PDE Techniques for Bayesian Inverse Problems

Sergios Agapiou

Mathematics Institute, University of Warwick, Coventry, UK

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

Computational and Applied Mathematics Seminar,
Chalmers University of Technology, 28 September 2011

Outline

Outline

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) is in general ill-posed:
  - The existence of a solution is not guaranteed
  - Solutions do not depend continuously on the data
- Tikhonov-Philips Regularization: u approximated by the minimizer of

$$J_0(u) := \frac{1}{2}\left\|C_1^{-\frac{1}{2}}(y - \mathcal{A}^{-1}u)\right\|^2 + \frac{\lambda}{2}\left\|C_0^{-\frac{1}{2}}u\right\|^2,$$

where $C_i : \mathcal{X} \to \mathcal{X},\ i = 0, 1$ self-adjoint, possibly compact on $\mathcal{X}$, positive definite linear operators.

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

Bayesian Approach

Adopt a Bayesian approach to solve problem (1): assume $\xi \sim \mathcal{N}(0, C_1)$, where $C_1 : \mathcal{X} \to \mathcal{X}$ selfadjoint positive definite but not necessarily trace class linear operator (white noise allowed).

- Likelihood: for fixed u the law of $y|u$, called the *likelihood*, is $\mathcal{N}(\mathcal{A}^{-1}u, \frac{1}{n}C_1)$.
- Prior: we choose a *prior* distribution for the unknown u, based on prior knowledge about u. Let $u \sim \mathcal{N}(0, \tau^2 C_0)$, where $C_0 : \mathcal{X} \to \mathcal{X}$ selfadjoint positive definite trace class linear operator.
- Posterior: in the Bayesian Approach, the solution of problem (1) is the distribution of $u|y$, called the *posterior* distribution.

Bayesian Approach - Finite Dimensions

- Bayes rule:

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

- In $\mathbb{R}^d$

$$\pi^y(u) \propto \rho(y - \mathcal{A}^{-1}u)\pi(u).$$

In general calculation of posterior difficult. Gaussian prior and noise case simpler.

- Assuming $\mathcal{X} = \mathbb{R}^d$, $\mathcal{A}$, C_1, $C_0 \in \mathbb{R}^{d\times d}$ and defining $\lambda = \frac{1}{n\tau^2}$ we have

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

$$nJ_0(u) = \frac{n}{2}\left\|C_1^{-\frac{1}{2}}(y - \mathcal{A}^{-1}u)\right\|^2 + \frac{1}{2\tau^2}\left\|C_0^{-\frac{1}{2}}u\right\|^2.$$

Bayesian Approach - Finite Dimensions

- This suggests that posterior is Gaussian $\mu^y = \mathcal{N}(m, \frac{1}{n}B_\lambda^{-1})$, where by completing the square

$$B_\lambda = \mathcal{A}^{-1}C_1^{-1}\mathcal{A}^{-1} + \lambda C_0^{-1} \tag{2}$$

$$B_\lambda m = \mathcal{A}^{-1}C_1^{-1}y. \tag{3}$$

- Observe B_λ depends on n and τ only through $\lambda = \frac{1}{n\tau^2}$.

- Observation: m is the minimizer of J hence also of J_0, that is the posterior mean is the Tikhonov-Philips solution of problem (1).

- We work in an infinite dimensional setting where formulae (2) and (3) can be justified (Theorem 1).

Bayesian Approach - Posterior Consistency

Assume observations

$$y^\dagger = \mathcal{A}^{-1}u^\dagger + \frac{1}{\sqrt{n}}\xi, \quad \xi \sim \mathcal{N}(0, C_1)$$

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

- This data model gives $\mu^{y^\dagger} = \mathcal{N}(m_\lambda^\dagger, \frac{1}{n}B_\lambda^{-1})$, where

$$B_\lambda m_\lambda^\dagger = \mathcal{A}^{-1}C_1^{-1}y^\dagger.$$

- We examine the consistency of the posterior mean in the sense

$$\mathbb{E}\left\|m_\lambda^\dagger - u^\dagger\right\|^2 \to 0,$$

as the noise disappears ($n \to \infty$) and for appropriate choice of scaling of prior $\tau = \tau(n)$ (Theorem 2). Assume a-priori known regularity of $u^\dagger$ and provide rates.

Outline

Hilbert Scale

- Recall we assume

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

$u \sim \mu_0 = \mathcal{N}(0, \tau^2 C_0)$ and $\xi \sim \mathcal{N}(0, C_1)$ implying $y|u \sim \mathcal{N}(\mathcal{A}^{-1}u, \frac{1}{n}C_1)$.

- We work in the Hilbert Scale $(X^s)_{s\in\mathbb{R}}$, for $X^s := \overline{\mathcal{M}}^{\|\cdot\|_s}$ where

$$\mathcal{M} = \bigcap_{k=0}^{\infty} \mathcal{D}(C_0^{-k}),\ \langle u, v\rangle_s := \left\langle C_0^{-\frac{s}{2}}u, C_0^{-\frac{s}{2}}v\right\rangle \quad \text{and} \quad \|u\|_s := \left\|C_0^{-\frac{s}{2}}u\right\|.$$

Operator Interrelations - Prior Dominates

We have a number of assumptions on $\mathcal{A}$, C_0 and C_1 reflecting the idea

$$C_1 \simeq C_0^{\beta}$$

and

$$\mathcal{A}^{-1} \simeq C_0^{\ell},$$

for some $\beta \geq 0, \ell > 0$, where $\simeq$ means they induce equivalent norms (precise later).

- Let $\Delta = 2\ell - \beta + 1$. We have

  $$B_\lambda \simeq C_0^{2\ell-\beta} + \lambda C_0^{-1}$$

  thus the sign of Δ determines which term dominates. We require $\Delta > 0$ which means that the prior is regularizing.

Posterior Identification

Theorem 1

Let $y = \mathcal{A}^{-1}u + \frac{1}{\sqrt{n}}\xi$, where $u \sim \mathcal{N}(0, \tau^2 C_0)$ and $\xi \sim \mathcal{N}(0, C_1)$. Then

$$\mu^y = \mathcal{N}(m, \frac{1}{n}B_\lambda^{-1}),$$

where m and B_λ are given by

$$B_\lambda = \mathcal{A}^{-1}C_1^{-1}\mathcal{A}^{-1} + \lambda C_0^{-1} \tag{2}$$

and

$$B_\lambda m = \mathcal{A}^{-1}C_1^{-1}y. \tag{3}$$

Posterior Consistency - Convergence Rates

Theorem 2

For every $\theta \in [0,1]$ for $\mu = (1-\theta)(\beta - 2\ell) + \theta$, by choosing

$$\tau(n) = n^{\frac{\theta_2 - \theta_1 - 1}{4 - 2\theta_2 + 2\theta_1}},$$

we have

$$\mathbb{E}\left\|m_\lambda^\dagger - u^\dagger\right\|_\mu^2 \leq c\mathbb{E}(\kappa^2) n^{\frac{\theta + \theta_2 - 2}{\theta_1 - \theta_2 + 2}},$$

where

$$\kappa = \max\left\{\|\xi\|_{2\beta - 2\ell - \mu_1}, \left\|u^\dagger\right\|_{2-\mu_2}\right\} \quad \text{for} \quad \mu_i = (1-\theta_i)(\beta - 2\ell) + \theta_i, \ i = 1,2$$

and where $\theta_1, \theta_2 \in [0,1]$ are chosen so that $\mathbb{E}(\kappa^2) < \infty$.

- - higher θ stronger error-norm gives slower rates
  - regular noise/truth allows smaller θ_1, θ_2, faster rates
  - limitations because $\theta_1, \theta_2 \in [0,1]$
  - choice of θ_1, θ_2 independent of θ

Outline

Setting - Regularity of Prior

- Let $\{\lambda_k^2, \phi_k\}_{k\in\mathbb{N}}$ orthonormal eigenbasis of C_0 in $\mathcal{X}$. Since C_0 trace class $\sum_{k=1}^{\infty} \lambda_k^2 < \infty$.

Assumption 1

Let $\sigma_0 \in (0,1]$ be such that $\sum_{k=1}^{\infty} \lambda_k^{2(1-\sigma)} < \infty \quad \forall \sigma < \sigma_0$.

eg. if $\sigma_0 = 1$, λ_k^2 decay faster than any negative power of k.

Lemma 1

Let ζ white noise wrt C_0. Then $\mathbb{E}\left\|C_0^{\frac{s}{2}}\zeta\right\|^2 < \infty$ for all $s > s_0 := 1 - \sigma_0$.

- Proof: ζ white noise thus $C_0^{\frac{s}{2}}\zeta \sim \mathcal{N}(0, C_0^s)$. To have $\mathbb{E}\left\|C_0^{\frac{s}{2}}\zeta\right\|^2 < \infty$, it suffices that C_0^s trace class, $\sum_{k=1}^{\infty} \lambda_k^{2s} < \infty$. By Assumption 1 need $s > 1 - \sigma_0$. $\square$

Assumptions - Norm Equivalence

Assumption 2

Suppose there exist $\beta \geq 0$, $\ell > 0$ and constants c_1-$c_6 > 0$ such that for $\Delta = 2\ell - \beta + 1$ we have

1. $\Delta > 2(1-\sigma_0)(> 0)$;
2. $\left\| C_1^{-\frac{1}{2}} \mathcal{A}^{-1} u \right\| \asymp \left\| C_0^{\ell - \frac{\beta}{2}} u \right\|, \ \forall u \in X^{\beta - 2\ell}$;
3. $\left\| C_0^{-\frac{\rho}{2}} C_1^{\frac{1}{2}} u \right\| \leq c_1 \left\| C_0^{\frac{\beta-\rho}{2}} u \right\|, \ \forall u \in X^{\rho-\beta}, \ \forall \rho < \beta + \sigma_0 - 1$;
4. $\left\| C_0^{\frac{s}{2}} C_1^{-\frac{1}{2}} u \right\| \leq c_2 \left\| C_0^{\frac{s-\beta}{2}} u \right\|, \ \forall u \in X^{\beta - s}, \ \forall s \in (s_0, 1]$;
5. $\left\| C_0^{-\frac{s}{2}} C_1^{-\frac{1}{2}} \mathcal{A}^{-1} u \right\| \leq c_3 \left\| C_0^{\frac{2\ell-\beta-s}{2}} u \right\|, \ \forall u \in X^{s+\beta-2\ell}, \ \forall s \in (s_0, 1]$;
6. $\left\| C_0^{\frac{\mu}{2}} \mathcal{A}^{-1} C_1^{-1} u \right\| \leq c_4 \left\| C_0^{\frac{\mu}{2}+\ell-\beta} u \right\|, \ \forall u \in X^{2\beta-2\ell-\mu}, \ \forall \mu \in [\beta - 2\ell, 1]$;
7. $\left\| \mathcal{A}^{-1} C_1^{-1} \mathcal{A}^{-1} u \right\| \leq c_5 \left\| C_0^{2\ell-\beta} u \right\|, \ \forall u \in X^{2\beta - 4\ell}$.

Making Sense of (3), $B_\lambda m = \mathcal{A}^{-1} C_1^{-1} y$

- Consider
$$B_\lambda w = r, \tag{4}$$
recalling $B_\lambda = \mathcal{A}^{-1} C_1^{-1} \mathcal{A}^{-1} + \lambda C_0^{-1}$.
- Let $r \in X^{-1}$. An element $z \in X^1$ is called a weak solution of (4), if $\forall v \in X^1$
$$\langle B_\lambda z, v \rangle = \langle r, v \rangle .$$

Proposition 1

Under the Assumptions 2, a unique weak solution of (4) exists.

- Proof: Apply Lax-Milgram since $B(u, v) := \langle B_\lambda u, v \rangle$ coercive and continuous bilinear form on $X^1 \times X^1$.$\square$

Additional Assumption

We make the following additional assumption needed to show that the posterior is Gaussian and justify the formulae (2) and (3).

Assumption 3

If $r \in X$ then the weak solution $w \in X^1$ of the equation (4) satisfies

$$\left\| \mathcal{A}^{-1} C_1^{-1} \mathcal{A}^{-1} w \right\| \leq c \left\| r \right\| ,$$

for some constant $c > 0$.

Outline

1. Introduction/Background
2. Statement of Main Results
3. Setting/Assumptions
4. Proof of Theorem 1
5. Proof of Theorem 2
6. Conclusions

Sketch of Proof of Theorem 1

Want to show for $y = \mathcal{A}^{-1}u + \frac{1}{\sqrt{n}}\xi$, where $u \sim \mathcal{N}(0, \tau^2 C_0)$, $\xi \sim \mathcal{N}(0, C_1)$ that $\mu^y = \mathcal{N}(m, \frac{1}{n}B_\lambda^{-1})$, where m, B_λ given by (3), (2).

- Characterize μ^y through Radon-Nikodym derivative wrt prior, μ_0:

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

$$\Phi(u,y) := \frac{n}{2}\left\| C_1^{-\frac{1}{2}}\mathcal{A}^{-1}u\right\|^2 - n\left\langle C_1^{-\frac{1}{2}}y, C_1^{-\frac{1}{2}}\mathcal{A}^{-1}u\right\rangle$$

  $Z(y)$ normalizing constant.
- This gives existence and Gaussianity of posterior.
- Finite dim approximation in Karhunen-Loeve basis of prior:

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

  $\Phi^N(u,y) = \Phi(P^N u; y)$, P^N projection onto $\mathcal{X}^N$:=span$\{\phi_1, \ldots, \phi_N\}$

Sketch of Proof of Theorem 1

- Lemma 1 gives $\mathcal{A}^{-1}C_1^{-1}y \in X^{-1}$ thus $\exists! m \in X^1$ s.t. $B_\lambda m = \mathcal{A}^{-1}C_1^{-1}y$.
- Let $C = \frac{1}{n}B_\lambda^{-1}$. Completing the square gives Galerkin Approximation of (3), (2): $\mu^{N,y} = \mathcal{N}(m^N, C^N)$ where

$$m^N = \begin{cases} \tilde{m}^N, & in \quad \mathcal{X}^N \\ 0, & in \quad (\mathcal{X}^N)^\perp \end{cases} \quad \text{and} \quad C^N = \begin{cases} \tilde{C}^N, & in \quad \mathcal{X}^N \\ C_0, & in \quad (\mathcal{X}^N)^\perp \end{cases},$$

$(\tilde{C}^N)^{-1} = P^N C^{-1} P^N$ and $\tilde{m}^N = n\tilde{C}^N P^N \mathcal{A}^{-1}C_1^{-1}y$.

- Pass to the limit
  - $\mu^{N,y} \Rightarrow \mu^y$
  - $\mu^{N,y} \Rightarrow \mathcal{N}(m, \frac{1}{n}B_\lambda^{-1})$.

  Uniqueness of weak limit completes the proof.

Characterization of Posterior Mean and Covariance

Proposition

i)

$$\lim_{N\to\infty} \left\| m^N - m \right\| = 0,$$

ii)

$$\lim_{N\to\infty} \left\| \sqrt{C^N} - \sqrt{C} \right\|_{HS} = 0.$$

i) Galerkin Orthogonality: $B(m - m^N, v) = 0,\ \forall v \in P^N X^1.$

For any $z \in P^N X^1$

$$c \left\| m - m^N \right\|_1^2 \le B(m - m^N, m - m^N) = B(m - m^N, m - z) \le c \left\| m - m^N \right\|_1 \left\| m - z \right\|_1$$

$$z = P^N m \quad \Rightarrow \quad \left\| m - m^N \right\|_1^2 \le c \left\| m - P^N m \right\|_1^2 = \sum_{k>N} \langle m, \phi_k \rangle^2 \lambda_k^{-2} \xrightarrow{N} 0,$$

since $m \in X^1$ thus $\sum_{k=1}^{\infty} \langle m, \phi_k \rangle^2 \lambda_k^{-2} < \infty$.

Posterior Justification

ii) We have

$$\left\|\sqrt{C^N}-\sqrt{C}\right\|_{HS}=\sum_{k>N}\left\|\left(\sqrt{C_0}-\sqrt{C}\right)\phi_k\right\|^2.$$

Let $r_k \in X^1, k \in \mathbb{N}$ be the unique weak solution of $C^{-1}r_k=\phi_k$. Then

$$\sum_{k>N}\left\|\left(\sqrt{C_0}-\sqrt{C}\right)\phi_k\right\|^2 \le \sum_{k>N}\left(\left\|\sqrt{C_0}\phi_k\right\|+\left\|\sqrt{C}\phi_k\right\|\right)^2$$

$$\le 2\left(\sum_{k>N}\langle C_0\phi_k,\phi_k\rangle+\sum_{k>N}\langle C\phi_k,\phi_k\rangle\right)=2\left(\sum_{k>N}\langle C_0\phi_k,\phi_k\rangle+\sum_{k>N}\langle r_k,\phi_k\rangle\right).$$

First sum goes to 0 as $N\to\infty$.

Posterior Justification

By definition of r_k, for $v_k = \tau^2 C_0 \phi_k \in X^1$

$$\sum_{k>N} \langle \phi_k, \tau^2 C_0 \phi_k \rangle = \sum_{k>N} \langle C^{-1} r_k, \tau^2 C_0 \phi_k \rangle$$

$$= \sum_{k>N} \langle n\mathcal{A}^{-1} C_1^{-1} \mathcal{A}^{-1} r_k, \tau^2 C_0 \phi_k \rangle + \sum_{k>N} \left\langle \frac{1}{\tau^2} C_0^{-1} r_k, \tau^2 C_0 \phi_k \right\rangle,$$

thus

$$\sum_{k>N} \langle r_k, \phi_k \rangle = \sum_{k>N} \langle \tau^2 C_0 \phi_k, \phi_k \rangle - \sum_{k>N} \langle n\tau^2 \mathcal{A}^{-1} C_1^{-1} \mathcal{A}^{-1} r_k, C_0 \phi_k \rangle$$

$$\leq \sum_{k>N} \left| \langle \tau^2 C_0 \phi_k, \phi_k \rangle \right| + \sum_{k>N} \left\| n\tau^2 \mathcal{A}^{-1} C_1^{-1} \mathcal{A}^{-1} r_k \right\| \left\| C_0 \phi_k \right\|$$

$$\leq \sum_{k>N} \left| \langle \tau^2 C_0 \phi_k, \phi_k \rangle \right| + \sum_{k>N} \left\| n\tau^2 \phi_k \right\| \left\| C_0 \phi_k \right\| \leq c \sum_{k>N} \left\| C_0 \phi_k \right\| \overset{N\to\infty}{\longrightarrow} 0,$$

where second inequality holds by Assumption 3 since $\phi_k \in X$. $\square$

Outline

Posterior Consistency - Convergence Rates

Recall we assume data $y^\dagger = \mathcal{A}^{-1}u^\dagger + \frac{1}{\sqrt{n}}\xi$ and want to estimate the error $m_\lambda^\dagger - u^\dagger$.

$$B_\lambda u^\dagger = \mathcal{A}^{-1}C_1^{-1}y^\dagger + \lambda C_0^{-1}u^\dagger - \frac{1}{\sqrt{n}}\mathcal{A}^{-1}C_1^{-1}\xi$$

$$B_\lambda m_\lambda^\dagger = \mathcal{A}^{-1}C_1^{-1}y^\dagger$$

therefore

$$m_\lambda^\dagger - u^\dagger = \frac{1}{\sqrt{n}}B_\lambda^{-1}\mathcal{A}^{-1}C_1^{-1}\xi - \lambda B_\lambda^{-1}C_0^{-1}u^\dagger.$$

Since ξ has zero mean, we have that for any $s \in \mathbb{R}$

$$\mathbb{E}\left\|m_\lambda^\dagger - u^\dagger\right\|_s^2 = \lambda^2\left\|B_\lambda^{-1}C_0^{-1}u^\dagger\right\|_s^2 + \frac{1}{n}\mathbb{E}\left\|B_\lambda^{-1}\mathcal{A}^{-1}C_1^{-1}\xi\right\|_s^2.$$

Operator Norm Bounds

Lemma 2

$$\left\|B_\lambda^{-1}\mathcal{A}^{-1}C_1^{-1}\right\|_{\mathcal{L}(X^{2\beta-2\ell-\mu_1},X^{\beta-2\ell})} \leq c\lambda^{-\frac{\theta_1}{2}}$$

and

$$\left\|B_\lambda^{-1}\mathcal{A}^{-1}C_1^{-1}\right\|_{\mathcal{L}(X^{2\beta-2\ell-\mu_1},X^{1})} \leq c\lambda^{-\frac{\theta_1+1}{2}},$$

for every $\theta_1 \in [0,1]$, where $\mu_1 = (1-\theta_1)(\beta-2\ell)+\theta_1$.

Lemma 3

$$\left\|B_\lambda^{-1}C_0^{-1}\right\|_{\mathcal{L}(X^{2-\mu_2},X^{\beta-2\ell})} \leq c\lambda^{-\frac{\theta_2}{2}}$$

and

$$\left\|B_\lambda^{-1}C_0^{-1}\right\|_{\mathcal{L}(X^{2-\mu_2},X^{1})} \leq c\lambda^{-\frac{\theta_2+1}{2}},$$

for every $\theta_2 \in [0,1]$, where $\mu_2 = (1-\theta_2)(\beta-2\ell)+\theta_2$.

Operator Norm Bounds

Proof of Lemma 3: Let $h \in X^{2-\mu_2}$. Then $r := C_0^{-1}h \in X^{-1}$. Indeed,

$$\left\| C_0^{-1}h \right\|_{-1} = \left\| C_0^{\frac{1}{2}} C_0^{-1}h \right\| = \|h\|_1 \leq c \, \|h\|_{2-\mu_2} ,$$

since $\mu_2 \leq 1$. Therefore, $\exists! z \in X^1$ weak solution of $B_\lambda w = r$, i.e. $\forall v \in X^1$

$$\langle B_\lambda z, v \rangle = \left\langle C_0^{-1}h, v \right\rangle .$$

Choose $v = z$

$$\left\langle \mathcal{A}^{-1} C_1^{-1} \mathcal{A}^{-1} z, z \right\rangle + \lambda \left\langle C_0^{-1} z, z \right\rangle = \left\langle C_0^{\frac{\mu_2}{2}-1} h, C_0^{-\frac{\mu_2}{2}} z \right\rangle$$

and by Assumption 2 and the Cauchy-Schwarz inequality on rhs, we get

$$\|z\|_{\beta-2\ell}^2 + \lambda \, \|z\|_1^2 \leq c \left\| C_0^{\frac{\mu_2}{2}-1} h \right\| \|z\|_{\mu_2} .$$

Operator Norm Bounds

Interpolate norm on z on the rhs between norms on z on the lhs, then use Cauchy with ε inequality and then Young's inequality for $p = \frac{1}{1-\theta_2}$, $q = \frac{1}{\theta_2}$ to get

$$\|z\|_{\beta-2\ell}^2 + \lambda \|z\|_1^2 \leq c \left\| C_0^{\frac{\mu_2}{2}-1} h \right\| \|z\|_{\beta-2\ell}^{1-\theta_2} \lambda^{-\frac{\theta_2}{2}} \left(\lambda^{\frac{1}{2}} \|z\|_1\right)^{\theta_2}$$

$$\leq \frac{c}{2\varepsilon}\left(\lambda^{-\theta_2} \left\| C_0^{\frac{\mu_2}{2}-1} h \right\|^2\right) + \frac{c\varepsilon}{2}\left(\|z\|_{\beta-2\ell}^{2(1-\theta_2)} \left(\lambda^{\frac{1}{2}} \|z\|_1\right)^{2\theta_2}\right)$$

$$\leq \frac{c}{2\varepsilon}\left(\lambda^{-\theta_2} \left\| C_0^{\frac{\mu_2}{2}-1} h \right\|^2\right) + \frac{c\varepsilon}{2}\left((1-\theta_2)\|z\|_{\beta-2\ell}^2 + \theta_2 \lambda \|z\|_1^2\right).$$

Choose $\varepsilon > 0$ small to get, for $c > 0$ independent of θ, λ,

$$\|z\|_{\beta-2\ell} \leq c\lambda^{-\frac{\theta_2}{2}} \left\| C_0^{\frac{\mu_2}{2}-1} h \right\| \quad \text{and} \quad \|z\|_1 \leq c\lambda^{-\frac{\theta_2+1}{2}} \left\| C_0^{\frac{\mu_2}{2}-1} h \right\|.$$

Replace $z = B_\lambda^{-1} C_0^{-1} h$ to get the result. □

Posterior Consistency - Convergence Rates

Proof of Theorem 2: We use the norm bounds to get

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

and

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

Optimize by choosing $\lambda = \lambda(n) = n^p$ such that $\lambda^{2-\theta_2} = \frac{1}{n}\lambda^{-\theta_1}$ i.e. $p = \frac{-1}{2-\theta_2+\theta_1}$ to obtain that

$$\mathbb{E}\left\|m_\lambda^\dagger - u^\dagger\right\|_{\beta-2\ell}^2 \leq c\mathbb{E}(\kappa^2)n^{\frac{\theta_2-2}{\theta_1-\theta_2+2}} \quad \text{and} \quad \mathbb{E}\left\|m_\lambda^\dagger - u^\dagger\right\|_1^2 \leq c\mathbb{E}(\kappa^2)n^{\frac{\theta_2-1}{\theta_1-\theta_2+2}}.$$

Interpolate between two last estimates to obtain the claimed rate.

□

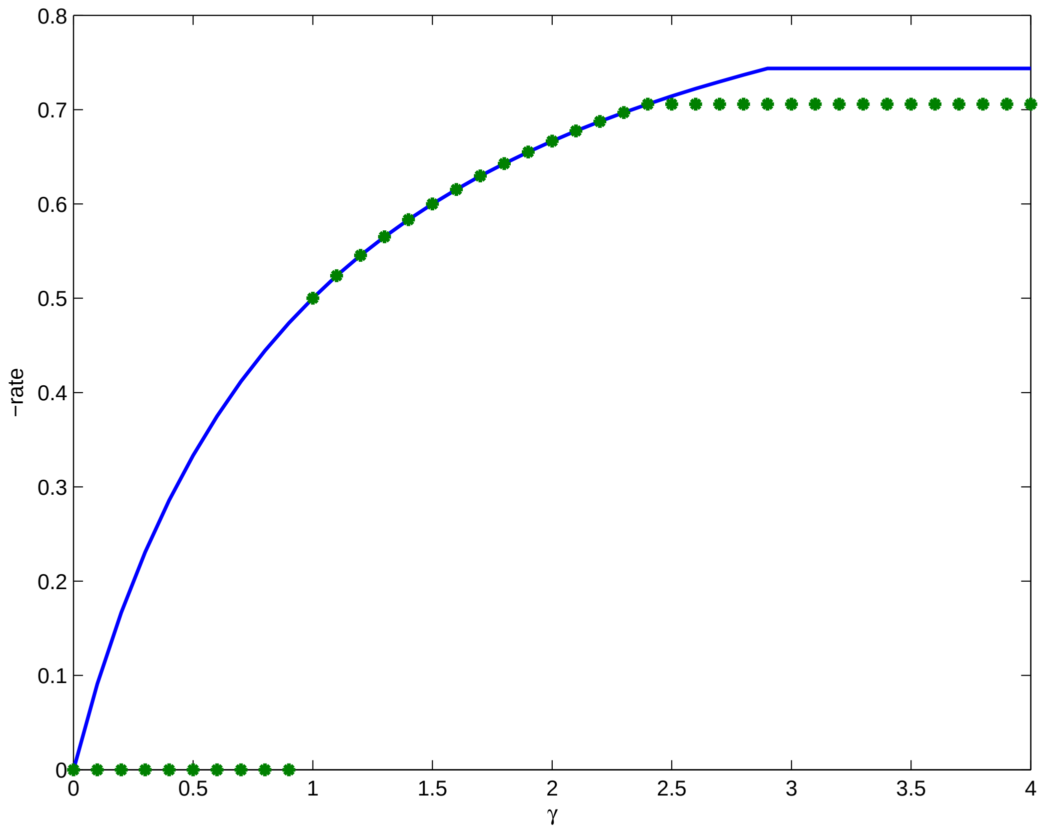
Outline

Optimality

Diagonal Case: $C_0 = (-\Delta)^{-1}$, $\mathcal{A}^{-1} = C_0^{\ell}$ and $C_1 = C_0^{\beta}$ gives sharp rates.

Our assumptions are satisfied trivially thus we can compare to our rates.

- We assume $u^{\dagger} \in X^{\gamma}$ and calculate the rate of convergence with both methods for $\ell = \beta = 1/2$.



Synopsis

We have

- Considered a linear Inverse Problem in infinite dimensions;
- Adopted a Bayesian approach with Gaussian prior and Gaussian noise;
- Identified the posterior as Gaussian and characterized its mean and covariance;
- Shown posterior consistency results in the small noise limit.

- Thank You!

References - Further Reading

- Agapiou, S., Larsson, S. and Stuart, A. M., *Posterior Consistency of the Bayesian Approach to Linear Ill-Posed Inverse Problems*, work in progress.
- Bogachev, Vladimir I., *Gaussian Measures*, Mathematical Surveys and Monographs, vol. 62, American Mathematical Society, 1998.
- Engl, Heinz W. and Hanke, Martin and Neubauer, Andreas, *Regularization of inverse problems*, Mathematics and its Applications, vol. 375, Kluwer Academic Publishers Group, Dordrecht, 1996.
- Knapik, B. T., van der Vaart, A. W., van Zanten, J. H., *Bayesian Inverse Problems*, to appear.
- Mandelbaum, Avi, *Linear estimators and measurable linear transformations on a Hilbert space*, Z. Wahrsch. Verw. Gebiete 65 (1984), no. 3, 385-397.
- Stuart, A. M., *Inverse problems: a Bayesian perspective*, Acta Numer. 19 (2010), 451-559.