The Annals of Statistics
1983, Vol. 11, No. 2, 530–551

ASYMPTOTICS FOR *M*-TYPE SMOOTHING SPLINES

BY DENNIS D. COX

University of Wisconsin

Limit theorems giving rates of convergence of nonparametric regression estimates obtained from smoothing splines are proved. The main emphasis is on nonlinear, robust smoothing splines, but new results are obtained for the usual (linear) case. It is assumed that the knots become asymptotically uniform in a vague sense. Convergence of derivatives is also investigated. The main mathematical tools are a linearization of the robust smoothing spline, and an approximation of the linear smoothing spline utilizing the Green's function of an associated boundary value problem.

1. Introduction. Let $\theta(\cdot)$ be an unknown real valued function defined on a compact interval, which may be assumed to be [0, 1]. Consider observations

$$z_i = \theta(t_i) + \varepsilon_i, \quad 1 \le i \le n, \tag{1.1}$$

where $0 \le t_1 < t_2 < \cdots < t_n \le 1$ are known, and the random errors $\varepsilon_1, \varepsilon_2, \cdots, \varepsilon_n$ are assumed independent and identically distributed (abbreviated i.i.d) with either $E\varepsilon_1 = 0$ or the law (ε_1) symmetric about 0. The goal is to estimate θ from the observations. If θ is assumed to belong to a space of functions which is parameterized by a finite number of parameters (e.g. the polynomials of degree m or less) then standard regression techniques may be applied. However, if we are only willing to assume that θ belongs to

$$W_2^m = \{f \in C^{m-1}[0, 1]: f^{(m-1)} \text{ is absolutely continuous, and } f^{(m)} \in L^2[0, 1]\}, \tag{1.2}$$

then use of a nonparametric regression technique is in order. One of the most popular such techniques is polynomial spline smoothing. For this, we take our estimate $\hat{\theta}$ as the minimizer over $\xi \in W_2^m$ of

$$\frac{1}{n} \sum_{i=1}^n \{z_i - \xi(t_i)\}^2 + \lambda \int_0^1 \{\xi^{(m)}(t)\}^2 \, dt, \tag{1.3}$$

where $\lambda > 0$ is a smoothing constant, either chosen by the investigator or determined from the data. It turns out that $\hat{\theta}$ is a polynomial spline function of degree $2m - 1$, i.e. $\hat{\theta} \in C^{2m-2}[0, 1]$ and $\hat{\theta}^{(2m-1)}$ is a step function. The basic idea of this estimate can be traced back to Whittaker (1923), who proposed a discrete version using finite differences in place of derivatives in (1.3). In a 1964 article, Schoenberg proposed replacement of the finite differences in Whittaker's objective function by derivatives. Schoenberg related this via Lagrange multipliers to the following constrained minimization problem: minimize $n^{-1} \sum \{z_i - \xi(t_i)\}^2$ subject to $\int \{\xi^{(m)}(t)\}^2 \, dt < M$, where M is a given constant. Under the assumption of normal errors, one recognizes this as a maximum likelihood estimation approach, where the parameter space is $\Theta(M) \equiv \{\xi \in W_2^m: \int (\xi^{(m)})^2 < M\}$. This is closely related to the method of sieves introduced by Grenander (1981). A complementary approach to Schoenberg's was given by Reinsch (1967), who considered the problem of minimizing $\int (\xi^{(m)})^2$ subject to an upper bound constraint on the residual sum of squares. We can think of no statistical justification for this approach, although it is typical of what is done in much of the numerical analysis literature. Kimeldorf and Wahba (1970a) showed that smoothing splines are optimal Bayes estimates for the squared error loss function under the assumptions that the errors are normal, and that θ is chosen according to a

Received November 1981; revised September 1982.

AMS 1970 *subject classifications.* Primary 62J99; secondary 62F35, 41A15, 62G05.

Key words and phrases. Smoothing splines, nonparametric regression, robust estimation.

Gaussian prior for which W_2^m is the reproducing kernel Hilbert space (see Kallianpur, 1971).

As with many statistical procedures which are optimal for normal errors, ordinary smoothing splines are highly sensitive to one or more outlying observations. This was noted quite early by Greville (1957), who proposed a numerical procedure that amounts to an iteratively reweighted least squares computation. More recently, Huber (1979) has utilized the following robustified smoothing spline estimate of θ: let $\hat{\theta}$ be the minimizer over $\xi \in W_2^m$ of

$$\frac{1}{n}\sum_{i=1}^{n} \rho(z_i - \xi(t_i)) + \lambda \int \{\xi^{(m)}(t)\}^2\, dt. \tag{1.4}$$

Here, ρ is a suitably chosen function, usually convex and symmetric about 0. Huber's favorite is

$$\rho(x) = \begin{cases} x^2 & \text{if } |x| \le C, \\ C(2|x| - C) & \text{if } |x| \ge C, \end{cases}$$

where C is a tuning constant usually determined from the data. Huber derived such estimates by looking at the Lagrangian equation for the (constrained) maximum likelihood problem with the parameter space $\Theta(M)$, but where the errors are assumed to have a density of the form $\exp\{-\rho(\varepsilon)\}$. Alternatively, one could argue that such estimates are Bayesian maximum *a posteriori*, wherein one uses Huber's error density and Kimeldorf and Wahba's prior. Such an approach involves some measure theoretic difficulties (see e.g. Larkin, 1972, and Leonard, 1978) without being particularly compelling. For the case $m = 2$, Utreras (1981) gives a numerical procedure for computing the minimizer of (1.4), which we refer to as an M-type smoothing spline. In Section 2 of this last reference, some of the basic properties of these estimates are described, such as, that the estimates are polynomial spline functions. We will review such properties that are important for us in Section 2 below.

Now one would hope that as the knots $t_1, t_2, \cdots, t_n$ become more numerous, the estimate $\hat{\theta}$ would converge in some sense to θ. This will indeed happen, provided $\lambda = \lambda_n$ is varied with n. The basic result, first stated by Craven and Wahba (1979) (abbreviated hereafter by CW) gives convergence rates for the least squares smoothing spline (the minimizer of (1.3)). If $t_i = t_{in} = i/n$, $\theta \in W_2^m$, and the limit relations

$$n \to \infty,\ \lambda \to 0,\ n^{2m}\lambda \to \infty \tag{1.5}$$

hold, then

$$E\frac{1}{n}\sum_{i=1}^{n} \{\hat{\theta}(t_i) - \theta(t_i)\}^2 = O(\lambda) + O(n^{-1}\lambda^{-1/2m}). \tag{1.6}$$

The proof in CW is lacking in rigor at one point, but this detail has been repaired by Utreras (1980b). One may find still more general results in Speckman (1981a).

Theorem 4.3 below extends the estimate (1.6) to fairly arbitrary knot sequences. Slightly stronger conditions than (1.5) are required, but the loss of generality is not serious, as is explained in the remarks following the statement of the theorem in Section 4. In order to describe the hypothesis of that theorem, let F_n be the cumulative distribution function (abbreviated c.d.f.) for the probability measure obtained by putting equal magnitude point masses at each knot. Then, the third limit relation in (1.5) is replaced by

$$\left(\int |t - F_n(t)|\, dt\right)\lambda^{-1/m} |\log \lambda| \to 0 \tag{1.7}$$

in the conditions for Theorem 4.3. If $m > 2$, then the factor $|\log \lambda|$ is not necessary in (1.7). Note that (1.7) requires that the knots be asymptotically uniform in a rather weak sense. For the equispaced knots case ($t_i = i/n$), the first factor in (1.7) is $O(n^{-1})$, but for the random knots case ($t_1, \cdots, t_n$ i.i.d. from the uniform distribution on [0, 1]), this factor is $O_p(n^{-1/2})$. The main tool used in the proof of Theorem 4.3 is given in Theorem 3.2. This latter result may be described heuristically as follows: First let $n \to \infty$ while holding λ fixed in the variational equation for $\hat{\theta}$ given in Proposition 2.2. (That one can do this is suggested by Lemma 2.3 of Cogburn and Davis (1974), and also by Theorem 2.3 of Speckman (1981a).) One then obtains a differential equation with boundary conditions for the estimate. Theorem 3.2 shows that the Green's function for this boundary value problem can be used to approximate the original estimate $\hat{\theta}$ (which minimized (1.3)) and gives bounds on the error. This result is used extensively in the sequel.

One would hope that an analog of (1.6) holds for general M-type smoothing splines. That this is the case for a wide class of ρ functions is stated in Theorem 5.1, but under the restriction $m = 2$ (cubic smoothing splines). We are quite sure that this holds for all m, it only being necessary to compute the appropriate Green's function, and verify certain asymptotic estimates on it. Theorem 5.1 hinges also on Theorem 3.2, and on a generalization of Huber's (1973) asymptotic linearization of robust regression estimates (see Theorem 3.1 below). Results for general m but in the periodic case are given in Theorem 4.4.

It is frequently of interest to estimate one or more derivatives of θ. The natural approach is to use $\hat{\theta}^{(p)}$, the pth order derivative of the smoothing spline, as an estimate of $\theta^{(p)}$. This topic is treated in Section 6, where it shows that under rather complicated conditions, the L^2 norm $\| \hat{\theta}^{(p)} - \theta^{(p)} \|_2$ does converge to 0 in probability. In particular we only look at the case of periodic smoothing (i.e. the unknown θ is assumed periodic, and $\hat{\theta}$ is taken as the minimizer of (1.4) over the space of periodic functions, see (2.8) below). Consistency results for derivatives of linear (least squares) smoothing splines (i.e. minimizers of (1.3)) have been obtained by Rice and Rosenblatt (1981b) in the general case, and Ragozin (1981) in the nonperiodic case. Related work by Stone (1980) suggests that the order of convergence bounds in Theorem 6.3 are tight, but we believe that significant improvement can be made in weakening the hypotheses of the theorem.

Indeed, all of our convergence theorems in Sections 4 through 6 contain rates of convergence which the uninitiated reader may find disappointingly slow. In the usual finite dimensional parameter estimation problems, all decent estimates converge to the true parameter at rate $n^{-1/2}$ in probability, no matter what distance measure is used. The papers by Stone (1980) and Speckman (1981b) show that this is not the case if the parameter space is infinite dimensional. Now our upper bounds, which are valid for an infinite dimensional space of functions Θ (defined by smoothness and boundary conditions), depend on the parameter λ and the sample size n. If one chooses λ as a function of n so as to minimize the upper bound in each case, then an "optimal" or "best possible" upper bound on the rate of convergence is obtained. We conjecture that in all cases treated here, the rate so obtained cannot be improved on by any estimator sequence which would be consistent for any element of Θ of it were true. See the reference of Stone's for a more precise definition of optimal convergence rates.

Finally, we mention two problems of great interest that are not treated in this work. Firstly, the problem of scale estimation is ignored throughout. The objective function (1.4) should be amended to

$$\frac{1}{n} \sum_{i=1}^{n} \rho(\{z_i - \xi(t_i)\}/S) + \lambda \int_0^1 \{\xi^{(m)}(t)\}^2 \, dt,$$

where S is an estimate of the scale of the errors, obtained usually from the residuals. We have tacitly assumed S is known, and hence incorporated it in the ρ function. A problem of somewhat greater magnitude is that of "estimating" the smoothing parameters, λ and m. The estimation of m has received little attention, but there is a voluminous literature on the "estimation" of λ, mostly by the use of cross validation (consult CW and the references cited therein). Suggestions for the estimation of both S and λ are given in Huber

(1979). Some results on consistency (even asymptotic optimality) of cross validated estimates related to smoothing splines have been given by Speckman (1981b).

2. Definitions and basic results. In this section, we establish some notations and give some characterizations of the solution to the minimization problem (1.4). This naturally leads to the formalism of reproducing kernel Hilbert spaces. The general results are specialized to the two cases of interest to us—namely, polynomial smoothing splines and periodic smoothing splines.

Let

$$\mathbf{Z}_n = (z_{1n}, z_{2n}, \cdots, z_{nn})', \qquad \boldsymbol{\varepsilon}_n = (\varepsilon_{1n}, \varepsilon_{2n}, \cdots, \varepsilon_{nn})'$$

be the data and error vectors, respectively, as given in (1.1). Then we may rewrite the observation model as

$$\mathbf{Z}_n = X_n\theta + \boldsymbol{\varepsilon}_n, \tag{2.1}$$

where X_n is a linear operator whose domain is some function space H and whose range is $\mathbb{R}^n$, viz.

$$X_n\theta = (\theta(t_{1n}), \theta(t_{2n}), \cdots, \theta(t_{nn}))'.$$

We are interested in the case where our function space is a Hilbert space, and where X_n is continuous. We are then naturally led to suppose that the parameter space H is a reproducing kernel Hilbert space (RKHS), i.e. a Hilbert space of functions (on [0, 1]) for which the evaluation functional $f \mapsto f(t)$ is continuous for each t. It follows from the classical RKHS theory (see e.g. Aronsajn, 1950) that there is a "reproducing kernel" $K: [0, 1] \times [0, 1] \to \mathbb{R}$ which satisfies: (i) for each fixed $t \in [0, 1]$, $K(\cdot, t)$ is in H; (ii) for each $t \in [0, 1]$ and each $f \in H$,

$$\langle K(\cdot, t), f \rangle = f(t),$$

where $\langle \cdot, \cdot \rangle$ denotes the inner product on H. Thus, $K(\cdot, t)$ is the element of H which represents evaluation at t (in the sense of Riesz representation).

Given the observations as in (2.1), we are interested in estimating θ. For convenience, let

$$y_i = K(\cdot, t_i), \quad 1 \le i \le n$$

(we drop the subscript n whenever possible). Then, according to the reproducing property, $h(t_i) = \langle h, y_i \rangle$. Hence, our estimate will be a minimizer of an objective function of the form

$$n^{-1} \textstyle\sum_{i=1}^{n} \rho(z_i - \langle y_i, h \rangle) + \lambda \, \| \mathscr{P} h \|^2 \tag{2.2}$$

with respect to the variable $h \in H$. Here, $\mathscr{P}$ is an orthogonal projection whose null space,

$$N = \{h \in H : \mathscr{P}h = 0\}, \tag{2.3a}$$

is finite dimensional. Note that $\mathscr{P}$ is self adjoint and idempotent (i.e. $\mathscr{P}^2 = \mathscr{P}$). Examples will be given shortly. Define the subspace

$$S_n = \operatorname{span}\{y_1, \cdots, y_n\}. \tag{2.3b}$$

If $w \in S_n^\perp \cap N$, then the value of the objective function (2.2) at $h + w$ is the same as at h. However, the usual situation is that $N \subseteq S_n$, in which case the existence of a minimizer implies the existence of a minimizer in S_n. Thus, the following result is obtained.

PROPOSITION 2.1. *If ρ is strictly convex and $N \subseteq S_n$, then there is a unique minimizer $\hat{\theta}$ of* (2.2), *and $\hat{\theta} \in S_n$.* □

The next proposition provides a variational equation for computing $\hat{\theta}$. It's proof follows by taking directional derivatives and setting them equal to 0.

PROPOSITION 2.2. *Suppose $\psi = \rho'$ exists everywhere; then a minimizer $\hat{\theta}$ of* (2.2) *satisfies*

$$-n^{-1} \sum_{i=1}^{n} y_i \psi(z_i - \langle y_i, \theta \rangle) + \lambda \mathscr{P} \hat{\theta} = 0. \square \tag{2.4}$$

We now specialize to the examples which are of interest to us. For the first one, the RKHS is the Sobolev space

$$W_2^m = W_2^m[0, 1] = \{f : f^{(0)}, \cdots, f^{(m-1)} \text{ are absolutely continuous on } [0, 1], \text{ and } f^{(m)} \in L^2[0, 1]\},$$

where m is a positive integer. If we use the inner product

$$\langle f, g \rangle = \sum_{i=0}^{m-1} f^{(i)}(0) g^{(i)}(0) + \int_0^1 f^{(m)}(t) g^{(m)}(t)\, dt,$$

then the reproducing kernel is

$$K(s, t) = \sum_{i=0}^{m-1} \frac{t^i s^i}{(i!)^2} + \int_0^1 \frac{(s-u)_+^{m-1} (t-u)_+^{m-1}}{\{(m-1)!\}^2}\, du,$$

where $u_+ = \max(u, 0)$. Let $\mathscr{P}$ be the orthogonal projection onto the closed subspace of functions f satisfying

$$f^{(i)}(0) = 0 \quad \text{for} \quad 0 \le i \le m-1.$$

Then we note that

$$\langle \mathscr{P} f \rangle = \int_0^1 \{f^{(m)}(t)\}^2\, dt, \tag{2.5}$$

so the objective function (2.2) to be minimized is

$$n^{-1} \sum_{i=1}^{n} \rho(z_i - h(t_i)) + \lambda \int_0^1 \{h^{(m)}(t)\}^2\, dt. \tag{2.6}$$

The space S_n of (2.3b) depends only on $t_1, \cdots, t_n$ and is the space of polynomial spline functions (see Section 3 of Greville, 1969). If $n \ge m$, then N given in (2.3a) is a subspace of S_n, so a minimizer $\hat{\theta}$ of (2.6) is in S_n. In fact, $\hat{\theta}$ is in

$$\eta_{2m-1}(t_1, t_2, \cdots, t_n) = \{x \in S_n : x \text{ is a polynomial of degree } 2m-1 \text{ in } [t_i, t_{i+1}],\ 1 \le i \le n-1,\ x \text{ is a polynomial of degree } m-1 \text{ in } [0, t_1] \text{ and } [t_n, 1] \text{ and } x \in C^{2m-2}[0, 1]\}. \tag{2.7}$$

This follows because the problem of minimizing

$$\int_0^1 \{h^{(m)}(t)\}^2\, dt$$

subject to constraints $h(t_i) = \hat{\theta}(t_i)$, $1 \le i \le n$, has solution $h = \hat{\theta}$. Hence, $\hat{\theta}$ is an interpolating spline, and belongs to η_{2m-1} by Theorems 6.2 and 7.3 of Greville (1969).

For the other case of interest, the RKHS is the periodic Sobolev space

$$K^{(m)} = \{f \in W_2^m : f^{(i)}(0) = f^{(i)}(1),\ 0 \le i \le m-1\}, \tag{2.8}$$

equipped with the inner product

$$\langle f, g\rangle = \left(\int_0^1 f(t)\,dt\right)\left(\int_0^1 g(t)\,dt\right) + \int_0^1 f^{(m)}(t)g^{(m)}(t)\,dt,$$

for which the reproducing kernel is conveniently given in the form

$$K(s, t) = 1 + 2\sum_{\nu=1}^{\infty} (2\pi\nu)^{-2m}\cos\{2\pi\nu(t - s)\}.$$

If the projection $\mathscr{P}$ is projection onto the subspace of functions with integral 0, then equation (2.5) holds in this setting. If $t_k = k/n$ for $1 \le k \le n$, then an explicit formula may be given for the solution to the least squares minimization problem (pages 1113 and 1114 of Cogburn and Davis, 1974). In this case of periodic splines we shall denote the space S_n of (2.3b) by $\zeta_{2m-1}(t_1, t_2, \cdots, t_n)$.

3. Approximation theorems. In this section, we give two technical results which form the basis for the rest of the paper. The first theorem shows that under certain conditions (Assumptions 1 and 2), a nonlinear M-type smoothing spline can be approximated by a linear smoothing spline acting on some transformed (unobservable) data. This is useful since the asymptotics for linear smoothing splines are somewhat easier to obtain. However, in order to check the conditions for this asymptotic linearization, it is required to know something of the limiting behavior of impulse response smoothing splines. These are smoothing splines to the data vector which is zero except for one component. The second theorem shows that these impulse responses can be approximated by a Green's function for a certain boundary value problem. The limiting behavior of the Green's function can be explicitly calculated in many instances, as is shown in the subsequent sections. Furthermore, this approximation theorem provides a new means for obtaining limit theorems for linear smoothing splines.

Let H denote the RKHS of the last section. Suppose θ is unknown and consider the observations

$$z_{in} = \theta(t_{in}) + \varepsilon_{in} \quad \text{for} \quad 1 \le i \le n, \tag{3.1}$$

where the ε_{in} are i.i.d. with common distribution having c.d.f. denoted by F. It will be convenient to assume that the knots $\{t_{in}: 1 \le i \le n\}$ at stage n are not necessarily a subset of the knots of stage $n + 1$ (e.g. $t_{in} = i/n$). Given any vector $\mathbf{x} \in \mathbb{R}^n$, let $s_n\mathbf{x}$ be the unique element in S_n such that $(s_n\mathbf{x})(t_{in}) = x_i$ (note that s_n is well defined if $S_n = \eta_{2m-1}$ or $S_n = \zeta_{2m-1}$, provided $n \ge m$). Let $\mathscr{P}$ be the projection of (2.2), and y_{in} the element of S_n representing evaluation at t_{in}. Define operators on S_n by

$$\mathscr{T}_n\xi = \frac{1}{n}\sum_{i=1}^n \xi(t_{in})y_{in} \tag{3.2}$$

$$\mathscr{G}_{n\lambda} = \mathscr{T}_n + 2(\lambda/E\psi')\mathscr{P}. \tag{3.3}$$

Note that $\mathscr{T}_n$ and $\mathscr{G}_{n\lambda}$ are self adjoint, and $\mathscr{T}_n$ is positive definite. We shall omit the subscripts n, λ wherever possible. Now consider the function $\Phi_{n\lambda}: S_n \to S_n$ given by

$$\Phi_{n\lambda}(\xi) = -\frac{1}{n}\sum_{i=1}^n y_{in}\psi(z_{in} - \xi(t_{in})) + 2\lambda\mathscr{P}\xi. \tag{3.4}$$

We are interested in solutions of $\Phi_{n\lambda}(\xi) = 0$. We shall denote such M-type smoothing spline estimates of θ by $\hat{\theta}_{n\lambda}$. Define pseudo data

$$\tilde{z}_{in} = \theta(t_{in}) + \psi(\varepsilon_{in})/(E\psi'),$$

and let $\tilde{\mathbf{z}}$ denote the associated vector. Consider the operator on S_n given by

$$\Psi_{n\lambda}(\xi) = -\mathscr{T}_n s_n\tilde{\mathbf{z}} + \mathscr{G}_{n\lambda}\xi. \tag{3.5}$$

Note that the solution $\tilde{\theta}_{n\lambda}$ of $\Psi_{n\lambda}(\xi) = 0$ is the minimizer of

$$\frac{1}{2n} \sum_{i=1}^{n} \{\tilde{z}_{in} - \xi(t_{in})\}^2 + (\lambda/E\psi') \| \mathscr{P}\xi \|^2, \tag{3.6}$$

where $\| \cdot \|$ denotes the norm of H. Note that $\mathscr{G}$ is invertible if $n \geq m$, since $\mathscr{G}\xi = 0$ is the equation for obtaining a least squares smoothing spline for identically zero data, and the solution is, uniquely, the zero spline. Also, note that for an arbitrary data vector **z**, the least squares smoothing spline on **z** is $\mathscr{G}^{-1}\mathscr{T}s\mathbf{z}$.

For continuous functions ξ_1 and ξ_2, define

$$d_n(\xi_1, \xi_2) = \left[\frac{1}{n} \sum_{k=1}^{n} \{\xi_1(t_{kn}) - \xi_2(t_{kn})\}^2\right]^{1/2}.$$

We will be interested in a sequence of semi-norms $\{\| \cdot \|_n : n \geq m\}$ given by a bilinear form, viz., $\| x \|_n = (x, x)_n^{1/2}$. Each $\| \cdot \|_n$ is supposed to be defined on a space containing S_n and $\{\theta\}$ such that $\| \cdot \|_n$ is a norm on S_n. An example is $\| \xi \|_n = d_n(\xi, 0)$. We also suppose that $\lambda = \lambda_n$ varies in a manner so that the limit conditions below hold. The following assumptions will be made:

ASSUMPTION 1. $\psi \in C^2(-\infty, \infty)$ and satisfies $M = \sup_{-\infty<t<\infty} | \psi''(t) | < \infty$, $E\psi' = \int \psi' \, dF \neq 0$, $E\psi = 0$, $\mathrm{Var}(\psi') = \int (\psi' - E\psi')^2 \, dF < \infty$, and $\mathrm{Var}(\psi) < \infty$.

ASSUMPTION 2. Define norms

$$||| \xi |||_{n\lambda} = \| \mathscr{G}_{n\lambda}^{-1} \xi \|_n$$

on S_n and constants

$$A_n = \sup\{d_n^2(\xi, 0)/\| \xi \|_n^2 : \xi \in S_n\}, \tag{3.7}$$

$$B_n = \sup\{||| y_{kn} |||_{n\lambda}^2 : 1 \leq k \leq n\}, \tag{3.8}$$

$$C_n = E \| \tilde{\theta}_{n\lambda} - \theta \|_n^2. \tag{3.9}$$

Assume that

$$\lim_{n\to\infty} n^{-1} A_n B_n = \lim_{n\to\infty} A_n^2 B_n C_n = 0. \tag{3.10}$$

REMARK. The constants A_n simply relate the two different norms on S_n, and the C_n's are the mean squared error of the linear estimate. The B_n's are somewhat more difficult to interpret; see (3.15) and the accompanying remarks. Theorem 3.2 and Corollary 3.3 will shed some light on the asymptotic behavior of the B_n's. If $\| \cdot \|_n = d_n(\cdot, 0)$, then the limit relations (3.10) are generally implied by the simpler requirements that $\lambda \to 0$ and $n\lambda^{1/m} \to \infty$. However, we shall need the requirement in (3.21) below, which is stronger than $n\lambda^{1/m} \to \infty$. Sufficient conditions for these limit relations to hold for other choices of $\| \cdot \|_n$ (involving derivatives) are given in the hypotheses of Theorem 6.3 below.

THEOREM 3.1. *Under Assumptions* 1 *and* 2, *we have that for any* $\delta > 0$, *there is an* n_0 *such that for all* $n \geq n_0$

$$P[\text{there is a solution } \hat{\theta}_{n\lambda} \text{ to } \Phi_{n\lambda}(\xi) = 0 \text{ satisfying } \| \hat{\theta}_{n\lambda} - \tilde{\theta}_{n\lambda} \|^2 \leq \delta C_n] \geq 1 - \delta.$$

REMARK. The conclusion says that with high probability, $\hat{\theta}_{n\lambda}$ and $\tilde{\theta}_{n\lambda}$ will be much closer than $\tilde{\theta}_{n\lambda}$ and θ. Hence, $\hat{\theta}_{n\lambda}$ will enjoy the same asymptotics as the least squares estimate $\tilde{\theta}_{n\lambda}$.

PROOF. We partially follow Huber (1973, pages 805–806). By Taylor's theorem

$$||| \Psi_{n\lambda}(\xi) - \Phi_{n\lambda}(\xi)/E\psi' |||_{n\lambda} \leq T_1 + T_2,$$

where

$$T_1 = ||| n^{-1} \textstyle\sum_{k=1}^{n} y_{kn}\{\theta(t_{kn}) - \xi(t_{kn})\}\{\psi'(\varepsilon_{kn}) - E\psi'\}/E\psi' |||_{n\lambda}$$

and

$$T_2 = ||| \tfrac{1}{2} n^{-1} \textstyle\sum_{k=1}^{n} y_{kn}\{\theta(t_{kn}) - \xi(t_{kn})\}^2 \psi''(\varepsilon_{kn} + a_{kn})/E\psi' |||_{n\lambda},$$

where a_{kn} is between 0 and $\theta(t_{kn}) - \xi(t_{kn})$. If $K \ge 8/\delta$, then

$$P[\| \tilde{\theta}_{n\lambda} - \theta \|_n < \tfrac{1}{2}(KC_n)^{1/2}] > 1 - \delta/2. \tag{3.11}$$

Define the sets

$$F_n = \{\xi \in S_n : \| \xi - \theta \|_n^2 \le KC_n\}.$$

Recalling that $\| \cdot \|_n$ and hence $||| \cdot |||_{n\lambda}$ is given by a bilinear form, we have for any $\xi \in S_n$ that

$$\begin{aligned} ET_1 &= E ||| n^{-1} \textstyle\sum_k y_{kn}\{\theta(t_{kn}) - \xi(t_{kn})\}\{\psi'(\varepsilon_{kn}) - E\psi'\}/E\psi' |||_{n\lambda}^2 \\ &= n^{-2} \textstyle\sum_k ||| y_{kn} |||_{n\lambda}^2 \{\theta(t_{kn}) - \xi(t_{kn})\}^2 \mathrm{Var}(\psi')/(E\psi')^2 \\ &\le n^{-1} B_n d_n(\xi, \theta)^2 \mathrm{Var}(\psi')/(E\psi')^2 \le n^{-1} B_n A_n \| \xi - \theta \|_n^2 \mathrm{Var}(\psi')/(E\psi')^2. \end{aligned}$$

By Markov's inequality, if $\xi \in F_n$ and

$$D = \{2\delta^{-1} \mathrm{Var}(\psi')(E\psi')^{-2}\}^{1/2},$$

then

$$P[T_1 \le D(n^{-1} K A_n B_n C_n)^{1/2}] > 1 - \delta/2. \tag{3.12}$$

If $\xi \in F_n$, then we have

$$\begin{aligned} T_2 &\le \tfrac{1}{2} M | E\psi' |^{-1} (\max_k ||| y_{kn} |||_{n\lambda}) d_n^2(\xi, \theta) \\ &\le D' B_n^{1/2} A_n \| \xi - \theta \|_n^2 \le D' K A_n B_n^{1/2} C_n, \end{aligned} \tag{3.13}$$

where $D' = \tfrac{1}{2} M | E\psi' |^{-1}$. Putting together (3.11), (3.12), and (3.13), we obtain an event of probability greater than $1 - \delta$ on which we have for all $\xi \in F_n$,

$$\begin{aligned} &||| \Phi_{n\lambda}(\xi)/E\psi' - \mathscr{G}_{n\lambda}(\xi - \theta) |||_{n\lambda} \\ &\qquad \le ||| \Phi_{n\lambda}(\xi)/E\psi' - \Psi_{n\lambda}(\xi) |||_{n\lambda} + ||| \mathscr{G}_{n\lambda}(\xi - \tilde{\theta}_{n\lambda}) - \mathscr{G}_{n\lambda}(\xi - \theta) |||_{n\lambda} \\ &\qquad \le \{D(n^{-1} A_n B_n)^{1/2} + D' K^{1/2} A_n B_n^{1/2} C_n^{1/2} + \tfrac{1}{2}\}(KC_n)^{1/2}. \end{aligned}$$

In view of the limit relations (3.10), the quantity in braces will be less than or equal to 1 for all n sufficiently large. For such n, if $x \in F_n - \theta$ and

$$U(x) = x - \mathscr{G}_{n\lambda}^{-1} \Phi(x + \theta)/E\psi'$$

then $\| U(x) \|_n^2 \le KC_n$, i.e. the continuous function U maps the compact, convex set $F_n - \theta$ into itself (note $F_n - \theta \subseteq S_n$, where S_n is finite dimensional). By Brouwer's theorem, U has a fixed point $\hat{x}$ in $F_n - \theta$. Putting $\hat{\theta}_{n\lambda} = \hat{x} + \theta$, we see that $\Phi_{n\lambda}(\hat{\theta}_{n\lambda}) = 0$. Furthermore,

$$\begin{aligned} ||| \Phi_{n\lambda}(\hat{\theta}_{n\lambda})/E\psi' - \Psi_{n\lambda}(\hat{\theta}_{n\lambda}) |||_{n\lambda} &= ||| \mathscr{G}_{n\lambda}(\hat{\theta}_{n\lambda} - \tilde{\theta}_{n\lambda}) |||_{n\lambda} = \| \hat{\theta}_{n\lambda} - \tilde{\theta}_{n\lambda} \|_n \\ &\le \{D(n^{-1} A_n B_n)^{1/2} + D' K^{1/2} A_n B_n^{1/2} C_n^{1/2}\}(KC_n)^{1/2}, \end{aligned} \tag{3.14}$$

where the inequality holds on an event of probability greater than $1 - \delta$. Applying (3.10) again will complete the proof of the theorem. □

We now turn to the problem of estimating the constant B_n in (3.8). First note that

$$\tilde{\theta}_{n\lambda j} \equiv \mathscr{G}_{n\lambda}^{-1} y_{jn} = \mathscr{G}^{-1} \mathscr{T} \mathscr{T}^{-1} y_{jn} = (\mathscr{G}^{-1} \mathscr{T}) \gamma_{jn} \tag{3.15}$$

where $\gamma_{jn} = ns(\mathbf{e}_{jn})$ and $\mathbf{e}_{jn} = (0, \cdots, 0, 1, 0, \cdots, 0)$ is the jth unit coordinate vector in $\mathbb{R}^n$. In the case of natural polynomial smoothing splines $\tilde{\theta}_{n\lambda j} \equiv \mathscr{G}_{n\lambda}^{-1} y_{jn}$ is the minimizer over ξ of

$$\frac{1}{n} \sum_{k=1}^{n} \{n\delta_{kj} - \xi(t_{kn})\}^2 + \lambda \int_0^1 \{\xi^{(m)}(t)\}^2 \, dt, \tag{3.16}$$

where δ_{kj} is Kronecker's delta. The next result shows $\tilde{\theta}_{n\lambda j}$ may be well approximated by the Green's function of a certain boundary value problem, a result suggested by the work of Utreras (1980) and Speckman (1981). Let $G_\lambda(t, \tau)$ be the Green's function for the differential operator $(-1)^m \lambda D^{2m} + 1$ acting on the subspace of $C^{2m}[0, 1]$ of functions satisfying the natural boundary conditions:

$$f^{(p)}(0) = f^{(p)}(1) = 0 \quad \text{for} \quad p = m, m+1, \cdots, 2m-1.$$

(For the case of periodic spline smoothing, we require periodic boundary conditions, i.e. $f^{(p)}(0) = f^{(p)}(1)$ for $p = 0, 1, 2, \cdots, 2m-1$.) That is to say, G_λ has the following properties:

$$\text{(i)} \qquad (-1)^m \lambda G_\lambda^{2m,0}(t, \tau) + G_\lambda(t, \tau) = \delta(t - \tau),$$

where $\delta(\cdot)$ is Dirac's delta 'function" (more correctly, generalized function), and

$$G_\lambda^{p,q}(t, \tau) = \frac{\partial^{p+q}}{\partial t^p \partial \tau^q} G_\lambda(t, \tau).$$

This property of G is interpreted as requiring that, when considered as a function of t, for fixed τ, $G(\cdot, \tau) \in C^{2m-2}[0, 1] \cap C^{2m}\{[0, \tau) \cup (\tau, 1]\}$, and that $G^{2m-1,0}(\cdot, \tau)$ has an appropriate jump discontinuity at $t = \tau$.

(ii) $G_\lambda(\cdot, \tau)$ satisfies the natural (or periodic, as the case may be) boundary conditions, to wit $G_\lambda^{p,0}(0, \tau) = G_\lambda^{p,0}(1, \tau) = 0$ for $m \le p \le 2m-1$. The existence of G_λ is assured, since $(-1)^m D^{2m}$ is a positive operator on the space of functions satisfying the boundary conditions (use m-fold integration by parts), so its eigenvalues are nonnegative.

Our approximation of $\tilde{\theta}_{n\lambda j}$ makes use of the following operator defined on $C[0, 1]$:

$$(\mathscr{R}_{n\lambda}\xi)(t) = \int_0^1 \xi(u) G_\lambda(t, u) \, d(u - F_n(u)). \tag{3.17}$$

Here, $F_n(u)$ denotes the c.d.f. of the probability measure which assigns mass $1/n$ to each knot t_{kn}. We shall require that

$$D_n \equiv \int_0^1 |t - F_n(t)| \, dt \tag{3.18}$$

converges to 0 as $n \to \infty$. This is equivalent to assuming F_n converges weakly to the uniform distribution, or just that $F_n(t) \to t$ for $\forall t \in [0, 1]$.

THEOREM 3.2. *Assume $n \ge m \ge 2$, and suppose there exist constants $C_{p,q} > 0$ such that for all $t, \tau \in [0, 1]$ and all $\lambda > 0$;*

$$|G_\lambda^{p,q}(t, \tau)| \le C_{p,q} \lambda^{-(p+q+1)/2m} \tag{3.19}$$

provided $p + q \le 2m - 2$, and

$$|G^{p,q}(t, \tau)| \le C_{p,q} \lambda^{-1}(L\lambda) \tag{3.20}$$

if $p + q = 2m - 1$, where

$$L\lambda = \max\{|\log \lambda|, 1\}.$$

Put

$$g_{n\lambda j}(t) = G_\lambda(t, t_{jn})$$

where G_λ is the Green's function defined above. Assume that λ varies with n in such a way that

$$\lim_{n\to\infty} D_n \lambda^{-1/m}(L\lambda) = 0. \tag{3.21}$$

Then for all $t \in [0, 1]$, if n is sufficiently large,

$$\tilde{\theta}_{n\lambda j}(t) = \textstyle\sum_{\nu=0}^{\infty} (\mathscr{R}_{n\lambda}^{\nu} g_{n\lambda j})(t). \tag{3.22}$$

Furthermore, if $p \le 2m - 2$, then $\tilde{\theta}_{n\lambda j}^{(p)}$ may be obtained by termwise differentiation of the series, and

$$\sup_t |\tilde{\theta}_{n\lambda j}^{(p)}(t) - g_{n\lambda j}^{(p)}(t)| \le K D_n \lambda^{-(p+3)/2m}(L\lambda) \tag{3.23}$$

where K is a constant depending only on m.

REMARK. That assumptions (3.19) and (3.20) concerning $|G_\lambda^{p,q}|$ are typically true will be seen in Sections 4 and 5.

PROOF. For any function $\xi \in C^1[0, 1]$, if $p \le 2m - 2$ then

$$(\mathscr{R}\xi)^{(p)}(t) = \int_0^1 G^{p0}(t, u)\xi(u)\, d(u - F_n(u)) = -\int_0^1 \{u - F_n(u)\}\left\{\frac{\partial}{\partial u} G^{p0}(t, u)\xi(u)\right\} du.$$

Consequently,

$$\sup_t |(\mathscr{R}\xi)^{(p)}(t)| \le D_n \sup_{t,u} \left|\frac{\partial}{\partial u} G^{p0}(t, u)\xi(u)\right|. \tag{3.24}$$

We will now use this bound to show by induction that

$$\sup_t |(\mathscr{R}_{n\lambda}^{\nu} g_{n\lambda j})^{(p)}(t)| \le K^\nu D_n^\nu \lambda^{-(p+2\nu+1)/2m}(L\lambda)^\nu \tag{3.25}$$

where $K = \max\{C_{pj}: 0 \le p \le 2m - 1 \text{ and } j = 0, 1\}$ depends only on m. For $\nu = 0$ the result is obvious from (3.19) and (3.20). Assume that for some ν,

$$\sup |(\mathscr{R}^\nu g_j)^{(p)}| \le K^\nu D_n^\nu \lambda^{-(p+2\nu+1)/2m}(L\lambda)^\nu.$$

Then

$$\begin{aligned}\sup |(\mathscr{R}^{\nu+1} g_j)^{(p)}| &\le D_n \sup |G^{p1}(\mathscr{R}^\nu g_j) + G^{p0}(\mathscr{R}^\nu g_j)^{(1)}| \\ &\le D_n^{\nu+1}[C_{p1}\lambda^{-(p+2)/2m}(L\lambda)K^\nu\lambda^{-(2\nu+1)/2m}(L\lambda)^\nu + C_{p0}\lambda^{-(p+1)/2m}(L\lambda)K^\nu\lambda^{-(2\nu+2)/2m}(L\lambda)^\nu] \\ &\le \max\{C_{p0}, C_{p1}\}K^\nu D_n^{\nu+1}\lambda^{-(p+2\nu+3)/2m}(L\lambda)^{\nu+1}\end{aligned}$$

and (3.25) follows since $\max(C_{p0}, C_{p1}) \le K$. Therefore, if only

$$D_n \lambda^{-1/m} L\lambda < K^{-1},$$

it follows that

$$\textstyle\sum_{\nu=0}^{\infty} (\mathscr{R}^\nu g_j)^{(p)}(t)$$

converges uniformly in $t \in [0, 1]$ by comparison with a geometric series. In particular, if (3.21) holds, the series $\sum \mathscr{R}^\nu g_j$ eventually defines a continuous function ξ_j whose first $2m - 2$ derivatives are continuous and computable by termwise differentiation. We need to show ξ_j is the unique minimizer of (3.16), and hence that $\xi_j = \tilde{\theta}_{n\lambda j}$. A necessary and sufficient condition for ξ_j to be that minimizer is that for all $h \in W_2^m$,

$$\lambda \int_0^1 \xi_j^{(m)}(t)h^{(m)}(t)\,dt + \int_0^1 \xi_j(t)h(t)\,dF_n(t) = h(t_{jn}). \tag{3.26}$$

The necessity follows from Proposition 2. If ξ_j satisfies the condition, then it is a minimizer since (3.16) defines a convex function $\mathscr{L}$ of ξ. We will show the unicity of any such minimizer by establishing strict convexity of $\mathscr{L}$. For ξ, $h \in W_2^m$ fixed and α a real variable

$$\frac{\partial^2}{\partial\alpha^2}\mathscr{L}(\xi + \alpha h) = 2\lambda \int_0^1 \{h^{(m)}(t)\}^2\,dt + 2\int_0^1 \{h(t)\}^2\,dF_n(t).$$

If this is zero, then $h^{(m)} = 0$ a.e., and so $h^{(m-1)}$ is a constant, since it is (absolutely) continuous. Hence, h is a polynomial of degree $m-1$ vanishing at the distinct points t_{n1}, $t_{n2}, \cdots, t_{nn}$. Since $n \ge m$, $h \equiv 0$. This proves that

$$\frac{\partial^2}{\partial\alpha^2}\mathscr{L}(\xi + \alpha h) > 0 \quad \text{for all} \quad h \in W_2^m, \quad h \not\equiv 0,$$

which implies strict convexity for $\mathscr{L}$.

To show that (3.26) holds, first note that by uniform convergence,

$$\xi_j = g_{n\lambda j} + \mathscr{R}_{n\lambda}\xi_j,$$

and since $m \ge 2$ (so that $m \le 2m-2$),

$$\begin{aligned}\lambda \int_0^1 \xi_j^{(m)}(t)h^{(m)}(t)\,dt &+ \int_0^1 \xi_j(t)h(t)\,dt \\ &= \int_0^1 \{\lambda G^{m,0}(t, t_{jn})h^{(m)}(t) + G(t, t_{jn})h(t)\}\,dt \\ &\quad + \int_0^1 \xi_j(\tau)\int_0^1 \{\lambda G^{m0}(t,\tau)h^{(m)}(t) + G(t,\tau)h(t)\}\,dt\,d(\tau - F_n(\tau)).\end{aligned} \tag{3.27}$$

Using the boundary conditions and m-fold integration by parts gives

$$\begin{aligned}\int_0^1 \{\lambda G^{m,0}(t,\tau)h^{(m)}(t) &+ G(t,\tau)h(t)\}\,dt \\ &= \int_0^1 \{(-1)^{m-1}\lambda G^{2m-1,0}(t,\tau)h'(t) + G(t,\tau)h(t)\}\,dt \\ &= (-1)^{m-1}\lambda[G^{2m-1,0}(\tau-,\tau) - G^{2m-1,0}(\tau+,\tau)]h(\tau) \\ &\quad + \left(\int_0^{\tau-} + \int_{\tau+}^1\right)\{(-1)^m\lambda G^{2m,0}(t,\tau) + G(t,\tau)\}h(t)\,dt = h(\tau),\end{aligned}$$

where the last equation follows from the definition of G as a Green's function. Substituting this result back into (3.27) gives

$$\lambda \int_0^1 \xi_j^{(m)}(t)h^{(m)}(t)\,dt + \int_0^1 \xi_j(t)h(t)\,dt = h(t_{jn}) + \int_0^1 \xi_j(t)h(t)(dt - dF_n(t)).$$

This proves ξ_j satisfies (3.26), and we need only show (3.23) to complete the proof of the theorem. If $p \le 2m - 2$, then from (3.25) and (3.21) we have that for all n sufficiently large,

$$\sup|\tilde{\theta}_{n\lambda j}^{(p)} - g_{n\lambda j}^{(p)}| \le \sum_{\nu=1}^{\infty} \sup|(\mathscr{R}_{n\lambda}^{\nu} g_{n\lambda j})^{(p)}| \le \sum_{\nu=1}^{\infty} K^{\nu} D_n^{\nu} \lambda^{-(p+2\nu+1)/2m}(L\lambda)^{\nu}$$
$$= KD_n\lambda^{-(p+3)/2m}(L\lambda)/\{1 - KD_n\lambda^{-1/m}(L\lambda)\}$$

which proves the result. □

COROLLARY 3.3. *Under the same assumptions as Theorem* 3.2, *if* $p \le 2m - 2$, *then for some constant* C_p, *depending only on p and m*,

$$\sup_t|\tilde{\theta}_{n\lambda j}^{(p)}(t)| \le C_p\lambda^{-(p+1)/2m}. \tag{3.28}$$

REMARK. The factor $(L\lambda)$ can be dispensed with in (3.21), (3.22), and (3.28) if $m \ge 3$ and $p \le 2m - 3$.

4. Periodic spline smoothing. Periodic spline smoothing was first introduced by Cogburn and Davis (1974). While these splines are seldom used in practice, they have figured prominently in the theory of smoothing splines (see e.g. Wahba, 1975 and Craven and Wahba, 1979). One reason for this is that circular symmetry greatly simplifies the calculations, as is demonstrated in the following.

PROPOSITION 4.1. *Suppose in addition to Assumption* 1 *that* $\psi' > 0$. *If* $\theta \in K^{(m)}$, $t_{jn} = j/n$ *for* $1 \le j \le n$, *and both of the following limits hold for a sequence* $\{\lambda_n\}$ *of positive numbers*:

$$\left\{\theta(u) - \int_0^1 \theta(t)\,dt\right\}\lim_{n\to\infty}\lambda_n = 0; \tag{4.1}$$

$$\lim_{n\to\infty} n\lambda_n^{1/m} = \infty; \tag{4.2}$$

then if $\lambda = \lambda_n$

$$d_n^2(\tilde{\theta}_{n\lambda}, \hat{\theta}_{n\lambda}) = o_p(E\, d_n^2(\tilde{\theta}_{n\lambda}, \theta)). \tag{4.3}$$

PROOF. The assumption $\psi' > 0$ implies strict convexity of ρ, so $\hat{\theta}_{n\lambda}$ exists uniquely. Computing the constants in Assumption 2, we clearly have $A_n = 1$, and using (3.15) and the definition of $|||\cdot|||_{n\lambda}$ gives

$$|||y_{in}|||_{n\lambda}^2 = d_n^2(\mathscr{G}_{n\lambda}^{-1}y_{in}, 0) = \frac{1}{n}\sum_{k=1}^{n}(nA_{ik}^2),$$

where the matrix $A = s^{-1}\mathscr{G}^{-1}\mathscr{T}s$ is the "influence" matrix of CW. Under the circular symmetry assumptions, A is a circulant matrix so for some constant K depending only on m

$$n\sum_k A_{ik}^2 = \text{Trace } A^2 \le K\lambda^{-1/2m},$$

(see Lemma 4.3 of CW). Hence

$$B_n \le K\lambda_n^{-1/2m}.$$

It is also proved in Section 4 of Craven and Wahba that

$$C_n = \begin{cases} O(n^{-1}\lambda_n^{-1/2m}) & \text{if } \theta \equiv \int_0^1 \theta(t)\,dt, \\ O(\lambda_n) + O(n^{-1}\lambda_n^{-1/2m}) & \text{otherwise.} \end{cases}$$

(The case when θ is constant is different since there is no bias.) It now follows that (4.1) and (4.2) imply (3.10), and hence the theorem. □

REMARK. The best obtainable rate of convergence according to Craven and Wahba (if $\theta \not\equiv$ constant) is

$$E\, d_n^2(\tilde{\theta}_{n\lambda}, \theta) = O(n^{-2m/(2m+1)}),$$

obtainable by using the deterministic sequence $\lambda = n^{-2m/(2m+1)}$. Using this in (4.3), we obtain the following asymptotic estimate as $n \to \infty$:

$$d_n^2(\tilde{\theta}_{n\lambda}, \hat{\theta}_{n\lambda}) = (E\, d_n^2(\tilde{\theta}_{n\lambda}, \theta))O_p(n^{-(4m-1)/(2m+1)}).$$

In order to eliminate the assumption of equispaced knots, and obtain theorems about convergence of the derivatives $\tilde{\theta}^{(p)}$, $\hat{\theta}^{(p)}$ to $\theta^{(p)}$, we will need to invoke Theorem 3.2. This requires the following.

LEMMA 4.2. *Let $H_\lambda(t, \tau)$ denote the Green's function for $(-1)^m\lambda D^{2m} + 1$ with periodic boundary conditions. Then, there exist constants C_{pq} such that for any $\lambda \in [0, \lambda_0]$,*

$$\sup_{t,\tau}| H_\lambda^{p,q}(t, \tau)| \le C_{p,q}\lambda^{-(p+q+1)/2m},$$

provided $p + q \le 2m - 2$, and if $p + q = 2m - 1$,

$$\sup_{t,\tau}|H_\lambda^{p,q}(t, \tau)| \le C_{p,q}\lambda^{-(p+q+1)/2m}L\lambda,$$

where $L\lambda = \max(| \log \lambda |, 1)$.

PROOF. H_λ is most conveniently given by

$$H_\lambda(t, \tau) = \textstyle\sum_{\nu=-\infty}^{\infty} \{1 + \lambda(2\pi\nu)^{2m}\}^{-1}\exp\{2\pi i\nu(t - \tau)\}. \tag{4.4}$$

This bilinear expansion may be easily obtained by the method described on page 363 of Courant and Hilbert (1953), since the eigenvalues and eigenfunctions of D^{2m} operating on $K^{(2m)}$ are $(2\pi i\nu)^{2m}$ and $\exp\{2\pi i\nu t\}$, respectively, for $\nu = \cdots, -1, 0, +1, \cdots$. It is readily verified that $H_\lambda^{p,q}$ may be computed by termwise differentiation if $p + q = 2m - 2$, and standard theorems on Fourier series in conjunction with estimates given below also allow this if $p + q = 2m - 1$, provided $t \ne \tau$ (see e.g. page 275 of Carslaw, 1930). If $r \equiv p + q \le 2m - 2$, then

$$\sup | H^{p,q} | \le \textstyle\sum_{\nu=-\infty}^{\infty} (2\pi\nu)^r\{1 + \lambda(2\pi\nu)^{2m}\}^{-1} \le 1 + 2\sum_{\nu=1}^{\infty} (2\pi\nu)^r\{1 + \lambda(2\pi\nu)^{2m}\}^{-1}. \tag{4.5}$$

Writing $\mu = 2\pi\lambda^{1/2m}$, define the step functions

$$f_\mu(x) = \textstyle\sum_{\nu=1}^{\infty} (\nu\mu)^r\{1 + (\nu\mu)^{2m}\}^{-1}I_{\nu\mu}(x)$$

for $x \ge 0$, where $I_{\nu\mu}$ is the indicator function for the interval $[(\nu - 1)\mu, \nu\mu)$. Then, for all $x \ge 0$,

$$\lim_{\mu\downarrow 0} f_\mu(x) = f(x) \equiv x^r(1 + x^{2m})^{-1},$$

and furthermore if $x^* = [r/(2m - r)]^{1/2m}$ then

$$f_\mu(x) \le \begin{cases} f(x^*) & \text{if } 0 \le x \le x^*, \\ f(x) & \text{if } x^* \le x. \end{cases}$$

Hence, by dominated convergence, as $\mu \downarrow 0$

$$\begin{aligned} \textstyle\sum_{\nu=1}^{\infty} (2\pi\nu)^r\{1 + \lambda(2\pi\nu)^{2m}\}^{-1} &= (2\pi)^r\mu^{-(r+1)} \int_0^\infty f_\mu(x)\, dx \\ &\sim (2\pi)^r\mu^{-(r+1)} \int_0^\infty f(x)\, dx, \end{aligned} \tag{4.6}$$

which proves the theorem for $p + q \le 2m - 2$.

The case $p + q = 2m - 1$ requires a more careful analysis. We begin with the well known Fourier series

$$\tfrac{1}{2} - y = \textstyle\sum_{\nu=-\infty,\nu\neq 0}^{\infty} (2\pi i\nu)^{-1}\exp(2\pi i\nu y),$$

valid if $0 < y < 1$. Writing $\mu = 2\pi\lambda^{1/2m}$ as before, we have

$$\left| H^{p,q}(t,\tau) - (-1)^{q+m}\lambda^{-1}\left(\frac{1}{2} - t + \tau\right)\right|$$

$$= \left| \textstyle\sum_{\nu=-\infty,\nu\neq 0}^{\infty} \left\{\dfrac{(-1)^q(2\pi i\nu)^{2m-1}}{1+\lambda(2\pi\nu)^{2m}} - \dfrac{(-1)^q i^{2m-1}}{\lambda(2\pi\nu)}\right\}\exp\{2\pi i\nu(t-\tau)\}\right|$$

$$\leq 2\lambda^{-1}\textstyle\sum_{\nu=1}^{\infty} (2\pi\nu)^{-1}\{1+\lambda(2\pi\nu)^{2m}\}^{-1}$$

$$\leq (\pi\lambda)^{-1}\textstyle\sum_{\nu=1}^{[1/\mu]} \nu^{-1} + (2\pi)^{-(2m+1)}\lambda^{-2}\sum_{[1/\mu]+1}^{\infty} \nu^{-(2m+1)},$$

where $[\cdot]$ denotes the greatest integer function. Now the first term in the last expression is $O(\lambda^{-1}\log(1/\lambda))$, while the second is $O(\lambda^{-1})$. The theorem is now obvious. □

As a byproduct, it is now possible to prove the following result, which considerably extends the class of knot sequences for which standard smoothing spline asymptotics are applicable.

THEOREM 4.3. *Assume $m \geq 2$. Let $\theta_{n\lambda}$ be the minimizer over $\xi \in W_2^m$ (or $\xi \in K^{(m)}$) of*

$$\frac{1}{n}\textstyle\sum_{k=1}^{n} \{z_{kn} - \xi(t_{kn})\}^2 + \lambda \displaystyle\int_0^1 \{\xi^{(m)}(t)\}^2\, dt,$$

where the data z_{kn} are given in (3.1). *Assume the errors $\{\varepsilon_{kn}\}$ are mean zero, uncorrelated, and with common variance $\sigma^2 < \infty$. Suppose that λ varies with n in such a way that $\lambda \downarrow 0$ and,* (3.21) *holds.*

(a) *If θ is a polynomial of degree at most $m - 1$,*

$$E\, d_n^2(\tilde{\theta}_{n\lambda}, \theta) = O(n\lambda^{-1/2m}),$$

as $n \to \infty$.

(b) *If θ is not a polynomial of degree at most $m - 1$, then*

$$E\, d_n^2(\tilde{\theta}_{n\lambda}, \theta) = O(\lambda) + O(n\lambda^{-1/2m}).$$

REMARKS. These rates of convergence are stated in CW, but the proof is not rigorous in one detail. However, their "heuristic" argument can be rigorized, provided the knots t_{kn} are equispaced. The conclusion also follows from the main theorem in Utreras (1980c) in conjunction with the CW argument, if $m = 2$ and the knots satisfy

$$\frac{\max_{1\leq k<n}(t_{k+1,n} - t_{kn})}{\min_{1\leq k<n}(t_{k+1,n} - t_{kn})} < C,$$

for some constant C. It should be noted that the results of CW and Utreras are slightly better than ours in the sense that they only require $n\lambda^{1/2m} \to \infty$, whereas (3.21) always implies $n\lambda^{1/m}\,|\log\lambda|^{-1} \to \infty$. However, this is not a loss of significant generality, since the optimal rate of convergence always requires $n\lambda \to \infty$ (and we only consider $m \geq 2$). For a completely rigorous version of the CW results (including non-uniform asymptotic knot distributions) see Speckman (1981).

PROOF. We recap the argument from CW. Let $\bar{\theta}_{n\lambda}$ be the smoothing spline obtained from perfect data, i.e. $\bar{\theta}_{n\lambda}$ minimizes

$$\frac{1}{n}\sum_{k=1}^{n}\{\theta(t_k)-\xi(t_k)\}^2+\lambda\int_0^1\{\xi^{(m)}(t)\}^2\,dt.$$

Then $E\,d_n^2(\tilde{\theta},\theta)$ can be decomposed into bias squared plus variance, to wit

$$E\,d_n^2(\tilde{\theta}_{n\lambda},\theta)=d_n^2(\bar{\theta}_{n\lambda},\theta)+\sigma^2\left(\frac{1}{n}\operatorname{tr}A^2\right) \tag{4.7}$$

(see page 389 of CW). Here, A denotes the influence matrix (a function of n and λ) whose (i,j)th entry is $n^{-1}\tilde{\theta}_{n\lambda j}(t_{in})$ (see (3.16)) i.e. if $\tilde{\boldsymbol{\theta}}=(\tilde{\theta}_{n\lambda}(t_{in}),\cdots,\tilde{\theta}_{n\lambda}(t_{nn}))'$ denotes the vector of values of the smooth, then $\tilde{\boldsymbol{\theta}}=A\mathbf{z}$, where $\mathbf{z}$ is the data vector. By Lemma 4.1 of CW,

$$d_n^2\,(\bar{\theta}_{n\lambda},\theta)\le\lambda\int_0^1\{\theta^{(m)}(t)\}^2\,dt.$$

Hence, the theorem is proved once we show that $\operatorname{tr}A^2=O(\lambda^{-1/2m})$.

An argument on page 400 of CW shows that it is sufficient to consider $\operatorname{tr}A_0^2$, where A_0 is almost the A matrix for periodic smoothing, denoted A_1. The only difference between A_0 and A_1 is in the eigenvalue corresponding to the constant eigenvector (i.e. $(1,1,\cdots,1)'$). This eigenvalue is 1 for A_1, and between 0 and 1 for A_0. Hence, it suffices to show that $\operatorname{tr}A_1^2=O(\lambda^{-1/2m})$. Now the eigenvalues of A_1 are all between 0 and 1, so $0\le\operatorname{tr}A_1^2\le\operatorname{tr}A_1$. By Theorem 3.2 and Lemma 4.2,

$$\operatorname{tr}A_1=\frac{1}{n}\sum_j H_\lambda(t_j,t_j)+O(D_n\lambda^{-3/2m}(L\lambda))$$

and the second term on the right hand side is $o(\lambda^{-1/2m})$ by (3.21). Also

$$\left|\frac{1}{n}\sum_j H_\lambda(t_j,t_j)-\int H_\lambda(t,t)\,dt\right|=\left|\int H_\lambda(t,t)\,d(t-F_n(t))\right|$$
$$\le D_n\sup|H_\lambda^{10}+H_\lambda^{01}|=D_nO(\lambda^{-2/2m}(L\lambda))=o(\lambda^{-1/2m}).$$

Since $\int H_\lambda(t,t)\,dt=O(\lambda^{-1/2m})$, the desired conclusion is obtained. □

REMARK. From the proof of Lemma 4.2 we have

$$\int H(t,t)\,dt=\sum_{\nu=-\infty}^{\infty}\{1+\lambda(2\pi\nu)^{2m}\}^{-1}\sim\lambda^{-1/2m}\left(\frac{1}{2\pi}\int_{-\infty}^{\infty}\frac{dx}{1+x^{2m}}\right),$$

and it follows that if (3.21) holds, then $\operatorname{tr}A_1$ is also asymptotically equivalent to the last quantity. Similar arguments show that

$$\left|\operatorname{tr}A_1^2-\int\int\{H_\lambda(t,\tau)\}^2\,dt\,d\tau\right|=O(D_n\lambda^{-2/m}(L\lambda)),$$

and

$$\int\int\{H_\lambda(t,\tau)\}^2\,dt\,d\tau\sim\lambda^{-1/2m}\left\{\frac{1}{2\pi}\int_{-\infty}^{\infty}\frac{dx}{(1+x^{2m})^2}\right\},$$

so if (3.21) is strengthened to

$$\lim_{n\to\infty}\{D_n\lambda^{-3/2m}(L\lambda)\}=0$$

then we have

$$\operatorname{tr} A_1^2 \sim \lambda^{-1/2m}\left\{\frac{1}{2\pi}\int_{-\infty}^{\infty}\frac{dx}{(1+x^{2m})^2}\right\}.$$

The next theorem partially extends this result to general M-type smoothing splines.

THEOREM 4.4. *Assume $m \geq 2$. Let $\hat\theta_n$ be the solution over $\xi \in \mathscr{K}^{(m)}$ of $\Phi_{n\lambda}(\xi) = 0$, where $\Phi_{n\lambda}$ is given by* (3.4). *Suppose in addition to Assumption* 1 *that $\psi' > 0$, so $\hat\theta_{n\lambda}$ exists uniquely.*

If $\theta \in \mathscr{K}^{(m)}$, and $\lambda_n > 0$ varies in such a way that

$$\lim_{n\to\infty} D_n\lambda_n^{-5/4m}(L\lambda_n) = \lim \lambda_n = 0, \tag{4.8}$$

then

$$d_n^2(\hat\theta_{n\lambda}, \theta) = O_p(\lambda_n) + O_p(n^{-1}\lambda_n^{-1/2m}). \tag{4.9}$$

PROOF. From the argument of Theorem 4.3, we have

$$C_n = E\, d_n^2(\tilde\theta_{n\lambda}, \theta) = O(\lambda_n) + O(n^{-1}\lambda_n^{-1/2m}). \tag{4.10}$$

Now, by (3.15)

$$B_n \equiv d_n^2\,(\mathscr{G}_{n\lambda}^{-1} y_{kn}, 0) = \max_k \frac{1}{n}\textstyle\sum_{j=1}^n \{\tilde\theta_{n\lambda k}(t_j)\}^2,$$

and by Lemma 4.2, writing $g_{n\lambda k} = H_\lambda(\cdot, t_k)$, we have

$$\frac{1}{n}\textstyle\sum_j \tilde\theta_{n\lambda k}(t_j)^2 \leq \frac{2}{n}\sum_j g_{n\lambda k}(t_j)^2 + \frac{2}{n}\sum_j \{\tilde\theta_{n\lambda k}(t_j) - g_{n\lambda k}(t_j)\}^2$$

$$\leq 2\int_0^1 \{H(t_k, u)\}^2\, du + 2\int \{H(t_k, u)\}^2\, d(u - F_n(u)) + 2KD_n^2\lambda^{-3/m}(L\lambda)^2.$$

The third term in the last expression is

$$\lambda^{-1/2m}\{D_n\lambda^{-5/4m}(L\lambda)\} = o(\lambda^{-1/2m})$$

by (4.8). Lemma 4.2 implies

$$\int \{H(t_k, u)\}^2\, d(u - F_n(u)) \leq 2D_n \sup|\, H(t, u)H^{01}(t, u)|$$

$$= O(D_n\lambda^{-3/2m}(L\lambda)) = o(\lambda^{-1/2m})$$

by (4.8). Finally,

$$\int_0^1 \{H_\lambda(t, u)\}^2\, du = \textstyle\sum_{\nu=-\infty}^{\infty} \{1 + \lambda(2\pi\nu)^{2m}\}^{-2} \sim (2\pi)^{-1}\lambda^{-1/2m}\displaystyle\int_{-\infty}^{\infty}(1 + x^{2m})^{-2}\, dx \tag{4.11}$$

as $\lambda \downarrow 0$. Hence $B_n = O(\lambda_n^{-1/2m})$. Since $A_n = 1$ for all n, we have

$$A_n^2 B_n C_n = O(\lambda_n\lambda_n^{-1/2m}) + O(n\lambda_n^{-1/m}) \to 0$$

as $n \to \infty$ by (4.8). Also

$$n^{-1}A_n B_n = O(n^{-1}\lambda^{-1/2m}) \to 0$$

as $n \to \infty$. Hence, by Theorem 3.1,

$$\sup_\lambda d_n^2(\hat{\theta}_{n\lambda}, \tilde{\theta}_{n\lambda}) = O_p(C_n),$$

which, together with (4.10), completes the proof. □

REMARKS. (1) If θ is a constant, then there is no bias, and the conclusion of the theorem may be simplified to

$$\sup_\lambda d_n^2(\hat{\theta}_{n\lambda}, \theta) = O_p(n^{-1}\lambda^{-1/2m}).$$

Of course, if it was known that $\theta \equiv$ constant, then one would set $\lambda = \infty$ (which forces $\hat{\theta}$ to be constant) and obtain $O_p(n^{-1})$ rate of convergence.

(2) Suppose λ_n is a power of n, say $\lambda_n = n^\alpha$. Also assume $D_n \approx n^{-1}$, as in the equispaced knots case. Then the assumptions of the theorem are equivalent to requiring $-4m/5 < \alpha < 0$.

(3) According to Speckman (1981b), the optimal convergence rate for the integrated mean squared error (abbreviated IMSE, meaning either $E\|\tilde{\theta} - \theta\|_2^2$ or $E\, d_n^2(\hat{\theta}, \theta)$) for a linear estimate of arbitrary $\theta \in \mathscr{K}^{(m)}$ is

$$\text{IMSE} = O(n^{-2m/(2m+1)}).$$

If we put $\lambda_n = n^{-2m/(2m+1)}$ in part (a), then the upper bound will be of this optimal order (note that this choice for λ_n is permitted by (4.8)).

5. Natural cubic spline smoothing. We now wish to extend the results of the previous section on periodic smoothing to smoothing by natural splines. Because of the complications involved in estimating the Green's function and its derivatives, we must unfortunately restrict ourselves to the case $m = 2$. If H_λ is given by (4.4) with $m = 2$, then the Green's function for the natural boundary conditions is

$$G_\lambda(t, \tau) = H_\lambda(t, \tau) - \textstyle\sum_{p=2}^{3} H^{p0}(0, \tau)\{\eta_{p0}(t) + \eta_{p1}(t)\},$$

where η_{pj} is the solution to the boundary value problem

$$\lambda\eta^{(4)} + \eta = 0, \quad \eta^{(q)}(k) = \delta_{pq}\delta_{jk} \quad \text{for} \quad q = 2, 3 \quad \text{and} \quad k = 0, 1.$$

By directly solving the differential equation, it can be shown that if $\gamma = 2^{-1/2}\lambda^{-1/4}$, then

$$\eta_{02}(t) = \lambda^{1/2}e^{-\gamma t}\{\cos(\gamma t) - \sin(\gamma\tau)\} + e^{-2\gamma}\Delta_{02} \tag{5.1}$$

$$\eta_{03}(t) = 2^{1/2}\lambda^{3/4}e^{-\gamma t}\cos(\gamma t) + e^{-2\gamma}\Delta_{03}, \tag{5.2}$$

where Δ_{0p} is a sum of products of $e^{\pm\gamma t}$ and powers of $\sin(\gamma t)$, $\cos(\gamma t)$. By the symmetry of the boundary value problem, it is easy to see that

$$\eta_{1p}(t) = (-1)^p\eta_{0p}(1 - t),$$

so that η_{12} and η_{13} are essentially defined by (5.1) and (5.2) as well. The estimates on G_λ needed for Theorem 3.2 now follow from these formulae in conjunction with Lemma 4.2.

THEOREM 5.1. *Assume $m = 2$. Let $\tilde{\theta}_{n\lambda}$ be the solution over $\xi \in W_2^2$ of $\Phi_{n\lambda}(\xi) = 0$, where $\Phi_{n\lambda}$ is given by* (3.4). *Suppose in addition to Assumption* 1 *that $\psi' > 0$. Let $\lambda_n > 0$ and suppose $\theta \in W_2^2$. If*

$$\lim_{n\to\infty} D_n\lambda_n^{-5/8}(L\lambda_n) = \lim_{n\to\infty}\lambda_n = 0, \tag{5.3}$$

then

$$d_n^2(\hat{\theta}_{n\lambda}, \theta) = O_p(\lambda_n) + O_p(n^{-1}\lambda_n^{-1/4}).$$

PROOF. In view of Theorem 4.3 and the remarks above, it is sufficient to show that

$$\int_0^1 \{G_\lambda(t_{kn}, u)\}^2 \, du = O(\lambda^{-1/4})$$

as $\lambda \downarrow 0$ (see the proof of Theorem 4.4). It is easily checked from (5.1) and (5.2) that

$$\int_0^1 \{\eta_{0p}(u)\}^2 \, du = O(\lambda^{(2p+1)/4}),$$

and since

$$\sup | H^{p0} |^2 = O(\lambda^{-2(p+1)/4}),$$

it follows from Lemma 4.2 that the required inequality holds. □

6. Estimates of derivatives. In this section, we obtain rates of convergence for the IMSE in estimating $\theta^{(p)}$ by $\hat{\theta}^{(p)}$. Again, we restrict ourselves to periodic smoothing splines.

We shall need two lemmas which give asymptotic estimates on the integrated bias squared and variance for periodic smoothing spline estimates of the pth derivative of θ. First, we extend the definition of $\mathscr{K}^{(m)}$ to arbitrary (noninteger) positive real numbers $r > 0$:

$$\mathscr{K}^{(r)} = \{ f \in L^2[0, 1] : \textstyle\sum_{\nu=-\infty}^{\infty} | f_\nu |^2 \, | 2\pi\nu |^{2r} < \infty \},$$

where

$$f_\nu = \int_0^1 f(t) \exp(-2\pi i \nu t) \, dt$$

are the Fourier coefficients of f.

LEMMA 6.1. *Assume $n \geq m \geq 2$, and that $\{\lambda_n\}$ is such that $\lambda_n \to 0$ and* (3.21) *holds. Suppose $\theta \in \mathscr{K}^{(r)}$ for some $r > 2$. Let $\bar{\theta}_{n\lambda}$ be the minimizer of*

$$\frac{1}{n} \textstyle\sum_{k=1}^{n} \{\theta(t_{kn}) - \xi(t_{kn})\}^2 + \lambda_n \displaystyle\int_0^1 \{\xi^{(m)}(t)\}^2 \, dt,$$

with respect to $\xi \in \mathscr{K}^{(m)}$. If $0 \leq p \leq \min(r, 2m-2)$, and $\max(p, 3/2) \leq q \leq \min(r, 2m+p)$, then for all n sufficiently large

$$\| \theta^{(p)} - \bar{\theta}_{n\lambda}^{(p)} \|_2^2 \leq \| \theta^{(q)} \|_2^2 \, \lambda_n^{(q-p)/m} \{K_1 + K_2 D_n \lambda^{-5/4m}(L\lambda)\}^2,$$

where K_1 and K_2 are constants independent of n, and depending on θ only through r.

PROOF. First note that

$$\bar{\theta}_{n\lambda} = \frac{1}{n} \textstyle\sum_{k=1}^{n} \theta(t_{kn}) \tilde{\theta}_{n\lambda k}$$

where $\tilde{\theta}_{n\lambda j}$ is the minimizer over $\xi \in \mathscr{K}^{(m)}$ of (3.16). We will drop the subscripts n and λ. By Theorem 3.2,

$$\begin{aligned} \bar{\theta} &= \frac{1}{n} \textstyle\sum_{k=1}^{n} \theta(t_k) \sum_{\nu=0}^{\infty} \mathscr{R}^\nu H(\cdot, t_k) = \sum_\nu \mathscr{R}^\nu \left\{ \displaystyle\int H(\cdot, \tau) \theta(\tau) \, dF_n(\tau) \right\} \\ &= \textstyle\sum_\nu \mathscr{R}^\nu \left\{ \displaystyle\int H(\cdot, \tau)\theta(\tau) \, d\tau - \mathscr{R}\theta \right\} = \theta + \textstyle\sum_\nu \mathscr{R}^\nu \phi_\lambda, \end{aligned} \tag{6.1}$$

where

$$\phi_\lambda = \int H(\cdot, \tau)\theta(\tau)\, d\tau - \theta,$$

and it is necessary to justify the last line of (6.1) by proving convergence of the two series $\Sigma \mathscr{R}^\nu \int H(\cdot, \tau)\theta(\tau)\, d\tau$ and $\Sigma \mathscr{R}^\nu\theta$. By a generalization of the argument used for (3.25), one can show that $\forall\ \xi \in C^1[0, 1]$ and any nonnegative integers ν and $p \le 2m - 2$,

$$\| (\mathscr{R}^\nu_{n\lambda}\xi)^{(p)} \|_\infty \le \{C_1 D_n \lambda^{-1/m}(L\lambda)\}^\nu C_2 \lambda^{-(p-1)/2m}(\lambda^{-1/2m} \| \xi \|_\infty + \| \xi' \|_\infty),$$

where C_1 and C_2 are constants depending only on p. Hence, by (3.21), both of the aforementioned series converge absolutely and uniformly for all n sufficiently large provided $\theta \in C^1$, which is the case if $\theta \in \mathscr{K}^{(r)}$ with $r > 2$. This justifies the calculation of (6.1), and furthermore yields that the series in the final expression of (6.1) may be differentiated term by term p times if p does not exceed $\min(2m - 2, r)$, and that for all n sufficiently large,

$$\sum_{\nu=1}^{\infty} \|(\mathscr{R}^\nu_{n\lambda}\phi_\lambda)^{(p)} \|_\infty \le C_3 D_n \lambda^{-(p+1)/2m}(L\lambda)(\lambda^{-1/2m} \| \phi_\lambda \|_\infty + \| \phi'_\lambda \|_\infty), \tag{6.2}$$

where $C_3 = 2C_1C_2$. Now for $p = 0$ or 1, and $p + \frac{1}{2} < q \le r$,

$$\begin{aligned} \| \phi_\lambda^{(p)} \|_\infty &\le \sum_{\nu=-\infty}^{\infty} | \theta_\nu | \lambda | 2\pi\nu |^{2m+p}/\{1 + \lambda(2\pi\nu)^{2m}\} \\ &\le [\sum_\nu \lambda^2 | 2\pi\nu |^{4m+2p-2q}/\{1 + \lambda(2\pi\nu)^{2m}\}^2]^{1/2}\{\sum_\nu | \theta_\nu |^2 | 2\pi\nu |^{2q}\}^{1/2} \\ &\sim C_4 \lambda^{(2q-2p-1)/4m} \| \theta^{(q)} \|_2, \end{aligned} \tag{6.3}$$

as $n \to \infty$, where

$$C_4^2 = \frac{1}{\pi} \int_0^\infty \frac{x^{4m+2p-2q}}{(1 + x^{2m})^2}\, dx.$$

This is finite provided $q > \frac{3}{2} \ge p + \frac{1}{2}$ for $p = 0, 1$. Finally, the last piece we shall need for the Lemma is the estimate

$$\begin{aligned} \| \phi_\lambda^{(p)} \|_2^2 &= \sum_{\nu=-\infty}^{\infty} | \theta_\nu |^2 | 2\pi\nu |^{2p} | \lambda(2\pi\nu)^{2m}/\{1 + \lambda(2\pi\nu)^{2m}\}|^2 \\ &= \sum_\nu | \theta_\nu |^2 | 2\pi\nu |^{2q} | \lambda(2\pi\nu)^{2m+p-q}/\{1 + \lambda(2\pi\nu)^{2m}\}|^2 \\ &\le \lambda^{(q-p)/m} \| \theta^{(q)} \|_2^2 C_5, \end{aligned} \tag{6.4}$$

where

$$C_5 = \sup_{x\ge0}\{x^{2m+p-q}/(1 + x^{2m})\}^2,$$

which is finite if $p \le q \le 2m + p$. Now collect (6.2) through (6.4) and put them into (6.1) to obtain that for all n sufficiently large

$$\begin{aligned} \| \bar{\theta}_{n\lambda}^{(p)} - \theta^{(p)} \|_2 &\le \| \phi \|_2 + \| \sum_{\nu=1}^{\infty} \mathbb{R}^\nu\phi \|_\infty \\ &\le C_5 \| \theta^{(q)} \|_2 \lambda^{(q-p)/2m} + C_3 D_n \lambda^{-(p+1)/2m}(L\lambda)(3C_4\lambda^{(2q-3)/4m} \| \theta^{(q)} \|_2), \end{aligned}$$

which proves the lemma if we take $K_1 = C_5$ and $K_2 = 3C_3C_4$.

LEMMA 6.2. *Let $\varepsilon_1, \varepsilon_2, \cdots$ be independent random variables with mean* 0 *and common variance $\sigma^2 < \infty$. Assume $n \ge m \ge 2$, and that $\lambda \to 0$ and* (3.21) *holds. Then, if $0 \le p \le 2m - 2$,*

$$E \left\| \frac{1}{n} \sum_{k=1}^{n} \varepsilon_k \tilde{\theta}_{n\lambda k} \right\|_2^2 \le K\sigma^2 n^{-1}\{\lambda^{-(2p+1)/2m} + D_n^2 \lambda^{-(p+3)/m}(L\lambda)^2\},$$

where K depends only on p and m.

PROOF. Fubini's theorem implies

$$E \parallel n^{-1} \textstyle\sum_{k=1}^{n} \varepsilon_k \tilde{\theta}_k^{(p)} \parallel_2^2 = n^{-2}\sigma^2 \sum_k \parallel \tilde{\theta}_k^{(p)} \parallel_2^2$$
$$\leq 2n^{-2}\sigma^2\{\textstyle\sum_k \parallel g_k^{(p)} \parallel_2^2 + \sum_k \parallel \tilde{\theta}_k^{(p)} - g_k^{(p)} \parallel_2^2\}.$$

Direct calculation using (4.4) yields

$$\parallel g_k^{(p)} \parallel_2^2 = \textstyle\sum_{\nu=-\infty}^{\infty} |(2\pi\nu)^p/\{1 + \lambda(2\pi\nu)^{2m}\}|^2,$$

and the argument used in proving (4.6) shows that the right hand side is asymptotically equivalent to

$$(2\pi)^{-1}\lambda^{-(2p+1)/2m} \int_0^\infty \{x^p/(1 + x^{2m})\}^2 \, dx,$$

as $\lambda \downarrow 0$. Now by (3.23),

$$\parallel \tilde{\theta}_k^{(p)} - g_k^{(p)} \parallel_2^2 \leq KD_n^2 \lambda^{-(p+3)/m}(L\lambda)^2$$

where K depends only on m. This completes the proof. □

THEOREM 6.3. *Let $\theta \in \mathscr{K}^{(r)}$ for some $r > 2$. Take $p \leq \min(r, 2m - 2)$. Suppose the following hold*:

$$2p + \tfrac{1}{2} < q = \min(r, 2m + p),$$

and

$$\lim_n \lambda_n = \lim_n n^{-1}\lambda_n^{-(2p+1)/m} = \lim_n D_n \lambda_n^{-5/4m}(L\lambda) = 0.$$

Then

$$\parallel \hat{\theta}_{n\lambda}^{(p)} - \theta^{(p)} \parallel_2^2 = O_p(\lambda_n^{(q-p)/m}) + O_p(n^{-1}\lambda^{-(2p+1)/2m}).$$

PROOF. Let $\parallel \xi \parallel_n^2 = d_n^2(\xi, 0) + \parallel \xi^{(p)} \parallel_2^2$. Now apply Lemmas 6.1 and 6.2 and note that the fact that $D_n\lambda^{-5/4m}(L\lambda) \to 0$ allows us to ignore the second terms in the upper bounds of both lemmas. This yields

$$C_n = O(\lambda_n^{(q-p)/m}) + O(n^{-1}\lambda_n^{-(2p+1)/2m}). \tag{6.5}$$

Also, if $p \leq 2m - 2$, by Theorem 3.2, we can estimate B_n as follows:

$$\parallel \tilde{\theta}_{n\lambda k}^{(p)} \parallel_2^2 \leq 2 \parallel g_{n\lambda k}^{(p)} \parallel_2^2 + 2 \parallel \tilde{\theta}_{n\lambda k}^{(p)} - g_{n\lambda k}^{(p)} \parallel_2^2$$
$$= 2 \textstyle\sum_{\nu=-\infty}^{\infty} (2\pi\nu)^{2p}\{1 + \lambda(2\pi\nu)^{2m}\}^{-1} + O(D_n^2\lambda^{-(p+3)/m}(L\lambda)^2)$$
$$= O(\lambda^{-(2p+1)/2m}).$$

Finally, $A_n \leq 1$ for all n, so

$$A_n^2 B_n C_n = O(\lambda_n^{(q-p)/m}\lambda_n^{-(2p+1)/2m}) + O(n^{-1}\lambda_n^{-(2p+1)/m}),$$

which tends to 0 by hypothesis. Also,

$$n^{-1}A_n B_n = O(n^{-1}\lambda^{-(2p+1)/2m}) \to 0.$$

The result now follows by Theorem 3.1 and (6.5). □

REMARKS. (1) If θ is constant, then the bias term $O(\lambda_n^{(q-p)/m})$ may be deleted.
(2) If $\lambda_n = n^\alpha$ and $D_n \approx n^{-1}$, then the limit assumptions reduce to the requirement that

$$-\min\{m/(2p + 1), \quad 4m/5\} < \alpha < 0.$$

(3) Under the conditions on α cited in (2), the best obtainable upper bound on the rate of convergence is

$$E\|\hat{\theta}^{(p)} = \theta^{(p)}\|_2^2 = O(n^{-2(m-p)/(2m+1)})$$

given by $\alpha = -2m/(2m+1)$, and with the provisions that $m \geq 2$, and $p < (m-1)/2$. This upper bound agrees with the optimal *pointwise* convergence rate of arbitrary derivative estimates obtained for a slightly different model by Stone (1980).

Acknowledgments. The author is deeply indebted to Professor Grace Wahba, who suggested this problem to him, and also to Paul Speckman and David Ragozin, with whom he had helpful discussions, however brief. The author also wishes to thank a referee who kindly pointed out several related references.

REFERENCES

ARONSZAJN, N. (1950). Theory of reproducing kernels. *Trans. Amer. Math. Soc.* **68** 337–404.

CARSLAW, H. (1930). *Introduction to the Theory of Fourier's Series and Integrals*, third revised edition. Dover, New York.

COGBURN, R. and DAVIS, H. (1974). Periodic splines and spectral estimation. *Ann. Statist.* **2** 1108–1126.

CRAVEN, P. and WAHBA, G. (1979). Smoothing noisy data with spline functions: estimating the correct degree of smoothing by the method of generalized cross validation. *Numerische Math.* **31** 377–403.

GRENANDER, U. (1981). *Abstract Inference.* Wiley, New York.

GREVILLE, T. (1957). On smoothing a finite table: a matrix approach. *J. Soc. Indust. Appl. Math.* **5** 137–154.

GREVILLE, T. (1969). Introduction to spline functions. In *Theory and Applications of Spline Functions.* Academic, New York.

HUBER, P. (1973). Robust regression: asymptotics, conjectures, and Monte Carlo. *Ann. Statist.* **1** 799–821.

HUBER, P. (1979). Robust smoothing. In *Robustness in Statistics*, ed. Launer and Wilkinson. Academic, New York.

KALLIANPUR, G. (1971). Abstract Wiener processes and their reproducing kernel Hilbert spaces. *Z. Wahrsch. verw. Gebiete* **17** 113–123.

KIMELDORF, G. and WAHBA, G. (1970a). A correspondence between Bayesian estimation on stochastic processes and smoothing by splines. *Ann. Math. Statist.* **41** 495–502.

KIMELDORF, G. and WAHBA, G. (1970b). Spline functions and stochastic processes. *Sankhyā Ser. A* **32** 173–180.

LARKIN, F. (1972). Gaussian measures in Hilbert space and applications in numerical analysis. *Rocky Mountain J. of Math.* **2** 379–421.

LEONARD, T. (1978). Density estimation, stochastic processes, and prior information. *J. Roy. Statist. Soc. B***40** 113–146.

RAGOZIN, D. (1981). Error bounds for derivative estimates based on spline smoothing of exact or noisy data, preprint from Dept. of Math., Univ. of Washington. *J. Approx. Theory*, to appear.

REINSCH, C. (1967). Smoothing by spline functions. *Numer. Math.* **24** 383–393.

RICE, J. and ROSENBLATT, M. (1981a). Integrated mean squared error of a smoothing spline. *J. Approx. Theory* **33** 353–369.

RICE, J. and ROSENBLATT, M. (1981b). Smoothing splines: regression, derivatives, and deconvolution. Unpublished manuscript.

SCHOENBERG, I. (1964). Spline functions and the problem of graduation. *Proc. Nat. Acad. Sci. U.S.A.* **52** 947–950.

SPECKMAN, P. (1981a). The asymptotic integrated mean square error for smoothing noisy data by splines. University of Oregon preprint.

SPECKMAN, P. (1981b). Spline smoothing and optimal rates of convergence in nonparametric regression models. University of Oregon preprint.

STONE, C. (1980). Optimal rates of convergence for nonparametric estimators. *Ann. Statist.* **8** 1348–1360.

UTRERAS, F. (1980a). Sur le choix du parametre d'adjustment dans le lissage par functions spline. *Numer. Math.* **34** 15–28.

UTRERAS, F. (1980b). Natural spline functions, its (sic) associated eigenvalue problem. Universidad de Chile preprint.

UTRERAS, F. (1980c). On the eigenvalue problem associated with cubic splines: the arbitrarily spaced knots case. Universidad de Chile preprint.

UTRERAS, F. (1981). On computing robust splines and applications. *SIAM J. Sci. Stat. Comp.* **2** 153–163.

WAHBA, G. (1975). Smoothing noisy data with spline functions. *Numerische Math.* **24** 383–393.

WHITTAKER, E. T. (1923). On a new method of graduation. *Proc. Edinburgh Math. Soc.* **41** 63–75.

DEPARTMENT OF STATISTICS
UNIVERSITY OF WISCONSIN
1210 WEST DAYTON STREET
MADISON, WISCONSIN 53706