Posterior consistency of the Bayesian approach to linear ill-posed inverse problems

Sergios Agapiou

Mathematics Institute, University of Warwick, Coventry, UK

Joint work with Stig Larsson (Göteborg) and Andrew Stuart (Warwick)

NSF-CBMS Conference, Small Deviation Probabilities: Theory and Applications, 4-8 June 2012, University of Alabama in Huntsville

Inverse Problems

- Inverse Problems are concerned with determining causes for a desired or an observed effect.
- $\mathcal{A} : \mathcal{D}(\mathcal{A}) \subset \mathcal{X} \to \mathcal{X}$ known self-adjoint, positive definite linear operator with bounded inverse, $\mathcal{X}$ separable Hilbert space.
- Linear inverse problem: find u from y, where y noisy observation of $\mathcal{A}^{-1}u$.
- Model:

$$y = \mathcal{A}^{-1}u + \frac{1}{\sqrt{n}}\xi, \tag{1}$$

$\frac{1}{\sqrt{n}}\xi$ additive noise.

Inverse Problems are Ill-Posed - Deterministic Approach

- Problem (1) ill-posed:
  - existence of solution issues
  - solutions do not depend continuously on the data

- Tikhonov-Phillips Regularization: u approximated by minimizer of

$$J_0(u) := \frac{1}{2}\left\|\mathcal{C}_1^{-\frac{1}{2}}(y - \mathcal{A}^{-1}u)\right\|^2 + \frac{\lambda}{2}\left\|\mathcal{C}_0^{-\frac{1}{2}}u\right\|^2,$$

$\mathcal{C}_i : \mathcal{X} \to \mathcal{X}$, self-adjoint, possibly compact, positive definite linear operators.

- λ regularization parameter, appropriate function of noise level $n^{-\frac{1}{2}}$ which shrinks to zero as $n \to \infty$ to recover unknown u.

Bayesian Approach

Bayesian framework: assume $\xi \sim \mathcal{N}(0, \mathcal{C}_1)$, $\mathcal{C}_1 : \mathcal{X} \to \mathcal{X}$ selfadjoint positive definite.

- Likelihood: for fixed u, $y|u \sim \mathcal{N}(\mathcal{A}^{-1}u, \frac{1}{n}\mathcal{C}_1)$.
- Prior: choose *prior* distribution for unknown u, encoding prior knowledge. Let $u \sim \mathcal{N}(0, \tau^2\mathcal{C}_0)$, $\mathcal{C}_0 : \mathcal{X} \to \mathcal{X}$ selfadjoint positive definite trace class.
- Posterior: in Bayesian Approach, solution of problem (1) is the distribution of $u|y$, called the *posterior* distribution μ^y.

Bayesian Approach via Precision Operators - Intuition

- Link: Bayes rule:

$$P(u|y) \propto P(y|u)P(u)$$

- Assume $\mathcal{X} = \mathbb{R}^d$,

$$\pi^y(u) \propto \exp\left(-nJ_0(u)\right)$$

$$nJ_0(u) = \frac{n}{2}\left\|\mathcal{C}_1^{-\frac{1}{2}}(y - \mathcal{A}^{-1}u)\right\|^2 + \frac{1}{2\tau^2}\left\|\mathcal{C}_0^{-\frac{1}{2}}u\right\|^2.$$

- This suggests that $u|y$; complete the square to find mean and covariance.

Bayesian Approach via Precision Operators - Main Result 1

Work in infinite dimensional setting where we can show:

Theorem (Agapiou, Larsson, Stuart)

The posterior is Gaussian, $\mu^y = \mathcal{N}(m, \frac{1}{n}\mathcal{B}_\lambda^{-1})$, where

$$\mathcal{B}_\lambda = \mathcal{A}^{-1}\mathcal{C}_1^{-1}\mathcal{A}^{-1} + \lambda\mathcal{C}_0^{-1}, \quad \lambda = \frac{1}{n\tau^2}$$

$$\mathcal{B}_\lambda m = \mathcal{A}^{-1}\mathcal{C}_1^{-1}y.$$

- $\mathcal{B}_\lambda$ depends on n and τ only through λ.
- Observation: m minimizer of J_0; posterior mean Tikhonov-Phillips solution of (1)!

Setting - Assumptions

- Hilbert Scale $(X^s)_{s\in\mathbb{R}}$, for $X^s = \mathcal{D}(\mathcal{C}_0^{-\frac{s}{2}})$ with $\langle u, v\rangle_s = \left\langle \mathcal{C}_0^{-\frac{s}{2}} u, \mathcal{C}_0^{-\frac{s}{2}} v \right\rangle$.
- Assumptions

$$\exists s_0 \in [0,1) \text{ s.t. } \mathrm{tr}(\mathcal{C}_0^s) < \infty \quad \forall s > s_0;$$

$$\mathcal{C}_1 \simeq \mathcal{C}_0^{\beta}, \quad \beta \geq 0;$$

$$\mathcal{A}^{-1} \simeq \mathcal{C}_0^{\ell}, \quad \ell > 0.$$

- We have

$$\mathcal{B}_\lambda = \mathcal{A}^{-1}\mathcal{C}_1^{-1}\mathcal{A}^{-1} + \lambda\mathcal{C}_0^{-1} \simeq \mathcal{C}_0^{2\ell-\beta} + \lambda\mathcal{C}_0^{-1}.$$

Assume $\Delta := 2\ell - \beta + 1 > 0$, i.e. prior regularizing.

Posterior Consistency

- Assume observations

$$y^\dagger = \mathcal{A}^{-1}u^\dagger + \frac{1}{\sqrt{n}}\xi, \quad \xi \sim \mathcal{N}(0, \mathcal{C}_1)$$

$u^\dagger \in \mathcal{X}$ fixed true solution.

- This data model gives $\mu^{y^\dagger}_{\lambda,n} := \mu^y|_{y=y^\dagger} = \mathcal{N}(m^\dagger_\lambda, \frac{1}{n}\mathcal{B}^{-1}_\lambda)$, where

$$\mathcal{B}_\lambda m^\dagger_\lambda = \mathcal{A}^{-1}\mathcal{C}_1^{-1}y^\dagger.$$

- AIM: Show that in small noise limit $(n \to \infty)$ posterior contracts to a Dirac centered on the true solution.

Posterior Consistency - Posterior Contraction

Assume $u^\dagger \in X^\gamma$. Determine rate $\varepsilon_n = \varepsilon_n(\gamma, \Delta, s_0)$ such that

$$\mathbb{E}^{y^\dagger} \mu_{\lambda,n}^{y^\dagger} \left\{u : \left\|u - u^\dagger\right\| \geq M_n \varepsilon_n\right\} \to 0, \quad \forall M_n \to \infty, \text{ as } n \to \infty.$$

- Markov Inequality

$$\mathbb{E}^{y^\dagger} \mu_{\lambda,n}^{y^\dagger} \left\{u : \left\|u - u^\dagger\right\| \geq M_n \varepsilon_n\right\} \leq \frac{1}{M_n^2 \varepsilon_n^2} \mathbb{E}^{y^\dagger} \int \left\|u - u^\dagger\right\|^2 \mu_{\lambda,n}^{y^\dagger}(du).$$

- Since $\mu_{\lambda,n}^{y^\dagger} = \mathcal{N}(m_\lambda^\dagger, \frac{1}{n}\mathcal{B}_\lambda^{-1})$, suffices to show

$$SPC := \underbrace{\mathbb{E}^{y^\dagger} \left\|m_\lambda^\dagger - u^\dagger\right\|^2}_{\text{MISE}} + \underbrace{\mathrm{tr}(\frac{1}{n}\mathcal{B}_\lambda^{-1})}_{\text{posterior spread}} \leq c\varepsilon_n^2.$$

Posterior Consistency - Main result 2

- Assume $\Delta \geq 1$, i.e. sufficiently ill-posed inverse problem.

Theorem (Agapiou, Larsson, Stuart)

Assume $u^\dagger \in X^\gamma$, $\gamma \geq 1$. Under our assumptions, we have the following rates of contraction, for appropriate choice of $\lambda = \lambda(n) \to 0$:

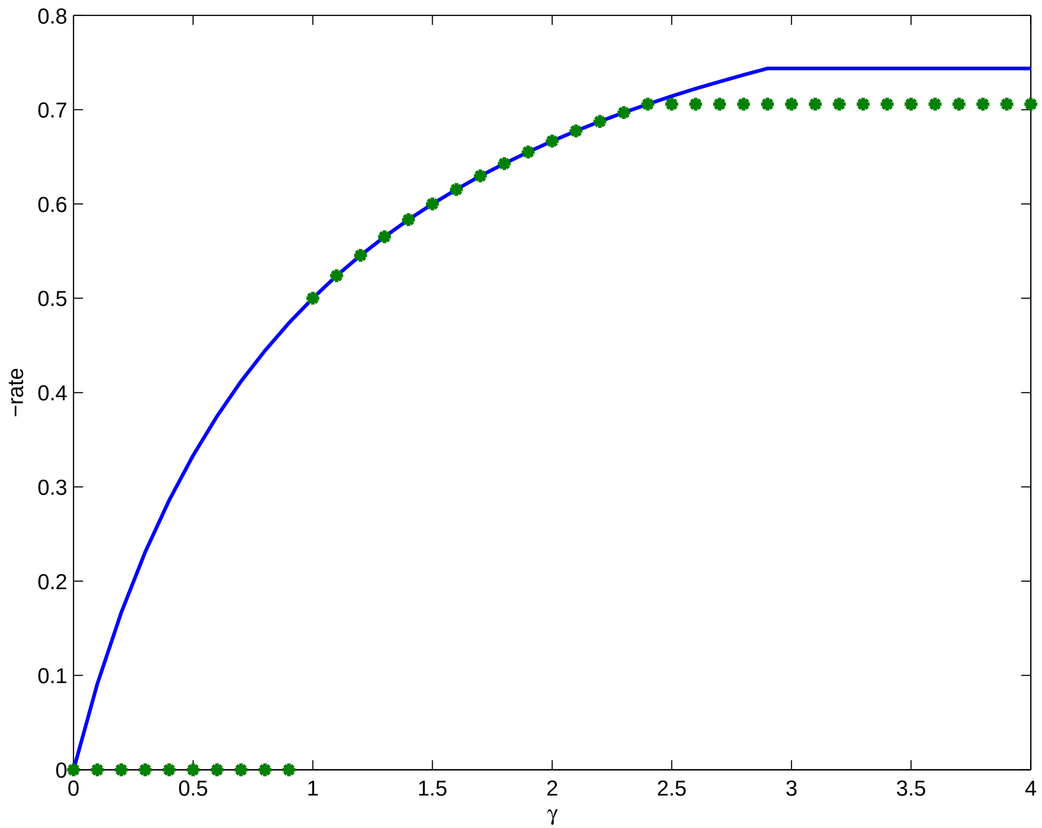
$$\varepsilon_n = \begin{cases} n^{-\frac{\gamma}{2(\Delta+\gamma-1+s_0)}}, & \text{if } \gamma \in [1, \Delta+1] \\ n^{-\frac{\Delta+1}{2(2\Delta+s_0)}}, & \text{if } \gamma > \Delta+1. \end{cases}$$

Optimality

Diagonal Case: $\mathcal{A}^{-1} = \mathcal{C}_0^{\ell}$ and $\mathcal{C}_1 = \mathcal{C}_0^{\beta}$ gives sharp rates.

Our operator similarity assumptions satisfied trivially.

- Assume $\mathcal{C}_0 \asymp \operatorname{diag}\{k^{-2}\}$, $u^{\dagger} \in X^{\gamma}$. Compare rates of convergence for $\ell = \beta = 1/2$.



Main Result 2 - proof idea

- Assumptions secure posterior spread bounded by MISE; suffices to bound MISE.

$$\mathcal{B}_\lambda m_\lambda^\dagger = \mathcal{A}^{-1}\mathcal{C}_1^{-1}y^\dagger = \underbrace{\mathcal{A}^{-1}\mathcal{C}_1^{-1}\mathcal{A}^{-1}u^\dagger} + \frac{1}{\sqrt{n}}\mathcal{A}^{-1}\mathcal{C}_1^{-1}\xi$$

$$\mathcal{B}_\lambda u^\dagger = \underbrace{\mathcal{A}^{-1}\mathcal{C}_1^{-1}\mathcal{A}^{-1}u^\dagger} + \lambda\mathcal{C}_0^{-1}u^\dagger.$$

Set $e = m_\lambda^\dagger - u^\dagger$

$$\mathcal{B}_\lambda e = \frac{1}{\sqrt{n}}\mathcal{A}^{-1}\mathcal{C}_1^{-1}\xi - \lambda\mathcal{C}_0^{-1}u^\dagger.$$

Main Result 2 - proof idea

- Testing against e, using norm equivalence and interpolation techniques

$$\|e\|^2_{\beta-2\ell} + \lambda \|e\|^2_1 \leq c(\frac{1}{n}\lambda^{-\theta_1} \|\xi\|^2_{\beta-\theta_1\Delta} + \lambda^{2-\theta_2} \left\|u^\dagger\right\|^2_{1+\Delta(1-\theta_2)}),$$

$\theta_1, \theta_2 \in [0, 1]$ chosen to make rhs finite.

- Choose $\lambda = \lambda(n)$ optimally to get rates for two error norms on lhs.

- Interpolate between two rates and take expectations to get the rate for MISE. □

Ongoing - Future Research

- Same methodology applied in *Pokern et al.* in nonparametric drift estimation for diffusion processes. Extension to an abstract setting which includes both cases as examples;
- Extension to non-Gaussian priors; Besov priors;
- Extension to nonlinear inverse problems.

References - Further Reading

- S. Agapiou, S. Larsson and A. M. Stuart, *Posterior Consistency of the Bayesian Approach to Linear Ill-Posed Inverse Problems*, http://arxiv.org/abs/1203.5753.
- B. T. Knapik, A. W. van der Vaart, J. H. van Zanten, *Bayesian inverse problems with Gaussian priors*, Ann. Statist. 39 (2011), 2626-2657.
- Y. Pokern, A. M. Stuart and J. H. Van Zanten, *Posterior consistency via precision operators for nonparametric drift estimation in SDEs*, http://arxiv.org/abs/1202.0976.
- A. M. Stuart, *Inverse problems: a Bayesian perspective*, Acta Numer. 19 (2010), 451-559.