Nonparametric Wavelet based Minimax Estimation in non Gaussian Models

F. Götze and B.A. Zalesky[1]

Abstract

We investigate nonparametric estimation procedures for images and functions with discontinuities in the presence of additive noise using wavelet decomposition methods. For functions and images belonging to Besov type spaces we derive bounds for the minimax risk in the class of linear and the class of arbitrary estimators for *non Gaussian* additive error models. The bounds are sharp for the linear minimax risks under second order moment assumptions on the noise. They are sharp up to a logarithmic factor for the general minimax risks under exponential moment conditions for the noise. Applications are made to quantization and image compression problems.

Key Words and Phrases. Haar bases, wavelets, minimax risk, linear estimators, threshold estimators, quantization errors.

[1]Supported by Deutsche Forschungsgemeinschaft, Sonderforschungsbereich 343 "Diskrete Strukturen in der Mathematik"

1 Introduction

In this paper the properties of the linear estimators and non-linear wavelet threshold estimators of functions (or images) are studied in the presence of *non Gaussian* additive noise. Bounds for the minimax risks are derived which are based on the tail and local dependence structure of the error models.

Let $I = [0,1]^d$ denote the unit cube of the Euclidean space R^d. For an integer r and a displacement $h \in R^d$ let $\Delta_h^r(f(x))$ denote the r-th order difference of f taken at the points $x + lh, l = 0, \ldots, r$. Using these differences one can define the r-th modulus of smoothness in $L^p(I)$ as

$$w_r(f,t)_p \stackrel{\text{def}}{=} \left(\sup_{|h| \le t} \int_I |\Delta_h^r(f(x))|^p \mathbf{1}_{\{x+rh \in I\}} dx \right)^{1/p},$$

and the Besov space $B_{p,q}^{\alpha,r}, r < \alpha + 1$ as the set of equivalence classes of measurable functions such that

$$||f||_{B_{p,q}^{\alpha,r}} \stackrel{\text{def}}{=} ||f||_{L^p(I)} + \left(\int_0^\infty \left[t^{-\alpha} w_r(f,t)_p \right]^q t^{-1} dt \right)^{1/q}$$

is finite.

Let $J_m \in I$ be dyadic cubes with sidelength 2^{-m}. Denote by D_m the set of all J_m and for some previously fixed integer l the union $D = \bigcup_{m=0}^l D_m$. We shall consider functions f belonging to subspaces N_l resp. $\bar{N}_l$ of the Besov spaces $B_{p,q}^{\alpha,r}$ resp. $B_{\tau,\tau}^{\alpha,r}$, $\tau = (\alpha/d + 1/2)^{-1}$, generated by the elements up to resolution 2^{-l} of an orthogonal compactly supported wavelet basis $\phi_{J_m}(x) = \phi(2^m x - j)$ i.e.

$$f = \sum_{m=0}^{l} \textstyle\sum_{J_m \in D_m} \beta_{J_m} \phi_{J_m}.$$

We shall investigate functions sampled with random noise ξ, i.e. we study the discrete time random processes y indexed by J_l given by the coefficients β_{J_l} sampled with random noise ξ_{J_l}

$$y_{J_l} = \beta_{J_l} + \xi_{J_l}.$$

Let ξ_{J_l} denote mean zero random variables with finite second moments $\mathbf{E}\xi_{J_l}^2 < \infty$ which are not necessary identically distributed. Let $B_{l,M}$ resp. $\bar{B}_{l,M}$ denote the Besov balls in N_l resp. $\bar{N}_l$ defined by $||f||_{B_{p,q}^{\alpha,r}} < M$ resp. $||f||_{B_{\tau,\tau}^{\alpha,r}} < M$. We shall investigate the minimax risks

$$\mathbf{R}_L(M,\Theta) = \sup_{\xi \in \Theta} \inf_{\hat{f}-\text{linear}} \sup_{B_{l,M}} \mathbf{E}||f - \hat{f}||_{L^2(I)}^2,$$

in the class of linear estimators, and

$$\mathbf{R}_{NL}(M,\Theta) = \sup_{\xi \in \Theta} \inf_{\hat{f}} \sup_{\bar{B}_{l,M}} \mathbf{E}||f - \hat{f}||_{L^2(I)}^2,$$

for the class of all measurable estimators.

First we investigate linear estimators in the case of independent random noise ξ with bounded second moments $\mathbf{E}\xi_{J_l}^2 \leq R$. For functions f belonging to $B_{l,M}$ and represented by Haar bases it is shown that
$\mathbf{R}_L(M,\Theta) \leq C2^{-\gamma}$, $\gamma = 2d\alpha l/(d+2\alpha)$, for $p \geq 2$ and $0 < \alpha < 1/p$, resp. for $1 \leq p < 2$ and $d(1/p - 1/2) < \alpha < 1/p$. In Theorem 2.2 and Corollary 2.3 we prove estimates for $\mathbf{R}_L(M,\Theta)$ in the case of correlated random noise.

In Theorem 2.4 we prove for Haar bases and regular wavelet bases
$\mathbf{R}_{NL}(M,\Theta) \leq C2^{-\gamma}l^{\gamma}$, $\gamma = 2\alpha dl/(2\alpha + d)$. provided that the random noise satisfies the Bernstein type exponential moment condition. This estimate is optimal up to a logarithmic factor. Furthermore estimates of $\mathbf{R}_{NL}(M,\Theta)$ are obtained for the case of random noise with finite polynomial moments
$\mathbf{E} \mid \xi_{J_l} \mid^p < R$, $p \geq 2$. These bounds are proved using the (nonlinear) threshold estimator obtained by truncating the higher order wavelet coefficients which are above a certain threshold to zero. Moreover, for both types of random noises we have precise estimates of the convergence rate of the square risk $\mathbf{E}\|f - \hat{f}_\epsilon\|_{L^2(I)}^2$ (see Corollary 2.5).

The (nonlinear) quantization estimators obtained by quantizing wavelet coefficients to a set of discrete levels have similar properties as the threshold estimators. Here the aim is to compress the information about a function contained in the wavelet coefficients by quantization and estimate the squared error loss of this compression which is obtained by replacing the random coefficients $\beta_{J_m}+\xi_{J_m}$ by $\delta_m[(\beta_{J_m}+\xi_{J_m})/\delta_m]$ where $[x]$ denotes the rounding towards the nearest integer and δ_m is chosen as $2^{md/2}\delta$.

For random noise satisfying an exponential Bernstein type moment condition we prove for the Besov spaces of index $\tau = (\alpha/d + 1/2)^{-1}$ the squared error loss can be bounded as $\sup_{\bar{B}_{l,M}} \mathbf{E}\|f - \hat{f}_r\|_{L^2(I)}^2 \leq C\delta^{2\tau\alpha/d}$ both for Haar bases and regular wavelets. Analogous results are proven for the case of random noise with finite polynomial moments.

Also we provide two examples illustrating some limits of image reconstruction. In Corollary 2.11 we prove that the existence of a second absolute moment for the random noise does not guarantee the convergence of threshold estimators whereas the linear estimators attain their optimal speed of convergence. In Example 2.9 we show that the so-called near-best estimator which represents the optimal rate reconstruction for images without random noise is not consistent for images sampled with random noise.

2 Results.

Linear estimators. Independent noise.

In order to apply our results to recover real images which are usually discontinuous functions we suppose that the recovering functions f are represented by Haar bases and belong to the Besov spaces $B_{p,q}^{\alpha,1}$ with $\alpha < 1/p$. Note that all functions represented by Haar bases and belonging to the Besov spaces with $\alpha \geq 1/p$ are constants. In this paragraph we formulate results for linear estimators and independent random variables ξ_{J_l}.

Theorem 2.1 *Assume that $p \geq 1,\ q > 0$, and that the set of the functions $\{\phi_{J_m} \mid \phi_{J_m} \in D\}$ denotes a Haar bases variables $\{\xi_{J_l} \mid \xi_{J_l}$ independent, $\mathbf{E}\xi_{J_l} = 0$, $\mathbf{E}\xi^2_{J_l} < R\}$. Then the risk for linear estimators is bounded as follows*

$$\mathbf{R}_L(M, \Theta) \ \leq \ C2^{-r}$$

where the constant C depends on the parameters $d,\ R,\ M$ only and $r = \dfrac{2d\alpha l}{d + 2\alpha}$, if $p \geq 2,\ 0 < \alpha < 1/p$ and $r = \dfrac{dl(d/2 + \alpha - d/p)}{d + \alpha - d/p}$, if $1 \leq p < 2,\ d(1/p - 1/2) < \alpha < 1/p$.

Remarks. (i) It is not difficult to show that under the conditions of Theorem 2.1 in the case $p \geq 2$ the minimax risks $\mathbf{R}_L(M, \Theta)$ and $\mathbf{R}_{NL}(M, \Theta)$ are proportional and the estimate of the theorem has an optimal convergence rate.

(ii) The results of the Theorem 2.1 in the case $d/p - d/2 < \alpha < 1/p$ can be applied to the function $f = P_l g$ with $g \notin L^2(I)$ where the operator P_l is the linear projector onto N_l.

Linear estimators. Correlated random noise.

The results of Theorem 2.1 can be generalized to the case of correlated noise ξ with finite second moment.

Theorem 2.2 *Assume that the conditions of Theorem 2.1 hold with $\Theta = \{\xi_{J_l} \mid \ \mathbf{E}\xi_{J_l} = 0,\ \mathbf{E}\xi^2_{J_l} < R\}$. Then the estimate*

$$\mathbf{R}_L(M, \Theta) \ \leq \ C2^{-2\alpha m_1}$$

with the constant C depending on $d,\ M\ \Theta$ only and

$$m_1 \ = \ \arg \min_{1 \leq m \leq l} \{2^{-2d(l-m)} \sum_{J_m \in D_m} \mathbf{E}(\sum_{j \in J_m} \xi_{J_l})^2 \ + \ 2^{-r_1}\}$$

where $r_1 = 2m\alpha$ in the case $p \geq 2,\ 0 < \alpha < 1/p$ and $r_1 = -(2md/p - md - 2\alpha m)$ in the case $1 \leq p < 2,\ d(1/p - 1/2) < \alpha < 1/p$ is valid.

If the correlation between the random variables ξ_{J_l} and $\xi_{J'_l}$ is not greater than some power of the distance $\rho(J_l, J'_l)$ between the index cubes J_l and J'_l, that is $\Theta = \{\xi_{J_l} \mid \ \mathbf{E}\xi_{J_l} = 0,\ |\mathbf{E}\xi_{J_l}\xi_{J'_l}| \leq c(1 + \rho(J_l, J'_l)^{-\gamma}\}$ (where c and γ are positive constants) the inequality holds with a constant C depending on $d,\ M,\ \rho$ only and with $m_1 = \dfrac{\hat{\gamma} l}{\hat{\gamma} + 2\alpha}$, where $\hat{\gamma} = \min\{\gamma, d\}$, if $p \geq 2,\ 0 < \alpha < 1/p$ or with $m_1 = \dfrac{\hat{\gamma} l(d + 2\alpha - 2d/p)}{2\alpha(\hat{\gamma} + d + 2\alpha - 2d/p)}$ if $1 \leq p < 2,\ d(1/p - 1/2) < \alpha < 1/p$.

Threshold estimators for non-Gaussian case.

To define the threshold estimator we introduce the linear subspaces $V_m = span\{\phi_{J_m} \mid J_m \in D_m\}$ and the linear orthogonal projectors $P_m : V_l \to V_m$, $1 \le m \le l$ and represent functions f in the form

$$f \ = \ P_l f \ = \ P_0 f + \sum_{m=1}^{l} (P_m - P_{m-1}) f \ = \ \sum_{m=0}^{l} \sum_{J_m \in D_m} \beta_{J_m} \phi_{J_m}$$

Then the threshold estimator can be written as

$$\hat{f}_\epsilon \ = \ \sum_{m=0}^{l} \sum_{J_m \in D_m} \hat{\beta}_{J_m} \phi_{J_m}$$

for a nonnegative parameter ϵ which will be defined below and $\hat{\beta}_{J_m} \ = \ (\beta_{J_m} + \xi_{J_m}) \mathbf{1}_{\{|\beta_{J_m} + \xi_{J_m}| > \ 2^{m/2}\epsilon\}}$.

It has been shown by Donoho and Johnstone (cf. Donoho (1992), Donoho and Johnstone (1992, 1993)) that in contrast to the case of Besov spaces $B^{\alpha,r}_{p,q}$ with parameter $p \ge 2$ threshold estimators outperform linear estimators in the case of the Besov spaces with parameter $p < 2$. Below we describe the properties of the threshold estimators for *non-Gaussian* random noise. It follows from our results that convergence of the minimax risk $\mathbf{R}_{NL}(M, \Theta)$ as well as square risk for threshold estimators depends on the behavior of the "tails" of the random noise. Recall that the random noise ξ satisfies Bernstein type condition if for some positive number t and $0 < g_j < R$, $\mathbf{E} \exp\{t\xi_{J_l}\} \le \exp\{\frac{1}{2} g_j t^2\}$.

Theorem 2.3 *If f is in the Haar bases subspace V_l, α satisfies $0 < \alpha < 1/p$, and the class of random variables is given by $\Theta = \{\xi_{J_l} \mid \xi_{J_l}$ independent, $\mathbf{E}\xi_{J_l} = 0$, satisfying Bernstein type condition$\}$, and $\hat{f}_\epsilon$ is the threshold estimator with thresholding level $\epsilon = A 2^{-dl/2} l^{-1/2}$ for a sufficiently large constant A, and $r = \dfrac{2\alpha}{2\alpha + d}$ then the following estimate holds*

$$\mathbf{R}_{NL}(M, \Theta) \ \le \ \sup_{\xi_{J_l} \in \Theta} \sup_{\bar{B}_{l,M}} \mathbf{E} ||f - \hat{f}_\epsilon||^2_{L^2(I)} \ \le \ C 2^{-rdl} l^r .$$

If $\{\phi_{J_m} \mid \phi_{J_m} \in W^s_\infty, \ J_m \in D\}$ is a compactly supported orthogonal wavelet with finite mask which satisfies the Strang-Fix condition of order r, then for $0 < \alpha < \min\{r, s\}$ the minimax risk can be bounded as

$$\mathbf{R}_{NL}(M, \Theta) \ \le \ \sup_{\xi_{J_l} \in \Theta} \sup_{\bar{B}_{l,M}} \mathbf{E} ||f - \hat{f}_\epsilon||^2_{L^2(I)} \ \le \ C 2^{-rdl} l^r .$$

Remarks. (i) Note that in the Haar bases case the minimax risk is defined for the Besov spaces $B^{\alpha,1}_{\tau,\tau}$ whereas in the regular wavelets case the Besov spaces $B^{\alpha,r}_{\tau,\tau}$ are used.

(ii) Note that noise models with *Gaussian* and bounded random variables ($|\xi_{J_l}| \leq R$) belong to the class Θ.

(iii) The result of Theorem 2.3 holds for random processes satisfying others Bernstein type conditions as well. For instance, assume that the random variables ξ_{J_l} belong to the class $\Theta = \{\xi_{J_l} \mid \ \xi_{J_l}$ *independent* $\mathbf{E}\xi_{J_l} = 0,\ |\mathbf{E}\xi_j^m| \leq 2^{-1}m!\sigma_j^2 R^{m-2}$ *for all random variables* ξ_j *and integers* $m = 2, 3, \ldots\}$

(iv) The estimate of square losses for the threshold estimator in Theorem 2.3 has the optimal order.

Corollary 2.4 *Under the conditions of Theorem 2.3 we have*

$$\inf_{\epsilon \geq 0} \sup_{\bar{B}_{l,M}} \mathbf{E}||f - \hat{f}_\epsilon||^2_{L^2(I)} \propto C(f) 2^{-rdl} l^r.$$

Below we consider the problem of recovering functions f from additive random noise ξ with finite polynomial moments $\mathbf{E} \mid \xi_{J_l} \mid^p,\ p \geq 2$. Recall that $\tau = (\alpha/d + 1/2)^{-1}$.

Theorem 2.5 *Assume that the conditions of Theorem 2.3 hold and the class of random variables is give by* $\Theta = \{\xi_{J_l} \mid \ \xi_{J_l}$ *independent,* $\mathbf{E}\xi_{J_l} = 0,\ \mathbf{E} \mid \xi_{J_l} \mid^p < R,\ p \geq 2\}$. *Consider thresholding levels* $\epsilon = 2^{-\delta},\ \delta = \dfrac{(p-2)dl}{2(2\tau\alpha/d + p - 2)}$ *for the threshold estimator* $\hat{f}_\epsilon$. *Define* $\gamma = \dfrac{\tau\alpha(p-2)l}{2\tau\alpha/d + p - 2}$. *Then there exists a constant* C *depending* d, α, M, R *only such that*

$$\mathbf{R}_{NL}(M, \Theta) \ \leq \ \sup_{\xi_{J_l} \in \Theta} \sup_{B_{l,M}} \mathbf{E}\|f - \hat{f}_\epsilon\|^2_{L^2(I)} \leq C2^{-\gamma},$$

in the Haar basis case, and

$$\mathbf{R}_{NL}(M, \Theta) \ \leq \ \sup_{\xi_{J_l} \in \Theta} \sup_{\bar{B}_{l,M}} \mathbf{E}\|f - \hat{f}_\epsilon\|^2_{L^2(I)} \leq C2^{-\gamma}.$$

in the regular wavelet case are valid.

Remarks. (i) In general the exponent γ can not be improved.

(ii) If the conditions of Theorem 2.6 are fulfilled and ξ_{J_m} are identically distributed random variables then

$$\sup_{\bar{B}_{l,M}} \mathbf{E}\|f - \hat{f}_\epsilon\|^2_{L^2(I)} = o(2^{-\gamma}),$$

where $\gamma = \dfrac{\tau\alpha(p-2)l}{2\tau\alpha/d + p - 2}$.

(iii) Below we prove that in this case the thresholding level essentially depends on the parameters of the Besov space.

Quantization algorithms.

In order to transmit images along a communications channel or to store them in computer memory we may compress them using quantized coefficients (the problem of quantization of decomposed images was studied for instance by DeVore, Jawerth and Lucier (1992)). Below we discuss how to choose appropriate quantization levels for functions sampled with random noise. Again the square loss criterion is used to measure the quantization errors.

Theorem 2.6 *If the conditions of Theorem 2.3 are valid then there exist a positive constant $A = A(R)$ such that for quantization estimators with quantization levels $\delta_m = 2^{md/2}\delta$ where $\delta > A2^{-ld/2}l^{1/2}$ the estimate*

$$\sup_{\xi_{J_l}\in\Theta}\sup_{B_{l,M}} \mathbf{E}\|f-\hat{f}_r\|^2_{L^2(I)} \leq C\delta^{2\tau\alpha/d}$$

holds in the Haar bases case and the estimate

$$\sup_{\xi_{J_l}\in\Theta}\sup_{\bar{B}_{l,M}} \mathbf{E}\|f-\hat{f}_r\|^2_{L^2(I)} \leq C\delta^{2\tau\alpha/d}$$

holds in the regular wavelet case with a constant C depending on d,α,M,R only.

Now we formulate two results for discontinuous and regular images under weaker moment conditions.

Theorem 2.7 *Suppose that the conditions of Theorem 2.5 are valid.*

For quantization levels $\delta_m = 2^{md/2}\delta$ with $\delta \geq A2^{-\gamma},\ \gamma = \dfrac{(p-2)dl}{2(2\tau\alpha/d+p-2)}$ and a constant C depending on $d,\ \alpha,\ M,\ R$ only the estimate

$$\sup_{\xi_{J_l}\in\Theta}\sup_{B_{l,M}} \mathbf{E}\|f-\hat{f}_r\|^2_{L^2(I)} \leq C\delta^{2\tau\alpha/d}$$

holds in the Haar basis case and the estimate

$$\sup_{\xi_{J_l}\in\Theta}\sup_{\bar{B}_{l,M}} \mathbf{E}\|f-\hat{f}_r\|^2_{L^2(I)} \leq C\delta^{2\tau\alpha/d}$$

holds in the regular wavelet case.

Some features of image compression.

It has been proved by Donoho and Johnstone(1992) for functions sampled with *Gaussian noise* that the threshold estimator possesses the minimax property in the case of L^2–risk under weaker smoothness conditions compared to linear estimators. They also proved that in the case of $L^p,\ 1<p<2$–risks the threshold estimator is minimax up to logarithmic factor whereas the linear estimator is not minimax.

The situation changes if we consider *non-Gaussian* random noise with finite polynomial moments. In this case the linear estimator can be more effective compared to the threshold one. There are examples of *non-Gaussian* noise when the L^2–risk of the threshold estimator does not tend to zero, but the linear estimator is minimax.

Corollary 2.8 *Assume that $p \geq 1,\ q > 0$, that the set of the functions $\{\phi_{J_m} \mid \phi_{J_m} \in D\}$ is a Haar bases in V_l and Θ denotes the class of random variables $\Theta = \{\xi_{J_l} \mid \xi_{J_l}$ independent, $\mathbf{E}\xi_{J_l} = 0,\ \mathbf{E}\xi_{J_l}^2 < R\}$. Then for any thresholding level $\epsilon > 0$ and the threshold estimator $\hat{f}_\epsilon$ we have*

$$\sup_{\xi_{J_l} \in \Theta} \sup_{B_{l,M}} \mathbf{E}\|f - \hat{f}_\epsilon\|^2_{L^2[0,1]^d} \geq C, \quad l \to \infty$$

Finally, to illustrate some features of the compression of images sampled with random noise we give one simple example.

Example. Suppose that $f_1 \equiv 1$ and we are interested in compression in $L^1(I)$. Then f_1 can be represented in the form $f_1 = \sum_{m=0}^{l} \sum_{J_m \in D_m} C_{J_m}(f_1) \cdot \phi_{J_m}$ with coefficients which are calculated from best approximations in the spaces V_m and V_{m-1}. In this case the best approximations are medians (see DeVore, Jawerth and Lucier (1992)), i.e. numbers a_{J_m} such that

$$\{x \in J_m \mid f(x) \geq a_{J_m}\} \ \geq \ 1/2$$

and

$$\{x \in J_m \mid f(x) \leq a_{J_m}\} \ \geq \ 1/2$$

Hence, $C_{J_m}(f_1) = 0,\ 1 \leq m \leq l$ and $C_{J_0}(f_1) = a_{J_0} = 1$. The threshold estimator $\hat{f}_1 = \sum_{m=0}^{l} \sum_{J_m \in D_m} C_{J_m}(f_1) \cdot \mathbf{1}_{\{|C_{J_m}| > 2^{d \cdot m/2} \cdot \epsilon\}} \cdot \phi_{J_m}$ is equal to 1 while $0 < \epsilon < 1$ and we have $||f_1 - \hat{f}_1||_{L^1(I)} = 0$.

Now assume that we observe a noised function $\tilde{f}(x) = f(x) + \sum_{J_l \in D_l} \xi_{J_l} \cdot \phi_{J_l}$ with independent Bernoulli random variables $\xi_{J_l}\ :\ \mathbf{P}(\xi_{J_l} = \pm 1) = 1/2$. One can easily prove that for any $0 < \epsilon < 1$ and the threshold estimator

$$\hat{f} \ = \ \sum_{m=0}^{l} \sum_{J_m \in D_m} (C_{J_m} + \xi_{J_m}) \cdot \mathbf{1}_{\{|C_{J_m} + \xi_{J_m}| > 2^{d \cdot m/2} \cdot \epsilon\}} \cdot \phi_{J_m}$$

we have

$$\mathbf{E}||f - \hat{f}||_{L^1(I)} \geq C$$

for an absolute constant C, since in this case $(C_{J_m} + \xi_{J_m}) \cdot \mathbf{1}_{\{|C_{J_m} + \xi_{J_m}| > 2^{d \cdot m/2} \cdot \epsilon\}} = 2$ or 0 with probability

$$1 \ - \ \mathbf{P}(\sum_{J_l \in D_l} \xi_{J_l} = 0) \geq C$$

(It is well known that $\mathbf{P}(\sum_{J_l \in D_l} \xi_{J_l} = 0) \approx C \cdot 2^{-d \cdot l/2})$). That is we get a good approximation of the image without noise by the thresholding method but a non shrinking distance between the noised image and its thresholding estimate.

3 Proofs.

Proof of Theorem 2.1. The case $p \geq 2$ can be proved easily. Obviously,

$$\mathbf{E}||f - \hat{f}||^2_{L^2(I)} = ||f - P_m f||^2_{L^2(I)} + \mathbf{E}||P_m \sum_{J_l} \xi_{J_l} \phi_{J_l}||^2_{L^2(I)} .$$

The last summand satisfies the inequality

$$\mathbf{E}||P_m \sum_{J_l} \xi_{J_l} \phi_{J_l}||^2_{L^2(I)} \leq 2^{-(l-m)d} R .$$

The first one can be estimated by using a known result of DeVore and Popov (1988) (see Theorem 4.5, p.404) and an equivalent definition of the Besov space norm

$$||f - P_m f||^2_{L^2(I)} \leq C(d) w_1(f, 2^{-m})^2_2 \leq 2^{-2\alpha m} C(d) ||f||^2_{B^{\alpha,1}_{p,q}}, \quad p \geq 2.$$

Then for $p \geq 2$ and any $1 \leq m \leq l$ for some constant $C(d)$

$$\mathbf{E}||f - \hat{f}||^2_{L^2(I)} \leq C(d)(2^{-(l-m)d} + 2^{-2\alpha m} ||f||^2_{B^{\alpha,1}_{p,q}})$$

and the estimate we need may be constructed by minimizing the last inequality in m.

Now let us consider the second case when $1 \leq p < 2$, $d/p - d/2 < \alpha < 1/p$. Here we have to use a slightly more complicated estimate for $||f - P_m f||_{L^2(I)}$. It is easy to see, that

$$||f - P_m f||_{L^2(I)} \leq \sum_{k=m}^{l-1} ||P_{k+1} f - P_k f||_{L^2(I)} \tag{1}$$

with

$$P_k f = \sum_{J_k \in D_k} f_{J_k} \phi_{J_k}, \quad f_{J_k} = 2^{-(l-k)d} \sum_{2^{-l} j \in J_k} f(2^{-l} j)$$

Assume $J_{k,\nu}$, $0 \leq \nu \leq d$ are the parallelepipeds the first sides of which have length $2^{-(k+1)}$ and the other ones have length 2^{-k} (if $\nu = 0$ then $J_{k,\nu} = J_k$ and if $\nu = d$ then $J_{k,\nu} = J_{k+1}$).

Denote by $D_{k,\nu}$ the union of all $J_{k,\nu}$ and $\phi_{J_{k,\nu}} = \mathbf{1}_{\{J_{k,\nu}\}}$. Also define the linear spaces $V_{k,\nu} = span\{\phi_{J_{k,\nu}} \mid J_{k,\nu} \in D_{k,\nu}\}$ and the projectors $P_{k,\nu} : V_l \to V_{k,\nu}$. Then

$$||P_{k+1} f - P_k f||_{L^2(I)} \leq \sum_{\nu=0}^{d-1} ||P_{k,\nu+1} f - P_{k,\nu} f||_{L^2(I)} \tag{2}$$

with

$$P_{k,\nu} f = \sum_{J_{k,\nu} \in D_{k,\nu}} f_{J_{k,\nu}} \phi_{J_{k,\nu}}, \quad f_{J_{k,\nu}} = v_{k,\nu} \sum_{2^{-lj} \in J_{k,\nu}} f(2^{-l} j),$$

$$v_{k,\nu} = 2^{(l-k-1)\nu - (l-k)(d-\nu)}.$$

Thus, we have

$$||P_{k,\nu+1}f \quad - \quad P_{k,\nu}f||_{L^2(I)} =$$

$$= \quad || \sum_{J_{k,\nu+1}\in D_{k,\nu+1}} f_{J_{k,\nu+1}}\phi_{J_{k,\nu+1}} \quad - \sum_{J_{k,\nu}\in D_{k,\nu}} f_{J_{k,\nu}}\phi_{J_{k,\nu}}||_{L^2(I)}$$

$$= \quad s^{1/2}_{k,\nu+1}(\sum_{J_{k,\nu}\in D_{k,\nu}} \quad \sum_{J_{k,\nu+1}\subset J_{k,\nu}} (f_{J_{k,\nu}} - f_{J_{k,\nu+1}})^2)^{1/2}$$

with $s_{k,\nu} = 2^{-(k+1)\nu-k(d-\nu)}$ (the sum $\sum_{J_{k,\nu+1}\subset J_{k,\nu}}$ consists of two summand only).

The last expansion does not exceed

$$2^{1/2}s^{1/2}_{k,\nu+1}(\sum_{J_{k,\nu}\in D_{k,\nu}} (f_{J_{k,\nu}} - f_{J_{k,\nu+1}})^2)^{1/2}$$

for any fixed $J_{k,\nu+1} \subset J_{k,\nu}$. Below we consider $\tilde{J}_{k,\nu+1}$ as the left (or lower) parallelepiped, from the two $J_{k,\nu+1}$ belonging to $J_{k,\nu}$ and the vector

$h_{k,\nu} \;=\; (\overbrace{0,\ldots,0}^{\nu-1}, 2^{-k-1}, 0,\ldots,0)$. Then for $x_j = 2^{-l}j$

$$|\, f_{J_{k,\nu}} - f_{\tilde{J}_{k,\nu+1}}| \;=\; v_{k,\nu+1}| \sum_{x_j\in\tilde{J}_{k,\nu+1}} (f(x_j+h) - f(x_j)) \quad |$$

and

$$||P_{k,\nu+1}f \quad - \quad P_{k,\nu}f||_{L^2(I)} \;=$$

$$= \quad 2^{1/2}s^{1/2}_{k,\nu+1}v_{k,\nu+1}(\sum_{J_{k,\nu}\in D_{k,\nu}} (\sum_{x_j\in\tilde{J}_{k,\nu+1}} (f(x_j+h) - f(x_j))^2)^{1/2}$$

$$\leq \quad 2^{1/2}s^{1/2}_{k,\nu+1}v_{k,\nu+1}$$

$$(\sum_{J_{k,\nu}\in D_{k,\nu}} v^{-2(p-1)/p}_{k,\nu+1}(\sum_{x_j\in\tilde{J}_{k,\nu+1}} |f(x_j+h) - f(x_j)|^p)^{2/p})^{1/2}$$

$$\leq \quad 2^{1/2+dl/p}s^{1/2}_{k,\nu+1}v^{1/p}_{k,\nu+1}w_1(f, 2^{-k-1})_p$$

Using the definition of the Besov space norm we can get the estimate

$$||P_{k,\nu+1}f - P_{k,\nu}f||_{L^2(I)} \;\leq\; 2^{kd/p-kd/2-\alpha k}C||f||_{B^{\alpha,r}_{p,q}}$$

with an absolute constant C.

It follows from estimates (1), (2) and the last inequality that

$$||f - P_m f||^2_{L^2(I)} \;=\; C(d)2^{2md/p-md-2\alpha m}||f||^2_{B^{\alpha,r}_{p,q}}$$

is valid as soon as $d/p - d/2 \leq \alpha$.

Hence

$$\mathbf{E}||f - \hat{f}_m||^2_{L^2(I)} \;\leq C(d)(2^{2md/p-md-2\alpha m}M^2 + 2^{-(l-m)d}R).$$

The inequality we need can be obtained by minimizing the right hand side of the last estimate in n.

□

To prove *Theorem 2.2* one can repeat the proof of Theorem 2.1 and choose the level m_1 such that approximation error and random error has the same convergence rate.

To construct minimax estimates of the square losses for threshold estimators we need a slightly modified result by DeVore, Jawerth and Lucier (1992) (Theorem 1, p.724) concerning the problem of accuracy of approximation of functions belonging to Besov spaces. We call a constant $\bar{a}$ is a "near best" approximation of a function $f \in L^p(I)$, $1 \leq p < \infty$ in a cube $J_m \in D_m$ if for some absolute constant c the inequality

$$\int_{J_m} |f - \bar{a}|^p dx \;\leq\; c \inf_a \int_{J_m} |f - a|^p dx$$

holds (DeVore, Jawerth and Lucier (1992)).

Consider a function $f_1 \;=\; \sum_{m=0}^{\infty} \sum_{J_m \in D_m} C_{J_m}(f_1)\phi_{J_m}$, its coefficients $C_{J_m}(f_1)$ can be written as differences of "near best" approximations of this function in spaces V_m and V_{m-1} (the coefficient $C_{J_0}(f_1)$ is a "near best" approximation in V_0). Let f_2 denote the function $f_2 \;=\; \sum_{m=0}^{\infty} \sum_{J_m \in D_m} \bar{C}_{J_m}\phi_{J_m}$ with coefficients $\bar{C}_{J_m}$ such that

$$||(C_{J_m}(f_1) \;-\; \bar{C}_{J_m})\phi_{J_m}||_{L^p(I)} \;\leq\; CN^{-1/\tau}$$

for an absolute constant C, positive numbers $N, p \geq 1, \alpha$ and $\tau = (\alpha/d + 1/p)^{-1}$, and such that $\bar{C}_{J_m} = 0$, if $||(C_{J_m}(f_1)\phi_{J_m}||_{L^p(I)} \leq N^{-1/\tau}$. Then the following theorem which is the modified variant of the result by DeVore and Popov (DeVore, Jawerth and Lucier (1992), Theorem 1, p.724) is valid.

Theorem 3.1 *If a function f_1 belonging to the Besov space $B^{\alpha,1}_{\tau,\tau}$, $0 < \alpha < 1/p$ is represented by Haar bases , or if belongs to the Besov space $B^{\alpha,r}_{\tau,\tau}$, $0 < \alpha < r+1$ and is represented by regular wavelets then there exists a constant C such that the inequalities*

$$||f_1 - f_2||_{L^p(I)} \leq CN^{-\alpha/d}||f||^{\tau/p}_{B^{\alpha,1}_{\tau,\tau}}$$

in the Haar bases case and

$$||f_1 - f_2||_{L^p(I)} \leq CN^{-\alpha/d}||f||^{\tau/p}_{B^{\alpha,r}_{\tau,\tau}}$$

in the regular case hold for any f_2 defined as above.

Proof of Theorem 2.3. We prove the theorem only in the case of Haar bases, since it is similar in the other case.

Let us fix the number

$$N = A(R)2^{\tau dl/2}l^{-\tau/2}$$

for sufficiently large $A(R)$ that will be chosen below and define $\epsilon = N^{-1/\tau}$.

The of the square loss does not exceed

$$\mathbf{E}||f - \hat{f}||^2_{L^2(I)} \ \leq \ 2I_1 + 2I_2$$

where

$$I_1 \ = \ \mathbf{E}||\sum_{m=0}^{l}\sum_{J_m\in D_m}(\beta_{J_m} - \hat{\beta}_{J_m})\mathbf{1}_{\{|\xi_{J_m}|\leq 2^{dm/2-1}\epsilon\}}\phi_{J_m}||^2_{L^2(I)}$$

$$I_2 \ = \ \mathbf{E}||\sum_{m=0}^{l}\sum_{J_m\in D_m}(\beta_{J_m} - \hat{\beta}_{J_m})\ \mathbf{1}_{\{|\xi_{J_m}|>2^{dm/2-1}\epsilon\}}\phi_{J_m}||^2_{L^2(I)}$$

with

$$\begin{aligned}
\xi_{J_m} \quad &= \quad 2^{-d(l-m)}\sum_{j\in J_m}\xi_{J_l} \quad - \quad 2^{-d(l-m+1)}\sum_{j\in \tilde{J}_{m-1}(J_m)}\xi_{J_l}, \quad 1\leq m\leq l \\
\xi_{J_0} \quad &= \quad 2^{-dl}\sum_{j\in I}\xi_{J_l}
\end{aligned}$$

One can see that

$$\begin{aligned}
(\beta_{J_m} - \hat{\beta}_{J_m}) \qquad & \mathbf{1}_{\{|\xi_{J_m}|\leq 2^{dm/2-1}\epsilon\}} = \\
&= \begin{cases} 0 & if \quad |\xi_{J_m}| > 2^{d\cdot m/2-1}\epsilon \\ \xi_{J_m} & if \quad \hat{\beta}_{J_m} > 2^{d\ m/2-1}\epsilon,\ |\xi_{J_m}|\leq 2^{dm/2-1}\epsilon \\ \beta_{J_m} & if \quad \hat{\beta}_{J_m}\leq 2^{dm/2-1}\epsilon,\ |\xi_{J_m}|\leq 2^{dm/2-1}\epsilon \end{cases}
\end{aligned} \tag{3}$$

Hence,

$$||(\beta_{J_m} - \hat{\beta}_{J_m})\mathbf{1}_{\{|\xi_{J_m}|\leq 2^{dm/2-1}\epsilon\}}\phi_{J_m}||_{L^2(I)} \leq 2\epsilon$$

and

$$\hat{\beta}_{J_m}\mathbf{1}_{\{|\xi_{J_m}|\leq 2^{dm/2-1}\epsilon\}} \ = \ 0, \quad \text{when } ||\beta_{J_m}\mathbf{1}_{\{|\xi_{J_m}|\leq 2^{dm/2-1}\epsilon\}}\phi_{J_m}||_{L^2(I)} \leq 2^{-1}\epsilon$$

Thus, we can apply Theorem 3.1 for functions

$$f_1 \ = \ \sum_{m=0}^{l}\sum_{J_m\in D_m}\beta_{J_m}\mathbf{1}_{\{|\xi_{J_m}|\leq 2^{dm/2-1}\epsilon\}}\phi_{J_m}$$

and

$$f_2 \ = \ \sum_{m=0}^{l}\sum_{J_m\in D_m}\tilde{\beta}_{J_m}\mathbf{1}_{\{|\xi_{J_m}|\leq 2^{dm/2-1}\epsilon\}}\phi_{J_m}$$

Since in our case

$$||f||^{\alpha}_{B^{\alpha,1}_{\tau,\tau}} \approx 2^{-\alpha\tau} \sum_{m=0}^{l} \sum_{J_m \in D_m} |\beta_{J_m}||J_m|^{\alpha/p},$$

(see DeVore and Popov (1987), p.244) the inequality $||f_1||^{\alpha}_{B^{\alpha,1}_{\tau,\tau}} \leq ||f||^{\alpha}_{B^{\alpha,1}_{\tau,\tau}}$ holds the estimate

$$I_1 \ < \ N^{-2\alpha/d} \cdot M^{\tau} \tag{4}$$

is valid. To estimate I_2 let us remark that the random variables $\xi_{J_m},\ 1 \leq m \leq l$ have zero mean $\mathbf{E}\xi_{J_m} = 0$ and variation $\mathbf{E}\xi^2_{J_m} = \sigma^2_{J_m} \leq R2^{-(l-m)d}(1-2^d)$ and ξ_{J_0} have zero mean and variation $\sigma^2_{J_0} \leq R2^{-ld}$. Then (see for instance, Shorack and Wellner (1986), p 850) for $\epsilon = N^{-1/\tau}$

$$\mathbf{P}(|\xi_{J_m}| \geq 2^{dm/2-1}\epsilon) \geq C\exp\{-2^{-3}R^{-1}(1-2^{-d})^{-1}A(R)^{2/\tau}l\}.$$

Now choose $A(R)$ so large, that the right hand side of the last equation will be less than $2^{-3\cdot l}$. Then

$$I_2 \ < \ 2^{-3l}(1+M^2) \tag{5}$$

Finally, to construct a minimax estimate of the square losses it will be enough to apply the inequality

$$\mathbf{E}||f-\hat{f}||^2_{L^2(I)} \ \leq \ C(M,R)2^{-2\alpha dl/(2\alpha+d)}l^{-2\alpha/(2\alpha+d)}$$

which is a consequence of relations (4) and (5).
□

Proof of Corollary 2.4. Suppose for simplicity that ξ_{J_l} is $(0,1)$-Gaussian and $||f||_{B^{\alpha,1}_{\tau,\tau}} = 1$. Recall that $\epsilon = \epsilon(l) = A^{-1/\tau} \cdot 2^{-d\cdot l/2} \cdot l^{1/2}$. Denote by D^+_m the set of parallelepipeds J_m such that $|\beta_{J_m}| \leq 2^{m\cdot d/2-1}\cdot\epsilon$. Using the representation (3) one can show the inequality

$$\begin{aligned} &\mathbf{E} \ \ ||f-\hat{f}(\epsilon)||^2_{L^2(I)} \geq \\ &\geq \ (||\sum_{m=0}^{l} \sum_{J_m \in D^+_m} \beta_{J_m}\cdot\phi_{J_m}||_{L^2(I)} \ - \ C(f)\cdot P(|\xi_{J_m}| \ > \ 2^{d\cdot l/2}\cdot\epsilon))^2. \end{aligned}$$

It follows from the results of the paper of DeVore and Popov (1987, p. 243) that there exists the function f with Besov space norm is equal to one and the property

$$||\sum_{m=0}^{l} \sum_{J_m \in D^+_m} \beta_{J_m}\cdot\phi_{J_m}||_{L^2(I)} \ \geq \ C(f)\cdot\epsilon^{\tau\cdot\alpha/d}.$$

Furthermore we use the well known Mill's ratio for Gaussian random variables and arbitrary number $s \geq 0$ (see Shorack and Wellner (1986), p 850))

$$(2\pi)^{-1/2}(s^{-1}-s^{-3})\exp\{-s^2/2\} \ < \ \mathbf{P}(|\xi_{J_l}| > s) \ < \ (2\pi)^{-1/2}s^{-1}\exp\{-s^2/2\}$$

Now the proof of our proposition follows by the last two inequalities for sufficiently a large constant A and Theorem 2.4.
□

Proof of Theorem 2.5. We shall give the proof in the case of Haar bases only since it is similar in the regular cases . We shall need the representation of the Besov space norm

$$\|f\|^{\tau}_{B^{\alpha,1}_{\tau,\tau}} \approx 2^{-\alpha\tau} \sum_{m=0}^{l} \sum_{J_m \in \mathcal{D}_{\Updownarrow}} 2^{-md\tau/2} \mid \beta_{J_m} \mid^{\tau}$$

(see DeVore and Popov (1987), p.244) and the definition of the coefficients

$$a_{J_m} = (\beta_{J_m} - \hat{\beta}_{J_m})\mathbf{1}_{\{|\xi_{J_m}|\leq\ 2^{md/2-1}\epsilon\}}$$

and

$$b_{J_m} = (\beta_{J_m} - \hat{\beta}_{J_m})\mathbf{1}_{\{|\xi_{J_m}|>\ 2^{md/2-1}\epsilon\}}$$

Obviously,

$$\mathbf{E}\|f - \hat{f}_\epsilon\|^2_{L^2(I)} \leq 2I_1 + 2I_2$$

with

$$I_1 = \mathbf{E}\| \sum_{m=0}^{l} \sum_{J_m \in D_m} a_{J_m}\phi_{J_m}\|^2_{L^2(I)} \ \ and \ \ I_2 = \mathbf{E}\| \sum_{m=0}^{l} \sum_{J_m \in D_m} b_{J_m}\phi_{J_m}\|^2_{L^2(I)}$$

Since

$$a_{J_m} = \begin{cases} \xi_{J_m} & \text{if} \mid \beta_{J_m} + \xi_{J_m} \mid > 2^{md/2}\epsilon, \mid \xi_{J_m} \mid < 2^{md/2-1}\epsilon \\ \beta_{J_m} & \text{if} \mid \beta_{J_m} + \xi_{J_m} \mid \leq 2^{md/2}\epsilon, \mid \xi_{J_m} \mid < 2^{md/2-1}\epsilon \end{cases}$$

the coefficients of the first term I_1 are less then

$$\mid \alpha_{J_m} \mid \leq 4\min\{\beta_{J_m}, 2^{md/2-1}\epsilon\}$$

Hence, by Theorem 3.1(DeVore and Popov)

$$I_1 \leq 8{\textstyle\sum}_{m=0}^{l}{\textstyle\sum}_{J_m\in D_m} \min\{2^{-md/2}\beta^2_{J_m}, 2^{-2}\epsilon^2\} \leq C(d,\alpha,M)\epsilon^{2\tau\alpha/d}$$

To estimate I_2 one can use the inequality

$$\|f\|_{L^2(I)} \leq C(d)\|f\|_{B^{\alpha,1}_{\tau,\tau}}$$

(see DeVore and Popov(1988)), and the estimate

$$\mathbf{P}(\mid \xi_{J_m} \mid > s) \leq C(p)(2^{(l-m)d/2}s)^{-p}(R+1)$$

that follows by Chebyshev's inequality together with Rosenthal's inequality.

Then

$$\begin{aligned} I_2 \ &\leq\ 4\mathbf{E}{\textstyle\sum}_{m=0}^{l}{\textstyle\sum}_{J_m\in D_m} 2^{-md}(\beta^2_{J_m} + \xi^2_{J_m})\mathbf{1}_{\{|\xi_{J_m}|>\ 2^{md/2}\epsilon\}} \\ &\leq\ C(2^{-pld/2}\epsilon^{-p}M^2(R+1) + 2^{-(p-2)ld/2}\epsilon^{-(p-2)}(R+1)) \end{aligned}$$

and

$$\sup_{\|f\|_{B^{\alpha,1}_{\tau,\tau}}\le M} \mathbf{E}\|f-\hat f_\epsilon\|_{L^2(I)} \le C(d,\alpha,M,R)(\epsilon^{2\tau\alpha/d}+2^{-(p-2)ld/2}\epsilon^{-(p-2)})$$

By minimization of the right hand side of the last inequality in ϵ the proof is finished. $\square$

The *proof of Theorem 2.6* coincides with the proof Theorem 2.3, and *the proof of Theorem 2.7* coincides with the proof Theorem 2.5.

To *prove Corollary 2.8* it is enough to consider the constant function $f\equiv 1$ which is sampled with random noise taking large values with small probabilities, so that $\mathbf{E}\xi^2_{J_l}=1$ and $\mathbf{P}(\xi_{J_l}\ge 2^{l/2}\epsilon)\ne 0$.

Acknowledgments. The authors would like to thank I.M. Johnstone for fruitful discussions and useful remarks.

References

[1] Ch.K. Chui and J.Z. Wang. *A general framework of compactly supported splines and wavelets.* J. Approx. Theory, **71**(1992), 263-304.

[2] A. Cohen, I. Daubechies, B. Jawerth, and P. Vial. *Multiresolution analysis, wavelets and fast algorithms on interval.* C.R. Acad. Sci. Paris, ser. I, **316**(1993), 417-421.

[3] R. Coifman, Y. Meyer and M.V. Wickerhauser. *Wavelet analysis and signal processing.* Wavelets and their Applications. Jones and Batlett Publishers, Boston, 1992.

[4] I. Daubecies. *Orthogonal bases of compactly supported wavelets.* Pure Appl. Math., **41**(1988), 909-996.

[5] R.A. DeVore, B. Jawerth, and B.J. Lucier. *Image compression through wavelet coding.* IEEE Trans. Inf. Theory, **38**,2(1992), 719-746.

[6] R.A. DeVore and B.J. Lucier. *Wavelets.* In Acta Numerica, **1**(1991), 1-56.

[7] R.A. DeVore and V. Popov. *Free multivariate splines.* Constr. Approx., **3**(1987), 239-248.

[8] R.A. DeVore and V. Popov. *Interpolation of Besov spaces.* Trans. American Math. Sos., **305**(1988), 397-414.

[9] D.L. Donoho. *De-noising by soft-thresholding.* Tech. Report Dep. Statistics, Stanford univ., 1992.

[10] D.L. Donoho and I.M. Johnstone. *Minimax estimation by wavelet shrinkage.* Tech. Report Dep. Statistics, Stanford univ., 1992.

[11] D.L. Donoho and I.M. Johnstone. *Minimax risk over l_p-balls for l_q-error.* Tech. Report Dep. Statistics, Stanford univ., 1993.

[12] G.R. Shorack and J.A.Wellner. *Empirical processes with applications to statistics.* John Wiley & Sons, 1986.