Deconvolution in white noise with a random blurring function

Thomas Willer

Deconvolution in white noise with a random blurring function

Thomas Willer

Laboratoire de Probabilités et Modèles Aléatoires, Université Paris VII (Denis Diderot), 175 rue de Chevaleret, F-75013 Paris, France.

Abstract

We consider the problem of denoising a function observed after a convolution with a random filter independent of the noise and satisfying some mean smoothness condition depending on an ill posedness coefficient. We establish the minimax rates for the L^p risk over balls of periodic Besov spaces with respect to the level of noise, and we provide an adaptive estimator achieving these rates up to log factors. Simulations were performed to highlight the effects of the ill posedness and of the distribution of the filter on the efficiency of the estimator.

Keywords: Adaptive estimation; Deconvolution; Inverse problem; Minimax risk; Nonparametric estimation; Wavelet decomposition.

1 Motivations and preliminaries

1.1 Inverse problems in practice

Deconvolution is a particularly important case in a more general setting of problems, known as inverse problems. They consist in recovering an unknown object f from an observation h_n corresponding to $H(f)$ corrupted by a white noise ξ, for some operator H. The model is of the kind:

$$h_n = H(f) + \sigma n^{-1/2}\xi, \quad \forall n \geq 1. \tag{1}$$

Inverse problems appear in many scientific domains. Several applications can be found for example in OFTA [1999] in various domains such as meteorology, thermodynamics and mecanics. Deconvolution, in particular, is a common problem in signal and image processing (see Bertero and Boccacci [1998]). It appears notably in light detection and ranging devices, computing distances to an object by measuring the lapse of time between the emission of laser pulses and the detection of the pulses reflected by the object. In the underlying model f is a distance to an object measured up to small gaussian errors after being blurred by a convolution phenomenon due to the fact that the system response function of the device is longer than the time resolution interval of the detector. Several papers deal with this application of deconvolution methods, for example Harsdorf and Reuter [2000] or Johnstone et al. [2004].

In some cases, it is difficult to know *a priori* the underlying operator which transformed the object to be determined into the observed data. This problem appears notably when the operator is sensitive to even slight changes in the experimental conditions, or is affected by external random effects that cannot be controlled, and thus changes for every observation. In

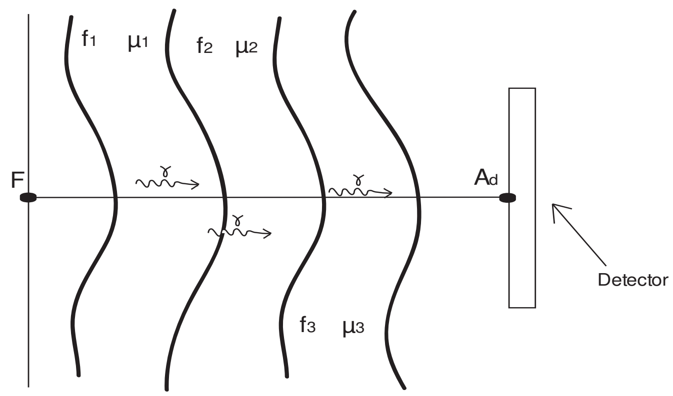

Figure 1: Reconstruction of a density of activity

these conditions, a framework with a random operator is more adapted than a setting with a fixed deterministic operator.

As an example let us consider an inverse problem of reconstruction in a tomographic imagery system, borrowed from OFTA [1999]. The problem is to find the density of activity f of a radioactive tracer by collecting the γ photons which it radiates on a detector. The framework is illustrated on Figure 1. The setting is such that only the photons transmitted perpendicularly to the detector are taken into account. A given pixel A_d of the detector collects a number of photons that depends on the density of activity f along some segment $[FA_d]$, where F is the focal point towards which A_d is headed. Each point M of this segment transmits a contribution $f(M)$ towards A_d but the pixel detects only $a(M, A_d)f(M)$ photons from M because the radiation diminishes after it has gone across the fluid between M and A_d. So the following quantity is observed on the pixel A_d:

$$X_\mu f(F, A_d) = \int_{M \in [F, A_d]} f(M)a(M, A_d)dM,$$

and the function a can be put in the following form :

$$a(M, A_d) = \exp\big[- \int_{M' \in [M, A_d]} \mu(M')dM' \big],$$

where μ is a coefficient quantifying the radiation fading around M'. On figure 1 several zones characterized by different densities of activity and different coefficients μ are represented. If μ is constant along the segment $[FA_d]$, then recovering f is a deconvolution problem.

In practice the cartography of μ is not well known *a priori*. There is a different function for each pixel and this function depends on the characteristics of the fluid where the tracers were injected. Complementary measures and reconstruction algorithms are necessary to obtain it. In this context a probabilistic model is useful, where μ is a random function determined *a posteriori* thanks to additionnal measures.

1.2 Estimation in inverse problems with random operators

In the case of deterministic operators, inverse problems have been studied in many papers in a general framework where (1) holds with some linear operator H. Two main methods of estimation are generally used to recover f from the observation: singular value decomposition (SVD) and Galerkin projection methods. The former uses a decomposition of f on a basis of eigenfunctions of $H^T H$, which can be hard to perform if H is difficult to diagonalize. The latter uses a decomposition of f on a fixed basis adapted to the kind of functions to be estimated and then consists in solving a finite linear system to recover the coefficients of f. Wavelet decomposition is a very useful tool in such settings, see Donoho [1995] and Abramovich and Silverman [1998].

Among others, a method combining wavelet-vaguelettes decompositions and Galerkin projections can be found in Cohen et al. [2002], whereas a sharp adaptive SVD estimator can be found in Cavalier and Tsybakov [2002]. Concerning the deconvolution problem, wavelet-based estimation techniques were developed in Pensky and Vidakovic [1999], Walter and Shen [1999], Fan and Koo [2002], Kalifa and Mallat [2003] and Johnstone et al. [2004]. Multidimensional situations have also been considered: minimax rates and estimation techniques can be found in Tsybakov [2001].

Generalisations of inverse problems to the case of random operators have been made in several recent papers. First, random operators enable to treat situations where, in practice, the operator modifying the object to be estimated is not exactly known because of errors of measure. In such settings, equation (1) holds with an unknown deterministic operator H, and additionnal noisy observations provide a random operator H_δ where δ is a level of noise : $H_\delta = H(f) + \delta\xi$. The problem is to build an estimator of f based on the data (h_n, H_δ) achieving minimax rates. Several adaptive estimation methods have been developed in this case. Some are based on SVD methods such as in Cavalier and Hengartner [2004], whereas estimators based on Galerkin projection methods were developed in Efromovich and Koltchinskii [2001] or Cohen et al. [2004].

Random operators also appear quite naturally in models where the evolution of a random process is influenced by its past. For example let us consider the problem of estimating an unknown function f thanks to the observation of X_n ruled by the following equation (called stochastic delay differential equation, SDDE in short):

$$\begin{aligned} dX_n(t) &= (\int_0^r X_n(t-s)f(s)ds)dt + \sigma n^{-1/2} dW(t) \quad \forall t \geq 0, \\ X_n(t) &= F(t) \quad \forall t \in [-r, 0]. \end{aligned}$$

This problem is close to problem (2): a convolution of the unknown function with the random filter X_n is observed with small errors. However this filter is not independent from W so our results do not apply to this particular problem. Numerous estimation results in SDDEs can be found in Reiss [2004] and in Reiss [2001], with a different asymptotic framework.

The organisation of the paper is as follows. Section 2, 3 and 4 present respectively the model, the estimator and the main results. Section 5 gives simulation results where the behaviour of the estimator is investigated for several distributions of the random filter, and section 6 gives the proofs of the theorems.

2 The model

We consider the following deconvolution problem. Let $(\Omega, \mathcal{A}, P)$ be a probability space and W a standard Wiener process on this space. For a given $n \in \mathbb{N}^*$ we observe the realizations of two processes X_n and Y linked in the following way:

$$\begin{cases} dX_n(t) &= f \star Y(t)dt + \sigma n^{-1/2} dW(t), \quad \forall t \in [0,1], \\ X_n(0) &= x_0, \end{cases} \tag{2}$$

where $\star$ denotes the convolution : $f \star Y(t) = \int_0^1 f(t-s)Y(s)ds$, x_0 is a deterministic initial condition and σ is a positive known constant.

The problem is to estimate the 1-periodic function f when Y is independent of W and satisfies some condition of smoothness.

2.1 The target function

We introduce functional spaces especially useful to describe the target functions. For a given $\rho > 1$, let us first denote by L^ρ the following space:

$$L^\rho([0,1]) = \{f : \mathbb{R} \mapsto \mathbb{R} \quad | \quad f \text{ is } 1-periodic, \quad \text{and} \int_0^1 |f|^\rho < \infty\}.$$

Secondly we use periodic Besov spaces which are defined thanks to the modulus of continuity in a similar way as in the non periodic case (see Johnstone et al. [2004] for the exact definition). They have the advantage of being very general, including spatially unsmooth functions, and of being very well suited to wavelet decompositions. Indeed, the following characterization holds under several conditions on the wavelet basis similar to the conditions in the general case (which can be found in Härdle et al. [1998]):

$$B^s_{p,q}([0,1]) = \{f \in L^p([0,1]) \quad | \quad \|f\|_{s,p,q} := \Big(\sum_{j\leq 0} 2^{j(s+1/2-1/p)q}\big(\sum_{0\leq k\leq 2^j} |\beta_{j,k}|^p\big)^{q/p}\Big)^{1/q} < \infty\}.$$

We investigate the maximal error when f can be any function in a ball of a periodic Besov space $B^s_{p,q}([0,1])$ of radius R and when the estimation error is measured by the L^ρ-loss. We suppose that $s > \frac{1}{p}$ so that f is continuous an hence its L^ρ-norm exists.

Definition 1. *For given $R > 0$, $p > 1$, $q > 1$ and $s > \frac{1}{p}$, define :*

$$M(s,p,q,R) = \{f \in B^s_{p,q}([0,1]) \quad | \quad \|f\|_{s,p,q} \leq R\}.$$

Our aim is to determine the rate of the following minimax risk for $\rho > 1$:

$$R_n := \inf_{\hat{f}_n} \sup_{f\in M(s,p,q,R)} E_f(\|\hat{f}_n - f\|_\rho),$$

where the infimum is taken over all $\sigma((X_n(t), Y(t))_{t\in[0,1]})-$measurable estimators $\hat{f}_n$.

2.2 The filter

We assume that the blurring function Y is a random process independent of n, f, and (in probabilistic terms) of the process W, and taking its values in $L^2([0,1])$.

Throughout this paper, we will use the following notations for two functions A and B depending on parameters p :

- $A \lesssim B$ means that there exists a positive constant C such that for all p, $\quad A(p) \leq CB(p)$,
- $A \gtrsim B$ means that $B \lesssim A$,
- $A \asymp B$ means that $A \lesssim B$ and $A \gtrsim B$.

For $j \in \mathbb{N}$ we introduce two random variables L_j^Y and U_j^Y (whenever they exist) linked to the smoothness of the process Y:

$$L_j^Y = \frac{\sum_{l=2^j}^{2^{j+1}-1} |Y_l|^2}{2^j}, \quad \text{and} \quad U_j^Y = \frac{\sum_{l=0}^{2^{j+1}-1} |Y_l|^{-2}}{2^j},$$

where $(Y_l)_{l \in \mathbb{Z}}$ are the Fourier coefficients of $(Y(t))_{t \in [0,1]}$.

To establish the lower (resp upper) bound of the minimax risk, we impose the following control on the distribution of L_j^Y (resp U_j^Y), which implies that the Fourier coefficients are not too large (resp small):

C_{low}: *There exists a constant* $\nu \geq 0$ *such that, for all* $j \in \mathbb{N}$:

$$E(L_j^Y) \lesssim 2^{-2\nu j}.$$

C_{up} : $\forall l \in \mathbb{Z}, Y_l \neq 0$ *almost surely, and there exist* $\nu \geq 0, c > 0, \alpha > 0$ *such that, for all* $j \in \mathbb{N}$:

$$\forall t \geq 0, \quad P\big(U_j^Y \geq t 2^{2\nu j}\big) \lesssim e^{-ct^\alpha}.$$

All those conditions are satisfied if the Fourier Transform $\hat{Y}$ of the process Y has the following form: $|\hat{Y}(w)| = \frac{T(w)}{(1+w^2)^{\nu/2}}$, where T is a positive random process with little probability of taking small or high values (for example bounded almost surely by deterministic constants). This case includes for example gamma probability distribution functions with some random scale parameter, which will be used further. On the contrary, condition C_{up} does not hold for filters with realizations belonging to supersmooth functions, ie Y such that $|\hat{Y}(w)| = T(w) \frac{e^{-B|w|^\beta}}{(1+w^2)^{\nu/2}}$, for some constants $B, \beta > 0$ and with T as before. Results on deconvolution of supersmooth functions can be found in Butucea [2004].

3 Adaptive estimators

We first build an adaptive estimator, nearly achieving the minimax rates exposed in the next section, which is close to the one developed in Johnstone et al. [2004] in the case of a deterministic filter Y. The method combines elements of the SVD methods (deconvolution thanks to the Fourier basis) and of the projection methods (decomposition on a wavelet basis adapted to the target functions).

Let us set $R_j = \{0, \ldots, 2^j - 1\}$ for all $j \in \mathbb{N}$, and let $(\Phi_{j,k}, \Psi_{j,k})_{j,k \in \mathbb{Z}}$ denote the periodized Meyer wavelet basis (see Meyer [1990] or Mallat [1998] for details). For convenience the following notations will be used further: $R_{-1} = \{0\}$ and $\Phi_{-1,0} = \Psi_{0,0}$. Any 1-periodic target function f belonging to $M(s,p,q,S)$ has an expansion of the kind:

$$f = \sum_{j \geq -1,\, k \in R_j} \beta_{j,k} \Psi_{j,k},$$

where

$$\beta_{j,k} = \int_0^1 f \Psi_{j,k}.$$

We estimate f by estimating its wavelet coefficients. Let $(e_l(t)) = (\exp(2\pi i l t))_{l\in\mathbb{Z}}$ denote the Fourier basis, and let $(\Psi_{j,k,l})_{l\in\mathbb{Z}}$, $(f_l)_{l\in\mathbb{Z}}$ and $(Y_l)_{l\in\mathbb{Z}}$ be the Fourier coefficients of the functions $\Psi_{j,k}$, f and Y. Set also: $W_l = \int_0^1 e_l(t)dW(t)$ and $X_l^n = \int_0^1 e_l(t)dX_n(t)$. Then by Plancherel's identity we have:

$$\beta_{j,k} = \sum_{l\in\mathbb{Z}} f_l \Psi_{j,k,l}.$$

Moreover $\int_0^1 (f \star Y)\bar{e}_l = f_l Y_l$, so equation (2) yields:

$$X_l^n = f_l Y_l + \sigma n^{-1/2} W_l,$$

and thus if we suppose that $Y_l \neq 0$ almost surely for all l, f_l can naturally be estimated by $\frac{X_l^n}{Y_l}$ and we set:

$$\hat{\beta}_{j,k} = \sum_{l\in\mathbb{Z}} \frac{X_l^n}{Y_l} \Psi_{j,k,l}.$$

Then a hard thresholding estimator is built with the following values for the thresholds λ_j and the highest resolution level j_1:

$$2^{j_1} = \{n/(\log n)^{1+\frac{1}{\alpha}}\}^{1/(1+2\nu)},$$

$$\lambda_j = \eta 2^{\nu j} \sqrt{(\log n)^{1+\frac{1}{\alpha}}/n},$$

where η is a positive constant larger than a threshold (which is determined in section 6).

Finally the following estimator achieves the minimax rates up to log factors when the filter satisfies condition C_{up}:

$$\hat{f}_n^D = \sum_{(j,k)\in\Lambda_n} \hat{\beta}_{j,k} I_{\{|\hat{\beta}_{j,k}| \geq \lambda_j\}} \Psi_{j,k}, \tag{3}$$

where $\Lambda_n = \{(j,k) \in \mathbb{Z}^2 \mid j \in \{-1, \ldots, j_1\},\ k \in R_j\}$.

Moreover we also introduce a slightly different estimator $\hat{f}_n^R$ with random thresholds instead of deterministic ones (hence the superscript R instead of D), ie with j_1 and λ_j replaced by j_2 and τ_j:

$$2^{j_2} = \{n/\log n\}^{1/(1+2\nu)},$$

$$\tau_j = \eta' \sqrt{U_j^Y \log n/n},$$

where η' is a large enough constant. The theoretical performances of $\hat{f}_n^R$ will be studied in a separate publication, here only a simulation study is provided.

4 Main results

Let $\rho > 1$, $R > 0$, $p > 1$, $q > 1$ and $s > 1/p$. We distinguish three cases for the regularity parameters characterizing the target functions according to the sign of $\epsilon = \frac{2s+2\nu+1}{\rho} - \frac{2\nu+1}{p}$:

the sparse case ($\epsilon < 0$), the critical case ($\epsilon = 0$) and the regular case ($\epsilon > 0$).

Let us introduce the two following rates:

$$r_n(s,\nu) = \left(\frac{1}{n}\right)^{\frac{s}{2s+2\nu+1}}, \quad s_n(s,p,\rho,\nu) = \left(\frac{\log(n)}{n}\right)^{\frac{s-1/p+1/\rho}{2s+2\nu+1-2/p}}.$$

Theorem 1. *Under condition C_{low} on Y:*

$$r_n(s,\nu)^{-1}R_n \gtrsim 1 \quad \textit{in the regular case,}$$
$$s_n(s,p,\rho,\nu)^{-1}R_n \gtrsim 1 \quad \textit{in the sparse and critical cases.}$$

Theorem 2. *Under condition C_{up} on Y:*

$$r_n(s,\nu)^{-1}R_n \lesssim 1 \quad \textit{in the regular case,}$$
$$s_n(s,p,\rho,\nu)^{-1}R_n \lesssim 1 \quad \textit{in the sparse case,}$$
$$s_n(s,p,\rho,\nu)^{-1}R_n \lesssim \log(n)^{(1-\frac{p}{\rho q})_+} \quad \textit{in the critical case.}$$

Theorem 3. *Under condition C_{up} on Y, for estimator $\hat{f}_n^D$ defined in* (3) *and if $q \leq p$ in the critical case:*

$$\sup_{f\in M(s,p,q,R)} E_f(\|\hat{f}_n^D - f\|_\rho) \lesssim \left(\frac{\log(n)^{1+\frac{1}{\alpha}}}{n}\right)^{\frac{s}{2s+2\nu+1}} \quad \textit{in the regular case,}$$

$$\sup_{f\in M(s,p,q,R)} E_f(\|\hat{f}_n^D - f\|_\rho) \lesssim \left(\frac{\log(n)^{1+\frac{1}{\alpha}}}{n}\right)^{\frac{s-1/p+1/\rho}{2s+2\nu+1-2/p}} \quad \textit{in the critical and sparse cases.}$$

When the filter satisfies C_{low} and C_{up} the rates of Theorems 1 and 2 match except in the critical case when $\rho > \frac{p}{q}$, where the upper bound contains an extra logarithmic factor. This is also observed in density estimation or regression problems (see Donoho et al. [1996] and Donoho et al. [1997]), and that factor is probably part of the actual rate of R_n: the lower bound is maybe too optimistic.

Analysing the effect of ν, we remark that the rates are similar to the ones established in the white noise model or other classical non-parametric estimation problems (examples can be found in Tsybakov [2004]), except that here an additional effect reflected by ν slows the minimax speed. Indeed the convolution blurs the observations, making the estimation all the more difficult as ν is large. This parameter is called ill-posedness coefficient, explanations about this notion can be found in Nussbaum and Pereverzev [1999] for example.

Concerning Theorem 3, we remark that estimator $\hat{f}_n^D$ is not optimal first by a log factor in the regular case, which is a common phenomenon for adaptive estimators as was highlighted in Tsybakov [2000], and secondly by log factors with exponents proportional to $\frac{1}{\alpha}$. This is due to the difficulty to control the deviation probability of the estimated wavelet coefficients when the probability of having small eigenvalues Y_l of the convolution operator is high (ie when α is small).

The main interest of these results is that bounds of the minimax risk are established in a random operator setting, for a wide scale of L^ρ losses, and over general functional spaces which include unsmooth functions. As far as we know, the lower bound has not been established in deconvolution problems for such settings even in the case of deterministic filters.

Let us also note that condition C_{up} imposed on the filter Y is similar to the conditions generally used in other inverse problems where the singular values of the operator are required to decrease polynomially fast. Moreover condition C_{up} concern means of eigenvalues over diadic blocs, which enables to include filters for which Fourier coefficients vary erratically individually, but not in mean, such as some boxcar filters (see Kerkyacharian et al. [2004]). The case of severely ill-posed inverse problems, where the singular values decrease exponentially fast, has also been studied in Cavalier et al. [2003] for example.

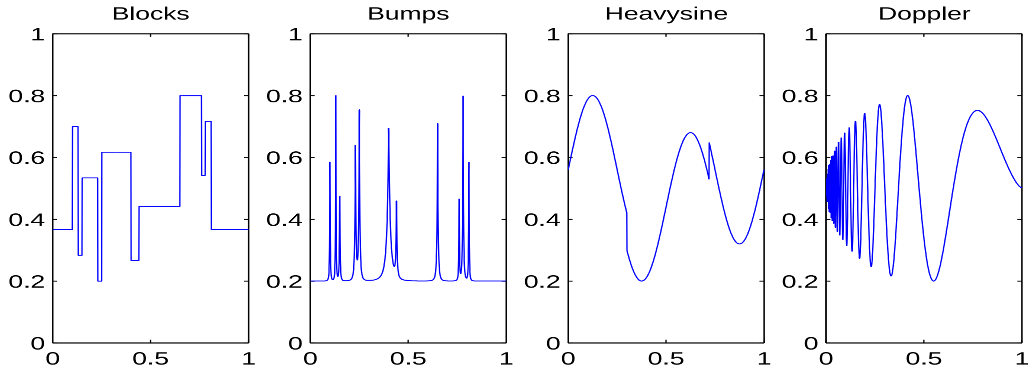

Figure 2: Target functions

5 Simulations

To illustrate the rates obtained for the upper bound, the behaviours of estimators $\hat{f}_n^D$ and $\hat{f}_n^R$ are examined in practice for the following settings. We consider the four target functions (Blocks, Bumps, Heavisine, Doppler) represented on figure 2, which were used by Donoho and Johnstone in a series of papers (Donoho and Johnstone [1994] for example). These functions are blurred by convolution with realizations of a random filter Y and by adding gaussian noise with root signal to noise ratio ($rsnr$) of three levels: $rsnr \in \{3, 5, 7\}$. Then the two estimators are computed in each case and their performances are examined, judging by the mean square error (MSE). For the simulation of the data and the implementation of the estimators, parts of the WaveD software package written by Donoho and Raimondo for Johnstone et al. [2004] were used.

5.1 Distribution of the filter

A simple way to represent the blurring effect is the convolution with a boxcar filter, ie at time t one observes the mean of the unknown function on an interval $[t-a, t]$ with a random width a. However these kinds of filters have various degrees of ill posedness depending on a. For some numbers called "badly approximable" numbers, this degree is constant and equal to 3/2. For other numbers the situation is more complicated, and the set of the badly approximable numbers has a Lebesgue measure equal to zero (more explanations can be found in Johnstone and Raimondo [2004] or Johnstone et al. [2004]). However new results have been found recently for almost all boxcar widths in Kerkyacharian et al. [2004] where the near optimal properties of several thresholding estimators are established.

So as to keep a fixed ill posedness coefficient boxcar filters are excluded, and one considers convolutions with periodized gamma functions with parameters ν and λ:

$$Y(t) = \frac{1}{\int_0^{+\infty} s^{\nu-1} e^{-\lambda s} ds} \sum_{l \in \mathbb{N}} (t+l)^{\nu-1} e^{-\lambda(t+l)},$$

where ν is a fixed shape parameter and λ is a *random* scale parameter with a probability distribution function F_α parametrized by some $\alpha > 0$:

$$F_\alpha(t) = \min\left(1, 2e^{-\frac{C_\alpha}{t^{2\alpha}}} \mathbb{I}(t \geq 0)\right),$$

where the constant C_α is set such that $E(\lambda) = 150$ for all α.

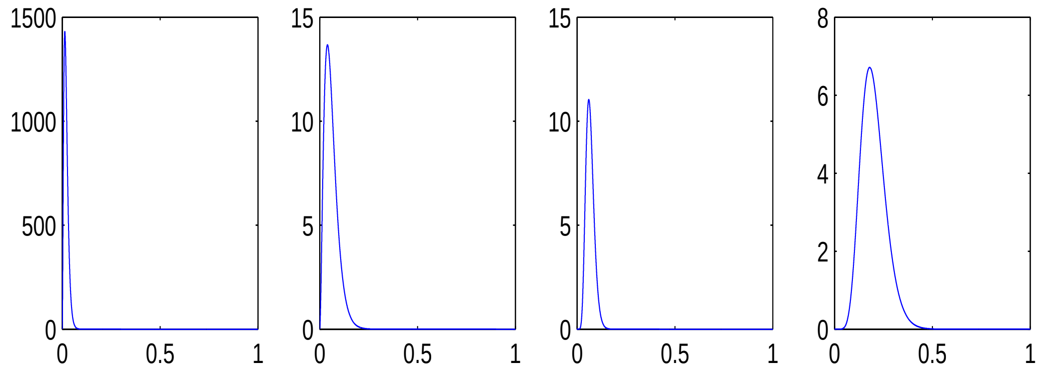

Figure 3: Examples of filters, from left to right: $(\nu, \lambda) \in \{(3, 150), (3, 50), (10, 150), (10, 50)\}$

Such a filter Y satisfies conditions C_{up} and C_{low}. Some examples of its shapes are given in figure 3: ν and λ can be interpreted respectively as a delay and a spreading parameter. According to the minimax rates, f should be (asymptotically) more difficult to estimate for large ν and for small α. This is checked in practice in the next section.

5.2 Results

First we focus on the effect of ν conditionnally to the filter Y. An example in medium noise for the Blocks target is given in figure 4, where the filter is kept constant with $\lambda = 150$: as expected, both estimators get less and less efficient when ν increases. Moreover in practice the thresholds of estimator $\hat{f}_n^D$ need to be rescaled for each ν, contrarily to those of estimator $\hat{f}_n^R$ which is thus more convenient. The same results were obtained for the other target functions and by examining the MSE of the estimators, the figures were not included for the sake of conciseness.

Next we set $\nu = 1$ and we investigate the effect of the distribution of the filter Y. Both estimators perform well for mean and high realizations of λ, but difficulties appear for small realizations which are all the more frequent as α is small: the worst case among 10 simulations is represented in figure 5 when $\alpha = 2$ and in figure 6 when $\alpha = 0.5$, and the two estimators perform more poorly in the last case. However they remain better in that case than a fixed threshold estimator (ie with thresholds completely independent of the filter) also represented in the figures.

More generally the MSE were computed for several values of α and for the three noise levels. The results are given in figure 7: the shape of the distribution of Y clearly affects estimator $\hat{f}_n^D$, and also $\hat{f}_n^R$ to a much lesser extent. The smaller α, the poorer they behave. Especially the Doppler and Bumps targets are not well estimated by $\hat{f}_n^D$ for small α, mainly because the high thresholds make it ignore many of the numerous details of these targets.

Finally estimator $\hat{f}_n^R$ proves more convenient than estimator $\hat{f}_n^D$ when the ill-posedness varies, and also less sensitive to the weight of the probability of small eigenvalues.

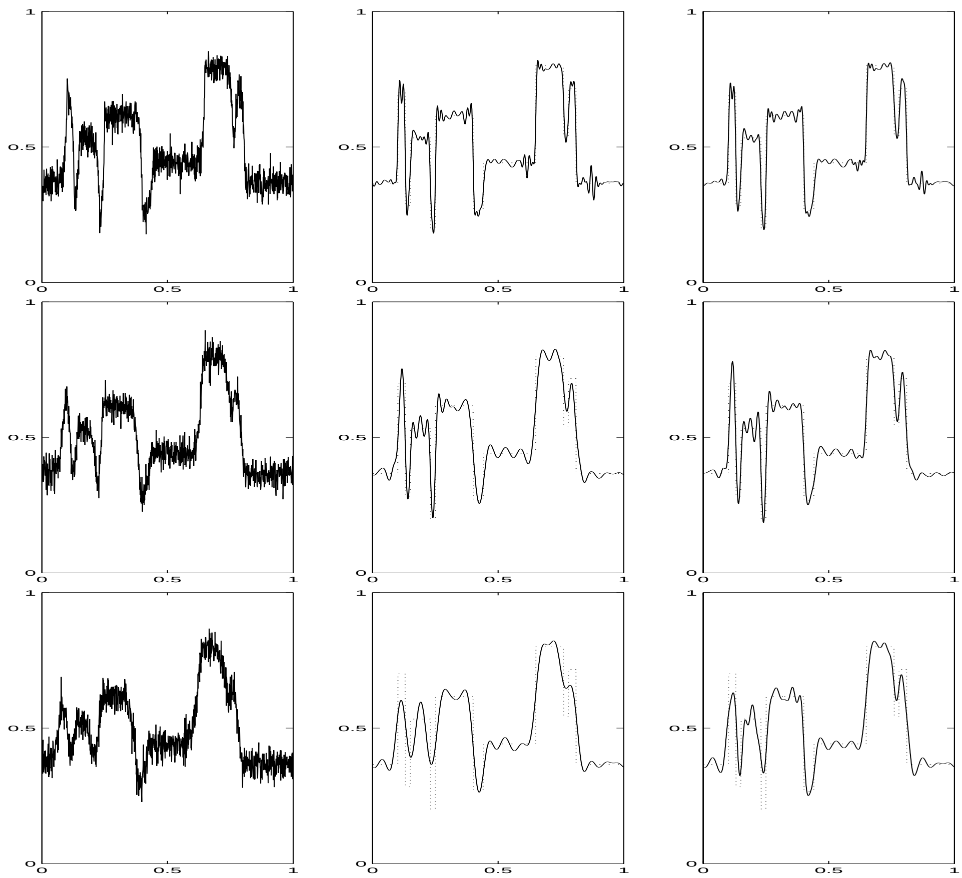

Figure 4: Data, estimator $\hat{f}_n^R$ and estimator $\hat{f}_n^D$ (left to right) for fixed $\lambda = 150$ and $\nu = 1$ (top), $\nu = 3$ (middle) and $\nu = 5$ (bottom)

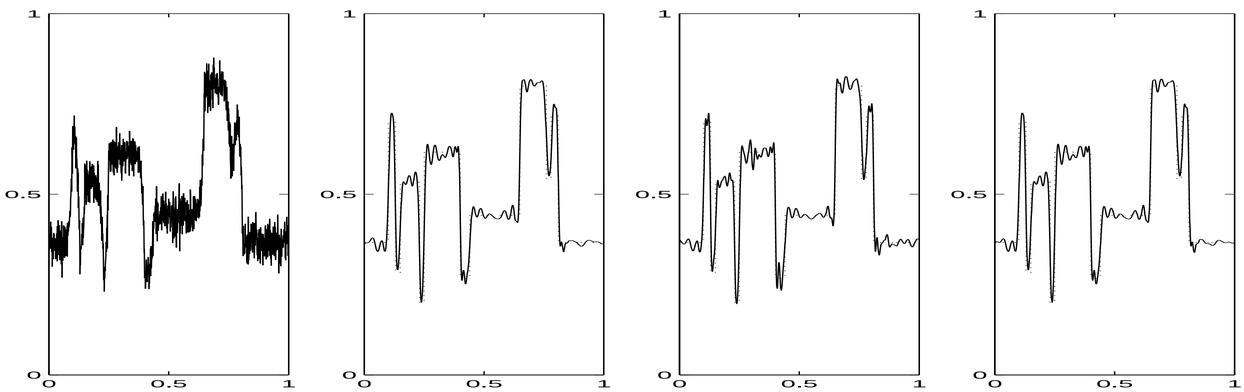

Figure 5: Data, estimator $\hat{f}_n^R$, estimator $\hat{f}_n^D$ and a fixed-threshold estimator (left to right) for $\alpha = 2$

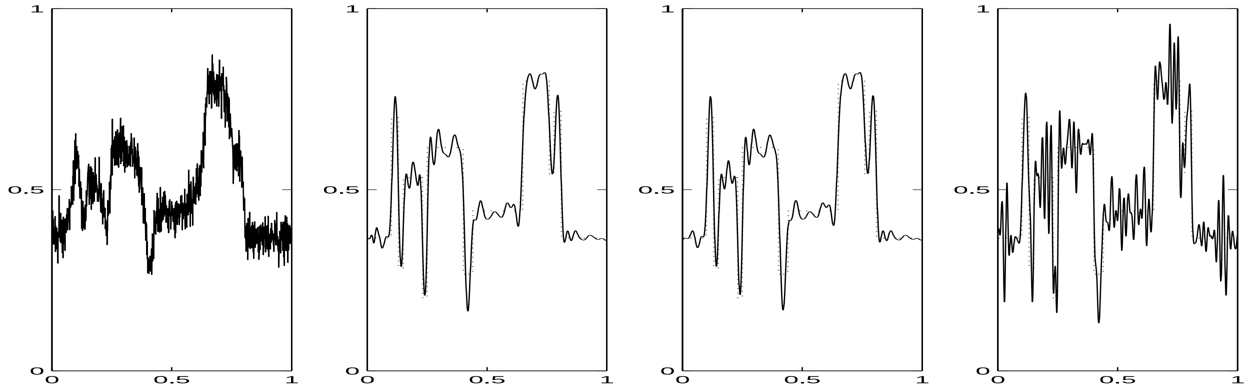

Figure 6: Data, estimator $\hat{f}_n^R$, estimator $\hat{f}_n^D$ and a fixed-threshold estimator (left to right) for $\alpha = 0.5$

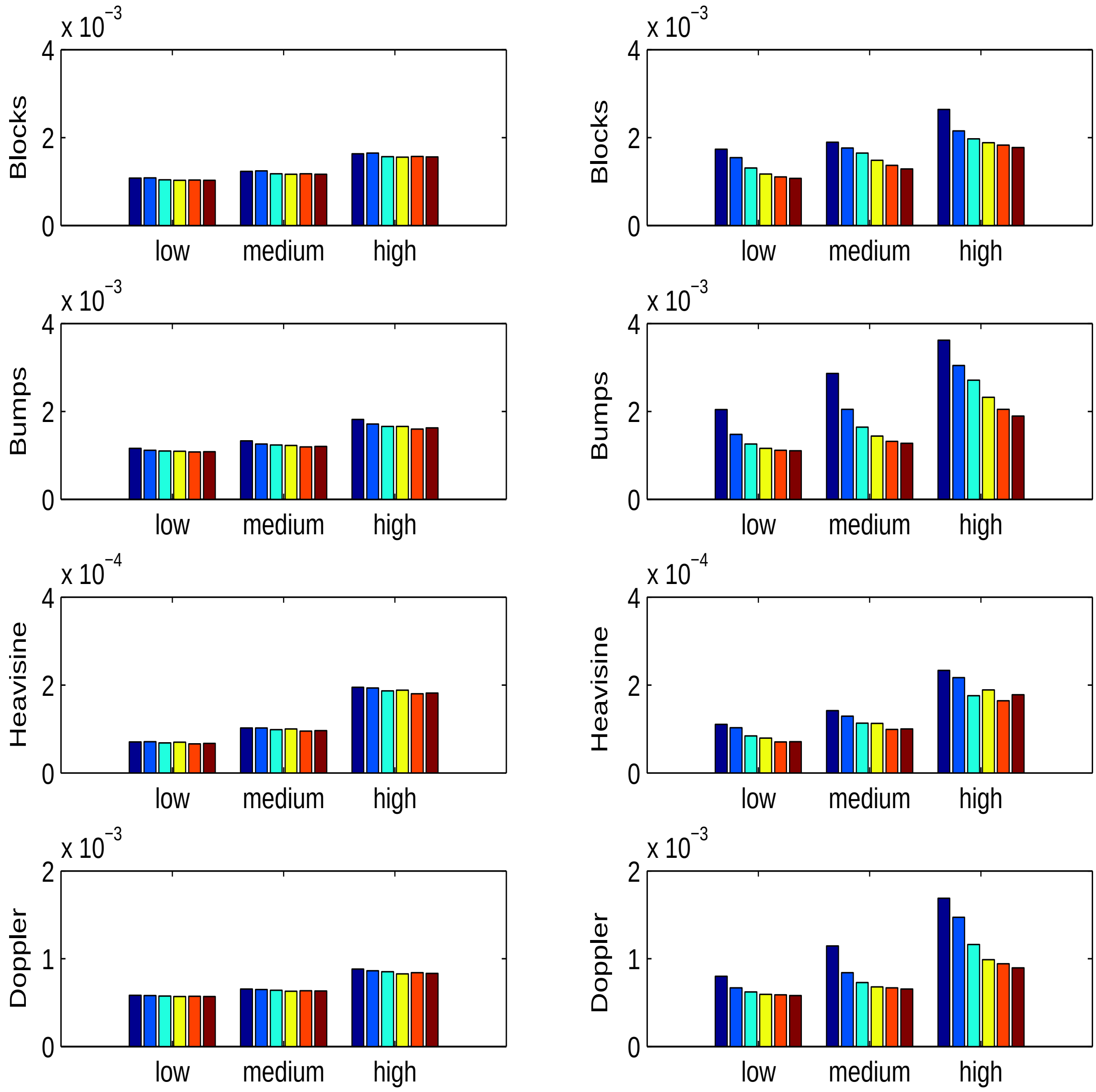

Figure 7: Effect of α on the MSE of estimator $\hat{f}_n^R$ (left) and estimator $\hat{f}_n^D$ (right) for each target, each level of noise and $\alpha \in \{0.5, 0.6, 0.7, 0.8, 0.9, 1\}$ (left to right in each group)

6 Proofs of the lower and upper bounds

6.1 Lower bound

6.1.1 Sparse case

We use a classical lemma on lower bounds (Korostelev and Tsybakov Korostelev and Tsybakov [1993]):

Lemma 1. *Let V a functionnal space, $d(.,.)$ a distance on V,*
for f, g belonging to V denote by $\Lambda_n(f,g)$ the likelihood ratio : $\Lambda_n(f,g) = \frac{dP_{X_n^{(f)}}}{dP_{X_n^{(g)}}}$ where $dP_{X_n^{(h)}}$ is the probability distribution of the process X_n if h is true.
If V contains functions $f_0, f_1, \ldots, f_K$ such that :

- $d(f_{k'}, f_k) \geq \delta > 0$ *for* $k \neq k'$,
- $K \geq \exp(\lambda_n)$ *for some* $\lambda_n > 0$,
- $\Lambda_n(f_0, f_k) = \exp(z_n^k - v_n^k)$, *where* z_n^k *is a random variable such that there exists* $\pi_0 > 0$ *with* $P(z_n^k > 0) \geq \pi_0$, *and* v_n^k *are constants,*
- $\sup_k v_n^k \leq \lambda_n$.

Then

$$\sup_{f\in V} P_{X_n^{(f)}}\left(d(\hat{f}_n, f) \geq \delta/2\right) \geq \pi_0/2,$$

for an arbitrary estimator $\hat{f}_n$.

To use this result, we build a finite set of functions belonging to $M(s,p,q,R)$ as follows. Let $(\psi_{j,k})_{j\geq -1, k\in\mathbb{Z}}$ be an $s-$regular Meyer wavelet basis, which we periodize according to:

$$\Psi_{j,k}(x) = \sum_{l\in\mathbb{Z}} \psi_{j,k}(x+l).$$

In the sequel we denote by $(\Psi_{j,k})_{(j,k)\in\Lambda}$ the periodized Meyer wavelet basis obtained this way, where $\Lambda = \{(j,k) \mid j \geq -1;\ k \in R_j\}$ and $R_j = \{0, \ldots, 2^j - 1\}$.

Now for a fixed level of resolution j set for any $k \in R_j$:

$$f_{j,k} = \gamma \Psi_{j,k},$$

with $\gamma \lesssim 2^{-j(s+1/2-1/p)}$ such that $\|f_{j,k}\|_{s,p,q} \leq R$. Set also $f_0 = 0$.

Let us choose for d the distance $d(f,g) = \|f-g\|_\rho$. Because of the relation between the L^ρ norm of a linear combination of wavelets of fixed resolution j and the l^ρ norm of the corresponding coefficients (see Meyer [1990]), we have for any $k, k' \in R_j$, $k \neq k'$:

$$d(f_{j,k'}, f_{j,k}) = \|\gamma\Psi_{j,k'} - \gamma\Psi_{j,k}\|_{L^\rho} \asymp \gamma 2^{j(1/2-1/\rho)}.$$

In this framework we have : $K = 2^j$ and $\delta \asymp \gamma 2^{j(1/2-1/\rho)}$. So as to apply the lemma, we have to find parameters $\gamma(n)$ and $j(n)$ such that the other hypotheses of the lemma are satisfied, which will be true if :

$$P_{f_{j,k}}\Big(\ln(\Lambda_n(f_0, f_{j,k})) \geq -j(n)\ln(2)\Big) \geq \pi_0 > 0,$$

uniformly for all $f_{j,k}$. Moreover we have :

$$P_{f_{j,k}}\Big(\ln(\Lambda_n(f_0,f_{j,k}))\geq -j(n)\ln(2)\Big)\geq 1-P_{f_{j,k}}\Big(|\ln(\Lambda_n(f_0,f_{j,k}))|>j(n)\ln(2)\Big)$$
$$\geq 1-E_{f_{j,k}}\Big(|ln(\Lambda_n(f_0,f_{j,k})|\Big)/(j(n)\ln(2)).$$

So the previous condition is satisfied when $\gamma(n)$ and $j(n)$ are chosen such that, with a constant $0<c<1$:

$$E_{f_{j,k}}\Big(|ln(\Lambda_n(f_0,f_{j,k}))|\Big)\leq cj(n)\ln(2). \tag{4}$$

Consider two hypotheses f_0 and $f_{j,k}$, and let us determine the likelihood ratio of the corresponding distributions of the observations $(X_n(t),Y(t))_{t\in[0,1]}$. Let F be a bounded measurable function. Since Y is assumed to be independent of W and free with respect to f in (2), we have:

$$E_{f_{j,k}}\big[F\big(X^n,Y\big)\big]=E\big[E\{F\big(\big(\int_0^t f_{j,k}\star Y(s)ds+\sigma n^{-1/2}W(t),Y(t)\big)_{t\in[0,1]}\big)\quad|Y\}\big]$$
$$=\int E\{F\big(\sigma n^{-1/2}\tilde{W},y\big)\}dP_Y(y),$$

where P_Y denotes the distribution of Y and $\tilde{W}(t)=W(t)+\int_0^t\sigma^{-1}n^{1/2}f_{j,k}\star y(s)ds$.

For a given function y let $h^y_{j,k}$ be defined by: $h^y_{j,k}(t)=\sigma^{-1}n^{1/2}f_{j,k}\star y(t)$. We assumed that Y takes its values in $L^2([0,1])$ so for each of its realization there exists a constant C_y such that for all $t\in[0,1]$, $\int_0^t(h^y_{j,k})^2(s)ds<C_y$ and we can apply the formula of Girsanov: the process $\tilde{W}$ is a Wiener process under the probability Q defined by

$$dQ=\exp\big[-\int_0^1 h^y_{j,k}(t)dW(t)-\frac{1}{2}\int_0^1(h^y_{j,k}(t))^2dt\big]dP.$$

Thus for any function y:

$$E_P\big[F\big(\sigma n^{-1/2}\tilde{W},y\big)\big]=E_Q\big[F\big(\sigma n^{-1/2}\tilde{W},y\big)\exp\big[\int_0^1 h^y_{j,k}(t)dW(t)+\frac{1}{2}\int_0^1(h^y_{j,k}(t))^2dt\big]\big]$$
$$=E_Q\big[F\big(\sigma n^{-1/2}\tilde{W},y\big)\exp\big[\int_0^1 h^y_{j,k}(t)d\tilde{W}(t)-\frac{1}{2}\int_0^1(h^y_{j,k}(t))^2dt\big]\big]$$
$$=E_P\big[F\big(\sigma n^{-1/2}W,y\big)\exp\big[\int_0^1 h^y_{j,k}(t)dW(t)-\frac{1}{2}\int_0^1(h^y_{j,k}(t))^2dt\big]\big].$$

So finally:

$$\Lambda_n(f_0,f_{j,k})=\exp\big[-\int_0^1\frac{f_{j,k}\star Y(t)}{\sigma n^{-1/2}}dW(t)+\frac{1}{2}\int_0^1\big(\frac{f_{j,k}\star Y(t)}{\sigma n^{-1/2}}\big)^2dt\big].$$

We can now examine under which conditions (4) is true. We have:

$$E|\ln(\Lambda_n(f_0,f_{j,k}))|=E|\frac{\gamma n^{1/2}}{\sigma}\int_0^1\Psi_{j,k}\star Y(t)dW(t)-\frac{\gamma^2 n}{2\sigma^2}\int_0^1(\Psi_{j,k}\star Y(t))^2dt|\leq A_n+B_n,\ \text{with:}$$

$$B_n=\frac{\gamma^2 n}{2\sigma^2}E\big(\int_0^1(\Psi_{j,k}\star Y(t))^2dt\big),$$

$$A_n=\frac{\gamma n^{1/2}}{\sigma}E|\int_0^1\Psi_{j,k}\star Y(t)dW(t)|\leq\frac{\gamma n^{1/2}}{\sigma}\big(E\big(\int_0^1\Psi_{j,k}\star Y(t)dW(t)\big)^2\big)^{1/2}\leq(2B_n)^{1/2},$$

where we used Jensen's inequality for A_n.

Let us find a bound for B_n. We introduce the Fourier coefficients of Y and $\Psi_{j,k}$ denoted by Y_l and $\Psi_{j,k,l}$ for all $l \in \mathbb{Z}$. Since the Fourier Transform of $\Psi_{j,k}$ is bounded by $2^{-j/2}$ we have:

$$B_n = \frac{\gamma^2 n}{2\sigma^2} E_{f_{j,k}}\big(\frac{1}{2\pi}\sum_{l\in\mathbb{Z}} |Y_l \Psi_{j,k,l}|^2\big) \lesssim \gamma^2 n 2^{-j} E_{f_{j,k}}\big(\sum_{l\in C_j} |Y_l|^2\big),$$

where C_j is the set of integers where the coefficients $\Psi_{j,k,l}$ are not equal to zero (it can easily be shown that this set does not depend on k).

The support of the Fourier transform of the Meyer wavelet is included in $[-\frac{2\pi}{3}, -\frac{8\pi}{3}] \cup [\frac{2\pi}{3}, \frac{8\pi}{3}]$. So $\Psi_{j,k,l} = 0$ as soon as $|2\pi 2^{-j} l| \in [\frac{2\pi}{3}, \frac{8\pi}{3}]^c$, and $C_j \subset [-2^{j+1}, -2^{j-2}] \cup [2^{j-2}, 2^{j+1}]$ for all j. Then under condition C_{low} and noticing that $Y_{-l} = Y_l$ we obtain:

$$B_n \lesssim \gamma^2 n 2^{-2\nu j}.$$

Finally, condition (4) holds if we choose γ and j such that:

$$\gamma^2 n 2^{-2\nu j} \lesssim j, \quad \text{and} \quad \gamma \lesssim 2^{-j(s+1/2-1/p)}.$$

We choose the following values that satisfy those two conditions:

$$\gamma \asymp 2^{-j(s+1/2-1/p)}, \quad \text{and} \quad 2^j \asymp (n/log(n))^{1/(2s+2\nu+1-2/p)}.$$

Finally, using the lemma and the inequality of Markov, for $\sigma((X_n(t), Y(t)), t \in [0,1])-$measurable estimators $\hat{f}_n$ the following bound holds:

$$\inf_{\hat{f}_n} \sup_{f\in M(s,p,q,S)} E_f(\|\hat{f}_n - f\|_\rho) \gtrsim \gamma 2^{j(1/2-1/\rho)} \asymp \big(\frac{\log(n)}{n}\big)^{\frac{s-1/p+1/\rho}{2s+2\nu+1-2/p}}.$$

6.1.2 Regular case

Here we consider another set of functions belonging to $M(s,p,q,R)$. We use the periodized Meyer wavelet basis $(\Psi_{j,k})$ like before. But now we set for any $\epsilon \in \{-1,+1\}^{R_j}$:

$$f_{j,\epsilon} = \gamma \sum_{k\in R_j} \epsilon_k \Psi_{j,k},$$

with $\gamma \lesssim 2^{-j(s+1/2)}$ such that $\|f_{j,\epsilon}\|_{s,p,q} \leq S$. We also set $I_{j,k} = [\frac{k}{2^j}, \frac{k+1}{2^j}]$.

We use an adaptation of lemma 10.2 in Härdle et al. [1998] to the case of Meyer wavelets (that do not have compact supports) and of the norm $\|.\|_\rho$:

Lemma 2. *Suppose the likelihood ratio satisfies for some constant λ:*

$$P_{f_{j,\epsilon}}\big(\Lambda_n(f_{j,\epsilon^k}, f_{j,\epsilon}) \geq e^{-\lambda}\big) \geq p_* > 0,$$

uniformly for all $f_{j,\epsilon}$ and all $k \in R_j$, where ϵ^k is equal to ϵ except for the k^{th} element which is multiplied by -1. Then the following bound holds:

$$\max_{\epsilon\in\{-1,+1\}^{R_j}} E_{f_{j,\epsilon}}(\|\hat{f}_n - f_{j,\epsilon}\|_\rho) \geq C 2^{j/2} \gamma e^{-\lambda} p_*,$$

where C is positive and depends only on ρ.

Similarly to the sparse case, the hypothesis of this lemma is satisfied if, for a small enough constant c:

$$E_{f_{j,\epsilon}} |\ln\big(\Lambda_n(f_{j,\epsilon^k}, f_{j,\epsilon})\big)| \leq c.$$

Now the log-likelihood is equal to:

$$\ln\big(\Lambda_n(f_{j,\epsilon^k}, f_{j,\epsilon})\big) = \frac{2\gamma n^{1/2}}{\sigma} \int_0^1 \Psi_{j,k} \star Y(t) dW(t) - \frac{2\gamma^2 n}{\sigma^2} \int_0^1 [\Psi_{j,k} \star Y(t)]^2 dt.$$

Like before, we only need to dominate the following quantity:

$$B_n = \gamma^2 n E_{f_{j,\epsilon}}\Big(\int_0^1 (\Psi_{j,k} \star Y(t))^2 dt\Big).$$

We use the same bound as in the sparse case, under assumption C_{low}. The parameters have to be chosen such that:

$$\gamma^2 n 2^{-2\nu j} \lesssim 1 \quad \text{and} \quad \gamma \lesssim 2^{-j(s+1/2)}.$$

Finally the regular rate is obtained for the following choices:

$$\gamma \asymp 2^{-j(s+1/2)}, \quad \text{and} \quad 2^j \asymp n^{1/(2s+2\nu+1)}.$$

Proof. of the lemma

The Meyer wavelet satisfies $\exists A > 0$ such that $|\psi(x)| \leq \frac{A}{1+|x|^2}$. Consequently:

$$\begin{aligned}
\Big(\int_{I_{j,k}} |\Psi_{j,k}(x) dx|^\rho\Big)^{1/\rho} &= 2^{j(\frac{1}{2}-\frac{1}{\rho})}\Big(\int_0^1 |\sum_{l\in\mathbb{Z}} \psi(x+2^j l)|^\rho dx\Big)^{1/\rho} \\
&\geq 2^{j(\frac{1}{2}-\frac{1}{\rho})}\Big(\int_0^1 |\psi(x)|^\rho dx - \sum_{l\in\mathbb{Z}^*} \int_0^1 |\psi(x+2^j l)|^\rho dx\Big)^{1/\rho} \\
&\geq 2^{j(\frac{1}{2}-\frac{1}{\rho})}\Big(\int_0^1 |\psi(x)|^\rho dx - \frac{A^\rho}{2^{2\rho j}} \sum_{l\in\mathbb{N}^*} \frac{1}{(l/2)^{2\rho}}\Big)^{1/\rho} \\
&\geq c 2^{j(\frac{1}{2}-\frac{1}{\rho})},
\end{aligned}$$

for j large enough and $c > 0$ depends only on ρ.

Then using a concavity inequality and similar arguments as in the compact support case, we have:

$$\max_\epsilon E_{f_{j,\epsilon}}(\|\hat{f}_n - f_{j,\epsilon}\|_\rho) \geq 2^{-2^j} \sum_\epsilon E_{f_{j,\epsilon}}[\sum_{k=0}^{2^j-1} \int_{I_{j,k}} |\hat{f}_n - f_{j,\epsilon}|^\rho]^{\frac{1}{\rho}}$$

$$\geq 2^{-2^j + j(\frac{1}{\rho}-1)} \sum_\epsilon \sum_{k=0}^{2^j-1} E_{f_{j,\epsilon}}[\int_{I_{j,k}} |\hat{f}_n - f_{j,\epsilon}|^\rho]^{\frac{1}{\rho}}$$

$$\geq 2^{-2^j + j(\frac{1}{\rho}-1)} \sum_{k=0}^{2^j-1} \sum_{\epsilon|\epsilon_k=1} E_{f_{j,\epsilon}}[(\int_{I_{j,k}} |\hat{f}_n - f_{j,\epsilon}|^\rho)^{\frac{1}{\rho}} + \Lambda_n(f_{j,\epsilon^k}, f_{j,\epsilon})(\int_{I_{j,k}} |\hat{f}_n - f_{j,\epsilon^k}|^\rho)^{\frac{1}{\rho}}]$$

$$\geq 2^{-2^j + j(\frac{1}{\rho}-1)} \sum_{k=0}^{2^j-1} \sum_{\epsilon|\epsilon_k=1} E_{f_{j,\epsilon}}[\delta I\{\int_{I_{j,k}} |\hat{f}_n - f_{j,\epsilon}|^\rho \geq \delta^\rho\} + \Lambda_n(f_{j,\epsilon^k}, f_{j,\epsilon})\delta I\{\int_{I_{j,k}} |\hat{f}_n - f_{j,\epsilon^k}|^\rho \geq \delta^\rho\}]$$

with $\delta = c\gamma 2^{j(\frac{1}{2}-\frac{1}{\rho})}$.

Noticing that

$$\Big(\int_{I_{j,k}} |\hat{f}_n - f_{j,\epsilon}|^\rho\Big)^{1/\rho} + \Big(\int_{I_{j,k}} |\hat{f}_n - f_{j,\epsilon^k}|^\rho\Big)^{1/\rho} \geq 2\gamma\Big(\int_{I_{j,k}} |\Psi_{j,k}(x)|^\rho\Big)^{1/\rho} \geq 2\gamma c 2^{j(\frac{1}{2}-\frac{1}{\rho})}$$

for j large enough, the end of the proof follows as in Härdle et al. [1998]. □

6.2 Upper bounds

6.2.1 Properties of the estimated wavelet coefficients

The performances of the thresholding estimators rest on the properties of the estimated wavelet coefficients $\hat{\beta}_{j,k}$. In the sequel we will also need properties for the estimators $\hat{\alpha}_{j,k}$ defined the same way as $\hat{\beta}_{j,k}$ in estimator (3) except with Φ instead of Ψ. We have the following results:

Proposition 1. *Under condition C_{up} we have for all $j \geq -1$, $k \in R_j$ and $r > 0$,*

$$E(|\hat{\beta}_{j,k} - \beta_{j,k}|^r) \lesssim \Big(\frac{2^{\nu j}}{\sqrt{n}}\Big)^r \quad \textit{and} \quad E(|\hat{\alpha}_{j,k} - \alpha_{j,k}|^r) \lesssim \Big(\frac{2^{\nu j}}{\sqrt{n}}\Big)^r,$$

and there exist positive constants κ, and κ' such that for all $\lambda \geq 1$,

$$P\Big(|\hat{\beta}_{j,k} - \beta_{j,k}| \geq \frac{2^{\nu j}}{\sqrt{n}}\lambda\Big) \lesssim 2^{-\kappa\lambda^{\frac{2\alpha}{\alpha+1}}} \quad \textit{and} \quad P\Big(|\hat{\beta}_{j,k} - \beta_{j,k}| \geq \sqrt{\frac{U_j^Y}{n}}\lambda\Big) \lesssim 2^{-\kappa'\lambda^2},$$

where the constants in the inequalities do not depend on j, k and λ.

Proof. of Proposition 1

Remark that conditionally to the process Y, $(\hat{\beta}_{j,k} - \beta_{j,k})$ is a centered gaussian variable with variance:

$$Var(|\hat{\beta}_{j,k} - \beta_{j,k}| \quad |Y) = E\Big[\frac{\sigma^2}{n}\sum_{l\in\mathbb{Z}} \Big|\frac{W_l}{Y_l}\Psi_{j,k,l}\Big|^2 \quad |Y\Big].$$

Since the Fourier transform of the Meyer wavelet is bounded by $2^{-j/2}$ and only $l \in [-(2^{j+1}-1), -2^{j-2}] \cup [2^{j-2}, 2^{j+1}-1]$ has to be considered, we have for some constant $C > 0$:

$$Var(|\hat{\beta}_{j,k} - \beta_{j,k}| \quad |Y) \leq C U_j^Y/n.$$

Thus the moment of order r of $(\hat{\beta}_{j,k} - \beta_{j,k})$ is bounded by

$$E(|\hat{\beta}_{j,k} - \beta_{j,k}|^r) \lesssim E[(Var(|\hat{\beta}_{j,k} - \beta_{j,k}| \quad |Y))^{r/2}] \lesssim E[(U_j^Y/n)^{r/2}],$$

and by similar arguments the same bound holds for $(\hat{\alpha}_{j,k} - \alpha_{j,k})$ because the support of the Fourier Transform of $\phi_{j,k}$ is $\frac{4\pi}{3}[-2^j, 2^j]$.

For the deviation probability we use a probabilistic inequality for a centered standard gaussian variable Z. Conditionally to Y we have:

$$P(|\hat{\beta}_{j,k} - \beta_{j,k}| > \frac{2^{\nu j}}{\sqrt{n}}\lambda \quad |Y) \leq P(|Z| \geq \lambda\sqrt{2^{2\nu j}/(CU_j^Y)} \quad |Y)$$
$$\lesssim \frac{1}{\lambda\sqrt{2^{2\nu j}/(CU_j^Y)}} \exp(-\frac{\lambda^2 2^{2\nu j}}{2CU_j^Y}).$$

Then we take the expectation over Y, by Cauchy Schwartz we obtain for $\lambda \geq 1$:

$$P(|\hat{\beta}_{j,k} - \beta_{j,k}| > \frac{2^{\nu j}}{\sqrt{n}}\lambda) \lesssim \sqrt{E(\frac{U_j^Y}{2^{2\nu j}})E(\exp(-\frac{\lambda^2 2^{2\nu j}}{CU_j^Y}))}.$$

The end of the proof is directly deducible from the lemma below, and the last part of Proposition 1 is easily proved by replacing $2^{\nu j}$ by $\sqrt{U_j^Y}$ in the three inequalities above.

Lemma 3. *Let X_j be the following random variable: $X_j = \frac{U_j^Y}{2^{2\nu j}}$. For all $j \geq 0$ there exists positive constants C', C'', $C(.)$ such that for all $r > 0$:*

$$E(e^{-\frac{r}{X_j}}) \leq C' e^{-C'' r^{\frac{\alpha}{\alpha+1}}}, \quad \textit{and} \quad E(X_j^r) \leq C(r).$$

Proof. of the lemma

For all $r > 0$ we have:

$$\begin{aligned} E(e^{-\frac{r}{X_j}}) &= \int_0^1 P(e^{-\frac{r}{X_j}} \geq u)du \\ &= r\int_0^{+\infty} P(X_j \geq 1/u)e^{-ru}du \\ &\leq r\int_0^1 P(X_j \geq 1/u)e^{-ru}du + e^{-r} \\ &\lesssim r\int_0^1 e^{-ru-c/u^\alpha}du + e^{-r}, \end{aligned}$$

and one can check that there exists $C'' > 0$ such that $\int_0^1 e^{-ru-c/u^\alpha}du \lesssim e^{-C'' r^{\frac{\alpha}{\alpha+1}}}$.

The second part of the lemma is easily proved by using similar arguments.

□

□

6.2.2 Proof of the sharp rates

In the regular and critical zones, estimator (3) is not optimal up to a logarithmic factor. In order to show that the rates of Theorem 1 are sharp, we exhibit estimators achieving the rates of Theorem 2. Those are not as interesting in practice as (3), since they depend on caracteristics of f, ie they are not adaptive.

We will use the following bound to estimate the risks, which holds for any $-1 \leq j_m \leq j_M \leq \infty$ and any set of random or deterministic coefficients $\tilde{\beta}_{j,k}$ such that the quantities below are finite:

$$E\| \sum_{j_m \leq j \leq j_M} \sum_{k \in R_j} \tilde{\beta}_{j,k} \Psi_{j,k} \|_\rho \lesssim \sum_{j_m \leq j \leq j_M} 2^{j(\frac{1}{2}-\frac{1}{\rho})} (\sum_{k \in R_j} E|\tilde{\beta}_{j,k}|^\rho)^{\frac{1}{\rho}}. \tag{5}$$

The proof is immediate by Minkowski inequality, the fact that $\| \sum_{k \in R_j} \tilde{\beta}_{j,k} \Psi_{j,k} \|_\rho \asymp 2^{j(\frac{1}{2}-\frac{1}{\rho})} \|\tilde{\beta}_{j,.}\|_{l_\rho}$ (established in Meyer [1990]) and a concavity argument.

Let us denote: $\nu' = \nu + 1/2$ and $\epsilon = ps - \nu'(\rho - p)$. We distinguish two cases: $\rho \leq p$ and $p < \rho$. In the first case $M(s,p,q,R)$ is included in the regular zone. By concavity we have:

$$\inf_{\hat{f}_n} \sup_{f \in M(s,p,q,R)} E_f \|\hat{f}_n - f\|_\rho \leq \inf_{\hat{f}_n} \sup_{f \in M(s,p,q,R)} E_f \|\hat{f}_n - f\|_p.$$

So seeing the expected rate only the case $\rho = p$ needs to be considered. We take the following linear estimator:

$$\hat{f}_n = \sum_{k \in R_{j_1}} \hat{\alpha}_{j_1,k} \Phi_{j_1,k}.$$

For any $f \in M(s,p,q,R)$ the risk is composed of a bias error and a stochastic error:

$$E_f \|\hat{f}_n - f\|_p \leq A_s + A_s,$$

with:

$$A_s = E\| \sum_{k \in R_{j_1}} (\hat{\alpha}_{j_1,k} - \alpha_{j_1,k}) \Phi_{j_1,k} \|_p \lesssim 2^{j_1(\frac{1}{2}-\frac{1}{p})} [\sum_{k \in R_{j_1}} E|\hat{\alpha}_{j_1,k} - \alpha_{j_1,k}|^p]^{\frac{1}{p}} \lesssim (\frac{2^{\nu j_1}}{\sqrt{n}}) 2^{\frac{j_1}{2}} = \frac{2^{\nu' j_1}}{\sqrt{n}},$$

$$A_b = \| \sum_{j > j_1} \sum_{k \in R_j} \beta_{j,k} \Psi_{j,k} \|_p \lesssim \sum_{j > j_1} 2^{j(\frac{1}{2}-\frac{1}{p})} (\sum_{k \in R_j} |\beta_{j,k}|^p)^{\frac{1}{p}} \lesssim \sum_{j > j_1} 2^{j(\frac{1}{2}-\frac{1}{p})} 2^{-j(s+\frac{1}{2}-\frac{1}{p})} \lesssim 2^{-j_1 s},$$

and we obtain the rate by choosing $j_1 = [\frac{log_2(n)}{2s+2\nu'}]$.

In the second case ($p < \rho$) we consider the following estimator:

$$\hat{f}_n = \sum_{k \in R_{j_1+1}} \hat{\alpha}_{j_1+1,k} \Phi_{j_1+1,k} + \sum_{j_1 < j < j_2} \sum_{k \in R_j} \hat{\beta}_{j,k} I_{\{|\hat{\beta}_{j,k}| \geq \lambda_j\}} \Psi_{j,k},$$

where:

$$2^{j_1} \approx n^{\frac{1}{2s+2\nu'}}, \quad 2^{j_2} \approx (\frac{n}{(\log n)^{I\{\epsilon < 0\}}})^{\frac{s}{(2s+\nu')(s-\frac{1}{p}+\frac{1}{\rho})}}, \quad \lambda_j = \eta \sqrt{U_j^Y (j - j_1)/n},$$

and $\eta > 2(\frac{2\rho\nu'}{\kappa'})^{\frac{1}{2}}$, so that we have by Proposition 1: $P(|\hat{\beta}_{j,k} - \beta_{j,k}| \geq \lambda_j) \lesssim 2^{-\kappa' \eta^2 (j - j_1)}$.

We proceed as in Donoho et al. [1996] by distinguishing six terms:

$$\begin{aligned}\hat{f}_n - f &= \sum_{k\in R_j}(\hat{\alpha}_{j_1,k}-\alpha_{j_1,k})\Phi_{j,k} + \sum_{j\geq j_2}\sum_{k\in R_j}\beta_{j,k}\Psi_{j,k}\\ &+ \sum_{j_1<j<j_2}\sum_{k\in R_j}(\hat{\beta}_{j,k}-\beta_{j,k})\Psi_{j,k}[I_{\{|\hat{\beta}_{j,k}|\geq\lambda_j,|\beta_{j,k}|<\lambda_j/2\}} + I_{\{|\hat{\beta}_{j,k}|\geq\lambda_j,|\beta_{j,k}|\geq\lambda_j/2\}}]\\ &+ \sum_{j_1<j<j_2}\sum_{k\in R_j}\beta_{j,k}\Psi_{j,k}[I_{\{|\hat{\beta}_{j,k}|<\lambda_j,|\beta_{j,k}|\geq 2\lambda_j\}} + I_{\{|\hat{\beta}_{j,k}|<\lambda_j,|\beta_{j,k}|<2\lambda_j\}}]\\ &= e_s + e_b + e_{bs} + e_{bb} + e_{sb} + e_{ss}.\end{aligned}$$

Like before the stochastic error is bounded by:

$$E(\|e_s\|_\rho) \lesssim \frac{2^{\nu' j_1}}{\sqrt{n}},$$

and by using Sobolev embeddings it is easy to see that:

$$E(\|e_b\|_\rho) \lesssim 2^{-j_2(s-\frac{1}{p}+\frac{1}{\rho})}.$$

The terms e_{bs} and e_{sb} can be grouped together because of the two following assertions: $\{|\hat{\beta}_{j,k}|<\lambda_j, |\beta_{j,k}|\geq 2\lambda_j\} \cup \{|\hat{\beta}_{j,k}|\geq\lambda_j, |\beta_{j,k}|<\lambda_j/2\} \subset \{|\hat{\beta}_{j,k}-\beta_{j,k}|>\lambda_j/2\}$, and $[|\hat{\beta}_{j,k}|<\lambda_j, |\beta_{j,k}|\geq 2\lambda_j] \Rightarrow [|\beta_{j,k}|\leq 2|\hat{\beta}_{j,k}-\beta_{j,k}|]$. Consequently:

$$\begin{aligned}E(\|e_{bs}\|_\rho + \|e_{sb}\|_\rho) &\lesssim \sum_{j_1<j<j_2} 2^{j(\frac{1}{2}-\frac{1}{\rho})}(E\sum_{k\in R_j}|\hat{\beta}_{j,k}-\beta_{j,k}|^\rho I_{\{|\hat{\beta}_{j,k}-\beta_{j,k}|>\lambda_j/2\}})^{\frac{1}{\rho}}\\ &\leq \sum_{j_1<j<j_2} 2^{j(\frac{1}{2}-\frac{1}{\rho})}(\sum_{k\in R_j}(E|\hat{\beta}_{j,k}-\beta_{j,k}|^{2\rho})^{\frac{1}{2}}(P\{|\hat{\beta}_{j,k}-\beta_{j,k}|>\lambda_j/2\})^{\frac{1}{2}})^{\frac{1}{\rho}}\\ &\lesssim \sum_{j_1<j<j_2} 2^{j(\frac{1}{2}-\frac{1}{\rho})}(\sum_{k\in R_j}\frac{2^{\rho\nu j}}{n^{\frac{\rho}{2}}}2^{-\frac{\kappa'(\eta/2)^2(j-j_1)}{2}})^{\frac{1}{\rho}}\\ &\leq \frac{2^{\nu' j_1}}{n^{\frac{1}{2}}}\sum_{0<j<j_2-j_1} 2^{(\nu'-\frac{\kappa'(\eta/2)^2}{2\rho})j}\\ &\lesssim \frac{2^{\nu' j_1}}{n^{\frac{1}{2}}},\end{aligned}$$

where we used Cauchy Schwartz inequality and Proposition 1.

For e_{bb} we use the characterization of Besov spaces:

$$\begin{aligned}E(\|e_{bb}\|_\rho) &\lesssim \sum_{j_1<j<j_2} 2^{j(\frac{1}{2}-\frac{1}{\rho})}(\sum_{k\in R_j}E|\hat{\beta}_{j,k}-\beta_{j,k}|^\rho I_{\{|\beta_{j,k}|\geq\lambda_j/2\}})^{\frac{1}{\rho}}\\ &\lesssim \sum_{j_1<j<j_2} 2^{j(\frac{1}{2}-\frac{1}{\rho})}(\sum_{k\in R_j}\frac{2^{\rho\nu j}}{n^{\frac{\rho}{2}}}(\frac{|\beta_{j,k}|}{\lambda_j/2})^p)^{\frac{1}{\rho}}\\ &\lesssim \sum_{j_1<j<j_2}(\frac{2^{j(\frac{\rho}{2}-1+(\rho-p)\nu)}}{n^{\frac{\rho-p}{2}}(j-j_1)^{\frac{p}{2}}}2^{-pj(s+\frac{1}{2}-\frac{1}{p})}(\|f\|^s_{p,\infty})^p)^{\frac{1}{\rho}}\\ &\lesssim \frac{1}{n^{\frac{\rho-p}{2\rho}}}\sum_{j_1<j<j_2}(\frac{2^{-\epsilon j}}{(j-j_1)^{\frac{p}{2}}})^{\frac{1}{\rho}}.\end{aligned}$$

Lastly for e_{ss} we remark that $|\beta_{j,k}|^\rho \le (2\lambda_j)^{\rho-p}|\beta_{j,k}|^p$ and we use again the characterization of Besov spaces:

$$\begin{aligned} E(\|e_{ss}\|_\rho) &\lesssim \sum_{j_1<j<j_2} 2^{j(\frac{1}{2}-\frac{1}{\rho})}\big((2\lambda_j)^{\rho-p}\sum_{k\in R_j}|\beta_{j,k}|^p\big)^{\frac{1}{\rho}} \\ &\lesssim \sum_{j_1<j<j_2}\big(\frac{2^{j(-ps+\nu'(\rho-p))}}{n^{\frac{\rho-p}{2}}}(j-j_1)^{\frac{\rho-p}{2}}(\|f\|^s_{p,\infty})^p\big)^{\frac{1}{\rho}} \\ &\lesssim \frac{1}{n^{\frac{\rho-p}{2\rho}}}\sum_{j_1<j<j_2}\big(2^{-\epsilon j}(j-j_1)^{\frac{\rho-p}{2}}\big)^{\frac{1}{\rho}} \end{aligned}$$

According to these bounds e_{bs}, e_{sb} and e_s are of the same order and e_{ss} dominates e_{bb}, so we choose j_1 and j_2 so as to balance the bounds of e_b, e_s and e_{ss}.
In the regular zone we have:

$$E(\|e_{ss}\|_\rho) \lesssim \big(\frac{2^{-\epsilon j_1}}{n^{\frac{\rho-p}{2}}}\big)^{\frac{1}{\rho}},$$

and in the sparse zone:

$$E(\|e_{ss}\|_\rho) \lesssim \big(\frac{j_2 2^{-\epsilon j_2}}{n^{\frac{\rho-p}{2}}}\big)^{\frac{1}{\rho}}.$$

Thus with the announced choices of j_1 and j_2 we get the prescribed rates in both zones.

Lastly in the critical zone we change the majoration of $(\beta_{j,k})$ in e_{bb} and e_{ss} by using:

$$\begin{aligned} \sum_{j_1<j<j_2}\big(2^{pj(s+\frac{1}{2}-\frac{1}{p})}\sum_{k\in R_j}|\beta_{j,k}|^p\big)^{\frac{1}{\rho}} &\lesssim (j_2-j_1)^{1-\frac{p}{\rho q}}(\|f\|^s_{p,q})^{\frac{p}{\rho}} \quad \text{if} \quad \frac{p}{\rho}<q, \\ &\lesssim (\|f\|^s_{p,q})^q \quad \text{if} \quad \frac{p}{\rho}\ge q. \end{aligned}$$

Here again e_{ss} is dominant and of the order: $E(\|e_{ss}\|_\rho) \lesssim (\frac{j_2}{n})^{\frac{\rho-p}{2\rho}} j_2^{(1-\frac{p}{\rho q})_+}$, hence the extra logarithmic factor.

6.2.3 Proof of the rates of the adaptive estimator

To prove Theorem 3 we use a theorem for thresholding algorithms established by Kerkyacharian and Picard (Theorem 3.1 in Kerkyacharian and Picard [2000]) which holds in a very general setting where one wants to estimate an unknown function f thanks to observations in a sequence of statistical models $(E_n)_{n\in\mathbb{N}}$. It uses the Temlyakov inequalities, let us first recall this notion.

Definition 2. *Let e_n be a basis in L^ρ. It satisfies the Temlyakov property if there are absolute constants c and C such that for all $\Lambda \in \mathbb{N}$:*

$$c\sum_{n\in\Lambda}\int|e_n(x)|^\rho dx \le \int\{\sum_{n\in\Lambda}\int|e_n(x)|^2\}^{\rho/2}dx \le C\sum_{n\in\Lambda}\int|e_n(x)|^\rho dx.$$

Now let $(\psi_{j,k})_{j,k}$ denote a periodized wavelet basis and let $\rho>1$ and $0<r<\rho$. Assume that there exist a positive value $\delta>0$, a positive sequence $(\sigma_j)_{j\ge -1}$, a positive sequence c_n tending to 0, and a subset Λ_n of $\mathbb{N}^2$ such that :

$$|\Lambda_n| \sim c_n^{-\delta} \text{ where } |S| \text{ denotes the cardinal of the set } S, \tag{6}$$

$$(\sigma_j \psi_{j,k})_{j,k} \text{ satisfies the Temlyakov property,} \tag{7}$$

$$\sup_n [\mu\{\Lambda_n\} c_n^\rho] < \infty, \tag{8}$$

where μ is the following measure on $\mathbb{N}^2$:

$$\mu(j,k) = \|\sigma_j \psi_{j,k}\|_\rho^\rho = 2^{j(\rho/2-1)} \sigma_j^\rho \|\psi\|_\rho^\rho.$$

Assume also that we have a statistical procedure yielding estimators $\hat{\beta}_{j,k}$ of the wavelet coefficients $\beta_{j,k}$ of f in the basis $(\psi_{j,k})_{j,k}$ and a positive value $\eta > 0$ such that for all $(j,k) \in \Lambda_n$:

$$E(|\hat{\beta}_{j,k} - \beta_{j,k}|^{2\rho}) \leq C(c_n \sigma_j)^{2\rho}, \tag{9}$$

$$P(|\hat{\beta}_{j,k} - \beta_{j,k}| \geq \eta \sigma_j c_n/2) \leq C \min(c_n^{2\rho}, c_n^4). \tag{10}$$

Finally let $l_{r,\infty}(\mu)$ and $A(c_n^{\rho-r})$ be the following spaces and let $\hat{f}_n$ be the following estimator:

$$l_{r,\infty}(\mu) = \{f, \quad \sup_{\lambda>0} [\lambda^q \mu\{(j,k)/|\beta_{j,k}| > \sigma_j \lambda\}] < \infty\},$$
$$A(c_n^{\rho-r}) = \{f, \quad c_n^{-(\rho-r)} \|f - \sum_{\kappa \in \Lambda_n} \beta_\kappa \psi_\kappa\|_\rho^\rho < \infty\},$$
$$\hat{f}_n = \sum_{j,k \in \Lambda_n} \hat{\beta}_{j,k} I_{\{|\hat{\beta}_{j,k}| \geq \eta \sigma_j c_n\}} \psi_{j,k}.$$

Theorem 4. *Using the objects defined above and under the hypotheses* (6) *to* (10)*, we have the following equivalence:*

$$E\|\hat{f}_n - f\|_\rho^\rho \lesssim c_n^{\rho-r} \quad \Longleftrightarrow \quad f \in l_{r,\infty}(\mu) \cap A(c_n^{\rho-r}).$$

We adapt this to estimator $\hat{f}_n^D$ by setting, for given $\rho > 1$, $p > 1$, $s > 1/p$ and $q > 1$:

$$c_n = \sqrt{\frac{\log(n)^{\frac{\alpha+1}{\alpha}}}{n}}, \quad \sigma_j = 2^{\nu j}, \quad 2^{j_1} \approx \{\frac{n}{\log(n)^{\frac{\alpha+1}{\alpha}}}\}^{\frac{1}{1+2\nu}}, \quad \Lambda_n = \{(j,k) \mid -1 \leq j \leq j_1,\ k \in R_j\}.$$

With these choices we have:

$$|\Lambda_n| \asymp 2^{j_1} \asymp c_n^{-2/(1+2\nu)},$$

$$\mu(\Lambda_n) = \sum_{j=0}^{j_1-1} 2^j 2^{j(\rho/2-1)} 2^{\rho\nu j} \asymp 2^{j_1 \rho(\nu+1/2)}.$$

Consequently (8) and (6) hold with $\delta = 2/(1+2\nu)$. Condition (7) is also satisfied, the proof can be found in Johnstone et al. [2004]. Moreover thanks to Proposition 1, it is easy to establish that the estimators $\hat{\beta}_{j,k}$ used by (3) satisfy (9) and (10) as soon as $\eta > 2(\frac{max(2,\rho)}{\kappa})^{\frac{\alpha+1}{2\alpha}}$.

Then we prove Theorem 3 by setting r such that the right hand side of the inequality in the first point of the theorem corresponds to the rates in the sparse and in the regular case, ie:

$$r = \rho - 2\rho \frac{s - 1/p + 1/\rho}{2s + 2\nu + 1 - 2/p},$$

or

$$r = \rho - 2\rho\frac{s}{2s + 2\nu + 1},$$

and by showing that the space over which the risk is maximized is included in the maxiset, if we add the condition $q \le p$ in the critical case $\frac{2s+2\nu+1}{\rho} = \frac{2\nu+1}{p}$:

$$M(s, p, q, R) \subset l_{r,\infty}(\mu) \cap A(c_n^{\rho-r}).$$

The inclusion $M(s, p, q, R) \subset A(c_n^{\rho-r})$ is established in Johnstone et al. [2004], and the following proof of $M(s, p, q, R) \subset l_{r,\infty}(\mu)$ uses the same arguments as Kerkyacharian et al. [2004] for the boxcar blur. We have:

$$\begin{aligned}
\mu\{(j, k) : |\beta_{j,k}| > 2^{\nu j}\lambda\} &= \sum_{j\ge 0,\, k\in R_j} 2^{j(\rho(\nu+1/2)-1)} I\{|\beta_{j,k}| > 2^{\nu j}\lambda\} \\
&\le \sum_j (2^{j\rho(\nu+1/2)}) \wedge (2^{j(\rho(\nu+1/2)-1)} \sum_k (|\beta_{j,k}|/(2^{\nu j}\lambda))^p \\
&\le \sum_j (2^{j\rho(\nu+1/2)}) \wedge (\frac{2^{-j(sp+\nu' p-\nu'\rho)}}{\lambda^p}\epsilon_j^p),
\end{aligned}$$

where $\nu' = \nu + 1/2$ and $\epsilon_j \in l_q$. We cut the sum at J such that $2^J \asymp \lambda^{-r/(\nu'\rho)}$.

In the regular case we have:

$$\mu\{(j, k) : |\beta_{j,k}| > 2^{\nu j}\lambda\} \le \lambda^{-r} + \frac{\lambda^{(sp-\nu'(\rho-p))\frac{r}{\nu'\rho}}}{\lambda^p},$$

and the power of λ in the second term is also exactly $-r$.

In the critical case we obtain, since $q \le p$:

$$\mu\{(j, k) : |\beta_{j,k}| > 2^{\nu j}\lambda\} \le \lambda^{-r} + \frac{\sum_j \epsilon_j^p}{\lambda^p} \lesssim \lambda^{-r} + \frac{\sum_j \epsilon_j^q}{\lambda^p} \lesssim \lambda^{-r} + \lambda^{-p},$$

and $r = p$ in this case.

Lastly in the sparse case (where $r \ge p$ is satisfied) we use the Sobolev embedding $B^s_{p,q} \subset B^{s'}_{r,q}$ with $s' = s - 1/p + 1/r$. We proceed as before by cutting the sum at J such that $2^J \asymp \lambda^{-r/(\nu'\rho)}$ and noticing that $s'r + \nu' r - \nu'\rho = 0$. There exists $\tilde\epsilon_j \in l_r$ such that:

$$\begin{aligned}
\mu\{(j, k) : |\beta_{j,k}| > 2^{\nu j}\lambda\} &\le \sum_j (2^{j\rho\nu'}) \wedge (2^{j(\rho\nu'-1)} \sum_k (|\beta_{j,k}|/(2^{\nu j}\lambda))^r \\
&\le \sum_j (2^{j\rho\nu'}) \wedge (\frac{\tilde\epsilon_j^r}{\lambda^r}) \\
&\lesssim \lambda^{-r}.
\end{aligned}$$

Thus $\mu\{(j, k) : 2^{\nu j}\lambda\} \lesssim 1/\lambda^r$ for both values of r, and finally using the equivalence in Theorem 4 and Jensen inequality we obtain the prescribed rates for $E\|\hat f_n^D - f\|_\rho$.

Acknowledgments

I would like to thank Professor Dominique Picard for introducing me to the subject and for her numerous and helpful comments.

References

F. Abramovich and B. Silverman. Wavelet decomposition approaches to statistical inverse problems. *Biometrika*, 85(1):115–129, 1998.

M. Bertero and P. Boccacci. Introduction to inverse problems in imaging. *Institute of Physics, Bristol and Philadelphia*, 1998.

C. Butucea. Deconvolution of supersmooth densities with smooth noise. *Canad. J. Statist.*, 32 (2):181–192, 2004.

L. Cavalier, Y. Golubev, O. Lepski, and A. Tsybakov. Block thresholding and sharp adaptive estimation in severely ill-posed inverse problems. *Theory of Probability and its Applications*, 48(3):534–556, 2003.

L. Cavalier and N. W. Hengartner. Adaptive estimation for inverse problems with noisy operators. *Manuscript*, 2004.

L. Cavalier and A. Tsybakov. Sharp adaptation for inverse problems with random noise. *Probab. Theory Related Fields*, 123(3):323–354, 2002.

A. Cohen, M. Hoffmann, and M. Reiss. Adaptive wavelet galerkin methods for linear inverse problems. *Preprint LPMA*, dec 2002.

A. Cohen, M. Hoffmann, and M. Reiss. On adaptive estimation in linear inverse problems with error in the operator. *Manuscript*, 2004.

D. Donoho. Nonlinear solution of linear inverse problems by wavelet-vaguelette decomposition. *Applied Computational and Harmonic Analysis*, 2:101–126, 1995.

D. Donoho and I. Johnstone. Ideal spatial adaptation by wavelet shrinkage. *Biometrika*, 81: 425–455, 1994.

D. Donoho, I. Johnstone, G. Kerkyacharian, and D. Picard. Density estimation by wavelet thresholding. *Ann. Statist.*, 2:508–539, 1996.

D. Donoho, I. Johnstone, G. Kerkyacharian, and D. Picard. Universal near minimaxity of wavelet shrinkage. *Festschrift for Lucien Le Cam*, pages 183–218, 1997.

S. Efromovich and V. Koltchinskii. On inverse problems with unknown operators. *IEEE*, 47(7): 2876–2894, 2001.

J. Fan and J. Koo. Wavelet deconvolution. *IEEE Transactions on Information Theory*, 48(3): 734–747, 2002.

W. Härdle, G. Kerkyacharian, D. Picard, and A. Tsybakov. *Wavelets, Approximation and Statistical Applications*. Springer-Verlag, 1998.

S. Harsdorf and R. Reuter. Stable deconvolution of noisy lidar signals. *Submitted to Earsel meeting*, 2000.

I. M. Johnstone, G. Kerkyacharian, D. Picard, and M. Raimondo. Wavelet deconvolution in a periodic setting. *Journal of the Royal Statistical Society*, 66(3):1–27, 2004.

I. M. Johnstone and M. Raimondo. Periodic boxcar deconvolution and diophantine approximation. *Annals of Statistics*, 32(5):1781–1804, 2004.

J. Kalifa and S. Mallat. Thresholding estimators for linear inverse problems and deconvolutions. *Annals of Statistics*, 31:58–109, 2003.

G. Kerkyacharian and D. Picard. Thresholding algorithms, maxisets and well concentrated bases. *Test*, 9(2), 2000.

G. Kerkyacharian, D. Picard, and M. Raimondo. Adaptive boxcar deconvolution on full lebesgue measure sets. *Preprint LPMA*, 2004.

V. Korostelev and A. Tsybakov. *Minimax theory of image reconstruction*. Springer-Verlag, 1993.

S. Mallat. *A wavelet tour of signal processing (2nd edition)*. Academic Press Inc., San Diego, CA, 1998.

Y. Meyer. *Ondelettes et Opérateurs-I*. Hermann, 1990.

M. Nussbaum and S. Pereverzev. The degree of ill-posedness in stochastic and deterministic noise models. *Preprint WIAS*, 1999.

OFTA. *Problèmes inverses : de l'expérimentation à la modélisation*. Observatoire Francais des Techniques Avancées, 1999.

M. Pensky and B. Vidakovic. Adaptive wavelet estimator for nonparametric density deconvolution. *Annals of Statistics*, 27:2033–2053, 1999.

M. Reiss. Nonparametric estimation for stochastic delay differential equations. *Ph.D. thesis, Humboldt Universität zu Berlin*, 2001.

M. Reiss. Adaptive estimation for affine stochastic delay differential equations. *Submitted to Bernoulli*, 2004.

A. Tsybakov. On the best rate of adaptive estimation in some inverse problems. *C.R. Acad. Sci.*, 1(330):835–840, 2000.

A. Tsybakov. Sharp adaptive estimation of linear functionals. *Ann. Statist.*, 29(6):1567–1600, 2001.

A. Tsybakov. *Introduction à l'estimation Non-paramétrique*. Springer, 2004.

G. Walter and X. Shen. Deconvolution using the meyer wavelet. *Journal of Integral Equations and Applications*, 11:515–534, 1999.