Reprinted from JOURNAL OF MATHEMATICAL ANALYSIS AND APPLICATIONS Vol. 33, No. 1, January 1971

Some Results on Tchebycheffian Spline Functions*

GEORGE KIMELDORF

Department of Statistics, Florida State University, Tallahassee, Florida 32306

AND

GRACE WAHBA

Department of Statistics, University of Wisconsin, Madison, Wisconsin 53706

Submitted by L. A. Zadeh

This report derives explicit solutions to problems involving Tchebycheffian spline functions. We use a reproducing kernel Hilbert space which depends on the smoothness criterion, but not on the form of the data, to solve explicitly Hermite-Birkhoff interpolation and smoothing problems. Sard's best approximation to linear functionals and smoothing with respect to linear inequality constraints are also discussed. Some of the results are used to show that spline interpolation and smoothing is equivalent to prediction and filtering on realizations of certain stochastic processes.

1. INTRODUCTION

Suppose we are given a closed interval $I = [a, b]$, a set $\{y_i\}$ of n constants, and a set $\{t_i\}$ of n distinct constants in (a, b). Consider the class of functions

$$\mathscr{H} = \{u : D^{m-1}u \text{ is absolutely continuous and } Lu \in \mathscr{L}_2(I)\} \tag{1.1}$$

where L is an m-th order linear differential operator. Of all functions $u \in \mathscr{H}$ satisfying

$$u(t_i) = y_i \qquad i = 1, 2,..., n, \tag{1.2}$$

we seek one, say $\hat{u}$, which minimizes $\int_a^b (Lu)^2$. It is now a classical result that when $L = D^m$ and $n \geqslant m$, then a solution $\hat{u}$ exists and is unique. The function $\hat{u}$ is called the $(2m - 1)$-th degree natural polynomial spline of interpolation to the data $\{(t_i, y_i)\}$.

* Sponsored by the Mathematics Research Center, United States Army, Madison, Wisconsin, under Contract No.: DA-31-124-ARO-D-462.

Recently (see, for example [3, 7]) spline problems have been placed in the context of an abstract Hilbert space. In this paper, the class $\mathscr{H}$ is made into a Hilbert space in which the linear functionals F_i defined by

$$F_i u = u(t_i) \tag{1.3}$$

are continuous, and hence have representers ψ_i so that (1.2) can be written

$$\langle \psi_i , u \rangle = y_i \qquad i = 1, 2,..., n. \tag{1.4}$$

Moreover, there exists a subspace $\mathscr{H}_1$ of $\mathscr{H}$ such that $\int (Lu)^2 = \| P_1 u \|^2$ for all $u \in \mathscr{H}$ where P_1 is the projection of $\mathscr{H}$ on $\mathscr{H}_1$.

There are several important advantages of placing spline problems in an abstract context. One is that existence and uniqueness proofs become consequences of straightforward results in the geometry of Hilbert space. A second advantage is the facility with which the results can be extended to more general operators than D^m and to more general linear functionals. An important disadvantage of an abstract approach is the nonconstructive nature of the proofs, for in general there is no way to construct the representer of a given linear functional.

De Boor and Lynch [4] suggested the use of reproducing kernel Hilbert spaces in problems involving splines. If the linear functional $N : f \to f(t)$ is continuous for each $t \in I$, then it is well known that there exists a reproducing kernel which, if known, allows one to construct the representer of any bounded linear functional. In this paper the reproducing kernel structure is combined with known results on total positivity to provide a unified approach to a variety of problems involving Tchebycheffian spline functions.

In Section 2, we exhibit the explicit reproducing kernel structure for $\mathscr{H}$. Section 3 uses the results of Section 2 to provide an explicit solution to the following generalized Hermite-Birkhoff problem: Given a set $\{N_i\}$ of continuous linear functionals on $\mathscr{H}$ and a set $\{y_i\}$ of scalars, find a function $\hat{u} \in \mathscr{H}$ which minimizes $\int_a^b (Lu)^2$ subject to the constraints $N_i u = y_i$. Section 4 discusses approximations to linear functionals.

Sections 5 and 6 consider variations of the interpolation problem. In Section 5 we treat the following smoothing problem: Rather than constrain the function according to $N_i u = y_i$ we insist only that $N_i u$ be "near" y_i. More precisely, we seek a function $\hat{u} \in \mathscr{H}$ which minimizes

$$\int_a^b (Lu)^2 + \sum (N_i u - y_i)\, s_{ij}(N_j u - y_j)$$

where $S = [s_{ij}]$ is a positive definite matrix.

Section 6 extends a problem recently solved by Ritter [11] of replacing the equations $N_i u = y_i$ by inequalities of the form

$$y_i \leqslant N_i u \leqslant z_i \tag{1.5}$$

where the y_i and z_i are prescribed. We seek a function $\hat{u} \in \mathscr{H}$ which minimizes $\int_a^b (Lu)^2$ subject to (1.5). The reproducing kernel structure is used to show that this problem is reducible to a standard quadratic programming problem.

In Section 7 we describe a stochastic process for which spline interpolation and smoothing are equivalent to minimum variance unbiased linear prediction and smoothing.

2. The Reproducing Kernel Structure

Without loss of generality we restrict our attention to the interval [0, 1]. Let $\{a_i\}$ be a set of m strictly positive functions such that $a_i \in C^i$ and without loss of generality we take $a_i(0) = 1$. Define the m-th order differential operator L by

$$L = D \frac{1}{a_1} D \frac{1}{a_2} \cdots D \frac{1}{a_m} \tag{2.1}$$

where D is the differentiation operator. Define operators M_i by

$$M_0 = I, \qquad M_i = D \frac{1}{a_{m+1-i}} M_{i-1}, \qquad i = 1, 2, \ldots, m-1 \tag{2.2}$$

where I is the identity operator. For $i = 1, 2, \ldots, m$, let the function ω_i be defined by

$$\begin{aligned}
\omega_1 &= a_m \\
\omega_2(t) &= a_m(t) \int_0^t a_m \quad t_{m-1})\, dt_{m-1} \\
&\vdots \\
\omega_m(t) &= a_m(t) \int_0^t a_{m-1}(t_{m-1}) \int_0^{t_{m-1}} a_{m-2}(t_{m-2}) \cdots \int_0^{t_2} a_1(t_1)\, dt_1 \cdots dt_{m-1}
\end{aligned} \tag{2.3}$$

so that we have

$$(M_i \omega_{j+1})(0) = \delta_{i,j}, \qquad i, j = 0, 1, \ldots, m-1 \tag{2.4}$$

where δ is the Kronecker delta. It is well known ([9] p. 379) that the set $\{\omega_i\}$ is an extended Tchebycheff system.

Let

$$\mathscr{H}_0 = \{u : D^{m-1}u \text{ is absolutely continuous and } Lu \equiv 0\}.$$

Hence $\mathscr{H}_0$ is an m-dimensional vector space spanned by the ω_i. From (2.2) and (2.4) we have that $\mathscr{H}_0$ is a Hilbert space with inner product

$$\langle u, v\rangle_0 = \sum_{i=0}^{m-1} [(M_i u)(0)]\,[(M_i v)(0)] \tag{2.5}$$

and orthonormal basis $\{\omega_i\}$. Let

$$\mathscr{H}_1 = \{u : D^{m-1}u \text{ is absolutely continuous}, Lu \in \mathscr{L}_2 \text{ and } (M_i u)(0) = 0, i = 0, 1,\ldots, m-1\}.$$

Clearly $\mathscr{H}_1$ is a Hilbert space with inner product

$$\langle u, v\rangle_1 = \int_0^1 (Lu)(Lv). \tag{2.6}$$

Let

$$\mathscr{H} = \{u : D^{m-1}u \text{ is absolutely continuous and } Lu \in \mathscr{L}_2\}$$

so that $\mathscr{H} = \mathscr{H}_0 \oplus \mathscr{H}_1$ is a Hilbert space with inner product

$$\langle u, v\rangle = \langle u, v\rangle_0 + \langle u, v\rangle_1 . \tag{2.7}$$

We now proceed to construct a reproducing kernel for $\mathscr{H}$; that is, we construct a function K defined on $[0, 1] \otimes [0, 1]$ such that for every fixed $s_0 \in [0, 1]$ we have: (1) $K(s_0, \cdot) \in \mathscr{H}$, and (2) $u \in \mathscr{H}$ if, and only if,

$$u(s_0) = \langle K(s_0, \cdot), u\rangle.$$

For an account of reproducing kernel Hilbert spaces, the reader is referred to Aronszajn [2].

Let K_1 be the symmetric positive definite function defined by

$$K_1(s, t) = a_m(s)\, a_m(t) \int_0^s \int_0^t \int_0^{s_{m-1}} \int_0^{t_{m-1}} \cdots \int_0^{s_2} \int_0^{t_2} \left[\prod_{j=1}^{m-1} a_j(s_j)\, a_j(t_j)\right] \min(s_1, t_1)\, ds_1\, dt_1 \cdots ds_{m-1}\, dt_{m-1} \tag{2.8}$$

and $G(s_0, \cdot) = LK_1(s_0, \cdot)$. Clearly, for fixed s_0, $K_1(s_0, \cdot) \in \mathscr{H}_1$. Moreover, it is known ([8] chap. 10) that G is the Green's function for the differential equation $Lu = b \in \mathscr{L}_2$ with boundary conditions $(M_i u)(0) = 0$, $i = 0, 1, \ldots, m-1$. Hence $u \in \mathscr{H}_1$ if, and only if, for any fixed $s_0 \in [0, 1]$

$$
\begin{aligned}
u(s_0) &= \int_0^1 G(s_0, t)\,[(Lu)(t)]\,dt \\
&= \int_0^1 [LK(s_0, t)]\,[Lu(t)]\,dt \\
&= \langle K_1(s_0, \cdot), u\rangle .
\end{aligned}
$$

Therefore, K_1 is the reproducing kernel for $\mathscr{H}_1$ and

$$K_1(s, t) = \int^1 G(s, r)\, G(t, r)\, dr.$$

Letting

$$K_0(s, t) = \sum_{i=1}^{m} \omega_i(s)\, \omega_i(t), \tag{2.9}$$

we verify readily that K_0 is the reproducing kernel for $\mathscr{H}_0$, and hence the function

$$K = K_0 + K_1 \tag{2.10}$$

is the reproducing kernel for $\mathscr{H}$.

The proof of the following lemma is an elementary consequence of the reproducing kernel structure.

LEMMA 2.1. *Let $\mathscr{H}$ be a Hilbert space with reproducing kernel Q. If N is a linear functional on $\mathscr{H}$ and ψ the function defined by $\psi(s_0) = NQ(s_0, \cdot)$, then N is continuous if, and only if, $\psi \in \mathscr{H}$, in which case $Nu = \langle \psi, u\rangle$ for all $u \in \mathscr{H}$ (i.e. ψ is the representer of N).*

The following lemma follows from the preceding.

LEMMA 2.2. *Let P_0 and P_1 be the projection operators in $\mathscr{H}$ onto $\mathscr{H}_0$ and $\mathscr{H}_1$ respectively. Then if N is a continuous linear functional with representer ψ, we have*

$$(P_0\psi)(s_0) = \sum_{i=1}^{m} \omega_i(s_0)\,(N\omega_i),$$

and

$$(P_1\psi)(s_0) = NK_1(s_0, \cdot).$$

We remark that the linear functional N defined by $Nu = u^{(j)}(t_0)$ for $0 < t_0 < 1$ and $j = 0, 1, \ldots, m-1$ is continuous on $\mathscr{H}$. This fact, whose proof follows from Lemma 2.1, will be used in the examples in the sequel.

3. The Generalized Hermite-Birkhoff Problem

We are now in a position to solve a generalized Hermite-Birkhoff problem. In order to apply the results of Section 2 we need the following lemma.

LEMMA 3.1. *Let $\mathscr{H} = \mathscr{H}_0 \oplus \mathscr{H}_1$ be the direct sum of an m-dimensional Hilbert space $\mathscr{H}_0$ with basis $\{\omega_1, \omega_2, \ldots, \omega_m\}$ and any Hilbert space $\mathscr{H}_1$. Let P_0 and P_1 be the projections onto $\mathscr{H}_0$ and $\mathscr{H}_1$ respectively and let $\psi_1, \psi_2, \ldots, \psi_n$ be $n \geqslant m$ elements of $\mathscr{H}$ such that*

(i) *The set $\{P_0\psi_i : i = 1, 2, \ldots, n\}$ spans $\mathscr{H}_0$, and*

(ii) *The set $\{P_1\psi_i : i = 1, 2, \ldots, n\}$ is linearly independent.*

If $\mathbf{y}' = (y_1, y_2, \ldots, y_n)$ is any n-tuple of scalars, then the unique element $\hat{u} \in \mathscr{H}$ which minimizes $\langle P_1 u, P_1 u\rangle$ subject to the constraints

$$\langle \psi_i, u\rangle = y_i, \qquad i = 1, 2, \ldots, n \tag{3.1}$$

is

$$\hat{u} = \boldsymbol{\omega}'(T\Sigma^{-1}T')^{-1}\, T\Sigma^{-1}\mathbf{y} + \boldsymbol{\xi}'\Sigma^{-1}[I - T'(T\Sigma^{-1}T')^{-1}\, T\Sigma^{-1}]\,\mathbf{y} \tag{3.2}$$

where T is the $m \times n$ matrix $[\langle \omega_i, \psi_j\rangle]$, Σ is the $n \times n$ matrix $[\langle P_1\psi_i, P_1\psi_j\rangle]$, $\boldsymbol{\omega} = (\omega_1, \omega_2, \ldots, \omega_m)'$, and $\boldsymbol{\xi} = (P_1\psi_1, P_1\psi_2, \ldots, P_1\psi_n)'$.

Proof. By assumption, T and Σ are of full rank. We can write $\hat{u} = \boldsymbol{\omega}'\boldsymbol{\alpha} + \boldsymbol{\xi}'\boldsymbol{\beta} + x$ where $\boldsymbol{\alpha}'$ and $\boldsymbol{\beta}'$ are some (as yet undetermined) m-tuple and n-tuple respectively of scalars, $x \in \mathscr{H}_1$, and $\langle x, \xi_i\rangle = 0$. Clearly we must take $x = 0$. The set of constraints $\langle \psi_1, \hat{u}\rangle = y_i$ is equivalent to $T'\boldsymbol{\alpha} + \Sigma\boldsymbol{\beta} = \mathbf{y}$ or

$$\boldsymbol{\beta} = \Sigma^{-1}(\mathbf{y} - T'\boldsymbol{\alpha}). \tag{3.3}$$

We have

$$\langle P_1\hat{u}, P_1\hat{u}\rangle = \boldsymbol{\beta}'\Sigma\boldsymbol{\beta}. \tag{3.4}$$

If (3.3) is substituted into (3.4), it is trivial to verify that (3.4) is minimized if and only if

$$\boldsymbol{\alpha} = (T\Sigma^{-1}T')^{-1}\, T\Sigma^{-1}\mathbf{y}$$

whence

$$\boldsymbol{\beta} = \Sigma^{-1}[I - T'(T\Sigma^{-1}T')^{-1}\, T\Sigma^{-1}]\,\mathbf{y}.$$

The explicit solution to the generalized Hermite-Birkhoff problem is stated in the following theorem.

THEOREM 3.1. *Let L be an m-th order differential operator of the form* (2.1), *let $\mathscr{H}$ be the Hilbert space defined in Section 2, let $\{\omega_i : i = 1, 2,..., m\}$ be as defined by* (2.3), *and let K_1 be defined by* (2.8). *Suppose $\{N_i : i = 1, 2,..., n\}$ is a set of $n \geqslant m$ continuous linear functionals on $\mathscr{H}$ such that*

(i) *The rank of the $m \times n$ matrix $T = [N_j\omega_i]$ is m, and*

(ii) *The n functions ξ_i defined for fixed s_0 by $\xi_i(s_0) = N_i K_1(s_0, \cdot)$ are linearly independent.*

If $\mathbf{y} = (y_1, y_2, ..., y_n)'$ is any vector of scalars, then the unique function $\hat{u} \in \mathscr{H}$ which minimizes $\int_0^1 (Lu)^2$ subject to the constraints

$$N_i u = y_i, \qquad i = 1, 2,..., n \tag{3.5}$$

is given by (3.2) *where $\boldsymbol{\omega} = (\omega_1, \omega_2, ..., \omega_m)'$, $\boldsymbol{\xi} = (\xi_1, \xi_2, ..., \xi_n)'$ and Σ is the $n \times n$ matrix $[N_i\xi_j] = [\langle \xi_i, \xi_j \rangle]$.*

The proof of Theorem 3.1 follows directly from Lemma 3.1 by letting ψ_i be the representer of N_i and using the results of Section 2.

EXAMPLE 3.1. Let N_i be $N_i : f \to f(t_i)$ where the $t_i \in (0, 1)$ are distinct. Condition (i) of the theorem is satisfied because the set $\{\omega_j\}$ forms a Tchebycheff system and condition (ii) is satisfied because K_1 is positive definite on $(0, 1) \otimes (0, 1)$. The unique function $\hat{u} \in \mathscr{H}$ which minimizes $\int_0^1 (Lu)^2$ subject to the constraints $u(t_i) = y_i$ is called the natural L-spline of interpolation to the points $\{(t_i, y_i)\}$.

EXAMPLE 3.2. Let N_i be defined by $N_i u = u^{(m_i)}(t_i)$ where $m_i < m$ and $0 < t_i < 1$. It can be shown that K_1 is a Green's function of the type considered by Karlin ([8] chap. 10, Secs. 7, 8), and hence that condition (ii) of the theorem is satisfied. In general, however, condition (i) will fail. In the case $L = D^m$, condition (i) is equivalent to the condition of Schoenberg [15] that the problem be m-poised. The concept of m-poisedness is also studied by Ferguson [5].

4. BEST APPROXIMATION OF CONTINUOUS LINEAR FUNCTIONALS

Let us adopt the notation of Theorem 3.1 and suppose $\{N_i : i = 1, 2,..., n\}$ satisfies the hypotheses of the theorem. If N is a given continuous linear functional on $\mathscr{H}$ we desire an approximation $\hat{N}$ to N of the form

$$\hat{N} = \sum_{i=1}^{n} c_i N_i \tag{4.1}$$

where the c_i are constants such that

$$\hat{N}u = Nu \qquad \text{if} \qquad u \in \mathscr{H}_0. \tag{4.2}$$

Equation (4.2) implies that the representer, say v, of $\hat{N} - N$ belongs to $\mathscr{H}_1$, and hence

$$(\hat{N} - N)u = \langle v, u\rangle = \int (Lv)(Lu). \tag{4.3}$$

Among all approximations $\hat{N}$ of the form (4.1) subject to the constraint (4.2), we seek one for which $\int (Lv)^2$ is minimized. This functional $\hat{N}$ is called the best approximation to N in the sense of Sard [12]. We shall prove that $\hat{N}$ exists, is unique, and satisfies

$$\hat{N}u = N\hat{u} \tag{4.4}$$

for all $u \in \mathscr{H}$ satisfying (3.5) where $\hat{u}$ is as in Theorem 3.1. We first need the following lemma.

LEMMA 4.1. *Adopting the notation and hypotheses of Lemma 3.1, we let* ψ *be any fixed element of* $\mathscr{H}$. *Then there exists a unique element* $\hat{\psi} \in \mathscr{H}$ *of the form* $\hat{\psi} = \sum c_i\psi_i$ *which minimizes* $\| P_1(\hat{\psi} - \psi)\|^2$ *subject to the constraint* $\| P_0(\hat{\psi} - \psi)\|^2 = 0$. *Furthermore, if* $u \in \mathscr{H}$ *satisfies* (3.1), *then*

$$\langle u, \hat{\psi}\rangle = \langle \hat{u}, \psi\rangle. \tag{4.5}$$

Proof. Writing $\hat{\psi} = \sum c_i\psi_i$, we seek a vector $\mathbf{c} = (c_1, ..., c_n)'$ of scalars which minimizes $\mathbf{c}'\Sigma\mathbf{c} - 2\mathbf{w}'\mathbf{c}$ subject to the constraint $\mathbf{v} = T\mathbf{c}$ where $\mathbf{v}$ is the vector $[\langle \omega_j, \psi\rangle]$ and $\mathbf{w}$ is the vector $[\langle P_1\psi_j, \psi\rangle]$. To verify the unique solution

$$\mathbf{c}' = \mathbf{v}'(T\Sigma^{-1}T')^{-1}\, T\Sigma^{-1} + \mathbf{w}\Sigma^{-1}[I - T'(T\Sigma^{-1}T')^{-1}\, T\Sigma^{-1}], \tag{4.6}$$

we let $\mathbf{e}$ be any vector such that $\mathbf{v} = T(\mathbf{c} + \mathbf{e})$. Hence we have $T\mathbf{e} = \mathbf{0}$ so that

$$(\mathbf{c}' + \mathbf{e}')\,\Sigma(\mathbf{c} + \mathbf{e}) - 2\mathbf{w}'(\mathbf{c} + \mathbf{e}) - \mathbf{c}'\Sigma\mathbf{c} + 2\mathbf{w}'\mathbf{c} = \mathbf{e}'\Sigma\mathbf{e},$$

which is non-negative, and which is zero if, and only if, $\mathbf{e} = \mathbf{0}$. Equation (4.5) can be verified directly using (3.2) and (4.6).

To prove (4.4) and the existence and uniqueness of $\hat{N}$, we use Lemma 4.1 and let N have representer ψ, N_i have representer ψ_i, and $\hat{N}$ have representer $\hat{\psi}$.

Equation (4.4) states that the "best" approximation to a functional N operating on a function u is the functional N operating on the spline which "interpolates" u. Schoenberg [14] proved (4.4) in the case when $L = D^m$

and $N_i u = u(t_i)$. Karlin and Ziegler [10] demonstrated (4.4) for L of the form (2.1) with $N_i u = u^{(m_i)}(t_i)$ for $0 \leqslant m_i \leqslant m - 1$, while Jerome and Schumaker [7] considered general continuous linear functionals with L of the form $L = \Sigma b_j D^j$ for $b_j \in C^j$.

5. The Generalized Smoothing Problem

Lemma 5.1. *Let $\mathscr{H} = \mathscr{H}_0 \oplus \mathscr{H}_1$ be the direct sum of an m-dimensional Hilbert space $\mathscr{H}_0$ with basis $\{\omega_1, \omega_2, \ldots, \omega_m\}$ and any Hilbert space $\mathscr{H}_1$. Let P_0 and P_1 be the projections onto $\mathscr{H}_0$ and $\mathscr{H}_1$ respectively. Suppose $S = [s_{ij}]$ is a $p \times p$ positive definite matrix and $\phi_1, \phi_2, \ldots, \phi_p$ are $p \geqslant m$ elements of $\mathscr{H}$ such that*

(i) *The set $\{P_0\phi_i : i = 1, 2, \ldots, p\}$ spans $\mathscr{H}_0$, and*

(ii) *The set $\{P_1\phi_i : i = 1, 2, \ldots, p\}$ is linearly independent.*

If $\mathbf{z}' = (z_1, z_2, \ldots, z_p)$ is any p-tuple of scalars, then the unique element $\hat{u} \in \mathscr{H}$ which minimizes

$$\sum\sum (\langle \phi_i, u\rangle - z_i)\, s_{ij}(\langle \phi_j, u\rangle - z_j) + \langle P_1 u, P_1 u\rangle \tag{5.1}$$

is

$$\hat{u} = \boldsymbol{\omega}'(UM^{-1}U')^{-1}\, UM^{-1}\mathbf{z} + \boldsymbol{\eta}'M^{-1}[I - U'(UM^{-1}U')^{-1}\, UM^{-1}]\,\mathbf{z} \tag{5.2}$$

where U is the $m \times p$ matrix $[\langle \omega_i, \phi_j\rangle]$, M is the $p \times p$ matrix

$[\langle P_1\phi_i, P_1\phi_j\rangle] + S^{-1}$, $\quad \boldsymbol{\eta} = (P_1\phi_1, \ldots, P_1\phi_p)'$ *and* $\quad \boldsymbol{\omega} = (\omega_1, \omega_2, \ldots, \omega_m)'$.

Proof. We can write $\hat{u} = \boldsymbol{\omega}'\boldsymbol{\alpha} + \boldsymbol{\eta}'\boldsymbol{\beta} + x$ where $\boldsymbol{\alpha}'$ and $\boldsymbol{\beta}'$ are some (as yet undetermined) m-tuple and p-tuple respectively of scalars, $x \in \mathscr{H}_1$, and $\langle x, \eta_i\rangle = 0$. Clearly we must take $x = 0$. The quantity (5.1) to be minimized is

$$(U'\boldsymbol{\alpha} + \Sigma\boldsymbol{\beta} - \mathbf{z})'\, S(U'\boldsymbol{\alpha} + \Sigma\boldsymbol{\beta} - \mathbf{z}) + \boldsymbol{\beta}'\Sigma\boldsymbol{\beta},$$

which is minimized if, and only if,

$$\boldsymbol{\alpha} = (UM^{-1}U')^{-1}\, UM^{-1}\mathbf{z} \qquad \text{and} \qquad \boldsymbol{\beta} = -\, M^{-1}(U'\boldsymbol{\alpha} - \mathbf{z}).$$

The explicit solution to the generalized smoothing problem provided by Section 2 and Lemma 5.1 is stated in the following theorem:

Theorem 5.1. *Let L be an m-th order differential operator of the form* (2.1), *$\mathscr{H}$ be the Hilbert space of Section 2, $\{\omega_i : i = 1, 2, \ldots, m\}$ be as defined by* (2.3), *and let K_1 be defined by* (2.8). *Suppose $\{R_i : i = 1, 2, \ldots, p\}$ is a set of $p \geqslant m$ continuous linear functionals on $\mathscr{H}$ such that*

(i) *The rank of the* $m \times p$ *matrix* $U = [R_j\omega_i]$ *is* m, *and*

(ii) *The* p *functions* η_i *defined by* $\eta_i(s_0) = R_iK_I(s_0, \cdot)$ *are linearly independent.*

If $\mathbf{z} = (z_1, ..., z_p)'$ *is any vector of scalars and* $S = [s_{ij}]$ *is a* $p \times p$ *positive define matrix, then the unique element* $\hat{u} \in \mathscr{H}$ *which minimizes*

$$\sum\sum (R_iu - z_i)\, s_{ij}(R_ju - z_j) + \int_0^1 (Lu)^2 \tag{5.3}$$

is given by (5.2) *where* $\boldsymbol{\omega} = (\omega_1, \omega_2, ..., \omega_m)'$, $\boldsymbol{\eta} = (\eta_1, \eta_2, ..., \eta_p)'$, *and* $M = [R_i\eta_j] + S^{-1}$.

We remark that the function $\hat{u}$ of Theorem 5.1 is a spline function in the sense that it solves a generalized Hermite-Birkhoff interpolation problem. In fact, if we let $z_i = R_i\hat{u}$, then it is clear that $u = \hat{u}$ is the unique function in $\mathscr{H}$ which minimizes $\int (Lu)^2$ subject to the constraints $R_iu = z_i$.

EXAMPLE 5.1. Let R_i be given by $R_i : u \to u(t_i)$ where the $t_i \in (0, 1)$ are distinct, and let $L = D^m$. Then $\hat{u}$ is a polynomial spline function. This result, where S is a diagonal matrix, was announced by Schoenberg [13] in 1964. An analogous result with L a general linear differential operator was announced by Greville and Schoenberg [6] in 1965. Abstract existence and uniqueness theorems for smoothing spline functions were proved by Anselone and Laurent [1] in 1967.

Theorem 3.1 presents a solution to a minimization problem in which the values of certain linear functionals are constrained, while in Theorem 5.1, the values of the functionals are not constrained, but appear in the term to be minimized. We state the following theorem, whose proof is analogous to that of the preceding results, of which Theorems 3.1 and 5.1 are special cases.

THEOREM 5.2. *In the notation of Theorems 3.1 and 5.1, suppose* $\{N_i : i = 1, 2, ..., n\}$ *and* $\{R_i : i = 1, 2, ..., p\}$ *are sets of continuous linear functionals on* $\mathscr{H}$ *such that*

(i) *The rank of the* $m \times (p + n)$ *matrix* $V = [T; U]$ *is* m, *and*

(ii) *The set* $\{\xi_i\} \cup \{\eta_j\}$ *of* $n + p$ *functions is linearly independent.*

If $\mathbf{w} = (y_1, y_2, ..., y_n; z_1, z_2, ..., z_p)'$ *is any vector of scalars and* $S = [s_{ij}]$ *is any positive definite* $p \times p$ *matrix, then the unique function* $\hat{u} \in \mathscr{H}$ *which minimizes*

$$\sum\sum (R_iu - z_i)\, s_{ij}(R_ju - z_j) + \int_0^1 (Lu)^2$$

subject to the constraints

$$N_i u = y_i \qquad i = 1, 2, \ldots, n$$

is

$$\hat{u} = \boldsymbol{\omega}(VM^{-1}V')^{-1}\, VM^{-1}\mathbf{w} + \begin{bmatrix}\boldsymbol{\xi}\\ \boldsymbol{\eta}\end{bmatrix}' M^{-1}[I - V(VM^{-1}V')\, VM^{-1}]\,\mathbf{w}$$

where M is the $(n+p) \times (n+p)$ partitioned matrix

$$M = \begin{bmatrix}\Sigma & \Lambda \\ \Lambda' & \Gamma + S^{-1}\end{bmatrix}$$

in which $\Sigma = [N_i\xi_j]$, $\Lambda = [N_i\eta_j]$, and $\Gamma = [R_i\eta_j]$.

6. Linear Inequality Constraints

Klaus Ritter [11] recently proposed replacing equations (3.5) by inequalities. He showed the resulting minimization problem to be reducible to a standard problem in quadratic programming. In particular, we have the following result.

Theorem 6.1. *Let L, $\mathscr{H}$, $\boldsymbol{\omega}$, K, T, Σ, and $\boldsymbol{\xi}$ be defined as in Theorem 3.1. Let $\{N_i\}$ be a set of n continuous linear functionals on $\mathscr{H}$, and let $\mathbf{y}_j = (y_{j1}, \ldots, y_{jn})'$ be n-tuples of scalars ($j = 1, 2$). Then the function $\hat{u} = \hat{\boldsymbol{\alpha}}'\boldsymbol{\omega} + \hat{\boldsymbol{\beta}}'\boldsymbol{\xi}$ minimizes $\int (Lu)^2$ subject to the constraints*

$$y_{1i} \leqslant N_i u \leqslant y_{2i}, \qquad i = 1, 2, \ldots, n \tag{6.1}$$

if and only if, $\boldsymbol{\alpha} = \hat{\boldsymbol{\alpha}}$ and $\boldsymbol{\beta} = \hat{\boldsymbol{\beta}}$ minimizes $\boldsymbol{\beta}'\Sigma\boldsymbol{\beta}$ subject to the constraints $\mathbf{y}_1 \leqslant T'\boldsymbol{\alpha} + \Sigma\boldsymbol{\beta} \leqslant \mathbf{y}_2$.

The proof of Theorem 6.1 follows immediately from the following lemma, whose proof is elementary.

Lemma 6.1. *Let $\mathscr{H}$, $\mathscr{H}_0$, $\mathscr{H}_1$, $\boldsymbol{\omega}$, P_0, and P_1 be as in Lemma 3.1. Let $\{\psi_i\}$ be n elements of $\mathscr{H}$, $\boldsymbol{\xi} = (P_1\psi_1, P_1\psi_2, \ldots, P_1\psi_n)'$, and let $\mathbf{y}_j = (y_{j1}, y_{j2}, \ldots, y_{jn})'$ be n-tuples of scalars for $j = 1, 2$. Then the element $\hat{u} = \boldsymbol{\alpha}'\boldsymbol{\omega} + \boldsymbol{\beta}'\boldsymbol{\xi} \in \mathscr{H}$ minimizes $\langle P_1 u, P_1 u\rangle$ subject to the constraints*

$$y_{1i} \leqslant \langle\psi_i, u\rangle \leqslant y_{2i}, \qquad i = 1, 2, \ldots, n \tag{6.2}$$

if and only if, $\boldsymbol{\alpha} = \hat{\boldsymbol{\alpha}}$ and $\boldsymbol{\beta} = \hat{\boldsymbol{\beta}}$ minimizes $\boldsymbol{\beta}'\Sigma\boldsymbol{\beta}$ subject to the constraints $\mathbf{y}_1 \leqslant T'\boldsymbol{\alpha} + \Sigma\boldsymbol{\beta} \leqslant \mathbf{y}_2$.

7. Spline Interpolation as Statistical Prediction and Filtering

Let $Y(t)$, $0 \leqslant t \leqslant 1$, be the stochastic process defined by

$$Y(t) = \sum_{i=1}^{m} \theta_i \omega_i(t) + X(t), \qquad 0 \leqslant t \leqslant 1 \tag{7.1}$$

where $\{\theta_i\}_{i=1}^m$ are random variables independent of $X(t)$, $0 \leqslant t \leqslant 1$, $\{\omega_i(t)\}_{i=1}^m$ are given by (2.3), and $X(t)$, $0 \leqslant t \leqslant 1$ is the zero mean Gaussian stochastic process with covariance given by

$$EX(s)\, X(t) = K_1(s, t), \tag{7.2}$$

$K_1(s, t)$ being given by (2.8). $X(t)$ has a representation as

$$X(t) = a_m(t) \int_0^t a_{m-1}(t_{m-1})\, dt_{m-1} \int_0^{t_{m-1}} a_{m-2}(t_{m-2})\, dt_{m-2} \cdots \int_0^{t_2} a_1(t_1)\, dW(t_1) \tag{7.3}$$

where $W(u)$, $0 \leqslant u \leqslant 1$, is the Wiener process. Hence, we may say that $Y(t)$ formally satisfies the stochastic differential equation

$$LY(t) = \frac{dW(t)}{dt}$$

with the (random) boundary conditions $M_i Y(0) = \theta_{i+1}$, $i = 0, 1, 2,..., m-1$. $dW(t)/dt$ is commonly referred to as "white noise". Let now $\{\theta_i\}_{i=1}^m$ be independent normal random variables with mean zero and variance 1. We may define the Hilbert space $\mathscr{G}$ spanned by the family of random variables $\{y(t), 0 \leqslant t \leqslant 1\}$ as all finite linear combinations of random variables of the form

$$\rho = \sum_\nu c_\nu Y(t)_\nu$$

plus the closure of this linear manifold under the norm induced by the inner product

$$\langle \rho_1, \rho_2 \rangle = E\rho_1\rho_2 . \tag{7.4}$$

(We have $EY(s)\, Y(t) = K(s, t)$, where K is given by (2.10).)

$\mathscr{G}$ is isomorphic to $\mathscr{H}$ under the correspondence induced by

$$Y(s_0) \leftrightarrow K(s_0, \cdot), \qquad s_0 \in [0, 1]. \tag{7.5}$$

Let $\rho_0, \rho_1, ..., \rho_n$ be n random variables in $\mathscr{G}$ and let $u, \psi_1, \psi_2, ..., \psi_n$ be the $n+1$ elements in $\mathscr{H}$ which correspond to $\rho_0, \rho_1, ..., \rho_n$ under correspondence (7.5). We shall call an estimate $\tilde{\rho}_0$ of ρ_0 unbiased with respect to $\theta = (\theta_1, \theta_2, \dots \theta_m)'$ if

$$E(\tilde{\rho}_0 \mid \theta) = E(\rho_0 \mid \theta). \tag{7.6}$$

Let $\hat{\rho}_0$ be the minimum variance unbiased estimate of ρ_0, based on data $\{\rho_i = y_i\}$, $i = 1, 2, \dots, M$, that is, $\hat{\rho}_0$ is that linear combination $\sum_{i=1}^n d_i \rho_i$ which minimizes

$$E\left(\hat{\rho}_0 - \sum_{i=1}^n d_i \rho_i\right)^2 \tag{7.7}$$

subject to

$$E(\hat{\rho}_0 - \rho_0 \mid \theta) = 0. \tag{7.8}$$

This last condition can be shown to be equivalent to

$$\sum_{i=1}^n \theta_i (\hat{N}_0 - N_0)\, w_i(\cdot) \equiv 0 \tag{7.9}$$

where $\hat{N}_0$ and N_0 are the continuous linear functionals (i.e. elements in the dual of $\mathscr{H}$) whose representers $\hat{u}$ and u correspond to the random variables $\hat{\rho}_0$ and ρ_0 under the correspondence (7.5). As a consequence of Lemma 4.1, and the isomorphism, we have

THEOREM 7.1. *Let Σ, the $n \times n$ matrix $[E(\rho_i \rho_j \mid \theta)]$, and T, the $m \times n$ matrix $[E\theta_i \rho_j]$, be of full rank. Then the minimum variance unbiased estimate $\hat{\rho}_0$ for ρ_0, based on data $\{\rho_i = y_i\}_{i=1}^n$ is given by*

$$\hat{\rho}_0 = N_0 \hat{u}$$

where $\hat{u}$ is given by (3.2) *and N_0 is the continuous linear functional whose representer corresponds to ρ_0.*

EXAMPLE 7.1. Let $\rho_i = Y(t_i)$ for $t_i \in (0, 1)$ distinct. Then the minimum variance unbiased estimate $\hat{Y}(t)$ for $Y(t)$ based on data $\{Y(t_i) = y_i\}_{i=1}^n$ is, considered as a function of t, the natural L-spline of interpolation to the points $\{(t_i, y_i)\}$.

The generalized smoothing problem of Section 5 also has a statistical interpretation. We cite as a theorem only one form, which displays the smoothing problem as equivalent to the extraction of signal from noise.

THEOREM 7.2. *Let*

$$Z(t) = Y(t) + \epsilon(t)$$

$Y(t)$

where $\epsilon(t)$ *is a Gaussian noise process independent of* $Z(t)$ *with* $E\epsilon(s)\,\epsilon(t) = A(s, t)$. *Let* $\hat{Y}(t)$ *be the minimum variance unbiased estimate of* $Y(t)$ *based on data* $Z(t_i) = z_i$, $i = 1, 2,..., p$. *Then* $\hat{Y}(t)$ *is given by*

$$Y(t) = \langle K(t, \cdot), \hat{u} \rangle = u(t)$$

where $\hat{u}$ *is given by* (5.2) *with* $\phi_i(\cdot) = K(t_i, \cdot)$, $i = 1, 2,..., p$ *and* S^{-1} *is the* $p \times p$ *matrix* $[A(t_i, t_j)]$.

References

1. P. M. Anselone and P. J. Laurent, A general method for the construction of interpolating or smoothing spline functions, MRC Technical Summary Report #834, 1967, Mathematics Research Center, Univ. of Wisconsin.
2. N. Aronszajn, Theory of reproducing kernels, *Trans. Amer. Math. Soc.* **68** (1950), 337–404.
3. M. Attéia, Generalisation de la définition et des propriétés des "spline-fonctions," *C. R. Acad. Sci. Paris* **260** (1965), 3550–3553.
4. C. deBoor and R. E. Lynch, On splines and their minimum properties, *J. Math. Mech.* **15** (1966), 953–969.
5. D. Ferguson, The question of uniqueness for G. D. Birkhoff interpolation problems, *J. Approximation Theory* **2** (1969), 1–28.
6. T. N. E. Greville and I. J. Schoenberg, Smoothing by generalized spline functions (Abstract), *SIAM Rev.* **7** (1965), 617.
7. J. W. Jerome and L. L. Schumaker, On Lg-splines, *J. Approximation Theory* **2** (1969), 29–49.
8. S. Karlin, "Total Positivity," Vol. I, Stanford University Press, 1968.
9. S. Karlin and W. J. Studden, "Tchebycheff Systems: with Applications in Analysis and Statistics," Interscience, 1966.
10. S. Karlin and Z. Ziegler, Chebyshevian spline functions, *SIAM J. Numer. Anal.* **3** (1966), 514–543.
11. K. Ritter, Splines and quadratic programming, Conference on Approximations, University of Wisconsin, 1969.
12. A. Sard, "Linear Approximation," Mathematical Surveys No. 9, American Mathematical Society, Providence, 1963.
13. I. J. Schoenberg, Spline functions and the problem of graduation, *Proc. Nat. Acad. Sci. U.S.A.* **52** (1964), 947–950.
14. I. J. Schoenberg, On best approximations of linear operators, *Indag. Math.* **26** (1964), 155–163.
15. I. J. Schoenberg, On the Ahlberg-Nilson extension of spline interpolation: The *g*-splines and their optimal properties, *J. Math. Anal. Appl.* **21** (1968), 207–231.

PRINTED IN BELGIUM BY THE ST. CATHERINE PRESS, TEMPELHOF, 37, BRUGES, LTD.