# Random tree besov priors – towards fractal imaging

Kekkonen, Hanne; Lassas, Matti; Saksman, Eero; Siltanen, Samuli

**DOI**
10.3934/ipi.2022059

**Publication date**
2023

**Document Version**
Final published version

**Published in**
Inverse Problems and Imaging

Inverse Problems and Imaging
Vol. 17, No. 2, April 2023, pp. 507-531
doi:10.3934/ipi.2022059

# RANDOM TREE BESOV PRIORS – TOWARDS FRACTAL IMAGING

Hanne Kekkonen[*1], Matti Lassas[2], Eero Saksman[2] and Samuli Siltanen[2]

[1]Delft University of Technology, Netherlands

[2]University of Helsinki, Finland

(Communicated by Martin Burger)

Abstract. We propose alternatives to Bayesian prior distributions that are frequently used in the study of inverse problems. Our aim is to construct priors that have similar good edge-preserving properties as total variation or Mumford-Shah priors but correspond to well-defined infinite-dimensional random variables, and can be approximated by finite-dimensional random variables. We introduce a new wavelet-based model, where the non-zero coefficients are chosen in a systematic way so that prior draws have certain fractal behaviour. We show that realisations of this new prior take values in Besov spaces and have singularities only on a small set $\tau$ with a certain Hausdorff dimension. We also introduce an efficient algorithm for calculating the MAP estimator, arising from the the new prior, in the denoising problem.

1. **Introduction.** Inverse problems arise from the need to extract information from indirect measurements. They are typically ill-posed, meaning that algorithmic recovery of information is sensitive to noise and modelling errors. Robust reconstruction methods are based on combining the measurement data with *a priori* knowledge about the unknown target. Formulating *a priori* knowledge mathematically is a core challenge in inverse problems research. Popular models for *a priori* information promote global smoothness, piecewise regularity, or sparsity in a given or learned basis or a more general collection of building blocks. They can be implemented using variational regularisation, providing a stable solution but little information about its uncertainties. Bayesian inversion offers an attractive alternative by also delivering information on how uncertainties in the data or model affect the point estimates.

Practical measurements are always finite and corrupted by noise, which can in many cases be reasonably modelled using independent Gaussian random variables. This gives rise to a discrete measurement model of the form

$$M_i = (Af)_i + w_i, \quad i = 1, \dots, n, \ w_i \overset{\text{iid}}{\sim} \mathcal{N}(0,1), \tag{1}$$

where $A$ describes the forward process. Computational solutions of the inverse problem require also a finite approximate model for the unknown $f$. It is advisable to design the models for $f$ and for the noise in a discretisation-invariant way [26, 32,

2020 *Mathematics Subject Classification.* Primary: 60F17, 65C20, 42C40; Secondary: 28A80.
*Key words and phrases.* Inverse problem, statistical inversion, Bayesian inversion, Besov priors, discretisation invariance, wavelets, fractals.

*Corresponding author: Hanne Kekkonen.

34, 33]. One way to achieve this is to build a continuous model which is discretised as late as possible in the analysis and solution procedure [18, 46]. In this paper we consider the case where $A : H_1 \to H_2$ is a linear operator between Banach spaces $H_1$ and $H_2$, and assume the continuous equivalent model (in the sense of [5, 43])

$$M = Af + \varepsilon \mathbb{W}, \tag{2}$$

where $\mathbb{W}$ is a Gaussian white noise process indexed by $H_2$.

Fractals have many applications in science since they often describe the real world better than traditional mathematical models. As Mandelbrot noted, "Clouds are not spheres, mountains are not cones, coastlines are not circles, and bark is not smooth, nor does lightning travel in a straight line." Fractal geometry and analysis of fractal dimension is a powerful tool that has been used to study turbulent mixing flows and atmospheric turbulence [21, 44], evaluate the risk of a patient developing cancer and predict tumour malignancy [37, 14, 6], screening for osteopenia [3], recognising Alzheimer's disease patients from magnetoencephalogram recordings [25], and in study of macular diseases [47].

We will next give a few examples of measurement models like (2) arising in practical applications where the unknown quantity has fractal properties. In each case we encounter a function $f$ that has singular support on a fractal set $\tau$. We emphasise that we mostly consider *random fractals:* sets produced by a random process whose (Hausdorff) dimension is not an integer.

As a one-dimensional example we consider functions $f : [0, 1] \to \mathbb{R}$ whose graphs have fractal properties or whose smoothness changes on a fractal set. Such functions appear, for example, in medicine (heart rate, stride, and breathing variability) [52] and financial data (local volatility models for asset prices where the volatility changes when the price drops below a threshold value) [38, 39, 36, 42]. Here the formula (2) can model e.g. interpolation or extrapolation problems. An example of two-dimensional problems of the form (2) are digital image processing tasks such as deblurring, inpainting and denoising an image $f : [0, 1]^2 \to \mathbb{R}$. If the underlying picture is a photograph containing clouds, coastlines or skylines with forest or mountains, some interfaces between image areas are fractals. See Figure 1 (a). X-ray tomography is a good example of a three-dimensional inverse problem that fits to our model. There, one records several two-dimensional X-ray images of a patient along different directions of projection. The goal is to reconstruct the X-ray attenuation coefficient $f : [0, 1]^3 \to \mathbb{R}$ from these images, where each pixel is considered as a line integral of $f$ along a ray. The structure of human lungs is known to be fractal [52], so our new model may be useful for lung tomography. See Figure 1 (b).

Most of the theory for infinite-dimensional Bayesian inverse problems focuses on the case were the unknown is assumed to follow a Gaussian prior. Gaussian inverse problems benefit from fast computational properties, but in many signal and image reconstruction problems the detection of edges and interfaces is crucial. In those cases basic Gaussian priors are not optimal. One way of circumventing the issue is to use hierarchical Gaussian models as in [7, 8]. Closely related to this paper are the hierarchical models whose *maximum a posteriori* (MAP) estimate converges to a minimiser of the Mumford-Shah functional [27, 28].

We follow the idea of using wavelet-based Besov priors introduced in [33] and further studied in [16]. Consistency of such priors has been considered in [2, 2, 24].

(a) (b)

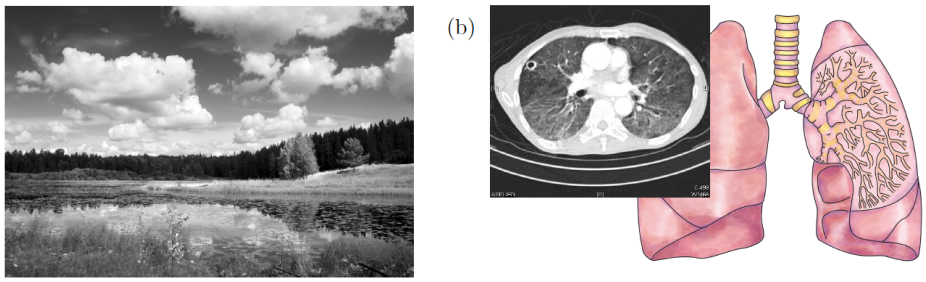

FIGURE 1. (a) Photograph featuring fractal-like structures such as boundaries of clouds. (b) The three-dimensional structure of human lungs follows a self-similar rule leading to a fractal dimension. Also shown is an X-ray tomography slice of lungs. Images courtesy of Wikimedia Commons.

These priors are especially useful since smooth functions with few local irregularities have a sparser expansion in the wavelet basis than, e.g., in the Fourier basis. If the unknown is assumed to be sparse in some basis, using a prior that encourages this sparsity results in a more efficient finite-dimensional approximation of the solution. Methods for recovering finite-dimensional estimates of the unknown based on wavelet bases are broadly studied in image and signal processing, and statistical literature, see e.g. [1, 13, 20, 22]. Priors based on small trees and structured wavelet shrinkage have also been used in density estimation, see e.g. [9, 12, 11, 10].

We assign a prior probability measure $\Pi$ to $f$. The full solution to the Bayesian inverse problem is the *posterior distribution*; i.e. conditional distribution of $f$ for a given realization of data. The mean or mode of the posterior can be used as a point estimator. A popular method for achieving edge-preserving solutions in image analysis is to employ the so-called total variation prior and use the mode of the posterior as a point estimate. In practice this means solving the minimisation problem

$$\min_{f\in H}\left\{\|Af-M\|_{L^2}^2+\beta\|\nabla f\|_{L^1}\right\}, \tag{3}$$

where $H$ is a finite-dimensional subspace of piece-wise smooth functions and $A : L^2 \to L^2$ is a continuous operator. Despite active research in this area, no natural infinite dimensional models have been found that would have (3) as a MAP estimate. In other words, the widely used formal prior

$$\pi(f) \underset{formally}{\propto} \exp(-\beta\|\nabla f\|_{L^1}) \tag{4}$$

is not known to correspond to any well-defined random variable. It is also known that the usual discrete total variation priors can converge to Gaussian smoothness priors as the discretization is refined, see [34].

The idea presented in [33] is to replace formula (4) by a well-defined prior

$$\pi(f) \underset{formally}{\propto} \exp(-\|\nabla f\|_{B^0_{pp}}^p),$$

where the Besov spaces $B^0_{pp}$ are closely related to $L^p$ spaces. Hence the Besov $B^0_{11}$-priors have similar properties to total variation prior, but correspond to well-defined infinite-dimensional random variables. The construction of the Besov prior

is done using the Karhunen-Loève expansion and can hence be approximated with finite-dimensional random variables.

We introduce a new wavelet-based computational model for *a priori* information about the fractal dimension of the edges of the unknown target. We will build on the idea from [33] but choose the non-zero wavelet coefficients in the Karhunen-Loève expansion in a systematic way, so that the resulting priors have a certain fractal dimension. This is done by introducing a new random variable $T$ that takes values in the space of 'trees'. The trees are chosen so that the realisations $f$ have singularities only on a small set. This opens up new possibilities in medical imaging and signal processing. Our new model allows rigorous analysis in the Bayesian inversion framework. Also, we present an efficient algorithm for the computation of MAP estimates for two- and three-dimensional denoising problems. The exact construction of the prior can be found in Section 3.

The rest of the paper is organised as follows. In Section 2 we introduce our Besov space setting for constructing the priors. We then define the new random-tree Besov priors and formulate the main results of the paper in Section 3. Section 4 is dedicated for the denoising examples and constructing an algorithm for calculating MAP estimators.

## 2. Priors in Besov-spaces.

### 2.1. Gaussian inverse problems.

We start by motivating the use of Besov priors by first considering the standard Gaussian prior. Let $H_1$ is a Banach space, and let $f$ be a $H_1$-valued random variable following a Gaussian distribution $\mu$. We denote by $\mathbb{W}$ the Gaussian white noise proces. For simplicity, we assume that $\mathbb{E}f = 0$ and denote by $C_f$ and $C_w$ the covariance operators of $f$ and $\mathbb{W}$, respectively. If the forward operator $A : H_1 \to H_2$ is assumed to be linear, then the posterior distribution $\mu(\cdot \,|\, M)$ is also Gaussian. It follows that the conditional mean estimate for the Bayesian inverse problem (2) coincides with the MAP estimate (under mild assumptions on $A$, see e.g. [17]) and is given by

$$f_{MAP} = (A^* C_w^{-1} A + C_f^{-1})^{-1} A^* C_w^{-1} M, \tag{5}$$

where $A^* : H_2 \to H_1$ is the adjoint of $A$ (see e.g. [35]).

**Example 1.** Here we consider a stereotype of a 1-dimensional linear Bayesian inverse problem. We use an *a priori* model that $f$ is a Brownian bridge on the interval $(0, \pi)$. Assume that $H_1 = L^2(0, \pi)$ with the usual inner product $\langle \cdot, \cdot \rangle$. Choose the basis $e_j(x) = \frac{1}{2}\sin(jx)$. Let $A_N = \{f \in H_1 : \ a_j < \langle f, e_j \rangle < b_j, \ j = 1, 2, \ldots, N\}$ be a cylinder and define $\mu$ by setting

$$\mu(f : (\langle f, e_j \rangle)_{j=1}^N \in A_N\}) = c_N \int_{a_1}^{b_1} \cdots \int_{a_N}^{b_N} \prod_{j=1}^N e^{-\lambda_j^{-1} t_j^2/2} dt_1 \ldots dt_N, \tag{6}$$

where $c_N$ are normalisation constants and $\lambda_j = j^{-2}$. We see that $f$ has a representation $f = \sum_{j=1}^\infty \lambda_j^{-1/2} f_j e_j$, where $f_j$ are independent normalised Gaussian variables. Furthermore, $f$ has the covariance operator $C_f$ defined by

$$\langle C_f \phi, \psi \rangle = \mathbb{E}\left(\langle \phi, f \rangle \langle f, \psi \rangle\right), \tag{7}$$

where $\phi, \psi \in L^2(0, \pi)$, which can easily be seen to be

$$C_f \phi = \sum_{j=1}^\infty \lambda_j \langle \phi, e_j \rangle e_j. \tag{8}$$

Now, if we consider the Dirichlet Laplacian, that is, the operator $-\Delta = -\frac{d^2}{dx^2}$ defined in the domain $\mathcal{D}(-\Delta) = H^2(0,\pi) \cap H_0^1(0,\pi)$, we see that $e_j$ are the complete system of orthonormal eigenvectors of $-\Delta$ corresponding to eigenvalues $\lambda_j^{-1}$. Thus we see from (8) that the covariance operator $C$ coincides with $-\Delta^{-1}$. In applications, the distribution (6) is (non-rigorously) expressed by saying that $f$ has the probability density function

$$\pi_f(f) = ce^{-\langle f, \Delta f\rangle/2}, \quad f \in L^2(0,\pi) \tag{9}$$

where $c$ is a normalisation constant. This notation becomes rigorous if we discretise the system, that is, consider $N+1$-dimensional vector $(f(0), f(\pi/N), \ldots, f(\pi))$ and approximate $-\Delta$ with the finite difference operator [31]. When $A : L^2(0,\pi) \to L^2(0,1)$ is a bounded linear operator, and $\mathbb{W}$ is normalised Gaussian white noise on the interval $(0,1)$, then the conditional mean estimate for inverse problem (2) is $f_{MAP} = (A^*A - \frac{1}{2}\Delta)^{-1}A^*m$, in accordance with (5). Here $m$ is the given measurement.

2.2. **Sobolev and Besov spaces.** Our aim is to combine the fast computational properties of the Gaussian inverse problems with the good edge-preserving properties of some non-Gaussian variables, in particular those generated by assuming a total variation prior. Analogously to (4) we would like to use the prior

$$\pi(f) \underset{formally}{\propto} \exp(-\|\nabla f\|_{L^p}^p), \quad p \geq 1.$$

As this is not a well-defined object for general $p$, we will replace it by

$$\pi(f) \underset{formally}{\propto} \exp(-\|\nabla f\|_{B_{pp}^0}^p)$$

or more generally by

$$\pi(f) \underset{formally}{\propto} \exp(-\|f\|_{B_{pp}^s}^p),$$

where the Besov space $B_{pp}^0$ has many similar properties to $L^p$ as we see below.

In this section we recall the Besov spaces $B_{pq}^s(\mathbb{R}^d)$, $1 \leq p, q < \infty$ and $s \in \mathbb{R}$, and the closely related Sobolev spaces $W^{s,p}(\mathbb{R}^d)$ (see e.g. [50] for general theory). Let $\mathcal{F}$ denote the Fourier transform and $\mathcal{S}'(\mathbb{R}^d)$ the space of tempered generalised functions (also called tempered distributions). The Sobolev spaces on $\mathbb{R}^d$ for $p \in (1,\infty)$ and $s \in \mathbb{R}$ are defined as

$$W^{s,p}(\mathbb{R}^d) = \{f \in \mathcal{S}'(\mathbb{R}^d) \,:\, \mathcal{F}^{-1}\big((1+|\cdot|^2)^{\frac{s}{2}}\mathcal{F}f\big) \in L^p(\mathbb{R}^d)\}$$

with norm

$$\|f\|_{W^{s,p}(\mathbb{R}^d)} := \big\|\mathcal{F}^{-1}\big((1+|\cdot|^2)^{\frac{s}{2}}\mathcal{F}f\big)\big\|_{L^p(\mathbb{R}^d)}.$$

For $s \in \mathbb{N}$ the above definition is equivalent to having derivatives of $f$ of order at most $s$, defined in the distributional sense, in $L^p$. This second definition can be used to consider also the Lebegue exponent $p = 1$. For example, $W^{1,1}(\mathbb{R}^d)$ consist of functions that together with their first derivatives lie in $L^1(\mathbb{R}^n)$.

The Besov spaces are slightly more complicated. One possible way to define them is to use Fourier multipliers, that is, frequency band filters in the Fourier space. First, let $\phi_0 \in C_0^\infty(\mathbb{R}^d)$ be a function for which

$$\operatorname{supp}(\phi_0) \subset \left\{\xi \in \mathbb{R}^d : |\xi| \leq 3/2\right\}, \quad \phi_0(\xi) = 1 \text{ if } |\xi| \leq 1.$$

We can then define

$$\phi_j(\xi) = \phi_0\left(2^{-j}\xi\right) - \phi_0\left(2^{-j+1}\xi\right), \quad \xi \in \mathbb{R}^d,\ j \geq 1,$$

and note that

$$\operatorname{supp}(\phi_j) \subset \left\{\xi \in \mathbb{R}^d : 2^{j-1} \leq |\xi| \leq 2^{j+1}\right\} \quad \text{and} \quad \sum_{j=0}^{\infty} \phi_j(\xi) = 1 \text{ if } \xi \in \mathbb{R}^d,$$

that is, $\phi_j$ is a partition of unity. We define the Fourier multipliers $F_j : \mathcal{S}'(\mathbb{R}^d) \to \mathcal{S}'(\mathbb{R}^d)$ as,

$$F_j f = \mathcal{F}^{-1}(\phi_j(\cdot)\mathcal{F}f).$$

The Besov space is the collection of all $f \in \mathcal{S}'$ such that

$$\|f\|_{B^s_{pq}(\mathbb{R}^d)} = \left(\sum_{j=0}^{\infty} 2^{sjq}\|F_j f\|^q_{L^p(\mathbb{R}^d)}\right)^{1/q} < \infty. \tag{10}$$

Below, we denote $\|f\|_1 \approx \|f\|_2$ when $c_1\|f\|_1 \leq \|f\|_2 \leq c_2\|f\|_1$ with some $c_1, c_2 > 0$. In particular, we are interested in $B^1_{1\,1}(\mathbb{R}^d)$ having norm

$$\|f\|_{B^1_{1\,1}(\mathbb{R}^d)} \approx \|f\|_{B^0_{1\,1}(\mathbb{R}^d)} + \|\nabla f\|_{B^0_{1\,1}(\mathbb{R}^d)} \approx \|f\|_{B^0_{1\,1}(\mathbb{R}^d)} + \sum_{j=0}^{\infty} \|\nabla(F_j f)\|_{L^1(\mathbb{R}^d)}, \tag{11}$$

(see the first theorem in [48, Section 2.3.8]) as the space $B^1_{1\,1}(\mathbb{R}^d)$ is relatively close to space of functions of bounded variations that satisfy $\|\nabla f\|_{L^1} < \infty$. For example, it is not difficult to show that (locally) all the functions $x_+^\alpha$ belong to $B^1_{1\,1}(\mathbb{R})$ for any $\alpha > 0$. Formula (11) also implies that

$$\begin{aligned} \|f\|_{B^1_{1\,1}(\mathbb{R}^d)} &\approx \|f\|_{B^0_{1\,1}(\mathbb{R}^d)} + \|\nabla f\|_{B^0_{1\,1}(\mathbb{R}^d)} \\ &\approx \sum_{j=0}^{\infty} \|F_j f\|_{L^1(\mathbb{R}^d)} + \sum_{j=0}^{\infty} \|\nabla(F_j f)\|_{L^1(\mathbb{R}^d)} \approx \sum_{j=0}^{\infty} \|F_j f\|_{W^{1,1}(\mathbb{R}^d)}. \end{aligned} \tag{12}$$

This means that if a function $f : \mathbb{R} \to \mathbb{R}$ is written as a sum its frequency band filtered components $F_j f$, that is, $f = \sum_{j=0}^{\infty} F_j f$, then $f$ is in the Besov space $B^1_{1\,1}(\mathbb{R}^d)$ if and only if the frequency band filtered components $F_j f$ are in $W^{1,1}(\mathbb{R})$ and the sum of the Sobolev norms of the components, $\sum_{j=0}^{\infty} \|F_j f\|_{W^{1,1}(\mathbb{R}^d)}$, is finite. Thus $B^1_{1\,1}(\mathbb{R})$ almost contains functions having Heaviside-type jump singularities.

Let us digress for a slightly more careful comparision of the $B^1_{1\,1}$-norm and the $W^{1,1}$-norm. As $f = \sum_{j=0}^{\infty} F_j f$, formula (12) implies that $\|f\|_{W^{1,1}(\mathbb{R}^d)} \leq \|f\|_{B^1_{1\,1}(\mathbb{R}^d)}$ which implies the embedding $B^1_{1\,1}(\mathbb{R}^d) \subset W^{1,1}(\mathbb{R}^d)$. On the other hand, [51, (7.3)] verifies that compactly supported signed Borel messures belong to $B^0_{1\,\infty}$, and a constant times their total variation norm gives an upper bound for the Besov norm. By combining this with [48, 2.3.8] we deduce that $W^{1,1}(\mathbb{R}^d) \subset B^1_{1\,\infty}$. Moreover, by using Hölders inequality and formula (10), we see that for any $\varepsilon > 0$ we have $B^1_{1\,\infty}(\mathbb{R}^d) \subset B^{1-\varepsilon}_{1\,1}(\mathbb{R}^d)$. The embedding $W^{1,1}(\mathbb{R}^d) \subset B^{1-\varepsilon}_{1\,1}(\mathbb{R}^d)$ follows from these. Thus the spaces $B^1_{1\,1}$ and $W^{1,1}$ are rather close to each other since an arbitrarily small drop of smoothness changes the direction of the inclusion of one space into other:

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

We finally mention that Besov norm for $p > 1$ can also be defined equivalently via suitable integrals of finite differences of the function $f$, see e.g. [49, p.8].

2.3. **Random variables with values in Besov spaces.** In this paper we will use the wavelet representation of Besov norms in dimension $d \geq 1$. For this, we recall that for any integer $r \geq 1$ there exist compactly supported $C^r$ functions $\phi$ and $\Psi^\ell$ $(\ell = 1, \ldots, 2^d - 1)$ which generate wavelets suitable for multi-resolution analysis of smoothness $C^r$. More specifically, if we denote

$$\psi^\ell_{jk}(x) = 2^{jd/2}\Psi^\ell(2^j x - k), \qquad k \in \mathbb{Z}^d,\ j \geq 0,\ \ell \in \mathcal{L} := \{1, \ldots, 2^d - 1\},$$
$$\phi_k(x) = \phi(x - k), \qquad k \in \mathbb{Z}^d,$$

we can write any function $f \in L^2(\mathbb{R}^d)$ using the wavelet representation

$$\begin{aligned} f(x) &= \sum_{k\in\mathbb{Z}^d} \langle f, \phi_k\rangle \phi_k(x) + \sum_{\substack{j\geq 0,\ k\in\mathbb{Z}^d \\ \ell\in\mathcal{L}}} \langle f, \psi^\ell_{jk}\rangle \psi^\ell_{jk}(x) \\ &= \sum_{k\in\mathbb{Z}^d} f_{-1k}\phi_k(x) + \sum_{\substack{j\geq 0,\ k\in\mathbb{Z}^d \\ \ell\in\mathcal{L}}} f^\ell_{jk}\psi^\ell_{jk}(x). \end{aligned}$$

Especially, Daubechies wavelets with vanishing moments up to order $N - 1 \in \mathbb{N}$ are suitable for $r$-regular multi-resolution analysis if $r < 0.1936(N-1)$. Daubechies wavelets $\phi, \Psi^\ell$ $(\ell = 1, \ldots, 2^d - 1)$ of order $N$ are supported in the cube $[-L, L]^d$, where we have $L := 2N - 1$. For more details about wavelets see [19, 41].

Note that for 1D signals $\ell = 1$ and the wavelet coefficients are naturally organised in a binary tree. However, for 2D image data $\ell \in \{1, 2, 3\}$ and the corresponding data structure consists of three parallel quadtrees, since the vertical, horizontal and diagonal subbands each have their own tree. We take the simplified approach of using only one logical tree where value 0 means that all subbands have a zero coefficient at that location, and value 1 means that all subbands are allowed to have a nonzero coefficient on that node. It is an interesting avenue for further investigations to consider separate trees and to generalise to more steerable frames than orthonormal wavelets, but such discussions are outside the scope of this paper.

We have the following equivalent representation to the Besov norm (10) in case $p = q$

$$\|f\|_{B^s_{pp}(\mathbb{R}^d)} \simeq \left( \sum_{j=-1}^{\infty} 2^{jp\left(s + \frac{d}{2} - \frac{d}{p}\right)} \|\mathbf{f}_j\|^p_{\ell^p} \right)^{1/p}, \tag{13}$$

where

$$\mathbf{f}_j = (f^\ell_{jk})_{k\in\mathbb{Z}^d,\ \ell\in\mathcal{L}} \quad \text{with} \quad \|\mathbf{f}_j\|_{\ell^p} = \left( \sum_{k\in\mathbb{Z}^d,\ \ell\in\mathcal{L}} |f^\ell_{jk}|^p \right)^{1/p} \qquad \text{if } j \geq 0, \quad \text{and}$$

$$\mathbf{f}_{-1} = (f_{-1k})_{k\in\mathbb{Z}^d} \quad \text{with} \quad \|\mathbf{f}_{-1}\|_{\ell^p} = \left( \sum_{k\in\mathbb{Z}^d} |f_{-1k}|^p \right)^{1/p}.$$

We may re-enumerate the full set of indices of our wavelets

$$\{(-1, k)_{k\in\mathbb{Z}^d}\} \cup \{(\ell, j, k)_{\ell\in\mathcal{L},\, j\in\mathbb{N},\, k\in\mathbb{Z}^d}\}$$

as $\{(\ell(l), j(l), k(l)) :\ l \in \mathbb{N}\}$ (if $j(l) = -1$ we may set set $k(l) = -1$ and interpret $(-1, -1, k) = (-1, k)$). In the case $p = q$ we have the following trivial but essential

consequence of (13) which states that considering random Besov-space valued variables is abstractly equivalent to considering random variables in a weighted $\ell^p$-space (or via another trivial isomorphism, in a non-weighted $\ell^p$-space).

**Proposition 2.1.** *Besov space $B^s_{pp}(\mathbb{R}^d)$ is isomorphic to the weighted $\ell^p$ space*

$$\ell^p_v = \{(a_l)_{l=1}^\infty : \ \|(a_l)\|^p_{\ell^p_v} = \sum_{l=1}^\infty |v_l a_l|^p < \infty\}$$

*where $v = (v_1, v_2, \dots)$,*

$$v_l = 2^{j(l)\left(s+\frac{d}{2}-\frac{d}{p}\right)}. \tag{14}$$

*The isomorphism is given by*

$$(a_l) \mapsto \sum_{l=1}^\infty a_l \psi^{\ell(l)}_{j(l),k(l)}.$$

For simplicity we will mainly consider the behaviour of random functions on the unit cube $D = [0,1]^d$. Hence, while defining random functions via the wavelet decompositions we may well set $f^\ell_{jk} = 0$ if $k = (k_1, \dots, k_d)$ and $|k_l| > L2^j - 1$ for some $l = 1, \dots, d$. Because of this, we call the remaining set of indices of possibly nonzero wavelet coefficients the *entire tree*, which is given by

$$\mathbf{T} = \{(j,k) \in \mathbb{N} \times \mathbb{Z}^d \,,\ j \geq 0,\ k = (k_1, \dots, k_d),\ |k_l| \leq L2^j - 1\}$$

Note that we consider the same tree for all $\ell \in \mathcal{L}$.

We emphasize that in our theoretical results on the behaviour of the Besov prior we aim for modelling the generic local behaviour of the prior, which in our setup takes place *in the interior of the unit cube* $D = [0,1]^d$. One may of course fine-tune the definition of the prior suitably in connection with different boundary conditions, and study its behaviour also near the boundary, but since this is generally application-specific we do not consider it in this paper.

Next we introduce random variables in Besov spaces, or equivalently in $\ell^p_v$. The generated measures are similar to the p-exponential measures whose consistency has been studied in [2].

**Definition 2.2.** Let $\{\psi^\ell_{jk}\}_{\ell\in\mathcal{L},\, j\in\mathbb{N},\, k\in\mathbb{Z}^d}$ be an $r$-regular wavelet basis for $L^2(\mathbb{R}^d)$. Let $(X^\ell_{jk})_{(j,k)\in\mathbf{T}}$ be an i.i.d. sequence of real random variables with probability density function $\pi(x) \propto \exp(-\frac{1}{2\kappa^p}|x|^p)$, $1 \leq p < \infty$, in which case we denote $X^\ell_{jk} \sim \mathcal{N}_p(0, \kappa^p)$. Let the random function $f$ be defined as

$$f(x) = \sum_{\substack{(j,k)\in\mathbf{T}\\ \ell\in\mathcal{L}}} h_j X^\ell_{j,k} \psi^\ell_{j,k}(x), \quad x \in D,$$

where $h_j = 2^{-j\left(s+\frac{d}{2}-\frac{d}{p}\right)}$, $s < r$, are deterministic constants. In light of (13) we call $f$ a $B^s_{pp}$-random variable.

We note that random variables defined in Definition 2.2 take values in Besov spaces $B^t_{pp}$, with $t < s - \frac{d}{p}$, a.s. and a realisation $f$ takes values in the space $B^{s-d/p}_{pp}$ only with probability zero. This follows directly from Theorem 3.3 below with $\gamma = d$. The space $B^s_{pp}$ plays a role similar to the Cameron-Martin space for Gaussian processes. Informally, $f$ has density proportional to $\exp(-\frac{1}{2\kappa^p}\|f\|^p_{B^s_{pp}})$,

see [2, 33, 16]. If $p = 2$, we obtain a Gaussian measure with Cameron-Martin space $B^s_{22} = H^s$. If $p = 1$, then $f$ is a semi-Laplace random variable, by which we mean that $f$ is a hierarchical random variable which is determined in the last stage as a Laplace random variable.

3. **Generalisation to random functions having singularities on random fractals.** In applications the strength of wavelets often appears in the fact that if a large portion of small wavelet coefficients of a given function are replaced by zero, the new function corresponding to these truncated wavelet coefficients approximates well the original function. In particular, the singularities of the original function are often preserved. Because of this we consider a model where a large part of wavelet coefficients are assumed to be zero. This idea has been previously studied e.g. by [1, 15, 29, 45] but here we emphasise the singularities by choosing the non-zero wavelet coefficients in a systematic way.

We will consider the following set of subtrees of $\mathbf{T}$:

$$\Gamma = \{T \subset \mathbf{T} \,|\, \text{if } (j,k) \in T \text{ and } j \geq 1 \text{ then } (j-1, [k/2]) \in T\}$$

where $[k/2] = ([k_1/2], \ldots, [k_d/2])$ is the vector which elements are the integer parts of $k_l/2$. We call $\Gamma$ *the set of proper subtrees.* The above definition means that if some node is in the tree $\Gamma$, then all of its ancestor nodes also belong to the tree. That is, all the branches are connected to the root node, see right hand side of Figure 2.

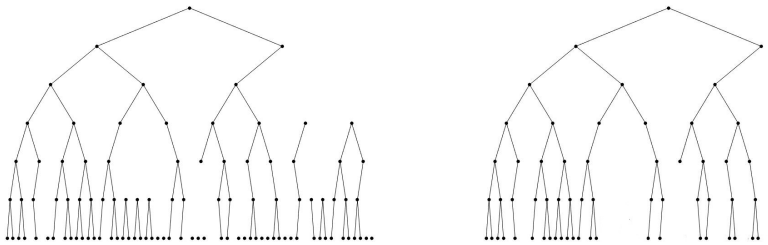

FIGURE 2. On the left: all the nodes where $t_{j,k} < \beta$, with $t_{j,k} \sim \mathcal{U}[0,1]$ and $\beta \in [0,1]$. On the right: the proper subtree where only nodes with direct connection to the root node are included.

**Definition 3.1.** Let $\{\psi^\ell_{jk}\}_{\ell \in \mathcal{L}, j \in \mathbb{N}, k \in \mathbb{Z}^d}$ be an $r$-regular wavelet basis for $L^2(\mathbb{R}^d)$, and $\beta = 2^{\gamma - d}$ with some $\gamma \in (-\infty, d]$. Consider pairs $(X^\ell, T)$ where $X^\ell$ is an $\mathbb{R}^{\mathbf{T}}$-valued random variable, and $T \in \Gamma$ is a random tree. We assume that $X$ and $T$ are independent random variables, having the following distributions

- The sequence $X^\ell$ consists i.i.d $X^\ell_{jk} \sim \mathcal{N}_p(0, \kappa^p)$, with probability density proportional to $\exp(-\frac{1}{2\kappa^p}|x|^p)$, $\kappa > 0$ and $1 \leq p < \infty$.
- A proper subtree $T$ is build recursively by choosing at each level new nodes into the tree with probability $\beta$. More precisely, $T$ is determined with the following rule: We assume that the root node is always chosen. For the rest of the levels; When for a given level $j$ all nodes $(j,k)$ in the tree $T$ are chosen we can move to level $j+1$. First draw $t_{j+1,l} \sim \mathcal{U}([0,1])$. The node $(j+1, l)$ is chosen to be in the tree $T$ if and only if its parent is in the tree and $t_{j+1,l} < \beta$.

Let $f$ be the random function

$$f(x) = \sum_{\substack{(j,k)\in T \\ \ell\in\mathcal{L}}} h_j X_{j,k}^{\ell} \psi_{j,k}^{\ell}(x), \quad x \in D,$$

where $h_j = 2^{-j\left(s+\frac{d}{2}-\frac{d}{p}\right)}$, $s < r$, are deterministic constants. We say that $f$ is a $B_{pp}^s$-random variable with stochastic fractality index $\beta$.

We call $\beta$ stochastic fractality index since, with certain positive probability, the Hausdorff dimension of the singular support of the random function $f$ turns out to be $\log_2(\beta) + d$, see Corollary 3.4. Notice that every node can have at most $2^d$ children in the tree $T$. This means that if we choose new nodes to $T$ with probability $\beta < 2^{-d}$ the recursion will stop at some point with probability one resulting in a finite tree. A random function $f$ created using a finite tree has the same regularity as the wavelet basis.

Note that for a single $(\ell, j, k)$ the random variable $X_{j,k}^{\ell} \sim \mu = \mathcal{N}_p(0, \kappa^p)$ has probability distribution

$$X_{jk}^{\ell} \sim P_j\mu + (1 - P_j)\delta_0,$$

i.e. it vanishes with a probability $1 - P_j$ (compare this with [1]).

Next we will employ theory of random fractals based on[23]. Let us consider dyadic cubes

$$Q_{j,k} = \{y \in [0,1]^d : \ [jy] = k\}.$$

For the random variable $T$ we define a random fractal (a set-valued random variable)

$$\tau = \bigcap_{j=1}^{\infty} \bigcup_{k\in R_j} \overline{Q}_{j,k}$$

where $R_j = \{k : (j,k) \in T\}$ is set of level $j$ elements in the random tree $T$.

We are interested in the behaviour of random draws $f$ from the random tree Besov prior and will show that the random fractal $\tau$ coincides with the $r$-singular support of $f$. The $r$-singular support of $f$ is the complement of the largest open set on which $f$ is $r$-smooth. We will also study the fractal dimension of the $r$-singular support of $f$. For this we recall the definition of the Hausdorff outer measure and Hausdorff dimension. Let $\eta \geq 0$, $\delta > 0$, and $A$ be any subset of $\mathbb{R}^d$. We define the $\eta$-dimensional Hausdorff outer measure as

$$H^{\eta}(A) = \lim_{\delta\to 0} \inf \left\{ \sum_{i=1}^{\infty} (\operatorname{diam}(U_i))^{\eta} : A \subset \cup_{i=1}^{\infty} U_i, \quad 0 < \operatorname{diam}(U_i) \leq \delta \right\}.$$

The Hausdorff dimension of $A$ is then given by

$$\dim_H(A) = \inf \{\eta \geq 0 : \mathcal{H}^{\eta}(A) = 0\}.$$

For a smooth shape or a shape with few corners, e.g., smooth curve or cube the Hausdorff dimension is an integer agreeing with the topological dimension. Certain irregular objects like fractals have non-integer Hausdorff dimension which describes their roughness. For example, Cantor set, which can be built by removing the central third of a line segment at each iteration creating a nowhere dense and non-countable set, has Hausdorff dimension $\log_3(2) \approx 0.63$. The Weierstrass function, which is an example of an everywhere continuous but nowhere differential function,

has Hausdorff dimension 1.5. It is also possible to construct space filling curves whose Hausdorff dimension equals to 2, see Figure 3.

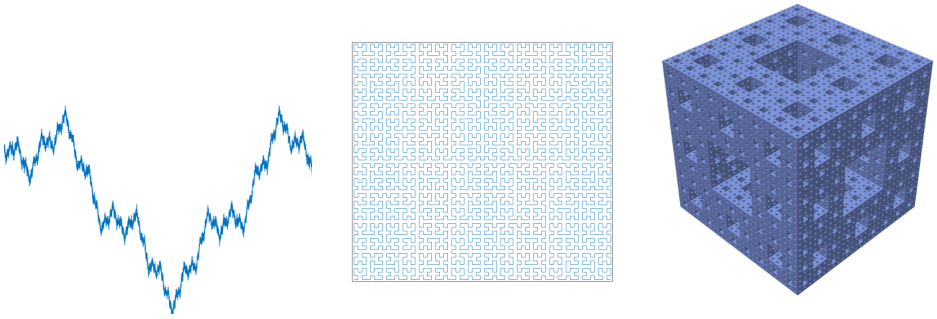

FIGURE 3. From left to right; Weierstrass function with Hausdorff dimension 1.5, sixth iteration of the space filling Hilbert curve with Hausdorff dimension 2, and Menger sponge with Hausdorff dimension $\log_3(20) \approx 2.7$.

The following theorem gives the Hausdorff dimension of the random fractal $\tau$, which is shown to coincide with the $r$ -singular support of the random variable $f$ in Theorem 3.3 below.

**Theorem 3.2.** *Let $\beta = 2^{\gamma-d}$, with $\gamma \in (-\infty, d]$, and $T$ be chosen as in Definition 3.1. If $\gamma \in (-\infty, 0]$ then $\tau$ is an empty set with probability one. If $\gamma \in (0, d]$ the set $\tau$ has Hausdorff dimension*

$$dim_H(\tau) = \gamma,$$

*with probability $1 - P_\beta$ and is empty with probability $P_\beta$, where $P_\beta$ is the solution to $P_\beta = ((1-\beta) + \beta P_\beta)^{2^d}$.*

*Proof.* This result is known from the basic theory of Galton-Watson trees, see e.g. [30, Section 8], but for the readers benefit we sketch part of the argument here in the one-dimensional case. We start by noting that $\tau$ is empty if and only if the tree $T$ terminates at some finite level, and we denote the probability for this by $P_\beta$. We will first look at the case $d = 1$. Since the nodes of the tree are chosen recursively and a new node can only be chosen if its parent node is chosen we can write

$$P_\beta = (1-\beta)^2 + 2(1-\beta)\beta P_\beta + \beta^2 P_\beta^2 = ((1-\beta) + \beta P_\beta)^2.$$

Solving the above for $P_\beta$ gives us that the probability for the tree $T$ being finite, and for $\tau$ being an empty set, is $\left(\frac{1-\beta}{\beta}\right)^2$. For a general $d$ the result follows by a similar argument. We note that if $\beta \leq 2^{-d}$ there is no solution $P_\beta < 1$ and the tree terminates almost surely at some finite level.

Consider the random process $v(j)$ which is the number of elements on $R_j = \{k : (j,k) \in T\}$. Then $v(0) = 1$ and the random variables $v(j)$, $j \geq 1$, follow binomial distributions $v(j) \sim \text{Bin}(2^d v(j-1), \beta)$. This means that $v(j)$ is the number of offspring at level $j$ in the induced Galton-Watson branching process and when $\beta > 2^{-d}$ we note that $\mathbb{E}\, v(j) = (2^d\beta)^j > 1$ for all $j \geq 1$.

We denote by $q_{jk}$ the ratio of the diameter of $Q_{jk}$ to the diameter of its parent when the parent is non-empty set, so that $q_{jk} = 1/2$ with probability $\beta$ and $q_{jk} = 0$ with probability $1 - \beta$. We then notice that

$$\mathbb{E}\Big( \sum_{k=\{0,1\}^d} q_{1k}^\gamma \Big) = 2^d \left(\frac{1}{2}\right)^\gamma \beta = 1$$

which implies that when $\tau$ is non-empty its Hausdorf dimension is $\gamma$ a.s. [40, Theorem 1.1]. □

Consider next a realisation of the random function $f$ corresponding to a realisation of $(X, T) = (X^\ell, T)_{\ell\in\mathcal{L}}$. Since the wavelets $\psi_{j,k}^\ell$ are in a Hölder space $C^r(\mathbb{R}^d)$ and the wavelet-representation is locally finite outside the closed set $\tau$, we see that

$$f|_{D\setminus\tau} \in C^r_{loc}(D \setminus \tau), \quad \text{i.e.} \quad \text{singsupp}_r(f) \subset \tau,$$

where $\text{singsupp}_r(f)$ is the $C^r$-singular support of $f$. Motivated by this, we show next that $f$ is a Besov-space $B^t_{pp}$, $t < s - \gamma/p$, valued function and that the $C^r$-singular support of $f$ is a.s. the random fractal $\tau$ corresponding to the realisation of the tree $T$.

**Theorem 3.3.** *Let $f$ be a $B^s_{pp}$-random variable with stochastic fractality index $\beta = 2^{\gamma-d}$, with $\gamma \in (0, d]$, as in Definition 3.1. Then, for all $t < s - \frac{\gamma}{p}$, $f$ takes values in $B^t_{pp}$ almost surely and $f \notin B^{s-\gamma/p}_{pp}$ on the event*

$$G := \{\omega \in \Omega \ : \ \tau \neq \emptyset\}$$

*Moreover,*

$$\mathit{singsupp}_r(f) = \tau$$

*almost surely.*

*Proof.* We consider again the random process $v(j) \sim \text{Bin}(2^d v(j-1), \beta)$ which is the number of offspring at level $j$ in the induced Galton-Watson branching process. Denote by $\bar{v} = \mathbb{E}\, v(1) = 2^d\beta > 1$, and $w(j) = v(j)/\bar{v}^j$. Next we consider the Besov norm of $f$, which according to (13) has the same distribution as

$$\|f\|^p_{B^t_{pp}} = \sum_{j=0}^\infty v(j) 2^{-jp(s-t)} \frac{\sum_{u=1}^{(2^d-1)v(j)} |Y_j^u|^p}{v(j)}, \tag{15}$$

where $Y_j^u$ are independent draws from $\mathcal{N}_p(0, \kappa^p)$.

Let us first consider the case $t < s - \frac{\gamma}{p}$. Denote $\mathbb{E}(|Y_j^u|^p) = c_p$. Then

$$\mathbb{E}\left( \frac{\sum_{u=1}^{(2^d-1)v(j)} |Y_j^u|^p}{v(j)} \right) = (2^d - 1)c_p \quad \text{and} \quad \sum_{j=0}^\infty \bar{v}^j 2^{-jp(s-t)} < \infty$$

and hence we see

$$\mathbb{E}\, \|f\|^p_{B^t_{pp}} < \infty.$$

This verifies that, for $t < s - \frac{\gamma}{p}$, we have $f \in B^t_{pp}$ almost surely.

Next we will consider the case $t = s - \frac{\gamma}{p}$. Since $\mathbb{E}\,(v(1))^2 < \infty$, the basic theory of Galton-Watson processes (see e.g. [4]) yields that the sequence $(w(j))_{j\geq 1}$ is an $L^2$-bounded martingale that has a limit $w_\infty$,

$$w_\infty = \lim_{j\to\infty} w(j)$$

satisfying

$$w_\infty \neq 0 \quad \text{a.s. on } G. \tag{16}$$

Naturally, $w_\infty = 0$ in $\Omega \setminus G$. Since $w(j)$ is then a uniformly integrable martingale we have, by Doob's theorem, $L^1(\Omega)$-convergence

$$\lim_{j\to\infty} \|w(j) - w_\infty\|_{L^1(\Omega)} = 0. \tag{17}$$

Write $Z_j := \sum_{u=1}^{v(j)(2^d-1)} |Y_j^u|^p$. Let $m_j \in \{1, \dots, 2^{jd}\}$ and consider the random variable $\widetilde{Z}_j$ that is the variable $Z_j$ conditioned on the set $\{\omega \,|\, v(j) = m_j\}$. Denoting $\mathrm{Var}(|Y_j^u|^p) = c_p'$ we get $\mathbb{E}\,(\widetilde{Z}_j) = c_p m_j$ and $\mathrm{Var}(\widetilde{Z}_j) = c_p' m_j$. Hence we can conclude

$$\mathbb{E}\,(\widetilde{Z}_j^2) = c_p' m_j + (c_p m_j)^2 \leq c_p'' m_j^2.$$

Write $A := \{\omega : \ \widetilde{Z}_j \geq c_p m_j/2\}$ and let $a = \mathbb{P}(A)$. Then the Cauchy–Schwarz inequality yields

$$\begin{aligned} c_p m_j = \mathbb{E}\,\widetilde{Z}_j &\leq (1-a) c_p m_j/2 + \int_A \widetilde{Z}_j \, d\mathbb{P} \\ &\leq (1-a) c_p m_j/2 + \Big( \int_\Omega \widetilde{Z}_j^2 d\mathbb{P} \Big)^{1/2} \Big( \int_\Omega \chi_A d\mathbb{P} \Big)^{1/2} \\ &\leq (1-a) c_p m_j/2 + m_j \sqrt{c_p'' a}. \end{aligned}$$

Thus $c_p(1+a) \leq 2\sqrt{c_p'' a}$ and we obtain that $a \geq a_0$, where $a_0 \in (0,1]$ depends only on $p$. Hence for all $m_j \in \{1, \dots, 2^{jd}\}$ it holds that

$$\mathbb{P}\big(\{Z_j \geq m_j c_p/2\} \mid v(j) = m_j\big) \geq a_0. \tag{18}$$

One may observe that we essentially reproved a lemma due to Paley-Zygmund.

Let $q$ and $m_1, m_2, \dots, m_q$ be given positive integers with $m_i \in \{1, \dots, 2^{id}\}$. This time we condition $Z_j$ on the set

$$A_{m_1,m_2,\dots,m_q} = \{\omega \in \Omega : \ v(1) = m_1, \dots, v(q) = m_q\}.$$

Notice that since the $Y_j^u$ are independent random variables, the estimate (18) holds for $j \leq q$ as well if we replace there the set $\{v(j) = m_j\}$ by $A_{m_1,m_2,\dots,m_q}$, i.e.,

$$\mathbb{P}(\{Y_j \geq m_j c_p/2\} \mid A_{m_1,m_2,\dots,m_q}) \geq a_0, \quad j \leq q. \tag{19}$$

To apply this, consider variables $W_j := Z_j/v(j)$. Observe that the variables $Z_j$, $1 \leq j \leq q$, conditioned on the set $A_{m_1,m_2,\dots,m_q}$, are independent. Thus (19) implies that

$$\mathbb{P}\bigg( \Big\{ \sum_{j=1}^q W_j < \frac{c_p q}{2} \Big\} \ \bigg| \ A_{m_1,m_2,\dots,m_q} \bigg) \leq (1-a_0)^q.$$

Let $C > 0$ and $\varepsilon > 0$ be chosen arbitrarily. Choosing $q$ so large that $q > 2Cc_p^{-1}$ and $(1-a_0)^q < \varepsilon$ we see that

$$\mathbb{P}\bigg( \Big\{ \sum_{j=1}^q W_j < M \Big\} \mid A_{m_1,m_2,\dots,m_q} \bigg) < \varepsilon.$$

This does not depend on $q$ nor $m_1, m_2, \dots, m_q$ and we deduce that

$$\mathbb{P}\bigg( \Big\{ \sum_{j=1}^\infty W_j = \infty \Big\} \ \bigg| \ G \bigg) = 1$$

which implies, in view of (16) and the the expression (15) for the Besov norm, that almost surely $\mathbb{P}(\{\|f\|^p_{B^{s-\frac{\gamma}{p}}_{pp}} = \infty\}) = 1$.

It remains to prove that $\operatorname{sing\,supp_r}(f) = \tau$ almost surely. By construction it is almost immediate that $\operatorname{sing\,supp_r}(f) \subset \tau$. Towards the other direction, given any dyadic subcube $Q$ of $D$, denote by $2Q$ the cube parallel to and with the same center as $Q$ but double the size of $Q$. Let $\varphi_Q$ denote a smooth cut-off function that is zero outside $2Q$ and one in a neighbourhood of $Q$. Because of the stochastic structure of the tree, the part of the tree (and the coefficients of our random Besov function) corresponding to $Q$ is (essentially) similar to the whole tree. Hence the above proof applies and we deduce that

$$\text{a.s.} \quad \|\varphi_Q f\|_{B^{s-\gamma/p}_{pp}} = \infty \quad \text{on the set} \quad \tau \cap \overline{Q} \neq \emptyset.$$

On the other hand, we have the equality of the following events:

$$\{\operatorname{sing\,supp_r}(f) \subsetneq \tau\} = \bigcup_Q \{\|\varphi_Q f\|_{B^r_{pp}} < \infty\} \cap \{\tau \cap \overline{Q} \neq \emptyset\},$$

and the claim follows combining these observations and the fact that the number of dyadic subcubes $Q$ is countable. □

**Corollary 3.4.** *Let $f$ be a $B^s_{pp}$-random variable with stochastic fractality index $\beta = 2^{\gamma-d}$, $\gamma \in (-\infty, d]$, as in Definition 3.1. Theorems 3.2 and 3.3 imply that for $\gamma \in (-\infty, 0]$ a.s. $f \in C^r$, and for $\gamma \in (0, d]$ $\operatorname{singsupp}_r(f)$ is an empty set with probability $P_\beta$, with $P_\beta$ being the solution of $P_\beta = ((1-\beta) + \beta P_\beta)^{2^d}$, and $\dim_H(\operatorname{singsupp}_r(f)) = \gamma$ with probability $1 - P_\beta$.*

## 4. MAP-estimate for denoising problem.

4.1. **Discrete wavelet decomposition model.** In this section we study signal and image denoising examples and show how the MAP estimator can be calculated explicitly. We start by introducing some notation for finite trees. For clarity we introduce the notation for the 1-dimensional case. The techniques readily generalise to 2-dimensional setting, and the image denoising example is introduced at the end of the section.

We define an entire finite tree as a set

$$\mathbf{T} = \{(j,k) \in \mathbb{N}^2 \mid 0 \le j \le j_{\max},\ 0 \le k \le 2^j - 1\},$$

where $j_{\max}$ is a chosen terminating depth. In practice $j_{\max}$ is is determined by the size of the signal. The discrete wavelet transform computation starts from the pixel resolution of the data and builds step by step from finer to coarser wavelet sub-bands. If the length of the signal is $n = 2^j$, we set $j_{\max} = \log_2(n)$. If the length of the signal is not $2^j$ for some $j \in \mathbb{N}$, then extra 'padding' can be added. This padding can be disregarded after the denoising has been done. We also consider finite proper subtrees

$$\Gamma = \{T \subset \mathbf{T} \,|\, \text{if } (j,k) \in T \text{ and } 1 \le j \le j_{max} \text{ then } (j-1, [k/2]) \in T\},$$

where $[k/2]$ is the integer part of $k/2$. We denote $(j',k') \trianglelefteq (j,k)$ if $(j,k) = (j',k')$ or $(j',k') \triangleleft (j,k)$ by which we mean that $(j,k)$ is an ancestor of $(j',k')$. We can then define the full subtree with a root node $(j,k)$ as

$$\mathbf{T}_{(j,k)} = \{(j',k') \in \mathbb{N}^2 \mid (j',k') \trianglelefteq (j,k)\}.$$

The size of a sub tree $\mathbf{T}_{(j,k)}$ (number of the nodes) is $s_j = 2^{j_{\max}-j+1} - 1$. The parent of a node $(j,k)$ is $P(j,k) = (j-1, [k/2])$, and its left and right child are $C_0(j,k) = (j+1, 2k)$ and $C_1(j,k) = (j+1, 2k+1)$ respectively. If the nodes $(j,k)$ and $(j',k')$ have the same parent we say that they are siblings and denote $(j',k') \sim (j,k)$.

We consider the denoising problem of recovering $f$ from

$$M = f + \mathbb{W},$$

where $\mathbb{W} = \sum w_{jk}\psi_{jk}$ with $w_{jk} \sim \mathcal{N}(0,1)$ is a white noise process independent of $f$. We employ a discrete version of the random tree Besov prior introduced in the previous section and assume that, with some $T \in \Gamma$, $f$ can be written as

$$\begin{aligned} f(x) &= \sum_{(j,k)\in T} \langle f, \psi_{jk}\rangle \psi_{jk}(x) \\ &= \sum_{(j,k)\in \mathbf{T}} \widetilde{t}_{jk} g_{jk} \psi_{jk}(x), \end{aligned} \tag{20}$$

where $g_{jk} \sim \mathcal{N}(0,1)$ or $g_{jk} \sim \text{Laplace}(0,a)$, and $\widetilde{t}_{jk} \in \{0,1\}$ defines if a node $(j,k) \in \mathbf{T}$ is chosen i.e. term $g_{jk}\psi_{jk}(x)$ is allowed to be non-zero. We assume that an independent node $t_{jk}$ is chosen with probability $\beta$, that is, $\mathbb{P}(t_{jk} = 1) = \beta$ and $\mathbb{P}(t_{jk} = 0) = 1 - \beta$. Then the sub tree $\Gamma$ contains the nodes for which $\widetilde{t}_{jk} = 1$, where $\widetilde{t}_{jk}$ is defined as

$$\widetilde{t}_{jk} = \prod_{(j',k') \trianglerighteq (j,k)} t_{j'k'}. \tag{21}$$

Notice that this means that a coefficient can only be chosen if all of its ancestors have been chosen. Otherwise the coefficient is zero. All the variables $g_{jk}, t_{jk}$ are assumed to be mutually independent.

The data can be written as $M = \sum m_{jk}\psi_{jk}(x)$, where

$$m_{jk} = \begin{cases} g_{jk} + w_{jk}, & \text{when } \widetilde{t}_{jk} = 1 \\ w_{jk}, & \text{when } \widetilde{t}_{jk} = 0. \end{cases} \tag{22}$$

The posterior distribution of $g = (g_{jk})_{(j,k)\in\mathbf{T}}$ and $t = (t_{jk})_{(j,k)\in\mathbf{T}}$, given data $m = (m_{jk})_{(j,k)\in\mathbf{T}}$ can then be written as

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

4.2. **Pruning and tree enforced soft thresholding algorithms.** In this section we will show how the MAP estimator for the above denoising problem can be calculated explicitly. If we assume that $g_{jk} \sim \mathcal{N}(0,1)$, then the result is a pruning algorithm where $\beta < 1/2$ acts as a regularisation parameter. Regularisation is achieved through turning branches of the wavelet tree on or off depending on whether they are important for the reconstruction. If we assume $g_{jk} \sim \text{Laplace}(0,a)$ instead, then the outcome is a mixture of the above-mentioned pruning algorithm and soft thresholding where the threshold is given by $a$.

We start by noting that $g_{jk}$ has density

$$\pi(g_{jk}) \propto \exp(-R(g_{jk})),$$

where $R(g_{jk}) = \frac{g_{jk}^2}{2}$ when $g_{jk} \sim \mathcal{N}(0,1)$, and $R(g_{jk}) = \frac{|g_{jk}|}{a}$ when $g_{jk} \sim \text{Laplace}(0,a)$. We define

$$z_{jk} = \pi(m_{jk} \mid g_{jk}, \widetilde{t}_{jk})\pi(g_{jk})\pi(t_{jk})$$

and note that, since the root node $(0,0)$ is always chosen,

$$z_{0\,0} = \pi(m_{0\,0} \mid g_{0\,0}, t_{0\,0} = 1)\pi(g_{0\,0}).$$

For general $(j,k) \in \mathbf{T}$ the value of $z_{jk}$ depends on whether the node is chosen or not and we denote

$$z_{jk}^1 = \pi(m_{jk} \mid g_{jk}, \widetilde{t}_{jk} = 1)\pi(g_{jk})\pi(t_{jk} = 1) = \exp\Big(-\frac{1}{2}(m_{jk} - g_{jk})^2 - R(g_{jk})\Big)\beta$$

$$z_{jk}^0 = \pi(m_{jk} \mid g_{jk}, \widetilde{t}_{jk} = 0)\pi(g_{jk})\pi(t_{jk} = 0) = \exp\Big(-\frac{1}{2}m_{jk}^2 - R(g_{jk})\Big)(1-\beta).$$

The problem of maximising $\pi(g,t \mid m)$ is equivalent to minimising $\Big(-\log\big(\prod_{(j,k)\in\mathbf{T}} z_{jk}\big)\Big)$. Write

$$\begin{aligned} F(m) &:= \min_{g,\,t}\Big(-\log\Big(\prod_{(j,k)\in\mathbf{T}} z_{jk}\Big)\Big) \\ &= \min_{g,\,t}\Big(-\log(z_{0\,0}) - \sum_{(j,k)\in\mathbf{T}\setminus(0,0)}\Big(\widetilde{t}_{jk}\log(z_{jk}^1) + (1-\widetilde{t}_{jk})\log(z_{jk}^0)\Big)\Big). \end{aligned}$$

Notice that if $t_{jk} = 0$ then $\widetilde{t}_{j'k'} = 0$ for all $(j',k') \trianglelefteq (j,k)$. This means that the whole branch is turned off and we attain the minimum

$$\min_g\Big(-\sum_{(j',k')\in\mathbf{T}_{(j,k)}}\log(z_{j'k'}^0)\Big) = \frac{1}{2}\|m|_{\mathbf{T}_{(j,k)}}\|_2^2 - s_j\log(1-\beta)$$

by choosing $g|_{\mathbf{T}_{(j,k)}} = 0$.

Denote the restriction of the function $F(m)$ to a sub tree $\mathbf{T}_{(j,k)}$ by $F_{(j,k)}(m) = F(m|_{\mathbf{T}_{(j,k)}})$. We can then rewrite the minimisation problem in the following recursive form

$$\begin{aligned} F(m) = \min_{g,\,t}\Big(&-\log(z_{0\,0}) - \sum_{k=0}^{1}\Big(t_{1\,k}\big(\log(z_{1\,k}^1) + \sum_{(j',k')\in\mathbf{T}_{(1,k)}\setminus(1,k)}\log(z_{j'k'})\big) \\ &+ (1-t_{1\,k})\sum_{(j',k')\in\mathbf{T}_{(1,k)}}\log(z_{j'k'}^0)\Big)\Big) \\ = \min_{t_{1\,0},t_{1\,1}}\Big(&-\log(z_{0\,0}) + \sum_{k=0}^{1}\Big(t_{1\,k}\big(F_{(1,k)}(m) - \log\beta\big) \\ &+ (1-t_{1\,k})\Big(\frac{1}{2}\|m|_{\mathbf{T}_{(1,k)}}\|^2 - s_1\log(1-\beta)\Big)\Big)\Big). \end{aligned}$$

The above recursion can be repeated until the lowest level $j_{max}$ is reached.

To find the minimising tree and the minimising coefficients we start from the bottom of the tree and calculate the minimising weights. Let $j = j_{\max}$. If $g_{jk} \sim \mathcal{N}(0,1)$ then

$$F_{(j,k)}(m) = \min_{g_{jk}}\Big(-\log\big(\pi(m_{jk} \mid g_{jk}, \widetilde{t}_{jk} = 1)\pi(g_{jk})\big)\Big)$$

$$= \min_{g_{jk}} \Big( \frac{1}{2}(m_{jk} - g_{jk})^2 + \frac{1}{2} g_{jk}^2 \Big) = \frac{1}{4} m_{jk}^2,$$

and the minimum is attained by $\hat{g}_{jk} = m_{jk}/2$. If $g_{jk} \sim \text{Laplace}(0, a)$ then

$$F_{(j,k)}(m) = \min_{g_{jk}} \Big( -\log\big(\pi(m_{jk} \mid g_{jk}, \widetilde{t}_{jk} = 1)\pi(g_{jk})\big)\Big)$$

$$= \min_{g_{jk}} \Big( \frac{1}{2}(m_{jk} - g_{jk})^2 + \frac{1}{a}|g_{jk}| \Big) = \begin{cases} \frac{1}{2} m_{jk}^2 & \text{when } |m_{jk}| \le \frac{1}{a} \\ \frac{1}{a}\big(m_{jk} - \frac{1}{2a}\big) & \text{when } m_{jk} > \frac{1}{a} \\ -\frac{1}{a}\big(m_{jk} + \frac{1}{2a}\big) & \text{when } m_{jk} < -\frac{1}{a} \end{cases}$$

and

$$\hat{g}_{(j,k)} = \arg\min_{g_{jk}} \Big( \frac{1}{2}(m_{jk} - g_{jk})^2 + \frac{1}{a}|g_{jk}| \Big) = \begin{cases} 0, & \text{when } |m_{jk}| \le \frac{1}{a} \\ m_{jk} - a & \text{when } m_{jk} > \frac{1}{a} \\ m_{jk} + a & \text{when } m_{jk} < -\frac{1}{a}. \end{cases}$$

We can now use the weights $F_{(j,k)}(m)$ to find the minimising tree structure. After calculating the weights for the nodes on the bottom level $j_{\max}$ we move one level up and test if using a nonzero $g_{jk}$, while paying the penalty $-\log\beta$ for doing so, gives a smaller value than setting $g_{jk}$ to zero. If choosing nonzero $g_{jk}$ produces a smaller value we put $\widetilde{t}_{jk} = 1$ otherwise $\widetilde{t}_{jk} = 0$. The same procedure is then carried out on every level. Note that $F_{(j,k)}(m)$ is the weight of a minimised tree with a root node $(j,k)$ and we only need to test it against the weight of an tree where all $g_{jk}$ are chosen to be zero. See Algorithm 1 below for case $g_{jk} \sim \mathcal{N}(0,1)$.

When $g_{jk} \sim \mathcal{N}(0,1)$ the denoising problem is only regularised by turning of branches that do not carry enough information. If $g_{jk} \sim \text{Laplace}(0, a)$ regularisation is achieved through soft thresholding, with threshold $1/a$, and by excluding branches with small wavelet coefficients. This allows us to combine the benefits of soft thresholding and the pruning algorithm described above.

We also note that due to the recursive nature of the algorithm its computational complexity is linear in the number of the wavelet coefficients. This can be seen by observing that in the algorithm one performs four "for-to" or "while" loops where a finite number of operations are performed once in all vertexes of the tree. The number of the vertexes of the tree is equal to the number of the wavelet coefficients that is of the same order as the length $n$ of the given signal.

4.2.1. *Some results for general linear inverse problems.* We will next show that when a wavelet branch does not carry enough information it will be turned off also in the case of a general linear forward operator $A$. We assume the random tree Besov prior introduced in (20). We can then write

$$\pi(f \mid M) \propto \pi(M - Af \,|\, f)\pi(g)\pi(t)$$

$$= \exp\Big( -\frac{1}{2}\|M - Af\|^2 \Big) \exp\big( -R(g)\big)\beta^{n(t)}(1-\beta)^{(s_1 - n(t))}\Big)$$

where $n(t)$ is the number of elements in a tree $\Gamma = \{(j,k) \in \mathbf{T} \mid \widetilde{t}_{jk} = 1\}$. We are interested in the minimisation problem

$$\min_f \big( -\log\pi(f \mid M)\big) = \min_{g,\,t} \Big( \frac{1}{2}\|M - Af\|^2 + R(g) - n(t)\log\beta - (s_1 - n(t))\log(1-\beta)\Big).$$

We will next show that if the measurement $m|_{T_{(j,k)}}$ has small enough $\ell^2$ norm in some $T_{(j,k)}$ the subtree will be turned off.

Input: $j_{\max}$, wavelet coefficients $m_{jk}$ of the measurement, parameter $\beta$
Output: wavelet coefficients $g_{jk}$ of the fractal-denoised signal

**for** $k = 0$ **to** $2^{j_{\max}} - 1$ **do**
  $F_{(j_{\max},k)} = \frac{1}{4}|m_{j_{\max}k}|^2$;
**end**
$j = j_{\max} - 1$;
**while** $1 \le j \le j_{\max} - 1$ **do**
  $s = 2^{j_{\max}-j} - 1$;
  **for** $k = 0$ **to** $2^j - 1$ **do**
    $$F_{(j,k)} = \frac{1}{4}|m_{jk}|^2;$$
    $j', k' \leftarrow$ indices of the left child of leaf $j, k$
    **if** $F_{(j',k')} - \log\beta < \frac{1}{2}\|m|_{\boldsymbol{T}_{(j',k')}}\|^2 - s\log(1-\beta)$ **then**
      $t_{j',k'} = 1$; $F_{(j,k)} = F_{(j,k)} + F_{(j',k')} - \log\beta$;
    **else**
      $t_{j',k'} = 0$; $F_{(j,k)} = F_{(j,k)} + \frac{1}{2}\|m|_{\mathbf{T}_{(j',k')}}\|^2 - s\log(1-\beta)$;
    **end**
    $j'', k'' \leftarrow$ indices of the right child of leaf $j, k$
    **if** $F_{(j'',k'')} - \log\beta < \frac{1}{2}\|m|_{\boldsymbol{T}_{(j'',k'')}}\|^2 - s\log(1-\beta)$ **then**
      $t_{j'',k''} = 1$; $F_{(j,k)} = F_{(j,k)} + F_{(j'',k'')} - \log\beta$;
    **else**
      $t_{j'',k''} = 0$; $F_{(j,k)} = F_{(j,k)} + \frac{1}{2}\|m|_{\mathbf{T}_{(j'',k'')}}\|^2 - s\log(1-\beta)$;
    **end**
  **end**
  $j = j - 1$;
**end**

$\widetilde{t}_{00} = 1$;
**for** $j = 1$ **to** $j_{\max}$ **do**
  **for** $k = 0$ **to** $2^j - 1$ **do**
    $j''', k''' \leftarrow$ indices of the parent of leaf $j, k$
    $\widetilde{t}_{jk} = t_{jk}\widetilde{t}_{j'''k'''}$;
  **end**
**end**

**for** $j = 0$ **to** $j_{\max}$ **do**
  **for** $k = 0$ **to** $2^j - 1$ **do**
    **if** $\widetilde{t}_{jk} = 1$ **then**
      $g_{jk} = m_{jk}/2$;
    **else**
      $g_{jk} = 0$;
    **end**
  **end**
**end**

**Algorithm 1:** Pseudocode for finding the minimising $t$ and $g$ recursively. In practical calculations we found that replacing $g_{jk} = m_{jk}/2$ by $g_{jk} = m_{jk}$ gives better results.

**Lemma 4.1.** *Let $M = Af + \mathbb{W}$, where $\mathbb{W}$ is a centred Gaussian noise process. We assume a random tree Besov prior defined in (20). If $\|m|_{T_{(j,k)}}\|^2 < \varepsilon$, where $0 < \varepsilon < \min\{\log(1/\beta - 1), 1\}$, $0 < \beta < 1/2$ then the tree maximising the posterior $\pi(f \mid M)$ is a subset of $\boldsymbol{T} \setminus \boldsymbol{T}_{(j,k)}$.*

*Proof.* The result follows directly from the fact that if $\|m\|^2 < \varepsilon$ the tree that maximises the posterior is an empty tree $T = \emptyset$. For an empty tree $n(t) = 0$ and we have

$$\min_g \big( -\log \pi(f \mid M)\big) = \frac{1}{2}\|m\|^2 - s_1 \log(1-\beta).$$

On the other hand, if $n(t) \geq 1$ we can write

$$\begin{aligned}\min_{g,t} \big( -\log \pi(f \mid M)\big) &\geq \min_t \big( -n(t)\log\beta - (s_1 - n(t))\log(1-\beta)\big) \\ &> \min_t \big(n(t)(\varepsilon - \log(1-\beta)) - (s_1 - n(t))\log(1-\beta)\big) \\ &> \|m\|^2 - s_1 \log(1-\beta)\end{aligned}$$

which concludes the proof. □

4.3. **Signal and image denoising examples.** In our first example we consider the blocks test data from [22] which is displayed in Figure 4. We assume that only a noisy signal, with noise ratio 3, is observed and want to denoise it. We employ semi-Gaussian random tree Besov prior with Haar wavelets, in which case the MAP estimator is given by the pruning algorithm. We also tested denoising the signal with several Matlab denoising packages of which hard thresholding performed the best. Figure 5 shows that the pruning algorithm performs better than hard thresholding. The $\ell^2$ error and root mean squared error between the original and our denoised signal are 9.5 and 0.21 respectively. For the thresholded signal the errors are 13.7 and 0.30. The denoised signals and the corresponding wavelet trees are shown in Figure 5 and one can clearly see how choosing a proper tree model is beneficial for the reconstruction.

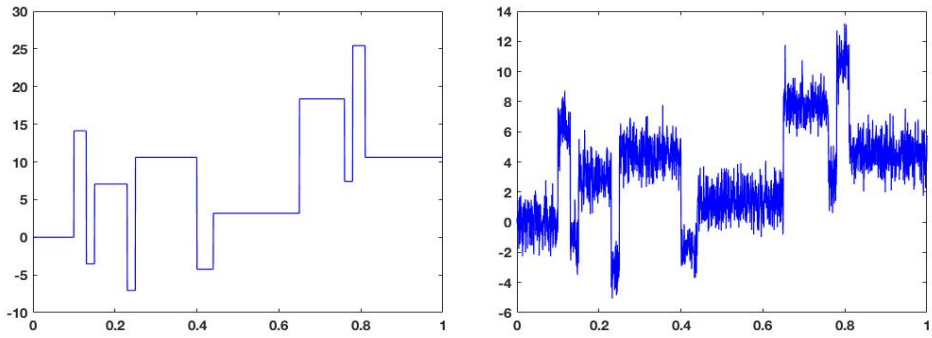


FIGURE 4. Original blocks signal and the observed noisy signal with signal to noise ratio 3.

In Figure 6 we present some prior draws from semi-Gaussian random tree Besov priors with Haar wavelets with different wavelet densities $\beta$. As expected, when $1/2 < \beta < 1$ is small the realisations are mostly flat with some small areas of large jumps, while with large $\beta$ the realisations are more erratic.

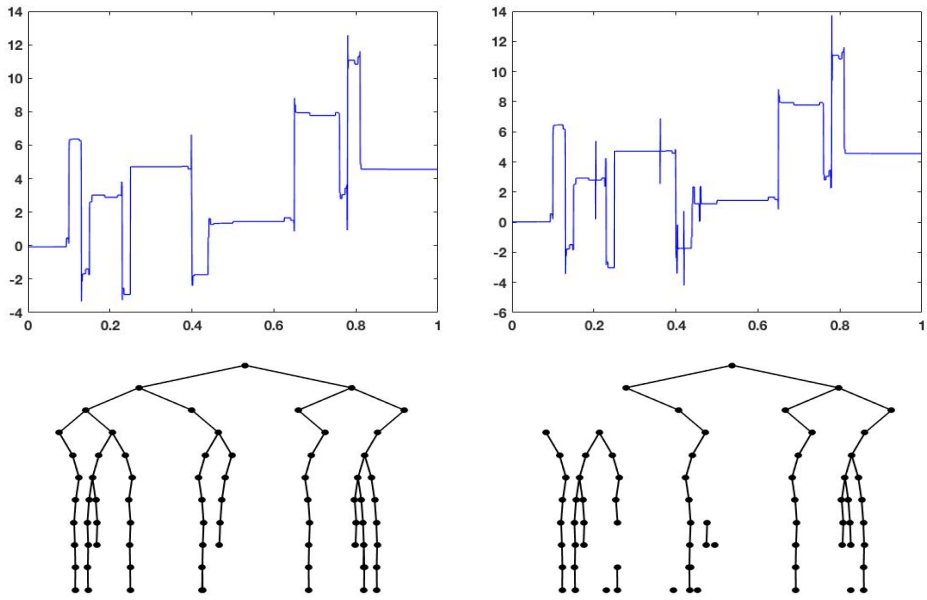

FIGURE 5. The denoised signal using pruning algorithm on top left and the denoised signal that was attained using hard thresholding on top right. Below them are the wavelet trees corresponding to the estimators.

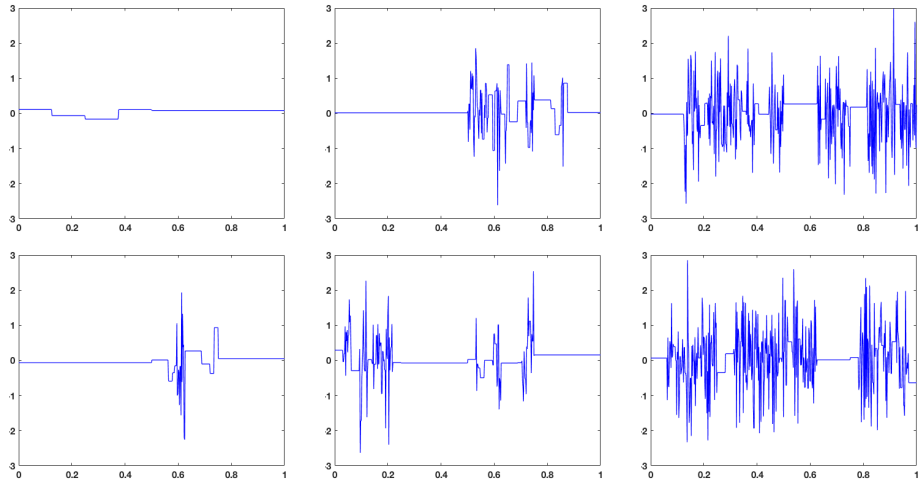

FIGURE 6. Prior draws from the 1-dimensional semi-Gaussian random tree Besov prior with Haar wavelets. The wavelet densities are $\beta = 0.6$ on the left, $\beta = 0.75$ on the middle and $\beta = 0.9$ on the right.

Next we consider real accelerometer data collected from a human subject with a wearable device. The accelerometer erroneously records slight movement even

when the device is still and our aim is to use the pruning algorithm with Haar wavelets to denoise the signal. The original and denoised signal are presented in Figure 7. The pruning algorithm returns a denoised signal where the areas of large acceleration have been left untouched while the parts where the device was still have been cleaned and are piecewise constant as one would expect.

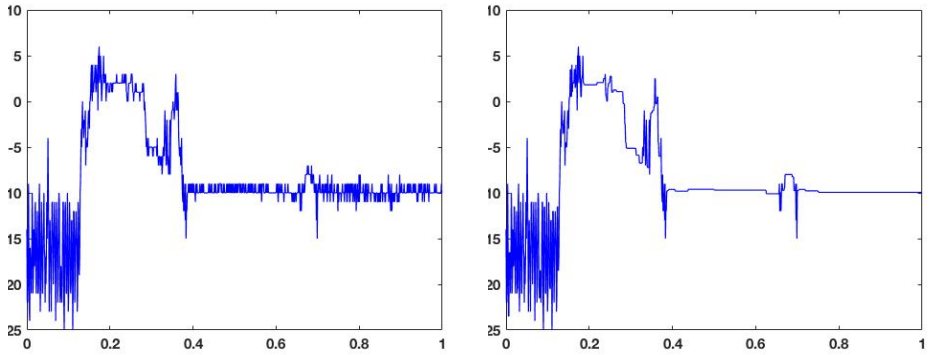


FIGURE 7. Measured accelerometer data on left and the denoised signal on right.

In our final example we study image denoising. The original and noisy image are shown in Figure 8. We consider the MAP estimators arising from semi-Gaussian and semi-Laplace random tree Besov priors with Daubechies 2 wavelets, which are given by pruning and tree enforced soft thresholding algorithms respectively. The image has been extended by mirroring to avoid boundary effects, and cut down to the original size after using the algorithm. In the tree enforced soft thresholding and soft thresholding algorithms the thresholding is only done for the deeper levels $j \geq 5$ containing finer details.

We use two error measures to quantify the reconstruction quality. The first one is the peak signal-to-noise ratio (PSNR) which is an expression for the ratio between the maximum possible value of an image and the power of distorting noise that affects the quality of its representation. The PSNR is given by $10 \log_{10}(R^2/MSE)$, where MSE is the squared error between the denoised and the original image, and $R$ is the maximum possible pixel value of the image. Higher values of PSNR indicate better reconstruction quality. The second error measure we use is the structural similarity index measure (SSIM) which measures the perceived change in structural information by assessing the visual impact of three characteristics of an image: luminance, contrast and structure. The higher the SSIM, the better the quality of the reconstructed image.

As expected the tree enforced soft thresholding performs better than the pruning algorithm since it allows two type of regularisation; pruning and soft thresholding. The PSNR (using the original image as a reference) for the pruned image and for the denoised image given by the tree enforced soft thresholding are 23.0 and 23.7 respectively. The SSIM are 0.69 and 0.71 respectively. We also denoised the image using soft thresholding with Daubechies 2 wavelets and the PSNR for the best achieved reconstruction is 23.6 and SSIM is 0.69. The three reconstructions are presented in Figure 9. As one can see from the close-ups of the images the tree

FIGURE 8. Original sharp image and image with Gaussian noise with variance 0.01.

FIGURE 9. From left to right; Denoised images and close-ups attained using pruning, tree enforced soft thresholding and soft thresholding algorithms. From the close-ups we can clearly see that the tree based methods denoise smooth areas better than soft thresholding.

based algorithms are better at denoising smooth areas like the leaves and the nose of the koala.

Random draws from the semi-Gaussian prior with Daubechies 2 wavelets and different wavelet densities can be found in Figure 10.

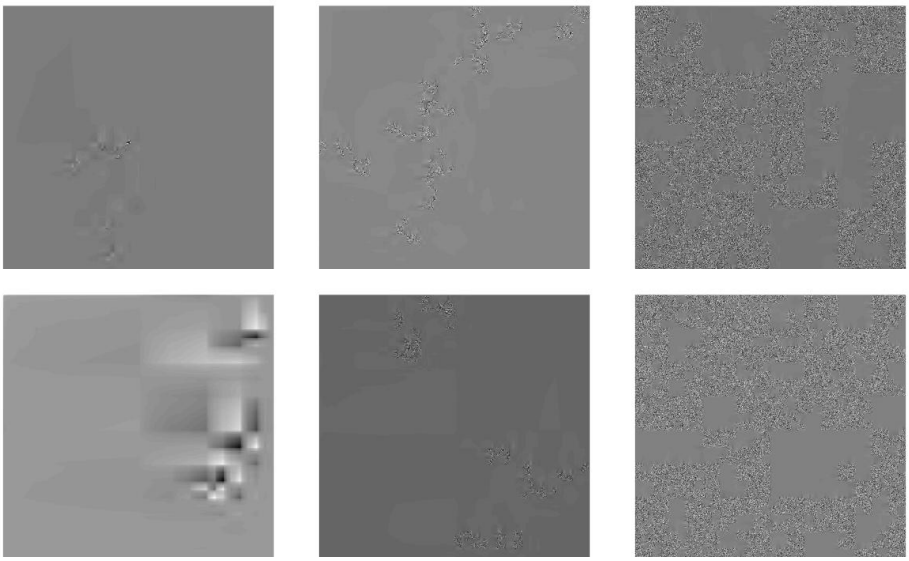

FIGURE 10. Prior draws from the 2-dimensional semi-Gaussian random tree Besov prior with Daubechies 2 wavelets. The wavelet densities are $\beta = 0.3$ on the left, $\beta = 0.6$ on the middle and $\beta = 0.9$ on the right.

**Acknowledgments.** H.K. acknowledges partial support by Cantab Capital Institute for the Mathematics of Information (RG83535). M.L. and S.S. were supported by the Academy of Finland through the Finnish Centre of Excellence in Inverse Modelling and Imaging 2018-2025, decision number 312339. E.S was supported by the Finnish Academy grant 1309940.

## REFERENCES

[1] F. Abramovich, T. Sapatinas and B. W. Silverman, Wavelet thresholding via a Bayesian approach, *J. R. Stat. Soc. Ser. B Stat. Methodol.*, **60** (1998), 725-749.

[2] S. Agapiou, M. Dashti and T. Helin, Rates of contraction of posterior distributions based on p-exponential priors, *Bernoulli*, **27** (2021), 1616-1642.

[3] A. Alman, L. Johnson, D. Calverley, G. Grunwald, D. Lezotte and J. Hokanson, Diagnostic capabilities of fractal dimension and mandibular cortical width to identify men and women with decreased bone mineral density, *Osteoporosis International*, **23** (2012), 1631-1636.

[4] K. B. Athreya and P. E. Ney, *Branching Processes*, Dover Publications, Inc., Mineola, NY, 2004.

[5] R. M. Brown, Global uniqueness in the impedance imaging problem for less regular conductivities, *SIAM Journal on Mathematical Analysis*, **27** (1996), 1049-1056.

[6] C. B. Caldwell, S. J. Stapleton, D. W. Holdsworth, R. A. Jong, W. J. Weiser, G. Cooke and M. J. Yaffe, Characterisation of mammographic parenchymal pattern by fractal dimension, *Physics in Medicine & Biology*, **35** (1990), 235.

[7] D. Calvetti and E. Somersalo, A Gaussian hypermodel to recover blocky objects, *Inverse Problems*, **23** (2007), 733-754.
[8] D. Calvetti and E. Somersalo, Hypermodels in the Bayesian imaging framework, *Inverse Problems*, **24** (2008), 034013, 20 pp.
[9] I. Castillo, Pólya tree posterior distributions on densities, *Ann. Inst. Henri Poincaré Probab. Stat.*, **53** (2017), 2074-2102.
[10] I. Castillo and R. Mismer, Spike and slab pólya tree posterior densities: Adaptive inference, *Ann. Inst. Henri Poincaré Probab. Stat.*, **57** (2021), 1521-1548.
[11] I. Castillo and T. Randrianarisoa, Optional polya trees: Posterior rates and uncertainty quantification, (2021), arXiv preprint, arXiv:2110.05265.
[12] I. Castillo and V. Rockova, *Multiscale Analysis of Bayesian CART*, University of Chicago, Becker Friedman Institute for Economics Working Paper, 2019.
[13] A. Chambolle, R. A. DeVore, N. yong Lee and B. J. Lucier, Nonlinear wavelet image processing: Variational problems, compression, and noise removal through wavelet shrinkage, *IEEE Transactions on Image Processing*, **7** (1998), 319-335.
[14] A. Chan and J. A. Tuszynski, Automatic prediction of tumour malignancy in breast cancer with fractal dimension, *Royal Society Open Science*, **3** (2016), 160558.
[15] S. G. Chang, B. Yu and M. Vetterli, Adaptive wavelet thresholding for image denoising and compression, *IEEE Trans. Image Process.*, **9** (2000), 1532-1546.
[16] M. Dashti, S. Harris and A. Stuart, Besov priors for Bayesian inverse problems, *Inverse Probl. Imaging*, **6** (2012), 183-200.
[17] M. Dashti, K. J. Law, A. M. Stuart and J. Voss, MAP estimators and their consistency in Bayesian nonparametric inverse problems, *Inverse Problems*, **29** (2013), 095017, 27 pp.
[18] M. Dashti and A. M. Stuart, The Bayesian approach to inverse problems, *Handbook of Uncertainty Quantification, Springer, Cham*, **1, 2, 3** (2017), 311-428.
[19] I. Daubechies, *Ten Lectures on Wavelets (Ninth printing, 2006)*, BMS-NSF Regional Conference Series in Applied Mathematics, 61. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 1992.
[20] I. Daubechies, M. Defrise and C. De Mol, An iterative thresholding algorithm for linear inverse problems with a sparsity constraint, *Comm. Pure Appl. Math.*, **57** (2004), 1413-1457.
[21] M. Davis and H. Li, Evaluation of fractal dimension for mixing and combustion by the schlieren method, *Experiments in Fluids*, **2**1 (1996), 248-258.
[22] D. L. Donoho and I. M. Johnstone, Ideal spatial adaptation via wavelet shrinkage, *Biometrika*, **81** (1994), 425-455.
[23] A. V. Dryakhlov and A. A. Tempelman, On Hausdorff dimension of random fractals, *New York J. Math.*, **7** (2001), 99-115.
[24] M. Giordano and K. Ray, Nonparametric Bayesian inference for reversible multidimensional diffusions, *Ann. Statist.*, **50** (2022), 2872-2898.
[25] C. Gómez, Á. Mediavilla, R. Hornero, D. Abásolo and A. Fernández, Use of the Higuchi's fractal dimension for the analysis of MEG recordings from Alzheimer's disease patients, *Medical Engineering & Physics*, **31** (2009), 306-313.
[26] H. Haario, M. Laine, M. Lehtinen, E. Saksman and J. Tamminen, Markov chain monte carlo methods for high dimensional inversion in remote sensing, *J. R. Stat. Soc. Ser. B Stat. Methodol.*, **66** (2004), 591-607.
[27] T. Helin, On infinite-dimensional hierarchical probability models in statistical inverse problems, *Inverse Probl. Imaging*, **3** (2009), 567-597.
[28] T. Helin and M. Lassas, Hierarchical models in statistical inverse problems and the mumford–shah functional, *Inverse Problems*, **27** (2011), 015008, 32 pp.
[29] M. Jansen, *Noise Reduction by Wavelet Thresholding*, Lecture Notes in Statistics, 161. Springer-Verlag, New York, 2001.
[30] S. Karlin and H. E. Taylor, *A First Course in Stochastic Processes*, Second edition, Academic Press, New York-London, 1975.
[31] S. Lasanen, Discretizations of generalized random variables with applications to inverse problems, *Ann. Acad. Sci. Fenn. Math. Diss.*, (2002), 64 pp.
[32] S. Lasanen, Measurements and infinite-dimensional statistical inverse theory, *PAMM: Proceedings in Applied Mathematics and Mechanics, Wiley Online Library*, **7** (2007), 1080101-1080102.
[33] M. Lassas, E. Saksman and S. Siltanen, Discretization-invariant Bayesian inversion and Besov space priors, *Inverse Problems and Imaging*, **3** (2009), 87-122.

[34] M. Lassas and S. Siltanen, Can one use total variation prior for edge-preserving Bayesian inversion?, *Inverse Problems*, **20** (2004), 1537-1563.
[35] M. Lehtinen, L. Päivärinta and E. Somersalo, Linear inverse problems for generalised random variables, *Inverse Problems*, **5** (1989), 599-612.
[36] A. Lejay and P. Pigato, A threshold model for local volatility: Evidence of leverage and mean reversion effects on historical data, *Int. J. Theor. Appl. Finance*, **22** (2019), 1950017, 24 pp.
[37] H. Li, M. L. Giger, O. I. Olopade and L. Lan, Fractal analysis of mammographic parenchymal patterns in breast cancer risk assessment, *Academic Radiology*, **14** (2007), 513-521.
[38] B. B. Mandelbrot, The inescapable need for fractal tools in finance, *Annals of Finance*, **1** (2005), 193-195.
[39] ———, Parallel cartoons of fractal models of finance, *Annals of Finance*, **1** (2005), 179-192.
[40] R. D. Mauldin and S. C. Williams, Random recursive constructions: Asymptotic geometric and topological properties, *Trans. Amer. Math. Soc.*, **295** (1986), 325-346.
[41] Y. Meyer, *Wavelets and Operators*, Cambridge Studies in Advanced Mathematics, 37. Cambridge University Press, Cambridge, 1992.
[42] P. Pigato, Extreme at-the-money skew in a local volatility model, *Finance and Stochastics*, **23** (2019), 827-859.
[43] M. Reiß, Asymptotic equivalence for nonparametric regression with multivariate and random design, *Annals of Statistics*, **36** (2008), 1957-1982.
[44] E. Ribak, C. Schwartz and G. Baum, Fractal wave fronts: Simulation and prediction for adaptive optics, *European Southern Observatory Conference and Workshop Proceedings*, **48**, (1994), 205.
[45] L. Sendur and I. W. Selesnick, Bivariate shrinkage functions for wavelet-based denoising exploiting interscale dependency, *IEEE Transactions on Signal Processing*, **50** (2002), 2744-2756.
[46] A. M. Stuart, Inverse problems: A Bayesian perspective, *Acta Numerica*, **19** (2010), 451-559.
[47] G. N. Thomas, S.-Y. Ong, Y. C. Tham, W. Hsu, M. L. Lee, Q. P. Lau, W. Tay, J. Alessi-Calandro, L. Hodgson, R. Kawasaki, et al., Measurement of macular fractal dimension using a computer-assisted program, *Investigative Ophthalmology & Visual Science*, **55** (2014), 2237-2243.
[48] H. Triebel, *Theory of Function Spaces*, Monographs in Mathematics, 78. Birkhäuser Verlag, Basel, 1983.
[49] H. Triebel, *Theory of Function Spaces. II*, Monographs in Mathematics, 84. Birkhäuser Verlag, Basel, 1992.
[50] H. Triebel, *Function Spaces and Wavelets on Domains*, EMS Tracts in Mathematics, 7. European Mathematical Society (EMS), Zürich, 2008.
[51] H. Triebel, *Theory of Function Spaces IV*, Monographs in Mathematics, 107. Birkhäuser/Springer, Cham, 2020.
[52] B. J. West, *Fractal Physiology and Chaos in Medicine*, Studies of Nonlinear Phenomena in Life Science, 16. World Scientific Publishing Co. Pte. Ltd., Hackensack, NJ, 2013.

Received May 2021; 1st revision November 2021; 2nd revision July 2022; early access November 2022.