Exact Constants for Pointwise Adaptive Estimation under the Riesz transform

Jussi Klemelä* and Alexandre B. Tsybakov

Universität Heidelberg and *Université Paris VI*

Abstract

We consider nonparametric estimation of a multivariate function and its partial derivatives at a fixed point when the Riesz transform of the function is observed in Gaussian white noise. We assume that the unknown function belongs to some Sobolev class and construct an estimation procedure which achieves the best asymptotic minimax risk when the smoothness of the function is unknown.

Mathematics Subject Classifications: 62G05, 62G20

Key Words: Adaptive curve estimation, Bandwidth selection, Deconvolution, Exact constants in nonparametric smoothing, Gaussian white noise, Inverse problems, Kernel estimation, Minimax risk.

Short title: Estimation under the Riesz transform

*Writing of this article was financed by Deutsche Forschungsgemeinschaft under project MA1026/6-2.

1 Introduction

For a function $f \in L_1(\mathbf{R}^d)$ and for $0 \le \gamma < d$ consider the operator $R_\gamma f$ defined by

$$(R_\gamma f)(x) = \begin{cases} \alpha_\gamma \int_{\mathbf{R}^d} f(y) \|x-y\|^{\gamma-d} dy & \text{if } 0 < \gamma < d, \\ f(x) & \text{if } \gamma = 0, \end{cases}$$

where $\|\cdot\|$ denotes the Euclidean norm in $\mathbf{R}^d$ and

$$\alpha_\gamma = (2\pi)^{-d/2} \pi^{\gamma - d/2} \, \frac{\Gamma((d-\gamma)/2)}{\Gamma(\gamma/2)}.$$

The operator $R_\gamma f$ is called the Riesz transform (see Stein (1970)).

We consider the statistical model

$$dY_\varepsilon(t) = (R_\gamma f)(t) dt + \varepsilon \, dW(t), \;\; t \in \mathbf{R}^d, \tag{1}$$

where W is the d-dimensional Brownian sheet and $0 < \varepsilon < 1$. Given a realisation of the process $Y_\varepsilon(t)$, the problem is to estimate f or partial derivatives of f at a fixed point.

The Riesz transform is a convolution with a function whose Fourier transform has polynomially decreasing tails. Its special cases are several integrals of the type of potential used in mathematical physics, for example, the Newtonian potentials. It is also useful in noise removal from a picture which is corrupted with focusing errors. For $\gamma = 2$ the problem of estimating f in the model (1) is equivalent to recovering the right hand side of the Poisson equation from noisy observations of its solution.

We will assume that f belongs to a Sobolev ball $\mathcal{F}_{\beta,L}$ where $\beta > 0$ is the smoothness index and $L > 0$ is the radius of the ball. When estimating partial derivatives of f of order $r \ge 0$, it can be shown that the optimal pointwise rate of convergence of estimators on $\mathcal{F}_{\beta,L}$ is

$$\left(\varepsilon^2\right)^{(\beta - r - d/2)/(2(\beta+\gamma))}$$

where γ is the index of the Riesz transform (see Ibragimov and Hasminskii (1984), Donoho and Low (1992)). Optimal pointwise rates of convergence are also established for inverse problems with other operators (Korostelev and Tsybakov (1991, 1993), Donoho and Low (1992), Chow and Khasminskii (1997), Chow, Ibragimov and Khasminskii (1999)). The problem of adaptation consists in constructing an estimator which is simultaneously optimal over as large scale of classes $\mathcal{F}_{\beta,L}$ as possible. We consider the scale $(\beta, L) \in [\beta_*, \infty) \times [L_*, L^*]$, where β_*, L_*, L^* are some fixed positive numbers. It can be shown that the minimax rate cannot be achieved simultaneously over this scale: one looses a logarithmic factor, and the optimal adaptive rate is

$$\varphi_{\beta,L} = \left(\varepsilon^2 \log \varepsilon^{-1}\right)^{(\beta-r-d/2)/(2(\beta+\gamma))}$$

(Goldenshluger (1999), Goldenshluger and Pereverzev (2000)). These results concern the rates of convergence. Our aim is to get sharper results, and to investigate the asymptotically exact behavior of the pointwise risks. We propose estimators that attain asymptotically optimal adaptation (with the exact constant) over the scale of classes $\mathcal{F}_{\beta,L}$ with $(\beta, L) \in [\beta_*, \beta_\varepsilon] \times [L_*, L^*]$ where $\beta_\varepsilon \to \infty$ sufficiently slowly, as $\varepsilon \to 0$. More precisely, optimality is defined in the following way.

- We find the constants $c_{\beta,L}$ such that

$$\lim_{\varepsilon\to 0} \inf_{T_\varepsilon} \sup_{(\beta,L)\in B_\varepsilon} (c_{\beta,L}\varphi_{\beta,L})^{-p} \sup_{f\in\mathcal{F}_{\beta,L}} E_f \left|T_\varepsilon - f^{(\alpha_0)}(x_0)\right|^p = 1 \qquad (2)$$

  where $B_\varepsilon = [\beta_*, \beta_\varepsilon] \times [L_*, L^*]$, $\inf_{T_\varepsilon}$ denotes the infimum over all estimators, $f^{(\alpha_0)}(x_0)$ is the partial derivative of order r at a point $x_0 \in \mathbf{R}^d$, and $p > 0$.

- We find the estimator that asymptotically attains the infimum in (2). This estimator is called asymptotically sharp adaptive over B_ε.

Note that in the following we formally fix x_0 and take $x_0 = 0$ to shorten the notation, but our results are valid for any x_0 with an obvious shift of the argument of the initial family of linear estimators (cf. Remark 1 below). So, in fact we consider pointwise adaptive estimation of f on the whole domain of its definition.

The results of this paper are closely related to our previous paper (Klemelä and Tsybakov (2001)) where the case of direct observations ($\gamma = 0$) is studied. Along with extending that result to an inverse problems framework, we allow a wider range of possible values of β. Namely, in Klemelä and Tsybakov (2001) it is assumed that $\beta \in [\beta_*, \beta^*]$ where the upper bound β^* on the smoothness index is supposed to be known: β^* is needed to construct the estimator in Klemelä and Tsybakov (2001) and it appears in the expression for the exact asymptotic constant. The prior knowledge of β^* might be an inconvenient assumption in practice, and an informal suggestion made in Klemelä and Tsybakov (2001) is to assume that β^* is large enough: it can be therefore suppressed in the terms appearing in the construction of the estimator, which leads to a suboptimal constant. Here we show that replacing a fixed β^* by an upper bound β_ε depending on ε ($\beta_\varepsilon \to \infty$ slowly enough, as $\varepsilon \to 0$) allows to get optimality of the constant. We consider here only the Sobolev classes $\mathcal{F}_{\beta,L}$, while in Klemelä and Tsybakov (2001) more general classes (such as Hölder, Besov etc) are covered. Similar framework with Sobolev classes has been studied in the paper by Tsybakov (1998) (for the special case $\gamma = 0$, $d = 1$, $r = 0$), and our results can be also considered as an extension of that paper. The exact constants of Tsybakov (1998) are obtained from our constants $c_{\beta,L}$ for these particular values of parameters. The difference is that in Tsybakov (1998) the function f is defined on the interval [0,1] rather than on the whole space and that the smoothness parameter β lies in a discrete set $\{\beta_1, \dots, \beta_\varepsilon\}$, while L is fixed and known. Butucea (2001a,b) extended a result of Tsybakov (1998) to the problem of density estimation and showed a successful behavior of the asymptotically sharp adaptive estimator on numerical examples. A related work by Lepski and Spokoiny (1997) considers the scale of Hölder classes with $\gamma = 0$, $d = 1$, $r = 0$, and the smoothness parameter that lies in a finite interval $[\beta_*, \beta^*] \times [L_*, L^*]$.

We also mention the results on sharp adaptive estimation for the statistical inverse problems obtained in a framework different from ours. Efromovich (1997a,b) considered a deconvolution problem with supersmooth errors (and logarithmic rates of convergence), and obtained sharp adaptive estimators both in pointwise and L_2 sense. This corresponds to the case where the eigenvalues of the operator of the inverse problem are exponentially decreasing while the signal to recover belongs to a Sobolev ellipsoid in Fourier domain. Cavalier, Golubev, Lepski and Tsybakov (2003) obtained sharp adaptive estimators for the case where the eigenvalues of the operator are exponential and the signal to recover belongs to an ellipsoid with exponential axes in Fourier domain. Cavalier and Tsybakov (2002), Cavalier, Golubev, Picard and Tsybakov (2002) investigated the inverse problems where the eigenvalues are decreasing as a power law; they suggested sharp adaptive estimators with respect to the L_2 risk based either on a penalized blockwise Stein rule or on the unbiased risk estimators, inculding the multidimensional case. These two papers develop asymptotically exact oracle inequalities that are used to prove sharp minimax results.

2 Definition of the estimator and the result

We consider the Gaussian white noise model (1). Without of loss generality, we consider estimation of the value of f or the value of its partial derivatives at $x_0 = 0$. That is, we study the estimation of the derivative $f^{(\alpha_0)}(0)$, where α_0 is a multi-index. For a multi-index $\alpha = (\alpha_1, \ldots, \alpha_d)$ and for $\omega = (\omega_1, \ldots, \omega_d) \in \mathbf{R}^d$ we denote $|\alpha| = \alpha_1 + \ldots + \alpha_d$ and $\omega^\alpha = \omega_1^{\alpha_1} \cdots \omega_d^{\alpha_d}$. We assume that $|\alpha_0| = r$, where $r \geq 0$ is an integer. We may write

$$f^{(\alpha_0)}(x) = i^{|\alpha_0|} \int_{\mathbf{R}^d} \omega^{\alpha_0} \hat{f}(\omega) \exp(ix^T\omega)\, d\omega$$

where $\hat{f}(\omega)$ denotes the Fourier transform of f,

$$\hat{f}(\omega) = \frac{1}{(2\pi)^d} \int_{\mathbf{R}^d} f(x) \exp(-ix^T\omega) dx$$

and i denotes the imaginary unit.

Scale of classes. Consider the Sobolev classes of the form

$$\mathcal{F}_{\beta,L} = \left\{ f \in L_1(\mathbf{R}^d) \,\Big|\, \rho_\beta^2(f) + \|R_\gamma f\|_2^2 \leq L^2 \right\} \tag{3}$$

where $\|\cdot\|_2$ is the $L_2(\mathbf{R}^d)$-norm, $\rho_\beta(\cdot)$ is the Sobolev semi-norm,

$$\rho_\beta^2(f) = (2\pi)^d \int_{\mathbf{R}^d} \|\omega\|^{2\beta} \left| \hat{f}(\omega) \right|^2 d\omega \tag{4}$$

and $\beta > r + d/2$. If β is an integer, then $\rho_\beta^2(f) = \sum_{|\alpha|=\beta} \int_{\mathbf{R}^d} \left| f^{(\alpha)} \right|^2$. For $f \in \mathcal{F}_{\beta,L}$ we have, by the Fourier convolution formula,

$$(R_\gamma f)^\wedge(\omega) = \hat{f}(\omega) \|\omega\|^{-\gamma}, \;\; \omega \in \mathbf{R}^d, \tag{5}$$

see Stein (1970), page 73.

Let the scale of classes $\{\mathcal{F}_{\beta,L}\}_{(\beta,L)\in B}$ be defined by (3) with

$$B = \{(\beta, L) : \;\beta_* \leq \beta < \infty, \; L_* \leq L \leq L^*\}$$

where $r + d/2 < \beta_* < \infty$, $0 < L_* < L^* < \infty$. This means that we are certain that $f \in \mathcal{F}_{\beta,L}$ for some $\beta \in [\beta_*, \infty)$ and $L \in [L_*, L^*]$. The values r and β_* are supposed to be known but L_*, L^* can be unknown: we do not need these values for the construction of our sharp adaptive estimators.

Kernel function. Define the kernel function of the estimator by

$$K_\beta(x) = b^{r+\gamma+d}\tilde{K}_\beta(bx), \;\; x \in \mathbf{R}^d, \tag{6}$$

where

$$\tilde{K}_\beta(x) = (2\pi)^{-d} i^r \int_{\mathbf{R}^d} \omega^{\alpha_0} \|\omega\|^\gamma (1 + \|\omega\|^{2(\beta+\gamma)})^{-1} \exp(ix^T\omega) d\omega \tag{7}$$

and

$$b = b(\beta) \overset{\text{def}}{=} \left(\frac{2(\beta - r) - d}{2(r+\gamma) + d} \right)^{1/(2(\beta+\gamma))}. \tag{8}$$

Note that $\tilde{K}_\beta$ is always real-valued: it is the directional derivative corresponding to the multi-index α_0 of the function whose Fourier transform is $(2\pi)^{-d}\|\omega\|^\gamma(1 + \|\omega\|^{2(\beta+\gamma)})^{-1}$.

Grid. We introduce a sufficiently fine grid on $[\beta_*, \beta_\varepsilon]$ where $\beta_\varepsilon \to \infty$ as $\varepsilon \to 0$, and consider a statistic $\hat{\beta}$ taking values on this grid. To each point of the grid we assign a linear kernel estimator. The statistic $\hat{\beta}$ chooses one of these estimators in a data-driven way. The grid is defined as

$$S = \{\beta_1, \dots, \beta_m\},$$

where

$$r' + d/2 < \beta_1 < \cdots < \beta_m = \beta_\varepsilon$$

with a fixed r' satisfying $r < r' \le \beta_* - d/2$.

We assume that there exist $k_2 > k_1 > 0$ and $\delta_1 \ge \delta > 1$ such that

$$k_1(\log \varepsilon^{-1})^{-\delta_1} \le \beta_{i+1} - \beta_i \le k_2(\log \varepsilon^{-1})^{-\delta}, \quad i = 0, \dots, m-1, \tag{9}$$

where $\beta_0 = r'$, and β_ε is chosen such that

$$\beta_\varepsilon = \left(\log \log \varepsilon^{-1}\right)^{\delta_2} \tag{10}$$

where $0 < \delta_2 < 1$. Note that the same grid S can be used for different values β_*, provided $\beta_* \ge r' + d/2$. In this sense the exact knowledge of β_* is not required for the construction of the estimator.

Scale of estimators. For any $h > 0$ denote

$$K_{\beta,h}(\cdot) = h^{-\gamma-r-d} K_\beta(\cdot/h). \tag{11}$$

Consider kernel estimators of the form $\int K_{\beta,h}(t)dY_\varepsilon(t)$ where h is a suitably chosen bandwidth. Denote

$$h(\beta,\varepsilon) = \varepsilon^{1/(\beta+\gamma)}, \tag{12}$$

and introduce the "effective noise level under adaptation":

$$\tilde{\varepsilon} = \tilde{\varepsilon}(\beta) = \varepsilon d_\varepsilon(\beta) = \left(\lambda(\beta)\varepsilon^2 \log \varepsilon^{-1}\right)^{1/2},$$

where

$$d_\varepsilon(\beta) = \left(\lambda(\beta) \log \varepsilon^{-1}\right)^{1/2} \tag{13}$$

and

$$\lambda(\beta) = \frac{2p(r+\gamma+d/2)}{\beta+\gamma}.$$

We use the bandwidth computed at the effective noise level:

$$h(\beta, \tilde{\varepsilon}(\beta)) = \tilde{\varepsilon}(\beta)^{1/(\beta+\gamma)} = \left(\lambda(\beta)\varepsilon^2 \log \varepsilon^{-1}\right)^{1/[2(\beta+\gamma)]}.$$

For any $\beta \in S$, where S is the grid defined previously, introduce the linear kernel estimator of the functional $f^{(\alpha_0)}(0)$:

$$T_{\beta,\varepsilon} = \int K_{\beta,h(\beta,\tilde{\varepsilon}(\beta))}(t)dY_\varepsilon(t).$$

Remark 1. For $x_0 \neq 0$ the linear estimator of $f^{(\alpha_0)}(x_0)$ should be taken in a shifted form: $T_{\beta,\varepsilon} = \int K_{\beta,h(\beta,\tilde{\varepsilon}(\beta))}(t-x_0)dY_\varepsilon(t)$. All the constructions and results that follow remain analogous, up to this modification.

Adaptive estimator. The sharp adaptive estimator has the form $T_{\hat{\beta},\varepsilon}$ where $\hat{\beta}$ is a suitably chosen statistic. To define $\hat{\beta}$ we act in the spirit of Lepski (1990, 1991, 1992). That is, the statistic $\hat{\beta}$ is defined as the largest of those β-values in the grid for which the estimator $T_{\beta,\varepsilon}$ does not differ significantly from the estimators corresponding to the smaller β-values. We choose

$$\hat{\beta} = \max\left\{\beta \in S : |T_{\beta,\varepsilon} - T_{\beta',\varepsilon}| \le \eta(\beta') \text{ for all } \beta' \in S,\ \beta' \le \beta\right\}$$

with the threshold

$$\eta(\beta) = d_\varepsilon(\beta)\,\sigma_\beta = \tilde{\varepsilon}(\beta)^{(\beta-r-d/2)/(\beta+\gamma)}\|K_\beta\|_2,$$

where σ_β is the standard deviation of $T_{\beta,\varepsilon}$,

$$\sigma_\beta = \varepsilon\|K_{\beta,h(\beta,\tilde{\varepsilon}(\beta))}\|_2 = \varepsilon h^{-\gamma-r-d/2}(\beta,\tilde{\varepsilon}(\beta))\|K_\beta\|_2. \tag{14}$$

We have that

$$\|K_\beta\|_2 = C_* \left(\frac{2(\beta-r)-d}{2(r+\gamma)+d}\right)^{(\beta+2\gamma+r+d/2)/(2(\beta+\gamma))} \tag{15}$$

where

$$C^* = \left[\frac{1}{2(\beta+\gamma)}\,B\left(1-\frac{2(r+\gamma)+d}{2(\beta+\gamma)}, 1+\frac{2(r+\gamma)+d}{2(\beta+\gamma)}\right) I(d,\alpha_0)\right]^{1/2} \tag{16}$$

where

$$I(d,\alpha_0) = (2\pi)^{-d}\int_{S_d} \xi^{2\alpha_0}d\mu(\xi),$$

$S_d = \{x \in \mathbf{R}^d : \|x\| = 1\}$ for $d = 2,3,\ldots$, $S_1 = [-1,1]$, μ is the Lebesgue measure on S_d so that $\mu(S_d) = 2\pi^{d/2}/\Gamma(d/2)$, $d = 1,2,\ldots$, $\Gamma(\cdot)$ denotes the gamma-function, and $B(a,b) = \int_0^1 t^{a-1}(1-t)^{b-1}dt$ denotes the beta-function.

Finally, define the estimator of $f^{(\alpha_0)}(0)$ as

$$T^*_\varepsilon = T_{\hat{\beta},\varepsilon}. \tag{17}$$

Optimal constant. We will show that the estimator T^*_ε is sharp adaptive and that the exact asymptotical constant $c_{\beta,L}$ for the minimax adaptive risk is given by the expression

$$\begin{aligned} c_{\beta,L} \quad &= \quad C^* L^{(\gamma+r+d/2)/(\beta+\gamma)}\left(\frac{p(2(r+\gamma)+d)}{\beta+\gamma}\right)^{(\beta-r-d/2)/(2(\beta+\gamma))} \\ &\times \left(\frac{2(r+\gamma)+d}{2(\beta-r)-d}\right)^{(\beta-r-d/2)/(2(\beta+\gamma))} \frac{2(\beta+\gamma)}{2(r+\gamma)+d} \end{aligned}$$

where C^* is defined in (16). For $\gamma = 0, d = 1, r = 0$ this yields the constant found in Tsybakov (1998).

The results. To state the results, we fix $p > 0$ and we introduce the maximal risk of an estimator T_ε:

$$\mathcal{R}_{\varepsilon,\beta,L}(T_\varepsilon) = \sup_{f \in \mathcal{F}_{\beta,L}} E_f\left(|T_\varepsilon - f^{(\alpha_0)}(0)|^p\right).$$

Consider the normalizing factor

$$\psi_{\beta,L} = c_{\beta,L}\left(\varepsilon^2 \log \varepsilon^{-1}\right)^{(\beta - r - d/2)/(2(\beta+\gamma))} = c_{\beta,L}\varphi_{\beta,L}. \tag{18}$$

Theorem 1 *Let $p > 0$ and denote $B_\varepsilon = [\beta_*, \beta_\varepsilon] \times [L_*, L^*]$. Then the estimator T_ε^* defined in (17) is sharp adaptive:*

$$\limsup_{\varepsilon \to 0} \sup_{(\beta,L) \in B_\varepsilon} \mathcal{R}_{\varepsilon,\beta,L}(T_\varepsilon^*)\psi_{\beta,L}^{-p} \leq 1 \tag{19}$$

and

$$\liminf_{\varepsilon \to 0} \inf_{T_\varepsilon} \sup_{(\beta,L) \in B_\varepsilon} \mathcal{R}_{\varepsilon,\beta,L}(T_\varepsilon)\psi_{\beta,L}^{-p} \geq 1. \tag{20}$$

Here $\inf_{T_\varepsilon}$ denotes the infimum over all estimators.

Proof of Theorem 1 is given in Section 4. Note that Theorem 1 shows optimality of the constant $c_{\beta,L}$ when the rate of convergence in (18) is chosen equal to $\varphi_{\beta,L}$. Such a choice is motivated by the fact that $\varphi_{\beta,L}$ is the best obtainable rate. More precisely, it is not hard to show that $\varphi_{\beta,L}$ is *adaptive rate of convergence on the scale of classes* $\{\mathcal{F}_{\beta,L}, (\beta, L) \in B_\varepsilon\}$ in the sense of Definition 3 in Tsybakov (1998). In fact, one has the following property which is even stronger than the one required by Definition 3 in Tsybakov (1998).

Theorem 2 *Let $p > 0$. If an estimator $\hat{T}_\varepsilon$ is such that, for some $\beta_0 \geq \beta_*,\ L > 0$,*

$$\limsup_{\varepsilon \to 0} \mathcal{R}_{\varepsilon,\beta_0,L}(\hat{T}_\varepsilon)\psi_{\beta_0,L}^{-p} < 1, \tag{21}$$

then there exists $\beta_0' > \beta_0$ such that

$$\frac{\mathcal{R}_{\varepsilon,\beta_0',L}(\hat{T}_\varepsilon)}{\mathcal{R}_{\varepsilon,\beta_0',L}(T_\varepsilon^*)} \geq \Gamma_\varepsilon \frac{\mathcal{R}_{\varepsilon,\beta_0,L}(T_\varepsilon^*)}{\mathcal{R}_{\varepsilon,\beta_0,L}(\hat{T}_\varepsilon)}, \tag{22}$$

where $\Gamma_\varepsilon \to \infty$, as $\varepsilon \to 0$.

Proof of Theorem 2 is given in Section 4. Theorem 2 means that if an estimator $\hat{T}_\varepsilon$ gains over T_ε^* at least for one smoothness β_0 in terms of a constant factor (cf.(21)), there exists another smoothness β_0' for which $\hat{T}_\varepsilon$ looses much more than it gains at β_0 (cf.(22)). This, together with Theorem 1, shows that the normalizing factor $\psi_{\beta,L}$ cannot be improved in the rate and in the constant.

As a consequence of Theorem 2 and of (19) we get, in particular,

$$\limsup_{\varepsilon\to 0} \mathcal{R}_{\varepsilon,\beta,L}(T_\varepsilon^*)\psi_{\beta,L}^{-p} = 1, \tag{23}$$

for any fixed $\beta \geq \beta_*,\ L > 0$, but this relation is not guaranteed if β depends on ε, for example, if $\beta = \beta_\varepsilon$.

Remark 2. Connection to optimal recovery. A connection of some nonparametric estimation problems to optimal recovery was established by Donoho and Liu (1991), Donoho and Low (1992), and Donoho (1994b). Lepski and Tsybakov (2000) noticed that such a connection exists also for the problems of testing of nonparametric hypotheses. We will explain how the kernel of our estimator is connected to optimal recovery. Let

$$T(f) = f^{(\alpha_0)}(0).$$

The kernel K_β defined in (6) can be characterized as a solution of the optimal recovery problem: K_β satisfies

$$\sup_{\rho_\beta(f)\leq 1,\ \|f-g\|_2\leq 1} \left| \int K_\beta(R_\gamma g) - T(f) \right| = \inf_K \sup_{\rho_\beta(f)\leq 1, |f-g\|_2\leq 1} \left| \int K(R_\gamma g) - T(f) \right|.$$

A way to find K_β is the following. Define the modulus of continuity by

$$\omega_{\beta,L}(\varepsilon) = \sup\left\{T(f) : \|R_\gamma f\|_2 \leq \varepsilon,\ \rho_\beta(f) \leq L\right\}. \tag{24}$$

There exists a function $g_{\beta,L,\varepsilon}$ which attains the supremum of the modulus of continuity, i.e.

$$T(g_{\beta,L,\varepsilon}) = \omega_{\beta,L}(\varepsilon). \tag{25}$$

Then we get K_β by

$$K_\beta = (-1)^r a^{-1} b^{r-2\gamma-d} R_\gamma g_{\beta,1,1} \tag{26}$$

where $a = b^{d/2-\beta}/C^*$ and C^* is defined in (16), and $b = b(\beta)$. We have that

$$g_{\beta,1,1}(x) = (-1)^r a\tilde{g}_\beta(bx) \tag{27}$$

where

$$\tilde{g}_\beta(x) = (2\pi)^{-d} i^r \int_{\mathbf{R}^d} \omega^{\alpha_0} \|\omega\|^{2\gamma} (1 + \|\omega\|^{2(\beta+\gamma)})^{-1} \exp(ix^T\omega) d\omega.$$

For the case where R_γ is equal to the identity operator ($\gamma = 0$), the extremal function $g_{\beta,1,1}$ is given by Taikov (1968) in dimension one ($d = 1$), and by Klemelä and Tsybakov (2001) for general d. The case with $\gamma > 0$, $d = 1$, and $r = 0$ was considered by Donoho and Low (1992, page 959). Klemelä (2003) considers optimal recovery under more general L_p smoothness conditions.

Remark 3. Modulus of continuity and the normalizing factor. We may write the normalizing factor with the help of the modulus of the continuity defined in (24). In fact, for $a \geq 0$, $b > 0$, semi-norm ρ_β satisfies

$$\rho_\beta(af(b\,\cdot\,)) = ab^{\beta-d/2}\rho_\beta(f),$$

functional $T(f) = f^{(\alpha_0)}(0)$ satisfies

$$T(af(b\,\cdot)) = ab^r T(f),$$

and operatior R_γ satisfies

$$R_\gamma(af(b\,\cdot\,)) = ab^{-\gamma}(R_\gamma f)(b\,\cdot\,). \tag{28}$$

According to the terminology of Donoho and Low (1992), ρ_β is a homogeneous functional with the dilation exponent $\beta - d/2$, T is a homogeneous functional with the dilation exponent r, and the operator R_γ is homogeneous with the dilation exponents $-\gamma$ and 1. Note that (28) implies that $\|R_\gamma(af(b\,\cdot\,))\|_2 = ab^{-\gamma-d/2}\|R_\gamma f\|_2$ where $\|\cdot\|_2$ is the $L_2(\mathbf{R}^d)$-norm. Thus, the functional $f \mapsto \|R_\gamma f\|_2$ is homogeneous with the dilation exponent $-\gamma - d/2$. These properties and a renormalization argument entail that

$$g_{\beta,L,\tilde{\varepsilon}(\beta)}(\cdot) = a_1 g_{\beta,1,1}(b_1\,\cdot) \tag{29}$$

where $a_1 = Lb_1^{d/2-\beta}$, $b_1 = (L/\tilde{\varepsilon}(\beta))^{1/(\beta+\gamma)}$, and $g_{\beta,1,1}$ is defined in (27). Thus, the modulus of continuity at the "effective noise" is equal to the normalizing factor,

$$\begin{aligned}\omega_{\beta,L}\left(\tilde{\varepsilon}(\beta)\right) &= T\left(g_{\beta,L,\tilde{\varepsilon}(\beta)}\right) = a_1 b_1^r T\left(g_{\beta,1,1}\right) \\ &= \tilde{\varepsilon}(\beta)^{(\beta-d/2-r)/(\beta+\gamma)} L^{(r+\gamma+d/2)/(\beta+\gamma)} T(g_{\beta,1,1}) \\ &= \psi_{\beta,L} \qquad (30)\end{aligned}$$

where we used the fact that

$$T(g_{\beta,1,1}) = C^* \left(\frac{2(r+\gamma)+d}{2(\beta-d/2-r)}\right)^{(\beta-d/2-r)/(2(\beta+\gamma))} \frac{2(\beta+\gamma)}{2(r+\gamma)+d}, \qquad (31)$$

where C^* is defined in (16).

Remark 4. The adaptive estimation procedure defined above can be transformed for the problems of nonparametric density or regression estimation in the same way as described in Section 4 of Klemelä and Tsybakov (2001).

3 Preliminary lemmas

For β, β' be such that $r' + d/2 \le \beta' \le \beta$ where $r < r' < \beta_* - d/2$ define

$$\tilde{\beta} = \tilde{\beta}(\beta, \beta') = \begin{cases} \beta, & \text{if } \beta/2 - r/2 + d/4 < \beta' \le \beta, \\ \beta' + r + d/2, & \text{if } r' + d/2 \le \beta' \le \beta/2 - r/2 + d/4. \end{cases}$$

The following lemma gives a bound for the bias of a kernel estimator.

Lemma 1 *(i) For $h > 0$,*

$$\sup_{f \in \mathcal{F}_{\beta,L}} \left| E_f \int K_{\beta',h} dY_\varepsilon - f^{(\alpha_0)}(0) \right| \le L h^{\tilde{\beta}-d/2-r} b_{\beta,\beta'}$$

where

$$b_{\beta,\beta'} = (b')^{r+d/2-\tilde{\beta}} \left[(2\pi)^{-d} \int_{\mathbf{R}^d} \frac{\omega^{2\alpha_0} \|\omega\|^{4(\beta'+\gamma)-2\tilde{\beta}}}{(1+\|\omega\|^{2(\beta'+\gamma)})^2} d\omega \right]^{1/2} \qquad (32)$$

with $b' = b(\beta')$, where b is defined in (8).

(ii)

$$\sup_{r'+d/2<\beta'\le\beta<\infty} b_{\beta,\beta'} < \infty, \tag{33}$$

(iii)

$$\limsup_{\delta\to 0} \sup_{\beta,\beta'\in[\beta_*,\infty):|\beta-\beta'|\le\delta} \frac{b_{\beta,\beta'}}{b_{\beta,\beta}} \le 1. \tag{34}$$

Proof. We have $\hat{K}_{\beta',h}(\omega) = (2\pi)^{-d} i^r \omega^{\alpha_0} \|\omega\|^{\gamma} (1 + \|h\omega/b'\|^{2(\beta'+\gamma)})^{-1}$. Hence, using the formula for the Fourier transform of $R_\gamma f$ given in (5), by the Cauchy inequality, for $f \in \mathcal{F}_{\beta,L}$,

$$\begin{aligned}
&\left| E_f \int K_{\beta',h} dY_\varepsilon - f^{(\alpha_0)}(0) \right| \\
&= \left| \int K_{\beta',h}(R_\gamma f) - f^{(\alpha_0)}(0) \right| \\
&= \left| \int_{\mathbf{R}^d} \hat{f}(\omega) \left((2\pi)^d \hat{K}_{\beta',h}(\omega) \|\omega\|^{-\gamma} - i^r \omega^{\alpha_0} \right) d\omega \right| \\
&= \left| \int_{\mathbf{R}^d} \hat{f}(\omega) \left(\frac{i^r \omega^{\alpha_0}}{1 + \|h\omega/b'\|^{2(\beta'+\gamma)}} - i^r \omega^{\alpha_0} \right) d\omega \right| \\
&= \left| \int_{\mathbf{R}^d} \hat{f}(\omega) i^r \omega^{\alpha_0} \frac{\|h\omega/b'\|^{2(\beta'+\gamma)}}{1 + \|h\omega/b'\|^{2(\beta'+\gamma)}} d\omega \right| \\
&\le (h/b')^{\tilde{\beta}-d/2-r} \left[(2\pi)^d \int_{\mathbf{R}^d} \|\omega\|^{2\tilde{\beta}} \left| \hat{f}(\omega) \right|^2 d\omega \right]^{1/2} \\
&\quad \times \left[(2\pi)^{-d} \int_{\mathbf{R}^d} \frac{\omega^{2\alpha_0} \|\omega\|^{4(\beta'+\gamma)-2\tilde{\beta}}}{(1 + \|\omega\|^{2(\beta'+\gamma)})^2} d\omega \right]^{1/2}.
\end{aligned}$$

Because $f \in \mathcal{F}_{\beta,L}$, we get that

$$\begin{aligned}
&(2\pi)^d \int_{\mathbf{R}^d} \|\omega\|^{2\tilde{\beta}} \left| \hat{f}(\omega) \right|^2 d\omega \\
&\le (2\pi)^d \left[\int_{\|\omega\|\le 1} \|\omega\|^{-2\gamma} \left| \hat{f}(\omega) \right|^2 d\omega + \int_{\|\omega\|>1} \|\omega\|^{2\beta} \left| \hat{f}(\omega) \right|^2 d\omega \right] \\
&\le \|R_\gamma f\|_2^2 + \rho_\beta^2(f) \le L^2.
\end{aligned}$$

We have proved (i). Statement in (ii) follows from

$$\begin{aligned}
\int_{\mathbf{R}^d} \frac{\omega^{2\alpha_0} \|\omega\|^{4(\beta'+\gamma)-2\tilde{\beta}}}{(1 + \|\omega\|^{2(\beta'+\gamma)})^2} d\omega &\le \int_{\mathbf{R}^d} \frac{\|\omega\|^{4(\beta'+\gamma)+2r-2\tilde{\beta}}}{(1 + \|\omega\|^{2(\beta'+\gamma)})^2} d\omega \\
&\le \mu(S_d) \int_0^\infty \frac{t^{4(\beta'+\gamma)+2r-2\tilde{\beta}+d-1}}{(1 + t^{2(\beta'+\gamma)})^2} dt \\
&\le \mu(S_d) \left(\int_0^1 t^{4(\beta'+\gamma)+2r-2\tilde{\beta}+d-1} dt + \int_1^\infty t^{2r-2\tilde{\beta}+d-1} dt \right).
\end{aligned}$$

This integral is finite because, by the definition of $\tilde{\beta}$, $r + d/2 < \tilde{\beta} < 2\beta' + r + d/2$. Statement (iii) follows from the representation

$$\begin{aligned} b_{\beta,\beta'} &= (b')^{r+d/2-\beta} I^{1/2}(d,\alpha_0) \qquad (35) \\ &\times \left[\frac{1}{2(\beta'+\gamma)} B\left(1 + \frac{2(r+\gamma+\beta'-\beta)+d}{2(\beta'+\gamma)}, 1 - \frac{2(r+\gamma+\beta'-\beta)+d}{2(\beta'+\gamma)}\right)\right]^{1/2} \end{aligned}$$

where $B(\cdot,\cdot)$ is the Beta-function. ■

We need an exponential bound for the stochastic part of the estimator. Define

$$Z_\beta = \varepsilon \int K_{\beta,h(\beta,\tilde{\varepsilon}(\beta))}(t)dW(t). \qquad (36)$$

Lemma 2 *Let σ_β be as defined in (14). Then for $u > 0$, $p \geq 0$,*

$$E\left[|Z_\beta|^p I(|Z_\beta| \geq u)\right] \leq D(p)\left(\sigma_\beta^p + u^p\right)\exp\left\{-\frac{u^2}{2\sigma_\beta^2}\right\}.$$

where $D(p) > 0$ is a constant depending only on p.

Proof of this lemma is straightforward since $Z_\beta \sim \mathcal{N}(0, \sigma_\beta^2)$.

We will need the the following facts concerning the kernel K_β and the bias constant $b_{\beta,\beta}$, defined in (32). For $\delta > 0$ set

$$\Omega(\delta) = \sup\{\|K_\beta - K_{\beta'}\|_2 : \ |\beta - \beta'| \leq \delta, \ \beta,\beta' \in [r' + d/2, \infty)\}. \qquad (37)$$

Lemma 3 *Let Ω be defined by (37). Then there exists a positive constant c_0 such that*

$$\Omega(\delta) \leq c_0\delta, \quad 0 < \delta \leq 1. \qquad (38)$$

There exist positive constants c_1, c_2 such that for $r' + d/2 \leq \beta < \infty$,

$$c_1 \leq \|K_\beta\|_2 \leq c_2\beta. \qquad (39)$$

Let $g_{\beta,1,1}$ be defined in (27) and $T(f) = f^{(\alpha_0)}(0)$. Then

$$T(g_{\beta,1,1}) = \|K_\beta\|_2 + b_{\beta,\beta}. \qquad (40)$$

Finally, there exist positive constants c_3, c_4 such that for $r' + d/2 \leq \beta < \infty$,

$$c_3 \leq T(g_{\beta,1,1}) \leq c_4. \qquad (41)$$

Proof. To prove (38), note that by the definition of the kernel function,

$$\|K_\beta - K_{\beta'}\|_2^2 = I(d,\alpha_0)\int_0^\infty \frac{t^{2\gamma+2r+d-1}((t/b)^{2\beta}-(t/b')^{2\beta'})^2}{(1+(t/b)^{2(\beta+\gamma)})^2(1+(t/b')^{2(\beta'+\gamma)})^2}\,dt$$

where $b = b(\beta), b' = b(\beta')$. Note that, for $r' + d/2 \le \beta < \infty$ and any $t > 0$, the following bound holds for the derivative:

$$\Big|\frac{d}{d\beta}\Big[\Big(\frac{t}{b(\beta)}\Big)^{2\beta}\Big]\Big| \le \Big(2|\log t| + \frac{C|\log\beta|}{\beta+1}\Big)\Big(\frac{t}{b(\beta)}\Big)^{2\beta} \le C(|\log t|+1)\Big(\frac{t}{b(\beta)}\Big)^{2\beta}. \quad (42)$$

Here and further in this proof we use the same notation C for different positive constants that depend only on γ, r', r, d. Assume w.l.o.g. that $\beta \le \beta' \le \beta+\delta \le \beta+1$ (the last inequality follows since $\delta \le 1$). From the definition of $b(\beta)$, it is easy to obtain that there exist constants $C > 0$, $c' > 0, c'' > 0$ depending only on γ, r', r, d, such that

$$b(\beta) \ge C, \quad b(\beta)^{2\beta} \ge C, \quad c' \le \frac{b(\beta)^{2\beta}}{b(\beta')^{2\beta'}} \le c'', \quad (43)$$

for $r' + d/2 \le \beta < \infty$, $\beta \le \beta' \le \beta + 1$.

Using (42) and (43), we find

$$\|K_\beta - K_{\beta'}\|_2^2 \le C|\beta-\beta'|^2 \int_0^\infty \frac{t^{2\gamma+2r+d-1}(|\log t|+1)^2 \max((t/b)^{4\beta},(t/b')^{4\beta'})}{(1+(t/b)^{2(\beta+\gamma)})^2(1+(t/b')^{2(\beta'+\gamma)})^2}\,dt$$

$$\le C|\beta-\beta'|^2\Big(\int_0^1 (|\log t|+1)^2 t^{2\gamma+2r+d-1}dt + \int_1^\infty (|\log t|+1)^2 t^a dt\Big)$$

where $a = 2\gamma + 2r + d - 1 - 4(\beta+\gamma) - 4(\beta'+\gamma) + 4\max(\beta,\beta')$. The last but one integral here is bounded since $2\gamma + 2r + d - 1 \ge 0$, while the last integral is bounded since the inequalities $\beta \ge r + d/2$, $\beta \le \beta'$ imply that $a \le -6\gamma - 2r - d - 1 < -1$. This finishes the proof of (38).

The proof of (39) follows from the equation (15). Equations (40) and (41) follow from (15), (31), and (35). ■

Finally we will need the following lemma.

Lemma 4 *Let $\beta, \beta' \in [r' + d/2, \infty)$, $r < r' < \infty$, $\beta' < \beta$, $L \in [L_*, L^*]$, and denote $\nu = (\beta, L)$. Then there exist positive constants $D_1, \ldots, D_4$, that can depend only on $\beta_*, L_*, L^*, r, r', \gamma, d, p$, such that, denoting*

$$\kappa(\beta) = (\beta - d/2 - r)/(\beta + \gamma),$$

we have

$$\frac{\psi_{\beta',L}}{\psi_\nu} \le D_1 \varepsilon^{\kappa(\beta')-\kappa(\beta)} = D_1 \exp\left\{\frac{1}{2p}\left[d_\varepsilon^2(\beta') - d_\varepsilon^2(\beta)\right]\right\}, \tag{44}$$

$$\frac{\psi_{\beta',L}}{\psi_\nu} \ge \frac{D_2}{\beta_\varepsilon + \gamma}\left(\varepsilon^2 \log \varepsilon^{-1}\right)^{(\kappa(\beta')-\kappa(\beta))/2}, \tag{45}$$

and

$$D_3 \le \frac{\psi_{\beta,L}}{\eta(\beta)} \le D_4. \tag{46}$$

Proof. From (30) we have $\psi_{\beta,L} = \tilde{\varepsilon}^{\kappa(\beta)} L^{(r+\gamma+d/2)/(\beta+\gamma)} T(g_{\beta,1,1})$ where $\tilde{\varepsilon} = \varepsilon d_\varepsilon(\beta)$. Thus,

$$\frac{\psi_{\beta',L}}{\psi_\nu} = \varepsilon^{\kappa(\beta')-\kappa(\beta)} \frac{d_\varepsilon(\beta')^{\kappa(\beta')}}{d_\varepsilon(\beta)^{\kappa(\beta)}} \frac{T(g_{\beta',1,1})}{T(g_{\beta,1,1})} \frac{L^{(r+\gamma+d/2)/(\beta'+\gamma)}}{L^{(r+\gamma+d/2)/(\beta+\gamma)}}.$$

By definition,

$$d_\varepsilon^2(\beta) = 2p(1-\kappa(\beta)) \log \varepsilon^{-1} = 2p\,\frac{r+\gamma+d/2}{\beta+\gamma}\,\log \varepsilon^{-1} \tag{47}$$

and thus, for $\beta \in [r' + d/2, \infty)$,

$$\frac{2p(r+\gamma+d/2)}{\beta_\varepsilon+\gamma}\,\log \varepsilon^{-1} \le d_\varepsilon^2(\beta) \le 2p \log \varepsilon^{-1}. \tag{48}$$

Therefore,

$$\frac{D_5}{\beta_\varepsilon+\gamma}\left(\log \varepsilon^{-1}\right)^{(\kappa(\beta')-\kappa(\beta))/2} \le \frac{d_\varepsilon(\beta')^{\kappa(\beta')}}{d_\varepsilon(\beta)^{\kappa(\beta)}} \le D_6 \left(\log \varepsilon^{-1}\right)^{(\kappa(\beta')-\kappa(\beta))/2}$$

for $D_5, D_6 > 0$. Observing that $(\kappa(\beta') - \kappa(\beta))/2 < 0$ and using (41) we obtain (44) and (45). To prove the bound (46) note that, by (30) and by the definition of $\eta(\beta)$,

$$\frac{\psi_{\beta,L}}{\eta(\beta)} = L^{(r+\gamma+d/2)/(\beta+\gamma)} \frac{T(g_{\beta,1,1})}{\|K_\beta\|_2}.$$

The lower bound follows from (40). The upper bound follows from (39) and (41). ■

4 Proofs

We denote, as before, $T(f) = f^{(\alpha_0)}(0)$. We will denote by $C, C', C_1, C_2, \ldots$ positive constants and by $\gamma_{\varepsilon i}, i = 1, 2, \ldots$ the functions of ε such that $\lim_{\varepsilon\to 0} \gamma_{\varepsilon i} = 0$. These constants and functions can depend only on $\beta_*, L_*, L^*, r, r', \gamma, d, p$ and may be different in different occasions.

4.1 Proof of the upper bound in Theorem 1.

Here we prove the bound (19). Consider $\nu = (\beta, L) \in B_\varepsilon$ and define $\beta^- = \beta^-(\beta)$ by

$$\beta^- = \beta - \frac{\beta_\varepsilon^+}{\log(1/\varepsilon)}$$

where

$$\beta_\varepsilon^+ = \left(\log\log\varepsilon^{-1}\right)^{\delta_3}$$

and $\delta_2 < \delta_3 < 1$ where δ_2 is from the definition of β_ε in (10). We have

$$\mathcal{R}_{\varepsilon,\nu}(T_\varepsilon^*)\psi_\nu^{-p} = \sup_{f\in\mathcal{F}_\nu} E_f\left(\psi_\nu^{-p}\left|T_\varepsilon^* - T(f)\right|^p\right) = R_{\varepsilon,\nu}^- + R_{\varepsilon,\nu}^+$$

where

$$R_{\varepsilon,\nu}^- = \sup_{f\in\mathcal{F}_\nu} E_f\left(\psi_\nu^{-p}\left|T_\varepsilon^* - T(f)\right|^p I\left(\hat\beta < \beta^-\right)\right),$$

$$R_{\varepsilon,\nu}^+ = \sup_{f\in\mathcal{F}_\nu} E_f\left(\psi_\nu^{-p}\left|T_\varepsilon^* - T(f)\right|^p I\left(\hat\beta \geq \beta^-\right)\right).$$

To show (19), we will prove that

$$\lim_{\varepsilon\to 0}\sup_{\nu\in B_\varepsilon} R_{\varepsilon,\nu}^- = 0 \tag{49}$$

and

$$\limsup_{\varepsilon\to 0}\sup_{\nu\in B_\varepsilon} R_{\varepsilon,\nu}^+ \leq 1. \tag{50}$$

4.1.1 Proof of (49).

Let $\beta \in [\beta_*, \beta_\varepsilon]$, $\beta' \in S$, $\beta' < \beta^-$, $L \in [L_*, L^*]$. Let $f \in \mathcal{F}_\nu$, $\nu = (\beta, L)$. Then for sufficiently small ε, using Lemma 1 and the fact that $\beta' \leq \tilde\beta = \tilde\beta(\beta, \beta')$, we get

$$|E_f T_{\beta',\varepsilon} - T(f)| \leq L h^{\tilde\beta - d/2 - r}(\beta', \tilde\varepsilon(\beta'))b_{\beta,\beta'} \leq L h^{\beta' - d/2 - r}(\beta', \tilde\varepsilon(\beta'))b_{\beta,\beta'}. \tag{51}$$

By definition,

$$h^{\beta - d/2 - r}(\beta, \tilde\varepsilon(\beta)) = \tilde\varepsilon^{\kappa(\beta)}(\beta, \tilde\varepsilon(\beta)) \tag{52}$$

where $\kappa(\beta) = (\beta - d/2 - r)/(\beta + \gamma)$. Comparing this to (30) and using (33), (41), we get

$$L h^{\beta' - d/2 - r}(\beta', \tilde\varepsilon(\beta'))b_{\beta,\beta'} \leq C_1\psi_{\beta',L}. \tag{53}$$

Then

$$|T_{\beta',\varepsilon} - T(f)| \leq |E_f T_{\beta',\varepsilon} - T(f)| + |Z_{\beta'}| \leq C_1 \psi_{\beta',L} + |Z_{\beta'}| \tag{54}$$

where $Z_{s'}$ is defined in (36). Define

$$\tau(\beta') = \sigma_{\beta'} \left[(d_\varepsilon^2(\beta') - d_\varepsilon^2(\beta))^{1/2} + \left(\log \varepsilon^{-1}\right)^{1/4} \right]$$

where $\sigma_{\beta'}$ is defined in (14) and $d_\varepsilon(\beta)$ is defined in (13). By definitions, (48), and by the definition of β_ε in (10),

$$\frac{\tau(\beta')}{\eta(\beta')} = \frac{(d_\varepsilon^2(\beta') - d_\varepsilon^2(\beta))^{1/2} + (\log \varepsilon^{-1})^{1/4}}{d_\varepsilon(\beta')} \leq C \left[1 + (\log \varepsilon^{-1})^{-1/2} (\beta_\varepsilon + \gamma)^{1/2} \right] \leq C_1.$$

Next, by (46),

$$\frac{\eta(\beta')}{\psi_\nu} \leq C_2 \frac{\psi_{\beta',L}}{\psi_\nu}.$$

Combining the two previous inequalities we get

$$\frac{\tau(\beta')}{\psi_\nu} \leq C_3 \frac{\psi_{\beta',L}}{\psi_\nu}. \tag{55}$$

Using this and (54) we find

$$R_{\varepsilon,\nu}^- \leq \sum_{\beta' \in S, \beta' < \beta^-} \sup_{f \in \mathcal{F}_\nu} E_f \left(\psi_\nu^{-p} \left| T_{\varepsilon,\beta'} - T(f) \right|^p I \left(\hat{\beta} = \beta' \right) \right) \leq g_1(\nu) + g_2(\nu)$$

where

$$g_1(\nu) = C \sum_{\beta' \in S, \beta' < \beta^-} \sup_{f \in \mathcal{F}_\nu} P_f \left(\hat{\beta} = \beta' \right) \left(\psi_\nu^{-1} \psi_{\beta',L} \right)^p \tag{56}$$

and

$$g_2(\nu) = C \sum_{\beta' \in S, \beta' < \beta^-} E \left[\psi_\nu^{-p} \left(\psi_{\beta',L} + |Z_{\beta'}| \right)^p I \left(|Z_{\beta'}| \geq \tau(\beta') \right) \right]. \tag{57}$$

Let us prove that the probability of underestimating largely the value of β by the statistic $\hat{\beta}$ is small, uniformly over $f \in \mathcal{F}_\nu$.

Lemma 5 *Let $\beta \in [\beta_*, \infty)$, $\beta' \in S$, $\beta' < \beta^-$, $L \in [L_*, L^*]$, and $\nu = (\beta, L)$. Then,*

$$\sup_{f \in \mathcal{F}_\nu} P_f \left(\hat{\beta} = \beta' \right) \leq Cm \exp \left\{ -\frac{1}{2} d_\varepsilon^2(\beta')(1 - \delta) \right\}$$

where $m = \mathrm{Card}(S)$, $\delta = \delta(\varepsilon) = \alpha_\varepsilon + \alpha'_\varepsilon$ with

$$\alpha_\varepsilon = C_4 \exp \left\{ -C_5 (\beta_\varepsilon + \gamma)^{-1} \beta_\varepsilon^+ \right\}$$

and

$$\alpha'_\varepsilon = C_6 \max\left\{\beta_\varepsilon \exp\left\{-\frac{2}{(\beta_\varepsilon+\gamma)^2}(\log \varepsilon^{-1})^{1/2}\right\}, c_0(\log\varepsilon^{-1})^{-1/2}\right\}$$

where c_0 is the constant from (38).

Proof. Since $\mathrm{Card}(S) = m$, we have

$$\begin{aligned}\sup_{f\in\mathcal{F}_\nu} P_f\left(\hat\beta = \beta'\right) &\le \sum_{\beta''\in S, \beta''\le\beta'} \sup_{f\in\mathcal{F}_\nu} P_f\left(\left|T_{\bar\beta',\varepsilon} - T_{\beta'',\varepsilon}\right| > \eta(\beta'')\right) \\ &\le m \max_{\beta''\in S,\beta''\le\beta'} \sup_{f\in\mathcal{F}_\nu} P_f\left(\left|T_{\bar\beta',\varepsilon} - T_{\beta'',\varepsilon}\right| > \eta(\beta'')\right) \qquad (58)\end{aligned}$$

where $\bar\beta' = \bar\beta'(\beta')$ is the smallest element of S greater than β'. Let $f \in \mathcal{F}_\nu$. Using Lemma 1 and (33), for $\beta'' \in S$, $\beta'' \le \beta'$,

$$|E_f T_{\beta'',\varepsilon} - T(f)| \le L h^{\tilde\beta - d/2 - r}(\beta'', \tilde\varepsilon(\beta''))b_{\beta,\beta''} \le C h^{\tilde\beta-d/2-r}(\beta'', \tilde\varepsilon(\beta''))$$

where $\tilde\beta = \tilde\beta(\beta, \beta'')$ is as in Lemma 1. We have that $\psi_{\beta'',L} \ge C\tilde\varepsilon(\beta'')^{(\beta''-d/2-r)/(\beta''+\gamma)}$ by (30) and (41). Thus,

$$\begin{aligned}\frac{h^{\tilde\beta-d/2-r}(\beta'', \tilde\varepsilon(\beta''))}{\psi_{\beta'',L}} &\le C\tilde\varepsilon(\beta'')^{(\tilde\beta-d/2-r)/(\beta''+\gamma)-(\beta''-d/2-r)/(\beta''+\gamma)} \\ &= \exp\left\{-\frac{\tilde\beta-\beta''}{\beta''+\gamma}\log\varepsilon^{-1}\right\} d_\varepsilon(\beta'')^{(\tilde\beta-\beta'')/(\beta''+\gamma)} \\ &\le \exp\left\{-\frac{\tilde\beta-\beta''}{\beta_\varepsilon+\gamma}\left(\log\varepsilon^{-1} - \log\log\varepsilon^{-1}\right)\right\}\end{aligned}$$

by (48). We have

$$\tilde\beta - \beta'' \ge \min\{\beta-\beta'', r+d/2\} \ge \min\{\beta-\beta^-, r+d/2\} \ge \frac{\beta_\varepsilon^+}{\log\varepsilon^{-1}}.$$

Using similar inference for $\left|E_f T_{\bar\beta',\varepsilon} - T(f)\right|$ we get that

$$\left|E_f T_{\bar\beta',\varepsilon} - T(f)\right| + |E_f T_{\beta'',\varepsilon} - T(f)| \le C\gamma_\varepsilon \psi_{\beta'',L}$$

where

$$\gamma_\varepsilon = C\exp\{-C(\beta_\varepsilon+\gamma)^{-1}\beta_\varepsilon^+\}.$$

Using (46) we obtain $\psi_{\beta'',L} \le D_5\eta(\beta'')$ and thus

$$\begin{aligned}\left|T_{\bar\beta',\varepsilon} - T_{\beta'',\varepsilon}\right| &\le \left|E_f T_{\bar\beta',\varepsilon} - T(f)\right| + |E_f T_{\beta'',\varepsilon} - T(f)| + \left|Z_{\bar\beta'} - Z_{\beta''}\right| \\ &\le C\gamma_\varepsilon\eta(\beta'') + \left|Z_{\bar\beta'} - Z_{\beta''}\right|.\end{aligned}$$

Hence, for sufficiently small ε,

$$P_f\left(\left|T_{\bar\beta',\varepsilon}-T_{\beta'',\varepsilon}\right|>\eta(\beta'')\right)\le P_f\left(\left|Z_{\bar\beta'}-Z_{\beta''}\right|>\eta(\beta'')(1-C\gamma_\varepsilon)\right). \qquad (59)$$

Denote $h_0=h(\beta'',\tilde\varepsilon(\beta''))$, $h_1=h(\bar\beta',\tilde\varepsilon(\bar\beta'))$, $\bar K_0=K_{\beta''}$, $\bar K_1=K_{\bar\beta'}$, and, as before, $\bar K_{i,h_i}=h_i^{-\gamma-r-d}\bar K_i(\cdot/h_i)$. Now $Z_{\bar\beta'}-Z_{\beta''}\sim\mathcal N(0,\varepsilon^2V^2)$ where

$$V=\left\|\bar K_{0,h_0}-\bar K_{1,h_1}\right\|_2\le\left\|\bar K_{0,h_0}-\bar K_{0,h_1}\right\|_2+\left\|\bar K_{0,h_1}-\bar K_{1,h_1}\right\|_2\overset{\text{def}}{=}V_1+V_2.$$

Now

$$V_1^2=h_0^{-2(\gamma+r)-d}\int\left[\bar K_0(x)-\left(\frac{h_0}{h_1}\right)^{\gamma+r+d}\bar K_0\left(\frac{h_0}{h_1}x\right)\right]^2dx \qquad (60)$$

$$=h_0^{-2(\gamma+r)-d}\left[\int\bar K_0^2+\left(\frac{h_0}{h_1}\right)^{2(r+\gamma)+d}\int\bar K_0^2-2\left(\frac{h_0}{h_1}\right)^{\gamma+r+d}\int\bar K_0(x)\bar K_0\left(\frac{h_0}{h_1}x\right)dx\right].$$

We have,

$$\begin{aligned}\frac{h_0}{h_1}&=\exp\left\{\frac{(\beta''-\bar\beta')}{(\beta''+\gamma)(\bar\beta'+\gamma)}\left(\log\varepsilon^{-1}-\frac12\log\log\varepsilon^{-1}\right)\right\}\frac{\lambda(\beta'')^{1/(2(\beta''+\gamma))}}{\lambda(\bar\beta')^{1/(2(\bar\beta'+\gamma))}}\\&\le\exp\left\{-\frac{C}{(\beta_\varepsilon+\gamma)^2}(\bar\beta'-\beta'')\log\varepsilon^{-1}\right\}\frac{\lambda(\beta'')^{1/(2(\beta''+\gamma))}}{\lambda(\bar\beta')^{1/(2(\bar\beta'+\gamma))}}.\end{aligned} \qquad (61)$$

Since, $\bar\beta'>\beta''$ and by the definition of the grid in (9) it suffices to consider the two cases:

$$(i)\quad \bar\beta'-\beta''\ge(\log\varepsilon^{-1})^{-1/2}$$

and

$$(ii)\quad \kappa_1(\log\varepsilon^{-1})^{-\delta_1}<\bar\beta'-\beta''<(\log\varepsilon^{-1})^{-1/2}$$

where $\delta_1>1$ is the parameter from the definition of the grid (9).

In the case (i), $\lim_{\varepsilon\to0}(h_0/h_1)=0$, and in the case (ii), $\limsup_{\varepsilon\to0}(h_0/h_1)\le1$. Thus, in the both cases, for sufficiently small ε,

$$\begin{aligned}&\left(\frac{h_0}{h_1}\right)^{-\gamma-r}\int\bar K_0(x)\bar K_0\left(\frac{h_0}{h_1}x\right)dx\\&=\int\frac{\omega^{2\alpha_0}\|\omega\|^{2\gamma}}{(1+\|\omega/b''\|^{2(\beta+\gamma)})(1+\|(h_0/h_1)(\omega/b'')\|^{2(\beta+\gamma)})}d\omega\\&\ge\frac12\int\frac{\omega^{2\alpha_0}\|\omega\|^{2\gamma}}{(1+\|\omega/b''\|^{2(\beta+\gamma)})^2}d\omega=\frac12\int\bar K_0^2\end{aligned} \qquad (62)$$

where $b'' = b(\beta'')$. From (60) and (62),

$$V_1^2 \leq h_0^{2(q-r)} \int \bar{K}_0^2. \tag{63}$$

Also,

$$V_2^2 = h_1^{-2(\gamma+r)-d} \left\| \bar{K}_0 - \bar{K}_1 \right\|_2^2 = h_0^{-2(\gamma-r)-d} \left\| \bar{K}_0 \right\|_2^2 \gamma'_\varepsilon \tag{64}$$

where

$$\gamma'_\varepsilon = \gamma'_\varepsilon(\bar{\beta}', \beta'') = \left(\frac{h_0}{h_1} \right)^{2(r+\gamma)+d} \frac{\left\| \bar{K}_0 - \bar{K}_1 \right\|_2^2}{\left\| \bar{K}_0 \right\|_2^2} \longrightarrow 0$$

as $\varepsilon \to 0$. Indeed, to prove this consider again the two cases: (i) $\bar{\beta}' - \beta'' \geq (\log \varepsilon^{-1})^{-1/2}$ and (ii) $\kappa_1 (\log \varepsilon^{-1})^{-\delta_1} < \bar{\beta}' - \beta'' < (\log \varepsilon^{-1})^{-1/2}$. If (i) holds then by (61),

$$\frac{h_0}{h_1} \leq C \exp \left\{ - \frac{C}{(\beta_\varepsilon + \gamma)^2} (\log \varepsilon^{-1})^{1/2} \right\},$$

and, using (39), we get

$$\gamma'_\varepsilon \leq C \beta_\varepsilon \exp \left\{ - \frac{C}{(\beta_\varepsilon + \gamma)^2} (\log \varepsilon^{-1})^{1/2} \right\}.$$

If (ii) holds, then (61) and the inequality $\bar{\beta}' > \beta''$ entail $h_0/h_1 \leq C$, while

$$\left\| \bar{K}_0 - \bar{K}_1 \right\|_2 \leq \Omega(\bar{\beta}' - \beta'') \leq \Omega((\log \varepsilon^{-1})^{-1/2}) \leq c_0 (\log \varepsilon^{-1})^{-1/2}$$

where Ω is defined in (37) and we used (38) (we assume here and later that ε is small enough to guarantee that $(\log \varepsilon^{-1})^{-1/2} \leq 1$). Combining cases (i) and (ii) we have

$$\gamma'_\varepsilon \leq C \max \left\{ \beta_\varepsilon \exp \left\{ - \frac{2}{(\beta_\varepsilon + \gamma)^2} (\log \varepsilon^{-1})^{1/2} \right\}, c_0 (\log \varepsilon^{-1})^{-1/2} \right\}.$$

From (63) and (64) it follows that for sufficiently small ε,

$$\varepsilon^2 V^2 \leq \varepsilon^2 h^{-2(\gamma+r)-d}(\beta'', \tilde{\varepsilon}(\beta'')) \|K_{\beta''}\|_2^2 (1 + \gamma'_\varepsilon) = \sigma_{\beta''}^2 (1 + \gamma'_\varepsilon).$$

Now, for sufficiently small ε,

$$\frac{\eta^2(\beta'')}{\varepsilon^2 V^2} \geq \frac{\eta^2(\beta'')}{\sigma_{\beta''}^2 (1 + \gamma'_\varepsilon)} = \frac{d_\varepsilon^2(\beta'')}{1 + \gamma'_\varepsilon} \geq \frac{d_\varepsilon^2(\beta')}{1 + \gamma'_\varepsilon}.$$

Thus, using Lemma 2 we get

$$
\begin{aligned}
P_f\left(\left|Z_{\bar\beta'} - Z_{\beta''}\right| > \eta(\beta'')(1-C\gamma_\varepsilon)\right) &\leq C\exp\left\{-\frac{1}{2\varepsilon^2V^2}\,\eta^2(\beta'')(1-C\gamma_\varepsilon)^2\right\} \\
&\leq C\exp\left\{-\frac{1}{2}\,d_\varepsilon^2(\beta')(1-C\gamma_\varepsilon)^2(1+\gamma_\varepsilon')^{-1}\right\} \\
&\leq C\exp\left\{-\frac{1}{2}\,d_\varepsilon^2(\beta')(1-\delta(\varepsilon))\right\}
\end{aligned}
$$

where $\delta(\varepsilon) = \gamma_\varepsilon' + 2C\gamma_\varepsilon$. Comparing this to (58) and (59) we get the lemma. ■

Lemma 6 *Let g_1 be defined in (56). Then,*

$$\lim_{\varepsilon\to 0}\sup_{\nu\in B_\varepsilon} g_1(\nu) = 0.$$

Proof. Let $\beta\in[\beta_*,\beta_\varepsilon]$, $\beta'\in S$, $\beta'<\beta^-$, $L\in[L_*,L^*]$, $\nu=(\beta,L)$. Using Lemma 5 and (44) we find

$$
\begin{aligned}
g_1(\nu) &\leq C\ m \sum_{\beta'\in S,\beta'<\beta^-}\left[\exp\left\{-\frac{1}{2}\,d_\varepsilon^2(\beta')(1-\delta(\varepsilon))\right\}\right]\exp\left\{\frac{1}{2}\,(d_\varepsilon^2(\beta')-d_\varepsilon^2(\beta))\right\} \\
&\leq C'\ m^2\exp\left\{-p(r+\gamma+d/2)\left(\frac{1}{\beta_\varepsilon+\gamma}-\frac{\delta(\varepsilon)}{r+\gamma+d/2}\right)\log\varepsilon^{-1}\right\}.
\end{aligned}
$$

By (9),

$$m \leq \beta_\varepsilon\kappa_1^{-1}(\log(1/\varepsilon))^{\delta_1}. \tag{65}$$

The Lemma follows now by using the definition of $\delta(\varepsilon)$ in the statement of Lemma 5, the assumption (10), and (65). ■

Lemma 7 *Let g_2 be defined in (57). Then,*

$$\lim_{\varepsilon\to 0}\sup_{\nu\in B_\varepsilon} g_2(\nu) = 0.$$

Proof. Let $\beta\in[\beta_*,\beta_\varepsilon]$, $\beta'\in S$, $\beta'<\beta^-$, $L\in[L_*,L^*]$, $\nu=(\beta,L)$. Now,

$$\frac{\tau^2(\beta')}{\sigma_{\beta'}^2} = \left[(d_\varepsilon^2(\beta')-d_\varepsilon^2(\beta))^{1/2}+\left(\log\varepsilon^{-1}\right)^{1/4}\right]^2 \geq d_\varepsilon^2(\beta')-d_\varepsilon^2(\beta)+\left(\log\varepsilon^{-1}\right)^{1/2}. \tag{66}$$

In particular,

$$\sigma_{\beta'} \leq \tau^2(\beta'). \tag{67}$$

Lemma 2, (67), and (55) yield

$$\begin{aligned} g_2(\nu) &\le C \sum_{\beta' \in S, \beta' < \beta^-} \psi_\nu^{-p} \left[\psi_{\beta',L} + \left(\sigma_{\beta'}^p + \tau^p(\beta')\right)\right] \exp\left\{-\frac{\tau^2(\beta')}{2\sigma_{\beta'}^2}\right\} \\ &\le C' \sum_{\beta' \in S, \beta' < \beta^-} \psi_\nu^{-p} \psi_{\beta',L} \exp\left\{-\frac{\tau^2(\beta')}{2\sigma_{\beta'}^2}\right\}. \end{aligned}$$

Applying (44) and (66) to this upper bound we get

$$g_2(\nu) \le C\, m\ \exp\left\{-\frac{1}{2}\left(\log \varepsilon^{-1}\right)^{1/2}\right\}.$$

Lemma follows from this, (65), and definition of β_ε in (10). ■

Lemmas 6 and 7 imply (49).

4.1.2 Proof of (50)

Define the bias term:

$$B(\beta, L, \varepsilon) = L h_l^{\beta - d/2 - r}(\beta, L, \varepsilon) b_{\beta,\beta}$$

and the standard deviation term:

$$R(\beta, L, \varepsilon) = \varepsilon h_l^{-\gamma - d/2 - r}(\beta, L, \varepsilon) \|K_\beta\|_2$$

where

$$h_l(\beta, L, \varepsilon) = (\varepsilon/L)^{1/(\beta+\gamma)}.$$

From (40) and from formula (30) we get a decomposition of the normalising factor to the bias and variance components:

$$\psi_\nu = B(\beta, L, \tilde\varepsilon(\beta)) + R(\beta, L, \tilde\varepsilon(\beta)). \tag{68}$$

Let $\beta \in [\beta_*, \beta_\varepsilon]$, $L \in [L_*, L^*]$, $\nu = (\beta, L)$. Let $\bar\beta = \bar\beta(\beta)$ be defined by

$$\bar\beta = \beta - \frac{2(\beta+\gamma)\log L}{2\log(1/\varepsilon) - \log\log(1/\varepsilon) + 2\log L}.$$

That is, $\bar\beta$ is chosen so that

$$\left(L^{-2}\varepsilon^2 \log \varepsilon^{-1}\right)^{1/(2(\beta+\gamma))} = \left(\varepsilon^2 \log \varepsilon^{-1}\right)^{1/(2(\bar\beta+\gamma))}. \tag{69}$$

Let $\beta^+ \in S$ be the largest grid point $\leq \bar{\beta}$. Denote $\mathcal{S}_1 = \mathcal{S}_1(\beta) = \{\beta' \in S : \beta^- \leq \beta' \leq \beta^+\}$ and $\mathcal{S}_2 = \mathcal{S}_2(\beta) = \{\beta' \in S : \beta^+ < \beta' \leq \beta_\varepsilon\}$. Assume that ε is small enough, so that $\beta^- < \beta^+$. We have

$$R^+_{\varepsilon,\nu} = \sup_{f\in\mathcal{F}_\nu} E_f \left(\psi_\nu^{-p} \left|T^*_\varepsilon - T(f)\right|^p I\left(\hat{\beta} \in \mathcal{S}_1 \cup \mathcal{S}_2\right)\right).$$

Let $\beta' \in \mathcal{S}_1$ and $f \in \mathcal{F}_\nu$. Assume that ε is so small that $\tilde{\beta}(\beta,\beta') = \beta$, with $\tilde{\beta}$ from Lemma 1. Using successively Lemma 1, the fact that $\beta' \leq \bar{\beta}$ and (69) we get (for details see Klemelä and Tsybakov, 2001),

$$|E_f T_{\beta',\varepsilon} - T(f)| \leq \Lambda(\beta,\beta') L h_l^{\beta-d/2-r}(\beta, L, \tilde{\varepsilon}(\beta)) b_{\beta,\beta'} \tag{70}$$

where $\Lambda(\beta,\beta') = \lambda(\beta')^{(\beta-d/2-r)/(2(\beta'+\gamma))}\lambda(\beta)^{-(\beta-d/2-r)/(2(\beta+\gamma))}$ and $\lambda(\beta) = 2p(r+\gamma+d/2)/(\beta+\gamma)$. Note that

$$|\beta - \beta'| \leq \frac{C\beta^+_\varepsilon}{\log(1/\varepsilon)}, \quad \text{for all } \beta' \in \mathcal{S}_1. \tag{71}$$

This and the uniform continuity of $\Lambda(\beta,\beta')$ in $\beta,\beta' \in [\beta_*,\infty)$ yields that $\Lambda(\beta,\beta') \leq 1+\gamma_{\varepsilon 1}$. Next, using (34) and (71), we get for every $\beta' \in \mathcal{S}_1, \beta \in [\beta_*,\beta_\varepsilon]$, $b_{\beta,\beta'} \leq b_{\beta,\beta}(1+\gamma_{\varepsilon 2})$. These remarks, (70) and the decomposition of the normalizing factor in (68), yield

$$\begin{aligned} |E_f T_{\beta',\varepsilon} - T(f)| &\leq L h_l^{\beta-d/2-r}(\beta, L, \tilde{\varepsilon}(\beta)) b_{\beta,\beta}(1+\gamma_{\varepsilon 3}) = B(\beta, L, \tilde{\varepsilon}(\beta))(1+\gamma_{\varepsilon 3}) \\ &\leq \psi_\nu(1+\gamma_{\varepsilon 3}), \quad \text{for all } \beta' \in \mathcal{S}_1. \end{aligned} \tag{72}$$

From (44) and (71) we have

$$\psi_{\beta',L}/\psi_\nu \leq C\exp\{C(\beta-\beta')\log\varepsilon^{-1}\} \leq C\exp\{C\beta^+_\varepsilon\}$$

for all $\beta' \in \mathcal{S}_1$. This and (46) entail

$$\eta(\beta') \leq D_3^{-1}\psi_{\beta',L} \leq D_3^{-1}C\exp\{C\beta^+_\varepsilon\}\psi_\nu, \quad \text{for all } \beta' \in \mathcal{S}_1,$$

and

$$\begin{aligned} \frac{\sigma_{\beta'}}{\psi_\nu} &\leq D_3^{-1}C\exp\{C\beta^+_\varepsilon\}\frac{\sigma_{\beta'}}{\eta(\beta')} = \frac{C\exp\{C\beta^+_\varepsilon\}}{D_3 d_\varepsilon(\beta')} \\ &\leq \frac{C\exp\{C\beta^+_\varepsilon\}}{D_3 d_\varepsilon(\beta_\varepsilon)} \leq \frac{C_7\exp\{C\beta^+_\varepsilon\}(\beta_\varepsilon+\gamma)^{1/2}}{\log^{1/2}(1/\varepsilon)} \stackrel{def}{=} A_\varepsilon, \end{aligned} \tag{73}$$

for all $\beta' \in \mathcal{S}_1$, where we used also (48). Note also that, since $\beta^+ \le \bar{\beta}$,

$$\eta(\beta^+) \le \tilde{\varepsilon}(\beta) h_l^{-\gamma-d/2-r}(\beta, L, \tilde{\varepsilon}(\beta)) \|K_\beta\|_2 (1+\gamma_{\varepsilon 5}) = R(\beta, L, \tilde{\varepsilon}(\beta))(1+\gamma_{\varepsilon 5}), \quad (74)$$

(for details see Klemelä and Tsybakov, 2001).

Now we are ready for the main argument of the proof. Let first $\hat{\beta} = \beta' \in \mathcal{S}_1$. Then, in view of (72),

$$|T_\varepsilon^* - T(f)| = |T_{\beta',\varepsilon} - T(f)| \le |E_f T_{\beta',\varepsilon} - T(f)| + |Z_{\beta'}| \le \psi_\nu (1+\gamma_{\varepsilon 3}) + |Z_{\beta'}| \,. \quad (75)$$

Next, let $\hat{\beta} = \beta' \in \mathcal{S}_2$. Then, using the definition of $\hat{\beta}$, (72), (74), and (68), we get

$$\begin{aligned} |T_\varepsilon^* - T(f)| &\le |T_{\beta',\varepsilon} - T_{\beta^+,\varepsilon}| + |T_{\beta^+,\varepsilon} - T(f)| \\ &\le \eta(\beta^+) + |E_f T_{\beta^+,\varepsilon} - T(f)| + |Z_{\beta^+}| \\ &\le R(\beta, L, \tilde{\varepsilon}(\beta))(1+\gamma_{\varepsilon 5}) + B(\beta, L, \tilde{\varepsilon}(\beta))(1+\gamma_{\varepsilon 3}) + |Z_{\beta^+}| \\ &\le \psi_\nu (1+\gamma_{\varepsilon 6}) + |Z_{\beta^+}| \,. \end{aligned}$$

This and (75) entail

$$\begin{aligned} & E_f \left(\psi_\nu^{-p} \left| T_\varepsilon^* - T(f) \right|^p I \left(\hat{\beta} \in \mathcal{S}_1 \cup \mathcal{S}_2 \right) \right) \\ \le & \sum_{\beta' \in \mathcal{S}_1} E_f \left(\left(1 + \gamma_{\varepsilon 3} + \psi_\nu^{-1} \left| Z_{\beta'} \right| \right)^p I \left(\hat{\beta} = \beta' \right) \right) \\ + & \sum_{\beta' \in \mathcal{S}_2} E_f \left(\left(1 + \gamma_{\varepsilon 6} + \psi_\nu^{-1} \left| Z_{\beta^+} \right| \right)^p I \left(\hat{\beta} = \beta' \right) \right) . \end{aligned}$$

Applying Lemma 2 and (73) we get, for any $\beta' \in \mathcal{S}_1$,

$$\begin{aligned} & E_f \left(\left(1 + \gamma_{\varepsilon 3} + \psi_\nu^{-1} \left| Z_{\beta'} \right| \right)^p I \left(\hat{\beta} = \beta' \right) \right) \\ \le & \left(1 + \gamma_{\varepsilon 3} + \sqrt{\sigma_{\beta'} \psi_\nu^{-1}} \right)^p P_f \left(\hat{\beta} = \beta' \right) \\ + & C \left[1 + \psi_\nu^{-p} \left((\sigma_{\beta'} \psi_\nu)^{p/2} + \sigma_{\beta'}^p \right) \right] \exp \left\{ - \frac{\psi_\nu}{2\sigma_{\beta'}} \right\} \\ \le & \left(1 + \gamma_{\varepsilon 3} + A_\varepsilon^{1/2} \right)^p P_f \left(\hat{\beta} = \beta' \right) + C_8 \exp \left\{ - \frac{1}{2A_\varepsilon} \right\} , \end{aligned}$$

where A_ε is defined in (73) (for details see Klemelä and Tsybakov, 2001). Since $\beta^+ \in \mathcal{S}_1$, analogous bound holds for $E_f \left((1 + \gamma_{\varepsilon 6} + \psi_\nu^{-1} |Z_{\beta^+}|)^p I \left(\hat{\beta} = \beta' \right) \right)$. We conclude therefore that

$$R_{\varepsilon,\nu}^+ = E_f \left(\psi_\nu^{-p} \left| T_\varepsilon^* - T(f) \right|^p I \left(\hat{\beta} \in \mathcal{S}_1 \cup \mathcal{S}_2 \right) \right)$$

$$\leq \left(1+\gamma_{\varepsilon 7}+A_\varepsilon^{1/2}\right)^p P_f\left(\hat{\beta}\in\mathcal{S}_1\cup\mathcal{S}_2\right)+2C_8 m\exp\left\{-\frac{1}{2A_\varepsilon}\right\} \tag{76}$$

where $\gamma_{\varepsilon 7}=\max\{\gamma_{\varepsilon 3},\gamma_{\varepsilon 6}\}$. It remains to note that (50) follows from (76) and (65) applying the definition of A_ε in (73) and definition of β_ε given in (10). ■

4.2 Proof of the lower bound in Theorem 1.

Here we prove the bound (20). We denote, as above, $T(f)=f^{(\alpha_0)}(0)$. Let $L\in[L_*,L^*]$, $\nu_0=(\beta_0,L)$, and $\nu'=(\beta',L)$ where $\beta_*\leq\beta_0<\beta'\leq\beta_\varepsilon$. Consider the functions

$$f_0\equiv 0,\ \ f_1=(1-\delta)g_{\beta_0,L,\tilde{\varepsilon}(\beta_0)}$$

where $0<\delta<1/2$, $g_{\beta_0,L,\tilde{\varepsilon}(\beta_0)}$ is defined in (29), and $\tilde{\varepsilon}(\beta_0)=\varepsilon d_\varepsilon(\beta_0)$. Obviously, $f_0\in\mathcal{F}_{\beta',L}$. Furthermore, we have by direct calculation

$$\|R_\gamma f_1\|_2=(1-\delta)\tilde{\varepsilon}(\beta_0)=(1-\delta)\varepsilon d_\varepsilon(\beta_0) \tag{77}$$

and

$$\rho_{\beta_0}(f_1)\leq(1-\delta)L. \tag{78}$$

Indeed, by the renormalization argument,

$$\|R_\gamma g_{\beta,1,1}\|_2=ab^{-\gamma-d/2}\|R_\gamma\tilde{g}_\beta\|_2=b^{-\beta-\gamma}(C^*)^{-1}\|\tilde{K}_\beta\|_2=1$$

where $g_{\beta,1,1}$ is defined in (27), $\tilde{K}_\beta$ is defined in (7), and

$$\|\tilde{K}_\beta\|_2^2=(C^*)^2\,\frac{2(\beta-r)-d}{2(\gamma+r)+d}$$

which is proved by applying the property $B(u,v)=B(u-1,v+1)(u-1)/v$ of the Beta function. Then,

$$\|R_\gamma g_{\beta,L,\epsilon}\|_2=a_1 b_1^{-\gamma-d/2}\|R_\gamma g_{\beta,1,1}\|_2=\epsilon$$

where a_1,b_1 are defined (analogously as after Equation (29)), as $a_1=Lb_1^{d/2-\beta}$, $b_1=(L/\varepsilon)^{1/(\beta+\gamma)}$. Thus (77) holds. Also, by the renormalization argument,

$$\rho_\beta(g_{\beta,1,1})=ab^{\beta-d/2}\rho_\beta(\tilde{g}_\beta)=(C^*)^{-1}\rho_\beta(\tilde{g}_\beta)=1$$

and

$$\rho_\beta(g_{\beta,L,\epsilon}) = a_1 b_1^{\beta-d/2} \rho_\beta(g_{\beta,1,1}) = L.$$

Thus (78) holds. Then

$$\rho^2_{\beta_0}(f_1) + \|R_\gamma f_1\|_2^2 \le (1-\delta)^2 \left(L^2 + \tilde{\varepsilon}(\beta_0)^2\right)$$

and $f_1 \in \mathcal{F}_{\beta_0,L}$ for sufficiently small ε. From equation (30), $T(f_1) = (1-\delta)\psi_{\nu_0}$. Also, $T(f_0) = 0$. Thus, for any estimator T_ε,

$$|T_\varepsilon - T(f_i)| = \psi_{\nu_0}\, D\left((1-\delta)^{-1}\psi_{\nu_0}^{-1} T_\varepsilon, i\right), \quad i = 0, 1,$$

where $D(u,v) = (1-\delta)|u-v|$, $u, v \in \mathbf{R}$. Denoting $Q = \psi_{\nu_0}/\psi_{\nu'}$ and $E_i = E_{f_i}$, we get

$$\begin{aligned} \inf_{T_\varepsilon} \sup_{\nu \in B_\varepsilon} \mathcal{R}_{\varepsilon,\nu}(T_\varepsilon)\psi_\nu^{-p} &= \inf_{T_\varepsilon} \max\left\{\mathcal{R}_{\varepsilon,\nu'}(T_\varepsilon)\psi_{\nu'}^{-p}, \mathcal{R}_{\varepsilon,\nu_0}(T_\varepsilon)\psi_{\nu_0}^{-p}\right\} \\ &\ge \inf_{T_\varepsilon} \max\left\{E_0\left(\psi_{\nu'}^{-p}\,|T_\varepsilon - T(f_0)|^p\right), E_1\left(\psi_{\nu_0}^{-p}\,|T_\varepsilon - T(f_1)|^p\right)\right\} \\ &= \inf_{T_\varepsilon} \max\left\{Q^p E_0\Big(D^p(T_\varepsilon, 0)\Big), E_1\Big(D^p(T_\varepsilon, 1)\Big)\right\}. \end{aligned} \tag{79}$$

Let

$$\tau = \exp\left\{-\frac{1-\delta}{2}\, d_\varepsilon^2(\beta_0)\right\}.$$

Denoting $P_i = P_{f_i}$, we obtain

$$P_1\left(\frac{dP_0}{dP_1} \ge \tau\right) = P\left(\exp\left\{\varepsilon^{-1}\|f_1\|_2\xi - \varepsilon^{-2}\|f_1\|_2^2/2\right\} \ge \tau\right) = 1 - \Phi(l_\varepsilon) \tag{80}$$

where $\xi \sim \mathcal{N}(0,1)$, $\Phi(\cdot)$ is a standard normal c.d.f. and

$$l_\varepsilon = \frac{\varepsilon}{\|f_1\|_2}\left(\log\tau + \varepsilon^{-2}\|f_1\|_2^2/2\right) = -\frac{\delta}{2}\, d_\varepsilon(\beta_0) \longrightarrow -\infty,$$

as $\varepsilon \to 0$. Using (80) and applying Theorem 6 in Tsybakov (1998), we get

$$\inf_{T_\varepsilon} \max\left\{Q^p E_0 D^p(T_\varepsilon, 0), E_1 D^p(T_\varepsilon, 1)\right\} \ge \frac{(1-\Phi(l_\varepsilon))(1-2\delta)^p \tau (Q\delta)^p}{(1-2\delta)^p + \tau(Q\delta)^p}. \tag{81}$$

In view of (18) and (47), where $\kappa(\beta) = (\beta - d/2 - r)/(\beta + \gamma)$, we obtain

$$\begin{aligned} \frac{\psi_{\nu_0}}{\psi_{\nu'}} &= \frac{(\varepsilon^2 \log\varepsilon^{-1})^{\kappa(\beta_0)/2}}{(\varepsilon^2\log\varepsilon^{-1})^{\kappa(\beta')/2}}\ \frac{c_{\nu_0}}{c_{\nu'}} \\ &= \exp\left\{\frac{1}{2p}\left[d_\varepsilon^2(\beta_0) - d_\varepsilon^2(\beta')\right] - \frac{(r+\gamma+d/2)(\beta'-\beta_0)}{2(\beta_0+\gamma)(\beta'+\gamma)}\log\log\varepsilon^{-1}\right\} \frac{c_{\nu_0}}{c_{\nu'}}. \end{aligned}$$

Thus,

$$\begin{aligned}
\tau Q^p &= \exp\left\{-\frac{1-\delta}{2}\, d_\varepsilon^2(\beta_0)\right\}\left(\frac{\psi_{\nu_0}}{\psi_{\nu'}}\right)^p \\
&= \exp\left\{\frac{p(r+\gamma+d/2)[\delta(\beta'+\gamma+d/2)-(\beta_0+\gamma+d/2)]}{(\beta_0+\gamma+d/2)(\beta'+\gamma+d/2)}\log\varepsilon^{-1}\right\} \\
&\quad\times \exp\left\{-\frac{p(r+\gamma+d/2)(\beta'-\beta_0)}{2(\beta_0+\gamma)(\beta'+\gamma)}\,\log\log\varepsilon^{-1}\right\}\left(\frac{c_{\nu_0}}{c_{\nu'}}\right)^p.
\end{aligned}$$

Thus $\tau Q^p \to \infty$, as $\varepsilon \to 0$, if $\delta > (\beta_0+\gamma+d/2)/(\beta'+\gamma+d/2)$, and under this condition on δ we have

$$\inf_{T_\varepsilon} \max\left\{\mathcal{R}_{\varepsilon,\nu'}(T_\varepsilon)\psi_{\nu'}^{-p}, \mathcal{R}_{\varepsilon,\nu_0}(T_\varepsilon)\psi_{\nu_0}^{-p}\right\} \geq (1-2\delta)^p(1+o(1)), \tag{82}$$

as $\varepsilon \to 0$. Choosing now $\beta_0 = \beta_*$, $\beta' = \beta_\varepsilon$ and $\delta = (\beta_0+\gamma+d/2)/(A\beta_0+\gamma+d/2)$ for $A > 0$ large enough to have $\delta < 1/2$, passing to the limit as $\varepsilon \to 0$ and then as $A \to \infty$, we get the result. ∎

4.3 Proof of Theorem 2.

Let $\beta_0, \beta', L, \nu_0, \nu'$ be as defined in Section 4.2. Assumption (21) implies that there exists $0 < \delta < 1/3$ such that

$$\mathcal{R}_{\varepsilon,\nu'}(\hat{T}_\varepsilon)\psi_{\nu'}^{-p} \leq (1-3\delta)^p \tag{83}$$

for all ε small enough where, as above, $\nu' = (\beta', L)$. Next, note that (82) holds with $\mathcal{R}_{\varepsilon,\nu'}(T_\varepsilon)$ replaced by $\mathcal{R}_{\varepsilon,\nu_0'}(T_\varepsilon)$ where $\nu_0' = (\beta_0', L)$ for any real number β_0' belonging to B_ε (in fact, the inequality in (79) remains valid if we replace $\mathcal{R}_{\varepsilon,\nu'}(T_\varepsilon)$ by $\mathcal{R}_{\varepsilon,\nu_0'}(T_\varepsilon)$ with arbitrary β_0' since the function $f_0 \equiv 0$ belongs to all the classes $\mathcal{F}_{\beta_0',L}$). Hence, for any $\beta_0' \in (\beta', \beta_\varepsilon]$ we have

$$\inf_{T_\varepsilon} \max\left\{\mathcal{R}_{\varepsilon,\nu_0'}(T_\varepsilon)\psi_{\nu'}^{-p}, \mathcal{R}_{\varepsilon,\nu_0}(T_\varepsilon)\psi_{\nu_0}^{-p}\right\} \geq (1-2\delta)^p(1+o(1)), \tag{84}$$

as $\varepsilon \to 0$, provided that β' is such that

$$\delta > (\beta_0+\gamma+d/2)/(\beta'+\gamma+d/2). \tag{85}$$

Now choose $\beta' > \beta_0$ large enough to guarantee (85) for the δ appearing in (83). Clearly, $\beta' \in B_\varepsilon$ for all sufficently small ε. Then (83) and (84) imply

$$\max\left\{\mathcal{R}_{\varepsilon,\nu_0'}(T_\varepsilon)\psi_{\nu'}^{-p}, (1-3\delta)^p\right\} \geq (1-2\delta)^p(1+o(1)) \geq (1-5\delta/2)^p$$

for all sufficently small ε. Therefore,

$$\mathcal{R}_{\varepsilon,\nu_0'}(T_\varepsilon) \geq (1-5\delta/2)^p \psi_{\nu'}^p \tag{86}$$

for all sufficently small ε. Next, it follows from Ibragimov and Hasminskii (1984), Donoho and Low (1992) that $\varepsilon^{\kappa(\beta_0)}$ is minimax rate of convergence for the class $\mathcal{F}_{\beta_0,L}$, in particular, there exists a constant $C_9 > 0$ such that

$$\inf_{T_\varepsilon} \mathcal{R}_{\varepsilon,\nu_0}(T_\varepsilon) \geq C_9 \varepsilon^{p\kappa(\beta_0)}. \tag{87}$$

Using (86), (87) and (19) we get, for any $\beta_0' > \beta'$,

$$\begin{aligned}
\frac{\mathcal{R}_{\varepsilon,\beta_0',L}(\hat{T}_\varepsilon)}{\mathcal{R}_{\varepsilon,\beta_0',L}(T_\varepsilon^*)} \frac{\mathcal{R}_{\varepsilon,\beta_0,L}(\hat{T}_\varepsilon)}{\mathcal{R}_{\varepsilon,\beta_0,L}(T_\varepsilon^*)} &\geq \frac{(1-5\delta/2)^p \psi_{\nu'}^p}{\psi_{\nu_0'}^p} \left(\frac{C_9 \varepsilon^{p\kappa(\beta_0)}}{\psi_{\nu_0}^p} \right) (1+o(1)) \\
&\geq C_{10} (\varepsilon^2 \log \varepsilon^{-1})^{p(\kappa(\beta')-\kappa(\beta_0'))/2} (\log \varepsilon^{-1})^{p\kappa(\beta_0)/2} \to \infty,
\end{aligned}$$

as $\varepsilon \to 0$. ■

Acknowledgements.

We wish to thank a referee from the helpful comments.

References

[1] Butucea, C. (2001a). Exact adaptive pointwise estimation on Sobolev classes of densities. *ESAIM: Probability and Statistics* **5** 1-31.

[2] Butucea, C. (2001b). Numerical results concerning a sharp adaptive density estimator. *Computational Statistics* **16** 271-298.

[3] Cavalier, L., Golubev, G.K., Lepski, O.V. and Tsybakov, A.B. (2003). Block thresholding and sharp adaptive estimation in severely ill-posed inverse problems. *Theory of Probability and its Applications*, in press.

[4] Cavalier, L., Golubev, G.K., Picard, D. and Tsybakov, A.B. (2002). Oracle inequalities for inverse problems. *Annals of Statistics* **30** 843-874.

[5] Cavalier, L. and Tsybakov, A.B. (2002). Sharp adaptation for inverse problems with random noise. *Probability Theory and Related Fields* **123** 323-354.

[6] Chow, P.L. and Khasminskii, R.Z. (1997). Statistical approach to dynamical inverse problems. *Comm. Math. Phys.* **189** 521-531.

[7] Chow, P.L., Ibragimov, I.A. and Khasminskii, R.Z. (1999). Statistical approach to some inverse problems for linear partial differential equations. *Probability Theory and Related Fields* **113**, 421-441.

[8] Donoho, D. (1994a). Asymptotic minimax risk for sup-norm loss: solution via optimal recovery. *Probab. Theory and Related Fields* **99** 145-170.

[9] Donoho, D. (1994b). Statistical estimation and optimal recovery. *Ann. Statist.* **22** 238-270.

[10] Donoho, D. and Low, M. (1992). Renormalization exponents and optimal pointwise rates of convergence. *Ann. Statist.* **20** 944-970.

[11] Efromovich, S. Y. (1997a). Density estimation for the case of supersmooth measurement error. *J. Amer. Statist. Assoc.* **92** 526-535.

[12] Efromovich S. (1997b). Robust and efficient recovery of a signal passed through a filter and then contaminated by non-Gaussian noise. *IEEE Trans. Inform. Theory* **43** 1184-1191.

[13] Goldenshluger A. (1998). On pointwise adaptive nonparametric deconvolution. *Bernoulli*, **5**, 907-926.

[14] Goldenshluger, A. and Pereverzev, S. V. (2000). Adaptive estimation of linear functionals in Hilbert scales from indirect white noise observations. *Probab. Theory Relat. Fields* **118** 169-186.

[15] Ibragimov, I. A. and Hasminskii, R. Z. (1984). On nonparametric estimation of the value of a linear functional in Gaussian white noise. *Theory Probab. Appl.* **29** 18-32.

[16] Klemelä, J. (2003). Optimal recovery and statistical estimation in L_p Sobolev classes. Submitted.

[17] Klemelä, J. and Tsybakov, A. B. (2001). Sharp adaptive estimation of linear functionals. *Annals of Statistics* **29** 1567-1600.

[18] Korostelev, A. P. and Tsybakov, A. B. (1991). Optimal rates of convergence of estimators in a probabilistic setup of tomography problem. *Problems of Information Transmission* **27** 73-81.

[19] Korostelev, A. P. and Tsybakov, A. B. (1993). *Minimax Theory of Image Reconstruction.* Lecture Notes in Statistics, v.82, Springer, New York.

[20] Lepski, O. V. (1990). One problem of adaptive estimation in Gaussian white noise. *Theory Probab. Appl.* **35** 454-466.

[21] Lepski, O. V. (1991). Asymptotically minimax adaptive estimation I: upper bounds. Optimal adaptive estimates. *Theory Probab. Appl.* **36** 682-697.

[22] Lepski, O. V. (1992). Asymptotically minimax adaptive estimation II: statistical models without optimal adaptation. Adaptive estimators. *Theory Probab. Appl.* **37** 433-468.

[23] Lepski, O. V. and Spokoiny, V. G. (1997). Optimal pointwise adaptive methods in nonparametric estimation. *Ann. Statist.* **25** 2512-2546.

[24] Lepski, O. V. and Tsybakov, A. B. (2000). Asymptotically exact nonparametric hypothesis testing in sup-norm and at a fixed point. *Probab. Theory and Related Fields*, **117**, 17-48.

[25] Stein, E. M. (1970). *Singular Integrals and Differentiability Properties of Functions.* Princeton University Press, Princeton, New Jersey.

[26] Taikov, L. V. (1968). Kolmogorov type inequalities and the best formulas of numerical differentiation. *Math. Notes* **4** 233-238.

[27] Tsybakov, A. B. (1998). Pointwise and sup-norm sharp adaptive estimation of functions on the Sobolev classes. *Ann. Statist.* **26** 2420-2469.

Jussi Klemelä
Institut für Angewandte Mathematik
Universität Heidelberg
Im Neuenheimer Feld 294
69120 Heidelberg, Germany
email: klemela@statlab.uni-heidelberg.de

Alexandre Tsybakov
Laboratoire de Probabilités et Modèles Aléatoires
Université Paris 6
4, pl.Jussieu, BP 188, 75252 Paris, France
email: tsybakov@ccr.jussieu.fr