The Annals of Mathematical Statistics
1970, Vol. 41, No. 2, 495–502

A CORRESPONDENCE BETWEEN BAYESIAN ESTIMATION ON STOCHASTIC PROCESSES AND SMOOTHING BY SPLINES

BY GEORGE S. KIMELDORF[1] AND GRACE WAHBA[2]

Mathematics Research Center, University of Wisconsin and University of Wisconsin

1. Introduction. Let

$$L = \sum_{j=0}^{m} a_j D^j \qquad m > 0, a_m \neq 0 \tag{1.1}$$

be a linear differential operator with real constant coefficients and denote the adjoint operator by

$$L^* = \sum_{j=0}^{m} a_j(-D)^j. \tag{1.2}$$

Let $\{t_i : i = 1, 2, \cdots, n\}$ be a set of distinct real constants.

DEFINITION. An L-spline with knots $\{t_i\}$ is a function $x \in C^{2m-2}$ for which

$$L^*Lx = 0 \tag{1.3}$$

on each open interval $(-\infty, t_1)$, (t_i, t_{i+1}), (t_n, ∞). Hence an L-spline consists of piecewise solutions of a $2m$th order linear homogeneous differential equation joined at the knots in such a manner as to maintain continuity of all derivatives up to and including the $(2m-2)$th. If we were to require that $x \in C^{2m-1}$, then x would satisfy (1.3) everywhere. Thus, an L-spline can be looked upon as the "most differentiable" function which satisfies (1.3) on the appropriate open intervals without satisfying it everywhere. Although operators of the form (1.1) are sufficiently general for our present purposes, it should be pointed out that L-splines are often defined and studied for other linear differential operators, in which case the domain of definition of x is taken to be a finite closed interval. References [1], [3] and [6] contain extensive bibliographies on splines.

Two common problems for which L-splines are solutions are the following:

(i) *Curve fitting*. Given data $\{(t_j, y_j): j = 1, 2, \cdots, n\}$ to find a function $\hat{x}(t)$ which minimizes

$$\int_{-\infty}^{\infty} (Lx)^2 \, dt \tag{1.4}$$

among all functions x in a certain class for which

$$x(t_j) = y_j \qquad j = 1, 2, \cdots, n. \tag{1.5}$$

If (1.4) is interpreted as a measure of the roughness of x, then $\hat{x}$, if it exists, is the smoothest interpolator to the data.

Received December 19, 1968.

[1] Research sponsored by the Mathematics Research Center, United States Army, Madison, Wisconsin, under Contract No. DA-31-124-ARO-D-462. Now at Florida State University.

[2] Research sponsored in part by the Wisconsin Alumni Research Foundation.

(ii) *Smoothing*. Given data $\{(t_j, y_j): j = 1, 2, \cdots, n\}$ to find a function $\hat{x}(t)$ which minimizes the sum of (1.4) and

$$\sum_{j=1}^{n} \sum_{k=1}^{n} [x(t_j) - y_j] b^{jk} [x(t_k) - y_k] \tag{1.6}$$

where $B^{-1} = [b^{jk}]$ is a given positive-definite matrix. If (1.6) is interpreted as a measure of the disparity of x with the data, then $\hat{x}$, if it exists, is a compromise between smoothness and fidelity to the data.

Conditions under which $\hat{x}$ exists, is unique, and is an L-spline are presented in Theorem 3.1 and Theorem 3.2 below for the smoothing and curve-fitting problem respectively.

The literature on spline functions is silent concerning criteria for selecting an appropriate operator L for a particular smoothing or curve-fitting problem. In practice the choice of L is based on computational convenience rather than theoretical considerations. Although the question of what constitutes a "smooth" function for a particular problem is necessarily subjective, it can be argued that a choice of smoothness criterion should rely on prior knowledge of the underlying system which generates the data. In this paper we present a stochastic model for curve fitting and smoothing in which the selection of a smoothing criterion corresponds to the specification of a prior probability measure over a function space. In particular, we exhibit classes of prior distributions for which the Bayes' estimate of an unknown function, given certain (perhaps error-plagued) observations, solves the curve-fitting or smoothing problem and is an L-spline.

2. The stochastic model. Consider a real random function $\{x(t), -\infty < t < \infty\}$ to be estimated based on prior information and a finite set of observations, perhaps containing random error. The prior information is summarized by a prior distribution on $\{x(t)\}$ which is Gaussian with covariance $K(s, t)$ and, without loss of generality, with mean zero. At fixed points $t_1, \cdots, t_n$, the random variables $y_j = x(t_j) + e_j$ are observed, where the vector of measurement errors e_j is independent of $\{x(t)\}$ and distributed normally with mean zero and covariance matrix $B = [b_{jk}]$, which is assumed known. For fixed t, the estimate

$$\hat{x}(t) = E[x(t) \mid y(t_1), \cdots, y(t_n)] \tag{2.1}$$

of $x(t)$ is taken to be the mean of the marginal conditional distribution of $x(t)$, given the observations. (Formally, $\hat{x}(t)$ is the solution to a standard filtering-prediction problem.)

Let us now state two known results which are needed in the sequel:

LEMMA 2.1. *Let $\{\xi_j\}$ be a set of n linearly independent elements of a real Hilbert space $\mathscr{H}$, let M be the matrix $[\langle \xi_j, \xi_k \rangle]$ of inner products among the ξ_j, and let $\{y_j\}$ be a set of n scalars. Then the unique element $\hat{u} \in \mathscr{H}$ which minimizes $\langle u, u \rangle$ subject to $\langle u, \xi_j \rangle = y_j$ for all j is*

$$\hat{u} = (\xi_1, \cdots, \xi_n) M^{-1} (y_1, \cdots, y_n)'. \tag{2.2}$$

LEMMA 2.2. *Under the hypotheses of Lemma* 2.1 *if* $B = [b_{jk}] = [b^{jk}]^{-1}$ *is a positive definite real matrix, then the unique element* $\hat{u} \in \mathscr{H}$ *which minimizes*

$$\sum\sum [\langle u, \xi_j \rangle - y_j] b^{jk} [\langle u, \xi_k \rangle - y_k] + \langle u, u \rangle \tag{2.3}$$

is

$$\hat{u} = [\xi_1, \cdots, \xi_n][M+B]^{-1}[y_1, \cdots, y_n]'. \tag{2.4}$$

As an application of Lemma 2.1, we derive the well-known formula for $\hat{x}(t)$ given by (2.1). Fix t and let $\mathscr{H}$ be the (finite dimensional) space spanned by the set of random variables $\{x(t), x(t_j), e_j\}$ with inner product $\langle z_1, z_2 \rangle = E[z_1 z_2]$. Then, if K is the covariance kernel, we have

$$\begin{aligned}\langle y_j, y_k \rangle &= \langle x(t_j), x(t_k) \rangle + \langle e_j, e_k \rangle \\ &= K(t_j, t_k) + b_{jk}\end{aligned}$$

and, if $\hat{x}(t)$ is to be the conditional expectation of $x(t)$ given $y_1, \cdots, y_n$, we must have the relations

$$E[\hat{x}(t) - x(t)]y_j = 0 \qquad j = 1, 2, \cdots, n, \tag{2.5a}$$

which are equivalent to the constraints

$$\langle \hat{x}(t), y_j \rangle = K(t_j, t), \qquad j = 1, 2, \cdots, n. \tag{2.5b}$$

Geometrically, (2.5a) states that $\hat{x}(t)$ is on the line through $x(t)$ perpendicular to the hyperplane spanned by $\{y_j\}$. But $\hat{x}(t)$, the projection of $x(t)$ onto this hyperplane, must lie in the hyperplane, and hence is the element in $\mathscr{H}$ of minimal norm subject to (2.5b). Therefore, by Lemma 2.1,

$$\hat{x}(t) = [y_1, y_2, \cdots, y_n][\Sigma + B]^{-1}[K(t_1, t), \cdots, K(t_n, t)]' \tag{2.6}$$

where

$$\Sigma = [K(t_j, t_k)].$$

For fixed observations $y_1, y_2, \cdots, y_n$, we can consider $\hat{x}(t)$ defined by (2.1) as a function of t. We then ask for what, if any, prior covariance function K (depending on L) will $\hat{x}(t)$ solve the smoothing problem (or, for $B = 0$, the curve-fitting problem). Reasoning heuristically we note that if functions x and z are related by $x(t) = z(t + t_0)$, t_0 fixed, then $\int (Lx)^2 \, dt = \int (Lz)^2 \, dt$. Hence if $K(s, t)$ is a prior covariance function which corresponds to (1.4) as a roughness criterion, then K is shift invariant. We therefore restrict our attention to stationary Gaussian processes.

3. The theorems. We use the following notation: We denote by F the Fourier–Plancherel transform, which is an isometry (1–1 onto inner-product preserving map) of $\mathscr{L}_2'$ onto $\mathscr{L}_2'$ defined by

$$(Fg)(\lambda) = (2\pi)^{-\frac{1}{2}} \int_{-\infty}^{\infty} e^{it\lambda} g(t) \, dt$$

where $\mathscr{L}_2' = \{u + iv : u, v \in \mathscr{L}_2\}$. If $\mathscr{H}$ is any Hilbert space of functions then, following Aronszajn [2], we call $\mathscr{H}$ a reproducing kernel Hilbert space (RKHS) with kernel $K(s, t)$ if $K(s_0, t) \in \mathscr{H}$ for all fixed s_0, and $\langle f(t), K(s_0, t) \rangle = f(s_0)$ for all $f \in \mathscr{H}$.

LEMMA 3.1. *Let ψ be any Hermitian bounded function in $\mathscr{L}_2'$ which is non-zero almost everywhere and define $\mathscr{K} = \{f \in \mathscr{L}_2 : (F^{-1}f)/\psi \in \mathscr{L}_2'\}$. Define the mapping $M:\mathscr{K} \to \mathscr{L}_2$ by $Mf = F[(F^{-1}f)/\psi]$. Then $\mathscr{K}$ is an* RKHS *with inner product*

$$\langle f, g\rangle_{\mathscr{K}} = \textstyle\int_{-\infty}^{\infty} (Mf)(Mg) = \langle Mf, Mg\rangle_{\mathscr{L}_2} \tag{3.1}$$

and with kernel

$$K(s,t) = (2\pi)^{-1} \textstyle\int_{-\infty}^{\infty} e^{i(t-s)\lambda} |\psi(\lambda)|^2 \, d\lambda. \tag{3.2}$$

PROOF. To show $\mathscr{K}$ is a Hilbert space with inner product (3.1) it is sufficient to show that $M:\mathscr{K} \to \mathscr{L}_2$ is 1–1 and onto. Since $Mf = 0 \Rightarrow F[(F^{-1}f)/\psi] = 0 \Rightarrow (F^{-1}f)/\psi = 0 \Rightarrow F^{-1}f = 0 \Rightarrow f = 0$, we have that M is 1–1. For any $h \in \mathscr{L}_2$, define $f = F[\psi(F^{-1}h)] \in \mathscr{L}_2$; hence M is onto. To prove that $K(s,t)$ is the reproducing kernel we note that for fixed s_0, $[F^{-1}K(s_0, t)]/\psi = (2\pi)^{-\frac{1}{2}} e^{-is_0\lambda}\overline{\psi}(\lambda) \in \mathscr{L}_2'$; hence $K(s_0, t) \in \mathscr{K}$. Also, for any $f \in \mathscr{K}$, we use Parseval's theorem to derive

$$\begin{aligned}\langle f(t), K(s_0, t)\rangle_{\mathscr{K}} &= \langle F[(F^{-1}f)/\psi], \quad F[(2\pi)^{-\frac{1}{2}} e^{-is_0\lambda}\overline{\psi}(\lambda)\rangle_{\mathscr{L}_2'} \\ &= (2\pi)^{-\frac{1}{2}}\langle (F^{-1}f)/\psi, \quad e^{-is_0\lambda}\overline{\psi}(\lambda)\rangle_{\mathscr{L}_2'} \\ &= (2\pi)^{-\frac{1}{2}} \textstyle\int_{-\infty}^{\infty} (F^{-1}f)(\lambda) \cdot e^{is_0\lambda} \, d\lambda \\ &= f(s_0)\end{aligned}$$

where the inner products on the right-hand side are in $\mathscr{L}_2'$.

LEMMA 3.2. *Under the hypotheses of Lemma* 3.1 *let $\{x(t), -\infty < t < \infty\}$ be a real stationary Gaussian stochastic process with mean zero and spectral density $f(\lambda) = (2\pi)^{-1} |\psi(\lambda)|^2$. If $y_j = x(t_j) + e_j$ where $e_j \sim N(0, B)$, $B = [b_{jk}] = [b^{jk}]^{-1}$ known, and the e_j are independent of $\{x(t), -\infty < t < \infty\}$, then the function $\hat{x}(t)$ defined by* (2.1) *is the unique function $x \in \mathscr{K}$ which minimizes the sum of* (1.6) *and* $\int_{-\infty}^{\infty} (Mx)^2 \, dt$.

PROOF. Use (2.6) and Lemma 3.1, and apply Lemma 2.2 with $\mathscr{H} = \mathscr{K}$, $\xi_j = K(t_j, t)$ and $M = \Sigma$.

LEMMA 3.3. *Under the hypotheses of Lemma* 3.2 *except that now $e_j \equiv 0$ for all j (i.e. $B = 0$), we have that $\hat{x}(t)$ is the unique function $x \in \mathscr{K}$ which minimizes $\int_{-\infty}^{\infty} (Mx)^2 \, dt$ subject to the constraints* (1.5).

PROOF. Use (2.6) with $B = 0$ and Lemma 3.1, and apply Lemma 2.1 with $\mathscr{H} = \mathscr{K}$, $\xi_j = K(t_j, t)$, and $M = \Sigma$.

Following Parzen [5] and Hájek [4], we note that the space $\mathscr{K}$ is isometric to the space $\mathscr{X}$ of random variables generated by $\{x(t)\}$ under the correspondence

$$u \in \mathscr{X} \leftrightarrow f \in \mathscr{K} \Leftrightarrow f(s) = E[u \cdot x(s)].$$

To relate the foregoing theory to L-splines, we seek a function ψ for which M is a differential operator L of the form (1.1). Let P be the complex polynomial

$$P(\lambda) = \textstyle\sum_{j=0}^{m} a_j (i\lambda)^j, \tag{3.3}$$

where the real coefficients a_j are as in (1.1), and let $\psi(\lambda) = 1/P(\lambda)$ so that $\{x(t)\}$ has spectral density

$$f(\lambda) = (2\pi)^{-1}|\psi(\lambda)|^2 = (2\pi)^{-1}|P(\lambda)|^{-2} \tag{3.4}$$

and covariance

$$K(s,t) = (2\pi)^{-1}\int_{-\infty}^{\infty} e^{i(t-s)\lambda}|P(\lambda)|^{-2}\,d\lambda. \tag{3.5}$$

In order that ψ satisfy the hypothesis of Lemma 3.1, we shall assume that $P(\lambda) \neq 0$ for all real λ.

That a spectral density of this form does in fact lead to L-splines which solve the curve-fitting and smoothing problems of Section 1 is the content of the following theorems.

THEOREM 3.1. *Let L, as defined by* (1.1), *be a linear differential operator with real, constant coefficients. Let the stationary Gaussian random function $\{x(t), -\infty < t < \infty\}$ have mean zero and spectral density f defined by* (3.4) *and* (3.3) *where P has no real zeros. If $y_j = x(t_j)+e_j$ where $e_j \sim N(0, B)$, $B = [b_{jk}] = [b^{jk}]^{-1}$ known, and the e_j are independent of $\{x(t), -\infty < t < \infty\}$, then the function $\hat{x}(t)$ defined by* (2.1) *is*

(a) *an L-spline with knots $\{t_j\}$, and*

(b) *the unique function $x \in \mathscr{L}_2$ having absolutely continuous* $(m-1)$th *derivative which minimizes*

$$\sum_{j=0}^{n}\sum_{k=0}^{n}[x(t_j)-y_j]b^{jk}[x(t_k)-y_k]+\int_{-\infty}^{\infty}(Lx)^2\,dt,$$

and hence solves the smoothing problem.

If we consider the Bayesian estimation problem in which the measurements have no error, we again get that $\hat{x}(t)$ is an L-spline. In particular, we have the following result:

THEOREM 3.2. *Under the hypotheses of Theorem* 3.1 *except that now $e_j \equiv 0$ for all j (i.e. $B = 0$), we have that $\hat{x}(t)$ is*

(a) *an L-spline with knots $\{t_j\}$, and*

(b) *the unique function $x \in \mathscr{L}_2$ having absolutely continuous* $(m-1)$th *derivative which minimizes*

$$\int_{-\infty}^{\infty}(Lx)^2\,dt \tag{3.6}$$

subject to the constraints

$$x(t_j) = y_j \qquad j = 1, 2, \cdots, n, \tag{3.7}$$

and hence solves the curve-fitting problem.

Let $\mathscr{H}' = \{f : f \in \mathscr{L}_2,\ f^{(m-1)}$ is absolutely continuous, $Lf \in \mathscr{L}_2\}$ with inner product

$$\langle f, g\rangle_{\mathscr{H}'} = \int_{-\infty}^{\infty}(Lf)(Lg) = \langle Lf, Lg\rangle_{\mathscr{L}_2}.$$

To prove part (b) of Theorems 3.1 and 3.2 we must show that (3.3) and (3.4) imply $\mathscr{K} = \mathscr{K}'$ and $M = L$. This assertion requires the following lemma.

LEMMA 3.4. *For all* $f \in \mathscr{K}'$ *we have*

$$F^{-1}(Lf) = P \cdot (F^{-1}f) \tag{3.8}$$

where P is given by (3.3).

PROOF. Define

$$\begin{aligned} \varphi(t) &= \exp[-(1-t^2)^{-1}] \quad &\text{if} \quad |t| < 1, \\ &= 0 &\text{if} \quad |t| \geqq 1; \end{aligned}$$

so that for $j = 0, 1, 2, \cdots$, $\varphi^{(j)}$ is absolutely continuous and vanishes outside of $[-1, 1]$; hence the same is true for the convolution $(f * \varphi)^{(j)}$ since

$$(f * \varphi)' = f' * \varphi = f * \varphi'. \tag{3.9}$$

Therefore, we can integrate by parts to get

$$\begin{aligned} F^{-1}[L(f * \varphi)] &= P \cdot F^{-1}(f * \varphi) \\ &= P \cdot F^{-1}f \cdot F^{-1}\varphi. \end{aligned} \tag{3.10}$$

But by (3.9), we have

$$\begin{aligned} F^{-1}[L(f * \varphi)] &= F^{-1}[(Lf) * \varphi] \\ &= F^{-1}(Lf) \cdot F^{-1}\varphi. \end{aligned} \tag{3.11}$$

Equating (3.10) and (3.11) proves the lemma since $F^{-1}\varphi \neq 0$.

Lemma 3.4 implies $\mathscr{K}' \subset \mathscr{K}$ and $Mf = Lf$ for all $f \in \mathscr{K}'$. To show $\mathscr{K} \subset \mathscr{K}'$ it is sufficient to show that $K(s_0, t) \in \mathscr{K}'$ for every s_0. By (3.5) we have $K(s_0, t) \in \mathscr{L}_2$, $K^{(m-1)}(s_0, t)$ is absolutely continuous and $LK(s_0, t) = F[(2\pi)^{-\frac{1}{2}} e^{-is_0\lambda}\bar{P}(\lambda)] \in \mathscr{L}_2$, thus completing the proof of part (b) of Theorems 3.1 and 3.2.

To prove part (a) of Theorems 3.1 and 3.2, we denote $K(s, t)$ by $K(t-s)$. By (2.6), it is sufficient to show

$$K(t) \in C^{2m-2}, \tag{3.12}$$

$$L^*LK(t) \equiv 0 \quad \text{on} \quad (0, \infty), \tag{3.13}$$

$$L^*LK(t) \equiv 0 \quad \text{on} \quad (-\infty, 0). \tag{3.14}$$

Formal differentiation of (3.5) $2j$ times yields

$$K^{(2j)}(t) = \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{i\lambda t} \frac{(i\lambda)^{2j}}{|P(\lambda)|^2}\, d\lambda \tag{3.15}$$

and (3.15) is valid if $(i\lambda)^j/P(\lambda) \in \mathscr{L}_2'$, which is true if $j \leqq m-1$ since P is a polynomial of degree m. Hence $K \in C^{2m-2}$. To prove (3.13), we write

$$P(\lambda) = \prod_{j=1}^{m} (i\lambda - c_j)$$

and let A denote the set of c_j with positive real part.

Then defining

$$Q(\lambda) = \textstyle\prod_{c_j \in A}(i\lambda - c_j)\prod_{c_j \notin A}(-i\lambda - c_j) \tag{3.16}$$

we have that $Q(\lambda)\bar{Q}(\lambda) = P(\lambda)\bar{P}(\lambda)$ so that by (3.5)

$$K(t) = \frac{1}{2\pi}\int_{-\infty}^{\infty} \frac{e^{it\lambda}}{Q(\lambda)\bar{Q}(\lambda)}\, d\lambda. \tag{3.17}$$

If N and N^* are the operators defined by formally replacing $i\lambda$ by D in Q and $\bar{Q}$ respectively, then $N^*N = L^*L$. For fixed $t > 0$, (3.17) yields

$$N^*K(t) = \frac{1}{2\pi}\int_{-\infty}^{\infty} \frac{e^{it\lambda}}{Q(\lambda)}\, d\lambda, \tag{3.18}$$

and for z in the upper half-plane, $1/Q(z)$ has no poles and $|e^{itz}| \leqq 1$; hence for $t > 0$ integration by residues yields $N^*K(t) = 0$. A similar argument establishes (3.14).

4. Interpretation of the theorems. The relationship between Bayesian estimation on a stochastic process and (a) curve-fitting or (b) smoothing is expressed by Theorems 3.2 or 3.1 respectively, whereby a choice of smoothing criterion L is equivalent to a prior probability measure for a random function. The restriction to $\mathscr{L}_2$ in parts (b) of the theorems has a natural interpretation. Loosely speaking, for t close to the points t_i at which data are available, the conditional expectation $\hat{x}(t)$ is more closely influenced by the observed data. Conversely, for large $|t|$, $\hat{x}(t)$ is close to the mean, zero, of the prior distribution. The restriction to $\mathscr{L}_2$ in parts (b) of the theorems expresses the rate at which $\hat{x}(t)$ approaches zero as $|t| \to \infty$.

An alternative way of characterizing statistically a well-behaved (i.e. smooth) function x is the extent to which $x(s+t)$ is predictable from knowledge of $x(s)$ for small t. This characterization is equivalent to the rate at which the conditional variance, $K(0)-K(t)$, of $x(s+t)$ given $x(s)$ approaches zero as $t \to 0$, i.e. the largest r, say r_0, for which $K^{(r)}(0)$ exists. From (3.15) we have $r_0 = 2m-2$, which is the largest r for which $\hat{x}^{(r)}(t)$ is continuous. Hence smoothness in this statistical sense is related to continuity of higher-order derivatives of the smoothing function $\hat{x}(t)$.

The class of L-splines for which $L = D^m$, $m > 0$, called ordinary (or polynomial) splines, is used almost exclusively (especially when $m = 2$) for applied smoothing and curve-fitting problems. A D^m-spline consists of piecewise polynomials of degree $2m-1$. It is meaningful, therefore, to inquire about the prior information structure implied by using D^m-splines. Let A be any set of functions and let $\mu(A)$, if it exists, denote the prior probability of A. For any constant c, let $A_c = \{f{:}f = g+c,\ g \in A\}$. Since $D^m(f+c) = D^m f$, adoption of a smoothing criterion D^m would imply

$$\mu(A+c) = \mu(A) \tag{4.1}$$

for all measurable A and real c. Clearly (4.1) cannot hold in any non-trivial probability space; hence using D^m-splines for curve-fitting or smoothing implies an improper prior. Since $P(\lambda) = (i\lambda)^m$ has the real root $\lambda = 0$, D^m-splines are excluded from our treatment in Section 3.

Let us approximate the operator D^m by $L_a = (D+a)^m$ for $a > 0$. Then since $P_a(\lambda) = (i\lambda + a)^m$ has no real roots, the theory of Sections 2 and 3 can be applied to define the conditional expectation $\hat{x}_a(t)$ which satisfies the conclusions of Theorem 3.1 or 3.2 with $L = (D+a)^m$. It can be shown that

$$\hat{x}_0(t) = \lim_{a \to 0} \hat{x}_a(t)$$

exists for all t and that $\hat{x}_0(t)$ is a D^m-spline. For $t < t_1$ and $t > t_n$, $\hat{x}_0(t)$ reduces to a polynomial of degree $m-1$. Moreover, among functions having absolutely continuous $(m-1)$th derivative, $\hat{x}_0(t)$ solves the curve-fitting or smoothing problem where $L = D^m$. Although $\hat{x}_a(t) \in \mathscr{L}_2(-\infty, \infty)$, it is not true that $\hat{x}_0(t) \in \mathscr{L}_2(-\infty, \infty)$ (since $\hat{x}_0(t)$ is a piecewise polynomial). Under a Bayesian statistical interpretation, D^m corresponds to an improper prior in which, for fixed t, $x(t)$ has infinite prior variance (hence one no longer requires that $\lim_t \hat{x}_0(t) = 0$) and the estimates $\hat{x}_0(t_i)$ are unbiased.

If in (1.4) we replace the infinite limits of integration by finite limits and L by D^m, the solutions to the corresponding minimization problems are well known to be ordinary (polynomial) splines with $2m-2$ continuous derivatives. In a forthcoming paper the present authors consider a stochastic model leading to these and more general spline functions.

Acknowledgment. We thank Horst Becker for suggesting the proof of Lemma 3.4 and Donald A. Jones for helpful conversations with one of the authors.

REFERENCES

[1] AHLBERG, J. H., NILSON, E. N., and WALSH, J. L. (1967). *The Theory of Splines and Their Application.* Academic Press, New York.

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

[3] GREVILLE, T. N. E., ed. (1969). *Theory and Applications of Spline Functions.* Academic Press, New York.

[4] HÁJEK, J. (1962). On linear statistical problems in stochastic processes. *Czechoslovak Math. J.* **12** 404–444.

[5] PARZEN, E. (1961). An approach to time series analysis. *Ann. Math. Statist.* **32** 951–989.

[6] SCHULTZ, M. H. and VARGA, R. S. (1967). *L*-splines. *Numer. Math.* **10** 345–369.