Edge preserving random tree Besov priors

Hanne Kekkonen

joint work with
Matti Lassas, Eero Saksman and Samuli Siltanen

Delft Institute of Applied Mathematics
Delft University of Technology

September 27, 2022

TU Delft

Outline

1. Bayesian inverse problems
2. Edge preserving Besov priors
3. Examples

Outline

1. Bayesian inverse problems
2. Edge preserving Besov priors
3. Examples

Bayesian approach to inverse problems

We want to recover the unknown f from a noisy measurement M;

$$M = Af + noise,$$

where A is a forward operator that usually causes loss of information.

- Consider observing data M drawn at random from some unknown probability distribution $P^M_{f^\dagger}$, and sample size n.
- Specify a prior distribution Π for the unknown f and assume

$$M \,|\, f \sim P^M_f.$$

- Using Bayes' theorem the prior distribution can be updated to a posterior distribution

$$f \,|\, M \sim \Pi(\cdot \,|\, M).$$

Gaussian priors are often used for inverse problems

- Assume measurement model

$$M = Af + \delta \mathbb{W},$$

  where A is a linear forward operator and $\mathbb{W} \sim \mathcal{N}(0, I)$.

- If we assume $f \sim \mathcal{N}(0, C_f)$ the posterior is also Gaussian and CM coincides with MAP estimate and is given by

$$\widehat{f}(M) = (A^*A + \delta^2 C_f^{-1})^{-1} A^* M.$$

- Standard Gaussian priors are often used in practice due to their fast computational properties.

Many applications require edge preservation

Noisy image

ℓ^2-regularised solution

TV-regularised solution

$$\pi_{TV}(f) \propto \exp\left(\alpha \sum_{i,j} |f_{i+1,j} - f_{i,j}| + |f_{i,j+1} - f_{i,j}|\right)$$

Is total variation prior consistent?

When discretisation gets finer the discrete total variation prior either diverges or the posterior distribution converges to a Gaussian distribution.
$\Rightarrow$ Not edge preserving with fine discretisation, Lassas and Siltanen 2004.

Is total variation prior consistent?

When discretisation gets finer the discrete total variation prior either diverges or the posterior distribution converges to a Gaussian distribution.
$\Rightarrow$ Not edge preserving with fine discretisation, Lassas and Siltanen 2004.

The widely used formal total variation prior

$$\pi_{pr}(f) \underset{formally}{\approx} \exp(-\alpha \|\nabla f\|_{L^1}), \quad f \in L^2.$$

is not known to correspond to any well defined random variable.

Is total variation prior consistent?

When discretisation gets finer the discrete total variation prior either diverges or the posterior distribution converges to a Gaussian distribution.
$\Rightarrow$ Not edge preserving with fine discretisation, Lassas and Siltanen 2004.

The widely used formal total variation prior

$$\pi_{pr}(f) \underset{formally}{\approx} \exp(-\alpha \|\nabla f\|_{L^1}), \quad f \in L^2.$$

is not known to correspond to any well defined random variable.

We want to

- Have similar edge preserving properties than total variation priors.
- Correspond to well defined infinite dimensional random variables.
- Can be approximated by finite dimensional random variables.

Outline

1. Bayesian inverse problems
2. Edge preserving Besov priors
3. Examples

Replacing TV prior by a Besov prior

We can replace the formal prior

$$\pi(f) \underset{\text{formally}}{\propto} \exp\left(-\|\nabla f\|_{L^1}\right)$$

by a well defined Besov prior

$$\pi(f) \underset{\text{formally}}{\propto} \exp\left(-\|\nabla f\|^p_{B^0_{11}}\right),$$

that was first introduced by Lassas, Saksman and Siltanen 2009, and further studied by Dashti, Harris and Stuart 2012.

How to form a random function?

Remember that 1, $\sqrt{2}\sin(kt)$ and $\sqrt{2}\cos(kt)$ form an orthonormal basis for $L^2[-\pi,\pi]$. A periodic signal $u(t)$, $t \in [-\pi,\pi]$, can be written as

$$u(t) = \frac{a_0}{2} + \sum_{k=1}^{\infty} a_k \sin(kt) + b_k \cos(kt).$$

How to form a random function?

Remember that 1, $\sqrt{2}\sin(kt)$ and $\sqrt{2}\cos(kt)$ form an orthonormal basis for $L^2[-\pi,\pi]$. A periodic signal $u(t)$, $t \in [-\pi,\pi]$, can be written as

$$u(t) = \frac{a_0}{2} + \sum_{k=1}^{\infty} a_k \sin(kt) + b_k \cos(kt).$$

Extension of this idea for random functions is given by

$$U(t) = \sum_{k=1}^{\infty} Z_k \psi_k(t),$$

where Z_k's are pairwise uncorrelated random variables and ψ_k is an orthonormal basis on $L^2[-\pi,\pi]$.

Karhunen–Loève expansion

We can construct random draws from a Gaussian measure;

- Let $\{\psi_k, \lambda_k\}$ be an orthonormal set of eigenvectors and eigenvalues for the covariance operator Σ.
- Take $\{\xi_k\}_{k=1}^{\infty}$ to be a sequence of independent random variables with $\xi_k \sim \mathcal{N}(0, 1)$.

Karhunen–Loève expansion

We can construct random draws from a Gaussian measure;

- Let $\{\psi_k, \lambda_k\}$ be an orthonormal set of eigenvectors and eigenvalues for the covariance operator Σ.
- Take $\{\xi_k\}_{k=1}^{\infty}$ to be a sequence of independent random variables with $\xi_k \sim \mathcal{N}(0, 1)$.

Then the random variable U given by the Karhunen–Loève expansion

$$U(t) = \sum_{k=1}^{\infty} \sqrt{\lambda_k} \xi_k \psi_k(t)$$

is distributed according to $\mathcal{N}(0, \Sigma)$.

Karhunen–Loève expansion

We can construct random draws from a Gaussian measure;

- Let $\{\psi_k, \lambda_k\}$ be an orthonormal set of eigenvectors and eigenvalues for the covariance operator Σ.
- Take $\{\xi_k\}_{k=1}^{\infty}$ to be a sequence of independent random variables with $\xi_k \sim \mathcal{N}(0, 1)$.

Then the random variable U given by the Karhunen–Loève expansion

$$U(t) = \sum_{k=1}^{\infty} \sqrt{\lambda_k} \xi_k \psi_k(t)$$

is distributed according to $\mathcal{N}(0, \Sigma)$.

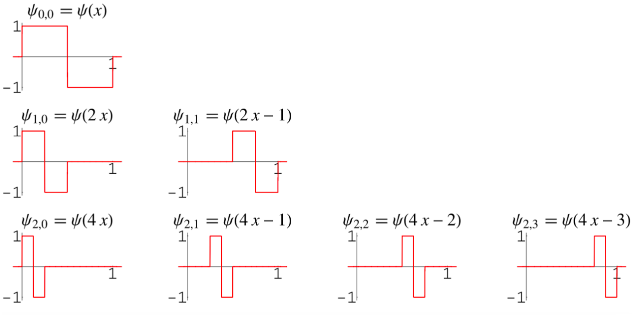
Example: If Σ^{-1} is a Laplace type operator the eigenvalues will grow like k^{-2}.

Wavelet basis

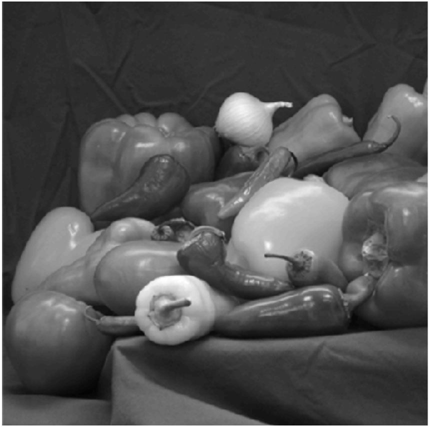
Let Ψ be the mother wavelet suitable for multi-resolution analysis of smoothness C^r and define wavelets

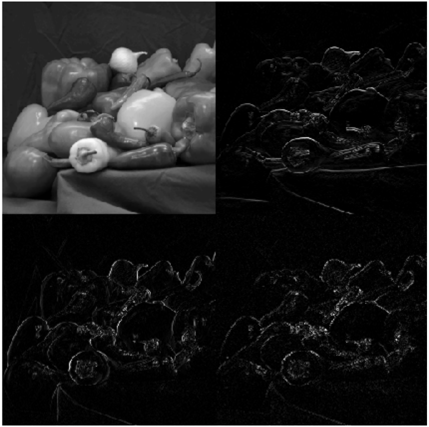
$$\psi_{j,k}(x) = 2^{j/2}\Psi(2^j x_1 - k_1, \ldots, 2^j x_d - k_d), \quad j \in \mathbb{N},\ k \in \mathbb{Z}^d.$$

We consider $f(x) = \sum_{j\in\mathbb{N},k\in\mathbb{Z}^d} f_{j,k}\psi_{j,k}(x), \quad f_{j,k} = \langle f, \psi_{j,k}\rangle.$

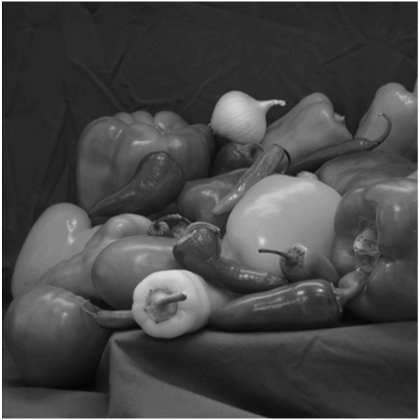


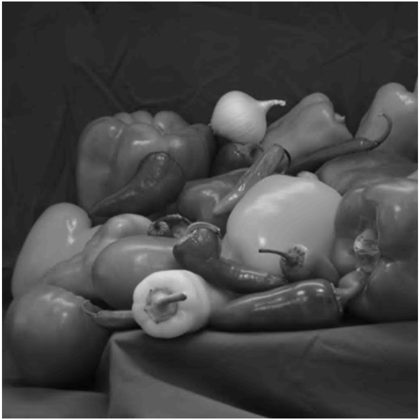
Discrete wavelet transform





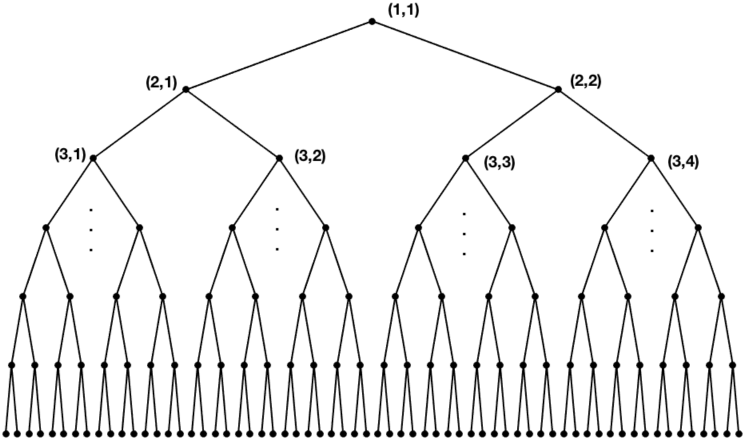
Thresholded peppers





Left: the original image. Right: 95% of the wavelet coefficients are set to zero using hard thresholding.

The wavelet coefficients can be placed into a tree ($d = 1$)



An entire tree is defined as a set of indices

$$\mathbf{T} = \{(j,k) \in \mathbb{N} \times \mathbb{N}^d \mid j \in \mathbb{N}_{\geq 1}, k = (k_1, \cdots, k_d),\ 1 \leq k_\ell \leq 2^{j-1}\},$$

Besov spaces B^s_{pp}

For $s < r$, the Besov norm can be defined as

$$\|f\|^p_{B^s_{pp}(\mathbb{R}^d)} = \sum_{j=0}^{\infty} 2^{jp(s+d(\frac{1}{2}-\frac{1}{p}))} \|F_j\|^p_{\ell^p} \qquad F_j = (f_{j,k})_{k\in\mathbb{Z}^d}.$$

Besov spaces B^s_{pp}

For $s < r$, the Besov norm can be defined as

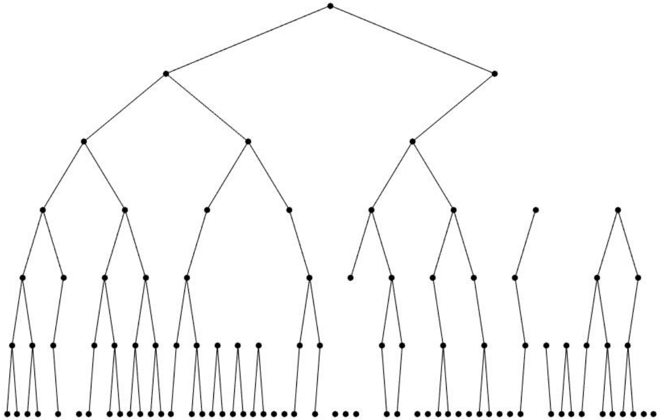
$$\|f\|^p_{B^s_{pp}(\mathbb{R}^d)} = \sum_{j=0}^{\infty} 2^{jp(s+d(\frac{1}{2}-\frac{1}{p}))} \|F_j\|^p_{\ell^p} \qquad F_j = (f_{j,k})_{k\in\mathbb{Z}^d}.$$

- Besov spaces $B^s_{22}(\mathbb{R}^d)$ coincide with the Sobolev spaces $H^s(\mathbb{R}^d)$.
- $B^1_{11}(\mathbb{R}^d)$ space is relatively close to space of functions with bounded variations, $\|\nabla u\|_{L^1} < \infty$.

We can show that:

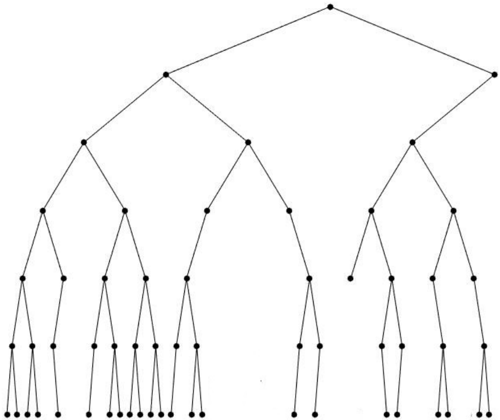
$$B^1_{11}(\mathbb{R}^d) \subset W^{1,1}_{loc}(\mathbb{R}^d) \subset B^{1-\varepsilon}_{11}(\mathbb{R}^d), \qquad \text{for all} \quad \varepsilon > 0.$$

Creating a proper subtree



Draw $t_{j,k} \sim \mathcal{U}[0,1]$ and set a node 1 if $t_{j,k} \leq \beta$, $\beta \in [0,1]$, and 0 otherwise.

Creating a proper subtree



Only choose nodes that are connected to the root node.

Random tree Besov priors ($d = 1$)

Let $\beta = 2^{\gamma-1}$, with $\gamma \in (-\infty, 1]$, and consider pairs (X, T), where X is a $\mathbb{R}^{\mathbf{T}}$-valued random variable and T is a random sub-tree and $X \perp T$.

- The sub-tree T is determined recursively: Let $t_{j,k} \sim \mathcal{U}[0, 1]$, i.i.d. When, for a given level j, all pairs (j, k) in T are chosen, we choose a pair $(j + 1, \ell)$ to be in tree T iff $(j, [\frac{\ell}{2}]) \in T$ and $t_{j+1,\ell} \leq \beta$.
- The sequence X consists of i.i.d $X_{j,k} \sim \mathcal{N}(0, 1)$ or $X_{j,k} \sim \text{Laplace}(0, a)$, $(j, k) \in \mathbf{T}$.

Random tree Besov priors ($d = 1$)

Let $\beta = 2^{\gamma-1}$, with $\gamma \in (-\infty, 1]$, and consider pairs (X, T), where X is a $\mathbb{R}^{\mathbf{T}}$-valued random variable and T is a random sub-tree and $X \perp T$.

- The sub-tree T is determined recursively: Let $t_{j,k} \sim \mathcal{U}[0, 1]$, i.i.d. When, for a given level j, all pairs (j, k) in T are chosen, we choose a pair $(j + 1, \ell)$ to be in tree T iff $(j, [\frac{\ell}{2}]) \in T$ and $t_{j+1,\ell} \leq \beta$.
- The sequence X consists of i.i.d $X_{j,k} \sim \mathcal{N}(0, 1)$ or $X_{j,k} \sim \text{Laplace}(0, a)$, $(j, k) \in \mathbf{T}$.

Let f be the random function

$$f(x) = \sum_{(j,k)\in T} h_j X_{j,k} \psi_{j,k}(x), \quad x \in D = [0, 1],$$

where $h_j = 2^{-j(s+\frac{1}{2}-\frac{1}{p})}$. Then f takes values in $B^{\tilde{s}}_{pp}$, $\tilde{s} < s - \frac{\gamma}{p}$.

Fractal dimension of the prior ($d = 1$)

The random tree Besov constructions creates non-smooth priors. We can also calculate the Hausdorff dimension of the singular support of the resulting prior.

Theorem 1 (K., Lassas, Saksman and Siltanen 2020)

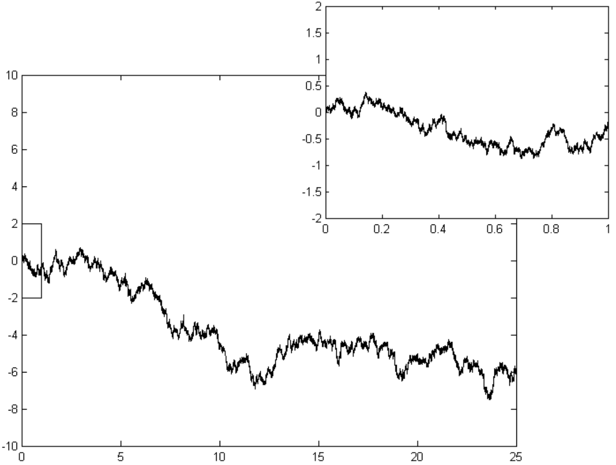
Let $\gamma \in (-\infty, 1]$ satisfy $\beta = 2^{\gamma-1}$ and T be a sub-tree chosen as above.

- If $\gamma < 0$ the sub-tree will terminate on some finite level with probability one $\Longrightarrow f \in C^r$ a.s.
- If $\gamma \in [0, 1]$

$$\mathbb{P}\big(\dim_H(singsupp_r(f)) = \gamma\big) = \frac{2\beta - 1}{\beta^2}$$

and $singsupp_r(f)$ is an empty set with probability $\left(\frac{1-\beta}{\beta}\right)^2$.

Zeros of a Wiener process have Hausdorff dimension 0.5



Wiener processes are often used for modelling stock prices.

Outline

1 Bayesian inverse problems

2 Edge preserving Besov priors

3 Examples

Example: signal denoising

Consider the denoising problem

$$M = f + W,$$

where $W = \sum w_{j,k} \psi_{j,k}$ is white noise, independent of f.

Example: signal denoising

Consider the denoising problem

$$M = f + W,$$

where $W = \sum w_{j,k}\psi_{j,k}$ is white noise, independent of f. We choose prior

$$f(x) = \sum_{(j,k)\in T} f_{j,k}\psi_{j,k}(x) = \sum_{(j,k)\in \mathbf{T}} \tilde{t}_{j,k} g_{j,k}\psi_{j,k}(x),$$

where $g_{j,k} \sim \mathcal{N}(0,1)$ or $g_{j,k} \sim \text{Laplace}(0,a)$ and $\tilde{t}_{j,k} \in \{0,1\}$ defines if a node $(j,k) \in \mathbf{T}$ is chosen. Denote $t_{j,k}$ an independent node, assume $\mathbb{P}(t_{j,k} = 1) = \beta$, $\beta < 1/2$, and

$$\tilde{t}_{j,k} = \prod_{(j',k')\trianglerighteq(j,k)} t_{j,k}.$$

The MAP estimator can be calculated explicitly

- The posterior distribution can be written in form

$$\begin{aligned}\pi(g,t \mid m) &\propto \pi(m \mid g,t)\pi(g)\pi(t) \\ &= \prod_{(j,k)\in\mathbf{T}} \pi(m_{j,k} \mid g_{j,k},\tilde{t}_{j,k})\pi(g_{j,k})\pi(t_{j,k})\end{aligned}$$

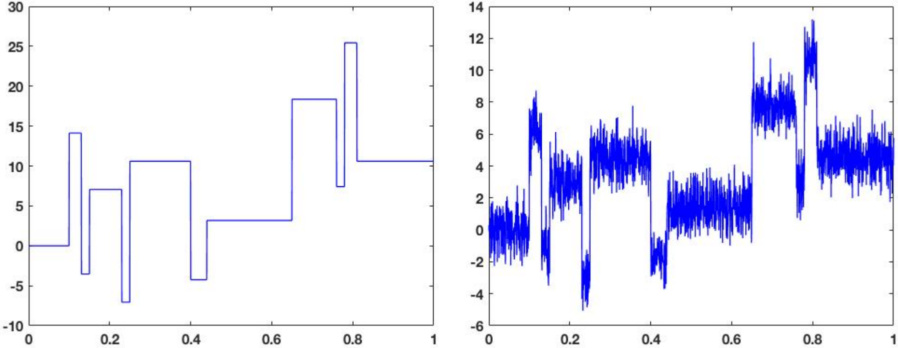
The MAP estimator can be calculated explicitly

- The posterior distribution can be written in form

$$\begin{aligned}\pi(g, t \mid m) &\propto \pi(m \mid g, t)\pi(g)\pi(t) \\ &= \prod_{(j,k)\in \mathbf{T}} \pi(m_{j,k} \mid g_{j,k}, \tilde{t}_{j,k})\pi(g_{j,k})\pi(t_{j,k})\end{aligned}$$

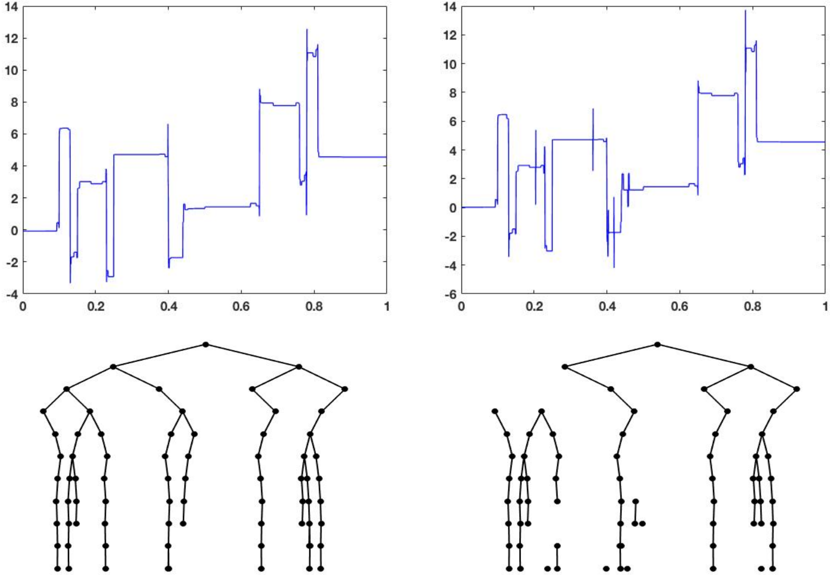
- The MAP estimator can then be calculated recursively.
- The wavelet density $\beta < \frac{1}{2}$ acts as a regularisation parameter.
- If $g_{jk} \sim \mathcal{N}(0, 1)$ the result is a wavelet pruning algorithm which either accepts a full branch or turns it off.
- If $g_{jk} \sim \text{Laplace}(0, a)$ then we get tree enforced soft thresholding algorithm with threshold a.

Example: signal denoising



Original blocks signal and a noisy signal with signal to noise ratio 3.

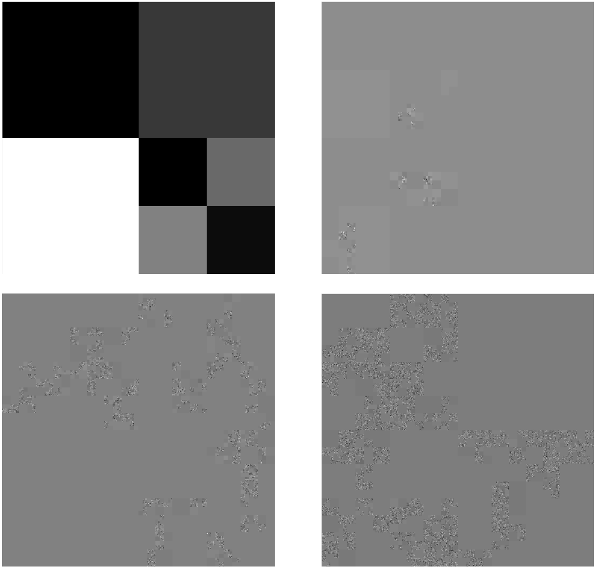
Denoising using wavelet pruning and thresholding



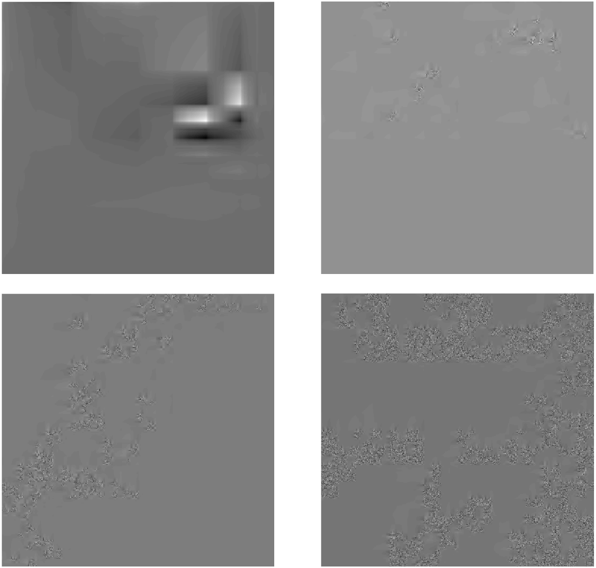
Example: Image denoising

Original and noisy image

Prior draws with Haar wavelets, Gaussian coefficients and wavelet densities 0.2, 0.4, 0.6 and 0.8



Prior draws with Daubechies 2 wavelets, Gaussian coefficients and wavelet densities 0.2, 0.4, 0.6 and 0.8



Denoising an image using wavelet pruning

Using Laplace prior leads to tree enforced soft thresholding

Comparing pruning, tree enforced soft thresholding and soft thresholding

In a nutshell

1. We proposed new edge preserving random tree Besov priors
   - Similar properties to TV priors
   - Correspond to well defined infinite dimensional random variables
2. We can calculate the fractal dimension of the singular support of a prior draw
3. We introduced a sparsity promoting algorithm for calculating the MAP estimator in denoising problem
   - Semi-Gaussian prior $\Longrightarrow$ Wavelet tree pruning
   - Semi-Laplace prior $\Longrightarrow$ Tree enforced soft thresholding