Abstract

A new variable bandwidth selector for kernel estimation is proposed. The application of this bandwidth selector leads to kernel estimates that achieve optimal rates of convergence over Besov classes. This implies that the procedure adapts to spatially inhomogeneous smoothness. In particular, the estimates share optimality properties with wavelet estimates based on thresholding of empirical wavelet coefficients.

1 Introduction

In nonparametric curve estimation the statistical analysis may focus on the inference of the qualitative structure of the analysed curve. Often, interesting features of the curve are connected with spatially inhomogeneous smoothness. In this case, curve estimates that are spatially adaptive are appropriate.

A variety of such procedures have been proposed in the literature. In Breiman, Friedman, Olshen and Stone (1983) piecewise constant least squares estimates are considered with a data adaptive choice of the pieces (CART). More generally, Friedman (1991) uses variable knot splines (MARS). Knot points are added, removed and allocated recursively using cross validation techniques. These methods have shown good performance in simulations and real data examples. However, no asymptotic theory is available.

Mammen and van de Geer (1993) discuss penalized least squares curve estimation for spatial inhomogeneous curves. They propose penality terms which allow more spatial inhomogeneity than the usual L^2-norms of derivatives of the curve. The estimates turn out to be variable knot splines (see also Mammen (1991)). Results on rates of convergence and a pointwise asymptotic distribution theory are given.

Müller and Stadtmüller (1987), Staniswalis (1989), and Brockmann, Gasser and Hermann (1993) propose kernel estimation with locally variable bandwidth selectors. The calculation of local bandwidths is based on pilot estimation of local smoothness characteristics. An asymptotic analysis is available here, however, only under additional smoothness conditions on the curve (for a discussion of this point see also Gijbels and Mammen, 1994). Spatially adaptive local polynomial regression estimates were introduced and discussed in Fan and Gijbels (1993). In a series of papers D. Donoho, I. Johnstone, G. Kerkyacharian and D. Picard have shown that wavelet analysis offers a powerful technology for spatial adaptive curve estimation. Curve estimates based on thresholding empirical wavelet coefficients are nearly minimal for a wide range of loss functions and smoothness classes (see Donoho et al., 1993, Kerkyacharian and Picard, 1993, Delyon and Juditsky, 1994). Up to a log factor the estimates achieve the same risk as a variable knot spline with optimally placed (deterministic) knot points (ideal spatial adaptation). This holds for every function (see Donoho and Johnstone, 1993). [For a comparison of wavelet estimates and local polynomial regression estimates with variable bandwidth selector see Fan et al., 1993].

In this paper, a new variable bandwidth kernel estimate is proposed. The bandwidth selector is based on a modification of a procedure for adaptive estimation due to Lepskii (1990, 1991, 1992). We show that this estimate is a reasonable alternative to wavelet estimates. It shares some decision theoretical optimality properties with wavelets. Furthermore, it possesses the simple mathematical interpretation due to kernel estimates. In particular, we prove near minimaxity and ideal spatial adaptation of this estimate.

Our model and our procedure will be described in the next section. Section 3 contains our results. The proofs are postponed to Section 4.

2 A data adaptive local bandwidth selector

In this paper we consider the white noise model

$$dY(t) = f(t)dt + \sigma dW(t) \;\; (0 \le t \le 1), \tag{2.1}$$

where $W(t) \;\; (0 \le t \le 1)$ is a Brownian motion and f is an unknown (regression) function. Performance of estimates of f is studied for $\sigma \to 0$. Model (2.1) gives an asymptotic description for density estimation with i.i.d. observations and for nonparametric regression with i.i.d. Gaussian errors and sample size of order σ^{-2} [see Brown and Low (1990), Low (1992), Nussbaum (1993)].

We will study kernel estimates $\hat{f}_h$ with kernel K and bandwidth h:

$$\hat{f}_h(x) = \int K_h(x-t)dY(t), \tag{2.2}$$

where $K_h(x) = h^{-1}K(x/h)$. We write also $f_h(x) = \int K_h(x-t)f(t)dt = E\hat{f}_h(x)$. We assume that the kernel K has compact support (say, $[-1,1]$), that it is continuous, and that $\int K(u)du = 1$ and $\int u^i K(u)du = 0$ (for $1 \le i \le k$) with k specified below. For $t < h$ and $t > 1-h$ the kernel K_h is replaced by boundary kernels K_h^t (kernels with support $[-t,h]$ and $[-h,1-t]$, respectively). We assume further that all K_h^t are uniformly bounded, fulfil $\int K_h^t(u)du = 0$, and have k vanishing moments. We set $d_K^2 = \sup\{\int L^2(u)du$: $L = K$ or $L = K^t$ for a t with $0 \le t < h$ or $1-h < t \le 1\}$. For simplicity, our notation will not take into account the modifications at the boundary, in particular we will skip the superscript t in K_h^t.

With fixed $a > 1$ and $0 < h_\sigma^* \le 1$ we define

$$\hat{h}_\sigma(t) = \sup\{h \in H_\sigma : |\hat{f}_h(t) - \hat{f}_\eta(t)| \le D\frac{\sigma}{\sqrt{\eta}}\sqrt{1 + \ln\frac{h_\sigma^*}{\eta}} \;\text{ for all }\;\; \eta < h, \eta \in H_\sigma\},$$

where H_σ is the grid

$$H_\sigma = \{h \in [\sigma^2, h_\sigma^*] : h = a^{-j}h_\sigma^*,\; j = 0,1,2,\ldots\}.$$

We write L_σ for the number $\#H_\sigma$ of elements of H_σ. The constant D will be chosen below.

We propose the estimate $\hat{f}(t) = \hat{f}_{\hat{h}(t)}(t)$. A modification of $\hat{f}$ based on piecewise constant choices of $\hat{h}$ is discussed in Lepskii and Spokoiny (1994). The construction of $\hat{h}(t)$ is a modification of a general approach for adaptation given in Lepskii (1991). The bandwidth $\hat{h}(t)$ has a nice statistical interpretation. It is the largest bandwidth h

such that $\hat{f}_h(t)$ does not differ "significantly" from kernel estimates with smaller bandwidth: One chooses a resolution level such that no significant features are visible on a finer resolution level. This approach has a principal difference to wavelet estimation techniques based on thresholding of empicial wavelet coefficients. Empirical wavelet coefficients are related to the values

$$Z_{j,\sigma}(t) = \hat{f}_{2^{-j}h^*_\sigma}(t) - \hat{f}_{2^{-j-1}h^*_\sigma}(t).$$

A kernel estimate analogue of the wavelet threshold estimates would look like

$$\tilde{f}(t) = \hat{f}_{h^*_\sigma}(t) + \sum_{j\geq 0} Z_{j,\sigma}(t)\mathbf{1}(|Z_{j,\sigma}(t)| \geq C_{j,\sigma})$$

with appropriate threshold values $C_{j,\sigma}$. In particular, in contrast to $\hat{f}$, this method is based on comparison of neighbored resolution levels. It may find that for arbitrary many resolution levels "significant" differences are present.

We will study the rate of convergence of $\hat{f}$ over balls $B^s_{p,q}(M)$ in Besov spaces

$$B^s_{p,q}\ (0 < M,\ 1 \leq p, q \leq +\infty,\ s > 0).$$

The following characterisation of a Besov ball will become helpful in our calculations.

$$B^s_{p,q}(M) = \{f : \|f\|_{B^s_{p,q}} \leq M\}, \tag{2.3}$$

where

$$\|f\|_{B^s_{p,q}} = \begin{cases} \|f\|_p + \left[\int_0^1 h^{-sq}\|\text{osc } f(\cdot,h)\|^q_p \frac{dh}{h}\right]^{1/q} & \text{if } q < \infty, \\ \|f\|_p + \sup\limits_{0\leq h\leq 1} h^{-s}\|\text{osc } f(\cdot,h)\|_p & \text{if } q = +\infty. \end{cases} \tag{2.4}$$

In (2.4) $\|f\|_p$ is the L_p-norm $\|f\|^p_p = \int_0^1 |f|^p$. Furthermore, for the definition of the local oscillation osc $f(x,h)$ of the function f an arbitrary $r \in I\!N$ with $r \geq s$ and a real u have to be chosen. The constant u has to fulfill

$$\begin{aligned} 1 \leq u \leq +\infty \quad & \text{if} \quad sp > 1, \\ 1 \leq u < +\infty \quad & \text{if} \quad sp = 1, \\ 1 \leq u < p(1-\text{sp})^{-1} \quad & \text{if} \quad sp < 1. \end{aligned}$$

With this choice of r and u the local oscillation osc $f(x,h)$ of f is defined as

$$\text{osc } f(x,h) = \begin{cases} \inf \sup\limits_{|y-x|\leq h} |f(y) - P(y)|, & \text{if } u = +\infty, \\ \inf \left[\frac{1}{2h}\int\limits_{|y-x|\leq h} |f(y)-P(y)|^u dy\right]^{1/u} & \text{if } u < +\infty. \end{cases} \tag{2.5}$$

The infimum in (2.5) is taken over all polynomials of order r.

A proof that $\| \quad \|_{B^s_{p,q}}$ is a norm of $B^s_{p,q}$ can be found in Triebel (1992) (Section 3.5.1). Other equivalent norms are discussed there, too.

We will study maximal $L_{p'}$ risks of $\hat{f}$ over $B^s_{p,q}$ balls [We make the additional restriction that the functions are uniformly bounded (say by L). For $sp > 1$ this holds automatically]:

$$R_\sigma(\hat{f}, B^s_{p,q}, p') = \sup_{\substack{f \in B^s_{p,q}(M), \\ |f| \leq L}} E_f \|\hat{f} - f\|^{p'}_{p'}. \tag{2.6}$$

For simplicity, our notation does not always indicate every dependence. For instance, remember that $\hat{f}$ depends on σ and the choice of D, a and h^*_σ. Furthermore, it depends on the kernel K (and its number k of vanishing moments).

3 Near minimaxity and ideal spatial adaptation

In this section we show that our curve estimate achieves optimal rates of convergence over Besov classes. For the parameters of the Besov classes we make the usual assumptions:

$$\begin{aligned} 1 &\leq p, q \leq +\infty, \\ 1 &\leq p' < +\infty, \\ s &> \Big(\frac{1}{p} - \frac{1}{p'}\Big)_+. \end{aligned}$$

For the case that $s \leq \frac{1}{p}$ together with $q < +\infty$ hold, we need the additional condition that the kernel K can be decomposed as

$$K(u) = 2M(u) - \frac{1}{2}M(\frac{u}{2}),$$

where M is a bounded function with compact support (say, $[-1/2, +1/2]$) and with $\int M(u)du = 1$. Without any indication in the notation, modifications of M are used again at the boundary. Note that $\int K(u)du = 1$ and $\int uK(u)du = 0$.

We are now ready to state our main result.

Theorem 1 *For the choices $h^*_\sigma = \sigma^{\frac{2}{2s+1}}$, $D > 2 + \sqrt{8d^2_K(p'+2)}$, and for $k > [s]$ the risks of $\hat{f}$ satisfy*

$$R_\sigma(\hat{f}, B^s_{p,q}, p') \leq \begin{cases} const. \sigma^{p'r} & \text{if } \; sp > \frac{p'-p}{2}, \\ const. \left(\sigma\sqrt{\ln 1/\sigma}\right)^{p'r'} [\ln 1/\sigma]^{\frac{4}{(p'-2)(2s+1)}} & \text{if } \; sp = \frac{p'-p}{2} \\ const. \left(\sigma\sqrt{\ln 1/\sigma}\right)^{p'r'} & \text{if } \; sp < \frac{p'-p}{2}. \end{cases} \tag{3.1}$$

if σ is small enough. Here

$$r = \frac{2s}{2s+1},$$

$$r' = \frac{2(s - \frac{1}{p} + \frac{1}{p'})}{2(s - \frac{1}{p}) + 1}$$

and const. is some absolute constant depending on p' only.

The exponent of σ in (3.1) gives the optimal rate. For $sp \neq \frac{p'-p}{2}$ this holds also for the logarithmic factor. Small choices of the class parameter p correspond to Besov classes that contain functions with spatially inhomogeneous smoothness. Because our estimates achieve optimal rates in all Besov classes this shows that the estimates adapt well to spatially inhomogeneous smoothness. For a discussion of minimax rates in Besov spaces we refer to Donoho et al. (1993) and Delyon and Juditsky (1994).

For the interpretation of the exponents in (3.1) let us shortly remark that for the case of $sp \leq \frac{p'-p}{2}$ we have $p' > 2$ and a positive denominator $2(s - \frac{1}{p}) + 1 > 0$ in the exponent. For this to become obvious, note that in this case we have $p' > p$, and applying our condition $s > \left(\frac{1}{p} - \frac{1}{p'}\right)_+$ we obtain: $\frac{p'}{2} - 1 \geq sp + \frac{p}{2} - 1 > \left(\frac{1}{p} - \frac{1}{p'}\right)p + \frac{p}{2} - 1 = \frac{p}{p'}\left(\frac{p'}{2} - 1\right)$. Because of $p' > p$ this implies $p' > 2$ and $sp - 1 + \frac{p}{2} > 0$.

The procedure $\hat{f}$ requires explicit knowledge of s. The next theorem helps to understand the performance of $\hat{f}$ in case of unknown degree s of smoothness.

Theorem 2 *For D and k as in Theorem 1 and for h^*_σ with $\sigma^{\frac{2}{2s+1}} \leq h^*_\sigma \leq 1$ one gets for σ small enough*

$$R_\sigma(\hat{f}, B^s_{p,q}, p') = \begin{cases} const.\left(\sigma\sqrt{\ln(1/\sigma)}\right)^{p'r} & \text{if } sp > \frac{p'-p}{2} \\ const.\left(\sigma\sqrt{\ln 1/\sigma}\right)^{p'r'} [\ln 1/\sigma]^{4(p'-2)^{-1}(2s+1)^{-1}} & \text{if } sp = \frac{p'-p}{2} \\ const.\left(\sigma\sqrt{\ln 1/\sigma}\right)^{p'r'} & \text{if } sp < \frac{p'-p}{2}. \end{cases} \tag{3.2}$$

Here r, r' are the same as in Theorem 1.

Using $h^*_\sigma = 1$ gives the optimal rate for $sp < \frac{p'-p}{2}$ and an additional logarithmic factor for $sp > \frac{p'-p}{2}$. The choice $h^*_\sigma = \sigma^{2/(2s'+1)}$ leads to an optimal estimation for $s = s' = s - \frac{1}{p} + \frac{1}{p'}$. The additional logarithmic factor appears only for $s < s'$ (and $sp > \frac{p'-p}{2}$).

It is known from Lepskii (1990) and Brown and Low (1992) that in the pointwise estimation one has to pay an additional logarithmic factor for not knowing s. However, here we consider global and not pointwise risks. We conjecture that the additional logarithmic factor in (3.2) can be removed when a more sophisticated adaptive curve estimate is used.

Now we turn to state a property of $\hat{f}$ which was been called ideal spatial adaptation in Donoho and Johnstone (1993). For quadratic loss we would like to compare the risk of $\hat{f}$ with $\inf E\|\hat{f}_{h(\cdot)} - f\|_2^2$, where the infimum runs over all (deterministic) variable bandwidth $h(\cdot)$. The minimizing $h(\cdot)$ was called an oracle in Donoho and Johnstone (1993). Note that $E\|\hat{f}_{h(\cdot)} - f\|_2^2 = \int_0^1 (f_{h(t)}(t) - f(t))^2 dt + \int_0^1 \mathrm{Var}\ \hat{f}_{h(t)}(t) dt$. Here it suffices to consider the pointwise risk $E(\hat{f}_{h(t)}(t)-f(t))^2$. We are only able to compare the risk of $\hat{f}(t)$ with $r_{\text{opt}}(t) = \inf_{0\le h\le 1}\Big[\sup_{0\le\eta\le h}(f_\eta(t)-f(t))^2+\text{Var}\,(\hat{f}_h(t))\Big]$.

We denote the minimizing bandwidth by h_{opt}.

Theorem 3 *Choose $h_\sigma^*=1$ and $D>0$. For all functions f and all variances σ^2 small enough it holds for $t\in(0,1)$ with uniform constants L_0, L_1:*

$$E(\hat{f}(t)-f(t))^2 \le (L_0+L_1\ln(\frac{1}{h_{\text{opt}}}))r_{\text{opt}}(t).$$

There exist versions of Theorem 3 for nonquadratic losses.

4 Proofs

Proof of Theorem 1. For $f\in B^s_{p,q}(M)$ and for $t\in[0,1]$ we define

$$h_\sigma^-(t,f)=\sup\{h\in H_\sigma : |f_\eta(t)-f(t)|\le \frac{\sigma}{\sqrt{\eta}}\sqrt{1+\ln\frac{h_\sigma^*}{\eta}}\ \text{ for all }\ \eta\le h\} \tag{4.1}$$

and

$$h_\sigma(t,f)=a\ h_\sigma^-(t,f).$$

For any $f\in B^s_{p,q}(M)$ we consider

$$R_\sigma^+(f)=\int_0^1 E_f|\hat{f}(t)-f(t)|^{p'}\mathbf{1}(A_\sigma(t,f))dt \tag{4.2}$$

and

$$R_\sigma^-(f)=\int_0^1 E_f|\hat{f}(t)-f(t)|^{p'}\mathbf{1}(A_\sigma^c(t,f))dt, \tag{4.3}$$

where $A_\sigma(t,f)$ denotes the random event $A_\sigma(t,f)=\{\hat{h}(t)\ge h_\sigma^-(t,f)\}$ and $A_\sigma^c(t,f)$ its complement.

Clearly, we obtain

$$R_\sigma(\hat{f},B^s_{p,q},p')\le \sup_{f\in B^s_{p,q}(M)} R_\sigma^+(f)+\sup_{f\in B^s_{p,q}(M)} R_\sigma^-(f). \tag{4.4}$$

We start by proving

$$\sup_{f\in B^s_{p,q}(M)} R_\sigma^-(f)\le \text{const}\ \left[\frac{\sigma}{\sqrt{h_\sigma^*}}\right]^{p'}. \tag{4.5}$$

Proof of (4.5). We fix now an arbitrary function $f\in B^s_{p,q}(M)$ and write

$$H_\sigma^-=\{h\in H_\sigma : h<h_\sigma^-(t,f)\}.$$

For any $h_1,h_2\in H_\sigma^-$ with $h_2<h_1$ we put

$$B_\sigma(t,h_1,h_2)=\left\{|\hat{f}_{h_1}(t)-\hat{f}_{h_2}(t)|>D\frac{\sigma}{\sqrt{h_2}}\sqrt{1+\ln\frac{h_\sigma^*}{h_2}}\right\}.$$

With this notation we get

$$A^c_\sigma(t,f) = \bigcup_{h\in H^-_\sigma} \{\hat{h}(t) = h\} = \bigcup_{h\in H^-_\sigma} \; \bigcup_{\eta\le h, \eta\in H^-_\sigma} B_\sigma(t,2h,\eta). \tag{4.6}$$

Using the Cauchy-Schwarz inequality we obtain

$$\begin{aligned} R^-_\sigma(f) &= \int_0^1 \sum_{h\in H^-_\sigma} E_f|\hat{f}_h(t) - f(t)|^{p'} \mathbf{1}(\hat{h}(t) = h)dt \\ &\le \int_0^1 \sum_{h\in H^-_\sigma} r_\sigma(h,t)\Big[\sum_{\substack{\eta\le h \\ \eta\in H^-_\sigma}} P_f\{B_\sigma(t,2h,\eta)\}\Big]^{1/2} dt, \end{aligned} \tag{4.7}$$

where

$$r_\sigma(h,t) = \{E_f|\hat{f}_h(t) - f(t)|^{2p'}\}^{1/2}.$$

Forgetting the modifications of K at the boundary we can write

$$\hat{f}_h(t) = f_h(t) + \frac{\sigma d_K}{\sqrt{h}} \xi_\sigma(t,h), \tag{4.8}$$

where $d_K^2 = \int_{-\infty}^{+\infty} K^2(u)du$ and

$$\xi_\sigma(t,h) = \frac{\sqrt{h}}{\sigma d_K}\Big[\sigma h^{-1} \int_0^1 K\left(\frac{t-u}{h}\right) dW(u)\Big].$$

Note that for $h \le t \le 1-h$ the random variable $\xi_\sigma(t,h)$ is standard Gaussian. [For $t < h$ and $t > 1-h$ it is a mean zero Gaussian variable with variance ≤ 1]. Because of (4.1) and (4.2) we have for $h < h^-_\sigma(t,f)$,

$$f_h(t) - f(t) \le \frac{\sigma}{\sqrt{h}}\sqrt{1 + \ln\frac{h^*_\sigma}{h}}.$$

This implies

$$|\hat{f}_h(t) - f(t)| \le \frac{\sigma}{\sqrt{h}}\sqrt{1 + \ln\frac{h^*_\sigma}{h}} + \frac{d_K\sigma}{\sqrt{h}}|\xi_\sigma(t,h)|.$$

Thus we obtain

$$r_\sigma(h,t) \le \text{const.}\ \left(\frac{\sigma}{\sqrt{h}}\sqrt{1 + \ln\frac{h^*_\sigma}{h}}\right)^{p'}. \tag{4.9}$$

Combining (4.7) and (4.9) gives

$$R^-_\sigma(f) \le \text{const.} \int_0^1 \left\{ \sum_{h\in H^-_\sigma} \left(\frac{\sigma}{\sqrt{h}}\sqrt{1+\ln\frac{h^*_\sigma}{h}}\right)^{p'} \left(\sum_{\substack{\eta\le h \\ \eta\in H^-_\sigma}} P_f(B_\sigma(t,2h,\eta)) \right)^{1/2} \right\} dt. \tag{4.10}$$

Using (4.8) we can bound

$$B_\sigma(t,2h,\eta) \subseteq \left\{ |f_{2h}(t) - f_\eta(t)| + \frac{\sigma d_K}{\sqrt{2h}}|\xi_\sigma(t,2h)| + \frac{\sigma d_K}{\sqrt{\eta}}|\xi_\sigma(t,\eta)| \ge \frac{D\sigma}{\sqrt{\eta}}\sqrt{1 + \ln\frac{h^*_\sigma}{\eta}} \right\}.$$

For $\eta \le h < h_\sigma(t,f)$ we have

$$|f_{2h}(t) - f_\eta(t)| \le |f_{2h}(t) - f(t)| + |f_\eta(t) - f(t)| \le \frac{2\sigma}{\sqrt{\eta}}\sqrt{1 + \ln\frac{h_\sigma^*}{\eta}}.$$

This gives

$$\begin{aligned} B_\sigma(t, 2h, \eta) \;\; &\subseteq \;\; \left\{\xi_\sigma(t, 2h) \ge \frac{D-2}{2d_K}\sqrt{1 + \ln\frac{h_\sigma^*}{\eta}}\right\} \\ &\cup \left\{\xi_\sigma(t, \eta) \ge \frac{D-2}{2d_K}\sqrt{1 + \ln\frac{h_\sigma^*}{\eta}}\right\} \end{aligned}$$

and

$$P_f(B_\sigma(t, 2h, \eta)) \le \text{const.}\left(\frac{\eta}{h_\sigma^*}\right)^{\frac{(D-2)^2}{8d_K^2}}.$$

Inserting this in (4.10) and using $\frac{(D-2)^2}{8d_K^2} \ge p' + 2$ and $\eta < h$ we have

$$R_\sigma^-(f) \le \text{const.} \sum_{h \in H_\sigma^-} \left[\frac{\sigma\sqrt{1 + \ln\frac{h_\sigma^*}{h}}}{\sqrt{h}}\right]^{p'} \cdot \left[\sum_{\substack{\eta \le h \\ \eta \in H_\sigma^-}} \left(\frac{\eta}{h_\sigma^*}\right)^{p'+2}\right]^{1/2}.$$

But

$$\sum_{\substack{\eta \le h \\ \eta \in H_\sigma^-}} \left(\frac{\eta}{h_\sigma^*}\right)^{p'+2} \le \text{const.}\left(\frac{h}{h_\sigma^*}\right)^{p'+2}$$

and

$$\begin{aligned} R_\sigma^-(f) \;\; &\le \;\; \text{const.}\left(\frac{\sigma}{\sqrt{h_\sigma^*}}\right)^{p'} \cdot \sum_{h \in H_\sigma^-} \left\{\left(\frac{h}{h_\sigma^*}\right)^{-p'/2}\left(1 + \ln\frac{h_\sigma^*}{h}\right)^{p'/2} \cdot \left(\frac{h}{h_\sigma^*}\right)^{p'/2+1}\right\} \\ &\le \;\; \text{const.}\left(\frac{\sigma}{\sqrt{h_\sigma^*}}\right)^{p'} \end{aligned}$$

where const. is some absolute constant depending on p' only. Thus the proof of (4.5) is complete. It remains to show

$$R_\sigma^+(f) \le \begin{cases} \text{const.}\, \sigma^{p'r} & \text{if } \; sp > \frac{p'-p}{2} \\ \text{const.}\left(\sigma\sqrt{\ln\frac{1}{\sigma}}\right)^{p'r'} \cdot [\ln 1/\sigma]^{4(p'-2)^{-1}(2s+1)^{-1}} & \text{if } \; sp = \frac{p'-p}{2}, \\ \text{const.}\left(\sigma\sqrt{\ln 1/\sigma}\right)^{p'r'} & \text{if } \; sp < \frac{p'-p}{2}. \end{cases} \tag{4.11}$$

Proof of (4.11). Note that, by means of (4.8) we obtain

$$\begin{aligned} R_\sigma^+(f) \;\; &\le \;\; \int_0^1 E_f\left\{\left[|\hat f(t) - \hat f_{h_\sigma^-(t,f)}(t)| + |f_{h_\sigma^-(t,f)}(t) - f(t)| + \frac{\sigma d_K |\xi_\sigma(t, h_\sigma^-(t,f))|}{\sqrt{h_\sigma^-(t,f)}}\right]^{p'}\right. \\ &\left. \mathbf{1}(A_\sigma(t,f))\right\} dt. \end{aligned} \tag{4.12}$$

By the definition of $\hat{h}(t)$, we have on $A_\sigma(t,f)$

$$|\hat{f}(t) - \hat{f}_{h_\sigma^-(t,f)}(t)| \le \frac{\sigma}{\sqrt{h_\sigma^-(t,f)}} \cdot \sqrt{1 + \ln \frac{h_\sigma^*}{h_\sigma^-(t,f)}}. \tag{4.13}$$

Moreover, by the definition of $h_\sigma(t,f)$ (see (4.1), (4.2)) we conclude for $h_\sigma^-(t,f) < h_\sigma(t,f)$

$$|f_{h_\sigma^-(t,f)}(t) - f(t)| \le \frac{\sigma}{\sqrt{h_\sigma^-(t,f)}} \sqrt{1 + \ln \frac{h_\sigma^*}{h_\sigma^-(t,f)}}. \tag{4.14}$$

By inserting (4.13) and (4.14) in (4.12) and by using that $\xi_\sigma(t, h_\sigma^-(t,f))$ has bounded moments we arrive at

$$R_\sigma^+(f) \le \text{const.} \int |\psi_\sigma(h_\sigma(t,f))|^{p'} dt, \tag{4.15}$$

where

$$\psi_\sigma(h) = \frac{\sigma}{\sqrt{h}} \sqrt{1 + \ln \frac{h_\sigma^*}{h}}.$$

The right-hand side of (4.15) can be written as

$$= \text{const.} \sum_{h \in H_\sigma} \int_{S_h} |\psi_\sigma(h)|^{p'} dt$$

where $S_h = \{t : h_\sigma(t,f) = h\}$. On S_h it holds that

$$\Delta_h(t) \ge \psi_\sigma(h), \tag{4.16}$$

where $\Delta_h(t) = \sup_{\eta \le h} |f_\eta(t) - f(t)|$.
This follows from the definition (4.1) of $h_\sigma(t,f)$ and the monotonicity of $\Delta_h(t)$ and $\psi_\sigma(h)$ in h.

We define now a function $p_1(h)$. For $sp > \frac{p'-p}{2}$ we put $p_1(h) \equiv \min\{p, p'\}$. For the case of $sp \le \frac{p'-p}{2}$ we put

$$p_1(h) = \begin{cases} 0 & \text{if} \quad h > h_1(\sigma) \\ p & \text{if} \quad h_1(\sigma) \ge h \ge h_2(\sigma) \\ p' & \text{if} \quad h < h_2(\sigma), \end{cases}$$

where $h_1(\sigma) = \left(\sigma\sqrt{\ln(1/\sigma)}\right)^{2/(2s+1)}$ and $h_2(\sigma) = \left(\sigma\sqrt{\ln(1/\sigma)}\right)^{1/(s-\frac{1}{p}+\frac{1}{2})}$. For $sp \le \frac{p'-p}{2}$ we have that $s - \frac{1}{p} + \frac{1}{2} > 0$ [see the remark after Theorem 1]. Therefore, the definition of $h_2(\sigma)$ makes sense. Using (4.16) we obtain

$$\begin{aligned} R_\sigma^+(f) &\le \text{const.} \sum_{h \in H_\sigma} |\psi_\sigma(h)|^{p'-p_1(h)} \int |\Delta_h(t)|^{p_1(h)} dt \\ &= \text{const.} \sum_{h \in H_\sigma} |\psi_\sigma(h)|^{p'-p_1(h)} \|\Delta_h\|_{p_1(h)}^{p_1(h)}. \end{aligned} \tag{4.17}$$

We will use now that

$$\sup_f \sup_{0\le h\le 1} h^{-s}\|\Delta_h\|_p < +\infty \tag{4.18}$$

$$\sup_f \sup_{0\le h\le 1} h^{-s'}\|\Delta_h\|_{p'} < +\infty \ \text{ if } \ sp \le \frac{p'-p}{2}. \tag{4.19}$$

Here the supremum runs over all functions f in the Besov ball $B^s_{pq}(M)$ [which are in supnorm uniformly bounded by L]. The quantity s' is defined as $s' = s - \frac{1}{p} + \frac{1}{p'}$.

Before we come to the proof of (4.18) and (4.19) let us show that these both statements imply Theorem 1. For $sp > \frac{p'-p}{2}$ and $p \ge p'$ we obtain from (4.17) and (4.18)

$$R^+_\sigma(f) \le \text{const.} \sum_{h\in H_\sigma} h^{sp'}. \tag{4.20}$$

The right-hand side of (4.20) is a geometric series. It can be bounded by const. $(h^*_\sigma)^{sp'} = \text{const.}\ \sigma^{p'\frac{2s}{2s+1}}$. This shows the statement of Theorem 1 for this case.

If $sp > \frac{p'-p}{2}$ but $p < p'$, then again by (4.17) and (4.18) we get

$$R^+_\sigma(f) \le \text{const.} \sum_{h\in H_\sigma} h^{sp}\psi_\sigma(h)^{p'-p} \le (h^*_\sigma)^{sp}|\psi_\sigma(h^*_\sigma)|^{p'-p} \sum_{h\in H_\sigma} \left(\frac{h}{h^*_\sigma}\right)^{sp} \left(\frac{\psi_\sigma(h)}{\psi_\sigma(h^*_\sigma)}\right)^{p'-p}.$$

By the definitions of h^*_σ and $\psi_\sigma(h)$ one gets $|h^*_\sigma|^{sp}|\psi_\sigma(h^*_\sigma)|^{p'-p} = \sigma^{p\frac{2s}{2s+1}}$. It remains to note that for $sp - \frac{p'-p}{2} > 0$

$$\sum_{h\in H_\sigma} \left(\frac{h}{h^*_\sigma}\right)^{sp} \left(\frac{\psi_\sigma(h)}{\psi_\sigma(h^*_\sigma)}\right)^{p'-p} = \sum_{h\in H_\sigma} \left(\frac{h}{h^*_\sigma}\right)^{sp-\frac{p'-p}{2}} \left(1 + \ln\frac{h^*_\sigma}{h}\right)^{\frac{p'-p}{2}} \le \text{const.}$$

For the case of $sp \le \frac{p'-p}{2}$ we split the summation on the right-hand side of (4.17) into three sums and apply (4.18) and (4.19). We obtain

$$R^+_\sigma(f) \le \text{const.}\ [R_1 + R_2 + R_3]$$

where

$$\begin{aligned}
R_1 &= \sum_{\substack{h>h_1(\sigma)\\ h\in H_\sigma}} \|\Delta_h\|^{p_1(h)}_{p_1(h)}\, |\psi_\sigma(h)|^{p'-p_1(h)} \\
&= \sum_{\substack{h\ge h_1(\sigma)\\ h\in H_\sigma}} |\psi_\sigma(h)|^{p'}, \\
R_2 &= \sum_{\substack{h_1(\sigma)\ge h\ge h_2(\sigma)\\ h\in H_\sigma}} \|\Delta_h\|^{p_1(h)}_{p_1(h)}\, |\psi_\sigma(h)|^{p'-p_1(h)} \\
&= \sum_{\substack{h_1(\sigma)\ge h\ge h_2(\sigma)\\ h\in H_\sigma}} h^{sp}|\psi_\sigma(h)|^{p'-p}, \\
R_3 &= \sum_{\substack{h<h_2(\sigma)\\ h\in H_\sigma}} \|\Delta_h\|^{p_1(h)}_{p_1(h)}\, |\psi_\sigma(h)|^{p'-p_1(h)} \\
&= \sum_{\substack{h<h_2(\sigma)\\ h\in H_\sigma}} h^{s'p'}.
\end{aligned}$$

Comparing ψ_σ for two neighboured elements in H_σ we obtain

$$\frac{\psi_\sigma(h)}{\psi_\sigma(h/a)} = \frac{1}{\sqrt{a}}\left(1 + \frac{\ln a}{1+\ln(h_\sigma^*/h)}\right).$$

For $h \in H_\sigma$ this is bounded away from 0 and 1. This implies

$$\begin{aligned} R_1 &\leq \text{const. } |\psi_\sigma(h_1(\sigma))|^{p'} \\ &\leq \text{const. } \left(\sigma\sqrt{\ln(1/\sigma)}\right)^{p'} |h_1(\sigma)|^{-p'/2} \\ &= \text{const. } \left(\sigma\sqrt{\ln(1/\sigma)}\right)^{p'r}. \end{aligned}$$

For $sp \leq \frac{p'-p}{2}$ this bound is always of smaller order than the right-hand side of (3.1) because of $r' \leq r$ for $sp \leq \frac{p'-p}{2}$.

The sum R_3 is a geometric series and can be bounded by

$$\text{const.}\,[h_2(\sigma)]^{s'p'} = \text{const.}\left[\sigma\sqrt{\ln(1/\sigma)}\right]^{2p'\frac{s-\frac{1}{p}+\frac{1}{p'}}{2(s-\frac{1}{p})+1}} = \text{const.}\left[\sigma\sqrt{\ln(1/\sigma)}\right]^{p'r'},$$

which, again, is of the same order than the right-hand side of (3.1) (for $sp \leq \frac{p'-p}{2}$).

It remains to bound R_2. We obtain

$$R_2 \leq \text{const. } \left(\sigma\sqrt{\ln(1/\sigma)}\right)^{p'-p} \sum_{\substack{h_2(\sigma)<h<h_1(\sigma)\\ h\in H_\sigma}} h^{sp-\frac{p'-p}{2}}.$$

For the case of $sp - \frac{p'-p}{2} < 0$ this gives

$$\begin{aligned} R_2 &\leq \text{const. } \left(\sigma\sqrt{\ln(1/\sigma)}\right)^{p'-p} |h_2(\sigma)|^{sp-\frac{(p'-p)}{2}} \\ &= \text{const. } \left(\sigma\sqrt{\ln(1/\sigma)}\right)^{2p'\frac{s-\frac{1}{p}+\frac{1}{p'}}{2(s-\frac{1}{p})+1}}. \end{aligned}$$

For the case of $sp - \frac{p'-p}{2} = 0$ we obtain $r'p' = p' - p$ and

$$R_2 \leq \text{const. } \left(\sigma\sqrt{\ln(1/\sigma)}\right)^{p'-p} \ln\frac{h_1(\sigma)}{h_2(\sigma)}.$$

The last two estimates give (3.1) for $sp \leq \frac{p'-p}{2}$.

We come now to the proofs of (4.18) and (4.19).

Proof of (4.18). For $sp > 1$ the definition (2.5) of local oscillations with $u = +\infty$ implies that for $0 \leq t \leq 1$ and for each $\varepsilon > 0$ there exists a polynomial $P_{t,h}$ of degree k with

$$\sup_{|x-t|\leq h} |f(x) - P_{t,h}(x)| \leq \operatorname{osc} f(t,h) + \varepsilon.$$

This implies

$$\sup_{|x-t|\le h} |f(x) - f(t) - P_{t,h}(x) + P_{t,h}(t)| \le 2\text{osc}\, f(t,h) + 2\varepsilon.$$

Since K has k vanishing moments, we obtain $\Delta_h(t) \le$ const. osc $f(t,h)$. This shows (4.18).

For $sp \le 1$ and $q = +\infty$ we apply the definition (2.5) of local oscillations with $u = 1$. Arguing similarly as above we obtain

$$|f_h(t) - f_{h/2}(t)| \le \text{const. osc}\, f(t,h).$$

Because of $\|f_\eta - f\|_p \to 0$ (for $\eta \to 0$) it holds that

$$\|f_h - f\|_p \le \sum_{i\ge 0} \|f_{2^{-i}h} - f_{2^{-i-1}h}\|_p.$$

Now $h^{-\beta}\|\text{osc}\, f(t,h)\|_p \le$ const. provides

$$\|f_h - f\|_p \le h^{+\beta} \sum_{i\ge 0} \text{const. } 2^{-i\beta} \le \text{const. } h^\beta.$$

This shows (4.18).

For $sp \le 1$ and $q < +\infty$ we recall that K can be decomposed as

$$K(x) = 2M(x) - \frac{1}{2}M(\frac{x}{2}).$$

Now

$$\begin{aligned} f_h(x) - f(x) &= \int M(t)[2f(x+ht) - f(x+2ht) - f(x)]dt \\ &\le \text{const.} \int_{|t|\le 1} |2f(x+ht) - f(x+2ht) - f(x)|dt. \end{aligned}$$

The equation (4.18) follows by application of Theorem 3.5.3 in Triebel (1992) and by using the embedding $B^s_{p,q} \subset B^s_{p,\infty}$.

Proof of (4.19). For $p' \ge p$ the Besov space $B^{s'}_{p',q}$ can be embedded into $B^s_{p,q}$ for all $q \ge 1$ (see Triebel, 1992). This means that

$$\sup_{f \in B^s_{p,q}(M)} \|f\|_{B^{s'}_{p',q}} < +\infty.$$

Note also that $s'p' < 1, = 1$, or > 1, if and only if $sp < 1, = 1$, and > 1, respectively. Thus, (4.19) can be shown by the same arguments as (4.18).

Proof of Theorem 2. We proceed similarly as in the proof of Theorem 1. The term $R^-_\sigma(f)$ can be bounded again by const. $\left(\frac{\sigma}{\sqrt{h^*_\sigma}}\right)^{p'}$. This is sufficient. For $sp \le \frac{p'-p}{2}$ also the term $R^+_\sigma(f)$ can be treated as in the proof of Theorem 1.

For the case of $sp > \frac{p'-p}{2}$ another definition of $p_1(h)$ will be used for estimating $R^+_\sigma(f)$. The choice

$$p_1(h) = \begin{cases} 0 & \text{if } h > \left(\sigma\sqrt{\ln(1/\sigma)}\right)^{2/(2s+1)} \\ p \quad x & \text{if } h \le \left(\sigma\sqrt{\ln(1/\sigma)}\right)^{2/2s+1} \end{cases}$$

will do.

Proof of Theorem 3. We fix t and write $r = r(t)$, $f = f(t)$, $\hat{f} = \hat{f}(t)$, $f_h = f_h(t)$. We define

$$\begin{aligned} h_\sigma &= \sup\{h \le 1 : |f_\eta - f| \le \frac{\sigma}{\sqrt{h}}\sqrt{1 + \ln(1/h)} \text{ for all } \eta \le h\}, \\ h^-_\sigma &= \sup\{h \in H_\sigma : h < h_\sigma\}. \end{aligned}$$

Note that the definition of h_σ differs slightly from the definition (4.1) of $h_\sigma(t, f)$. We write

$$\begin{aligned} r_\sigma &= E_f|\hat{f} - f|^2 = r^-_\sigma + r^+_\sigma, \quad \text{where} \\ r^-_\sigma &= E_f|\hat{f} - f|^2 \mathbf{1}(\hat{h} < h^-_\sigma), \\ r^+_\sigma &= E_f|\hat{f} - f|^2 \mathbf{1}(\hat{h} \ge h^-_\sigma). \end{aligned}$$

Using the arguments given in the proof of Theorem 1 we obtain

$$r^-_\sigma \le \text{const. } \sigma^2$$

and

$$r^+_\sigma \le \text{const. } \frac{\sigma^2}{h_\sigma}(1 + \ln(1/h_\sigma)).$$

Combining of these inequalities provides

$$r_\sigma \le \text{const. } \frac{\sigma^2}{h_\sigma}(1 + \ln(1/h_\sigma)). \tag{4.21}$$

We put now

$$r^{(1)}(h) = \sup_{0 \le \eta \le h} (f_\eta - f)^2$$

and

$$r^{(2)}_\sigma(h) = \text{Var } (\hat{f}_h).$$

Then

$$r_{\text{opt}} = r_{\text{opt}}(t) = \inf_{0 \le h \le 1} r^{(1)}(h) + r^{(2)}_\sigma(h).$$

Suppose that the infimum is attained at h_{opt}. For h_σ we get

$$r^{(2)}_\sigma(h_\sigma)(1 + \ln(1/h_\sigma)) \cdot c_0 = r^{(1)}_\sigma(h_\sigma), \tag{4.22}$$

where $c_0 = \int K^2(u)du$.

We treat now the cases of $h_{\rm opt} \geq h_0$ and $h_{\rm opt} < h_0$ separately. Note that $r^{(1)}(h)$ is monotone increasing in h and that $r_\sigma^{(2)}(h)$ is monotone decreasing in h. Suppose first that $h_{\rm opt} \geq h_0$. Applying (4.22) gives

$$\frac{\sigma^2}{h_\sigma} \cdot c_0 = r_\sigma^{(2)}(h_\sigma) = \frac{r_\sigma^{(1)}(h_\sigma)}{c_0(1+\ln(1/h_\sigma))} \leq \frac{r_\sigma^{(1)}(h_{\rm opt})}{c_0(1+\ln(1/h_\sigma))} \leq \frac{r_{\rm opt}}{c_0(1+\ln(1/h_\sigma))}. \qquad (4.23)$$

For the case that $h_{\rm opt} \leq h_0$ we have

$$\frac{\sigma^2}{h_\sigma} c_0 \leq r_{\rm opt}. \qquad (4.24)$$

The formulas (4.21), (4.23), and (4.24) give

$$r_\sigma \leq \text{const. } r_{\rm opt}(1+\ln(1/h_{\rm opt})).$$

This is the statement of Theorem 3.

References

Brockmann, M., Gasser, T. and Herrmann, E. (1993). Locally adaptive bandwidth choice for kernel regression estimators. *J. Amer. Statist. Assoc.* **88** 1302-1309.

Brown, L.D. and Low, M.G. (1990). Asymptotic equivalence of nonparametric regression and white noise. *Technical Report*, Cornell University.

Brown, L.D. and Low, M.G. (1992). Superefficiency and lack of adaptability in functional estimation. Preprint.

Delyon, B. and Juditsky, A. (1994). Wavelet estimators, global error measures revisited. *Technical Report*, IRISA, Rennes.

Donoho, D.L. and Johnstone, I.M. (1992). Minimax estimation via wavelet shrinkage. *Biometrika*, to appear.

Donoho, D.L. and Johnstone, I.M. (1993). Ideal spatial adaptation by wavelet shrinkage. *Biometrika*, to appear.

Donoho, D.L., Johnstone, I.M., Kerkyacharian, G. and Picard, D. (1994). Wavelet shrinkage: asymptopia? *J. Royal Statist. Soc.*, to appear.

Gijbels, I. and Mammen, E. (1994). On local adaptivity of kernel estimates with plug-in local bandwidth selectors. Preprint, Sonderforschungsbereich 373, Humboldt Universität, Berlin

Fan, J. and Gijbels, I. (1993). Data-driven bandwidth selection in local polynomial fitting: variable bandwidth and spatial adaptation. *Discussion Paper # 9313*, Institute of Statistics, Catholic University of Louvain, Louvain-la-Neuve, Belgium.

Fan, J., Hall, P., Martin, M. and Patil, P. (1993). Adaptation to high spatial inhomogeneity based on wavelets and on local linear smoothing. *Research report SMS-59-93*, Australian National University.

Kerkyacharian, G. and Picard, D. (1993). Density estimation by kernel and wavelet method, Optimality in Besov space. *Statistics and Probab.* Letters 18, 327-336.

Lepskii, O.V. (1990). One problem of adaptive estimation in Gaussian white noise. *Theory Probab. Appl.* **35** N. 3 459-470

Lepskii, O.V. (1991). Asymptotic minimax adaptive estimation. 1. Upper bounds. *Theory Probab. Appl.* **36** N. 4 654-659

Lepskii, O.V. (1992). Asymptotic minimax adaptive estimation. 2. Statistical model without optimal adaptation. Adaptive estimators. *Theory Probab. Appl.* **37** N. 3 468-481

Lepskii, O.V. and Spokoiny, V.G. (1994). Local adaptivity to inhomogeneous smoothness. 1. Resolution level. Preprint, Institut für Angewandte Analysis und Stochastik, Berlin, Germany

Low, M.G. (1992). Renormalization and white noise approximation for nonparametric functional estimation. *Ann. Statist.* **20** 545-544.

Mammen, E. (1991). Nonparametric regression under qualitative smoothness assumptions. *Ann. Statist.* **19** 741-759.

Mammen, E. and van de Geer, S. (1993). Locally adaptive regression splines. *Technical Report*, Humboldt-Universität Berlin.

Müller, H.-G. and Stadtmüller, U. (1987). Variable bandwidth kernel estimators of regression curves. *Ann. Statist.* **15** 182-201.

Nussbaum, M. (1993). Asymtptotic equivalence of density estimation in white noise. *Technical Report*, Institute of Applied Analysis and Stochastics, Berlin.

Staniswalis, J.G.S. (1989). Local bandwidth selection for kernel estimates. *J. Amer. Statist. Assoc.* **84** 284-288.

Triebel, H. (1992). Theory of function spaces II. Birkhäuser, Basel.