The horseshoe prior for edge-preserving Bayesian inversion

Yiqiu Dong

DTU Compute

Technical University of Denmark

Joint work with: Per Christian Hansen, Monica Pragliola and Felipe Uribe

THE VELUX FOUNDATIONS
VILLUM FONDEN ✕ VELUX FONDEN

CUQI

Linear inverse problems

Consider a linear inverse problem in the form:

$$\mathbf{y} = A\bar{\mathbf{x}} + \mathbf{e}.$$

- Noisy observational data: $\mathbf{y} \in \mathbb{R}^m$.
- Forward operator: $A \in \mathbb{R}^{m \times n}$.
- Unknown variables: $\mathbf{x} \in \mathbb{R}^n$ with the ground-truth $\bar{\mathbf{x}}$.
- Measurement noise: $\mathbf{e} \in \mathbb{R}^m$ follows $\mathcal{N}(\mathbf{0}, \sigma^2 \mathbf{I}_m)$.

Goal: Find a good approximation to $\bar{\mathbf{x}}$, which is robust with respect to the measurement noise.

Bayesian inverse problems

- The objective in a Bayesian inverse problem is to find or characterize the *posterior* probability density, defined through Bayes' Theorem as

$$\pi_{\text{pos}}(\mathbf{x}|\mathbf{y}) = \frac{1}{Z}\,\pi_{\text{lk}}(\mathbf{y}|\mathbf{x})\,\pi_{\text{pr}}(\mathbf{x}).$$

Bayesian inverse problems

- The objective in a Bayesian inverse problem is to find or characterize the *posterior* probability density, defined through Bayes' Theorem as

$$\pi_{\mathrm{pos}}\,(\mathbf{x}|\mathbf{y}) = \frac{1}{Z}\,\pi_{\mathrm{lk}}(\mathbf{y}|\mathbf{x})\,\pi_{\mathrm{pr}}(\mathbf{x}).$$

- The *likelihood* density follows from the noise assumption:

$$\pi_{\mathrm{lk}}(\mathbf{y}|\mathbf{x}) = \frac{1}{(2\pi)^{m/2}\sigma^m}\exp\left(-\frac{1}{2\sigma^2}\,\|\mathbf{y} - A\mathbf{x}\|_2^2\right).$$

Bayesian inverse problems

- The objective in a Bayesian inverse problem is to find or characterize the *posterior* probability density, defined through Bayes' Theorem as

$$\pi_{\mathrm{pos}}\left(\mathbf{x}|\mathbf{y}\right) = \frac{1}{Z}\,\pi_{\mathrm{lk}}(\mathbf{y}|\mathbf{x})\,\pi_{\mathrm{pr}}(\mathbf{x}).$$

- The *likelihood* density follows from the noise assumption:

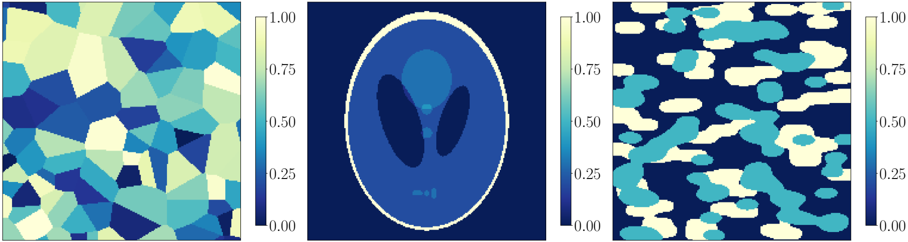
$$\pi_{\mathrm{lk}}(\mathbf{y}|\mathbf{x}) = \frac{1}{(2\pi)^{m/2}\sigma^m}\exp\left(-\frac{1}{2\sigma^2}\left\|\mathbf{y} - A\mathbf{x}\right\|_2^2\right).$$

- The *prior* density can be hierarchical:

$$\pi_{\mathrm{pr}}(\mathbf{x}, \tau, \boldsymbol{\theta}) = \pi_{\mathrm{pr}}(\mathbf{x}|\tau, \boldsymbol{\theta})\pi_{\mathrm{hpr}}(\tau)\pi_{\mathrm{hpr}}(\boldsymbol{\theta}),$$

Applications that require edge-preserving prior

A few examples (crystallography, medical imaging, geophysics):



Prior models for edge preservation

- Heavy-tailed Markov random fields: increase the probability of large jump events by imposing heavy-tailed distributions on the increments. Some examples: Laplace[1] and Cauchy Markov random fields[2].

[1] J. M. Bardsley. "Laplace-distributed increments, the Laplace prior, and edge-preserving regularization". In: *Journal of Inverse and Ill-Posed Problems* 20.3 (2012), pp. 271–285.

[2] M. Markkanen et al. "Cauchy difference priors for edge-preserving Bayesian inversion". In: *Journal of Inverse and Ill-posed Problems* 27.2 (2019), pp. 225–240.

Prior models for edge preservation

- Heavy-tailed Markov random fields: increase the probability of large jump events by imposing heavy-tailed distributions on the increments. Some examples: Laplace and Cauchy Markov random fields .

- Random fields with jump discontinuities: include Besov space priors, Gaussian process, level-set priors[1], etc.

[1] M. M. Dunlop et al. "Hierarchical Bayesian level set inversion". In: *Statistics and Computing* 27 (2017), pp. 1555–1584.

Prior models for edge preservation

- Heavy-tailed Markov random fields: increase the probability of large jump events by imposing heavy-tailed distributions on the increments. Some examples: Laplace and Cauchy Markov random fields .
- Random fields with jump discontinuities: include Besov space priors, Gaussian process, level-set priors , etc.
- Machine learning-based models: plug-and-play priors and Bayesian neural nets with heavy-tailed weights which promotes edge-preservation[1].

[1] C. Li et al. "Bayesian neural network priors for edge-preserving inversion". In: *Inverse Problems and Imaging* 0 (2022), pp. 1–26.

Prior models for edge preservation

- Heavy-tailed Markov random fields: increase the probability of large jump events by imposing heavy-tailed distributions on the increments. Some examples: Laplace and Cauchy Markov random fields .

- Random fields with jump discontinuities: include Besov space priors, Gaussian process, level-set priors , etc.

- Machine learning-based models: plug-and-play priors and Bayesian neural nets with heavy-tailed weights which promotes edge-preservation .

- Shrinkage priors: are popular in sparse statistics. These models are hierarchical by nature and include for instance, elastic net, spike-slab, Horseshoe, discrete Gaussian mixtures, along with many others[1].

[1] N. G. Polson and V. Sokolov. “Bayesian regularization: from Tikhonov to horseshoe”. In: *WIREs Computational Statistics* 11.4 (2019), e1463.

Horseshoe prior

- The standard horseshoe model[2] imposes a conditionally Gaussian prior to $\mathbf{x}$:

$$\pi_{\text{pr}}(\mathbf{x}|\tau, \boldsymbol{\theta}) = \frac{1}{(2\pi)^{\frac{n}{2}} \det(\Sigma_{\tau,\boldsymbol{\theta}})^{\frac{1}{2}}} \exp\left(-\frac{1}{2}\mathbf{x}^T \Sigma_{\tau,\boldsymbol{\theta}}^{-1} \mathbf{x}\right),$$

where $\Sigma_{\tau,\boldsymbol{\theta}} = \tau^2 \text{diag}(\theta_1^2, \dots, \theta_n^2) \in \mathbb{R}^{n\times n}$ is a prior covariance matrix depending on hyperparameters $\tau \in \mathbb{R}_{>0}$ (global) and $\boldsymbol{\theta} = [\theta_1, \dots, \theta_n]^T \in \mathbb{R}^n_{>0}$ (local).

[2] C. M. Carvalho et al. "The horseshoe estimator for sparse signals". In: *Biometrika* 97.2 (2010), pp. 465–480.

Horseshoe prior

- The standard horseshoe model[2] imposes a conditionally Gaussian prior to $\mathbf{x}$:

$$\pi_{\mathrm{pr}}(\mathbf{x}|\tau,\boldsymbol{\theta}) = \frac{1}{(2\pi)^{\frac{n}{2}}\det(\Sigma_{\tau,\boldsymbol{\theta}})^{\frac{1}{2}}}\exp\left(-\frac{1}{2}\mathbf{x}^T\Sigma_{\tau,\boldsymbol{\theta}}^{-1}\mathbf{x}\right),$$

  where $\Sigma_{\tau,\boldsymbol{\theta}} = \tau^2\mathrm{diag}(\theta_1^2,\ldots,\theta_n^2) \in \mathbb{R}^{n\times n}$ is a prior covariance matrix depending on hyperparameters $\tau \in \mathbb{R}_{>0}$ (global) and $\boldsymbol{\theta} = [\theta_1,\ldots,\theta_n]^T \in \mathbb{R}^n_{>0}$ (local).

- This hierarchical prior imposes half-Cauchy hyperpriors on the hyperparameters:

$$\pi_{\mathrm{hpr}}(\tau) \propto \frac{1}{\tau_0^2+\tau^2} \quad \text{and} \quad \pi_{\mathrm{hpr}}(\boldsymbol{\theta}) \propto \prod_{i=1}^{n}\frac{1}{1+\theta_i^2} \qquad \text{with } \tau,\theta_i > 0,$$

  where τ_0 is a scale parameter.

---

[2] C. M. Carvalho et al. “The horseshoe estimator for sparse signals”. In: *Biometrika* 97.2 (2010), pp. 465–480.

Horseshoe prior

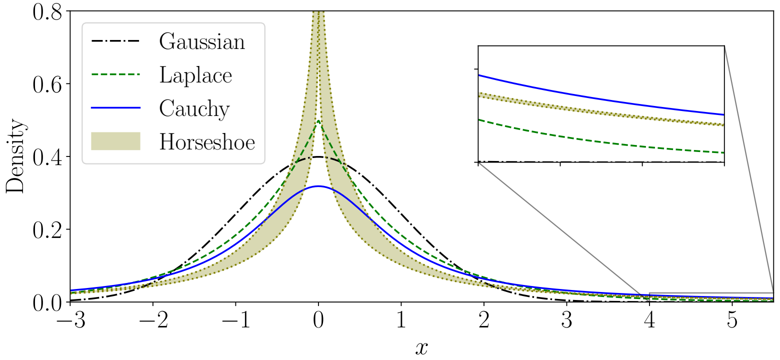


Figure: Comparison of the horseshoe prior bounds with other probability densities. The zoom-in highlights the distributions at the tails.

No closed form for the horseshoe prior, but upper and lower bounds:

$$\frac{1}{2\sqrt{2\pi^3}} \log\left(1 + \frac{4}{x^2}\right) \leq \pi_{\mathrm{pr}}(x) \leq \frac{1}{\sqrt{2\pi^3}} \log\left(1 + \frac{2}{x^2}\right).$$

Shrinkage coefficients

Set $\tau = 1$. The shrinkage coefficients are defined according to the covariance matrix:

$$\phi_i = \frac{1}{1+\theta_i^2} \in [0,1], \quad i = 1,\ldots,n.$$

Shrinkage coefficients

Set $\tau = 1$. The shrinkage coefficients are defined according to the covariance matrix:

$$\phi_i = \frac{1}{1+\theta_i^2} \in [0,1]\,, \quad i = 1,\dots,n\,.$$

- As $\phi_i \to 0$, there is no shrinkage and x_i is a non-vanishing component.
- As $\phi_i \to 1$, the shrinkage occurs.

Shrinkage coefficients

Set $\tau = 1$. The shrinkage coefficients are defined according to the covariance matrix:

$$\phi_i = \frac{1}{1+\theta_i^2} \in [0,1]\,, \quad i = 1, \ldots, n\,.$$

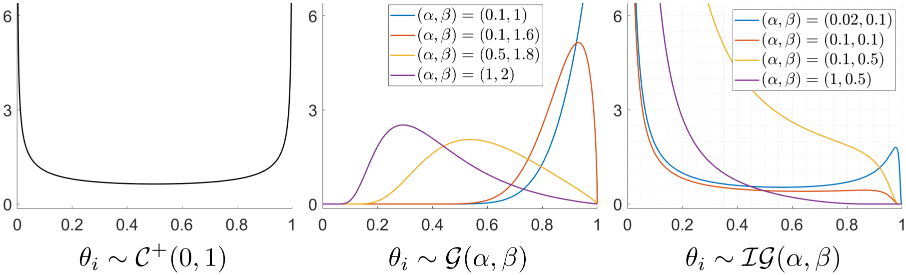
- As $\phi_i \to 0$, there is no shrinkage and x_i is a non-vanishing component.
- As $\phi_i \to 1$, the shrinkage occurs.

Since θ_i follows a half Cauchy distribution with parameter $(0, 1)$, we can derive the pdf of ϕ_i:

$$\pi(\phi_i) = \frac{1}{\pi} \frac{1}{\sqrt{\phi_i}} \frac{1}{\sqrt{1-\phi_i}}\,,$$

that is, ϕ_i follows a Beta distribution with a shape parameter equal to $1/2$.

Shrinkage coefficients



Hierarchical posterior density

$$\pi_{\text{pos}}\left(\mathbf{x}, \tau, \boldsymbol{\theta} | \mathbf{y}\right) \propto \pi_{\text{lk}}(\mathbf{y}|\mathbf{x})\, \pi_{\text{pr}}(\mathbf{x}, \tau, \boldsymbol{\theta}).$$

- The *likelihood* density follows from the noise assumption:

$$\pi_{\text{lk}}(\mathbf{y}|\mathbf{x}) \propto \exp\left(-\frac{1}{2\sigma^2}\left\|\mathbf{y} - A\mathbf{x}\right\|_2^2\right).$$

- The *prior* density is hierarchical horseshoe prior on $D\mathbf{x}$:

$$\pi_{\text{pr}}(\mathbf{x}, \tau, \boldsymbol{\theta}) = \pi_{\text{pr}}(\mathbf{x}|\tau, \boldsymbol{\theta})\pi_{\text{hpr}}(\tau)\pi_{\text{hpr}}(\boldsymbol{\theta}),$$

where we have

- $\pi_{\text{pr}}(\mathbf{x}|\tau, \boldsymbol{\theta}) \propto \dfrac{1}{\det(\Sigma_{\tau,\boldsymbol{\theta}})^{\frac{1}{2}}} \exp\left(-\frac{1}{2}(D\mathbf{x})^T \Sigma_{\tau,\boldsymbol{\theta}}^{-1}(D\mathbf{x})\right)$,
- $\pi_{\text{hpr}}(\tau) \propto \dfrac{1}{\tau_0^2 + \tau^2}$ with $\tau > 0$,
- $\pi_{\text{hpr}}(\boldsymbol{\theta}) \propto \prod_{i=1}^{n} \dfrac{1}{1+\theta_i^2}$ with $\boldsymbol{\theta} > 0$.

Hierarchical posterior density

$$\pi_{\text{pos}}\left(\mathbf{x},\tau,\boldsymbol{\theta}|\mathbf{y}\right) \propto \frac{1}{\det(\Sigma_{\tau,\boldsymbol{\theta}})^{\frac{1}{2}}} \exp\left(-\frac{1}{2\sigma^2}\left\|\mathbf{y}-A\mathbf{x}\right\|_2^2 - \frac{1}{2}(D\mathbf{x})^T\Sigma_{\tau,\boldsymbol{\theta}}^{-1}(D\mathbf{x})\right)$$

$$\times \frac{1}{\tau_0^2+\tau^2}\prod_{i=1}^{n}\frac{1}{1+\theta_i^2} \qquad \text{with} \quad \tau,\boldsymbol{\theta}>0.$$

The main challenges to explore this posterior:

- the dimension of the parameter space is increased;
- the hyperparameters are endowed with heavy-tailed distributions.

MAP estimate

A commonly used point estimate for the posterior density is the maximum a posteriori (MAP) estimate, where one sets the mode of the posterior as the single point representative of the whole density function:

$$\{\mathbf{x}^*, \tau^*, \boldsymbol{\theta}^*\} \in \arg\max_{\mathbf{x},\tau,\boldsymbol{\theta}} \pi_{\text{pos}}\left(\mathbf{x}, \tau, \boldsymbol{\theta} | \mathbf{y}\right) = \arg\min_{\mathbf{x},\tau,\boldsymbol{\theta}} -\ln \pi_{\text{pos}}\left(\mathbf{x}, \tau, \boldsymbol{\theta} | \mathbf{y}\right).$$

MAP estimate

A commonly used point estimate for the posterior density is the maximum a posteriori (MAP) estimate, where one sets the mode of the posterior as the single point representative of the whole density function:

$$\min_{\mathbf{x},\tau>0,\boldsymbol{\theta}>0} \quad \mathcal{J}(\mathbf{x},\tau,\boldsymbol{\theta}) := \frac{1}{2\sigma^2}\|\mathbf{y}-A\mathbf{x}\|_2^2 + \frac{1}{2}\|\Sigma_{\tau,\boldsymbol{\theta}}^{-\frac{1}{2}}D\mathbf{x}\|_2^2 + \sum_{i=1}^{n}\ln\theta_i + \sum_{i=1}^{n}\ln(1+\theta_i^2) + n\ln\tau + \ln(\tau_0^2+\tau^2).$$

- $\mathcal{J}$ is quadratic with respect to $\mathbf{x}$.
- $\mathcal{J}$ is non-convex with respect to τ and $\boldsymbol{\theta}$. But the global minimizers of the τ- and $\boldsymbol{\theta}$-subproblems have closed-form.
- $\mathcal{J}$ is non-convex with respect to $(\mathbf{x},\tau,\boldsymbol{\theta})$.
- We can prove that the alternating minimization algorithm converges to a stationary point of $\mathcal{J}$.

CT reconstruction

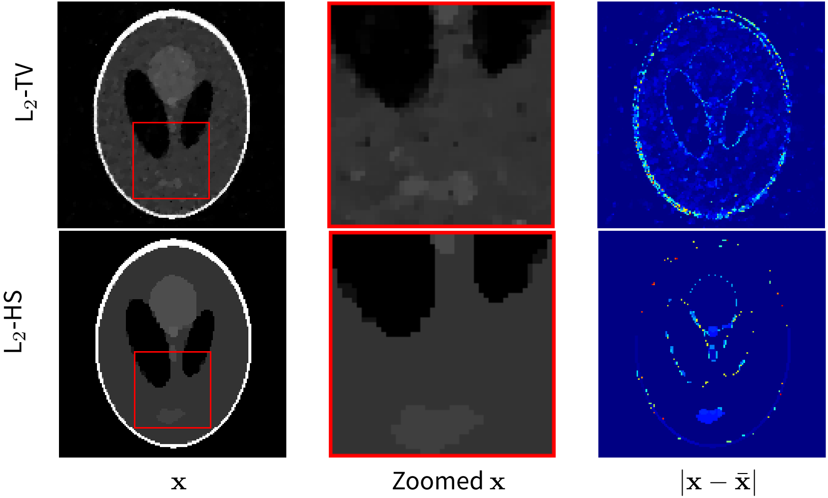


Figure: CT geometry: 45 equidistance angles, 170 detector pixels and 150-by-150 reconstruction resolutions. SNR and SSIM values are: (16.4663, 0.9067) for L_2-TV and (22.8216, 0.9797) for L_2-HS.

Explore the posterior by sampling

$$\pi_{\mathrm{pos}}\left(\mathbf{x}, \tau^2, \boldsymbol{\theta}^2 | \mathbf{y}\right) \propto \pi_{\mathrm{lk}}(\mathbf{y}|\mathbf{x})\, \pi_{\mathrm{pr}}(\mathbf{x}, \tau^2, \boldsymbol{\theta}^2)$$
$$\propto \pi_{\mathrm{lk}}(\mathbf{y}|\mathbf{x})\, \pi_{\mathrm{pr}}(\mathbf{x}|\tau^2, \boldsymbol{\theta}^2)\, \pi_{\mathrm{hpr}}(\tau^2)\, \pi_{\mathrm{hpr}}(\boldsymbol{\theta}^2)$$

We can use Gibbs sampling method to characterize the posterior.

1. Sample $\pi(\mathbf{x}|\mathbf{y}, \tau, \boldsymbol{\theta}) \propto \pi_{\mathrm{lk}}(\mathbf{y}|\mathbf{x})\, \pi_{\mathrm{pr}}(\mathbf{x}|\tau, \boldsymbol{\theta})$;
2. Sample $\pi(\tau|\mathbf{x}, \boldsymbol{\theta}) \propto \pi_{\mathrm{pr}}(\mathbf{x}|\tau, \boldsymbol{\theta})\pi_{\mathrm{hpr}}(\tau)$;
3. Sample $\pi(\boldsymbol{\theta}|\mathbf{x}, \tau) \propto \pi_{\mathrm{pr}}(\mathbf{x}|\tau, \boldsymbol{\theta})\pi_{\mathrm{hpr}}(\boldsymbol{\theta})$.

Explore the posterior by sampling

$$\pi_{\text{pos}}\left(\mathbf{x},\tau^2,\boldsymbol{\theta}^2|\mathbf{y}\right) \propto \pi_{\text{lk}}(\mathbf{y}|\mathbf{x})\,\pi_{\text{pr}}(\mathbf{x},\tau^2,\boldsymbol{\theta}^2)$$
$$\propto \pi_{\text{lk}}(\mathbf{y}|\mathbf{x})\,\pi_{\text{pr}}(\mathbf{x}|\tau^2,\boldsymbol{\theta}^2)\,\pi_{\text{hpr}}(\tau^2)\,\pi_{\text{hpr}}(\boldsymbol{\theta}^2)$$

We can use Gibbs sampling method to characterize the posterior.

1. Sample $\pi(\mathbf{x}|\mathbf{y},\tau,\boldsymbol{\theta}) \propto \pi_{\text{lk}}(\mathbf{y}|\mathbf{x})\,\pi_{\text{pr}}(\mathbf{x}|\tau,\boldsymbol{\theta})$;
2. Sample $\pi(\tau|\mathbf{x},\boldsymbol{\theta}) \propto \pi_{\text{pr}}(\mathbf{x}|\tau,\boldsymbol{\theta})\pi_{\text{hpr}}(\tau)$;
3. Sample $\pi(\boldsymbol{\theta}|\mathbf{x},\tau) \propto \pi_{\text{pr}}(\mathbf{x}|\tau,\boldsymbol{\theta})\pi_{\text{hpr}}(\boldsymbol{\theta})$.

Main challenge is to sample $\pi(\tau|\mathbf{x},\boldsymbol{\theta})$ and $\pi(\boldsymbol{\theta}|\mathbf{x},\tau)$.

Extended horseshoe prior

DTU

Scale mixture representation of half-Student's t-distribution

If A and B are random variables such that

$$(A^2|B) \sim \mathrm{IG}\left(\frac{\nu}{2}, \frac{\nu}{B}\right) \text{ and } B \sim \mathrm{IG}\left(\frac{1}{2}, \frac{1}{c^2}\right),$$

then $A \sim \mathrm{t}^+(\nu, 0, c)$.

- $\mathrm{IG}(\cdot,\cdot)$ denotes the inverse Gamma distribution depending on shape and scale parameters;
- $\mathrm{t}^+(\cdot,\cdot,\cdot)$ is the half-Student's t-distribution depending on degrees of freedom, location and scale parameters;
- $\mathrm{t}^+(\nu, 0, c)$ with $\nu = 1$ is identical to a half-Cauchy distribution.

M. P. Wand et al. "Mean field variational Bayes for elaborate distributions". In: *Bayesian Analysis* 6.4 (2011), pp. 847–900.

Extended horseshoe prior

- Extended hierarchical horseshoe prior is

$$\pi_{\mathrm{pr}}(\mathbf{x}|\tau^2, \boldsymbol{\theta}^2) = \mathcal{N}\left(\mathbf{0}, (D^T \boldsymbol{\Sigma}_{\tau,\boldsymbol{\theta}}^{-1} D)^{-1}\right),$$

$$\pi_{\mathrm{hpr}}(\tau^2|\gamma) = \mathrm{IG}\left(\frac{1}{2}, \frac{1}{\gamma}\right), \qquad \pi_{\mathrm{hpr}}(\gamma) = \mathrm{IG}\left(\frac{1}{2}, \frac{1}{\tau_0^2}\right),$$

$$\pi_{\mathrm{hpr}}(\theta_i^2|\xi_i) = \mathrm{IG}\left(\frac{1}{2}, \frac{1}{\xi_i}\right), \qquad \pi_{\mathrm{hpr}}(\xi_i) = \mathrm{IG}\left(\frac{1}{2}, 1\right).$$

Extended horseshoe prior

- Extended hierarchical horseshoe prior is

$$\pi_{\mathrm{pr}}(\mathbf{x}|\tau^2,\boldsymbol{\theta}^2) = \mathcal{N}\left(\mathbf{0}, (D^T\boldsymbol{\Sigma}_{\tau,\boldsymbol{\theta}}^{-1}D)^{-1}\right),$$

$$\pi_{\mathrm{hpr}}(\tau^2|\gamma) = \mathrm{IG}\left(\frac{1}{2},\frac{1}{\gamma}\right), \qquad \pi_{\mathrm{hpr}}(\gamma) = \mathrm{IG}\left(\frac{1}{2},\frac{1}{\tau_0^2}\right),$$

$$\pi_{\mathrm{hpr}}(\theta_i^2|\xi_i) = \mathrm{IG}\left(\frac{1}{2},\frac{1}{\xi_i}\right), \qquad \pi_{\mathrm{hpr}}(\xi_i) = \mathrm{IG}\left(\frac{1}{2},1\right).$$

- The new posterior density is

$$\pi_{\mathrm{pos}}(\mathbf{x},\tau^2,\boldsymbol{\theta}^2,\gamma,\boldsymbol{\xi}|\mathbf{y}) \propto \pi_{\mathrm{lk}}(\mathbf{y}|\mathbf{x})\,\pi_{\mathrm{pr}}(\mathbf{x}|\tau^2,\boldsymbol{\theta}^2)\,\pi_{\mathrm{hpr}}(\tau^2|\gamma)\,\pi_{\mathrm{hpr}}(\boldsymbol{\theta}^2|\boldsymbol{\xi})\,\pi_{\mathrm{hpr}}(\gamma)\,\pi_{\mathrm{hpr}}(\boldsymbol{\xi}).$$

Gibbs sampler

$$\pi_1(\mathbf{x}|\mathbf{y}, \tau^2, \boldsymbol{\theta}^2) \propto \pi_{\mathrm{lk}}(\mathbf{y}|\mathbf{x})\, \pi_{\mathrm{pr}}(\mathbf{x}|\tau^2, \boldsymbol{\theta}^2)$$

$$\pi_2(\tau^2|\mathbf{x}, \boldsymbol{\theta}^2, \gamma) \propto \pi_{\mathrm{pr}}(\mathbf{x}|\tau^2, \boldsymbol{\theta}^2)\pi_{\mathrm{hpr}}(\tau^2|\gamma)$$

$$\pi_3(\boldsymbol{\theta}^2|\mathbf{x}, \tau^2, \boldsymbol{\xi}) \propto \pi_{\mathrm{pr}}(\mathbf{x}|\tau^2, \boldsymbol{\theta}^2)\pi_{\mathrm{hpr}}(\boldsymbol{\theta}^2|\boldsymbol{\xi})$$

$$\pi_4(\gamma|\tau^2) \propto \pi_{\mathrm{hpr}}(\tau^2|\gamma)\pi_{\mathrm{hpr}}(\gamma),$$

$$\pi_5(\boldsymbol{\xi}|\boldsymbol{\theta}^2) \propto \pi_{\mathrm{hpr}}(\boldsymbol{\theta}^2|\boldsymbol{\xi})\pi_{\mathrm{hpr}}(\boldsymbol{\xi}).$$

Gibbs sampler

$$\pi_1(\mathbf{x}|\mathbf{y},\tau^2,\boldsymbol{\theta}^2) \propto \pi_{\mathrm{lk}}(\mathbf{y}|\mathbf{x})\,\pi_{\mathrm{pr}}(\mathbf{x}|\tau^2,\boldsymbol{\theta}^2)$$
$$\pi_2(\tau^2|\mathbf{x},\boldsymbol{\theta}^2,\gamma) \propto \pi_{\mathrm{pr}}(\mathbf{x}|\tau^2,\boldsymbol{\theta}^2)\pi_{\mathrm{hpr}}(\tau^2|\gamma)$$
$$\pi_3(\boldsymbol{\theta}^2|\mathbf{x},\tau^2,\boldsymbol{\xi}) \propto \pi_{\mathrm{pr}}(\mathbf{x}|\tau^2,\boldsymbol{\theta}^2)\pi_{\mathrm{hpr}}(\boldsymbol{\theta}^2|\boldsymbol{\xi})$$
$$\pi_4(\gamma|\tau^2) \propto \pi_{\mathrm{hpr}}(\tau^2|\gamma)\pi_{\mathrm{hpr}}(\gamma),$$
$$\pi_5(\boldsymbol{\xi}|\boldsymbol{\theta}^2) \propto \pi_{\mathrm{hpr}}(\boldsymbol{\theta}^2|\boldsymbol{\xi})\pi_{\mathrm{hpr}}(\boldsymbol{\xi}).$$

- $\pi_1(\mathbf{x}|\mathbf{y},\tau^2,\boldsymbol{\theta}^2)$ follows Gaussian distribution with the mean $\widetilde{\boldsymbol{\mu}}$ and the covariance $\widetilde{\Lambda}$:

$$\widetilde{\Lambda}_{\tau,\boldsymbol{\theta}} = \frac{1}{\sigma^2}A^TA + D^T\Sigma_{\tau,\boldsymbol{\theta}}^{-1}D, \qquad \widetilde{\boldsymbol{\mu}}_{\tau,\boldsymbol{\theta}} = \widetilde{\Lambda}_{\tau,\boldsymbol{\theta}}^{-1}\left(\frac{1}{\sigma^2}A^T\mathbf{y}\right).$$

- The conditional densities on all hyperparameters, i.e., π_2 to π_5, follow inverse Gamma distribution with closed forms.

1D deconvolution

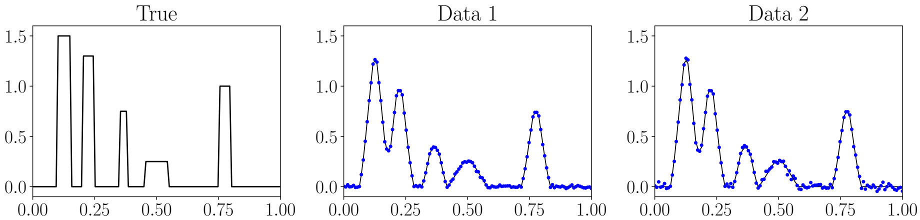


Figure: The original signal and noisy observed data with 2% and 5% noise, respectively.

- A is from Gaussian blurring.
- $\mathbf{x}$ is sparse under the first order derivative, so we use the first order derivative operator as D.
- Samples: $n_s = 2 \times 10^4$, $n_b = 2 \times 10^3$ and $n_t = 40$.

Hyperparameters τ and θ

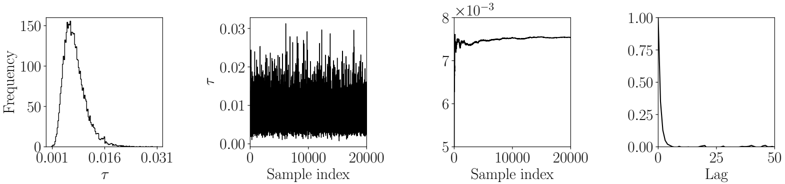


Figure: Posterior of τ with low noise level.

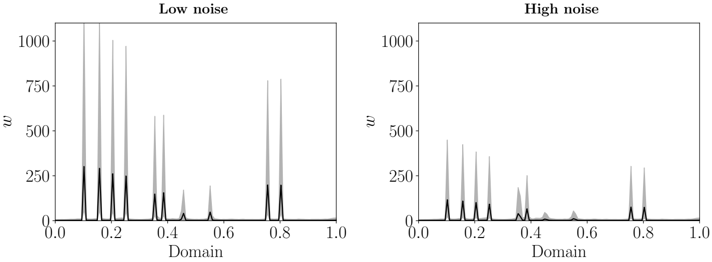


Figure: Posterior mean and 95% CI for $\boldsymbol{\theta}$.

Target parameter x

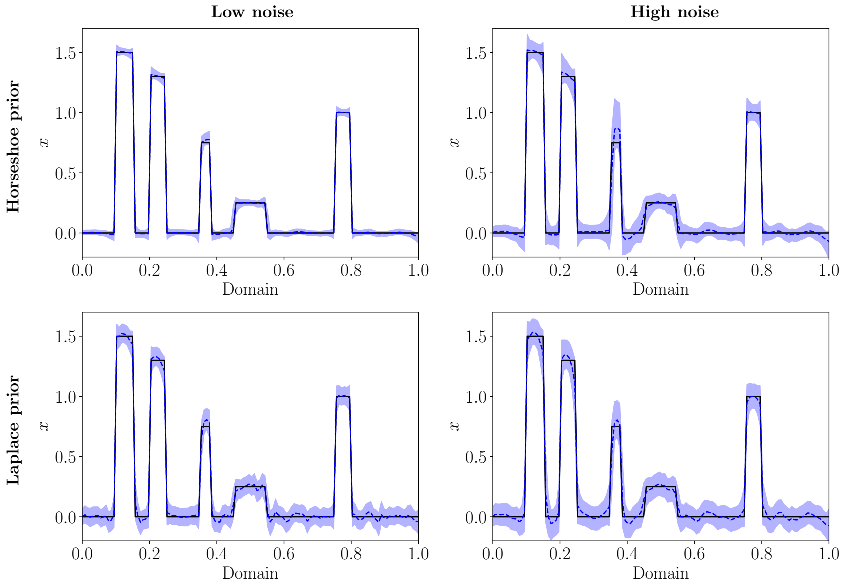


Figure: The relative errors are: (0.0154, 0.0663) for HS and (0.0536, 0.0927) for Laplace.

CT reconstruction

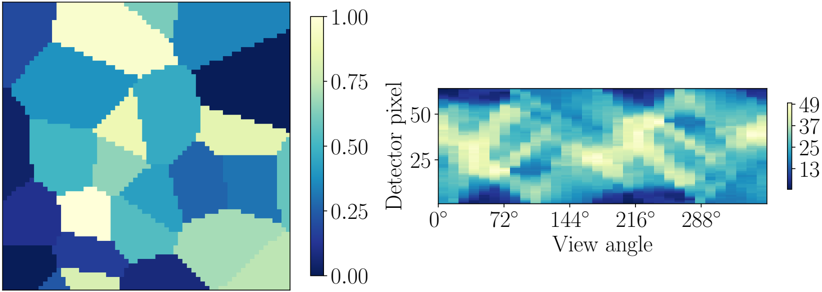
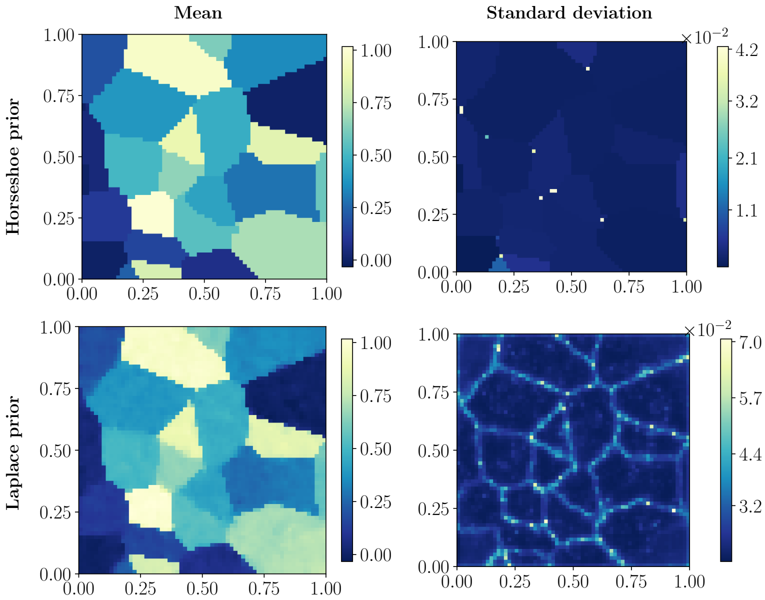


Figure: The ground truth and the sinogram.

- A is from Radon transform with 32 equidistance angles and 64 detector pixels.
- $\mathbf{x}$ is sparse under the first order derivatives, so we use the gradient operator as D.
- The resolution of the reconstruction is 64-by-64. The noise level is 1%.

CT reconstruction: Target parameter $\mathbf{x}$



Thank you!

For your interest:

- **Reference**:
  - Uribe, F., Dong, Y. and Hansen, P. C.: Horseshoe priors for edge-preserving linear Bayesian inversion, *SIAM Journal on Scientific Computing*, Vol. 45(3), pp. 337-365, 2023.
  - Dong, Y. and Pragliola, M.: Inducing sparsity via horseshoe prior in imaging problems, *Inverse Problems*, Vol. 39(7), 074001, 2023.
- **CUQI project**: https://www.compute.dtu.dk/english/cuqi

https://www.icms.org.uk/UQIPI24

UQIPI24: UQ for Inverse Problems and Imaging

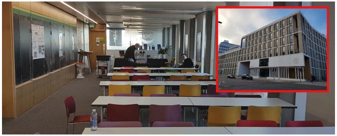
16 - 20 Sep 2024

Apply to attend here

ICMS, Bayes Centre, Edinburgh

Open in google maps

Enquiries



This workshop will bring together specialists in UQ for inverse problems and imaging, and we invite talks related to the development of **theory**, **methodology**, and **software**. We also invite talks about interesting **applications** of UQ in imaging. The goal is to stimulate networking and collaboration between researchers and students in these areas, and to present state-of-the-art research results.

CUQIpy Software Tranining Course

CUQIpy is a python software package for computational uncertainty quantification for inverse problems, developed in the CUQI research project.

Before the main workshop, we give a training course on this software. Participants will learn to use CUQIpy to model statistical inverse problems and perform UQ on them. The course includes hands-on tutorials (bring your laptop!) with examples from image deblurring, X-ray CT, and inverse problems based on partial differential equations. Half of the course is devoted to working on a small use-case with CUQIpy, and participants are encouraged to bring their own case and data.

Programme

Monday morning is devoted to a brief tutorial on Bayesian inference and UQ for inverse problems.

The CUQIpy training course lasts from Monday noon until Tuesday noon; Monday evening is available for the nerds.

The core workshop lasts from Tuesday noon until Friday noon, and consists of plenary talks, contributed talks, and poster sessions. There will be a welcome reception on Tuesday evening, and on Thursday evening there will be a guided tour followed by the workshop dinner.

For those who stay on Friday afternoon, we arrange a social event – perhaps a visit to a whisky or gin distillery.

Plenary Speakers

- Yoann Altmann, Heriot-Watt University
- Tatiana Bubba, University of Bath
- Per Christian Hansen, Technical University of Denmark
- Aku Seppänen, University of Eastern Finland
- Julián Tachella, CNRS and ENS de Lyon
- Faouzi Triki, Grenoble-Alpes University

	Monday	Tuesday	Wednesday	Thursday	Friday
Morning	UQ tutorial	CUQIpy course	Workshop	Workshop	Workshop
After-noon	CUQIpy course	Workshop	Workshop	Workshop	Social event
Evening	CUQIpy course for the nerds	Reception		Guided tour & workshop dinner	