A Bayesian Approach to Linear Inverse Problems

Sergios Agapiou

Mathematics Institute, University of Warwick, Coventry, UK

Joint work with Andrew Stuart

Young Researchers in Mathematics, 14-16 April 2011

THE UNIVERSITY OF WARWICK

Summary

- Aim: To provide a brief introduction to the Bayesian Approach to Inverse Problems.

- Inverse Problems
- The Deterministic Approach - Tikhonov Regularization
- The Bayesian Approach - the Gaussian prior and Gaussian noise case
  - Finite dimensions
  - Infinite dimensions - Separable Hilbert space setting
- Posterior Consistency

- We use the Laplacian Inverse Problem as a guiding example.

Inverse Problems

- Inverse Problems are concerned with determining causes for a desired or an observed effect.

- Let $K:\mathcal{X}\rightarrow\mathcal{Y}$ be a bounded linear operator, $\mathcal{X},\mathcal{Y}$ separable Hilbert spaces. For simplicity assume K is injective. Suppose
$$y = Ku, \tag{1}$$
where y is considered as the *data* and u as the *unknown parameter*.

- Want to invert (1), i.e. find u from possibly noisy observations of y, which we model as
$$y^{\delta} = Ku + \eta, \tag{2}$$
where η is some additive noise and where we assume that $||\eta||_{\mathcal{Y}} \leq \delta$ (deterministic approach) or that η has known statistics (Bayesian approach).

Inverse Problems are Ill-Posed - Deterministic Approach

- Problem (1)-(2) is in general ill-posed:
- The existence of a solution is not guaranteed:
  - noisy observations of y, may not live in $\mathcal{R}(K)$.
- Solutions do not depend continuously on the data:
  - important, since we want for small size of the noise to have a good approximation of u.

- Tikhonov Regularization: define the Tikhonov regularized approximate solution

$$u^{\lambda,\delta} := \underset{u\in\mathcal{X}}{\text{argmin}} \left(\frac{1}{2} ||y^{\delta} - Ku||^2_{\mathcal{Y}} + \frac{\lambda}{2} ||u||^2_{\mathcal{X}} \right)$$

where $\lambda > 0$ is called the *regularization parameter*.

Theorem

Let $||\eta||_{\mathcal{Y}} \leq \delta$. For $\lambda = \lambda(\delta)$ appropriately chosen $u^{\lambda,\delta} \rightarrow u$ in $\mathcal{X}$, as $\delta \rightarrow 0$.

Example 1: The Inverse Laplacian Problem

- Suppose $K = (-\boldsymbol{\Delta})^{-1}$ where $\boldsymbol{\Delta} : H^2(0,1) \cap H_0^1(0,1) \to L^2(0,1)$ is the Dirichlet-Laplacian. Let $\{ck^2, \phi_k\}_{k\in\mathbb{N}}$ be a complete orthonormal eigensystem and define the *"Sobolev-like"* spaces
$$\mathcal{H}^s := \left\{ u \in L^2(0,1) : \quad ||u||_s := \sum_{k=1}^{\infty} k^{2s}\langle u, \phi_k\rangle^2 < \infty \right\}, \quad s \geq 0.$$

- Since $(-\boldsymbol{\Delta})^{-1}$ is injective, we can invert iff $y^\delta \in \mathcal{R}((-\boldsymbol{\Delta})^{-1}) = \mathcal{H}^2$ in which case
$$u^\delta := -\boldsymbol{\Delta} y^\delta = \sum_{k=1}^{\infty} k^2 \langle y^\delta, \phi_k\rangle \phi_k.$$
Since $k^2 \to \infty$ the instability in the data is apparent.

- The minimizer of the Tikhonov functional is
$$u^{\lambda,\delta} = \sum_{k=1}^{\infty} \frac{k^{-2}}{k^{-4}+\lambda} \langle y^\delta, \phi_k\rangle \phi_k,$$
i.e. we filter out the components corresponding to large k's and as $\lambda \to 0$ we recover u^δ.

Generalized Tikhonov Regularization

- We can do even better, by using different norms to penalize attributes which we know that the unknown does not possess:

$$u^{\lambda,\delta} := \underset{u\in\mathcal{E}}{\text{argmin}} \left(\frac{1}{2}||y^{\delta} - Ku||_{\mathcal{Z}}^2 + \frac{\lambda}{2}||u||_{\mathcal{E}}^2 \right),$$

where $\mathcal{E}$ and $\mathcal{Z}$ are compactly embedded in $\mathcal{X}$ and $\mathcal{Y}$ respectively.

- The standard practice is to assume some *a-priori* known information on the regularity of the solution and a norm bound of the noise and determine convergence rates

$$||u^{\lambda,\delta} - u|| = O(f(\delta)).$$

Theorem

In Example 1, assume $\mathcal{E} = \mathcal{H}^{\alpha}, \mathcal{Z} = \mathcal{H}$ and $u \in \mathcal{H}^{\gamma}$, $||\eta|| \leq \delta$, where $\alpha \geq 0, \gamma > 0$. Then for $\lambda = \lambda(\delta)$ appropriately chosen

$$||u^{\lambda,\delta} - u|| = O(\delta^{\frac{\gamma}{2+\gamma}}), \quad \text{for} \quad \gamma \leq 4 + 2\alpha,$$

$$||u^{\lambda,\delta} - u|| = O(\delta^{\frac{4+2\alpha}{6+2\alpha}}), \quad \text{for} \quad \gamma > 4 + 2\alpha \quad \text{(saturation rate)}.$$

Disadvantages of the Deterministic Theory

- The Deterministic approach can be criticized:
- convergence results depend on a norm bound of the noise which is a worst-case scenario;
- we often have more available information than a norm bound of the noise, or the space where the solution lives in;
- we may have information on the statistics of the noise and of the solution.

A change of perspective: the Bayesian Approach

Consider the additive noise model for the inverse problem (1)

$$y = Ku + \eta. \tag{3}$$

- Main innovation: we express all the quantities in the model as *random variables*. We always assume $u \perp \eta$;
- express prior beliefs about the solution in the form of the *prior* distribution, μ_0, which is the distribution of u;
- express knowledge of the noise in the form of the noise distribution, P. By equation (3) we obtain the distribution of $y|u$, P^u, called the *data likelihood*;
- the solution is the distribution of $u|y$, called the *posterior* distribution, μ^y.

Example 2: Finite Dimensional Linear Case

- Suppose we have the inverse problem,

$$y = Ku, \quad u, y \in \mathbb{R}^n, K \in \mathbb{R}^{n\times n}$$

and consider the model

$$y = Ku + \eta, \quad \eta \in \mathbb{R}^n.$$

- Assume π_0 is the p.d.f. of the prior μ_0 and π^y is the p.d.f. of the posterior μ^y.
- Let the noise η be a random variable with density ρ. Then the data likelihood has density

$$\rho(y|u) = \rho(y - Ku).$$

- By the *Bayes* formula, we have that

$$\pi^y(u) \propto \rho(y|u)\pi_0(u) = \rho(y - Ku)\pi_0(u).$$

- In general the calculation of the posterior is difficult. In the Gaussian prior, Gaussian noise case things are simpler.

Example 2: The Gaussian Case in finite dimensions

- Suppose $\eta \sim \mathcal{N}(0, B)$ and $\mu_0 = \mathcal{N}(0, \Sigma_0)$, where $\mathcal{B}, \Sigma_0$ are $n \times n$ positive definite matrices. We have

$$\pi_0(u) \propto \exp\left(-\frac{1}{2}\left|\Sigma_0^{-\frac{1}{2}}u\right|^2\right), \qquad \rho(y-Ku) \propto \exp\left(-\frac{1}{2}\left|B^{-\frac{1}{2}}(y-Ku)\right|^2\right)$$

  thus

$$\pi^y(u) \propto \exp\left(-\frac{1}{2}\left|B^{-\frac{1}{2}}(y-Ku)\right|^2 - \frac{1}{2}\left|\Sigma_0^{-\frac{1}{2}}u\right|^2\right) := exp(-\Phi(u;y)),$$

  where Φ is a Tikhonov functional for $(\mathcal{Z}, ||.||_{\mathcal{Z}}) = (\mathbb{R}^n, |B^{-\frac{1}{2}}.|)$, $(\mathcal{E}, ||.||_{\mathcal{E}}) = (\mathbb{R}^n, |\Sigma_0^{-\frac{1}{2}}.|)$.

- Since $\Phi(u;y)$ is quadratic in u, μ^y is also Gaussian $\mathcal{N}(m, \Sigma)$, where m and Σ can be calculated by completing the square:

$$m = \Sigma_0 K^*(B + K\Sigma_0 K^*)^{-1} y = \operatorname*{argmin}_{u \in \mathcal{E}} \Phi(u, y), \qquad \text{(random variable)}$$

$$\Sigma = \Sigma_0 - \Sigma_0 K^*(B + K\Sigma_0 K^*)^{-1} K\Sigma_0. \qquad \text{(deterministic)}$$

The Gaussian Case in Infinite Dimensions

- This generalizes to the infinite dimensional setting:

Theorem - Setup 1: An Infinite dimensional Gaussian prior - Gaussian noise Bayesian setup

In (3), assume $u \sim \mathcal{N}(0, \tau^2 C_0)$ and $\eta \sim \mathcal{N}(0, \delta^2 C_1)$ where C_0, C_1 are positive definite trace class operators. Then $\mu^y = \mathcal{N}(m^\delta, C^\delta)$ where

$$m^\delta = \tau^2 C_0 K^* \left(\delta^2 C_1 + \tau^2 K C_0 K^*\right)^{-1} y := L^\delta y$$

and

$$C^\delta = \tau^2 C_0 - \tau^2 L K C_0.$$

The posterior mean is (formally) the minimizer of a generalized Tikhonov functional

$$m^\delta = \operatorname*{argmin}_{u \in \mathcal{E}} \left(\frac{1}{2} \left|\left| y^\delta - Ku \right|\right|_{\mathcal{Z}}^2 + \frac{\delta^2}{2\tau^2} \left|\left| u \right|\right|_{\mathcal{E}}^2 \right),$$

where $\mathcal{E} = \mathcal{R}(C_0^{\frac{1}{2}})$ with norm $||.||_{\mathcal{E}} = \left|\left| C_0^{-\frac{1}{2}} . \right|\right|$ and $\mathcal{Z} = \mathcal{R}(C_1^{\frac{1}{2}})$ with norm $||.||_{\mathcal{Z}} = \left|\left| C_1^{-\frac{1}{2}} . \right|\right|$.

Posterior Consistency - Posterior Contraction

- In Setup 1, consider observations of the form

$$y_0^\delta = Ku_0 + \eta^\delta$$

where $u_0 \in \mathcal{X}$ is a fixed *true solution* and $\eta^\delta \sim \mathcal{N}(0, \delta^2 C_1)$.

- This choice of data model gives as posterior the measure $\mu^{y_0^\delta} = \mathcal{N}(m_0^\delta, C^\delta)$, where $m_0^\delta = L^\delta y_0^\delta$.

- Reasonable to examine whether as the noise vanishes, the posterior distribution converges in some sense to the true solution.

- Define the square posterior contraction

$$SPC := \mathbb{E}\left|\left|m_0^\delta - u_0\right|\right|^2 + tr(C^\delta).$$

The first term measures how far is the posterior mean from the true solution, while the second term measures how concentrated the posterior distribution is.

Example 1 revisited: Posterior Consistency

- Consider the Setup 1 in Example 1, for $C_0 = (-\mathbf{\Delta})^{-\frac{1}{2}-\alpha}$ and $C_1 = (-\mathbf{\Delta})^{-\beta}$, where $\beta \geq \frac{1}{2}, \alpha \geq 0$. Then $\mu^y = \mathcal{N}(m^\delta, C^\delta)$ where $m^\delta = L^\delta y$.

- Formally, the posterior mean is

$$m^\delta = \underset{u \in \mathcal{E}}{\text{argmin}} \left(\frac{1}{2} ||y^\delta - Ku||^2_{\mathcal{Z}} + \frac{\delta^2}{2\tau^2} ||u||^2_{\mathcal{E}} \right),$$

where $\mathcal{E} = \mathcal{R}(C_0^{\frac{1}{2}}) = \mathcal{H}^{\frac{1}{2}+\alpha}$ and $\mathcal{Z} = \mathcal{R}(C_1^{\frac{1}{2}}) = \mathcal{H}^\beta$ with norms the corresponding "Sobolev-like" norms.

- Suppose we have observations of the form

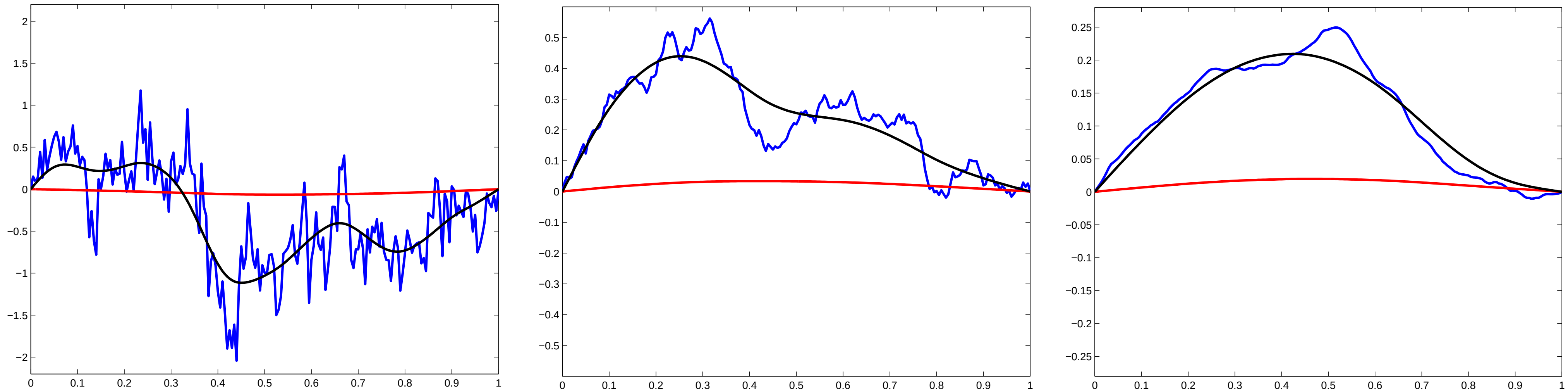
$$y_0^\delta = (-\mathbf{\Delta})^{-1} u_0 + \eta^\delta, \quad \text{where} \quad \eta^\delta \sim \mathcal{N}(0, \delta^2 C_1).$$

Then the posterior distribution is $\mathcal{N}(m_0^\delta, C^\delta)$ where

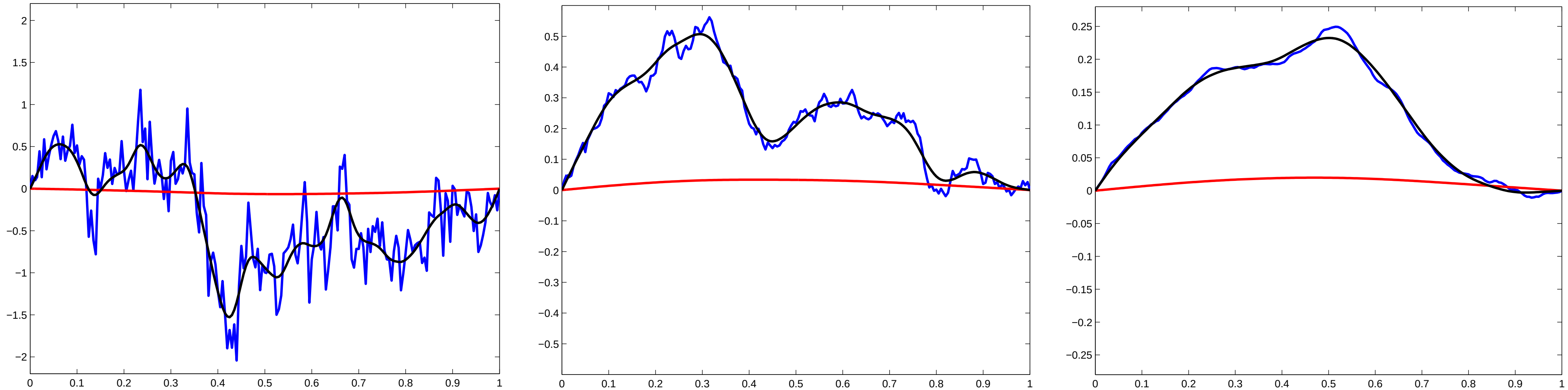
$$m_0^\delta = L^\delta \left((-\mathbf{\Delta})^{-1} u_0 + \eta^\delta \right).$$

- As before we assume to have the *a-priori* information that $u_0 \in \mathcal{H}^\gamma$, $\gamma > 0$.

Example 1: Posterior Consistency



(a) $\gamma = \frac{1}{4}$, $\delta = 10^{-3}$, $||m_0^\delta - u_0|| \approx 0.85$ (b) $\gamma = \frac{3}{4}$, $\delta = 10^{-3}$, $||m_0^\delta - u_0|| \approx 0.15$ (c) $\gamma = \frac{5}{4}$, $\delta = 10^{-3}$, $||m_0^\delta - u_0|| \approx 0.050$



(d) $\gamma = \frac{1}{4}$, $\delta = 10^{-4}$, $||m_0^\delta - u_0|| \approx 0.65$ (e) $\gamma = \frac{3}{4}$, $\delta = 10^{-4}$, $||m_0^\delta - u_0|| \approx 0.066$ (f) $\gamma = \frac{5}{4}$, $\delta = 10^{-4}$, $||m_0^\delta - u_0|| \approx 0.016$

Figure: Realizations of the true solution (blue), the posterior mean (black) and the observation (red) for $\alpha = \frac{1}{4}$, $\beta = \frac{3}{5}$, for three values of γ and for $\delta = 10^{-3}$ (1st line) and $\delta = 10^{-4}$ (2nd line). We have a better reconstruction when the truth is more regular.

Example 1 revisited: Posterior Contraction

- Define $R = \frac{5}{4} + \frac{\alpha}{2} - \frac{\beta}{2}$, and assume that $R > 0$ (the prior is indeed regularizing).
- We study optimal rates of contraction for the different values of $\gamma > 0$:

Theorem

i) If the noise and the true solution are not too smooth, i.e. if $\beta \leq \frac{5}{2}$ and $\gamma < 4R$, for $\tau = \tau(\delta)$ chosen appropriately, we have

$$SPC \asymp \delta^{\frac{2\gamma}{2R+\gamma-\alpha}}.$$

If the true solution is too smooth, $\gamma \geq 4R$ then the rate does not improve as γ increases, i.e. the rate saturates at

$$\delta^{\frac{8R}{6R-\alpha}}.$$

ii) If the noise is too smooth, $\beta > \frac{5}{2}$, then the rate of contraction is δ^2.

- Conclusion: We have posterior consistency in Example 1, for $\gamma, R > 0$.

Example 1: Posterior Contraction

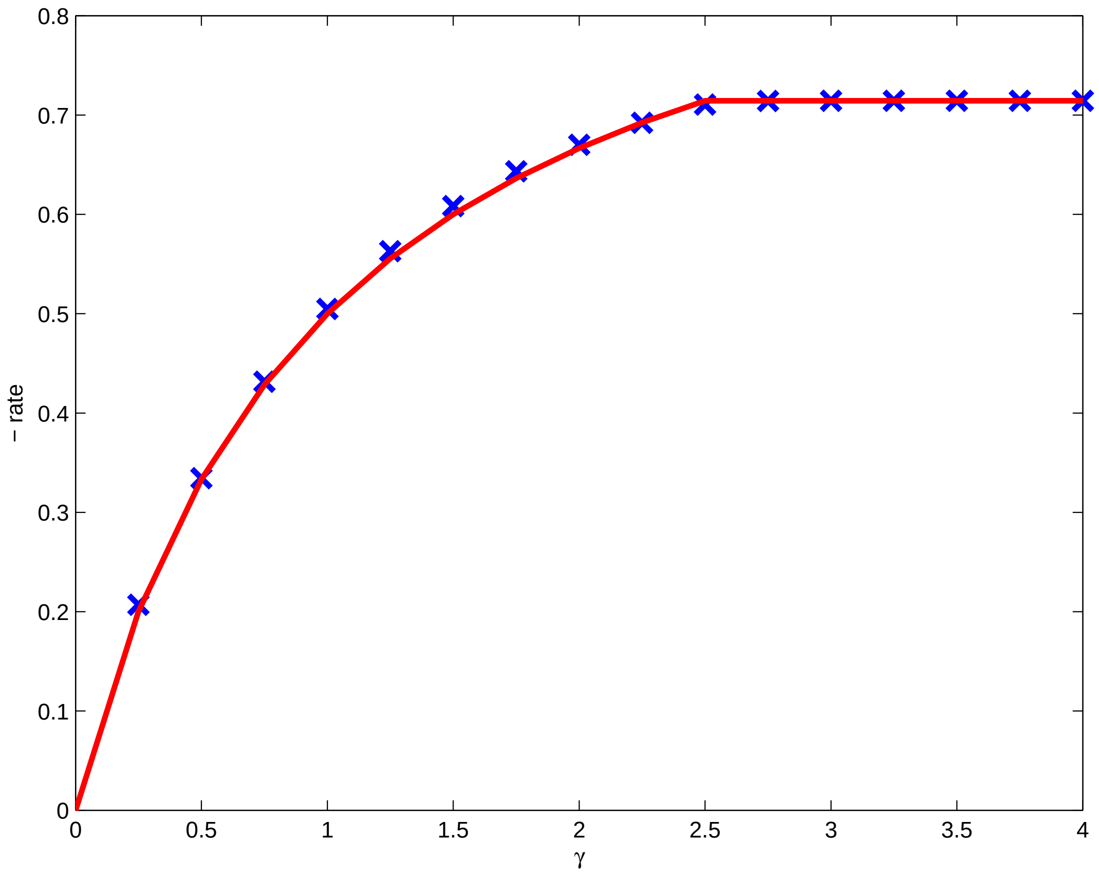


Figure: Rates of contraction plotted against γ. In red we have the rates predicted by the theory and in blue we have estimated rates using the average of 160 realizations.

References - Further Reading

- Agapiou, S., Larsson, S. and Stuart, A. M., *A PDE method for examining the Posterior Consistency of the Bayesian Approach to Linear Ill-Posed Inverse Problems*, work in progress.
- Diaconis, Persi and Freedman, David, *On the consistency of Bayes Estimates*, Ann. Statist. 14 (1986), no. 1, 1-67.
- 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.