Optimal spatial adaptation
to inhomogeneous smoothness:
An approach based on kernel estimates
with variable bandwidth selectors

Oleg Lepskii¹, Enno Mammen², Vladimir G. Spokoiny³

submitted: 31st October 1995

¹ Institute for System Analysis
Prospekt 60–Let Oktjabrja, 9
Moscow, 117312
Russia

² Humboldt Universität Berlin
Mathematisch-Naturwissenschaftliche
Fakultät II
Institut für Mathematik
Unter den Linden 6
D – 10099 Berlin
Germany

³ Weierstraß-Institut
für Angewandte Analysis
und Stochastik
Mohrenstraße 39
D – 10117 Berlin
Germany

Preprint No. 191
Berlin 1995

This work was supported by the Deutsche Forschungsgemeinschaft, Sonderforschungsbereich 373 “Quantifikation und Simulation ökonomischer Prozesse”, Berlin, Germany. The research of the first author was supported by the Deutsche Forschungsgemeinschaft under a personal grant 436 RUS 17/241/93.
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 penalty 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

$$dY(t) = f(t)dt + \sigma dW(t) \quad (0 \leq t \leq 1),$$

where $W(t)$ $(0 \leq t \leq 1)$ 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 $\sigma^{-2}$ [see Brown and Low (1990), Low (1992), Nussbaum (1993)].

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

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

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

With fixed $a > 1$ and $0 < h^* \leq 1$ we define

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

where $H_\sigma$ is the grid

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

We write $L_\sigma$ for the number $\#H_\sigma$ of elements of $H_\sigma$. 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 empirical 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

\[
\hat{f}(t) = \hat{f}_{h_\sigma}(t) + \sum_{j \geq 0} Z_{j,\sigma}(t) \mathbf{1}(|Z_{j,\sigma}(t)| \geq C_{j,\sigma})
\]

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 \( B_{p,q}^s(M) \) in Besov spaces

\[
B_{p,q}^s(M) = \{ f : \| f \|_{B_{p,q}^s} \leq M \},
\]

where

\[
\| f \|_{B_{p,q}^s} = \begin{cases} 
\| f \|_p + \left[ \int_0^1 h^{-sp} \| \text{osc}_f(\cdot, h) \|_{L_p} \right]^{1/q} & \text{if } q < \infty, \\
\| f \|_p + \sup_{0 \leq h \leq 1} h^{-sp} \| \text{osc}_f(\cdot, h) \|_p & \text{if } q = +\infty.
\end{cases}
\]

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

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

With this choice of \( r \) and \( u \) the local oscillation \( \text{osc}_f(x, h) \) of \( f \) is defined as

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

The infimum in (2.5) is taken over all polynomials of order \( r \).
A proof that \( \|B_{p,q}^s \| \) is a norm of \( B_{p,q}^s \) 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 \( B_{p,q}^s \) balls [We make the additional restriction that the functions are uniformly bounded (say by \( L \)). For \( sp > 1 \) this holds automatically]:

\[
R_\sigma(f, B_{p,q}^s, p') = \sup_{f \in B_{p,q}^s(M)} E_f \| f - \hat{f} \|_{p'}^p. \tag{2.6}
\]

For simplicity, our notation does not always indicate every dependence. For instance, remember that \( \hat{f} \) depends on \( \sigma \) and the choice of \( D, a \) and \( h_0^* \). 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:

\[
1 \leq p, q \leq +\infty, \\
1 \leq p' < +\infty, \\
s > \left( \frac{1}{p} - \frac{1}{p'} \right)^+. \notag
\]

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 \( h_0^* = \sigma^{\frac{2}{p'+1}} \), \( D > 2 + \sqrt{8dK(p' + 2)} \), and for \( k > [s] \) the risks of \( \hat{f} \) satisfy

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

if \( \sigma \) 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 \( \sigma \) in (3.1) gives the optimal rate. For \( sp \neq \frac{p' - 2}{2} \) 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 Delion and Juditsky (1994).

For the interpretation of the exponents in (3.1) let us shortly remark that for the case of \( sp \leq \frac{p' - 2}{2} \) 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: \( \frac{p'_2 - 1}{2} \geq sp + \frac{p}{2} - 1 > \left( \frac{1}{p} - \frac{1}{p'} \right)p + \frac{p}{2} - 1 = p_2 \left( \frac{p'}{2} - 1 \right) \). 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 \( h^*_\sigma \) with \( \sigma_{\frac{3}{2}+1} \leq h^*_\sigma \leq 1 \) one gets for \( \sigma \) small enough

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

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

Using \( h^*_\sigma = 1 \) gives the optimal rate for \( sp < \frac{p' - 2}{2} \) and an additional logarithmic factor for \( sp > \frac{p' - 2}{2} \). 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 \( sp > \frac{p' - 2}{2} \)).

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 \( E\| \hat{f}_h (\cdot) - f \|_2^2 = f_0^1 (f(\hat{f}_h(t)) - f(t))^2dt + f_0^1 \text{Var } \hat{f}_h(t)dt \). Here it
suffices to consider the pointwise risk \( E(\hat{f}_h(t) - f(t))^2 \). We are only able to compare the risk of \( \hat{f}(t) \) with \( r_{opt}(t) = \inf_{h \leq h_{\tau}} \left[ \sup_{h \leq h_{\tau}} (f_{\eta}(t) - f(t))^2 + \text{Var}(\hat{f}_h(t)) \right] \).

We denote the minimizing bandwidth by \( h_{\text{opt}} \).

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

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

There exist versions of Theorem 3 for nonquadratic losses.

4 Proofs

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

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

and

\[
h_{\sigma}(t,f) = h_{\sigma}^-(t,f).
\]

For any \( f \in B_{p,q}^*(M) \) we consider

\[
R_{\sigma}^+(f) = \int_0^1 E_f|\hat{f}(t) - f(t)|^{p'} 1(A_{\sigma}(t,f))dt
\]

and

\[
R_{\sigma}^-(f) = \int_0^1 E_f|\hat{f}(t) - f(t)|^{p'} 1(A_{\sigma}^c(t,f))dt,
\]

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

Clearly, we obtain

\[
R_{\sigma}(\hat{f}, B_{p,q}^*(M)) \leq \sup_{f \in B_{p,q}^*(M)} R_{\sigma}^+(f) + \sup_{f \in B_{p,q}^*(M)} R_{\sigma}^-(f).
\tag{4.4}
\]

We start by proving

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

**Proof of (4.5).** We fix now an arbitrary function \( f \in B_{p,q}^*(M) \) 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

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

6
With this notation we get
\[ A_{\sigma}^{-}(t, f) = \bigcup_{h \in H_{\sigma}} \{ \hat{h}(t) = h \} = \bigcup_{h \in H_{\sigma}} \bigcup_{\eta \leq h, \eta \in H_{\sigma}} B_{\sigma}(t, 2h, \eta). \] (4.6)

Using the Cauchy-Schwarz inequality we obtain
\[ R_{\sigma}^{-}(f) = \int_{0}^{1} \sum_{h \in H_{\sigma}} E[f|\hat{h}(t) - f(t)|^p] 1(\hat{h}(t) = h) dt \]
\[ \leq \int_{0}^{1} \sum_{h \in H_{\sigma}} r_{\sigma}(h, t) \left( \sum_{\eta \leq h} \sum_{\eta \in H_{\sigma}} P_f(B_{\sigma}(t, 2h, \eta)) \right)^{1/2} dt, \]
where
\[ r_{\sigma}(h, t) = \{ E[f|\hat{h}(t) - f(t)|^{2p'}] \}^{1/2}. \]

Forgetting the modifications of \( K \) at the boundary we can write
\[ \hat{f}_{h}(t) = f_{h}(t) + \frac{\sigma d_{K}}{\sqrt{h}} \xi_{\sigma}(t, h), \] (4.8)

where \( d_{K} = \int_{-\infty}^{\infty} K^{2}(u) du \) and
\[ \xi_{\sigma}(t, h) = \frac{\sqrt{h}}{\sigma d_{K}} \left[ \frac{\sigma h^{-1}}{\sqrt{h}} \int_{0}^{1} K \left( \frac{t - u}{h} \right) dW(u) \right]. \]

Note that for \( h \leq t \leq 1 - h \) 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 \( \leq 1 \).] Because of (4.1) and (4.2) we have for \( h < h_{\sigma}^{-}(t, f) \),
\[ f_{h}(t) - f(t) \leq \frac{\sigma}{\sqrt{h}} \left[ 1 + \ln \frac{h_{\sigma}^{*}}{h} \right] . \]

This implies
\[ |\hat{f}_{h}(t) - f(t)| \leq \frac{\sigma}{\sqrt{h}} \left[ 1 + \ln \frac{h_{\sigma}^{*}}{h} + \frac{d_{K} \sigma}{\sqrt{h}} |\xi_{\sigma}(t, h)|. \]

Thus we obtain
\[ r_{\sigma}(h, t) \leq \text{const.} \left( \frac{\sigma}{\sqrt{h}} \left[ 1 + \ln \frac{h_{\sigma}^{*}}{h} \right] \right)^{p'}. \] (4.9)

Combining (4.7) and (4.9) gives
\[ R_{\sigma}^{-}(f) \leq \text{const.} \int_{0}^{1} \left\{ \sum_{h \in H_{\sigma}} \left( \frac{\sigma}{\sqrt{h}} \left[ 1 + \ln \frac{h_{\sigma}^{*}}{h} \right] \right)^{p'} \left( \sum_{\eta \leq h} \sum_{\eta \in H_{\sigma}} P_f(B_{\sigma}(t, 2h, \eta)) \right)^{1/2} \right\} dt. \] (4.10)

Using (4.8) we can bound
\[ B_{\sigma}(t, 2h, \eta) \subseteq \left\{ |f_{2h}(t) - f_{\eta}(t)| + \frac{\sigma d_{K}}{\sqrt{2h}} |\xi_{\sigma}(t, 2h)| + \frac{\sigma d_{K}}{\sqrt{\eta}} |\xi_{\sigma}(t, \eta)| \geq \frac{D \sigma}{\sqrt{\eta}} \left[ 1 + \ln \frac{h_{\sigma}^{*}}{\eta} \right] \right\}. \]
For $\eta \leq h < h_\sigma(t, f)$ we have

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

This gives

$$B_\sigma(t, 2h, \eta) \subseteq \left\{ \xi_\sigma(t, 2h) \geq \frac{D - 2}{2d_K} \sqrt{1 + \ln \frac{h_\sigma^*}{\eta}} \right\}$$

$$\cup \left\{ \xi_\sigma(t, \eta) \geq \frac{D - 2}{2d_K} \sqrt{1 + \ln \frac{h_\sigma^*}{\eta}} \right\}$$

and

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

Inserting this in (4.10) and using $(D-2)^2 \geq p' + 2$ and $\eta < h$ we have

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

But

$$\sum_{\eta \leq h} \left( \frac{\eta}{h_\sigma^*} \right)^{p'+2} \leq \text{const.} \left( \frac{h}{h_\sigma^*} \right)^{p'+2}$$

and

$$R_\sigma^-(f) \leq \text{const.} \left( \frac{\sigma}{\sqrt{h_\sigma^*}} \right)^{p'} \sum_{h \in \mathcal{H}_\sigma} \left\{ \left( \frac{h}{h_\sigma^*} \right)^{-p'/2} \left( 1 + \ln \frac{h_\sigma^*}{h} \right)^{p'/2} \left( \frac{h}{h_\sigma^*} \right)^{p'/2+1} \right\}$$

$$\leq \text{const.} \left( \frac{\sigma}{\sqrt{h_\sigma^*}} \right)^{p'}$$

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

$$R_\sigma^+(f) \leq \left\{ \begin{array}{ll}
\text{const.} \sigma^p & \text{if } sp > \frac{p'-p}{2} \\
\text{const.} \left( \frac{\sigma \sqrt{1/\sigma}}{\sqrt{h_\sigma^*}} \right)^{p/2} \cdot \left[ \ln 1/\sigma \right]^{p'(p'-2)-(2s+1)-1} & \text{if } sp = \frac{p'-p}{2}, \\
\text{const.} \left( \frac{\sigma \sqrt{1/\sigma}}{\sqrt{h_\sigma^*}} \right)^{p/2} & \text{if } sp < \frac{p'-p}{2}.
\end{array} \right. \quad (4.11)$$

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

$$R_\sigma^+(f) \leq \int_0^1 E_f \left\{ \left[ \hat{f}(t) - \hat{f}_{h_\sigma}(t, f)(t) \right] + |f_{h_\sigma}(t, f) - f(t)| \right\} dt.$$ 

$$1(A_{\sigma}(t, f)) \right\} dt. \quad (4.12)$$
By the definition of \( \hat{h}(t) \), we have on \( A_\sigma(t, f) \)

\[
|\dot{\hat{h}}(t) - \dot{h}_{\sigma}(t, f)(t)| \leq \frac{\sigma}{\sqrt{h_{\sigma}(t, f)}} \sqrt{1 + \ln \frac{h_{\sigma}(t, f)}{h_{\sigma}(t, f)}}. \tag{4.13}
\]

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

\[
|h_{\sigma}(t, f)(t) - f(t)| \leq \frac{\sigma}{\sqrt{h_{\sigma}(t, f)}} \sqrt{1 + \ln \frac{h_{\sigma}(t, f)}{h_{\sigma}(t, f)}}. \tag{4.14}
\]

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

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

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

\[
= \text{const. } \sum_{h \in H_\sigma S_h} \int |\psi_{\sigma}(h)| |p'| dt
\]

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

\[
\Delta_h(t) \geq \psi_{\sigma}(h), \tag{4.16}
\]

where \( \Delta_h(t) = \sup_{\eta \leq h} |f_\eta(t) - f(t)| \).

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 \( sp > \frac{\sigma - p}{2} \) we put \( p_1(h) \equiv \min\{p, p'\} \). For the case of \( sp \leq \frac{\sigma - p}{2} \) we put

\[
p_1(h) = \begin{cases} 
0 & \text{if } h > h_1(\sigma) \\
p & \text{if } h_1(\sigma) \geq h \geq h_2(\sigma) \\
p' & \text{if } h < h_2(\sigma),
\end{cases}
\]

where \( h_1(\sigma) = \left(\sigma \sqrt{\ln(1/\sigma)}\right)^{2/(2s+1)} \) and \( h_2(\sigma) = \left(\sigma \sqrt{\ln(1/\sigma)}\right)^{1/(s-p+1/2)} \). For \( sp \leq \frac{\sigma - p}{2} \) 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

\[
R_{\sigma}^+(f) \leq \text{const. } \sum_{h \in H_\sigma} |\psi_{\sigma}(h)| |p' - p_1(h)| \int |\Delta_h(t)|^{p_1(h)} dt
\]

\[
= \text{const. } \sum_{h \in H_\sigma} |\psi_{\sigma}(h)| |p' - p_1(h)| \|\Delta_h||^{p_1(h)} \tag{4.17}
\]
We will use now that
\[
\sup_f \sup_{0 \leq h \leq 1} h^{-s} \| \Delta_h \|_p < +\infty \quad (4.18)
\]
\[
\sup_f \sup_{0 \leq h \leq 1} h^{-s'} \| \Delta_h \|_{p'} < +\infty \text{ if } sp \leq \frac{p' - p}{2}. \quad (4.19)
\]
Here the supremum runs over all functions \( f \) in the Besov ball \( B_{pq}^s(M) \) [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 \( p \geq p' \) we obtain from (4.17) and (4.18)
\[
R_s^+(f) \leq \text{const.} \sum_{h \in H_\sigma} h^{sp'}. \quad (4.20)
\]
The right-hand side of (4.20) is a geometric series. It can be bounded by const. \((h_\sigma^*)^{sp'} = \text{const.} \sigma^{p'/2+1}\). 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
\[
R_s^+(f) \leq \text{const.} \sum_{h \in H_\sigma} h^{sp} \psi_s(h)^{p' - p} \leq (h_\sigma^*)^{sp} |\psi_\sigma(h_\sigma^*)|^{p' - p} \sum_{h \in H_\sigma} \left( \frac{h}{h_\sigma^*} \right)^{sp} \left( \frac{\psi_\sigma(h)}{\psi_\sigma(h_\sigma^*)} \right)^{p' - p}. \]
By the definitions of \( h_\sigma^* \) and \( \psi_\sigma(h) \) one gets \( |h_\sigma^*|^{sp} |\psi_\sigma(h_\sigma^*)|^{p' - p} = \sigma^{p'/2+1} \). It remains to note that for \( sp - \frac{p' - p}{2} > 0 \)
\[
\sum_{h \in H_\sigma} \left( \frac{h}{h_\sigma^*} \right)^{sp} \left( \frac{\psi_\sigma(h)}{\psi_\sigma(h_\sigma^*)} \right)^{p' - p} = \sum_{h \in H_\sigma} \left( \frac{h}{h_\sigma^*} \right)^{sp - \frac{p' - p}{2}} \left( 1 + \ln \frac{h_\sigma^*}{h} \right)^{\frac{p' - p}{2}} \leq \text{const.}
\]
For the case of \( sp \leq \frac{p' - p}{2} \) we split the summation on the right-hand side of (4.17) into three sums and apply (4.18) and (4.19). We obtain
\[
R_s^+(f) \leq \text{const.} [R_1 + R_2 + R_3]
\]
where
\[
R_1 = \sum_{h \in H_\sigma} \| \Delta_h \|_{p_1(h)} \| \psi_\sigma(h) \|_{p' - p_1(h)} = \sum_{h \in H_\sigma} |\psi_\sigma(h)|^{p'},
\]
\[
R_2 = \sum_{h \in H_\sigma} \| \Delta_h \|_{p_1(h)} \| \psi_\sigma(h) \|_{p' - p_1(h)} = \sum_{h \in H_\sigma} h^{sp} |\psi_\sigma(h)|^{p' - p},
\]
\[
R_3 = \sum_{h \in H_\sigma} \| \Delta_h \|_{p_1(h)} \| \psi_\sigma(h) \|_{p' - p_1(h)} = \sum_{h \in H_\sigma} h^{sp'}. \]
Comparing $\psi_\sigma$ for two neighboured elements in $H_\sigma$ 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

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

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

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

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

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

It remains to bound $R_2$. We obtain

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

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

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

For the case of $sp - \frac{p'-2}{2} = 0$ 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 $sp \leq \frac{p'-2}{2}$.

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

$$\sup_{|x-t| \leq h} |f(x) - P_{t,h}(x)| \leq \text{osc } f(t,h) + \varepsilon.$$
This implies
\[ \sup_{|x-t| \leq h} |f(x) - f(t) - P_{t,h}(x) + P_{t,h}(t)| \leq 2 \text{osc } f(t, h) + 2\varepsilon. \]

Since \( K \) has \( k \) vanishing moments, we obtain \( \Delta_h(t) \leq \text{const. osc } f(t, h). \) This shows (4.18).

For \( sp \leq 1 \) and \( q = +\infty \) we apply the definition (2.5) of local oscillations with \( u = 1. \) Arguing similarly as above we obtain
\[ |f_h(t) - f_{h/2}(t)| \leq \text{const. osc } f(t, h). \]

Because of \( \|f_\eta - f\|_p \to 0 \) (for \( \eta \to 0 \)) it holds that
\[ \|f_h - f\|_p \leq \sum_{i \geq 0} \|f_{2^{-i}h} - f_{2^{-i-1}h}\|_p. \]

Now \( h^{-\beta}\|\text{osc } f(t, h)\|_p \leq \text{const. provides} \)
\[ \|f_h - f\|_p \leq h^{+\beta} \sum_{i \geq 0} \text{const. } 2^{-i\beta} \leq \text{const. } h^\beta. \]

This shows (4.18).

For \( sp \leq 1 \) and \( q < +\infty \) we recall that \( K \) can be decomposed as
\[ K(x) = 2M(x) - \frac{1}{2}M(x/2). \]

Now
\[ f_h(x) - f(x) = \int M(t)[2f(x + ht) - f(x + 2ht) - f(x)]dt \]
\[ \leq \text{const. } \int_{|t| \leq 1} |2f(x + ht) - f(x + 2ht) - f(x)|dt. \]

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

**Proof of (4.19).** For \( p' \geq p \) the Besov space \( B_{p',q}^s \) can be embedded into \( B_{p,q}^s \) for all \( q \geq 1 \) (see Triebel, 1992). This means that
\[ \sup_{f \in B_{p,q}^s(M)} \|f\|_{B_{p',q}^{s'}} < +\infty. \]

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 \( \text{const. } (\frac{\sigma}{\sqrt{h^p}})^{p'}. \) This is sufficient. For \( sp \leq \frac{p'-2}{2} \) also the term \( R_\sigma^+ (f) \) can be treated as in the proof of Theorem 1.
For the case of $sp > \frac{p-2}{2}$ another definition of $p_1(h)$ will be used for estimating $R^+_\sigma(f)$. The choice

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

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

$$h_\sigma = \sup\{h \leq 1 : |f_\eta - f| \leq \frac{\sigma}{\sqrt{h}} \sqrt{1 + \ln(1/h)} \text{ for all } \eta \leq h\},$$

$$h^-_\sigma = \sup\{h \in H_\sigma : h < h_\sigma\}.$$

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

$$r_\sigma = E_f|\hat{f} - f|^2 = r^-_\sigma + r^+_\sigma,$$

where

$$r^-_\sigma = E_f|\hat{f} - f|^2 1(\hat{h} < h^-_\sigma),$$

$$r^+_\sigma = E_f|\hat{f} - f|^2 1(\hat{h} \geq h^-_\sigma).$$

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

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

and

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

Combining of these inequalities provides

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

We put now

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

and

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

Then

$$r_{opt} = r_{opt}(t) = \inf_{0 \leq h \leq 1} r^{(1)}(h) + r^{(2)}(h).$$

Suppose that the infimum is attained at $h_{opt}$. For $h_\sigma$ we get

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

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

13
We treat now the cases of $h_{\text{opt}} \geq h_0$ and $h_{\text{opt}} < h_0$ separately. Note that $r^{(1)}(h)$ is monotone increasing in $h$ and that $r^{(2)}(h)$ is monotone decreasing in $h$. Suppose first that $h_{\text{opt}} \geq h_0$. Applying (4.22) gives

$$
\frac{\sigma^2}{h_\sigma} \cdot c_0 = r^{(2)}_{\sigma}(h_\sigma) = \frac{r^{(1)}_{\sigma}(h_\sigma)}{c_0(1 + \ln(1/h_\sigma))} \leq \frac{r^{(1)}_{\sigma}(h_{\text{opt}})}{c_0(1 + \ln(1/h_\sigma))} \leq \frac{r_{\text{opt}}}{c_0(1 + \ln(1/h_\sigma))}.
$$

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

$$
\frac{\sigma^2}{h_\sigma} c_0 \leq r_{\text{opt}}.
$$

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

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

This is the statement of Theorem 3.
References


15


Recent publications of the
Weierstraß–Institut für Angewandte Analysis und Stochastik

Preprints 1995


164. Dmitry Ioffe: A note on the extremality of the disordered state for the Ising model on the Bethe lattice.


169. Pierluigi Colli, Jürgen Sprekels: Remarks on the existence for the one-dimensional Frémond model of shape memory alloys.


172. Reiner Lauterbach: Symmetry breaking in dynamical systems.


175. Alexander A. Gushchin: On efficiency bounds for estimating the offspring mean in a branching process.

177. Vladimir Maz'ya, Gunther Schmidt: “Approximate approximations” and the cubature of potentials.

178. Sergey V. Nepomnyaschikh: Preconditioning operators on unstructured grids.

179. Hans Babovsky: Discretization and numerical schemes for stationary kinetic model equations.


182. Martin Krupa, Björn Sandstede, Peter Szmolyan: Fast and slow waves in the FitzHugh–Nagumo equation.


186. Björn Sandstede: Center manifolds for homoclinic solutions.

187. Steven N. Evans, Klaus Fleischmann: Cluster formation in a stepping stone model with continuous, hierarchically structured sites.

188. Sybille Handrock-Meyer: Identifiability of distributed parameters for a class of quasilinear differential equations.


190. Wolfgang Härdle, Vladimir G. Spokoiny, Stefan Sperlich: Semiparametric single index versus fixed link function modelling.