On local adaptivity of kernel estimates with plug-in local bandwidth selectors*

Irène Gijbels [1] and Enno Mammen [2]

[1] Institut de Statistique
Université Catholique de Louvain
Voie du Roman Pays 34
B–1348 Louvain-la-Neuve
Belgium

[2] Institut für Stochastik
Fachbereich Mathematik
Humboldt-Universität
Unter den Linden 6
10099 Berlin
Germany

Abstract

In nonparametric function estimation selection of a smoothing parameter is one of the most decisive issues. The performance of smoothing techniques highly depends on the choice of this parameter. Preferable the bandwidth should be determined via a data-driven procedure. In this paper we consider kernel estimators in a white noise model, and investigate whether locally adaptive plug-in bandwidths can achieve optimal global rates of convergence. We hereby consider various classes of functions: Sobolev classes, bounded variation function classes, classes of convex functions and classes of monotone functions. We study the situations of pilot estimation with oversmoothing and without oversmoothing. Our main conclusion is that simple local plug-in bandwidth selectors can not adapt fully to spatial inhomogenity of the regression function.

AMS 1991 subject classification: Primary 62G07; secondary 60G15.
Key words and phrases: kernel estimates, local variable bandwidth, optimal rates, plug-in local bandwidths, spatial inhomogenity, total variation function classes, white noise model.
Short title: Plug-in local bandwidths.

* This paper was completed while the second author was visiting the Institute of Statistics, Catholic University of Louvain, Louvain-la-Neuve, Belgium.

1 Introduction

Nonparametric function estimation can be conducted using various smoothing techniques, such as kernel estimation, local polynomial fitting, nearest neighbour estimation, smoothing splines among others. An important issue in all these procedures is the choice of the smoothing parameter or bandwidth. Choosing this bandwidth in some optimal way is by no means an easy task. A global or constant bandwidth can be considered, but does not incorporate a flexibility to adapt locally to the structure of the underlying unknown curve. In case of curves with locally changing smoothness it is preferable to consider a variable bandwidth: the proportion of smoothing should depend on the local degree of smoothness of the underlying curve to be estimated. An asymptotical optimal local variable bandwidth $h(x)$ is obtained by minimizing the asymptotic expression for the mean square error. This optimal local bandwidth does however depend on unknown quantities, such as for example the density function and its second derivative in case of density estimation, or the second derivative of the regression function in case of estimation of the regression function. One possible approach is to use preliminary estimates, so-called pilot estimates, for the unknown quantities and to substitute those into the expression for the desired optimal local bandwidth. This then reveals a data-adaptive approximation for the theoretical optimal bandwidth $h(x)$.

The question then arises how those plug-in type local variable bandwidth estimators are performing. Do they achieve optimal rates of convergence? Over which classes of functions? Those are the questions which will be addressed in this paper for kernel estimators in the setup of a white noise model.

Using plug-in kernel estimators in practice requires some additional work. Indeed, one is faced with the fact that there are two bandwidths to be selected: the initial smoothing parameter g used for obtaining the pilot estimates and the final data-adaptive bandwidth $\hat{h}(x)$. Hall (1993) studies theoretically the effects of the choice of the initial smoothing parameter on the final kernel estimate in the density estimation setup. Practical plug-in bandwidth procedures have been proposed by Müller and Stadtmüller (1987), Staniswalis (1989) and Brockmann, Gasser and Hermann (1993) in the fixed-design regression model. Müller and Stadtmüller (1987) and Staniswalis (1989) used cross-validation techniques for choosing the pilot bandwidth. Brockmann, Gasser and Hermann (1993) introduce a local adaptive bandwidth selection procedure in this fixed-design regression model using convolution type estimators. Their iterative selection procedure starts with calculating a data-adaptive global bandwidth. The pilot bandwidth is then based on inflation of this global bandwidth. A non plug-in adaptive bandwidth procedure was introduced and discussed by Fan and Gijbels (1993) for local polynomial regression estimators in the random design model. In Donoho and Johnstone (1992, 1993) it has been pointed out that curve estimates based on thresholding empirical wavelet coefficients are closely related to kernel estimates with local variable bandwidths. For comparison of local smoothing techniques and wavelet thresholding methods see Fan et al. (1993a, b).

In order to study the local adaptivity property of kernel estimates we will concentrate on a simple plug-in selection procedure:

- an initial bandwidth g is used to estimate the second derivative of the regression function $m''(x)$. The resulting estimator is denoted by $\tilde{m}''_g(x)$,
- substitution of the derivative estimator in the expression for the optimal local bandwidth leads to the adaptive choice $\hat{h}(x)$.

Application of this simple plug-in rule requires however some caution. One can think immediately of three potential problems which may possibly occur:

(i) Suppose that the pilot bandwidth g is of higher order of magnitude than the final bandwidth $\hat{h}(x)$. This means that for the final calculation of the estimator we work in a neighbourhood which is smaller than the initial neighbourhood determined by g, and hence we can refer to this as a case of *intrapolation*.

(ii) Diametrically to that we have the case of *extrapolation* which happens if the pilot bandwidth g is of smaller order of magnitude than $\hat{h}(x)$. The second derivative estimator is based on data points lying in a smaller interval than that considered in calculating the final estimator.

(iii) Some features in the unknown regression curve, such as wiggles, might be smoothed out in the pilot estimation step leading to a too large local bandwidth $\hat{h}(x)$.

The objective of this paper is to investigate the theoretical behavior of the above described simple plug-in kernel estimators in a white noise model. More precisely, we will point out whether this type of plug-in kernel estimates can still achieve optimal rates of convergence over certain classes of functions. In our study we consider, for $c > 0$, Sobolev classes

$$\mathcal{M}_c^{\mathrm{Sob}} = \{m : \int m''(x)^2 dx \le c^2\}, \tag{1.1}$$

classes of regression functions with derivatives of bounded variation

$$\mathcal{M}_c^{\mathrm{TV}} = \{m : TV(m') \le c\}, \tag{1.2}$$

classes of convex regression functions in $\mathcal{M}_c^{\mathrm{TV}}$ which will be denoted by $\mathcal{M}_c^{\mathrm{conv}}$, and finally the classes $\mathcal{M}_c^{\mathrm{mon}}$ of monotone functions on $[0,1]$ with $|m(1)-m(0)| \le c$. Note that $\mathcal{M}_c^{\mathrm{Sob}} \subset \mathcal{M}_c^{\mathrm{TV}}$. We are interested in the behaviour of kernel estimates $\hat{m}_h(x)$, with local bandwidth $h = \hat{h}(x)$, for regression functions m with locally changing smoothness. As prototype of such a function class we use $\mathcal{M}_c^{\mathrm{TV}}$. Note that $\{m : m$ two times differentiable, $\int |m''| \le c\}$ is dense in $\mathcal{M}_c^{\mathrm{TV}}$ (with respect to the supnorm). It is well-known that the optimal rate for estimation in $\mathcal{M}_c^{\mathrm{TV}}$ is the same as that in the corresponding Sobolev class. Estimation can be based on thresholding empirical wavelet coefficients (see Donoho and Johnstone, 1992). In a regression model Mammen and van de Geer (1993) discuss regression splines with data-adaptive placed knot points, which achieve in $\mathcal{M}_c^{\mathrm{TV}}$ the optimal rate. Kernel estimates $\hat{m}_h$ with global bandwidth do not achieve optimal rates in this class. The main finding in this paper is that the same holds for kernel estimates with simple plug-in variable bandwidths, i.e. they can not adapt fully to spatial inhomogenity.

In this paper we discuss how kernel estimates with locally data-adaptive plug-in bandwidth behave over the classes $\mathcal{M}_c^{\mathrm{Sob}}$, $\mathcal{M}_c^{\mathrm{TV}}$, $\mathcal{M}_c^{\mathrm{conv}}$ and $\mathcal{M}_c^{\mathrm{mon}}$. The paper is organized as follows. In the next section we specify the white noise model which is considered. Furthermore, we establish helpful bounds for estimating squared risks of kernel estimates with random bandwidth. In Section 3 we investigate what happens in case of pilot estimation with oversmoothing, whereas the findings on pilot estimation without oversmoothing are established in Section 4. The proofs of the presented results are deferred until the last section. It turns out that problem (ii), the extrapolation, is a serious one. Indeed, in case of oversmoothing the bandwidth $\hat{h}(x)$ may be so large that the estimation procedure breaks down even in Sobolev classes: the optimal rate of convergence is not achieved. Therefore a protection against extrapolation is needed. Truncation of the bandwidth $\hat{h}(x)$ in an appropriate way deals with the problem of extrapolation. With this truncated version intrapolation is also no problem. However, also after truncation, for classes of convex functions oversmoothing of the pilot estimation is fatal and results in estimators which do not achieve the optimal rate of convergence. Only without oversmoothing in the pilot estimation things work out fine for these classes of functions. Similar remarks can be made for classes of monotone functions. For classes of functions with derivatives of bounded variation the estimation procedures (with and without oversmoothing in the pilot estimation) break always down. This is explained by the occurence of problem (iii). Intuitively, this can be seen as follows: suppose that the regression function shows two consecutive breaks with opposite direction. This possibly is smoothed away in the pilot estimation if the distance between the breaks is small in comparison with the pilot bandwidth g. Two neighbouring breaks with opposite direction correspond to an important shape characteristic of the curve. Roughly they may appear as a jump. This may be a feature of big importance in practical data analysis. For classes of convex functions such a consecutive occurence of breaks is excluded and this explains why problem (iii) is not appearing here.

2 Preliminaries

Let us consider the white noise model

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

where $W(t)$ $(t \ge 0)$ is a Brownian motion and m is an unknown (regression) function. We make asymptotic considerations for $\sigma \to 0$. Note that model (2.1) corresponds to a regression model with i.i.d. errors and sample size $n = \sigma^{-2}$. In fact Brown and Low (1990), Low (1992) and Nussbaum (1993) showed that the white noise model, the density estimation model and the nonparametric regression model with Gaussian errors are asymptotically equivalent.

We will study the asymptotic behaviour of kernel estimates $\hat{m}_h$ in model (2.1):

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

where $K_h(x) = h^{-1}K(x/h)$ is the kernel and h denotes the bandwidth of the kernel estimate $\hat{m}_h$. The local optimal choice of h is $h(x) = \kappa\sigma^{2/5}|m''(x)|^{-2/5}$ where κ is a constant depending only on the kernel K. We consider plug-in bandwidth selectors $\bar{h}(x)$ which depend only on $\tilde{m}_g''(x)$ (and deterministic constants), where $\tilde{m}_g(x)$ is a kernel estimate with kernel $L_g(x) = g^{-1}L(x/g)$ and bandwidth g. For the kernels K and L we suppose

> **(K1).** K and L are two times continuously differentiable, positive and symmetric with compact support ($[-1,1]$, say).

Put $m_{L,g}(x) = E\tilde{m}_g(x) = \int L_g(x-y)m(y)dy$ and $m_{K,h}(x) = E\hat{m}_h(x)$. Then $m''_{L,g}(x) = \int L_g''(x-y)m(y)dy$, which is also defined when m'' does not exist. For simplicity we write also $m_h(x) = m_{K,h}(x)$.

For calculating the rate of convergence of $\hat{m}_h(x)$ with $h = \bar{h}(x)$ the following result will be helpful.

Proposition 1. *Suppose* (K1). *For a kernel estimate $\hat{m}_h$ with bandwidth $h = \bar{h}(x)$ depending only on $\tilde{m}_g''(x)$ the following bounds for the quadratic risk apply*

$$B_\ell(x) \le E(\hat{m}_{\bar{h}(x)}(x) - m(x))^2 \le 2B_u(x) + 2S(x), \tag{2.2}$$

where

$$\begin{aligned}
B_\ell(x) &= E\left(m_{\bar{h}(x)}(x) - m(x) + (\tilde{m}_g''(x) - m''_{L,g}(x)) \cdot \frac{\int K_{\bar{h}(x)}L_g''}{\int (L_g'')^2}\right)^2, \\
B_u(x) &= E(m_{\bar{h}(x)}(x) - m(x))^2, \\
S(x) &= \sigma^2[\int (L'')^2]^{-1}[\max|L''|]^2 g^{-1} + \sigma^2[\int K^2]E\bar{h}^{-1}(x).
\end{aligned}$$

For functions m with existing second derivative $B_u(x)$ can be calculated by

$$B_u(x) = E\bar{h}(x)^4[\int M_{\bar{h}(x)}(x-y)m''(y)dy]^2, \tag{2.3}$$

where M is the kernel defined by

$$M(u) = \begin{cases} -\int_u^1 (u-\nu)K(\nu)d\nu & u \ge 0, \\ \int_{-1}^u (u-\nu)K(\nu)d\nu & u \le 0. \end{cases}$$

After partial integration the right-hand side of (2.3) is also defined when m is not two times differentiable.

3 Pilot estimation with oversmoothing

We now consider plug-in estimates $\tilde{h}(x)$ of the asymptotically optimal bandwidth $h(x) = \kappa\sigma^{2/5}|m''(x)|^{-2/5}$ with a constant κ depending on the kernel K. The local plug-in bandwidth may be chosen as $\tilde{h}(x) = \kappa\sigma^{2/5}|\tilde{m}''_g(x)|^{-2/5}$. We consider choices of the pilot bandwidth $g = \gamma\sigma^{+2/5-\alpha}$ with $\alpha \geq 0$. We will make the distinction between pilot estimation with oversmoothing, i.e. $\alpha > 0$, and pilot estimation without oversmoothing, i.e. the case $\alpha = 0$. In this section we concentrate on the case of oversmoothing in the pilot estimation step.

The estimator $\tilde{m}''_g(x)$ has a normal distribution with expectation $m''_{L,g}(x)$ and variance of order $\sigma^{5\alpha}$. Therefore the bandwidth $\tilde{h}(x)$ is of stochastic order $\sigma^{2/5}|m''_{L,g}(x)|^{-2/5} \wedge g$. The operation characteristics of a kernel estimate $\hat{m}_h(x)$ depend on the smoothness of the curve m in a neighbourhood of x which has length of order h. When using the bandwidth $h = \tilde{h}(x)$ this choice of h is based on estimating the local amount of smoothness in a neighbourhood of length of order g. When g and $\tilde{h}(x)$ are of different order then $\tilde{h}(x)$ may be an unsuitable choice of the bandwidth h. Let us discuss the cases $\tilde{h}(x) \ll g$ and $\tilde{h}(x) \gg g$. In the introduction we referred to these cases as intrapolation, respectively extrapolation. In the following theorem we show that the estimate $\hat{m}_h$ with the choice $h = \tilde{h}(x)$ breaks down even in Sobolev classes when we allow that the chosen bandwidth becomes too large (the case of extrapolation). It should be remarked that in these classes kernel estimates with data-adaptive chosen global bandwidth work quite well. This is a striking fact because a local bandwidth choice is thought of as an improvement over global choices. The second statement of the following theorem says that after truncation also the local bandwidth choice $\tilde{h}(x)$ works in Sobolev classes. However, the truncation constant must be of order $\sigma^{2/5}$, i.e. the order of the optimal global bandwidth. That using a truncated version of $\tilde{h}(x)$ works shows also that intrapolation is no problem. One has to take care only of extrapolation.

For the formulation of the theorem we define for $\beta \geq 0$, $\delta > 0$, the truncated bandwidth

$$\hat{h}_\beta(x) = \tilde{h}(x) \wedge \delta\sigma^{2/5-\beta}. \tag{3.1}$$

The region of integration $I_{(\sigma)}$ appearing in the theorem equals $I_{(\sigma)} = \{x \in [0,1] : \delta\sigma^{2/5-\beta} \leq x \leq 1 - \delta\sigma^{2/5-\beta}\}$. This integration region $I_{(\sigma)}$ has been used to avoid boundary effects.

Theorem 1 . *Suppose that the kernels K and L fulfill (K1) and that the bandwidth $\hat{h}_\beta(x)$ is chosen according to (3.1). Define by $\hat{m}^\beta(x)$ the kernel estimate with kernel K and bandwidth $\hat{h}_\beta(x)$. For the maximal mean integrated square error over Sobolev classes we get*

$$\sigma^{-8/5} \sup_{m \in \mathcal{M}_c^{\mathrm{Sob}}} E \int_{I_{(\sigma)}} (\hat{m}^\beta(x) - m(x))^2 dx \to +\infty, \tag{3.2}$$

when $\beta > 0$ *and*

$$\limsup_{\sigma\to 0} \sigma^{-8/5} \sup_{m\in\mathcal{M}_c^{\text{Sob}}} E \int_{I_{(\sigma)}} (\hat{m}^\beta(x) - m(x))^2 dx < +\infty, \tag{3.3}$$

when $\beta = 0$.

Also after truncation the estimator $\hat{m}^\beta(x)$ breaks down for $\beta = 0$ in the class $\mathcal{M}_c^{\text{conv}}$, as is shown in the next theorem. Note that this class is a subclass of the class of functions $\mathcal{M}_c^{\text{TV}}$. In this subclass neighbouring breaks with opposite directions are excluded. Hence, this breakdown must be explained again by an extrapolation implicitly done between the pilot estimation step and final estimation step (see point (ii) in the introduction).

Theorem 2 . *Suppose* (K1) *and*

$$\textbf{(K2)}: \quad \min(L(x) : |x| \le \delta) > 0.$$

Denote by $\mathcal{M}_c^{\text{conv}}$ *the class of convex functions in* $\mathcal{M}_c^{\text{TV}}$. *Let* $\hat{m}$ *denote the kernel estimate with bandwidth* $\hat{h}(x) = \kappa\sigma^{2/5}|\tilde{m}_g''(x)|^{-2/5} \wedge \delta\sigma^{2/5}$. *The pilot bandwidth is* $g = \gamma\sigma^{2/5-\alpha}$. *Then,*

$$\sigma^{-8/5} \sup_{m\in\mathcal{M}_c^{\text{conv}}} E \int_{I_{(\sigma)}} (\hat{m}(x) - m(x))^2 dx \to \infty.$$

So even for this subclass of the class $\mathcal{M}_c^{\text{TV}}$ the estimator breaks down in case of pilot estimation with oversmoothing.

4 Pilot estimation without oversmoothing

In this section we investigate the case of pilot estimation without oversmoothing, i.e. $\alpha = 0$. Hence, the pilot bandwidth is $g = \gamma\sigma^{2/5}$. We put $\beta = 0$, skip the superscript β and define again $\hat{m}(x)$ as the kernel estimate with bandwidth $\hat{h}(x) = \kappa\sigma^{2/5}|\tilde{m}_g''(x)|^{-2/5} \wedge \delta\sigma^{2/5}$.

As in Section 3 one can easily see that $\hat{m}(x)$ works over the Sobolev classes $\mathcal{M}_c^{\text{Sob}}$ in case of pilot estimation without oversmoothing.

Theorem 3 . *Suppose that the kernels K and L fulfill (K1). For the maximal mean integrated square error over Sobolev classes we get*

$$\limsup_{\sigma\to 0} \sigma^{-8/5} \sup_{m\in\mathcal{M}_c^{\text{Sob}}} E \int_{I_{(\sigma)}} (\hat{m}(x) - m(x))^2 dx < +\infty.$$

The kernel estimate with local data-adaptive bandwidth breaks down in $\mathcal{M}_c^{\text{TV}}$, as is established in Theorem 4, but achieves optimal rates of convergence over the smaller class $\mathcal{M}_c^{\text{conv}}$ (see Theorem 5). Clearly the major reason for this breakdown is problem (iii) mentioned in the introduction.

Theorem 4 . *Suppose* (K1). *Then for the mean integrated square loss we have*

$$\sigma^{-8/5} \sup_{m \in \mathcal{M}_c^{\mathrm{TV}}} E \int_{I_{(\sigma)}} (\hat{m}(x) - m(x))^2 dx \to +\infty.$$

We exclude now the case of neighbouring breaks with opposite direction, and consider the smaller class of convex functions. We show that $\hat{m}(x)$ achieves the optimal rate when the pilot bandwidth g is of order $\sigma^{2/5}$. For this choice $\tilde{m}_g''$ is not consistent (This is a normal variable with variance $\sigma^2 g^{-5} \int (L'')^2$.). The asymptotic distribution of $\hat{m}(x)$ is affected by the random nature of its bandwidth: it is mixed normal instead of normal.

Theorem 5 . *If* (K1) *and* (K2) *hold, then*

$$\limsup_{\sigma \to 0} \sigma^{-8/5} \sup_{m \in \mathcal{M}_c^{\mathrm{conv}}} E \int_{I_{(\sigma)}} (\hat{m}(x) - m(x))^2 < +\infty.$$

Given $\tilde{m}_g''(x)$ the estimate $\hat{m}$ has a normal distribution with mean

$$(\tilde{m}_g''(x) - m_{L,g}''(x))[\int K_{\hat{h}(x)} L_g''][\int (L_g'')^2]^{-1}$$

and variance $\sigma^2 \{\int K_{\hat{h}(x)}^2 - [\int K_{\hat{h}(x)} L_g''][\int (L_g'')^2]^{-1}$.

Up to now we have only considered smoothness classes of degree two. Analogous considerations could be applied to other function classes. Let us mention here only a result, similar to Theorem 5, for monotone functions.

Theorem 6 . *Suppose* (K1). *Denote by $\mathcal{M}_c^{\mathrm{mon}}$ the class of monotone functions m on $[0,1]$ with $|m(1) - m(0)| \leq c$ (i.e. with total variation bounded by c). The curve estimate $\hat{m}$ is now a kernel estimate with bandwidth $\hat{h}(x) = \kappa \sigma^{2/3} |\tilde{m}_g'(x)|^{-1/3} \wedge \delta \sigma^{2/3}$ and pilot bandwidth $g = \gamma \sigma^{2/3}$. Then we get*

$$\limsup_{\sigma \to 0} \sigma^{-4/3} \sup_{m \in \mathcal{M}_c^{\mathrm{mon}}} E \int_{I_{(\sigma)}} (\hat{m}(x) - m(x))^2 dx < +\infty.$$

Given $\tilde{m}_g'(x)$ the estimate $\hat{m}$ has a normal distribution with mean

$$(\tilde{m}_g'(x) - m_{L,g}'(x))[\int K_{\hat{h}(x)} L_g'][\int (L_g')^2]^{-1}$$

and variance

$$\sigma^2 [\int K_{\hat{h}(x)}^2 - \int K_{\hat{h}(x)} L_g'][\int (L_g')^2]^{-2}.$$

The kernel estimate $\hat{m}$ can be compared with the least squares monotone estimate $\hat{m}_{LS}$, defined as an minimizer of $\int (dY(t) - m(t)dt)^2/dt$. Like $\hat{m}$, the estimate $\hat{m}_{LS}$ achieves the optimal rate $\sigma^{4/3}$ in $\mathcal{M}_c^{\mathrm{mon}}$. However, it has a much more complicated asymptotic distribution (see Groeneboom, 1985, 1989). Also $\hat{m}_{LS}$ adapts to locally varying smoothness of the curve (see Birgé, 1987 and Mammen, 1991).

5 Proofs

Proof of Proposition 1. For every (deterministic) choice of h and g the variable $(\hat{m}_h(x)-m_h(x), \tilde{m}''_g(x)-m''_{L,G}(x))$ has a normal distribution with mean $(0,0)$ and variance-covariance matrix $\begin{pmatrix} s_{11}(h) & s_{12}(h) \\ s_{21}(h) & s_{22} \end{pmatrix}$, where $s_{11}(h) = \sigma^2 \int (K_h)^2$, $s_{22} = \sigma^2 \int (L''_g)^2$ and $s_{12}(h) = s_{21}(h) = \sigma^2 \int K_h L''_g$. Conditionally, given $\tilde{m}''_g(x) - m''_{L,g}(x) = \nu$, the variable $\hat{m}_h(x) - m_h(x)$ has a normal distribution with mean $\nu s_{12}(h)/s_{22}$ and variance $s_{11}(h) - s_{12}^2(h)/s_{22}$. Conditionally, given $\tilde{m}''_g(x) - m_{L'',g}(x) = \nu$, the bandwidth $\bar{h}(x)$ is fixed. Therefore, conditionally the variable $\hat{m}_{\bar{h}(x)}(x) - m_{\bar{h}(x)}(x)$ has a normal distribution with mean $\nu\ s_{12}(\bar{h}(x))/s_{22}$ and variance $s_{11}(\bar{h}(x)) - s_{12}^2(\bar{h}(x))/s_{22}$. Let us write E^*_ν for the conditional expectation, given $\tilde{m}''_g(x)-m''_{L,g}(x) = \nu$. With an $N(0, s_{22})$ distributed random variable V we get

$$E(\hat{m}_{\bar{h}(x)}(x) - m(x))^2 \geq EE^*_V \left(\hat{m}_{\bar{h}(x)}(x) - E^*_V \hat{m}_{h(x)}(x)\right)^2 = B_\ell(x)$$

and

$$\begin{aligned} E(\hat{m}_{\bar{h}(x)}(x) - m(x))^2 &\leq 2E(\hat{m}_{\bar{h}(x)}(x) - m_{\bar{h}(x)}(x))^2 \\ &\quad +2E(m_{\bar{h}(x)}(x) - m(x))^2 \\ &= 2EE^*_V(\hat{m}_{\bar{h}(x)}(x) - m_{\bar{h}(x)}(x))^2 + 2B_u(x) \\ &= 2EV^2 s_{12}^2(\bar{h}(x))s_{22}^{-2} + s_{11}(\bar{h}(x)) - s_{12}^2(\bar{h}(x))s_{22}^{-1} + 2B_u(x) \\ &\leq 2S(x) + 2B_u(x) \end{aligned}$$

because of

$$\sigma^{-2} s_{12}(\bar{h}(x)) = \int K_{\bar{h}(x)}(x) L''_g(x) dx \leq \max_x |L''_g(x)| = g^{-3} \max_x |L''(x)|.$$

Equation (2.3) follows from

$$m_h(x) - m(x) = h^2 \int M_h(x-y) m''(y) dy.$$

□

Proof of Theorem 1. We start with the proof of (3.2). For $\varepsilon < \frac{4}{9}\alpha \wedge \beta$ we choose a function m^σ in $\mathcal{M}_c^{\text{Sob}}$ with

$$\begin{aligned} &\frac{d^2}{dx^2} m^\sigma(x) = 1 && \text{if } x \text{ lies for an } i \text{ in the interval } [2i\sigma^{2/5-\varepsilon}, (2i+1)\sigma^{2/5-\varepsilon}], \\ &\frac{d^2}{dx^2} m^\sigma(x) = -1 && \text{elsewhere,} \end{aligned} \tag{5.1}$$

$$\frac{d}{dx} m^\sigma(0) = -\frac{1}{2}\sigma^{2/5-\varepsilon}, \tag{5.2}$$

$$m^\sigma(0) = 0. \tag{5.3}$$

The function m^σ is uniquely defined by (5.1) — (5.3). Simple calculus shows that $\sup_x |m''_{L,g}(x)| = O(\sigma^{2(\alpha-\varepsilon)})$. Because of $\sup_x |\tilde{m}''_g(x) - m''_{L,g}(x)| = o_P(\sigma^{2\alpha})$, this implies $\sup_x |\tilde{m}''_g(x)| = O_P(\sigma^{2(\alpha-\varepsilon)})$. The bandwidth $\hat{h}_\beta(x)$ is defined as $\hat{h}_\beta(x) = \kappa\sigma^{2/5}|\tilde{m}''_g(x)|^{-2/5} \wedge \delta\sigma^{2/5-\beta}$. We arrive at

$$\sup_x \sigma^{2/5-\varepsilon}\hat{h}_\beta(x)^{-1} = O_P(\sigma^{\frac{4}{5}\alpha-\frac{9}{5}\varepsilon} \vee \sigma^{\beta-\varepsilon}). \tag{5.4}$$

For the function $m = m^\sigma$ there exists a constant C_1 such that for $h \geq \sigma^{2/5-\varepsilon}$ it holds that

$$\sup_x |m_h(x)| \leq C_1 \left(\frac{\sigma^{2/5-\varepsilon}}{h}\right)^2 (\sigma^{2/5-\varepsilon})^2. \tag{5.5}$$

Expressions (5.4) and (5.5)imply that

$$\sup_x |m_{\hat{h}_\beta(x)}| = O_P(\sigma^{4/5-2\varepsilon}). \tag{5.6}$$

On the other hand on a subset of $[0,1]$ with Lebesgue measure $\geq 1/2$ it holds that $|m^\sigma(x)| \geq \frac{3}{32}\sigma^{4/5-2\varepsilon}$. We now want to apply the lower bound for squared risks established in Proposition 1. For this purpose it remains to bound

$$\begin{aligned}
&\sup_x |\tilde{m}''_g(x) - m''_{L,g}(x)| \cdot \frac{\int K_{\hat{h}_\beta(x)} L''_g}{\int (L''_g)^2} \\
&= \sup_x |\tilde{m}''_g(x) - m''_{L,g}(x)| \cdot \frac{\sup_x |L''_g(x)]}{\int (L''_g)^2} \\
&= \sup_x |\tilde{m}''_g(x) - m''_{L,g}(x)| \cdot O(g^2) \\
&= o_P(\sigma^{4/5}).
\end{aligned}$$

This shows that for x in a subset of $[0,1]$ with Lebesgue measure $\geq \frac{1}{2}$ we get with a constant C_2 for σ small enough that

$$B_\ell(x) \geq C_2\sigma^{+8/5-4\varepsilon}.$$

This proves (3.2).

We next prove (3.3). Because of (2.2) and (2.3) it suffices to show that

$$\int E\hat{h}_\beta(x)^{-1} = O(\sigma^{2/5}) \tag{5.7}$$

$$\int E\hat{h}_\beta(x)^4 \left[\int M_{\hat{h}_\beta(x)}(x-y)m''(y)dy\right]^2 = O(\sigma^{8/5}). \tag{5.8}$$

Statement (5.7) follows from

$$\hat{h}_\beta(x)^{-1} = \kappa^{-1}\sigma^{-2/5}|\tilde{m}_g''(x)|^{2/5} \vee \delta^{-1}\sigma^{-2/5}$$

and

$$E\int |\tilde{m}_g''(x)|^{2/5} = O(1).$$

Statement (5.8) follows from

$$\begin{aligned}
&\int E\hat{h}_\beta(x)^4 \left[\int M_{\hat{h}_\beta(x)}(x-y)m''(y)dy\right]^2 \\
&= O(\sigma^{8/5})\int \left[\sigma^{-2/5}\int |m''(y)|\,\mathbf{1}[|x-y|\le \delta\sigma^{2/5}]dy\right] 2dx \\
&= O(\sigma^{8/5})\sigma^{-2/5}\int\int |m''(y)|\,|m''(z)|\,\mathbf{1}[|y-z|\le 2\delta\sigma^{2/5}]dydz \\
&= O(\sigma^{8/5}).
\end{aligned}$$

□

Proof of Theorem 2. For simplicity suppose $\gamma = \delta = 1$. Then $g = \sigma^{2/5-\alpha}$. Define $m^\sigma(x)$ as the piecewise linear interpolation of $m^0(x) = (c/2)x^2$ with interpolation points $x_i = i\cdot\sigma^{2/5-\alpha/2}$ $(i = 0,1,\ldots)$. We write also m instead of m^σ. The functions m^σ are in $\mathcal{M}_c^{\mathrm{TV}}$. The estimate $\tilde{m}_g''(x)$ is a Gaussian variable with expectation $m_{L,g}''(x) = \int L_g''(x-u)m^\sigma(x)dx$. We have $m_{L,g}''(x) = c + O(\sigma^\alpha)$ (uniformly in x). This follows from $|m^\sigma(x) - m^0(x)| = O(\sigma^{4/5-\alpha})$. The variance of $\tilde{m}_g''(x)$ is of order $O(\sigma^2 g^{-5}) = O(\sigma^{5\alpha})$. Therefore for every sequence d_σ' with $d_\sigma' \to 0$ (slowly enough for $\sigma \to 0$) it holds that

$$\sup_{x\in I(\sigma)} |P\left(c(1-d_\sigma') \le \tilde{m}_g''(x) \le c(1+d_\sigma')\right) - 1| \to 0 \qquad (\text{for } \sigma \to 0).$$

Put $\Delta_\sigma(x) = \mathbf{1}[(1-d_\sigma)h_0 \le \hat{h}(x) \le (1+d_\sigma)h_0]$ and $h_0 = \kappa\sigma^{2/5}c^{-2/5}$. For $d_\sigma \to 0$ (slowly enough) we get

$$\sup_{x\in I(\sigma)} |P(\Delta_\sigma(x) = 1) - 1) \to 0 \qquad (\text{for } \sigma \to 0).$$

Using (2.2) we get the following lower bound for the quadratic risk

$$\begin{aligned}
E(\hat{m}(x) - m(x))^2 &\ge B_\ell(x) \\
&\ge E[m_{\hat{h}(x)}(x) - m(x) + R_\sigma(x)(\tilde{m}_g''(x) - m_{L,g}''(x))]^2\Delta_\sigma(x).
\end{aligned}$$

Here $R_\sigma(x)$ is the factor $\int(K_{\hat{h}(x)}L_g'')[\int(L_g'')^2]^{-1}$. We apply now the inequality $E[X+Y]^2 \ge [[EX^2]^{1/2} - [EY^2]^{1/2}]^2$. This results in

$$[E(\hat{m}(x)-m(x))^2]^{1/2} \ge [E(m_{\hat{h}(x)} - m(x))^2\Delta_\sigma(x)]^{1/2} - [ER_\sigma^2(x)(\tilde{m}_g''(x) - m_{L,g}''(x))^2]^{1/2}.$$

The second term does not depend on m. The factor $|R_\sigma(x)|$ can be bounded by

$$\sup_{-g\le x\le g} |L_g''(x)|[\int (L_g'')^2]^{-1} = O(g^2).$$

This implies

$$\sup_{x\in I(\sigma)} ER_\sigma^2(x)(\tilde m_g''(x) - m_{L,g}''(x))^2 = O(\sigma^2 g^{-1}) = o(\sigma^{8/5}).$$

For the mean integrated square error we get the following lower bound:

$$\begin{aligned}\int\limits_{I(\sigma)} E(\hat m(x) - m(x))^2 dx &\ge \int\limits_{I(\sigma)} \left\{[E(m_{\hat h(x)}(x) - m(x))^2\Delta_\sigma(x)]^{1/2} - o(\sigma^{8/5})\right\}^2 dx \\ &\ge \int\limits_{I(\sigma)} \left\{\inf_{(1-d_\sigma)h_0\le h\le(1+d_\sigma)h_0} |m_h(x) - m(x)| - o(\sigma^{8/5})\right\}^2 dx.\end{aligned}$$

Consider $I_{\varepsilon,\sigma} = \{x \in I(\sigma) : |x - x_i| \le \varepsilon\sigma^{2/5}$ for an $i \ge 0\}$. For ε small enough it holds that

$$\liminf_{\sigma\to 0} \sigma^{-4/5+\alpha/2} \inf_{x\in I_{\varepsilon,\sigma}} |m_h(x) - m(x)| > 0.$$

Because the Lebesgue measure of $I_{\varepsilon,\sigma}$ is of order $\sigma^{2/5}\sigma^{-2/5+\alpha/2} = \sigma^{\alpha/2}$ this gives

$$\liminf_{\sigma\to 0} \sigma^{-8/5+\alpha/2} \int\limits_{I(\sigma)} E(\hat m(x) - m(x))^2 dx \ge \liminf_{\sigma\to 0} \int\limits_{I_{\varepsilon,\sigma}} E(\hat m(x) - m(x))^2 dx > 0.$$

□

Proof of Theorem 3. Theorem 3 follows as the proof of (3.3) in Theorem 1.

Proof of Theorem 4. Suppose $g = \sigma^{2/5}$ (w.l.o.g.). For $\varepsilon < 2/15$ we write $s = \sigma^{2/5+\varepsilon}$ and we define m in $\mathcal{M}_c^{\mathrm{TV}}$ such that $m'(\frac{1}{2}) = m(\frac{1}{2}) = 0$ and

$$m''(x) = \frac{c}{2s}\left[\mathbf{1}[\frac{1}{2} < x < \frac{1}{2} + s] - \mathbf{1}[\frac{1}{2} - s < x < \frac{1}{2}]\right] \quad \text{(a.s.)}$$

The expectation of $\tilde m_g''(x)$ is now

$$m_{L,g}''(x) = \int L(v)m''(x - gv)dv.$$

For x with $|x - \frac{1}{2}| \le g$ this can be written as

$$\frac{c}{2s}\Big[2F_L\Big(\frac{x-\frac{1}{2}}{g}\Big) - F_L\Big(\frac{x-\frac{1}{2}-s}{g}\Big) - F_L\Big(\frac{x-\frac{1}{2}+s}{g}\Big)\Big],$$

where $F_L(u) = \int\limits_{-1}^{u} L(v)dv$. Expansion of F_L gives for x with $|x - \frac{1}{2}| \le s$:

$$m''_{L,g}(x) = -\frac{c}{8s}\, L''(0)\, \frac{x - \frac{1}{2}}{g}\, (\frac{s}{g})^2 + o(\frac{s^2}{g^3}).$$

For $s/2 \le |x - \frac{1}{2}| \le s$ this is of order $\sigma^{2\varepsilon - 2/5}$. As above, the standard deviation of $\tilde{m}''_g(x)$ is of lower order. Arguing as above we get for sequences d_σ with $d_\sigma \to 0$ slowly enough (for $\sigma \to 0$):

$$\sup_{s/2 \le |x - \frac{1}{2}| \le s} |P(\Delta_\sigma(x) = 1) - 1| \to 0 \quad (\text{for } \sigma \to 0),$$

where now

$$\Delta_\sigma(x) = \mathbf{1}[(1 - d_\sigma)h_0(x) \le \hat{h}(x) \le (1 + d_\sigma)h_0(x)]$$

and

$$h_0(x) = \kappa |\frac{c}{8} L''(0)|^{-2/5} \sigma^{2/5 - 6\varepsilon/5} |x - \frac{1}{2}|^{-2/5}.$$

The quantity $h_0(x)$ is of order $\sigma^{6/25 - 8\varepsilon/5}$ (for $s/2 \le |x - \frac{1}{2}| \le s$). Again, we use the bound

$$\begin{aligned}[E(\hat{m}(x) - m(x))^2]^{1/2} &\ge [E(m_{\hat{h}(x)}(x) - m(x))^2 \Delta_\sigma(x)]^{1/2} \\ &\quad - [E R_\sigma^2(x)(\tilde{m}''_g(x) - m''_{L,g}(x))^2]^{1/2}.\end{aligned}$$

The second term can be bounded by $O(\sigma^{8/5})$ (uniformly in x with $s/2 \le |x - 1/2| \le s$). Calculation and expansion of $m_h(x)$ gives

$$m_h(x) = \frac{c}{2}\, (x - \frac{1}{2}) + o(s).$$

This expansion holds uniformly for $s/2 \le |x - \frac{1}{2}| \le s$ and $c_0 \sigma^{6/25 - 8\varepsilon/5} \le h \le c_1 \sigma^{6/25 - 8\varepsilon/5}$ (with fixed $c_0, c_1 > 0$). On the other hand we get for $|x - \frac{1}{2}| \le s$ that $m(x) = \frac{c}{4s}|x - \frac{1}{2}|(x - \frac{1}{2})$.
This gives $|m_h(x) - m(x)| \ge c/4 |x - \frac{1}{2}| + o(s) \ge cs/8 + o(s)$. We get

$$\begin{aligned}&\int E(m_{\hat{h}(x)}(x) - m(x))^2 \Delta_\sigma(x) dx \\ &\quad \ge \int 1(s/2 \le |x - \frac{1}{2}| \le s) E(m_{\hat{h}(x)}(x) - m(x))^2 \Delta_\sigma(x) dx \ge s(cs/8)^2(1 + 0(1)).\end{aligned}$$

This is of order $\sigma^{6/5 + 3\varepsilon}$. In particular, because of $\varepsilon < 2/15$ we have $[\sigma^{6/5+3\varepsilon}][\sigma^{8/5}]^{-1} \to \infty$. $\square$

Proof of Theorem 5. W.l.o.g. we suppose that m is two times differentiable. We apply Proposition 1. First note that

$$\begin{aligned}\int S(x)dx &= O(\sigma^2 g^{-1}) + O(\sigma^2) \int E\hat{h}^{-1}(x) \\ &= O(\sigma^{8/5}) + O(\sigma^{8/5}) \int E(\tilde{m}''_g(x))^{2/5} dx.\end{aligned}$$

Now

$$\int E|\tilde{m}_g''(x)|^{2/5} \le 1+\int E|\tilde{m}_g''(x)|dx = O(1)+\int |m_{L,g}''(x)|dx \le O(1)+\int |m''(x)|dx = O(1).$$

It remains to bound $\int B_u(x)dx$. We use formula (2.3) and get for $c_1 > c_0 > 0$

$$\text{(5.9)}\quad \int B_u(x)dx = \int E\hat{h}(x)^4[\int M_{\hat{h}(x)}(x-y)m''(y)dy]^2dx$$
$$= \int E\left[\ \dots \mathbf{1}[\hat{h}(x) < c_0 g]\right] dx + \int E\left[\ \dots \mathbf{1}[c_0 g \le \hat{h}(x) \le c_1 g]\right] dx.$$

First we bound the second term. Because of (K2) we get for h with $c_0 g \le h \le c_1 g$:

$$\text{(5.10)}\qquad \int M_h(x-y)m''(y)dy \le C\int L_g(x-y)m''(y)dy = Cm_{L,g}''(x)$$

where C depends on c_0, but not on m. Now $\hat{h}(x) \ge c_0 g$ implies $\tilde{m}_g''(x) \le C'c_0^{-5/2}$ (with a constant C'). $\tilde{m}_g''(x)$ is normal with expectation $m_{L,g}''(x)$. Therefore

$$\text{(5.11)}\qquad P(\hat{h}(x) \ge c_0 g) \le \Phi(C'c_0^{-5/2} - m_{L,g}''(x)).$$

With (5.10) and (5.11) one gets for the second term of the right-hand side of (5.9):

$$\int E\left[\ \dots \mathbf{1}[c_0 g \le \hat{h}(x) \le c_1 g]\right] dx = O(1)\cdot g^4 \int m_{L,g}''(x)P(\hat{h}(x) \ge c_0 g)dx$$
$$= O(g^4)\int m_{L,g}''(x)\Phi(C'c_0^{-5/2} - m_{L,g}''(x))dx = O(\sigma^{8/5}).$$

We bound now the first term of the right-hand side of (5.9). For $h \ge c_1 g$ one gets (compare (5.10)) with a constant C'':

$$\text{(5.12)}\qquad \int M_h(x-y)m''(y)dy \le C''gh^{-1}m_{L,g}''(x).$$

The inequality $\hat{h}(x) < c_0 g$ holds if and only if $\tilde{m}_g''(x) > \kappa^{5/2}c_0^{-5/2}g^{-5/2}\sigma$. Now we get with (5.11) and (5.12)

$$\int E\ \hat{h}(x)^4[\int M_{\hat{h}(x)}(x-y)\ m''(y)dy]^2dx = \int E\hat{h}(x)^4\ m_{M,\hat{h}(x)}''(x)^2dx$$
$$\le C''^{3/2}E\ \mathbf{1}[\hat{h}(x) < c_0 g]\ \hat{h}(x)^4(g/\hat{h}(x))^{3/2}(m_{L,g}''(x))^{3/2}(m_{M,\hat{h}(x)}''(x))^{1/2}dx + O(\sigma^{8/5})$$
$$\le C''^{3/2}\int E\ \mathbf{1}[\hat{h}(x) < c_0 g]\cdot g^{3/2}\cdot \frac{\kappa^{5/2}\sigma}{\tilde{m}_g''(x)}(m_{L,g}''(x))^{3/2}(m_{M,\hat{h}(x)}'')^{1/2}dx + O(\sigma^{8/5})$$
$$= O(\sigma^{8/5})[\int E\ \mathbf{1}[\hat{h}(x) < c_0 g]\ m_{L,g}''(x)^3\tilde{m}_g''(x)^{-2}dx]^{1/2}$$
$$[\int E\ \mathbf{1}[\hat{h}(x) < c_0 g]\ m_{M,\hat{h}(x)}''(x)dx]^{1/2} + O(\sigma^{8/5}).$$

It remains to show

$$\sup_{m\in\mathcal{M}_c^{\mathrm{conv}}} \int E\ \mathbf{1}[\hat{h}(x) < c_0 g]\ m''_{L,g}(x)^3\ \tilde{m}''_g(x)^{-2} dx < \infty, \tag{5.13}$$

$$\sup_{m\in\mathcal{M}_c^{\mathrm{conv}}} \int E\ \mathbf{1}[\hat{h}(x) < c_0 g]\ m''_{M,\hat{h}(x)}(x) dx < +\infty. \tag{5.14}$$

□

Proof of (5.13). We put $U(x) = \tilde{m}''_g(x) - m''_{L,g}(x)$. $U(x)$ has a normal distribution with mean 0 and variance $\sigma^2 \int (L''_g)^2$.
For $\hat{h}(x) < c_0 g$ we have $\tilde{m}''_g(x) > \kappa^{5/2} c_0^{-5/2} g^{-5/2} \sigma$. This implies

$$\begin{aligned} &\mathbf{1}[\hat{h}(x) < c_0 g]\ m''_{L,g}(x)\ \tilde{m}''_g(x)^{-1} \\ &\quad \le \mathbf{1}[\hat{h}(x) < c_0 g]\ (\tilde{m}''_g(x) - U(x))\ \tilde{m}''_g(x)^{-1} \\ &\quad \le \mathbf{1}[\hat{h}(x) < c_0 g]\ (1 + O(1)|U(x)|). \end{aligned} \tag{5.15}$$

Therefore we can bound the left-hand side of (5.13) by

$$O(1) \cdot \sup_{m\in\mathcal{M}_c^{\mathrm{conv}}} \int m''_{L,g}(x) dx.$$

This is bounded because of

$$\int m''_{L,g}(x) \le c.$$

□

Proof of (5.14). We estimate

$$T_1(x) = \int E\ \mathbf{1}[\hat{h}(x) < c_0 g]\ m''_{M,\hat{h}(x)}(x)\ \mathbf{1}[U(x) < -\frac{1}{2} m''_{L,g}(x)] dx$$

and

$$T_2(m) = \int E\ \mathbf{1}[\hat{h} < c_0 g]\ m''_{M,\hat{h}(x)}(x)\ \mathbf{1}[U(x) \ge -\frac{1}{2} m''_{L,g}(x)] dx.$$

Because of (5.12) we get with a constant $C''' > 0$

$$\begin{aligned} T_1(m) &\le C'' \int E \left[g \hat{h}(x)^{-1} m''_{L,g}(x)\ \mathbf{1}[U(x) < -\frac{1}{2} m''_{L,g}(x)] \right] dx \\ &= O(1) \cdot \int E \left[m''_{L,g}(x) + U(x))^{2/5} m''_{L,g}(x)\ \mathbf{1}[U(x) < -\frac{1}{2} m''_{L,g}(x)] \right] dx \\ &= O(1) \int \exp(-C''' m''_{L,g}(x)^2) dx = O(1). \end{aligned}$$

For the treatment of $T_2(m)$ note first that $U(x) \geq -m''_{L,g}(x)/2$ implies $\tilde{m}''_g(x) \geq m''_{L,g}(x)/2$. Then

$$h_g(x) = \kappa\sigma^{2/5}|m''_{L,g}(x)|^{-2/5} \geq \hat{h}(x)/2.$$

This gives with a constant $C^{(4)}$

$$\begin{aligned}
T_2(m) &= O(1)\int E\ \mathbf{1}[\hat{h}(x) < c_0 g]\ \mathbf{1}[U(x) \geq -\frac{1}{2}m''_{L,g}(x)]\ \hat{h}(x)^{-1}m''(y)\ \mathbf{1}[|x-y| \leq \hat{h}(x)]dxdy \\
&= O(1)\int E\ \mathbf{1}[\hat{h}(x) < c_0 g]\ \tilde{m}''_g(x)^{2/5}(m''_{L,g}(x) \vee C^{(4)})^{-2/5}(2h_g(x) \wedge g)^{-1} \\
&\qquad \times m''(y)\ \mathbf{1}[|x-y| \leq h_g(x) \wedge c_0 g]dxdy \\
&= O(1)\int (h_g(x) \wedge g)^{-1}m''(y)\ \mathbf{1}[|x-y| \leq 2h_g(x) \wedge c_0 g]dxdy.
\end{aligned}$$

The theorem follows from the following lemma. □

Lemma:

$$\sup_{m \in \mathcal{M}_c^{\text{conv}}} \int (h_g(x) \wedge g)^{-1}m''(y)\ \mathbf{1}[|x-y| \leq h_g(x) \wedge c_0 g]dxdy < +\infty$$

for c_0 small enough.

Proof. We consider

$$S_1(m) = \int (h_g(x) \wedge g)^{-1}m''(y)\ \mathbf{1}[|x-y| \leq c_0 g)1(h_g(x) \geq c_0 g]dxdy$$

and

$$S_2(m) = \int h_g(x)^{-1}m''(y)\ \mathbf{1}[|x-y| \leq h_g(x))1(h_g(x) < c_0 g]dxdy.$$

For $S_1(m)$ we get

$$S_1(m) \leq \int g^{-1}m''(y)\mathbf{1}[|x-y| \leq c_0 g] \leq c_0 c.$$

It remains to bound $S_2(m)$. First we can write

$$S_2(m) = \int \sigma^{-2/5}\kappa^{-1}m''_{L,g}(x)^{2/5}m''(y)\ \mathbf{1}[|x-y| \leq \sigma^{2/5}\kappa|m''_{L,g}(x)|^{-2/5} < c_0 g]dxdy.$$

Denote by N the uniform kernel $N(x) = \frac{1}{2}\mathbf{1}[|x| \leq 1]$. Then because of (K2) we get for $|x-y| < c_0 g$ with a constant C':

$$(1/C')m''_{N,(c_1-c_0)g}(y) \leq m''_{L,g}(x) \leq C'm''_{N,(1+c_0)g}(y).$$

This gives

$$S_2(m) = O(1)\int \frac{m''_{U,(1+c_0)g}(y)^{2/5}}{m''_{N,(c_1-c_0)g}(y)^{2/5}}m''(y)dy.$$

Choose now $c_3 = (c_1 - c_0)/2$ and put $y_i = ic_3 g$. We get for constants $D, D', D'', \ldots$

$$
\begin{aligned}
S_2(m) &\leq D \sum_i \int_{|y-y_i| \leq c_3 g/2} \frac{m''_{N,(1+c_0)g}(y)^{2/5}}{m''_{N,(c_1-c_0)g}(y)^{2/5}} m''(y) dy \\
&\leq D' \sum_i \int_{|y-y_i| \leq c_3 g/2} \frac{m''_{N,(1+c_0+c_3/2)g}(y_i)^{2/5}}{m''_{N,(c_1-c_0-c_3/2)g}(y_i)^{2/5}} m''(y) dy \\
&\leq D'' g \sum_i m''_{2N,(1+c_0+c_3/2)g}(y_i)^{2/5} m''_{N,c_3 g/2}(y_1)^{3/5} \\
&\leq D''' \int m''_{2N,(1+c_0+c_3/2)g}(y) dy \leq D''' \cdot c.
\end{aligned}
$$

□

References

Birgé, L. (1987). Estimating a density under order restrictions: Nonasymptotic minimax risk. *Ann. Statist.*, **15**, 995 – 1012.

Brockmann, M., Gasser, T. and Hermann, 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.

Donoho, D.L. and Johnstone, I.M. (1992). Minimax estimation via wavelet shrinkage. Unpublished manuscript.

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

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. (1993a). Adaptation to high spatial inhomogeneity based on wavelets and on local linear smoothing. *Research report SMS-59-93*, Australian National University.

Fan, J., Hall, P., Martin, M. and Patil, P. (1993b). On the local smoothing of nonparametric curve estimators. *Research report SMS-75-93*, Australian National University.

Groeneboom, P. (1985). Estimating a monotone density. *Proceedings of the Berkeley Conferences in Honor of Neyman, J. and Kiefer, J.* (ed. by Le Cam, L.M. and Olshen, R.A.) II 539 – 555.

Groeneboom, P. (1989). Brownian motion with a parabolic drift in Airy functions. *Probab. Th. Rel. Fields*, **81**, 79 – 109.

Hall, P. (1993). On plug-in rules for local smoothing of density estimators. *Ann. Statist.*, **21**, 694 – 710.

Low, M.G. (1992). Renormalization and white noise approximation for nonparametric functional estimation. *Ann. Statist.* , **20**, 545 – 554.

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). Asymptotic 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.