

FINITE DIMENSIONAL NONLINEAR ESTIMATION
IN CONTINUOUS AND DISCRETE TIME

Steven I. Marcus
Department of Electrical Engineering
The University of Texas at Austin

Sanjoy K. Mitter
Department of Electrical Engineering and Computer Science
Massachusetts Institute of Technology

Daniel Ocone
Department of Mathematics
Massachusetts Institute of Technology

1. Introduction

In [1]-[3] we have shown that, for certain classes of nonlinear stochastic systems in both continuous and discrete time, the optimal conditional mean estimator of the system state given the past observations can be computed with a recursive filter of fixed finite dimension. The typical nonlinear system in these classes consists of a linear system with linear measurements and white Gaussian noise processes, which feeds forward into a nonlinear system described by a certain type of Volterra series expansion or by a bilinear or state-linear system satisfying certain algebraic conditions. It is our purpose in this paper to consider estimation problems similar to those in [1]-[3], to present simpler proofs that the estimators are indeed finite dimensional, to provide deeper insight into these problems by relating them to the homogeneous chaos of Wiener and to orthogonal polynomial expansions [4]-[8],[24], to explain the similarities and differences between the continuous and discrete time cases, and to prove some extensions of our previous results. The existence of polynomials in the innovations in the discrete time recursive estimator, in contrast to the continuous time estimator (as noted in [2]), is interpreted

in terms of the homogeneous chaos. The existence of such polynomials in the innovations in the optimal filter suggests that suboptimal filter design in discrete time could be improved by incorporating such structure; this is in contrast to most discrete time estimator designs, such as the extended Kalman filter, in which the updated estimate is linear in the innovations (exceptions are the quasi-moment estimators of [9] and [10]) and the higher measurement space filter of [23].

2. Problem Statement

As in [1]-[3], the classes of systems considered in this paper are described as follows. It will be assumed that all random variables and processes are defined on a probability space (Ω, B, P) . In continuous time, we consider systems of the form, for $t \in [0, T]$,

$$dx(t) = A(t)x(t)dt + B(t)dw(t) \quad (1)$$

$$dy(t) = f(x(t), y(t), t)dt \quad (2)$$

$$dz_1(t) = C(t)x(t)dt + R^{\frac{1}{2}}dv(t) \quad (3)$$

where $x(t) \in \mathbb{R}^n$, $y(t) \in \mathbb{R}^m$, $z(t) \in \mathbb{R}^p$, w and v are standard vector Wiener processes, $R > 0$, $x(0)$ is Gaussian, $\{x(0), y(0), w(t), v(s)\}$ are independent for all t and s , f is an analytic function of x and y , and $[A(t), B(t), C(t)]$ is completely controllable and observable.

The discrete time systems to be considered are of the form, for $t \in \{0; T\}$,

$$x(t+1) = A(t)x(t) + B(t)w(t) \quad (4)$$

$$y(t+1) = f(x(t), y(t), t) \quad (5)$$

$$z_2(t) = C(t)x(t) + R^{\frac{1}{2}}v(t) \quad (6)$$

where $T \in \mathbb{Z}^+$, the set of positive integers, and $\{s; t\}$ is the set of integers $\{s, s+1, \dots, t\}$. The assumptions in (4)-(6) are the same as those in (1)-(3),

except that $w(t)$ and $v(t)$ are zero-mean Gaussian white noise processes. Motivation for the study of systems of the form (1)-(3) and (4)-(6) is presented in [3].

The optimal estimate, with respect to a wide variety of criteria (including minimum mean square error), of $x(t)$ given the past observations $z_1^t \triangleq \{z_1(s), 0 \leq s \leq t\}$ or $z_2^t \triangleq \{z_2(s), s \in \{0;t\}\}$, is the conditional mean $\hat{x}(t|t)$ of $x(t)$ given the σ -field $F_t^{z_i}$ generated by z_i^t , also denoted $E[x(t)|z_i^t]$. It is assumed that all the relevant random variables are in $L_2(\Omega, B, P)$, so the conditional expectation $\hat{x}(t|t)$ can also be interpreted as the orthogonal projection of $x(t)$ onto the subspace $L_2(\Omega, F_t^{z_i}, P)$ [14, App. A.]; this interpretation will be used in the sequel. Predicted and smoothed estimates will also be used extensively, so we introduce the equivalent notations $\hat{x}(s|t) \triangleq E[x(s)|z_i^t] \triangleq E^t[x(s)] \triangleq E[x(s)|F_t^z]$. Thus our objective is the recursive computation of $\hat{x}(t|t)$ and $\hat{y}(t|t)$. The computation of $\hat{x}(t|t)$ can be performed by the recursive n -dimensional (linear) Kalman filter in continuous or discrete time. It is, in general, not possible to compute $\hat{y}(t|t)$ with a recursive estimator of fixed finite dimension. It has been proved in [1]-[3] that if the nonlinear system (2) or (4) is characterized by a certain type of finite series expansion or by certain bilinear or state-affine equations, then $\hat{y}(t|t)$ can be computed by such a recursive finite dimensional estimator. Some of the major results can be summarized as follows.

Let the Volterra series expansions for the i^{th} components of $y(t)$ in (2) and (4) be given by

$$y_i(t) = w_{0i}(t) + \sum_{k=1}^{\infty} \int_0^t \dots \int_0^t \sum_{\alpha_1, \dots, \alpha_k=1}^n w_{ki}^{(\alpha_1, \dots, \alpha_k)}(t, \sigma_1, \dots, \sigma_k) \cdot x_{\alpha_1}(\sigma_1) \dots x_{\alpha_k}(\sigma_k) d\sigma_1 \dots d\sigma_k \quad (7)$$

and

$$y_i(t) = w_{0i}(t-1) + \sum_{k=1}^{\infty} \sum_{\ell_1, \dots, \ell_k=0}^{t-1} \sum_{\alpha_1, \dots, \alpha_k=1}^n w_{ki}^{(\alpha_1, \dots, \alpha_k)}(t, \ell_1, \dots, \ell_k) \cdot x_{\alpha_1}(\ell_1) \dots x_{\alpha_k}(\ell_k), \quad (8)$$

respectively. Here $w_{ki}^{(\alpha_1, \dots, \alpha_k)}$ is called a k^{th} order kernel, and a finite Volterra series expansion of order q is one such that all k^{th} order kernels are zero for $k > q$. In the continuous case (7), we consider, without loss of generality [11], only triangular kernels which satisfy

$w_{ki}^{(\alpha_1, \dots, \alpha_k)}(t, \sigma_1, \dots, \sigma_k) = 0$ unless $0 < \sigma_1 < \dots < \sigma_k < t$. Such a kernel is separable if it can be expressed as a finite sum

$$w_{ki}^{(\alpha_1, \dots, \alpha_k)}(t, \sigma_1, \dots, \sigma_k) = \sum_{j=1}^m \gamma_{j0}(t) \gamma_{j1}(\sigma_1) \dots \gamma_{jk}(\sigma_k). \quad (9)$$

Similar definitions can be made in discrete time [2], but they are more complicated (this difficulty is related to the fact that the solution of a discrete time system may not be defined backward in time [12],[21]). Brockett [11] and Gilbert [13] have shown that the kernels in (7) are separable if f is analytic. Using variational expansions similar to those of Gilbert [13], it is straightforward to show that the kernels of the Volterra series (8) are also separable in the sense of [2],[12]; this is basically due to the fact that the kernels arise from the variational equations as products of pulse responses of linear systems. Brockett [11] has also shown that a continuous time finite Volterra series has a bilinear realization if and only if it has separable kernels. The separability and realizability results are crucial in the proofs of the following two theorems.

Theorem 1 [1]: Consider the system (1)-(3), and assume that (2) has a finite Volterra series expansion. Then $\hat{y}(t|t)$ can be computed with a finite dimensional recursive estimator--i.e., by a finite set of nonlinear stochastic differential equations driven by the innovations

$$v_1(t) \triangleq z_1(t) - \int_0^t C(s)\hat{x}(s|s)ds. \quad (10)$$

Theorem 2 [2]: Consider the system (4)-(6), and assume that (5) has a finite Volterra series expansion. Then $\hat{y}(t|t)$ can be computed with a finite set of nonlinear difference equations driven by the innovations

$$v_2(t) = z_2(t) - C(t)A(t-1)\hat{x}(t-1|t-1). \quad (11)$$

The basic technique employed in [1]-[3] to prove these theorems is the augmentation of the state of the original system with additional states which arise as smoothed statistics of the original state. For the classes of systems considered here, it is shown that only a finite number of additional states (smoothed statistics) are required. We will see here, from a different point of view, how the additional filter states arise. In addition, we will prove results similar to Theorems 1 and 2 for some systems in which equations (2) and (4) for $y(t)$ contain an additive noise term.

In this paper both the continuous and discrete time problems will be considered in a unified framework. It is useful first to contrast these problems with the estimation and prediction problems considered by Huang and Cambanis [8]. There the problem is that of estimating a nonlinear functional y of a Gaussian process $\{x(t), t \in S\}$, given observations of $\{x(t), t \in \tilde{S}\}$, where \tilde{S} is a subset of S . In our problem the objective is

to recursively estimate a nonlinear functional $y(t)$ of $x(\cdot)$, given observations of linear functionals of $x(\cdot)$ plus noise. Although the elegant formulas of Huang and Cambanis cannot be applied here, the approach of utilizing the homogeneous chaos, or, equivalently, the Cameron-Martin orthogonal series decomposition of a Gaussian process [4]-[8], will prove to be quite useful in unifying and simplifying our results.

By employing the "innovations approach" [15],[16], the conditional expectations $\hat{y}(t|t)$ of Theorems 1 and 2 can equivalently be viewed as projections on Hilbert spaces generated by the innovations instead of the observations. For the discrete time problem (4)-(6) it can easily be shown recursively that $F_t^{z_2} = F_t^{v_2}$, so that $\hat{y}(t|t)$ is the projection of $y(t)$ onto $L_2(\Omega, F_t^{v_2}, P)$; in fact $v_2(\cdot)$ is just obtained from the Gram-Schmidt orthogonalization of the sequence $z_2(\cdot)$. It has been shown [15],[16] for continuous time Gaussian processes (as in (1),(3)) that $F_t^{z_1} = F_t^{v_1}$; hence, $\hat{y}(t|t)$ is the projection of $y(t)$ onto $L_2(\Omega, F_t^{v_1}, P)$. The innovations process $v_1(t)$ is a Wiener process with the same covariance as $R^{\frac{1}{2}} v(t)$ [14]-[16]; the innovations process $v_2(t)$ is a zero-mean Gaussian white noise sequence with $E[v_2(t)v_2(t)'] = C(t)P(t|t-1)C'(t) + R$, where $P(t|t-1)$ is the Kalman filter one-step error covariance matrix [17]. In both cases, the linear and nonlinear innovations are equal. Hence the estimation problem (1)-(3) or (4)-(6) can be reformulated as that of estimating $y(t)$, a nonlinear L_2 -functional of the Gaussian process x^t ; the estimate $\hat{y}(t|t)$ is the nonlinear L_2 -functional of the innovations process (either v_1^t or v_2^t) which minimizes the mean square error. The expansion of such L_2 -functionals of Gaussian processes is the subject of [4]-[8], and the application of these results to our recursive estimation problem is presented in the next section, where a new proof of Theorem 1 is presented and the corresponding

proof of Theorem 2 is outlined.

3. L_2 -Functionals of Gaussian Processes and Finite Dimensional Estimation

Kallianpur [7] has generalized the earlier results of Cameron and Martin [5] and Ito [6] on the orthogonal decomposition of L_2 -functionals of a Gaussian process. We will not require all of the isomorphisms presented in [7]; only the following decomposition in terms of Hermite polynomials will be utilized here [8]. Let $x(t)$, $t \in S$, be any zero-mean second order Gaussian process defined on (Ω, B, P) ; for our purposes S will be either an interval $[0, T]$ or the discrete time set $\{0; T\}$. Define the two Hilbert spaces associated with x : the nonlinear space $L_2(x) \triangleq L_2(\Omega, F^x, P)$, where F^x is the σ -algebra generated by $x(t)$, $t \in S$; and the linear space $H(x)$, the closed subspace of $L_2(x)$ spanned by $x(t)$, $t \in S$.

Lemma 1 [7],[8]: If $\{\xi_{\gamma}, \gamma \in \Gamma\}$ (Γ linearly ordered) is a complete orthonormal set (CONS) in $H(x)$, then the family

$$(p_{\gamma_1}! \dots p_{\gamma_k}!)^{-\frac{1}{2}} H_{p_1}(\xi_{\gamma_1}) \dots H_{p_k}(\xi_{\gamma_k}),$$

$$p \geq 0, k \geq 1, p_1 + \dots + p_k = p, \gamma_1 < \dots < \gamma_k,$$

is a CONS in $L_2(x)$, where H_n is the n^{th} normalized Hermite polynomial.

That is, any L_2 -functional θ of $x(\cdot)$ has the orthogonal series expansion

$$\theta = \sum_{p \geq 0} \sum_{\substack{p_1 + \dots + p_k = p \\ \gamma_1 < \dots < \gamma_k}} a_{\gamma_1 \dots \gamma_k}^{p_1 \dots p_k} H_{p_1}(\xi_{\gamma_1}) \dots H_{p_k}(\xi_{\gamma_k}). \quad (12)$$

Remark: If x has nonzero mean, the representation of Lemma 1 can be written with respect to a centered CONS, and the coefficients in (12) will depend on the mean of x .

Corollary 1 [6]: If $x(t)$, $t \in [0, T]$ is a standard Wiener process, then any $\theta \in L_2(x)$ has the orthogonal expansion

$$\theta = \sum_{p \geq 0} \int_0^T \int_0^t \dots \int_0^{t_2} f_p(t_1, \dots, t_p) dx(t_1) \dots dx(t_{p-1}) dx(t_p) \\ \triangleq \sum_{p \geq 0} I_p(f_p) \quad (13)$$

where the integrals in (13) are iterated stochastic integrals; also, $I_p(f_p)$ and $I_q(f_q)$ are orthogonal for all $p \neq q$.

Now we consider the estimation problems of Theorems 1 and 2 in this framework. Assume throughout this section, for simplicity of notation, that x, y, z_1 , and z_2 are all scalars; the following results also hold in the vector case. The state $y(t)$ (as given by (2) or (4)) is a nonlinear functional of x^t ; assume that $y(t)$ has a finite Volterra series expansion (of the form (7) or (8)) of order q . It is then clear that $y(t)$ has a finite orthogonal series expansion (12) of order q -- i.e., with $a_{\gamma_1 \dots \gamma_k}^{p_1 \dots p_k} = 0$ for $p > q$. In the continuous case, the $\{\xi_i\}$ are centered versions of functionals of the form $\int_0^t \phi_i(s) x(s) ds$, while in the discrete time the $\{\xi_i\}$ are just centered linear combinations of the $x(s)$, $s \in \{1; t\}$. The estimate $\hat{y}(t|t)$ is a nonlinear L_2 -functional of the Gaussian innovations process; thus it also has an orthogonal expansion of the form (12). In continuous time $v_1(t)$ is a Wiener process, so $\hat{y}(t|t)$ has the expansion (13) with $x(t) = R^{-1/2} v_1(t)$. In discrete time, $v_2(t)$ is an orthogonal sequence, so $\eta(t) \triangleq [C(t)^2 P(t|t-1) + R]^{-1/2} v_2(t)$ is a CONS in $H(v_2^t)$, and the expansion (12) is valid with $\xi_i = \eta(i)$.

Thus Theorems 1 and 2 can be proved by showing that: (a) the orthogonal series expansion of $\hat{y}(t|t)$ has only a fixed finite number of

terms for all t ; and (b) such a finite orthogonal series can be realized as the output of a finite dimensional recursive system (i.e., a system in state-space form). The states of this finite dimensional system are the additional filter states referred to in Section 2. The following theorem proves (a) for a more general formulation; the proof of (b) must be done separately for continuous and discrete time, and involves the calculation and separability of the Volterra kernels.

Theorem 3: Let $x(t)$, $z(t)$, $t \in S$ be zero-mean jointly Gaussian second order processes, and assume that $y \in L_2(x)$ has an orthogonal series expansion of order q . Let the orthogonal expansion of $\hat{y} \triangleq E[y|F^Z]$ be given by

$$\hat{y} = \sum_{r \geq 0} \sum_{\substack{r_1 + \dots + r_j = r \\ \beta_1 < \dots < \beta_j}} b_{\beta_1 \dots \beta_j}^{r_1 \dots r_j} H_{r_1}(\eta_{\beta_1}) \dots H_{r_j}(\eta_{\beta_j}) \quad (14)$$

where $\{\eta_\delta, \delta \in \Delta_1\}$ is a CONS in $H(z)$. Then

$$b_{\beta_1 \dots \beta_j}^{r_1 \dots r_j} = 0, \quad r > q;$$

that is, \hat{y} also has an orthogonal expansion of order q .

Proof: Consider $H(x, z)$, the linear space spanned by $\{x(t), z(t); t \in S\}$. Since $\{\eta_\delta, \delta \in \Delta_1\}$ is an orthonormal set in $H(x, z)$, it can be completed by adding elements $\{\eta_\delta, \delta \in \Delta_2\}$ in $H(x, z)$ to form the CONS $\{\eta_\delta, \delta \in \Delta_1 \cup \Delta_2\}$ in $H(x, z)$. The orthogonal expansion for y can then be rewritten in terms of this CONS in $H(x, z)$; the new expansion is clearly also of order q :

$$y = \sum_{p=0}^q \sum_{\substack{p_1 + \dots + p_k = p \\ \gamma_1 < \dots < \gamma_k}} c_{\gamma_1 \dots \gamma_k}^{p_1 \dots p_k} H_{p_1}(\eta_{\gamma_1}) \dots H_{p_k}(\eta_{\gamma_k}) \quad (15)$$

where $\{\gamma_i\} \in \Delta_1 \cup \Delta_2$. Now \hat{y} is the orthogonal projection of y onto $L_2(z)$;

that is, by Lemma 1 \hat{y} is just the projection of y onto the space spanned by the products $H_{p_1}(\eta_{\gamma_1}) \dots H_{p_k}(\eta_{\gamma_k})$ with $\gamma_1, \dots, \gamma_k \in \Delta_1$. The orthogonality of such products in $L_2(x, z)$ (see Lemma 1) then yields

$$\hat{y} = \sum_{p=0}^q \sum_{\substack{p_1 + \dots + p_k = p \\ \gamma_1 < \dots < \gamma_k}} c_{\gamma_1 \dots \gamma_k}^{p_1 \dots p_k} \tilde{H}_{p_1}(\eta_{\gamma_1}) \dots H_{p_k}(\eta_{\gamma_k}) \quad (16)$$

where $\{\gamma_i\} \in \Delta_1$, thus proving the theorem.

Theorem 3 also holds for nonzero-mean and vector-valued processes, with obvious modifications in the proof. This theorem then applies to $\hat{y}(t|t)$ of Theorems 1 and 2. It remains only to prove that the finite orthogonal series expansion for $\hat{y}(t|t)$ is realizable with a nonlinear recursive system of fixed finite dimension. Consider first the continuous time problem (1)-(3).

Proof of Theorem 1: Assume that $y(t)$ has a finite Volterra series expansion of order q . Then Theorem 3 implies that $\hat{y}(t|t)$ has the orthogonal expansion

$$\hat{y}(t|t) = \sum_{p=0}^q \int_0^t \int_0^{s_p} \dots \int_0^{s_2} f_p(t, s_1, \dots, s_p) dv(s_1) \dots dv(s_p) \quad (17)$$

where $v(t) \triangleq R^{-\frac{1}{2}} dv_1(t)$. The projection theorem and the orthogonality of the iterated stochastic integrals [6] imply that, for $s_1 < \dots < s_p < t$,

$$f_p(t, s_1, \dots, s_p) = \frac{1}{p!} \frac{\partial^p}{\partial s_1 \dots \partial s_p} E[y(t)v(s_1) \dots v(s_p)] \quad (18)$$

(the proof of (18) is analogous to that of Davis [14, p. 95] for the best

linear estimate).¹ A proof identical to that of Brockett [11] for the deterministic case shows that, if the kernels (18) are separable (see (9)), then $\hat{y}(t|t)$ in (17) can be generated as the output of a finite dimensional bilinear system driven by the innovations $v(t)$. Hence, Theorem 1 is proved if the kernels in (18) are separable.

Lemma 2: The triangular kernels $f_p(t, s_1, \dots, s_p)$ given by (18) are separable for $s_1 < \dots < s_p < t$ under the hypotheses of Theorem 1.

Proof: Let $y(t)$ be given by one k^{th} order term in the finite Volterra series (7); the proof generalizes in the obvious way. Since the kernels of (7) are separable due to the analyticity assumption in (2), we can assume that

$$y(t) = \int_0^t \int_0^{\tau_k} \dots \int_0^{\tau_2} \gamma_1(\tau_1) \dots \gamma_k(\tau_k) x(\tau_1) \dots x(\tau_k) d\tau_1 \dots d\tau_k \quad (19)$$

Thus, by the Fubini theorem (see [1],[3])

$$f_p(t, s_1, \dots, s_p) = \frac{1}{p!} \int_0^t \int_0^{\tau_k} \dots \int_0^{\tau_2} \gamma_1(\tau_1) \dots \gamma_k(\tau_k) \cdot \frac{\partial^p}{\partial s_1 \dots \partial s_p} E[x(\tau_1) \dots x(\tau_k) v(s_1) \dots v(s_p)] d\tau_1 \dots d\tau_k \quad (20)$$

Since $x(\tau_1), \dots, x(\tau_k), v(s_1), \dots, v(s_p)$ are jointly Gaussian, the expectation in (20) can be expanded via Lemma B.1 of [1], resulting in a sum of products of terms of the form: $E[x(\tau_i)], E[v(s_i)],$

¹In general, whenever the linear innovations $v(t)$ in a nonlinear estimation problem form a Wiener process, then an (infinite) orthogonal expansion of the form (17) will hold for the estimate of each L_2 -state $y(t)$, and the kernels are calculated via (18). The sum of the first two terms ($p=0,1$) in (17) is the best linear estimate, the sum of the terms for $p=0,1,2$ yields the best quadratic estimate, etc. These are not necessarily realizable with finite dimensional recursive filters.

$\text{cov}[x(\tau_i), x(\tau_j)]$, $\text{cov}[x(\tau_i), v(s_j)]$, and $\text{cov}[v(s_i), v(s_j)]$. Notice that $E[v(s_i)] = 0$, so all products involving such terms are zero. If

$\text{cov}[v(s_i), v(s_j)]$ arises, it results in a term of the form

$\frac{\partial^2}{\partial s_i \partial s_j} \text{cov}[v(s_i), v(s_j)]$, which can be shown to be zero for $s_i \neq s_j$.

Also, $\text{cov}[x(\tau_i), x(\tau_j)]$ is the covariance function of the state of the linear system (1); hence, for $\tau_i > \tau_j$,

$$\text{cov}[x(\tau_i), x(\tau_j)] = \exp\left[\int_{\tau_j}^{\tau_i} A(\sigma) d\sigma\right] \text{cov}[x(\tau_j), x(\tau_j)] \quad (21)$$

Finally, consider

$$\begin{aligned} R^{\frac{1}{2}} \text{cov}[x(\tau_i), v(s_j)] &= \text{cov}[x(\tau_i), z(s_j) - \int_0^{s_j} C(\sigma) \hat{x}(\sigma|\sigma) d\sigma] \\ &= \text{cov}[x(\tau_i), \int_0^{s_j} C(\sigma) (x(\sigma) - \hat{x}(\sigma|\sigma)) d\sigma + v(s_j)] \\ &= \text{cov}[x(\tau_i), \int_0^{s_j} C(\sigma) (x(\sigma) - \hat{x}(\sigma|\sigma)) d\sigma] \quad (22) \end{aligned}$$

since $x(\tau_i)$ and $v(s_j)$ are independent. This gives rise in (20) to

$$\begin{aligned} \frac{\partial}{\partial s_j} \text{cov}[x(\tau_i), v(s_j)] &= R^{-\frac{1}{2}} \text{cov}[x(\tau_i), C(s_j) (x(s_j) - \hat{x}(s_j|s_j))] \\ &= C(s_j) R^{-\frac{1}{2}} \text{cov}[x(\tau_i), x(s_j) - \hat{x}(s_j|s_j)], \quad (23) \end{aligned}$$

which is the covariance function of a finite (two-) dimensional linear system with states $x(t)$ and $x(t) - \hat{x}(t|t)$, and is thus also separable.

Lemma B.1 of [1] and the separability of the relevant covariance functions imply that there exist functions $\{\alpha_{\ell_i}, \beta_{\ell_i}\}$ such that (20) can be written as

$$\begin{aligned}
f_p(t, s_1, \dots, s_p) &= \frac{1}{p!} \int_0^t \int_0^{\tau_k} \dots \int_0^{\tau_2} \gamma_1(\tau_1) \dots \gamma_k(\tau_k) \\
&\quad \cdot \left(\sum_{\ell=1}^m \alpha_{\ell 1}(\tau_1) \dots \alpha_{\ell k}(\tau_k) \beta_{\ell 1}(s_1) \dots \beta_{\ell p}(s_p) \right) d\tau_1 \dots d\tau_k \\
&= \frac{1}{p!} \sum_{\ell=1}^m \alpha_{\ell}(t) \beta_{\ell 1}(s_1) \dots \beta_{\ell p}(s_p), \tag{24}
\end{aligned}$$

and f_p is separable as claimed; this also completes the proof of Theorem 1.

An example in which the kernels and the recursive estimator are computed explicitly is presented in the next section. The discrete time result which is analogous to Lemma 2 can be used to prove Theorem 2, but for the sake of brevity we will only present an example of the procedure (Section 5).

4. A Continuous Time Example

Before discussing the example, we present an extension of Theorem 1 to a class of systems in which $y(t)$ contains process noise; the analogous extension of Theorem 2 is proved in the same manner.

Theorem 4: Consider the system (1)-(3), and assume that (2) has a finite Volterra series expansion. Assume that there is an additional state $y_1(t)$ satisfying

$$dy_1(t) = (F(t)y_1(t) + G(t)y(t))dt + H(t)d\tilde{w}(t) \tag{25}$$

where \tilde{w} is a Wiener process and $\{x(0), y(0), y_1(0), w(t_1), v(t_2), \tilde{w}(t_3)\}$ are independent for all t_1, t_2, t_3 . Then $\hat{y}_1(t|t)$ can also be computed with a finite dimensional recursive estimator.

Proof: The solution of (25) is

$$\begin{aligned}
y_1(t) &= \Phi(t,0)y_1(0) + \int_0^t \Phi(t,s)G(s)y(s)ds + \int_0^t \Phi(t,s)H(s)d\tilde{w}(s) \\
&\triangleq \tilde{y}_1(t) + \int_0^t \Phi(t,s)H(s)d\tilde{w}(s)
\end{aligned} \tag{26}$$

where Φ is the state transition matrix for F . Since $\tilde{w}(\cdot)$ and $z(\cdot)$ are independent, $\hat{y}_1(t|t) \triangleq E[y_1(t)|F_t^Z] = E[\tilde{y}_1(t)|F_t^Z]$, and $\tilde{y}_1(t)$ is just described by a finite Volterra series expansion in x . The theorem then follows from Theorem 1.

Example 1: Consider the scalar system

$$dx(t) = -\alpha x(t)dt + dw_1(t) \tag{27}$$

$$dy(t) = (-\gamma y(t) + x^2(t))dt + dw_2(t) \tag{28}$$

$$dz(t) = x(t)dt + dv(t) \tag{29}$$

with the same assumptions as in Theorem 4. The solution of (28) is

$$y(t) = e^{-\gamma t} y(0) + \int_0^t e^{-\gamma(t-\sigma)} x^2(\sigma)d\sigma + \int_0^t e^{-\gamma(t-\sigma)} dw_2(\sigma) \tag{30}$$

By Theorems 3 and 4, it follows that

$$\hat{y}(t|t) = f_0(t) + \int_0^t f_1(t,s)dv(s) + \int_0^t \int_0^{s_2} f_2(t,s_1,s_2)dv(s_1)dv(s_2), \tag{31}$$

where $v(t) = z(t) - \int_0^t \hat{x}(s|s)ds$. Using (18) to compute the kernels as in Lemma 2, we have (since $y(0)$ and \tilde{w} are independent of v)

$$f_0(t) = E[y(t)] = e^{-\gamma t} E[y(0)] + \int_0^t e^{-\gamma(t-\sigma)} E[x^2(\sigma)]d\sigma \tag{32}$$

$$\begin{aligned}
f_1(t,s) &= \frac{\partial}{\partial s} E[y(t)v(s)] = \frac{\partial}{\partial s} \int_0^t e^{-\gamma(t-\sigma)} E[x^2(\sigma)v(s)] d\sigma \\
&= 2 \int_0^t e^{-\gamma(t-\sigma)} m(\sigma) \frac{\partial}{\partial s} E[x(\sigma) \int_0^s (x(\tau) - \hat{x}(\tau|\tau)) d\tau] d\sigma \\
&= 2 \int_0^t e^{-\gamma(t-\sigma)} m(\sigma) E[x(\sigma)(x(s) - \hat{x}(s|s))] d\sigma \quad (33)
\end{aligned}$$

where $m(\sigma) = E[x(\sigma)] = e^{-\alpha\sigma} E[x(0)]$. It can be shown, using Lemma 2.2 of [1], that

$$E[x(\sigma)(x(s) - \hat{x}(s|s))] = \begin{cases} e^{-\alpha(\sigma-s)} P(s), & \sigma \geq s \\ K(s,\sigma) P(s), & \sigma < s \end{cases} \quad (34)$$

where $K(s,\sigma) = \exp[\alpha(s-\sigma) - \int_\sigma^s P^{-1}(\tau) d\tau]$ and $P(t)$ is the Kalman filter error covariance for $x(t)$. Thus

$$f_1(t,s) = 2 \left[\int_0^s e^{-\gamma(t-\sigma)} m(\sigma) K(s,\sigma) d\sigma + \int_s^t e^{-\gamma(t-\sigma)} m(\sigma) e^{-\alpha(\sigma-s)} d\sigma \right] P(s) \quad (35)$$

Similarly,

$$\begin{aligned}
f_2(t,s_1,s_2) &= \left[\int_0^{s_1} e^{-\gamma(t-\sigma)} K(s_1,\sigma) K(s_2,\sigma) d\sigma \right. \\
&\quad + \int_{s_1}^{s_2} e^{-\gamma(t-\sigma)} e^{-\alpha(\sigma-s_1)} K(s_2,\sigma) d\sigma \\
&\quad \left. + \int_{s_2}^t e^{-\gamma(t-\sigma)} e^{-\alpha(\sigma-s_1)} e^{-\alpha(\sigma-s_2)} d\sigma \right] P(s_1) P(s_2) \quad (36)
\end{aligned}$$

(recall that $0 < s_1 < s_2 < t$).

These kernels are obviously separable, so $\hat{y}(t|t)$ can be realized as the output of a finite dimensional bilinear system driven by the

innovations. However, it may not be efficient to realize each term in (31) individually. In fact, one efficient recursive realization of $\hat{y}(t|t)$ is readily derived via the procedure of [1, Example 2.1]; a recursive 3-state filter which computes both $\hat{x}(t|t)$ and $\hat{y}(t|t)$ is constructed as follows. First, augment the state $x(t)$ of (27) with the additional state $\xi(t)$ given by

$$\dot{\xi}(t) = (\alpha - \gamma - P^{-1}(t))\xi(t) + x(t); \quad \xi(0) = 0 \quad (37)$$

Then the Kalman-Bucy 2-state filter for the linear system (27), (37) with observations (29) recursively computes $\hat{x}(t|t)$ and $\hat{\xi}(t|t)$. Finally, $\hat{y}(t|t)$ is computed by

$$\begin{aligned} d\hat{y}(t|t) &= (-\gamma\hat{y}(t|t) + [\hat{x}(t|t)]^2 + P(t))dt + 2P(t)\hat{\xi}(t|t)d\nu(t) \\ \hat{y}(0|0) &= 0 \end{aligned} \quad (38)$$

To check that this filter has the series expansion (31), (32), (35), (36) is straightforward.

It should also be noted that if $x(t)$ has zero mean, then the best linear estimate of $y(t)$ given z^t (the first two terms in (31)) is equal to the a priori mean of $y(t)$. This is due to the fact that, in this case, y and z are uncorrelated. However, since y and z are not independent, the best quadratic estimator (which is equal to the conditional mean in this example) can in fact offer significant improvement in estimator performance (see [18] for some case studies and further analysis along these lines).

5. A Discrete Time Example

Example 2: Consider the scalar discrete time system

$$x(t+1) = \alpha x(t) + w_1(t) \quad (39)$$

$$y(t+1) = \gamma y(t) + x^2(t) + w_2(t) \quad (40)$$

$$z(t) = x(t) + v(t) \quad (41)$$

with the same assumptions as in Theorem 2; also, w_2 is a discrete time white noise process independent of $x(0)$, $y(0)$, w_1 , and v . The solution of (40) is

$$y(t) = \gamma^t y(0) + \sum_{i=0}^{t-1} \gamma^{t-i-1} x^2(i) + \sum_{i=0}^{t-1} \gamma^{t-i-1} w_2(i) \quad (42)$$

By Theorem 3, it follows from (16) that

$$\begin{aligned} \hat{y}(t|t) &= c_0(t) + \sum_{j=0}^t c_1(t,j) H_1(\eta(j)) \\ &+ \sum_{\substack{j,k=0 \\ j < k}}^t c_{11}(t,j,k) H_1(\eta(j)) H_1(\eta(k)) \\ &+ \sum_{j=0}^t c_2(t,j) H_2(\eta(j)) \end{aligned} \quad (43)$$

where $\eta(t) = [P(t|t-1) + 1]^{-\frac{1}{2}} [z(t) - \alpha \hat{x}(t-1|t-1)]$ are the normalized innovations. By orthogonality (Lemma 1) and the projection theorem,

$$c_{p_1 \dots p_k}(t, j_1, \dots, j_k) = \frac{1}{p_1! \dots p_k!} E[y(t) H_{p_1}(\eta(j_1)) \dots H_{p_k}(\eta(j_k))] \quad (44)$$

These kernels can, as in Example 1, be explicitly evaluated. They are indeed separable, and a 3-state filter can be constructed as follows using the methods of [1],[2]. First, augment the state $x(t)$ of (39) with

$$\xi(t+1) = \frac{\alpha \gamma P(t|t)}{P(t+1|t)} \xi(t) + \frac{\gamma P(t|t)}{P(t+1|t)} x(t); \quad \xi(0) = 0 \quad (45)$$

Then $\hat{x}(t|t)$ and $\hat{\xi}(t|t)$ can be calculated by a 2-state Kalman filter.

Finally,

$$\begin{aligned}
\hat{y}(t+1|t+1) &= \alpha \hat{y}(t|t) + \hat{x}(t|t)^2 + \delta(t) \\
&+ 2M(t, t+1) [\hat{x}(t|t) + \gamma \hat{\xi}(t|t)] v(t+1) \\
&+ \left[\sum_{i=0}^t \gamma^{t-i} M(i, t+1)^2 \right] v(t+1)^2 \\
\hat{y}(0|0) &= 0
\end{aligned} \tag{46}$$

$$M(i, t+1) = \frac{\alpha^{t-i+1} P(i|i) \dots P(t+1|t+1)}{P(i+1|i) \dots P(t+1|t)} \tag{47}$$

and $\delta(t)$ are deterministic functions, and $v(t) = z(t) - \alpha \hat{x}(t-1|t-1)$ is the unnormalized innovations process.

Notice that the recursive optimal estimator (46) contains a final term which is quadratic in the innovations. In general, if $y(t)$ contains a Volterra series of order q in $x(t)$, then the recursive estimator for $\hat{y}(t|t)$ will contain polynomials of degree q in $v(t)$. This result was proved in [2], but it also follows naturally from the orthogonal series decomposition (16) of $\hat{y}(t|t)$ -- if $y(t)$ has a Volterra series of order q , (16) will contain terms such as $H_q(\eta(t))$, or polynomials of order q in $\eta(t)$. This phenomenon does not occur in continuous time estimation problems with observations corrupted by "Gaussian white noise", in which the optimal recursive estimator is always linear in the innovations.

In [2], this contrast is explained by means of the different martingale representation theorems in continuous and discrete time [19], [20]. However, a simple explanation is provided by the representation (16). In continuous time, the elements η_γ of the CONS in $L_2(z^t)$ are of the form $\int_0^t \phi_\gamma(s) dv_1(s)$, and the series (16) can be expressed in terms of iterated stochastic integrals as in (17). Given separability, the series can then be realized with a finite dimensional bilinear system -- that is,

the stochastic differential equations in the realization are linear in $dv_1(t)$. In the discrete time case, the elements η_γ of the CONS in $L_2(z^t)$ are given by the normalized discrete time innovations $\eta(t)$; the series (16) then gives rise to a finite Volterra series in the innovations $v_2(t)$ which contains polynomials in $v_2(t)$. Given the appropriate realizability conditions, this series can be realized by a finite dimensional state-affine system [2],[21] -- that is, the recursive equations in the realization contain polynomials in $v_2(t)$. Hence state-affine equations containing polynomials in $v_2(t)$ arise in a very natural way as realizations of the finite series expansion of $\hat{y}(t|t)$.

Acknowledgement

This research was supported in part by the DoD Joint Services Electronics Program through the Air Force Office of Scientific Research (AFSC) Contract F49620-77-C-0101 and in part by the National Science Foundation under Grant ENG 76-11106. The research of the second author was supported by the Air Force of Scientific Research under contract AFOSR 77-3281.

References

1. S.I. Marcus and A.S. Willsky, "Algebraic structure and finite dimensional nonlinear estimation," SIAM J. Math. Anal., 9 (1978) 312-327.
2. S.I. Marcus, "Discrete time optimal nonlinear estimation," submitted for publication.
3. S.I. Marcus, "Estimation and analysis of nonlinear stochastic systems," Ph.D. Thesis, Dept. of Electrical Engineering, M.I.T., Cambridge, MA., June 1975.
4. N. Wiener, "The homogeneous chaos," Amer. J. Math., 60 (1938) 897-936.
5. R.H. Cameron and W.T. Martin, "The orthogonal development of nonlinear functionals in series of Fourier-Hermite functionals," Ann. Math., 48 (1947), 385-392.
6. K. Ito, "Multiple Wiener integrals," J. Math. Soc. Japan, 13 (1951) 157-169.
7. G. Kallianpur, "The role of reproducing kernel Hilbert spaces in the study of Gaussian processes," in Advances in Probability and Related Topics, P. Ney, ed., New York: Marcel Dekker, 1970, 49-83.

8. S.T. Huang and S. Cambanis, "Estimation and prediction of nonlinear functionals of Gaussian processes," Proc. 1977 Johns Hopkins Conf. on Inform. Sciences and Systems, Baltimore, Md., 140-144.
9. D.E. Gustafson, "On optimal estimation and control of linear systems with state-dependent and control-dependent noise," Ph.D. Thesis, Dept. of Aeronautics and Astronautics, M.I.T., Cambridge, MA., June 1972.
10. H.W. Sorenson and A.R. Stubberud, "Nonlinear filtering by approximation of the a posteriori density," Int. J. Control, 8 (1968), 33-51.
11. R.W. Brockett, "Volterra series and geometric control theory," Automatica, 12 (1976), 167-176.
12. E.G. Gilbert, "Bilinear and 2-power input-output maps: Finite dimensional realizations and the role of functional series," IEEE Trans. Aut. Control, AC-23 (1978), 418-425.
13. E.G. Gilbert, "Functional expansions for the response of nonlinear differential systems," IEEE Trans. Aut. Control, AC-22, (1977), 909-921.
14. M.H.A. Davis, Linear Estimation and Stochastic Control. London: Chapman and Hall, 1977.
15. M.H.A. Davis, "A direct proof of innovations/observations equivalence for Gaussian processes," IEEE Trans. Inform. Theory, IT-24 (1978), 252-254.
16. T. Kailath, "An innovations approach to least-squares estimation, part I: Linear filtering in additive white noise," IEEE Trans. Aut. Control, AC-13 (1968), 646-655.
17. J.S. Meditch, Stochastic Optimal Linear Estimation and Control. New York: McGraw-Hill, 1969.
18. C.-H. Liu, "A comparison of optimal and suboptimal estimators and estimation lower bounds," M.S. Thesis, Dept. of Electrical Engineering, University of Texas at Austin, August 1978.
19. A. Segall, "Stochastic processes in estimation theory," IEEE Trans. Inform. Theory, IT-22 (1976), 275-286.
20. P.M. Brémaud and J.H. Van Schuppen, "Discrete time processes: Parts I and II," preprint, June 1976.
21. E.D. Sontag, "On the internal realization of polynomial response maps," Ph.D. Dissertation, University of Florida, Gainesville, 1976.
22. K.S. Miller, Multidimensional Gaussian Distributions. New York: Wiley, 1965.
23. W.H. Kroy, "A higher measurement space filter for passive tracking," Proc. 3rd Symp. on Nonlinear Estimation and Its Applications, San Diego, Calif., 138-141, 1972.
24. A.V. Balakrishnan, "A general theory of nonlinear estimation problems in control systems," J. Math. Anal. Appl., 8 (1964), 4-30.