ASYMPTOTICALLY MINIMAX NON–PARAMETRIC FUNCTION ESTIMATION WITH POSITIVITY CONSTRAINTS I

Lubomir Dechevsky
Université de Montréal
Département de mathématiques et de statistique
dechevsk@DMS.UMontreal.CA

Brenda MacGibbon
GERAD et Université du Québec à Montréal
Département de mathématiques
brenda@math.uqam.ca

Cahier de GÉRAD G-99-24
avril 1999

Abstract

The spatially inhomogeneous smoothness of the non–parametric density or regression–function to be estimated by non–parametric methods is often modelled by Besov- and Triebel–type smoothness constraints. For such problems, Donoho and Johnstone (1992), Delyon and Juditsky (1993) studied minimax rates of convergence for wavelet estimators with thresholding, while Lepskii, Mammen and Spokoiny (1995) proposed a variable bandwidth selection for kernel estimators that achieved optimal rates over Besov classes. However, a second challenge in many real applications of non–parametric curve estimation is that the function must be positive. Here, we show how to construct estimators under positivity constraints that satisfy these constraints and also achieve minimax rates over the appropriate smoothness class.

Résumé

Le lissage spatial non-homogène d'une densité non-paramétrique ou d'une fonction de régression qui doit être estimé par des méthodes non-paramétriques est souvent modelisé avec des contraintes de lissage de type Besov ou Triebel. Pour de tels problèmes, Donoho et Johnstone (1992), Delyon et Juditsky (1993) ont étudiés le taux de convergence minimax pour des estimateurs d'ondelettes avec des seuils, alors que Lepskii, Mammen and Spokoiny (1995) ont proposés une sélection de bande passante variable pour un estimateur à noyau qui atteint un taux optimal pour les classes de Besov. Toutefois un deuxième défi dans plusieurs applications réelles d'estimation de courbes non-paramétriques est représenté par le fait que les fonctions doivent être positives. Ici, on montre comment construire des estimateurs sous la contrainte de positivite qui satisfont ces contraintes et aussi atteignent le taux minimax pour le lissage des fonctions des classes appropriées.

Mathematics Subject Classification: Primary 62G07; Secondary 41A65, 46B40, 46B42, 46B70, 46E30, 46E35, 62G15

Key Words and Phrases: wavelet estimator, kernel estimator, estimator with positivity constraints, asymptotic-minimax optimal rate, Besov space, Triebel-Lizorkin space, Wiener-Young bounded variation

1 Introduction

Our goal is to develop methods that yield estimators which are asymptotically optimal for non–parametric density and regression function estimation when faced with both the problem of spatially inhomogeneous smoothness as well as order constraints. The problems of "smooth monotone density estimation" and "isotonic regression" come immediately to mind. Here we solve the simplest problem when the "spatially inhomogeneous smooth" function must be non-negative, or, more generally bounded from above and/or below by known constants. The methodology presented here works for d-variate real-valued functions, $d = 1, 2, \ldots$.

In the case of density estimation there have been some recent results about wavelet–based estimators in [Penev *et al.* (1997)] and [Pinheiro *et al.* (1997)] which preserve positivity, perform well on moderate samples, and are relatively computationally inexpensive. However, they are narrowly specialized to preserve only a one–sided restriction, and this restriction has to be a constant, typically $v(x) \equiv 0$. Moreover, asymptotic results are limited to the Hellinger metric. Our study of the general problem shows that already the case where both upper and lower constraints of the form $w(x) \equiv c_1$, $v(x) \equiv c_0$, with known constants $c_0 < c_1$, are given, is essentially more difficult; its study is not amenable to these previously published techniques. In this paper we shall restrict our considerations to this latter case.

More precisely, for an unknown density or regression function f whose graph is *a priori* known to be bounded by those of v and/or w, given an asymptotically optimal 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 more general case when v and/or w is a function, can also be studied by the methods developed here, but because of their technical complexity they will not be included here and will be the subject of subsequent work.

The standard unconstrained non-parametric function estimation problem is to estimate an unknown function f on the basis of N (not necessarily uncorrelated) noisy observations

$$y_i = f(t_i) + \sigma\varepsilon_i,\ i = 1, \ldots, N,$$

where σ and ε_i represent variance and error terms, respectively. There are many different approaches to this estimation problem, but we will concentrate on two of them which are known to produce asymptotically optimal results if the regularity conditions on the function can be expressed in terms of Besov classes. These classes contain the well-known Sobolev and Hölder classes but, in addition, the spatially inhomogeneous class of functions of bounded variation can be sandwiched between two Besov spaces and asymptotic optimality results can also be deduced for this class (see Donoho and Johnstone (1998)). Donoho and Johnstone (1998), Delyon and Juditsky (1996) studied minimax rates of convergence for wavelet estimators with thresholding, while Lepski, Mammen and Spokoiny (1997) proposed a variable bandwidth selector for kernel estimators. Both types of estimators achieve optimal rates over Besov classes.

We can outline our method as follows. If $\hat{f}$ is one of the above-mentioned asymptotically minimax estimators, then we can construct $\hat{f}^+$ which also achieves the optimal rates with respect to Besov classes. Asymptotic minimax optimality of $\hat{f}^+$ is proved with respect to Triebel-Lizorkin and Wiener-Young functions classes, too. When $\hat{f}$ is the estimator proposed by Lepski *et al.* (which is ideally spatially adaptive, but may be non-smooth and even discontinuous at some points, i.e., need obey neither the *upper/lower* bounds nor the *smoothness* constraints), the corresponding $\hat{f}^+$ does obey *all* constraints.

Let us consider the case of a d-variate function f satisfying $f \geq v$ (v a constant). Let $\hat{f}$ denote an estimator that satisfies the "smoothness" criteria but not necessarily the positivity constraint. Clearly, $\hat{f}_+(x) := \max\{v, \hat{f}(x)\}$ will satisfy the positivity constraint but not necessarily the smoothness criteria. Consider the following

Definition 1. *The pos* $-\,\varepsilon$ *smooth of* $\hat{f}$*, denoted by* $\hat{f}_{+,\varepsilon}$*, is*

$$\hat{f}_{+,\varepsilon}(x) := \Phi_\varepsilon * \hat{f}_+(x), \tag{1}$$

where $*$ *denotes the convolution of* $\hat{f}_+$ *with a "suitably smooth" approximate identity* Φ_ε*; that is,*

$$\Phi_\varepsilon(x) := \frac{1}{\varepsilon^d}\Phi(\frac{x}{\varepsilon}),\ x \in \mathbb{R}^d,\ \varepsilon > 0,$$

where Φ *is a "suitably smooth" symmetric positive Lebesgue measurable function with integral on* $\mathbb{R}^d$ *equal to one (See Appendix 0 for details and examples).*

It is this "pos-ε smooth" of f which will be shown to be the desired estimator $\hat{f}^+$ obeying the smoothness and order constraints and achieving the optimal asymptotic rates.

The paper is organized as follows. The introduction constitues Section 1. All preliminary notation and definitions which are not explained in the main text are given in Appendix 0. In section 2, we present the main results about $\hat{f}^+ = \hat{f}_{+,\varepsilon}$. In Section 3, these results are applied to wavelet estimators and variable–bandwidth kernel estimators for density and non-parametric regression estimation problems. The proofs are given in Section 4. Section 5 contains some remarks including directions for future research.

2 A general result about constrained estimation

Here we present a method for adjusting asymptotically optimal estimators in order to satisfy positivity constraints under very general conditions. The main result can be stated as follows:

Theorem 1. *Let* $f : \mathbb{R}^d \to \mathbb{R}$ *be such that there exists a constant* v *with* $f(x) \geqq v$ *a.e. with respect to Lebesgue measure. We have* N *observations* $X_1, \ldots, X_N$ *with c.d.f.* F^N *and let* $\hat{f} = \hat{f}_N : \mathbb{R}^d \to \mathbb{R}$ *denote an estimator of* f *based on these observations* X_j*,* $j = 1, \ldots, N$*;*

We assume $f \in L_{p_1}(\mathbb{R}^d)$*,* $\hat{f} \in \mathcal{L}_\rho(\mathbb{R}^d, L_{p_1}(\mathbb{R}^d))^\rho$ *with* $1 \leqq p_1 \leqq \infty$*,* $0 < \rho < \infty$*.*

Let W *denote a (quasi-)Banach space such that*

$$W \ \hookrightarrow\ L_{p_1}(\mathbb{R}^d) + \dot{W}^2_{p_1}(\mathbb{R}^d),$$

and such that there exists $\delta_0 > 0$, $\beta \in \Omega(\delta_0)$, satistfying

$$\omega_2(g;\delta)_{L_{p_1}} \leqq \beta(\delta)\|g\|_W, \quad \forall g \in W \ \forall \delta \in (0,\delta_0]; \tag{2}$$

If we assume $f \in W$ and $\|f\|_W \leqq L$ for some $L \in (0,\infty)$, and that there exists $\alpha \in \Omega(\delta_0)$, $K = K(p_1,\rho,W,L,\alpha) \in (0,\infty)$ and $N_0 = N_0(p_1,\rho,W,L,\alpha) \in (0,\infty)$, such that

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

Consider $\hat{f}_{+,\varepsilon}$, as defined in (1).

Then,

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

where

$\rho_1 = \max\{1,\rho\}$,

$$\varepsilon_N = \beta^{-1}\left(\frac{2}{L} \cdot (3 + 2^{\rho/\rho_1})^{\rho_1/\rho} \ \cdot K^{1/\rho} \cdot \alpha\left(\frac{1}{N}\right)^{1/\rho}\right). \tag{4}$$

Remark 1. *If in Theorem 1, we let*

$v = 0$,

$\int_{\mathbb{R}^d} f(x)\,dx = 1$,

$X_1, \ldots, X_N$ *are i.i.d. with c.d.f.* $F : \mathbb{R} \to [0,1]$,

F is absolutely continuous relative to the Lebesgue measure in $\mathbb{R}^d$, with Radon-Nikodym derivative f, then this corresponds to (asymptotic) density estimation with i.i.d. observations.

The next result provides explicit computation of β when W in Theorem 1 is a Besov space.

Lemma 1. *Assume*

$d \in \mathbb{N}$,

$1 \leqq p_1 \leqq \infty$,

$1 \leqq p \leqq \infty$, $0 < q \leqq \infty$,

$d\left(\frac{1}{p} - \frac{1}{p_1}\right)_+ < s < \infty$. *Assume further that*

for $p \leqq p_1$ $W = B^s_{pq}(\mathbb{R}^d)$;

for $p_1 < p$ there exists $R : 0 < R < \infty$ such that $W = W_R = \{f \in B^s_{pq}(\mathbb{R}^d)$, $supp\ f \subset \{x \in \mathbb{R}^d : |x| \leqq R\}\}$ and W_R is endowed with the (quasi-)norm of $B^s_{pq}(\mathbb{R}^d)$.

Then,

there exists

$$K_1 = \begin{cases} K_1(p,q,s,d) \in (0,\infty), & p \leqq p_1 \\ K_1(p,q,s,d,R) \in (0,\infty), & p > p_1, \end{cases}$$

such that

(2) is fulfilled for any $\delta_0 \in (0,1]$*;*

$$(5) \qquad \beta(\delta) \leqq K_1 \cdot \begin{cases} \delta^{\min\{2, s-d\left(\frac{1}{p}-\frac{1}{p_1}\right)_+\}}, & s \neq 2+d\left(\frac{1}{p}-\frac{1}{p_1}\right)_+ \\ \delta^2 \left(ln\frac{1}{\delta}\right)^{\left(1-\frac{\min\{p,2\}}{q}\right)_+}, & s = 2+d\left(\frac{1}{p}-\frac{1}{p_1}\right)_+, p<\infty \\ \delta^2 \left(ln\frac{1}{\delta}\right)^{(1-1/q)_+}, & s = 2+d\left(\frac{1}{p}-\frac{1}{p_1}\right)_+, p=\infty. \end{cases}$$

Remark 2. *It can be proved that* W_R*, as defined in Lemma 1, is a closed subspace of the (quasi–) Banach space* $B^s_{pq}(\mathbb{R}^d)$ *and, therefore,* W_R *is a (quasi–) Banach space itself.*

Remark 3. *The additional condition for* W *in the case* $p > p_1$ *of Lemma 1 could be weakened in various ways, so that (5) continues to hold true. Our choice was motivated by the fact that the simple condition about compactness of the support of* f *is sufficient to cover all the results in [Delyon* et al. *(1996)], [Lepski* et al. *(1997)] on which our applications in section 3 are based.*

Remark 4. *The restriction* $\delta_0 \in (0,1]$ *in Lemma 1 is not essential. It can be replaced by "for any* $\delta_0 \in (0,\infty)$*" but then* K_1 *will have to depend on* δ_0*, too (not on* δ*!).*

The next result is an analogue of Lemma 1 when W in Theorem 1 is a Triebel-Lizorkin space. To avoid additional technical considerations about equivalent norms in the proof, we include here only the case $1 < p_1 < \infty$, $1 < p < \infty$.

Lemma 2. *Under the assumptions of Lemma 1, assume that*

for $1 \leqq p \leqq p_1 < \infty$ $W = F^s_{pq}(\mathbb{R}^d)$*;*

for $1 < p_1 < p < \infty$ $W = W_R = \{f \in F^s_{pq}(\mathbb{R}^d) \ : \ supp\, f \subset \{x \in \mathbb{R}^d \ : \ |x| \leqq R\}\}$*, endowed with the (quasi–)norm in* $F^s_{pq}(\mathbb{R}^d)$*, where* $0 < s < r$*,* $0 < q \leqq \infty$*.*

Then, the conclusion of Lemma 1 holds, with

$$(6) \qquad \beta(\delta) \leqq C_1 \cdot \begin{cases} \delta^{\min\{2, s-d\left(\frac{1}{p}-\frac{1}{p_1}\right)_+\}}, & s \neq 2+d\left(\frac{1}{p}-\frac{1}{p_1}\right)_+, \ 0<q<\infty \\ \delta^2 & s = 2+d\left(\frac{1}{p}-\frac{1}{p_1}\right)_+, \ 0<q\leqq 2. \end{cases}$$

Remark 5. *As in Lemma 1, in the case* $s = 2+d\left(\frac{1}{p}-\frac{1}{p_1}\right)_+$*,* $2 < q \leqq \infty$*, an additional factor appears which is logarithmic in* $1/\delta$*. The proof is, however, more technically involved than with Besov spaces and will not be discussed here.*

The next lemma provides an analogous computation of β when W is (essentially) $T\vee_p^k$, $1 \leqq p < \infty$, $k = 0, 1, 2, \ldots$

Lemma 3. *Under the assumptions of Lemma 1, assume that*

$d = 1$;

$k = 0, 1, 2, \ldots$;

for $1 \leqq p \leqq p_1 < \infty$ $\quad W = L_p \curvearrowright (T\vee_p^k)^{1/p}$;

for $1 \leqq p_1 \leqq p < \infty$ $\quad W = W_R = \{f \in W_p^k \cap T\vee_p^k,\ supp\ f \subset [-R, R]\}$ and W_R is endowed with the norm in $W_p^k \curvearrowright (T\vee_p^k)^{1/p}$.

Then,

the conclusion of Lemma 1 holds, with

$$\beta(\delta) \;=\; \delta^{\min\{2,\, k + \frac{1}{\max\{p,\, p_1\}}\}}. \tag{7}$$

3 Applications

We shall consider applications to wavelet and kernel estimators, as considered in [Delyon *et al.* (1996)] and [Lepski *et al.* (1997)]. In all these cases $X_1, \ldots, X_N$ are independent, identically distributed from a c.d.f. $F : \mathbb{R}^d \to \mathbb{R}$ and, therefore,

$$dF^N(x_1, \ldots, x_N) = dF(x_1) \ldots dF(x_N),\ x_\nu \in \mathbb{R}^d.$$

3.1 Wavelet estimators

In this subsection σ and r are defined as in [Delyon *et al.* (1996)];

Corollary 1. *Assume that:*

for f and $\hat{f} = \hat{f}_N$ the hypotheses of Theorem 1 in [Delyon et al. *(1996)] are fulfilled, with $\sigma = 0$;*

$\hat{f}_{+,\varepsilon}$ is defined as in (1), with $\Phi \in C_0^r(\mathbb{R}^d)$;

W is as in Lemma 2.1.

Then,

the conclusions of Theorem 2.1 hold, where

$\rho = \rho_1 = 2$,

β is defined via (5);

(A) for p_1 : $1 \leqq p \leqq p_1 < \infty$

K is as in Theorem 1 in [Delyon et al. *(1996)] for $p_1 < \infty$;*

$$\alpha\left(\frac{1}{N}\right) = \begin{cases} \left(\frac{1}{N}\right)^{2\frac{s}{2s+d}}, & \theta > 0 \\ \left(\frac{lnN}{N}\right)^{2\frac{s-\frac{d}{p}+\frac{d}{p_1}}{2\left(s-\frac{d}{p}\right)+d}} (lnN)^{2\left(1-\frac{2p}{qp_1}\right)_+}, & \theta = 0 \\ \left(\frac{lnN}{N}\right)^{2\frac{s-\frac{d}{p}+\frac{d}{p_1}}{2\left(s-\frac{d}{p}\right)+d}}, & \theta < 0, \end{cases}$$

where

$$\theta = \frac{s}{p_1} - \frac{d}{2}\left(\frac{1}{p} - \frac{1}{p_1}\right); \tag{8}$$

(B) for $p_1 = \infty$

K is as in Theorem 1 in [Delyon et al. *(1996)] for $p = \infty$;*

$$\alpha\left(\frac{1}{N}\right) = \left(\frac{lnN}{N}\right)^{2\frac{s-d/p}{2(s-d/p)+d}}$$

The choice of ε_N in Corollary 1 is determined via (4). Therefore, it depends on the explicit knowledge of p, p_1, q and s. The next statement suggests a (smaller) value of ε_N so that the conclusions of Corollary 1 hold simultaneously for a broad range of the relevant parameters, uniformly in $\varepsilon \in (0, \varepsilon_N]$.

Corollary 2. *Under the hypotheses of Corollary 1 (A) or (B), respectively, assume that, instead of being computed via (4) for concrete values p_1, p, q and s,*

ε_N is such that there exists $K_2 : 0 < K_2 < \infty$, so that

$$\varepsilon_N = K_2 \cdot N^{-\frac{r}{r_0(2r+d)}},$$

where $r_0 : 0 < r_0 < 2$, $r : \max\{2, r_0 + d\} < r$.

Then,

there exists $N_1 = N_1(r_0, r, L, K_2)$ such that, for all $N > \max\{N_0, N_1\}$,

the conclusions of Corollary 1 (A) hold simultaneously for all $p, p_1 : 1 \leqq p \leqq p_1 < \infty$, and

the conclusions of Corollary 1 (B) hold simultaneously for all $p : 1 \leqq p \leqq p_1 = \infty$, respectively uniformly in

$q : 0 < q \leqq \infty$,

$\varepsilon : 0 < \varepsilon \leqq \varepsilon_N$,

$s : r_0 + d\left(\frac{1}{p} - \frac{1}{p_1}\right)_+ \leqq s \leqq r$.

In the case $p_1 = 2$ there are simpler and somehow more precise versions of Corollaries 1 and 2.

Corollary 3. *Assume that for f and $\hat{f} = \hat{f}_N$ the hypotheses of Theorem 2 in [Delyon* et al. *(1996)] are fulfilled, with $\sigma = 0$. $\hat{f}_{+,\varepsilon}$ is defined as in Corollary 1, $p_1 = 2$. Then, the conclusions of Theorem 2.1 hold, where $\rho = \rho_1 = 2$, β is defined via (5), K is defined as in Theorem 2 in [Delyon* et al. *(1996)],*

$$\alpha\left(\frac{1}{N}\right) = \left(\frac{1}{N}\right)^{2\frac{s}{2s+d}}.$$

Corollary 4. *Under the premises of Corollary 3, assume that, instead of being computed via (4) for concrete values of p, q and s,*

ε_N is such that there exists $K_2' : 0 < K_2' < \infty$ with

$$\varepsilon = K_2' \begin{cases} N^{-\frac{r}{\min\{2,\, \max\{d/2,\, r_0\}\} \cdot (2r+d)}}, & d \neq 4 \\ \beta_1^{-1}\left(\frac{1}{N}\right), & d = 4 \end{cases},$$

where $r_0 : 0 < r_0 < 2$, $r : \max\{2, r_0 + \frac{d}{2}\} < r$, $\beta_1(\delta) = \delta^2 ln \frac{1}{\delta}$.

Then,

there exists $N_1 = N_1(r_0, r, L, K_2')$ such that, for all $N > \max\{N_0, N_1\}$,

the conclusions of Corollary 3 hold for all $p : 1 \leqq p \leqq 2 = p_1$,

uniformly in

$q : 0 < q \leqq \infty$,

$\varepsilon : 0 < \varepsilon \leqq \varepsilon_N$,

$s : r_0 + d\left(\frac{1}{p} - \frac{1}{p_1}\right)_+ \leqq s \leqq r.$

Remark 1. *The upper bound for the constant K in (3) obtained via Corollaries 3, 4 is much smaller than the one obtained via Corollaries 1, 2 for $p_1 = 2$, because the constants in Theorem 2 in [Delyon* et al. *(1996)] are much better than the ones obtained from Theorem 1 in [Delyon* et al. *(1996)] by setting $p_1 = 2$.*

Corollary 5 (Density estimation). *Assume that*

for f and $\hat{f} = \hat{f}_N$ the hypotheses of Theorem 3 in [Delyon et al. *(1996)] are fulfilled, with $\sigma = 0$;*

$\hat{f}_{+,\varepsilon}$ is defined as in Corollary 1.

Then,

the conclusions of Corollaries 1 (A) and (B) hold, with

K as in Theorem 3 in [Delyon et al. *(1996)].*

Corollary 6. *Under the assumptions of Corollary 5, for ε_N as in Corollary 2, the conclusions of Corollary 2 hold, with K as in Theorem 3 in [Delyon* et al. *(1996)].*

Corollary 7 (Nonparametric regression). *Assume that for f and $\hat{f} = \hat{f}_N$ the hypotheses of Theorem 4 in [Delyon* et al. *(1996)] are fulfilled, with $\sigma = 0$, $\hat{f}_{+,\varepsilon}$ is defined as in Corollary 1. Then, the conclusions of Corollaries 1 (A) and (B) hold, with K as in Theorem 4 in [Delyon* et al. *(1996)].*

Corollary 8. *Under the assumptions of Corollary 7, for ε_N as in Corollary 2, the conclusions of Corollary 2 hold, with K as in Theorem 4 in [Delyon* et al. *(1996)].*

Remark 2. *Note that in Theorem 3, 4 in [Delyon* et al. *(1996)] supp $f \subset [0,1]^d$ holds.*

Next, we consider a model analogue of Corollary 1 when f belongs to a Triebel–Lizorkin space. We shall restrict the consideration to $d = 1$, $1 < p < \infty$, $p_1 = 2$. More general results can be obtained in the same spirit, but would require more varied techniques in the proofs.

Corollary 9. *Under the hypotheses of Corollary 1, assume that W is as in Lemma 2.2 for $d = 1$, $p_1 = 2$, $1 < p < \infty$, $0 < q \leqq \infty$, $r > \frac{1}{\min\{2,p,q\}}$, $\frac{1}{\min\{2,p,q\}} < s < r$. Then, the conclusions of Theorem 2.1 hold, with $\rho = \rho_1 = 2$; β defined via (6).*

Corollary 10. *Under the hypotheses of Corollary 9, assume that, instead of being computed via (4) for concrete values $p_1 = 2$, p, q and s, ε_N is such that there exists $C : 0 < C < \infty$ such that*

$$\varepsilon_N = C \cdot N^{-\frac{2r}{2r+1}}.$$

Then, there exists $N_1 = N_1(r, L, C)$ such that, for all $N > \max\{N_0, N_1\}$, the conclusions of Corollary 9 hold simultaneously for $p : 1 < p < \infty$, $q : 0 < q \leqq \infty$, $\varepsilon : 0 < \varepsilon \leqq \varepsilon_N$, $s : \frac{1}{\min\{2,p,q\}} < s < r$.

Finally, let us consider spaces of functions with bounded variation.

Corollary 11. *Under the hypotheses of Corollary 1, assume that W is as in Lemma 2.3, $d = 1$, $p_1 = 2$, $1 \leqq p < \infty$, $k = 0, 1, 2, \ldots$ Then, the conclusions of Theorem 2.1 hold, with $\rho = \rho_1 = 2$, β defined via (7).*

Because of the discreteness of the range of k and Lemma 0.7 we do not consider a "uniform" analogue of Corollary 11.

3.2 Kernel estimators

In this subsection we consider the white-noise model in the setting of [Lepski *et al.* (1997)]. For density estimation with i.i.d. observations and nonparametric regression with i.i.d. Gaussian errors this model corresponds to the models in sections 3.2 and 3.3 of [Delyon *et al.* (1996)], respectively, with $d = 1$ and σ in [Lepski *et al.* (1997)] corresponding to $\frac{C}{\sqrt{N}}$, $C > 0$ in [Delyon *et al.* (1996)]. (The meanings of σ in these two references are completely unrelated!) We note that, as in sections 3.2 and 3.3 in [Delyon *et al.* (1996)], in [Lepski *et al.* (1997)] the condition supp $f \subset [0,1]$ is required (see (2.1) in [Lepski *et al.* (1997)]).

In [Lepski *et al.* (1997)] the Banach case $1 \leqq p \leqq \infty$, $1 \leqq q \leqq \infty$ is assumed, with $p_1 : 1 \leqq p_1 < \infty$, $s > \left(\frac{1}{p} - \frac{1}{p_1}\right)_+$, $d = 1$ and the case $p_1 < p$ is also included.

Corollary 1. *Suppose that for f and $\hat{f}$ the assumptions of Theorem 1 in [Lepski* et al. *(1997)] are fulfilled, $\hat{f}_{+,\varepsilon}$ is defined as in (1), with $\Phi \in C_0^r(\mathbb{R})$. Then, the conclusions of Theorem 2.1 hold, where $\rho = \rho_1 = p_1$, β is defined via (3), K is as in Theorem 1 in [Lepski* et al. *(1997)],* $1 \leqq p_1 < \infty$, *$d = 1$,*

$$\alpha\left(\frac{1}{N}\right) = \begin{cases} \left(\frac{1}{N}\right)^{p_1 \frac{s}{2s+d}}, & \theta > 0 \\ \left(\frac{1}{N}\right)^{p_1 \frac{s-1/p+1/p_1}{2(s-1/p)+1}} \ (lnN)^{p_1\left[\frac{1}{2}\cdot\frac{s-1/p+1/p_1}{2(s-1/p)+1}+\frac{1}{p_1}\right]}, & \theta = 0 \\ \left(\frac{1}{N}\right)^{p_1 \frac{s-1/p+1/p_1}{2(s-1/p)+1}} \ (lnN)^{\frac{p_1}{2}\ \frac{s-1/p+1/p_1}{2(s-1/p)+1}}, & \theta < 0 \end{cases},$$

where θ is as defined in (8).

Corollary 2. *Suppose that for f and $\hat{f}$ the assumptions of Theorem 2 in [Lepski* et al. *(1997)] are fulfilled; $\hat{f}_{+,\varepsilon}$, ρ are as in Corollary 1, K is as in Theorem 2 in [Lepski* et al. *(1997)]. Suppose also that, instead of being computed via (4) for concrete values of p, q and s, ε_N is such that there exists $K_3 : 0 < K_3 < \infty$ so that*

$$\varepsilon_N = K_3 \cdot N^{-\frac{r}{r_0(2r+1)}},$$

where $r_0 : 0 < r_0 < 2$, $r : \max\{2, r_0 + 1\} < r$. Then, the conclusions of Corollary 3.1.2 are fulfilled, with $d = 1$, $p : 1 \leqq p \leqq \infty$, $p_1 : 1 \leqq p_1 < \infty$, $q : 1 \leqq q \leqq \infty$, $\varepsilon : 0 < \varepsilon \leqq \varepsilon_N$, $s : r_0 + \left(\frac{1}{p} - \frac{1}{p_1}\right)_+ \leqq s \leqq r$.

Remark 1. *Note that $\hat{f}$, as defined in [Lepski* et al. *(1997)], may be non–smooth, even discontinuous. Thus, $\hat{f}_{+,\varepsilon}$ in this case plays a twofold role – it preserves positive order and improves upon the regularity of $\hat{f}$.*

3.3 Asymptotic - minimax optimality of constrained estimation

The rates in Theorems 1-4 in [Delyon *et al.* (1996)] and in Theorems 1, 2 in [Lepski *et al.* (1997)] are optimal within the "asymptopia" almost optimality paradigm in [Donoho *et al.* (1995)]. Moreover, for $\theta \neq 0$ the results in Theorems 1, 2 in [Lepski *et al.* (1997)] are also sharp with respect to the logarithmic factor. The fact that these are upper minimax rates and that, obviously, constrained estimation cannot have better upper minimax rates, means that $\hat{f}_{+,\varepsilon}$, as defined in subsections 3.1, 3.2, is asymptotically optimal.

As a matter of fact, our main result - Theorem 2.1 – focuses our attention on a much more general observation: under very general assumptions about the space scale (including Besov, Sobolev and Triebel-Lizorkin scales) the upper asymptotic-minimax rates of constrained estimation for this space scale coincide, including the logarithmic factors, with the respective optimal rates for non-constrained estimation. Moreover, there exists a general constructive definition of a "constrained" estimator – $\hat{f}_{+,\varepsilon}$ – which achieves the optimal rates simultaneously for a broad range of function spaces and error measures.

In the forthcoming paper treating general non–constant constraints $v(x)$ and/or $w(x)$, we shall see that the above basic observation also continues to hold true in this general context.

4 Proofs

4.1 Preliminary lemmata

Lemmas 0.1-5 in Appendix 0 are easy and well-known. Let us, for completeness, outline the proofs of some of them.

Proof of Lemma 0.2.

$$\omega_k(f;\delta)_{L_p} = \sup_{|h|\leqq\delta} \left\| \sum_{\nu=0}^{k} \binom{k}{\nu} (-1)^{k-\nu} f(\cdot+\nu h) \right\|_{L_p} \leqq$$

$$\leqq \sup_{|h|\leqq\delta} \left(\sum_{\nu=0}^{k} \binom{k}{\nu} \left\| f(\cdot+\nu h) \right\|_{L_p} \right) = \left[\sum_{\nu=0}^{k} \binom{k}{\nu} \right] \|f\|_{L_p} = 2^k \|f\|_{L_p}.$$

□

Proof of Lemma 0.4. Direct application of the generalized Minkowski inequality (see, e.g., [Nikol'skiĭ (1975)]) which in the case of convolution is equivalent to Young's inequality (see [Bergh *et al.* (1976)], [Nikol'skiĭ (1975)], [Reed *et al.* (1975)]). □

Proof of Lemma 0.5. By the symmetry of Φ, $\int\limits_{\mathbb{R}^d} \Phi(x)\,dx = 1$, supp $\Phi_\varepsilon \subset \{|x| \leqq \varepsilon\}$.

$$\left\| f(\cdot) - \int\limits_{\mathbb{R}^d} \Phi_\varepsilon(\xi) f(\cdot - \xi)\,d\xi \right\|_{L_p} =$$

$$= \left\| f(\cdot) - \int\limits_{\mathbb{R}^{d-1}\times[0,\infty)} \Phi_\varepsilon(\xi) f(\cdot - \xi)\,d\xi - \int\limits_{\mathbb{R}^{d-1}\times[0,\infty)} \Phi_\varepsilon(\xi) f(\cdot + \xi)\,d\xi \right\|_{L_p} =$$

$$= \left\| \int\limits_{\mathbb{R}^{d-1}\times[0,\infty)} \Phi_\varepsilon(\xi) \Big(f(\cdot + \xi) - 2f(\cdot) + f(\cdot - \xi) \Big)\,d\xi \right\|_{L_p} \leqq \text{(by generalized Minkowski inequality)}$$

$$\leqq \int\limits_{\mathbb{R}^{d-1}\times[0,\infty)} \Phi_\varepsilon(\xi) \left\| \Delta_\xi^2 f(\cdot - \xi) \right\|_{L_p} d\xi = \int\limits_{\mathbb{R}^{d-1}\times[0,\infty)} \Phi_\varepsilon(\xi) \left\| \Delta_\xi^2 f(\cdot) \right\|_{L_p} d\xi \leqq$$

$$\leqq \omega_2(f;\varepsilon)_{L_p} \int\limits_{\mathbb{R}^{d-1}\times[0,\infty)} \Phi_\varepsilon(\xi)\,d\xi = \frac{1}{2}\omega_2(f;\varepsilon)_{L_p}.$$

□

The proofs of Lemma 0.6 and Lemma 0.7 are easy.

Proof of Lemma 0.8. First of all, the inequality

$$\omega_{k+1}(f;\delta)_p \leqq \delta^k\,\omega_1(f^{(k)};\delta)_p, \quad 1 \leqq p \leqq \infty,$$

is involved (see, e.g., [Petrushev *et al.* (1987)], [Sendov *et al.* (1988)]) Then one makes use of the inequality

$$\omega_1(g;\delta)_p \leqq \delta^{1/p}\,(\vee_p g)^{1/p}, \quad 1 \leqq p < \infty. \tag{9}$$

For $p = 1$ (9) is proved, e.g., in [Petrushev *et al.* (1987)], [Sendov *et al.* (1988)]. In the general case $1 \leqq p < \infty$ the proof of (9) can be carried out analogously. □

The right embedding in Lemma 0.9 follows directly from Lemma 0.8. The left one requires additional technical consideration and will not be discussed here.

4.2 Proofs of main results

Proof of Theorem 2.1. For $\rho_1 = \max\{1, \rho\}$, $\varepsilon > 0$, where for simplicity of notation here in the proof p replaces the p_1 in the statement of Theorem 2.1 in Section 2,

$$\|f - \hat{f}_{+,\varepsilon}\|^{1/\rho_1}_{\mathcal{L}_\rho\left(\mathbb{R}^d,\, L_p(\mathbb{R}^d)\ \right)^\rho} = \left(E_{F^N}\left\|(f-\hat{f}) + (\hat{f} - \hat{f}_\varepsilon) + (\hat{f}_\varepsilon - \hat{f}_{+,\varepsilon})\right\|^\rho_{L_p}\right)^{1/\rho_1} \leqq$$

$$\leqq \left(E_{F^N}\left\|f - \hat{f}\right\|^\rho_{L_p}\right)^{1/\rho_1} + \left(E_{F^N}\left\|\hat{f} - \hat{f}_\varepsilon\right\|^\rho_{L_p}\right)^{1/\rho_1} + \left(E_{F^N}\left\|\hat{f}_\varepsilon - \hat{f}_{+,\varepsilon}\right\|^\rho_{L_p}\right)^{1/\rho_1} =$$

$$= I_1 + I_2 + I_3,$$

where $\hat{f}_\varepsilon = (\hat{f})_\varepsilon$.

By assumptions about α, $\forall N > N_0$

$$I_1 \leqq K^{1/\rho_1} \cdot \alpha\left(\frac{1}{N}\right)^{1/\rho_1}.$$

By Lemma 0.5 and Lemma 0.1,

$$I_2 \leqq 2^{-\rho/\rho_1}\left(E_{F^N}\ \omega_2\big(\hat{f}, \varepsilon\big)^\rho_{L_p}\right)^{1/\rho_1} = 2^{-\rho/\rho_1}\left(E_{F^N}\ \omega_2\Big((\hat{f} - f) + f; \varepsilon\Big)^\rho_{L_p}\right)^{1/\rho_1} \leqq$$

$$\leqq 2^{-\rho/\rho_1}\left[\left(E_{F^N}\ \omega_2\big(\hat{f} - f; \varepsilon\big)^\rho_{L_p}\right)^{1/\rho_1} + \left(E_{F^N}\omega_2(f;\varepsilon)^\rho_{L_p}\right)^{1/\rho_1}\right] = 2^{-\rho/\rho_1}\,[I_{21} + I_{22}].$$

By Lemma 0.2 and by assumption about α

$$I_{21} \leqq 2^{2\rho/\rho_1} \cdot \left(E_{F^N}\left\|\hat{f} - f\right\|^\rho_{L_p}\right)^{1/\rho_1} \leqq 2^{2\rho/\rho_1}\ K^{1/\rho_1} \cdot \alpha\left(\frac{1}{N}\right)^{1/\rho_1}.$$

By assumptions about β, L

$$I_{22} = \omega_2\big(f;\varepsilon\big)^{\rho/\rho_1}_{L_p} \leqq \beta(\varepsilon)^{\rho/\rho_1} \cdot \big\|f\big\|^{\rho/\rho_1}_W \leqq \beta(\varepsilon)^{\rho/\rho_1} \cdot L^{\rho/\rho_1}.$$

Therefore,

$$I_2 \leqq 2^{\rho/\rho_1}\ K^{1/\rho_1} \cdot \alpha\left(\frac{1}{N}\right)^{1/\rho_1} + 2^{-\rho/\rho_1}\ \beta(\varepsilon)^{\rho/\rho_1}\,L^{\rho/\rho_1}.$$

By Lemma 0.3 and Lemma 0.4,

$$I_3 = \left(E_{F^N}\left\|(\hat{f} - \hat{f}_+)_\varepsilon\right\|^\rho_{L_p}\right)^{1/\rho_1} \leqq \left(E_{F^N}\left\|\hat{f} - \hat{f}_+\right\|^\rho_{L_p}\right)^{1/\rho_1}$$

$$= \left(E_{F^N}\left\|(\hat{f} - f) + (f - \hat{f}_+\right\|^\rho_{L_p}\right)^{1/\rho_1} \leqq$$

$$\left(E_{F^N}\|\hat{f} - f\|^\rho_{L_p}\right)^{1/\rho_1} + \left(E_{F^N}\|f - \hat{f}_+\|^\rho_{L_p}\right)^{1/\rho_1} \leqq$$

$$\leqq\ K^{1/\rho_1}\ \cdot \alpha\left(\frac{1}{N}\right)^{1/\rho_1} + \left(E_{F^N}\|f-\hat{f}_+\|^\rho_{L_p}\right)^{1/\rho_1}\ =\ K^{1/\rho_1}\ \cdot \alpha\left(\frac{1}{N}\right)^{1/\rho_1}\ +\ I_{31}.$$

By the lattice property and by assumption about α,

$$I_{31}\ \leqq\ \left(E_{F^N}\|f-\hat{f}\|^\rho_{L_p}\right)^{1/\rho_1}\ \leqq\ K^{1/\rho_1}\ \cdot\ \alpha\left(\frac{1}{N}\right)^{1/\rho_1}.$$

Therefore,

$$I_3\ \leqq\ 2K^{1/\rho_1}\ \cdot \alpha\left(\frac{1}{N}\right)^{1/\rho_1},$$

$$I_1+I_2+I_3\ \leqq\ (3+2^{\rho/\rho_1})\cdot K^{1/\rho_1}\cdot \alpha\left(\frac{1}{N}\right)^{1/\rho_1}\ +\ 2^{-\rho/\rho_1}\ L^{\rho/\rho_1}\ \beta(\varepsilon)^{\rho/\rho_1}$$

and, requiring that the contribution of the term containing $\beta(\varepsilon)$ does not exceed that of the term containing $\alpha\left(\frac{1}{N}\right)$, we obtain $\varepsilon:\ 0<\varepsilon\leqq\varepsilon_N$,

$$\varepsilon_N:\ 2^{-\rho/\rho_1}\ L^{\rho/\rho_1}\ \beta(\varepsilon_N)^{\rho/\rho_1} = (3+2^{\rho/\rho_1})\cdot\ K^{1/\rho_1}\cdot\ \alpha\left(\frac{1}{N}\right)^{1/\rho_1}, \tag{10}$$

i.e., (4).

Inequality (3) follows from the definition of ε_N in (10). □

Proof of Lemma 2.1. First let $p=p_1$.

If $s\in(0,2)$, then, by using an appropriate equivalent norm in $B^s_{p\infty}$

$$\omega_2\big(f;\delta\big)_{L_p}\ \leqq\ C_1\cdot\delta^s\ \|f\|_{B^s_{p\infty}}. \tag{11}$$

Hence, by the embedding $B^s_{pq}\hookrightarrow B^s_{p\infty}$, $0<q\leqq\infty$,

$$\omega_2\big(f;\delta\big)_{L_p}\ \leqq\ C_2\cdot\delta^s\ \|f\|_{B^s_{pq}}. \tag{12}$$

If $s\in(2,\,r)$, then (12) follows from

$$\omega_2\big(f;\delta\big)_{L_p}\ \leqq\ C_3\cdot\delta^2\ \|f\|_{W^2_p} \tag{13}$$

and the embeddings $W^2_p \hookleftarrow B^2_{p1}\hookleftarrow B^s_{pq}$, which are valid for any $p,\ q,\ s\ :\ 1\leqq p\leqq\infty$, $0<q\leqq\infty$, $s>2$.

If $s=2$ and either $1\leqq p<\infty$, $1\leqq q\leqq\min\{p,2\}$ or $p=\infty$, $q=1$ then (12) follows from (13) and the embeddings $W^2_p\hookleftarrow B^2_{p,\min\{p,2\}}$, $1\leqq p\leqq\infty$, and $W^2_\infty\hookleftarrow B^2_{\infty 1}$, respectively.

If $s=2$ and either $1\leqq p<\infty$, $\min\{p,2\}<q\leqq\infty$ or $p=\infty$, $1<q\leqq\infty$, then one utilizes the equivalence of $\omega_k\big(f;\delta\big)_{L_p}$ to the K-functional $K\big(\delta^k,f;L_p,\dot{W}^k_p\big)$ (see [Johnen *et al.* (1977)]) for $k=2$, leading to the Marchaud-type inequality ((3.2) in [Johnen *et al.* (1977)])

$$\omega_2\big(f;\delta\big)_{L_p}\sim\ K\big(\delta^2,f;L_p,\dot{W}^2_p\big)\ \ \leqq \tag{14}$$

$$\leqq \; C_4\,\delta^2 \left(\|f\|_{L_p} + \int\limits_{\delta}^{1} \tau^{-3}\, K\big(\tau^k, f; L_p, \dot W_p^k\big)\, d\tau \right) \;\sim\; \delta^2 \left(\|f\|_{L_p} + \int\limits_{\delta}^{1} \tau^{-3}\, \omega_k\big(f;\tau\big)_{L_p}\, d\tau \right),$$

where $k = [r] + 1$, $1 \leqq p \leqq \infty$. Applying Hölder's inequality for $q = \infty$ to the right-hand side in (14) yields

$$C_5\,\delta^2 \left(\|f\|_{L_p} + \int\limits_{\delta}^{1} \tau^{-3} \omega_k\big(f;\tau\big)_{L_p}\, d\tau \right) \;\leqq \tag{15}$$

$$\leqq\; C_5\,\delta^2 \left(\|f\|_{L_p} + \left(\int\limits_{\delta}^{1} \frac{d\tau}{\tau} \right) \cdot \left[\sup_{\tau > 0} \Big(\tau^{-2}\, \omega_k\big(f;\tau\big)_{L_p} \Big) \right] \right) \;\leqq$$

$$\leqq\; C_5\,\delta^2 \left(\|f\|_{L_p} \;+\; C_6 \,\cdot\, \left(ln\frac{1}{\delta} \right) \,\cdot\, \|f\|_{B^2_{p\infty}} \right) \;\leqq\; C_7\,\delta^2 \left(ln\frac{1}{\delta} \right) \,\cdot\, \|f\|_{B^2_{p\infty}},$$

where for the last inequality the embedding $L_p \hookleftarrow B^2_{p\infty}$ was used. (15) proves (5) in the case $s = 2$, $1 \leqq p \leqq \infty$, $q = \infty$.

Let us consider now the case $\min\{p,\,2\} < q < \infty$. If $1 \leqq p < \infty$, then it was already proved that

$$\omega_2\big(f;\delta\big)_{L_p} \;\leqq\; C_8\,\delta^2\, \|f\|_{B^2_{p,\min\{p,2\}}}. \tag{16}$$

Since $\omega_2\big(f;\delta\big)_{L_p}$ is an equivalent seminorm in the space $L_p + \dot W_p^2$ with the seminorm $K\big(\delta^2, f; L_p, \dot W_p^2\big)$, one can apply the real interpolation method between (16) and (15) (more precisely, Theorem 6.4.5 (2) in [Bergh *et al.* (1976)]) for $\eta = 1 - \dfrac{\min\{p,2\}}{q}$, so that $\dfrac{1}{q} = \dfrac{1-\eta}{\min\{p,2\}} + \dfrac{\eta}{\infty} = \dfrac{1-\eta}{\min\{p,2\}}$. The result is

$$\omega_2\big(f;\delta\big)_{L_p} \leqq C_9\,\delta^2 \left(ln\frac{1}{\delta} \right)^{1 - \frac{\min\{p,2\}}{q}} \|f\|_{B^2_{pq}}, \tag{17}$$

i.e., (5) for $1 \leqq p < \infty$.

If $p = \infty$, then in the place of (16) one uses (14), noting that

$$\|f\|_{L_\infty} + \int\limits_{0}^{1} \tau^{-3} \omega_k\big(f;\tau\big)_{L_\infty}\, d\tau \;\leqq\; C_{10}\, \|f\|_{B^2_{\infty 1}}, \tag{18}$$

by one of the theorems about equivalent norms in B^s_{pq}. Combining (14) and (18) yields

$$\omega_2\big(f;\delta\big)_{L_\infty} \leqq C_{10}\;\delta^2\, \|f\|_{B^2_{\infty 1}}. \tag{19}$$

After this, (5) for $p = \infty$ is obtained from (19, 15) by real interpolation analogously to (17). Thus, Lemma 2.1 is completely proved in the case $p_1 = p$.

For $p < p_1$, we use the embedding

$$B^s_{pq} \hookrightarrow B^{s - d\left(\frac{1}{p} - \frac{1}{p_1}\right)}_{p_1 q},\;\; 0 < p \leqq p_1 \leqq \infty,\; 0 < q \leqq \infty,\; s \in \mathbb{R}$$

to reduce consideration to the already proven case $p = p_1$.

Finally, let $p > p_1$. It is easy to verify that, by Hölder's inequality,

$$\omega_2\big(f;\delta\big)_{L_{p_1}} \leqq C(d,\,p,\,p_1)\,R^{\left(\frac{1}{p_1}-\frac{1}{p}\right).d}\,\omega_2\big(f;\delta\big)_{L_p} \tag{20}$$

which once again reduces the consideration to the case $p = p_1$. □

Remark 4.2.1. *All facts from the theory of function spaces used in the proof of Lemma 2.1 can be collected from [Bergh* et al. *(1976)], [Johnen* et al. *(1977)], [Triebel (1983)].*

Proof of Lemma 2.2. The case $s \neq 2 + d\left(\frac{1}{p} - \frac{1}{p_1}\right)_+$ follows from the embeddings

$$B^s_{p,\,\min\{p,\,q\}}(\mathbb{R}^d) \hookrightarrow F^s_{pq}(\mathbb{R}^d) \hookrightarrow B^s_{p,\,\max\{p,\,q\}}(\mathbb{R}^d),\ \ 0 < q \leqq \infty,\ 0 < p < \infty,\ s \in \mathbb{R}$$

(see, e.g., [Triebel (1983)]), and subsequent application of Lemma 2.1.

In the case $s = 2 + d\left(\frac{1}{p} - \frac{1}{p_1}\right)_+$ consider first $p = p_1$. Then $s = 2$, $q \leqq 2$ implies

$$F^s_{pq}(\mathbb{R}^d) \hookrightarrow F^2_{p2}(\mathbb{R}^d) \rightleftharpoons W^2_p(\mathbb{R}^d)$$

and (6) follows from the well–known inequality

$$\omega_2\big(f;\delta\big)_p \leqq C_{11}\delta^2\,\|f\|_{\dot{W}^2_p(\mathbb{R}^d)} \leqq C_{11}\delta^2\,\|f\|_{W^2_p(\mathbb{R}^d)}. \tag{21}$$

For $p < p_1$ one invokes

$$F^s_{pq} \hookrightarrow F^{s-d\left(\frac{1}{p}-\frac{1}{p_1}\right)}_{p_1q_1},\ \ 0 < p < p_1 < \infty,\ 0 < q \leqq \infty,\ 0 < q_1 \leqq \infty,\ s \in \mathbb{R},$$

and for $p > p_1$ – (20), which reduces the consideration to the case $p = p_1$, and the lemma follows in the same way as Lemma 2.1. □

Proof of Lemma 2.3. Let first $k = 2, 3, \ldots$ In this case (7) follows from the embeddings

$$W^2_p \hookleftarrow W^k_p \hookleftarrow W^k_p \curvearrowright (TV^k_p)^{1/p}$$

and (21).

For $k = 0, 1$ (7) follows from

$$(TV^k_p)^{1/p} \hookleftarrow W^k_p \curvearrowright (TV^k_p)^{1/p}$$

and Lemma 0.8. □

4.3 Proofs of corollaries

4.3.1. Corollaries in 3.1

Proof of Corollary 3.1.1. By invoking the embeddings (cf. the proof of Lemma 2.1) $B^{\sigma}_{p_1,\min\{p_1,2\}} \hookrightarrow W^{\sigma}_{p_1}$, $1 \leqq p_1 < \infty$, and $B^{\sigma}_{\infty 1} \hookrightarrow W^{\sigma}_{\infty}$, $\sigma > 0$, one verifies that the error measure $E\|f - \hat{f}\|^2_{B^{\sigma}_{p_1 u}}$ can be replaced by $E\|f - \hat{f}\|^2_{W^{\sigma}_{p_1}}$, where $\sigma \geqq 0$, $u = \min\{p_1, 2\}$ for $p_1 < \infty$ and $u = 1$ for $p_1 = \infty$. The corollary now follows by applying Theorem 2.1 and Lemma 2.1 with $\alpha\left(\frac{1}{N}\right)$ given by the rates in Theorem 1 in [Delyon *et al.* (1996)]. $\square$

Proof of Corollary 3.1.2. For fixed p_1, p, q and s, $\beta(\delta)$ is given by (5). To obtain the desired uniform estimate of ε_N, take the highest possible rate for $\alpha\left(\frac{1}{N}\right)$:

$$\alpha\left(\frac{1}{N}\right) \;=\; O\left(N^{-\frac{2r}{2r+d}}\right)$$

and the lowest possible rate for $\beta(\delta)$:

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

By, taking $\varepsilon_N = O\big(\beta^{-1}\big(\alpha\left(\frac{1}{N}\right)^{1/2}\big)\big)$ in (4) for the whole range of p_1, p, q and s the inequality (3) is preserved for $\square$

Proof of Corollary 3.1.3. The relevant embedding from Corollary 3.1.1 is in this case an isomorphism: $B^0_{22} \rightleftharpoons L_2$. The rest is analogous to Corollary 3.1.1. $\square$

Proof of Corollary 3.1.4. Analogous to Corollary 3.1.2, but in this case the minimal rate for $\beta(\delta)$ is

$$\beta(\delta) \;=\; O\left(\delta^{\min\left\{2,\, \max\{d/2,\, r_0\}\right\}}\right).$$

$\square$

The proofs of Corollaries 3.1.5, 7 are the same as for Corollary 3.1.1, with different values of K coming from the different assumption in Theorems 3 and 4 in [Delyon *et al.* (1996)]. The same applies to Corollaries 3.1.6, 8 which are respective analogues of Corollary 3.1.2.

Proof of Corollary 3.1.9. The proof is analogous to the one of Corollary 3.1.1, but $\alpha\left(\frac{1}{N}\right)$ is given by the rates in Corollary 1 in [Donoho *et al.* (1998)], and $\beta(\delta)$ is given by Lemma 2.2. $\square$

Proof of Corollary 3.1.10. For $p_1 = 2$ and fixed p, q, s, $\beta(\delta)$ is given by (6). The highest possible rate for $\alpha\left(\frac{1}{N}\right)$ is

$$\alpha\left(\frac{1}{N}\right) = O\Big(N^{-\frac{2r}{2r+1}}\Big).$$

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

$$\beta(\delta) = O\Big(\delta^{\inf_{s,\,p,\,q}\,\min\left\{2,\, s-\left(\frac{1}{p}-\frac{1}{2}\right)_+\right\}}\Big) = O\Big(\delta^{\inf_{p,\,q}\,\min\left\{2,\, \frac{1}{\min\{2,\,p,\,q\}}-\left(\frac{1}{p}-\frac{1}{2}\right)_+\right\}}\Big) =$$

$$= O\Big(\delta^{\inf_{p}\,\min\left\{2,\, \frac{1}{\min\{2,\,p\}}-\left(\frac{1}{p}-\frac{1}{2}\right)_+\right\}}\Big) = O\Big(\delta^{\min\{2,\,\frac{1}{2}\}}\Big) = O\Big(\delta^{1/2}\Big).$$

Now the proof is completed in the same way as the one of Corollary 3.1.2 by taking $\varepsilon_N = O\Big(\beta^{-1}\Big(\alpha\left(\frac{1}{N}\right)^{1/2}\Big)\Big)$. □

Remark 4.3.1.1. *From the proof of Corollary 3.1.10 it becomes clear that the range of* $q:\ 0 < q \leq \infty$ *can be reduced to* $0 < q \leqq 2$ *without affecting the value of* ε_N *obtained.*

Proof of Corollary 3.1.11. The proof is analogous to the one of Corollary 3.1.1 but $\alpha\left(\frac{1}{N}\right)$ is given by Corollary 1 in [Donoho *et al.* (1998)] (for $\sigma = k + 1/p$ in the notation there) and $\beta(\delta)$ is given by Lemma 2.3. Actually, in [Donoho *et al.* (1998)] the rates are given for $k = 0$ only but the general case $k = 0, 1, 2, \ldots$ can be dealt with analogously, using Lemma 0.9. □

4.3.2. Corollaries in 3.2

Proof of Corollary 3.2.1. The proof is the same as the proof of Corollaries 3.1.5, 7 but instead of Theorems 3, 4 in [Delyon *et al.* (1996)] one invokes Theorem 1 in [Lepski *et al.* (1997)]. □

Proof of Corollary 3.2.2. The prrof is the same as the proof of Corollaries 3.1.6, 8 but instead of Theorems 3, 4 in [Delyon *et al.* (1996)], Theorem 2 in [Lepski *et al.* (1997)]. □

5 Remarks

It should be noted that in the proof of Theorem 2.1 the lattice property plays an essential role. This means that the conclusion of Theorem 2.1 cannot be generalized for $\sigma > 0$ (where σ has the meaning in [Delyon *et al.* (1996)]).

In separate forthcoming papers we study the properties of the (variable-kernel) estimators obtained by replacing the integral in the definition of $\hat{f}_{+,\,\varepsilon}$ with quadrature formulae. These new estimators of kernel type preserve positivity, too, and prove to have the same regularity as the kernel Φ in the definition of approximate identity (see Appendix 0). They depend on ε and also on the parameters of the quadrature formula. As with ε, an optimal range can be found for each of these additional parameters, so that the new estimators retain asymptotic-minimax optimality. Special attention is given to the use of spline-wavelets, spline-kernels and a B-spline kernel Φ. In this case it is possible, at least for $d = 1$, to compute the integral explicitly, by closed formulae which can be obtained either manually or by automatic symbolic integration. The use of quadrature formulae would be equally effective as long as the quadrature formulae are chosen to be exact of order not less than the order of the spline. This can also facilitate renormalization, if, as in the case of density estimation , the integral of the renormalized estimator is required to be 1.

It is easy to see that if the condition $f(x) \geqq v$ be replaced by $f(x) \leqq w$ or by $v \leqq f(x) \leqq w$ a.e. on $\mathbb{R}^d$, Theorem 2.1 still holds true, *mutatis mutandis*. We note that this proves to be true also in the general case of non–constant restrictions $v(x) \leqq f(x) \leqq w(x)$.

Theorem 2.1 and the applications in 3.1 can in principle be extended also for $p_1 : \; 0 < p_1 < 1$ and/or $p : \; 0 < p < 1$, but the respective definition of the modulus of smoothness has to be more sophisticated (see [Dechevsky *et al.* (1997)] for the appropriate definition in the case $d = 1$).

It should be noted that the spaces appearing in the error measure in [Delyon *et al.* (1996)] for $p : \; 0 < p \leqq 1$ (in particular, see Lemma 1 in [Delyon *et al.* (1996)]) are not the classical Sobolev spaces but only closed subspaces of theirs. In particular for $\sigma = 0$ (in the notations of [Delyon *et al.* (1996)]) the related spaces are not L_p but H_p – the Hardy spaces which are essentially narrower. In view of this, we note that Lemma 1 in [Delyon *et al.* (1996)] is well-known (as an embedding between the spaces F^s_{p2} and B^s_{pp}). It can be obtained, e.g., from [Triebel (1983)].

Comparing the rates in Corollaries 3.1.1 and 3.2.1 indicates that for $\theta \leqq 0$ there is a difference in the logarithmic factors. This difference comes, at least partially, from the difference in the choice of ρ in [Delyon *et al.* (1996)] and [Lepski *et al.* (1997)] ($\rho = 2$ and $\rho = p_1$, respectively).

The classes V_p of functions with bounded p-variation considered in [Donoho *et al.* (1998)] are defined in the spirit of the work of Bergh and Peetre (for a detailed reference we recommend the original [Bergh *et al.* (1976)]). Comparing V_p to the Wiener-Young classes shows that V_p is "almost" TV°_p, but the former is slightly bigger, in the sense that $TV^\circ_p \hookrightarrow V_p$. In the present paper we gave preference to considerations of the Wiener-Young classes, because they are the classical ones, and because their norm is simpler to compute for concrete functions. This paper's results about TV°_p are automatically valid for V_p, too. In fact, the result about V_p is slightly more precise, because in the place of Lemma 0.9, for $k = 0$, the following (slightly more precise) embedding holds:

$$B^{1/p}_{p1} \hookrightarrow L_p \curvearrowright V_p \hookrightarrow B^{1/p}_{p\infty}, \; 1 \leqq p < \infty.$$

Let us note that it would be interesting to develop analogues of the results in [Delyon *et al.* (1996)], [Lepski *et al.* (1997)], etc., in terms of the *homogeneous* Besov, Sobolev, etc., spaces, rather than the inhomogeneous ones. This would allow a better assessment of the role, say, of a density's tail weight versus its regularity where interesting new results could be expected in the case of densities with non–compact support (cf. Remark 2.2.4 in [Dechevsky *et al.* (1998)]).

The performance of $\hat{f}_{+,\,\varepsilon}$ can be further improved by making the parameter $\varepsilon > 0$ space–dependent: $\varepsilon = \varepsilon(x)$, $x \in \mathbb{R}^d$. This would certainly improve the behaviour of $\hat{f}_{+,\,\varepsilon}$ on moderate samples. All the results in the present study can be extended to this more general case, if $\varepsilon(x)$ has the same regularity as the kernel $\Phi(x)$ and if $\|\varepsilon\|_{L_\infty} \leqq \varepsilon_0$. The basic ideas of the proofs are the same, but their realization is more technically involved. The nature of W may impose some additional restrictions on $\varepsilon(x)$.

Finally, due to the regularization properties of the convolution in the definition of $\hat{f}_{+,\varepsilon}$, Gibbs phenomenon with $\hat{f}_{+,\,\varepsilon}$ is expected to be considerably weaker compared to that of $\hat{f}$. These expectations have been confirmed by our numerical experiments.

In conclusion, the "pos-ε smooth" of f seems to be ideally suited to preserving asymptotic optimality over certain regularity classes.

Appendix 0. Preliminaries and notation

We recall the following general definitions from functional analysis necessary for our work here. We cite as a general reference [Bergh et al. (1976)].

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

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

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

Quasi-norm (see [Bergh *et al.* (1976)], [Dechevsky *et al.* (1997)]) : $\exists c \geqq 1;\ \|a+b\| \leqq c(\|a\|+\|b\|)$.

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

A – quasi(semi)normed abelian group – see [Bergh *et al.* (1976)].

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

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

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

$A \hookrightarrow B$: continuous embedding of A in B (for quasinormed abelian groups) – see [Bergh *et al.* (1976)].

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

$A \frown B = \{a \in A \cap B\ :\ \|a\|_{A \frown B}\ :=\ \max\{\|a\|_A, \|a\|_B\}$.

$A+B : A, B$ – quasinormed abelian groups: $\|a\|_{A+B} = \inf_{a=\alpha+\beta}\{\|\alpha\|_A+\|\beta\|_B\}$ (see [Bergh *et al.* (1976)]).

The K-functional: $K(t,a;A,B) = \inf_{a=\alpha+\beta}\big(\|\alpha\|_A + t\|\beta\|_B\big)$, $0 < t < \infty$, plays a big rolw in this paper. (It should be noted that the K-functional is an equivalent quasinorm on $A+B$ (see [Bergh *et al.* (1976)], [Johnen *et al.* (1977)] for details and properties).

We use the following notation for function spaces:

$L_p = L_p(\mathbb{R}^d)$, $1 \leqq p \leqq \infty$ – Lebesgue spaces over $\mathbb{R}^d$ with respect to the Lebesgue measure in $\mathbb{R}^d$ (notations as in [Bergh *et al.* (1976)]).

$W_p^k = W_p^k(\mathbb{R}^d)$, $\dot{W}_p^k = \dot{W}_p^k(\mathbb{R}^d)$, $1 \leqq p \leqq \infty$, $k = 0, 1, \ldots$ – the inhomogeneous and homogeneous Sobolev spaces, respectively (see [Adams (1975)], [Bergh *et al.* (1976)], [Johnen *et al.* (1977)], [Triebel (1983)], [Triebel (1992)]).

$B_{pq}^s = B_{pq}^s(\mathbb{R}^d)$, $\dot{B}_{pq}^s = \dot{B}_{pq}^s(\mathbb{R}^d)$ – the inhomogeneous and homogeneous Besov spaces, $0 < p \leqq \infty$, $0 < q \leqq \infty$, $s \in \mathbb{R}$ – see [Bergh *et al.* (1976)], [Triebel (1983)], [Triebel (1992)] (we shall be dealing with the case $1 \leqq p \leqq \infty$, $s > 0$).

$F_{pq}^s = F_{pq}^s(\mathbb{R}^d)$, $\dot{F}_{pq}^s = \dot{F}_{pq}^s(\mathbb{R}^d)$ - the inhomogeneous and homogeneous Triebel-Lizorkin spaces, respectively, $0 < p \leqq \infty$, $0 < q \leqq \infty$, $s \in \mathbb{R}$ (see [Bergh *et al.* (1976)], [Triebel (1983)], [Triebel (1992)]).

For more details on the properties of B_{pq}^s and F_{pq}^s we refer to [Bergh *et al.* (1976)], [Triebel (1983)], [Triebel (1992)].

The integral modulus of smoothness of $f \in L^p(\mathbb{R}^d)$ is defined, for $\delta > 0$, as:

$$\omega_k(f;\delta)_{L_p(\mathbb{R}^d)} = \sup_{|h|\leqq\delta} \|\Delta_h^k f\|_{L_p(\mathbb{R}^d)},$$

where $k \in \mathbb{N}$, $h \in \mathbb{R}^d$, $\Delta_h^k f(x)$ is the k-th finite difference with step h at x:

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

A very important property of $\omega_k(f;\delta)_{L_p}$ which we will use is its equivalence to a K-functional (see [Johnen *et al.* (1977)]): $\omega_k(f;\delta)_{L_p} \sim K(\delta^k, f, L_p, \dot{W}_p^k)$, $\delta > 0$, with equivalence constants independent of δ.

Throughout $(X_1, \ldots, X_N)$ will denote the random vector of N (not necessarily uncorrelated) observations which has a cumulative distribution function F^N. The error measures in which the risk will be measured, are the quasi-norms on $(0 < p \leqq \infty, 0 < \rho \leqq \infty)$, $\mathcal{L}_\rho(\mathbb{R}^d, L_p(\mathbb{R}^d))$, which is the complete quasi-normed abelian group of trajectories of stochastic processes $G(x)$, $x \in \mathbb{R}^d$, based on measurable transformations of the random vector $(X_1, \ldots, X_N)$ of N (not necessarily uncorrelated) random variables from a cumulative distribution function (c.d.f.) $F^N : \mathbb{R}^{dN} \to [0,1]$ and

$$\|G\|_{\mathcal{L}_\rho(L_p)} = \Big(E_{F^N}\Big(\int_{\mathbb{R}^d} |G(x)|^p \, dx\Big)^{\rho/p}\Big)^{1/\rho};$$

$L_p(\mathbb{R}^d, \mathcal{L}_\rho(\mathbb{R}^d))$ is the complete quasi-normed abelian group of functions $g(x)$, defined on $\mathbb{R}^d$, taking random values depending on the random vector $(X_1, \ldots, X_N)$ and and

$$\|g\|_{L_p(\mathcal{L}_\rho)} = \Bigg(\int\limits_{\mathbb{R}^d} \Big(E_{F^N} |g(x)|^\rho \Big)^{p/\rho} dx \Bigg)^{1/p}.$$

(In the definitions of $\mathcal{L}_\rho(L_p)$ and $L_p(\mathcal{L}_\rho)$ for $p = \infty$ the integral over $\mathbb{R}^d$ is replaced by Lebesgue-ess sup, $x \in \mathbb{R}^d$.) By Fubini's theorem $\mathcal{L}_p(\mathbb{R}^d, L_p(\mathbb{R}^d))$ and $L_p(\mathbb{R}^d, \mathcal{L}_p(\mathbb{R}^d))$ are isometric and for $\rho < p$ and $\rho > p$ there are respective embeddings (see [Dechevsky *et al.* (1998)]). For this reason we restrict consideration to $\mathcal{L}_p(\mathbb{R}^d, L_p(\mathbb{R}^d))$ in its slightly generalized form $\mathcal{L}_\rho(\mathbb{R}^d, L_p(\mathbb{R}^d))^\rho$, where $\rho \in (0, \infty)$. This is convenient in order to present all applications in section 3 in a unified way. In this generalized form $\mathcal{L}_\rho(L_p)^\rho$ is still a complete quasi-normed abelian group. We note that it has the lattice property, i.e., from $|G(x)| \leqq G_1(x)$ Lebesgue a.e. on $\mathbb{R}^d$, with $G_1 \in \mathcal{L}_\rho(L_p)^\rho$, it follows $G \in \mathcal{L}_\rho(L_p)^\rho$ and $\|G\|_{\mathcal{L}_\rho(L_p)^\rho} \leqq \|G_1\|_{\mathcal{L}_\rho(L_p)^\rho}$.

Let $\Psi \in L_1(\mathbb{R}^d)$, $\Psi \geqq 0$ Lebesgue-a.e. on $\mathbb{R}^d$, the support of Ψ, *supp* $\Psi \subset \{x \in \mathbb{R}^d : |x| \leqq 1\}$, $\Psi(-x) = \Psi(x)$, $x \in \mathbb{R}^d$. By considering

$$\Phi(x) := \frac{\Psi(x)}{\int\limits_{\mathbb{R}^d} \Psi(t)\, dt},$$

we obtain a function Φ which enjoys all the above properties and, additionally,

$$\int\limits_{\mathbb{R}^d} \Phi(x)\, dx = 1.$$

Define the approximate identity (cf. [Reed *et al.* (1975)])

$$\Phi_\varepsilon(x) := \frac{1}{\varepsilon^d} \Phi\left(\frac{x}{\varepsilon}\right),\ x \in \mathbb{R}^d.$$

Clearly,

$$\int\limits_{\mathbb{R}^d} \Phi_\varepsilon(x)\, dx = 1,\ \forall \varepsilon > 0.$$

For any $f \in L_p(\mathbb{R}^d)$, $1 \leqq p \leqq \infty$, the convolution $f_\varepsilon := \Phi_\varepsilon * f(x)$, $x \in \mathbb{R}^d$ is in $L_p(\mathbb{R}^d)$ (by Young's inequality – see [Bergh *et al.* (1976)], [Nikol'skiĭ (1975)], [Reed *et al.* (1975)]). $f_\varepsilon(x)$ is called (see [Nikol'skiĭ (1975)]) the Sobolev ε-mean of f.

Let $C^k(\mathbb{R}^d)$ (and $C_0^k(\mathbb{R}^d)$), $k = 0, 1, \ldots, \infty$, denote, as usual, the k-times continuously differentiable functions on $\mathbb{R}^d$ (with compact support respectively). A classical construction of an approximate identity $\Phi_\varepsilon \in C_0^\infty(\mathbb{R}^d)$ is as follows:

- consider

$$\phi_1(x) = \begin{cases} e^{\frac{1}{x-1} - \frac{1}{x}} & 0 < x < 1 \\ 0 & x \in (-\infty, 0] \cup [1, \infty); \end{cases}$$

- define

$$\phi_2(x) = \frac{1}{\int_0^1 \phi_1(t)\,dt} \cdot \int_x^1 \phi_1(t)\,dt;$$

- take $\Psi(x) = \phi_2(|x|^2)$, $x \in \mathbb{R}^d$, and so on.

A similar classical construction is to take (see [Nikol'skiĭ (1975)])

$$\phi_3(x) = \begin{cases} e^{\frac{x^2}{x^2-1}}, & -1 < x < 1 \\ 0, & x \in (-\infty, -1] \cup [1, \infty). \end{cases}$$

and $\Psi(x) = \phi_3(|x|)$, $x \in \mathbb{R}^d$.

Another way of constructing $\Phi_\varepsilon \in W_1^r(\mathbb{R}^d) \cap W_\infty^r(\mathbb{R}^d)$ is to start with $\phi_1(x)$, $x \in \mathbb{R}^d$, where ϕ_1 is a r-th degree B-spline ([de Boor (1978)]).

In the following lemmata $p\,:\,1 \leqq p \leqq \infty$.

Lemma 1. *$\omega_k(f;\delta)_{L_p}$ is a semi-norm, $\forall\delta > 0$.*

Lemma 2. *$\omega_k(f;\delta)_{L_p} \leqq 2^k \|f\|_{L_p}$, $\forall\delta > 0$.*

Lemma 3. *$f \mapsto f_\varepsilon$ is a linear operator, $\forall\varepsilon > 0$.*

Lemma 4. *$\|f_\varepsilon\|_{L_p} \leqq \|f\|_{L_p}$, $\forall\varepsilon > 0$.*

Lemma 5. *$\|f - f_\varepsilon\|_{L_p} \leqq \frac{1}{2}\omega_2(f;\varepsilon)_{L_p}$, $\forall\varepsilon > 0$.*

Next we consider functions with bounded total variation on $\mathbb{R}$, $d = 1$. Let $\omega \in \Omega(\infty)$. Define

$$T\vee_\omega \;=\; \{f \;:\; \mathbb{R} \to \mathbb{R},\;\; \vee_\omega f < \infty\}$$

where

$$\vee_\omega f \;=\; \sup\{\sum_{\nu=1}^{n} \omega\left(|f(x_\nu) - f(x_{\nu-1})|\right) \;:\; -\infty < x_1 < \ldots < x_n < \infty,\; n \in \mathbb{N}\}$$

is the ω–variation of f on $\mathbb{R}$ in the sense of Young [Young (1937)]. The case $\omega = \omega(t) = t$ corresponds to the usual Jordan variation, while the case $\omega = \omega(t) = t^p$, $1 \leqq p < \infty$, corresponds to p–variation in the sense of Wiener [Wiener (1924)].

The following result is well–known.

Lemma 6. *Let* $\omega_i \in \Omega(\infty)$, $i = 1, 2$, *and*

$$\limsup_{t \to 0+} \frac{\omega_1(t)}{\omega_2(t)} < \infty.$$

Then,

$$T\vee_{\omega_2} \subset T\vee_{\omega_1}.$$

In particular, for the Wiener-Young classes one has

Lemma 7. $(T\vee_{p_1})^{1/p_1} \hookrightarrow (T\vee_{p_2})^{1/p_2}$, $1 \leqq p_1 \leqq p_2 < \infty$.

For $k = 0, 1, 2, \ldots$ consider the quasiseminormed abelian group

$$T\vee_p^k = \{f \in \dot{W}_p^k(\mathbb{R}) \ : \ \|f\|_{T\vee_p^k} := \vee_p f^{(k)} < \infty\}.$$

Lemma 8. *Let* $k = 0, 1, 2, \ldots$, $1 \leqq p < \infty$, $f \in T\vee_p^k$. *Then,*

$$\omega_{k+1}(f; \delta)_p \leqq C(k)\delta^{k+1/p}(\vee_p f^{(k)})^{1/p}.$$

Lemma 9. $B_{p\infty}^{s+1/p} \hookrightarrow L_p \curvearrowright T\vee_p^k \hookrightarrow B_{p\infty}^{k+1/p}$, $\forall s > k$.

Acknowledgement: This research has been supported by the Natural Sciences and Engineering Research Council of Canada. We are thankful to Dr. Spiridon Penev (University of New South Wales, Sydney, Australia) for making the manuscripts [Delyon *et al.* (1996)], [Lepski *et al.* (1997)] available to us while they were still in preprint form.

References

[Adams (1975)] R.A. Adams. *Sobolev spaces.* Academic Press, 1975.

[Bergh *et al.* (1976)] J. Bergh, J. Löfström. *Interpolation spaces. An introduction.* Grundlehren, **223**. Springer, 1976.

[de Boor (1978)] C. de Boor. *A practical guide to splines.* Springer, 1978.

[Dechevsky *et al.* (1997)] L. Dechevsky, S. Penev. *On shape–preserving probabilistic wavelet approximators. Stochastic Analysis and Applications*, **15**, **2**, (1997), 187-215.

[Dechevsky *et al.* (1998)] L. T. Dechevsky, S. I. Penev. *On shape preserving wavelet estimators of cumulative distribution functions and densities. Stochastic Analysis and Applications*, **16**, **3**, (1998), 428-469.

[Delyon *et al.* (1996)] B. Delyon, A. Juditsky. *On minimax wavelet estimators.* Applied and Computational Harmonic Analysis, **3**, (1996), 215-228.

[Donoho *et al.* (1998)] D. L. Donoho, I. M. Johnstone. *Minimax estimation via wavelet shrinkage.* Annals of Statistics, **26**, **3**, (1998), 879-921.

[Donoho *et al.* (1995)] D. L. Donoho, I. M. Johnstone, G. Kerkyacharian, D. Picard. *Wavelet shrinkage: Asymptopia? Journal of the Royal Statistical Society*, Series **B**, **57**, **2**, (1995), 301-369.

[Johnen *et al.* (1977)] H. Johnen, K. Scherer. *On the equivalence of the K-functional and moduli of continuity and some applications.* Lecture Notes in Mathematics, **571**. Springer, 1977, 119-140.

[Lepski *et al.* (1997)] O. V. Lepski, E. Mammen, V. G. Spokoiny. *Optimal spatial adaptation to inhomogeneous smoothness: an approach based on kernel estimators with variable bandwidth selectors.* Annals of Statistics, **25**, **3**, (1997), 929-947.

[Nikol'skiĭ (1975)] S. M. Nikol'skiĭ. *Approximation of functions of several variables and imbedding theorems.* Grundlehren, **205**. Springer, 1975.

[Penev *et al.* (1997)] S. Penev, L. Dechevsky. *On non–negative wavelet–based density estimators.* Nonparam. Stat., **7**, (1997), 365–394.

[Petrushev *et al.* (1987)] P. P. Petrushev, V. A. Popov. *Rational approximation of real functions.* Cambridge Univ. Press, 1987.

[Pinheiro *et al.* (1997)] A. Pinheiro, B. Vidakovic. *Estimating the square root of a density via compactly supported wavelets.* Computational Statistics and Data Analysis, **25**, (1997), 399-415.

[Reed *et al.* (1975)] M. Reed, B. Simon. *Methods of modern mathematical physics.* Vol. 2. Academic Press, 1975.

[Sendov *et al.* (1988)] B. Sendov, V. A. Popov. *The Averaged Moduli of Smoothness: Applications in Numerical Methods and Approximation.* Wiley, 1988.

[Triebel (1983)] H. Triebel. *Theory of function spaces.* Monographs in Mathematics, **78**. Birkhäuser, 1983.

[Triebel (1992)] H. Triebel. *Theory of function spaces II.* Monographs in Mathematics, **84**. Birkhäuser, 1992.

[Wiener (1924)] N. Wiener. *The quadratic variation of a function and its Fourier coefficients.* Massachusetts Journal of Mathematical Physics, **3**, (1924), 72–94.

[Young (1937)] L. C. Young. *Sur une généralisation de la notion de puissance p-ième bornée au sens de Wiener, et sur la convergence de séries de Fourier*. Compt. Rend. Acad. Sci., **204**, **7**, (1937), 470–472.