Sankhyā : The Indian Journal of Statistics
2011, Volume 73-A, Part 2, pp. 231-244

On the Estimation of the Function and Its Derivatives in Nonparametric Regression: A Bayesian Testimation Approach

Athanasia Petsa and Theofanis Sapatinas
University of Cyprus, Nicosia, Cyprus

Abstract

We consider the problem of estimating the unknown response function and its derivatives in the standard nonparametric regression model. Recently, Abramovich et al. (2010) applied a Bayesian testimation procedure in a wavelet context and proved asymptotical minimaxity of the resulting adaptive level-wise maximum a posteriori wavelet testimator of the unknown response function and its derivatives in the Gaussian white noise model. Using the boundary-modified coiflets of Johnstone and Silverman (2004), we show that dicretization of the data does not affect the order of magnitude of the accuracy of a discrete version of the suggested level-wise maximum a posteriori wavelet testimator, obtaining thus its adaptivity and asymptotical minimaxity in the standard nonparametric regression model that is usually considered in practical applications. Simulated examples are used to illustrate the performance of the developed wavelet testimation procedure and compared with three recently proposed empirical Bayes wavelet estimators and a block thresholding wavelet estimator.

AMS (2000) *subject classification.* Primary 62G05; Secondary 62G08.
Keywords and phrases. Adaptive estimation, Besov spaces, boundary wavelets, coiflets, Gaussian white noise model, multiple testing, nonparametric regression model, thresholding, wavelet analysis.

1 Introduction

Asymptotical minimaxity in nonparametric function estimation is usually considered under the Gaussian white noise model, driven by the following stochastic differential equation

$$dY_N(t) = f(t)dt + \frac{\sigma}{\sqrt{N}}\,dW(t), \quad t \in T = [0,1], \tag{1.1}$$

where $\sigma > 0$ is assumed to be known and finite, $f \in L^2(T)$ is the unknown response function and W is a standard Wiener process. According to Ibragimov and Khasminskii (1981, p. 5), the Gaussian white noise model (1.1) has been initially introduced as a statistical model by Kotel'nikov (1959). Since then, it has been extensively studied in the nonparametric literature and is considered as an idealized model that provides an approximation to many nonparametric models. We call it the *continuous data model*.

In particular, under some smoothness assumptions on f, the continuous data model (1.1) is asymptotically equivalent (in Le Cam sense) to the nonparametric regression model (see Brown and Low, 1996). In the standard nonparametric regression model, one observes Gaussian random variables governed by

$$Y_i = f(i/N) + \sigma\, \epsilon_i, \quad i = 1, 2, \ldots, N, \tag{1.2}$$

where $\sigma > 0$ is assumed to be known and finite, $f \in L^2(T)$ is the unknown response function and ϵ_i, $i = 1, 2, \ldots, N$, are independent $N(0, 1)$ random variables. We call it the *discrete data model*.

The statistical estimation problem is to estimate the unknown response function f based on observations from either the continuous data model or the discrete data model. In order to derive our minimax results, we shall assume that f belongs to a Besov ball $B^s_{p,q}(M)$ of a radius $M > 0$. The parameter s measures the degree of smoothness while p and q specifiy the type of norm used to measure the smoothness. Besov classes contain various traditional smoothness spaces such as Hölder and Sobolev spaces as special cases. However, they also include different types of spatially inhomogeneous functions (see, e.g., Meyer, 1992). As in Johnstone and Silverman (2005), we consider the cases where $q = \infty$, $1 \le p \le \infty$ and either (i) $s > 1/p$ or (ii) $s = p = 1$. The restriction (i) ensures that the corresponding Besov balls are embedded in balls of the space of continuous functions (ensuring also that the point evaluation functionals $f \mapsto f(t_0)$ are continuous, so that the discrete data model makes sense). As in Donoho and Johnstone (1998) and Johnstone and Silverman (2005), the restriction (ii) is included since a ball in the space of functions of bounded variation is contained in a ball of the Besov space $B^s_{1,\infty}(M)$ and contains a ball of the Besov space $B^s_{1,1}(M)$ (in this case, we can make sense of point evaluation functionals $f \mapsto f(t_0)$ by agreeing to use, say, the left continuous versions of f in the space of functions of bounded variation).

The fact that wavelet series constitute unconditional bases for Besov spaces has caused various wavelet-based estimation procedures to be widely used for estimating the unknown response f, assuming it lies in these spaces, in either the continuous data model or the discrete data model. The standard wavelet approach for the estimation of f is based on finding the empirical wavelet coefficients of the data and denoising them, usually by some type of a thresholding rule. Transforming them back to the function space then yields the resulting estimate. The main statistical challenge in such an approach is a proper choice of a thresholding rule. A series of various wavelet thresholds originated by different ideas has been proposed in the literature during the last decade, e.g., the universal threshold (see Donoho and Johnstone, 1994), Stein's unbiased risk estimation threshold (see Donoho and Johnstone, 1995), the false discovery rate threshold (see Abramovich and Benjamini, 1996), the cross-validation threshold (see Nason, 1996), the Bayes threshold (see Abramovich, Sapatinas and Silverman, 1998) and the empirical Bayes threshold (see Johnstone and Silverman, 2005).

Abramovich and Benjamini (1996) demonstrated that thresholding can be seen as a multiple hypothesis testing procedure, where one first simultaneously tests the wavelet coefficients of the unknown response function f, for significance. The coefficients concluded to be significant are then estimated by the corresponding empirical wavelet coefficients of the data, while the non-significant ones are discarded. Such a testimation procedure evidently mimics a hard thresholding rule. Here, we proceed along the lines of the testimation approach, where we utilize the recently developed maximum a posteriori Bayesian multiple testing procedure of Abramovich and Angelini (2006). Their hierarchical prior model is based on imposing a prior distribution on the number of false null hypotheses.

Recently, Abramovich, Grinshtein and Pensky (2007) applied the Bayesian 'testimation' approach of Abramovich and Angelini (2006) in order to recover an unknown high-dimensional Gaussian mean vector, and upper error bounds were obtained over various weak and strong l_p-balls, $0 \leq p < \infty$. Their results were extended to more general cases by Abramovich et al. (2010) who then applied the Bayesian testimation procedure in a wavelet context in order to show that a level-wise maximum a posteriori wavelet testimator of the response function f is adaptive and asymptotical minimax in the continuous data model. Their results were also extended to the estimation of the derivatives of f. A small simulation study was also conducted to illustrate the performance of the proposed wavelet testimator in practice, assuming that the discrete data model holds.

Because of its orthonormality properties, carrying out a wavelet decomposition of $dY_N(t)$ described by the continuous data model, we obtain independent observations $Y_{jk} \sim N(\theta_{jk}, 1/N)$. In practice, however, instrumentally acquired data that is digitally processed is typically discrete. Such settings can be represented by the discrete data model. In this case, the discrete wavelet transform of the sequence $N^{-1/2}Y_i$ yields independent observations $\tilde{Y}_{jk} \sim N(\tilde{\theta}_{jk}, 1/N)$. In much of the existing literature, however, the difference between Y_{jk} and $\tilde{Y}_{jk}$ is often ignored. Estimators are usually motivated, derived and analyzed in the continuous data model and then, in practical applications, are applied to the discrete data model. An interesting problem is to investigate the risk bounds of an estimator based on the observations $\tilde{Y}_{jk}$ from the discrete data model. Donoho and Johnstone (1999) used Deslauriers-Dubucs interpolation to pass from the discrete data model to the continuous data model. Jonhstone and Silverman (2004) used boundary-modified coiflets to show that the discrete wavelet transform of finite data from the discrete data model asymptotically provides a close approximation to the wavelet transform of the data from the continuous data model. These results were then used in Johnstone and Silverman (2005) to prove that discretization of the data does not affect the asymptotic convergence rates of the upper risk bounds of their proposed empirical Bayes estimators. As pointed out above, Abramovich et al. (2010) derived and theoretically studied their proposed adaptive level-wise maximum a posteriori wavelet testimator of the response function and its derivatives in the continuous data model and then they applied it in the discrete data model, revealing nice finite sample properties. However, a theoretical justification allowing one to do this is lacking from the theoretical findings of Abramovich et al. (2010).

Our aim is to fill this gap. In what follows, we use the Bayesian testimation methodology proposed by Abramovich, Grinshtein and Pensky (2007), and further extended by Abramovich et al. (2010), for estimating an unknown high-dimensional mean vector in a Gaussian sequence model. We refer to these papers for more details. In Section 2, using the boundary-modified coiflets of Johnstone and Silverman (2004), we show that discretization of the data does not affect the order of magnitude of the accuracy of a discrete version of the suggested level-wise maximum a posteriori wavelet testimator, obtaining thus its adaptivity and asymptotical minimaxity in the discrete data model that is usually considered in practical applications. In Section 3, we present a simulation study to illustrate the performance of the developed level-wise maximum a posteriori wavelet testimator, and compare it with three recently proposed empirical Bayes wavelet estimators and a block thresholding wavelet estimator. Finally, the proofs of the theoretical results are given in Appendix A.

2 Level-wise MAP Wavelet Testimator in the Discrete Data Model

In this section, we consider the boundary-corrected version of the level-wise maximum a posteriori (MAP) wavelet testimator, constructed under the continuous data model, in order to estimate the unknown response function f and its derivatives in the discrete data model, and prove its adaptivity and asymptotical minimaxity.

2.1. A brief overview of the MAP testimator. Consider the multiple hypothesis testing problem, where we wish to simultaneously test

$$H_{0i} : \mu_i = 0 \quad \text{versus} \quad H_{1i} : \mu_i \neq 0, \quad i = 1, 2, \ldots, n.$$

A configuration of true and false null hypotheses is uniquely defined by the indicator vector $x = (x_1, ..., x_n)'$, where $x_i = \mathbb{I}(\mu_i \neq 0)$ and $\mathbb{I}(A)$ denotes the indicator function of the set A. Let $\kappa = x_1 + ... + x_n = ||\mu||_0$ be the number of non-zero μ_i, i.e., $||\mu||_0 = \#\{i : \mu_i \neq 0\}$. Assume some prior distribution π_n on κ with $\pi_n(\kappa) > 0$, $\kappa = 0, \ldots, n$. For a given κ, all the corresponding different vectors x are assumed to be equally likely a priori. After some simple calculations, the proposed Bayesian multiple testing procedure leads to finding $\hat{\kappa}$ which maximizes the logarithm of the posterior distribution. The data corresponding to the $\hat{\kappa}$ largest absolute values survive and the rest are set to 0, yielding the MAP testimator. For a detailed review of the MAP methodology, we refer to Abramovich, Grinshtein and Pensky (2007) and Abramovich et al. (2010).

In what follows, using the boundary-modified coiflets of Johnstone and Silverman (2004), the MAP methodology is applied in a wavelet context to derive adaptive and asymptotically minimax wavelet estimators of the unknown response function f in the discrete data model.

2.2. Estimation of the unknown response function and its derivatives. Consider the discrete data model. Let R be the number of continuous derivatives of the scaling function ϕ. Suppose that the wavelets and scaling functions are modified by the boundary construction described in Section 5.3 of Johnstone and Silverman (2005). Assume that, for $N = 2^J$, we have sufficient observations to evaluate the preconditioned sequence $P_J Y$, where $Y = (Y_1, Y_2, \ldots, Y_N)$, and let $\tilde{Y}$ be the boundary corrected discrete wavelet transform of $N^{-1/2} P_J Y$, defined in Section 5.4 of Johnstone and Silverman (2005).

When $Y = (Y_1, Y_2, \ldots, Y_N)$ is a vector of independent Gaussian random variables with common variance σ^2, as in the discrede data model, then $\tilde{Y}$ has a multivariate Gaussian distribution whose elements have variances bounded by $c_A \sigma^2/N$, where c_A is a positive constant. Furthermore, the array of interior wavelet coefficients $\tilde{Y}^I$ is an array of independent Gaussian random variables with common variance σ^2/N. For furter details, see Sections 5.3 and 5.4 in Johnstone and Silverman (2005).

Take L such that $2^L \geq 6(S-1)$ for some appropriate $S > 0$ (see Sections 5.3 and 5.4 in Johnstone and Silverman, 2005). We also write θ^I for the wavelet coefficients θ_{jk} with $j \geq L$ and $k \in \mathcal{K}_j^I$ (i.e., the interior wavelet coefficients), θ^B for the wavelet coefficients θ_{jk} with $j \geq L$ and $k \in \mathcal{K}_j^B$ (i.e., the boundary wavelet coefficients). For further details on θ^I, $\mathcal{K}_j^I$, θ^B and $\mathcal{K}_j^B$, we refer to Sections 5.3 and 5.4 in Johnstone and Silverman (2005).

Requiring some conditions on the prior and using appropriate thresholds for estimating the boundary wavelet coefficient array θ^B, allows one to adaptively estimate the unknown response function f and its derivatives by the adaptive level-wise MAP wavelet testimator $\hat{f}_N$ and its corresponding derivatives, respectively, under the discrete data model at the optimal convergence rates. Such a plug-in estimation of $f^{(r)}$ by $\hat{f}_N^{(r)}$, where $r \geq 0$ is an integer, is, in fact, along the lines of the vaguelette-wavelet decomposition approach of Abramovich and Silverman (1998). (Here, and in what follows, $f^{(s)}$, where $s \geq 0$ is an integer, denotes the s-th derivative of f and, by convention, $f^{(0)} = f$.)

Define the estimated coefficient array $\hat{\theta}$ as follows:

(I) Estimate the coarse scaling coefficients by their observed values, i.e., set $\hat{\theta}_{L-1} = \tilde{Y}_{L-1}$.

(II) Estimate the interior wavelet coefficients θ^I by the corresponding $\hat{\theta}^I$. This is accomplished by applying the MAP procedure on $\tilde{Y}^I$ for each $L \leq j \leq J-1$ under the assumptions on the prior in Abramovich et al. (2010, Section 2.2, Collorary 1). (Note that $\beta = 0$ when we estimate f, that is when $r = 0$.)

(III) Threshold the boundary wavelet coefficient array θ^B separately to obtain $\hat{\theta}^B$. Specifically, at level j, use a hard threshold of $\tau_A \sqrt{j/N}$, where $\tau_A^2 \geq c_A(1 + \beta)\log 2$ for some $c_A > 0$ and for $\beta \geq 2r$, where $r \geq 0$ is an integer, so that for each $L \le j \le J-1$ and $k \in K_j^B$, the estimated boundary wavelet coefficients are given by

$$\hat{\theta}_{jk} = \tilde{Y}_{jk}\mathbb{I}\big(|\tilde{Y}_{jk}| > \tau_A\sqrt{j/N}\big).$$

(IV) For unobserved levels $j \ge J$, set $\hat{\theta}_{jk} = 0$.

In order to establish our theoretical properties, we follow Johnstone and Silverman (2005) and consider again functions f whose wavelet coefficient array θ falls in the sequence Besov ball $b^s_{p,\infty}(M)$, where $M > 0$, $s > \max(0, 1/p - 1/2)$ and $0 < p \le \infty$.

We note first that the proposed level-wise MAP wavelet testimator (described by the steps (I)-(IV) above) does not rely on the knowledge of the parameters s, p, and M of a specific sequence Besov ball $b^s_{p,\infty}(M)$ and it is, therefore, inherently adaptive.

Furthermore, as in Johnstone and Silverman (2005), we measure the risk of the proposed adaptive level-wise MAP wavelet testimator as an estimate of the wavelet expansion of the unknown respone function f itself by

$$R^*_{N,r}(f) = \mathbb{E}\|\hat{\theta}_{L-1} - \theta_{L-1}\|_2^2 + \sum_{j=L}^{\infty} 2^{2rj}\mathbb{E}\|\hat{\theta}_j - \theta_j\|_2^2, \quad r = 0, 1, 2, \dots .$$

THEOREM 2.1. *Assume that the scaling function ϕ and the mother wavelet ψ have R continuous derivatives and support $[-S+1, S]$ for some integer $S > R$, and that $\int x^m\phi(x)dx = \int x^m\psi(x)dx = 0$ for $m = 1, 2, \dots, R-1$, $R \ge 2$. Assume that the wavelets and scaling functions are modified by the boundary construction described in Section 5.3 of Johnstone and Silverman (2005). Consider the construction of the adaptive level-wise MAP wavelet testimator described by the steps (I)-(IV) above. Suppose that $0 < s < R$, $0 < p \le \infty$, $2r \le \beta$, where $r \ge 0$ is an integer and $r < \min\big(s, \big(s + 1/2 - 1/p\big)p/2\big)$, and either i) $s > 1/p$ or ii) $s = p = 1$. Let $\mathcal{F}(\mathcal{M})$ be the set of functions f whose wavelet coefficient array θ falls in the sequence Besov ball $b^s_{p,\infty}(M)$. Then, there is a constant C, independent of N, such that*

$$\sup_{f \in \mathcal{F}(\mathcal{C})} R^*_{N,r}(f) \le CN^{-2(s-r)/(2s+1)}.$$

REMARK 2.1. Using wavelets with bounded support and vanishing moments up to order $R-1$, $R \ge 2$, (e.g., as those considered in Theorem 2.1), provided that $1 \le p \le \infty$ and $\max(0, 1/p - 1/2) < s < R$, the sequence of Besov balls $b^s_{p,\infty}(M)$, $M > 0$, are equivalent to the corresponding Besov balls $B^s_{p,\infty}(M) = \{f |\ f \in L_p,\ ||f||_{B^s_{p,\infty}} \le M\}$ of the functions themselves (see, e.g., Appendix D in Johnstone, 2002). Also, according to Donoho et al. (1997) and Johnstone and Silverman (2005), as $N \to \infty$, the asymptotical adaptive minimax convergence rate for the L^2-risk of estimating the rth derivative of the unknown response function f in the discrete data model over Besov balls $B^s_{p,q}(M)$, where $M > 0$, $0 < s < R$,

$0 \le r < \min\{s, (s+1/2-1/p)p/2\}$, $1 \le p, q \le \infty$ and $s > 1/p$, is given by

$$\inf_{\tilde{f}_N} \sup_{f \in B^s_{p,q}(M)} \mathbb{E}||\tilde{f}^{(r)}_N - f^{(r)}||^2_2 \asymp N^{-2(s-r)/(2s+1)},$$

where the infimum is taken over all possible estimators $\tilde{f}_N$ (i.e., measurable functions) from the discrete data model. (Here, and in what follows, $g_1(N) \asymp g_2(N)$ denotes $0 < \liminf\{g_1(N)/g_2(N)\} \le \limsup\{g_1(N)/g_2(N)\} < \infty$ as $N \to \infty$.)

Note that, in view of Proposition 1 in Johnstone and Silverman (2005) showing that, for any integer r such that $0 \le r \le s$, and any function f with a boundary-corrected wavelet expansion (see (32) in Johnstone and Silverman, 2005), one has

$$\int_0^1 |f^{(r)}(t)|^2 dt \le \|\theta_{L-1}\|^2_2 + \sum_{j=L}^{\infty} 2^{2rj}\|\theta_j\|^2_2, \quad r = 0, 1, 2, \dots .$$

Consider now the estimator

$$\hat{f}_N(t) = \sum_{k \in \mathcal{K}_{L-1}} \hat{\theta}_{L-1,k}\phi_{Lk}(t) + \sum_{j=L}^{J-1}\sum_{k=0}^{2^j-1} \hat{\theta}_{jk}\psi_{jk}(t), \quad t \in [0,1], \tag{2.1}$$

where the estimated wavelet coefficient array $\hat{\theta}$ is obtained by the steps (I)-(IV) above. (For the definition of the set of indices $\mathcal{K}_{L-1}$, associated with the boundary-modified coiflets, we refer to Section 5.3 in Johnstone and Silverman, 2005.) Then, Theorem 2.1 also establishes, as a by-product, the asymptotical minimaxity for the L^2-risk of the rth derivative of the proposed adaptive level-wise MAP wavelet testimator $\hat{f}^{(r)}_N$ over Besov balls $B^s_{p,\infty}(M)$, where $M > 0$, $0 < s < R$, $0 \le r < \min\{s, (s+1/2-1/p)p/2\}$, $1 \le p \le \infty$ and $s > 1/p$.

Remark 2.2. According to Donoho and Johnstone (1998), as $N \to \infty$, the asymptotical adaptive minimax convergence rate for the L^2-risk of estimating the unknown response function f in the discrete data model over Besov balls $B^s_{p,q}(M)$, where $M > 0$, $0 < s < R$, $1 \le p, q \le \infty$ and either $s > 1/p$ or $s = p = q = 1$, is given by

$$\inf_{\tilde{f}_N} \sup_{f \in B^s_{p,q}(M)} \mathbb{E}||\tilde{f}_N - f||^2_2 \asymp N^{-2s/(2s+1)},$$

where the infimum is taken over all possible estimators $\tilde{f}_N$ (i.e., measurable functions) from the discrete data model.

Consider now the wavelets mentioned in Remark 2.1 and the estimator $\hat{f}_N$ defined in (2.1). Therefore, in view of the above, on noting that $R^*_{N,0}(f) = \mathbb{E}||\hat{f}_N - f||^2_2$ (due to Parseval's equality), Theorem 2.1 also establishes, as a by-product, the asymptotical minimaxity for the L^2-risk of the proposed adaptive level-wise MAP wavelet testimator $\hat{f}_N$ over Besov balls $B^s_{p,\infty}(M)$, where $M > 0$, $0 < s < R$, $1 \le p \le \infty$ and either $s > 1/p$ or $s = p = 1$.

Remark 2.3. It is easy to see that the assumptions on π_j in Step 2 above for $r = 0$ and $\beta = 0$ are satisfied by, e.g., the truncated geometric prior TrGeom$(1-q_j)$, with probability of success $p_j = 1 - q_j$, given by

$$\pi_j(\kappa) = \frac{(1-q_j)q_j^{\kappa}}{\left(1-q_j^{2^j-2S+3}\right)}, \quad \kappa = 0, 1, \ldots, 2^j - 2S + 2, \quad q_j \sim e^{-c(\gamma_j)}.$$

3 Simulation Study

In this section, we present a simulation study to illustrate the performance of the developed level-wise MAP wavelet testimator and compare it with three empirical Bayes wavelet estimation procedures and one block thresholding wavelet estimation method, namely, the posterior mean (PostMean) and posterior median (PostMed) wavelet estimators proposed in Johnstone and Silverman (2005), the Bayes Factor (BF) wavelet estimator proposed in Pensky and Sapatinas (2007) and the Neigh-Block (Block) wavelet thresholding estimator proposed in Cai and Silverman (2001). We note that all estimators are adaptive to the unknown smoothness and attain the optimal convergence rate, except for the Block estimator that is near optimal (up to a logarithmic factor). For PostMean, PostMed and BF wavelet estimators we used the Double-exponential prior, where the corresponding prior parameters were estimated level-by-level by marginal likelihood maximization, as described in Johnstone and Silverman (2005). The prior parameters for the level-wise MAP wavelet testimator were estimated by conditional likelihood maximization described in Section 4.2 of Abramovich et al. (2010). For the Block wavelet estimator, the lengths of the overlapping and non-overlapping blocks and the value of the thresholding coefficient, associated with the method, were selected as suggested by Cai and Silverman (2001). Finally, for all competing methods, σ was estimated by the median of the absolute value of the empirical wavelet coefficients at the finest resolution level divided by 0.6745.

In the simulation study, we evaluated the above five estimators for a series of test functions. We present the results for the *Corner, Sharp Peak, Blip* and *Spikes* test functions defined on $[0,1]$. The formulae for *Corner*, *Blip* and *Spikes* can be found in Antoniadis, Bigot and Sapatinas (2001), while the formula for *Sharp Peak* is given by

$$f(t) = \begin{cases} 2t, & \text{if} \quad 0 \le t \le 0.5, \\ 2-2t, & \text{if} \quad 0.5 \le t \le 1. \end{cases}$$

Note that a thorough and detailed simulation study was presented in Abramovich et al. (2010, Section 4.3), where five different estimators were compared with MAP wavelet testimator for the *Bumps*, *Blocks*, *Heavisine*, *Doppler*, *Peak* and *Wave* test functions.

For each test function, $M = 100$ samples were generated by adding independent Gaussian noise $\varepsilon \sim N(0, \sigma^2)$ to $n = 1024$ equally spaced points on [0,1]. The value

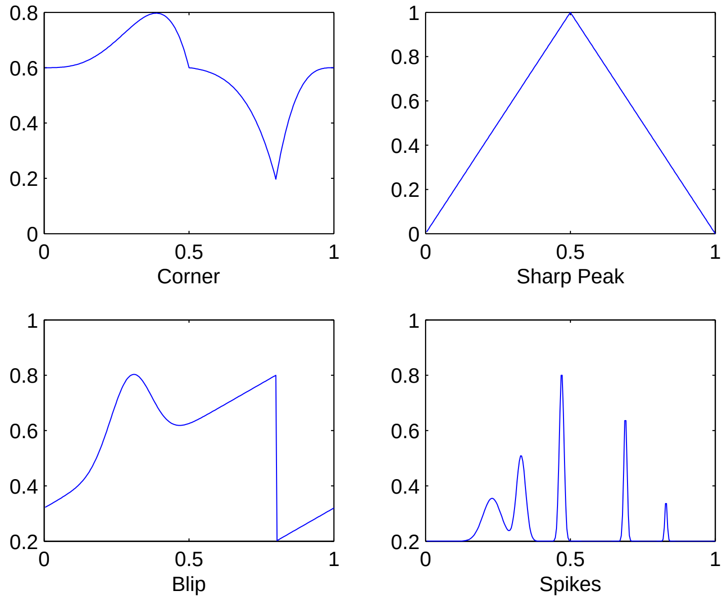

Figure 1: Corner, Sharp Peak, Blip and Spikes test functions, sampled at $n = 1024$ points.

of the (root) signal-to-noise ratio (SNR) was taken to be 3 (high noise level), 5 (moderate noise level) and 7 (low noise level), where

$$SNR(f, \sigma) = \sigma^{-1} \left(\frac{1}{n} \sum_{i=1}^{n} (f(t_i) - \bar{f})^2 \right)^{1/2} \quad and \quad \bar{f} = \frac{1}{n} \sum_{i=1}^{n} f(t_i).$$

The goodness-of-fit for an estimator $\hat{f}$ of f in a single replication was measured by its mean squared error (MSE), defined as

$$MSE(f, \hat{f}) = \frac{1}{n} \sum_{i=1}^{n} (\hat{f}(t_i) - f(t_i))^2.$$

We report the results for $n = 1024$ using the compactly supported mother wavelet *Coiflet 3* (see Daubechies, 1992, p. 258) and the primary resolution level $j_0 = 4$ (different choices of wavelet functions and resolution levels yielded basically similar results in magnitude). The sample distributions of MSE over replications for all estimators in simulation studies were typically highly asymmetrical and affected by outliers. Therefore, we preferred the sampled medians of MSEs rather than

Table 1: Relative median MSE for the Corner, Sharp Peak, Blip and Spikes test functions, sampled at $n = 1024$ data points and using three values of SNR (3, 5 and 7), for the various wavelet estimators. The minimal relative median MSE for each estimator is bold.

n	signal	SNR	MAP	BF	Postmed	Postmean	Block
1024	Corner	3	0.8155	0.2931	1	0.8382	**0.7427**
		5	0.8091	0.378	1	0.9367	0.8008
		7	0.8697	0.3924	1	0.9541	0.7436
1024	Sharp Peak	3	**0.7542**	**0.121**	**0.7296**	**0.556**	1
		5	0.8802	0.1215	0.814	0.6328	1
		7	0.787	0.182	0.8345	0.6761	1
1024	Blip	3	0.9149	0.7199	0.9974	1	0.874
		5	0.9218	0.8328	1	0.9778	0.865
		7	0.9449	0.8611	0.9778	1	0.9793
1024	Spikes	3	0.8761	0.6093	9937	1	0.9396
		5	0.8394	0.5466	0.953	0.9775	1
		7	0.8187	0.5865	0.9501	1	0.9596

means to gauge the estimators' goodness-of-fit. For each estimator, test function and noise level, we calculated the median MSE over all 100 replications. To quantify the comparison between estimators over various test functions and noise levels, for each considered model we found the best estimator among the five ones, i.e., the one achieving the minimum median MSE, and evaluated the *relative* median MSE of the i-th estimator defined as $\min_{1 \leq j \leq 5}\{\text{Median}(\text{MSE}_j)\}/\text{Median}(\text{MSE}_i)$, $i = 1, 2, \ldots, 5$ (see Table 1).

Evidence from Table 1 (see also Table 1 in Abramovich et al., 2010) indicates that there is no 'uniformly best' estimator. The relative performance of each estimator depends on a specific test function and the noise level. However, the MAP testimator results in the highest *minimal* relative median MSE over all cases among the considered five estimators (see the bold numbers in Table 1). The minimal relative median MSE of an estimator reflects its inefficiency at the most challenging combination of a test function and SNR level and is a natural measure of its robustness. Additionally, we compared the competing estimators in terms of sparsity, measured by the total number of non-zero wavelet coefficients (averaged over 100 replications) surviving after thresholding. These results are given in Table 2 below. The proposed method is sparser than the empirical Bayes estimators (note that PostMean is not included in this comparison since is a non-linear shrinkage, hence all wavelet coefficients survive). The sparsity of the Neighblock thresholding estimator depends on the signal.

Table 2: Sparsity, averaged over 100 simulations, of the various wavelet methods for the Corner, Sharp Peak, Blip and Spikes functions, sampled at $n = 1024$ data points and using three values of SNR (3, 5 and 7).

n	signal	SNR	MAP	BF	Postmed	Block
1024	Corner	3	65.57	143.21	93.46	18.58
		5	57.72	154.54	90.62	22.7
		7	55.24	144.8	86.57	27.95
1024	Sharp Peak	3	63.42	184.95	90.85	16.87
		5	59.95	178.56	96.59	16.64
		7	73.71	172.75	97.24	16.93
1024	Blip	3	103.29	114.55	109.67	170.4
		5	114.12	129.02	127.49	216.41
		7	129.9	141.96	141.13	240.17
1024	Spikes	3	86.71	158.73	107.66.13	72.14
		5	96.6	171.88	112.26	78.17
		7	117.7	170.56	119.12	82.9

Acknowledgement. The authors would like to thank the Editor, the Associate Editor and the anonymous referee for helpful comments on improvements to this paper.

References

Abramovich, F. and Angelini, C. (2006). Bayesian maximum a posteriori multiple testing procedure. *Sankhyā*, **68**, 436–460.

Abramovich, F. and Benjamini, Y. (1996). Adaptive thresholding of wavelet coefficients. *Comput. Statist. Data Anal.*, **22**, 351–361.

Abramovich, F. and Silverman, B.W. (1998). Wavelet decomposition approaches to statistical inverse problems. *Biometrika*, **85**, 115–129.

Abramovich, F., Sapatinas, T. and Silverman, B.W. (1998). Wavelet thresholding via a Bayesian approach. *J. Roy. Stat. Soc. Ser. B*, **60**, 725–749.

Abramovich, F., Grinshtein, V. and Pensky, M. (2007). On optimality of Bayesian testimation in the normal means problem. *Ann. Statist.*, **35**, 2261–2286.

Abramovich, F., Grinshtein, V., Petsa, A. and Sapatinas, T. (2010). On Bayesian testimation and its application to wavelet thresholding. *Biometrika*, **97**, 181–198.

Antoniadis, A., Bigot, J. and Sapatinas, T. (2001). Wavelet estimators in nonparametric regression: a comparative simulation study. *J. Statist. Software*, **6**(6), 1–83.

Brown, L.D. and Low, M.G. (1996). Asymptotic equivalence of nonparametric regression and white noise. *Ann. Statist.*, **24**, 2384–2398.

Cai, T.T. and Silverman, B.W. (2001). Incorporating information on neighboring coefficients into wavelet estimation. *Sankhyā Ser. B*, **63**, 127–148.

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

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

Donoho, D.L. and Johnstone, I.M. (1995). Adapting to unknown smoothness via wavelet shrinkage. *J. Amer. Statist. Assoc.*, **90**, 1200–1224.

Donoho, D.L. and Johnstone, I.M. (1998). Minimax estimation via wavelet shrinkage. *Ann. Statist.*, **26**, 879–921.

Donoho, D.L. and Johnstone, I.M. (1999). Asymptotic minimaxity of wavelet estimators with sampled data. *Statist. Sinica*, **9**, 1–32.

Donoho, D.L., Johnstone, I.M., Kerkyacharian, G. and Picard, D. (1997). Universal near minimaxity of wavelet shrinkage. In *Festschrift for Lucien Le Cam* (D. Pollard, E. Torgersen and G.L. Yang, eds), pp. 183–218. Springer-Verlag, Berlin.

Folland, G.B. (1999). *Real Analysis.* 2nd Edition. John Wiley and Sons, New York.

Ibragimov, I.A. and Khasminskii, R.Z. (1981). *Statistical Estimation: Asymptotic Theory.* Springer-Verlag, Berlin.

Johnstone, I.M. (2002). *Function Estimation in Gaussian Noise: Sequence Models.* Unpublished Monograph. Available at `http://www-stat.stanford.edu/~imj/`.

Johnstone, I.M. and Silverman, B.W. (2004). Boundary coiflets for wavelet shrinkage in function estimation. *J. Appl. Probab.*, **41A**, 81–98.

Johnstone, I.M. and Silverman, B.W. (2005). Empirical Bayes selection of wavelet thresholds. *Ann. Statist.*, **33**, 1700–1752.

Kotel'nikov, V. (1959). *The Theory of Optimum Noise Immunity.* McGraw Hill, New York.

Meyer, Y. (1992). *Wavelets and Operators.* Cambridge University Press, Cambridge.

Nason, G.P. (1996). Wavelet shrinkage using cross-validation. *J. Roy. Stat. Soc. Ser. B*, **58**, 463–479.

Pensky, M. and Sapatinas, T. (2007). Frequentist optimality of Bayes factor estimators in wavelet regression models. *Statist. Sinica*, **17**, 599–633.

A Appendix

Recall that the symbol C is used for a generic positive constant, independent of N, which may take different values at different places. Also, recall that we write θ^I for the wavelet coefficients θ_{jk} with $j \geq L$ and $k \in \mathcal{K}_j^I$ (i.e., the interior wavelet coefficients), θ^B for the wavelet coefficients θ_{jk} with $j \geq L$ and $k \in \mathcal{K}_j^B$ (i.e., the boundary wavelet coefficients) and set $\theta_j^I = \{\theta_{jk}, k \in \mathcal{K}_j^I\}$ and $\theta_j^B = \{\theta_{jk}, k \in \mathcal{K}_j^B\}$ for each $j \geq L$.

A.1. Proof of Theorem 2.1. Following the steps in the proof of Theorem 3.3 in Abramovich, Grinshtein, Petsa and Sapatinas (2010), it is easy to show that

$$\sum_{j=L}^{J-1} 2^{2rj}\mathbb{E}(\|\hat{\theta}_j^I - \tilde{\theta}_j^I\|_2^2) + \sum_{j=J}^{\infty} 2^{2rj}\|\theta_j\|_2^2 \leq CN^{-2(s-r)/(2s+1)}. \tag{A.1}$$

Using Proposition 2 of Johnstone and Silverman (2005), we get

$$\sum_{j=L}^{J-1} 2^{2rj}\mathbb{E}\|\hat{\theta}_j^B - \tilde{\theta}_j^B\|_2^2 \leq CN^{-2(s-r)/(2s+1)}. \tag{A.2}$$

Discretization bias: The risks in (A.1) and (A.2) all quantify errors around the vector of discretized coefficients $\tilde{\theta}$. To control the difference in risk norm between $\tilde{\theta}$ and the vector of true coefficients θ, define $\Delta = \max(0, 1/2 - 1/p)$ and $r^{''} = s - \Delta$. Obviously, for $2 \leq p \leq \infty$, $r^{''} = s'$, while for $0 < p < 2$, $r^{''} = s$. Using the well-known norm inequality in infinite-dimensional spaces (see, e.g., Propositions 6.11 and 6.12 in Folland (1999)),

$$\|x\|_q \leq \|x\|_p \leq \|x\|_q n^{1/p-1/q},$$

for any $0 < p < q \leq \infty$ and $x = (x_1, x_2, \ldots, x_n)$, we immediately have

$$\|\tilde{\theta}_j - \theta_j\|_2 \leq 2^{j\Delta}\|\tilde{\theta}_j - \theta_j\|_p, \quad L-1 \leq j \leq J-1.$$

Hence, using Proposition 5 in Johnstone and Silverman (2004), we get

$$\begin{aligned}
\sum_{j=L-1}^{J-1} 2^{2rj}\|\tilde{\theta}_j - \theta_j\|_2^2 &\leq \sum_{j=L-1}^{J-1} 2^{2(r+\Delta)j}2^{2\Delta j}\|\tilde{\theta}_j - \theta_j\|_p^2 \\
&\leq \sum_{j=L-1}^{J-1} CM^2 2^{2\Delta j}2^{-2\tilde{a}(J-j)}2^{-2js'} \\
&= CM^2 2^{-2\tilde{a}J}\sum_{j=L-1}^{J-1} 2^{2j(\tilde{\alpha}-r^{''})},
\end{aligned}$$

where $\tilde{\alpha} = s - \max(0, \frac{1}{p} - 1)$. If $\tilde{a} = r^{''}$, then

$$\sum_{j=L-1}^{J-1} 2^{2rj}\|\tilde{\theta}_j - \theta_j\|_2^2 \leq CM^2 2^{-2\tilde{a}J}J.$$

On the other hand, if $\tilde{a} < r^{''}$, then

$$\begin{aligned}
\sum_{j=L-1}^{J-1} 2^{2rj}\|\tilde{\theta}_j - \theta_j\|_2^2 &\leq CM^2 2^{-2\tilde{a}J}\sum_{j=L-1}^{J-1} 2^{2j(\tilde{a}-r^{''})} \\
&= CM^2 2^{-2\tilde{a}J}2^{(L-1)(\tilde{a}-r^{''})}\left[\frac{1-2^{2(\tilde{a}-r^{''})(J-L-1)}}{1-2^{2(\tilde{a}-r^{''})}}\right] \\
&\leq CM^2 2^{-2\tilde{a}J}.
\end{aligned}$$

Therefore, combining the above, we arrive at the following bound

$$\sum_{j=L-1}^{J-1} 2^{2rj}\|\tilde{\theta}_j - \theta_j\|_2^2 \leq CM^2 J^{\lambda^{'}} 2^{-2\tilde{a}J}, \quad \tilde{a} \leq r^{''},$$

with $\lambda^{'} = 1$ if and only if $\tilde{a} = r^{''}$ and $\lambda^{'} = 0$ otherwise. Consider now the case $\tilde{a} > r^{''}$. Then,

$$\sum_{j=L-1}^{J-1} 2^{2rj}\|\tilde{\theta}_j - \theta_j\|_2^2 = CM^2 2^{-2\tilde{a}J}\sum_{j=L-1}^{J-1} 2^{2j(\tilde{a}-r^{''})}$$

$$
\begin{aligned}
&= CM^2 2^{-2\tilde{a}J} 2^{2(\tilde{a}-r^{''})(L-1)} [2^{2(\tilde{a}-r^{''})(J-L-1)} - 1] \\
&\leq CM^2 2^{-2\tilde{a}J} 2^{2(\tilde{a}-r^{''})(J-2)} \leq CM^2 2^{-2Jr^{''}} \\
&= C\frac{M^2}{N^{2r^{''}}}, \quad \tilde{a} > r^{''}.
\end{aligned}
$$

Letting $r^{'''} = \min\{r^{''}, \tilde{a}\}$, and combining all cases above, we arrive at the the bound

$$
\sum_{j=L-1}^{J-1} 2^{2rj} \|\tilde{\theta}_j - \theta_j\|_2^2 \leq CM^2 N^{-2r^{'''}} (\log N)^{\lambda^{'}}, \tag{A.3}
$$

with $\lambda^{'} = 1$ if and only if $\tilde{a} = r^{''}$ and $\lambda^{'} = 0$ otherwise.

Combining (4), (5) and (6) we arrive at

$$
R^*_{N,r}(f) \leq CN^{-2(s-r)/(2s+1)},
$$

completing thus the proof of Theorem 2.1.

Athanasia Petsa and Theofanis Sapatinas
Department of Mathematics and Statistics
University of Cyprus
P.O. Box 20537
CY 1678 Nicosia
Cyprus
E-mail: petsa.athanasia@ucy.ac.cy
fanis@ucy.ac.cy

Paper received December 2009; revised October 2010.