Wavelet deconvolution in a periodic setting*

Iain M. Johnstone, Gérard Kerkyacharian,
Stanford University, Université de Paris X-Nanterre
Dominique Picard and Marc Raimondo
Universités de Paris VI-VII and University of Sydney

25 February 2003

Abstract

In this paper, we present an inverse estimation procedure which combines Fourier analysis with wavelet expansion. In the periodic setting, our method can recover a blurred function observed in white noise. The blurring process is achieved through a convolution operator which can either be smooth (polynomial decay of the Fourier transform) or irregular (such as the convolution with a box-car). The proposal is non-linear and does not require any prior knowledge of the smoothness class; it enjoys fast computation and is spatially adaptive. This contrasts with more traditional filtering methods which demand a certain amount of regularisation and often fail to recover non-homogeneous functions. A fine tuning of our method is derived via asymptotic minimax theory which reveals some key differences with the direct case of Donoho et al. (1995): (a) band-limited wavelet families have nice theoretical and computing features; (b) the high frequency cut off depends on the spectral characteristics of the convolution kernel; (c) thresholds are level dependent in a geometric fashion. We tested our method using simulated *lidar* data for underwater remote sensing. Both visual and numerical results show an improvement over existing methods. Finally, the theory behind our estimation paradigm gives a complete characterisation of the 'Maxiset' of the method i.e. the set of functions where the method attains a near-optimal rate of convergence for a variety of L^p loss functions.

1 De-convolution in white noise

Suppose we observe the random process $Y_n(.)$

$$Y_n(dt) = f \star g(t)dt + n^{-1/2}W(dt), \quad t = (t_1, \ldots, t_n) \in T = [0,1], \tag{1}$$

where $W(.)$ is a Gaussian white noise and

$$f \star g(t) = \int_T f(t-u)g(u)du\,. \tag{2}$$

Our goal is to recover the unknown function f from the noisy-blurred observations (1). The function g which drives the shape of the blurring through the convolution (2) is assumed to be known. Further, we assume that the functions f and g are periodic on the unit interval T and that g has a certain degree of smoothness. Our assumptions, illustrated Figure 2, are the following:

*Key Words and Phrases. Adaptive estimation, deconvolution, non-parametric regression, Meyer wavelet.
Short Title: Wavelet deconvolution.
AMS 2000 Subject Classification: primary 62G05; secondary 62G08

(A1) *Ordinary smooth convolution*; the Fourier coefficients of g decay in a polynomial fashion i.e. there exists some positive constants C and ν such that $g_l \sim C|l|^{-\nu}$.

(A2) *Irregular convolution (box-car)*; $g(x) = \frac{1}{2a}\mathbb{I}_{[-a,a]}(x)$, where a is a Badly Approximable (BA) irrational number (those includes quadratic irrationals like $\sqrt{5}$ see section 6.3). Under assumption (A2) we will put as a convention (which will be justified later) $\nu = 3/2$.

Our aim in this paper is to present a non-linear wavelet estimation method which, up to log-factors, is minimax optimal for a wide class of target functions in various L^p-metrics. For example, measuring the performance with respect to the mean integrated square error, our proposal can recover the unknown function f with an accuracy of order :

$$\Big(\frac{\log n}{n}\Big)^{\alpha}, \quad \text{where} \quad \alpha = \frac{2s}{s+\nu+\frac{1}{2}}. \tag{3}$$

Here s plays the role of a smoothness index while n denotes the usual sample size.

Comments 1. Our proposal is 'adaptive' that is regardless of the target function (taken in a very large class which includes non-homogeneous functions (see section 2.4) it will achieve near optimal rates such as (3), see section 4.1 for a detailed discussion.

2. Our method is computationally sound taking full advantage of the Fast Fourier and Fast Wavelet Transforms available in `matlab`. A key ingredient in the implementation of our procedure is the use of band-limited wavelet basis. Of particular relevance is the availability of a fast algorithm to compute the periodised Meyer wavelet transforms, see Kolaczyk (1993).

3. Our assumptions (A1) and (A2) do not include the case of so-called super-smooth deconvolution (exponential decay of the Fourier transform). Indeed our estimation procedure can be applied in the latter case too. However, recent results of Fan & Koo (2002), obtained in the density setting, show that in the super-smooth case non-linear wavelet estimators can not improve upon linear estimators deriving a much poorer rate than that described at (3).

4. The smoother is the convolution the harder it is recover the original function f as seen from the effect of ν in (3). This contrasts with the practical need for convolution kernels to act as low-pass filters. In practice, Gaussian kernels are often used although theory shows that only poor results can be expected then (see comment 3 above). In this light the box-car kernel is interesting as it combines low-pass characteristics with good recovery rates ($\nu = 3/2$).

5. If the box-car width is rational some frequencies are impossible to recover. This phenomenon can be avoided by considering box-car with irrational support as in our assumption (A2). In this case, Johnstone & Raimondo (2002) have shown that the best-possible recovery rate in the case of a known smoothness s is only a log-term faster than (3) with $\nu = 3/2$.

6. Box-car and ordinary smooth convolutions are often used to model motion blur in digital images, Bertero & Boccacci (1998). While the methodology presented in this paper focuses on *1-d* deconvolution model (1); it is worth mentioning that our proposal can be extented to the deconvolution of *2-d* objects such as blurred digital images. This work is currently in progress.

We begin in section 2 by presenting an application of statistical deconvolution to remote sensing. It is followed by a short introduction on periodised Meyer wavelets, Besov spaces and a review of wavelet shrinkage. In section 3, we introduce our inverse estimation paradigm which builts on recent work of Fan & Koo (2002). After presenting our method into details we compared its numerical performances to more classical filtering procedures. In the implementation of our method choices of tuning parameters are driven by asymptotic minimax theory which is discussed in Section 4. Proofs are summarised in section 5 and 6.

2 Motivations and preliminaries

2.1 Application to remote sensing

Deconvolution is a common problem in many areas of signal and image processing, see e.g. Jain (1989). Here we shall focus on *lidar* remote sensing applications as in Je Park et al. (1997), Harsdorf & Reuter (2000). '*Lidar*' (*LI*ght *D*etection *A*nd *R*anging) 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 distance and position of the reflecting material. Accordingly, the distance resolution is limited by the time resolution of the *lidar*. If the system response function of the *lidar* is longer than the time resolution interval 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* signal are corrupted by additional noise which renders direct deconvolution impossible. Borrowed from Harsdorf & Reuter (2000) ([HR] in the sequel), we have depicted an ideal lidar signal in Figure 1; this will be our target function f for numerical illustrations throughout this paper.

The system response function of the *lidar* (denoted $g(t)$ in (1)) should have a strong low-pass

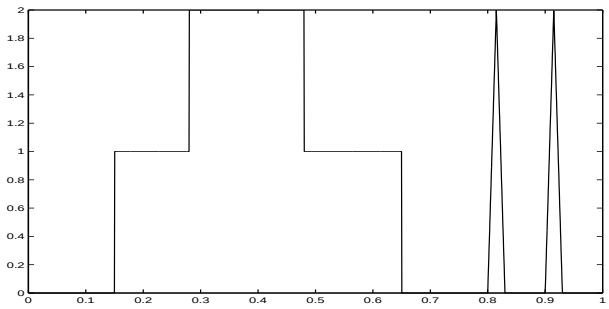

Figure 1: Ideal lidar signal as in [HR]; corresponds to data for an underwater lidar.

characteristic with a frequency cut-off at about 200 MHz. While [HR] uses a Gaussian shape to model the system response function we shall use a Gamma or a Box-Car shape both with low-pass characteristics, see Figure 2. The reason for such a choice is that a much better recovery rate is achievable in the latter cases, see comments 3,4 in section 1. Finally, in Figure 3 we depicted simulated *lidar* data for a combination a different noise levels and system response functions.

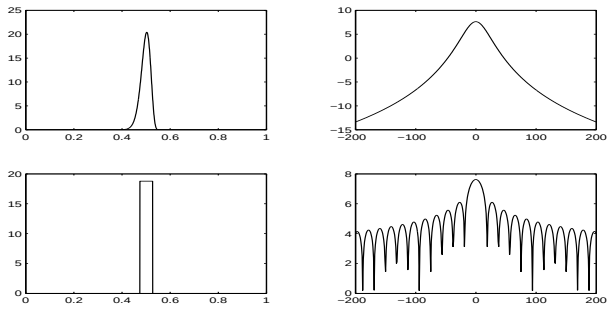

Figure 2: System response functions g (left panels) and their log-|spectrum| (right panels). Top: Gamma shape g-based on the pdf of a $\Gamma(10, 0.0065)$. Bottom: Box-car shape with $a = 1/\sqrt{353}$.

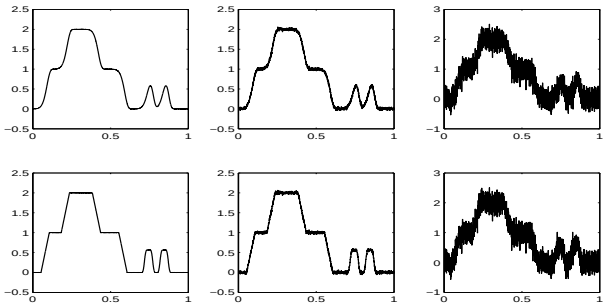

Figure 3: Simulated lidar signals (1) with $t_i = i/n, n = 2048$ corresponding to the system response functions of Figure 2. Top plots: Gamma shape blurring. Bottom plots: Box-Car blurring. From left to right: low (sd=0.002)-medium (sd=0.02)-high (sd=0.2) noise level.

2.2 Periodised Meyer wavelet transforms

In this section we review some basic properties associated with the Meyer wavelet which plays a key role in our procedure. Let (ϕ, ψ) be the Meyer scaling and wavelet function respectively, see Meyer (1990). Recall that

$$\psi_\kappa(x) = \psi_{j,k}(x) = 2^{j/2}\psi(2^j x - k), k \in \mathbb{Z}, j \in \mathbb{Z}, \tag{4}$$

is the dilated-translated wavelet at resolution level j and time position $k/2^j$; here and below κ denotes the bivariate index (j, k). A basic property of wavelet functions is to defined a so-called multi-resolution analysis of $L^2(\mathbb{R})$; for any $f \in L^2(\mathbb{R})$ the following expansion holds

$$f = \sum_k \alpha_{j_0,k}\phi_{j_0,k} + \sum_{j>j_0}\sum_k \beta_{j,k}\psi_{j,k} \tag{5}$$

where

$$\alpha_\kappa = \int f\,\phi_\kappa, \quad \beta_\kappa = \int f\,\psi_\kappa \tag{6}$$

are the wavelet coefficients of f. Quite naturally, one can define a similar multi-resolution analysis for periodic functions in $L^2(T)$, $T = [0, 1]$. This is done by periodising the basis functions

$$\Phi_\kappa(x) = \sum_{l\in\mathbb{Z}} \phi_\kappa(x + l), \quad \Psi_\kappa(x) = \sum_{l\in\mathbb{Z}} \psi_\kappa(x + l). \tag{7}$$

Here and in the sequel of the paper (Φ, Ψ) will denote the periodised Meyer scaling and wavelet functions. Thus, for any periodic function f an expansion similar to (5) holds with periodised basis functions $(\Phi_\kappa, \Psi_\kappa)$ and bivariate index κ restricted to the set $I = \{(j, k) : j \geq 0$ and $k = 0, 1, \ldots, 2^j - 1\}$. Our main motivations for using this wavelet basis are the following:

- The Meyer wavelet is band limited. In particular, we have $Supp(\mathcal{F}\psi) = Supp(\mathcal{F}\Psi) = \{w : |w| \in [2\pi/3, 8\pi/3]\}$, where $\mathcal{F}f$ denotes the Fourier transform of f.
- An efficient algorithm, due to Kolaczyk (1993), is available to compute the periodised Meyer wavelet transforms. It requires only $O(n(log(n))^2)$ steps to compute empirical version of the coefficients (6) from a sample of size n of f.

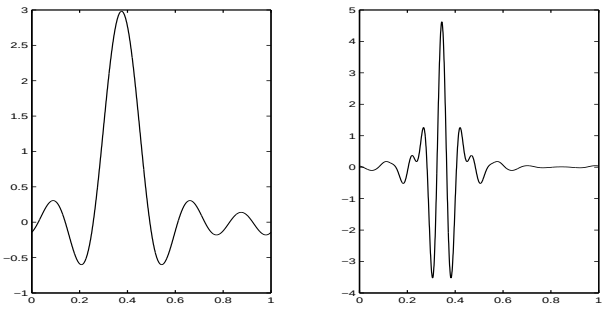

Figure 4: Periodised Meyer scaling and wavelet functions. Left: $\Phi_{3,4}$; right: $\Psi_{4,5}$

2.3 A wide class of target functions

Let us first introduce the standard Besov spaces of periodic functions $B^s_{\pi,r}(T)$, $s > 0, \pi \geq 1$ and $r \geq 1$. For this purpose, we define, for every measurable function f

$$\Delta_\varepsilon f(x) = f(x+\varepsilon) - f(x)$$

then, recursively, $\Delta^2_\varepsilon f(x) = \Delta_\varepsilon(\Delta_\varepsilon f)(x)$ and identically, for $N \in \mathbb{N}^*$, $\Delta^N_\varepsilon f(x)$. Let

$$\rho^N(t, f, \pi) = \sup_{|\varepsilon| \leq t} \left(\int_0^1 |\Delta^N_\varepsilon f(u)|^\pi du \right)^{1/\pi}.$$

Then for $N > s$, we define :

$$B^s_{\pi,r}(T) = \{ f \ periodic : \ \left(\int_0^1 \left(\frac{(\rho^N(t,f,\pi)}{t^s} \right)^r \frac{dt}{t} \right)^{1/r} < \infty \}.$$

(with the usual modifications for r or $\pi = \infty$.)

In the sequel, we shall also keep the standard notation : $\psi_{-1} = \phi$. In this setting, we also recall that the Besov spaces are characterised by the behaviour of the wavelet coefficients (as soon as the wavelet is periodic and has enough vanishing moments). We have in particular, for $f \in L^\pi(T)$,

$$f = \sum \beta_{j,k} \psi_{j,k} \ \in B^s_{\pi,r}(T) \iff \sum_{j \geq 0} 2^{j(s+1/2-1/\pi)r} \Big[\sum_{k \leq 2^j} |\beta_{j,k}|^\pi \Big]^{r/\pi} < \infty. \tag{8}$$

The Besov spaces have proved to be a very interesting scale of spaces for studying the properties of statistical procedures. 's' usually indicates the degree of smoothness of the function. But, due to the effect of other parameters π and r, they are able to model very different kind of smoothness features in a function including non homogeneous behaviour, see Donoho et al. (1995).

2.4 Wavelet shrinkage

Wavelet shrinkage is now a well established statistical procedure used for non-parametric estimation. A wavelet estimator of an unknown function $f \in L^2(T)$ will be written as

$$\hat{f} = \sum_{\kappa \in I_0} \hat{\alpha}_\kappa I_{\{|\hat{\alpha}_\kappa| \geq \lambda_{j_0}\}} \, \Phi_\kappa \ + \sum_{\kappa \in I_1} \hat{\beta}_\kappa \, I_{\{|\hat{\beta}_\kappa| \geq \lambda_j\}} \, \Psi_\kappa \tag{9}$$

where $\hat{\alpha}_\kappa, \hat{\beta}_\kappa$ are estimated wavelet coefficients and I_0, I_1 are set of indices. $I_0 = \{(j_0, k) : k = 0, 1, ..., 2^{j_0} - 1\}$ corresponds to a coarse resolution level j_0 and $I_1 = \{(j, k) : k = 0, 1, ..., 2^j - 1,\ j_0 < j \leq j_1\}$ corresponds to details up to a fine resolution level j_1 . Note that the procedure (9) is non-linear since only statistically significant coefficients: $|\hat{\beta}_\kappa| \geq \lambda_j$ are kept; here λ_j is a threshold parameter. Choices of parameters j_0, j_1, λ_j as well as estimators $\hat{\alpha}_\kappa, \hat{\beta}_\kappa$ depend on the problem at hand. For deconvolution problem (1) this will be discussed in next section.

3 Deconvolution via Fourier for Wavelets

3.1 Inverse estimation paradigm

Due to the ill-posedness nature of the problem, wavelet coefficients (6) can not be recovered directly and some Fourier analysis is required. One of the basic properties of the Fourier transform is to interchange convolution and multiplication. Let $e_l(t) = e^{2\pi i l t}$, $l \in \mathbb{Z}$ and denote $f_l = \langle f, e_l \rangle, g_l = \langle f, e_l \rangle$ the Fourier coefficients of f, g respectively. Letting $h = f \star g$ we have

$$h_l = \langle f \star g, e_l \rangle = f_l \times g_l. \tag{10}$$

Calculating Fourier coefficients in (1):

$$y_l = h_l + n^{-1/2} z_l \tag{11}$$

where $y_l = \langle Y_n(dt), e_l \rangle$ are observable Fourier coefficients and $z_l = \langle W(dt), e_l \rangle$ zero-mean Gaussian random variables. We denote Ψ_l^κ the Fourier coefficients of Ψ_κ i.e. $\Psi_l^\kappa = \langle \Psi_\kappa, e_l \rangle$; combining (10) with Plancherel's identity we obtain

$$\int_T f \bar{\Psi}_\kappa = \sum_l f_l \bar{\Psi}_l^\kappa = \sum_l \Big(\frac{h_l}{g_l}\Big) \bar{\Psi}_l^\kappa. \tag{12}$$

Noting that $\bar{\Psi}_\kappa = \Psi_\kappa$, one can recover wavelet coefficients

$$\beta_\kappa = \int_T f \Psi_\kappa = \sum_l \Big(\frac{h_l}{g_l}\Big) \bar{\Psi}_l^\kappa. \tag{13}$$

Here g_l and Ψ_l^κ are known Fourier coefficients but the h_l's are not directly observable; in (11) we take y_l as an (unbiased) estimator of h_l and let

$$\hat{\beta}_\kappa = \sum_l \Big(\frac{y_l}{g_l}\Big) \bar{\Psi}_l^\kappa \tag{14}$$

be our estimator of β_κ which can be computed from the observations (1). Of course, an estimator $\hat{\alpha}_\kappa$ of α_κ is defined in a similar fashion with Φ in place of Ψ.

3.2 Optimal tuning

Given some estimated coefficients $\hat{\alpha}_\kappa, \hat{\beta}_\kappa$, the wavelet-based estimator (9) requires only three input parameters: j_0, j_1 and λ; a fine tuning is required, however. This is particularly important for the threshold parameter λ and the high frequency cut-off j_1. Asymptotic minimax theory is used to derive optimal choice of tuning parameters. For example, in direct estimation problems Donoho et al. (1995) have shown that the following tuning was the 'best-possible':

$$\lambda_j := \lambda_n = \hat{\sigma} \sqrt{(2 \log n)/n} \tag{15}$$

where $\hat{\sigma}$ is an estimated scale from the data and

$$2^{j_1} = O(\log(n)/n). \tag{16}$$

For deconvolution problems (1) we will use wavelet based estimator (9) with coefficients (14). Our main result (proposition 1, section 4.1) states that the following tuning is the 'best-possible' for a wide variety of target function (section 2.3)

$$\lambda_j := \lambda_{n,j} = \eta\,\hat{\sigma}\,\tau_j\,\sqrt{(\log n)/n} \tag{17}$$

and

$$2^{j_1} = O\Big((n/\log n)^{1/(1+2\nu)}\Big). \tag{18}$$

where τ_j is a level-by-level scale which depends on the function g and η is a constant. Let $|C_j|$ denote the cardinal of $C_j = \{l : \Psi_l^\kappa \neq 0\}$ and take

$$\tau_j^2 = |C_j|^{-1} \sum_{C_j} g_l^{-2}. \tag{19}$$

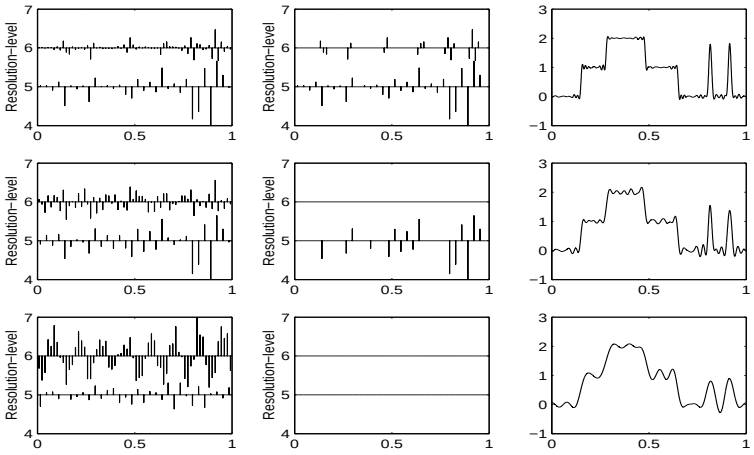

Figure 5: The WaveD-method. From top to bottom: Deconvolution of lidar signal of figure 3 (g =Box-Car) in low-medium-high noise. Left panels: estimated wavelet coefficients (13). Middle panels estimated wavelet coefficients (13) after shrinkage (17); right panels: reconstructed lidar signal.

3.3 The WaveD-method

We summarise the main steps of our Wavelet-Deconvolution procedure which are illustrated in Figure 5. Here and in the sequel of the paper we refer to this method as the *WaveD*-method.

(a) Compute Fourier coefficients y_l, g_l, Ψ_l^κ and recover wavelet coefficients (14), see left panels of Figure 5.

(b) Compute an estimate of the noise standard deviation, $\hat{\sigma}$ say . For direct data, a commonly used estimator is found in Donoho et al. (1995); here we applied this method to the set of wavelet coefficients $\langle Y_n(dt), \Psi_{J,k}\rangle$, where J is the largest accessible resolution level. Combining this with g_l and (19) we compute the threshold $\lambda_j := \lambda_{n,j}$ as shown in (17). Note that for the Meyer wavelet we have $|C_j| = 4\,\pi\,2^j$. Numerical results for $j = 5, 6, 7$ and $\eta = 1$ are given in Table 1.

Table 1: *Level-by-level thresholds λ_j defined by (17) with $\eta = 1$ for different noise levels (Indicated in brackets are the percentage of shrinked coefficients at the corresponding resolution level).*

Noise level	$\hat{\sigma}$	$j = 5$	$j = 6$	$j = 7$
Low : $sd = 0.002$	0.0019	0.0018 (%12)	0.0106 (%79)	0.0118 (%100)
Medium : $sd = 0.002$	0.0185	0.0180 (%62)	0.1031(%100)	0.1146 (%100)
High : $sd = 0.2$	0.1898	0.1846 (%100)	1.0578 (%100)	1.1753 (%100)

(d) Apply thresholding $|\beta_\kappa| > \lambda_j$, see middle panels of Figure 5. Finally, invert the wavelet transform to obtain an estimate of f, see right panels of figure 5.

All computations were carried on the `wavelab` package on `MATLAB`, see Buckheit et al. (1995). For information on the Meyer toolbox we refer to Kolaczyk (1993).

3.4 Competing approaches

Wiener filter Regularised Deconvolution (WiRD) is a commonly used method to recover a function f from noisy-blurred observations (1). Denoting $\mathcal{F}_h$ the Fourier transform of the function h, the WiRD-estimator of f is defined in the Fourier domain :

$$f_\alpha(\omega) := G_\alpha(\omega)\mathcal{F}_Y(\omega) \tag{20}$$

where

$$G_\alpha(\omega) := \Big(\frac{1}{\mathcal{F}_g(\omega)}\Big)\frac{|\mathcal{F}_g(\omega)|^2|\mathcal{F}_W(\omega)|^2}{|\mathcal{F}_g(\omega)|^2|\mathcal{F}_z(\omega)|^2 + \alpha\sigma^2}. \tag{21}$$

Then we take $\hat{f}_\alpha = \mathcal{F}^{-1}(f_\alpha)$ as an estimator of f. Here α is a regularisation parameter which balances noise suppression with signal distortion, see Galatsanos & Katsaggelos (1992). Small values of α give an unbiased but noisy estimate whereas large values of α suppress the noise but also distort the signal. For $\alpha = 1$, this corresponds a Linear Time Invariant *Wiener Filter* which is known to be optimal, in L^2, provided that the input signal f is stationary in a wide sense, see e.g. Jain (1989).

Table 2: *Monte-Carlo approximations to* $\mathbb{E}\|\hat{f} - f\|_p^p$. The results are based on 1000 independent simulations of the model (1) with $n = 2048$ as in Figure 3 (box-car blurring). (*) for $sd = 0.2$ we reported median values as both WiRD and WaRD means were highly affected by some very large values.

sd	Method	$p = 1.05$	$p = 2$	$p = 5$
0.002	WaveD	**0.0301**	**0.0037**	*0.0002*
0.002	WiRD	0.0441	0.0058	*0.0002*
0.002	WaRD	0.0383	0.0080	0.0005
0.02	WaveD	**0.0761**	**0.0143**	**0.0007**
0.02	WiRD	0.1502	0.0593	0.0219
0.02	WaRD	0.0870	0.0232	0.0206
$0.2^\star$	WaveD	**0.1643**	**0.0730**	**0.0316**
$0.2^\star$	WiRD	0.5648	0.4935	0.9829
$0.2^\star$	WaRD	0.7563	0.8710	3.48

Wavelet Regularised Deconvolution (WaRD) is an hybrid Fourier/Wavelet deconvolution method, see Neelamani et al. (2000). The method is to first apply *Wiener Filtering* (20) deriving an estimator $\hat{f}_\alpha$ of f. In a second step, one further smooth $\hat{f}_\alpha$ using direct wavelet thresholding (9) as in Donoho et al. (1995). Here too α plays the role of a regularisation parameter which balances noise level and signal distortion. The motivation behind WaRD is to remove the remaining noise in a WiRD estimator. For typical non-homogeneous signals such as in Figure 1, the WaRD estimator has better visual quality and often a smaller mean square error than a WiRD estimator.

3.5 Numerical performances

We compared the WaveD method to the WiRD and WaRD methods in an extensive simulation study. Performances were recorded with respect to various L^p-metrics for different noise levels. For each combination of noise level and L^p-metric we computed the Monte-Carlo approximation to $\mathbb{E}\|\hat{f} - f\|_p^p$. For each simulation, the regularisation parameters α_1 and α_2 (of the WiRD and WaRD method respectively) were chosen numerically to minimise the observed error i.e. $\hat{\alpha} = \arg\min_{\alpha\in(0,1000)} \sum_i |\hat{f}_\alpha(i/n) - f(i/n)|^p$. Of course, this is only possible when knowing the target function f (in practice one could only approximate such a value). We used several functions f, g and different sample sizes from $n = 256$ to $n = 2048$. We observed a pattern similar to that of Table 2 where we summarised the results we obtained for the *lidar* example.

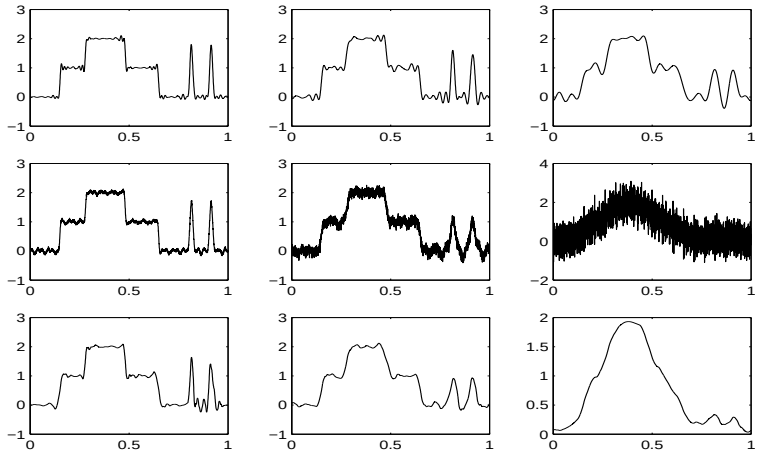

Figure 6: From top to bottom: WaveD-WiRD-WaRD estimator; from left to right low (sd=0.002)-medium (sd=0.02) -high (sd=0.2) noise level.

Comment on the results. The visual superiority of the WaveD-method, as seen Figure 6, is confirmed by our simulation results (Table 2). While for low noise-level all methods behave fairly well there are striking differences in medium to high noise level in particular for metrics with $p > 2$. We see that the WaveD-estimator outperformed both the WiRD and the WaRD methods in all cases but one. These are quite promising results given that the WaveD-method does not required any choice of regularisation parameter. As seen in Table 1, the fraction of shrinked coefficients automatically adjust to the noise level: for low noise level some wavelet coefficients are kept up-to resolution $j = 6$ whereas for high noise level all wavelet coefficients above $j_0 = 5$ are shrinked resulting in a linear wavelet filter. In other words, the WaveD-method provides an adaptive-non-linear-wavelet-filter which, unlike *Wiener-Filtering*, preserves sharp features in the original signal while suppressing high level of noise.

4 Asymptotic theory

4.1 Near optimality for a wide range of smoothness classes

The following proposition states the near optimal properties (up to logarithmic factors) of our procedure. We refer to section 1 for assumptions (A1) and (A2).

Proposition 1. *Suppose that we observe the random process (1) under assumptions (A1) or (A2). Let $p > 1$ be an arbitrary number. If f belongs to $B^s_{\pi,r}(T)$ with $\pi \geq 1$, $s \geq 1/\pi$ and $0 < r \leq \max(\frac{p(2\nu+1)}{2(\nu+s)+1}, \frac{(2\nu+1)p-2}{2(s+\nu)-2/\pi+1})$ then, for η large enough the wavelet based estimator (9) with threshold (17) and maximum level (18), is such that :*

$$\mathbb{E}\|\hat{f} - f\|_p^p \leq C[n^{-1}\log(n)]^\alpha, \tag{22}$$

$$\alpha = \frac{2sp}{1+2(\nu+s)}, \quad \textit{if } s \geq (2\nu+1)(\frac{p}{2\pi} - \frac{1}{2}) \tag{23}$$

and

$$\alpha = \frac{2(s-1/\pi+1/p)p}{1+2(\nu+s-1/\pi)}, \quad \textit{if } \frac{1}{\pi} - \frac{1}{2} - \nu \leq s < (2\nu+1)(\frac{p}{2\pi} - \frac{1}{2}). \tag{24}$$

Remarks 1. There is an elbow in the rate at $s = (2\nu+1)(\frac{p}{2\pi} - \frac{1}{2})$. 'Elbow' phenomenon are commonly observed for direct models; recently such effects have been observed in a deconvolution problem, see Johnstone & Raimondo (2002). We notice that the rates in the dense (23) and sparse (24) cases are very comparable to those observed in the direct case, with the additional presence of the parameter ν sometimes referred as the Degree of Ill-Posedness.

2. Concerning the rate for $p = 2$ and under assumption (A1), let us mention results similar to (23) obtained by Pensky & Vidakovic (1999), Fan & Koo (2002) in the context of deconvolution in a density model. Indeed, up-to-log factors, such results can not be improved by any estimators.

3. Under assumption (A2) our result extends that of Johnstone & Raimondo (2002). It provides an adaptive estimator which is near-optimal in various L^p-metrics.

4. Interestingly enough the implementation of our method bears elementary differences from traditional wavelet estimators used in the direct case Donoho et al. (1995) or in the indirect case Fan & Koo (2002). Particular features include: (a) the use of band-limited wavelets which renders possible the computation of coefficients in the Fourier domain where the wavelets are compactly supported; (b) the finest resolution level j_1 depends on the Degree of Ill-Posedness of the problem, see condition (18); (c) level-by-level thresholds depend on the system response function g (and the wavelet) which allows fine tuning even for box-car deconvolution.

5. In practice, we used the threshold (17) with $\eta = 1$ which gave good results in our simulation study. Small variation in the choice of η did not affect our results. As for the finest resolution j_1, we used a data-driven method: we took j_1 to be first level j with %100 of shrinked coefficients. For example in our simulation study this would corresponds to $j_1 = 5, 6$ or 7 depending on the level of noise (Note that these are smaller values than those used the in the direct case). This is consistent with our condition (18) which suggests a smaller j_1 for ill-posed problems. Since, in every cases, the optimal level (i.e. the level where bias and stochastic terms are balanced) is always smaller than j_1, introducing unnecessary levels of resolution can only induce instability in the estimator. This was confirmed by poorer numerical performances when we tested the method with larger values of j_1.

4.2 The Maxiset-approach

Near-optimal properties of our proposal are direct applications of the following theorem borrowed from Kerkyacharian & Picard (2000). This theorem gives the 'Maxiset' (condition (29)) for a general wavelet estimator of the form (28). It will be applied directly to our procedure as outlined in section 5. We refer to the appendix for condition (54) (known as the Temlyakov property). First, we introduce some notation: μ will denote the measure such that for $j \in \mathbb{N},\ k \in \mathbb{Z}$,

$$\mu\{(j,k)\} = \|\sigma_j \Psi_{j,k}\|_p^p = \sigma_j^p 2^{j(\frac{p}{2}-1)} \|\Psi\|_p^p$$

$$l_{q,\infty}(\mu) = \left\{ f,\ \sup_{\lambda>0} \lambda^q \mu\{(j,k)/\ |\beta_{j,k}| > \sigma_j \lambda\} < \infty \right\}$$

Theorem 1. *Let $p > 1$, $0 < q < p$, $\{\ \psi_{j,k}, j \geq -1, k \in \mathbb{N}\}$ be a periodised wavelet basis of $L^2(T)$ and σ_j be a positive sequence such that the heteroscedastic basis $\sigma_j \psi_{j,k}$ satisfies property (54). Suppose that Λ_n is a set a pairs (j,k) and c_n is a deterministic sequence tending to zero with*

$$\sup_n \mu\{\Lambda_n\}\, c_n^p < \infty. \tag{25}$$

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

$$\mathbb{E}|\hat{\beta}_\kappa - \beta_\kappa|^{2p} \ \leq\ C\,(\sigma_j\, c_n)^{2p} \tag{26}$$

$$P\Big(|\hat{\beta}_\kappa - \beta_\kappa| \geq \eta\, \sigma_j\, c_n/2\Big) \ \leq\ C\,(c_n^{2p} \wedge c_n^4) \tag{27}$$

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

$$\hat{f}_n = \sum_{\kappa \in \Lambda_n} \hat{\beta}_\kappa\, \psi_\kappa\, I\{|\hat{\beta}_\kappa| \geq \eta\, \sigma_j\, c_n\} \tag{28}$$

is such that, for all positive integers n,

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

if and only if :

$$\begin{array}{rcl} f & \in & l_{q,\infty}(\mu), \quad \textit{and}, \\ \sup_n c_n^{q-p} & \| & f - \sum_{\kappa \in \Lambda_n} \beta_\kappa \psi_\kappa \|_p^p < \infty. \end{array} \tag{29}$$

Remarks 1. Through condition (29) and in the light of section 2.3 the theorem gives the 'Maxiset' of the method i.e. the set of functions where the method attains a given rate of convergence. This way of measuring the performances of statistical procedures has been particularly successful in the non-parametric framework. It has often the advantage of giving less arbitrary and pessimistic comparisons of procedures than the minimax approach.

2. We will prove (see section 6.1) that the Besov spaces $B^s_{\pi,r}(T)$ are included in the Maxiset defined in (29) for the rate given in (22). In particular, we have :

$$\|f - \sum_{\kappa \in \Lambda_n} \beta_\kappa \Psi_\kappa\|_p \asymp \|f - P_{V_{j_1}} f\|_p \tag{30}$$

where $P_{V_{j_1}}$ denotes the projection on the space V_{j_1} of the multi-resolution analysis associated to the wavelet basis. In this case it appears more clearly that the second part of (29) is in fact directly linked to the well-known belonging of a standard Besov space. This part is responsible of the condition $s \geq 1/\pi$ in the assumptions of proposition 1.

5 Proofs

Outline of the proof of proposition 1. We will prove that proposition 1 follows from Theorem 1. For this purpose, we will consider the wavelet based estimator (9) with threshold (17) and maximum level (18), that is to say

$$\Lambda_n = \{(j,k),\ -1 \le j \le j_1, 0 \le k \le 2^j\}, \quad c_n = \left(\frac{\log n}{n}\right)^{1/2}, \quad 2^{j_1} = O\Big(\frac{n}{\log n}\Big)^{\frac{1}{1+2\nu}}. \tag{31}$$

In this setting, and under assumptions (A1) and (A2), we will prove the following claims :

(C1): Inequalities (26) and (27) hold with $\sigma_j = \tau_j$ as in (19) and η large enough.

(C2): The basis $(\sigma_j \Psi_{j,k})$ satisfies condition (54) (see section 6.2) as soon as there exists a constant C, such that for any finite subset Λ of $\mathbb{N}$:

$$\sum_{j\in\Lambda} \quad 2^j\sigma_j^2 \le C \sup_{j\in\Lambda} 2^j\sigma_j^2 \qquad if\ 2 < p < \infty \tag{32}$$

$$\sum_{j\in\Lambda} \quad 2^{jp/2}\sigma_j^p \le C \sup_{j\in\Lambda} 2^{jp/2}\sigma_j^p \quad if\ 1 < p < 2 \tag{33}$$

Note also that for $p = 2$ condition (54) holds without any condition on σ_j.

(C3): If we take $\sigma_j = \tau_j$ as defined by (19) then conditions (32) and (33) are satisfied.

(C4): The condition (25) is verified as soon as we have $\sigma_j \asymp C2^{j\nu}$.

Assuming claims (C1),...,(C4) we see that theorem 1 applies to the wavelet based estimator (9) with the setting (31). This combined with remarks following theorem 1 proves proposition 1. To complete the proof we shall now prove the claims.

Proof. (C1): First, we derive the bias and variance of $\hat{\beta}_\kappa$. Taking expectation in (11) we have

$$\mathbb{E}\,\hat{\beta}_\kappa = \sum_l \Big(\frac{h_l}{g_l}\Big)\Psi_l^\kappa = \beta_\kappa \tag{34}$$

$$\mathrm{Var}\hat{\beta}_\kappa = \mathbb{E}\Big(\sum_l \Big(\frac{y_l - h_l}{g_l}\Big)\Psi_l^\kappa\Big)^2 = \mathbb{E}\Big(n^{-1/2}\sum_l \Big(\frac{z_l}{g_l}\Big)\Psi_l^\kappa\Big)^2 = n^{-1}\sum_l \Big(\frac{\Psi_l^\kappa}{g_l}\Big)^2. \tag{35}$$

Note that $C_j = \{l : \Psi_l^\kappa \neq 0\}$ and that $|\Psi_l^\kappa| \le C\,2^{-j/2}$:

$$\mathrm{Var}\hat{\beta}_\kappa \le C\,n^{-1}\,2^{-j}\sum_{l\in C_j}\Big(\frac{1}{g_l}\Big)^2.$$

Hence recalling the notation (19), we have :

$$\mathrm{Var}\hat{\beta}_\kappa \le C'\,n^{-1}\,\tau_j^2. \tag{36}$$

As the $\hat{\beta}_\kappa$'s are Gaussian :

$$\mathbb{E}|\hat{\beta}_\kappa - \beta_\kappa|^{2p} \le C_{2p}[\mathrm{Var}\hat{\beta}_\kappa]^{p/2} \tag{37}$$

which combined with (36)

$$\mathbb{E}|\hat{\beta}_\kappa - \beta_\kappa|^{2p} \leq c_2 \left(\frac{\tau_j^2}{n}\right)^{p/2} \tag{38}$$

where $c_2 = C'^{p/2} C_{2p}$; hence proving inequality (26). As again, the $\hat{\beta}_\kappa$'s are Gaussian

$$P\Big(|\hat{\beta}_\kappa - \beta_\kappa| \geq \eta \sigma_j c_n\Big) \leq \frac{2}{\eta c_n} \exp\left(-\frac{2\eta^2 \log n}{2C}\right)$$

Hence for $\eta = \eta_\gamma$ large enough, we have, for any $\gamma > 0$

$$P\Big(|\hat{\beta}_\kappa - \beta_\kappa| \geq \lambda_j\Big) \leq c_2 n^{-\gamma} \tag{39}$$

which proves inequality (27). □

Proof. (C2): This is a direct application of theorem 2 (see appendix). □

Proof. (C3): Clearly condition (32),(33) will be true will be true for any σ_j of the form $2^{j\nu}$. We shall prove that for $\sigma_j = \tau_j$ as in (19) the latter condition holds. Recall that

$$\tau_j^2 = |C_j|^{-1} \sum_{C_j} g_l^{-2}.$$

Here Ψ is band-limited hence $C_j = \{l : 2^j \leq |l| \leq 2^{j+r}\}$, for some fixed $r > 0$. To simplify the exposition we shall further assume that $C_j = \{l : 2^j \leq l \leq 2^{j+r}\}$; noting that, by symmetry, bounds below hold for negative values of l too. Under assumption (A1): $g_l \sim C|l|^\nu$,

$$\tau_j^2 = |C_j|^{-1} \sum_{l=2^j}^{2^{j+r}} l^{2\nu} \asymp 2^{-j} 2^{j(2\nu+1)} \asymp 2^{2j\nu}, \tag{40}$$

which proves (C3) under (A1). The proof under (A2) is a little bit more evolved as for the box-car function the Fourier coefficients are given by

$$g_l = \frac{\sin \pi l a}{\pi l a} \tag{41}$$

The behaviour of (41) is driven by Diophantine approximation

$$||la|| := \inf\{|la - p|, p \in \mathbb{N}\}\ ,\ l = 1, 2, \ldots \tag{42}$$

where $\|x\|$ denote the distance from $x \in \mathbb{R}$ to the nearest integer. Indeed, combining the periodic behaviour of the sine function together with a first order (linear) approximation, we find

$$\frac{2}{\pi} \frac{\|la\|}{la} \leq g_l \leq \frac{\|la\|}{la}. \tag{43}$$

Although $g_l, l = 1, 2, ..$ behave in an erratic fashion (see Figure 2), the average behaviour is much less susceptible to fluctuation and we shall prove that

$$\tau_j^2 = |C_j|^{-1} \sum_{l=2^j}^{2^{j+r}} g_l^{-2} \asymp 2^{3j\nu}. \tag{44}$$

Result (44) follows from (43) and the following lemma (see Johnstone & Raimondo (2002)). We refer to the appendix for the notion of BA number.

Lemma 1. *Let p/q and p'/q' be successive principal convergents in the continued fraction expansion of a real number a. Let N be a non-negative integer with $N+q<q'$. Then, for BA number a,*

$$\sum_{l=N+1}^{N+q} \|la\|^{-2} \asymp q^2 \tag{45}$$

Starting at (19) and using (43), we see that

$$\tau_j^2 \asymp 2^{-j} \sum_{l\in C_j} \Big(\frac{l}{\|la\|}\Big)^2 \asymp 2^{-j}\, 2^{2j} \sum_{l\in C_j} \|la\|^{-2} \asymp 2^j \sum_{l\in C_j} \|la\|^{-2}. \tag{46}$$

Let m be the smallest index such that $q_m \geq 2^j$. Recall that r is fixed and that $C_j = \{l : 2^j \leq l \leq 2^{j+r}\}$. The geometric growth of the denominators q_n (compare (60)) implies that $q_{m+2r} \geq 2^r q_m > 2^{j+r}$, so that

$$C_j \subset \mathbb{N} \cap [1, q_{m+2r}).$$

Introduce intervals $D_0 = \mathbb{N}\cap[1,q_m)$ and $D_\tau = \mathbb{N}\cap[q_{m+\tau-1}, q_{m+\tau})$ for $\tau = 1,\dots,2r$ which together cover C_j. Since a is BA, there is an integer $K = K(a)$ such that $q_{n+1} \leq K q_n$ for all n. Hence there are at most K disjoint blocks of length $q_{m+\tau}$ that cover D_τ. Apply Lemma 1 to each of these blocks:

$$\sum_{l\in D_\tau} \|la\|^{-2} \asymp q_{m+\tau}^2, \qquad\qquad 0 \leq \tau \leq 2r-1.$$

Since $q_{m+\tau} \leq K^\tau q_m$, we combine over τ to get

$$\sum_{l\in D_1} \|la\|^{-2} \leq \sum_{l\in C_j} \|la\|^{-2} \leq \sum_{\tau=0}^{2r-1} \sum_{l\in D_\tau} \|la\|^{-2} \leq C q_m^2 \sum_{\tau=0}^{2r-1} K^{\tau+1}.$$

Combining this with (46) and noting that $q_m \leq K q_{m-1} \leq K 2^j$, we recover (44):

$$\tau_j^2 \asymp 2^j\, (2^j)^2 \asymp 2^{3j}.$$

□

Proof. (C4): Since $\nu > 0$, in the case where $\sigma_j \asymp C 2^{j\nu}$ we have

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

Notice that for $p>1$, $p\nu + p/2 > 1$ is equivalent to $\nu > 1/p - 1/2$. Now by definition (18) of j_1

$$2^{j_1 \frac{p}{2}(2\nu+1)} \asymp c_n^{-p} \asymp (\frac{n}{\log n})^{p/2}$$

which proves (25). □

6 Appendix

6.1 Embedding of Besov spaces

Our aim here is to investigate which particular periodic Besov space may be embedded in the spaces $l_{q,\infty}(\mu)$ as well as imply the condition :

$$\sup_n c_n^{q-p} \|f - \sum_{(j,k)\in\Lambda_n} \beta_{j,k}\Psi_{j,k}\|_p^p < \infty. \tag{47}$$

Let us remind that we will concentrate on the case where

$$\mu(j,k) = \tau_j^p 2^{j(p/2-1)}, \quad \tau_j = 2^{j\nu}, \; 2^{j_1} = \Big(\frac{n}{\log n}\Big)^{\frac{1}{2\nu+1}}, \; c_n = \Big(\frac{\log n}{n}\Big)^{1/2}.$$

First, we observe that (47) will be satisfied when f belongs to $B_{p,\infty}^{(\nu+1/2)(1-q/p)}(T)$. Hence, we only need to prove that $B_{\pi,r}^s(T)$ is included in the latter. To this aim we will use two types of Besov embeddings, setting appropriate conditions on s, π, r and q :

- In the periodic setting, we have :

$$B_{\pi,r}^s(T) \subset B_{\rho,r}^s(T), \quad if \; 0 < \rho \leq \pi. \tag{48}$$

- In the general case, we have the standard 'Sobolev embeddings' :

$$B_{\pi,r}^s(T) \subset B_{\rho,r}^{\sigma'}(T), \quad if \; \rho > \pi, \quad and \; s - 1/\pi = \sigma' - 1/\rho. \tag{49}$$

To prove (47), we are interested in taking $\rho = p$. Hence for the case $p \leq \pi$, we need to prove (since $s > 0$ implies that only the dense case (23) can occur) that $s \geq (\nu + 1/2)(1 - q/p) = (\nu + 1/2)2s/(1 + 2\nu + 2s)$ (since, $1 - q/p = 2s/(1 + 2\nu + 2s)$). This is always true for $s > 0$.

For the case, $p > \pi$, we have to prove in the dense case (23) that $\sigma' \geq (\nu+1/2)2s/(1+2\nu+2s)$. This is equivalent to $2s(\sigma') + (1 + 2\nu)(1/p - 1/\pi) \geq 0$. But in this case the left hand side is greater than $(1 + 2\nu)(p/\pi - 1)(s - 1/\pi) \geq 0$. In the sparse case (24), we have to check that : $\sigma' \geq (2\nu + 1)s'/(1 + 2\nu + 2s')$, but this is equivalent to $(2\nu + 1)p/[(2\nu + 1)p - 2 + 2p\sigma'] \leq 1$ or $s \geq 1/\pi$.

Now let us turn to the problem of embedding a particular space $B_{\pi,r}^s(T)$ into $l_{q,\infty}(\mu)$. First let us mention that we will simplify the problem by considering the embedding into

$$l_q(\mu) = \{f = \sum_{jk} (\frac{|\beta_{j,k}|}{\tau_j})^q \mu(j,k) < \infty\}.$$

Using Markov inequality : $l_q(\mu) \subset l_{q,\infty}(\mu)$. We observe that in the case where

$$s = (2\nu + 1)(\frac{p}{2q} - \frac{1}{2}) \tag{50}$$

we have

$$l_q(\mu) = B_{q,q}^s(T).$$

Hence deriving the advertised rate of convergence since here $p - q = \frac{2sp}{1+2(\nu+s)}$.
It remains to study the more intricated cases where we have not $\pi = r = q$.

Proposition 2. *1. Let q be defined by the relation (50), if $0 < r \leq q$ and*

$$s \geq (2\nu+1)(\frac{p}{2\pi} - \frac{1}{2}) \tag{51}$$

then

$$B^s_{\pi,r}(T) \hookrightarrow l_q(\mu) = B^s_{q,q}(T).$$

2. Let q be defined by

$$p - q = \frac{2s'p}{1+2(\nu+s')}, \quad s' = \frac{s - 1/\pi + 1/p}{1 - 2/[(2\nu+1)p]} \tag{52}$$

if $0 < r \leq q$ and

$$\frac{1}{\pi} - \frac{1}{2} - \nu < s < (2\nu+1)(\frac{p}{2\pi} - \frac{1}{2}) \tag{53}$$

then

$$B^s_{\pi,r}(T) \hookrightarrow l_q(\mu) = B^{s'}_{q,q}(T).$$

Remark. The case (53) implies that

$$p > \frac{2}{1+2\nu}$$

For $\nu \geq 1/2$ this not a restriction, since we are considering $1 < p < \infty$. Moreover, in this case the first member of the inequality (53) is always true if we deal with $1 \leq \pi$, as $\frac{1}{\pi} - \frac{1}{2} - \nu \leq 0$.

Proof. We have :

$$\begin{aligned} \sum_{j,k} (\frac{|\beta_{j,k}|}{\tau_j})^q \mu(j,k) &= \sum_{j,k} (\frac{|\beta_{j,k}|}{\tau_j})^q \tau_j^p 2^{j(p/2-1)} = \\ \sum_{j,k} |\beta_{j,k}|^q \tau_j^{p-q} 2^{j(p/2-1)} &= \sum_{j,k} |\beta_{j,k}|^q 2^{j((\nu+1/2)p - \nu q - 1)} \end{aligned}$$

recalling that q has been chosen in such a way that

$$l_q(\mu) = B^s_{q,q}(T),$$

and using (8), we obtain the following characterisation of $B^s_{q,q}(T)$:

$$\sum_{j,k} |\beta_{j,k}|^q 2^{j(s+1/2)q-1} < \infty.$$

We will now use embeddings (48) and (49) taking $\rho = q$:

1. If $s \geq (2\nu+1)(\frac{p}{2\pi} - \frac{1}{2}), r \leq q$ we have $q \leq \pi$, hence using (48) :

   $$B^s_{\pi,r}(T) \hookrightarrow l_q(\mu) = B^s_{q,q}(T)$$

   (Let us observe in addition that $p > q \iff s > 0$.)

2. If $s < (2\nu+1)(\frac{p}{2\pi} - \frac{1}{2})$, $r \leq q$ we will use (49) to find an embedding with a different order of smoothness. This explains the our definition of q. Solving

   $$s - \frac{1}{\pi} = s' - \frac{1}{q}, \quad s' = (2\nu+1)(\frac{p}{2q} - \frac{1}{2}), \quad \pi < q.$$

   and using (49) gives :

   $$B^s_{\pi,r}(T) \hookrightarrow B^{s'}_{q,q}(T) = l_q(\mu).$$

   We now have to check that $p > q$, but this is equivalent to : $\frac{1}{\pi} - \frac{1}{2} - \nu < s$.

□

6.2 Temlyakov inequalities

Let us recall the Temlyakov property for a basis $e_n(x)$ in L^p is the following : there exists absolute constants c, C such that for all $\Lambda \subset \mathbb{N}$,

$$c\sum_{n\in\Lambda}\int |e_n(x)|^p d\mu \leq \int (\sum_{n\in\Lambda}|e_n(x)|^2)^{p/2} d\mu \leq C\sum_{n\in\Lambda}\int |e_n(x)|^p d\mu$$

or, equivalently :

$$c'\|(\sum_{n\in\Lambda}|e_n(x)|^p)^{1/p}\|_p \leq \|(\sum_{n\in\Lambda}|e_n(x)|^2)^{1/2}\|_p \leq C'\|(\sum_{n\in\Lambda}|e_n(x)|^p)^{1/p}\|_p \tag{54}$$

Obviously the left hand side is always true for $p \geq 2$ with $c = 1$, whereas the right hand side is always true for $p \leq 2$ with $C = 1$. In this section, we will prove the following result :

Theorem 2. *Let ϕ be a scaling function of a multi-resolution analysis and ψ the associated wavelet. Let us assume that*

$$|\phi(x)| + |\psi(x)| \leq \frac{C}{1+|x|}$$

If there exists a constant $C < \infty$, such that for all $A \subset \mathbb{N}$

$$\Big(\sum_{j\in A}(2^{j/2}\sigma_j)^{p\wedge 2}\Big)^{\frac{1}{p\wedge 2}} \leq C'\Big(\sum_{j\in A}(2^{j/2}\sigma_j)^{p\vee 2}\Big)^{\frac{1}{p\vee 2}} \tag{55}$$

then the weighted wavelet basis $\{\sigma_j 2^{j/2}\psi(2^j x - k)\ j \in \mathbb{N}, k \in \mathbb{Z}\} \cup \{\sigma_0\phi(x-k)k \in \mathbb{Z}\}$ satisfies Temlyakov property.

Proof. We start by proving the theorem for the Haar basis. Introducing the weighted Haar basis $(\sigma_j 2^{j/2}h_\kappa)$ where as usual $h_\kappa(x) = h_{j,k}(x) = h(2^j x - k)$, $h(x) = 1_{[0,1]}(2x) - 1_{[0,1]}(2x-1)$.

Let us suppose first that $p < 2$ and there exists $C < \infty$, such that for all $A \subset \mathbb{N}$,

$$\Big(\sum_{j\in A}(2^{j/2}\sigma_j)^2\Big)^{1/2} \leq C\Big(\sum_{j\in A}(2^{j/2}\sigma_j)^p\Big)^{1/p}. \tag{56}$$

Typically this is true when $\sigma_j = 2^{j\nu}$. If (56) is true, we have for all $\Lambda \subset \mathbb{N}\times\mathbb{Z}$, point-wise :

$$\Big(\sum_{\kappa\in\Lambda}|\sigma_j 2^{j/2}h_\kappa(x)|^2\Big)^{1/2} \leq C\Big(\sum_{\kappa\in\Lambda}|\sigma_j 2^{j/2}h_\kappa(x)|^p\Big)^{1/p}$$

so in this case

$$\|(\sum_{\kappa\in\Lambda}|\sigma_j 2^{j/2}h_\kappa(x)|^2)^{1/2}\|_p \leq C\|(\sum_{\kappa\in\Lambda}|\sigma_j 2^{j/2}h_\kappa(x)|^p)^{1/p}\|_p.$$

Using (56) for $p \leq 2$,

$$\|(\sum_{\kappa\in\Lambda}|\sigma_j 2^{j/2}h_\kappa(x)|^p)^{1/p}\|_p \leq \|(\sum_{\kappa\in\Lambda}|\sigma_j 2^{j/2}h_\kappa(x)|^2)^{1/2}\|_p \leq C\|(\sum_{\kappa\in\Lambda}|\sigma_j 2^{j/2}h_\kappa(x)|^p)^{1/p}\|_p$$

Now we suppose that $p > 2$ and that there exists $C' < \infty$, such that for all $A \subset \mathbb{N}$,

$$\Big(\sum_{j\in A}(2^{j/2}\sigma_j)^p\Big)^{1/p} \leq C'\Big(\sum_{j\in A}(2^{jp/2}\sigma_j)^2\Big)^{1/2}. \tag{57}$$

then again we have point-wise, for all $\Lambda \subset \mathbb{N} \times \mathbb{Z}$:

$$\Big(\sum_{\kappa \in \Lambda} |\sigma_j 2^{j/2} h_\kappa(x)|^p \Big)^{1/p} \leq C' \Big(\sum_{\kappa \in \Lambda} |\sigma_j 2^{j/2} h_\kappa(x)|^2 \Big)^{1/2}$$

so in this case

$$\|(\sum_{\kappa \in \Lambda} |\sigma_j 2^{j/2} h_\kappa(x)|^p)^{1/p}\|_p \leq C \|(\sum_{\kappa \in \Lambda} |\sigma_j 2^{j/2} h_\kappa(x)|^2)^{1/2}\|_p$$

using (57) for $p > 2$:

$$\frac{1}{C'} \|(\sum_{\kappa \in \Lambda} |\sigma_j 2^{j/2} h_\kappa(x)|^p)^{1/p}\|_p \leq \|(\sum_{\kappa \in \Lambda} |\sigma_j 2^{j/2} h_\kappa(x)|^2)^{1/2}\|_p \leq \|(\sum_{\kappa \in \Lambda} |\sigma_j 2^{j/2} h_\kappa(x)|^p)^{1/p}\|_p.$$

Now we will extend this result to a general wavelet using the *Transfer* lemma (below). For any locally measurable function let us recall the definition of the Hardy-Littlewood maximal function, let I denote an interval of $\mathbb{R}$ and $|I|$ its Lebesgue measure, for all $x \in \mathbb{R}$,

$$f^*(x) = \sup_{I, x \in I} \frac{1}{|I|} \int_I |f(y)| dy.$$

Lemma 2. *(Transfer) Let us consider two sequences of functions $(f_n(x))_{n \in \mathbb{N}}$, and $(e_n(x))_{n \in \mathbb{N}}$. Suppose that the sequence $(f_n(x))_{n \in \mathbb{N}}$ satisfies the Temlyakov property and that there exists $A < \infty$ such that for all $n \in \mathbb{N}$*

$$|f_n(x)| \leq A\, e_n^*(x) \text{ a.e. and } |e_n(x)| \leq A\, f_n^*(x) \text{ a.e.}$$

Then the sequence $(e_n(x))_{n \in \mathbb{N}}$ satisfies also the Temlyakov property.

Theorem 2 follows from the lemma since for $f = 1_{[0,1]}$, $f^*(x) \asymp c(1 \wedge \frac{1}{|x|})$, and obviously, for all $x \in \mathbb{R}$, $|f(x)| \leq f^*(x)$. Combining this with the assumption of theorem 2 we have $|\psi(x)| \leq C h^*(x)$ and $|h(x)| \leq C \psi^*(x)$. □

To complete the proof we derive the lemma.

Proof. The key tool for deriving the *Transfer* lemma is the Fefferman-Stein inequality : for all $1 < p < \infty$, $1 < q \leq \infty$ there exists $C_{p,q} < \infty$, such that

$$\|(\sum_n |f_n(x)|^q)^{1/q}\|_p \leq \|(\sum_n (f_n^*)^q(x))^{1/q}\|_p \leq C_{p,q} \|(\sum_n |f_n(x)|^q)^{1/q}\|_p.$$

Using our assumption and the previous inequality we have for all $1 < q \leq \infty$ and for all $\Lambda \subset \mathbb{N}$

$$\|(\sum_{n \in \Lambda} |f_n(x)|^q)^{1/q}\|_p \leq A \|(\sum_{n \in \Lambda} |e_n^*(x)|^q)^{1/q}\|_p \leq A C_{p,q} \|(\sum_{n \in \Lambda} |e_n(x)|^q)^{1/q}\|_p$$

$$\leq A^2 C_{p,q} \|(\sum_{n \in \Lambda} |f_n^*(x)|^q)^{1/q}\|_p \leq A^2 C_{p,q}^2 \|(\sum_{n \in \Lambda} |f_n(x)|^q)^{1/q}\|_p$$

So for all $1 < q \leq \infty$

$$\|(\sum_{n \in \Lambda} |f_n(x)|^q)^{1/q}\|_p \asymp \|(\sum_{n \in \Lambda} |e_n(x)|^q)^{1/q}\|_p$$

So using the previous computation for $q = 2$ and $q = p$, we have

$$\|(\sum_{n \in \Lambda} |f_n(x)|^p)^{1/p}\|_p \asymp \|(\sum_{n \in \Lambda} |f_n(x)|^2)^{1/2}\|_p$$

and so

$$\|(\sum_{n \in \Lambda} |e_n(x)|^p)^{1/p}\|_p \asymp \|(\sum_{n \in \Lambda} |e_n(x)|^2)^{1/2}\|_p$$

□

6.3 Diophantine approximation

We recall some basic properties of continued fractions, referring to Lang (1966) and Khinchin (1992) ([L] and [K] respectively, below) for further details. Any real number a that is not an integer may be uniquely determined by its continued fraction expansion

$$a = [a_0; a_1, a_2, \ldots] = a_0 + \cfrac{1}{a_1 + \cfrac{1}{a_2 + \cfrac{1}{a_3 + \ldots}}}$$

where a_0 is an integer and $a_1, a_2, \ldots$ are strictly positive integers. The convergents $p_k(a)/q_k(a)$, $k = 0, 1, ...$ of a are those rational numbers whose continued fraction expansion terminate at stage k, thus $p_0/q_0 = a_0, p_1/q_1 = a_0 + 1/a_1, p_2/q_2 = a_0 + 1/(a_1 + 1/a_2)$, and so on. The convergents have the property of *Best approximation*: for $n \geq 1$,

$$\inf_{1 \leq k \leq q_n} ||ka|| = |q_n a - p_n| = ||q_n a||, \tag{58}$$

We recall the *law of formation* of the convergents:

$$q_n = a_n q_{n-1} + q_{n-2} \quad , \quad p_n = a_n p_{n-1} + p_{n-2}$$

from which follow some basic properties of the convergents of all irrational numbers a:

(i) The denominators q_n grow at least geometrically:

$$q_{n+i} \geq 2^{(i-1)/2} q_i, \quad i = 2k+1, k > 0 \tag{59}$$

$$q_{n+i} \geq 2^{i/2} q_i, \quad i = 2k, k > 0 \tag{60}$$

(ii) For all $n \geq 0$, $a_n < q_n/q_{n-1} \leq a_n + 1$

We say that an irrational number a is Badly Approximable (BA) if $\sup_n a_n(a) < \infty$. The set of all BA's contains quadratic irrationals (e.g. $\sqrt{5}$).

Acknowledgements. This project began while Gérard Kerkyacharian and Dominique Picard visited the University of Sydney, partly funded by the University of Sydney.

References

Bertero, M. & Boccacci, P. (1998), *Introduction to Inverse Problems in Imaging*, Institute of Physics, Bristol and Philadelphia.

Buckheit, J., Chen, S., Donoho, D. & Johnstone, I. (1995), 'Wavelab reference manual', *Stanford University* . http://www-stat.stanford.edu/ wavelab/.

Donoho, D. L., Johnstone, I. M., Kerkyacharian, G. & Picard, D. (1995), 'Wavelet shrinkage: Asymptopia?', *Journal of the Royal Statistical Society, Series B* **57**, 301–369. With Discussion.

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

Galatsanos, N. & Katsaggelos, A. (1992), 'Methods for hoosting the regularisation parameter and estimating the noise variance in image processing and their relation', *IEEE Trans. on Image Proc.* **1**, 322–336.

Harsdorf, S. & Reuter, R. (2000), 'Stable deconvolution of noisy lidar signals'. http://las.physik.uni-oldenburg.de/projekte/earsel/4th_workshop_paper/harsdorf.pdf.

Jain, A. (1989), *Fundamentals of Digital Image Processing*, Prentice Hall.

Je Park, Y., Whoe Dho, S. & Jin Kong, H. (1997), 'Deconvolution of long-pulse lidar signals with matrix formulation', *Applied Optics* **36**, 5158–5161.

Johnstone, I. M. & Raimondo, M. (2002), Periodic boxcar deconvolution and diophantine approximation, Technical Report SRR02-006, Mathematical Sciences Institute. the Australian National University.

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

Khinchin, A. Y. (1992), *Continued Fractions*, Dover. Original english version, U. Chicago Press, 1964.

Kolaczyk, E. (1993), Wavelet methods for the inversion of certain homogeneous linear operators in the presence of noisy data. PhD dissertation. Department of Statistics, Stanford University, Stanford.

Lang, S. (1966), *Introduction to Diophantine Approximations*, Springer-Verlag.

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

Neelamani, R., Choi, H. & Baranuik, R. (2000), 'Wavelet-based deconvolution for ill-conditioned systems'. http://www-dsp.rice.edu/publications/pub/neelsh98icassp.pdf.

Pensky, M. & Vidakovic, B. (1999), 'Adaptive wavelet estimator for nonparametric density deconvolution', *Annals of Statistics* **27**, 2033–2053.