Gauss versus Laplace rates of contraction under Besov regularity

Sergios Agapiou

Department of Mathematics and Statistics, University of Cyprus

December 2020, International Bayes Club

University of Cyprus

Joint work with

- S. Agapiou, M. Dashti and T. Helin, *Rates of contraction of posterior distributions based on p-exponential priors*, arXiv:1811.12244 (to appear in Bernoulli).

- S. Agapiou and A. Savva, *Adaptive rates of contraction based on p-exponential priors*, in preparation.

- S. Agapiou and S. Wang, *Frequentist rates of contraction in Bayesian inverse problems with Laplace priors*, in preparation.

Outline

Outline

Multiscale features in images

Wavelet expansions

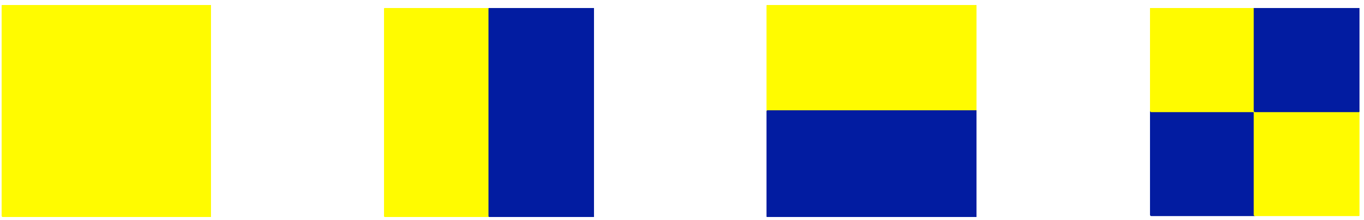
- $\{\psi_\ell\}_{\ell=1}^{\infty}$ orthonormal basis for $L_2(\mathbb{T}^m)$

$$u(x) = \sum_{\ell=1}^{\infty} u_\ell \psi_\ell(x), \quad u_\ell = \langle u, \psi_\ell \rangle.$$

- e.g. $\{\psi_\ell\}$ is the Fourier basis

- For functions with multiscale features, better use wavelet bases $\{\psi_{kl}\}$

$$u(x) = \sum_{k=1}^{\infty} \sum_{l=1}^{2^k} u_{kl} \psi_{kl}(x), \quad u_{kl} = \langle u, \psi_{kl} \rangle.$$

e.g. 2D Haar 

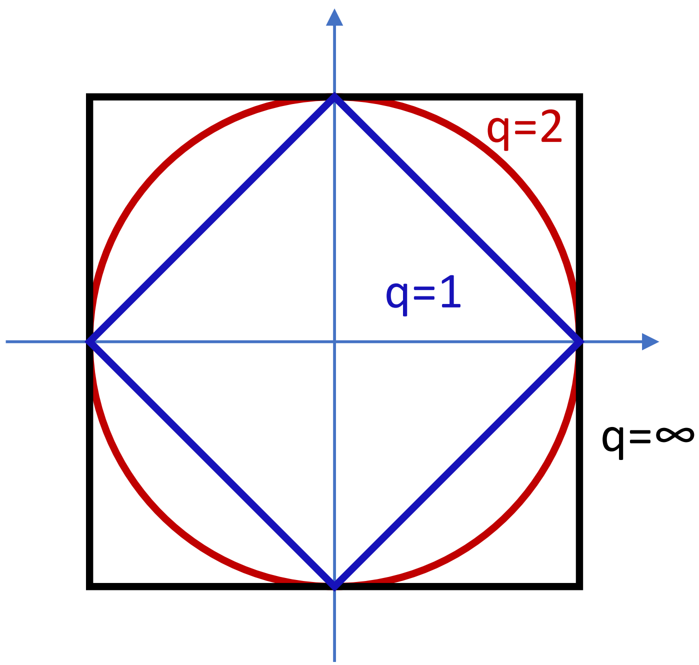
Besov Spaces

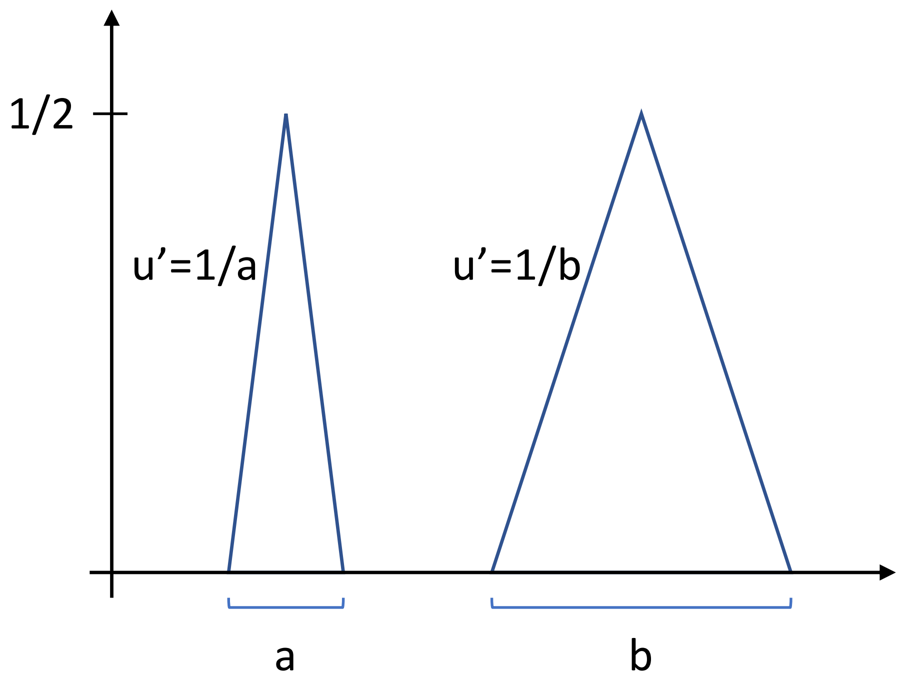
- Functions identified with expansion coefficients $(u_\ell) \in \ell_2$ or $(u_{kl}) \in \ell_2$
- Besov space of smoothness $s \in \mathbb{R}$, with integrability parameter $q \geq 1$

$$B^s_{qq} = \left\{ u \in \mathbb{R}^\infty : \sum_{\ell=1}^{\infty} \ell^{q(\frac{s}{d}+\frac{1}{2})-1} |u_\ell|^q < \infty \right\}, \quad \|u\|_{B^s_{qq}} = \left(\sum_{\ell=1}^{\infty} \ell^{q(\frac{s}{d}+\frac{1}{2})-1} |u_\ell|^q \right)^{\frac{1}{q}}.$$

- $q = 2$: $B^s_{22} = H^s$, Sobolev Hilbert spaces
- $q = \infty$, $s \notin \mathbb{N}$: $B^s_{\infty\infty} = C^s$, Hölder spaces
- Smaller q associated with sparsity and spatial inhomogeneity

Besov Spaces - Intuition





$$\|u'\|_{L_1} = 2, \ \|u'\|_{L_2} = \sqrt{\frac{1}{a} + \frac{1}{b}}, \ \|u'\|_{L_\infty} = \frac{1}{a}$$

I.M. Johnstone, *Gaussian estimation: sequence and wavelet models*, draft book.

Function-space priors via random series expansions

$$u(x) = \sum_{\ell=1}^{\infty} u_\ell \psi_\ell(x)$$

- Randomize coefficients: $u_\ell = \gamma_\ell \xi_\ell$ where $\xi_\ell \overset{iid}{\sim} f$, $\gamma_\ell > 0$ decaying scalings
- Choice of wavelet basis, distribution f, decay scaling
- eg if f has finite second moments, then $u \in L_2$ almost surely iff $(\gamma_\ell) \in \ell_2$
- B^s_{11}-Besov priors: $\xi_\ell \overset{iid}{\sim} Laplace(0,1)$ and $\gamma_\ell = \ell^{\frac{1}{2}-\frac{s}{d}}$, s smoothness parameter

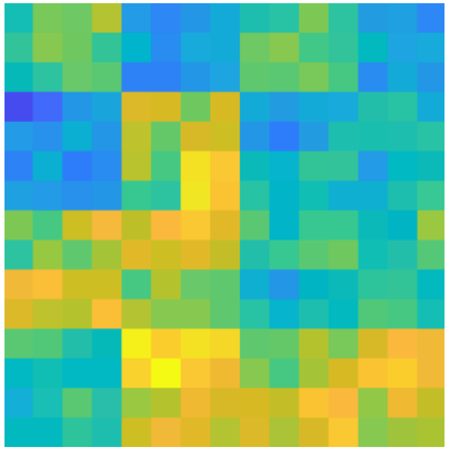
$$\text{"}\pi(u) \propto \exp(-\|u\|_{B^s_{11}})\text{"}$$

M. Lassas, E. Saksman and S. Siltanen, *Discretization-invariant Bayesian inversion and Besov space priors, Inverse Problems and Imaging* 2009

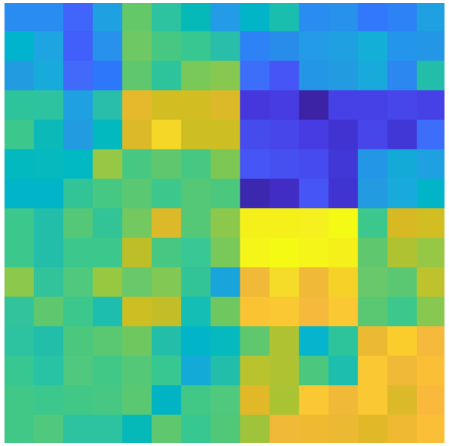
V. Kolehmainen, M. Lassas, K. Niinimaki and S. Siltanen, *Sparsity-promoting Bayesian inversion,* Inverse Prob 2012

M. Dashti, S. Harris and A. Stuart, *Besov priors for Bayesian inverse problems,* Inverse Problems and Imaging, 2013

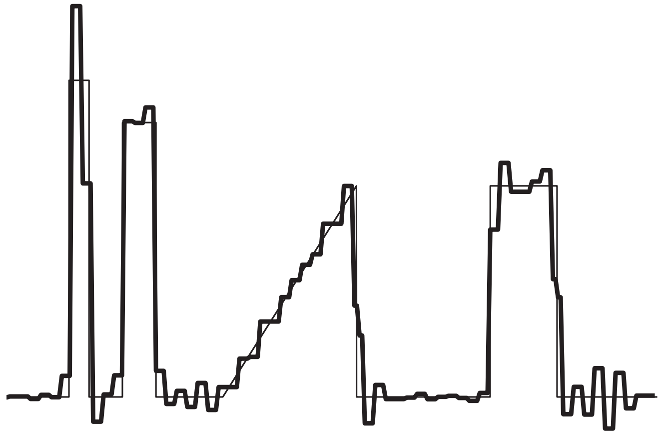
Priors via random series expansions - Haar wavelets



Gaussian



Laplace (B^s_{11})



Kolehmainen et al. 2012

- S. Agapiou, M. Burger, M. Dashti and T. Helin, *Sparsity-promoting and edge-preserving MAP estimators in nonparametric Bayesian inverse problems*, Inverse Problems, 2018.
- S. Agapiou, M. Dashti and T. Helin, *Rates of contraction of posterior distributions based on p-exponential priors*, to appear in Bernoulli.

Outline

White noise model - Minimax estimation rates

- Observe solution to

$$dY_t^{(n)} = u(t)dt + \frac{1}{\sqrt{n}}dW_t, \quad t \in [0,1]$$

$Y_0^{(n)} = 0$, W_t is a sBM

- $u \in L_2[0,1]$ unknown
- $P_u^{(n)}$ distribution of $Y_t^{(n)}$
- Interested in small noise limit $n \to \infty$

White noise model - Minimax estimation rates

- Minimax risk in L_2-loss over class $\mathcal{F} \subset L_2[0,1]$

$$R_n(\hat{u}, u) = \min_{\hat{u}} \max_{u \in \mathcal{F}} \mathbb{E}_{P_u^{(n)}} \|\hat{u} - u\|_{L_2}^2$$

- Minimax rate in L_2-risk over $\mathcal{F}$: fastest rate of decay of above minimax risk, as $n \to \infty$

- Linear minimax rate in L_2-risk over $\mathcal{F}$: restrict to linear estimators

WNM - Minimax estimation rates under Besov regularity

Theorem (Donoho + Johnstone '98)

In the WNM for $\beta > \frac{1}{q}$ or $\beta \geq 1$ for $q = 1$,

- Minimax rate in L_2-loss over B^{β}_{qq}

$$m_n = n^{-\frac{\beta}{1+2\beta}}$$

- Linear minimax rate in L_2-loss over B^{β}_{qq}

$$l_n = n^{-\frac{\beta-\gamma/2}{1+2\beta-\gamma}},$$

where $\gamma = \frac{2}{q} - \frac{2}{q \vee 2} \geq 0$.

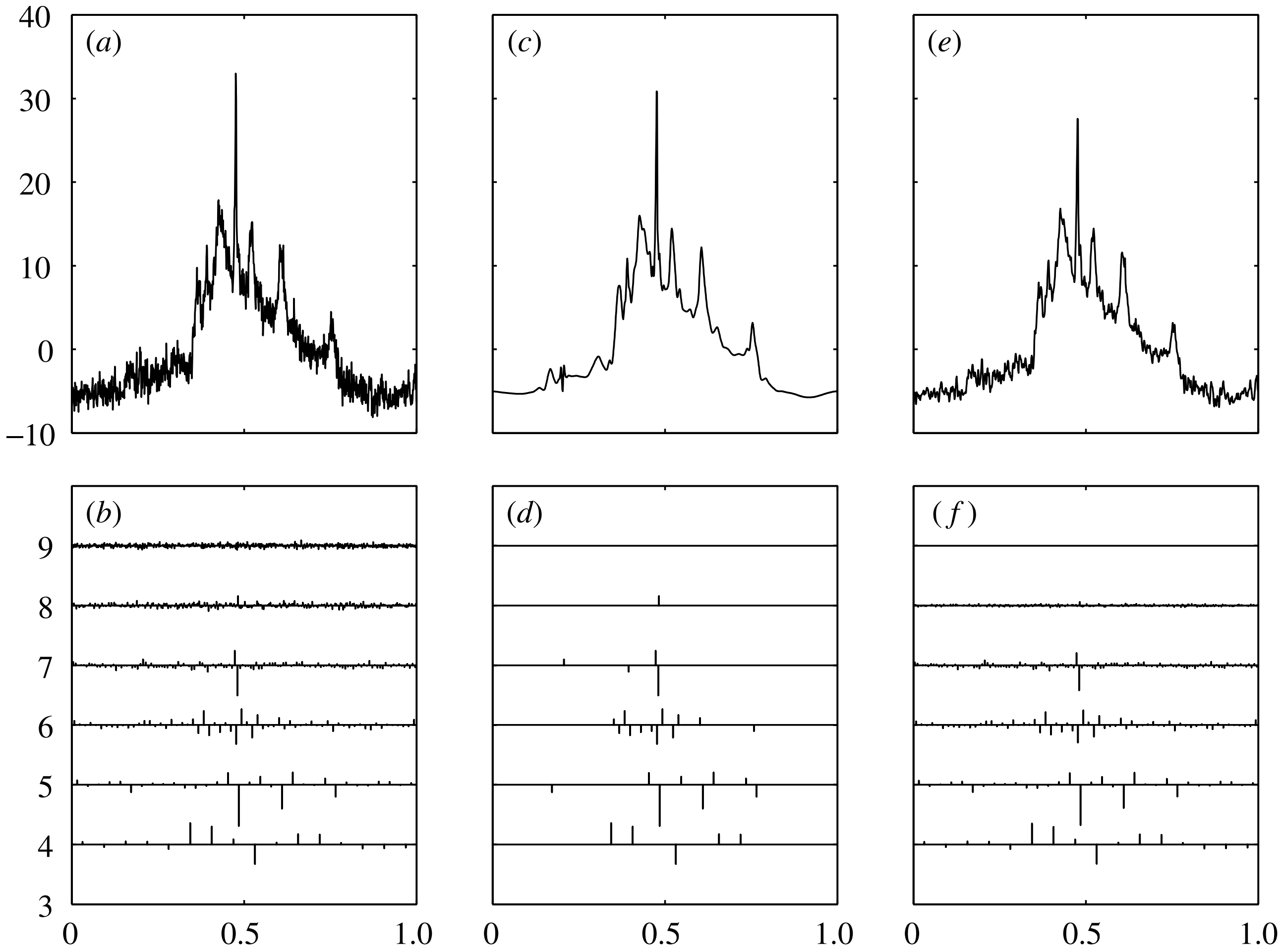
- For $q < 2$ (spatially inhomogeneous unknowns) linear estimators sub-optimal

$$\text{e.g. } q = 1, \quad l_n = n^{\frac{1-2\beta}{4\beta}}$$

- Same result holds in Gaussian regression setting

D. Donoho and I. Johnstone, *Minimax estimation via wavelet shrinkage*, Annals of Statistics, 1998.

NMR data denoising



Linear methods either oversmooth irregular part, or undersmooth regular part or both

I. Johnstone, *Wavelets and the theory of non-parametric function estimation*, Phil. tans. R. Soc. Lond. A, 1999.

Outline

p-exponential product measure

- $\xi_\ell \overset{iid}{\sim} f_p,\ f_p(x) = c_p e^{-\frac{|x|^p}{p}}, \quad p \in [1,2]$
- (γ_ℓ) decaying positive scalings
- Define *p*-exponential measure

$$\mu = \mathcal{L}\big((\gamma_\ell \xi_\ell)\big)$$

- μ log-concave (unimodal, exponential moments, ...)

p-exponential product measure

- e.g. $(\gamma_\ell) \in \ell_2$
- $\{\psi_\ell\}$ orthonormal basis in separable Hilbert-space X
- μ identified with measure in X via expansion

$$u(x) = \sum_{\ell=1}^{\infty} \gamma_\ell \xi_\ell \psi_\ell(x)$$

Shift space

Proposition (A., Dashti, Helin '20)

The space of admissible shifts of μ is the Hilbert space

$$\mathcal{Q}(\mu) = \{h \in \mathbb{R}^\infty : \|h\|_{\mathcal{Q}} < \infty\},$$

where

$$\|h\|_{\mathcal{Q}} = \Big(\sum_{\ell=1}^{\infty} \frac{h_\ell^2}{\gamma_\ell^2}\Big)^{\frac{1}{2}}.$$

For $h \in \mathcal{Q}(\mu)$

$$\frac{d\mu(\cdot - h)}{d\mu}(u) = \lim_{N\to\infty} \prod_{\ell=1}^{N} \frac{f_p(u_\ell - h_\ell)}{f_p(u_\ell)} = \lim_{N\to\infty} e^{\frac{1}{p}\sum_{\ell=1}^N \left(|\frac{u_\ell}{\gamma_\ell}|^p - |\frac{u_\ell - h_\ell}{\gamma_\ell}|^p\right)}.$$

L. Shepp, *Distinguishing a sequence of random variables from a translate of itself*, Annals of Mathematical Statistics, 1965.

S. Kakutani, *On equivalence of infinite product measures*, Annals of Mathematics, 1948.

Another important subspace

- Let $\mathcal{Z}(\mu) = \{h \in \mathbb{R}^\infty : \|h\|_{\mathcal{Z}} < \infty\}$, where

$$\|h\|_{\mathcal{Z}} = \Big(\sum_{\ell=1}^{\infty} |\frac{h_\ell}{\gamma_\ell}|^p\Big)^{\frac{1}{p}}$$

- $\mathcal{Z}$ Banach space

- $\mathcal{Z} \subset \mathcal{Q}$, both null sets (e.g. $\|(\gamma_\ell \xi_\ell)\|^2_{\mathcal{Q}} = \sum_{\ell=1}^\infty \xi_\ell^2$)

- For Gaussian μ: $\mathcal{Z} = \mathcal{Q} = \mathcal{H}$, $\mathcal{H}$ RKHS

- e.g. $X = L_2[0,1]$, $\mathcal{Z} \subset \mathcal{Q}$ identified with spaces of functions with higher regularity

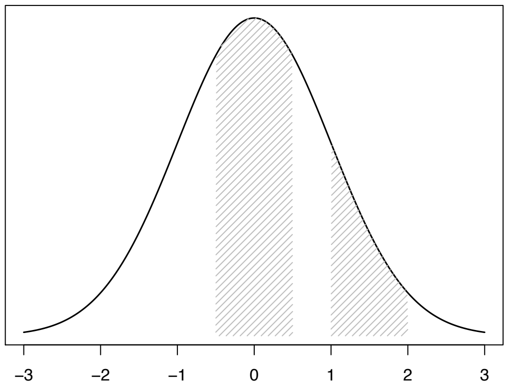
$$X = \mathrm{supp}(\mu) = \overline{\mathcal{Q}}^{\|\cdot\|_X} = \overline{\mathcal{Z}}^{\|\cdot\|_X}$$

Lower bound on probability of non-centered balls

Theorem (A., Dashti, Helin '20)

For any $h \in \mathcal{Z}$

$$\mu(\epsilon B_X + h) \geq e^{-\frac{1}{p}\|h\|_{\mathcal{Z}}^p} \mu(\epsilon B_X).$$



For proof:

- Use expression for $\frac{d\mu(\cdot - h)}{d\mu}$
- Exploit symmetry and convexity (important that $p \in [1, 2]$)

Concentration function

- Define the concentration function for μ a p-exponential measure at $w \in X$

$$\phi_w(\epsilon) = \inf_{h \in \mathcal{Z}: \|h-w\|_X \leq \epsilon} \frac{1}{p}\|h\|_{\mathcal{Z}}^p - \log \mu(\epsilon B_X)$$

- ϕ_0 measures probability of ϵ-balls around 0, $\mu(\epsilon B_X) = e^{-\phi_0(\epsilon)}$
- Last theorem + approximation:

  ϕ_w controls probability of ϵ-balls around $w \in X$ from below

- $\phi_w(\epsilon)$ blows up as $\epsilon \to 0$

Talagrand's two level concentration inequality

Lemma

There exists $K > 0$ depending only on p, s.t. for any $\epsilon > 0$ and any $M > 0$

$$\mu(\epsilon B_X + M^{\frac{p}{2}} B_{\mathcal{Q}} + M B_{\mathcal{Z}}) \geq 1 - \frac{1}{\mu(\epsilon B_X)} e^{-\frac{M^p}{K}}.$$

M. Talagrand, *The supremum of some canonical processes*, American J. of Mathematics, 1994.

For Gaussian μ, get Borell's concentration inequality

$$\mu(\epsilon B_X + M B_{\mathcal{H}}) \geq 1 - \frac{1}{\mu(\epsilon B_X)} e^{-\frac{M^2}{K}}$$

C. Borell, *The Brunn-Minkowski inequality in Gauss space*, Inventiones Mathematicae, 1975.

Rates of contraction

Bayesian context with *p*-exponential priors

- Last two results allow to apply Ghosal and van der Vaart's ROC theory
  - Lower bound on prior probability around truth
  - Sieve set of bounded complexity, capturing most of prior mass

S. Ghosal and A. van der Vaart, *Convergence rates of posterior distributions for noniid observations*, Annals of Statistics, 2007.

A. van der Vaart and H. van Zanten, *Rates of contraction of posterior distributions based on Gaussian process priors*, Annals of Statistics, 2008.

- Control probability around truth using the concentration function
- Use $\epsilon B_X + M^{\frac{p}{2}} B_{\mathcal{Q}} + M B_{\mathcal{Z}}$ as sieve set
  - Captures most of prior mass (Talagrand)
  - Complexity bound tricky because of having two large balls
  - If $\mathcal{Q}$ approximated well by $\mathcal{Z}$ in X, can handle by embedding in $2\epsilon B_X + cMB_{\mathcal{Z}}$

Outline

WNM - Gaussian sequence model

- WNM or equivalently Gaussian sequence model

$$y_\ell^{(n)} \overset{ind}{\sim} N(u_\ell, 1/n), \quad \ell \in \mathbb{N}$$

- $u = (u_\ell) \in \ell_2$ unknown
- $P_u^{(n)}$ distribution of $y^{(n)} = (y_\ell^{(n)})$
- Interested in small noise limit, $n \to \infty$

α-regular p-exponential priors

- Prior $\Pi = \mathcal{L}((\gamma_\ell \xi_\ell))$, $\xi_\ell \overset{iid}{\sim} f_p$, $p \in [1, 2]$
- $\gamma_\ell = \tau \ell^{-\frac{1}{2}-\alpha}$ $\quad \left(\gamma_{kl} = \tau 2^{-(\frac{1}{2}+\alpha)k}\right)$
- $\tau > 0$ scaling parameter
- $\alpha > 0$ regularity parameter

Lemma

For any $q \geq 1$, it holds $\Pi(B^s_{qq}) = 1$ for all $s < \alpha$ and $\Pi(B^s_{qq}) = 0$ for all $s \geq \alpha$.

- For $p = 1$, $B^{\alpha+1}_{11}$-Besov prior

α-regular p-exponential priors

- Space of admissible shifts

$$\mathcal{Q} := \mathcal{Q}_\alpha = \{h \in \mathbb{R}^\infty : \|h\|_{\mathcal{Q}_\alpha} < \infty\}, \quad \|h\|_{\mathcal{Q}_\alpha} = \tau^{-1}\left(\sum_{\ell=1}^{\infty} \ell^{1+2\alpha} h_\ell^2\right)^{\frac{1}{2}}$$

- Space determining mass-loss for noncentered balls

$$\mathcal{Z} := \mathcal{Z}_\alpha = \{h \in \mathbb{R}^\infty : \|h\|_{\mathcal{Z}_\alpha} < \infty\}, \quad \|h\|_{\mathcal{Z}_\alpha} = \tau^{-1}\left(\sum_{\ell=1}^{\infty} \ell^{\frac{p}{2}+p\alpha} |h_\ell|^p\right)^{\frac{1}{p}}$$

- Identified with Besov spaces $\mathcal{Q}_\alpha = B_{22}^{\alpha+\frac{1}{2}}$ and $\mathcal{Z}_\alpha = B_{pp}^{\alpha+\frac{1}{p}}$
- Scaling parameter τ appears in norms

WNM - Rates of contraction

- Concentration function

$$\phi_w(\epsilon) = \inf_{h \in B_{pp}^{\alpha+\frac{1}{p}} : \|h-w\|_{\ell_2} \leq \epsilon} \frac{\mathcal{T}^{-p}}{p} \|h\|^p_{B_{pp}^{\alpha+\frac{1}{p}}} - \log \Pi(\epsilon B_{\ell_2})$$

Theorem (A., Dashti, Helin '20)

Let $u_0 \in \ell_2$ and let $\epsilon_n \to 0$ such that

$$\phi_{u_0}(\epsilon_n) \leq n\epsilon_n^2.$$

Then as $n \to \infty$

$$\Pi_n(u \; : \; \|u - u_0\|_{\ell_2} > M\epsilon_n \,|\, y^{(n)}) \to 0,$$

in $P_{u_0}^{(n)}$-probability, for some $M > 0$.

Estimating the concentration function

- Centered small ball probabilities: for any $\tau > 0$, $\alpha > 0$ and $p \in [1, 2]$

$$-\log \Pi(\epsilon B_{\ell_2}) \asymp (\epsilon/\tau)^{-\frac{1}{\alpha}}$$

F. Aurzada, *On the lower tail probabilities of some random sequences in ℓ_p*, J. Theoretical Probability, 2007.

- Decentering:

$$\inf_{h \in B_{pp}^{\alpha+\frac{1}{p}} : \|h - u_0\|_{\ell_2} \leq \epsilon} \frac{\tau^{-p}}{p} \|h\|^p_{B_{pp}^{\alpha+\frac{1}{p}}}$$

  - $h_{1:L}$ truncation of u_0 up to L, $u_{1:L} \in B_{pp}^{\alpha+\frac{1}{p}}$
  - Depending on regularity of u_0, for large enough L, $\|h_{1:L} - u_0\|_{\ell_2} \leq \epsilon$
  - Depending on regularity of u_0, get bound on $\|h_{1:L}\|_{B_{pp}^{\alpha+\frac{1}{p}}}$ hence also on infimum

Rates under Sobolev regularity - no rescaling

Theorem (A., Dashti, Helin '20)

Assume $u_0 \in B^{\beta}_{22}$. Consider α-regular p-exponential priors $p \in [1,2]$, with fixed $\tau > 0$. Then the posterior contracts at rate

$$\epsilon_n = \begin{cases} n^{-\frac{\beta}{1+2\beta+p(\alpha-\beta)}}, & \text{if } \alpha > \beta \\ n^{-\frac{\alpha}{1+2\alpha}}, & \text{if } \alpha \leq \beta \end{cases}$$

- For $\alpha = \beta$ minimax rate $m_n = n^{-\frac{\beta}{1+2\beta}}$
- For $\alpha < \beta$ rate independent of p

L. Zhao, *Bayesian aspects of some nonparametric problems*, Annals of Statistics, 2000.

I. Castillo and R. Nickl, *Nonparametric Bernstein - von Mises theorems in Gaussian white noise*, Annals of Statistics, 2013.

- For $\alpha > \beta$ the smaller p is the faster the rate

Rates under Sobolev regularity - rescaling

Theorem (A., Savva '21)

Assume $u_0 \in B^{\beta}_{22}$. Consider α-regular p-exponential priors $p \in [1,2]$, with $\tau_n = \tau_n(\alpha, \beta, p)$ chosen optimally. Then the posterior contracts at rate

$$\epsilon_n := \begin{cases} n^{-\frac{\beta}{1+2\beta}}, & \text{if } \alpha > \beta - 1/p \\ n^{-\frac{\beta}{1+2\beta}} \log^{\frac{2-p}{2\beta p+p}} n, & \text{if } \alpha = \beta - 1/p \\ n^{-\frac{1+\alpha p}{2+p+2\alpha p}}, & \text{if } \alpha < \beta - 1/p. \end{cases}$$

- The smaller p is, the more undersmoothing the prior can be while still achieving the minimax rate
- Same picture for $u_0 \in B^{\beta}_{qq}$, $q > 2$

Rates under spatially inhomogeneous truth - no rescaling

Theorem (A., Dashti, Helin '20)

Assume $u_0 \in B^{\beta}_{qq}$, $q < 2$, $\beta > 0 \vee (\frac{1}{q} - \frac{1}{2})$. Consider α-regular p-exponential priors $p \in [1, 2]$, with fixed $\tau > 0$. Then the best ROC ϵ_n is achieved for

$$\alpha = \alpha(\beta, p, q) < \beta$$

and

- If $p \leq q$
$$m_n \ll \epsilon_n \ll \; l_n.$$
As p grows towards q the best rate improves.
- If $q < p < 2$
$$m_n \ll \epsilon_n \ll \; l_n.$$
As p decreases towards q the best rate improves.
- If $p = 2$
$$\epsilon_n = l_n.$$

("$\ll$" = polynomially faster, l_n linear minimax rate)

Rates under spatially inhomogeneous truth - with rescaling

Theorem (A., Dashti, Helin '20)

Assume $u_0 \in B^{\beta}_{qq}$, $q < 2$, $\beta > \frac{1}{p} \vee \frac{1}{q}$. Consider α-regular p-exponential priors $p \in [1, 2]$, with $\tau_n = \tau_n(\alpha, \beta, p, q)$ chosen optimally. Then the posterior contracts at rate ϵ_n s.t.:

- For $p = q$, $\alpha = \beta - \frac{1}{p}$

$$\epsilon_n = m_n \quad \text{for} \quad \tau_n = n^{-\frac{1}{p(1+2\beta)}}.$$

- For $p < q$, $\alpha = \beta - \frac{1}{p}$

$$\epsilon_n = m_n \log^{\frac{q-p}{pq(1+2\beta)}} n.$$

- In all other cases

$$m_n \ll \epsilon_n.$$

- For $p = 2$ the best achievable rate is $\epsilon_n = l_n$, achieved for any $\alpha \geq \beta - \frac{1}{q}$.

e.g. $u_0 \in B^1_{11}$ can use rescaled α-regular Laplace prior, with $\alpha > 0$ arbitrarily small, to get $\epsilon_n = m_n n^{\delta}$ for $\delta > 0$ arbitrarily small.

Difficulty for $q < 2$

Assume $u_0 \in B^{\beta}_{qq}$

$$\sum_{\ell=1}^{\infty} \ell^{\beta q + \frac{q}{2} - 1} |u_{0,\ell}|^q < \infty$$

- $h_{1:L}$ truncation of u_0 up to L

$$\|h_{1:L} - u_0\|^2_{\ell_2} = \sum_{\ell > L} u^2_{0,\ell} = \sum_{\ell > L} \ell^{-2\beta - 1 + \frac{2}{q}} \ell^{2\beta + 1 - \frac{2}{q}} u^2_{0,\ell}$$

- If $q = 2$ can bound by $\|u_0\|^2_{B^{\beta}_{22}} L^{-2\beta}$
- If $q > 2$ can use Hölder inequality $(\frac{q}{2}, \frac{q}{q-2})$ to bound by $\|u_0\|^2_{B^{\beta}_{qq}} L^{-2\beta}$
- If $q < 2$ cannot use Hölder inequality, forced to use crude bound

$$|u_{0,\ell}| \leq \|u_0\|_{B^{\beta}_{qq}} \ell^{-\beta - \frac{1}{2} + \frac{1}{q}}$$

- Estimating $\|h_{1:L}\|_{B^{\alpha + \frac{1}{p}}_{pp}}$: if $p \leq q$ can use Hölder, if $p > q$ need crude bound

Difficulty for $q < 2$

$$\phi_{u_0}(\epsilon) \leq \tau^{-p}\epsilon^{-s} + (\epsilon/\tau)^{-\frac{1}{\alpha}}, \qquad s = s(\alpha, \beta, p, q) > 0$$

- Inf-term worse for larger α (s increases with α)
- Centered small ball term better for larger α
- To achieve minimax rate need centered term to dominate for $\alpha \leq \beta$
- Crude bound: inf-term dominates already for large enough $\alpha < \beta$
- Address by choosing small α s.t. centered term dominates, inf-term better
  - Correct by sending $\tau \to 0$ which improves centered term
  - Pay penalty τ^{-p} in infimum term, hopefully less than choosing larger α
  - For $p = 2$ penalty identical to increasing α, no improvement
  - For $q < p < 2$ improvement not enough to get minimax rate
  - For $p \leq q$ enough improvement to get minimax rate (up to logs for $p < q$)

Are Gaussian priors limited by the linear minimax rate?

In general minimax rate in L_2-risk benchmark for ROC in L_2

- ROC ϵ_n implies center of smallest ball containing half posterior mass, converges at same rate in probability

  S. Ghosal, J. Ghosh and A. van der Vaart, *Convergence rates of posterior distributions*, Annals of Statistics, 2000.

- Lower bounds on minimax rates in L_2-risk proved by showing lower bounds in probability (stronger statement due to Markov inequality)

Are Gaussian priors limited by the linear minimax rate?

For Gaussian priors, hence Gaussian posteriors

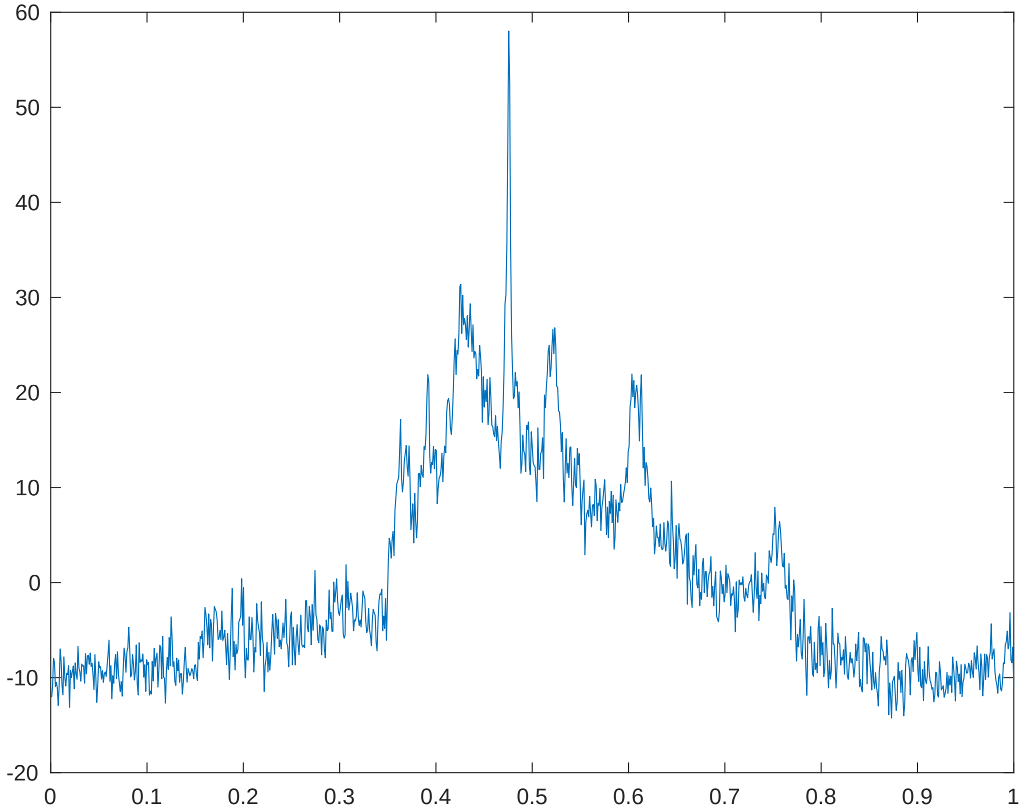
- Posterior mean = MAP estimator, linear estimator
- ROC ϵ_n implies posterior mean converges at same rate in probability
- Lower bounds on linear minimax rate derived in L_2-risk directly
- Need to show that posterior mean converges at same rate also in L_2-risk

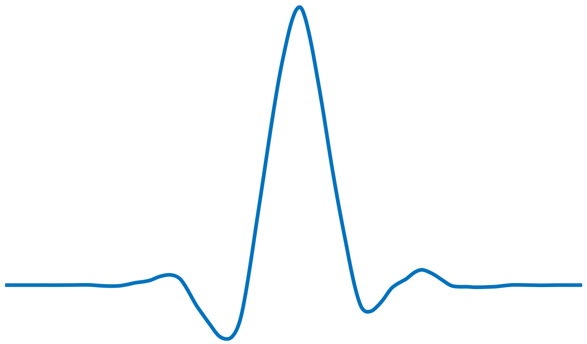
(S. Wang)

Outline

NMR data

- Nuclear Magnetic Resonance data, available in WaveLab 850
- Signal expanded in Symlet 6 orthonormal wavelet basis $\{\psi_{kl}\}$ truncated at $k = 9$





Bayesian Denoising of NMR data

- Model wavelet coefficients as

$$y_{kl} = u_{kl} + \frac{1}{\sqrt{\delta}} z_{kl}, \quad z_{kl} \overset{iid}{\sim} N(0,1)$$

- Rescaled α-regular p-exponential prior on unknown $u = (u_{kl})$, with $p = 1$ or 2

$$u_{kl} = \tau 2^{-(\frac{1}{2}+\alpha)k} \xi_{kl}, \quad \xi_{kl} \overset{iid}{\sim} f_p, \;\; p = 1 \text{ or } 2$$

- Hyperprior on prior-rescaling τ: $\tau^{-2} \sim \mathrm{Gamma}(a_1, b_1)$

- Hyperprior on noise-precision δ: $\delta \sim \mathrm{Gamma}(a_2, b_2)$

- a_1, a_2, b_1, b_2 chosen so that hyperpriors non-informative for τ, δ

Bayesian Denoising of NMR data - Gaussian prior

- Conditional conjugacy
  - $u_{kl}|y_{kl}, \tau, \delta \sim N\big(m_{kl}\ ,\ c_{kl}\big)$
  - $\tau^{-2}|u, y \sim \mathrm{Gamma}\big(a_1'\ ,\ b_1'(u)\big)$
  - $\delta|u, y \sim \mathrm{Gamma}\big(a_2'\ ,\ b_2'(u, y)\big)$
- Can use simple Gibbs Sampler to sample posterior
- Normally in high-dim τ-chain mixes poorly $\big($$u$ and τ a-priori strongly dependent$\big)$

  $\rightarrow$ use non-centered parametrization $u = \tau v$, and work with v instead of u

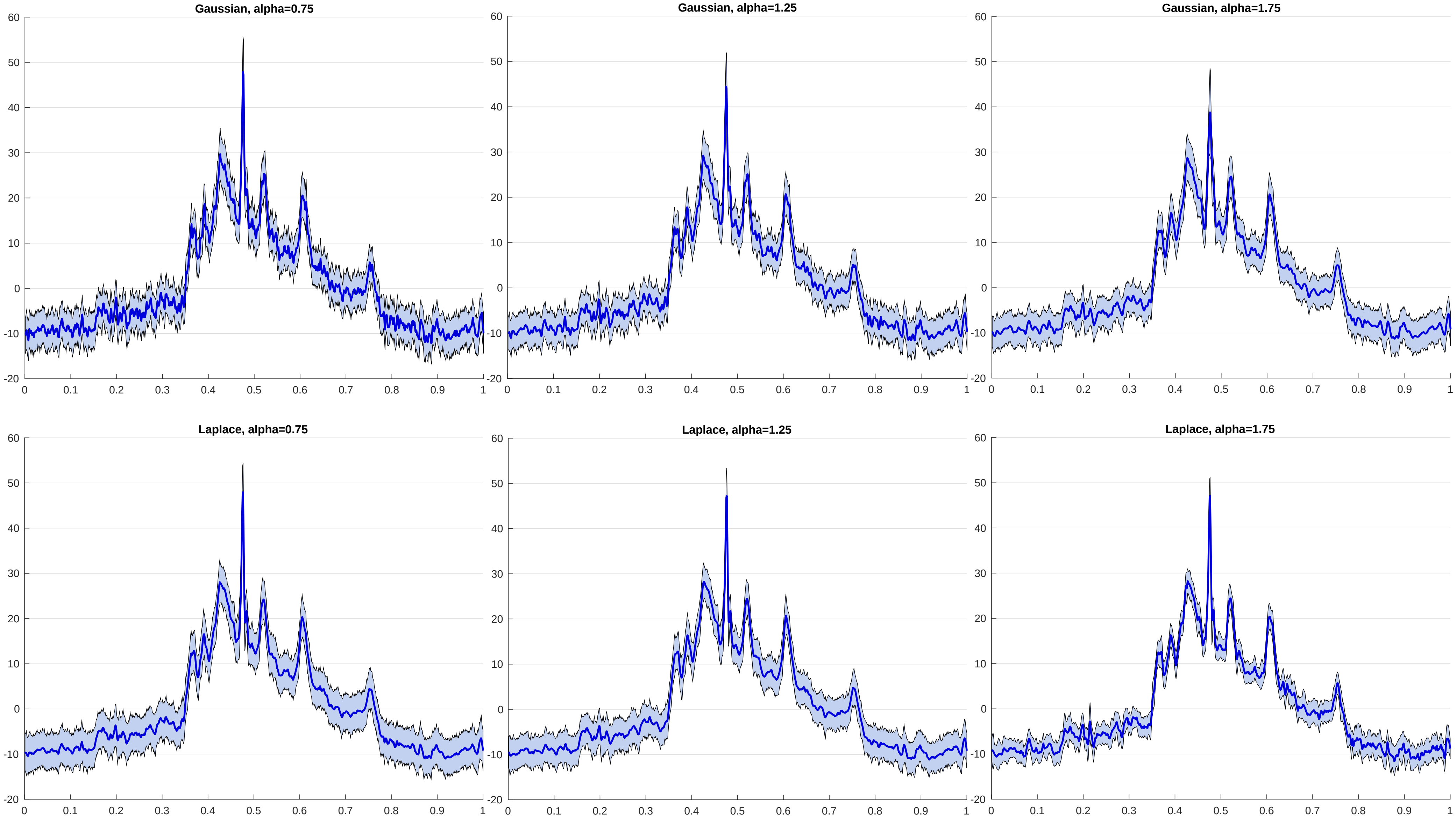
S. Agapiou, J. Bardsley, O. Papaspiliopoulos, A. Stuart *Analysis of the Gibbs Sampler for Hierarchical Inverse Problems*, SIAM/ASA Journal on UQ, 2014.

Bayesian Denoising of NMR data - Laplace prior

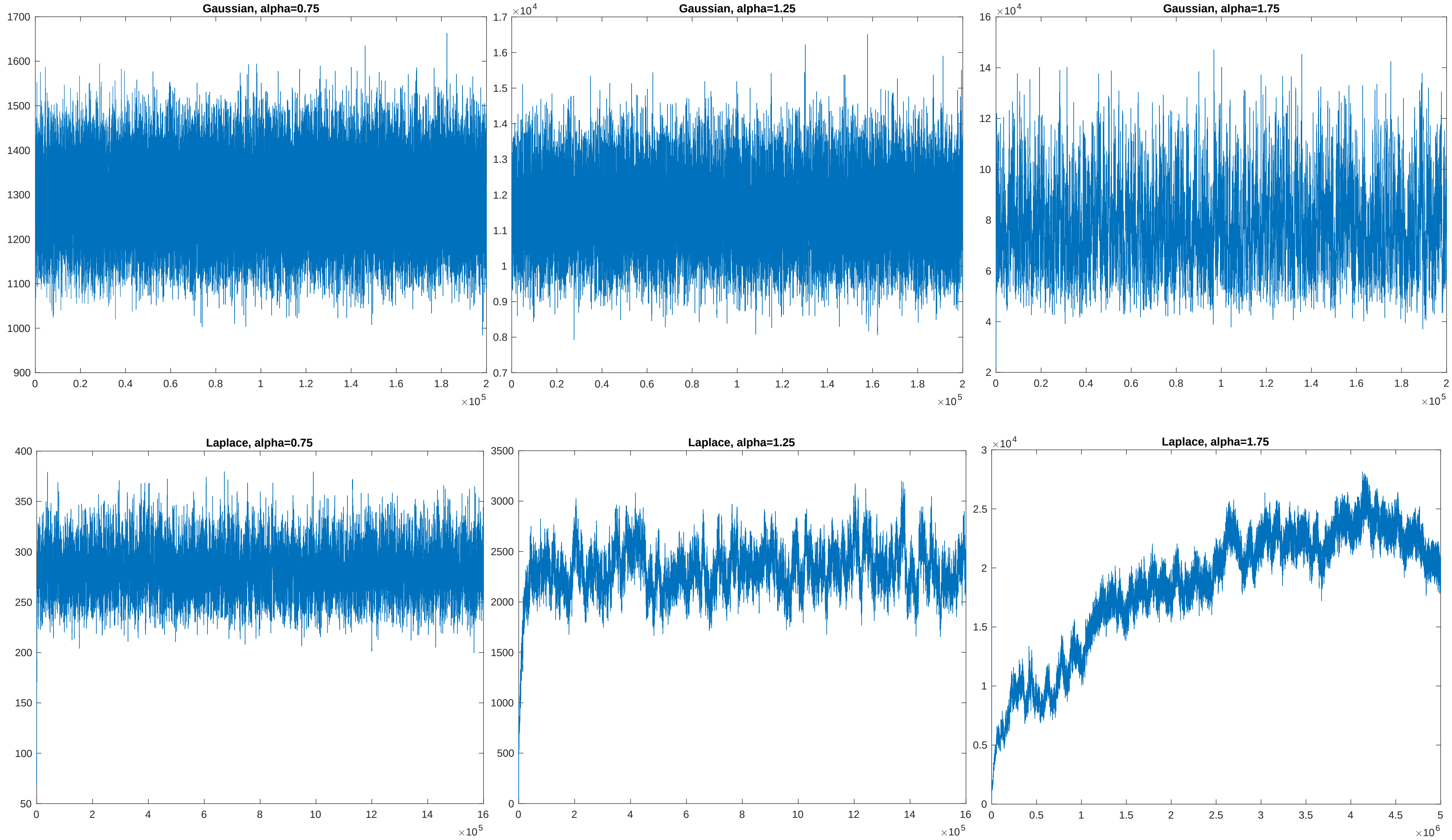
- No conditional conjugacy (only for $\delta|u, y$)
- Need to use Metropolis within Gibbs
- pCN dimension-robust for Gaussian priors
- Again u, τ a-priori strongly dependent
- Use non-centered pCN within Gibbs
  - Write $u = T(\zeta, \tau)$ such that ζ, τ a-priori independent and ζ is Gaussian WN
  - Sample iteratively $\zeta|y, \tau$ (pCN) and $\tau|y, \zeta$ (independence sampler)

V. Chen, M. Dunlop, O. Papaspiliopoulos, A. Stuart *Dimension-Robust MCMC in Bayesian Inverse Problems*, arXiv:1803.03344.

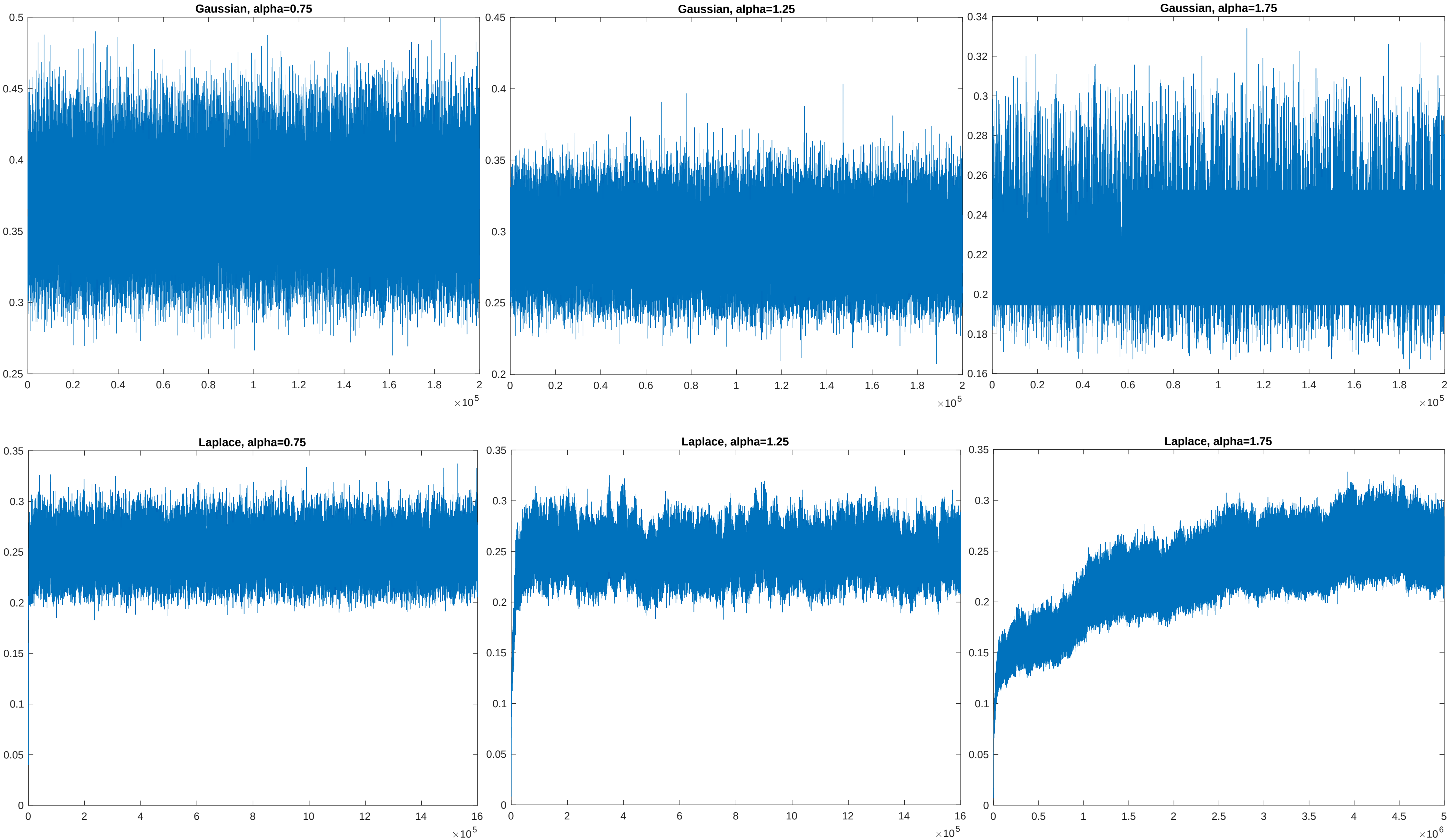
NMR data - Gauss vs Laplace priors



NMR data - Gauss vs Laplace priors - τ-chains



NMR data - Gauss vs Laplace priors - δ-chains



Outline

Summary and open questions

- Laplace priors appear to outperform Gaussian priors over Besov regularity
- Sharpness of rates
- For benefit to be realized need better algorithms
- Adaptation over Besov spaces (with A. Savva)
- ROC for Bayesian inverse problems with Besov-priors (with S. Wang)

Summary and open questions

- Laplace priors appear to outperform Gaussian priors over Besov regularity
- Sharpness of rates
- For benefit to be realized need better algorithms
- Adaptation over Besov spaces (with A. Savva)
- ROC for Bayesian inverse problems with Besov-priors (with S. Wang)

THANK YOU!

http://www.sergiosagapiou.com/

- S. Agapiou, M. Dashti and T. Helin, *Rates of contraction of posterior distributions based on p-exponential priors*, arXiv:1811.12244 (to appear in Bernoulli).

- S. Agapiou, M. Burger, M. Dashti and T. Helin, *Sparsity-promoting and edge-preserving MAP estimators in nonparametric Bayesian inverse problems*, Inverse Problems, 2018.

- I.M. Johnstone, *Gaussian estimation: sequence and wavelet models*, draft book.

- E. Gine and R. Nickl, *Mathematical Foundations of Infinite-Dimensional Statistical Models*, Cambridge Series in Statistical and Probabilistic Mathematics, 2016.

- S. Ghosal and A. van der Vaart, *Fundamentals of nonparametric Bayesian inference*, Cambridge Series in Statistical and Probabilistic Mathematics, 2017.

- M. Ledoux, *The concentration of measure phenomenon*, AMS mathematical surveys and monographs, 2005