Justification of the posterior in Bayesian linear inverse problems using precision operators

Sergios Agapiou

Mathematics Institute, University of Warwick, Coventry, UK

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

Bayesian Inverse Problems Day,
Monday May 14th 2012, University of Warwick

Linear Inverse Problems

- $\mathcal{A} : \mathcal{D}(\mathcal{A}) \subset \mathcal{X} \to \mathcal{X}$ 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

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

$\mathcal{C}_i : \mathcal{X} \longrightarrow \mathcal{X}$, 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 $n \longrightarrow \infty$ to recover unknown u.

Bayesian Approach

Bayesian framework: assume $\frac{1}{\sqrt{n}}\xi \sim \mathbb{P}_0 = \mathcal{N}(0, \frac{1}{n}\mathcal{C}_1)$, $\mathcal{C}_1 : \mathcal{X} \to \mathcal{X}$ selfadjoint positive definite.

- Likelihood: for fixed u, $y|u \sim \mathbb{P} = \mathcal{N}(\mathcal{A}^{-1}u, \frac{1}{n}\mathcal{C}_1)$.
- Prior: assume $u \sim \mu_0 = \mathcal{N}(0, \tau^2\mathcal{C}_0)$, $\mathcal{C}_0 : \mathcal{X} \to \mathcal{X}$ selfadjoint positive definite trace class.
- Posterior: *Mandelbaum ('84)* and *Lehtinen et al. ('89)* $u|y \sim \mu^y = \mathcal{N}(m, \mathcal{C})$, where

$$\mathcal{C} = \tau^2\mathcal{C}_0 - \tau^2\mathcal{C}_0\mathcal{A}^{-1}(\mathcal{A}^{-1}\mathcal{C}_0\mathcal{A}^{-1} + \lambda\mathcal{C}_1)^{-1}\mathcal{A}^{-1}\mathcal{C}_0$$

$$m = \mathcal{C}_0\mathcal{A}^{-1}(\mathcal{A}^{-1}\mathcal{C}_0\mathcal{A}^{-1} + \lambda\mathcal{C}_1)^{-1}y$$

Bayesian Approach via Precision Operators - Main Result

Work in infinite dimensional setting where we show:

Theorem

The posterior is Gaussian, $u|y \sim \mu^y = \mathcal{N}(m, \mathcal{C})$, where

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

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

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

Bayesian Approach via Precision Operators - Intuition

- 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 $u|y \sim \mathcal{N}(m, \mathcal{C})$, where by completing the square

$$\mathcal{B}_\lambda := \frac{1}{n}\mathcal{C}^{-1} = \mathcal{A}^{-1}\mathcal{C}_1^{-1}\mathcal{A}^{-1} + \lambda\mathcal{C}_0^{-1}$$
$$\mathcal{B}_\lambda m = \mathcal{A}^{-1}\mathcal{C}_1^{-1}y.$$

Setting - Assumptions

- $\{\phi_k\}_{k\in\mathbb{N}}$ eigenfunctions of $\mathcal{C}_0$, orthonormal basis of $\mathcal{X}$.
- Hilbert Scale $(X^t)_{t\in\mathbb{R}}$, for $X^t = \mathcal{D}(\mathcal{C}_0^{-\frac{t}{2}})$ with $\langle u, v\rangle_t = \left\langle \mathcal{C}_0^{-\frac{t}{2}} u, \mathcal{C}_0^{-\frac{t}{2}} v \right\rangle$.
- Assumptions:

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

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

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

- 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.

Setting - Assumptions

- Define

$$\nu(du, dy) = \mathbb{P}(dy|u)\mu_0(du)$$

$$\nu_0(du, dy) = \mathbb{P}_0(dy) \otimes \mu_0(du).$$

- Assumptions allow to determine in which X^t draws from the measures $\mu_0, \mathbb{P}_0, \mathbb{P}, \nu, \nu_0$, belong to almost surely.

  Key: for $\zeta \sim \mathcal{N}(0, \mathrm{Id})$, we have $\mathbb{E}\|\zeta\|_{-s}^2 = \mathrm{tr}(\mathcal{C}_0^{\mathrm{s}}) < \infty, \ \forall \mathrm{s} > \mathrm{s}_0$.

Proof of Main result - Overview

Steps:

- make sense of the mean equation weakly in X^1;
- characterize the posterior through Radon-Nikodym derivative wrt the prior;
- finite dimensional approximation of posterior in the Karhunen-Loeve basis of prior; approximation is Gaussian and converges weakly to posterior;
- the Gaussian family is closed under weak convergence, hence posterior is Gaussian. We justify the equations for mean and covariance by showing that mean and covariance of the approximation converge to m and $\mathcal{C}$.

Step 1 - Making sense of the posterior mean equation

$$\mathcal{B}_\lambda m = \mathcal{A}^{-1}\mathcal{C}_1^{-1}y$$
$$\mathcal{B}_\lambda = \mathcal{A}^{-1}\mathcal{C}_1^{-1}\mathcal{A}^{-1} + \lambda\mathcal{C}_0^{-1}.$$

- Define bilinear form $B : X^1 \times X^1 \to \mathbb{R}$

$$B(u,v) := \left\langle \mathcal{C}_1^{-\frac{1}{2}}\mathcal{A}^{-1}u, \mathcal{C}_1^{-\frac{1}{2}}\mathcal{A}^{-1}v \right\rangle + \lambda \left\langle \mathcal{C}_0^{-\frac{1}{2}}u, \mathcal{C}_0^{-\frac{1}{2}}v \right\rangle.$$

- B coercive and continuous; by Lax-Milgram $\forall r \in X^{-1}\ \ \exists! e \in X^1$

$$B(e,v) = \langle r, v\rangle,\ \forall v \in X^1.$$

- Assumptions secure that $\mathcal{A}^{-1}\mathcal{C}_1^{-1}y \in X^{-1}$ ν-a.s., thus $\exists! m \in X^1$ weak solution of mean equation ν-a.s.

Step 2: Posterior well defined and characterized using μ_0

Proposition

The posterior measure μ^y is absolutely continuous wrt the prior μ_0 and

$$\frac{d\mu^y}{d\mu_0}(u) = \frac{1}{Z(y)}\exp(-\Phi(u,y)),$$

where

$$\Phi(u,y) := \frac{n}{2}\left\|\mathcal{C}_1^{-\frac{1}{2}}\mathcal{A}^{-1}u\right\|^2 - n\left\langle \mathcal{C}_1^{-\frac{1}{2}}y, \mathcal{C}_1^{-\frac{1}{2}}\mathcal{A}^{-1}u\right\rangle$$

and $Z(y) \in (0,\infty)$ is the normalizing constant.

Step 2: Posterior well defined and characterized using μ_0

Sketch of proof of proposition:

- Assumptions secure $\mathcal{A}^{-1}u \in \mathcal{D}(\mathcal{C}_1^{-\frac{1}{2}})$ μ_0-a.s., thus by Cameron-Martin formula

$$\frac{d\mathbb{P}}{d\mathbb{P}_0}(y|u) = \exp(-\Phi(u,y)) \quad \mu_0\text{-a.s.},$$

  hence

$$\frac{d\nu}{d\nu_0}(u,y) = \exp(-\Phi(u,y)).$$

- By Lemma 5.3 *Hairer et al. (2007)*, since $\nu_0(u|y) = \mu_0(u)$ exists, the result holds provided $Z(y) = \int \exp(-\Phi(u,y)) d\mu_0(u) > 0$ ν-a.s.

- Since Gaussian measures in separable Hilbert spaces are full, assumptions secure $Z(y) > 0$ ν-a.s. □

Step 3: Finite dimensional approximation of posterior

- Consider finite dimensional approximation in Karhunen-Loeve basis of prior:

$$\frac{d\mu^{N,y}}{d\mu_0}(u) = \frac{1}{Z^N}\exp(-\Phi(P^N u, y))$$

P^N projection onto $\mathcal{X}^N := \mathrm{span}\{\phi_1, .., \phi_\mathrm{N}\}$.

- $\mu^{N,y} = \mathcal{N}(m^N, \mathcal{C}^N)$, where

$$P^N \mathcal{C}^{-1} P^N m^N = n P^N \mathcal{A}^{-1} \mathcal{C}_1^{-1} y$$

$$\mathcal{C}^N = P^N \mathcal{C} P^N + (I - P^N)\mathcal{C}_0(I - P^N).$$

- Assumptions, using Dominated Convergence and the Fernique Theorem, secure $\mu^{N,y} \Rightarrow \mu^y$ in $\mathcal{X}$, ν-a.s.

Final Step: Posterior is Gaussian $\mu^y = \mathcal{N}(m, \mathcal{C})$

- μ^y is ν-a.s. Gaussian, $\mu^y = \mathcal{N}(\overline{m}, \overline{\mathcal{C}})$. We show $\overline{m} = m$ and $\overline{\mathcal{C}} = \mathcal{C}$.

- Since m, m^N satisfy mean equation weakly in $\mathcal{X}^N$, Standard Galerkin method shows $m^N \longrightarrow m$ in $\mathcal{X}$.

- Weak convergence of Gaussian measures in separable Hilbert spaces implies convergence of the means, thus $m^N \rightarrow \overline{m}$ in $\mathcal{X}$. Uniqueness of limit $\overline{m} = m$.

Final Step: Posterior is Gaussian $\mu^y = \mathcal{N}(m, \mathcal{C})$

- Fix k. For $N > k$ and $w \in \mathcal{X}$

$$\left|\left\langle w, \mathcal{C}^N \phi_k \right\rangle - \langle w, \mathcal{C}\phi_k \rangle\right| = \left|\left\langle w, (P^N - I)\mathcal{C}\phi_k \right\rangle\right| \leq \left\|(P^N - I)w\right\| \|\mathcal{C}\phi_k\|$$

  where rhs converges to zero, thus $\mathcal{C}^N \phi_k \rightharpoonup \mathcal{C}\phi_k, \ \forall k \in \mathbb{N}$.

- Weak convergence of Gaussian measures in separable Hilbert spaces implies strong convergence of the covariance operators, thus $\mathcal{C}^N \phi_k \rightharpoonup \overline{\mathcal{C}}\phi_k, \ \forall k \in \mathbb{N}$.

- $\overline{\mathcal{C}}\phi_k = \mathcal{C}\phi_k, \ \forall k \in \mathbb{N}$. Since $\{\phi_k\}_k$ orthonormal basis of $\mathcal{X}$, $\overline{\mathcal{C}} = \mathcal{C}$.

$\square$

Synopsis

We have

- Considered a linear Inverse Problem in infinite dimensions;
- Adopted a Bayesian approach with Gaussian prior and Gaussian noise;
- Shown that the posterior is absolutely continuous with respect to the prior;
- Seen that the posterior is Gaussian with mean the minimizer of a Tikhonov-Phillips functional;
- Worked with precision operators, which allows us to use tools from analysis and numerical analysis to examine posterior consistency (next talk by Yuan-Xiang Zhang)

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
- Bogachev, Vladimir I., *Gaussian Measures*, Mathematical Surveys and Monographs, vol. 62, American Mathematical Society, 1998.
- M. Hairer, A. M. Stuart and J. Voss, *Analysis of SPDEs arising in path sampling. II. The nonlinear case*, Ann. Appl. Probab. 17 (2007), no. 5-6, 1657-1706.
- 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.