Numerical Methods for Asymptotically Minimax Non–parametric Function Estimation with Positivity Constraints I

September 17, 1998

Lubomir Dechevsky
Département de mathématiques et de statistique
Université de Montréal

Brenda MacGibbon
Département de mathématiques
Université du Québec à Montréal

Spiridon Penev
Department of Statistics
University of New South Wales

Abstract

One important challenge in nonparametric density and regression-function estimation is spatially inhomogeneous smoothness. This is often modelled by Besov-type smoothness constraints. Relevant to this type of constraints, Donoho and Johnstone (1992), Delyon and Juditsky (1993) studied asymptotic-minimax optimal wavelet estimators with thresholding, while Lepski, Mammen and Spokoiny (1995) proposed a variable-bandwidth selection for kernel estimators that also achieved the asymptotic-minimax rates. However, a second challenge in many applications of nonparametric curve estimation is that the function must be nonnegative. In [DM1] Dechevsky and MacGibbon (1998) constructed wavelet- and kernel-based estimators under positivity constraints that satisfied these constraints and also achieved asymptotic-minimax rates over the appropriate smoothness classes. These estimators involved integration. In a sequence of two papers, of which this is the first one, we show how to replace the integral in their definition by a quadrature formula, so that the new "quadrature" estimators enjoy the positivity- and smoothness-preserving properties of the ones in [DM1], and are also asymptotic-minimax optimal.

0 Introduction

One important challenge in nonparametric density and regression-function estimation is spatially inhomogeneous smoothness. This is often modelled by Besov-type smoothness constraints. Donoho and Johnstone (1992), Delyon and Juditsky (1993) studied wavelet estimators with thresholding, while Lepski, Mammen and Spokoiny (1995) proposed a variable-bandwidth selection for kernel estimators; all these wavelet and kernel estimators achieve asymptotic-minimax optimal rates over Besov classes. However, a second challenge in many applications of nonparametric curve estimation is that the function must also be nonnegative, or, more generally, it must be bounded from below and/or above by known functions v an/or w. In [DM1] and [DM2] Dechevsky and MacGibbon (1998) propose a general method to achieve the following: for an unknown density or regression function f whose graph is a priori known to be bounded by the one(s) of v and/or w - known functions with certain regularity - given an estimator $\hat{f}$ whose graph need not obey the bound(s), construct a *sufficiently smooth* estimator $\hat{f}^+$ whose graph obeys the bound(s) and achieves *the same* asymptotic rates as $\hat{f}$, with respect to *the same* metric. The definition of the new estimator involves a convolution with a smooth nonnegative symmetric kernel with certain bandwidth ε which is being chosen in an optimal way to preserve the rates. In some cases, it is possible to compute the convolution's integral exactly, by a closed formula, but in general one needs to do numerical integration by a quadrature rule (thus replacing the convolution by a kernel-type estimator). The aim of the present paper is to study the properties of the resulting new "quadrature" kernel estimator of f and to compare them to the properties of the underlying $\hat{f}$ and $\hat{f}^+$. In particular: (1) the quadrature formula must be so selected that the new estimator is *as smooth as* $\hat{f}^+$; (2) the quadrature formula depends on *additional small parameters* determining its precision of approximating the integral; these additional parameters must, together with ε, be chosen in an optimal way so as to preserve *the same* asymptotic rates; (3) the choice of these parameters can be made optimal only with respect to certain values of the metric indices p, q and the smoothness index s of the Besov space B^s_{pq} to which f is assumed to belong; in practice, only approximate values or a "confidence" range of p, q, s is known; therefore, optimal (smaller) values of the quadrature parameters and of ε have to be determined in such a way that the "quadrature" estimator attains *the same* rates as $\hat{f}$ and $\hat{f}^+$ *simultaneously on the entire* "confidence" range of p, q and s.

In the present article we consider the quadrature discretization of the convolution under the assumptions and relying on the results of [DM1]. The theory in [DM1] is developed under the restrictive assumption that v and w are constants. The general case when v and/or w are non-constant functions is being addressed in [DM2], which is an extension and development of [DM1]. Parallel to [DM2], we are working on [DMP] where the quadrature version of the estimator from [DM2] is being studied for general functions v and w. The basic approach in [DM2] and [DMP] is essentially the same as in [DM1] and the present paper; however, its realization when v and/or w are not constants is much more technically involved, thus necessitating a separate consideration.

In [DM1] the multivariate case $f : \mathbb{R}^d \to \mathbb{R}$ was considered. In the present paper we address the univariate case $d = 1$ only. It is possible to extend, quite straightforwardly, the present results to the multivariate case, too. However, formulating and proving them in a multivariate setting would be rather spacious. It would be also more difficult for the reader to extract the essentials of our approach from the more technical presentation needed for $d.1$. That is why we have given preference to a "univariate exposition"; the most important technical differences in the multivariate case are being commented upon in appropriate places later in the text.

The organization of this article is as follows. In section 2 we consider a general class of variable-kernel estimators which includes in itself all asymptotically minimax-optimal wavelet and kernel estimators as well as the respective $\hat{f}^+$ and its new quadrature versions. Section 3 contains the main results. In Section 4 some applications are discussed in the context of [DJ] and [LMS]. Section 5 addresses the important practical question of selecting a concrete appropriate quadrature rule. The graphic results of some model numerical simulations are displayed in Section 6. The proofs are in section 7. Some final remarks are given in section 8. Section 1 contains preliminary material only and is designed to be used just as a reference source while reading the rest of the text.

1 Preliminaries

In this section we provide a list of basic notations, definitions and orientation remarks.

$[\tau]$ – integer part of τ (for $\tau > 0$).

$x_+ = \max\{0, x\}$.

$\Omega(\delta_0) = \{\omega : [0, \delta_0] \to [0, \infty)\,;\ \omega(t_1) > \omega(t_2),\ t_1 > t_2;\ \exists \lim_{t\to 0+} \omega(t) = \omega(0+) = 0;\ \ \omega(0) = 0\}$.

Quasi-norm (see [BL], [DP1]): $\exists c \geq 1;\ \|a+b\| \leq c\,(\|a\|+\|b\|)$.

Semi-norm : the norm property $\|a\|=0 \Leftrightarrow a=0$ is not necessarily fulfilled.

A - quasi(semi) normed abelian group - see [BL].

A - quasi-Banach space, if A is a complete quasi-normed abelian group, a linear space, and the quasinorm is homogeneous: $\|\alpha a\|_A = |\alpha|\|a\|_A$.

$A^\rho = \left\{a \in A : \|a\|_{A^\rho} := \|a\|_A^\rho < \infty\right\},\ \rho \in (0,\infty)$.

$\rho A = \{a \in A : \|a\|_{\rho A} := \rho\|a\|_A < \infty\},\ \rho \in (0,\infty)$.

$A \hookrightarrow B$ ($B \hookleftarrow A$): continuous embedding of A in B (for quasinormed abelian groups) - see [BL].

$A \rightleftharpoons B$: $A \hookrightarrow B$ and $B \hookrightarrow A$. In this case $\|\cdot\|_A$ and $\|\cdot\|_B$ are equivalent - denoted by $\|a\|_A \sim \|a\|_B$.

For the relevant properties of Besov spaces we refer to [BL], [T1] and [T2].

The integral modulus of smoothness (on $\mathbb{R}^d$) : for $\delta > 0$,

$\omega_k(f;\delta)_{L_p(\mathbb{R}^d)} = \sup\limits_{|h|\leq\delta} \|\Delta_h^k f\|_{L_p(\mathbb{R}^d)}$, where $k \in \mathbb{N}$, $\quad h \in \mathbb{R}^d$, $\Delta_h^k f(x)$ is the k-th finite difference with step h at x :

$\Delta_h^k f(x) = \sum_{\nu=0}^k \binom{\nu}{k}(-1)^{k+\nu} f(x+\nu h), \quad \Delta_h^k f(x) = \Delta_h^1 \Delta_h^{k-1} f(x)$.

The error measure in which the risk will be measured is the same as in [DM1, [DJ] and [LMS], that is, the quasi-norm of the quasi-normed abelian group $\mathcal{L}_\rho(L_p)^\rho = \mathcal{L}_\rho\left(\mathbb{R}^d, L_p(\mathbb{R}^d)\right)^\rho$, defined in [DM1], Section 0, $0 < p \leq \infty$, $0 < \rho < \infty$. Recall the definition of this quasi-norm:

$$\|h\|_{\mathcal{L}_\rho(L_p)^\rho} = E_{F^N}\left(\|h\|_{L_p}^\rho\right),$$

where $h = h(x)$, $\quad x \in \mathbb{R}^d$, is a trajectory of a stochastic process based on measurable transformation of the random vector $(X_1,\dots,X_N)$ of N (not necessarily uncorelated) random variables from a c.d.f. $F^N : \mathbb{R}^{Nd} \to [0,1]$. In applications $(X_1,\dots,X_N)$ are most often i.i.d. (see, e.g., [DJ], [LMS]).

In our considerations about wavelet estimators based on orthonormal wavelets on $\mathbb{R}^d$ (see 2.11, 13, 15) we use exactly the same notations as in [DJ]. In particular, φ and ψ are the respective scaling function and "mother" wavelet; $\varphi,\psi \in C_0^r(\mathbb{R}^d)$, where $C_0^r(\mathbb{R}^d)$ is comprised, as usual, by all $f \in C^r(\mathbb{R}^d)$ (r-times continuously differentiable) with compact support. Here $r \in \mathbb{N}$ is the upper bound for the smoothness index s of the Besov spaces in consideration (the same as in [DJ]).

The kernels $\Phi(x)$ and $\Phi_\varepsilon(x)$ (sometimes called approximate identity), $x \in \mathbb{R}^d$, $\varepsilon > 0$, are defined in [DM1], Section 1. Notice that $\Phi \in C_0^r(\mathbb{R}^d)$ where r has the above-said meaning.

As usual,

$L_{1,loc}(\mathbb{R}^d) = \{f : \mathbb{R}^d \to \mathbb{R},\ f|_\Omega \in L_1(\Omega)$ for any compact $\Omega \subset \mathbb{R}^d\}$.

For $f \in L_{1,loc}(\mathbb{R}^d)$, the convolution $f_\varepsilon(x) = \Phi_\varepsilon * f(x)$, $0 < \varepsilon < \infty$, is in $C^r(\mathbb{R}^d)$ (and in $L_p(\mathbb{R}^d)$ if $f \in L_p(\mathbb{R}^d)$). f_ε is called the Sobolev $\varepsilon-$mean of f (see [N], [DM1]).

For $f \in L_{1,loc}(\mathrm{I\!R}),\ 0 < \varepsilon < \infty$,

$$f_{r,\varepsilon}(x) = (-\varepsilon)^{-r} \underbrace{\int_0^{\varepsilon} \cdots \int_0^{\varepsilon}}_{r} \left[\sum_{\nu=0}^{r-1} (-1)^{\nu+r} f\left(x + \frac{r-\nu}{r} \sum_{\mu=1}^{r} \theta_\mu \right) \right] d\theta_1 \cdots d\theta_r,$$

is the Steklov (r,ε)-mean of f (see [SP], [PP], [DP1]; for the multivariate case $d > 1$, see [JS]).

For the definition of the *integral modulus* of smoothness of $f \in L_{1,loc}(\mathrm{I\!R})$, of order r, in $L_p(\mathrm{I\!R}^d),\ 1 \leq p \leq \infty$, with step $\varepsilon > 0 : \omega_r(f;\varepsilon)_p = \omega_r(f;\varepsilon)_{L_p(\mathrm{I\!R}^d)}$, and its properties, we refer to [N], [SP], [PP], [DP1], [DP2], [DM1], [JS], [BL], [T1].

For $\varepsilon > 0,\ 0 < p \leq \infty$, consider (see [D], [D1])

$A_{p,\varepsilon}(\mathrm{I\!R}) = \{f : \mathrm{I\!R} \to \mathrm{I\!R},\ \|f\|_{A_{p,\varepsilon}(\mathrm{I\!R})} < \infty\}$,

where

$\|f\|_{A_{p,\varepsilon}(\mathrm{I\!R})} = \|S(\varepsilon, |f|; \cdot)\|_{L_p(\mathrm{I\!R})}$,

$S(\varepsilon, f; x) = \sup\{f(y) : y \in [x-\varepsilon, x+\varepsilon]\}$ (upper Baire's function).

The *local modulus of smoothness* is a generalisation of $S(\frac{\varepsilon}{2}, |f|; x)$ (which coincides with if for $r = 0$):

$\omega_r(f, x; \varepsilon) = \sup\{|\Delta_\theta^r f(y)| : y, y + r\theta \in [x - \frac{r\varepsilon}{2}, x + \frac{r\varepsilon}{2}]\}$, (see [SP], [D1], [PP]).

The *averaged modulus of smoothness* (or τ*-modulus*) of $f : \mathrm{I\!R} \to \mathrm{I\!R}$, of order r, in $L_p(\mathrm{I\!R})$, $0 < p \leq \infty$, with step $\varepsilon > 0$, is defined by

$\tau_r(f;\varepsilon)_p = \tau_r(f;\varepsilon) = \|\omega(f, \cdot; \varepsilon)\|_{L_p(\mathbb{R})}$

(see [SP], [D1], [PP]).

For the definition of the Wiener - Young variation $\bigvee_{p\,-\infty}^{+\infty}$ of f, $1 \leq p \leq \infty$, and for its relevant properties, we refer to [DP1], [DP2], [SP], [PP], as well as [DM1], [D]. In the sequel we make implicit use of some of the properties of $A_{p,\varepsilon}$ and the three types of moduli of smoothness; all these properties, including the relevant Marchaud-type inequalities and bounds by the Wiener - Young variation, can be collected within the above-listed reference sources.

An equivalent norm in $B_{pq}^s(\mathrm{I\!R})$ can be defined via integral moduli of smoothness (actually this is the "classical" definition):

$\|f\|_{B_{pq}^s(\mathrm{I\!R})} \sim \|f\|_{L_p(\mathrm{I\!R})} + \left[\int_0^{c_0} \left(\xi^{-s}\omega_r(f;\xi)_{L_p(\mathrm{I\!R})}\right)^q \frac{d\xi}{\xi}\right]^{\frac{1}{q}}$, $0 < s < r$, $1 \leq p \leq \infty$,

where $c_0 : 0 < c_0 \leq \infty$ (the constants of equivalence depend on c_0; note that $c_0 = \infty$ is an admissible value).

The so-called A-spaces A_{pq}^s have been defined by V. A. Popov analogously to the above definition of B_{pq}^s; the quasi-norm in $A_{pq}^s(\mathrm{I\!R})$, $0 < s < r$, $0 < p \leq \infty$, $0 < q \leq \infty$ is defined by

$\|f\|_{B_{pq}^s(\mathbb{R})} = \|f\|_{A_{pq}^s(\mathbb{R})} + \left[\int_0^{c_0} \left(\xi^{-s}\tau_r(f;\xi)_{L_p(\mathrm{I\!R})}\right)^q \frac{d\xi}{\xi}\right]^{\frac{1}{q}}$,

(see, e.g., [D], [D1]). In [D1], Section 2, if has been proved that $A_{pq}^s(\mathrm{I\!R}) \rightleftharpoons B_{pq}^s(\mathrm{I\!R})$, $0 < p \leq \infty$, $0 < q \leq \infty$, $s > \frac{1}{p}$. (It can be proved that this isomorphism continues to hold true for $s = \frac{1}{p}$, $q = \min\{1, p\}$, too.)

Like integral moduli, averaged moduli can also be defined in the multivariate case $d > 1$; the above isomorphism between Besov- and A-spaces continues to hold with $\frac{1}{p}$ being replaced by $\frac{d}{p}$.

There is a close relatonship between the Steklov (r,ε)-mean and the integral, local and averaged moduli of smoothness. Its essence is that the Steklov mean can be used to *simultaneously quasi-linearize* (see [BL]) the equivalent K-functionals (see [BL], [DP1], [T1], [PP]) to all these moduli of smoothness. Its form is in terms of upper bounds for the L_p- and $A_{p,\varepsilon}$-norm of $f - f_{\varepsilon,r}$ and $f_{\varepsilon,r}^{(\nu)}$, $\nu = 1, \ldots, r$, via integral and averaged moduli of smoothness (see [SP], [D], [D1], [DP1], [PP]). For the definition of the estimators $\hat{h}_+$ and $\hat{h}_{+,\varepsilon}$, see (3.12,13) and also [DM1].

2 Generalized kernel estimators

Our aim in this section is to propose an unified systematic approach to kernel and wavelet methods for density and regression–function estimation in which new positivity–preserving methods can also be included in a natural way. We shall show that the attempts to unite the notions of a kernel and a wavelet estimator, and in particular, their asymptotic–minimax versions, lead to a natural generalization which we call a "generalized kernel estimator" (GKE) – see Definition 2.1 below. Then, in the end of this section and the beginning of the next one, it will be shown that $\widehat{f}_{+,\varepsilon}$ and its versions with discretization of the integral are also GKE.

Definition 2.1. *A generalized kernel estimators of a (a) density f is:*

(2.1) $\widehat{f}(x) := \frac{1}{N} \sum\limits_{\nu=1}^{N} K(x, X_\nu; X_1, \ldots, X_N),\ x \in \mathbb{R}^d,$

which is an empirical version of

(2.2) $f_1(x) = \int\limits_{\mathbb{R}^d} K(x,t;X_1,\ldots,X_N)\ f(t)\ dt;$

(b) regression function $g : \mathbb{R}^d \to \mathbb{R}$ is

(2.3) $\widehat{g}(x) := \frac{1}{N} \sum\limits_{\nu=1}^{N} Y_\nu . K(x, X_\nu; X_1, \ldots, X_N),$

which is an empirical version of

(2.4) $g_1(x) = \int\limits_{\mathbb{R}^d} K(x,t;X_1,\ldots,X_N)\ g(t)\ dF(t).$

In (2.2,4), for any concrete sample $(X_1, \ldots, X_N)$

(2.5) $K(.;X_1,\ldots,X_N) \in L_{1,loc}(\mathbb{R}^{2d})$ *as a function of* (x,t).

In other words, (2.1,3) are kernel–type estimators whose kernel $K(x,t)$ is variable in t and sample–dependent.

Example 2.1. *(A) Constant–bandwidth kernel estimator*

(a) density:

(2.6) $\widehat{f}(x) := \frac{1}{N} \sum\limits_{\nu=1}^{N} \Psi(x - X_\nu);$

(b) regression:

(2.7) $\widehat{g}(x) := \frac{1}{N} \sum_{\nu=1}^{N} Y_\nu . \Psi(x - X_\nu)$;

(2.8) $K(x,t) = \Psi(x-t)$ *and* K *is not sample–dependent.*

(B) Non–thresholded wavelet estimator

(a) density:

(2.9) $\widehat{f}(x) := \frac{1}{N} \sum_{\nu=1}^{N} K(x, X_\nu)$;

(b) regression:

(2.10) $\widehat{g}(x) := \frac{1}{N} \sum_{\nu=1}^{N} Y_\nu . K(x, X_\nu)$;

(2.11) $K(x,t) \;=\; \sum_{k\in\mathbb{Z}^d} \varphi_{j_0}(x).\varphi_{j_0 k}(t) \;+\; \sum_{j=j_0}^{j_1} \sum_{k\in\mathbb{Z}^d} \sum_{i=1}^{2^d-1} \psi_{jk}^{(i)}(x).\psi_{jk}^{(i)}(t)$

and K *is not sample–dependent.*

Example 2.2. *(Asymptotically minimax estimators)*

(A) Variable–bandwidth kernel estimator

(a) density: $\widehat{f}$ *is defined by (2.1);*

(b) regression (see [LMS]): $\widehat{g}$ *is defined by (2.2);*

(2.12) $K(x,t;X_1,\dots,X_N) \;=\; \frac{1}{\widehat{\chi}(X_1,\dots,X_N)}.\Psi\left[\frac{x-t}{\widehat{\chi}(X_1,\dots,X_N)}\right] \;=\; \Psi_1(x-t;X_1,\dots,X_N)$,

where $\widehat{\chi}$ *is an optimally selected bandwidth (in [LMS] it is denoted by* $\widehat{h}$ *– not to be mixed with* $\widehat{h}$ *in our notations!).*

(B) Thresholded wavelet estimator (see [DJ])

(a) density: $\widehat{f}$ *is defined via (2.1),*

$$
\begin{aligned}
(2.13) \quad K(x,t;X_1,\dots,X_N) \quad &= \quad \sum_{k\in\mathbb{Z}^d} \varphi_{j_0 k}(x)\ \varphi_{jk}(t)\ + \\
&+ \quad \sum_{j=j_0}^{j_1} \sum_{k\in\mathbb{Z}^d} \sum_{i=1}^{2^d-1} \mu_{jk}^{(i)}(X_1,\dots,X_N)\ \psi_{jk}^{(i)}(x)\ \psi_{jk}^{(i)}(t),
\end{aligned}
$$

$$
(2.14) \quad \mu_{jk}^{(i)}(X_1,\dots,X_N) \;=\; \begin{cases} \dfrac{\delta\left[\frac{1}{N}\sum_{\nu=1}^{N} \psi_{jk}^{(i)}(X_\nu);\lambda_j\right]}{\frac{1}{N}\sum_{\nu=1}^{N}\psi_{jk}^{(i)}(X_\nu)}, & \textit{if } \frac{1}{N}\sum_{\nu=1}^{N}\psi_{jk}^{(i)}(X_\nu) \neq 0, \\ 0, & \textit{otherwise}, \end{cases}
$$

where $\delta(\alpha;\lambda_j)$ *is the thresholding rule with thresholding value* λ_j*, as considered in [DJ];*

(b) regression: $\widehat{g}$ *is defined by (2.2);*

$K(x,t;X_1,\dots,X_N)$ is defined by (2.13), but with modified values of $\mu_{jk}^{(i)}$:

$$(2.15)\quad \mu_{jk}^{(i)}(X_1,\dots,X_N) = \begin{cases} \dfrac{\delta\left[\frac{1}{N}\sum\limits_{\nu=1}^{N} Y_\nu.\gamma_{jk,\nu}^{(i)}\psi_{jk}^{(i)}(X_\nu);\lambda_j\right]}{\frac{1}{N}\sum\limits_{\nu=1}^{N} Y_\nu.\gamma_{jk,\nu}^{(i)}\psi_{jk}^{(i)}(X_\nu)}, & \text{if } \frac{1}{N}\sum\limits_{\nu=1}^{N} Y_\nu\ .\ \gamma_{jk,\nu}^{(i)}\ .\ \psi_{jk}^{(i)}(X_\nu)\neq 0, \\ 0, & \text{otherwise,} \end{cases}$$

$$(2.16)\quad \gamma_{jk,\nu}^{(i)} = \begin{cases} 1, & \text{if } |Y_\nu.\psi_{jk}^{(i)}(X_\nu)| \leqq M \\ 0, & \text{otherwise,} \end{cases}$$

$M = \sqrt{\dfrac{K.N}{\ln N}}$, $K = const$, is the truncation value, as considered in [DJ]. (In [DJ] the value of M is misprinted.)

Remark 2.1. *Comparing Examples 2.1,2, the following two observations can be made:*
(i) Requirements for achieving asymptotic–minimax rates bring about variability and sample–dependence of the estimator's kernel.

(ii) From the point of view of Definition 2.1, the only essential difference between homologous kernel and wavelet estimators is that kernel estimators $K(x,t)$ depend on $x-t$ only, which implies that kernel estimators commute with translations over the whole of $\mathrm{I\!R}^d$, whileas the respective wavelet estimators commute with shifts only over the discrete group $2^{-j_1}.\mathbb{Z}^d$.

Proposition 2.1. *If $\widehat{f}$ (or $\widehat{g}$) is a GKE of the density f (resp. the regression function g), with kernel $K(x,t;X_1,\dots,X_N)$, then $\widehat{f}_{+,\varepsilon}$ (resp. $\widehat{g}_{+,\varepsilon}$) is also a GKE, with kernel*

$$(2.17)\quad K_\varepsilon(x,t;X_1,\dots,X_N) = \left[I_{\{\xi\in\mathrm{I\!R}^d\ :\ h(\xi)>0\}}(\cdot)\ K(.,t;X_1,\dots,X_N)\right]_\varepsilon(x),$$

where $h := \widehat{f}$ (resp. $h := \widehat{g}$).

3 Main results

In this section we first recall the main result of [DM1]. In brief, it gives the answer to the following question: "given an estimator $\widehat{h}$ of h achieving certain rate of estimation with respect to a certain metric when h belongs to a certain function space, what is the range of ε in $\widehat{h}_{+,\varepsilon}$, so that $\widehat{h}_{+,\varepsilon}$ achieves the same rates of estimation of h as $\widehat{h}$, with respect to the same metric and the same function space?"

The definition of $\widehat{h}_{+,\varepsilon}$ involves an integral. In some cases it is possible to compute the integral explicitly, in closed form, but in general one has to apply numerical integration, and in this way new estimators of h are generated from $\widehat{h}_{+,\varepsilon}$. These estimators are the main object of study in this paper. We shall be considering in detail the case $d=1$ only, but the results and the ideas of the proofs can be extended to the multivariate case, too. The few aspects, in which the multivariate case differs considerably from the univariate one, will be addressed in the concluding remarks.

As a method for numerical integration, we utilize quadrature formulae. The new estimators obtained in this way are proved to be GKE as $\widehat{h}_{+,\varepsilon}$ itself (see section 2), and to have the same pointwise regularity as the kernel ϕ in the definition of approximate identity (see section 1). Besides their dependence on ε, these new estimators are also functions of certain additional parameters of the quadrature formula. This justifies the general notation $\widehat{h}_{+,\varepsilon,\mathrm{quadr}}$ for all these estimators. The main results in this paper are several theorems which are analogue of the main result in [DM1] – they give reply to the following question: "what is the range of ε and the relevant parameters of the quadrature formula, so that $\widehat{h}_{+,\varepsilon,\mathrm{quadr}}$ achieves the same rates of estimation of h as $\widehat{h}$, with respect to the same metric and the same function space?"

The first of these theorems is about non–adaptive quadrature formulae. This is the simplest algorithmically, but the most computationally expensive estimator $\widehat{h}_{+,\varepsilon,\mathrm{quadr}}$. The next theorems suggest several ways, ordered in increasing efficiency, to decrease the volume of computations by utilizing adaptive quadrature formulae, depending on the concrete $\widehat{h}$.

Finally, in the end of this section we consider possible choices of the wavelet or the kernel in the definition of $\widehat{h}$, and of the kernel Φ in the definition of approximate identity, for which the integral in $\widehat{h}_{+,\varepsilon}$'s definition can be computed exactly, in closed form. This method is so computationally inexpensive that, at least theoretically, it is comparable in its numerical performance with $\widehat{h}$ itself.

The main result of [DM1] is

Theorem 3.A (Theorem 2.1 in [DM1]). *Assume that*

(3.1) $d \in \mathbb{N}$;

$h : \mathrm{I\!R}^{\mathrm{d}} \to \mathrm{I\!R}$ *is such that there exists a constant* $v \in \mathrm{I\!R}$, *so that*

(3.2) $h(x) \geqq v$

holds for Lebesgue–almost all $x \in \mathrm{I\!R}^{\mathrm{d}}$.

(3.3) $N \in \mathbb{N}$; X_j, $j = 1, \dots, N$ *are i.i.d. with c.d.f.* F;

(3.4) $\widehat{h} = \widehat{h}_N : \mathrm{I\!R}^{\mathrm{d}} \to \mathrm{I\!R}$ *is an estimator of* h *based on the i.i.d. observations* X_j, $j = 1, \dots, N$;

(3.5) $1 \leqq p_1 \leqq \infty$;

$0 < \rho < \infty$;

(3.6) $h \in L_{p_1}(\mathrm{I\!R}^{\mathrm{d}})$, $\widehat{h} \in \mathcal{L}_\rho(\mathrm{I\!R}^{\mathrm{d}}, \mathrm{L}_{\mathrm{p}_1}(\mathrm{I\!R}^{\mathrm{d}}))^\rho$;

W is a (quasi–) Banach space such that

(3.7) $W \hookrightarrow L_{p_1}(\mathrm{I\!R}^{\mathrm{d}}) \ + \ \dot{W}^2_{p_1}(\mathrm{I\!R}^{\mathrm{d}})$,

there exist $\delta_0 > 0$, $\beta \in \Omega(\delta_0)$, *so that*

(3.8) $\omega_2(h_1, \delta)_{L_{p_1}} \leqq \beta(\delta)\ \|h_1\|_W$, $\forall h_1 \in W \qquad \forall \delta \in (0, \delta_0]$;

(3.9) $h \in W$ *and* $\|h\|_W \leqq L$ *for some* $L \in (0, \infty)$;

there exist $\alpha \in \Omega(\delta_0)$, $K = K(p_1, \rho, W, L, \alpha) \in (0, \infty)$ *and*

(3.10) $N_0 = N_0(p_1, \rho, W, L, \alpha) \in (0, \infty)$, *such that*

(3.11) $E_{F^N} \|h - \widehat{h}_N\|^\rho_{L_{p_1}} \leqq K \ . \ \alpha\left(\frac{1}{N}\right), \qquad \forall N > N_0.$

Consider

(3.12) $\widehat{h}_+ \ : \ \widehat{h}_+(x) = \max\{v, \widehat{h}(x)\}$, $x \in \mathbb{R}^d$;

(3.13) $\widehat{h}_{+,\varepsilon} \ : \ \widehat{h}_{+,\varepsilon}(x) = \Phi_\varepsilon * \widehat{h}_+(x)$, $\forall x \in \mathbb{R}^d$, $\forall \varepsilon > 0$,

where Φ *satisfies the requirements for approximate identity.*

Then,

(3.14) $E_{F^N} \|h - \widehat{h}_{+,\varepsilon}\|^\rho_{L_{p_1}} \leqq 2^{\rho_1} \ (3 + 2^{\rho/\rho_1})^{\rho_1} \ . \ K \ . \ \alpha\left(\frac{1}{N}\right), \ \ \forall N > N_0, \ \ \forall \varepsilon \in (0, \varepsilon_N],$

where

(3.15) $\rho_1 = \max\{1, \rho\}$,

(3.16) $\varepsilon_N = \beta^{-1}\left(\frac{2}{L}. \ (3 + 2^{\rho/\rho_1})^{\rho_1/\rho}.K^{1/\rho}.\alpha\left(\frac{1}{N}\right)^{1/\rho}\right).$

In the context of the regression model $h = g$; in the density model $h = f$ and $v \equiv 0$ (see Remark 2.1 in [DM1]).

Remark 3.1. *The quasinorm in* W *is homogeneous, therefore, without loss of generality in (3.9) we may assume* $\|h\|_W = L$. *We shall be essentially relying on this observation in our proofs, which, in particular, means that the requirement about* W *being a quasi–normed linear space (not just a quasi–normed abelian group) is essential. In this text we consider in detail the case when* W *is a Besov space. Analogous asymptotic theory can be developed when* W *belongs to other scales of relevant function spaces, e.g., the Triebel–Lizorkin and Wiener–Young spaces (see [DM1]). We shall give more details in Section 8.*

In most cases it is possible to compute $\widehat{h}_{+,\varepsilon}$ only numerically. In the sequel we consider numerical approximations to $\widehat{h}_{+,\varepsilon}$; for conciseness of presentation the case $d = 1$ and $v \equiv 0$ is assumed throughout. The general cases for d and v are discussed in separate remarks. The numerical approximation to $\widehat{h}_{+,\varepsilon}$ is achieved by replacing the convolution integral in the definition of $\widehat{h}_{+,\varepsilon}$ by a composite quadrature formula in the following way:

(3.17) $\widehat{h}_{+,\varepsilon}(x) = \frac{1}{\varepsilon} \int\limits_{-\infty}^{\infty} \Phi\left(\frac{x-t}{\varepsilon}\right) \ \widehat{h}_+(t) \ dt = \frac{1}{\varepsilon} \sum\limits_{\mu=-\infty}^{\infty} \int\limits_{\Delta_\mu} \Phi\left(\frac{x-t}{\varepsilon}\right) \ \widehat{h}_+(t) \ dt \ \mapsto$

$\mapsto \ \widehat{h}_{+,\varepsilon,\mathrm{quadr}}(x) := \ \frac{1}{\varepsilon} \sum\limits_{\mu=-\infty}^{\infty} \sum\limits_{k=1}^{n_\mu} A_{\mu k} \ . \ \Phi\left(\frac{x-t_{\mu k}}{\varepsilon}\right). \ \widehat{h}_+(t_{\mu k}),$

where:

(3.18) $\{\Delta_\mu\}_{\mu\in\mathbb{Z}}$ is a decomposition of $\mathbb{R}$ into non–intersecting intervals;

(3.19) $0 < \inf\limits_\mu |\Delta_\mu| \quad \text{and} \quad \sup\limits_\mu |\Delta_\mu| < \infty$;

(3.20) $t_{\mu k} \in \Delta_\mu$, $A_{\mu k} \geqq 0$,

(3.21) $\inf\limits_{\mu,\mu_1,k,k_1} |t_{\mu_1 k_1} - t_{\mu k}| = C_0 > 0.$

Proposition 3.1. *If $\widehat{f}$ (or $\widehat{g}$) is a GKE of the density f (resp. the regression function g), with kernel $K(x,t;X_1,\dots,X_N)$, then $\widehat{f}_{+,\varepsilon,quadr}$ (resp. $\widehat{g}_{+,\varepsilon,quadr}$) is also a GKE, with kernel*

$$(3.22)\quad K_{\varepsilon,quadr}(x,t;X_1,\dots,X_N) = \Big[I_{\{\xi\in\mathbb{R}\,:\,\mathrm{h}(\xi)>0\}}(\cdot)\; K(.,t;X_1,\dots,X_N)\Big]_{\varepsilon,quadr}(x),$$

where $h := \widehat{f}$ (resp. $h := \widehat{g}$); moreover,

$$(3.23)\quad \widehat{f}_{+,\varepsilon,quadr} \text{ is non–negative on } \mathbb{R} \text{ (resp. } \widehat{g}_{+,\varepsilon,quadr}\text{) is non–negative on } \mathbb{R}).$$

Theorem 3.1. *Let $r \in \mathbb{N}$ and $\Phi \in C_0^r(\mathbb{R})$, supp $\Phi = [-1,1]$, be the kernel in the definition of approximate identity. Then,*

$$(3.24)\quad \widehat{f}_{+,\varepsilon,quadr} \in C^r(\mathbb{R}),\ \text{resp. } \widehat{g}_{+,\varepsilon,quadr} \in C^r(\mathbb{R}).$$

If, moreover, $\widehat{f} \in C_0^r(\mathbb{R})$, resp. $\widehat{g} \in C_0^r(\mathbb{R})$, then also

$$(3.24')\quad \widehat{f}_{+,\varepsilon,quadr} \in C_0^r(\mathbb{R}),\ \text{resp. } \widehat{g}_{+,\varepsilon,quadr} \in C_0^r(\mathbb{R}).$$

In other words, quadrature approximations of $\widehat{f}_{+,\varepsilon}$ and $\widehat{g}_{+,\varepsilon}$ generated via (3.17) have the same regularity as the smoothing kernel Φ, and have compact support as soon as $\widehat{f}$ and $\widehat{g}$ are compactly supported.

If no additional study of the graph of $\widehat{h} = \widehat{f}$ (or $\widehat{h} = \widehat{g}$) is made, under the assumptions of Theorem 3.1 the only knowledge about $\widehat{h}_+$ is that it is C^r–regular just as $\widehat{h}$ itself, everywhere on $\mathbb{R}$ except in those x in which $\widehat{h}$ changes its sign. In the later x, $\widehat{h}_+$ is continuous but not necessarily C^1–smooth. Such "bad" x will henceforward be called "critical points of $\widehat{h}$ (of the first kind)". These are the kind of "bad" points one can encounter with the wavelet estimator considered in [DJ]. In the case of the kernel estimator in [LMS], however, the set of critical points must include also "critical points of the second kind": those x where a jump of the bandwidth has occured (see the definition in [LMS] of the bandwidth (in their notations) $\widehat{h}_\sigma(t)$ – not to be mixed with $\widehat{h}$ in our notations!). In these special points the estimator $\widehat{h}$ of Lepski *et al.* can be non–smooth, even discontinuous. Thus, no matter how big the regularity index r of $\widehat{h}$ is, $\widehat{h}_+$ is at most absolutely continuous on $\mathbb{R}$ in the case of [DJ] and piece–wise absolutely continuous with bounded variation on $\mathbb{R}$ in the case of [LMS].

The simplest from algorithmic point of view (and the most expensive from computational point of view) $\widehat{h}_{+,\varepsilon,\mathrm{quadr}}$ is obtained if the decomposition of $\mathbb{R}$ into $\{\Delta_\mu\}$ and the choice of the quadrature formula in each Δ_μ is done without preliminary localization of the critical points of $\widehat{h}_+$. Then the mesh can be chosen uniform:

$$(3.25)\quad \Delta_\mu = [\mu\delta\varepsilon, (\mu+1)\delta\varepsilon],\ \mu \in \mathbb{Z}, \qquad \text{where } 0 < \delta \leqq 1.$$

It does not make sense to use a quadrature formula which is accurate of high order, because in general $\widehat{h}_+$ does not have regularity higher than 1 in Δ_μ. (We recall that a quadrature formula on Δ is called accurate of order $r_1 \in \mathbb{N}$, if the error of replacing the integral over Δ by the quadrature formula is zero for any integrand which is an algebraic polynomial of degree not exceeding $r_1 - 1$

on Δ, but there exists polynomial integrand of degree r_1 on Δ for which this error is not zero.) In other words, one has to consider a very dense mesh and a very simple quadrature formula (it is enough to be accurate of order $r_1 = 1$). The estimator $\widehat{h}_{+,\varepsilon,\mathrm{quadr}}$ depends on ε and δ in (3.25), so it will be denoted by $\widehat{h}_{+,\varepsilon,\delta}$:

(3.26) $\widehat{h}_{+,\varepsilon,\delta} := \widehat{h}_{+,\varepsilon,\mathrm{quadr}}$ when Δ_μ is given by (3.25) and the quadrature formula is accurate of order 1.

Together with ε, δ must be selected so that $\widehat{h}_{+,\varepsilon,\delta}$ retains asymptotic–minimax optimality.

Theorem 3.2. *Under the conditions of Theorem 3.A, for $\widehat{h}$ being the wavelet density or regression–function estimator from [DJ], assume: $d = 1$, $v \equiv 0$ on $\mathbb{R}$;*

(3.27) $W = B^s_{pq}$, $1 \leqq p \leqq \infty$, $1 \leqq q \leqq \infty$, $s_0 \leqq s < r$, *where* $q = 1$ *if* $s = s_0$, *and*

(3.28) $r \in \mathbb{N}$, $r \geqq 2$, $1 \geqq s_0 \geqq \max\{1/p,\ r_0\}$, $\quad 0 < r_0 \leqq 1$;

(3.29) $\Phi \in C_0^r(\mathbb{R})$, *supp* $\Phi = [-1, 1]$, *is as in the definition of approximate identity;*

(3.30) *supp* h $\cup$ *supp* $\widehat{h}$ $\subset [-R/2, R/2]$, $\quad 0 < R < \infty$, *where R does not depend on the sample size N;*

(3.31) $0 < \varepsilon \leqq 1$, $0 < \delta \leqq 1$ *(for convenience of presentation only);*

(3.32) *the quadrature formula on Δ_μ is accurate of order 1, $\forall \mu \in \mathbb{Z}$.*

Then, $\exists N_1 \geqq N_0$:

(3.33) $$E_{F^N} \|h - \widehat{h}_{+,\varepsilon,\delta}\|^\rho_{L_{p_1}} \leqq 2^{2\rho_1} (3 + 2^{\rho/\rho_1})^{\rho_1} \ .\ K \ .\ \alpha\left(\frac{1}{N}\right),$$

$\forall N > N_1$, $\forall \varepsilon \in (0, \varepsilon_N]$, $\forall \delta \in (0, \delta_N]$,

where:

(3.34) p_1, ρ, ρ_1, K, N_0, ε_N *are as defined in Theorem 3.A;* $\rho = \rho_1 = 2$;

(3.35) $$\delta_N = C^{1+\left(1/s_0 - 1/s_0^*\right)_+} \ .\ \left[\alpha\left(\frac{1}{N}\right)^{1/\rho} \beta^{-1} \left(C_1 \alpha\left(\frac{1}{N}\right)^{1/\rho}\right)^{1-1/p_1}\right]^{1+\left(1/s_0 - 1/s_0^*\right)_+};$$

(3.36) $C_1 = \dfrac{2}{L} \ .\ (3 + 2^{\rho/\rho_1})^{\rho_1/\rho} \ .\ K^{1/\rho}$ *(see (4) in [DM1]); L is as in Theorem 3.A;*

(3.37) $$C = C_2(p, p_1, \rho, r_0, s, L) \ .\ \frac{K^{1/\rho}}{R^{1-1/p} \, \|\Phi\|_{W^1_{p_1}}};$$

(3.38) $$s_0^* = 1 - \frac{\log_N \left[\beta^{-1}\left(C_1 \alpha\left(\frac{1}{N}\right)^{1/\rho}\right)\right]}{\log_N C + \frac{1}{\rho}\log_N \alpha\left(\frac{1}{N}\right) + \left(2 - \frac{1}{p_1}\right) \log_N \left[\beta^{-1}\left(C_1 \alpha\left(\frac{1}{N}\right)^{1/\rho}\right)\right]} \in$$
$\in (0, 1)$.

Remark 3.2. *The above theorem holds true also if $\widehat{h}$ is the kernel estimator from [LMS], with the following modifications:*

- *(3.33) becomes*

(3.33′) $E_{F^N}\|h-\widehat{h}_{+,\varepsilon,\delta}\|^{\rho}_{L_{p_1}} \;=\; O\left(\alpha\left(\frac{1}{N}\right)\right)$; $N\to\infty$; $\rho=p_1$;

- *(3.35) becomes*

(3.35′) $\delta_N = O\left\{\left[\alpha\left(\frac{1}{N}\right)^{1/\rho}\beta^{-1}\ \left(C_1\alpha\left(\frac{1}{N}\right)^{1/\rho}\right)^{1-1/p_1}\right]^{1+\left(1/s_0-1/s_0^*\right)_+}\right\}$,

$N\to\infty$.

The constants in the RHS of (3.33', 35') depend on $\bigvee_{-\infty^p}^{\infty}\widehat{h}$. *The idea of the proof of Theorem 3.2 allows obtaining explicit bounds for the two constants, involving* $\bigvee_{-\infty^p}^{\infty}\widehat{h}$, *and, therefore, these bounds are available for any concrete* $\widehat{h}$.

Let us analyze this result. The rates are in terms of p, p_1, α, β. This means that information about s and p (W is a Besov space) must be available. p_1 is also fixed. If we want to achieve simultaneous asymptotic–minimax optimality for all p, p_1: $1\leqq p\leqq\infty$, $1\leqq p_1\leqq\infty$, and for all $s\in[s_0,r)$, then the proofs of the corollaries in the next section yield

(3.39) $\alpha\left(\frac{1}{N}\right) \;=\; N^{-\frac{2r}{2r+1}}$, $\beta(\gamma)=\gamma^{r_0}$.

For simplicity, assume

(3.40) $r_0=1$.

(This is actually the most common choice of r_0 in practical problems.) Then, $s_0=1$ and, by Corollary 4.2 (4.11), (3.41) $\delta_N=O\left[N^{-(2-1/p_1)\,\frac{r}{2r+1}}\right]$, for fixed p_1,

or, by Corollary 4.3 (4.11'),

(3.42) $\delta_N=O\left(N^{-\frac{2r}{2r+1}}\right)$, uniformly in p_1: $1\leqq p_1\leqq\infty$.

This small value of δ_N shows that this simplest estimator $\widehat{h}_{+,\varepsilon,\delta}$ has high computational complexity, compared to the underlying estimator $\widehat{h}$.

The general way to decrease the volume of computations is clear: one has to take into consideration the concrete properties of the concrete $\widehat{h}$ and $\widehat{h}_+$ and only then define an *adaptive* mesh and/or quadrature formula depending on the concrete $\widehat{h}$. Below we suggest several ways for achieving this, ordered in increasing efficiency, expressed by respective increase of the value of δ_N (and the ultimate one being the *exact* computation of $\widehat{h}_{+,\varepsilon}$ for spline–wavelets and spline–kernels).

An important question arising here is: what methods should be used in the preliminary stage of studying the properties of $\widehat{h}$'s graph, and what is their own computational complexity? Fortunately, in practice there is a reasonable simple answer to it. On the one hand, what we need to achieve in this preliminary stage is localize the critical points of $\widehat{h}$. On the other hand, we usually plot the graph of $\widehat{h}$. The data accumulated in the plotting–data file are sufficient for the localization of all the critical points (both the "first" and the "second" kind of them, as was discussed above). Thus, assuming that plotting our estimators' graphs is "*conditio sine qua non*", we get the localization of critical points essentially for free. In section 5 we continue the discussion of this question.

Let us assume first that only a "rough" localization of the critical points of $\widehat{h}$ has been carried out. More precisely, consider the uniform cover with disjoints interiors

(3.43) $\quad \mathbb{R} = \cup_{k=-\infty}^{\infty} \ [k\varepsilon, (k+1)\varepsilon],$

and assume that it is known for which $k \in \mathbb{Z}$ $\qquad [k\varepsilon, (k+1)\varepsilon]$ contains at least one critical point of $\widehat{h}$.

Let us define $\{\Delta_\mu\}$. Consider δ, δ_1: $\delta_1 \geqq \delta > 0$, $\frac{1}{\delta} \in \mathbb{N}$, $\frac{1}{\delta_1} \in \mathbb{N}$, whose values will be specified later on.

For any $k \in \mathbb{Z}$ such that $[k\varepsilon, (k+1)\varepsilon]$ contains a critical point of $\widehat{h}$, consider the refinement

(3.44) $\quad [k\varepsilon, (k+1)\varepsilon] \ = \ \cup_{\nu=1}^{1/\delta} \ [(k + (\nu-1)\delta)\varepsilon, \ (k+\nu\delta)\varepsilon].$

For any $k_1 \in \mathbb{Z}$ such that $[k_1\varepsilon, (k_1+1)\varepsilon]$ does not contain a critical point of $\widehat{h}$, consider

(3.45) $\quad [k_1\varepsilon, (k_1+1)\varepsilon] \ = \ \cup_{\nu_1=1}^{1/\delta_1} \ [(k_1 + (\nu_1-1)\delta_1)\varepsilon, \ (k_1+\nu_1\delta_1)\varepsilon].$

Denote $\Lambda = \{k \in \mathbb{Z} : \ (3.44) \text{ holds}\}$, $\Lambda_1 = \{k_1 \in \mathbb{Z} : \ (3.45) \text{ holds}\}$. The intervals $\{\Delta_\mu\}$, $\mu \in \mathbb{Z}$, are the intervals $[(k + (\nu-1)\delta)\varepsilon, \ (k+\nu\delta)\varepsilon]$, $k \in \Lambda$, and $[(k_1 + (\nu_1-1)\delta)\varepsilon, \ (k_1+\nu_1\delta)\varepsilon]$, $k_1 \in \Lambda_1$, reordered together, by increasing abscissae.

Define

(3.46) $\quad I := \{\mu \in \mathbb{Z} : \ \Delta_\mu \text{ is of type } [(k + (\nu-1)\delta)\varepsilon, \ (k+\nu\delta)\varepsilon], \ k \in \Lambda \ \},$

(3.46′) $\quad I_1 := \{\mu \in \mathbb{Z} : \ \Delta_\mu \text{ is of type } [(k_1 + (\nu-1)\delta_1)\varepsilon, \ (k_1+\nu\delta_1)\varepsilon], \ k_1 \in \Lambda_1 \ \}.$

In each Δ_μ, $\mu \in I$, consider, as previously, a quadrature formula which is accurate of order 1; if $\mu \in I_1$, then consider a quadrature formula accurate of order r_1. Here r_1 is any integer between 1 and r : $1 \leqq r_1 \leqq r$, and is the same for all $\mu \in I_1$. Given $\widehat{h}$, denote

(3.26′) $\quad \widehat{h}_{+,\varepsilon,\delta,\delta_1} \ := \ \widehat{h}_{+,\varepsilon,\text{quadr}},$

where $\widehat{h}_{+,\varepsilon,\text{quadr}}$ corresponds to the above–defined mesh and composite quadrature formula.

Theorem 3.3. *Under the conditions of Theorem 3.2, assume that (3.43-46, 46', 26') holds ($\widehat{h}$ being the wavelet estimator in [DJ]). Suppose further that*

(3.47) $\quad r_1 \in \mathbb{N}, \ 1 \leqq r_1 \leqq r;$

(3.32) holds only for $\mu \in I$; for $\mu_1 \in I_1$ there holds

(3.32′) $\quad$ *the quadrature formula on Δ_μ is accurate of order r_1,*

$\forall \mu_1 \in I_1$

$s:\ \max\{s_0,\ \sigma\} < s < r,\ \sigma:\ \max\{1/p,\ r_0\} \leqq \sigma < r.$

Then, $\exists N_2 \geqq N_1$*:*

$$(3.48)\quad E_{F^N}\|h - \widehat{h}_{+,\varepsilon,\delta,\delta_1}\|^{\rho}_{L_{p_1}} \ \leqq\ 2^{2\rho_1}\ (3+2^{\rho/\rho_1})^{\rho_1}\ .\ K\ .\ \alpha\left(\frac{1}{N}\right),$$

$\forall N > N_2,\ \forall \varepsilon \in (0, \varepsilon_N],\ \forall \delta \in (0, \delta_N],\ \forall \delta_1 \in (0, \delta_{1,N}],$

where:

(3.49) $\varepsilon_N = \min\{1,\ \varepsilon'_N\}$, ε'_N *being the value of* ε_N *as of Theorem 3.2;*

$$(3.50)\quad \delta_N = \min\left\{1,\ C^{1+\left(1/s_0 - 1/s_0^*\right)_+}\ .\ \left[\alpha(1/N)^{1/\rho}\beta^{-1}\ \left[C_1\ \alpha(1/N)^{1/\rho}\right]^{1-1/p_1}\right]^{1+\left(1/s_0-1/s_0^*\right)_+}\right\};$$

$$(3.51)\quad \delta_{1,N} = \min\left\{1,\ C_2^{1/r_1+(1/\sigma^*-1/\sigma^*)_+}\ .\ \left[\alpha(1/N)^{1/\rho}\beta^{-1}\ \left[C_1\ \alpha(1/N)^{1/\rho}\right]^{1-1/p_1}\right]^{1/r_1+(1/\sigma-1/\sigma^*)_+}\right\};$$

(3.52) $C,\ C_1$ *are as in Theorem 3.2;*

(3.53) $C_2 = C_2(r, r_1, p, q, \rho, R, \phi)$ *can be evaluated from the proof, similarly to* C*;*

$$(3.54)\quad \sigma^* \ =\ \left[1\ -\ \frac{r_1\ \log_N \varepsilon_N}{\log_N C_2\ +\ \frac{1}{\rho}\log_N \alpha\left(\frac{1}{N}\right)\ +\ \left(r_1 + 1 - \frac{1}{p_1}\right)\ \log_N \varepsilon_N}\right]\ .\ r_1\ \in$$

$\in\ (0, r_1)$*;* s_0^* *is defined via (3.38).*

Remark 3.3. *The above theorem is also valid if* $\widehat{h}$ *is the kernel estimator from [LMS], with the following modifications:*

- *(3.48) becomes*

  $$(3.48')\quad E_{F^N}\|h - \widehat{h}_{+,\varepsilon,\delta,\delta_1}\|^{\rho}_{L_{p_1}} \ =\ O\left(\alpha\left(\frac{1}{N}\right)\right);\ N \to \infty;$$

- *(3.50) becomes*

  $$(3.50')\quad \delta_N = O\left\{\left[\alpha(1/N)^{1/\rho}\beta^{-1}\ \left(C_1\ \alpha(1/N)^{1/\rho}\right)^{1-1/p_1}\right]^{1+\left(1/\sigma_0-1/s_0^*\right)_+}\right\},$$

  $\delta_N \leqq 1$*;*

- *(3.51) becomes*

  $$(3.51')\quad \delta_{1,N} = O\left\{\left[\alpha(1/N)^{1/\rho}\beta^{-1}\ \left(C_1\ \alpha(1/N)^{1/\rho}\right)^{1-1/p_1}\right]^{1/r_1+(1/\sigma-1/\sigma^*)_+}\right\},$$

  $\delta_{1,N} \leqq 1.$

As in Remark 3.2, the constants in the O–estimates in (3.48',50',51') depend on $\bigvee_{-\infty^p}^{\infty} \widehat{h}$.

Remark 3.4. *The assumption $\varepsilon_N \leqq 1$, $\delta_N \leqq 1$, $\delta_{1,N} \leqq 1$ is made only for the sake of simplicity of presentation. If more general bounds for these parameters are considered, these bounds' values will also appear in the expressions for the constants involved in the theorem.*

Assume now that not only have been the critical points of $\widehat{h}$ roughly localized, but that they have been computed with sufficiently higher precision than the step δ in Theorem 3.3. Then, we may consider the following limiting case of $\widehat{h}_{+,\varepsilon,\delta,\delta_1}$.

Each interval between two neighbouring critical points $x_1, x_2 : \ x_1 < x_2$ is being subdivided into ν intervals of equal length, where

$$(3.55) \quad \nu \ = \ \left[\frac{x_2 - x_1}{\delta_1}\right] \ + \ 1.$$

The intervals obtained by this procedure form a cover of $[x_1, x_l]$, where x_1 is the leftmost, and x_l – the rightmost, critical point of $\widehat{h}$. This cover is being completed to a cover of $\mathbb{R}$ by uniform δ_1–partition of $(-\infty, x_1]$ and $[x_l, \infty)$. Assuming that $l = l_N$ remains bounded as $N \to \infty$ (which should be the case, unless the graph of h has fractal properties); the cover obtained is a δ_1– quasi–uniform partition of $\mathbb{R}$ into $\{\Delta_\mu\}$: there exist $0 < K_1(l_0) \leqq K_2(l_0) < \infty$, independent of N, so that

$$(3.56) \quad K_1(l_0)\ \delta_1 \ \leqq \ |\Delta_\mu| \ \leqq \ K_2(l_0)\ \delta_1, \quad \forall \mu \in \mathbb{Z},$$

where l_0 is the upper bound for l_N. The so–defined Δ_μ all enjoy the property of not having a critical point of $\widehat{h}$ in their interiors. Therefore a quadrature formula accurate of high order can be applied for any Δ_μ. This leads to the following limiting case of the estimator $\widehat{h}_{+,\varepsilon,\delta_1}$ and of Theorem 3.3:

$$(3.57) \quad \widehat{h}_{+,\varepsilon,\delta_1} \ = \ \widehat{h}_{+,\varepsilon,\mathrm{quadr}} \ .$$

Theorem 3.4. *Under the conditions of Theorem 3.3, assume that $I = \emptyset$, $I_1 = \mathbb{Z}$, and $\{\Delta_\mu\}$ is generated in the above way, (3.56) holding true, and all critical points of $\widehat{h}$ being among the boundary points of Δ_μ, $\mu \in \mathbb{Z}$. Then, $\exists N_2' \geqq N_1$:*

$$(3.58) \quad E_{F^N} \|h - \widehat{h}_{+,\varepsilon,\delta_1}\|^\rho_{L_{p_1}} \ \leqq \ 2^{2\rho_1}\ (3 + 2^{\rho/\rho_1})^{\rho_1} \ . \ K \ . \ \alpha(1/N),$$
$$\forall N > N_2', \ \forall \varepsilon \in (0, \varepsilon_N], \ \forall \delta_1 \in (0, \delta_{1,N}],$$

where ε_N and $\delta_{1,N}$ are defined as in Theorem 3.3.

Remark 3.5. *Theorem 3.4 is also valid if $\widehat{h}$ is the kernel estimator from [LMS], with (3.48') for $\widehat{h}_{+,\varepsilon,\delta_1}$ and (3.51') holding.*

Let us analyze the result of Theorem 3.4 and compare it to Theorem 3.2. For simplicity, consider only the case $\sigma \geqq \sigma^*$, i.e.; $\sigma^* < s < r$ in Theorem 3.4.

Assume, also for simplicity, $r_1 = r$, $r_0 = 1$.

With α, β, r_0 as in (3.39, 40), Corollaries 4.2,3 (4.12,12') yield

(3.59) $$\sigma^* = \frac{o(1) + 1 + (1 - 1/p_1)\ /r_0}{o(1) + 1 + (r_1 + 1 - 1/p_1)/r_0} \cdot r_1 = \frac{o(1) + 2 - 1/p_1}{o(1) + r + 2 - 1/p_1} \cdot r \leqq \frac{2}{r+2} \cdot r,$$

(3.60) $$\delta_{1,N} = O\left[N^{-\frac{r(r_0+1-1/p_1)}{r_0 r_1 (2r+1)}}\right], \qquad \text{for fixed } \ p_1 : \ 1 \leqq p_1 \leqq \infty,$$

or, for all $p_1 \in [1, \infty]$ and for $r_0 = 1$, $r_1 = r$, $\frac{2}{r+2} \cdot r < s < r$,

(3.61) $$\delta_{1,N} = O\left(N^{-\frac{2}{2r+1}}\right).$$

Comparing (3.61) to (3.42) shows that $\widehat{h}_{+,\varepsilon,\delta_1}$ is significantly less computationally expensive than $\widehat{h}_{+,\varepsilon,\delta}$, while achieving the same asymptotic–minimax rates. In fact, the computation of $\widehat{h}_{+,\varepsilon,\delta_1}$ for, say, $N \leqq 10^4$ and the biggest possible value for ε : $\varepsilon = \varepsilon_N$, is within very reasonable real–time limits, as long as the underlying estimator $\widehat{h}$ is not too computationally complex itself. The smaller the value $0 < \varepsilon \leqq \varepsilon_N$ is selected, the faster the computation of $\widehat{h}_{+,\varepsilon,\delta_1}$.

Remark 3.6. *In particular, Theorems 3.2-4 imply that, not surprisingly, the bigger the values of s_0 and σ (i.e., the more restricted the consideration is to smooth functions only), the bigger the values of ε_N, δ_N, $\delta_{1,N}$, i.e., the easier it is to compute $\widehat{h}_{+,\varepsilon}$ by quadratures with sufficient precision, and the more advantageous it is to utilize $\widehat{h}_{+,\varepsilon,\delta,\delta_1}$ or $\widehat{h}_{+,\varepsilon,\delta_1}$ rather than $\widehat{h}_{+,\varepsilon,\delta}$.*

So far we have been considering estimators $\widehat{h}_{+,\varepsilon,\text{quadr}}$ whose computational complexity, compared to that of $\widehat{h}$, increases with the increase of N. More precisely, (3.42,61) imply

(3.62) $$\text{comput. compl.}(\widehat{h}_{+,\varepsilon,\delta}) \ = \ O\left(N^{\frac{2r}{2r+1}}\right) \times \text{comput. compl.}(\widehat{h}),$$

(3.63) $$\text{comput. compl.}(\widehat{h}_{+,\varepsilon,\delta_1}) \ = \ O\left(N^{\frac{2}{2r+1}}\right) \times \text{comput. compl.}(\widehat{h}).$$

Now we shall discuss a method for exact computation of $\widehat{h}_{+,\varepsilon}$ whose computational complexity may go down to $O(1)$. comput. compl.$(\widehat{h})$. Implementation of this method for wavelet estimators is being discussed below. The non–negative smoothing kernel Φ is chosen to be a B–spline of degree r ($\Phi(x) > 0$, $\forall x \in (-1, 1)$, $\operatorname{supp} \Phi = [-1, 1]$); an orthonormal or biorthonormal basis of spline–wavelets of degree r is being used. For example, consider the Battle–Lemarié spline–wavelets of degree r (see [Chui], [Daub], [W]). From the functional equations obtained for the orthonormal scaling function and wavelet in terms of shifts of the initial, non–orthogonal scaling function (see, e.g. [Daub, p. 147]), it is seen that the knots of both the scaling function and the wavelet are exactly all integers. (Recall that these wavelets do not have compact support, but their "tails" are exponentially light.) This ensures that a wavelet expansion whose finest level is j_1 is an r-th degree spline with knots at $2^{-j_1}\mathbb{Z}$, hence it is an r-th degree polynomial on each interval $[k_1 2^{-j_1}, (k_1 + 1)2^{-j_1}]$, $k_1 \in \mathbb{Z}$. On the other hand, $\Phi\left(\frac{x-t}{\varepsilon}\right)$ is an r-th degree spline, whose knots are concentrated in $[x - \varepsilon, x + \varepsilon]$ and whose number depends on r only. In order to compute $\widehat{h}_{+,\varepsilon}(x)$,

given x, the following steps are required:
1) sorting together, in increasing order, the knots of $\Phi\left(\frac{x-t}{\varepsilon}\right)$ and $\widehat{h}(t)$;
2) in each of the subintervals between two neighbouring points, obtained in step 1, $\Phi\left(\frac{x-t}{\varepsilon}\right)\ \widehat{h}(t)$ is a $2r$-th degree polynomial, whose zeros coincide with the zeros of $\widehat{h}(t)$ where $\Phi\left(\frac{x-t}{\varepsilon}\right)>0$, or fill the whole subinterval where $\Phi\left(\frac{x-t}{\varepsilon}\right)\equiv 0$. In each subinterval Δ where $\Phi\left(\frac{x-t}{\varepsilon}\right)>0$, the eventual roots of $\widehat{h}(t)=0$, $t\in\Delta$, are found. For $r=1,2,3,4$, the roots can be computed by closed formulae; for $r \geqq 5$ one may apply Sturm's theorem or Budan's theorem to count the roots in Δ. For $r\geqq 5$, if Δ contains root(s), one can apply additional approximate methods for sufficiently precise localization of the roots. But in practice, sufficiently precise approximate values of the roots can be deduced from the plotting data, and in this case the roots' abscissae are sorted together with the sorted array of abscissae obtained at step 1. The output of step 2 is an array of abscissae in increasing order, containing all entries at step 1 and the zeros of $\widehat{h}$.
3) In each of the subintervals between two neighbouring points obtained on step 2, $\Phi\left(\frac{x-t}{\varepsilon}\right)\ \widehat{h}_+(t)$ is a polynomial of degree $\leqq 2r$; the convolution integral in the definition of $\widehat{h}_{+,\varepsilon}$ is being decomposed in a sum over these subintervals; each of these integrals can be computed exactly by closed–form formulae, or, equivalently, by a quadrature formula which is accurate of order $2r+1$. In case that on step 2 only presence/absence of roots in each subinterval is being detected, on step 3 the integrals over subintervals containing roots can be solved approximately by a low–order quadrature formula as in the model in Theorem 3.3. It can be seen that if $\varepsilon \leqq C\ 2^{j_1(N)}$, where C does not depend on N, then the computational complexity of this method is $O_{\tau,C}(1)$ times that of $\widehat{h}$.

In principle, this method can be implemented for kernel estimators, too, but essential modifications are needed. We shall not discuss this version in details here.

4 Applications

The results in the previous section can be applied for concrete estimators to design optimal strategies of selecting ε,δ and δ_1, so that the quadrature approximation of $\hat{h}_{+,\varepsilon}$ has the same asymptotic rates as $\hat{h}_{+,\varepsilon}$ itself, in a broad class of error measures and function spaces simultaneously. Examples of such results can be obtained, e.g., by combining Theorems 3.2-4 with Corollaries 3.1.1-8 in [DM1] for wavelet estimators, and with Corollaries 3.2.1,2 in the same paper for kernel estimators. Some of these examples are considered below as model corollaries which can be used as patterns for deriving further results in this direction.

Corollary 4.1. *Assume that*
(4.1) $\quad h\in B^s_{pq}(\mathbb{R}),\ 1\leq p\leq\infty,\ 1\leq p_1\leq\infty,\ 1\leq q\leq\infty,\ \max\left\{\frac{1}{p}, r_0+\left(\frac{1}{p}-\frac{1}{p_1}\right)_+\right\}<s<r,$
where $0<r_0\leq 1,\quad r\in\mathbb{N}:\ r\geq 2,\quad r\neq 2+(\frac{1}{p}-\frac{1}{p_1})_+,\quad r_1=r;$
(4.2) $\quad \hat{h}^{(w)}$ *is the wavelet estimator from Theorem 1 in [DJ];*

(4.2′) $\hat{h}^{(k)}$ *is the variable-bandwidth kernel estimator from [LMS];*

(4.3) $\hat{h}_{+,\varepsilon,\delta,\delta_1} = \hat{h}^{(w)}_{+,\varepsilon,\delta,\delta_1}$ *or* $\hat{h}^{(k)}_{+,\varepsilon,\delta,\delta_1}$ *is the respective quadrature estimator from Theorem 3.3;*

(4.4) $0 < \varepsilon \le \varepsilon_N,\ 0 < \delta \le \delta_N,\ 0 < \delta_1 \le \delta_{1,N}$, *where the values of* ε_N, δ_N *and* $\delta_{1,N}$ *will be specified below;*

(4.5) $\theta = \frac{s}{p_1} - \frac{1}{2}\left(\frac{1}{p} - \frac{1}{p_1}\right)_+$;

(4.6) $\tau_1 = \tau_1(s,p,p_1,r_0,r) = \frac{[1+o(1)].\max\{A,[o(1)+1-\frac{1}{p_1}+B],[o(1)+1-\frac{1}{p_1}+B]\cdot r-A\}}{ABr}$, *as* $N \to \infty$;

(4.6′) $\tau = \tau_1(s,p,p_1,r_0,1)$;

(4.7) $A = A(p,p_1,r_0) = \max\left\{\frac{1}{p}, r_0 + \left(\frac{1}{p} - \frac{1}{p_1}\right)_+\right\}$;

(4.8) $B = B(s,p,p_1) = \min\left\{2, s - \left(\frac{1}{p} - \frac{1}{p_1}\right)_+\right\}$.

If, additionally:

(A) $\theta > 0,\ \varepsilon_N = O\left(N^{-\frac{s}{2s+1}\cdot\frac{1}{B}}\right),\ \delta_N = O\left(N^{-\frac{s}{2s+1}\cdot\tau}\right),\ \delta_{1,N} = O\left(N^{-\frac{s}{2s+1}\cdot\tau_1}\right)$,

then

(A1) $\left(E_{F^N}\|h - \hat{h}^{(k)}_{+,\varepsilon,\delta,\delta_1}\|^{p_1}_{L_{p_1}}\right)^{\frac{1}{p_1}} = O\left(N^{-\frac{s}{2s+1}}\right), 1 \le p_1 < \infty$;

(A2) *if, moreover,* $p \le p_1$,

then

$$\left(E_{F^N}\|h - \hat{h}^{(w)}_{+,\varepsilon,\delta,\delta_1}\|^2_{L_{p_1}}\right)^{\frac{1}{2}} = O\left(N^{-\frac{s}{2s+1}}\right);$$

(B) $\theta = 0$ *and if*

(B1) $\varepsilon_N = O\left\{\left[\left(\frac{\sqrt{\ln N}}{N}\right)^{\frac{s-\frac{1}{p}+\frac{1}{p_1}}{2(s-\frac{1}{p})+1}} \cdot (\ln N)^{\frac{1}{p_1}}\right]^{\frac{1}{B}}\right\}$,

$$\delta_N = O\left\{\left[\left(\frac{\sqrt{\ln N}}{N}\right)^{\frac{s-\frac{1}{p}+\frac{1}{p_1}}{2(s-\frac{1}{p})+1}} \cdot (\ln N)^{\frac{1}{p_1}}\right]^{\tau}\right\},$$

$$\delta_{1,N} = O\left\{\left[\left(\frac{\sqrt{\ln N}}{N}\right)^{\frac{s-\frac{1}{p}+\frac{1}{p_1}}{2(s-\frac{1}{p})+1}} \cdot (\ln N)^{\frac{1}{p_1}}\right]^{\tau_1}\right\},$$

then

$$\left(E_{F^N}\|h - \hat{h}^{(k)}_{+,\varepsilon,\delta,\delta_1}\|^{p_1}_{L_{p_1}}\right)^{\frac{1}{p_1}} = O\left[\left(\frac{\sqrt{\ln N}}{N}\right)^{\frac{s-\frac{1}{p}+\frac{1}{p_1}}{2(s-\frac{1}{p})+1}} \cdot (\ln N)^{\frac{1}{p_1}}\right], \quad 1 \le p_1 < \infty;$$

(B2) $p \le p_1$,

$$\varepsilon_N = O\left\{\left[\left(\frac{\ln N}{N}\right)^{\frac{s-\frac{1}{p}+\frac{1}{p_1}}{2(s-\frac{1}{p})+1}} \cdot (\ln N)^{\left(1-\frac{2p}{qp_1}\right)_+}\right]^{\frac{1}{B}}\right\},$$

$$\delta_N = O\left\{\left[\left(\frac{\ln N}{N}\right)^{\frac{s-\frac{1}{p}+\frac{1}{p_1}}{2(s-\frac{1}{p})+1}} \cdot (\ln N)^{\left(1-\frac{2p}{qp_1}\right)_+}\right]^{\tau}\right\},$$

$$\delta_{1,N} = O\left\{\left[\left(\frac{\ln N}{N}\right)^{\frac{s-\frac{1}{p}+\frac{1}{p_1}}{2(s-\frac{1}{p})+1}} \cdot (\ln N)^{\left(1-\frac{2p}{qp_1}\right)_+}\right]^{\tau_1}\right\},$$

then

$$\left(E_{F^N}\|h-\hat{h}^{(w)}_{+,\varepsilon,\delta,\delta_1}\|^2_{L_{p_1}}\right)^{\frac{1}{2}}=O\left[\left(\frac{\ln N}{N}\right)^{\frac{s-\frac{1}{p}+\frac{1}{p_1}}{2(s-\frac{1}{p})+1}}\cdot(\ln N)^{\left(1-\frac{2p}{qp_1}\right)_+}\right];$$

(C) $\theta<0$ *and if*

(C1) $\varepsilon_N=O\left[\left(\frac{\sqrt{\ln N}}{N}\right)^{\frac{s-\frac{1}{p}+\frac{1}{p_1}}{2(s-\frac{1}{p})+1}\cdot\frac{1}{B}}\right]$,

$$\delta_N=O\left[\left(\frac{\sqrt{\ln N}}{N}\right)^{\frac{s-\frac{1}{p}+\frac{1}{p_1}}{2(s-\frac{1}{p})+1}\cdot\tau}\right],$$

$$\delta_{1,N}=O\left[\left(\frac{\sqrt{\ln N}}{N}\right)^{\frac{s-\frac{1}{p}+\frac{1}{p_1}}{2(s-\frac{1}{p})+1}\cdot\tau_1}\right],$$

then

$$\left(E_{F^N}\|h-\hat{h}^{(k)}_{+,\varepsilon,\delta,\delta_1}\|^{p_1}_{L_{p_1}}\right)^{\frac{1}{p_1}}=O\left[\left(\frac{\sqrt{\ln N}}{N}\right)^{\frac{s-\frac{1}{p}+\frac{1}{p_1}}{2(s-\frac{1}{p})+1}}\right],\quad 1\le p_1<\infty;$$

(C2) $p\le p_1$,

$$\varepsilon_N=O\left[\left(\frac{\ln N}{N}\right)^{\frac{s-\frac{1}{p}+\frac{1}{p_1}}{2(s-\frac{1}{p})+1}\cdot\frac{1}{B}}\right],$$

$$\delta_N=O\left[\left(\frac{\ln N}{N}\right)^{\frac{s-\frac{1}{p}+\frac{1}{p_1}}{2(s-\frac{1}{p})+1}\cdot\tau}\right],$$

$$\delta_{1,N}=O\left[\left(\frac{\ln N}{N}\right)^{\frac{s-\frac{1}{p}+\frac{1}{p_1}}{2(s-\frac{1}{p})+1}\cdot\tau_1}\right],$$

then

$$\left(E_{F^N}\|h-\hat{h}^{(w)}_{+,\varepsilon,\delta,\delta_1}\|^2_{L_{p_1}}\right)^{\frac{1}{2}}=O\left[\left(\frac{\ln N}{N}\right)^{\frac{s-\frac{1}{p}+\frac{1}{p_1}}{2(s-\frac{1}{p})+1}}\right].$$

Remark 4.1. *In all cases considered in Corollary 4.1* $\hat{h}^{(k)}_{+,\varepsilon,\delta,\delta_1}$ *and* $\hat{h}^{(w)}_{+,\varepsilon,\delta,\delta_1}$ *attain the same asymptotic rates as* $\hat{h}^{(k)}$, $\hat{h}^{(k)}_{+,\varepsilon}$ *and* $\hat{h}^{(w)}$, $\hat{h}^{(w)}_{+,\varepsilon}$, *respectively. This means that for* $\theta>0$ *all these estimators attain the asymptotic-minimax rates, and for* $\theta\le 0$ *they all are asymptotic-minimax within the "asymptopia" paradigm (see [DJKP]).*

The choice of ε_N, δ, δ_1 *in Corollary 4.1 depends on the explicit knoledge of* p_1,p,q *and* s. *Of these, only* p_1 *is "a priori" known, in general. The next statement suggests (smaller) values for* $\varepsilon_N=\varepsilon_N(p_1)$, $\delta_N=\delta_N(p_1)$ *and* $\delta_{1,N}=\delta_{1,N}(p_1)$, *so that the conclusions of Corollary 4.1 hold simultaneously for the whole range of admissible* p,q *and* s.

Corollary 4.2. *Assume that*

(4.9) $p_1:1\le p_1\le\infty;\quad 0<r_0\le 1;\quad r\in\mathbb{N},\quad r\ge 2;\quad r_1\in\mathbb{N},$
$1\le r_1\le r;\quad s_0:r_0\le s_0\le 1;\quad \sigma:r_0\le\sigma<r;$

(4.10) $\varepsilon_N=O\left[N^{-\frac{r}{r_0(2r+1)}}\right];$

(4.11) $\delta_N = O\left[N^{-\frac{r}{r_0(2r+1)}\cdot(r_0+1-\frac{1}{p_1})\cdot\max\left\{\frac{1}{s_0}-\frac{1}{o(1)+r_0+1-\frac{1}{p_1}},1\right\}}\right]$;

(4.12) $\delta_{1,N} = O\left[N^{-\frac{r}{r_0(2r+1)}\cdot(r_0+1-\frac{1}{p_1})\cdot\max\left\{\frac{1}{\sigma}-\frac{1}{o(1)+r_0+1-\frac{1}{p_1}},\frac{1}{r_1}\right\}}\right]$.

Then,

(A) $\hat{h}^{(k)}_{+,\varepsilon,\delta,\delta_1}$ *is asymptotic-minimax optimal (within the "asymptopia" paradigm) in the risk norm* $\mathcal{L}_{p_1}(L_{p_1})$, $1 \le p_1 < \infty$, *uniformly in:*

(4.13) $p : 1 \le p \le \infty, \quad \max\left\{\frac{1}{p}, r_0 + \left(\frac{1}{p}-\frac{1}{p_1}\right)_+\right\} \le \min\{s_0,\sigma\};$

(4.14) $q : 1 \le q \le \infty, \quad \max\{s_0,\sigma\} < s < r;$

(B) $\hat{h}^{(w)}_{+,\varepsilon,\delta,\delta_1}$ *is also such in* $\mathcal{L}_2(L_{p_1})$, *uniformly in* q, s *as in (4.14) and*

(4.13′) $p : 1 \le p \le p_1, \quad \max\left\{\frac{1}{p}, r_0 + \frac{1}{p}-\frac{1}{p_1}\right\} \le \min\{s_0,\sigma\}.$

The following corollary gives (smaller) values of δ_N and $\delta_{1,N}$, so that asymptotic-minimax optimality is preserved uniformly in p_1, too.

Corollary 4.3. *Under the assumptions of (4.9) for* r_0, r, s_0 *and* σ, *suppose that* ε_N *is given by (4.10) and*

(4.11′) $\delta_N = O\left[N^{-\frac{r}{r_0(2r+1)}\cdot(r_0+1)\cdot\max\left\{\frac{1}{s_0}-\frac{1}{o(1)+r_0+1},1\right\}}\right]$;

(4.12′) $\delta_{1,N} = O\left[N^{-\frac{r}{r_0(2r+1)}\cdot(r_0+1)\cdot\max\left\{\frac{1}{\sigma}-\frac{1}{o(1)+r_0+1},\frac{1}{r_1}\right\}}\right]$.

Then, the conclusions (A,B) of Corollary 4.2 hold, uniformly in $p_1 : 1 \le p_1 < \infty$ *for* $\hat{h}^{(k)}_{+,\varepsilon,\delta,\delta_1}$ *and* $1 \le p_1 \le \infty$ *for* $\hat{h}^{(w)}_{+,\varepsilon,\delta,\delta_1}$ *respectively.*

From (4.11,12) it is seen that the "critical" choices of s_0 and σ, when the rate "switches orders", are:

(4.15) $s_0 = 1 + \dfrac{1}{o(1)+r_0+1-\frac{1}{p_1}}, \quad \sigma = r_1 + \dfrac{1}{o(1)+r_0+1-\frac{1}{p_1}},$

which are exactly the respective values of s_0^* and σ^*, computed at $s = r$ and $p_1 = \infty$.

The most important particular choice of r_1 is $r_1 = r$.

The above corollaries follow from Theorem 3.3. The results about δ_N and $\delta_{1,N}$ are the same in the respective corollaries from Theorems 3.2 and 3.4.

Remark 4.2. *The restriction* $p \le p_1$ *for* $\hat{h}^{(w)}$ *in Corollaries 4.1-3 comes from the formulation of Theorem 1 in [DJ]. We make the conjecture that an extension of this theorem can be proved without this restriction, thereby relaxing the assumptions about* p_1 *for the wavelet estimator in Corollaries 4.1-3.*

5 Numerical methods

In this section we discuss which classes of quadrature formulae are most appropriate to apply in our context. Before all, Hermit–Birkhoff and Padé – type quadratures are inadmissible because they involve values of derivatives, while we would like to achieve simultaneous asymptotic–minimax estimation in Besov spaces for a broad range of p, q, s and, in particular, including values $s < 1$. This limits the considerations to essentially 3 large classes of quadrature formulae:
1) Newton–Cotes quadratures;
2) Gaussian quadratures;
3) Quadrature processes, based on Runge principle (Richardson extrapolation). (For references on quadratures, we refer to [SP], [BZ], [R].)

Newton–Cotes quadratures seem to be the most appropriate ones for computation of $\widehat{h}_{+,\varepsilon,\delta}$, as considered in Theorem 3.2. As a concrete quadrature formula for approximate computation of $\int_a^b f(t)\, dt$ in this theorem we suggest "the formula of rectangles" ($n = 1$, $A_1 = b - a$, $t_1 = \frac{a+b}{2}$); or "the trapezoid rule" ($n = 2$, $A_1 = A_2 = \frac{b-a}{2}$, $t_1 = a$, $t_2 = b$). (The former is also the simplest Gaussian formula.) Both of them are accurate of order 2 (in our notations), while in Theorem 3.2 only accuracy of order 1 is required, but the above formulae provide additional local symmetry which is useful to have in excess when N is not large, for better visual performance of the estimator. The same quadrature formulae are recommended for numerical integration on the intervals $[(k + (\nu - 1)\delta)\varepsilon,\ (k + \nu\delta)\varepsilon]$, $k \in \Lambda$, in Theorem 3.3.

In the numerical integration over the intervals $[(k_1 + (\nu_1 - 1)\delta_1)\varepsilon,\ (k_1 + \nu_1\delta_1)\varepsilon]$, $k_1 \in \Lambda_1$ in Theorem 3.3 and for all intervals in Theorem 3.4 quadrature formulae which are accurate of order r_1 are needed. Here Newton–Cotes formulae compete with Gaussian ones. Both of them have advantages and disadvantages of their own. For example, let $r_1 = 4 \leqq r$. Then, a strong candidate from the Newton–Cotes class is the well–known "Simpson's rule" ($n = 3$, $A_1 = A_3 = \frac{b-a}{6}$, $A_2 = \frac{2}{3}(b - a)$, $t_1 = a$, $t_2 = \frac{a+b}{2}$, $t_3 = b$). An equally strong candidate from the Gaussian class is the formula with $n = 2$, $A_1 = A_2 = \frac{b-a}{2}$, $t_1 = \frac{1}{2}\left(1 + \frac{1}{\sqrt{3}}\right) a + \frac{1}{2}\left(1 - \frac{1}{\sqrt{3}}\right) b$, $t_2 = \frac{1}{2}\left(1 - \frac{1}{\sqrt{3}}\right) a + \frac{1}{2}\left(1 + \frac{1}{\sqrt{3}}\right) b$. The two formulae are both accurate of order $r_1 = 4$; the Gaussian one uses values at 2 points, while the Simpson's rule – at 3; the error of the Gaussian formula is slightly smaller. However, Simpson's rule uses values on a uniform mesh, and in practice should be the preferred one in most cases.

Unfortunately, with the increase of r_1, the numerical performance of both Newton–Cotes and Gaussian quadratures gets worse. One of the problems with Newton–Cotes rules is that the absolute values of their coefficients $A_k(r_1)$ increase quickly with the increase of r_1. It should be noted also that for higher values of r_1, some of the $A_{\mu k}$–coefficients of the Newton–Cotes formulae may be negative, i.e., (3.20) fails. With Gaussian quadratures the $A_{\mu k}$–coefficients are always positive. The problem with Gaussian quadratures is different: its knots are the zeros of orthogonal polynomials (in our case, Legendre polynomials) of order $\left[\frac{r_1}{2}\right]+1$; with the increase of r_1 these zeros (which are irrational reals) have to be computed with ever increasing precision; the volume of computations thus rises very quickly. For these reasons, if we use Newton–Cotes or Gaussian rules in the numerical integration over the intervals $[(k_1+(\nu_1-1)\delta_1)\varepsilon,\ (k_1+\nu_1\delta_1)\varepsilon],\ k_1\in\Lambda_1$ in Theorem 3.3 or in the context of Theorem 3.4, and if r is large, we cannot afford to take the most desirable value of $r_1:\ r_1=r$, but instead have to content ourselves with an intermediate value between 1 and r, which slows down the computation of the asymptotic–minimax $\widehat{h}_{+,\varepsilon,\mathrm{quadr}}$.

Fortunately, these problems can be completely overcome by using quadrature processes (i.e., iterative methods for generating the knots and coefficients of the quadrature formula). To ensure that the quadrature process converges to the exact value of the integral with sufficiently high order of accuracy r_1 (presumably, $r_1=r$, our general recommendation is to repeatedly apply Aitken–Steffensen's method of accelerating the convergence of the iterative process. This approach results in a procedure which cab be found under different names in various reference sources on numerical integration, e.g., as Runge principle or Richardson extrapolation. Various types of Richardson extrapolation techniques can be used in our context but, to be concrete, we suggest Romberg integration with the "trapezoid rule" as initial formula for the iterative process. Thanks to exploiting the structure of the remainder in the Euler–McLaurin summation formula, the recursive formulae for this particular type of Richardson extrapolation are very simple. In brief, given the order of accuracy r_1, and assuming that f has integrable derivative of even order not less than r_1, $\int_a^b f(x)\,dx$ can be written as

$$(5.1)\quad \int_a^b f(x)\,dx = \frac{b-a}{\mu}\left(\frac{1}{2}f_0+\sum_{i=1}^{\mu-1}f_i+\frac{1}{2}f_\mu\right)+\sum_{i=1}^{\nu}a_i\mu^{-2i}+R_\mu\mu^{-2(\nu+1)},$$

where

$$\nu=\left[\frac{r_1+1}{2}\right]-1,$$

$$(5.2)\quad f_i=f\left(a+i\frac{b-a}{\mu}\right),\ i=0,1,\dots,\mu,$$

$$(5.3)\quad a_i=(b-a)^{2i}\,.\,B_{2i}\,.\,[f^{(2i-1)}(b)-f^{(2i-1)}(a)]/(2i)!$$

$$(5.4)\quad R_\nu=\frac{(b-a)^{2\nu+2}\mu^{-2\nu-2}}{(2\nu+2)!}\int_0^{b-a}B_{2\nu+2}^{\mathrm{per}}\left(\frac{\mu x}{b-a}\right)f^{(2\nu+2)}(a+x)\,dx,$$

B_l is the l–th Bernoulli number, $B_l^{\mathrm{per}}(x)$ is the 1–periodic version of the Bernoulli polynomials:

$$(5.5)\quad B_l^{\mathrm{per}}(x)=\begin{cases}B_l(x), & 0\leqq x\leqq 1\\ B_l^{\mathrm{per}}(x-1), & x>1\\ B_l^{\mathrm{per}}(x+1), & x<0\end{cases}$$

Note that the expression containing $f_0,f_1,\dots,f_\mu$ in (5.1) is exactly the composite quadrature formula obtained by applying the "trapezoid rule" to every interval of the form $\left[a+(i-1)\frac{b-a}{\mu},\ a+i\frac{b-a}{\mu}\right]$, $i=1,\dots,\mu$.

For $\mu = 2^n$, $n = 0, 1, 2, \dots$, consider

$$(5.6) \quad I_{0,n} = (b-a)\, 2^{-n} \left(\frac{1}{2} f_0 + \sum_{i=1}^{2^n-1} f_i + \frac{1}{2} f_{2^n} \right),$$

$$(5.7) \quad I_{\nu,n} = \frac{1}{2^{2\nu}-1} \left(2^{2\nu} I_{\nu-1,\, n+1} - I_{\nu-1,\, n} \right).$$

It can be shown that the iterative computation of $I_{\nu,\, n}$ is equivalent to iterative computation of the coefficients and knots of a quadrature formula, starting with (5.6). Computing $I_{\nu,\, n}$ is equivalent to eliminating the coefficients a_i in (5.1) and the resulting quadrature formula for $I_{\nu,\, n}$ is accurate of order $2\nu+2:\ r_1 \leqq 2\nu+2 \leqq r_1+1$. The organization of the computation is usually the following: first I_{00} is being computed, then $I_{0,1}$, then I_{10} via (5.7), etc. On the ν–th step, first $I_{0,\, \nu}$ is being computed via (5.6), and then $I_{k,\, \nu-k}$, $k = 1, 2, \dots, \nu$ are being calculated consecutively via (5.7). The procedure stops at $I_{n,0}$, $n = \left[\frac{r_1+1}{2}\right] - 1$, which is accepted as the value of the quadrature formula for $\int_a^b f(x)\, dx$. An additional advantage is that the values of f used in the computation of $I_{0,\nu}$ can be used in the computation of $I_{0,\nu+1}$, too.

We recommend this type of Romberg integration as the algorithm for numerical integration in all cases discussed in this paper. Here are some of our arguments in its favour:

- this algorithm works well for both small and large values of r_1. This allows to always choose the most favourable value of $r_1:\ r_1 = r$ in the models of Theorems 3.3, 4, no matter how big r is.
- it works equally well with Daubechies wavelets and spline–wavelets and, in fact, with any kind of dyadic wavelets, since it uses only values of f involved in iterative refinement via the dyadic functional equation whose fundamental interpolant the wavelet f is.
- essentially, this is a *dyadic pyramidal algorithm* which can in certain sense be considered as an extension of Mallat's pyramidal algorithms "on subcoefficient level".
- if the quadrature formula in Theorem 3.2 and for numerical integration on intervals in I in Theorem 3.3 (see (3.46)) is chosen to be the "trapezoid rule", then all numerical integrations are being done via a unified algorithm. For example, in the model considered in Theorem 3.3, if $\Delta_\mu \in I$, then 0 iterations of Romberg integration are being carried out (the "trapezoid rule"); if $\Delta_\mu \in I_1$, then (for $r_1 = r$) $\left[\dfrac{r+1}{2}\right] - 1$ iterations of Romberg integration are being performed.
- it can be proved that the coefficients in the quadrature formula corresponding to $I_{\nu,\, n}$ are non–negative for any ν, n.

6 Some numerical examples

7 Proofs

Proof of Proposition 2.1. Let us prove the proposition for g. The proof for f is analogous.

By the definition $\widehat{g}_{+,\varepsilon}$, $\widehat{g}_+$ and $\widehat{g}$, after changing the order of summation and integration, and using also the definition of GKE,

$$(7.1)\quad \widehat{g}_{+,\varepsilon}(x) = \varepsilon^{-d} \int\limits_{\mathrm{I\!R^d}} \Phi\left(\frac{x-\theta}{\varepsilon}\right) \widehat{g}_+(\theta)\, d\theta =$$

$$= \varepsilon^{-d} \int\limits_{\mathrm{I\!R^d}} \Phi\left(\frac{x-\theta}{\varepsilon}\right) \widehat{g}_+(\theta)\, I_{\{\xi\in\mathrm{I\!R^d}\,:\,\widehat{\mathrm{g}}(\xi)>0\}}(\theta)\, d\theta =$$

$$= \varepsilon^{-d} \int\limits_{\mathrm{I\!R^d}} \Phi\left(\frac{x-\theta}{\varepsilon}\right) \left[\frac{1}{N}\sum_{\nu=1}^{N} Y_\nu \,.\, K(\theta, X_\nu; X_1, \dots, X_N)\right] . \, I_{\{\xi\in\mathrm{I\!R^d}\,:\,\widehat{\mathrm{g}}(\xi)>0\}}(\theta)\, d\theta =$$

$$= \frac{1}{N}\sum_{\nu=1}^{N} Y_\nu \,.$$

$$. \left[\frac{1}{\varepsilon^d}\int\limits_{\mathrm{I\!R^d}} \Phi\left(\frac{x-\theta}{\varepsilon}\right) . \, I_{\{\xi\in\mathrm{I\!R^d}\,:\,\widehat{\mathrm{g}}(\xi)>0\}}(\theta) \,.\, K(\theta, X_\nu; X_1, \dots, X_N)\, d\theta\right] =$$

$$= \frac{1}{N}\sum_{\nu=1}^{N} Y_\nu \,.\, \left[I_{\{\xi\in\mathrm{I\!R^d}\,:\,\mathrm{h}(\xi)>0\}}(\cdot)\; K(\cdot, t; X_1, \dots, X_N)\right]_\varepsilon \Big|_{t=X_\nu} =$$

$$= \frac{1}{N}\sum_{\nu=1}^{N} Y_\nu \,.\, K_\varepsilon(x, X_\nu; X_1, \dots, X_N) \implies (2.17).$$

□

Proof of Proposition 3.1. The proof of (3.22) is analogous to the proof of (2.17) in Proposition 2.1. The proof of (3.23) follows from the definition of $\widehat{f}_{+,\varepsilon,\mathrm{quadr}}$ ($\widehat{g}_{+,\varepsilon,\mathrm{quadr}}$) and $A_{\mu k} \geqq 0$ in (3.20). □

Proof of Theorem 3.1. We give the proof for $\widehat{g}$; the one for $\widehat{f}$ is analogous.

$$(7.2)\quad \widehat{g}_{+,\varepsilon,\mathrm{quadr}}(x) = (\widehat{g}_+)_{\varepsilon,\mathrm{quadr}} =$$

$$= \frac{1}{N}\sum_{\nu=1}^{N} Y_\nu \,.$$

$$. \left[\frac{1}{\varepsilon}\sum_{\mu=-\infty}^{\infty}\sum_{k=1}^{n_\mu} A_{\mu k}\, \Phi\left(\frac{x-t_{\mu k}}{\varepsilon}\right) . \, I_{\{\xi\in\mathrm{I\!R}\,:\,\widehat{\mathrm{g}}(\xi)>0\}}(t_{\mu k}) \,.\, K(t_{\mu k}, X_\nu; X_1, \dots, X_N)\right]$$

where only a finite number of summands in $\sum\limits_{\mu}\sum\limits_{k}$ is non–zero. Namely, $\Phi\left(\frac{x-t_{\mu k}}{\varepsilon}\right)$ is non–zero for only those $t_{\mu k}$ for which $|x - t_{\mu k}| < \varepsilon$.

For every $x_0 \in \mathbb{R}$ consider $I_{x_0,\ \varepsilon} \subset \mathbb{Z} \times \mathbb{N}$, defined by

(7.3) $\quad I_{x_0,\ \varepsilon} = \{(\mu, k) :\ |x_0 - t_{\mu k}| \leqq \varepsilon\}.$

Observe that, for C_0 defined in (3.21),

(7.4) $\quad \forall x :\ |x - x_0| \leqq C_0/2 \implies I_{x,\ \varepsilon} = I_{x_0,\ \varepsilon}\ .$

Therefore, in a $C_0\ \varepsilon/2$ – neighbourhood of any $x_0 \in \mathbb{R}$ $(\widehat{g}_+)_{\varepsilon,\text{quadr}}$ is the same finite linear combination of translates $\Phi\left(\frac{x-t_{\mu k}}{\varepsilon}\right)$, hence

(7.5) $$\widehat{g}_{+,\varepsilon,\text{quadr}}\Big|_{[x_0-\varepsilon C_0/2,\ x_0+\varepsilon C_0/2]} \in C^r([x_0 - \varepsilon C_0/2,\ x_0 + \varepsilon C_0/2]), \quad \forall x_0 \in \mathbb{R},$$

which easily implies

(7.6) $\quad \widehat{g}_{+,\varepsilon,\text{quadr}} \in C^r(\mathbb{R}) \implies (3.24).$

If $\operatorname{supp} \widehat{g} \subset [-R, R]$ for some $R :\ 0 < R < \infty$, then $\operatorname{supp} \Phi = [-1, 1]$ and (7.2) imply

(7.7) $\quad \operatorname{supp} \widehat{g}_{+,\varepsilon,\text{quadr}} \subset [R - \varepsilon, R + \varepsilon] \quad \implies (3.24').$ $\square$

Proof of Theorem 3.2. For ρ_1 as in (3.15),

(7.8) $$(E_{F^N}\|h - \widehat{h}_{+,\varepsilon,\delta}\|^{\rho}_{L_{p_1}})^{1/\rho_1} \leqq (E_{F^N}\|h - \widehat{h}_{+,\varepsilon}\|^{\rho}_{L_{p_1}})^{1/\rho_1} + (E_{F^N}\|\widehat{h}_{+,\varepsilon} - \widehat{h}_{+,\varepsilon,\delta}\|^{\rho}_{L_{p_1}})^{1/\rho_1} \leqq$$
$$\leqq 2(3 + 2^{\rho/\rho_1})\ .\ K^{1/\rho_1}\ .\ \alpha\left(\frac{1}{N}\right)^{1/\rho_1} + (E_{F^N}\|\widehat{h}_{+,\varepsilon} - \widehat{h}_{+,\varepsilon,\delta}\|^{\rho}_{L_{p_1}})^{1/\rho_1}\ .$$

We shall carry out the proof in details for $1 \leqq p_1 < \infty$. For $p_1 = \infty$ the proof is analogous (but simpler).

(7.9) $$\|\widehat{h}_{+,\varepsilon} - \widehat{h}_{+,\varepsilon,\delta}\|_{L_{p_1}(\mathbb{R})} =$$
$$= \left(\int\limits_{-\infty}^{\infty} \left|\frac{1}{\varepsilon} \sum_{\mu=-\infty}^{\infty} \left[\int\limits_{\Delta_\mu} \Phi\left(\frac{x-t}{\varepsilon}\right) \widehat{h}_+(t)\ dt - \sum_{\nu=1}^{n_\mu} A_{\mu\nu}\ \Phi\left(\frac{x - t_{\mu\nu}}{\varepsilon}\right) \widehat{h}_+(t_{\mu\nu})\right]\right|^{p_1} dx\right)^{1/p_1} =:$$
$$=:\ T$$

By the accuracy of the quadrature formula of order 1,

(7.10) $$\sum_{\nu=1}^{n_\mu} A_{\mu\nu} = \delta\varepsilon = \int\limits_{\Delta_\mu} dt\ .$$

Hence,

(7.11) $$T_2 := \int\limits_{\Delta_\mu} \Phi\left(\frac{x-t}{\varepsilon}\right) \widehat{h}_+(t)\ dt - \sum_{\nu=1}^{n_\mu} A_{\mu\nu}\ \Phi\left(\frac{x-t_{\mu\nu}}{\varepsilon}\right) \widehat{h}_+(t_{\mu\nu}) =$$
$$= \frac{1}{\delta\varepsilon} \int\limits_{\Delta_\mu} \sum_{\nu=1}^{n_\mu} A_{\mu\nu} \left(\Phi\left(\frac{x-t}{\varepsilon}\right) \widehat{h}_+(t) - \Phi\left(\frac{x-t_{\mu\nu}}{\varepsilon}\right) \widehat{h}_+(t_{\mu\nu})\right) dt =$$
$$= \frac{1}{\delta\varepsilon} \int\limits_{\Delta_\mu} \left\{ \sum_{\nu=1}^{n_\mu} A_{\mu\nu} \left[\Phi\left(\frac{x-t}{\varepsilon}\right) \left(\widehat{h}_+(t) - \widehat{h}_+(t_{\mu\nu})\right) + \right.\right.$$

$$+ \left(\Phi\left(\frac{x-t}{\varepsilon}\right) - \Phi\left(\frac{x-t_{\mu\nu}}{\varepsilon}\right) \right) \widehat{h}_+(t_{\mu\nu}) \Bigg] \Bigg\} \, dt$$

By (7.9, 11), the generalized and the "standard" Minkowski's inequality, a change of variable in Δ_μ yields, in view of the translation–invariance of $L_{p_1}(\mathbb{R})$, the definition of local, integral and average moduli of smoothness (3.32, 20) and (7.10),

$$(7.12) \quad T_1 = \left(\int\limits_{-\infty}^{\infty} \left| \frac{1}{\varepsilon} \sum_{\mu=-\infty}^{\infty} \frac{1}{\delta\varepsilon} \int\limits_{\Delta_\mu} \left\{ \sum_{\nu=1}^{n_\mu} A_{\mu\nu} \Bigg[\Phi\left(\frac{x-t}{\varepsilon}\right) (\widehat{h}_+(t) - \widehat{h}_+(t_{\mu\nu})) + \right.\right.$$

$$\left.\left. + \left(\Phi\left(\frac{x-t}{\varepsilon}\right) - \Phi\left(\frac{x-t_{\mu\nu}}{\varepsilon}\right) \right) \widehat{h}_+(t_{\mu\nu}) \Bigg] \right\} dt \right|^{p_1} dx \right)^{1/p_1} \leqq$$

$$\leqq \frac{1}{\varepsilon} \sum_{\mu=-\infty}^{\infty} \frac{1}{\delta\varepsilon} \int\limits_{\Delta_\mu} \left\{ \sum_{\nu=1}^{n_\mu} |A_{\mu\nu}| \left[\int\limits_{-\infty}^{\infty} \left| \Phi\left(\frac{x-t}{\varepsilon}\right) (\widehat{h}_+(t) - \widehat{h}_+(t_{\mu\nu})) + \right.\right.\right.$$

$$\left.\left.\left. + \left(\Phi\left(\frac{x-t}{\varepsilon}\right) - \Phi\left(\frac{x-t_{\mu\nu}}{\varepsilon}\right) \right) \widehat{h}_+(t_{\mu\nu}) \right|^{p_1} dx \right)^{1/p_1} \right] \right\} dt =$$

$$= \frac{1}{\delta\varepsilon^2} \sum_{\mu=-\infty}^{\infty} \int\limits_{-\delta\varepsilon/2}^{\delta\varepsilon/2} \left[\sum_{\nu=1}^{n_\mu} |A_{\mu\nu}| \left(\int\limits_{-\infty}^{\infty} \left| \Phi\left(\frac{x-(\mu+1/2)\delta\varepsilon-\theta}{\varepsilon}\right) \cdot \right.\right.\right.$$

$$\cdot \left(\widehat{h}_+\left(\left(\mu+\tfrac{1}{2}\right)\delta\varepsilon+\theta\right) - \widehat{h}_+(t_{\mu\nu}) \right) +$$

$$\left.\left.\left. + \left(\Phi\left(\frac{x-(\mu+1/2)\delta\varepsilon-\theta}{\varepsilon}\right) - \Phi\left(\frac{x-t_{\mu\nu}}{\varepsilon}\right) \right) \widehat{h}_+(t_{\mu\nu}) \right|^{p_1} dx \right)^{1/p_1} \right] d\theta \leqq$$

$$\leqq \frac{1}{\delta\varepsilon^2} \sum_{\mu=-\infty}^{\infty} \int\limits_{-\delta\varepsilon/2}^{\delta\varepsilon/2} \left\{ \sum_{\nu=1}^{n_\mu} |A_{\mu\nu}| \left[\left| \widehat{h}_+\left(\left(\mu+\frac{1}{2}\right)\delta\varepsilon+\theta\right) - \widehat{h}_+(t_{\mu\nu}) \right| \cdot \right.\right.$$

$$\cdot \left\| \Phi\left(\frac{\cdot - (\mu+1/2)\delta\varepsilon-\theta}{\varepsilon} \right) \right\|_{L_{p_1}} +$$

$$\left.\left. + |\widehat{h}_+(t_{\mu\nu})| \cdot \left\| \Phi\left(\frac{\cdot - (\mu+1/2)\delta\varepsilon-\theta}{\varepsilon}\right) - \Phi\left(\frac{\cdot - t_{\mu\nu}}{\varepsilon}\right) \right\|_{L_{p_1}} \right] \right\} d\theta =$$

$$= \frac{1}{\delta\varepsilon^2} \cdot \left\| \Phi\left(\frac{\cdot}{\varepsilon}\right) \right\|_{L_{p_1}} \cdot$$

$$\cdot \sum_{\mu=-\infty}^{\infty} \sum_{\nu=1}^{n_\mu} |A_{\mu\nu}| \int\limits_{-\delta\varepsilon/2}^{\delta\varepsilon/2} \left| \widehat{h}_+\Big((\mu+1/2)\delta\varepsilon + \theta\Big) - \widehat{h}_+(t_{\mu\nu}) \right| d\theta +$$

$$+ \frac{1}{\delta\varepsilon^2} \cdot \sum_{\mu=-\infty}^{\infty} \sum_{\nu=1}^{n_\mu} |A_{\mu\nu}| \, . \, |\widehat{h}_+(t_{\mu\nu})| \cdot$$

$$\cdot \int\limits_{-\delta\varepsilon/2}^{\delta\varepsilon/2} \left\| \Phi\left(\frac{\cdot + t_{\mu\nu} - (\mu+1/2)\delta\varepsilon-\theta}{\varepsilon}\right) - \Phi\left(\frac{\cdot}{\varepsilon}\right) \right\|_{L_{p_1}} d\theta =$$

$$\leqq \frac{1}{\varepsilon} \left\| \Phi\left(\frac{\cdot}{\varepsilon}\right)\right\|_{L_{p_1}} \sum_{\mu=-\infty}^{\infty} \int\limits_{-\delta\varepsilon/2}^{\delta\varepsilon/2} \omega_1(\widehat{h}_+,\ (\mu+1/2)\delta\varepsilon \ + \ \theta_1;\ 2\delta\varepsilon)\ d\theta_1 \ +$$

$$+ \frac{1}{\varepsilon}\,\omega_1\left(\Phi\left(\frac{\cdot}{\varepsilon}\right); \delta\varepsilon\right)_{L_{p_1}} \cdot \sum_{\mu=-\infty}^{\infty} \sum_{\nu=1}^{n_\mu} |A_{\mu\nu}| \,.\, |\widehat{h}_+(t_{\mu\nu})| \ \leqq$$

$$= \frac{1}{\varepsilon}\left[\left\| \Phi\left(\frac{\cdot}{\varepsilon}\right)\right\|_{L_{p_1}} \int\limits_{-\infty}^{\infty} \omega_1(\widehat{h}_+, t; 2\delta\varepsilon)\ dt \ + \ T_3 \ .\ \omega_1\left(\Phi\left(\frac{\cdot}{\varepsilon}\right); \delta\varepsilon\right)_{L_{p_1}}\right] \ =$$

$$= \frac{1}{\varepsilon}\left[\left\| \Phi\left(\frac{\cdot}{\varepsilon}\right)\right\|_{L_{p_1}} .\ \tau_1(\widehat{h}_+; 2\delta\varepsilon)_{L_1} \ + \ T_3 \ .\ \omega_1\left(\Phi\left(\frac{\cdot}{\varepsilon}\right); \delta\varepsilon\right)_{L_{p_1}}\right],$$

where

(7.13) $$T_3 \ = \ \sum_{\mu=-\infty}^{\infty} \sum_{\nu=1}^{n_\mu} |A_{\mu\nu}| \,.\, |\widehat{h}_+(t_{\mu\nu})| \ .$$

Let us evaluate T_3. First of all, clearly for any $t \in \Delta_\mu$,

(7.14) $$|\widehat{h}_+(t_{\mu\nu})| \ \leqq \ |\widehat{h}(t_{\mu\nu})| \ \leqq \ |\widehat{h}(t_{\mu\nu}) - \widehat{h}(t)| + |\widehat{h}(t)|$$

Integrating (7.14) for $t \in \Delta_\mu$ yields

(7.15) $$|\widehat{h}_+(t_{\mu\nu})| \ \leqq \ \frac{1}{|\Delta_\mu|} \int\limits_{\Delta_\mu} \left(|\widehat{h}(t_{\mu\nu}) - \widehat{h}(t)| + |\widehat{h}(t)|\right)\ dt \ \leqq$$

$$\leqq \ \tfrac{1}{|\Delta_\mu|} \int\limits_{\Delta_\mu} [\omega_1(\widehat{h}, t; \delta\varepsilon) + |\widehat{h}(t)| \ ]\ dt$$

Now (7.13, 15), (3.32, 20), (7.10) and $|\Delta_\mu| = \delta\varepsilon$ yield

(7.16) $$T_3 \ \leqq \ \sum_{\mu=-\infty}^{\infty} \int\limits_{\Delta_\mu} [\omega_1(\widehat{h}, t; \delta\varepsilon) + |\widehat{h}(t)| \ ]\ dt \ = \ \tau_1(\widehat{h}; \delta\varepsilon)_{L_1} + \|\widehat{h}\|_{L_1}$$

Next we show that

(7.17) $$\tau_1(\widehat{h}_+; 2\delta\varepsilon)_{L_1} \ \leqq \ 2\tau_1(\widehat{h}; \delta\varepsilon)_{L_1}$$

Indeed, by one of the basic properties of τ_1,

(7.18) $\tau_1(\varphi; 2d)_{L_q} \ \leqq \ 2\tau_1(\varphi; d)_{L_q}$, $\forall q \in [1,\infty]$, $\forall d > 0$, for every bounded measurable $\varphi :\ \mathbb{R} \to \mathbb{R}$.

It remains to prove

(7.17′) $$\tau_1(\widehat{h}_+; \delta\varepsilon)_{L_1} \ \leqq \ \tau_1(\widehat{h}; \delta\varepsilon)_{L_1}.$$

(7.17') follows from the definition of local and average moduli of smoothness and the following inequality, where $a, b, c \in \mathbb{R}$,

(7.19) $$|\max\{a,b\} - \max\{a,c\}| \ = \ \left|\frac{1}{2}(a+b+|a-b|) \ - \ \frac{1}{2}(a+c+|a-c|)\right| \ =$$

$$= \ \frac{1}{2}\Big|((b-c) + (|a-b| - |a-c|)\Big| \ \leqq \ \frac{1}{2}\left[|b-c| + \Big||a-b| - |a-c|\Big|\right] \ \leqq$$

$\leqq \dfrac{1}{2}\left[|b-c| + \Big|(a-b)-(a-c)\Big|\right] \;=\; |b-c|,$
which completes the proof of (7.17).

(7.12,13,16,17) imply, together with Hölder's inequality and (3.30),

$$(7.20)\quad T_1 \;\leqq\; \frac{1}{\varepsilon}\left[\, 2\left\|\, \Phi\left(\frac{\cdot}{\varepsilon}\right)\right\|_{L_{p_1}} \,.\, \tau_1(\widehat{h};\delta\varepsilon)_{L_1} \,+ \right.$$
$$\left. +\, \left(\|\widehat{h}\|_{L_1} \,+\, \tau_1(\widehat{h};\delta\varepsilon)_{L_1}\right)\, \omega_1\left(\,\Phi\left(\frac{\cdot}{\varepsilon}\right); \delta\varepsilon\right)_{L_{p_1}} \;\right] \leqq$$
$$\leqq \frac{R^{1-1/p}}{\varepsilon}\left[\, 2\left\|\,\Phi\left(\tfrac{\cdot}{\varepsilon}\right)\right\|_{L_{p_1}}\; \tau_1(\widehat{h};\delta\varepsilon)_{L_p} \,+ \right.$$
$$\left. +\, \left(\|\widehat{h}\|_{L_p} \,+\, \tau_1(\widehat{h};\delta\varepsilon)_{L_p}\right)\, \omega_1\left(\,\Phi\left(\frac{\cdot}{\varepsilon}\right); \delta\varepsilon\right)_{L_{p_1}} \;\right].$$

Hence, from (7.8,9,20), and the triangle inequality, available for $\mathcal{L}_{p_1}(L^{\rho}_{p_1})^{p_1/\rho_1}$

$$(7.21)\quad \left(E_{F^N}\|h-\widehat{h}_{+,\varepsilon,\delta}\|^{\rho}_{L_{p_1}}\right)^{1/\rho_1} \;\leqq\; 2(3+2^{\rho/\rho_1})\;.\;K^{1/\rho_1}\;.\;\alpha\left(\frac{1}{N}\right)^{1/\rho_1} \;+$$
$$+\,\frac{R^{\left(1-\frac{1}{p}\right)\frac{\rho}{\rho_1}}}{\varepsilon^{\rho/\rho_1}}\left\{\, 2^{\rho/\rho_1}\;.\;\left\|\,\Phi\left(\frac{\cdot}{\varepsilon}\right)\right\|^{\rho/\rho_1}_{L_{p_1}}\;\left[E_{F^N}\;\tau_1(\widehat{h};\delta\varepsilon)^{\rho}_{L_p}\right]^{1/\rho_1} \;+\right.$$
$$\left. +\,\omega_1\left(\,\Phi\left(\frac{\cdot}{\varepsilon}\right);\delta\varepsilon\right)^{\rho/\rho_1}_{L_{p_1}}\;\left[(E_{F^N}\;\|\widehat{h}\|^{\rho}_{L_p})^{1/\rho_1} \,+\, (E_{F^N}\;\tau_1(\widehat{h};\delta\varepsilon)^{\rho}_{L_p})^{1/\rho_1}\;\right]\;\right\}.$$

We begin evaluating the components of the RHS in (7.21).

$$(7.22)\quad \left\|\Phi\left(\frac{\cdot}{\varepsilon}\right)\right\|_{L_{p_1}} \;=\; \varepsilon^{1/p_1}\;\|\Phi\|_{L_{p_1}}.$$

Using some basic properties of integral moduli of smoothness, we get

$$(7.23)\quad \omega_1\left(\,\Phi\left(\frac{\cdot}{\varepsilon}\right);\;\delta\varepsilon\right)_{L_{p_1}} \;\leqq\; \min\left\{2\left\|\Phi\left(\frac{\cdot}{\varepsilon}\right)\,\right\|_{L_{p_1}}\;,\qquad \delta\varepsilon\;\left\|\left[\Phi\left(\frac{\cdot}{\varepsilon}\right)\right]'\right\|_{L_{p_1}}\right\} \;=$$
$$=\;\min\left\{2\varepsilon^{1/p_1}\;\|\Phi\|_{L_{p_1}},\;\delta\;\left\|\Phi'\left(\tfrac{\cdot}{\varepsilon}\right)\right\|_{L_{p_1}}\right\} \;=$$
$$=\;\varepsilon^{1/p_1}\;\min\left\{2\|\Phi\|_{L_{p_1}},\;\delta\|\Phi'\|_{L_{p_1}}\right\} \;\leqq\; \varepsilon^{1/p_1}\;\min\{2,\;\delta\}\;\|\Phi\|_{W^1_{p_1}}$$

So far we have been making all considerations for a very general estimator $\widehat{h}$, including the wavelet estimator (of both density or regression function) from [DJ], as well as the kernel estimator (of a regression function) from [LMS]. Let now $\widehat{h}$ be the wavelet estimator from [DJ]. By the triangle inequality about $\mathcal{L}_{p_1}(L^{\rho}_{p_1})^{p_1/\rho_1}$, and because moduli of smoothness are semi–norms,

$$(7.24)\quad \left(E_{F^N}\;\tau_1(\widehat{h};\delta\varepsilon)^{\rho}_{L_p}\right)^{1/\rho_1} \;\leqq\; \tau_1(h;\delta\varepsilon)^{\rho/\rho_1}_{L_p} \;+\; \left(E_{F^N}\;\tau_1(\widehat{h}-h;\delta\varepsilon)^{\rho}_{L_p}\right)^{1/\rho_1}.$$

By (3.28), $\dfrac{1}{p} \leqq s_0 \leqq 1$; by properties of the τ_1 – modulus and by embedding theorems about Besov spaces,

$$(7.25)\quad \tau_1(h;\delta\varepsilon)^{\rho_1/\rho}_{L_p} \;\leqq\; C_1(p,p_1,\rho,r_0,s)\;(\delta\varepsilon)^{\frac{\rho s_0}{\rho_1}}\;\|h\|^{\frac{\rho}{\rho_1}}_{B^{s_o}_{p_1}} \;\leqq$$
$$\leqq\; C_2(p,p_1,\rho,r_0,s)\;(\delta\varepsilon)^{\frac{\rho s_0}{\rho_1}}\;\|h\|^{\rho/\rho_1}_{B^s_{pq}}.$$

By the same property of the τ_1 – modulus, and by the respective rate in Theorem 1 in [DJ] (for $\sigma = 1/p$, $\pi = p$, $u = 1$, $q = q$, $\rho = \rho_1 = 2$ in their notations)

$$(7.26)\quad \left(E_{F^N}\ \tau_1(\widehat{h}-h;\delta\varepsilon)^{\rho}_{L_p}\right)^{1/\rho_1} \leqq C_2(p,p_1,\rho,r_0,s)\ .\ (\delta\varepsilon)^{\frac{\rho s_0}{\rho_1}} \left(E_{F^N}\ \|\widehat{h}-h\|^{\rho}_{B^{s_0}_{p_1}}\right)^{1/\rho_1} \leqq$$
$$\leqq C_3(p,p_1,\rho,r_0,s,L)\ (\delta\varepsilon)^{\frac{\rho s_0}{\rho_1}}\ N^{-\frac{s-s_0}{2s+1}\cdot\frac{\rho}{\rho_1}},$$

where $L = \|h\|_{B^s_{pq}}$.

Analogously,

$$(7.27)\quad \left(E_{F^N}\ \|\widehat{h}\|^{\rho}_{L_p}\right)^{1/\rho_1} \leqq \|h\|^{\rho/\rho_1}_{L_p} + \left(E_{F^N}\ \|\widehat{h}-h\|^{\rho}_{L_p}\right)^{1/\rho_1} \leqq$$
$$\leqq \|h\|^{\rho/\rho_1}_{B^s_{pq}} + \left(E_{F^N}\ \|\widehat{h}-h\|^{\rho}_{B^0_{p_1}}\right)^{1/\rho_1} \leqq L^{\rho/\rho_1} + C_4(p,\rho,s,L)\ N^{-\frac{s}{2s+1}\cdot\frac{\rho}{\rho_1}}.$$

(7.21–27) now imply

$$(7.28)\quad \left(E_{F^N}\|h-\widehat{h}_{+,\varepsilon,\delta}\|^{\rho}_{L_{p_1}}\right)^{1/\rho_1} \leqq 2(3+2^{\rho/\rho_1})\ .\ K^{1/\rho_1}\ .\ \alpha\left(\frac{1}{N}\right)^{1/\rho_1} +$$
$$+ \frac{R^{\left(1-\frac{1}{p}\right)\frac{\rho}{\rho_1}}}{\varepsilon^{\rho/\rho_1}} \left\{ 2^{\rho/\rho_1}\ .\ \varepsilon^{\frac{\rho}{\rho_1 p_1}}.\ \|\Phi\|_{L_{p_1}}\ .\ (\delta\varepsilon)^{\frac{\rho s_0}{\rho_1}}\ .\ \left[C_2(p,p_1,\rho,r_0,s)\ .\ L^{\rho/\rho_1} + \right.\right.$$
$$\left. + C_3(p,p_1,\rho,r_0,s,L)\ N^{-\frac{s-s_0}{2s+1}\cdot\frac{\rho}{\rho_1}}\right] +$$
$$+ \varepsilon^{\frac{\rho}{\rho_1 p_1}}\ .\ \delta^{\frac{\rho}{\rho_1}}\ .\ \|\Phi\|^{\rho/\rho_1}_{W^1_{p_1}}\ .\ \left[L^{\rho/\rho_1} + C_4(p,q,s,L)\ N^{-\frac{s}{2s+1}} + \right.$$
$$\left.\left. + (\delta\varepsilon)^{\frac{\rho s_0}{\rho_1}} \left(C_2(p,p_1,\rho,r_0,s)\ L^{\rho/\rho_1} + C_3(p,p_1,\rho,r_0,s,L)\ N^{-\frac{s-s_0}{2s+1}\cdot\frac{\rho}{\rho_1}}\right)\right]\right\} \leqq$$
$$\leqq 2(3+2^{\rho/\rho_1})\ .\ K^{1/\rho_1}\ .\ \alpha\left(\frac{1}{N}\right)^{1/\rho_1} +$$
$$+ \left[R^{1-1/p}\ .\ \|\Phi\|_{W^1_{p_1}}\ .\ L\ .\varepsilon^{-(1-1/p_1)}\ .\ \delta^{s_0}\right]^{\rho/\rho_1}\ .$$
$$\cdot \left\{ 2^{\rho/\rho_1}\ .\ \varepsilon^{s_0\rho/\rho_1}\ .\ \left[C_2 + o_N(1)\right] + \right.$$
$$\left. + \delta^{(1-s_0)\rho/\rho_1}\ .\ \left[1 + o_N(1) + (\delta\varepsilon)^{s_0\rho/\rho_1}\left(C_2 + o_N(1)\right)\right]\right\} \leqq$$
$$\leqq \left[C_5(p,p_1,\rho,r_0,s,L)\ .\ R^{1-1/p}\ .\ \|\Phi\|_{W^1_{p_1}}\ .\ L\ .\ \varepsilon^{-(1-1/p_1)}\ .\ \delta^{s_0}\ .\ \max\{\varepsilon^{s_0},\ \delta^{1-s_0}\}\right]^{\frac{\rho}{\rho_1}} +$$
$$+ 2(3+2^{\rho/\rho_1})\ .\ K^{1/\rho_1}\ .\ \alpha\left(\frac{1}{N}\right)^{1/\rho_1},$$
$$\forall N > N_1 \text{ (invoking also (3.31))}.$$

Requiring that for $N > N_1$ (i.e., asymptotically) the contributions of the two summands in the RHS of (7.28) are the same, we get

(7.29) $\left[\varepsilon_N^{-(1-1/p_1)} \ . \ \delta_N^{s_0} \ . \ \max\{\varepsilon_N^{s_0}, \ \delta_N^{1-s_0}\} \ \right]^{\rho/\rho_1} \ = \ C_6 \ \alpha(1/N)^{1/\rho_1}$,

where

(7.30) $C_6 \ = \ C_6(p, p_1, \rho, r_0, s, L, K, R, \phi) \ =$

$$= \ \frac{2(3+2^{\rho/\rho_1}) \ . \ K^{1/\rho_1}}{[C_5(p,p_1,\rho,r_0,s,L) \ . \ R^{1-1/p} \ . \ \|\Phi\|_{W^1_{p_1}} \ . \ L]^{\rho/\rho_1}} \ .$$

Now we have to determine $s_0^* = s_0^*(1/N)$, such that simultaneously

(7.31) $\varepsilon_N^{s_0^*} \ = \ \delta_N^{1-s_0^*}$

$\varepsilon_N^{-(1-1/p_1)} \ . \ \delta_N \ = \ C_6^{\rho_1/\rho} \ . \ \alpha(1/N)^{1/\rho}$

This yields

(7.32) $$s_0^* \ = \ \frac{\log_N \delta_N}{\log_N \varepsilon_N + \log_N \delta_N}$$

and, in view of $\delta_N \ = \ C_6^{\rho_1/\rho} \ \alpha(1/N)^{1/\rho} \ \varepsilon_N^{1-1/p_1}$, we get

(7.33) $$s_0^* \ = \ \frac{\frac{\rho_1}{\rho}\log_N C_6 + \frac{1}{\rho}\log_N \alpha(1/N) + (1-1/p_1)\log_N \varepsilon_N}{\frac{\rho_1}{\rho}\log_N C_6 + \frac{1}{\rho}\log_N \alpha(1/N) + (2-1/p_1)\log_N \varepsilon_N}$$

¿From (7.33, 29), for $\varepsilon_N = \beta^{-1}(C_7 \ \alpha(1/N)^{1/\rho})$, we obtain

(7.34) $$\log_N \delta_N \ = \ \begin{cases} \frac{\rho_1}{\rho}\log_N C_6 + \frac{1}{\rho}\log_N \alpha(1/N) + \\ + (1-1/p_1)\log_N \ [\beta^{-1}(C_7 \ \alpha(1/N)^{1/\rho})], & s_0^* \leqq s_0 \leqq 1 \\ \frac{1}{s_0}\Big(\frac{\rho_1}{\rho}\log_N C_6 \ + \ \frac{1}{\rho}\log_N \alpha(1/N) \ + \\ + \ \Big(-1 + \frac{1-1/p_1}{s_0}\Big) \ \log_N \ [\beta^{-1}(C_7 \ \alpha(1/N)^{1/\rho})]\Big), & r_0 \leqq s_0 \leqq s_0^* \leqq 1 \end{cases}$$

(recall that $s_0 = \min\Big\{1, \ \max\{1/p, r_0\}\Big\}, \ \ 0 < r_0 \leqq 1$),

or, equivalently to (7.34),

(7.35) $$\log_N \delta_N \ = \ \left\{\frac{\rho_1}{\rho}\log_N C_6 \ + \ \frac{1}{\rho}\log_N \alpha(1/N) \ + \ \left(1 - \frac{1}{p_1}\right) \ \log_N \left[\beta^{-1}\left(C_7 \ \alpha\left(\frac{1}{N}\right)^{1/\rho}\right)\right]\right\} \ . \\ . \ \left[1 + \left(\frac{1}{s_0} - \frac{1}{s_0^*}\right)_+\right]$$

i.e.,

(7.36) $$\delta_N \ = \ C_6^{\left[1+\left(\frac{1}{s_0}-\frac{1}{s_0^*}\right)_+\right]\frac{\rho_1}{\rho}} \ . \ \left\{\alpha\left(\frac{1}{N}\right)^{1/\rho} \ . \ \beta^{-1}\left[C_7 \ \alpha\left(\frac{1}{N}\right)^{\frac{1}{\rho}}\right]^{1-1/p_1}\right\}^{1+\left(\frac{1}{s_0}-\frac{1}{s_0^*}\right)_+} ,$$

(7.37) $$s_0^* \ = \ 1 \ - \ \frac{\log_N\left[\beta^{-1}\left(C_7 \ \alpha\left(\frac{1}{N}\right)^{1/\rho}\right)\right]}{\frac{\rho_1}{\rho}\log_N C_6 \ + \ \frac{1}{\rho}\log_N \alpha\left(\frac{1}{N}\right) \ + \ \left(2 - \frac{1}{p_1}\right) \ \log_N\left[\beta^{-1}\left(C_7 \ \alpha\left(\frac{1}{N}\right)^{1/\rho}\right)\right]} \ \in$$

$\in (0,1)$,

i.e., (3.35, 38) hold. From (7.36) and (3.35) also (3.37) is found to hold true. □

Remark 7.1. *For the variable–kernel regression–function estimator from [LMS] asymptotic minimax rates are available only when the loss function in the risk is the L_{p_1}– norm. It cannot be replaced by any $B^{s_1}_{p_1 q_1}$– norm for any $s_1 :\ 0 < s_1 < s$, because the estimator from [LMS] may be non–smooth, even discontinuous in a finite number of points (if h has compact support) (or locally finite, if supp h is not compact). Because of this, the bounds (7.25–27) are unavailable for this estimator. However, if $s_0 = 1/p$ (cf. (3.28)), then $\tau_1(\widehat{h};\delta\varepsilon)^{\rho/\rho_1}_{L_p}$ can be evaluated in the following way; corresponding to (7.24–26):*

$$(7.24_1)\quad \tau_1(\widehat{h};\delta\varepsilon)_{L_p} \leqq C_p \,.\, (\delta\varepsilon)^{1/p} \,.\, \left[\bigvee_{-\infty^p}^{\infty} \widehat{h}\right]^{1/p} = C_p \,.\, (\delta\varepsilon)^{s_0} \,.\, O_N(1)$$

In other words, although the estimator may be discontinuous, it has always bounded Jordan variation, hence p–variation, too. Note that in this case $\rho = p_1$.

Without going into details, let us note that on the basis of (7.24_1) one can deduce that C_5 in the RHS of (7.28) and C_6 in (7.29) should be replaced by $O_N(1)$. Thus Theorem 3.2 holds true for the estimator from [LMS] with the same rates, but without deterministic bounds for the constant factors to those rates, as was the case with wavelets.

Proof of Theorem 3.3. The idea is similar to the one of the proof of Theorem 3.2, but the realization of this idea is more technically involved. Because of this, we shall only sketch the analogous moments and will concentrate the exposition on the new elements in the proof.

$$(7.8')\quad \left(E_{F^N}\|h-\widehat{h}_{+,\varepsilon,\delta,\delta_1}\|^{\rho}_{L_{p_1}}\right)^{1/\rho_1} \leqq 2(3+2^{\rho/\rho_1})K^{1/\rho_1}\alpha\left(\frac{1}{N}\right)^{1/\rho_1} + \left(E_{F^N}\|\widehat{h}_{+,\varepsilon}-\widehat{h}_{+,\varepsilon,\delta,\delta_1}\|^{\rho}_{L_{p_1}}\right)^{1/\rho_1},$$

$$(7.9')\quad \|\widehat{h}_{+,\varepsilon}-\widehat{h}_{+,\varepsilon,\delta,\delta_1}\|_{L_{p_1}} \leqq \widehat{\mathcal{J}}+\widehat{\mathcal{J}}_1,$$

where

$$(7.9'_1)\quad \widehat{\mathcal{J}} = \left(\int_{-\infty}^{\infty}\left|\frac{1}{\varepsilon}\sum_{\mu\in I}\left[\int_{\Delta_\mu}\Phi\left(\frac{x-t}{\varepsilon}\right)\widehat{h}_+(t)\,dt - \sum_{\nu=1}^{n_\mu}A_{\mu\nu}\,\Phi\left(\frac{x-t_{\mu\nu}}{\varepsilon}\right)\widehat{h}_+(t_{\mu\nu})\right]\right|^{p_1}dx\right)^{1/p_1},$$

$$(7.9'_2)\quad \widehat{\mathcal{J}}_1 = \left(\int_{-\infty}^{\infty}\left|\frac{1}{\varepsilon}\sum_{\mu_1\in I_1}\left[\int_{\Delta_{\mu_1}}\Phi\left(\frac{x-t}{\varepsilon}\right)\widehat{h}_+(t)\,dt - \sum_{\nu=1}^{n_{\mu_1}}A_{\mu_1\nu}\,\Phi\left(\frac{x-t_{\mu_1\nu}}{\varepsilon}\right)\widehat{h}_+(t_{\mu_1\nu})\right]\right|^{p_1}dx\right)^{1/p_1}.$$

Here I and I_1 have the meaning of (3.46, 46'), respectively. The estimation of $\left(E_{F^N}\|\widehat{\mathcal{J}}\|^{\rho}_{L_{p_1}}\right)^{1/\rho_1}$ goes along the lines of the proof of Theorem 3.2. We only need to evaluate expressions of the type

$\sum\limits_{\mu\in I}\int\limits_{\Delta_\mu} \omega_1(\widehat{h}_+,\theta,\delta\varepsilon)\ d\theta$, which is easy:

$$(7.38)\quad \sum_{\mu\in I}\int\limits_{\Delta_\mu} \omega_1(\widehat{h}_+,\theta,\delta\varepsilon)\ d\theta \ \leqq \ \sum_{\mu=-\infty}^{\infty}\int\limits_{\Delta_\mu} \omega_1(\widehat{h}_+,\theta,\delta\varepsilon)\ d\theta \ = \ \tau_1(\widehat{h}_+,\delta\varepsilon)_{L_1}.$$

Evaluating $(7.9_2')$ is more technically involved. First of all, observe that

$$(7.39)\quad \widehat{h}_+\Big|_{\Delta_{\mu_1}} \ \equiv \ \widehat{h}\Big|_{\Delta_{\mu_1}}, \qquad \forall\mu_1\in I_1.$$

Therefore, $\widehat{\mathcal{J}}_1$ can be rewritten as

$$(7.9'')\quad \widehat{\mathcal{J}}_1 = \left(\int\limits_{-\infty}^{\infty}\left|\frac{1}{\varepsilon}\sum_{\mu_1\in I_1}\left[\int\limits_{\Delta_{\mu_1}}\Phi\left(\frac{x-t}{\varepsilon}\right)\ \widehat{h}(t)\ dt - \sum_{\nu=1}^{n_{\mu_1}(r_1)} A_{\mu_1\nu}\ \Phi\left(\frac{x-t_{\mu_1\nu}}{\varepsilon}\right)\ \widehat{h}(t_{\mu_1\nu})\right]\right|^{p_1} dx\right)^{1/p_1}.$$

(Recall that for $\mu_1\in I_1$ $\qquad \Delta_{\mu_1} \ = \ [(k_{1\mu}+(\nu_{1\mu}-1)\delta_1)\varepsilon,(k_{1\mu}+\nu_{1\mu}\delta_1)\varepsilon].$)

Consider the Steklov mean $(\widehat{h})_{r_1,\ \delta_1\varepsilon/r_1}$ of $\widehat{h}$. In the definition of $(\widehat{h})_{r_1,\ \delta_1\varepsilon/r_1}$ values of $\widehat{h}(t)$ are being utilized for $t\in\Delta_\mu+[0,\ \delta_1\varepsilon]$.

$$(7.11')\quad \mathcal{J}_{1,\mu_1}(x) \ := \int\limits_{\Delta_{\mu_1}}\Phi\left(\frac{x-t}{\varepsilon}\right)\ \widehat{h}(t)\ dt \ - \ \sum_{\nu=1}^{n_{\mu_1}(r_1)} A_{\mu_1\nu}\ \Phi\left(\frac{x-t_{\mu_1\nu}}{\varepsilon}\right)\ \widehat{h}(t_{\mu_1\nu}) \ =$$

$$= \int\limits_{\Delta_{\mu_1}}\Phi\left(\frac{x-t}{\varepsilon}\right)\ [\widehat{h}(t)-(\widehat{h})_{r_1,\ \delta_1\varepsilon/r_1}(t)]\ dt \ +$$

$$+\int\limits_{\Delta_{\mu_1}}\Phi\left(\frac{x-t}{\varepsilon}\right)\ (\widehat{h})_{r_1,\ \delta_1\varepsilon/r_1}(t) \ - \ \sum_{\nu=1}^{n_{\mu_1}(r_1)} A_{\mu_1\nu}\ \Phi\left(\frac{x-t_{\mu_1\nu}}{\varepsilon}\right)\ (\widehat{h})_{r_1,\ \delta_1\varepsilon/r_1}(t_{\mu\nu}) \ +$$

$$+\sum_{\nu=1}^{n_{\mu_1}(r_1)} A_{\mu_1\nu}\ \Phi\left(\frac{x-t_{\mu_1\nu}}{\varepsilon}\right)\ [(\widehat{h})_{r_1,\ \delta_1\varepsilon/r_1}(t_{\mu_1\nu})-\widehat{h}(t_{\mu_1\nu}) \ =$$

$$= \ \mathcal{J}'_{1,\ \mu_1}(x)+\mathcal{J}''_{1,\ \mu_1}(x)+\mathcal{J}'''_{1,\ \mu_1}(x)$$

$$(7.40)\quad |\mathcal{J}'_{1,\ \mu_1}(x)| \ \leqq \int\limits_{\Delta_{\mu_1}}\left|\Phi\left(\frac{x-t}{\varepsilon}\right)\right|\ \omega_{r_1}(\widehat{h},t;\ \delta_1\varepsilon)\ dt$$

$$(7.41)\quad |\mathcal{J}'''_{1,\ \mu_1}(x)| \ \leqq \ \sum_{\nu=1}^{n_\mu(r_1)} |A_{\mu_1\nu}|\ \left|\Phi\left(\frac{x-t_{\mu_1\nu}}{\varepsilon}\right)\right|\ \omega_{r_1}(\widehat{h},t_{\mu\nu};\ \delta_\varepsilon).$$

For the evaluation of $|\mathcal{J}''_{1,\ \mu_1}(x)|$ we expand the function

$$(7.42)\quad G(x,t) \ := \ \Phi\left(\frac{x-t}{\varepsilon}\right)\ (\widehat{h})_{r_1,\ \delta_1\varepsilon/r_1}(t)$$

as a function of t in a local Taylor expansion with integral remainder involving $\dfrac{\partial^{r_1}}{\partial t^{r_1}}G(x,t)$. Note that, by definition of Φ, by $r\geqq r_1$, and by definition of the Steklov mean,

$$(7.43)\quad \frac{\partial^{r_1}}{\partial t^{r_1}}G(x,\cdot) \ \in \ L_1(\Delta_{\mu_1}+[0,\ \delta_1\varepsilon]), \qquad \forall x\in\mathbb{R}\ .$$

$$(7.11'')\quad \mathcal{J}''_{1,\ \mu_1}(x) \ = \int\limits_{\Delta_{\mu_1}} G(x,t)\ dt \ - \sum_{\nu=1}^{n_{\mu_1}(r_1)} A_{\mu_1\nu}\ G(x,t_{\mu_1\nu}) \ =$$

$$= \int\limits_{\Delta_{\mu_1}} P_{r_1-1,\ x}(t)\ dt \ - \sum_{\nu=1}^{n_{\mu_1}(r_1)} A_{\mu_1\nu}\ P_{r_1-1,\ x}(t_{\mu_1\nu}) \ +$$

$$+ \frac{1}{(r_1-1)!} \int\limits_0^1 (1-\theta)^{r_1-1} \Bigg[\int\limits_{\Delta_\mu} (t-t_0)^{r_1} \ \frac{\partial^{r_1}}{\partial t^{r_1}}\ G(x,t_0+\theta(t-t_0))\ dt \ -$$

$$- \sum_{\nu=1}^{n_{\mu_1}(r_1)} A_{\mu_1\nu}\ (t_{\mu_1\nu}-t_0)^{r_1}\ \frac{\partial^{r_1}}{\partial t^{r_1}}\ G(x_0+\theta(t_{\mu_1\nu}-t_0)) \Bigg]\ d\theta,$$

where:

(7.44) $t_0 \ = \ [k_{1\mu_1} + (\nu_{1\mu_1} - 1/2)\ \delta_1]\ \varepsilon$ is the centre of Δ_{μ_1};

(7.45) $P_{r_1-1,\ x}(t)$ is a polynomial in t of degree at most $r_1 - 1$ and with coefficients depending on x.

Now we make use of the accuracy of the quadrature formula of order r_1, and, upon applying the triangle inequality and

(7.46) $|t-t_0| \ \leqq \ \delta_1\varepsilon_1/2$ for $t \in \Delta_{\mu_1}$, $|t_{\mu_1\nu}-t_0| \ \leqq \ \delta_1\varepsilon_1/2, \quad \nu = 1,\dots,n_{\mu_1}(r_1)$,

obtain

$$(7.11''')\quad |\mathcal{J}''_{1\mu_1}(x)| \ = \ \frac{1}{(r_1-1)!} \Bigg| \int\limits_0^1 (1-\theta)^{r_1-1}$$

$$\Bigg[\int\limits_{\Delta_{\mu_1}} (t-t_0)^{r_1}\ \frac{\partial^{r_1}}{\partial t^{r_1}}\ G(x,t_0+\theta(t_{\mu_1\nu}-t_0))\ dt \ -$$

$$- \sum_{\nu=1}^{n_{\mu_1}(r_1)} A_{\mu_1\nu}\ (t_{\mu_1\nu}-t_0)^{r_1}\ \frac{\partial^{r_1}}{\partial t^{r_1}}\ G(x,t_0+\theta(t_{\mu_1\nu}-t_0)) \Bigg]\ d\theta \Bigg| \ \leqq$$

$$\leqq \ \frac{2^{-r_1}(\delta_1\varepsilon)^{r_1}}{(r_1-1)!} \int\limits_0^1 (1-\theta)^{r_1-1} \ . \ \Bigg[\int\limits_{\Delta_{\mu_1}} \left| \frac{\partial^{r_1}}{\partial t^{r_1}}\ G(x,t_0+\theta(t-t_0)) \right| dt \ +$$

$$+ \sum_{\nu=1}^{n_{\mu_1}(r_1)} |A_{\mu_1\nu}| \ . \ \left| \frac{\partial^{r_1}}{\partial t^{r_1}}\ G(x,t_0+\theta(t_{\mu_1\nu}-t_0)) \right| \Bigg]\ d\theta \ =: \ T_4$$

Next, Leibniz rule and properties of the Steklov function imply

$$(7.11^{\mathrm{iv}})\quad T_4 \ = \ \frac{2^{-r_1}(\delta_1\varepsilon)^{r_1}}{(r_1-1)!} \ . \int\limits_0^1 (1-\theta)^{r_1-1} \ .$$

$$. \ \Bigg[\int\limits_{\Delta_{\mu_1}} \left| \sum_{j=0}^{r_1} \binom{r_1}{j} \left(\frac{-\theta}{\varepsilon}\right)^{r_1-j} \Phi^{(r_1-j)} \left(\frac{x-t_0+\theta(t-t_0)}{\varepsilon}\right) \ . \ \theta^j (\widehat{h})^{(j)}_{r_1,\ \delta_1\varepsilon/r_1}(t_0+\theta(t- \right.$$

$$\left. t_0)) \right|\ dt \ +$$

$$+\sum_{\nu=1}^{n_{\mu_1}(r_1)} |A_{\mu_1\nu}| \,.\left|\sum_{j=0}^{r_1}\binom{r_1}{j}\left(\frac{-\theta}{\varepsilon}\right)^{r_1-j}\Phi^{(r_1-j)}\left(\frac{x-t_0+\theta(t_{\mu_1\nu}-t_0)}{\varepsilon}\right).\,\theta^j(\widehat{h})^{(j)}_{r_1,\ \delta_1\varepsilon/r_1}(t_0+\right.$$

$$\left.\left.\theta(t_{\mu_1\nu}-t_0))\right|\right]d\theta \leqq$$

$$\leqq \frac{2^{-r_1}(\delta_1\varepsilon)^{r_1}}{(r_1-1)!}\,.\int_0^1(1-\theta)^{r_1-1}\,.\,\theta^{r_1}\,.\left\{\sum_{j=0}^{r_1}\binom{r_1}{j}\varepsilon^{j-r_1}\,.\right.$$

$$.\left[\int_{\Delta_{\mu_1}}\left|\Phi^{(r_1-j)}\left(\frac{x-t_0+\theta(t-t_0)}{\varepsilon}\right)\right|.\right.$$

$$.\,C(r_1,j)\ (\delta_1\varepsilon)^{-j}\ \omega_j(\widehat{h},t_0+\theta(t-t_0);\ \delta_1\varepsilon)\ dt+$$

$$+\sum_{\nu=1}^{n_{\mu_1}(r_1)}|A_{\mu_1\nu}|\,.\left|\Phi^{(r_1-j)}\left(\frac{x-t_0-\theta(t_{\mu_1\nu}-t_0)}{\varepsilon}\right)\right..$$

$$\left.\left.\cdot\, C(r_1,j)\ (\delta_1\varepsilon)^{-j}\ \omega_j\Big(\widehat{h},\ t_0+\theta(t_{\mu_1\nu}-t_0);\ \delta_1\varepsilon\Big)\right]\right\}d\theta$$

which is the desired bound for $|\mathcal{J}''_{1\mu_1}(x)|$.

Returning now to the evaluation of $\mathcal{J}_1$, we obtain by similar arguments as in (7.12–20), after some straightforward computations, making also use of (7.10) and (3.20),

$$(7.12')\quad \widehat{\mathcal{J}}_1 \leqq \sum_{l=1}^{3}\left[\int_{-\infty}^{\infty}\left(\frac{1}{\varepsilon}\sum_{\mu_1\in I_1}\left|\mathcal{J}^{[l]}_{1,\mu}(x)\right|\right)^{p_1}dx\right]^{1/p_1}\leqq$$

$$\leqq \frac{1}{\varepsilon}\left\{2\left(\sum_{\mu_1\in I_1}\int_{\Delta_{\mu_1}}\omega_{r_1}(\widehat{h},t;\ \delta_1\varepsilon)\ dt\right)\left\|\Phi\left(\frac{\cdot}{\varepsilon}\right)\right\|_{L_{p_1}}+\right.$$

$$+\frac{2^{-r_1}\delta_1^{r_1}}{(r_1-1)!}\left(\int_0^1(1-\theta)^{r_1-1}\,.\,\theta^{r_1}\ d\theta\right).\sum_{j=0}^{r_1}\binom{r_1}{j}.\,C(r_1,j).\delta_1^{-j}\,.\left\|\Phi^{(r_1-j)}\left(\frac{\cdot}{\varepsilon}\right)\right\|_{L_{p_1}}$$

$$\left..\,2\,.\left(\sum_{\mu_1\in I_1}\int_{\Delta_{\mu_1}}\omega_j(\widehat{h},t;\ 2\delta_1\varepsilon)\ dt\right)\right\}\leqq$$

$$\leqq \frac{2}{\varepsilon}\left[C_1(r)\left\|\Phi\left(\frac{\cdot}{\varepsilon}\right)\right\|_{L_{p_1}}\tau_{r_1}(\widehat{h};\ \delta_1\varepsilon)_{L_1}+\frac{2^{-r_1}\delta_1^{r_1}}{(r_1-1)!}\cdot\frac{(r_1-1)!\ r_1!}{(2r_1-1)!}\,\cdot\right.$$

$$\left.\cdot\sum_{j=0}^{r_1}\binom{r_1}{j}\delta_1^{-j}\,.\,C(r_1,j)\,.\,C_1(j)\,.\left\|\Phi^{(r_1-j)}\left(\frac{\cdot}{\varepsilon}\right)\right\|_{L_{p_1}}.\,\tau_j(\widehat{h};\ \delta_1\varepsilon)_{L_1}\right]=$$

$$=2\varepsilon^{-(1-1/p_1)}\left[C_1(r_1)\ \|\Phi\|_{L_{p_1}}\ \tau_{r_1}(\widehat{h};\ \delta_1\varepsilon)_{L_1}+\right.$$

$$\left.+\frac{2^{-r_1}\,.\,(r_1!)}{(2r_1-1)!}\,\delta_1^{r_1}\sum_{j=0}^{r_1}\binom{r_1}{j}\delta_1^{-j}\,.\,C(r_1,j)\,.\,C_1(j)\,.\left\|\Phi^{(r_1-j)}\right\|_{L_{p_1}}.\,\tau_j(\widehat{h};\ \delta_1\varepsilon)_{L_1}\right]\leqq$$

$$\leqq 2\varepsilon^{-(1-1/p_1)}R^{1-1/p_1}\left[C_1(r_1)\ \|\Phi\|_{L_{p_1}}\ \tau_{r_1}(\widehat{h};\ \delta_1\varepsilon)_{L_p}+\right.$$

$$\left.+\frac{2^{-r_1}\,.\,(r_1!)}{(2r_1-1)!}\,\delta_1^{r_1}\sum_{j=0}^{r_1}\binom{r_1}{j}\delta_1^{-j}\,.\,C(r_1,j)\,.\,C_1(j)\,.\left\|\Phi^{(r_1-j)}\right\|_{L_{p_1}}.\,\tau_j(\widehat{h};\ \delta_1\varepsilon)_{L_p}\right]=:T_5$$

Now we apply Marchaud–type inequalities of first kind for r and $j = 0, 1, \ldots, r_1 - 1$, and if $r_1 < r$, then also $j = r_1$.
For $j \geqq 1$ we get

$$(7.47) \quad \tau_j(\widehat{h};\ \delta_1\varepsilon)_{L_p(\mathbb{R})} \ \leqq\ C_2(r,j)(\delta_1\varepsilon)^j \int\limits_{\delta_1\varepsilon}^{\infty} t^{-j}\tau_r(f;t)_{L_p(\mathbb{R})}\ \frac{dt}{t}\ ,$$

For $j = 0$, by the properties of $A_{p,\ \delta_1\varepsilon}$ and (7.47) for $j = 1$,

$$(7.47') \quad \tau_0(\widehat{h};\ \delta_1\varepsilon)_{L_p} \ =\ \|f\|_{A_{p,\ \delta_1\varepsilon}} \ \leqq\ \|\widehat{h}\|_{L_p} \ +\ \tau_1(\widehat{h};\ \delta_1\varepsilon)_{L_p} \ \leqq$$
$$\leqq\ \|\widehat{h}\|_{L_p(\mathbb{R})} \ +\ C_2(r,1)\ \delta_1\varepsilon \int\limits_{\delta_1\varepsilon}^{\infty} t^{-1}\tau_r(\widehat{h};t)_{L_p(\mathbb{R})}\ \frac{dt}{t}\ .$$

Like in Theorem 3.2 (see the comment after (7.23)), we have so far been considering the estimator $\widehat{h}$ in a very general setting. Now let us concentrate our attention on the wavelet estimator in [DJ]. We use again (cf. (7.24)) the semi–norm property of the τ_r – modulus. Let us consider the case $j \geqq 1$ (the case $j = 0$ is treated in the same way, by reducing it to the case $j = 1$)

$$(7.24') \quad \left(E_{F^N}\ \tau_j(\widehat{h};\delta_1\varepsilon)^{\rho}_{L_p}\right)^{1/\rho_1} \ \leqq$$
$$\leqq \left\{E_{F^N}\left[C_2(r,j)(\delta_1\varepsilon)^j \left(\int\limits_{\delta_1\varepsilon}^{\infty} t^{-j}\tau_r(h;t)_{L_p}\ \frac{dt}{t} \ +\ \int\limits_{\delta_1\varepsilon}^{\infty} t^{-j}\tau_r(\widehat{h}-h;t)_{L_p}\ \frac{dt}{t}\right)\right]^{\rho}\right\}^{1/\rho_1} =:$$
$$=:\ T_6\ .$$

Let us find an upper bound for T_6 using $\frac{1}{p} \leqq \sigma < s < r$, and the properties of τ–moduli and Besov spaces.

If $\sigma < j$, then

$$(7.48) \quad (\delta_1\varepsilon)^j \int\limits_{\delta_1\varepsilon}^{\infty} t^{-j}\tau_r(h;t)_{L_p}\ \frac{dt}{t} \ \leqq\ (\delta_1\varepsilon)^j \int\limits_{\delta_1\varepsilon}^{\infty} t^{-j+\sigma-1}\ dt\ \cdot\ C_3(p,\sigma,r)\ \|h\|_{B^{\sigma}_{p\infty}} \ \leqq$$
$$\leqq\ \frac{C_3(p,\sigma,r)}{j-\sigma}\ (\delta_1\varepsilon)^{\sigma}\ \|h\|_{B^{\sigma}_{p\infty}} \ \leqq\ C_4(p,\sigma,s,j,r)\ .\ (\delta_1\varepsilon)^{\sigma}\ \|h\|_{B^s_{pq}}$$

If $\sigma \geqq j$, then, in view of $s > \sigma \geqq j \geqq 1 \geqq 1/p$,

$$(7.49) \quad (\delta_1\varepsilon)^j \int\limits_{\delta_1\varepsilon}^{\infty} t^{-j}\tau_r(h;t)_{L_p}\ \frac{dt}{t} \ \leqq\ (\delta_1\varepsilon)^j \int\limits_{0}^{\infty} t^{-j}\tau_r(h;t)\ \frac{dt}{t} \ \leqq$$
$$\leqq\ (\delta_1\varepsilon)^j\ .\ \|h\|_{B^j_{p1}} \ \leqq\ C_5(p,q,s,j,r)\ (\delta_1\varepsilon)^j\ \|h\|_{B^s_{pq}}\ .$$

Combining (7.48, 49) yields

$$(7.50) \quad (\delta_1\varepsilon)^j \int\limits_{\delta_1\varepsilon}^{\infty} t^{-j}\tau_r(h;t)_{L_p}\ \frac{dt}{t} \ \leqq\ C_6(p,q,\sigma,s,j,r)\ .\ (\delta_1\varepsilon)^{\min\{\sigma,\ j\}}\ \|h\|_{B^s_{pq}}\ .$$

T_6 can now be evaluated by

$$(7.51) \quad T_6 \ \leqq\ C_7(p,q,\sigma,s,j,r,\rho)\ .\ (\delta_1\varepsilon)^{\min\{\sigma,\ j\}.\rho/\rho_1}\ \|h\|^{\rho/\rho_1}_{B^s_{pq}} \ +$$

$$+ C_2(r,j)^{\rho/\rho_1} \ . \ \left\{ E_{F^N} \left[(\delta_1\varepsilon)^j \int\limits_{\delta_1\varepsilon}^{\infty} t^{-j} \tau_r(\widehat{h}-h;t)_{L_p} \ \frac{dt}{t} \right]^{\rho} \right\}^{1/\rho_1}$$

Applying again the idea of (7.49, 50) to the integral in the RHS of (7.51) and using the respective rate in Theorem 1 in [DJ], analogously to Theorem 3.2, one obtains

$$(7.52) \quad C_2(r,j)^{\rho/\rho_1} \ . \ \left\{ E_{F^N} \left[(\delta_1\varepsilon)^j \int\limits_{\delta_1\varepsilon}^{\infty} t^{-j} \tau_r(\widehat{h}-h;t)_{L_p} \ \frac{dt}{t} \right]^{\rho} \right\}^{1/\rho_1} \leqq$$
$$\leqq \ C_8(p,q,\sigma,s,j,r,\rho) \ . \ (\delta_1\varepsilon)^{\min\{\sigma,j\}} \ \|h\|_{B^s_{pq}} \ . \ o_N(1)$$

It follows from (7.50, 52) that there exist $C_{10} = C_{10}(p,q,\sigma,s,j,r,\rho)$ such that for any $N > N_1$

$$(7.53) \quad T_6 \ \leqq \ C_{10}(p,q,\sigma,s,j,r,\rho) \ . \ (\delta_1\varepsilon)^{\min\{\sigma,j\}.\rho/\rho_1} \ \|h\|^{\rho/\rho_1}_{B^s_{pq}}$$

holds, $j \geqq 1$. The case $j = 0$ can also be included, and the proof is easier than for $j \geqq 1$.

$(7.12^{'}, 53)$ now imply

$$(7.21') \quad \left(E_{F^N} \|\widehat{h} - \widehat{h}_{+,\varepsilon,\delta,\delta_1}\|^{\rho}_{L_{p_1}} \right)^{1/\rho_1} \ \leqq \ 2(3+2^{\rho/\rho_1}) \ K^{1/\rho_1} \ \alpha \left(\frac{1}{N}\right)^{1/\rho} \ +$$
$$+ \left(E_{F^N} \|\widehat{\mathcal{J}}\|^{\rho}_{L_{p_1}} \right)^{1/\rho_1} \ + \ \left(E_{F^N} \|\widehat{\mathcal{J}}_1\|^{\rho}_{L_{p_1}} \right)^{1/\rho_1} \ ,$$

where

$$(7.28') \quad \left(E_{F^N} \|\widehat{\mathcal{J}}\|^{\rho}_{L_{p_1}} \right)^{1/\rho_1} \ \leqq \ \left[2\varepsilon^{-(1-1/p_1)} \ R^{1-1/p} \right]^{\rho/\rho_1} \ . \ C_{11}(p,q,\sigma,s,r,\rho,L) \ .$$
$$\cdot \left[\|\Phi\|^{\rho/\rho_1}_{L_{p_1}} \ . \ \delta_1^{\min\{\sigma, \ r_1\}} \ \varepsilon^{\min\{\sigma, \ r_1\}} \ + \right.$$
$$\left. + \sum_{j=0}^{r_1} \ \|\Phi^{(r_1-j)}\|^{\rho/\rho_1}_{L_{p_1}} \ . \ \delta_1^{\min\{r_1-j+\sigma, \ r_1\}} \ \varepsilon^{\min\{\sigma, \ j\}} \right] \ \leqq$$
$$\leqq \ \left[2\varepsilon^{-(1-1/p_1)} \ R^{1-1/p} \ \|\Phi\|_{W^{r_1}_{p_1}} \right]^{\rho/\rho_1} \ . \ C_{12}(p,q,\sigma,s,r,\rho,L) \ .$$
$$\cdot \ \max_{j=0,1,\ldots, \ r_1} \ \left\{ \delta_1^{\min\{r_1+\sigma-j, \ r_1\}} \ \varepsilon^{\min\{\sigma, \ j\}} \right\}^{\rho/\rho_1} \ =$$
$$= \ \left(\ R^{1-1/p} \ \|\Phi\|_{W^{r_1}_{p_1}} \right)^{\rho/\rho_1} \ . \ C_{13}(p,q,\sigma,s,r,\rho,L) \ .$$
$$\cdot \left(\varepsilon^{-(1-1/p_1)} \ . \ \max\left\{ \delta_1^{\min\{\sigma, \ r_1\}} \ \varepsilon^{\min\{\sigma, \ r_1\}}, \ \delta_1^{r_1} \right\} \right)^{\rho/\rho_1} \ =$$
$$= \ C_{13}(p,q,\sigma,s,r,\rho,L) \ \left(\ R^{1-1/p} \ \|\Phi\|_{W^{r_1}_{p_1}} \right)^{\rho/\rho_1} \ .$$
$$\cdot \ \left[\varepsilon^{-(1-1/p_1)} \ \delta_1^{\min\{\sigma, \ r_1\}} \ \max\left\{ \varepsilon^{\min\{\sigma, \ r_1\}}, \ \delta_1^{(r_1-\sigma)_+} \ \right\} \ \right]^{\rho/\rho_1} \ = \ T_7$$

Now determine $\sigma^* = \sigma^*(1/N)$ such that simultaneously

$$(7.31') \quad \varepsilon_N^{\min\{\sigma, \ r_1\}} \ = \ \delta_{1,N}^{(r_1-\sigma)_+}$$
$$\varepsilon_N^{-(1-1/p_1)} \ \delta^{r_1}_{1, \ N} \ = \ C_{14}^{\rho_1/\rho} \ . \ \alpha(1/N)^{1/\rho}$$

where

$$(7.30') \quad C_{14} \ = \ C_{14}(p,p_1,q,\rho,r,r_1,\sigma,s,K,L) \ =$$

$$= \frac{2(3+2^{\rho/\rho_1}) \, . \, K^{1/\rho_1}}{C_{13}(p,q,\sigma,s,r,\rho,L) \, \left(R^{1-1/p} \, \|\Phi\|_{W^{r_1}_{p_1}} \right)^{\rho/\rho_1}} =$$

$$= \frac{2(3+2^{\rho/\rho_1}) \, . \, K^{1/\rho_1}}{\left[C_{15}(p,q,\sigma,s,\rho,L) \;\; R^{1-1/p} \, . \, \|\Phi\|_{W^{r_1}_{p_1}} \, . \, L \right]^{\rho/\rho_1}}$$

$(7.31')$ yields

(7.31″) $\min\{\sigma^*, \; r_1\} \, . \, \log_N \varepsilon_N \; = \; (r_1 - \sigma^*)_+ \, . \, \log_N \delta_{1,\,N}$.

If $\sigma^* \geqq r_1$, then $r_1 . \log_N \varepsilon_N = 0$, $\forall N > N_1$, and since $r_1 > 0$, it follows that $\log_N \varepsilon_N = 0$, i.e., $\varepsilon_N = 1$, $\forall N > N_1$. But $\lim\limits_{\varepsilon_N \to 0} \; \varepsilon_N = 0$.

Contradiction. Therefore, $\sigma^* < r_1$. Now $(7.31'')$ implies

(7.32′) $\sigma^* \; = \; r_1 \; \cdot \; \dfrac{\log_N \delta_{1,\,N}}{\log_N \varepsilon_N \; + \; \log_N \delta_{1,\,N}} \qquad (< r_1)$.

$(7.32', 31')$ implies (3.51,53,54). □

Remark 7.2. *(3.48′,50′,51′) can be obtained relying on the same arguments as in Remark 7.1.*

Proof of Theorem 3.4. Both Theorem 3.4 and Theorem 3.2 can be obtained as limiting cases of Theorem 3.3 and its proof. These limiting cases are $I = \emptyset$ and $I_1 = \emptyset$, respectively. □

Proof of Corollary 4.1. Under the assumption $r \neq 2 + (1/p - 1/p_1)_+$, it follows

(7.53) $\beta(\delta) = \delta^B$.

$(3.51, 51')$ now imply

(7.54) $\delta_{1,N} = O\{\alpha(1/N)^{\frac{1}{\rho}(1+\frac{1-1/p_1}{B})[\frac{1}{r}+(\frac{1}{\sigma}-\frac{1}{\sigma^*})_+]}\}$,

(7.55) $\delta_N = O\{\alpha(1/N)^{\frac{1}{\rho}(1+\frac{1-1/p_1}{B})[1+(\frac{1}{s_0}-\frac{1}{s_0^*})_+]}\}$,

where $\rho = 2$ for $\hat{h}^{(w)}$ and $\rho = p_1$ for $\hat{h}^{(k)}$; $s_0 = \sigma = A$; for $\hat{h}^{(w)}$, $\alpha(1/N)$ is as of Corollary 3.1.1 in [DM1]; for $\hat{h}^{(k)}$, $\alpha(1/N)$ is as of Corollary 3.2.1 in the same paper.

After computations,

(7.56) $(1 + \frac{1-1/p_1}{B}).[\frac{1}{r} + (\frac{1}{A} - \frac{1}{\sigma^*})_+] = \tau_1$,

(7.57) $(1 + \frac{1-1/p_1}{B}).[1 + (\frac{1}{A} - \frac{1}{s_0^*})_+] = \tau$. □

Proof of Corollary 4.2. The highest rate for $\alpha(1/N)^{1/\rho}$ is for $s = r$:

(7.58) $\alpha(1/N)^{1\rho} = O(N^{-\frac{r}{2r+1}})$.

The lowest possible rate for $\beta(\delta)$ is for $s = r_0$:

(7.59) $\beta(\delta) = O(\delta^{r_0})$.

s_0^* and σ^*, as computed in $(3.54, 38)$, now depend on s, and increase together with s. Therefore, for fixed s_0 and σ,

(7.60) $\quad 1 + (\frac{1}{s_0} - \frac{1}{s_0^*})_+ \ , \ \ \frac{1}{r_1} + (\frac{1}{\sigma} - \frac{1}{\sigma^*}{}_+$

also increase together with s, so they are biggest for $s = r$.

On the other hand, $(1 - 1/p_1)/B$ decreases with the increase of s, and is bounded from above by its value at $s = r_0$, uniformly in all admissible choices of s_0 and σ. Combining $(7.58 - 60)$ yields $(4.10 - 12)$, and (A, B) are fulfilled. □

Proof of Corollary 4.3. Follows from Corollary 4.2 by noticing that the quantities in (7.60), as well as $(1 - 1/p_1)/B$, all attain their maximal values at $p_1 = \infty$. □

8 Concluding remarks

The results in this article were considered only in the univariate case. It is possible to extend them to the multivariate case, too, but there are additional facts to be taken into consideration, for example,

- integration in $\widehat{h}_{+,\varepsilon}$ can be with respect to the ε–cube in $\mathbb{R}^d$, instead of the ε–ball, which makes it possible to reduce the d–dimensional integral to a d–tuple integral, which is particularly easy to solve for wavelets (recall that there is a wavelet basis in $\mathbb{R}^d$ which is comprised by tensor products of univariate functions). Additional difficulties appear in the computation of $\widehat{h}_+$. In any case computational complexity increases dramatically with the increase of d, whether d–dimensional quadratures are involved, or the problem is somehow reduced to a sequence of univariate quadratures.
- in the present paper we solve all problems entirely within the framework of Besov spaces. This is possible *only* for $d = 1$. However, if asymptotic theory of $\widehat{h}_{+,\varepsilon}$ is studied in A–*spaces*, then the results in one and several dimensions can be formulated in an unified way. This indicates that the scale of A–spaces may be the "true" one in which asymptotic theory of wavelet estimators should be studied. Recall that for $s > d/p$ the two scales of spaces coincide.

Generalizations of the result from the present paper, when v and or w are no longer constants, but functions, will be included in [DMP]. They are much more technically involved.

Acknowledgment: This research has been supported by the Natural Sciences and Engineering Research Council of Canada and by the Australian Research Council. The authors also wish to thank GERAD, where parts of this paper were written in a very hospitable atmosphere while the third author was visiting the first two in Montreal.

References

[A] R.A. Adams. *Sobolev Spaces.* Academic Press , 1975.

[BZ] I.S. Berezim, N.P. Zhidkov. *Computing Methods.* Pergamon Press, 1965.

[BL] J. Bergh. Löfström. *Interpolation Spaces. An Introduction.* Grundlehren der math. Wiss., 223. Springer, 1976.

[Chui] C.K. Chui. On Cardinal spline wavelets. In: Ruskai *et al.* (Eds.). *Wavelets and their Applications.* Jones & Barlett, (1992), 419-438.

[Daub] I. Daubechies. *Ten Lectures on Wavelets.* SIAM, 1992.

[DR] P.J. Davis, P. Rabinowitz. *Methods of Numerical Integration.* Academic Press, 1975.

[D] L.T. Dechevski. *Ph.D. Dissertation.* Sofia University, 1988.

[D1] L.T. Dechevski. τ-Moduli and interpolation. In: Peetre *et al.* (Eds.). *Theory of Function Spaces and Applications.* Lecture Notes in Math., 1302. Springer, (1988), 177-190.

[DM1] L.T. Dechevsky, B. MacGibbon. Asymptotically minimax non-parametric function estimation with positivity constraints I. Preprint, 1998.

[DM2] L.T. Dechevsky, B. MacGibbon. Asymptotically minimax non-parametric function estimation with positivity constraints II. (In preparation).

[DMP] L.T. Dechevsky, B. MacGibbon, S.I. Penev. Numerical methods for asymptotically minimax non-parametric function estimation with positivity constraints II. (In preparation).

[DP1] L.T. Dechevsky, S.I. Penev. On shape-preserving probabilistic wavelet approximators. Stochast. Anal. and Appl., 15(2) (1997), 187-215.

[DP2] L.T. Dechevsky, S.I. Penev. On shape-preserving wavelet estimators of cumulative distribution functions and densities. Stochast. Anal. and Appl., 16(3) (1998), 428-469.

[DJ] B. Delyon, A. Juditsky. On minimax wavelet estimators. ACHA, 3 (1996), 215-228.

[DoJo] D.L. Donoho, I.M. Johnstone. Minimax estimation via wavelet shrinkage. (To appear in The Annals of Statistics).

[JS] H. Johnen, K. Scherer. On the equivalence of the K-functional and moduli of continuity and some applications. In: Schempp *et al.* (Eds.) *Constructive Theory of Functions of Several Variables.* Lecture Notes in Math., 571. Springer (1977), 119-140.

[LMS] O.V. Lepski, E. Mammen, V.G. Spokoiny. Optimal spatial adaptation to inhomogeneous smoothness: an approach based on kernel estimators with variable bandwidth selectors. The Annals of Statist., 25(3) (1977), 929-947.

[N] S.M. Nikol'skiĭ. *Approximation of Functions of Several Variables and Imbedding Theorems.* Grundlehren der math. Wiss., 205. Springer, 1975.

[PP] P.P. Petrushev, V.A. Popov. *Rational Approximation of Real Functions.* Cambridge Univ. Press, 1987.

[RR] A. Ralston, P. Rabinowitz. *A First Course in Numerical Analysis.* McGraw-Hill, 1978.

[SP] Bl. Sendov, V.A. Popov. *The Averaged Moduli of Smoothness: Applications in Numerical Methods and Approximation.* Wiley, 1988.

[SP1] Bl. Sendov, V.A. Popov. *Numerical Methods, v. 1.* Nauka i Izkustvo, 1976. (In Bulgarian).

[T1] H. Triebel. *Theory of Function Spaces.* Monogr. in Math., 78. Birkhäuser, 1983.

[T2] H. Triebel. *Theory of Function Spaces II.* Monogr. in Math., 84. Birkhäuser, 1992.

[W] G.G. Walter. *Wavelets and Other Orthogonal Systems with Applications.* CRC Press, 1994.