

# Asymptotically minimax estimation of a function with jumps

CATHARINA G.M. OUDSHOORN

*Institute for Business and Industrial Statistics, University of Amsterdam, Plantage Muidersgracht 24, 1018 TV Amsterdam, The Netherlands. e-mail: karino@fwi.uva.nl*

Asymptotically minimax nonparametric estimation of a regression function observed in white Gaussian noise over a bounded interval is considered, with respect to a  $L_2$ -loss function. The unknown function  $f$  is assumed to be  $m$  times differentiable except for an unknown although finite number of jumps, with piecewise  $m$ th derivative bounded in  $L_2$  norm. An estimator is constructed, attaining the same optimal risk bound, known as Pinsker's constant, as in the case of smooth functions (without jumps).

*Keywords:* jump-point estimation; nonparametric regression; optimal constant; tapered orthogonal series estimator

1350–7265 © 1998 Chapman & Hall

## 1. Introduction

In the 1980s, optimal rates of convergence in nonparametric regression estimation problems have been thoroughly examined, following the book of Ibragimov and Hasminskii (1981), the ground-breaking papers due to Stone (1982) and Birgé (1983) and others. Later the interest shifted to finding not only the optimal rates but also the asymptotic optimal constants, determining the risk of optimal estimators. This interest was greatly initiated by the pioneering paper of Pinsker (1980). Such results have been obtained for different observation schemes, involving smooth functions confined to balls in a Sobolev space, with  $L_2$  losses (see, for example, Nussbaum 1985; Golubev and Nussbaum 1990; Speckman 1985; and Efroimovich 1994), for functions restricted to Hölder balls in the case of  $L_\infty$  losses (Korostelev 1994; Donoho 1994) and for analytic functions with different types of loss function (Golubev *et al.* 1995).

Recently there has been growing interest in estimating functions with isolated singularities, stimulated by a variety of applications, such as change-point problems, spatially inhomogeneous data, image fragments restoration as well by their mathematical meaningfulness (see Müller 1992 and further references therein). At first glance, the known results on the optimal rates of convergence may suggest that only slow rates of convergence can be achieved for such functions with discontinuities. This, however, turns out not to be the case. For functions which are smooth except for a few points the discontinuities do not affect the convergence rate of the optimal estimators (Hall and Patil 1995).

In this paper we show that, in the Sobolev-type setting with  $L_2$  losses, regardless of the

presence of an unknown although finite number of jumps in the unknown regression function even the same asymptotic optimal constant can be attained as in the case of smooth functions without jumps. This optimal constant is known as Pinsker's constant.

In order to obtain this optimal result we need accurate estimators of the jump points. To detect the number and location of the jumps we construct an estimator depending on a sample version of the jump sizes  $f(x+) - f(x-)$ . For a similar kind of estimator we refer to (among others) Müller (1992), who used boundary kernels for estimating the location of a jump and its size. Wang (1995) detected jumps using wavelets.

We shall work with the Gaussian white noise model, as will be defined below in (2.1). Signal recovery in Gaussian white noise with variance tending to zero has already served for some time as a representative model for nonparametric curve estimation, having all the essential traits in a pure form. In contrast, in particular with the nonparametric regression model, with observations on a discrete grid, it entails minimal technical nuisance. This is reflected by the fact that, roughly speaking for the corresponding derivations that one makes in these models, in the discrete model one has to deal with summations and in the Gaussian model with integrals; hence the latter gives more elegant and transparent arguments. We conjecture that an approximation in the sense of Le Cam's deficiency distance should make it precise. The models are then asymptotically equivalent for all purposes of statistical decision with bounded loss. A first result of this kind has recently been established by Brown and Low (1996). They proved that the nonparametric regression model with observations on a discrete grid is asymptotically equivalent to the Gaussian white-noise model for Hölder classes with smoothness  $\beta > \frac{1}{2}$ . Although our results concern regression for Sobolev classes with discontinuities, we might expect from their work that similar results should hold also in our context. Hence the results obtained in this paper should hold also in the discrete-type model, but that requires even more technical proofs. Moreover it gives the idea what the corresponding asymptotically minimax estimator is in that model.

## 2. The model and main result

Suppose that we observe a random process  $X_\epsilon(t)$  satisfying the stochastic differential equation

$$dX_\epsilon(t) = f(t) dt + \epsilon d\mathscr{W}(t), \quad t \in [0, 1], \quad (2.1)$$

with some prescribed initial value  $X(0)$ , which is either constant or a random variable independent of  $\mathscr{W}$ ;  $\mathscr{W}(t)$  is a standard Wiener process and  $\epsilon$  is a known parameter, assumed to be small. The space of square-integrable functions (or signals) on  $[0, 1]$  is denoted by  $\mathbf{L}_2 = \mathbf{L}_2[0, 1]$  and  $\|\cdot\|$  is the usual  $\mathbf{L}_2$  norm.

Let  $m$  be a positive integer and  $Q > 0$  a constant, both given. Assume that the unknown signal  $f$  belongs to a class of functions  $\Phi_{B,\kappa,L}(m, Q)$ , for which there exist (not necessarily known) positive constants  $B, \kappa, L$  such that, for all  $f \in \Phi_{B,\kappa,L}(m, Q)$ , the following hold.

$$(A1) \sup_{t \in [0,1]} |f(t)| < B.$$

(A2) There exist jump points  $b_j, j = 0, \dots, S+1$  (unknown),  $0 = b_0 < \dots < b_{S+1} = 1$ , such that

(A2.1) the points are at least distance  $\kappa$  apart,

(A2.2) the jumps have at least size  $L$ , i.e.

$$L \leq \left| \lim_{t \downarrow b_j} f(t) - \lim_{t \uparrow b_j} f(t) \right| =: |f(b_{j+}) - f(b_{j-})|, \quad j = 1, \dots, S,$$

(A2.3)  $f$  is  $m$ -times differentiable on  $]b_j, b_{j+1}[$ ,  $j = 0, \dots, S$ , and

(A2.4)  $f$  belongs to a piecewise Sobolev ‘ellipsoid’, i.e.,

$$\sum_{j=0}^S \int_{b_j}^{b_{j+1}} \{f^{(m)}(t)\}^2 dt \leq Q.$$

Note that the existence of the limits in (A2.2) follows from (A2.4). Moreover, conditions (A1) and (A2.4) imply that the derivatives  $f^{(i)}(t)$ ,  $i = 1, \dots, m-1$  are bounded everywhere except at the jump points  $b_j$ . Obviously the functions in  $\Phi_{B,\kappa,L}(m, Q)$  are square integrable.

Define the quadratic risk for an estimator  $\hat{f}$  of  $f$  as follows:

$$R(\hat{f}, f) = \mathbf{E}_f \|\hat{f} - f\|^2. \quad (2.2)$$

Furthermore denote the minimax quadratic risk in estimating  $f$  with respect to the class of functions  $\Phi$  by

$$r_\epsilon(\Phi) = \inf_{\hat{f}} \sup_{f \in \Phi} R(\hat{f}, f), \quad (2.3)$$

where the infimum is taken over all estimators  $\hat{f}$ .

We derive the exact asymptotic behaviour of this minimax quadratic risk for the class  $\Phi_{B,\kappa,L}(m, Q)$ , described by the conditions (A1)–(A2.4), formulated in the following theorem.

**Theorem 2.1.** *Let  $\gamma(m, Q) = \{Q(2m+1)\}^{1/(2m+1)}(m/\pi(m+1))^{2m/(2m+1)}$ . The minimax quadratic risk of the above-defined model satisfies*

$$\lim_{\epsilon \downarrow 0} \epsilon^{-4m/(2m+1)} r_\epsilon(\Phi_{B,\kappa,L}(m, Q)) = \gamma(m, Q), \quad (2.4)$$

for arbitrary although fixed  $B, \kappa$  and  $L$ .

Note that the right-hand side of (2.4) is independent of  $B, \kappa$  and  $L$ . The proof of this theorem is outlined in Section 2.4. There also a projection-type estimator  $\tilde{f}$  is given which attains this optimal constant  $\gamma(m, Q)$ , Pinsker’s constant.

For the lower bound on the minimax risk in the setting (2.1) we refer to Pinsker (1980), and to a more recent paper by Belitser and Levit (1996) for the corresponding discrete setting. Note that our model allows functions  $f$  without jumps, while the additional restriction (A1) does not affect the technique used in these papers for obtaining the lower bounds in estimating such functions.

First in Section 2.1 we give the motivation of the chosen basis. Second in Section 2.2 we generalize the ideas of Pinsker (1980), such that an asymptotically minimax estimator  $\hat{f}(t) = \hat{f}(t; b_1, \dots, b_s)$ , in the case when the number and locations of the jump points are known, is obtained. Of course this estimator depends on the (unknown) jump points  $b_j$ . Then we estimate these jump points (see Section 2.3). Finally in Section 2.4 we argue that replacement, in the estimator  $\hat{f}$  for the case of known jump points, of the unknown jump points  $b_j$  by their estimators, results in an asymptotically minimax estimator  $\tilde{f}$  in the case when the jump points are unknown (cf. (2.20)). We stress again that we do not have to know the number of jump points, but because of the choice of  $\kappa$  this number is assumed to be finite.

## 2.1. Choice of the basis

Let us consider the basis arising from the following boundary value problem on the interval  $[a, b]$ , with Neumann conditions on the boundary

$$\begin{aligned} \left(\frac{-d^2}{dt^2}\right)^m u &= \lambda u, & t \in [a, b], \\ u^{(s)}(a) &= u^{(s)}(b) = 0, & s = m, \dots, 2m-1. \end{aligned} \tag{2.5}$$

The corresponding differential operator  $L = (-d^2/dt^2)^m$  is self-adjoint and semi positive definite on the space  $D = \{u \in C^{2m}[a, b]: u^{(s)}(a) = u^{(s)}(b) = 0 \text{ for } s = m, \dots, 2m-1\}$ . Furthermore the null space  $D_0$  of  $L$  is spanned by the polynomials of degree at most  $m-1$ . On  $D \setminus D_0$ ,  $L$  is compactly invertible and positive definite. Hence owing to the spectral theorem the normed eigenfunctions  $\varphi_m, \varphi_{m+1}, \dots$  of (2.5) corresponding to positive eigenvalues  $\lambda_m, \lambda_{m+1}, \dots$  and supplemented with an orthonormal set of polynomials  $\varphi_0, \dots, \varphi_{m-1}$  on  $[a, b]$  with degree at most  $m-1$ , with corresponding eigenvalues  $\lambda_0 = 0, \dots, \lambda_{m-1} = 0$ , provide a basis for  $D$ . Moreover it is an orthonormal basis for the whole of  $L_2[a, b]$ , as  $D$  is dense in  $L_2[a, b]$  (compare, for example, Coddington and Levinson 1955, Chapter 7).

The statistical importance of the eigenfunctions of the boundary value problem (2.5), in the Sobolev-type setting, was apparently first recognized by Golubev and Nussbaum (1990). Its discrete counterpart, the so-called Demmler–Reinsch basis, appeared in the non-parametric methods based on splines even earlier (see, for example, Speckman 1985).

In the theory of differential equations, extensive study of the asymptotic behaviour of the eigenvalues and eigenfunctions of the boundary value problem (2.5), usually in the broader setting of general linear differential operators of order  $m$ , was initiated by G.D. Birkhoff in 1908. A detailed description of these properties has been presented in the monographs by Neumark (1967, Section II.4) and Dunford and Schwartz (1971, Section XIX).

The particular form of (2.5) allows a more straightforward and detailed account of these properties. Professor J.J. Duistermaat from the University of Utrecht kindly agreed to review these properties at our request. In his recent report (Duistermaat 1995), sharper asymptotics for the eigenvalues and eigenfunctions of the boundary value problem (2.5), compared with

those presented in the references above, are obtained and a further study of the asymptotic properties of the corresponding eigenfunctions is made. In particular the sharper asymptotics of the eigenfunctions is necessary to prove our result.

Here we summarize the results that we use later.

**Theorem 2.2 (Duistermaat 1995).** *The non-zero eigenvalues  $\lambda_k$  of the boundary value problem (2.5) satisfy the relation*

$$\lambda_k = \rho_k^{2m} = \{k\pi(b-a)^{-1}\}^{2m}\{1 + o(1)\}, \quad k \rightarrow \infty, \quad (2.6)$$

and the corresponding eigenfunctions  $\varphi_k$  equal (for coefficients  $A_k, B_k$  such that they are orthonormalized)

$$A_k\{\cos(\rho_k t) + r_k(t)\}, \text{ for } k \text{ odd, and } B_k\{\sin(\rho_k t) + r_k(t)\} \text{ for } k \text{ even,}$$

where the functions  $r_k(t)$  satisfy

$$|r_k^{(i)}(t)| \leq C_{i,m} \rho_k^i (e^{-\rho_k \alpha_m (t-a)} + e^{-\rho_k \alpha_m (b-t)}),$$

for any  $i$  and some constants  $C_{i,m}$  depending on  $i, m$  and  $b-a$  and  $\alpha_m > 0$  depending only on  $m$ .

Note that for  $m=1$  the eigenvalues equal  $\{k\pi/(b-a)\}^2$  and the eigenfunctions are exactly cosines.

Observe that the functions  $\{(\lambda_k)^{-1/2} \varphi_k^{(m)}(\cdot)\}_{k \geq m}$  are also orthonormal. Therefore, for a function  $f \in C^m[a, b]$  satisfying

$$\int_a^b \{f^{(m)}(t)\}^2 dt \leq Q,$$

we can derive by partial integration and Bessel's inequality that the Fourier coefficients  $\theta_k = \int_a^b f(t) \varphi_k(t) dt$  belong to the ellipsoid

$$\sum_{k=m}^{\infty} \lambda_k \theta_k^2 = \sum_{k=0}^{\infty} \lambda_k \theta_k^2 \leq Q. \quad (2.7)$$

We note that, in the paper by Pinsker (1980), among other things, an optimal estimator is constructed for signals with a restriction on Fourier coefficients of precisely this type. Our goal here is to develop a similar but broader technique, incorporating the piecewise smooth functions.

## 2.2. Optimal estimation with known jump points

Given (2.7) we follow and generalize the ideas of Pinsker in deriving an optimal estimator of  $f$  (see also Belitser and Levit 1996). Denote the corresponding orthonormal eigenvalues and eigenfunctions of (2.5) on the interval  $[b_j, b_{j+1}]$  by  $\lambda_{k,j} = \lambda_k (b_{j+1} - b_j)^{-2m}$  and  $\varphi_{k,j}(t) =$

$(b_{j+1} - b_j)^{-1/2} \varphi_k((b_{j+1} - b_j)^{-1}(t - b_j))$  for  $j = 0, \dots, S$ , where  $\varphi_k$  and  $\lambda_k$  denote here the corresponding quantities related to the standard interval  $[0, 1]$ .

Let us emphasize the fact that the jump points  $b_1$  up to  $b_S$  are unknown. However, let us pretend for a moment that they are known. Rewrite the observation process (2.1) into the following equivalent sequence of observations ( $j = 0, \dots, S$ ):

$$Y_{k,j} = \theta_{k,j} + \epsilon \xi_{k,j}, \quad k = 0, 1, \dots, \quad (2.8)$$

with

$$Y_{k,j} = \int_{b_j}^{b_{j+1}} \varphi_{k,j}(t) dX_\epsilon(t), \quad \theta_{k,j} = \int_{b_j}^{b_{j+1}} f(t) \varphi_{k,j}(t) dt$$

and

$$\xi_{k,j} = \int_{b_j}^{b_{j+1}} \varphi_{k,j}(t) d\mathcal{W}'(t).$$

Here  $\xi_{1,0}, \dots, \xi_{1,S}, \xi_{2,0}, \dots, \xi_{2,S}, \dots$  are independent and standard Gaussian random variables.

Consider the following tapered orthogonal series estimator of  $f$ :

$$\hat{f}(t) = \sum_{k=0}^{\infty} \bar{\theta}_{k,j} \varphi_{k,j}(t), \quad t \in [b_j, b_{j+1}[ , \quad j = 0, \dots, S \quad (2.9)$$

(the last subinterval is  $[b_S, 1]$  instead of  $[b_S, 1[$ ), where

$$\bar{\theta}_{k,j} = h_{k,j} Y_{k,j}, \quad 0 \leq h_{k,j} \leq 1,$$

is an estimate of  $\theta_{k,j}$  and the tapering coefficients  $h_{k,j}$  are chosen such that

$$h_{k,j} = \begin{cases} 1, & k = 0, \dots, m-1, \\ (1 - c_j \lambda_{k,j}^{1/2})_+, & k \geq m, \end{cases} \quad (2.10)$$

with  $c_j$  the solutions of the equations

$$c_j (b_{j+1} - b_j) Q = \epsilon^2 \sum_{k=0}^{\infty} \lambda_{k,j}^{1/2} (1 - c_j \lambda_{k,j}^{1/2})_+. \quad (2.11)$$

These coefficients are shown below to be optimal in the sense that the maximal quadratic risk of the pseudo-estimator  $\hat{f}$  over  $\Phi_{B,K,L}(m, Q)$  attains asymptotically the optimal bound (2.4). In order to explain this observe first that the risk  $R(\hat{f}, f)$ , defined in (2.2), of an estimator  $\hat{f}$  of the form (2.9) with arbitrary tapering coefficients  $h_{k,j}$  is equal to

$$\sum_{j=0}^S \sum_{k=0}^{\infty} \mathbf{E}_f (\bar{\theta}_{k,j} - \theta_{k,j})^2$$

or, written out,

$$\sum_{j=0}^S \sum_{k=0}^{\infty} \{\epsilon^2 h_{h,j}^2 + \theta_{k,j}^2 (h_{k,j} - 1)^2\}.$$

Therefore the maximal quadratic risk of  $\hat{f}$  over  $\Phi_{B,\kappa,L}(m, Q)$  for arbitrary tapering coefficients  $h_{k,j}$  can be bounded as follows (cf. (2.7)):

$$\begin{aligned} \sup_{f \in \Phi_{B,\kappa,L}(m, Q)} R(\hat{f}, f) &= \sup_{f \in \Phi_{B,\kappa,L}(m, Q)} \sum_{j=0}^S \sum_{k=0}^{\infty} \{\epsilon^2 h_{k,j}^2 + \theta_{k,j}^2 (h_{k,j} - 1)^2\} \\ &\leq \sup_{\{\theta_{k,j}: \sum_j \sum_k \lambda_{k,j} \theta_{k,j}^2 \leq Q\}} \sum_{j=0}^S \sum_{k=0}^{\infty} \{\epsilon^2 h_{k,j}^2 + \theta_{k,j}^2 (h_{k,j} - 1)^2\} \\ &\leq \sup_{\{Q_j: \sum_{j=0}^S Q_j = Q\}} \sup_{\{\theta_{k,j}: \sum_k \lambda_{k,j} \theta_{k,j}^2 \leq Q_j\}} \sum_{j=0}^S \sum_{k=0}^{\infty} \{\epsilon^2 h_{k,j}^2 + \theta_{k,j}^2 + \theta_{k,j}^2 (h_{k,j} - 1)^2\} \\ &\leq \sup_{\{Q_j: \sum_{j=0}^S Q_j = Q\}} \sum_{j=0}^S \left( \epsilon^2 \sum_{k=0}^{\infty} h_{k,j}^2 + Q_j \sup_{k \geq m} \lambda_{k,j}^{-1} (h_{k,j} - 1)^2 \right). \end{aligned}$$

Taking the infimum over all tapering coefficients  $0 \leq h_{k,j} \leq 1$  for  $k \geq m$  in the last expression we see that the minimax risk in estimating  $f$ , as defined in (2.3), is bounded by

$$r_{\epsilon}(\Phi_{B,\kappa,L}(m, Q)) \leq \inf_{c_j} \inf_{\{h_{k,j}: (h_{k,j} - 1)^2 \leq c_j \lambda_{k,j}\}} \sup_{Q_j} \sum_{j=0}^S \left( \epsilon^2 \sum_{k=m}^{\infty} h_{k,j}^2 + c_j^2 Q_j \right) + (S+1)m\epsilon^2. \quad (2.12)$$

If the class  $\Phi_{B,\kappa,L}(m, Q)$  comprises only those functions  $f$  with given values of  $Q_0, \dots, Q_S$  (i.e. the distribution of the power of the signal for the intervals  $[b_j, b_{j+1}[$  is known), then it is not difficult to see (Belitser and Levit 1996) that the optimal tapering coefficients  $h_{k,j}$  in (2.12) are given by (2.10) and

$$c_j Q_j = \epsilon^2 \sum_{k=m}^{\infty} \lambda_{k,j}^{1/2} (1 - c_j \lambda_{k,j}^{1/2})_+ \quad (2.13)$$

(cf. (2.11)). However, since the  $Q_j$  are not known, we replace them by

$$\tilde{Q}_j = (b_{j+1} - b_j) Q \quad (2.14)$$

in (2.13), i.e., we presume for the moment that the ‘worst’ possible distribution of the signals power  $Q_j$  over the intervals  $[b_j, b_{j+1}[$  occurs when the  $Q_j$  are proportional to their lengths (constant power per unit of time). This hypothesis can be easily backed up by an elementary

calculation. In Section 3 we show that, with  $c_j$  and  $h_{k,j}$  defined above, asymptotically the  $c_j$  do not depend on  $j$  (cf. (3.13)) and

$$\sup_{\{Q_j: \sum_{j=0}^S Q_j = Q\}} \sum_{j=0}^S \left( \epsilon^2 \sum_{k=m}^{\infty} h_{k,j}^2 + c_j^2 Q_j \right) = \gamma(m, Q) \epsilon^{4m/(2m+1)} \{1 + o(1)\}, \quad \epsilon \rightarrow 0. \quad (2.15)$$

Denote therefore from now on  $c_j$  by  $c$ .

Thus the estimator  $\hat{f}$  is an asymptotically minimax estimator, in the case when the jump points are known, i.e.,

$$\sup_{f \in \Phi_{B,K,L}(m,Q)} R(\hat{f}, f) \leq \gamma(m, Q) \epsilon^{4m/(2m+1)} \{1 + o(1)\}, \quad \epsilon \rightarrow 0. \quad (2.16)$$

To summarize, using the method of Pinsker on each of the subintervals  $[b_j, b_{j+1}] \subset [0, 1]$  we obtain, in the case of *known* jump points, the same rate and even the same constant as in the case *without* jump points.

However, the jump points on which the construction of the estimator  $\hat{f}$  above heavily depends (note that all the quantities involved, namely  $h_{k,j}$ ,  $\varphi_{k,j}$  and  $\lambda_{k,j}$  in the above relations (2.5), (2.8), (2.10), (2.13) and (2.14) depend on them) are actually not known.

Our next goal is to show that combining the method described above with the use of sufficiently accurate estimators  $\hat{b}_j$  of  $b_j$  will enable us to obtain asymptotically the same quality of estimation as in (2.16). Therefore we now explain how we estimate the (number of) jump points.

### 2.3. Estimation of the jump points

There exist estimators of the jump points that converge with rate  $\epsilon^2$  to the true jump points (Korostelev 1987). However, we define here jump-point estimators which are easy to calculate and are nearly optimal but converge fast enough to the true jump points for our purposes.

In the sequel the bandwidth parameter  $h$  equals  $\epsilon^2 (\ln \epsilon^{-2})^{1+\delta}$  for an arbitrary but fixed  $\delta > 0$ . Furthermore suppose that  $\epsilon$  is small enough such that  $2h < \kappa$ . Define a set of grid points, separated at distance  $\epsilon^2$  by

$$A = \{a_i = i\epsilon^2, i = \lceil h\epsilon^{-2} \rceil, \dots, \lfloor (1-h)\epsilon^{-2} \rfloor\}$$

on the interval  $[h, 1-h]$  and let  $N$  be the number of elements of  $A$ . For every grid point  $a_i \in A$  we calculate the quantity

$$T(a_i) = h^{-1} \int_{a_i-h}^{a_i} dX_\epsilon(t) - h^{-1} \int_{a_i}^{a_i+h} dX_\epsilon(t).$$

We expect that  $|T(a_i)|$  is ‘large’ only if there is a jump point  $b_j$  in the interval  $[a_i - h, a_i + h]$ . More precisely the following result will be proved in Section 3.



**Lemma 2.3.** For every  $\gamma > 0$  and  $0 < \alpha < \delta/2$ , uniformly in  $f \in \Phi_{B,\kappa,L}(m, Q)$

(i)

$$\mathbf{P}_f\{|T(a_i)| > (\ln \epsilon^{-2})^{-\alpha}\} = o(\epsilon^\gamma) \quad \epsilon \rightarrow 0,$$

for every grid point  $a_i \in A$  such that there are no jump points  $b_j$  in the interval  $[a_i - h, a_i + h]$ ;

(ii)

$$\mathbf{P}_f\{|T(a_i)| < (\ln \epsilon^{-2})^{-\alpha}\} = o(\epsilon^\gamma), \quad \epsilon \rightarrow 0$$

for every grid point  $a_i \in A$  such that for some  $j = 1, \dots, S$ ,  $|a_i - b_j| < h/2$ .

However, it is possible that there are more than one grid point  $a_i$  ‘near’ a jump point  $b_j$  that have a large value  $|T(a_i)|$ . Thus we have to classify those  $a_i$  detecting the same  $b_j$ .

Fix an  $\alpha$  between 0 and  $\delta/2$ . Neighbouring grid points  $a_i$  and  $a_{i+1}$  are called connected if both  $|T(a_i)|$  and  $|T(a_{i+1})|$  exceed  $(\ln \epsilon^{-2})^{-\alpha}$ . Consider all subsets of  $A$ , consisting of more than  $\lceil (\log \epsilon^{-2})^{1+\delta} \rceil$  connected grid points and denote these subsets (or so-called cluster sets) by  $T_j$ ,  $j = 1, \dots, \tilde{S}$ . We use here the convention that, if such subsets do not exist, we put  $\tilde{S} = 0$ . Loosely speaking these sets contain the candidates for jump points. Finally we define estimator(s)  $\hat{b}_j$  of possible jump points as follows:

$$\hat{b}_j = \arg \max_{a_i \in T_j} \{|T(a_i)|\}, \quad j = 1, \dots, \tilde{S}, \quad (2.17)$$

and  $\hat{b}_0 = 0$ ,  $\hat{b}_{\tilde{S}+1} = 1$ . Note that, if  $\tilde{S}$  is larger than  $S$ , then at least for one  $a_i$  there is no jump  $b_j$  in the interval  $[a_i - h, a_i + h]$  and nevertheless  $|T(a_i)| > (\ln \epsilon^{-2})^{-\alpha}$ . Fortunately owing to Lemma 2.3(i), with  $\gamma = 2 + 4m/(2m + 1)$ , this happens only with small probability, namely

$$\mathbf{P}_f\{S < \tilde{S}\} \leq \sum_{a_i: \text{no jump in } [a_i - h, a_i + h]} \mathbf{P}_f\{|T(a_i)| > (\ln \epsilon^{-2})^{-\alpha}\} = o\{(\epsilon^2)^{2m/(2m+1)}\}. \quad (2.18)$$

In the case when  $\tilde{S} < S$  there is at least one  $b_j$  such that there exists an  $a_i$  within distance  $h/2$  of  $b_j$  having  $|T(a_i)|$  value smaller than  $(\ln \epsilon^{-2})^{-\alpha}$ . Applying Lemma 2.3(ii) with  $\gamma = 4m/(2m + 1)$  yields

$$\mathbf{P}_f\{S > \tilde{S}\} \leq \sum_{b_j} \mathbf{P}_f\{|T(a_i)| < (\ln \epsilon^{-2})^{-\alpha}\} = o\{(\epsilon^2)^{2m/(2m+1)}\}. \quad (2.19)$$

Indeed with high probability the right number of jump points is estimated. Note that, since  $2h > \kappa$ , the minimal distance of the jump points, it cannot happen that two real jump points are ‘seen’ as one. The following bound concerning the accuracy of estimating  $b_j$  is frequently used in the sequel.

**Corollary 2.4** *Uniformly in  $f \in \Phi_{B,\kappa,L}(m, Q)$  we have*

$$\mathbf{E}_f |\hat{b}_j - b_j|^2 \leq O[\{\epsilon^2 (\ln \epsilon^{-2})^{1+\delta}\}^2], \quad \epsilon \rightarrow 0.$$

Observe that the above expectation is obviously bounded by  $(2h)^2 + \mathbf{P}_f\{|\hat{b}_j - b_j| > 2h\}$ . Use Lemma 2.3 to finish the proof of this corollary.

## 2.4. The proposed estimator and the framework of the proof

Substitute  $\hat{b}_j$  and  $\hat{b}_{j+1}$  in (2.5) and (2.8). Denote the eigenfunctions  $\varphi_k(\cdot, \hat{b}_j, \hat{b}_{j+1})$  by  $\hat{\varphi}_{k,j}$  and  $\lambda_{k,j}(\hat{b}_j, \hat{b}_{j+1})$  by  $\hat{\lambda}_{k,j}$ . The observation process (2.1) can be rewritten into the sequence of observations

$$\hat{Y}_{k,j} = \hat{\theta}_{k,j} + \epsilon \hat{\xi}_{k,j}, \quad k = 0, 1, \dots,$$

where

$$\hat{Y}_{k,j} = \int_{\hat{b}_j}^{\hat{b}_{j+1}} \hat{\varphi}_{k,j}(t) dX_c(t), \quad \hat{\theta}_{k,j} = \int_{\hat{b}_j}^{\hat{b}_{j+1}} f(t) \hat{\varphi}_{k,j}(t) dt$$

and

$$\hat{\xi}_{k,j} = \int_{\hat{b}_j}^{\hat{b}_{j+1}} \hat{\varphi}_{k,j}(t) d\mathcal{W}(t).$$

Furthermore substitute  $\hat{b}_j$  and  $\hat{b}_{j+1}$  also in (2.10), (2.11) and (2.14). As suggested in the discussion in Section 2.2 we propose the following estimator  $\tilde{f}$  for the unknown regression function  $f$ :

$$\tilde{f}(t) = \sum_{k=0}^{\infty} \tilde{\theta}_{k,j} \hat{\varphi}_{k,j}(t), \quad t \in [\hat{b}_j, \hat{b}_{j+1}[], \quad j = 0, \dots, \tilde{S}, \quad (2.20)$$

where the estimator  $\tilde{\theta}_{k,j}$  is equal to  $\hat{h}_{k,j} \hat{Y}_{k,j}$  with the (estimated) tapering coefficients  $\hat{h}_{k,j}$  defined by the relations

$$\hat{h}_{k,j} = \{1 - (c \hat{\lambda}_{k,j})^{1/2}\}_+, \quad k = 0, 1, \dots, \quad j = 0, \dots, \tilde{S},$$

and  $c$  is given by (3.13) below (note that again the first  $m$  tapering coefficients equal 1).

As the regression function is supposed to be bounded by some (unknown) constant  $B$  we continue the proof with the following truncated version of  $\tilde{f}$ :

$$\tilde{f}_{\text{tr}}(t) = \begin{cases} -B(\epsilon), & \tilde{f} < -B(\epsilon), \\ \tilde{f}(t), & |\tilde{f}| \leq B(\epsilon), \\ B(\epsilon), & \tilde{f} > B(\epsilon). \end{cases}$$

Here  $B(\epsilon)$  is a sequence tending ‘slowly’ to infinity, e.g.,  $B(\epsilon) = \ln \epsilon^{-2}$ ; thus, for  $\epsilon$  small enough, it exceeds  $B$ . The risk of this estimator is bounded by  $[2B(\epsilon)]^2$ . Owing to the

inequalities (2.18) and (2.19) we know that the probability that  $S$  is not equal to  $\tilde{S}$  is sufficiently small, namely,

$$\mathbf{P}_f\{S \neq \tilde{S}\} = o\{(\epsilon^2)^{2m/(2m+1)}\},$$

as  $\epsilon$  tends to zero. Moreover we have, according to Corollary 2.4,

$$\mathbf{E}_f|\hat{b}_j - b_j|^2 \leq O[\{\epsilon^2(\ln \epsilon^{-2})^{1+\delta}\}^2] = o\{(\epsilon^2)^{2m/(2m+1)}\}, \quad \epsilon \rightarrow 0. \quad (2.21)$$

Combining these two facts we see that the complement of the event  $F = \{S = \tilde{S}\} \cap \{|\hat{b}_j - b_j| < 2h, j = 1, \dots, S\}$  has small probability. Thus we derive for the risk of  $\tilde{f}_{\text{tr}}$ , provided that  $\epsilon$  is small enough,

$$\begin{aligned} R(\tilde{f}_{\text{tr}}, f) &= \mathbf{E}_f \|\tilde{f}_{\text{tr}} - f\|^2 \mathbf{I}_F + \mathbf{E}_f \|\tilde{f}_{\text{tr}} - f\|^2 \mathbf{I}_{F^c} \\ &\leq \mathbf{E}_f \|\tilde{f} - f\|^2 \mathbf{I}_F + o\{(\epsilon^2)^{2m/(2m+1)}\}, \quad \epsilon \rightarrow 0. \end{aligned}$$

Hence it suffices to restrict ourselves to the case  $S$  equals  $\tilde{S}$  and we can assume that  $\hat{b}_j$  estimates  $b_j$  within distance  $2h$ . From now on we take expectations conditioned on the event  $F$  without mentioning it. Applying the Parseval equality we derive for the risk of  $\tilde{f}$

$$\begin{aligned} R(\tilde{f}, f) &= \sum_{j=0}^S \sum_{k=0}^{\infty} \mathbf{E}_f (\tilde{\theta}_{k,j} - \hat{\theta}_{k,j})^2 \\ &= \sum_{j=0}^S \sum_{k=0}^{\infty} \mathbf{E}_f \{\epsilon \hat{h}_{k,j} \hat{\xi}_{k,j} - (1 - \hat{h}_{k,j}) \hat{\theta}_{k,j}\}^2 \\ &= \sum_{j=0}^S \mathbf{E}_f \hat{L}_{\epsilon,j}(f) + R_{1,j} + R_{2,j} - R_{3,j}, \end{aligned}$$

where

$$\begin{aligned} \hat{L}_{\epsilon,j}(f) &= \sum_{k \geq m} (1 - \hat{h}_{k,j})^2 \hat{\theta}_{k,j}^2 + \epsilon^2 \hat{h}_{k,j}^2, \\ R_{1,j} &= R_{1,j}(f, \hat{b}_j, \hat{b}_{j+1}) = \epsilon^2 \sum_{k=0}^{m-1} \mathbf{E}_f \hat{\xi}_{k,j}^2, \\ R_{2,j} &= R_{2,j}(\hat{b}_j, \hat{b}_{j+1}) = \epsilon^2 \sum_{k \geq m} \mathbf{E}_f \hat{h}_{k,j}^2 (\hat{\xi}_{k,j}^2 - 1), \\ R_{3,j} &= R_{3,j}(f, \hat{b}_j, \hat{b}_{j+1}) = 2\epsilon \sum_{k \geq m} \mathbf{E}_f (1 - \hat{h}_{k,j}) \hat{h}_{k,j} \hat{\theta}_{k,j} \hat{\xi}_{k,j}. \end{aligned}$$

For the remainder terms  $R_{l,j}$  ( $l = 1, 2, 3$ ), we deduce in Section 3 the following bounds for  $\epsilon$  tending to 0, all uniformly for  $f \in \Phi_{B,\kappa,L}(m, Q)$ :

$$R_{1,j}(f, \hat{b}_j, \hat{b}_{j+1}) = o[\epsilon^2 \ln^3(\epsilon^{-2})], \quad (2.22)$$

$$R_{2,j}(f, \hat{b}_j, \hat{b}_{j+1}) = o(1)\mathbf{E}_f \hat{L}_{\epsilon,j}(f) + O\{\epsilon^2 \ln^3(\epsilon^{-2})\} \quad (2.23)$$

and

$$|R_{3,j}(f, \hat{b}_j, \hat{b}_{j+1})| = o(1)\mathbf{E}_f \hat{L}_{\epsilon,j}(f) + O\{\epsilon^2 \ln^3(\epsilon^{-2})\}. \quad (2.24)$$

This implies uniformly in  $f \in \Phi_{B,\kappa,L}(m, Q)$

$$R(\tilde{f}, f) = \{1 + o(1)\} \sum_{j=0}^S \mathbf{E}_f \hat{L}_{\epsilon,j}(f) + O\{\epsilon^2 \ln^3(\epsilon^{-2})\}, \quad \epsilon \rightarrow 0.$$

Denote by  $L_{\epsilon,j}(f)$  the quantity

$$L_{\epsilon,j}(f) = \sum_{k=m}^{\infty} (1 - h_{k,j})^2 \theta_{k,j}^2 + \epsilon^2 h_{k,j}^2,$$

with  $h_{k,j}$  defined in (2.10) and (2.11). The next step in the proof is to show that, uniformly in  $f \in \Phi_{B,\kappa,L}(m, Q)$ , we have for  $j = 0, \dots, S$

$$\mathbf{E}_f \hat{L}_{\epsilon,j}(f) \leq L_{\epsilon,j}(f) \{1 + o(1)\} + o(\epsilon^{4m/(2m+1)}), \quad \epsilon \rightarrow 0 \quad (2.25)$$

(cf. Section 3). Thus for the maximal quadratic risk of  $\tilde{f}$  we have

$$\sup_{f \in \Phi_{B,\kappa,L}(m, Q)} R(\tilde{f}, f) \leq \sup_{\Phi_{B,\kappa,L}(m, Q)} \sum_{j=0}^S L_{\epsilon,j}(f) \{1 + o(1)\} + o(\epsilon^{4m/(2m+1)}), \quad \epsilon \rightarrow 0. \quad (2.26)$$

Finally note that in Section 2.2 we have explained that

$$\sup_{\Phi_{B,\kappa,L}(m, Q)} \sum_{j=0}^S L_{\epsilon,j}(f) \leq \gamma(m, Q) \epsilon^{4m/(2m+1)} \{1 + o(1)\}, \quad \epsilon \rightarrow 0. \quad (2.27)$$

Combination of (2.26) and (2.27) finishes the proof of Theorem 2.1.

### 3. Proofs

**Proof of Lemma 2.3.** Suppose that  $a_i \in A$ , the set of grid points is such that there are no jump points  $b_j$  in the interval  $[a_i - h, a_i + h]$ . For any such grid point we bound the bias of  $T(a_i)$ , uniformly in  $f \in \Phi_{B,\kappa,L}(m, Q)$ , as follows:

$$|\mathbf{E}_f T(a_i)| = h^{-1} \left| \int_{a_i-h}^{a_i} f(t) dt - \int_{a_i}^{a_i+h} f(t) dt \right| = \text{const. } h \leq \frac{1}{2} (\ln \epsilon^{-2})^{-\alpha}, \quad (3.1)$$

provided that  $\epsilon$  is sufficiently small.  $T(a_i)$  is normally distributed with expectation  $\mathbf{E}_f T(a_i)$  and variance  $2\epsilon^2 h^{-1}$ . Therefore, for any  $\gamma > 0$ , using the ‘tail’ approximation of a normal distribution and (3.1) we derive Lemma 2.3(i) with the following steps:

$$\begin{aligned}
\mathbf{P}_f\{|T(a_i)| > (\ln \epsilon^{-2})^{-\alpha}\} &\leq \mathbf{P}_f\{|T(a_i) - \mathbf{E}_f T(a_i)| > (\ln \epsilon^{-2})^{-\alpha} - |\mathbf{E}_f T(a_i)|\} \\
&\leq \mathbf{P}_f\{|T(a_i) - \mathbf{E}_f T(a_i)| > \frac{1}{2}(\ln \epsilon^{-2})^{-\alpha}\} \\
&\leq 4\{(\pi h)^{-1}\epsilon^2\}^{1/2}(\ln \epsilon^{-2})^\alpha \exp\{-4\epsilon^{-2}(\ln \epsilon^{-2})^{-2\alpha}h\} \\
&= 4\pi^{-1/2}(\ln \epsilon^{-2})^{\alpha-(\delta/2)-(1/2)} \exp\{-4^{-2}(\ln \epsilon^{-2})^{1+\delta-2\alpha}\} \\
&= o(\epsilon^\gamma),
\end{aligned}$$

if  $\alpha$  is chosen to be smaller than  $\delta/2$ .

Let  $a_i$  now be a grid point such that  $|a_i - b_j| < h/2$ . For any such  $a_i$  we bound the bias of  $T(a_i)$  from below by  $L/3$ :

$$|\mathbf{E}_f T(a_i)| = h^{-1} \left| \int_{a_i-h}^{a_i} f(t) dt - \int_{a_i}^{a_i+h} f(t) dt \right| = |f(b_{j+}) - f(b_{j-})| \{1 + o(1)\} > \frac{L}{3}. \quad (3.2)$$

Using (3.2) and taking arbitrary  $\gamma > 0$  we finally prove Lemma 2.3(ii) as follows:

$$\begin{aligned}
\mathbf{P}_f\{|T(a_i)| < (\ln \epsilon^{-2})^{-\alpha}\} &\leq \mathbf{P}_f\left\{|T(a_i) - \mathbf{E}_f T(a_i)| > \frac{L}{4}\right\} \\
&\leq 8\{(\pi h)^{-1}\epsilon^2\}^{1/2}L^{-1} \exp\left\{-\left(\frac{L}{8\epsilon}\right)^2 h\right\} \\
&= 8\{\pi^{1/2}L(\ln \epsilon^{-2})^{(1+\delta)}\}^{-1} \exp\left\{-\left(\frac{L}{8}\right)^2 (\ln \epsilon^{-2})^{1+\delta}\right\} \\
&= o(\epsilon^\gamma). \quad \square
\end{aligned}$$

**Proof of (2.22).** Define for  $i, j$ , each varying from  $\lceil h\epsilon^{-2} \rceil$  to  $\lfloor (1-h)\epsilon^{-2} \rfloor$ , the (not necessarily independent but) standard Gaussian distributed random variables

$$\xi_{k,i,l} = \int_{a_i}^{a_l} \varphi_k(t; a_i, a_l) d\mathcal{W}(t), \quad k = 0, \dots, m-1, m, \dots,$$

where  $a_k, a_l$  are grid points taken from the set  $A$  defined in Section 2.3. Observe that for every  $j = 0, \dots, S$  we have

$$R_{1,j}(f, \hat{b}_j, \hat{b}_{j+1}) = 4\epsilon^2 \sum_{k=0}^{m-1} \mathbf{E}_f \ln e^{\xi_{k,j}^2/4}.$$

Note that for every  $j$  there are indices  $i$  and  $l$  such that  $\hat{\xi}_{k,j}$  is equal to  $\xi_{k,i,l}$  as the estimators  $\hat{b}_j$  and  $\hat{b}_{j+1}$  belong to the set of grid points  $A$ . Therefore we have

$$R_{1,j}(f, \hat{b}_j, \hat{b}_{j+1}) \leq 4\epsilon^2 \sum_{k=0}^{m-1} \mathbf{E}_f \ln \left( \sum_{i,l} e^{\xi_{k,i,l}^2/4} \right).$$

By the Jensen inequality we finally obtain that the remainder term  $R_1$  is  $o\{\epsilon^2 \ln^3(\epsilon^{-2})\}$  (for  $\epsilon$  tending to 0 and because  $N \leq \epsilon^{-2}$ ):

$$\begin{aligned} R_{1,j}(f, \hat{b}_j, \hat{b}_{j+1}) &\leq 4\epsilon^2 \sum_{k=0}^{m-1} \ln \left( \sum_{i,l} \mathbf{E}_f e^{\xi_{k,i,l}^2/4} \right) \\ &= 4\epsilon^2 m \ln(2^{1/2} N^2) = o\{\epsilon^2 \ln^3(\epsilon^{-2})\}. \quad \square \end{aligned}$$

**Proof of (2.23).** Using (2.10) and (2.11) we can associate tapering coefficients  $h_{k,i,l}$  with any pair  $a_i, a_l$  in the same way that it has been done for the pairs  $b_j, b_{j+1}$ . In particular,  $\hat{h}_{k,j} = h_{k,i,l}$  if  $\hat{b}_j = a_i$  and  $\hat{b}_{j+1} = a_l$ . Denote by  $\|x\|$  the  $l_2$  norm of a sequence  $(x_k)_{k \geq m}$ . Recall the random variables  $\xi_{k,i,l}$  for  $k = m, \dots$ , defined above.

Applying first the Cauchy–Schwarz inequality and then the Jensen inequality to the concave function  $\ln^2(x + e)$  (for  $x$  positive) we bound  $R_{2,j}$  as follows:

$$\begin{aligned} R_{2,j}^2 &\leq (2\epsilon)^4 \mathbf{E}_f \|\hat{h}_{\cdot,j}\|^2 \mathbf{E}_f \{(4\|\hat{h}_{\cdot,j}\|\}^{-1} \left( \sum_{k=m}^{\infty} \hat{h}_{k,j}^2 (\xi_{k,j}^2 - 1) \right)^2 \\ &\leq (2\epsilon)^4 \mathbf{E}_f \|\hat{h}_{\cdot,j}\|^2 \mathbf{E}_f \ln^2 \sum_{i,l} \exp \left( (4\|h_{\cdot,i,l}^2\|)^{-1} \sum_{k=m}^{\infty} h_{k,i,l}^2 (\xi_{k,i,l}^2 - 1) \right) \\ &\leq (2\epsilon)^4 \mathbf{E}_f \|\hat{h}_{\cdot,j}\|^2 \ln^2 \left\{ \sum_{i,l} \mathbf{E}_f \exp \left( (4\|h_{\cdot,i,l}^2\|)^{-1} \sum_{k=m}^{\infty} h_{k,i,l}^2 (\xi_{k,i,l}^2 - 1) \right) + e \right\} \\ &\leq (2\epsilon)^4 \mathbf{E}_f \|\hat{h}_{\cdot,j}\|^2 \ln^2 \left\{ N^2 \max_{i,l} \mathbf{E}_f \exp \left( (4\|h_{\cdot,i,l}^2\|)^{-1} \sum_{k=m}^{\infty} h_{k,i,l}^2 (\xi_{k,i,l}^2 - 1) \right) + e \right\}. \quad (3.3) \end{aligned}$$

According to the distribution of the random variables  $\xi_{k,i,l}$  we see that the second expectation  $\mathbf{E}_f \exp \{(4\|h_{\cdot,i,l}^2\|)^{-1} \sum_{k \geq m} h_{k,i,l}^2 (\xi_{k,i,l}^2 - 1)\}$  equals

$$\prod_{k=m}^{\infty} \exp \left[ -h_{k,i,l}^2 (4\|h_{\cdot,i,l}^2\|)^{-1} - \frac{1}{2} \ln \{1 - h_{k,i,l}^2 (2\|h_{\cdot,i,l}^2\|)^{-1}\} \right]$$

and using the elementary inequality  $\ln(1 - x) \geq -x - x^2$  for  $|x| \leq \frac{1}{2}$  this is bounded by  $\prod_{k \geq m} \exp \{h_{k,i,l}^4 (8\|h_{\cdot,i,l}^2\|^2)^{-1}\}$ . Substituting the last expression into (3.3) we obtain

$$\begin{aligned} R_{2,j}^2 &\leq (2\epsilon)^4 \mathbf{E}_f \|\hat{h}_{\cdot,j}\|^2 \ln^2 \left\{ N^2 \max_{i,l} \exp \left( (8\|h_{\cdot,i,l}^2\|^2)^{-1} \sum_{k=m}^{\infty} h_{k,i,l}^4 \right) + e \right\} \\ &\leq (2\epsilon)^4 \mathbf{E}_f \|\hat{h}_{\cdot,j}\|^2 \ln^2 \{e(N^2 + 1)\}. \end{aligned}$$

Using the Cauchy inequality,

$$2ab \leq \gamma^{-1} a^2 + \gamma b^2, \quad \gamma > 0, \quad a, b \in \mathbf{R}, \quad (3.4)$$

and choosing  $\gamma = \ln \{e(N^2 + 1)\}$  we finish the proof of (2.23) as follows:

$$\begin{aligned} R_{2,j} &\leq 4\epsilon^2 \ln \{e(N^2 + 1)\} (\mathbf{E}_f \|\hat{h}_{\cdot,j}\|^2)^{1/2} \\ &\leq 2\epsilon^2 [\ln \{e(N^2 + 1)\}]^{-1} \mathbf{E}_f \|\hat{h}_{\cdot,j}\|^2 \ln^3 \{e(N^2 + 1)\} \\ &= o(1) \mathbf{E}_f \hat{L}_{\epsilon,j}(f) + O(\epsilon^2 \ln^3 \epsilon^{-2}). \end{aligned} \quad \square$$

**Proof of (2.24).** Denote the sequence  $\{(1 - \hat{h}_{k,j})\hat{h}_{k,j}\hat{\theta}_{k,j}\}_{k \geq m}$  by  $(\hat{\mu}_{k,j})_{k \geq m}$ . Furthermore define, for  $k = m, \dots$ ,

$$\mu_{k,i,l} = (1 - h_{k,i,l})h_{k,i,l} \int_{a_i}^{a_l} f(t)\varphi_k(t, a_i, a_l) dt.$$

Using the Cauchy–Schwarz inequality and the Jensen inequality we obtain as above

$$\begin{aligned} R_{3,j}^2 &\leq (2\epsilon)^2 \mathbf{E}_f \|\hat{\mu}_{\cdot,j}\|^2 \mathbf{E}_f \left( \|\hat{\mu}_{\cdot,j}\|^{-1} \sum_{k=m}^{\infty} \hat{\mu}_{k,j} \hat{\xi}_{k,j} \right)^2 \\ &\leq (2\epsilon)^2 \mathbf{E}_f \|\hat{\mu}_{\cdot,j}\|^2 \ln^2 \left\{ N^2 \max_{i,l} \mathbf{E}_f \exp \left( \|\mu_{\cdot,i,l}\|^{-1} \sum_{k=m}^{\infty} \mu_{k,i,l} \xi_{k,i,l} \right) + e \right\} \\ &\leq (2\epsilon)^2 \ln^2 \{e(N^2 + 1)\} \mathbf{E}_f \|\hat{\mu}_{\cdot,j}\|^2. \end{aligned}$$

Hence, according to (3.4) with  $\gamma = \ln \{e(N^2 + 1)\}$  we have

$$\begin{aligned} |R_{3,j}| &\leq 2\epsilon \ln \{e(N^2 + 1)\} (\mathbf{E}_f \|\hat{\mu}_{\cdot,j}\|^2)^{1/2} \\ &\leq \ln^{-1} \{e(N^2 + 1)\} \mathbf{E}_f \|\hat{\mu}_{\cdot,j}\|^2 + \epsilon^2 \ln^3 \{e(N^2 + 1)\} \\ &= o(1) \mathbf{E}_f \sum_{k=m}^{\infty} (1 - \hat{h}_{k,j})^2 \hat{\theta}_{k,j}^2 + O(\epsilon^2 \ln^3 \epsilon^{-2}) \\ &= o(1) \mathbf{E}_f \hat{L}_{\epsilon,j}(f) + O\{\epsilon^2 \ln^3 (\epsilon^{-2})\}. \end{aligned} \quad \square$$

**Proof of (2.25).** In this section  $j = 0, \dots, S$  is arbitrary but fixed. Obviously the expectation of  $\hat{L}_{\epsilon,j}(f)$  is equal to

$$\mathbf{E}_f \sum_{k=m}^{\infty} (1 - h_{k,j} + h_{k,j} - \hat{h}_{k,j})^2 \hat{\theta}_{k,j}^2 + \epsilon^2 (h_{k,j} - h_{k,j} + \hat{h}_{k,j})^2.$$

For arbitrary  $0 < \gamma < 1$  the simple inequality  $(a + b)^2 \leq (1 - \gamma)^{-1} a^2 + \gamma^{-1} b^2$ ,  $a, b \in \mathbf{R}$ , holds. Combining this inequality and the fact that the squared Fourier coefficient  $\hat{\theta}_{k,j}^2$  is bounded by  $\|f\|^2$  we bound  $\mathbf{E}_f \hat{L}_{\epsilon,j}(f)$  by

$$\sum_{k=m}^{\infty} [(1 - \gamma)^{-1} \mathbf{E}_f \{(1 - h_{k,j})^2 \hat{\theta}_{k,j}^2 + \epsilon^2 h_{k,j}^2\} + \gamma^{-1} (\|f\|^2 + \epsilon^2) \mathbf{E}_f (\hat{h}_{k,j} - h_{k,j})^2].$$

Rewrite this last expression as

$$(1 - \gamma)^{-1} \{L_{c,j}(f) + R_1\} + \gamma^{-1} R_2, \quad (3.5)$$

with

$$R_1 = \sum_{k=m}^{\infty} (1 - h_{k,j})^2 (\mathbf{E}_f \hat{\theta}_{k,j}^2 - \theta_{k,j}^2), \quad R_2 = (\|f\|^2 + \epsilon^2) \sum_{k=m}^{\infty} \mathbf{E}_f (\hat{h}_{k,j} - h_{k,j})^2.$$

Observe that the desired term  $L_{c,j}(f)$  turns up in (3.5), besides two (remainder) terms  $R_1$  and  $R_2$  that are bounded.

Because  $\mathbf{E}_f \hat{\theta}_{k,j}^2 - \theta_{k,j}^2 = \mathbf{E}_f (\hat{\theta}_{k,j} - \theta_{k,j})^2 + 2\theta_{k,j}(\mathbf{E}_f \hat{\theta}_{k,j} - \theta_{k,j})$  and  $h_{k,j}^2 - 2h_{k,j} \leq 0$  for all  $k$  the first remaining term  $R_1$  obviously does not exceed

$$\sum_{k=m}^{\infty} (\mathbf{E}_f \hat{\theta}_{k,j}^2 - \theta_{k,j}^2) + 2 \sum_{k=m}^{\infty} (h_{k,j}^2 - 2h_{k,j}) \theta_{k,j} (\mathbf{E}_f \hat{\theta}_{k,j} - \theta_{k,j}),$$

which we rewrite as

$$\sum_{k=0}^{m-1} (\theta_{k,j}^2 - \mathbf{E}_f \hat{\theta}_{k,j}^2) + 2 \sum_{k=m}^{\infty} (h_{k,j}^2 - 2h_{k,j}) \theta_{k,j} (\mathbf{E}_f \hat{\theta}_{k,j} - \theta_{k,j}), \quad (3.6)$$

as both  $\sum_{k \geq 0} \hat{\theta}_{k,j}^2$  and  $\sum_{k \geq 0} \theta_{k,j}^2$  equal  $\|f\|^2$ . In order to estimate further we establish some approximations of the differences  $|\hat{\theta}_{k,j} - \theta_{k,j}|$ .

Recall that the jump points  $b_j$  are at least a distance  $\kappa$  from each other. Moreover  $\hat{b}_j$  differs from  $b_j$  by more than  $2h$  only with a small probability (cf. (2.21) and (2.17)). Therefore we can assume that the estimated jump points  $\hat{b}_j$  are also separated from each other at least by  $\kappa/2$ .

In view of the definitions of the functions  $\varphi_{k,j}$  and  $\hat{\varphi}_{k,j}$  on the subintervals  $[b_j, b_{j+1}]$  and  $[\hat{b}_j, \hat{b}_{j+1}]$ , respectively, and Theorem 2.2 we deduce that there exists a constant  $C_1$  such that for  $j = 0, \dots, S$  and  $t \in [\max(b_j, \hat{b}_j), \min(b_{j+1}, \hat{b}_{j+1})]$  the following approximations hold:

$$|\hat{\varphi}_{k,j}(t) - \varphi_{k,j}(t)| \leq C_1 (|\hat{b}_{j+1} - b_{j+1}| + |\hat{b}_j - b_j|), \quad k = 0, \dots, m-1,$$

and

$$|\hat{\varphi}_{k,j}(t) - \varphi_{k,j}(t)| \leq C_1 k (|\hat{b}_{j+1} - b_{j+1}| + |\hat{b}_j - b_j|), \quad k = m, \dots$$

On the intervals  $[\min(\hat{b}_j, b_j), \max(\hat{b}_j, b_j)]$  and  $[\min(\hat{b}_{j+1}, b_{j+1}), \max(\hat{b}_{j+1}, b_{j+1})]$  either  $\hat{\varphi}_{k,j}$  or  $\varphi_{k,j}$  is zero and the other is bounded. This implies that there exists a constant  $C_2$  such that

$$|\hat{\theta}_{k,j} - \theta_{k,j}| \leq C_2 \sup_{j=0, \dots, S} |\hat{b}_j - b_j|, \quad k = 0, \dots, m-1 \quad (3.7)$$

and

$$|\hat{\theta}_{k,j} - \theta_{k,j}| \leq C_2 k \sup_{j=0, \dots, S} |\hat{b}_j - b_j|, \quad k = m, \dots \quad (3.8)$$



Denote  $\sup_{j=0,\dots,S} |\hat{b}_j - b_j|$  by  $D$ . Given (3.7) and (3.8) we return to the estimation of  $R_1$ . The first term of (3.6) equals

$$\sum_{k=0}^{m-1} \mathbf{E}_f(\theta_{k,j} - \hat{\theta}_{k,j})^2 + 2\theta_{k,j}(\theta_{k,j} - \mathbf{E}_f \hat{\theta}_{k,j})$$

and is therefore bounded by  $C_2 m(C_2 \mathbf{E}_f D^2 + 2\|f\| \mathbf{E}_f D)$ . As  $|h(h-2)| \leq 2h$  the second term of (3.6) is bounded by  $4C_2 \mathbf{E}_f D \sum_{k \geq m} h_{k,j} k |\theta_{k,j}|$ . According to condition (A2.3) and (2.6) the sum  $\sum_{k \geq m} k^2 \theta_{k,j}^2$  is bounded. Furthermore Corollary 2.4 gives a bound for  $\mathbf{E}_f D^2$ . Hence for some constant  $C_3$  we have, by the Cauchy–Schwarz inequality and (3.4) with  $\gamma = \epsilon^2 (\ln \epsilon^{-2})^{-1}$  that

$$\begin{aligned} R_1 &\leq C_2 \mathbf{E}_f D \left( 4 \sum_{k=m}^{\infty} h_{k,j} k |\theta_{k,j}| + 2m \|f\| \right) + m C_2^2 \mathbf{E}_f D^2 \\ &\leq C_3 \|h_j\| \left( \sum_{k=m}^{\infty} k^2 \theta_{k,j}^2 \right)^{1/2} (\mathbf{E}_f D^2)^{1/2} + O(\mathbf{E}_f D) \\ &\leq C_3 \left( (\ln \epsilon^{-2})^{-1} \epsilon^2 \sum_{k=m}^{\infty} h_{k,j}^2 + \epsilon^{-2} \ln \epsilon^{-2} \mathbf{E}_f D^2 \right) + O(\mathbf{E}_f D) \\ &= o(1) L_{\epsilon,j}(f) + O\{\epsilon^2 \ln^{3+2\delta}(\epsilon^{-2})\}, \quad \epsilon \rightarrow 0. \end{aligned} \tag{3.9}$$

It remains to estimate  $R_2$ . Below we shall see that there exists a constant  $C_4$  such that the coefficients  $h_{k,j}$  vanish for  $k > C_4 \epsilon^{-2/(2m+1)}$  (cf. (3.11) and (3.13)). Furthermore from (2.6) and (2.10) it is clear that the tapering coefficients  $h_{k,j}$  behave well in the sense that we can assume that  $h_{k,j}$  is Lipschitz with respect to  $b_{j+1} - b_j$ . Hence for  $\epsilon$  tending to zero we have

$$\begin{aligned} R_2 &= (\|f\| + \epsilon^2) \sum_{k=m}^{C_4 \epsilon^{-2/(2m+1)}} \mathbf{E}_f (\hat{h}_{k,j} - h_{k,j})^2 \\ &= O(\epsilon^{-2/(2m+1)} \mathbf{E}_f D^2) = O\{\epsilon^{2+4m/(2m+1)} \ln^{2+2\delta}(\epsilon^{-2})\}. \end{aligned} \tag{3.10}$$

Substituting (3.9), (3.10) and  $\gamma = \epsilon$  in (3.5) we obtain (2.25).  $\square$

**Proof of (2.15).** According to (2.6), the number of non-zero summations in (2.11) is finite. Denote these numbers by  $N_j = N_j(c_j)$ . Note that the solutions  $c_j = c_j(\epsilon)$  of (2.11) tend to zero, as  $\epsilon$  does. Indeed if  $c_j(\epsilon)$  stays away from zero, the same would happen to the left-hand sides of (2.11) while their respective right-hand sides would tend to zero. Therefore, again according to (2.6),

$$N_j = \frac{b_{j+1} - b_j}{\pi} c_j^{-1/m} \{1 + o(1)\}, \quad \epsilon \rightarrow 0, \tag{3.11}$$

and using this, together with the asymptotic relation (for  $j = 0, \dots, S$  and  $\alpha > 0$ ),

$$\sum_{k=m}^{N_j} \lambda_{k,j}^\alpha = \left( \frac{\pi}{b_{j+1} - b_j} \right)^{2m\alpha} \frac{N_j^{2m\alpha+1}}{2m\alpha+1} \{1 + o(1)\}, \quad N_j \rightarrow \infty, \quad (3.12)$$

(2.11) become

$$c_j^{(2m+1)/m} Q = \frac{m\epsilon^2}{\pi(m+1)(2m+1)} \{1 + o(1)\}, \quad \epsilon \rightarrow 0.$$

Note that the  $c_j$  asymptotically do not depend on  $j$ , i.e.,  $c_j = c\{1 + o(1)\}$  where

$$c = c(\epsilon) = \left( \frac{m\epsilon^2}{\pi Q(m+1)(2m+1)} \right)^{m/(2m+1)} \quad (3.13)$$

and, as the calculations below show, we can just substitute this value of  $c$  into (2.15). Indeed with such a choice of  $c$ , we have, according to (3.11)–(3.13), for  $\epsilon$  tending to zero,

$$\begin{aligned} \sum_{j=0}^S \left( \epsilon^2 \sum_{k=m}^{\infty} h_{k,j}^2 + c_j^2 Q_j \right) &= \sum_{j=0}^S \left( \epsilon^2 \sum_{k=m}^{N_j} (1 - c\lambda_{k,j}^{1/2})^2 + c^2 Q_j \right) \\ &= \epsilon^2 c^{-1/m} \frac{2m^2}{\pi(m+1)(2m+1)} \sum_{j=0}^S (b_{j+1} - b_j) + c^2 Q \\ &= \gamma(m, Q) \epsilon^{4m/(2m+1)} \{1 + o(1)\}. \end{aligned} \quad \square$$

## Acknowledgements

I thank B. Levit for proposing this problem and for his supervision during this work, and Y. Golubev for stimulating discussions and his indispensable help with the proofs. Moreover I wish to thank H. Duistermaat and R. Gill for their helpful comments and useful discussions.

## References

- Belitser, E. and Levit, B. (1996) Asymptotically minimax nonparametric regression in  $L_2$ . *Statistics*, **28**, 105–122.
- Birgé, L. (1983) Approximation dans les espaces métriques et théorie de l'estimation. *Z. Wahrscheinlich Keitstheorie Verw. Geb.*, **65**, 181–237.
- Brown, L. and Low, M. (1996) Asymptotic equivalence of nonparametric regression and white noise. *Ann. Statist.*, **24**, 2384–2398.
- Coddington, E. and Levinson, N. (1955) *Theory of Ordinary Differential Equations*. New York: McGraw-Hill.
- Donoho, D. (1994) Asymptotic risk for sup-norm loss: solution via optimal recovery. *Probab. Theory Related Fields*, **99**, 145–170.

- Duistermaat, J. (1995) The Sturm–Liouville problem for the operator  $(-d^2/dx^2)^m$ , with Neumann or Dirichlet boundary conditions. Technical Report 899, Department of Mathematics, University of Utrecht.
- Dunford, N. and Schwartz, J. (1971) *Linear operators*, Part III: *Spectral Operators*. New York: Wiley Interscience.
- Efroimovich, S.Y. (1994) On nonparametric regression for i.i.d. observations in general setting. Technical Report, University of New Mexico.
- Golubev, Y.K., Levit, B. and Tsybakov, A. (1995) Asymptotically efficient estimation of analytic functions in Gaussian noise. Technical Report 894, Department of Mathematics, University of Utrecht.
- Golubev, Y.K. and Nussbaum, M. (1990) A risk bound in Sobolev class regression. *Ann. Statist.*, **18**, 758–778.
- Hall, P. and Patil, P. (1995) Formulae for mean integrated squared error of nonlinear wavelet-based density estimators. *Ann. Statist.*, **23**, 905–928.
- Ibragimov, I. and Hasminskii, R. (1981) *Statistical Estimation: Asymptotic Theory*. Berlin: Springer-Verlag.
- Korostelev, A. (1987) On minimax estimation of a discontinuous signal. *Theory Probab. Appl.*, **32**, 727–730.
- Korostelev, A. (1994) Exact asymptotic minimax estimate of a nonparametric regression in the uniform norm. *Theory Probab. Appl.*, **38**, 737–743.
- Müller, H. (1992) Change-points in nonparametric regression analysis. *Ann. Statist.*, **20**, 737–761.
- Neumark, M. (1967) *Lineare differential Operatoren*. Berlin: Akademie-Verlag.
- Nussbaum, M. (1985) Spline smoothing in regression models and asymptotic efficiency in  $L_2$ . *Ann. Statist.*, **13**, 984–997.
- Pinsker, M. (1980) Optimal filtration of square-integrable signals of Gaussian noise. *Prob. Inf. Trans.*, **16**, 120–133.
- Speckman, P. (1985) Spline smoothing and optimal rates of convergence in nonparametric regression models. *Ann. Statist.*, **13**, 970–983.
- Stone, C. (1982) Optimal global rates of convergence for nonparametric regression. *Ann. Statist.*, **10**, 1040–1053.
- Wang, Y. (1995) Jump and sharp cusp detection by wavelets. *Biometrika*, **82**, 385–397.

Received November 1995 and revised January 1997