Wavelets-based non-parametric regression: optimal rate in the sup-norm

C.G.M. Oudshoorn

March 5, 1997

1 Introduction

The model we will study in this paper is a non-parametric regression model on the unit interval $[0,1]$ with equidistant deterministic grid design. The unknown regression function is assumed to belong to some Hölder class with the smoothness parameter $\beta > \frac{1}{2}$. Furthermore the regression function is observed in a Gaussian noise (precise definitions are given in SECTION 2).

Ibragimov and Hasminskii [7] have studied a similar kind of model. They considered a more general grid, namely a stochastic grid design, and the class of functions they took was the class of periodic functions on the interval $[0, 2\pi]$. Stone [12] obtained optimal rates of convergence where the regression function was assumed to belong to some Hölder class of functions. The same kind of model, but with the emphasis on getting the exact constants were studied in Korostelev [8] (for β in $]\frac{1}{2}, 1]$) and in Donoho [2] (for $\beta > \frac{1}{2}$). In both articles the optimal rates where obtained with kernel methods.

Our goal in this article is to obtain optimal rates by using wavelet estimators. Recently wavelet estimators were studied in the context of density estimation (cf. Picard and Kerkyacharian [11]) and in estimation the diffusion coefficient of a diffusion process (cf. Genon-Catalot, Laredo and Picard [5]). In Donoho and Johnstone [3] a non-linear wavelet estimator was used for estimating functions with jumps.

Regression has some specific problems, the treatment of the neighbourhoods of the end-points being one of them. For example in Müller (see [10]) this problem has been solved by taking special kernels at the neighbourhoods of the boundary. We will give another way of treating this problem by defining the estimator in the neighbourhoods of the endpoints as a Taylor-polynomial up to the order corresponding to the smoothness of the regression function (see SECTION 4).

The lower bound for the rate of convergence will be derived in SECTION 5.2 by a technic which uses the ideas close to the Hajeks derivation of the local asymptotic minimax lower bounds, as generalised in Ibragimov and Hasminskii [6] for the multivariate case (cf. also Korostelev [8]). In SECTION 5.1 we will prove that a wavelet estimator modified in the neighbourhoods of the endpoints achieves this lower bound. For the convenience of the reader we will give a summary of the theory of wavelets (in SECTION 3).

The author would like to use this opportunity to thank B. Ya. Levit for his stimulating ideas and frequent useful discussions.

2 Definitions and the model

We study the non-parametric regression problem with deterministic regular grid design on the interval $[0,1]$. Let $y_1, \ldots, y_n$ be observations from the following model:

$$\begin{cases} y_i = f(\frac{i}{n}) + \xi_i \\ \\ \xi_i \sim \mathcal{N}(0,\sigma^2) \end{cases} \qquad i = 1, \ldots, n \tag{2.1}$$

where

$$f \in \mathcal{F}^\beta = \left\{ f : \sup_{x,y\in[0,1]} \frac{|f^m(x) - f^m(y)|}{|x-y|^\alpha} + \sup_{x\in[0,1]} |f(x)| \leq L \right\}, \tag{2.2}$$

for some $L > 0$, $\frac{1}{2} < \beta = m + \alpha$, $0 < \alpha \leq 1$ and $m = [\beta]$ (here $[a]$ is notation for the biggest integer smaller than a). In the sequel $\|\cdot\|$ will be a shorthand notation for $\sup_{x\in[0,1]} |f(x)|$ and $\|\cdot\|_\infty$ for $\sup_{x\in\mathbf{R}} |f(x)|$.

We want to study optimal rate wavelets-based estimates of the unknown regression function f in the uniform norm. The notion of optimality we use below is that of Stone [12]:

Definition 2.1

- a) Φ_n *is a lower rate of convergence if there exists a positive* C_{low} *such that*

$$\lim_{n\to\infty} \inf_{f_n} \sup_{f\in\mathcal{F}^\beta} \mathbf{P}\left\{\|f_n - f\| \geq C_{\text{low}}\Phi_n\right\} = 1 \tag{2.3}$$

  where the infimum is taken over all estimators f_n *of* f.

- b) Φ_n *is an achievable rate of convergence if there exists a sequence* $\{f_n\}_{n\geq 1}$ *of estimators and a positive constant* C_{upp} *such that*

$$\lim_{n\to\infty} \sup_{f\in\mathcal{F}^\beta} \mathbf{P}\left\{\|f_n - f\| \geq C_{\text{upp}}\Phi_n\right\} = 0 \tag{2.4}$$

  The sequence f_n *satisfying (2.4) will be called* Φ_n*-rate consistent.*

- c) Φ_n *is called an optimal rate of convergence if it is both a lower and an achievable rate of convergence.*

When Φ_n is the optimal rate of convergence and a sequence of estimators $\{f_n\}$ satisfies (2.4), the estimators f_n, $n \geq 1$, are said to be asymptotically optimal (cf. [12]).

The lower rates of convergence for simular models have been obtained by Stone [12] and Ibragimov & Khasminskii [7] . More recently Korostelev [8] found the exact optimal constants for β between $\frac{1}{2}$ and 1. This result has been extended by Donoho [2] for $\beta > 1$. In SECTION 5 we generalize the method by Korostelev to give a new and more elementary proof of the following result (cf. Stone [12] , Ibragimov & Khasminskii [7]):

Theorem 2.2 $\Phi_n = \left(\dfrac{\log n}{n}\right)^{\frac{\beta}{2\beta+1}}$ *is a lower rate of convergence.*

Remark. Note that the optimal rate in L^p-norm, $1 < p < \infty$, is different from the L^∞-norm, namely $n^{-\frac{\beta}{2\beta+1}p}$.

Theorem 2.2 can be applied to the following class of risk functions:

$$R(f_n, f) = \mathbf{E}\left[\Phi_n^{-1}\|f_n - f\|\right]^\alpha \qquad \alpha > 0, \tag{2.5}$$

in the following way:

Corollary 2.3 $\lim_{n\to\infty} \inf_{f_n} \sup_{f\in\mathcal{F}^\beta} R(f_n, f) > 0.$

A wavelet type estimator achieving the optimal rate Φ_n will be described in SECTION 4. SECTION 3 presents a short summary of wavelet theory to be used later.

3 Wavelets: a summary

The easiest access to the theory of wavelets is provided by the notion of multiresolution analysis (see Meyer [9]):

Definition 3.1 *A multiresolution analysis (MRA) is a sequence of subspaces $(V_j)_{j\in\mathbf{Z}}$ of $L^2(\mathbf{R})$ with the following conditions:*

(i) $\ldots \subset V_{-1} \subset V_0 \subset V_1 \subset \ldots$

(ii) $\bigcap_{j\in\mathbf{Z}} V_j = \{0\}$; $\overline{\bigcup_{j\in\mathbf{Z}} V_j} = L^2(\mathbf{R})$

(iii) $f(\cdot) \in V_j \Leftrightarrow f(2\cdot) \in V_{j+1} \quad \forall j \in \mathbf{Z}$

(iv) $f(\cdot) \in V_0 \Rightarrow f(\cdot + k) \in V_0 \quad \forall k \in \mathbf{Z}$

(v) There exists a function φ, called a scaling function, such that $\{\varphi(\cdot - k), k \in \mathbf{Z}\}$ forms an orthonormal basis of V_0.

It is possible to assume that φ is of class C^r and compactly supported on $[0, 2N-1]$ (in that case it is proved that there exists a constant γ such that the length of the support is of order $\gamma \cdot r$). If φ has regularity r then the corresponding MRA is said to have regularity r. From (iii), (iv) and (v) it easily follows that $\left\{\varphi_{j,k} = 2^{\frac{j}{2}}\varphi(2^j\cdot - k);\ \ k \in \mathbf{Z}\right\}$ is an orthonormal basis of V_j. φ is usually called the mother wavelet.

Furthermore it can be proved that for the multiresolution analysis, the orthogonal complements W_j, defined as $V_{j+1} = W_j \bigoplus V_j$, have the same kind of properties as the V_j's. Thus there exists a function ψ, the father wavelet, such that $\psi \in C^r$, ψ is compactly supported and W_0 is spanned by the collection $\{\psi(\cdot - k);\ \ k \in \mathbf{Z}\}$. Then the collection of functions $\left\{\psi_{j,k} = 2^{\frac{j}{2}}\psi(2^j\cdot - k);\ \ j, k \in \mathbf{Z}\right\}$ is an orthonormal basis of $L^2(\mathbf{R})$.

From now on we will assume that a mother wavelet φ has been chosen such that $r \geq \max(1, \beta)$. The Daubechies wavelets provide well known examples of such functions (cf. [1]). Following notations will be used in the sequel, for $f \in L^2(\mathbf{R})$:

- $P_j f(x) =$ Projection of f on $V_j = \sum_{k\in\mathbf{Z}} c_{j,k}\varphi_{j,k}(x)$
  where $c_{j,k} = \int_{\mathbf{R}} f(y)\varphi_{j,k}(y)\,dy$
- $D_j f(x) =$ Projection of f on $W_j = \sum_{k\in\mathbf{Z}} d_{j,k}\psi_{j,k}(x)$
  where $d_{j,k} = \int_{\mathbf{R}} f(y)\psi_{j,k}(y)\,dy$.

Since φ and ψ are compactly supported the above sum comprises, for each x, only finite number of terms (though depending on N).

According to the definition and properties of the MRA a function $f \in L^2(\mathbf{R})$ can be decomposed for any j as follows:

$$f(x) = \sum_k c_{j,k}\varphi_{j,k}(x) + \sum_{j'\geq j}\sum_{k\in\mathbf{Z}} d_{j',k}\psi_{j',k}(x) = P_j f(x) + \sum_{j'\geq j} D_{j'} f(x) \tag{3.1}$$

with convergence in $L^2(\mathbf{R})$. Moreover for f in $\mathcal{F}^\beta$ (see (2.2)) we have:

Proposition 3.2 *If $(V_j)_{j\in\mathbf{Z}}$ is a MRA of regularity $r \geq \beta$ then the following implication holds:*

$$f \in \mathcal{F}^\beta \Rightarrow P_0 f \in L^\infty \text{ and } \|D_j f\|_\infty \leq C_{\mathrm{p}} 2^{-j\beta}$$

where C_{p} is a constant (to be determined explicitely in the proof).

Proof. The proof is a slight modification of the similar result of Meyer, proved in [9] for the so called Zygmund class of functions. The two classes differ only for the integer values of β (see Meyer p. 53).

Define $D(x,y) = \sum_k \psi(x-k)\overline{\psi}(y-k)$. This kernel has the property that it is perpendicular to all polynomials up to degree β (i.e. $\int D(x,y)y^\gamma\,dy = 0$ for $|\gamma| \leq \beta$). As we have taken a MRA of regularity at least β, $|D(x,y)| \leq C_l(1+|x-y|)^{-l}$, for $l \in \mathbf{N}$ (see Meyer [9]). Let $D_{(m)}$ be the kernel such that $\frac{\partial^m}{\partial y^m} D_{(m)} = D(x,y)$. For $D_{(m)}$ the same bound as for $D(x,y)$ holds. Taking this into account one can deduce:

$$\begin{aligned}
\|D_j f\|_\infty &= \sup_{x\in\mathbf{R}} \left| \int 2^j D(2^j x, 2^j y) f(y)\,dy \right| \\
&= \sup_{x\in\mathbf{R}} \left| \int 2^j D(2^j x, 2^j y)(f(y) - f(x))\,dy \right| \\
&= \sup_{x\in\mathbf{R}} \left| \int D(x,y)(f(2^{-j}y) - f(2^{-j}x))\,dy \right| \\
&\leq \sup_{x\in\mathbf{R}} 2^{-j\beta} L \int |D_{(m)}(x,y)|\,|y-x|^\alpha\,dy \\
&= 2^{-j\beta} C_{\mathrm{p}},
\end{aligned}$$

where the constant $C_{\mathrm{p}} = C_{\mathrm{p}}(L,\alpha,m)$ equals $L\,\|\int |D_{(m)}(\cdot,y)|\,|y-\cdot|^\alpha\,dy\|_\infty$. □

In the sequel we also need the following Bernstein's type inequality:

Proposition 3.3 *(Meyer [9] p. 47) If $(V_j)_{j\in\mathbf{Z}}$ is a MRA of regularity $r \geq \beta$ and $f \in \mathcal{F}^\beta$ then there exist constants C_1 and C_2 such that for $m = [\beta]$:*

$$C_1 2^{jm}\|D_j f\|_\infty \leq \|(D_j f)^{(m)}\|_\infty \leq C_2 2^{jm}\|D_j f\|_\infty \tag{3.2}$$

4 The Estimator and main Theorems

First we replace f by a function $\overline{f}$ such that for arbitrary, but fixed $\epsilon > 0$:

i) $f = \overline{f}$ on $[0, 1]$

ii) $\overline{f} \in L^2(\mathbf{R})$

iii) $\operatorname{supp} \overline{f} \subset [-\epsilon, 1 + \epsilon]$

iv) $\overline{f} \in \mathcal{F}^\beta([-\epsilon, 1 + \epsilon])$

Let $P_j\overline{f}$ be the projection of $\overline{f}$ on V_j where j, which we will define shortly, depends on n. As was mentioned in the previous SECTION we choose a sufficiently smooth wavelet φ which is compactly supported with support $[0, 2N - 1]$. N depends on the smoothness of φ.

For the projection $P_j\overline{f}$, evaluated in x, where x is in the interval $[2^{-j}(2N-1), 1-2^{-j}(2N-1)]$, we only need a finite number of $\varphi_{j,k}$, namely those for k runs from 0 to $2^j - 2N + 1$. Thus for the corresponding coefficients we have:

$$
\begin{aligned}
c_{j,k} &= \int_{\mathbf{R}} \varphi_{j,k}(x)\overline{f}(x)\,dx \\
&= \int_0^1 \varphi_{j,k}(x)\overline{f}(x)\,dx \\
&= \int_0^1 \varphi_{j,k}(x)f(x)\,dx.
\end{aligned}
$$

The obvious idea is now to estimate f by estimating the projection $P_j\overline{f}$. For this projection we have to estimate the corresponding coefficients $c_{j,k}$. Let us abbreviate $2^{-j}(2N-1)$ by h. A natural estimator $\hat{c}_{j,k}$ for $c_{j,k}$ is the following sum:

$$
\hat{c}_{j,k} = \frac{1}{n}\sum_{i=1}^{n} \varphi_{j,k}\left(\tfrac{i}{n}\right) y_i. \tag{4.1}
$$

Thus the estimator of f for $x \in [h, 1-h]$ is:

$$
f_n^{\mathrm{w}}(x) = \sum_{k\in\mathbf{Z}} \varphi_{j,k}(x)\hat{c}_{j,k}, \qquad x \in [h, 1-h]. \tag{4.2}
$$

It remains to define the estimator for $x \in [0, h] \cup [1-h, 1]$. Therefore we propose to extrapolate f_n^{w} based on the values $(f_n^{\mathrm{w}})^{(l)}(h)$ and $(f_n^{\mathrm{w}})^{(l)}(1-h)$ (for $l = 0, \dots, m$) by a Taylor polynomial of degree m, corresponding to the smoothness of f. Note that $(f_n^{\mathrm{w}})^{(l)}$ equals $2^{j(l+\frac{1}{2})}\sum_{k\in\mathbf{Z}} \varphi^{(l)}(2^j h - k)\hat{c}_{j,k}$. The precise definition of the proposed estimator is:

$$
f_n^{\mathrm{w}}(x) = \begin{cases}
\sum_{l=0}^{m} (f_n^{\mathrm{w}})^{(l)}(h)\frac{(x-h)^l}{l!} & x \in [0, h] \\
\sum_{k\in\mathbf{Z}} \varphi_{j,k}(x)\hat{c}_{j,k} & x \in [h, 1-h] \\
\sum_{l=0}^{m} (f_n^{\mathrm{w}})^{(l)}(1-h)\frac{(x-(1-h))^l}{l!} & x \in [1-h, 1]
\end{cases} \tag{4.3}
$$

Note that h here plays the role comparable to that of the bandwidth for the Kernel-type estimators. Therefore we call it, with some abuse of terminology, henceworth bandwidth. Let

$$h = \left(\frac{\log n}{n}\right)^{\frac{1}{2\beta+1}} \tag{4.4}$$

or equivalently $j = \lceil {}^2\log(2N-1) - {}^2\log h \rceil$ ($\lceil a \rceil$ is a shorthand notation for the smallest integer larger or equal to a). Now we can state the main Theorem:

Theorem 4.1 *The estimator f_n^{w} defined through (4.1), (4.2), (4.3) and (4.4) is $\left(\frac{\log n}{n}\right)^{\frac{\beta}{2\beta+1}}$-rate consistent.*

In SECTION 2 we have defined a risk function. Generalizing the proof of Theorem 4.1 one can show the following:

Theorem 4.2 $$\lim_{n\to\infty} \sup_{f\in\mathcal{F}^\beta} R(f_n^{\mathrm{w}}, f) < \infty$$

The proofs will be furnished in SECTION 5. Also an explicit upper bound for the asymptotic risk will be given.

5 Proofs

5.1 proof of Theorem 2.2

First we reduce the non-parametric problem to a parametric problem by choosing a special parametric subfamily of $\mathcal{F}^\beta$. Second we derive the sample distribution and describe sufficient statistics of the sample. Third we bound the minimax risk from below by the Bayes risk.

Suppose that $K : \mathbf{R} \to \mathbf{R}$ is a kernel with the following properties:

1 $K(0) = 1$

2 $K(u) = 0$ for $|u| \geq 1$

3 $K \in \mathcal{F}^\beta$

We have to introduce some definitions and notations: Let h, the bandwidth, equals

$$(n)^{-1} \lfloor nh_o \left(\frac{\log n}{n}\right)^{\frac{1}{2\beta+1}} \rfloor$$

where h_o will be defined later ($\lfloor a \rfloor$ is a notation for the largest integer smaller than or equal to a).

Define a grid on the interval $[0,1]$ as follows:

$$x_j = (2j-1)h \qquad j = 1, \ldots, J = \lfloor \tfrac{1}{2h} \rfloor.$$

By definition of the bandwidth, hn is an integer smaller than n, therefore the above grid is a subset of the sample points. The parametric family, depending on the vector parameter $\mathbf{c} = (c_1, \ldots, c_J)$, is defined as follows :

$$f_{\mathbf{c}}(x) = \sum_{j=1}^{J} h^\beta c_j K\left(\frac{x-x_j}{h}\right) \qquad |c_j| \leq 1.$$

Note that $f_{\mathbf{c}}(x_j) = h^\beta c_j$.

Of course we have to check whether $f_{\mathbf{c}}$ indeed belongs to $\mathcal{F}^\beta$. Therefore we introduce for $x, y \in [0,1]$ arbitrary, j^* such that $|x - x_{j^*}| \leq h$ and j' the same but with y instead of x. By construction of K and the grid there is a $\tilde{y}$ such that $|\tilde{y} - x_{j^*}| \leq h$ and $K^{(m)}\left(\frac{\tilde{y}-x_{j^*}}{h}\right) = K^{(m)}\left(\frac{y-x_{j'}}{h}\right)$. With these tools we can deduce:

$$\begin{aligned}
\left|f^{(m)}(x) - f^{(m)}(y)\right| &= \left|\sum_{j=1}^{J} h^\alpha c_j \left(K^{(m)}\left(\frac{x-x_j}{h}\right) - K^{(m)}\left(\frac{y-x_j}{h}\right)\right)\right| \\
&= \left|h^\alpha \left(c_{j^*} K^{(m)}\left(\frac{x-x_{j^*}}{h}\right) - c_{j'} K^{(m)}\left(\frac{y-x_{j'}}{h}\right)\right)\right| \\
&= \left|h^\alpha c_{j^*} \left(K^{(m)}\left(\frac{x-x_{j^*}}{h}\right) - K^{(m)}\left(\frac{\tilde{y}-x_{j^*}}{h}\right)\right)\right| \\
&\leq Lh^\alpha |c_{j^*}| \left|\frac{x-\tilde{y}}{h}\right|^\alpha \\
&\leq L\,|x-y|^\alpha .
\end{aligned}$$

Therefore our subfamily $\mathcal{F}_{\mathbf{c}} = \{f_{\mathbf{c}} \ : \ |c_j| \leq 1\}$ belongs to $\mathcal{F}^\beta$.

Now we can bound the supremum over the whole class of Hölder continous functions by the supremum over the above defined subclass. For an arbitrary sequence of estimators $\{f_n\}_{n=1}^\infty$, for an arbitrary but fixed $\epsilon > 0$ and any constant C we have:

$$\begin{aligned}
\Delta &\stackrel{\mathrm{d}}{=} \sup_{f\in\mathcal{F}^\beta} \mathbf{P}_f \left(\Phi_n^{-1} \|f_n - f\| > C\,(1-\epsilon)\right) \\
&\geq \sup_{f\in\mathcal{F}^\beta} \mathbf{P}_f \left(\Phi_n^{-1} \max_{j=1,\ldots,J} |f_n(x_j) - f(x_j)| > C\,(1-\epsilon)\right) \\
&\geq \sup_{f\in\mathcal{F}_{\mathbf{c}}} \mathbf{P}_{f_{\mathbf{c}}} \left(\Phi_n^{-1} \max_{j=1,\ldots,J} |f_n(x_j) - f(x_j)| > C\,(1-\epsilon)\right) \\
&= \sup_{\mathbf{c}\in[-1,1]^J} \mathbf{P}_{f_{\mathbf{c}}} \left(\max_{j=1,\ldots,J} |\tau_j - c_j| > C\,(1-\epsilon)\,h_o^{-\beta}\right)
\end{aligned}$$

where we set $\tau_j = f_n(x_j)h^{-\beta}$. If we now choose $C = C_{\mathrm{low}} = h_o^\beta$ then we have for Δ:

$$\Delta \geq \sup_{\mathbf{c}\in[-1,1]^J} \mathbf{P}_{f_{\mathbf{c}}} \left(\max_{j=1,\ldots,J} |\tau_j - c_j| > (1-\epsilon)\right). \tag{5.1}$$

Let A_j be the set defined as follows:

$$A_j = \{i \ | \ |\tfrac{i}{n} - x_j| \leq h\}.$$

It can be seen that for $i \in A_j$ the sample can be rewritten as follows:

$$y_i = f_{\mathbf{c}}\left(\tfrac{i}{n}\right) + \xi_i = c_j h^\beta K\left(\frac{\frac{i}{n}-x_j}{h}\right) + \xi_i \stackrel{\mathrm{d}}{=} c_j\alpha_{ij} + \xi_i,$$

thus for each i the observation y_i bears information about just one of the coefficients c_j where $j = j(i)$. The joint density of the sample is therefore the following:

$$p\,(\mathbf{y},\mathbf{c}) = \prod_{i=1}^{n} \left(\sqrt{2\pi}\sigma\right)^{-1} \exp\left\{\frac{-\,(y_i - c_j\alpha_{ij})^2}{2\sigma^2}\right\} \stackrel{\mathrm{d}}{=} \prod_{i=1}^{n} \varphi_{\sigma^2}(y_i - c_j\alpha_{ij})$$

where $\mathbf{y} = (y_1, \ldots, y_n)$ and $\mathbf{c} = (c_1, \ldots, c_J)$.

By straightforward calculations one can prove that the statistics $\mathbf{T} = (T_1, \ldots, T_J)$ where:

$$T_j = \frac{\sum\limits_{i \in A_j} \alpha_{ij} y_i}{\sum\limits_{i \in A_j} \alpha_{ij}^2}$$

are independent, normally distributed with expectations c_j and variances $\sigma^2 / \sum_{i \in A_j} \alpha_{ij}^2$. As $(nh)^{-1} \sum_{i \in A_j} K^2 \left(\frac{\frac{i}{n} - x_j}{h} \right) = \int K^2(u)\, du\, (1 + o(1)) \stackrel{d}{=} \mu\, (1 + o(1))$ it follows immediately that the Fisher Information I_j of T_j,is independent of j and equals

$$\frac{\mu \log n\, h_o^{2\beta+1}}{\sigma^2} (1 + o(1)).$$

Choosing now $h_o = \left(\frac{2\sigma^2}{(2\beta+1)\mu} \right)^{\frac{1}{2\beta+1}}$ we have for each j:

$$I_j = I = \frac{2 \log n}{2\beta + 1} (1 + o(1)), \tag{5.2}$$

The statistics T_j are sufficient for the parameter $\mathbf{c}$ of the family $p(\mathbf{y}, \mathbf{c})$. Moreover, T_j is sufficient for c_j $(j = 1, \ldots, J)$, i.e. $p(\mathbf{y}, \mathbf{c})$ can be written as $g(\mathbf{y}) \prod_j p_j(T_j, c_j)$.

Using (5.1) we can continue by bounding the minimax risk by the Bayes risk. The Bayes risk is calculated w.r.t the uniform prior on $[-1, 1]^J$:

$$\begin{aligned}
\Delta &\geq \sup_{f\mathbf{c}} \left(1 - \mathbf{P}_{f\mathbf{c}} \left\{ \max_{j=1,\ldots,J} |\tau_j - c_j| \leq 1 - \epsilon \right\} \right) \\
&\geq 2^{-J} \int\limits_{[-1,1]^J} \left(1 - \mathbf{P}_{f\mathbf{c}} \left\{ \max_{j=1,\ldots,J} |\tau_j - c_j| \leq 1 - \epsilon \right\} \right) d\mathbf{c} \\
&= 1 - 2^{-J} \int\limits_{[-1,1]^J} \mathbf{P}_{f\mathbf{c}} \left\{ \max_{j=1,\ldots,J} |\tau_j - c_j| \leq 1 - \epsilon \right\} d\mathbf{c} \\
&\geq 1 - \max_{\tau(\cdot)} 2^{-J} \int\limits_{[-1,1]^J} \mathbf{P}_{f\mathbf{c}} \left\{ \max_{j=1,\ldots,J} |\tau_j - c_j| \leq 1 - \epsilon \right\} d\mathbf{c} \\
&= 1 - \max_{\tau(\cdot)} 2^{-J} \int\limits_{[-1,1]^J} \mathbf{E}_{f\mathbf{c}} \prod_{j=1}^{J} \mathbf{I}\left(|\tau_j - c_j| \leq 1 - \epsilon \right) d\mathbf{c} \\
&= 1 - \max_{\tau_j(\cdot)} \prod_{j=1}^{J} \int\limits_{-\infty}^{\infty} \int\limits_{-1}^{1} \mathbf{I}\left(|\tau_j - c_J| \leq 1 - \epsilon \right) \varphi_{I^{-1}}\left(T_j - c_j \right) dc_j\, dT_j \\
&= 1 - \max_{\tau_j = \tau_j(T_j)} 2^{-J} \prod_{j=1}^{J} \int\limits_{[-1,1]^J} \mathbf{I}\left(|\tau_j(T_j) - c_j| \leq 1 - \epsilon \right) \varphi_{I^{-1}}(T_j - \mathbf{c}_j)\, dT_j\, dc_j, \\
&= 1 - \left(\max_{\tau(T)} 2^{-1} \int\limits_{-\infty}^{\infty} dT \left(\int\limits_{-1}^{1} \mathbf{I}\left(|\tau - c| \leq 1 - \epsilon \right) \varphi_{I^{-1}}(T - c)\, dc \right) \right)^J
\end{aligned}$$

It is not difficult to realise that the function τ^*, defined by:

$$\tau^* = \begin{cases} -\epsilon & T \le -\epsilon \\ T & |T| \le \epsilon \\ \epsilon & T \ge \epsilon \end{cases},$$

is the Bayes estimator of c_j. Therefore we have derived the following inequality for Δ:

$$\Delta \ \ge \ 1 - \left(2^{-1} \int_{-\infty}^{\infty} dT \int_{-1}^{1} \mathbf{I}\left(|\tau^* - c| \le 1-\epsilon\right) \varphi_{I^{-1}}(T-c)\, dc \right)^J .$$

(cf. Korostelev [8]).

Now we continue bounding Δ from below as follows:

$$\begin{aligned}
\Delta \ &\ge \ 1 - \left(1 - 2^{-1} \int dT \int_{-1}^{1} \mathbf{I}\left(|\tau^* - c| > 1-\epsilon\right) \varphi_{I^{-1}}(T-c)\, dc \right)^J \\
&\ge \ 1 - \left(1 - 2^{-1} \int_{-\frac{\epsilon}{2}}^{\frac{\epsilon}{2}} dT \int_{1-\epsilon \le |T-c| \le 1-\frac{\epsilon}{2}} \varphi_{I^{-1}}(T-c)\, dc \right)^J \\
&= \ 1 - \left(1 - \epsilon \int_{1-\epsilon}^{1-\frac{\epsilon}{2}} \varphi_{I^{-1}}(\xi)\, d\xi \right)^J . \qquad (5.3)
\end{aligned}$$

Using the inequality:

$$\frac{1}{\sqrt{2\pi}\,(x+1)} \exp\left\{-\tfrac{1}{2}x^2\right\} \le (2\pi)^{-\frac{1}{2}} \int_x^{\infty} \exp\left\{-\tfrac{1}{2}y^2\right\} dy \le \frac{1}{\sqrt{2\pi}x} \exp\left\{-\tfrac{1}{2}x^2\right\} \qquad (5.4)$$

for $x > 0$ (see Feller [4]), we can deduce for the integral in expression (5.3):

$$\begin{aligned}
\int_{1-\epsilon}^{1-\frac{\epsilon}{2}} \varphi_{I^{-1}}(\xi)\, d\xi \ &\ge \ (2\pi)^{-\frac{1}{2}} \left(\frac{\exp\left\{-\frac{1}{2}I(1-\epsilon)^2\right\}}{\sqrt{I}\,(1-\epsilon)+1} - \frac{\exp\left\{-\frac{1}{2}I(1-\epsilon/2)^2\right\}}{\sqrt{I}\,(1-\epsilon/2)} \right) \\
&= \ (2\pi)^{-\frac{1}{2}} \left(\frac{\exp\left\{-\frac{\log n}{2\beta+1}(1-\epsilon)^2\right\}}{\sqrt{\frac{\log n}{2\beta+1}}\,(1-\epsilon)+1} - \frac{\exp\left\{-\frac{\log n}{2\beta+1}(1-\epsilon/2)^2\right\}}{\sqrt{\frac{\log n}{2\beta+1}}\,(1-\epsilon/2)} \right)(1+o(1)) \\
&= \ (2\pi)^{-\frac{1}{2}} \left(\frac{n^{\frac{-(1-\epsilon)^2}{2\beta+1}}}{\sqrt{\frac{\log n}{2\beta+1}}\,(1-\epsilon)+1} - \frac{n^{\frac{-(1-\epsilon/2)^2}{2\beta+1}}}{\sqrt{\frac{\log n}{2\beta+1}}\,(1-\epsilon/2)} \right)(1+o(1)) \\
&\ge \ \tfrac{1}{2}\, n^{-\frac{(1-\epsilon)^2}{2\beta+1}} (\log n)^{-\frac{1}{2}} \\
&\ge \ \tfrac{1}{2}\, n^{-\frac{(1-\epsilon)}{2\beta+1}} (\log n)^{-\frac{1}{2}}
\end{aligned}$$

for $n > n_o(\epsilon)$. By definition we have for J:

$$J = \lfloor \tfrac{1}{2hb} \rfloor \sim \text{const} \left(\frac{\log n}{n} \right)^{-\frac{1}{2\beta+1}} < n^{\frac{1}{2\beta+1}} . \qquad (5.5)$$

Therefore we can conclude the proof by the following:

$$
\begin{aligned}
\Delta &\geq 1-\left(1-\frac{\epsilon}{2} n^{-\frac{(1-\epsilon)}{2\beta+1}}(\log n)^{-\frac{1}{2}}\right)^{J} \\
&\geq 1-\left(1-\frac{\epsilon}{2} n^{-\frac{(1-\epsilon)}{2\beta+1}}(\log n)^{-\frac{1}{2}}\right)^{n^{\frac{1}{2\beta+1}}} \\
&\geq 1-\exp\left\{-\frac{\epsilon}{2} n^{-\frac{(1-\epsilon)}{2\beta+1}}(\log n)^{-\frac{1}{2}} \cdot n^{\frac{1}{2\beta+1}}\right\} \\
&= 1-\exp\left\{-\frac{\epsilon}{2} n^{\frac{\epsilon}{2\beta+1}}(\log n)^{-\frac{1}{2}}\right\} \\
&\rightarrow 1 \qquad\qquad \text{as } n \rightarrow \infty .
\end{aligned}
$$

Remark that the constant C_{low}, as was mentioned in definition 2.1(a), equals $\left(\frac{2\sigma^2}{\mu(2\beta+1)}\right)^{\frac{1}{2\beta+1}}$.

5.2 Proof of Corollary 2.3

Proof. For $Y_n = \Phi_n^{-1}\|f_n - f\|$ and f_n an arbitrary sequence of estimators $\{f_n\}_{n\geq 1}$ of f we have:

$$
\begin{aligned}
R(f_n, f) &= \mathbf{E}_f Y_n^{\alpha} \\
&= \int_0^{\infty} \alpha y^{\alpha-1} \mathbf{P}_f(Y_n > y)\, dy \\
&\geq \int_0^{C_{\text{low}}} \alpha y^{\alpha-1} \mathbf{P}_f(Y_n > y)\, dy \\
&\geq \mathbf{P}_f\left(Y_n > C_{\text{low}}\right)\left(C_{\text{low}}\right)^{\alpha}.
\end{aligned}
$$

This holds uniformly over f in $\mathcal{F}^{\beta}$, which immediately gives the corollary (by applying Theorem 2.2). $\square$

5.3 proof of Theorem 4.1

The quantity $\sup_{x\in[0,1]}|f_n^{\mathrm{w}}(x) - f(x)|$ will be bounded from above by bounding seperately $Z_n(x) \stackrel{\mathrm{d}}{=} f_n^{\mathrm{w}}(x) - \mathbf{E}f_n^{\mathrm{w}}(x)$, a variance term, and $b_n(x) \stackrel{\mathrm{d}}{=} \mathbf{E}f_n^{\mathrm{w}}(x) - f(x)$, a bias term which is deterministic. Furthermore, as we defined our estimator differently on the neighbourhoods of the edges, we split the interval as follows: $[0,1] = [0,h] \cup [h, 1-h] \cup [1-h, 1]$. Precisely for any C:

$$
\begin{aligned}
\mathbf{P}\left\{\sup_{x\in[0,1]}|f_n^{\mathrm{w}}(x) - f(x)| > C\,\Phi_n\right\} \leq\ & \mathbf{P}\left\{\sup_{x\in[0,h]}|Z_n(x)| > C\,\Phi_n - \sup_{x\in[0,h]}|b_n(x)|\right\} \\
&+ \mathbf{P}\left\{\sup_{x\in[h,1-h]}|Z_n(x)| > C\,\Phi_n - \sup_{x\in[h,1-h]}|b_n(x)|\right\} \\
&+ \mathbf{P}\left\{\sup_{x\in[1-h,1]}|Z_n(x)| > C\,\Phi_n - \sup_{x\in[1-h,1]}|b_n(x)|\right\}.
\end{aligned}
$$

First we deal with the bias term and after that we complete the proof by deriving an upper bound for the variance term.

Lemma 5.1 *Under the conditions of Theorem 4.1 there exists a constant* $C_{\text{bias}} = C_{\text{bias}}(\beta, L, N)$ *such that uniformly over* $\mathcal{F}^\beta$*:*

$$\sup_{x\in[0,1]} |b_n(x)| \le C_{\text{bias}} \left(\frac{\log n}{n}\right)^{\frac{\beta}{2\beta+1}}.$$

Proof. Again, as just above we split the interval and first we examine the bias on the interval $[h, 1-h]$. Remember that f_n^{w} is an estimator of the projection $P_j\overline{f}$ of $\overline{f}$ on V_j (see SECTION 2). Recall that Proposition 3.2 states that for Hölder spaces $\mathcal{F}^\beta$ the difference $P_j\overline{f} - \overline{f}$ tends to zero at the rate $2^{-j(s\wedge\beta)}$ where s is the smoothness of the choosen scaling function φ. As $2^{-j\beta}$ equals $\Phi_n(2N-1)^{-\beta}$ the difference has precisely the rate we want it to have. The question remains what is the rate of the supnorm of the difference between the expected value of the estimator and the projection. Note that in our case $\hat{c}_{j,k}$ is not an unbaised estimator for the coefficient $c_{j,k}$ (because $\mathbf{E}\hat{c}_{j,k}$ is a sum and $c_{j,k}$ is an integral, contrary to the situation in the density estimation, cf. [11]). It will turn out that we'll have to know how good the sum $\frac{1}{n}\sum_{i=1}^n \varphi_{j,k}\left(\frac{i}{n}\right) f\left(\frac{i}{n}\right)$ approximates $c_{j,k} = \int_0^1 \varphi_{j,k} f(x)\,dx$.

After these heuristics, we proceed with the following calculations:

$$\begin{aligned}\sup_{x\in[h,1-h]} |\mathbf{E}f_n^{\text{w}}(x) - f(x)| &\le \sup_{x\in[h,1-h]} \left|\mathbf{E}f_n^{\text{w}}(x) - P_j\overline{f}(x)\right| + \sum_{j'\ge j} \|D_{j'}\overline{f}\| \\ &\le \sup_{x\in[h,1-h]} \left|\mathbf{E}f_n^{\text{w}}(x) - P_j\overline{f}(x)\right| + \frac{2^\beta C_{\text{p}}}{2^\beta - 1} 2^{-j\beta}\end{aligned}$$

where C_{p} is the constant which arised in Proposition 3.2. Therefore it remains to examine $\|\mathbf{E}f_n^{\text{w}} - P_j\overline{f}\|$. This term equals:

$$\sup_{x\in[h,1-h]} \left| \sum_{k=\lfloor 2^j x - 2N+1\rfloor}^{\lfloor 2^j x\rfloor} \varphi_{j,k}(x) \cdot \left(\frac{1}{n}\sum_{i=1}^n \varphi_{j,k}\left(\tfrac{i}{n}\right) f\left(\tfrac{i}{n}\right) - \int_0^1 \varphi_{j,k}(x) f(x)\,dx \right) \right|$$

which is smaller than:

$$\sup_{x\in[h,1-h]} \sum_{k=\lfloor 2^j x - 2N+1\rfloor}^{\lfloor 2^j x\rfloor} |\varphi_{j,k}(x)| \left| \frac{1}{n}\sum_{i=1}^n \varphi_{j,k}\left(\tfrac{i}{n}\right) f\left(\tfrac{i}{n}\right) - \int_0^1 \varphi_{j,k}(x) f(x)\,dx \right|. \tag{5.6}$$

Let us denote by Δ the following difference:

$$\frac{1}{n}\sum_{i=1}^n \varphi_{j,k}\left(\tfrac{i}{n}\right) f\left(\tfrac{i}{n}\right) - \int_0^1 \varphi_{j,k}(x) f(x)\,dx.$$

Recall that $j = j(n)$ was defined implicitely through (4.4). For Δ we can derive the following appropriate upper bound:

Lemma 5.2 *For* $k = \{0, 1, \ldots, 2^j - 2N + 1\}$ *and* $\beta > \frac{1}{2}$ *:*

$$|\Delta| \le 2^{\frac{j}{2}} n^{-1} \left(\tfrac{1}{2}(2N-1)\|\varphi'\|\,\|f\| + o(1)\right) \qquad n \to \infty.$$

Proof. The proof only involves straightforward calculations with a Taylor expansion:

$$
\begin{aligned}
|\Delta| &= \Big| \int 2^{\frac{j}{2}} \varphi\left(2^j x - k\right) f(x)\, dx - n^{-1} \sum_{i=1}^{n} 2^{\frac{j}{2}} \varphi\left(2^j \tfrac{i}{n} - k\right) f\left(\tfrac{i}{n}\right) \Big| \\
&= \Big| \sum_{i=\lfloor 2^{-j} kn \rfloor}^{\lfloor 2^{-j} n(k+2N-1)\rfloor} \int_{\frac{i-1}{n}}^{\frac{i}{n}} 2^{\frac{j}{2}} \left(\varphi\left(2^j x - k\right) f(x) - \varphi\left(2^j \tfrac{i}{n} - k\right) f\left(\tfrac{i}{n}\right)\right) dx \Big| \\
&= \Big| \sum_{i=\lfloor 2^{-j} kn \rfloor}^{\lfloor 2^{-j} n(k+2N-1)\rfloor} \int_{\frac{i-1}{n}}^{\frac{i}{n}} 2^{\frac{j}{2}} \left(\varphi\left(2^j x - k\right) \left(f(x) - f\left(\tfrac{i}{n}\right)\right) + 2^j \varphi'\left(2^j \xi_i - k\right) f(x) \left(x - \tfrac{i}{n}\right)\right) dx
\end{aligned}
$$

where $\xi_i \in]\frac{i-1}{n}, \frac{i}{n}[$.

Let $\sigma = \min(1, \beta)$ and $C_\sigma = \begin{cases} L \text{ the Hölder constant} & \text{for} \quad \sigma = \beta \\ \|f'\| & \text{for} \quad \sigma = 1 \end{cases}$ then:

$$
\begin{aligned}
|\Delta| &\leq \sum_{i=\lfloor 2^{-j} kn \rfloor}^{\lfloor 2^{-j} n(k+2N-1)\rfloor} \left(2^{\frac{j}{2}} \|\varphi\| \int_{\frac{i-1}{n}}^{\frac{i}{n}} C_\sigma |x - \tfrac{i}{n}|^\sigma \, dx + 2^{3\frac{j}{2}} \|\varphi'\|_\infty \|f\| \int_{\frac{i-1}{n}}^{\frac{i}{n}} |x - \tfrac{i}{n}| dx \right) \\
&= (2N-1)\, 2^{-j} n \left(2^{\frac{j}{2}} \|\varphi\| \frac{C_\sigma}{\sigma + 1} n^{-(\sigma+1)} + 2^{3\frac{j}{2}} \|\varphi'\| \|f\| \tfrac{1}{2} n^{-2} \right) \\
&= 2^{\frac{j}{2}} n^{-1} \left(\tfrac{1}{2} (2N-1) \|\varphi'\|_\infty \|f\| + o(1) \right)
\end{aligned}
$$

□

Remark that due to the last equation we have to require that β exceeds $\frac{1}{2}$. Using (5.6) and Lemma 5.2 we get:

$$
\sup_{x \in [h, 1-h]} \left| \mathbf{E} f_n^{\mathrm{w}}(x) - P_j \overline{f}(x) \right| \leq (2N-1) \|\varphi\|\, 2^{\frac{j}{2}} n^{-1} \left(\tfrac{2N-1}{2} \|\varphi'\|\, \|f\| + o(1) \right) \qquad n \to \infty
$$

which is negligable compared to $\sup_{x \in [h, 1-h]} \left| P_j \overline{f}(x) - \overline{f}(x) \right|$.

It remains to evaluate the risk at the neighbourhoods of the edges. Due to symmetry it is sufficient to look at the case $x \in [0, h]$. We need the following Lemma:

Lemma 5.3 $\displaystyle\sup_{x,y \in [0,h]} \left| (P_j \overline{f})^{(m)}(x) - (P_j \overline{f})^{(m)}(y) \right| < L_P h^\alpha$

where the constant L_P *equals* $L + \frac{2 C_2 (2N-1)^{-\alpha}}{1 - 2^{-\alpha}}$.

Proof. Using equation (3.1) we have:

$$
\begin{aligned}
|(P_j \overline{f})^{(m)}(x) - (P_j \overline{f})^{(m)}(y)| &= |(\overline{f})^{(m)}(x) - \sum_{j' \geq j} (D_{j'} \overline{f})^{(m)}(x) + \sum_{j' \geq j} (D_{j'} \overline{f})^{(m)}(y) - (\overline{f})^{(m)}(y)| \\
&\leq |(\overline{f})^{(m)}(x) - (\overline{f})^{(m)}(y)| + |\sum_{j' \geq j} (D_{j'} \overline{f})^{(m)}(y) - \sum_{j' \geq j} (D_{j'} \overline{f})^{(m)}(x)|.
\end{aligned}
$$

Remark first that due to Proposition 3.3 and 3.2 we are allowed to differentiate each term in the summation above. Second, remark that $\overline{f}^{(m)}$ is Hölder continuous. If we again apply Proposition 3.3 and Proposition 3.2 we obtain:

$$\begin{aligned}
\sup_{x,y\in[0,h]} |(P_j\overline{f})^{(m)}(x) - (P_j\overline{f})^{(m)}(y)| &\leq Lh^\alpha + 2C_2 \sum_{j'\geq j} 2^{j'm}\|D_{j'}\overline{f}\| \\
&\leq Lh^\alpha + 2C_2 \sum_{j'\geq j} 2^{j'\alpha} \\
&\leq Lh^\alpha + \frac{2^{-j\alpha+1}C_2}{1-2^{-\alpha}} \\
&= \left(L + \frac{2\,C_2(2N-1)^{-\alpha}}{1-2^{-\alpha}}\right) h^{\alpha}
\end{aligned}$$

which concludes the proof of Lemma 5.3. □

Due to this Lemma we have:

$$\begin{aligned}
\sup_{x\in[0,h]} \left|\mathbf{E}f_n^{\mathrm{w}}(x) - P_j\overline{f}(x)\right| &= \sup_{x\in[0,h]} \left|\sum_{l=0}^{m} \left(\mathbf{E}(f_n^{\mathrm{w}})^{(l)}(h) - (P_j\overline{f})^{(l)}(h)\right) \frac{(x-h)^l}{l\,!}\right. \\
&+ \left.\left((P_j\overline{f})^{(m)}(\xi) - (P_j\overline{f})^{(m)}(x)\right) \frac{(x-h)^m}{m\,!}\right| \\
&\leq \sum_{l=0}^{m} \left|\mathbf{E}(f_n^{\mathrm{w}})^{(l)}(h) - (P_j\overline{f})^{(l)}(h)\right| \frac{h^l}{l\,!} + \frac{h^\beta}{m\,!}L_P \\
&= \sum_{l=0}^{m} \frac{h^l}{l\,!}\left|\sum_{k=0}^{2N-1} 2^{j(l+\frac{1}{2})}\varphi^{(l)}(2^j h-k)\left(\mathbf{E}\hat{c}_{j,k} - c_{j,k}\right)\right| + \frac{h^\beta}{m\,!}L_P \\
&\leq \sum_{l=0}^{m}(2N-1)2^{j(l+1)}n^{-1}\frac{h^l}{l\,!}\|\varphi^{(l)}\|\left(\tfrac{2N-1}{2}\|\varphi'\|\|f\| + o(1)\right) + \frac{h^\beta}{m\,!}L_P \\
&= \sum_{l=0}^{m} \tfrac{(2N-1)^{-l}}{l\,!\,hn}\|\varphi^{(l)}\|\left(\tfrac{1}{2}(2N-1)\|\varphi'\|\,\|f\| + o(1)\right) + \frac{h^\beta}{m\,!}L_P \\
&= \frac{h^\beta}{m\,!}L_P(1+o(1)) \quad \text{as} \quad n\to\infty.
\end{aligned}$$

This enables us to conclude for $x \in [0,h]$ as follows:

$$\begin{aligned}
\sup_{x\in[0,h]} |\mathbf{E}f_n^{\mathrm{w}}(x) - f(x)| &\leq \sup_{x\in[0,h]} \left|\mathbf{E}f_n^{\mathrm{w}}(x) - (P_j\overline{f})(x)\right| + \sup_{x\in[0,h]} \left|(P_j\overline{f})(x) - f(x)\right| \\
&\leq \frac{h^\beta}{m\,!}L_P(1+o(1)) + \frac{C_{\mathrm{p}}2^\beta}{2^\beta-1}2^{-j\beta} \\
&= \Phi_n\left(\frac{L_P}{m\,!}(1+o(1)) + \frac{C_{\mathrm{p}}2^\beta}{2^\beta-1}(2N-1)^{-\beta}\right).
\end{aligned}$$

Finally one obtains for the bias:

$$\sup_{x\in[0,1]} |b_n(x)| \leq \Phi_n\left(\frac{2L_P}{m\,!}(1+o(1)) + \frac{3\,C_{\mathrm{p}}2^\beta}{2^\beta-1}(2N-1)^{-\beta}\right) \tag{5.7}$$

where C_{p} was the constant which arise in Proposition 3.2, Thus

$$C_{\mathrm{bias}} = \frac{2L_P}{m!} + \frac{3\,C_{\mathrm{p}} 2^{\beta}}{(2^{\beta}-1)(2N-1)^{\beta}} + o(1) \tag{5.8}$$

and this completes the proof of Lemma 5.3. □

To proceed further with the proof of Theorem 4.1 we need an upper bound for the large deviations of a Gaussian process in supnorm. We will state and prove a more concise upper bound which will be found useful in the proof of Theorem 4.2:

Lemma 5.4 *With* $h = \left(\frac{\log n}{n}\right)^{\frac{1}{2\beta+1}}$ *and* $\Phi_n = h^{\beta}$ *we have:*

$$\begin{aligned}
(a)\qquad \Delta_1(u) &\stackrel{\mathrm{d}}{=} \mathbf{P}\left\{\sup_{x\in[h,1-h]} |Z_n(x)| > u\,\Phi_n\right\} \\
&\leq \left(\frac{\log n}{n}\right)^{-\frac{1}{2\beta+1}} \frac{(2(2N-1))^{\frac{5}{2}}\|\varphi\|\,\sigma}{u\sqrt{\pi}\,\log n} \exp\left\{-\frac{u^2\log n}{8\sigma^2\|\varphi\|^2(2N-1)^3}\right\}\cdot\left(1+O\left(2^j n^{-1}\right)\right)
\end{aligned}$$

$$\begin{aligned}
(b)\qquad \Delta_2(u) &\stackrel{\mathrm{d}}{=} \mathbf{P}\left\{\sup_{x\in[0,h]} |Z_n(x)| > u\,\Phi_n\right\} \\
&\leq \sum_{l=0}^{m}\sqrt{\frac{2}{\pi}}\frac{(2N-1)^{\frac{2i+3}{2}} m\sigma\|\varphi^{(l)}\|\|\varphi\|}{l!\,u\sqrt{\log n}} \exp\left\{-\frac{(l!)^2u^2\log n}{m^2\sigma^2\|\varphi^{(l)}\|^2(2N-1)^{(2l+3)}}\right\}
\end{aligned}$$

and a similar bound holds for $x \in [1-h, 1]$.

Proof. Let $A_j = \{1, \ldots, \lfloor h^{-1}\rfloor - 1\}$ and $\Delta_p = [\,ph, \min((p+1)h, 1-h)]$ for $p \in A_j$. Then we have:

$$\begin{aligned}
\Delta_1(u) &\leq \sum_{p\in A_j} \mathbf{P}\left\{\sup_{x\in\Delta_p} |f_n^{\mathrm{w}}(x) - \mathbf{E}f_n^{\mathrm{w}}(x)| > u\,\Phi_n\right\} \\
&= \sum_{p\in A_j} \mathbf{P}\left\{\sup_{x\in\Delta_p}\left|\sum_{k=\lfloor 2^j x-2N+1\rfloor}^{\lfloor 2^j x\rfloor} \varphi_{j,k}(x)\tfrac{1}{n}\sum_{i=1}^{n}\varphi_{j,k}\left(\tfrac{i}{n}\right)\xi_i\right| > u\,\Phi_n\right\} \\
&\leq \sum_{p\in A_j} \mathbf{P}\left\{\sum_{k=(p-1)(2N-1)}^{(p+1)(2N-1)} \tfrac{1}{n}\left|\sum_{i=1}^{n}\varphi_{j,k}\left(\tfrac{i}{n}\right)\xi_i\right| > u\,\Phi_n 2^{-\frac{j}{2}}\|\varphi\|^{-1}\right\} \\
&\leq \sum_{p\in A_j}\sum_{k=(p-1)(2N-1)}^{(p+1)(2N-1)} \mathbf{P}\left\{\tfrac{1}{n}\left|\sum_{i=1}^{n}\varphi_{j,k}\left(\tfrac{i}{n}\right)\xi_i\right| > \frac{u\,\Phi_n 2^{-\frac{j}{2}}}{2\,\|\varphi\|(2N-1)}\right\}.
\end{aligned} \tag{5.9}$$

In our model the ξ_i's are assumed to be normally distributed. Therefore the variables $\nu_{n,k} = \frac{1}{n}\sum_{i=1}^{n}\varphi_{j,k}\left(\frac{i}{n}\right)\xi_i$ are normally distributed with mean zero and variance

$$\frac{\sigma^2}{n^2}\sum_{i=1}^{n}\varphi_{j,k}^2\left(\tfrac{i}{n}\right) = \frac{\sigma^2}{n}\left(1 + O\left(\frac{2^j}{n}\right)\right) \qquad n\to\infty. \tag{5.10}$$

The last equality can be proved by applying similar kind of calculations as in the proof of Lemma 5.2, based on the fact that $\int \varphi^2 = 1$. By using (5.10) and the following inequality:

$$\int_x^\infty (2\pi)^{-\frac{1}{2}} \exp\left\{-\tfrac{1}{2}y^2\right\} dy \le \frac{1}{x\sqrt{2\pi}} \exp\left\{-\tfrac{1}{2}x^2\right\} \tag{5.11}$$

(cf. Feller ([4], p. 175), for $x = \frac{u\,\Phi_n 2^{-\frac{j}{2}}}{\|\varphi\|\, 2\,(2N-1)}$ we obtain the appropriate upperbound for the tail probability of $\nu_{n,k}$, $\mathbf{P}\left\{|\nu_{n,k}| > \frac{u\,\Phi_n 2^{-\frac{j}{2}}}{\|\varphi\|\, 2\,(2N-1)}\right\}$:

$$\frac{(2\,(2N-1))^{\frac{3}{2}}\,\|\varphi\sigma\|}{u\sqrt{\pi}\,\log n} \exp\left\{-\frac{u^2 \log n}{8\,\sigma^2\|\varphi\|^2(2N-1)^3}\right\}\left(1 + O\left(2^j n^{-1}\right)\right) \tag{5.12}$$

Finally we complete the proof of (a) by substituting (5.12) in (5.9). Furthermore choosing $u^2 \ge \frac{8\sigma^2\|\varphi\|^2(2N-1)^3}{2\beta+1}$ one finds that $\Delta_1(u) \to 0$, as $n \to \infty$.

Turning to (b) we have:

$$\begin{aligned}\Delta_2(u) &\le \mathbf{P}\left\{\sum_{l=0}^m \left|(f_n^{\mathrm{w}})^{(l)}(h) - \mathbf{E}(f_n^{\mathrm{w}})^{(l)}(h)\right| \frac{h^l}{l\,!} > u\,\Phi_n\right\} \\ &\le \sum_{l=0}^m \mathbf{P}\left\{\left|(f_n^{\mathrm{w}})^{(l)}(h) - \mathbf{E}(f_n^{\mathrm{w}})^{(l)}(h)\right| > \frac{u\,\Phi_n l\,!}{m h^l}\right\}.\end{aligned}$$

Here $Z_n^{(l)}(h) \stackrel{\mathrm{d}}{=} (f_n^{\mathrm{w}})^{(l)}(h) - \mathbf{E}(f_n^{\mathrm{w}})^{(l)}(h)$ is normally distributed with variance $\sigma_l^2(h)$, where:

$$\sigma_l^2(h) \le (2N-1)^{2l+3} h^{-(2l+1)} n^{-1} \|\varphi^{(l)}\|^2 \|\varphi\|^2 \sigma^2. \tag{5.13}$$

Again by using the inequality (5.11) one finds for any $u > 0$:

$$\begin{aligned}\Delta_2(u) &\le \sum_{l=0}^m \mathbf{P}\left\{\frac{|Z_n^{(l)}(h)|}{\sigma_l(h)} > \frac{u\,\Phi_n l\,!}{m\sigma_l(h)\,h^l}\right\} \\ &\le \sum_{l=0}^m \sqrt{\frac{2}{\pi}} \frac{m\,\sigma_l(h)\,h^l}{l\,!\;u\,\Phi_n} \exp\left\{-\frac{\Phi_n^2(l\,!\,u)^2}{2\sigma_l^2(h)m^2}\right\} \\ &\le \sum_{l=0}^m \sqrt{\frac{2}{\pi}} \frac{(2N-1)^{\frac{2l+3}{2}}\|\varphi^{(l)}\|\|\varphi\| m\,\sigma}{u\,l\,!\sqrt{\log n}} \exp\left\{-\frac{(l\,!)^2 u^2 \log n}{m^2\sigma^2\|\varphi^{(l)}\|^2(2N-1)^{(2l+3)}}\right\}.\end{aligned}$$

Therefore $\Delta_2(u) = o(1)(n \to \infty)$, for any $u > 0$. □

To complete the proof of Theorem 4.1 we apply Lemma 5.4. Choose therefore $C_{\mathrm{upp}}^{\mathrm{w}} = C_{\mathrm{bias}} + u$ (see (5.8)) with $u = \left(\frac{8\sigma^2\|\varphi\|^2(2N-1)^3}{2\beta+1}\right)^{\frac{1}{2}}$ and we can establish:

$$\begin{aligned}\mathbf{P}\left\{\sup_{x\in[0,1]} |f_n^{\mathrm{w}}(x) - f(x)| > C_{\mathrm{upp}}^{\mathrm{w}}\,\Phi_n\right\} &\le \mathbf{P}\left\{\sup_{x\in[0,h]} |Z_n(x)| > (C_{\mathrm{upp}}^{\mathrm{w}} - C_{\mathrm{bias}})\,\Phi_n\right\} \\ &+ \mathbf{P}\left\{\sup_{x\in[h,1-h]} |Z_n(x)| > (C_{\mathrm{upp}}^{\mathrm{w}} - C_{\mathrm{bias}})\,\Phi_n\right\} \\ &+ \mathbf{P}\left\{\sup_{x\in[1-h,1]} |Z_n(x)| > (C_{\mathrm{upp}}^{\mathrm{w}} - C_{\mathrm{bias}})\,\Phi_n\right\} \\ &\to 0 \qquad \text{as } n \to \infty.\end{aligned}$$

5.4 Proof of Theorem 4.2

Proof. First let us denote $\Phi_n^{-1}\|f_n^{\mathrm{w}} - f\|$ by Y_n^{w}. For the risk of f_n^{w} the following holds:

$$R(f_n^{\mathrm{w}}, f) \leq (C_{\mathrm{upp}}^{\mathrm{w}})^{\alpha} + \int_{C_{\mathrm{upp}}^{\mathrm{w}}}^{\infty} \alpha y^{\alpha-1} \mathbf{P}\left(Y_n^{\mathrm{w}} > y\right) dy. \tag{5.14}$$

So it remains to proof that this integral converges to 0, uniformly over our class $\mathcal{F}^{\beta}$ (see (2.2) if n tends to ∞. Remark the fact that:

$$\mathbf{P}(Y_n^{\mathrm{w}} > y) \leq \mathbf{P}(\Phi_n^{-1}\|Z_n\| > y - C_{\mathrm{bias}}). \tag{5.15}$$

Substituting $y = C_{\mathrm{upp}} + v$ for $v > 0$ in Lemma 5.4 we obtain for (5.15):

$$\begin{aligned}
(5.15) \quad \leq \quad & \frac{C_\alpha \left(\frac{\log n}{n}\right)^{-\frac{1}{2\beta+1}}}{(C_{\mathrm{upp}} + v - C_{\mathrm{bias}})\sqrt{\log n}} \exp\left\{-C_\beta (C_{\mathrm{upp}} + v - C_{\mathrm{bias}})^2 \log n\right\} \\
+ \quad & \sum_{i=0}^{m} \frac{C_\gamma(l)}{(C_{\mathrm{upp}} + v - C_{\mathrm{bias}})\sqrt{\log n}} \exp\left\{-C_\delta(l)(C_{\mathrm{upp}} + v - C_{\mathrm{bias}})^2 \log n\right\}
\end{aligned}$$

Therefore it remains to study the following integral:

$$\begin{aligned}
& \int_0^\infty \alpha\,(C_{\mathrm{upp}} + v)^{\alpha-1} \Bigg(\frac{C_\alpha \left(\frac{\log n}{n}\right)^{-\frac{1}{2\beta+1}}}{(C_{\mathrm{upp}} + v - C_{\mathrm{bias}})\sqrt{\log n}} \exp\left\{-C_\beta (C_{\mathrm{upp}} + v - C_{\mathrm{bias}})^2 \log n\right\} \\
& \quad + \sum_{i=0}^{m} \frac{C_\gamma(l)}{(C_{\mathrm{upp}} + v - C_{\mathrm{bias}})\sqrt{\log n}} \exp\left\{-C_\delta(l)(C_{\mathrm{upp}} + v - C_{\mathrm{bias}})^2 \log n\right\} \Bigg)
\end{aligned}$$

(here C_α, C_β, C_γ and C_δ are the constants which appeared in Lemma 5.4) and this integral is smaller than:

$$\begin{aligned}
& e^{-\min(C_\beta, C_\delta(l))\,(C_{\mathrm{upp}} - C_{\mathrm{bias}})^2} \int_0^\infty \alpha\,(C_{\mathrm{upp}} + v)^{\alpha-1} \Bigg(\frac{C_\alpha \left(\frac{\log n}{n}\right)^{-\frac{1}{2\beta+1}} \exp\left\{-C_\beta v^2 \log n\right\}}{(C_{\mathrm{upp}} + v - C_{\mathrm{bias}})\sqrt{\log n}} \\
& \quad + \sum_{i=0}^{m} \frac{C_\gamma(l)}{(C_{\mathrm{upp}} + v - C_{\mathrm{bias}})\sqrt{\log n}} \exp\left\{-C_\delta(l)\, v^2 \log n\right\} \Bigg)
\end{aligned}$$

Our problem reduces now to showing that the following integral converges to 0 (uniformly over the class $\mathcal{F}^{\beta}$), by using the above obtained upper bound for $\mathbf{P}(\Phi_n^{-1}\|Z_n\| > y - C_{\mathrm{bias}})$:

$$\int_{C_{\mathrm{bias}}^{\mathrm{w}}}^{\infty} \alpha y^{\alpha-1} \mathbf{P}(\Phi_n^{-1}\|Z_n\| > y - C_{\mathrm{bias}})\, dy$$

Using the fact that $\int_0^\infty x^{p-1} e^{-x}\, dx < \infty$ if $p > 0$ it can be shown, with straightforward calculations, that the above integral converges to 0, uniformly over the class $\mathcal{F}^{\beta}$. Thus we can conclude that $\lim_{n\to\infty} \sup_{f\in\mathcal{F}^\beta} R(f_n^{\mathrm{w}}, f) \leq (C_{\mathrm{upp}}^{\mathrm{w}})^{\alpha}$ as n tends to ∞. □

References

[1] Daubechies, I. (1992). *Ten Lectures on Wavelets.* CBMS- NSF Regional Conference series in Applied Mathematics. SIAM.

[2] Donoho, D.L. (1992). Asymptotic Minimax Risk (for Sup-norm loss): Solution via Optimal Recovery. *Technical Report.*

[3] Donoho, D.L., Johnstone, I.M., (1992). Ideal spatial adaptation by wavelet shrinkage. *Technical Report.*

[4] Feller, W. (1950). *An Introduction to Probability Theory and Its Applications*, vol.1, Wiley.

[5] Genon-Catalot, V., Laredo, C. and Picard, D. (1992). Non-parametric Estimation of the Diffusion Coefficient by Wavelets Methods. *Scand. J. Stat*, **19**, 317-335.

[6] Ibragimov, I.A. and Hasminskii, R.Z. (1981). *Statistical Estimation: Asymptotic Theory.* Springer-Verlag, New York, Heidelberg, Berlin.

[7] Ibragimov,I.A. and Hasminskii, R.Z. (1982). Bounds for the risks of non-parametric regression estimates. *Theor. of Prob. and its Appl.*, **27-1**, 84-99.

[8] Korostelev, A.P. (1991). Exact asymptotical minimax estimate for a nonparametric regression in the uniform norm. To appear in *Theory of Probability and Its Applications.*

[9] Meyer, Y. (1992). *Wavelets and Operateurs.* vol. 37 Cambridge studies in advanced mathematics. Cambridge University Press.

[10] Müller, H.G. (1991). Smooth optimum kernel estimators near endpoints. *Biometrika.* **78-3**, 521-530.

[11] Picard, D. and Kerkyacharian, G. (1992). Density estimation in besov spaces. *Statistics & Probability letters*, **13**, 15-24.

[12] Stone, C.J. (1982). Optimal global rates of convergence for nonparametric regression. *Annals of Statistics*, **10-4**, 1040-1053.