SIAM J. MATH. ANAL.
Vol. 9, No. 2, April 1978
Copyright © 1978 Society for Industrial and Applied Mathematics

ALGEBRAIC STRUCTURE AND FINITE DIMENSIONAL NONLINEAR ESTIMATION*

STEVEN I. MARCUS† AND ALAN S. WILLSKY‡

Abstract. The algebraic structure of certain classes of nonlinear systems is exploited in order to prove that the optimal estimators for these systems are recursive and finite dimensional. These systems are represented by certain Volterra series expansions or by bilinear systems with nilpotent Lie algebras. In addition, an example is presented, and the steady-state estimator for this example is discussed.

1. Introduction. Optimal recursive state estimators have been derived for very general classes of nonlinear stochastic systems [14], [7]. The optimal estimator requires, in general, an infinite dimensional computation to generate the conditional mean of the system state given the past observations. This computation involves either the solution of a stochastic partial differential equation for the conditional density or an infinite set of coupled ordinary stochastic differential equations for the conditional moments. However, the class of linear stochastic systems with linear observations and white Gaussian plant and observation noises has a particularly appealing structure, because the optimal state estimator consists of a finite dimensional linear system—the Kalman–Bucy filter [12].

In this paper we exploit the algebraic structure of certain other classes of systems, in order to prove that the optimal estimators for these systems are finite dimensional. The general class of systems is given by a linear Gauss–Markov process ξ which feeds forward into a nonlinear system with state x. Our goal is to estimate ξ and x given noisy linear observations of ξ. Specifically, consider the system

$$d\xi(t)=F(t)\xi(t)\,dt+G(t)\,dw(t), \tag{1.1}$$

$$dx(t)=a_0(x(t))\,dt+\sum_{i=1}^{N} a_i(x(t))\xi_i(t)\,dt, \tag{1.2}$$

$$dz(t)=H(t)\xi(t)\,dt+R^{1/2}(t)\,dv(t), \tag{1.3}$$

where $\xi(t)$ is an n-vector, $x(t)$ is a k-vector, $z(t)$ is a p-vector, w and v are independent standard Brownian motion processes, $R>0$, $\xi(0)$ is a Gaussian random variable independent of w and v, $x(0)$ is independent of $\xi(0)$, w, and v, and $\{a_i, i=0,\cdots,N\}$ are analytic functions of x. It will be assumed that $[F(t), G(t), H(t)]$ is completely controllable and observable. Also we define $Q(t)\triangleq G(t)G'(t)$.

The optimal estimate, with respect to a wide variety of criteria, of $x(t)$ given the observations $z^t\triangleq\{z(s), 0\leqq s\leqq t\}$, is the conditional mean $\hat{x}(t|t)$, also denoted by $E^t[x(t)]$ or $E[x(t)|z^t]$ [8] (henceforth we will freely interchange these three notations for the conditional expectation given the σ field $\sigma\{z(s), 0\leqq s\leqq t\}$ generated by the observation process up to time t). Thus our objective is the computation of $\hat{\xi}(t|t)$ and $\hat{x}(t|t)$. The computation of $\hat{\xi}(t|t)$ can be performed by the finite dimensional (linear)

* Received by the editors November 14, 1975.

† Department of Electrical Engineering, University of Texas at Austin, Austin, Texas 78712. The work of this author was supported in part by a National Science Foundation Fellowship, in part by the DoD Joint Services Electronics Program under AFSC Contract F44620-71-C-0091, and in part by the National Science Foundation under Grant ENG 76-11106.

‡ Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139. The work of this author was supported in part by the National Science Foundation under Grant GK-42090 and in part by the Air Force Office of Scientific Research under Grant 72-2273.

Kalman–Bucy filter; moreover, the conditional density of $\xi(t)$ given z^t is Gaussian with mean $\hat{\xi}(t|t)$ and nonrandom covariance $P(t)$ [12], [8]. However, the computation of $\hat{x}(t|t)$ requires in general an infinite dimensional system of equations. The purpose of this paper is to show that if $x(t)$ is characterized by a certain type of Volterra series expansion, or if $x(t)$ satisfies a certain type of bilinear equation, then $\hat{x}(t|t)$ can be computed with a *finite dimensional* nonlinear estimator.

This research is related to the recent work of Brockett [1]–[3] on algebraic and geometric methods in control theory and the work of Lo and Willsky [17], [25] on estimation for bilinear systems.

2. Volterra series and finite dimensional estimation. As shown by Brockett [2], [3] and d'Alessandro, Isidori and Ruberti [5] in the deterministic case, considerable insight can be gained by considering the Volterra series expansion of the system (1.2). The Volterra series expansion for the ith component of x is given by

$$x_i(t)=w_{0i}(t)+\sum_{j=1}^{\infty}\int_0^t\cdots\int_0^t\sum_{k_1,\cdots,k_j=1}^{n} w_{ji}^{(k_1,\cdots,k_j)}(t,\sigma_1,\cdots,\sigma_j)\cdot\xi_{k_1}(\sigma_1)\cdots\xi_{k_j}(\sigma_j)\,d\sigma_1\cdots d\sigma_j \tag{2.1}$$

where the jth *order kernel* $w_{ji}^{(k_1,\cdots,k_j)}$ is a locally bounded, piecewise continuous function. We will consider, without loss of generality [2], only *triangular kernels* which satisfy $w_{ji}^{(k_1,\cdots,k_j)}(t,\sigma_1,\cdots,\sigma_j)=0$ if $\sigma_{l+m}>\sigma_m$; $l,m=1,2,3,\cdots$. We say that a kernel $w(t,\sigma_1,\cdots,\sigma_j)$ is *separable* if it can be expressed as a finite sum

$$w(t,\sigma_1,\cdots,\sigma_j)=\sum_{i=1}^{m}\gamma_0^i(t)\gamma_1^i(\sigma_1)\gamma_2^i(\sigma_2)\cdots\gamma_j^i(\sigma_j). \tag{2.2}$$

Brockett [2] discusses the convergence of (2.1) in the deterministic case, but we will not consider this question in the general stochastic case. We will be more concerned with the case in which the linear-analytic system (1.2) has a finite Volterra series—that is, the expansion (2.1) has a finite number of terms. Brockett shows that a finite Volterra series has a bilinear realization if and only if the kernels are separable. Hence, a proof similar to that of Martin [20] of the existence and uniqueness of solutions to a bilinear system drived by the Gauss–Markov process (1.1) implies that a finite Volterra series in ξ with separable kernels is well defined in the mean-square sense.

With these preliminary concepts, the major results can be stated. The proofs are contained in this section and Appendix B; an example follows.

THEOREM 2.1. *Consider the linear system described by* (1.1) *and* (1.3), *and define the scalar-valued process*

$$x(t)=e^{\xi_i(t)}\eta(t) \tag{2.3}$$

where η is a finite Volterra series in ξ with separable kernels. Then $\hat{\eta}(t|t)$ and $\hat{x}(t|t)$ can be computed with a finite dimensional recursive system of nonlinear stochastic differential equations driven by the innovations $d\nu(t)\triangleq dz(t)-H(t)\hat{x}(t|t)\,dt$.

THEOREM 2.2. *Consider the linear system described by* (1.1) *and* (1.3), *and define the scalar-valued processes*

$$\eta(t)=\int_0^t\int_0^{\sigma_1}\cdots\int_0^{\sigma_{j-1}}\xi_{k_1}(\sigma_{m_1})\cdots\xi_{k_l}(\sigma_{m_l})\gamma_1(\sigma_1)\cdots\gamma_j(\sigma_j)\,d\sigma_1\cdots d\sigma_j, \tag{2.4}$$

$$x(t)=e^{\xi_i(t)}\eta(t) \tag{2.5}$$

where $\{\gamma_i\}$ *are deterministic functions of time and* $i>j$. *Then* $\hat{\eta}(t|t)$ *and* $\hat{x}(t|t)$ *can be computed with a finite dimensional recursive system of nonlinear stochastic differential equations driven by the innovations.*

The distinction between Theorems 2.1 and 2.2 lies in the fact that $i>j$ in (2.4)—i.e., there are more ξ_k's than integrals. On the other hand, each term in the finite Volterra series in (2.3) has $i=j$ and the σ_{m_k} are distinct. As Brockett [2] remarks, we can consider (2.4) as a single term in Volterra series if the kernel is allowed to contain impulse functions. As we will show in Lemma B.2, a term (2.4) with $i<j$ (more integrals than ξ_k's) can be rewritten as a Volterra term with $i=j$; so Theorem 2.1 also applies in this case.

The basic technique employed in the proofs of Theorems 2.1 and 2.2 is the augmentation of the state of the original system with the processes which are required in the nonlinear filtering equation (A.5)–(A.6) for $\hat{x}(t|t)$. For the classes of systems considered here, it is shown that only a finite number of additional states are required.

Proof of Theorem 2.1. We consider one term in the finite Volterra series; since the kernels are separable, we can assume without loss of generality that this term has the form

$$\eta(t)=\int_0^t\int_0^{\sigma_1}\cdots\int_0^{\sigma_{j-1}}\xi_{k_1}(\sigma_1)\cdots\xi_{k_j}(\sigma_j)\gamma_1(\sigma_1)\cdots\gamma_j(\sigma_j)\,d\sigma_1\cdots d\sigma_j. \tag{2.6}$$

The theorem is proved by induction on j, the *order* of the Volterra term (2.6). We now give the proof for $j=1$; the proof by induction is given in Appendix B. If $j=1$, then

$$\eta(t)=\int_0^t\gamma_1(\sigma_1)\xi_{k_1}(\sigma_1)\,d\sigma_1 \tag{2.7}$$

and $\eta(t)$ is *linear* function of ξ. Hence, if the state ξ of (1.1) is augmented with η, the resulting system is also linear. Then the Kalman–Bucy filter for the system described by (1.1), (1.3), and (2.7) generates $\hat{\xi}(t|t)$ and $\hat{\eta}(t|t)$. In order to prove that $\hat{x}(t|t)$ is "finite dimensionally computable" (FDC), we need the following lemma. First we define, for $\sigma_1,\sigma_2\leqq t$, the conditional cross-covariance matrix

$$P(\sigma_1,\sigma_2,t)=E[(\xi(\sigma_1)-\hat{\xi}(\sigma_1|t))(\xi(\sigma_2)-\hat{\xi}(\sigma_2|t))'|z^t] \tag{2.8}$$

(where $\hat{\xi}(\sigma|t)=E[\xi(\sigma)|z^t]$).

LEMMA 2.1. *The joint conditional density* $p_{\xi(\sigma_1),\xi(\sigma_2)}(\nu,\nu'|z^t)$ *is Gaussian with nonrandom conditional cross-covariance* $P(\sigma_1,\sigma_2,t)$*—i.e.,* $P(\sigma_1,\sigma_2,t)$ *is independent of* $\{z(s),0\leqq s\leqq t\}$.

Proof. First, the conditional density is Gaussian because ξ^t and z^t are jointly Gaussian random processes. Assume $\sigma_1>\sigma_2$; then

$$p_{\xi(\sigma_1),\xi(\sigma_2)}(\nu,\nu'|z^t)=p_{\xi(\sigma_1)}(\nu|\xi(\sigma_2)=\nu',z^t)p_{\xi(\sigma_2)}(\nu'|z^t) \tag{2.9}$$

$$=p_{\xi(\sigma_1)}(\nu|\xi(\sigma_2)=\nu',z^t_{\sigma_2})p_{\xi(\sigma_2)}(\nu'|z^t) \tag{2.10}$$

where $z^t_{\sigma_2}=\{z(s),\sigma_2\leqq s\leqq t\}$.

Here (2.9) follows by the definition of the conditional density, and (2.10) is due to the Markov property of the process (ξ,z) [8]. Each of the densities in (2.10) is the result of a linear smoothing operation; hence, each is Gaussian with nonrandom covariance $P_{\sigma_1|\sigma_2}(t)$ and $P(\sigma_2,\sigma_2,t)$, respectively [16]. Also, for $\sigma>0$, [11] $P(\sigma,\sigma,t)=[P^{-1}(\sigma)+P_B^{-1}(\sigma)]^{-1}$ where P_B is the error covariance of a Kalman filter running backward in time from t to σ, and $P_B^{-1}(t)\triangleq 0$. Due to the controllability of $[F,G]$, $P(\sigma)$

is invertible for all $\sigma>0$ and $P_B(\sigma)$ is invertible for all $\sigma<t$ [28]; consequently, $P(\sigma,\sigma,t)$ is invertible for all $0<\sigma\leqq t$. By the formula for the conditional covariance of a Gaussian distribution [8], we have for $0\leqq\sigma_1<\sigma_2\leqq t$

$$P_{\sigma_1|\sigma_2}(t)=P(\sigma_1,\sigma_1,t)-P(\sigma_1,\sigma_2,t)P^{-1}(\sigma_2,\sigma_2,t)P'(\sigma_1,\sigma_2,t). \tag{2.11}$$

Since $P(\sigma_1,\sigma_2,t)$, $0\leqq\sigma_1<\sigma_2<t$, can be computed from (2.11), it is also nonrandom; and since we have shown previously that $P(0,0,t)$ is nonrandom, $P(\sigma_1,\sigma_2,t)$ is nonrandom for all $0\leqq\sigma_1,\sigma_2\leqq t$. $\square$

This lemma allows the off-line computation of $P(\sigma_1,\sigma_2,t)$ via the equations of Kwakernaak [15] (for $\sigma_1\leqq\sigma_2$)

$$\begin{aligned} P(\sigma_1,\sigma_2,t)&=P(\sigma_1)\Psi'(\sigma_2,\sigma_1)\\ &\quad -P(\sigma_1)\left[\int_{\sigma_2}^{t}\Psi'(\tau,\sigma_1)H'(\tau)R^{-1}(\tau)H(\tau)\Psi(\tau,\sigma_2)\,d\tau\right]P(\sigma_2),\end{aligned} \tag{2.12}$$

$$\frac{d}{dt}\Psi(t,\tau)=[F(t)-P(t)H'(t)R^{-1}(t)H(t)]\Psi(t,\tau);\ \Psi(\tau,\tau)=I \tag{2.13}$$

where the Kalman filter error covariance matrix $P(t)\triangleq P(t,t,t)$ is computed via the Riccati equation

$$\begin{aligned}\dot P(t)&=F(t)P(t)+P(t)F'(t)+Q(t)-P(t)H'(t)R^{-1}(t)H(t)P(t),\\ P(0)&=P_0.\end{aligned} \tag{2.14}$$

Recall [8] that the characteristic function of a Gaussian random vector y with mean m and covariance P is given by

$$M_y(u)=E[\exp(iu'y)]=\exp[iu'm-\tfrac{1}{2}u'Pu]. \tag{2.15}$$

Hence, by taking partial derivatives of the characteristic function (see Lemma B.1), we have

$$\begin{aligned} E^t[x(t)]&=\int_0^t\gamma_1(\sigma)E^t[e^{\xi_j(t)}\xi_{k_1}(\sigma)]\,d\sigma\\ &=\int_0^t\gamma_1(\sigma)[\hat\xi_{k_1}(\sigma|t)+P_{k_1,j}(\sigma,t,t)]\;e^{\hat\xi_j(t|t)+(1/2)P_{jj}(t)}\,d\sigma,\\ &=\left\{\int_0^t\gamma_1(\sigma)P_{k_1,j}(\sigma,t,t)\,d\sigma+E^t\left[\int_0^t\gamma_1(\sigma)\xi_{k_1}(\sigma)\,d\sigma\right]\right\}\cdot e^{\hat\xi_j(t|t)+(1/2)P_{jj}(t)}\\ &=\left\{\int_0^t\gamma_1(\sigma)P_{k_1,j}(\sigma,t,t)\,d\sigma+\hat\eta(t|t)\right\}e^{\hat\xi_j(t|t)+(1/2)P_{jj}(t)}.\end{aligned} \tag{2.16}$$

Since the first term in (2.16) is nonrandom and $\hat\eta(t|t)$ and $\hat\xi(t|t)$ can be computed with a Kalman–Bucy filter, $\hat x(t|t)$ is indeed FDC for the case $j=1$.

The induction step of the proof of Theorem 2.1 is given in Appendix B. A crucial component of the proof is Lemma B.1, which expresses higher order moments of a Gaussian distribution in terms of the lower moments. Notice that in equation (2.16) we have interchanged the operations of integration and conditional expectation. This is justified by the version of the Fubini theorem proved in [18]; since we will be dealing only with integrals of products of Gaussian random processes, the use of the Fubini theorem is easily justified, and we will use it without further comment.

The proof of Theorem 2.2 is almost identical to that of Theorem 2.1; the differences are explained in Appendix B. We now present an example to illustrate the basic concepts of these theorems; this example is a special case of Theorem 2.2. However, we will need one preliminary lemma.

LEMMA 2.2. *The conditional cross-covariance satisfies*

$$P(\sigma, t, t) = K(t, \sigma)P(t) \tag{2.17}$$

where

$$\frac{d}{dt}K'(t, \sigma) = -[F'(t) + P^{-1}(t)Q(t)]K'(t, \sigma); \; K'(\sigma, \sigma) = I. \tag{2.18}$$

Proof. Let

$$\tilde{P}(\sigma, t) \triangleq E[(\xi(\sigma) - \hat{\xi}(\sigma|\sigma))(\xi(t) - \hat{\xi}(t|t))']$$

and consider

$$P(\sigma, t, t) - \tilde{P}(\sigma, t) = E[(\hat{\xi}(\sigma|\sigma) - \hat{\xi}(\sigma|t))(\xi(t) - \hat{\xi}(t|t))'|z^t].$$

Since $\hat{\xi}(\sigma|\sigma) - \hat{\xi}(\sigma|t)$ is measurable with respect to the σ-field $\sigma(z^t)$, the projection theorem [22] implies that $P(\sigma, t, t) - \tilde{P}(\sigma, t) = 0$. The proof is concluded by noting that [11] $\tilde{P}(\sigma, t) = K(t, \sigma)P(t)$. □

Example 2.1. Consider the system described by

$$\begin{bmatrix} d\xi_1(t) \\ d\xi_2(t) \end{bmatrix} = \begin{bmatrix} -\alpha & 0 \\ 0 & -\beta \end{bmatrix} \begin{bmatrix} \xi_1(t) \\ \xi_2(t) \end{bmatrix} dt + \begin{bmatrix} dw_1(t) \\ dw_2(t) \end{bmatrix}, \tag{2.19}$$

$$dx(t) = (-\gamma x(t) + \xi_1(t)\xi_2(t))\, dt, \tag{2.20}$$

$$\begin{bmatrix} dz_1(t) \\ dz_2(t) \end{bmatrix} = \begin{bmatrix} \xi_1(t) \\ \xi_2(t) \end{bmatrix} dt + \begin{bmatrix} dv_1(t) \\ dv_2(t) \end{bmatrix} \tag{2.21}$$

where $\alpha, \beta, \gamma > 0$, w_1, w_2, v_1, and v_2 are independent, zero mean, unit variance Wiener processes, $\xi_1(0)$ and $\xi_2(0)$ are independent Gaussian random variables which are also independent of the noise processes, and $x(0) = 0$.

The conditional expectation $\hat{x}(t|t)$ satisfies the nonlinear filtering equation (A.5)–(A.6):

$$\begin{aligned} d\hat{x}(t|t) = E^t[-\gamma x(t) + \xi_1(t)\xi_2(t)]\, dt + \Big\{ & E^t\Big[\int_0^t e^{-\gamma(t-s)}\xi_1(s)\xi_2(s)\, ds \cdot \xi'(t)\Big] \\ & - E^t\Big[\int_0^t e^{-\gamma(t-s)}\xi_1(s)\xi_2(s)\, ds\Big]\hat{\xi}'(t|t)\Big\}\, d\nu(t) \end{aligned} \tag{2.22}$$

where $\xi(t) = [\xi_1(t), \xi_2(t)]'$ and the innovations process ν is given by

$$d\nu(t) = dz(t) - \hat{\xi}(t|t)\, dt. \tag{2.23}$$

Recall that the conditional covariance $P(t)$ of $\xi(t)$ given z^t satisfies the Riccati equation (2.14). Since $\xi_1(0)$ and $\xi_2(0)$ are independent, it is not difficult to show that $P_{12}(t) = P_{21}(t) = 0$ for all t. From (2.17)–(2.18) we can compute

$$P(\sigma, t, t) = \begin{bmatrix} P_{11}(t)\exp[\alpha(t-\sigma) - \int_\sigma^t P_{11}^{-1}(s)\, ds] & 0 \\ 0 & P_{22}(t)\exp[\beta(t-\sigma) - \int_\sigma^t P_{22}^{-1}(s)\, ds] \end{bmatrix}. \tag{2.24}$$

These facts and equation (B.3a) imply that the transpose of the *gain term* in (2.22) is

$$E^t\left[\int_0^t e^{-\gamma(t-s)}\xi_1(s)\xi_2(s)\xi(t)\,ds\right]-E^t\left[\int_0^t e^{-\gamma(t-s)}\xi_1(s)\xi_2(s)\,ds\right]\hat{\xi}(t|t)$$

$$=\int_0^t e^{-\gamma(t-s)}(E^t[\xi_1(s)\xi_2(s)\xi(t)]-E^t[\xi_1(s)\xi_2(s)]E^t[\xi(t)])\,ds$$

$$=E^t\left\{\int_0^t e^{-\gamma(t-s)}\begin{bmatrix}0 & P_{11}(s,t,t)\\ P_{22}(s,t,t) & 0\end{bmatrix}\begin{bmatrix}\xi_1(s)\\ \xi_2(s)\end{bmatrix}ds\right\} \tag{2.25a}$$

$$=E^t\begin{bmatrix}\eta_1(t)P_{11}(t)\\ \eta_2(t)P_{22}(t)\end{bmatrix} \tag{2.25b}$$

where

$$\begin{bmatrix}\dot{\eta}_1(t)\\ \dot{\eta}_2(t)\end{bmatrix}=\begin{bmatrix}\alpha-\gamma-P_{11}^{-1}(t) & 0\\ 0 & \beta-\gamma-P_{22}^{-1}(t)\end{bmatrix}\begin{bmatrix}\eta_1(t)\\ \eta_2(t)\end{bmatrix}+\begin{bmatrix}0 & 1\\ 1 & 0\end{bmatrix}\begin{bmatrix}\xi_1(t)\\ \xi_2(t)\end{bmatrix}, \tag{2.26}$$
$$\eta_1(0)=\eta_2(0)=0.$$

In other words, the argument of the conditional expectation in (2.25a) can be realized as the output of a finite dimensional linear system with state $\eta(t)=[\eta_1(t),\ \eta_2(t)]'$ satisfying (2.26).

Thus the finite dimensional optimal estimator for the system (2.19)–(2.21) is constructed as follows (see Fig. 1). First we augment the state ξ of (2.19) with the state η of (2.26). Then the Kalman–Bucy filter for the linear system (2.19), (2.26), with observations (2.21), computes the conditional expectations $\hat{\xi}(t|t)$ and $\hat{\eta}(t|t)$. Finally,

$$d\hat{x}(t|t)=[-\gamma\hat{x}(t|t)+\hat{\xi}_1(t|t)\hat{\xi}_2(t|t)]\,dt+\hat{\eta}'(t|t)P(t)\,d\nu(t), \tag{2.27}$$
$$\hat{x}(0|0)=0.$$

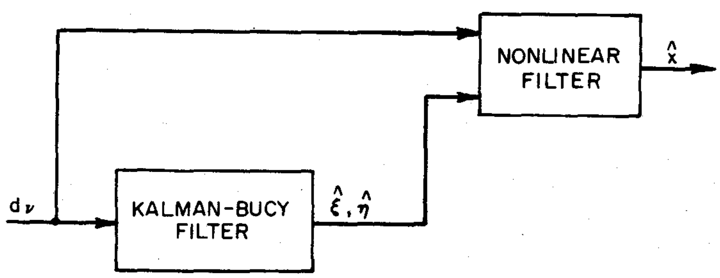

FIG. 1. *Block diagram of the optimal filter for Example* 2.1.

We now discuss the steady-state behavior of the optimal filter. Since the linear system (2.19) is asymptotically stable (and hence detectable) and controllable, the Riccati equation (2.14) has a unique positive-definite steady-state solution P [28]; a simple computation shows that

$$P=\begin{bmatrix}P_{11} & 0\\ 0 & P_{22}\end{bmatrix}=\begin{bmatrix}-\alpha+\sqrt{\alpha^2+1} & 0\\ 0 & -\beta+\sqrt{\beta^2+1}\end{bmatrix}. \tag{2.28}$$

Thus, in steady-state, the augmented linear system (2.19), (2.26) is time-invariant. Now consider the eigenvalues of (2.26) in steady-state:

$$\alpha-\gamma-P_{11}^{-1}=\alpha-\gamma-(-\alpha+\sqrt{\alpha^2+1})^{-1}=-\gamma-\sqrt{\alpha^2+1},$$
$$\beta-\gamma-P_{22}^{-1}=\beta-\gamma-(-\beta+\sqrt{\beta^2+1})^{-1}=-\gamma-\sqrt{\beta^2+1}.$$

Consequently, the augmented linear system is also asymptotically stable and controllable in steady-state. Let the conditional covariance matrix of the augmented state $[\xi(t), \eta(t)]$ given z^t be denoted by $S(t)$. Then the Riccati equation satisfied by $S(t)$ has a unique positive-definite steady-state solution S (notice that $S_{11}=P_{11}$ and $S_{22}=P_{22}$).

The steady-state Kalman–Bucy filter [8] for the augmented system (2.19), (2.26) is easily computed to be

$$\begin{bmatrix} d\hat{\xi}_1(t|t)\\ d\hat{\xi}_2(t|t)\\ d\hat{\eta}_1(t|t)\\ d\hat{\eta}_2(t|t)\end{bmatrix}=\begin{bmatrix}-\alpha & 0 & 0 & 0\\ 0 & -\beta & 0 & 0\\ 0 & 1 & -\gamma-\sqrt{\alpha^2+1} & 0\\ 1 & 0 & 0 & -\gamma-\sqrt{\beta^2+1}\end{bmatrix}\begin{bmatrix}\hat{\xi}_1(t|t)\\ \hat{\xi}_2(t|t)\\ \hat{\eta}_1(t|t)\\ \hat{\eta}_2(t|t)\end{bmatrix}dt$$
$$(2.29)\qquad +\begin{bmatrix}P_{11} & 0\\ 0 & P_{22}\\ 0 & S_{23}\\ S_{14} & 0\end{bmatrix}\begin{bmatrix}d\nu_1(t)\\ d\nu_2(t)\end{bmatrix}$$

where

$$S_{14}=\frac{P_{11}P_{22}}{P_{11}P_{22}+(\alpha-\beta+\gamma)P_{22}+1},\qquad S_{23}=\frac{P_{11}P_{22}}{P_{11}P_{22}+(\beta-\alpha+\gamma)P_{11}+1}$$

(here P_{11} and P_{22} are defined in (2.28)). The conditional expectation $\hat{x}(t|t)$ is computed according to

$$(2.30)\qquad \begin{aligned} &d\hat{x}(t|t)=[-\gamma\hat{x}(t|t)+\hat{\xi}_1(t|t)\hat{\xi}_2(t|t)]\,dt+\hat{\eta}'(t|t)\,P\,d\nu(t),\\ &\hat{x}(0|0)=0\end{aligned}$$

which is a nonlinear, time-invariant equation.

We note that the stability of the original linear system is not necessary for the existence of the steady-state optimal filter in this example; in fact, a weaker sufficient condition is the detectability [28] of the linear system (2.19), (2.21) and the positivity of γ in (2.20). The generalization of this result to other systems is presently being investigated.

3. Finite dimensional estimators for bilinear systems. In this section the results of the previous section are applied, with the aid of some concepts from the theory of Lie algebra [23], to prove that the optimal estimators for certain bilinear systems are finite dimensional. Consider the system described by (1.1), (1.3), and the bilinear system [1], [10]

$$(3.1)\qquad \dot{X}(t)=\left(A_0+\sum_{i=1}^{N}\xi_i(t)A_i\right)X(t);\qquad X(0)=I$$

where X is a $k\times k$ matrix. We associate with (3.1) the Lie algebra $\mathscr{L}\triangleq\{A_0, A_1, \ldots, A_N\}_{LA}$, the smallest Lie algebra containing $A_0, A_1, \ldots, A_N$; the ideal $\mathscr{L}_0$ in $\mathscr{L}$ generated by $\{A_1, \ldots, A_N\}$; the group $G\triangleq\{\exp\mathscr{L}\}_G$, the smallest group generated by $\{\exp A\}$ for all $A\in\mathscr{L}$; and the subgroup $G_0\triangleq\{\exp\mathscr{L}_0\}_G$ [10], [18], [19], [26], [27].

DEFINITION 3.1 [23]. A Lie algebra $\mathscr{L}$ is *solvable* if the *derived series* of ideals

$$\mathscr{L}^{(0)}=\mathscr{L},$$
$$\mathscr{L}^{(n+1)}=[\mathscr{L}^{(n)},\mathscr{L}^{(n)}]=\{[A,B]\,|\,A,B\in\mathscr{L}^{(n)}\},\qquad n\geqq 0,$$

is the trivial ideal $\{0\}$ for some n. $\mathscr{L}$ is *nilpotent* if the *lower central* series of ideals

$$\mathscr{L}^{0}=\mathscr{L},$$
$$\mathscr{L}^{n+1}=[\mathscr{L},\mathscr{L}^{n}]=\{[A,B]\,|\,A\in\mathscr{L}^{n}\},\qquad n\geqq 0,$$

is $\{0\}$ for some n. $\mathscr{L}$ is Abelian if $\mathscr{L}^{(1)}=\mathscr{L}^1=\{0\}$. Note that Abelian$\Rightarrow$nilpotent$\Rightarrow$ solvable, but none of the reverse implications hold in general.

A useful structural result for nilpotent Lie algebras is presented in the following lemma [23, p. 224].

LEMMA 3.1. *A matrix Lie algebra $\mathscr{L}$ is nilpotent if and only if there exists a (possibly complex-valued) nonsingular matrix P such that, for all $A\in\mathscr{L}$, PAP^{-1} has the block diagonal form*

$$\begin{bmatrix} \begin{bmatrix}\Phi_1(A) & & *\\ 0 & \ddots & \Phi_1(A)\end{bmatrix} & & & 0\\ & \begin{bmatrix}\Phi_2(A) & & *\\ & \ddots & \Phi_2(A)\end{bmatrix} & \\ 0 & & & \ddots \end{bmatrix}\tag{3.2}$$

(this will be called the *nilpotent canonical form*). *The functions $\Phi_k\colon\mathscr{L}\to C$ are linear. Furthermore, $\Phi_k([\mathscr{L},\mathscr{L}])=\{0\}$.*

It is easy to show, using Brockett's results [2] on finite Volterra series, that each term in (2.3) can be realized by a bilinear system of the form

$$\dot{x}(t)=\xi_i(t)x(t)+\sum_{k=1}^{n}A_k(t)\xi_k(t)x(t)\tag{3.3}$$

where x is a k-vector and the A_i are strictly upper triangular (zero on and below the main diagonal). For such systems, the Lie algebra $\mathscr{L}_0$ is nilpotent. In this section we will show conversely that if the Lie algebra $\mathscr{L}_0$ corresponding to the bilinear system (3.1) is nilpotent, then each component of the solution to (3.1) can be written as a finite sum of terms of the form (2.3). Hence, such systems also have finite dimensional estimators; this result is summarized in the next theorem.

THEOREM 3.1. *Consider the system described by* (1.1), (1.3), *and* (3.1) *and assume that $\mathscr{L}_0$ is a nilpotent Lie algebra. Then the conditional expectation $\hat{X}(t|t)$ can be computed with a finite dimensional system of nonlinear differential equations driven by the innovations.*

Remarks. (i) It can easily be shown that if $\mathscr{L}_0$ is nilpotent, then $\mathscr{L}$ is solvable; however, the converse is not true. Hence, $\mathscr{L}$ is always solvable in Theorem 3.1.

(ii) Theorem 3.1 provides a generalization of the work of Lo and Willsky [17] (in which $\mathscr{L}$ is Abelian) and Willsky [25]. The Abelian discrete-time problem is also considered by Johnson and Stear [9].

(iii) The model considered in Theorem 3.1 is motivated by a problem in strapdown inertial navigation [18], [26]. However, in the navigation problem $\mathcal{L}_0$ is not nilpotent (in fact, $\mathcal{L} = so(3)$ is simple [23]), so Theorem 3.1 does not apply.

(iv) Using the notation of Brockett [2], it is easily seen that the pth order moments $X^{[p]}(t)$ satisfy an equation of the form (3.1) (with different coefficient matrices $A_{i_{[p]}}$), and hence $X^{[p]}(t|t)$ can also be computed with a finite dimensional estimator. In particular, the performance of the estimator of Theorem 3.1 can be evaluated by computing the conditional covariance of $X(t)$ given z^t in this manner.

Theorem 3.1 is proved via a series of lemmas which reduce the estimation problem to the case in which $\mathcal{L}$ is a particular nilpotent Lie algebra. The first lemma generalizes a result of Willsky [25], Brockett [1], and Krener [13] (the proof is analogous and will be omitted).

LEMMA 3.2. *Consider the system described by* (1.1), (1.3), *and* (3.1) *and define the* $k \times k$ *matrix-valued process*

$$Y(t) = e^{-A_0 t} X(t). \tag{3.4}$$

Then there exists a deterministic matrix-valued function $D(t)$ *such that* Y *satisfies*

$$\dot{Y}(t) = \left[\sum_{i=1}^{M} H_i y_i(t)\right] Y(t); \qquad Y(0) = I \tag{3.5}$$

where $\{H_1, \cdots, H_M\}$ *is a basis for* $\mathcal{L}_0$ *and*

$$y(t) = D(t)\xi(t). \tag{3.6}$$

In addition, $\hat{X}$ *can be computed according to*

$$\hat{X}(t|t) = e^{A_0 t}\hat{Y}(t|t). \tag{3.7}$$

Lemma 3.2 enables us, without loss of generality, to examine the estimation problem for $Y(t)$ evolving on the subgroup $G_0 = \{\exp \mathcal{L}_0\}_G$, rather than for $X(t)$ evolving on the full Lie group $G = \{\exp \mathcal{L}\}_G$. Thus, we need only consider the case in which $A_0 = 0$ and $\mathcal{L} = \mathcal{L}_0$ is nilpotent in order to prove Theorem 3.1.

By means of Lemma 3.1 the problem can be further reduced to the consideration of Lie algebras in nilpotent canonical form.

LEMMA 3.3. *Consider the system described by* (1.1), (1.3), *and* (3.1), *where* $A_0 = 0$ *and* $\mathcal{L}$ *is nilpotent. Then there exists a (possibly complex-valued) nonsingular matrix* P *such that*

$$\hat{X}(t|t) = P^{-1}\hat{Y}(t|t)P \tag{3.8}$$

where Y *satisfies* (3.5) *and* [$\{H_1, \cdots, H_M\}$ *are in nilpotent canonical form.*

Proof. According to Lemma 3.1, there exists a nonsingular matrix P such that $P\mathcal{L}P^{-1}$ is in nilpotent canonical form. If we define $H_i = PA_iP^{-1}$, then $X(t) = P^{-1}Y(t)P$, where Y satisfies (3.5). Hence, $\hat{X}(t|t) = P\hat{Y}(t|t)P^{-1}$ and the lemma is proved. □

Finally, by means of the following trivial lemma, we reduce the problem to the consideration of one block in the nilpotent canonical form.

LEMMA 3.4. *Consider the system described by* (1.1), (1.3), *and* (3.1), *where* $A_0 = 0$ *and* $\{A_1, \cdots, A_N\}$ *are in nilpotent canonical form. Then* $X(t)$ *has a block diagonal form conformable with that of* $\{A_1, \cdots, A_N\}$.

Let $gn(m)$ denote the Lie algebra of upper triangular $m \times m$ matrices with equal diagonal elements. Then Lemma 3.4 implies that the bilinear system (3.1) can be viewed as the "direct sum" of a number of decoupled k_j-dimensional subsystems

$$\dot{X}^j(t) = \left[\sum_{i=1}^{N} \xi_i(t) A_i^j\right] X^j(t); \qquad X^j(0) = I \tag{3.9}$$

where $A_1^j, \cdots, A_N^j$ belong to $gn(k_j)$. Hence, Theorem 3.1 will be established when we prove the following lemma.

LEMMA 3.5. *Consider the system described by* (1.1), (1.3), *and* (3.1), *where* $A_0 = 0$ *and* $\{A_1, \cdots, A_N\} \in gn(k)$. *Then each element of the solution* $X(t)$ *of* (3.1) *can be expressed in the form*

$$\exp\left(\sum_{i=1}^{N} \alpha_i \int_0^t \xi_i(s)\, ds\right) \eta(t) \tag{3.10}$$

where η *is a finite Volterra series in* ξ *with separable kernels. Hence, Theorem* 2.1 *implies that* $\hat{X}(t|t)$ *can be computed with a finite dimensional system of nonlinear stochastic differential equations.*

Proof. Since $\{A_1, \cdots, A_N\} \in gn(k)$, the bilinear equation (3.1) can be rewritten in the form

$$\dot{X}(t) = \left[\left(\sum_{i=1}^{N} \alpha_i \xi_i(t)\right) I + \sum_{i=1}^{N} \xi_i(t) B_i\right] X(t) \tag{3.11}$$

where α_i are constants, I denotes the $k \times k$ identity matrix, and $B_1, \cdots, B_N$ are *strictly* upper triangular (zero on the diagonal). It is easy to show that

$$X(t) = \exp\left(\sum_{i=1}^{N} \alpha_i \int_0^t \xi_i(s)\, ds\right) Y(t)$$

where Y satisfies

$$\dot{Y}(t) = \left[\sum_{i=1}^{N} \xi_i(t) B_i\right] Y(t); \qquad Y(0) = I. \tag{3.12}$$

Since the $\{B_i\}$ are strictly upper triangular, the solution of (3.12) can be written as a finite Peano–Baker (Volterra) series [2], and each element of $X(t)$ can be expressed in the form (3.10). □

4. Conclusions. It is shown in [18] that if $\mathcal{L}_0$ is not nilpotent, then the optimal estimator for (1.1), (1.3), and (3.1) is infinite dimensional. Thus, the results of this paper cannot be generalized to much larger classes of systems.

However, the papers of Fliess [6] and Sussmann [24] show that, in the deterministic case with bounded inputs, any causal and continuous input-output map on a finite interval can be uniformly approximated by a bilinear system of the form (3.1) in which $A_0, A_1, \cdots, A_N$ are all strictly upper triangular. For such a bilinear system both $\mathcal{L}_0$ and $\mathcal{L}$ are nilpotent Lie algebras. Stochastic analogues of this result are currently being investigated. The implication of such a result would be that suboptimal estimators for a large class of nonlinear stochastic systems could be constructed using the results of this paper.

Appendix A. General nonlinear filtering equations. In this Appendix we state some results on nonlinear filtering [7], [8], [14]. Consider a model in which the state evolves according to the Ito stochastic differential equation

$$dx(t) = f(x(t), t)\, dt + G(x(t), t)\, dw(t) \tag{A.1}$$

and the observed process is the solution of the vector Ito equation

$$dz(t) = h(x(t), t)\,dt + R^{1/2}(t)\,dv(t). \tag{A.2}$$

Here $x(t)$ is an n-vector, $z(t)$ is a p-vector, $R^{1/2}$ is the unique positive definite square root of the positive definite matrix R, and v and w are independent Brownian motion (Wiener) processes such that

$$E[w(t)w'(s)] = \int_0^{\min(t,s)} Q(\tau)\,d\tau, \tag{A.3}$$

$$E[v(t)v'(s)] = \min(t, s) \cdot I. \tag{A.4}$$

For any integrable random process $\alpha(t)$, we denote $E(\alpha(t)|z(s),\ 0 \leqq s \leqq t)$ by $\hat{\alpha}(t|t)$ or $E^t[\alpha(t)]$. Then, [7], [8], [14], the conditional mean $\hat{x}(t|t)$ satisfies

$$d\hat{x}(t|t) = E^t[f(x(t), t)]\,dt + \{E^t[x(t)h'(x(t), t)] - \hat{x}(t|t)E^t[h'(x(t), t)]\}R^{-1}(t)\,d\nu(t) \tag{A.5}$$

where the *innovations process* ν is defined by

$$d\nu(t) = dz(t) - E^t[h(x(t), t)]\,dt. \tag{A.6}$$

Appendix B. Proofs of Theorems 2.1 and 2.2.

B.1. Preliminary results. In this section we present some preliminary results which are crucial in the proofs of Theorems 2.1 and 2.2. The first lemma follows easily from some identities of Miller [21].

LEMMA B.1. *Let $x = [x_1, \cdots, x_k]'$ be a Gaussian random vector with mean m, covariance matrix P, and characteristic function M_x. Then, if $l \leqq k$,*

$$\begin{aligned}\frac{\partial^l}{\partial u_1 \cdots \partial u_l} M_x(u_1, \cdots, u_k) = \Big\{&\varepsilon_1 \cdots \varepsilon_l - \sum P_{j_1 j_2}\varepsilon_{j_3} \cdots \varepsilon_{j_l} \\ &+ \sum P_{j_1 j_2}P_{j_3 j_4}\varepsilon_{j_5} \cdots \varepsilon_{j_l} - \cdots\Big\} M_x(u_1, \cdots, u_k)\end{aligned} \tag{B.1}$$

where

$$\varepsilon_j = im_j - \sum_{n=1}^{k} u_n P_{jn} \tag{B.2}$$

and the sums in (B.1) are over all possible combinations of pairs of the $\{j_i,\ i = 1, \cdots, l\}$. Also,

$$E[x_1 x_2 \cdots x_k] = E[x_k]E[x_1 x_2 \cdots x_{k-1}] + \sum_{j_1=1}^{k-1} P_{kj_1}E[x_{j_2}x_{j_3} \cdots x_{j_{k-1}}] \tag{B.3a}$$

$$\begin{aligned}&= E[x_1 \cdots x_i]E[x_{i+1} \cdots x_k] + \sum P_{j_1 l_{i+1}}E[x_{j_2} \cdots x_{j_i}]E[x_{l_{i+2}} \cdots x_{l_k}] \\ &\quad + \sum P_{j_1 l_{i+1}}P_{j_2 l_{i+2}}E[x_{j_3} \cdots x_{j_i}]E[x_{l_{i+3}} \cdots x_{l_k}] + \cdots\end{aligned} \tag{B.3b}$$

$$\begin{aligned}&= m_1 \cdots m_k + \sum P_{j_1 j_2}m_{j_3} \cdots m_{j_k} \\ &\quad + \sum P_{j_1 j_2}P_{j_3 j_4}m_{j_5} \cdots m_{j_k} + \cdots\end{aligned} \tag{B.3c}$$

where the sums in (B.3b, c) are defined as in (B.1); also, in (B.3b), $\{j_\alpha,\ \alpha = 1, \cdots, i\}$ is a permutation of $\{1, \cdots, i\}$ and $\{l_\alpha,\ \alpha = i+1, \cdots, k\}$ is a permutation of $\{i+1, \cdots, k\}$.

In the remainder of this Appendix it will be assumed that ξ and z are Gauss–Markov processes satisfying (1.1) and (1.3), respectively. We now define classes of random processes which occur as the jth order term in a Volterra series expansion in ξ

with separable kernels, and we prove some lemmas relating these to other relevant processes.

DEFINITION B.1. The space Λ_j of *Volterra terms of order* j is the vector space over R consisting of all scalar-valued random processes λ_j of the form

$$\lambda_j(t)=\sum_{i=1}^{N}\gamma_0^i(t)\lambda_j^i(t) \tag{B.4}$$

where

$$\lambda_j^i(t)=\int_0^t\int_0^{\sigma_1}\cdots\int_0^{\sigma_{j-1}}\gamma_1^i(\sigma_1)\cdots\gamma_j^i(\sigma_j)\xi_{k_{1,i}}(\sigma_1)\cdots\xi_{k_{j,i}}(\sigma_j)\,d\sigma_1\cdots d\sigma_j \tag{B.5}$$

where for each i, $\{\xi_{k_{1,i}},\cdots,\xi_{k_{j,i}}\}$ are not necessarily distinct elements of ξ, and $\{\gamma_l^i\}$ are locally bounded, piecewise continuous, deterministic functions of time. We denote by $\hat{\Lambda}_j$ the space of all processes

$$\hat{\lambda}_j(t|t)\triangleq E[\lambda_j(t)|z^t],\qquad \text{where } \lambda_j\in\Lambda_j.$$

The next lemma, which is due to Brockett [4], shows that terms of the form (2.4) with $i<j$ (more integrals than ξ_k's) are in fact elements of Λ_i.

LEMMA B.2. *Let ξ satisfy* (1.1) *and consider the scalar-valued process*

$$\eta(t)=\int_0^t\int_0^{\sigma_1}\cdots\int_0^{\sigma_{j-1}}\gamma_1(\sigma_1)\cdots\gamma_j(\sigma_j)\xi_{k_1}(\sigma_{m_1})\cdots\xi_{k_i}(\sigma_{m_i})\,d\sigma_1\cdots d\sigma_j \tag{B.6}$$

where γ_i are as in Definition B.1, *$m_n\neq m_l$ for $n\neq l$, and $i<j$. Then $\eta\in\Lambda_i$.*

Proof. It is easy to show using the construction of Brockett [2, Thm. 4] that $\eta(t)$ has a realization as a time-varying bilinear system

$$\dot{x}(t)=A(t)x(t)+\sum_{l=1}^{i}\xi_{k_l}(t)B_l(t)x(t), \tag{B.7}$$

$$\eta(t)=x_1(t) \tag{B.8}$$

where $A(t)$ and $\{B_l(t)\}$ are strictly upper triangular matrices. The Volterra series for (B.7) can be expressed via the Peano–Baker series [2], and the Volterra series is finite because $A(t)$ and $\{B_l(t)\}$ are upper triangular. In fact, because the original expression (B.6) contains only the product of i components of ξ, the Volterra expansion of $\eta(t)=x_1(t)$ will contain *only* an ith order term

$$\eta(t)=\int_0^t\int_0^{\sigma_1}\cdots\int_0^{\sigma_{i-1}}\left[\sum_{l=1}^{m}\gamma_1^l(\sigma_1)\cdots\gamma_i^l(\sigma_i)\right]\xi_{n_1}(\sigma_1)\cdots\xi_{n_i}(\sigma_i)\,d\sigma_1\cdots d\sigma_i \tag{B.9}$$

where $\{n_l, l=1,\cdots,i\}$ is a permutation of the $\{k_l, l=1,\cdots,i\}$ of (B.6). Hence $\eta\in\Lambda_i$. □

Recall that the conditional cross-covariance $P(\sigma_1,\sigma_2,t)$ (defined in (2.8)) was shown to be nonrandom in Lemma 2.1; it can be computed from Kwakernaak's equations (2.12)–(2.14). The following lemma shows that $P_{ij}(\sigma_1,\sigma_2,t)$ is a separable kernel.

LEMMA B.3. *$P_{ij}(\sigma_1,\sigma_2,$ t) is a separable kernel; i.e., it can be expressed in the form*

$$P_{ij}(\sigma_1,\sigma_2,t)=\sum_{k=1}^{m}\gamma_0^k(t)\gamma_1^k(\sigma_1)\gamma_2^k(\sigma_2). \tag{B.10}$$

Proof. Assume $\sigma_1 \leqq \sigma_2 \leqq t$. Then it follows from (2.12) that, for arbitrary real numbers α, β, and δ,

$$P(\sigma_1, \sigma_2, t) = P(\sigma_1)\Psi'(\alpha, \sigma_1)\Big[\Psi'(\sigma_2, \alpha) - \int_{\sigma_2}^{\beta} \Psi'(\tau, \alpha)H'(\tau)R^{-1}(\tau)H(\tau)\Psi(\tau, \sigma_2)\, d\tau \cdot P(\sigma_2)$$

$$- \int_{\beta}^{t} \Psi'(\tau, \alpha)H'(\tau)R^{-1}(\tau)H(\tau)\Psi(\tau, \delta)\, d\tau \cdot \Psi(\delta, \sigma_2)P(\sigma_2)\Big] \tag{B.11}$$

$$\triangleq A(\sigma_1)[B(\sigma_2) + C(t)D(\sigma_2)].$$

Hence, if e_i denotes the ith unit vector in R^n, it is obvious from (B.11) that

$$P_{ij}(\sigma_1, \sigma_2, t) = e_i' P(\sigma_1, \sigma_2, t)e_j \tag{B.12}$$

has the form (B.10) for some functions $\{\gamma_i^m(t)\}$. □

The next lemma proves that certain processes which occur in the proof of Theorem 2.1 are elements of Λ_j.

LEMMA B.4. *Let ξ satisfy* (1.1), *and consider the scalar-valued process*

$$\eta(t) = \int_0^t \int_0^{\sigma_1} \int_0^{\sigma_{j-1}} P_{n_1 n_2}(\sigma_{m_1}, \sigma_{m_2}, t) \cdots P_{n_{l-1} n_l}(\sigma_{m_{l-1}}, \sigma_{m_l}, t) \cdot \gamma_1(\sigma_1) \cdots \gamma_j(\sigma_j)\xi_{k_1}(\sigma_1) \cdots \xi_{k_j}(\sigma_j)\, d\sigma_1 \cdots d\sigma_j \tag{B.13}$$

where the m_i are arbitrary integers in $\{1, \cdots, i\}$ *and $P_{n_{i_1} n_{i_2}}$ are arbitrary elements of P. Then $\eta \in \Lambda_j$.*

Proof. Since we have shown in Lemma B.3 that $P_{n_{i_1} n_{i_2}}(\sigma_{m_{i_1}}, \sigma_{m_{i_2}}, t)$ is a separable kernel, the kernel of the integral (B.13) is also a separable kernel. Hence, $\eta \in \Lambda_j$. □

Lemma B.4 implies that if $\hat{\lambda}_j(t|t)$ can be computed with a finite dimensional estimator for all $\lambda_j \in \Lambda_j$, then $\hat{\eta}(t|t)$ where η is defined by (B.13)) is also "*finite dimensionally computable*" (FDC).

B.2. Proofs of Theorems 2.1 and 2.2. The proofs of these two theorems are almost identical. We will prove Theorem 2.1; then we will explain how this proof is modified to prove Theorem 2.2.

Proof of Theorem 2.1. As stated in § 2, we consider the jth order Volterra term

$$\eta(t) = \int_0^t \int_0^{\sigma_1} \int_0^{\sigma_{j-1}} \gamma_1(\sigma_1) \cdots \gamma_j(\sigma_j) \cdot \xi_{k_1}(\sigma_1) \cdots \xi_{k_j}(\sigma_j)\, d\sigma_1 \cdots d\sigma_j. \tag{B.14}$$

The theorem is proved by induction on j, the order of the Volterra term. The proof for $j = 1$ is presented in § 2. We now assume the theorem holds for $j \leqq i - 1$ (i.e., we assume that $E^t[e^{\xi_l(t)}\eta(t)]$ is FDC, where $\eta \in \Lambda_j$, for $j \leqq i-1$), and prove that it holds for $j = i$.

The proof is in two steps. We first reduce the problem to the computation of the elements of $\hat{\Lambda}_i$ (see Definition B.1). We then show by induction that all of the processes in $\hat{\Lambda}_i$ can be computed with finite dimensional estimators.

(i) We first consider the computation of $\hat{x}(t|t)$, where

$$x(t) = e^{\xi_l(t)}\eta(t). \tag{B.15}$$

Now

$$\hat{x}(t|t) = \int_0^t \int_0^{\sigma_1} \int_0^{\sigma_{i-1}} \gamma_1(\sigma_1) \cdots \gamma_i(\sigma_i) \cdot E^t[e^{\xi_l(t)}\xi_{k_1}(\sigma_1) \cdots \xi_{k_i}(\sigma_i)]\, d\sigma_1 \cdots d\sigma_i. \tag{B.16}$$

By (B.1) and the definition of the characteristic function, it follows that

$$
\begin{aligned}
&E^t[e^{\xi_l(t)}\xi_{k_1}(\sigma_1)\cdots\xi_{k_i}(\sigma_i)] \\
\text{(B.17)} \qquad &= e^{\hat{\xi}_l(t|t)+(1/2)P_{ll}(t)}\{\delta_1(\sigma_1)\cdots\delta_i(\sigma_i) \\
&\quad+\sum P_{j_1j_2}(\sigma_{m_1},\sigma_{m_2},t)\delta_{j_3}(\sigma_{m_3})\cdots\delta_{j_i}(\sigma_{m_i})+\cdots\}
\end{aligned}
$$

where

$$\text{(B.18)} \qquad \delta_{j_\alpha}(\sigma_{m_\alpha})=\hat{\xi}_{j_\alpha}(\sigma_{m_\alpha}|t)+P_{l,j_\alpha}(t,\sigma_{m_\alpha},t)$$

and $\{j_\alpha, \alpha=1,\cdots,i\}$ is a permutation of $\{k_\alpha, \alpha=1,\cdots,i\}$.

Equation (B.3) implies that (B.17) can be rewritten as

$$
\begin{aligned}
&E^t[e^{\xi_l(t)}\xi_{k_1}(\sigma_1)\cdots\xi_{k_i}(\sigma_i)] = e^{\hat{\xi}_l(t|t)+(1/2)P_{ll}(t)} \\
\text{(B.19)} \qquad &\cdot\{E^t[\xi_{k_1}(\sigma_1)\cdots\xi_{k_i}(\sigma_i)]+\sum P_{l,j_1}(t,\sigma_{m_1},t)E^t[\xi_{j_2}(\sigma_{m_2})\cdots\xi_{j_i}(\sigma_{m_i})] \\
&+\sum P_{l,j_1}(t,\sigma_{m_1},t)P_{l,j_2}(t,\sigma_{m_2},t)E^t[\xi_{j_3}(\sigma_{m_3})\cdots\xi_{j_i}(\sigma_{m_i})] \\
&+\cdots+\sum P_{l,k_1}(t,\sigma_1,t)\cdots P_{l,k_i}(t,\sigma_i,t)\}.
\end{aligned}
$$

Hence, Lemmas B.2 and B.4 imply that the computation of $\hat{x}(t|t)$ involves only the computation of elements in $\hat{\Lambda}_j$, $j=1,\cdots,i$. However, the induction hypothesis implies that the elements of $\hat{\Lambda}_j$, $j=1,\cdots,i-1$ are FDC, so we need only prove that the elements of $\hat{\Lambda}_i$ are FDC.

(ii) Assume that $\eta\in\Lambda_i$ is defined by (B.14) (where $j=i$). Then the nonlinear filtering equation (A.5)–(A.6) for $\hat{\eta}(t|t)$ is

$$\text{(B.20)} \quad d\hat{\eta}(t|t)=E^t[\gamma_1(t)\xi_{k_1}(t)\lambda(t)]+\{E^t[\eta(t)\xi'(t)]-\hat{\eta}(t|t)\hat{\xi}'(t|t)\}H'(t)R^{-1}(t)\,d\nu(t)$$

where

$$\text{(B.21)} \qquad d\nu(t)=dz(t)-H(t)\hat{\xi}(t|t)\,dt$$

and

$$\text{(B.22)} \qquad \lambda(t)=\int_0^t\int_0^{\sigma_2}\cdots\int_0^{\sigma_{i-1}}\gamma_2(\sigma_2)\cdots\gamma_i(\sigma_i)\xi_{k_2}(\sigma_2)\cdots\xi_{k_i}(\sigma_i)\,d\sigma_2\cdots d\sigma_i$$

is an element of Λ_{i-1}; thus, by the induction hypothesis $\hat{\lambda}(t|t)$ is FDC. The first term in (B.20) (the *drift* term) is (see (B.3a))

$$
\begin{aligned}
&E^t[\gamma_1(t)\xi_{k_1}(t)\lambda(t)]=\gamma_1(t)\hat{\xi}_{k_1}(t|t)\hat{\lambda}(t|t) \\
\text{(B.23)} \qquad &+\gamma_1(t)E^t\Big[\sum_{l=2}^{i}\int_0^t\int_0^{\sigma_2}\cdots\int_0^{\sigma_{i-1}}P_{k_1,k_i}(t,\sigma_i,t)\gamma_2(\sigma_2)\cdots\gamma_i(\sigma_i) \\
&\cdot\xi_{k_2}\cdots\xi_{k_{l-1}}\xi_{k_{l+1}}\cdots\xi_{k_i}\,d\sigma_2\cdots d\sigma_i\Big].
\end{aligned}
$$

The first term in (B.23) is FDC by the induction hypothesis, and the second term, by Lemmas B.2 and B.4, is also FDC (i.e., it is an element of $\hat{\Lambda}_{i-2}$).

Equation (B.3a) implies that the *gain* term in (B.20) is the row vector (here $P_i(\sigma,t,t)$ denotes the ith row of $P(\sigma,t,t)$)

$$
\begin{aligned}
&E^t[\eta(t)\xi'(t)]-\hat{\eta}(t|t)\hat{\xi}'(t|t) \\
\text{(B.24)} \qquad &=\sum_{l=1}^{i}E^t\Big[\int_0^t\int_o^{\sigma_1}\cdots\int_0^{\sigma_{i-1}}\gamma_1(\sigma_1)\cdots\gamma_i(\sigma_i) \\
&\cdot\xi_{k_1}(\sigma_1)\cdots\xi_{k_{l-1}}(\sigma_{l-1})\xi_{k_{l+1}}(\sigma_{l+1})\cdots\xi_{k_i}(\sigma_i)P_{k_l}(\sigma_l,t,t)\,d\sigma_1\cdots d\sigma_k\Big]
\end{aligned}
$$

each element of which, by Lemmas B.2 and B.4, is an element of $\hat{\Lambda}_{i-1}$. Thus, by the induction hypothesis, the gain term, and hence the nonlinear equation (B.20) for $\hat{\eta}(t|t)$ is FDC. This completes the proof of Theorem 2.1. □

Proof of Theorem 2.2. This proof is identical to the proof of Theorem 2.1, except for the computation of the drift term in (B.20), so we will consider only that aspect of the proof. Assume that η is defined as in (2.4)—i.e., η is given by

$$\text{(B.25)} \quad \eta(t)=\int_0^t \int_0^{\sigma_1} \cdots \int_0^{\sigma_{j-1}} \xi_{k_1}(\sigma_{m_1})\cdots\xi_{k_i}(\sigma_{m_i})\gamma_1(\sigma_1)\cdots\gamma_j(\sigma_j)\, d\sigma_1\cdots d\sigma_j$$

where $i>j$; we also assume that $m_1=\cdots=m_\alpha=1$ and $m_\beta\neq 1$ for $\beta>\alpha$. In this proof, the induction is on j, the number of integrals in (B.25). That is, we assume that the theorem is true when η contains $\leqq j-1$ integrals, and prove that the theorem holds if η contains j integrals.

The nonlinear filtering equation yields

$$\text{(B.26)} \quad \begin{aligned} d\hat{\eta}(t|t) &= E^t[\gamma_1(\sigma_1)\xi_{k_1}(t)\cdots\xi_{k_\alpha}(t)\lambda(t)] \\ &\quad +\{E^t[\eta(t)\xi'(t)]-\hat{\eta}(t|t)\hat{\xi}'(t|t)\}H'(t)R^{-1}(t)\, d\nu(t) \end{aligned}$$

where $d\nu$ is defined in (B.21) and

$$\text{(B.27)} \quad \lambda(t)=\int_0^t \int_0^{\sigma_2} \cdots \int_0^{\sigma_{j-1}} \gamma_2(\sigma_2)\cdots\gamma_j(\sigma_j)\xi_{k_{\alpha+1}}(\sigma_{m_{\alpha+1}})\cdots\xi_{k_i}(\sigma_{m_i})\, d\sigma_2\cdots d\sigma_j.$$

The drift term in (B.26) is, from (B.3b),

$$\text{(B.28)} \quad \begin{aligned} &E^t[\gamma_1(t)\xi_{k_1}(t)\cdots\xi_{k_\alpha}(t)\lambda(t)] \\ &\quad =\gamma_1(t)E^t[\xi_{k_1}(t)\cdots\xi_{k_\alpha}(t)]\hat{\lambda}(t|t) \\ &\quad\quad +\gamma_1(t)\sum\Big\{E^t[\xi_{l_2}(t)\cdots\xi_{l_\alpha}(t)] \\ &\quad\quad\quad \cdot E^t\Big[\int_0^t\int_0^{\sigma_2}\cdots\int_0^{\sigma_{j-1}} \gamma_2(\sigma_2)\cdots\gamma_j(\sigma_j)P_{l_1 l_{\alpha+1}}(t,\sigma_{m_{\alpha+1}},t) \\ &\quad\quad\quad\quad \cdot\xi_{l_{\alpha+2}}(\sigma_{m_{\alpha+2}})\cdots\xi_{l_i}(\sigma_{m_i})\, d\sigma_2\cdots d\sigma_j\Big]\Big\}+\cdots \end{aligned}$$

where $\{l_1,\cdots,l_\alpha\}$ is a permutation of $\{k_1,\cdots,k_\alpha\}$ and $\{l_{\alpha+1},\cdots,l_i\}$ is a permutation of $\{k_{\alpha+1},\cdots,k_i\}$. The first term of (B.28) is FDC by the induction hypothesis, and the other terms, by Lemmas B.2 and B.4 and the induction hypothesis, are also FDC. We have also used the fact that the conditional distribution of $\xi(t)$ given z^t is Gaussian (Lemma 2.1) in order to conclude that $E^t[\xi_{k_1}(t)\cdots\xi_{k_\alpha}(t)]$ can be computed (via (B.3c)) as a memoryless function of $\hat{\xi}(t|t)$ and $P(t)$.

The gain term in (B.26) is also FDC; the proof is identical to that of Theorem 2.1. Hence $\hat{\eta}(t|t)$ is FDC, and Theorem 2.2 is proved. □

Acknowledgment. The authors would like to thank Professor Roger Brockett of Harvard University for many helpful discussions and for suggesting the use of Volterra series in the present context.

REFERENCES

[1] R. W. BROCKETT, *System theory on group manifolds and coset spaces*, SIAM J. Control, 10 (1972), pp. 265–284.

[2] ———, *Volterra series and geometric control theory*, Automatica, 12 (1976), pp. 167–176.

[3] ———, *Nonlinear systems and differential geometry*, Proc. IEEE, 64 (January 1976), pp. 61–72.

[4] ———, Personal communication.

[5] P. D'ALESSANDRO, A. ISIDORI AND A. RUBERTI, *Realizations and structure theory of bilinear dynamical systems*, SIAM J. Control, 12 (1974), pp. 517–535.

[6] M. FLIESS, *Un outil algebrique: les series formelles non commutatives*, Mathematical Systems Theory, G. Marchesini and S. K. Mitter, eds., Springer-Verlag, New York, 1976, pp. 122–148.

[7] M. FUJISAKI, G. KALLIANPUR, AND H. KUNITA, *Stochastic differential equations for the nonlinear filtering problem*, Osaka J. Math., 9 (1972), pp. 19–40.

[8] A. H. JAZWINSKI, *Stochastic Processes and Filtering Theory*, Academic Press, New York, 1970.

[9] C. JOHNSON AND E. B. STEAR, *Optimal filtering in the presence of multiplicative noise*, Proc. Fifth Symp. Nonlinear Estimation Theory and Its Applications, September 1974, Western Periodicals, North Hollywood, CA, pp. 124–134.

[10] V. JURDJEVIC AND H. J. SUSSMAN, *Control systems on Lie groups*, J. Differential Equations, 12 (1972), pp. 313–329.

[11] T. KAILATH AND P. FROST, *An innovations approach to least-squares estimation, Part II: Linear smoothing in additive white noise*, IEEE Trans. Automatic Control, AC-13 (1968), pp. 655–660.

[12] R. E. KALMAN AND R. S. BUCY, *New results in linear filtering and prediction theory*, J. Basic Engr. (Trans. ASME), 83D (1961) pp. 95–108.

[13] A. J. KRENER, *On the equivalence of control systems and linearization of nonlinear systems*, SIAM J. Control, 11 (1973), pp. 670–676.

[14] H. J. KUSHNER, *Dynamical equations for optimal nonlinear filtering*, J. Differential Equations, 3 (1967), pp. 179–190.

[15] H. KWAKERNAAK, *Optical filtering in linear systems with time delays*, IEEE Trans. Automatic Control, AC-12 (1967), pp. 169–173.

[16] C. T. LEONDES, J. B. PELLER AND E. B. STEAR, *Nonlinear smoothing theory*, IEEE Trans. Systems Sci. Cybernet., SSC-6 (1970), p. 63–71.

[17] J. T. LO AND A. S. WILLSKY, *Estimation for rotational processes with one degree of freedom I: introduction and continuous time processes*, IEEE Trans. Automatic Control, AC-20 (1975), pp. 10–21.

[18] S. I. MARCUS, *Estimation and analysis of nonlinear stochastic systems*, Ph.D. thesis, Dept. of Electrical Engineering, Mass. Inst. of Tech., Cambridge, June 1975; M.I.T. Electronic Systems Laboratory Rep. ESL-R-601, June 1975.

[19] S. I. MARCUS AND A. S. WILLSKY, *A class of finite dimensional optimal nonlinear estimators*, Proc. Fifth Symp. Nonlinear Estimation Theory and Its Applications, September 1974, Western Periodicals, North Hollywood, CA, pp. 158–168.

[20] D. N. MARTIN, *Stability criteria for systems with colored multiplicative noise*, Ph.D. thesis, Dept. of Electrical Engineering, Mass. Inst. of Tech. Cambridge, June 1974; M.I.T. Electronic Systems Laboratory Rep. ESL-R-567, June 1974.

[21] K. S. MILLER, *Multidimensional Gaussian Distributions*, John Wiley, New York, 1965.

[22] I. B. RHODES, *A tutorial introduction to estimation and filtering*, IEEE Trans. Automatic Control, AC-16 (1971), pp. 688–706.

[23] A. A. SAGLE AND R. E. WALDE, *Introduction to Lie Groups and Lie Algebras*, Academic Press, New York, 1973.

[24] H. J. SUSSMANN, *Semigroup representations, bilinear approximation of input-output maps, and generalized inputs*, Mathematical Systems Theory, G. Marchesini and S. K. Mitter, eds., Springer-Verlag, New York, 1976, pp. 172–192.

[25] A. S. WILLSKY, *Some estimation problems on Lie groups*, Geometric Methods in System Theory, R. W. Brockett and D. W. Mayne, eds., Reidel, Hingham, MA, 1973.

[26] A. S. WILLSKY AND S. I. MARCUS, *Estimation for bilinear stochastic systems*, Proc. of the 1974 Conf. on Bilinear Systems, Lecture Notes in Economics and Mathematical Systems, R. R. Mohler and A. Ruberti, eds., Springer-Verlag, New York, 1975, pp. 116–137.

[27] ———, *Analysis of bilinear noise models in circuits and devices*, J. Franklin Inst., 301 (1976), pp. 103–122.

[28] W. M. WONHAM, *On a matrix Riccati equation of stochastic control*, SIAM J. Control, 6 (1968), pp. 681–697.