Les Cahiers du GERAD ISSN: 0711–2440

Asymptotically Minimax Non–Parametric Function Estimation with Positivity Constraints: The General Case of Variable Order Constraints

L. Dechevsky
B. MacGibbon

G–2009–72

November 2009

Asymptotically Minimax Non–Parametric Function Estimation with Positivity Constraints: The General Case of Variable Order Constraints

Lubomir Dechevsky

Narvik University College
2 Lodve Langes St.
P.O.B. 385
N-8505 Narvik, Norway

ltd@hin.no

Brenda MacGibbon

GERAD, HEC Montréal
3000, chemin de la Côte-Sainte-Catherine
Montréal (Québec) Canada, H3T 2A7
& Université du Québec à Montréal

macgibbon.brenda@uqam.ca

November 2009

Les Cahiers du GERAD

G–2009–72

Abstract

A challenge in many applications of non-parametric curve estimation is that the function must satisfy some (lower and/or upper) variable order constraints (for example, a density is constrained to lie between two functions). At the same time the spatially inhomogeneous smoothness of the function is modelled by Besov and Triebel-type smoothness constraints. Donoho and Johnstone (1998) and Delyon and Juditsky (1996) studied minimax rates of convergence for wavelet estimators with thresholding, while Lepski et al. (1997) proposed a variable bandwidth selection for kernel estimators that achieved optimal rates over the scale of Besov spaces. Here we show how to construct estimators that satisfy the variable order constraints and also achieve minimax rates over the appropriate smoothness class. This generalizes results of Dechevsky and MacGibbon (1999) for the case of constant constraints. The parameters of the new constrained estimator (when the constraints are functions) are shown here to depend on the regularity of the constraint functions, except when the lower constraint function is convex and/or the upper constraint function is concave. A preliminary announcement of some of the results of the present work (without proofs) was made in Dechevsky (2007) as a part of a survey on the state of the art and ongoing research in shape-preserving wavelet approximation.

Résumé

Le défi dans plusieurs applications de l'estimation non-paramétrique des fonctions est représenté par le fait que les fonctions doivent également satisfaire des contraintes variables d'ordre (par exemple, une fonction de densité doit se situer entre deux fonctions). En même temps, le lissage spatialement non-homogène de la fonction est modélisé par les contraintes de lissage de type Besov et Triebel. Donoho and Johnstone (1998) and Delyon and Juditsky (1996) ont étudié le taux de convergence minimax pour les estimateurs d'ondelettes avec des seuils, alors que Lepski et al. (1997) ont proposé une sélection de fenêtre variable pour un estimateur à noyau qui atteint un taux optimal sur l'échelle des espaces de Besov. Dans ce travail nous démontrons comment construire des estimateurs qui satisfont des contraintes variables d'ordre et aussi atteignent le taux minimax sur des classes de lissage appropriées. Ces résultats généralisent ceux de Dechevsky and MacGibbon (2010) pour le cas où les contraintes d'ordre sont des constantes. Les paramètres de ces nouveaux estimateurs dépendent de la régularité des fonctions de contrainte, sauf dans le cas où la borne inférieure est convexe et/ou la borne supérieure est concave. Une annonce préliminaire (sans démonstration) des résultats a été faite dans Dechevsky (2007) dans un survol de la recherche récente de l'approximation par ondelettes qui conservent la forme.

Acknowledgments: This research has been supported by the Natural Sciences and Engineering Research Council of Canada, and, in its final stage, by the 2007 Annual Research Grant of the R&D Group for Mathematical Modelling, Numerical Simulation and Computer Visualization at Narvik University College, Norway. The authors also wish to thank GERAD (Groupe d'études et de recherche en analyse des décisions – HEC Montréal, École Polytechnique, McGill University, UQAM).

1 Introduction

The spatially inhomogeneous smoothness of nonparametric methods is often modelled by Besov and Triebel-type smoothness constraints. For such problems, Donoho and Johnstone (1998) and Delyon and Juditsky (1996) studied minimax rates of convergence for wavelet estimators with thresholding, while Lepski et al. (1997) proposed a variable bandwidth selection for kernel estimators that achieved optimal rates over Besov classes. However, a second challenge in many applications of non-parametric curve estimation is that the function must also satisfy some (lower and/or upper) variable order constraints (for example, a density must be non-negative or a density is constrained to lie between two functions). Here we show how to construct estimators under order constraints that satisfy these constraints and also achieve minimax rates over the appropriate smoothness class. In the preprint Dechevsky and MacGibbon (1999), recently published in Dechevsky and MacGibbon (2009), we studied the case when the lower and/or upper constraints are constants; in the present article, which follows closely the exposition of the previously unpublished preprint Dechevsky and MacGibbon (2001), we consider the general case when these constraints are functions. The parameters of the new constrained estimator are shown here to depend on the regularity of the constraint functions, except when the lower constraint function is convex and/or the upper constraint function is concave.

Our goal is to develop a general method yielding estimators which are asymptotically minimax-rate optimal for non-parametric density and regression-function estimation when faced with this problem of spatially inhomogeneous smoothness in the presence of variable order constraints. This method works in the univariate case, as well as in the multivariate case.

For density estimation, Penev and Dechevsky (1997) and Pinheiro and Vidakovic (1997) have obtained results using wavelet-based estimators which preserve positivity, perform well on moderate samples, and are relatively computationally inexpensive. However, they are narrowly specialized to preserve only a one-sided constraint, and this constraint has to be a constant (typically, a lower bound $v(x) \equiv 0$). Moreover, asymptotic results are limited to the Hellinger metric. Our study of the general problem showed that already the case where both lower and upper constraints of the form, $v(x) \equiv c_0$ and $w(x) \equiv c_1$ with known constants $c_0 < c_1$, are given, is essentially more difficult; its study is not amenable to these previously published techniques. In Dechevsky and MacGibbon (1999, 2009) we constructed and studied the properties of a new constrained estimator for this latter case.

In the present article we extend the definition of the constrained estimator from Dechevsky and MacGibbon (1999, 2009) to the general case when v and/or w are non-constant functions: $v = v(x)$, $w = w(x)$. More precisely, for an unknown density or regression function $f(x)$ whose graph is a priori known to be bounded by those of $v(x)$ and/or $w(x)$, given a rate-optimal estimator $\widehat{f}$ whose graph need not obey the bound(s), we construct a sufficiently smooth estimator $\widehat{f}^+(x)$ whose graph obeys the bounds and achieves the same asymptotic rate as $\widehat{f}(x)$, with respect to the same metric (and with the constant factor associated with the rate for $\widehat{f}^+(x)$ depending only on the factor for $\widehat{f}(x)$ and the chosen metric).

The standard unconstrained d-dimensional non-parametric regression-function estimation problem is to estimate an unknown function $f(x)$ on the basis of a sample of N (not necessarily uncorrelated) noisy observations

$$Y_i = f(t_i) + \sigma\varepsilon_i, \quad i = 1, \ldots, N, \tag{1}$$

where σ and ε_i represent the variance and error terms respectively. The points t_i may belong to a (uniform or non-uniform) deterministic design ($t_i = x_i$, $i = 1, \ldots, N$), or to a random design ($t_i = X_i$, $i = 1, \ldots, N$) where X_i, $i = 1, \ldots, N$, are (not necessarily uncorrelated) d-dimensional random vectors from a cumulative distribution function (c.d.f.) $F^N : \mathbb{R}^{\mathrm{Nd}} \to [0, 1]$. In the case of density estimation, it is assumed in addition that there exists an absolutely continuous density, with Radon-Nikodym derivative $f = F'$, and the problem is to estimate the density f based on the sample X_i, $i = 1, \ldots, N$. In many applications, it is often assumed that the samples, $\{X_i\}$ and $\{\varepsilon_i\}$ each consist of independent, identically distributed (i.i.d.) random variables and the Y_i's are independent (cf., e.g., Delyon and Juditsky (1996); Lepski et al. (1997)).

There are many different approaches to these estimation problems but we will concentrate on two of them which are known to yield asymptotically optimal results when the regularity of the estimated function is measured in terms of the Besov class to which it belongs. These classes contain the well-known Hölder (or "Lipschitz-α") classes. In addition, classes of spatially inhomogeneous functions with bounded Jordan and Wiener-Young variation, as well as Triebel-Lizorkin classes (including the well-known Sobolev spaces), are sandwiched between two Besov spaces and asymptotic optimality results can also be deduced for these classes (cf. Donoho and Johnstone (1998); Donoho et al. (1995)). Donoho and Johnstone (1998); Delyon and Juditsky (1996) studied minimax rates for wavelet estimators with thresholding while Lepski et al. (1997) proposed a variable-bandwidth selector for kernel estimators. Both types of estimators achieve optimal rates over Besov classes. For each of these rate-optimal unconstrained estimators $\widehat{f}$ our method allows us to construct a constrained estimator $\widehat{f}^+$ which is also rate-optimal. When $\widehat{f}$ is the kernel estimator from Lepski et al. (1997) (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 $\widehat{f}^+$ does obey all constraints.

Here we will outline our method in the case of constant order constraints (cf. Dechevsky and MacGibbon (1999, 2009)), as follows. Consider the case of a d-variate function satisfying the order constraints $v \leq f \leq w$ (or $v \leq f$, or $f \leq w$), with $v \equiv c_0$, $w \equiv c_1$ (c_0 and c_1 constants with $c_0 \leq c_1$). Let $\widehat{f}$ denote an estimator that satisfies the smoothness criteria, but not necessarily the order constraints. Clearly,

$$\widehat{f}_+(x) := \min\left\{w, \max\left\{v, \widehat{f}(x)\right\}\right\}, \qquad x \in \mathbb{R}^{\mathrm{d}}, \tag{2}$$

(or $\widehat{f}_+(x) := \max\left\{v, \widehat{f}(x)\right\}$, or $\widehat{f}_+(x) := \min\left\{w, \widehat{f}(x)\right\}$, respectively) will satisfy the order constraints but not necessarily the smoothness constraints. In Dechevsky and MacGibbon (1999, 2009) we achieved simultaneous matching of both types of constraints by proposing the following.

Definition 1.1 *The constrained ε-smooth of $\widehat{f}$, denoted by $\widehat{f}_{+,\varepsilon}$ is,*

$$\widehat{f}_{+,\varepsilon}(x) := \Phi_\varepsilon * \widehat{f}_+(x), \qquad x \in \mathbb{R}^{\mathrm{d}}, \tag{3}$$

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

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

where Φ is a "suitably smooth" compactly supported symmetric non-negative Lebesgue-measurable function with integral on $\mathbb{R}^{\mathrm{d}}$ equal to one and with the diameter of its support $= C_\Phi$ (see the Appendix for details and references).

Remark 1.1 *Because $v \leq w$, $\min\{w, \max\{v,t\}\} = \max\{v, \min\{w,t\}\}$ holds for any $t \in \mathbb{R}$.*

In Dechevsky and MacGibbon (1999, 2009) it was shown that this "constrained ε-smooth" of $\widehat{f}$, with a suitable selection of ε, can be chosen to be the estimator with the desired smoothness properties for the case of constant constraints v and w. Here, the underlying property on which the simultaneous achievement of rate optimality, smoothness and bounded order is founded is that if v and w are constants and $v \leq \widehat{f}_+ \leq w$, then $v \leq \Phi_\varepsilon * \widehat{f}_+ \leq w$ also holds. What happens, however, if in (2) the bounds $v = v(x)$, $w = w(x)$ are variable ? Then, it can be shown that, for any $\varepsilon > 0$,

$$v(x) \leq \widehat{f}_+(x) \leq w(x) \qquad \text{for all} \quad x \in \mathbb{R}^{\mathrm{d}}, \tag{4}$$

does not imply, in general,

$$v(x) \leq \widehat{\Phi}_\varepsilon * \widehat{f}_+(x) \leq w(x) \qquad \text{for all} \quad x \in \mathbb{R}^{\mathrm{d}}. \tag{5}$$

However, there is an important special case of variable $v(x)$, $w(x)$ (also including the case of constant v and w), when (4) does imply (5). This is the first important new observation in the present paper, and we summarize it as follows.

Definition 1.2 *Using the same assumptions as in Definition 1.1, let us suppose more generally, that $v = v(x)$ and $w = w(x)$ in (2) are (non-constant) functions such that the lower bound $v(x)$ is convex, and the upper bound $w(x)$ is concave.*

As it will be shown in Lemma 2.1 below, under the more general assumptions of Definition 1.2, (4) does imply (5) which permits us to extend essentially all the results of Dechevsky and MacGibbon (1999, 2009) to the more general case of non-constant $v(x)$ and $w(x)$ as given in Definition 1.2.

In the general case of variable, not necessarily convex, $v = v(x)$, and variable, not necessarily concave, $w = w(x)$, in order to ensure that (4) implies (5), we need to modify the definition of $\widehat{f}_+$ as follows.

Definition 1.3 *Under the assumptions of Definition 1.1, suppose that $v = v(x)$ and $w = w(x)$ are non-constant functions, Φ is compactly supported and that, for any $\varepsilon > 0$, $\widehat{f}_+(x) = \widehat{f}_+^{\varepsilon}(x)$ is defined by*

$$\widehat{f}_+^{\varepsilon}(x) := \min\left\{I(w, x; 2C_\Phi\varepsilon),\ \max\left\{S(v, x; 2C_\Phi\varepsilon),\ \widehat{f}(x)\right\}\right\}, \qquad x \in \mathbb{R}^{\mathrm{d}}, \tag{6}$$

where: $I(w, x; 2C_\Phi\varepsilon)$ is the lower Baire function of w at x, with step $2C_\Phi\varepsilon$; $S(v, x; 2C_\Phi\varepsilon)$ is the upper Baire function of v at x, with step $2C_\Phi\varepsilon$ (see the Appendix); $C_\Phi = $ diam supp Φ.

As we shall see, the results of Dechevsky and MacGibbon (1999, 2009) do extend even to this very general case about variable order constraints $v(x)$ and $w(x)$. The new element here is the more complex theory of the selection of the optimal range for ε in (3) which now reads

$$\widehat{f}_{+,\varepsilon}(x) := \Phi_\varepsilon * \widehat{f}_+^{\varepsilon}(x), \qquad x \in \mathbb{R}^{\mathrm{d}}, \tag{7}$$

and where, in order to achieve simultaneously rate-optimality, sufficient smoothness, and bounded order, the selection of ε depends not only on the regularity of f, but also on the regularity of v and w.

The paper is organized as follows. This introduction constitutes Section 1. The extension of the results of Dechevsky and MacGibbon (1999, 2009) to the case of convex variable lower bound $v(x)$ and/or concave variable upper bound $w(x)$ is given in Section 2. Section 3 treats the case of general variable order constraints. In Section 4 our methods are applied to the wavelet and kernel estimators considered by Delyon and Juditsky (1996) and Lepski et al. (1997), respectively. Section 5 contains the proofs. Some concluding remarks are collected in Section 6. Preliminary notation and definitions which are not explained in the main text can be found in the Appendix.

Some of the main results in Dechevsky and MacGibbon (2001) and the present paper were previously announced, without proofs or details, as part of Dechevsky (2007), a survey paper on the state of the art and ongoing research in shape-preserving wavelet-based approximation.

2 Convex lower and concave upper constraints

The results of Dechevsky and MacGibbon (1999, 2009) are valid for a lower constraint $v \equiv c_0$ and/or an upper constraint $w \equiv c_1$, where $c_0$and c_1 are constants such that $c_0 \le c_1$ holds. In this section we shall extend the main results of Dechevsky and MacGibbon (1999, 2009) to the case of variable lower bound $v(x)$ and/or upper bound $w(x)$ with $v(x) \le w(x)$, $x \in \mathbb{R}^{\mathrm{d}}$, and such that v is convex and w is concave. The key fact allowing for this extension is given in the following Lemma 2.1.

For the closed subset $A \subset \mathbb{R}^{\mathrm{d}}$, let A^{conv} be its convex hull, i.e.,

$$A^{conv} = \left\{y \in \mathbb{R}^{\mathrm{d}} : \mathrm{y} = (1-\alpha)\mathrm{x}_1 + \alpha\mathrm{x}_2,\ \mathrm{x}_1, \mathrm{x}_2 \in \mathrm{A},\ \alpha \in [0, 1]\right\}.$$

Let $D(f,\widehat{f}) = \left(supp\ f \cup supp\ \widehat{f}\right)^{conv}$ and let, for $\varepsilon > 0$, $D(f,\widehat{f};\varepsilon)$ be the closed $C_\Phi\frac{\varepsilon}{2}$-neighbourhood of $D(f,\widehat{f})$; i.e.,

$$D(f,\widehat{f};\varepsilon) = \left\{y \in \mathrm{IR^d} : \left(\exists \mathrm{x} \in \mathrm{D(f,\widehat{f})} : |\mathrm{y}-\mathrm{x}| \leq \frac{1}{2}\mathrm{C}_\Phi\varepsilon\right)\right\},$$

where $C_\Phi = diam\ supp\ \Phi$.

Lemma 2.1 *Let $v(x)$ and $w(x)$ be Lebesgue-measurable functions defined for every $x \in \mathrm{IR^d}$, with $v(x) \leq w(x)$, and let v be convex and w be concave on $D(f,\widehat{f};\varepsilon)$, with ε as in (5). Then, (4) implies (5).*

Now we are in a position to extend the main result of Dechevsky and MacGibbon (1999, 2009) (Theorem 2.1) as follows.

Theorem 2.1 *Let $f : \mathrm{IR^d} \to \mathrm{IR}$ be such that there exists on $\mathrm{IR^d}$ a Lebesgue-measurable function $v = v(x)$ with $v(x) \leq f(x)$, $x \in \mathrm{IR^d}$, and/or there exists on $\mathrm{IR^d}$ a Lebesgue-measurable function $w = w(x)$ with $f(x) \leq w(x)$, $x \in \mathrm{IR^d}$, such that v is convex, and w is concave, on $D(f,\widehat{f};\varepsilon)$, where $\widehat{f}$ and ε will be specified below.*

Assume also that W is a (quasi-) Banach space such that

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

and such that there exist $\delta_0 > 0$, $\beta \in \Omega(\delta_0)$ (cf. Appendix), such that the integral modulus of smoothness (cf. Appendix) satisfies

$$\omega_2(g;\delta)_{L_{p_1}} \leq \beta(\delta)\|g\|_W, \qquad \forall g \in W \forall \delta \in (0,\delta_0]. \tag{8}$$

Let us also assume that $f \in L_{p_1}(\mathrm{IR^d}) \cap \mathrm{W}$ and $\|f\|_W \leq L$ for some $L \in (0,\infty)$, $p_1 : 1 \leq p_1 \leq \infty$. Let $\widehat{f} = \widehat{f}_N \in \mathcal{L}_p\left(\mathrm{IR^d}, \mathrm{L_{p_1}(IR^d)}\right)^\rho$,$(0 < \rho < \infty)$ depend on either the observations $\{X_i\}_{i=1}^N$ or $\{(X_i,Y_i)\}_{i=1}^N$, where $\{X_i\}_{i=1}^N$ are discussed in the Appendix, while $\{Y_i\}_{i=1}^N$ are given in formula (1).

Let us also assume that there exists $\alpha \in \Omega(\delta_0)$, $k = k(\rho_1,\rho,W,L,\alpha) \in (0,\infty)$ and $N_0 = N_0(p_1,\rho,W,L,\alpha) \in (0,\infty)$, such that $E\|f-\widehat{f}_N\|^\rho_{L_{p_1}} \leq K\alpha\left(\frac{1}{N}\right), \forall N > N_0$.

Now let us consider $\widehat{f}_{+,\varepsilon}$, as defined in Formulas (2)–(5).

Then,

$$E\|f-\widehat{f}_{+,\varepsilon}\|^\rho_{L_{p_1}} \leq 2^{\rho_1}\left(3+2^{\frac{\rho}{\rho_1}}\right)^{\rho_1} K\alpha\left(\frac{1}{N}\right), \quad \forall N > N_0, \quad \forall \varepsilon \in (0,\varepsilon_N], \tag{9}$$

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

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

3 General variable order constraints

In this section the lower bound $v(x)$ and the upper bound $w(x)$ are Lebesgue-measurable functions defined everywhere on $\mathrm{IR^d}$, with $v(x) < w(x)$ for $x \in \mathrm{IR^d}$. (Note that here it will be essential that this inequality is strict.) In fact, we will even require the stronger condition:

$$S(v,x;2C_\Phi\varepsilon) \leq I(w,x;2C_\Phi\varepsilon), \quad \forall x \in supp\ f \subset \mathrm{IR^d}. \tag{11}$$

We note that if $supp\ f$ is compact and if v and w are continuous, then $v(x) < w(x)\ |_{supp\ f}$ implies that there exists $\varepsilon' = \varepsilon'(f,v,w)$ such that, for every $\varepsilon \in (0,\varepsilon']$, Formula (11) holds true.

The analogue of Lemma 2.1 in this general situation is:

Lemma 3.1 *Under the conditions of Lemma 2.1, without the assumption about convexity of v or concavity of w, suppose that (11) holds. Then, the following analogue of Formula (5) holds*

$$v(x) \leq \Phi_\varepsilon * \widehat{f}_+^\varepsilon(x) \leq w(x), \qquad x \in \mathbb{R}^{\mathrm{d}}. \tag{12}$$

Now we can formulate the analogue of Theorem 2.1 which will be valid for general variable order constraints, as follows.

Theorem 3.1 *Under the conditions of Theorem 2.1, assume, more generally, that v is not necessarily convex on $D(f, \widehat{f}; \varepsilon)$ and/or that w is not necessarily concave on $D(f, \widehat{f}; \varepsilon)$.*

Assume also that W_- and/or W_+ is a (quasi-) Banach space, such that $W_- \hookrightarrow L_{p_1}(\mathbb{R}^{\mathrm{d}})$, and/or $W_+ \hookrightarrow L_{p_1}(\mathbb{R}^{\mathrm{d}})$ and such that there exists $\gamma_- \in \Omega(\delta_0)$, and/or $\gamma_+ \in \Omega(\delta_0)$, satisfying

$$\tau_1(v;\delta)_{L_{p_1}} \leq \gamma_-(\delta)\|v\|_{W_-}, \text{ and/or } \tau_1(w;\delta)_{L_{p_1}} \leq \gamma_+(\delta)\|w\|_{W_+}, \quad \forall \delta \in (0, \delta_0]. \tag{13}$$

Consider $\widehat{f}_{+,\varepsilon}$, as defined in Formulas (6) and (7).

Then, the inequality (9) of Theorem 2.1 holds true, with the factor 2^{ρ_1} in the RHS of (9) being replaced by 4^{ρ_1}, and with (10) replaced by

$$\begin{aligned} \varepsilon_N \quad = \quad & \min\left\{\beta^{-1}\left(\frac{2}{L}\left(3+2^{\frac{\rho}{\rho_1}}\right)^{\frac{\rho_1}{\rho}} K^{\frac{1}{\rho}}\alpha\left(\frac{1}{N}\right)^{\frac{1}{\rho}}\right),\right. \\ & \frac{1}{C_\Phi}\gamma_-^{-1}\left(\frac{2}{\|v\|_{W_-}}\left(3+2^{\frac{\rho}{\rho_1}}\right)^{\frac{\rho_1}{\rho}} K^{\frac{1}{\rho}}\alpha\left(\frac{1}{N}\right)^{\frac{1}{\rho}}\right), \\ & \left.\frac{1}{C_\Phi}\gamma_+^{-1}\left(\frac{2}{\|w\|_{W_+}}\left(3+2^{\frac{\rho}{\rho_1}}\right)^{\frac{\rho_1}{\rho}} K^{\frac{1}{\rho}}\alpha\left(\frac{1}{N}\right)^{\frac{1}{\rho}}\right)\right\}. \end{aligned} \tag{14}$$

Remark 3.1 *If in Theorem 3.1 one of the order constraints $v(x) \leq f(x)$ or $f(x) \leq w(x)$ is absent, then the factor 4^{ρ_1} in the new version of (9) can be replaced by 3^{ρ_1}, and the corresponding term in the RHS of (14) can be omitted.*

In the remaining part of this section, we provide explicit computation of the rates $\beta(\delta)$, $\gamma_-(\delta)$ and $\gamma_+(\delta)$ when the spaces W, W_- and W_+ are Besov spaces, Triebel-Lizorkin spaces, A-spaces, or spaces with bounded Wiener-Young p-variation. For the case of $\beta(\delta)$, this computation was made in Dechevsky and MacGibbon (1999, 2009), and here we only list the corresponding results.

In the first lemma, W is a Besov space.

Lemma 3.2 (see Lemma 2.1 in Dechevsky and MacGibbon (1999, 2009)) *Assume that $d \in \mathbb{N}$, $1 \leq p_1 \leq \infty$, $1 \leq p \leq \infty$, $0 < q \leq \infty$, $d\left(\frac{1}{p} - \frac{1}{p_1}\right)_+ < s < \infty$.*

Assume further that, for $p \leq p_1$, $W = B^s_{pq}(\mathbb{R}^{\mathrm{d}})$, and for $p_1 < p$ there exists $R : 0 < R < \infty$, such that

$$W = W_R = \left\{f \in B^s_{pq}(\mathbb{R}^{\mathrm{d}}),\ \mathrm{supp\ f} \subset \left\{\mathrm{x} \in \mathbb{R}^{\mathrm{d}}; |\mathrm{x}| \leq \mathrm{R}\right\}\right\}$$

and W_R is endowed with the (quasi-) norm of $B^s_{pq}(\mathbb{R}^{\mathrm{d}})$.

Then, there exists

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

such that (8) is fulfilled for any $\delta_0 \in (0, 1]$*, and*

$$\beta(\delta) \leq K_1 \cdot \begin{cases} \delta^{\min\{2, s-d(\frac{1}{p}-\frac{1}{p_1})_+\}}, & s \neq 2 + d\left(\frac{1}{p} - \frac{1}{p_1}\right)_+, \\ \delta^2 \left(ln\frac{1}{\delta}\right)^{(1-\frac{\min\{p,2\}}{q})_+}, & s = 2 + d\left(\frac{1}{p} - \frac{1}{p_1}\right)_+, \quad p < \infty \\ \delta^2 \left(ln\frac{1}{\delta}\right)^{(1-\frac{1}{q})_+}, & s = 2 + d\left(\frac{1}{p} - \frac{1}{p_1}\right)_+, \quad p = \infty \end{cases} \tag{15}$$

In the next lemma, W is a Triebel-Lizorkin space.

Lemma 3.3 (see Lemma 2.2 in Dechevsky and MacGibbon (1999, 2009)) *Under the conditions of Lemma 3.2, assume that, for* $1 \leq p \leq p_1 < \infty$*,* $W = F^s_{pq}(\mathrm{I\!R}^{\mathrm{d}})$*, and, for* $1 < p_1 < p < \infty$*,*

$$W = W_R = \left\{f \in F^s_{pq}(\mathrm{I\!R}^{\mathrm{d}}),\ \mathrm{supp\ f} \subset \left\{\mathrm{x} \in \mathrm{I\!R}^{\mathrm{d}}; |\mathrm{x}| \leq \mathrm{R}\right\}\right\},$$

endowed with the (quasi-) norm in $F^s_{pq}(\mathrm{I\!R}^{\mathrm{d}})$*, where* $0 < s < r$*,* $0 < q \leq \infty$*.*

Then, the conclusion of Lemma 3.2 holds, with

$$\beta(\delta) \leq K_1 \cdot \begin{cases} \delta^{\min\{2, s-d(\frac{1}{p}-\frac{1}{p_1})_+\}}, & s \neq 2 + d\left(\frac{1}{p} - \frac{1}{p_1}\right)_+, \quad 0 < q < \infty, \\ \delta^2, & s = 2 + d\left(\frac{1}{p} - \frac{1}{p_1}\right)_+, \quad 0 < q \leq 2. \end{cases} \tag{16}$$

The next lemma is restricted to the case $d = 1$, and W is (essentially) the space TV^k_p, $1 \leq p < \infty$, $k = 0, 1$, of functions f whose k-th derivative $f^{(k)}$ is with bounded Wiener-Young p-variation.

Remark 3.2 *It is possible, in principle, to extend the definition of the space* TV^k_p *(hence, also the range of validity of the next lemma) for any dimension* $d \in \mathbb{N}$*, however, for a function of several variables the concept of variation can be extended in several essentially diverse ways, thus leading to a diversity of rates in the next lemma. Because of this, and in order to keep the exposition of the present study sufficiently concise and focused on the essentials, here we restrict the consideration of the space* TV^k_p *only to the case* $d = 1$*.*

Lemma 3.4 (see Lemma 2.3 in Dechevsky and MacGibbon (1999, 2009)) *Under the conditions of Lemma 3.2, assume that* $d = 1$*,* $k = 0, 1$*. Assume further that for* $1 \leq p \leq p_1 < \infty$*,*

$$W = L_p \frown \left(TV^k_p\right)^{\frac{1}{p}},$$

and, for $1 \leq p_1 \leq p < \infty$*,*

$$W = W_R = \left\{f \in W^k_p \cap TV^k_p,\ supp\ f \subset [-R, R]\right\}$$

and W *is endowed with the norm in* $W^k_p \frown \left(TV^k_p\right)^{\frac{1}{p}}$*.*

Then, the conclusion of Lemma 3.2 holds, with

$$\beta(\delta) = \delta^{k+\frac{1}{\max\{p,p_1\}}}. \tag{17}$$

Remark 3.3 *Some important details about the results in Lemmas 3.2–3.4 and their proofs are given in Remarks 2.1–2.4 in Section 2 of Dechevsky and MacGibbon (1999, 2009).*

The computation of the rate $\gamma_-(\delta)$ depending on W_-, and of $\gamma_+(\delta)$ depending on W_+, is essentially the same (cf. Formula (13), so it suffices to compute $\gamma_+(\delta)$ for suitable spaces W_+.

In the next lemma W_+ is an A-space.

Lemma 3.5 *Assume that* $d \in \mathbb{N}$, $1 \le p_1 \le \infty$, $1 \le p \le \infty$, $0 < q \le \infty$, $d\left(\frac{1}{p} - \frac{1}{p_1}\right)_+ < s < \infty$.

Assume further that, for $p \le p_1$, $W_+ = A^s_{pq}(\mathbb{R}^d)$, *and for* $p_1 < p$ *there exists* $R : 0 < R < \infty$, *such that*

$$W_+ = W_{+,R} = \left\{w \in A^s_{pq}(\mathbb{R}^d),\ \mathrm{supp\ w} \subset \left\{\mathrm{x} \in \mathbb{R}^d; |\mathrm{x}| \le \mathrm{R}\right\}\right\}$$

and $W_{+,R}$ *is endowed with the (quasi-)norm of* $A^s_{pq}(\mathbb{R}^d)$.

Then, there exists

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

such that (13) is fulfilled for w *for any* $\delta_0 \in (0,1]$ *and*

$$\gamma_+(\delta) \le K_1' \cdot \begin{cases} \delta^{\min\{1, s-d(\frac{1}{p}-\frac{1}{p_1})_+\}}, & s \ne 1 + d\left(\frac{1}{p} - \frac{1}{p_1}\right)_+, \\ \delta\left(ln\frac{1}{\delta}\right)^{(1-\frac{1}{q})_+}, & s = 1 + d\left(\frac{1}{p} - \frac{1}{p_1}\right)_+. \end{cases} \tag{18}$$

In the next three lemmas, we refine the rates, obtained under the assumptions of Lemma 3.5, for the particular case $d = 1$.

In the next lemma W_+ is a Besov space.

Lemma 3.6 *Under the conditions of Lemma 3.5, assume that* $d = 1$ *and that, when* $p_1 = 1$, $0 < q \le 1$ *holds. Impose on* s *the stronger restriction* $\max\left\{\frac{1}{p_1}, \left(\frac{1}{p} - \frac{1}{p_1}\right)_+\right\} < s < \infty$ *where* $p_1 > 1$, *and* $1 \le s < \infty$ *when* $p_1 = 1$.

Assume that, for $p \le p_1$, $W_+ = B^s_{pq}(\mathbb{R})$, *and for* $p_1 < p$ *there exists* $R : 0 < R < \infty$, *such that*

$$W_+ = W_{+,R} = \left\{w \in B^s_{pq}(\mathbb{R}),\ \mathrm{supp\ w} \subset [-\mathrm{R}, \mathrm{R}]\right\},$$

endowed with the (quasi-) norm of $B^s_{pq}(\mathbb{R})$.

Then, the conclusion of Lemma 3.5 holds in the following more refined form:

$$\gamma_+(\delta) \le K_1'' \cdot \begin{cases} \delta^{\min\{1, s-(\frac{1}{p}-\frac{1}{p_1})_+\}}, & s \ne 1 + \left(\frac{1}{p} - \frac{1}{p_1}\right)_+, \\ \delta\left(ln\frac{1}{\delta}\right)^{(1-\frac{\min\{p,2\}}{q})_+}, & s = 1 + \left(\frac{1}{p} - \frac{1}{p_1}\right)_+, \quad p < \infty, \\ \delta\left(ln\frac{1}{\delta}\right)^{(1-\frac{1}{q})_+}, & s = 1 + \left(\frac{1}{p} - \frac{1}{p_1}\right)_+, \quad p = \infty. \end{cases} \tag{19}$$

In the next lemma W_+ is a Triebel-Lizorkin space.

Lemma 3.7 *Under the conditions of Lemma 3.5, assume that* $d = 1$ *and that* $\max\left\{\frac{1}{p_1}, \left(\frac{1}{p} - \frac{1}{p_1}\right)_+\right\} < s < \infty$.

Assume also that, for $1 \le p \le p_1 < \infty$, $W_+ = F^s_{pq}(\mathbb{R})$, *and for* $1 < p_1 < p < \infty$,

$$W_+ = W_{+,R} = \left\{w \in F^s_{pq}(\mathbb{R}),\ \mathrm{supp\ w} \subset [-\mathrm{R}, \mathrm{R}]\right\},$$

endowed with the (quasi-) norm in $F^s_{pq}(\mathbb{R})$, *where* $0 < s < r$, $0 < q \le \infty$.

Then, the conclusion of Lemma 3.5 holds, with

$$\gamma_+(\delta) \le K_1'' \cdot \begin{cases} \delta^{\min\{1, s-(\frac{1}{p}-\frac{1}{p_1})_+\}}, & s \ne 1 + \left(\frac{1}{p} - \frac{1}{p_1}\right)_+, \quad 0 < q < \infty, \\ \delta, & s = 1 + \left(\frac{1}{p} - \frac{1}{p_1}\right)_+, \quad 0 < q \le 2. \end{cases} \tag{20}$$

Finally, in the next lemma W_+ is (essentially) the space TV_p^0, $1 \le p < \infty$.

Lemma 3.8 *Under the conditions of Lemma 3.5, assume that* $d = 1$.

For $1 \le p \le p_1 < \infty$, *let*

$$W_+ = L_p \curvearrowright \left(TV_p^0\right)^{\frac{1}{p}},$$

and for $1 \le p_1 \le p < \infty$, *let*

$$W_+ = W_{+,R} = \left\{w \in L_p \cap TV_p^0,\ supp\ w \subset [-R, R]\right\},$$

endowed with the norm in $L_p \curvearrowright \left(TV_p^0\right)^{\frac{1}{p}}$.

Then, the conclusion of Lemma 3.5 holds, with

$$\gamma_+(\delta) = \delta^{\frac{1}{\max\{p,p_1\}}}. \tag{21}$$

4 Applications

The range of application of our results is very general. One aspect of this broad range of applicability is that our method permits achieving rate-optimal constrained estimators of curves, surfaces and manifolds not only when the data consists of i.i.d. random vectors, but also for time series and trajectories of more general, not necessarily Markovian, stochastic processes. Another aspect of the generality of the method is that it can be applied to any kind of regression-function or density estimator (spline, kernel, wavelet estimator, etc.). All that is needed for the method to work for a given estimator $\widehat{f}$ is to provide the (global) risk estimation rate for $\widehat{f}$.

Several authors have also considered the problem of hazard rate estimation with censored data and have established rate-optimal estimators when the function was assumed to belong to certain smoothness classes (cf. Huber and MacGibbon (2004); Liang et al. (2005); Li (2004); Li et al. (2008) have also obtained wavelet estimators which are optimal over a large range of Besov function classes for nonparametric regression with censored data. In some applied problems it is quite conceivable that variable upper and lower constraints (possibly based on historical studies) could be found for the hazard function. The resulting constrained estimation problem would be amenable to our methods presented here.

Here we shall only consider some applications of our results to the wavelet and kernel estimators, as considered in Delyon and Juditsky (1996) and Lepski et al. (1997), respectively. In both cases the observations are i.i.d., but we shall not be relying on this fact. The results obtained can be considered as refinement, generalization and extension of the results in Section 3 of Dechevsky and MacGibbon (1999, 2009).

The rates $\alpha\left(\frac{1}{N}\right)$ for the estimator $\widehat{f}$ and Delyon and Juditsky (1996) and Lepski et al. (1997) are, as follows:

- for the wavelet (density or regression-function) estimator from Delyon and Juditsky (1996), for the d-dimensional case, with $1 \le p \le p_1 \le \infty$, $\frac{d}{p} < s$, $\rho = 2$, $\theta = \frac{s}{p_1} - \frac{d}{2}\left(\frac{1}{p} - \frac{1}{p_1}\right)$,

$$\alpha\left(\frac{1}{N}\right) = \begin{cases} \left(\frac{1}{N}\right)^{2\frac{s}{2s+d}}, & p_1 < \infty,\ \ \theta > 0, \\ \left(\frac{lnN}{N}\right)^{2\frac{s-d\left(\frac{1}{p}-\frac{1}{p_1}\right)}{2\left(s-\frac{d}{p}\right)+d}} (lnN)^{2(1-\frac{2p}{qp_1})_+}, & p_1 < \infty,\ \ \theta = 0, \\ \left(\frac{lnN}{N}\right)^{2\frac{s-d\left(\frac{1}{p}-\frac{1}{p_1}\right)}{2\left(s-\frac{d}{p}\right)+d}}, & p_1 < \infty,\ \ \theta < 0, \\ \left(\frac{lnN}{N}\right)^{2\frac{s-\frac{d}{p}}{2\left(s-\frac{d}{p}\right)+d}}, & p_1 = \infty, \end{cases} \tag{22}$$

  where $f \in B^s_{pq}(\mathbb{R}^{\mathrm{d}})$ with $q = \min\{2, p_1\}$;

- for the kernel (regression-function) estimator from Lepski et al. (1997), with $d = 1$, p, p_1, s and θ as in (22), $\rho = p_1 < \infty$,

$$\alpha\left(\frac{1}{N}\right) = \begin{cases} \left(\frac{1}{N}\right)^{p_1 \frac{s}{2s+1}}, & \theta > 0, \\ \left(\frac{1}{N}\right)^{p_1 \frac{s-\left(\frac{1}{p}-\frac{1}{p_1}\right)}{2\left(s-\frac{1}{p}\right)+1}} (lnN)^{p_1\left(\frac{1}{2}\cdot\frac{s-\left(\frac{1}{p}-\frac{1}{p_1}\right)}{2\left(s-\frac{1}{p}\right)+1}+\frac{1}{p_1}\right)}, & \theta = 0, \\ \left(\frac{1}{N}\right)^{p_1 \frac{s-\left(\frac{1}{p}-\frac{1}{p_1}\right)}{2\left(s-\frac{1}{p}\right)+1}} (lnN)^{\frac{p_1}{2}\cdot\frac{s-\left(\frac{1}{p}-\frac{1}{p_1}\right)}{2\left(s-\frac{1}{p}\right)+1}}, & \theta < 0, \end{cases} \tag{23}$$

  assuming that $f \in B^s_{pq}(\mathbb{R})$, with $q = p_1$.

In Section 3 of Dechevsky and MacGibbon (1999, 2009) we used the rates (22), (23) for $\alpha\left(\frac{1}{N}\right)$ and the constrained rates (15)–(17) for $\beta(\delta)$ to develop asymptotic theory for the respective wavelet and kernel-based estimators $\widehat{f}_{+,\varepsilon}$ in the case of constant lower and/or upper bounds v, w, by computing the range $(0, \varepsilon_N]$ of admissible values of ε. Our first new observation here is that all results in Section 3 or Dechevsky and MacGibbon (1999, 2009) automatically extend to the case of convex variable $v = v(x)$ and/or concave variable $w = w(x)$, as follows.

Corollary 4.1 *In the context of Section 3 of Dechevsky and MacGibbon (1999, 2009), assume, more generally, that $v = v(x)$ is a convex function, and $w = w(x)$ is a concave function, on $D(f, \widehat{f}; \varepsilon)$. If $\widehat{f}$ is the wavelet estimator from Delyon and Juditsky (1996), then Corollaries 1–11 in Subsection 3.1 of Dechevsky and MacGibbon (1999, 2009) continue to hold true. If $\widehat{f}$ is the kernel estimator from Lepski et al. (1997), then Corollaries 1, 2 in Subsection 3.2 of Dechevsky and MacGibbon (1999, 2009) continue to hold true.*

The results from Section 3 of Dechevsky and MacGibbon (1999, 2009) can also be extended to the general case of non-convex $v(x)$ or non-concave $w(x)$, and the proof of this extension is similar to the proof of Corollary 4.1, where the rates (18)–(21) for γ_+ and γ_- must now also be taken into consideration. It should be noted that, due to the logarithmic factors in the rates for $\alpha\left(\frac{1}{N}\right)$, $\beta(\delta)$, γ_+ and γ_-, it is not always possible to give explicitly the corresponding rates for ε_N in terms of elementary functions. That is why here we shall present the rate for ε_N not in the case when $\widehat{f}_{+,\varepsilon}$ achieves exactly the same rates as $\widehat{f}$, but rather when $\widehat{f}_{+,\varepsilon}$ achieves these rates within the "asymptopia" convention (see Donoho et al. (1995)), i.e., up to a logarithmic factor. Under this convention, it is possible to ignore the logarithmic factors in the rates (15)–(23) and work instead only with the polynomial factors in these rates. Now the rates for ε_N are polynomial in all cases, and can be given explicitly, as follows.

Corollary 4.2 *Under the assumptions of Theorem 3.1, let $\widehat{f}$ be the wavelet estimator from Delyon and Juditsky (1996), with $\alpha\left(\frac{1}{N}\right)$ satisfying (22), $p_1 \le \infty$, $\rho = 2$, or let $\widehat{f}$ be the kernel estimator from Lepski et al. (1997), with $\alpha = 1$, $\alpha\left(\frac{1}{N}\right)$ satisfying (23), $p_1 < \infty$, $\rho = p_1$. Assume further that $W = B^s_{pq}\left(\mathbb{R}^{\mathrm{d}}\right)$, $W_- = W_+ = A^{s'}_{p'q'}\left(\mathbb{R}^{\mathrm{d}}\right)$, where $1 \le p \le p_1$, $0 < q \le \infty$, $\frac{d}{p} < s < r$, $1 \le p' \le \infty$, $0 < q' \le \infty$, $d\left(\frac{1}{p'} - \frac{1}{p_1}\right)_+ < s' \le r$, $r \ge 2$. Suppose also that $\widehat{f} \in C^r_0\left(\mathbb{R}^{\mathrm{d}}\right)$ and $\Phi \in C^r_0\left(\mathbb{R}^{\mathrm{d}}\right)$. Then,*

(A) if $v = v(x)$ ($w = w(x)$) is convex (concave) on $D(f, \widehat{f}; \varepsilon)$, and if $\widehat{f}_{+,\varepsilon}$ is defined via (2), (3) (i.e., in the sense of Theorem 2.1), then $\widehat{f}_{+,\varepsilon}$ attains the rate of $\widehat{f}$ within the "asymptopia" convention if ε_N has the following rate:

$$\varepsilon_N = O\left(N^{-\frac{1}{2\left(s-\frac{d}{p}\right)+d}}\right), \tag{24}$$

for $\frac{d}{p_1} < s - d\left(\frac{1}{p} - \frac{1}{p_1}\right) \le \min\left\{2, p_1 d\left(\frac{1}{2} - \frac{1}{p_1}\right)\left(\frac{1}{p} - \frac{1}{p_1}\right)\right\}$;

$$\varepsilon_N = O\left(N^{-\frac{s}{(2s+d)\left(s-d\left(\frac{1}{p}-\frac{1}{p_1}\right)\right)}}\right), \tag{25}$$

for $\max\left\{\frac{d}{p_1}, p_1 d\left(\frac{1}{2}-\frac{1}{p_1}\right)\left(\frac{1}{p}-\frac{1}{p_1}\right)\right\} < s - d\left(\frac{1}{p}-\frac{1}{p_1}\right) \leq 2;$

$$\varepsilon_N = O\left(N^{-\frac{s-d\left(\frac{1}{p}-\frac{1}{p_1}\right)}{2\left(2\left(s-\frac{d}{p}\right)+d\right)}}\right), \tag{26}$$

for $\max\left\{\frac{d}{p_1}, 2\right\} < s - d\left(\frac{1}{p}-\frac{1}{p_1}\right) \leq \min\left\{r - d\left(\frac{1}{p}-\frac{1}{p_1}\right), p_1 d\left(\frac{1}{2}-\frac{1}{p_1}\right)\left(\frac{1}{p}-\frac{1}{p_1}\right)\right\};$

$$\varepsilon_N = O\left(N^{-\frac{s}{2(2s+d)}}\right), \tag{27}$$

for $\max\left\{\frac{d}{p_1}, 2, p_1 d\left(\frac{1}{2}-\frac{1}{p_1}\right)\left(\frac{1}{p}-\frac{1}{p_1}\right)\right\} < s - d\left(\frac{1}{p}-\frac{1}{p_1}\right).$

(B) if $v = v(x)$ ($w = w(x)$) is not necessarily a convex (concave) function on $D(f, \widehat{f}; \varepsilon)$, and if $\widehat{f}_{+,\varepsilon}$ is defined via (6), (7) (i.e., in the sense of Theorem 3.1), then $\widehat{f}_{+,\varepsilon}$ attains the rate of $\widehat{f}$ within the "asymptopia" convention, if ε_N has the following rate:

$$\varepsilon_N = O\left(N^{-\frac{1}{2\left(s-\frac{d}{p}\right)+d}}\right), \tag{28}$$

for $\frac{d}{p_1} < s - d\left(\frac{1}{p}-\frac{1}{p_1}\right) \leq \min\left\{1, p_1 d\left(\frac{1}{2}-\frac{1}{p_1}\right)\left(\frac{1}{p}-\frac{1}{p_1}\right), s' - d\left(\frac{1}{p'}-\frac{1}{p_1}\right)_+\right\};$

$$\varepsilon_N = O\left(N^{-\frac{s}{(2s+d)\left(s-d\left(\frac{1}{p}-\frac{1}{p_1}\right)\right)}}\right), \tag{29}$$

for $\max\left\{\frac{d}{p_1}, p_1 d\left(\frac{1}{2}-\frac{1}{p_1}\right)\left(\frac{1}{p}-\frac{1}{p_1}\right)\right\} < s - d\left(\frac{1}{p}-\frac{1}{p_1}\right) \leq \min\left\{1, s' - d\left(\frac{1}{p'}-\frac{1}{p_1}\right)_+\right\};$

$$\varepsilon_N = O\left(N^{-\frac{s-d\left(\frac{1}{p}-\frac{1}{p_1}\right)}{2\left(s-\frac{d}{p}\right)+d}\cdot\max\left\{1, \frac{1}{s'-d\left(\frac{1}{p'}-\frac{1}{p_1}\right)_+}\right\}}\right), \tag{30}$$

for $\max\left\{\frac{d}{p_1}, \min\left\{1, s' - d\left(\frac{1}{p'}-\frac{1}{p_1}\right)_+\right\}\right\} < s - d\left(\frac{1}{p}-\frac{1}{p_1}\right) \leq p_1 d\left(\frac{1}{2}-\frac{1}{p_1}\right)\left(\frac{1}{p}-\frac{1}{p_1}\right);$

$$\varepsilon_N = O\left(N^{-\frac{s}{2s+d}\cdot\max\left\{1, \frac{1}{s'-d\left(\frac{1}{p'}-\frac{1}{p_1}\right)_+}\right\}}\right), \tag{31}$$

for $\max\left\{\frac{d}{p_1}, p_1 d\left(\frac{1}{2}-\frac{1}{p_1}\right)\left(\frac{1}{p}-\frac{1}{p_1}\right), \min\left\{1, s' - d\left(\frac{1}{p'}-\frac{1}{p_1}\right)_+\right\}\right\} < s - d\left(\frac{1}{p}-\frac{1}{p_1}\right).$

As seen from Corollary 4.2, the rates in the case of general variable constraints (see (28)–(31)) achieve the rates in the special case of convex lower and concave upper order constraints (see (24)–(27)) (and are independent of the regularity of the order constraints) only when the order constraints $v(x)$, $w(x)$ are "more regular" than $f(x)$ itself, in the sense that $s - d\left(\frac{1}{p}-\frac{1}{p_1}\right) \leq \min\left\{1, s' - d\left(\frac{1}{p'}-\frac{1}{p_1}\right)_+\right\}$ must hold.

Remark 4.1 *In Corollary 4.2 we considered the case $W = B^s_{pq}\left(\mathbb{R}^d\right)$ and $W_- = W_+ = A^{s'}_{p'q'}\left(\mathbb{R}^d\right)$, where for the rates we use Lemmas 3.2 and 3.5. Under different restrictions for the regularity parameters, we can use each of Lemmas 3.2–3.4 for W versus each of Lemmas 3.5–3.8 for $W_- = W_+$ to obtain analogous results to those of Corollary 4.2 for all possible couples of space classes (W, $W_- = W_+$) (Besov spaces, A-spaces, Triebel-Lizorkin spaces, spaces of functions with bounded Wiener-Young p-variation). Details will not be provided here.*

Remark 4.2 *Using the rates in parts A and B of Corollary 4.2 (and its analogues discussed in Remark 4.1) it is possible to obtain the (generally, faster) sharp rates for ε_N under which $\widehat{f}_{+,\varepsilon}$ attains the rates of $\widehat{f}$ (within the "asymptopia") uniformly in (some, or all of) the regularity parameters s, s', p_1, p, p', q, q' in the whole of their respective ranges of admissibility. We note that even in the simplest case of constant v and w, these "uniform" results constitute sharpening of the respective results in Section 3 of* Dechevsky and MacGibbon *(1999, 2009) (for the case of "asymptopia"). We omit the technical elaboration of the details here.*

5 Proofs

Proof of Lemma 2.1. By the definition of Φ_ε, the integral in (3) is a convergent limit of convex combinations of values of $\widehat{f}_+(x)$ at some points $x_l \in \mathbb{R}^d$ where Formula (4) is satisfied. At the same time, again by definition of Φ_ε, the same convex combinations of the points x_l themselves converge to the centre of mass with coordinates

$$x_{0,i} = \int_{\mathbb{R}^d} \Phi_\varepsilon(x-\xi)\xi_i \, d\xi, \qquad i = 1, \ldots, d.$$

From this and from the symmetry of Φ and $\int_{\mathbb{R}^d} \Phi(\xi)\, d\xi = 1$ it follows that $x_{0,i} = x_i$, $i = 1, \ldots, d$, where $(x_1, \ldots, x_d)^T = x$.

This, together with (4), the convexity of v, and the concavity of w, implies (5). $\square$

Proof of Theorem 2.1. This follows the lines of the proof of Theorem 2.1 in Dechevsky and MacGibbon (1999, 2009) where, in the last stage (invoking the lattice property), one now uses Lemma 2.1 instead. $\square$

Proof of Lemma 3.1. By the properties of the lower and upper Baire functions, it follows from (6) that, for any $x \in \mathbb{R}^d$,

(P1) $\quad S(v, x; C_\Phi\varepsilon) \le S(v, \xi; 2C_\Phi\varepsilon) \le \widehat{f}^\varepsilon_+(\xi) \le I(w, \xi; 2C_\Phi\varepsilon) \le I(w, x; C_\Phi\varepsilon)$

holds for any $\xi : |x - \xi| \le C_\Phi\varepsilon$. Multiplying (P1) by $\Phi(x - \xi) \ge 0$, using the fact that *supp* $\Phi_\varepsilon(x - \cdot) \subset \{\xi : |x - \xi| \le C_\Phi\varepsilon)\}$, and that $\int_{\mathbb{R}^d} \Phi(y)\, dy = 1$, we obtain after integrating in ξ and using the properties of the lower and upper Baire functions once again

(P2) $\quad v(x) \le S(v, x; C_\Phi\varepsilon) \le \Phi_\varepsilon * \widehat{f}^\varepsilon_+(x) \le I(w, x; C_\Phi\varepsilon) \le w(x),$

which is the desired analogue (12) of (5). $\square$

Proof of Theorem 3.1. To avoid misunderstanding of the notation in this proof, we stress here that in the present Theorem 3.1 $\widehat{f}_{+,\varepsilon}$ has the meaning of (7), which means that $\widehat{f}_{+,\varepsilon} = \left(\widehat{f}^\varepsilon_+\right)_\varepsilon$ with $\widehat{f}^\varepsilon_+$ defined in (6). In contrast to this (and with slight abuse of notation), in Theorem 2.1 $\widehat{f}_{+,\varepsilon}$ has the meaning of (3), i.e., there $\widehat{f}_{+,\varepsilon} = \left(\widehat{f}_+\right)_\varepsilon$, where $\widehat{f}_+$ is defined in (2). (In the context of the current Theorem 3.1 there is no special notation for the quantity $\left(\widehat{f}_+\right)_\varepsilon$.) With this in mind, we obtain

$$\text{(P3)}\quad \left(E\left\|f-\widehat{f}_{+,\varepsilon}\right\|_{L_{p_1}}^{\rho}\right)^{\frac{1}{\rho_1}} = \left(E\left\|f-\left(\widehat{f}_{+}^{\varepsilon}\right)_{\varepsilon}\right\|_{L_{p_1}}^{\rho}\right)^{\frac{1}{\rho_1}}$$
$$= \left(E\left\|f-\left(\widehat{f}_{+}\right)_{\varepsilon}+\left(\widehat{f}_{+}\right)_{\varepsilon}-\left(\widehat{f}_{+}^{\varepsilon}\right)_{\varepsilon}\right\|_{L_{p_1}}^{\rho}\right)^{\frac{1}{\rho_1}}$$
$$\le \left(E\left\|f-\left(\widehat{f}_{+}\right)_{\varepsilon}\right\|_{L_{p_1}}^{\rho}\right)^{\frac{1}{\rho_1}} + \left(E\left\|\left(\widehat{f}_{+}\right)_{\varepsilon}-\left(\widehat{f}_{+}^{\varepsilon}\right)_{\varepsilon}\right\|_{L_{p_1}}^{\rho}\right)^{\frac{1}{\rho_1}} = \mathcal{J}_1+\mathcal{J}_2.$$

The evaluation of $\mathcal{J}_1$ now goes on in exactly the same way as the proof of Theorem 2.1 (In fact, it is the same proof and we shall not discuss it here in detail). For the evaluation of $\mathcal{J}_2$ we use consecutively Lemmas A.3 and A.4 in Dechevsky and MacGibbon (1999, 2009) to obtain:

$$\text{(P4)}\quad \mathcal{J}_2 = \left(E\left\|\left(\widehat{f}_{+}-\widehat{f}_{+}^{\varepsilon}\right)_{\varepsilon}\right\|_{L_{p_1}}^{\rho}\right)^{\frac{1}{\rho_1}} \le \left(E\left\|\widehat{f}_{+}-\widehat{f}_{+}^{\varepsilon}\right\|_{L_{p_1}}^{\rho}\right)^{\frac{1}{\rho_1}}$$

and, by (P4) and Lemma P1 (given immediately after this proof),

$$\text{(P1)}\quad \mathcal{J}_2 \le \left(E\left\|\max\left\{\omega_1(v,\cdot;C_\Phi\varepsilon),\omega_1(w,\cdot;C_\Phi\varepsilon)\right\}\right\|_{L_{p_1}}^{\rho}\right)^{\frac{1}{\rho_1}}$$
$$\le \left(E\left\|\omega_1(v,\cdot;C_\Phi\varepsilon)+\omega_1(w,\cdot;C_\Phi\varepsilon)\right\|_{L_{p_1}}^{\rho}\right)^{\frac{1}{\rho_1}}$$
$$\le \tau_1\left(v;C_\Phi\varepsilon\right)_{L_{p_1}}^{\frac{\rho}{\rho_1}} + \tau_1\left(w;C_\Phi\varepsilon\right)_{L_{p_1}}^{\frac{\rho}{\rho_1}} = \mathcal{J}_{21}+\mathcal{J}_{22}.$$

(If one of the order constraints is absent, then so is the corresponding term on the RHS of (P5) also absent.) Requiring now that the contribution of each of $\mathcal{J}_{21}$ and/or $\mathcal{J}_{22}$ does not exceed $\left(3+2^{\frac{\rho}{\rho_1}}\right)K^{\frac{1}{\rho_1}}\alpha\left(\frac{1}{N}\right)^{\frac{1}{\rho_1}}$ (cf. Formula (10) in Dechevsky and MacGibbon (1999, 2009)), after simple computations, we obtain (14). □

The following lemma has been used in the proof of Theorem 3.1.

Lemma P1 *For any $\varepsilon>0$, it is true that*

$$\text{(P6)}\quad \left|\widehat{f}_{+}-\widehat{f}_{+}^{\varepsilon}\right| \le \max\left\{\omega_1(v,x;C_\Phi\varepsilon),\omega_1(w,x;C_\Phi\varepsilon)\right\}, \quad \forall x\in\mathbb{R}^{\mathrm{d}}.$$

If one of the order constraints is absent, then so is absent the corresponding term in the RHS of (P6).

Proof of Lemma P1. Denote $a=v(x)$, $a_1=S(v,x;C_\Phi\varepsilon)$, $b=w(x)$, $b_1=I(w,x;C_\Phi\varepsilon)$, where, by the assumptions made in the beginning of Section 3, especially Formula (11), $a\le a_1\le b_1\le b$ holds. Denote also $\lambda=\widehat{f}(x)$. Then,

$$\text{(P7)}\quad \left|\widehat{f}_{+}-\widehat{f}_{+}^{\varepsilon}\right| = \begin{cases} a_1-a, & \lambda\le a\le a_1\le b_1\le b,\\ \lambda-a_1\le a_1-a, & a\le\lambda\le a_1\le b_1\le b,\\ 0, & a\le a_1\le\lambda\le b_1\le b,\\ b_1-\lambda\le b-b_1, & a\le a_1\le b_1\le\lambda\le b,\\ b-b_1, & a\le a_1\le b_1\le b\le\lambda \end{cases}$$

and, by the definitions of the lower and upper Baire functions and the local modulus of smoothness, $a_1-a=|a_1-a|\le\omega_1(v,x;C_\Phi\varepsilon)$, $b-b_1=|b-b_1|\le\omega_1(w,x;C_\Phi\varepsilon)$, and the lemma follows from (P7). □

Lemmas 3.2–3.4 have been proved in Dechevsky and MacGibbon (1999, 2009).

Proof of Lemma 3.5. Consider first the case $p=p_1$. If $s\in(0,1)$, we have

$$\text{(P8)}\quad \tau_1(w;\delta)_{L_p}\le c_1\delta^s\|w\|_{A_{p\infty}^s}\le c_2\delta^s\|w\|_{A_{pq}^s}.$$

If $1<s<r<\infty$, then (P8) follows from the Marchaud-type inequality of second type for τ-moduli, with $k=[r]+1$

(P9) $\tau_1(w;\delta)_{L_p} \le c_3'\tau_k(w;\delta)_{L_p} + c_3''\delta\|w\|_{L_p} + c_3'''\delta\int_\delta^1 \xi^{-2}\omega_k(w;\xi)_{L_p}\, d\xi$

(see Dechevski (1988a), Subsection 1.3, Theorem 2, part (a), or case (i) of Theorem 5 in the R&D Report Dechevsky (2008) or its published version Dechevsky (2007a)), and the embedding $A^1_{p1} \hookleftarrow A^s_{pq}$, $1 \le p \le \infty$, $0 < q \le \infty$, $s > 1$.

If $s = 1$, $1 \le p \le \infty$ and $0 < q \le 1$, then (P8) follows from (P9) and the embedding $A^1_{p1} \hookleftarrow A^1_{pq}$.

If $s = 1$, $1 \le p \le \infty$ and $1 < q \le \infty$, then a simple upper bound of the RHS of (P9) gives, by also using Hölder's inequality for $q = \infty$,

$$\text{(P10)} \quad \tau_1(w;\delta)_{L_p} \le c_4\delta\left(\|w\|_{L_p} + \int_\delta^1 \xi^{-2}\tau_k(w;\xi)_{L_p}\, d\xi\right) \le$$
$$c_5\delta\left(\|w\|_{L_p} + \left(\int_\delta^1 \frac{d\xi}{\xi}\right)\cdot\left[\sup_{\xi>0}\left(\xi^{-1}\tau_k(w;\xi)_{L_p}\right)\right]\right) \le$$
$$c_5\delta\left(\|w\|_{L_p} + c_6\left(ln\tfrac{1}{\delta}\right)\|w\|_{A^1_{p\infty}}\right) \le c_7\delta\left(ln\tfrac{1}{\delta}\right)\cdot\|w\|_{A^1_{p\infty}},$$

where for the last inequality the embedding $L_p \hookleftarrow A^1_{p\infty}$ was used. (P10) proves (18) in the case $s = 1$, $1 \le p \le \infty$, $q = \infty$. Let us consider now the case $1 < q < \infty$. Since $\tau_1(\cdot;\delta)_{L_p}$ is an equivalent semi-norm in the space $A_{p,\delta} + \dot{W}^1_{p,\delta}$ with the semi-norm $K\left(\delta,\cdot;A_{p,\delta},\dot{W}^1_{p,\delta}\right)$, we can apply the real interpolation method between (P8) (for $q = 1$) and (P10). (For the relevant details about the use of the real interpolation method in this context, see Dechevski (1988,a); Dechevsky (2008, 2007a) and Theorem 6.4.5 (2) in Bergh and Löfström (1976), valid also for A-spaces.) Namely, we apply the real (η, q)-method, $\eta \in (0,1)$, $q \in (1,\infty)$ where $\eta = 1 - \frac{1}{q} \in (0,1)$, so that $\frac{1}{q} = \frac{1-\eta}{1} + \frac{\eta}{\infty} = 1 - \eta$. As a result, we obtain

$$\tau_1(w;\delta) \le c_8\delta\left(ln\frac{1}{\delta}\right)^{1-\frac{1}{q}}\|w\|_{A^1_{pq}},$$

i.e., Formula (18) for $s = 1$, $1 \le p \le \infty$, $1 < q < \infty$. Thus, Lemma 3.5 is completely proven in the case $p_1 = p$.

For $p < p_1$ we use the Sobolev-type embedding $A^s_{pq} \hookrightarrow A^{s-d(\frac{1}{p}-\frac{1}{p_1})}_{p_1 q}$, $0 < p \le p_1 \le \infty$, $0 < q \le \infty$, $s > d(\frac{1}{p} - \frac{1}{p_1})_+$, (which can be proved similarly to the proof of the analogous Sobolev-type embedding for Besov spaces). In this way, we reduce the consideration of the case $p < p_1$ to the already proven case $p = p_1$.

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

$$\tau_1(w;\delta)_{L_{p_1}} \le c(d,p,p_1)R^{\left(\frac{1}{p_1}-\frac{1}{p}\right)d}\tau_1(w;\delta)_{L_p},$$

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

Proof of Lemma 3.6 (outline). Let $p_1 = p$ first. Then, since $d = 1$ and $s > \frac{1}{p}$ for $p > 1$, the isomorphism $A^s_{pq} \rightleftharpoons B^s_{pq}$ holds for any $p : 1 \le p \le \infty$, $0 < q \le \infty$. When $p = 1$, there is also the isomorphism $A^s_{1q} \rightleftharpoons B^s_{1q}$, $s \ge 1$, $0 < q \le 1$. Now the proof follows (with respective modifications) the proof of Lemma 3.2 (Lemma 1 in Section 2 of Dechevsky and MacGibbon (1999, 2009)). □

Proof of Lemma 3.7 (outline). Let $p_1 = p$. Since $d = 1$ and $s > \frac{1}{p}$, $\tau_1(w;\delta)_{L_{p_1}} \le c_9\delta^s\|w\|_{F^s_{p2}}$ holds for $s \in (\frac{1}{p}, 1)$. The rest of the proof is similar to the proof of Lemma 3.3 (Lemma 2 in Section 2 of Dechevsky and MacGibbon (1999, 2009)). □

Proof of Lemma 3.8 (outline). Again let $p_1 = p$. Since $d = 1$, $\tau_1(w;\delta)_{L_{p_1}} \le (\delta V_p w)^{\frac{1}{p}}$ holds. The rest of the proof is similar to the proof of Lemma 3.4 (Lemma 3 in Section 2 of Dechevsky and MacGibbon (1999, 2009)). □

Proof of Corollary 4.1 (outline). The proof follows the lines of the proofs of the corresponding results in Section 3 of Dechevsky and MacGibbon (1999, 2009) where, instead of Theorem 1 in Section 2 of Dechevsky and MacGibbon (1999, 2009), we use the more general Theorem 2.1 of the present paper. □

Proof of Corollary 4.2. Since, by the "asymptopia" convention, the rates for $\widehat{f}$ and $\widehat{f}_{+,\varepsilon}$ have to coincide only up to the logarithmic factor, in the place of $\alpha\left(\frac{1}{N}\right)$, $\beta(\delta)$, and $\gamma_-(\delta)=\gamma_+(\delta)$, we may consider their main, polynomial-factor, rate

$$\text{(P11)}\quad \tilde{\alpha}\left(\tfrac{1}{N}\right)=\begin{cases}\left(\frac{1}{N}\right)^{\rho\frac{s}{2s+d}}, & \theta>0,\\ \left(\frac{1}{N}\right)^{\rho\frac{s-d\left(\frac{1}{p}-\frac{1}{p_1}\right)}{2\left(s-\frac{d}{p}\right)+d}}, & \theta\le 0,\end{cases}$$

$$\text{(P12)}\quad \tilde{\beta}(\delta)=\delta^{\min\{2,s-d(\frac{1}{p}-\frac{1}{p_1})\}}.$$

$$\text{(P13)}\quad \tilde{\gamma}_-(\delta)=\tilde{\gamma}_+(\delta)=\delta^{\min\{1,s'-d(\frac{1}{p'}-\frac{1}{p_1})_+\}},$$

in view of Formulas (22), (23), (15) and (18). Now we can compute $\tilde{\beta}^{-1}$ and $\tilde{\gamma}^{-1}$ explicitly from (P12) and (P13):

$$\text{(P14)}\quad \delta=\tilde{\beta}^{-1}(\tilde{\beta}(\delta))=\tilde{\beta}^{\max\left\{\frac{1}{2},\frac{1}{s-d\left(\frac{1}{p}-\frac{1}{p_1}\right)}\right\}},$$

$$\text{(P15)}\quad \delta=\tilde{\gamma}_+^{-1}(\tilde{\gamma}_+(\delta))=\tilde{\gamma}^{\max\left\{1,\frac{1}{s'-d(\frac{1}{p'}-\frac{1}{p_1})_+}\right\}}.$$

Then, in part A of the corollary the rate for ε_N is $O\left(\tilde{\beta}^{-1}\left(\tilde{\alpha}\left(\frac{1}{N}\right)^{\frac{1}{\rho}}\right)\right)$, and in part B the rate for ε_N is the higher of the rates $O\left(\tilde{\beta}^{-1}\left(\tilde{\alpha}\left(\frac{1}{N}\right)^{\frac{1}{\rho}}\right)\right)$ and $O\left(\tilde{\gamma}_+^{-1}\left(\tilde{\alpha}\left(\frac{1}{N}\right)^{\frac{1}{\rho}}\right)\right)$.

Consider part A first. We have four cases depending on whether $\theta\le 0$ or $\theta>0$ and on whether $s-d\left(\frac{1}{p}-\frac{1}{p_1}\right)\le 2$ or $s-d\left(\frac{1}{p}-\frac{1}{p_1}\right)>2$. After computations, using (P11) and (P14), we obtain (24)–(27). The proof of part A is complete.

Consider now part B. We now have eight cases, depending on the signs of the quantities in part A and, additionally, on whether $s'-d\left(\frac{1}{p'}-\frac{1}{p_1}\right)_+\le 1$ or $s'-d\left(\frac{1}{p'}-\frac{1}{p_1}\right)_+>1$. In the analysis of these eight cases, using (P11), (P14) and (P15), we note that some of the cases, when $\min\left\{1,s'-d\left(\frac{1}{p'}-\frac{1}{p_1}\right)_+\right\}$ $\le\min\left\{2,s-d\left(\frac{1}{p}-\frac{1}{p_1}\right)\right\}$, can be united, and also that two of the cases, when $\min\left\{2,s-d\left(\frac{1}{p}-\frac{1}{p_1}\right)\right\}$ $>\min\left\{1,s'-d\left(\frac{1}{p'}-\frac{1}{p_1}\right)\right\}$, are void. After computation, this leads to (28)–(31), which completes the proof of part B. □

6 Concluding remarks

Here we provide the concluding remarks given in Section 6 of Dechevsky and MacGibbon (2001), together with some recent updates.

Remark 6.1 *In the proof of Theorems 2.1 and 3.1 (the evaluation of $\mathcal{J}_1$), the lattice property plays an essential role (see the proof of Theorem 1 in Section 2 of Dechevsky and MacGibbon (1999, 2009) for the details). This means that (unlike the rates in the $B^{s_1}_{p_1q_1}$-metric, $(s_1>0)$ for $\widehat{f}$ as in Delyon and Juditsky (1996)) for $\widehat{f}_{+,\varepsilon}$ the L_{p_1}-norm in the risk in Theorems 2.1 and 3.1 cannot be replaced by a $B^{s_1}_{p_1q_1}$-norm for any $s_1>0$.*

Remark 6.2 *We note that $\widehat{f}$, as defined in* Lepski et al. (1997)*, may be nonsmooth, and even discontinuous at some points. Thus, in this case $\widehat{f}$ may violate both the order and the smoothness constraints, while $\widehat{f}_{+,\varepsilon}$ obeys all of these constraints.*

Remark 6.3 *Due to the regularization effect of the convolution with Φ_ε in the definition of $\widehat{f}_{+,\varepsilon}$, Gibbs phenomenon with $\widehat{f}_{+,\varepsilon}$ is much reduced compared to that of $\widehat{f}$.*

Remark 6.4 *In* Dechevsky et al. (2001) *and in the forthcoming* Dechevsky and MacGibbon (2010) *we study the properties of the estimators obtained by replacing the convolution integral in the definition of $\widehat{f}_{+,\varepsilon}$ with quadrature formulae. These new estimators of kernel type preserve the order constraints just as $\widehat{f}_{+,\varepsilon}$, and they have the same regularity as the kernel $\Phi \in C_0^r$ (see Appendix 0). These quadrature-based kernel estimators 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. In the case when spline-wavelets, spline-kernels and a B-spline kernels Φ are used, it is possible, at least for $d = 1$, to compute the integral explicitly, in closed form, which can be obtained manually or by automatic symbolic integration (for more details, see* Dechevsky et al. (2001)*, the end of Section 3). The use of quadrature formulae in this case also leads to the exact value of the integral, as long as the quadrature formulae are chosen to be exact of order not less than the order of the spline. In the case of density estimation, it can also facilitate normalization. In the forthcoming paper* Dechevsky and MacGibbon (2010) *we consider the problem in the case of general variable constraints, and make use of the results of Theorems 2.1 and 3.1 of the present work.*

Remark 6.5 *Besov spaces have been widely used in recent years to model the smoothness constraints in nonparametric density and regression-function estimation. Here we would like to stress that, in our opinion, Besov spaces are a natural space scale for these types of estimation problems only for density estimation and when the regression problem is with random design. The natural space scale for regression problems with deterministic design are A-spaces. In this context Besov spaces appear only as far as $B^s_{pq} \rightleftharpoons A^s_{pq}$ for $s > \frac{d}{p}$ (see* Dechevsky et al. (1999)*, Appendix B, B12 for further details, as well as Appendix 0).*

Remark 6.6 *It would be interesting to develop analogues of the results of* Delyon and Juditsky (1996) *and* Lepski et al. (1997) *in terms of homogeneous Besov, Sobolev, etc., spaces. For instance, this would allow a better assessment of the role of the density's tail weight versus its regularity, where interesting new results can be expected in the case of densities with non-compact support (see also Remark 2.2.4 in* Dechevsky and Penev (1997)*).*

Remark 6.7 *For density estimation, a natural lower bound for f is $v \equiv 0$ (or, eventually, $v(x) \geq 0$ with $\int_{\mathbb{R}^d} v(x)\ dx < 1$). In this case it is desirable that the estimator also integrates to 1 on $\mathbb{R}^d$, i.e., is a density itself. If there is no upper constraint $w(x)$, it suffices to normalize the estimator $\widehat{f}_{+,\varepsilon}$ by a factor $c = \frac{1}{\int_{\mathbb{R}^d} \widehat{f}_{+,\varepsilon}(x)\ dx}$. However, if there is also an upper bound w, it may happen that $c\widehat{f}_{+,\varepsilon}(x) > w(x)$ for some x. This difficulty can be overcome by a modification in the definition of $\widehat{f}_+$ in (2) and $\widehat{f}^\varepsilon_+$ in (6), as follows.*

- *In the special case of convex/concave lower/upper constraints:*

$$\widehat{f}_{c,+}(x) = \min\left\{w(x), c \cdot \max\left\{0, \widehat{f}(x)\right\}\right\},$$

  when $v \equiv 0$ (is convex non-negative), $w(x) \geq 0$ is concave on $D(f, \widehat{f}; \varepsilon)$, with $\int_{\mathbb{R}^d} w(x)\ dx > 1$.

- *In the general case of variable constraints:*

$$\widehat{f}^\varepsilon_{c,+}(x) = \min\left\{I(w, x, 2C_\Phi\varepsilon), c \cdot \max\left\{S(v, x; 2C_\Phi\varepsilon), \widehat{f}(x)\right\}\right\},$$

  for the case of general variable constraints $0 \leq v(x) < w(x)$ with (11), $\int_{\mathbb{R}^d} v(x)\ dx < 1$, and $\int_{\mathbb{R}^d} w(x)\ dx > 1$.

The value for c is the only solution of $\int_{\mathbb{R}^d} \widehat{f}_{c,+}(x)\, dx = 1$*, respectively* $\int_{\mathbb{R}^d} \widehat{f}^{\varepsilon}_{c,+}(x)\, dx = 1$*. Convolving with* Φ_ε *(which is a density) ensures that* $\Phi_\varepsilon * \widehat{f}_{c,+}$*, respectively* $\Phi_\varepsilon * \widehat{f}^{\varepsilon}_{c,+}$ *is a density itself, too.*

Remark 6.8 *If the parameter* $\varepsilon = \varepsilon(x)$ *for* $x \in \mathbb{R}^d$ *in the definition of* $\widehat{f}_{+,\varepsilon}$ *were made space-dependent, the estimator* $\widehat{f}_{+,\varepsilon}$ *can be made not only rate-optimal as* $\widehat{f}$*, but also ideally spatially adaptive for moderate samples, if* $\widehat{f}$ *is such itself. (Both the wavelet estimator from Delyon and Juditsky (1996) and the kernel estimator from Lepski et al. (1997) enjoy such ideal spatial adaptivity.) For example, let us consider the adaptive kernel estimator* $\widehat{f}(x) = \widehat{f}_{\widehat{h}(x)}(x)$ *of Lepski et al. (1997), with variable bandwidth* $\widehat{h}(x)$*. Denote* $\widehat{h}_{min} = \min_x \widehat{h}(x)$ *(by the construction of* $\widehat{h}(x)$ *in Lepski et al. (1997) it follows that* $\widehat{h}_{min}$ *always exists and is strictly positive, when* f *is compactly supported). One simple data-adaptive choice of* ε *can be* $\varepsilon = \min\left\{\frac{\widehat{h}_{min}}{2}, \varepsilon_N\right\}$*; a more refined choice, involving variable bandwidth* $\varepsilon(x)$ *is* $\varepsilon(x) = \min\left\{\frac{\widehat{h}(x)}{2}, \varepsilon_N\right\}$*. Further improvement of the adaptivity of* $\widehat{f}_{+,\varepsilon(\cdot)}(\cdot)$ *can be achieved, if we make use of the fact that the ideas of the proofs of Theorems 2.1, 3.1 and Lemmas 3.2–3.8 can be modified for local pointwise estimates. (In fact, tracing our proofs shows that obtaining local estimates for* $\widehat{f}_{+,\varepsilon}$ *and* $\widehat{f}^{\varepsilon}_{+,\varepsilon}$ *is an intermediate stage of these proofs; when there is a local pointwise rate available for* $\widehat{f}(x)$*, this stage is sufficient; when only global rate in* L_{p_1}*-metric is known for* $\widehat{f}$*, then this stage has to be followed by a stage of* L_{p_1}*-global averaging of the local rates for* $\widehat{f}_{+,\varepsilon}$*, as is the case with Theorems 2.1 and 3.1.) In the case of local rates available for* $\widehat{f}(x)$ *the variable bandwidth can be defined by* $\bar{\varepsilon}(x) = \min\left\{\frac{\widehat{h}(x)}{2}, \varepsilon_N(x)\right\}$*. At this point it should be recalled that the method of defining* $\widehat{h}(x)$ *in Lepski et al. (1997) generates piecewise continuous* $\widehat{h}(x)$ *which is a step function. To obtain sufficiently smooth varying bandwidth* $\tilde{\varepsilon}(x)$*, it now suffices to smooth* $\bar{\varepsilon}$*, e.g., by using the kernel* $\Phi(x)$ *and bandwidth* $\frac{\widehat{h}_{min}}{2}$*. This would guarantee that* $\widehat{f}_{+,\tilde{\varepsilon}(\cdot)}(\cdot) \in W$*.*

Remark 6.9 *The papers Dechevsky and Penev (1997, 1998); Penev and Dechevsky (1997); Pinheiro and Vidakovic (1997); Dechevsky et al. (1999); Dechevsky and MacGibbon (1999); Dechevsky et al. (2000, 2001); Dechevsky and MacGibbon (2001) appeared between 1997 and 2001; in the period since then several new publications relevant to shape-preserving approximation/estimation have appeared. Here we shall mention explicitly the following ones.*

1. *In his Nobel Lecture McFadden (2000), McFadden noted the need for extending the results of Anastassiou and Yu (1992); Dechevsky and Penev (1997) to high-dimensional shape-preserving approximation.*
2. *Mammen et al. (2001) proposed a general projection framework for constrained smoothing, the emphasis being on preserving monotonicity and related topics. The topic of positivity constraints (i.e., constant lower and/or upper constraints) is mentioned very briefly in the concluding section, where a reference to Dechevsky and MacGibbon (1999); Dechevsky et al. (2000) is provided. Clearly, at that time the authors have been unaware of the more general results in Dechevsky and MacGibbon (2001) which treat also the case of variable lower and/or upper constraints. Another highly relevant previous study which has not been considered in Mammen et al. (2001) is Appendix B in Dechevsky et al. (1999).*
3. *Cosma et al. (2007) developed a multivariate extension of the results in Dechevsky and Penev (1997, 1998), thereby partially filling the gap indicated by McFadden (2000). The results in Cosma et al. (2007) are, however, valid only for a compactly supported density/c.d.f. To the best of our knowledge, a full-scale extension of the results Dechevsky and Penev (1997, 1998) (involving also assessment of the role of the tailweights of a density/c.d.f. with non-compact support) has not been achieved yet.*
4. *The survey article Dechevsky (2007) provided a relatively recent overview (as of the end of 2007) of the state of the art and ongoing research in shape-preserving wavelet-based approximation, giving references to, among others, Anastassiou and Yu (1992); Dechevsky and Penev (1997, 1998); Penev and Dechevsky (1997); Pinheiro and Vidakovic (1997); Dechevsky et al. (1999); Dechevsky and MacGibbon (1999); Dechevsky et al. (2000, 2001); McFadden (2000); Mammen et al. (2001); Cosma et al. (2007), providing also a preliminary announcement (without proofs or details) of some of the main results Dechevsky*

and MacGibbon (2001), and reporting the ongoing research in the forthcoming paper Dechevsky and MacGibbon (2010).

In conclusion, our estimator $\widehat{f}_{+,\varepsilon}$ and its various modifications considered here seem to be ideally suited to preserving optimal estimation rates over a broad variety of smoothness classes, while also obeying order constraints of a quite general kind.

Appendix: Preliminaries and notation

For the relevant facts, definitions and standard notation related to the theory of function spaces, we cite as a general reference Bergh and Löfström (1976).

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

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

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

Quasi-norm (see Bergh and Löfström (1976), Dechevsky and Penev (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 and Löfström (1976).

A – quasi-Banach space, if A is a complete quasi–normed abelian group, a linear space, and the quasi-norm 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 quasi-normed abelian groups) – see Bergh and Löfström (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 – quasi-normed abelian groups: $\|a\|_{A+B} = \inf_{a=\alpha+\beta}\{\|\alpha\|_A + \|\beta\|_B\}$ (see Bergh and Löfström (1976)).

The K-functional: $K(t, a; A, B) = \inf_{a=\alpha+\beta}\big(\|\alpha\|_A + t\|\beta\|_B\big)$, $0 < t < \infty$, plays an important role in this paper. (It should be noted that the K-functional is one possible equivalent quasinorm on $A + B$ (see Bergh and Löfström (1976), Johnen and Scherer (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 and Löfström (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 and Löfström (1976), Johnen and Scherer (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 and Löfström (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 and Löfström (1976), Triebel (1983), Triebel (1992)).

For more details on the properties of B^s_{pq} and F^s_{pq} we refer to Bergh and Löfström (1976), Triebel (1983), Triebel (1992).

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

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

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

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

A very important property of $\omega_k(f;\delta)_{L_p}$ which we use is its equivalence to a K-functional (see Johnen and Scherer (1977)): $\omega_k(f;\delta)_{L_p} \sim K(\delta^k, f, L_p, \dot{W}^k_p)$, $\delta > 0$, 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

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

$L_p(\mathbb{R}^d, \mathcal{L}_\rho(\mathbb{R}^d))$ 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

$$\|g\|_{L_p(\mathcal{L}_\rho)} = \left(\int_{\mathbb{R}^d} \left(E_{F^N} |g(x)|^\rho \right)^{p/\rho} dx \right)^{1/p}.$$

(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 based on the generalized Minkowski inequality, see, e.g., Nikol'skiĭ (1975). (These embeddings are discussed, e.g., in the preliminaries of Dechevsky and Penev (1998).) For this reason here we restrict the 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 results in Section 3 and applications in Section 4 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

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

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

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

Define the approximate identity (cf. Reed and Simon (1975))

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

Clearly,

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

For ways to define Φ in such a way that $\Phi \in C_0^\infty(\mathrm{I\!R}^{\mathrm{d}})$ and Φ is symmetric with respect to the origin, see the Appendix in Dechevsky and MacGibbon (1999, 2009).

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

$$T\vee_\omega = \{f : \mathrm{I\!R} \to \mathrm{I\!R},\ \vee_\omega \mathrm{f} < \infty\}$$

where

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

is the ω-variation of f on $\mathrm{I\!R}$ in the sense of Young (1937). The case $\omega = \omega(t) = t$ corresponds to the usual Jordan variation, while the case $\omega = \omega(t) = t^p$, $1 \le p < \infty$, corresponds to p-variation in the sense of Wiener (1924).

To define the local and averaged moduli of smoothness, let us first consider the case $d = 1$.

For $\varepsilon > 0$, $0 < p \le \infty$, (see Dechevski (1988,a); Dechevsky (2008, 2007a)) consider

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

where

$$\|f\|_{A_{p,\varepsilon}(\mathrm{I\!R})} = \|S(\varepsilon, |f|; \cdot)\|_{L_p(\mathrm{I\!R})},$$
$$S(\varepsilon, f; x) = sup\{f(y) : y \in [x - \varepsilon, x + \varepsilon]\},$$

where S is often called an upper Baire's function.

The local modulus of smoothness is a generalization of $S\left(\frac{\varepsilon}{2}, |f|; x\right)$ (which corresponds to $r = 0$):

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

(see Sendov and Popov (1988); Petrushev and Popov (1987); Dechevski (1988,a); Dechevsky (2008, 2007a)).

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

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

(see Sendov and Popov (1988); Petrushev and Popov (1987); Dechevski (1988,a); Dechevsky (2008, 2007a)).

An equivalent norm in $B^s_{pq}(\mathrm{I\!R})$ can be defined via integral moduli of smoothness

$$\|f\|_{B^s_{pq}(\mathrm{I\!R})} \sim \|f\|_{L_p(\mathrm{I\!R})} + \left[\int_0^{c_0} \left(\xi^{-s} \omega_r(f; \xi)_{L_p(\mathrm{I\!R})}\right)^q \frac{d\xi}{\xi}\right]^{\frac{1}{q}},$$

$0 < s < r$, $1 \le p \le \infty$, where $c_0 : 0 < c_0 \le \infty$ (the constants of equivalence depend on c_0; note that $c_0 = \infty$ is an admissible value). The so-called A-spaces A^s_{pq} have been defined by V. A. Popov analogously to the above definition of B^s_{pq}; the quasi-norm $A^s_{pq}(\mathrm{I\!R})$, $0 < s < r$, $0 < p \le \infty$ is defined by

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

(see, e.g., Dechevski (1988,a); Dechevsky (2008, 2007a)). In Dechevski (1988a), Section 1.4 (see also Dechevski (1988), or Section 4.2 in Dechevsky (2008, 2007a)), it has been proved that

$$A^s_{pq}(\mathrm{I\!R}) \leftrightharpoons \mathrm{B}^{\mathrm{s}}_{\mathrm{pq}}(\mathrm{I\!R}),\ 0 < \mathrm{p} \le \infty,\ 0 < \mathrm{q} \le \infty,\ \mathrm{s} > \frac{1}{\mathrm{p}}.$$

(It can also be proved that this isomorphism continues to hold true for $s = \frac{1}{p}$, $q = \min\{1, p\}$.) As well as 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 $\frac{d}{p}$.

For a general reference on all considered types of moduli of smoothness, see Sendov and Popov (1988); Petrushev and Popov (1987). Additional relevant information can be found in Dechevski (1988,a); Dechevsky et al. (1999); Johnen and Scherer (1977); Dechevsky and MacGibbon (1999, 2009); Dechevsky (2008, 2007a).

In the multivariate case integral, local and averaged moduli of smoothness can all be defined by finite differences, too, but it is somehow more convenient to work with the equivalent K-functionals most of the time. For the integral moduli and their equivalent K-functionals in the multivariate case, see Johnen and Scherer (1977). These K-functionals are between L_p-spaces and the homogeneous Sobolev spaces $\dot{W}_p^k$. The analogous K-functionals for local moduli are just the previous moduli for L_∞ and $\dot{W}_\infty^k$ over local neighbourhoods, rather than the whole domain. As for the averaged moduli, their equivalent K-functionals are in terms of the spaces $A_{p,\varepsilon}$ and $\dot{W}_{p,\varepsilon}^k$ (where the quasinorm in $\dot{W}_{p,\varepsilon}^k$ is obtained by replacing the L_p-quasinorm with the $A_{p,\varepsilon}$-quasinorm. There follow the detailled definitions in the case of local and averaged moduli.

Denote $\bar{B}_d(x;\varepsilon) = \left\{\xi \in \mathrm{I\!R}^{\mathrm{d}}, \|\mathrm{x} - \xi\| \le \varepsilon\right\}, \quad x \in \mathrm{I\!R}^{\mathrm{d}}, \varepsilon > 0$, where $\|\cdot\|$ is an arbitrary , henceforward fixed, norm in $\mathrm{I\!R}^{\mathrm{d}}$. (The definitions, given below, are equivalent for different choices of $\|\cdot\|$. For concreteness, let $\|\cdot\|$ be the usual Hilbert norm $|x| = \sum_{i=1}^d x_i^2$.)

Let $\Omega \subset \mathrm{I\!R}^{\mathrm{d}}$ be open, simply connected set, or the closure of such a set. The local modulus of smoothness of $f : \Omega \to \mathrm{I\!R}$ of order $k \in \mathbb{N}$, with step ε, at $x \in \mathrm{I\!R}^{\mathrm{d}}$ (or, rather, the equivalent K-functional to this modulus) is defined by

$$\omega_k(f, x; \varepsilon) = K\left(\varepsilon^k, f; L_\infty\left(\bar{B}_d(x;\varepsilon) \cap \Omega\right), \dot{W}_\infty^k\left(\bar{B}_d(x;\varepsilon) \cap \Omega\right)\right).$$

The upper Baire function of f, with step ε, at $x \in \mathrm{I\!R}^{\mathrm{d}}$:

$$S(f, x; \varepsilon) = sup\left\{f(\xi) : \xi \in \bar{B}_d(x;\varepsilon) \cap \Omega\right\}.$$

The lower Baire function of f, with step ε, at $x \in \mathrm{I\!R}^{\mathrm{d}}$:

$$I(f, x; \varepsilon) = inf\left\{f(\xi) : \xi \in \bar{B}_d(x;\varepsilon) \cap \Omega\right\}.$$

For $0 < p \le \infty$, $\varepsilon > 0$

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

where

$$\|f\|_{A_{p,\varepsilon}(\Omega)} = \|S(|f|, \cdot; \varepsilon)\|_{L_p(\Omega)}.$$

For $k \in \mathbb{N}$, $0 < p \le \infty$, $\varepsilon > 0$,

$$\dot{W}_{p,\varepsilon}^k(\Omega) = \{(f : \Omega \to \mathrm{I\!R}) : \|\mathrm{f}\|_{\dot{\mathrm{W}}_{\mathrm{p},\varepsilon}^{\mathrm{k}}(\Omega)} < \infty\},$$

where

$$\|f\|_{\dot{W}_{p,\varepsilon}^k(\Omega)} = \sum_{|\nu|=k} \|D^\nu f\|_{A_{p,\varepsilon}(\Omega)},$$

where ν is a multiindex.

The averaged modulus of smoothness (together with its equivalent K-functional) of $f : \Omega \to \mathrm{I\!R}$, of order $k \in \mathbb{N}$, metric L_p, $0 < p \le \infty$, and step $\varepsilon > 0$:

$$\tau_k(f;\varepsilon)_{L_p(\Omega)} := \|\omega_k(f,\cdot;\varepsilon)\|_{L_p(\Omega)} \sim K\left(\varepsilon^k, f; A_{p,\varepsilon}(\Omega), \dot{W}^k_{p,\varepsilon}(\Omega)\right),\ k \in \mathbb{N},$$
$$\tau_0(f;\varepsilon)_{L_p(\Omega)} := \|f\|_{A_{p,\varepsilon}(\Omega)}.$$

The definitions of Besov and A-spaces in terms of the integral with respect to the Haar measure $\frac{dt}{t}$, given above for $d = 1$, extends to any $d \in \mathbb{N}$. For properties of the moduli, as well as for the Marchaud-type inequalities, we refer to Sendov and Popov (1988); Petrushev and Popov (1987); Johnen and Scherer (1977) and Dechevski (1988a); Dechevsky (2008, 2007a), as well as the lemmas in Appendix 0 of Dechevsky and MacGibbon (1999, 2009).

References

R.A. Adams. *Sobolev Spaces*. Academic Press, London, 1975.

G.A. Anastassiou, X.M. Yu. Monotone and probabilistic wavelet approximation. *Stochastic Analysis and Applications*, **10(3)** (1992), 251–264.

J. Bergh, J. Löfström. *Interpolation Spaces. An Introduction.* Grundlehren **223**. Springer, Berlin, 1976.

A. Cosma, O. Scaillet, R. von Sachs. Multivariate wavelet-based shape-preserving estimation for dependent observations. *Bernoulli*, **13(2)** (2007), 301–329.

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

L.T. Dechevski. *Some Applications of the Theory of Function Spaces to Numerical Analysis.* Ph.D. Dissertation. Sofia University, 1989. (In Bulgarian.)

L.T. Dechevsky. Shape preserving wavelet approximation: state of the art and ongoing research. In: F. Truchetet, O. Laligant (Eds.) *Wavelet Applications in Industrial Processing V.* Proceedings of SPIE, **6763** (2007), article 67630X, 1–14.

L.T. Dechevsky. Properties of function spaces generated by the averaged moduli of smoothness. R&D Report 1/2008 (ISSN 1890-923X) Applied Mathematics, Narvik University College, Narvik, Norway, 2008.

L.T. Dechevsky. Properties of function spaces generated by the averaged moduli of smoothness. International Journal of Pure and Applied Mathematics, **41(9)** (2007), 1305–1375.

L.T. Dechevsky, B. MacGibbon. Asymptotically minimax non-parametric function estimation with positivity constraints I. *Les Cahiers du GERAD* (ISSN 0711-2440) G–99–24, Montreal, Canada, 1999.

L.T. Dechevsky, B. MacGibbon. Asymptotically minimax non-parametric function estimation with positivity constraints I. *Statistics*, **43(5)** (2009), 423–442.

L.T. Dechevsky, B. MacGibbon. Asymptotically minimax non-parametric function estimation with positivity constraints II. *Les Cahiers du GERAD* (ISSN 0711-2440) G–2001–34, Montreal, Canada, 2001.

L.T. Dechevsky, B. MacGibbon, S.I. Penev. Numerical methods for asymptotically minimax non-parametric function estimation with positivity constraints I. *Les Cahiers du GERAD* (ISSN 0711-2440) G–2000–33, Montreal, Canada, 2000.

L.T. Dechevsky, B. MacGibbon, S.I. Penev. Numerical methods for asymptotically minimax non-parametric function estimation with positivity constraints I. *Sankhya Series B: The Indian Journal of Statistics, Special issue on Wavelets*, **63(2)** (2001), 149–181.

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

L.T. Dechevsky, S.I. Penev. On shape–preserving probabilistic wavelet approximators. *Stochastic Analysis and Applications*, **15(2)** (1997), 187–215.

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.

L.T. Dechevsky, J.O. Ramsay, S.I. Penev. Penalized wavelet estimation with Besov regularity constraints. *Math. Balkanica (N. S.)*, **13(3-4)** (1999), 257–376.

B. Delyon, A. Juditsky. On minimax wavelet estimators. *Applied and Computational Harmonic Analysis*, **3** (1996), 215–228.

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

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.

C. Huber, B. MacGibbon. Lower bounds for estimating a hazard. In: N. Balakrishnan, C.R. Rao (Eds.). Handbook of Statistics **23**. Elsevier B.V., Amsterdam (2004), 209–226.

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

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.

L. Li. On the minimax optimality of wavelet estimation with censored data. *Journal of Statistical Planning and Inference*, **137** (2007), 1138–1150.

L. Li, B. MacGibbon, C. Valenta. On the optimality of wavelet-based nonparametric regression with censored data. *Journal of Applied Probability and Statistics*, **2** (2008), 243–261.

H.-Y. Liang, V. Mammitzsch, J. Steineback. Nonlinear wavelet density and hazard rate estimation for censored data under dependent observations. *Statistics and Decisions*, **23** (2005), 161–180.

D. McFadden. Economic Choices. *Nobel Lecture in Economics*, 2000.

E. Mammen, J.S. Marron, B.A. Turlach, M.P. Wand. A general projection framework for constrained smoothing. *Statistical Science*, **16(3)** (2001), 232–248.

S.M. Nikol'skiĭ. *Approximation of Functions of Several Variables and Imbedding Theorems*. Grundlehren **205**. Springer, 1975.

S.I. Penev, L.T. Dechevsky. On non–negative wavelet–based density estimators. *Nonparametric Statistics*, **7** (1997), 365–394.

P.P. Petrushev, V.A. Popov. *Rational Approximation of Real Functions*. Cambridge University Press, Cambridge, 1987.

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

M. Reed, B. Simon. *Methods of Modern Mathematical Physics*, Vol. 2. Academic Press, London, 1975.

B. Sendov, V.A. Popov. *The Averaged Moduli of Smoothness: Applications in Numerical Methods and Approximation*. Wiley, New York, 1988.

H. Triebel. *Theory of Function Spaces*. Monographs in Mathematics **78**. Birkhäuser, Boston, 1983.

H. Triebel. *Theory of Function Spaces II*. Monographs in Mathematics **84**. Birkhäuser, Boston, 1992.

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

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.