Smoothing Splines and Shape Restrictions

E. MAMMEN

Universität Heidelberg

C. THOMAS-AGNAN

Université des Sciences Sociales

ABSTRACT. Constrained smoothing splines are discussed under order restrictions on the shape of the function m. We consider shape constraints of the type $m^{(r)} \geq 0$, i.e. positivity, monotonicity, convexity, …. (Here for an integer $r \geq 0$, $m^{(r)}$ denotes the rth derivative of m.) The paper contains three results: (1) constrained smoothing splines achieve optimal rates in shape restricted Sobolev classes; (2) they are equivalent to two step procedures of the following type: (a) in a first step the unconstrained smoothing spline is calculated; (b) in a second step the unconstrained smoothing spline is "projected" onto the constrained set. The projection is calculated with respect to a Sobolev-type norm; this result can be used for two purposes, it may motivate new algorithmic approaches and it helps to understand the form of the estimator and its asymptotic properties; (3) the infinite number of constraints can be replaced by a finite number with only a small loss of accuracy, this is discussed for estimation of a convex function.

Key words: convexity, monotonicity, rates of convergence, shape restrictions, smoothing splines

1. Introduction

In this paper, constrained smoothing splines are discussed under restrictions on the shape of the underlying function m of the form $m^{(r)} \geq 0$ (or $m^{(r)} \leq 0$). [Here for an integer $r \geq 0$, $m^{(r)}$ denotes the rth derivative of m.] In particular, this includes positivity, monotonicity and convexity constraints. Shape restrictions of this type arise in many applications. The constraints may be given by the context, e.g. convexity for production functions or Engel curves, monotonicity of failure rates. Often, inference on the qualitative shape of a curve may be based on the comparison of constrained and unconstrained estimators. An overview on curve estimation under shape restrictions can be found in Delecroix & Thomas-Agnan (1997). Constrained spline estimates are considered in Villa-lobos & Wahba (1987) and Utreras (1985). For a discussion of unconstrained splines, see e.g. Eubank (1988) and Wahba (1990).

We consider the regression model:

$$Y_i = m_0(x_i) + \epsilon_i, \tag{1}$$

where $m_0: [0, 1] \to \mathbb{R}$ is an unknown regression function, $x_i \in [0, 1]$ are deterministic design points $[x_1 \leq \cdots \leq x_n]$, ϵ_i are independent errors with expectation $E(\epsilon_i) = 0$ for $i = 1, \ldots, n$.

Under the constraint $m^{(r)}(x) \geq 0$ for $x \in [0, 1]$, estimation of m may be done by the constrained smoothing spline $\hat{m}$ of order k. For an integer $k \geq 1$, a constant $0 < D \leq \infty$ and a sequence of penalty weights $\lambda_n > 0$ this estimate is defined as the solution of the optimization problem:

$$\hat{m}_{n,D}^{CS} = \arg \min_{m \in \mathcal{M}_{k,r}(D)} \left[\frac{1}{n} \sum_{i=1}^{n} (Y_i - m(x_i))^2 + \lambda_n \int_0^1 m^{(k)}(x)^2 \, dx \right], \tag{2}$$

where the argmin runs over all functions m that lie in the following function class $\mathcal{M}_{k,r}(D)$:

$$\mathcal{M}_{k,r}(D) = \{m: m^{(r-1)} \text{ exists a.s. and is monotone}, |m^{(r-1)}| \leq D,$$
$$m^{(k-1)} \text{ exists and is absolutely continuous with}$$
$$\int m^{(k)}(x)^2 \, dx < \infty\} \quad \text{if } r \geq 1, \tag{3}$$

$$\mathcal{M}_{k,r}(D) = \{m: m \text{ is positive},$$
$$m^{(k-1)} \text{ exists and is absolutely continuous with}$$
$$\int m^{(k)}(x)^2 \, dx < \infty\} \quad \text{if } r = 0.$$

We write $\mathcal{M}_{k,r}$ for $\mathcal{M}_{k,r}(\infty)$. For $n > k$ the argmin in (2) is uniquely defined, see Utreras (1985). For simplicity of notation, the dependence of $\hat{m}_{n,D}^{CS}$ on r and k will not be indicated in the notation. We write $\hat{m}_n^{CS}$ for $\hat{m}_{n,\infty}^{CS}$.

The asymptotic behaviour of this estimate will be studied in the next section for different choices of k and r. It will be shown that this estimate achieves optimal rates of convergence if λ_n is chosen of an appropriate order.

Furthermore, when $k \geq r + 1$, we will show that the estimate coincides with the unconstrained smoothing spline with probability tending to one. In the case $k = r$, the differences between the constrained and unconstrained estimate do not vanish asymptotically.

In section 3, we show that the constrained smoothing spline is equivalent to the projection (with respect to a Sobolev-type norm) of the unconstrained smoothing spline onto the constrained set. This result helps to understand the asymptotic results of section 2. Furthermore, it can be used to discuss the relation of the constrained smoothing spline to a modified estimator proposed in Delecroix *et al.* (1996). Constrained smoothing splines with infinitely many constraints [like $m^{(r)}(x) \geq 0$ for all x] are difficult to compute (see Elfving & Anderson, 1988, for $k = 2$, $r \leq 2$). We will show that these constraints can be replaced by finitely many constraints without a large loss of accuracy in the calculation of $\hat{m}_n^{CS}$. Proofs of the results can be found in section 4.

2. Rates of convergence

In this section, we show that the constrained smoothing spline $\hat{m}_{n,D}^{CS}$ achieves optimal rates of convergence in constrained Sobolev classes. Our first result (proposition 1) gives the rates of the constrained smoothing spline. Our second result (proposition 2) shows that these rates cannot be improved by other estimates. It will turn out that for $k \geq r$ the optimal rates for the constrained and the unconstrained case coincide. Furthermore, for $k \leq r$, we get the same optimal rate as if only the shape restriction $m^{(r)} \geq 0$ is assumed [and no smoothness assumptions $\int_0^1 m^{(k)}(x)^2 < \infty$ are made.] For $k > r$ the constrained smoothing spline and the unconstrained smoothing spline coincides with probability tending to one if $m^r(x) \neq 0$ for all $x \in [0, 1]$. This is the content of proposition 3. The limiting case $k = r$ is considered in proposition 4 for $k = r = 2$. It will be shown that for this case there is a first order difference between the constrained smoothing spline and the unconstrained smoothing spline.

We will measure the accuracy of curve estimates by the L_2-distance and by the empirical norm:

$$\|g\|_n^2 = \frac{1}{n}\sum_{i=1}^n g^2(x_i).$$

We will assume that the underlying true regression function m_0 lies in the restricted Sobolev class $\mathcal{M}_{k,r}$, see (3). For the error distributions we suppose that they have (uniform) subexponential tails, i.e. there exist constants $C < +\infty$ and $t_0 > 0$ with

$$\mathrm{E}(\exp t|\epsilon_i|) < C \quad \text{for } 0 < t < t_0,\ 1 \leq i \leq n,\ n \geq 1. \tag{4}$$

Proposition 1

For an integer $r \geq 0$ and an integer $k \geq 1$, assume model (1.1) with m_0 in $\mathcal{M}_{k,r}$ and subexponential error distribution (see (4)). Put $p = \max(k, r)$. The penalty weight λ_n is assumed to be a random sequence of order $n^{-2p/(2p+1)}$ (i.e. $\lambda_n = \mathcal{O}_P(n^{-2p/(2p+1)})$ and $\lambda_n^{-1} = \mathcal{O}_P(n^{2p/(2p+1)})$).

Then, for $D < \infty$ large enough, we get:

$$\|\hat{m}_{n,D}^{CS} - m_0\|_n = \mathcal{O}_P(n^{-p/(2p+1)}) \tag{5}$$

and

$$\int_0^1 \left\{ \frac{\partial^k}{(\partial x)^k} \hat{m}_{n,D}^{CS}(x) \right\}^2 dx = \mathcal{O}_P(1). \tag{6}$$

For the case $r \leq k$, (5) and (6) hold with $\hat{m}_{n,D}^{CS}$ replaced by $\hat{m}_n^{CS}$.

This proposition can easily be shown using empirical process methods developed e.g. in van de Geer (1990). For details see section 4. Proposition 1 can be generalized to the case that the underlying regression function m_0 depends on n. Then the statement of proposition 1 remains valid if $\int_0^1 m_0^{(k)}(x)^2 dx$ and $\sup_x |m_0^{(r-1)}(x)|$ are uniformly bounded for all n. This shows that the rate $n^{-p/(2p+1)}$ is uniformly attained over classes $\mathcal{M}_{k,r}(A, D)$ [see (7), below]. For another generalization one can consider the case that shape constraints of different order are assumed at the same time (e.g. estimation of a convex monotone function). In particular, the statement of proposition 1 remains valid if the set $\mathcal{M}_{k,r}$ is replaced by $\mathcal{M}_{k,r} \cap \{m: m^{(s)}$ is monotone for $s \in I\}$, where I is a subset of $\{0, \ldots, r-2\}$. Furthermore, proposition 1 can be applied to the case of random design: $Y_i = m_0(X_i) + \epsilon_i$ with independent tuples $(X_1, \epsilon_1), \ldots, (X_n, \epsilon_n)$ where $\mathrm{E}(\epsilon_i|X_i) = 0$. For this purpose it suffices to replace assumption (4) by $\sup_{x, 1 \leq i \leq n} \mathrm{E}(\exp t|\epsilon_i|\,|\,X_i = x) < C$ (a.s.) for $0 < t < t_0$. Then the statement of proposition 1 follows for this model of random design by a simple conditioning argument.

Proposition 1 generalizes a result of Utreras (1985) where this rate of convergence has been shown for the uniform design for $k \geq r$. We show now that the rate $\mathcal{O}_P(n^{-2p/(2p+1)})$ cannot be improved. For $A > 0$ and $D > 0$, we consider constrained Sobolev balls:

$$\mathcal{M}_{k,r}(A, D) = \left\{ m \in \mathcal{M}_{k,r}(D): \int_0^1 m^{(k)}(x)^2 dx \leq A \right\}. \tag{7}$$

The optimal rate for estimation of m in $\mathcal{M}_{k,r}(A, D)$ is $n^{-p/(2p+1)}$. This follows from the following proposition and from proposition 1 [note also that proposition 1 holds for regression functions m_0 in $\mathcal{M}_{k,r}(A, D)$ that may depend on n, see the remark after proposition 1].

Proposition 2

Assume model (1) *with* $m_0 \in \mathscr{M}_{k,r}(A, D)$ *and with normal i.i.d. errors* $\epsilon_1, \ldots \epsilon_n$. *Suppose that with* $x_0 = 0$ *and* $x_{n+1} = 1$

$$\liminf_{n \to \infty} \inf_{0 \leqslant i \leqslant n} n|x_{i+1} - x_i| > 0$$

and

$$\limsup_{n \to \infty} \sup_{0 \leqslant i \leqslant n} n|x_{i+1} - x_i| < \infty.$$

Then there exists no estimate with faster rate than $n^{-p/(2p+1)}$, *i.e.*

$$\liminf_{n \to \infty} n^{2p/(2p+1)} \inf_{\tilde{m}_n} \sup_{m_0 \in \mathscr{M}_{k,r}(A,D)} E_{m_0} \|\tilde{m}_n - m_0\|_n^2 > 0 \tag{8}$$

and

$$\liminf_{n \to \infty} n^{2p/(2p+1)} \inf_{\tilde{m}_n} \sup_{m_0 \in \mathscr{M}_{k,r}(A,D)} D_{m_0} \int_0^1 |\tilde{m}_n(x) - m_0(x)|^2 \, dx > 0, \tag{9}$$

where the infimum runs over all curve estimates $\tilde{m}_n$ *based on* $Y_1, \ldots Y_n$.

The rate of unconstrained smoothing splines is $\mathscr{O}_P(n^{-2k/(2k+1)})$. Propositions 1 and 2 imply that no faster rates can be achieved by adding shape constraints as long as $r \leqslant k$. Furthermore, for $r \geqslant k$, the constrained smoothing spline achieves the same rate as a shape restricted least squares estimate (rates of shape restricted least squares estimates have been considered in Mammen, 1991). Here, no faster rate is achieved by the additional smoothness assumption $\int m_0^{(k)}(x)^2 \, dx < \infty$.

For $r < k$, shape restrictions have a negligible influence. The following proposition states that constrained and unconstrained smoothing splines coincide with probability tending to one for the case that $r < k$ and $m_0^{(r)}(x) \neq 0$.

Proposition 3

Suppose $r < k$ *and assume model* (1) *where the regression function* m_0 *fulfills that* $\int m_0^{(k)}(x)^2 \, dx$ *is finite and that* $m_0^{(r)}(x) \neq 0$ *for* $x \in [0, 1]$. *Furthermore, it is assumed that* $\sup_{1 \leqslant i < n}(x_{i+1} - x_i) = o(1)$ *and that errors have subexponential tails (see* (4)). *Then, if* λ_n *is a random sequence of order* $n^{-2k/(2k+1)}$, *we get:*

$$P(\hat{m}_n^{CS}(x) = \hat{m}_n^S(x) \forall x \in [0, 1]) \to 1.$$

Here $\hat{m}_n^S$ *is the unconstrained smoothing spline:*

$$\hat{m}_n^S = \arg \min_{m \in \mathscr{H}_k} \left[\frac{1}{n} \sum_{i=1}^n (y_i - m(x_i))^2 + \lambda_n \int_0^1 m^{(k)}(x)^2 \, dx \right], \tag{10}$$

where $\mathscr{H}_k = \{m: m^{(k-1)}$ *exists and is absolutely continuous with* $\int_0^1 m^{(k)}(x)^2 \, dx < \infty\}$.

We consider now the case $k = r$. We will show that, if $k = r = 2$, there is with positive probability a non-negligible difference between the constrained and the unconstrained smoothing spline. The proof of this result makes use of the asymptotic representation of smoothing splines as linear kernel smoothers for $k = 2$ given in Silverman (1982). We conjecture that our result holds also for other choices of $k = r$. For a proof of this conjecture generalizations of the results in Silverman (1982) for other choices of k are required. A discussion of such generalizations can be found in Messer (1991) and Nychka (1995).

Proposition 4

Suppose $r = k = 2$ and assume model (1) *with Gaussian i.i.d. errors. The empirical distribution function F_n of the design points $x_1, \ldots, x_n$ is assumed to converge to a distribution function F:*

$$n^{1/5} \sup_{x \in [0,1]} |F_n(x) - F(x)| \to 0.$$

The derivative f of F is assumed to be bounded away from 0 and to have an absolutely bounded derivative. Then, if λ_n is a deterministic sequence of order $n^{-4/5}$ and $D \le \infty$, there exists $\delta > 0$ such that

$$\liminf_{n \to \infty} P(\|\hat{m}_{n,D}^{CS} - \hat{M}_n^S\|_n > \delta n^{-2/5}) > 0.$$

3. Modifications of constrained smoothing splines

In this section we show that for the constrained smoothing spline $\hat{m}_n^{CS}$ the following holds

$$\hat{m}_n^{CS} = \arg \min_{m \in \mathscr{M}_{k,r}} \left[\|m(x) - \hat{m}_n^S(x)\|_n^2 + \lambda_n \int_0^1 \left(m^{(k)}(x) - \frac{\partial^k}{(\partial x)^k} \hat{m}_n^S(x) \right)^2 dx \right]. \tag{11}$$

The estimate $\hat{m}_n^S$ is the unconstrained smoothing spline, see (10). The equivalence (11) is stated in the following proposition 5.

Proposition 5

The relation (11) *holds.*

Equation (11) has the following interpretation. The estimate $\hat{m}_n^{CS}$ is a two steps estimate:

1. In a first step the unconstrained smoothing spline $\hat{m}_n^S$ (see (10)) is calculated.
2. In a second step this estimate is "projected" onto the constrained set. The projection is calculated with respect to the Sobolev-type norm $\|g\|_n^2 + \lambda_n \int_0^1 \{g^{(k)}(x)\}^2 dx$, see (11).

For a similar result on a general class of constrained smoothers, see Mammen *et al.* (1998).
In Delecroix *et al.* (1996), another two steps estimate $\tilde{m}_n^{CS}$ has been proposed:

$$\tilde{m}_n^{CS} = \arg \min_{m \in \mathscr{M}_{k,r}} \left[\int_0^1 \{m(x) - \hat{m}_n^S(x)\}^2 dx + \lambda_n \int_0^1 \left\{ m^{(k)}(x) - \frac{\partial^k}{(\partial x)^k} \hat{m}_n^S(x) \right\}^2 dx \right].$$

[To be more precise, in Delecroix *et al.* (1996), a discretized version of the constraints was used for computational simplifications]. Our proposition 5 shows now that $\tilde{m}_n^{CS}$ is similarly defined as $\hat{m}_n^{CS}$, the only difference being that the integrated norm $\int_0^1 g^2(x) dx$ is replaced by the empirical norm $\|g\|_n^2$. This difference is asymptotically negligible for equidistant design, as is shown in the following corollary.

Corollary 1

Suppose that $x_i = (i - 1/2/n)$, that $k \ge 2$, and that the assumptions of proposition 1 hold, then we get:

$$\|\hat{m}_n^{CS} - \tilde{m}_n^{CS}\|_n^2 = \mathcal{O}_P(n^{-6k/(2k+1)}) \tag{12}$$

and

$$\int_0^1 (\hat{m}_n^{CS}(x) - \tilde{m}_n^{CS}(x))^2 \, dx = \mathcal{O}_P(n^{-6k/(2k+1)}). \tag{13}$$

Computation of constrained estimates can be speeded up by restricting the constraints to a discrete set. For $k = r = 2$, we consider the following discretized modification of $\hat{m}_n^{CS}$. For a grid $T_n = \{t_1, \ldots t_s\} \subset [0, 1]$, with $t_1 = 0$, $t_s = 1$, we define

$$\hat{m}_n^{RCS} = \arg\min_m \left[\frac{1}{n} \sum_{i=1}^n (Y_i - m(x_i))^2 + \lambda_n \int_0^1 m^{(k)}(x)^2 \, dx \right],$$

where the argmin runs over all functions m whose restrictions to T_n are convex. Arguing as in the proof of proposition 5 one can show that

$$\hat{m}_n^{RCS} = \arg\min_m \left[\|m(x) - \hat{m}_n^S(x)\|_n^2 + \lambda_n \int_0^1 (m^{(k)}(x) - \frac{\partial^k}{(\partial x)^k} \hat{m}_n^S(x))^2 \, dx \right],$$

where again the argmin runs over all functions m whose restrictions to T_n are convex. The next proposition describes how far away $\hat{m}_n^{RCS}$ is from the class of functions that are convex on the whole interval $[0, 1]$.

Proposition 6

Suppose the conditions of proposition 1, that $k \geqslant 2$ and that for a δ_n with $\delta_n \to 0$, it holds that $\sup_i |t_{i+1} - t_i| = \mathcal{O}(\delta_n)$. Then we get that

$$\inf_{m \text{ convex}} \int_0^1 \{m(x) - \hat{m}_n^{RCS}(x)\}^2 \, dx = \mathcal{O}_P(\delta_n^4).$$

4. Proofs

Proof of proposition 1. The proposition can be proved similiarly as th. 6.2 in van de Geer (1990), th. 5 in Mammen & van de Geer (1997a), and lem. 3.1 in Mammen & van de Geer (1997b). We give here the basic idea. Denote by $\langle, \rangle_n$ the scalar product corresponding to the norm $\| \|_n$, i.e. $\langle g, h \rangle_n = n^{-1} \sum_{i=1}^n g(x_i) h(x_i)$. We write $\mathscr{P}_{s,n}^\perp$ for the orthogonal complement of the set of all polynomials of degree $(s - 1)$ [with respect to the scalar product $\langle, \rangle_n$]. First note that for

$$\mathscr{M}^0 = \{m: m^{(r-1)} \text{monotone}, \ |m^{(r-1)}| \leqslant 1\} \cap \mathscr{P}_{r,n}^\perp$$

and

$$\mathscr{M}^1 = \{m: \int_0^1 m^{(k)}(x)^2 \, dx \leqslant 1\} \cap \mathscr{P}_{k,n}^\perp$$

we have the following bounds for entropies with bracketing:

$$\log N_{2,B}(\delta, \|.\|_n, \mathscr{M}^0) \leqslant C_0 \delta^{-1/r}, \tag{14}$$

$$\log N_{2,B}(\delta, \|.\|_n, \mathscr{M}^1) \leqslant C_1 \delta^{-1/k}, \tag{15}$$

where C_0 and C_1 are positive constants and $r, k \geqslant 1$. $N_{2,B}(\delta, \|.\|_n, \mathscr{M}^i)$ denotes the smallest number N of pairs $(g_{1,j}, t_{2,g})$: $j = 1, \ldots, N$ with (i) $\|g_{1,j} - g_{2,j}\|_n \leqslant \delta$, (ii) $g_{1,j}, g_{2,j} \in \mathscr{M}^i$, (iii) for every $g \in \mathscr{M}^i$ there exists a j with $g_{1,j} \leqslant g \leqslant g_{2,j}$. Equations (14) and (15) follow from Birman & Solomjak (1967), see van de Geer (1990, 1993) and Mammen (1991).

We define now $\mathscr{M}$ to be the intersection of $\mathscr{M}^0$ and $\mathscr{M}^1$ if $r > k$ and $\mathscr{M} = \mathscr{M}^1$ if $r \leqslant k$. Then we have

$$\log(N_{2,B}(\delta, \|\cdot\|_n, \mathscr{M})) \leqslant C_2 \delta^{-1/p} \tag{16}$$

for a $C_2 > 0$. Inequality (16) implies:

$$\sup_{m \in \mathscr{M}} \frac{|n^{-1/2} \sum_{i=1}^n m(x_i) \epsilon_i|}{[\min\{\|m\|_n, n^{-[p-2]/[2p]}\}]^{2p/[2p-1]}} = \mathscr{O}_P(1) \tag{17}$$

[For errors with subGaussian tails this has been stated in lem. 3.5 in van de Geer (1990). For errors with subexponential tails this follows from an additional application of a result in Birgé & Massart (1993), see van de Geer (1995).] For the proof of equations (5) and (6) one proceeds as in Mammen & van de Geer (1997a, b).

Proof of proposition 2. We choose $I = I_n$ as the largest integer $\leqslant n^{1/(2p+1)}$. For $i = 1, \ldots, I$, we consider the intervals: $R_{i,n} = [(i-1)/I_n, i/I_n]$. We choose a function $g: [0, 1] \to \mathbb{R}^+$ which is p times continuously differentiable and with $g^{(s)}(0) = g^{(s)}(1) = 0$ for $s = 0, \ldots, p$ and $\int_0^1 g(x)^2 \, dx > 0$. For $\theta \in \{0, 1\}^I$ we put

$$m_\theta(x) = ax^r + b\theta_i n^{-p/(2p+1)} g\{I[x - (i-1)/I_n]\}$$

for $x \in R_{i,n}$, where a, b are chosen such that $m_\theta \in \mathscr{M}_{k,r}(A, D)$ for $\theta \in \{0, 1\}^I$. For the proof of (9) one notes first that

$$\inf_{\tilde{m}_n} \sup_{m_0 \in \mathscr{M}_{k,r}(A,D)} E_{m_0} \int_0^1 |\tilde{m}_n(x) - m_0(x)|^2 \, dx \geqslant \inf_{\tilde{m}_n} \sup_{\theta \in \{0,1\}^I} E_{m_\theta} \int_0^1 |\tilde{m}_n(x) - m_\theta(x)|^2 \, dx, \tag{18}$$

where the infimum runs over all curve estimates $\tilde{m}_n$ based on $Y_1, \ldots Y_n$. The right hand side of (18) can be bounded from below by standard techniques based on Assouad's lemma. We refer to sect. 2.6 and 2.7 in Korostelev & Tsybakov (1993) where this has been done for Hölder function classes. This shows (9). The proof of (8) follows analogously.

Proof of proposition 3. It suffices to show that

$$P\left(\frac{\partial^{r-1}}{(\partial x)^{r-1}} \hat{m}_n^S \text{ is monotone}\right) \to 1.$$

Because under our assumptions $m_0^{(r)}$ is continuous and therefore bounded away from 0, this follows from

$$\sup_x \left| \frac{\partial^r}{(\partial x)^r} \hat{m}_n^S(x) - m_0^{(r)}(x) \right| = o_P(1). \tag{19}$$

It remains to show (19). From proposition 1, we know that

$$\|\hat{m}_n^S - m_0\|_n = o_P(1)$$

and

$$\int \left\{ \frac{\partial^k}{(\partial x)^k} \hat{m}_n^S(x) \right\}^2 \, dx = \mathscr{O}_P(1).$$

Because of $\int m_0^{(k)}(x)^2 \, dx < \infty$ and $\sup_i(x_{i+1} - x_i) = o(1)$ this implies $\int (\hat{m}_n^S(x) - m_0(x))^2 \, dx = o_P(1)$. The interpolation inequality (see Agmon, 1965) gives for $0 < \theta < 1$ with a constant $C > 0$ for $1 \leqslant q \leqslant k$:

$$\int \left\{ \frac{\partial^q}{(\partial x)^q} \hat{m}_n^S(x) - m_0^{(q)}(x) \right\}^2 dx \geqslant C\theta^{-2q} \int \{\hat{m}_n^S(x) - m_0(x)\}^2 dx$$

$$+ C\theta^{2k-2q} \int \left\{ \frac{\partial^k}{(\partial x)^k} \hat{m}_n^S(x) - m_0^{(k)}(x) \right\}^2 dx. \quad (20)$$

Application with $q = r$ and $q = r + 1$ gives for $\Delta(x) = (\partial^r/(\partial x)^r)\hat{m}_n^S(x) - m_0^{(r)}(x)$ that $\int |\Delta'(x)|^2 dx = \mathcal{O}_P(1)$ and $\int \Delta(x)^2 dx = o_P(1)$. Because of $\int |\Delta'(x)|^2 dx = \mathcal{O}_P(1)$, application of an embedding theorem (see Adams, 1975, p. 97) gives

$$\sup_{x,y} |\Delta(x) - \Delta(y)|/|x - y|^{1/2} = \mathcal{O}_P(1).$$

This equality and $\int \Delta(x)^2 dx = o_P(1)$ implies $\sup |\Delta(x)| = o_P(1)$. This shows (19).

Proof of proposition 4. For simplicity we consider only the case var $\epsilon_i = 1$, $\lambda_n = n^{-4/5}$ and $D = \infty$. For the proof we make use of the following lemma.

Lemma 1

For a subset $\mathscr{X}$ of $\mathbb{R}$ and a point $x_0 \in \mathscr{X}$ we put $\mathscr{X}^- = \{x \in \mathscr{X}: x \leqslant x_0\}$ and $\mathscr{X}^+ = \{x \in \mathscr{X}: x > x_0\}$. We consider a Hilbert space $\mathscr{H}$ of functions $h: \mathscr{X} \to \mathbb{R}$ with norm $\|h\|^2 = \int_{\mathscr{X}} h(x)^2 dx$ and scalar product $\langle h_1, h_2 \rangle = \int_{\mathscr{X}} h_1(x)h_2(x) dx$. For a function $g \in \mathscr{H}$ we define:

$$g_I = \arg\min\{\|h - g\|: h \in \mathscr{H}, h \text{ increasing}\},$$

$$g_{PC} = \arg\min\{\|h - g\|: h \in \mathscr{H}, h \text{ is constant on } \mathscr{X}^- \text{ and on } \mathscr{X}^+\},$$

$$g_{PCI} = \arg\min\{\|h - g_{PC}\|: h \in \mathscr{H}, h \text{ increasing}\}.$$

With these definitions the following holds

$$\|g - g_I\| \geqslant \|g_{PC} - g_{PCI}\|. \quad (21)$$

The proof of lemma 1 will be given after the proof of proposition 4. For the proof of proposition 4 we apply the lemma for $1 \leqslant j \leqslant 0.5n^{1/5}$ with

$$\mathscr{X}^- = \mathscr{X}_j^- = [0.25 + (j-1)n^{-1/5}, 0.25 + (j-1/2)n^{-1/5}],$$

$$\mathscr{X}^+ = \mathscr{X}_j^+ = (0.25 + (j-1/2)n^{-1/5}, 0.25 + jn^{-1/5}],$$

$$\mathscr{X} = \mathscr{X}_j = \mathscr{X}_j^- \cup \mathscr{X}_j^+,$$

norm $\|h\|^2 = \int_{\mathscr{X}_j} h(x)^2 dx$, and $g = g_j$ equal to $(\partial/\partial x)\hat{m}_n^S(x)$ restricted to $\mathscr{X}_j$. Lemma 1 implies that

$$\int_0^1 \left\{ \frac{\partial}{\partial x} \hat{m}_n^S(x) - \frac{\partial}{\partial x} \hat{m}_n^{CS}(x) \right\}^2 dx \geqslant \sum_{1 \leqslant j \leqslant 0.5n^{1/5}} \int_{\mathscr{X}_j} \left\{ \frac{\partial}{\partial x} \hat{m}_n^S(x) - \frac{\partial}{\partial x} \hat{m}_n^{CS}(x) \right\}^2 dx$$

$$\geqslant \sum_{1 \leqslant j \leqslant 0.5n^{1/5}} \int_{\mathscr{X}_j} \{g_j(x) - g_{j,I}(x)\}^2 dx$$

$$\geqslant \sum_{1 \leqslant j \leqslant 0.5n^{1/5}} \int_{\mathscr{X}_j} \{g_{j,PC}(x) - g_{j,PCI}(x)\}^2 dx$$

$$= S, \quad (22)$$

where

$$S = n^{1/5} \sum_{1 \leqslant j \leqslant 0.5 n^{1/5}} Z_{j,+}^2,$$

$$Z_j = \hat{m}_n^S(0.25 + (j-1)n^{-1/5}) + \hat{m}_n^S(0.25 + jn^{-1/5}) - 2\hat{m}_n^S(0.25 + (j-1/2)n^{-1/5}),$$

$$Z_{j,+} = Z_j \mathbf{1}(Z_j \geqslant 0).$$

We will show that for $C' > 0$ small enough

$$ES \geqslant C' n^{-2/5}. \tag{23}$$

We apply now the interpolation inequality (20). With

$$\theta^2 = \min\left\{\frac{1}{2}, \frac{R_1}{2CR_2}\right\},$$

$$R_0 = \int_0^1 \{\hat{m}_n^S(x) - \hat{m}_n^{CS}(x)\}^2 \, dx,$$

$$R_1 = \int_0^1 \left\{\frac{\partial}{\partial x}\hat{m}_n^S(x) - \frac{\partial}{\partial x}\hat{m}_n^{CS}(x)\right\}^2 \, dx,$$

$$R_2 = \int_0^1 \left\{\frac{\partial^2}{(\partial x)^2}\hat{m}_n^S(x) - \frac{\partial^2}{(\partial x)^2}\hat{m}_n^{CS}(x)\right\}^2 \, dx$$

this gives

$$R_0 \geqslant \min\left\{\frac{R_1}{4C}, \frac{R_1^2}{4C^2 R_2}\right\}.$$

The inequalities (22) and (23) and $R_2 = \mathcal{O}_P(1)$ imply the statement of proposition 4.

Proof of (23). We write $m_n^S(x) = E\hat{m}_n^S(x)$. Because spline smoothing is linear in the observations, the following holds:

$$m_n^S = \arg\min_m \left[\frac{1}{n}\sum_{i=1}^n (m_0(x_i) - m(x_i))^2 + \lambda_n \int_0^1 m''(x)^2 \, dx\right].$$

This shows

$$\int_0^1 \left\{\frac{\partial^2}{(\partial x)^2}m_n^S(x)\right\}^2 dx \leqslant \frac{1}{n\lambda_n}\sum_{i=1}^n (m_0(x_i) - m_n^S(x_i))^2 + \int_0^1 \left\{\frac{\partial^2}{(\partial x)^2}m_n^S(x)\right\}^2 dx \leqslant \rho, \tag{24}$$

where $\rho = \int_0^1 m_0''(x)^2 \, dx$. Put $\rho_j = \int_{\mathcal{X}_j}\{(\partial^2/(\partial x)^2)m_n^S(x)\}^2 \, dx$. Inequality (24) implies $\sum_{1 \leqslant j \leqslant 0.5 n^{1/5}}\rho_j \leqslant \rho$. This shows that the set $J_n = \{1 \leqslant j \leqslant 0.5 n^{1/5}: \rho_j \leqslant 4n^{-1/5}\rho\}$ has at least $0.25 \ n^{1/5} - 1$ elements. We show now that there exist positive constants C_1 and C_2 such that for $j \in J_n$

$$|EZ_j| \leqslant C_1 n^{-2/5}, \tag{25}$$

$$\operatorname{var} Z_j \geqslant C_2 n^{-4/5}. \tag{26}$$

Because Z_j has a Gaussian distribution this implies

$$\min_{j \in J_n} EZ_{j,+}^2 \geqslant C_3 n^{-4/5},$$

for $C_3 > 0$ small enough. This shows (23). It remains to prove (25), (20), and lemma 1.

Proof of (25). We get for $j \in J_n$

$$
|EZ_j| = |m_n^S(0.25 + (j-1)n^{-1/5}) + m_n^S(0.25 + jn^{-1/5})
$$

$$
- 2m_n^S(0.25 + (j - 1/2)n^{1/5})|
$$

$$
= \left| \int_{\mathscr{X}_j^+} \frac{\partial}{\partial x} m_n^S(x)\, dx - \int_{\mathscr{X}_j^-} \frac{\partial}{\partial x} m_n^S(x)\, dx \right|
$$

$$
\leqslant \int_{\mathscr{X}_j^-} \int_x^{x+0.5n^{-1/5}} \left| \frac{\partial^2}{(\partial u)^2} m_n^S(u) \right| du\, dx
$$

$$
\leqslant \int_{\mathscr{X}_j^-} \left[\int_{\mathscr{X}_j} \left| \frac{\partial^2}{(\partial u)^2} m_n^S(u) \right|^2 du \right]^{1/2} \left[\frac{1}{2} n^{-1/5} \right]^{3/2} dx
$$

$$
\leqslant \frac{1}{2} \rho_j^{1/2} n^{-3/10} \leqslant \rho^{1/2} n^{-2/5}.
$$

Proof of (26). According to th. A in Silverman (1984) we have under our conditions

$$
\hat{m}_n^S(s) = \frac{1}{n} \sum_{i=1}^n G_n(s, x_i) Y_i,
$$

with a function G_n that fulfills

$$
\sup |n^{-1/5} f(x)^{-1/4} G_n(x + n^{-1/5} f(x)^{-1/4} t, x) - \kappa(t) f(x)^{-1}| \to 0.
$$

Here for a sequence δ_n with $n^{1/5}\delta_n \to \infty$ and $\delta_n \to 0$, the supremum runs over all t and x with $x + n^{-1/5} f(x)^{-1/4} t \in [0, 1]$ and $x \in [\delta_n, 1 - \delta_n]$. The function κ is defined as

$$
\kappa(t) = \frac{1}{2} \exp(-|u|/\sqrt{2}) \sin(|u|/\sqrt{2} + \pi/4).
$$

Put $L_n(x) = \{i: 1 \leqslant i \leqslant n, x_i \in [\delta_n, 1 - \delta_n], |x - x_i| \leqslant n^{-1/5} f(x_i)^{-1/4}]\}$. From this result we get for $j \in J_n$:

$$
n^{4/5} \operatorname{var} Z_j = n^{-6/5} \sum_{i=1}^n \{G_n((j-1)n^{-1/5}, x_i) + G_n(jn^{-1/5}, x_i) - 2G_n((j - 1/2)n^{-1/5}, x_i)\}^2
$$

$$
\geqslant n^{-6/5} \sum_{i \in L_n(jn^{-1/5})} \{G_n((j-1)n^{-1/5}, x_i) + G_n(jn^{-1/5}, x_i) - 2G_n((j - 1/2)n^{-1/5}, x_i)\}^2
$$

$$
= n^{-4/5} \sum_{i \in L_n(jn^{-1/5})} \left\{ \kappa[(j - 1 - n^{1/5} x_i) f(x_i)^{1/4}] + \kappa[j - n^{1/5} x_i) f(x_i)^{1/4}]. \right.
$$

$$
\left. -2\kappa \left[\left(j - \frac{1}{2} - n^{1/5} x_i \right) f(x_i)^{1/4} \right] \right\}^2 f(x_i)^{-3/2} + o(1)
$$

$$
= \int_{-1}^1 \left\{ \kappa[(u - 1)\tau_j^{1/4}] + \kappa[u\tau_j^{1/4}] - 2\kappa \left[\left(u - \frac{1}{2} \right) \tau_j^{1/4} \right] \right\}^2 du\, \tau_j^{-1/2} + o(1),
$$

where $\tau_j = f(jn^{-1/5})$. This inequality shows claim (26).

It remains to show lemma 1.

Proof of lemma 1. For a closed convex cone C denote the projection onto C by P_C. The polar cone C^* of C is defined by $C^* = \{v: P_C(v) = 0\}$. Lemma 1 is a consequence of the following geometric property.

Lemma 2

If C is a closed convex cone and L a linear subspace, then the following two conditions are equivalent:

$$\|P_C(P_L(v))\| \le \|P_C(v)\| \quad \text{for all } v, \tag{27}$$

$$P_L(C^*) \subset C^*. \tag{28}$$

For the proof of lemma 1 it is enough to apply lemma 2 to the cone C^* equal to the set of increasing functions of $\mathscr{H}$ and to the subspace L equal to the set of functions of $\mathscr{H}$ constant on $\mathscr{X}^+$ and $\mathscr{X}^-$. It remains to check that the projection of an increasing function onto L is increasing. However, this is clear because in the projection the values of the function on both intervals are replaced by the interval averages.

We come now to the proof of lemma 2.

Proof of lemma 2. Although this lemma is quite simple we are not aware of a reference in the literature on convex analysis.

We show first that (27) implies (28). If (27) holds, and $P_C v = 0$ we have

$$\|P_C P_L v\| \le \|P_C v\| = 0,$$

so that $P_C P_L v = 0$, i.e. (28) holds.

Conversely, assume now that (28) holds. Then for all v, because of $P_L P_{C^*} v \in C^*$, it holds that $P_C P_L P_{C^*} v = 0$. This and $v = P_C v + P_{C^*} v$ implies

$$\|P_C P_L v\| = \|P_C P_L P_C v + P_C P_L P_{C^*} v\| = \|P_C P_L P_C v\| \le \|P_C v\|,$$

i.e. (27) holds.

Proof of proposition 5. Note that for all functions g with $\int g^{(k)}(x)^2\, dx < \infty$ we have

$$\frac{1}{n}\sum_{i=1}^{n}(\hat{m}_S(x_i) - Y_i)^2 + \lambda_n \int \hat{m}_S^{(k)}(x)^2\, dx \le \frac{1}{n}\sum_{i=1}^{n}(g(x_i) - Y_i)^2 + \lambda_n \int g^{(k)}(x)^2\, dx. \tag{29}$$

For all functions m with $\int m^{(k)}(x)^2\, dx < \infty$ we get by application of (29) for $g = \hat{m}_S + \alpha(\hat{m}_S - m)$ with $\alpha \to 0$

$$\frac{1}{n}\sum_{i=1}^{n}(\hat{m}_S(x_i) - m(x_i))(\hat{m}_S(x_i) - Y_i) + \lambda_n \int \hat{m}_S^{(k)}(x)(\hat{m}_S^{(k)}(x) - m^{(k)}(x))\, dx = 0. \tag{30}$$

Equation (30) shows

$$\frac{1}{n}\sum_{i=1}^{n}(m(x_i) - Y_i)^2 + \lambda_n \int m^{(k)}(x)^2\, dx = \frac{1}{n}\sum_{i=1}^{n}(\hat{m}_S(x_i) - m(x_i))^2 + \lambda_n \int (\hat{m}_S^{(k)}(x) - m^{(k)}(x))^2\, dx$$

$$+ \frac{1}{n}\sum_{i=1}^{n}(\hat{m}_S(x_i) - Y_i)^2 + \lambda_n \int \hat{m}_S^{(k)}(x)^2\, dx$$

$$= \|\hat{m}_S - m\|_n^2 + \lambda_n \int (\hat{m}_S^{(k)}(x) - m^{(k)}(x))^2\, dx + C(Y),$$

where $C(Y)$ is a quantity that does not depend on m. This shows the statement of the proposition.

Proof of corollary 1. For $\hat{m}_n^{CS}$, we have

$$\|\hat{m}_n^{CS} - m_0\|_n^2 = \mathcal{O}_P(n^{-2k/(2k+1)}).$$

Because of $\int\{(\partial^k/(\partial x)^k)\hat{m}_n^{CS}(x)\}^2\, dx = \mathcal{O}_P(1)$ and $\int\{m_0^{(k)}(x)\}^2\, dx < \infty$, this implies $\int\{\hat{m}_n^{CS}(x) - m_0(x)\}^2\, dx = \mathcal{O}_P(n^{-2k/(2k+1)})$. The interpolation inequality (30) implies for $q \leqslant 2$,

$$\int_0^1 \left\{ \frac{\partial^q}{(\partial x)^q} \hat{m}_n^{CS}(x) - m_0^{(q)}(x) \right\}^2 dx = \mathcal{O}_P(n^{-(2k-2q)/(2k+1)}). \tag{31}$$

Similarly one gets for $q \leqslant 2$,

$$\int_0^1 \left\{ \frac{\partial^q}{(\partial x)^q} \hat{m}_n^S(x) - m_0^{(q)}(x) \right\}^2 dx = \mathcal{O}_P(n^{-(2k-2q)/(2k+1)}). \tag{32}$$

Equations (31) and (32) imply for $q \leqslant 2$,

$$\int_0^1 \left\{ \frac{\partial^q}{(\partial x)^q} \hat{m}_n^{CS}(x) - \frac{\partial^q}{(\partial x)^q} \hat{m}_n^S(x) \right\}^2 dx = \mathcal{O}_P(n^{-(2k-2q)/(2k+1)}). \tag{33}$$

We apply now that for a function h and for $C > 0$ large enough it holds for our choice of x_i, $i = 1, \ldots, n$ that

$$\left| \int_0^1 h(x)\, dx - \frac{1}{n} \sum_{i=1}^n h(x_i) \right| \leqslant C n^{-2} \int_0^1 |h'(x)| + |h''(x)|\, dx.$$

(This follows from

$$\left| \int_a^b h(x)\, dx - \{b - a\}\{h(a) + h(b)\}/2 \right| \leqslant C'\{b - a\}^2 \int_a^b |h''(x)|\, dx,$$

$$\left| \int_0^b h(x)\, dx - bh(b) \right| \leqslant C'b^2|h'(b)| + C'b^2 \int_0^b |h''(x)|\, dx$$

$$\leqslant C'b^2 \int_0^1 |h'(x)| + 2|h''(x)|\, dx,$$

$$\left| \int_a^1 h(x)\, dx - \{1 - a\}h(a) \right| \leqslant C'\{1 - a\}^2 \int_0^1 |h'(x)| + 2|h''(x)|\, dx$$

for C' large enough.) With $h = g^2$ this gives

$$\left| \int_0^1 g(x)^2\, dx - \|g\|_n^2 \right| \leqslant C n^{-2} \int_0^1 \left(\left| \frac{\partial^2(g^2)}{(\partial x)^2}(x) \right| + \left| \frac{\partial(g^2)}{\partial x}(x) \right| \right) dx.$$

Using the Cauchy–Schwarz inequality one can show for C'' large enough

$$\left| \int_0^1 g(x)^2 \, dx - \|g\|_n^2 \right| \leqslant C'' n^{-2} \left(\int_0^1 g'(x)^2 \, dx + \sqrt{\int_0^1 g(x)^2 \, dx \int_0^1 g''(x)^2 \, dx} \right.$$

$$\left. + \sqrt{\int_0^1 g(x)^2 \, dx \int_0^1 g'(x)^2 \, dx} \right). \tag{34}$$

Because of (33) this shows for $g = \hat{m}_n^{CS} - \hat{m}_n^S$

$$\left| \|\hat{m}_n^{CS} - \hat{m}_n^S\|_n^2 - \int_0^1 \{\hat{m}_n^{CS}(x) - \hat{m}_n^S(x)\}^2 \, dx \right| = \mathcal{O}_P(n^{-6k/(2k+1)}). \tag{35}$$

By definition of $\tilde{m}_n^{CS}$ and because of $\lambda_n = \mathcal{O}_P(n^{-2k/(2k+1)})$ we have

$$\int_0^1 \{\tilde{m}_n^{CS}(x) - \tilde{m}_n^S(x)\}^2 \, dx + \lambda_n \int_0^1 \left\{ \frac{\partial^k}{(\partial x)^k} \tilde{m}_n^{CS}(x) - \frac{\partial^k}{(\partial x)^k} \hat{m}_n^S(x) \right\}^2 \, dx$$

$$\leqslant \int_0^1 \{\tilde{m}_n^{CS}(x) - \tilde{m}_n^S(x)\}^2 \, dx + \lambda_n \int_0^1 \left\{ \frac{\partial^k}{(\partial x)^k} \hat{m}_n^{CS}(x) - \frac{\partial^k}{(\partial x)^k} \hat{m}_n^S(x) \right\}^2 \, dx$$

$$= \mathcal{O}_P(n^{-2k/(2k+1)}).$$

For $g = \tilde{m}_n^{CS} - \hat{m}_n^s$ this shows $\int g(x)^2 \, dx = \mathcal{O}_P(n^{-2k/(2k+1)})$ and $\int g^{(k)}(x)^2 \, dx = \mathcal{O}_P(n^{-2k/(2k+1)})$. With interpolation inequality (20) this gives for $q \leqslant 2$, $\int_0^1 g^{(q)}(x)^2 \, dx = \mathcal{O}_P(n^{(2k-2q)/(2k+1)})$. Using (34) again we get

$$\left| \|\tilde{m}_n^{CS} - \hat{m}_n^S\|_n^2 - \int_0^1 \{\tilde{m}_n^{CS}(x) - \hat{m}_n^S(x)\}^2 \, dx \right| = \mathcal{O}_P(n^{-6k/(2k+1)}). \tag{36}$$

Using (35) and (36) one can show (12) and (13) by a geometrical argument.

Proof of proposition 6. Choose $\check{g}$ as the linear interpolant of $\hat{m}_n^{RCS}$ with interpolation points $t_1 < \cdots < t_s$. We will show that

$$\int_0^1 \{\check{g}(x) - \hat{m}_n^{RCS}(x)\}^2 \, dx = \mathcal{O}_P(\delta_n^4).$$

Proceeding as in the proof of proposition 2, we get that $\int_0^1 \{(\partial^2/(\partial x)^2)\hat{m}_n^{RCS}(x)\}^2 \, dx = \mathcal{O}_P(1)$. Put $\Delta(u) = \hat{m}_n^{RCS}(u) - \check{g}(u)$. Note that $\Delta(t_i) = 0$ for $i = 1, \ldots, s$. For $t_i < x < t_{i+1}$ (note that for all i there exists a u_i with $\Delta'(u_i) = 0$), we get

$$|\Delta'(x)| \leqslant \int_{t_i}^{t_{i+1}} |\Delta''(u)| \, du \leqslant (t_{i+1} - t_i)^{1/2} \left\{ \int_{t_i}^{t_{i+1}} \Delta''(u)^2 \, du \right\}^{1/2}.$$

This gives

$$|\Delta(x)| \leqslant (t_{i+1} - t_i)^{3/2} \left\{ \int_{t_i}^{t_{i+1}} \Delta''(u)^2 \, du \right\}^{1/2}$$

and

$$\int_{t_i}^{t_{i+1}} \Delta^2(u) \, du \leqslant |t_{i+1} - t_i|^4 \int_{t_i}^{t_{i+1}} \Delta''(u)^2 \, du \leqslant \delta_n^4 \int_{t_i}^{t_{i+1}} \Delta''(u)^2 \, du.$$

Because of $\int_0^1 \{(\partial^2/(\partial x)^2)\hat{m}_n^{RCS}(x)\}^2 \, dx = \mathcal{O}_P(1)$, this shows the statement of the proposition.

Acknowledgement

References

Adams, R. A. (1975). *Sobolev spaces*. Academic Press, New York.

Agmon, S. (1965). *Lectures on elliptic boundary value problems.* D. van Nostrand, Princeton, NJ.

Birgé, L. & Massart, P. (1993). Rates of convergence for minimum contrast estimators. *Probab. Theory Related Fields* **97**, 113–150.

Birman, M. S. & Solomjak, M. J. (1967). Piecewise polynomial approximations of functions of the classes W_p^α. *Mat. Sb.* **73**, 295–317.

Delecroix, M. & Thomas-Agnan, C. (1997). Kernel and spline smoothing under shape restrictions. In *Smoothing and regression: approaches, computation and application* (ed. M. Schimek). Wiley, New York. (To appear.)

Delecroix, M., Simioni, S. & Thomas-Agnan, C. (1996). Functional estimation under shape constraints. *J. Nonparametr. Statist.* **6**, 69–89.

Eubank, R. L. (1988). *Spline smoothing and nonparametric regression*. Marcel Dekker, New York.

Elfving, T. & Andersson, L. E. (1988). An algorithm for computing constrained smoothing spline functions. *Numer. Math.* **52**, 583–595.

Korostelev, A. P. & Tsybakov, A. B. (1993). *Minimax theory of image reconstruction.* Lecture Notes in Statistics **82**, Springer, New York.

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

Mammen, E., Marron, J. S., Turlach, B. A. & Wand M. P. (1998). A general framework for constrained smoothing. (Preprint.)

Mammen, E. & van de Geer, S. (1997a). Locally adaptive regression splines. *Ann. Statist.* **25**, 387–413.

Mammen, E. & van de Geer, S. (1997b). Penalized quasi-likelihood estimation in partial linear models. *Ann. Statist.* **25**, 1014–1035.

Messer, K. (1991). A comparison of a spline estimate to its equivalent kernel estimate. *Ann. Statist.* **19**, 817–829.

Nychka, D. (1995). Splines as local smoothers. *Ann. Statist.* **23**, 1175–1197.

Silverman, B. W. (1982). Spline smoothing: the equivalent variable kernel method. *Ann. Statist.* **12**, 898–916.

Utreras, F. (1985). Smoothing noisy data under monotonicity constraints: existence, characterization and convergence rates. *Numer. Math.* **47**, 611–625.

van de Geer, S. (1990). Estimating a regression function. *Ann. Statist.* **18**, 907–924.

van de Geer, S. (1993). Hellinger-consistency of certain nonparametric maximum likelihood estimators. *Ann. Statist.* **21**, 14–44.

van de Geer, S. (1995). A maximal inequality for the empirical process. Technical Report TW95-05, University of Leiden.

Villalobos, M. & Wahba, G. (1987). Inequality constrained multivariate smoothing splines with application to the estimation of posterior probabilities. *J. Amer. Statist. Assoc.* **82**, 239–248.

Wahba, G. (1990). *Spline models for observational data*. CBMS–NSF Regional Conference Series in Applied Mathematics. SIAM, Philadelphia.

Enno Mammen, Institut für Angewandte Mathematik, Universität Heidelberg, Im Neuenheimer Feld 294, 69120 Heidelberg, Germany.