The Annals of Statistics
1983, Vol. 11, No. 3, 1011–1017

ON THE SMOOTHNESS PROPERTIES OF THE BEST LINEAR UNBIASED ESTIMATE OF A STOCHASTIC PROCESS OBSERVED WITH NOISE[1,2]

BY ROBERT KOHN AND CRAIG F. ANSLEY

University of Chicago

Suppose $x(t)$ is a vector stochastic process generated by a first order differential equation and $f(t)$ is a linear combination of the elements of $x(t)$. Functionals of $x(t)$ are observed with noise. We obtain the smoothness properties of the best linear unbiased estimate of $f(t)$, and those of its derivatives that exist. In addition we obtain the smoothness properties of their mean squared errors.

1. Introduction. Let $f(t)$ be a scalar stochastic process described by

$$f(t) = c'x(t) \tag{1}$$

$$dx(t)/dt = A(t)x(t) + b(t)\, dW(t)/dt. \tag{2}$$

(2) is a vector stochastic differential equation driven by a scalar continuous time uncorrelated increment process.

Let $\lambda_i(x)$, $i = 1, \cdots, N$, be given by

$$\lambda_i(x) = c_i'x(t_i). \tag{3}$$

As discussed later in this section, and also in Section 4, $\lambda_i(x)$ will often be a linear functional of $f(\cdot)$. Suppose that we have N discrete observations

$$y_i = \lambda_i(x) + e_i \tag{4}$$

with e_i a sequence of observational errors.

The main purpose of this paper is to obtain the smoothness properties of the best linear unbiased estimate of $f(t)$ and its mean squared error, given the observations y_1 to y_N.

The results of this paper are motivated by work connecting smoothing splines and best linear unbiased estimates of a stochastic process. See, for example, Kimeldorf and Wahba (1970a, 1970b, 1971), Wahba (1978), Weinert and Kailath (1974), Weinert and Sidhu (1978), and Weinert, Byrd and Sidhu (1980). In fact the smoothness properties of smoothing and interpolating Lg splines can be deduced from Corollary 2 (to Theorem 1) of Section 4. See Remark 3 of Section 4 for details.

The paper is structured as follows. The assumptions are stated in Section 2, the main results are obtained in Section 3, and these results are then applied to a scalar stochastic differential equation in Section 4.

When the functionals $c_i'x(t_i) = f(t_i)$ for $i = 1$ to N, the y_i are discrete observations of a continuous time process observed with noise. Such models are often used in engineering and statistics. For example, the scalar stochastic differential equation of Section 4 is often used to model a continuous time series. See, for example, Hannan (1970), page 405.

For some further examples consider the scalar stochastic equation

$$df(t)/dt + \rho f(t) = dW(t)/dt. \tag{5}$$

If in (5) we take $W(\cdot)$ as a Wiener process, then $f(t)$ is the Ornstein-Uhlenbeck process;

Received March 1982; revised December 1982.

[1] Please do not quote without permission.

[2] First Draft May 1981.

AMS 1980 *subject classifications*. Primary, 60635.

Key words and phrases. Smoothness properties, stochastic process, best linear unbiased estimate.

Breiman (1968), page 349.

If in (5),

$$W(t) = \textstyle\sum_{i=1}^{N(t)} Z_i, \quad N(t) > 0$$
$$= 0, \quad N(t) = 0$$

where $N(t)$, $t \geq t_1$, is a Poisson process and the Z_i are independent random variables, then $W(t)$ is a compound Poisson process (Parzen, 1968, page 128), and (5) becomes a model of shot noise (Parzen, 1968, page 128).

2. Assumptions.

ASSUMPTION 1. (i) In this paper t lies in the interval $[T_0, T_i]$.

(ii) $T_0 < t_1 \leq t_2 \leq \cdots \leq t_N < T_1$.

ASSUMPTION 2. (i) $W(t)$ is a scalar uncorrelated increment process with $dE[W(t)]/dt = m(t)$ and $d\,\mathrm{var}[W(t)]/dt = r(t)$.

(ii) β is a positive integer greater than or equal to 1. $r(t)$ and $m(t)$ have $\beta - 1$ continuous derivatives. If $\beta = 1$, then $r(t)$ and $m(t)$ are continuous.

ASSUMPTION 3. (i) $A(t)$ is an $n \times n$ matrix function of t with continuous elements. $b(t)$ is an $m \times 1$ vector whose elements have $\max(1, \beta - 2)$ continuous derivatives.

(ii) c and c_i, $i = 1, \cdots, N$, are $n \times 1$ constant vectors. For $j \geq 0$, define $\zeta_j(t)$ and $\eta_j(t)$ by

$$\zeta_0(t) = c, \quad \eta_0(t) = b(t)$$
$$\zeta_{j+1}(t) = d\zeta_j(t)/dt + A(t)'\zeta_j(t)$$
$$\eta_{j+1}(t) = d\eta_j(t)/dt - A(t)\eta_j(t).$$

ASSUMPTION 4. (i) With α a positive integer and β defined as above, $\zeta_j(t)$ is differentiable for $j = 1$ to $\alpha + \beta - 3$ and continuous for $j = 1$ to $\alpha + \beta - 2$. $\eta_j(t)$ is differentiable for $j = 1$ to $\beta - 2$ and continuous for $j = \beta - 2$.

(ii)

$$c_i'\eta_j(t_i) = 0 \quad \text{for all } i \text{ and } j = 0 \text{ to } \beta - 2$$
$$\zeta_j'(t)b(t) = 0 \quad \text{for all } t \text{ and } j = 0 \text{ to } \alpha - 2$$
$$\zeta_j'(t)\eta_1(t) = 0 \quad \text{for } j = 0 \text{ to } \alpha - 3.$$

ASSUMPTION 5. (i) The e_i, $i = 1$ to n, have zero mean and finite variance.

(ii) e_i is uncorrelated with $W(t)$ for $i = 1$ to N and $T_0 \leq t \leq T_1$.

(iii) $W(t) - W(T_0)$ is uncorrelated with $x(T_0)$ for all $T_0 \leq t \leq T_1$.

3. Smoothness results. Let $F(t)$ be a $n \times n$ fundamental matrix solution of

$$dF(t)/dt = A(t)F(t)$$

with the columns of $F(t)$ linearly independent; Coddington and Levinson (1955), page 69. Put

$$D(t, s) = F(t)F(s)^{-1}.$$

Then,

$$\partial D(t, s)/\partial t = A(t)D(t, s) \tag{6}$$

$$\partial D(t, s)/\partial s = -D(t, s)A(s). \tag{7}$$

Solving for $x(t)$ from Equation (2) we obtain

$$x(t) = D(t, T_0)x(T_0) + \int_{T_0}^{t} D(t, s)b(s)W(ds); \tag{8}$$

see Coddington and Levinson (1955), page 74.

Put $y^{(N)} = (y_1, \cdots, y_N)'$, and let $f(t|N)$ and $x(t|N)$ be the best linear unbiased estimates of $f(t)$ and $x(t)$, respectively, given $y^{(N)}$.

Put

$$\textstyle\sum (t|N) = E\{(x(t) - x(t|N))(x(t) - x(t|N))'\},$$
$$\textstyle\sum_f (t|N) = E\{(f(t) - f(t|N))^2\}.$$

Throughout this section we will assume that Assumptions 1 to 5 hold. Then,

THEOREM 1. *Subject to Assumptions* 1 *to* 4, $f(t|N)$ *and* $\sum_f (t|N)$ *have* $\alpha + \beta - 2$ *continuous derivatives in* (T_0, T_1).

To prove Theorem 1 we need to establish a number of preliminary results. Purely for convenience, we will assume that all random variables are jointly Gaussian. This enables us to avoid cumbersome notation by working with conditional expectations rather than having to always refer to best linear unbiased estimates.

LEMMA 1. (i) *For* $t > T_0$,

$$\operatorname{cov}\{f(t), y_i\} = c'D(t, T_0)\operatorname{cov}\{x(T_0), y_i\} + c'Q_i(t)c_i; \tag{9}$$

where

$$Q_i(t) = \int_{T_0}^{t\wedge t_i} D(t, s)b(s)b(s)'D(t_i, s)'r(s)\, ds, \tag{10}$$

and $t \wedge t_i = \min(t, t_i)$.

(ii) *Under Assumptions* 1 *to* 5, $\operatorname{cov}(f(t), y_i)$ *has* $\alpha + \beta - 2$ *continuous derivatives.*

PROOF. From (8) and our assumptions it is easy to establish (i).

(ii) For $t < t_i$ and $j = 1$ to $\alpha - 1$ we can establish that

$$\begin{aligned} d^j\{c'Q_i(t)c_i\}/dt^j &= \zeta_j(t)'Q_i(t)c_i + \zeta_{j-1}(t)'b(t)b(t)'D(t_i, t)'r(t) \\ &= \zeta_j(t)'Q_i(t)c_i, \end{aligned}$$

because $\zeta_j'(t)b(t) = 0$ for $j = 0$ to $\alpha - 2$. We can also establish for $j = \alpha$ to $\alpha + \beta - 2$ that

$$\frac{d^j\{c'Q_i(t)c_i\}}{dt^j} = \zeta_j(t)'Q_i(t)c_i + \textstyle\sum_{\ell=0}^{j-\alpha} \dfrac{d^\ell}{dt^\ell}\{\zeta_{j-1-\ell}'(t)b(t)c_i'D(t_i, t)b(t)r(t)\}. \tag{11}$$

Furthermore,

$$\frac{d^\ell}{dt^\ell}\{\zeta_{j-1-\ell}'(t)b(t)c_i'D(t_i, t)b(t)r(t)\}$$
$$= \textstyle\sum_{u+k+v=\ell} \dfrac{\ell!}{u!k!v!}\left\{\dfrac{d^u}{dt^u}(\zeta_{j-1-\ell}'b)\right\}\left\{\dfrac{d^k}{dt^k}(c_i'Db)\right\} \times \left\{\dfrac{d^v r}{dt^v}\right\}.$$

Now we can check that

$$d^k[c_i'D(t_i, t)b(t)]/dt^k = c_i'D(t_i, t)\eta_k(t). \tag{12}$$

Because $c_i'\eta_k(t_i) = 0$ for $k = 0$ to $\beta - 2$ and $D(t_i, t_i) = I_n$, letting $t \uparrow t_i$ makes the right side of (12), and hence the left, tend to zero (for $k = 0$ to $\beta - 2$). It follows that the second term on the right side of (11) tends to zero as $t \uparrow t_i$ for $j = \alpha$ to $\alpha + \beta - 2$. Because $Q_i(t)$ and $\zeta_j(t)$ are continuous it follows that for $t \uparrow t_i$,

$$d^j(c'Q_i(t)c_i)/dt^j \to \zeta_j(t_i)'Q_i(t_i)c_i.$$

For $t > t_i$, it is not hard to show that for $j = 0$ to $\alpha + \beta - 2$,

$$d^j(c'Q_i(t)c_i)/dt^j = \zeta_j(t)'Q_i(t)c_i.$$

Hence $c'Q_i(t)c_i$ has $\alpha + \beta - 2$ continuous derivatives at $t = t_i$, and from the above derivation also for all other t.

Because

$$d^jc'D(t, T_0)/dt^j = \zeta_j'(t)D(t, T_0)$$

for $j = 1, \cdots, \alpha + \beta - 2$, $c'D(t, T_0)$ has $\alpha + \beta - 2$ continuous derivatives. The conclusion of Lemma 1(ii) now follows. □

LEMMA 2. *$E\{f(t)\}$ and $\operatorname{var}\{f(t)\}$ have $\alpha + \beta - 2$ continuous derivatives.*

PROOF. First consider $E\{f(t)\}$. Put

$$h(t) = D(t, T_0)E\{x(T_0)\} + \int_{T_0}^{t} D(t, s)b(s)m(s)\, ds.$$

Then, $E\{f(t)\} = c'h(t)$. As in the proof of Lemma 1, for $j = 1$ to $\alpha - 1$,

$$d^jE\{f(t)\}/dt^j = \zeta_j(t)'h(t) + \zeta_{j-1}'(t)b(t)m(t) = \zeta_j(t)'h(t).$$

For $j = \alpha$ to $\alpha + \beta - 2$

$$\frac{d^jE\{f(t)\}}{dt^j} = \zeta_j(t)'h(t) + \textstyle\sum_{\ell=0}^{j-\alpha} \frac{d^\ell}{dt^\ell}\{\zeta_{j-1-\ell}'bm\}.$$

It now follows from our assumptions that $E(f)$ has $\alpha + \beta - 2$ continuous derivatives.

The proof for $\operatorname{var}\{f(t)\}$ is similar and we omit it. □

PROOF OF THEOREM 1. We know that

$$f(t\,|\,N) = E\{f(t)\} + \operatorname{cov}\{f(t), y^{(N)}\}V(y^{(N)})^{-1}\{y^{(N)} - E(y^{(N)})\} \tag{13}$$

$$\textstyle\sum_f(t\,|\,N) = V_f(t) - \operatorname{cov}\{f(t), y^{(N)}\}V(y^{(N)})^{-1}\operatorname{cov}\{f(t), y^{(N)}\}, \tag{14}$$

where $V_f(t) = \operatorname{var}\{f(t)\}$. The proof now follows from Lemmas 1 and 2. □

COROLLARY 1.

$$f(t\,|\,N) = c'x(t\,|\,N)$$

and for $j = 1$ to $\alpha - 1$,

$$d^jf(t\,|\,N)/dt^j = \zeta_j'(t)x(t\,|\,N)$$

$$\frac{d^j \sum_f(t\,|\,N)}{dt^j} = \textstyle\sum_{i=0}^{j} \binom{j}{i} \zeta_i'(t) \sum (t\,|\,N)\zeta_{j-i}(t). \tag{15}$$

We now look at the smoothness of $f(t)$ itself and the best linear unbiased predictor of $f^{(j)}(t)$ given $y^{(N)}$.

THEOREM 2. *For* $\alpha \geq 2$,

(i) $f^{(j)}(t) = \zeta_j'(t)x(t)$ *for* $j = 0$ *to* $\alpha - 2$.

(ii) *The best linear predictor of* $f^{(j)}(t)$ *given* $y^{(N)}$ *is* $\zeta_j(t)'x(t \mid N)$ *with mean squared error*

$$\zeta_j(t)' \textstyle\sum (t \mid N)\zeta_j(t). \tag{16}$$

(iii) *If* $W(t)$ *is continuous, then* $f^{(\alpha-1)}(t)$ *is also continuous.*

PROOF. Using integration by parts, we deduce from (8) that

$$x(t) = g(t) + b(t)W(t) \tag{17}$$

where

$$g(t) = D(t, T_0)\{x(T_0) - b(T_0)W(T_0)\} + \int_{T_0}^{t} D(t, s)\eta_1(s)W(s)\, ds.$$

Hence $f(t) = c'g(t)$. Now

$$\frac{dh(t)}{dt} = A(t)g(t) + \eta_1(t)W(t)$$

so that we can check that $f^{(j)}(t) = \zeta_j'(t)g(t)$, for $j = 1$ to $\alpha - 2$, because $\zeta_{j-1}'(t)\eta_1(t) = 0$ for $j = 1$ to $\alpha - 2$. Hence

$$f^{(j)}(t) = \zeta_j'(t)x(t)$$

because $\zeta_j(t)'\{x(t) - g(t)\} = 0$ from (17). (i) and (ii) now follow.

To obtain (iii) note that

$$f^{(\alpha-1)}(t) = \zeta_{\alpha-1}'(t)g(t) + \zeta_{\alpha-2}'(t)\eta_1(t)W(t)$$

so that if $W(t)$ is continuous so is $f^{(\alpha-1)}(t)$. □

REMARK 1. (i) Given $x(t \mid N)$ and $\sum (t \mid N)$, (15) and (16) give us computing formulae for $d^j f(t \mid N)/dt^j$, and its mean squared error.

4. Application to a scalar differential equation. In this section we apply our theory to the stochastic scalar differential equation

$$Lf(t) = W(dt), \tag{18}$$

where

$$L = \frac{d^n}{dt^n} + a_{n-1}(t)\frac{d^{n-1}}{dt^{n-1}} + \cdots + a_0(t),$$

with $a_j(t)$ having j continuous derivatives and $a_0(t)$ continuous.

For $i = 1$ to N, let

$$\lambda_i(x) = \lambda_i(f) = \textstyle\sum_{j=0}^{n-\beta} \gamma_{ij} f^{(j)}(t_i) \tag{19}$$

where $1 \leq \beta \leq n$. Then as in Weinert, Byrd and Sidhu (1980), we can rewrite (18) and (19) in the form of equations (1) to (4), with $x_j(t) = f^{(j-1)}(t)$, $j = 1$ to n,

$$c = i_1, \quad b = i_n, \quad c_i' = (\gamma_{i0}, \cdots, \gamma_{i,n-\beta}, 0, \cdots, 0)$$

and

$$A(t) = \begin{pmatrix} 0 & \vdots & I_{n-1} \\ \cdots & \vdots & \cdots\cdots \\ -a_0 & \vdots & -a_1 \cdots -a_{n-1} \end{pmatrix}.$$

For $j = 1$ to n, i_j is an n dimensional vector having 1 in the jth position and zeros elsewhere.

COROLLARY 2. *Put $\alpha = n$ and assume that if $\beta \geq 3$, then for $j = 0$ to $n - 1$, $a_j(t)$ has $\beta - 1$ continuous derivatives. Then:*

(i) *The results of Theorems 1 and 2 and Corollary 1 hold for (18) and (19).*

(ii) *Let L^* be the adjoint of the differential operator. Then, if $m(t) \equiv 0$,*

$$L^* r(t)^{-1} L f(t \mid N) = 0, \quad \text{for} \quad t \neq t_i, \quad i = 1 \cdots N. \tag{20}$$

For $t > t_N$

$$L f(t \mid N) = 0. \tag{21}$$

PROOF. (i) We can check that $\zeta_i = i_{j+1}$ for $j = 0$ to $n - 1$, and that the first $n - j - 1$ elements of η_j are zero. It follows that $\zeta_j' b = 0$ for $j = 0$ to $\alpha - 2$, $\zeta_j' \eta_1 = 0$ for $j = 0$ to $\alpha - 3$, and $c_i' \eta_j = 0$ for all i and $j = 0$ to $\beta - 2$. Now

$$\zeta_n'(t) = -(a_0(t), \cdots, a_{n-1}(t))$$

so that if $\alpha + \beta - 2 > n$, i.e., $\beta \geq 3$, $\zeta_j(t)$ exists and is continuous for $j = 2$ to $\alpha + \beta - 2$.

Having checked that Assumptions 1 to 4 hold, it will follow that (i) holds.

(ii) $m(t) = 0$ implies that $E\{f(t)\} = 0$. From (13) it will suffice to show that

$$L^* r(t)^{-1} L \operatorname{cov}(f(t), y_i) = 0, \quad i = 1 \text{ to } N, \quad t \neq t_i.$$

We will use the expression for $\operatorname{cov}(f(t), y_i)$ given by (9) and (10). First note that $L\{c' D(t, T_0)\} = 0$ for all t. Next for $t < t_i$,

$$dQ_i(t) c_i / dt - A(t) Q_i(t) c_i = bb' D(t_i, t)' c_i r(t)$$

with Q_i defined as in (10). It follows that

$$L(c' Q_i(t) c_i) = i_n' \left\{ \frac{d}{dt} Q_i(t) c_i - A(t) Q_i(t) c_i \right\} = b' D(t_i, t)' c_i r(t).$$

Hence

$$r(t)^{-1} L(c' Q_i(t) c_i) = i_n' D(t_i, t)' c_i.$$

From (7), we can check that

$$\frac{d}{dt} D(t_i, t)' c_i + A(t)' D(t_i, t)' c_i = 0$$

so that (Levinson and Coddington, 1955, page 85)

$$L^* i_n' D(t_i, t)' c_i = 0.$$

Therefore

$$L^* r(t)^{-1} L\{c' Q_i(t) c_i\} = 0,$$

so that (20) holds. Using the same technique as above, we can similarly show that for $t > t_i$, $L(c' Q_i(t) c_i) = 0$. Hence (20) holds for all i and all $t \neq t_i$. (21) follows easily from the above discussion.

REMARK 2. If the coefficients a_j, $j = 0, \cdots, n - 1$, are constant, then a thorough analysis of the smoothness properties of $f(t \mid N)$ is given by WDS. In fact their analysis considers a rational differential operator with constant coefficients. See Weinert, Desai and Sidhu (1979), Theorem 6.1. It is straightforward to apply our Theorem 1 to obtain their results.

REMARK 3. Weinert, Byrd and Sidhu (1980) show that the optimal Lg smoothing spline equals $f(t \mid N)$ where the stochastic process $f(t)$ is described in this section, $W(t)$ is a Wiener process, and in addition,

$$\begin{aligned} \operatorname{cov}(f(t), e_i) &= -\rho_i z_i(t), \quad i = 1, \cdots, n; \\ &= 0 \qquad\qquad\quad i > n. \end{aligned}$$

The $z_i(t)$, $i = 1, \cdots, n$ form the basis of the null space of L.

Now

$$\operatorname{cov}(f(t), y_i) = \operatorname{cov}(f(t), \lambda_i(f)) + \operatorname{cov}(f(t), e_i),$$

and $z_i(t)$, $i = 1, \cdots, n$, has all the smoothness properties described in Corollary 2. It follows that Corollary 2 holds for $f(t \mid N)$ and $\sum_f (t \mid N)$ with $f(t)$ as just described and hence for Lg smoothing splines.

Acknowledgment. We would like to thank a referee for several useful suggestions and corrections.

REFERENCES

BREIMAN, L. (1968). *Probability.* Addison-Wesley, Reading, Mass.

CODDINGTON, E. A., and LEVINSON, N. (1955). *Theory of Ordinary Differential Equations.* McGraw-Hill, New York.

HANNAN, E. J. (1970). *Multiple Time Series.* Wiley, New York.

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

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

KIMELDORF, G., and WAHBA, G. (1971). Some results on Tchebycheffian spline functions. *J. Multivariate Anal. Appl.* **33** 82–95.

PARZEN, E. (1962). *Stochastic Processes.* Holden Day, San Francisco.

WAHBA, G. (1978). Improper priors, spline smoothing and the problem of guarding against model errors in regression. *J. Roy. Statist. Soc. Ser. B* **40** 364–72.

WEINERT, H. L., BYRD, R. H., and SIDHU, G. S. (1980). A stochastic framework for recursive computation of spline functions: Part II, Smoothing Splines. *J. Optim. Theor. Appl.* **30** 255–268.

WEINERT, H. L., DESAI, U. B., AND SIDHU, G. S. (1979). ARMA splines, system inverses, and least squares estimates. *Siam J. Control Optim.* **17** 525–536.

WEINERT, H. L., and KAILATH, T. (1974). Stochastic interpretations and recursive algorithms for spline functions. *Ann. Statist.* **7** 787–794.

WEINERT, H. L. and SIDHU, G. S. (1978). A stochastic framework for recursive computation of spline functions: Part I, Interpolating Splines. *IEEE Trans. Inform. Theory* **24** 45–50.

GRADUATE SCHOOL OF BUSINESS
UNIVERSITY OF CHICAGO
1101 EAST 58TH STREET
CHICAGO, ILLINOIS 60637