Sharp Adaptive Estimation by a Blockwise Method

T. Tony Cai, Mark G. Low and Linda H. Zhao

Department of Statistics

University of Pennsylvania

Abstract

We consider a blockwise James-Stein estimator for nonparametric function estimation in suitable wavelet or Fourier bases. The estimator can be readily explained and implemented. We show that the estimator is asymptotically sharp-adaptive in minimax risk over any Sobolev ball containing the true function. Further, for a moderately broad range of bounded sets in Besov space our estimator is asymptotically nearly sharp adaptive in the sense that it comes within the Donoho-Liu constant, 1.24, of being exactly sharp adaptive. Other parameter spaces are also considered. The paper concludes with a Monte-Carlo study comparing the performance of our estimator to that of three other popular wavelet estimators. Our procedure generally (but not always) outperforms two of these and is overall comparable, or perhaps slightly superior, to the third.

Keywords: Adaptive estimation; Besov space; Empirical Bayes; Fourier series; Minimax; Nonparametric regression; Sobolev space; Wavelets; White noise.

AMS 1991 Subject Classification: Primary: 62G99; Secondary: 62F12, 62F35, 62M99.

1 Introduction

Various levels of asymptotic adaptivity have become recognized as an important feature of useful nonparametric function estimators. In the present paper we construct an easily understood and implemented adaptive estimator. We investigate the adaptivity of the procedure by using the minimax approach.

Under the minimax criteria, our estimator is sharp-adaptive over Sobolev spaces. This is to say that its risk for estimating an unknown regression function, f, is asymptotically bounded by the minimax risk over any Sobolev ball containing this function. This property can even be shown to hold uniformly over a broad spectrum of Sobolev balls (See (12)).

For a moderately broad range of bounded sets in Besov space our estimator is nearly sharp-adaptive, since it comes within the Donoho-Liu constant, $1.24\ldots$, of being exactly sharp adaptive. The range of Besov spaces over which our adaptivity holds is more restricted than for the adaptivity property of the popular VisuShrink, SureShrink and BlockJS wavelet estimators. (Donoho and Johnstone (1994, 1995); Cai (1999).) On the other hand our estimator is more nearly sharp adaptive over suitable Besov sets and is as easy, or easier, to implement.

Our adaptive estimator can be readily explained, as follows. Express the nonparametric regression problem in terms of an empirical wavelet basis or Fourier basis. Group these coefficients in blocks of slowly growing size. Apply the James-Stein shrinkage estimator within each individual block. The details of this construction are described in steps 1-3 of Section 2.

The idea of block thresholding can be traced back to Efromovich (1985) in orthogonal series estimators; and Kerkyacharian, Picard and Tribouley (1996), for wavelet density estimation. The closest in spirit to the blockwise James-Stein estimator proposed here is the BlockJS procedure mentioned above. A similar construction was earlier used in Brown, Low and Zhao (1997) to produce superefficient estimators. Goldenschluger and Tsybakov (2003) have used a related construction for a nonparametric prediction problem.

The above mentioned wavelet estimators provide a level of adaptivity across a broad range of bounded sets. There are a variety of other adaptive proposals that could be mentioned. Most of these involve a more restricted range of sets. The earliest sharp-adaptive estimators are those in Efromovich and Pinsker (1984). Adaptive estimators built from kernel estimators are described in Lepski, Mammen and Spokoiny (1997). Efromovich

(2000) describes another type of adaptive estimator in Sobolev spaces.

Zhang (2005) contains some broadly applicable results describing the existence of exactly sharp-adaptive estimators across a range of sets that includes the Besov spaces in our result. Our estimator is more easily implemented than his, and we suspect ours may have superior numerical performance for small to moderately large sample sizes.

We conclude this paper with a Monte Carlo examination of the performance of our estimator for sample sizes ranging from 512 to 8192. For four standard test functions, performance of our estimator in terms of integrated squared error is compared with that of three other popular adaptive wavelet procedures. It can be seen from Figure 1 that our procedure generally (but not always) outperforms two of these and is overall comparable, or perhaps slightly superior, to the third.

2 The estimator

We consider the canonical white noise model which is asymptotically equivalent to the conventional formulation of nonparametric regression. See Brown and Low (1996), Brown, Cai, Low, and Zhang (2002) and Brown and Zhao (2000). There is also a slightly less direct equivalence to nonparametric density estimation. See Nussbaum (1996), Klemelä and Nussbaum (1999), and Brown, Carter, Low and Zhang (2004).

Under the white noise model, we observe a stochastic processes $Y_n(t)$ governed by

$$dY_n(t) = f(t)dt + n^{-1/2}dW(t), \ \ 0 \le t \le 1$$

where $W(t)$ is a standard Brownian motion. We wish to estimate the drift function f. The accuracy of an estimator $\hat{f}$ is measured by the mean integrated square error:

$$R(f, \ \hat{f}) = E\|\hat{f} - f\|_2^2.$$

Suppose $\{\beta_i(t)\}_{i=1}^\infty$ is an orthonormal basis of $\mathcal{L}^2[0,1]$. Let $y_i = \int \beta_i(t)dY_n(t)$ and $\theta_i = \int f(t)\beta_i(t)dt$. Then the function estimation problem is exactly equivalent to the following sequence model.

Observe

$$y_i = \theta_i + n^{-1/2}\epsilon_i, \ \epsilon_i \overset{iid}{\sim} N(0,1), \ \ i = 1,\dots \tag{1}$$

and wish to estimate θ under risk

$$R(\theta,\hat{\theta}) = E_\theta\|\hat{\theta} - \theta\|_{\ell^2}^2. \tag{2}$$

An estimator $\hat{\theta}$ of the coefficient sequence θ directly provides an estimator

$$\hat{f}(t) = \sum_{i=1}^{\infty} \hat{\theta}_i \, \beta_i(t)$$

of the function f with an isometry of risk $R(f, \, \hat{f}) = R(\theta, \, \hat{\theta})$.

The estimation problem (1) - (2) is an infinite dimensional normal mean problem. Among the many traditional shrinkage estimators developed in the multivariate normal decision theory, the James-Stein estimator is perhaps the best-known and will be used below. Efron and Morris (1973) showed that the (positive part) James-Stein estimator does more than just demonstrate the inadequacy of the maximum likelihood estimator. It is a member of a class of good shrinkage rules, all of which may be useful in different estimation problems. The James-Stein estimator can also be regarded as truncated empirical Bayes rules. See Efron and Morris (1973).

We propose a blockwise James-Stein estimator for estimating the infinite dimensional normal mean θ. The estimator is constructed as follows:

1. Choose a "base" multiplier, $b = b_n > 1$. Corresponding to b choose an initial block length $I_0 = [b^{K_0}]$ with K_0 an integer chosen so that $I_0 \geq 3$ and $[b^k] - [b^{k-1}] \geq 3$ for all $k \geq K_0 + 1$. Let $K_1 = K_1(n) = [\log_b n] - 1$ and $I_1 = [b^{K_1}]$. Note $I_1 \geq n/b$.

2. Let $G_{K_0} = \{i : 1 \leq i \leq I_0 = [b^{K_0}]\}$, $G_k = \{i : [b^{k-1}] < i \leq [b^k]\}$, $k \geq K_0 + 1$. Let $m_k = [b^k] - [b^{k-1}]$ denote the number of indices in G_k and $m_{K_0} = I_0$. Note $m_k \geq 3$ for $k \geq K_0$. Let $\|Y\|_{(k)}^2 = \sum_{i \in G_k} y_i^2$, and $\|\theta\|_{(k)}^2 = \sum_{i \in G_k} \theta_i^2$.

3. Define the corresponding estimator $\tilde{\theta}$ by

$$\tilde{\theta}_i = \begin{cases} (1 - \frac{m_k - 2}{n\|Y\|_{(k)}^2})_+ y_i & \text{if } i \in G_k \text{ for some } k = K_0, \dots, K_1(n) \\ 0 & \text{if } i > I_1. \end{cases} \tag{3}$$

For our main results, we choose the base multiplier $b = b_n = 1 + 1/\log n$ and denote the corresponding estimator of θ by $\tilde{\theta}^*$ and the corresponding estimator of f by $\tilde{f}^*$. Other choices of b are discussed in Sections 4 and 5.

3 Sharp adaptivity over Sobolev classes

We now investigate the minimaxity of the blockwise James-Stein estimator $\tilde{f}^*$. For our first basic result we consider the Sobolev function class $\mathcal{F}$ which consists of periodic functions

defined by the Sobolev condition

$$\mathcal{F}(\alpha, M) = \{f : \ \|f\|_2^2 + \|f^{(\alpha)}\|_2^2 \le M.\} \tag{4}$$

Choose $\phi_1 = 1, \phi_{2i} = \sqrt{2}\cos(2\pi it)$ and $\phi_{2i+1}(t) = \sqrt{2}\sin(2\pi it)$. Then the problem of estimating f under $\mathcal{L}_2$ loss is exactly equivalent to the sequence model with $\theta \in \Theta$ where Θ is defined as

$$\Theta = \Theta(\alpha, B) = \{\theta : \sum_{i=1}^{\infty} c_i \theta_i^2 \le B\} \tag{5}$$

with $c_{2i} = 1 + \pi^{2\alpha}(2i)^{2\alpha} = c_{2i+1}$ and $M = B$. Efromovich and Pinsker(1982) and Efromovich (1998) show the minimax risk over $\Theta(\alpha, B)$ (or $\mathcal{F}(\alpha, M)$) with

$$c_i = Ci^{2\alpha}(1 + o(1)) \quad \text{as } i \to \infty \tag{6}$$

is

$$R_*(\alpha, B) = \inf_{\hat{\theta}} \sup_{\theta \in \Theta(\alpha,B)} E\|\hat{\theta} - \theta\|^2 = C^*(\alpha)B^{\frac{1}{1+2\alpha}} C^{-\frac{1}{1+2\alpha}} n^{-\frac{2\alpha}{1+2\alpha}}(1 + o(1)) \tag{7}$$

where $C^*(\alpha)$ is a constant depending on α; see Pinsker (1980) for the explicit formula for $C^*(\alpha)$.

They also show this same expression yields the asymptotic linear minimax risk (i.e. the value for the expression on the left of (7) when the procedures $\hat{\theta}$ are restricted to be linear). If the periodicity restriction on f in (4) is dropped one needs to use a modification of the standard Fourier basis, above. With the modified basis (5), (6) and (7) are still valid.

Theorem 1 *The estimator $\tilde{\theta}^*$ satisfies*

$$\lim_{n\to\infty} \frac{\sup_{\theta \in \Theta(\alpha,B)} R(\theta, \tilde{\theta}^*)}{R_*(\alpha, B)} = 1 \tag{8}$$

for all $\alpha > 0, B > 0$.

Before we prove Theorem 1 we state a standard risk bound for the James-Stein estimator (see, e.g., Donoho and Johnstone (1995)). Let $Z_i \overset{ind}{\sim} N(w_i, \sigma^2)$, $i = 1, \ldots, m \ge 3$, with σ^2 known. Let $\tilde{w}_i = (1 - (m-2)\sigma^2/s^2)_+ Z_i$, where $s^2 = \sum_{i=1}^m Z_i^2$. Then

$$E_w(\|\tilde{w} - w\|^2) \le \frac{\|w\|^2(m-2)\sigma^2}{\|w\|^2 + (m-2)\sigma^2} + 2\sigma^2 \le \frac{\|w\|^2 m\sigma^2}{\|w\|^2 + m\sigma^2} + 2\sigma^2. \tag{9}$$

Proof of Theorem 1 : Fix $\alpha > 0, B > 0$. For convenience we take $C = 1$ and $c_i = i^{2\alpha}$. Only minor modifications are needed under (6). As noted R_* is attained by linear estimators.

Further, Donoho, Liu and MacGibbon (1990) show it suffices to consider linear estimators of the form $\hat{\theta}_i = d_{ni}Y_i$, where each multiplier, d_{ni} depends on n (and α, B). We have

$$
\begin{aligned}
R_*(\alpha, B) &\geq \{1 + o(1)\} \inf_{d_{ni}} \sup_{\theta \in \Theta(\alpha, B)} \sum_{i=1}^{\infty} E(d_{ni}Y_i - \theta_i)^2 \\
&\geq \{1 + o(1)\} \sup_{\sum_{K_0+1}^{K_1} b^{2k\alpha}\|\theta\|_{(k)}^2 \leq B} \inf_{d_{ni}} \sum_{K_0+1}^{K_1} \sum_{i \in G_k} E(d_{ni}Y_i - \theta_i)^2.
\end{aligned}
$$

Further, it is easy to verify that

$$
\sup_{\|\theta_k\|^2 \leq t_k^2} \inf_{d_{ni}} \sum_{i \in G_k} E(d_{ni}y_i - \theta_i)^2 = \frac{t_k^2 m_k n^{-1}}{t_k^2 + m_k n^{-1}}.
$$

Thus

$$
R_*(\alpha, B) \geq \{1 + o(1)\} \sup_{\sum_{K_0+1}^{K_1} b^{2k\alpha}t_k^2 \leq B} \sum_{K_0+1}^{K_1} \frac{t_k^2 m_k n^{-1}}{t_k^2 + m_k n^{-1}}, \tag{10}
$$

since $\{\theta : \sum_{K_0+1}^{K_1} b^{2k\alpha}\|\theta\|_{(k)}^2 \leq B,\ \theta_i = 0 \text{ for } i \notin (I_0, I_1]\} \subset \Theta(\alpha, B)$. On the other hand

$$
\begin{aligned}
R(\theta, \tilde{\theta}) &= \sum_{k=K_0}^{\infty} \sum_{i \in G_k} E\|\tilde{\theta} - \theta\|_{(k)}^2 \leq m_{K_0} n^{-1} + \sum_{K_0+1}^{K_1} \frac{\|\theta\|_{(k)}^2 m_k n^{-1}}{\|\theta\|_{(k)}^2 + m_k n^{-1}} + 2K_1 n^{-1} + \sum_{i=I_1+1}^{\infty} \theta_i^2 \\
&= \sum_{K_0+1}^{K_1} \frac{\|\theta\|_{(k)}^2 m_k n^{-1}}{\|\theta\|_{(k)}^2 + m_k n^{-1}} + o(n^{-\frac{2\alpha}{1+2\alpha}}),
\end{aligned}
$$

since $K_1 = O(\log n)$, m_{K_0} bounded, and

$$
\sum_{i=I_1+1}^{\infty} \theta_i^2 \leq \frac{1}{(I_1 + 1)^{2\alpha}} \sum_{i}^{\infty} i^{2\alpha}\theta_i^2 = O(n^{-2\alpha}) = o(n^{-\frac{2\alpha}{1+2\alpha}}).
$$

Thus

$$
\sup_{\theta \in \Theta(\alpha, B)} R(\theta, \tilde{\theta}) \leq \sup_{\sum_{K_0+1}^{K_1} b^{2(k-1)\alpha}t_k^2 \leq B} \sum_{K_0+1}^{K_1} \frac{t_k^2 m_k n^{-1}}{t_k^2 + m_k n^{-1}} + o(n^{-\frac{2\alpha}{1+2\alpha}}), \tag{11}
$$

since $\{\theta : \sum_{K_0+1}^{K_1} b^{2(k-1)\alpha}\|\theta\|_{(k)}^2 \leq B\} \supset \Theta(\alpha, B)$. The theorem follows from comparison of (10) and (11). More precisely, by using the Lagrange multiplier, it is not difficult to show that

$$
\sup_{\sum_{K_0+1}^{K_1} b^{2k\alpha}t_k^2 \leq B} \sum_{K_0+1}^{K_1} \frac{m_k t_k^2 n^{-1}}{t_k^2 + m_k n^{-1}} + o(n^{-\frac{2\alpha}{1+2\alpha}}) \sim C^*(\alpha) B^{\frac{1}{1+2\alpha}} n^{-\frac{2\alpha}{1+2\alpha}},
$$

where $C^*(\alpha)$ is the minimax constant. (See also Efromovich (1998)). Hence

$$
\lim_{n \to \infty} \frac{\sup_{\theta \in \Theta(\alpha, B)} R(\theta, \tilde{\theta})}{R_*(\alpha, B)} \leq \lim_{n \to \infty} \frac{C^*(\alpha)(b_n^{2\alpha}B)^{\frac{1}{1+2\alpha}} n^{\frac{2\alpha}{1+2\alpha}}}{C^*(\alpha)B^{\frac{1}{1+2\alpha}} n^{\frac{2\alpha}{1+2\alpha}}} = 1. \quad \blacksquare
$$

The expression (8) shows in a strong sense that $\tilde{\theta}^*$ is adaptively asymptotically minimax. Under slightly stronger assumptions an even stronger conclusion is possible. Take $C = 1$ in (6), without loss of generality, and assume the $o(1)$ term in (6) is uniformly bounded over the allowable range of α, B. Restrict that range to $S = \{\alpha, B : 0 < \alpha_0 \leq \alpha \leq \alpha_0^{-1}, 0 < B_0 \leq B \leq B_0^{-1}\}$. Then in place of (8) one can write

$$\lim_{n \to \infty} \sup_{\alpha, B \in S} \frac{\sup_{\theta \in \Theta(\alpha, B)} R(\theta, \tilde{\theta})}{R_*(\alpha, B)} = 1 \tag{12}$$

We omit the details of the proof, much of which closely follows that of Theorem 1.

Remark: If the blockwise James-Stein estimators $\tilde{\theta}$ uses a constant base multiplier b, then the estimator satisfies

$$\varlimsup_{n \to \infty} \frac{\sup_{\theta \in \Theta(\alpha, B)} R(\theta, \tilde{\theta})}{R_*(\alpha, B)} \leq b^{\frac{2\alpha}{1+2\alpha}}.$$

As mentioned in the introduction a blockwise James-Stein estimator, called BlockJS, was proposed in the wavelet context by Cai (1999). This estimator is closely related to the one developed in (3). There are two major differences between the two procedures. One is that the block size in BlockJS is fixed at $\log n$ instead of weakly geometrically growing as in (3), and another is that the constant $\lambda_k \equiv m_k - 2$ in (3) is replaced by $\lambda = 4.505\ldots$ (solution to the equation $\lambda - \log \lambda - 3 = 0$). The BlockJS estimator was proven in Cai (1999) to have excellent asymptotic adaptivity. Although BlockJS enjoys many desirable properties, it is not asymptotically sharp adaptive over Sobolev balls. That is, Theorem 1 does not hold for the BlockJS estimator.

4   Wavelet estimation

Wavelet bases and the associated Besov space have a slightly different structure. There (1) is

$$y_{jk} = \theta_{jk} + n^{-1/2}\epsilon_{jk}, k = 1, \ldots, 2^j; \ j \geq j_0; \ \epsilon_{jk} \overset{iid}{\sim} N(0, 1). \tag{13}$$

As mentioned earlier, The BlockJS estimator enjoys excellent asymptotic properties over a very wide range of Besov spaces. Here we restrict attention to adaptivity over a somewhat restricted range of Besov spaces. Of interest are those spaces that are quadratically convex. These are the spaces $B_{p,q}^\alpha$ with $p \geq 2, q \geq 2$. See Donoho and Johnstone (1998, p899-900).

Consider Y_{jk} as in (13) with

$$\Theta = B_{p,q}^\alpha(B) = \{\theta_{jk} : \sum_{j=j_0}^\infty (2^{j(\alpha+1/2-1/p)} (\sum_{k=1}^{2^j} |\theta_{jk}|^p)^{1/p})^q \le B\}.$$

Let $R_{*W}(\alpha, p, q, B)$ denote the corresponding minimax risk

$$R_{*W}(\alpha, p, q, B) = \inf_{\hat\theta} \sup_{\theta \in B_{p,q}^\alpha(B)} R(\theta, \hat\theta).$$

Donoho and Johnstone (1998) calculate that the minimax rate is $n^{-2\alpha/(2\alpha+1)}$, i.e.,

$$R_{*W}(\alpha, p, q, B) \asymp n^{-2\alpha/(2\alpha+1)} \quad as \quad n \to \infty \tag{14}$$

A wavelet basis also provides a natural blocking strategy. Use the resolution levels as blocks and let $\|y\|_{(j)}^2 = \sum_{k=1}^{2^j} y_{jk}^2$. Define $\tilde\theta^w$

$$\tilde\theta^w = \begin{cases} (1 - \frac{(2^j-2)n^{-1}}{\|y\|_{(j)}^2})_+ y_{jk}, & \text{if } 1 \le j \le \log_2 n \\ 0 & \text{if } j > \log_2 n. \end{cases} \tag{15}$$

The following result shows that this estimator is nearly sharp-adaptive over a range of Besov balls; it comes within the Donoho-Liu constant, $1.24\ldots$, of being exactly sharp adaptive.

Theorem 2 *For $B > 0$, $p \ge 2$, and $q \ge 2$*

$$\overline{\lim_{n\to\infty}} \sup_{\theta \in B_{p,q}^\alpha(B)} \frac{R(\theta, \tilde\theta^w)}{R_{*W}(\alpha, p, q, B)} \le K_{p,q}, \tag{16}$$

where $1 \le K_{p,q} \le 1.25$, and $K_{22} = 1$.

Proof: As in (22) the linear estimators are within $K_{p,q}$ of being asymptotically minimax. For these we have

$$\inf_{\hat\theta_\mathcal{L}} \sup_{\theta \in B_{p,q}^\alpha(B)} R(\theta, \hat\theta_\mathcal{L}) = \sup_{\theta \in B_{p,q}^\alpha(B)} \sum_{j=j_0}^\infty \sum_{k=1}^{2^j} \frac{\theta_{jk}^2 n^{-1}}{\theta_{jk}^2 + n^{-1}} \tag{17}$$

$$\le \sup_{\{t_j : \sum_{j=j_0}^\infty (2^{j(\alpha+1/2-1/p)}t_j)^q < B\}} \sum_{j=j_0}^{\log_2 n} \frac{t_j^2 2^j n^{-1}}{t_j^2 + 2^j n^{-1}} + o(n^{-2\alpha/(2\alpha+1)})$$

since the supremum occurs when each $\theta_{jk}^2 = t_j^2 2^{-j}$ and then $\sum_{j=1}^{2^j} \theta_{jk}^p = t_j^p 2^{j(1-p/2)}$. At the same time, letting $\|\theta\|_{(j)}^2 = \sum_{k=1}^{2^j} \theta_{jk}^2$ we have

$$\sup_{\theta \in B_{p,q}^\alpha(B)} R(\theta, \tilde\theta) = \sup_{\theta \in B_{p,q}^\alpha(B)} \{ \sum_{j=j_0}^{\log_2 n} \frac{\|\theta\|_{(j)}^2 2^j n^{-1}}{\|\theta\|_{(j)}^2 + 2^j n^{-1}} + \sum_{j>\log_2 n} \|\theta\|_{(j)}^2 + 2n^{-1}\log_2 n \}$$

$$= \sup_{\theta \in B_{p,q}^\alpha(B)} \sum_{j=j_0}^\infty \sum_{k=1}^{2^j} \frac{\theta_{jk}^2 n^{-1}}{\theta_{jk}^2 + n^{-1}} + o(n^{-2\alpha/(2\alpha+1)}). \tag{18}$$

since $\sup_{\theta \in B^{\alpha}_{p,q}(B)} \sum_{j > \log_2 n} \|\theta\|^2_{(j)} = o(n^{-2\alpha/(2\alpha+1)})$. Combining $(18), (17)$ and (14) yields (16). ∎

Remark: The estimator (15) has also been considered in Donoho and Johnstone (1995). It is shown that the estimator does not perform well numerically. The main reason is that the block sizes grow too fast and the estimator is not well localized.

Instead of using resolution levels as natural blocks, we can also apply the blocking method used in Section 3 in the wavelet setting. In a wavelet basis, the coefficients $\theta_{j,k}$ can be matched to (1) by using a lexicographic correspondence. To do this set $\theta_i = \theta_{jk}$ with $i = (2^j - 2^{j_0}) + k$. Then, for example, for Besov balls in $B^{\alpha}_{2,2}$ (5) holds with $c_i = 2^{2\alpha j}$. Note that these c_i have a slightly different structure from (6) since (6) implied $c_{i+1}/c_i \to 1$. The c_i here are constant over the range corresponding to $2^j - 2^{j_0} + k$, $k = 1, \ldots, 2^j$ which is clearly consistent with this, but $c_{2^j - 2^{j_0}+1}/c_{2^j - 2^{j_0}} = 2^{-2\alpha}$, which is not.

In general, we can construct a blockwise shrinkage estimator in a wavelet basis by first piecing together the empirical wavelet coefficients at different resolution levels (in the order of the lowest to the highest) as a vector, then applying the Steps 1-3 to the vector of the empirical coefficients with a base multiplier $b_n = 1 + 1/\log n$. Denote the resulting estimator by $\tilde{\theta}$. Then by using a similar proof as that of Theorem 1, it can be shown that the estimator is adaptively rate optimal over a range of Besov balls. We omit the details of the proof here.

Theorem 3 *For $B > 0$, $p \geq 2$, and $q \geq 1$*

$$\overline{\lim_{n \to \infty}} \sup_{\theta \in B^{\alpha}_{p,q}(B)} \frac{R(\theta, \tilde{\theta})}{R_{*W}(\alpha, p, q, B)} = M_{p,q},$$

where $M_{p,q}$ is a constant.

Our simulation results in Section 6 show that this rate-adaptive estimator performs well numerically.

5   Other parameter spaces

Various generalizations to a wider variety of parameter spaces are possible, and follow easily from the steps used in the preceding proofs. One may then conclude that $\tilde{\theta}$ is adaptively

9

minimax in rate and provide very good bounds on the relative constants multiplying the rate. Here are two such generalizations, plus sketches for their proofs. For the first we consider the parameter set

$$\Theta_1(\alpha, B) = \{\theta : \sum e^{2\alpha i}\theta_i^2 \leq B\}. \tag{19}$$

This function class has been considered in Efromovich and Pinsker (1982), and Donoho and Liu (1991). In this problem it may be calculated as in Donoho, Liu and MacGibbon (1990) that the minimax risk

$$R_{*1}(\alpha, B) \doteq \inf_{\hat{\theta}} \sup_{\theta \in \Theta_1(\alpha, B)} R(\theta, \hat{\theta})$$

satisfies

$$R_{*1}(\alpha, B) = \frac{\log n}{2\alpha n}(1 + o(1)) \ \ as \ \ n \to \infty \tag{20}$$

uniformly for all $0 < B_0 \leq B \leq B_0^{-1}$.

Theorem 4 *Consider the problem* $(1)-(2)$ *with* $\Theta = \Theta_1$ *as in* (19). *Let* $\tilde{\theta}$ *be the estimator defined in* (3) *with a constant base multiplier b. Let* $B_0 > 0$. *Then for each* α

$$\varlimsup_{n \to \infty} \frac{\sup_{\theta \in \Theta_1(\alpha, B)} R(\theta, \tilde{\theta})}{R_{*1}(\alpha, B)} \leq b + \frac{4\alpha}{\log b} \tag{21}$$

uniformly for $B \in [B_0, B_0^{-1}]$.

Proof: Reasoning as in (11) yields

$$\sup_{\theta \in \Theta_1(\alpha, B)} R(\theta, \tilde{\theta}) \leq \sup_{\sum_{K_0+1}^{K_1} e^{2\alpha b^{k-1}} t_k^2 \leq B} \sum_{K_0+1}^{K_1} \frac{t_k^2 m_k n^{-1}}{t_k^2 + m_k n^{-1}} + \frac{2\log_b n}{n} + o(\frac{\log n}{n}).$$

(The term $2\log_b n/n$ derives from the additional $2\sigma^2$ term on the right of (9).) A standard Lagrange multiplier bound then yields

$$\sup_{\theta \in \Theta_1(\alpha, B)} R(\theta, \tilde{\theta}) \leq \frac{b\log n}{2\alpha n} + \frac{2\log n}{n\log b} + o(\frac{\log n}{n}).$$

(21) follows from this and (20). ∎

For the second example we choose a particular case of a quadratically convex Θ. Choosing only a particular case as we have done should hopefully make clear how to handle the general situation, but without requiring detailed general definitions. Let

$$\Theta = \Theta_2(\alpha, B) = \{\theta : \sum i^{4\alpha}\theta_i^4 \leq B\}.$$

10

Let $\mathcal{L}$ denote the class of linear estimators . Then

$$R_{**2}(\alpha, B) \doteq \inf_{\hat{\theta}_{\mathcal{L}} \in \mathcal{L}} \sup_{\theta \in \Theta_2(\alpha,B)} R(\theta, \hat{\theta}_{\mathcal{L}}) = \sup_{\theta \in \Theta_2(\alpha,B)} \sum \frac{\theta_i^2 n^{-1}}{\theta_i^2 + n^{-1}} = C_2(\alpha) B^{\frac{1}{4\alpha+1}} n^{-\frac{4\alpha-1}{4\alpha+1}}(1+o(1)).$$

Since Θ is quadratically convex Donoho, Liu and MacGibbon (1990) show that

$$R_{**2}(\alpha, B) \leq 1.25 R_{*2}(\alpha, B) \doteq 1.25 \inf_{\hat{\theta}} \sup_{\theta \in \Theta_2(\alpha,B)} R(\theta, \hat{\theta}) \tag{22}$$

(It is possible with more care to replace the constant 1.25 in (22) by a slightly smaller value, but still larger than 1. It is also possible to evaluate $C_2(\alpha)$.)

Theorem 5 *Let $\tilde{\theta}$ be a blockwise James-Stein estimator with a constant base multiplier b. Then for each $\alpha > 1/4$, $B > 0$*

$$\varlimsup_{n \to \infty} \frac{\sup_{\theta \in \Theta_2(\alpha,B)} R(\theta, \tilde{\theta})}{R_{*2}(\alpha, B)} \leq 1.25 \varlimsup_{n \to \infty} \frac{\sup_{\theta \in \Theta_2(\alpha,B)} R(\theta, \tilde{\theta})}{R_{**2}(\alpha, B)} \leq 1.25 b^{\frac{4\alpha}{1+4\alpha}} \leq 1.25 b \tag{23}$$

For the estimator $\tilde{\theta}^*$ the upper bound is 1.25.

Proof: First note that

$$\sup_{\sum_{i \in G_k} \theta_i^4 = t^4} \sum_{i \in G_k} \frac{\theta_i^2 n^{-1}}{\theta_i^2 + n^{-1}} = \frac{t^2 m_k n^{-1}}{t^2 + m_k^{1/2} n^{-1}} = \sup_{\sum_{i \in G_k} \theta_i^4 = t^4} \frac{\|\theta\|_{(k)}^2 m_k n^{-1}}{\|\theta\|_{(k)}^2 + m_k n^{-1}}$$

since both suprema occur when each $\theta_i = m_k^{-1/4} t$. Following the plan of proof of Theorem 1 we can thus write

$$R_{**2}(\alpha, B) \geq \sup_{\{t: \sum_{K_0+1}^{K_1} b^{4k\alpha} t_k^4 \leq B\}} \sum_{K_0+1}^{K_1} \frac{t_k^2 m_k n^{-1}}{t_k^2 + m_k^{1/2} n^{-1}} = C_2^*(\alpha) B^{\frac{1}{1+4\alpha}} n^{-\frac{4\alpha-1}{4\alpha+1}}(1 + o(1)) \tag{24}$$

and

$$\begin{aligned} R(\theta, \tilde{\theta}) &\leq (1+o(1)) \sup_{\{t: \sum_{K_0+1}^{K_1} b^{4k\alpha} t_k^4 \leq B b^{4\alpha}\}} \sum_{K_0+1}^{K_1} \frac{t_k^2 m_k n^{-1}}{t_k^2 + m_k^{1/2} n^{-1}} \\ &= C_2^*(\alpha) B^{\frac{1}{1+4\alpha}} b^{\frac{4\alpha}{1+4\alpha}} n^{-\frac{4\alpha-1}{4\alpha+1}}(1 + o(1)). \end{aligned} \tag{25}$$

Taking the ratio of (24) and (25) yields the second inequality in (23). ∎

6 Numerical results

In the earlier sections it is shown that the blockwise James-Stein estimator as described in Steps 1-3 in Section 2 enjoys a number of desirable adaptivity properties. In contrast to some of the other adaptive minimax estimators in the literature, our estimator is very easy to implement with $O(n)$ computational cost. In this section we study the numerical performance of the estimator $\tilde{f}^*$ in a wavelet basis. In implementing $\tilde{f}^*$ in a wavelet basis, we first piece the empirical wavelet coefficients at different resolution levels (in the order of the lowest to the highest) together as a vector, then the Steps 1-3 are applied to the vector of the empirical coefficients with a base multiplier $b_n = 1 + 1/\log n$. The blocking scheme can be adjusted so that all blocks fall within single resolution levels. There might be a slight advantage in doing so, but we choose to implement our estimators in the simplest way.

We compare the MSE of $\tilde{f}^*$ with those of VisuShrink (Donoho and Johnstone (1994)), translation-invariant denoising (Coifman and Donoho (1995)), and BlockJS (Cai (1999)). See the original papers for details of the methods. We use the four well-known test functions of Donoho and Johnstone (1994) to compare the performance of the estimators. The test functions are normalized so that all of the functions have equal $s.d.(f) = 10$. Each of the four methods is applied to noisy versions of the test functions. Daubechies compactly supported wavelet $Symmlet$ 8 is used. Table 1 reports the mean squared error over 100 replications with sample sizes ranging from $n = 512$ to $n = 8192$ and signal-to-noise ratio (SNR) 3. Figure 1 provides a graphical comparison of the mean squared error of $\tilde{f}^*$ with those of the other three estimators. In Figure 1, the vertical bars represent the ratios of the MSEs of various estimators to the corresponding MSE of $\tilde{f}^*$. The higher the bar the better the relative performance of $\tilde{f}^*$, and a value of one means that the estimators have equal performance. For each signal the bars are ordered from left to right by the sample sizes ($n = 512, 1024, 2048, 4096, 8192$).

Simulation shows that the estimator $\tilde{f}^*$ overall enjoys significantly better performance relative to both VisuShrink and TI-denoising , especially when the underlying function has large spatial variability. $\tilde{f}^*$ outperforms VisuShrink in 17 out of 20 cases, and is better than TI-denoising in 15 out of 20 cases. In many cases the reduction in MSE is substantial. See Figure 1. For three of the four test functions, Doppler, Bumps, and Blocks, the estimator has better precisions with sample size n than VisuShrink with sample size $2 \cdot n$ for all n

from 512 to 8192 (see Table 1).

n	$\tilde{f}^*$	Visu	TI	BlockJS	n	$\tilde{f}^*$	Visu	TI	BlockJS
Doppler					*Bumps*				
512	2.25	5.43	4.33	2.49	512	6.42	18.21	14.46	9.22
1024	1.26	3.76	3.01	1.49	1024	4.93	11.73	8.98	4.56
2048	0.87	2.38	1.87	0.85	2048	2.94	8.10	5.78	2.54
4096	0.46	1.36	1.05	0.56	4096	1.83	4.32	2.94	1.81
8192	0.31	0.90	0.67	0.31	8192	1.04	2.68	1.76	0.92
Blocks					*HeaviSine*				
512	4.67	9.02	8.08	6.16	512	1.25	1.06	0.90	1.05
1024	3.49	6.68	5.91	3.94	1024	0.82	0.64	0.56	0.64
2048	2.36	4.82	4.27	2.58	2048	0.54	0.45	0.38	0.41
4096	1.66	2.78	2.51	2.10	4096	0.35	0.35	0.26	0.35
8192	1.13	1.97	1.71	1.35	8192	0.20	0.25	0.18	0.24

Table 1: Mean Squared Error From 100 Replications (SNR = 3).

Figure 1: Comparison of MSEs. For each signal the bars are ordered from left to right by the sample sizes (n=512 to 8192). The higher the bar the better the relative performance of $\tilde{f}^*$.

It is interesting to compare the numerical performance of $\tilde{f}^*$ with that of BlockJS in terms of mean squared error. Overall, the estimator proposed in this paper outperforms BlockJS for Doppler, Blocks, and HeaviSine (in the large sample size case), and under-performs BlockJS for Bumps and HeaviSine (in the small sample size case). See Table 1 and Figure 1. Recall that the block size in BlockJS is fixed at $\log n$ for all blocks instead of growing as in (3). At the higher resolution levels, BlockJS has smaller block size than $\tilde{f}^*$ and can thus more readily adjust to the local smoothness properties of the underlying function. The Bumps signal is spatially inhomogeneous. Its wavelet coefficients vary significantly and the large coefficients at high resolution levels cluster in group around the peaks of the function. BlockJS is well suited for such a spatially inhomogeneous signal. On the other hand, wavelet coefficients of relatively smooth signals do not vary significantly at high resolution levels. In such a case, our new estimator $\tilde{f}^*$ performs very well.

13

References

Brown, L.D., Cai, T., Low, M.G. and Zhang, C. (2002), 'On Asymptotic Equivalence of White Noise Model and Nonparametric Regression with Random Designs', *The Annals of Statistics*, 30, 688-707.

Brown, L.D. and Low, M.G. (1996), 'Asymptotic Equivalence of Nonparametric Regression and White Noise', *The Annals of Statistics*, 24, 2384-2398.

Brown, L. D., Carter, A. V., Low, M. G. and Zhang, C.-H. (2004), 'Equivalence Theory for Density Estimation, Poisson Processes and Gaussian White Noise with Drift', *The Annals of Statistics*, 32, 2074-2097.

Brown, L.D., Low, M.G. and Zhao, L. (1997), 'Superefficiency in Nonparametric Function Estimation', *The Annals of Statistics*, 28, 2607-25.

Brown, L.D. and Zhao, L. (2000), 'Convenient Asymptotic Near Equivalence of Nonparametric Regression and the Series Version of White Noise with Drift', Technical Report, Department of Statistics, University of Pennsylvania.

Cai, T. (1999), 'Adaptive Wavelet Estimation: A Block Thresholding and Oracle Inequality Approach', *The Annals of Statistics*, 898-924.

Coifman, R.R. and Donoho, D.L. (1995), 'Translation Invariant Denoising', In A. Antoniadis and G. Oppenheim (eds), *Wavelets and Statistics*, Lecture Notes in Statistics, 103, New York: Springer-Verlag, 125-150.

Donoho, D.L. and Johnstone, I.M. (1994), 'Ideal Spatial Adaptation via Wavelet Shrinkage', *Biometrika*, 81, 425-55.

Donoho, D.L. and Johnstone, I.M. (1995), 'Adapting to Unknown Smoothness via Wavelet Shrinkage', *Journal of the American Statistical Association*, 90, 1200–1224.

Donoho, D.L. and Johnstone, I.M. (1998), 'Minimax Estimation via Wavelet Shrinkage', *The Annals of Statistics*, 26, 879-921.

Donoho, D. L. and Liu, R. G. (1991), 'Geometrizing Rates of Convergence III'. *The Annals of Statistics*, 19, 668-701.

Donoho, D.L., Liu, R.C. and MacGibbon, B. (1990), 'Minimax Risk over Hyperrectangles, and Implications', *The Annals of Statistics*, 18, 1416-37.

Efromovich, S.Y. (1985), ' Nonparametric Estimation of a Density of Unknown Smoothness', *Theory of Probability and Its Applications*, 30, 557-661.

Efromovich, S. Y.(1998), Simultaneous Sharp Estimation of Functions and Their Derivatives. *The Annals of Statistics*, 26, 273-278.

Efromovich, S. Y.(2000), 'Can Adaptive Estimators for Fourier Series Be of Interest to Wavelets?', *Bernoulli*, 6, 699-708.

Efromovich, S. Y. and Pinsker, M. S. (1982), 'Estimation of Square-Integrable Probability Density of a Random Variable', *Problems of Information Transmission*, 18, 19-38.

Efromovich, S. Y. and Pinsker, M. S. (1984), 'An Adaptive Algorithm of Nonparametric Filtering', *Automation And Remote Control*, 11, 58-65.

Efron, B. and Morris, C. (1973), 'Stein's Estimation Rule and Its Competitors - an Empirical Bayes Approach', *Journal of the American Statistical Association*, 68, 117-130.

Goldenshluger, A. and Tsybakov, A. (2003), 'Adaptive Prediction and Estimation in Linear Regression with Infinite Many Parameters', *Journal of Multivariate Analysis*,84, 40-60.

Klemelä, J. and Nussbaum, M. (1999), 'Constructive Asymptotic Equivalence of Density Estimation and Gaussian White Noise', Discussion paper No. 53, *Sonderforschungsbereich* 373, Humboldt University, Berlin.

Kerkyacharian, G., Picard, D. and Tribouley, K (1996), 'L_p Adaptive Density Estimation', *Bernoulli*, 2, 229-247.

Lepski, O. V., Mammen, E. and Spokoiny, V. G. (1997), 'Optimal Spatial Adaptation to Inhomogeneous Smoothness: an Approach Based on Kernel Estimates with Variable Bandwidth Selectors', *The Annals of Statistics*, 25, 929 - 947.

Nussbaum, M. (1996), 'Asymptotic Equivalence of Density Estimation and Gaussian White Noise', *The Annals of Statistics*, 24, 2399 - 2430.

Pinsker, M. S. (1980), 'Optimal Filtering of Square Integrable Signals in Gaussian White Noise', *Problems of Information Transmission*, 16, 120-133.

Zhang, C. (2005), 'General Empirical Bayes Wavelet Methods and Exactly Adaptive Minimax Estimation', *The Annals of Statistics*, 33, 54-100.