Adaptive rates of contraction for spatially inhomogeneous unknowns

Sergios Agapiou

Department of Mathematics and Statistics, University of Cyprus

May 2022, AUEB Statistics Seminar

University of Cyprus

Joint work with

- S. Agapiou, M. Dashti and T. Helin, *Rates of contraction of posterior distributions based on p-exponential priors*, Bernoulli, 2021.
- S. Agapiou and S. Wang, *Laplace priors and spatial inhomogeneity in Bayesian inverse problems*, arXiv:2112.05679.
- S. Agapiou and A. Savva, *Adaptive rates of contraction based on p-exponential priors*, in preparation.

Outline

Outline

Multiscale features in images

Wavelet expansions

- $\{\psi_\ell\}_{\ell=1}^{\infty}$ orthonormal basis for L_2

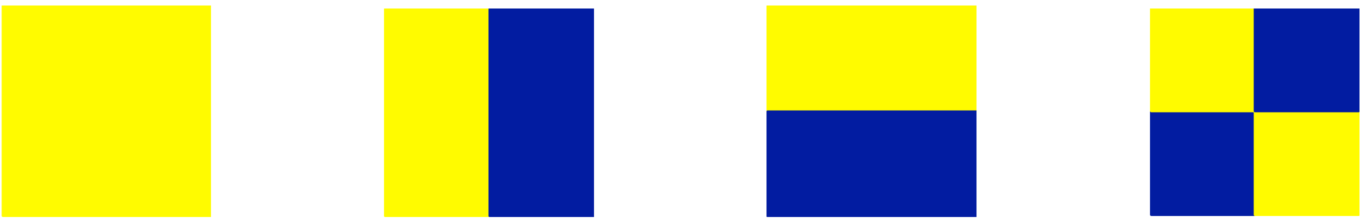
$$u(x) = \sum_{\ell=1}^{\infty} u_\ell \psi_\ell(x), \quad u_\ell = \langle u, \psi_\ell \rangle.$$

- 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



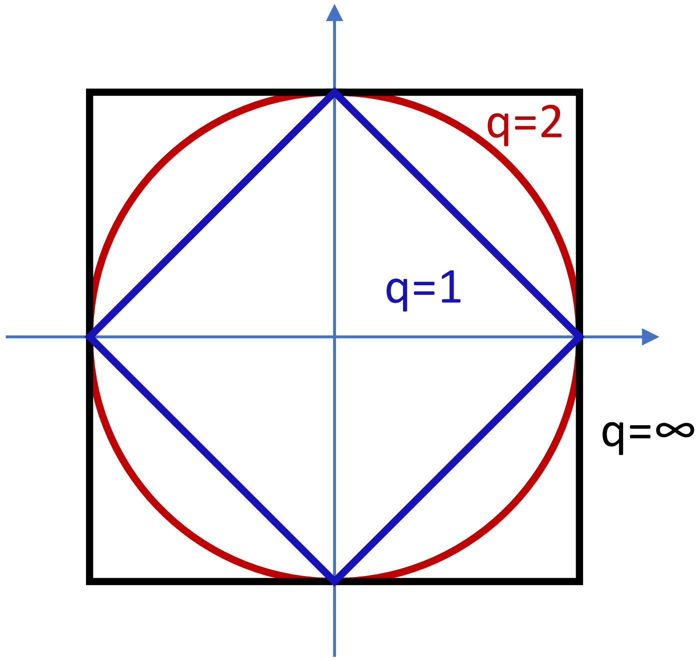
Besov Spaces

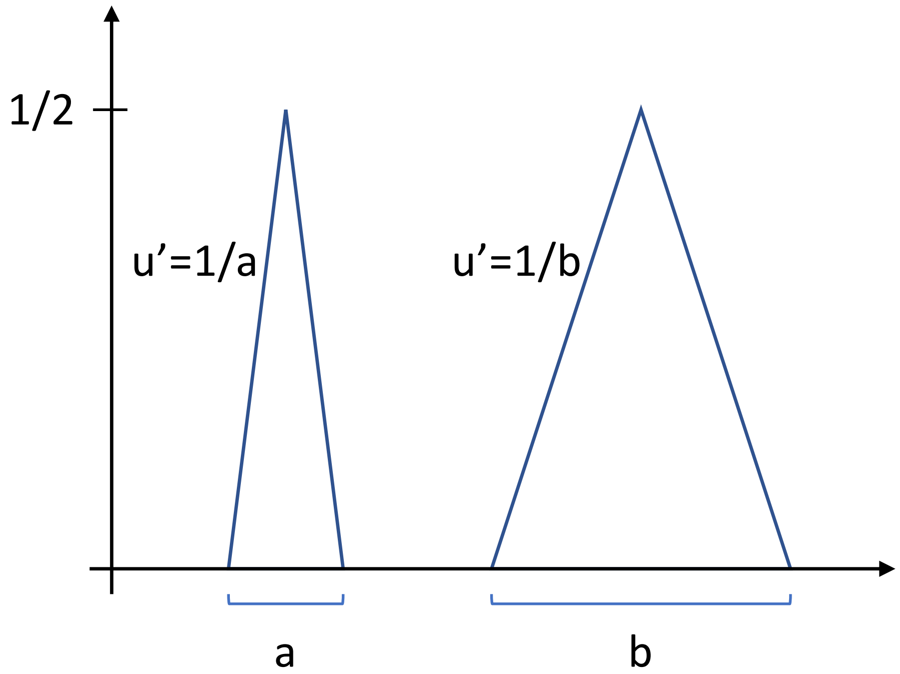
- Functions identified with expansion coefficients $(u_\ell) \in \ell_2$ or $(u_{kl}) \in \ell_2$
- Besov space of smoothness $s \in \mathbb{R}$, with integrability parameter $q \geq 1$

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

- $q = 2$: $B^s_{22} = H^s$, 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

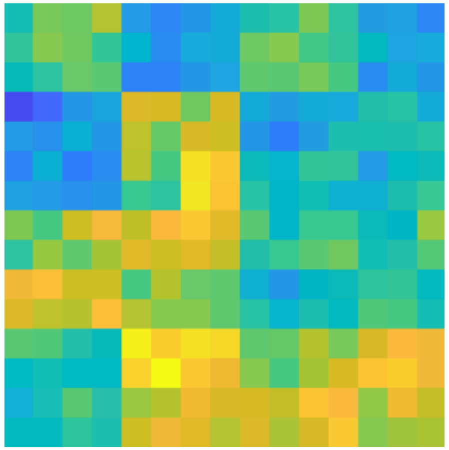
$$u(x) = \sum_{\ell=1}^{\infty} u_\ell \psi_\ell(x)$$

- 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

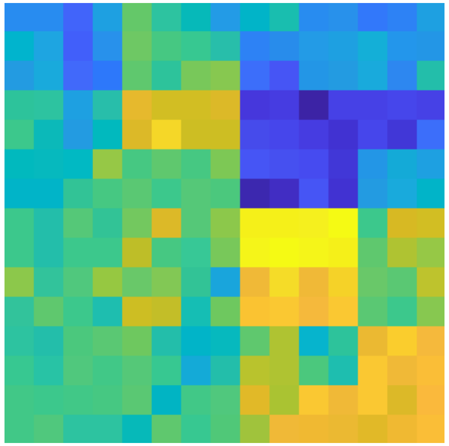
$$\text{"}\pi(u) \propto \exp(-\|u\|_{B^s_{11}})\text{"}$$

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

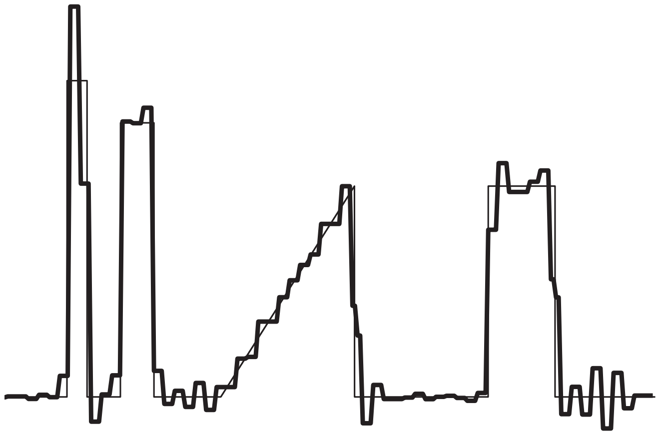
Priors via random series expansions - Haar wavelets



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.

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

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, work in ℓ_2

- P_u^n distribution of Y^n

- Interested in small noise limit $n \to \infty$

White noise model - Minimax estimation rates

- Minimax risk in ℓ_2-loss over class $\mathcal{F} \subset \ell_2$

$$R_n(\hat{u}, u) = \min_{\hat{u}} \max_{u \in \mathcal{F}} \mathbb{E}_{P_u^n} \|\hat{u} - u\|_2^2$$

- Minimax rate in ℓ_2-risk over $\mathcal{F}$: fastest rate of decay of 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}

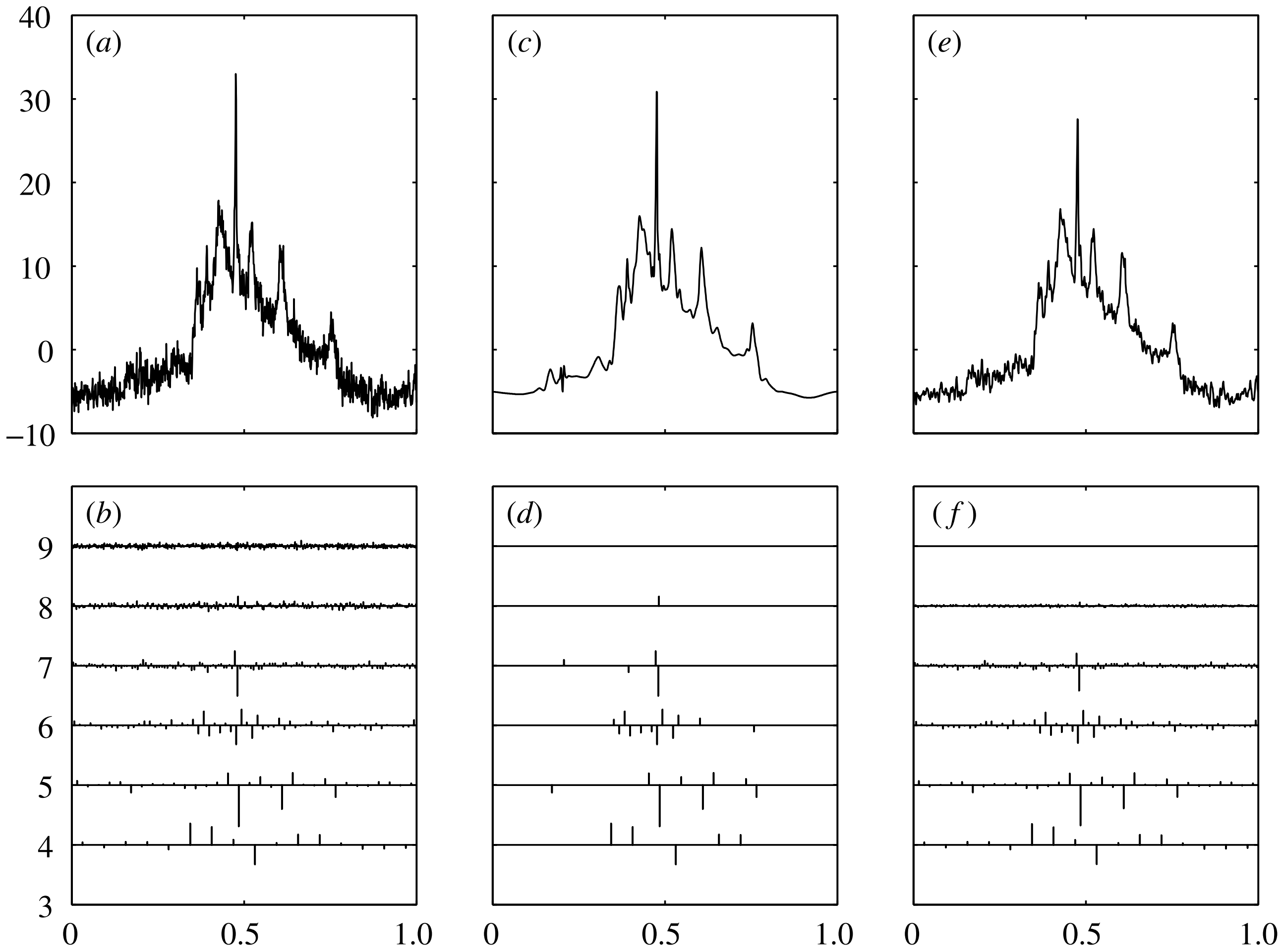
$$l_n = n^{-\frac{\beta-\gamma/2}{1+2\beta-\gamma}},$$

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

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

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

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

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

Frequentist performance of posterior

- Prior Π on $u \in \ell_2$
- Posterior $\Pi(\cdot|Y^n)$ on u
- Frequentist assumption: observations Y^n in WNM generated from fixed $u_0 \in \ell_2$
- ϵ_n is a posterior contraction rate at u_0, if $\exists M > 0$ such that as $n \to \infty$

$$\Pi(u : \|u - u_0\|_2 \geq M\epsilon_n | Y^n) \to 0$$

in $P^n_{u_0}$-probability

- Do Gaussian priors perform better for Sobolev truths?
- Do Laplace priors perform better for spatially inhomogeneous truths?

Rates of contraction - General Theory

General contraction theory $\rightarrow$ rate ϵ_n depends on

- Prior putting a certain minimum mass on small ℓ_2-balls around u_0
- Existence of sieve sets such that:
  - capture the bulk of prior's mass
  - their elements can be tested against u_0 with good enough type I & type II errors

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

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

Shift space

Proposition (A., Dashti, Helin '21)

The space of admissible shifts of Π is the Hilbert space

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

where

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

For $h \in \mathcal{Q}$

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

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 $\mathcal{Z} = \{h \in \mathbb{R}^\infty : \|h\|_\mathcal{Z} < \infty\}$, where

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

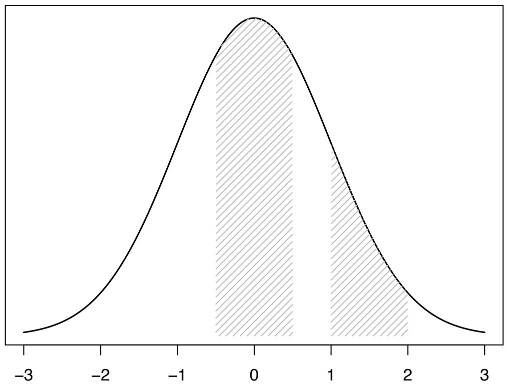
- $\mathcal{Z}$ Banach space
- $\mathcal{Z} \subset \mathcal{Q}$, both null sets (e.g. $\|(\gamma_\ell \xi_\ell)\|_\mathcal{Q}^2 = \sum_{\ell=1}^\infty \xi_\ell^2$)
- For Gaussian Π: $\mathcal{Z} = \mathcal{Q} = \mathcal{H}$, $\mathcal{H}$ RKHS

Lower bound on probability of non-centered balls

Theorem (A., Dashti, Helin '21)

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

$$\Pi(\epsilon B_{\ell_2} + h) \geq e^{-\frac{1}{p}\|h\|_{\mathcal{Z}}^p} \Pi(\epsilon B_{\ell_2}).$$



For proof:

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

Concentration function

- Define the concentration function for Π a p-exponential measure at $w \in \ell_2$

$$\phi_w(\epsilon) = \inf_{h \in \mathcal{Z}: \|h-w\|_2 \le \epsilon} \frac{1}{p}\|h\|_{\mathcal{Z}}^p - \log \Pi(\epsilon B_{\ell_2})$$

- ϕ_0 measures probability of ϵ-balls around 0, $\Pi(\epsilon B_{\ell_2}) = e^{-\phi_0(\epsilon)}$
- Last theorem + approximation:

ϕ_w controls probability of ϵ-balls around $w \in \ell_2$ from below

- Note that $\phi_w(\epsilon) \to 0$, 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 $M > 0$

$$\Pi(\epsilon B_{\ell_2} + M^{\frac{p}{2}} B_{\mathcal{Q}} + M B_{\mathcal{Z}}) \geq 1 - \frac{1}{\Pi(\epsilon B_{\ell_2})} e^{-\frac{M^p}{K}}.$$

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

For Gaussian Π, get Borell's concentration inequality

$$\Pi(\epsilon B_{\ell_2} + M B_{\mathcal{H}}) \geq 1 - \frac{1}{\Pi(\epsilon B_{\ell_2})} e^{-\frac{M^2}{K}}$$

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

Rates of contraction

- Recall 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
- Control probability around truth using the concentration function
- Use $\epsilon B_{\ell_2} + M^{\frac{p}{2}} B_{\mathcal{Q}} + M B_{\mathcal{Z}}$ as sieve set
  - Captures most of prior mass (Talagrand)
  - Concentration function turns out to control complexity as well

Outline

α-regular τ-scaled p-exponential priors

- Prior $\Pi = \mathcal{L}((\gamma_\ell \xi_\ell)),\ \ \xi_\ell \overset{iid}{\sim} f_p,\ \ p \in [1, 2]$
- $\gamma_\ell = \tau \ell^{-\frac{1}{2}-\alpha} \qquad (\gamma_{kl} = \tau 2^{-(\frac{1}{2}+\alpha)k})$
- $\tau > 0$ 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 $s \geq \alpha$.

α-regular τ-scaled p-exponential priors

- Space of admissible shifts $\mathcal{Q} := \mathcal{Q}_\alpha$

$$\|h\|_{\mathcal{Q}_\alpha} = \tau^{-1}\left(\sum_{\ell=1}^{\infty} \ell^{1+2\alpha} h_\ell^2\right)^{\frac{1}{2}}$$

- Space determining mass-loss for noncentered ball $\mathcal{Z} := \mathcal{Z}_\alpha$

$$\|h\|_{\mathcal{Z}_\alpha} = \tau^{-1}\left(\sum_{\ell=1}^{\infty} \ell^{\frac{p}{2}+p\alpha} |h_\ell|^p\right)^{\frac{1}{p}}$$

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

- Concentration function

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

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

Theorem (A., Dashti, Helin '21)

Assume $u_0 \in B^{\beta}_{22}$ and consider an α-regular τ-scaled p-exponential prior $p \in [1, 2]$. Then if either

- $\alpha = \beta$ with $\tau > 0$ fixed, or
- $\alpha > \beta - \frac{1}{p}$ and $\tau = \tau(n; \alpha, \beta, p)$ chosen optimally

the posterior contracts at the minimax rate $m_n = n^{-\frac{\beta}{1+2\beta}}$.

Rates under Sobolev regularity - adaptation

Same rates with data driven choice of α or τ (no a priori knowledge of β required)

- Hierarchical Bayes on smoothness α, e.g. using exponential hyper-prior
- Hierarchical Bayes on scaling τ, e.g. using inverse gamma hyper-priors
- Empirical Bayes, estimate α or τ using the maximum marginal likelihood estimator
- In preparation, with A. Savva

B. T. Szabó, A. W. van der Vaart, and J. H. van Zanten, *Empirical Bayes scaling of Gaussian priors in the white noise model,* Electronic Journal of Statistics, 2013.

B. T. Knapik, B. Szabó, A. W. van der Vaart, and J. van Zanten, *Bayes procedures for adaptive inference in inverse problems for the white noise model,* Probability Theory and Related Fields, 2016.

J. Rousseau, B. Szabó, *Asymptotic behaviour of the empirical Bayes posteriors associated to maximum marginal likelihood estimator,* The Annals of Statistics, 2017.

Rates under spatially inhomogeneous truth

Theorem (A., Dashti, Helin '21)

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

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

$$\epsilon_n = m_n.$$

- 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

$$\epsilon_n \gg m_n.$$

- For $p = 2$ the best achievable rate is $\epsilon_n = l_n \gg m_n$ (l_n linear minimax).

Appropriately tuned Laplace priors achieve minimax rate for $q < 2$ (up to logs if $q > 1$)

Suboptimality of Gaussian priors for spatial inhomogeneity

Theorem (A. and Wang '21)

Assume $\beta > 1/q$, $1 \leq q < 2$ or $\beta = q = 1$, and let $\delta_n \downarrow 0$ as $n \to \infty$.

Let $(\Pi_n : n \in \mathbb{N})$ be mean-zero Gaussian priors supported on L_2, such that for all $\eta > 0$

$$\sup_{u_0 : \|u_0\|_{B^\beta_{qq}} \leq 1} P^n_{u_0}\Big(\Pi_n\big(u : \|u - u_0\|_2 \geq \delta_n \big| Y_n\big) \geq \eta\Big) \xrightarrow{n\to\infty} 0.$$

Then there exists some constant $c > 0$ such that

$$\delta_n \geq c l_n, \quad n \in \mathbb{N}.$$

- Uniform statement on contraction rate required to link to minimax
- Tuned Laplace priors satisfy uniform contraction with $\delta_n = m_n \ll l_n$!

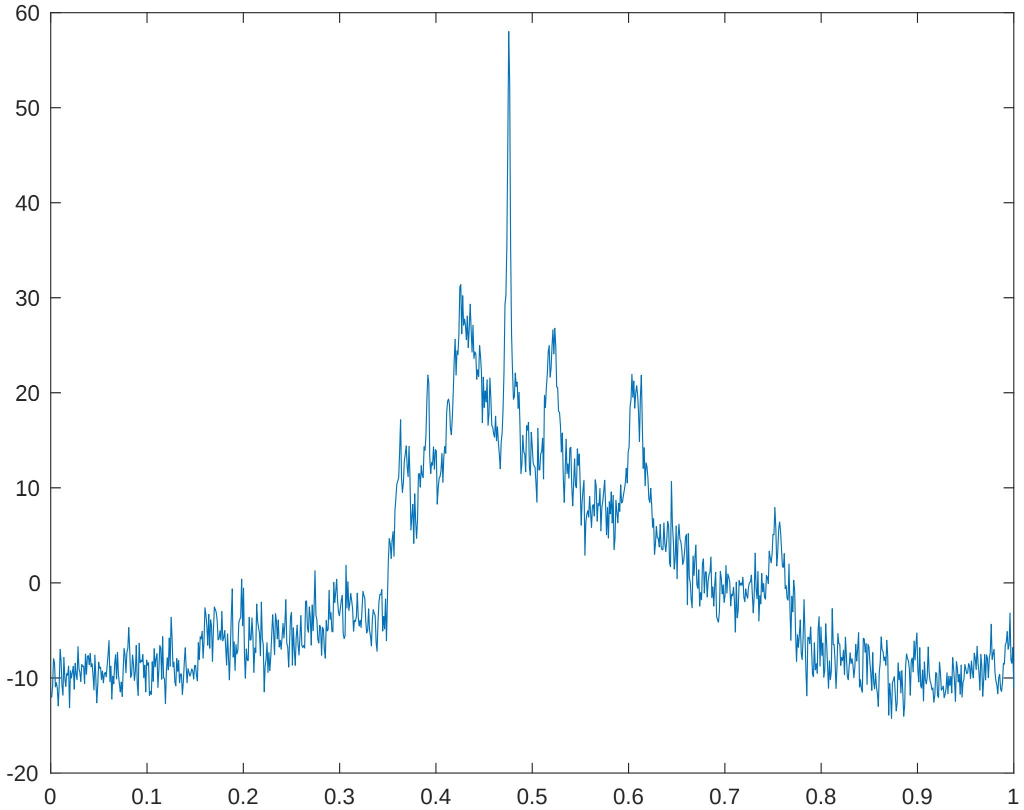
Gaussian vs Laplace priors

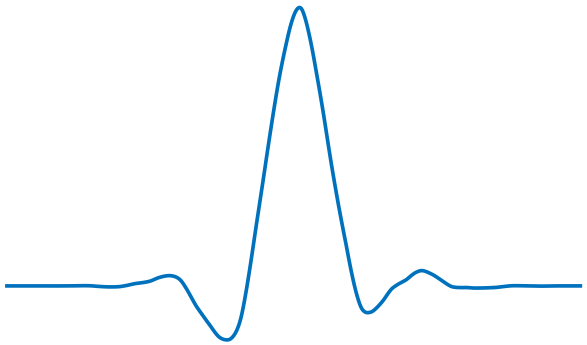
- For Sobolev truths Gaussian and Laplace priors have similar performance
- For spatially inhomogeneous truths, tuned Laplace priors outperform Gaussians
- Tuning, smoothness and scaling simultaneously, can be performed adaptively using Hierarchical or Empirical Bayes approach (in preparation with A. Savva)

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 (u and τ a-priori strongly dependent)

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

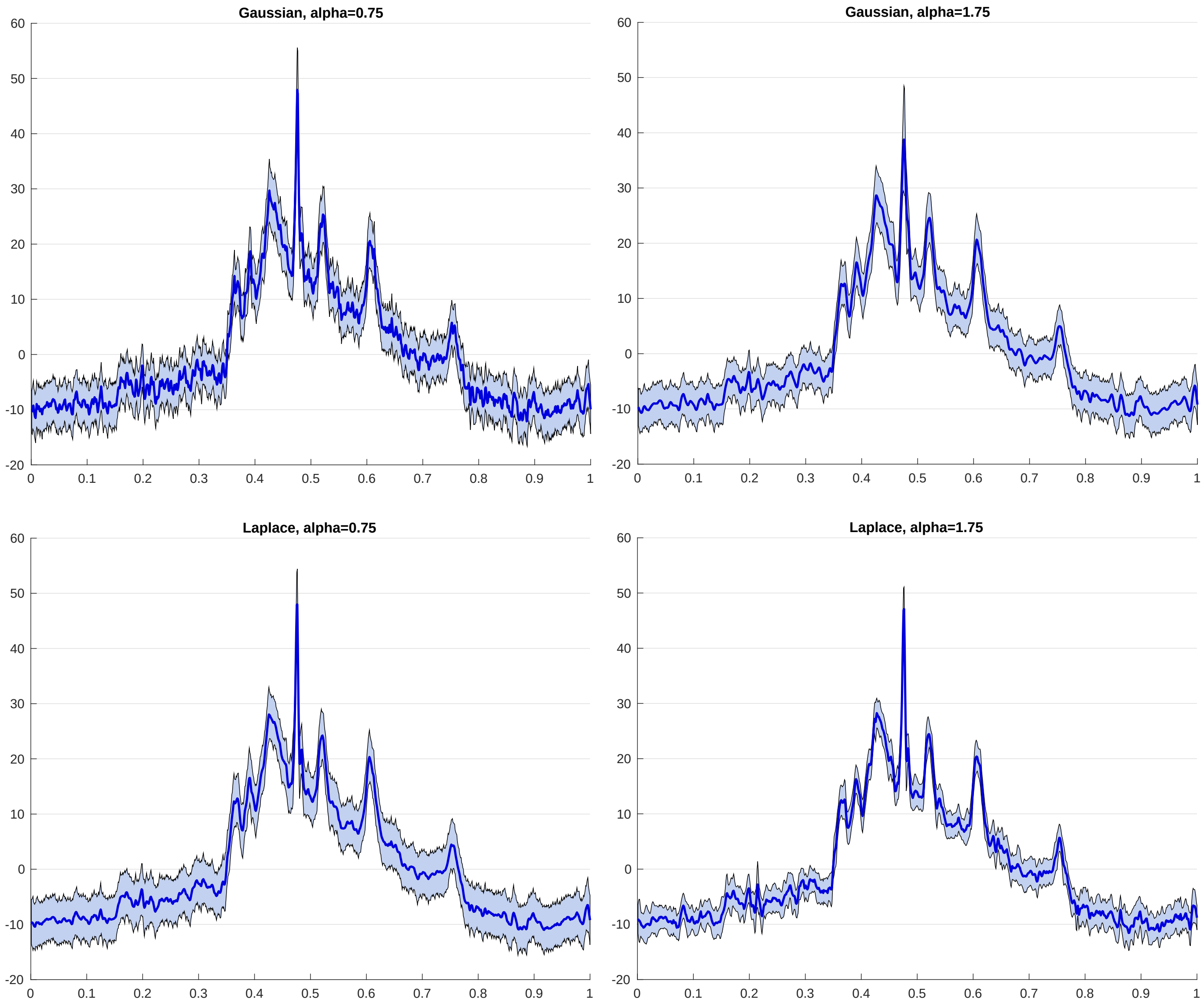
S. Agapiou, J. Bardsley, O. Papaspiliopoulos, A. Stuart *Analysis of the Gibbs Sampler for Hierarchical Inverse Problems*, SIAM/ASA Journal on UQ, 2014.

Bayesian Denoising of NMR data - Laplace prior

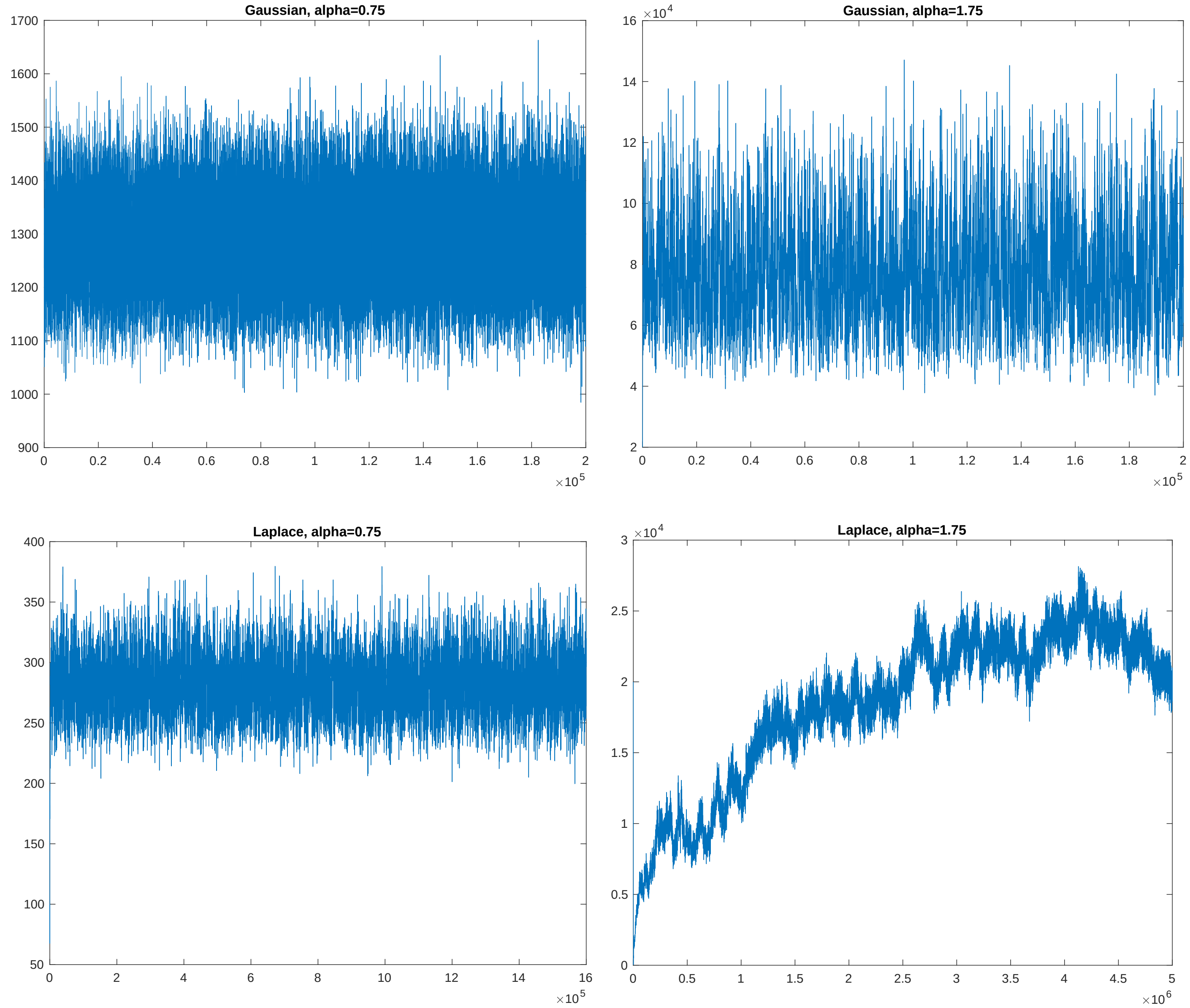
- No conditional conjugacy (only for $\delta | u, y$)
- Need to use Metropolis within Gibbs
- pCN dimension-robust for Gaussian priors
- Again u, τ a-priori strongly dependent
- Use non-centered pCN within Gibbs
  - Write $u = T(\zeta, \tau)$ such that ζ, τ a-priori independent and ζ is Gaussian WN
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

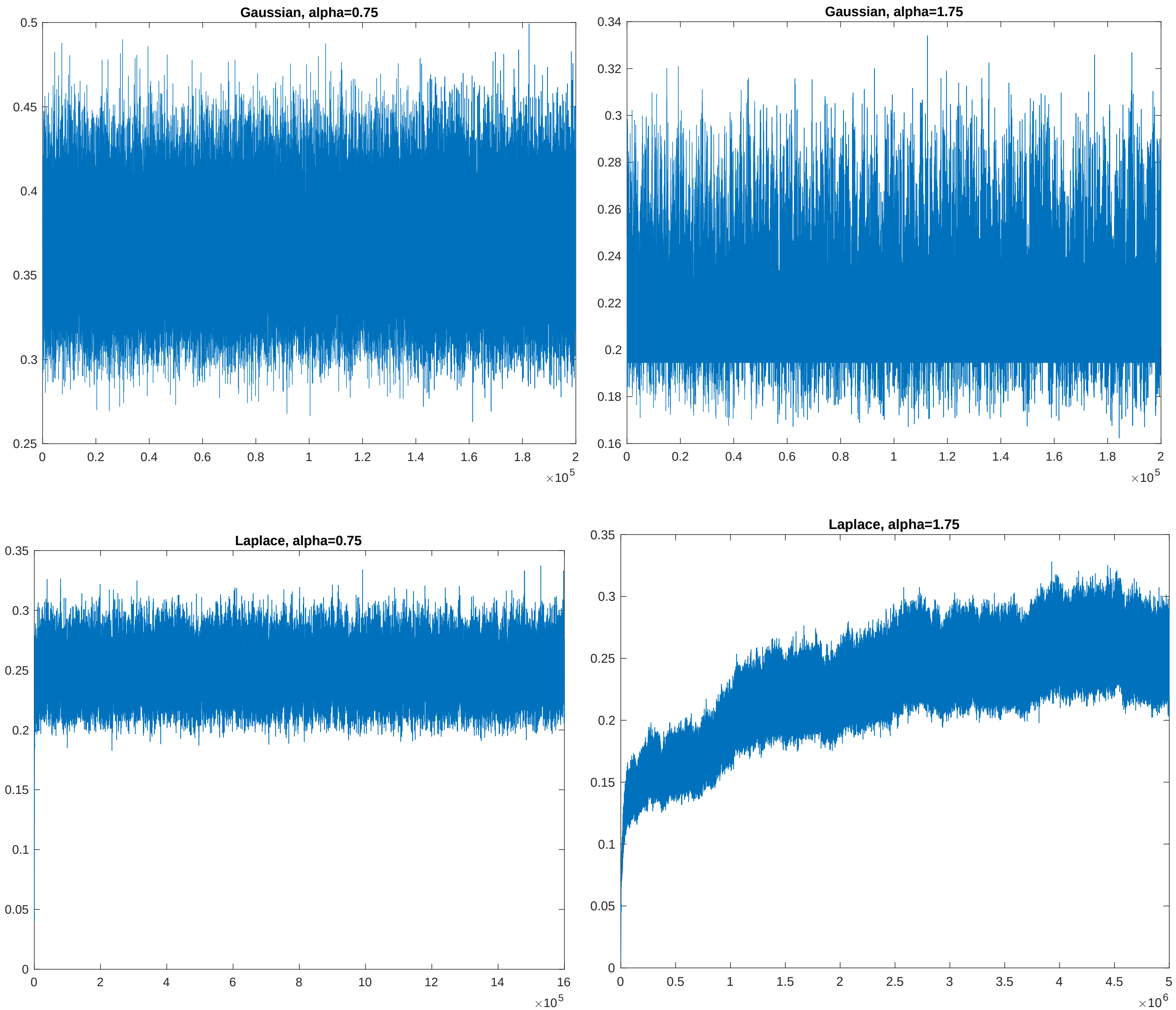
NMR data - Gauss vs Laplace priors



NMR data - Gauss vs Laplace priors - τ-chains



NMR data - Gauss vs Laplace priors - δ-chains



Outline

Summary and open questions

- Laplace priors outperform Gaussian priors over Besov regularity
- Adaptation over Besov spaces (with A. Savva)
- ROC for Bayesian inverse problems with Besov-priors (with S. Wang)
- Sharpness of rates, do we really need scaling and regularity tuning?
- Experimenting with heavier-tailed priors (with I. Castillo)
- For benefit to be realized need better algorithms

Summary and open questions

- Laplace priors outperform Gaussian priors over Besov regularity
- Adaptation over Besov spaces (with A. Savva)
- ROC for Bayesian inverse problems with Besov-priors (with S. Wang)
- Sharpness of rates, do we really need scaling and regularity tuning?
- Experimenting with heavier-tailed priors (with I. Castillo)
- For benefit to be realized need better algorithms

THANK YOU!

http://www.mas.ucy.ac.cy/∼sagapi01/

- S. Agapiou, M. Dashti and T. Helin, *Rates of contraction of posterior distributions based on p-exponential priors*, Bernoulli, 2021.
- S. Agapiou and S. Wang, *Laplace priors and spatial inhomogeneity in Bayesian inverse problems*, arXiv:2112.05679.
- S. Agapiou and A. Savva, *Adaptive rates of contraction based on p-exponential priors*, in preparation.