Statistica Sinica **13**(2003), 881-902

RATES OF CONVERGENCE AND ADAPTATION OVER BESOV SPACES UNDER POINTWISE RISK

T. Tony Cai

University of Pennsylvania

Abstract: Function estimation over the Besov spaces under pointwise ℓ^r ($1 \le r < \infty$) risks is considered. Minimax rates of convergence are derived using a constrained risk inequality and wavelets. Adaptation under pointwise risks is also considered. Sharp lower bounds on the cost of adaptation are obtained and are shown to be attainable by a wavelet estimator. The results demonstrate important differences between the minimax properties under pointwise and global risk measures. The minimax rates and adaptation for estimating derivatives under pointwise risks are also presented. A general ℓ^r-risk oracle inequality is developed for the proofs of the main results.

Key words and phrases: Adaptability, adaptive estimation, Besov spaces, constrained risk inequality, minimax estimation, nonparametric functional estimation, oracle inequality, rate of convergence, wavelets, white noise model.

1. Introduction

Besov spaces occur naturally in many fields of analysis. They contain as special cases a number of traditional smoothness spaces such as Hölder and Sobolev spaces. The Besov space $B^\alpha_{p,q}$, defined in detail in Section 2, contains functions having α bounded derivatives in L^p norm, the third parameter q gives a finer gradation of smoothness. Over the last few years, with the development of wavelet thresholding techniques, statistical function estimation over the Besov spaces has been of considerable interest. Most of the research, however, is focused on estimation under global risk measures, especially the mean integrated squared error (MISE). Asymptotic properties under pointwise risk measures are still mostly unknown.

Minimax rates of convergence over a ball in the Besov space $B^\alpha_{p,q}$ under MISE are derived in Donoho and Johnstone (1998) for the white noise model, and in Donoho (1995) for inverse problems. In all these cases, the global minimax rates of convergence depend solely on one smoothness parameter, α. It is also known that under MISE it is possible to have rate-optimal estimators converging faster than the minimax rate at some points in the parameter space. Indeed, it is even possible to be superefficient at every point in a Besov ball, see Cai (2000) and Zhang (2000). See also Brown, Low and Zhao (1997).

In this paper we consider function estimation over Besov classes under pointwise ℓ^r risk for all $1 \leq r < \infty$. The results reveal some interesting features about minimax properties under pointwise risk measures which are significantly different from minimax properties under the global MISE. Unlike the global minimax rates, pointwise minimax convergence rates over a ball in the Besov space $B^\alpha_{p,q}$ depends not only on α but on p as well. More interestingly, in a sharp contrast to the global risk measure, under pointwise ℓ^r risks any rate-optimal estimator must uniformly attain the same "flat" rate at every f in the parameter space. Thus superefficiency without penalty is impossible. That is, if an estimator has risk converging faster than the minimax rate at some f in the parameter space, the risk of the estimator must converge at slower than the minimax rate at other functions in the parameter space.

We take a new approach in deriving pointwise minimax lower bounds in this paper. There are several existing methods: modulus of continuity, metric entropy, information inequality, and renormalization. See, e.g., Farrell (1972), Hasminskii (1979), Stone (1980), Ibragimov and Hasminskii (1984), Donoho and Liu (1991), Brown and Low (1991), Low (1992), Donoho and Low (1992) and Birgé and Massart (1995). Our method is based on a constrained ℓ^r-risk inequality which is a generalization of an ℓ^2 version introduced in Brown and Low (1996b). The constrained risk inequality specifies, in a general setting, a lower bound on the ℓ^r risk at one parameter point subject to the constraint of a given risk at another parameter point. This approach has the advantage of allowing for simultaneous study of lower bounds on minimax rate of convergence and the cost of adaptation. To establish the minimax upper bounds, we use wavelet estimators.

Adaptation is now an important part of nonparametric function estimation problems. In virtually all practical situations, the smoothness parameters are unknown. Therefore, adaptation to unknown smoothness is essential. In the case of more traditional smoothness spaces such as Lipschitz and Sobolev classes, the adaptation problem has been considered by Lepski (1990), Brown and Low (1996b), Lepski and Spokoiny (1997) and Tsybakov (1998) under squared error loss. In this paper, we consider the adaptability problem under pointwise ℓ^r risk over Besov classes. It is shown that, with a certain condition on the smoothness parameters α and p, adaptation can be achieved without cost over a collection of Besov classes. That is, the exact minimax rate of convergence can be attained simultaneously by a single estimator over each of the Besov classes in the collection. Under such a condition a wavelet projection estimator can attain the optimal rate over all the Besov classes in the collection. In general, however, adaptation for free is impossible and one has to pay a penalty for not knowing the underlying smoothness. Lower bounds on the cost of adaptation are derived using the constrained ℓ^r-risk inequality.

Wavelet estimators are used to show that the lower bounds are sharp. That is, it is possible to achieve adaptation with the minimum cost given by the lower bounds. The approach we take can be applied to other function estimation problems as well. The minimax rates of convergence and adaptation for estimating derivatives under pointwise ℓ^r risk are also considered.

It is now well understood that oracle inequalities are an effective tool to study asymptotic properties of wavelet estimators. See, e.g., Donoho and Johnstone (1994), Cai (1999), Antoniadis and Fan (2001) and Johnstone (1998). However, most of the discussions have so far been restricted to the standard case of mean squared error and are thus not applicable to our problems. To prove our main results, we develop a general oracle inequality for ℓ^r risks and for arbitrary noise distributions. This general oracle inequality can be of independent interest. As a special case of the general ℓ^r-risk oracle inequality we obtain an ℓ^r-risk oracle inequality for Gaussian noise which serves as one of our main tools for the proofs. The oracle inequalities are derived using the approach of optimal recovery.

The paper is organized as follows. After Section 2, in which basic notation and definitions of the Besov spaces and wavelet bases are reviewed, we derive the minimax rates of convergence over Besov classes in Section 3. Minimax rates are established in two steps. First minimax lower bounds are obtained by using a constrained risk inequality, then a wavelet projection estimator is constructed and is shown to converge at the same rates as the lower bounds. It is also shown that any rate-optimal estimator must uniformly attain the same rate at every fixed point in the parameter space. The adaptation problem is investigated in Section 4. Sharp lower bounds on the cost of adaptation are derived, and it is shown that a wavelet estimator is adaptive with the minimum cost over a wide range of Besov classes under pointwise ℓ^r risk for $1 \le r < \infty$. We consider in Section 5 the minimax rates and adaptation for estimating derivatives over the Besov classes under pointwise risks. Section 6 presents the general ℓ^r-risk oracle inequalities. The proofs of the main results are postponed to Section 7.

2. Besov Spaces and Wavelets

A Besov space $B^\alpha_{p,q}$ has three parameters: α measures degree of smoothness, p and q specify the type of norm used to measure the smoothness. These spaces arise naturally in many fields of analysis. They contain many traditional smoothness spaces such as Hölder(-Zygmund) and Sobolev spaces as special cases.

There are several ways of defining the Besov spaces. For the present paper, we will use two versions of Besov norms: one is defined through the modulus of smoothness and another is based on the wavelet coefficients.

For $f \in L^p[0,\ 1]$ and $h > 0$, denote the Kth difference by $\Delta_h^{(K)} f(t) = \sum_{k=0}^{K} (-1)^k f(t+kh)$. The modulus of smoothness of order K of f is $\omega_{K,p}(f,h) = \|\Delta_h^{(K)} f\|_{L^p[0,1-Kh]}$. The Besov norm of index (α, p, q) is defined for $K > \alpha$ by

$$\|f\|_{B^\alpha_{p,q}} = \begin{cases} \|f\|_p + \left(\int_0^1 [h^{-\alpha}\omega_{K,p}(f,h)]^q \frac{dh}{h}\right)^{1/q} & \text{for } q < \infty \\ \|f\|_p + \|h^{-\alpha}\omega_{K,p}(f,h)\|_\infty & \text{for } q = \infty. \end{cases} \tag{1}$$

The Besov space $B^\alpha_{p,q}$ on $[0,\ 1]$ is a Banach space consisting of functions with finite Besov norm $\|\cdot\|_{B^\alpha_{p,q}}$. The Besov class $B^\alpha_{p,q}(M)$ is defined to be a ball of radius M in the Besov space $B^\alpha_{p,q}$: $B^\alpha_{p,q}(M) = \{f : \|f\|_{B^\alpha_{p,q}} \le M\}$. The Besov spaces are very rich function spaces containing both smooth and nonsmooth functions. For example, with $p = q = \infty$ and $\alpha = 1$, there exists a dense open subset of $B^1_{\infty,\infty}$ which is composed of nowhere differentiable functions, see Meyer (1992). The Besov spaces on the real line $I\!R$ can be defined analogously and we denote by $B^\alpha_{p,q}(I\!R)$ and $B^\alpha_{p,q}(I\!R,\ M)$ a Besov space and a Besov class on the line, respectively. See Triebel (1983, 1992) and Meyer (1992) for more on Besov spaces. Also see Donoho and Johnstone (1998) for discussions on the relevance of Besov spaces to scientific problems.

The Besov spaces can also be defined based on the sequence norm of wavelet coefficients. An orthonormal wavelet basis of $L^2[0,1]$ is generated from dilation and translation of two basic functions, a "father" wavelet ϕ and a "mother" wavelet ψ. In the present paper, the functions ϕ and ψ are assumed to be compactly supported and $\int \phi = 1$. We call a wavelet ψ *K-regular* if ψ has K vanishing moments and K continuous derivatives. Let $\phi_{jk}(t) = 2^{j/2}\phi(2^j t - k)$, $\psi_{jk}(t) = 2^{j/2}\psi(2^j t - k)$. The collection $\{\phi_{j_0 k},\ k = 1, \ldots, 2^{j_0};\ \psi_{jk},\ j \ge j_0,\ k = 1, \ldots, 2^j\}$ with appropriate treatments at the boundaries is then an orthonormal basis of $L^2[0,1]$, provided the primary resolution level j_0 is large enough to ensure that the support of the scaling functions and wavelets at level j_0 is not the whole of $[0,1]$. See Cohen, Daubechies, Jawerth and Vial (1993), Daubechies (1994) and Meyer (1991) for further details on wavelet bases on the unit interval $[0,\ 1]$. For wavelets on the line, see Daubechies (1992) and Meyer (1992).

For a function $f : [0,\ 1] \to I\!R$, denote $\xi_{jk} = \int_0^1 f(t)\phi_{jk}(t)\,dt$ and $\theta_{jk} = \int_0^1 f(t)\psi_{jk}(t)\,dt$. Define the sequence norm of wavelet coefficients of f by

$$\|f\|_{b^\alpha_{p,q}} = \|\xi_{j_0,k}\|_{\ell^p} + \left(\sum_{j=j_0}^{\infty}\left(2^{js}\left(\sum_k |\theta_{j,k}|^p\right)^{1/p}\right)^q\right)^{1/q}, \tag{2}$$

where $s = \alpha + 1/2 - 1/p$. The standard modification applies for the cases $p, q = \infty$.

Suppose the wavelet ψ is *K-regular* with $K > \alpha$. Let $1 \le p \le \infty$. Then the Besov function norm defined in (1) is equivalent to the Besov sequence norm

(2) for every $f \in L^p[0,1]$. The equivalence means that there exist constants $c^* \geq c_* > 0$ independent of f and such that $c_* \leq \|f\|_{B^\alpha_{p,q}}/\|f\|_{b^\alpha_{p,q}} \leq c^*$. See Meyer (1992) and DeVore and Popov (1988).

3. Minimax Rate under Pointwise ℓ^r Risk

Consider the white noise model in which we observe Gaussian processes $Y_n(t)$ governed by

$$dY_n(t) = f(t)dt + n^{-1/2}dW(t), \quad 0 \leq t \leq 1, \tag{3}$$

where $W(t)$ is a standard Brownian motion and f is an unknown function of interest. The white noise model is asymptotically equivalent to the conventional formulation of nonparametric regression. See Brown and Low (1996a), Brown, Cai, Low and Zhang (2002). The white noise model is also equivalent to nonparametric density estimation. See Nussbaum (1996), Klemelä and Nussbaum (1999) and Brown, Low and Zhang (2000).

Although f is unknown in detail, we assume in this section that f belongs to a known Besov class $B^\alpha_{p,q}(M)$. We wish to estimate f under the pointwise ℓ^r risk

$$R_r(\hat{f},\ f,\ t_0) = E_f|\hat{f}(t_0) - f(t_0)|^r, \tag{4}$$

where $t_0 \in (0,\ 1)$ is any fixed point and $1 \leq r < \infty$.

The difficulty of the estimation problem is measured by the minimax risk

$$R^*_r(B^\alpha_{p,q}(M);\ t_0, n) = \inf_{\hat{f}_n} \sup_{f \in B^\alpha_{p,q}(M)} E_f|\hat{f}_n(t_0) - f(t_0)|^r. \tag{5}$$

We wish to determine the rate of convergence of the minimax risk as $n \to \infty$.

The minimax risk over a fixed Besov class has been studied by Donoho and Johnstone (1998) when the risk measure is the global mean integrated squared error (MISE). Donoho and Johnstone (1998) show that the minimax MISE risk over a Besov class $B^\alpha_{p,q}(M)$ is of the order $n^{-2\alpha/(1+2\alpha)}$, i.e., $\inf_{\hat{f}_n} \sup_{f \in B^\alpha_{p,q}(M)} E\|\hat{f}_n - f\|_2^2 \asymp n^{-\frac{2\alpha}{1+2\alpha}},\ \ n \to \infty$. In particular, the global minimax rate is determined solely by the smoothness index α and does not depend on the other two parameters p and q of the Besov class, provided that $\alpha > 1/p$.

As we show below, the pointwise risk behaves differently from the global risk. The pointwise minimax risk depends on two smoothness parameters of the Besov class, α and p, and it converges at a rate slower than the corresponding global rate. Also pointwise rate-optimal estimators have some interesting properties. See Theorem 2 below.

The minimax convergence rate under the pointwise risk (4) is derived in two steps. First we establish the lower bounds for the minimax risk. We denote by $\nu = \alpha - 1/p$ and assume $\nu > 0$ and $1 \leq p \leq \infty$ in the rest of the paper.

Theorem 1.(Lower Bound) *The minimax risk of estimating f over the Besov class $B^\alpha_{p,q}(M)$ under the ℓ^r risk* (4) *is bounded below by $R^*_r(B^\alpha_{p,q}(M);\ t_0,n) \geq Cn^{-r\nu/(1+2\nu)}$ for some fixed constant $C > 0$. Equivalently, for any estimator $\hat{f}_n$ and any $B_n \to \infty$,*

$$\overline{\lim_{n\to\infty}}\, n^{\frac{r\nu}{1+2\nu}} B_n \sup_{f\in B^\alpha_{p,q}(M)} E_f|\hat{f}_n(t_0) - f(t_0)|^r = \infty. \tag{6}$$

Theorem 1 states that the minimax risk converges no faster than $n^{-\frac{r\nu}{1+2\nu}}$. We obtain this minimax lower bound by using a different approach from more conventional methods. The main tool we use is a general constrained ℓ^r-risk inequality which is also used to study the adaptability problem.

Second, we show that the lower bound given in Theorem 1 can be attained. In fact, as given in the proof, the lower bound on the rate of convergence can be attained by a simple wavelet projection estimator.

Theorem 2.(Upper Bound) *There exist estimators $\hat{f}_n$ of f attaining the convergence rate of $n^{r\nu/(1+2\nu)}$ over the Besov class $B^\alpha_{p,q}(M)$, i.e.,*

$$\overline{\lim_{n\to\infty}}\, n^{\frac{r\nu}{1+2\nu}} \sup_{f\in B^\alpha_{p,q}(M)} E_f|\hat{f}_n(t_0) - f(t_0)|^r < \infty. \tag{7}$$

Furthermore, for any estimator $\hat{f}_n$ satisfying (7), *the estimator must also satisfy*

$$\underline{\lim_{n\to\infty}}\, n^{\frac{r\nu}{1+2\nu}} E_f|\hat{f}_n(t_0) - f(t_0)|^r > 0 \tag{8}$$

for any fixed $f \in B^\alpha_{p,q}(M)$.

The minimax rate of convergence is established by combining Theorems 1 and 2.

Corollary 1. *The minimax rate of convergence of estimating f over the Besov class $B^\alpha_{p,q}(M)$ under the ℓ^r risk* (4) *is $n^{r\nu/(1+2\nu)}$. That is,*

$$0 < \underline{\lim_{n\to\infty}}\, n^{\frac{r\nu}{1+2\nu}} R^*_r(B^\alpha_{p,q}(M);\ t_0,n) \leq \overline{\lim_{n\to\infty}}\, n^{\frac{r\nu}{1+2\nu}} R^*_r(B^\alpha_{p,q}(M);\ t_0,n) < \infty.$$

Unlike the minimax rate under MISE risk, the minimax rate under pointwise risk not only depends on the smoothness index α, but also depends on the parameter p.

Remark. The second part of Theorem 2 says that any rate-optimal estimator must attain the same rate at every f in the parameter space. In other words, the rate of convergence has to be "flat" over $B^\alpha_{p,q}(M)$. In contrast, under the global MISE risk, rate-optimal estimators over $B^\alpha_{p,q}(M)$ can achieve a much faster rate at some parameter points. Indeed, it is possible to have estimators which converge at a rate faster than the minimax rate at every fixed function in $B^{\alpha}_{p,q}(M)$, see Cai (2000) and Zhang (2000). See also Brown, Low and Zhao (1997).

4. Adaptation

Adaptive estimation has become an important part of nonparametric function estimation problems. Adaptation to unknown smoothness is essential because the smoothness parameters of the underlying functions are unknown in virtually all practical situations.

We consider in this section the adaptability problem in estimating f under pointwise risk over a wide range of Besov classes $B^{\alpha}_{p,q}(M)$. Let $\mathcal{F} = \{B^{\alpha}_{p,q}(M) : (\alpha, p, q, M) \in \Sigma\}$ be a collection of Besov classes, where Σ is some index set. The adaptability problem concerns whether it is possible to find an estimator sequence $\hat{f}_n$ to attain the optimal convergence rate simultaneously over every $B^{\alpha}_{p,q}(M) \in \mathcal{F}$:

$$\overline{\lim_{n\to\infty}}\, n^{\frac{r\nu}{1+2\nu}} \sup_{f\in B^{\alpha}_{p,q}(M)} R_r(\hat{f}_n,\ f,\ t_0) < \infty, \quad \text{for all } (\alpha, p, q, M) \in \Sigma.$$

If such an estimator exists, we say that adaptation for free is possible. When this is impossible, it is of interest to know the minimum cost for adaptation. Adaptability over Besov classes has been considered in Cai (2000) under global MISE risk. Once again, the behaviors of the estimators under global risk and pointwise risk are quite different. We consider two cases.

4.1. When adaptation for free is possible

We first consider the case in which $\alpha - 1/p$ are the same for all $B^{\alpha}_{p,q}(M) \in \mathcal{F}$. Let $\mathcal{F}_\nu = \{B^{\alpha}_{p,q}(M) :\ \alpha - 1/p = \nu\}$ be a collection of Besov classes with the same value of $\alpha - 1/p = \nu$ and thus the same minimax convergence rates. In this case adaptation for free is possible.

Theorem 3. *There exists an estimator $\hat{f}_n$ of f attaining the convergence rate of $n^{r\nu/(1+2\nu)}$ over every Besov class $B^{\alpha}_{p,q}(M) \in \mathcal{F}_\nu$ under the pointwise ℓ^r risk* (4):

$$\overline{\lim_{n\to\infty}}\, n^{\frac{r\nu}{1+2\nu}} \sup_{f\in B^{\alpha}_{p,q}(M)} E_f|\hat{f}_n(t_0) - f(t_0)|^r < \infty \quad \text{for all } B^{\alpha}_{p,q}(M) \in \mathcal{F}_\nu. \tag{9}$$

Remark. The minimax convergence rates over all Besov classes in $\mathcal{F}_\nu$ are the same and Theorem 3 shows that in this case the exact minimax rate can be attained adaptively over each of the Besov classes in $\mathcal{F}_\nu$. However, for a general collection of parameter spaces with the same minimax convergence rate, adaptation for free is not always possible. There exist function classes with the same minimax rate over which adaptation for free is not possible. See Cai and Low (2002).

4.2. When adaptation for free is impossible

The more interesting and more general case is that there are some Besov classes in $\mathcal{F}$ with different values of $\alpha - 1/p$. We begin by considering the simplest case: the collection $\mathcal{F}$ consists of only two Besov classes $B^{\alpha_1}_{p_1,q_1}(M_1)$ and $B^{\alpha_2}_{p_2,q_2}(M_2)$ with $\nu_1 \neq \nu_2$, where $\nu_i \equiv \alpha_i - 1/p_i$ for $i = 1, 2$. We wish to know if it is possible to have an estimator f_n such that

$$\max_{i=1,2} \overline{\lim_{n\to\infty}}\, n^{\frac{r\nu_i}{1+2\nu_i}} \sup_{f\in B^{\alpha_i}_{p_i,q_i}(M_i)} R_r(\hat{f}_n,\ f,\ t_0) < \infty.$$

It is not difficult to see that in this case the second part of Theorem 2 implies that adaptation for free is impossible, since the two parameter spaces have nonempty intersection, $B^{\alpha_1}_{p_1,q_1}(M_1) \cap B^{\alpha_2}_{p_2,q_2}(M_2) \neq \emptyset$. Therefore, one must pay a price for not knowing the smoothness of the underlying function classes even in the simple case that the parameter space is one of only two possible Besov classes. Now the question is: what is the minimum cost for adaptation?

Theorem 4. *Let $B_n \to \infty$, $n/\log B_n \to \infty$ and let $\hat{f}_n$ be an estimator sequence of f based on* (3). *If f_0 is a function in $B^{\alpha}_{p,q}(M')$ with $M' < M$ satisfying*

$$\overline{\lim_{n\to\infty}}\, n^{\frac{r\nu}{1+2\nu}} B_n E_{f_0} |\hat{f}_n(t_0) - f_0(t_0)|^r < \infty, \tag{10}$$

$$\underline{\lim_{n\to\infty}} \left(\frac{n}{\log B_n}\right)^{\frac{r\nu}{1+2\nu}} \sup_{f\in B^{\alpha}_{p,q}(M)} E_f|\hat{f}_n(t_0) - f(t_0)|^r > 0. \tag{11}$$

Theorem 4 shows that even if an estimator $\hat{f}_n$ is superefficient at only one function f_0 in the interior of $B^{\alpha}_{p,q}(M)$, i.e., the risk of $\hat{f}_n$ converges faster than the minimax rate at $f_0(t_0)$, then it must pay a penalty of not being rate optimal over $B^{\alpha}_{p,q}(M)$.

The following is an immediate consequence of Theorem 4.

Corollary 2. *Consider two Besov classes $B^{\alpha_i}_{p_i,q_i}(M_i)$ with $\nu_i \equiv \alpha_i - 1/p_i$ for $i = 1, 2$. Let $\nu_1 > \nu_2 > 0$. If an estimator $\hat{f}_n$ attains a rate of n^{ρ} over $B^{\alpha_1}_{p_1,q_1}(M_1)$ with $\rho > r\nu_2/(1+2\nu_2)$, in particular, if $\hat{f}_n$ is rate-optimal over $B^{\alpha_1}_{p_1,q_1}(M_1)$, then*

$$\underline{\lim_{n\to\infty}} \left(\frac{n}{\log n}\right)^{\frac{r\nu_2}{1+2\nu_2}} \sup_{f\in B^{\alpha_2}_{p_2,q_2}(M_2)} E_f|\hat{f}_n(t_0) - f(t_0)|^r > 0.$$

Therefore the minimum cost for adaptation is at least a logarithmic factor. In the case of the Lipschitz and Sobolev classes under ℓ^2 risk, this problem has been considered by Lepski (1990), Brown and Low (1996b), Lepski and Spokoiny (1997) and Tsybakov (1998). It is not surprising that our result agrees with the results found in Lepski (1990) and Brown and Low (1996b), because both the risk measure and the function classes in those two papers are special cases of what we consider here.

Can the rate $(n/\log n)^{r\nu/(1+2\nu)}$ be adaptively attained? The answer to this question is yes and we will call this rate the adaptive minimax rate for estimating f over $B^{\alpha}_{p,q}(M)$ under the pointwise ℓ^r risk.

4.3. Adaptation with minimum cost

Using an orthonormal wavelet basis, the function f can be expanded into a wavelet series

$$f(t)=\sum_{k=1}^{2^{j_0}}\xi_{j_0k}\phi_{j_0k}(t)+\sum_{j=j_0}^{\infty}\sum_{k=1}^{2^j}\theta_{jk}\psi_{jk}(t) \tag{12}$$

with $\xi_{j_0,k}=\int_0^1 f(t)\phi_{j_0,k}(t)\,dt$ and $\theta_{j,k}=\int_0^1 f(t)\psi_{j,k}(t)\,dt$.

Let $y_{j,k}=\int\psi_{j,k}(t)\,dY_n(t)$ and $z_{j,k}=\int\psi_{j,k}(t)\,dW(t)$, and define $\tilde{y}_{j_0,k}$ and $\tilde{z}_{j_0,k}$ similarly. The white noise model (3) is equivalent to a sequence model in which one observes an empirical wavelet coefficient sequence:

$$\tilde{y}_{j_0,k}=\xi_{j_0,k}+n^{-1/2}\tilde{z}_{j_0,k},\quad k=1,2,\ldots,2^{j_0}, \tag{13}$$

$$y_{j,k}=\theta_{j,k}+n^{-1/2}z_{j,k},\quad k=1,2,\ldots,2^j,\ j\ge j_0, \tag{14}$$

where $\tilde{z}_{j_0,k}$ and $z_{j,k}$ are i.i.d. $N(0,1)$.

Let J be an integer satisfying $n\le 2^J<2n$. We apply the soft threshold rule to the empirical wavelet coefficients $y_{j,k}$ for $j<J$, to obtain estimated coefficients $\hat{\theta}_{j,k}$. Denote the soft threshold rule by

$$\eta_\beta(y)=sgn(y)(|y|-\beta)_+. \tag{15}$$

Define the estimators of the coefficients by

$$\begin{aligned}
\hat{\xi}_{j_0,k}&=\tilde{y}_{j_0,k}, && k=1,2,\ldots,2^{j_0},\\
\hat{\theta}_{j,k}&=\eta_{\lambda\sigma}(y_{j,k}), && k=1,2,\ldots,2^j,\ j_0\le j<J,\\
\hat{\theta}_{j,k}&=0 && j\ge J,
\end{aligned} \tag{16}$$

where $\lambda=(r\log n)^{1/2}$ and $\sigma=n^{-1/2}$. The estimator of f is given by

$$\hat{f}_n(t)=\sum_{k=1}^{2^{j_0}}\hat{\xi}_{j_0k}\phi_{j_0k}(t)+\sum_{j=j_0}^{J-1}\sum_{k=1}^{2^j}\hat{\theta}_{jk}\psi_{jk}(t). \tag{17}$$

Note that when $r = 2$ the estimator (16) is the usual soft threshold estimator of Donoho and Johnstone (1994). The wavelet estimator constructed in (17) achieves adaptation with the minimum cost in the sense that it attains the lower bound on the cost of adaptation given in Theorem 4.

Theorem 5. *Let $\hat{f}_n$ be the wavelet estimator of f given in* (17). *Suppose the wavelet ψ is K-regular. Then the estimator $\hat{f}_n$ simultaneously attains the adaptive convergence rate under the pointwise risk* (4) *for all $B^\alpha_{p,q}(M)$ with $\alpha < K$, $\nu \equiv \alpha - 1/p > 0$, $1 \le p \le \infty$, $0 < q \le \infty$, and $M > 0$. That is,*

$$\overline{\lim_{n\to\infty}} \left(\frac{n}{\log n}\right)^{\frac{r\nu}{1+2\nu}} \sup_{f\in B^\alpha_{p,q}(M)} E_f|\hat{f}_n(t_0) - f(t_0)|^r < \infty. \tag{18}$$

Remark. Suppose we consider a range of Besov classes $B^\alpha_{p,q}(M)$ with $0 < \nu \le \nu^*$. Then Theorem 4 and Corollary 2 do not exclude the possibility of having an estimator adaptively attaining the rate of $(n/\log n)^{r\nu/(1+2\nu)}$ for $\nu < \nu^*$ and the exact minimax rate $n^{r\nu^*/(1+2\nu^*)}$ for $\nu = \nu^*$. In fact, this is can be achieved by the wavelet estimator given in (16) and (17) with the lowest resolution level j_0 depending on n and satisfying $2^{j_0} \asymp n^{1/(1+2\nu^*)}$. Then, in view of Theorem 4 and Corollary 2, this is the best possible adaptive estimator at the level of the convergence rate.

5. Estimating Derivatives

The approach used in the preceding sections can also be applied to other function estimation problems. In this section we consider the problem of estimating the derivatives of f under pointwise risk measures.

The problem of estimating the first derivative f' over Besov classes under the global MISE, and other related inverse problems, has been considered in, e.g., Donoho (1995), Abramovich and Silverman (1998) and Cai (2002). In particular, Donoho (1995) showed that the minimax convergence rate of estimating f' under MISE is $n^{2(\alpha-1)/(1+2\alpha)}$ when $\alpha + 1/2 - 3/p > 0$.

Suppose we observe the Gaussian process (3) and we wish to estimate the mth derivative of f, $h = f^{(m)}$, under the pointwise risk

$$R_r(\hat{h}_n,\ f^{(m)},\ t_0) = E_f|\hat{h}_n(t_0) - f^{(m)}(t_0)|^r. \tag{19}$$

Again, $t_0 \in (0,\ 1)$ is any fixed point and $1 \le r < \infty$.

The minimax risk is defined in a similar way:

$$R^*_r(B^\alpha_{p,q}(M), m;\ t_0, n) = \inf_{\hat{h}_n} \sup_{f\in B^\alpha_{p,q}(M)} E_f|\hat{h}_n(t_0) - f^{(m)}(t_0)|^r. \tag{20}$$

The results on estimating $f^{(m)}$ under the pointwise ℓ^r risk parallel those on estimating f and are summarized as follows.

Theorem 6.

(i) *The minimax rate of convergence of estimating $f^{(m)}$ over the Besov class $B^\alpha_{p,q}(M)$ with $\nu \equiv \alpha - 1/p > m$ under the pointwise ℓ^r risk* (19) *is $n^{\frac{r(\nu-m)}{1+2\nu}}$, i.e.,*

$$0 < \varliminf_{n\to\infty} n^{\frac{r(\nu-m)}{1+2\nu}} R_r^*(B^\alpha_{p,q}(M), m;\ t_0, n) \leq \varlimsup_{n\to\infty} n^{\frac{r(\nu-m)}{1+2\nu}} R_r^*(B^\alpha_{p,q}(M), m;\ n) < \infty. \tag{21}$$

(ii) *If $\hat{h}_n$ is a rate-optimal estimator of $f^{(m)}$, then for any fixed $f \in B^\alpha_{p,q}(M)$,*

$$\varliminf_{n\to\infty} n^{\frac{r(\nu-m)}{1+2\nu}} E_f|\hat{h}_n(t_0) - f^{(m)}(t_0)|^r > 0. \tag{22}$$

That is, a rate-optimal estimator must attain the "flat rate" at every $f \in B^\alpha_{p,q}(M)$.

(iii) *The adaptive minimax rate of convergence is $(n/\log n)^{\frac{r(\nu-m)}{1+2\nu}}$. This rate can be attained adaptively by the soft threshold wavelet estimator given in* (16)−(17) *with threshold $\lambda = (r(2m+1)\log n)^{1/2}$.*

6. General ℓ^r Risk Oracle Inequalities

Oracle inequalities have been an effective tool in the study of asymptotic properties of wavelet estimators. See, e.g., Donoho and Johnstone (1994), Cai (1999), Antoniadis and Fan (2001) and Johnstone (1998). However, most of the results in the literature are restricted to the standard case of mean squared error and Gaussian noise. The results are not applicable to our problems.

In this section we first present a general ℓ^r-risk oracle inequality for arbitrary error distributions. This general inequality can be of independent interest. It is derived using the approach of optimal recovery. In the special case of Gaussian noise, the ℓ^r risk oracle inequality serves as a main tool for the proofs of Theorems 5 and 6.

Theorem 7.(ℓ^r Risk Oracle Inequality) *Let z_i be random variables with mean* 0 *and standard deviation* 1, *and let*

$$y_i = \theta_i + \sigma z_i \quad \text{for } i = 1, \ldots, n. \tag{23}$$

Let $\hat{\theta}_i = \eta_{\lambda\sigma}(y_i) = sgn(y_i)(|y_i| - \lambda\sigma)_+$ be a soft threshold estimator of θ_i. Then for any $1 \leq r < \infty$,

$$E|\hat{\theta}_i - \theta_i|^r \leq \min(|\theta_i|^r,\ 2^r\lambda^r\sigma^r) + 2^r\sigma^r E|z_i|^r I(|z_i| > \lambda), \tag{24}$$

$$E\sum_{i=1}^{n}|\hat{\theta}_i-\theta_i|^r \le \sum_{i=1}^{n}\left\{\min(|\theta_i|^r,\ 2^r\lambda^r\sigma^r)+2^r\sigma^r E|z_i|^r I(|z_i|>\lambda)\right\}. \tag{25}$$

In the "oracular" form,

$$E\sum_{i=1}^{n}|\hat{\theta}_i-\theta_i|^r \le 2^r\lambda^r\sum_{i=1}^{n}\min(|\theta_i|^r,\ \sigma^r)+2^r\sigma^r\sum_{i=1}^{n}E|z_i|^r I(|z_i|>\lambda). \tag{26}$$

The first term in (26) involves the optimal tradeoff between the signal ("bias") and noise ("variance"), and the second term is the risk bound for the case when all θ_i are zero. For most applications, the thresholding constant λ is appropriately chosen so the second term in (26) is sufficiently small. The oracle inequality turns the problem of bounding the ℓ^r risk into the problem of calculating the rth tail moment of the error distributions, which is often straightforward.

As an important special case, we consider below the case of Gaussian noise. Denote by $[r]$ the integer part of r.

Corollary 3. *Suppose in* (23) $z_i \sim N(0,\ 1)$ *for all* i *and suppose* $\lambda = (r(2m+1)\log n)^{1/2}$ *for some* $m \ge 0$*, then*

$$E|\hat{\theta}_i-\theta_i|^r \le \min\{|\theta_i|^r, 2^r(r(2m+1))^{r/2}(\log n)^{r/2}\sigma^r\}+H(n,r)n^{-\frac{r}{2}(2m+1)}\sigma^r, \tag{27}$$

$$\begin{aligned} E\sum_{i=1}^{n}|\hat{\theta}_i-\theta_i|^r \le& \sum_{i=1}^{n}\min\{|\theta_i|^r,\ 2^r(r(2m+1))^{r/2}(\log n)^{r/2}\sigma^r\} \\ &+H(n,r)n^{-\frac{r}{2}(2m+1)+1}\sigma^r, \end{aligned} \tag{28}$$

where $H(n,r)$*, given in* (44)*, is a polynomial of* $(\log n)^{1/2}$ *of degree* $[r]$ *and so* $H(n,r) \le C(r)\cdot(\log n)^{[r]/2}$ *for some constant* $C(r)>0$*.*

In particular, when $\lambda = (r\log n)^{1/2}$*, i.e.,* $m=0$*,*

$$E\sum_{i=1}^{n}|\hat{\theta}_i-\theta_i|^r \le \sum_{i=1}^{n}\min\{|\theta_i|^r,\ (4r)^{r/2}(\log n)^{r/2}\sigma^r\}+H(n,r)n^{-\frac{r}{2}+1}\sigma^r. \tag{29}$$

Remark. The inequality (29) extends the oracle inequality for the ℓ^2 risk given in Donoho and Johnstone (1994) to the general ℓ^r risk for $1 \le r < \infty$.

7. Proofs

We prove the results in the order of Theorems 4, 1, 2, 3, 7, 5 and 6. Throughout this section, C is a generic positive constant which may vary from place to place. We first collect some necessary tools.

7.1. Preparatory results

For Theorems 4 and 6, we need a result on the Besov norm.

Lemma 1. *Let the Besov norm be defined as in* (1) *with some fixed* $K > \alpha$. *Suppose* $g \in B^{\alpha}_{p,q}(I\!R)$ *is a compactly supported function. Let* $f(t) = ag(bt)$ *with* $a > 0$ *and* $b > 1$. *Suppose* f *is supported on* $[0,\ 1]$. *Then*

$$\|f\|_{B^{\alpha}_{p,q}([0,1])} \le ab^{\alpha-1/p}\|g\|_{B^{\alpha}_{p,q}(I\!R)}. \tag{30}$$

Proof. We prove the case of $p < \infty$ and $q < \infty$, other cases are similar.

It is easy to see that $\|f\|_p = ab^{-1/p}\|g\|_p$ and

$$\begin{aligned}\omega_{K,p}(f,h) &= (\int_0^{1-Kh} |\sum_{k=0}^{K}(-1)^k f(t+kh)|^p\ dt)^{1/p} \\ &= a(\int_0^{1-Kh} |\sum_{k=0}^{K}(-1)^k g(bt+kbh)|^p\ dt)^{1/p} \\ &\le ab^{-1/p}(\int_0^{\infty} |\sum_{k=0}^{K}(-1)^k g(t+kbh)|^p\ dt)^{1/p} \\ &= ab^{-1/p}\omega_{K,p}(g,bh).\end{aligned}$$

Hence,

$$\begin{aligned}\|f\|_{B^{\alpha}_{p,q}([0,1])} &= \|f\|_{L^p} + \left(\int_0^1 [h^{-\alpha}\omega_{K,p}(f,h)]^q \frac{dh}{h}\right)^{1/q} \\ &\le ab^{-1/p}\|g\|_p + ab^{-1/p}\left(\int_0^1 [h^{-\alpha}\omega_{K,p}(g,bh)]^q \frac{dh}{h}\right)^{1/q} \\ &\le ab^{-1/p}\|g\|_p + ab^{\alpha-1/p}\left(\int_0^{\infty} [t^{-\alpha}\omega_{K,p}(g,t)]^q \frac{dt}{t}\right)^{1/q} \\ &\le ab^{\alpha-1/p}\|g\|_{B^{\alpha}_{p,q}(I\!R)}.\end{aligned}$$

A main tool for the proof of Theorems 4 and 6 is a constrained risk inequality stated below. The risk inequality is a generalization to ℓ^r risk of an inequality introduced in Brown and Low (1996b) under mean squared error. A further generalization and its applications are presented in Cai, Low and Zhao (2001). The constrained risk inequality gives a lower bound for the ℓ^r risk at one parameter value subject to having a small risk at another parameter value. This type of constrained risk inequality is a useful technical tool for providing lower bounds for the cost of adaptation.

Let X be a random variable having either distribution P_{θ_0} with density f_{θ_0} or distribution P_{θ_1} with density f_{θ_1}, with respect to some dominating measure. For any estimator δ based on X its ℓ^r risk is defined by $R_r(\delta,\theta) = E_\theta|\delta(X) - \theta|^r$. Denote by $\kappa(x) = f_{\theta_1}(x)/f_{\theta_0}(x)$ the ratio of the two density functions. ($\kappa(x) = \infty$ for some x is possible, with the obvious interpretation $\kappa(x)f_{\theta_0}(x) = f_{\theta_1}(x)$.)

For $1 \le r < \infty$, denote by r^* the value satisfying $1/r + 1/r^* = 1$. Let $I_{r^*} = I_{r^*}(\theta_0,\ \theta_1) = (E_{\theta_0}\kappa^{r^*}(X))^{1/r^*}$ with obvious change for $r^* = \infty$. The quantity I_{r^*} is a measure of distance between the two distributions P_{θ_0} and P_{θ_1}.

Lemma 2. *Suppose $R_r(\delta,\ \theta_0) \le \epsilon_r^r$. Denote $\Delta = |\theta_1 - \theta_0|$.*

(i) *If $r > 1$ and $\Delta > \epsilon_r I_{r^*}$, then $R_r(\delta,\ \theta_1) \ge (\Delta - \epsilon_r I_{r^*})^r \ge \Delta^r(1 - r\epsilon_r I_{r^*}/\Delta)$.*

(ii) *Let $r = 1$ and suppose there exists a measurable set Λ_0 such that $\omega \equiv P_{\theta_1}(\Lambda_0) > 0$ and $\tilde I_\infty \equiv \|\kappa(x)\, I(x \in \Lambda_0)\|_\infty < \infty$, where the supnorm is taken with respect to P_{θ_0}. Suppose $\Delta = |\theta_1 - \theta_0| > \epsilon_1 \tilde I_\infty/\omega$, then $R_1(\delta,\ \theta_1) \ge \omega\,\Delta(1 - (\epsilon_1 \tilde I_\infty/\omega\,\Delta))$.*

Sketch of Proof. A more general version of this result, together with its proof, is presented in Cai, Low and Zhao (2001). For completeness, we give an outline of the proof of Lemma 2 for the case $r > 1$ here.

Jensen's inequality and the triangle inequality yield $(R_r(\delta,\ \theta_1))^{1/r} \ge |\theta_1 - \theta_0| - |E_{\theta_1}(\delta(X) - \theta_0)|$. It then follows from Hölder's inequality and the assumption $R_r(\delta,\ \theta_0) \le \epsilon_r^r$ that $|E_{\theta_1}(\delta(X) - \theta_0)| \le (E_{\theta_0}|\delta(X) - \theta_0|^r)^{1/r}(E_{\theta_0}\kappa^{r^*}(X))^{1/r^*} \le \epsilon_r I_{r^*}$ and so $R_r(\delta,\ \theta_1) \ge (\Delta - \epsilon_r I_{r^*})^r$.

7.2. Proof of the main results

Proof of Theorem 4. We first outline the main ideas. The constrained risk inequality given in Lemma 2 implies that if an estimator has a small ℓ^r risk ϵ_r^r at one parameter value θ_0 and $|\theta_1 - \theta_0| \gg \epsilon_r I_{r^*}$ then its risk at the other parameter value θ_1 must be "large". Now the assumption (10) means that the estimator $\hat f_n(t_0)$ has a small risk at $\theta_0 = f_0(t_0)$. If we can construct a sequence of functions $f_n \in B^\alpha_{p,q}(M)$ such that f_n is "close" to f_0 in the sense that $\|f_n - f_0\|_2^2$ is "small" (so that I_{r^*} is small) and at the same time $\Delta = |f_n(t_0) - f_0(t_0)|$ is "large", then it follows from the constrained risk inequality that $\hat f_n(t_0)$ must have a "large" risk at $\theta_1 = f_n(t_0)$. So the first step of the proof is a construction for such a sequence of functions f_n.

We divide the proof into two cases: $r > 1$ and $r = 1$.

Case (i).($r > 1$) Let g be a compactly supported function satisfying $g(0) > 0$, $\|g\|_2^2 > 0$ and $g \in B^\alpha_{p,q}(M - M')$. Such a function is easy to construct either directly or by using wavelets.

Denote by $b = 2(1 - 1/r)\|g\|_2^{-2}$ and let $\gamma_n = (n/(b\log B_n))^{\nu/(1+2\nu)}$ and $\beta_n = (n/(b\log B_n))^{1/(1+2\nu)}$. Then $\gamma_n^2\beta_n = n/b\log B_n$ and $\gamma_n^{-1}\beta_n^\nu = 1$. Let

$$f_n(t) = \gamma_n^{-1} g(\beta_n(t - t_0)) + f_0(t). \tag{31}$$

We show that f_n has the desired properties. It follows from Lemma 1 that $f_n \in B^\alpha_{p,q}(M)$ since $\|f_n\|_{B^\alpha_{p,q}} \le \gamma_n^{-1}\beta_n^\nu\|g\|_{B^\alpha_{p,q}} + \|f_0\|_{B^\alpha_{p,q}} \le M$. Note also that for sufficiently large n, say $n \geq N_1$, $\rho_n = n\|f_n - f_0\|_2^2 = n\gamma_n^{-2}\beta_n^{-1}\|g\|_2^2 = 2(1 - 1/r)\log B_n$ and $|f_n(t_0) - f_0(t_0)| = \gamma_n^{-1}g(0) = g(0)(b\log B_n/n)^{\nu/(1+2\nu)}$.

Write P_0^n and P_1^n for the probability measure associated with the white noise with drift process (3) with $f = f_0$ and $f = f_n$, respectively. Then a sufficient statistic for the family of measures $\{P_0^n,\ P_1^n\}$ is given by $T_n = \log(dP_1^n/dP_0^n)$. Note that for $n \geq N_1$

$$T_n \sim \begin{cases} N(-\rho_n/2,\ \rho_n) & \text{under } \ P_0^n \\ N(\rho_n/2,\ \rho_n) & \text{under } \ P_1^n, \end{cases}$$

see, for example, Brown and Low (1996b). Let $\delta_n = \hat{f}_n(t_0)$, $\theta_0 = f_0(t_0)$ and $\theta_1 = f_n(t_0)$. It follows from the assumption (10) of the theorem that there exist constants $C,\ N_2 > 0$ such that for all $n \geq N_2$, $E_{\theta_0}|\delta_n - \theta_0|^r \leq Cn^{-r\nu/(1+2\nu)}\ B_n^{-1}$. Let $\delta_n^* = E(\delta_n|T_n)$. Since T_n is sufficient for $\{P_0^n,\ P_1^n\}$, it follows from the Rao-Blackwell Theorem that for $i = 0,\ 1$, $E_{\theta_i}|\delta_n^* - \theta_i|^r \leq E_{\theta_i}|\delta_n - \theta_i|^r$. Hence $E_{\theta_0}|\delta_n^* - \theta_0|^r \leq Cn^{-r\nu/(1+2\nu)}\ B_n^{-1}$. We now apply Lemma 2(i) with $\theta_0 = f_0(t_0)$, $\theta_1 = f_n(t_0)$, f_{θ_0} the density of T_n under P_0^n and f_{θ_1} the density of T_n under P_1^n. Noting that $(r-1)(r^*-1) = 1$, we have, for $n \geq N_1$, $I_{r^*}(\theta_0,\ \theta_1) = e^{\rho_n\cdot(r^*-1)/2} = e^{2(1-1/r)\log B_n\cdot(r^*-1)/2} = B_n^{1/r}$. Lemma 2 now yields for $n \geq \max(N_1,\ N_2)$,

$$\begin{aligned} E_{\theta_1}|\delta_n^* - \theta_1|^r &\geq \left(\frac{g(0)}{\gamma_n}\right)^r \left(1 - rC^{\frac{1}{r}}n^{-\frac{\nu}{1+2\nu}}\ B_n^{-\frac{1}{r}} \cdot B_n^{\frac{1}{r}} \cdot (g(0))^{-1}\gamma_n\right) \\ &= (bg(0))^{\frac{r\nu}{1+2\nu}} \left(\frac{\log B_n}{n}\right)^{\frac{r\nu}{1+2\nu}} \left(1 - rC^{\frac{1}{r}}(g(0))^{-1}(b\log B_n)^{-\frac{\nu}{1+2\nu}}\right) \\ &= (bg(0))^{\frac{r\nu}{1+2\nu}} \left(\frac{\log B_n}{n}\right)^{\frac{r\nu}{1+2\nu}} (1 + o(1)). \end{aligned}$$

Hence, $E_{\theta_1}|\delta_n - \theta_1|^r \geq E_{\theta_1}|\delta_n^* - \theta_1|^r \geq (bg(0))^{\frac{r\nu}{1+2\nu}}(\log B_n/n)^{\frac{r\nu}{1+2\nu}}(1 + o(1))$.

Case (ii).($r = 1$) In this case, let $f_n(t) = \gamma_n^{-1}g(\beta_n(t - t_0)) + f_0(t)$ with the function g same as in Case (i) and $\gamma_n = (n/\log B_n)^{\nu/(1+2\nu)}$ and $\beta_n = (n/\log B_n)^{1/(1+2\nu)}$. Then $f \in B_{p,q}^\alpha(M)$ and $\rho_n = n\|f_n - f_0\|_2^2 = \log B_n$. Noting again that T_n is sufficient for $\{P_0^n,\ P_1^n\}$, we may apply Lemma 2(ii) here. Using the same notation as in Lemma 2(ii), $\kappa(x) = e^x$. Let $\Lambda_0 = \{T_n \leq \rho_n\}$, then $\omega = P_{\theta_1}(\Lambda_0) = P_{\theta_1}(T_n \leq \rho_n) = \Phi(\rho_n^{1/2}/2) \geq 1/2$, and $\tilde{I}_\infty = \|\kappa(T_n)I(T_n \leq \rho_n)\|_\infty = e^{\rho_n} = B_n$. It follows from the assumption (10) in the theorem with $r = 1$ that $E_{\theta_0}|\delta_n - \theta_0| \leq Cn^{-\nu/(1+2\nu)}\ B_n^{-1}$. Noting that $|\theta_1 - \theta_0| = g(0)\gamma_n^{-1}$, Lemma 2(ii) now yields

$$\begin{aligned} E_{\theta_1}|\delta_n - \theta_1| &\geq \frac{1}{2}g(0)\gamma_n^{-1}\left(1 - \frac{Cn^{-\frac{\nu}{1+2\nu}}B_n^{-1} \cdot B_n}{\frac{1}{2}g(0)\gamma_n^{-1}}\right) \\ &= \frac{1}{2}g(0)\left(\frac{\log B_n}{n}\right)^{\frac{\nu}{1+2\nu}}(1 + o(1)). \end{aligned}$$

The theorem is proved.

Proof of Theorem 1. Theorem 1 follows directly from Theorem 4. Suppose (6) does not hold for some estimator $\hat{f}_n$ and some sequence $B_n \to \infty$. Then

$$\overline{\lim_{n\to\infty}}\, n^{\frac{r\nu}{1+2\nu}} B_n \sup_{f\in B^\alpha_{p,q}(M)} E_f|\hat{f}_n(t_0) - f(t_0)|^r < \infty. \tag{32}$$

Then for any $f_0 \in B^\alpha_{p,q}(M)$, $\overline{\lim}_{n\to\infty}\, n^{\frac{r\nu}{1+2\nu}} B_n E_{f_0}|\hat{f}_n(t_0) - f_0(t_0)|^r < \infty$. Theorem 4 yields that $\underline{\lim}_{n\to\infty}(n/\log B_n)^{\frac{r\nu}{1+2\nu}} \sup_{f\in B^\alpha_{p,q}(M)} E_f|\hat{f}_n(t_0) - f(t_0)|^r > 0$, and thus $\overline{\lim}_{n\to\infty}\, n^{\frac{r\nu}{1+2\nu}} \sup_{f\in B^\alpha_{p,q}(M)} E_f|\hat{f}_n(t_0) - f(t_0)|^r = \infty$, which contradicts the assumption made in (32).

Proof of Theorem 2. The convergence rate of $n^{r\nu/(1+2\nu)}$ can be attained by a linear wavelet estimator. Let $\{\phi,\ \psi\}$ be a pair of compactly supported wavelets generating an orthonormal basis in $L^2[0,1]$. Let ψ be K-regular with $K > \alpha$, and let $\tilde{y}_{j_0,k}$ and $y_{j,k}$ be the empirical wavelet coefficients as given in (13) and (14). Denote by J_1 the largest integer satisfying $2^{J_1} \le n^{1/(1+2\nu)}$ and let the estimator of f be

$$\hat{f}_n(t) = \sum_{k=1}^{2^{j_0}} \tilde{y}_{j_0k}\phi_{j_0k}(t) + \sum_{j=j_0}^{J_1-1}\sum_{k=1}^{2^j} y_{jk}\psi_{jk}(t). \tag{33}$$

For simplicity, assume the lowest level j_0 is chosen so that all the boundary wavelets are vanishing at t_0. This can be easily accomplished for any fixed $t_0 \in (0,1)$. See, e.g., Cohen, Daubechies, Jawerth and vial (1993).

We now show that this estimator is rate-optimal over $B^\alpha_{p,q}(M)$. Applying the elementary inequality

$$E|\sum_{i=1}^n X_i|^r \le (\sum_{i=1}^n (E|X_i|^r)^{1/r})^r, \tag{34}$$

where X_i are random variables, we have

$$\begin{aligned}
&E_f|\hat{f}_n(t_0)-f(t_0)|^r = E\left|\sum_{k=1}^{2^{j_0}}(\hat{\xi}_{j_0k}-\xi_{j_0k})\phi_{j_0k}(t_0)+\sum_{j=j_0}^{\infty}\sum_{k=1}^{2^j}(\hat{\theta}_{jk}-\theta_{jk})\psi_{jk}(t_0)\right|^r\\
&\le \Bigg[\sum_{k=1}^{2^{j_0}}|\phi_{j_0k}(t_0)|(E|\tilde{y}_{j_0k}-\xi_{j_0k}|^r)^{\frac{1}{r}} + \sum_{j=j_0}^{J_1-1}\sum_{k=1}^{2^j}|\psi_{jk}(t_0)|(E|y_{jk}-\theta_{jk}|^r)^{\frac{1}{r}}\\
&\quad + \sum_{j=J_1}^{\infty}\sum_{k=1}^{2^j}|\theta_{jk}\psi_{jk}(t_0)|\Bigg]^r\\
&\equiv (Q_1+Q_2+Q_3)^r.
\end{aligned}$$

It is easy to see that

$$Q_1 = \sum_{k=1}^{2^{j_0}} |\phi_{j_0 k}(t_0)|(E|\tilde{y}_{j_0 k} - \xi_{j_0 k}|^r)^{1/r} = O(n^{-1/2}). \tag{35}$$

Since the wavelets are compactly supported, say on $[-L/2, L/2]$, there are at most L basis functions ψ_{jk} at each resolution level j that are nonvanishing at t_0. Let $K_j(t_0) = \{k : \psi_{j,k}(t_0) \neq 0\}$. Then $\text{Card}(K_j(t_0)) \le L$. For $f \in B^\alpha_{p,q}(M)$, using the wavelet sequence norm (2), we have $|\theta_{j,k}| \le M_* 2^{-j(\alpha+1/2-1/p)}$ for all (j,k), with some constant $M_* > 0$ not depending on f. Hence

$$Q_3 = \sum_{j=J_1}^{\infty} \sum_{k \in K_j(t_0)} |\theta_{jk}\psi_{jk}(t_0)| \le \sum_{j=J_1}^{\infty} L\|\psi\|_\infty 2^{j/2} M_* 2^{-j(\alpha+1/2-1/p)} = Cn^{-\nu/(1+2\nu)}. \tag{36}$$

Let $b_r = E|Z|^r$ where $Z \sim N(0,1)$. Then $E|y_{jk} - \theta_{jk}|^r = n^{-r/2} b_r$. So,

$$Q_2 = \sum_{j=j_0}^{J-1} \sum_{k \in K_j(t_0)} 2^{j/2}\|\psi\|_\infty (E|y_{jk} - \theta_{jk}|^r)^{1/r} \le C \sum_{j=j_0}^{J_1-1} 2^{j/2} n^{-1/2} = Cn^{-\nu/(1+2\nu)}. \tag{37}$$

Combining (35), (36) and (37), we have $E_f|\hat{f}_n(t_0) - f(t_0)|^r \le Cn^{-r\nu/(1+2\nu)}$.

Proof of Theorem 3. The estimator given in (33) depends on the smoothness index $(\alpha,\ p,\ q)$ only through $\nu = \alpha - 1/p$. So the estimator attains the optimal rate simultaneously over all $B^\alpha_{p,q}(M)$ with a fixed value of $\alpha - 1/p = \nu$.

Proof of Theorem 7. It suffices to consider the univariate case in which one observes $y = \theta + \sigma z$ and wishes to estimate θ under the risk $E_\theta|\delta(y) - \theta|^r$.

To derive the oracle inequality we first consider an optimal recovery problem. Suppose

$$y = \theta + \beta \cdot u, \tag{38}$$

where u is deterministic and $|u| \le 1$. We wish to estimate θ with small ℓ^r error.

Lemma 3. *Suppose y is observed as in* (38). *Let $\hat{\theta} = \eta_\beta(y) = sgn(y)(|y| - \beta)_+$. Then for $r \ge 1$,*

$$\sup_{|u|\le 1} |\hat{\theta} - \theta|^r \le \min(|\theta|^r,\ 2^r\beta^r). \tag{39}$$

Proof. It is easy to see that $\hat{\theta}$ has the same sign as θ and $|\hat{\theta}| \le |\theta|$. We first show that

$$\sup_{|u|\le 1} |\hat{\theta} - \theta|^r \le |\theta|^r. \tag{40}$$

When $\hat{\theta} = 0$, (40) holds trivially. When $\hat{\theta} \neq 0$, since $\hat{\theta}$ and θ have the same sign and $|\hat{\theta}| \le |\theta|$, $|\hat{\theta} - \theta|^r \le (|\theta| - |\hat{\theta}|)^r \le |\theta|^r$. Rewriting $\hat{\theta}$ as $\hat{\theta} = y - \text{sgn}(y) \cdot \min(\beta, |y|)$, we have $|\hat{\theta}-y| \le \beta$. So, applying the elementary inequality $|a+b|^r \le 2^{r/r^*}(|a|^r + |b|^r)$, we have for $r > 1$,

$$|\hat{\theta}-\theta|^r \le 2^{r/r^*}(|\hat{\theta}-y|^r + |y-\theta|^r) \le 2^r\beta^r. \tag{41}$$

For $r=1$, the triangle inequality yields $|\hat{\theta}-\theta| \le |\hat{\theta}-y| + |y-\theta| \le 2\beta$.

We now use Lemma 3 to prove Theorem 7. First separate the risk into two parts:

$$E_\theta|\eta_{\lambda\sigma}(y)-\theta|^r = E_\theta|\eta_{\lambda\sigma}(y)-\theta|^r I(|z| \le \lambda) + E_\theta|\eta_{\lambda\sigma}(y)-\theta|^r I(|z| > \lambda).$$

We use Lemma 3 to bound the first term. Write $y = \theta + \sigma z = \theta + \lambda\sigma z_1$ with $z_1 = z/\lambda$. Then $|z| \le \lambda$ is equivalent to $|z_1| \le 1$ and so Lemma 3 yields that

$$E_\theta|\eta_{\lambda\sigma}(y)-\theta|^r I(|z| \le \lambda) \le \sup_{|z_1|\le 1} |\eta_{\lambda\sigma}(y)-\theta|^r \le \min(|\theta|^r,\ 2^r\lambda^r\sigma^r). \tag{42}$$

For the second term, applying the inequality $|a+b|^r \le 2^{r/r^*}(|a|^r + |b|^r)$ and noting that $|\eta_{\lambda\sigma}(y)-y| \le \lambda\sigma$ we have, for $r>1$,

$$\begin{aligned} E_\theta|\eta_{\lambda\sigma}(y)-\theta|^r I(|z| > \lambda) &\le 2^{r/r^*}E_\theta\{(|\eta_{\lambda\sigma}(y)-y|^r + |y-\theta|^r)I(|z|>\lambda)\} \\ &\le 2^{r/r^*}\{\lambda^r\sigma^r P(|z|>\lambda) + \sigma^r E|z|^r I(|z|>\lambda)\} \\ &\le 2^r\sigma^r E|z|^r I(|z|>\lambda). \end{aligned} \tag{43}$$

The case of $r=1$ can be verified in a similar way. The inequality (25) follows by combining (42) and (43).

When the distribution of z is given, $E|z|^r I(|z|>\lambda)$ can be evaluated explicitly. Suppose $z \sim N(0,\ 1)$. Let $\lambda \ge 1$, then

$$\begin{aligned} E|z|^r I(|z|>\lambda) = 2Ez^r I(z>\lambda) &= \left(\frac{2}{\pi}\right)^{\frac{1}{2}} \sigma^r \int_\lambda^\infty x^r e^{-\frac{x^2}{2}}\, dx \\ &\le \left(\frac{2}{\pi}\right)^{\frac{1}{2}} \sigma^r \int_\lambda^\infty x^{[r]+1} e^{-\frac{x^2}{2}}\, dx. \end{aligned}$$

Integration by parts yields, for odd k,

$$\int_\lambda^\infty x^k e^{-\frac{x^2}{2}}\, dx = \{\lambda^{k-1} + (k-1)\lambda^{k-3} + \cdots + (k-1)(k-3)\cdots 2\} \cdot e^{-\lambda^2/2},$$

and for even k,

$$\begin{aligned} \int_\lambda^\infty x^k e^{-\frac{x^2}{2}}\, dx &= \{\lambda^{k-1} + (k-1)\lambda^{k-3} + \cdots + (k-1)(k-3)\cdots 3\lambda\} \cdot e^{-\lambda^2/2} \\ &\quad + (k-1)(k-3)\cdots 1 \int_\lambda^\infty e^{-\frac{x^2}{2}}\, dx \\ &\le \{\lambda^{k-1} + (k-1)\lambda^{k-3} + \cdots + (k-1)(k-3)\cdots 3\lambda + 1\} \cdot e^{-\lambda^2/2}. \end{aligned}$$

Let

$$H(\lambda,r)=\begin{cases}2^{r+1/2}\pi^{-1/2}\{\lambda^{[r]}+[r]\lambda^{[r]-2}+\cdots+[r]\cdot([r]-2)\cdots 2\} & \text{if } [r] \text{ is even},\\ 2^{r+1/2}\pi^{-1/2}\{\lambda^{[r]}+[r]\lambda^{[r]-2}+\cdots+[r]\cdot([r]-2)\cdots 3\lambda+1\} & \text{if } [r] \text{ is odd}.\end{cases} \tag{44}$$

It is easy to see that with $\lambda=(r(2m+1)\log n)^{1/2}$, one has $H(n,r)\le C(r)(\log n)^{[r]/2}$ for some constant $C(r)>0$, and $E_\theta|\eta_\lambda(y)-\theta|^r\le\min(|\theta|^r, 2^r\lambda^r\sigma^r)+C(r)(\log n)^{[r]/2}$ $n^{-\frac{r}{2}(2m+1)}\sigma^r$.

Proof of Theorem 5. The proof is similar to that of Theorem 2. Let $\hat{f}$ be given as in (16)-(17) with $\lambda=(r\log n)^{1/2}$. Again, for simplicity, we assume the lowest level j_0 is chosen so that all the boundary wavelets are vanishing at t_0. Only very minor modifications of the proof is needed if this is not the case.

Applying the inequality (34), we have

$$\begin{aligned}E|\hat{f}_n(t_0)-f(t_0)|^r &= E\left|\sum_{k=1}^{2^{j_0}}(\hat{\xi}_{j_0k}-\xi_{j_0k})\phi_{j_0k}(t_0)+\sum_{j=j_0}^{\infty}\sum_{k=1}^{2^j}(\hat{\theta}_{jk}-\theta_{jk})\psi_{jk}(t_0)\right|^r\\ &\le \left[\sum_{k=1}^{2^{j_0}}|\phi_{j_0k}(t_0)|(E|\hat{\xi}_{j_0k}-\xi_{j_0k}|^r)^{1/r}+\sum_{j=j_0}^{J-1}\sum_{k=1}^{2^j}|\psi_{j_0k}(t_0)|(E|\hat{\theta}_{jk}-\theta_{jk}|^r)^{1/r}\right.\\ &\quad\left.+\sum_{j=J}^{\infty}\sum_{k=1}^{2^j}|\theta_{jk}\psi_{jk}(t_0)|\right]^r\\ &\equiv (Q_1+Q_2+Q_3)^r.\end{aligned}$$

Similar as in the proof of Theorem 2, both Q_1 and Q_3 are small:

$$Q_1=O(n^{-1/2}) \quad \text{and} \quad Q_3=O(n^{-\nu}). \tag{45}$$

Again, denote $K_j(t_0)=\{k:\ \psi_{j,k}(t_0)\neq 0\}$. Then $\mathrm{Card}(K_j(t_0))\le L$, where L is the support length of ψ. For $f\in B^\alpha_{p,q}(M)$, $|\theta_{j,k}|\le C2^{-j(\alpha+1/2-1/p)}$ for all (j,k), with some constant $C>0$ not depending on f. Let J_2 be the largest integer satisfying $2^{J_2}\le(n/\log n)^{1/(1+2\nu)}$. Separate a simple bound of Q_2 into two parts:

$$\begin{aligned}Q_2 &\le \sum_{j=j_0}^{J_2-1}\sum_{k\in K_j(t_0)}2^{\frac{j}{2}}\|\psi\|_\infty(E|\hat{\theta}_{jk}-\theta_{jk}|^r)^{1/r}+\sum_{j=J_2}^{J-1}\sum_{k\in K_j(t_0)}2^{\frac{j}{2}}\|\psi\|_\infty(E|\hat{\theta}_{jk}-\theta_{jk}|^r)^{1/r}\\ &\equiv Q_{21}+Q_{22}.\end{aligned}$$

Applying the oracle inequality (27) in Corollary 3 with $m=0$ and $\sigma=n^{-1/2}$, together with the elementary inequality $(a+b)^{1/r}\le a^{1/r}+b^{1/r}$ for $a,b\ge 0$, we have

$$Q_{21}\le C\sum_{j=j_0}^{J_2-1}2^{\frac{j}{2}}(\log n)^{\frac12}n^{-\frac12}+C(\log n)^{\frac12}n^{-\frac12}=C\left(\frac{n}{\log n}\right)^{-\frac{\nu}{1+2\nu}}(1+o(1)), \tag{46}$$

$$Q_{22} \le C \sum_{j=J_2}^{J-1} 2^{\frac{j}{2}} 2^{-j(\alpha+\frac{1}{2}-\frac{1}{p})} + C(\log n)^{\frac{1}{2}} n^{-\frac{1}{2}} = C\left(\frac{n}{\log n}\right)^{-\frac{\nu}{1+2\nu}} (1+o(1)). \quad (47)$$

Combining (45), (46) and (47), we have $E_f|\hat{f}_n(t_0)-f(t_0)|^r \le C(n/\log n)^{-r\nu/(1+2\nu)}$.

Proof of Theorem 6. The proof of this theorem is similar to the combinations of the proofs of Theorems 1 − 5. We only highlight the main changes here. The key is to prove a similar result as in Theorem 4 for estimating derivatives.

Theorem 8. *Let $B_n \to \infty$, $n/\log B_n \to \infty$. Suppose $\nu = \alpha - 1/p > m$. Let $\hat{h}_n$ be an estimator of $f^{(m)}$ and $f_0 \in B^\alpha_{p,q}(M')$ with $M' < M$. If*

$$\overline{\lim_{n\to\infty}}\, n^{r(\nu-m)/(1+2\nu)} B_n E_{f_0}|\hat{h}_n(t_0) - f_0^{(m)}(t_0)|^r < \infty, \quad (48)$$

$$\underline{\lim_{n\to\infty}} \left(\frac{n}{\log B_n}\right)^{r(\nu-m)/(1+2\nu)} \sup_{f\in B^\alpha_{p,q}(M)} E_f|\hat{h}_n(t_0) - f^{(m)}(t_0)|^r > 0. \quad (49)$$

To prove Theorem 8, let $g \in B^\alpha_{p,q}(I\!R,\ M - M')$ be a compactly supported function on $I\!R$ with $g^{(m)}(t_0) > 0$ and $\|g\|_2^2 > 0$. Such a function is easy to construct. Let the function sequence f_n be given as in (31) in the proof of Theorem 4, with the same choices of γ_n and β_n. Denote by $\theta_0 = f_0^{(m)}(t_0)$ and $\theta_1 = f_n^{(m)}(t_0) = \gamma_n^{-1}\beta_n^m g^{(m)}(t_0) + f_0^{(m)}(t_0)$. Then $|\theta_1 - \theta_0| = g^{(m)}(t_0)(b\log B_n/n)^{\frac{\nu-m}{1+2\nu}}$. Following the same steps as in the proof of Theorem 4 we have, for sufficiently large n, $E_{f_n}|\hat{h}_n(t_0) - f_n^{(m)}(t_0)|^r \ge C(\log B_n/n)^{\frac{r(\nu-m)}{1+2\nu}}(1+o(1))$.

Theorem 8 yields immediately part (ii) of Theorem 6 as well as the lower bounds on the minimax convergence rate and adaptive minimax convergence rate. The attainment of the two lower bounds can be shown in a similar way as in the proofs of Theorems 2 and 5.

Let the estimator $\hat{h}_{1,n} = (\hat{f}_n)^{(m)}$ where $\hat{f}_n$ is the projection estimator defined in (33), and let $\hat{h}_{2,n} = (\hat{f}_n)^{(m)}$ where $\hat{f}_n$ is the soft threshold estimator given in equations (16)-(17) with $\lambda = (r(2m+1)\log n)^{1/2}$. Using the same arguments as in the proofs of Theorems 2 and 5, it is easy to show that $\hat{h}_{1,n}$ attains the minimax convergence rate of $n^{r(\nu-m)/(1+2\nu)}$ under the pointwise risk (19) as an estimator of $f_n^{(m)}$ over $f \in B^\alpha_{p,q}(M)$ when the smoothness parameters (α, p, q) are fixed and known, and $\hat{h}_{2,n}$ attains the adaptive minimax convergence rate of $(n/\log n)^{r(\nu-m)/(1+2\nu)}$ over $B^\alpha_{p,q}(M)$ for unknown smoothness parameters (α, p, q).

Acknowledgement

This research supported in part by NSF Grant DMS-0072578.

References

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

Antoniadis, A. and Fan, J. (2001). Regularization of wavelet approximations. *J. Amer. Stat. Assoc.* **96**, 939-967.

Birgé, L. and Massart, P. (1995). Estimation of integral functionals of a density. *Ann. Statist.* **23**, 11-29.

Brown, L. D., Cai, T., Low, M. G. and Zhang, C. (2002). Asymptotic equivalence theory for nonparametric regression with random design. *Ann. Statist.* **30**, 688-707.

Brown, L. D. and Low, M. G. (1991). Information inequality bounds on the minimax risk (with an application to nonparametric regression). *Ann. Statist.* **19**, 239-337.

Brown, L. D. and Low, M. G. (1996a). Asymptotic equivalence of nonparametric regression and white noise. *Ann. Statist.* **24**, 2384-98.

Brown, L. D. and Low, M. G. (1996b). A constrained risk inequality with applications to nonparametric functional estimation. *Ann. Statist.* **24**, 2524-35.

Brown, L. D., Low, M. G. and Zhang, C. (2000). An asymptotic equivalence map between nonparametric density estimation and white noise. Technical Report, Department of Statistics, University of Pennsylvania.

Brown, L. D., Low, M. G. and Zhao, L. (1997). Superefficiency in nonparametric functional estimation. *Ann. Statist.* **25**, 2607-25.

Cai, T. (1999). Adaptive wavelet estimation: a block thresholding and oracle inequality approach. *Ann. Statist.* **27**, 898-924.

Cai, T. (2000). On adaptability and information-pooling in nonparametric function estimation. Technical Report, Department of Statistics, University of Pennsylvania. Submitted.

Cai, T. (2002). On adaptive wavelet estimation of a derivative and other related linear inverse problems. *J. Statist. Plann. Inference* **108**, 329-349.

Cai, T. and Low. M. (2002). On modulus of continuity and adaptability in nonparametric functional estimation. Technical Report, Department of Statistics, University of Pennsylvania.

Cai, T., Low, M. G. and Zhao, L. (2001). Tradeoffs between global and local risks in nonparametric function estimation. Technical Report, Department of Statistics, University of Pennsylvania. Available at www-stat.wharton.upenn.edu/~tcai/

Cohen, A., Daubechies, I., Jawerth, B. and Vial, P. (1993). Multiresolution analysis, wavelets, and fast algorithms on an interval. *Comptes Rendus Acad. Sci. Paris* (A). **316**, 417-421.

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

Daubechies, I. (1994). Two recent results on wavelets: wavelet bases for the interval, and biorthogonal wavelets diagonalizing the derivative operator. In *Recent Advances in Wavelet Analysis*, (Edited by Schumaker L. L. and Webb G.), 237-258. Academic Press.

DeVore, R. and Popov, V. (1988). Interpolation of Besov spaces. *Trans. Amer. Math. Soc.* **305**, 397-414.

Donoho, D. L. (1995). Nonlinear solution of linear inverse problems by wavelet-vaguelette decomposition. *Appl. Comput. Harm. Anal.* **2**, 101-126.

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. (1998). Minimax estimation via wavelet shrinkage. *Ann. Statist.* **26**, 879-921.

Donoho, D. L. and Low, M. G. (1992). Renormalization exponents and optimal pointwise rates of convergence. *Ann. Statist.* **20**, 944-970.

Farrell, R. H. (1972). On the best obtainable asymptotic rates of convergence in estimation of a density function at a point. *Ann. Math. Statist.* **43**, 170-180.

Hasminskii, R. Z. (1979). Lower bound for the risks of nonparametric estimates of the mode. In *Contribution to Statistics: Jaroslav Hájek Memorial Volume* (Edited by J. Jureckova), 91-97. Academia, Prague.

Ibragimov, I. A. and Hasminskii, R. Z. (1984). Nonparametric estimation of the values of a linear functional in Gaussian white noise. *Theory Probab. Appl.* **29**, 18-32.

Johnstone, I. M. (1998). Oracle inequalities and nonparametric function estimation. *Documenta Mathematica: Extra Volume ICM* **III**, 267-278.

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.

Lepski, O. V. (1990). On a problem of adaptive estimation in white Gaussian noise. *Theory Probab. Appl.* **35**, 454-466.

Lepski, O. V. and Spokoiny, V. G. (1997). Optimal pointwise adaptive methods in nonparametric estimation. *Ann. Statist.* **25**, 2512-2546.

Low, M. G. (1992). Renormalization and white noise approximation for nonparametric functional estimation problems. *Ann. Statist.* **20**, 545-554.

Meyer, Y. (1991). Ondelettes sur l'intervalle. *rev. Mat. Iberoamericana* **7**, 115-133.

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

Nussbaum, M. (1996). Asymptotic equivalence of density estimation and Gaussian white noise. *Ann. Statist.* **24**, 2399–2430.

Stone, C. J. (1980). Optimal rate of convergence for nonparametric estimators. *Ann. Statist.* **8**, 1348-1360.

Triebel, H. (1983). *Theory of Function Spaces.* Birkhäuser Verlag, Basel.

Triebel, H. (1992). *Theory of Function Spaces II.* Birkhäuser Verlag, Basel.

Tsybakov, A. B. (1998). Pointwise and sup-norm sharp adaptative estimation of functions on the Sobolev classes. *Ann. Statist.* **26**, 2420-2469.

Zhang, C. (2000). General empirical Bayes wavelet methods and exactly adaptive minimax estimation. Technical Report, Department of Statistics, Rutgers University.

Department of Statistics, The Wharton School, University of Pennsylvania, Philadelphia, PA 19104, U.S.A.

E-mail: tcai@wharton.upenn.edu

(Received February 2002; accepted April 2003)