



Multichannel deconvolution with long range dependence: Upper bounds on the L^p -risk ($1 \leq p < \infty$)



Rafal Kulik^a, Theofanis Sapatinas^b, Justin Rory Wishart^{c,*}

^a Department of Mathematics and Statistics, University of Ottawa, 585 King Edward Avenue, Ottawa ON K1N 6N5, Canada

^b Department of Mathematics and Statistics, University of Cyprus, P.O. Box 20537, CY 1678 Nicosia, Cyprus

^c Department of Mathematics and Statistics, University of New South Wales, Sydney, NSW 2052, Australia

ARTICLE INFO

Article history:

Received 6 August 2013

Received in revised form 15 April 2014

Accepted 27 April 2014

Available online 30 April 2014

Communicated by Guy Nason

MSC:

62G08

62G05

62G20

Keywords:

Besov spaces

Brownian motion

Deconvolution

Fourier analysis

Fractional Brownian motion

Meyer wavelets

Multichannel deconvolution

Thresholding

Wavelet analysis

ABSTRACT

We consider multichannel deconvolution in a periodic setting with long-memory errors under three different scenarios for the convolution operators, i.e., super-smooth, regular-smooth and box-car convolutions. We investigate global performances of linear and hard-thresholded non-linear wavelet estimators for functions over a wide range of Besov spaces and for a variety of loss functions defining the risk. In particular, we obtain upper bounds on convergence rates using the L^p -risk ($1 \leq p < \infty$). Contrary to the case where the errors follow independent Brownian motions, it is demonstrated that multichannel deconvolution with errors that follow independent fractional Brownian motions with different Hurst parameters results in a much more involved situation. An extensive finite-sample numerical study is performed to supplement the theoretical findings.

© 2014 Elsevier Inc. All rights reserved.

1. Introduction

We study multichannel deconvolution with errors following independent fractional Brownian motions (fBms). More specifically, consider the problem of recovering $f(\cdot) \in L^2(T)$, $T = [0, 1]$, on the basis of observing the following noisy convolutions, with known blurring functions $g_\ell(\cdot)$,

* Corresponding author.

E-mail addresses: rkulik@uottawa.ca (R. Kulik), fanis@ucy.ac.cy (T. Sapatinas), j.wishart@unsw.edu.au (J.R. Wishart).

$$dY_\ell(t) = K_\ell f(t)dt + \frac{\sigma_\ell}{n^{\alpha_\ell/2}} dB_{H_\ell}(t), \quad t \in T, \ell = 1, 2, \dots, M, \quad (1)$$

where σ_ℓ are known positive constants and the convolution operators K_ℓ are defined as

$$K_\ell f(t) := f * g_\ell(t) = \int_T g_\ell(t-x)f(x)dx, \quad t \in T, \ell = 1, 2, \dots, M. \quad (2)$$

Here, $B_{H_\ell}(\cdot)$ are independent standard fBms with *Hurst* parameters $H_\ell = 1 - \alpha_\ell/2 \in [1/2, 1)$, $\ell = 1, 2, \dots, M$; that is, for each $\ell = 1, 2, \dots, M$; $B_{H_\ell}(\cdot)$ is a Gaussian process with zero mean and covariance function

$$\mathbb{E}(B_{H_\ell}(s)B_{H_\ell}(t)) = \frac{1}{2}(|s|^{2H_\ell} + |t|^{2H_\ell} - |t-s|^{2H_\ell}), \quad s, t \in T, \ell = 1, 2, \dots, M.$$

The case where $M = 1$ corresponds to the fractional Gaussian noise model that can also be viewed as an approximation to the nonparametric regression model with long-range dependence (LRD) (cf. [1,2]). On the other hand, the case $H_\ell = 1/2$, $\ell = 1, \dots, M$; becomes the *multichannel* deconvolution with independent standard Brownian motion errors. This model has received attention in studies by [3–5] and [6].

We consider the following scenarios for the convolution operators K_ℓ , $\ell = 1, 2, \dots, M$; given by (2) in the Fourier domain where $\tilde{f}(m) := \int_{\mathbb{R}} e^{-2\pi imx} f(x) dx$.

1. *Smooth* convolutions such that, in the Fourier domain,

$$|\widetilde{K_\ell f}(m)| \asymp |m|^{-\nu_\ell} \exp\{-\theta_\ell |m|^{\beta_\ell}\} |\tilde{f}(m)|, \quad (3)$$

where $m \in \mathbb{R}$, $\ell = 1, 2, \dots, M$; $\beta_\ell > 0$ and $\theta_\ell \geq 0$. In particular, $\nu_\ell \in \mathbb{R}$ if $\theta_\ell > 0$ and $\nu_\ell > 0$ if $\theta_\ell = 0$. The key parameter is θ_ℓ , controlling the severity of the decay. The so-called super-smooth deconvolution or exponential decay occurs when $\theta_\ell > 0$ and the regular-smooth or polynomial case occurs when $\theta_\ell = 0$. In the regular-smooth case, each $\nu_\ell > 0$ corresponds to the so-called *degree of ill-posedness* (DIP) index with $\nu_\ell = 0$ representing the *direct* (or *well-posed*) case.

2. *Box-car* convolutions such that, in the Fourier domain,

$$|\widetilde{K_\ell f}(m)| = \frac{\sin(\pi mc_\ell)}{\pi mc_\ell} |\tilde{f}(m)|, \quad m \in \mathbb{R}, \ell = 1, 2, \dots, M; \quad (4)$$

where $c_\ell > 0$ for each $\ell = 1, 2, \dots, M$.

Deconvolution is a common problem in many areas of signal and image processing which include, for instance, light detection and ranging (LIDAR) remote sensing and reconstruction of blurred images. LIDAR is a laser device which emits pulses, reflections of which are gathered by a telescope aligned with the laser. The return signal is used to determine the distance and the position of the reflecting material. However, if the system response function of the LIDAR is longer than the time resolution interval, then the measured LIDAR signal is blurred and the effective accuracy of the LIDAR decreases. This loss of precision can be corrected by deconvolution. In practice, measured LIDAR signals are corrupted by additional noise which renders direct deconvolution impossible. Moreover, if $M \geq 2$ (finite) LIDAR devices are used to recover a signal, then we talk about a *multichannel* deconvolution problem. The case where $M \geq 2$ in (1)–(2) and $H_\ell = 1/2$, $\ell = 1, \dots, M$; i.e., the problem of considering systems of convolution equations with independent errors, was first considered by [7] in order to evade the ill-posedness of the standard deconvolution model.

In the standard Brownian motion error case, a statistical use of the above idea was investigated by [8,3] who proposed adaptive wavelet thresholding estimators. In particular, if K_ℓ are regular-smooth convolutions,

they showed that an *adaptive* wavelet thresholding estimator based on the output from the M channels “picks” the convergence rate according to “the best” operator K_ℓ , i.e., the one with the smallest ν_ℓ , $\ell = 1, 2, \dots, M$. Consequently, adding more channels does not improve the convergence rate of the suggested estimator. On the other hand, if K_ℓ , $\ell = 1, 2, \dots, M$; are box-car convolutions, they showed that adding new channels improves the convergence rate. To be more specific, [3] showed, in particular, that the true signal $f(\cdot)$ can be recovered with accuracy (within a logarithmic factor),

$$n^{-2s/(2s+2\nu+1)} \quad \text{and} \quad n^{-2s/(2s+(2M+1)/M+1)},$$

in the regular-smooth and box-car convolutions, respectively. Here, $s > 0$ is the smoothness of the underlying signal, $\nu = \min\{\nu_1, \dots, \nu_M\}$ and the *accuracy of estimation* is measured with respect to an upper bound on the L_2 -risk. In [3] the authors did not consider the super-smooth convolutions.

However, real data do not always meet the independence assumption and scientist in diverse fields have observed empirically that correlations between observations that are far apart decay to zero at a slower rate than one would expect from independent data (or, in more general situation, where one deals with short-range dependent data). These fields include astronomy, agronomy, economics chemistry, etc. (see, e.g., [9]).

Therefore, our aim is to study the multichannel deconvolution with errors following fBMs. In fact, we show that the situation in this case is much more involved than in the case where the errors follow standard Brownian motions. In particular, we show that in multichannel deconvolution with errors following fBms, the true signal $f(\cdot)$ can be recovered with respect to an upper bound on the L^p -risk ($1 \leq p < \infty$) with accuracy,

$$n^{-s\alpha_{\ell_*}p/(2s+2\nu_*+1)}, \quad (\log n)^{-ps^*/\beta_{\ell_*}} \quad \text{and} \quad n^{-s\alpha_*p/(2s+2\tilde{\nu}_*+1)}$$

for regular-smooth, super-smooth and box-car deconvolutions respectively (the regular smooth and box-car scenarios are within a logarithmic factor). The parameters in the case of smooth (both regular-smooth and super-smooth) convolutions are defined with

$$\ell_* := \arg \min_{1 \leq \ell \leq M} n^{-\alpha_\ell} 2^{(\alpha_\ell+2\nu_\ell)} e^{2\theta_\ell 2^{\beta_\ell}} \tag{5}$$

for the optimal channel and ν_* is defined for the regular-smooth case as

$$\nu_* := \nu_{\ell_*} + \frac{\alpha_{\ell_*}}{2} - \frac{1}{2}. \tag{6}$$

For the case of box-car convolutions the parameters are defined with

$$\alpha_* := \min\{\alpha_1, \dots, \alpha_M\} \quad \text{and} \quad \alpha^* := \max\{\alpha_1, \dots, \alpha_M\}, \tag{7}$$

$$\tilde{\nu}_* := \frac{2M+1}{2M} + \frac{\alpha^*}{2} - \frac{1}{2}. \tag{8}$$

Consequently, the conclusions of [3] are no longer valid here. Even in case of $M = 2$, there are different possibilities for the *best scenario*, depending on a complicated relationship between s , M , ν_ℓ , α_ℓ , θ_ℓ and β_ℓ , as we illustrate in Section 4.

1.1. Modification of the WaveD method

Along with theoretical results, a comparison with the existing WaveD method is presented to examine the effect of LRD and multiple channels. Let us compare our modification of the WaveD to the standard

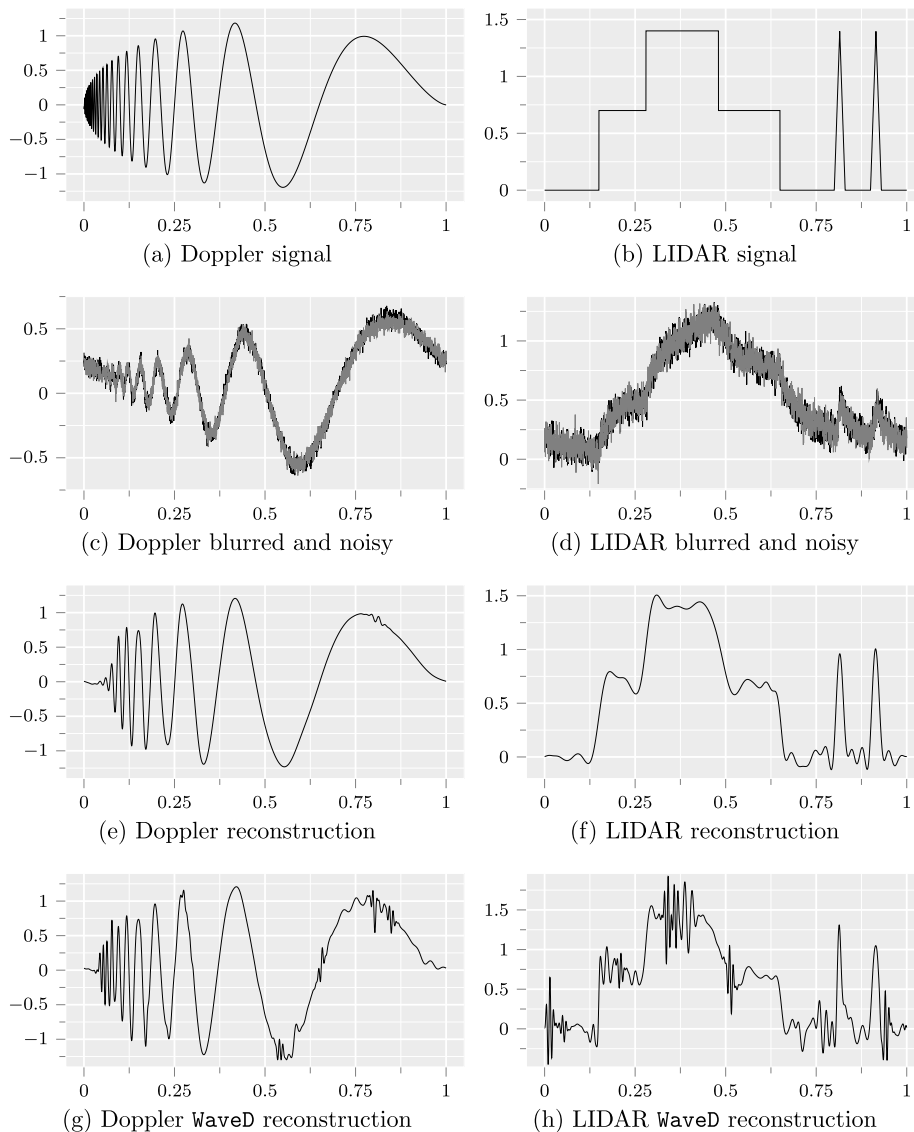


Fig. 1. Top row: original Doppler and LIDAR signal; 2nd row: corresponding blurred and noisy signals, $\nu = 0.5$, $\alpha = 0.5$ (black line: first channel; grey line: second channel); 3rd row: reconstructed signal using the proposed method with $M = 2$ channels; 4th row: reconstructed signal using the standard R-package WaveD using the best channel (see (31), for the notion of ‘best channel’).

R-package WaveD of [10]. In particular, the four signals, LIDAR, Doppler, Bumps and Blocks are used as candidate signals in estimation.

For mild levels of LRD ($1/2 < \alpha < 1$) there is not too much difference between the both approaches. However, an improvement is visible for a stronger dependence ($0 < \alpha < 1/2$), as illustrated on Fig. 1 and Fig. 2. For the parameters $\alpha = 0.5$, $\nu = 0.5$ and $M = 2$, in the third row, a signal is reconstructed using the proposed multichannel method while the fourth row shows the standard WaveD approach using the best channel.

Clearly, the standard WaveD approach does not remove artificial noise, which is due to LRD (cf. Fig. 1). We modify the WaveD approach and achieve more reliable estimation by appropriately modified tuning parameters and also truncating the wavelet expansion at an appropriate lower scale level. This truncation is particularly important when there is severe LRD but does not universally yield better estimates (cf. Fig. 2) and is discussed in more depth in the numerical section later.

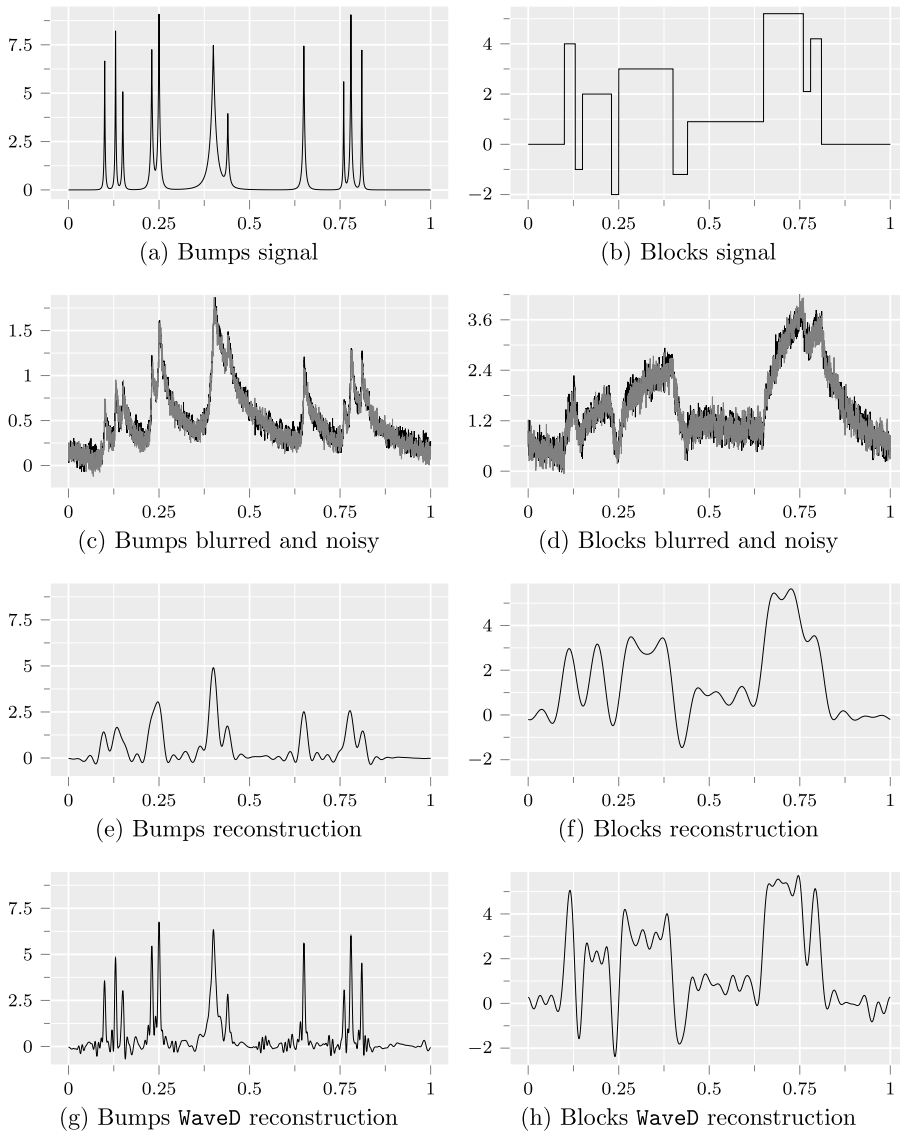


Fig. 2. Top row: original Bumps and Blocks signal; 2nd row: corresponding blurred and noisy signals, $\nu = 0.5$, $\alpha = 0.5$ (black line: first channel; grey line: second channel); 3rd row: reconstructed signal using the proposed method with $M = 2$ channels; 4th row: reconstructed signal using the standard R-package WaveD using the best channel (see (31), for the notion of ‘best channel’).

1.2. Related works

The case where $M = 1$ and $H_1 = 1/2$ in (1)–(2) refers to the so-called *standard deconvolution model* which attracted attention of a number of researchers. (Note that the standard deconvolution model is typically *ill-posed* in the sense of Hadamard: the inversion does not depend continuously on the observed data, i.e., small noise in the convolved signal leads to a significant error in the estimation procedure.) After a rather rapid progress in this problem in late eighties–early nineties, authors turned to adaptive wavelet solutions of the problem that are optimal (in the *minimax* or the *maxiset* sense), or near-optimal within a logarithmic factor, in a wide range of Besov balls and for a variety of loss functions defining the risk, and under mild conditions on the blurring function (see, e.g., [11–18]).

The case $M = 1$ and $H_\ell > 1/2$ (i.e., standard deconvolution with LRD errors) has been investigated in [1,2,19] and [20].

The case where $\alpha_\ell = 1$ for each $\ell = 1, 2, \dots, M$; (i.e., the case where in the multichannel deconvolution model (1) the errors follow independent standard Brownian motions) was first considered in [3] (extending the results obtained in [14] for the case $M = 1$).

The case of the multichannel deconvolution with errors following LRD sequences was investigated in [21] using the minimax approach, extending results obtained in [4,5] and [6].

The case of nonparametric density estimation for the errors-in-variables problem with LRD has been studied by [22]. In particular, it was shown that LRD has no impact on the optimal convergence properties in the super-smooth scenario. We show similar results for the multichannel deconvolution model presented here.

Finally, for more information regarding the LIDAR device, the reader is referred to, e.g., [23] and [24].

1.3. Structure of the paper

The paper is organised as follows. Section 2 contains some preliminaries on the periodised Meyer wavelets and Besov spaces on the unit interval T . Section 3 provides the construction of the proposed adaptive wavelet thresholding estimators while Section 4 contains the corresponding upper bound results over a wide range of Besov spaces and for a variety of loss functions defining the risk, for regular-smooth, super-smooth and box-car convolutions. An extensive simulation study to supplement the theoretical findings of Section 4 is performed in Section 5. Conclusions and discussion are given in Section 6 and the proofs of the theoretical results and auxiliary results given in Section 7 and Appendix A.

2. Preliminaries

2.1. Periodised Meyer wavelets and Besov spaces on the unit interval

To avoid edge problems and unnecessary technicalities arising in defining wavelet basis on the unit interval T , we will assume that $f(\cdot)$ and $g_\ell(\cdot)$, $\ell = 1, 2, \dots, M$; are periodic on T . Moreover, not only for theoretical reasons but also for practical convenience (see, e.g., [14], Sections 2.3, 3.1–3.2), we use band-limited wavelet basis, and in particular the periodised Meyer wavelet basis for which fast algorithms exist (see, e.g., [25] and [15]). Specifically, let $\phi(\cdot)$ and $\psi(\cdot)$ be the Meyer scaling and mother wavelet functions, respectively, on the real line $\mathbb{R} = (-\infty, \infty)$ (see, e.g., [26] or [27]). As usual,

$$\phi_{j,k}(t) = 2^{j/2}\phi(2^j t - k), \quad \psi_{j,k}(t) = 2^{j/2}\psi(2^j t - k), \quad j \geq 0, k \in \mathbb{Z}, t \in \mathbb{R},$$

are, respectively, the dilated and translated Meyer scaling and wavelet functions at resolution level j and scale position $k/2^j$. Similarly to Section 2.3 in [14], we obtain a periodised version of Meyer wavelet basis by periodising the basis functions $\{\phi(\cdot), \psi(\cdot)\}$ on \mathbb{R} , i.e., for $j \geq 0$ and $k = 0, 1, \dots, 2^j - 1$,

$$\Phi_{j,k}(t) = \sum_{i \in \mathbb{Z}} 2^{j/2}\phi(2^j(t+i) - k), \quad \Psi_{j,k}(t) = \sum_{i \in \mathbb{Z}} 2^{j/2}\psi(2^j(t+i) - k), \quad t \in T.$$

In the periodic setting, we recall that Besov spaces are characterised by the behaviour of the wavelet coefficients (see, e.g., [14], Section 2.4), i.e.,

Definition 1. For $f(\cdot) \in L^{\pi_0}(T)$, $1 \leq \pi_0 < \infty$,

$$f(\cdot) \in \mathcal{B}_{\pi_0, r}^s(T) \iff \sum_{j=0}^{\infty} 2^{j(s+1/2-1/\pi_0)r} \left[\sum_{k=0}^{2^j-1} |b_{j,k}|^{\pi_0} \right]^{r/\pi_0} < \infty, \tag{9}$$

with the usual modification if $\pi_0 = \infty$ and/or $r = \infty$.

As usual, the wavelet coefficients $b_{j,k}$ are obtained by $b_{j,k} = \int_T f(t)\psi_{j,k}(t)dt$. The parameter $s > 0$ can be thought of as related to the number of derivatives of $f(\cdot)$. With different values of π_0 ($1 \leq \pi_0 \leq \infty$) and r ($1 \leq r \leq \infty$), the Besov spaces $\mathcal{B}_{\pi_0,r}^s(T)$ capture a variety of smoothness features in a function including spatially inhomogeneous behaviour.

In the sequel, κ will denote the multiple index (j, k) and, adopting standard convention, $\Phi(\cdot) = \Psi_{-1}(\cdot)$, where $\Phi(\cdot)$ corresponds to the periodised scaling function associated with the Meyer wavelet basis mentioned above.

3. Construction of the adaptive wavelet thresholding and linear estimators

The estimation of f is approached differently for the different deconvolution types. Namely, for regular-smooth and box-car convolutions a wavelet non-linear (hard thresholding) estimator is used while for the super-smooth convolutions a wavelet linear (projection) estimator is used.

To simplify the overall problem, the estimation procedure is considered in the Fourier domain to reduce the convolution operator to a product of Fourier coefficients. Denote the Fourier basis functions, $e_m(t) := e^{2\pi imt}$, $m \in \mathbb{Z}$, with the corresponding inner product operator, $\langle f_1, f_2 \rangle = \int f_1(x)\overline{f_2(x)} dx$ where \overline{f} denotes the complex conjugate of f . Let $h = f * g_\ell$. Denote the relevant Fourier coefficients,

$$\begin{aligned} \Phi_{mj_0k} &= \langle \Phi_{j_0,k}, e_m \rangle, & \Psi_m^\kappa &= \Psi_{mj_k} = \langle \Psi_{j,k}, e_m \rangle, \\ h_{m,\ell} &= \langle h_\ell, e_m \rangle, & y_{m,\ell} &= \int_{\mathbb{R}} \overline{e_m}(t) dY_\ell(t), & z_{m,\ell} &= \int_{\mathbb{R}} \overline{e_m}(t) dB_{H_\ell}(t), \\ f_m &= \langle f, e_m \rangle, & g_{m,\ell} &= \langle g_\ell, e_m \rangle, & \ell &= 1, 2, \dots, M. \end{aligned} \tag{10}$$

Applying the Fourier transform to (1), we get the following sequence space model

$$y_{m,\ell} = h_{m,\ell} + \frac{\sigma_\ell}{n^{\alpha_\ell/2}} z_{m,\ell}, \quad m \in \mathbb{Z}, \ell = 1, 2, \dots, M; \tag{11}$$

$$h_{m,\ell} = g_{m,\ell} f_m, \quad m \in \mathbb{Z}, \ell = 1, 2, \dots, M; \tag{12}$$

where, for each ℓ , σ_ℓ are known positive constants and the structure of the Fourier coefficients, $g_{m,\ell} f_m = \widetilde{K_\ell} f(m)$, is given by (3) and (4) for the smooth-type and box-car convolutions respectively. Following a similar procedure to [3], weights $\gamma_{m,\ell} \overline{g_{m,\ell}}$ are multiplied to the $h_{m,\ell}$ coefficients and added together (where $\gamma_{m,\ell}$ are weights to be specified later). Thus (12) leads to the following expression for the target function coefficients,

$$f_m = \frac{\sum_{\ell=1}^M \gamma_{m,\ell} \overline{g_{m,\ell}} h_{m,\ell}}{\sum_{\ell=1}^M \gamma_{m,\ell} |g_{m,\ell}|^2}, \quad m \in \mathbb{Z}.$$

Furthermore using the Parseval identity one can obtain the wavelet coefficients,

$$b_\kappa = \int_T f(t)\Psi_\kappa(t) dt = \sum_{m \in \mathbb{Z}} f_m \overline{\Psi_m^\kappa} = \sum_{m \in \mathbb{Z}} \frac{\sum_{\ell=1}^M \gamma_{m,\ell} \overline{g_{m,\ell}} h_{m,\ell}}{\sum_{\ell=1}^M \gamma_{m,\ell} |g_{m,\ell}|^2} \overline{\Psi_m^\kappa}$$

which can be estimated using (11) with

$$\widehat{b}_\kappa = \sum_{m \in C_j} \frac{\sum_{\ell=1}^M \gamma_{m,\ell} \overline{g_{m,\ell}} y_{m,\ell}}{\sum_{\ell=1}^M \gamma_{m,\ell} |g_{m,\ell}|^2} \overline{\Psi_m^\kappa}, \tag{13}$$

where C_j denotes the domain of the Meyer wavelet in the Fourier domain,

$$C_j = \left\{ a \in \mathbb{Z} : \pm a \in \left\{ \left\lceil \frac{2^j}{3} \right\rceil, \left\lceil \frac{2^j}{3} \right\rceil + 1, \dots, \left\lfloor \frac{2^{j+2}}{3} \right\rfloor \right\} \right\}, \tag{14}$$

where $j \geq 0$. The scaling coefficients $a_\kappa = \int_T f(t)\Phi_\kappa(t) dt$ and their estimates \hat{a}_κ are defined in a similar manner.

Estimators: A non-linear estimator $\hat{f}_n(\cdot)$ of $f(\cdot)$ based on hard thresholding of a wavelet expansion is as follows:

$$\hat{f}_n(t) = \sum_{k=0}^{2^{j_0}-1} \hat{a}_{j_0,k} \Phi_{j_0,k}(t) + \sum_{\kappa \in \Lambda} \hat{b}_\kappa \mathbb{1}_{\{|\hat{b}_\kappa| \geq \lambda\}} \Psi_\kappa(t), \quad t \in T, \tag{15}$$

where $\mathbb{1}_A$ denotes the indicator function of the set A , the index range, $\Lambda = \Lambda_n$, the coarse scale level j_0 and the threshold parameter $\lambda = \lambda_j$ are forthcoming.

A linear (projection) wavelet estimator $\hat{f}_n(\cdot)$ of $f(\cdot)$ with coarse scale level j_0 is

$$\hat{f}_n(t) = \sum_{k=0}^{2^{j_0}-1} \hat{a}_{j_0,k} \Phi_{j_0,k}(t). \tag{16}$$

Resolution levels: The range of resolution levels (frequencies) is given by

$$\Lambda_n = \{(j, k), j_0 \leq j \leq j_1, 0 \leq k \leq 2^j - 1\}.$$

The coarse scale j_0 is defined in the super-smooth case as,

$$2^{j_0} \asymp \left(\frac{(\alpha_{\ell_*} - \epsilon) \log n}{2\theta_{\ell_*}} \right)^{1/\beta_{\ell_*}} \tag{17}$$

where $\epsilon > 0$ is small, θ_ℓ is the super-smooth parameter defined in (3) and ℓ_* is given by (5). For the regular-smooth and box-car case the parameter j_0 is not important for the asymptotic convergence of the estimator and we set $j_0 = -1$. The fine scale level j_1 is important for the asymptotic convergence results in these cases and is set to be,

$$2^{j_1} \asymp \left(\frac{n^{\alpha_{\ell_*}}}{\log n} \right)^{1/(2\nu_*+1)} \tag{18}$$

for regular-smooth convolutions and

$$2^{j_1} \asymp \left(\frac{n^{\alpha_*}}{\log n} \right)^{1/(2\tilde{\nu}_*+1)} \tag{19}$$

for box-car convolutions, where α_* , ν_* , $\tilde{\nu}_*$ and ℓ_* are defined in (7), (6), (8) and (5) respectively. The fine resolution level j_1 in (18) coincides with the level given by [20] for the case when $M = 1$, $\nu_1 = \nu$ and $\alpha_1 = \alpha$.

Thresholds: To ease the presentation and include both the regular-smooth and box-car cases, define

$$\xi = \begin{cases} \alpha_{\ell_*}, & \text{in the case of regular-smooth deconvolutions;} \\ \alpha_*, & \text{in the case of box-car deconvolutions.} \end{cases}$$

Then the scale level threshold values $\lambda = \lambda_j$ are given by

$$\lambda_j = \zeta \tau_j c_n, \tag{20}$$

where the three input parameters are specified as:

- ζ : a smoothing parameter, $\zeta > 2\sqrt{(p \vee 2)2\xi}$.
- c_n : a sample size-dependent scaling factor,

$$c_n = \sqrt{\frac{\log n}{n^\xi}}. \tag{21}$$

- τ_j : a level-dependent scaling factor,

$$\tau_j^2 = n^\xi \sum_{m \in C_j} |\Psi_m^\kappa|^2 \left(\sum_{\ell=1}^M \sigma_\ell^{-2} n^{\alpha_\ell} |m|^{2H_\ell-1} |g_{m,\ell}|^2 \right)^{-1}. \tag{22}$$

In practical applications, the noise levels σ_ℓ ; $\ell = 1, 2, \dots, M$; are usually unknown. In this case, estimate each σ_ℓ by $\widehat{\sigma}_\ell$ and define

$$\widehat{\tau}_j^2 = n^\xi \sum_{m \in C_j} |\Psi_m^\kappa|^2 \left(\sum_{\ell=1}^M \widehat{\sigma}_\ell^{-2} n^{\alpha_\ell} |m|^{2H_\ell-1} |g_{m,\ell}|^2 \right)^{-1}. \tag{23}$$

This expression is used in the simulation study conducted in Section 4. Note that the above thresholds λ_j defined in (20) coincide with the ones defined in [3] ($M \geq 2, \alpha_* = 1$).

4. Upper bound results of the adaptive wavelet thresholding and linear estimators

Consider first the smooth convolutions scenario. In this case, the regular-smooth and super-smooth cases are handled when $\theta_\ell = 0$ or $\theta_\ell > 0$ respectively. The super-smooth case is similar to estimating analytic functions with a slow convergence rate. In this scenario linear estimators obtain the optimal (in the minimax sense) convergence rates and hence a linear (projection) wavelet estimator with an appropriate primary resolution level j_0 suffices.

Theorem 1. Consider the model described by (1) with $f \in \mathcal{B}_{\pi_0, r}^s(T)$ with $\pi_0 \geq 1, s \geq \frac{1}{\pi_0}$. If $\theta_\ell = 0$ for each $\ell = 1, 2, \dots, M$; (regular-smooth case) then consider

$$0 < r \leq r_0 = \min \left\{ \frac{p(2\nu_* + 1)}{2(\nu_* + s) + 1}, \frac{(2\nu_* + 1)p - 2}{2(\nu_* + s) - 2/\pi_0 + 1} \right\} \tag{24}$$

and the adaptive wavelet estimator \widehat{f}_n defined in (15) with the index range $\Lambda = \Lambda_n$ defined by (18) and threshold value $\lambda = \lambda_j$ defined by (20) for some $\zeta > 2\sqrt{(p \vee 2)2\alpha_{\ell_*}}$ with τ_j and c_n given, respectively, by (22) and (21). If $\theta_\ell > 0$ for each $\ell = 1, 2, \dots, M$; (super-smooth case) then consider $r > 0$ and the linear projection wavelet estimator defined in (16) with coarse scale level, j_0 , given by (17). Let $p > 1$ be an arbitrary finite real number. Then, there exists a constant $C > 0$ such that for all $n \geq 1$,

$$\mathbb{E} \|\widehat{f}_n - f\|_p^p \leq C \left(\frac{\log n}{n^\delta} \right)^q,$$

where in the regular-smooth case $\theta_{\ell_*} = 0$ and $\delta = 1$ with

$$\varrho = \frac{\alpha_{\ell_*} sp}{2(s + (2\nu_* + 1)/2)}, \quad \text{if } s \geq \frac{(2\nu_* + 1)}{2} \left(\frac{p}{\pi_0} - 1 \right); \tag{25}$$

$$\varrho = \frac{\alpha_{\ell_*} p(s - 1/\pi_0 + 1/p)}{2(s - 1/\pi_0 + (2\nu_* + 1)/2)}, \quad \text{if } \frac{1}{\pi_0} - \nu_* - \frac{1}{2} \leq s < \frac{(2\nu_* + 1)}{2} \left(\frac{p}{\pi_0} - 1 \right); \tag{26}$$

while in the super-smooth case, $\theta_{\ell_*} > 0$ and $\delta = 0$ with,

$$\varrho = -ps^*/\beta_{\ell_*} \quad \text{where } \beta_{\ell_*} > 0 \tag{27}$$

and $s^* = s + 1/p - 1/\min(p, \pi_0)$ and ℓ_* is defined with (5).

Now, consider box-car convolutions scenario. Recall α_* defined by (7) and ν_* is now replaced with $\tilde{\nu}_*$ defined by (8). For the definitions of the ‘Badly Approximable’ (BA) irrational number and the BA irrational tuple that we used in the following statement, see, e.g., p. 22 and p. 42 of [28].

Theorem 2. Consider the model described by (1) and the wavelet estimator \hat{f}_n defined in (15) with the index range $\Lambda = \Lambda_n$ defined by (19) and threshold value $\lambda = \lambda_j$ defined by (20) for some $\zeta > 2\sqrt{(p \vee 2)2\alpha_*}$ with τ_j and c_n given, respectively, by (22) and (21). Let $p > 1$ be an arbitrary finite real number and assume that one of the c_1, c_2, \dots, c_M is a BA irrational number and that c_1, c_2, \dots, c_M ($M \geq 2$) is a BA irrational tuple. If $f \in \mathcal{B}_{\pi_0, r}^s(T)$ with $\pi_0 \geq 1$, $s \geq \frac{1}{\pi_0} - \tilde{\nu}_* - 1/2$ and r satisfying (24) with ν_* replaced with $\tilde{\nu}_*$, then, in this case, the result of Theorem 1 still holds with $\delta = 1$ and

$$\varrho = \frac{\alpha_* sp}{2(s + (2\tilde{\nu}_* + 1)/2)}, \quad \text{if } s \geq \frac{(2\tilde{\nu}_* + 1)}{2} \left(\frac{p}{\pi_0} - 1 \right); \tag{28}$$

$$\varrho = \frac{\alpha_* p(s - 1/\pi_0 + 1/p)}{2(s - 1/\pi_0 + (2\tilde{\nu}_* + 1)/2)}, \quad \text{if } \frac{1}{\pi_0} - \tilde{\nu}_* - \frac{1}{2} \leq s < \frac{(2\tilde{\nu}_* + 1)}{2} \left(\frac{p}{\pi_0} - 1 \right); \tag{29}$$

where $\tilde{\nu}_*$ is defined by (8) and α_* is defined by (7).

Remark 1. There is an elbow effect or phase transition in the upper bound on the L^p -risk ($1 \leq p < \infty$) in both the regular-smooth and box-car convolutions. Namely, in the regular-smooth case switching from (25) to (26) as the assumed smoothness decreases; and similarly switching from (28) to (29) in the box-car case. The two regimes are usually referred to as the dense and sparse regions respectively (see [14] and [3] for the case with independent Brownian motion errors). The upper bound results obtained in Theorem 1 for the regular-smooth case and in Theorem 2 for the box-car case show that the boundary region of s depends on the LRD indices α_ℓ , $\ell = 1, 2, \dots, M$; and the sparse region is smaller in the case where the errors follow independent fBms.

Remark 2. Single Channel, $M = 1$: For $\nu_* = \nu_1 = 0$, the upper bounds obtained on the L^p -risk ($1 \leq p < \infty$) in Theorem 1 agree with existing optimal rate results (up to a logarithmic factor) for wavelet regression with long-memory errors obtained by [1], (minimax L^2 -risk) and [19] (upper bounds on the L^p -risk, $1 \leq p < \infty$). Similarly, when $\nu_* = \nu_1 > 0$ the results also agree with [2] (minimax L^2 -risk, $p = 2$) and [20] (upper bounds on the L^p -risk, $1 \leq p < \infty$). Multichannel, $M > 1$: Our results generalise the results in [3] and include the results of their case when $\alpha_* = 1$ (upper bounds on the L^p -risk, $1 \leq p < \infty$).

Remark 3. As expected, the upper bounds deteriorate in the regular-smooth and box-car cases when ν_{ℓ_*} increases (larger DIP) or when α_{ℓ_*} decreases (stronger LRD). The combined effect of ν_{ℓ_*} and α_{ℓ_*} on the location of the elbow is reverse as the sparse region increases with both ν_{ℓ_*} and α_{ℓ_*} . Consistent with the literature, the super-smooth case has a logarithmic convergence rate with indices that depend on the

underlying smoothness in s^* and the severity of the super-smooth decay in β_{ℓ^*} . The upper bounds on the L^p -risk ($1 \leq p < \infty$) in the super-smooth case do not depend on ν_{ℓ^*} or α_{ℓ^*} .

Remark 4. Our upper bounds on the L^p -risk (for $p = 2$) are not directly comparable to the maximal upper bounds obtained in [21]. In that paper the framework is different whereby the number of channels M depends on the number of total observations, n , in each channel (i.e. $M = M_n$). However, in our case the number of channels is fixed and not dependent on n . Our results are comparable to the works of [3,20] demonstrating both the effects of the number of channels and the LRD on the upper bounds on the L^p -risk ($1 \leq p < \infty$).

5. Simulation study

A simulation study for $p = 2$ is conducted for the regular-smooth scenario and is heavily based on the algorithm in the WaveD R-package of [10]. In the regular smooth scenario, it is crucial to know ℓ_* , the ‘best channel’, since it appears in both the smoothing parameter ζ and in the fine scale level j_1 . The fine scale parameter is particularly important since it truncates the wavelet expansion early enough to ensure an accurate yet reliable algorithm. Methods have been established for choosing j_1 in practice for the single channel regular Brownian motion case by [29] and expanded to the single channel LRD case by [20]. The method is sketched below and the interested reader is referred to those papers for a more in-depth treatment.

The method assumes the practitioner can pass the Fourier basis, $f = \{e_u\}_{u \in \mathbb{Z}}$, into (1) and denote this new information with,

$$d\check{Y}_\ell(x) = g_\ell * e_u(x) + \sigma_\ell n^{-\alpha_\ell/2} dB_{H_\ell}(x).$$

Due to the orthogonality of the Fourier basis, the Fourier domain representation of \check{Y}_ℓ is

$$\check{y}_{m,\ell} = \int_{\mathbb{R}} e_m(x) d\check{Y}_\ell(x) = g_{m,\ell} + \sigma_\ell n^{-\alpha_\ell/2} w_{m,\ell}$$

where $w_{m,\ell}$ is identically distributed to but independent of $z_{m,\ell}$ (recall (10) for the definition of $z_{m,\ell}$). Then an estimate of $j_{\ell,1} \asymp (n^{\alpha_\ell} / \log n)^{1/(\alpha_\ell + 2\nu_\ell)}$ is constructed with,

$$\hat{j}_{\ell,1} = \lfloor \log_2 F_\ell \rfloor - 1$$

where the stopping time F_ℓ is determined in the Fourier domain with $F_\ell = \min\{\omega, \omega > 0 : |\check{y}_{\omega,\ell}| \leq \omega^{\alpha_\ell/2} \varepsilon_\ell \log \varepsilon_\ell^{-2}\}$ and $\varepsilon_\ell := \sigma_\ell n^{-\alpha_\ell/2}$. The estimate $\hat{j}_{\ell,1}$ is close to $j_{\ell,1}$ with high probability due to Lemma 1 in [20].

Then define the overall fine scale estimator with,

$$\hat{j}_1 = \max_{1 \leq \ell \leq M} \hat{j}_{\ell,1}, \tag{30}$$

since the optimal channel defined with

$$\ell^* := \arg \max_{1 \leq \ell \leq M} \left\{ \left(\frac{n^{\alpha_\ell}}{\log n} \right)^{1/(\alpha_\ell + 2\nu_\ell)} \right\},$$

is equivalent to the optimal channel ℓ_* defined in (5). For the same reason, the best channel is estimated as the one with the largest stopping time,

$$\hat{\ell}_* = \arg \max_{1 \leq \ell \leq M} F_\ell. \tag{31}$$

The theory suggests that the smoothing parameter should satisfy the bound, $\zeta > 4\sqrt{\alpha_{\ell^*}}$ when $p = 2$. However, as will become evident in the simulations a smaller choice for ζ results in improved numerical performance. This smaller choice of smoothing parameter compared to the theory is consistent with other numerical results of [14] and [20]. The signals used in the simulations are the standard **LIDAR**, **Doppler**, **Bumps** and **Blocks** functions that have been used consistently throughout the literature (cf. [30,29]).

The steps for a simulation study are then as follows:

1. We choose $f(\cdot)$ to be the **Doppler**, **LIDAR**, **Bumps** or **Blocks** functions.
2. Choose M , n , ν_ℓ and the set of dependence parameters α_ℓ for each $l = 1, 2, \dots, M$; and $n = 2^J$ for $J = 12$.
3. Generate M independent FARIMA sequences of length n . Each sequence is standardised, to have the same signal-to-noise ratio,

$$\text{SNR} = 10 \log_{10}(\|g_\ell * f\|^2 / \sigma_\ell^2)$$

for three scenarios where SNR = 10 dB (high noise), 20 dB (medium noise) or 30 dB (low noise). This means that the level of noise compared to the blurred signal is standardised. To simulate the dependent sequence, we use the R-package `fracdiff` and the R-function `fracdiff.sim`.

4. Estimate the highest permissible scale level, \hat{j}_1 by using the estimator \hat{j}_1 defined in (30).
5. Estimate the ‘best channel’ from the noisy data using (31) with σ_ℓ replaced with $\hat{\sigma}_\ell$. Then set the smoothing parameter ζ .
6. Compute \hat{b}_κ using the formula (13) with level-depending thresholds $\lambda_j = \zeta \hat{\tau}_j c_n$ defined in (20), where $\hat{\tau}_j$ and c_n are given (23) and (21), respectively. The noise level in each channel is estimated using the MAD of the wavelet coefficients at the highest scale level ($J - 1$).
7. Compute the above estimates repeatedly to obtain an empirical version of the RMSE with,

$$\widehat{MSE}(\hat{f}, f) = \widehat{\mathbb{E}}\|\hat{f} - f\|_2 = \frac{1}{m} \sum_{i=1}^m \|\hat{f}_i - f\|_2$$

where $m = 1024$.

The results of the simulations are populated in Tables 1–4.

Comments and analysis

The numerical study is considered for three particular contexts. Namely, the effect of the severity of LRD, the effect of multiple channels and the degree of ill-posedness. The method is also compared with using the standard **WaveD** estimator on the ‘best channel’ in the sense of the algorithm posed at the start of this section. The results are contained in Tables 1–4. Simulations were conducted for a large range of noise levels with SNR = 10, 15, 20, 25 and 30 dB but are omitted due to space constraints. The estimates at other noise levels showed similar results to those displayed here and did not add further to the concepts being discussed below.

Performance of our method (and the **WaveD** method) is reliant on two key steps. The most important step is choosing the fine scale level j_1 to truncate the expansion at the highest allowable level before performance deteriorates. A less important but still crucial step is to choose the smoothing parameter ζ appropriately (the smoothing parameter η for the **WaveD** algorithm is fixed at its default of $\sqrt{6}$).

To demonstrate both the role of j_1 and ζ , the RMSE of the estimators in all the forthcoming contexts are presented inside the cells of the tables with the average fine scale level \hat{j}_1 shown in parenthesis. The values

Table 1
RMSE for estimates when $\nu = 0.3$ at mild levels of strong dependence when the number of channels (M) increases.

$\nu = 0.3$	$\alpha = 1$			$\alpha = 0.8$			$\alpha = 0.6$		
	$M = 1$	$M = 2$	$M = 3$	$M = 1$	$M = 2$	$M = 3$	$M = 1$	$M = 2$	$M = 3$
LIDAR: SNR 20 dB									
$\sqrt{\alpha_{\ell_*}}$	0.054 (7)	0.045 (7)	0.041 (7)	0.064 (7)	0.052 (7)	0.046 (7)	0.081 (7)	0.064 (7)	0.056 (7)
$4\sqrt{\alpha_{\ell_*}}$	0.073 (7)	0.06 (7)	0.052 (7)	0.08 (7)	0.065 (7)	0.057 (7)	0.093 (7)	0.074 (7)	0.065 (7)
WaveD	0.06 (7)	0.059 (7)	0.059 (7)	0.064 (7.6)	0.064 (7.8)	0.064 (7.9)	0.083 (8)	0.084 (8)	0.084 (8)
Doppler: SNR 20 dB									
$\sqrt{\alpha_{\ell_*}}$	0.039 (8)	0.03 (8)	0.026 (8)	0.048 (8)	0.036 (8)	0.031 (8)	0.062 (8)	0.046 (8)	0.039 (8)
$4\sqrt{\alpha_{\ell_*}}$	0.056 (8)	0.044 (8)	0.036 (8)	0.059 (8)	0.046 (8)	0.038 (8)	0.064 (8)	0.05 (8)	0.041 (8)
WaveD	0.046 (8)	0.045 (8)	0.045 (8)	0.047 (8)	0.047 (8)	0.047 (8)	0.057 (8)	0.058 (8)	0.058 (8)
Bumps: SNR 20 dB									
$\sqrt{\alpha_{\ell_*}}$	0.275 (7)	0.27 (7)	0.268 (7)	0.28 (7)	0.273 (7)	0.27 (7)	0.288 (7)	0.278 (7)	0.274 (7)
$4\sqrt{\alpha_{\ell_*}}$	0.279 (7)	0.273 (7)	0.271 (7)	0.282 (7)	0.276 (7)	0.273 (7)	0.289 (7)	0.279 (7)	0.276 (7)
WaveD	0.276 (7)	0.276 (7)	0.275 (7)	0.253 (7.2)	0.231 (7.4)	0.215 (7.6)	0.189 (8)	0.188 (8)	0.188 (8)
Blocks: SNR 20 dB									
$\sqrt{\alpha_{\ell_*}}$	0.373 (6)	0.365 (6)	0.363 (6)	0.384 (6)	0.371 (6)	0.367 (6)	0.502 (5)	0.492 (5)	0.489 (5)
$4\sqrt{\alpha_{\ell_*}}$	0.397 (6)	0.373 (6)	0.366 (6)	0.414 (6)	0.382 (6)	0.372 (6)	0.508 (5)	0.495 (5)	0.49 (5)
WaveD	0.376 (6)	0.376 (6)	0.376 (6)	0.385 (6)	0.385 (6)	0.385 (6)	0.408 (6)	0.408 (6)	0.409 (6)

Table 2
RMSE for estimates when $\nu = 0.3$ at severe levels of strong dependence when the number of channels (M) increases.

$\nu = 0.3$	$\alpha = 0.5$			$\alpha = 0.3$			$\alpha = 0.1$		
	$M = 1$	$M = 2$	$M = 3$	$M = 1$	$M = 2$	$M = 3$	$M = 1$	$M = 2$	$M = 3$
LIDAR: SNR 20 dB									
$\sqrt{\alpha_{\ell_*}}$	0.094 (7)	0.073 (7)	0.063 (7)	0.115 (6)	0.089 (6)	0.077 (6)	0.192 (6)	0.141 (6)	0.118 (6)
$4\sqrt{\alpha_{\ell_*}}$	0.102 (7)	0.081 (7)	0.07 (7)	0.122 (6)	0.098 (6)	0.084 (6)	0.164 (6)	0.126 (6)	0.107 (6)
WaveD	0.103 (8)	0.105 (8)	0.105 (8)	0.168 (8)	0.169 (8)	0.171 (8)	0.271 (8)	0.273 (8)	0.273 (8)
Doppler: SNR 20 dB									
$\sqrt{\alpha_{\ell_*}}$	0.072 (7.7)	0.054 (7.9)	0.045 (8)	0.091 (7)	0.073 (7)	0.066 (7)	0.153 (7)	0.113 (7)	0.097 (7)
$4\sqrt{\alpha_{\ell_*}}$	0.068 (7.7)	0.053 (7.9)	0.044 (8)	0.08 (7)	0.065 (7)	0.059 (7)	0.111 (7)	0.086 (7)	0.076 (7)
WaveD	0.069 (8)	0.07 (8)	0.07 (8)	0.107 (8)	0.109 (8)	0.109 (8)	0.17 (8)	0.171 (8)	0.172 (8)
Bumps: SNR 20 dB									
$\sqrt{\alpha_{\ell_*}}$	0.294 (7)	0.281 (7)	0.276 (7)	0.469 (6)	0.461 (6)	0.458 (6)	0.496 (6)	0.475 (6)	0.467 (6)
$4\sqrt{\alpha_{\ell_*}}$	0.294 (7)	0.282 (7)	0.278 (7)	0.467 (6)	0.461 (6)	0.458 (6)	0.489 (6)	0.472 (6)	0.466 (6)
WaveD	0.201 (8)	0.201 (8)	0.202 (8)	0.246 (8)	0.247 (8)	0.248 (8)	0.329 (8)	0.331 (8)	0.33 (8)
Blocks: SNR 20 dB									
$\sqrt{\alpha_{\ell_*}}$	0.511 (5)	0.497 (5)	0.492 (5)	0.82 (4)	0.808 (4)	0.804 (4)	1.116 (3)	1.094 (3)	1.087 (3)
$4\sqrt{\alpha_{\ell_*}}$	0.518 (5)	0.5 (5)	0.494 (5)	0.823 (4)	0.809 (4)	0.805 (4)	1.116 (3)	1.094 (3)	1.087 (3)
WaveD	0.43 (6)	0.43 (6)	0.43 (6)	0.506 (6)	0.507 (6)	0.507 (6)	0.68 (6.1)	0.691 (6.3)	0.697 (6.3)

of ζ are given in the first column (with WaveD denoting the standard WaveD estimator in the best possible channel).

Theoretical arguments suggest that ζ should be at least greater than $4\sqrt{\alpha_{\ell_*}}$ for $p = 2$. Simulations were conducted for more liberal and conservative choices of ζ with $\zeta \in (\sqrt{\alpha_{\ell_*}}, 8\sqrt{\alpha_{\ell_*}})$. In almost all cases, the performance was optimal using the smaller choice of $\zeta = \sqrt{\alpha_{\ell_*}}$. The exceptions generally being when the dependence was considerably strong ($\alpha < 0.3$) and $M = 1$.

As is consistent with [20], allowing higher scales can capture more transient features of a signal but can be at the cost of spurious effects of LRD noise being included. Sometimes early truncation can be beneficial to performance or detrimental to performance based on the features of the signal. For example, the estimation performance on the LIDAR and Doppler signals benefits from the earlier truncation but is detrimental to the estimation of the Bumps and Blocks signals. In the latter estimated signals, the captured transient features at higher scales outweigh the potential loss incurred from including spurious LRD noise effects. A potential reason that the LIDAR signal is estimated well in the multichannel method in simulations compared to the

Table 3RMSE for estimates when $\nu = 0.5$ at mild levels of strong dependence when the number of channels (M) increases.

$\nu = 0.5$	$\alpha = 1$			$\alpha = 0.8$			$\alpha = 0.6$		
	$M = 1$	$M = 2$	$M = 3$	$M = 1$	$M = 2$	$M = 3$	$M = 1$	$M = 2$	$M = 3$
LIDAR: SNR 20 dB									
$\sqrt{\alpha_{\ell_*}}$	0.073 (6)	0.062 (6)	0.056 (6)	0.085 (6)	0.07 (6)	0.063 (6)	0.104 (6)	0.085 (6)	0.075 (6)
$4\sqrt{\alpha_{\ell_*}}$	0.094 (6)	0.081 (6)	0.073 (6)	0.103 (6)	0.088 (6)	0.08 (6)	0.122 (6)	0.1 (6)	0.09 (6)
WaveD	0.083 (6)	0.082 (6)	0.083 (6)	0.086 (6)	0.086 (6)	0.086 (6)	0.105 (6.1)	0.106 (6.2)	0.107 (6.2)
Doppler: SNR 20 dB									
$\sqrt{\alpha_{\ell_*}}$	0.059 (7)	0.053 (7)	0.051 (7)	0.067 (7)	0.058 (7)	0.054 (7)	0.084 (6.9)	0.067 (7)	0.061 (7)
$4\sqrt{\alpha_{\ell_*}}$	0.076 (7)	0.061 (7)	0.056 (7)	0.082 (7)	0.065 (7)	0.059 (7)	0.092 (6.9)	0.071 (7)	0.063 (7)
WaveD	0.065 (7)	0.064 (7)	0.064 (7)	0.067 (7)	0.067 (7)	0.067 (7)	0.078 (7)	0.078 (7)	0.078 (7)
Bumps: SNR 20 dB									
$\sqrt{\alpha_{\ell_*}}$	0.457 (6)	0.455 (6)	0.453 (6)	0.461 (6)	0.457 (6)	0.455 (6)	0.467 (6)	0.46 (6)	0.458 (6)
$4\sqrt{\alpha_{\ell_*}}$	0.457 (6)	0.456 (6)	0.455 (6)	0.461 (6)	0.457 (6)	0.456 (6)	0.467 (6)	0.46 (6)	0.458 (6)
WaveD	0.457 (6)	0.457 (6)	0.457 (6)	0.441 (6.1)	0.429 (6.2)	0.418 (6.3)	0.332 (6.9)	0.319 (7)	0.318 (7)
Blocks: SNR 20 dB									
$\sqrt{\alpha_{\ell_*}}$	0.494 (5)	0.488 (5)	0.486 (5)	0.505 (5)	0.494 (5)	0.49 (5)	0.807 (4)	0.801 (4)	0.8 (4)
$4\sqrt{\alpha_{\ell_*}}$	0.506 (5)	0.493 (5)	0.489 (5)	0.519 (5)	0.501 (5)	0.494 (5)	0.811 (4)	0.803 (4)	0.801 (4)
WaveD	0.497 (5)	0.497 (5)	0.497 (5)	0.506 (5)	0.506 (5)	0.506 (5)	0.527 (5)	0.528 (5)	0.528 (5)

Table 4RMSE for estimates when $\nu = 0.5$ at severe levels of strong dependence when the number of channels (M) increases.

$\nu = 0.5$	$\alpha = 0.5$			$\alpha = 0.3$			$\alpha = 0.1$		
	$M = 1$	$M = 2$	$M = 3$	$M = 1$	$M = 2$	$M = 3$	$M = 1$	$M = 2$	$M = 3$
LIDAR: SNR 20 dB									
$\sqrt{\alpha_{\ell_*}}$	0.11 (5)	0.094 (5)	0.087 (5)	0.142 (5)	0.115 (5)	0.103 (5)	0.215 (4)	0.184 (4)	0.172 (4)
$4\sqrt{\alpha_{\ell_*}}$	0.134 (5)	0.109 (5)	0.097 (5)	0.163 (5)	0.128 (5)	0.113 (5)	0.214 (4)	0.185 (4)	0.173 (4)
WaveD	0.134 (6.5)	0.14 (6.7)	0.142 (6.9)	0.243 (7)	0.246 (7)	0.246 (7)	0.399 (7)	0.401 (7)	0.4 (7)
Doppler: SNR 20 dB									
$\sqrt{\alpha_{\ell_*}}$	0.104 (6)	0.097 (6)	0.094 (6)	0.122 (6)	0.107 (6)	0.101 (6)	0.183 (5)	0.166 (5)	0.16 (5)
$4\sqrt{\alpha_{\ell_*}}$	0.106 (6)	0.098 (6)	0.095 (6)	0.115 (6)	0.103 (6)	0.099 (6)	0.172 (5)	0.161 (5)	0.156 (5)
WaveD	0.091 (7)	0.092 (7)	0.092 (7)	0.14 (7)	0.141 (7)	0.141 (7)	0.216 (7)	0.218 (7)	0.218 (7)
Bumps: SNR 20 dB									
$\sqrt{\alpha_{\ell_*}}$	0.688 (5.2)	0.643 (5.3)	0.611 (5.4)	0.742 (5)	0.736 (5)	0.734 (5)	0.88 (4)	0.873 (4)	0.871 (4)
$4\sqrt{\alpha_{\ell_*}}$	0.689 (5.2)	0.643 (5.3)	0.611 (5.4)	0.743 (5)	0.736 (5)	0.734 (5)	0.88 (4)	0.873 (4)	0.871 (4)
WaveD	0.334 (7)	0.334 (7)	0.334 (7)	0.387 (7)	0.389 (7)	0.389 (7)	0.487 (7)	0.488 (7)	0.488 (7)
Blocks: SNR 20 dB									
$\sqrt{\alpha_{\ell_*}}$	0.813 (4)	0.805 (4)	0.802 (4)	1.085 (3)	1.078 (3)	1.075 (3)	1.126 (3)	1.098 (3)	1.089 (3)
$4\sqrt{\alpha_{\ell_*}}$	0.819 (4)	0.807 (4)	0.803 (4)	1.086 (3)	1.078 (3)	1.075 (3)	1.129 (3)	1.099 (3)	1.089 (3)
WaveD	0.548 (5)	0.549 (5)	0.549 (5)	0.625 (5)	0.626 (5)	0.625 (5)	0.796 (5)	0.798 (5)	0.8 (5)

similar Blocks signal is the close proximity of the jumps combined with the early truncation (small j_1) in the expansion. The WaveD does not truncate early to avoid the LRD effects and hence captures the jumps better (cf. Figs. 1 and 2). Finally in the Bumps signal, the WaveD method consistently outperformed the multichannel estimator (except with the liberal choice with $\zeta = 1$ when $\alpha_{\ell_*} = 1$). This makes sense since the captured high frequency local features of the Bumps signal used with a larger j_1 outweigh the loss incurred by spurious LRD effects. All of the aforementioned points are evident across Tables 1–4 and shown visually as particular cases in Figs. 1 and 2.

Supporting the theory and being consistent with previous results in the literature, as the degree of ill-posedness increases (ν increases), the performance of estimation deteriorates. This is demonstrated by comparing results from Tables 1–2 with the results in Tables 3–4.

In the same vein, as the level of dependence increases (α decreases), the performance deteriorates. Studying Tables 1–4 in more detail, consider the effect of α while keeping M fixed and ν fixed. As is consistent

with the theoretical upper bound on rates of convergence established in Section 4, the convergence rate deteriorates as the level of dependence increases (α decreases).

The theory also suggests that the convergence rate only relies on the best available channel. However, numerically this doesn't seem to be the case. Interestingly, when keeping the dependence and DIP levels fixed across multiple channels, the inclusion of more channels (increasing M) generally results in improved estimation performance for the multichannel estimator in all signals while the WaveD estimator has the same performance across multiple channels. This should not seem surprising since the WaveD estimator is only used the 'best channel' meaning only $n = 4096$ observations are being used each time. The multichannel estimator though is using a weighted average of all channels using 4096, 8192 and 12288 observations respectively in the $M = 1, 2$ and 3 scenarios.

6. Conclusion

In this paper we considered multichannel deconvolution with errors following fractional Brownian motions, with different Hurst parameters. We established upper bounds on the L^p -risk ($1 \leq p < \infty$) for the non-linear wavelet estimators for regular-smooth and box-car convolutions and linear wavelet estimator for super-smooth convolutions. In particular, we extended the findings from [3] and demonstrated that they are no longer valid in the LRD set-up. That is, in the box-car case adding new channels is beneficial for the upper bound only if the additional channel isn't outweighed by the dependence in the sense of $\tilde{\nu}_*$ defined in (8) and the upper bound in Theorem 2. While in the regular-smooth case, adding new channels might perhaps improve the upper bound. An improved upper bound would arise if the α and DIP parameters in the new channel are better in the sense of (5). In both regular-smooth and box-car cases though, LRD affects upper bounds which is consistent with previous findings in [2,19] and [20]. In the super-smooth case, adding new channels is also beneficial, however, the upper bounds do not involve LRD.

We supported our theoretical findings by extensive simulations studies for the regular-smooth case using the L^p -risk for $p = 2$. We found that adding new channels improves performance, especially for severe levels of LRD. On the other hand, the optimal choice of threshold level was in some instances different than the one suggested by the theory. The optimal choice highly depends on the underlying signal. One has to remember though, that the established theory is *asymptotic* in nature, whereas simulations studies are based on finite sample properties. This explains the aforementioned discrepancy.

A possible direction for future research is to explore and extend our upper bounds to minimax type rates towards the direction of [21] obtained for the L^2 -risk in the discrete model when the number of channels, M , also depend on the total number of observations n , i.e., $M = M_n$.

7. Proofs

We provide technical details for the proofs of Theorems 1 and 2. In the regular-smooth and box-car cases, the proofs are based on the maxiset theorem (see Theorem 6.1 in [31]). The steps are similar to those of [14] and [3], with necessary modifications. In the super-smooth case we do not need the maxiset theorem but proceed according to [32] and consider the L^p -risk ($1 \leq p < \infty$) directly.

7.1. Stochastic analysis of estimated wavelet coefficients

By definition, it is clearly seen that the estimated wavelet coefficients have no bias. Consider now the covariance structure of the $z_{\cdot,\ell}$ process where $z_{m,\ell} = \int_{\mathbb{R}} \bar{e}_m(t) dB_{H_\ell}(t)$. It is assumed that, B_{H_ℓ} is independent of $B_{H_{\ell'}}$ for $\ell \neq \ell'$. This has the immediate consequence that, $\text{Cov}(z_{m\ell}, z_{m'\ell'}) = 0$ for $\ell \neq \ell'$. Using the results of Section 5.2 of [20], the covariance of the fBm coefficients within each channel is,

$$\text{Cov}(z_{m\ell}, z_{m'\ell}) = |mm'|^{1/2-H_\ell} \sum_{\kappa'} \psi_m^{\kappa'} \overline{\psi_{m'}^{\kappa'}}, \tag{32}$$

where ψ is the Meyer wavelet and $\kappa' = (j', k')$.

The result in (32) would seem to imply that the covariance matrix of $z_{m\ell}$ is non-trivial. However, applying Lemma 1, the covariance matrix reduces to

$$\text{Cov}(z_{m\ell}, z_{m'\ell}) = |mm'|^{1/2-H_\ell} \sum_{j \in \mathbb{Z}} \mathbb{1}_{\{\frac{m-m'}{2^j} \in \mathbb{Z}\}} \psi_{m2^{-j}} \overline{\psi_{m'2^{-j}}}. \tag{33}$$

Thus we are in a position to bound the variance of the estimated wavelet coefficients (recall $\gamma_{m,\ell}$ are weighting constants),

$$\begin{aligned} \text{Var}(\widehat{b}_\kappa) &= \text{Var} \left(b_\kappa + \sum_{m \in C_j} \sum_{\ell=1}^M \frac{\gamma_{m,\ell} n^{-\alpha_\ell/2} \sigma_\ell \overline{g_{m,\ell}} z_{m\ell} \overline{\Psi_m^\kappa}}{\sum_{\ell=1}^M \gamma_{m,\ell} |g_{m,\ell}|^2} \right) \\ &= \sum_{\ell=1}^M \sum_{m, m' \in C_j} \frac{\gamma_{m,\ell} \gamma_{m',\ell} \sigma_\ell^2 n^{-\alpha_\ell} |mm'|^{1/2-H_\ell} \overline{g_{m,\ell}} g_{m',\ell} \overline{\Psi_m^\kappa} \Psi_{m'}^\kappa}{\left(\sum_{\ell=1}^M \gamma_{m,\ell} |g_{m,\ell}|^2\right) \left(\sum_{\ell=1}^M \gamma_{m',\ell} |g_{m',\ell}|^2\right)} \\ &\quad \times \sum_{j' \in \mathbb{Z}} \mathbb{1}_{\{\frac{m-m'}{2^{j'}} \in \mathbb{Z}\}} \psi_{m2^{-j'}} \overline{\psi_{m'2^{-j'}}}, \end{aligned} \tag{34}$$

where the second last line follows by (33) and the independence of the fBms. Apply Lemma 2 to (34) yields,

$$\text{Var}(\widehat{b}_\kappa) = \sum_{\ell=1}^M \sum_{m \in C_j} \frac{\gamma_{m,\ell}^2 \sigma_\ell^2 n^{-\alpha_\ell} |m|^{1-2H_\ell} |g_{m,\ell}|^2 |\Psi_m^\kappa|^2}{\left(\sum_{\ell=1}^M \gamma_{m,\ell} |g_{m,\ell}|^2\right)^2}. \tag{35}$$

Using the Cauchy Schwarz-inequality we have,

$$\left(\sum_{\ell=1}^M \gamma_{m,\ell} |g_{m,\ell}|^2 \right)^2 \leq \left(\sum_{\ell=1}^M \gamma_{m,\ell}^2 \sigma_\ell^2 n^{-\alpha_\ell} |m|^{1-2H_\ell} |g_{m,\ell}|^2 \right) \left(\sum_{\ell=1}^M \sigma_\ell^{-2} n^{\alpha_\ell} |m|^{2H_\ell-1} |g_{m,\ell}|^2 \right)$$

with equality only if $\gamma_{m,\ell} = \gamma_{m,\ell}^* := n^{\alpha_\ell} \sigma_\ell^{-2} |m|^{2H_\ell-1}$. Use these choice of optimal weights, $\gamma_{m,\ell}^*$, starting with the case of regular-smooth convolution,

$$\begin{aligned} \text{Var}(\widehat{b}_\kappa) &= \sum_{m \in C_j} |\Psi_m^\kappa|^2 \sum_{\ell=1}^M \frac{\gamma_{m,\ell}^2 \sigma_\ell^2 n^{-\alpha_\ell} |m|^{1-2H_\ell} |g_{m,\ell}|^2}{\left(\sum_{\ell=1}^M \gamma_{m,\ell} |g_{m,\ell}|^2\right)^2} \\ &= \sum_{m \in C_j} |\Psi_m^\kappa|^2 \left(\sum_{\ell=1}^M \sigma_\ell^{-2} n^{\alpha_\ell} |m|^{2H_\ell-1} |g_{m,\ell}|^2 \right)^{-1} \\ &\leq C \int_{\mathbb{R}} |\Psi(x)|^2 dx \left(\sum_{\ell=1}^M n^{\alpha_\ell} \inf_{x \in C_j} |x|^{2H_\ell-1} \inf_{y \in C_j} |g_{y,\ell}|^2 \right)^{-1} \\ &= \mathcal{O} \left(\left(\sum_{\ell=1}^M n^{\alpha_\ell} 2^{j(1-\alpha_\ell-2\nu_\ell)} \right)^{-1} \right) \\ &= \mathcal{O} \left(\min_{1 \leq \ell \leq M} n^{-\alpha_\ell} 2^{-j(1-\alpha_\ell-2\nu_\ell)} \right). \end{aligned}$$

Consider the case of box-car convolution. In particular, for $x \in \mathbb{R}$ define the distance to the nearest integer, $\|x\| := \inf\{|x - r| : r \in \mathbb{Z}\}$. Then bounds can be given on the box-car Fourier coefficients with,

$$\left| \frac{2\|mc_\ell\|}{\pi mc_\ell} \right| \leq |g_{m,\ell}| \leq \left| \frac{\|mc_\ell\|}{mc_\ell} \right|,$$

(see for example, p. 298 of [3]). Using this bound with the same optimal weights $\gamma_{m,\ell}^*$ and the bound $|\Psi_m^\kappa| \leq 2^{-j}$ with (35),

$$\begin{aligned} \text{Var}(\widehat{b}_\kappa) &= \sum_{m \in C_j} |\Psi_m^\kappa|^2 \left(\sum_{\ell=1}^M \sigma_\ell^{-2} n^{\alpha_\ell} |m|^{2H_\ell-1} |g_{m,\ell}|^2 \right)^{-1} \\ &\leq \frac{2}{\pi} 2^{-j} n^{-\alpha_*} \sum_{m \in C_j} m^2 \left(\sum_{\ell=1}^M c_\ell^{-2} \sigma_\ell^{-2} |m|^{2H_\ell-1} \|mc_\ell\|^2 \right)^{-1} \\ &= \mathcal{O} \left(2^{j(\alpha_*-2)} n^{-\alpha_*} \sum_{m \in C_j} m^2 \left(\sum_{\ell=1}^M \|mc_\ell\|^2 \right)^{-1} \right) \\ &= \mathcal{O}(n^{-\alpha_*} j 2^{j(1+\alpha_*+1/M)}). \end{aligned}$$

The last bound follows from a result in the proof of Lemma 4 in [3] where,

$$\sum_{m \in C_j} m^2 \left(\sum_{\ell=1}^M \|mc_\ell\|^2 \right)^{-1} = \mathcal{O}(j 2^{j(3+1/M)}).$$

Consider the final case of the super smooth convolution. Using similar arguments it can be shown,

$$\begin{aligned} \text{Var}(\widehat{u}_\kappa) &= \sum_{m \in C_j} |\Phi_m^\kappa|^2 \left(\sum_{\ell=1}^M \sigma_\ell^{-2} n^{\alpha_\ell} |m|^{2H_\ell-1} |g_{m,\ell}|^2 \right)^{-1} \\ &\leq C \sum_{m \in C_j} |\Phi_m^\kappa|^2 \left(\sum_{\ell=1}^M n^{\alpha_\ell} \inf_{x \in C_j} |x|^{2H_\ell-1} \inf_{y \in C_j} |g_{y,\ell}|^2 \right)^{-1} \\ &= \mathcal{O} \left(\left(\sum_{\ell=1}^M n^{\alpha_\ell} 2^{j(1-\alpha_\ell)} \inf_{y \in C_j} |y|^{-2\gamma_\ell} e^{-2\theta_\ell |y|^{\beta_\ell}} \right)^{-1} \right) \\ &= \mathcal{O} \left(\min_{1 \leq \ell \leq M} n^{-\alpha_\ell} 2^{-j(1-\alpha_\ell-2\nu_\ell)} e^{2\theta_\ell 2^{j\beta_\ell}} \right). \end{aligned} \tag{36}$$

7.2. The maxiset theorem

For completeness, we give the statement of the following theorem that is borrowed from Theorem 6.1 in [31]. We also refer to Section 7.3 below for the definition of the Temlyakov property. First, we introduce some notation: μ will denote the measure such that for $j \in \mathbb{N}$, $k \in \mathbb{N}$ and $0 < q < p$,

$$\mu\{(j, k)\} = \|\tau_j \psi_{j,k}\|_p^p = \tau_j^p 2^{j(\frac{p}{2}-1)} \|\psi\|_p^p,$$

$$l_{q,\infty}(\mu) = \left\{ f \in L^p, \sup_{\lambda>0} \lambda^q \mu\{(j, k) : |b_{j,k}| > \tau_j \lambda\} < \infty \right\}.$$

Theorem 3. Let $p > 1$, $0 < q < p$, $\{\psi_{j,k}, j \geq -1, k = 0, 1, \dots, 2^j\}$ be a periodised wavelet basis of $L^2(T)$, $T = [0, 1]$, and τ_j be a positive sequence such that the heteroscedastic basis $\tau_j \psi_{j,k}$ satisfies the Temlyakov property. Suppose that Λ_n is a set of pairs (j, k) and that c_n is a deterministic sequence tending to zero with

$$\sup_n \mu\{\Lambda_n\} c_n^p < \infty. \tag{37}$$

If, for any n and any pair $\kappa = (j, k) \in \Lambda_n$, we have

$$\mathbb{E}|\widehat{b}_\kappa - b_\kappa|^{2p} \leq C(\tau_j c_n)^{2p}, \tag{38}$$

$$\mathbb{P}(|\widehat{b}_\kappa - b_\kappa| \geq \eta \tau_j c_n / 2) \leq C(c_n^{2p} \wedge c_n^A), \tag{39}$$

for some positive constants η and C , then, the wavelet based estimator

$$\widehat{f}_n(t) = \sum_{\kappa \in \Lambda_n} \widehat{b}_\kappa \mathbb{1}_{\{|\widehat{b}_\kappa| \geq \eta \tau_j c_n\}} \psi_\kappa(t), \quad t \in T, \tag{40}$$

is such that, for all positive integers n ,

$$\mathbb{E}\|\widehat{f}_n - f\|_p^p \leq C c_n^{p-q},$$

if and only if

$$f(\cdot) \in l_{q,\infty}(\mu), \tag{41}$$

and

$$\sup_n c_n^{q-p} \left\| f - \sum_{\kappa \in \Lambda_n} b_\kappa \psi_\kappa \right\|_p^p < \infty. \tag{42}$$

This theorem identifies the ‘maxiset’ of a general wavelet thresholding estimator of the form (40). This is done by using conditions (41) and (42) for an appropriate choice of q . In the proofs of Theorems 1 and 2, we will choose q according to the *dense* or the *sparse* regions as follows

$$q = q_d := \frac{(2\nu_* + 1)p}{2s + 2\nu_* + 1}, \quad \text{if } s \geq \frac{2\nu_* + 1}{2} \left(\frac{p}{\pi_0} - 1 \right) \tag{43}$$

$$q = q_s := \frac{(2\nu_* + 1)p/2 - 1}{s - 1/\pi_0 + (2\nu_* + 1)/2}, \quad \text{if } s \leq \frac{2\nu_* + 1}{2} \left(\frac{p}{\pi_0} - 1 \right). \tag{44}$$

We first verify (37). Consider first the case of regular-smooth convolutions. Using (18), simple algebra shows that

$$\begin{aligned} \mu(\{\Lambda_n\}) &= \sum_{j \leq j_1} \sum_{k=0}^{2^j-1} \mu\{(j, k)\} = \sum_{j \leq j_1} 2^j \mu\{(j, k)\} \\ &= \mathcal{O}(1) \sum_{j \leq j_1} 2^j 2^{j(p/2-1)} \tau_j^p = \mathcal{O}(1) \sum_{j \leq j_1} 2^{jp(1/2+\nu_*)} \\ &= \mathcal{O}(2^{j_1 p(1/2+\nu_*)}) = \mathcal{O}(c_n^{-p}), \end{aligned}$$

where c_n is given by (21), since it is easily seen in this case that $\tau_j^2 = \mathcal{O}(2^{2j\nu_*})$ with ν_* given by (6) (compare also with p. 306 of [3]). A similar bound can be shown for the box-car case with ν_* replaced with $\tilde{\nu}_*$ given by (8).

We now verify (38) and (39). Since the random variables $\widehat{b}_\kappa - b_\kappa$ follow a Gaussian distribution, the higher moment bounds (38) follow from the variance inequality. Similarly, denoting Z to be a standard Gaussian distributed random variable,

$$\begin{aligned} \mathbb{P}(|\widehat{b}_\kappa - b_\kappa| > \zeta \tau_j c_n / 2) &= 2\mathbb{P}\left(Z \geq \frac{\zeta \sqrt{\log n}}{2}\right) \\ &\leq \frac{4n^{-\zeta^2/8}}{\zeta \sqrt{\log n}} \\ &= O(c_n^4 \wedge c_n^{2p}), \end{aligned}$$

as long as $\zeta > 2\sqrt{(p \vee 2)2\xi}$. This proves (39).

7.3. Temlyakov property

As seen in Appendix A in [14], the basis $\{\tau_j \psi_{j,k}(\cdot)\}$ satisfies the Temlyakov property as soon as

$$\sum_{j \in \Lambda_n} 2^j \tau_j^2 \leq C \sup_{j \in \Lambda_n} (2^j \tau_j^2) \tag{45}$$

and

$$\sum_{j \in \Lambda_n} 2^{jp/2} \tau_j^p \leq C \sup_{j \in \Lambda_n} (2^{jp/2} \tau_j^p), \quad 1 \leq p < 2. \tag{46}$$

Recall that $\tau_j^2 = \mathcal{O}(2^{2\nu_* j})$ (regular-smooth convolutions) and $\tau_j^2 = \mathcal{O}(j 2^{2\tilde{\nu}_* j})$ (box-car convolutions) ν_* and $\tilde{\nu}_*$ given by (6) and (8). Hence, (45) and (46) are verified by direct calculations.

7.4. Besov embedding and maxiset conditions

We recall that

$$\mathcal{B}_{\pi_0,r}^s \subseteq \mathcal{B}_{p,r}^{s''} \quad \text{if } \pi_0 \geq p, \quad s \geq s''. \tag{47}$$

$$\mathcal{B}_{\pi_0,r}^s \subseteq \mathcal{B}_{p,r}^{s''} \quad \text{if } \pi_0 < p, \quad s - 1/\pi_0 = s'' - 1/p. \tag{48}$$

For both *dense* (43) and *sparse* (44) regions, we look for a Besov scale δ such that $\mathcal{B}_{\pi_0,r}^\delta \subseteq l_{q,\infty}$. As usual, we note that it is easier to work with

$$l_q(\mu) = \left\{ f(\cdot) \in L^p(T) : f = \sum_{j,k} b_{j,k} \psi_{j,k} \text{ such that } \sum_{j,k \in A_j} \frac{|\beta_{j,k}|^q}{\tau_j^q} \|\tau_j \psi_{j,k}\|_p^p < \infty \right\},$$

where A_j is a set of cardinality proportional to 2^j . Using (22) and the fact that

$$\|\tau_j \psi_{j,k}\|_p^p = \tau_j^p 2^{j(p/2-1)} = 2^{j((2\nu_*+1)p/2-1)},$$

we see that $f(\cdot) \in l_q(\mu)$ if,

$$\sum_{j \geq 0} 2^{j((2\nu_*+1)p-2-2\nu_*q)/2} \sum_{k=0}^{2^j-1} |b_{j,k}|^q = \sum_{j \geq 0} 2^{jq[\frac{(2\nu_*+1)(p-q)}{2q} + \frac{1}{2} - \frac{1}{q}]} \sum_{k=0}^{2^j-1} |b_{j,k}|^q < +\infty.$$

From (9), the latter condition holds when

$$f(\cdot) \in \mathcal{B}_{q,q}^\delta(T) \quad \text{for } \delta = \frac{(2\nu_* + 1)}{2} \left(\frac{p}{q} - 1 \right). \tag{49}$$

Now, depending on whether we are in the *dense* (43) or *sparse* (44) regions, we look for s and π such that

$$\mathcal{B}_{\pi_0,r}^s \subseteq \mathcal{B}_{q,q}^\delta. \tag{50}$$

This embedding can be found by exploiting the known monotonicity of Besov balls, namely for $0 < r \leq q$, $\mathcal{B}_{\pi_0,r}^s \subseteq \mathcal{B}_{\pi_0,q}^s$, along with (47) or (48).

The dense region. By definition (43) of $q = q_d$, we have $s \geq (\nu_* + 1/2)(p/\pi_0 - 1)$. Eliminate p by substituting $p = q_d(2s + 2\nu_* + 1)/(2\nu_* + 1)$ yields $\pi_0 \geq q_d$. Hence, (50) follows from (47) as long as $s \geq \delta = \frac{(2\nu_*+1)}{2} \left(\frac{p}{q} - 1 \right)$, which is always true in this *dense* region since $\delta = s > 0$ when $q = q_d$.

The sparse region. Take $q = q_s$ and $\delta = \frac{(2\nu_*+1)}{2} \left(\frac{p}{q_s} - 1 \right) = (2\nu_* + 1) \frac{sp-p/\pi_0+1}{(2\nu_*+1)p-2}$. We consider two cases. If $\pi_0 > q_s$, we use the embedding (47). We have to check that $s > \delta = (2\nu_* + 1) \frac{sp-p/\pi_0+1}{(2\nu_*+1)p-2}$ which is equivalent to $s < \frac{(2\nu_*+1)}{2} \left(\frac{p}{\pi_0} - 1 \right)$, which is true in the *sparse* region. Note that we require $\delta > 0$ which implies either (i) $p > 2/(2\nu_* + 1)$ and $s > 1/\pi_0 - 1/p$ or (ii) $p < 2/(2\nu_* + 1)$ and $s < 1/\pi_0 - 1/p$. The (ii) scenario is impossible since $p < 2/(2\nu_* + 1)$ and $s < 1/\pi_0 - 1/p$ is a contradiction of $s \geq 1/\pi_0 - \nu_* - 1/2$. On the other hand, by definition, when in the sparse phase, $1/\pi_0 - \nu_* - 1/2 < (\nu_* + 1/2)(p/\pi_0 - 1)$ which implies $p > 2/(2\nu_* + 1)$ and consequently verifies that $s > 1/\pi_0 - 1/p$ since $s > 1/\pi_0 - \nu_* - 1/2$. Thus we established (50) for $q_s < \pi_0 < q_d$. By definition (44) of $q = q_s$, if $\pi_0 \leq q_s$, the corresponding δ fulfils $s - 1/\pi_0 = s' - 1/q$. In this case, (48) and (49) ensure that

$$\mathcal{B}_{\pi_0,r}^s \subseteq \mathcal{B}_{q,q}^{s'} \equiv l_q(\mu),$$

as had to be proved.

To apply Theorem 3, (42) needs to be verified. Therefore we need to find a $\delta > 0$ such that for any $f \in \mathcal{B}_{p,r}^\delta$, (42) is satisfied.

$$c_n^{q-p} \left\| f - \sum_{j,k} b_{j,k} \Psi_{j,k} \right\|_p^p = c_n^{q-p} 2^{-j_1 \delta p} \|f\|_{\mathcal{B}_{p,r}^\delta}^p = c_n^{q-p+2\delta p/(2\nu_*+1)} \|f\|_{\mathcal{B}_{p,r}^\delta}^p.$$

The above is bounded uniformly in n if we choose $\delta = 1/2(2\nu_* + 1)(1 - q/p)$. Now we need to find s, π_0 such that $\mathcal{B}_{\pi_0,r}^s \subseteq \mathcal{B}_{p,r}^\delta$.

Consider the first case $\pi_0 \geq p$. This case cannot occur in the sparse phase due to (26) and (29) with the assumption that s is positive. In the dense phase, use embedding (47) with $\gamma = \delta$ and $q = q_d$. Therefore, (47) holds if $s \geq 1/2(2\nu + \alpha)(1 - q_d/p)$. This implies,

$$\begin{aligned} s &\geq 1/2(2\nu_* + 1)(1 - q_d/p) \\ &= \frac{(2\nu_* + 1)s}{2s + 2\nu_* + 1}, \end{aligned}$$

which always holds under the assumption that $s > 0$.

Now consider the dense case when $\pi_0 < p$. In this scenario use embedding (48) by defining $s - 1/\pi_0 = s'' - 1/p$ which ensures $\mathcal{B}_{\pi_0,r}^s \subseteq \mathcal{B}_{p,r}^{s''}$. Then complete the embedding using (47) (namely, $\mathcal{B}_{p,r}^{s''} \subseteq \mathcal{B}_{p,r}^\delta$) which requires $s'' \geq \delta$ with $q = q_d$ or equivalently after rearrangement, $2s'' + (2\nu_* + 1)(1/p - 1/\pi_0) \geq 0$. The left hand side is greater than $(s - 1/\pi_0)(p/\pi_0 - 1)(2\nu_* + 1) \geq 0$ when $s \geq (\nu_* + 1/2)(p/\pi_0 - 1)$ (which is true in the dense phase).

The last case to consider is the sparse case when $\pi_0 < p$. Again introduce a new Besov scale s'' defined with, $s - 1/\pi_0 = s'' - 1/p$ and apply a similar argument to above which requires that, $s'' \geq \delta$ with $q = q_s$. This is satisfied if $s > 1/\pi_0$, which always holds.

7.5. Proofs of Theorem 1 and Theorem 2

The proofs of Theorems 1 and 2 are a direct application of Theorem 3 with j_1, ζ, τ_j , and c_n of Section 2. Combining results of Sections 7.3 and 7.4, we see that all conditions of Theorem 3 are satisfied. Using the embedding results of Section 7.4, we derive the rate exponent $\gamma = \gamma_S$ or $\gamma = \gamma_B$ given by (25) and (28) respectively for smooth and boxcar convolutions for any $f(\cdot) \in \mathcal{B}_{\pi_0,r}^s$ using (43) for q when $s \geq \frac{(2\nu_*+1)}{2}(p/\pi_0 - 1)$ and the rate exponent $\gamma = \gamma_S$ and $\gamma = \gamma_B$ given by (26) and (29) respectively for smooth and boxcar convolutions for any $f(\cdot) \in \mathcal{B}_{\pi_0,r}^s$ using (44) for q when $1/\pi_0 \leq s < \frac{(2\nu_*+1)}{2}(p/\pi_0 - 1)$, with ν_* given either by (6) (regular-smooth convolutions) or (8) (box-car convolutions).

For the super-smooth scenario in Theorem 1 we appeal to the same arguments used in the proof of [32, Theorem 4.2]. Consider the moment bound directly with the estimator (16),

$$\begin{aligned} \mathbb{E} \|\widehat{f}_n - f\|_p^p &\leq 2^{p-1} \mathbb{E} \left\| \sum_{k=0}^{2^{j_0}-1} (\widehat{a}_{j_0,k} - a_{j_0,k}) \Phi_{j_0,k}(t) \right\|_p^p \\ &\quad + 2^{p-1} \left\| \sum_{j=j_0}^{\infty} \sum_{k=0}^{2^j-1} b_{j,k} \Psi_{j,k}(t) \right\|_p^p \end{aligned} \tag{51}$$

The two terms in (51) can be bounded separately with (36) and the scale level (17),

$$\begin{aligned} \mathbb{E} \left\| \sum_{k=0}^{2^{j_0}-1} (\widehat{a}_{j_0,k} - a_{j_0,k}) \Phi_{j_0,k}(t) \right\|_p^p &\leq C 2^{j_0(p/2-1)} \sum_{k=0}^{2^{j_0}-1} \mathbb{E} |\widehat{a}_{j_0,k} - a_{j_0,k}|^p \\ &\leq C n^{-\alpha_{\ell_*} p/2} 2^{j_0 p/2(\alpha_{\ell_*} + 2\nu_{\ell_*})} e^{a_{\ell_*} p 2^{j_0 \beta_{\ell_*}}} \\ &\leq C n^{-\epsilon p/2} (\log n)^{p/2(\alpha_{\ell_*} + 2\nu_{\ell_*})} \\ &= o((\log n)^{-s^* p/\beta_{\ell_*}}). \end{aligned} \tag{52}$$

For the next term use the property of Besov spaces,

$$\begin{aligned} \left\| \sum_{j=j_0}^{\infty} \sum_{k=0}^{2^j-1} b_{j,k} \Psi_{j,k}(t) \right\|_p^p &\leq \left(\sum_{j=j_0}^{\infty} C 2^{-j(s+1/p-1/\min(\pi_0,p))} \right)^p \\ &= \mathcal{O}((\log n)^{-s^* p/\beta_{\ell_*}}). \end{aligned} \tag{53}$$

The result of (27) follows combining the results (51), (52) and (53).

Acknowledgments

The authors would like to thank the Editor and two anonymous reviewers whose comments and suggestions lead to an improved version and presentation of the paper.

Appendix A. Meyer wavelet proofs

Lemma 1. *Let $\omega \in \mathbb{Z}$ and $j \in \mathbb{Z}^+$, then the following identity holds for the sum of the dyadic rationals on the complex unit circle.*

$$\sum_{k=0}^{2^j-1} e^{2\pi i \omega k 2^{-j}} = 2^j \mathbb{1}_{\{\frac{\omega}{2^j} \in \mathbb{Z}\}}$$

Proof. The proof relies on the trigonometric components (real and imaginary parts) of the complex exponential. Namely,

$$e^{2\pi i \omega k 2^{-j}} = \cos(2\pi \omega k 2^{-j}) + i \sin(2\pi \omega k 2^{-j}).$$

The case when $\omega = 0$ follows immediately due to the identities that $\cos(0) = 1$ and $\sin(0) = 0$. Consider now the case when $\omega \neq 0$. Starting with the real part, partition the summation into halves with,

$$\begin{aligned} \sum_{k=0}^{2^j-1} \cos(2\pi \omega k 2^{-j}) &= \sum_{k=0}^{2^{j-1}-1} \cos(2\pi \omega k 2^{-j}) + \sum_{k=2^{j-1}}^{2^j-1} \cos(2\pi \omega k 2^{-j}) \\ &= \sum_{k=0}^{2^{j-1}-1} \cos(2\pi \omega k 2^{-j}) + \sum_{k=0}^{2^j-2^{j-1}-1} \cos(2\pi \omega k 2^{-j} + \pi \omega) \\ &= (1 + (-1)^\omega) \sum_{k=0}^{2^{j-1}-1} \cos(2\pi \omega k 2^{-j}). \end{aligned}$$

If ω is odd, $\omega = 2s + 1$ for some $s \in \mathbb{Z}$ then the above result is zero. Therefore consider $\omega = 2s$ for some $s \in \mathbb{Z}$ (ω is even).

$$\begin{aligned} \sum_{k=0}^{2^j-1} \cos(2\pi \omega k 2^{-j}) &= (1 + (-1)^\omega) \sum_{k=0}^{2^{j-1}-1} \cos(2\pi \omega k 2^{-j}) \\ &= 2 \sum_{k=0}^{2^{j-1}-1} \cos(2\pi \omega k 2^{-j}) \\ &= 2 \left\{ \sum_{k=0}^{2^{j-2}-1} \cos(2\pi \omega k 2^{-j}) + \sum_{k=2^{j-2}}^{2^{j-1}-1} \cos(2\pi \omega k 2^{-j}) \right\} \\ &= 2 \left\{ \sum_{k=0}^{2^{j-2}-1} \cos(2\pi \omega k 2^{-j}) + \sum_{k=0}^{2^{j-2}-1} \cos\left(2\pi \omega k 2^{-j} + \frac{\pi \omega}{2}\right) \right\} \\ &= 2 \left\{ \sum_{k=0}^{2^{j-2}-1} \cos(2\pi \omega k 2^{-j}) + \sum_{k=0}^{2^{j-2}-1} \cos(2\pi \omega k 2^{-j} + \pi s) \right\} \end{aligned}$$

$$= 2(1 + (-1)^s) \sum_{k=0}^{2^{j-2}-1} \cos(2\pi\omega k 2^{-j}).$$

If s is odd then the above result is zero. This process can be repeated until we reach the last possible result where $\omega = C2^j$ for some $C \in \mathbb{Z}$ and,

$$\sum_{k=0}^{2^j-1} \cos(2\pi\omega k 2^{-j}) = 2^j \mathbb{1}_{\{k=0\}} \cos(2\pi\omega k 2^{-j}) = 2^j.$$

A similar proof applies for the imaginary part except the final step has,

$$\sum_{k=0}^{2^j-1} \sin(2\pi\omega k 2^{-j}) = 2^j \mathbb{1}_{\{k=0\}} \sin(2\pi\omega k 2^{-j}) = 0. \quad \square$$

Lemma 2. *Let (ϕ, ψ) be the Meyer wavelet basis. That is, the mother Meyer wavelet defined in the Fourier domain with,*

$$\psi_m = \int_{\mathbb{R}} e^{-2\pi i m x} \psi(x) dx = e^{i\pi m} \begin{cases} \sin(\frac{\pi}{2}\nu(3|m| - 1)) & \text{for } \frac{1}{3} \leq |m| \leq \frac{2}{3}; \\ \cos(\frac{\pi}{2}\nu(\frac{3}{2}|m| - 1)) & \text{for } \frac{2}{3} \leq |m| \leq \frac{4}{3}; \\ 0 & \text{otherwise,} \end{cases} \quad (\text{A.1})$$

where $\nu(x)$ is a polynomial that controls the vanishing moment properties of the wavelet basis. In particular, the Meyer wavelet has the defining property that the polynomial satisfies,

$$\nu(x) + \nu(1 - x) = 1. \quad (\text{A.2})$$

Then the matrix $\mathbf{M} = (M_{m,m'})_{m,m' \in C_j}$ defined with entries

$$M_{m,m'} = \sum_{j' \in \mathbb{Z}} \mathbb{1}_{\{\frac{m-m'}{2^j} \in \mathbb{Z}\}} \psi_{m2^{-j}} \overline{\psi_{m'2^{-j'}}$$

is the identity matrix.

Proof. Using the definition of (A.1) and considering $a \in \mathbb{Z}$ such that $(m - m')2^{-j} = a$ we can write,

$$\mathbb{1}_{\{\frac{m-m'}{2^j} \in \mathbb{Z}\}} \psi_{m2^{-j}} \overline{\psi_{m'2^{-j}}} = e^{ia\pi} f(|m2^{-j}|) f(|m'2^{-j}|)$$

where f is defined by the piece wise trigonometric functions given in (A.1). Recall the support of the Meyer wavelet at scale j in the Fourier domain is C_j defined in (14). Define a partition of the domain at level j with the domain at the surrounding scales, $j - 1$ and $j + 1$ with,

$$C_j \cap C_{j-1} = \left\{ a \in \mathbb{Z} : \frac{1}{3} \leq |a2^{-j}| \leq \frac{2}{3} \right\} =: C_{j-1}^{\sin}$$

$$C_j \cap C_{j+1} = \left\{ a \in \mathbb{Z} : \frac{2}{3} \leq |a2^{-j}| \leq \frac{4}{3} \right\} =: C_{j+1}^{\cos}.$$

The sets are named C_{j-1}^{\sin} and C_{j+1}^{\cos} respectively since they refer to those trigonometric parts of the Meyer wavelet at scale j respectively (see (A.1)) and it is the domain where the coefficients are in both $C_j \cap C_{j-1}$ and $C_j \cap C_{j+1}$ respectively. To ease the tedious nature of the forthcoming argument, consider a particular ordering of the two sets C_{j+1}^{\cos} and C_{j-1}^{\sin} .

$$C_{j-1}^{\sin} = \left\{ -\left\lfloor \frac{2^{j+1}}{3} \right\rfloor, -\left\lfloor \frac{2^{j+1}}{3} \right\rfloor + 1, \dots, -\left\lfloor \frac{2^j}{3} \right\rfloor, \left\lfloor \frac{2^j}{3} \right\rfloor, \left\lfloor \frac{2^j}{3} \right\rfloor + 1, \dots, \left\lfloor \frac{2^{j+1}}{3} \right\rfloor \right\}.$$

$$C_{j+1}^{\cos} = \left\{ -\left\lfloor \frac{2^{j+2}}{3} \right\rfloor, -\left\lfloor \frac{2^{j+2}}{3} \right\rfloor + 1, \dots, -\left\lfloor \frac{2^{j+1}}{3} \right\rfloor, \left\lfloor \frac{2^{j+1}}{3} \right\rfloor, \left\lfloor \frac{2^{j+1}}{3} \right\rfloor + 1, \dots, \left\lfloor \frac{2^{j+2}}{3} \right\rfloor \right\}.$$

Further partition these sets into the positive and negative parts with $C_{j-1}^{\sin+} := C_{j-1}^{\sin} \cap \mathbb{Z}^+$, $C_{j-1}^{\sin-} := C_{j-1}^{\sin} \cap \mathbb{Z}^-$, $C_{j+1}^{\cos+} := C_{j+1}^{\cos} \cap \mathbb{Z}^+$ and $C_{j+1}^{\cos-} := C_{j+1}^{\cos} \cap \mathbb{Z}^-$ where \mathbb{Z}^+ and \mathbb{Z}^- are the positive and negative integers respectively.

Write the matrix M in the following way,

$$M = \begin{array}{c} \begin{array}{c} C_{j+1}^{\cos-} \\ C_{j-1}^{\sin-} \\ C_{j-1}^{\sin+} \\ C_{j+1}^{\cos+} \end{array} \\ \left(\begin{array}{c|c|c|c} C_{j+1}^{\cos-} & C_{j-1}^{\sin-} & C_{j-1}^{\sin+} & C_{j+1}^{\cos+} \\ \mathbf{E}_1 & \mathbf{0} & \mathbf{0} & \mathbf{E}_2 \\ \hline \mathbf{0} & \mathbf{R}_1 & \mathbf{R}_2 & \mathbf{0} \\ \hline \mathbf{0} & \mathbf{R}_3 & \mathbf{R}_4 & \mathbf{0} \\ \hline \mathbf{E}_3 & \mathbf{0} & \mathbf{0} & \mathbf{E}_4 \end{array} \right) \end{array}$$

where the outer sets denote the values of m, m' inside the M matrix. With a slight abuse of notation we will refer to the elements of M using $m, m' \in C_j$. For example, the first element $M_{1,1}$ (the top left matrix entry of \mathbf{E}_1) has $m = m' = -\lfloor \frac{2^{j+1}}{3} \rfloor$, the first element of $C_{j+1}^{\cos-}$.

The matrix M is composed of block matrix components where $\mathbf{0}$ denotes a matrix of zeros of appropriate size implied by the cardinalities of $C_{j+1}^{\cos-}$, $C_{j-1}^{\sin-}$, $C_{j-1}^{\sin+}$ and $C_{j+1}^{\cos+}$. The zero matrices follow since a value $m \in C_j$ cannot be in both C_{j-1} and C_{j+1} since $C_{j-1} \cap C_{j+1} = \emptyset$. The overall result follows by showing that the other block matrices simplify to the following: $\mathbf{E}_1 = \mathbf{R}_1 = \mathbf{R}_4 = \mathbf{E}_4 = \mathbf{I}$ and $\mathbf{E}_2 = \mathbf{R}_2 = \mathbf{R}_3 = \mathbf{E}_3 = \mathbf{0}$ where \mathbf{I} is the identity matrix of appropriate size. To show these results for each case, one needs to first consider the values of $m, m' \in C_j$ such that $m - m'$ is a factor of 2^k for $k \in \{j, j - 1, j + 1\}$ and then compute the sum $\sum_{k=j-1}^{j+1} \psi_{m2^{-k}} \psi_{m2^{-k}}$ for these values.

Before proceeding, some notation is defined. For $x \in \mathbb{R}$, let $\{x\}$ denote the fractional part of x . Then we have,

$$\lfloor x \rfloor = x - \{x\} \quad \text{and} \quad \lceil x \rceil = x + 1 - \{x\}.$$

Case \mathbf{R}_1 and \mathbf{R}_4

We will consider here only the case for \mathbf{R}_4 , the case for \mathbf{R}_1 follows by symmetry. In this context, $m, m' \in C_{j-1}^{\sin+}$ where it is possible that $m = m'$ is a solution to $m - m' = s2^j$ with $s = 0$. This is in fact the only solution since the cardinality of $C_{j-1}^{\sin+} < 2^{j-1}$. Indeed,

$$\begin{aligned} |C_{j-1}^{\sin+}| &= \left\lfloor \frac{2^{j+1}}{3} \right\rfloor - \left\lfloor \frac{2^j}{3} \right\rfloor + 1 \\ &= \frac{2^{j+1}}{3} - \left\{ \frac{2^{j+1}}{3} \right\} - \frac{2^j}{3} - 1 + \left\{ \frac{2^j}{3} \right\} + 1 \\ &= \frac{2^j - (-1)^j}{3} < 2^{j-1}. \end{aligned}$$

Therefore the only value of $s \in \mathbb{Z}$ such that $m - m' = s2^j$ or $m - m' = s2^{j-1}$ is $s = 0$ ($m = m'$). This scenario occurs along the diagonal of \mathbf{R}_4 , therefore the off-diagonal elements are zero. Computing the diagonal elements, we have $m, m' \in C_{j-1}^{\sin+} \implies 2m, 2m' \in C_{j+1}^{\cos+}$ and $\frac{m}{2}, \frac{m'}{2} \notin C_j$. Therefore only the scales $j - 1$ and j are used in the summation. Consider these diagonal elements of \mathbf{R}_4 with,

$$\begin{aligned}
 R_4 &= \mathbb{1}_{\{m,m' \in C_{j-1}^{\sin+} : m=m'\}} \{ \psi_{m2^{-j}} \overline{\psi_{m'2^{-j}}} + \psi_{m2^{-j+1}} \overline{\psi_{m'2^{-j+1}}} \} \\
 &= \mathbb{1}_{\{m,m' \in C_{j-1}^{\sin+} : m=m'\}} \psi_{m2^{-j}} \overline{\psi_{m'2^{-j}}} \\
 &\quad + \mathbb{1}_{\{2m,2m' \in C_{j+1}^{\cos+} : m=m'\}} \psi_{2m2^{-j}} \overline{\psi_{2m'2^{-j}}} \\
 &= \mathbb{1}_{\{m=m'\}} \left(\sin^2 \left(\frac{\pi}{2} \nu (3|m2^{-j}| - 1) \right) + \cos^2 \left(\frac{\pi}{2} \nu \left(\frac{3}{2} |m2^{-j+1}| - 1 \right) \right) \right) \\
 &= 1
 \end{aligned}$$

since $\sin^2 \theta + \cos^2 \theta = 1$ for all $\theta \in \mathbb{R}$.

Case **R₂** and **R₃**

Similarly, we will consider here only the case for **R₃**, the case for **R₂** follows by symmetry. In this context, $m \in C_{j-1}^{\sin+}$ and $m' \in C_{j-1}^{\sin-}$. Consider the values $m - m'$ along the main diagonal of **R₃** which are identical since the values $m \in C_{j-1}^{\sin+}$ and $m' \in C_{j-1}^{\sin-}$ are consecutive. The first diagonal element is when $m = \lfloor \frac{2^{j+1}}{3} \rfloor$ and $m' = -\lceil \frac{2^j}{3} \rceil$ yielding,

$$\begin{aligned}
 m - m' &= \left\lfloor \frac{2^{j+1}}{3} \right\rfloor + \left\lceil \frac{2^j}{3} \right\rceil \\
 &= \frac{2^{j+1}}{3} - \left\{ \frac{2^{j+1}}{3} \right\} + \frac{2^j}{3} + 1 - \left\{ \frac{2^j}{3} \right\} \\
 &= 2^j - \left\{ \frac{2^{j+1}}{3} \right\} + 1 - \left\{ \frac{2^j}{3} \right\} \\
 &= 2^j - \left\{ 2^j - \frac{2^j}{3} \right\} - \left\{ \frac{2^j}{3} \right\} + 1 \\
 &= 2^j - 1 + 1 = 2^j.
 \end{aligned}$$

Similarly the maximum and minimum distances are, $\max_{m \in C_{j-1}^{\sin+}, m' \in C_{j-1}^{\sin-}} (m - m') = 2 \lfloor \frac{2^{j+1}}{3} \rfloor$ and $\min_{m \in C_{j-1}^{\sin+}, m' \in C_{j-1}^{\sin-}} (m - m') = 2 \lceil \frac{2^j}{3} \rceil$. Therefore the range of possible distances between m and m' are of length $2(\lfloor \frac{2^{j+1}}{3} \rfloor - \lceil \frac{2^j}{3} \rceil) = \frac{2^{j+1} - 2(-1)^j}{3} - 2 < 2^j$. Therefore $m - m' = s2^j$ for some $s \in \mathbb{Z}$ only on the diagonal and $s = 1$ in this case ($m = m' + 2^j$). Thus again, **R₃** is a diagonal matrix. Computing these values we have, $m \in C_{j-1}^{\sin+} \implies 2m \in C_{j+1}^{\cos+}$ and $\frac{m}{2} \notin C_j$. Similar cases apply to m' . Therefore only the scales $j - 1$ and j are used in the summation. Consider these diagonal elements of **R₃**,

$$\begin{aligned}
 R_3 &= \mathbb{1}_{\{m \in C_{j-1}^{\sin+}, m' \in C_{j-1}^{\sin-} : m=m'+2^j\}} \{ \psi_{m2^{-j}} \overline{\psi_{m'2^{-j}}} + \psi_{m2^{-j+1}} \overline{\psi_{m'2^{-j+1}}} \} \\
 &= \mathbb{1}_{\{m=m'+2^j\}} \psi_{m2^{-j}} \overline{\psi_{m'2^{-j}}} \mathbb{1}_{\{m \in C_{j-1}^{\sin+}, m' \in C_{j-1}^{\sin-}\}} \\
 &\quad + \mathbb{1}_{\{m=m'+2^j\}} \psi_{2m2^{-j}} \overline{\psi_{2m'2^{-j}}} \mathbb{1}_{\{2m \in C_{j+1}^{\cos+}, 2m' \in C_{j+1}^{\cos-}\}} \\
 &= \mathbb{1}_{\{m=m'+2^j\}} e^{i\pi(m-m')2^{-j}} \sin \left(\frac{\pi}{2} \nu (3|m2^{-j}| - 1) \right) \sin \left(\frac{\pi}{2} \nu (3|m'2^{-j}| - 1) \right) \\
 &\quad + \mathbb{1}_{\{m=m'+2^j\}} e^{i\pi(m-m')2^{-j-1}} \cos \left(\frac{\pi}{2} \nu \left(\frac{3}{2} |m2^{-j+1}| - 1 \right) \right) \cos \left(\frac{\pi}{2} \nu \left(\frac{3}{2} |m'2^{-j+1}| - 1 \right) \right) \\
 &= -\mathbb{1}_{\{m=m'+2^j\}} \sin \left(\frac{\pi}{2} \nu (3|m2^{-j}| - 1) \right) \sin \left(\frac{\pi}{2} \nu (3|m'2^{-j}| - 1) \right) \\
 &\quad + \mathbb{1}_{\{m=m'+2^j\}} \cos \left(\frac{\pi}{2} \nu (3|m2^{-j}| - 1) \right) \cos \left(\frac{\pi}{2} \nu (3|m'2^{-j}| - 1) \right)
 \end{aligned}$$

$$= \mathbb{1}_{\{m=m'+2^j\}} \cos\left(\frac{\pi}{2}\nu(3|m2^{-j}| - 1) + \frac{\pi}{2}\nu(3|m'2^{-j}| - 1)\right).$$

Exploit now the fact that $\mathbb{1}_{\{m \in C_{j-1}^{\sin+}, m' \in C_{j-1}^{\sin-}\}}$ implying $m > 0$ and $m' < 0$ along with the defining property of the Meyer wavelet in (A.2) with the specific choice $x = 3m2^{-j} - 1$,

$$\begin{aligned} R_3 &= \mathbb{1}_{\{m=m'+2^j\}} \cos\left(\frac{\pi}{2}\nu(3|m2^{-j}| - 1) + \frac{\pi}{2}\nu(3|m'2^{-j}| - 1)\right) \\ &= \mathbb{1}_{\{-m'2^{-j}=1-m2^{-j}\}} \cos\left(\frac{\pi}{2}\nu(3m2^{-j} - 1) + \frac{\pi}{2}\nu(-3m'2^{-j} - 1)\right) \\ &= \cos\left(\frac{\pi}{2}\nu(3m2^{-j} - 1) + \frac{\pi}{2}\nu(3 - 3m2^{-j} - 1)\right) \\ &= \cos\left(\frac{\pi}{2}\right) = 0. \end{aligned}$$

Therefore $R_3 = 0$.

Case E_1 and E_4

Similarly, we will consider here only the case for E_4 , the case for E_1 follows by symmetry. In this context, $m, m' \in C_{j+1}^{\cos+}$ and we apply a similar argument used in the cases for R_1 and R_4 . Again, $m = m'$ is the only solution to $m - m' = s2^j$ with $s = 0 \in \mathbb{Z}$. Indeed,

$$\begin{aligned} |C_{j+1}^{\cos+}| &= \left\lfloor \frac{2^{j+2}}{3} \right\rfloor - \left\lfloor \frac{2^{j+1}}{3} \right\rfloor + 1 \\ &= \frac{2^{j+2}}{3} - \left\{ \frac{2^{j+2}}{3} \right\} - \frac{2^{j+1}}{3} - 1 + \left\{ \frac{2^{j+1}}{3} \right\} + 1 \\ &= \frac{2^{j+1} - (-1)^{j+1}}{3} < 2^j. \end{aligned}$$

Therefore the only value of $s \in \mathbb{Z}$ such that $m - m' = s2^j$ or $m - m' = s2^{j+1}$ is $s = 0$ ($m = m'$). This scenario occurs along the diagonal of E_4 which is therefore zero on the off-diagonal. Computing these values we have, $m, m' \in C_{j+1}^{\cos+} \implies \frac{m}{2}, \frac{m'}{2} \in C_{j+1}^{\sin+}$ and $2m, 2m' \notin C_j$. Therefore only the scales j and $j + 1$ are used in the summation. Consider these diagonal elements of E_4 with,

$$\begin{aligned} E_4 &= \mathbb{1}_{\{m, m' \in C_{j+1}^{\cos+}, m=m'\}} \{ \psi_{m2^{-j}} \overline{\psi_{m'2^{-j}}} + \psi_{m2^{-(j+1)}} \overline{\psi_{m'2^{-(j+1)}}} \} \\ &= \mathbb{1}_{\{m=m'\}} e^{i\pi(m-m')2^{-j}} \cos\left(\frac{\pi}{2}\nu\left(\frac{3}{2}|m2^{-j}| - 1\right)\right) \cos\left(\frac{\pi}{2}\nu\left(\frac{3}{2}|m'2^{-j}| - 1\right)\right) \\ &\quad + \mathbb{1}_{\{m=m'\}} e^{i\pi(m-m')2^{-(j+1)}} \sin\left(\frac{\pi}{2}\nu(3|m2^{-(j+1)}| - 1)\right) \sin\left(\frac{\pi}{2}\nu(3|m'2^{-(j+1)}| - 1)\right) \\ &= 1 \end{aligned}$$

since $\sin^2 \theta + \cos^2 \theta = 1$ for all $\theta \in \mathbb{R}$.

Case E_2 and E_3

Lastly, consider the case for E_3 , the case for E_2 follows by symmetry. In this context, $m \in C_{j+1}^{\cos+}$ and $m' \in C_{j+1}^{\cos-}$. Consider the differences $m - m'$ along the main diagonal of E_3 which are identical since the values of $m \in C_{j+1}^{\cos+}$ and $m' \in C_{j+1}^{\cos-}$ are consecutive. The first diagonal element is when $m = \lceil \frac{2^{j+1}}{3} \rceil$ and $m' = -\lfloor \frac{2^{j+2}}{3} \rfloor$ yielding,

$$\begin{aligned}
 m - m' &= \left\lceil \frac{2^{j+1}}{3} \right\rceil + \left\lfloor \frac{2^{j+2}}{3} \right\rfloor \\
 &= \frac{2^{j+1}}{3} + 1 - \left\{ \frac{2^{j+1}}{3} \right\} + \frac{2^{j+2}}{3} - \left\{ \frac{2^{j+2}}{3} \right\} \\
 &= 2^{j+1} - \left\{ \frac{2^{j+1}}{3} \right\} - \left\{ \frac{2^{j+2}}{3} \right\} + 1 \\
 &= 2^{j+1} - \left\{ \frac{2^{j+1}}{3} \right\} - \left\{ 2^{j+1} - \frac{2^{j+1}}{3} \right\} + 1 \\
 &= 2^{j+1} - 1 + 1 = 2^{j+1}.
 \end{aligned}$$

Similarly the maximum and minimum distances are, $\max_{m \in C_{j+1}^{\text{cos}+}, m' \in C_{j+1}^{\text{cos}-}} (m - m') = 2 \lfloor \frac{2^{j+2}}{3} \rfloor$ and $\min_{m \in C_{j+1}^{\text{cos}+}, m' \in C_{j+1}^{\text{cos}-}} (m - m') = 2 \lceil \frac{2^{j+1}}{3} \rceil$. Therefore the range of possible distances between m and m' are of length $2(\lfloor \frac{2^{j+2}}{3} \rfloor - \lceil \frac{2^{j+1}}{3} \rceil) = \frac{2^{j+2} - 2(-1)^{j+1}}{3} - 2 < 2^{j+1}$. Therefore $m - m' = s2^j$ for some $s \in \mathbb{Z}$ only on the diagonal and $s = 2$ in this case ($m = m' + 2^{j+1}$). Thus again, \mathbf{E}_3 is a diagonal matrix. Computing these values we have, $m \in C_{j+1}^{\text{cos}+} \implies \frac{m}{2} \in C_{j+1}^{\text{sin}+}$ and $2m \notin C_j$. Similar cases apply to m' . Therefore only the scales j and $j + 1$ are used in the summation. Consider these diagonal elements of \mathbf{E}_3 ,

$$\begin{aligned}
 E_3 &= \mathbb{1}_{\{m \in C_{j+1}^{\text{cos}+}, m' \in C_{j+1}^{\text{cos}-} : m = m' + 2^{j+1}\}} (\psi_{m2^{-j}} \overline{\psi_{m'2^{-j}}} + \psi_{m2^{-(j+1)}} \overline{\psi_{m'2^{-(j+1)}}}) \\
 &= \mathbb{1}_{\{m = m' + 2^{j+1}\}} e^{i\pi(m-m')2^{-j}} \cos\left(\frac{\pi}{2}\nu\left(\frac{3}{2}|m2^{-j}| - 1\right)\right) \cos\left(\frac{\pi}{2}\nu\left(\frac{3}{2}|m'2^{-j}| - 1\right)\right) \\
 &\quad + \mathbb{1}_{\{m = m' + 2^{j+1}\}} e^{i\pi(m-m')2^{-j-1}} \sin\left(\frac{\pi}{2}\nu\left(3\left|\frac{m}{2}2^{-j}\right| - 1\right)\right) \sin\left(\frac{\pi}{2}\nu\left(3\left|\frac{m'}{2}2^{-j}\right| - 1\right)\right) \\
 &= -\mathbb{1}_{\{m = m' + 2^{j+1}\}} \sin\left(\frac{\pi}{2}\nu\left(\frac{3}{2}|m2^{-j}| - 1\right)\right) \sin\left(\frac{\pi}{2}\nu\left(\frac{3}{2}|m'2^{-j}| - 1\right)\right) \\
 &\quad + \mathbb{1}_{\{m = m' + 2^{j+1}\}} \cos\left(\frac{\pi}{2}\nu\left(\frac{3}{2}|m2^{-j}| - 1\right)\right) \cos\left(\frac{\pi}{2}\nu\left(\frac{3}{2}|m'2^{-j}| - 1\right)\right) \\
 &= \mathbb{1}_{\{m = m' + 2^{j+1}\}} \cos\left(\frac{\pi}{2}\nu\left(\frac{3}{2}|m2^{-j}| - 1\right)\right) + \frac{\pi}{2}\nu\left(\frac{3}{2}|m'2^{-j}| - 1\right).
 \end{aligned}$$

Again, exploit the fact that $\mathbb{1}_{\{m \in C_{j+1}^{\text{cos}+}, m' \in C_{j+1}^{\text{cos}-}\}}$ implying $m > 0$ and $m' < 0$ along with the Meyer polynomial property in (A.2) with the specific choice $x = \frac{3}{2}m2^{-j} - 1$,

$$\begin{aligned}
 E_3 &= \mathbb{1}_{\{m = m' + 2^{j+1}\}} \cos\left(\frac{\pi}{2}\nu\left(\frac{3}{2}|m2^{-j}| - 1\right)\right) + \frac{\pi}{2}\nu\left(\frac{3}{2}|m'2^{-j}| - 1\right) \\
 &= \mathbb{1}_{\{-m'2^{-j} = 2 - m2^{-j}\}} \cos\left(\frac{\pi}{2}\nu\left(\frac{3}{2}m2^{-j} - 1\right)\right) + \frac{\pi}{2}\nu\left(-\frac{3}{2}m'2^{-j} - 1\right) \\
 &= \cos\left(\frac{\pi}{2}\nu\left(\frac{3}{2}m2^{-j} - 1\right)\right) + \frac{\pi}{2}\nu\left(3 - \frac{3}{2}m2^{-j} - 1\right) \\
 &= \cos\left(\frac{\pi}{2}\right) = 0.
 \end{aligned}$$

Therefore $\mathbf{E}_3 = \mathbf{0}$ which completes the proof. \square

References

- [1] Y. Wang, Function estimation via wavelet shrinkage for long-memory data, *Ann. Statist.* 24 (1996) 466–484.
- [2] Y. Wang, Minimax estimation via wavelets for indirect long-memory data, *J. Statist. Plann. Inference* 64 (1997) 45–55.
- [3] D. De Canditiis, M. Pensky, Simultaneous wavelet deconvolution in periodic setting, *Scand. J. Stat.* 33 (2006) 293–306.
- [4] M. Pensky, T. Sapatinas, Functional deconvolution in a periodic setting: uniform case, *Ann. Statist.* 37 (2009) 73–104.
- [5] M. Pensky, T. Sapatinas, On convergence rates equivalency and sampling strategies in functional deconvolution models, *Ann. Statist.* 38 (2010) 1793–1844.
- [6] M. Pensky, T. Sapatinas, Multichannel boxcar deconvolution with growing number of channels, *Electron. J. Stat.* 5 (2011) 53–82.
- [7] S.D. Casey, D.F. Walnut, Systems of convolution equations, deconvolution, Shannon sampling, and the wavelet and Gabor transforms, *SIAM Rev.* 36 (1994) 537–577.
- [8] D. De Canditiis, M. Pensky, Discussion on the meeting on ‘statistical approaches to inverse problems’, *J. R. Stat. Soc. Ser. B Stat. Methodol.* 66 (2004) 627–652.
- [9] J. Beran, Y. Feng, S. Ghosh, R. Kulik, *Long-Memory Processes: Probabilistic Properties and Statistical Methods*, Springer, Berlin, Heidelberg, 2013, <http://dx.doi.org/10.1007/978-3-642-35512-7>.
- [10] M. Raimondo, M. Stewart, The waved transform in r: performs fast translation-invariant wavelet deconvolution, *J. Stat. Softw.* 21 (2007) 1–28.
- [11] D.L. Donoho, Nonlinear solution of linear inverse problems by wavelet-vaguelette decomposition, *Appl. Comput. Harmon. Anal.* 2 (1995) 101–126.
- [12] F. Abramovich, B.W. Silverman, Wavelet decomposition approaches to statistical inverse problems, *Biometrika* 85 (1998) 115–129.
- [13] J. Kalifa, S. Mallat, Thresholding estimators for linear inverse problems and deconvolutions, *Ann. Statist.* 31 (2003) 58–109.
- [14] I.M. Johnstone, G. Kerkycharian, D. Picard, M. Raimondo, Wavelet deconvolution in a periodic setting, *J. R. Stat. Soc. Ser. B Stat. Methodol.* 66 (2004) 547–573.
- [15] D.L. Donoho, M.E. Raimondo, Translation invariant deconvolution in a periodic setting, *Int. J. Wavelets Multiresolut. Inf. Process.* 2 (2004) 415–431.
- [16] I.M. Johnstone, M. Raimondo, Periodic boxcar deconvolution and Diophantine approximation, *Ann. Statist.* 32 (2004) 1781–1804.
- [17] R. Neelamani, H. Choi, R. Baraniuk, ForWaRD: Fourier-wavelet regularized deconvolution for ill-conditioned systems, *IEEE Trans. Signal Process.* 52 (2004) 418–433.
- [18] G. Kerkycharian, D. Picard, M. Raimondo, Adaptive boxcar deconvolution on full Lebesgue measure sets, *Statist. Sinica* 17 (2007) 317–340.
- [19] R. Kulik, M. Raimondo, L^p -wavelet regression with correlated errors and inverse problems, *Statist. Sinica* 19 (2009) 1479–1489.
- [20] J.R. Wishart, Wavelet deconvolution in a periodic setting with long-range dependent errors, *J. Statist. Plann. Inference* 143 (2013) 867–881.
- [21] R. Benhaddou, R. Kulik, M. Pensky, T. Sapatinas, Multichannel deconvolution with long-range dependence: a minimax study, *J. Statist. Plann. Inference* 148 (2014) 1–19 (Invited Paper).
- [22] R. Kulik, Nonparametric deconvolution problem for dependent sequences, *Electron. J. Stat.* 2 (2008) 722–740.
- [23] Y.J. Park, S.W. Dho, H.J. Kong, Deconvolution of long-pulse lidar signals with matrix formulation, *Appl. Optim.* 36 (1997) 5158–5161.
- [24] S. Harsdorf, R. Reuter, Stable deconvolution of noisy lidar signals, in: *Proceedings of EARSeL-SIG-Workshop LIDAR, Dresden/FRG, 2000*, pp. 16–17.
- [25] E. Kolaczyk, Wavelet methods for the inversion of certain homogeneous linear operators in the presence of noisy data, Ph.D. thesis, Department of Statistics, Stanford University, 1994.
- [26] Y. Meyer, *Wavelets and Operators*, Cambridge Studies in Advanced Mathematics, vol. 37, Cambridge University Press, Cambridge, 1992, translated from the 1990 French original by D.H. Salinger.
- [27] S. Mallat, *A Wavelet Tour of Signal Processing*, Academic Press Inc., San Diego, CA, 1999.
- [28] W.M. Schmidt, *Diophantine Approximation*, Lecture Notes in Mathematics, vol. 785, Springer, Berlin, 1980.
- [29] L. Cavalier, M. Raimondo, Wavelet deconvolution with noisy eigenvalues, *IEEE Trans. Signal Process.* 55 (2007) 2414–2424.
- [30] D.L. Donoho, I.M. Johnstone, G. Kerkycharian, D. Picard, Wavelet shrinkage: asymptopia?, *J. R. Stat. Soc. Ser. B Stat. Methodol.* 57 (1995) 301–369, with discussion and a reply by the authors.
- [31] G. Kerkycharian, D. Picard, Thresholding algorithms, maxisets and well-concentrated bases, *TEST* 9 (2000) 283–344, With comments, and a rejoinder by the authors.
- [32] A. Petsa, T. Sapatinas, Minimax convergence rates under the L^p -risk in the functional deconvolution model, *Statist. Probab. Lett.* 79 (2009) 1568–1576.