WAVELET SHRINKAGE FOR NONEQUISPACED SAMPLES

by

Professor T. Tony Cai
Purdue University
and
Professor Lawrence D. Brown
University of Pennsylvania

Technical Report #97-06

Department of Statistics
Purdue University
West Lafayette, IN 47907 USA

April, 1997

*

Wavelet Shrinkage For Nonequispaced Samples

Lawrence D. Brown
Department of Statistics, University of Pennsylvania
T. Tony Cai
Department of Statistics, Purdue University

Abstract

Standard wavelet shrinkage procedures for nonparametric regression are restricted to equispaced samples. There, data are transformed into empirical wavelet coefficients and threshold rules are applied to the coefficients. The estimators are obtained via the inverse transform of the denoised wavelet coefficients. In many applications, however, the samples are nonequispaced. It can be shown that these procedures would produce suboptimal estimators if they were applied directly to nonequispaced samples.

We propose a wavelet shrinkage procedure for nonequispaced samples. We show that the estimate is adaptive and near optimal. For global estimation, the estimate is within a logarithmic factor of the minimax risk over a wide range of piecewise Hölder classes, indeed with a number of discontinuities that grows polynomially fast with the sample size. For estimating a target function at a point, the estimate is optimally adaptive to unknown degree of smoothness within a constant. In addition, the estimate enjoys a smoothness property: if the target function is the zero function, then with probability tends to 1 the estimate is also the zero function.

Keywords: wavelets, multiresolution analysis, wavelet approximation, nonparametric regression, minimax, adaptivity, piecewise Hölder class.

AMS 1991 Subject Classification: Primary 62G07, Secondary 62G20.

1 Introduction

Suppose we are given data:

$$y_i = f(t_i) + \epsilon z_i \tag{1}$$

i = 1, 2, ..., n, $0 < t_1 < t_2 < ... < t_n = 1$, and z_i are independently and identically distributed as N(0, 1).

The function f is an unknown function of interest. We wish to estimate the function f globally or to estimate f at a point. In the case of recovering the entire function f on [0, 1], one can measure the performance of an estimate $\hat{f}$, for example, by the global squared L_2 norm risk:

$$R(\hat{f}, f) = E \int_0^1 (\hat{f}(t) - f(t))^2 \, dt$$

The goal is to construct estimates that have "small" risk. In order to have some meaningful estimate according to this criterion, one must assume certain regularity conditions on the unknown function f, such as f belongs to some Hölder classes, Sobolev classes, or Besov classes, etc.

The more traditional approaches to nonparametric regression include fixed-bandwidth kernel methods, orthogonal series methods, and linear spline smoothers. These methods are not adaptive. That is, the estimators based on these methods may achieve substantially slower rate of convergence if the smoothness of the underlying regression functions is misspecified. In recent years, more efforts have been made to develop adaptive procedures. A variety of adaptive methods have been proposed, such as variable-bandwidth kernel methods and variable-knot spline smoothers.

The recent development of wavelet bases based on the multiresolution analyses suggests new techniques for non-parametric function estimation. Wavelets offer a degree of localization both in space and in frequency. This gives great advantage over the traditional Fourier basis. In the recent few years, wavelet theory has been widely applied to the fields of signal and image processing, as well as statistical estimation.

The application of wavelet theory to the field of statistical function estimation was pioneered by David Donoho and Iain Johnstone. In a series of important papers ([11], [12], [14]), Donoho and Johnstone and co-authors present a coherent set of procedures that are spatially adaptive and near optimal over a range of function spaces of inhomogeneous smoothness. . Wavelet procedures achieve adaptivity through thresholding of the empirical wavelet coefficients. They enjoy excellent mean squared error properties when used to estimate functions that are only piecewise smooth and have near optimal convergence rates over large function classes. In contrast, traditional linear estimators typically achieve good performance only for relatively smooth functions.

Despite their considerable advantages, however, standard wavelet procedures have limitations. One serious limitation is the requirement of equispaced samples. Standard wavelet procedures are restricted to equispaced samples, i.e. t_i in (1) are equally spaced on [0, 1]. In practice, however, there are many interesting applications in statistics that the samples are not equispaced. In some wavelet software packages, non-equispaced samples are currently treated same as equispaced. As we shall explain later, non-equispaced samples should not in general be treated as equispaced. Otherwise the convergence rate could be far below the optimal rate. Different treatments are needed. So how to apply the wavelet shrinkage method to non-equispaced samples is of practical interest.

We formulate nonequispaced regression model as follows:

$$y_i = f(t_i) + \epsilon z_i \tag{2}$$

with i = 1, 2, ..., n, n = 2^J, $t_i = H^{-1}(i/n)$ for some cumulative density function H on [0, 1], and $z_i \overset{iid}{\sim}$ N(0, 1).

We develop an adaptive wavelet threshold procedure for non-equispaced model based on multiresolution analysis and projection as well as nonlinear thresholding. The algorithm for implementing the procedure has the following ingredients:

1. Precondition the data by a matrix;
2. Transform the preconditioned data by the discrete wavelet transform;
3. Denoise the noisy wavelet coefficients;

The function with the denoised wavelet coefficients is our estimate of the function f that we intend to recover. If one is interested in estimating the function at the sample points, two more steps are needed:

4. Apply the inverse wavelet transform to the denoised coefficients;
5. Postcondition the data by a matrix to get the estimate at the sample points.

Both preconditioning and postconditioning matrices are sparse matrices containing $O(n)$ nonzero entries. Comparing with Donoho and Johnstone's VisuShrink, this procedure has two additional steps, preconditioning and postconditioning, to account for the irregular spacing of the sample points. The procedure agrees with the VisuShrink when the sample is in fact equispaced.

The procedure is adaptive and near optimal. We investigate the adaptivity of the estimators over a wide range of piecewise Hölder classes, indeed with a number of discontinuities that increases polynomially fast with the sample size. We show in Section 4 that the rate of convergence for estimating regression function f globally over the function classes is a logarithmic factor away from the minimax risk. Furthermore, for estimating a target function at a point, the estimate is optimally adaptive to unknown degree of smoothness within a constant factor. The estimate also enjoys a smoothness property. If the target function is the zero function, then the estimate will also be the zero function with probability tends to 1. Therefore, the procedure removes pure noise completely with high probability.

The rest of the paper is organized as follows. Section 2 describes wavelet basis, multiresolution analysis and wavelet approximation. Section 3 introduces the nonequispaced procedure. Optimality of the estimators will be presented in Section 4. Further discussion about the procedure and related topics are given in Section 5. Section 6 contains proofs of the main results.

2 Wavelets And Wavelet Approximation

We summarize in this section the basics on wavelets and multiresolution analysis that will be needed in later sections. Further details on wavelet theory can be found in Daubechies [10] and Meyer [20].

An orthonormal wavelet basis is generated from dilation and translation of two basic functions, a "father" wavelet ϕ and a "mother" wavelet ψ. The functions ϕ and ψ are assumed to be compactly supported. Assume that $supp(\phi) = supp(\psi) = [0, N]$. Also we assume that ϕ satisfies $\int \phi = 1$. We call a wavelet ψ *r-regular* if ψ has r vanishing moments and r continuous derivatives.

Let

$$\phi_{jk}(t) = 2^{j/2}\phi(2^j t - k), \quad \psi_{jk}(t) = 2^{j/2}\psi(2^j t - k)$$

And denote the periodized wavelets

$$\phi^p_{jk}(t) = \sum_{l\in\mathcal{Z}} \phi_{jk}(t-l), \quad \psi^p_{jk}(t) = \sum_{l\in\mathcal{Z}} \phi_{jk}(t-l), \qquad \text{for } t \in [0,1]$$

For simplicity in exposition, we use the periodized wavelet bases on [0, 1] in the present paper. The collection $\{\phi^p_{j_0 k}, k = 1, ..., 2^{j_0}; \psi^p_{jk}, j \geq j_0, k = 1, ..., 2^j\}$ constitutes such an orthonormal basis of $L_2[0,1]$. Note that the basis functions are periodized at the boundary. The superscript "p" will be suppressed from the notations for convenience.

An wavelet basis has an associated multiresolution analysis on $[0,1]$. Let V_j and W_j be the closed linear subspaces generated by $\{\phi_{jk}, k = 1, ..., 2^j\}$ and $\{\psi_{jk}, k = 1, ..., 2^j\}$ respectively. Then

1. $V_{j_0} \subset V_{j_0+1} \subset \cdots \subset V_j \subset \cdots$;
2. $\cup_{j=j_0}^{\infty} V_j = L_2([0,1])$;
3. $V_{j+1} = V_j \oplus W_j$

The nested sequence of closed subspaces $V_{j_0} \subset V_{j_0+1} \subset \cdots$ is called a multiresolution analysis on $[0,1]$.

An orthonormal wavelet basis has an associated exact orthogonal Discrete Wavelet Transform (DWT) that transforms sampled data into wavelet coefficient domain. A crucial point is that the transform is not implemented by matrix multiplication, but by a sequence of finite-length filtering which produce an order $O(n)$ orthogonal transform. See Daubechies ([10]) and Strang ([21]) for further details about the discrete wavelet transform.

For a given square-integrable function f on $[0, 1]$, denote

$$\xi_{jk} = \langle f, \phi_{jk} \rangle, \qquad \theta_{jk} = \langle f, \psi_{jk} \rangle$$

So the function f can be expanded into a wavelet series:

$$f(x) = \sum_{k=1}^{2^{j_0}} \xi_{j_0 k} \phi_{j_0 k}(x) + \sum_{j=j_0}^{\infty} \sum_{k=1}^{2^j} \theta_{jk} \psi_{jk}(x) \tag{3}$$

Wavelet transform decomposes a function into different resolution components. In (3), $\xi_{j_0 k}$ are the coefficients at the coarsest level. They represent the gross structure of the function f. And θ_{jk} are the wavelet coefficients. They represent finer and finer structures of the function f as the resolution level j increases.

We note that the DWT is an orthogonal transform, so it transforms i.i.d. Gaussian noise to i.i.d. Gaussian noise and it is norm-preserving. This important property of DWT allows us to transform the problem in the function domain into a problem in the sequence domain of the wavelet coefficients with isometry of risks.

Wavelets provide smoothness characterization of function spaces. Many traditional smoothness spaces, for example Hölder spaces, Sobolev spaces and Besov spaces, can be completely characterized by wavelet coefficients. See Meyer [20]. In the present paper, we consider the estimation problem over a range of piecewise Hölder classes. Therefore, we are interested in the properties of wavelet coefficients of functions in piecewise Hölder classes.

Definition 1 *A piecewise Hölder class $\Lambda^\alpha(M, B, m)$ on $[0, 1]$ with at most m discontinuous jumps consists of functions f satisfying the following conditions:*

1. *f is bounded by B, i.e. $|f| \le B$;*
2. *There exist $l \le m$ points $0 \le a_1 < \cdots < a_l \le 1$ such that for all $a_i \le x,\ y < a_{i+1}$, $i = 0, 1, \cdots, l$ (with $a_0 = 0$ and $a_{l+1} = 1$),*

$$(i). \quad |f(x) - f(y)| \le M\,|x - y|^\alpha \qquad \text{if } \alpha \le 1;$$

$$(ii). \quad |f^{(\lfloor\alpha\rfloor)}(x) - f^{(\lfloor\alpha\rfloor)}(y)| \le M\,|x - y|^{\alpha'} \text{ and } |f'(x)| \le B \qquad \text{if } \alpha > 1$$

where $\lfloor\alpha\rfloor$ is the largest integer less than α and $\alpha' = \alpha - \lfloor\alpha\rfloor$.

In words, the function class $\Lambda^\alpha(M, B, m)$ consists of functions that are piecewise Hölder with the number of discontinuities bounded by m. In our main results, the maximum number of jump points, m, is allowed to grow polynomially fast with the sample size. This enables the function classes $\Lambda^\alpha(M, B, m)$ to model functions of significant spatial inhomogeneity. The following are the upper bounds of wavelet coefficients of functions in a piecewise Hölder classes $\Lambda^\alpha(M, B, m)$. Throughout, C denotes a generic constant not depending on function f and sample size n and the standard notation $\langle\ ,\ \rangle$ denotes inner product in L_2 space.

Lemma 1 *Let $f \in \Lambda^\alpha(M, B, m)$ and let the wavelet function ψ is r-regular with $r \geq \alpha$. Then*

(i). If $supp(\psi_{jk})$ does not contain any jump points of f, then

$$\theta_{jk} \equiv |\langle f, \psi_{jk}\rangle| \leq C \cdot 2^{-j(1/2+\alpha)} \quad (4)$$

(ii). If $supp(\psi_{jk})$ contains at least one jump point of f, then

$$\theta_{jk} \equiv |\langle f, \psi_{jk}\rangle| \leq C \cdot 2^{-j/2} \quad (5)$$

Now suppose we have a dyadically sampled function $\{f(k/n)\}_{k=1}^n$ with $n = 2^J$. We can utilize a wavelet basis and the associated multiresolution analysis to get a good approximation of the entire function f. Let us begin with the following result. The proof is straightforward.

Lemma 2 *Suppose that $f \in \Lambda^\alpha(M, B, m)$ and let $\xi_{Jk} \equiv \langle f, \phi_{Jk}\rangle$, then*

(i). If $supp(\phi_{Jk})$ does not contain any jump points of f, then

$$|n^{-1/2} f(k/n) - \xi_{Jk}| \leq C \cdot n^{-(1/2+\alpha)} \quad (6)$$

(ii). If $supp(\phi_{Jk})$ contains jump points of the function f, then

$$|n^{-1/2} f(k/n) - \xi_{Jk}| \leq C \cdot n^{-1/2} \quad (7)$$

According to this result, we may use $n^{-1/2} f(k/n)$ as an approximation of $\xi_{Jk} = \langle f, \phi_{Jk}\rangle$. It means that if a dyadically sampled function is given, we may use a multiresolution analysis to get an approximation of the projection of the function f onto subspace V_J because ξ_{Jk} are the coefficients of the projection. This in turn provides a good approximation of the entire function f. More specifically, we may use $f_n(t) = \sum_{k=1}^n n^{-1/2} f(k/n)\phi_{Jk}(t)$ as an approximation of f. Based on Lemma 1 and Lemma 2, simple calculation shows that the approximation error $\|f_n - f\|_2^2$ is in the order of $n^{-2\alpha}$ for functions in the piecewise Hölder class $\Lambda^\alpha(M, B, m)$ with fixed α, M, B and m.

3 The Nonequispaced Procedure

3.1 The Estimator

Suppose now that we observe the data $\{y_i\}$ as in (2) and we wish to recover the regression function f. Our estimation method is based on multiresolution analysis and projection method. The motivation of the method will be given in Section 3.2 from the approximation point of view. Let $\tilde{g}(t) = n^{-1/2} \sum_{i=1}^n y_i \phi_{Ji}(t)$ and let

$$\tilde{f}_J(t) = Proj_{V_J}\tilde{g}(H(t)) = n^{-1/2} \sum_{k=1}^{2^{j_0}} \tilde{\xi}_{j_0 k}\phi_{j_0 k}(t) + \sum_{j=j_0}^{J-1} \sum_{k=1}^{2^j} \tilde{\theta}_{jk}\psi_{jk}(t)$$

where

$$\tilde{\xi}_{jk} = n^{-1/2}\sum_{i=1}^{n} y_i \langle \phi_{Ji} \circ H,\ \phi_{jk}\rangle, \quad \tilde{\theta}_{jk} = n^{-1/2}\sum_{i=1}^{n} y_i \langle \phi_{Ji} \circ H,\ \psi_{jk}\rangle \tag{8}$$

We can regard $\tilde{\xi}_{j_0k}$ and $\tilde{\theta}_{jk}$ as noisy observations of the true wavelet coefficients ξ_{j_0k} and θ_{jk}. Indeed, we estimate θ_{jk} by thresholding $\tilde{\theta}_{jk}$. Let

$$\hat{\xi}_{j_0k} = \tilde{\xi}_{j_0k} \quad \hat{\theta}_{jk} = sgn(\tilde{\theta}_{jk})(|\tilde{\theta}_{jk}| - \lambda_{jk})_+ \tag{9}$$

be estimate of the wavelet coefficients of f where the threshold λ_{jk} is derived in Section 3.3. Then a soft-thresholded wavelet estimator of f is given as follows:

$$\hat{f}_n^*(t) = \sum_{k=1}^{2^{j_0}} \hat{\xi}_{j_0k}\phi_{j_0k}(t) + \sum_{j=j_0}^{J-1}\sum_{k=1}^{2^j} \hat{\theta}_{jk}\psi_{jk}(t) \tag{10}$$

Similarly, a hard-thresholded estimator can be obtained by setting the coefficients in (10) as

$$\hat{\xi}_{j_0k} = \tilde{\xi}_{j_0k} \quad \hat{\theta}_{jk} = \tilde{\theta}_{jk} I(|\tilde{\theta}_{jk}| > \lambda_{jk}) \tag{11}$$

with the same threshold λ_{jk} as in (9).

The coefficients $\hat{\xi}_{j_0k}$ contain gross structure of function f and we do not threshold these coefficients. The risk of the estimate (10) can be decomposed as approximation error and estimation error. From Theorem 1, it is easy to see that the dominant term is the estimation error. We will show in Section 4 that the estimation error is comparable to the equispaced samples and the estimate enjoys the same convergence rate as the Donoho-Johnstone's VisuShrink estimate in the equispaced case.

Remark: We consider here the case of fixed design variables t_i. The method can be extended to random designs. The case of random designs have also been studied by Hall and Turlach (1996). Their methods are based on linear interpolation.

3.2 Approximation

Let us see why the estimation method makes sense. We first consider the problem of approximating a whole function based a noiseless nonequispaced sample. Denote $\Lambda^1(h)$ the collection of Lipshitz functions f satisfying

$$|f(x) - f(y)| \le h\,|x - y|, \qquad\qquad \text{for } x,\ y \in [0,1]$$

Suppose we are given a sampled function $\{f(t_i), i = 1, 2, \cdots, n(= 2^J)\}$ with $t_i = H^{-1}(i/n)$ where H is a strictly increasing cumulative density function on [0, 1] and $H^{-1} \in \Lambda^1(h)$ for some constant h. How to approximate the function f via multiresolution analysis ?

If t_i are equispaced, it follows from Lemmas 1 and 2 that $f_n(t) = \sum_{k=1}^{n} n^{-1/2} f(t_k)\phi_{Jk}(t)$ is a good approximation. When t_i are nonequispaced, an approximation using a multiresolution analysis can be derived by the following consideration. One can first approximate

$f(H^{-1}(t))$ by $\tilde{g}_n(t) = \sum_{k=1}^{n} n^{-1/2} f(t_k)\phi_{Jk}(t)$, then use the projection of $\tilde{g}_n(H(t))$ onto multiresolution space V_J as the approximation of f. To be more specific, let

$$\xi'_{Ji} = n^{-1/2} \sum_{k=1}^{2^J} f(t_k) \langle \phi_{Jk} \circ H, \phi_{Ji} \rangle \tag{12}$$

and let

$$f_n(t) = \sum_{i=1}^{2^J} \xi'_{Ji} \phi_{Ji}(t) \tag{13}$$

be an approximation of the function f. Note that f_n is in the multiresolution approximation space V_J. An upper bound for the approximation error is shown in the following result.

Theorem 1 *Suppose that a sampled function $\{f(t_i), i = 1, 2, \cdots, n(= 2^J)\}$ is given with $t_i = H^{-1}(i/n)$ where H is a strictly increasing cumulative density function on [0, 1] with $H^{-1} \in \Lambda^1(h)$ Let the wavelet function ψ be r-regular with $r > \alpha$. Let ξ'_{Ji} and f_n be given as in (12) and (13) respectively. Then the approximation error $\|f_n - f\|_2^2$ satisfies*

$$\sup_{f \in \Lambda^\alpha(M,B,m)} \|f_n - f\|_2^2 = o(n^{-2\alpha/(1+2\alpha)}) \tag{14}$$

where the maximum number of jump discontinuities $m = Cn^\gamma$ with constants $C > 0$ and $0 < \gamma < 1/(1 + 2\alpha)$.

Theorem 1 shows that the approximation error over function class $\Lambda^\alpha(M, B, m)$ is of higher order than $n^{-2\alpha/(1+2\alpha)}$ even when the number of jump points increases polynomially with the sample size. Because the optimal convergence rate for estimating f over uniform Hölder class $\Lambda^\alpha(M, B, 0)$ under the model (2) is $n^{-2\alpha/(1+2\alpha)}$, the approximation error is smaller in order than the minimax risk for statistical estimation.

3.3 The Threshold

Approximation result (14) implies that $\tilde{\xi}_{jk}$ and $\tilde{\theta}_{jk}$ in (8) have the "correct" means. In order to make thresholding work, we need to know the noise level of each coefficient $\tilde{\theta}_{jk}$.

The function H^{-1} is strictly increasing, so H^{-1} is differentiable almost everywhere. Denote $\tilde{h}(t)$ the derivative of $H^{-1}(t)$. Then

$$0 < \tilde{h}(t) \leq h, \qquad \text{for almost all } t \in [0, 1]$$

It is easy to see from (8) that

$$\sigma_{jk}^2 \equiv var(\tilde{\theta}_{jk}) = n^{-1}\epsilon^2 \sum_{i=1}^{n} (\langle \phi_{Ji} \circ H, \psi_{jk} \rangle)^2 \leq n^{-1}\epsilon^2 \int \psi_{jk}^2(t)\tilde{h}(t)dt \equiv u_{jk}^2 \tag{15}$$

Note that the inequality in (15) is asymptotically sharp, $\sigma_{jk} \to u_{jk}$, as $n \to \infty$. We set the threshold

$$\lambda_{jk} = u_{jk} \, (2 \log n)^{1/2} \tag{16}$$

Remark: This procedure generalizes the Donoho and Johnstone's VisuShrink for equispaced samples. When the samples are in fact equispaced, i.e., when H is identity function and thus $h = 1$, then $\tilde{\xi}_{j_0 k}$ and $\tilde{\theta}_{jk}$ are discrete wavelet transform of $\{n^{-1/2} y_i\}$ and $\lambda_{jk} = \epsilon(2n^{-1}\log n)^{1/2}$. Therefore the procedure agrees with the VisuShrink when the sample is equispaced.

4 Optimality Results

In this section, we discuss the properties of the wavelet estimate (10) given in Section 3.1. We begin by showing that the estimate enjoys a smoothness property. If the target function is the zero function, then the estimate is also the zero function with high probability. Specifically we have

Theorem 2 *If the regression function is the zero function $f \equiv 0$, then there exist a sequence of constants P_n such that*

$$P(\hat{f}_n^* \equiv 0) \geq P_n \to 1, \quad as \quad n \to \infty \tag{17}$$

Therefore, with high probability, the estimate remove pure noise completely. We then prove below that the estimate enjoys near-minimaxity for global estimation and the estimate optimally adapts to unknown degree of local smoothness within a constant factor when used for estimating a function at a point.

4.1 Global Estimation

We investigate the adaptivity of the wavelet estimate constructed in Section 3.1 over a range of piecewise Hölder classes $\Lambda^\alpha(M, B, m)$, where the maximum number of jump discontinuities is allowed to increase polynomially with the sample size. This enhances the power of the function classes $\Lambda^\alpha(M, B, m)$ for modeling spatially inhomogeneous functions. We show that the estimate (10) is near optimal. The convergence rate is within a logarithmic factor of the minimax rate over a range of function classes $\Lambda^\alpha(M, B, m)$.

Theorem 3 *Suppose we observe $\{y_i,\ i = 1, 2, \cdots n(= 2^J)\}$ as in (2). Let $\hat{f}_n^*$ be either the soft-thresholded or hard-thresholded wavelet estimator of f given in (10). Suppose that the wavelet function ψ is r-regular. Then the estimator $\hat{f}_n^*$ is near-optimal:*

$$\sup_{f \in \Lambda^\alpha(M,B,m)} E\,\|\hat{f}_n^* - f\|_2^2 \leq C \cdot (\log n/n)^{2\alpha/(1+2\alpha)}(1 + o(1)) \tag{18}$$

for all $0 < \alpha < r$ and all $m \leq Cn^\gamma$ with constants $C > 0$ and $0 < \gamma < 1/(1 + 2\alpha)$.

4.2 Estimation At A Point

Theorem 3 gives the convergence rate of global estimation. Now we turn our attention to local estimation. The adaptive estimation in this case is similar to global estimation, but with a very interesting distinction. The adaptive minimax rate for estimation at a point is different from that for estimation of a whole function.

By the results of Brown and Low [5] and Lepski [19], an estimator adaptive to unknown smoothness without loss of efficiency is impossible for pointwise estimation, even when the function is known to belong to one of two Hölder classes. Therefore, local adaptation can not be achieved "for free". The minimum loss of efficiency is a $(\log n)^{2\alpha/(1+2\alpha)}$ factor for estimating a function of unknown degree of local Hölder smoothness at a point. See Brown and Low [5] and Lepski [19]. We call $(\log n/n)^{2\alpha/(1+2\alpha)}$ adaptive minimax rate. Donoho and Johnstone ([13]) discuss pointwise performance of wavelet estimate for equispaced samples. They show that the VisuShrink estimate attains adaptive minimax rate for estimating functions at a point. see [13] for details.

We show below that the estimator given in Theorem 3 attains the exact adaptive minimax rate for estimating a function in a Hölder class at a fixed point. Therefore, the estimator is optimally adaptive to unknown degree of smoothness within a constant factor. To be more precise, we have the following:

Theorem 4 *For any fixed* $t_0 \in [0,1]$, *let* $\hat{f}_n^*(t)$ *be given as in (10). Suppose the wavelet* ψ *is r-regular. Then*

$$\sup_{f\in\Lambda^\alpha(M,B,0)} E(\hat{f}_n^*(t_0) - f(t_0))^2 \leq C \cdot (\log n/n)^{2\alpha/(1+2\alpha)}(1+o(1)) \tag{19}$$

for all $0 < \alpha < r$.

We state here the result in the case of uniform smoothness without jumps for the sake of simplicity. The wavelet procedure is locally adaptive, the result hold for general piecewise Hölder classes so long as the jump points are away from a fixed neighborhood of t_0.

5 Discussion

5.1 Why not treat nonequispaced samples same as equispaced ?

The nonequispaced model (2) is reduced to equispaced model when H is the identity function. But for general H, one can still "pretend" the sample as equispaced. Let $g = f \circ H^{-1}$. Then the sample is equispaced in terms of the function g. One can use the standard wavelet shrinkage procedure to estimate g by $\hat{g}$ and then use $\hat{g} \circ H$ as an estimator of f. This is what we mean by treating nonequispaced sample as equispaced. Here the estimators do not depend on the distribution of t_i.

This type of estimators do not perform well. This can be shown by a formal calculation of asymptotic risk as well as by simulation. One can show that in many situations, the convergence rates of the estimators are suboptimal if nonequispaced samples are simply treated as equispaced. See Cai (1996) for more details.

5.2 Choice of Threshold

In (16), we set the threshold $\lambda_{jk} = u_{jk}(2\log n)^{1/2}$ where $u_{jk} = (\int \psi_{jk}^2(t)\tilde{h}(t)dt)^{1/2}$. It is clear that

$$u_{jk}^2 \leq n^{-1}\epsilon^2 h_{jk} \tag{20}$$

where $h_{jk} = ess.sup\{\tilde{h}(t) : t \in [H(2^{-j}k), H(2^{-j}(k+N))]\}$.

We may replace the threshold λ_{jk} by

$$\lambda'_{jk} = \epsilon(2h_{jk}n^{-1}\log n)^{1/2}. \tag{21}$$

The optimality results hold with λ'_{jk} as the threshold. The threshold λ'_{jk} has computational advantage over the threshold λ_{jk}.

5.3 Implementation

In this section we address the issue of numerical implementation of the procedure we propose in Section 3.1.

Let $P_H = (p_{ki})$ be a matrix with entries

$$p_{ki} = \langle \phi_{Ji} \circ H, \ \phi_{Jk} \rangle$$

In general, it is not easy to calculate the exact value of p_{ki}, because we do not have the analytic expression for the father wavelet ϕ. But based on Lemma 2, we may use $n^{-1/2}\phi_{Ji}(H(k/n)) = \phi(n\,H(k/n) - i)$ as an approximation of p_{ki}. The cascade algorithm (see Daubechies (1992)) can be used to compute ϕ. The cascade algorithm converges exponentially fast.

In practice, H is sometimes unknown. In this case, one can use the piecewise linear empirical $\hat{H}$ in place of the "true" H. Here $\hat{H}$ is the piecewise linear function satisfying $H(t_i) = i/n$.

Let W be the discrete wavelet transform and let

$$\tilde{\Theta} = (\tilde{\xi}_{j_0 1}, \cdots, \tilde{\xi}_{j_0 2^{j_0}}, \tilde{\theta}_{j_0 1}, \cdots, \tilde{\theta}_{j_0 2^{j_0}}, \cdots, \tilde{\theta}_{J-1,1}, \cdots, \tilde{\theta}_{J-1,2^{J-1}})'$$

where $\tilde{\xi}_{j_0 k}$ and $\tilde{\theta}_{jk}$ are given as in (8). We can view P_H as a preconditioning matrix because

$$\tilde{\Theta} = W \cdot (P_H \cdot n^{-1/2} Y)$$

Our algorithm for implementing the procedure has the following steps:

- Step 1: Use the cascade algorithm to compute P_H, then preconditioning the data $n^{-1/2}Y$ by P_H, say $Y_p = P_H \cdot n^{-1/2}Y$;
- Step 2: Apply discrete wavelet transform to the preconditioned data to get the noisy wavelet coefficients, let $\tilde{\Theta} = W \cdot Y_p$;
- Step 3: Threshold the noisy wavelet coefficients, denote $\hat{\theta}_{jk} = \eta_{\lambda_{jk}}(\tilde{\theta}_{jk})$; where $\eta_{\lambda_{jk}}$ is either the hard or soft thresholding function.

Then

$$\hat{f}_n(t) = \sum_{k=1}^{2^j} \hat{\xi}_{j_0 k}\, \phi_{j_0 k}(t) + \sum_{j=j_0}^{J-1} \sum_{k=1}^{2^j} \hat{\theta}_{jk}\, \psi_{jk}(t)$$

is our estimate of the target function f.

If one is also interested in estimating the function at sample points, then two more steps are needed to get there:

- Step 4: Apply the inverse wavelet transform to the denoised wavelet coefficients to get $W^{-1} \cdot \hat{\Theta}$;
- Step 5: Compute P^H by using the cascade algorithm, where $P^H(k,i) = \phi_{Ji}(t_k)$, then apply this postconditioning transform to $W^{-1} \cdot \hat{\Theta}$ to get the estimate of $f(t_i)$.

$$\hat{f}_n = P^H \cdot (W^{-1} \cdot \hat{\Theta}) \tag{22}$$

Note that both the preconditioning matrix P_H and the postconditioning matrix P^H are sparse matrices with only $O(n)$ non-zero entries.

6 Proofs

This section contains proofs of the main results. We begin with a brief proof of Theorem 1 by using Lemmas 1 and 2.

Proof of Theorem 1: Let $g(t) = f(H^{-1}(t))$. Denote $0 < a_1 < a_2 \cdots < a_m < 1$ the jump discontinuities of the function f, then $0 < b_1 < b_2 \cdots < b_m < 1$ are the jump points of g where $b_i = H^{-1}(a_i)$. Let $s(\alpha) = min(\alpha,\ 1)$, $b_0 = 0$ and $b_{m+1} = 1$. Then on each interval $[b_i, b_{i+1})$, $i = 0,\ 1,\ \cdots,\ m$,

$$|g(x) - g(y)| \leq hM|x-y|^{s(\alpha)}$$

Therefore, $g \in \Lambda^{s(\alpha)}(hM, B, m)$. Now $f_n(t) = Proj_{V_J} g_n(H(t))$. It follows from Lemma 1 and Lemma 2 that

$$\begin{aligned} \|f_n - f\|_2^2 &\leq \|Proj_{V_J}(g_n \circ H - g \circ H)\|_2^2 + \|Proj_{V_J} f - f\|_2^2 \\ &\leq C\, n^{-2s(\alpha)} + C\, m\, n^{-1} = o(n^{-2\alpha/(1+2\alpha)}) \end{aligned}$$ ∎

The proof of Theorem 2 is straightforward. For the reason of spaces, we omit the proof of the theorem. Before we prove Theorem 3 and Theorem 4, let us consider the problem of estimating a univariate normal mean.

Let $y \sim N(\theta, \sigma^2)$ be a normal variable with known variance σ^2. We are interested in estimating the mean θ with threshold estimator and we wish to assess the risk of the estimator. Let $\lambda = a\,\sigma$ with $a \geq 1$. And let $\hat{\theta}^h_\lambda = yI(|y| > \lambda)$ be a hard threshold estimator and let

$$\hat{\theta}^s_\lambda = sgn(y)(|y| - \lambda)_+$$

be a soft threshold estimator of the mean θ. We recall the following results on the risk upper bound of the threshold estimator $\hat{\theta}$ from Cai (1996).

Lemma 3 *Suppose $y \sim N(\theta, \sigma^2)$. Let $\hat{\theta}^s_\lambda$ and $\hat{\theta}^h_\lambda$ be soft and hard threshold estimator of θ respectively. Let $\lambda = a\,\sigma$ with $a \geq 1$. Then*

$$(i). \quad E(\hat{\theta}^s_\lambda - \theta)^2 \leq (a^2+1)\sigma^2 \wedge (2\theta^2 + e^{-a^2/2}\sigma^2) \tag{23}$$

$$(ii). \quad E(\hat{\theta}^h_\lambda - \theta)^2 \leq (2a^2+2)\sigma^2 \wedge (2\theta^2 + 2ae^{-a^2/2}\sigma^2) \tag{24}$$

The proofs of Theorem 3 and Theorem 4 are given only for soft threshold estimators. The proofs for hard threshold estimators are similar.

Proof of Theorem 3: We follow the notations in Section 3.1. Let $g(t) = f(H^{-1}(t))$ and $\tilde{g}(t) = n^{-1/2}\sum_{i=1}^n y_i\phi_{Ji}(t)$ and let $\tilde{f}(t) = \tilde{g}(H(t))$ Then

$$\begin{aligned}\tilde{f}(t) &= \tilde{n}^{-1/2}\sum_{i=1}^n f(t_i)\phi_{Ji}(H(t)) + n^{-1/2}\epsilon\sum_{i=1}^n z_i\phi_{Ji}(H(t)) \\ &= f(t) + \Delta(t) + r(t)\end{aligned}$$

where $\Delta(t) = n^{-1/2}\sum_{i=1}^n f(t_i)\phi_{Ji}(H(t)) - f(t)$ is the approximation error and $r(t) = n^{-1/2}\epsilon\sum_{i=1}^n z_i\phi_{Ji}(H(t))$. Now project $\tilde{f}$ onto multiresolution space V_J and decompose the orthogonal projection $\tilde{f}_J(t) = Proj_{V_J}\tilde{f}(t)$ into three terms:

$$\tilde{f}_J(t) = f_J(t) + \Delta_J(t) + r_J(t) \tag{25}$$

where $f_J = Proj_{V_J}f$, $\Delta_J = Proj_{V_J}\Delta$ and $r_J = Proj_{V_J}r$ respectively. Theorem 1 yields that

$$\|\Delta_J\|_2^2 = o(n^{-2\alpha/(1+2\alpha)}) \tag{26}$$

Denote $\tilde{\theta}_{jk} = \langle \tilde{f}_J, \psi_{jk}\rangle$. In the same fashion as in (25), we decompose $\tilde{\theta}_{jk}$ into three parts:

$$\tilde{\theta}_{jk} = \theta_{jk} + d_{jk} + r_{jk} \quad \text{for } k = 1, \cdots, 2^j, \;\; j = j_0, \cdots, J-1$$

where $\theta_{jk} = \langle f, \psi_{jk}\rangle$ is the true wavelet coefficient of f, $d_{jk} = \langle \Delta_J, \psi_{jk}\rangle$ is the approximation error and $r_{jk} = \langle r_J, \psi_{jk}\rangle$ is the noise. Similarly separate $\tilde{\xi}_{j_0k} = \langle \tilde{f}_J, \phi_{j_0k}\rangle$ into three terms:

$$\tilde{\xi}_{j_0k} = \xi_{j_0k} + d'_{j_0k} + r'_{j_0k} \qquad \text{for } k = 1, \cdots, 2^{j_0}$$

Let $\hat{\xi}_{j_0k}$ and $\hat{\theta}_{jk}$ are given as in (9). Note that

$$\sum_{k=1}^{2^{j_0}}(d'_{j_0k})^2 + \sum_{j=j_0}^{J-1}\sum_{k=1}^{2^j} d_{jk}^2 = \|\Delta_J\|_2^2 = o(n^{-2\alpha/(1+2\alpha)}). \tag{27}$$

and

$$\begin{aligned} \sigma_{jk}^2 &\equiv var(r_{jk}) = n^{-1}\epsilon^2\sum_{i=1}^{n}(\langle\phi_{Ji}\circ H, \psi_{jk}\rangle)^2 \le n^{-1}\epsilon^2 h \\ (\sigma'_{j_0k})^2 &\equiv var(r'_{j_0k}) \le n^{-1}\epsilon^2 h \end{aligned}$$

By the orthonormality of the wavelet basis, we have the isometry between the L_2 function norm and the ℓ_2 wavelet sequence norm.

$$\begin{aligned} E\|\hat{f}_n^* - f\|^2 &= \sum_{k=1}^{2^{j_0}} E(\hat{\xi}_{j_0k} - \xi_{j_0k})^2 + \sum_{j=j_0}^{J-1}\sum_{k=1}^{2^j} E(\hat{\theta}_{jk} - \theta_{jk})^2 + \sum_{j=J}^{\infty}\sum_{k=1}^{2^j}\theta_{jk}^2 \\ &\equiv S_1 + S_2 + S_3 \end{aligned}$$

It is easy to see that

$$S_1 \le 2^{j_0} n^{-1}\epsilon^2 h + \sum_{k=1}^{2^{j_0}}(d'_{j_0k})^2 = o(n^{-2\alpha/(1+2\alpha)}) \tag{28}$$

At each resolution level j, denote

$$G_j \equiv \{k : supp(\psi_{jk}) = [2^{-j}k, 2^{-j}(N+k)] \text{contains at least one jump point of } f\}$$

Then $card(G_j) \le N(m+2)$ (counting two end points 0 and 1 as jump points as well). Lemma 1 yields

$$|\theta_{jk}| \le C2^{-j(1/2+\alpha)} \qquad \text{for } k \notin G_j; \tag{29}$$

$$|\theta_{jk}| \le C2^{-j/2} \qquad \text{for } k \in G_j; \tag{30}$$

where C is a constant not depending on f. Therefore,

$$\begin{aligned} S_3 &= \sum_{j=J}^{\infty}\sum_{k\in G_j}\theta_{jk}^2 + \sum_{j=J}^{\infty}\sum_{k\notin G_j}\theta_{jk}^2 \le \sum_{j=J}^{\infty} N(m+2)C^2 2^{-j} + \sum_{j=J}^{\infty}\sum_{k=1}^{2^j} C^2 2^{-j(1+2\alpha)} \\ &= o(n^{-2\alpha/(1+2\alpha)}) \end{aligned} \tag{31}$$

Now we consider S_2. First note from (15) that $\sigma_{jk} \le u_{jk}$ and $\lambda_{jk} = u_{jk}(2\log n)^{1/2}$, so $a_{jk} \equiv \lambda_{jk}/\sigma_{jk} \ge (2\log n)^{1/2}$, it follows from (23) that

$$E(\hat{\theta}_{jk} - \theta_{jk})^2 \le (4\log n + 2)h\epsilon^2 n^{-1} \wedge (8\theta_{jk}^2 + 2h\epsilon^2 n^{-2}) + 10 d_{jk}^2 \tag{32}$$

Write

$$\begin{aligned} S_2 &= \sum_{j=j_0}^{J-1} \sum_{k \in G_j} E(\hat{\theta}_{jk} - \theta_{jk})^2 + \sum_{j=j_0}^{J-1} \sum_{k \notin G_j} E(\hat{\theta}_{jk} - \theta_{jk})^2 \\ &\equiv S_{21} + S_{22} \end{aligned}$$

Since $card(G_j) \le N(m+2)$, so it follows from (32) that

$$S_{21} \le \sum_{j=j_0}^{J-1} N(m+2) \cdot [(4 \log n + 2) h \epsilon^2 n^{-1} + 10 d_{jk}^2] = o(n^{-2\alpha/(1+2\alpha)}) \tag{33}$$

Now let J_1 be an integer satisfying $2^{J_1(1+2\alpha)} = n/\log n$. (For simplicity we assume the existence of such an integer. In general, choose $J_1 = \lfloor 1/(1+2\alpha) \log_2(n/\log n) \rfloor$.) From (32), we have

$$\begin{aligned} E(\hat{\theta}_{jk} - \theta_{jk})^2 &\le 5\epsilon^2 n^{-1} \log n + 10 d_{jk}^2 & \text{for } j_0 \le j \le J_1 - 1,\ k \notin G_j \quad (34) \\ E(\hat{\theta}_{jk} - \theta_{jk})^2 &\le 8C^2 2^{-j(1+2\alpha)} + 2h\epsilon^2 n^{-2} + 10 d_{jk}^2 & \text{for } J_1 \le j \le J - 1,\ k \notin G_j \quad (35) \end{aligned}$$

Therefore,

$$\begin{aligned} S_{22} &\le \sum_{j=j_0}^{J_1-1} \sum_{k \notin G_j} 5\epsilon^2 n^{-1} \log n + \sum_{j=J_1}^{J-1} \sum_{k \notin G_j} (8C^2 2^{-j(1+2\alpha)} + 2h\epsilon^2 n^{-2}) + 10 \sum_{j=j_0}^{J-1} \sum_{k=1}^{2^j} d_{jk}^2 \\ &= C (\log n / n)^{2\alpha/(1+2\alpha)} (1 + o(1)) \qquad (36) \end{aligned}$$

We finish the proof by putting (28),(31), (33) and (36) together.

$$E\|\hat{f}_n^* - f\|_2^2 \le C\ (\log n/n)^{2\alpha/(1+2\alpha)} (1 + o(1)) \tag{37}$$

∎

Proof of Theorem 4 : First we recall a simple but useful inequality.

Lemma 4 *Let X_i be random variables, $i = 1, \cdots, n$. Then*

$$E(\sum_{i=1}^{n} X_i)^2 \le (\sum_{i=1}^{n} (EX_i^2)^{1/2})^2 \tag{38}$$

Now apply the inequality (38), we have

$$\begin{aligned} E(f_n^*(t_0) - f(t_0))^2 &= E\left[\sum_{k=1}^{2^{j_0}} (\hat{\xi}_{j_0 k} - \xi_{j_0 k}) \phi_{j_0 k}(t_0) + \sum_{j=j_0}^{\infty} \sum_{k=1}^{2^j} (\hat{\theta}_{jk} - \theta_{jk}) \psi_{jk}(t_0)\right]^2 \\ &\le \left[\sum_{k=1}^{2^{j_0}} (E(\hat{\xi}_{j_0 k} - \xi_{j_0 k})^2 \phi_{j_0 k}^2(t_0))^{1/2} + \sum_{j=j_0}^{J-1} \sum_{k=1}^{2^j} (E(\hat{\theta}_{jk} - \theta_{jk})^2 \psi_{jk}^2(t_0))^{1/2} + \sum_{j=J}^{\infty} \sum_{k=1}^{2^j} |\theta_{jk} \psi_{jk}(t_0)|\right]^2 \\ &\equiv (Q_1 + Q_2 + Q_3)^2 \end{aligned}$$

Now consider the three terms separately. Note that at each resolution level j, there are at most N basis functions ψ_{jk} that are nonvanishing at t_0, where N is the length of the support of wavelet functions ϕ and ψ. Therefore,

$$Q_1 = \sum_{k=1}^{2^{j_0}} (E(\hat{\xi}_{j_0k} - \xi_{j_0k})^2)^{1/2} |\phi_{j_0k}(t_0)| \le C\,(Nn^{-1}\epsilon^2 h + \sum_{k=1}^{2^{j_0}} (d'_{jk})^2)^{1/2} = o(n^{-\alpha/(1+2\alpha)}) \quad (39)$$

For the third term, it follows from Lemma 1(i) that

$$Q_3 = \sum_{j=J}^{\infty} \sum_{k=1}^{2^j} |\theta_{jk}||\psi_{jk}(t_0)| \le \sum_{j=J}^{\infty} N\|\psi\|_\infty 2^{j/2} C 2^{-j(1/2+\alpha)} \le Cn^{-\alpha} \quad (40)$$

Now let us consider the term S_2. First, note that for function $f \in \Lambda^\alpha(M, B, 0)$, the approximation error $\Delta(t)$ satisfies $\sup_t |\Delta(t)| \le Cn^{-s(\alpha)}$. This yields that

$$|d_{jk}| = |\langle \Delta, \psi_{jk} \rangle| \le C_1 2^{-j/2} n^{-s(\alpha)}$$

where the constant C_1 does not depend on f. Let the integer J_1 be given as in the proof of Theorem 3. Apply (34) and (35), then

$$\begin{aligned} Q_2 &\le N\|\psi\|_\infty \sum_{j=j_0}^{J_1-1} (5\epsilon^2 n^{-1}\log n + 10C_1^2 2^{-j} n^{-2s(\alpha)})^{1/2} \\ &\quad + N\|\psi\|_\infty \sum_{j=J_1}^{J-1} (8C^2 2^{-j(1+2\alpha)} + 2h\epsilon^2 n^{-2} + 10C_1^2 2^{-j} n^{-2s(\alpha)})^{1/2} \\ &= C(\frac{\log n}{n})^{\frac{\alpha}{1+2\alpha}}(1+o(1)) \end{aligned} \quad (41)$$

Combining (39), (40) and (41), we have

$$E(f_n^*(t_0) - f(t_0))^2 \le C(\log n/n)^{2\alpha/(1+2\alpha)}(1+o(1)) \quad (42)$$

∎

Acknowledgements: Part of the research was completed when the second author was a visitor at the Department of Statistics at the University of Pennsylvania. The second author would like to thank the first author and the department for their (financial and other) support.

References

[1] Andersson, L., Hall, N., Jawerth, B., & Peters, G. (1994). Wavelets on Closed Subsets of the Real Line. In *Recent Advances in Wavelet analysis* (Larry L. Shumaker and Glenn Webb, eds.), Academic Press, New York, 1–61.

[2] Bickel, P. J. (1983). Minimax Estimation of a Normal Mean Subject to Doing Well at a Point. In *Recent Advances in Statistics* (M. H. Rizvi, J. S. Rustagi, and D. Siegmund, eds.), Academic Press, New York, 511–528.

[3] Breiman, L., Friedman, J.H., Olshen, R.A.,& Stone, C.J. (1983). *CART: Classification and Regression Trees*. Wadsworth: CBelmont, CA.

[4] Brown, L.D. & Low, M.G. (1996). Asymptotic Equivalence of Nonparametric Regression and White Noise. *Annals of Statistics.* **24** 2384 - 2398.

[5] Brown, L.D. & Low, M.G. (1992). A Constrained Risk Inequality with Applications to Nonparametric Functional Estimations. *Annals of Statistics.* **24** 2524 - 2535.

[6] Cai, T. (1996). Nonparametric Function Estimation via Wavelets. Ph.D Thesis, Cornell University.

[7] Casella, G. & Strawderman, W.E. (1981). Estimating a Bounded Normal Mean. *Annals of Statistics.* **9** 870-878.

[8] Cohen, A., Daubechies, I., Jawerth, B. & Vial, P. (1993). Multiresolution analysis, wavelets, and fast algorithms on an interval. *Comptes Rendus Acad. Sci. Paris* (A). **316.**, 417–421.

[9] Daubechies, I. (1988). Orthonormal bases of compactly supported wavelets. *Communications in Pure and Applied Mathematics* **41**, Nov. 1988, pp. 909-996.

[10] Daubechies, I. (1992). *Ten Lectures on Wavelets* SIAM: Philadelphia.

[11] Donoho, D.L. & Johnstone, I.M. (1992). Ideal Spatial Adaptation via Wavelet Shrinkage. *Biometrika,* ***81***, 425–455.

[12] Donoho, D.L. & Johnstone, I.M. (1995). Adapting to Unknown Smoothness via Wavelet Shrinkage. *JASA,* **90(432)**, 1200 - 1224.

[13] Donoho, D.L. & Johnstone, I.M. (1995). Neo-Classic Minimax Problems, Thresholding, and Adaptation. *Bernoulli,* ***2***, 39 - 62.

[14] Donoho, D.L., Johnstone, I.M., Kerkyacharian, G. & Picard, D. (1995). Wavelet Shrinkage: Asymptopia?, *J. Roy. Stat. Soc. Ser. B,* ***57***, 301–369.

[15] Hall, P. & Turlach, B.A. (1996). Interpolation Methods For Nonlinear Wavelet Regression with Irregularly Spaced Design. Manuscript.

[16] Härdle, W. (1990). *Applied Nonparametric Regression.* Cambridge University Press, Cambridge.

[17] Ibragimov, I.A. & Has'minskii, R.Z. (1984). Nonparametric Estimation of the Value of a Linear Functional in a Gaussian White Noise. *Theory of Probability and its Applications* **29**, 1-17.

[18] Johnstone, I.M. & Silverman, B.W. (1994). Wavelet Threshold Estimators For Data with Correlated Noise. Technical Report, Stanford University.

[19] Lepski, O.V. (1990). On a Problem of Adaptive Estimation on White Gaussian Noise. *Theory of Probability and Appl.* **35**, 3, 454-466

[20] Meyer, Y. (1990). *Ondelettes et Opérateurs: I. Ondelettes.* Hermann et Cie, Paris.

[21] Strang, G. (1989). Wavelet And Dilation Equations: A Brief Introduction. *SIAM Review,* 31(4), 614 - 627.