\mathbb{W}}$, $\dot{\mathbb{W}}$ white noise (or regression)

u_f solution to a PDE that depends on parameter f

"inverse" $u_f \mapsto f$ not differentiable

"nonlinear" $f \mapsto u_f$ nonlinear (e.g. $\frac{1}{2}\Delta u_f = f u_f$)

Regularisation $\hat{f} = \operatorname*{argmin}_f \Big[\|Y_n - u_f\|^2 + \lambda^2 \mathrm{pen}(f)\Big]$

Bayesian Regularisation $f \sim$ prior

$\longrightarrow$ posterior distribution $f|\,Y_n$

$\longrightarrow$ posterior mean $\mathrm{E}(f|\,Y_n)$

Lem If $f \sim$ Gaussian with RKHS $\mathbb{H}$, then

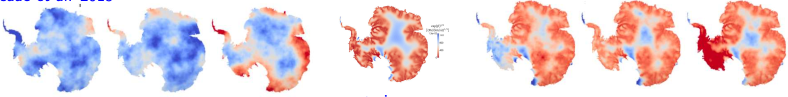
$$\text{posterior mode} = \operatorname*{argmin}_f \Big[\|Y_n - u_f\|^2 + \sigma^2 \|f\|_{\mathbb{H}}^2\Big]$$

Posterior $f|\,Y_n$ also has spread $\longrightarrow$ uncertainty quantification

INVERSE PROBLEMS – EXAMPLES

- Schrödinger [Nickl 2020] $\frac{1}{2}\Delta u = fu$
- Heat with absorption [Kekkonen 2022] $\partial_t u - \frac{1}{2}\Delta u = fu$
- X-ray transform [Monard Nickl Paternain 2019, 2021] $\log u(x,v) = \int_0^{\tau v} f(x+tv)\,dt$
- Divergence/Darcy [Abraham Nickl 2019, Bohr 2022] $\nabla \cdot (f \nabla u) = g$
- Navier-Stokes [Nickl Titi 2023] $\begin{cases} \partial_t u_t - \nu \Delta v + u \cdot \nabla u + \nabla p = 0, \\ \nabla \cdot u = 0, \\ u(0,\cdot) = f \end{cases}$

Isaac et al. 2015



3 draws from prior posterior mean 3 draws from posterior

A.M. Stuart, 2010. Inverse problems: a Bayesian perspective

R. Nickl, 2023. Bayesian non-linear statistical inverse problems

NONPARAMETRIC BAYESIAN INFERENCE

- $f \sim \Pi$ prior distribution Π
- $Y_n | f \sim p_{n,f}$ likelihood
- $f | Y_n$ posterior distribution $\Pi_n(\cdot | Y_n)$

Bayes rule: $d\Pi_n(f | Y_n) \propto p_{n,f}(Y_n)\, d\Pi(f)$

Computation

- MCMC (e.g. preconditioned Crank-Nicolson)
- Distributed posterior

Frequentist Bayesian theory

- Contraction rate
  Fastest $\varepsilon_n \downarrow 0$ so that $\mathrm{E}_{f_0}\Pi_n\big(f: \|f - f_0\| \lesssim \varepsilon_n | Y_n\big) \to 1$?
- Uncertainty quantification
  If $\Pi_n\big(C_n(Y_n) | Y_n\big) \geq 0.95$, then $\mathrm{P}_{f_0}\big(f_0 \in C_n(Y_n)\big) \gg 0$?

CONTRACTION RATE

Fastest $\varepsilon_n \downarrow 0$ so that $\mathrm{E}_{f_0}\Pi_n\big(f\colon \|f - f_0\| \lesssim \varepsilon_n \mid Y_n\big) \to 1$?

- Depends on combination of prior and f_0
- Good prior gives contraction at minimax rate
- Hierarchical prior or prior with data-based bandwidth gives adaptive contraction rate

Benchmark rate for (inverse) curve fitting:

A function f of d variables with bounded derivatives of order β is estimable based on n observations at rate

$$\varepsilon_{n,\beta} := n^{-\beta/(2\beta+d+2p)}$$

A good prior gives a posterior such that

$$\forall \beta\colon \quad \forall f_0 \in \mathcal{F}_\beta\colon \quad \mathrm{E}_{f_0}\Pi_n\big(f\colon \|f - f_0\| \lesssim \varepsilon_{n,\beta} \mid Y_n\big) \to 1$$

CONTRACTION IN GAUSSIAN REGRESSION

Centered small ball exponent

For Gaussian process W in Banach space:

$$\phi_0(\varepsilon) = -\log \mathrm{P}\big(\|W\| < \varepsilon\big).$$

Decentering function

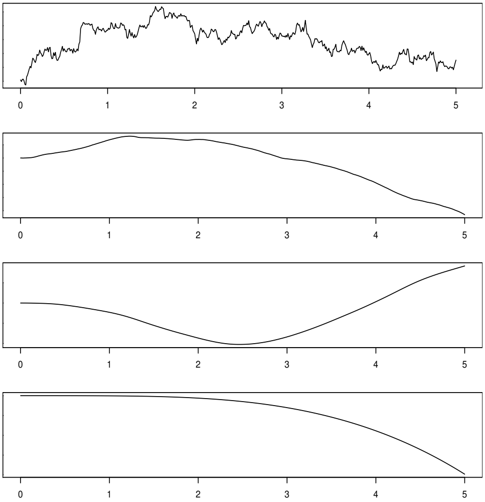
For Gaussian process in Banach space with RKHS $\mathbb{H}$ and given f_0:

$$D(\varepsilon; f_0) = \inf_{h \in \mathbb{H}: \|h - f_0\| < \varepsilon} \|h\|_{\mathbb{H}}^2$$

Thm Contraction rate with prior $f \sim W$ for nonparametric regression $Y_n = f + \frac{1}{\sqrt{n}}\dot{\mathbb{W}}$ is ε_n if

$$\phi_0(\varepsilon_n) + D(\varepsilon_n; f_0) \lesssim n\varepsilon_n^2$$

EXAMPLE: INTEGRATED BROWNIAN MOTION



$x \mapsto B_x$

$x \mapsto (I^1 B)_x = \int_0^x B_s \, ds$

$x \mapsto (I^2 B)_x$

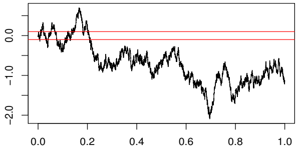
$x \mapsto (I^3 B)_x$

0, 1, 2 and 3 times integrated Brownian motion

EXAMPLE: INTEGRATED BROWNIAN MOTION

k times integrated Brownian motion, as map in $C[0,1]$ has

- $\mathbb{H} = H_0^{k+1}[0,1]$
- $\phi_0(\varepsilon) \asymp (1/\varepsilon)^{2/(2k+1)}$ relative to uniform norm
- $D(\varepsilon; f_0) \asymp (1/\varepsilon)^{(2k+2-2\beta)/\beta}$ if $f_0 \in C^\beta[0,1]$



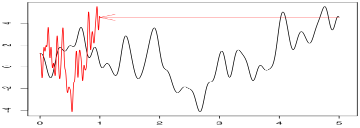
$$\Pi\big(\|W\| < \varepsilon\big) = e^{-\phi_0(\varepsilon)}$$

Thm Contraction rate $n^{-(\beta\wedge(k+1/2))/(2k+2)}$ if $f_0 \in C_0^\beta[0,1]$

$k + 1/2 < \beta \to$ prior too rough, $\phi_0(\varepsilon)$ dominates
$k + 1/2 = \beta \to$ prior optimal
$k + 1/2 > \beta \to$ prior too smooth, $D(\varepsilon; f_0)$ dominates

LENGTH SCALE – ADAPTATION

Priors can be made to adapt to unknown smoothness by changing the length scale: use prior $x \mapsto W_{\tau x}$ for f instead of $x \mapsto W_x$



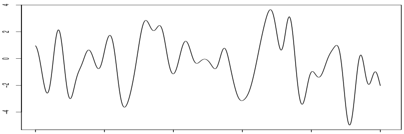
$\tau \gg 1$

Length scale τ can be data-based by $(W_{\tau\cdot} \sim \Pi_\tau)$

- Full Bayes: ordinary Bayes with $\tau \sim$ prior and $f \,|\, \tau \sim \Pi_\tau$
- Empirical Bayes: $\tau = \underset{\tau}{\mathrm{argmax}} \int p_{n,f}(Y_n)\, d\Pi_\tau(f)$

"Thm" Optimal contraction rate attained if $f_0 \in C^\beta[0,1]^d$ and $\beta \leq$ prior smoothness $+1/2$

EXAMPLE: SQUARE EXPONENTIAL PRIOR



$\tau^2 = 1$

$$\mathrm{cov}(W_x, W_y) = e^{-\|x-y\|^2}$$

$$\phi_0(\varepsilon) \asymp \Big(\log \frac{1}{\varepsilon}\Big)^{1+d/2}$$

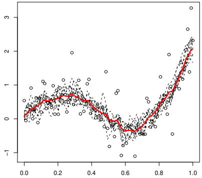
Thm If $\tau^d \sim \Gamma(a,b)$,

- if $f_0 \in C^\beta[0,1]^d$, then contraction rate nearly $n^{-\beta/(2\beta+d)}$
- if f_0 is analytic, then contraction rate nearly $n^{-1/2}$

UNCERTAINTY QUANTIFICATION

If $\Pi_n\big(C_n(Y_n)|\, Y_n\big) \geq 0.95$, then $\mathrm{P}_{f_0}\big(f_0 \in C_n(Y_n)\big) \gg 0$?

- Spread of posterior should indicate "remaining uncertainty"
- Visualisation through draws from the posterior
- Credible sets are "Bayesian confidence sets" (?)



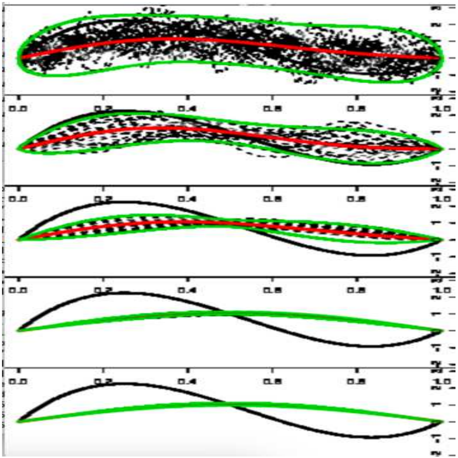
A good prior gives a posterior with credible sets of high coverage

UNCERTAINTY QUANTIFICATION with GAUSSIAN PRIORS

Two very different cases:

- Fixed length scale
  Credible sets cover if and only if prior undersmoothes truth
- Data-driven length scale
  Credible sets cover "most but not all truths"

FIXED LENGTH SCALE: COVERAGE REQUIRES UNDERSMOOTHING



True f_0 (black), posterior mean (red)

- $f(x) = \sum_{i=1}^{\infty} f_i e_i(x)$
- Truth:
  $f_{0,i} \asymp i^{-1-2\beta}$
- Prior:
  $f_i \overset{\text{ind}}{\sim} N(0, i^{-1-2\alpha})$

Top to bottom: increasing α

Black: truth

Green: bands

DATA-DRIVEN LENGTH SCALE: TRICKED BY INCONVENIENT TRUTHS

Prior on length scale τ gives adaptive recovery:
faster contraction for smoother f_0

Theory implies that data-driven posteriors *must be tricked* by some inconvenient truths and sometimes be misleading in their uncertainty quantification

- Estimation: $\forall \beta$: $\forall f \in \mathcal{F}_\beta$: rate $\varepsilon_{n,\beta}$
- Uncertainty: $\forall f \in \cup_\beta \mathcal{F}_\beta : \mathrm{P}_f\big(f \in C_n(Y_n)\big) \geq 0.95$

Inconvenient truths will not be covered.
Example for series prior:

$$f_1, f_2, \ldots, f_{N_1}, 0, 0, \ldots, 0, f_{n_2}, f_{n_2+1}, \ldots, f_{N_2}, 0, 0, \ldots, 0, f_{n_3}, \ldots, f_{N_3}, 0, \ldots$$

(Lengths of zero runs $0, 0, \ldots, 0$ are increasing)

CONVENIENT TRUTHS for SERIES PRIORS

Def $f = \sum_{i=1}^{\infty} f_i e_i$ satisfies the *polished tail condition* if

$$\sum_{i=N}^{1000N} f_i^2 \geq 0.001 \sum_{i=N}^{\infty} f_i^2, \qquad \forall \text{ large } N.$$

Polished tail sequences are "generic"

Non-polished tail series are "rare"

"Thm" Slighly enlarged credible sets of posteriors with data-driven length scale cover convenient truths

CONVENIENT TRUTHS for SERIES PRIORS – details

Def $f = \sum_{i=1}^\infty f_i e_i$ satisfies the *polished tail condition* if

$$\sum_{i=N}^{1000N} f_i^2 \geq 0.001 \sum_{i=N}^{\infty} f_i^2, \qquad \forall \text{ large } N.$$

"Everything" is polished tail...:

- For the *topologist* [Giné,Nickl 2010]
  Non polished tail sequences are meagre in a natural topology
- For the *minimax expert*:
  Intersecting the usual models with polished tail sequences decreases the minimax risk by at most a logarithmic factor
- For the *Bayesian*:
  Every f from a prior $f_i \overset{\text{ind}}{\sim} N(0, ci^{-\alpha-1/2})$ is polished tail

"Thm" If f_0 polished tail, M large, α set by FB or EB, then $\mathrm{P}_{f_0}\big(f_0 \in C_n(Y_n, M\hat{r}_n)\big) \to 1$ for $C_n(Y_n, \hat{r}_n)$ credible ball.

INVERSE PROBLEMS

In direct nonparametric problems, as regression or density estimation, the prior regularises the parameter

- Contraction rates known for many priors
- Credible sets are less studied

$Y_n = u_f + \frac{1}{\sqrt{n}}\dot{\mathbb{W}},$ $\dot{\mathbb{W}}$ white noise (or regression)

u_f solution to a PDE that depends on parameter f

In inverse problems the prior must regularise both the parameter and the inverse map $u_f \mapsto f$

- Much remains to be studied

LINEAR FOR NONLINEAR

$Y_n = u_f + \frac{1}{\sqrt{n}}\dot{\mathbb{W}},$ $\dot{\mathbb{W}}$ white noise (or regression)

Assume $\begin{cases} \mathcal{L}u_f = c(f, u_f), & \text{on } \mathcal{O} \subset \mathbb{R}^d, \\ u_f = g, & \text{on } \Gamma \subseteq \partial\mathcal{O} \end{cases}$

$f = e(\mathcal{L}u_f)$

- Recover $\mathcal{L}u_f$ from $Y_n = K(\mathcal{L}u_f) + \frac{1}{\sqrt{n}}\dot{\mathbb{W}},\quad K = \mathcal{L}^{-1}$
- Recover f from $\mathcal{L}u_f$

EXAMPLE

$\Delta u_f = 2fu_f$ $\quad c(f, u_f) = 2fu_f$

$\mathcal{L} = \Delta$ $\quad f = \frac{\Delta u_f}{2u_f} =: e(\Delta u_f)$

LINEAR PROBLEM

$Y_n = u_f + \frac{1}{\sqrt{n}}\dot{\mathbb{W}}$, $\dot{\mathbb{W}}$ white noise (or regression)

$$\tilde{g} \text{ solves } \begin{cases} \mathcal{L}\tilde{g} = 0, & \text{on } \mathcal{O} \subset \mathbb{R}^d, \\ \tilde{g} = g, & \text{on } \Gamma \subseteq \partial\mathcal{O} \end{cases}$$

$$K \text{ solves } \begin{cases} \mathcal{L}Ku = u, & \text{on } \mathcal{O} \subset \mathbb{R}^d, \\ Ku = 0, & \text{on } \Gamma \subseteq \partial\mathcal{O} \end{cases}$$

- $u_f = K\mathcal{L}u_f + \tilde{g}$
- $\tilde{Y}_n := Y_n - \tilde{g} = K\mathcal{L}u_f + \frac{1}{\sqrt{n}}\dot{\mathbb{W}}$

$v = \mathcal{L}u_f$ $\quad f = e(\mathcal{L}u_f) = e(v)$

(L) $\tilde{\Pi}_n(v \in \cdot \mid \tilde{Y}_n)$ based on $\tilde{Y}_n = Kv + \frac{1}{\sqrt{n}}\dot{\mathbb{W}}$

(N) $\Pi_n(f \in \cdot \mid Y_n)$ based on $Y_n = u_f + \frac{1}{\sqrt{n}}\dot{\mathbb{W}}$

CONNECTING LINEAR AND NONLINEAR

$Y_n = K\mathcal{L}u_f + \tilde{g} + \frac{1}{\sqrt{n}}\dot{\mathbb{W}}$, $\dot{\mathbb{W}}$ white noise (or regression)

$v = \mathcal{L}u_f$ $f = e(\mathcal{L}u_f) = e(v)$

(L) $\tilde{\Pi}_n(v \in \cdot \,|\, \tilde{Y}_n)$ based on $\tilde{Y}_n = Kv + \frac{1}{\sqrt{n}}\dot{\mathbb{W}}$

(N) $\Pi_n(f \in \cdot \,|\, Y_n)$ based on $Y_n = u_f + \frac{1}{\sqrt{n}}\dot{\mathbb{W}}$

Prop If $\tilde{\Pi}_n(v \in V_n|\, \tilde{Y}_n) \to 1$ and $e\colon (V_n, \|\cdot\|) \to L_2$ is Lipschitz, then contraction rate for (L) is also contraction rate for (N)

CONNECTING LINEAR AND NONLINEAR

$Y_n = K\mathcal{L}u_f + \tilde{g} + \frac{1}{\sqrt{n}}\dot{\mathbb{W}}$, $\dot{\mathbb{W}}$ white noise (or regression)

$v = \mathcal{L}u_f$ $f = e(\mathcal{L}u_f) = e(v)$

(L) $\tilde{\Pi}_n(v \in \cdot \mid \tilde{Y}_n)$ based on $\tilde{Y}_n = Kv + \frac{1}{\sqrt{n}}\dot{\mathbb{W}}$

(N) $\Pi_n(f \in \cdot \mid Y_n)$ based on $Y_n = u_f + \frac{1}{\sqrt{n}}\dot{\mathbb{W}}$

Prop If $\tilde{\Pi}_n(v \in V_n \mid \tilde{Y}_n) \to 1$ and $e\colon (V_n, \|\cdot\|) \to L_2$ is Lipschitz, then contraction rate for (L) is also contraction rate for (N)

$$C_n(Y_n) := \{f : \mathcal{L}u_f \in \tilde{C}_n(Y_n - \tilde{g})\} = \{e(v) : v \in \tilde{C}_n(Y_n - \tilde{g})\}$$

Prop

- Credible and confidence levels of $C_n(Y_n)$ are those of $\tilde{C}_n(\tilde{Y}_n)$
- If $e\colon V_n \to L_2$ uniformly Lipschitz at $\bar{v}_n \in \tilde{C}_n(Y_n) \subset V_n$, then diameters of $C_n(Y_n)$ and $\tilde{C}_n(\tilde{Y}_n)$ are of same order

CONNECTING LINEAR AND NONLINEAR

$v = \mathcal{L}u_f \qquad f = e(\mathcal{L}u_f) = e(v)$

(L) $\tilde{\Pi}_n(v \in \cdot \mid \tilde{Y}_n)$ based on $\tilde{Y}_n = Kv + \frac{1}{\sqrt{n}}\dot{\mathbb{W}}$

(N) $\Pi_n(f \in \cdot \mid Y_n)$ based on $Y_n = u_f + \frac{1}{\sqrt{n}}\dot{\mathbb{W}}$

- Connection works for every prior on f
- Effective use of linearity easier with "standard" priors on $\mathcal{L}u_f$
- Need good results for linear problems

LINEAR INVERSE PROBLEM

$\tilde{Y}_n = Kv + \frac{1}{\sqrt{n}}\dot{\mathbb{W}}$, $\quad \dot{\mathbb{W}}$ white noise, $\quad K\colon L_2(\mathcal{O}) \to L_2(\mathcal{O})$ linear

$(h_i)_{i\in\mathbb{N}}$ orthonormal basis of $L_2(\mathcal{O})$, $O \subset \mathbb{R}^d$
$\|v\|^2_{G^s} := \sum_{i\in\mathbb{N}} v_i^2 i^{2s/d}$

Assume $\|Kv\|_2 \asymp \|v\|_{G^{-p}}$

Galerkin projection

$$\begin{array}{cccc} G^0 \ni v & \xrightarrow{K} & vf & \in L \\ & & \downarrow Q_j & \\ K^{-1}Q_jKv & \xleftarrow{K^{-1}} & Q_jKv & \in K\,\mathrm{lin}\,(e_1,\ldots,e_j) \end{array}$$

Thm If $\exists j_n \lesssim n\varepsilon_n^2$ and $\eta_n \gtrsim \varepsilon_n j_n^p \vee j_n^{-\beta}$ with

$$\Pi\big(v\colon \|Kv - Kv_0\|_2 < \varepsilon_n\big) \gtrsim e^{-n\varepsilon_n^2},$$
$$\Pi\big(v\colon \|K^{-1}Q_{j_n}Kv - v\|_2 > \eta_n\big) \lesssim e^{-4n\varepsilon_n^2},$$

then $\mathrm{E}_{v_0}\Pi_n\big(v\colon \|v - v_0\|_2 \lesssim \eta_n \,|\, \tilde{Y}_n\big) \to 1$

LINEAR INVERSE PROBLEM

$\tilde{Y}_n = Kv + \frac{1}{\sqrt{n}}\dot{\mathbb{W}}$, $\quad \dot{\mathbb{W}}$ white noise, $\quad K\colon L_2(\mathcal{O}) \to L_2(\mathcal{O})$ linear

$(h_i)_{i\in\mathbb{N}}$ orthonormal basis of $L_2(\mathcal{O})$, $O \subset \mathbb{R}^d$
$\|v\|_{G^s}^2 := \sum_{i\in\mathbb{N}} v_i^2 i^{2s/d}$.

Assume $\|Kv\|_2 \asymp \|v\|_{G^{-p}}$

$$v = \sum_{i\in\mathbb{N}} v_i h_i, \qquad v_i \overset{\text{ind}}{\sim} N(0, i^{-1-2\alpha/d})$$

Thm

- If $v_0 \in G^\beta$, then $\mathrm{E}_{v_0}\Pi_n\Big(v\colon \|v - v_0\|_{G^\delta} \lesssim n^{-\frac{\alpha\wedge\beta-\delta}{2\alpha+2p+d}} \,|\, \tilde{Y}_n\Big) \to 1$,
  for $\delta \in [-p, \alpha\wedge\beta)$
- If also $\sup_i i^{u/d}\|Kh_i\|_\infty < \infty$, some $u > d/2 - \alpha\wedge\beta$, then $\mathrm{E}_{v_0}\Pi_n\Big(v\colon \|Kv - Kv_0\|_\infty \le \varepsilon \,|\, \tilde{Y}_n\Big) \to 1$, $\forall \varepsilon > 0$

LINEAR INVERSE PROBLEM – EIGEN PRIOR

$\tilde{Y}_n = Kv + \frac{1}{\sqrt{n}}\dot{\mathbb{W}}$, $\quad \dot{\mathbb{W}}$ white noise, $\quad K\colon L_2(\mathcal{O}) \to L_2(\mathcal{O})$ linear

$(h_i)_{i\in\mathbb{N}}$ eigen basis of $K^T K$, eigenvalues $\kappa_i^2 \asymp i^{-2p/d}$

$$v = \sum_{i\in\mathbb{N}} v_i h_i, \qquad v_i |\, \alpha \overset{\text{ind}}{\sim} N(0, i^{-1-2\alpha/d})$$
$$\alpha \sim \text{prior or } \alpha = EB - MLE$$

$$\tilde{C}_n(\tilde{Y}_n) = \{v\colon \|v - \hat{v}_n\|_2 \leq r\xi_{n,\gamma}\}$$

Thm

- If $v_0 \in G^\beta$, then $\mathrm{E}_{v_0}\Pi_n\big(v\colon \|v - v_0\|_{G^\delta} \lesssim n^{-\frac{\beta-\delta}{2\beta+2p+d}} \,|\, \tilde{Y}_n\big) \to 1$
- If also $\sup_i i^{u/d}\|Kh_i\|_\infty < \infty$, some $u > d/2 - \beta$, then $\mathrm{E}_{v_0}\Pi_n\big(v\colon \|Kv - Kv_0\|_\infty \leq \varepsilon |\, \tilde{Y}_n\big) \to 1$, $\forall \varepsilon > 0$
- If v_0 polished tail, then $\mathrm{P}_{v_0}\big(v_0 \in \tilde{C}_n(\tilde{Y}_n)\big) \to 1$

EXAMPLE – SCHRÖDINGER

$$\begin{cases} \frac{1}{2}\Delta u_f = f\, u_f & \text{on } \mathcal{O} = (0,1)^d, \\ u_f = g, & \text{on } \Gamma \subseteq \partial\mathcal{O} \end{cases}$$

$\mathcal{L} = \Delta, \qquad e(v) = \frac{v}{Kv+\tilde{g}}$

Eigenbasis $h_{i_1\ldots i_d}(x) = 2^{d/2}\prod_{j=1}^d \sin(i_j \pi x_j), \qquad \kappa_{i_1\ldots i_d} = \frac{1}{(\sum i_j^2)\pi^2}$

Thm If $\Delta u_{f_0} \in G^\beta$, $\beta > d/2$, $u_{f_0} \geq 2c_0 > 0$, $\alpha + 2 > d/2$:

- L_2-contraction rate $n^{-\frac{\alpha\wedge\beta}{2\alpha+4+d}}$ for α-regular prior on Δu_f
- L_2-contraction rate $n^{-\frac{\beta}{2\beta+4+d}} \log^2 n$ for HB/EB-prior
- coverage for α-regular prior with $\alpha \leq \beta - 1/\log n$

EXAMPLE – SCHRÖDINGER

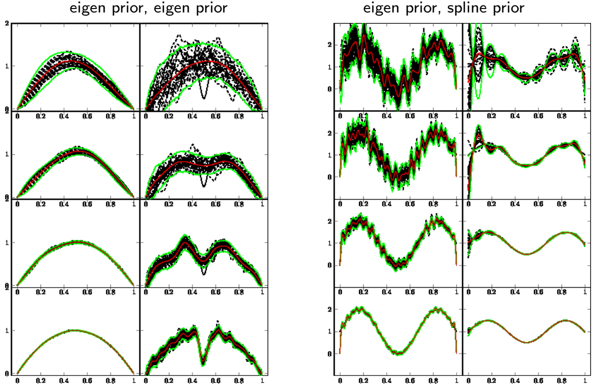
$$\begin{cases} \frac{1}{2}\Delta u_f = f u_f, & \text{on } \mathcal{O} = (0,1), \\ u_f = g, & \text{on } \partial\mathcal{O} \end{cases} \qquad f = e(\Delta u_f) := \frac{\Delta u_f}{2f}$$

$n = 10^4$

$n = 10^6$

$n = 10^8$

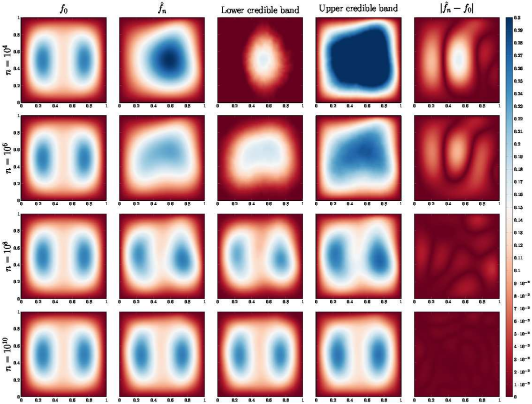
$n = 10^{10}$



black: f_0 red: posterior mean green: 95% credible band dashed: posterior draws

EXAMPLE – SCHRÖDINGER

$$\begin{cases} \frac{1}{2}\Delta u_f = f u_f, & \text{on } \mathcal{O} = (0,1)^2, \\ u_f = g, & \text{on } \partial\mathcal{O} \end{cases} \qquad f = e(\Delta u_f) := \frac{\Delta u_f}{2f}$$



$$f_0(x,y) = 2x(x-1)y(y-1)(2+\sin(3+\pi x)\sin(\pi y)$$

$$g(x,y) = 3 + xy^2 + 2y\sin(2\pi x) + x\cos(3\pi y)$$

OTHER EXAMPLES

Heat with absorption
$$\begin{cases} \frac{\partial}{\partial t}u - \frac{1}{2}\Delta u = fu, & \text{on } (0,1)^d \times (0,1]; \\ u = g, & \text{on } \partial(0,1)^d \times [0,1]; \\ u = u_0, & \text{on } (0,1)^d \times \{0\} \end{cases}$$

$\mathcal{L} = \frac{\partial}{\partial t} - \frac{1}{2}\Delta,$ $e(v) = \frac{v}{Kv+\tilde{g}}$

Exponentiated Volterra
$$\begin{cases} u_f' = fu_f, & \text{on } (0,1), \\ u_f = g, & \text{at } 0 \end{cases}$$

$\mathcal{L}u = u',$ $e(v) = \frac{v}{Kv+g}$ $Kv(x) = \int_0^x v(s)\,ds$

Darcy
$$\begin{cases} \nabla \cdot (f\nabla u_f) = h, & \text{on } \mathcal{O}, \\ u_f = g, & \text{on } \partial O \end{cases}$$

$\mathcal{L} = \Delta$ e numerical

EXAMPLE – HEAT WITH ABSORPTION – details

$$\begin{cases} \dfrac{\partial}{\partial t} u - \dfrac{1}{2}\Delta u = fu, & \text{on } (0,1)^d \times (0,1]; \\ u = g, & \text{on } \partial(0,1)^d \times [0,1]; \\ u = u_0, & \text{on } (0,1)^d \times \{0\} \end{cases}$$

$\mathcal{L} = \frac{\partial}{\partial t} - \frac{1}{2}\Delta, \qquad e(v) = \frac{v}{Kv + \tilde{g}}$

Eigenbasis $h_{i_1 \ldots i_d, k}, \qquad \kappa_{i_1 \ldots i_d, k} \asymp \frac{1}{\sum i_j^2/4 + k^2\pi^2}$

Thm Assume $\mathcal{L}u_{f_0} \in G^\beta$, $\beta > d+1$, $u_{f_0} \geq c_0 > 0$, $\alpha > d+1$

- L_2-contraction rate $n^{-\frac{\alpha \wedge \beta}{2\alpha + 4 + d + 1}}$ for α-regular prior on $\mathcal{L}u_f$
- L_2-contraction rate $n^{-\frac{\beta}{2\beta + 4 + d + 1}} \log^2 n$ for HB/EB-prior
- coverage for α-regular prior with $\alpha \leq \beta - 1/\log n$

EXAMPLE – EXPONENTIATED VOLTERRA – details

$$\begin{cases} u_f' = f u_f, & \text{on } (0,1), \\ u_f = g, & \text{at } 0 \end{cases}$$

$$\mathcal{L}u = u', \qquad e(v) = \frac{v}{Kv+g} \qquad Kv(x) = \int_0^x v(s)\,ds$$

Thm

- L_2-contraction for $v_0 = u_0'$ implies same L_2-contraction at f_0
- coverage and L_2-diameter of credible sets preserved

EXAMPLE – DARCY – details

$$\begin{cases} \nabla \cdot (f \nabla u_f) = h, & \text{on } \mathcal{O}, \\ u_f = g, & \text{on } \partial O \end{cases}$$

$\mathcal{L} = \Delta$

For $d > 1$ there is no explicit solution map $e \colon \Delta u_f \mapsto f$, but for given boundary conditions, if $\Delta u_f + \|\nabla u_f\|^2 \geq c_0 > 0$, then

$$\|f - f_0\|_{L_2(\mathcal{O})} \lesssim \|u_f - u_{f_0}\|_{H^2(\mathcal{O})}$$

Our method can be implemented together with a numerical implementation of the map $e \colon \Delta u_f \to f$

DISTRIBUTED POSTERIOR DISTRIBUTION

To speed up computations:

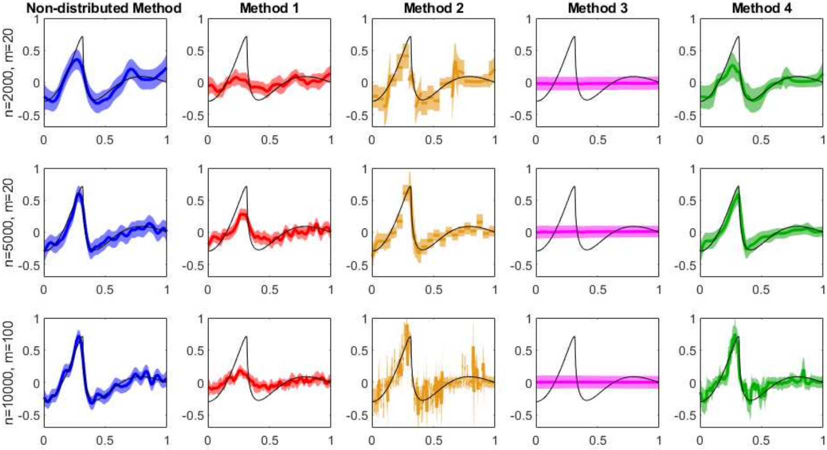
- Divide data in m batches $Y_n^{(j)}$ of size n/m
- On each batch j compute a posterior distribution
- Compute law of $f = m^{-1} \sum_{j=1}^{m} f_j$, for $f_j \overset{\text{ind}}{\sim} \Pi_{n/m}(\cdot \mid Y_n^{(j)})$

Thm

- For prior with fixed length scale contraction rates are retained provided prior variance is cut by $1/m$
- For prior with data-driven length scale adaptation and coverage are retained when using spatially defined batches

(n,m)	(1000, 5)	(5000, 10)	(10000, 100)
Benchmark	14.84s (5.23s)	633.8s (241.2s)	4741s (1069s)
Random	2.37s (1.14s)	24.7s (5.1s)	22.0s (6.1s)
Spatial	2.55s (1.11s)	25.2s (5.4s)	22.6s (5.8s)

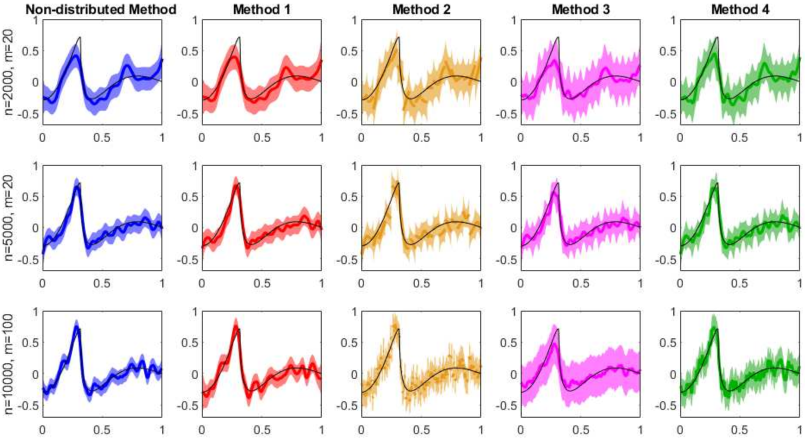
DISTRIBUTED POSTERIOR DISTRIBUTION – ADAPTIVE BANDWIDTH



Method 1: random data split
Method 2: spatial data split
Methods 3,4: spatial data split + smoothing

black: f_0
color: posterior mean + bands

DISTRIBUTED POSTERIOR DISTRIBUTION – FIXED BANDWIDTH



Method 1: random data split
Method 2: spatial data split
Methods 3,4: spatial data split + smoothing

black: f_0
color: posterior mean + bands

DISCUSSION

Our idea is to isolate a linear operator $\mathcal{L}$ so that, for a nice e,

$$f = e(\mathcal{L}u_f)$$

Statistical problem becomes linear
Nonlinear inversion becomes deterministic

- Is this is always possible?
- Numerical methods to compute e may be expensive?
- Works best if a standard prior is placed on u_f?
- Good posterior behaviour of smooth functionals?

Inverse problems in general:

- Uncertainty quantification?
- Boundary conditions?

CREDITS

Thank you

and thanks to co-authors:

Subhashis Ghoshal

Harry van Zanten

Botond Szabo

Bartek Knapik

Suzanne Sniekers

Dong Yan

Amine Hadji

Geerten Koers