Regularization of Linear Inverse Problems by Bayesian Methodology

Sergios Agapiou

Mathematics Institute, University of Warwick, Coventry, UK

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

Numerical Analysis Postgraduate Seminar Day,
Friday March 30th 2012, University of Reading

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.
- Consider the inverse problem to find u from y, where y noisy observation of $\mathcal{A}^{-1}u$.
- Model:

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

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

Inverse Problems are Ill-Posed - Deterministic Approach

- Problem (1) ill-posed:
  - Existence of solution not guaranteed
  - 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 the Bayesian Approach, the solution of problem (1) is the distribution of $u|y$, called the *posterior* distribution.
- Link: Bayes rule

$$\mathbb{P}(u|y) \propto \mathbb{P}(y|u)\mathbb{P}(u)$$

Bayesian Approach - link to Tikhonov

- Work in infinite dimensional setting where we can show

$$\mathbb{P}(u|y) = \mathcal{N}(m, \frac{1}{n}\mathcal{B}_\lambda^{-1})$$

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

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

- $\mathcal{B}_\lambda$ depends on n and τ only through λ.
- Observation: m minimizer of J_0; posterior mean is 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 on $\mathcal{A}, \mathcal{C}_0$ and $\mathcal{C}_1$

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

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

- 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 is regularizing.

Bayesian Approach - Posterior Consistency

- Assume observations

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

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

- This data model gives $\mathbb{P}(u|y = y^\dagger) = \mathcal{N}(m_\lambda^\dagger, \frac{1}{n}\mathcal{B}_\lambda^{-1})$, where

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

Posterior Consistency - Error Equation

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

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

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

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

- Assume $u^\dagger \in X^\gamma$. Want rates

$$\mathbb{E}\,\|e\|_\eta^2 = \mathcal{O}(f(n,\eta,\gamma,\ell,\beta)).$$

Main result

- Set $\eta_i = (1-\theta_i)(\beta - 2\ell) + \theta_i$, where $\theta_i \in [0,1]$.

  Let $\kappa^2 = \max\{\|\xi\|^2_{2\beta-2\ell-\eta_1}, \|u^\dagger\|^2_{2-\eta_2}\}$ and choose θ_1, θ_2 s.t. $\mathbb{E}(\kappa^2) < +\infty$.

Main result

- Set $\eta_i = (1-\theta_i)(\beta - 2\ell) + \theta_i$, where $\theta_i \in [0,1]$.

  Let $\kappa^2 = \max\{\|\xi\|^2_{2\beta-2\ell-\eta_1}, \|u^\dagger\|^2_{2-\eta_2}\}$ and choose θ_1, θ_2 s.t. $\mathbb{E}(\kappa^2) < +\infty$.

Theorem (Agapiou, Larsson, Stuart)

For any $\theta \in [0,1]$,

$$\mathbb{E}\,\|e\|^2_\eta \leq c\mathbb{E}(\kappa^2) n^{\frac{2-\theta_2-\theta}{\theta_2-\theta_1-2}}$$

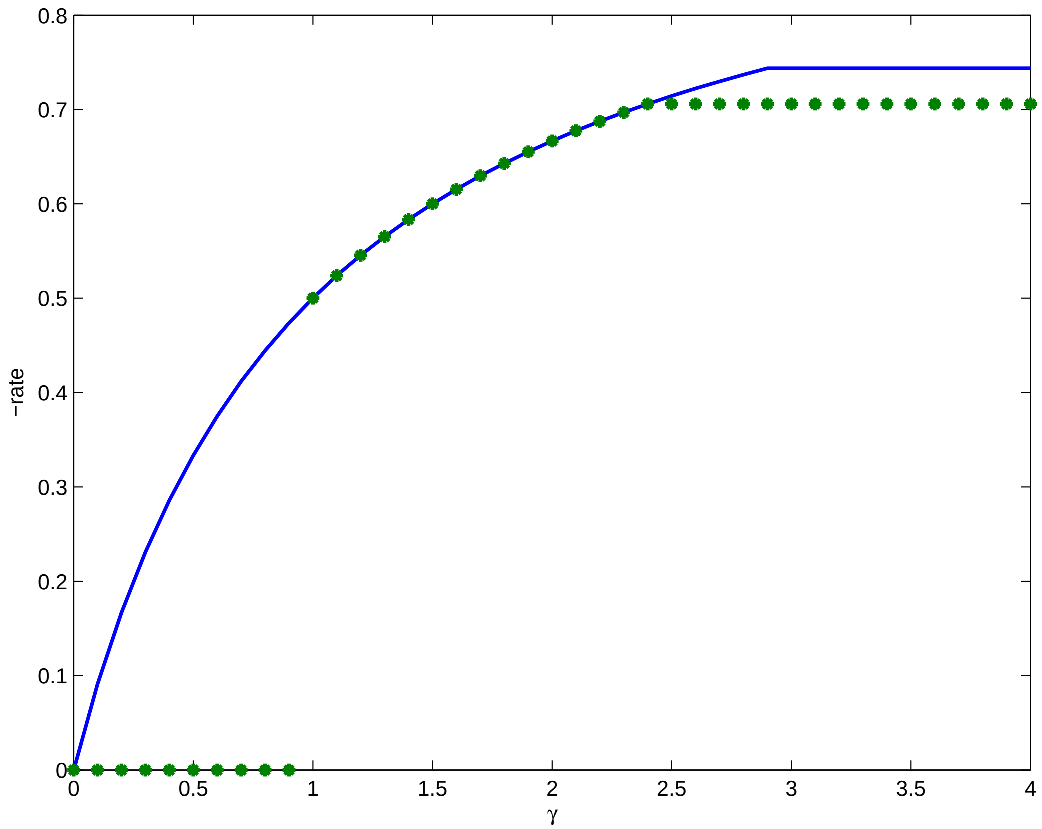
where $\eta = (1-\theta)(\beta - 2\ell) + \theta$.

Optimality

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

Our assumptions satisfied trivially.

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



Proof - Making Sense of the error equation

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

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

- By Lax-Milgram $\forall r \in X^{-1} \;\; \exists! e \in X^1$

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

- For $r =$ r.h.s of (err) $\in X^{-1}$, $v = e \in X^1$

$$\left\| \mathcal{C}_1^{-\frac{1}{2}} \mathcal{A}^{-1} e \right\|^2 + \lambda \left\| \mathcal{C}_0^{-\frac{1}{2}} e \right\|^2 = \frac{1}{\sqrt{n}} \left\langle \mathcal{A}^{-1} \mathcal{C}_1^{-1} \xi, e \right\rangle - \lambda \left\langle \mathcal{C}_0^{-1} u^\dagger, e \right\rangle$$

Proof - Analysing the error equation

- Norm equivalence

$$\|e\|^2_{\beta-2\ell} + \lambda \|e\|^2_1 \leq c \left(\frac{1}{\sqrt{n}} \langle \mathcal{A}^{-1} \mathcal{C}_1^{-1} \xi, e \rangle - \lambda \langle \mathcal{C}_0^{-1} u^\dagger, e \rangle \right).$$

- Strategy: Interpolate norms on e on the r.h.s. between norms on e on the l.h.s. Recall $\eta_i = (1-\theta_i)(\beta - 2\ell) + \theta_i, \ \theta_i \in [0,1]$.

- Use norm equivalence, Cauchy-Schwarz, interpolation inequality, Cauchy with ε, Young's inequality

$$\|e\|^2_{\beta-2\ell} + \lambda \|e\|^2_1 \leq c \left(\frac{1}{\sqrt{n}} \left\langle \mathcal{C}_0^{\frac{\eta_1}{2}} \mathcal{A}^{-1} \mathcal{C}_1^{-1} \xi, \mathcal{C}_0^{-\frac{\eta_1}{2}} e \right\rangle - \lambda \left\langle \mathcal{C}_0^{\frac{\eta_2}{2}-1} u^\dagger, \mathcal{C}_0^{-\frac{\eta_2}{2}} e \right\rangle \right)$$

$$\leq c \left(\frac{1}{\sqrt{n}} \|\xi\|_{2\beta-2\ell-\eta_1} \|e\|_{\eta_1} + \lambda \left\| u^\dagger \right\|_{2-\eta_2} \|e\|_{\eta_2} \right)$$

Proof - Analysing the error equation

$$\leq c\left(\frac{1}{\sqrt{n}}\lambda^{-\frac{\theta_1}{2}}\|\xi\|_{2\beta-2\ell-\eta_1}\|e\|_{\beta-2\ell}^{1-\theta_1}(\lambda^{\frac{1}{2}}\|e\|_1)^{\theta_1}+\lambda^{1-\frac{\theta_2}{2}}\|u^\dagger\|_{2-\eta_2}\|e\|_{\beta-2\ell}^{1-\theta_2}(\lambda^{\frac{1}{2}}\|e\|_1)^{\theta_2}\right)$$

$$\leq \frac{c}{2\varepsilon}\left(\frac{1}{n}\lambda^{-\theta_1}\|\xi\|_{2\beta-2\ell-\eta_1}^2+\lambda^{2-\theta_2}\|u^\dagger\|_{2-\eta_2}^2\right)$$
$$+\frac{c\varepsilon}{2}\left(\|e\|_{\beta-2\ell}^{2-2\theta_1}(\lambda^{\frac{1}{2}}\|e\|_1)^{2\theta_1}+\|e\|_{\beta-2\ell}^{2-2\theta_2}(\lambda^{\frac{1}{2}}\|e\|_1)^{2\theta_2}\right)$$

$$\leq \frac{c}{2\varepsilon}\left(\frac{1}{n}\lambda^{-\theta_1}\|\xi\|_{2\beta-2\ell-\eta_1}^2+\lambda^{2-\theta_2}\|u^\dagger\|_{2-\eta_2}^2\right)$$
$$+\frac{c\varepsilon}{2}\left((2-\theta_1-\theta_2)\|e\|_{\beta-2\ell}^2+(\theta_1+\theta_2)\lambda\|e\|_1^2\right).$$

Proof - Analysing the error equation

- For ε small enough

$$\|e\|^2_{\beta-2\ell} + \lambda \|e\|^2_1 \leq c(\frac{1}{n}\lambda^{-\theta_1} \|\xi\|^2_{2\beta-2\ell-\eta_1} + \lambda^{2-\theta_2} \|u^\dagger\|^2_{2-\eta_2}).$$

- Then

$$\mathbb{E} \|e\|^2_{\beta-2\ell} \leq c\mathbb{E}(\kappa^2)(\frac{1}{n}\lambda^{-\theta_1} + \lambda^{2-\theta_2})$$

$$\mathbb{E} \|e\|^2_1 \leq c\mathbb{E}(\kappa^2)(\frac{1}{n}\lambda^{-1-\theta_1} + \lambda^{1-\theta_2}).$$

- Optimize choosing $\lambda = n^p$ and interpolate. $\square$

Synopsis

We have

- Considered a linear Inverse Problem in infinite dimensions;
- Adopted a Bayesian approach with Gaussian prior and Gaussian noise;
- Seen that the posterior is Gaussian with mean the minimizer of a Tikhonov-Phillips functional;
- Shown posterior consistency results in the small noise limit.

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
- Heinz W. Engl, Martin Hanke and Andreas Neubauer, *Regularization of inverse problems*, Mathematics and its Applications, vol. 375, Kluwer Academic Publishers Group, Dordrecht, 1996.
- 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.