AISM (2007) 59: 425–434
DOI 10.1007/s10463-006-0071-7

Felix Abramovich · Claudia Angelini
Daniela De Canditiis

Pointwise optimality of Bayesian wavelet estimators

Received: 1 April 2005 / Revised: 9 September 2005 /
Published online: 19 July 2006

Abstract We consider pointwise mean squared errors of several known Bayesian wavelet estimators, namely, posterior mean, posterior median and Bayes Factor, where the prior imposed on wavelet coefficients is a mixture of an atom of probability zero and a Gaussian density. We show that for the properly chosen hyperparameters of the prior, all the three estimators are (up to a log-factor) asymptotically minimax within any prescribed Besov ball $B^s_{p,q}(M)$. We discuss the Bayesian paradox and compare the results for the pointwise squared risk with those for the global mean squared error.

Keywords Bayes Factor · Bayes model · Bayesian paradox · Besov spaces · Minimax rates · Nonparametric regression · Point estimation · Posterior mean · Posterior median · Wavelets

1 Introduction

Consider the standard "signal + white noise" model:

$$\mathrm{d}Y(t) = f(t)\mathrm{d}t + \sigma_n \mathrm{d}W(t), \quad t \in [0, 1] \tag{1}$$

where $\sigma_n = \sigma_0 n^{-1/2}$, W is a standard Wiener process and an unknown f belongs to a Besov ball $B^s_{p,q}(M)$ of radius M on [0, 1]. Besov classes include, in particular, Hölder ($p = q = \infty$) and Sobolev ($p = q = 2$) classes of smooth functions,

F. Abramovich (✉)
Department of Statistics and Operations Research, Tel Aviv University, Tel Aviv 69978, Israel
E-mail: felix@post.tau.ac.il

C. Angelini
Istituto per le Applicazioni del Calcolo, CNR Via P. Castellino 111, 80131 Naples, Italy

D. De Canditiis
Istituto per le Applicazioni del Calcolo, CNR V.le del Policlinico, 137-00161 Rome, Italy

as well as various classes of spatial inhomogeneous functions like the class of functions of bounded variation sandwiched between $B^1_{1,1}$ and $B^1_{1,\infty}$ (see Meyer, 1992 for more details). Different wavelet-based estimators have been intensively studied in the literature in the last decade (see Antoniadis 1997; Vidakovic, 1999; Abramovich et al., 2000 for comprehensive reviews). Among them there has been proposed a whole series of various Bayesian wavelet estimators (e.g., Chipman et al., 1997; Abramovich et al., 1998; Clyde et al., 1998; Vidakovic 1998), see also Müller and Vidakovic (1999) for an overview.

Numerous simulation studies demonstrated good performances of Bayesian estimators (e.g., Abramovich et al., 1998; Antoniadis et al., 2001). However, until recently their frequentist properties (e.g., minimaxity) have not been studied. Abramovich et al. (2004) investigated frequentist performance of several Bayesian estimators in terms of the global integrated mean squared error loss (MISE). Pensky (2006) explored the minimaxity of the posterior mean estimator with respect to the MISE under more general families of noise distributions and priors. Johnstone and Silverman (2005) considered the frequentist optimality of *empirical* Bayes wavelet estimators.

In this paper we explore the minimaxity of Bayesian wavelet estimators under the *pointwise* risk. We derive the pointwise convergence rates within a range of Besov spaces for the Bayesian estimators considered in Abramovich et al. (2004) and show that for the properly chosen hyperparameters of the prior they achieve the minimax rate up to a log-factor. The optimal choices of the hyperparameters for the global and pointwise risks are, however, generally different.

The paper is organized as follows. In Sect. 2 we present the prior model on wavelet coefficients and several Bayesian wavelet estimators corresponding to different Bayesian rules. Formulation and discussion of the main results on their pointwise convergence rates and asymptotic minimaxity over Besov classes are given in Sect. 3. In Sect. 4 we discuss the well-known Bayesian paradox when a prior yielding an optimal Bayesian estimator over a certain class of functions lies outside this class. In particular, we answer (positively!) the open question raised by Li and Zhao (2002) on the existence of a prior on Sobolev (or more general Besov) spaces whose Bayes procedures attain the optimal pointwise rates. The proofs are left to the Appendix.

2 The model

2.1 The prior

For simplicity of exposition we assume that f is periodic and work with periodic orthonormal wavelet bases on [0, 1] generated by a compactly supported scaling function φ and a corresponding mother wavelet ψ (e.g., Daubechies, 1992, Sect. 9.3). Then, f can be expanded in wavelet series as

$$f(t) = w_{-10}\varphi(t) + \sum_{j=0}^{\infty}\sum_{k=0}^{2^j-1} w_{jk}\psi_{jk}(t), \tag{2}$$

where $w_{-10} = \int_0^1 f(t)\varphi(t)dt$ and $w_{jk} = \int_0^1 f(t)\psi_{jk}(t)\mathrm{d}t$.

Consider the following prior on w_{jk} (Abramovich et al., 1998):

$$w_{jk} \sim \pi_j N(0, \tau_j^2) + (1 - \pi_j)\delta(0), \quad j \geq 0; \quad k = 0, \ldots, 2^j - 1, \tag{3}$$

where

$$\tau_j^2 = c_1 2^{-\alpha j} \quad \text{and} \quad \pi_j = \min(1, c_2 2^{-\beta j}), \quad j \geq 0, \tag{4}$$

α and β are non-negative constants, $c_1,\ c_2 > 0$.

Some intuitive understanding of the model implied by (3), (4) can be found in Abramovich et al. (1998) where they also established a relationship between the choice of α and β and the parameters of Besov spaces within which realizations from the prior will fall with probability one (see Sect. 4 below). Similar priors but with different forms for the hyperparameters are considered in Clyde et al. (1998).

2.2 Bayesian wavelet estimators

Let $Y_{jk} = \int_0^1 \psi_{jk}(t) dY(t)$. Subject to the prior (3)–(4), the posterior distribution of $w_{jk}|Y_{jk}$ is also a mixture of a corresponding posterior normal distribution and $\delta(0)$ with the posterior cumulative distribution function

$$F(w_{jk} \mid Y_{jk}) = \frac{1}{1+\eta_{jk}} \Phi\left(\frac{w_{jk} - Y_{jk}\tau_j^2/(\sigma_n^2 + \tau_j^2)}{\sigma_n \tau_j / \sqrt{\sigma_n^2 + \tau_j^2}}\right) + \frac{\eta_{jk}}{1+\eta_{jk}} \mathrm{I}(Y_{jk} \geq 0), \tag{5}$$

where Φ is the standard normal cumulative distribution function and the posterior odds ratio for the component at zero

$$\eta_{jk} = \frac{1 - \pi_j}{\pi_j} \frac{\sqrt{\tau_j^2 + \sigma_n^2}}{\sigma_n} \exp\left(-\frac{\tau_j^2 {Y_{jk}}^2}{2\sigma_n^2(\tau_j^2 + \sigma_n^2)}\right). \tag{6}$$

To derive a Bayesian rule one should specify the loss function. Different losses lead to different Bayesian estimators. The traditional L^2-loss yields the posterior mean (e.g., Chipman et al., 1997; Clyde et al., 1998; Vidakovic, 1998). From (5) and (6) one has

$$\widehat{w}_{jk} = E(w_{jk} \mid Y_{jk}) = \frac{1}{1+\eta_{jk}} \frac{\tau_j^2}{\tau_j^2 + \sigma_n^2} Y_{jk} \tag{7}$$

that mimics a nonlinear smoothing shrinkage.

Abramovich et al. (1998) proposed the posterior median estimator that corresponds to the L^1-loss and can be obtained in the following closed form

$$\tilde{w}_{jk} = \mathrm{Med}(w_{jk} \mid Y_{jk}) = \mathrm{sign}(Y_{jk}) \quad \max(0, \zeta_{jk}), \tag{8}$$

where

$$\zeta_{jk} = \frac{\tau_j^2}{\sigma_n^2 + \tau_j^2}|Y_{jk}| - \frac{\tau_j \sigma_n}{\sqrt{\sigma_n^2 + \tau_j^2}} \Phi^{-1}\left(\frac{1 + \min(\eta_{jk}, 1)}{2}\right). \tag{9}$$

The quantity ζ_{jk} is negative for all Y_{jk} in some implicitly defined interval $[-\lambda_j^{PM}, \lambda_j^{PM}]$, and hence the posterior median is a 'shrink' or 'kill' thresholding rule with the level-dependent thresholds λ_j^{PM}.

The Bayes Factor estimator of Vidakovic (1998) is based on hypothesis testing ideas and corresponds to the 0/1-loss : given Y_{jk}, test the hypothesis $H_0 : w_{jk} = 0$ against a two-sided alternative $H_1 : w_{jk} \neq 0$. If the hypothesis H_0 is rejected, w_{jk} is estimated by Y_{jk}, otherwise $w_{jk} = 0$, that is

$$\check{w}_{jk} = Y_{jk} I(\eta_{jk} < 1), \tag{10}$$

where the posterior odds ratio $\eta_{jk} = P(H_0 \mid Y_{jk})/P(H_1 \mid Y_{jk})$ is given by (6). The Bayes Factor rule (10) mimics the level-dependent 'keep' or 'kill' (hard) thresholding rule:

$$\check{w}_{jk} = Y_{jk} I\left(|Y_{jk}| \geq \lambda_j^{BF}\right),$$

where

$$(\lambda_j^{BF})^2 = \frac{2\sigma_n^2(\sigma_n^2 + \tau_j^2)}{\tau_j^2} \log\left(\frac{1-\pi_j}{\pi_j}\frac{\sqrt{\sigma_n^2 + \tau_j^2}}{\sigma_n}\right). \tag{11}$$

Abramovich et al. (2004, Lemma 1) showed that the sequences of level-dependent thresholds λ_j^{PM} and λ_j^{BF} are asymptotically 'similar' ($\lambda_j^{\mathrm{PM}} \sim \lambda_j^{\mathrm{BF}}$) in the sense that there exist two positive constants $0 < C_1 \leq C_2 < \infty$ such that $C_1 \leq \lambda_j^{\mathrm{PM}}/\lambda_j^{\mathrm{BF}} \leq C_2$ for all j and

$$\lambda_j^{PM} \sim \lambda_j^{BF} \sim \begin{cases} \sqrt{\frac{\log n}{n}}, & j \leq J_\alpha, \\ \frac{\sqrt{j 2^{\alpha j}}}{n}, & j > J_\alpha, \end{cases} \tag{12}$$

where $J_\alpha = (1/\alpha)\log_2 n$.

The resulting Bayesian wavelet estimators $\hat{f}$, $\tilde{f}$ and $\check{f}$ are obviously obtained by substituting $\hat{w}_{jk}$, $\tilde{w}_{jk}$ and $\check{w}_{jk}$ respectively instead of w_{jk}, and Y_{-10} instead of w_{-10} in (2).

3 Main results

Consider again the white noise model (1), where $f \in B_{p,q}^s(M)$, $s > 1/p$, $p, q \geq 1$. The difficulty of the pointwise estimation of f at a point $t_0 \in [0, 1]$ is usually measured by the pointwise (local) minimax mean squared error:

$$R_l(n, t_0, B_{p,q}^s(M)) = \inf_{f^{\mathrm{est}}} \sup_{f \in B_{p,q}^s(M)} E(f^{\mathrm{est}}(t_0) - f(t_0))^2$$

Cai (2003) established the minimax pointwise rates over a Besov class $B_{p,q}^s(M)$:

$$R_l(n, t_0, B_{p,q}^s(M)) \asymp n^{-2(s-1/p)/(2(s-1/p)+1)}. \tag{13}$$

We derive now the pointwise rates for the three Bayesian estimators introduced in Sect. 2 and compare them with the optimal rate in (13).

Theorem 1 *Let a mother wavelet ψ have regularity r, $1/p < s < r$, $p,\ q \geq 1$ and $\alpha > 1$. Let f^* be any of the $\widehat{f}$, $\tilde{f}$ or $\check{f}$. Then for any fixed $t_0 \in [0, 1]$*

$$\sup_{f \in B^s_{p,q}(M)} E(f^*(t_0) - f(t_0))^2 = \mathcal{O}\left(\sqrt{\log n}\ \ n^{-(\alpha-1)/\alpha}\right) + \mathcal{O}\left(n^{-2(s-1/p)/\alpha}\right). \quad (14)$$

The following Corollary 1 is an immediate consequence of Theorem 1:

Corollary 1 *Let f^* be any of the $\widehat{f}$, $\tilde{f}$ or $\check{f}$ with $\alpha = 2(s - 1/p) + 1$. Then, under the conditions of Theorem 1, for any fixed $t_0 \in [0, 1]$*

$$\sup_{f \in B^s_{p,q}(M)} (f(t_0) - f^*(t_0))^2 = \mathcal{O}\left(\sqrt{\log n}\, n^{-2(s-1/p)/(2(s-1/p)+1)}\right)$$

Corollary 1 shows that with $\alpha = 2(s - 1/p) + 1$ all the three Bayesian estimators up to a log-factor achieve the optimal rate (13).

It is interesting to compare the obtained results for the local pointwise risk with the corresponding results for the global mean squared error. Donoho and Johnstone (1998) derived the minimax rates for the global (integrated) mean squared error over a Besov class $B^s_{p,q}(M)$:

$$R_g(n, B^s_{p,q}(M)) = \inf_{f^{\text{est}}} \sup_{f \in B^s_{p,q}(M)} E||f^{\text{est}} - f||^2_{L^2[0,1]} \asymp n^{-2s/(2s+1)}. \quad (15)$$

Abramovich et al. (2004) showed that for $p \geq 2$, all the above Bayesian estimators attain the optimal global rate (15) (up to a log-factor) with $\alpha = 2s + 1$. These results still hold for the Bayes Factor estimator for $p \geq (2s + 2)/(2s + 1)$, while the posterior mean and posterior median estimators for $1 \leq p < 2$ can achieve only the best possible rates for *linear* estimators with $\alpha = 2(s + 1 - 1/p)$. This is caused by too large thresholds (or too strong extent of shrinkage for the posterior mean) implied by the procedures for large j (see (12)). They become 'too severe' towards significant coefficients present on high resolution levels for spatially inhomogeneous functions with $1 \leq p < 2$. To get optimal global rates for $1 \leq p < 2$ one should replace a Gaussian nonzero part of the prior (3) by heavier-tailed distributions (Johnstone and Silverman, 2005; Pensky, 2006).

On the contrary, all three Bayesian estimators achieve the minimax pointwise rates for any $p \geq 1$ with the same $\alpha = 2(s - 1/p) + 1$ which is less than those for the global risk. Note also that under the pointwise risks both the minimax rates and the optimal α always depend not only on s but on p as well.

Different (unless $p = \infty$) choices for α for the local and global risks are not, in fact, surprising. Similar results are known for Bayesian Fourier estimators within Sobolev spaces (Zhao, 2000; Li and Zhao, 2002). Cai et al. (2006) discussed the significant differences in minimax properties under the pointwise risk measures and the global MISE. They showed that for $p < \infty$ no estimator can simultaneously have minimax rate both globally and locally at each point. Substituting $\alpha = 2s + 1$ into (14) shows that the penalty on the maximum *local* risk of the *globally* optimal Bayesian estimators is a power of n. The same is also true for the maximum *global* risk of the *locally* optimal Bayesian estimators (apply Theorem 2 of Abramovich et al., 2004 for $\alpha = 2(s - 1/p) + 1$). Cai et al. (2006) proved that while such a penalty is inherent for *any* locally optimal estimator, there exists a globally optimal estimator that attains the pointwise minimax rate up to a log-factor only (instead of power). It is achieved by a wavelet-based estimator with thresholds smaller on high levels than those in (12) (see Cai et al., 2006 for details).

For $p = \infty$, the local and global minimax rates coincide and with $\alpha = 2s + 1$ all the three Bayesian estimators are both locally and globally optimal (up to log-factors).

4 Bayesian paradox

There is a well-known Bayesian paradox when a prior (and, hence, a posterior) leading to a Bayesian estimator that attains the minimax rate over a certain space of functions, has a zero measure on this space (e.g., Wahba, 1983; Zhao, 2000; Li and Zhao, 2002). The additional requirements on the hyperparameters of the prior (3)–(4) are needed to avoid such a paradox. Whereas $\alpha = 2(s - 1/p) + 1$ yields the optimal rates (13) for *any* β, the choice of β should guarantee that the prior is also supported on the assumed Besov space $B^s_{p,q}$.

From (3) and (4) the expected number of non-zero wavelet coefficients on the jth level is $c_2 2^{j(1-\beta)}$. Applying the first Borel–Cantelli lemma, in the case $\beta > 1$, the number of non-zero coefficients in the wavelet expansion is finite almost surely and, hence, with probability one, f will necessarily belong to the same Besov space as the mother wavelet ψ. Consider the more interesting case $0 \le \beta \le 1$. For $\beta = 1$ the expected number of non-zero wavelet coefficients is the same on each level which is typical for piecewise polynomial functions. The case $\beta = 0$ assumes the same probability of being non-zero for all coefficients on all levels that characterizes self-similar processes such as white noise or Brownian motion. Abramovich et al. (1998, Theorem 1) showed that for $0 \le \beta \le 1$ realizations from such a prior will fall (with probability one) within a Besov space $B^s_{p,q}$ if and only if either

$$s + \left(\frac{1}{2}\right) - \left(\frac{\beta}{p}\right) - \left(\frac{\alpha}{2}\right) < 0, \tag{16}$$

or:

$$s + \frac{1}{2} - \frac{\beta}{p} - \frac{\alpha}{2} = 0 \quad \text{and} \quad 0 \le \beta < 1, \quad 1 \le p < \infty, \ \ q = \infty. \tag{17}$$

Abramovich et al. (2004) used these results to find the ranges of admissible β for the globally optimal Bayesian estimators.

The situation is different for their locally optimal counterparts. For the pointwise optimal $\alpha = 2(s - 1/p) + 1$, (16) and (17) immediately imply that the corresponding Bayesian estimators with probability one will lie outside the considered Besov space and the Bayesian paradox is, therefore, unavoidable. These results are similar to those of Li and Zhao (2002) for Bayesian Fourier estimators over Sobolev spaces which are the particular cases of Besov classes with $p = q = 2$. The existence of a prior on Sobolev (or more general Besov) spaces whose Bayes procedures attain the optimal pointwise rates (at least up to a log-factor) has been left as a conjecture in Li and Zhao (2002). It turns out that a modification of the prior (3)–(4) allows one to construct such a prior. Following Abramovich et al.

Table 1 Conditions for parameter γ to produce sample path within the specified Besov space

	$p, q < \infty$	$p = \infty, q < \infty$	$p < \infty, q = \infty$	$p, q = \infty$
$0 \le \beta < 1$	$\gamma < -2/q$	$\gamma < -1 - 2/q$	$\gamma \le 0$	$\gamma \le -1$
$\beta = 1$	$\gamma < -2/q$		$\gamma < 0$	

(1998), consider a more delicate dependence of the variance τ_j^2 in (4) on the level j by adding a third hyperparameter $\gamma : \tau_j^2 = c_1 2^{-\alpha j} j^\gamma,\ \gamma \le 0$. We omit here the straightforward calculus analogous to the proof of Theorem 1 but, intuitively, it is clear that the exponential term $2^{-\alpha j}$ dominates the asymptotic behavior of τ_j^2 and $\alpha = 2(s - 1/p) + 1$ still implies the optimal pointwise rate (13) up to a log-factor (depending now on γ as well).

On the other hand, the additional hyperparameter γ allows one more flexibility to satisfy the requirement on the prior to be within a given Besov space. Abramovich et al. (1998, Theorem 2) extended the conditions (16) and (17) for the modified three-parameter prior. They showed that for $0 \le \beta \le 1$ and $\gamma \le 0$ the corresponding realizations will almost surely lie within a Besov space $B^s_{p,q}(M)$ if and only if either

$$s + \left(\frac{1}{2}\right) - \left(\frac{\beta}{p}\right) - \left(\frac{\alpha}{2}\right) < 0,$$

or: $s + 1/2 - \beta/p - \alpha/2 = 0$ and γ satisfies the conditions summarized in Table 1.

Thus, for the optimally chosen $\alpha = 2(s - 1/p) + 1$, setting $\beta = 1$ and any $\gamma < -2/q$ guarantees that the resulting prior will belong with probability one to a Besov space $B^s_{p,q}$ and essentially solves the Bayesian paradox for the pointwise optimal (up to a log-factor) Bayesian estimators.

Appendix: Proof of Theorem 1

Throughout the proof we use C to denote a generic positive constant, not necessarily the same each time it is used, even within a single equation.

Using Minkowski inequality we have

$$\begin{aligned} E(f^*(t_0) - f(t_0))^2 &= E\Bigg((Y_{-10} - w_{-10})\varphi_{-10}(t_0) \\ &\qquad + \sum_{j\ge 0}\sum_{k=0}^{2^j-1}(w^*_{jk} - w_{jk})\psi_{jk}(t_0)\Bigg)^2 \\ &\le \Bigg(\sqrt{E(Y_{-10} - w_{-1,0})^2}|\varphi_{-10}(t_0)| \\ &\qquad + \sum_{j\ge 0}\sum_{k=0}^{2^j-1}\sqrt{E(w^*_{jk} - w_{jk})^2}|\psi_{jk}(t_0)|\Bigg)^2 \end{aligned}$$

$$:= (T_1 + T_2)^2,$$

where, immediately, $T_1 = \mathcal{O}(1/\sqrt{n})$.

Consider now the second term T_2. Let L be the support length of ψ and $K_j(t_0) = \{k : \psi_{jk}(t_0) \neq 0\}$. At each resolution level j there are at most L basis functions ψ_{jk} whose supports contain t_0 and, therefore, $Card(K_j(t_0)) \leq L$. Exploiting also that $|\psi_{jk}(t_0)| \leq C\ 2^{j/2}$ we have

$$\begin{aligned} T_2 &= \sum_{j\geq 0}\sum_{k=0}^{2^j-1}\sqrt{E(w^*_{jk}-w_{jk})^2}|\psi_{jk}(t_0)| \\ &\leq C\left(\sum_{j=0}^{J_\alpha}\sum_{k\in K(j,t_0)} 2^{j/2}\sqrt{E(w^*_{jk}-w_{jk})^2}\right. \\ &\qquad \left. + \sum_{j=J_\alpha+1}^{\infty}\sum_{k\in K(j,t_0)} 2^{j/2}\sqrt{E(w^*_{jk}-w_{jk})^2}\right) \\ &:= A_1 + A_2 \end{aligned} \tag{18}$$

Denote $b_j = \tau_j^2/(\tau_j^2+\sigma_n^2)$ and $J_\alpha = (1/\alpha)\log_2 n$ as in (12). A simple calculus yields

$$b_j \sim \begin{cases} 1, & j \leq J_\alpha \\ 2^{-\alpha j}n, & j > J_\alpha \end{cases} \quad \text{and} \quad 1-b_j \sim \begin{cases} 2^{\alpha j}/n, & j \leq J_\alpha, \\ 1, & j > J_\alpha. \end{cases} \tag{19}$$

From now on we consider the three cases separately using several intermediate technical results obtained in the proof of Theorem 2 of Abramovich et al. (2004) for the global risk.

Posterior mean. From the proof of Theorem 2 of Abramovich et al. (2004) for the posterior mean we have

$$E(\hat{w}_{jk}-w_{jk})^2 \leq C\left(b_j^2\sigma_n^2 + 2b_j^2 w_{jk}E\left(\frac{\eta_{jk}Y_{jk}}{1+\eta_{jk}}\right) + (1-b_j)^2 w_{jk}^2\right)$$

$E(\eta_{jk}Y_{jk}/(1+\eta_{jk}))$ is a symmetric (nonlinear) shrinkage of Y_{jk} and it is easy to verify that $w_{jk}E(\eta_{jk}Y_{jk}/(1+\eta_{jk})) \geq 0$. Thus,

$$\begin{aligned} T_2 &\leq C\sum_{j\geq 0}\sum_{k\in K(j,t_0)} 2^{j/2}\sqrt{b_j^2\sigma_n^2 + 2b_j^2 w_{jk}E\left(\frac{\eta_{jk}Y_{jk}}{1+\eta_{jk}}\right) + (1-b_j)^2 w_{jk}^2} \\ &\leq C\sum_{j\geq 0}\sum_{k\in K(j,t_0)} 2^{j/2}\left(b_j\sigma_n + b_j\sqrt{2w_{jk}E\left(\frac{\eta_{jk}Y_{jk}}{1+\eta_{jk}}\right)}\right. \\ &\qquad \left. +(1-b_j)|w_{jk}|\right) \end{aligned} \tag{20}$$

The assumption $f \in B^s_{p,q}$ implies $|w_{jk}| \leq C\,2^{-j(s+1/2-1/p)}$. In addition, in the proof of their Lemma 4, Abramovich et al. (2004) showed that

$w_{jk}E\left(\eta_{jk}Y_{jk}/(1+\eta_{jk})\right) \sim \log n/n$ for $j \leq J_\alpha$, while for $j > J_\alpha$ we use the trivial inequality $w_{jk}E\left(\eta_{jk}Y_{jk}/(1+\eta_{jk})\right) \leq w_{jk}^2$. We then have from (18)–(20):

$$A_1 \leq C\left(\sigma_n \sum_{j=0}^{J_\alpha} 2^{j/2} + \sqrt{\frac{\log n}{n}} \sum_{j=0}^{J_\alpha} 2^{j/2} + \frac{1}{n}\sum_{j=0}^{J_\alpha} 2^{j(\alpha-s+1/p)}\right)$$
$$= \mathcal{O}\left(\sqrt{\log n}\, n^{-((\alpha-1)/2\alpha)}\right) + \mathcal{O}\left(n^{-((s-1/p)/\alpha)}\right)$$

and

$$A_2 \leq C\left(\sigma_n n \sum_{j=J_\alpha+1}^{\infty} 2^{-j(\alpha-1/2)} + n \sum_{j=J_\alpha+1}^{\infty} 2^{-j(\alpha+s-1/p)} + \sum_{j=J_\alpha+1}^{\infty} 2^{-j(s-1/p)}\right)$$
$$= \mathcal{O}\left(n^{-((\alpha-1)/2\alpha)}\right) + \mathcal{O}\left(n^{-((s-1/p)/\alpha)}\right).$$

Posterior median. Abramovich et al. (2004, Proof of Theorem 2) showed that

$$E(\tilde{w}_{jk} - w_{jk})^2 \leq C\left(b_j^2(\lambda_j^{\mathrm{PM}})^2 + (1-b_j)^2 w_{jk}^2 + \sigma_n^2 b_j^2\right),$$

where λ_j^{PM} are given in (12). Repeating the arguments for the posterior mean case we have

$$A_1 \leq C\left(\sqrt{\frac{\log n}{n}} \sum_{j=0}^{J_\alpha} 2^{j/2} + \frac{1}{n}\sum_{j=0}^{J_\alpha} 2^{j(\alpha-s+1/p)} + \frac{1}{\sqrt{n}}\sum_{j=0}^{J_\alpha} 2^{j/2}\right)$$
$$= \mathcal{O}\left(\sqrt{\log n}\, n^{-((\alpha-1)/2\alpha)}\right) + \mathcal{O}\left(n^{-((s-1/p)/\alpha)}\right)$$

and

$$A_2 \leq C\left(\sum_{j=J_\alpha+1}^{\infty} \sqrt{j} 2^{-j(\alpha-1)/2} + \sum_{j=J_\alpha+1}^{\infty} 2^{-j(s-1/p)} + \sqrt{n} \sum_{j=J_\alpha+1}^{\infty} 2^{-j(\alpha-1/2)}\right)$$
$$= \mathcal{O}\left(\sqrt{\log n}\, n^{-((\alpha-1)/2\alpha)}\right) + \mathcal{O}\left(n^{-((s-1/p)/\alpha)}\right)$$

Bayes Factor From the Proof of Theorem 2 of Abramovich et al. (2004) it follows that

$$E(\check{w}_{jk} - w_{jk})^2 \leq C\left(\min((\lambda_j^{\mathrm{BF}})^2, w_{jk}^2) + \frac{\sigma_n^4}{(\lambda_j^{\mathrm{BF}})^2}\right),$$

where λ_j^{BF} are also given in (12). The arguments analogous to those used for the previous two cases imply

$$A_1 \leq C\left(\sqrt{\frac{\log n}{n}} \sum_{j=0}^{J_\alpha} 2^{j/2} + \sqrt{\frac{1}{n\log n}} \sum_{j=0}^{J_\alpha} 2^{j/2}\right) = \mathcal{O}\left(\sqrt{\log n}\, n^{-((\alpha-1)/2\alpha)}\right)$$

and

$$A_2 \leq C \left(\sum_{j=J_\alpha+1}^{\infty} 2^{-j(s-1/p)} + \sum_{j=J_\alpha+1}^{\infty} 2^{-j(\alpha-1)/2} j^{-1/2} \right)$$
$$= \mathcal{O}\left(n^{-((s-1/p)/\alpha)}\right) + \mathcal{O}\left(n^{-((\alpha-1)/2\alpha)}\right).$$

Acknowledgments The authors gratefully acknowledge the financial support of Consiglio Nazionale delle Ricerche. We are delighted to thank Umberto Amato for stimulating discussions. Felix Abramovich would like to thank Claudia Angelini and Daniela De Canditiis for excellent hospitality while visiting Naples and Rome to carry out part of this work.

References

Abramovich, F., Amato, U., & Angelini, C. (2004). On optimality of Bayesian wavelet estimators. *Scandinavian Journal of Statistics*, *31*, 217–234

Abramovich, F., Bailey, T. C., & Sapatinas, T. (2000). Wavelet analysis and its statistical applications. *The Statistician*, *49*, 1–29

Abramovich, F., & Sapatinas, T. (1999). Bayesian approach to wavelet decomposition and shrinkage. In P., Müller, & Vidakovic, (Eds.) *Bayesian inference in wavelet based models, Lecture Notes in Statistics, vol. 141* (pp. 33–50). New York: Springer

Abramovich, F., Sapatinas, T., & Silverman, B. W. (1998). Wavelet thresholding via a Bayesian approach. *Journal of the Royal Statistical Society, Series B*, *60*, 725–749

Antoniadis, A. (1997). Wavelets in statistics: a review (with discussion). *Journal of the Italian Statistical Society*, *6*, 97–144

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

Cai, T. (2003). Rates of convergence and adaptation over Besov spaces under pointwise risk. *Statistica Sinica*, *13*, 881–902

Cai, T., Low, M. G., & Zhao L. H. (2006). Tradeoffs between global and local risks in nonparametric function estimation, *Bernoulli* (to appear)

Chipman, H. A., Kolaczyk, E. D., & McCulloch, R. E. (1997) Adaptive Bayesian wavelet shrinkage. *Journal of the American Statistical Association*, *92*, 1413–1421

Clyde, M., Parmigiani, G., & Vidakovic, B. (1998). *Multiple shrinkage and subset selection in wavelets. Biometrika*, *85*, 391–401

Daubechies, I. (1992). *Ten lectures on wavelets*, Philadelphia: SIAM

Donoho, D.L., & Johnstone, I. M. (1998). Minimax estimation via wavelet shrinkage. *Annals of Statistics*, *26*, 879–921

Johnstone, I. M., & Silverman, B. W. (2005). Empirical Bayes selection of wavelet thresholds. *Annals of Statistics*, *33*, 1700–1752

Li, X., & Zhao, L. (2002). Bayesian nonparametric point estimation under a conjugate prior. *Statistics & Probability Letters*, *58*, 23–30

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

Müller, P., & Vidakovic, B. (Eds.) (1999). Bayesian inference in wavelet-based models. In *Lecture Notes in Statistics, vol. 141*. New York: Springer

Pensky, M. (2006). Frequentist optimality of Bayesian wavelet shrinkage rules for Gaussian and non-Gaussian noise. *Annals of Statistics*, *34*, 769–807

Vidakovic, B. (1998). Non-linear wavelet shrinkage with Bayes rules and Bayes factors. *Journal of the American Statistical Association*, *93*, 173–179

Vidakovic, B. (1999). *Statistical modeling by wavelets*. New York: Wiley

Wahba, G. (1983). Bayesian confidence intervals for the cross-validated smoothing spline. *Journal of the Royal Statistical Society, Series B*, *45*, 133–150

Zhao, L. H. (2000). Bayesian aspects of some nonparametric problems. *Annals of Statistics*, *28*, 532–552