THE UNIVERSITY OF WISCONSIN
MATHEMATICS RESEARCH CENTER, UNITED STATES ARMY

Contract No.: DA-31-124-ARO-D-462

ESTIMATING DERIVATIVES
FROM OUTER SPACE

Grace Wahba

This document has been approved for public release and sale; its distribution is unlimited.
MRC Technical Summary Report # 989
May 1969

Madison, Wisconsin 53706
Let $f(t)$, $0 \leq t \leq 1$, be an unknown function in the class $\mathcal{H}_m: \{f: f^{(m)} \in L^2[0,1], f^{(m-1)} \text{ absolutely continuous }\}$. We say $f_1 \in \mathcal{H}_m$ is smoother than $f_2 \in \mathcal{H}_m$ if

$$\int_0^1 [f_1^{(m)}(t)]^2 \, dt < \int_0^1 [f_2^{(m)}(t)]^2 \, dt$$

Values of $f(t)$ for $t = t_1, t_2, \ldots, t_n$, $n > m$ are observed on a spacecraft, but, due to transmission limitations it is only known on the ground that

$$a_i \leq f(t_i) \leq b_i, \quad i = 1, 2, \ldots, n$$

where the intervals $[a_i, b_i]$ are some choice of the quantization levels of the analog to digital conversion device in the spacecraft. It is desired to estimate $f'(t_0)$ for some $t_0$. An estimate of the form $\hat{f}'(t_0)$ where $\hat{f}(t)$ is a smoothest function in $\mathcal{H}_m$ satisfying

$$a_i \leq \hat{f}(t_i) \leq b_i$$

is suggested. The smoothest functions $\hat{f}(t)$ satisfying (*) are characterized via a theorem of Ritter, and their uniqueness discussed. A stochastic model for $f(t)$ is constructed for which $\hat{f}'(t_0)$ is a certain type of maximum likelihood for $f'(t_0)$. The smoothness criteria is generalized and the resulting solutions $\hat{f}(t)$ characterized, and analogous stochastic models constructed. The associated stochastic models essentially consist of any zero mean Gaussian stochastic processes continuous in quadratic mean.
1. Introduction

A continuous signal \( f(t) \), \( 0 \leq t \leq 1 \) exists in nature and is partly measured by an experimental device installed in a space probe. For example \( f(t) \) is the interplanetary magnetic field strength as a function of time, in the vicinity of a space probe on its way to Jupiter. \( f(t) \) is measured by a magnetometer at times \( t = t_1, t_2, \ldots, t_n \). However, due to severe limitations on the rate at which data can be transmitted from great distances, the signal \( f(t_i) \), \( i = 1, 2, \ldots, n \) observed in the space probe, is coded into one of \( 2^k \) possible "bit levels", and the experimenter on the ground recovers only the information that

\[
a_i \leq f(t_i) \leq b_i , \quad i = 1, 2, \ldots, n .
\]

In this example the interval \([a_i, b_i]\) is one of \( 2^k \) possible intervals which depend on the instrument design. The typical situation in experiments carried on space probes is for \( k \) to be small, say 4 or 5. Thus the effect of so-called quantization may be severe and it may be inappropriate to treat this discretization as simple roundoff error.

Actually, the experimenter is interested in estimating \( f'(t_0) \) at \( t_0 = .321 \), because at that time, a disturbance in, say, the solar wind, was known to occur and it is desired to see whether this affected the rate of change of the magnetic field strength. More generally, the experimenter may be interested in cross correlating the magnetic field strength with say, the solar wind number.

Sponsored by the Mathematics Research Center, United States Army, Madison, Wisconsin, under Contract No. DA-31-124-ARO-D-462.
density observed from the same vehicle, and hence is interested in estimating $f(t)$ for all $0 \leq t \leq 1$.

Let

$$a = (a_1, a_2, \ldots, a_n)$$
$$b = (b_1, b_2, \ldots, b_n)$$

with

$$a_i \leq b_i$$

and let $\mathcal{R}_{ab}$ be the interval in Euclidean $n$-space containing all $n$-dimensional vectors with $i$th coordinate $y_i$ satisfying

$$a_i \leq y_i \leq b_i.$$

Let

$$\vec{f} = (f(t_1), \ldots, f(t_n)).$$

We say we observe the "data" $\mathcal{R}_{ab}$ if we observe that $\vec{f} \in \mathcal{R}_{ab}$. In Section 2 we suggest a simple and natural criterion for estimating $f'(t_0)$ given the data $\mathcal{R}_{ab}$. We show how the estimate is to be calculated from the data. The estimate, which may not be unique, is the solution to a quadratic programming problem, which may be solved numerically by standard methods. In Section 3 we construct a statistical model for the unknown function $f(t)$; that is we consider that it is a stochastic process. The model so constructed leads precisely to the estimate of Section 2, which was obtained without reference to any statistical model.

In Section 4 we show how the criteria of Sections 2 and 3 generalize to a wide class of models, and a simple general example is given. In practice, the experimenter must select his model based on his understanding or prior opinion of the mechanism generating $f(t)$. 
2. The Criteria and Algorithm for Estimating $f'(t_0)$

Let $m$ be a fixed integer $>1$, and consider the class $\mathcal{H}_m$ of all functions defined on $[0, 1]$ as follows $\mathcal{H}_m = \{f: f^{(m-1)} \text{ absolutely continuous}, \ f^{(m)} \text{ square integrable on } [0, 1]\}$. We will say that $f_1(t)$ is smoother than $f_2(t)$ if

$$\int_0^1 [f_1^{(m)}(t)]^2 \, dt \leq \int_0^1 [f_2^{(m)}(t)]^2 \, dt.$$  \hspace{1cm} (2.1)

Let $\hat{f}(t)$ be a smoothest function $\in \mathcal{H}_m$ satisfying the inequalities

$$a_i \leq \hat{f}(t_i) \leq b_i, \quad i = 1, 2, \ldots, n$$ \hspace{1cm} (2.2)

for $n > m$ specified points $t_1 < t_2 < \ldots < t_n$.

We will take $\hat{f}(t_0)$ as our estimate of $f(t_0)$. It is clear that there may be data $\mathcal{R}_{ab}$ for which $\hat{f}(t)$ is not unique. Any set $\mathcal{R}_{ab}$ which admits more than one polynomial of degree $m-1$ is such an example. The class of all solutions $\hat{f}(t)$ is given by the following

**Theorem 1 (Ritter).**

Any solution $\hat{f}(t)$ to the problem of finding $\hat{f}(t) \in \mathcal{H}_m$ to minimize

$$\int_0^1 (f^{(m)}(t))^2 \, dt$$ \hspace{1cm} (2.2a)

subject to

$$a_i \leq \hat{f}(t_i) \leq b_i, \quad i = 1, 2, \ldots, n$$ \hspace{1cm} (2.2b)

is given by

$$\hat{f}(t) = \sum_{\nu=1}^{m} \phi_\nu(t) \theta_\nu + \sum_{i=1}^{n} K(t_i, t)x_i^*$$ \hspace{1cm} (2.3)

where

$$\phi_\nu(t) = \frac{t^{\nu-1}}{\nu!}, \quad \nu = 1, 2, \ldots, m$$  \hspace{1cm} (2.4)
\[ K(s, t) = \int_{0}^{1} G_m(s, u)G_m(t, u)du, \quad G_m(r, u) = \frac{(r-u)^{m-1}}{(m-1)!}, \quad (2.5) \]

\[
(u)_+ = u, \quad u \geq 0 \\
= 0 \quad \text{otherwise}
\]

\[ \theta^* = (\theta_1^*, \theta_2^*, \ldots, \theta_m^*) \]

\[ x^* = (x_1^*, x_2^*, \ldots, x_n^*) \]

and \( \theta^*, x^* \) are any solutions to the quadratic programming problem: Find \( x^*, \theta^* \) to minimize

\[ x^* \sum x^* \quad (2.6a) \]

subject to

\[ a \leq T \theta^* + \sum x^* \leq b \quad (2.6b) \]

where

\[ a' = (a_1', a_2', \ldots, a_n') \]

\[ b' = (b_1', b_2', \ldots, b_n') \]

\( T \) is the \( n \times m \) matrix with \( i \)th entry \( \phi_i(t_i) \), \( \sum \) is the \( n \times n \) matrix with \( ij \)th entry \( K(t_i, t_j) \) and the inequalities in (2.6b) are interpreted component wise.

This theorem was proved by K. Ritter [9]. We present a complete proof here in a form that will generalize easily. We remark that any solution \( \hat{f}(t) \) of the form of (2.3) can be shown to be a polynomial of degree at most \( 2m-1 \) in each of the intervals \( (t_i, t_{i+1}) \), \( i = 1, 2, \ldots, n-1 \), a polynomial of order at most \( m-1 \) in \([0, t_1) \) and \((t_n, 1] \) and \( \hat{f}(t) \) possess \( 2m-2 \) continuous derivatives. It is a polynomial spline function, about which there is a considerable literature (see, for example [1]).
Proof: It may be verified that $\mathcal{H}_m$ is a Hilbert space with inner product
$$<g, h> = \left\{ \sum_{\nu=1}^{m} \left( \frac{1}{(\nu-1)!} \right)^2 g^{(\nu-1)}(0) h^{(\nu-1)}(0) \right\}$$
$$+ \left\{ \int_{0}^{1} g^{(m)}(u) h^{(m)}(u) \, du \right\}$$

$\mathcal{H}_m$ may be decomposed as the direct sum of $\mathcal{Q}_0$ and $\mathcal{Q}_1$ where $\mathcal{Q}_0$ is the $m$ dimensional space spanned by the polynomials $\{\phi_{\nu}(t)\}_{\nu=1}^{m}$ with inner product given by the first term in brackets in (2.7) and $\mathcal{Q}_1$ is the Hilbert space of functions $\{f:t^{(\nu)}(0)=0, \, \nu = 0, 1, 2 \ldots m-1\}$ with inner product given by the second term in brackets in (2.7). The $\{\phi_{\nu}(t)\}_{\nu=1}^{m}$ are an orthonormal set with this inner product, that is
$$<\phi_{\nu}, \phi_{\mu}> = \delta_{\mu\nu}.$$  (2.8)

We may, by integration by parts, obtain that, for any $f(t) \in \mathcal{H}_m$,
$$f(t) = \left\{ \sum_{\nu=1}^{m} \phi_{\nu}(t) \frac{f^{(\nu-1)}(0)}{(\nu-1)!} \right\} + \left\{ \int_{0}^{1} G_{m}(t, u) f^{(m)}(u) \, du \right\}.$$  (2.9)

Letting $P_0$ and $P_1$ be the projection operators onto $\mathcal{Q}_0$ and $\mathcal{Q}_1$, the two terms in brackets in (2.9), are then obviously $P_0 f$ and $P_1 f$ respectively.

Let
$$H(s, t) = \sum_{\nu=1}^{m} \hat{\phi}_{\nu}(s) \phi_{\nu}(t) + K(s, t),$$  (2.10)

and let
$$\psi_{\nu}(t) = H(t_{\nu}, t).$$  (2.11)

Also note that $\frac{\partial^{\nu}}{\partial t^{\nu}} K(s, t) \big|_{t=0} = 0, \, \nu = 0, 1, 2 \ldots m-1$  (2.12)
\[
\frac{\partial^m}{\partial t^m} K(s, t) = G_m(s, t) .
\]

By setting \( t = t_i \) in (2.9) and using (2.12) and (2.13) we may obtain the formula

\[
f(t_i) = \langle f, \psi_i \rangle , \quad i = 1, 2 \ldots n .
\]

The problem may now be written: Find \( \hat{\xi} \in \mathcal{H}_m \) to minimize

\[
\langle P_1 \hat{\xi}, P_1 \hat{\xi} \rangle
\]

subject to

\[
a_i \leq \langle \hat{\xi}, \psi_i \rangle \leq b_i , \quad i = 1, 2 \ldots n .
\]

Now let

\[
k_i(t) = K(t_i, t) .
\]

By (2.12),

\[
P_0 k_i = 0 , \quad i = 1, 2, \ldots n .
\]

Thus, it follows from the properties of Hilbert spaces that any \( f \in \mathcal{H}_m \) has a representation

\[
f(t) = \sum_{\nu=1}^{m} \phi_\nu(t) \theta_\nu + \sum_{i=1}^{n} k_i(t) x_i + \rho(t)
\]

where

\[
\langle \rho, k_i \rangle = \langle \rho, \phi_\nu \rangle = 0 , \quad i = 1, 2, \ldots n , \quad \nu = 1, 2 \ldots m
\]

(2.9) and (2.18) imply that \( \rho(t_i) = 0 , \quad i = 1, 2 \ldots n \). The representation (2.17) is unique, since it can be verified that, for \( \{t_i \} \) distinct, \( t_i > 0 \), the \( \{k_i(t)\}_{i=1}^{m} \) are linearly independent. Using (2.13) and (2.16) gives

\[
\langle k_i, k_j \rangle = K(t_i, t_j) .
\]
Using (2.16), (2.17), (2.18) and (2.19) we now have

\[ < P_1 f, P_1 f > = x' \sum x + < \rho, \rho > \]  \hspace{1cm} (2.20)

\[ < f, \psi_1 > = \sum_{\nu=1}^{m} \phi_\nu (t_\nu) \theta_\nu + \sum_{j=1}^{n} K(t_1, t_j) x_j . \]  \hspace{1cm} (2.21)

Since we evidently take \( \rho = 0 \) to minimize (2.20), the theorem is proved, by substituting (2.20) with \( \rho = 0 \) and (2.21) into (2.2a)' and (2.2b)'.

The minimization problem defined by (2.6a) and (2.6b) is a standard quadratic programming problem for which there is an extensive literature and well known algorithms, which we shall not discuss here. See, for example [3], [11].

We will be interested in

**Lemma 1.** Suppose \( T \) is of rank \( m \). Then the solutions \( x^* \) and \( \theta^* \) to the problem

\[ \text{minimize } x'^* \sum x^* \]

subject to

\[ a \leq T \theta^* + \sum x^* \leq b \]

are given by

\[ \theta^* = (T' \sum_{-1}^{-1} T) \sum_{-1}^{-1} y^* \]  \hspace{1cm} (2.22)

\[ x^* = (\sum_{-1}^{-1} - \sum_{-1}^{-1} T(T' \sum_{-1}^{-1} T)^{-1} T') \sum_{-1}^{-1} y^* \]  \hspace{1cm} (2.23)

where

\[ y^* = (y_1^*, y_2^*, \ldots, y_n^*) , \]  \hspace{1cm} (2.24)

is any vector minimizing the quadratic form

\[ y'(I - \sum_{-1}^{-1} T(T' \sum_{-1}^{-1} T)^{-1} T') \sum_{-1}^{-1} (I - T(T' \sum_{-1}^{-1} T)^{-1} T') \sum_{-1}^{-1} y \]  \hspace{1cm} (2.25)
subject to \[ a \leq y \leq b \] \tag{2.26}

Remark: The quadratic form appearing in (2.25) is of rank \( n-m \).

Proof: For any \( \theta' = (\theta_1', \theta_2', \ldots, \theta_m') \), \( x' = (x_1', \ldots, x_n') \) define
\[ T \theta + \sum x = y. \] \tag{2.27}

Then
\[ x' \sum x = (y - T \theta)' \sum^{-1} (y - T \theta) \] \tag{2.28}
and for any given \( y \), the vector \( \hat{\theta} = \hat{\theta}(y) \) which minimizes (2.28) is given by
\[ \hat{\theta}(y) = (T \sum^{-1} T)^{-1} T \sum^{-1} y. \] \tag{2.29}

Substituting \( \hat{\theta}(y) \) for \( \theta \) in (2.28) gives the result.

Remark: A tedious but trivial calculation obtained by substituting (2.22) and (2.23) into (2.3) shows that \( \hat{f}(t) = y^* \).

Lemma 2.

\[ x^* \] is unique

Proof. Let us make Euclidean n-space a (finite-dimensional) Hilbert space with inner product
\[ < u, v > = u' \sum^{-1} v \] \tag{2.30}
where \( u, v \) are n-vectors.

Let
\[ M = I - T' (T' \sum^{-1} T)^{-1} T \sum^{-1} \).

Then (2.25) may be written
\[ < My, My > \] \tag{2.31}

Let \( E_0 \) be the \( m \) dimensional subspace spanned by the columns of \( T \) and let \( E_1 = E_0^\perp \) with the given inner product. By observing that \( MT = 0 \) and the rank of \( M \) is \( n-m \), we have that \( < My, My > = 0 \) for \( y \in E_0 \) and is strictly greater than 0 for \( y \in E_1 \). Any element \( y^* \) in the solution set may
be written \( y^* = y^{1*} + y^{0*} \) where \( y^{1*} \) and \( y^{0*} \) are the projections of \( y^* \) onto \( E_1 \) and \( E_0 \) respectively (in the given inner product). Letting \( R_{ab}^{1} \) be the projection of \( R_{ab} \) onto \( E_1 \), we have that \( y^{1*} \) is a solution to the problem: Find \( y \in R_{ab}^{1} \) to minimize \( <M y^{1*}, M y^{1*}> \). The solution to this problem is, however, well known to be unique since \( <M y^{1*}, M y^{1*}> \) is strictly greater than 0 for \( y^{1*} \neq 0 \), and \( R_{ab}^{1} \) is a closed convex set.

But a simple calculation shows that \( < \sum x^*, T \theta > = 0 \) for any \( \theta \), that is, \( \sum x^* \in E_1 \). Thus \( y^{0*} = T \theta^* \), \( y^{1*} = \sum x^* \), and \( x^* \) is uniquely determined.

It is not hard to establish when a solution \( y^* \) is unique. Define the number of sign changes of \( y^* \) with respect to \( R_{ab} \) as follows:

Let

\[
\epsilon_i = +1 \text{ if } y_i = b_i \\
\epsilon_i = -1 \text{ if } y_i = a_i \\
\epsilon_i = 0 \text{ if } a_i < y_i < b_i.
\]

Let \( \epsilon = (\epsilon_1, \epsilon_2, \ldots, \epsilon_k) \) where \( \epsilon_1 \) is the first non zero \( \epsilon_i \), \( \epsilon_2 \) is the second non zero \( \epsilon_i \), \ldots and \( \epsilon_k \) is the last non zero \( \epsilon_i \). The number of sign changes of \( y^* \) with respect to \( R_{ab} \) is then defined as the number of sign changes of the sequence in \( \epsilon \). We have

**Lemma 3.** \( y^* \) is unique iff it has more than \( m \) sign changes with respect to \( R_{ab} \).

**Proof:** By Lemma 2, if \( y^* \) and \( y^{**} \) are both solutions to the minimization problem, then

\[
y^* - y^{**} = T \theta
\]

#989
for some \( \theta \), or

\[
y^*_i - y^{**}_i = \sum_{v=1}^{m} \theta^*_v \phi_v(t_i)
\]  

(2.32)

\( y^* \) is thus unique iff there does not exist a polynomial

\[
P_m(t) = \sum_{v=1}^{m} \theta^*_v \frac{t^{v-1}}{(v-1)!}
\]  

(2.33)

of degree at most \( m-1 \) such that

\[
a_i \leq y^*_i - P_m(t_i) \leq b_i.
\]  

(2.34)

But, for (2.34) to hold, the sign of \( P_m(t_i) \) must equal the sign of \( \epsilon_i \), whenever \( \epsilon_i \neq 0 \), so that the lemma is equivalent to the existence of a polynomial of degree \( m-1 \) with at least the required number of sign changes.

3. A Statistical Model Leading to the Algorithm of Section 2

To explore the statistical implications of algorithms of the type of Section 2 we make the following definition. Let \( Y_0 \) and \( Y = (Y_1, Y_2, \ldots, Y_n) \) be a set of \( n+1 \) random variables with joint density

\[
P_{\theta}(y_0, y), \quad y' = (y_1, y_2, \ldots, y_n)
\]

depending on an (unknown) parameter \( \theta' = (\theta_1, \theta_2, \ldots, \theta_m) \), \( \theta \in \Theta \). Let \( \mathcal{R} \) be a region in Euclidean \( n \)-space. Then we say that a maximum likelihood estimate \( \hat{Y}_0 \) for the random variable \( Y_0 \), given data \( y \in \mathcal{R} \), is any number \( \hat{Y}_0 \) for which

\[
\sup_{\theta \in \Theta, \ y \in \mathcal{R}} P_{\theta}(y_0, y)
\]  

(3.1)

is attained.

This definition leads to the following

Lemma 4.

Let \( Y \) be an \( n \) dimensional normally distributed random vector with mean vector \( T \theta \) and covariance matrix \( \sum \) and let \( Y_0 \) be a normally
distributed random variable with mean $\sum_{v=1}^{m} u_v \theta_v$ and variance $\sigma_{00}$, where 
$\{u_v\}_{v=1}^{m}$ are any fixed $m$ real numbers, $\sigma_{00}$ is given, and 
\[
\text{cov}(Y_i, Y_0) = \sigma_{i0}, \quad i = 1, 2, \ldots, n \tag{3.2}
\]
where $\{\sigma_{i0}\}_{i=1}^{n}$ are given.

Then a maximum likelihood estimate $\hat{Y}_0$ for $Y_0$, given data $Y \in \mathbb{R}$ is given by 
\[
\hat{Y}_0 = \sum_{v=1}^{m} u_v \theta_v^* + \sum_{i=1}^{n} \sigma_{i0} x_i^* \tag{3.3}
\]
where $\theta^* = (\theta_1^*, \theta_2^*, \ldots, \theta_m^*)$ and $x^* = (x_1^*, x_2^*, \ldots, x_n^*)$ are any vectors minimizing 
\[
x^* \overset{\text{min}}{\sum x^*} \tag{3.4}
\]
subject to 
\[
T \theta^* + \sum x^* \in \mathbb{R} \tag{3.5}
\]

Proof. Let $s^* = (\sigma_{10}, \sigma_{20}, \ldots, \sigma_{n0})$, $u^* = (u_1^* u_2^* \ldots u_m^*)$. Then 
\[
P_\theta(y_0, y) = P_\theta(y_0 \mid y) P_\theta(y) = [(2\pi)^{n+1} |\sigma_{00} - s^* \sum^{-1} s^*| |\sum|^{-1/2} \cdot 
\exp - \frac{1}{2} [(\sigma_{00} - s^* \sum^{-1} s^*)^{-1} \cdot (y_0 - u^* \theta + s^* \sum^{-1} (y - T\theta))] 
\cdot \exp - \frac{1}{2} [(y - T\theta)^* \sum^{-1} (y - T\theta)] \tag{3.6}
\]

Lemma 4 now follows directly from Lemma 1, since (3.6) is evidently minimized by setting $\theta = \hat{\theta}(y)$ as given by (2.29), maximizing (2.25) subject to $y \in \mathbb{R}$, and letting 
\[
\hat{Y}_0 = u^* \theta^* + s^* \sum^{-1} (y^* - T\theta^*) = u^* \theta^* + s^* x^* \tag{3.7}
\]
where $\theta^*$ is given by (2.22) and $x^* = \sum^{-1} (y^* - T\theta^*)$. 

\#989
We are now going to define a stochastic process for which the algorithm of Section 2 provides maximum likelihood estimates for quantized data.

Let

\[ Y(t) = \sum_{\nu=1}^{m} \theta_{\nu} \phi_{\nu}(t) + X(t) \]  

(3.8)

where \( \phi_{\nu}(t) \) is given by (2.4), \( \theta = (\theta_1, \ldots, \theta_m) \) is a parameter and \( X(t) \) is the m-fold integrated Weiner process, that is

\[ X(t) = \int_{0}^{t} dt_{m-1} \int_{0}^{t_{m-1}} dt_{m-2} \cdots \int_{0}^{t_{2}} dt_{1} \int_{0}^{t_{1}} dW(t) \]  

(3.9)

where \( W(t), \ 0 \leq t \leq 1 \) is the Weiner process, \( W(0) = 0, \ E W(t) = 0, \ E W(s) W(t) = \min(s, t) \). It is natural to adopt the formalism

\[ \frac{d^m}{dt^m} Y(t) = \frac{dW(t)}{dt} \]  

(3.10)

since \( \frac{d^m}{dt^m} \phi_{\nu}(t) = 0 \), although \( Y(t) \), as is well known, has only \( m-1 \) quadratic mean derivatives. \( \frac{dW(t)}{dt} \) is commonly called "white noise". By formal integration by parts, it may be verified that \( X(t) \) has the representation

\[ X(t) = \int_{0}^{1} G_m(t, u) \ dW(u) \]  

(3.11)

where \( G_m(t, u) \) is given by (2.5). (See also Shepp [10].) We have the following

**Theorem 2.** Let \( Y(t) \) be the stochastic process defined by (3.8).

Let \( Y_0 = Y(t_0), \ t_0 \in [0, 1] \) and let

\[ Y = (Y(t_1), \ldots, Y(t_n)) \]  

(3.12)
Then any maximum likelihood estimate \( \hat{y}_0 \) for \( Y_0 \) based on the data \( Y \in \mathbb{R}_{ab} \) is given by

\[
\hat{y}_0 = \hat{f}(t_0)
\]  

(3.13)

where \( \hat{f}(t) \) is any function minimizing

\[
\int_0^1 (\hat{f}(m)(t))^2 \, dt
\]

subject to

\[
a_i \leq \hat{f}(t) \leq b_i
\]

(3.15)

Proof. The proof follows directly upon making the correct associations, as follows:

From (3.8) and (3.11)

\[
E \, Y(t) = \sum_{\nu=1}^m \theta_{\nu} \phi_{\nu}(t)
\]

(3.16)

\[
\text{cov}(Y(s)Y(t)) = E X(s)X(t) = \int_0^1 G_m(s, u) G_m(t, u) \, du = K(s, t)
\]

(3.17)

Hence \( Y \) as given by (3.12) satisfies

\[
Y \sim h(T \Theta, \sum)
\]

(3.18)

and

\[
\text{cov}(Y(t), Y_0) = E X(t)X'(t_0) = \int_0^1 \frac{d}{ds} G_m(t, u) G_m(s, u) \bigg|_{s=t_0} \, du
\]

(3.19)

By setting \( u_{\nu} = \frac{d}{dt} \phi_{\nu}(t_0) \), \( \sigma_{10} = \frac{d}{dt} K(t_0, t) \bigg|_{t=t_0} \) in (3.3) and comparing

(3.3) with (2.3) gives the theorem.

#989
Remarks. In choosing this model, the experimenter is, in an intuitive sense, favoring the assumption that the signal is in the null space of the linear differential operator \( L_m = \frac{d^m}{dt^m} \) and his smoothness criterion \( \int_0^1 [L_m f(t)]^2 dt \) is a measure of the "distance" of a function from this null space. The model (3.8) is a rather particular one, though intuitively appealing for many applications. However Theorems 1 and 2 can be generalized to a vast array of smoothing criteria, which we will do in Section 4.

We now remark that we have set \( Y_0 \) as \( Y'(t_0) \) in Theorem 2, but there is in fact nothing sacred about derivatives. Let \( N \) be a continuous linear functional defined on \( H_m \). For this it is sufficient that

\[
N(s) G_m(s, u) = \eta(u) \tag{3.20}
\]

be defined everywhere and \( \eta(u) \in L_2 [0,1] \); and \( N \phi \) be defined for each \( \nu \). The symbol \( N(s) \) means \( N \) is applied to \( G_m(s, u) \) considered as a function of \( s \).

Let \( x_0 \) be the random variable

\[
x_0 = \int_0^1 \eta(s) \, dW(s) \tag{3.21}
\]

where \( W(s) \) is the same Weiner process appearing in the representation (3.11), and let

\[
Y_0 = (NY(t)) = N \sum_{\nu=1}^{m} \theta_{\nu} \phi_{\nu}(t) + \int_0^1 \eta(s) dW(s) \tag{3.22}
\]

Then

\[
EY_0 = \sum_{\nu=0}^{m} \theta_{\nu} (N \phi_{\nu}) \tag{3.23}
\]

\[
cov(Y(t)Y_0) = \int_0^1 G_m(t, u) \eta(u) du = N(s) K(t, s) \tag{3.24}
\]
and a step by step mimic of the proofs of Theorems 1 and 2 shows that

\[ \hat{N} \hat{f} \]

is a maximum likelihood estimate for \( Y_0 \) based on data \( Y \in \mathcal{R}_{ab} \).

Random variables of the form

\[ Y_0 = Y^{(v)}(t_0), \quad v = 1, 2 \ldots m-1, \ t_0 \text{ fixed, } t_0 \in [0, 1] \quad (3.25) \]

and

\[ Y_0 = \int_0^1 w(t) Y(t) \, dt \quad (3.26) \]

where \( w(t) \) is a piecewise continuous weight function, are included. By Lemma 2, we have that if \( N \) has the property that \( N \phi_v = 0, \ v = 1, 2, \ldots, m \), then the maximum likelihood estimate for \( Y_0 = Y_0(N) \) will be unique.

4. Generalizations to Other Models

The model \((3.8)\) is a fairly restricted one. However Theorems 1 and 2 have broad generalizations, essentially to models of the form \((3.8)\) where \( X(t) \) is only supposed to be a zero mean Gaussian process that is continuous in quadratic mean. In this section we state the generalizations, and point the potential user to concrete examples.

Let now \( H(s, t) \) be any continuous symmetric positive definite kernel on \([0, 1] \times [0, 1]\). By the Moore-Aronszajn Theorem (see [2]), to every such kernel there corresponds a unique Hilbert space \( \mathcal{H} \) of functions on \([0, 1]\) with the following properties

1) \( H_t(s) \in \mathcal{H}, \quad \forall t \in [0, 1] \quad (4.1) \)

where \( H_t = H_t(s) \) is that function of \( s \) given by \( H_t(s) = H(s, t) \)

2) For every \( f \in \mathcal{H}, \) we have

\[ \langle f, H_t \rangle = f(t), \quad 0 \leq t \leq 1 \quad (4.2) \]
$\mathcal{H}_m$ of Section 2 with inner product given by (2.7) and kernel $H(s, t)$ given by (2.10) is such a space. Equations (2.9) and (2.14) constitute a verification of property 2). Property 2) implies that

$$\langle H_t, H_s \rangle = H(s, t) \tag{4.3}$$

Hence, $H(s, t)$ is called a reproducing kernel, and $\mathcal{H}$ a reproducing kernel Hilbert space. By 2) and the Riesz representation theorem, reproducing kernel spaces of functions defined on $[0, 1]$ are characterized by the fact that the linear functional $f(t_0)$ is continuous for every $t_0 \in [0, 1]$. Let now $\{\phi_v(t)\}_{v=1}^m$ be $m < n$ orthonormal functions in $\mathcal{H}$, let $K(s, t)$ be defined by

$$H(s, t) = \sum_{v=1}^m \phi_v(s) \phi_v(t) + K(s, t) \tag{4.4}$$

Let $Q_0$ be the subspace of $\mathcal{H}$ spanned by $\{\phi_v(t)\}_{v=1}^m$, let $Q_1 = Q_0 ^\perp$ and let $P_1$ be the projection operator onto $Q_1$. $K(s, t)$ is non-negative definite, and can be shown to be the reproducing kernel for $Q_1$.

**Theorem 3.**

Let the $n \times m$ matrix $T$ with $i$th entry $\phi_v(t_i)$ be of rank $m$, let $k_i(t) = K(t_i, t)$ and let the $n \times n$ matrix $\sum$ with $ij$th entry

$$\langle k_i, k_j \rangle = K(t_i, t_j) \tag{4.5}$$

be of rank $n$. Then any solution to the problem: Find $\hat{f} \in \mathcal{H}$ to minimize

$$\langle P_1 \hat{f}, P_1 \hat{f} \rangle \tag{4.6}$$

subject to

$$a_i \leq \hat{f}(t_i) \leq b_i, \quad i = 1, 2 \ldots n \tag{4.7}$$

is

$$\hat{f}(t) = \sum_{v=1}^m \phi_v(t) \theta_v^* + \sum_{i=1}^n K(t_i, t) x_i^* \tag{4.8}$$
where $\theta^* = (\theta_1^*, \theta_2^*, \ldots, \theta_m^*)$ and $x^* = (x_1^*, x_2^*, \ldots, x_n^*)$ are any solutions to the problem

$$\text{minimize } x^* \sum x^*$$

subject to

$$a \leq T \theta^* + \sum x^* \leq b,$$

$$a' = (a_1, a_2, \ldots, a_n), \quad b' = (b_1, b_2, \ldots, b_n).$$

The proof mimics that of Theorem 1 step by step and is omitted.

Now let $X(t)$ be the zero mean Gaussian stochastic processes defined by

$$E X(s) X(t) = K(s, t)$$

where $K(s, t)$ is given by (4.4). We may define a Hilbert space $\HH_X$ of random variables as follows (see [8]). All finite linear combinations of the form

$$\rho = \sum_t c_t X(t)$$

form a linear manifold. $\HH_X$ is the closure of this linear manifold with the inner product

$$<\rho_1, \rho_2> = E \rho_1 \rho_2$$

$\HH_X$ is isomorphic to $C_1$ under the isomorphism

$$X(t) \leftrightarrow K_t(s)$$

where, for each fixed $t$, $K_t(s)$ is that function of $s$, given, for fixed $t$, by $K_t(s) = K(s, t)$. We have

$$<X(s_1), X(s_2)> = E X(s_1) X(s_2) = K(s_1, s_2) = <K_{s_1}, K_{s_2}>$$

Let $\rho$ be any (fixed) random variable in $\HH_X$. It is well known [4][8] that if

$$E \rho X(t) = \eta(t)$$

$\#989$
then \( \eta(t) \in Q_1 \) and we have the correspondence

\[
\rho \leftrightarrow \eta(t) \quad \tag{4.15}
\]

under the isomorphism (4.12). Let \( N \) be the continuous linear functional defined on \( \mathcal{H} \) by

\[
N f = < \eta + \phi_0, f > \quad \tag{4.16}
\]

where \( \eta = \eta(t) \) of (4.14) and \( \phi_0 \) is a fixed element in \( Q_0 \). By the reproducing property (4.3) and the fact that \( \phi_0 \in Q_0 \Rightarrow < \phi_0, K_t > = 0 \), all \( t \), we have

\[
\eta(t) = N(s) K(s, t). \quad \tag{4.17}
\]

It is now legitimate to define the random variable \( Y_0 = (N Y(t)) \) by

\[
(N Y(t)) = N \left( \sum_{\nu=1}^{m} \phi_\nu(t) \theta_\nu \right) + (N X(t)) \quad \tag{4.18}
\]

where

\[
(N X(t)) = \rho. \quad \tag{4.19}
\]

We have

\[
E \ Y_0 = N \left( \sum_{\nu=1}^{m} \phi_\nu(t) \theta_\nu \right) \quad \tag{4.20}
\]

and

\[
\text{cov} (Y(t), Y_0) = E \rho X(t) = \eta(t) = N(s) K(s, t). \quad \tag{4.21}
\]

We have

**Theorem 4.**

Any maximum likelihood estimate for the random variable \( Y_0 \) based on data \( Y = (Y(t_1), \ldots, Y(t_n)) \in \mathcal{R}_{ab} \) is given by

\[
N \hat{f}(t) \quad \tag{4.22}
\]

where \( \hat{f}(t) \) is given by (4.8).

**Proof.** With the use of (4.20) and (4.21) the proof mimics that of Theorem 2, and is omitted.
Remarks and Examples. We have assumed that the \( n \times m \)
regression matrix with \( i \) \( \nu \)th entry given by \( \phi_{\nu}(t_i) \) is always of rank \( m \), (with
\( n \geq m \)). Any system \( \{ \phi_{\nu}(t) \}_{\nu=1}^{m} \) which has this property for any distinct choice
\( t_1 < t_2 \ldots t_n \) is said to constitute a Tchebychev system. Such systems have
come in for a substantial amount of study recently. ([5][6]).

Let \( \Gamma \) be any fixed non-singular \( m \times m \) matrix with \( \mu \nu \)th entry \( \gamma_{\mu \nu} \)
and let the linear functionals \( M_{\mu}, \mu = 1, 2 \ldots m \) be defined by
\[
M_{\mu} f = \sum_{\nu=1}^{m} \gamma_{\mu \nu} f^{(\nu-1)}(0), \quad \mu = 1, 2 \ldots m .
\] (4.23)

Let \( L_m \) be an \( m \)th order linear differential operator on \([0, 1]\) whose null space
is spanned by some Tchebychev system \( \{ \tilde{\phi}_{\nu}(s) \}_{\nu=1}^{m} \), and let \( G_m(t, u) \) be the
Green's function for the equation
\[
L_m f = p
\] (4.24)
with boundary conditions \( M_{\nu} f = 0, \nu = 1, 2, \ldots m \). We remind the reader that
the Green's function solves the equation (4.24), as
\[
f(t) = \int_{0}^{1} G_m(t, u) p(u) \, du
\] (4.25)
with \( M_{\nu} f = 0, \nu = 1, 2 \ldots m \). For \( \Gamma \) non-singular, \( G_m(t, u) \) does not
depend on \( \Gamma \). For an introduction to Green's functions, see [7]. In Section 2
we had \( L_m = \frac{d^m}{dt^m}, \quad \{ \tilde{\phi}_{\nu}(t) \}_{\nu=1}^{m} \}
\[
= \{ \frac{t^{\nu-1}}{(\nu-1)!} \}_{\nu=1}^{m} \quad \text{and} \quad G_m(t, u) = \frac{(t-u)^{m-1}}{(m-1)!} .
\]
A straightforward procedure for generating examples may be found in Karlin [5],
p. 276, and another example is given at the end of this section. We may
always construct a set \( \{ \phi_{\nu} \}_{\nu=1}^{m} \) from \( \{ \tilde{\phi}_{\nu} \}_{\nu=1}^{m} \) satisfying conditions
\[
M_{\nu} \phi_{\mu} = 1 \quad \nu = \mu
\]
\[
= 0 \quad \nu \neq \mu
\]
For examples for Theorems 3 and 4, we let \( \mathcal{Q}_0 \) be the \( m \) dimensional Euclidean space spanned by the \( \{\phi_\nu\} \) above. They are an orthonormal set if we endow \( \mathcal{Q}_0 \) with the inner product
\[
<g,h>_{\mathcal{Q}_0} = \sum_{\nu=1}^{m} (M_\nu g)(M_\nu h).
\]

Take \( \mathcal{Q}_1 \) as the Hilbert space of functions
\[
\{f : f^{(m-1)} \text{ absolutely continuous, } L_m f \in L_2[0,1], M_\nu f = 0, \nu = 1, 2, \ldots, m\}.
\]
The inner product in \( \mathcal{Q}_1 \) is
\[
<g,h>_{\mathcal{Q}_1} = \int_0^1 (L_m g(t))(L_m h(t)) \, dt \quad g, h \in \mathcal{Q}_1
\]
and the "smoothness criterion" is given by
\[
<\Psi_1 g, \Psi_1 g> = \int_0^1 (L_m g(t))^2 \, dt \quad g \in \Psi.
\]
The choice of \( \Gamma \), which is equivalent to a choice of inner product on \( \Psi_0 \), is irrelevant to the calculation of \( \hat{f} \). To see this, start with any non-singular \( m \times m \) matrix \( A \) with \( \mu \nu \)th entry \( a_{\mu \nu} \), and the representation for \( \hat{f}(t) \) based on (2.3), (2.22), (2.23) and (2.25). Replace \( \{\phi_\nu(t)\}_{\nu=1}^{m} \) by \( \{\Phi_\nu(t)\}_{\nu=1}^{m} \) defined by
\[
\Phi_\nu(t) = \sum_{\mu=1}^{m} a_{\nu \mu} \phi_\mu(t)
\]
and replace the matrix \( T' \) with \( v \)th entry \( \phi_\nu(t) \) by \( \tilde{T'} = A T' \) where the \( v \)th entry of \( \tilde{T'} \) is given by \( \tilde{\phi}_v(t) \). Note that (2.23), (2.25) and
\[
\sum_{\nu=1}^{m} \phi_\nu(t) \theta^*_\nu, \quad \text{where } \theta^*_\nu \text{ is given by (2.22) are invariant under this transformation}
\]
and hence so is \( \hat{f}(t) \).
This provides us with another relatively simple example. Let

$\alpha_1, \alpha_2, \ldots, \alpha_m$ be $m$ distinct real positive numbers. The functions

$$\phi_v(t) = e^{-\alpha_v t}, \quad v = 1, 2, \ldots, m$$

span the null space of the linear differential operator $L_m$ defined by

$$L_m f = \prod_{\nu=1}^{m} (D + \alpha_\nu) f$$

where

$$(D + \alpha_\nu) f = f' + \alpha_\nu f$$

and the appropriate Green's function on $[0, 1]$ is

$$G_m(s, t) = \sum_{\nu=1}^{m} \prod_{\nu \neq \nu'} (\alpha_\nu - \alpha_\nu')^{-1} e^{-\alpha_\nu (s-t)}, \quad s > t$$

$$= 0 \quad \text{otherwise.}$$
REFERENCES


