BLOCK THRESHOLDING AND WAVELET ESTIMATION FOR NONEQUISPACED SAMPLES

by

Eric Chicken
Purdue University

Technical Report # 01-05

Department of Statistics
Purdue University
West Lafayette, IN USA

March 2001

Block Thresholding and Wavelet Estimation for Nonequispaced Samples

Eric Chicken
Purdue University

Abstract

For samples with the design points occuring as a Poisson process or having a uniform distribution, the wavelet method of block thresholding can be applied directly to the data as though it was equispaced without sacrificing adaptivity or optimality. When the underlying true function is in certain Besov and Hölder classes, the resulting estimator achieves the minimax rate of convergence. Simulation results are examined.

1. Introduction

Wavelets have been shown to be very successful in nonparametric function estimation. Specifically, they excel in the areas of spatial adaptivity, optimality, and low computational cost. Typically, wavelet analysis is performed through the use of term-by-term thresholding of wavelet coefficients, such as the VisuShrink method of Donoho and Johnstone (1994). There, a noisy signal is transformed into empirical wavelet coefficients by the discrete wavelet transform, these coefficients are denoised by comparison with a specified thresholding rule, and the underlying function is estimated by applying the inverse discrete wavelet transform to these denoised coefficients. This method is adaptive and is within a logarithmic factor of the optimal minimax convergence rate over large classes of Besov functions.

Hall et al. (1999) proposed a method of wavelet analysis whereby the optimal minimax convergence rate is attained without the logarithm penalty found in the term-by-term methods. Using block thresholding, where the empirical wavelet coefficients are thresholded in blocks rather than individually, they achieved this optimal rate and maintained adaptivity over a large class of Hölder functions. By looking at coefficients in blocks rather than individually, more precise comparisons between the coefficients and the threshold is allowed, resulting in improved rates. Cai (1998) has extended this idea. Using a James-Stein thresholding rule, he has shown that this block-thresholded wavelet estimator attains the optimal convergence rate from both a global and local estimation perspective. A specific block length and threshold value are established that attain these rates. Additionally, this estimate is easy to implement and has a low computational cost of $O(n)$.

MSC2000: Primary 62G07; secondary 62G20.

Key words and phrases. Wavelets, nonparametric regression, adaptive, Besov class, Hölder class, Poisson process.

The above methods have been developed for data that is equispaced. Little emphasis has been placed on sample data that is not equispaced, however. Cai and Brown (1998) investigated wavelet methods on samples with fixed, nonequispaced designs via an approximation approach. They showed that applying the methods devised for equispaced data directly to nonequispaced data can lead to suboptimal estimators. They then proposed a method that was adaptive and near optimal. Hall and Turlach (1997) used interpolation methods to deal with samples with random design. Unfortunately, these methods are much more complex from a computational standpoint than their equispaced counterparts.

Cai and Brown (1999) have also examined convergence rates when the unknown function is in a Hölder class with exponent α and the positions of the sample points are distributed as independent uniform random variables. Using term-by-term thresholding, they showed that the equispaced wavelet method can be directly applied to the nonequispaced data without a loss in the rate of convergence, i.e., to within a logarithmic factor of the optimal convergence rate of $n^{-\frac{2\alpha}{2\alpha+1}}$. This method maintains the computational efficiency and simplicity of the equispaced algorithm.

In this paper, it is shown that these results of Cai and Brown (1999) for uniform design can be improved upon for Hölder classes and extended to many Besov classes through the use of block thresholding. Additionally, when the sample points occur as a Poisson process, the same optimal rates are attained. Adaptivity is kept, and the computational cost remains low since the equispaced algorithm is used.

In Section 2 of this paper, the method of thresholding the wavelet coefficients is stated, and the theorems on convergence rates are put forth. Section 3 discusses simulation results and section 4 contains the proofs.

2. Methodology

2.1 Wavelets

ϕ and ψ will represent the father and mother wavelets, respectively. Both are assumed to be compactly supported. Let ϕ_{jk} and ψ_{jk} be the translations and dilations of ϕ and ψ:

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

In this paper, wavelets periodized to the interval $[0,1]$ will be used. The set

$$\{\phi^p_{j_0k} : k = 1, ..., 2^{j_0}\} \cup \{\psi^p_{jk} : j \geq j_0, k = 1, ..., 2^j\}$$

is an orthonormal basis of $L^2[0,1]$, where

$$\begin{aligned} \phi^p_{jk}(x) &= \sum_{l=-\infty}^{\infty} \phi_{jk}(x-l), \\ \psi^p_{jk}(x) &= \sum_{l=-\infty}^{\infty} \psi_{jk}(x-l), \end{aligned}$$

for x in $[0,1]$. These periodized wavelets will be used for the rest of the paper with the superscript suppressed. For further simplicity, the functions to be estimated will be assumed to have domain of $[0,1]$, also. Minimal changes are necessry if the period of the wavelets or the support of the functions is changed to something other than $[0,1]$.

The wavelet coefficients for the discrete wavelet transform of a function f are the usual inner product:

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

$$\theta_{jk} = \langle f, \psi_{jk} \rangle.$$

Hence, f can be expressed as an infinite series

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

In terms of resolution, the $\xi_{j_0 k}$ coefficients represent the coarsest, smoothest portions of f, and the θ_{jk} are the coefficients representing the detailed structure of f.

2.2 The Estimator

Suppose a noisy signal is received with an unknown underlying function f. We assume the noise is normally distributed with mean zero. The observed signal gives rise to a bivariate vector consisting of the signal arrival time and the signal's value at this time:

$$\{(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)\},$$

where y_i is

$$y_i = f(x_i) + \sigma \varepsilon_i, \tag{1}$$

$i = 1, 2, \ldots, n$, and the ε_i's are independent, standard normal random variables. The parameter σ is constant and known. In much of the literature, n is constant and the x_i's are equidistant apart. For example, $x_i = \frac{i}{n}$, $i = 1, 2, \ldots, n$. Here, two cases will be considered. First, the sample points are uniformly distributed over the interval $[0,1]$ and independent of the ε_i's, and the number of points in the interval is fixed. Second, the sample points follow a Poisson process and independent of the ε_i's, and the number of points in the interval is random.

When estimating f by $\hat{f}$, the goal is to find an estimate with a small mean integrated squared error:

$$E\|f - \hat{f}\|_2^2.$$

By ordering the sample points and performing the equivalent reordering of the y_i's and ε_i's, the signal now looks like

$$y_i = f(s_i) + \sigma \varepsilon_i.$$

In the uniform case, the s_i are the ordered, uniform random variables $x_{(i)}$, and $i = 1, 2, \ldots, n$. In the Poisson process case the s_i are the arrival times, and $i = 1, 2, \ldots, N$,

where N is the number of arrivals in $[0,1]$. Given $N = n$, the arrival times are distributed as ordered uniform [0,1] random variables, so the following discussion will be restricted to this case.

The observed data vector is now (using the appropriately relabeled y_i's)

$$\{(x_{(1)}, y_1), (x_{(2)}, y_2), \ldots, (x_{(n)}, y_n)\}.$$

These randomly spaced data points will be treated as equidistant in the wavelet algorithm. Each $x_{(i)}$ will be replaced with its expected value, $\frac{i}{n+1}$. The data vector can then be viewed as though it were sampled at the equidistant points

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

Provided that $n = 2^J$, the wavelet algorithm can be applied to the vector $y = \{y_1, y_2, \ldots, y_n\}$. Since in the uniform design case the number of points is known, n can be assumed to be of this form. In the Poisson case, if $n \neq 2^J$, the signal will be extended to the next multiple of 2 by reflecting the signal abouts its endpoint. The underlying, unknown function is then assumed to be extended similary on $[1, 1 + \frac{k}{n}]$, where $k = 2^J - n$ is the number of points added to the signal.

Let

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

be the discrete wavelet transform of $n^{-1/2}y$. The $\tilde{\xi}_{j_0 k}$ are the empirical wavelet coefficients of the father wavelet corresponding to the coarsest resolution level. They will not be thresholded. The $\tilde{\theta}_{jk}$ are the empirical wavelet coefficients from the mother wavelet, and represent the detailed structure of the wavelet expansion of f.

Usually, the thresholding at this point is done on a term-by-term basis. Each coefficient is compared against some threshold and the coefficient is kept and scaled or discarded (set to 0) depending on whether it is larger or smaller than the thresholding value. For example, in Donoho and Johnstone's VisuShrink algorithm with the soft threshold, each coefficient is compared to $\sigma\sqrt{2n^{-1}\log n}$. This method has its drawbacks, however. The size of the optimal threshold is $n^{-\frac{1}{2}}$. But, the error in estimating the coefficient is also of size $n^{-\frac{1}{2}}$. Therefore, accurate thresholding with the optimal threshold is not possible. This difficulty is avoided by introducing the logarithm term seen in the soft threshold above. Unfortunately, this introduces a factor of $\log n$ in the convergence rate. See Hall et al. (1998).

Instead of term-by-term thresholding, the $\tilde{\theta}_{jk}$ are thresholded in blocks using the James-Stein threshold. At each resolution level j, the $\tilde{\theta}_{jk}$ are grouped into non-overlapping blocks of length L. Let B_{jb} be the indices of the $\tilde{\theta}_{jk}$ in the bth block in resolution level j. The James-Stein threshold rule is

$$\hat{\theta}_{jk} = \left(1 - \frac{\lambda L \sigma^2}{n \sum_{i \in B_{jb}} \tilde{\theta}_{ji}^2}\right)_+ \cdot \tilde{\theta}_{jk},$$

for all k in the bth block of resolution level j. In effect, this threshold compares the average bias of the coefficients in the block with a multiple of the variance of a coefficient.

It keeps all coefficients in a block (after scaling them) whenever their average bias is larger than the variance of the coefficient, otherwise the entire block's coefficients are set to 0. By pooling the information of neighboring coefficents together, a better decision is made in terms of which components to retain. This added information allows the improvement in the convergence rates. Cai (1998) has shown that the optimal rate of convergence for equispaced samples is attained when $L = \log(n)$ and $\lambda = 4.50524$.

The function f (or the extended version of f in the Poisson case) is then estimated by

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

In the Poisson process case, N could be any non-negative integer. To avoid computational problems, the estimate of f above is valid for any $N \in \{5, 6, \ldots, 2\mu\}$, where μ is the rate of the process. For other values of N, we set $\hat{f}$ to 0. The upper limit may be arbitrarily raised without affecting the results of the following theorem. The lower limit of 5 is set in order to avoid calculating the trivial wavelet transform on the data when using the Symmlet "s8" wavelet. Other wavelets will require modifications of this lower bound.

The rate of convergence of $\hat{f}$ to f is the main result of this paper.

Theorem 1 *Suppose the sample points follow a Poisson process with rate $\mu > 0$ and the sample $\{(x_1, y_1), (x_2, y_2), \ldots\}$ is collected with $y_i = f(x_i) + \sigma\varepsilon_i$ as in (1) above. Further, ψ has r vanishing moments, $\alpha \in [\frac{1}{2}, r]$, $2 \leq p \leq \infty$, $1 \leq q \leq \infty$, and $0 < M < \infty$. Then for $\hat{f}$ as constructed above,*

$$\sup_{f \in F^{\alpha}_{p,q}(M)} E\|\hat{f} - f\|_2^2 \leq C\mu^{-\frac{2\alpha}{2\alpha+1}}.$$

Theorem 1 impies that for all functions in $F^{\alpha}_{p,q}(M)$, to be defined in the section 2.3, data sampled as a Poisson process may be treated as equispaced data without a loss in the rate of convergence. An easy consequence of theorem 1 is the following ($\Lambda^{\alpha}(M)$ will be defined in section 2.3):

Corollary 1 *Under the conditions of theorem 1,*

$$\sup_{f \in \Lambda^{\alpha}(M)} E\|\hat{f} - f\|_2^2 \leq C\mu^{-\frac{2\alpha}{2\alpha+1}}.$$

If the samples are not following the Poisson process, but are instead uniformly spaced throughtout the interval [0,1], similar results hold:

Theorem 2 *Suppose the sample $\{(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)\}$ is collected with $y_i = f(x_i) + \sigma\varepsilon_i$ as in (1) above. Further, ψ has r vanishing moments, $\alpha \in [\frac{1}{2}, r]$, $2 \leq p \leq \infty$, $1 \leq q \leq \infty$, and $0 < M < \infty$. Then for $\hat{f}$ as constructed above,*

$$\sup_{f \in F^{\alpha}_{p,q}(M)} E\|\hat{f} - f\|_2^2 \leq Cn^{-\frac{2\alpha}{2\alpha+1}}.$$

Corollary 2 *Under the conditions of theorem 2,*

$$\sup_{f\in\Lambda^\alpha(M)} E\|\hat{f}-f\|_2^2 \le Cn^{-\frac{2\alpha}{2\alpha+1}}.$$

Corollary 2 shows that by using block thresholding on Hölder spaces, the rate of convergence has increased in comparison to the term-by-term thresholding method of Cai and Brown (1999). In that case, the rate is $O\left(\left(\frac{\log n}{n}\right)^{\frac{2\alpha}{2\alpha+1}}\right)$.

2.3 Function Spaces

In order to define the function space $F^\alpha_{p,q}(M)$, two other spaces are needed. The first, and more general, of the two is the Besov space $B^\alpha_{p,q}$, where $0 < p,q \le \infty$ and $\alpha > 0$. A function f is said to be in this space if its Besov norm is finite:

$$\|f\|_{B^\alpha_{p,q}} < \infty,$$

where, for $0 < \alpha \le 1$,

$$\|f\|_{B^\alpha_{p,q}} = \|f\|_{L^p} + \begin{cases} \left(\int_0^\infty \frac{1}{h}\left(\frac{1}{h^\alpha}\|f(\cdot+h)-f(\cdot)\|_{L^p}\right)^q dh\right)^{\frac{1}{q}}, & q<\infty, \\ \sup_{0<h}\frac{\|f(\cdot+h)-f(\cdot)\|_{L^p}}{h^\alpha}, & q=\infty. \end{cases}$$

For $\alpha > 1$, $\alpha = \lfloor\alpha\rfloor + s, 0 < s \le 1$,

$$\|f\|_{B^\alpha_{p,q}} = \sum_{m=0}^{\lfloor\alpha\rfloor}\|f^{(m)}\|_{B^s_{p,q}}.$$

The Hölder space Λ^α is a special case of a Besov space. Specifically, $\Lambda^\alpha = B^\alpha_{\infty,\infty}$. This paper considers functions whose Besov (or Hölder) norms are bounded. For any $0 < M < \infty$, we define the Besov or Hölder ball as:

$$B^\alpha_{p,q}(M) = \left\{f : \|f\|_{B^\alpha_{p,q}} \le M\right\},$$

$$\Lambda^\alpha(M) = \left\{f : \|f\|_{\Lambda^\alpha} \le M\right\}.$$

Wavelet coefficients for functions in Besov spaces have the property that the Besov norm of the function f can be represented as a sequence norm in terms of its wavelet coefficients (see Meyer (1990)). If $f \in B^\alpha_{p,q}(M)$ and $\Theta = \left\{\tilde{\xi}_{j_0,1},\ldots,\tilde{\xi}_{j_0,2^{j_0}},\tilde{\theta}_{j_0,1},\ldots\right\}'$ is the vector of wavelet coefficents of f, then

$$\|f\|_{B^\alpha_{p,q}} = \begin{cases} \left(\sum_{j=j_0}^\infty\left(2^{j\left(\alpha+\frac{1}{2}-\frac{1}{p}\right)}\left(\sum_{k=1}^{2^j}|\theta_{jk}|^p\right)^{\frac{1}{p}}\right)^q\right)^{\frac{1}{q}}, & q<\infty \\ \sup_{j\ge j_0} 2^{j\left(\alpha+\frac{1}{2}-\frac{1}{p}\right)}\left(\sum_{k=1}^{2^j}\mid\theta_{jk}\mid^p\right)^{\frac{1}{p}}, & q=\infty. \end{cases}$$

When dealing with the coefficients of a Besov function, we will use the notation $\Theta^{\alpha}_{p,q}(M)$ to refer to the space.

Define $F^{\alpha}_{p,q}(M)$ to be

$$F^{\alpha}_{p,q}(M) = B^{\alpha}_{p,q}(M) \cap \Lambda^{\frac{2\alpha}{2\alpha+1}}(M).$$

Since $B^{\alpha}_{p,q}(M) \subseteq \Lambda^{\alpha-1/p}(M)$, when $\alpha > \frac{1}{p}$, the assumption that $F^{\alpha}_{p,q}(M)$ be in $\Lambda^{\frac{2\alpha}{2\alpha+1}}(M)$ only requires an increase in the Hölder exponent of

$$\left(\frac{1}{p} - \frac{2\alpha^2 - \alpha}{2\alpha + 1}\right)_+.$$

So for many cases, the assumption that f be in $\Lambda^{\frac{2\alpha}{2\alpha+1}}(M)$ is redundant.

3. Simulation

To compare the efficiency of the block estimator under the different sampling designs, it was run on equispaced, uniform, and Poisson process samples and the results were compared. Also, the block estimator was compared to the VisuShrink estimator with universal term thresholding on uniform distributed sample points.

The test functions used are those specified in Donoho and Johnstone (1994). These eight functions represent varying degrees of spatial inhomogeneity. Sample sizes ranged from $n = 512$ to $n = 8192$, and a signal-to-noise ratio (SNR) of 5 was simulated. For the Poisson case, simlulations were run for μ taking the same values as n above. Each of the functions has been normalized to give a standard deviation of 10 before noise was added. The MSE was estimated using 200 replications. In the Poisson case, this is equivalent to observing the process for 200 intervals, and estimating the function on each interval. The Symmlet "s8" wavelet was used. The block length was set to the greatest dyadic integer less than or equal to $\log(n)$, and the number of levels was fixed to ensure that the block length would evenly divide the number of coefficients at each level. Additionally, the number of coefficients at the coarsest level is the same for all n. Plots and formulae of the test functions are in appendix A.

Table 1 shows the results of the simulations. For the block estimator, the uniformly spaced estimate (BU) and the equispaced estimate (B) are not significantly different ($\alpha = 0.05$ significance level) in five of the 40 test cases. In two of the cases, the uniformly spaced estimate outperformed the equispaced one, and in the remaining cases, the equispaced estimate is better. In no instance does the uniformly spaced estimate have more than twice the MSE of the equispaced one. And, in general, the percentage difference between these two estimates tends to decrease as n increases. The block estimator performed least well on the "dopppler" function.

The estimator performs equivalently on the Poisson (BP) and uniform (BU) designs with the exception of the "heavisine" and "doppler" functions. Here, the MSE for the Poisson case is never more than 11% above the uniform case. This is not surprising since the conditional Poisson samples are equivalent to the uniform samples. The main differences noted for "heavisine" and "doppler" are most likely due to a few instances of low numbers of events in some of the 200 intervals simulated.

The block estimator (BU) compares quite favorably to the VisuShrink estimator (VU) for uniform distributed sample points. The block estimate has a lower MSE in 36 of the 40 cases and a higher one in four cases. For all test functions, the percentage difference between the estimators increases in favor of the block estimator as n increases.

The block estimate is generally better than the VisuShrink estimate on equispaced samples. The block estimate is better in 37 of 40 cases, not significantly different from VisuShrink in 1 case, and worse in 2 cases ("corner" and "heavisine" functions). In these two instances of VisuShrink outperforming the block estimator, it is better by no more than 7%.

In the table, the difference is set to zero when the MSE's are not significantly different at the 95% confidence level.

Figure 1 shows some examples of the reconstructions. The dashed line is the actual function, the solid line is the estimate. The visual quality of the reconstructions is comparable for both cases, though the uniformly spaced reconstruction shows more oscillating behavior than the other due to the uneveness of the sample point placement. For these functions, $n = 1024$ and a signal-to-noise ratio of 5 was used.

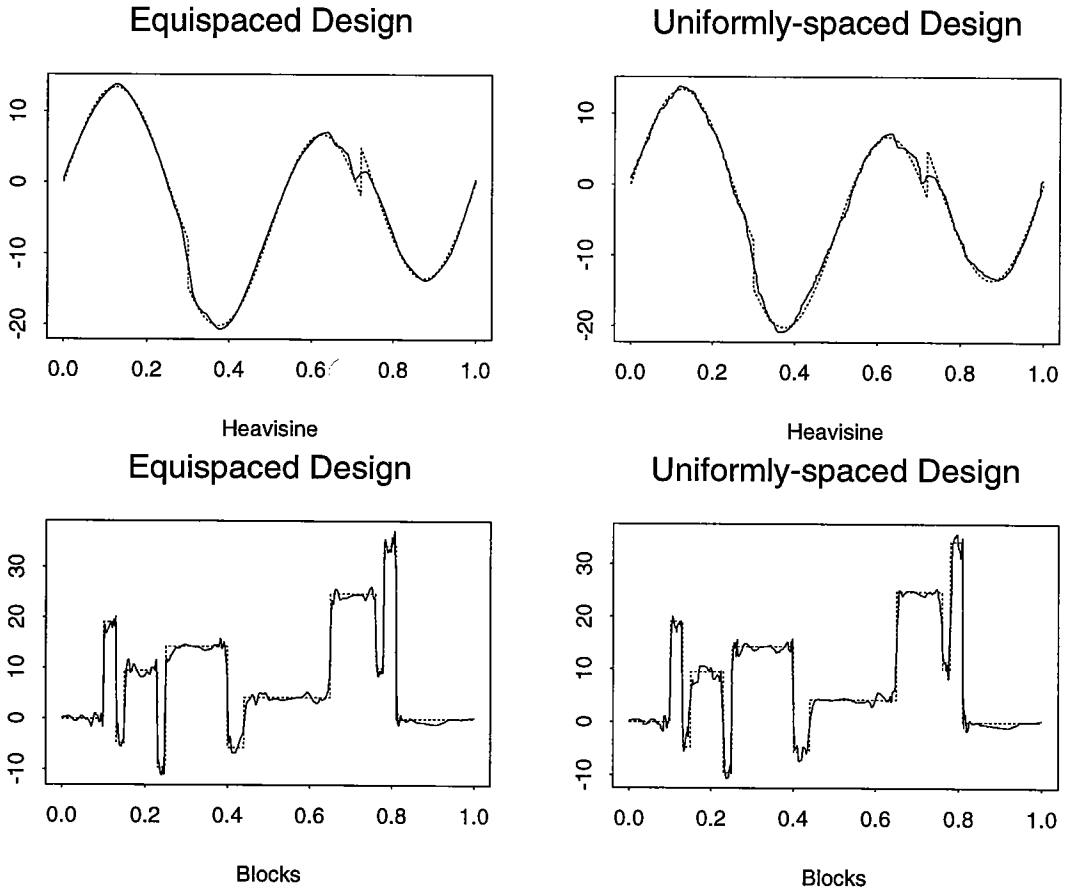

Figure 1: Reconstructions of Functions.

Function n	B	BU	BP	V	VU	B vs BU	BU vs VU	B vs BP	BU vs BP	B vs V
Blip										
512	0.60	0.63	0.63	0.86	0.91	5%	-31%	5%	0%	43%
1024	0.34	0.35	0.36	0.59	0.61	5%	-42%	7%	0%	75%
2048	0.20	0.21	0.21	0.37	0.41	9%	-49%	5%	0%	91%
4096	0.11	0.11	0.12	0.25	0.27	8%	-58%	10%	0%	139%
8192	0.06	0.06	0.06	0.17	0.17	0%	-64%	4%	0%	170%
Blocks										
512	3.02	3.02	3.05	5.27	5.33	0%	-43%	0%	0%	74%
1024	1.85	1.85	1.83	3.69	3.85	0%	-52%	0%	0%	100%
2048	1.15	1.09	1.07	2.56	2.67	-5%	-59%	-7%	0%	123%
4096	0.64	0.64	0.64	1.72	1.82	0%	-65%	0%	0%	169%
8192	0.37	0.37	0.37	1.16	1.20	-2%	-70%	-1%	0%	211%
Bumps										
512	3.15	3.24	3.29	9.43	8.33	0%	-61%	4%	0%	199%
1024	1.66	1.98	2.01	5.92	6.03	20%	-67%	21%	0%	258%
2048	0.93	1.21	1.20	3.78	4.01	31%	-70%	29%	0%	307%
4096	0.51	0.70	0.70	2.30	2.51	38%	-72%	38%	0%	354%
8192	0.29	0.38	0.39	1.40	1.51	35%	-75%	35%	0%	392%
Corner										
512	0.40	0.51	0.51	0.38	0.44	27%	17%	27%	0%	-6%
1024	0.20	0.28	0.28	0.22	0.25	39%	10%	38%	0%	8%
2048	0.10	0.14	0.14	0.12	0.15	47%	-5%	46%	0%	27%
4096	0.05	0.07	0.08	0.08	0.10	52%	-25%	57%	0%	60%
8192	0.03	0.04	0.04	0.05	0.06	40%	-36%	48%	6%	73%
Doppler										
512	1.07	2.13	2.11	2.85	3.90	99%	-45%	98%	0%	166%
1024	0.68	1.21	1.21	1.88	2.54	78%	-52%	78%	0%	177%
2048	0.38	0.65	0.67	1.20	1.59	70%	-59%	73%	0%	212%
4096	0.22	0.34	0.34	0.80	0.99	56%	-65%	56%	0%	263%
8192	0.12	0.18	0.18	0.51	0.60	51%	-70%	51%	0%	335%
Heavisine										
512	0.59	0.74	0.82	0.55	0.62	26%	19%	40%	11%	-7%
1024	0.40	0.45	0.48	0.40	0.43	14%	6%	21%	7%	0%
2048	0.24	0.26	0.28	0.29	0.28	6%	-9%	16%	10%	21%
4096	0.14	0.15	0.15	0.21	0.20	5%	-24%	9%	4%	46%
8192	0.08	0.08	0.09	0.14	0.13	7%	-37%	11%	3%	72%
Spikes										
512	1.05	1.75	1.80	2.07	3.34	67%	-48%	71%	0%	97%
1024	0.56	0.96	0.96	1.38	2.11	72%	-55%	72%	0%	148%
2048	0.33	0.50	0.50	0.90	1.28	50%	-61%	50%	0%	173%
4096	0.18	0.26	0.26	0.59	0.75	38%	-66%	38%	0%	220%
8192	0.09	0.13	0.13	0.38	0.44	39%	-71%	40%	0%	315%
Wave										
512	0.98	1.46	1.49	2.23	2.56	50%	-43%	52%	0%	128%
1024	0.39	0.66	0.71	1.47	1.65	70%	-60%	83%	8%	281%
2048	0.18	0.31	0.34	0.85	1.00	69%	-69%	87%	10%	369%
4096	0.10	0.16	0.17	0.48	0.59	61%	-73%	72%	7%	389%
8192	0.06	0.09	0.10	0.27	0.34	46%	-73%	54%	5%	339%

Table 1: Mean-squared errors from 200 replications.

4. Proof

Before the main theorems are proved, some preliminaries are necessary. First, a result about the order statistics of the uniformly spaced sample points is stated.

Lemma 1 *Let x_i be independent, uniform(0,1) random variables. Let $x_{(1)} < x_{(2)} < \ldots < x_{(n)}$ be their order statistics. Then $x_{(k)}$ is a Beta$(k, n-k+1)$ random variable with mean $\frac{k}{n+1}$ and variance $\frac{(n+1)k-k^2}{(n+1)^2(n+2)}$.*

The following result from Cai (1998) will also be needed.

Lemma 2 *Suppose $\theta_{jk} \in \Theta^{\alpha}_{p,q}(M)$ and*

$$y_{jk} = \theta_{jk} + \frac{1}{\sqrt{n}}\sigma\varepsilon_{jk}, j \geq j_0, k = 1, 2, \ldots, 2^j,$$

where z_{jk} are iid standard normal random variables. Let

$$\hat{\theta}_{jk} = \left(1 - \frac{\lambda L\sigma^2}{n\sum_{i\in B_{jb}} y_{ji}^2}\right)_+ \cdot y_{jk},$$

be the result of applying the James-Stein threshold rule to the y_{jk}. Then

$$E\|\hat{\theta} - \theta\|_2^2 \leq Cn^{-\frac{2\alpha}{2\alpha+1}}.$$

Lemma 3 *Let f be in $B^{\alpha}_{p,q}(M), p \geq 2, q \geq 1$ and θ_{jk} be the wavelet coefficents. Then*

$$\sum_{j=J}^{\infty}\sum_{k=1}^{2^j} \theta_{jk}^2 \leq Cn^{-2\alpha}.$$

Proof: Using the Besov sequence norm and the inequality

$$\|x\|_u \leq m^{1/u-1/v}\|x\|_v$$

for $x \in R^m$, $u \leq v$, we have

$$\sum_{k=1}^{2^j} \theta_{jk}^2 = \|\theta_{j\cdot}\|_2^2 \leq \left[2^{j\left(\frac{1}{2}-\frac{1}{p}\right)}\|\theta_{j\cdot}\|_p\right]^2.$$

This implies

$$\sum_{j=J}^{\infty}\sum_{k=1}^{2^j} \theta_{jk}^2 \leq M^2\sum_{j=J}^{\infty} 2^{2j\left(\frac{1}{2}-\frac{1}{p}\right)}2^{-2j\left(\alpha+\frac{1}{2}-\frac{1}{p}\right)} \leq M^2 2^{-2J\alpha}.$$

Since $2^J = n$, the proof is finished.

Proof of theorem 2. Let $\tilde{f}(x) = \sum_{i=1}^{2^J} \frac{1}{\sqrt{n}} y_i \phi_{Ji}(x)$ where y_i are as in (1). Then

$$\tilde{f}(x) = \sum_{i=1}^{2^J} \left(\xi_{Ji} + \left(\frac{1}{\sqrt{n}} f(x_{(i)}) - \xi_{Ji} \right) + \frac{1}{\sqrt{n}} \sigma \varepsilon_i \right) \phi_{Ji}(x)$$

$$= \sum_{k=1}^{2^{j_0}} \left(\xi_{j_0 k} + \nu_{j_0 k} + \eta_{j_0 k} \right) \phi_{j_0 k}(x) + \sum_{j=j_0}^{\infty} \sum_{k=1}^{\infty} \left(\theta_{jk} + \zeta_{jk} + \delta_{jk} \right) \psi_{jk}(x),$$

where $\xi_{j_0 k}$ and θ_{jk} are the coefficients for $\sum_{i=1}^{2^J} \xi_{j_0 k} \phi_{Ji}(x)$, $\nu_{j_0 k}$ and ζ_{jk} are the coefficients for $\sum_{i=1}^{2^J} \left(\frac{1}{\sqrt{n}} f(x_{(i)}) - \xi_{Ji} \right) \phi_{Ji}(x)$, and $\eta_{j_0 k}$ and δ_{jk} are the coefficients for $\sum_{i=1}^{2^J} \frac{1}{\sqrt{n}} \sigma \varepsilon_i \phi_{Ji}(x)$.

Let $\tilde{\xi}_{j_0 k} = \xi_{j_0 k} + \nu_{j_0 k} + \eta_{j_0 k}$ be the coarse coefficients. In our estimate of f, these will not be thresholded. Let $\theta'_{jk} = \theta_{jk} + \zeta_{jk} + \delta_{jk}$. Since the wavelet transform is an orthogonal transform when n is a power of 2, $\theta'_{jk} \sim \mathrm{N}(\theta_{jk} + \zeta_{jk}, \frac{\sigma^2}{n})$. Applying the James-Stein threshold with optimal λ to θ'_{jk} yields $\hat{\theta}_{jk}$. The estimate of f then is

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

and the error is

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

From lemma 3,

$$\sum_{j=J}^{\infty} \sum_{k=1}^{2^j} \theta_{jk}^2 \leq C n^{-2\alpha} \leq C n^{-\frac{2\alpha}{2\alpha+1}}. \tag{2}$$

From lemma 2, with $\widetilde{X} = (x_{(1)}, x_{(2)}, \ldots, x_{(n)})$,

$$\sum_{j=j_0}^{J-1} \sum_{k=1}^{2^j} E(\hat{\theta}_{jk} - \theta_{jk})^2 \leq 2 E_{(\widetilde{X})} E_{(\cdot|\widetilde{X})} \sum_{j=j_0}^{J-1} \sum_{k=1}^{2^j} \left[(\hat{\theta}_{jk} - (\theta_{jk} + \zeta_{jk}))^2 + \zeta_{jk}^2 \right]$$

$$\leq 2E \| \hat{\theta}_{jk} - (\theta_{jk} + \zeta_{jk}) \|_2^2 + 2E_{(\widetilde{X})} \sum_{j=j_0}^{J-1} \sum_{k=1}^{2^j} \zeta_{jk}^2 \tag{3}$$

$$\leq C n^{-\frac{2\alpha}{2\alpha+1}} + 2E_{(\widetilde{X})} \sum_{j=j_0}^{J-1} \sum_{k=1}^{2^j} \zeta_{jk}^2.$$

And, since $\eta_{j_o k} \sim \mathrm{N}(0, \frac{\sigma^2}{n})$ due to the othogonality of the wavelet transform,

$$\begin{aligned}\sum_{k=1}^{2^{j_o}} E(\tilde{\xi}_{j_o k} - \xi_{j_o k})^2 &= \sum_{k=1}^{2^{j_o}} E_{(\tilde{X})} E_{(\cdot|\tilde{X})} (\nu_{j_o k} + \eta_{j_o k})^2 \\ &= 2^{j_o} \frac{\sigma^2}{n} + E_{(\tilde{X})} \sum_{k=1}^{2^{j_o}} (\nu_{j_o k})^2. \end{aligned} \tag{4}$$

Using (2), (3) and (4),

$$E\|\hat{f} - f\|_2^2 \leq C n^{-\frac{2\alpha}{2\alpha+1}} + C E_{(\tilde{X})} \left[\sum_{k=1}^{2^{j_0}} (\nu_{j_0 k})^2 + \sum_{j=j_0}^{J-1} \sum_{k=1}^{2^j} \zeta_{jk}^2 \right]. \tag{5}$$

Since $\sum_{k=1}^{2^{j_0}} (\nu_{j_0 k})^2 + \sum_{j=j_0}^{J-1} \sum_{k=1}^{2^j} \zeta_{jk}^2 = \sum_{i=1}^{2^J} \left(\frac{1}{\sqrt{n}} f(x_{(i)}) - \xi_{Ji} \right)^2$,(5) becomes

$$E\|\hat{f} - f\|_2^2 \leq C n^{-\frac{2\alpha}{2\alpha+1}} + C E_{(\tilde{X})} \sum_{i=1}^{2^J} \left(\frac{1}{\sqrt{n}} f(x_{(i)}) - \xi_{Ji} \right)^2. \tag{6}$$

The theorem will be proved once we have shown the following lemma.

Lemma 4 *If $f \in F_{p,q}^{\alpha}(M)$, $\alpha \in [\frac{1}{2}, r]$, $2 \leq p \leq \infty$, and $1 \leq q \leq \infty$, then*

$$E_{(\tilde{X})} \sum_{i=1}^{2^J} \left(\frac{1}{\sqrt{n}} f(x_{(i)}) - \xi_{Ji} \right)^2 \leq C \left(\frac{1}{n} \right)^{\frac{2\alpha}{2\alpha+1}}.$$

Proof: Assume the range of ϕ is contained in $[-R, R]$ and $q = \infty$. Then

$$\begin{aligned} \left| \frac{1}{\sqrt{n}} f(x_{(i)}) - \xi_{Ji} \right| &\leq \left| \frac{1}{\sqrt{n}} f(x_{(i)}) - \int_{-\frac{R}{n}+\frac{i}{n}}^{\frac{R}{n}+\frac{i}{n}} f(y) \phi_{Ji}(y) dy \right| \\ &= \left| \int_{-\frac{R}{n}+\frac{i}{n}}^{\frac{R}{n}+\frac{i}{n}} (f(x_{(i)}) - f(y)) \phi_{Ji}(y) dy \right| \\ &\leq \sqrt{\left\{ \int_{-\frac{R}{n}+\frac{i}{n}}^{\frac{R}{n}+\frac{i}{n}} (f(x_{(i)}) - f(y))^2 dy \right\} \left\{ \int_{-\frac{R}{n}+\frac{i}{n}}^{\frac{R}{n}+\frac{i}{n}} \phi_{Ji}^2(y) dy \right\}} \\ &= \sqrt{\int_{-\frac{R}{n}+\frac{i}{n(n+1)}}^{\frac{R}{n}+\frac{i}{n(n+1)}} \left[f(x_{(i)}) - f\left(y + \frac{i}{n+1} \right) \right]^2 dy}. \end{aligned}$$

Therefore,

$$E_{(\tilde{X})}\sum_{i=1}^{2^J}\left(\frac{1}{\sqrt{n}}f(x_{(i)})-\xi_{Ji}\right)^2$$

$$\leq E_{(\tilde{X})}\sum_{i=1}^{2^J}\int_{-\frac{R}{n}+\frac{i}{n(n+1)}}^{\frac{R}{n}+\frac{i}{n(n+1)}}\left[f(x_{(i)})-f\left(y+\frac{i}{n+1}\right)\right]^2 dy$$

$$\leq 2E_{(\tilde{X})}\sum_{i=1}^{2^J}\int_{-\frac{R}{n}+\frac{i}{n(n+1)}}^{\frac{R}{n}+\frac{i}{n(n+1)}}\left[f(x_{(i)})-f(y+x_{(i)})\right]^2 dy \tag{7}$$

$$+\,2E_{(\tilde{X})}\sum_{i=1}^{2^J}\int_{-\frac{R}{n}+\frac{i}{n(n+1)}}^{\frac{R}{n}+\frac{i}{n(n+1)}}\left[f(x_{(i)}+y)-f\left(y+\frac{i}{n+1}\right)\right]^2 dy.$$

To bound the first term on the right side of (7), note that

$$E_{(\tilde{X})}\sum_{i=1}^{2^J}\int_{-\frac{R}{n}+\frac{i}{n(n+1)}}^{\frac{R}{n}+\frac{i}{n(n+1)}}\left[f(x_{(i)})-f(y+x_{(i)})\right]^2 dy\leq\sum_{i=1}^{2^J}\int_0^1\int_{-\frac{R+1}{n}}^{\frac{R+1}{n}}\left[f(x_i)-f(y+x_i)\right]^2 dydx_i,$$

where the x_i are now independent uniform$(0,1)$ random variables, no longer ordered. Applying Fubini's theorem, and using the properties of the Besov function norm,

$$\sum_{i=1}^{2^J}\int_0^1\int_{-\frac{R+1}{n}}^{\frac{R+1}{n}}\left[f(x_i)-f(y+x_i)\right]^2 dydx_i$$

$$\leq\sum_{i=1}^{2^J}\int_{-\frac{R+1}{n}}^{\frac{R+1}{n}}\int_0^1\left[f(x_i)-f(y+x_i)\right]^2 dx_idy$$

$$\leq\sum_{i=1}^{2^J}\int_{-\frac{R+1}{n}}^{\frac{R+1}{n}}\left(\int_0^1\left|f(x_i)-f(y+x_i)\right|^p dx_i\right)^{\frac{2}{p}}dy \tag{8}$$

$$\leq M^2n\int_{-\frac{R+1}{n}}^{\frac{R+1}{n}}y^{2(\alpha\wedge 1)}dy$$

$$\leq Cn^{-2(\alpha\wedge 1)}.$$

Since $f\in\Lambda^{\frac{2\alpha}{2\alpha+1}}(M)$, the second term on the right side of (7) is easily bounded by lemma

1 and Jensen's inequality:

$$E_{(\tilde{X})} \sum_{i=1}^{2^J} \int_{-\frac{R}{n}+\frac{i}{n(n+1)}}^{\frac{R}{n}+\frac{i}{n(n+1)}} \left[f(x_{(i)}+y) - f\left(y+\frac{i}{n+1}\right)\right]^2 dy$$

$$\leq E_{(\tilde{X})} \sum_{i=1}^{2^J} \int_{-\frac{R+1}{n}}^{\frac{R+1}{n}} \left[\left(x_{(i)} - \frac{i}{n+1}\right)^{\frac{2\alpha}{2\alpha+1}}\right]^2 dy$$

$$\leq C\left(\frac{1}{n}\right) \sum_{i=1}^{2^J} E_{(\tilde{X})} \left[\left(x_{(i)} - \frac{i}{n+1}\right)^2\right]^{\frac{2\alpha}{2\alpha+1}} \tag{9}$$

$$= C\left(\frac{1}{n}\right) \sum_{i=1}^{n} \left(\frac{(n+1)i - i^2}{(n+1)^2(n+2)}\right)^{\frac{2\alpha}{2\alpha+1}}$$

$$\leq Cn^{-\frac{2\alpha}{2\alpha+1}}$$

Since $\alpha \geq 1/2$ implies $2(\alpha \wedge 1) \geq 2\alpha/2\alpha+1$, lemma 4 is proved for $q = \infty$. For $1 \leq q < \infty$, the same results hold because of the inclusion $B^{\alpha}_{p,q}(M) \subseteq B^{\alpha}_{p,\infty}(M)$.

The proof of theorem 2 now follows from (6) and lemma 4.

Proof of theorem 1. Let N be the number of sample points in the interval $[0,1]$. Then N is distributed as a Poisson(μ) random variable. Recall that if N is smaller than 5 or larger than 2μ, then $\hat{f}$ is set to zero. If N is not a power of 2, the signal will be reduced rather than extended to a dyadic integer to simplify the proof. In applying the estimator to data, however, the signal will be extended as described in section 2.3. Given N, the arrival times $x_{(i)}$ have the distribution of N ordered uniform statistics. Randomly permuting these times yields a set of N i.i.d. uniform[0,1] random variables. Let $N^* = 2^J \leq N < 2^{J+1}$, the smallest dyadic integer less than or equal to N. Randomly choose N^* of the i.i.d. uniforms, and order them. The estimator will be applied to this new, smaller signal vector of N^* ordered uniform random variables. By theorem 2 and the fact that $f \in L^p(M)$,

$$E\left(\|\hat{f} - f\|_2^2 \mid N\right) \leq \begin{cases} 2M^2 & N < 5, N > 2\mu, \\ CN^{*-\frac{2\alpha}{2\alpha+1}}, & N = 5, 6, \ldots, 2\mu. \end{cases}$$

Since $N \leq 2N^*$,

$$E\left(\|\hat{f} - f\|_2^2 \mid N\right) \leq \begin{cases} 2M^2 & N < 5, N > 2\mu, \\ CN^{-\frac{2\alpha}{2\alpha+1}}, & N = 5, 6, \ldots, 2\mu. \end{cases}$$

This implies that

$$E\left(E\left(\|\hat{f}-f\|_2^2 \mid N\right)\right) \leq 2M^2P(N<5, N>2\mu) + \sum_{n=5}^{2\mu} Cn^{-\frac{2\alpha}{2\alpha+1}} P(N=n)$$

For a Poisson(μ) random variable, $P(N<5, N>2\mu) \leq C\mu^{-1}$. Therefore,

$$\begin{aligned}
E\left(E\left(\|\hat{f}-f\|_2^2 \mid N\right)\right) &\leq C\mu^{-1} + \sum_{n=5}^{\infty} Cn^{-\frac{2\alpha}{2\alpha+1}} P(N=n) \\
&\leq C\mu^{-1} + (1-e^{-\mu}) \sum_{n=1}^{\infty} Cn^{-\frac{2\alpha}{2\alpha+1}} \frac{P(N=n)}{1-e^{-\mu}} \\
&\leq C\mu^{-1} + C \cdot E(N^{-\frac{2\alpha}{2\alpha+1}} \mid N \geq 1) \\
&\leq C\mu^{-1} + C\left(E(N^{-1} \mid N \geq 1)\right)^{\frac{2\alpha}{2\alpha+1}} \\
&\leq C\mu^{-\frac{2\alpha}{2\alpha+1}}.
\end{aligned}$$

Appendix

The eight test functions are displayed in figure 2. The formulas for "doppler", "heavisine", "bumps", and "blocks" can be found in Donoho and Johnstone (1994). The rest are given below. In the simulations, these formula were modified by a constant to give a standard deviation of 10.

Spikes:

$$\begin{aligned}
f(x) = 15.6676[&e^{-500(x-0.23)^2} + 2e^{-2000(x-0.33)^2} \\
&+ 4e^{-8000(x-0.47)^2} + 3e^{-16000(x-0.69)^2} + e^{-32000(x-0.83)^2}]
\end{aligned}$$

Blip:

$$f(x) = (0.32 + 0.6x + 0.3e^{-100(x-0.3)^2})I_{(0,.8]}(x) + (-0.28 + 0.6x + 0.3e^{-100(x-1.3)^2})I_{(.8,1]}(x)$$

Corner:

$$f(x) = 10x^3(1-4x^2)I_{(0,.5]} + 3(0.125 - x^3)x^4 I_{(.5,.8]}(x) + 59.4432(x-1)^3 I_{(.8,1]}(x)$$

Wave:

$$f(x) = 0.5 + 0.2\cos(4\pi x) + 0.1\cos(24\pi x)$$

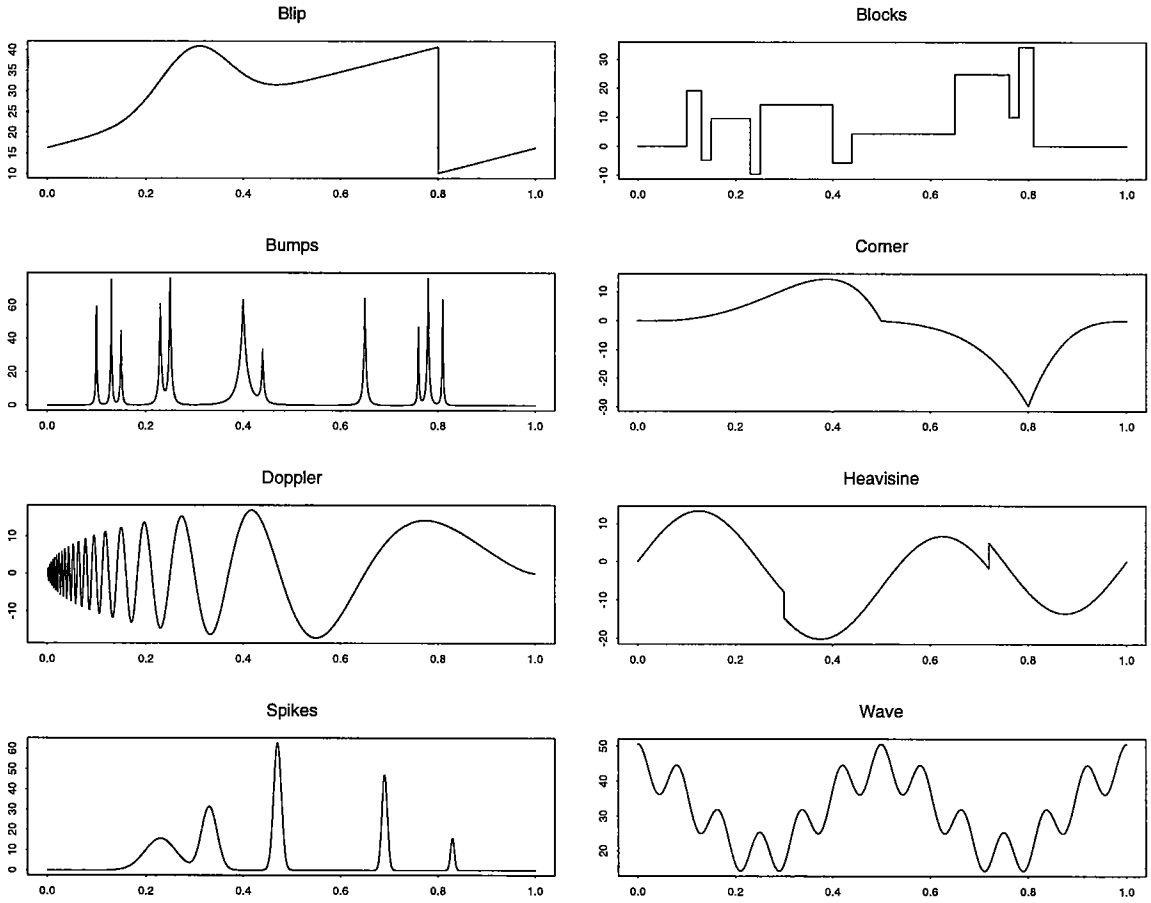

Figure 2: Test Functions

References

CAI, T. (1998). Adaptive wavelet estimation: A block thresholding and oracle inequality approach. *Ann. Statist.* **27** 898–924.

CAI, T. and BROWN, L. (1998). Wavelet shrinkage for noneqispaced samples. *Ann. Statist.* **26** 1783–1799.

CAI, T. and BROWN, L. (1999). Wavelet estimation for samples with random uniform design. *Statist. Probab Lett.* **42** 313–321.

DONOHO, D. and JOHNSTONE, I. (1994). Ideal spatial adaptation via wavelet shrinkage. *Biometrika* **81** 425–455.

HALL, P., KERKYACHARIAN, G. and PICARD, D. (1998). Block threshold rules for curve estimation using kernel and wavelet methods. *Ann. Statist.* **26** 922–942.

HALL, P., KERKYACHARIAN, G. and PICARD, D. (1999). On the minimax optimality of block thresholded wavelet estimators. *Statist. Sinica* **9** 33–50.

HALL, P. and TURLACH, B. (1997). Interpolation methods for nonlinear wavelet regression with irregularly sapced design. *Ann. Statist.* **25** 1912–1925.

MEYER, Y. (1990). *Ondelettes et Opérateurs: I. Ondelettes.* Hermann et Cies, Paris.