Nonparametric Regression on Random Processes and Designs

Eric Chicken
Florida State University

Abstract

Conditions are given on random processes and distributions for sample point placement that result in optimal rates of convergence for nonparametric regression when using a block thresholded wavelet estimator. The estimator is adaptive over a large range of Hölder function spaces and the convergence rate exhibited is an improvement over term-by-term wavelet estimators. Threshold selection is implemented via minimzing Stein's unbiased risk estimate, rather than using a fixed threshold as is usually done in block thresholding. This estimator with data dependent threshold ("BlockSure") is compared against fixed threshold block estimators and the more traditional term-by-term threshold wavelet estimators on fixed, equally spaced sample points and on several random design schemes. Simulations show BlockSure is superior to the other wavelet estimators in terms of mean squared error.

1 Introduction

Suppose the following nonparametric problem is being considered:

$$y_i = f(x_i) + \sigma\varepsilon_i, \quad i = 1, 2, \dots, n, \tag{1}$$

where the y_i are observed data, x_i are fixed, equally spaced sample points, ε_i are iid standard normal random variables, and σ is known. When wavelets are being used, this problem has been examined in such papers as Donoho and Johnstone (1994), Cai (1998), and Hall et al. (1998), to name a few. Donoho and Johnstone solve (1) through the use of term-by-term thresholding of wavelet coefficients, and show that such estimators as SureShrink or VisuShrink come to within $\log n$ of the minimax rate of convergence, i.e.,

$$\sup_{f\in\Lambda^\alpha(M)} \mathbb{E}\|f-\hat{f}\|_2^2 \le C(\log n)n^{-2\alpha/(2\alpha+1)}.$$

Here, $\hat{f}$ is the estimate of f derived by term-by-term thresholding of the coefficients of the discrete wavelet transform (DWT) and $\Lambda^\alpha(M)$ is a Hölder space of functions (to be defined in Section 2). These estimators are adaptive in the sense that they work for α unknown, but in a specified interval which can be made arbitrarily large. Additionally, their estimators possess fast computation speed (on the order of $O(n)$ for a fixed threshold level).

MSC2000: Primary 62G07; secondary 62G20.
Key words and phrases. Wavelets, thresholding, nonparametric regression, adaptive, Hölder class

Hall et al. (1998) and Cai (1998) have devised wavelet estimators based on thresholding several wavelet coefficients simultaneously rather than individually as in term-by-term estimators. For the regression problem as given in (1), these block thresholded estimators maintain the speed and adaptivity of the term-by-term methods, but remove the logarithm penalty associated with the convergence rate:

$$\sup_{f \in \Lambda^\alpha(M)} \mathbb{E}\|f - \hat{f}\|_2^2 \leq Cn^{-2\alpha/(2\alpha+1)}. \tag{2}$$

When non-standard assumptions on the x_i are introduced, less work has been done. For example, when the sample points x_i are not fixed or equispaced, but uniformly placed over an interval, Cai and Brown (1999) have shown that term-by-term methods can still be applied without loss of computation speed, convergence rates, or adaptivity. Chicken (2003b) extended their results by using block thresholding. There it was shown that when the sample points are distributed as uniform random variables or placed on an interval in accordance with a Poisson process, the convergence rate is improved over the term-by-term method by removing the $\log n$ term, and adaptivity and computation speed are preserved.

These results on the random placement of sample points can be broadened. In this paper, theorems are stated that give general conditions on the distributions and random processes for the placement of sample points x_i in model (1) for which block thresholded wavelet regression maintains adaptivity and the minimax rates of convergence at (2). The uniform placement of sample points and the placement of sample points per a Poisson process are seen to be special instances of these theorems.

Block thresholding in this paper will not implemented as it has been in the past. The block thresholded estimators cited above all use a constant threshold parameter indicated by theoretical results. The threshold is used for all wavelet coefficients across all resolution levels. This is similar to Donoho and Johnstone's VisuShrink estimator, where a "universal" threshold is used. In Chicken (2003a), a different block threshold estimator was proposed, "BlockSure". In that article, which dealt with fixed rather than random sample points, a threshold was determined by choosing a value that minimizes Stein's unbiased risk estimate. This data-dependent threshold was used for all coefficients in all resolution levels of the DWT. Here, this method of threshold selection is extended to finding a threshold for each resolution level of wavelet coefficients. The theorems hold for each type of threshold: constant or data-dependent.

Simulation results are reported that compare how well the block and term-by-term thresholded estimators perform on some random sample point placement schemes. Additionally, BlockSure is compared to fixed, block threshold estimators on equally spaced, fixed sample points, something lacking in Chicken (2003a). BlockSure is shown to be superior to all the other estimators, whether term-by-term or block thresholded, in all simulated examples.

Section 2 of this paper gives some background on wavelets and the function spaces under consideration. Section 3 details the estimator and gives the theoretical results of this paper. Simulation results are laid out and discussed in Section 4. The theorems are proved in section 5.

2 Wavelets and function spaces

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

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

The wavelets used in this paper will be assumed to have compact support, say, $[0, s]$ for s a positive integer. The functions f to be estimated will be assumed to have domain of $[0, 1]$. Minimal changes are necessary if the support of the functions is changed to something other than $[0, 1]$.

The wavelet coefficients for the wavelet transform of a function $f \in L_2$ are the usual inner product:

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

Hence, f can be expressed as an infinite series

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

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. These coefficients are not known since f is unknown. Estimates of $\hat{\theta}$ and $\hat{\xi}$ will be found via the DWT. The range of the subscripts k depends on the level j. For a given j there are exactly $2^j + s - 1$ indices k where the wavelet coefficients are not 0.

The function space of interest here is a Hölder space with smoothness parameter α. A function f is in the space $\Lambda^\alpha(M)$ $(0 < M < \infty)$ if, for $0 < \alpha \leq 1$,

$$|f(x) - f(y)| \leq M\,|x - y|^\alpha\,,$$

and, for $\alpha > 1$, $\alpha = \lfloor \alpha \rfloor + s$, where $\lfloor \alpha \rfloor$ is the largest integer less than α, $0 < s \leq 1$, and $|f'(x)| \leq M$,

$$\left| f^{(\lfloor \alpha \rfloor)}(x) - f^{(\lfloor \alpha \rfloor)}(y) \right| \leq M\,|x - y|^s\,.$$

In particular, for all $\alpha > 0$,

$$|f(x) - f(y)| \leq M\,|x - y|^{\alpha \wedge 1}\,.$$

Wavelet coefficients for functions in Hölder spaces have the following properties (see Meyer (1990)). If $f \in \Lambda^\alpha(M)$, $\Theta = \left\{\xi_{j_0,1}, \ldots, \xi_{j_0,2^{j_0}}, \theta_{j_0,1}, \ldots\right\}'$ is the vector of wavelet coefficents of f, and ψ has r vanishing moments $(r \geq \alpha)$, then for all integers k

$$|\xi_{j_0,k}| \leq C,$$

and

$$|\theta_{j,k}| \leq C 2^{-j(\alpha+1/2)}$$

where the constant C depends only on M and ψ.

For the results on random processes, it will also be necessary for the function f to have bounded L^2 norm. f is in $L^2(M)$ if

$$\|f\|_2 = \sqrt{\int f^2} \leq M.$$

3 The estimator

Suppose a signal is received on an interval in accordance with one of the following:

$$y_i = f(X_i) + \sigma\varepsilon_i, \quad i = 1, 2, \ldots, n, \tag{3}$$

where the X_i are iid random variables on the interval $[0, 1]$, independent of the error terms ε_i, and σ is known, or,

$$y_i = f(S_i) + \sigma\varepsilon_i, \quad i = 1, 2, \ldots, N, \tag{4}$$

where the S_i are the arrival times with respect to a random process on $[0, 1]$ independent of the ε_i, N is a non-negative integer-valued random variable (independent of ε_i), and σ is again known. In (3), the random variables X_i will be ordered and the model is rewritten as

$$y_i = f(X_{(i)}) + \sigma\varepsilon_i, \quad i = 1, 2, \ldots, n, \tag{5}$$

where the ε_i and y_i have been appropriately reordered.

The estimator will treat the data points in (4) or (5) as though they were equispaced. However, while the number of points n in (5) can be taken to be dyadic (only the placement of points is random, the experimenter chooses the number), in (4) this cannot be done. Since the DWT requires a dyadic number of points, some modifications must be done. Reflecting the data by an equal amount about each endpoint to the next highest dyadic integer is one possibility, but it induces some dependency to the sample points. Another possibility would be to use the points $x_1, x_2, \ldots, x_{2^k}$, where 2^k is the largest dyadic integer less than or equal to N, to estimate f on the appropriate subinterval. Then, use the last 2^k points $x_{N-2^k+1}, \ldots, x_N$ to estimate f on the remaining portion of the interval. Such a method would maintain the independence of the points on each of the two partial estimates of f. The estimates are then combined to give the full estimate over the entire interval. In either case, n or N may now be assumed to be dyadic.

Given dyadic number of points n (or N) and treating the data as equispaced, the estimate of f will be determined via block thresholding the DWT coefficients. Let

$$\Theta = \begin{pmatrix} \hat{\xi} \\ \hat{\theta} \end{pmatrix} = n^{-1/2} W y$$

be the DWT of the data (4) or (5). Since W is an orthogonal transform when n is dyadic, Θ is normal with covariance $\sigma^2/n \cdot I_{n\times n}$. This $n \times 1$ vector of estimated coefficients will be thresholded. The thresholding rule considered in this paper is

$$\tilde{\theta}_{jk} = \eta(\hat{\theta}_{jk}, \lambda) = \hat{\theta}_{jk} \cdot \left(1 - \frac{\lambda L \sigma^2/n}{\sum_{i\in B} \hat{\theta}_{ji}^2}\right)_+, \tag{6}$$

where B is the block of coefficients $\hat{\theta}_{ji}$ containing the index jk. L is the number of coefficients in the block, λ is a threshold parameter (to be determined), and σ^2/n is the variance of a wavelet coefficient. This rule keeps, but shrinks, all coefficients in a block when the estimated information in the block exceeds a multiple of the noise in a block. When this does not occur, all coefficients in a block are set to zero. Setting

$$\tilde{\Theta} = \begin{pmatrix} \hat{\xi} \\ \tilde{\theta} \end{pmatrix}$$

to be the thresholded DWT coefficients (note that the coarse coefficients $\hat{\xi}$ are not thresholded), the estimate of f is then

$$\hat{f} = n^{1/2} W' \tilde{\Theta},$$

where $W' = W^{-1}$ is the inverse DWT. The noise level σ is not known, but easily estimated. For example, the median absolute deviation of the finest level of detail coefficients is a consistent estimate. The parameter λ also needs to be determined.

One choice for λ is the value found by Cai (1998), $\lambda = 4.50524$. This number is found to be asymptotically optimal for block length $L = \log n$ in terms of convergence rates and adaptivity. Such a threshold is similar to $\lambda = \sqrt{2 \log n}$ found in term-by-term universal thresholding rules. In Chicken (2003a), another choice for λ has been proposed. It is a data-adaptive value determined by minimizing Stein's unbiased risk estimate (Stein (1981)). The block estimator using this threshold will be referred to as BlockSure.

Suppose X is multivariate normal with mean μ and and covariance matrix V. If $\hat{\mu} = \hat{\mu}(X)$ is an estimate of μ and $g(X) = \hat{\mu}(X) - X$, then the mean squared error of $\hat{\mu}$ is

$$E\|\hat{\mu} - \mu\|_2^2 = E\|g(X) + X - \mu\|_2^2 = E\left(tr(V) + \sum_i g_i^2(X) + 2tr(V \cdot dg(X)\right),$$

where $dg(X)$ is the matrix whose ijth entry is determined by $\partial g_i(X)/\partial x_j$. For the rule η, letting $X = \Theta$, $V = \sigma^2/n \cdot I_{n\times n}$, μ be the true wavelet coefficient vector for f, and $\hat{\mu}(X) = \tilde{\Theta}$, minimizing Stein's unbiased risk estimate is equivalent to choosing λ which minimizes

$$\sum_b (S_b^2 - 2L\sigma^2/n) I(S_b^2 \le \lambda L\sigma^2/n) + \textstyle\sum_b \left[(\lambda L\sigma^2/n)^2 + 2\lambda L(\sigma^2/n)^2(2-L)\right]/S_b^2 I(S_b^2 > \lambda L\sigma^2/n), \tag{7}$$

where the summation over b is the summation over all blocks of thresholded coefficients. Alternatively, a variation of BlockSure can be found be minimizing

$$(S_b^2 - 2L\sigma^2/n) I(S_b^2 \le \lambda_b L\sigma^2/n) + \left[(\lambda_b L\sigma^2/n)^2 + 2\lambda_b L(\sigma^2/n)^2(2-L)\right]/S_b^2 I(S_b^2 > \lambda_b L\sigma^2/n) \tag{8}$$

for each block b of thresholded coefficients. In this case, the threshold is a vector of values, one λ_b for each block.

The theorems are now ready to be stated.

Theorem 1 *Let X_i be iid positive random variables with*

$$\mathbb{E}X_i = 1/n, \quad varX_i \le c/n^2,$$

and

$$S_i = \sum_{k=1}^{i} X_k.$$

Suppose a sample $\{(S_1, y_1), (S_2, y_2), \ldots, (S_N, y_N)\}$ is collected on an interval of length 1 with

$$y_i = f(S_i) + \sigma\varepsilon_i,$$

where ε_i are independent standard normal random variables, independent of the S_i. If ψ has r vanishing moments, then for $\lambda \ge 1$, $\alpha \in [1/2, r]$, $0 < M < \infty$, and $\hat{f}$ as given above,

$$\sup_{f \in \Lambda^\alpha(M) \cap L^2(M)} \mathbb{E}\|f - \hat{f}\|_2^2 \le C n^{-2\alpha/(s\alpha+1)}$$

on each interval.

The block thresholded wavelet estimator therefore achieves the minimax rate at (2), an improvement over the rate term-by-term estimates can attain, over a wide range of Hölder spaces when the sample points arrive according to a random process subject to the restrictions stated in the theorem. This is the same rate as for equispaced sample point placement. This rate is achieved for any method of choosing the threshold λ, though it will be seen from the simulations that some λ selection methods seem superior to others. In particular, choosing the threshold as done in BlockSure generally gives the best results in terms of low mean squared error. BlockSure thus gives the optimal rate in terms of theoretical results and performs the best in simulations.

If the X_i are exponential $(1/n)$ random variables, then the S_i are a Poisson process with rate n. The results in Chicken (2003b) are therefore a special case of this theorem.

Theorem 2 *Let X_i be iid random variables on $[0, 1]$ and $X_{(i)}$ their ordered counterparts with*

$$\left|\mathbb{E}X_{(i)} - i/n\right| \le C n^{-1/2},$$

and

$$varX_{(i)} \le C/n.$$

Suppose a sample $\left\{(X_{(1)}, y_1), (X_{(2)}, y_2), \ldots, (X_{(n)}, y_n)\right\}$ is collected with

$$y_i = f(X_{(i)} + \sigma\varepsilon_i,$$

where ε_i are independent standard normal random variables, independent of the S_i. If ψ has r vanishing moments, then for $\lambda \ge 1$, $\alpha \in [1/2, r]$, $0 < M < \infty$, and $\hat{f}$ as given above,

$$\sup_{f \in \Lambda^\alpha(M)} \mathbb{E}\|f - \hat{f}\|_2^2 \le C n^{-2\alpha/(s\alpha+1)}$$

on each interval.

The comments following Theorem 1 all apply to Theorem 2 as well. Note that this theorem does not require f to be in $L^2(M)$ as is the case for the previous theorem. If the bound condition on the variance of the order statistics is lowered to C/n^2, then the range of α can be broadened to $(0, r]$ vice $[1/2, r]$.

Since the expected values of the order statistics are approaching the sample quantiles of a uniform $(0, 1)$ random variable, this theorem in effect holds for "quasi-uniform" random variables. The case where the X_i are uniform $(0, 1)$ is certainly such a random variable, and so the results of Chicken (2003b) are a special case of this theorem. Additionally, this theorem is an improvement over that found in Cai and Brown (1999). There, only uniform random variable were considered, and the rate achieved was not within a constant of minimax, but only within $\log n$ of minimax.

The conditions in each of these theorems are sufficient, but have not been shown to be necessary.

A note on computation speed. The block estimators are $O(n)$ given the threshold λ. However, just as is the case with the term-by-term SureShrink estimator, computational complexity is increased when λ must be determined for BlockSure.

4 Simulation results

Simulations were run to examine the effectiveness of the varying estimators on different types of random designs. Test functions from Marron et al. (1998) and Donoho and Johnstone (1994) were used, see Appendix. The test functions were scaled so each has a standard deviation of 10. Noise was added to each function with signal-to-noise ratios of 5 and 7. The MSE is estimated using 40 replications for each combination of simulation parameters.

For the random process case, rates of $n = 256, 512, \ldots, 4096$ were implemented. Since the actual number of points in $[0, 1]$ will not necessarily be n, the signal was extended by reflection about each endpoint to the next largest dyadic. The function is then estimated on this extended interval, but error is measured only on the original interval. Three processes were simulated. The interarrival times were uniform $(0, 2/n)$, exponential $(1/n)$, or beta $(1, n-1)$. Each of these random variables meets the requirements of the theorem on random processes.

Five estimators were compared for each simulation case. Two are term-by-term estimators devised by Donoho and Johnstone, and three are block estimators. The term-by-term estimators are RiskShrink (R) and RiskShrink with threshold determined separately for each level of thresholded coefficients (R_l). The VisuShrink estimator was not considered since its MSE is known to be larger than that of RiskShrink. WaveThresh3 software (Nason (1998)) was used to implement these estimators (as well as to genereate the DWT for all estimators).

The block estimators are Cai's "universal" thresholded estimator (B_c) where the threshold is fixed at 4.50524, an optimal theoretical value for blocks of length $\log n$. BlockSure (B), as given in Chicken (2003a) and Section 3, chooses a threshold by minimizing the risk for all thresholded coefficients. A similar estimator minimizes the risk within each level of coefficients and will be denoted as (B_l). The noise level σ for the block estimators was determined by finding the median absolute deviation of the highest level of detail coefficients. The results of these simulations are reported in Tables 1 through 6. Only Table 1 will be discussed in detail.

In that table, BlockSure (B) has the lowest mean-squared error in all but 5 cases, where it is the second lowest. Twice it is outperformed by BlockSure with level-dependent thresholding (B_l) and three times by RiskShrink (R_l). The rate n in each of these cases is at the low end, $n = 256$ four times and $n = 512$ once. At the higher rates, (B) always has the lowest MSE. This is not surprising given the theoretical results on convergence rates: the term-by-term estimators have an additional $\log n$ term in their convergence rate. Of the three block estimators, (B) seems to be the most efficient, while (R_l) is the apparent best of the term-by-term estimators. The other combinations of simulation parameters (results in Tables 2 through 6) follow this same general pattern.

Some examples of the reconstructions are given in Figures 1 and 2. Figure 1 shows reconstructions of "Blip". The signal-to-noise ratio is 5 and the rate of the process is 500. The interarrival times for the process follow a uniform $(0, 2/n)$ distribution. The solid line is the wavelet estimate and the dashed line is the true function. The block estimators are preferable in terms of lower MSE, and this may be seen visually in each of the block reconstructions of this test function. In fact, the MSE for BlockSure for this particular example is 0.480, for Cai's block estimator B_c it is 0.844, for the RiskShrink estimator R_l it is 1.531, and for the VisuShrink estimator it is 1.961.

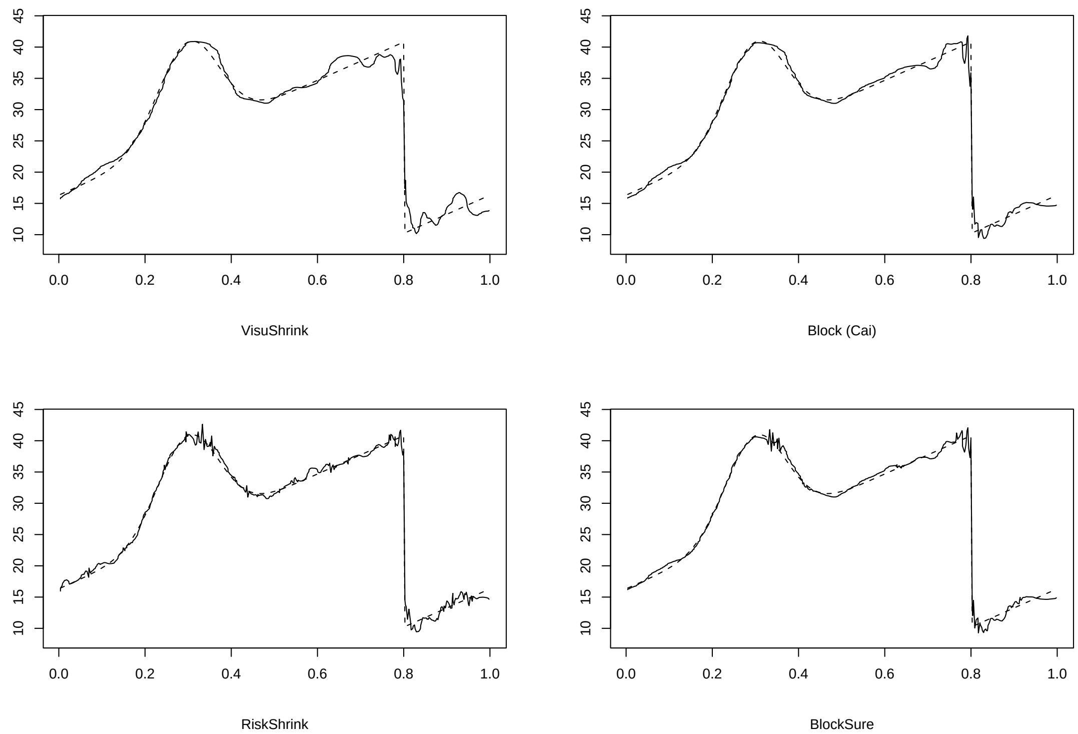

Figure 1: Reconstruction of "Blip" test function.

From the figure (and the tables, as well), some parallels may be drawn between term-by-term and block estimators. For each type of estimator (term-by-term or block), the data-dependent threshold (R_l or BlockSure) has lower MSE than its non data-dependent counterpart (VisuShrink or B_c). However, the non data-dependent estimators are visually smoother. BlockSure and RiskShrink each have "spiky" artifacts throughout their reconstructions which detract from their overall appearance, though these spikes are fewer and less pronounced with BlockSure (just as B_c is smoother than VisuShrink). But, regardless of whether the analyst is interested in a smooth reconstruction or a low MSE, the block estimators still surpass the term-by-term estimators.

Figure 2 shows more reconstructions. Here, the test function is "doppler", the signal-to-noise ratio is 7, and the interarrival distribution is exponential $(1/n)$, where $n = 1000$. The comments made about Figure 1 apply here, also.

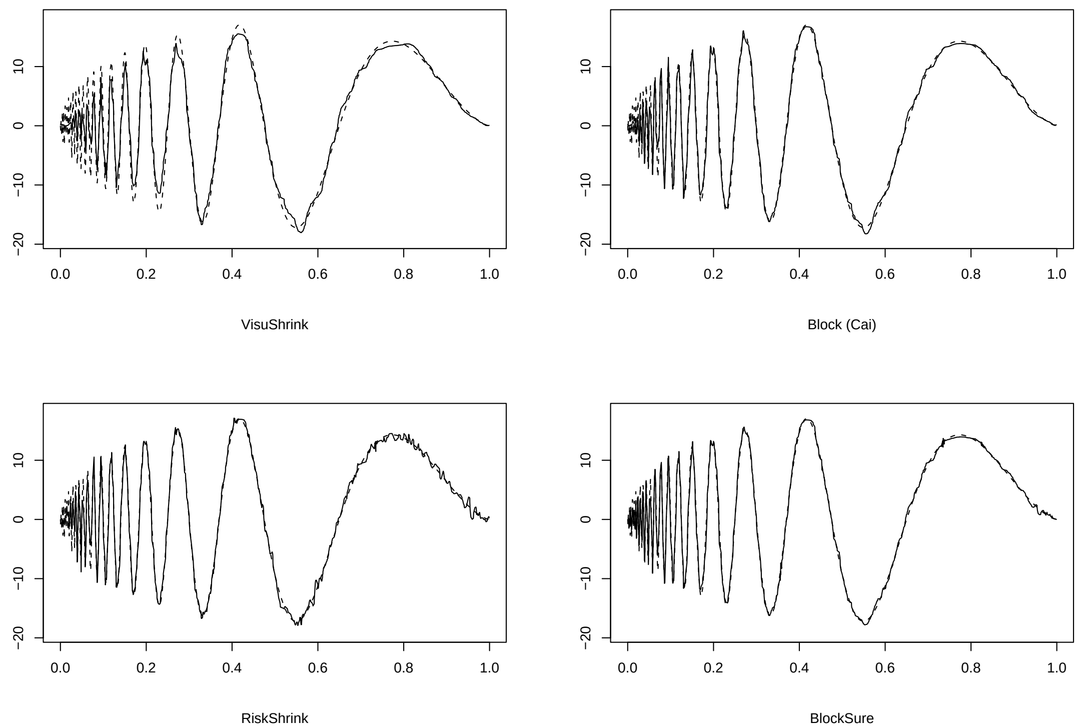

Figure 2: Reconstruction of "Doppler" test function.

For the case where the points are randomly distributed on an interval, the same five estimators were compared. Additionally, the performance of BlockSure on equispaced data (B_E) is given to show how close the the nonequispaced case comes to the equispaced one. As before, the signal-to-noise ratio is either 5 or 7, and sample sizes range from $n = 256$ to $n = 4096$. The ditribution used was uniform $(0, 1)$. The results are given in Tables 7 and 8.

In Table 7, the BlockSure estimator is best (or tied for best) in all but four cases. The term-by-term estimators are never better. BlockSure is outperformed in these simulations by only by (B_l), and only at low sample sizes as was the case with the random processes. BlockSure was tied with Cai's block estimator (B_c) in three cases. Similar conclusions can be drawn from Table 8.

Function n	R	R_l	B_c	B	B_l
Blip					
256	1.345	1.125	1.156	0.980	1.262
512	0.868	0.722	0.655	0.533	0.737
1024	0.586	0.551	0.391	0.365	0.522
2048	0.395	0.462	0.275	0.222	0.270
4096	0.250	0.285	0.177	0.129	0.158
Blocks					
256	3.893	2.861	6.071	2.442	2.301
512	2.636	2.144	3.420	1.759	1.775
1024	1.799	1.593	2.055	1.231	1.297
2048	1.246	1.270	1.512	0.874	0.903
4096	0.855	0.844	1.049	0.590	0.616
Bumps					
256	4.567	2.895	6.447	2.662	2.485
512	3.089	2.392	3.550	1.892	1.889
1024	2.172	1.787	2.065	1.321	1.344
2048	1.415	1.275	1.369	0.900	0.942
4096	0.949	0.904	0.863	0.591	0.622
Corner					
256	1.266	0.758	0.906	0.730	0.959
512	0.781	0.463	0.499	0.457	0.664
1024	0.521	0.302	0.272	0.258	0.389
2048	0.272	0.206	0.163	0.150	0.227
4096	0.173	0.139	0.116	0.097	0.137
Doppler					
256	2.808	2.177	3.043	1.737	1.753
512	1.880	1.543	1.777	1.202	1.229
1024	1.193	1.030	0.954	0.746	0.858
2048	0.776	0.636	0.496	0.419	0.499
4096	0.485	0.421	0.258	0.212	0.258
Heavisine					
256	1.741	0.829	1.131	0.987	1.167
512	1.275	0.596	0.815	0.649	0.773
1024	0.799	0.426	0.496	0.412	0.521
2048	0.564	0.290	0.318	0.254	0.308
4096	0.341	0.200	0.221	0.160	0.187
Spikes					
256	2.762	2.002	2.950	1.822	1.902
512	1.750	1.259	1.523	1.130	1.279
1024	1.091	0.739	0.788	0.650	0.827
2048	0.668	0.447	0.436	0.376	0.453
4096	0.416	0.279	0.256	0.201	0.236
Wave					
256	2.522	1.521	2.339	1.537	1.880
512	1.450	0.903	1.097	0.844	1.157
1024	0.927	0.735	0.499	0.449	0.614
2048	0.578	0.662	0.223	0.214	0.295
4096	0.347	0.480	0.119	0.115	0.158

Table 1: Mean-squared errors from 40 replications, signal to noise ratio of 5, sample points following a uniform $(0, 2/n)$ interarrival time.

Function n	R	R_l	B_c	B	B_l
Blip					
256	1.594	1.181	1.323	1.002	1.197
512	0.887	0.727	0.750	0.572	0.688
1024	0.623	0.544	0.448	0.372	0.494
2048	0.407	0.399	0.289	0.221	0.297
4096	0.257	0.339	0.197	0.143	0.181
Blocks					
256	4.019	2.863	5.862	2.519	2.349
512	2.626	2.212	3.438	1.799	1.794
1024	1.761	1.711	2.048	1.201	1.282
2048	1.261	1.345	1.486	0.857	0.918
4096	0.850	1.023	1.046	0.594	0.628
Bumps					
256	4.097	2.738	5.594	2.456	2.372
512	2.993	2.329	3.536	1.900	1.876
1024	2.220	1.877	2.265	1.376	1.392
2048	1.502	1.314	1.513	0.952	0.986
4096	1.008	0.933	0.958	0.632	0.666
Corner					
256	1.311	0.816	1.026	0.842	1.014
512	0.858	0.549	0.574	0.504	0.675
1024	0.500	0.333	0.307	0.280	0.404
2048	0.313	0.212	0.185	0.161	0.206
4096	0.185	0.147	0.115	0.097	0.127
Doppler					
256	3.022	2.395	3.748	2.021	1.957
512	2.123	1.727	2.192	1.424	1.450
1024	1.402	1.195	1.190	0.892	0.966
2048	0.904	0.738	0.662	0.526	0.598
4096	0.564	0.469	0.357	0.282	0.315
Heavisine					
256	1.862	1.017	1.318	1.115	1.259
512	1.201	0.698	0.810	0.682	0.795
1024	0.797	0.439	0.493	0.433	0.572
2048	0.556	0.304	0.356	0.276	0.323
4096	0.344	0.213	0.227	0.174	0.211
Spikes					
256	3.045	2.142	3.338	1.939	1.960
512	1.856	1.389	1.783	1.230	1.332
1024	1.223	0.906	1.000	0.783	0.908
2048	0.772	0.555	0.560	0.450	0.524
4096	0.474	0.330	0.320	0.256	0.303
Wave					
256	3.004	1.834	2.891	1.888	1.987
512	1.731	1.174	1.396	1.103	1.305
1024	1.023	0.752	0.643	0.595	0.731
2048	0.643	0.591	0.313	0.289	0.341
4096	0.391	0.472	0.159	0.153	0.192

Table 2: Mean-squared errors from 40 replications, signal to noise ratio of 5, sample points following an exponential $(1/n)$ interarrival time.

Function n	R	R_l	B_c	B	B_l
Blip					
256	1.520	1.150	1.252	0.969	1.143
512	0.902	0.816	0.733	0.609	0.758
1024	0.592	0.545	0.405	0.334	0.450
2048	0.407	0.433	0.283	0.224	0.294
4096	0.262	0.287	0.189	0.140	0.170
Blocks					
256	3.880	2.905	5.600	2.458	2.378
512	2.717	2.173	3.408	1.791	1.780
1024	1.853	1.704	2.139	1.238	1.288
2048	1.245	1.343	1.493	0.850	0.896
4096	0.864	0.999	1.055	0.599	0.637
Bumps					
256	4.419	2.918	5.721	2.496	2.410
512	3.144	2.387	3.555	1.948	1.920
1024	2.226	1.879	2.229	1.364	1.376
2048	1.467	1.332	1.452	0.937	0.971
4096	1.013	0.946	0.964	0.636	0.662
Corner					
256	1.326	0.816	0.939	0.877	1.129
512	0.800	0.522	0.552	0.528	0.685
1024	0.489	0.310	0.331	0.292	0.395
2048	0.306	0.223	0.196	0.177	0.236
4096	0.175	0.152	0.112	0.099	0.128
Doppler					
256	3.346	2.458	4.004	2.197	2.100
512	2.088	1.730	2.195	1.413	1.438
1024	1.406	1.133	1.237	0.912	0.965
2048	0.882	0.732	0.665	0.531	0.585
4096	0.559	0.473	0.347	0.277	0.310
Heavisine					
256	1.596	1.019	1.260	1.081	1.342
512	1.183	0.700	0.873	0.704	0.838
1024	0.902	0.446	0.568	0.449	0.531
2048	0.549	0.305	0.339	0.275	0.347
4096	0.324	0.211	0.236	0.170	0.199
Spikes					
256	2.890	2.067	3.053	1.895	1.904
512	1.935	1.416	1.874	1.293	1.361
1024	1.217	0.905	0.988	0.765	0.899
2048	0.755	0.532	0.545	0.442	0.509
4096	0.466	0.318	0.323	0.251	0.285
Wave					
256	2.930	1.868	2.915	1.981	2.061
512	1.735	1.161	1.392	1.102	1.313
1024	1.048	0.804	0.657	0.566	0.725
2048	0.630	0.625	0.311	0.299	0.397
4096	0.384	0.486	0.160	0.159	0.208

Table 3: Mean-squared errors from 40 replications, signal to noise ratio of 5, sample points following a beta $(1, n-1)$ interarrival time.

Function n	R	R_l	B_c	B	B_l
Blip					
256	0.772	0.757	0.619	0.491	0.595
512	0.502	0.525	0.344	0.289	0.391
1024	0.331	0.369	0.208	0.191	0.253
2048	0.212	0.372	0.140	0.114	0.137
4096	0.138	0.303	0.090	0.070	0.089
Blocks					
256	2.343	2.127	3.212	1.320	1.244
512	1.515	1.657	1.875	0.952	0.958
1024	1.006	1.485	1.103	0.648	0.666
2048	0.692	1.282	0.797	0.459	0.482
4096	0.474	0.989	0.552	0.319	0.336
Bumps					
256	2.607	1.581	3.191	1.404	1.320
512	1.828	1.444	1.828	1.043	1.025
1024	1.163	1.139	1.035	0.689	0.704
2048	0.799	0.894	0.737	0.487	0.508
4096	0.532	0.695	0.476	0.329	0.341
Corner					
256	0.642	0.408	0.519	0.464	0.584
512	0.386	0.239	0.304	0.260	0.347
1024	0.241	0.137	0.157	0.144	0.195
2048	0.157	0.100	0.103	0.092	0.124
4096	0.098	0.076	0.071	0.054	0.071
Doppler					
256	1.717	1.525	1.847	1.035	0.993
512	1.117	1.085	1.063	0.684	0.680
1024	0.712	0.670	0.578	0.444	0.470
2048	0.461	0.419	0.307	0.245	0.282
4096	0.278	0.251	0.158	0.128	0.143
Heavisine					
256	1.023	0.682	0.805	0.605	0.680
512	0.686	0.364	0.474	0.364	0.440
1024	0.492	0.254	0.294	0.233	0.284
2048	0.261	0.177	0.185	0.146	0.175
4096	0.143	0.124	0.125	0.092	0.114
Spikes					
256	1.508	1.219	1.652	0.977	1.000
512	1.005	0.824	0.897	0.644	0.712
1024	0.642	0.473	0.472	0.379	0.429
2048	0.379	0.278	0.255	0.212	0.254
4096	0.234	0.169	0.143	0.118	0.135
Wave					
256	1.590	1.052	1.350	0.931	1.062
512	0.858	0.871	0.599	0.498	0.642
1024	0.533	0.906	0.285	0.268	0.356
2048	0.320	0.713	0.132	0.135	0.185
4096	0.189	0.864	0.075	0.072	0.090

Table 4: Mean-squared errors from 40 replications, signal to noise ratio of 7, sample points following a uniform $(0, 2/n)$ interarrival time.

Function n	R	R_l	B_c	B	B_l
Blip					
256	0.788	0.734	0.679	0.523	0.603
512	0.494	0.490	0.361	0.307	0.399
1024	0.342	0.403	0.222	0.196	0.253
2048	0.223	0.341	0.153	0.129	0.157
4096	0.137	0.261	0.094	0.073	0.092
Blocks					
256	2.184	1.944	3.199	1.324	1.257
512	1.487	1.776	1.872	0.946	0.961
1024	1.013	1.445	1.089	0.643	0.670
2048	0.692	1.463	0.816	0.462	0.481
4096	0.468	1.056	0.545	0.312	0.332
Bumps					
256	2.523	1.614	3.120	1.362	1.283
512	1.702	1.410	1.845	1.018	0.993
1024	1.249	1.253	1.158	0.733	0.739
2048	0.849	0.997	0.819	0.529	0.540
4096	0.583	0.709	0.550	0.362	0.378
Corner					
256	0.730	0.508	0.609	0.510	0.622
512	0.425	0.296	0.328	0.306	0.412
1024	0.258	0.176	0.186	0.166	0.237
2048	0.160	0.117	0.109	0.093	0.123
4096	0.102	0.069	0.067	0.056	0.071
Doppler					
256	1.978	1.757	2.484	1.261	1.163
512	1.341	1.271	1.397	0.870	0.845
1024	0.876	0.834	0.814	0.574	0.599
2048	0.558	0.499	0.444	0.338	0.361
4096	0.337	0.335	0.223	0.184	0.198
Heavisine					
256	1.042	0.778	0.890	0.671	0.708
512	0.704	0.455	0.536	0.416	0.512
1024	0.487	0.335	0.334	0.267	0.306
2048	0.273	0.187	0.216	0.157	0.189
4096	0.159	0.122	0.137	0.098	0.115
Spikes					
256	1.673	1.342	1.767	1.061	1.065
512	1.121	0.926	1.081	0.709	0.737
1024	0.706	0.580	0.603	0.443	0.489
2048	0.441	0.364	0.345	0.273	0.303
4096	0.276	0.214	0.200	0.156	0.168
Wave					
256	2.012	1.445	1.914	1.242	1.205
512	1.065	1.051	0.894	0.693	0.793
1024	0.641	0.896	0.433	0.384	0.475
2048	0.380	0.835	0.214	0.199	0.237
4096	0.219	0.806	0.112	0.102	0.121

Table 5: Mean-squared errors from 40 replications, signal to noise ratio of 7, sample points following an exponential $(1/n)$ interarrival time.

Function n	R	R_l	B_c	B	B_l
Blip					
256	0.795	0.712	0.700	0.546	0.633
512	0.505	0.516	0.364	0.310	0.393
1024	0.329	0.324	0.205	0.193	0.253
2048	0.226	0.406	0.157	0.125	0.153
4096	0.144	0.248	0.095	0.071	0.086
Blocks					
256	2.250	1.961	3.259	1.296	1.225
512	1.483	1.619	1.894	0.966	0.968
1024	0.996	1.494	1.088	0.633	0.659
2048	0.681	1.306	0.792	0.464	0.480
4096	0.470	1.057	0.565	0.322	0.339
Bumps					
256	2.357	1.447	2.843	1.344	1.253
512	1.721	1.358	1.885	1.039	1.027
1024	1.241	1.249	1.174	0.734	0.731
2048	0.855	0.883	0.840	0.536	0.555
4096	0.577	0.809	0.548	0.351	0.364
Corner					
256	0.829	0.536	0.605	0.562	0.659
512	0.438	0.288	0.333	0.290	0.390
1024	0.261	0.158	0.180	0.169	0.237
2048	0.168	0.116	0.107	0.092	0.114
4096	0.100	0.079	0.067	0.054	0.070
Doppler					
256	1.878	1.721	2.466	1.221	1.156
512	1.326	1.246	1.406	0.860	0.854
1024	0.860	0.827	0.805	0.564	0.583
2048	0.553	0.556	0.451	0.340	0.363
4096	0.336	0.301	0.227	0.178	0.196
Heavisine					
256	1.159	0.759	0.916	0.704	0.750
512	0.768	0.521	0.556	0.451	0.538
1024	0.438	0.307	0.322	0.262	0.310
2048	0.260	0.199	0.205	0.156	0.187
4096	0.176	0.125	0.134	0.095	0.112
Spikes					
256	1.590	1.275	1.700	1.038	1.034
512	1.097	0.982	1.115	0.734	0.773
1024	0.711	0.618	0.623	0.456	0.493
2048	0.459	0.388	0.361	0.278	0.310
4096	0.267	0.215	0.193	0.152	0.173
Wave					
256	2.064	1.471	1.954	1.245	1.170
512	1.084	0.975	0.916	0.713	0.805
1024	0.638	0.795	0.419	0.378	0.441
2048	0.378	0.787	0.207	0.193	0.233
4096	0.225	0.913	0.113	0.102	0.123

Table 6: Mean-squared errors from 40 replications, signal to noise ratio of 7, sample points following a beta $(1, n-1)$ interarrival time.

Some reconstructions are shown in Figure 3. The same general pattern as discussed for Figures 1 and 2 is apparent here. The block estimators are smoother than their term-by-term counterparts, and the data-dependent estimators have lower MSE than the non data-dependent estimators.

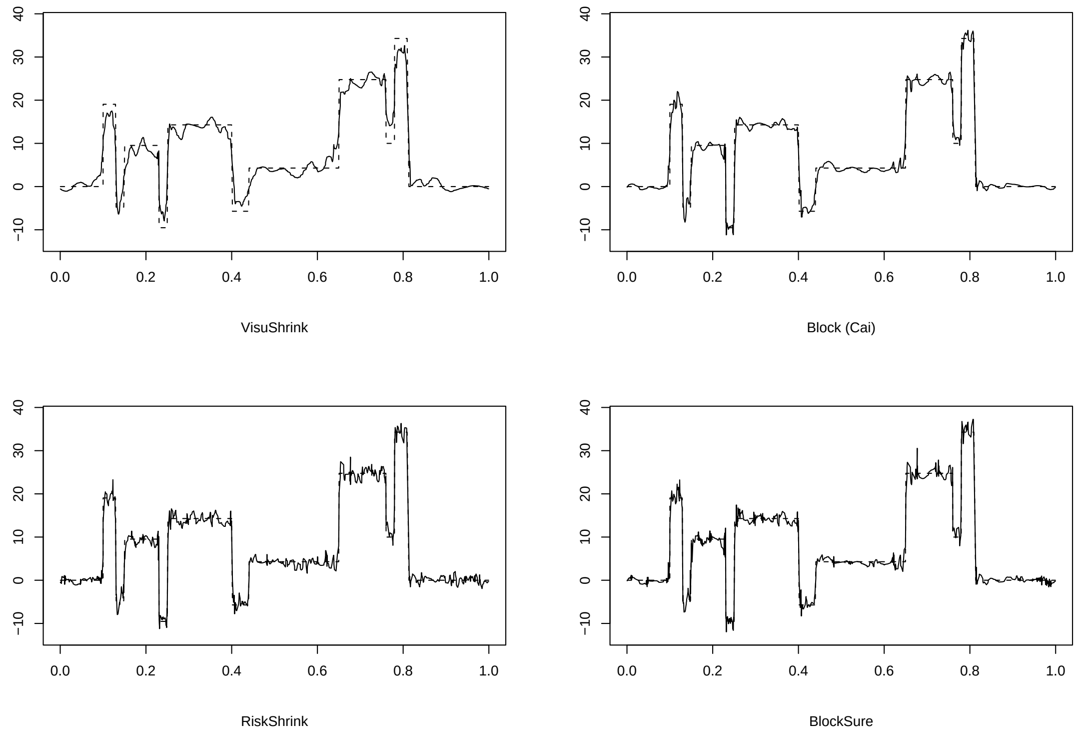

Figure 3: Reconstruction of "Block" test function.

Finally, a comparison of the estimators on equispaced samples is given. There is little to be said here other than that the block estimators outperform the term-by-term estimators, and that BlockSure is the best of the block estimators.

5 Proofs

5.1 Preliminaries

Before the theorems are proved, some preliminary results are necessary. First, for f in $\Lambda^{\alpha}(M)$, there is a useful approximation to ξ_{Jk}, where $2^J = n$ is a dyadic integer and $0 \leq k \leq n-1$. (In this lemma and others, C denotes a constant whose value depends only on the wavelets and M.)

For simplicity, it is assumed the support length of ϕ (and hence ψ) is 1. Therefore, for a fixed level j, the only indices k that have non-zero coarse coefficients are $k = 0, 1, \ldots, 2^j - 1$. Minimal changes are necessary if this changed to $[0, s]$. Then, there are $2^j + s - 1$ non-zero coefficients, $-s+1, -s+2, \ldots, 2^j - 1$. Additionally, renumber the observations $y_i, i = 1, 2, \ldots, n$ to $y_i, i = 0, 1, \ldots, n-1$ in order to simplify notation.

Function n	R	R_l	B_c	B	B_l	B_E
Blip						
256	1.361	1.162	1.166	0.916	1.226	0.954
512	0.857	0.761	0.621	0.549	0.730	0.667
1024	0.569	0.543	0.377	0.330	0.466	0.406
2048	0.383	0.381	0.207	0.186	0.267	0.193
4096	0.247	0.269	0.166	0.123	0.154	0.125
Blocks						
256	3.713	2.593	5.846	2.482	2.335	2.663
512	2.438	1.911	3.381	1.711	1.778	1.891
1024	1.729	1.361	2.017	1.204	1.259	1.254
2048	1.199	1.043	1.193	0.779	0.837	0.840
4096	0.845	0.682	1.017	0.586	0.614	0.612
Bumps						
256	4.203	2.831	5.750	2.533	2.384	2.789
512	3.203	2.324	3.628	1.906	1.872	2.126
1024	2.175	1.776	2.201	1.325	1.337	1.324
2048	1.507	1.257	1.247	0.894	0.949	0.861
4096	0.998	0.865	0.930	0.621	0.652	0.551
Corner						
256	1.551	0.801	0.922	0.789	1.061	0.621
512	0.886	0.506	0.487	0.450	0.657	0.381
1024	0.405	0.331	0.250	0.250	0.381	0.211
2048	0.253	0.237	0.131	0.131	0.211	0.138
4096	0.158	0.187	0.089	0.081	0.109	0.082
Doppler						
256	2.908	2.309	3.651	2.043	2.036	1.350
512	1.944	1.562	2.050	1.370	1.383	0.908
1024	1.267	1.035	1.136	0.858	0.933	0.530
2048	0.802	0.648	0.619	0.516	0.601	0.297
4096	0.490	0.407	0.313	0.263	0.309	0.138
Heavisine						
256	1.765	0.894	1.255	1.010	1.237	0.963
512	1.187	0.579	0.779	0.633	0.737	0.582
1024	0.778	0.382	0.483	0.403	0.472	0.364
2048	0.548	0.258	0.286	0.236	0.307	0.232
4096	0.374	0.176	0.203	0.145	0.188	0.152
Spikes						
256	2.975	2.052	3.188	1.874	1.937	1.414
512	1.823	1.352	1.755	1.165	1.260	0.766
1024	1.199	0.865	0.958	0.750	0.842	0.506
2048	0.753	0.514	0.488	0.433	0.533	0.275
4096	0.453	0.295	0.275	0.225	0.261	0.141
Wave						
256	2.805	1.818	2.816	1.873	2.023	1.113
512	1.621	1.017	1.399	1.073	1.269	0.698
1024	0.924	0.608	0.667	0.549	0.657	0.344
2048	0.555	0.491	0.322	0.279	0.379	0.175
4096	0.345	0.638	0.150	0.133	0.166	0.071

Table 7: Mean-squared errors from 40 replications, signal to noise ratio of 5, sample points following a uniform $(0, 1)$ distribution.

Function n	R	R_l	B_c	B	B_l	B_E
Blip						
256	0.734	0.683	0.580	0.523	0.620	0.483
512	0.491	0.546	0.338	0.298	0.381	0.321
1024	0.323	0.373	0.196	0.181	0.224	0.189
2048	0.215	0.260	0.108	0.102	0.142	0.110
4096	0.139	0.203	0.087	0.068	0.082	0.069
Blocks						
256	2.110	1.681	2.931	1.255	1.206	1.400
512	1.293	1.312	1.650	0.878	0.902	0.963
1024	0.987	1.142	1.041	0.627	0.651	0.672
2048	0.680	0.737	0.621	0.424	0.460	0.447
4096	0.470	0.534	0.556	0.314	0.332	0.338
Bumps						
256	2.380	1.549	2.986	1.383	1.289	1.477
512	1.725	1.351	1.902	0.996	0.974	1.085
1024	1.204	1.114	1.144	0.714	0.730	0.727
2048	0.855	0.874	0.739	0.511	0.530	0.477
4096	0.586	0.622	0.554	0.354	0.363	0.299
Corner						
256	0.721	0.466	0.565	0.465	0.639	0.333
512	0.391	0.299	0.290	0.261	0.343	0.197
1024	0.257	0.253	0.160	0.154	0.217	0.130
2048	0.151	0.182	0.086	0.088	0.131	0.085
4096	0.090	0.137	0.052	0.045	0.061	0.045
Doppler						
256	1.771	1.509	2.208	1.155	1.091	0.684
512	1.179	1.132	1.344	0.826	0.839	0.474
1024	0.817	0.720	0.797	0.552	0.577	0.280
2048	0.510	0.432	0.422	0.333	0.369	0.162
4096	0.307	0.235	0.216	0.172	0.192	0.075
Heavisine						
256	1.422	0.637	0.937	0.671	0.752	0.595
512	0.886	0.366	0.541	0.412	0.466	0.352
1024	0.556	0.221	0.299	0.244	0.293	0.211
2048	0.389	0.142	0.176	0.150	0.179	0.143
4096	0.248	0.085	0.122	0.088	0.111	0.093
Spikes						
256	1.693	1.359	1.872	1.040	1.040	0.803
512	1.088	0.933	1.088	0.714	0.726	0.450
1024	0.695	0.578	0.580	0.427	0.469	0.276
2048	0.452	0.338	0.315	0.266	0.304	0.146
4096	0.266	0.174	0.180	0.142	0.164	0.076
Wave						
256	1.796	1.261	1.867	1.207	1.168	0.645
512	1.004	0.730	0.865	0.668	0.762	0.343
1024	0.577	0.530	0.413	0.372	0.468	0.167
2048	0.347	0.484	0.204	0.188	0.234	0.087
4096	0.200	0.951	0.095	0.089	0.105	0.040

Table 8: Mean-squared errors from 40 replications, signal to noise ratio of 7, sample points following a uniform $(0, 1)$ distribution.

Function n	R	R_l	B_c	B	B_l
Blip					
256	1.254	1.106	1.164	0.842	1.173
512	0.828	0.721	0.588	0.464	0.739
1024	0.527	0.480	0.332	0.289	0.447
2048	0.349	0.350	0.203	0.183	0.273
4096	0.232	0.221	0.149	0.119	0.160
Blocks					
256	3.342	2.571	5.252	2.357	2.267
512	2.448	1.948	3.445	1.738	1.744
1024	1.667	1.335	1.976	1.142	1.199
2048	1.191	1.009	1.169	0.756	0.802
4096	0.819	0.680	1.014	0.587	0.628
Bumps					
256	4.230	2.953	5.725	2.513	2.460
512	3.002	2.408	3.672	1.917	1.943
1024	2.046	1.829	2.044	1.270	1.310
2048	1.331	1.276	1.030	0.734	0.801
4096	0.854	0.986	0.696	0.492	0.529
Corner					
256	1.595	0.651	0.714	0.694	0.938
512	1.029	0.394	0.407	0.356	0.546
1024	0.594	0.266	0.215	0.240	0.405
2048	0.268	0.180	0.133	0.130	0.207
4096	0.151	0.124	0.097	0.084	0.110
Doppler					
256	2.167	1.950	2.358	1.429	1.511
512	1.533	1.296	1.351	0.913	1.036
1024	1.040	0.877	0.751	0.569	0.678
2048	0.662	0.528	0.382	0.317	0.406
4096	0.417	0.349	0.198	0.162	0.213
Heavisine					
256	1.527	0.785	1.088	0.933	1.156
512	1.206	0.523	0.698	0.534	0.666
1024	0.892	0.369	0.434	0.331	0.419
2048	0.604	0.258	0.242	0.199	0.270
4096	0.395	0.185	0.190	0.139	0.168
Spikes					
256	2.418	1.697	2.138	1.542	1.696
512	1.360	0.959	0.925	0.806	1.026
1024	0.845	0.557	0.539	0.481	0.603
2048	0.583	0.360	0.327	0.282	0.375
4096	0.363	0.228	0.200	0.156	0.198
Wave					
256	2.081	1.245	2.314	1.406	1.846
512	1.373	0.675	1.083	0.720	1.012
1024	0.835	0.547	0.398	0.333	0.536
2048	0.521	0.535	0.167	0.164	0.283
4096	0.325	0.559	0.079	0.077	0.122

Table 9: Mean-squared errors from 40 replications, signal to noise ratio of 5, sample points equispaced over $(0, 1)$.

Function n	R	R_l	B_c	B	B_l
Blip					
256	0.662	0.575	0.481	0.382	0.588
512	0.461	0.385	0.281	0.240	0.345
1024	0.296	0.245	0.164	0.152	0.219
2048	0.198	0.159	0.110	0.099	0.130
4096	0.134	0.103	0.083	0.064	0.079
Blocks					
256	2.108	1.991	3.402	1.346	1.233
512	1.392	1.230	1.826	0.924	0.922
1024	0.993	0.935	1.050	0.632	0.652
2048	0.676	0.681	0.608	0.411	0.434
4096	0.468	0.485	0.577	0.318	0.335
Bumps					
256	2.316	1.591	2.499	1.341	1.295
512	1.624	1.405	1.799	0.971	0.981
1024	1.079	1.348	0.907	0.633	0.670
2048	0.735	1.135	0.523	0.407	0.438
4096	0.461	0.986	0.371	0.265	0.285
Corner					
256	0.861	0.298	0.431	0.380	0.521
512	0.346	0.174	0.230	0.219	0.349
1024	0.213	0.105	0.130	0.131	0.185
2048	0.140	0.067	0.084	0.077	0.119
4096	0.084	0.039	0.064	0.047	0.061
Doppler					
256	1.282	1.218	1.408	0.773	0.807
512	0.831	0.759	0.701	0.492	0.548
1024	0.578	0.505	0.406	0.299	0.350
2048	0.381	0.292	0.221	0.181	0.217
4096	0.240	0.163	0.109	0.088	0.109
Heavisine					
256	1.288	0.525	0.799	0.550	0.613
512	0.871	0.336	0.434	0.297	0.378
1024	0.591	0.234	0.235	0.189	0.249
2048	0.389	0.171	0.140	0.118	0.152
4096	0.135	0.124	0.116	0.084	0.099
Spikes					
256	1.349	0.926	1.274	0.833	0.880
512	0.760	0.474	0.522	0.433	0.533
1024	0.505	0.304	0.312	0.272	0.364
2048	0.317	0.179	0.160	0.141	0.195
4096	0.204	0.096	0.101	0.081	0.101
Wave					
256	1.108	0.524	1.088	0.663	0.949
512	0.732	0.378	0.426	0.347	0.522
1024	0.444	0.564	0.176	0.170	0.290
2048	0.281	0.772	0.084	0.088	0.148
4096	0.170	0.949	0.042	0.043	0.068

Table 10: Mean-squared errors from 40 replications, signal to noise ratio of 7, sample points equispaced over $(0, 1)$.

Lemma 1 *For $f \in \Lambda^\alpha(M)$, $\alpha > 0$, and $M < \infty$,*

$$\left|n^{-1/2} f\left((k+1)/(n+1)\right) - \xi_{Jk}\right| \le C n^{-1/2-(\alpha \wedge 1)}.$$

Proof: This follows from Daubechies (1992). Since $\int \phi = 1$,

$$\int \phi_{Jk}(x)dx = \int 2^{J/2}\phi(2^J x - k)dx = 2^{-J/2}\int \phi(x)dx = 2^{-J/2} = n^{-1/2}.$$

Using this,

$$\begin{aligned}
\left|n^{-1/2} f\left(\frac{k+1}{n+1}\right) - \xi_{Jk}\right| &= \left|n^{-1/2} f\left(\frac{k+1}{n+1}\right)\int_0^1 \phi(x)dx \right.\\
&\qquad \left. - \int_{k/2^J}^{(1+k)/2^J} f(x)\phi_{Jk}(x)dx\right| \\
&= \left|n^{-1/2} f\left(\frac{k+1}{n+1}\right)\int_0^1 \phi(x)dx \right.\\
&\qquad \left. -2^{-J/2}\int_0^1 f\left(2^{-J}(x+k)\right)\phi(x)dx\right| \\
&= n^{-1/2}\left|\int_0^1 \phi(x)\left[f\left(\frac{k+1}{n+1}\right) - f\left(\frac{x+k}{n}\right)\right]dx\right| \\
&\le n^{-1/2}\int_0^1 |\phi(x)|\cdot\left|f\left(\frac{k+1}{n+1}\right) - f\left(\frac{x+k}{n}\right)\right|dx \\
&\le Cn^{-1/2}\int_0^1 \left|\frac{k+1}{n+1} - \frac{x+k}{n}\right|^{\alpha\wedge 1} dx \\
&= Cn^{-1/2}\int_0^1 \left|\frac{x}{n} + \frac{k+n}{n(n+1)}\right|^{\alpha\wedge 1} dx \\
&\le Cn^{-1/2}\int_0^1 (x/n)^{\alpha\wedge 1} dx.
\end{aligned}$$

Since the function f has finite domain and ϕ can be assumed to be bounded, this implies

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

Lemma 1 can be used to find a reasonable approximation to f. See Cai (1996).

Lemma 2 *Let $f_n(x) = \sum_{k=0}^{n-1} n^{-1/2} f\left(\frac{k+1}{n+1}\right)\phi_{Jk}(x)$. For $f \in \Lambda^\alpha(M)$, $\alpha > 0$, and $M < \infty$,*

$$\|f_n - f\|_2^2 \le Cn^{-2(\alpha\wedge 1)}.$$

While this gives a good convergence rate theoretically, it is not very useful in a practical sense. All the benefits of decomposing a function into its coarse and detail parts is lost.

Proof: Using the orthogonality of the wavelet functions,

$$\begin{aligned}
\|f_n - f\|_2^2 &= \int_0^1 (f_n(x) - f(x))^2\, dx \\
&= \int_0^1 \left(\sum_{k=0}^{n-1} n^{-1/2} f\left(\frac{k+1}{n+1}\right) \phi_{Jk}(x) - \sum_{k=0}^{n-1} \xi_{Jk}\phi_{Jk}(x) \right. \\
&\quad \left. - \sum_{j=J}^{\infty} \sum_k \theta_{jk}\psi_{jk}(x) \right)^2 dx \\
&\le 2\int_0^1 \left(\sum_{k=0}^{n-1} n^{-1/2} f\left(\frac{k+1}{n+1}\right) \phi_{Jk}(x) - \sum_{k=0}^{n-1} \xi_{Jk}\phi_{Jk}(x) \right)^2 dx \\
&\quad +2\int \left(\sum_{j=J}^{\infty} \sum_k \theta_{jk}\psi_{jk}(x) \right)^2 dx \\
&= C\int \sum_{k=0}^{n-1} \left(n^{-1/2} f\left(\frac{k+1}{n+1}\right) - \xi_{Jk} \right)^2 \phi_{Jk}^2(x) dx \\
&\quad +C\int \sum_{j=J}^{\infty} \sum_k \theta_{jk}^2 \psi_{jk}^2(x) dx.
\end{aligned}$$

Now, $\int \phi = \int \phi^2 = \int \phi_{jk}^2 = \int \psi^2 = \int \psi_{jk}^2 = 1$. Therefore,

$$\|f_n - f\|_2^2 \le C\sum_{k=0}^{n-1} \left(n^{-1/2} f\left(\frac{k+1}{n+1}\right) - \xi_{Jk} \right)^2 + C\sum_{j=J}^{\infty} \sum_k \theta_{jk}^2.$$

Using lemma 1 and the decay rate of wavelet coefficients,

$$\begin{aligned}
\|f_n - f\|_2^2 &\le C\sum_{k=0}^{n-1} \left(n^{-1/2-(\alpha\wedge 1)} \right)^2 + C\sum_{j=J}^{\infty} \sum_k 2^{-2j(\alpha+1/2)} \\
&= Cn^{-2(\alpha\wedge 1)} + C\sum_{j=J}^{\infty} 2^{-2j\alpha} \\
&= Cn^{-2(\alpha\wedge 1)} + Cn^{-2\alpha} \\
&\le Cn^{-2(\alpha\wedge 1)}.
\end{aligned}$$

Now, f_n is a good approximation to f by lemma 2. However, it assumes that the function is known at the n equally spaced points $(k+1)/(n+1)$. Since only the y_k are observed in (4) or (5), the values of the function at these points are not known. But, these observed values of y_k will be used to approximate f. The analog of f_n using y_k rather than $f((k+1)/(n+1))$ is

$$\tilde{f}(x) = \sum_{k=0}^{n-1} n^{-1/2} y_k \phi_{Jk}(x).$$

This function has a noise component, so thresholding will be necessary. The proofs of Theorems 1 and 2 are now ready to commence.

5.2 Theorem 2

Let $\tilde{f}(x)$ be as above. Then

$$\tilde{f}(x) = \sum_{k=0}^{n-1} \left(\xi_{Jk} + \left(n^{-1/2} f(X_{(k)}) - \xi_{Jk}\right) + n^{-1/2}\sigma\varepsilon_k\right) \phi_{Jk}(x).$$

Using the multiresolution properties of wavelets, the DWT of $\tilde{f}$ can be written as (for some fixed $j_0 \leq J$)

$$\tilde{f}(x) = \sum_{i=0}^{2^{j_0}-1} (\xi_{j_0 i} + \nu_{j_0 i} + \eta_{j_0 i})\, \phi_{j_0 i}(x) + \sum_{j=j_0}^{J-1} \sum_i (\theta_{ji} + \zeta_{ji} + \delta_{ji})\, \psi_{ji}(x).$$

The $\xi_{j_0 i}$ and θ_{ji} are the DWT coefficients associated with $\sum_{k=0}^{n-1} \xi_{Jk}\phi_{Jk}(x)$, $\nu_{j_0 i}$ and ζ_{ji} are for $\sum_{k=0}^{n-1} \left(n^{-1/2} f(X_{(k)}) - \xi_{Jk}\right) \phi_{Jk}(x)$, and $\eta_{j_0 i}$ and δ_{ji} go with $\sum_{k=0}^{n-1} n^{-1/2}\sigma\varepsilon_k\phi_{Jk}(x)$. Note that $\xi_{j_0 i}$ and θ_{ji} have no random component, $\nu_{j_0 i}$ and ζ_{ji} have a random component due to the x_k, and $\eta_{j_0 i}$ and δ_{ji} have a random component due to the noise. The coefficients $\xi_{j_0 i}$, $\nu_{j_0 i}$, and $\eta_{j_0 i}$ represent the coarse part of $\tilde{f}$ and will not be thresholded. Let $\hat{\xi}_{j_o i} = \xi_{j_0 i} + \nu_{j_0 i} + \eta_{j_0 i}$ be the sum of these coarse coefficients. Let $\hat{\theta}_{ji} = \theta_{ji} + \zeta_{ji} + \delta_{ji}$ be the sum of the detail coefficients. By the orthogonality of the DWT, $\hat{\theta}_{ji} \sim \mathrm{N}(\theta_{ji} + \zeta_{ji}, \sigma^2/n)$. These $\hat{\theta}_{ji}$ will be thresholded via the rule at (6), resulting in new detail coefficients $\tilde{\theta}_{ji}$. The estimate of f is then

$$\hat{f}(x) = \sum_{i=0}^{2^{j_0}-1} \hat{\xi}_{j_0 i}\phi_{j_0 i}(x) + \sum_{j=j_0}^{J-1} \sum_i \tilde{\theta}_{ji}\psi_{ji}(x)$$

and the error is

$$\begin{aligned}
\mathbb{E}\|\hat{f} - f\|_2^2 &= \mathbb{E}\left\|\hat{f} - \left(\sum_{k=0}^{2^J-1} \xi_{Jk}\phi_{Jk} + \sum_{j=J}^{\infty}\sum_k \theta_{jk}\psi_{jk}\right)\right\|_2^2 \\
&\leq C\mathbb{E}\left\|\hat{f} - \sum_{k=0}^{2^J-1} \xi_{Jk}\phi_{Jk}\right\|_2^2 + C\mathbb{E}\left\|\sum_{j=J}^{\infty}\sum_k \theta_{jk}\psi_{jk}\right\|_2^2 \\
&\leq C\mathbb{E}\left\|\hat{f} - \sum_{i=0}^{2^{j_0}-1} \xi_{j_0 i}\phi_{j_0 i} - \sum_{j=j_0}^{J-1}\sum_i \theta_{ji}\psi_{ji}\right\|_2^2 + Cn^{-2\alpha} \\
&= C\sum_{i=0}^{2^{j_0}-1} \mathbb{E}(\hat{\xi}_{j_0 i} - \xi_{j_0 i})^2 + C\sum_{j=j_0}^{J-1}\sum_i \mathbb{E}(\tilde{\theta}_{ji} - \theta_{ji})^2 + Cn^{-2\alpha}
\end{aligned}$$

Using orthogonality of the DWT again, $\nu_{j_0 i} + \eta_{j_0 i} \sim \mathrm{N}(\nu_{j_0 i}, \sigma^2/n)$. Letting $\mathbb{E}_X$ be expectation with respect to the data, and $\mathbb{E}_{(\cdot|X)}$ be expectation given the data, the coarse part of the error is

$$\begin{aligned}
\sum_{i=0}^{2^{j_0}-1} \mathbb{E}(\hat{\xi}_{j_0 i} - \xi_{j_0 i})^2 &= \mathbb{E}_X \sum_{i=0}^{2^{j_0}-1} \mathbb{E}_{(\cdot|X)}(\nu_{j_0 i} + \eta_{j_0 i})^2 \\
&= \mathbb{E}_X \sum_{i=0}^{2^{j_0}-1} (\nu_{j_0 i}^2 + \sigma^2/n) \\
&= 2^{j_0}\sigma^2/n + \mathbb{E}_X \sum_{i=0}^{2^{j_0}-1} \nu_{j_0 i}^2 \\
&= Cn^{-1} + \mathbb{E}_X \sum_{i=0}^{2^{j_0}-1} \nu_{j_0 i}^2.
\end{aligned}$$

For the detail portion of the error, write

$$\begin{aligned}
\sum_{j=j_0}^{J-1} \sum_i \mathbb{E}(\tilde{\theta}_{ji} - \theta_{ji})^2 &= \mathbb{E}_X \sum_{j=j_0}^{J-1} \sum_{b=1}^{B_j} \mathbb{E}_{(\cdot|X)} \sum_k (\tilde{\theta}_{ji} - \theta_{ji})^2 \\
&= \mathbb{E}_X \sum_{j=j_0}^{J-1} \sum_{b=1}^{B_j} \mathbb{E}_{(\cdot|X)} \sum_k (\tilde{\theta}_{ji} - (\theta_{ji} + \zeta_{ji}) + \zeta_{ji})^2 \\
&\leq 2\mathbb{E}_X \sum_{j=j_0}^{J-1} \sum_{b=1}^{B_j} \mathbb{E}_{(\cdot|X)} \sum_k (\tilde{\theta}_{ji} - (\theta_{ji} + \zeta_{ji}))^2 + 2\mathbb{E}_X \sum_{j=j_0}^{J-1} \sum_k \zeta_{ji}^2
\end{aligned}$$

where the summation b is over all B blocks of coefficients in a resolution level j and k is the summation within a particular block. Since the block length is $L = \log n$, there are no more than $2^j/\log n$ blocks in each resolution level j. Before bounding this detail portion of the error, a theorem from Cai (1998) is stated.

Theorem 3 *Let $x_i = \mu_i + \sigma z_i$, $i = 1, 2, \ldots, L$ with $z_i \sim$ iid normal (0,1) random variables. Let*

$$\hat{\mu}_i = \left(1 - \frac{\lambda L \sigma^2}{S^2}\right)_+ x_i,$$

where $\lambda \geq 1$ is a constant, and $S^2 = \sum_{j=1}^{L} x_i^2$. Then,

$$\mathbb{E}\|\hat{\mu} - \mu\|_2^2 \leq \left(\|\mu\|^2 \wedge \lambda L \sigma^2\right) + 4\sigma^2 P(\chi_L^2 > \lambda L).$$

If $L = \log n$ and $\lambda \geq 4.50524$, then

$$\mathbb{E}\|\hat{\mu} - \mu\|_2^2 \leq \left(\|\mu\|^2 \wedge \lambda \sigma^2 \log n\right) + 4\sigma^2/n.$$

Using this theorem with $x_i = \hat{\theta}_{ji}$, $\mu_i = \theta_{ji} + \zeta_{ji}$, $\hat{\mu}_i = \tilde{\theta}_{ji}$, and σ^2/n in place of σ^2,

$$\begin{aligned}
\mathbb{E}_{(\cdot|X)} \sum_k (\tilde{\theta}_{ji} - (\theta_{ji} + \zeta_{ji}))^2 &= \mathbb{E}_{(\cdot|X)} \|\tilde{\theta}_{j\cdot} - (\theta_{j\cdot} + \zeta_{j\cdot})\|^2 \\
&\leq \left(\|\theta_{j\cdot} + \zeta_{j\cdot}\|^2 \wedge \lambda \sigma^2 n^{-1} \log n\right) + 4\sigma^2 n^{-2} \\
&\leq \left(2\|\theta_{j\cdot}\|^2 \wedge \lambda \sigma^2 n^{-1} \log n\right) + 2\|\zeta_{j\cdot}\|^2 + Cn^{-2} \\
&\leq C\left[\left(\|\theta_{j\cdot}\|^2 \wedge n^{-1} \log n\right) + \|\zeta_{j\cdot}\|^2 + n^{-2}\right]
\end{aligned}$$

Therefore,

$$\sum_{j=j_0}^{J-1}\sum_i \mathbb{E}(\tilde{\theta}_{ji}-\theta_{ji})^2 \;\; \le \;\; C\mathbb{E}_X\sum_{j=j_0}^{J-1}\sum_{b=1}^{B_j}\left[\left(\|\theta_{j\cdot}\|^2\wedge n^{-1}\log n\right)+n^{-2}\right]+C\mathbb{E}_X\sum_{j=j_0}^{J-1}\sum_k \zeta_{ji}^2.$$

Let j^* be the integer such that $2^{j^*}\le n^{1/(2\alpha+1)}<2^{j^*+1}$. Then

$$\begin{aligned}
\mathbb{E}_X\sum_{j=j_0}^{J-1}\sum_{b=1}^{B_j}\left(\|\theta_{j\cdot}\|^2\wedge n^{-1}\log n\right) &\le \sum_{j=j_0}^{j^*-1}\sum_{b=1}^{B_j} n^{-1}\log n+\sum_{j=j^*}^{J-1}\sum_{b=1}^{B_j}\sum_k \theta_{jk}^2\\
&\le \sum_{j=j_0}^{j^*-1} 2^j(\log n)^{-1}n^{-1}\log n+\sum_{j=j^*}^{J-1}\sum_k\theta_{jk}^2\\
&\le 2^{j^*}n^{-1}+\sum_{j=j^*}^{J-1}\sum_k C2^{-2j(\alpha+1/2)}\\
&\le n^{-2\alpha/(2\alpha+1)}+C\sum_{j=j^*}^{J-1}2^j2^{-2j(\alpha+1/2)}\\
&\le n^{-2\alpha/(2\alpha+1)}+C2^{-2j^*\alpha}\\
&\le Cn^{-2\alpha/(2\alpha+1)}.
\end{aligned}$$

The detail portion of the error is then bounded:

$$\sum_{j=j_0}^{J-1}\sum_i \mathbb{E}(\tilde{\theta}_{ji}-\theta_{ji})^2\le Cn^{-2\alpha/(2\alpha+1)}+C\mathbb{E}_X\sum_{j=j_0}^{J-1}\sum_k\zeta_{ji}^2,$$

and the entire error is then bounded by

$$\begin{aligned}
\mathbb{E}\|\hat{f}-f\|_2^2 &\le Cn^{-2\alpha/(2\alpha+1)}+C\mathbb{E}_X\left(\sum_{i=0}^{2^{j_0}-1}\nu_{j_0i}^2+\sum_{j=j_0}^{J-1}\sum_k\zeta_{ji}^2\right)\\
&= Cn^{-2\alpha/(2\alpha+1)}+C\mathbb{E}_X\int\left(\sum_{i=0}^{2^{j_0}-1}\nu_{j_0i}\phi_{j_0i}(y)+\sum_{j=j_0}^{J-1}\sum_k\zeta_{ji}\psi_{ji}(y)\right)^2dy\\
&= Cn^{-2\alpha/(2\alpha+1)}+C\mathbb{E}_X\int\left(\sum_{k=0}^{n-1}\left(n^{-1/2}f(X_{(k)})-\xi_{Jk}\right)\phi_{Jk}(y)\right)^2dy\\
&= Cn^{-2\alpha/(2\alpha+1)}+C\mathbb{E}_X\int\sum_{k=0}^{n-1}\left(n^{-1/2}f(X_{(k)})-\xi_{Jk}\right)^2\phi_{Jk}^2(y)dy\\
&= Cn^{-2\alpha/(2\alpha+1)}+C\mathbb{E}_X\sum_{k=0}^{n-1}\left(n^{-1/2}f(X_{(k)})-\xi_{Jk}\right)^2.
\end{aligned}$$

All that remains is to bound $\mathbb{E}_X \sum_{k=0}^{n-1} \left(n^{-1/2} f(X_{(k)}) - \xi_{Jk}\right)^2$.

$$\begin{aligned}
\left|n^{-1/2} f(X_{(k)}) - \xi_{Jk}\right| &= \left|n^{-1/2} f(X_{(k)}) - \int_{k/n}^{(1+k)/n} f(y)\phi_{Jk}(y)dy\right| \\
&= \left|\int_{k/n}^{(1+k)/n} \left[f(X_{(k)}) - f(y)\right]\phi_{Jk}(y)dy\right| \\
&\leq \sqrt{\left\{\int_{k/n}^{(1+k)/n} \left[f(X_{(k)}) - f(y)\right]^2 dy\right\}\left\{\int_{k/n}^{(1+k)/n} \phi_{Jk}^2(y)dy\right\}} \\
&= \sqrt{\int_0^{1/n} \left[f(X_{(k)}) - f(y + k/n)\right]^2 dy}.
\end{aligned}$$

This implies

$$\begin{aligned}
\mathbb{E}_X \sum_{k=0}^{n-1} \left(n^{-1/2} f(X_{(k)}) - \xi_{Jk}\right)^2 &\leq \mathbb{E}_X \sum_{k=0}^{n-1} \int_0^{1/n} \left[f(X_{(k)}) - f(y + k/n)\right]^2 dy \\
&\leq \mathbb{E}_X \sum_{k=0}^{n-1} \int_0^{1/n} \left[f(X_{(k)}) - f(y + X_{(k)})\right]^2 dy \\
&\quad + \mathbb{E}_X \sum_{k=0}^{n-1} \int_0^{1/n} \left[f(y + X_{(k)}) - f(y + k/n)\right]^2 dy.
\end{aligned}$$

The first term on the right is easily bounded,

$$\begin{aligned}
\mathbb{E}_X \sum_{k=0}^{n-1} \int_0^{1/n} \left[f(X_{(k)}) - f(y + X_{(k)})\right]^2 dy &\leq C \sum_{k=0}^{n-1} \int_0^{1/n} (y^2)^{\alpha\wedge 1} dy \\
&= C \int_0^{1/n} n (y^2)^{\alpha\wedge 1} dy \\
&= C\mathbb{E}(Y^2)^{\alpha\wedge 1}.
\end{aligned}$$

where Y is a uniform $(0, 1/n)$ random variable. Using Jensen's inequality,

$$\mathbb{E}(Y^2)^{\alpha\wedge 1} \leq \left(\mathbb{E}Y^2\right)^{\alpha\wedge 1} = Cn^{-2(\alpha\wedge 1)}.$$

For the second term,

$$
\begin{aligned}
\mathbb{E}_X \sum_{k=0}^{n-1} \int_0^{1/n} \left[f(y+X_{(k)}) - f(y+k/n)\right]^2 dy &\leq C\mathbb{E}_X \sum_{k=0}^{n-1} \int_0^{1/n} \left[(X_{(k)} - k/n)^2\right]^{\alpha\wedge 1} dy \\
&= Cn^{-1} \sum_{k=0}^{n-1} \mathbb{E}_X \left[(X_{(k)} - k/n)^2\right]^{\alpha\wedge 1} \\
&\leq Cn^{-1} \sum_{k=0}^{n-1} \mathbb{E}_X \left[(X_{(k)} - \mu_{(k)})^2\right]^{\alpha\wedge 1} \\
&\quad + Cn^{-1} \sum_{k=0}^{n-1} \mathbb{E}_X \left[(\mu_{(k)} - k/n)^2\right]^{\alpha\wedge 1} \\
&\leq Cn^{-1} \sum_{k=0}^{n-1} \left(\mathbb{E}_X (X_{(k)} - \mu_{(k)})^2\right)^{\alpha\wedge 1} \\
&\quad + Cn^{-1} \sum_{k=0}^{n-1} (\mu_{(k)} - k/n)^{2(\alpha\wedge 1)},
\end{aligned}
$$

where $\mu_{(k)} = \mathbb{E}_X X_{(k)}$. Since $var(X_{(k)}) \leq C/n$ and $|\mu_{(k)} - k/n| \leq Cn^{-1/2}$, this term is now bounded,

$$
\mathbb{E}_X \sum_{k=0}^{n-1} \int_0^{1/n} \left[f(y+X_{(k)}) - f(y+k/n)\right]^2 dy \leq Cn^{-(\alpha\wedge 1)}.
$$

Note that if the variance was bounded by C/n^2, the bound above would be of order $n^{-2(\alpha\wedge 1)}$. Putting all these bounds together, we have

$$
\mathbb{E}\|\hat{f} - f\|_2^2 \leq Cn^{-2\alpha/(2\alpha+1)} + Cn^{-(\alpha\wedge 1)}.
$$

If $\alpha \geq 1/2$, this is

$$
\mathbb{E}\|\hat{f} - f\|_2^2 \leq Cn^{-2\alpha/(2\alpha+1)}.
$$

If the tighter variance bound is used, this rate is achieved for $\alpha > 0$. This completes the proof of Theorem 1.

5.3 Theorem 1

The rate of the process is n, not necessarily a dyadic integer as in Theorem 2, say $2^l < n \leq 2^{l+1}$. Recall that the sample points are labeled $S_0, S_1, \ldots, S_k, \ldots$. Let N is the largest dyadic integer such that $S_{N-1} \leq 1$ and $S_{2N-1} > 1$, and N^* be the largest integer such that $S_{N^*-1} \leq 1$ and $S_{N^*} > 1$. Clearly $N \leq N^* < 2N$. If $N \geq 2n$ or $N \leq n/4$, f will be estimated as 0 on the entire interval. For $N \in (n/4, 2n)$, the function will be estimated in two pieces, first on the interval $[0, S_{N-1}]$ using the points $S_0, S_1, \ldots S_{N-1}$, then on $[S_{N^*-N}, 1]$ using the points $S_{N^*-N}, S_{N^*-N+1}, \ldots S_{N^*-1}$. On each of these intervals, there are N points S_i. The theorem will be proved for the first interval. The proof for the second interval follows by symmetry.

In general, the process wll not have its starting point at the beginning of the interval. Therefore, the first arrival time is actually a truncated value and S_k would be written as

$$S_k = X_0^* + \sum_{i=1}^{k} X_i$$

where X_0^* is the (possibly) truncated value of X_0. Then

$$0 \le \mathbb{E}X_0^* \le \mathbb{E}X_0 = 1/n,$$

and

$$0 \le \mathbb{E}(X_0^*)^2 \le \mathbb{E}(X_0)^2.$$

Therefore,

$$\begin{aligned} var(S_0) &= var(X_0^*) \\ &= \mathbb{E}(X_0^*)^2 - (\mathbb{E}X_0^*)^2 \\ &\le \mathbb{E}X_0^2 - (\mathbb{E}X_0^*)^2 \\ &= \mathbb{E}X_0^2 - (\mathbb{E}X_0)^2 + (\mathbb{E}X_0)^2 - (\mathbb{E}X_0^*)^2 \\ &\le var(X_0) + n^{-2} \\ &\le C/n^2. \end{aligned}$$

For simplicity, the proof assumes that the first interarrival time is not truncated. The proof is easily modified if this is not the case.

Bounds are needed for $P(N \ge 2n)$ and $P(N \le n/4)$. For the first probability,

$$P(N \ge 2n) \le P(N^* \ge 2n) = P(S_{2n-1} < 1)$$

Now,

$$\mathbb{E}S_{2n-1} = 2 \quad \text{and} \quad var(S_{2n-1}) \le C/n.$$

Using Markov's inequality,

$$\begin{aligned} P(S_{2n-1} < 1) &= P(S_{2n-1} - 2 < -1) \\ &\le P(|S_{2n-1} - 2| > 1) \\ &\le var(S_{2n-1}) \\ &\le C/n. \end{aligned}$$

For the second probability, $N \le n/4$ and $N^* < 2N$ imply that $N^* < n/2$. Therefore,

$$P(N \le n/4) \le P(N^* < n/2) = \begin{cases} P(S_{n/2} > 1), & n \text{ even} \\ P(S_{(n+1)/2} > 1), & n \text{ odd} \end{cases}$$

For n even,

$$\mathbb{E}S_{n/2} = 1/2 + 1/n \quad \text{and } var(S_{n/2}) \le C(1/(2n) + 1/n^2) \le C/n,$$

and Markov's inequality gives

$$\begin{aligned} P(S_{n/2} > 1) &= P(S_{n/2} - 1/2 - 1/n > 1 - 1/2 - 1/n) \\ &\le P(|S_{n/2} - 1/2 - 1/n| > 1/2 - 1/n) \\ &\le var(S_{n/2})\,(1/2 - 1/n))^{-2} \\ &\le C/n \end{aligned}$$

provided $n \ge 3$. Similar results hold for n odd. Therefore, $P(N \le n/4, N \ge 2n) \le C/n$. Since $2^l < n \le 2^{l+1}$, restricting attention to $N \in (n/4, 2n)$ allows N to only take on the values 2^{l-1},2^l, and 2^{l+1}.

The estimator first uses the N points $X_0, X_1, \ldots, X_{N-1}$. Let $T = S_{N-1} \le 1$. If $T < N/(2n)$, then set the estimate $\hat{f}$ to 0. Otherwise, if $T > N/(2n)$ rescale the arrival times from $[0,T]$ to $[0,1]$ by dividing each point X_i by T, $X'_i = X_i/T$. Given T,

$$\mathbb{E}X'_i = 1/(nT) \quad \text{and} \quad var(X'_i) \le C/(n^2T^2).$$

The new function $f_T(x) = f(xT)$ still remains in the Hölder space $\Lambda^\alpha(M)$:

$$|f_T(x) - f_T(y)| = |f(xT) - f(yT)| \le T^\alpha M|x-y|^\alpha \le M|x-y|^\alpha$$

for $\alpha \le 1$. For $\alpha > 1$, $f_T^{(i)}(x) = T^i f^{(i)}(xT)$ and

$$\begin{aligned} |f_T^{(\lfloor\alpha\rfloor)}(x) - f_T^{(\lfloor\alpha\rfloor)}(y)| &= T^{(\lfloor\alpha\rfloor)}|f^{(\lfloor\alpha\rfloor)}(xT) - f^{(\lfloor\alpha\rfloor)}(yT)| \\ &\le T^{(\lfloor\alpha\rfloor)}M|xT - yT|^s \\ &\le M|x-y|^s. \end{aligned}$$

For the remainder of this proof, the subscript T will dropped from the scaled f.

By conditioning on N and T, and replacing $X_{(k)}$ with the rescaled S'_k, the proof of Theorem 2 is unchanged up to the bound

$$\mathbb{E}\|f - \hat{f}\|_2^2 \le CN^{-2(\alpha\wedge 1)} + C\mathbb{E}\sum_{k=0}^{N-1}\left(N^{-1/2}f(S'_k) - \xi_{Jk}\right)^2.$$

Here, $J = \log_2 N$. Note that the randomness in this expression is due to X, T, and N, but not the noise. As before,

$$\begin{aligned} \mathbb{E}\sum_{k=0}^{N-1}\left(N^{-1/2}f(S'_k) - \xi_{Jk}\right)^2 &\le \mathbb{E}\sum_{k=0}^{N-1}\int_0^{1/N}[f(S'_k) - f(y+S'_k)]^2\,dy \\ &\quad + \mathbb{E}\sum_{k=0}^{N-1}\int_0^{1/N}[f(y+S'_k) - f(y+k/N)]^2\,dy. \end{aligned}$$

The first term on the right is bounded by $CN^{-2(\alpha\wedge 1)}$ (conditional on N) just as in theorem 2. The second term, given N and T, is bounded by

$$\begin{aligned} \mathbb{E}\sum_{k=0}^{N-1}\int_0^{1/N}[f(y+S'_k) - f(y+k/N)]^2\,dy &\le C\mathbb{E}N^{-1}\sum_{k=0}^{N-1}\left((S'_k - \mu_k)^2\right)^{\alpha\wedge 1} \\ &\quad + CN^{-1}\sum_{k=0}^{N-1}(\mu_k - k/N)^{2(\alpha\wedge 1)}, \end{aligned}$$

where

$$\mu_k = \mathbb{E}S'_k = \mathbb{E}\sum_{i=0}^{k} X_i/T = \frac{k+1}{nT}.$$

Given N and T and using Jensen's inequality,

$$\begin{aligned}
\mathbb{E}N^{-1}\sum_{k=0}^{N-1}\left((S'_k-\mu_k)^2\right)^{\alpha\wedge 1} &\leq N^{-1}\sum_{k=0}^{N-1}\left(\mathbb{E}\left(S'_k-\mu_k\right)^2\right)^{\alpha\wedge 1}\\
&= N^{-1}\sum_{k=0}^{N-1}\left(var(S'_k)\right)^{\alpha\wedge 1}\\
&\leq CN^{-1}\sum_{k=0}^{N-1}\left(k/(n^2T^2)\right)^{\alpha\wedge 1}\\
&= C(N/(n^2T^2))^{\alpha\wedge 1}N^{-1}\sum_{k=0}^{N-1}(k/N)^{\alpha\wedge 1}\\
&\leq C\left(Nn^{-2}T^{-2}\right)^{\alpha\wedge 1}.
\end{aligned}$$

Now to bound $N^{-1}\sum_{k=0}^{N-1}(\mu_k - k/N)^{2(\alpha\wedge 1)}$.

$$\begin{aligned}
N^{-1}\sum_{k=0}^{N-1}(\mu_k - k/N)^{2(\alpha\wedge 1)} &= N^{-1}\sum_{k=0}^{N-1}\left[\left(\frac{k+1}{nT}\right)-\frac{k}{N}\right]^{2(\alpha\wedge 1)}\\
&\leq 2N^{-1}\sum_{k=0}^{N-1}\left[\frac{1}{nT}\right]^{2(\alpha\wedge 1)} + 2N^{-1}\sum_{k=0}^{N-1}\left[\frac{k}{nT}-\frac{k}{N}\right]^{2(\alpha\wedge 1)}.
\end{aligned}$$

The index k runs from 0 to $N-1$, so this is bounded by

$$N^{-1}\sum_{k=0}^{N-1}(\mu_k - k/N)^{2(\alpha\wedge 1)} \leq 2(nT)^{-2(\alpha\wedge 1)} + 2\left[\frac{N-nT}{nT}\right]^{2(\alpha\wedge 1)}.$$

To bound the overall error, let

$$A = \{N \leq n/2, N \geq 2n\}, B = \{T = S_{N-1} < N/(2n)\}.$$

Then

$$\begin{aligned}
\mathbb{E}\left(\|\hat{f}-f\|_2^2 \,\middle|\, N, T\right) &\leq \|f\|_2^2 I(A\cup A^cB) + CN^{-2(\alpha\wedge 1)}I(A^cB^c)\\
&\quad + C\left[\left((N+1)n^{-2}T^{-2}\right)^{\alpha\wedge 1} + \left(\frac{N-nT}{nT}\right)^{2(\alpha\wedge 1)}\right] I(A^cB^c),
\end{aligned}$$

and,

$$\begin{aligned}
\mathbb{E}\|\hat{f}-f\|_2^2 &\leq \|f\|_2^2 P(A\cup A^cB) + C\mathbb{E}N^{-2(\alpha\wedge 1)}I(A^cB^c)\\
&\quad + C\mathbb{E}\left[\left((N+1)n^{-2}T^{-2}\right)^{\alpha\wedge 1} + \left(\frac{N-nT}{nT}\right)^{2(\alpha\wedge 1)}\right] I(A^cB^c).
\end{aligned}$$

Since f is bounded in L_2,

$$\begin{aligned}\|f\|_2^2 P(A\cup A^cB) &\le C\left(P(A)+P(A^cB)\right)\\ &\le C/n + C\cdot P(A^cB).\end{aligned}$$

To find $P(A^cB)$, recall that $2^l < n \le 2^{l+1}$. Then

$$\begin{aligned}P(A^cB) &= P(T<N/(2n), n/4<N<2n)\\ &= P(S_{N-1}<N/(2n), N\in\{2^{l-1},2^l,2^{l+1}\})\\ &= P(S_{2^{l-1}-1}<2^{l-1}/(2n), N=2^{l-1}) + P(S_{2^l-1}<2^l/(2n), N=2^l)\\ &\quad +P(S_{2^{l+1}-1}<2^{l+1}/(2n), N=2^{l+1})\\ &\le P(S_{2^{l-1}-1}<2^{l-1}/(2n)) + P(S_{2^l-1}<2^l/(2n)) + P(S_{2^{l+1}-1}<2^{l+1}/(2n)).\end{aligned}$$

The bound for each of these three pieces is similar. Only the first is found here. Using

$$\mathbb{E}S_{2^{l-1}-1} = 2^{l-1}/n$$

and

$$var(S_{2^{l-1}-1}) \le C2^{l-1}/n^2,$$

$$\begin{aligned}P(S_{2^{l-1}-1}<2^{l-1}/(2n)) &= P(S_{2^{l-1}-1}-2^{l-1}/n < 2^{l-1}/(2n)-2^{l-1}/n)\\ &\le P(|S_{2^{l-1}-1}-2^{l-1}/n|) > 2^{l-1}/(2n))\\ &\le C(2^{l-1}/n^2)/(2^{l-1}/(2n))^2\\ &\le C/2^{l-1}\\ &\le C/n.\end{aligned}$$

Therefore,

$$\|f\|_2^2 P(A\cup A^cB) \le C/n.$$

To bound the next piece,

$$\begin{aligned}C\mathbb{E}N^{-2(\alpha\wedge1)}I(A^cB^c) &\le C\mathbb{E}N^{-2(\alpha\wedge1)}I(A^c)\\ &\le Cn^{-2(\alpha\wedge1)}\cdot P(A^c)\\ &\le Cn^{-2(\alpha\wedge1)}.\end{aligned}$$

For the third piece,

$$\begin{aligned}C\mathbb{E}\left((N+1)n^{-2}T^{-2}\right)^{\alpha\wedge1} I(A^cB^c) &\le C\mathbb{E}\left((N+1)n^{-2}(N/2n)^{-2}\right)^{\alpha\wedge1} I(A^cB^c)\\ &\le C\mathbb{E}\left((N+1)N^{-2}\right)^{\alpha\wedge1} I(A^c)\\ &\le C\mathbb{E}N^{-\alpha\wedge1}I(A^c)\\ &\le Cn^{-(\alpha\wedge1)}.\end{aligned}$$

And the final piece:

$$\begin{aligned}C\mathbb{E}\left(\frac{N-nT}{nT}\right)^{2(\alpha\wedge1)} I(A^cB^c) &\le C\mathbb{E}\left[\left(\frac{N-nT}{n}\right)^2\right]^{\alpha\wedge1} I(A^cB^c)\\ &\le Cn^{-2(\alpha\wedge1)}\mathbb{E}\left(N-nT\right)^{2(\alpha\wedge1)} I(A^c).\end{aligned}$$

To find $\mathbb{E}\,(N-nT)^2\,I(A^c)$, rewrite A^c:

$$\begin{aligned}\mathbb{E}\,(N-nT)^2\,I(A^c) &= \mathbb{E}\,(N-nT)^{2(\alpha\wedge 1)}\,I(N\in\{2^{l-1},2^l,2^{l+1}\}) \\ &= \mathbb{E}\,(N-nT)^{2(\alpha\wedge 1)}\,I(N=2^{l-1}) \\ &\quad +\mathbb{E}\,(N-nT)^{2(\alpha\wedge 1)}\,I(N=2^{l}) \\ &\quad +\mathbb{E}\,(N-nT)^{2(\alpha\wedge 1)}\,I(N=2^{l+1}).\end{aligned}$$

Each of these three expectations is bounded in a similar fashion. Here, only the first bound is found.

$$\begin{aligned}\mathbb{E}\,(N-nT)^{2(\alpha\wedge 1)}\,I(N=2^{l-1}) &= \mathbb{E}\left(2^{l-1}-nS_{2^{l-1}-1}\right)^{2(\alpha\wedge 1)}I(N=2^{l-1}) \\ &\le \left[\mathbb{E}\left(2^{l-1}-nS_{2^{l-1}-1}\right)^2\right]^{\alpha\wedge 1} \\ &= \left[var\,(nS_{2^{l-1}-1})\right]^{\alpha\wedge 1} \\ &= n^{2(\alpha\wedge 1)}\left[var\,(S_{2^{l-1}-1})\right]^{\alpha\wedge 1} \\ &\le Cn^{2(\alpha\wedge 1)}\left[2^{l-1}/n^2\right]^{\alpha\wedge 1} \\ &\le Cn^{\alpha\wedge 1}.\end{aligned}$$

So the last piece is bounded by

$$C\mathbb{E}\left(\frac{N-nT}{nT}\right)^{2(\alpha\wedge 1)} I(A^cB^c)\le Cn^{-(\alpha\wedge 1)}.$$

Putting these four pieces together,

$$\mathbb{E}\|f-\hat{f}\|_2^2\le Cn^{-(\alpha\wedge 1)}\le Cn^{-2\alpha/(2\alpha+1)}$$

for $\alpha \ge 1/2$. This bounds the estimate of the function using the first N data points scaled to $[0,1]$. Rescaling gives the same bound (to within a constant) for the interval $[0,S_{N-1}]$. Using the last N data points gives an estimate on $[S_{N^*-N},1]$ with the same error bound. Combining these two estimates on the entire interval $[0,1]$ results in the desired error, to within a constant.

6 Appendix

The eight test functions used in the simulations are displayed in Figure 4. The functions “doppler”, “heavisine”, “bumps”, and “blocks” are from Donoho and Johnstone (1994). The rest are from Marron et al. (1998). In the simulations, these formula were modified by a constant to give a standard deviation of 10.

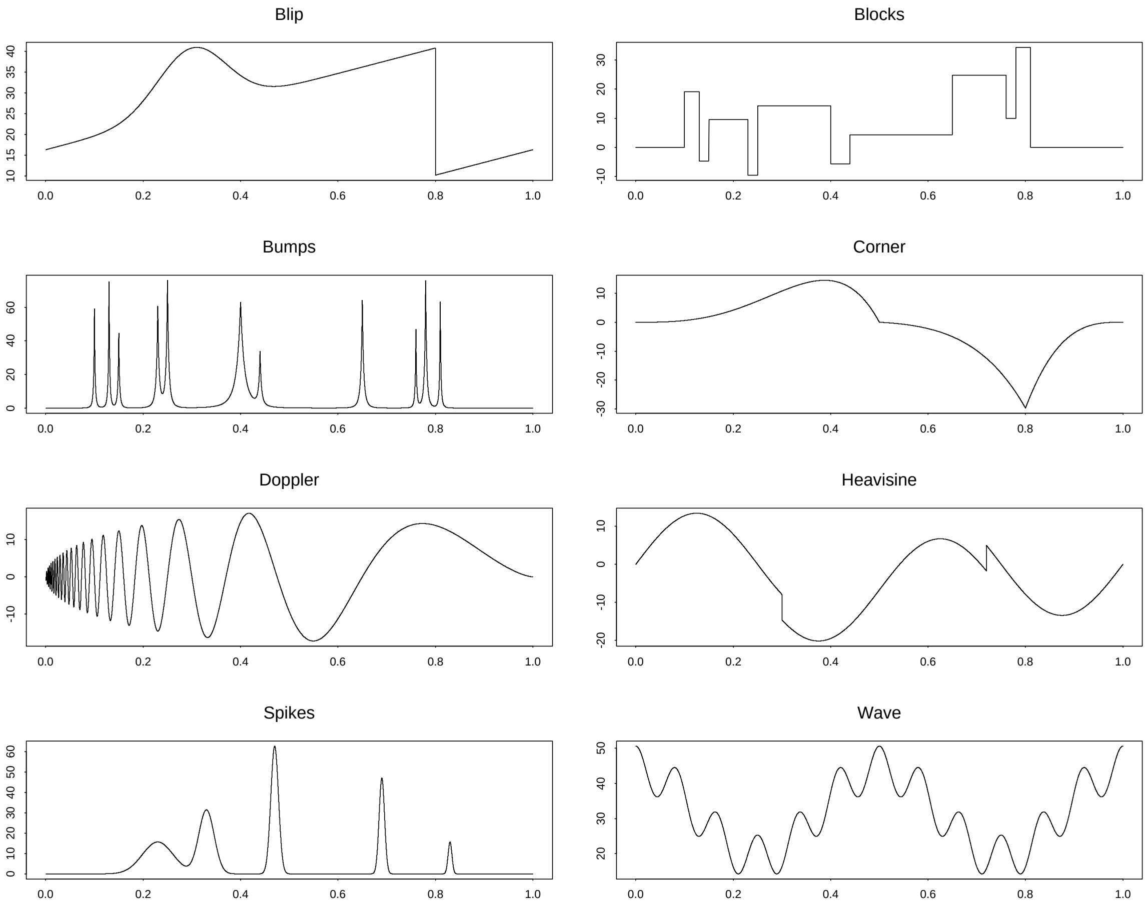

Figure 4: Test Functions

References

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

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

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

Chicken, E. (2003a). Asymptotic rates for coefficient-dependent and block-dependent thresholding in wavelet regression. *Technical Report M960, Department of Statistics, Florida State University* .

Chicken, E. (2003b). Block thresholding and wavelet estimation for nonequispaced samples. *J. Stat. Plan. Inference* **116** 113–129.

Daubechies, I. (1992). *Ten Lectures on Wavelets.* SIAM, Philadelphia.

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.

Marron, S., Adak, S., Johnstone, I., Neuman, M. and Patil, P. (1998). Exact risk analysis of wavelet regression. *J. Comput. Graph. Statist.* **7** 278–309.

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

Nason, G. (1998). *WaveThresh3 Software.* Department of Mathematics, University of Bristol, Bristol, UK.

Stein, C. (1981). Estimation of the mean of a multivariate normal distribution. *Ann. Stat.* **9** 1135–1151.