Ann. Inst. Statist. Math.
Vol. 56, No. 2, 205–223 (2004)

ANALYSIS OF BLOCKWISE SHRINKAGE WAVELET ESTIMATES VIA LOWER BOUNDS FOR NO-SIGNAL SETTING

SAM EFROMOVICH*

Department of Mathematics and Statistics, University of New Mexico, 468 Humanities Bldg, Albuquerque, NM 87131-1141, U.S.A.

(Received January 29, 2003; revised July 9, 2003)

Abstract. A blockwise shrinkage is a popular procedure of adaptation that has allowed the statisticians to establish an impressive bouquet of asymptotic mathematical results and develop softwares for solving practical problems. Traditionally risks of the estimates are studied via upper bounds that imply sufficient conditions for a blockwise shrinkage procedure to be minimax. This article suggests to analyze the estimates via exact (non-asymptotic) lower bounds established for a no-signal setting. The approach complements the familiar minimax, Bayesian and numerical analysis, it allows to find necessary conditions for a procedure to attain desired rates, and it sheds a new light on popular choices of blocks and thresholds recommended in the literature. Mathematical results are complemented by a numerical study.

Key words and phrases: Adaptation, asymptotic, nonparametric, minimax, Stein shrinkage, oracle, regression, small sample.

1. Introduction

A blockwise shrinkage nonparametric estimation, suggested in the early 1980s for a data-driven nonparametric Fourier series estimation and then in the middle 1990s for wavelets, has shown to be well suited for many applied settings of nonparametric curve estimation including filtering, regression, density and spectral density estimation. The estimator is data-driven, rate minimax over vast classes of smooth and spatially inhomogeneous functions, it is also robust, superefficient, can "mimic" oracles, implies optimal plug-in estimation of functionals and solving ill-posed problems. On the top of this, a large variety of particular blocks and shrinkage procedures can imply all the above-mentioned statistical properties; see the discussion in Brown *et al.* (1997), Downie and Silverman (1998), Hall *et al.* (1998, 1999), Härdle *et al.* (1998), Cai (1999), Efromovich ((1999), ch. 7), Cai and Silverman (2000), Nemirovski (2000), Cavalier and Tsybakov (2001), Abramovich *et al.* (2002), Tsybakov (2002), Zhang (2002), Chiken (2003), and DeCanditiis and Vidakovic (2004).

To describe the method, let us consider a classical homoscedastic regression model with observations $Y_l = g(l/n) + v^{1/2}\epsilon_l$, $l = 1, 2, \dots, n$, where g is a bounded regression function estimated on $[0, 1]$, $\epsilon_1, \dots, \epsilon_n$ are independent standard Gaussian random variables and v is known. Under mild assumptions, this model is equivalent to a series model written in Fourier, wavelet or any other orthogonal basis domain.

*Supported in part by NSF Grants DMS-9971051 and DMS-0243606.

In this article a periodized orthonormal wavelet basis on $[0,1]$ is considered. It is supposed that the basis is generated by a pair of compactly supported scaling function (father wavelet) ϕ and wavelet function (mother wavelet) ψ. (The reader interested in Fourier bases can use a single index instead of two wavelet indices—this will imply the corresponding Fourier results; see the discussion in Efromovich (2000).) For a j-th resolution scale set $\phi_{j,k}(x) = 2^{j/2}\phi(2^j x - k)$, $\psi_{j,k}(x) = 2^{j/2}\psi(2^j x - k)$, $x \in [0,1]$. Then the collection $\{\phi_{j_0,k}, k = 1,2,\ldots,2^{j_0}; \psi_{j,k}, j \geq j_0 \geq 0, k = 1,2,\ldots,2^j\}$ is an orthonormal basis on $[0,1]$ provided the primary resolution level j_0 is large enough to ensure that the supports of the scale and wavelet functions at level j_0 are not the whole of the unit interval. Then the regression function g, which by assumption is bounded on the unit interval, can be expanded into a wavelet series

$$g(x) = \sum_{k=1}^{2^{j_0}} \kappa_{j_0,k}\phi_{j_0,k}(x) + \sum_{j=j_0}^{\infty}\sum_{k=1}^{2^j} \theta_{j,k}\psi_{j,k}(x), \qquad x \in [0,1], \tag{1.1}$$

where $\kappa_{j_0,k} = \int_0^1 g(x)\phi_{j_0,k}(x)dx$ and $\theta_{j,k} = \int_0^1 g(x)\psi_{j,k}(x)dx$ are scaling and wavelet coefficients, respectively.

The statistical assumption is a traditional one: the sample size n is dyadic and for all $j_0 \leq j \leq \log_2(n/\ln(n))$ and all k considered there exist statistics $\tilde\theta_{j,k}$ (empirical wavelet coefficients) and $\tilde\kappa_{j_0,k}$ (empirical scaling coefficients) such that $\tilde\theta_{j,k} = \theta_{j,k} + v^{1/2}n^{-1/2}\xi_{j,k}$ and $\tilde\kappa_{j_0,k} = \kappa_{j_0,k} + v^{1/2}n^{-1/2}\eta_{j_0,k}$ where $(\{\xi_{j,k}\},\{\eta_{j_0,k}\})$ are independent standard normal random variables. The interested reader can find more about wavelet bases and the assumptions in Johnstone (1998), Mallat (1999) and Vidakovic (1999).

Let us now briefly review blockwise shrinkage estimates. Scaling coefficients are always estimated by empirical ones. Wavelet coefficients are estimated by using special blockwise shrinkage procedures. Consider a particular wavelet coefficient $\theta_{j,k}$ and assume that it is estimated using a block $\{\tilde\theta_{j,s}, s \in T_{j,m}\}$ of empirical wavelet coefficients. In what follows it is always assumed that $k \in T_{j,m}$, each wavelet coefficient (index) belongs to a single block and only consecutive coefficients from the same scale are included into a block. Let us denote cardinality (size, length) of the block by $L_{j,m}$. Then all blockwise procedures, suggested so far in the literature, try to mimic the benchmark

$$\hat\theta^*_{j,k} = \frac{L_{j,m}^{-1}\sum_{s\in T_{j,m}}\theta^2_{j,s}}{L_{j,m}^{-1}\sum_{s\in T_{j,m}}\theta^2_{j,s} + vn^{-1}}\tilde\theta_{j,k}. \tag{1.2}$$

This benchmark is referred to as a blockwise shrinkage oracle because: it is based on both data and underlying wavelet coefficients; it always shrinks empirical coefficients; the shrinkage procedure is based on a block of estimated coefficients. Also note that (1.2) is a blockwise version of the ideal Wiener filter; see the discussion in Mallat (1999).

It is easy to see that the term $\sum\theta^2_{j,s}$ can be estimated by the unbiased statistic $\sum(\tilde\theta^2_{j,s} - vn^{-1})$, and this is the method used by all shrinkage estimates known in the literature. It is also well known that a shrinkage factor should be nonnegative because otherwise the estimate is inadmissible; see Lehmann and Casella (1998).

Thus, instead of plugging in the unbiased estimate in (1.2), two blockwise shrinkage procedures with nonnegative shrinkage factors have been suggested in the nonparametric literature. The former is

$$(1.3) \qquad \hat{\theta}_{j,k} = \frac{L_{j,m}^{-1} \sum_{s \in T_{j,m}} \tilde{\theta}_{j,s}^2 - vn^{-1}}{L_{j,m}^{-1} \sum_{s \in T_{j,m}} \tilde{\theta}_{j,s}^2} I \left(L_{j,m}^{-1} \sum_{s \in T_{j,m}} \tilde{\theta}_{j,s}^2 > (1 + \lambda_{j,m}^2) vn^{-1} \right) \tilde{\theta}_{j,k},$$

and it is referred to as Efromovich-Pinsker (EP) blockwise shrinkage. Here $\lambda_{j,m}^2$ are some nonnegative constants that are referred to as thresholds. Then a particular EP estimate is defined by its set of blocks and thresholds $\{T_{j,m}, \lambda_{j,m}^2\}$. This adaptive procedure was originally suggested for Fourier bases in Efromovich and Pinsker (1984, 1986) and Efromovich (1985, 1986) for sharp minimax filtering and estimation of spectral density, probability density and regression function, respectively; for wavelets and multiwavelets it was explored in Efromovich (1999, 2000, 2001). Note that EP procedure combines a naive mimicking of (1.2) with the idea of hard thresholding.

The latter blockwise shrinkage procedure is

$$(1.4) \qquad \bar{\theta}_{j,k} = \frac{L_{j,m}^{-1} \sum_{s \in T_{j,m}} \tilde{\theta}_{j,s}^2 - (1 + \mu_{j,m}) vn^{-1}}{L_{j,m}^{-1} \sum_{s \in T_{j,m}} \tilde{\theta}_{j,s}^2} \times I \left(L_{j,m}^{-1} \sum_{s \in T_{j,m}} \tilde{\theta}_{j,s}^2 > (1 + \mu_{j,m}) vn^{-1} \right) \tilde{\theta}_{j,k}.$$

Constants $\mu_{j,m}$, that are not necessarily nonnegative, also referred to as thresholds. This shrinkage procedure combines a naive mimicking of (1.2) with the idea of soft thresholding.

For Fourier bases this adaptive procedure was suggested in Efromovich and Pinsker (1996) where its sharp minimaxity was established for a heteroscedastic regression model with random or fixed design. For wavelet bases the procedure was pioneered by Donoho and Johnstone ((1995), p. 1216) who established its rate minimax optimality for the studied homoscedastic regression model. Other instrumental references are Brown *et al.* (1997), Cai (1999, 2002), Cai *et al.* (2000), Cai and Silverman (2000), Cavalier and Tsybakov (2001), Abramovich *et al.* (2002), Tsybakov (2002), Zhang (2002) and DeCanditiis and Vidakovic (2004).

If $\mu_{j,m} = -2/L_{j,m}$, $L_{j,m} > 2$ then in the parametric literature (see Lehmann and Casella (1998), ch. 5) the shrinkage (1.4) is called the (positive-part) James-Stein one, and the shrinkage with $\mu_{j,m} \geq 0$, $L_{j,m} > 4$ is called the Stein shrinkage. Moreover, the classical Stein shrinkage uses $\mu_{j,m} = 0$, so to emphasize cases of the zero and positive thresholds, we may write Stein(0) and Stein(> 0). This terminology, motivated by the classical parametric theory, will be used in this article.

For simplicity in exposition and because the studied estimates are different only in the estimation of wavelet coefficients, from now on we are considering only the problem of estimation of

$$(1.5) \qquad f(x) = \sum_{j=j_0}^{\infty} \sum_{k=1}^{2^j} \theta_{j,k} \psi_{j,k}(x).$$

We can do this because for any blockwise shrinkage estimate $\check{g}$ and any fixed point of interest $x_0 \in [0,1]$ we can write

$$(1.6) \qquad E(\check{g}(x_0) - g(x_0))^2 = 2^{j_0} v n^{-1} \sum_{k=1}^{2^{j_0}} \phi^2(2^{j_0} x_0 - k) + E(\check{f}(x_0) - f(x_0))^2,$$

and

$$(1.7) \qquad E \int_0^1 (\check{g}(x) - g(x))^2 dx = 2^{j_0} v n^{-1} + E \int_0^1 (\check{f}(x) - f(x))^2 dx,$$

where $\check{f}(x) = \check{g}(x) - \sum_{k=1}^{2^{j_0}} \tilde{\kappa}_{j_0,k} \phi_{j_0,k}(x)$.

As it has been mentioned, some of the blockwise shrinkage estimates can attain adaptive minimax rates under the global and/or pointwise approaches. Function classes, typically considered in the literature, include parametric, analytic, Hölder, Sobolev and Besov classes. Recall that an adaptive minimax rate R can be written as

$$(1.8) \qquad R = (\ln^{\gamma}(n)/n)^{\beta},$$

where, depending on a function class and an approach used (pointwise or global), the parameters (γ, β) satisfy $\gamma \geq 0$ and $0 < \beta \leq 1$. These rates indicate a required minimax accuracy of adaptive estimation which, of course, depends on an underlying function class (regression function). In particular, a good adaptive estimate should attain the rate (1.8) with parameters: $(\gamma = 0, \beta = 1)$ for classical parametric functions; $(\gamma = 1, \beta = 1)$ for analytic functions; $(\gamma = 0, \beta < 1)$ for Hölder functions under the global approach, and $(\gamma = 1, \beta < 1)$ for Hölder functions under the pointwise approach. The interested reader can find more about adaptive minimax rates in Härdle *et al.* (1998), Johnstone (1998) and Efromovich (1999).

The aim of this article is to find necessary conditions for EP, James-Stein and Stein blockwise shrinkage estimates to attain those minimax rates. Our tool will be exact (not-asymptotic) lower bounds for mean squared and mean integrated squared errors of the estimates for the case of no-signal setting, that is, $f(x) = f_0(x)$ where $f(x) \equiv 0$. This function plays the pivotal role in the minimax literature (see Efromovich (1999), ch. 7) and it belongs to all(!) function classes considered in the literature, that is, it is parametric, analytic, Hölder, Sobolev, etc. As a result, if an adaptive procedure fails to perform well for this function, that procedure should raise eyebrows. Also, recall that this function is in the core of the universal threshold paradigm of Donoho and Johnstone. And the lower bound for f_0 is automatically a minimax lower bound for any of the above-mentioned function classes.

As we shall see, this intuitively clear approach of analyzing risks via the no-signal setting will produce a wealth of information about blockwise estimates.

The content of the article is as follows. Lower bounds are presented in Section 2. Section 3 is devoted to a discussion of the results. Section 4 presents a numerical study. Proofs are deferred to the Appendix.

2. Lower bounds

Let us denote by $\hat{f}^*$ the blockwise shrinkage oracle, by $\hat{f}$ the EP estimator, and by $\bar{f}_{JS}$ and $\bar{f}_S$ estimators based on the James-Stein shrinkage and the Stein shrinkage, respectively. For instance,

$$\hat{f}(x) = \sum_{j=j_0}^{\log_2(n/\ln(n))} \sum_m \sum_{k \in T_{j,m}} \hat{\theta}_{j,k}\psi_{j,k}(x), \qquad x \in [0,1], \tag{2.1}$$

where $\hat{\theta}_{j,k}$ are defined in (1.3), $\hat{f}^*$ is defined by (2.1) with $\hat{\theta}_{j,k}$ replaced by $\hat{\theta}^*_{j,k}$ defined in (1.2), $\bar{f}_{JS}$ is also defined by (2.1) with $\hat{\theta}_{j,k}$ replaced by $\bar{\theta}_{j,k}$ defined in (1.4) with $\mu_{j,m} = -2/L_{j,m}$ and the understanding that $L_{j,m} > 2$, etc. In (2.1) and in what follows $\sum_m$ is the summation over a set M_j such that the corresponding blocks $\{T_{j,m}, m \in M_j\}$ cover the entire j-th scale; for example $\sum_m L_{j,m} = 2^j$.

In what follows C's denote generic positive constants, and a pointwise estimation is considered at a point $x_0 \in [0,1]$, and it is always assumed that the underlying wavelet component of an estimated regression function is $f_0(x) \equiv 0$, that is, we consider the no-signal setting.

We begin with lower bounds for the EP estimate (2.1).

THEOREM 2.1. *The EP estimator $\hat{f}$ satisfies*

$$\begin{aligned} &E(\hat{f}(x_0) - \hat{f}^*(x_0))^2 \\ &\quad = E(\hat{f}(x_0) - f_0(x_0))^2 \\ &\quad > Cn^{-1}v \sum_{j=j_0}^{\log_2(n/\ln(n))} \sum_m \left[\frac{\lambda^2_{j,m}}{1+\lambda^2_{j,m}} + \frac{1}{(1+\lambda^2_{j,m})^2 L_{j,m}}\right] \frac{1}{L^{1/2}_{j,m}} \\ &\qquad \times e^{j\ln(2) - [\lambda^2_{j,m} - \ln(1+\lambda^2_{j,m})]L_{j,m}/2} \sum_{k \in T_{j,m}} \psi^2(2^j x_0 - k). \end{aligned} \tag{2.2}$$

Remark 2.1. The pointwise lower bound (2.2) depends on sums of squared wavelet functions. Let us comment on how the statistician can evaluate them. Define a function

$$\Psi(j,x) = \sum_{k=1}^{2^j} \psi^2(2^j x - k), \qquad j \geq j_0, \tag{2.3}$$

and let us describe some of its properties. First of all, for Haar basis this function is identically equal to 1 (note that this basis also makes the analysis of (2.2) trivial). Second, recall that we consider only wavelet functions with a compact support. Let s be the rounded up length of the support of a particular ψ. Then at each scale j there are at most s number of wavelets $\psi(2^j x - k)$ whose support includes x (the interested reader can find more in Mallat (1999), p. 243). Third, it is easy to see that neither maximum nor minimum of $\Psi(j,x)$ depends on j. Fourth, the maximum is always finite. Can we say that the minimum is bounded below from zero? Figure 1 exhibits the function for several popular wavelets (recall that we are considering only periodized bases). As we see, in all these cases the function is separated from zero and the maximum/minimum ratio is reasonable. The author also checked that the same outcome held for all wavelets supported by S-PLUS wavelet package. Finally, note that $\max_k \min_x \psi^2(2^j x - k) \geq \min_x \Psi(j,x)/s$.

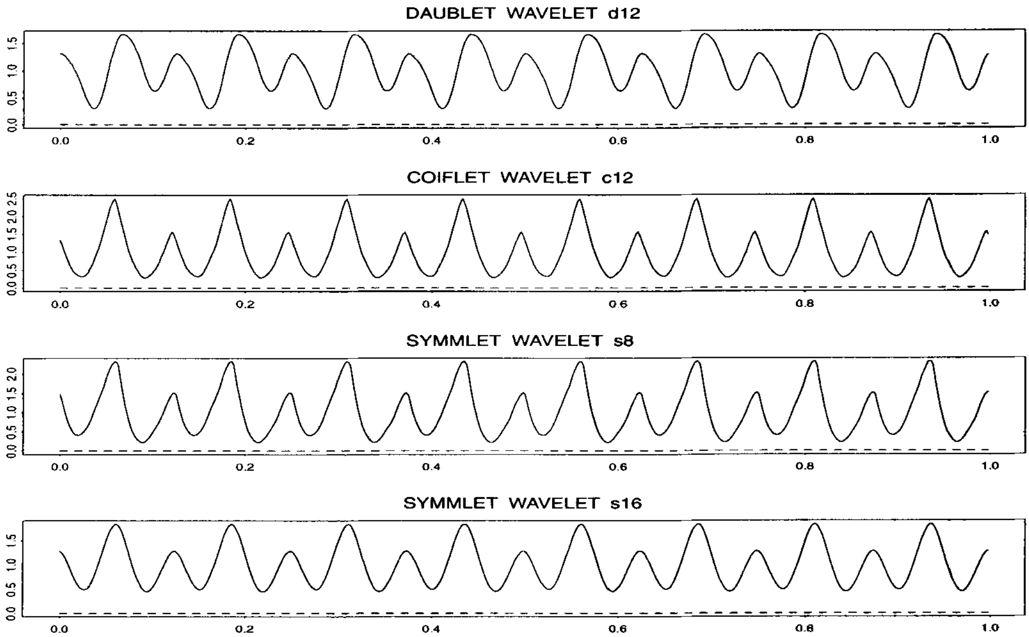

Fig. 1. Function $\Psi(3, x)$ for several popular wavelet functions. The dashed line exhibits the zero level.

Remark 2.1 implies that it is reasonable to assume that a considered wavelet function ψ satisfies $\min_x \Psi(j, x) = c^* > 0$, $j \geq j_0$. From now on only such wavelet functions are considered.

The last assumption together with Theorem 2.1 implies the following pointwise lower bound.

COROLLARY 2.1. *Assume that $L_{j,m} \equiv L_j$ and $\lambda_{j,m}^2 \equiv \lambda_j^2$, that is, at each scale only identical block lengths and threshold levels are used. Then*

$$
\begin{aligned}
(2.4) \qquad E(\hat{f}(x_0) - \hat{f}^*(x_0))^2 &= E(\hat{f}(x_0) - f_0(x_0))^2 \\
&> Cn^{-1}v \sum_{j=j_0}^{\log_2(n/\ln(n))} \left[\frac{\lambda_j^2}{1+\lambda_j^2} + \frac{1}{(1+\lambda_j^2)^2 L_j} \right] \frac{1}{L_j^{1/2}} \\
&\quad \times e^{j\ln(2) - [\lambda_j^2 - \ln(1+\lambda_j^2)]L_j/2}.
\end{aligned}
$$

The pointwise result (2.2) also implies the following global lower bounds.

COROLLARY 2.2. *The EP estimate $\hat{f}$ satisfies*

$$
\begin{aligned}
(2.5) \qquad & E\int_0^1 (\hat{f}(x) - \hat{f}^*(x))^2 dx \\
&= E\int_0^1 (\hat{f}(x) - f_0(x))^2 dx
\end{aligned}
$$

$$> Cn^{-1}v \sum_{j=j_0}^{\log_2(n/\ln(n))} \sum_m \left[\frac{\lambda_{j,m}^2}{1+\lambda_{j,m}^2} + \frac{1}{(1+\lambda_{j,m}^2)^2 L_{j,m}} \right] L_{j,m}^{1/2}$$
$$\times e^{-[\lambda_{j,m}^2 - \ln(1+\lambda_{j,m}^2)]L_{j,m}/2}.$$

In particular, if $L_{j,m} \equiv L_j$ *and* $\lambda_{j,m}^2 \equiv \lambda_j^2$, *that is, at each scale only identical block lengths and threshold levels are used, then*

$$(2.6)\qquad E\int_0^1 (\hat{f}(x) - \hat{f}^*(x))^2 dx = E\int_0^1 (\hat{f}(x) - f_0(x))^2 dx$$
$$> Cn^{-1}v \sum_{j=j_0}^{\log_2(n/\ln(n))} \left[\frac{\lambda_j^2}{1+\lambda_j^2} + \frac{1}{(1+\lambda_j^2)^2 L_j} \right] \frac{1}{L_j^{1/2}}$$
$$\times e^{j\ln(2) - [\lambda_j^2 - \ln(1+\lambda_j^2)]L_j/2}.$$

Let us make three preliminary comments about the results obtained. First of all, the lower bounds are not asymptotic. Moreover, proofs presented in the Appendix allow the interested reader to evaluate the generic constants C's used in the bounds. Second, we got lower bounds on how well EP estimate can mimic the blockwise shrinkage oracle $\hat{f}^*$. Finally, for the case of identical block lengths and threshold levels at each scale, the pointwise and global lower bounds coincide up to a constant factor.

Now let us present lower bounds for the James-Stein estimate.

THEOREM 2.2. *Assertions of Theorem* 2.1 *and Corollaries* 2.1–2.2 *hold for a blockwise estimator with the James-Stein shrinkage if one formally sets* $\lambda_{j,m}^2 \equiv 0$, *that is,*

$$(2.7)\qquad E(\bar{f}_{JS}(x_0) - \hat{f}^*(x_0))^2 = E(\bar{f}_{JS}(x_0) - f_0(x_0))^2$$
$$> Cn^{-1}v \sum_{j=j_0}^{\log_2(n/\ln(n))} \sum_m 2^j L_{j,m}^{-3/2} \sum_{k\in T_{j,m}} \psi^2(2^j x_0 - k),$$

and

$$(2.8)\qquad E\int_0^1 (\bar{f}_{JS}(x) - \hat{f}^*(x))^2 dx = E\int_0^1 (\bar{f}_{JS}(x) - f_0(x))^2 dx$$
$$> Cn^{-1}v \sum_{j=j_0}^{\log_2(n/\ln(n))} \sum_m L_{j,m}^{-1/2}.$$

In particular, if $L_{j,m} \equiv L_j$ *(at each scale all blocks have the same length) then*

$$(2.9)\qquad \min\Bigg(E(\bar{f}_{JS}(x_0) - \hat{f}^*(x_0))^2, E(\bar{f}_{JS}(x_0) - f_0(x_0))^2,$$
$$E\int_0^1 (\bar{f}_{JS}(x) - \hat{f}^*(x))^2 dx, E\int_0^1 (\bar{f}_{JS}(x) - f_0(x))^2 dx \Bigg)$$
$$> Cn^{-1}v \sum_{j=j_0}^{\log_2(n/\ln(n))} 2^j L_j^{-3/2}.$$

Note how simple the lower bounds are.

Now let us consider a Stein estimator. Recall that in this case: threshold levels $\mu_{j,m}$ are nonnegative and they proxy the thresholds $\lambda^2_{j,m}$ used in the EP estimator.

THEOREM 2.3. *Stein*(0) *estimate satisfies* (2.2) *and* (2.4)–(2.6) *with* $\lambda^2_{j,m} \equiv 0$. *For Stein*(>0) *estimate consider a nonnegative sequence* $\gamma_{j,m}$. *Then*

$$\begin{aligned}
(2.10)\qquad & E(\bar{f}_S(x_0) - \hat{f}^*(x_0))^2 \\
&= E(\bar{f}_S(x_0) - f_0(x_0))^2 \\
&> Cn^{-1}v \sum_{j=j_0}^{\log_2(n/\ln(n))} \sum_m \frac{\gamma^2_{j,m}}{(1+\gamma_{j,m}+\mu_{j,m})^2} \frac{1}{L^{1/2}_{j,m}} \\
&\quad \times e^{j\ln 2 - [\gamma_{j,m}+\mu_{j,m} - \ln(1+\gamma_{j,m}+\mu_{j,m})]L_{j,m}/2} \sum_{k \in T_{j,m}} \psi^2(2^j x_0 - k),
\end{aligned}$$

and

$$\begin{aligned}
(2.11)\qquad E\int_0^1 (\bar{f}_S(x) - \hat{f}^*(x))^2 dx &= E\int_0^1 (\bar{f}_S(x) - f_0(x))^2 dx \\
&> Cn^{-1}v \sum_{j=j_0}^{\log_2(n/\ln(n))} \sum_m \frac{\gamma^2_{j,m}}{(1+\gamma_{j,m}+\mu_{j,m})^2} L^{1/2}_{j,m} \\
&\quad \times e^{-[\gamma_{j,m}+\mu_{j,m} - \ln(1+\gamma_{j,m}+\mu_{j,m})]L_{j,m}/2}.
\end{aligned}$$

If additionally $\mu_{j,m} \equiv \mu_j$ *and* $L_{j,m} \equiv L_j$ *then*

$$\begin{aligned}
(2.12)\qquad & \min\Big(E(\bar{f}_S(x_0) - \hat{f}^*(x_0))^2, E(\bar{f}_S(x_0) - f_0(x_0))^2, \\
& E\int_0^1 (\bar{f}_S(x) - \hat{f}^*(x))^2 dx, E\int_0^1 (\bar{f}_S(x) - f_0(x))^2 dx\Big) \\
&> Cn^{-1}v \sum_{j=j_0}^{\log_2(n/\ln(n))} \frac{\gamma^2_j}{(1+\gamma_j+\mu_j)^2} \frac{1}{L^{1/2}_j} \\
&\quad \times e^{j\ln 2 - [\gamma_j+\mu_j - \ln(1+\gamma_j+\mu_j)]L_j/2}.
\end{aligned}$$

Remark 2.2. It follows from the proofs that all pointwise lower bounds, presented in this section, hold for a more general class of functions $f_0(x)$ whose wavelet coefficients $\theta_{j,k} = 0$ for $(j,k) \in \{(j,k) : \psi(2^j x_0 - k) \neq 0\}$.

3. Discussion of results

Lower bounds allow us to present necessary conditions for blockwise shrinkage estimates to be minimax over a wide variety of function classes that include, in particular, parametric, Hölder and Besov classes. Recall that the corresponding rates were highlighted in Introduction.

3.1 *James-Stein estimator*

Assume that a statistician would like to use a James-Stein estimator $\bar{f}_{JS}$. Then the statistician must to choose a set of blocks. There are mixed recommendations in the literature on how to do this. Donoho and Johnstone (1995) recommend dyadic blocks and show that under the global approach they imply Hölder minimax rates. On the other hand, that article also highlights a bad performance of the estimate in simulated examples. In numerical studies much smaller blocks are typically used and recommended. See a discussion in Cai *et al.* (2000) and DeCanditiis and Vidakovic (2004).

Thus, it is of interest to understand how small blocks can be to give the typical rates of convergence highlighted in (1.8). Let us use Theorem 2.2 for finding the answer.

To simplify the discussion, let us restrict our attention to blocks of the same length at each scale ($L_{j,m} \equiv L_j$) and $L_j \le L_{j+1}$. These are the typical blocks considered in the literature. Set $L_j =: (2^j b_j)^{2/3}$, $b_j > 0$. Then (2.9) yields

$$\tag{3.1} \begin{aligned} &\min\left(E(\bar{f}_{JS}(x_0) - f_0(x_0))^2, E\int_0^1 (\bar{f}_{JS}(x) - f_0(x))^2 dx\right) \\ &\quad \ge Cn^{-1} v \sum_{j=j_0}^{\log_2(n/\ln(n))} b_j^{-1}. \end{aligned}$$

This lower bound implies that the necessary condition for a James-Stein estimate to attain the parametric rate n^{-1} is $\sum_j b_j^{-1} < \infty$. This immediately implies that L_j should increase faster than $2^{(2/3)j}$. For attaining the analytic rate $\ln(n)n^{-1}$ the block lengths must be of order $2^{(2/3)j}$. The blocks can be smaller a bit for Besov spaces where the rate is $(\ln^\gamma(n)/n)^\beta$. However, the parameter $\beta < 1$ depends on the underlying Besov space which is unknown to the statistician and thus β is also unknown. This implies that for any $\alpha > 0$ the block lengths should satisfy

$$n^{-\alpha} \sum_{j=j_0}^{\log_2(n/\ln(n))} b_j^{-1} \to 0 \quad \text{as} \quad n \to \infty.$$

As we see, the slower minimax Besov rates do not help a lot in decreasing the necessary block lengths unless the statistician knows an upper bound for β.

As it has been explained earlier, because the considered f_0 belongs to all classical function spaces, the conclusion made also holds for the corresponding minimax approaches. Indeed, it is easy to see that

$$\tag{3.2} \begin{aligned} &\sup_{g \in \mathcal{F}} E\int_0^1 (\bar{f}_{JS}(x) - f(x))^2 dx \\ &\quad \ge E\int_0^1 (\bar{f}_{JS}(x) - f(x))^2 dx \quad \text{whenever} \quad f \in \mathcal{F}, \end{aligned}$$

and obviously a similar conclusion is valid for the pointwise approach.

We may conclude that using the classical James-Stein shrinkage, that plays the prominent role in the parametric shrinkage theory, requires employment of at least geometrically increasing blocks whenever the statistician wants to achieve the classical rates. This conclusion will be complemented by a numerical study presented in the next section.

3.2 *Stein*(0) *estimate*

According to Theorem 2.3, discussion of the necessity to employ geometrically increasing blocks for obtaining the classical rates is absolutely similar to the above-presented case for the James-Stein estimator. Thus let us make several complementing remarks.

Tsybakov (2002) proved that Stein(0) estimator with a special set of geometrically increasing blocks implies a minimax estimation over a Sobolev class. This implies sharpness of the lower bound obtained.

Also note that if a statistician would like to use Stein(0) estimator with constant blocks, then to get the classical rates it is necessary to use blocks proportional to $n/\ln(n)$.

Finally, let us recall one more time that typically large blocks imply a poor performance for small datasets. On the other hand, these blocks dramatically simplify proofs of asymptotic results and make the asymptotic results more transparent. Thus, the statistician should be aware about good and bad properties of a particular estimate and then use the estimate correspondingly.

3.3 *EP estimate*

It is clear from the previous discussion, Theorem 2.1 and Corollaries 2.1–2.2 that blocks should not be too small for attaining minimax rates. Thus, let us begin the discussion of EP estimate with the case of identical (over all scales) blocks and thresholds ($L_{j,m} \equiv L$, $\lambda_{j,m} \equiv \lambda$) studied in Cai (1999) under the global approach. Corollary 2.2 implies that the necessary condition for EP estimate to attain the parametric rate n^{-1} is

$$\left[\frac{\lambda^2}{1+\lambda^2} + \frac{1}{(1+\lambda^2)^2 L}\right] L^{-1/2} e^{-[\lambda^2 - \ln(1+\lambda^2)]L/2} < Cn^{-1}\ln(n). \tag{3.3}$$

Thus either λ^2 or L should be large. For instance, if we consider the "limit" case $L = 1$ then we get the classical $\lambda^2 \geq 2\ln(n)[1 - o(1)]$. This together with the familiar upper bound for the risk of the universal threshold estimator of Donoho and Johnstone (1994) implies sharpness of the lower bound.

If $0 < c_1 < \lambda^2 < c_2 < \infty$ then (3.3) yields

$$[\lambda^2 - \ln(1+\lambda^2)]L + \ln(L) > 2\ln(n)[1 - \ln(C\ln(n))/\ln(n)]. \tag{3.4}$$

Thus the block length L is necessarily at least logarithmic and this coincides with the conclusion of Cai (2000) obtained from a minimax study.

Let us continue the exploration of (3.4) and solve the equation $[\lambda^2 - \ln(1+\lambda^2)]L = 2\ln(n)$ for L being the rounded up $\ln(n)$. The solution is $\lambda_*^2 = 3.50524$, which is exactly the famous Cai's optimal threshold μ_* for Stein($>$ 0) estimator with identical logarithmic blocks. This is an interesting outcome because Cai ((1999), p. 910) got this threshold level from solving a very special optimization problem motivated by ideas of Wahba (1990), Donoho and Johnstone (1994) and Donoho (1995). The methodology of the present article has absolutely no connection with the Cai (1999) approach, but interestingly enough it implies the same optimal threshold level. We shall continue the discussion of this "coincidence" in the next subsection.

If the goal is to attain a Besov minimax rate $n^{-\beta}$, $0 < \beta < 1$ (here we skip a possible logarithmic factor to simplify the discussion) instead of the parametric rate n^{-1} then blocks and thresholds should satisfy

$$[\lambda^2 - \ln(1+\lambda^2)]L + \ln(L) > 2\beta\ln(n)[1 - \ln(C\ln(n))/\ln(n)]. \tag{3.5}$$

Thus, if it is known that $\beta < \beta_0 < 1$ then λ and/or L can be slightly decreased. Unfortunately, typically β_0 is unknown and thus even the slower Besov rates do not help a lot in the decreasing of the lower limit on block lengths and threshold levels.

Another important set of blocks-thresholds used in applications is where blocks and thresholds are the same within each scale. In this case the necessary condition for obtaining the parametric rate n^{-1} is that for some $j^* \geq j_0$

$$[\lambda_j^2 - \ln(1+\lambda_j^2)]L_j + (1/2)\ln(L_j) > (j - j^*)\ln(4). \tag{3.6}$$

Interestingly, for the "limit" case $L_j \equiv 1$ the equality in (3.7) implies the familiar minimax estimates suggested by Delyon and Juditsky (1996) and Juditsky (1997), see also the discussion in Vidakovic (1999).

3.4 *Stein(> 0) estimate*

All the conclusions, made for EP estimate, hold here as well. At least from the lower bounds point of view, the estimates perform similarly. Thus, let us instead of repeating the above-formulated conclusions make several new remarks.

If, similarly to Cai (1999), we restrict our attention to identical blocks $L_{j,m} \equiv L := L(n)$ and a constant threshold level $\mu > 1$ then the inequality

$$\frac{n}{\ln(n)L^{1/2}}\frac{\gamma^2}{(1+\mu+\gamma)^2}\exp\{-[\gamma+\mu-\ln(1+\gamma+\mu)]L/2\} < C \tag{3.7}$$

is the necessary condition for attaining the parametric rate n^{-1} under the pointwise and global approaches. Plainly

$$[\gamma+\mu-\ln(1+\gamma+\mu)]L \geq 2\ln(n) \tag{3.8}$$

implies (3.7) and $[\gamma+\mu-\ln(1+\gamma+\mu)]L < 2\ln(n)(1-c_0)$, $c_0 > 0$ does not. In particular, consider the Cai's choice $L = \ln(n)$. If we set, for instance, $\gamma = \ln^{-1}(n)$, then this together with the equality in (3.8) yields the Cai's optimal threshold $\mu_* = 3.50523$.

The lower bound approach implies that μ_* is simply the *minimal* identical threshold level preserving the parametric rate of convergence for identical logarithmic blocks and no-signal setting. Similarly to the previous subsections, we can also conclude that μ_* has the same meaning for the other classical function classes. Thus, the Cai's optimality, based on the Wahba approach, is equivalent to the minimal threshold level that preserves the classical rates for the case of no-signal.

Now let us consider a different set of thresholds suggested by Cavalier and Tsybakov (2001). The authors were interested in the study of blocks and estimates that imply the parametric rate of convergence. They used an oracle inequality (that is, an upper bound that includes, as an additive term, mean integrated squared error of an oracle) to get a sufficient condition for attaining this rate. In particular, they recommended thresholds

$$\mu^*_{j,m} = C^*[\ln(L_{j,m})/L_{j,m}]^{1/2}. \tag{3.9}$$

The interested reader is referred to that interesting article for details; in particular, Remark 2 and the discussion on p. 253 of the article are of a special interest.

The recommendation (3.9) is based on the analysis of a specific sum with terms including exponential factors $\exp\{-L_{j,m}\mu_{j,m}^2/4(1+\mu_{j,m})\}$. Let us show that the sum in our lower bound has similar exponential factors. Keeping in mind that these factors are the main "players" in the analysis of blocks and thresholds, the lower bound supports the Cavalier-Tsybakov oracle inequality.

Recall that the lower bound is

$$
\begin{aligned}
&E\int_0^1 (\bar{f}_S(x) - f_0(x))^2 dx \\
&\quad \geq Cn^{-1}v \sum_{j=0}^{\log_2(n/\ln(n))} \sum_m \frac{\gamma_{j,m}^2 L_{j,m}^{1/2}}{(1+\gamma_{j,m}+\mu_{j,m})^2} \\
&\quad\quad \times \exp\{-[\gamma_{j,m}+\mu_{j,m} - \ln(1+\gamma_{j,m}+\mu_{j,m})]L_{j,m}/2\}.
\end{aligned}
$$

To analyze the exponential factors, we use the elementary inequality $-z + \ln(1 + z) + z^2/2 > 0$, $z > 0$ which is easily verified by the fact that at $z = 0$ the left side is zero and its derivative is positive for the considered z. This yields that

$$
\exp\{-[\gamma_{j,m}+\mu_{j,m} - \ln(1+\gamma_{j,m}+\mu_{j,m})]L_{j,m}/2\} \geq \exp\{-(\gamma_{j,m}+\mu_{j,m})^2 L_{j,m}/4\}.
$$

Set $\gamma_{j,m} = o(1)\mu_{j,m}$ where $o(1) \to 0$ as $n \to \infty$. Then comparison of the last exponential term with the Cavalier-Tsybakov exponential factor establishes the wished similarity.

Finally, let us present an example of how the lower bounds may help the statistician to analyze a familiar conjecture. Cavalier and Tsybakov ((2001), p. 269) conjecture that because the recommended threshold levels (3.9) are typically small, the difference between Stein(0) and Stein(>0) estimates is not very strong. Sure enough, Tsybakov (2002) proved that Stein(0) estimate can be asymptotically minimax. On the other hand, as we have seen, Stein(0) is dramatically less flexible in terms of a possible pool of blocks necessary for a minimax estimation.

4. Numerical study

Let us complement the discussion of Section 3 by a numerical study of the James-Stein estimate.

It has been explained in Subsection 3.1 that this estimate requires employing of very large blocks for attaining the classical asymptotic minimax rates. Is this conclusion crucial for relatively small datasets considered in the literature? What will be if we employ traditionally recommended identical blocks of small sizes? To answer these questions, consider a numerical experiment shown in Fig. 2. Here the S-PLUS supported regression function "GAUSS", that has perfect "no-signal" tails and a smooth bell-shaped part, is restored by James-Stein estimates with identical blocks of lengths $L = 3, 4, 8$ and, for comparison, by the default universal soft threshold estimate supported by S-PLUS. In this and all other experiments, default S-PLUS parameters (including periodized wavelets) are used. The wavelet function is Symmlet-8. The signal-to-noise ratio is denoted as "snr" and it together with the sample size n is exhibited in the titles.

We see that the result of this particular simulation supports the theory: James-Stein shrinkage performs better with larger blocks. Note that the universal thresholding, which satisfies (3.9), implies a dramatically better estimation. On the other hand, even this

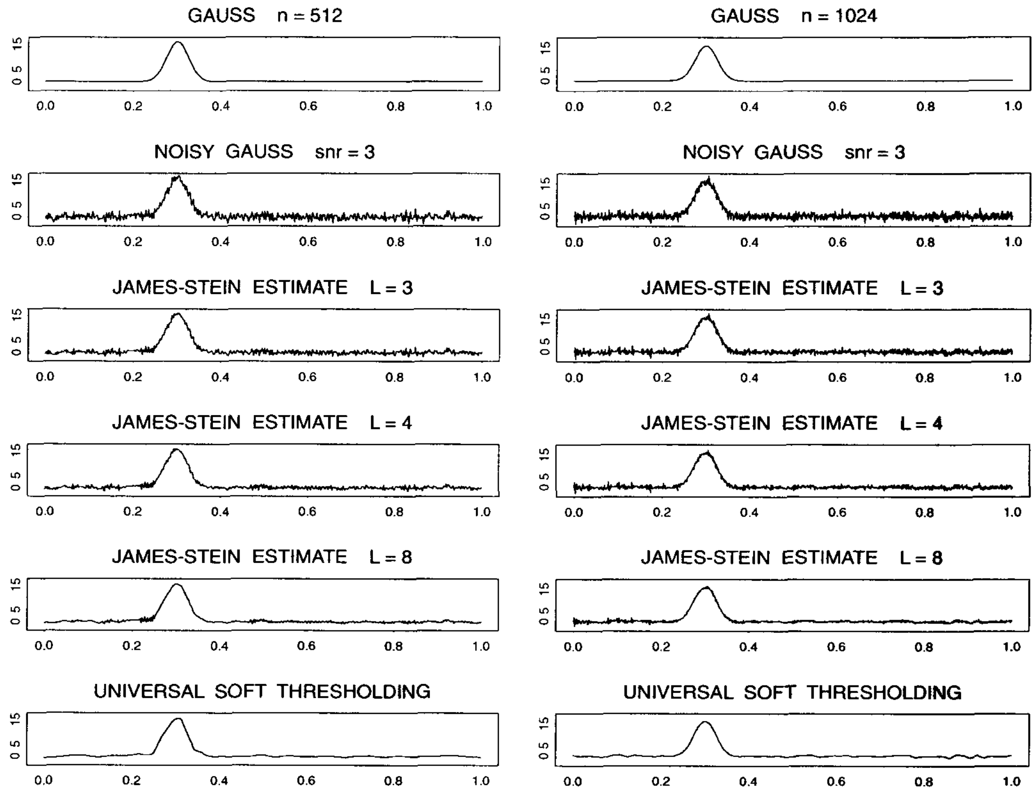

Fig. 2. Performance of the James-Stein estimate with different identical blocks of length L.

estimate cannot perfectly restore the flat tails. This fact explains why estimation of the zero (or any constant) function is not a trivial problem and why the no-signal function f_0 is of the special interest.

The conclusion made is drawn from the single experiment. What will be if this simulation is repeated 300 times? The results are summarized in Table 1 where simulations for two other sample sizes are also presented. The "ARMISE" row shows average ratios of the square root of MISE of a James-Stein estimate to the corresponding value of the universal estimate. The "ARMSE1" raw shows average ratios of the absolute deviation of a James-Stein estimate from the underlying function at point $x_1 = .3$ (the peak of "GAUSS") to the corresponding value of the universal estimate. The "ARMSE2" raw shows average ratios calculated at point $x_2 = .5$ (the flat part of "GAUSS").

As we see, the intensive numerical study supports the conclusions made from the analysis of Fig. 2. Table 1 also sheds additional light on the James-Stein shrinkage estimate. First of all, the necessity of using large blocks becomes more urgent as the sample size increases. Secondly, for the smaller samples the relative quality of estimation (James-Stein versus the universal soft thresholding) at the point $x_2 = .5$, where the "GAUSS" function is flat, is worse than at the peak point $x_1 = .3$. Returning to Fig. 2, we may note that this property could be forecasted.

Table 1. Summary of results of numerical experiments based on Fig. 2.

n	256	256	256	512	512	512	1024	1024	1024	2048	2048	2048
L	3	4	8	3	4	8	3	4	8	3	4	8
$ARMISE$	1.73	1.46	1.13	2.28	1.89	1.36	2.81	2.27	1.52	2.81	2.23	1.41
$ARMSE1$	0.55	0.58	0.61	1.49	1.29	1.04	4.39	3.15	1.94	5.11	3.93	2.12
$ARMSE2$	4.60	3.40	1.76	5.01	3.71	1.89	2.63	1.95	1.59	4.65	3.54	1.90

5. Conclusion

This article suggests to complement the traditional minimax and Bayesian analysis of blockwise shrinkage estimates by exploring exact (not-asymptotic) lower bounds for risks calculated for the case of a no-signal setting. Because this case is included in all classical minimax settings, the lower bounds also imply necessary conditions for an estimate to be minimax. Moreover, if a procedure fails to perform well for the no-signal setting, this procedure should raise eyebrows. Interestingly, this simple and intuitively appealing theoretical tool yields a wealth of information about blockwise threshold procedures. In particular, it proves that Stein(0) and James-Stein estimates require geometrically increasing blocks to attain classical minimax rates, it sheds a new light on the famous Cai (1999)'s optimal threshold level for logarithmic blocks, and it supports several familiar oracle inequalities (upper bounds). Due to the not-asymptotic nature of the lower bounds, the suggested approach also complements methods of the numerical analysis. The developed methodology can be recommended for the analysis of a wide spectrum of adaptive nonparametric procedures.

Appendix: Proofs

PROOF OF THEOREM 2.1. For the considered f_0 (or f_0 defined in Remark 2.2) we can write

$$\begin{aligned} E(\hat{f}(x_0) - \hat{f}^*(x_0))^2 &= E(\hat{f}(x_0) - f_0(x_0))^2 \\ &= E\left(\sum_j \sum_k \hat{\theta}_{j,k}\psi_{j,k}(x_0)\right)^2 \\ &= n^{-1}vE\left(\sum_j \sum_m \sum_{k\in T_{j,m}} \frac{\sum_{s\in T_{j,m}}(\xi_{j,s}^2 - 1)}{\sum_{s\in T_{j,m}}\xi_{j,s}^2}\right. \\ &\qquad \left. \times I\left(\sum_{s\in T_{j,m}} \xi_{j,s}^2 > (1+\lambda_{j,m}^2)L_{j,m}\right)\psi_{j,k}(x_0)\xi_{j,k}\right)^2. \end{aligned}$$

Because $\xi_{j,k}$ are independent and symmetrically distributed about zero,

$$\begin{aligned} &E(\hat{f}(x_0) - \hat{f}^*(x_0))^2 \\ &= n^{-1}vE\left\{\sum_j \sum_m \sum_{k\in T_{j,m}} \frac{[\sum_{s\in T_{j,m}}(\xi_{j,s}^2 - 1)]^2}{[\sum_{s\in T_{j,m}}\xi_{j,s}^2]^2}\right. \end{aligned} \tag{A.1}$$

$$\times I\left(\sum_{s\in T_{j,m}} \xi_{j,s}^2 > (1+\lambda_{j,m}^2)L_{j,m}\right)\psi_{j,k}^2(x_0)\xi_{j,k}^2\Bigg\}.$$

The fact that $\xi_{j,s}$ are independent and identically distributed implies that for $k \in T_{j,m}$

$$E\left\{\frac{[\sum_{s\in T_{j,m}}(\xi_{j,s}^2-1)]^2}{[\sum_{s\in T_{j,m}}\xi_{j,s}^2]^2} I\left(\sum_{s\in T_{j,m}} \xi_{j,s}^2 > (1+\lambda_{j,m}^2)L_{j,m}\right)\xi_{j,k}^2\right\} \tag{A.2}$$
$$= L_{j,m}^{-1} E\left\{\frac{[\sum_{s\in T_{j,m}}(\xi_{j,s}^2-1)]^2}{\sum_{s\in T_{j,m}}\xi_{j,s}^2} I\left(\sum_{s\in T_{j,m}} \xi_{j,s}^2 > (1+\lambda_{j,m}^2)L_{j,m}\right)\right\}.$$

Let us estimate the right side of (A.2). Note that $\sum_{k\in T_{j,s}}\xi_{j,s}^2$ has central chi-squared distribution with $L_{j,m}$ degrees of freedom. Denote by χ_L^2 a central chi-squared random variable with L degrees of freedom and write

$$E\{(\chi_L^2 - L)^2(\chi_L^2)^{-1}I(\chi_L^2 > (1+\lambda^2)L)\}$$
$$= E\{[\chi_L^2 - 2L + L^2/\chi_L^2]I(\chi_L^2 > (1+\lambda^2)L)\}.$$

Recall that χ_L^2 has the density

$$p_L(y) = \frac{1}{2^{L/2}\Gamma(L/2)} y^{L/2-1}e^{-y/2}, \qquad y > 0,$$

denote $a := (1+\lambda^2)L$ and write using integration by parts

$$\begin{aligned} &E\{\chi_L^2 I(\chi_L^2 > a)\} \\ &= \int_a^\infty y p_L(y)dy \\ &= [2^{L/2}\Gamma(L/2)]^{-1}\int_a^\infty y^{L/2}e^{-y/2}dy \\ &= [2^{L/2}\Gamma(L/2)]^{-1}\left[2a^{L/2}e^{-a/2} + 2(L/2)\int_a^\infty y^{L/2-1}e^{-y/2}dy\right] \\ &= \frac{a^{L/2}e^{-a/2}}{2^{L/2-1}\Gamma(L/2)} + LP(\chi_L^2 > a). \end{aligned} \tag{A.3}$$

Similarly

$$\begin{aligned} LP(\chi_L^2 > a) &= L[2^{L/2}\Gamma(L/2)]^{-1} \\ &\quad \times \left[2a^{L/2-1}e^{-a/2} + 2(L/2-1)\int_a^\infty y^{-1}y^{L/2-1}e^{-y/2}dy\right] \\ &= La^{L/2-1}e^{-a/2}[2^{L/2-1}\Gamma(L/2)]^{-1} \\ &\quad + (L^2-2L)E\{(\chi_L^2)^{-1}I(\chi_L^2 > a)\}. \end{aligned} \tag{A.4}$$

These two relations imply

$$E\{[\chi_L^2 - 2L + L^2/\chi_L^2]I(\chi_L^2 > a)\}$$
$$= a^{L/2}e^{-a/2}[2^{L/2-1}\Gamma(L/2)]^{-1}(1 - L/a) + 2LE\{(\chi_L^2)^{-1}I(\chi_L^2 > a)\}.$$

Note that $1 - L/a = \lambda^2/(1+\lambda^2)$, and then integration by parts yields

$$2LE\{(\chi_L^2)^{-1}I(\chi_L^2 > a)\} \geq 4L[2^{L/2}\Gamma(L/2)]^{-1}a^{L/2-2}e^{-a/2}. \tag{A.5}$$

Combining the results we get

$$\begin{aligned} &E\{(\hat{f}(x_0) - f(x_0))^2\} \\ &\geq n^{-1}v\sum_j\sum_m \frac{(1+\lambda_{j,m}^2)^{L_{j,m}/2}L_j^{L_{j,m}/2}e^{-(1+\lambda_{j,m}^2)L_{j,m}/2}}{L_{j,m}2^{L_{j,m}/2-1}\Gamma(L_{j,m}/2)} \\ &\times\left[\frac{\lambda_{j,m}^2}{1+\lambda_{j,m}^2} + \frac{2}{(1+\lambda_{j,m}^2)^2L_{j,m}}\right]\sum_{k\in T_{j,m}}\psi_{j,k}^2(x_0). \end{aligned} \tag{A.6}$$

According to Robbins (1955),

$$0 < c_1 < \frac{\Gamma(L/2)}{(L/2)^{L/2-1/2}e^{-L/2}} < c_2 < \infty. \tag{A.7}$$

Then a simple calculation implies

$$\begin{aligned} &\frac{(1+\lambda^2)^{L/2}L^{L/2}e^{-(1+\lambda^2)L/2}}{L2^{L/2-1}\Gamma(L/2)}\left[\frac{\lambda^2}{1+\lambda^2} + \frac{2}{(1+\lambda^2)^2L}\right] \\ &\quad\geq C[\lambda_j^2(1+\lambda_j^2)^{-1} + (1+\lambda^2)^{-2}L^{-1}][L^{-1/2}e^{(L/2)(\ln(1+\lambda^2)-\lambda^2)}]. \end{aligned}$$

This together with (A.6) and $\psi_{j,s}^2(x_0) = 2^j\psi^2(2^jx_0 - s)$ proves Theorem 2.1.

PROOF OF COROLLARY 2.1. Note that under the assumption about the wavelet function we have

$$\sum_m\sum_{k\in T_{j,m}}\psi^2(2^jx_0 - k) = \Psi(j, x_0) \geq \min_x\Psi(j,x) = c^* > 0.$$

Using this relation together with (2.2) implies

$$\begin{aligned} &E(\hat{f}(x_0) - f_0(x_0))^2 \\ &\quad > Cn^{-1}v\sum_{j=j_0}^{\log_2(n/\ln(n))}\left[\frac{\lambda_j^2}{1+\lambda_j^2} + \frac{1}{(1+\lambda_j^2)^2L_j}\right] \\ &\quad\times\frac{1}{L_j^{1/2}}e^{j\ln(2)-[\lambda_j^2-\ln(1+\lambda_j^2)]L_j/2}\sum_m\sum_{k\in T_{j,m}}\psi^2(2^jx_0-k) \\ &\quad > Cn^{-1}v\sum_{j=j_0}^{\log_2(n/\ln(n))}\left[\frac{\lambda_j^2}{1+\lambda_j^2} + \frac{1}{(1+\lambda_j^2)^2L_j}\right]\frac{1}{L_j^{1/2}}e^{j\ln(2)-[\lambda_j^2-\ln(1+\lambda_j^2)]L_j/2}. \end{aligned}$$

Corollary 2.1 is verified.

PROOF OF COROLLARY 2.2. Note that the generic constant C in (2.2) does not depend on x_0. Then verification of the lower bounds is based on taking integrals on the right and left sides of (2.2) and using the fact that $2^{j/2}\psi(2^j x - k)$ are elements of an orthonormal basis on $[0,1]$. This yields (2.5). Then there are 2^j wavelet coefficients at the j-th scale, this together with (2.2) yields (2.6). Corollary 2.2 is proved.

PROOF OF THEOREM 2.2. For the underlying regression function $f_0(x)$ the James-Stein shrinkage does not decrease mean squared error of EP shrinkage with the zero threshold $\lambda^2 = 0$. Indeed, in this case

$$E\left\{\left(\frac{\hat{\Theta} + (2/L)vn^{-1}}{\hat{\Theta} + vn^{-1}}\right)_+ \tilde{\theta}\right\}^2 \geq E\left\{\left(\frac{\hat{\Theta}}{\hat{\Theta} + vn^{-1}}\right)_+ \tilde{\theta}\right\}^2$$

where $\hat{\Theta} := L^{-1}\sum_{s\in T}\tilde{\theta}_s^2 - vn^{-1}$ and T is a particular block. This together with Theorem 2.1 yields the verified assertion.

PROOF OF THEOREM 2.3. Similarly to (A.1)–(A.2) we can write

$$\begin{aligned} &E(\bar{f}_S(x_0) - f_0(x_0))^2 \\ &\quad = n^{-1}v \sum_{j=j_0}^{\ln(n/\ln(n))} \sum_m L_{j,m}^{-1} E\left\{\frac{[\chi^2_{L_{j,m}} - (1+\mu_{j,m})L_{j,m}]^2}{\chi^2_{L_{j,m}}}\right. \\ &\qquad\qquad \left. \times I(\chi^2_{L_{j,m}} > (1+\mu_{j,m})L_{j,m})\right\} \sum_{k\in T_{j,m}} \psi^2_{j,k}(x_0). \end{aligned}$$

For a nonnegative γ using (A.5) and (A.7) implies

$$\begin{aligned} &L^{-1}E\left\{\frac{[\chi^2_L - (1+\mu)L]^2}{\chi^2_L} I(\chi^2_L > (1+\mu)L)\right\} \\ &\quad \geq L^{-1}E\left\{\frac{(\gamma L)^2}{\chi^2_L} I(\chi^2_L > (1+\gamma+\mu)L)\right\} \\ &\quad \geq C\gamma^2 L \frac{2^{L/2}[(1+\gamma+\mu)L]^{L/2-2}e^{-(1+\gamma+\mu)L/2}}{2^{L/2}L^{L/2-1/2}e^{-L/2}} \\ &\quad = C\frac{\gamma^2 L^{-1/2}}{(1+\gamma+\mu)^2} e^{-[\gamma+\mu-\ln(1+\gamma+\mu)]L/2}. \end{aligned}$$

Combining the results we get that for any sequence $\gamma_{j,m} \geq 0$

$$\begin{aligned} &E(\bar{f}_S(x_0) - f_0(x_0))^2 \\ &\quad \geq Cn^{-1}v \sum_{j=j_0}^{\ln(n/\ln(n))} \sum_m \frac{\gamma^2_{j,m} L_{j,m}^{-1/2}}{(1+\gamma_{j,m}+\mu_{j,m})^2} \\ &\quad\quad \times e^{j\ln(2) - [\gamma_{j,m}+\mu_{j,m} - \ln(1+\gamma_{j,m}+\mu_{j,m})]L_{j,m}/2} \sum_{k\in T_{j,m}} \psi^2(2^j x_0 - k). \end{aligned}$$

This inequality verifies (2.10). The proof of (2.11) and (2.12) is identical to the above-presented proof of Corollaries 2.1–2.2. Theorem 2.3 is verified.

References

Abramovich, F., Besbeas, P. and Sapatinas, T. (2002). Empirical Bayes approach to block wavelet function estimation, *Computational Statistics and Data Analysis*, **39**, 435–451.

Brown, L. D., Low, M. G. and Zhao, L. H. (1997). Superefficiency in nonparametric function estimation, *The Annals of Statistics*, **25**, 2607–2625.

Cai, T. (1999). Adaptive wavelet estimation: a block thresholding and oracle inequality approach, *The Annals of Statistics*, **27**, 898–924.

Cai, T. (2002). On block thresholding in wavelet regression: Adaptivity, block size, and threshold level, *Statistical Sinica*, **12**, 1321–1273.

Cai, T. and Silverman, B. (2000). Incorporating information on neighboring coefficients into wavelet estimation, *Sankhyā Series B*, **63**, 127–148.

Cai, T., Low, M. and Zhao, L. (2000). Sharp adaptive estimation by a blockwise method, Tech. Report, University of Pennsylvania, Philadelphia.

Cavalier, L. and Tsybakov, A. (2001). Penalized blockwise Stein's method, monotone oracles and sharp adaptive estimation, *Mathematical Methods of Statistics*, **10**, 247–282.

Chiken, E. (2003). Block thresholding and wavelet estimation for nonequispaced samples, *Journal of Statistical Planning and Inference*, **16**, 113–129.

DeCanditiis, D. and Vidakovic, B. (2004). Wavelet Bayesian block shrinkage via mixture of normal-inverse gamma priors, *Journal of Computational and Graphical Statistics*, **12** (to be published).

Delyon, B. and Juditsky, A. (1996). On minimax wavelet estimators, *Journal of Applied and Computational Harmonic Analysis*, **3**, 215–228.

Donoho, D. L. (1995). De-Noising by soft thresholding, *IEEE Transactions of Information Theory*, **41**, 613–627.

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

Donoho, D. L. and Johnstone, I. M. (1995). Adapting to unknown smoothness via wavelet shrinkage, *Journal of American Statistical Association*, **90**, 1200–1224.

Downie, T. R. and Silverman, B. M. (1998). The discrete multiple discrete transform and threshold methods, *IEEE Transactions on Signal Processing*, **46**, 2558–2561.

Efromovich, S. (1985). Nonparametric estimation of a density with unknown smoothness, *Theory of Probability and Applications*, **30**, 557–568.

Efromovich, S. (1986). Adaptive algorithm for nonparametric regression, *Procedures of Second IFAC Symposium on Stochastic Control*, 112–114, Science, Vilnuis.

Efromovich, S. (1999). *Nonparametric Curve Estimation: Methods, Theory and Applications*, Springer, New York.

Efromovich, S. (2000). Can adaptive estimators for Fourier series be of interest to wavelets?, *Bernoulli*, **6**, 699–709.

Efromovich, S. (2001). Multiwavelets and signal denoising, *Sankhyā Series B*, **23**, 367–393.

Efromovich, S. and Pinsker, M. (1984). A learning algorithm for nonparametric filtering, *Automation and Remote Control*, **24**, 1434–1440.

Efromovich, S. and Pinsker, M. (1986). Adaptive algorithm of minimax nonparametric estimating spectral density, *Problems of Information Transmission*, **22**, 62–76.

Efromovich, S. and Pinsker, M. (1996). Sharp-optimal and adaptive estimation for heteroscedastic nonparametric regression, *Statistica Sinica*, **6**, 925–945.

Hall, P., Kerkyacharian, G. and Picard, D. (1998). Block threshold rules for curve estimation using kernel and wavelet methods, *The Annals of Statistics*, **26**, 922–942.

Hall, P., Kerkyacharian, G. and Picard, D. (1999). On the minimax optimality of block thresholded wavelet estimators, *Statistica Sinica*, **9**, 33–50.

Härdle, W., Kerkyacharian, G., Picard, D. and Tsybakov, A. (1998). *Wavelets, Approximation, and Statistical Applications*, Springer, New York.

Johnstone, I. (1998). *Function Estimation in Gaussian Noise. Sequence Models*, Draft of Monograph, University of Stanford, California.

Juditsky, A. (1997). Wavelet estimators: Adapting to unknown smoothness, *Mathematical Methods of Statistics*, **6**, 1–25.

Lehmann, E. L. and Casella, G. (1998). *Theory of Point Estimation*, 2nd ed., Springer, New York.

Mallat, S. (1999). *A Wavelet Tour of Signal Processing*, University Press, New York.

Nemirovski, A. (2000). *Topics in Non-Parametric Statistics*, Lecture Notes in Mathematics, **1738**, Springer, New York.

Robbins, H. (1955). A remark on Stirling's formula, *American Mathematics Monthly*, **62**, 26–29.

Tsybakov, A. (2002). Discussion on random rates in anisotropic regression, *The Annals of Statistics*, **30**, 379–385.

Vidakovic, B. (1999). *Statistical Modeling by Wavelets*, Springer, New York.

Wahba, G. (1990). *Spline Models for Observational Data*, SIAM, Philadelphia.

Zhang, C.-H. (2002). Minimax compound estimation, Tech. Report, Rutgers University, New Jersey.