Wavelet estimation for samples with random uniform design

T. Tony Cai [a,*] , Lawrence D. Brown [b]

[a] *Department of Statistics, Purdue University, West Lafayette, IN 47907, USA*
[b] *Department of Statistics, University of Pennsylvania, PA 19104, USA*

Abstract

We show that for nonparametric regression if the samples have random uniform design, the wavelet method with universal thresholding can be applied directly to the samples as if they were equispaced. The resulting estimator achieves within a logarithmic factor from the minimax rate of convergence over a family of Hölder classes. Simulation result is also discussed. © 1999 Elsevier Science B.V. All rights reserved

1. Introduction

Wavelet shrinkage methods have been very successful in nonparametric regression. But so far most of the wavelet regression methods have been focused on 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. The most widely used wavelet shrinkage method for equispaced samples is the Donoho–Johnstone's VisuShrink procedure (Donoho and Johnstone, 1992; Donoho et al., 1995). It has three steps:

1. transform the noisy data via the discrete wavelet transform;
2. denoise the empirical wavelet coefficients by "hard" or "soft" thresholding rules with threshold $\lambda = \sigma \sqrt{2 \log n}$;
3. estimate function f at the sample points by inverse discrete wavelet transform of the denoised wavelet coefficients.

---

* Corresponding author.

This procedure is adaptive and easy to implement. The computational cost is of O(n). And with high probability, VisuShrink estimator is at least as smooth as the target function. The estimator produced by the procedure achieves minimax convergence rates up to a logarithmic penalty over a wide range of function classes.

In many statistical applications, however, the samples are nonequispaced. It is shown that the procedure might produce suboptimal estimator if it is applied directly to nonequispaced samples (Cai, 1996). Wavelet methods for samples with nonequispaced designs have been studied by Cai and Brown (1998) and Hall and Turlach (1997). Cai and Brown (1998) introduced a wavelet shrinkage method for samples with fixed nonequispaced designs based on approximation approach. It is shown that the estimator attains near-minimaxity across a range of piecewise Hölder classes. Hall and Turlach (1997) proposed interpolation methods for samples with random designs. They used samples with random uniform design as examples for their methods. Despite the asymptotic near-optimality for these nonequispaced methods, the estimators are computationally much harder to implement than VisuShrink for equispaced samples.

In the present paper, we consider the special case of samples with random uniform design. We show that in this special case the samples can in fact be treated as if they were equispaced. That is, the VisuShrink procedure of Donoho and Johnstone can be applied directly to the data and the resulting estimator adaptively achieves within a logarithmic factor of the optimal convergence rate across a range of Hölder classes. Therefore, we have a fast estimation procedure for samples with random uniform design. Simulation is conducted to evaluate the numerical performance of the method. It is shown that the mean-squared error is comparable to that of the samples with truly equispaced designs.

In Section 2 we describe the method and state the asymptotic optimality property of the estimator. Section 3 summarizes the simulation results. Some relevant results on wavelet approximation is presented in Section 4. Section 5 contains a concise proof of the main results.

2. Methodology

2.1. Wavelets

Let ϕ and ψ denote the orthogonal father and mother wavelet functions. The functions ϕ and ψ are assumed to be compactly supported with associated discrete wavelet transform W. Assume ψ has r vanishing moments and ϕ satisfies $\int \phi = 1$. Let

$$\phi_{jk}(x) = 2^{j/2}\phi(2^j x - k), \qquad \psi_{jk}(x) = 2^{j/2}\psi(2^j x - k).$$

And denote the periodized wavelets

$$\phi_{jk}^{\mathrm{p}}(x) = \sum_{l \in \mathscr{Z}} \phi_{jk}(x - l), \qquad \psi_{jk}^{\mathrm{p}}(x) = \sum_{l \in \mathscr{Z}} \phi_{jk}(x - l) \quad \text{for } x \in [0,1].$$

For the purposes of this paper, we use the periodized wavelet bases on $[0,1]$. The collection $\{\phi_{j_0 k}^{\mathrm{p}}, k = 1, \ldots, 2^{j_0}; \psi_{jk}^{\mathrm{p}}, j \geqslant j_0, k = 1, \ldots, 2^j\}$ 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. This basis has an associated exact orthogonal discrete wavelet transform (DWT) that transforms data into wavelet coefficient domains.

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

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

Wavelet transform decomposes a function into different resolution components. In (1), $\xi_{j_0 k}$ 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.

2.2. The estimator

Consider the nonparametric regression model

$$y_i = f(x_i) + \sigma z_i, \tag{2}$$

$i = 1, 2, \ldots, n \ (= 2^J)$, x_i's are independently uniformly distributed on $[0, 1]$, z_i's are independent $N(0,1)$ variables and independent of x_i's, and the noise level σ is fixed and known.

The function f is an unknown function of interest. We wish to estimate f globally with small integrated mean-squared error

$$R(\hat{f}, f) = E \int_0^1 (\hat{f}(x) - f(x))^2 \, \mathrm{d}x.$$

Let $0 \leqslant x_{(1)} < x_{(2)} < \cdots < x_{(n)} \leqslant 1$ be the order statistics of the x_i's. Now relabel y_i's and z_i's according the order of the x_i's. For convenience, we use the same label. So

$$y_i = f(x_{(i)}) + \sigma z_i. \tag{3}$$

Now we observed $(x_{(1)}, y_1), (x_{(2)}, y_2), \ldots, (x_{(n)}, y_n)$ with x_i independently uniformly distributed on $[0, 1]$. So $x'_{(i)}s$ are not equispaced in general. But we pretend that $x_{(i)}$ is $Ex_{(i)} = i/(n+1)$. That is, we pretend to have an equispaced sample

$$\left(\frac{1}{n+1}, y_1 \right), \left(\frac{2}{n+1}, y_2 \right), \ldots, \left(\frac{n}{n+1}, y_n \right).$$

We apply Donoho and Johnstone's VisuShrink procedure directly to $y = \{y_1, y_2, \ldots, y_n\}$.

Let $\tilde{\theta} = W n^{-1/2} y$ be the discrete wavelet transform of $n^{-1/2} y$. Write

$$\tilde{\theta} = (\tilde{\xi}_{j_0 1}, \ldots, \tilde{\xi}_{j_0 2^{j_0}}, \tilde{\theta}_{j_0 1}, \ldots, \tilde{\theta}_{j_0 2^{j_0}}, \ldots, \tilde{\theta}_{J-1,1}, \ldots, \tilde{\theta}_{J-1, 2^{J-1}})^{\mathrm{T}}.$$

Here $\tilde{\xi}_{j_0 k}$ are the empirical coefficients of the father wavelets at the lowest resolution level. They represent the gross structure of the function and are usually not thresholded. The coefficients $\tilde{\theta}_{jk}$ $(j = j_0, \ldots, J-1, \ k = 1, \ldots, 2^j)$ are fine structure wavelet terms.

The empirical wavelet coefficients is denoised via soft thresholding:

$$\hat{\theta}_{jk} = \eta_\lambda^s(\tilde{\theta}_{jk}) = \mathrm{sgn}(\tilde{\theta}_{jk})(|\tilde{\theta}_{jk}| - \lambda)_+ \quad \text{where } \lambda = \sigma \sqrt{2n^{-1} \log n}.$$

The whole function f is estimated by

$$\hat{f}_*(x) = \sum_{k=1}^{2^{j_0}} \tilde{\xi}_{j_0 k} \phi_{j_0 k}(x) + \sum_{j=j_0}^{J-1} \sum_{k=1}^{2^j} \hat{\theta}_{jk} \psi_{jk}(x).$$

If one is interested in estimating the function at the sample points, apply the inverse discrete wavelet transform to the denoised wavelet coefficients:

$$(f_*\widehat{(x_{(k)})})_{k=1}^n = W^{-1}n^{1/2}\hat{\theta}.$$

The estimator is adaptive and easy to implement.

Theorem 1. *Suppose that the sample* $(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)$ *is observed as in* (2) *and the mother wavelet* ψ *has r vanishing moments. Then the estimator constructed above achieves within a logarithmic factor of the optimal convergence rate over the range of Hölder classes* $\Lambda^\alpha(M)$ *(defined in Section* 4*) with* $\frac{1}{2} \leqslant \alpha \leqslant r$. *That is,*

$$\sup_{f \in \Lambda^\alpha(M)} E \parallel \hat{f}_* - f \parallel_2^2 \leqslant C \left(\frac{\log n}{n} \right)^{2\alpha/(1+2\alpha)}, \tag{4}$$

$$\sup_{f \in \Lambda^\alpha(M)} \frac{1}{n} \sum E \parallel \widehat{f_*(x_i)} - f(x_i) \parallel_2^2 \leqslant C \left(\frac{\log n}{n} \right)^{2\alpha/(1+2\alpha)} \tag{5}$$

for all $M \in (0, \infty)$ *and* $\alpha \in [\frac{1}{2}, r]$.

Remark. The same result holds for hard threshold estimator. The result shows that in the case of random design with uniformly distributed x_i's, we can treat it as if they are fixed equispaced design. The constraint $\alpha \geqslant \frac{1}{2}$ is due to the approximation of $f(x_{(i)})$ by $f(i/(n+1))$.

3. Simulations

A simulation study is conducted to compare the estimator based on random-x samples with the estimator based on truly equispaced samples. The results show that the quality of the estimator based on random-x samples is comparable to the estimator based on equispaced samples.

We consider four test functions of Donoho and Johnstone (1994) representing different level of spatial variability. The test functions are plotted in Fig. 1. For each of the four objects under study, we compare the estimators at two noise levels, one with signal-to-noise ratio SNR $= 5$ and another with SNR $= 7$. Sample sizes from $n = 512$ to 8192 are considered.

Table 1 reports the mean-squared errors over 200 replications of the four test functions: Doppler, Heavi-Sine, Bumps and Blocks. The wavelet used is the Symmlet "s8". The conventional t-test is used to test the significance of the differences between the MSEs of random design and the equispaced design. Table 1 shows that the MSEs of random-x are worse than those of equispaced-x in 29 out of 40 cases, and are significantly worse in the case of Doppler function. Random design is better in six cases, and the differences in MSE between the two designs are insignificant in five cases at 95% level according to the t-test.

The following plots compare the visual quality of the reconstructions (see Fig. 2). The solid line is the estimator and the dotted line is the true function. The sample size is 1024 and SNR $= 7$. For each function, one is based on a sample with random-x and another is based on a sample with equispaced-x. One can see from the plots, the visual quality of the estimators are comparable, with the random-x reconstruction a little wobblier due to the stochastic nature of the design. For more simulation results, the readers are referred to Cai and Brown (1997).
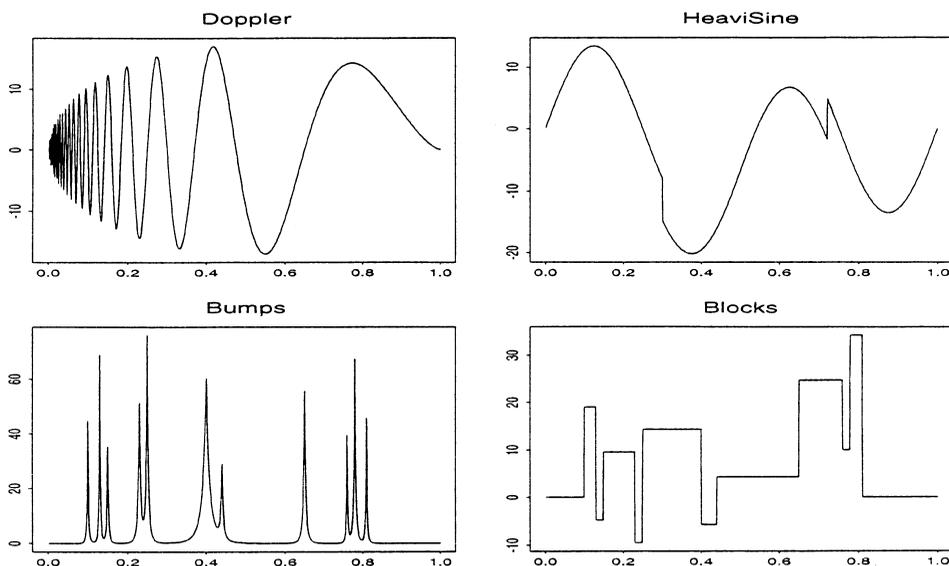
Fig. 1. Test functions.

Table 1
Mean-squared errors from 200 replications. The "% diff." columns are the percentage differences between the MSEs of random-x and of equispaced-x. The percentage differences are set to 0 when the differences are insignificant at 95% level according to the conventional t-test

| n	SNR = 5			SNR = 7		
	Uniform-x	Equispaced-x	% diff.	Uniform-x	Equispaced-x	% diff.
Doppler						
512	3.94	2.88	37	2.85	1.80	58
1024	2.54	1.88	35	1.79	1.19	50
2048	1.61	1.21	33	1.10	0.76	44
4096	0.86	0.69	24	0.57	0.43	34
8192	0.53	0.44	18	0.34	0.26	29
HeaviSine						
512	0.62	0.55	13	0.45	0.40	14
1024	0.42	0.40	4	0.30	0.29	4
2048	0.28	0.30	−4	0.20	0.20	0*
4096	0.17	0.20	−11	0.11	0.12	−6
8192	0.11	0.12	−7	0.08	0.08	0*
Bumps						
512	8.27	9.31	−11	5.26	5.79	−9
1024	6.07	5.97	0*	3.80	3.61	5
2048	4.02	3.82	5	2.51	2.26	11
4096	2.22	1.98	12	1.39	1.16	20
8192	1.34	1.21	11	0.83	0.71	18
Blocks						
512	5.39	5.34	0*	3.39	3.42	0*
1024	3.83	3.69	4	2.37	2.28	4
2048	2.66	2.55	4	1.63	1.58	3
4096	1.54	1.43	8	0.95	0.89	7
8192	1.04	0.99	4	0.64	0.62	3
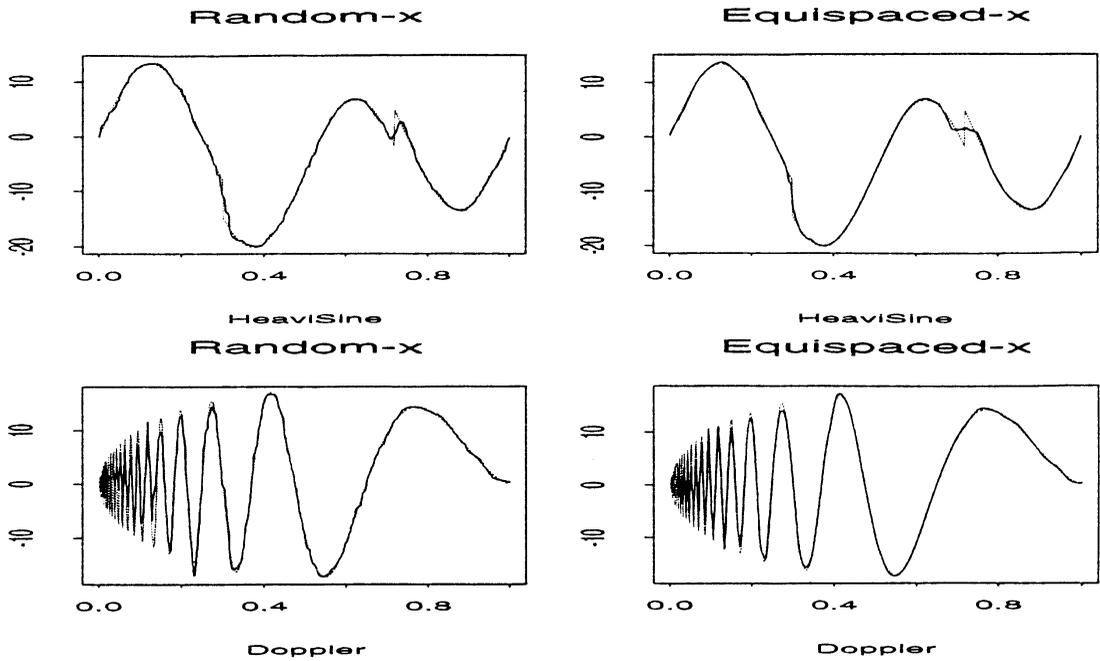
Fig. 2. Comparisons of the reconstructions.

4. Wavelet approximation

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 co-efficients. See Meyer (1990). In the present paper, we consider the estimation problem over a range of Hölder classes.

Definition 1. We define the following Hölder classes $\Lambda^\alpha(M)$:

(i) if $\alpha \leqslant 1$, $\Lambda^\alpha(M) = \{f : |f(x) - f(y)| \leqslant M|x - y|^\alpha\}$;

(ii) if $\alpha > 1$, $\Lambda^\alpha(M) = \{f : |f^{(\lfloor\alpha\rfloor)}(x) - f^{(\lfloor\alpha\rfloor)}(y)| \leqslant M|x - y|^{\alpha'}$ and $|f'(x)| \leqslant M\}$ where $\lfloor\alpha\rfloor$ is the largest integer less than α and $\alpha' = \alpha - \lfloor\alpha\rfloor$.

The wavelet coefficients of functions in a Hölder class $\Lambda^\alpha(M)$ decay exponentially as the resolution level j increases (see Daubechies, 1992).

Lemma 1. *Let $f \in \Lambda^\alpha(M)$ and let the wavelet function ψ has r vanishing moments with $r \geqslant \alpha$. Let $\theta_{jk} = \langle f, \psi_{jk} \rangle$ be wavelet coefficients of f. Then*

$$|\theta_{jk}| \leqslant C2^{-j(1/2+\alpha)}, \tag{6}$$

where C is a constant depending on M and the wavelet basis only.

If one has a sampled function $\{f(k/(n+1))\}_{k=1}^n$ with $n = 2^J$, one can utilize a wavelet basis to get a good approximation of the entire function f. Denote $s(\alpha) = \min(\alpha, 1)$. We have the following (also see Daubechies, 1992).

Proposition 1. *Suppose that $f \in \Lambda^\alpha(M)$, and let $\xi_{Jk} = \langle f, \phi_{Jk} \rangle$, then*

$$\left| n^{-1/2} f\left(\frac{k}{n+1} \right) - \xi_{Jk} \right| \leqslant Cn^{-(1/2+s(\alpha))}. \tag{7}$$

According to this result, we may use $n^{-1/2} f(k/(n+1))$ as an approximation of ξ_{Jk}. This means that if an equispaced sampled function is given, we can use a wavelet basis to get an approximation of the entire function f. To be more specific, we can use $f_n(x) = \sum_{k=1}^n n^{-1/2} f(k/(n+1)) \phi_{Jk}(x)$ as an approximation of f. Furthermore, the approximation error can be bounded based on the sample size and the smoothness of the function. We quote the following result from Cai (1996).

Proposition 2. *Suppose that $f \in \Lambda^\alpha(M)$. Let $f_n(x) = \sum_{k=1}^n n^{-1/2} f(k/(n+1)) \phi_{Jk}(x)$ Then the approximation error satisfies*

$$\| f_n - f \|_2^2 \leqslant Cn^{-2s(\alpha)}. \tag{8}$$

5. Proof

We need some preparations before we prove the theorem. First some well-known results on the order statistics of uniform variables.

Lemma 2. *Let x_i be i.i.d. uniform random variables on $[0,1]$. And let $0 \leqslant x_{(1)} < x_{(2)} < \cdots < x_{(n)} \leqslant 1$ be the order statistics. Then $x_{(k)}$ is distributed as $Beta(k, n-k+1)$. In particular,*

$$Ex_{(k)} = \frac{k}{n+1}, \qquad Ex_{(k)}^2 = \frac{k+k^2}{(n+1)(n+2)}, \qquad \mathrm{Var}\{x_{(k)}\} = \frac{(n+1)k - k^2}{(n+1)^2(n+2)}.$$

Now let us consider the noiseless case. We want to simply use $f(x_{(i)})$ as an approximation of $f(i/(n+1))$ and wish to know the approximation error. Denote E_1 the conditional expectation given $x_1, x_2, \ldots, x_n$ and denote E_x the expectation with respect to $x_1, x_2, \ldots, x_n$.

Lemma 3. *The upper bound of the approximation error is*

$$\sup_{f \in \Lambda^\alpha(M)} \frac{1}{n} \sum E_x \left(f(x_{(k)}) - f\left(\frac{k}{n+1} \right) \right)^2 \leqslant Cn^{-s(\alpha)}. \tag{9}$$

Proof. For a fixed $f \in \Lambda^\alpha(M)$, we have $|f(x) - f(y)| \leqslant C |x-y|^{s(\alpha)}$. Some algebra shows that

$$\frac{1}{n} \sum_k E_x \left(f(x_{(k)}) - f\left(\frac{k}{n+1} \right) \right)^2 \leqslant \frac{C}{n} \sum_k \left[E_x \left(x_{(k)} - \frac{k}{n+1} \right)^2 \right]^{s(\alpha)}$$

$$\leqslant \frac{C(n+1)^{1+s(\alpha)}}{n(n+1)^{s(\alpha)}(n+2)^{s(\alpha)}} \leqslant Cn^{-s(\alpha)}. \qquad \square$$

To prove the main result, we also need the following upper bound of the risk of threshold estimator of a univariate normal mean. Similar bound holds for hard threshold. The proof can be found in Cai (1996).

Lemma 4. *Suppose that $y \sim N(\theta, n^{-1}\sigma^2)$. Then $\hat{\theta} = \eta_\lambda^s(y)$ with $\lambda = \sigma\sqrt{2n^{-1} \log n}$ satisfies*

$$E(\hat{\theta} - \theta)^2 \leqslant (2\theta^2 + n^{-2}\sigma^2) \wedge (2\log n + 1)n^{-1}\sigma^2. \tag{10}$$

Proof of Theorem 1. We give the proof of (4) only. The proof of (5) is similar. First, some notations. We use ξ_{jk} as coefficients of ϕ_{jk} (the "father wavelets"), and use θ_{jk} as coefficients of ψ_{jk} (the "mother wavelets"). The $\tilde{\xi}_{j_0k}$ are the empirical coefficients at the coarsest level. They represent the gross structure of the function and they are usually not thresholded. The discrete wavelet transform W is an orthogonal transform, so it is norm-preserving.

Let $\tilde{f}(x) = \sum_i n^{-1/2} y_i \phi_{Jk}(x)$. Then $\tilde{f}(x)$ can be written as

$$\tilde{f}(x) = \sum_i \left[\xi_{Ji} + \overbrace{\left(n^{-1/2} f\left(\frac{i}{n+1}\right) - \xi_{Ji} \right)}^{A} + \overbrace{\left(n^{-1/2} f(x_{(i)}) - n^{-1/2} f\left(\frac{i}{n+1}\right) \right)}^{B} + \overbrace{n^{-1/2}\sigma z_i}^{R} \right] \phi_{Ji}(x)$$

$$= \sum_k [\xi_{j_0k} + \tilde{a}_{j_0k} + \tilde{b}_{j_0k} + \tilde{r}_{j_0k}] \phi_{j_0k}(x) + \sum_j \sum_k [\theta_{jk} + a_{jk} + b_{jk} + r_{jk}] \psi_{jk}(x).$$

Here the ξ_{j_0k} and θ_{jk} are the discrete wavelet transform of ξ_{Ji}, and likewise $\tilde{a}_{j_0k}$ and a_{jk} the transform of the term A, $\tilde{b}_{j_0k}$ and b_{jk} the transform of B and $\tilde{r}_{j_0k}$ and r_{jk} the transform of R.

Let $\tilde{\xi}_{j_0k} = \xi_{j_0k} + \tilde{a}_{j_0k} + \tilde{b}_{j_0k} + \tilde{r}_{j_0k}$ be the coefficients of gross structure terms and set $\hat{\xi}_{j_0k} = \tilde{\xi}_{j_0k}$. Let $\theta'_{jk} = \theta_{jk} + a_{jk} + b_{jk}$ and let $\tilde{\theta}_{jk} = \theta'_{jk} + r_{jk}$ be the noisy empirical wavelet coefficients. Then $\tilde{\theta}_{jk} \sim N(\theta'_{jk}, n^{-1}\sigma^2)$. Now denote $\lambda = \sigma\sqrt{2n^{-1}\log n}$ and let $\hat{\theta}_{jk} = \text{sgn}(\tilde{\theta}_{jk})(|\tilde{\theta}_{jk}| - \lambda)_+$. Now the estimator of the regression function f is given by

$$\hat{f}_*(x) = \sum_k \hat{\xi}_{j_0k} \phi_{j_0k}(x) + \sum_{j=j_0}^{J-1} \sum_k \hat{\theta}_{jk} \psi_{jk}(x)$$

and the risk function can be written as

$$E \| \hat{f}_* - f \|_2^2 = \sum_k E_x(E_1(\hat{\xi}_{j_0k} - \xi_{j_0k})^2) + \sum_{j=j_0}^{J-1} \sum_k E_x(E_1(\hat{\theta}_{jk} - \theta_{jk})^2) + \sum_{j=J}^{\infty} \sum_k \theta_{jk}^2.$$

Lemma 1 yields that

$$\sum_{j=J}^{\infty} \sum_k \theta_{jk}^2 = O(n^{-2\alpha}). \tag{11}$$

Also, we have

$$\sum_k E_1(\hat{\xi}_{j_0k} - \xi_{j_0k})^2 = 2^{j_0} n^{-1}\sigma^2 + \sum_k (\tilde{a}_{j_0k} + \tilde{b}_{j_0k})^2 \leqslant 2^{j_0} n^{-1}\sigma^2 + 2\sum_k \tilde{a}_{j_0k}^2 + 2\sum_k \tilde{b}_{j_0k}^2. \tag{12}$$

Applying Lemma 4 to the term $E_1(\hat{\theta}_{jk} - \theta_{jk})^2$, we have

$$E_1(\hat{\theta}_{jk} - \theta_{jk})^2 \leqslant 2E_1(\hat{\theta}_{jk} - \theta'_{jk})^2 + 2a_{jk}^2 + 2b_{jk}^2$$

$$\leqslant 8\theta_{jk}^2 \wedge 3n^{-1}\sigma^2 \log n + 10a_{jk}^2 + 10b_{jk}^2 + n^{-2}\sigma^2.$$

Let J_1 be an integer satisfying $2^{J_1} \asymp (n/\log n)^{1/(1+2\alpha)}$. Then,

$$\sum_{j,k} E_1(\hat{\theta}_{jk} - \theta_{jk})^2 \leqslant \sum_{j=j_0}^{J_1-1} \sum_k 3n^{-1}\sigma^2 \log n + \sum_{j=J_1}^{J-1} \sum_k 8\theta_{jk}^2 + 10 \sum_{j=j_0}^{J-1} \sum_k (a_{jk}^2 + b_{jk}^2) + n^{-1}\sigma^2$$

$$\leqslant C(n^{-1}\log n)^{2\alpha/(1+2\alpha)} + 10 \sum_{j=j_0}^{J-1} \sum_k a_{jk}^2 + 10 \sum_{j=j_0}^{J-1} \sum_k b_{jk}^2. \tag{13}$$

It follows from Proposition 1 and Lemma 3 that

$$\sum_k \tilde{a}_{j_0 k}^2 + \sum_{j=j_0}^{J-1} \sum_k a_{jk}^2 = \sum_i \left(n^{-1/2} f\left(\frac{i}{n+1}\right) - \xi_{Ji} \right)^2 \leqslant Cn^{-2s(\alpha)}, \tag{14}$$

$$E_x \left(\sum_k \tilde{b}_{j_0 k}^2 + \sum_{j=j_0}^{J-1} \sum_k b_{jk}^2 \right) = \frac{1}{n} \sum_{i=1}^n E_x \left(f(x_{(i)}) - f\left(\frac{i}{n+1}\right) \right)^2 \leqslant Cn^{-s(\alpha)}. \tag{15}$$

For $\alpha \geqslant 1/2$, $s(\alpha) \geqslant 2\alpha/(1+2\alpha)$. Now it follows from (11)–(15) that

$$E \parallel \hat{f}_* - f \parallel_2^2 \leqslant C \left(\frac{\log n}{n} \right)^{2\alpha/(1+2\alpha)}. \qquad \square$$

Acknowledgements

Appendix A.

The four test functions (Fig. 1) represent different degrees of spatial variability. The functions are normalized so that every function has standard deviation 10. Formulae of the test functions can be found in Donoho and Johnstone (1994).

References

Cai, T., Brown, L.D., 1997. Wavelet regression for random uniform design. Technical Report #97-15, Department of Statistics, Purdue University.

Cai, T., Brown, L.D., 1998. Wavelet shrinkage for nonequispaced samples. Ann. Statist., in press.

Cai, T., 1996. Nonparametric function estimation via wavelets. Ph.D Thesis, Cornell University.

Daubechies, I., 1992. Ten Lectures on Wavelets. SIAM, Philadelphia.

Donoho, D.L., Johnstone, I.M., 1994. Ideal spatial adaptation via wavelet shrinkage. Biometrika 81, 425–455.

Donoho, D.L., Johnstone, I.M., Kerkyacharian, G., Picard, D., 1995. Wavelet shrinkage: asymptopia? J. Roy. Stat. Soc. Ser. B 57, 301–369.

Hall, P., Turlach, B.A., 1997. Interpolation methods for nonlinear wavelet regression with irregularly spaced design. Ann. Statist. 25, 1912–1925.

Meyer, Y., 1990. Ondelettes et Opérateurs: I. Ondelettes. Hermann et Cie, Paris.