Analysis of the Gibbs Sampler for Hierarchical Inverse Problems

Sergios Agapiou

Mathematics Institute, University of Warwick

Joint work with Johnathan Bardsley (Montana), Omiros Papaspiliopoulos (Barcelona) and Andrew Stuart (Warwick)

CSC@lunchtime Seminar, 2nd December 2013

Outline

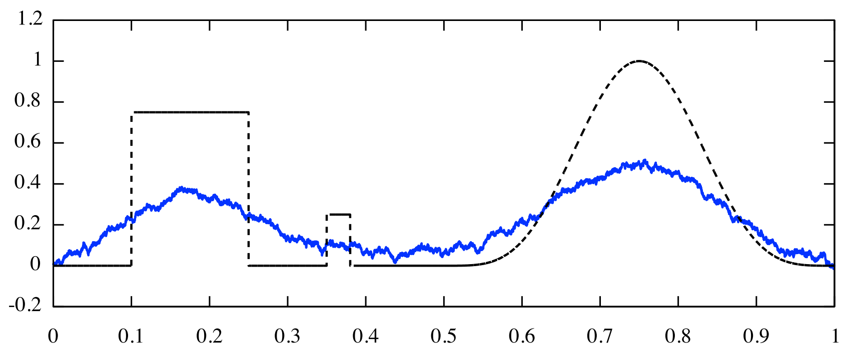
Linear Inverse Problems

- $(X, \langle \cdot, \cdot \rangle, \| \cdot \|)$ separable Hilbert space.

- AIM: recover u from blurred, noisy, observation, y.

$$y = Ku + \xi,$$

$K : X \to X$ blurring operator, ξ additive noise.



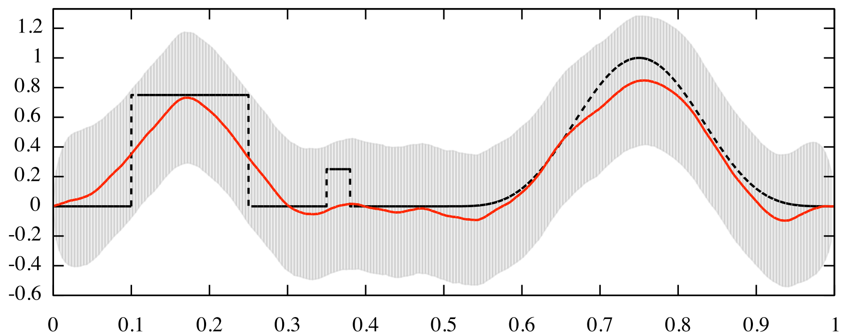
- Problem ill-posed: recovery of u very sensitive to errors in observation y (regularization).

Bayesian Approach to Inverse Problems

- *Likelihood*: distribution of $y|u$.
- *Prior*: distribution of u, encoding prior beliefs.
- *Posterior*: distribution of $u|y$, object of interest.

- Link: *Bayes'* theorem

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



Hierarchical Linear Inverse Problems

Assume $\xi \sim N(0, \lambda^{-1}\mathcal{C}_1)$, $\lambda > 0$ models amplitude, operator $\mathcal{C}_1$ regularity of noise.

- Likelihood is $y|u \sim N(Ku, \lambda^{-1}\mathcal{C}_1)$.
- Hierarchical prior $u|\delta \sim N(0, \delta^{-1}\mathcal{C}_0)$, where $\delta \sim \mathrm{Ga}(\alpha, \beta)$.

($\mathrm{Ga}(\alpha, \beta)$ has pdf $\delta^{\alpha-1}e^{-\beta\delta}$)

Hierarchical Linear Inverse Problems

- Implement in $\mathbb{R}^N$, discretize operators $\mathcal{C}_0, \mathcal{C}_1, K$.
- *Bayes'* theorem gives density of posterior on $\mathbb{R}^N \times \mathbb{R}$

$$p(u, \delta | y) \propto \delta^{\frac{N}{2}+\alpha-1} \exp\Big(-\beta\delta - \frac{\lambda}{2}\|\mathcal{C}_1^{-\frac{1}{2}}(Ku - y)\|^2 - \frac{\delta}{2}\|\mathcal{C}_0^{-\frac{1}{2}}u\|^2\Big)$$

- $u|y, \delta \sim N(m_\delta, \mathcal{C}_\delta)$ - complete the square
- $\delta|y, u \sim \mathrm{Ga}\left(\alpha + \frac{N}{2},\ \beta + \frac{1}{2}\|\mathcal{C}_0^{-\frac{1}{2}}u\|^2\right)$

- Challenge: Extract information from the posterior on $u, \delta | y$.

Sampling probability measures

- For independent samples $Z_1, Z_2, \ldots \sim \mu$. Monte Carlo principle:

$$\int f(z)\mu(dz) \approx \frac{1}{n}\sum_{k=1}^{n} f(Z_k)$$

- MCMC: construct Markov chain whose stationary distribution is μ. Under mild conditions, sample path $\{Z_1, Z_2, \ldots\}$ is approximate and dependent sample from μ and

$$\frac{1}{n}\sum_{k=1}^{n} f(Z_k) \xrightarrow{n} \int f(z)\mu(dz)$$

Strong dependence in chain can lead to very slow convergence of ergodic averages.

- Often random variable $Z \sim \mu$ admits natural partition $Z = (X, Y)$. Gibbs sampler alternates between sampling $X|Y$ and $Y|X$.

Gibbs Sampler

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

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

2. Draw $\delta^{(k+1)} \sim \mathrm{Ga}(\alpha + \frac{N}{2}, \beta + \frac{1}{2}\|\mathcal{C}_0^{-\frac{1}{2}} u^{(k)}\|^2)$;

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

Key Scalings

$$\delta^{(k+1)} \sim \mathrm{Ga}(\alpha + \frac{N}{2}, \beta + \frac{1}{2}\|\mathcal{C}_0^{-\frac{1}{2}} u^{(k)}\|^2)$$

- $\mathrm{Ga}(\alpha + \frac{N}{2}, \beta + b^{-1}\frac{N}{2}) \approx \mathrm{Dirac(b)}$, for large N.
- Assumptions ensuring that prior is regularizing together with the fact that prior is strongly dependent on δ, secure $\|\mathcal{C}_0^{-\frac{1}{2}} u^{(k)}\|^2 \approx (\delta^{(k)})^{-1} N$
- Hence $\delta^{(k+1)} \approx \delta^{(k)}$, for large N.

Key Scalings

$$\delta^{(k+1)} \sim \mathrm{Ga}(\alpha + \frac{N}{2}, \beta + \frac{1}{2}\|\mathcal{C}_0^{-\frac{1}{2}} u^{(k)}\|^2)$$

- $\mathrm{Ga}(\alpha + \frac{N}{2}, \beta + b^{-1}\frac{N}{2}) \approx \mathrm{Dirac(b)}$, for large N.
- Assumptions ensuring that prior is regularizing together with the fact that prior is strongly dependent on δ, secure $\|\mathcal{C}_0^{-\frac{1}{2}} u^{(k)}\|^2 \approx (\delta^{(k)})^{-1}N$
- Hence $\delta^{(k+1)} \approx \delta^{(k)}$, for large N.
- Strong dependence in (u, δ)-chain as N increases.

Simulation Example

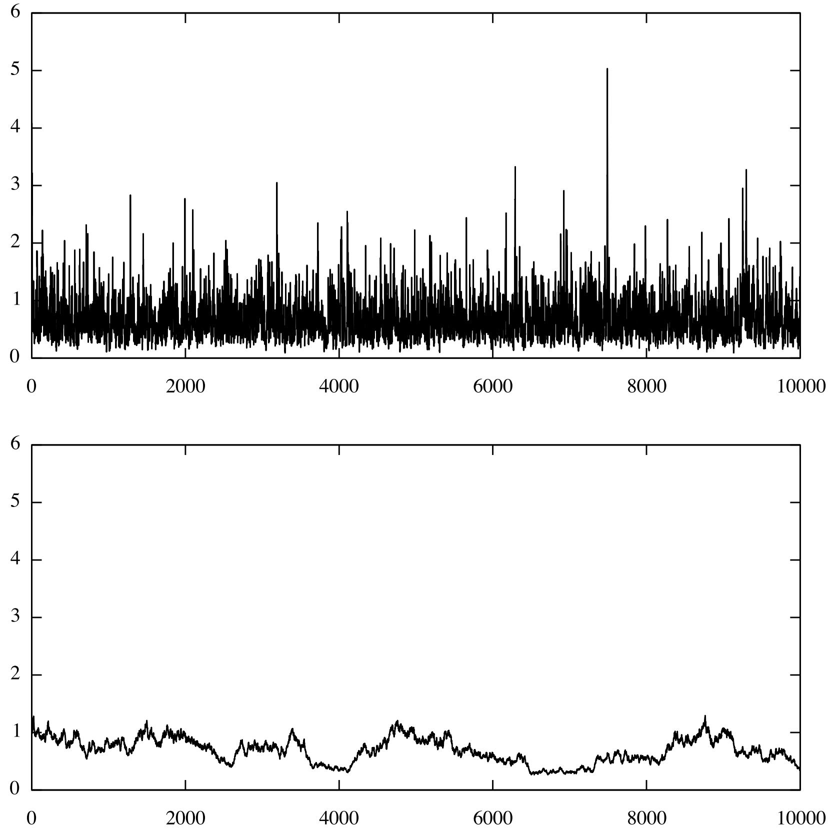


Figure: δ-samples for $N = 32$ (top) and $N = 8192$ (bottom)

Simulation Example

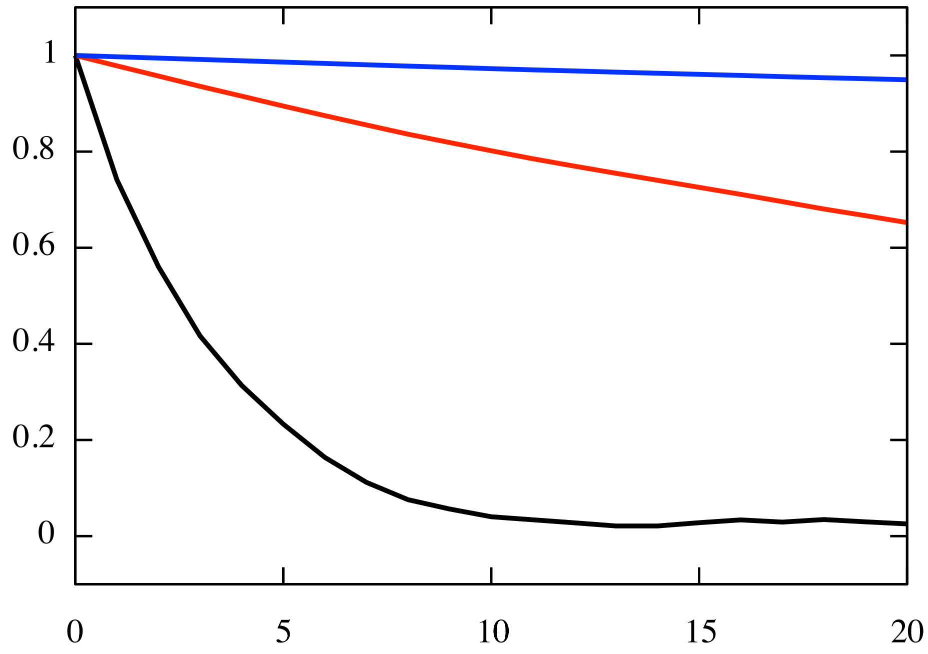


Figure: Autocorrelation functions for $N = 32$ (black), 512 (red) and 8192 (blue)

δ Evolves Slowly

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

In the limit $N \to \infty$, almost surely with respect to y

$$\delta^{(k+1)} - \delta^{(k)} \approx \frac{2}{N}\Big((\alpha + 1)\delta^{(k)} - f(\delta^{(k)}; y)(\delta^{(k)})^2\Big) + \sqrt{\frac{4\delta^2}{N}}\Xi,$$

where Ξ is a real random variable with mean zero and variance one.

Suggests diffusion limit to

$$d\delta = \big(\alpha + 1 - f(\delta; y)\delta\big)\delta dt + \sqrt{2}\delta dW$$

Partial Solution - Reparametrization

- New variables: $u = \delta^{-\frac{1}{2}}v$, where $v \sim N(0, \mathcal{C}_0)$, $\delta \sim \mathrm{Ga}(\alpha, \beta)$, $v \perp \delta$.
- Reparametrization of prior breaks dependence which slows down preceding algorithm.
- *Bayes'* theorem gives density of posterior on $\mathbb{R}^N \times \mathbb{R}$

$$p(u, \delta | y) \propto \delta^{\alpha - 1} \exp\Big(-\beta\delta - \frac{\lambda}{2} \| \mathcal{C}_1^{-\frac{1}{2}} (\delta^{-\frac{1}{2}} K v - y) \|^2 - \frac{1}{2} \| \mathcal{C}_0^{-\frac{1}{2}} v \|^2 \Big)$$

- $v | y, \delta$ Gaussian, but $\delta | y, v$ has more complicated distribution.
- Use Gibbs sampler alternating between updating $v | y, \delta$ and $\delta | y, v$; Metropolis step for updating $\delta | y, v$.

Simulation Example (Reparametrization)

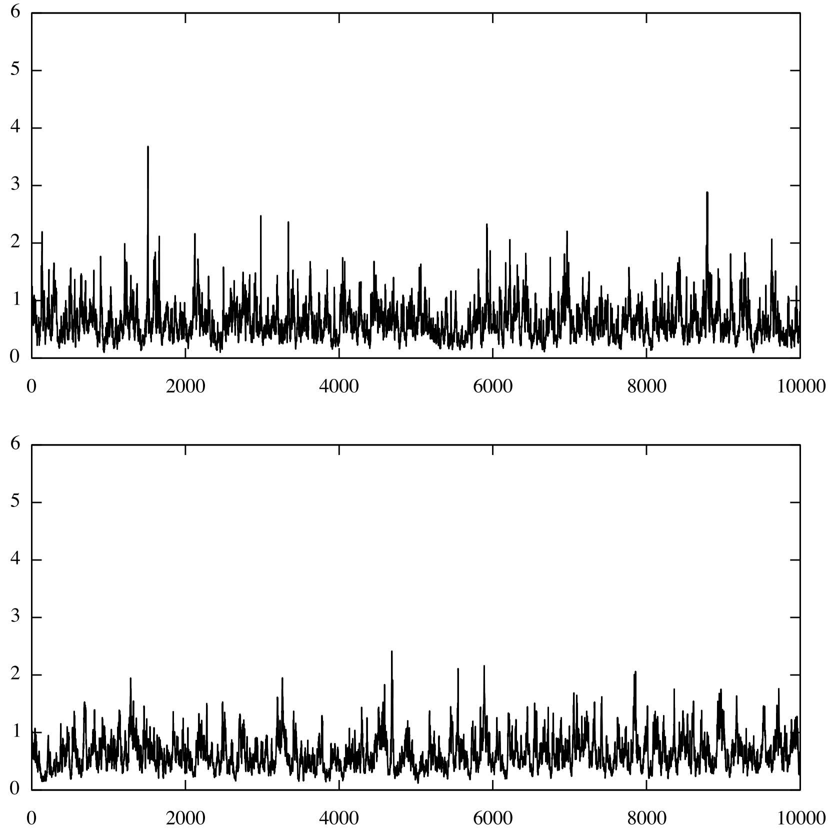


Figure: δ-samples for $N = 32$ (top) and $N = 8192$ (bottom)

Simulation Example (Reparametrization)

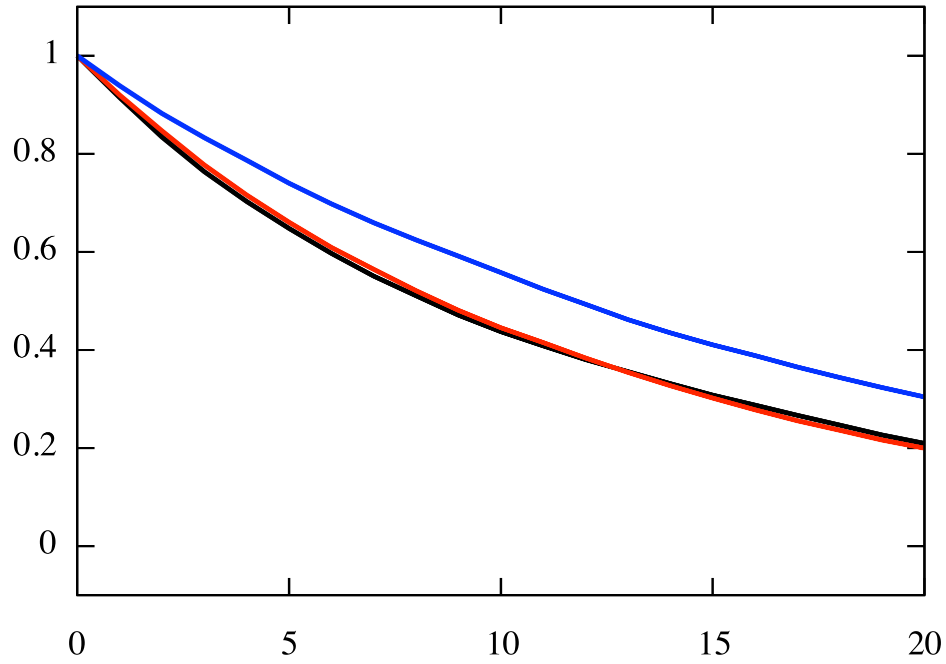
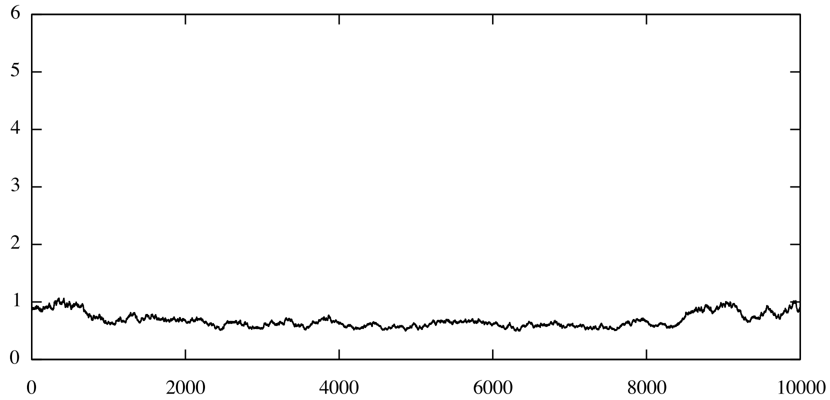


Figure: Autocorrelation functions for $N = 32$ (black), 512 (red) and 8192 (blue)

Why Partial?

- v and δ are a posteriori dependent due to their need to explain the data.
- When noise is small (λ large) then $\delta^{-\frac{1}{2}}Kv \approx y$ hence v and δ are strongly dependent.



Future/Ongoing Work

- Prove diffusion limit in the standard algorithm.
- Find parametrizations which work well in both the high dimensional and small noise limits.
- Intuition applies to other priors which are mixtures of Gaussians, e.g. when attempting to learn regularity of prior.
- Intuition applies to other nonlinear settings, e.g. nonparametric drift estimation of SDE's.

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

- S. Agapiou, J. Bardsley, O. Papaspiliopoulos and A. M. Stuart, *Analysis of the Gibbs sampler for hierarchical inverse problems*, http://arxiv.org/abs/1311.1138
- S. Agapiou, S. Larsson and A. M. Stuart, *Posterior contraction rates for the Bayesian approach to linear ill-posed inverse problems*, Stoch. Proc. Appl., 123(10):3828-3860, 2013
- 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
- Stuart, A. M., *Inverse problems: a Bayesian perspective*, Acta Numer. 19 (2010), 451-559.
- O. Papaspiliopoulos, G. O. Roberts, and M. Sköld, *A general framework for the parametrization of hierarchical models*, Statist. Sci., 22(1):59-73, 2007