On pointwise adaptive nonparametric deconvolution

ALEXANDER GOLDENSHLUGER

Department of Statistics, University of Haifa, Mount Carmel, 31905 Haifa, Israel.
E-mail: goldensh@rstat.haifa.ac.il

We consider estimating an unknown function f from indirect white noise observations with particular emphasis on the problem of nonparametric deconvolution. Nonparametric estimators that can adapt to unknown smoothness of f are developed. The adaptive estimators are specified under two sets of assumptions on the kernel of the convolution transform. In particular, kernels having Fourier transform with polynomially and exponentially decaying tails are considered. It is shown that the proposed estimates possess, in a sense, the best possible abilities for pointwise adaptation.

Keywords: adaptive estimation; deconvolution; rates of convergence

1. Introduction

This paper investigates the problem of pointwise adaptive nonparametric estimation from indirect white noise observations. Let $f \in L_2(\mathbb{R})$ be an unknown function, and let $K : \mathscr{D}(K) \mapsto \mathscr{R}(K)$ be a linear transformation, $\mathscr{D}(K) \subseteq L_2(\mathbb{R})$, $\mathscr{R}(K) \subseteq L_2(\mathbb{R})$. We can observe a process $y(t)$ characterized by

$$\mathrm{d}y(t) = (Kf)(t)\,\mathrm{d}t + \varepsilon\,\mathrm{d}W(t), \quad t \in \mathbb{R}, \tag{1}$$

where $\varepsilon > 0$ is a known parameter, and $W(t)$ is the standard two-sided Wiener process on $\mathbb{R}$. This scheme is equivalent to observing Kf with added white noise. Our goal is to estimate f at a single given point t_0. We study here the white noise model (1) because it is mathematically more tractable and gives some suggestions as to what could hold in related nonparametric estimation problems.

Nonparametric estimation from indirect noisy observations arises in a wide variety of applications. The usual approach to such problems is based on certain regularized operator inversions; see, for example, O'Sullivan (1986), Wahba (1990), Nychka and Cox (1989), Mair and Ruymgaart (1996) and references therein. These methods assume that the function to be estimated belongs to a known pre-specified class of smooth functions. In practice, however, specifying the functional class presents severe difficulties. Besides this, functions may contain high-frequency oscillations, or discontinuities. Therefore it is desired to develop nonparametric estimators which can easily be adapted to varying levels of smoothness. In what follows such estimators are referred to as *adaptive*. Recently much attention has been concentrated on developing adaptive nonparametric estimators both for direct and indirect observations. In Donoho (1995) the *wavelet–vaguelette decomposition*

(WVD) was proposed for estimating functions with inhomogeneous smoothness from indirect observations. The basic idea of this proposal is to expand the unknown function f as a wavelet series, to find a corresponding vaguelette series for Kf, and then to estimate the WVD coefficients using an appropriate threshold technique. An alternative *vaguelette–wavelet decomposition* (VWD) for estimating functions with inhomogeneous smoothness was developed by Abramovich and Silverman (1998). Here, contrary to WVD, empirical wavelet coefficients of Kf are estimated by the threshold technique; the VWD estimator of f is obtained by mapping that wavelet expansion into the original space. It was shown that both WVD and VWD estimators adapt with 'minimal cost' to the unknown smoothness of f over a wide range of Besov scales. It should be noticed that these wavelet-based estimators apply only to linear transformations which are homogeneous with respect to dilatation.

In this paper we develop another adaptive estimator for nonparametric estimation from indirect observations. Our construction exploits the *linear functional strategy* for the solution of ill-posed inverse problems (Anderssen 1980; Goldberg 1979) along with the *general adaptation scheme* (Lepskii 1991; 1992) proposed for the nonparametric regression problem. In general, this approach can be applied to a wide variety of linear transformations K, not necessarily homogeneous with respect to dilatation. In what follows we will study the case of the convolution transform only, that is,

$$g(t) = (Kf)(t) = \int_{-\infty}^{\infty} K(t-s)f(s)\,\mathrm{d}s.$$

Some conditions on the kernel K should be imposed to ensure that f is identifiable. Throughout the paper we assume that the Fourier transform

$$\hat{K}(\lambda) = \int_{-\infty}^{\infty} K(t)\,\mathrm{e}^{-\mathrm{i}t\lambda}\,\mathrm{d}t$$

of K does not vanish. Under this condition estimation accuracy depends heavily on the rate at which $\hat{K}(\lambda)$ tends to zero as $|\lambda| \to \infty$. We consider two important cases where $\hat{K}(\cdot)$ has polynomially and exponentially decaying tails. In the former case we propose an adaptive estimator and prove that it is near-optimal for estimating at a single point up to a factor logarithmic in ε^{-1}. In other words, if the function f is smooth at the point t_0, then the estimation accuracy coincides, up to the logarithmic factor, with the best accuracy we could achieve in the case of known smoothness. If kernel K is Green's function of a linear differential operator, then the deconvolution problem is equivalent to estimating the linear differential operator of Kf from the direct white noise observations. For such a problem in the case of unknown smoothness the aforementioned logarithmic factor cannot be reduced for estimating at single point (see Lepskii 1992). Thus, in this specific case our estimator is the best possible for pointwise adaptation. In the case of exponentially decaying tails of $\hat{K}(\cdot)$ we establish a lower bound on the estimation accuracy and show that pointwise adaptation to unknown smoothness can be achieved by 'direct tuning' of the smoothing parameter. Here the optimal rates of convergence can be obtained without a priori information on smoothness of the underlying function f. It should be pointed out that our results are concerned with estimation of f at a single point. They can serve as a basis for proving that the same

estimators are also adaptive over a wide range of the global accuracy measures. The techniques of Lepskii *et al.* (1997) and Goldenshluger and Nemirovski (1997) seem to be appropriate for this purpose.

The rest of the paper is organized as follows. In Section 2 we describe our main idea for construction of adaptive estimators from indirect observations. In Section 3 we present our main results on pointwise adaptive deconvolution for polynomially and exponentially decaying tails of $\hat{K}(\cdot)$. Proofs are given in Section 4.

2. Construction of estimator

Let $\langle \cdot, \cdot \rangle$ denote the standard inner product on $L_2(\mathbb{R})$. To estimate $f(t_0)$ we will construct estimates for certain linear functionals $\langle f, \varphi_\delta \rangle$. The functions $\varphi_\delta(\cdot)$ should meet the following two requirements. First, the linear functionals $\langle f, \varphi_\delta \rangle$ should be suitable approximations to $f(t_0)$; for this purpose, it is quite natural to choose $\varphi_\delta(\cdot) = \delta^{-1}\varphi(\delta^{-1}(\cdot - t_0))$ with some fixed function $\varphi(\cdot)$ and $\delta > 0$ converging to zero. Second, according to the *linear functional strategy* we have to choose $\varphi_\delta(\cdot)$ in such a way that the linear functionals $\langle f, \varphi_\delta \rangle$ could be stably recovered from Kf. To this end, the functions $\varphi_\delta \in L_2(\mathbb{R})$ should lie in the range of the adjoint to K, operator K^*. Indeed, if $\mathscr{D}(K)$ is dense in $L_2(\mathbb{R})$ and $\varphi_\delta \in \mathscr{R}(K^*)$, then there exists a function $h_\delta \in L_2(\mathbb{R})$ such that $\varphi_\delta = K^* h_\delta$, and

$$\langle f, \varphi_\delta \rangle = \langle f, K^* h_\delta \rangle = \langle Kf, h_\delta \rangle = \langle g, h_\delta \rangle, \quad \forall f \in \mathscr{D}(K).$$

Thus, if functions $\varphi_\delta(\cdot)$ satisfy the above requirements, then $\hat{f}_\delta(t_0) = \int_{-\infty}^{\infty} h_\delta(t)\, \mathrm{d}y(t)$ can be chosen as an estimate of $f(t_0)$. To define the estimate completely, we should specify the *bandwidth* δ. As usual in nonparametric estimation, δ controls the trade-off between the bias and variance of the estimate. Typically the optimal choice of the bandwidth requires a priori information on the smoothness of the function to be estimated, and for this reason is not practical. We apply the *general adaptation scheme* to treat this problem.

Observe that for the convolution transform one has

$$\mathscr{R}(K^*) = \left\{ f : \int_{-\infty}^{\infty} |\hat{f}(\lambda)|^2 |\hat{K}(\lambda)|^{-2}\, \mathrm{d}\lambda < \infty \right\}.$$

As we will see, in this case the *linear functional strategy* results in the kernel deconvolution estimator extensively studied by Stefanski and Carrol (1990), Carrol and Hall (1988), Fan (1991a; 1991b) and Masry (1991) for the problem of density deconvolution. It is worth noting that in the construction of the WVD estimator, Donoho (1995) also applies the *linear functional strategy*, but the functions φ_δ are chosen to be an orthonormal wavelet basis, and the resulting functionals $\langle f, \varphi_\delta \rangle$ are the corresponding wavelet coefficients.

3. Pointwise adaptive deconvolution

Consider the family of functions $\varphi_\delta(\cdot) = \delta^{-1}\varphi(\delta^{-1}(\cdot - t_0))$ indexed by a positive real number δ and associated with some function $\varphi(\cdot)$. Given $\delta > 0$, we define the estimate for $f(t_0)$ as

$$\hat{f}_\delta(t_0) = \int_{-\infty}^{\infty} h_\delta(t)\,\mathrm{d}y(t), \tag{2}$$

where the function $h_\delta(\cdot)$ is defined via its Fourier transform as

$$\hat{h}_\delta(\lambda) = \mathrm{e}^{-i\lambda t_0}\,\frac{\hat{\varphi}(\delta\lambda)}{\hat{K}(-\lambda)}. \tag{3}$$

The following assumption will be used throughout the paper:

Assumption K1. $|\hat{K}(\lambda)| \neq 0$, *for all* λ.

We now consider separately two important cases.

3.1. Kernel with polynomially decaying Fourier transform

Here our assumptions on the kernel K are as follows.

Assumption K2. *There exist positive real numbers A, α and c_0 such that*

$$|\hat{K}(\lambda)| \geqslant c_0|\lambda|^{-\alpha}, \quad \forall|\lambda| > A, \tag{4}$$

$$\min_{|\lambda|\leqslant A} |\hat{K}(\lambda)| = m > 0. \tag{5}$$

Green's functions of linear differential operators are important examples of kernels K satisfying Assumption K2. For example, let

$$v(t) = \begin{cases} \exp(t), & -\infty < t < 0 \\ \frac{1}{2} & t = 0 \\ 0, & 0 < t < \infty, \end{cases}$$

and for non-vanishing real constants $a_1, \ldots, a_k$, let $v_j(t) = |a_j|v(a_jt)$, $j = 1, \ldots, k$. Define $K = v_1 \star v_2 \star \cdots \star v_k$, where $\star$ denotes convolution on $\mathbb{R}$. For such a kernel

$$\hat{K}(\lambda) = \left[\prod_{j=1}^{k}\left(1 - \frac{i\lambda}{a_j}\right)\right]^{-1},$$

and (4) holds with $\alpha = k$. In this case f can be recovered from Kf by applying the linear differential operator

$$f(t) = \left[\prod_{j=1}^{k} \left(1 - a_j^{-1} \frac{d}{dt} \right) \right] (Kf)(t), \quad t \in \mathbb{R}$$

see, for example, Hirschmann and Widder (1995, Chapter II). Here the estimation problem is equivalent to estimating the linear differential operator of order α from the direct white noise observations.

Let us fix a positive integer number s and choose a function φ satisfying the following assumptions:

Assumption A1. *φ is a symmetric, supported on $[-1, 1]$, square-integrable function such that*

$$\int_{-\infty}^{\infty} |\hat{\varphi}(\lambda)|^2 |\lambda|^{2\alpha} \, d\lambda = B^2 < \infty, \tag{6a}$$

$$\int_{-\infty}^{\infty} |\hat{\varphi}(\lambda)| \, |\lambda|^{\alpha} \, d\lambda < \infty. \tag{6b}$$

Assumption A2. *φ has s vanishing moments, that is,*

$$\int_{-1}^{1} \varphi(t) t^j \, dt = \begin{cases} 1, & j = 0, \\ 0, & j = 1, \ldots, s. \end{cases} \tag{7}$$

If φ has square-integrable derivative of the order $\lceil \alpha \rceil$ (where $\lceil \alpha \rceil$ denotes the smallest integer number greater than α), then inequality (6a) holds. The moment conditions (7) are quite standard for nonparametric kernel estimates. Assumptions similar to A1 and A2 have also been considered in Fan (1991a; 1991b), where the density deconvolution problem has been studied. A function φ satisfying Assumptions A1 and A2 can be easily constructed using Jacobi orthonormal polynomials on the interval $[-1, 1]$.

The following lemma asserts that under Assumptions K1, K2 and A1 the estimate $\hat{f}_\delta(t_0)$ given by (2)–(3) is well defined for any $\delta > 0$.

Lemma 1. *Let Assumptions K1, K2 and A1 hold; then for any $\delta > 0$ one has $\varphi_\delta \in \mathscr{R}(K^*)$, and*

$$\|h_\delta\|_2^2 \leq \frac{2A\|\varphi\|_1^2}{m^2} + \frac{B^2}{c_0^2 \delta^{2\alpha+1}}. \tag{8}$$

Proof. It suffices to establish inequality (8). We have

$$\|h_\delta\|_2^2 = \int_{-\infty}^{\infty} \left| \frac{\hat{\varphi}(\delta\lambda)}{\hat{K}(\lambda)} \right|^2 d\lambda = \int_{|\lambda| \leq A} \left| \frac{\hat{\varphi}(\delta\lambda)}{\hat{K}(\lambda)} \right|^2 d\lambda + \int_{|\lambda| > A} \left| \frac{\hat{\varphi}(\delta\lambda)}{\hat{K}(\lambda)} \right|^2 d\lambda. \tag{9}$$

It follows immediately from (5) that

$$\int_{|\lambda| \leq A} \left| \frac{\hat{\varphi}(\delta\lambda)}{\hat{K}(\lambda)} \right|^2 d\lambda \leq m^{-2} \int_{|\lambda| \leq A} |\hat{\varphi}(\delta\lambda)|^2 \, d\lambda \leq \frac{2A\|\varphi\|_1^2}{m^2}.$$

Further, taking into account (4) and (6a), we obtain

$$\int_{|\lambda|>A} \left| \frac{\hat{\varphi}(\delta\lambda)}{\hat{K}(\lambda)} \right|^2 \mathrm{d}\lambda \leq c_0^{-2} \delta^{-2\alpha-1} \int_{|\lambda|>A\delta} |\hat{\varphi}(\lambda)|^2 |\lambda|^{2\alpha} \, \mathrm{d}\lambda \leq \frac{B^2}{c_0^2 \delta^{2\alpha+1}}.$$

Combining these inequalities with (9) gives (8). Note also that (6b) implies $\int |\hat{h}_\delta(\lambda)| \, \mathrm{d}\lambda < \infty$, so that h_δ can be obtained from (3) by standard Fourier inversion.   □

Now we are ready to define the adaptive estimate $\hat{f}_*(t_0)$ we are actually interested in. The main idea underlying the construction of our adaptive estimator is the following. Consider a family of estimates $\{\hat{f}_\delta(t_0)\}_{\delta\in\Delta}$, where Δ is a finite ordered set of possible bandwidths. We choose the maximal bandwidth $\delta_* \in \Delta$ such that the estimate associated with this choice does not differ 'significantly' from the estimates with smaller bandwidths from Δ. The extent to which $\hat{f}_{\delta_*}(t_0)$ should be close to the estimates with smaller bandwidths can be evaluated without any a priori information on the function to be estimated.

Let

$$r(\delta) = \left(\frac{2A\|\varphi\|_1^2}{m^2} + \frac{B^2}{c_0^2 \delta^{2\alpha+1}} \right)^{1/2}, \tag{10}$$

and define

$$\overline{\delta} = \left(\frac{B^2 m^2}{2A c_0^2 \|\varphi\|_1^2} \right)^{1/(2\alpha+1)}, \qquad \underline{\delta} = \varepsilon^{2/(2\alpha+1)} \overline{\delta}. \tag{11}$$

Fix a real number $a > 1$, and consider the family of estimates $\hat{f}_\delta(t_0)$ indexed by $\delta > 0$ from the set

$$\Delta = \{\delta \in [\underline{\delta}, \overline{\delta}] : \delta = a^j \underline{\delta}, \, j = 0, 1, 2, \ldots\}.$$

With every estimate $\hat{f}_\delta(t_0)$ we associate the interval

$$I_\delta = [\hat{f}_\delta(t_0) - 2\kappa\varepsilon r(\delta), \, \hat{f}_\delta(t_0) + 2\kappa\varepsilon r(\delta)],$$

where $\kappa \geq 1$ is a fixed real number. Let δ_* denote *the maximal* $\delta \in \Delta$ such that

$$\bigcap_{\delta\in\Delta, \delta\leq\delta_*} I_\delta \neq \varnothing. \tag{12}$$

Then we define our estimate as $\hat{f}_*(t_0) = \hat{f}_{\delta_*}(t_0)$. Note that δ_* is well defined, in particular the minimal $\delta_* = \underline{\delta}$ always satisfies (12). Observe that our construction depends on observed trajectory $y(t)$, on information on the kernel K, on ε, and on three 'design' parameters s, a and κ. In the following κ will be chosen as function of s, a and ε, so that actually our estimator depends on two 'design parameters' s and a only.

The accuracy of the estimate $\hat{f}_*(t_0)$ will be analysed under the following assumption on the function f to be estimated.

Assumption F. *There exist a positive integer number* $l \geqslant 1$, *positive real numbers L and* $p \in [1, \infty]$, *and an interval* $D = [t_0 - d, t_0 + d]$ *such that* $pl > 1$, *and*

$$\left(\int_D |f^{(l)}(t)|^p \, dt \right)^{1/p} \leqslant L|D|^{1/p}. \tag{13}$$

In the case of $p = \infty$ *the left-hand side of* (13) *is the usual* $\|\cdot\|_\infty$*-norm on D. For functions satisfying this assumption we write* $f \in W_p^l(L, D)$.

We would like to stress here that parameters L, D, l, p of the class $W_p^l(L, D)$ are not involved in the construction of our estimate, and in fact the estimate does not use any a priori quantitative information on f.

Theorem 1. *Let Assumptions K1, K2, A1 and F hold. Assume that A2 holds with* $s + 1 \geqslant l$, *and let* $q = l - 1/p$ *and* $\beta = \alpha + \frac{1}{2}$. *Let* ε *be small enough so that*

$$\varepsilon \leqslant \min\left\{ \frac{c_0\|\varphi\|_1}{B}(L|D|^{1/p})^{-\beta/q}\left(\frac{\sqrt{2A}}{m}\right)^{1+\beta/q}, \frac{d^\beta\sqrt{2A}c_0\|\varphi\|_1}{mB}, a^{-3} \right\}. \tag{14}$$

Then there exists a constant $C = C(a, s)$ *depending on a and s only such that for our estimate* $\hat{f}_*(t_0)$ *associated with the choice* $\kappa = C(a, s)\sqrt{\ln(\varepsilon^{-1})}$ *one has*

$$[\mathbb{E}|\hat{f}_*(t_0) - f(t_0)|^2]^{1/2} \leqslant O(1)\left[(L|D|^{1/p}\|\varphi\|_1)^{\beta/(\beta+q)}\left(\frac{B}{c_0}\varepsilon\sqrt{\ln\frac{1}{\varepsilon}}\right)^{q/(\beta+q)} + \frac{\sqrt{A}\|\varphi\|_1}{m}\varepsilon\sqrt{\ln\frac{1}{\varepsilon}} \right], \tag{15}$$

where $O(1)$ *is a constant depending on a and s only.*

A proof is given in Section 4.

If we knew in advance parameters p, l, L, D of the class $W_p^l(L, D)$, we could achieve the minimax rate of convergence $O(\varepsilon^{q/(\beta+q)})$. Therefore the quality of our adaptive estimate coincides, up to a factor logarithmic in ε^{-1}, with the lower bound on the minimax risk corresponding to the case of known smoothness. It is well known (Lepskii 1992) that in the case of unknown smoothness the indicated factor cannot be reduced for estimating derivatives at a single point from the direct white noise observations. In Lepskii (1992) the corresponding lower bound was established on the minimax risk uniformly over a functional family containing at least two classes $W_\infty^{l_1}(L, [0, 1])$ and $W_\infty^{l_2}(L, [0, 1])$, with $l_1 \neq l_2$. As was noticed before, if K is Green's function of a linear differential operator of order α, then the deconvolution problem is equivalent to estimating the linear differential operator of Kf from direct observations. Observe also that the upper bound (15) holds for *all* f from *all* classes $W_p^l(L, D)$ with parameters p, l, L and D satisfying $p \in [1, \infty]$, $pl > 1$, and conditions (14). Thus, our estimate adapts optimally to the unknown smoothness, whenever K is Green's function of a linear differential operator.

3.2. Kernel with exponentially decaying Fourier transform

Now consider the case of exponentially decaying tails of the Fourier transform $\hat{K}(\lambda)$. We assume that K1 holds, and that instead of K2 the following condition is fulfilled:

Assumption K3. *There exist positive real numbers A, α, γ and c_0 such that*

$$|\hat{K}(\lambda)| \geqslant c_0 \exp(-\gamma|\lambda|^\alpha), \quad \forall|\lambda| > A, \quad and \quad \min_{|\lambda|\leqslant A} |\hat{K}(\lambda)| = m > 0.$$

Gaussian and Cauchy densities are the simplest examples of kernels K satisfying Assumption K3.

Fix a positive integer number s, and choose a function φ satisfying the following assumptions:

Assumption A3. *$\hat{\varphi}$ is a symmetric function, supported on $[-1, 1]$, and*

$$\int_{-\infty}^{\infty} |\varphi(t)|\, \mathrm{d}t = \|\varphi\|_1 < \infty, \qquad \int_{-\infty}^{\infty} |\varphi(t)|\, |t|^{s+1}\, \mathrm{d}t \equiv B < \infty.$$

Assumption A4.

$$\hat{\varphi}(0) = 1, \hat{\varphi}'(0) = \ldots = \hat{\varphi}^{(s)}(0) = 0.$$

Note that Assumption A4 is equivalent to the moment conditions (7).

For $\delta > 0$, let $\hat{f}_\delta(t_0)$ denote the estimate given by (2)–(3). The following lemma shows that under Assumptions K1, K3 and A3 the estimate $\hat{f}_\delta(t_0)$ is well defined.

Lemma 2. *Let Assumptions K1, K3 and A3 hold; then for any $\delta > 0$ one has $\varphi_\delta \in \mathscr{R}(K^*)$, and*

$$\|h_\delta\|_2^2 \leqslant \begin{cases} \dfrac{2A\|\varphi\|_1^2}{m^2} + \dfrac{2\|\varphi\|_1^2}{c_0^2\delta} \exp(2\gamma\delta^{-\alpha}), & \text{if } A\delta < 1, \\[2ex] \dfrac{2A\|\varphi\|_1^2}{m^2}, & \text{otherwise.} \end{cases}$$

The proof of the lemma resembles that of Lemma 1 and is therefore omitted. We note only that in this case the integral $\int_{|\lambda|>A} |\hat{\varphi}(\delta\lambda)|^2 |\hat{K}(\lambda)|^{-2}\, \mathrm{d}\lambda$ vanishes if $A\delta \geqslant 1$.

Let us set

$$\delta_* = \left[\gamma^{-1}\ln\frac{1}{\varepsilon} - \frac{(s+2)}{\alpha\gamma}\ln\left(\gamma^{-1}\ln\frac{1}{\varepsilon}\right)\right]^{-1/\alpha}, \tag{16}$$

and let $\hat{f}_*(t_0)$ stand for the estimate $\hat{f}_\delta(t_0)$ associated with the choice $\delta = \delta_*$. Note that this choice of δ_* does not require any a priori information on the smoothness of f to be estimated; only the 'order' s is a 'design' parameter for the estimate $\hat{f}_\delta(t_0)$.

Theorem 2. *Let Assumptions K1, K3, A3, A4 and F hold; let, in addition,*

$$|f(t)| \leqslant M, \quad \forall t \in \mathbb{R}. \tag{17}$$

Define $\beta = 1 + l/(s + 1 - l + 1/p)$ *and assume that* ε *is small enough so that*

$$\ln \frac{1}{\varepsilon} \geqslant 2\gamma \max\left\{ 1, \frac{2(s+2)^2}{\alpha^2 \gamma^2}, d^{-\alpha}, d^{-\alpha\beta} \right\}. \tag{18}$$

Then for the estimate $\hat{f}_*(t_0)$ *associated with the bandwidth* δ_* *one has*

$$[\mathrm{E}|\hat{f}_*(t_0) - f(t_0)|^2]^{1/2} \leqslant O(1)\left[\tilde{B}(\tilde{L}|D|^{1/p} + c_0^{-1})\left(\frac{1}{2\gamma} \ln \frac{1}{\varepsilon}\right)^{-(l-1/p)/\alpha} + \frac{\|\varphi\|_1 \sqrt{A}}{m}\varepsilon \right], \tag{19}$$

where $O(1)$ *is an absolute constant,* $\tilde{L} = \max\{L, M\}$ *and* $\tilde{B} = B + \|\varphi\|_1$.

A proof of the theorem is given in Section 4.

In the case of known smoothness for the density deconvolution problem it was proved (Carrol and Hall 1988; Fan 1991b) that the minimax pointwise rate of convergence is logarithmic in the number of observations. In particular, for $p = \infty$ it equals $O([\ln n]^{-l/\alpha})$, where n is the number of observations. These results are compatible with the upper bound (19) with the usual calibration $\varepsilon = n^{-1/2}$. In a recent paper Efromovich (1997) shows that for the white noise model with $p = 2$ the pointwise minimax rate of convergence is equal to $O([(2\gamma)^{-1} \ln(1/\varepsilon)]^{-(l-1/2)/\alpha})$. Theorem 2 shows that in this case adaptation to unknown smoothness can be achieved by 'direct tuning' of the bandwidth, without any additional adaptation rule. In Efromovich (1997) a similar result was obtained for an estimator based on orthogonal series. In fact, the rate of convergence given by Theorem 2 is optimal for every $p \in [1, \infty]$, as Theorem 3 asserts.

Without loss of generality, we assume that $t_0 = 0$, and replace K3 with the following:

Assumption K3'. *There exist real numbers* c_0, c_1, γ, α *such that*

$$c_0 \exp(-\gamma|\lambda|^\alpha) \leqslant |\hat{K}(\lambda)| \leqslant c_1 \exp(-\gamma|\lambda|^\alpha), \quad \forall\lambda.$$

The function f is assumed to satisfy

$$\left(\int_{-\infty}^{\infty} |f^{(l)}(t)|^p \, dt \right)^{1/p} \leqslant L, \qquad pl > 1, \qquad p \in [1, \infty].$$

For such functions we write $f \in W_p^l(L)$.

Theorem 3. *Let Assumptions K1 and K3' hold, and* $p \in [1, \infty]$. *Then there exist constants* $C_1(p, l, L, K)$ *and* $C_2(p, l, L, K)$ *depending on* $p, l, L,$ *and on the kernel* K *only such that for* ε *satisfying*

$$\ln \frac{1}{\varepsilon} \geqslant C_1(p, l, L, K) \tag{20}$$

one has

$$\inf_{\hat{f}} \sup_{f \in W_p^l(L)} \mathrm{E}_f \left[|\hat{f}(0) - f(0)|^2 \right]^{1/2} \geq C_2(p,\, l,\, L,\, K) \left[\gamma^{-1} \ln \frac{1}{\varepsilon} \right]^{-(l-1/p)/\alpha},$$

where the infimum is taken over all possible estimators.

A proof of the theorem is given in Section 4.

Theorem 3, along with Theorem 2, implies that for every $p \in [1,\, \infty]$ our estimator is the best possible for pointwise adaptation. Thus, in the case of exponentially decaying tails of $\hat{K}(\lambda)$ adaptation to unknown smoothness can be achieved at the same 'cost' in the terms of rates of convergence.

4. Proofs

4.1. Proof of Theorem 1

The main idea of the proof is the following. Let

$$\eta_h(\delta,\, t_0) = \int_{-\infty}^{\infty} h_\delta(t)\, \mathrm{d}W(t),$$

where h_δ is defined in (3). Lemma 1 implies that $h_\delta \in L_2(\mathbb{R})$, and therefore $\eta_h(\delta,\, t_0)$ is the Gaussian random variable $\mathcal{N}(0,\, \|h_\delta\|_2^2)$. By the definition of $\hat{f}_\delta(t_0)$, for any $\delta > 0$ the estimation accuracy can be bounded from above as follows:

$$|\hat{f}_\delta(t_0) - f(t_0)| \leq |f(t_0) - \langle f,\, \varphi_\delta \rangle| + \varepsilon |\eta_h(\delta,\, t_0)|, \tag{21}$$

where the first term on the right-hand side of (21) is the deterministic bias and the second is the stochastic error. First, we obtain the upper bound given in (15) on the set where the stochastic error behaves 'typically'. Then we prove that the aforementioned set is of 'large' probability, so that the influence of the complementary 'bad' event is negligible.

We start with some notation. Given $\kappa \geq 1$, define the event

$$\Omega_\kappa(t_0) = \left\{ \omega \in \Omega : \max_{\delta \in \Delta} [r(\delta)]^{-1} |\eta_h(\delta,\, t_0)| \leq \kappa \right\}.$$

Let $d_* = \min\{d,\, \overline{\delta}\}$, and

$$b_f(\delta,\, t_0) = \delta^{l-1/p} \int_{-1}^{1} |\varphi(\tau)| \left(\int_{t_0-\delta\tau}^{t_0+\delta\tau} |f^{(l)}(\zeta)|^p\, \mathrm{d}\zeta \right)^{1/p} \mathrm{d}\tau.$$

We define *the ideal bandwidth* $\delta_+(t_0)$ as follows. If the solution δ to the *balance equation*

$$\frac{\kappa \varepsilon B}{c_0 \delta^{\alpha+1/2}} = b_f(\delta,\, t_0) \tag{22}$$

does not exceed d_*, then we set $\delta_+(t_0) = \delta$; otherwise we set $\delta_+(t_0) = d_*$. The ideal

bandwidth $\delta_+(t_0)$ defined in such a manner balances the bias and stochastic error of the estimate. For brevity, we will write δ_+ instead of $\delta_+(t_0)$. Assumption F implies that, for any $\delta \leq d$,

$$b_f(\delta, t_0) \leq \delta^{l-1/p} L |D|^{1/p} \|\varphi\|_1. \tag{23}$$

By definition $\delta_+ \leq \overline{\delta}$; let us verify now that under the premise of the theorem $\delta_+ \geq \underline{\delta}$. Assume that $\delta_+ < d_*$; then

$$\delta_+ \geq \left(\frac{\kappa \varepsilon B}{c_0 L |D|^{1/p} \|\varphi\|_1} \right)^{1/(l+\alpha+1/2-1/p)} \tag{24}$$

and $\delta_+ \geq \underline{\delta}$ due to the first inequality in (14). If $\delta_+ = d_*$, then either $\delta_+ = \overline{\delta}$ or $\delta_+ = d$. In the former case $\delta_+ \geq \underline{\delta}$ by definition, while in the latter $\delta_+ \geq \underline{\delta}$ due to the second inequality in (14).

We bound from above the first term on the right-hand side of (21):

$$|f(t_0) - \langle f, \varphi_\delta \rangle| = \left| f(t_0) - \int_{-1}^{1} f(t_0 + \delta\tau)\varphi(\tau)\,d\tau \right|$$

$$= \left| \int_{-1}^{1} \varphi(\tau) \left[\int_{t_0}^{t_0+\delta\tau} f^{(l)}(\zeta) \frac{(\zeta - t_0)^{l-1}}{(l-1)!}\,d\zeta \right] d\tau \right|$$

$$\leq \delta^{l-1/p} \int_{-1}^{1} |\varphi(\tau)| \left[\int_{t_0}^{t_0+\delta\tau} |f^{(l)}(\zeta)|^p\,d\zeta \right]^{1/p} d\tau \leq b_f(\delta, t_0)$$

(here we have applied the Hölder inequality to the reminder of the $(l-1)$th-order Taylor expansion of f at t_0, and taken into account the moment conditions (7)). Observe that $b_f(\delta, t_0)$ is a monotone increasing function of δ. By the definition of the ideal bandwidth, if $\delta_+ < d_*$, then δ_+ is the solution to the balance equation (22). If $\delta_+ = d_*$, then the solution to the balance equation (22) is greater than $\delta_+ = d_*$. Thus, for any bandwidth $\delta \leq \delta_+$ we have

$$b_f(\delta, t_0) \leq \frac{\kappa \varepsilon B}{c_0 \delta^{\alpha+1/2}} \leq \kappa \varepsilon r(\delta). \tag{25}$$

Assume now that the event $\Omega_\kappa(t_0)$ holds; then for $\delta \leq \delta_+$ one has

$$|\hat{f}_\delta(t_0) - f(t_0)| \leq b_f(\delta, t_0) + \kappa \varepsilon r(\delta) \leq \frac{\kappa \varepsilon B}{c_0 \delta^{\alpha+1/2}} + \kappa \varepsilon r(\delta) \leq 2\kappa \varepsilon r(\delta) \tag{26}$$

(here we have taken into account (21), the definition of $\Omega_\kappa(t_0)$, and (25)). Thus, all segments

$$I_\delta = [\hat{f}_\delta(t_0) - 2\kappa \varepsilon r(\delta), \hat{f}_\delta(t_0) + 2\kappa \varepsilon r(\delta)]$$

associated with $\delta \leq \delta_+$ have at least one point in common, namely $f(t_0)$, and therefore $\bigcap_{\delta \in \Delta, \delta \leq \delta_+} I_\delta \neq \varnothing$. This implies that on the set $\Omega_\kappa(t_0)$ the bandwidth δ_*, given by (12), will be greater than or equal to δ_+. Therefore, we have

$$|\hat{f}_*(t_0) - f(t_0)| \leqslant |\hat{f}_*(t_0) - \hat{f}_{\delta_+}(t_0)| + |\hat{f}_{\delta_+}(t_0) - f(t_0)|$$

$$\leqslant 2\kappa\varepsilon r(\delta_+) + 2\kappa\varepsilon r(\delta_*) + 2\kappa\varepsilon r(\delta_+) \leqslant 6\kappa\varepsilon r(\delta_+).$$

The second and the third inequalities follow from (26), and the fact that $I_{\delta_+} \cap I_{\delta_*} \neq \varnothing$ by definition of δ_* (the distance between centres of the overlapping intervals I_{δ_*} and I_{δ_+} is smaller than sum of their half-widths). The third inequality is an immediate consequence of the fact that $\delta_+ \leqslant \delta_*$ on the set $\Omega_\kappa(t_0)$. Thus, if the event $\Omega_\kappa(t_0)$ holds then

$$|\hat{f}_*(t_0) - f(t_0)| \leqslant 6\kappa\varepsilon r(\delta_+) = 6\kappa\varepsilon \left(\frac{2A\|\varphi\|_1^2}{m^2} + \frac{B^2}{c_0^2 \delta_+^{2\alpha+1}} \right)^{1/2}. \tag{27}$$

Now consider the case of $\omega \in \Omega_\kappa^c(t_0)$. Observe that, for any $\delta \leqslant \delta_*$, $\delta \in \Delta$ and independently of the event $\Omega_\kappa(t_0)$, one has

$$|\hat{f}_*(t_0) - f(t_0)| \leqslant |\hat{f}_*(t_0) - \hat{f}_\delta(t_0)| + |\hat{f}_\delta(t_0) - f(t_0)|$$

$$\leqslant 4\kappa\varepsilon r(\delta) + |\hat{f}_\delta(t_0) - f(t_0)|. \tag{28}$$

This follows from the fact that intersection of I_{δ_*} and I_δ is not empty. By construction, the minimal bandwidth $\delta = \underline{\delta}$ is smaller than or equal to the bandwidth δ_* given by our rule. Therefore applying (28) to $\delta = \underline{\delta}$, we can write

$$|\hat{f}_*(t_0) - f(t_0)| \leqslant 4\kappa\varepsilon r(\underline{\delta}) + |\hat{f}_{\underline{\delta}}(t_0) - f(t_0)|$$

$$\leqslant \frac{8\sqrt{A}\|\varphi\|_1}{m}\kappa + b_f(\underline{\delta}, t_0) + \varepsilon \max_{\delta \in \Delta} |\eta_h(\delta, t_0)|$$

(here we have used (21), (10) and (11)). Further, due to the second inequality in (14) we have $\underline{\delta} \leqslant d$, so that $b_f(\underline{\delta}, t_0) \leqslant b_f(d_*, t_0)$, and, in view of (23),

$$|\hat{f}_*(t_0) - f(t_0)| \leqslant \frac{8\sqrt{A}\|\varphi\|_1}{m}\kappa + \overline{\delta}^{l-1/p} L |D|^{1/p} \|\varphi\|_1 + \varepsilon \max_{\delta \in \Delta} |\eta_h(\delta, t_0)|. \tag{29}$$

Define

$$\theta(t_0) = \max_{\delta \in \Delta} [\|h_\delta\|_2^{-1} |\eta_h(\delta, t_0)|]. \tag{30}$$

Since $\|h_\delta\|_2^{-1} \eta_h(\delta, t_0)$ is the standard Gaussian random variable, and the cardinality of the set Δ does not exceed $N \equiv 2(\ln a)^{-1} \ln(\varepsilon^{-1})$, we can write

$$P(\theta(t_0) > \tau) \leqslant N \int_\tau^\infty \exp(-x^2/2)\,dx, \qquad \forall \tau > 0. \tag{31}$$

Due to Lemma 1, for any $\delta \geqslant \underline{\delta}$ we have

$$\|h_\delta\|_2 \leqslant \left(\frac{2A\|\varphi\|_1^2}{m^2} + \frac{B^2}{c_0^2 \underline{\delta}^{2\alpha+1}} \right)^{1/2} \leqslant \frac{2\sqrt{A}\|\varphi\|_1}{\varepsilon m}.$$

Therefore we obtain from (29) and (30) that

$$|\hat{f}_*(t_0) - f(t_0)| \leqslant \frac{8\sqrt{A}\|\varphi\|_1}{m}\kappa + \overline{\delta}^{l-1/p}L|D|^{1/p}\|\varphi\|_1 + \frac{2\sqrt{A}\|\varphi\|_1}{m}\theta(t_0). \tag{32}$$

Now observe that $\theta(t_0) > \kappa$ on the set $\Omega_\kappa^c(t_0)$. Indeed, since $\|h_\delta\|_2 \leqslant r(\delta)$, we have

$$\Omega_\kappa^c(t_0) = \left\{\omega : \max_{\delta \in \Delta}[r(\delta)]^{-1}|\eta_h(\delta, t_0)| > \kappa\right\} \subseteq \left\{\omega : \max_{\delta \in \Delta}\|h_\delta\|_2^{-1}|\eta_h(\delta, t_0)| > \kappa\right\}$$

$$= \{\omega : \theta(t_0) > \kappa\}. \tag{33}$$

Therefore, it follows from (32) that, on the set $\Omega_\kappa^c(t_0)$,

$$|\hat{f}_*(t_0) - f(t_0)| \leqslant \frac{10\sqrt{A}\|\varphi\|_1}{m}\theta(t_0) + \overline{\delta}^{l-1/p}L|D|^{1/p}\|\varphi\|_1. \tag{34}$$

In addition, (31) and (33) imply that

$$P(\Omega_\kappa^c(t_0)) \leqslant N\int_\kappa^\infty \exp(-x^2/2)\,dx. \tag{35}$$

Now we complete proof of the theorem. It follows from (31) that

$$E[\theta(t_0)]^4 \leqslant O(1)(\ln N)^2 \tag{36}$$

with an absolute constant $O(1)$. Indeed, using (31) and integrating by parts, we have for every $c > 0$

$$E[\theta(t_0)]^4 = \int_0^\infty 4x^3 P\{\theta(t_0) > x\}\,dx \leqslant c^4 + 4N\int_c^\infty x^2\exp(-x^2/2)\,dx$$

$$\leqslant c^4 + 4N\left(c\exp(-c^2/2) + \int_c^\infty \exp(-x^2/2)\,dx\right).$$

Setting $c = \sqrt{2\ln N}$ we come to (36). Let $\chi\{\cdot\}$ be the indicator function of a set. Then taking into account (34), the definition of $\overline{\delta}$, and the first inequality in (14), we obtain

$$[E|\hat{f}_*(t_0) - f(t_0)|^2\chi\{\Omega_\kappa^c(t_0)\}]^{1/2}$$

$$\leqslant O(1)\left[\frac{A\|\varphi\|_1^2}{m^2}\{E[\theta(t_0)]^4 P(\Omega_\kappa^c(t_0))\}^{1/2} + (\overline{\delta}^{l-1/p}L|D|^{1/p}\|\varphi\|_1)^2 P(\Omega_\kappa^c(t_0))\right]^{1/2}$$

$$\leqslant O(1)\left[\frac{\sqrt{A}\|\varphi\|_1}{m}\sqrt{\ln N}[P(\Omega_\kappa^c(t_0))]^{1/4} + \varepsilon^{-(l-1/p)/(\alpha+1/2)}\frac{\sqrt{A}\|\varphi\|_1}{m}[P(\Omega_\kappa^c(t_0))]^{1/2}\right]$$

$$\leqslant O(1)\frac{\sqrt{A}\|\varphi\|_1}{m}[\sqrt{\ln N}[P(\Omega_\kappa^c(t_0))]^{1/4} + \varepsilon^{-2(s+1)}[P(\Omega_\kappa^c(t_0))]^{1/2}].$$

Due to (35), we can choose the constant $C(a, s)$, depending on a and s only, in such a way that for $\kappa = C(a, s)\sqrt{\ln(\varepsilon^{-1})}$ one has

$$\sqrt{\ln N}[P(\Omega_\kappa^c(t_0))]^{1/4} + \varepsilon^{-2(s+1)}[P(\Omega_\kappa^c(t_0))]^{1/2} \leqslant \varepsilon\sqrt{\ln \frac{1}{\varepsilon}}.$$

With the $C(a, s)$ in question we have

$$[E|\hat{f}_*(t_0) - f(t_0)|^2\chi\{\Omega_\kappa^c(t_0)\}]^{1/2} \leqslant O(1)\frac{\sqrt{A}\|\varphi\|_1}{m}\varepsilon\sqrt{\ln \frac{1}{\varepsilon}}, \tag{37}$$

where $O(1)$ depends on a and s only. On the other hand, due to (27), we have

$$[E|\hat{f}_*(t_0) - f(t_0)|^2\chi\{\Omega_\kappa(t_0)\}]^{1/2} \leqslant O(1)\left[\varepsilon\sqrt{\ln \frac{1}{\varepsilon}\left(\frac{A\|\varphi\|_1^2}{m^2} + \frac{B^2}{c_0^2\delta_+^{2\alpha+1}}\right)^{1/2}}\right]. \tag{38}$$

Substituting the lower bound (24) on δ_+ into (38), and combining (37) and (38), we come to (15).

4.2. Proof of Theorem 2

First we show that, under the premise of the theorem, δ_* given in (16) is well defined. Due to the first inequality in (18) we have $\varepsilon \leqslant \exp(-2\gamma)$, and therefore $\gamma^{-1}\ln(\varepsilon^{-1}) \geqslant 2$. Using this fact and the second inequality in (18), we obtain

$$\frac{s+2}{\alpha\gamma}\ln\left(\gamma^{-1}\ln\frac{1}{\varepsilon}\right) \leqslant \frac{s+2}{\alpha\gamma}\sqrt{\gamma^{-1}\ln\frac{1}{\varepsilon}} \leqslant \frac{1}{2}\gamma^{-1}\ln\frac{1}{\varepsilon}.$$

Thus, δ_* is positive for ε obeying (18) and

$$\left[\gamma^{-1}\ln\frac{1}{\varepsilon}\right]^{-1/\alpha} \leqslant \delta_* \leqslant \left[\frac{1}{2}\gamma^{-1}\ln\frac{1}{\varepsilon}\right]^{-1/\alpha}. \tag{39}$$

We have

$$E(|\hat{f}_*(t_0) - f(t_0)|^2)^{1/2} \leqslant \sqrt{2}[|f(t_0) - \langle f, \varphi_{\delta_*}\rangle| + \varepsilon(E|\eta_h(\delta_*, t_0)|^2)^{1/2}], \tag{40}$$

where $\eta_h(\delta_*, t_0) = \int_{-\infty}^\infty h_{\delta_*}(t)\,dW(t)$ is a well-defined Gaussian random variable (see Lemma 2). Now we bound from above the bias and the stochastic error of our estimate.

We have, for any $\delta > 0$,

$$|f(t_0) - \langle f, \varphi_\delta\rangle| = \left|f(t_0) - \int_{-\infty}^\infty f(t_0 + \delta\tau)\varphi(\tau)\,d\tau\right|$$

$$\leqslant \left|\int_{|\tau| > d\delta^{-1}} [f(t_0) - f(t_0 + \delta\tau)]\varphi(\tau)\,d\tau\right|$$

$$+ \left|\int_{|\tau| \leqslant d\delta^{-1}} [f(t_0) - f(t_0 + \delta\tau)]\varphi(\tau)\,d\tau\right|$$

$$= J_1 + J_2. \tag{41}$$

Due to (17) and the second condition in Assumption A3,

$$J_1 \leqslant 2M \int_{|\tau|>d\delta^{-1}} |\varphi(\tau)| \, d\tau \leqslant 2MB(\delta/d)^{s+1}. \tag{42}$$

Further, by the Taylor expansion formula,

$$J_2 \leqslant \left| \sum_{j=0}^{l-1} \frac{f^{(j)}(t_0)}{j!} \delta^j \int_{|\tau|\leqslant d\delta^{-1}} \varphi(\tau)\tau^j \, d\tau \right| + \left| \int_{|\tau|\leqslant d\delta^{-1}} \varphi(\tau) \int_{t_0}^{t_0+\delta\tau} \frac{f^{(l)}(\zeta)(\zeta - t_0)^{l-1}}{(l-1)!} \, d\zeta \, d\tau \right|.$$

$$= J_{21} + J_{22}.$$

First we bound J_{21} from above. Note that, due to Assumption A4,

$$J_{21} = \left| \sum_{j=0}^{l-1} \frac{f^{(j)}(t_0)}{j!} \delta^j \int_{|\tau|>d\delta^{-1}} \varphi(\tau)\tau^j \, d\tau \right|.$$

We have

$$J_{21} \leqslant \left| \sum_{j=0}^{l-1} \frac{f^{(j)}(t_0)}{j!} \delta^j \int_{|\tau|>d\delta^{-1}} \varphi(\tau)\tau^j \left(\frac{\tau\delta}{d}\right)^{s+1-j} \, d\tau \right|$$

$$= \left(\frac{\delta}{d}\right)^{s+1} \left| \sum_{j=0}^{l-1} \frac{f^{(j)}(t_0)}{j!} d^j \int_{|\tau|>d\delta^{-1}} \varphi(\tau)\tau^{s+1} \, d\tau \right|$$

$$\leqslant \left(\frac{\delta}{d}\right)^{s+1} \left| \left[f(t_0 + d) - f(t_0) - \int_{t_0}^{t_0+d} \frac{f^{(l)}(\zeta)(\zeta - t_0)^{l-1}}{(l-1)!} \, d\zeta \right] \int_{|\tau|>d\delta^{-1}} \varphi(\tau)\tau^{s+1} \, d\tau \right|$$

$$\leqslant \left(\frac{\delta}{d}\right)^{s+1} B \left(2M + \left| \int_{t_0}^{t_0+d} \frac{f^{(l)}(\zeta)(\zeta - t_0)^{l-1}}{(l-1)!} \, d\zeta \right| \right).$$

Here the second inequality follows from the Taylor expansion formula for $f(t_0 + d)$ near t_0, and the third inequality is an immediate consequence of Assumption A3 and inequality (17). Applying the Hölder inequality and using Assumption F, we conclude that

$$J_{21} \leqslant \left(\frac{\delta}{d}\right)^{s+1} B(2M + L|D|^{1/p} d^{l-1/p}). \tag{43}$$

To bound J_{22}, we apply the Hölder inequality and take into account Assumption A3:

$$J_{22} \leqslant \delta^{l-1/p} \int_{|\tau| \leqslant d\delta^{-1}} |\varphi(\tau)| \, |\tau|^{l-1/p} \left[\int_{t_0}^{t_0+\delta\tau} |f^{(l)}(\zeta)|^p \, d\zeta \right]^{1/p} d\tau$$

$$\leqslant \delta^{l-1/p} L |D|^{1/p} \left(\int_{|\tau| \leqslant 1} |\varphi(\tau)| \, d\tau + \int_{|\tau| \geqslant 1} |\varphi(\tau)| \, |\tau|^{l-1/p} \, d\tau \right)$$

$$\leqslant \delta^{l-1/p} L |D|^{1/p} (B + \|\varphi\|_1). \tag{44}$$

Thus, combining (41)–(44), we obtain, for any $\delta > 0$,

$$|f(t_0) - \langle f, \varphi_\delta \rangle| \leqslant 4MB(\delta/d)^{s+1} + L|D|^{1/p}(B + \|\varphi\|_1)[(\delta/d)^{s+1}d^{l-1/p} + \delta^{l-1/p}].$$

It follows from the third and forth inequalities in (18) and from (39) that $(\delta_*/d)^{s+1}d^{l-1/p} \leqslant \delta_*^{l-1/p}$ and $(\delta_*/d)^{s+1} \leqslant \delta_*^{l-1/p}d^{1/p}$, respectively. Therefore, taking into account (39), we have

$$|f(t_0) - \langle f, \varphi_{\delta_*} \rangle| \leqslant 4MB|D|^{1/p}\delta_*^{l-1/p} + 2L|D|^{1/p}(B + \|\varphi\|_1)\delta_*^{l-1/p}$$

$$\leqslant 6\tilde{L}|D|^{1/p}(B + \|\varphi\|_1)\left(\frac{1}{2}\gamma^{-1} \ln \frac{1}{\varepsilon} \right)^{-(l-1/p)/\alpha}, \tag{45}$$

where $\tilde{L} = \max\{L, M\}$. Due to Lemma 2, we have

$$\varepsilon(\mathrm{E}|\eta_h(\delta_*, t_0)|^2)^{1/2} = \varepsilon\|h_{\delta_*}\|_2 \leqslant \sqrt{2}\|\varphi\|_1 \varepsilon \left(\frac{A}{m^2} + \frac{\exp(2\gamma\delta_*^{-\alpha})}{c_0^2\delta_*} \right)^{1/2}.$$

Substituting δ_* from (16), we conclude that

$$\varepsilon(\mathrm{E}|\eta_h(\delta_*, t_0)|^2)^{1/2} \leqslant \sqrt{2}\|\varphi\|_1 \left[\frac{\varepsilon\sqrt{A}}{m} + \frac{1}{c_0}\left(\gamma^{-1} \ln \frac{1}{\varepsilon} \right)^{-\frac{s+3/2}{\alpha}} \right]. \tag{46}$$

Combining (46), (45) and (40), and taking into account that $s + 3/2 > l - 1/p$, we come to (19).

4.3. Proof of Theorem 3

The proof is based on the standard technique for deriving minimax lower bounds in nonparametric estimation problems; see, for example, Korostelev and Tsybakov (1993, Chapter 2). In what follows C_i, $i = 1, 2, \ldots$, denote positive constants depending on l, p, L and K only.

Let $\mathscr{S}(\mathbb{R})$ denote the class of infinitely differentiable rapidly decreasing functions on $\mathbb{R}$, that is, f belongs to $\mathscr{S}(\mathbb{R})$ if $t^k f^{(m)}(t)$ tends to zero as $|t| \to \infty$ for every non-negative k and m. It is well known that the Fourier transform maps $\mathscr{S}(\mathbb{R})$ onto itself. Pick a function $\hat{\psi} \in \mathscr{S}(\mathbb{R})$ such that:

  (i) $\hat{\psi}$ is symmetric, $\hat{\psi}(-\lambda) = \hat{\psi}(\lambda)$;

(ii) $\hat{\psi}$ is equal to 1 on an interval slightly smaller than $[1, 2]$ (and $[-2, -1]$), and vanishes outside $[-2, -1] \cup [1, 2]$.

For a real number $\delta > 0$, define $f_\delta(t) = \theta \delta^{-1} \psi(t/\delta)$, where θ is a positive real constant to be chosen. Observe that $\hat{f}_\delta(\lambda) = \theta \hat{\psi}(\delta \lambda)$, and f_δ can be recovered from its Fourier transform by the standard inversion formula

$$f_\delta(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \hat{f}_\delta(\lambda)\, e^{i\lambda t}\, d\lambda = \frac{\theta}{2\pi\delta} \int_{1 \leq |\lambda| \leq 2} \hat{\psi}(\lambda)\, e^{i\lambda t/\delta}\, d\lambda.$$

First, we verify that for appropriate choice of θ the function f_δ belongs to the class $W_p^l(L)$. We have

$$f_\delta^{(l)}(t) = \theta \delta^{-l-1} \psi^{(l)}(t/\delta),$$

and, for any $p \in [1, \infty]$,

$$\left(\int_{-\infty}^{\infty} |f_\delta^{(l)}(t)|^p\, dt \right)^{1/p} \leq \theta \delta^{-l-1} \left(\int_{-\infty}^{\infty} |\psi^{(l)}(t/\delta)|^p\, dt \right)^{1/p}$$

$$\leq C_1 \theta \delta^{-l-1+1/p}$$

(here we have taken into account that $\psi \in \mathscr{S}(\mathbb{R})$). Thus, if we choose

$$\theta = C_2 L \delta^{l+1-1/p} \tag{47}$$

then $f_\delta \in W_p^l(L)$.

Now consider two hypotheses $H_0 : f = f_0 \equiv 0$ and $H_1 : f = f_1 \equiv f_\delta$, and let P_0 and P_1 be the probability measures associated with the corresponding processes

$$dy_0 = \varepsilon\, dW(t), \quad t \in \mathbb{R},$$

$$dy_1 = (Kf_1)(t)\, dt + \varepsilon\, dW(t), \quad t \in \mathbb{R}.$$

We have

$$|f_0(0) - f_1(0)| = |f_\delta(0)| = \theta \delta^{-1} |\psi(0)| \geq \theta \delta^{-1} = C_2 L \delta^{l-1/p} \equiv r. \tag{48}$$

For an arbitrary estimator $\hat{f}(0)$ of $f(0)$,

$$\sup_{f \in W_p^l(L)} E_f |\hat{f}(0) - f(0)|^2 \geq \max_{f = f_0, f_1} E_f |\hat{f}(0) - f(0)|^2$$

$$\geq \frac{r^2}{4} \max_{i=0,1} P_i \{ |\hat{f}(0) - f(0)| \geq r/2 \}. \tag{49}$$

Now we bound from below the last probability using Proposition 2.2.2 from Korostelev and Tsybakov (1993). Observe that, for any $\delta > 0$,

$$\|Kf_\delta\|_2^2 = \int_{-\infty}^{\infty} |\hat{K}(\lambda)|^2 |\hat{f}_\delta(\lambda)|^2 \, d\lambda$$

$$\leqslant 2c_1^2\theta^2\delta^{-1} \int_1^2 \exp(-2\gamma|\lambda|^\alpha\delta^{-\alpha})|\hat{\psi}(\lambda)|^2 \, d\lambda$$

$$\leqslant C_3 L^2 \delta^{-2l+1-2/p} \exp(-2\gamma\delta^{-\alpha}) < \infty \tag{50}$$

(here we have taken into account Assumption K3$'$ and equation (47)). Therefore the probability measures P_0 and P_1 are equivalent, and the likelihood ratio is given by

$$\Lambda(f_0, f_\delta; y_0) = \frac{dP_1}{dP_0}(y_0(t), \, t \in \mathbb{R})$$

$$= \exp\left\{ \frac{1}{\varepsilon} \int_{-\infty}^{\infty} (Kf_\delta)(t) \, dW(t) - \frac{1}{2\varepsilon^2} \int_{-\infty}^{\infty} |(Kf_\delta)(t)|^2 \, dt \right\}$$

(cf. Liptser and Shiryayev 1977, §7.1). If we choose

$$\delta = \delta_* = C_4^{-1/\alpha} \left[\gamma^{-1} \ln\frac{1}{\varepsilon} + \frac{(l+1/2-1/p)}{\alpha\gamma} \ln\left(\gamma^{-1} \ln\frac{1}{\varepsilon} \right) \right]^{-1/\alpha}, \tag{51}$$

then, by (50) and (20), $\varepsilon^{-2}\|Kf_{\delta_*}\|_2^2$ remains bounded as ε tends to zero. In addition, by Proposition 2.2.2 of Korostelev and Tsybakov (1993), one can choose C_4 in such a way that the probability in (49) will be greater than, say, $\frac{1}{4}$. Substituting (51) into (49) and taking into account (48), we complete the proof.

Acknowledgements

References

Abramovich, F. and Silverman, B. (1998) Wavelet decomposition approaches to statistical inverse problems. *Biometrika*, **85**, 115–129.

Anderssen, R.S. (1980) On the use of linear functionals for Abel-type integral equations. In R. Anderssen, F. De Hoog and M. Lucas (eds), *The Application and Numerical Solution of Integral Equations*. Amsterdam: Sijthoff and Noordhof International Publishers.

Carrol, R.J. and Hall, P. (1988) Optimal rates of convergence for deconvolving a density. *J. Amer. Statist. Assoc.*, **83**, 1184–1186.

Donoho, D.L. (1995) Non-linear solution of linear inverse problems by wavelet–vaguelette decomposition. *Appl. Comput. Harmon. Anal.*, **2**, 101–126.

Efromovich, S. (1997) 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.

Fan, J. (1991a) Global behavior of deconvolution kernel estimates. *Statist. Sinica*, **19**, 541–551.

Fan, J. (1991b) On the optimal rates of convergence for nonparametric deconvolution problems. *Ann. Statist.*, **19**, 1257–1272.

Goldberg, M. (1979) A method of adjoints for solving some ill-posed equations of the first kind. *Appl. Math. Comput.*, **5**, 123–130.

Goldenshluger, A. and Nemirovski, A. (1997) On spatially adaptive estimation of nonparametric regression. *Math. Methods Statist.*, **6**(2), 135–170.

Hirschmann, I. and Widder, D. (1955) *The Convolution Transform*. Princeton, NJ: Princeton University Press.

Korostelev, A. and Tsybakov, A. (1993) *Minimax Theory of Image Reconstruction*, Lecture Notes in Statist. 82. New York: Springer-Verlag.

Lepskii, O. (1991) Asymptotically minimax adaptive estimation I: Upper bounds. Optimally adaptive estimates. *Theory Probab. Appl.*, **36**, 682–697.

Lepskii, O. (1992) Asymptotically minimax adaptive estimation II: Schemes without optimal adaptation. Adaptive estimators. *Theory Probab. Appl.*, **37**, 433–448.

Lepskii, O., Mammen, E. and Spokoiny, V. (1997) Optimal spatial adaptation to inhomogeneous smoothness: an approach based on kernel estimates with variable bandwidth selectors. *Ann. Statist.*, **25**(3), 929–947.

Liptser, R. and Shiryayev, A. (1977) *Statistics of Random Processes I*. New York: Springer-Verlag.

Mair, B. and Ruymgaart, F.H. (1996) Statistical inverse estimation in Hilbert scale. *SIAM J. Appl. Math.*, **56**, 1424–1444.

Masry, E. (1991) Multivariate probability density deconvolution for stationary random processes. *IEEE Trans. Inform. Theory*, **37**, 1105–1115.

Nychka, D. and Cox, D.D. (1989) Convergence rates for regularized solutions of integral equations from discrete noisy data. *Ann. Statist.*, **17**, 556–572.

O'Sullivan, F. (1986) A statistical perspective on ill-posed inverse problems. *Statist. Sci.*, **1**, 502–527.

Stefanski, L. and Carrol, R.J. (1990) Deconvoluting kernel density estimators. *Statistics*, **21**, 169–184.

Wahba, G. (1990) *Spline Models for Observational Data*. Philadelphia: Society for Industrial and Applied Mathematics.