Exploring a posterior with Besov prior through sampling

Andreas Horst

CUQI, Section for Scientific Computing, Technical University of Denmark (DTU)

Joint work with: Babak Afkham, Yiqiu Dong, and Jakob Lemvig.

CUQI VILLUM FONDEN

Outline

- Besov prior
- Sampling with Randomize-Then-Optimize
- Numerical Simulations

Introduction

- Linear inverse problem

$$y = Af + \epsilon, \tag{1}$$

where $y, \epsilon \in \mathbb{R}^m$, $A \in \mathbb{R}^{m \times n}$, and $f \in \mathbb{R}^n$.

Introduction

- Linear inverse problem
$$y = Af + \epsilon, \tag{1}$$
where $y, \epsilon \in \mathbb{R}^m$, $A \in \mathbb{R}^{m \times n}$, and $f \in \mathbb{R}^n$.
- Assume $\epsilon \sim N(0, \sigma I)$ with $\sigma > 0$.

Introduction

- Linear inverse problem

$$y = Af + \epsilon, \tag{1}$$

where $y, \epsilon \in \mathbb{R}^m$, $A \in \mathbb{R}^{m \times n}$, and $f \in \mathbb{R}^n$.

- Assume $\epsilon \sim N(0, \sigma I)$ with $\sigma > 0$.
- Bayesian approach where the posterior distribution is given by Bayes' Rule

$$\pi_{\text{post}}(f|y) \propto \pi_{\text{like}}(y|f)\pi_{\text{pr}}(f). \tag{2}$$

Introduction

- Linear inverse problem

$$y = Af + \epsilon, \tag{1}$$

  where $y, \epsilon \in \mathbb{R}^m$, $A \in \mathbb{R}^{m \times n}$, and $f \in \mathbb{R}^n$.

- Assume $\epsilon \sim N(0, \sigma I)$ with $\sigma > 0$.
- Bayesian approach where the posterior distribution is given by Bayes' Rule

$$\pi_{\text{post}}(f|y) \propto \pi_{\text{like}}(y|f)\pi_{\text{pr}}(f). \tag{2}$$

- Uncertainty Quantification.

Introduction

- Linear inverse problem

$$y = Af + \epsilon, \tag{1}$$

  where $y, \epsilon \in \mathbb{R}^m$, $A \in \mathbb{R}^{m \times n}$, and $f \in \mathbb{R}^n$.
- Assume $\epsilon \sim N(0, \sigma I)$ with $\sigma > 0$.
- Bayesian approach where the posterior distribution is given by Bayes' Rule

$$\pi_{\text{post}}(f|y) \propto \pi_{\text{like}}(y|f)\pi_{\text{pr}}(f). \tag{2}$$

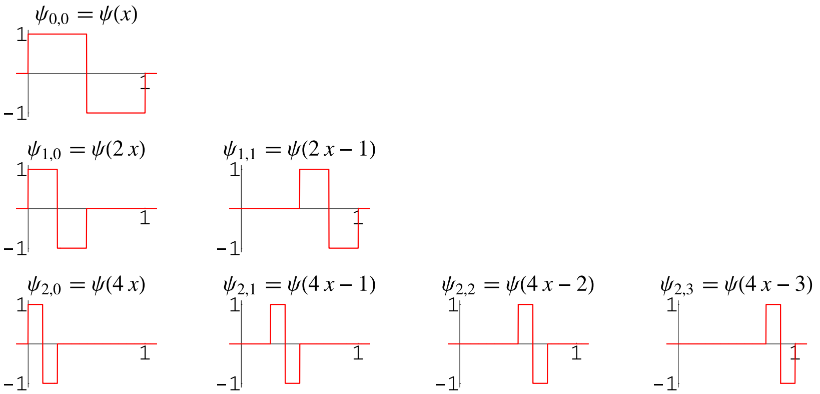
- Uncertainty Quantification.

How does the choice of Besov prior influence the posterior distribution?

Wavelet characterization of Besov spaces I

Let ψ, ϕ be the wavelet and scaling function of a multiresolution analysis in $L^2(\mathbb{R})$ and define basis functions by

$$\psi_{j,k}(x) = \psi(2^j x - k), \quad \phi_k(x) = \phi(x - k), \quad j \in \mathbb{N}, k \in \mathbb{Z}. \tag{3}$$



Wavelet characterization of Besov spaces II

If $\psi, \phi \in \mathcal{C}^r(\mathbb{R})$ with $r > 0$, then Besov spaces $B^s_{p,q}(\mathbb{R})$ is given by the functions

$$f(x) = \sum_{k\in\mathbb{Z}} \langle f, \phi_k \rangle_{L^2(\mathbb{R})} \phi_k(x) + \sum_{j=0}^{\infty} \sum_{k\in\mathbb{Z}} 2^{j/2} \langle f, \psi_{j,k} \rangle_{L^2(\mathbb{R})} \psi_{j,k}(x), \tag{4}$$

with

$$\|f\|_{B^s_{p,q}} = \left(\sum_{k\in\mathbb{Z}} \left| \langle f, \phi_k \rangle_{L^2(\mathbb{R})} \right|^p \right)^{1/p} + \left(\sum_{j=0}^{\infty} 2^{jq(s+1/2-1/p)} \left(\sum_{k\in\mathbb{Z}} \left| \langle f, \psi_{j,k} \rangle_{L^2(\mathbb{R})} \right|^p \right)^{q/p} \right)^{1/q} < \infty, \tag{5}$$

where $1 \leq p, q < \infty$ and $r > s > 0$.

Wavelet characterization of Besov spaces II

If $\psi, \phi \in \mathcal{C}^r(\mathbb{R})$ with $r > 0$, then Besov spaces $B^s_{p,q}(\mathbb{R})$ is given by the functions

$$f(x) = \sum_{k\in\mathbb{Z}} \langle f, \phi_k \rangle_{L^2(\mathbb{R})} \phi_k(x) + \sum_{j=0}^{\infty} \sum_{k\in\mathbb{Z}} 2^{j/2} \langle f, \psi_{j,k} \rangle_{L^2(\mathbb{R})} \psi_{j,k}(x), \tag{4}$$

with

$$\|f\|_{B^s_{p,q}} = \left(\sum_{k\in\mathbb{Z}} \left| \langle f, \phi_k \rangle_{L^2(\mathbb{R})} \right|^{p} \right)^{1/p} + \left(\sum_{j=0}^{\infty} 2^{jq(s+1/2-1/p)} \left(\sum_{k\in\mathbb{Z}} \left| \langle f, \psi_{j,k} \rangle_{L^2(\mathbb{R})} \right|^{p} \right)^{q/p} \right)^{1/q} < \infty, \tag{5}$$

where $1 \leq p, q < \infty$ and $r > s > 0$.

Wavelet characterization of Besov spaces II

If $\psi, \phi \in \mathcal{C}^r(\mathbb{R})$ with $r > 0$, then Besov spaces $B^s_{p,q}(\mathbb{R})$ is given by the functions

$$f(x) = \sum_{k\in\mathbb{Z}} \langle f, \phi_k \rangle_{L^2(\mathbb{R})} \phi_k(x) + \sum_{j=0}^{\infty} \sum_{k\in\mathbb{Z}} 2^{j/2} \langle f, \psi_{j,k} \rangle_{L^2(\mathbb{R})} \psi_{j,k}(x), \tag{4}$$

with

$$\|f\|_{B^s_{p,q}} = \left(\sum_{k\in\mathbb{Z}} \left| \langle f, \phi_k \rangle_{L^2(\mathbb{R})} \right|^p \right)^{1/p} + \left(\sum_{j=0}^{\infty} 2^{jq(s+1/2-1/p)} \left(\sum_{k\in\mathbb{Z}} \left| \langle f, \psi_{j,k} \rangle_{L^2(\mathbb{R})} \right|^p \right)^{q/p} \right)^{1/q} < \infty, \tag{5}$$

where $1 \leq p, q < \infty$ and $r > s > 0$.

Wavelet characterization of Besov spaces II

If $\psi, \phi \in \mathcal{C}^r(\mathbb{R})$ with $r > 0$, then Besov spaces $B^s_{p,q}(\mathbb{R})$ is given by the functions

$$f(x) = \sum_{k\in\mathbb{Z}} \langle f, \phi_k \rangle_{L^2(\mathbb{R})} \phi_k(x) + \sum_{j=0}^{\infty} \sum_{k\in\mathbb{Z}} 2^{j/2} \langle f, \psi_{j,k} \rangle_{L^2(\mathbb{R})} \psi_{j,k}(x), \tag{4}$$

with

$$\|f\|_{B^s_{p,q}} = \left(\sum_{k\in\mathbb{Z}} \left| \langle f, \phi_k \rangle_{L^2(\mathbb{R})} \right|^p \right)^{1/p} + \left(\sum_{j=0}^{\infty} 2^{jq(s+1/2-1/p)} \left(\sum_{k\in\mathbb{Z}} \left| \langle f, \psi_{j,k} \rangle_{L^2(\mathbb{R})} \right|^p \right)^{q/p} \right)^{1/q} < \infty, \tag{5}$$

where $1 \leq p, q < \infty$ and $r > s > 0$.

The Besov parameters

The Besov space parameters p, q, and s determines properties of the Besov space $B^s_{p,q}(\mathbb{R})$.

- The parameter p determines that $B^s_{p,q}(\mathbb{R}) \subset L^p(\mathbb{R})$.

The Besov parameters

The Besov space parameters $p, q,$ and s determines properties of the Besov space $B^s_{p,q}(\mathbb{R})$.

- The parameter p determines that $B^s_{p,q}(\mathbb{R}) \subset L^p(\mathbb{R})$.
- The parameter s determines the order of smoothness in $L^p(\mathbb{R})$.

The Besov parameters

The Besov space parameters p, q, and s determines properties of the Besov space $B^s_{p,q}(\mathbb{R})$.

- The parameter p determines that $B^s_{p,q}(\mathbb{R}) \subset L^p(\mathbb{R})$.
- The parameter s determines the order of smoothness in $L^p(\mathbb{R})$.
- The q parameter is a weak smoothness parameter

$$B^{s_1}_{p,q_1}(\mathbb{R}) \subset B^{s_2}_{p,q_2}(\mathbb{R}), \quad \text{if } s_1 \geq s_2 \text{ and for any } q_1, q_2 \geq 1. \tag{6}$$

The Besov parameters

The Besov space parameters p, q, and s determines properties of the Besov space $B^s_{p,q}(\mathbb{R})$.

- The parameter p determines that $B^s_{p,q}(\mathbb{R}) \subset L^p(\mathbb{R})$.
- The parameter s determines the order of smoothness in $L^p(\mathbb{R})$.
- The q parameter is a weak smoothness parameter

$$B^{s_1}_{p,q_1}(\mathbb{R}) \subset B^{s_2}_{p,q_2}(\mathbb{R}), \quad \text{if } s_1 \geq s_2 \text{ and for any } q_1, q_2 \geq 1. \tag{6}$$

- Besov spaces are generalizations of Sobolev and Hölder spaces

$$B^s_{2,2}(\mathbb{R}) = H^s(\mathbb{R}). \tag{7}$$

$$B^s_{p,p}(\mathbb{R}) = W^{s,p}(\mathbb{R}), \quad s \notin \mathbb{N}. \tag{8}$$

$$B^s_{\infty,\infty}(\mathbb{R}) = \mathcal{C}^s(\mathbb{R}), \quad s \notin \mathbb{N}. \tag{9}$$

The Besov prior

The Besov prior as defined in (Lassas et al. 2009)[1].

- Let $p = q$ and set the domain to be $\mathbb{T}$, that is, we only consider 1-periodic functions on $\mathbb{R}$.

[1] Matti Lassas, Eero Saksman, and Samuli Siltanen. "Discretization-invariant Bayesian inversion and Besov space priors". In: *Inverse Problems and Imaging* 3.1 (2009), pp. 87–122.

The Besov prior

The Besov prior as defined in (Lassas et al. 2009)[1].

- Let $p = q$ and set the domain to be $\mathbb{T}$, that is, we only consider 1-periodic functions on $\mathbb{R}$.
- Choose $\{\xi_{j,k}\}$ to be independent, identical distributed according to the density $\pi(x) \propto \exp(-\frac{1}{p}\,|x|^p)$.

[1]Matti Lassas, Eero Saksman, and Samuli Siltanen. "Discretization-invariant Bayesian inversion and Besov space priors". In: *Inverse Problems and Imaging* 3.1 (2009), pp. 87–122.

The Besov prior

The Besov prior as defined in (Lassas et al. 2009)[1].

- Let $p = q$ and set the domain to be $\mathbb{T}$, that is, we only consider 1-periodic functions on $\mathbb{R}$.
- Choose $\{\xi_{j,k}\}$ to be independent, identical distributed according to the density $\pi(x) \propto \exp(-\frac{1}{p}|x|^p)$.
- Let ϕ and $\{\psi_{j,k}\}$ be the scaling function and wavelets that form a basis for the 1-periodic Besov space $B^s_{p,p}(\mathbb{T})$.

[1] Matti Lassas, Eero Saksman, and Samuli Siltanen. "Discretization-invariant Bayesian inversion and Besov space priors". In: *Inverse Problems and Imaging* 3.1 (2009), pp. 87–122.

The Besov prior

The Besov prior as defined in (Lassas et al. 2009)[1].

- Let $p = q$ and set the domain to be $\mathbb{T}$, that is, we only consider 1-periodic functions on $\mathbb{R}$.
- Choose $\{\xi_{j,k}\}$ to be independent, identical distributed according to the density $\pi(x) \propto \exp(-\frac{1}{p}|x|^p)$.
- Let ϕ and $\{\psi_{j,k}\}$ be the scaling function and wavelets that form a basis for the 1-periodic Besov space $B^s_{p,p}(\mathbb{T})$.

  The random function U given by

$$U(x) = \xi_{-1,0}\phi(x) + \sum_{j=0}^{\infty} \sum_{k=0}^{2^j-1} 2^{-j(s+1/2-1/p)} \xi_{j,k}\psi_{j,k}(x) \tag{10}$$

  takes values in some $B^t_{p,p}(\mathbb{T})$ with $t < s - 1/p$. We say that U is distributed according to a $B^s_{p,p}$ Besov measure.

[1] Matti Lassas, Eero Saksman, and Samuli Siltanen. "Discretization-invariant Bayesian inversion and Besov space priors". In: *Inverse Problems and Imaging* 3.1 (2009), pp. 87–122.

Properties of the Besov prior

- Defines a probability measure on some Besov space $B^t_{p,p}(\mathbb{T})$.

Properties of the Besov prior

- Defines a probability measure on some Besov space $B^t_{p,p}(\mathbb{T})$.
- Can be approximated by finitely many 1D random variables by truncation

$$U(x) \approx \xi_{-1,0}\phi(x) + \sum_{j=0}^{J_{\text{max}}-1} \sum_{k=0}^{2^j-1} 2^{-j(s+1/2-1/p)} \xi_{j,k} \psi_{j,k}(x), \tag{11}$$

for some integer $J_{\text{max}} \geq 1$.

Properties of the Besov prior

- Defines a probability measure on some Besov space $B^t_{p,p}(\mathbb{T})$.
- Can be approximated by finitely many 1D random variables by truncation

$$U(x) \approx \xi_{-1,0}\phi(x) + \sum_{j=0}^{J_{\text{max}}-1} \sum_{k=0}^{2^j-1} 2^{-j(s+1/2-1/p)} \xi_{j,k}\psi_{j,k}(x), \tag{11}$$

  for some integer $J_{\text{max}} \geq 1$.
- Discretization invariant, that is, posterior estimates converges as discretization becomes finer.

Properties of the Besov prior

- Defines a probability measure on some Besov space $B^t_{p,p}(\mathbb{T})$.
- Can be approximated by finitely many 1D random variables by truncation

$$U(x) \approx \xi_{-1,0}\phi(x) + \sum_{j=0}^{J_{\text{max}}-1} \sum_{k=0}^{2^j-1} 2^{-j(s+1/2-1/p)} \xi_{j,k}\psi_{j,k}(x), \tag{11}$$

  for some integer $J_{\text{max}} \geq 1$.
- Discretization invariant, that is, posterior estimates converges as discretization becomes finer.
- Can be chosen to be edge-preserving.

Properties of the Besov prior

- Defines a probability measure on some Besov space $B^t_{p,p}(\mathbb{T})$.
- Can be approximated by finitely many 1D random variables by truncation

$$U(x) \approx \xi_{-1,0}\phi(x) + \sum_{j=0}^{J_{\mathsf{max}}-1} \sum_{k=0}^{2^j-1} 2^{-j(s+1/2-1/p)} \xi_{j,k}\psi_{j,k}(x), \tag{11}$$

  for some integer $J_{\mathsf{max}} \geq 1$.
- Discretization invariant, that is, posterior estimates converges as discretization becomes finer.
- Can be chosen to be edge-preserving.

  We assume that f is samples of a function U that follows a Besov prior, which leads to a posterior on the form

$$\pi_{\mathsf{post}}(f|y) \propto \exp\left(-\frac{1}{2\sigma^2}\|Af - y\|_2^2 - \frac{1}{p}\|Bf\|_p^p\right), \tag{12}$$

  where $Bf \approx \{2^{j(s+1/2-1/p)}\langle U, \psi_{j,k}\rangle_{L^2(\mathbb{T})}\}$.

Sampling approach

Transform the prior from Besov to Gaussian which enables Gaussian samplers.

- Let $\pi_p(x) \propto \exp\left(-\frac{1}{p}\left|x\right|^p\right)$ be a centered p-Gaussian distribution with inverse CDF Φ_p^{-1} and let Φ_{Gauss} be the CDF of a standard Gaussian .

Sampling approach

Transform the prior from Besov to Gaussian which enables Gaussian samplers.

- Let $\pi_p(x) \propto \exp\left(-\frac{1}{p}|x|^p\right)$ be a centered p-Gaussian distribution with inverse CDF Φ_p^{-1} and let Φ_{Gauss} be the CDF of a standard Gaussian .
- Let $g : \mathbb{R}^n \to \mathbb{R}^n$ be the inverse CDF transform defined by

$$g(x) = \left[\Phi_p^{-1}(\Phi_{\text{Gauss}}(x_1)), \ldots, \Phi_p^{-1}(\Phi_{\text{Gauss}}(x_n))\right]^T . \tag{13}$$

Sampling approach

Transform the prior from Besov to Gaussian which enables Gaussian samplers.

- Let $\pi_p(x) \propto \exp\left(-\frac{1}{p}|x|^p\right)$ be a centered p-Gaussian distribution with inverse CDF Φ_p^{-1} and let Φ_{Gauss} be the CDF of a standard Gaussian .
- Let $g : \mathbb{R}^n \to \mathbb{R}^n$ be the inverse CDF transform defined by

$$g(x) = \left[\Phi_p^{-1}(\Phi_{\text{Gauss}}(x_1)), \ldots, \Phi_p^{-1}(\Phi_{\text{Gauss}}(x_n))\right]^T . \quad (13)$$

- The prior transform from standard Gaussian h to Besov f is then

$$Bf = g(h). \quad (14)$$

Sampling approach

Transform the prior from Besov to Gaussian which enables Gaussian samplers.

- Let $\pi_p(x) \propto \exp\left(-\frac{1}{p}|x|^p\right)$ be a centered p-Gaussian distribution with inverse CDF Φ_p^{-1} and let Φ_{Gauss} be the CDF of a standard Gaussian .
- Let $g : \mathbb{R}^n \to \mathbb{R}^n$ be the inverse CDF transform defined by

$$g(x) = \left[\Phi_p^{-1}(\Phi_{\text{Gauss}}(x_1)), \dots, \Phi_p^{-1}(\Phi_{\text{Gauss}}(x_n))\right]^T . \tag{13}$$

- The prior transform from standard Gaussian h to Besov f is then

$$Bf = g(h). \tag{14}$$

  The transformed posterior

$$\pi_{\text{post}}(h|y) \propto \exp\left(-\frac{1}{2\sigma^2}\|AB^{-1}g(h) - y\|_2^2 - \frac{1}{2}\|h\|_2^2\right). \tag{15}$$

Randomize-Then-Optimize sampling I

In (Bardsley et al. 2014)[2] the Randomize-Then-Optimize(RTO) algorithm is presented as a method to sample from a Least-Square target posterior

$$\pi_{\text{target}}(h|y) \propto \exp(-\frac{1}{2}\|F(h)\|_2^2). \tag{16}$$

RTO samples are computed by solving a stochastic optimization problem

$$h^{\text{RTO}} = \arg\min_h \|Q^T\left(F(h) - v\right)\|_2^2, \quad v \sim N(0, I), \tag{17}$$

where Q is an orthogonal matrix.
The samples h^{RTO} comes from the RTO proposal distribution

$$\pi_{\text{RTO}}(h|y) \propto \left|Q^T J_F(h)\right| \exp\left(-\frac{1}{2}\|Q^T F(h)\|_2^2\right). \tag{18}$$

[2] Johnathan M. Bardsley et al. “Randomize-Then-Optimize: A Method for Sampling from Posterior Distributions in Nonlinear Inverse Problems”. In: *SIAM Journal on Scientific Computing* 36.4 (2014), A1895–A1910.

Randomize-Then-Optimize sampling II

Metropolis-Hastings step

$$p = \frac{\pi_{\text{target}}(h^{\text{RTO}}|y)\pi_{\text{RTO}}(h^{\text{target}}|y)}{\pi_{\text{target}}(h^{\text{target}}|y)\pi_{\text{RTO}}(h^{\text{RTO}}|y)}. \tag{19}$$

To increase acceptance rate the orthogonal matrix Q is computed as $J_F(\overline{h}) = QR$ with

$$\overline{h} = \arg\min_h \frac{1}{2}\|F(h)\|_2^2. \tag{20}$$

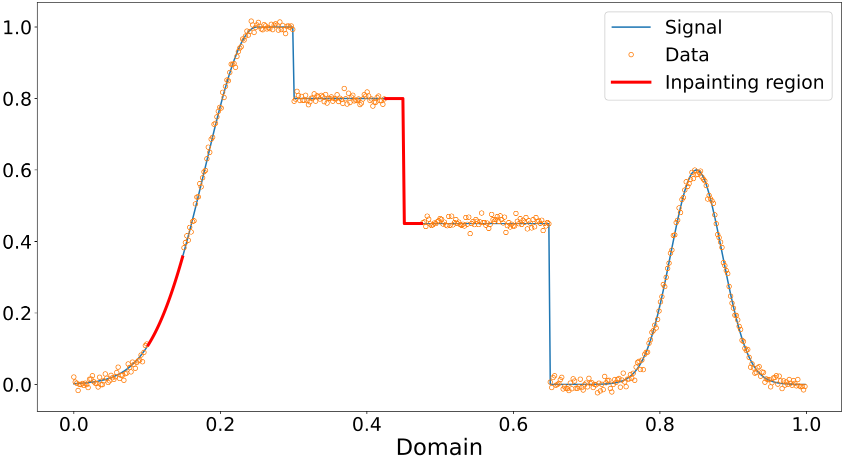
Sampling procedure

1. Use Randomize-Then-Optimize sampling to compute a sample h^{RTO} from π_{RTO}.
2. Use the h^{RTO} sample as proposal for Metropolis-Hastings acceptance-rejection
3. Transform the resulting sample back to the original Besov variable $f^{\mathsf{RTO}} = B^{-1}g(h^{\mathsf{RTO}})$.

This procedure also works when A is non-linear if the map $Q^T F(x) = Q^T \begin{bmatrix} x \\ A\left(B^{-1}g(x)\right) \end{bmatrix}$ is admissible for change of Gaussian variables.

Exploring the posterior

The inpainting problem

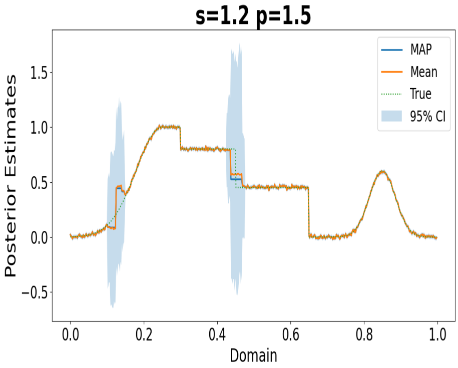


Numerical setup

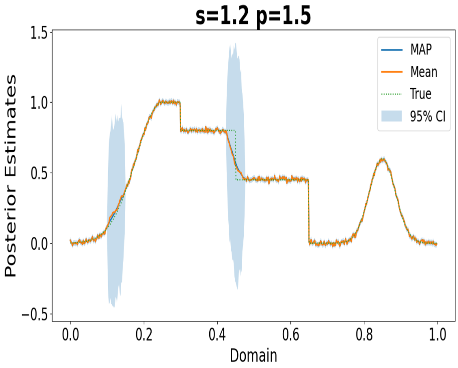
Numerical tests:

- Varying the parameters $s = \{0.8, 1.4, 2.0\}$ and $p = \{1.0, 1.5, 2.0\}$.
- Varying the wavelet between Haar wavelet and db8 wavelet
- For each test we compute the maximum a posteriori(MAP) estimate and 10000 posterior samples. We use the posterior samples to estimate the mean and the 95% credible interval(CI).

Changing the wavelet

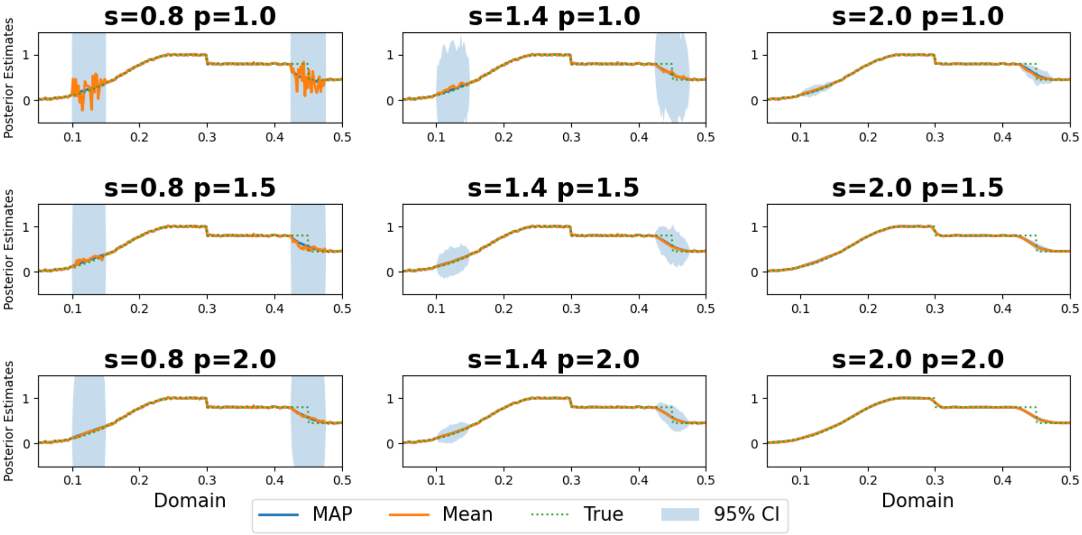


(a) Haar-wavelet result



(b) db8-wavelet result

Inpainting - db8 wavelet - Varying the parameters



Conclusion and Future work

- Mean and MAP are close in most cases
- The parameters s and p makes the posterior more concentrated around the mean.
- The wavelet basis controls the edge-preserving property of the prior.
- Future work: Extend to 2D non-linear problem.

The End

Thank you!
Questions, Comments, and Suggestions.