Wavelet method and asymptotically minimax estimation of regression

Golubev Georgii

Weierstraß–Institut für Angewandte Analysis und Stochastik
Mohrenstaße 39
D–10117, Berlin, Deutschland
e-mail: golubev@wias-berlin.de

[1] *1991 Mathematics Subject Classifications.* Primary 62G05; Secondary 62G20.

Key words and phrases. Minimax risk, entire analytic functions, wavelet method, hard thresholding.

This research was partially supported by RFFI 96-01-00684 and INTAS 93-1585, 94-4194.

Abstract

We attempt to recover a regression function from noisy data. It is assumed that the underlying function is a piecewise entire analytic function. Types and the number of singularities are assumed to be unknown. We show how to chose smoothing parameters and a wavelet basis to achieve the asymptotically minimax risk up to the constant.

1 Introduction

Efficient computational implementation has made wavelets very popular in non-parametric estimation. It is well-known that in most cases wavelet based estimators are almost rate optimal. See e.g. Donoho & Johnstone (1995), (1998), Donoho & Johnstone & Kerkyacharian & Picard (1996). In this paper we give an example of a regression problem showing that a sharp minimax asymptotic can be also achieved by the standard wavelet method based on the thresholding idea proposed by D. Donoho, Y. Johnstone, G. Kerkyacharian, and D. Picard in early nineties. Suppose that we are given noisy data

$$\mathrm{d}X_n(t) = f(t)\,\mathrm{d}t + \frac{\sigma}{\sqrt{n}}\,\mathrm{d}w(t), \quad t \in [0,1], \tag{1}$$

where $w(t),\ t \geq 0$ is the standard Wiener process. Our goal is to recover the unknown function $f(t),\ t \in [0,1]$ based on the observations $X_n(t),\ t \in [0,1]$. We will assume from now on that $n \to \infty$.

In order to develop a nontrivial theory of regression estimation, one usually specifies some functional class $\mathcal{F}$, to which $f(\cdot)$ is assumed to belong. Hölder's, Sobolev's, and Besov's functional classes are often used in non-parametric estimation. Along with these functional classes, entire analytic functions of a finite order constitute an interesting functional class which is also very popular in engineering applications (see Gallager (1968)). This functional class is deeply studied and commonly used in data transmission theory. In statistical usage it was introduced by Ibragimov & Khasminskii (1982) in the context of density estimation. Formally, this class is defined as the set of all real-valued functions f such that

$$f(t) = \int_{-\pi W}^{\pi W} \mathrm{e}^{\mathrm{i}\,t\xi} g(\xi)\,\mathrm{d}\xi, \quad \text{where } \int_{-\pi W}^{\pi W} |g(\xi)|^2\,\mathrm{d}\xi \leq P.$$

Denote this functional class by $\mathcal{F}^0(W,P)$. A remarkable property of $\mathcal{F}^0(W,P)$ showing why it is popular in applications is due to the so-called Sampling Theorem. This theorem states that any function $f(\cdot)$ from $\mathcal{F}^0(W,P)$ admits the representation

$$f(t) = \sum_{k=-\infty}^{\infty} f\left(\frac{k}{W}\right) \frac{\sin[\pi(tW+k)]}{\pi(tW+k)}. \tag{2}$$

The above formula provides us with a heuristic idea of approximation of the function $f(t)$ from $\mathcal{F}^0(W,P)$ on the unit interval $[0,1]$

$$f(t) \sim \sum_{k=0}^{W} f\left(\frac{k}{W}\right) \frac{\sin[\pi(tW+k)]}{\pi(tW+k)}, \quad t \in (0,1).$$

So, we can say that $f(t)$ has on the interval $[0,1]$ approximately W degrees of freedom.

Mathematical theory of the minimax filtering of functions from $\mathcal{F}^0(W,P)$ is nowadays well-developed. The most precise results can be obtained by combining Pinsker (1980) minimax theorem and asymptotic theory of prolate spheroidal wave functions (see Slepjan & Pollak (1961), Landau & Pollak (1961), Landau & Pollak (1962), Slepjan (1965)). Define the minimax $\mathbf{L}_2$–risk over a functional set $\mathcal{F}$ by

$$r_n(\mathcal{F}) = \inf_{\tilde{f}} \sup_{f\in\mathcal{F}} \mathbf{E}_f \int_0^1 (\tilde{f}(t) - f(t))^2 \, \mathrm{d}x,$$

where *inf* is taken over all estimators. In order to compute the minimax risk over entire analytic functions denote by $\Psi_k(x)$ and λ_k the eigen functions and the eigen values of the compact operator on $[0,1]$ with the kernel

$$K_W(x) = \frac{\sin(\pi W x)}{\pi x}. \tag{3}$$

In other words $\Psi_k(x)$ and λ_k are defined as solutions of the equation

$$\int_0^1 K_W(x-y)\Psi_k(y)\,\mathrm{d}y = \lambda_k \Psi_k(x),\ \ x\in[0,1].$$

The following theorem (cf. Pinsker (1980)) describes the asymptotic behavior of the minimax risk and the asymptotically minimax estimator.

Theorem 1 *Let* $W_n \geq W_0 > 0$.

- *As* $n \to \infty$

$$r_n(\mathcal{F}^0(W_n,P)) = (1+o(1))\frac{\sigma^2}{n}\sum_{k=1}^{\infty}\left[1-\mu\lambda_k^{-1/2}\right]_+,$$

where $[x]_+ = \max(0,x)$ *and* μ *is a root of the equation*

$$\frac{\sigma^2}{n}\sum_{k=1}^{\infty}\lambda_k^{-1}\left[\mu^{-1}\lambda_k^{-1/2}-1\right]_+ = P.$$

- *The asymptotically minimax estimator is*

$$f_n(t) = \sum_{k=1}^{\infty}\left[1-\mu\lambda_k^{-1/2}\right]_+ \Psi_k(t)\int_0^1 \Psi_k(u)\,\mathrm{d}X_n(u).$$

In order to apply the above theorem one has to know the asymptotic behavior of the eigen values λ_k. Fortunately the prolate spheroidal wave functions and the corresponding eigen values are well-studied. In particular, from Slepjan (1965) it follows that if

$$k = W_n + 1 + \frac{\alpha}{\pi^2}\log(2\pi W_n)$$

then as $n \to \infty$

$$\lambda_k = (1 + o(1)) \frac{1}{1 + \mathrm{e}^{\alpha}}.$$

Using this result we can simplify Theorem 1 provided that

$$\lim_{n \to \infty} \frac{W_n}{\log n \log W_n} = \infty. \tag{4}$$

Theorem 2 *Let (4) is satisfied. Then as $n \to \infty$*

$$r_n(\mathcal{F}^0(W_n, P)) = (1 + o(1)) \frac{\sigma^2 W_n}{n}.$$

The condition (4) is very valuable in non-parametric estimation. It shows when errors due to boundary effects are negligible with respect to the minimax risk. For instance, if the bandwidth $W = W_n$ is fixed and does not depend on n then asymptotic behavior of the risk is completely defined by the boundary effects. In this case $\lambda_n = O(\exp(-Cn))$ and $\mathbf{L}_2$-risk has the order of $\log n / n$ (see Theorem 1). Comparing this result with one from Ibragimov & Khasminskii (1982) we see the difference between density and regression estimation. For the density estimation problem on the hole real line the minimax $\mathbf{L}_2$-risk converge to zero with the rate W/n.

Unfortunately Theorem 2 does not say how to construct a simple asymptotically efficient estimator. Formally one can use the minimax estimator from Theorem 1. But since it involves the prolate spheroidal wave function its practical usage is very restrictive. The main difficulty in the construction of the minimax regression estimators is connected with the boundary problem. In modern statistic there are a lot of methods to overcome this difficulty. Boundary correction kernels, local polynomials and splines are one of them. But the most attractive from theoretical point of view results were obtained for wavelet method. In Donoho & Johnstone (1995) was shown that boundaries does not affect on the asymptotic performance of the wavelet method. In Hall & McKay & Turlach (1996) a detail behavior of $\mathbf{L}_2$-risk was decsribed for a fixed regression function with discontinuities. Here we demonstrate that a wavelet estimator not only overcome the boundary problem but simultaneously provide an asymptotically efficient estimators over the functional class $\mathcal{F}^0(W, P)$. In order to get such an estimator one has to use a special type of compactly supported wavelets (e.g. Daubechie's wavelet). The flexibility of wavelet methods gives us a possibility to consider also a broader functional class. Roughly speaking, this functional class consists of smooth functions with a finite number of discontinuities. Later on we deal with the functional class $\mathcal{F}^M(W, P)$, which can be defined in the following way. Let $\mathcal{A}^M = \{A_k\}_1^M$ be a partition of the interval $[0, 1]$

$$[0, 1] = \bigcup_{k=1}^{M} A_k^M, \quad A_k \cap A_j = \emptyset, \ k \neq j, \quad A_k^M = [t_k, t_{k+1}).$$

We say that function f belongs to $\mathcal{F}^M(W, P)$ if there exists a partition $\mathcal{A}^M$ possibly depending on f such that

$$f(t) = \sum_{k=1}^{M} f_k(t) \mathbf{1}\{t \in A_k\},$$

where $f_k \in \mathcal{F}^0(W,P)$. For simplicity it is assumed in the sequel that M is fixed but not known.

2 Main result

Let $\hat{g}(\xi)$ be the Fourier transform of the function $g(x) \in \mathbf{L}_2(-\infty,\infty)$

$$\hat{g}(\xi) = \lim_{\alpha\to 0}\int_{-\infty}^{\infty} \mathrm{e}^{-\mathrm{i}\,x\xi-\alpha^2x^2} g(x)\,\mathrm{d}x.$$

Orthogonal wavelet transform is ordinary based on two $\mathbf{L}_2$-orthogonal functions $\varphi(x)$ and $\psi(x)$. These functions are called the father and the mother functions and chosen in such a way that the functions

$$\varphi(x-k), \quad 2^{j/2}\psi(2^jx-l), \quad k,l \in (-\infty,\infty),\ j \geq 0$$

form an orthonormal basis in $\mathbf{L}_2(-\infty,\infty)$. This requirement leads to the well-known formulas (cf. Daubechies (1995))

$$\hat{\varphi}(\xi) = m_0\left(\frac{\xi}{2}\right)\hat{\varphi}\left(\frac{\xi}{2}\right), \tag{5}$$

$$\hat{\psi}(\xi) = \bar{m}_0\left(\frac{\xi}{2}+\pi\right)\hat{\varphi}\left(\frac{\xi}{2}\right)\mathrm{e}^{-\mathrm{i}\,\xi/2}, \tag{6}$$

where $m_0(\xi)$ is a 2π-periodic function such that

$$|m_0(\xi)|^2 + |m_0(\xi+\pi)|^2 = 1 \tag{7}$$

and $\bar{m}_0(\xi)$ denotes complex conjugate of $m_0(\xi)$. Further we will assume that $m_0(\xi) = m_0^N(\xi)$, $N \geq N_0$ is a sequence of functions satisfying (7) and the following conditions:

i) $m_0^N(\xi)$ is a trigonometric polynomial of the degree N,

ii) $|m_0^N(\xi)|^2 \geq 1 - \left(\frac{\xi}{2}\right)^{2N}$ uniformly in $|\xi| \leq 1$,

iii) $\max_{\xi\in[\pi(1+h)/2,\pi]} |m_0^N(\xi)|^2 \leq \exp(-A_0Nh^2)$,

where A_0 does not depend on N.

It is not very difficult to check that such a sequence exists. Consider for instance Daubechies wavelets. In this case

$$|m_0^N(\xi)|^2 = C_N\int_{\xi}^{\pi}\sin^{2N-1}(x)\,\mathrm{d}x,$$

where the constant C_N is chosen such that $|m_0^N(0)|^2 = 1$. The Laplace method easily reveals that as $N \to \infty$

$$C_N = \sqrt{\frac{2N-1}{2\pi}}(1+o(1)).$$

Therefore for $|\xi| \le \pi/2$ and for a sufficiently large N one obtains

$$\begin{aligned} |m_0^N(\xi)|^2 &= 1 - C_N \int_0^\xi \sin^{2N-1}(x)\,\mathrm{d}x \ge 1 - C_N \int_0^\xi \left(\frac{2x}{\pi}\right)^{2N-1} \mathrm{d}x \\ &\ge 1 - \left(\frac{\xi}{2}\right)^{2N}. \end{aligned}$$

On the other hand for $\xi \in [\pi(1+h)/2, \pi]$ one has

$$\begin{aligned} |m_0^N(\xi)|^2 &\le C_N \frac{\pi}{2} \sin^{2N-1}\left(\frac{\pi}{2} + \frac{\pi h}{2}\right) \le C_N \frac{\pi}{2} \cos^{2N-1}\left(\frac{\pi h}{2}\right) \\ &\le C_N \frac{\pi}{2} \left(1 - \left(\frac{\pi h}{4}\right)^2\right)^{2N-1} \le \sqrt{2N-1}\exp(-N\pi^2 h^2/8) \\ &\le \exp(-Nh^2). \end{aligned}$$

To stress that the father and the mother functions associated with $m_0^N(\xi)$ depend on N we supply them by the superscript N.

Denote for brevity $\varphi_k^N(x) = \varphi^N(x+k)$ and $\psi_{jk}^N(x) = 2^{j/2}\psi^N(2^j x + k)$. Then any function $f \in \mathbf{L}_2(-\infty, \infty)$ admits the following wavelet decomposition

$$f(t) = \sqrt{W} \sum_{k=-\infty}^{\infty} \alpha_k \varphi_k^N(tW) + \sqrt{W} \sum_{j=0}^{\infty} \sum_{k=-\infty}^{\infty} \beta_{jk} \psi_{jk}^N(tW), \tag{8}$$

where

$$\alpha_k = \sqrt{W} \int_{-\infty}^{\infty} f(x) \varphi_k^N(xW)\,\mathrm{d}x, \quad \beta_{jk} = \sqrt{W} \int_{-\infty}^{\infty} f(x) \psi_{jk}^N(xW)\,\mathrm{d}x.$$

Formula (8) provides us with a naive idea for recovering the regression function from noisy data. This idea is very simple. Let us replace α_k and β_{jk} by their empirical counterparts

$$\tilde{\alpha}_k = \sqrt{W} \int_0^1 \varphi_k^N(xW)\,\mathrm{d}X_n(x), \quad \tilde{\beta}_{jk} = \sqrt{W} \int_0^1 \psi_{jk}^N(xW)\,\mathrm{d}X_n(x).$$

Evidently that some additional filtering procedure is required in order to suppress the noise at high frequencies and to get an estimator with good statistical properties. One can do this by the using of the very popular hard thresholding procedure proposed by Donoho, Johnstone, Kerkyacharian, and Picard in the years 1990–1994. They proposed the estimator

$$f_n(t) = \sqrt{W} \sum_{k=-\infty}^{\infty} \tilde{\alpha}_k \varphi_k^N(tW) + \sqrt{W} \sum_{j=0}^{j_1} \sum_{k=-\infty}^{\infty} \tilde{\beta}_{jk}^* \psi_{jk}^N(tW), \tag{9}$$

where β_{jk}^* are the hard thresholded empirical mother coefficients

$$\beta_{jk}^* = \mathbf{1}\left\{|\tilde{\beta}_{jk}| \ge t_n\right\} \tilde{\beta}_{jk}, \quad \text{with } t_n = \sqrt{\frac{2\sigma^2 \log n}{n}}.$$

For motivation of this procedure we refer the reader to the nice book by Härdle& Kerkyacharian & Picard & Tsybakov (1998). We show that under appropriate choice of the smoothing parameters N, j_1, W the estimator (9) will be asymptotically efficient.

Theorem 3 *Let*

$$\lim_{n\to\infty} \frac{W_n}{\log^3 n} = \infty$$

and the parameters of the estimator (9) be

$$W = \frac{W_n}{1-h}, \quad N = \frac{(1+h)\log n}{A_0 h^2}, \quad j_1 = \log_2 \frac{n}{W_n}, \tag{10}$$

where $h > 0$. Then there exists sufficiently large integer $n(h)$ such that uniformly in $n \geq n(h)$

$$\sup_{f\in\mathcal{F}^M(W_n,P)} \mathbf{E}_f \|f_n - f\|^2 \leq (1+2h)\frac{\sigma^2 W_n}{n}.$$

3 Proof of Theorem 3

Let Π^Q be the projector

$$\Pi^Q g(x) = \int_{-\infty}^{\infty} K_Q(x-y)g(y)\,\mathrm{d}y, \tag{11}$$

with kernel $K_Q(\cdot)$ defined by (3). Denote for brevity $\Pi^Q g(x) = g^Q(x)$. By the convolution theorem

$$\hat{g}^Q(\xi) = \hat{g}(\xi)\mathbf{1}\{|\xi| \leq \pi Q\}. \tag{12}$$

In the following lemma we estimate from above the $\mathbf{L}_2$–distance between f and f^Q.

Lemma 1 *Let $Q \geq 2W_n$. Then uniformly in $f \in \mathcal{F}^M(W_n, P)$*

$$\|f - f^Q\|^2 \leq CP/Q. \tag{13}$$

Proof. By the definition, the function $f(t) \in \mathcal{F}^M(W_n, P)$ admits the following representation

$$f(t) = \sum_{l=1}^{M} f_l(t)\mathbf{1}\{t \in A_l\}, \tag{14}$$

where $\mathrm{supp}\hat{f}_l(\xi) \in [-\pi W_n, \pi W_n]$. Let $\chi_{A_l}(x) = \mathbf{1}\{x \in A_l\}$. By the Parseval formula and (12) one obtains

$$\begin{aligned} \|\chi_{A_l} - \Pi^{Q/2}\chi_{A_l}\|^2 &\leq \frac{1}{2\pi}\int_{-\infty}^{\infty} (1 - \mathbf{1}\{|\xi| \leq \pi Q/2\})^2 \,\frac{\sin^2(\mathrm{mes}A_l\xi)}{\xi^2}\,\mathrm{d}\xi \\ &\leq \frac{1}{2\pi}\int_{|\xi|>\pi Q/2} \xi^{-2}\,\mathrm{d}\xi \leq \frac{1}{\pi^2 Q}. \end{aligned} \tag{15}$$

Since $Q \geq 2W_n$ we have $\Pi^Q f_l \Pi^{Q/2}\chi_{A_l}(x) = f_l(x)\Pi^{Q/2}\chi_{A_l}(x)$. Hence from (15) we see that

$$\begin{aligned} \|f_l\chi_{A_l} - \Pi^Q f_l\chi_{A_l}\| &= \|f_l\chi_{A_l} - \Pi^Q f_l\Pi^{Q/2}\chi_{A_l} - \Pi^Q(f_l\chi_{A_l} - f_l\Pi^{Q/2}\chi_{A_l})\| \\ &\leq 2\|f_l\|\|\chi_{A_l} - \Pi^{Q/2}\chi_{A_l}\| \leq C\sqrt{Q/P}. \end{aligned}$$

The above inequality and (14) prove (13). □

The main idea in the proving of Theorem 3 is to show that the wavelet decomposition provides a good approximation to the ideal low pass filter. In order to prove this we begin with the following auxiliary lemma, which describes a behavior of $|\hat{\varphi}^N(\xi)|^2$.

Lemma 2 *Assume that the function $m_0^N(\xi)$ satisfies conditions ii)–iii). Then for sufficiently small h*

$$\max_{\xi\in K_h}(1 - |\hat{\varphi}^N(\xi)|^2) \leq C\exp(-A_0Nh^2),$$

where $K_h = [-(1-h)\pi, (1-h)\pi]$.

Proof. By (5) we have

$$|\hat{\varphi}^N(\xi)|^2 = \prod_{j=1}^{\infty}\left|m_0^N\left(\frac{\xi}{2^j}\right)\right|^2.$$

Therefore on easily obtains

$$\begin{aligned} 1 - |\hat{\varphi}^N(\xi)|^2 &= 1 - \left|m_0^N\left(\frac{\xi}{2}\right)\right|^2 \prod_{j=2}^{\infty}\left|m_0^N\left(\frac{\xi}{2^j}\right)\right|^2 \\ &= 1 - \left|m_0^N\left(\frac{\xi}{2}\right)\right|^2 - \left|m_0^N\left(\frac{\xi}{2}\right)\right|^2\left(\prod_{j=2}^{\infty}\left|m_0^N\left(\frac{\xi}{2^j}\right)\right|^2 - 1\right) \\ &\leq 1 - \left|m_0^N\left(\frac{\xi}{2}\right)\right|^2 + 1 - \prod_{j=2}^{\infty}\left|m_0^N\left(\frac{\xi}{2^j}\right)\right|^2 \\ &\leq \sum_{j=1}^{\infty}\left(1 - \left|m_0^N\left(\frac{\xi}{2^j}\right)\right|^2\right). \end{aligned}$$

Hence it follows from ii) – iii) and (7) that

$$\begin{aligned} \max_{\xi\in K_h}(1 - |\hat{\varphi}^N(\xi)|^2) &\leq \sum_{j=1}^{\infty}\max_{\xi\in 2^{-j}K_h}(1 - |m_0^N(\xi)|^2) \\ &\leq \max_{\xi\in K_h/2}(1 - |m_0^N(\xi)|^2) + \sum_{j=2}^{\infty}\max_{\xi\in 2^{-j}K_h}(1 - |m_0^N(\xi)|^2) \\ &\leq \exp(-A_0Nh^2) + \left(\frac{\pi}{4}\right)^{2N}, \end{aligned}$$

thus proving the lemma. □

Let us consider a function $f \in \mathcal{F}^0(W_n, P)$. Our next step is to show that this function can be well approximated in $\mathbf{L}_2(-\infty,\infty)$ by its projection on the space spanned by functions $\varphi_k(Wx)$. Define the approximation error

$$r_n(t,f) = \sqrt{W}\sum_{j=0}^{\infty}\sum_{k=-\infty}^{\infty}\beta_{jk}\psi_{jk}^N(Wt) = f(t) - \sqrt{W}\sum_{k=-\infty}^{\infty}\alpha_k\varphi_k^N(Wt). \tag{16}$$

Lemma 3 *Let $W = W_n/(1-h)$. Then uniformly in $f \in \mathcal{F}^0(W_n, P)$*

$$\int_{-\infty}^{\infty} r_n^2(t,f)\,\mathrm{d}t \le CP\exp(-A_0Nh^2).$$

Proof. The Poisson and the Parseval formulas reveal that (see also (16))

$$\begin{aligned}
\hat{r}_n(\xi,f) &= \hat{f}(\xi) - \hat{\varphi}^N(\xi/W)\sum_{k=-\infty}^{\infty}\hat{f}(\xi-2\pi kW)\hat{\varphi}^N(\xi/W-2\pi k)\\
&= \hat{f}(\xi)\left(1-\left|\hat{\varphi}^N(\xi/W)\right|^2\right)\\
&\quad - \hat{\varphi}^N(\xi/W)\sum_{k\neq 0}\hat{f}(\xi-2\pi kW)\hat{\varphi}^N(\xi/W-2\pi k).
\end{aligned}\tag{17}$$

The first term in the right-hand side of the above equation is evaluated by Lemma 2. Thus we have

$$\begin{aligned}
&\int_{-\infty}^{\infty}|\hat{f}(\xi)|^2\left(1-\left|\hat{\varphi}^N(\xi/W)\right|^2\right)^2\mathrm{d}\xi\\
&\quad= \int_{-\pi W_n}^{\pi W_n}|\hat{f}(\xi)|^2\left(1-\left|\hat{\varphi}^N(\xi/W)\right|^2\right)^2\mathrm{d}\xi\\
&\quad\le 2\pi P\max_{\xi\in[-\pi(1-h),\pi(1-h)]}(1-|\hat{\varphi}^N(\xi)|^2)^2 \le CP\exp(-2A_0Nh^2).
\end{aligned}$$

Next using the well-known formula (cf. (5), (7))

$$\sum_{k=-\infty}^{\infty}|\hat{\varphi}^N(\xi/W+2\pi k)|^2 = 1$$

and noting that the supports of the functions $\hat{f}(\xi-2\pi kW)$, $\hat{f}(\xi-2\pi lW)$, $k\neq l$ do not intersect, one obtains

$$\begin{aligned}
&\int_{-\infty}^{\infty}\left|\hat{\varphi}^N(\xi/W)\sum_{k\neq 0}\hat{f}(\xi-2\pi kW)\hat{\varphi}^N(\xi/W-2\pi k)\right|\mathrm{d}\xi\\
&\quad= \sum_{k\neq 0}\int_{-\infty}^{\infty}|\hat{f}(\xi-2\pi kW)|^2\left|\hat{\varphi}^N(\xi/W-2\pi k)\right|^2\left|\hat{\varphi}^N(\xi/W)\right|^2\mathrm{d}\xi\\
&\quad= \sum_{k\neq 0}\int_{-\infty}^{\infty}|\hat{f}(\xi)|^2\left|\hat{\varphi}^N(\xi/W+2\pi k)\right|^2\left|\hat{\varphi}^N(\xi/W)\right|^2\mathrm{d}\xi\\
&\quad= \int_{-\infty}^{\infty}|\hat{f}(\xi)|^2\left(1-\left|\hat{\varphi}^N(\xi/W)\right|^2\right)\left|\hat{\varphi}^N(\xi/W)\right|^2\mathrm{d}\xi\\
&\quad\le \int_{-\pi W_n}^{\pi W_n}|\hat{f}(\xi)|^2\left(1-\left|\hat{\varphi}^N(\xi/W)\right|^2\right)\mathrm{d}\xi\\
&\quad\le 2\pi P\max_{\xi\in[-\pi(1-h),\pi(1-h)]}(1-|\hat{\varphi}^N(\xi)|^2) \le CP\exp(-A_0Nh^2).
\end{aligned}$$

This inequality together with the Parseval formula and (17), (18) completes the proof. □

Two methods of noise reduction have been used in the estimator (9). They are the truncation of the infinite series in (8) starting from the resolution level j_1+1, and the thresholding. The truncation is equivalent to the projection on the space spanned by the functions $\varphi^N_{j_1k}(Wx)=\varphi^N(2^{j_1}Wx-k),\ k\in(-\infty,\infty)$. Therefore our next goal is to control the approximation error which is due to the truncation. In other words we have to estimate from above the $\mathbf{L}_2$-norm of the function

$$\rho_n(t,f)=\sqrt{W}\sum_{j\ge j_1}^{\infty}\sum_{k=-\infty}^{\infty}\beta_{jk}\psi^N_{jk}(Wt)=f(t)-\sqrt{W}\sum_{k=-\infty}^{\infty}\alpha_{j_1k}\varphi^N_{j_1k}(Wt), \tag{18}$$

where

$$\alpha_{j_1k}=\sqrt{W}\int_{-\infty}^{\infty}\varphi^N_{j_1k}(Wx)f(t)\,\mathrm{d}t.$$

Lemma 4 *Uniformly in* $f\in\mathcal{F}^0(W_n,P)$

$$\int_{-\infty}^{\infty}\rho_n^2(t,f)\,\mathrm{d}t\le CP2^{-j_1}W^{-1}+CP\exp(-A_0Nh^2).$$

Proof. Denote for brevity $W_1=2^{j_1}W,\ Q=(1-h)W_1$. Let $f^Q(t)=\Pi^Qf(t)$ (see (11)) and $\rho^Q_n(t)=\rho_n(t,f^Q)$. Since $\sqrt{W}\psi_{jk}(tW)$ is the orthonormal system in $\mathbf{L}_2(-\infty,\infty)$ we have from (18)

$$\|\rho_n(\cdot,f)\|\le\|f-f^Q\|+\|\rho^Q_n\|. \tag{19}$$

Since $Q\ge 2W_n$, the first term in right-hand side of this inequality is evaluated by Lemma 1 as

$$\|f-f^Q\|^2\le CP/Q. \tag{20}$$

In order to estimate $\|\rho^Q_n\|^2$ we apply the same arguments as in the proving of Lemma 3. Thus we have (cf. (17))

$$\begin{aligned}\hat{\rho}^Q_n(\xi,f) &= \hat{f}^Q(\xi)\left(1-\left|\hat{\varphi}^N(\xi/W_1)\right|^2\right)\\ &\quad-\hat{\varphi}^N(\xi/W_1)\sum_{k\ne0}\hat{f}^Q(\xi-2\pi kW_1)\bar{\hat{\varphi}}^N(\xi/W_1-2\pi k).\end{aligned} \tag{21}$$

By Lemma 2 we have

$$\begin{aligned}&\int_{-\infty}^{\infty}\left|\hat{f}^Q(\xi)\right|^2\left(1-\left|\hat{\varphi}^N(\xi/W_1)\right|^2\right)^2\mathrm{d}\xi\\ &\le\max_{\xi\in[-\pi Q,\pi Q]}\left(1-\left|\hat{\varphi}^N(\xi/W_1)\right|^2\right)^2\int_{-\pi Q}^{\pi Q}\left|\hat{f}^Q(\xi)\right|^2\mathrm{d}\xi\le CP\exp(-2A_0Nh^2).\end{aligned}$$

The $\mathbf{L}_2$-norm of the remainder term in (21) is estimated from above as

$$\int_{-\infty}^{\infty}\left|\hat{\varphi}^N(\xi/W_1)\sum_{k\ne0}\hat{f}(\xi-2\pi kW_1)\bar{\hat{\varphi}}^N(\xi/W_1-2\pi k)\right|^2\mathrm{d}\xi$$

$$= \sum_{k\neq 0} \int_{-\infty}^{\infty} |\hat{f}(\xi)|^2 \left|\hat{\varphi}^N(\xi/W_1)\right|^2 \left|\hat{\varphi}^N(\xi/W_1 + 2\pi k)\right|^2 \,\mathrm{d}\xi$$
$$\leq \int_{-\pi Q}^{\pi Q} |\hat{f}(\xi)|^2 \left(1 - \left|\hat{\varphi}^N(\xi/W_1)\right|^2\right) \mathrm{d}\xi$$
$$\leq CP \max_{\xi\in[-\pi(1-h),\pi(1-h)]} (1 - |\hat{\varphi}^N(\xi)|^2) \leq CP\exp(-A_0 N h^2).$$

Thus combining these inequalities with (19)–(21) one arrives at the assertion of the lemma. □

The thresholding idea is based on the simple and intuitively clear property of the wavelet transform. If the underlying function is sufficiently smooth then the mother coefficients are small. In the following lemma we give a strong motivation for this heuristic idea.

Lemma 5 *Let $W = W_n/(1-h)$. Assume that for some j,k the support of the function $\psi_{jk}^N(Wx)$ belongs to some A_l. Then*

$$|\beta_{jk}| \leq C\sqrt{P}2^{-j/2}\exp(-A_0 h^2 N/2).$$

Proof. From the Parseval formula, the Caushy-Schwartz inequality, and (14) it follows that

$$\begin{aligned} |\beta_{jk}| &\leq \frac{1}{2\pi}\left(2^j W\right)^{-1/2} \int_{-\infty}^{\infty} |\hat{f}_l(\xi)| \left|\psi^N\left(\frac{\xi}{2^j W}\right)\right| \mathrm{d}\xi \\ &\leq \frac{1}{2\pi}\left(2^j W\right)^{-1/2} \sqrt{P} \left(\int_{-\pi W_n}^{\pi W_n} \left|\psi^N\left(\frac{\xi}{2^j W}\right)\right| \mathrm{d}\xi\right)^{1/2} \\ &\leq C\sqrt{P}2^{-j/2} \max_{\xi\in[-\pi(1-h),\pi(1-h)]} |\psi^N(\xi)|. \qquad (22) \end{aligned}$$

By (6) we have $|\psi^N(\xi)| \leq |m_0^N(\xi/2+\pi)|$. Therefore from the condition iii) one obtains

$$\begin{aligned} \max_{\xi\in[-\pi(1-h),\pi(1-h)]} |\psi^N(\xi)| &\leq \max_{\xi\in[\pi-\pi(1-h)/2,\pi+\pi(1-h)/2]} |m_0^N(\xi)| \\ &\leq \exp(-A_0 h^2 N/2). \end{aligned}$$

This inequality together with (22) completes the proof. □

Proof of Theorem 3. We decompose the risk of the estimator $f_n(t)$ from (9) as follows

$$\begin{aligned} \mathbf{E}\|f_n - f\|^2 &\leq \frac{\sigma^2}{n} \sum_{k=-\infty}^{\infty} \mathbf{1}\{\mathrm{supp}\ \varphi_k^N(xW) \cap [0,1] \neq \emptyset\} \\ &\quad + \sum_{j=0}^{j_1} \beta_{jk}^2 \mathbf{P}\left\{|\tilde{\beta}_{jk}| \leq t_n\right\} \\ &\quad + \sum_{j=0}^{j_1} \mathbf{E}(\tilde{\beta}_{jk} - \beta_{jk})^2 \mathbf{1}\left\{|\tilde{\beta}_{jk}| > t_n\right\} + W \sum_{j=j_1+1}^{\infty} \beta_{jk}^2 \\ &= R_1 + R_2 + R_3 + R_4. \qquad (23) \end{aligned}$$

According to i) the support of the father function $\varphi(xW)$ belongs to an interval with the length of the order of N/W. Therefore

$$R_1 \leq \frac{\sigma^2 W}{n} + O\left(\frac{N}{n}\right). \tag{24}$$

The second term in (23) can be evaluated from above by the following way. Denote

$$\eta_{jk} = \sqrt{W}\int_0^1 \psi_{jk}(tW)\,\mathrm{d}w(t).$$

Then we have

$$\begin{aligned} R_2 &= \sum_{j=0}^{j_1}\sum_{k=-\infty}^{\infty}\beta_{jk}^2 \mathbf{E}\,\mathbf{1}\left\{\left|\beta_{jk}+\frac{\sigma}{\sqrt{n}}\eta_{jk}\right|\leq t_n\right\} \\ &= \sum_{j=0}^{j_1}\sum_{k=-\infty}^{\infty}\beta_{jk}^2 \mathbf{E}\,\mathbf{1}\left\{\left|\beta_{jk}+\frac{\sigma}{\sqrt{n}}\eta_{jk}\right|\leq t_n\right\}\mathbf{1}\left\{|\eta_{jk}|\geq\sqrt{2\log n}\right\} \\ &\quad+\sum_{j=0}^{j_1}\sum_{k=-\infty}^{\infty}\beta_{jk}^2 \mathbf{E}\mathbf{1}\left\{\left|\beta_{jk}+\frac{\sigma}{\sqrt{n}}\eta_{jk}\right|\leq t_n\right\}\mathbf{1}\left\{|\eta_{jk}|<\sqrt{2\log n}\right\} \\ &\leq \sum_{j=0}^{j_1}\sum_{k=-\infty}^{\infty}\beta_{jk}^2\mathbf{1}\left\{|\beta_{jk}|\leq 2t_n\right\}+\sum_{j=0}^{j_1}\sum_{k=-\infty}^{\infty}\beta_{jk}^2\mathbf{P}\left\{|\eta_{jk}|\geq\sqrt{2\log n}\right\}. \end{aligned} \tag{25}$$

Since

$$\mathbf{P}\left\{|\eta_{jk}|\geq\sqrt{2\log n}\right\} = O\left(n^{-1}\right)$$

we get the following upper bound for the last term in (25)

$$\sum_{j=0}^{j_1}\sum_{k=-\infty}^{\infty}\beta_{jk}^2\mathbf{P}\left\{|\eta_{jk}|\geq\sqrt{2\log n}\right\}\leq O\left(n^{-1}\right)P. \tag{26}$$

In order to evaluate the first term in the right-hand side (25) denote by K^j the set of all integers such that if $k\in K^j$ then the support of $\psi_{jk}(xW)$ belongs to A_l for some $l\in[1,M]$. Then one easily arrives at

$$\sum_{j=0}^{j_1}\sum_{k=-\infty}^{\infty}\beta_{jk}^2\mathbf{1}\left\{|\beta_{jk}|\leq 2t_n\right\}\leq\sum_{j=0}^{j_1}\sum_{k\in K^j}\beta_{jk}^2+\frac{4\sigma^2\log n}{n}\sum_{j=0}^{j_1}\sum_{k\notin K^j}\mathbf{1}\{\beta_{jk}\neq 0\}. \tag{27}$$

By Lemma 3 and (10) one obtains

$$\sum_{j=0}^{j_1}\sum_{k\in K^j}\beta_{jk}^2\leq CP\exp(-A_0Nh^2)=O\left(n^{-1}\right)P.$$

Since the support of $\psi_{jk}(xW)$ belongs to an interval with the length of the order of $N/(2^jW)$ we obviously have

$$\sum_{k\notin K^j}\mathbf{1}\{\beta_{jk}\neq 0\}\leq O\left(N\right).$$

Thus we obtain by (25)–(27)

$$R_2 \le O\left(\log^3 n/n\right). \tag{28}$$

Our next step is to evaluate from above the term R_3 in (23). By a simple algebra we have

$$\begin{aligned} R_3 &= \frac{\sigma^2}{n}\sum_{j=0}^{j_1}\sum_{k=-\infty}^{\infty}\mathbf{E}\eta_{jk}^2\mathbf{1}\left\{\left|\beta_{jk}+\frac{\sigma}{\sqrt{n}}\eta_{jk}\right|>t_n\right\} \\ &\le \frac{\sigma^2}{n}\sum_{j=0}^{j_1}\sum_{k\notin K^j}\mathbf{E}\eta_{jk}^2+\frac{\sigma^2}{n}\sum_{j=0}^{j_1}\sum_{k\in K^j}\mathbf{E}\eta_{jk}^2\mathbf{1}\left\{\left|\beta_{jk}+\frac{\sigma}{\sqrt{n}}\eta_{jk}\right|>t_n\right\}. \end{aligned} \tag{29}$$

Once again noting that $\psi_{jk}(xW)$ has a support of length of the order of $N/(2^jW)$, we see that

$$\sum_{k\notin K^j}\mathbf{E}\eta_{jk}^2 \le O(N). \tag{30}$$

The last term in (29) is estimated from above by Lemma 5

$$\begin{aligned} &\sum_{k\in K^j}\mathbf{E}\eta_{jk}^2\mathbf{1}\left\{\left|\beta_{jk}+\frac{\sigma}{\sqrt{n}}\eta_{jk}\right|>t_n\right\} \\ &\quad\le \sum_{k\in K^j}\mathbf{E}\eta_{jk}^2\mathbf{1}\left\{|\eta_{jk}|\ge\sqrt{2\log n}+O(n^{-h/2})\right\} \le\ C2^jWn^{-1}\sqrt{2\log n}. \end{aligned}$$

This inequality together with (29), (30) and (10) implies

$$R_3 \le O\left(\frac{\sigma^2\log n}{h^2 n}\right). \tag{31}$$

Finally we apply Lemma 3 to evaluate the remainder term R_4 in (23). So, by (10) one obtains that $R_4 \le O\left(n^{-1}\right)$. Now the assertion of the theorem easy follows from (23), (24), (28) and (31). □

References

[1] Daubechies, I. (1995). *Ten Lectures on Wavelets*, SIAM, Philadelphia.

[2] Donoho, D., Johnstone, I. (1994). Ideal spatial adaptation by wavelet shrinkage, *Biometrica* **81**, 425–455.

[3] Donoho, D., Johnstone, I. (1995). Adapting to Unknown Smoothness via Wavelet Shrinkage, *Journal of the American Statistical Association*. no. 432, 1200–1224.

[4] Donoho, D., Johnstone, I. (1998). Minimax Estimation via Wavelet Shrinkage, *Ann. Statist.* (to appear).

[5] Donoho, D., Johnstone, I., Kerkyacharian, G., Picard, D. (1995). Wavelet shrinkage: Asymtopija?, *Journal of the Royal Statistical Society*, Series B, **57**, 301–369.

[6] Donoho, D., Johnstone, I., Kerkyacharian, G., Picard, D. (1996). Density estimation by wavelet thresholding, *Ann. Statist.* **24**, 508–539.

[7] Gallager, R. (1968). *Information Theory and Reliable Communications*, John Wiley and sons, Inc. New York, London, Sydney, Toronto.

[8] Hall, P. McKay, I. Turlach, B. (1996). Performance of wavelet methods for functions with many discontinuities, *Ann. Statist.* **24**, 2462–2476.

[9] Härdle, W. Kerkyacharian, G. Picard, D. Tsybakov, A. (1998). *Wavelets, Approximation and Statistical Applications*. Springer Verlag (to appear).

[10] Ibragimov, I.A. Khasminskii R.Z. (1982) An estimate of the density of a distribution belonging to a class of entire functions. *Teor. Veroyatnost. i Primen.* **27**, 517–524.

[11] Landau, H.J. Pollak, H.Q. (1961). Prolate spheroidal wave function, Fourier analysis, and uncertainty – II. *Bell System Tech. J.* **40**, 65–84.

[12] Landau, H.J. Pollak, H.Q. (1962). Prolate spheroidal wave function, Fourier analysis, and uncertainty – III. *Bell System Tech. J.* **41**, 1295–1336.

[13] Pinsker, M.S. (1980). Optimal filtering of square integrable signals in Gaussian white noise *Problems of Information Transmission*, 120-133.

[14] Slepjan, D. (1965). Some asymptotic expansions for prolate spheroidal wave functions. *J. Math. and Rhys.* **44**, 99–140.

[15] Slepjan, D. Pollak, H.Q. (1961). Prolate spheroidal wave functions, Fourier analysis, and uncertainty – I. *Bell System Tech. J.* **40**, 43–64.