The Annals of Statistics
1985, Vol. 13, No. 3, 984–997

SPLINE SMOOTHING IN REGRESSION MODELS AND ASYMPTOTIC EFFICIENCY IN L_2

BY MICHAEL NUSSBAUM

Akademie der Wissenschaften der DDR, Berlin

For nonparametric regression estimation on a bounded interval, optimal rates of decrease for integrated mean square error are known but not the best possible constants. A sharp result on such a constant, i.e., an analog of Fisher's bound for asymptotic variances is obtained for minimax risk over a Sobolev smoothness class. Normality of errors is assumed. The method is based on applying a recent result on minimax filtering in Hilbert space. A variant of spline smoothing is developed to deal with noncircular models.

1. Introduction and main result. There is an important class of nonparametric curve estimation problems where asymptotic efficiency is usually formulated in terms of "optimal rates of convergence." Typically, an estimator is called asymptotically optimal if its risk decreases with rate $n^{-\varepsilon}$ as sample size n tends to infinity, and if a better rate cannot be achieved by any estimator. More precisely, one shows that the risk of the given estimator does not exceed $C_1 n^{-\varepsilon}(1 + o(1))$, and that no estimator can be better than $C_2 n^{-\varepsilon}(1 + o(1))$, where C_1, C_2 are some constants. On the other hand, in regular parametric estimation problems a much stronger result is available: not only it is known that n^{-1} is an optimal rate of convergence for quadratic loss, but one knows both constants to be equal to Fisher's bound for asymptotic variances. In both situations a minimax risk over a parameter set is considered; for curve estimation this is mostly a class of smooth functions.

Let us remark that for an appropriate choice of the loss, this class of curve estimation problems can also be treated in a limit experiment framework where results are analogous to the parametric case. However this amounts to estimating the distribution function or its regression analog (cp. Millar, 1982). We are concerned here with a squared L_2-loss where the rate n^{-1} is not attained.

Certainly it is desirable to strengthen the optimal rate results by finding constants $C_1 = C_2$, and to obtain thus an analog of Fisher's bound (or of the Hajek-LeCam asymptotic minimax theorem). Until recently, for problems like density estimation under smoothness information, such a possibility seemed remote. However a solution was found by Pinsker (1980) for a filtering problem in Hilbert space. It is essential there that loss is squared Hilbertian distance, and that the parameter space is an ellipsoid. Note that smoothness classes of the Sobolev type can be described as ellipsoids. The essence of Pinsker's method consists in showing that minimax linear estimators over an ellipsoid are asymptotically minimax in the class of all estimators. Such an exact asymptotic

Received October 1982; revised December 1984.
AMS 1980 *subject classifications*. Primary 62G20, 62G05; secondary 41A15, 65D10.
Key words and phrases. Smooth nonparametric regression, asymptotic minimax risk, linear spline estimation, boundary effects.

minimax bound, as distinct from the weaker rate optimality, has subsequently been established for other models, including probability density estimation (Ibragimov and Hasminski, 1981; Efroimovich and Pinsker, 1982).

The first result of this type in a regression context is due to Golubev (1984). However it concerns a somewhat nonstandard model with an asymptotics for an expanding interval of observation. The purpose of this paper is to establish such a bound for a nonparametric regression model on a fixed interval.

Consider observations

$$y_{jn} = f(x_{jn}) + \xi_j, \quad x_{jn} \in [0, 1], \quad j = 1, \cdots, n,$$

where $\{\xi_j\}$ are independent random variables with distribution $N(0, 1)$, and the function f is to be estimated. The design is assumed to be equidistant:

$$x_{jn} = (j - 1)/(n - 1), \quad j = 1, \cdots, n.$$

Let $L_2 = L_2([0, 1])$ be the Hilbert space of square-integrable functions on [0, 1], and let $\| \cdot \|$ denote the usual norm therein. We shall also use the seminorm $\| \cdot \|_n$, defined for real-valued functions f by

$$\| f \|_n^2 = n^{-1} \textstyle\sum_{j=1}^n f^2(x_{jn}).$$

Let, for natural m and $f \in L_2$, $D^m f$ denote the derivative of order m in the distributional sense, and let

$$W_2^m = \{f \in L_2;\ D^m f \in L_2\}$$

be the corresponding Sobolev space on the unit interval. The nonparametric class of functions to which f is assumed to belong is

$$W_2^m(P) = \{f \in W_2^m;\ \| D^m f \|^2 \le P\}$$

for given m and $P > 0$. Let $\mathscr{F}_n$ be the class of all estimators of f for given n, i.e., measurable mappings $\hat{f}$: $\mathbb{R}^n \times [0, 1] \to \mathbb{R}$.

THEOREM 1. *Let* $\Theta = W_2^m(P)$ *and*

$$\gamma(m, P) = (P(2m + 1))^{1/(2m+1)}(m/\pi(m + 1))^{2m/(2m+1)}.$$

Then

(i) $\lim_n \inf_{\hat{f} \in \mathscr{F}_n} \sup_{f \in \Theta} n^{2m/(2m+1)} E_f \| \hat{f} - f \|_n^2 = \gamma(m, P)$.

(ii) *Relation* (i) *holds also with the seminorm* $\| \cdot \|_n$ *replaced by the* L_2*-norm* $\| \cdot \|$.

To comment, observe that the rate of decay of the unnormed minimax risk, for a priori smoothness m, is known to be $n^{-2m/(2m+1)}$. The theorem specifies the constants C_1, C_2 in the optimal rate result as $C_1 = C_2 = \gamma(m, P)$. In Section 4 an estimator will be exhibited which attains the above bound, and which is hence optimal in a stronger sense than the usual rate-optimal ones. These results are based on applying the optimal filtering method of Pinsker (1980) to the regression model. We introduce a variant of spline smoothing to attain the new risk bound in the context of discrete measurements and noncircularity.

This paper is organized as follows. In Section 2 we review the basic optimal filtering result. In Section 3 the tools necessary for the regression case are developed, in particular some eigenvalue estimates in spline theory. Section 4 uses these to achieve the proof of the main theorem 1. Some discussion then follows.

Throughout the paper we shall use the convention that C denotes a positive constant independent of n whose value may change at each occurrence, even on the same line.

For any number x let $[x]$ denote its entire part and $(x)_+$ its nonnegative part. The symbol $\|\cdot\|$ denotes both standard vector and L_2 norms. Frequently but not as a rule the dependence subscript n will be dropped from notation.

2. Minimax filtering in Gaussian white noise. Let us summarize the basic result of Pinsker (1980). Suppose observations

$$\eta_j = \vartheta_j + n^{-1/2}\xi_j, \quad j \in \mathbb{Z},$$

where $\{\xi_j\}$ is discrete Gaussian white noise with unit intensity and $\vartheta = \{\vartheta_j\}$ is an unknown parameter. Assume that $\vartheta \in \Theta$, where Θ is an ellipsoid:

$$\Theta = \{\vartheta \in \mathbb{R}^{\mathbb{Z}}; \textstyle\sum_j a_j\vartheta_j^2 \leq P\}.$$

If $\Theta \subset l_2$ then the model is equivalent to a stochastic differential

$$d\eta(t) = \vartheta(t)\,dt + n^{-1/2}\,dW(t), \quad t \in [0, 1],$$

ϑ being a function from L_2 with coordinates in Θ, $W(t)$ being the standard Wiener process. The problem of nonparametric estimation of ϑ in this model, with a low noise asymptotics $n \to \infty$, has many traits in common with nonparametric density and regression estimation. For optimal rates of convergence, see Ibragimov and Hasminski (1981), Chapter 7. If $a_j = (2\pi j)^{2m}$ and ϑ_j are the classical Fourier coefficients of ϑ, then Θ can be identified with a periodic Sobolev class in L_2, i.e.,

$$\Theta = \tilde{W}_2^m(P) := \{f \in W_2^m(P); D^l f(0) = D^l f(1), l = 0, \cdots, m-1\},$$

in which m periodic boundary conditions are present. For this parameter space, Pinsker (1980) proved

$$\lim_n \inf_{\hat\vartheta} \sup_{\vartheta\in\Theta} n^{2m/(2m+1)} E_\vartheta \|\hat\vartheta - \vartheta\|^2 = \gamma(m, P)$$

(infimum over all estimators). This is a special case of a result for general ellipsoids. We find it convenient to state it for the following model with a finite number of observations:

$$\eta_j = \vartheta_j + n^{-1/2}\xi_j, \quad j = 1, \cdots, n. \tag{2.1}$$

Let a double array $\{a_{jn}, j = 1, \cdots, n\}_{n\in\mathbb{N}}$ be given such that $0 \leq a_{1n} \leq a_{2n} \leq \cdots \leq a_{nn}$, and also a number $P > 0$. Define an ellipsoid in $\mathbb{R}^n$ by

$$\Theta_n = \{x \in \mathbb{R}^n; \textstyle\sum_{j=1}^n a_{jn}x_j^2 \leq P\}. \tag{2.2}$$

Consider the estimation problem for ϑ, assuming $\vartheta \in \Theta_n$, for quadratic loss in $\mathbb{R}^n$, and the minimax risk over all estimators. We present a proof of the optimal filtering result which is inspired by Golubev (1982). The conditions are adapted to our case where the a_{jn} are close to but not equal to $(\pi j)^{2m}$.

Define a function g and a number μ by

$$g(x) = (1 - (\pi x)^m)_+ (\operatorname{sgn} x)_+, \quad x \in \mathbb{R}, \tag{2.3}$$
$$\mu^{2m+1} P = \int g(x)(1 - g(x))\, dx.$$

We calculate

$$\gamma(m, P) = \mu^{-1} \int g(x)\, dx. \tag{2.4}$$

Henceforth in the paper we put $k = n^{1/(2m+1)}$, $\tilde{k} = [k \log k]$.

THEOREM 2.1. *Suppose there is a sequence* $\{\delta_j\}$, *not depending on* n, *such that* $\lim_j \delta_j = 1$ *and*

$$\sup_{1 \le j \le \tilde{k}} \delta_j a_{jn} (\pi j)^{-2m} \le 1 + o(1), \quad n \to \infty.$$

Then

$$r := \liminf_n \inf_{\hat{\vartheta}} \sup_{\vartheta \in \Theta_n} k^{2m} E_\vartheta \| \hat{\vartheta} - \vartheta \|^2 \ge \gamma(m, P).$$

PROOF. Let τ and ε be real numbers from (0, 1), and define $\tilde{\mu}$ by

$$\tilde{\mu}^{2m+1} \tau P = \int g(x)(1 - g(x))\, dx. \tag{2.5}$$

Consider functions h, p, q on $\mathbb{R}$

$$h(x) = \tilde{\mu}^{-2m} g(x\tilde{\mu})(1 - g(x\tilde{\mu})),$$
$$p(x) = \max(\varepsilon, (\pi x)^{2m}), \quad q(x) = p^{-1}(x) h(x).$$

By virtue of (2.5), we have

$$\int h(x)\, dx = \tau P. \tag{2.6}$$

Define also

$$\tilde{\Theta}_n = \{x \in \mathbb{R}^n;\ \textstyle\sum_{j=1}^{\tilde{k}} p(jk^{-1}) x_j^2 \le \tau P,\ x_j = 0,\ j = \tilde{k} + 1, \cdots, n\}.$$

Now the condition of the theorem implies that for $n > C$

$$a_{jn} \le \tau^{-1} k^{2m} p(jk^{-1}), \quad j = 1, \cdots, \tilde{k}, \quad k^{-m} \tilde{\Theta}_n \subset \Theta_n.$$

Define $t = k^m \vartheta$, and consider the estimation problem for t, under an assumption $t \in \tilde{\Theta}_n$, from observations

$$\tilde{\eta}_j := k^m \eta_j = t_j + k^{-1/2} \xi_j, \quad j = 1, \cdots, n. \tag{2.7}$$

Then

$$r \ge \liminf_n \inf_{\hat{t}} \sup_{t \in \tilde{\Theta}_n} E_t \| \hat{t} - t \|^2. \tag{2.8}$$

Let $\delta \in (0, 1)$ and Q_n be the distribution of the random n-vector u having independent components

$$u_j \sim N(0, \delta k^{-1} q(jk^{-1})), \quad j = 1, \cdots, \tilde{k}, \quad u_j \equiv 0, \quad j = \tilde{k} + 1, \cdots, n.$$

Let us demonstrate that for the complement of $\tilde{\Theta}_n$

$$Q_n(\tilde{\Theta}_n^c) = o(1), \quad n \to \infty. \tag{2.9}$$

Let ρ be the measure having mass k^{-1} in points jk^{-1}, $j = 1, \cdots, \tilde{k}$. Since h is continuous with compact support, we obtain from (2.6)

$$\int h \, d\rho \to \int h(x) \, dx = \tau P, \quad \int h^2 \, d\rho \to \int h^2(x) \, dx < \infty.$$

Hence, for $\beta_j \sim N(0, 1)$, independent,

$$\begin{aligned} Q_n(\tilde{\Theta}_n^c) &= \Pr(k^{-1} \textstyle\sum_{j=1}^{\tilde{k}} h(jk^{-1}) \delta \beta_j^2 > \tau P) \\ &= \Pr(k^{-1} \textstyle\sum_{j=1}^{\tilde{k}} h(jk^{-1})(\beta_j^2 - 1) > \tau P \delta^{-1} - \displaystyle\int h d\rho) \\ &\leq 2k^{-1} \int h^2 \, d\rho (\tau P(\delta^{-1} - 1))^{-2}(1 + o(1)) = o(1). \end{aligned}$$

Furthermore let us demonstrate that

$$E \| u \|^4 = O(1), \quad n \to \infty. \tag{2.10}$$

Indeed, since q is continuous with compact support,

$$\begin{aligned} E \| u \|^4 = E(\textstyle\sum_{j=1}^{\tilde{k}} u_j^2)^2 &\leq \textstyle\sum_{j=1}^{\tilde{k}} E u_j^4 + (E \textstyle\sum_{j=1}^{\tilde{k}} u_j^2)^2 \\ &\leq 3k^{-1} \int q^2 d\rho + \left(\int q \, d\rho \right)^2 = O(1). \end{aligned}$$

Observe now that the right-hand side of (2.8) is not changed if only estimators with values in $\tilde{\Theta}_n$ are included, since $\tilde{\Theta}_n$ is closed and convex. For any such estimator t^*

$$\begin{aligned} &\sup_{t \in \tilde{\Theta}_n} E_t \| t^* - t \|^2 \\ &\geq \inf_{\hat{t}} \int E_t \| \hat{t} - t \|^2 Q_n(dt) - 2 \int_{\tilde{\Theta}_n^c} E_t(\| t^* \|^2 + \| t \|^2) Q_n(dt). \end{aligned} \tag{2.11}$$

Here

$$\| t^* \|^2 \leq \varepsilon^{-1} \textstyle\sum_{j=1}^{\tilde{k}} p(jk^{-1}) t_j^{*2} \leq \varepsilon^{-1} \tau P.$$

Hence by (2.9)

$$\int_{\tilde{\Theta}_n^c} E_t \| t^* \|^2 Q_n(dt) \leq \varepsilon^{-1} P Q_n(\tilde{\Theta}_n^c) = o(1). \tag{2.12}$$

Furthermore (2.9) and (2.10) imply

$$\int_{\tilde{\Theta}_n^c} \| t \|^2 Q_n(dt) \leq (E \| u \|^4 Q_n(\tilde{\Theta}_n^c))^{1/2} = o(1). \tag{2.13}$$

The first summand on the right-hand side of (2.11) is the Bayes risk in model (2.7) for a prior Q_n; this is easily found as

(2.14) $$\sum_{j=1}^{\tilde{k}} k^{-1}\delta q(jk^{-1})k^{-1}(k^{-1} + k^{-1}\delta q(jk^{-1}))^{-1} \geq \delta \int q(1+q)^{-1}\, d\rho.$$

Now we have

$$q(x)(1+q(x))^{-1} = g(x\tilde{\mu}) \quad \text{if} \quad (\pi x)^{2m} \geq \varepsilon.$$

Hence (2.14) implies for $\varepsilon' = \varepsilon^{1/2m}\pi^{-1}$

(2.15) $$\begin{aligned} \inf_{\hat{t}} \int E_t \| \hat{t} - t \|^2 Q_n(dt) &\geq \delta \int_{\varepsilon'}^{\infty} g(x\tilde{\mu})\rho(dx) \\ &= \delta\tilde{\mu}^{-1} \int_{\varepsilon'\tilde{\mu}}^{\infty} g(x)\, dx\, (1 + o(1)). \end{aligned}$$

Collecting (2.8), (2.11) – (2.15) we find

$$r \geq \delta\tilde{\mu}^{-1} \int_{\varepsilon'\tilde{\mu}}^{\infty} g(x)\, dx.$$

Now let $\delta \to 1$, $\varepsilon \to 0$, $\tau \to 1$ (hence $\tilde{\mu} \to \mu$) and recall (2.4). □

We now describe the asymptotically minimax estimate. Define $k^* = [k/\log k]$ and

(2.16) $$g_{jn} = 1, \quad j = 1, \cdots, k^*, \quad g_{jn} = g(\mu j k^{-1}), \quad j = k^* + 1, \cdots, n.$$

Consider the estimator

$$\tilde{\vartheta}_n = (g_{jn}\eta_j)_{j=1,\cdots,n}.$$

THEOREM 2.2. *Suppose there is a sequence $\{\delta_j\}$, not depending on n, such that $\lim_j \delta_j = 1$ and*

$$\sup_{1\leq j\leq \tilde{k}} \delta_j a_{jn}(\pi j)^{-2m} \geq 1 + o(1), \quad n \to \infty.$$

Then

$$\limsup_n \sup_{\vartheta\in\Theta_n} k^{2m} E_\vartheta \| \tilde{\vartheta}_n - \vartheta \|^2 \leq \gamma(m, P).$$

PROOF. The condition implies that $a_{jn} > 0$ for $k^* < j \leq \tilde{k}$, $n > C$. Then

$$\begin{aligned} (1 - g_{jn})^2 a_{jn}^{-1} &\leq (1 + o(1)) \sup_{k^* < j \leq \tilde{k}} (1 - g_{jn})^2 (\pi j)^{-2m} \\ &\leq (1 + o(1))(\mu k^{-1})^{2m} \sup_{x>0} (1 - g(x))^2 (\pi x)^{-2m} \\ &= (\mu k^{-1})^{2m}(1 + o(1)) \end{aligned}$$

uniformly for $k^* < j \leq \tilde{k}$. For $\tilde{k} < j \leq n$, we have $g_{jn} = 0$ and hence

$$(1 - g_{jn})^2 a_{jn}^{-1} \leq a_{\tilde{k}n}^{-1} \leq (\pi\tilde{k})^{-2m} C \leq (\mu k^{-1})^{2m}$$

for $n > C$.

Recalling (2.3), we obtain uniformly for $\vartheta \in \Theta_n$

$$\begin{aligned}\| E_\vartheta \tilde{\vartheta}_n - \vartheta \|^2 &= \textstyle\sum_{j=1}^n (1 - g_{jn})^2 \vartheta_j^2 \\ &\le (\mu k^{-1})^{2m} (1 + o(1)) \textstyle\sum_{j=k^*+1}^n a_{jn} \vartheta_j^2 \\ &\le (\mu k^{-1})^{2m} P(1 + o(1)) \\ &= k^{-2m} \mu^{-1} \int g(x)(1 - g(x))\, dx (1 + o(1)).\end{aligned}$$

Furthermore

$$k^{2m} E_\vartheta \| \tilde{\vartheta}_n - E_\vartheta \tilde{\vartheta}_n \|^2 = \textstyle\sum_{j=1}^n g_{jn}^2 k^{-1}.$$

In this term, the first k^* summands are negligible; hence it is

$$\int g^2(\mu x) \rho(dx)(1 + o(1)) = \mu^{-1} \int g^2(x)\, dx (1 + o(1)).$$

The last three relations and (2.4) yield

$$\begin{aligned}E_\vartheta \| \tilde{\vartheta}_n - \vartheta \|^2 &= E_\vartheta \| \tilde{\vartheta}_n - E_\vartheta \tilde{\vartheta}_n \|^2 + \| E_\vartheta \tilde{\vartheta}_n - \vartheta \|^2 \\ &\le k^{-2m} \mu^{-1} \int g(x)\, dx (1 + o(1)) \\ &= k^{-2m} \gamma(m, P)(1 + o(1))\end{aligned}$$

uniformly for $\vartheta \in \Theta_n$. □

3. Splines and noncircularity. The result available for the continuous observation case concerned the periodic Sobolev class $\tilde{W}_2^m(P)$, with boundary conditions. We are concerned with the discrete regression model, and also with the larger class $W_2^m(P)$.

The optimal convergence rate does not depend on the presence of boundary conditions. Indeed, the lower risk bound is established for subclasses of $\tilde{W}_2^m(P)$(Ibragimov and Hasminski, 1980; Stone, 1982). It then remains to show that there is an estimator which attains the rate $n^{-2m/(2m+1)}$ also over $W_2^m(P)$. It is well known that estimators of standard kernel type are not suitable. For our particular model, attainment has been shown for smoothing splines (Wahba, 1978; Utreras, 1983; Cox, 1983) and for a boundary modified kernel method (Gasser and Müller, 1979).

To deal with the problem of the asymptotic minimax constant, we ask what kind of restriction the smoothness assumption $f \in W_2^m(P)$ implies for the function values $f^{(n)} = (f(x_{jn}))_{j=1,\cdots,n}$. The answer to this is provided by spline theory. Consider the minimization problem

$$\min\{ \| D^m g \|^2 \mid g \in W_2^m,\ g(x_{jn}) = f(x_{jn}),\ j = 1, \cdots, n\}.$$

The solution is known to be unique for $n \ge m$, and is the natural polynomial interpolation spline, denoted $S(f^{(n)})$. Then

$$\| D^m S(f^{(n)}) \|^2 \le \| D^m f \|^2 \le P.$$

Now $S(f^{(n)})$ is known to be linear in $f^{(n)}$. Hence $\| D^m S(f^{(n)}) \|^2$ is a quadratic form, with matrix Γ_n, say, and we have

$$f^{(n)\prime}\Gamma_n f^{(n)} \le P. \tag{3.1}$$

Thus we know in fact that the parameter space for the vector $f^{(n)}$ is an ellipsoid in $\mathbb{R}^n$. We are then in the framework of Section 2, and it remains to specify the behaviour of the main axes of this ellipsoid for $n \to \infty$. The matrix Γ_n, associated with spline interpolation, and its eigenvalues have been studied by Craven and Wahba (1979) and Utreras (1980, 1983). Although it has been found that these eigenvalues, if multiplied by n, tend to behave like $\{(\pi j)^{2m}\}$, no estimates sufficiently sharp for our purpose seem to be available. We shall obtain an improved estimate which is the key auxiliary result in this paper.

Multiplying the regression data by $n^{-1/2}$, we obtain a model (2.1), (2.2), up to an orthogonal transformation. With the a_j determined, application of minimax filtering is straightforward and will be carried out in the next section.

We shall rely on the theory of interpolation and smoothing splines as outlined by Craven and Wahba (1979)(hereafter abbreviated by CW). From the known structure of the matrix Γ we shall deduce our spectrum estimate.

Consider functions on [0, 1]

$$k_0(t) = 1, \quad k_1(t) = t - \tfrac{1}{2},$$

$$k_r(t) = -\textstyle\sum_{j\in\mathbb{Z},j\neq 0} (2\pi i j)^{-r}\exp(2\pi i j t), \quad r = 2, 3, \cdots,$$

which are the rth degree Bernoulli polynomials, up to a factor $r!$. Let

$$K_{jl} = (-1)^{m-1}k_{2m}(x_l - x_j), \quad j, l = 1, \cdots, n,$$

where $\{x_1, \cdots, x_n\}$ is our regression design. Define real symmetric matrices

$$K = (K_{jl})_{j=1,\cdots,n}^{l=1,\cdots,n}, \quad K^0 = (K_{jl})_{j=2,\cdots,n}^{l=2,\cdots,n}$$

and numbers

$$\lambda_j = (n-1) \textstyle\sum_{l\in\mathbb{Z}} (2\pi(j + l(n-1)))^{-2m}, \quad j = 1, \cdots, n-2$$

$$\lambda_{n-1} = (n-1) \textstyle\sum_{l\in\mathbb{Z},l\neq 0} (2\pi l(n-1))^{-2m}.$$

LEMMA 3.1. *We have*

(i) $\lambda_j = \lambda_{n-j-1}, \quad j = 1, \cdots, n-2$
(ii) $\lambda_1 \ge \lambda_2 \ge \cdots \ge \lambda_{[(n-1)/2]}$
(iii) $\lambda_{[(n-1)/2]} \ge \lambda_{n-1}$.

PROOF. (i) is obvious. For (ii), note that for $l = 0, 1, \cdots$ the functions

$$\varphi_l(t) = (t + l(n-1))^{-2m} + (t - (l+1)(n-1))^{-2m}$$

are nonincreasing in t for $t \in [0, (n-1)/2]$, and write

$$\lambda_j = (n-1)(2\pi)^{-2m} \textstyle\sum_{l=0}^{\infty} \varphi_l(j), \quad j = 1, \cdots, n-2.$$

For (iii), note that for $l = 0, 1, \cdots$ the functions

$$\tilde{\varphi}_l(t) = (t + l(n-1))^{-2m} + (t - (l+2)(n-1))^{-2m}$$

are nonincreasing in t for $t \in [0, n-1]$, and write

$$\lambda_j = (n-1)(2\pi)^{-2m}((j-n+1)^{-2m} + \textstyle\sum_{l=0}^{\infty} \tilde{\varphi}_l(j)), \quad j = 1, \cdots, n-2,$$

which implies $\lambda_j \geq \lambda_{n-1}, j = 1, \cdots, n-2$. □

For any real symmetric $l \times l$-matrix A, let $\lambda_j(A), j = 1, \cdots, l$, be the l eigenvalues of A, ordered as $\lambda_1(A) \leq \cdots \leq \lambda_l(A)$. In their Lemma 2.1, CW show that the numbers $\lambda_j, j = 1, \cdots, n-1$, are eigenvalues of K^0. In conjunction with Lemma 3.1 this implies

$$\lambda_{n-j}(K^0) = \lambda_{[(j+1)/2]}, \quad j = 1, \cdots, n-2. \tag{3.2}$$

We now describe the structure of the matrix Γ.

Lemma 3.2. *For each $n \geq m$ there is*

(i) *an $n \times (n-m)$ matrix E such that $E'E = I_{n-m}$ and*
(ii) *a vector $e \in \mathbb{R}^n$ such that* $\operatorname{rank}(E'(K + ee')E) = n - m$

such that for all $z \in \mathbb{R}^n$

$$\| D^m S(z) \|^2 = z'\Gamma z$$

where

$$\Gamma = E(E'(K + ee')E)^{-1}E'.$$

Proof. It is well known that $S(z)$ has a representation

$$S(z)(t) = \textstyle\sum_{l=0}^{m} \theta_l k_l(t) + (-1)^{m-1} \sum_{j=1}^{n} \alpha_j k_{2m}(t - x_j),$$

see CW, formula (2.8a). Although CW treat smoothing splines, it is valid also for the interpolation case. According to formula (2.11) in CW we have, with $\alpha = (\alpha_j)_{j=1,\cdots,n}$,

$$\| D^m S(z) \|^2 = \alpha' K \alpha + \theta_m^2. \tag{3.3}$$

Define

$$T = (k_l(x_j))_{j=1,\cdots,n}^{l=0,\cdots,m-1}, \quad e = (k_m(x_j))_{j=1,\cdots,n},$$
$$\theta = (\theta_j)_{j=0,\cdots,m-1}.$$

The restrictions of interpolation for $S(z)$ can now be written

$$K\alpha + e\theta_m + T\theta = z. \tag{3.4}$$

$S(z)$ is obtained by minimizing (3.3) subject to (3.4). Since k_l is a polynomial of degree l and $n \geq m$, we have $\operatorname{rank}(T) = m$. Let E be a $n \times (n-m)$ matrix such that $E'T = O_{(n-m)\times m}$, $E'E = I_{n-m}$. Eliminating θ from (3.4), we obtain

$$E'(K\alpha + e\theta_m) = E'z.$$

Let $K^{1/2}$ be the unique symmetric square root matrix of the symmetric matrix K, and let

$$\tilde{K} = (K^{1/2} \mid e), \quad \tilde{\alpha} = \begin{pmatrix} K^{1/2}\alpha \\ \theta_m \end{pmatrix}.$$

Then the problem transforms to

$$\min\{\| \tilde{\alpha} \|^2 \mid \tilde{\alpha} \in \mathbb{R}^{n+1}, E'\tilde{K}\tilde{\alpha} = E'z\}. \tag{3.5}$$

We know that a solution of (3.4) exists for all $z \in \mathbb{R}^n$; hence this also holds for $E'\tilde{K}\tilde{\alpha} = E'z$. As rank $(E) = n - m$, it follows that $\operatorname{rank}(E'\tilde{K}) = n - m$. Then the value of (3.5) is

$$z'E(E'\tilde{K}\tilde{K}'E)^{-1}E'z = z'\Gamma z. \quad \square$$

Now we can bound the eigenvalues of Γ by successive application of separation theorems.

LEMMA 3.3. *For $n \geq m + 5$ we have*

$$\lambda_{[(j-m-1)/2]}^{-1} \leq \lambda_j(\Gamma) \leq \lambda_{[(j+1)/2]}^{-1}, \quad j = m + 3, \cdots, n - 2.$$

PROOF. Since K^0 is a submatrix of K we can apply the Sturmian or the Poincaré separation theorem (see Rao, 1973, page 64) to obtain

$$\lambda_{j-1}(K^0) \leq \lambda_j(K) \leq \lambda_j(K^0), \quad j = 2, \cdots, n - 1. \tag{3.6}$$

Let $\tilde{K}$ be as above; then

$$\lambda_j(\tilde{K}'\tilde{K}) = \lambda_{j-1}(K + ee'), \quad j = 2, \cdots, n + 1. \tag{3.7}$$

But analogously to (3.6) we have

$$\lambda_{j-1}(K) \leq \lambda_j(\tilde{K}'\tilde{K}) \leq \lambda_j(K), \quad j = 2, \cdots, n. \tag{3.8}$$

The Poincaré separation theorem yields

$$\lambda_j(K + ee') \leq \lambda_j(E'(K + ee')E) \leq \lambda_{j+m}(K + ee'), \quad j = 1, \cdots, n - m. \tag{3.9}$$

Collecting (3.6)–(3.9) we obtain

$$\lambda_{j-1}(K^0) \leq \lambda_j(E'(K + ee')E) \leq \lambda_{j+m+1}(K^0), \quad j = 2, \cdots, n - m - 2.$$

K^0 is nonsingular since λ_{n-1} is its smallest eigenvalue. Hence

$$\lambda_{j-2}(K^{0-1}) \leq \lambda_j((E'(K + ee')E)^{-1} \leq \lambda_{j+m}(K^{0-1}), \quad j = 3, \cdots, n - m - 1;$$

consequently

$$\lambda_{j-m-2}(K^{0-1}) \leq \lambda_j(\Gamma) \leq \lambda_j(K^{0-1}), \quad j = m + 3, \cdots, n - 1.$$

Finally note that (3.2) implies

$$\lambda_j(K^{0-1}) = \lambda_{[(j+1)/2]}^{-1}, \quad j = 1, \cdots, n - 2. \quad \square$$

The next lemma states our final result on the eigenvalues of Γ. Define

$$a_{jn} = n\lambda_j(\Gamma_n), \quad j = 1, \cdots, n. \tag{3.10}$$

LEMMA 3.4. *There is a sequence* $\{\delta_j\}$, *not depending on* n, $\delta_j = o(1)$, *such that for* $n > C, j = m + 3, \cdots, n - 2$

(i) $a_{jn}(\pi_j)^{-2m} \le (1 + \delta_j)(1 - n^{-1})^{-1}$
(ii) $a_{jn}(\pi j)^{-2m} \ge (1 - \delta_j)(1 + C(jn^{-1})^{2m})^{-1}$.

PROOF. According to Lemma 3.3 we have for $j = m + 3, \cdots, n - 2$

$$(1 - n^{-1})a_j(\pi j)^{-2m} \le ((n - 1)^{-1}\lambda_{[(j+1)/2]}(\pi j)^{2m})^{-1}$$

and furthermore

$$(n - 1)^{-1}\lambda_{[(j+1)/2]}(\pi j)^{2m} = \sum_{l\in\mathbb{Z}} (2[(j + 1)/2]j^{-1} + 2l(n - 1)j^{-1})^{-2m}$$
$$\ge (j/2[(j + 1)/2])^{2m}$$

which proves (i). Analogously

$$(1 - n^{-1})a_j(\pi j)^{-2m} \ge ((n - 1)^{-1}\lambda_{[(j-m-1)/2]}(\pi j)^{2m})^{-1}$$

and furthermore

$$(n - 1)^{-1}\lambda_{[(j-m-1)/2]}(\pi j)^{2m} = \sum_{l\in\mathbb{Z}} (2[(j - m - 1)/2]j^{-1} + 2l(n - 1)j^{-1})^{-2m}$$
$$\le (j/2[(j - m - 1)/2])^{2m} + (jn^{-1})^{2m}C$$

which implies (ii). □

The result may be compared to the one of Utreras (1983), who established $C < a_{jn}(\pi j)^{-2m} < C$ uniformly for $j, j \ge m$, and showed convergence of each $\{a_{jn}\}_{n\in\mathbb{N}}$ to an eigenvalue of a differential operator (Utreras, 1980).

4. Minimax spline smoothing. We now prove the main results by applying the optimal filtering scheme of Section 2. Let d_j, $j = 1, \cdots, n$, be an orthonormal eigensystem of Γ so that

$$n\Gamma = \sum_{j=1}^{n} a_j d_j d_j'.$$

Define

$$G = \sum_{j=1}^{n} g_j d_j d_j', \quad Y = (y_j)_{j=1,\cdots,n}$$

where g_j are the coefficients of the optimal smoother from (2.16), and

$$\tilde{f}^{(n)} = GY.$$

$\tilde{f}^{(n)}$ is the estimate of the function values in the design points x_{jn}. Let

$$\tilde{f}_n = S(\tilde{f}^{(n)}).$$

$\tilde{f}_n$ is the smoothing spline for which we claim optimality. For the rest of the paper set $\Theta = W_2^m(P)$.

PROOF OF THEOREM 1(i). Define

$$M = (n^{-1/2}(d_1, \cdots, d_n))'.$$

Then, because of (3.1), $Mf^{(n)}$ varies in an ellipsoid (2.2) where the a_j are given by (3.10). Moreover MY has a distribution $N(Mf^{(n)}, n^{-1}I_n)$; hence MY follows a model of type (2.1). For any estimator $\hat{f} \in \mathscr{F}_n$, we have

$$\|\hat{f} - f\|_n^2 = n^{-1}\|\hat{f}^{(n)} - f^{(n)}\|^2 = \|M\hat{f}^{(n)} - Mf^{(n)}\|^2.$$

We are therefore in the framework of Section 2. For the lower risk bound, refer to Theorem 2.1 and note that its condition is fulfilled by virtue of Lemma 3.4(i). The estimator $M\tilde{f}_n$ coincides with $\tilde{\vartheta}_n$ of Theorem 2.2; the condition there is implied by Lemma 3.4(ii). Hence

$$\sup_{f\in\Theta} E_f\|\tilde{f}_n - f\|_n^2 \le k^{2m}\gamma(m, P)(1 + o(1)). \quad \square \tag{4.1}$$

To obtain the corresponding statement for the L_2 norm, we need first an approximation-theoretic result. In the sequel, C shall denote constants which, in addition, do not depend on functions $f \in W_2^m$.

LEMMA 4.1. *For any function* $f \in W_2^m$

$$|\,\|f\|^2 - \|f\|_n^2| \le n^{-1}C(\|f\|_n^2 + \|D^m f\|^2), \quad n \ge C.$$

PROOF. The inequality

$$|\,\|f\|^2 - \|f\|_n^2| \le n^{-1}C(\|f\|^2 + \|D^m f\|^2)$$

is well known. It implies that

$$\|f\|^2 \le C(\|f\|_n^2 + \|D^m f\|^2), \quad n \ge C$$

from which the lemma follows. $\square$

For the following lower risk bound, let it be understood that loss is infinite if the estimate is not in L_2.

THEOREM 4.1.

$$\inf_{\hat{f}\in\mathscr{F}_n}\sup_{f\in\Theta} E_f\|\hat{f} - f\|^2 \ge k^{2m}\gamma(m, P)(1 + o(1)).$$

PROOF. Let $\hat{f} \in \mathscr{F}_n$ be an estimator with realizations in L_2. Since Θ is a closed convex subset of L_2, a projection argument justifies an assumption that $\hat{f}$ takes values in Θ. Then by Lemma 4.1

$$E_f\|\hat{f} - f\|^2 \ge E_f\|\hat{f} - f\|_n^2(1 - n^{-1}C) - n^{-1}CE_f\|D^m(\hat{f} - f)\|^2.$$

The second summand is bounded in modulus by n^{-1}CP. Now refer to the already established result for the norm $\|\cdot\|_n$. $\square$

LEMMA 4.2. *sup*$_{f\in\Theta} E_f\|D^m\tilde{f}_n\|^2 \le C.$

PROOF. Define

$$s = [k/\mu\pi] + 1, \quad \tilde{\Gamma} = n^{-1}\sum_{j=1}^{s} a_j d_j d_j'.$$

Then $g_j = 0, j = s, \cdots, n$, and

$$E_f\| D^m\tilde{f}_n\|^2 = E_f\tilde{f}^{(n)\prime}\Gamma\tilde{f}^{(n)} = E_f\tilde{f}^{(n)\prime}\tilde{\Gamma}\tilde{f}^{(n)}$$
$$\leq 2E_f(\tilde{f}^{(n)} - f^{(n)})'\tilde{\Gamma}(\tilde{f}^{(n)} - f^{(n)}) + 2f^{(n)\prime}\Gamma f^{(n)}$$
$$\leq 2n^{-1}a_s E_f\|\tilde{f}^{(n)} - f^{(n)}\|^2 + 2P.$$

Lemma 3.4(i) implies that $a_s \leq Ck^{2m}$. Invoking relation (4.1) completes the proof. □

THEOREM 4.2. $sup_{f\in\Theta}E_f\|\tilde{f}_n - f\|^2 \leq k^{2m}\gamma(m, P)(1 + o(1))$.

PROOF. By Lemma 4.1

$$E_f\|\tilde{f}_n - f\|^2 \leq E_f\|\tilde{f}_n - f\|_n^2(1 + n^{-1}C) + n^{-1}CE_f\|D^m(\tilde{f}_n - f)\|^2.$$

Now use Lemma 4.2 and relation (4.1). □

Note that part (ii) of Theorem 1 follows from Theorems 4.1 and 4.2.

REMARKS.

(1) The optimal estimator $\tilde{f}_n$ is linear in the data and may be viewed as a smoothing spline. It seems that the conventional smoothing spline, i.e., the g which minimizes

$$n^{-1}\sum_{j=1}^n (y_j - g(x_j))^2 + \lambda\|D^m g\|^2$$

is not asymptotically minimax in the present sense, whatever the choice of the smoothing parameter λ. Indeed it corresponds to a filter $\{(1 + \lambda(\pi j)^{2m})^{-1}\}$ (approximately) rather than to the optimal one (2.16).

(2) Rice and Rosenblatt (1983) analyze the behaviour of smoothing splines in the noncircular model of the present paper. They state that the integrated mean square error is dominated by contributions from the boundary, so that the rate of convergence is affected. However this does not contradict the rate optimality of smoothing splines; only the order of the spline has to be selected sufficiently large. The interplay between smoothness, boundary conditions and rates of convergence is clarified by Cox (1984).

(3) Sacks and Strawderman (1982) consider estimation of $f(x)$ for a given point x. They show that for some cases when f is from a smoothness class, linear estimators achieve the optimal rate for mean square error $E_f(\hat{f}(x) - f(x))^2$ but not the best constant in the minimax sense. The crucial difference in the present paper is the global L_2-loss. For a loss $(\hat{f}(x) - f(x))^2$ in our model, the conventional smoothing spline is minimax among linear estimators if $\lambda = (nP)^{-1}$ (see Li, 1982). It attains the optimal rate $n^{-(2m-1)/2m}$ (see Speckman, 1981).

(4) Cox (1984) established L_2-rate optimality of the smoothing spline for bounded domains of $\mathbb{R}^d$. Using some of the methods of this paper, our result can be generalized to such domains and to nonequidistant designs.

Acknowledgment. The author would like to express his gratitude to Professor R. Z. Hasminski and to G. K. Golubev for stimulating and helpful discussions during his stay at Moscow.

REFERENCES

COX, D. D. (1983). Asymptotics for M-type smoothing splines. *Ann. Statist.* **11** 530–551.

COX, D. D. (1984). Multivariate smoothing spline functions. *SIAM J. Numer. Anal.* **21** 789–813.

CRAVEN, P. and WAHBA, G. (1979). Smoothing noisy data with spline functions. *Numer. Math.* **31** 377–403.

EFROIMOVICH, S. YU. and PINSKER, M. S. (1982). Estimating a square integrable probability density of a random variable (in Russian). *Problems Inform. Transmission* **18**, No. 3, 19–38.

GASSER, T. and MÜLLER, H.-G. (1979). Kernel estimation of regression functions. In *Smoothing Techniques for Curve Estimation. Lecture Notes in Math.* **757** 23–68.

GOLUBEV, G. K. (1982). On minimax filtering of functions in L_2 (in Russian). *Problems Inform. Transmission* **18**, No. 4, 67–75.

GOLUBEV, G. K. (1984). On minimax estimation of regression (in Russian). *Problems Inform. Transmission* **20**, No. 1, 56–64.

IBRAGIMOV, I. A. and HASMINSKI, R. Z. (1980). Asymptotic efficiency bounds for nonparametric estimation of a regression function in L_p (in Russian). *Zap. Naučn. Sem. LOMI* **97** 88–101.

IBRAGIMOV, I. A. and HASMINSKI, R. Z. (1981). *Statistical Estimation: Asymptotic Theory.* Springer-Verlag, New York.

IBRAGIMOV, I. A. and HASMINSKI, R. Z. (1982). On density estimation within a class of entire functions (in Russian). *Theory Probab. Appl.* **27** 514–524.

LI, KER-CHAU (1982). Minimaxity of the method of regularization on stochastic processes. *Ann. Statist.* **10** 937–942.

MILLAR, P. W. (1982). Optimal estimation of a general regression function. *Ann. Statist.* **10** 717–740.

PINSKER, M. S. (1980). Optimal filtering of square integrable signals in Gaussian white noise (in Russian). *Problems Inform. Transmission* **16**, No. 2, 52–68.

RAO, C. R. (1973). *Linear Statistical Inference and its Applications.* Wiley, New York.

RICE J. and ROSENBLATT, M. (1983). Smoothing splines: regression, derivatives and deconvolution. *Ann. Statist.* **11** 141–156.

SACKS, J. and STRAWDERMAN, W. (1982). Improvements on linear minimax estimates. In *Statistical Decision Theory and Related Topics* III, *2* (S. Gupta, ed.) Academic, New York.

SPECKMAN, P. (1985). Spline smoothing and optimal rates of convergence in nonparametric regression models. *Ann. Statist.* **13** 970–983.

STONE, C. J. (1982). Optimal global rates of convergence for nonparametric regression. *Ann. Statist.* **10** 1040–1053.

UTRERAS, F. (1980). Sur le choix du paramètre d'ajustement par le lissage par fonctions spline. *Numer. Math.* **34** 15–28.

UTRERAS, F. (1983). Natural spline functions, their associated eigenvalue problem. *Numer. Math.* **42** 107–117.

WAHBA, G. (1978). Improper priors, spline smoothing and the problem of guarding against model errors in regression. *J. Roy. Statist. Soc. Ser. B* **40** 364–372.

AKADEMIE DER WISSENSCHAFTEN DER DDR
INSTITUT FÜR MATHEMATIK
DDR-1086 BERLIN
MOHRENSTRASSE 39
GERMAN DEMOCRATIC REPUBLIC