High Dimensional Bayesian Inverse Problems

Sergios Agapiou

Mathematics Institute
University of Warwick, UK

9^{th} Conference on Bayesian Nonparametrics
10-14 June 2013, Amsterdam

Outline

Bayesian Inverse Problems

- $(X, \langle \cdot, \cdot \rangle, \| \cdot \|)$ separable Hilbert space.
- Probabilistic approach to problem of recovering u from noisy, indirect observations, y.
- Work with Gaussian measures in X.
- Regularity of $u \sim N(m, \mathcal{C})$ determined by decay of spectrum of $\mathcal{C}$.

Bayesian Linear Inverse Problem

Consider linear case

$$y = Ku + \xi,$$

$K : X \to X$ bounded, selfadjoint, positive definite.

Assume $\xi \sim N(0, \lambda^{-1}\mathcal{C}_1)$, $\mathcal{C}_1 : X \to X$ selfadjoint, positive definite.

- *Likelihood*: $y|u \sim N(Ku, \lambda^{-1}\mathcal{C}_1)$.
- *Prior*: $u \sim \mu_0 = N(0, \delta^{-1}\mathcal{C}_0)$, $\mathcal{C}_0 : X \to X$ selfadjoint positive definite trace class.
- *Posterior* distribution on $u|y$, μ^y.

Bayesian Linear Inverse Problem - Posterior

Formally

$$\mu^y(u) \propto \exp(-\frac{\lambda}{2}\|\mathcal{C}_1^{-\frac{1}{2}}(Ku-y)\|^2 - \frac{\delta}{2}\|\mathcal{C}_0^{-\frac{1}{2}}u\|^2).$$

Theorem 1 (A., Larsson, Stuart '12)

Under reasonable assumptions $\mu^y \ll \mu_0$ and $\mu^y = N(m, \mathcal{C})$, where

$$\mathcal{C}^{-1} = \lambda K \mathcal{C}_1^{-1} K + \delta \mathcal{C}_0^{-1},$$

$$\mathcal{C}^{-1} m = \lambda K \mathcal{C}_1^{-1} y.$$

(precision operators)

- m minimizer of penalized least squares functional

$$J(u) = \|\mathcal{C}_1^{-\frac{1}{2}}(Ku-y)\|^2 + \frac{\delta}{\lambda}\|\mathcal{C}_0^{-\frac{1}{2}}u\|^2.$$

Bayesian Linear Inverse Problem - Posterior

- Mandelbaum '84 and Lehtinen et al. '89 show posterior is Gaussian and provide formulae for mean and covariance. We give alternative formulae using unbounded precision operators (Lax-Milgram).
- To apply standard conditioning argument, require likelihood to be equivalent to noise distribution: $Ku \in \mathcal{D}(\mathcal{C}_1^{-\frac{1}{2}})$ μ_0-almost surely.
- Cover non diagonal class of mildly ill-posed problems under norm equivalence assumptions relating $K, \mathcal{C}_0, \mathcal{C}_1$.
- Cover diagonal case for exponentially ill-posed problems.
- Method used in Pokern et al. '12 for Nonparametric Drift Estimation using local time.

Frequentist Posterior Consistency

- Consider

$$y^\dagger = Ku^\dagger + \xi,$$

$u^\dagger$ underlying truth, $\xi \sim N(0, \lambda^{-1}\mathcal{C}_1)$. Gives $\mu^{y^\dagger} = N(m^\dagger, \mathcal{C})$.

- Posterior Consistency: can we recover truth in small noise limit, $\lambda \to \infty$?

- Given γ regularity of the truth, does the posterior contract to the truth at the *minimax rate*, mrate=mrate(γ)?

Example

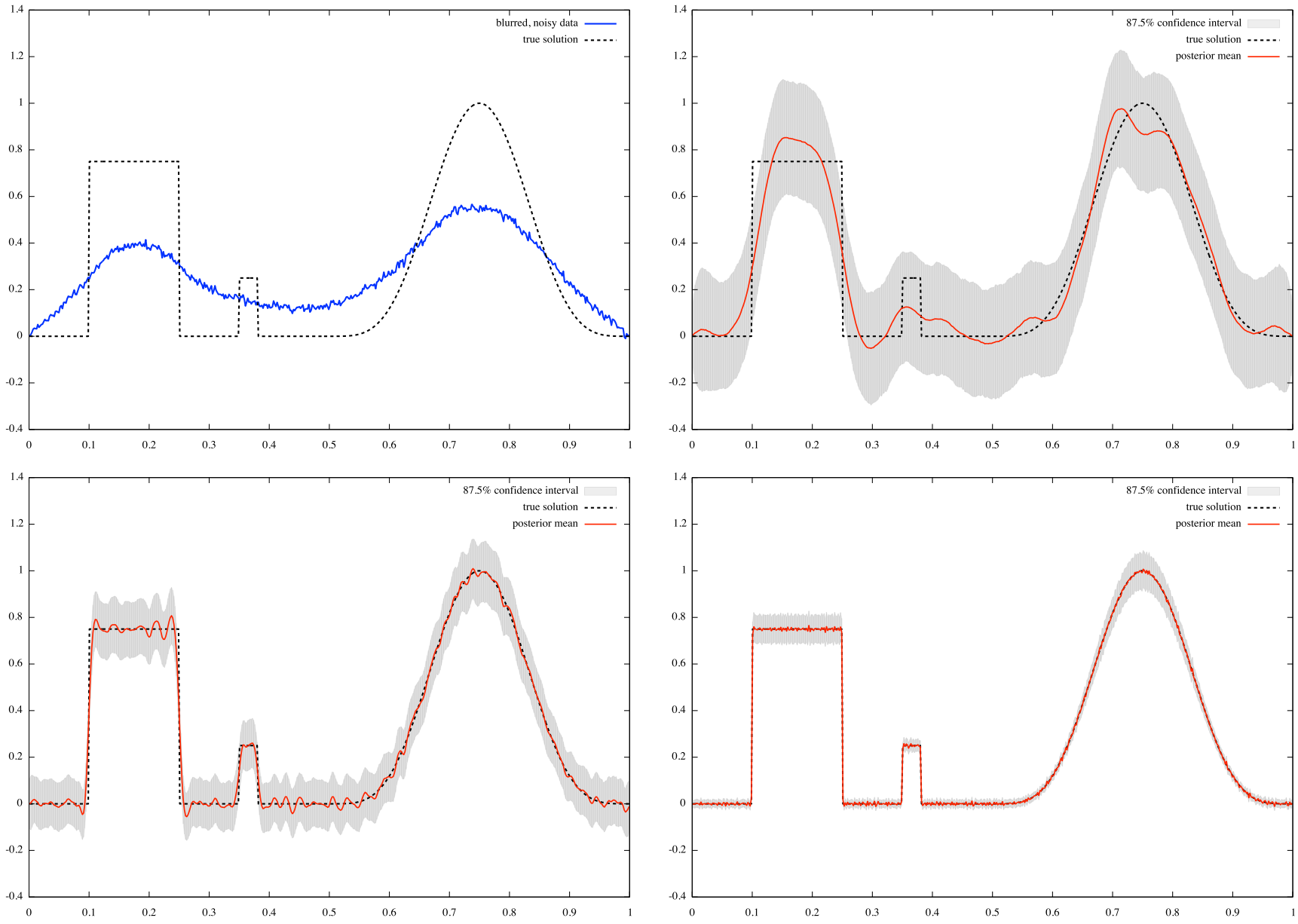


Figure: As λ increases, ad hoc choice of δ to achieve good reconstruction.

Frequentist Posterior Consistency

- s regularity of prior.

Theorem 2 (A., Larsson, Stuart '12 & A., Stuart, Zhang '12)

For γ *a priori* known regularity of $u^\dagger$,

$$\mathbb{E}\|m^\dagger - u^\dagger\|^2 + \mathrm{Tr}\,(\mathcal{C}) = \mathcal{O}(\lambda^{-\text{mrate}\,(+\varepsilon)}),$$

when

- either prior matches regularity of the truth $s = \gamma$, δ fixed;
- or we appropriately choose $\delta = \delta(\lambda; \gamma, s)$.

- Rate-adaptation achieved by rescaling the prior.
- Generalize Knapik et al. '11 and '12.

Frequentist Posterior Consistency - Non-diagonal Method

- Mean

$$(\lambda\mathcal{C})^{-1}m^\dagger = K\mathcal{C}_1^{-1}y^\dagger = \underbrace{K\mathcal{C}_1^{-1}Ku^\dagger} + K\mathcal{C}_1^{-1}\xi.$$

- Truth

$$(\lambda\mathcal{C})^{-1}u^\dagger = \underbrace{K\mathcal{C}_1^{-1}Ku^\dagger} + \frac{\delta}{\lambda}\mathcal{C}_0^{-1}u^\dagger.$$

- Error $e = m^\dagger - u^\dagger$ satisfies weak equation

$$(\lambda\mathcal{C})^{-1}e = K\mathcal{C}_1^{-1}\xi - \frac{\delta}{\lambda}\mathcal{C}_0^{-1}u^\dagger.$$

Frequentist Posterior Consistency - Non-diagonal Method

Hilbert Scale $X^t = \mathcal{D}(\mathcal{C}_0^{-\frac{t}{2}})$, norm $\|\cdot\|_t = \|\mathcal{C}_0^{-\frac{t}{2}}\cdot\|$.

As $\lambda \to \infty$ and $\frac{\delta}{\lambda} \to 0$

- terms in parenthesis go to zero in weak spaces: for t_i sufficiently large

$$\mathbb{E}\|K\mathcal{C}_1^{-\frac{1}{2}}\xi\|_{-t_1} = \mathcal{O}(\lambda^{-\frac{1}{2}}), \qquad \|\frac{\delta}{\lambda}\mathcal{C}_0^{-1}u^\dagger\|_{-t_2} = \mathcal{O}\Big(\frac{\delta}{\lambda}\Big)$$

- operator $\lambda\mathcal{C} = (K\mathcal{C}_1^{-1}K + \frac{\delta}{\lambda}\mathcal{C}_0^{-1})^{-1}$ blows-up in these spaces.

Frequentist Posterior Consistency - Non-diagonal Method

Lemma (A., Larsson, Stuart 12)

Interpolation between two terms in $\lambda\mathcal{C}$ yields for $t \in [t_0, 1]$

$$\|\lambda\mathcal{C}\|_{\mathcal{L}(X^{-t},X)} = \mathcal{O}\big(\big(\frac{\delta}{\lambda}\big)^{-c}\big),$$

$c \in (0,1)$ increasing in t.

- Rates of convergence obtained by comparing the rates of decay and blow-up

$$\mathbb{E}\|e\| = \mathcal{O}\big(\big(\frac{\delta}{\lambda}\big)^{-c_1}\lambda^{-\frac{1}{2}}\big) + \mathcal{O}\big(\big(\frac{\delta}{\lambda}\big)^{1-c_2}\big).$$

- Rates optimized by choosing $\delta = \delta(\lambda; \gamma, s)$ to balance two contributions.

Frequentist Posterior Consistency - Abstract Method

- Other conjugate-Gaussian problems (NP Drift Estimation via local time in Pokern et al. '12, NP Drift Estimation via Fourier series).
- Posterior consistency as quality of data y_λ improves, $\lambda \to \infty$.
- General weak error equation

$$e = \lambda \mathcal{C}(\sigma(y_\lambda; \lambda) - \frac{\delta}{\lambda}\mathcal{C}_0^{-1}u^\dagger)$$

$$\lambda\mathcal{C} = \left(Q(y_\lambda; \lambda) + \frac{\delta}{\lambda}\mathcal{C}_0^{-1}\right)^{-1}$$

  $\sigma(y_\lambda, \lambda)$ "noise" term, $Q(y_\lambda; \lambda)$ lower order.
- As $\lambda \to \infty$, $\frac{\delta}{\lambda} \to 0$ require results showing decay of $\sigma(y_\lambda, \lambda)$ in weak spaces.
- Interpolation yields blow-up rates for $\lambda\mathcal{C}$.

http://homepages.warwick.ac.uk/∼mariba/


S. Agapiou, S. Larsson and A. M. Stuart, *Posterior contraction rates for the Bayesian approach to linear ill-posed inverse problems*, Accepted Stoch. Proc. Appl., http://arxiv.org/abs/1203.5753


S. Agapiou, A. M. Stuart and Y. X. Zhang, *Bayesian posterior contraction rates for linear severely ill-posed inverse problems*, http://arxiv.org/abs/1210.1563


S. Agapiou, A. M. Stuart and Y. X. Zhang, *An abstract framework for the study of posterior contraction in Bayesian inverse problems*, in preparation.

Hierarchical Bayesian Linear Inverse Problem

As Bardsley '12, use *hierarchical setup* for inference on unknown and scaling parameters.

- $\xi|\lambda \sim N(0, \lambda^{-1}\mathcal{C}_1)$, where $\lambda \sim \mathrm{Ga}(\alpha_\lambda, \beta_\lambda)$.
- Hierarchical prior: $u|\delta \sim N(0, \delta^{-1}\mathcal{C}_0)$, where $\delta \sim \mathrm{Ga}(\alpha_\delta, \beta_\delta)$.

Implement in $\mathbb{R}^N$, *Bayes'* theorem

$$\mathbb{P}(u, \lambda, \delta|y) \propto \lambda^{\frac{N}{2}+\alpha_\lambda-1}\delta^{\frac{N}{2}+\alpha_\delta-1}\exp\left(-\beta_\lambda\lambda - \beta_\delta\delta - \frac{\lambda}{2}\|\mathcal{C}_1^{-\frac{1}{2}}(Ku-y)\|^2 - \frac{\delta}{2}\|\mathcal{C}_0^{-\frac{1}{2}}u\|^2\right)$$

- Conditional conjugacy, Gibbs Sampler natural.
- AIM: Study robustness of Gibbs Sampler as $N \to \infty$.

Gibbs Sampler

0. Initialize $\delta^{(0)}$ and $\lambda^{(0)}$, and set $k = 0$;

1. Compute $u^{(k)} \sim N(m_{\lambda^{(k)},\delta^{(k)}}, \mathcal{C}_{\lambda^{(k)},\delta^{(k)}})$;

2. Compute $\lambda^{(k+1)} \sim \mathrm{Ga}(\alpha_\lambda + \frac{N}{2}, \beta_\lambda + \frac{1}{2}\|\mathcal{C}_1^{-\frac{1}{2}}(Ku^{(k)} - y)\|^2)$;

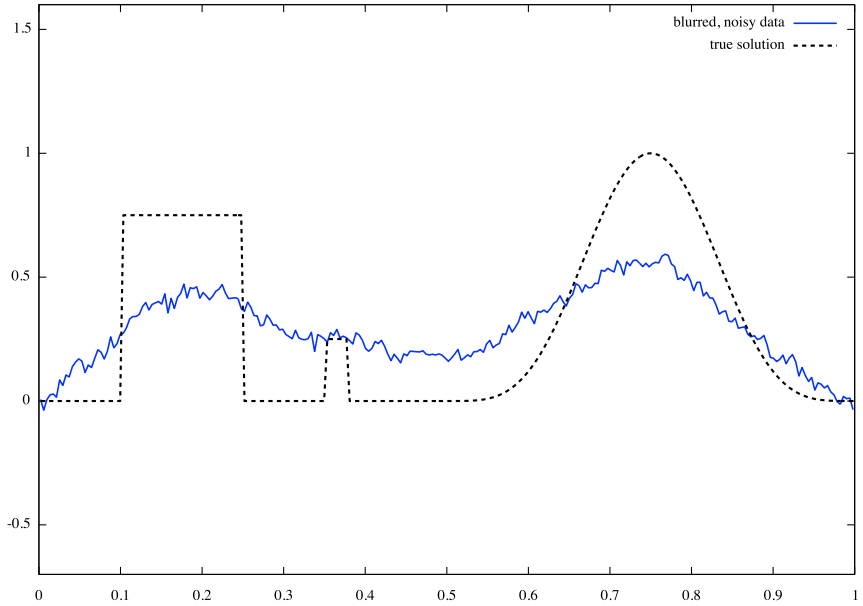
3. Compute $\delta^{(k+1)} \sim \mathrm{Ga}(\alpha_\delta + \frac{N}{2}, \beta_\delta + \frac{1}{2}\|\mathcal{C}_0^{-\frac{1}{2}}u^{(k)}\|^2)$;

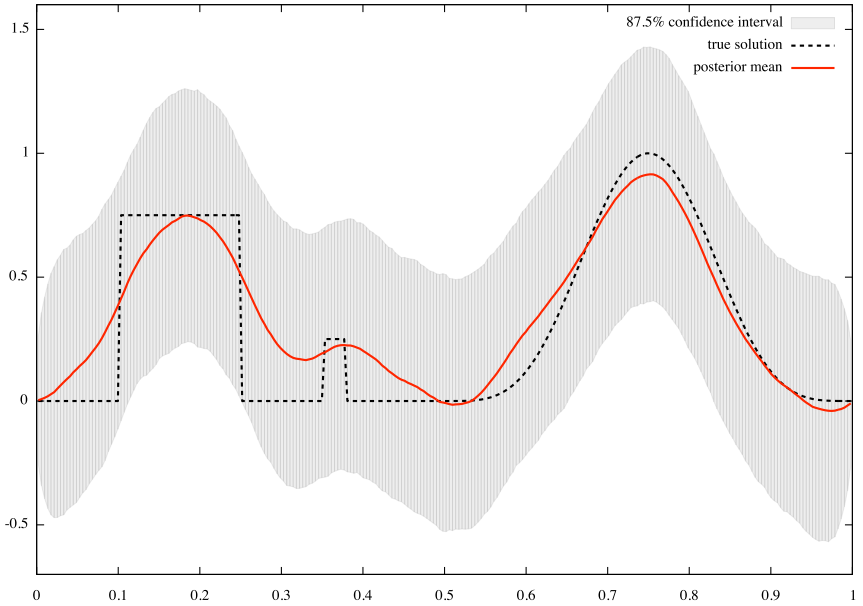
4. Set $k = k + 1$. If $k < k_{max}$ return to step 1, otherwise stop.

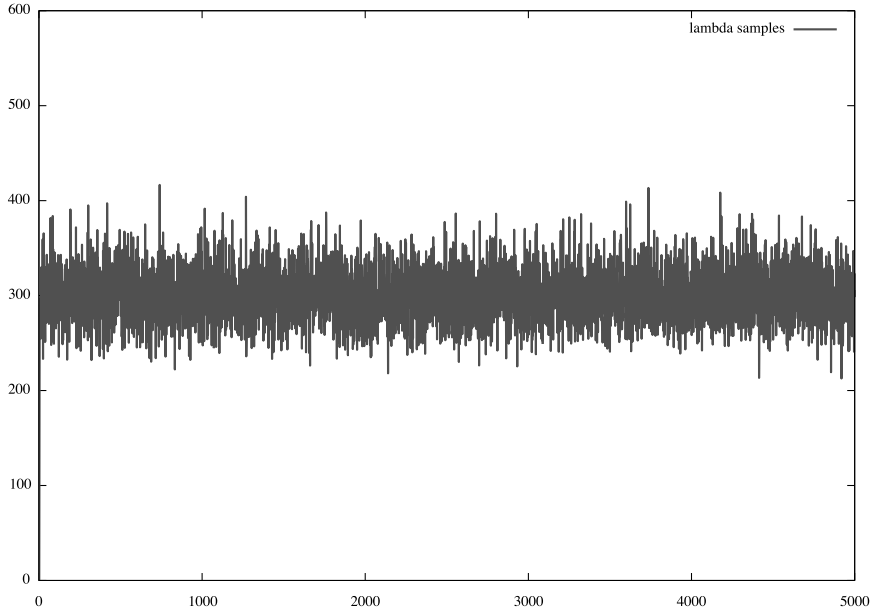
Consider

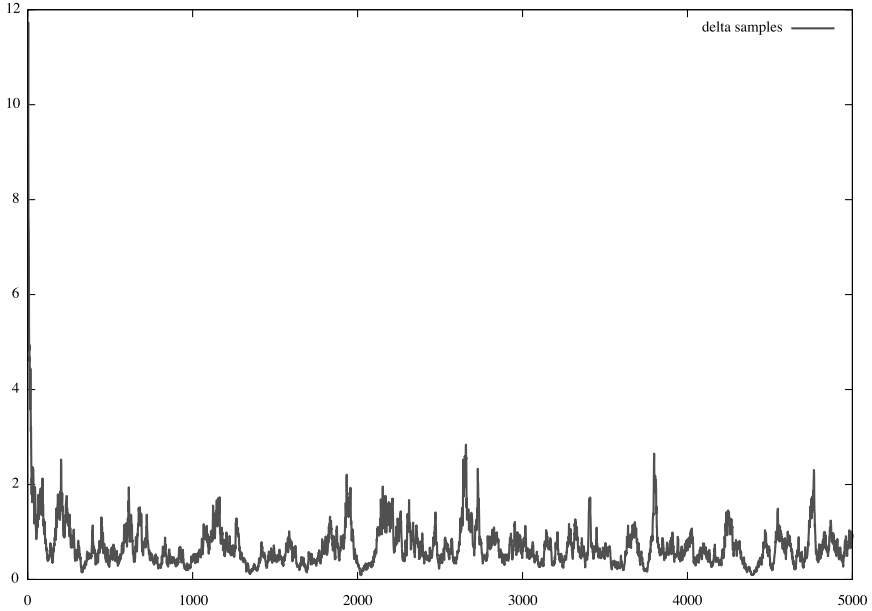
$$y^\dagger = Ku^\dagger + \lambda_0^{-\frac{1}{2}}\xi,\ \xi \sim N(0, \mathcal{C}_1).$$

Example - Hierarchical Linear Inverse Problem N=256

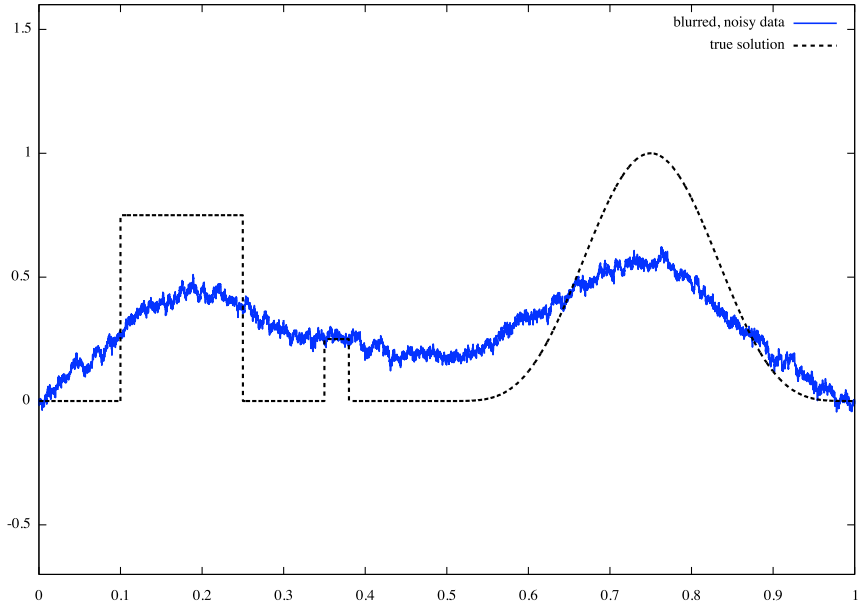


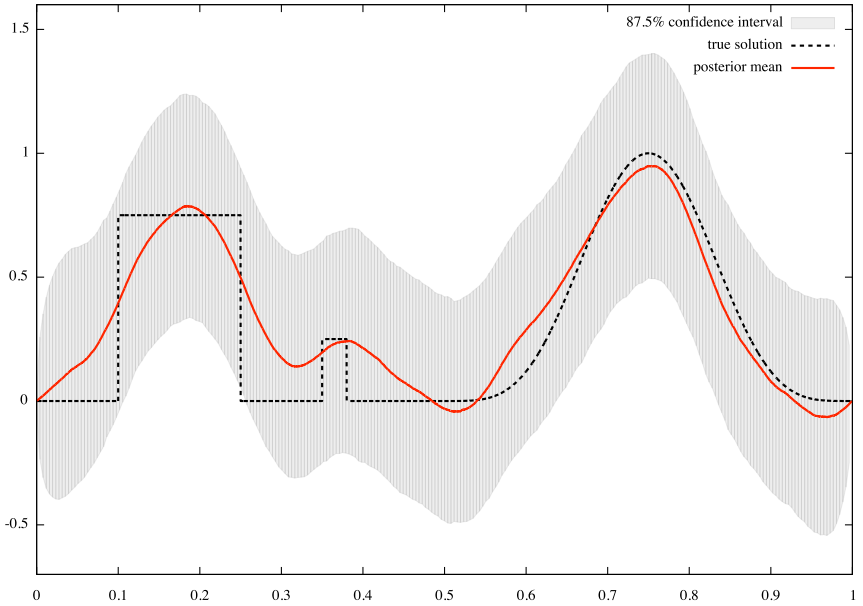


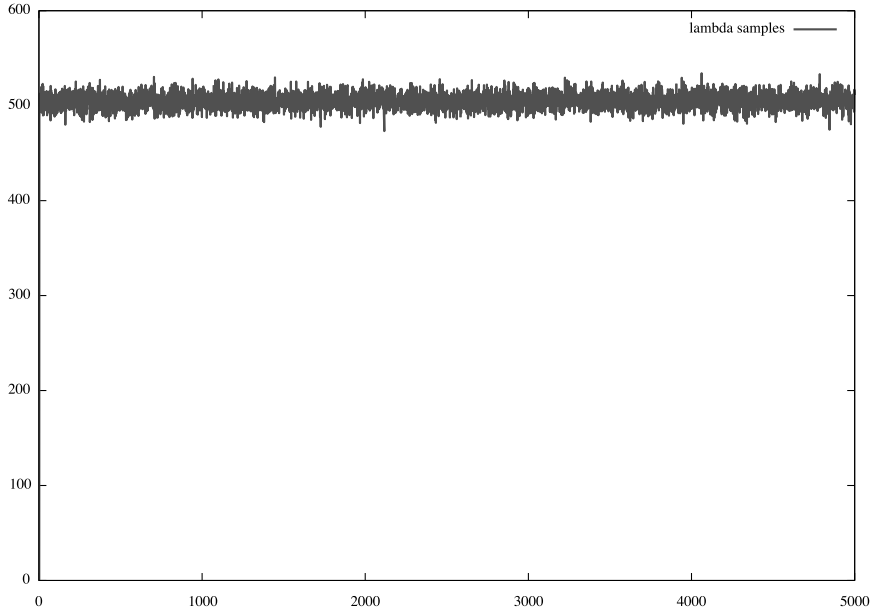


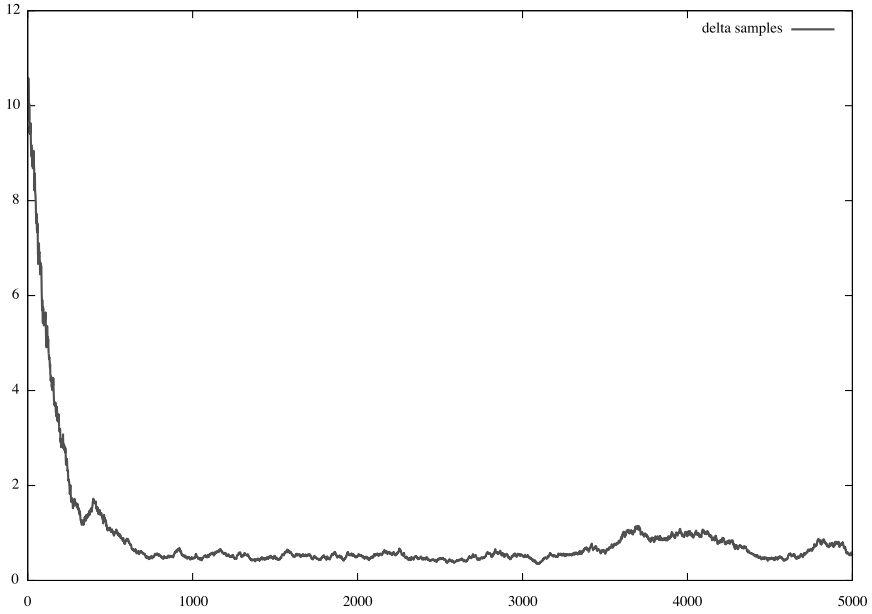


Example - Hierarchical Linear Inverse Problem N=8192



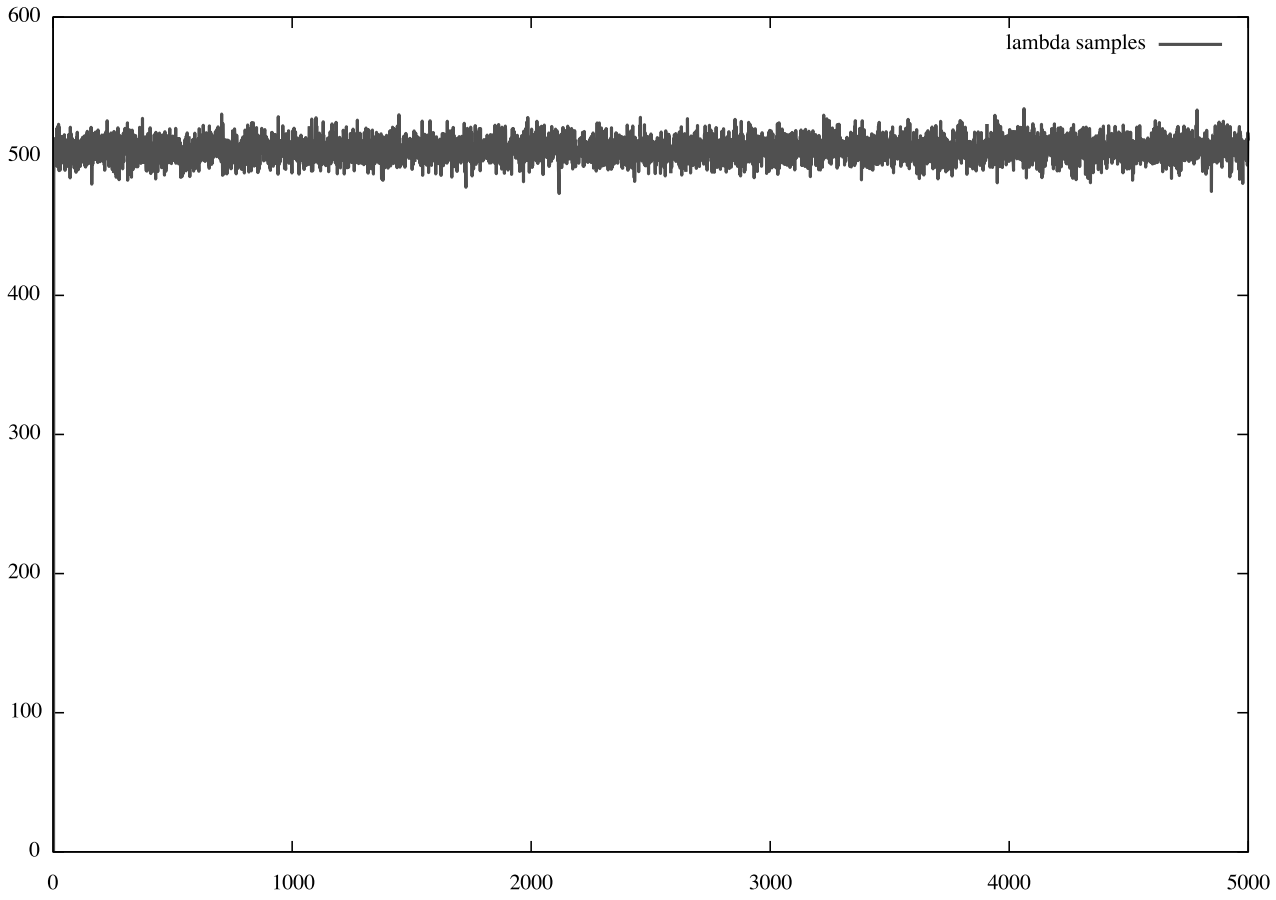






Behaviour of λ - Intuition

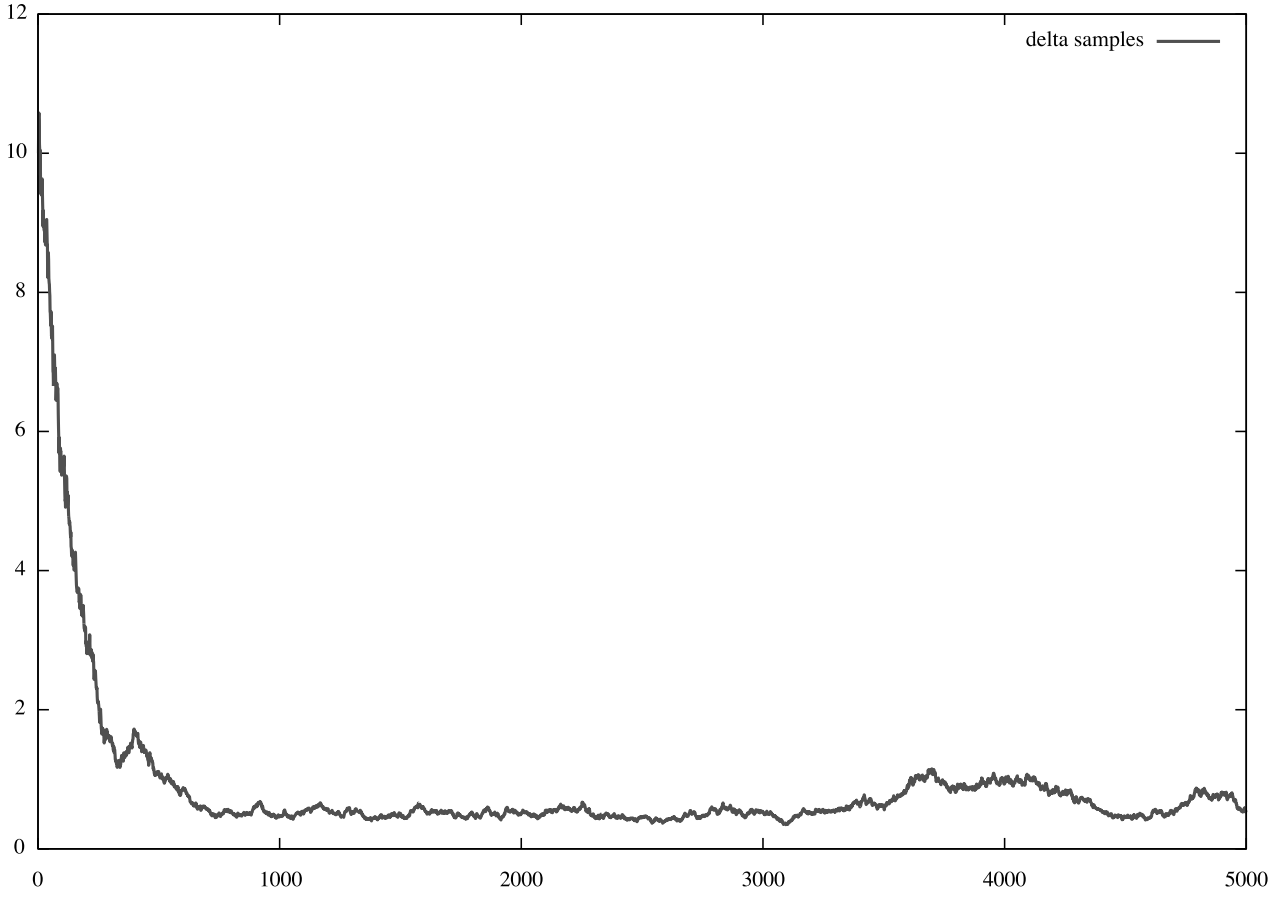
- Step 2 algorithm tries to find λ pretending to know y, u, δ.



Key point: infinite dimensional data contains infinite info about scaling of noise, instant learning of λ.

Behaviour of δ - Intuition

- Step 3 algorithm tries to find δ pretending to know y, u, λ.



Key point: infinite dimensional draw from posterior contains infinite info about scaling of prior, strong dependence between δ and u draws.

Different behaviour of λ and δ - Main Result

Theorem 3 (A., Bardsley, Papaspiliopoulos, Stuart '13)

In the limit $N \to \infty$

- For fixed δ

$$\lambda_N^{(k+1)} \overset{\mathcal{L}}{\simeq} \lambda_0(1 + \eta\sqrt{\frac{4}{N}}), \quad \eta \sim N(0,1);$$

- For fixed λ, $\{\delta_N^{(k)}\}_{k\in\mathbb{N}}$ approximated by

$$d\delta_s = \mu(\delta_s, \lambda)ds + \delta_s dB_s.$$

Similar conclusions hold when attempting to learn other parameters of prior and noise.

Further/Ongoing Work

- posterior consistency for drift estimation using Fourier expansion;
- expand algorithm analysis to nonparametric drift estimation;
- expand algorithm analysis to hierarchical setup for prior and noise regularity;
- rigorous proof of diffusion limits for the slow components;
- drawing intuition from Papaspiliopoulos et al. '07, propose reparametrizations to alleviate effect. e.g. algorithm suffers from dependence between components hence make them a priori independent. Components are a posteriori dependent since need to explain data, for very informative data algorithm deteriorates again.

http://homepages.warwick.ac.uk/∼mariba/


S. Agapiou, S. Larsson and A. M. Stuart, *Posterior contraction rates for the Bayesian approach to linear ill-posed inverse problems*, Accepted Stoch. Proc. Appl., http://arxiv.org/abs/1203.5753


S. Agapiou, A. M. Stuart and Y. X. Zhang, *Bayesian posterior contraction rates for linear severely ill-posed inverse problems*, http://arxiv.org/abs/1210.1563


S. Agapiou, J. Bardsley, O. Papaspiliopoulos and A. M. Stuart, *Dimension dependence of sampling algorithms, in Bayesian inverse problems*, in preparation.


O. Papaspiliopoulos, G. O. Roberts, and M. Sköld, *A general framework for the parametrization of hierarchical models*, Statist. Sci., 22(1):59-73, 2007


J. M. Bardsley, *MCMC-based image reconstruction with uncertainty quantification*, SIAM J. Sci. Comput. 34 (2012)


Y. Pokern, A. M. Stuart and J. H. Van Zanten, *Posterior consistency via precision operators for nonparametric drift estimation in SDEs*, Accepted Stoch. Proc. Appl., http://arxiv.org/abs/1202.0976