Recovery of the initial condition for the heat equation
Bayesian inverse problems

PEDS–II Workshop – Eurandom, June 4–6, 2012

Bartek Knapik

Joint work with Aad van der Vaart and Harry van Zanten

4 June 2012

Differential equations as inverse problems

Solutions to PDEs depend on the initial condition

$$u = K\mu,$$

where u is the solution, K is the solution operator, and μ is the initial condition

Such problems are often ill-posed

The solution operator of the differential equation does not have a well-behaved, continuous inverse, e.g., K is a compact operator

Statistical inverse problems

We observe the solution perturbed by the noise or measurement errors

$$Y = K\mu + \varepsilon Z$$

where Y is the observed solution, Z is the white noise, ε is the level of the noise

Estimator for μ

Because of the noise Z, Y is not in the range of K, and therefore $K^{-1}Y$ is not an answer.

Regularization

Find an operator A “close to” K^{-1} such that AY is well defined and then estimate μ by

$$\hat{\mu} = AY$$

Bayesian approach

Bayesian paradigm

- We choose a distribution for $\mu \longrightarrow$ prior
- Data Y have distribution given $\mu \longrightarrow Y|\mu$
- Bayes theorem gives distribution of μ given $Y \longrightarrow \mu|Y$ posterior

Bayesian answer – Posterior distribution

- Point estimator – posterior mean
- Confidence sets – credible sets

Our results

Our questions

- Does the posterior distribution center at μ?
- Are some priors better than others?
- Does the posterior correctly describe the order of the error?

Our answers

- Yes, we get consistent estimators.
- Performance depends on combinations of the characteristics of the prior, the true parameter, and the known transformation. It can be very good and very bad.

Recovery of the initial condition for the heat equation

The Dirichlet problem for the heat equation

$$\frac{\partial}{\partial t}u(x,t) = \frac{\partial^2}{\partial x^2}u(x,t), \quad u(x,0) = \mu(x), \quad u(0,t) = u(1,t) = 0,$$

where u is defined on $[0,1] \times [0,T]$ and the function $\mu \in L_2[0,1]$ satisfies $\mu(0) = \mu(1) = 0$

The solution is given by

$$u(x,T) = \sqrt{2}\sum_{i=1}^{\infty} \mu_i e^{-i^2\pi^2 T} \sin(i\pi x) = \sum_{i=1}^{\infty} \kappa_i \mu_i e_i(x),$$

Inverse problem

Recall

$$u = K\mu$$

Here

$$u(x,T) = \sum_{i=1}^{\infty} \kappa_i \mu_i e_i(x),$$

where $e_i(x) = \sqrt{2}\sin(i\pi x)$ and $\kappa_i = e^{-i^2\pi T}$ form the eigensystem of the solution operator K, and μ_i are the coordinates of μ in the eigenbasis of the operator K

Recovery of the initial condition for the heat equation

Sequence of the noisy, transformed Fourier coefficients of μ

$$Y_i = \kappa_i \mu_i + \frac{1}{\sqrt{n}} Z_i, \qquad i = 1, 2, \dots,$$

where $Z_1, Z_2, \dots$ are independent, standard normal random variables

Product prior on μ of the form

$$\Pi = \bigotimes_{i=1}^{\infty} N(0, \lambda_i)$$

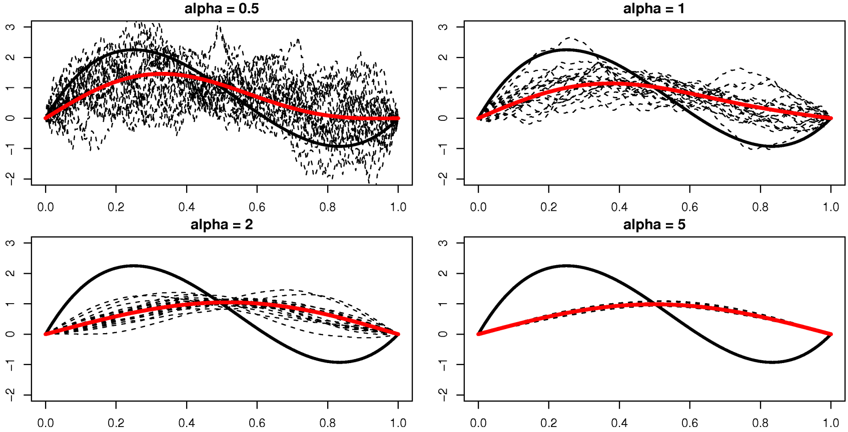
with two types of prior

$$\lambda_i^{(1)} = i^{-1-2\alpha}, \qquad \lambda_i^{(2)} = \exp(-\alpha i^2), \quad \text{for some } \alpha > 0.$$

Coordinates of the true μ_0 satisfy for some $\beta > 0$

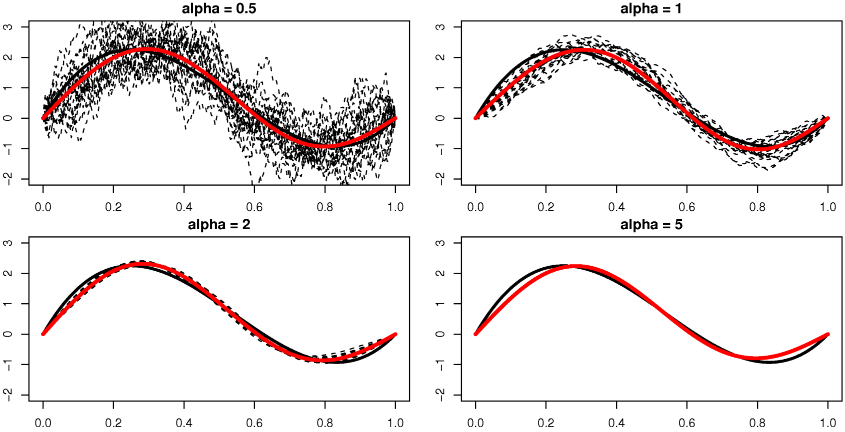
$$\sum_{i=1}^{\infty} \mu_{0,i}^2 i^{2\beta} < \infty$$

Example – $\lambda_i^{(1)} = i^{-1-2\alpha}$, $n = 10^4$



Black curve – true μ_0, red curve – posterior mean, dashed curves – 20 draws from the posterior

Example – $\lambda_i^{(1)} = i^{-1-2\alpha}$, $n = 10^8$



Black curve – true μ_0, red curve – posterior mean, dashed curves – 20 draws from the posterior

Rate of contraction – polynomial prior

Theorem (K, van der Vaart, van Zanten (2011))

If $\lambda_i = \tau_n^2 i^{-1-2\alpha}$ *for some* $\alpha > 0$ *and* $\tau_n > 0$ *such that* $n\tau_n^2 \to \infty$, *then for every* $M_n \to \infty$

$$\mathrm{E}_{\mu_0}\Pi_n(\mu : \|\mu - \mu_0\| \geq M_n\varepsilon_n \,|\, Y) \to 0,$$

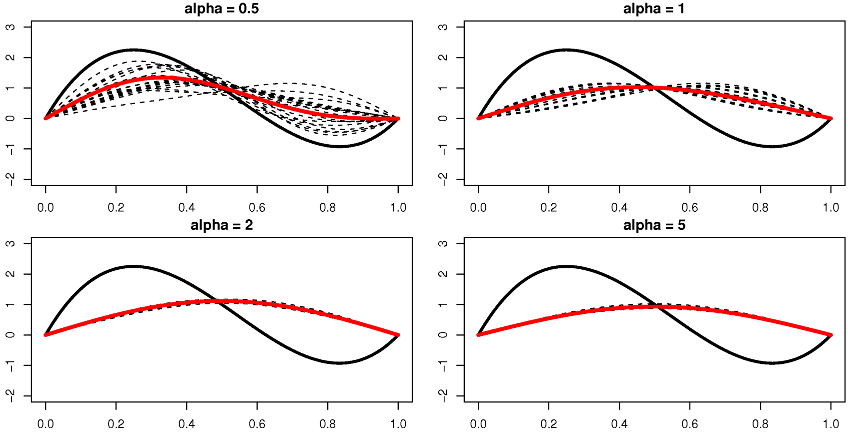
where

$$\varepsilon_n = \big(\log(n\tau_n^2)\big)^{-\beta/2} + \tau_n\big(\log(n\tau_n^2)\big)^{-\alpha/2}.$$

The rate is optimal when

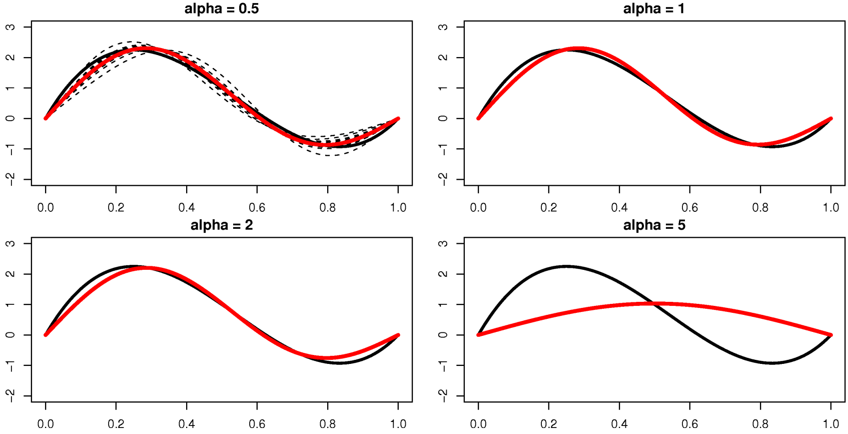
- $\tau_n \equiv 1$ and $\alpha \geq \beta$
- $n^{-1/2+\delta} \lesssim \tau_n \lesssim (\log n)^{(\alpha-\beta)/2}$, for some $\delta > 0$

Example – $\lambda_i^{(2)} = \exp(-\alpha i^2)$, $n = 10^4$



Black curve – true μ_0, red curve – posterior mean, dashed curves – 20 draws from the posterior

Example – $\lambda_i^{(2)} = \exp(-\alpha i^2)$, $n = 10^8$



Black curve – true μ_0, red curve – posterior mean, dashed curves – 20 draws from the posterior

Rate of contraction – exponential prior

Theorem (K, van der Vaart, van Zanten (2011))

If $\lambda_i = \exp(-\alpha i^2)$ *for some* $\alpha > 0$, *then for every* $M_n \to \infty$

$$\mathrm{E}_{\mu_0}\Pi_n(\mu : \|\mu - \mu_0\| \geq M_n\varepsilon_n \,|\, Y) \to 0,$$

where

$$\varepsilon_n = (\log n)^{-\beta/2}.$$

The rate is always optimal

Credible sets

Credible ball

A credible ball centered at the posterior mean $\hat{\mu}$ takes the form

$$\hat{\mu} + B(r_{n,\gamma}) = \{\mu \in \ell_2 : \|\mu - \mu_0\| < r_{n,\gamma}\}$$

where the radius $r_{n,\gamma}$ is determined such that

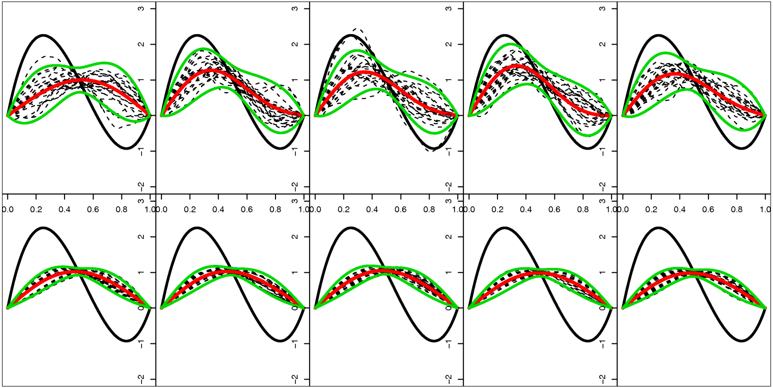
$$\Pi_n(\hat{\mu} + B(r_{n,\gamma}) \,|\, Y) = 1 - \gamma$$

Credible sets as confidence sets

The frequentist coverage of the credible set is

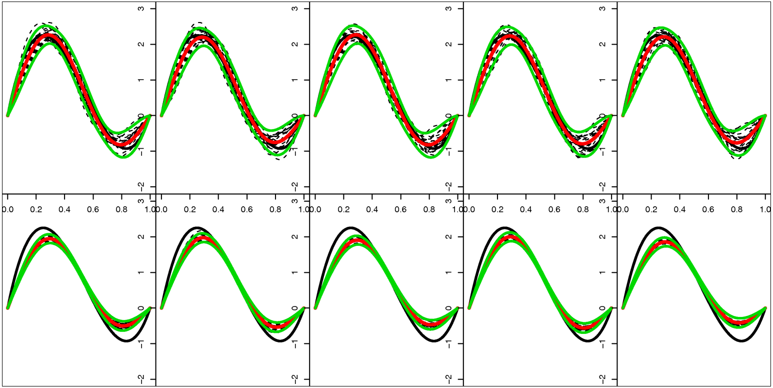
$$\mathrm{P}_{\mu_0}\big(\mu_0 \in \hat{\mu} + B(r_{n,\gamma})\big)$$

Example – $\lambda_i^{(1)} = i^{-1-2\alpha}$, $n = 10^4$



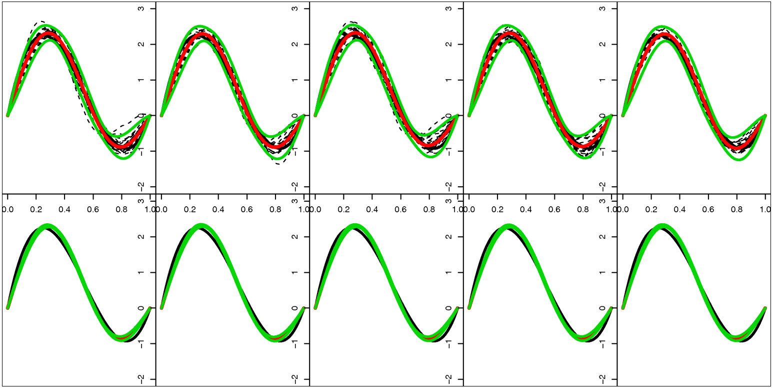
Top panels $\alpha = 1.5$, bottom panels $\alpha = 3$

Example – $\lambda_i^{(1)} = i^{-1-2\alpha}$, $n = 10^6$



Top panels $\alpha = 1.5$, bottom panels $\alpha = 3$

Example – $\lambda_i^{(1)} = i^{-1-2\alpha}$, $n = 10^8$



Top panels $\alpha = 1.5$, bottom panels $\alpha = 3$

Credibility – polynomial prior

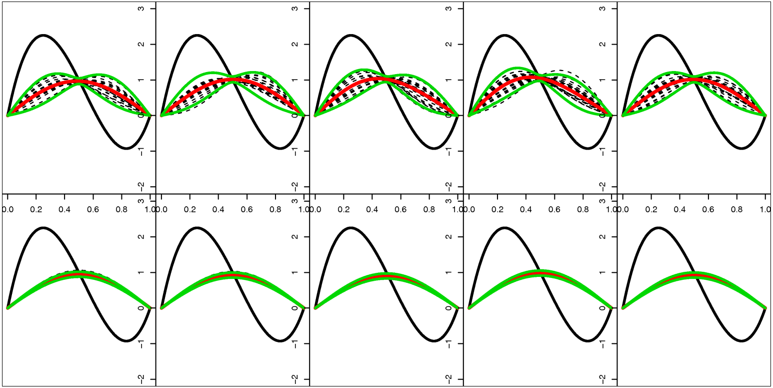
Theorem (K, van der Vaart, van Zanten (2011))

If $\lambda_i = i^{-1-2\alpha}$ for some $\alpha > 0$, then asymptotic coverage of the credible ball centered at the posterior mean is

- 1, *when* $\alpha < \beta$*; here* $r_{n,\gamma}/\tilde{r}_{n,\gamma} \to \infty$,
- 0, *when* $\alpha > \beta$,
- 1 *or* 0, *depending on the norm of* μ_0, *when* $\alpha = \beta$.

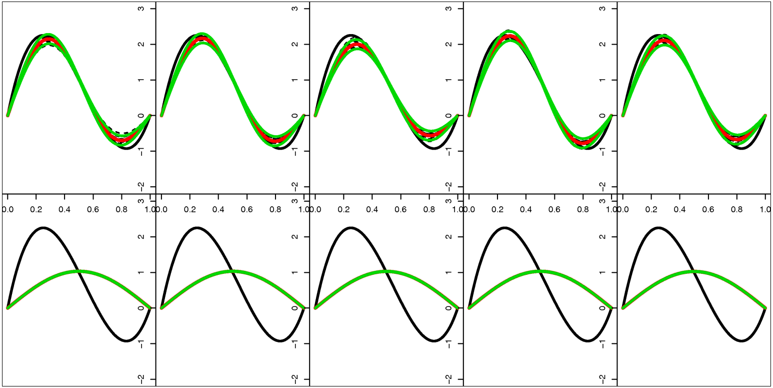
Undersmoothing leads to very conservative sets, oversmoothing has disastrous consequences for the coverage.

Example – $\lambda_i^{(2)} = \exp(-\alpha i^2)$, $n = 10^4$



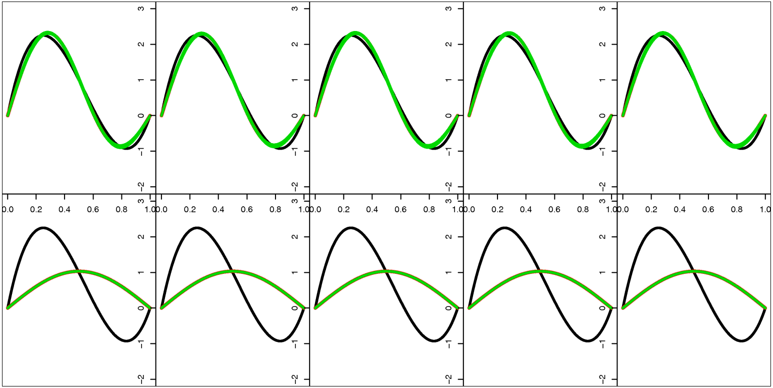
Top panels $\alpha = 1$, bottom panels $\alpha = 5$

Example – $\lambda_i^{(2)} = \exp(-\alpha i^2)$, $n = 10^6$



Top panels $\alpha = 1$, bottom panels $\alpha = 5$

Example – $\lambda_i^{(2)} = \exp(-\alpha i^2)$, $n = 10^8$



Top panels $\alpha = 1$, bottom panels $\alpha = 5$

Credibility – exponential prior

Theorem (K, van der Vaart, van Zanten (2011))

If $\lambda_i = \exp(-\alpha i^2)$ *for some* $\alpha > 0$*, then asymptotic coverage of the credible ball centered at the posterior mean is* 0*, for every* μ_0 *such that* $|\mu_{0,i}| \gtrsim \exp(-ci^2)$ *for some* $c < \alpha/2$*.*

Very bad confidence sets for a wide range of μ_0

Concluding remarks

- Inverse problems can be solved using Bayesian procedures
- Many priors yield optimal rate of recovery
- However, one should be (very) careful with credible sets

Thank you!