ESAIM: PS 18 (2014) 1–41
DOI: 10.1051/ps/2012024

ESAIM: Probability and Statistics
www.esaim-ps.org

NONPARAMETRIC REGRESSION ESTIMATION BASED ON SPATIALLY INHOMOGENEOUS DATA: MINIMAX GLOBAL CONVERGENCE RATES AND ADAPTIVITY

Anestis Antoniadis[1], Marianna Pensky[2] and Theofanis Sapatinas[3]

Abstract. We consider the nonparametric regression estimation problem of recovering an unknown response function f on the basis of spatially inhomogeneous data when the design points follow a known density g with a finite number of well-separated zeros. In particular, we consider two different cases: when g has zeros of a polynomial order and when g has zeros of an exponential order. These two cases correspond to moderate and severe data losses, respectively. We obtain asymptotic (as the sample size increases) minimax lower bounds for the L^2-risk when f is assumed to belong to a Besov ball, and construct adaptive wavelet thresholding estimators of f that are asymptotically optimal (in the minimax sense) or near-optimal within a logarithmic factor (in the case of a zero of a polynomial order), over a wide range of Besov balls. The spatially inhomogeneous ill-posed problem that we investigate is inherently more difficult than spatially homogeneous ill-posed problems like, *e.g.*, deconvolution. In particular, due to spatial irregularity, assessment of asymptotic minimax global convergence rates is a much harder task than the derivation of asymptotic minimax local convergence rates studied recently in the literature. Furthermore, the resulting estimators exhibit very different behavior and asymptotic minimax global convergence rates in comparison with the solution of spatially homogeneous ill-posed problems. For example, unlike in the deconvolution problem, the asymptotic minimax global convergence rates are greatly influenced not only by the extent of data loss but also by the degree of spatial homogeneity of f. Specifically, even if $1/g$ is non-integrable, one can recover f as well as in the case of an equispaced design (in terms of asymptotic minimax global convergence rates) when it is homogeneous enough since the estimator is "borrowing strength" in the areas where f is adequately sampled.

Mathematics Subject Classification. 62G08, 62G05, 62G20.

Received March 27, 2012. Revised August 25, 2012.

1. Introduction

Applicability of majority of techniques for estimation in the nonparametric regression model rests on the assumption that data is equispaced and complete. These assumptions were mainly adopted by signal processing

Keywords and phrases. Adaptivity, Besov spaces, inhomogeneous data, minimax estimation, nonparametric regression, thresholding, wavelet estimation.

[1] Laboratoire Jean Kuntzmann, Universite Joseph Fourier, 38041 Grenoble Cedex 9, France. Anestis.Antoniadis@imag.fr

[2] Department of Mathematics, University of Central Florida, Orlando, 32816-1364, USA. Marianna.Pensky@ucf.edu

[3] Department of Mathematics and Statistics, University of Cyprus, P.O. Box 20537, CY 1678 Nicosia, Cyprus. fanis@ucy.ac.cy

Article published by EDP Sciences

© EDP Sciences, SMAI 2013

https://doi.org/10.1051/ps/2012024 Published online by Cambridge University Press

community where the signal is assumed to be recorded at equal intervals in time. However, in reality, due to unexpected losses of data or limitations of data sampling techniques, data may fail to be equispaced and complete. To this end, we consider the problem of recovering an unknown response function $f \in L^2([0,1])$ on the basis of irregularly spaced observations, *i.e.*, when one observes y_i governed by

$$y_i = f(x_i) + \sigma\,\xi_i, \quad i = 1, 2, \ldots, n, \tag{1.1}$$

where $x_i \in [0,1]$, $i = 1, 2, \ldots, n$, are fixed (non-equidistant) or random points, ξ_i, $i = 1, 2, \ldots, n$, are independent standard Gaussian random variables and $\sigma^2 > 0$ (the noise level) is assumed to be known and finite. Model (1.1) can be viewed as a problem of recovering a signal when part of data is lost (*e.g.*, in cell phone use) or unavailable (*e.g.*, in military applications). Model (1.1) is also intimately connected to the problem of missing data since points x_i, $i = 1, 2, \ldots, n$, can be viewed as the remainder of N equidistant points j/N, $j = 1, 2, \ldots, N$, after observations at $(N - n)$ points have been lost. However, there is a great advantage in treating the missing data problem as a particular case of a nonparametric regression problem: with the last two decades seeing tremendous advancement in the field of nonparametric statistics, a nonparametric regression approach to incomplete data brings along all the modern tools in this field such as asymptotic minimax convergence rates, Besov spaces, wavelets and adaptive estimators.

The problem of estimating an unknown response function in the context of wavelet thresholding in the nonparametric regression setting with irregular design has been now addressed by many authors, see, *e.g.*, Hall and Turlach [15], Antoniadis and Pham [2], Cai and Brown [5], Sardy *et al.* [35], Kovac and Silverman [24], Pensky and Vidakovic [34], Brown *et al.* [4], Zhang *et al.* [39], Kohler [22] and Amato *et al.* [1]. Several tools were suggested for attacking the problem; here, we shall review only few of them. For instance, the procedure of Kovac and Silverman [24] relies upon a linear interpolation transformation R to the observed data vector $y = (y_1, y_2, \ldots, y_n)$ that maps it to a new vector of size 2^J ($2^{J-1} < n \leq 2^J$), corresponding to a new design with equispaced points. After the transformation, the new vector is multivariate normal with mean Rf and covariance matrix which is assumed to have a finite bandwidth, so that the computational complexity of their algorithm is of order n. Cai and Brown [5] attacked the problem by using multiresolution analysis, projection and wavelet nonlinear thresholding while Sardy *et al.* [35] applied an isometric method. Pensky and Vidakovic [34] estimated the conditional expectation $\mathbb{E}(Y|X)$ directly by constructing its wavelet expansion, while Amato *et al.* [1] applied a reproducing kernel Hilbert space (RKHS) approach in the spirit of Wahba [38]. However, until very recently, all studies have been carried out under the assumption that the nonequispaced design still possesses some regularity, namely, the density function g of the design points x_i, $i = 1, 2, \ldots, n$, is uniformly bounded from below, *i.e.*, $\inf_{x\in[0,1]} g(x) \geq c$ for some constant $c > 0$. In this case, asymptotically, model (1.1) is equivalent to the case of the standard (equispaced) nonparametric regression model, as long as the design density function g is known (see, *e.g.*, Brown *et al.* [4]).

Recently, an attempt has been made of more advanced investigations of the problem. Kerkyacharian and Picard [21] introduced warped wavelets to construct estimators of the unknown response function f under model (1.1) when the design density function g has zeros of polynomial order. They, however, measured the error of their suggested estimator in the warped Besov spaces which is, practically, equivalent to measuring the error of the estimator at the design points only. For this reason, the derived estimators posses the usual asymptotic (as the sample size increases) minimax global convergence rates which do not depend on the order of the zeros of the design density function g. This line of investigation was continued by Chesneau [6] who constructed asymptotic minimax lower bounds over a wide range of Besov balls, under the assumption that the design density function g is known and that $1/g$ is integrable, and, furthermore, suggested adaptive wavelet thresholding estimators for the unknown response function f. However, in Kerkyacharian and Picard [21] and Chesneau [6], the assumptions on the design density function g are restrictive enough so that the asymptotic minimax global convergence rates of any estimator coincide with the asymptotic minimax global convergence rates under the assumption that g is bounded from below, *i.e.*, the corresponding nonparametric estimation problem is a *well-posed* problem.

https://doi.org/10.1051/ps/2012024 Published online by Cambridge University Press

Gaïffas [9,11] was the first author who considered nonparametric regression estimation on the basis of spatially inhomogeneous data as an *ill-posed* problem. In particular, he constructed pointwise adaptive estimators of f on the basis of local polynomials when $1/g$ is non-integrable and showed that the asymptotic minimax local convergence rates of the suggested estimators are slower than in the case when g is bounded from below, hence, demonstrating that the aforementioned estimation problem is an ill-posed problem. Since his techniques are intended for local reconstruction and depend on cross-validation at each point, they become too involved when one tries to adapt them to the whole domain of f. Furthermore, Gaïffas [10, 12] studied asymptotic minimax uniform convergence rates. However, these rates are expressed in a very complex form which is very hard to obtain for f belonging to standard functional classes (see Rem. 6.4). Note also that some of his results were recently extended to the multivariate case by Guillou and Klutchnikoff [14].

Our objective is to study how the zeros of the design density function g affect the asymptotic minimax global convergence rates of f in model (1.1), and to construct adaptive wavelet thresholding estimators of f which attain these rates, over a wide range of Besov balls. As we show below (see Rem. 2.2), assessing asymptotic minimax global convergence rates is a much harder task than assessing asymptotic minimax local convergence rates. Model (1.1) can be viewed as a spatially inhomogeneous ill-posed problem which is inherently more difficult than spatially homogeneous ill-posed problems like, *e.g.*, deconvolution, especially in the case when the unknown response function is spatially homogeneous. To the best of our knowledge, so far, there are no results for asymptotic minimax global convergence rates in the case of spatially inhomogeneous ill-posed problems when its solution is spatially homogeneous since this problem is usually avoided by restricting attention to the case when the estimated function is spatially inhomogeneous, or, at most, belongs to a Sobolev ball (see, *e.g.*, Hoffmann and Reiss [17]).

In what follows, we address these issues. In particular, we mainly consider two different cases: when g has zeros of a polynomial order and when g has zeros of an exponential order. We obtain asymptotic (as the sample size increases) minimax lower bounds for the L^2-risk when f is assumed to belong to a Besov ball, and construct adaptive wavelet thresholding estimators of f that are asymptotically optimal (in the minimax sense) or near-optimal within a logarithmic factor (in the case of a zero of a polynomial order), over a wide range of Besov balls. Due to spatial irregularity, the suggested estimators exhibit very different behavior and asymptotic minimax global convergence rates in comparison with the solution of spatially homogeneous ill-posed problems (see Rem. 3.2). Specifically, even if $1/g$ is non-integrable, one can recover f as well as in the case of an equispaced design (in terms of asymptotic minimax global convergence rates) when the function is homogeneous enough since the estimator is "borrowing strength" in the areas where f is adequately sampled. These features lead to a different structure of estimators of f described in Section 4. The complementary case when $1/g$ is integrable has been partially handled by Chesneau [6] who showed that the problem is well-posed (*i.e.*, data loss does not affect the asymptotic minimax global convergence rates) when f is spatially homogeneous. A complete study of the case when $1/g$ is integrable is considered in Section 7. In depth discussion of the differences of the spatial features in the spatially inhomogeneous ill-posed problem considered in this paper is presented in Section 8.

To address spatial irregularity of the design in the case when the design density function g has a zero of a polynomial order, we develop a novel, two-stage, adaptive wavelet thresholding estimator. This estimator consists of a linear part which is taken at a resolution level that is chosen adaptively by Lepski's method and which estimates f in the neighborhood of the zero of g. We refer to this as the *zero-affected* part of the estimator. The second part is nonlinear (thresholding) and is used outside the immediate neighborhood of the zero of g. We refer to this as the *zero-free* part of the estimator. The lowest resolution level of the nonlinear part coincides with the resolution level of the linear part of the estimator, so that the sum of the two parts represents f correctly. If $1/g$ is integrable, then the zero-affected portion of the estimator vanishes and f can be estimated by an adaptive wavelet thresholding estimator in the spirit of Chesneau [6].

We limit our attention only to the L^2-risk since the consideration of a wider class of risk functions will make the exposition of the present work even longer; all results, however, obtained can be extended to the case of L^u-risks, $1 \leq u < \infty$. Moreover, we consider only the univariate case, leaving generalizations to the multivariate case for future investigation.

https://doi.org/10.1051/ps/2012024 Published online by Cambridge University Press

The rest of the paper is organized as follows. Section 2 discusses the formulation of the nonparametric regression estimation problem of the unknown response function f on the basis of spatially inhomogeneous data, in particular when the design density function g has either a zero of a polynomial order or a zero of an exponential order. Section 3 contains the asymptotic minimax lower bounds for the L^2-risk when f is assumed to belong to a Besov ball. Section 4 talks about estimation strategies when $1/g$ is non-integrable, in particular, about partitioning f and its estimator into the zero-affected and zero-free parts. Section 5 elaborates on the estimation of the zero-affected and the zero-free parts, and is followed by Section 6 which discusses the choice of adaptive resolution level and derives the asymptotic minimax upper bounds for the L^2-risk in the case when $1/g$ is non-integrable. Section 7 studies the complementary case when g has zeros but $1/g$ is still integrable. Section 8 concludes the paper with a discussion. Finally, Section 9 contains the proofs of the statements in the earlier sections.

2. Formulation of the problem

Consider the nonparametric regression model (1.1). Since the noise level is assumed to be known and finite, without loss of generality, we set $\sigma = 1$. Therefore, from now onwards, we work with observations y_i governed by equation (1.1), where $f \in L^2([0,1])$ is the unknown response function to be recovered, $x_i \in [0,1]$, $i = 1, 2, \ldots, n$, are random design points with the underlying density function g, and ξ_i, $i = 1, 2, \ldots, n$, are independent standard Gaussian random variables, independent of x_i, $i = 1, 2, \ldots, n$. Furthermore, we assume that the design density function g is known and has a finite number of zeros which are well-separated, *i.e.*, there exist a constant $\delta > 0$ such that the distance between two consecutive zeros is at least δ. The last assumption is motivated by the following considerations. If g vanishes on an interval $[a, b] \subset [0,1]$, $a < b$, then consistent estimation of $f(x)$, for $x \in [a, b]$, is impossible. Also, g has an infinite number of zeros on $[0,1]$ only in the case when g is highly oscillatory, which is not a very likely scenario. Finally, the assumption that g has low values on a part of its domain but is still separated from zero is not an interesting case to consider, since the lower bound on g will appear in the constant of the well-known expressions for the asymptotic minimax convergence rates (see, *e.g.*, Tsybakov [37], Chapts. 1–2).

Note that the above assumptions are not restrictive. If the noise level σ is unknown, it can be easily estimated with parametric precision using observations in the region where g is separated from zero. The assumption that the design points x_i, $i = 1, 2, \ldots, n$, are random is not confining either. In fact, with small modifications of the theory below, one can consider fixed points $0 \le x_1 < x_2 < \cdots < x_n \le 1$, generated by an increasing and continuously differentiable function G such that $G(0) = 0$, $G(1) = 1$ and $G(x_i) = i/n$, $i = 1, 2, \ldots, n$. Then, the function G plays the role of a "surrogate" distribution function with density function g; the design points x_i, $i = 1, 2, \ldots, n$, can be then obtained as $x_i = G^{-1}(i/n)$, $i = 1, 2, \ldots, n$.

Moreover, since the design density function g is assumed to be known with a finite number of zeros that are also well-separated, one can partition the interval $[0,1]$ into subintervals in such a manner that each subinterval contains only one zero of g. For this reason, in what follows, without loss of generality, we assume that g has only one zero $x_0 \in [0,1]$, and that the following condition holds.

Assumption A. Let the design density g be a continuous function on the interval $[0,1]$ with $g(x_0) = 0$, $x_0 \in [0,1]$. Then, there exists constants $\alpha \in \mathbb{R}$, $b \ge 0$ ($\alpha > 0$ if $b = 0$), $\beta > 0$ and $C_g > 0$ such that, for any x, with $x, x + x_0 \in [0,1]$,

$$\lim_{x \to 0} g(x_0 + x)|x|^{-\alpha} \exp(b|x|^{-\beta}) = C_g. \tag{2.1}$$

If $b = 0$, we shall say that x_0 is a zero of *polynomial order*. If $b > 0$, we shall say that x_0 is a zero of *exponential order*. Observe that (2.1) implies that there exist some constants $0 < C_{g1} < C_g < C_{g2}$ such that for any x, with $x, x + x_0 \in [0,1]$ and $x_0 \in [0,1]$, one has

$$g(x_0 + x) \le C_{g2}|x|^{\alpha} \exp(-b|x|^{-\beta}), \quad g(x_0 + x) \ge C_{g1}|x|^{\alpha} \exp(-b|x|^{-\beta}). \tag{2.2}$$

https://doi.org/10.1051/ps/2012024 Published online by Cambridge University Press

Note that the two cases in Assumption A correspond to the situations of moderate ($b = 0$) and severe ($b > 0$) data losses, respectively. Chesneau [6] showed that in the case of a moderate loss ($b = 0$) with $0 < \alpha < 1$ (*i.e.*, $1/g$ is integrable), and for a response function f that is spatially homogeneous, f can be estimated with the same asymptotic minimax global convergence rates as in the case of $b = 0$ with $\alpha = 0$ (*i.e.*, g is uniformly bounded from below); hence, in this case, the nonparametric regression estimation problem turns out to be a *well-posed* problem.

Therefore, we shall be mainly interested only in the complementary situation when $1/g$ is non-integrable: (i) moderate losses (*i.e.*, $b = 0$) with $\alpha \geq 1$ and (ii) severe losses (*i.e.*, $b > 0$) with $\alpha \in \mathbb{R}$ and $\beta > 0$. As we shall see below, usually in those cases, the asymptotically optimal (in the minimax sense) estimation procedures yield estimators with lower convergence rates than in the case of equispaced observations, so that the corresponding nonparametric regression estimation problem under model (1.1) becomes *ill-posed* (see Rem. 2.1), with the degree of ill-posedeness growing as $\alpha \geq 1$ increases when $b = 0$ or as $\beta > 0$ increases when $b > 0$.

In what follows, we use the symbol C for a generic positive constant, independent of the sample size n, which may take different values at different places.

Remark 2.1 (Risk functions and design)**.** As indicated above, we shall measure the precision of any estimator $\hat{f}_n$ of f by its L^2-risk, *i.e.*,

$$\Delta(\hat{f}_n) = \mathbb{E}\|\hat{f}_n - f\|^2.$$

If the design points $x_i \in [0,1]$, $i = 1, 2 \ldots, n$, in model (1.1) are treated as fixed (*i.e.*, non-random), then, the above risk, evaluated at the equispaced design $\{i/n\}$, $i = 1, 2, \ldots, n$, corresponds to

$$\Delta^d(\hat{f}_n) = \frac{1}{n}\sum_{i=1}^{n} \mathbb{E}[\hat{f}_n(i/n) - f(i/n)]^2,$$

and leads to an *ill-posed* nonparametric regression estimation problem. However, it is instructive to note that if one measures the precision of an estimator $\hat{f}_n$ at the design points $x_i \in [0,1]$, $i = 1, 2, \ldots, n$, only, by calculating

$$\Delta^d_{\text{fixed}}(\hat{f}_n, x_i) = \frac{1}{n}\sum_{l=1}^{n} \mathbb{E}[\hat{f}_n(x_i) - f(x_i)]^2,$$

as it was done in, *e.g.*, Amato *et al.* [1], then the problem ceases to be ill-posed. Moreover, in this case, no special treatment is necessary to account for the irregular design. To confirm that, note that model (1.1) can be re-written as

$$y_i = F(i/n) + \xi_i, \quad i = 1, 2, \ldots, n, \tag{2.3}$$

where $F(x) = f(G^{-1}(x))$, $x \in [0,1]$, and G is the "surrogate" distribution function mentioned earlier. Construct now an estimator $\hat{F}_n$ of F using, *e.g.*, any of the standard wavelet thresholding techniques, and set $\hat{f}_n(x) = \hat{F}_n(G(x))$, $x \in [0,1]$. Then,

$$\hat{F}_n(x) = \hat{f}_n(G^{-1}(x)), \quad x \in [0,1],$$

and $\Delta^d_{\text{fixed}}(\hat{f}_n, x_i)$ takes the form

$$\Delta^d_{\text{fixed}}(\hat{f}_n, x_i) = \frac{1}{n}\sum_{i=1}^{n} \mathbb{E}[\hat{F}_n(i/n) - F(i/n)]^2.$$

Therefore, if the observed data vector $y = (y_1, y_2, \ldots, y_n)$ is treated as if the measurements were carried out at equispaced design points, then, by using, *e.g.*, available wavelet denoising algorithms, the resulting estimator $\hat{F}_n$ of function F will be adaptive and it will lead to the smallest possible risk $\Delta^d_{\text{fixed}}(\hat{f}_n, x_i)$. This phenomenon was noticed earlier by Cai and Brown [5], Sardy *et al.* [35] and Brown *et al.* [4].

https://doi.org/10.1051/ps/2012024 Published online by Cambridge University Press

Remark 2.2 (Local *versus* global convergence rates)**.** The nonparametric regression estimation problem of recovering f globally, on the basis of spatially inhomogeneous data, is a much more difficult task than the corresponding problem of estimating f locally, say at a given point a. Indeed, if G, the distribution function associated with the design density function g, is known, then $F(G(a)) = f(a)$ and, hence, one can estimate F at the point $G(a)$ instead of estimating f at the point a, where $F(x) = f(G^{-1}(x))$, $x \in [0,1]$, and F is equispaced sampled, as in (2.3). Hence, local estimation can be reduced to a well-addressed pointwise regression estimation problem. If $g(a) \neq 0$, then the problem is well-posed and has been extensively studied before. If, instead, $a = x_0$ is a zero of g, then one can deduce asymptotic minimax pointwise convergence rates directly from considerations of Remark 2.1 and straightforward calculus. Let, for simplicity, $x_0 = 0$ and $g(x) = (\alpha+1)x^\alpha$, so that $G(x) = x^{\alpha+1}$ and $G^{-1}(x) = x^{1/(\alpha+1)}$, $x \in [0,1]$. Let f satisfy a Hölder condition of order s at x_0, *i.e.*, $|f(x) - f(x_0)| \leq C|x - x_0|^s$. Then, since $x_0 = 0$, $F(x) = f(G^{-1}(x))$, $x \in [0,1]$, satisfies a Hölder condition of order $s' = s/(\alpha+1)$ at 0, *i.e.*, for $x_0 = 0$,

$$|F(x) - F(x_0)| = |f(G^{-1}(x)) - f(G^{-1}(x_0))| \leq C|G^{-1}(x) - G^{-1}(x_0)|^s = C|x - x_0|^{s/(\alpha+1)}.$$

Since, for $x_0 = 0$, $f(x_0) = F(0)$, one can set $\hat{f}(x_0) = \hat{F}(0)$ and obtain asymptotic minimax pointwise convergence rates for $\hat{f}(x_0)$, on noting that

$$\mathbb{E}\|\hat{f}(x_0) - f(x_0)\|^2 = \mathbb{E}\|\hat{F}(0) - F(0)\|^2 \leq C\, n^{-\frac{2s'}{2s'+1}} = O\left(n^{-\frac{2s}{2s+\alpha+1}}\right),$$

which coincides with the asymptotic minimax pointwise convergence rates obtained by Gaiffas [9]. The whole argument here rests on the fact that $f(x_0) = F(G(x_0))$, $x_0 \in [0,1]$, so one can estimate F at $G(x_0)$ instead of estimating f at the x_0. This, however, cannot be accomplished when a global estimation procedure is required since, in such a case, a Taylor expansion is needed, that can be applied only locally.

3. Minimax lower bounds for the L^2-risk over Besov balls

Before constructing an adaptive estimator of the unknown response function f under model (1.1), we first derive the asymptotic minimax lower bounds for the L^2-risk over a wide range of Besov balls.

Among the various characterizations of Besov spaces $B^s_{p,q}$ in terms of wavelet bases, we recall that for an r-regular multiresolution analysis (see, *e.g.*, Meyer, [31], Chapt. 2, pp. 21–25), with $0 < s < r$, and for a Besov ball $B^s_{p,q}(A)$ defined as

$$B^s_{p,q}(A) = \{f \in L^p([0,1]) :\ f \in B^s_{p,q},\ \|f\|_{B^s_{p,q}} \leq A\},$$

of radius $A > 0$ with $1 \leq p, q \leq \infty$, one has, with $s' = s + 1/2 - 1/p$,

$$B^s_{p,q}(A) = \left\{ f \in L^p([0,1]) : \left(\sum_{k=0}^{2^m-1} |a_{mk}|^p \right)^{1/p} + \left(\sum_{j=m}^{\infty} 2^{js'q} \left(\sum_{k=0}^{2^j-1} |b_{jk}|^p \right)^{q/p} \right)^{1/q} \leq A \right\}, \tag{3.1}$$

with respective sum(s) replaced by maximum if $p = \infty$ and/or $q = \infty$, where $s' = s + 1/2 - 1/p$ (see, *e.g.*, Johnstone *et al.* [20]). We study below the L^2-risk over Besov balls $B^s_{p,q}(A)$ defined as

$$R_n(B^s_{p,q}(A)) = \inf_{\tilde{f}_n} \sup_{f \in B^s_{p,q}(A)} \mathbb{E}\|\tilde{f}_n - f\|^2,$$

where $\|h\|$ is the L^2-norm of a function h defined on $[0,1]$, and the infimum is taken over all possible square-integrable estimators (*i.e.*, measurable functions) $\tilde{f}_n$ of f based on observations y_i from model (1.1).

The following statement provides the asymptotic minimax lower bounds for the L^2-risk.

https://doi.org/10.1051/ps/2012024 Published online by Cambridge University Press

Theorem 3.1. *Let $1 \leq p, q \leq \infty$ and $\max(1/p, 1/2) \leq s < r$, and let Assumption A (with $\alpha > 0$ if $b = 0$, and $\alpha \in \mathbb{R}$ and $\beta > 0$ if $b > 0$) hold. Then, as $n \to \infty$,*

$$R_n(B^s_{p,q}(A)) \geq \begin{cases} C\, n^{-\frac{2s}{2s+1}} & \text{if } \ b = 0,\ \alpha s < s', \\ C\, n^{-\frac{2s'}{2s'+\alpha}} & \text{if } \ b = 0,\ \alpha s \geq s', \\ C\, (\ln n)^{-\frac{2s'}{\beta}} & \text{if } \ b > 0. \end{cases} \tag{3.2}$$

Note that the asymptotic minimax lower bound for the L^2-risk in the first part of (3.2) is obtained by the arguments in Theorem 3.1 of Chesneau [6].

Remark 3.2 (Global convergence rates)**.** As we shall show below, the asymptotic minimax lower bounds for the L^2-risk obtained in Theorem 3.1 are attainable for $b > 0$ and are attainable up to a logarithmic factor for $b = 0$. If $\alpha s = s'$, the asymptotic minimax global convergence rates in the first and second parts of (3.2) coincide. Hence, whenever $\alpha s \leq s'$, the aforementioned nonparametric regression estimation problem is not ill-posed but well-posed, in the sense that the asymptotic minimax global convergence rates are the same as in the case of an equispaced design. For $\alpha \geq 1$, this relation can take place only if $2 \leq p \leq \infty$, *i.e.*, when the function is spatially homogeneous. In particular, $\alpha s \leq s'$ holds true for any α such that $1 \leq \alpha \leq 1+(1/2-1/p)/s$, *i.e.*, when f is very spatially homogeneous (p is large, in particular, when $p > 2/(1-(\alpha-1)s)$ provided that $1 < \alpha < 1 + 1/s$), so that even a relatively severe data loss does not lead to the reduction of asymptotic minimax global convergence rates. If $0 < \alpha < 1$, then the considered nonparametric regression estimation problem is always well-posed whenever f is spatially homogeneous ($p \geq 2$) and also when f is spatially inhomogeneous ($1 \leq p < 2$) and $0 < \alpha < 1 - (1/p - 1/2)/s$. Therefore, even if f is spatially inhomogeneous, the aforementioned nonparametric regression estimation problem is well-posed whenever data loss is very limited ($0 < \alpha < 1 - (1/p - 1/2)/s$).

4. Estimation strategies when $1/g$ is non-integrable

We consider a scaling function φ^* and a mother wavelet ψ^* that generate an orthonormal wavelet basis in $L^2(\mathbb{R})$, as those obtained from, *e.g.*, an r-regular multiresolution analysis of $L^2(\mathbb{R})$, for some $r > 0$. We shall also assume that φ^* and ψ^* are both compactly supported, with integer bounds on their supports so that, for some $L_{\varphi^*},\ U_{\varphi^*},\ L_{\psi^*},\ U_{\psi^*} \in \mathbb{Z}$, with $L_{\varphi^*} < U_{\varphi^*}$, $L_{\psi^*} < U_{\psi^*}$,

$$\mathrm{supp}(\varphi^*) = [L_{\varphi^*}, U_{\varphi^*}], \quad \mathrm{supp}(\psi^*) = [L_{\psi^*}, U_{\psi^*}], \quad L_{\varphi^*} \leq 0,\ U_{\varphi^*} \geq 0,\ U_{\varphi^*} - L_{\varphi^*} \geq 4.$$

(For instance, the Daubechies or Symmlets scaling functions φ^* and mother wavelets ψ^*, with filter number (number of vanishing moments) $N \geq 3$, satisfy (4.2) with $L_{\varphi^*} = 0$, $U_{\varphi^*} = 2N - 1$, $L_{\psi^*} = 1 - N$ and $U_{\psi^*} = N$, see, *e.g.*, Mallat [30], Sect. 7.2).

We then obtain a periodized version of the wavelet basis on the unit interval, *i.e.*, for $j \geq 0$ and $k = 0, 1, \ldots, 2^j - 1$, as

$$\varphi_{jk}(x) = \sum_{i \in \mathbb{Z}} 2^{j/2} \varphi^*(2^j(x+i) - k), \quad \psi_{jk}(x) = \sum_{i \in \mathbb{Z}} 2^{j/2} \psi^*(2^j(x+i) - k), \quad x \in [0,1],$$

so that, for any $m \geq 0$, the set

$$\{\varphi_{mk},\ \psi_{jk} :\ j \geq m,\ k = 0, 1, \ldots, 2^j - 1\},$$

where

$$\varphi_{mk}(x) = 2^{m/2} \varphi(2^m x - k), \quad \psi_{jk}(x) = 2^{j/2} \psi(2^j x - k), \quad x \in [0,1],$$

forms an orthonormal wavelet basis for $L^2([0,1])$ (see, *e.g.*, Mallat [30], Thm. 7.16). Hence, for any $m \geq 0$, any $f \in L^2([0,1])$, can be expanded as

$$f(x) = \sum_{k=0}^{2^m - 1} a_{mk} \varphi_{mk}(x) + \sum_{j=m}^{\infty} \sum_{k=0}^{2^j - 1} b_{jk} \psi_{jk}(x), \quad x \in [0,1], \tag{4.1}$$

https://doi.org/10.1051/ps/2012024 Published online by Cambridge University Press

where

$$a_{mk} = \int_0^1 f(x)\varphi_{mk}(x)\,\mathrm{d}x, \quad k = 0, 1, \ldots, 2^m - 1,$$

$$b_{jk} = \int_0^1 f(x)\psi_{jk}(x)\,\mathrm{d}x, \quad j \geq m, \quad k = 0, 1, \ldots, 2^j - 1.$$

Denote by $L_\varphi, U_\varphi, L_\psi$ and U_ψ the support bounds of the periodic scaling function φ and mother wavelet ψ. Note that the supports of φ^*_{mk} and φ_{mk} coincide if and only if $2^m > U_{\varphi^*} - L_{\varphi^*}$, and, similarly, the supports of ψ^*_{jk} and ψ_{jk} coincide if and only if $2^m > U_{\psi^*} - L_{\psi^*}$. Choose the lowest resolution level m_1 such that $2^{m_1} > \max(U_{\varphi^*} - L_{\varphi^*}, U_{\psi^*} - L_{\psi^*})$, so that supports of periodic and non-periodic wavelets coincide. In this case, we obtain that

$$L_{\varphi^*} = L_\varphi,\ U_{\varphi^*} = U_\varphi,\ L_{\psi^*} = L_\psi,\ U_{\psi^*} = U_\psi,\ L_\varphi \leq 0,\ U_\varphi \geq 0,\ U_\varphi - L_\varphi \geq 4. \tag{4.2}$$

For any integer $l \geq 1$, denote $k_{0l} = 2^l x_0$. (Note that k_{0l} is not necessarily a rational quantity and can take any value). At each resolution level, we partition the set of all indices into the indices which are *zero-affected* and *zero–free*. In particular, let K^φ_{0m} and K^ψ_{0j} be the sets such that, for any integer $m \geq m_1$ and $j = m, m+1, \ldots$,

$$K^\varphi_{0m} = \{k : 0 \leq k \leq 2^m - 1, L_\varphi - 1 < k_{0m} - k < U_\varphi + 1\},$$
$$K^\psi_{0j} = \{k : 0 \leq k \leq 2^j - 1, L_\psi - 1 < k_{0j} - k < U_\psi + 1\}$$

and let

$$K^\varphi_{0mc} = \{k : 0 \leq k \leq 2^m - 1, k \notin K^\varphi_{0m}\}, \quad K^\psi_{0jc} = \left\{k : 0 \leq k \leq 2^m - 1, k \notin K^\psi_{0j}\right\}.$$

Simple calculations yield that $k \in K^\varphi_{0mc}$ and $k \in K^\psi_{0jc}$ imply that $x_0 \notin \operatorname{supp} \varphi_{mk}$ and $x_0 \notin \operatorname{supp} \psi_{jk}$, respectively, so that the sets K^φ_{0mc} and K^ψ_{0jc} are zero–free while the sets K^φ_{0m} and K^ψ_{0j} are zero-affected.

With the above notation it is easy to see that, for any $m \geq m_1$ and $j = m, m+1, \ldots$, f can be partitioned as the sum of zero-affected and zero–free parts, *i.e.*,

$$f(x) = f_{0,m}(x) + f_{c,m}(x), \quad x \in [0, 1],$$

where

$$f_{0,m}(x) = \sum_{k \in K^\varphi_{0m}} a_{mk}\varphi_{mk}(x) + \sum_{j=m}^{\infty} \sum_{k \in K^\psi_{0j}} b_{jk}\psi_{jk}(x), \quad x \in [0, 1], \tag{4.3}$$

$$f_{c,m}(x) = \sum_{k \in K^\varphi_{0mc}} a_{mk}\varphi_{mk}(x) + \sum_{j=m}^{\infty} \sum_{k \in K^\psi_{0jc}} b_{jk}\psi_{jk}(x), \quad x \in [0, 1]. \tag{4.4}$$

We then construct estimators $\hat{f}_{0,m}$ and $\hat{f}_{c,m}$ of $f_{0,m}$ and $f_{c,m}$, respectively, and estimate f by

$$\hat{f}_m(x) = \hat{f}_{0,m}(x) + \hat{f}_{c,m}(x), \quad x \in [0, 1]. \tag{4.5}$$

(We emphasize the unusual feature in the construction of $\hat{f}_m$: as we shall see below, $\hat{f}_{0,m}$ is a linear wavelet estimator while $\hat{f}_{c,m}$ is a nonlinear (thresholding) wavelet estimator with the lowest resolution level m determined by the linear part).

By observing that, for any function $u \in L^2[0, 1]$, we have

$$\int_0^1 u(x)f(x)\mathrm{d}x = \mathbb{E}\left(\frac{f(X)u(X)}{g(X)}\right),$$

https://doi.org/10.1051/ps/2012024 Published online by Cambridge University Press

when the random variable $X \sim g$, and setting, for any $m \geq m_1$ and $j = m, m+1, \ldots$, $u(x) = \varphi_{mk}(x)$ and $u(x) = \psi_{jk}(x)$, $x \in [0,1]$, in turn, similarly to (3.3) in Chesneau [6], we estimate a_{mk}, $k \in K^{\varphi}_{0mc}$, and b_{jk}, $k \in K^{\psi}_{0jc}$, respectively, by

$$\hat{a}_{mk} = \frac{1}{n}\sum_{i=1}^{n} \frac{\varphi_{mk}(x_i) y_i}{g(x_i)}, \quad k \in K^{\varphi}_{0mc}, \qquad \tilde{b}_{jk} = \frac{1}{n}\sum_{i=1}^{n} \frac{\psi_{jk}(x_i) y_i}{g(x_i)}, \quad k \in K^{\psi}_{0jc}. \tag{4.6}$$

Hence, we can construct an estimator $\hat{f}_{c,m}$ of $f_{c,m}$ by estimating a_{mk}, $k \in K^{\varphi}_{0mc}$, and b_{jk}, $k \in K^{\psi}_{0jc}$, by $\hat{a}_{mk}$, $k \in K^{\varphi}_{0mc}$, and $\tilde{b}_{jk}$, $k \in K^{\psi}_{0jc}$, respectively, given in (4.6), along with a thresholding step (see below).

Note that since $1/g$ is non-integrable, the estimators given in (4.6) would have infinite variances if $k \in K^{\varphi}_{0m}$ or $k \in K^{\psi}_{0j}$, so that one cannot construct an analogous estimator $\hat{f}_{0,m}$ of $f_{0,m}$ by direct estimation of the appropriate scaling and wavelet coefficients. Instead, in this case, we shall use a linear estimator with the lowest resolution level m estimated from the data. In what follows, we shall consider the estimation of $f_{0,m}$ and $f_{c,m}$ separately.

5. Estimation of the zero-free and the zero-affected parts

Consider first the estimation of the zero-free part. In order to estimate $f_{c,m}$, we construct a wavelet thresholding estimator $\hat{f}_{c,m}$ as

$$\hat{f}_{c,m}(x) = \sum_{k \in K^{\varphi}_{0mc}} \hat{a}_{mk}\varphi_{mk}(x) + \sum_{j=m}^{J-1}\sum_{k \in K^{\psi}_{0jc}} \hat{b}_{jk}\psi_{jk}(x), \quad m_1 \leq m \leq J-1, \ x \in [0,1], \tag{5.1}$$

where $\hat{a}_{mk}$ are given in (4.6), J is defined below in (5.3), while the coefficients $\hat{b}_{jk}$ are thresholded estimators of the wavelet coefficients b_{jk} defined as

$$\hat{b}_{jk} = \begin{cases} \tilde{b}_{jk}\, \mathbb{I}(\tilde{b}_{jk}^2 > d^2 n^{-1} \ln n\, 2^{j\alpha}\, |k - k_{0j}|^{-\alpha}) & \text{if} \quad b = 0, \\ \tilde{b}_{jk}\, \mathbb{I}(|k - k_{0j}| > 2^{j-m}) & \text{if} \quad b > 0. \end{cases} \tag{5.2}$$

Here, $d > 0$ is a constant, $\tilde{b}_{jk}$ are defined by (4.6) and m is such that $m_1 \leq m \leq J-1$, where

$$2^{m_1} = \max\left(U_{\varphi^*} - L_{\varphi^*}, U_{\psi^*} - L_{\psi^*}\right) + 1, \qquad 2^J = \begin{cases} (n/\ln n)^{1/(\alpha+1)} & \text{if} \quad b = 0, \\ (\ln n)^{2/\beta} & \text{if} \quad b > 0. \end{cases} \tag{5.3}$$

Consider now the estimation of the zero-affected part. Since the estimators $\hat{a}_{mk}$ of a_{mk}, given in (4.6), have infinite variances when $k \in K^{\varphi}_{0m}$, we estimate those coefficients by solving a system of linear equations. Note that there is a finite known number of indices in K^{φ}_{0m}, at most, $w_\phi = U_\varphi - L_\varphi$ indices. For any given m, such that $m_1 \leq m \leq J-1$, denote

$$f_m(x) = \sum_{k=0}^{2^m - 1} a_{mk}\varphi_{mk}(x), \quad \varepsilon_m(x) = \sum_{j=m}^{\infty}\sum_{k=0}^{2^j-1} b_{jk}\psi_{jk}(x), \quad x \in [0,1], \tag{5.4}$$

and observe that $f(x) = f_m(x) + \varepsilon_m(x)$, so that

$$\sum_{k \in K^{\varphi}_{0m}} a_{mk}\varphi_{mk}(x) = f_m(x) - \varepsilon_m(x) - \sum_{k \in K^{\varphi}_{0mc}} a_{mk}\varphi_{mk}(x), \quad x \in [0,1]. \tag{5.5}$$

Denote $\Omega_\delta = [L_\varphi + \delta_b, U_\varphi - \delta_b]$, and choose δ_b such that

$$\delta_b = \begin{cases} 0 < \delta_b < 1/2, \ \varphi(L_\varphi + \delta_b) \neq 0, \ \varphi(U_\varphi - \delta_b) \neq 0, & \text{if} \quad b > 0, \\ 0, & \text{if} \quad b = 0. \end{cases} \tag{5.6}$$

https://doi.org/10.1051/ps/2012024 Published online by Cambridge University Press

Introduce also a finite set of indices

$$K_{0m}^{*} = \{k : 0 \le k \le 2^m - 1,\ 2L_\varphi - U_\varphi \le k_{0m} - k < L_\varphi \ \text{ or } \ U_\varphi < k_{0m} - k \le 2U_\varphi - L_\varphi\}. \tag{5.7}$$

Now, multiply both sides of formula (5.5) by $g(x)\,\varphi_{ml}(x)\,\mathbb{I}(2^m x - l \in \Omega_\delta)$, $l \in K_{0m}^{\varphi}$, where $\mathbb{I}(x \in \Omega)$ is the indicator of set Ω, and integrate. As a result, obtain the following system of linear equations

$$\mathbf{A}^{(m)}\mathbf{u}^{(m)} = \mathbf{c}^{(m)} - \boldsymbol{\varepsilon}^{(m)} - \mathbf{B}^{(m)}\mathbf{v}^{(m)}. \tag{5.8}$$

Here, matrices $\mathbf{A}^{(m)}$ and $\mathbf{B}^{(m)}$ and vectors $\mathbf{c}^{(m)}$, $\boldsymbol{\varepsilon}^{(m)}$, $\mathbf{u}^{(m)}$ and $\mathbf{v}^{(m)}$ have, respectively, elements

$$A_{lk}^{(m)} = \int_0^1 \varphi_{mk}(x)\varphi_{ml}(x)g(x)\,\mathbb{I}(2^m x - l \in \Omega_\delta)\mathrm{d}x, \quad k,l \in K_{0m}^{\varphi}, \tag{5.9}$$

$$B_{lk}^{(m)} = \int_0^1 \varphi_{mk}(x)\varphi_{ml}(x)g(x)\,\mathbb{I}(2^m x - l \in \Omega_\delta)\mathrm{d}x, \quad l \in K_{0m}^{\varphi},\ k \in K_{0m}^{*}, \tag{5.10}$$

$$c_l^{(m)} = \int_0^1 f(x)\varphi_{ml}(x)g(x)\,\mathbb{I}(2^m x - l \in \Omega_\delta)\mathrm{d}x, \quad l \in K_{0m}^{\varphi}, \tag{5.11}$$

$$\varepsilon_l^{(m)} = \int_0^1 \varepsilon_m(x)\varphi_{ml}(x)g(x)\,\mathbb{I}(2^m x - l \in \Omega_\delta)\mathrm{d}x, \quad l \in K_{0m}^{\varphi}, \tag{5.12}$$

$$u_k^{(m)} = a_{mk}, \quad k \in K_{0m}^{\varphi}, \qquad v_k^{(m)} = a_{mk}, \quad k \in K_{0m}^{*}. \tag{5.13}$$

(Note that the matrices $\mathbf{A}^{(m)}$ and $\mathbf{B}^{(m)}$ are completely known, and also observe that $B_{lk}^{(m)} \neq 0$ only if $k \in K_{0m}^{*}$, since, for $k \notin K_{0m}^{*}$, one has $\varphi_{mk}(x)\varphi_{ml}(x) = 0$).

Since $K_{0m}^{*} \subset K_{0mc}^{\varphi}$, it follows from (5.13) that components $v_k^{(m)}$ of vector $\mathbf{v}^{(m)}$ can be estimated by

$$\hat{v}_k^{(m)} = \hat{a}_{mk}, \quad k \in K_{0mc}^{\varphi},$$

using (4.6). We also estimate $c_l^{(m)}$ by

$$\hat{c}_l^{(m)} = \frac{1}{n}\sum_{i=1}^{n} y_i\,\varphi_{ml}(x_i)\,\mathbb{I}(2^m x_i - l \in \Omega_\delta), \quad l \in K_{0m}^{\varphi}, \tag{5.14}$$

and ignore vector $\boldsymbol{\varepsilon}$ in (5.8), thus, replacing (5.8) by the following system of linear equations

$$\mathbf{A}^{(m)}\hat{\mathbf{u}}^{(m)} = \hat{\mathbf{c}}^{(m)} - \mathbf{B}^{(m)}\hat{\mathbf{v}}^{(m)}. \tag{5.15}$$

Since matrix $\mathbf{A}^{(m)}$ is a positive definite matrix of non-asymptotic size, $\det(\mathbf{A}^{(m)}) \neq 0$ and we obtain the solution

$$\hat{\mathbf{u}}^{(m)} = (\mathbf{A}^{(m)})^{-1}(\hat{\mathbf{c}}^{(m)} - \mathbf{B}^{(m)}\hat{\mathbf{v}}^{(m)})$$

of the system of linear equations (5.15).

Finally, for any given m, such that $m_1 \le m \le J-1$, we set $\hat{a}_{mk} = \hat{u}_k^{(m)}$, $k \in K_{0m}^{\varphi}$, and estimate $f_{0,m}$ by the following linear wavelet estimator

$$\hat{f}_{0,m}(x) = \sum_{k \in K_{0m}^{\varphi}} \hat{a}_{mk}\varphi_{mk}(x), \quad x \in [0,1]. \tag{5.16}$$

The following statement provides the asymptotic upper bounds for the bias and the variance of the estimator $\hat{f}_{0,m}$ given in (5.16).

https://doi.org/10.1051/ps/2012024 Published online by Cambridge University Press

Lemma 5.1. *Denote* $f_{0,m}(x) = \sum_{k \in K_{0m}^{\varphi}} a_{mk}\varphi_{mk}(x)$ *and let* $m = m(n)$ *be a non-random, non-negative integer, quantity, such that* $m(n) \to \infty$ *as* $n \to \infty$. *Let the estimator* $\hat{f}_{0,m}$ *be defined by* (5.16). *Then, as* $n \to \infty$,

$$\|\mathbb{E}\hat{f}_{0,m} - f_{0,m}\|^2 = O\left(2^{-2ms'}\right), \quad \mathbb{E}\|\hat{f}_{0,m} - \mathbb{E}\hat{f}_{0,m}\|^2 = O\left(n^{-1}2^{m\alpha}\ \exp(b\,2^{m\beta}[2^{\beta+1}+1])\right). \tag{5.17}$$

Moreover, if $b = 0$, *then, as* $n \to \infty$, $\mathbb{E}\|\hat{f}_{0,m} - \mathbb{E}\hat{f}_{0,m}\|^4 = o(1)$.

Define m_0 to be such that

$$2^{m_0} = \begin{cases} n^{\frac{1}{2s'+\alpha}} & \text{if} \quad b = 0, \\ \left(b^{-1}2^{-(\beta+2)}\ln n\right)^{\frac{1}{\beta}} & \text{if} \quad b > 0. \end{cases} \tag{5.18}$$

It follows from Lemma 5.1 that, if $m = m_0$, the error $\mathbb{E}\|\hat{f}_{0,m} - f_{0,m}\|^2$ of the estimator $\hat{f}_{0,m}$ attains the asymptotic minimax lower bounds for the L^2-risk obtained in Theorem 3.1. Since α, b and β in (5.18) are known, the value of m_0 is also known in the case of $b > 0$. Therefore, one can select m_0 as the lowest resolution level in the estimator of the zero-free part (5.1).

On the other hand, the following lemma demonstrates that the wavelet thresholding estimator $\hat{f}_{c,m}$, defined in (5.1) with $m = m_0$ given in (5.18), attains the asymptotic minimax lower bounds for the L^2-risk obtained in Theorem 3.1, in the case of $b > 0$.

Lemma 5.2. *Let* $1 \le p, q \le \infty$ *and* $\max(1/2, 1/p) \le s < r$, *and let Assumption A (with* $b > 0$, $\beta > 0$ *and* $\alpha \in \mathbb{R}$*) hold. Let the estimator* $\hat{f}_{c,m}$ *be defined by* (5.1) *with* $m = m_0$ *given in* (5.18), *Then, as* $n \to \infty$,

$$\sup_{f \in B_{p,q}^s(A)} \mathbb{E}\|\hat{f}_{c,m_0} - f_{c,m_0}\|^2 \le C\,(\ln n)^{-\frac{2s'}{\beta}}. \tag{5.19}$$

Unfortunately, this idea cannot be implemented in the case of $b = 0$. Indeed, though α in (5.18) is known, the value of s' is unknown and, therefore, the estimator $\hat{f}_{0,m}$, defined in (5.1) with $m = m_0$ given in (5.18), is not realizable if $b = 0$. In this case, we need to adequately choose a resolution level, say $\hat{m}$, which approximates m_0 in some sense, and then estimate f by $\hat{f}(x) = \hat{f}_{0,\hat{m}}(x) + \hat{f}_{c,\hat{m}}(x)$. The choice of such resolution level is a rather difficult task. On the one hand, $\hat{m}$ should not be too small since, otherwise, the linear portion of the estimator would have bias that will be too large. On the other hand, since $\hat{f}_{0,m}$ is the linear estimator, in order to represent $f = f_{0,\hat{m}} + f_{c,\hat{m}}$ adequately, $\hat{m}$ has to be used as the lowest resolution level in $\hat{f}_{c,m}$.

The following lemma provides the asymptotic minimax upper bounds for the L^2-risk of the wavelet thresholding estimator $\hat{f}_{c,m}$, defined in (5.1), in the case of $b = 0$. In particular, it shows that this risk contains the component $n^{-1}\ 2^{m\alpha}$, so that in order to attain the asymptotic minimax lower bounds for the L^2-risk in the case of $b = 0$, obtained in Theorem 3.1 (up to a logarithmic factor), one needs $\hat{m} \le m_0$ with high probability.

Lemma 5.3. *Let* $1 \le p, q \le \infty$ *and* $\max(1/2, 1/p) \le s < r$, *and let Assumption A (with* $b = 0$ *and* $\alpha \ge 1$*) hold. Let the estimator* $\hat{f}_{c,m}$ *be defined by* (5.1), *where* m *is such that* $m_1 \le m \le J - 1$, *with* m_1 *and* J *defined in* (5.3). *Let* $\hat{b}_{jk}$ *be given by* (5.2) *with* $d > 4C_d$, *where* C_d *is given by*

$$C_d = 8C_\psi C_{g1}^{-1}\max\left(2, 2\|f\|_\infty^2, \|f\|_\infty\|\psi\|_\infty/3, \|\psi\|_\infty\right) \quad \textit{with} \quad C_\psi = [2\max(|L_\psi|, |U_\psi|)]^\alpha. \tag{5.20}$$

Then, as $n \to \infty$,

$$\sup_{f \in B_{p,q}^s(A)} \mathbb{E}\|\hat{f}_{c,m} - f_{c,m}\|^2 \le \begin{cases} C\left(n^{-1}\ 2^{m\alpha}\,(\ln n)^{\mathbb{I}(\alpha=1)} + n^{-\frac{2s}{2s+1}}\ (\ln n)^{\mu_1}\right) & \text{if} \quad b = 0,\ \alpha s < s', \\ C\left(n^{-1}\ 2^{m\alpha}\,(\ln n)^{\mathbb{I}(\alpha=1)} + n^{-\frac{2s'}{2s'+\alpha}}\ (\ln n)^{\mu_2}\right) & \text{if} \quad b = 0,\ \alpha s \ge s', \end{cases} \tag{5.21}$$

where,

$$\mu_1 = \frac{2s(1+\mathbb{I}(\alpha=1))}{2s+1} \quad \textit{and} \quad \mu_2 = \frac{2s'(1+\mathbb{I}(\alpha=1))}{2s'+\alpha} + \mathbb{I}\left(\frac{s'}{s} = \alpha > 1\right).$$

https://doi.org/10.1051/ps/2012024 Published online by Cambridge University Press

Moreover, as $n \to \infty$,

$$\mathbb{E}\|\hat{f}_{c,m} - f_{c,m}\|^4 = o(1). \tag{5.22}$$

Remark 5.4 (The case of an unknown design density)**.** So far, we have made the assumption that the design density function g is known. In many practical situations, however, this may not be true. Nevertheless, the suggested method can be applied to the case of an unknown g. In particular, one should start with the construction of lower and upper confidence limits $\hat{g}_L$ and $\hat{g}_U$, respectively, for the unknown g. This can be accomplished by using a variety of nonparametric methodologies for constructing simultaneous confidence intervals of a probability density function (see, *e.g.*, Tribouley [36], Bissantz *et al.* [3] and Giné and Nickl [13]). The lower estimator confidence limit $\hat{g}_L$ allows to assess the areas where g vanishes. If there are several distinct areas like that, we partition the interval $[0,1]$ into subintervals, so that each of the intervals contains only one zero of g. After that, we can estimate the location of the zero of g as the middle of the interval where the lower confidence bound for g is equal to zero. From this point onwards, without loss of generality, we assume that g vanishes at only one point of the interval $[0,1]$. We shall also limit our attention to the case of zero of a polynomial order, since, in the case of exponential zero, data loss around zero is so severe that in practice f cannot be adequately estimated. In order to implement our estimators, we need to assess the value of α and the constants C_{g1} and C_{g2} in (2.2). For this purpose, note that whenever z is small, one has the following relation for the distribution function G of g

$$G(x_0 + z) - G(x_0 - z) \approx 2C_g(\alpha+1)^{-1} z^{\alpha+1}.$$

Therefore, $\alpha + 1$ and C_g can be recovered using linear regression of $\log[\hat{G}(x_0 + z) - \hat{G}(x_0 - z)]$ onto $\log z$ for small values of z (*i.e.*, using observations surrounding x_0), where $\hat{G}$ is the empirical distribution function of G based on $x_1, x_2, \ldots, x_n$. As the value of α has been estimated by $\hat{\alpha}$, the constants C_{g1} and C_{g2} can be estimated by

$$\widehat{C}_{g1} = \min_{1\le i\le n} \frac{|\hat{G}(x_i) - \hat{G}(x_0)|(\hat{\alpha}+1)}{|x_i - x_0|^{\hat{\alpha}+1}}, \qquad \widehat{C}_{g2} = \max_{1\le i\le n} \frac{|\hat{G}(x_i) - \hat{G}(x_0)|(\hat{\alpha}+1)}{|x_i - x_0|^{\hat{\alpha}+1}}.$$

Note that the estimated values $\hat{\alpha}$, $\widehat{C}_{g1}$ and $\widehat{C}_{g2}$ of α, C_{g1} and C_{g2}, respectively, are necessary for finding the highest resolution level J and for the construction of the involved thresholds. Once the above estimates haven been obtained, we then estimate the zero-affected part of f. This procedure is relatively easy to generalize to the case of an unknown g: one just needs to replace the elements $A_{lk}^{(m)}$ and $B_{lk}^{(m)}$ of the matrices $\mathbf{A}^{(m)}$ and $\mathbf{B}^{(m)}$, given by (5.9) and (5.10), respectively, by their corresponding unbiased estimators

$$\hat{A}_{lk}^{(m)} = n^{-1}\sum_{i=1}^{n} \varphi_{mk}(x_i)\varphi_{ml}(x_i), \quad k,l \in K_{0m}^{\varphi},$$

$$\hat{B}_{lk}^{(m)} = n^{-1}\sum_{i=1}^{n} \varphi_{mk}(x_i)\varphi_{ml}(x_i), \quad l \in K_{0m}^{\varphi},\ k \in K_{0m}^{*},$$

to solve the corresponding system of linear equations for various values of m and to carry out Lepski's procedure (see Sect. 6) to choose a suitable value of $\hat{m}$. Subsequently, we estimate the wavelet coefficients using an estimator $\hat{g}$ of g in (4.6). Note that we only need to evaluate $\hat{g}$ at the points where g cannot vanish. Moreover, since we need to use $\hat{g}$ only for the "zero-free" part, we need estimators of g away from its zero where the density of observations is reasonably high.

6. Adaptive estimation and the minimax upper bounds for the L^2-risk when $1/g$ is non-integrable

In order to construct an adaptive wavelet estimator of f in the case of $b = 0$, we shall use the technique of optimal tuning parameter selection pioneered by Lepski [26,27] and further exploited in Lepski and Spokoiny [29]

https://doi.org/10.1051/ps/2012024 Published online by Cambridge University Press

and Lepski *et al.* [29]. The idea behind this technique is to construct estimators for various values of the tuning parameter in question (m, in our case), and then choose an optimal value of the tuning parameter by regulating the differences between the estimators constructed with different values of the parameter.

In particular, if $b = 0$, for various values of m, we construct versions of the system of equations (5.15), where the estimators $\hat{\mathbf{v}}^{(m)}$ are constructed as before, solve those systems and obtain the estimators (5.16), where $\hat{a}_{mk} = \hat{u}_k^{(m)}$, $k \in K_{0m}^{\varphi}$. We then construct an estimator $\hat{f}_m$ of f using formula (4.5), where $\hat{f}_{0,m}$ and $\hat{f}_{c,m}$ are given by (5.16) and (5.1), respectively, where m is the lowest resolution level of $\hat{f}_{c,m}$. The choice of the optimal resolution level is driven by the zero-affected part of f rather than the zero-free part. For this reason, for any resolution level $m > 0$, we define a neighborhood Ξ_m of x_0 as

$$\Xi_m = \left\{x : \ 2^{-m}[\min(L_\varphi, L_\psi) - U_\varphi] < x - x_0 < 2^{-m}[\max(U_\varphi, U_\psi) - L_\varphi]\right\} \tag{6.1}$$

and observe that Ξ_m is designed so that $\mathrm{supp}(f_{0,m}) \subseteq \Xi_m$, $\mathrm{supp}(\hat{f}_{0,m}) \subseteq \Xi_m$ and $\Xi_j \subset \Xi_m$ if $j > m$.

For $b = 0$, choose $m = \hat{m}$ such that $m_1 \le m \le J - 1$, where m_1 and J are defined in (5.3) and

$$\hat{m} = \min\ \left\{m : \quad \|(\hat{f}_m - \hat{f}_j)\mathbb{I}(\Xi_m)\|^2 \le \lambda^2\, 2^{j\alpha}\, n^{-1} \ln n \ \text{ for all } \ j,\ m < j \le J - 1\right\}, \tag{6.2}$$

where $\lambda > 0$ is a constant to be defined below. For completeness, define

$$\hat{m} = m_0 \qquad \text{if} \qquad b > 0, \tag{6.3}$$

where m_0 is defined in (5.18).

The construction of $\hat{m}$ for $b = 0$ is based on the following idea. Note that, when $\hat{m} \le m_0$, then, for $m = \hat{m}$, one has

$$\mathbb{E}\|\hat{f}_m - f\|^2 \le 2\left[\mathbb{E}\|\hat{f}_m - \hat{f}_{m_0}\|^2 + \mathbb{E}\|\hat{f}_{m_0} - f\|^2\right]. \tag{6.4}$$

The first component in (6.4) is small due to definition of the resolution level $\hat{m}$ while the second component is calculated at the optimal resolution level m_0 and, hence, tends to zero at the optimal (in the minimax sense) global convergence rate (up to a logarithmic factor). On the other hand, if $m = \hat{m} > m_0$, then there exists $j > m_0$ such that $\|(\hat{f}_{m_0} - \hat{f}_j)\mathbb{I}(\Xi_{m_0})\|^2 > \lambda^2\, 2^{j\alpha}\, n^{-1} \ln n$.

The following lemma shows that, if λ is large enough, the probability of the above-mentioned event is infinitesimally small. (Here, $||h||_\infty$ is the uniform norm of a bounded function h defined on $[0, 1]$).

Lemma 6.1. *Let $b = 0$ and let m_0 and $\hat{m}$ be given by (5.18) and (6.2), respectively. Denote*

$$C_{\lambda 0} = 4\sqrt{2(U_\varphi - L_\varphi + 1)}, \quad C_{\lambda 1} = C_{\lambda 0}(\sqrt{2}\, C_{g2})^{-1}\, \|(\mathbf{A}^*)^{-1}\|, \quad C_{\lambda 2} = C_{\lambda 0}\, \|(\mathbf{A}^*)^{-1}\mathbf{B}^*\|. \tag{6.5}$$

Let C_d be given by (5.20) and

$$C_\lambda = \max\left(2C_u, C_\tau C_{\lambda 0}\right), \tag{6.6}$$

where

$$C_u = \max\left(C_{\lambda 1} C_\kappa, C_{\lambda 2} C_\tau\right), \tag{6.7}$$

$$C_\tau = 8 C_\varphi C_{g1}^{-1} \max\left(2, 2\|f\|_\infty^2, \|f\|_\infty \|\varphi\|_\infty / 3, \|\varphi\|_\infty\right), \tag{6.8}$$

$$C_\kappa = \min_{a>0} \max\left(16 C_\varphi C_{g2} \|f\|_\infty, 16a, \frac{8\|f\|_\infty \|\varphi\|_\infty}{3}, 16 C_\varphi C_{g2}, \frac{4 C_\varphi C_{g2} \|\varphi\|_\infty}{a^2}, \frac{4\|\varphi\|_\infty^2}{3a}\right), \tag{6.9}$$

C_{g1} and C_{g2} are defined by (2.2) and $C_\varphi = [2\max(|L_\varphi|, |U_\varphi|)]^\alpha$. If $\lambda \ge \max\left(C_{\lambda 1}, C_{\lambda 2}\right)$, then, as $n \to \infty$,

$$\mathbb{P}\left(\hat{m} > m_0\right) = O\left(n^{-\frac{\lambda}{C_\lambda}}\right) + O\left(n^{\frac{1}{\alpha+1} - \frac{d}{2C_d}}\right). \tag{6.10}$$

https://doi.org/10.1051/ps/2012024 Published online by Cambridge University Press

Lemma 6.1 confirms that indeed $m = \hat{m}$ can be chosen as the lowest resolution level in the nonlinear part of the estimator, so that we estimate f by

$$\hat{f}(x) = \hat{f}_{0,\hat{m}}(x) + \hat{f}_{c,\hat{m}}(x), \quad x \in [0,1], \tag{6.11}$$

where $\hat{f}_{0,m}(x)$ and $\hat{f}_{c,m}(x)$ are defined in (5.16) and (5.1), respectively.

The following statement confirms that, when $1/g$ is non-integrable, the adaptive wavelet thresholding estimator $\hat{f}$ defined by (6.11) attains (up to a logarithmic factor if $b = 0$) the asymptotic minimax lower bounds for the L^2-risk obtained in Theorem 3.1.

Theorem 6.2. *Let $1 \le p, q \le \infty$ and $\max(1/2, 1/p) \le s < r$, and let Assumption A (with $\alpha \ge 1$ if $b = 0$ and $\alpha \in \mathbb{R}$ if $b > 0$) hold. Let $\hat{f}$ be the estimator defined by (6.11) with $\lambda > \max(2C_\lambda, C_{\lambda 1}, C_{\lambda 2})$ in (6.2) and $d > 2(\alpha+1)^{-1}(2\alpha+3)C_d$, where C_λ is defined in (6.6), $C_{\lambda 1}, C_{\lambda 2}$ are defined in (6.5), and C_d is defined in (5.20). Then, as $n \to \infty$,*

$$\sup_{f \in B^s_{p,q}(A)} \mathbb{E}\|\hat{f} - f\|^2 \le \begin{cases} C\, n^{-\frac{2s}{2s+1}}\, (\ln n)^{\frac{2s(1+\mathbb{I}(\alpha=1))}{2s+1}} & \text{if} \quad b = 0,\ \alpha s < s', \\ C\, n^{-\frac{2s'}{2s'+\alpha}}\, (\ln n)^{\frac{2s'(1+\mathbb{I}(\alpha=1))}{2s'+\alpha} + \mathbb{I}\left(\frac{s'}{s} = \alpha > 1\right)} & \text{if} \quad b = 0,\ \alpha s \ge s', \\ C\, (\ln n)^{-\frac{2s'}{\beta}} & \text{if} \quad b > 0. \end{cases}$$

Remark 6.3 (Adaptivity)**.** Theorems 3.1 and 6.2 demonstrate that, for severe data losses ($b > 0$), the adaptive wavelet thresholding estimator $\hat{f}$ given by (6.11) attains the asymptotically optimal (in the minimax sense) global convergence rates. For moderate data losses ($b = 0$ with $\alpha \ge 1$), however, the adaptive wavelet thresholding estimator $\hat{f}$ given by (6.11) is asymptotically near-optimal (up to a logarithmic factor). Moreover, if p is large and $\alpha > 1$ is relatively small ($1 < \alpha < (1/2 - 1/p)/s$), then data loss does not affect the asymptotic minimax global convergence rates and they coincide with the asymptotic minimax global convergence rates obtained in the absence of data losses.

Remark 6.4 (Relation to local and uniform convergence rates)**.** The suggested estimation of the zero-affected part of f is somewhat similar to the procedure of Gaïffas [9,11], with the difference that he used local polynomials while we are using wavelets. However, the significant difference is that we use this estimator only for the zero-affected part and not for the whole function f. Another difference between our and Gaïffas' studies is that, first, we are able to formulate the asymptotic minimax convergence rates explicitly, in a simple meaningful way, and, due to the fact that we are using thresholding of wavelet coefficients rather than solution of the system of equations as in Gaïffas, our estimator can adapt to the case when the estimated function is spatially inhomogeneous. Moreover, we should point out that our asymptotic minimax convergence rates are global and over a wide range of Besov balls compare to Gaiffas that are local or uniform and only for Hölder spaces. In particular, Gaïffas [9, 11] deals only with estimation of f at x_0, the zero of the design density function g. The asymptotic minimax local convergence rates of his estimator can be expressed explicitly *via* α and the parameters of the Hölder ball that f belongs. However, as we pointed out in Remark 2.2, this problem is much easier than the global estimation problem we considered and can be solved by straightforward calculus. Furthermore, Gaïffas [10, 12] studied asymptotic minimax uniform convergence rates. The derived convergence rates are formulated in terms of a solution of a nonlinear equation, and there are no explicit expressions for these rates in a general situation. For instance, the only example, which appears in Gaïffas [12], is produced for the simplest situation when $\sigma = 1$, f belongs to a Hölder class with parameters $s = L = 1$ and the design density g is of the form $g(x) = 4|x - 1/2|$, *i.e.*, $\alpha = 1$. In this case, the asymptotic minimax uniform convergence rates are given by

$$r_n(x) = (\log n/n)^{\alpha_n(x)},$$

https://doi.org/10.1051/ps/2012024 Published online by Cambridge University Press

where

$$\alpha_n(x) = \begin{cases} \frac{1}{3}\left(1 - \frac{1-2x)}{\log(\log n/n)}\right), & x \in [0, 0.5 - (\log n/2n)^{1/4}), \\ \frac{\left([(x-0.5)^4 + 4\log n/n]^{1/2} - (x-0.5)^2\right) - \log 2}{2\log(\log n/n)}, & x \in [0.5 - (\log n/2n)^{1/4}, 0.5 - (\log n/2n)^{1/4}], \\ \frac{1}{3}\left(1 - \frac{2x-1)}{\log(\log n/n)}\right), & x \in (0.5 + (\log n/2n)^{1/4}, 1]. \end{cases}$$

On the other hand, in the case when $\alpha > 1$ or in the case when α is not an integer, a solution of the corresponding equation which produces the asymptotic minimax uniform convergence rates, as well as derivation of the explicit expression for these rates, require very nontrivial investigation.

7. Adaptive estimation and the minimax upper bounds for the L^2-risk when $1/g$ is integrable

The case when $1/g$ is integrable, *i.e.*, when g has a zero of a polynomial order α, $0 < \alpha < 1$, has been considered by Chesneau [6] who demonstrated that the problem is well-posed when f is spatially homogeneous, *i.e.*, when $p \geq 2$. However, the lower bounds in Theorem 3.1 show that the problem becomes ill-posed when $\alpha s > s'$, *i.e.*, when $1 \leq p < (s(1-\alpha) + 1/2)^{-1}$. Hence, by considering only spatially homogeneous regression functions ($p \geq 2$), Chesneau [6] missed the "elbow rate" when f is spatially inhomogeneous and the fact that the problem becomes ill-posed in this case. However, since the estimators (4.6) (of the scaling and wavelet coefficients) have finite variances for $0 < \alpha < 1$, one can construct an adaptive estimator, similar to the one considered in Chesneau [6], by simply thresholding wavelet coefficients. In particular, set

$$\hat{f}(x) = \sum_{k=0}^{2^{m_1}-1} \hat{a}_{m_1 k}\varphi_{m_1 k} + \sum_{j=m_1}^{J-1}\sum_{k=0}^{2^j-1} \hat{b}_{jk}\psi_{jk}(x), \tag{7.1}$$

where $\hat{a}_{mk}$ and $\hat{b}_{jk}$ are defined in (5.2), and m_1 and J are defined in (5.3) with $b = 0$.

The following statement confirms that estimator (7.1) attains (up to a logarithmic factor) the asymptotic minimax lower bounds for the L^2-risk obtained in Theorem 3.1.

Theorem 7.1. *Let $1 \leq p < \infty, 1 \leq q \leq \infty$, $\max(1/2, 1/p) \leq s < r$, and let d in (5.2) satisfy*

$$d > \frac{2C_d(3\alpha + 5)}{(1-\alpha)(1+\alpha)}, \tag{7.2}$$

where C_d is given by (5.20). Let Assumption A (with $b = 0$ and $0 < \alpha < 1$) hold, and let $\hat{f}$ be the estimator defined by (7.1). Then, as $n \to \infty$,

$$\sup_{f \in B^s_{p,q}(A)} \mathbb{E}\|\hat{f} - f\|^2 \leq \begin{cases} C\, n^{-\frac{2s}{2s+1}} (\ln n)^{\frac{2s}{2s+1}} & \text{if} \quad \alpha s < s', \\ C\, n^{-\frac{2s'}{2s'+\alpha}} (\ln n)^{\frac{2s'}{2s'+\alpha} + \mathbb{I}(\alpha s = s')} & \text{if} \quad \alpha s \geq s'. \end{cases} \tag{7.3}$$

Theorem 7.1 shows that, for $b = 0$ and $0 < \alpha < 1$, the aforementioned estimation problem is well-posed as long as $p > (s(1-\alpha) + 1/2)^{-1}$ and it becomes ill-posed when $p < (s(1-\alpha) + 1/2)^{-1}$. Therefore, even when data loss is very moderate ($b = 0$ and $0 < \alpha < 1$), the estimation problem becomes ill-posed whenever f is rather spatially inhomogeneous ($p < (s(1-\alpha) + 1/2)^{-1}$).

Remark 7.2 (Integrable and non-integrable design density)**.** For $b = 0$, the asymptotic minimax global convergence rates in Theorems 2 and 3 are the same, except for $\alpha = 1$. The reason for this lies in the fact that these rates are not driven by the fact whether $1/g$ is integrable ($b = 0$ and $0 < \alpha < 1$) or non-integrable ($b = 0$ and $\alpha \geq 1$) but by the relation between αs and s'. This clearly follows from Theorem 3.1 which establishes the asymptotic minimax lower bounds for the L^2-risk.

https://doi.org/10.1051/ps/2012024 Published online by Cambridge University Press

8. Discussion

We considered the nonparametric regression estimation problem of recovering an unknown response function f on the unit interval $[0,1]$ on the basis of incomplete data when the design density function g is known and has a zero $x_0 \in [0,1]$ of a polynomial or an exponential order. We investigated the asymptotic (as the sample size increases) global estimation (in the minimax sense and for an L^2-risk) of f over a wide range of Besov balls $B^s_{p,q}(A)$ of radius $A > 0$, where $1 \leq p, q \leq \infty$ and $\max(1/p, 1/2) \leq s < r$, where $r > 0$ is the regularity parameter associated with the wavelet system. The aforementioned global nonparametric regression estimation problem is a much harder problem than the local nonparametric regression estimation problem studied by Gaiffas [9, 11], since it cannot be reduced to the estimation of a related regularly-sampled function (see Rems. 2.1 and 2.2). As a spatially inhomogeneous ill-posed problem, the resulting estimators demonstrate completely different patterns of behavior in comparison with spatially homogeneous ill-posed problems like, *e.g.*, deconvolution.

We studied various regimes of data loss, ranging from relatively minor data losses (when g has a zero of polynomial order $0 < \alpha < 1$, so that $1/g$ is integrable) to moderate data losses (when g has a zero of polynomial order $\alpha \geq 1$ and, hence, $1/g$ is non-integrable) and, last, to severe data losses (when g has a zero of exponential order, so that $1/g$ is non-integrable).

Asymptotic minimax global convergence rates in the case of minor data losses ($0 < \alpha < 1$) were studied by Chesneau [6] who showed that the problem is well-posed (the asymptotic minimax global convergence rates are the same as in the absence of data loss) whenever the regression function f is spatially homogeneous ($p \geq 2$). As our study shows, the problem remains well-posed even if f is spatially inhomogeneous as long as the data loss is very minor ($0 < \alpha < 1 - (1/p - 1/2)/s$) or the function is relatively smooth ($p > (1/2 - s(\alpha - 1))^{-1}$). When $\alpha \geq 1 - (1/p - 1/2)/s$ ($p \leq (1/2 - s(\alpha - 1))^{-1}$), the problem becomes ill-posed.

Now, consider the situation when data loss is moderate ($b = 0$ and the zero of g is of a polynomial order $\alpha \geq 1$). The problem is now ill-posed if $\alpha \geq (1/2 - 1/p)/s$, *i.e.*, it is always ill-posed when f is spatially inhomogeneous ($1 \leq p < 2$). However, as Remark 3.2 points out, when f is very spatially homogeneous (p is rather large) and data loss is relatively moderate ($1 < \alpha < (1/2 - 1/p)/s$), the estimation problem of f ceases to be ill-posed and exhibits asymptotic minimax global convergence rates observed when g is bounded from below. Thus, in the case when f is very spatially homogeneous, the estimator of f is "borrowing strength" in the areas where f is adequately sampled and exhibits asymptotic minimax global convergence rates common for regularly spaced regression estimation problems. This is very dissimilar to spatially homogeneous ill-posed problems (*e.g.*, deconvolution) where there is a change point in the asymptotic minimax global convergence rates (the, so-called, elbow effect) when f is spatially inhomogeneous ($1 \leq p < 2$) and they are independent of p when it is spatially homogeneous ($2 \leq p \leq \infty$). On the contrary, in the case of spatially inhomogeneous ill-posed problems, like the one considered herein, the asymptotic minimax global convergence rates depend on p even when the function is spatially homogeneous ($2 < p \leq \infty$) as long as $\alpha \geq (1/2 - 1/p)/s$. Thus, the elbow effect occurs when $p > 2$, in particular, when $p > 2/(1 - (\alpha - 1)s)$ provided that $1 < \alpha < 1 + 1/s$.

In the case when data loss is severe ($b > 0$ and the zero of g is of an exponential order $\beta > 0$), the asymptotic minimax global convergence rates grow with p, *i.e.*, the more spatially homogeneous f is, the better it can be estimated. This is unlike spatially homogeneous ill-posed problems (*e.g.*, deconvolution) where the minimax global convergence rates improve when p is growing when $1 < p < 2$ and are independent of p when f is spatially homogeneous ($2 \leq p \leq \infty$) (see Pensky and Sapatinas [32, 33]).

The unusual behavior of the asymptotic minimax global convergence rates in the case of the spatially inhomogeneous ill-posed problem considered above calls for different adaptive estimation strategies. In particular, whenever data loss is moderate or severe, we partition f into zero-affected and zero-free parts. First, we construct a linear wavelet estimator of the zero-affected part where the lowest resolution level $m = \hat{m}$ is independent of the unknown parameters of the Besov ball that f is assumed to belong and, therefore, known when $b > 0$, and is chosen using Lepski's method when $b = 0$. After that, we construct a nonlinear (thresholding) wavelet estimator of the zero-free part of f starting from the lowest resolution level $m = \hat{m}$. Note that the nonlinear estimator is required even if g has a zero of exponential order ($b > 0$). This is very different from the case of spatially

https://doi.org/10.1051/ps/2012024 Published online by Cambridge University Press

homogeneous ill-posed problems (*e.g.*, deconvolution), where in the case of exponentially growing eigenvalues, a linear estimator usually attains asymptotically optimal (in the minimax sense) global convergence rates (see Pensky and Sapatinas [32, 33]).

We should also mention that there is a significant difference between asymptotic minimax local and asymptotic minimax global convergence rates. Note that asymptotic minimax local convergence rates at a zero of g are always affected by loss of data, even for moderate data losses. The asymptotic minimax global convergence rates, however, are not affected when data loss is limited and the regression function is very spatially homogeneous ($1 < \alpha < 1 + 1/s$ and $p > 2/(1 - (\alpha - 1)s)$).

Finally, we point out that some of the logarithmic factors which appear in Theorems 1–3 could be possibly removed by using block thresholding rather than the considered term-by-term thresholding of wavelet coefficients. Furthermore, due to its construction, the suggested adaptive wavelet thresholding estimator is not easily computable, so it is of limited practical use. Therefore, it is desirable to construct an alternative, more computational feasible, adaptive estimator which attains the asymptotic minimax global convergence rates, that was the aim of this work. This is the project for future work that we hope to address elsewhere.

9. Proofs

Since the paper contains a large number of statements, below we give a road map of Section 9. Section 9.1 contains the proof of the asymptotic minimax lower bounds for the L^2-risk. Theorem 6.2, which provides the asymptotic minimax upper bounds for the L^2-risk of an adaptive estimator of f in the case of $\alpha \geq 1$ if $b = 0$ and $\alpha \in \mathbb{R}$ if $b > 0$, is proved in Section 9.7. The Proof of Theorem 6.2 is based on Lemmas 5.1–6.1. In particular, Lemma 5.1, which is proved in Section 9.5, gives an asymptotic minimax upper bound for the L^2-risk of the zero-affected portion of the estimator at a fixed resolution level m. Lemmas 5.2 and 5.3 provide asymptotic minimax upper bounds for the L^2-risk of the zero-free part of the estimator when estimation is carried out (in the case of an exponential zero) or is started (in the case of a polynomial zero) at a fixed resolution level chosen in advance. Last, Lemma 6.1 proves that, with high probability, the resolution level chosen by Lepski's procedure is not higher than the optimal resolution level. The Proof of Lemma 5.1 is included in Section 9.5 while the proofs of Lemmas 5.2 and 5.3 are given in Section 9.6. Section 9.4 contains the Proof of Lemma 6.1 as well as large deviation results for a wavelet or scaling coefficient (Lem. 9.6) or the right-hand side of the system of linear equations (Lem. 9.7).

Sections 9.2 and 9.3 contain supplementary statements which are used in the proofs of Lemmas 5.1–5.3. In particular, Lemma 9.2, proved in Section 9.2, provides upper bounds for moments and covariances of wavelet and scaling coefficients. In Section 9.3, Lemma 9.3 contains a purely technical auxiliary result, while Lemma 9.4 provides upper and lower bounds for the entries of the matrices which appear in the system of linear equations which is used for the construction of the zero-affected part of the estimator.

Finally, Theorem 7.1, which delivers the asymptotic minimax upper bounds for the L^2-risk of an adaptive estimator of f in the case of $b = 0$ and $0 < \alpha < 1$, is proved in Section 9.8. The proof of this theorem requires a technical result provided by Lemma 9.9 which precedes Theorem 7.1 in Section 9.8.

9.1. Proof of the asymptotic minimax lower bounds for the L^2-risk

Proof of Theorem 3.1. On noting that the asymptotic minimax lower bounds for the L^2-risk in Theorem 3.1 of Chesneau [6] is also true when $b = 0$ and $1/g$ is non-integrable ($\alpha \geq 1$), the asymptotic minimax lower bounds for the L^2-risk in the first part of (3.2) can be obtain by the arguments of Chesneau [6] and, hence, we need to prove only the asymptotic minimax lower bounds for the L^2-risk in the second and third parts of (3.2). For this purpose, we consider functions f_{jk} be of the form $f_{jk} = \gamma_j \psi_{jk}$ and let $f_0 \equiv 0$. Note that by (3.1), in order $f_{jk} \in B^s_{p,q}(A)$, we need $\gamma_j \leq A 2^{-js'}$. Set $\gamma_j = c 2^{-js'}$, where c is a positive constant such that $c < A$, and apply the following classical lemma on lower bounds:

Lemma 9.1 (Härdle, Kerkyacharian, Picard & Tsybakov [16], Lem. 10.1)**.** *Let V be a functional space, and let $d(\cdot, \cdot)$ be a distance on V. For $f, g \in V$, denote by $\Lambda_n(f, g)$ the likelihood ratio $\Lambda_n(f, g) = d\mathbb{P}_{X_n^{(f)}} / d\mathbb{P}_{X_n^{(g)}}$, where*

https://doi.org/10.1051/ps/2012024 Published online by Cambridge University Press

$d\mathbb{P}_{X_n^{(h)}}$ is the probability distribution of the process X_n when h is true. Let V contains the functions $f_0, f_1, \ldots, f_{\aleph}$ such that

(*a*) $d(f_k, f_{k'}) \geq \delta > 0$ *for* $k = 0, 1, \ldots, \aleph,\ k \neq k'$;
(*b*) $\aleph \geq \exp(\lambda_n)$ *for some* $\lambda_n > 0$;
(*c*) $\ln \Lambda_n(f_0, f_k) = u_{nk} - v_{nk}$, *where* v_{nk} *are constants and* u_{nk} *is a random variable such that there exists* $\pi_0 > 0$ *with* $\mathbb{P}_{f_k}(u_{nk} > 0) \geq \pi_0$;
(*d*) $\sup_k v_{nk} \leq \lambda_n$.

Then, for an arbitrary estimator $\hat{f}$,

$$\sup_{f \in V} \mathbb{P}_{X_n^{(f)}}\big(d(\hat{f}, f) \geq \delta/2\big) \geq \pi_0/2.$$

Let now $V = \{f_{jk} : |k - k_{0j}| \leq K/2\}$, where $K > 2$ is a fixed positive constant, so that $\aleph = K$. Choose $d(f, g) = \|f - g\|$, where, as before, $\|\cdot\|$ denotes the L^2-norm on the interval $[0, 1]$. Then, $d(f_{jk}, f_0) = \gamma_j = \delta$. Let $v_{nk} = \lambda_n = \ln K$ and $u_{nk} = \ln \Lambda_n(f_0, f_{jk}) + \ln K$. Now, in order to apply Lemma 9.1, we need to show that for some $\pi_0 > 0$, uniformly for all f_{jk}, we have

$$\mathbb{P}_{f_{jk}}(u_{nk} > 0) = \mathbb{P}_{f_{jk}}\left(\ln \Lambda_n(f_0, f_{jk}) > -\ln K\right) \geq \pi_0 > 0.$$

Since, by Chebychev's inequality,

$$\mathbb{P}_{f_{jk}}\left(\ln \Lambda_n(f_0, f_{jk}) > -\ln K\right) \geq 1 - \frac{\mathbb{E}_{f_{jk}} \left|\ln \Lambda_n(f_0, f_{jk})\right|}{\ln K},$$

we need to find a uniform upper bound for $\mathbb{E}_{f_{jk}} |\ln \Lambda_n(f_0, f_{jk})|$.

Note that

$$-2 \ln \Lambda_n(f_0, f_{jk}) = \sum_{i=1}^{n} \gamma_j^2 \psi_{jk}^2(x_i) + 2 \sum_{i=1}^{n} \gamma_j \psi_{jk}(x_i) \xi_i$$

where ξ_i, $i = 1, 2, \ldots, n$, are independent standard Gaussian random variables. Thus,

$$\mathbb{E}| - 2 \ln \Lambda_n(f_0, f_{jk})| \leq A_n + 2B_n,$$

where

$$A_n = \mathbb{E}| \sum_{i=1}^{n} \gamma_j^2 \psi_{jk}^2(x_i)| = n\gamma_j^2 \int_0^1 \psi_{jk}^2(x) g(x) \mathrm{d}x, \quad B_n = \mathbb{E}| \sum_{i=1}^{n} \gamma_j \psi_{jk}(x_i) \xi_i|.$$

Note that by Jensen's inequality,

$$\begin{aligned} B_n &= \mathbb{E}\left\{ \mathbb{E}\left[\left| \sum_{i=1}^{n} \gamma_j \psi_{jk}(x_i) \xi_i \right| \,\middle|\, x_1, x_2, \ldots, x_n \right] \right\} \\ &\leq \mathbb{E}\left\{ \mathbb{E}\left[\left(\sum_{i=1}^{n} \gamma_j \psi_{jk}(x_i) \xi_i \right)^2 \,\middle|\, x_1, x_2, \ldots, x_n \right] \right\}^{1/2} = \sqrt{A_n}, \end{aligned}$$

so that one needs uniform upper bounds for A_n only.

If j is large enough, A_n can be presented as

$$A_n = n\gamma_j^2 \int_{L_\psi}^{U_\psi} \psi^2(z) g(x_0 + 2^{-j}(z + k - k_{0j})) \mathrm{d}z,$$

https://doi.org/10.1051/ps/2012024 Published online by Cambridge University Press

where $k_{0j} = 2^j x_0$. Observe that condition (2.1) implies that one has

$$g(x_0 + x) \leq C|x|^{\alpha} \exp(-b|x|^{-\beta}).$$

Let $M_\psi = \max(|L_\psi|, |U_\psi|)$. Then, for a finite value of K, one has

$$A_n \leq Cn\gamma_j^2 2^{-j\alpha} \ (M_\psi^\alpha + K^\alpha) \exp\left(-b2^{j\beta}(M_\psi + K)^{-\beta}\right).$$

Now, recall that $\gamma_j = c2^{-js'}$ and choose the smallest possible value of j such that A_n are uniformly bounded. Simple calculation yield that

$$A_n = O\left(2^{-j(2s'+\alpha)} \exp\left(- b2^{j\beta}[M_\psi + K]^{-\beta}\right)\right),$$

so that $2^j = O\left(n^{1/(2s'+\alpha)}\right)$ if $b = 0$ and $2^j = O\left((\ln n)^{1/\beta}\right)$ if $b > 0$. Now, applying Lemma 9.1 and Chebyshev inequality, we finally obtain

$$\inf_{\tilde{f}_n} \sup_{f \in B^s_{p,q}(A)} \mathbb{E}\|\tilde{f}_n - f\|^2 \geq \inf_{\tilde{f}_n} \sup_{f \in V} (\gamma_j^2/4) \ \mathbb{P}(\|\tilde{f}_n - f\| > \gamma_j/2) \geq \pi_0 \gamma_j^2/8,$$

which, on noting that

$$\frac{2s'}{2s'+\alpha} < \frac{2s}{2s+1} \quad \text{if and only if} \quad s' < \alpha s, \tag{9.1}$$

completes the proof of the theorem. □

9.2. Properties of the estimators of scaling and wavelet coefficients

Consider the quantity

$$J_{mkl} = \int 2^m |\varphi(2^m x - k)\varphi(2^m x - l)| \, g^{-1}(x) \, \mathrm{d}x. \tag{9.2}$$

Lemma 9.2. *Let $m = m(n)$ be a non-random, non-negative integer, quantity, with $m(n) \to \infty$ as $n \to \infty$, and let $\hat{a}_{mk}$ be defined by (4.6). Then, for $k, l \in K^{\varphi}_{0mc}$, as $n \to \infty$,*

$$|Cov(\hat{a}_{mk}, \hat{a}_{ml})| = O\left(n^{-1} \ (J_{mkl} + 1)\right), \tag{9.3}$$

where

$$J_{mkl} = O\left(n^{-1} \, 2^{m\alpha} |k - k_{0m}|^{-\alpha} \exp(b2^{m\beta}|k - k_{0m}|^{-\beta})\right) \qquad \text{if} \qquad |k - l| \leq U_\varphi - L_\varphi, \tag{9.4}$$

and $J_{mkl} = 0$ otherwise. Moreover, if $b = 0$, then, as $n \to \infty$,

$$\begin{aligned} Var(\hat{a}_{mk}) &= O\left(n^{-1} \, 2^{m\alpha} |k - k_{0m}|^{-\alpha}\right), \\ \mathbb{E}(\hat{a}_{mk} - a_{mk})^4 &= O\left(n^{-3} \, 2^{m(3\alpha+1)} |k - k_{0m}|^{-3\alpha}\right) + O\left(n^{-2} \, 2^{2m\alpha} |k - k_{0m}|^{-2\alpha}\right). \end{aligned} \tag{9.5}$$

Similarly, if $k, l \in K^{\psi}_{0jc}$ and $b = 0$, then $\tilde{b}_{jk}$, defined in (4.6), satisfy, as $n \to \infty$,

$$Var(\tilde{b}_{jk}) = O\left(n^{-1} \, 2^{j\alpha} \ |k - k_{0j}|^{-\alpha}\right), \tag{9.6}$$

$$\mathbb{E}(\tilde{b}_{jk} - b_{jk})^4 = O\left(n^{-3} \, 2^{j(3\alpha+1)} |k - k_{0j}|^{-3\alpha}\right) + O\left(n^{-2} \, 2^{2j\alpha} |k - k_{0j}|^{-2\alpha}\right), \tag{9.7}$$

$$\begin{aligned} \mathbb{E}(\tilde{b}_{jk} - b_{jk})^6 = {} & O\left(n^{-5} \, 2^{j(5\alpha+2)} |k - k_{0j}|^{-5\alpha}\right) + O\left(n^{-4} \, 2^{j(4\alpha+1)} |k - k_{0j}|^{-4\alpha}\right) \\ & + O\left(n^{-3} \, 2^{3j\alpha} |k - k_{0j}|^{-3\alpha}\right). \end{aligned} \tag{9.8}$$

https://doi.org/10.1051/ps/2012024 Published online by Cambridge University Press

Proof of Lemma 9.2. Let us first prove formula (9.4). Changing variables $z = 2^m(x - x_0)$ in the integral in (9.2), and using inequality (2.2), derive that

$$J_{mkl} \leq \frac{2^{m\alpha}}{C_{g1}} \int_{L_\varphi}^{U_\varphi} \frac{|\varphi(z)||\varphi(z+k-l)|\, dz}{|z+k-k_{0m}|^\alpha \exp\left(-b2^{m\beta}|z+k-k_{0m}|^{-\beta}\right)}.$$

It is easy to note that $J_{mkl} = 0$ if $|k-l| > U_\varphi - L_\varphi$. Also, $k \in K^\psi_{0jc}$ implies that $k_{0m} - k \leq L_\varphi - 1$ or $k_{0m} - k \geq U_\varphi + 1$, so that one has $|z + k - k_{0m}| \geq 1$ and, hence, $|z + k - k_{0m}| \propto |k - k_{0m}|$ which proves (9.4). Now, by direct calculations we obtain that

$$\mathrm{Cov}(\hat{a}_{mk}, \hat{a}_{ml}) = n^{-1} \left\{ \int [\sigma^2 + f^2(x)]\ \varphi_{mk}(x)\varphi_{ml}(x) g^{-1}(x) \mathrm{d}x - a_{mk}a_{ml} \right\},$$

so that (9.3) is valid.

Since the proofs for the scaling and the wavelet coefficients in Lemma 9.2 are similar, we shall prove only formulae (9.6)-(9.8). Observe that, due to (2.2) and the fact that $k \in K^\psi_{0jc}$ implies $k_{0j} - k \leq L_\psi - 1$ or $k_{0j} - k \geq U_\psi + 1$, by considerations similar to the ones provided above, for integers $r_1, r_2 > 0$, one has

$$\int (g(x))^{-r_2}\, (\psi_{jk}(x))^{2r_1}\, \mathrm{d}x \leq C\, 2^{j(r_1-1)}\ 2^{jr_2\alpha}\ |k - k_{0j}|^{-r_2\alpha}. \tag{9.9}$$

Now, to complete the proof of (9.6)–(9.8), as $n \to \infty$, apply (9.9) to the following formulae

$$\begin{aligned}
\mathrm{Var}(\tilde{b}_{jk}) &= O\left(n^{-1} \int g^{-1}(x)\psi^2_{jk}(x)\mathrm{d}x \right), \\
\mathbb{E}(\tilde{b}_{jk} - b_{jk})^4 &= O\left(n^{-3} \int g^{-3}(x)\psi^4_{jk}(x)\mathrm{d}x + n^{-2} \left[\int g^{-1}(x)\psi^2_{jk}(x)\mathrm{d}x \right]^2 \right), \\
\mathbb{E}(\tilde{b}_{jk} - b_{jk})^6 &= O\left(n^{-5} \int g^{-5}(x)\psi^6_{jk}(x)\mathrm{d}x + n^{-3} \left[\int g^{-1}(x)\psi^2_{jk}(x)\mathrm{d}x \right]^3 \right. \\
&\quad \left. + \ n^{-4} \int g^{-3}(x)\psi^4_{jk}(x)\mathrm{d}x \int g^{-1}(x)\psi^2_{jk}(x)\mathrm{d}x \right).
\end{aligned}$$

□

9.3. Proofs of the supplementary statements used in the Proof of Lemma 5.1.

Lemma 9.3. *Let* $\delta_0 = 0.5\ 3^{\beta+1}\ (2\ 3^{\beta+1} + (2M)^{\beta+1})^{-1}$, *where* $\beta > 0$ *and* $M > 0$, *and let* $0 < \delta < \delta_0$ *and* $a, b \in [2 - \delta, M]$. *Let* $c > 0$ *be such that* $c \leq \min(a, b) + \delta$ *and* $c \leq \max(a, b) - (1 - 2\delta)$. *Then,*

$$a^{-\beta} + b^{-\beta} - 2c^{-\beta} \leq -0.5\, \beta M^{-(\beta+1)}.$$

Proof of Lemma 9.3. Note that $\delta_0 < 1/2$ and that $\delta < \delta_0$ implies $\delta < 1/2$. Let, without loss of generality, $a \leq b$. Then, $c \leq a + \delta$, $c \leq b - (1 - 2\delta)$ and

$$\begin{aligned}
a^{-\beta} - c^{-\beta} &\leq a^{-(\beta+1)}\beta\delta \leq (2 - \delta)^{-(\beta+1)}\beta\delta \\
b^{-\beta} - c^{-\beta} &\leq b^{-\beta} - (b - 1 + 2\delta)^{-\beta} \leq -\beta M^{-(\beta+1)}\,(1 - 2\delta).
\end{aligned}$$

Therefore, taking into account that $2 - \delta > 3/2$ and $0 < \delta < \delta_0$, we obtain

$$\begin{aligned}
a^{-\beta} + b^{-\beta} - 2c^{-\beta} &\leq (2 - \delta)^{-(\beta+1)}\beta\delta - \beta M^{-(\beta+1)}\,(1 - 2\delta) \\
&< -\beta M^{-(\beta+1)}(3/2)^{-(\beta+1)}[(3/2)^{\beta+1} - \delta_0(M^{\beta+1} + 2(3/2)^{\beta+1})] = -0.5\beta M^{-(\beta+1)},
\end{aligned}$$

which proves the lemma. □

https://doi.org/10.1051/ps/2012024 Published online by Cambridge University Press

Lemma 9.4. *Let $A_{kl}^{(m)}$, $B_{kl}^{(m)}$, $c_l^{(m)}$ and $\hat{c}_l^{(m)}$ be given by (5.9)–(5.14), respectively. Then, $Var(\hat{c}_k^{(m)}) = O\left(n^{-1}\, A_{kk}^{(m)}\right)$, and, for some constants $C_1 > 0$ and $C_2 > 0$, one has*

$$C_1 2^{-m\alpha} \exp(-b2^{\beta(m+1)}) \le A_{kk}^{(m)} \le C_2 2^{-m\alpha} \exp(-b\, M_\varphi^{-\beta}\, 2^{m\beta}), \tag{9.10}$$

where $M_\varphi = U_\varphi - L_\varphi + \max(|U_\varphi|, |L_\varphi|)$. Moreover, if $b > 0$ and $0 < \delta_b < \delta_0$ for $\delta_0 = 0.5\ 3^{\beta+1}\ [2\ 3^{\beta+1} + (U_\varphi + L_\varphi)^{\beta+1}]^{-1}$, then

$$\frac{|A_{kl}^{(m)}|}{\sqrt{A_{kk}^{(m)}}\sqrt{A_{ll}^{(m)}}} \le C \exp\left(-0.25\ b\,(U_\varphi + L_\varphi)^{-(\beta+1)}\ 2^{m\beta}\right). \tag{9.11}$$

In addition, if $b = 0$ and $m_1 \le m \le J-1$, then, as $n \to \infty$,

$$\|\mathbf{B}^{(m)}\| = O\left(2^{-m\alpha/2}\right), \qquad \mathbb{E}(\hat{c}_k^{(m)} - c_k^{(m)})^4 = O\left(n^{-2}2^{-2m\alpha}\right), \tag{9.12}$$

where $\|\mathbf{B}^{(m)}\|$ is the spectral norm of matrix $\mathbf{B}^{(m)}$.

Proof of Lemma 9.4. First, note that, by (5.9), one has

$$\mathrm{Var}(\hat{c}_k^{(m)}) = n^{-1} \int \varphi_{mk}^2(x)(f^2(x) + \sigma^2) g(x)\, \mathbb{I}(2^m x - l \in \Omega_\delta)\mathrm{d}x = O\left(n^{-1}\, A_{kk}^{(m)}\right).$$

If $b = 0$, then

$$\begin{aligned}\mathbb{E}(\hat{c}_k^{(m)} - c_k^{(m)})^4 &= O\left(n^{-3} \int \varphi_{mk}^4(x) g(x)\mathrm{d}x + n^{-2}\left[\int \varphi_{mk}^2(x) g(x)\mathrm{d}x\right]^2\right) \\ &= O\left(n^{-3}\, 2^m 2^{-m\alpha} + n^{-2}2^{-2m\alpha}\right) = O\left(n^{-2}2^{-2m\alpha}\right),\end{aligned}$$

since $n^{-1}2^{m(1+\alpha)} < 1$ for $m_1 \le m \le J-1$, which completes the proof of the second half of (9.12). Now, observe that, as $n \to \infty$,

$$A_{kl}^{(m)} = \int \varphi(z + k_{0m} - k)\varphi(z + k_{0m} - l) g(x_0 + 2^{-m}z)\mathbb{I}(z + k_{0m} - l \in \Omega_\delta)\mathrm{d}z \tag{9.13}$$

$$\sim C_g\ 2^{-m\alpha} \int \varphi(z + k_{0m} - k)\varphi(z + k_{0m} - l)|z|^\alpha \exp(-b2m\beta|z|^{-\beta})\,\mathrm{d}z, \quad k, l \in K_{0m}^\varphi, \tag{9.14}$$

and $B_{kl}^{(m)}$ has a similar expression, just with $k \in K_{0m}^*$ and $l \in K_{0m}^\varphi$, where K_{0m}^* is defined in (5.7). Recalling that $b = 0$ and the quantities $|k - k_{0m}|$ and $|l - k_{0m}|$ are uniformly bounded for $k \in K_{0mc}^\varphi$ and $l \in K_{0m}^\varphi$, obtain (for $b = 0$) that $|B_{kl}^{(m)}| = O(2^{-m\alpha})$, so that the first statement in (9.12) is true due to the fact that matrix $\mathbf{B}^{(m)}$ is finite dimensional.

Now, let $b > 0$ and let us prove (9.10). Observe that $L_\varphi \le z + k_{0m} - k \le U_\varphi$ and $k \in K_{0m}^\varphi$ imply $|z| \le M_\varphi$. Hence, the upper bound in (9.10) follows from (2.2) and (9.13). In order to prove the lower bound in (9.10), note that

$$A_{kk}^{(m)} \ge C_{g1} 2^{-m\alpha} \int_{\Omega_\delta^*} \varphi^2(z)|z - (k_{0m} - k)|^\alpha \exp(-b2m\beta|z - (k_{0m} - k)|^{-\beta})\,\mathrm{d}z$$

where $\Omega_\delta^* = (L_\varphi + \delta_b, (L_\varphi + U_\varphi - 1)/2) \cup ((L_\varphi + U_\varphi + 1)/2, U_\varphi - \delta_b)$ and δ_b is defined in (5.6). Since $|z - (k_{0m} - k)| \ge 1/2$ for $z \in \Omega_\delta^*$, and by (4.2), $(L_\varphi + U_\varphi - 1)/2 - (L_\varphi + \delta_b) \ge 1$ and $(U_\varphi - \delta_b) - (L_\varphi + U_\varphi + 1)/2 \ge 1$, one has

$$A_{kk}^{(m)} \ge C_{g1} 2^{-\alpha(m+1)}\ \exp\left(-b2^{\beta(m+1)}\right)\ \min\left(\int_{L_\varphi + \delta_b}^{(L_\varphi + U_\varphi - 1)/2} \varphi^2(z)\mathrm{d}z,\ \int_{(L_\varphi + U_\varphi + 1)/2}^{U_\varphi - \delta_b} \varphi^2(z)\mathrm{d}z\right),$$

which completes the proof of (9.10).

https://doi.org/10.1051/ps/2012024 Published online by Cambridge University Press

Finally, let us prove (9.11). Note that asymptotic value of the integral in (9.13) is defined by the value at a point which maximizes the argument of the exponential function. Recall that (see, *e.g.*, Dingle [8]) if $F(\lambda) = \int_a^b h(x)\exp(\lambda S(x))\mathrm{d}x$, where $\max S(x)$ is achieved at $x = a$ and $S(x)$ is a decreasing function of x, and if the functions $f(x)$ and $S(x)$ are continuous for $x \in [a, b]$ and infinitely differentiable in the neighborhood of $x = a$ with $S'(a) \neq 0$, then, as $\lambda \to \infty$, $F(\lambda)$ has the following asymptotic expression

$$F(\lambda) \sim \exp(\lambda S(a)) \sum_{k=o}^{\infty} c_k \lambda^{-(k+1)} \quad \text{with} \quad c_k = -D^k(h(x)/S'(x)), \tag{9.15}$$

where D is the differential operator of the form $D = -\frac{1}{S'(x)}\frac{\mathrm{d}}{\mathrm{d}x}$.

It is easy to calculate that $\exp(-b2m\beta|z|^{-\beta})$ takes its maximum value at $z = z_{\max}^{(l,k,\delta)} = \max(u_\delta, v_\delta)$, where $u_\delta^{(l,k)} = \max(L_\varphi + k - k_{0m}, L_\varphi + \delta_b + l - k_{0m})$, $v_\delta^{(l,k)} = \min(U_\varphi + k - k_{0m}, U_\varphi - \delta_b + l - k_{0m})$ and $L_\varphi \leq k - k_{0m}, l - k_{0m} \leq U_\varphi$. In what follows, we shall drop the superscripts whenever it does not cause confusion.

First, consider the case of $k = l$. Then, by examining the cases $k_{0m} - l \leq (L_\varphi + U_\varphi)/2$ and $k_{0m} - l > (L_\varphi + U_\varphi)/2$ separately, one can easily conclude that

$$z_{\max}^{(l,l,\delta)} = \begin{cases} |L_\varphi + \delta_b + l - k_{0m}|, & \text{if} \quad k_{0m} - l > (L_\varphi + U_\varphi)/2, \\ |U_\varphi - \delta_b + l - k_{0m}|, & \text{if} \quad k_{0m} - l \leq (L_\varphi + U_\varphi)/2, \end{cases} \tag{9.16}$$

where, by (4.2), $z_{\max}^{(l,l,\delta)} \geq (U_\varphi - L_\varphi - 2\delta_b)/2 \geq 2 - \delta_b > 1$ in both cases. Hence, since $\varphi(z_{\max}^{(l,l,\delta)}) \neq 0$ by definition of δ_b, formula (9.15) yields

$$\begin{aligned} A_{ll}^{(m)} &\sim C_g(b\beta)^{-1}\varphi^2(z_{\max}^{(l,l,\delta)} + k_{0m} - l)\ |z_{\max}^{(l,l,\delta)}|^\alpha\ 2^{-m(\alpha+\beta)}\ \exp(-b2^{m\beta}|z_{\max}^{(l,l,\delta)}|^{-\beta}) \\ &\geq K_1 2^{-m(\alpha+\beta)}\ \exp(-b2^{m\beta}|z_{\max}^{(l,l,\delta)}|^{-\beta}). \end{aligned} \tag{9.17}$$

If $k \neq l$, then $|k - l| \geq 1$ and one has four cases, depending on whether $k_{0m} - k$ and $k_{0m} - l$ are smaller or greater than $(L_\varphi + U_\varphi)/2$. We shall consider two of those since the other two cases are similar. In what follows, we denote by $z_{\max}^{(k,k,0)}$ the value of $z_{\max}^{(k,k,\delta)}$ obtained if $\delta = \delta_b = 0$.

If $k_{0m} - l \leq (L_\varphi + U_\varphi)/2$ and $k_{0m} - k \leq (L_\varphi + U_\varphi)/2$ then $|z_{\max}^{(l,l,\delta)}| = U_\varphi - \delta_b + l - k_{0m}$, $|z_{\max}^{(k,k,0)}| = U_\varphi + k - k_{0m}$ and, since $\delta_b < 1/2$,

$$|z_{\max}^{(l,k,\delta)}| = \begin{cases} U_\varphi - \delta_b + l - k_{0m}, & \text{if} \quad l > k \\ U_\varphi + k - k_{0m}, & \text{if} \quad l < k. \end{cases}$$

Therefore, taking into account that $|z_{\max}^{(k,k,\delta)}| = |z_{\max}^{(k,k,0)}| - \delta_b$, one derives that

$$\max\left(\left|z_{\max}^{(l,l,\delta)}\right|, \left|z_{\max}^{(k,k,\delta)}\right|\right) - |z_{\max}^{(l,k,\delta)}| \geq 1 - 2\delta_b. \tag{9.18}$$

Now, consider the case when $k_{0m} - l \leq (L_\varphi + U_\varphi)/2$ and $k_{0m} - k > (L_\varphi + U_\varphi)/2$. In this situation, $|z_{\max}^{(l,l,\delta)}| = U_\varphi - \delta_b + l - k_{0m}$, $|z_{\max}^{(k,k,0)}| = k_{0m} - k - L_\varphi$ and $|z_{\max}^{(l,k,\delta)}| = \max(|U_\varphi + k - k_{0m}|, |L_\varphi + \delta_b + l - k_{0m}|)$, so that relation (9.18) is again true. Cases when $k_{0m} - l > (L_\varphi + U_\varphi)/2$ can be examined in a similar manner and it can be shown that (9.18) is valid.

The asymptotic expression for $A_{kl}^{(m)}$ as $m \to \infty$ can be obtained using formula (9.15)

$$A_{kl}^{(m)} \sim C_g K(\varphi, b, \beta, z_{\max})\ \ 2^{-m(\alpha+\beta)}\, 2^{-m\beta r^*}\ \exp(-b2m\beta|z_{\max}|^{-\beta}), \tag{9.19}$$

where $K(\varphi, b, \beta, z_{\max})$ depends on φ, b, β and $z_{\max}$ only and, hence, uniformly bounded, $r^* = 0$ if $z_{\max}$ does not coincide with L_φ or U_φ and $r^* = r_0 + 1$ if it does. (Here, r_0 is the number of continuous derivatives of φ).

We are now ready to complete the proof of the lemma. Recall that

$$|z_{\max}^{(l,k,\delta)}| \leq \min(|z_{\max}^{(l,l,\delta)}|, |z_{\max}^{(k,k,0)}|) \leq \min(|z_{\max}^{(l,l,\delta)}|, |z_{\max}^{(k,k,\delta)}|) + \delta_b,$$

https://doi.org/10.1051/ps/2012024 Published online by Cambridge University Press

and, by (9.18), that

$$|z_{\max}^{(l,k,\delta)}| \le \max(|z_{\max}^{(l,l,\delta)}|, |z_{\max}^{(k,k,\delta)}|) - (1 - 2\delta_b).$$

Since $|z_{\max}^{(l,l,\delta)}| > 2-\delta_b$ and $|z_{\max}^{(k,k,\delta)}| > 2-\delta_b$, an application of Lemma 9.3, with $\delta = \delta_b$, $a = |z_{\max}^{(l,l,\delta)}|$, $b = |z_{\max}^{(k,k,\delta)}|$, $c = |z_{\max}^{(l,k,\delta)}|$ and $M = (L_\varphi + U_\varphi)/2$, completes the proof of the lemma. □

Lemma 9.5. *Let* $\mathbf{A}$ *be the matrix with the entries given by* (5.9) *with* g *satisfying Assumption A and let* $\mathbf{D}$ *be the diagonal matrix* $\mathbf{D} = \sqrt{diag(\mathbf{A})}$. *Denote* $\mathbf{Q} = \mathbf{D}^{-1}\mathbf{A}\mathbf{D}^{-1}$. *Then, for any* $b \ge 0$, *one has* $\|\mathbf{Q}^{-1}\| = O(1)$ *as* $m \to \infty$, *where* $\|\cdot\|$ *denotes the spectral norm. Moreover, if* $b > 0$, *then* $\mathbf{Q}^{-1} = \mathbf{I} + \mathbf{H}$, *where*

$$\|\mathbf{H}\| = O\left(\exp(-0.125\ b\delta_0^2 2^{m\beta})\right), \quad m \to \infty,$$

and δ_0 *is defined in Lemma* 9.4, i.e., $\mathbf{Q}^{-1} = \mathbf{I}\,(1 + o(1))$ *as* $m \to \infty$.

Proof of Lemma 9.5. Note that matrix $\mathbf{Q}$ is an $(U_\varphi - L_\varphi + 1)$-dimensional positive definite matrix with a unit main diagonal and smaller off-diagonal entries, so that, it has a non-asymptotic bounded inverse $\mathbf{Q}^{-1}$ with $\|\mathbf{Q}^{-1}\| = O(1)$.

If $b > 0$, then $Q_{kk} = 1$, so that $\mathbf{Q} = \mathbf{I} + \mathbf{H}$. Here, by Lemma 9.4, $\mathbf{H}$ is a finite dimensional matrix with elements $H_{lk} = O\left(\exp\left\{-0.25\ b\,(U_\varphi + L_\varphi)^{-(\beta+1)}\ 2^{m\beta}\right\}\right)$, as $m \to \infty$. Hence, $\|\mathbf{H}\| \le C_H \exp\left\{-0.25\ b\,(U_\varphi + L_\varphi)^{-(\beta+1)}\ 2^{m\beta}\right\}$ for some $C_H > 0$, so that $\|\mathbf{H}\| \to 0$ as $m \to \infty$. To complete the proof of the lemma, it suffices to note that

$$\mathbf{Q}^{-1} = \mathbf{I} + \sum_{k=1}^{\infty}(-1)^k\mathbf{H}^k, \quad \text{where} \quad \left\|\sum_{k=1}^{\infty}(-1)^k\mathbf{H}^k\right\| \le \sum_{k=1}^{\infty}\|\mathbf{H}\|^k = O\left(\|\mathbf{H}\|\right) \to 0 \quad (m \to \infty).$$ □

9.4. Proofs of the large deviation results

Denote

$$\varrho_n = n^{-1/2}\sqrt{\ln n}. \tag{9.20}$$

In order to prove Lemma 6.1, we need the following three large deviation results (Lems. 9.6–9.8). (We note that the slightly unusual formulation of Lemma 9.6 is due to the fact that we are planning to use it with both $w = \varphi$ and $w = \psi$).

Lemma 9.6. *Let* $b = 0$. *Let* w *be a bounded function with a compact support* $[W_1, W_2]$ *and a unit* L^2*-norm. Denote* $w_{jk}(x) = 2^{j/2}w(2^j x - k)$ *and set*

$$\beta_{jk} = \int w_{jk}(x)f(x)\mathrm{d}x, \qquad \hat\beta_{jk} = n^{-1}\sum_{l=1}^{n}\frac{w_{jk}(x_i)y_i}{g(x_i)},$$

where f *is the unknown response function in model* (1.1). *Let* C_{g1} *be defined in* (2.2), *and let*

$$C_w = [2\max(|W_1|, |W_2|)]]^\alpha, \quad C_\tau = 8C_wC_{g1}^{-1}\max\left(2, 2\|f\|_\infty^2, \|f\|_\infty\|w\|_\infty/3, \|w\|_\infty\right). \tag{9.21}$$

Let m_1 *and* J *be defined by* (5.3), *let* ϱ_n *be defined by* (9.20), *and let*

$$K_{0jc}^w = \left\{k:\ 0 \le k \le 2^j - 1,\ x_0 \notin supp\ w_{jk}\right\}.$$

Then, for $m_1 \le j \le J - 1$, $k \in K_{0jc}^w$ *and* $\tau \ge 1$, *as* $n \to \infty$,

$$\mathbb{P}\left(|\hat\beta_{jk} - \beta_{jk}| > \tau\,\varrho_n\,2^{j\alpha/2}|k - k_{0j}|^{-\alpha/2}\right) = O\left(n^{-\frac{\tau}{C_\tau}}\right). \tag{9.22}$$

https://doi.org/10.1051/ps/2012024 Published online by Cambridge University Press

Proof of Lemma 9.6. The proof of the lemma is based on ideas presented in Chesneau [6]. Observe that

$$\mathbb{P}\left(|\hat{\beta}_{jk} - \beta_{jk}| > \tau\, \varrho_n\, 2^{j\alpha/2}|k - k_{0j}|^{-\alpha/2}\right) \leq P_1 + P_2,$$

where

$$P_1 = \mathbb{P}\left(\left|n^{-1}\ \sum_{i=1}^{n}[g(x_i)]^{-1}\, w_{jk}(x_i)f(x_i) - \beta_{jk}\right| > 0.5\,\tau\, \varrho_n\, 2^{j\alpha/2}|k - k_{0j}|^{-\alpha/2}\right),$$

$$P_2 = \mathbb{P}\left(\left|n^{-1}\ \sum_{i=1}^{n}[g(x_i)]^{-1}\, w_{jk}(x_i)\xi_i\right| > 0.5\,\tau\, \varrho_n\, 2^{j\alpha/2}|k - k_{0j}|^{-\alpha/2}\right).$$

The proof of the statement is now based on Bernstein's inequality, that we recall for completeness,

$$\mathbb{P}\left(\left|n^{-1}\sum_{i=1}^{n}\eta_i\right| > z\right) \leq 2\ \exp\left(-\frac{nz^2}{2(\sigma^2 + \|\eta\|_\infty\, z/3)}\right), \tag{9.23}$$

where $\eta_i,\ i = 1, 2, \ldots, n$, are independent and identically distributed random variables with $\mathbb{E}\eta_i = 0$, $\mathbb{E}\eta_i^2 = \sigma^2$ and $\|\eta_i\| \leq \|\eta\|_\infty < \infty$. First, let us construct an upper bound for P_1. Note that, for $k \in K_{0jc}^w$, and for $i = 1, 2, \ldots, n$, one has, for $x_i \in \operatorname{supp} w_{jk}$,

$$g(x_i) \geq C_w^{-1}\, C_{g1} 2^{-j\alpha}|k - k_{0j}|^\alpha. \tag{9.24}$$

Let $\eta_i = [g(x_i)]^{-1}\, w_{jk}(x_i)f(x_i) - \beta_{jk}$. Then, $\mathbb{E}\eta_i = 0$, and, by (9.24), we derive $\|\eta\|_\infty \leq C_w C_{g1}^{-1}\, |k - k_{0j}|^{-\alpha}\, 2^{j(\alpha+1/2)}\, \|w\|_\infty \|f\|_\infty$, so that

$$\mathbb{E}\eta_i^2 = \int \frac{w_{jk}^2(x)f^2(x)}{g(x)}\,\mathrm{d}x \leq \frac{\|f\|_\infty^2}{C_{g1}} \int_{W_1}^{W_2} \frac{w^2(t)2^{j\alpha}}{|t + k - k_{0j}|^\alpha} dt \leq \frac{\|f\|_\infty^2\, 2^{j\alpha} C_w}{C_{g1}\, |k - k_{0j}|^\alpha}.$$

Now, applying Bernstein's inequality and recalling that $m_1 \leq j \leq J - 1$, $2^{J(\alpha+1)} = n/\ln n$ and $|k - k_{0j}| \geq 1$, we obtain

$$P_1 \leq 2\exp\left(-\frac{C_{g1}\ \tau^2 \ln n}{8C_w\, \|f\|_\infty(\|f\|_\infty + \|w\|_\infty\, \tau/6)}\right).$$

Using the inequality $a/(b + c) \geq \min\,(a/(2b), a/(2c))$, where $a, b, c > 0$, and taking into account that $\tau^2 \geq \tau$ for $\tau \geq 1$, we obtain

$$P_1 \leq 2\exp(-\tau \ln n/D_1) \quad \text{with} \quad D_1 = 8\, C_w C_{g1}^{-1}\ \max(2\|f\|_\infty^2, \|f\|_\infty \|w\|_\infty/3). \tag{9.25}$$

In order to construct an upper bound for P_2, note that, conditionally on $(x_1, x_2, \ldots, x_n)$, one has, for $x_i \in \operatorname{supp} w_{jk}$,

$$n^{-1}\ \sum_{i=1}^{n}(g(x_i))^{-1}\, w_{jk}(x_i)\xi_i \sim \mathcal{N}(0, s_{jk}^2),$$

where, by (9.24) and $\sigma = 1$,

$$s_{jk}^2 = \frac{1}{n^2}\sum_{i=1}^{n}\frac{w_{jk}^2(x_i)}{g^2(x_i)} \leq \frac{C_w 2^{j\alpha}}{C_{g1}|k - k_{0j}|^\alpha\ n^2}\ \sum_{i=1}^{n}\frac{w_{jk}^2(x_i)}{g(x_i)}.$$

Hence, conditionally on $(x_1, x_2, \ldots, x_n)$,

$$\mathbb{P}\left(\left|n^{-1}\ \sum_{i=1}^{n}[g(x_i)]^{-1}\, w_{jk}(x_i)\,\xi_i\right| > \frac{\tau\, \varrho_n\, 2^{j\alpha/2}}{2\,|k - k_{0j}|^{\alpha/2}}\middle| x_1, x_2, \ldots, x_n\right) \leq \exp\left(-\frac{\tau^2 2^{j\alpha}\ \ln n}{8n|k - k_{0j}|^\alpha s_{jk}^2}\right).$$

https://doi.org/10.1051/ps/2012024 Published online by Cambridge University Press

Now, consider the following two sets:

$$\Omega_\upsilon(x_1,x_2,\dots,x_n)=\left\{(x_1,x_2,\dots,x_n):\left|n^{-1}\sum_{i=1}^n\frac{w_{jk}^2(x_i)}{g(x_i)}-1\right|\ge\upsilon\right\},$$

and its complementary, $\Omega_\upsilon^c(x_1,x_2,\dots,x_n)$. Then, $P_2\le P_{21}+P_{22}$, where

$$P_{21}=\mathbb{E}\left[\mathbb{P}\left(\left|n^{-1}\sum_{i=1}^n(g(x_i))^{-1}w_{jk}(x_i)\xi_i\right|>\frac{\tau\,\varrho_n\,2^{j\alpha/2}}{2\,|k-k_{0j}|^{\alpha/2}}\middle|x_1,x_2,\dots,x_n\right)\mathbb{I}(\Omega_\upsilon^c(x_1,x_2,\dots,x_n))\right],$$
$$P_{22}=\mathbb{E}\left[\,\mathbb{I}(\Omega_\upsilon(x_1,x_2,\dots,x_n))\,\right],$$

and $\mathbb{I}(\Omega)$ is the indicator of the set Ω. Since, for $\Omega_\upsilon^c(x_1,x_2,\dots,x_n)$, we have

$$n^{-1}\sum_{i=1}^n[g(x_i)]^{-1}w_{jk}^2(x_i)\le\left(1+\left|n^{-1}\sum_{i=1}^n[g(x_i)]^{-1}w_{jk}^2(x_i)-1\right|\right)\le\upsilon+1,$$

it is easy to check that

$$P_{21}\le\exp(-\tau^2\ln n/D_2)\quad\text{with}\quad D_2=8\,C_wC_{g1}^{-1}\,(\upsilon+1).\tag{9.26}$$

In order to find an upper bound for P_{22}, we apply Bernstein's inequality with $Z_i=[g(x_i)]^{-1}w_{jk}^2(x_i)-1$. Note that $\mathbb{E}Z_i=0$, $\mathbb{E}Z_i^2\le C_wC_{g1}^{-1}\,\|w\|_\infty 2^{j(\alpha+1)}|k-k_{0j}|^{-\alpha}$, $\|Z\|_\infty\le 2C_wC_{g1}^{-1}\,\|w\|_\infty 2^{j(\alpha+1)}|k-k_{0j}|^{-\alpha}$. Application of (9.23) with $z=\upsilon$, yields

$$P_{22}\le 2\exp(-\upsilon^2\ln n/D_3)\quad\text{with}\quad D_3=2\,\|w\|_\infty^2\,C_wC_{g1}^{-1}\,(1+2\upsilon/3).\tag{9.27}$$

Now, set $\upsilon=0.5\,\tau\|w\|_\infty$ and observe that, for $\tau\ge1$, one has

$$4\|w\|_\infty^{-2}\,(1+2\upsilon/3)^{-1}\,\upsilon^2\ge\tau^2/(\upsilon+1)\ge\tau\cdot\min(1/2,\|w\|_\infty^{-1}).$$

To complete the proof, we only need to combine (9.25)–(9.27). □

Lemma 9.7. *Let $b=0$, $C_\varphi=[2\max(|L_\varphi|,|U_\varphi|)]^\alpha$ and C_{g2} be defined in (2.2). Let m be an integer such that $m_1\le m\le J-1$, and let $k\in K_{0m}^\varphi$. Let ϱ_n be defined ny (9.20) and let C_κ be given by (6.9). Then, for $c_l^{(m)}$ and $\hat c_l^{(m)}$ given by (5.11) and (5.14), respectively, and an arbitrary constant $\kappa\ge1$,*

$$\mathbb{P}\left(|\hat c_l^{(m)}-c_l^{(m)}|>\kappa\,\varrho_n\,2^{-\frac{m\alpha}{2}}\right)=O\left(n^{-\frac{\kappa}{C_\kappa}}\right),\quad n\to\infty.\tag{9.28}$$

Proof of Lemma 9.7. The proof is very similar to the Proof of Lemma 9.6, therefore, we shall just provide its outline. Partition the probability in (9.28) into P_1 and P_2 with

$$P_1=\mathbb{P}\left(\left|n^{-1}\sum_{i=1}^n\varphi_{mk}(x_i)f(x_i)-c_k^{(m)}\right|>0.5\,\kappa\,\varrho_n\,2^{-\frac{m\alpha}{2}}\right),$$
$$P_2=\mathbb{P}\left(\left|n^{-1}\sum_{i=1}^n\varphi_{mk}(x_i)\xi_i\right|>0.5\,\kappa\,\varrho_n\,2^{-\frac{m\alpha}{2}}\right).$$

An upper bound for P_1, obtained by applying Bernstein's inequality, is of the form

$$P_1\le 2\exp(-\kappa\ln n/D_4)\quad\text{with}\quad D_4=8\,\|f\|_\infty\,\max(2C_{g2}C_\varphi,\|\varphi\|_\infty/3).\tag{9.29}$$

https://doi.org/10.1051/ps/2012024 Published online by Cambridge University Press

In order to derive an upper bound for P_2, introduce a set

$$\Theta_v(x_1,x_2,\ldots,x_n)=\left\{(x_1,x_2,\ldots,x_n):\left|n^{-1}\sum_{i=1}^n\varphi_{mk}^2(x_i)-\int\varphi_{mk}^2(x)g(x)\mathrm{d}x\right|\geq v2^{-m\alpha}\right\},$$

and its complementary, $\Theta_v^c(x_1,x_2,\ldots,x_n)$. Then, similarly to the Proof of Lemma 9.6, obtain $P_2\leq P_{21}+P_{22}$, where

$$\begin{aligned}P_{21}&=\mathbb{E}\left[\mathbb{P}\left(\left|n^{-1}\sum_{i=1}^n\varphi_{mk}(x_i)\xi_i\right|>0.5\,\kappa\,\varrho_n\,2^{-m\alpha/2}\middle|x_1,x_2,\ldots,x_n\right)\mathbb{I}(\Theta_v^c(x_1,x_2,\ldots,x_n))\right]\\&\leq\exp\left(-\frac{\kappa^2\ln n}{8(v+C_\varphi C_{g2})}\right).\end{aligned}$$

Also, application of (9.23) with $\eta_i=\left(\varphi_{mk}^2(x_i)-\int\varphi_{mk}^2(x)g(x)\mathrm{d}x\right)$, yields

$$P_{22}=\mathbb{E}\left[\mathbb{I}(\Theta_v(x_1,x_2,\ldots,x_n))\right]\leq 2\,\exp\left(-\frac{nv^2\,2^{-m(1+\alpha)}}{2\|\varphi\|_\infty^2(C_\varphi C_{g2}+v/3)}\right).$$

Setting $v=a\kappa$, noting that for any $A,B,C>0$ one has $A/(B+C)\geq\min(A/(2B),A/(2C))$, and recalling that $\kappa\geq 1$ and $n2^{-m(1+\alpha)}\geq\ln n$ by (5.3), we derive

$$P_2\leq 2\exp(-\kappa\ln n/D_5)\quad\text{with}\quad D_5=\max\left(16a,16C_\varphi C_{g2},\frac{4C_\varphi C_{g2}\|\varphi\|_\infty}{a},\frac{4\|\varphi\|_\infty^2}{3a}\right).\tag{9.30}$$

To complete the proof, it suffices to note that $a>0$ is arbitrary. □

Lemma 9.8. *Let $b=0$, let m_0 and $\hat{m}$ be given by (5.18) and (6.2), respectively. Consider the non-asymptotic finite dimension matrices $\mathbf{A}^*$ and $\mathbf{B}^*$ with elements*

$$A_{kl}^*=\int\varphi(z+k_{0m}-k)\varphi(z+k_{0m}-l)|z|^\alpha\mathrm{d}z,\quad k,l\in K_{0m}^\varphi,\tag{9.31}$$

$$B_{lk}^*=\int\varphi(z+k_{0m}-k)\varphi(z+k_{0m}-l)|z|^\alpha\mathrm{d}z,\quad l\in K_{0m}^\varphi,\ k\in K_{0m}^*.\tag{9.32}$$

Let $\hat{\mathbf{u}}^{(m)}$ be the solution of the system of equations (5.15). Let $C_{\lambda1}$ and $C_{\lambda2}$ be defined by (6.5) and let C_u be defined by (6.7). If $\lambda\geq\max(C_{\lambda1},C_{\lambda2})$, then, as $n\to\infty$,

$$\mathbb{P}\left(\|\hat{\mathbf{u}}^{(m)}-\mathbb{E}\hat{\mathbf{u}}^{(m)}\|>\lambda\,\varrho_n\,2^{m\alpha/2}\right)=O\left(n^{-\frac{2\lambda}{C_u}}\right).\tag{9.33}$$

Proof of Lemma 9.8. Observe that for any m, by (5.15), one has

$$\|\hat{\mathbf{u}}^{(m)}-\mathbb{E}\hat{\mathbf{u}}^{(m)}\|\leq\|(\mathbf{A}^{(m)})^{-1}(\hat{\mathbf{c}}^{(m)}-\mathbf{c}^{(m)})\|+\|(\mathbf{A}^{(m)})^{-1}\mathbf{B}^{(m)}(\hat{\mathbf{v}}^{(m)}-\mathbf{v}^{(m)})\|,$$

so that

$$\begin{aligned}&\mathbb{P}\left(\|\hat{\mathbf{u}}^{(m)}-\mathbb{E}\hat{\mathbf{u}}^{(m)}\|>\lambda\,\varrho_n\,2^{m\alpha/2}\right)\leq\mathbb{P}\left(\|(\mathbf{A}^{(m)})^{-1}(\hat{\mathbf{c}}^{(m)}-\mathbf{c}^{(m)})\|>0.5\,\lambda\,\varrho_n\,2^{m\alpha/2}\right)\\&+\mathbb{P}\left(\|(\mathbf{A}^{(m)})^{-1}\mathbf{B}^{(m)}(\hat{\mathbf{v}}^{(m)}-\mathbf{v}^{(m)})\|>0.5\,\lambda\,\varrho_n\,2^{m\alpha/2}\right)\equiv P_1+P_2.\end{aligned}$$

Now note that, by assumption (2.1) and the dominated convergence theorem, as $n\to\infty$, one has $\mathbf{A}^{(m)}=C_g2^{-m\alpha}\mathbf{A}^*(1+o(1))$ and $\mathbf{B}^{(m)}=C_g2^{-m\alpha}\mathbf{B}^*(1+o(1))$, where the matrices $\mathbf{A}^*$ and $\mathbf{B}^*$, defined in (9.31)

https://doi.org/10.1051/ps/2012024 Published online by Cambridge University Press

and (9.32), are independent of m, since the sets K^{φ}_{0m} and K^{*}_{0m} are defined in terms of $k-k_{0m}$ and $l-k_{0m}$. Therefore, $\|(\mathbf{A}^{(m)})^{-1}\| = C_g^{-1}\,2^{m\alpha}\,\|(\mathbf{A}^{*})^{-1}\|(1+o(1))$ and $\|(\mathbf{A}^{(m)})^{-1}\mathbf{B}^{(m)}\| = \|(\mathbf{A}^{*})^{-1}\,\mathbf{B}^{*}\|(1+o(1))$. Hence, setting $\kappa = C_{\lambda 1}^{-1}\,\lambda$ in Lemma 9.7, where $C_{\lambda 1}$ is defined in (6.5), and taking into account that the set K^{φ}_{0m} contains no more than $U_\varphi - L_\varphi + 1$ indices, we obtain

$$
\begin{aligned}
P_1 &\le \mathbb{P}\left(\|\hat{\mathbf{c}}^{(m)} - \mathbf{c}^{(m)}\| > \frac{C_{g2}\,\lambda\,\varrho_n\,2^{-m\alpha/2}}{2\,\|(\mathbf{A}^{*})^{-1}\|}\right) \le \sum_{k\in K^{\varphi}_{0m}} \mathbb{P}\left(|\hat{c}_k^{(m)} - c_k^{(m)}| > \frac{C_{g2}\,\lambda\,\varrho_n\,2^{-m\alpha/2}}{2\,\sqrt{U_\varphi - L_\varphi + 1}\,\|(\mathbf{A}^{*})^{-1}\|}\right) \\
&= \sum_{k\in K^{\varphi}_{0m}} \mathbb{P}\left(|\hat{c}_k^{(m)} - c_k^{(m)}| > 2C_{\lambda 1}^{-1}\,\lambda\,\varrho_n\,2^{-m\alpha/2}\right) = O\left(n^{-2(C_\kappa\,C_{\lambda 1})^{-1}\,\lambda}\right).
\end{aligned}
$$

Similarly, using Lemma 9.6 with $w=\varphi$ and C_τ given by (6.8), and recalling the definitions of $\hat{\mathbf{v}}^{(m)}$ and $\mathbf{v}^{(m)}$, one can derive an upper bound for P_2 as

$$
\begin{aligned}
P_2 &\le \mathbb{P}\left(\|\hat{\mathbf{v}}^{(m)} - \mathbf{v}^{(m)}\| > \frac{\lambda\,\varrho_n\,2^{m\alpha/2}}{2\,\|(\mathbf{A}^{*})^{-1}\,\mathbf{B}^{*}\|}\right) \le \sum_{k\in K^{*}_{0m}} \mathbb{P}\left(|\hat{a}_{mk} - a_{mk}| > \frac{\lambda\,\varrho_n\,2^{m\alpha/2}}{2\,\sqrt{U_\varphi - L_\varphi + 1}\,\|(\mathbf{A}^{*})^{-1}\,\mathbf{B}^{*}\|}\right) \\
&= O\left(n^{-2(C_\tau\,C_{\lambda 2})^{-1}\,\lambda}\right),
\end{aligned}
$$

which completes the proof of the lemma. □

Proof of Lemma 6.1. Note that by definition of $\hat{m}$, whenever $\hat{m} > m_0$, there exists $j > m_0$ such that $\|(\hat{f}_{m_0} - \hat{f}_j)\mathbb{I}(\Xi_{m_0})\|^2 > \lambda^2\,2^{j\alpha}\,\rho_n^2$, where ϱ_n is defined in (9.20). Therefore,

$$
\mathbb{P}(\hat{m} > m_0) \le \sum_{j=m_0}^{J-1} \mathcal{P}_j \quad \text{with} \quad \mathcal{P}_j = \mathbb{P}\left(\|(\hat{f}_{m_0} - \hat{f}_j)\mathbb{I}(\Xi_{m_0})\|^2 > \lambda^2\,2^{j\alpha}\,\rho_n^2\right). \tag{9.34}
$$

Observe that since

$$
\begin{aligned}
\|(\hat{f}_{m_0} - \hat{f}_j)\mathbb{I}(\Xi_{m_0})\| &\le \|(\hat{f}_{0,j} - f_{0,j})\mathbb{I}(\Xi_{m_0})\| + \|(\hat{f}_{c,j} - f_{c,j})\mathbb{I}(\Xi_{m_0})\| \\
&\quad + \|(\hat{f}_{0,m_0} - f_{0,m_0})\mathbb{I}(\Xi_{m_0})\| + \|(\hat{f}_{c,m_0} - f_{c,m_0})\mathbb{I}(\Xi_{m_0})\|,
\end{aligned}
$$

one has the following upper bound for $\mathcal{P}_j$ defined in (9.34):

$$
\mathcal{P}_j \le \mathcal{P}_{0,j,m_0} + \mathcal{P}_{0,j,j} + \mathcal{P}_{c,j,m_0} + \mathcal{P}_{c,j,j},
$$

where, for any $m_0 \le m \le j$,

$$
\begin{aligned}
\mathcal{P}_{0,j,m} &= \mathbb{P}\left(\|(\hat{f}_{0,m} - f_{0,m})\mathbb{I}(\Xi_{m_0})\| > 0.25\,\lambda\,2^{j\alpha/2}\,\rho_n\right), \\
\mathcal{P}_{c,j,m} &= \mathbb{P}\left(\|(\hat{f}_{c,m} - f_{c,m})\mathbb{I}(\Xi_{m_0})\| > 0.25\,\lambda\,2^{j\alpha/2}\,\rho_n\right).
\end{aligned}
$$

Since $\mathrm{supp}(f_{0,m}) \subseteq \Xi_m \in \Xi_{m_0}$ for $m \ge m_0$, one has

$$
\begin{aligned}
\|(\hat{f}_{0,m} - f_{0,m})\mathbb{I}(\Xi_{m_0})\|^2 &= \|(\hat{f}_{0,m} - f_{0,m})\mathbb{I}(\Xi_m)\|^2 = \|\hat{f}_{0,m} - f_{0,m}\|^2 \\
&\le \|\hat{\mathbf{u}}^{(m)} - \mathbf{u}^{(m)}\|^2 + 2(U_\varphi - L_\varphi + 1)A^2 2^{-2ms'}.
\end{aligned} \tag{9.35}
$$

Hence, by (9.35) and Lemma 9.8, since $m_0 \le m \le j$, one derives

$$
\mathcal{P}_{0,j,m} \le \mathbb{P}\left(\|\hat{\mathbf{u}}^{(m)} - \mathbf{u}^{(m)}\| > 0.25\,\lambda\,2^{j\alpha/2}\,\rho_n - A\sqrt{2(U_\varphi - L_\varphi + 1)}2^{-js'}\right) = O\left(n^{-\frac{\lambda}{2C_u}}\right).
$$

https://doi.org/10.1051/ps/2012024 Published online by Cambridge University Press

Now, let us consider the second term, $\mathcal{P}_{c,j,m}$. Note that $\mathrm{supp}(\varphi_{mk})$ and Ξ_{m_0} have non-empty intersection if and only if $k \in \tilde{K}_{m,m_0}$, where

$$\tilde{K}_{m,m_0} = \left\{k:\ 2^{m-m_0}[\min(L_\varphi, L_\psi) - U_\varphi] - U_\varphi < k - k_{0m} < 2^{m-m_0}[\max(U_\varphi, U_\psi) - L_\varphi] - L_\varphi\right\}.$$

Hence, for $m \geq m_0$,

$$\|(\hat{f}_{c,m} - f_{c,m})\mathbb{I}(\Xi_{m_0})\|^2 \leq \|\hat{\mathbf{v}}^{(m)} - \mathbf{v}^{(m)}\|^2 + \sum_{j'=m}^{J-1} \sum_{k\in\tilde{K}_{j',m_0}} (\hat{b}_{j'k} - b_{j'k})^2 + \sum_{j=J}^{\infty} \sum_{k\in\tilde{K}_{j,m_0}} b_{jk}^2.$$

Here, by (9.57), we have

$$\sum_{j=J}^{\infty} \sum_{k\in\tilde{K}_{j,m_0}} b_{jk}^2 \leq A^2 2^{-2Js^*} = O\left(n^{-\frac{2s'}{2s'+\alpha}} (\ln n)^{\frac{2s'}{2s'+\alpha}}\right),$$

where s^* is defined in (9.56). Also,

$$(\hat{b}_{j'k} - b_{j'k})^2 \leq (\hat{b}_{j'k} - b_{j'k})^2 \mathbb{I}(|\hat{b}_{j'k} - b_{j'k}| > 0.5d2^{j\alpha/2}|k - k_{0j'}|^{-\alpha/2}) + b_{j'k}^2$$

since $\mathbb{I}(|\hat{b}_{j'k}| > d2^{j\alpha/2}|k - k_{0j'}|^{-\alpha/2}) \leq \mathbb{I}(|b_{j'k}| > 0.5d2^{j\alpha/2}|k - k_{0j'}|^{-\alpha/2}) + \mathbb{I}(|\hat{b}_{j'k} - b_{j'k}| > 0.5d2^{j\alpha/2}|k - k_{0j'}|^{-\alpha/2})$ and, for $j \geq m_0$ and n large enough, $\mathbb{I}(|b_{j'k}| > 0.5d2^{j\alpha/2}|k - k_{0j'}|^{-\alpha/2}) = 0$. Denote $C_{LU} = \max(|\min(L_\varphi, L_\psi) - 2U_\varphi|, |\max(U_\varphi, U_\psi) - 2L_\varphi|)$ and observe that $\tilde{K}_{j',m_0} \subset \left\{k:\ |k - k_{0j'}| < 2^{j'-m_0}C_{LU}\right\}$. Hence, using the Cauchy–Schwarz inequality and (9.56), one obtains

$$\sum_{j'=m}^{J-1} \sum_{k\in\tilde{K}_{j',m_0}} b_{j'k}^2 \leq A^2(2C_{LU})^{(1-2/p)_+}\ 2^{-2m_0s'}2^{-2s^*(m-m_0)}.$$

Combining all inequalities above, we derive that for any $m \geq m_0$,

$$\begin{aligned}\|(\hat{f}_{c,m} - f_{c,m})\mathbb{I}(\Xi_{m_0})\|^2 \leq{}& \|\hat{\mathbf{v}}^{(m)} - \mathbf{v}^{(m)}\|^2 + A^2 2^{-2Js^*} + A^2(2C_{LU})^{(1-2/p)_+}\ 2^{-2m_0s'}2^{-2s^*(m-m_0)} \\ &+ \sum_{j'=m}^{J-1} \sum_{k\in\tilde{K}_{j',m_0}} (\hat{b}_{j'k} - b_{j'k})^2 \mathbb{I}(|\hat{b}_{j'k} - b_{j'k}| > 0.5d2^{j\alpha/2}|k - k_{0j'}|^{-\alpha/2}).\end{aligned}$$

Now, by Lemma 9.6 with $w = \varphi$ and $w = \psi$, obtain

$$\begin{aligned}\mathcal{P}_{c,j,m} &\leq \mathbb{P}\left(\|\hat{\mathbf{v}}^{(m)} - \mathbf{v}^{(m)}\| > 0.25\,\lambda\, 2^{j\alpha/2}\,\rho_n - A^2 2^{-2Js^*} + A^2(2C_{LU})^{(1-2/p)_+}\ 2^{-2m_0s'}2^{-2s^*(m-m_0)}\right) \\ &\quad + \sum_{j'=m}^{J-1} \sum_{k\in\tilde{K}_{j',m_0}} \mathbb{P}(|\hat{b}_{j'k} - b_{j'k}| > 0.5d2^{j\alpha/2}|k - k_{0j'}|^{-\alpha/2}) \\ &= O\left(n^{-(C_\tau C_{\lambda 0})^{-1}\lambda}\right) + O\left(n^{\frac{1}{\alpha+1} - \frac{d}{2C_d}}\right),\end{aligned}$$

which completes the proof. □

9.5. Proofs of the statements in Section 5: the zero-affected part of the wavelet thresholding estimator

Proof of Lemma 5.1. Note that

$$\Delta_1 = \|\mathbb{E}\hat{f}_0^{(m)} - f_0^{(m)}\|^2 = \sum_{j=m}^{\infty} \sum_{k\in K_{0m}^{\varphi}} b_{jk}^2, \quad \Delta_2 = \mathbb{E}\|\hat{f}_0^{(m)} - \mathbb{E}\hat{f}_0^{(m)}\|^2 = \sum_{k\in K_{0m}^{\varphi}} \mathbb{E}(\hat{a}_{mk} - a_{mk})^2,$$

https://doi.org/10.1051/ps/2012024 Published online by Cambridge University Press

where $\hat{a}_{mk} = \hat{u}_k^{(m)}$ for $k \in K_{0m}^{\varphi}$. From the characterization (3.1) of Besov spaces, it follows that, for any k, one has $b_{jk}^2 \le A 2^{-2js'}$, and, therefore, since the number of indices in the set K_{0m}^{φ} is finite,

$$\Delta_1 = O\left(\sum_{j=m}^{\infty} 2^{-2js'}\right) = O\left(2^{-2ms'}\right). \tag{9.36}$$

Now, consider Δ_2. Let, as in Lemma 9.5, $\mathbf{A}^{(m)}$ be the matrix with the entries given by (5.9), $\mathbf{D}^{(m)} = \sqrt{\mathrm{diag}(\mathbf{A}^{(m)})}$ and $\mathbf{Q}^{(m)} = (\mathbf{D}^{(m)})^{-1}\mathbf{A}^{(m)}(\mathbf{D}^{(m)})^{-1}$. In the following proof, for the sake of clarity, we shall suppress the index m. Rewrite the systems of equations (5.8) and (5.15), respectively, as

$$\mathbf{Q}\,\mathbf{D}\,\mathbf{u} = \mathbf{D}^{-1}\,\mathbf{c} + \mathbf{D}^{-1}\,\boldsymbol{\varepsilon} - \mathbf{D}^{-1}\,\mathbf{B}\mathbf{v}, \qquad \mathbf{Q}\,\mathbf{D}\,\hat{\mathbf{u}} = \mathbf{D}^{-1}\,\hat{\mathbf{c}} - \mathbf{D}^{-1}\,\mathbf{B}\hat{\mathbf{v}}, \tag{9.37}$$

so that

$$\hat{\mathbf{u}} - \mathbf{u} = \mathbf{D}^{-1}\mathbf{Q}^{-1}\mathbf{D}^{-1}(\hat{\mathbf{c}} - \mathbf{c}) - \mathbf{D}^{-1}\mathbf{Q}^{-1}\mathbf{D}^{-1}\mathbf{B}(\hat{\mathbf{v}} - \mathbf{v}) + \mathbf{D}^{-1}\mathbf{Q}^{-1}\mathbf{D}^{-1}\boldsymbol{\varepsilon}. \tag{9.38}$$

Therefore,

$$\Delta_2 = \mathbb{E}\|\hat{\mathbf{u}} - \mathbf{u}\|^2 = O\left(\Delta_{21} + \Delta_{22} + \Delta_{23}\right), \tag{9.39}$$

with

$$\Delta_{21} = \mathbb{E}\|\mathbf{D}^{-1}\mathbf{Q}^{-1}\mathbf{D}^{-1}(\hat{\mathbf{c}} - \mathbf{c})\|^2, \quad \Delta_{22} = \mathbb{E}\|\mathbf{D}^{-1}\mathbf{Q}^{-1}\mathbf{D}^{-1}\mathbf{B}(\hat{\mathbf{v}} - \mathbf{v})\|^2, \quad \Delta_{23} = \|\mathbf{D}^{-1}\mathbf{Q}^{-1}\mathbf{D}^{-1}\boldsymbol{\varepsilon}\|^2. \tag{9.40}$$

By Lemma 9.4, one has

$$\mathbf{D}_{ii} \ge C 2^{-m\alpha/2} \exp(-0.5\, b 2^{\beta(m+1)}),$$

and since $\mathbf{D}$ is the finite-dimensional diagonal matrix, the latter implies

$$\|\mathbf{D}^{-1}\| = O\left(2^{m\alpha/2} \exp(0.5\, b 2^{\beta(m+1)})\right). \tag{9.41}$$

Therefore, since the set K_{0m}^{φ} is finite, by Lemma 9.4, one has

$$\mathbb{E}\|\mathbf{D}^{-1}(\hat{\mathbf{c}} - \mathbf{c})\|^2 = \sum_{k \in K_{0m}^{\varphi}} \mathrm{Var}(\hat{c}_k^{(m)})/A_{kk}^{(m)} = O\left(n^{-1}\right),$$

so that we derive

$$\Delta_{21} = O\left(\|\mathbf{D}^{-1}\|^2 \|\mathbf{Q}^{-1}\|^2\ \mathbb{E}\|\mathbf{D}^{-1}(\hat{\mathbf{c}} - \mathbf{c})\|^2\right) = O\left(n^{-1}\ 2^{m\alpha}\ \exp(b 2^{\beta(m+1)})\right). \tag{9.42}$$

In order to derive an upper bound for Δ_{22}, note that from (9.10), (9.40) and considerations above, it follows that

$$\Delta_{22} = O\left(\|\mathbf{D}^{-1}\|^4 \|\mathbf{Q}^{-1}\|^2\ \mathbb{E}\|\mathbf{B}(\hat{\mathbf{v}} - \mathbf{v})\|^2\right) = O\left(2^{2m\alpha} \exp(2b\, 2^{\beta(m+1)})\ \mathbb{E}\|\mathbf{B}(\hat{\mathbf{v}} - \mathbf{v})\|^2\right).$$

Since $\exp(-b 2^{m\beta}|z|^{-\beta})$ is an increasing function of $|z|$ and, for $L_\varphi \le z + k_{0m} - k \le U_\varphi$ and $k \in K_{0m}^*$, one has $|z| \le 2(U_\varphi - L_\varphi)$, for $k \in K_{0m}^*$, we derive

$$\begin{aligned} C_{kk} = \int \varphi_{mk}^2(x) g(x) \mathbb{I}(2^m x - l \in \Omega_\delta) \mathrm{d}x &= O\left(2^{-m\alpha} \int \varphi^2(z + k_{0m} - k)\, |z|^\alpha \exp\left\{-b 2^{m\beta}|z|^{-\beta}\right\} \mathrm{d}z\right) \\ &= O\left(2^{-m\alpha}\ \exp\left\{-b 2^{m(\beta-1)}(U_\varphi - L_\varphi)^{-\beta}\right\}\right). \end{aligned}$$

https://doi.org/10.1051/ps/2012024 Published online by Cambridge University Press

Hence, since the sets K^{φ}_{0m} and K^{*}_{0m} are finite, by definition of vector $\hat{\mathbf{v}}$, Lemmas 9.2 and 9.4 and the Cauchy–Schwarz inequality, we obtain

$$\begin{aligned}\mathbb{E}\|\mathbf{B}(\hat{\mathbf{v}}-\mathbf{v})\|^2 &= \sum_{h\in K^{\varphi}_{0m}}\sum_{k,l\in K^{*}_{0m}} B_{hk}B_{hl}\mathrm{Cov}(\hat{a}_k,\hat{a}_l) \le n^{-1}\sum_{h\in K^{\varphi}_{0m}}\sum_{k,l\in K^{*}_{0m}} J_{mkl}A_{hh}\sqrt{C_{kk}C_{ll}} \\ &= O\left(n^{-1}2^{-m\alpha}\exp\left\{b2^{m\beta}-b2^{m(\beta-1)}(U_\varphi-L_\varphi)^{-\beta}-b2^{m\beta}\,M_\varphi^{-\beta}\right\}\right),\end{aligned}$$

where M_φ is defined in Lemma 9.4. Since $U_\varphi - L_\varphi \ge 4$, we finally obtain

$$\Delta_{22} = O\left(n^{-1}\ 2^{m\alpha}\ \exp(b\,2^{m\beta}[2^{\beta+1}+1])\right). \tag{9.43}$$

Now, for the function $\varepsilon_m(x)$ defined in (5.4), one has

$$\Delta_{23} = O\left(\|\mathbf{D}^{-1}\|^2\,\|\mathbf{Q}^{-1}\|^2\,\|\mathbf{D}^{-1}\boldsymbol{\varepsilon}\|^2\right), \tag{9.44}$$

where

$$\|\mathbf{D}^{-1}\boldsymbol{\varepsilon}\|^2 = \sum_{k\in K^{\varphi}_{0m}} A_{kk}^{-2}\left[\int \varepsilon_m(x)\,\varphi^{*}_{mk}(x)\,g(x)\mathrm{d}x\right]^2. \tag{9.45}$$

If $b=0$, by Cauchy–Schwarz inequality, one obtains $\|\mathbf{D}^{-1}\boldsymbol{\varepsilon}\|^2 \le \sum_{k\in K^{\varphi}_{0m}} A_{kk}^{-2}\|\varepsilon_{m0}\|^2\,\|\varphi_{mk}\,g\|^2$, where $\varepsilon_{m0}(x) = \varepsilon_m(x)\mathbb{I}(|x-x_0|\le C2^{-m})$. By calculations similar to Proof of Lemma 9.4 in the case of $b=0$, one can show that $\|\varphi_{mk}\,g\|^2 = \int \varphi^2_{mk}(x)g(x)\mathrm{d}x = O\left(2^{-2m\alpha}\right)$. Also, since $b_{jk}^2 \le A2^{-2js'}$, one has

$$\begin{aligned}\|\varepsilon_{m0}\|^2 &= O\left(\sum_{j=m}^{\infty}\sum_{|k-k_{0j}|\le C2^{j-m}}\right) \\ &= O\left(\sum_{j=m}^{\infty} 2^{-2js'}2^{(j-m)(1-2/p)}\right) = O\left(2^{-2ms'}\right).\end{aligned}$$

Recalling (9.10), we obtain in the case of $b=0$,

$$\Delta_{23} = O\left(2^{-2ms'}\right). \tag{9.46}$$

Now, let us consider the case of $b>0$. Denote $\varphi^{*}_{mk}(x) = \varphi_{mk}(x)\mathbb{I}(2^m x - k \in \Omega_\delta)$, $I_{jmkl} = \int \varphi^{*}_{mk}(x)\psi_{jl}(x)g(x)\mathrm{d}x$ and let

$$z_{\max}(\varphi^{*}_{mk},\psi_{jl}) = \arg\max_x[\varphi^{*}_{mk}(x)\psi_{jl}(x)g(x)]. \tag{9.47}$$

Observe that since $\varphi^{*}_{mk}(z_{\max}) \ne 0$, we have $I_{jmkl}/A^{(m)}_{kk} = O(1)$. Consider the collection of indices

$$\mathcal{L}_{mjk} = \left\{l:\ 0\le l\le 2^j-1,\ \mathrm{supp}(\varphi^{*}_{mk})\cap\mathrm{supp}(\psi_{jl})\ne\emptyset\right\}.$$

It is easy to see that $\mathcal{L}_{mjk} \subseteq [2^{j-m}(L_\varphi+\delta_b+k)-U_\psi,\ 2^{j-m}(U_\varphi+\delta_b+k)-L_\psi]$, so, for each k, there are $O(2^{j-m})$ terms such that $l\in\mathcal{L}_{mjk}$. Note that $|z_{\max}(\varphi^{*}_{mk},\psi_{jl})| \le |z_{\max}(\varphi^{*}_{mk})|$ and, for each k, there is only finite number of terms such that $|z_{\max}((\varphi^{*}_{mk},\psi_{jl})| = |z_{\max}((\varphi^{*}_{mk})|$. Indeed, straightforward calculation shows that

$$z_{\max}(\varphi^{*}_{mk},\psi_{jl}) = \min[(U_\varphi-\delta_b+k-k_{0m}), 2^{m-j}(U_\psi+l-k_{0j})] \quad \text{if} \quad k_{0m}-k<0.5\,(U_\varphi+L_\varphi),$$

and

$$z_{\max}(\varphi^{*}_{mk},\psi_{jl}) = \max[(L_\varphi+\delta_b+k-k_{0m}), 2^{m-j}(L_\psi+l-k_{0j})] \quad \text{if} \quad k_{0m}-k\ge 0.5\,(U_\varphi+L_\varphi).$$

https://doi.org/10.1051/ps/2012024 Published online by Cambridge University Press

Hence, $|z_{\max}(\varphi^*_{mk},\psi_{jl})| = |z_{\max}(\varphi^*_{mk})|$ if $l \geq 2^{j-m}(U_\varphi - \delta_b + k) - U_\psi$ or $l \leq 2^{j-m}(L_\varphi + \delta_b + k) - L_\psi$. Since we also need $l \in \mathcal{L}_{mjk}$, we obtain that $|z_{\max}((\varphi^*_{mk},\psi_{jl})| = |z_{\max}((\varphi^*_{mk})|$ if $l \in \mathcal{L}^*_{mjk}$ where

$$\mathcal{L}^*_{mjk} \subseteq [2^{j-m}(L_\varphi + \delta_b + k) - U_\psi, 2^{j-m}(L_\varphi + \delta_b + k) - U_\psi] \cup [2^{j-m}(U_\varphi + \delta_b + k) - U_\psi, 2^{j-m}(U_\varphi + \delta_b + k) - L_\psi]$$

and, thus, $\mathcal{L}^*_{mjk}$ contains at most $2(U_\psi - L_\psi)$ values of l for each k. If $l \in \mathcal{L}_{mjk} \setminus \mathcal{L}^*_{mjk} = \mathcal{L}^c_{mjk}$, then

$$2^{j-m}(L_\varphi + \delta_b + k) - L_\psi < l < 2^{j-m}(U_\varphi - \delta_b + k) - U_\psi. \tag{9.48}$$

Then, by (9.15),

$$\begin{aligned} I_{jmkl} &\sim 2^{-\frac{j-m}{2}} \int \varphi^*(2^{m-j}t + k_{0m} - k)\,\psi(t + k_{0j} - l)2^{-j\alpha}\,|t|^\alpha \exp(-b|t|^{-\beta}2^{j\beta})dt \\ &= O\left(2^{(m-j)/2}2^{-j\alpha}2^{-j\beta}2^{(j-m)\alpha}\exp\left\{-b2^{j\beta}|t^{(k,l)}_{\max}|^{-\beta}\right\}\right), \end{aligned}$$

where

$$t^{(k,l)}_{\max} = U_\psi + l - k_{0j} \quad \text{if} \quad k_{0m} - k < (U_\varphi + L_\varphi)/2$$

and

$$t^{(k,l)}_{\max} = L_\psi + l - k_{0j} \quad \text{if} \quad k_{0m} - k \geq (U_\varphi + L_\varphi)/2.$$

Using formula (9.17), we derive that

$$(A^{(m)}_{kk})^{-1}\,|I_{jmkl}| = O\left(2^{(j-m)(\beta+1/2)}\exp\left\{-b2^{j\beta}\left[|t^{(k,l)}_{\max}|^{-\beta} - 2^{-(j-m)\beta}|z^{(k,k,\delta)}_{\max}|^{-\beta}\right]\right\}\right), \tag{9.49}$$

where $z^{(k,k,\delta)}_{\max}$ is defined in (9.16).

Denote $h_{jmkl} = |t^{(k,l)}_{\max}| - 2^{(j-m)}|z^{(k,k,\delta)}_{\max}|$ and observe that

$$h_{jmkl} = 2^{(j-m)}(U_\varphi - \delta_b + k) - U_\psi - l \quad \text{if} \quad k_{0m} - k < (U_\varphi + L_\varphi)/2,$$

and

$$h_{jmkl} = l - 2^{(j-m)}(L_\varphi + \delta_b + k) + L_\psi \quad \text{if} \quad k_{0m} - k \geq (U_\varphi + L_\varphi)/2.$$

Comparing the latter formulae with definition of $\mathcal{L}^c_{mjk}$, we derive that, for $l \in \mathcal{L}^c_{mjk}$, $0 < h_{jmkl} < C_h 2^{j-m}$ for every value of k, where $C_h > 0$ is a constant which depends only on the choice of the wavelet basis. Now, for any $0 < x < y$ and $\beta > 0$,

$$x^{-\beta} - y^{-\beta} \geq \beta(y - x)y^{-(\beta+1)}.$$

Applying the above inequality with $x = |t^{(k,l)}_{\max}|$ and $y = 2^{(j-m)}|z^{(k,k,\delta)}_{\max}|$, we obtain that

$$|t^{(k,l)}_{\max}|^{-\beta} - 2^{-(j-m)\beta}|z^{(k,k,\delta)}_{\max}|^{-\beta} \geq \beta h_{jmkl}2^{-(j-m)(\beta+1)}|z^{(k,k,\delta)}_{\max}|^{-(\beta+1)},$$

and, thus, for $l \in \mathcal{L}^c_{mjk}$, we have

$$(A^{(m)}_{kk})^{-1}\,|I_{jmkl}| = O\left(2^{(j-m)(\beta+1/2)}\exp\left\{-b\beta|z^{(k,k,\delta)}_{\max}|^{-(\beta+1)}2^{m\beta}2^{-(j-m)}h_{jmkl}\right\}\right). \tag{9.50}$$

Now, it follows from (9.44) that $\Delta_{23} = \Delta_{231} + \Delta_{232}$, where

$$\Delta_{231} = O\left(\sum_{k\in K^\varphi_{0m}}\left[\sum_{j=m}^{\infty}\sum_{l\in\mathcal{L}^*_{mjk}}(A^{(m)}_{kk})^{-1}\,|I_{jmkl}||b_{jl}|\right]^2\right),$$

$$\Delta_{232} = O\left(\sum_{k\in K^\varphi_{0m}}\left[\sum_{j=m}^{\infty}\sum_{l\in\mathcal{L}^c_{mjk}}(A^{(m)}_{kk})^{-1}\,|I_{jmkl}||b_{jl}|\right]^2\right).$$

https://doi.org/10.1051/ps/2012024 Published online by Cambridge University Press

Using the facts thta the set K_{0m}^{φ} is finite, $|b_{jl}| = O(2^{-js'}$ and $(A_{kk}^{(m)})^{-1}\,|I_{jmkl}| = O(1)$, we derive that, as $m \to \infty$,

$$\Delta_{231} = O\left(2^{-2ms'}\right). \tag{9.51}$$

For Δ_{232}, using (9.50) and taking into account that h_{jmkl} changes by unit increments, we obtain

$$\begin{aligned}\Delta_{232} &= O\left(\sum_{k\in K_{0m}^{\varphi}}\left[\sum_{j=m}^{\infty}\sum_{h_{jmkl}=0}^{C_h 2^{j-m}} 2^{-js'}2^{(j-m)(\beta+1/2)}\exp\left\{-b\beta|z_{\max}^{(k,k,\delta)}|^{-(\beta+1)}2^{m\beta}2^{-(j-m)}h_{jmkl}\right\}\right]^2\right)\\ &= O\left(-2ms'\left[\sum_{j=m}^{\infty}2^{-(j-m)(\beta-1/2+s')}\right]^2\right) = O\left(2^{-2ms'}\right)\end{aligned} \tag{9.52}$$

since $s' \geq 1/2$ and $\beta > 0$. Finally, combining expressions (9.39), (9.42), (9.43), (9.51) and (9.52), we obtain

$$\Delta_2 = O\left(n^{-1}\ 2^{m\alpha}\ \exp(2b\,2^{\beta(m+1)}) + 2^{-2ms'}\right). \tag{9.53}$$

To complete the proof of (5.17), set $m = m_0$, where m_0 is defined in (5.18) and combine (9.53) with (9.36)–(9.43). □

Now, we need to show that $\mathbb{E}\|\hat{f}_0^{(m)} - \mathbb{E}\hat{f}_0^{(m)}\|^4 = o(1)$. Note that it follows from (9.37)–(9.40) that $\Delta^* = O\left(\Delta_1^* + \Delta_2^* + \Delta_3^*\right)$ where, similarly to the case of squared difference,

$$\begin{aligned}\Delta_1^* &= O\left(\|\mathbf{D}^{-1}\|^8\|\mathbf{Q}^{-1}\|^4\ \mathbb{E}\|\hat{\mathbf{c}} - \mathbf{c}\|^4\right),\\ \Delta_2^* &= O\left(\|\mathbf{D}^{-1}\|^8\ \|\mathbf{B}\|^4\|\mathbf{Q}^{-1}\|^4\ \mathbb{E}\|\hat{\mathbf{v}} - \mathbf{v}\|^4\right),\\ \Delta_3^* &= O\left(\|\mathbf{D}^{-1}\|^4\ \|\mathbf{Q}^{-1}\|^4\ \|\mathbf{D}^{-1}\boldsymbol{\varepsilon}\|^4\right).\end{aligned}$$

Applying Lemma 9.5 and using (9.12) and (9.41) with $b = 0$, we obtain $\Delta_1^* = O\left(n^{-2}2^{2m\alpha}\right) = o(1)$ and $\Delta_2^* = O\left(2^{2m\alpha}\mathbb{E}\|\hat{\mathbf{v}} - \mathbf{v}\|^4\right)$. Also, similarly to (9.44) and (9.46), $\Delta_3^* = O\left(2^{-4ms'}\right)$. To complete the proof of the lemma, recall the definitions of $\hat{\mathbf{v}}$ and $\mathbf{v}$, apply (9.5) with $k \in K_{0m}^*$, and note that, for $k \in K_{0m}^*$, one has $|k - k_{0m}| = O(1)$.

9.6. Proofs of the statements in Section 5: the zero-free part of the wavelet threholding estimator

Proof of Lemma 5.2. Let $R = \mathbb{E}\|\hat{f}_{c,m_0} - f_{c,m_0}\|^2 = R_1 + R_2 + R_3$, where

$$R_1 = \sum_{k\in K_{0mc}^{\varphi}}\mathrm{Var}(\hat{a}_{m_0k}), \quad R_2 = \sum_{j=J}^{\infty}\sum_{k\in K_{0jc}^{\psi}} b_{jk}^2, \quad R_3 = \sum_{j=m_0}^{J-1}\sum_{k\in K_{0jc}^{\psi}}\mathbb{E}(\hat{b}_{jk} - b_{jk})^2.$$

By Lemma 9.2 we derive that, as $n \to \infty$,

$$\begin{aligned}R_1 &= O\left(n^{-1}\,2^{m_0\alpha}\sum_{k\in K_{0m_0c}^{\varphi}}\left[|k - k_{0m_0}|^{-\alpha}\exp(b2^{m_0\beta}|k - k_{0m_0}|^{-\beta})\right]\right)\\ &= O\left(n^{-1}2^{m_0(1+\alpha)}\exp(2^{-(\beta+1)}\ln n)\right) = o\left((\ln n)^{-\frac{2s'}{\beta}}\right).\end{aligned}$$

https://doi.org/10.1051/ps/2012024 Published online by Cambridge University Press

Using (5.3) and (9.56), we derive that

$$R_2 = O\left(2^{-2Js^*}\right) = O\left((\ln n)^{-\frac{4s^*}{\beta}}\right) = O\left((\ln n)^{-\frac{2s'}{\beta}}\right),$$

since $s^* = s'$ for $1 \le p \le 2$ and $s^* = s \ge (s + 1/2 - 1/p)/2$ for $2 < p \le \infty$ due to $s \ge 1/2$. For R_3, we have

$$\begin{aligned} R_3 &= \sum_{j=m_0}^{J-1} \sum_{|k-k_{0j}| \le 2^{j-m_0}} b_{jk}^2 + \sum_{j=m_0}^{J-1} \sum_{|k-k_{0j}| > 2^{j-m_0}} \mathrm{Var}(\tilde{b}_{jk}) \\ &= O\left(\sum_{j=m_0}^{J-1} \left[2^{-2js'} (2^{j-m_0})^{1-2/p} + n^{-1}\, 2^j\, 2^{\alpha m_0} \exp\left(b 2^{\beta m_0}\right) \right] \right) \\ &= O\left(2^{-m_0(s+2/p-1)} + (\ln n)^{(2+\alpha)/\beta} n^{-1+2^{-(\beta+1)}}\right) = O\left((\ln n)^{-\frac{2s'}{\beta}}\right). \end{aligned}$$

To complete the proof of the lemma, note that the upper bounds are uniform for $f \in B^s_{p,q}(A)$. □

Proof of Lemma 5.3. Note that

$$R = \mathbb{E}\|\hat{f}_{c,m} - f_c\|^2 = R_1 + R_2 + R_3 + R_4, \tag{9.54}$$

where

$$\begin{aligned} R_1 &= \sum_{k \in K^{\varphi}_{0mc}} \mathbb{E}(\hat{a}_{mk} - a_{mk})^2, \quad R_2 = \sum_{j=J}^{\infty} \sum_{k \in K^{\psi}_{0jc}} b_{jk}^2, \\ R_3 &= \sum_{j=m}^{J-1} \sum_{k \in K^{\psi}_{0jc}} \mathbb{E}\left[(\tilde{b}_{jk} - b_{jk})^2\, \mathbb{I}(\tilde{b}_{jk}^2 > d^2\, \varrho_n\, 2^{j\alpha}\, |k - k_{0j}|^{-\alpha}) \right], \\ R_4 &= \sum_{j=m}^{J-1} \sum_{k \in K^{\psi}_{0jc}} b_{jk}^2\, \mathbb{P}(\tilde{b}_{jk}^2 \le d^2\, \varrho_n\, 2^{j\alpha}\, |k - k_{0j}|^{-\alpha}), \end{aligned}$$

with ϱ_n defined in (9.20). Using Lemma 9.2, we obtain

$$R_1 = O\left(n^{-1}\, 2^{m\alpha} \sum_{k \in K^{\varphi}_{0mc}} |k - k_{0m}|^{-\alpha} \right) = O\left(n^{-1}\, 2^{m\alpha} (\ln n)^{\mathbb{I}(\alpha=1)} \right), \tag{9.55}$$

since the set K^{φ}_{0mc} contains $O(\ln n)$ terms and the sum $\sum_{k \in K^{\varphi}_{0mc}} |k - k_{0m}|^{-\alpha}$ is uniformly bounded if $\alpha > 1$.

It is well-known (see, *e.g.*, Johnstone [19], Lem. 19.1) that if $f \in B^s_{p,q}(A)$, then for some constant $c^\star > 0$, dependent on p, q, s and A only, one has

$$\sum_{k=0}^{2^j - 1} b_{jk}^2 \le c^\star 2^{-2js^*} \quad \text{with} \quad s^* = \min(s, s'). \tag{9.56}$$

Therefore, an upper bound for R_2 is of the form

$$R_2 = \sum_{j=J}^{\infty} \sum_{k=0}^{2^j - 1} b_{jk}^2 = O\left(2^{-2Js'}\right).$$

https://doi.org/10.1051/ps/2012024 Published online by Cambridge University Press

If $1 \le p \le 2$, then $s^* = s'$ and $R_2 = O\left(n^{-\frac{2s'}{2s'+\alpha}} \ (\ln n)^{\frac{2s'}{2s'+\alpha}}\right)$. If $2 \le p \le \infty$, then $s^* = s$ and, since $s \ge 1/2$, one has $p > (4s - 2\alpha - 2)/(4s^2 - \alpha - 1)$. Hence,

$$2s/(\alpha+1) > 2s'/(2s'+\alpha),$$

so that, for $1 \le p \le \infty$, one has

$$R_2 = O\left(n^{-\frac{2s'}{2s'+\alpha}} \ (\ln n)^{\frac{2s'}{2s'+\alpha}}\right). \tag{9.57}$$

In oder to obtain an upper bound for R_3 and R_4, note that

$$R_3 \le R_{31} + R_{32}, \quad R_4 \le R_{41} + R_{42}, \tag{9.58}$$

where

$$\begin{aligned}
R_{31} &= \sum_{j=m}^{J-1} \sum_{k \in K^{\psi}_{0jc}} \mathbb{E}\left[(\tilde{b}_{jk} - b_{jk})^2 \ \mathbb{I}((\tilde{b}_{jk} - b_{jk})^2 > 0.25\, d^2\, \varrho_n\, 2^{j\alpha}\, |k - k_{0j}|^{-\alpha})\right], \\
R_{32} &= \sum_{j=m}^{J-1} \sum_{k \in K^{\psi}_{0jc}} \mathbb{E}\left[(\tilde{b}_{jk} - b_{jk})^2 \ \mathbb{I}(b_{jk}^2 > 0.25\, d^2\, \varrho_n^2\, 2^{j\alpha}\, |k - k_{0j}|^{-\alpha})\right], \\
R_{41} &= \sum_{j=m}^{J-1} \sum_{k \in K^{\psi}_{0jc}} b_{jk}^2 \ \mathbb{P}((\tilde{b}_{jk} - b_{jk})^2 > 0.25\, d^2\, \varrho_n^2\, 2^{j\alpha}\, |k - k_{0j}|^{-\alpha}), \\
R_{42} &= \sum_{j=m}^{J-1} \sum_{k \in K^{\psi}_{0jc}} b_{jk}^2 \ \mathbb{I}(b_{jk}^2 \le 2.25\ d^2\, \varrho_n^2\, 2^{j\alpha}\, |k - k_{0j}|^{-\alpha}).
\end{aligned} \tag{9.59}$$

Applying Lemma 9.6 with $w(\cdot) = \psi(\cdot)$ and $\tau = 0.5d$, we obtain

$$\mathbb{P}((\tilde{b}_{jk} - b_{jk})^2 > 0.25\, d^2\, \varrho_n^2\, 2^{j\alpha}\, |k - k_{0j}|^{-\alpha}) = O\left(n^{-0.5d/C_d}\right),$$

where C_d is given by (5.20). Hence, by Lemma 9.2 and inequality $\sqrt{a+b} \le \sqrt{a} + \sqrt{b}$, for $d > 4C_d$, as $n \to \infty$, we obtain

$$\begin{aligned}
R_{31} &\le \sum_{j=m}^{J-1} \sum_{k \in K^{\psi}_{0jc}} \left[\mathbb{E}(\tilde{b}_{jk} - b_{jk})^4 \ \cdot \ \mathbb{P}((\tilde{b}_{jk} - b_{jk})^2 > 0.25\, d^2\, \varrho_n^2\, 2^{j\alpha}\, |k - k_{0j}|^{-\alpha})\right]^{1/2} \\
&= O\left(n^{-\frac{d}{4C_d}} \left[n^{-\frac{3}{2}} 2^{\frac{j(3\alpha+1)}{2}} \sum_{k \in K^{\psi}_{0jc}} |k - k_{0j}|^{-\frac{3\alpha}{2}} + n^{-1} 2^{j\alpha} \sum_{k \in K^{\psi}_{0jc}} |k - k_{0j}|^{-\alpha}\right]\right) \\
&= O\left(n^{-\frac{d}{4C_d}}\right) = o\left(n^{-\frac{2s'}{2s'+\alpha}} \ (\ln n)^{\frac{2s'}{2s'+\alpha}}\right).
\end{aligned} \tag{9.60}$$

Similarly, by (9.56),

$$R_{41} = O\left(n^{-\frac{d}{2C_d}}\right) \sum_{j=m}^{J-1} \sum_{k \in K^{\psi}_{0jc}} b_{jk}^2 = o(n^{-1}). \tag{9.61}$$

https://doi.org/10.1051/ps/2012024 Published online by Cambridge University Press

Now, consider R_{32} and R_{42}. Note that it follows from Lemma 9.2 that

$$R_{32} = O\left(\sum_{j=m}^{J-1}\sum_{k\in K^{\psi}_{0jc}}\left[n^{-1}2^{j\alpha}|k-k_{0j}|^{-\alpha}\mathbb{I}(b^2_{jk} > 0.5\,d^2\,n^{-1}\ln n\,2^{j\alpha}\,|k-k_{0j}|^{-\alpha})\right]\right)$$
$$= O\left(\sum_{j=m}^{J-1}\sum_{k\in K^{\psi}_{0jc}}\min\left[(\ln n)^{-1}b^2_{jk}, n^{-1}2^{j\alpha}|k-k_{0j}|^{-\alpha}\right]\right)$$

and, similarly,

$$R_{42} = O\left(\sum_{j=m}^{J-1}\sum_{k\in K^{\psi}_{0jc}}\min\left[b^2_{jk}, n^{-1}\ln n\,2^{j\alpha}|k-k_{0j}|^{-\alpha}\right]\right).$$

Hence,

$$R_{32} = O\left((\ln n)^{-1}R_{42}\right) = O\left(R_{42}\right) \tag{9.62}$$

so that one needs to study only R_{42}. Partition R_{42} as $R_{42} = R_{421} + R_{422} + R_{423}$, where

$$R_{421} = \sum_{j=m}^{j_1}\sum_{k\in K^{\psi}_{0jc}}\left[n^{-1}\ln n\,2^{j\alpha}|k-k_{0j}|^{-\alpha}\right], \quad R_{422} = \sum_{j=j_2}^{J-1}\sum_{k\in K^{\psi}_{0jc}} b^2_{jk},$$
$$R_{423} = \sum_{j=j_1+1}^{j_2-1}\left[\sum_{|k-k_{0j}|>N_j} n^{-1}\ln n\,2^{j\alpha}|k-k_{0j}|^{-\alpha} + \sum_{|k-k_{0j}|\le N_j} b^2_{jk}\right],$$

and the values of j_1, j_2 and N_j will be defined later. It is easy to see that, by (9.56),

$$R_{421} = O\left(n^{-1}\ln n\,2^{j_1\alpha}(\ln n)^{\mathbb{I}(\alpha=1)}\right), \quad R_{422} = O\left(2^{-2j_2s^*}\right), \tag{9.63}$$

$$R_{423} = O\left(\sum_{j=j_1+1}^{j_2-1}\left[n^{-1}\ln n\,2^{j\alpha}N_j^{1-\alpha}(\ln n)^{\mathbb{I}(\alpha=1)} + 2^{-2js'}N_j^{1-2/p}\right]\right). \tag{9.64}$$

If $\alpha \neq 1$, the two terms in (9.64) are equal to each other when

$$N_j = \left(n^{-1}\ln n 2^{j(2s'+\alpha)}\right)^{1/(\alpha-2/p)},$$

and, for this value of N_j, one has

$$R_{423} = O\left(\sum_{j=j_1+1}^{j_2-1}(n/\ln n)^{\frac{2/p-1}{\alpha-2/p}}2^{\frac{2j(s'-\alpha s)}{\alpha-2/p}}\right). \tag{9.65}$$

Therefore, R_{423} behave differently when $\alpha s \ge s'$ and $\alpha s < s'$, and we consider those cases separately.

First, consider the case when $\alpha s = s'$. Then

$$R_{423} = O\left((j_2-j_1)(n/\ln n)^{\frac{2/p-1}{\alpha-2/p}}\right) = O\left((\ln n/n)^{\frac{2s'}{2s'+\alpha}}\ln n\right) \qquad \text{if} \qquad \alpha s = s'.$$

https://doi.org/10.1051/ps/2012024 Published online by Cambridge University Press

If $\alpha > 1$, $\alpha s > s'$, choose j_1 and j_2 such that

$$2^{j_1} = (n/\ln n)^{\frac{1}{2s'+\alpha}}, \quad 2^{j_2} = (n/\ln n)^{\frac{s'}{s^*(2s'+\alpha)}}.$$

Note that if $1 \le p \le 2$, one has $s^* = s \ge s'$, so that $j_2 \le j_1$ and $R_{423} = 0$. If $2 < p \le \infty$, then $j_2 > j_1$. Also, it follows from (9.63) and (9.65) that $R_{423} = O\left(n^{\frac{2/p-1}{\alpha-2/p}} (\ln n)^{\frac{1-\alpha}{\alpha-2/p}} 2^{\frac{2j_1(s'-\alpha s)}{\alpha-2/p}}\right)$. Hence, $R_{421} = O\left((n/\ln n)^{-\frac{2s'}{2s'+\alpha}}\right)$, $R_{422} = O\left((n/\ln n)^{-\frac{2s'}{2s'+\alpha}}\right)$ and $R_{423} = O\left((n/\ln n)^{-\frac{2s'}{2s'+\alpha}}\right)$, so that

$$R_{42} = O\left((n/\ln n)^{-\frac{2s'}{2s'+\alpha}}\right) \qquad \text{if} \qquad \alpha s \ge s',\ \alpha > 1. \tag{9.66}$$

Similarly, if $\alpha > 1$, $\alpha s < s'$, choose j_1 and j_2 such that

$$2^{j_1} = (n/\ln n)^{\frac{1}{\alpha(2s+1)}}, \quad 2^{j_2} = (n/\ln n)^{\frac{1}{2s+1}}.$$

In this case, $R_{423} = O\left(n^{\frac{2/p-1}{\alpha-2/p}} (\ln n)^{\frac{1-\alpha}{\alpha-2/p}} 2^{\frac{2j_2(s'-\alpha s)}{\alpha-2/p}}\right)$, and direct calculations yield $R_{421} = O\left((n/\ln n)^{-\frac{2s}{2s+1}}\right)$, $R_{422} = O\left((n/\ln n)^{-\frac{2s}{2s+1}}\right)$ and $R_{423} = O\left((n/\ln n)^{-\frac{2s}{2s+1}}\right)$, so that

$$R_{42} = O\left((n/\ln n)^{-\frac{2s}{2s+1}}\right) \qquad \text{if} \qquad \alpha s < s',\ \alpha > 1. \tag{9.67}$$

Finally, if $\alpha = 1$, set $j_1 = j_2$ such that

$$2^{j_1} = (n/\ln^2 n)^{\frac{1}{2s^*+1}}$$

and obtain

$$R_{42} = O\left(n^{-\frac{2s^*}{2s^*+1}} (\ln n)^{\frac{4s^*-1}{2s^*+1}}\right) \qquad \text{if} \qquad \alpha = 1. \tag{9.68}$$

Now, to complete the proof of (5.21), one just need to combine (9.54)–(9.68), and to note that all upper bounds are uniform for $f \in B^s_{p,q}(A)$. □

In order to prove (5.22), note that

$$R^* = \mathbb{E}\|\hat{f}_{c,m} - f_c\|^4 \le R_1^* + R_2^* + R_3^*,$$

where

$$R_1^* = O\left(\mathbb{E}\left\|\sum_{k\in K^{\varphi}_{0mc}} (\hat{a}_{mk} - a_{mk})\varphi_{mk}(x)\right\|^4\right), \quad R_2^* = O\left(\left\|\sum_{j=m}^{\infty}\sum_{k\in K^{\psi}_{0jc}} b_{jk}\psi_{jk}(x)\right\|^4\right),$$

$$R_3^* = O\left(\mathbb{E}\left[\sum_{j=m}^{J-1}\sum_{k\in K^{\psi}_{0jc}} (\tilde{b}_{jk} - b_{jk})^2\, \mathbb{I}(\tilde{b}_{jk}^2 > d^2\, \varrho_n\, 2^{j\alpha}\, |k - k_{0j}|^{-\alpha})\right]^2\right).$$

Observe that, by Lemma 9.2, since $2^{m(\alpha+1)} = o(n/\ln n)$, as $n \to \infty$,

$$\begin{aligned} R_1^* &= O\left(2^m \sum_{k\in K^{\varphi}_{0mc}} \mathbb{E}(\hat{a}_{mk} - a_{mk})^4\right) = O\left(n^{-3}\, 2^{m(3\alpha+2)} + n^{-2}\, 2^{m(2\alpha+1)}\right) \\ &= O\left(n^{-2}\, 2^{m(2\alpha+1)}\right) = o(1). \end{aligned}$$

https://doi.org/10.1051/ps/2012024 Published online by Cambridge University Press

For R_2^*, by (9.56), we have

$$R_2^* = O\left(\left[\sum_{j=m}^{\infty}\sum_{k\in K_{0jc}^{\psi}} b_{jk}^2\right]^2\right) = O\left(\left[2^{-2ms'}\right]^2\right) = o(1).$$

Finally, similarly to (9.59), partition R_3^* as $R_3^* = R_{31}^* + R_{32}^*$ with R_{31}^* and R_{32}^* corresponding to $\mathbb{I}(|\tilde{b}_{jk} - b_{jk}|^2 > 0.25\, d^2\, \varrho_n\, 2^{j\alpha}\, |k - k_{0j}|^{-\alpha})$ and $\mathbb{I}(b_{jk}^2 > 0.25\, d^2\, \varrho_n\, 2^{j\alpha}\, |k - k_{0j}|^{-\alpha})$, respectively. For R_{31}^*, applying Lemmas 9.2 and 9.6 with $w = \psi$ and C_d given by (5.20), and also noting that $\sum_{k\in K_{0jc}^{\psi}} |k - k_{0j}|^{-l} = O(1)$ for $l > 1$, we derive

$$\begin{aligned}
R_{31}^* &= O\left(n\sum_{j=m}^{J-1}\sum_{k\in K_{0jc}^{\psi}} \mathbb{E}\left[|\tilde{b}_{jk} - b_{jk}|^4\, \mathbb{I}(|\tilde{b}_{jk} - b_{jk}|^2 > 0.25\, d^2\, \varrho_n\, 2^{j\alpha}\, |k - k_{0j}|^{-\alpha})\right]\right)\\
&= O\left(n\sum_{j=m}^{J-1}\sum_{k\in K_{0jc}^{\psi}} \left[\mathbb{E}|\tilde{b}_{jk} - b_{jk}|^6\right]^{2/3} \left[\mathbb{P}(|\tilde{b}_{jk} - b_{jk}|^2 > 0.25\, d^2\, \varrho_n\, 2^{j\alpha}\, |k - k_{0j}|^{-\alpha})\right]^{1/3}\right)\\
&= O\left(n\sum_{j=m}^{J-1} n^{-d/(3C_d)} \left[n^{-10/3}\, 2^{j(10\alpha+4)/3} + n^{-8/3}\, 2^{j(8\alpha+2)/3} + n^{-2}\, 2^{2j\alpha}\right]\right)\\
&= o\left(n^{1-d/(3C_d)}\right) = o(1), \quad n \to \infty,
\end{aligned}$$

since $d > 3C_d$. For R_{32}^*, using Lemmas 9.2 and (9.56), we derive that

$$\begin{aligned}
R_{32}^* &= O\left(n\sum_{j=m}^{J-1}\sum_{k\in K_{0jc}^{\psi}} \mathbb{E}\left[|\tilde{b}_{jk} - b_{jk}|^4\, \mathbb{I}(b_{jk}^2 > 0.25\, d^2\, \varrho_n\, 2^{j\alpha}\, |k - k_{0j}|^{-\alpha})\right]\right)\\
&= O\left(n\sum_{j=m}^{J-1}\sum_{k\in K_{0jc}^{\psi}} [2^j\, \ln^{-3} n\, b_{jk}^6 + \ln^{-2} n\, b_{jk}^4]\right) = o\left(n\sum_{j=m}^{J-1}[2^{j(1-6s')} + 2^{-4js'}]\right)
\end{aligned}$$

since $n^{-1}\, 2^{j\alpha}|k - k_{0j}|^{-\alpha} < 0.25\, b_{jk}^2/(d^2 \ln n)$. Note that $m \geq m_0$ implies $2^m \geq n^{1/(2s'+\alpha)}$, so that

$$R_{32}^* = o\left(n^{-\frac{6s'-1}{2s'+\alpha}} + n^{-\frac{4s'}{2s'+\alpha}}\right) = o(1),$$

which completes the proof of the lemma.

9.7. Proof of the asymptotic minimax upper bounds for the L^2-risk in Section 6

Proof of Theorem 6.2. Since $\hat{m} = m_0$ for $b > 0$, the validity of Theorem 6.2 for $b > 0$ follows directly from Lemma 5.2. For $b = 0$, observe that

$$\begin{aligned}
\Delta = \mathbb{E}[\|\hat{f}_m - f\|^2 &= \sum_{m=m_1}^{m_0} \mathbb{E}[\|\hat{f}_m - f\|^2\ \ \mathbb{I}(\hat{m} = m \leq m_0)] + \mathbb{E}[\|\hat{f}_m - f\|^2\ \ \mathbb{I}(\hat{m} = m > m_0)]\\
&\equiv \Delta_1 + \Delta_2,
\end{aligned}$$

https://doi.org/10.1051/ps/2012024 Published online by Cambridge University Press

and consider terms Δ_1 and Δ_2 separately.

Denote

$$R(n) = \begin{cases} O\left(n^{-\frac{2s}{2s+1}} (\ln n)^{\mu_1}\right) & \text{if} \quad b=0,\ \alpha s < s', \\ O\left(n^{-\frac{2s'}{2s'+\alpha}} (\ln n)^{\mu_2}\right) & \text{if} \quad b=0,\ \alpha s \geq s', \end{cases} \tag{9.69}$$

and note that, for any $m \geq m_1$,

$$\mathbb{E}\|\hat{f}_m - f\|^2 \leq 2[\,\mathbb{E}\|\hat{f}_{m_0} - f\|^2 + \mathbb{E}\|(\hat{f}_m - \hat{f}_{m_0})\mathbb{I}(x \in \Xi_m)\|^2 + \mathbb{E}\|(\hat{f}_m - \hat{f}_{m_0})\mathbb{I}(x \in \Xi_m^c)\|^2]$$

where m_1 is defined in (5.3) and set Ξ_m is defined in (6.1). By Lemmas 5.1 and 5.3, we obtain

$$\mathbb{E}\|\hat{f}_{m_0} - f\|^2 = O\left(n^{-\frac{2s'}{2s'+\alpha}} + R(n)\right).$$

If $\hat{m} = m \leq m_0$, then by definition of $\hat{m}$, we derive that

$$\mathbb{E}\|(\hat{f}_m - \hat{f}_{m_0})\mathbb{I}(x \in \Xi_m)\|^2 \leq \lambda^2\, 2^{m_0\alpha}\, n^{-1} \ln n = O\left(n^{-\frac{2s'}{2s'+\alpha}}\right).$$

Now, recall that Ξ_m is defined in such a way that $\mathrm{supp}(f_{0,m}) \in \Xi_m$ for any m, and that $\Xi_{j1} \subset \Xi_{j2}$ for $j_1 > j_2$, so that, for $m \leq m_0$,

$$\begin{aligned} \mathbb{E}\|(\hat{f}_m - f)\mathbb{I}(x \in \Xi_m^c)\|^2 &= \mathbb{E}\|(\hat{f}_{0,m} + \hat{f}_{c,m} - f_{0,m} - f_{c,m})\mathbb{I}(x \in \Xi_m^c)\|^2 \\ &= \mathbb{E}\|(\hat{f}_{c,m} - f_{c,m})\mathbb{I}(x \in \Xi_m^c)\|^2 \leq \mathbb{E}\|\hat{f}_{c,m} - f_{c,m}\|^2 = O\left(R(n)\right) \end{aligned}$$

as $n \to \infty$. Noting that

$$\mathbb{E}\|(\hat{f}_m - \hat{f}_{m_0})\mathbb{I}(x \in \Xi_m^c)\|^2 \leq 2\left[\mathbb{E}\|(\hat{f}_m - f)\mathbb{I}(x \in \Xi_m^c)\|^2 + \mathbb{E}\|(\hat{f}_{m_0} - f)\mathbb{I}(x \in \Xi_m^c)\|^2\right]$$

and combining all formulae above, we obtain that $\Delta_1 = O\left(R(n)\right)$ as $n \to \infty$.

By Lemmas 5.1 and 5.3, one has $\mathbb{E}\|\hat{f}_{0,m} - f_{0,m}\|^4 = o\,(1)$ and $\mathbb{E}\|\hat{f}_{c,m} - f_{c,m}\|^4 = o\,(1)$. Then, Lemma 6.1 yields

$$\Delta_2 \leq \sqrt{\mathbb{E}[\|\hat{f}_m - f\|^4}\ \sqrt{\mathbb{P}(\hat{m} = m > m_0)} = O\left(n^{-\frac{\lambda}{2C_\lambda}} + n^{\frac{1}{2(\alpha+1)} - \frac{d}{4C_d}}\right) = O(n^{-1}),$$

provided $\lambda \geq \max\left(C_{\lambda 1}, C_{\lambda 2}, 2C_\lambda\right)$ and $d > 2(\alpha+1)^{-1}(2\alpha+3)C_d$, which completes the Proof of Theorem 6.2. □

9.8. Proof of the asymptotic minimax upper bounds for the L^2-risk in Section 7

The Proof of Theorem 7.1 is based on the following lemma.

Lemma 9.9. *Let Assumption A hold with $b = 0$ and $0 < \alpha < 1$. Then,*

$$\begin{aligned} \mathit{Var}(\tilde{b}_{jk}) &= O\left(n^{-1} 2^{j\alpha}\ \min(1, |k - k_{0j}|^{-\alpha})\right), \\ \mathbb{E}|\tilde{b}_{jk} - b_{jk}|^{\frac{\alpha+3}{\alpha+1}} &= O\left(n^{-\frac{2}{\alpha+1}} 2^{j\frac{(\alpha+3)}{2(\alpha+1)}} + n^{-\frac{\alpha+3}{2(\alpha+2)}} 2^j\right). \end{aligned}$$

Proof of Lemma 9.9. Proof of the first statement is very similar to the proof of validity of formula (9.6). Proof of the second statement is based on Lemma 3.1. in Chesneau [6] which states that whenever $\int [g(x)]^{1-\nu} \mathrm{d}x < \infty$ for some $\nu > 2$, one has

$$\mathbb{E}|\tilde{b}_{jk} - b_{jk}|^\nu = O\left(n^{1-\nu} \int |\psi_{jk}(x)|^\nu\, [g(x)]^{1-\nu} \mathrm{d}x + n^{-\nu/2} \int \psi_{jk}^2(x)\, [g(x)]^{-\nu/2} \mathrm{d}x\right). \tag{9.70}$$

To complete the proof, note that for $\nu = 1 + 2/(\alpha+1) > 2$ one has $\int [g(x)]^{1-\nu} \mathrm{d}x < \infty$ and apply (9.70). □

https://doi.org/10.1051/ps/2012024 Published online by Cambridge University Press

Proof of Theorem 7.1. The proof of this statement is similar to the Proof of Lemma 5.3. Indeed, similarly to the Proof of Lemma 5.3, partition the risk as $R = \mathbb{E}\|\hat{f} - f\|^2 = R_1 + R_2 + R_3 + R_4$ where, similarly to the Proof of Lemma 5.3,

$$R_1 = \sum_{k=0}^{2^{m_1}-1} \mathbb{E}(\hat{a}_{m_1 k} - a_{m_1 k})^2, \quad R_2 = \sum_{j=J}^{\infty} \sum_{k=0}^{2^j-1} b_{jk}^2,$$

$$R_3 = \sum_{j=0}^{J-1} \sum_{k=0}^{2^j-1} \mathbb{E}\left[(\tilde{b}_{jk} - b_{jk})^2 \, \mathbb{I}(\tilde{b}_{jk}^2 > d^2 \, \varrho_n \, 2^{j\alpha} \, |k - k_{0j}|^{-\alpha})\right],$$

$$R_4 = \sum_{j=0}^{J-1} \sum_{k=0}^{2^j-1} b_{jk}^2 \, \mathbb{P}(\tilde{b}_{jk}^2 \le d^2 \, \varrho_n \, 2^{j\alpha} \, |k - k_{0j}|^{-\alpha})$$

with ϱ_n defined in (9.20). Since $1/g$ is integrable and m_1 in (5.3) is finite, it is easy to show that $R_1 = O\left(n^{-1}\right)$. Also, same as before, $R_2 = O\left(2^{-2Js^*}\right)$. If $p > 2$, then $\alpha + 1 < 2s + 1$ since $s \ge \max(1/2, 1/p)$ and $\alpha < 1$, so that $R_2 = O\left(n^{-2s/(2s+1)}\right)$. If $1 \le p \le 2$, then $s^* = s'$ and $2s'/(1+\alpha) > \max\{2s'/(2s'+\alpha), 2s/(2s+1)\}$, so that

$$R_2 = O\left(\max\left\{n^{-2s/(2s+1)}, n^{-2s'/(2s'+\alpha)}\right\}\right).$$

Now, similarly to the Proof of Lemma 5.3, partition R_3 and R_4 as $R_3 \le R_{31} + R_{32}$ and $R_4 \le R_{41} + R_{42}$. Using Lemma 9.9, as $n \to \infty$, obtain upper bounds

$$\begin{aligned} R_{31} &\le \sum_{j=0}^{J-1} \sum_{k=0}^{2^j-1} \left[\mathbb{P}((\tilde{b}_{jk} - b_{jk})^2 > 0.25\, d^2 \, \varrho_n^2 \, 2^{j\alpha} \, |k - k_{0j}|^{-\alpha})\right]^{1-2/\nu} \left[\mathbb{E}|\tilde{b}_{jk} - b_{jk}|^{\nu}\right]^{2/\nu} \\ &= O\left(\sum_{j=0}^{J-1} 2^j \, n^{-d(1-2/\nu)/(2C_d)} \left[n^{1-\nu} 2^{j\nu/2} + n^{-\nu/2} \, 2^j\right]^{2/\nu}\right) \\ &= O\left(2^j \, n^{-d(1-2/\nu)/(2C_d)} \left[n^{-\nu/2} \, 2^J\right]^{2/\nu}\right) = O\left(n^{-1}\right), \end{aligned}$$

provided (7.2) holds, and also

$$R_{41} = O\left(n^{-\frac{d}{2C_d}}\right) \sum_{j=m}^{J-1} \sum_{k \in K_{0jc}^{\psi}} b_{jk}^2 = o(n^{-1}).$$

Now, same as before, $R_{32} = O\left((\ln n)^{-1} R_{42}\right) = O\left(R_{42}\right)$, so that we need to construct upper bounds for R_{42} only. Partition R_{42} as $R_{42} = R_{421} + R_{422} + R_{423}$ where

$$R_{421} = \sum_{j=0}^{j_1} \sum_{k=0}^{2^j-1} \left[n^{-1} \, \ln n \, 2^{j\alpha} |k - k_{0j}|^{-\alpha}\right], \quad R_{422} = \sum_{j=j_2}^{J-1} \sum_{k=0}^{2^j-1} b_{jk}^2,$$

$$R_{423} = \sum_{j=j_1+1}^{j_2-1} \left\{ \sum_{|k-k_{0j}|>N_j} |b_{jk}|^p \left[n^{-1} \, \ln n \, 2^{j\alpha} N_j^{-\alpha}\right]^{1-p/2} + \sum_{|k-k_{0j}| \le N_j} n^{-1} \, \ln n \, 2^{j\alpha} N_j^{1-\alpha} \right\},$$

and the values of j_1, j_2 and N_j will be defined later. It is easy to see that, same as before, $R_{421} = O\left(n^{-1} \, \ln n \, 2^{j_1\alpha}\right)$ and $R_{422} = O\left(2^{-2j_2 s^*}\right)$. For R_{423} we can write the following expression

$$R_{423} = \sum_{j=j_1+1}^{j_2-1} \left[2^{-js'p} \left(\frac{\ln n}{n} \, 2^{j\alpha} N_j^{-\alpha}\right)^{1-p/2} + \frac{\ln n}{n} \, 2^{j\alpha} N_j^{1-\alpha}\right].$$

https://doi.org/10.1051/ps/2012024 Published online by Cambridge University Press

If $p \geq 2$, we choose $j_1 = j_2$ such that $2^{j_1} = (\ln n/n)^{1/(2s+1)}$ and obtain $R_{42} = O\left((\ln n/n)^{2s/(2s+1)}\right)$. If $1 \leq p < 2$, we choose N_j which equalize the two terms in R_{423} and obtain, similarly to (9.65),

$$R_{423} = \begin{cases} O\left((n/\ln n)^{\frac{2/p-1}{\alpha-2/p}} 2^{\frac{2j_2(s'-\alpha s)}{\alpha-2/p}}\right) & \text{if} \quad \alpha s < s' \\ O\left((n/\ln n)^{\frac{2/p-1}{\alpha-2/p}} 2^{\frac{2j_1(s'-\alpha s)}{\alpha-2/p}}\right) & \text{if} \quad \alpha s > s' \\ O\left((j_2 - j_1)\,(n/\ln n)^{\frac{2/p-1}{\alpha-2/p}}\right) & \text{if} \quad \alpha s = s' \end{cases} \tag{9.71}$$

If $\alpha s < s'$, then choose

$$2^{j_1} = (n/\ln n)^{\frac{1}{2s+1}}, \quad 2^{j_2} = (n/\ln n)^{\frac{s}{s'(2s+1)}},$$

so that $j_1 < j_2$. Direct calculations show that in this case

$$R_{42} = O\left((n/\ln n)^{\zeta_1}\right) \quad with \quad \zeta_1 = \frac{2/p-1}{2/p-\alpha} + \frac{2(s'-\alpha s)}{(2s+1)(2/p-\alpha)} = \frac{2s}{2s+1}$$

and $R_{42} = O\left((\ln n/n)^{2s/(2s+1)}\right)$. If $\alpha s > s'$, then set

$$2^{j_1} = (n/\ln n)^{\frac{\alpha}{2s'+\alpha}}, \quad 2^{j_2} = (n/\ln n)^{\frac{1}{2s'+\alpha}},$$

so that again $j_1 < j_2$. Here we have

$$R_{42} = O\left((n/\ln n)^{\zeta_2}\right) \quad with \quad \zeta_2 = \frac{2/p-1}{2/p-\alpha} - \frac{2(\alpha s - s')}{(2s'+\alpha)(2/p-\alpha)} = \frac{2s'}{2s'+\alpha}$$

and $R_{42} = O\left((\ln n/n)^{2s'/(2s'+\alpha)}\right)$. If $\alpha s = s'$, then note that $j_2 - j_1 = O(\ln n)$, so that $R_{42} = O\left((\ln n/n)^{2s'/(2s'+\alpha)}\right) = O\left((\ln n/n)^{2s/(2s+1)}\right)$. Now, to complete the proof, just combine the expressions for R_1, R_2, R_{31}, R_{41}, R_{32} and R_{42}. □

Acknowledgements. Marianna Pensky was supported in part by National Science Foundation (NSF), grant DMS-1106564. The authors want to thank Yuri Golubev for valuable discussion of results and methodologies used in this paper. We would like also to thank the referees for constructive comments that led to improvements.

References

[1] U. Amato, A. Antoniadis and M. Pensky, Wavelet kernel penalized estimation for non-equispaced design regression. *Statist. Comput.* **16** (2006) 37–55.
[2] A. Antoniadis and D.T. Pham, Wavelet regression for random or irregular design. *Comput. Statist. Data Anal.* **28** (1998) 353–369.
[3] N. Bissantz, L. Dumbgen, H. Holzmann and A. Munk, Nonparametric confidence bands in deconvolution density estimation. *J. Royal Statist. Soc. Series B* **69** (2007) 483–506.
[4] L.D. Brown, T.T. Cai, M.G. Low and C.H. Zhang, Asymptotic equivalence theory for nonparametric regression with random design. *Annal. Statist.* **30** (2002) 688–707.
[5] T.T. Cai and L.D. Brown, Wavelet shrinkage for nonequispaced samples. *Annal. Statist.* **26** (1998) 1783–1799.
[6] C. Chesneau, Regression in random design: a minimax study. *Statist. Probab. Lett.* **77** (2007) 40–53.
[7] I. Daubechies, Ten Lectures on Wavelets. *Philadelphia: SIAM* (1992).
[8] R.B. Dingle, *Asymptotic Expansions: Their Derivation and Interpretation*. London: Academic Press (1973).
[9] S. Gaïffas,. Convergence rates for pointwise curve estimation with a degenerate design. *Math. Meth. Statist.* **14** (2005) 1–27.
[10] S. Gaïffas, Sharp estimation in sup norm with random design. *Statist. Probab. Lett.* **77** (2006) 782–794.
[11] S. Gaïffas, On pointwise adaptive curve estimation based on inhomogeneous data. *ESAIM: Probab. Statist.* **11** (2007) 344–364.
[12] S. Gaïffas, Uniform estimation of a signal based on inhomogeneous data. *Statistica Sinica* **19** (2009) 427–447.
[13] E. Giné and R. Nickl, Confidence bands in density estimation. *Annal. Statist.* **38** (2010) 1122–1170.

https://doi.org/10.1051/ps/2012024 Published online by Cambridge University Press

[14] A. Guillou and N. Klutchnikoff, Minimax pointwise estimation of an anisotropic regression function with unknown density of the design. *Math. Methods Statist.* **20** (2011) 30–57.
[15] P. Hall and B.A. Turlach, Interpolation methods for nonlinear wavelet regression with irregularly spaced design. *Annal. Statist.* **25** (1997) 1912–1925.
[16] W. Härdle, G. Kerkyacharian, D. Picard and A. Tsybakov, Wavelets, Approximation, and Statistical Applications, vol. 129 of *Lect. Notes Statist.* Springer-Verlag, New York (1998).
[17] M. Hoffmann and M. Reiss, Nonlinear estimation for linear inverse problems with error in the operator. *Annal. Statist.* **36** (2008) 310–336.
[18] I.M. Johnstone, Minimax Bayes, asymptotic minimax and sparse wavelet priors, in *Statistical Decision Theory and Related Topics*, edited by S.S. Gupta and J.O. Berger. Springer-Verlag, New York (1994) 303–326,
[19] I.M. Johnstone, *Function Estimation and Gaussian Sequence Models*. Unpublished Monograph (2002). http://statweb.stanford.edu/∼imj/
[20] I.M. Johnstone, G. Kerkyacharian, D. Picard and M. Raimondo, Wavelet deconvolution in a periodic setting. *J. Royal Statist. Soc. Series B* **66** (2004) 547–573 (with discussion, 627–657).
[21] G. Kerkyacharian and D. Picard, Regression in random design and warped wavelets. *Bernoulli* **10** (2004) 1053–1105.
[22] M. Kohler, Nonlinear orthogonal series estimation for random design regression. *J. Statist. Plann. Inference* **115** (2003) 491–520.
[23] A.P. Korostelev and A.B. Tsybakov, Minimax Theory of Image Reconstruction, vol. 82 of *Lect. Notes Statist.* Springer-Verlag, New York (1993).
[24] A. Kovac and B.W. Silverman, Extending the scope of wavelet regression methods by coefficient-dependent thresholding. *J. Amer. Statist. Assoc.* **95** (2000) 172–183.
[25] R. Kulik and M. Raimondo, Wavelet regression in random design with heteroscedastic dependent errors. *Annal. Statist.* **37** (2009) 3396–3430.
[26] O.V. Lepski, A problem of adaptive estimation in Gaussian white noise. *Theory Probab. Appl.* **35** (1990) 454–466.
[27] O.V. Lepski, Asymptotically minimax adaptive estimation I: Upper bounds. Optimally adaptive estimators. *Theory Probab. Appl.* **36** (1991) 682–697.
[29] O.V. Lepski, E. Mammen and V.G. Spokoiny, Optimal spatial adaptation to inhomogeneous smoothness: an approach based on kernel estimates with variable bandwidth selectors *Annal. Statist.* **25** (1997) 929–947.
[29] O. Lepski and V. Spokoiny, Optimal pointwise adaptive methods in nonparametric estimation. *Annal. Statist.* **25** (1997) 2512–2546.
[30] S. Mallat, *A Wavelet Tour of Signal Processing*. 2nd Edition, Academic Press, San Diego (1999).
[31] Y. Meyer, *Wavelets and Operators*. Cambridge: Cambridge University Press (1992).
[32] M. Pensky and T. Sapatinas, Functional deconvolution in a periodic case: uniform case. *Annal. Statist.* **37** (2009)73–104.
[33] M. Pensky and T. Sapatinas, On convergence rates equivalency and sampling strategies in functional deconvolution models. *Annal. Statist.* **38** (2010) 1793–1844.
[34] M. Pensky and B. Vidakovic, On non-equally spaced wavelet regression. *Annal. Instit. Statist. Math.* **53** (2001) 681–690.
[35] S. Sardy, D.B. Percival, A.G. Bruce, H.-Y. Gao and W. Stuelzle, Wavelet shrinkage for unequally spaced data. *Statist. Comput.* **9** (1999) 65–75.
[36] K. Tribouley, Adaptive simultaneous confidence intervals in non-parametric estimation. *Statist. Probab. Lett.* **69** (2004) 37–51.
[37] A.B. Tsybakov, *Introduction to Nonparametric Estimation*. Springer-Verlag, New York (2009).
[38] G. Wahba, *Spline Models for Observational Data*. SIAM, Philadelphia (1990).
[39] S. Zhang, M.-Y. Wong and Z. Zheng, Wavelet threshold estimation of a regression function with random design. *J. Multivariate Anal.* **80** (2002) 256–284.

https://doi.org/10.1051/ps/2012024 Published online by Cambridge University Press