INSTITUT DE STATISTIQUE

UNIVERSITÉ CATHOLIQUE DE LOUVAIN

DISCUSSION PAPER

0117

SMOOTH DESIGN-ADAPTED WAVELETS FOR NONPARAMETRIC STOCHASTIC REGRESSION

V. DELOUILLE, J. SIMOENS and R. von SACHS

http://www.stat.ucl.ac.be

Smooth Design-adapted Wavelets for Nonparametric Stochastic Regression.

V. Delouille[1,4] J. Simoens[2] R. von Sachs[3,4]

May 16, 2001

Abstract

In the setting of nonparametric stochastic regression, we introduce a new way to build smooth design-adapted wavelets. Starting from the Unbalanced Haar basis, we use the lifting scheme framework to build improved biorthogonal filters. A weighted average interpolation scheme allows us to construct wavelets with a higher number of vanishing analysing moments. We include a step which stabilizes the transform by local semi-orthogonalisation. The achievement of this article is to provide a uniform solution to the usual criticisms of wavelet estimators. Indeed, our transform automatically adapts to the nature of the regression problem, that is, to the irregularity of the design, to data on the interval, and to an arbitrary sample size (which does not need to be a power of two). We propose a wavelet thresholding algorithm and show its numerical performance both on real data and simulations including white, correlated and heteroscedastic noise.

Keywords: biorthogonal wavelet transform, heteroscedastic data, irregular design, lifting scheme, nonstationary subdivision, weighted average-interpolation.

1 Introduction

Nonparametric curve estimation by wavelets has been treated in a large number of papers in various different set-ups. These are reaching from the simple Gaussian i.i.d. error situation to more complicated data structures which have often called for a specific algorithm tailored to the problem at hand. Among these wavelet estimators, nonlinear thresholding schemes [20, 18, 19, 17] have been proved powerful and conceptually simple in order to have estimators which automatically adapt to spatially varying degrees of regularity of the underlying function to estimate. In practical terms, this offers the possibility to reconstruct regions of the function with possible discontinuities (jumps or cusps) as well as regions of higher smoothness, without needing to solve a local smoothing parameter problem, as it is the case in kernel estimation, for example.

[1]Research Assistant of the National Fund for Scientific Research – Wallonie, Belgium (F.N.R.S.). Université catholique de Louvain, Institut de Statistique, Louvain-la-Neuve, Belgium

[2]Research Assistant of the Fund for Scientific Research – Flanders, Belgium (F.W.O.). Katholieke Universiteit Leuven, Department of Computer Science, Celestijnenlaan 200A, B-3001 Leuven, Belgium

[3]Professor. Université catholique de Louvain, Institut de Statistique, Louvain-la-Neuve, Belgium

[4]Financial support from the contract 'Projet d'Actions de Recherche Concertes' no. 98/03-217 of the Belgian Government is gratefully acknowledged.

Despite these crucial advantages, the limitations of the first generation wavelet algorithms are well known. Indeed, the simplicity of the 'first generation' wavelet algorithms has to be given up when, for example, the design on which the regression curve is observed is no more fixed and equidistant, or the sample size (which is needed to be high in comparison with, e.g., kernel methods) is no more a power of two. Moreover, when the data are observed on a finite interval, a non-trivial adjustment to the presence of the boundaries is necessary.

The objective of this article is to propose a uniform solution to all these criticisms by providing a new nonlinear wavelet denoising algorithm in the situation of stochastic design regression. In contrast to most existing works [2, 4, 3, 15, 27, 30, 31] on wavelets for non-equidistant design, our wavelet method directly adapts to the design at hand, and as such avoids any pre-processing steps, such as binning for example.

To be more specific, we treat in this article the following model:

$$Y_i \;=\; f(X_i) \;+\; \epsilon_i \,,\; i = 1, \ldots, n \;, \tag{1.1}$$

where the X_i's are random variables, (Y_i, X_i) are independent pairs of observations, and the errors ϵ_i are independent of the X_i's with $E(\epsilon_i) = 0, \mathrm{Var}(\epsilon_i) < \infty$. Note that we do not assume homoscedasticity of the errors. Our target function is the conditional mean $f(x) = E[Y|X = x]$. This set-up offers also the possibility to treat nonlinear autoregression, by setting $X_i := Y_{i-1}$, and to estimate models like, e.g. autoregressive threshold or nonparametric ARCH models.

The main idea behind our procedure is to build a wavelet basis of the weighted space $\mathbf{L}_2(d\widehat{F}_n)$, where $\widehat{F}_n$ is the empirical measure associated to the stochastic design given by the $\{X_i\}_{i=1}^n$. To achieve this goal, we use the framework of the *lifting scheme* [36, 37]. More precisely, starting from a preliminary non-smooth Haar-like wavelet basis of $\mathbf{L}_2(d\widehat{F}_n)$ [22, 14], we improve this basis with the help of a *weighted average-interpolation* subdivision scheme. This allows us to construct new wavelet filters having a certain number of analysing vanishing moments. In this process, orthogonality has to be given up in favour of biorthogonality, and this may cause some stability problems, in particular near the boundaries of the observation interval. Hence, in a second step, we stabilise the transform by making it closer to an orthogonal one.

With this construction we therefore address the two following questions:

1. We furnish a smoothing method of curves which adapts automatically to whatever type of design (irregular, stochastic or even autoregressive) and to small and non-dyadic sample sizes. As a wavelet method, this procedure handles easily the presence of discontinuities in the regression function. Moreover, our estimator is adapted to the interval, no particular treatment of the boundary is necessary. This is the first and foremost achievement of this article.

2. Our second achievement is to show that the method proposed in [34, 33] is particularly efficient to improve the stability of our transform. This effect of the stabilization shows up in particular near the boundary. It thus becomes even more crucial for *small* sample sizes to have a transform that behaves well near the boundary. Indeed, for $n \leq 100$, say, a non-negligible part of the wavelet coefficients are influenced by the presence of the boundary and will benefit from the stabilizing update step. Compare our example, shown in Figure 8, where we succeeded in treating a data set with only 83 observations.

We finally compare in slightly more detail our new methodology with existing approaches. In contrast to our algorithm, almost all existing methods use some preliminary step to get back to the equidistant design situation and then apply the traditional discrete orthogonal wavelet transform algorithm. Methods to achieve this include projecting the data [8], interpolation [23, 27, 15], or approximating the scaling function by a recursion equation [5]. These approaches are somewhat hybrid and have to deal with an additional approximation error due to this first pre-processing step. The method of [9] avoids this but only treats the case of a random *uniform* design.

In the class of wavelet-type estimators for irregular designs which do not use some kind of pre-processing, we name [25], where the lifting scheme used is based on interpolation rather than average-interpolation. This construction differs from ours in that it does not start from an orthonormal wavelet basis such as the Unbalanced Haar basis developed in [22] and used in [14] for regression purpose. Using biorthogonal wavelets, the authors in [25] were confronted to the same type of stability problem as described above and proposed some solutions in [38]. In contrast, our numerical examples did not show instabilities as severe as the ones in [25].

The remaining part of this paper is organized as follows. After some preliminaries given in Section 2, the notion of design-adapted wavelets is introduced in Section 3, together with one example, the Unbalanced Haar wavelets. The way to improve this first basis, using a weighted average-interpolating scheme and a stabilizing update step is explained in Section 4. A wavelet thresholding scheme that can handle heteroscedastic noise is proposed in Section 5 and its performance is tested on simulated data sets in Section 6. Finally, Section 7 gives two applications of our wavelet algorithm on real data.

2 Preliminaries

Consider a *multiresolution analysis* (*MRA*) in a general $\mathbf{L}_2$ space, i.e. a strictly increasing and dense sequence $\mathcal{V} := \{V_j\}_{j\geq j_0}$ of closed subspaces of $\mathbf{L}_2$,

$$V_j \subset V_{j+1}\,,\ j \in \mathbb{Z}\,, \quad \text{and} \quad \operatorname{clos} \bigcup_{j=j_0}^{\infty} V_j = \mathbf{L}_2 \ \text{ for any } \ j_0 \ \in \ \mathbb{N}$$

Between successive spaces in $\mathcal{V}$, construct algebraic complements W_j so that

$$V_{j+1} = V_j \oplus W_j \ ,$$

where '$\oplus$' denotes the inner sum of disjoint linear spaces. The complement space W_j is not necessarily orthogonal to V_j. A fine-resolution space V_j can then be written as a telescopic decomposition into a coarser-resolution space and intermediate complement spaces,

$$V_j = V_{j_0} \oplus \bigoplus_{i=j_0}^{j-1} W_i \ . \tag{2.1}$$

With the notational convention that $W_{j_0-1} := V_{j_0}$, we call the sequence $\{W_j\}_{j\geq j_0-1}$ a *multiscale decomposition* (*MSD*).

The spaces V_j and W_j are equipped with bases,

$$\begin{aligned} V_j &= \operatorname{clos}_{\mathbf{L}_2} \operatorname{span} \Phi_j, \quad \Phi_j := \{\, \varphi_{jk} \mid k \in \mathcal{K}_j \,\} \ , \\ W_j &= \operatorname{clos}_{\mathbf{L}_2} \operatorname{span} \Psi_j, \quad \Psi_j := \{\, \psi_{jm} \mid m \in \mathcal{M}_j \,\} \ , \end{aligned}$$

with $\mathcal{K}_j$ and $\mathcal{M}_j$ some index sets; on the interval we use ranges of integers. We refer to any basis Φ_j for the space V_j as a set of *scaling functions*, and any basis Ψ_j for any type of complement space W_j, $j \geq 0$, is called a set of *wavelets* at level j.

Due to the multiresolution structure of the spaces V_j and W_j, there exist refinement coefficients $\{h_{jlk}\}$ and $\{g_{jlm}\}$ such that

$$\varphi_{jk} = \sum_{l \in \mathcal{K}_{j+1}} h_{jlk}\, \varphi_{j+1,l} \ , \quad \psi_{jm} = \sum_{l \in \mathcal{K}_{j+1}} g_{jlm}\, \varphi_{j+1,l} \ . \tag{2.2}$$

In the classical situation where the design points are equispaced, all sequences of refinement coefficients h_{l-2k} and g_{l-2k} will be the same for all scales and locations. In the case of irregularly spaced data however, we need to allow the refinement mask to depend both on location and scale. This releases some degrees of freedom that will be further used to adapt the transform to the irregular mesh. Building wavelets on an irregular mesh is a particular instance of the so-called second generation wavelets.

It is convenient to rewrite the refinement relations (2.2) as matrix expressions

$$\Phi_j = \Phi_{j+1} H_j \quad \text{and} \quad \Psi_j = \Phi_{j+1} G_j \ . \tag{2.3}$$

If a_j is the vector of coefficients of an element of V_j in the basis Φ_j, then the vector of coefficients of the same element in the basis Φ_{j+1} is given by $a_{j+1} = H_j a_j$. When repeated, this is a (nonstationary) *subdivision scheme*. See [13, 28] for a discussion of nonstationary subdivision schemes.

In this paper, we will use biorthogonal bases for our regression purpose. A biorthogonal basis is a generalization of an orthogonal basis, in the sense that two different bases are used for analysis and synthesis of the signal. More precisely, a primal MSD, generated by the basis $\Phi_{j_0} \cup \bigcup_{j=j_0}^{\infty} \Psi_j$ is used to reconstruct the signal, whereas a so-called *dual* MSD, with spaces $\tilde{V}_{j_0}$ and $\tilde{W}_j$ spanned by $\tilde{\Phi}_{j_0}$ and $\tilde{\Psi}_j$, respectively, is used for decomposition. Both MSD's must be linked through biorthogonality relations, i.e.,

$$\begin{aligned} \left\langle \varphi_{jk} \,, \tilde{\varphi}_{jk'} \right\rangle &= \delta_{kk'} \ \forall\, j \ , & \left\langle \varphi_{jk} \,, \tilde{\psi}_{jm'} \right\rangle &= 0 \ \forall\, j, k, m' \ , \\ \left\langle \psi_{jm} \,, \tilde{\varphi}_{jk'} \right\rangle &= 0 \ \forall\, j, m, k' \ , & \left\langle \psi_{jm} \,, \tilde{\psi}_{jm'} \right\rangle &= \delta_{mm'} \ \forall\, j \ , \end{aligned} \tag{2.4}$$

where $\delta_{kk'}$ is the Kronecker symbol. If there exists a dual MSD satisfying (2.4), the resulting pair of biorthogonal wavelet bases allows the following decomposition of any function f.

$$f(x) = \sum_k \langle f \,, \tilde{\varphi}_{j_0,k} \rangle\, \varphi_{j_0,k}(x) + \sum_{j,k} \langle f \,, \tilde{\psi}_{jm} \rangle\, \psi_{jm}(x). \tag{2.5}$$

The best constants c and C for which

$$c\, \|d\|_{\ell_2} \leq \|f\|_{\mathbf{L}_2} \leq C\, \|d\|_{\ell_2}$$

with $d := \{d_{jm}\}$, $d_{j_0,m} := \langle f \,, \tilde{\varphi}_{j_0,m} \rangle$ and $d_{jm} := \langle f \,, \tilde{\psi}_{jm} \rangle$, $j > j_0$, give the *condition number* $\kappa := C/c$ of the basis.

The dual refinement operators $\tilde{H}_j$ and $\tilde{G}_j$ are defined analogously to (2.3):

$$\tilde{\Phi}_j = \tilde{\Phi}_{j+1} \tilde{H}_j \quad \text{and} \quad \tilde{\Psi}_j = \tilde{\Phi}_{j+1} \tilde{G}_j \ . \tag{2.6}$$

Filling in the refinement relations in (2.4) gives the conditions to have *biorthogonal filters*

$$\begin{aligned} \tilde{H}_j^* H_j &= I \ , & \tilde{G}_j^* H_j &= 0 \ , \\ \tilde{H}_j^* G_j &= 0 \ , & \tilde{G}_j^* G_j &= I \ . \end{aligned} \tag{2.7}$$

Similarly to [33], we define the order of a MRA in the second-generation setting as follows.

Definition 2.1. *The* (polynomial) order *of a univariate MRA is given by*

$$\tilde{N} := \max\{\, n \mid \exists j_\star \text{ such that } \forall j \geq j_\star : \Pi_n \subset V_j \,\} \ ,$$

where Π_n *is the space of polynomials of degree at most* $n-1$.

When the order of a MRA is $\tilde{N}$ and an associated MSD exists, the analysing wavelets $\tilde{\psi}_{jm}$ with $j \geq j_\star$ have $\tilde{N}$ vanishing moments; in other words, the space $\tilde{W}_j$ is orthogonal to $\Pi_{\tilde{N}}$ for $j \geq j_\star$. Symmetrically, the primal wavelets are said to have N *vanishing moments* if, for j greater than a given level j', W_j is orthogonal to Π_N. In order to have a sparser representation of a smooth function, only $\tilde{N}$ needs to be increased. Hence in Sections 6 and 7, we use biorthogonal wavelets with $\tilde{N} = 3$ and $N = 1$. The property of vanishing moments of $\tilde{\Psi}_j$ being completely missing in the levels that are strictly smaller than $j_\star$, one should not discard in a denoising process the wavelet coefficients related to $\tilde{\Psi}_j$ for j smaller than $j_\star$. In other words, it will be reasonable to consider in the denoising algorithm of Sections 5 a primary resolution level j_0 larger or equal to $j_\star$.

3 Design-adapted wavelets

3.1 Nonparametric regression with design-adapted wavelets

Wavelets are traditionally built to form an orthonormal or biorthogonal basis in $\mathbf{L}_2(\mathbb{R})$ [11, 12]. However, when dealing with a stochastic design in a nonparametric regression context, we would like to build a wavelet basis (and hence a wavelet-based estimator) which automatically adapts to the design at hand. To achieve this goal, we need to consider a more general measure than the Lebesgue measure. That is, we construct in this paper some biorthogonal bases on a weighted space $\mathbf{L}_2(I, \Sigma, \mu)$, where I is a subset of $\mathbb{R}$, Σ is the σ-field of I and μ is a σ-finite measure on Σ. This measure μ can be very general; in particular it can be purely non-atomic (i.e. it does not contain point masses), purely atomic (it contains only point masses), or a mixture of both situations.

The first element needed to build wavelets adapted to a weighted space (I, Σ, μ) is a partitioning of the interval $I \in \mathbb{R}$ we are working on. This concept of *partitioning* is defined in [22] and recalled below. It can be thought of as the replacement of the dyadic intervals on the real line.

Definition 3.1. *A partitioning on* $I \in \mathbb{R}$ *is a subset of intervals* $I_{jk}, j \in \mathbb{Z}, k \in \mathbb{Z}$ *such that:*

(P1) $I = \cup_k \{I_{jk} | j \text{ fixed}, k \in \mathbb{Z}\}, \ \forall \ j;$

(P2) $I_{j,k_1} \cap I_{j,k_2} = \emptyset \ \forall k_1 \neq k_2$ *and* $\forall j$;

(P3) $I_{jk} = I_{j+1,l} \cup I_{j+1,l+1}$ *or* $I_{jk} = I_{j+1,l}$ *for some* $l \in \mathbb{Z}$;

(P4) $\sigma(I_{jk} | j \in \mathbb{Z}, k \in \mathbb{Z}) = \widetilde{\Sigma}$, *where* $\widetilde{\Sigma}$ *is the sub-σ-field of* Σ *such that the μ-completion of* $(I, \widetilde{\Sigma})$ *is* (I, Σ).

To present the idea behind design-adapted wavelets, consider the nonparametric regression model with stochastic design given by equation (1.1). In the simulation study of Section 6, we consider three types of noise: white, correlated and heteroscedastic. Let F_x denote the cumulative distribution function of the regressors $\{X_i\}$. In this paper, this function is supposed to be continuous.

By 'design-adapted' wavelets, we mean wavelets which form a basis of the space $\mathbf{L}_2(dF_x)$, where dF_x is the Borel-Stieltjes measure associated to F_x. In this space, the measure of an interval $[a, b)$ is given by

$$F_x(b) - F_x(a) = P\left(X_i \in [a, b)\right);\ i = 1, \ldots, n.$$

In words, we take into account in the measure of the space the information contained in the distribution of the regressors.

In case where F_x is strictly increasing on I (that is the density of the regressors is bounded away from zero), the $\sigma-$field $\widetilde{\Sigma}$ of Definition 3.1 equals the Borel field on I, i.e. $\mathcal{B}(I)$. To proceed, we need to specify which partitioning to use. Two possibilities have to be considered, depending on whether or not F_x is known.

If the distribution F_x is considered to be known, it is possible to build what we call a *quantile partitioning*, denoted by I°_{jk}, where the endpoints of the intervals are the quantiles of the distribution F_x. Here and in the following we choose as usual $F_x^{-1}(u) = \inf\{x :\ F_x(x) \geq u\}$ as the quantile function. By continuity of F_x, we always have $F_x(F_x^{-1}(u)) = u$. The elements of the partitioning are then defined as

$$I^\circ_{jk} = \left[F_x^{-1}\left(\frac{k-1}{2^j}\right),\ F_x^{-1}\left(\frac{k}{2^j}\right)\right)\ , \tag{3.1}$$

where $j = 1, \ldots, J_1,\ \ k = 1, \ldots, 2^j$; with J_1 denoting the finest scale considered in the decomposition. In this scheme, the measure of an interval $I^o_{j,k}$ is simply equal to 2^{-j}, that is the points are equispaced in $L_2(dF_x)$. For this partitioning, condition *(P4)* in Definition 3.1 is fulfilled with $\widetilde{\Sigma} = \mathcal{B}(\mathbb{R})$.

In practice however, we only have information about the empirical distribution $\hat{F}_n$ of the given realisation $X_1, \ldots, X_n$. Note that even for the purely atomic measure associated to $\hat{F}_n$, the condition *(P4)* in Definition 3.1 does hold. This allows the construction of biorthogonal wavelet bases in $\mathbf{L}_2(d\hat{F}_n)$ described below. Given the realisation $X_1, \ldots, X_n$, we can define the *empirical quantile partitioning*, noted I_{jk}, where we use the empirical quantiles instead of the theoretical quantiles in order to define the endpoints of the intervals. If the sample size n is a power of two, $n = 2^J$, then the random intervals I_{jk} are determined by the data as:

$$\begin{aligned} I_{jk} = \left[X_{((k-1)l_j+1)},\ X_{(kl_j+1)}\right)\ ,\ \ k = 1, \ldots, 2^j\ , l_j = n2^{-j}, \\ j_0 \leq j \leq J = \log_2(n) \end{aligned} \tag{3.2}$$

where $X_{(1)} \leq \ldots \leq X_{(n)}$ denote the order statistics of the design variables and, as a convention, $X_{(n+1)} := X_{(n)} + (X_{(n)} - X_{(1)})/n$. Similarly to the quantile partitioning case, the *empirical* measure $\hat{\mu}_n$ of the random interval I_{jk} is given by $\hat{\mu}_n(I_{jk}) = 2^{-j}$, $k = 1, \ldots, 2^j$. When n is not a power of two, the construction of the empirical quantile partitioning is generalized as follows. If at a given level $j+1$ we have an odd number $2p+1$ of intervals, then the first $2p$ intervals will produce p intervals at the next level j, as in condition *(P3)* in Definition 3.1. The last interval $I_{j+1,2p+1}$ will be passed unchanged to the level j, see Figure 1.

The next section presents a first example of an orthonormal wavelet basis on a weighted space.

3.2 Unbalanced Haar wavelets

In [22], Girardi and Sweldens introduced the Unbalanced Haar basis. Here we instantiate their very general construction to Euclidean spaces $\mathbf{L}_2(\mu)$. Given a partitioning $\{I_{jk}\}$ of I

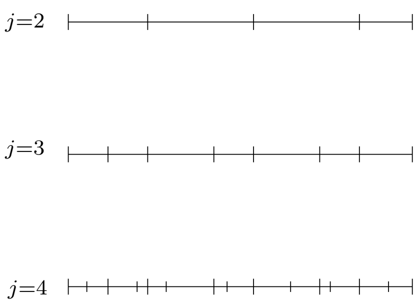


Figure 1: Coarsening procedure when n is not a power of two. At the level $j = 4$, we combine the 14 intervals two by two to produce 7 intervals at level $j = 3$. The first 6 intervals are further combined to produce the first three intervals at the coarser level, the last interval being passed unchanged to the level $j = 2$.

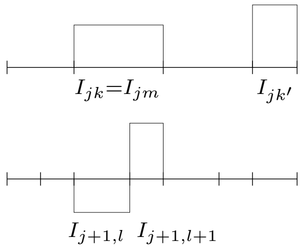


Figure 2: Two scaling functions and one wavelet in the Haar MSD on irregular meshes on the interval. Above, the mesh at level j is drawn, and below, the next finer mesh at level $j+1$. Unlike the interval $I_{jk'}$, the interval I_{jk} is split in two subintervals at level $j+1$, so besides a scaling function there is also a wavelet associated with it.

and a measured space $\mathbf{L}_2(\mu)$, the Unbalanced Haar scaling functions are defined as:

$$\check{\varphi}_{jk}(x) := \frac{1}{\mu(I_{jk})^{\frac{1}{2}}} 1_{jk}(x) \tag{3.3}$$

and the corresponding wavelets are given by

$$\check{\psi}_{jm}(x) := \frac{1}{\mu(I_{jm})^{\frac{1}{2}}} \left(\sqrt{\frac{\mu(I_{j+1,l})}{\mu(I_{j+1,l+1})}} 1_{j+1,l+1}(x) - \sqrt{\frac{\mu(I_{j+1,l+1})}{\mu(I_{j+1,l})}} 1_{j+1,l}(x) \right), \tag{3.4}$$

where 1_{jk} denotes the indicator function of the interval I_{jk} and where the index m is used to denote the location of intervals I_{jm} such that $I_{jm} = I_{j+1,l} \cup I_{j+1,l+1}$ for some l. On such intervals, it is possible to construct a wavelet $\check{\psi}_{jm}$ as is shown in Figure 2. Scaling and wavelets functions at level j are conveniently indexed by associating each of them to an interval in the partitioning, so that $\mathcal{M}_j \subseteq \mathcal{K}_j$. In our setting, $\mathcal{K}_j$ and $\mathcal{M}_j$ contain the indices of the locations of scaling and wavelet functions, respectively, present at scale j. Girardi and Sweldens proved in [22] that these wavelets $\{\check{\psi}_{jm}\}$ form an orthonormal basis of the space $\mathbf{L}_2(I, \Sigma, \mu)$ and have the ability to generate a multiresolution analysis, the order of this MRA as defined in Definition 2.1 being equal to 1.

Suppose the regressors $X_1, \ldots, X_n$ are coming from an unknown distribution, and consider the empirical quantile partitioning $\{I_{jk}\}$ defined in (3.2). If n is a power of two, equations (3.3) and (3.4) then reduce to a form similar to the classical Haar wavelets:

$$\begin{aligned}\check{\varphi}_{jk}(x) &= 2^{j/2}1_{jk}(x), \;\; j = j_0, \ldots, J, \;\; k = 1, \ldots, 2^j, \\ \check{\psi}_{jm}(x) &= \frac{1}{\sqrt{2}}\left(\check{\varphi}_{j+1,2m}(x) - \check{\varphi}_{j+1,2m+1}(x)\right), \\ j &= 1, \ldots, J-1, k = 1, \ldots, 2^j \;.\end{aligned} \tag{3.5}$$

The corresponding Haar filters associated to $\check{\Phi}_j$ and $\check{\Psi}_j$ are denoted by $\check{H}_j$ and $\check{G}_j$, respectively. When n is not a power of two, the coarsening is done as described in Section 3.1. The Haar basis is then given by the general equations (3.3) and (3.4), where the measure μ is replaced by the empirical measure $\hat{\mu}_n$. This measure is computed as $1/n$ times the number of data points falling into the interval,

$$\hat{\mu}_n(I_{jk}) = \frac{\sum_{i=1}^n 1_{jk}(X_i)}{n} \,.$$

We define the empirical wavelet coefficients of this last basis as

$$\widehat{d}_{j,m} = \frac{1}{n}\sum_{i=1}^{n} Y_i \check{\psi}_{j,m}(X_i), \qquad \widehat{s}_{j,k} = \frac{1}{n}\sum_{i=1}^{n} Y_i \check{\varphi}_{j,k}(X_i). \tag{3.6}$$

For $\alpha-$Hölder continuous functions, with $\frac{1}{2} < \alpha < 1$, the nonlinear threshold estimator based on the Haar wavelet basis built from the empirical quantile partitioning has been proved to have a near-optimal rate of convergence of the minimax $\mathbf{L}_2-$risk, see [14].

4 Wavelets built with weighted average-interpolation

4.1 The lifting scheme

The previous section gave a very simple instance of a design-adapted wavelet basis. The lifting scheme [37] can be used to modify the set of refinement matrices, yielding higher regularity and a higher order for the primal MRA, while maintaining $\mathbf{L}_2(\mu)$-biorthogonality.

The observation behind the lifting scheme is that if (H_j^o, G_j^o) and $(\tilde{H}_j^o, \tilde{G}_j^o)$ are an initial set of biorthogonal filter pairs, then for any P_j, new pairs of biorthogonal filters can be found as

$$(H_j := H_j^o + G_j^o P_j \,, \; G_j := G_j^o) \quad \text{and} \quad (\tilde{H}_j := \tilde{H}_j^o \,, \; \tilde{G}_j := \tilde{G}_j^o - \tilde{H}_j^o P_j^*) \,, \tag{4.1}$$

with P_j^* denoting the Hermitian conjugate of P_j. It suffices to check that the biorthogonality relations (2.7) are still satisfied. In this operation, one obtains new primal refinement matrices H_j, while the dual refinement matrices $\tilde{H}_j^o$ remain unchanged. The H_j define new primal scaling functions as limit functions under subdivision. Convergence is not guaranteed for all choices of the P_j, but if the new primal scaling functions exist then by (2.3) and (2.6) new primal and dual wavelets exist, and of course have the same smoothness as their corresponding scaling functions.

Naturally, the roles of the primal and the dual side in (4.1) can be interchanged, giving for any operator U_j

$$(H_j := H_j^o \,, \; G_j := G_j^o - H_j^o U_j) \quad \text{and} \quad (\tilde{H}_j := \tilde{H}_j^o + \tilde{G}_j^o U_j^* \,, \; \tilde{G}_j := \tilde{G}_j^o) \,. \tag{4.2}$$

The operations (4.1) and (4.2) are commonly called *predict* and *update* lifting steps, respectively.

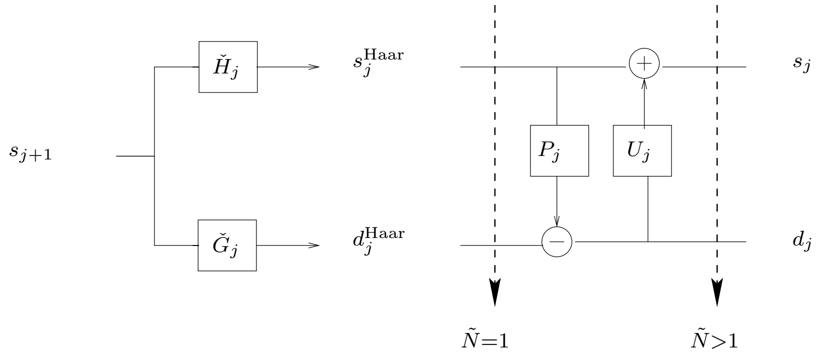


Figure 3: Filter bank diagram for our analysing wavelet transform. A Haar transform is performed first, producing detail and scaling coefficients corresponding to the initialisation wavelet basis of order $\tilde{N} = 1$. This basis is further improved by two lifting steps: the predict operator P_j produces detail coefficients with better approximation properties, the update operator U_j stabilizes the transform.

Our particular construction starts from the Haar basis and comprises a predict step followed by an update step. This yields the analysing wavelet transform as represented in Figure 3. The predict step aims at providing a new primal MRA with an higher order $\tilde{N}$. This is achieved through average-interpolation. The update step, on the other hand, is designed to improve the stability, that is, to reduce the condition number κ of the wavelet basis.

4.2 Average-Interpolation

A collection of functions $\{\varphi_k\}$ is *average-interpolating* on a partition $\{I_k\}$ if

$$\int_{I_k} \varphi_{k'} \, d\mu = \mu(I_k)^{\frac{1}{2}} \, \delta_{k,k'} \ .$$

The scaling factor is included in order to make this notion equivalent to the property that $\{\varphi_k\}$ is biorthogonal to the set of normalized indicator functions on the same partition, that is, to the set of Haar scaling functions defined on $\{I_k\}$.

Scaling functions $\{\varphi_{jk} \mid j_0 \leq j \leq J\}$ that are levelwise biorthogonal to the Haar scaling functions $\{\check{\varphi}_{jk}\}$ can be recovered by a single lifting operation from the Haar bases as follows. Assume we have at our disposal a refinement matrix H_j corresponding to Φ_j. If H_j and $\check{H}_j$ are biorthogonal, satisfying $\check{H}_j^* H_j = I$, using the orthogonality of the Haar filters $(\check{H}_j, \check{G}_j)$ yields

$$H_j = \begin{bmatrix} \check{H}_j & \check{G}_j \end{bmatrix} \begin{bmatrix} \check{H}_j^* \\ \check{G}_j^* \end{bmatrix} H_j = \begin{bmatrix} \check{H}_j & \check{G}_j \end{bmatrix} \begin{bmatrix} I \\ \check{G}_j^* H_j \end{bmatrix} = \check{H}_j + \check{G}_j (\check{G}_j^* H_j) \ . \tag{4.3}$$

Hence, a predict step from the Haar basis with $P_j = \check{G}_j^* H_j$ gives $\{\varphi_{jk}\}$ as new primal scaling functions. Moreover, once we have the predict operator P_j, using the relationships (4.1), (2.3) and (2.6) we find the corresponding primal and dual wavelet bases.

It remains to construct refinement matrices H_j corresponding to average-interpolating scaling functions on the interval. This construction is well known. The average-interpolating subdivision algorithm of Donoho [16] is easily generalized to irregular partitions [37] and to the present weighted irregular case. We present this algorithm in its general form and then explain how it translates to the particular weighted space $\mathbf{L}_2(d\hat{F}_n)$.

4.2.1 Average-interpolating subdivision in general

In an average-interpolating subdivision of order p, each finer-level coefficient $a_{j+1,l}$ is calculated from a set of $p = 2d + 1$ consecutive coarse-level coefficients:

$$\begin{array}{ccccc} a_{j,k(l)-d} & \cdots & a_{j,k(l)} & \cdots & a_{j,k(l)+d} \\ & \searrow & \downarrow & \swarrow & \\ & & a_{j+1,l} & & \end{array}$$

where $k(l)$ is chosen for each l.

Since scaling functions are associated with intervals, symmetry leads us to consider only odd orders. To find $a_{j+1,l}$, first set up the polynomial $\lambda_{j,k(l)}$ of degree $p-1$ such that the averages of $\lambda_{j,k(l)}$ over the intervals $I_{j,k(l)+i}$ correspond to the values $a_{j,k(l)+i}$,

$$\langle \lambda_{j,k} \, , \check{\varphi}_{j,k+i} \rangle = a_{j,k+i} \ , \quad -d \leq i \leq d \ , \tag{4.4}$$

with $k = k(l)$. Whenever possible the associated intervals are chosen symmetrically around $I_{j+1,l}$, i.e., $I_{j,k(l)} \supset I_{j+1,l}$. Close to the boundary this cannot always be satisfied, so that the coarse-level intervals have to be chosen asymmetrically around $I_{j+1,l}$. Then use the average over $I_{j+1,l}$ as the value for $a_{j+1,l}$,

$$a_{j+1,l} := \left\langle \lambda_{j,k(l)} \, , \check{\varphi}_{j+1,l} \right\rangle \ . \tag{4.5}$$

It is easily checked that the above procedure does indeed give average-interpolating scaling functions. As the Haar scaling functions are normalized indicator functions, one has that $\langle \sum_{l \in \mathcal{K}_{j+1}} a_{j+1,l} \, \check{\varphi}_{j+1,l} \, , \check{\varphi}_{jk} \rangle = a_{jk}$, and this implies the same property for all $\sum_{l \in \mathcal{K}_i} a_{il} \, \check{\varphi}_{il}$, $i > j$ and indeed for the subdivision limit function $\tilde{f} = \lim_{i \to \infty} \sum_{l \in \mathcal{K}_i} a_{il} \, \check{\varphi}_{il}$. The scaling function φ_{jk} is the limit function under subdivision from to the Kronecker sequence $\{\delta_{\cdot - k}\}$.

As mentioned earlier, one of the purposes of the lifting operation is to provide the new primal MRA with a higher order. The primal MRA being of order $\tilde{N}$ is equivalent to the property that for any polynomial $m \in \Pi_{\tilde{N}}$, and for $j \geq j_\star$ there is some sequence a_j which produces by subdivision the polynomial m. For $a_{jk} = \langle m \, , \check{\varphi}_{jk} \rangle$ with $m \in \Pi_p$, the polynomial λ_{jk} in (4.4) coincides with m for all k. Then also all subsequent coefficients a_{il}, $i > j$, equal $\langle m \, , \check{\varphi}_{il} \rangle$, and the subdivision process converges to m. Therefore the new primal MRA has order p.

As a remark, note that at the coarsest levels in the mesh where $\#\mathcal{K}_j < p$, the p coarse-level coefficients needed to perform order-p average-interpolating subdivision are not available, so that the order of average-interpolation needs to be lowered. This determines the level $j_\star$ in Definition 2.1.

4.2.2 Average-interpolating wavelet transform in $\mathbf{L}_2(d\hat{F}_n)$

When considering the empirical measure, the inner products of the previous section become sums. For any monomial x^q,

$$\langle x^q \, , \check{\varphi}_{jk} \rangle_{d\hat{F}_n} = \frac{1}{n} \sum_{X_i \in I_{jk}} \frac{X_i^q}{\hat{\mu}_n(I_{jk})^{\frac{1}{2}}} \ .$$

We now have all the tools needed to build a fast biorthogonal wavelet transform for use in nonparametric stochastic regression. We give a detailed description of the analysing wavelet transform that corresponds to a MRA of order $\tilde{N} = 2d + 1$ for a given natural d.

For each level $j = J-1, \ldots, j_0$ first perform a Haar transform, yielding the vectors s_j^{Haar} and d_j^{Haar}, then for each $l \in \mathcal{K}_{j+1}$ proceeds as follows.

1. Let $I_{j,k(l)}$ be the interval closest to $I_{j+1,l}$ for which $I_{j,k(l)-d}$ and $I_{j,k(l)+d}$ exist. Away from the boundary, this will give $I_{j,k(l)} \supset I_{j+1,l}$.

2. Compute $a_{j+1,l} = \sum_{k=k(l)-d}^{k(l)+d} h_{jlk}\, s_{jk}^{\mathrm{Haar}}$ satisfying (4.5). This is done by solving the $\tilde{N} \times \tilde{N}$ system

$$\begin{bmatrix} \langle x^0 , \check{\varphi}_{j,k(l)-d} \rangle_{d\hat{F}_n} & \cdots & \langle x^0 , \check{\varphi}_{j,k(l)+d} \rangle_{d\hat{F}_n} \\ \vdots & & \vdots \\ \langle x^{\tilde{N}-1} , \check{\varphi}_{j,k(l)-d} \rangle_{d\hat{F}_n} & \cdots & \langle x^{\tilde{N}-1} , \check{\varphi}_{j,k(l)+d} \rangle_{d\hat{F}_n} \end{bmatrix} \begin{bmatrix} h_{j,l,k(l)-d} \\ \vdots \\ h_{j,l,k(l)+d} \end{bmatrix} = \begin{bmatrix} \langle x^0 , \check{\varphi}_{j+1,l} \rangle_{d\hat{F}_n} \\ \vdots \\ \langle x^{\tilde{N}-1} , \check{\varphi}_{j+1,l} \rangle_{d\hat{F}_n} \end{bmatrix} .$$

The coefficients h_{jlk} give the elements of the new refinement matrix H_j.

3. By (4.3), the predict operator is found as $P_j = \check{G}_j^* H_j$ with $\check{G}_j$ the Haar high-pass filter, which simply takes the difference between two successive coefficients. Therefore, applying the P_j to s_j^{Haar} gives the sequence p_j, $p_{jk} = a_{j+1,l+1} - a_{j+1,l}$, where $I_{jk} = I_{j+1,l} \cup I_{j+1,l+1}$. The new detail coefficients d_{jk} are then computed as $d_{jk} = d_{jk}^{\mathrm{Haar}} - p_{jk}$.

For example, if $\tilde{N} = 3$, quadratic polynomials will belong to V_j for $j \geq j^*$, j^* being as in Definition 2.1. Indeed, in this case, $a_{j+1,l+1}$ will be exactly equal to $s_{j+1,l+1}^{\mathrm{Haar}}$ and hence the details coefficients are zero:

$$d_{jk} = (s_{j+1,l+1}^{\mathrm{Haar}} - s_{j+1,l}^{\mathrm{Haar}}) - (a_{j+1,l+1} - a_{j+1,l}) = 0 \; .$$

Figure 3 shows a filter bank diagram for the wavelet transform with predict step.

4.3 The Update Step

After the average-interpolating prediction step, the spaces V_j and W_j are no longer orthogonal. We may improve the stability of the new wavelet basis by performing an update step as described in [34].

The most intuitive stabilisation is probably an orthogonalisation of each wavelet ψ_{jm}^{new} with respect to V_j by subtracting its orthogonal projection onto V_j,

$$\psi_{jm}^{\mathrm{new}} = \psi_{jm} - P_{V_j} \psi_{jm} \, .$$

If $\tilde{N}$ is the order of the primal MRA, then the new wavelets have $\tilde{N}$ vanishing moments. However, they no longer have local support, since in general the projection onto V_j is globally supported. Therefore one might consider a local orthogonalisation, i.e. an orthogonalisation not with respect to the whole of V_j but only to a chosen finite set of scaling functions $\Phi_{jm} := \{\varphi_{jk}\}$ close to ψ_{jm}, that is, with $|k - m|$ small enough. One obtains

$$\psi_{jm}^{\mathrm{new}} = \psi_{jm} - P_{V_{jm}} \psi_{jm} \, ,$$

with $V_{jm} := \operatorname{span} \Phi_{jm}$. This preserves the local support property, but all vanishing moments the old wavelets may posses will be lost.

The fact that at least one vanishing moment is required for stability of a wavelet basis leads us to a constrained version of local orthogonalisation. After the predict step described above, the primal wavelets have one vanishing moment. This is due to biorthogonality and the property that the predict step leaves the dual scaling functions unchanged. We wish to preserve this property. The update operation we use is such that the angle between the new wavelets and the span of a finite set of close-by scaling functions is maximised under the constraint that the new wavelets still have one vanishing moment. With $V_{jm}^{+} := V_{jm} \cap \Pi_1^{\perp}$ the space of linear combinations from Φ_{jm} belonging to $\Pi_1^{\perp}$, that is, having a vanishing weighted mean, this reduces to

$$\psi_{jm}^{\text{new}} = \psi_{jm} - P_{V_{jm}^{+}} \psi_{jm} \,,$$

an operation which is slightly more expensive than local orthogonalisation.

5 Nonlinear threshold estimator

5.1 Introduction

The previous section described a wavelet transform that automatically adapts to the stochastic design at hand. Given a vector of regressors X, let $\mathcal{W}$ be the corresponding matrix that transforms the observations Y into a sequence of detail coefficients $\widehat{d}_{J-1}, \ldots, \widehat{d}_{j_0}$ plus a vector of coarse level coefficients $\widehat{s}_{j_0}$ In this section, we investigate through a simulation study the performance of our method for three error models. In the first model, the errors $\epsilon_i^{(1)}$'s are i.i.d. $N(0, \sigma^2)$. In the second, the $\epsilon_i^{(2)}$'s follow a stationary dependent $AR(1)$ model, $\epsilon_i^{(2)} = 0.5\, \epsilon_{i-1}^{(2)} + e_i$, e_i i.i.d.$N(0,1)$. In the third model, the errors are heteroscedastic: $\epsilon_i^{(3)} = \sigma(X_i) e_i$, with e_i i.i.d.$N(0,1)$.

Consider for a moment the heteroscedastic model, which we rewrite in vector form $Y = f(X) + \epsilon$, where $Y = [Y_1, \ldots, Y_n]^T$, $f(X) = [f(X_1), \ldots, f(X_n)]^T$, $\epsilon = [\epsilon_1, \ldots, \epsilon_n]^T$. In the wavelet domain, we have

$$\mathcal{W}Y = \mathcal{W}f(X) + \mathcal{W}\epsilon \,, \tag{5.1}$$

where $\mathcal{W}Y$ is the vector containing the estimated scaling and wavelet coefficients in Mallat's ordering, $\mathcal{W}Y = [\hat{s}_{j_0}, \hat{d}_{j_0}, \ldots, \hat{d}_{J-1}]^T$. For thresholding, we do not need to consider the variance of $\mathcal{W}f(X)$ in (5.1), but only the variance of the noise component $\mathcal{W}\epsilon$. Let Z be the matrix of ϵ. In general, Z is unknown. Then

$$\mathrm{Var}(\mathcal{W}\epsilon) = \mathcal{W}Z\mathcal{W}^T \tag{5.2}$$

To find an adequate threshold, we will need to estimate this variance. To this end, we suppose that Z is a diagonal matrix. This assumption is not met by the $\epsilon_i^{(2)}$ in the second model, but we will see that even for small positive correlation between the errors, our algorithm still works reasonably well. For negative correlation, the situation is as easy as in the case of i.i.d. errors, since the 'parasite trend' in the errors has a zero slope. In case of highly correlated errors, a non negligible parasite trend appears, which our algorithm can not remove.

We explore in the next sections the properties of two denoising schemes. One is optimal only when the errors are i.i.d., the other is tailored to the case of heteroscedastic errors. In the following, we use a noise-proportional thresholding rule,

$$t_{jk} = \sigma_{jk}\tau, \tag{5.3}$$

where σ_{jk}^2 is the variance of the wavelet coefficient given the value of the regressors, $\sigma_{jk}^2 = \mathrm{Var}(\hat{d}_{jk}|X)$. We first describe how to estimate σ_{jk}, then we indicate a possible choice for τ.

5.2 Estimation of the variance of the coefficients

When the errors are i.i.d. $N(0, \sigma^2)$, the variance (5.2) simplifies to

$$\mathrm{Var}(\mathcal{W}\epsilon) = \sigma^2 \mathcal{W}\mathcal{W}^T.$$

Hence a natural way of estimating σ^2 is to consider a modified version of estimating the median absolute deviation from the median (MAD), similarly to [26]. Let ω_{J-1} be the elements of the diagonal of $\mathcal{W}\mathcal{W}^T$ corresponding to the wavelet coefficients d_{J-1}. The variance of the element of $\mathcal{W}\epsilon$ that corresponds to the detail coefficient $\widehat{d}_{jk}$ can be estimated as $\hat{\sigma}^2_{\mathrm{MAD},jk}$ where

$$\begin{aligned} \hat{\sigma}^2_{\mathrm{MAD},jk} &= \hat{\sigma}^2_{\mathrm{MAD}}\, \mathrm{diag}(\mathcal{W}\mathcal{W}^T)_{jk} \\ \hat{\sigma}_{\mathrm{MAD}} &= \mathrm{MAD}(\frac{\hat{d}_{J-1}}{\sqrt{\omega_{J-1}}})/0.6745\,, \end{aligned} \tag{5.4}$$

In case of heteroscedastic noise, let $\hat{\sigma}^2(X)$ be an estimator of the variance function $\sigma^2(X)$. Then Z is estimated as the diagonal matrix $\hat{Z} = \mathrm{diag}(\hat{\sigma}^2(X_1), \ldots, \hat{\sigma}^2(X_n))$ and

$$\widehat{\mathrm{Var}}(\mathcal{W}\epsilon) = \mathcal{W}\hat{Z}\mathcal{W}^T \tag{5.5}$$

$$\{\hat{\sigma}^2_{jk}\} = \mathrm{diag}(\mathcal{W}\hat{Z}\mathcal{W}^T)\ . \tag{5.6}$$

Note that in case the wavelet transform $\mathcal{W}$ does not perform any update step, we have an analytic expression for $\widehat{\sigma}^2_{jk}$:

$$\widehat{\sigma}^2_{jk} = \frac{1}{n^2} \sum_{l, \tilde{g}_{j,l,m} \neq 0} \sum_{X_i \in I_{j+1,l}} \frac{\widehat{\sigma}^2(X_i)\ \tilde{g}^2_{j,l,m}}{\hat{\mu}_n(I_{j+1,l})}\ .$$

5.3 Choice of τ

We now need to specify which type of threshold to use. A whole theory has been developed in the case of orthogonal transforms to find adequate thresholding rules that fulfill some given requirements like, e.g., minimizing the unbiased estimate of the risk (SureShrink estimator [19]), minimizing the risk defined in a Bayesian framework [1] or providing a reconstruction as smooth as the underlying function (VisuShrink estimator [18, 17]).

In our case however, it is necessary to take into account the biorthogonal aspect of the transform. Hence we adopt the thresholding rule proposed by Berkner *et al.* [6] for a nonorthogonal transform. The authors derive a thresholding scheme that generalizes the VisuShrink (or universal threshold) estimator of Donoho and Johnstone [18] by incorporating a measure of the correlations between biorthogonal wavelet coefficients. More precisely, the threshold τ is chosen as

$$\tau := \sqrt{2(1+\delta_{j_0})\log n}\ ,$$

where $\delta_{j_0} \in [0,1]$ is defined as the bound on the cross-correlation between wavelet coefficients up to level j_0, i.e.

$$\delta_{j_0} := \max_{j \geq j_0, j' \geq j_0, j \neq j', k, k'} (|\mathrm{Corr}(\hat{d}_{jk}, \hat{d}_{j',k'})|) \tag{5.7}$$

In order to compute δ_{j_0}, we must first evaluate the correlation matrix of $\mathcal{W}\epsilon$

$$\mathrm{Corr}(\mathcal{W}\epsilon) = D^{-\frac{1}{2}}\mathcal{W}\hat{Z}\mathcal{W}^T D^{-\frac{1}{2}}\ , \tag{5.8}$$

where D is the diagonal matrix containing the diagonal elements of $\mathcal{W}\hat{Z}\mathcal{W}^T$. The maximum over all elements of (5.8) gives δ_{j_0} The matrix (5.8) can be formed as the analysing transform progresses. A fast implementation is thus possible; see also [26].

5.4 One- and three-step procedures

To summarize, in the simulations of Section 6, we explore the performance of two denoising algorithms.

The first algorithm can be applied when the errors are i.i.d. Gaussian. There the variance of the noise is estimated by means of expression (5.4) and we take the modified universal threshold

$$t_{jk} = \hat{\sigma}_{\mathrm{MAD},jk}\sqrt{2(1+\delta_{j_0})\log n}$$

to denoise the function.

The second algorithm, tailored to the case of heteroscedastic errors, proceeds in three steps.

1. Obtain a pilot estimator of $f(x)$ using a linear weighted average-interpolating (WAI) wavelet estimator, where the cut-off scale j_1 is chosen as $j_1 := \lfloor \log_2(n)/2 \rfloor$. Denote this pilot estimator by $\widehat{f}_0(x)$.

2. Take the residuals $r_i := Y_i - \widehat{f}_0(X_i)$ and estimate $\sigma^2(\cdot)$ from the data set $(X_i, r_i^2), i = 1, \ldots, n$ using a *robust* linear WAI estimator. That is, we choose a cutting level $j_2 := \lfloor \log_2(n)/3 \rfloor$ and we discard the 1% of highest value of the residuals. Call this estimate $\hat{\sigma}^2(x)$.

3. Finally, estimate f by a nonlinear WAI thresholding estimator. We compute the threshold as
$$t_{jk} = \hat{\sigma}_{jk}\sqrt{2(1+\delta_{j_0})\log n}\ ,$$
where $\hat{\sigma}^2_{jk}$ and δ_{j_0} are defined in (5.6) and (5.7), respectively.

In the third step, the nonlinear wavelet estimator is:

$$\hat{f}(x) = \sum_k \hat{s}_{j_0,k}\varphi_{j_0,k}(x) + \sum_{j=j_0}^{J-1}\sum_m \eta_{(\cdot)}(\hat{d}_{jm}, t_{jm})\psi_{jm}(x)\ , \tag{5.9}$$

where $\eta_{(\cdot)}(w,\lambda)$ denotes either 'hard' or 'soft' threshold nonlinearities:

$$\eta_H(w,\lambda) = w1_{\{|w|>\lambda\}}\ ; \quad \eta_S(w,\lambda) = sgn(w)(|w|-\lambda)_+\ .$$

In the expression (5.9), the coefficients can be written as $\hat{s}_{j_0,k} = \langle Y\,, \tilde{\varphi}_{j_0,k}\rangle_{d\hat{F}_n}$ and $\hat{d}_{jm} = \left\langle Y\,, \tilde{\psi}_{jm}\right\rangle_{d\hat{F}_n}$. Here we use an average-interpolation scheme of order $p = 3$. The MRA corresponding to the wavelet transform will thus be of order $\tilde{N} = 3$. The value of $j_0 = 2$ is chosen such that $j_0 \geq j^\star$, with $j^\star$ introduced in Definition 2.1.

More precisely, in the expression (5.9), the $\tilde{\varphi}_{j_0,k}$ are the Haar scaling functions, and the $\varphi_{j_0,k}$ are smooth scaling functions with $\Pi_3 \subset V_{j_0}$. The wavelet functions ψ_{jm} have the same regularity as their corresponding scaling function, but have only one vanishing moment. Finally, the analysing wavelets $\tilde{\psi}_{jm}$ before update are piecewise constant functions with three vanishing moments. After the update, they are no longer piecewise constant, but conserve their vanishing moments. This construction corresponds to detail coefficients being zero if the original signal is a quadratic polynomial.

6 Simulation study

6.1 Parameters of the simulation study

Throughout our simulation study, the regressors $\{X_t\}$ are independent and normally distributed random variables, $X_t \sim N(0.5, (0.2)^2)$. We take four test functions: a sine function (Sine), a function having one discontinuity (Jump), another having a discontinuity in its first derivative (Cusp), and a damped sine function (Damped Sine), see Figure 4. The sample sizes selected were $n = 200$ and $n = 500$, the primary resolution level has been set to $j_0 = 2$, and we use a hard thresholding rule.

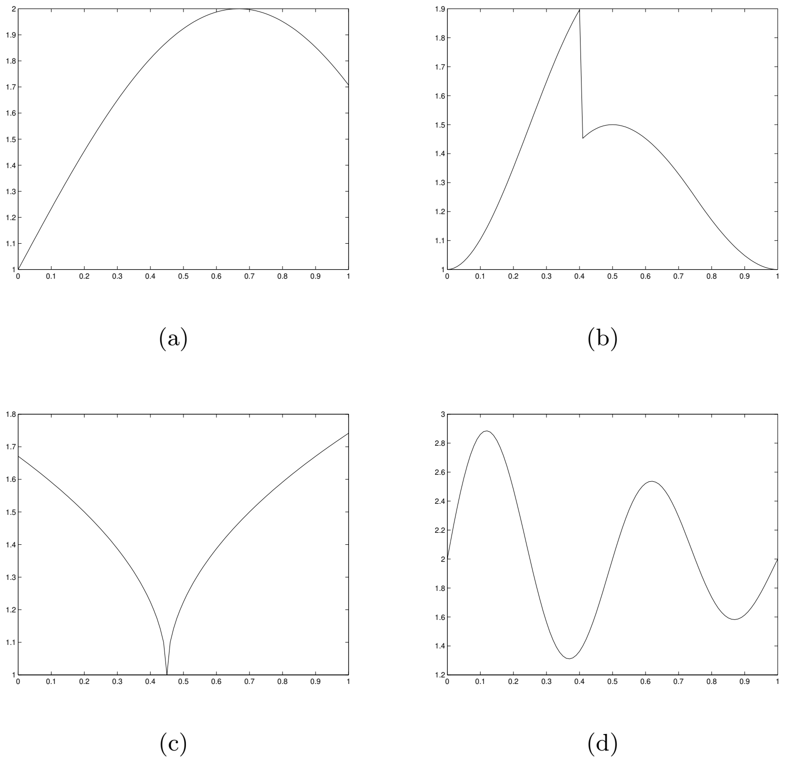


Figure 4: Test functions: (a) Sine (b) Jump (c) Cusp (d) Damped Sine

With the first algorithm, we compare our two methods (with and without stabilizing update) with the one proposed by Kovac [27, 26]. Kovac's method first computes a new data set by linear interpolation of the original data on a regular grid. A classical orthogonal wavelet transform is then performed on these new data. The detail coefficients above the primary resolution level are thresholded, taking into account the heteroscedasticity of the coefficients due to the linear interpolation step, and the estimate on the regular grid is found by inverting the wavelet transform. In order to make a fair comparison, we finally compute the value of the smoothed data on the original grid to evaluate the performance criterion defined in (6.2). For this comparison, we use the original code of Kovac, available in WaveThresh3 [29].

With the second algorithm, we show by some empirical evidence the importance of a stabilizing update step. The three error models of Section 5.1 are considered. In order to specify the variance σ_ϵ^2 of the errors, we define a signal-to-noise ratio as SNR := $\mathrm{sd}(f)/\sigma_\epsilon$, where

$$\mathrm{sd}(f)^2 = \frac{1}{n-1}\sum_{i=1}^{n}(f(x_i) - \overline{f})^2, \tag{6.1}$$

for a realization $\{x_i\}_{i=1}^n$ of the random variable $X \sim N(0.5, 0.04)$, and $\overline{f}$ is equal to the mean value of the $f(x_i)$, $i = 1, \ldots, n$. We set the SNR equal to two, which allows us to specify σ_ϵ. In the model with errors $\epsilon_i^{(1)}$, we took $\sigma := \sigma_\epsilon$. For the second model, let $\sigma_\epsilon := \mathrm{sd}(\epsilon^{(2)})$, with $\mathrm{sd}(\cdot)$ computed as in (6.1). Finally, for the heteroscedastic model, $\sigma_\epsilon := \int \sigma(x)dx$ and this integral is approximated by averaging $\sigma(x)$ over discrete x as in [21]. The chosen function $\sigma(x)$ is piecewise constant: $\sigma(x) = C(1_{\{x \leq 0.4\}}(x) + 2_{\{x>0.4\}}(x))$, where the constant C is determined to have a SNR $= 2$.

To evaluate the performance of the methods on one given data set, we use as a criterion the relative mean square error:

$$\mathrm{MSE} = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{\widehat{f}(x_i) - f(x_i)}{|f(x_i)|}\right)^2, \tag{6.2}$$

where $\widehat{f}(x_i)$ is the estimated function and $f(x_i)$ is the real underlying function computed at x_i. When computing equation (6.2) for a large number B of data sets, we need to summarize the information given by these B runs. We choose to compute the median, first and third quartiles over the B relative MSE values. We then took the square root of these three quantities (RMSE). In all cases, $B = 500$ runs were performed.

6.2 Simulation results

We now present and discuss the results of the simulation studies. In the graphs, we represent realisations where the WAI estimate with update has a value of the RMSE criterion close to the median RMSE value reported in the corresponding table.

6.2.1 Simulated data with white noise errors

Table 1 gives the results of the simulation study for the one step algorithm based on the variance $\hat{\sigma}^2_{\mathrm{MAD},jk}$ as defined in (5.4). It compares our two methods with the interpolation algorithm of Kovac [27].

We first comment on the lengths of the interquartile intervals, that is, on the robustness of the three methods with respect to both the randomness of the design and the variability of the noise part. The interquartile intervals for the algorithm with update and for the interpolation methods are in general relatively small compared to those obtained without update. Indeed, the average-interpolation method without update is sometimes less performant near the boundaries, this explains the greater variability in the values of the RMSE.

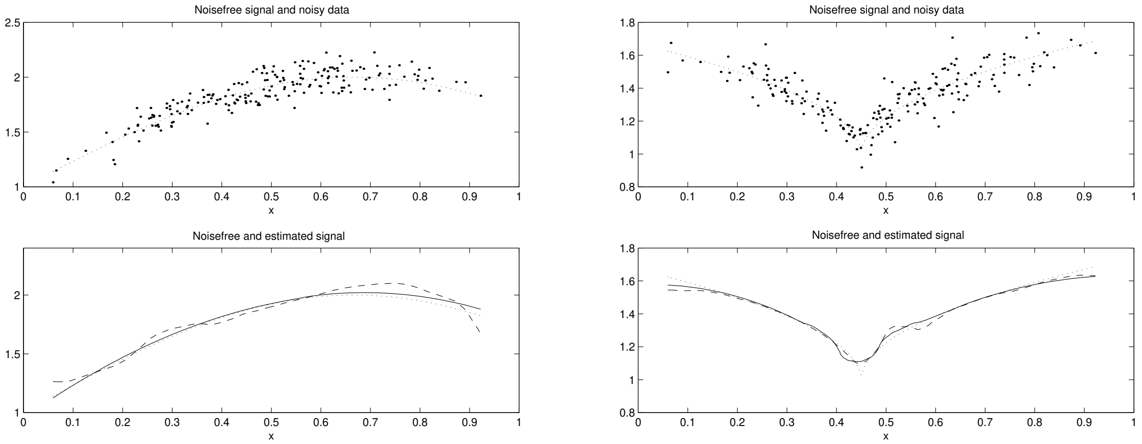
The figures in Table 1 clearly demonstrate the superiority of the average-interpolating scheme followed by a stabilizing update step, over the two other methods. In particular, the interquartile intervals of the Update and Interpolation methods never overlapped, which tends to indicate that the improvement in the estimation is significant.

The reason of this better performance with respect to the interpolation algorithm of Kovac can be explained by the fact that our design-adapted wavelet algorithm is optimal for piecewise quadratic polynomials and does not require any pre-processing step. In contrast, the Daubechies Extremal Phase wavelet basis with four vanishing moments was used in the interpolation algorithm. This combined with the pre- and post-processing steps results in a more wiggly estimator, see Figure 5. In this figure, the WAI estimators without update are not represented, since they are very close to the estimator resulting from a transform with an update step.

Moreover, like most wavelet-based estimators, our method handles the presence of a discontinuity in the function or in one of its derivatives reasonably well. See Figure 5(b) and 6 for an illustration. This illustrates our first claim.

As a remark, note that as we used a modified version of the VisuShrink estimator of Donoho and Johnstone, the cusp in Figure 5(b) is somewhat oversmoothed. A smaller value of the threshold would result in a better detection of the cusp.

The average-interpolation algorithm without update shows most of the time better results than the interpolation algorithm. However, without update, a large bias may appear near the boundaries, like, e.g., in the case of the damped sine function. This bias deteriorates the value of the RMSE criterion, see Table 1. The presence of a stabilizing update step corrects for this bias near the boundary, which illustrates our second claim expressed in the Introduction.



(a) (b)

Figure 5: Estimation of the Sine (a) and Cusp (b) functions with the one-step algorithm. Data are represented by dots and the underlying signal by a dotted line. The WAI estimator with update is represented as a plain line, the estimator from interpolation as a dashed line. The RMSE values for the WAI and interpolation methods, respectively, are equal to 0.94% and 2.92% in (a), and 1.83% and 2.00% in (b).

6.2.2 Iterative procedure to deal with heteroscedastic data

Table 2 shows the results of the simulation study using the three-step algorithm described in Section 5.4. We see that in all cases the addition of an update step improves the value of the RMSE.

For data with i.i.d. errors, the three steps procedure gives results similar to the one-step algorithm. When the errors are positively correlated, we see an increase in the RMSE values for both methods. Indeed, the dependency of the errors produces a parasite trend in the data, as can be seen in Figure 6 for $X_i \in [0.4, 0.5]$. A possible remedy to this problem is

Table 1: Results of the simulation study for white noise errors models using the one step algorithm (with and without update) and the interpolation algorithm of Kovac. For each function and each sample size, the first line gives the RMSE, i.e. the square root of the median of the relative MSE. The numbers between brackets give the square root of the inter-quartile intervals of the MSE. Values are expressed in percent.

Sine			**Cusp**		
$n = 200$			$n = 200$		
With Update	Without Update	Interp.	With Update	Without Update	Interp.
0.84	1.12	3.00	1.78	1.95	3.11
[0.69; 1.06]	[0.85; 1.52]	[2.55; 3.54]	[1.69; 1.91]	[1.78; 2.28]	[2.19; 3.58]
$n = 500$			$n = 500$		
With Update	Without Update	Interp.	With Update	Without Update	Interp.
0.63	1.04	2.50	1.58	1.77	1.93
[0.5; 0.78]	[0.73; 1.44]	[2.14; 3.15]	[1.26; 1.68]	[1.64; 1.94]	[1.79; 2.07]
Jump			**Damped Sine**		
$n = 200$			$n = 200$		
With Update	Without Update	Interp.	With Update	Without Update	Interp.
2.26	2.87	3.39	3.88	8.46	6.03
[2.12; 2.42]	[2.64; 3.27]	[3.11; 3.75]	[3.60; 4.23]	[6.13; 9.41]	[5.08; 6.91]
$n = 500$			$n = 500$		
With Update	Without Update	Interp.	With Update	Without Update	Interp.
2.5	2.79	2.90	3.06	5.98	5.01
[2.35; 2.63]	[2.53; 5.63]	[2.69; 3.08]	[2.80; 3.30]	[5.08; 7.82]	[3.83; 6.50]

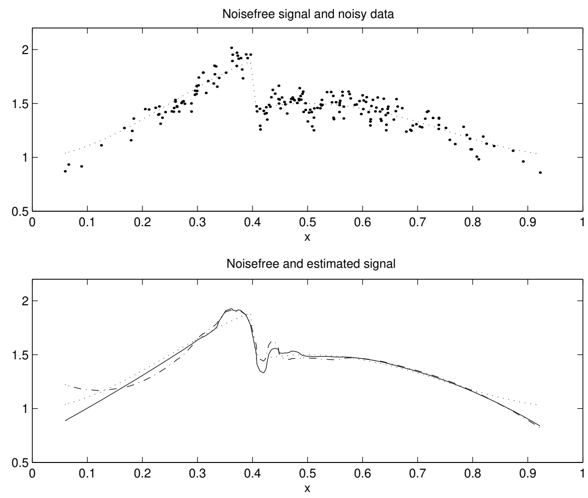


Figure 6: Estimation of the Jump function with the three-step algorithm in case of correlated errors. The dotted line represents the underlying signal. The WAI estimators with and without update are plotted in continuous and in dashed line, respectively. The RMSE equals 3.39% for the estimator with update and 4.19% without update.

testing for a serial correlation between the errors and to estimate this correlation. This can then be taken into account in the computation of $\hat{\sigma}^2_{jk}$ in (5.6). Finally, the results for the heteroscedastic data show that this algorithm can handle changes in the variance function $\sigma(x)^2$.

The comparison between both proposed methods in Table 2 shows once more the importance of the update step, which always improves the estimation near the boundary, and hence the value of the RMSE. The improvement is particularly striking for the damped sine function. We illustrate this in Figure 7, in the case of i.i.d. and heteroscedastic errors.

The results for the Cusp and Jump function in Table 2 and Figure 6 allow us to conclude that even in the presence of correlated errors or heteroscedasticity, our wavelet algorithm can correctly detect jumps or cusps in the estimated function.

7 Application on real data sets

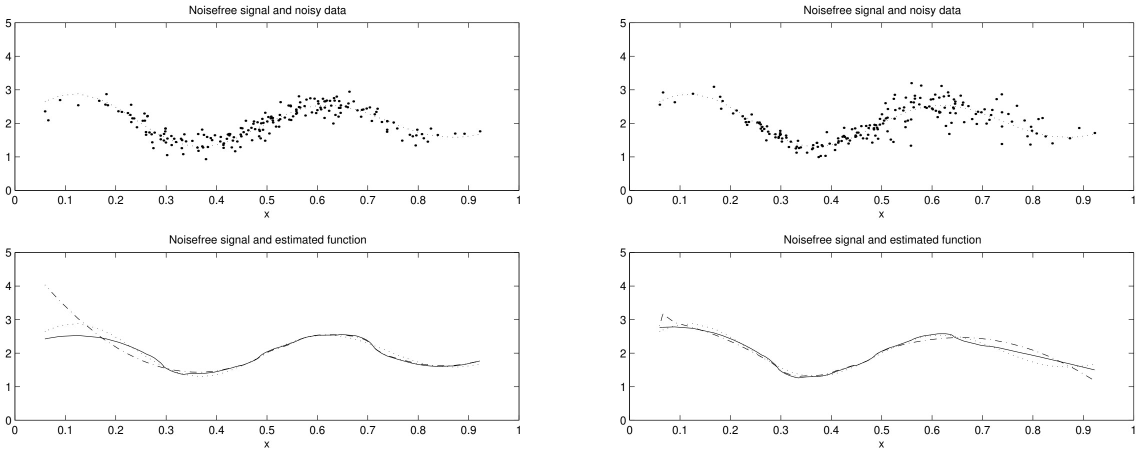
In this section, two data sets taken from real experiments are treated. In order to deal with repeated data, we replaced all observations made at a same design point by their median.

Once a threshold rule is chosen (like the modified VisuShrink rule), the only effective tuning parameter that needs to be specified in our wavelet procedure is the primary resolution level j_0, which we set equal to two, following a common practice in the literature. Note that, if the threshold is location and scale dependent, this is only due to the heteroscedasticity of the data and to the biorthogonality of the transform. Unlike the choice of a local bandwidth in kernel or local polynomial smoother, the optimal determination of a threshold t_{jk} does not involve the estimation of the smoothness of the underlying function.

We compared our wavelet estimator with a symmetric $k-$nearest neighbour linear least squares procedures, called 'Super Smoother' in Splus and implemented under the *supsmu* function. In this method, $k/2$ data points on each side of x are used in a linear regression to predict the value at x. Cross–validation is used to choose an optimal value of k for each data

Table 2: Results of the simulation study for the three error models (i.i.d, dependent, heteroscedastic), using the three-step algorithm with update (denoted U) and without update (denoted NU). For each function and each sample size, the first line presents the square root of the median of the relative MSE. The numbers between brackets give the square root of the inter-quartile intervals of the MSE. Values are expressed in percent.

Sine					
$n = 200$					
i.i.d.(U)	i.i.d.(NU)	Dep.(U)	Dep.(NU)	Hete.(U)	Hete.(NU)
0.84	1.09	2.56	3.26	1.59	2.18
[0.66; 1.06]	[8.80; 1.47]	[2.14; 3.01.]	[2.56; 4.00]	[0.93; 2.13]	[1.57; 2.78]
$n = 500$					
i.i.d.(U)	i.i.d.(NU)	Dep.(U)	Dep.(NU)	Hete.(U)	Hete.(NU)
0.61	0.99	2.44	3.01	1.17	1.74
[0.50; 0.75]	[0.69; 1.32]	[2.10; 2.77]	[2.56; 3.53]	[0.70; 1.50]	[1.36; 2.20]
Jump					
$n = 200$					
i.i.d.(U)	i.i.d.(NU)	Dep.(U)	Dep.(NU)	Hete.(U)	Hete.(NU)
2.25	2.84	3.62	4.37	2.96	3.91
[2.13; 2.46]	[2.60; 3.21]	[3.24; 3.94]	[3.86; 5.07]	[2.37; 3.48]	[3.28; 4.60]
$n = 500$					
i.i.d.(U)	i.i.d.(NU)	Dep.(U)	Dep.(NU)	Hete.(U)	Hete.(NU)
2.47	2.74	3.52	4.09	2.75	3.29
[2.32; 2.62]	[2.48; 3.50]	[3.22; 3.80]	[3.68; 4.57]	[2.48; 3.07]	[2.812; 3.73]
Cusp					
$n = 200$					
i.i.d.(U)	i.i.d.(NU)	Dep.(U)	Dep.(NU)	Hete.(U)	Hete.(NU)
1.78	1.96	2.54	3.08	1.89	2.24
[1.69; 1.91]	[1.81; 2.30]	[2.15; 2.86]	[2.49; 3.80]	[1.75; 2.27]	[1.89; 2.95]
$n = 500$					
i.i.d.(U)	i.i.d.(NU)	Dep.(U)	Dep.(NU)	Hete.(U)	Hete.(NU)
1.59	1.83	2.03	2.34	1.61	1.93
[1.26; 1.67]	[1.65; 2.00]	[1.79; 2.30]	[2.05; 2.65]	[1.46; 1.77]	[1.69; 2.18]
Damped Sine					
$n = 200$					
i.i.d.(U)	i.i.d.(NU)	Dep.(U)	Dep.(NU)	Hete.(U)	Hete.(NU)
3.93	8.53	5.34	8.44	4.62	9.54
[3.61; 4.22]	[6.09; 9.53]	[4.56; 6.13]	[6.87; 9.80]	[3.91; 6.26]	[7.24; 10.16]
$n = 500$					
i.i.d.(U)	i.i.d.(NU)	Dep.(U)	Dep.(NU)	Hete.(U)	Hete.(NU)
3.11	6.55	4.37	7.45	3.27	8.97
[2.81; 3.31]	[5.28; 9.72]	[3.75; 5.07]	[5.69; 10.14]	[2.96; 3.72]	[6.17; 9.98]



(a) (b)

Figure 7: Estimation of the Damped Sine function in case of (a) i.i.d. errors (b) heteroscedasticity. The dotted line represents the underlying signal. The WAI estimators with and without update are plotted in continuous and in dashed line, respectively. The RMSE equals 3.84% in (a), 4.11% in (b) for the estimator with update, and 7.32% in (a), 6.70% in (b) without update.

point x, see [39]. This method is thus comparable to a local linear polynomial estimator with a *local* choice of the bandwidth.

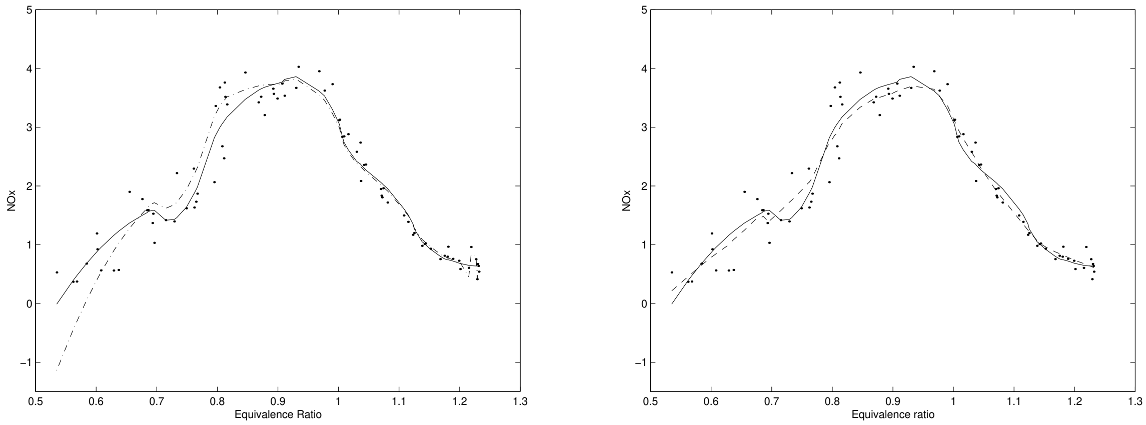
7.1 Measurement of exhaust from burning ethanol

The first data set, which has already been discussed by many authors [7, 10, 24, 26], consists of 88 measurements from an experiment in which ethanol was burned in a one-cylinder automobile engine. The concentration (labeled 'NOx') of the sum of nitric oxide (NO) and nitrogen dioxide ($N0_2$) is related to the equivalence ratio, a measure of the richness of the air/ethanol mix.

When dealing with repeated data as described above, we obtain a data set of sample size $n = 83$. Figure 8(a) shows the application of our three-step algorithm, using WAI wavelet transform with and without update. For such small sample size, the update step becomes crucial.

In Figure 8(b), we represent our wavelet estimator together with the $k-$nearest neighbour estimator as described above. We see that both resulting estimates have similar smoothness and shape. Note, however, how the wavelet estimator is sensitive to some local patterns, like what happens around $x = 0.75$.

When looking at the detail coefficients which have survived the VisuShrink threshold (for the method with update), we see that at the level j_0 all the detail coefficient are nonzero. Only one detail coefficient at level $j_0 + 1$ has survived the threshold and the levels above $j_0 + 1$ are all empty. Indeed, the detail coefficients at level j_0 give the general shape of a quadratic polynomial, and the non-zero coefficient at level $j_0 + 1$ is responsible for the small bump arising around $x = 0.75$.



(a) (b)

Figure 8: The *ethanol* data set is represented in dots. In (a) the WAI estimator with (plain line) and without (interrupted dashed line) update resulting from the three-step procedure is shown. In (b) the WAI estimator with update (plain line) is represented together with the $k-$nearest neighbour linear least squares, using cross-validation for a local choice of k (dashed line).

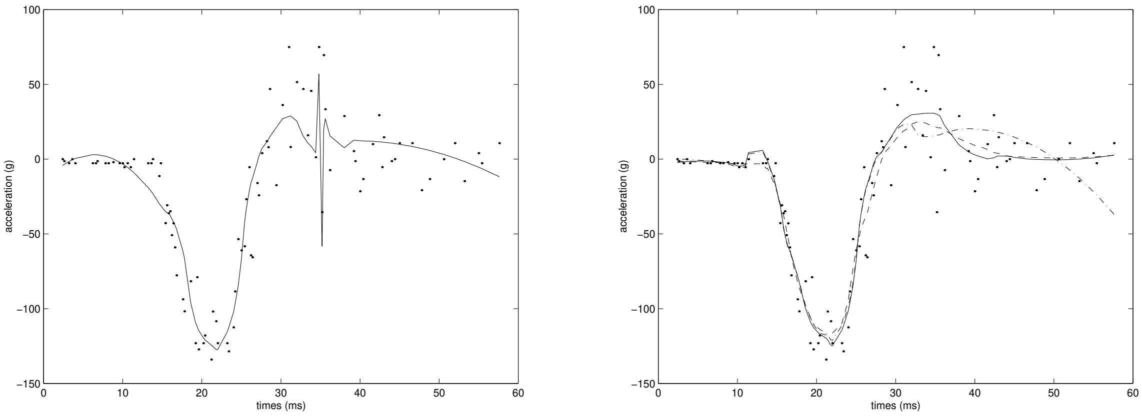
7.2 Motorcycle acceleration data

We also applied our nonparametric regression method to the motorcycle acceleration data taken from Silverman [32]. It consists of 133 observations from a crash test and shows the acceleration of some motorcyclists' head during a crash. The explanatory variable is time (in milliseconds) and the dependent variable is the head acceleration (in g). After dealing with repeated data, the effective sample size was then reduced to $n = 94$ observations.

A blind application of our three-step algorithm with update produces the curve in Figure 9(a). It exhibits a high frequency feature due to a very large variance of the data between 30 and 40 ms. Consequently, some detailed coefficients at the finest level $J-1$ have survived even the VisuShrink threshold. To solve this problem, we add a first stage in our procedure. We begin by computing a linear wavelet estimate with cutting level $j_1 = J-1$ (which removes the high frequency detail coefficients). Calling $\tilde{Y}$ the resulting smoothed data, we apply our three-step algorithm on the data set $(X, \tilde{Y})$. Figure 9(b) shows the resulting estimators, with and without update. In this Figure, we also represented the $k-$nearest neighbour estimator.

With such a small sample size, the need for an update step is again clear. It has already been discussed in the literature [35] that the underlying signal virtually exhibits two discontinuities in the first derivative, one being located around 23 ms and the other around 32 ms. Unlike the $k-$nearest neighbour estimator, both wavelet estimators do indeed seem to detect a cusp around 23 ms, see Figure 9(b).

When looking at the coefficients that have survived the thresholding procedure (for the method with update), we observe that at level j_0, all coefficients are nonzero. On level j_0+1, two coefficients out of six have survived the threshold, one of these two coefficients being responsible for the small cusp around 23 ms. The levels above j_0+1 are all empty.



(a) (b)

Figure 9: Motorcycle acceleration data are represented in dots. In (a) the WAI estimator with update using the three-step procedure is shown. In (b) we first removed all details at level $J-1$, then applied the iterative algorithm with update (plain line) and without update (interrupted dashed line). The $k-$nearest neighbour estimate is in dashed line. Both wavelet estimators seem to indicate a change point in the first derivative around 23 ms.

8 Conclusion

In this paper, we propose a wavelet regression type estimator which addresses the criticisms inherent to the first generation wavelet estimators. Our wavelet estimator provides a uniform solution in the sense that it does not need to be tuned to the situation at hand (particular design or sample size, presence of boundaries or of a discontinuity in the underlying signal). It is nevertheless comparable in performance with the most elaborate smoothing techniques used in nonparametric regression such as local linear polynomial or $k-$nearest neighbour linear least squares procedures having an automatic selection, by cross-validation, of the local 'bandwidth' parameter.

We use a thresholding rule adapted to the biorthogonality of our wavelet transform, which has proved in several numerical examples to work well in the presence of Gaussian noise, be it i.i.d., slightly correlated or heteroscedastic. Moreover, due to the presence of a stabilizing update step, datasets with sample size as small as 80 can still be processed in an efficient way, that is, the resulting estimates do not have a large bias near the boundaries.

In future work, we plan to confirm the performance of our estimator on a theoretical side. To this end, we intend to use the approach proposed by Donoho and Johnstone [18] based on the definition of an 'ideal' risk. They have shown by this investigation that wavelet estimators are never much worse than any of the estimators based on local bandwidth parameters.

Acknowledgements

The authors would like to thank M. Jansen, A. Kovac, M. Neumann and W. Sweldens for helpful discussions. V. Delouille was funded as a Research Assistant of the National Fund for Scientific Research-Wallonie, Belgium (F.N.R.S.). J. Simoens was funded as a Research

Assistant of the Fund for Scientific Research-Flanders, Belgium (F.W.O.). Partial support from the contract 'Projet d'Actions de Recherche Concertées' ARC 98/03-217 of the Belgian Government is as well gratefully acknowledged.

References

[1] F. Abramovich, T. Sapatinas, and B.W. Silverman. Wavelet thresholding via a Bayesian approach. *Computational Statistics and Data Analysis*, 22:351–361, 1996.

[2] A. Antoniadis and J. Fan. Regularization of wavelets approximations. Technical report, IMAG, Grenoble, France, 2000.

[3] A. Antoniadis, G. Grégoire, and I. McKeague. Wavelet methods for curve estimation. *Journ. Amer. Statist. Assoc*, 89:1340–1353, 1994.

[4] A. Antoniadis, G. Grégoire, and P. Vial. Random design wavelet curve smoothing. *Statistics and Probability Letters*, 35:225–232, 1997.

[5] A. Antoniadis and D.T. Pham. Wavelet regression for random or irregular design. *Computational Statistics and Data Analysis*, 28:353–369, 1998.

[6] K. Berkner and R.O. Wells. A correlation-dependent model for denoising via nonorthogonal wavelet transforms. Technical report, Department of Mathematics, Rice University, 1998.

[7] N.D. Brinkman. Ethanol - a single-cylinder engine study of efficiency in exhaust emissions. *SAE Transactions*, 90:1410–1424, 1981.

[8] T. Cai and L.D. Brown. Wavelet shrinkage for nonequispaced samples. *The Annals of Statistics*, 26(5):1783–1799, 1998.

[9] T. Cai and L.D. Brown. Wavelet regression for random uniform design. *Statistics and Probability Letters*, 42:313–321, 1999.

[10] W.S. Cleveland, E. Grosse, and W.M. Shyu. Local regression model. In J.M. Chambers and T.J. Hastie, editors, *Statistical Models in S*, pages 309–376. Wadsworth and Brooks, Pacific Grove, California, 1992.

[11] A. Cohen, I. Daubechies, and J. Feauveau. Bi-orthogonal bases of compactly supported wavelets. *Comm. Pure Appl. Math.*, 45:485–560, 1992.

[12] I. Daubechies. *Ten Lectures on Wavelets*. Philadelphia: Society for Industrial and Applied Mathematics, 1992.

[13] Ingrid Daubechies, Igor Guskov, and Wim Sweldens. Regularity of irregular subdivision. *Constructive Approximation*, 15(3):381–426, 1999.

[14] V. Delouille, J. Franke, and R. von Sachs. Nonparametric stochastic regression with design-adapted wavelets. Technical report, DP0019, Institut de Statistique, UCL, Belgium, 2000.

[15] B. Delyon and A. Juditsky. On the computation of wavelet coefficients. *Journal of Approximation Theory*, 88:47–79, 1997.

[16] D.L. Donoho. Smooth wavelet decompositions with blocky coefficient kernels. In L. Schumaker and G. Webb, editors, *Recent advances in Wavelet Analysis*, pages 259–308. Academic Press, 1993.

[17] D.L. Donoho. De-noising via soft-thresholding. *IEEE Transactions on Information Theory*, 41:613–627, 1995.

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

[19] D.L. Donoho and I.M. Johnstone. Adapting to unknown smoothness via wavelet shrinking. *Journ. Americ. Statist. Assoc.*, 90:1200–1224, 1995.

[20] D.L. Donoho and I.M. Johnstone. Minimax estimation via wavelet shrinkage. *Annals of Statistics*, 26:879–921, 1998.

[21] H.-Ye Gao. Wavelet shrinkage estimates for heteroscedastic regression models. Technical report, MathSoft, Inc. Seattle., 1997.

[22] M. Girardi and W. Sweldens. A new class of unbalanced Haar wavelets that form an unconditional basis for L_p on general measure spaces. *J. Fourier Anal. Appl.*, 3(4):457–474, 1997.

[23] P. Hall and B.A. Turlach. Interpolation methods for nonlinear wavelet regression with irregularly spaced design. *Annals of Statistics*, 25(5):1912–1925, 1997.

[24] T.J. Hastie. Generalized additive models. In J.M. Chambers and T.J. Hastie, editors, *Statistical Models in S*, pages 195–247. Wadsworth and Brooks, Pacific Grove, California, 1992.

[25] M. Jansen and A. Bultheel. Smoothing irregularly sampled signals using wavelets and cross validation. Technical report, Dept. of Computer Science, K.U.Leuven, Belgium, 1999.

[26] A. Kovac. *Wavelet Thresholding for Unequally Spaced Data*. PhD thesis, University of Bristol, March 1998.

[27] A. Kovac and B. W. Silverman. Extending the scope of wavelet regression methods by coefficient-dependent thresholding. *Journ. Amer. Statist. Assoc*, 95:172–183, 2000.

[28] David Levin. Using Laurent polynomial representation for the analysis of non-uniform binary subdivision schemes. *Advances in Comp. Math.*, 11:41–54, 1999.

[29] G.P. Nason. *WaveThresh3 Software*. Department of Mathematics, University of Bristol, Bristol, UK, 1998. Available from the StatLib archive.

[30] M. Pensky and B. Vidakovic. On non-equally spaced wavelet regression. To appear in *Annals of the Institute of Statistical Mathematics*, 2000.

[31] S. Sardy, D.B. Percival, A.G. Bruce, H.-Y. Gao, and W. Stuetzle. Wavelet de-noising for unequally spaced data. *Statistics and Computing*, 9:65–75, 1999.

[32] B.W. Silverman. Some aspects of the spline smoothing approach to non-parametric regression curve fitting. *J. Royal Statistical Society, Series B.*, 47:1–52, 1985.

[33] J. Simoens and S. Vandewalle. Average-interpolating wavelet bases on irregular meshes on the interval. Technical report, Dept. of Computer Science, K.U.Leuven, Belgium, 2000.

[34] J. Simoens and S. Vandewalle. On the stability of wavelet bases in the lifting scheme. Technical report, Dept. of Computer Science, K.U.Leuven, Belgium, 2000.

[35] P.L. Speckman. Fitting curves with features: semiparametric change-point methods. *Computing Science*, 26:257–264, 1995.

[36] W. Sweldens. The lifting scheme: A new philosophy in biorthogonal wavelet constructions. In A. F. Laine and M. Unser, editors, *Wavelet Applications in Signal and Image Processing III*, pages 68–79. Proc. SPIE 2569, 1995.

[37] W. Sweldens. The lifting scheme: A construction of second generation wavelets. *SIAM J. Math. Anal.*, 29(2):511–546, 1997.

[38] E. Vanraes, M. Jansen, and A. Bultheel. Stabilized wavelet transforms for non-equispaced data smoothing. Technical report, Dept. of Computer Science, K.U.Leuven, Belgium, 2001.

[39] W.N. Venables and B.D. Ripley. *Modern Applied Statistics with S-Plus*. Springer-Verlag, New York, 1994.