Wavelet shrinkage in nonparametric regression models with positive noise

Alex Rodrigo dos S. Sousa and Nancy Lopes Garcia

Department of Statistics, State University of Campinas, Campinas, Brazil

ABSTRACT

Wavelet shrinkage estimators are widely applied to analyse datasets in wavelet domain from several fields of science. They typically act by reducing magnitudes of empirical coefficients in a discrete wavelet transformation to estimate wavelet coefficients. In nonparametric regression problems, most of the shrinkage rules are derived from models composed of an unknown function with additive Gaussian noise. Although Gaussian noise assumption is reasonable in several real data analysis, mainly for large sample sizes, it is not general. Contaminated data with positive noise can occur in practice and nonparametric regression models with positive noise bring challenges in wavelet shrinkage point of view. In this sense, this work proposes bayesian shrinkage rules to estimate wavelet coefficients from a nonparametric regression framework with additive and strictly positive noise under exponential and lognormal distributions. Computational aspects are discussed and simulation studies to analyse the performances of the proposed shrinkage rules and compare them with standard techniques are done. An application for winning times Boston Marathon dataset is also provided.

1. Introduction

Wavelet-based methods have been applied in several fields of statistics such as time series modelling, functional data analysis, computational methods and nonparametric regression, for example, and their success can be justified by several mathematical and computational reasons. In nonparametric regression, the focus of this work, it is possible to represent an unknown squared integrable function in orthogonal wavelet basis, which are composed of dilations and translations of a specified function $\psi$ that is called wavelet function or mother wavelet. A wavelet function satisfies some properties, such as its integral is equals to zero. There are several wavelet functions and examples of them are shown in Figure 1 for Daubechies wavelet functions with one (Haar or Daub1), two (Daub2), four (Daub4) and ten (Daub10) null moments.

The wavelet representation of a function allows us the visualization of data in different resolution levels and to perform a multiresolution analysis by the application of a discrete
Wavelet coefficients are essentially sparse at smooth locations of the function, but in practice, after the application of the discrete wavelet transformation on the data, the empirical coefficients are contaminated with random noise. Thus, to estimate the wavelet coefficients by denoising the empirical ones, thresholding and shrinkage methods are usually applied on the empirical coefficients to reduce their magnitudes. There are several nonlinear thresholding and shrinkage methods available in the literature, most of them are based on the seminal works of Donoho [5,6], Donoho [7,8], Donoho and Johnstone [9,10] and Donoho and Johnstone [11], with the proposition of the so-called soft and hard thresholding rules. Bayesian shrinkage procedures have also been successfully proposed in the last years. These methods allow the incorporation of prior information regarding to the coefficients, such as their sparsity, support, dispersion and extreme values by means of a prior probabilistic distribution. In this context, the proposed priors are usually composed of a mixture of a highly concentrated distribution around zero to assign sparsity and a symmetric distribution around zero. Proposed prior distributions to the wavelet coefficients include mixtures of normals by Chipman et al. [12], mixtures of a point mass function at
zero and double exponential distribution by Vidakovic and Ruggeri [13], Bickel prior by Angelini and Vidakovic [14], double Weibull by Reményi and Vidakovic [15], Dirichlet-Laplace priors by Bhattacharya et al. [16] and, recently, logistic and beta priors by Sousa [17] and Sousa et al. [18] respectively. For a general overview about wavelet shrinkage and thresholding techniques, see Jansen [19].

Although the standard thresholding and bayesian shrinkage methods are well succeeded in general, most of them suppose that the data points from the underlying function are contaminated with additive normal random noise. Despite the normality assumption that can occur in practice and implies in several good estimation properties, it is not general mainly under small sample sizes, where central limit theorem cannot be applied. In this sense, little attention has been given for wavelet denoising problems in nonparametric regression models under non-normal random noise or, specifically, additive strictly positive random noise.

Neumann and von Sachs [20] discuss normal approximations to the wavelet empirical coefficients for thresholding without the normality supposition of the noises and independent and identically distributed (iid) assumption of them. Leporini and Pesquet [21] proposed the use of Besov priors on the wavelet coefficients to derive a bayesian thresholding rule under a possible resolution level dependent generalized normal distributed noise in the wavelet domain. Antoniadis et al. [22] provided explicit bayesian thresholding rules based on the Maximum a Posteriori (MAP) estimation procedure under exponential power distribution prior on the wavelet coefficients and supposing exponential power and Cauchy distributions to the noise in the wavelet domain. Averkamp and Houdré [23] analysed the ideal denoising in the sense of Donoho and Johnstone [11] by considering some classes of noise, including identically distributed symmetric around zero noises in the wavelet domain. Thresholding under compactly support noises in the wavelet domain is also discussed. Thus, the above cited works dealt with non-Gaussian noise but, no one of them assumes positive noise in the original model. Further, the noise distribution assumptions occur directly in the wavelet domain, after the discrete wavelet transform application on the original data.

In this sense, this paper proposes bayesian shrinkage procedures to be applied on empirical wavelet coefficients in nonparametric regression models with strictly positive random noise contamination in the original data. We assume that random noises are iid with exponential or lognormal distribution. The adopted priors are the mixture of a point mass function at zero and the logistic prior proposed by Sousa [17] and beta prior proposed by Sousa et al. [18], both works were developed under the classical gaussian noise structure.

Additive and positive random noise in the original nonparametric model brings several challenges in estimation point of view. First, independent noises property is lost after wavelet transformation, i.e., noises in the wavelet domain are possibly correlated. The consequence of this fact is that the wavelet coefficient estimation can not be performed individually as usually done under gaussian noise assumption, but jointly by a joint posterior distribution of the wavelet coefficients vector, which requires computational methods, such as Markov Chain Monte Carlo (MCMC) methods to sample from the joint posterior distribution. Further, noises in the wavelet domain are not necessarily positive, but only linear combinations of them. Finally, several statistical models with multiplicative positive noise were proposed and dealt with by logarithmic transformations, but models with additive positive noise are not so common in the literature, although additive positive noise can
be observed in a wide variety of real measurements. For example, arrival times of radio or waves measures typically contain positive errors due to possibly delays of equipment detection. See Radnosrati et al. [24] for an interesting study of the classical estimation theory of models with additive positive noise and a nice application involving global navigation satellite systems (GNSS) with positive noise arrival times.

Thus, the main novelty of this work is to perform wavelet shrinkage under additive positive noise in the original nonparametric model. To do so, mixtures of a point mass function at zero and logistic or beta priors are assigned to the wavelet coefficients. Logistic prior is suitable for coefficients with support in the Real set. Its scale hyperparameter has easy and direct interpretation in terms of shrinkage, as can be seen in Sousa [17]. The beta prior [18] is a good choice for bounded coefficients and its well-known shape flexibility brings advantages in modelling.

This paper is organized as follows: the considered statistical models are defined in Section 2 and their associated shrinkage rules with computational aspects are described in Section 3. Parameters and hyperparameters choices are discussed in Section 4. Simulation studies to obtain the performances of the shrinkage rules and to compare them with standard shrinkage/thresholding techniques are analysed in Section 5. A real data application involving winning times of the Boston Marathon is done in Section 6. The paper is concluded with final considerations in Section 7.

2. Statistical models

We consider \( n = 2^J, J \in \mathbb{N} \), points \((x_1, y_1), \ldots, (x_n, y_n)\) from the nonparametric regression model

\[
y_i = f(x_i) + e_i, \quad i = 1, \ldots, n \tag{1}
\]

where \( f \in L^2(\mathbb{R}) = \{f : \int f^2 < \infty\} \) is an unknown function and \( e_i \)'s are independent and identically distributed (iid) random noises such that \( e_i > 0, i = 1, \ldots, n \). The goal is to estimate \( f \) without assumptions about its functional structure, i.e., the estimation procedure will take only the data points into account. In this work, we consider random noise with exponential and lognormal distributions. Their densities are given by

- **Exponential distributed noise:** \( e_i \sim \text{Exp}(\lambda) \)

\[
h(e_i; \lambda) = \lambda \exp\{-\lambda e_i\} 1_{(0,\infty)}(e_i), \quad \lambda > 0, \tag{2}
\]

- **Lognormal distributed noise:** \( e_i \sim \text{LN}(0, \sigma) \)

\[
h(e_i; \sigma) = \frac{1}{e_i \sigma \sqrt{2\pi}} \exp\left\{ -\frac{\log^2(e_i)}{2\sigma^2} \right\} 1_{(0,\infty)}(e_i), \quad \sigma > 0, \tag{3}
\]

where \( 1_A(\cdot) \) is the usual indicator function on the set \( A \) and \( \log(\cdot) \) is the natural logarithm. We suppose both the noise distribution parameters \( \lambda \) and \( \sigma \) as known, although a brief discussion for the unknown case is provided in Section 4.
The unknown function $f$ can be represented by

$$f(x) = \sum_{j,k \in \mathbb{Z}} \theta_{j,k} \psi_{j,k}(x),$$  \hspace{1cm} (4)

where $\{\psi_{j,k}(x) = 2^{j/2} \psi(2^j x - k), j, k \in \mathbb{Z}\}$ is an orthonormal wavelet basis for $L^2(\mathbb{R})$ constructed by dilations $j$ and translations $k$ of a function $\psi$ called wavelet or mother wavelet and $\theta_{j,k}$ are wavelet coefficients that describe features of $f$ at spatial location $2^{-j} k$ and scale $2^j$ or resolution level $j$. In this context, the data points $(x_1, y_1), \ldots, (x_n, y_n)$ can be viewed as an approximation of $f$ at the finest resolution level $J$ with additive and positive noise contamination. As an example, Figure 2 displays a Donoho-Johnstone (D-J) test function called Blocks, which will be defined in Section 5, and $n = 1024 = 2^{10}$ data points generated from this function with additive exponential distributed random noises.

The estimation process of $f$ is done by the estimation of the wavelet coefficients. In vector notation, model (1) can be written as

$$y = f + e,$$  \hspace{1cm} (5)

where $y = [y_1, \ldots, y_n]'$, $f = [f(x_1), \ldots, f(x_n)]'$ and $e = [e_1, \ldots, e_n]'$. A discrete wavelet transform (DWT), which is typically represented by an orthonormal transformation matrix $W_{n \times n} = (w_{ij})_{1 \leq i, j \leq n}$, is applied on both sides of (5), which gives us the following model in wavelet domain

$$d = \theta + e,$$  \hspace{1cm} (6)

where $d = Wy$ is called empirical coefficients vector, $\theta = Wf$ is the wavelet coefficients vector and $e = We$ is the random noises vector. Although $W$ is used as DWT representation, fast algorithms are applied to perform DWT in practice, which are more computationally efficient, see Mallat [4]. When $e_i$'s are assumed to be iid normal distributed in model (1), the distribution of the random noises in wavelet domain remains normal, i.e.,
\( \epsilon_i \)'s are iid normal with the same scale parameter of the noises in the time domain. This property brings several estimation advantages, once the problem of estimating \( \theta \) in this context is equivalent of estimating a location parameter of a normal distribution. Moreover, as the noises in wavelet domain remain independent, the estimation of \( \theta \) can be done individually. When \( \epsilon_i \)'s are positive, however, most of these advantages are lost. Actually, \( \epsilon_i \)'s are correlated, not necessarily positive and their distribution is not the same as their counterparts in time domain. The main impact of these facts is that the estimation of \( \theta \) cannot be performed individually, but according to a joint posterior distribution of \( \theta \).

The wavelet coefficients vector \( \theta \) is estimated by application of a shrinkage rule \( \delta(d) \) on the empirical coefficients vector \( d \). This procedure essentially acts on the observed coefficients by reducing their magnitudes in order to estimate the wavelet coefficients. After the estimation \( \hat{\theta} = \delta(d) \), \( f \) is estimated by the inverse discrete wavelet transform (IDWT), \( \hat{f} = W^t \hat{\theta} \).

In this work, we apply a bayesian shrinkage procedure by assuming prior distributions to a single wavelet coefficient \( \theta \) (the subindices are dropped by simplicity). The priors have the general structure

\[
\pi(\theta; \alpha, \eta) = \alpha \delta_0(\theta) + (1 - \alpha) g(\theta; \eta),
\]

for \( \alpha \in (0, 1) \), \( \delta_0(\cdot) \) is the point mass function at zero and \( g(\cdot; \eta) \) is a probability distribution defined according to a hyperparameters vector \( \eta \). The choice of \( g(\cdot; \eta) \) can be made according to the support of \( \theta \). We consider in this work two quite flexible distributions \( g(\cdot; \eta) \), the symmetric around zero logistic distribution proposed by Sousa [17] that is given by

\[
g(\theta; \tau) = \frac{\exp \{-\theta \tau\}}{\tau \left(1 + \exp \{-\theta \tau\}\right)^2} \mathbb{I}_{\mathbb{R}}(\theta), \quad \tau > 0,
\]

and the beta distribution on the interval \([-m, m]\) proposed by Sousa et al. [18] that is given by

\[
g(\theta; a, b, m) = \frac{(\theta + m)^{a-1}(m - \theta)^{b-1}}{(2m)^{a+b-1}B(a, b)} \mathbb{I}_{[-m, m]}(\theta), \quad a, b, m > 0,
\]

where \( B(\cdot, \cdot) \) is the beta function. Sousa [17] and Sousa et al. [18] developed shrinkage rules under logistic and beta priors respectively under the standard Gaussian noise framework. Figure 3(a,b) shows logistic and beta densities for several hyperparameters values respectively. The beta densities are considered on interval \([-3, 3]\).

The logistic prior centred at zero is suitable in bayesian wavelet shrinkage for real-valued wavelet coefficients, i.e., when \( \theta \in \mathbb{R} \). Further, its hyperparameter \( \tau \) has an important role in determining the degree of shrinkage to be applied on the empirical coefficients, as described in Sousa [17]. The beta prior offers great flexibility in modelling bounded wavelet coefficients, i.e., when \( \theta \in [-m, m] \), once it allows symmetric (\( a = b \)) and asymmetric (\( a \neq b \)) distributions around zero. As the logistic prior, its hyperparameters \( a \) and \( b \) control the amount of shrinkage of the associated bayesian rule. For \( b = a \), bigger values of \( a \) imply the increase of the shrinkage level imposed on the data, i.e., the associated rule tends to a severe reduction of the empirical coefficients’ magnitudes. More details about beta priors on wavelet coefficients can be found in Sousa et al. [18]. Thus, logistic and beta
priors are convenient choices for \( g \) in (7) to model several prior information about the wavelet coefficients to be estimated, such as their support, symmetry and sparsity.

### 3. Shrinkage rules and computational aspects

The general shrinkage rules \( \delta \) associated to the models (1), (2), (3), (6) and (7) under squared loss function are obtained by the posterior expected value, i.e., \( \delta(d) = [\delta_1(d), \ldots, \delta_n(d)]' = \mathbb{E}_{\pi}(\theta | d) \). Once it is infeasible to obtain the posterior expected value analytically, we use an adaptive Markov Chain Monte Carlo (MCMC) method to be described later to generate \( L \) samples \( \theta_1, \theta_2, \ldots, \theta_L \) from the joint posterior distribution \( \pi(\cdot | d) \) of \( \theta \) and estimate a particular wavelet coefficient \( \theta_i \) by the sample mean,

\[
\hat{\theta}_i = \delta_i(d) \approx \frac{1}{L} \sum_{l=1}^{L} \theta_{li},
\]

where \( \theta_{li} \) is the \( i \)th element of the generated sample \( \theta_l, l = 1, \ldots, L \) and \( i = 1, \ldots, n \).

We use the Robust Adaptive Metropolis (RAM) algorithm proposed by Vihola [25] and implemented computationally in the adaptMCMC R package by Scheiddegger [26] to obtain samples of the joint posterior distribution of \( \theta \). The algorithm estimates the shape of the target distribution \( \pi(\cdot | d) \) and simultaneously coerces the mean acceptance rate of the process. For each iteration of the chain generation, a single shape matrix \( S \) is adaptively updated. Let \( S_1 \in \mathbb{R}^{n \times n} \) be a lower-diagonal matrix with positive diagonal elements, \( \{\eta_l\}_{l \geq 1} \subset (0, 1] \) be a sequence decaying to zero, \( \gamma \in (0, 1) \) be the target mean acceptance rate and \( \theta_1 \) such that \( \pi(\theta_1 | d) > 0 \), the RAM algorithm works according to Algorithm 1.

We applied \( \eta_l = \min\{1, n l^{-2/3}\} \) and \( \gamma = 0.234 \) as suggested by Vihola [25] along the simulation studies and real data application to obtain the posterior distribution samples of the wavelet coefficients. The next subsections provide the posterior distributions that are considered as target distributions in RAM algorithm.

#### 3.1. Posterior distributions under exponential noise

Under the exponential noise model (1), (2) and the model after DWT application (6), it is straightforward to obtain the likelihood function of the empirical coefficients \( \mathcal{L}(d | \theta) \) by
Algorithm 1: Robust Adaptive Metropolis (RAM) algorithm.

**Data:** $S_1, \theta_1, \{\eta_l\}_{l \geq 1} \subset (0, 1), \gamma \in (0, 1)$ and $L$

**Result:** Samples $\theta_1, \ldots, \theta_L$ from $\pi(\theta | d)$

1. Set $l = 2$. Generate $\theta^*_l = \theta_{l-1} + S_{l-1} U_l$, where $U_l \sim N_n(0, I_n)$ and $I_n$ is the identity matrix of dimension $n \times n$. Do $\theta_l = \theta^*_l$ with probability

   $$\gamma_l = \min \left(1, \frac{\pi(\theta^*_l | d)}{\pi(\theta_{l-1} | d)}\right),$$

   or $\theta_l = \theta_{l-1}$ else. Compute the lower diagonal matrix $S_l$ with positive diagonal elements satisfying the equation

   $$S_l S_{l-1}^T = S_{l-1} \left(I + \eta_l (\gamma_l - \gamma) \frac{U_l U_l^T}{\|U_l\|^2}\right) S_{l-1}^T.$$

If $l \leq L$, return to step 2.

The application of the Jacobian method to the transformation $d = \theta + W e$. The likelihood function is given by

$$L(d | \theta) = |W|^\lambda \exp \left\{-\lambda \sum_i \sum_j w_{ji} (d_j - \theta_j) \right\} \prod_i I_{(0, \infty)} \left(\sum_j w_{ji} (d_j - \theta_j)\right).$$ (11)

The posterior distribution of $\theta | d$ can be obtained by the well-known relationship

$$\pi(\theta | d) \propto \pi(\theta) L(d | \theta).$$ (12)

Thus, after the application of (12) for (11) and the logistic prior model (7) and (8), we have the following posterior distribution to the wavelet coefficients given the empirical ones under a logistic prior model and exponential noise on the original data,

$$\pi(\theta | d) \propto \prod_i \left[ \alpha \delta_0(\theta_i) + (1 - \alpha) \frac{\exp \left\{-\frac{\theta_i}{\tau}\right\}}{\tau \left(1 + \exp \left\{-\frac{\theta_i}{\tau}\right\}\right)^2} \right] \times \exp \left\{-\lambda \sum_i \sum_j w_{ji} (d_j - \theta_j) \right\} \times \prod_i I_{(0, \infty)} \left(\sum_j w_{ji} (d_j - \theta_j)\right).$$ (13)

Similarly, we can have the posterior distribution of $\theta | d$ under beta prior model and exponential noise on the original data by considering now (9) instead of (8). The posterior
distribution is given by

\[
\pi(\theta \mid d) \propto \prod_i \left[ \alpha \delta_0(\theta_i) + (1 - \alpha) \frac{(\theta_i + m)^{a-1} (m - \theta_i)^{b-1}}{(2m)^{a+b-1} B(a, b)} \right] \\
\times \exp \left\{ -\lambda \sum_i \sum_j w_{ji}(d_j - \theta_j) \right\} \\
\times \prod_i \mathbb{I}_{[-m,m]}(\theta_i) \times \prod_i \mathbb{I}_{(0,\infty)} \left( \sum_j w_{ji}(d_j - \theta_j) \right).
\] (14)

### 3.2. Posterior distributions under lognormal noise

The likelihood function of the empirical coefficients for the model under lognormal noise (1), (3) and the model after DWT application (6) is obtained as described in Subsection 3.1 and given by

\[
\mathcal{L}(d \mid \theta) = \frac{|W|}{(\sigma \sqrt{2\pi})^n \prod_i \left[ \sum_j w_{ji}(d_j - \theta_j) \right]} \times \exp \left\{ -\frac{1}{2\sigma^2} \sum_i \log^2 \left( \sum_j w_{ji}(d_j - \theta_j) \right) \right\} \\
\times \prod_i \mathbb{I}_{(0,\infty)} \left( \sum_j w_{ji}(d_j - \theta_j) \right).
\] (15)

Thus, the posterior distribution of \( \theta \mid d \) under lognormal noise in the original data and logistic prior model (7) and (8) is obtained by application of (12) for the likelihood function (15) and given by

\[
\pi(\theta \mid d) \propto \prod_i \left[ \alpha \delta_0(\theta_i) + (1 - \alpha) \frac{\exp \left\{ -\frac{\theta_i}{\tau} \right\}}{\tau \left( 1 + \exp \left\{ -\frac{\theta_i}{\tau} \right\} \right)^2} \right] \\
\times \exp \left\{ -\frac{1}{2\sigma^2} \sum_i \log^2 \left( \sum_j w_{ji}(d_j - \theta_j) \right) \right\} \\
\times \prod_i \mathbb{I}_{(0,\infty)} \left( \sum_j w_{ji}(d_j - \theta_j) \right),
\] (16)
Algorithm 2: Function points and wavelet coefficients estimation procedures under exponential noise and shrinkage rule under logistic prior

Data: $y$, $W$, $\lambda$, $\tau$ and $L$.

Result: Estimated wavelet coefficients vector $\hat{\theta} = [\hat{\theta}_1, \cdots, \hat{\theta}_n]'$ and function points $\hat{f} = [\hat{f}(x_1), \cdots, \hat{f}(x_n)]'$. 

1. Do $d = Wy$ (DWT application). Obtain $\pi(\theta \mid d)$ according to (13). Generate $L$ samples $\theta_1, \cdots, \theta_L$ from $\pi(\theta \mid d)$ according to Algorithm 1 (RAM algorithm). Do $\hat{\theta}_i = \frac{1}{L} \sum_{l=1}^{L} \theta_{li}, i = 1, \cdots, n$. Do $\hat{f} = W^t\hat{\theta}$ (IDWT application).

The posterior distribution of $\theta \mid d$ under beta prior model (7) and (9) is

$$
\pi(\theta \mid d) \propto \prod_i \left[ \alpha \delta_0(\theta_i) + (1 - \alpha) \frac{(\theta_i + m)^{a-1}(m - \theta_i)^{b-1}}{(2m)^{a+b-1}B(a, b)} \right]
\times \exp \left\{ -\frac{1}{2\sigma^2} \sum_i \log^2 \left( \sum_j w_{ji}(d_j - \theta_j) \right) \right\}
\times \prod_i I_{[-m,m]}(\theta_i) \times \prod_i I_{(0,\infty)} \left( \sum_j w_{ji}(d_j - \theta_j) \right). \tag{17}
$$

Therefore, the posterior distributions (13) and (14) of $\theta \mid d$ are the considered target distributions under logistic and beta prior models respectively in RAM algorithm to be sampled to estimate the wavelet coefficients by the shrinkage rule (10) for original data contaminated by exponential noise. Similarly, the posterior distributions (16) and (17) are the target ones under logistic and beta priors respectively for lognormal noise-contaminated observations. Algorithm 2 resumes the proposed function points and wavelet coefficients estimation procedures under exponential noise and shrinkage rule under logistic prior application. The procedures under lognormal noise and shrinkage rule under beta prior are similar, with suitable changes in the noise parameter and hyperparameters.

4. Parameters elicitation

The performance of the bayesian procedure is closely related to a good choice or estimation of the involved parameters and hyperparameters of the models. The proposed shrinkage rules depend on the parameters $\lambda$ and $\sigma$ of the noise exponential and lognormal distributions respectively, which were considered as known throughout the paper, the weight $\alpha$ of the point mass function of the prior models and the hyperparameters $\tau$ and $(a, b, m)$ of the logistic and beta priors respectively.
Angelini and Vidakovic [14] proposed the hyperparameters $\alpha$ and $m$ be dependent on the resolution level $j$ according to the expressions

$$\alpha = \alpha(j) = 1 - \frac{1}{(j - J_0 + 1)^r},$$

(18)

$$m = m(j) = \max_k |d_{jk}|,$$

(19)

where $J_0 \leq j \leq J - 1$, $J_0$ is the primary resolution level, $J$ is the number of resolution levels, $J = \log_2(n)$ and $r > 0$. They also suggest that in the absence of additional information, $r = 2$ can be adopted.

The choices of the hyperparameters $\tau$ and $(a, b)$ are discussed respectively by Sousa [17] and Sousa et al. [18]. In fact, their values have a direct impact on the shrinkage level of the associated rule. Higher denoising level on empirical coefficients requires higher values of $\tau$ and $(a, b)$. Moreover, these hyperparameters can be resolution level dependent, such as $\alpha$ and $m$. As default values, $\tau = a = b = 5$ can be used. Further discussion about how to choose $(a, b)$ of a beta prior distribution can also be seen in Chaloner and Duncan [27] and Duran and Booker [28].

The noise distribution parameters $\lambda$ and $\sigma$ of exponential and lognormal respectively, although considered as known, can be included in the Bayesian framework, independently of the wavelet coefficients, by attributing suitable priors to them, such inverse gamma prior for example. In this case, the general prior model (7) under exponential noise could be updated by

$$\pi(\theta, \lambda; \alpha, \eta, \zeta) = \pi(\theta; \alpha, \eta) \times \pi(\lambda; \zeta)$$

(20)

where $\pi(\lambda; \zeta)$ is the prior distribution of $\lambda$ and $\zeta$ is its hyperparameter vector. Analogous procedure can be done for the lognormal noise case.

5. Simulation studies

The performances of the proposed shrinkage rules were obtained in simulation studies and compared against standard shrinkage/thresholding techniques. The so-called Donoho-Johnstone (D-J) test functions [11] were considered as underlying functions to be estimated, which are composed of four test functions called Bumps, Blocks, Doppler and Heavisine defined on $[0, 1]$ by,

- **Bumps**

$$f(x) = \sum_{l=1}^{11} h_l K \left( \frac{x - x_l}{w_l} \right),$$

where

$$K(x) = (1 + |x|)^{-4};$$

$$(x_l)_{l=1}^{11} = (0.1, 0.13, 0.15, 0.23, 0.25, 0.40, 0.44, 0.65, 0.76, 0.78, 0.81);$$

$$(h_l)_{l=1}^{11} = (4, 5, 3, 4, 5, 2.1, 4.3, 3.1, 5.1, 4.2) \text{and}$$

$$(w_l)_{l=1}^{11} = (0.005, 0.005, 0.006, 0.01, 0.01, 0.03, 0.01, 0.01, 0.005, 0.008, 0.005).$$
Blocks

\[ f(x) = \sum_{l=1}^{11} h_l K(x - x_l), \]

where

\[ K(x) = (1 + \text{sgn}(x))/2; \]

\[(x_l)_{l=1}^{11} = (0.1, 0.13, 0.15, 0.23, 0.25, 0.40, 0.44, 0.65, 0.76, 0.78, 0.81) \text{ and} \]

\[(h_l)_{l=1}^{11} = (4, -5, 3, -4, 5, -4.2, 2.1, 4.3, -3.1, 2.1, -4.2). \]

Doppler

\[ f(x) = \sqrt{x(1-x)} \sin \left( \frac{2.1\pi}{x + 0.05} \right). \]

Heavisine

\[ f(x) = 4 \sin(4\pi x) - \text{sgn}(x - 0.3) - \text{sgn}(0.72 - x). \]

The functions are presented in Figure 4. In fact, the D-J functions have important local features to be recovered in the wavelet shrinkage process. Bumps function has spikes with different magnitudes, Blocks function is piecewise constants and requires the detection of the discontinuity locations, Doppler function is composed of oscillations and Heavisine function has two discontinuities points and a cusp. These local features usually occur in practice and are not well recovered by other basis function representations such as splines and their extensions. On the other hand, local features can be well localized in wavelet domain through significant wavelet coefficients. Moreover, smooth regions of the functions are represented in a sparse way by the coefficients.

For a particular test function, data were generated by adding exponential and lognormal noises to the function points according to two signal-to-noise ratio (SNR) values, SNR = 3 and 9 and two sample sizes, \( n = 32 \) and 64. Further, the noise parameters \( \lambda \) and \( \sigma \) of exponential and lognormal distributions respectively in the simulation were set according to the defined SNR values. Since values of equally spaced points of the Donoho-Johnstone functions have the same standard deviation, i.e., the signal standard deviation value is the same for the four functions, the noise parameters were \( \lambda = 0.4286 \) and \( \sigma = 1.0296 \) for data generation with SNR = 3 and \( \lambda = 1.2857 \) and \( \sigma = 0.5949 \) for SNR = 9.

Each scenario of underlying function, SNR and sample size data generation was replicated \( M = 100 \) times and the averaged mean squared error (AMSE) was calculated as a performance measure for each method, which is given by

\[
\text{AMSE} = \frac{1}{Mn} \sum_{m=1}^{M} \sum_{i=1}^{n} [\hat{f}^{(m)}(x_i) - f(x_i)]^2,
\]

where \( \hat{f}^{(m)}(\cdot) \) is the estimate of the function at a particular point in the \( m \)th replication, \( m = 1, \ldots, M = 100 \) by a specific method. Under this performance measure, the best method of a given scenario has the smallest AMSE. For each replication, \( L = 10,000 \) samples of the posterior distributions (13), (14), (16) and (17) were obtained by RAM algorithm.
and the proposed shrinkage rules under logistic (LOGISTIC) and beta (BETA) priors were calculated by (10). The hyperparameters were $a = b = 5$ and $m$ according to (19) for beta prior and $\tau = 5$ for logistic prior. In both rules, $\alpha$ was calculated by (18).

The performances of the proposed shrinkage rules under logistic and beta priors were compared against four extensively used shrinkage methods of thresholding type, Universal thresholding (UNIV) proposed by Donoho and Johnstone (1994), Cross-Validation (CV) proposed by Nason [29], False Discovery Rate (FDR) proposed by Abramovich and Benjamini [30] and Stein Unbiased Risk Estimator (SURE) proposed by Donoho and Johnstone [11]. These standard methods allow the determination of optimal threshold values ($\xi > 0$) in such a way that empirical wavelet coefficients smaller than these threshold values are set to zero, according to the soft thresholding rule of Donoho [7],

$$\delta(d) = \begin{cases} 0, & \text{if } |d| \leq \xi \\ \text{sgn}(d)(|d| - \xi), & \text{if } |d| > \xi \end{cases}$$

Note that the soft thresholding rule acts coefficient by coefficient, as opposed to the proposed shrinkage rules, which act in the entire empirical coefficients vector. See Vidakovic [1] for a full description of the standard wavelet thresholding methods.

### 5.1. Simulation under exponential noise

Table 1 shows the AMSEs of the shrinkage and thresholding rules under exponential noise simulated data. In fact, the proposed shrinkage rules had great performances in terms of AMSE in almost all the scenarios. The shrinkage rule under logistic prior was the best estimator for all the scenarios with sample size $n = 32$ and for most of the times when $n = 64$, being the best estimator in general. The shrinkage rule under beta prior was the
Table 1. AMSE of the shrinkage/thresholding rules in the simulation study for DJ-test functions under exponential noise.

| Signal   | n  | Method | SNR = 3 | SNR = 9 | Signal   | n  | Method | SNR = 3 | SNR = 9 |
|----------|----|--------|---------|---------|----------|----|--------|---------|---------|
| Bumps    | 32 | UNIV   | 18.721  | 2.882   | Blocks   | 32 | UNIV   | 17.631  | 3.292   |
|          |    | CV     | 38.439  | 23.175  |          |    | CV     | 21.504  | 15.457  |
|          |    | FDR    | 31.603  | 12.530  |          |    | FDR    | 21.684  | 16.227  |
|          |    | SURE   | 30.872  | 6.287   |          |    | SURE   | 21.841  | 16.211  |
|          |    | LOGISTIC | 7.069  | 0.787   |          |    | LOGISTIC | 5.960  | 0.748   |
|          |    | BETA   | 7.081   | 1.140   |          |    | BETA   | 6.542   | 0.769   |
|          |    | CV     | 28.317  | 9.140   |          |    | CV     | 24.277  | 16.021  |
|          |    | FDR    | 20.496  | 4.562   |          |    | FDR    | 21.586  | 7.864   |
|          |    | SURE   | 12.325  | 1.718   |          |    | SURE   | 24.277  | 8.419   |
|          |    | LOGISTIC | 8.449  | 1.028   |          |    | LOGISTIC | 8.303  | 1.033   |
|          |    | BETA   | 8.408   | 1.110   |          |    | BETA   | 8.903   | 1.022   |
| Doppler  | 32 | UNIV   | 11.977  | 1.881   | Heavisine| 32 | UNIV   | 7.374   | 1.146   |
|          |    | CV     | 12.795  | 3.573   |          |    | CV     | 7.429   | 1.150   |
|          |    | FDR    | 17.121  | 4.993   |          |    | FDR    | 7.564   | 1.161   |
|          |    | SURE   | 11.207  | 1.312   |          |    | SURE   | 7.526   | 1.148   |
|          |    | LOGISTIC | 6.422  | 0.834   |          |    | LOGISTIC | 6.373  | 0.779   |
|          |    | BETA   | 8.488   | 1.109   |          |    | BETA   | 8.410   | 0.995   |
|          |    | CV     | 12.566  | 3.556   |          |    | CV     | 6.436   | 1.004   |
|          |    | FDR    | 13.281  | 2.517   |          |    | FDR    | 6.460   | 1.019   |
|          |    | SURE   | 10.735  | 1.235   |          |    | SURE   | 6.439   | 1.046   |
|          |    | LOGISTIC | 8.230  | 1.031   |          |    | LOGISTIC | 8.194  | 1.045   |
|          |    | BETA   | 9.780   | 1.124   |          |    | BETA   | 9.736   | 1.145   |

best for Bumps function, SNR = 3 and n = 64 and Blocks, SNR = 9 and n = 64. Even when beta shrinkage rule was not the best one, its performance was close to the logistic rule in general. Then, the shrinkage rule under beta prior had the second-best general performance.

Moreover, the proposed rules worked much better against the standard rules in some of the cases. For example, the AMSEs of logistic and beta rules for Bumps function, SNR = 9 and n = 32, were 0.787 and 1.140 respectively. The third best estimator in that scenario was SURE, with AMSE = 6.287, almost 8 times the AMSE of logistic rule. Only for heavisine function and n = 64 we did not have the proposed rules as the best ones. They were outperformed by UNIV and CV methods, but even in these cases, their performances were close to these ones. It should also be noted the good behavior of the proposed rules for low signal to noise ratio, i.e., for SNR = 3, which is evidence of good work for high noise datasets. In fact, for the eight scenarios of SNR = 3, the shrinkage rule under logistic prior had the best AMSE in six of them and the rule under beta prior was the best in one scenario.

Figure 5 presents the estimates obtained by the shrinkage rule under logistic prior for n = 64 and SNR = 9. The main features of each test function were captured by the estimates, such as spikes of Bumps, piecewise constant regions of Blocks, oscillations of Doppler and the discontinuity points of Heavisine function. Boxplots of the mean squared errors (MSE) of the replications are also provided in Figure 6 and show low variation of the proposed shrinkage rules MSEs in all the scenarios. The standard methods however had high variability in terms of MSE in the replications. For instance, CV method the highest variation for Bumps and Doppler functions while SURE method had considerably high MSE variance for Blocks function.
Figure 5. Estimates of the D-J test functions by the shrinkage rule under logistic prior in the simulation study for \( n = 64 \), SNR = 9 and for simulated points under exponential noise.

Figure 6. Boxplots of the mean square errors (MSE) of the shrinkage and thresholding rules in the simulation study for \( n = 64 \), SNR = 9 and for simulated points under exponential noise. The associated rules are: 1-UNIV, 2-CV, 3-FDR, 4-SURE, 5-LOGISTIC and 6-BETA.

Finally, Table 2 provides the averaged running times of the proposed shrinkage rules according to SNR and sample size values. As expected, the sample size had an impact on the running time of the proposed rules. For \( n = 32 \), the shrinkage rule under logistic prior took 4.78 s in average by replication for scenarios with SNR = 3 and 4.61 s for SNR =
Table 2. Averaged running times (in seconds) of the proposed shrinkage rules under logistic and beta priors in the simulation studies under exponential noise.

| n  | Method  | SNR = 3 | SNR = 9 |
|----|---------|---------|---------|
| 32 | LOGISTIC | 4.78    | 4.61    |
|    | BETA    | 4.75    | 4.49    |
| 64 | LOGISTIC | 9.12    | 9.21    |
|    | BETA    | 8.95    | 9.11    |

Notes: The averaged running times of the standard methods were less than 0.01 s.

9. The shrinkage rule under beta prior took 4.75 s in average by replication for SNR = 3 and 4.49 s in average for SNR = 9. For n = 64, the averaged running times of the logistic rule were 9.12 and 9.21 s for SNR = 3 and 9, respectively and the beta rule took 8.95 and 9.11 s for scenarios with SNR = 3 and SNR = 9, respectively. Thus, it was not observed a significant difference in the averaged running times between the proposed shrinkage rules. Further, the SNR values did not have an impact on the running times.

The standard methods run instantaneously, i.e., they are very fast. However, it should be noted that these methods act coefficient by coefficient and do not require MCMC applications. On the other hand, the proposed shrinkage rules work on the entire empirical coefficients vector and need to perform MCMC algorithm to obtain samples of the joint posterior distributions of the wavelet coefficients vector. Although the proposed methods run considerably slower than the standard ones, they are built under positive noises (in opposite to the standard methods) and had better general performances in terms of AMSE measure. Moreover, their running times are not prohibitive for the considered sample sizes under computational and practical points of view.

5.2. Simulation under lognormal noise

The obtained results for simulated data under lognormal noise are available in Table 3. In general, the shrinkage rule under logistic prior had the best performance in terms of AMSE and outperformed the other estimators in practically all scenarios with SNR = 9. The rule under beta prior also presented good performance in this context, with AMSEs close to the logistic rule ones and was the best for Blocks function, n = 64.

The proposed shrinkage rules however did not have good performances in scenarios with low signal-to-noise ratio (SNR = 3), where they were outperformed by the standard methods. After a deeper investigation of these results, we found out that the proposed shrinkage rules did not perform enough denoising in the generated datasets under SNR = 3. Thus, it would be necessary that the proposed rules shrink the empirical coefficients in a more severe way. It can be achieved by changing the hyperparameters values of the logistic and beta priors, once they control the shrinkage level of the rules. For instance, when a = b = 8 and τ = 10 for beta and logistic priors respectively, we observed a great improvement in the performance of the associated rules in terms of AMSE under SNR = 3.

Although logistic rule was the best in general, it should be observed that the behaviors of the standard rules under lognormal noise were better in general than themselves under exponential noise. For example, considering data with SNR = 3, SURE was the best
for Bumps and Doppler underlying functions, while UNIV was the best one for Blocks and Heavisine. Under exponential noise, these rules were dominated by the proposed estimators for these same functions and scenarios.

Figure 7 shows the estimates of the D-J functions by the shrinkage rule under logistic prior, for $n = 64$ and SNR = 9. As occurred in the exponential noise context, the estimates captured well the main characteristics of the test functions. Boxplots of the MSEs are shown in Figure 8, where it is possible to note low MSE variation for the proposed shrinkage rules and high variations for CV, FDR and SURE methods.

The averaged running times of the proposed rules under lognormal noise are shown in Table 4. As in the exponential noise context, the sample size had a great impact on the running times and SNR did not. For $n = 32$, the logistic rule averaged running times were equal to 7.04 and 7.42 s by replication in scenarios under SNR = 3 and SNR = 9, respectively. The beta rule was faster than logistic one, with 6.64 and 6.38 s in average for datasets with SNR = 3 and 9, respectively. As expected, for $n = 64$, the running times were higher than the ones for $n = 32$. Logistic rule run in 10.17 and 9.92 s and beta rule run in 9.61 and 9.11 s in average by replication for SNR = 3 and 9, respectively. The standard methods averaged running times were practically equals to zero, as in the exponential noise case.

In general, the proposed shrinkage rules under logistic and beta priors had good performances in the simulation studies under exponential and lognormal noises. In fact, the proposed rules outperformed in terms of AMSE the standard methods in most of the scenarios under exponential noise (for both sample sizes and SNR values) and under lognormal noise in datasets with SNR = 9. Further, the proposed rules worked in a reasonable averaged running times, although higher than the standard methods.
Figure 7. Estimates of the D-J test functions by the shrinkage rule under logistic prior in the simulation study for \( n = 64, \text{SNR} = 9 \) and for simulated points under lognormal noise.

The hyperparameters values impact on the shrinkage level of the rule, i.e., in how severely the rule shrinks the empirical coefficients in the estimation (denoising) process. In the simulation studies, the hyperparameters choices allowed a suitably shrinkage action of the rules in most of the scenarios, once the estimated functions recovered well the local features of the original underlying Donoho-Johnstone test functions. Finally, it should be noted that the proposed work assumes known noise parameters. Although unknown noise parameters contexts can be overcome under the bayesian framework by (20) for instance, the performances of the shrinkage rules might be influenced by the noise parameters estimation procedure.

6. Real data application

Boston Marathon is one of the most important marathon of the world. It occurs yearly since 1897 with a trajectory of 42,195 m between Hopkinton and Boston cities, at US Massachusetts state. As mentioned in the introduction, arrival times measurements are classical examples of positive noise contamination due to possible delays of detection by instruments. Further, although the presence of positive noise in the data, the contamination level is usually small in this context due to the high precision of the instrument. In this sense, it is reasonable to consider a high signal-to-noise ratio in the dataset.

In this sense, we applied the proposed shrinkage rules with logistic and beta priors under exponential noise assumption to denoise \( n = 64 \) winning times (in minutes) of Boston Marathon Men’s Open Division from 1953 to 2016. The data is publicly available on Boston Athletic Association (BAA) webpage https://www.baa.org/races/boston-marathon/results/champions. We used a DWT with Daub10 basis and the prior hyperparameters were adopted according to (18) for \( \sigma \) in both shrinkage rules, \( \tau = 5 \) in the logistic shrinkage rule, \( a = b = 5 \) and \( m \) according to (19) in the beta shrinkage rule. For comparison purposes, we also applied the cross-validation thresholding to denoise the time series.
Figure 8. Boxplots of the mean square errors (MSE) of the shrinkage and thresholding rules in the simulation study for $n = 64$, $SNR = 9$ and for simulated points under lognormal noise. The associated rules are 1-UNIV, 2-CV, 3-FDR, 4-SURE, 5-LOGISTIC and 6-BETA.

Figure 9 shows original and denoised data by the shrinkage rules with logistic and beta priors under exponential noise and by the cross-validation thresholding rule. As expected, the denoised winning times under logistic and beta shrinkage rules are less than or equal to the measured ones, depending on the shrinkage level. Due to the good precision of measured times for this competition, the shrinkage rules did not reduce the empirical coefficients severely, i.e., it was not necessarily for a high shrinkage level. On the other hand, the original time series was denoised excessively by the cross-validation thresholding rule, thus its denoised time series version lost some important features of the data, such as peak magnitudes. Further, there are denoised winning times under cross validation rule greater than the original measures, which it is expected since standard methods are not built to deal with positive noise.

The empirical wavelet coefficients (represented by vertical bars) by resolution level and the differences between them and the estimated coefficients by the shrinkage rule under logistic prior, $d - \hat{\theta}$, are shown in Figure 10(a,b) respectively. It is possible to note that, although residuals in original data are positive, which can be seen in Figure 11(a), their counterparts in the wavelet domain are not necessarily positive, i.e. there are estimated coefficients bigger than their respective empirical ones.

Finally, Figure 11(b) presents the histogram (with area equals to 1) of the residuals in time domain of the denoised time series by the shrinkage rule under logistic prior, i.e., $y - \hat{y}$, with a superposed exponential density curve, for $\hat{\lambda} = n/\sum_i(y_i - \hat{y}_i) = 3.987$, the maximum likelihood estimate. In fact, the one-sample Kolmogorov-Smirnov test for exponential distribution with $\lambda = 3.987$ of the residuals provided a $p$-value $= 0.7057$, not
Table 4. Averaged running times (in seconds) of the proposed shrinkage rules under logistic and beta priors in the simulation studies under lognormal noise.

| n  | Method | SNR = 3 | SNR = 9 |
|----|--------|---------|---------|
| 32 | LOGISTIC | 7.04    | 7.42    |
|    | BETA    | 6.64    | 6.38    |
| 64 | LOGISTIC | 10.17   | 9.92    |
|    | BETA    | 9.61    | 9.11    |

Note: The averaged running times of the standard methods were less than 0.01 s.

Figure 9. Original and denoised winning times of Boston Marathon Men’s Open Division between 1953–2016. The denoised versions were obtained by the proposed shrinkage rules with logistic and beta priors under exponential noise model and the cross-validation thresholding rule.

rejecting the null hypothesis under 5% of significance level. Thus, the exponential noise assumption for these dataset seems to be reasonable.

7. Final considerations

We proposed bayesian wavelet shrinkage rules to estimate wavelet coefficients under non-parametric regression models with exponential and lognormal additive noise. The adopted priors to the wavelet coefficients were mixtures of a point mass function at zero with logistic and beta distributions. Under the standard Gaussian noise assumption, the noise distribution is preserved on wavelet domain, i.e., the random noises after discrete wavelet transform application on original data remain iid with normal distribution and the estimation process is done coefficient by coefficient. Under positive noise in the original model, however, the preservation of the distribution does not necessarily occur. Moreover, random noises on wavelet domain are not necessarily positive and are correlated. In this sense, wavelet shrinkage under positive noise in the original model is performed
on the empirical coefficients vector and it requires the application of a robust adaptive MCMC algorithm to calculate posterior expectations from joint posterior distributions of the wavelet coefficients.

The proposed shrinkage rules outperformed the standard shrinkage and thresholding techniques in most of the scenarios for different underlying functions, sample sizes and signal-to-noise ratios of the simulation studies in terms of averaged mean squared error (AMSE) measure. Although the rules are more computationally expensive than the standard methods, their performances in simulation studies indicate them to be considered as wavelet shrinkage rules in contaminated data with positive noise in practice. Finally, the proposed rules provided suitable estimates in the real dataset application. Since the winning times measurements of Boston Marathon have positive noises due to possible delays,
their denoised measures must be less than the observed ones. It occurred under the proposed rules but not under standard methods, once they were built to deal with datasets under Gaussian noises.

The behaviour of the shrinkage rules for other positive support distributed noises and the impact of the wavelet basis choice to perform DWT are suggested as important questions to be studied in future works.

**Disclosure statement**

No potential conflict of interest was reported by the author(s).

**Funding**

This work was supported by a CAPES (Coordination of Superior Level Staff Improvement, Brazil) fellowship.

**ORCID**

Alex Rodrigo dos S. Sousa  [http://orcid.org/0000-0001-5887-3638](http://orcid.org/0000-0001-5887-3638)

**References**

[1] Vidakovic B. Statistical modeling by wavelets. New York: Wiley; 1999.
[2] Nason GP. Wavelet methods in statistics with R. New York (NY): Springer; 2008.
[3] Daubechies I. Ten lectures on wavelets. Philadelphia (GA): SIAM; 1992.
[4] Mallat SG. A wavelet tour of signal processing. San Diego (CA): Academic Press; 1998.
[5] Donoho DL. Nonlinear wavelet methods of recovery for signals, densities, and spectra from indirect and noisy data. Vol. 47, Proceedings of Symposia in Applied Mathematics. Providence (RI): American Mathematical Society; 1993.
[6] Donoho DL. Unconditional bases are optimal bases for data compression and statistical estimation. Appl Comput Harmonic Anal. 1993;1:100–115.
[7] Donoho DL. De-noising by soft-thresholding. IEEE Trans Inf Theory. 1995a;41:613–627.
[8] Donoho DL. Nonlinear solution of linear inverse problems by wavelet-vaguelette decomposition. App Comp Harm Anal. 1995b;2:101–26.
[9] Donoho DL, Johnstone IM. Ideal denoising in an orthonormal basis chosen from a library of bases. Compt Rend Acad Sci Paris A. 1994;319:1317–1322.
[10] Donoho DL, Johnstone IM. Ideal spatial adaptation by wavelet shrinkage. Biometrika. 1994;81:425–455.
[11] Donoho DL, Johnstone IM. Adapting to unknown smoothness via wavelet shrinkage. J Am Statist Assoc. 1995;90:1200–1224.
[12] Chipman H, Kolaczyk E, McCulloch R. Adaptive Bayesian wavelet shrinkage. J Am Statist Assoc. 1997;92:1413–1421.
[13] Vidakovic B, Ruggeri F. BAMS method: theory and simulations. Sankhya: Indian J Stat Ser B. 2001;63(00):234–249.
[14] Angelini C, Vidakovic B. Gama-minimax wavelet shrinkage: a robust incorporation of information about energy of a signal in denoising applications. Stat Sin. 2004;14:103–125.
[15] Reményi N, Vidakovic B. Wavelet shrinkage with double weibull prior. Commun Stat Simul Comput. 2015;44(1):88–104.
[16] Bhattacharya A, Pati D, Pillai N, et al. Dirichlet-Laplace priors for optimal shrinkage. J Am Stat Soc. 2015;110:1479–1490.
[17] Sousa ARS. Bayesian wavelet shrinkage with logistic prior. Commun Stat B Simul Comput. 2020;51(8):4700–4714.
[18] Sousa ARS, Garcia NL, Vidakovic B. Bayesian wavelet shrinkage with beta prior. Comput Stat. 2020;36:1341–1363.
[19] Jansen M. Noise reduction by wavelet thresholding. New York: Springer; 2001.
[20] Neumann MH, von Sachs R. Wavelet thresholding: beyond the gaussian I.I.D. situation. Springer: Wavelets and Statistics; 1995.
[21] Leporini D, Pesquet JC. Bayesian wavelet denoising: Besov priors and non-Gaussian noises. Signal Processing. 2001;81:55–67.
[22] Antoniadis A, Leporini D, Pesquet JC. Wavelet thresholding for some classes of non-Gaussian noise. Stat Neerl. 2002;56(4):434–453.
[23] Averkamp R, Houdré C. Wavelet thresholding for non necessarily gaussian noise: idealism. Ann Stat. 2003;31(1):110–151.
[24] Radnosrati K, Hendeby G, Gustafsson F. Exploring positive noise in estimation theory. IEEE Trans Signal Process. 2020;68:3590–3602.
[25] Vihola M. Robust adaptive Metropolis algorithm with coerced acceptance rate. J Stat Comput. 2012;22:997–1008.
[26] Scheidegger A. adaptMCMC: implementation of a generic adaptive Monte Carlo Markov Chain. R package version 1.4. 2021.
[27] Chaloner KM, Duncan GT. Assessment of a beta prior distribution: PM elicitation. J Royal Stat Soc Ser D. 1983;32:174–80.
[28] Duran BS, Booker JM. A Bayes sensitivity analysis when using the beta distribution as prior. IEEE Trans Reliab. 1988;37(2):239–247.
[29] Nason GP. Wavelet shrinkage using cross-validation. J R Stat Soc B. 1996;58:463–479.
[30] Abramovich F, Benjamini Y. Adaptive thresholding of wavelet coefficients. Comp Stat Data Anal. 1996;22:351–361.