Filtering of Stochastic Nonlinear Differential Systems via a Carleman Approximation Approach

Alfredo Germani, *Member, IEEE*, Costanzo Manes, and Pasquale Palumbo

***Abstract*—This paper deals with the state estimation problem for stochastic nonlinear differential systems, driven by standard Wiener processes, and presents a filter that is a generalization of the classical Extended Kalman-Bucy filter (EKBF). While the EKBF is designed on the basis of a first order approximation of the system around the current estimate, the proposed filter exploits a Carleman-like approximation of a chosen degree $\nu \geq 1$. The approximation procedure, applied to both the state and the measurement equations, allows to define an approximate representation of the system by means of a bilinear system, for which a filtering algorithm is available from the literature. Numerical simulations on an example show the improvement, in terms of sample error covariance, of the filter based on the first-order, second-order and third-order system approximations $(\nu = 1, 2, 3)$.**

***Index Terms*—Carleman approximation, extended Kalman-Bucy filter, nonlinear filtering, Polynomial filtering.**

I. Introduction

This note considers the filtering problem for nonlinear stochastic differential systems described by the Itô equations

$$\begin{aligned} dx(t) &= \phi(x(t), u(t))dt + F dW^1(t), \quad x(0) = x_0 \\ dy(t) &= h(x(t), u(t))dt + G dW^2(t), \quad y(0) = 0, \quad a.s., \end{aligned} \tag{1}$$

defined on a probability space $(\Omega, \mathcal{F}, \mathcal{P})$, where $x(t) \in I\!\!R^n$ is the state vector, $u(t) \in I\!\!R^p$ is a known deterministic input, $y(t) \in I\!\!R^q$ is the measured output, $W^1(t) \in I\!\!R^s$ and $W^2(t) \in I\!\!R^q$ are independent standard Wiener processes with respect to a family of increasing σ-algebras $\{\mathcal{F}_t,\ t \geq 0\}$ (i.e., the components of vectors $W^1(t)$ and $W^2(t)$ are a set of independent standard Wiener processes). $\phi : I\!\!R^n \times I\!\!R^p \mapsto I\!\!R^n$ and $h : I\!\!R^n \times I\!\!R^p \mapsto I\!\!R^q$ are smooth nonlinear maps. The initial state x_0 is an $\mathcal{F}_0$-measurable random variable, independent of both W^1 and W^2. In order to avoid singular filtering problems, see [5], the standard assumption of nonsingular output-noise covariance is made here, i.e., $\mathrm{rank}(GG^T) = q$. However, the approach presented in [8] could be followed when the covariance of the measurement noise is singular or even zero.

It is well known that the minimum variance state estimate requires the knowledge of the conditional probability density, whose computation, in the general case, is a difficult infinite-dimensional problem [4], [23], [24], [32]. For this reason a great deal of work has been made in the past to study implementable approximations of the optimal estimator [11], [12], [16], [18].

Another approach to state estimation consists in considering the time discretization of the original system and then to apply known filtering procedures, like the Extended Kalman Filter (EKF), the most widely used algorithm in nonlinear filtering problems (see, e.g., [13], [20]), particle filters [26], Gaussian sum approximations [19], the Unscented Kalman Filter (UKF) [21]. When measurements are considered in continuous time, the Extended Kalman Bucy Filter (EKBF) (see [13]) can be applied. More recently, a *polynomial* extension of the EKF (denoted PEKF) has been proposed in [15], which is based on the application of the optimal polynomial filter of [6], [7] to the Carleman approximation of the nonlinear discrete-time system [22].

This paper presents a procedure for the filter design for systems of the type (1) that generalizes the standard extended Kalman-Bucy approach, and avoids time discretization. The main step consists in computing a Carleman-like approximation of a chosen degree ν of the original nonlinear stochastic differential system (1), in the form of a bilinear system (linear drift and multiplicative noise) with respect to a suitably defined extended state. In general, there may be good reasons for computing bilinear approximations of nonlinear systems [3], and therefore bilinearization techniques have been used in the past in problems of systems approximation [28], [29], [31].

Once the Carleman bilinear approximation of a system is computed, the equations of the optimal linear filter for stochastic bilinear differential systems presented in [9] can be applied without any further approximation (note that the paper [9] presents the optimal polynomial filter specifically worked out for bilinear systems, and therefore the technique can not be directly applied to nonlinear systems of the form (1)). When $\nu = 1$ the proposed filtering algorithm reduces to the classical Extended Kalman-Bucy filter (EKBF), consisting of the Kalman-Bucy filter equations applied to the linear approximation of the differential system. Better performances of filters designed using higher order system approximations are expected.

The paper is organized as follows: the next section presents the Carleman approximation of stochastic nonlinear differential systems of the type (1); in Section Three the optimal linear filter for the Carleman bilinear approximation is derived; Section Four displays some numerical results where the performances of the proposed algorithm are compared with those of an EKBF.

II. Carleman Approximation

In order to compute the Carleman approximation of a chosen order ν of the system (1), with ν positive integer, it is necessary to assume that the random vector x_0 has finite moments up to the degree 2ν

$$\zeta_i = I\!\!E\left\{x_0^{[i]}\right\} < +\infty, \quad i = 1, \ldots, 2\nu, \tag{2}$$

where the square brackets at the exponent denote the Kronecker powers (see [7] for a quick survey on the Kronecker product and its main properties).

Under standard analyticity hypotheses, both the state and the output equations can be written by using the Taylor polynomial approximation around a suitably chosen state $\bar{x} \in I\!\!R^n$. When using the Taylor approximation for the filter design, the state $\bar{x}$ can be chosen as a *fixed* state that defines a nominal working point of the system or, as an alternative, as the current state estimate (more details are given in Remark 1, at the end of Section III). According to the Kronecker formalism, the differential system in (1) becomes

$$\begin{aligned} dx(t) &= \sum_{i=0}^{\infty} \Phi_i(\bar{x}, u(t))(x(t) - \bar{x})^{[i]} dt + \sum_{j=1}^{s} F_j dW_j^1(t) \\ dy(t) &= \sum_{i=0}^{\infty} H_i(\bar{x}, u(t))(x(t) - \bar{x})^{[i]} dt + \sum_{j=1}^{q} G_j dW_j^2(t) \end{aligned} \tag{3}$$

where F_j and G_j are the columns of F and G, respectively, and

$$\Phi_i(x, u) = \frac{1}{i!}\left(\nabla_x^{[i]} \otimes \phi\right), \quad H_i(x, u) = \frac{1}{i!}\left(\nabla_x^{[i]} \otimes h\right). \tag{4}$$

Manuscript received March 13, 2007; revised July 27, 2007. Recommended by Associate Editor M.-Q. Xiao. This work was supported by the Italian Ministry of University and Research (MIUR) under Grant PRIN 2005097521_003, and the National Research Council (CNR).

A. Germani and C. Manes are with the Dipartimento di Ingegneria Elettrica e dell'Informazione, Università degli Studi dell'Aquila, 67040 L'Aquila, Italy (e-mail: germani@ing.univaq.it; manes@ing.univaq.it).

P. Palumbo is with the Istituto di Analisi dei Sistemi ed Informatica del CNR "A. Ruberti," 00185 Roma, Italy (e-mail: palumbo@iasi.rn.cnr.it).

Digital Object Identifier 10.1109/TAC.2007.908347

The differential operator $\nabla_x^{[i]}\otimes$ applied to a function $\psi = \psi(x) : \mathbb{R}^n \mapsto \mathbb{R}^p$ is defined as follows

$$\nabla_x^{[0]} \otimes \psi = \psi, \quad \nabla_x^{[i+1]} \otimes \psi = \nabla_x \otimes \left(\nabla_x^{[i]} \otimes \psi\right), \quad i \geq 1 \quad (5)$$

with $\nabla_x = [\partial/\partial x_1 \ \cdots \ \partial/\partial x_n]$ and $\nabla_x \otimes \psi$ the Jacobian of the vector function ψ.

For the reader's convenience, some useful properties of the Kronecker product and of the differential operator (5) are reported below. The following property is intensively used in this paper: for any matrices A, B, C, D of suitable dimensions it is

$$(A \cdot B) \otimes (C \cdot D) = (A \otimes C) \cdot (B \otimes D) \quad (6)$$

where $A \cdot B$ denotes the standard matrix product. Recall that the Kronecker product is not commutative. Given a pair of integers (a, b), the symbol $C_{a,b}$ denotes a orthonormal *commutation matrix* in $\{0,1\}^{ab\times ab}$ such that, given any two matrices $A \in \mathbb{R}^{r_a \times c_a}$ and $B \in \mathbb{R}^{r_b \times c_b}$

$$B \otimes A = C_{r_a,r_b}^T (A \otimes B) C_{c_a,c_b}. \quad (7)$$

The Kronecker power of a binomial, $(a+b)^{[h]}$, allows the following expansion:

$$(a+b)^{[h]} = \sum_{j=0}^{h} M_j^h \left(a^{[j]} \otimes b^{[h-j]}\right), \quad \forall a, b \in \mathbb{R}^n \quad (8)$$

with M_j^h suitably defined matrix coefficients in $\mathbb{R}^{n\times n}$ (see [7]). Throughout the paper, the symbol I_n will denote the identity matrix of order n.

Lemma 1: For any $x \in \mathbb{R}^n$ the following identities hold:

$$\nabla_x \otimes x^{[h]} = U_n^h \left(I_n \otimes x^{[h-1]}\right), \quad h \geq 1 \quad (9)$$

$$\nabla_x^{[2]} \otimes x^{[h]} = O_n^h \left(I_{n^2} \otimes x^{[h-2]}\right), \quad h > 1 \quad (10)$$

where U_n^h and O_n^h, for $h > 1$ are recursively computed as

$$U_n^h = I_{n^h} + C_{n^{h-1},n}^T \left(U_n^{h-1} \otimes I_n\right)$$
$$O_n^h = U_n^h C_{n^{h-1},n}^L \left(\left(U_n^{h-1} C_{n^{h-2},n}^L\right) \otimes I_n\right) C_{n^2,n^{h-2}}^T \quad (11)$$

with initial value $U_n^1 = I_n$ (the proof is reported in the Appendix).

In the derivation of the Carleman approximation of system (1) the Itô formula for the computation of stochastic differentials written using the Kronecker formalism is required. Using (70), proved in the Appendix along with Lemma 1, the differential of the Kronecker power $x^{[k]}$ can be written as

$$d\left(x^{[k]}(t)\right) = \left(\nabla_x \otimes x^{[k]}\right)\phi(x(t), u(t))dt + \frac{1}{2}\left(\nabla_x^{[2]} \otimes x^{[k]}\right) F_0\, dt + \left(\nabla_x \otimes x^{[k]}\right) F dW^1(t) \quad (12)$$

where

$$F_0 = \sum_{j=1}^{s} F_j^{[2]}. \quad (13)$$

The use of identities (9) and (10) gives the following expression for the differentials $d(x^{[k]}(t))$, for $k \geq 2$:

$$d\left(x^{[k]}(t)\right) = U_n^k \left(I_n \otimes x^{[k-1]}(t)\right)\phi(x(t), u(t))dt + \frac{1}{2} O_n^k \left(I_{n^2} \otimes x^{[k-2]}(t)\right) F_0 dt + U_n^k \left(I_n \otimes x^{[k-1]}\right) F dW^1(t). \quad (14)$$

Lemma 2: Consider system (1). For $k \geq 2$, the differential of the k-th order Kronecker power of the state $x(t)$ can be written as

$$d\left(x^{[k]}(t)\right) = \sum_{r=0}^{\infty} A_r^k(\bar{x}, u(t))(x(t) - \bar{x})^{[r]} dt + \sum_{j=1}^{s} \mathbf{B}_{jk} x^{[k-1]}(t) dW_j^1(t) \quad (15)$$

with

$$A_r^k(\bar{x}, u) = A_r^{k,a}(\bar{x}, u) + A_r^{k,b}(\bar{x}) \quad (16)$$

where

$$A_r^{k,a}(\bar{x}, u) = \sum_{i=0 \vee (r-k+1)}^{r} U_n^k (\Phi_i(\bar{x}, u) \otimes I_{n^{k-1}}) \cdot \left(I_{n^i} \otimes M_{r-i}^{k-1}\right)\left(I_{n^r} \otimes \bar{x}^{[k-r+i-1]}\right) \quad (17)$$

$$A_r^{k,b}(\bar{x}) = \frac{1}{2} O_n^k (F_0 \otimes I_{n^{k-2}}) M_r^{k-2} \left(\bar{x}^{[k-2-r]} \otimes I_{n^r}\right), \quad r \leq k-2 \quad (18)$$

$$A_r^{k,b}(\bar{x}) = 0, \quad r > k-2 \quad (19)$$

and

$$\mathbf{B}_{jk} = U_n^k (F_j \otimes I_{n^{k-1}}). \quad (20)$$

Proof: Consider the first term of (14) and replace $\phi(x, u)$ with the power expansion given in (3). The repeated use of (6) gives

$$\begin{aligned} U_n^k &\left(I_n \otimes x^{[k-1]}\right)\phi(x, u) \\ &= \sum_{i=0}^{\infty} U_n^k \left(I_n \otimes x^{[k-1]}\right)\Phi_i(\bar{x}, u)(x - \bar{x})^{[i]} \\ &= \sum_{i=0}^{\infty} U_n^k \left(\left(\Phi_i(\bar{x}, u)(x - \bar{x})^{[i]}\right) \otimes x^{[k-1]}\right) \\ &= \sum_{i=0}^{\infty} U_n^k (\Phi_i(\bar{x}, u) \otimes I_{n^{k-1}}) \left((x - \bar{x})^{[i]} \otimes x^{[k-1]}\right) \\ &= \sum_{i=0}^{\infty} U_n^k (\Phi_i(\bar{x}, u) \otimes I_{n^{k-1}}) \\ &\quad \times \left((x - \bar{x})^{[i]} \otimes ((x - \bar{x}) + \bar{x})^{[k-1]}\right). \end{aligned} \quad (21)$$

By applying (8) to the $(k-1)$-th Kronecker power in (21)

$$\begin{aligned} U_n^k\left(I_n \otimes x^{[k-1]}\right)\phi(x,u) &= \sum_{i=0}^{\infty} U_n^k\left(\Phi_i(\bar{x},u)\otimes I_{n^{k-1}}\right) \\ &\quad \times \left((x-\bar{x})^{[i]} \otimes \sum_{j=0}^{k-1} M_j^{k-1}\left((x-\bar{x})^{[j]} \otimes \bar{x}^{[k-1-j]}\right)\right) \\ &= \sum_{i=0}^{\infty}\sum_{j=0}^{k-1} U_n^k(\Phi_i(\bar{x},u)\otimes I_{n^{k-1}})\left(I_{n^i}\otimes M_j^{k-1}\right) \\ &\quad \times \left((x-\bar{x})^{[i+j]} \otimes \bar{x}^{[k-1-j]}\right) \\ &= \sum_{i=0}^{\infty}\sum_{j=0}^{k-1} U_n^k(\Phi_i(\bar{x},u)\otimes I_{n^{k-1}})\left(I_{n^i}\otimes M_j^{k-1}\right) \\ &\quad \times \left(I_{n^{i+j}} \otimes \bar{x}^{[k-1-j]}\right)(x-\bar{x})^{[i+j]}. \end{aligned} \tag{22}$$

After the change of index $r = i + j$, the sums in (22) become

$$\begin{aligned} U_n^k\left(I_n \otimes x^{[k-1]}\right)\phi(x,u) = \sum_{i=0}^{\infty}\sum_{r=i}^{i+k-1} U_n^k(\Phi_i(\bar{x},u)\otimes I_{n^{k-1}}) \\ \times \left(I_{n^i}\otimes M_{r-i}^{k-1}\right)\left(I_{n^r}\otimes \bar{x}^{[k-r+i-1]}\right)(x-\bar{x})^{[r]}. \end{aligned} \tag{23}$$

From the following equivalence between summations:

$$\sum_{i=0}^{\infty}\sum_{r=i}^{i+k-1}(\,)_{i,k,r} = \sum_{r=0}^{\infty}\sum_{i=0\vee(r-k+1)}^{r}(\,)_{i,k,r}, \tag{24}$$

it follows that

$$U_n^k\left(I_n \otimes x^{[k-1]}\right)\phi(x,u) = \sum_{r=0}^{\infty} A_r^{k,a}(\bar{x},u)(x-\bar{x})^{[r]} \tag{25}$$

where $A_r^{k,a}(\bar{x},u)$ is given in (17). Now consider the second term in (14). It can be written as

$$\begin{aligned} \frac{1}{2}O_n^k\left(I_{n^2}\otimes x^{[k-2]}\right)F_0 &= \frac{1}{2}O_n^k\left(F_0 \otimes x^{[k-2]}\right) \\ &= \frac{1}{2}O_n^k(F_0\otimes I_{n^{k-2}})x^{[k-2]}. \end{aligned} \tag{26}$$

Since it is

$$\begin{aligned} x^{[k-2]} &= (\bar{x}+(x-\bar{x}))^{[k-2]} \\ &= \sum_{r=0}^{k-2} M_r^{k-2}\left(\bar{x}^{[k-2-r]}\otimes(x-\bar{x})^{[r]}\right) \\ &= \sum_{r=0}^{k-2} M_r^{k-2}\left(\bar{x}^{[k-2-r]}\otimes I_{n^r}\right)(x-\bar{x})^{[r]}, \end{aligned} \tag{27}$$

from (26) and (27) it follows

$$\begin{aligned} \frac{1}{2}O_n^k\left(I_{n^2}\otimes x^{[k-2]}(t)\right)F_0 = \sum_{r=0}^{k-2}\frac{1}{2}O_n^k(F_0\otimes I_{n^{k-2}})M_r^{k-2} \\ \times\left(\bar{x}^{[k-2-r]}\otimes I_{n^r}\right)(x-\bar{x})^{[r]}, \end{aligned} \tag{28}$$

i.e., from definitions (18)

$$\frac{1}{2}O_n^k\left(I_{n^2}\otimes x^{[k-2]}(t)\right)F_0 = \sum_{r=0}^{\infty} A_r^{k,b}(\bar{x})(x-\bar{x})^{[r]}. \tag{29}$$

Consider now the last term in (14). By repeated use of the property (6) it can be written as

$$\begin{aligned} U_n^k&\left(I_n\otimes x^{[k-1]}\right)FdW^1 \\ &= \sum_{j=1}^{s} U_n^k\left(I_n\otimes x^{[k-1]}\right)(F_j\otimes 1)dW_j^1 \\ &= \sum_{j=1}^{s} U_n^k\left(F_j\otimes x^{[k-1]}\right)dW_j^1 \\ &= \sum_{j=1}^{s} U_n^k\left((F_j\cdot 1)\otimes(I_{n^{k-1}}\cdot x^{[k-1]})\right)dW_j^1 \\ &= \sum_{j=1}^{s} U_n^k(F_j\otimes I_{n^{k-1}})x^{[k-1]}dW_j^1, \end{aligned} \tag{30}$$

i.e., considering the definition (20)

$$U_n^k\left(I_n\otimes x^{[k-1]}\right)FdW^1 = \sum_{j=1}^{s}\mathbf{B}_{jk}x^{[k-1]}dW_j^1. \tag{31}$$

The Lemma is proved by substitution of (25), (29), and (31) into (14). ■

By neglecting in the summations in (3) and (15) the higher order terms, greater than a chosen degree ν, the differentials $d(x^{[k]}(t))$, $k = 1, 2, \ldots$, given by (1) for $k = 1$ and by (14) for $k \geq 2$, and $dy(t)$ are approximated as follows:

$$\begin{aligned} d\left(x^{[k]}(t)\right) &\simeq \sum_{r=0}^{\nu} A_r^k(\bar{x},u(t))(x(t)-\bar{x})^{[r]}dt \\ &\quad + \sum_{j=1}^{s}\left(\mathbf{B}_{jk}x^{[k-1]}(t)+\mathbf{F}_{jk}\right)dW_j^1(t), \\ dy(t) &\simeq \sum_{i=0}^{\nu} H_i(\bar{x},u(t))(x(t)-\bar{x})^{[i]}dt + GdW^2(t) \end{aligned} \tag{32}$$

with $A_r^k(\bar{x},u)$ and $\mathbf{B}_{jk}$, for $k \geq 2$ defined by (16)–(20), and for $k = 1$

$$\begin{aligned} A_r^1(\bar{x},u) = \Phi_r(\bar{x},u) \quad \forall r = 0,\ldots,\nu \\ \mathbf{B}_{j1} = 0 \quad \forall j = 1,\ldots,s \end{aligned} \tag{33}$$

moreover

$$\mathbf{F}_{j1} = F_j, \quad \mathbf{F}_{jk} = 0, \quad \forall j = 1,\ldots,s, \quad \forall k > 1. \tag{34}$$

Consider now the following equivalence

$$\sum_{r=0}^{\nu} A_r^k(\bar{x},u)(x-\bar{x})^{[r]} = \sum_{j=0}^{\nu}\mathbf{A}_j^k(\bar{x},u)x^{[j]} \tag{35}$$

with

$$\mathbf{A}_j^k(\bar{x},u)=\sum_{r=j}^{\nu}(-1)^{r-j}A_r^k(\bar{x},u)M_j^r\left(I_{n^j}\otimes\bar{x}^{[r-j]}\right) \quad (36)$$

derived from the following identities

$$\begin{aligned}\sum_{r=0}^{\nu}&A_r^k(\bar{x},u)(x-\bar{x})^{[r]}\\ &=\sum_{r=0}^{\nu}\sum_{j=0}^{r}A_r^k(\bar{x},u)M_j^r\left(x^{[j]}\otimes(-\bar{x})^{[r-j]}\right)\\ &=\sum_{j=0}^{\nu}\sum_{r=j}^{\nu}(-1)^{r-j}A_r^k(\bar{x},u)M_j^r\left(I_{n^j}\otimes\bar{x}^{[r-j]}\right)x^{[j]}. \quad (37)\end{aligned}$$

Analogously, it comes that

$$\sum_{i=0}^{\nu}H_i(\bar{x},u)(x-\bar{x})^{[i]}dt=\sum_{j=0}^{\nu}\mathbf{C}_j(\bar{x},u)x^{[j]}dt \quad (38)$$

with

$$\mathbf{C}_j(\bar{x},u)=\sum_{i=j}^{\nu}(-1)^{i-j}H_i(\bar{x},u)M_j^i\left(I_{n^j}\otimes\bar{x}^{[i-j]}\right). \quad (39)$$

It follows that the approximation (32) of the differentials $d(x^{[k]}(t))$ and $dy(t)$ can be rewritten, for $k\geq 1$, as

$$\begin{aligned}d\left(x^{[k]}(t)\right)&\simeq\sum_{j=0}^{\nu}\mathbf{A}_j^k\left(\bar{x},u(t)\right)x^{[j]}(t)\,dt\\ &\quad+\sum_{j=1}^{s}\left(\mathbf{B}_{jk}x^{[k-1]}(t)+\mathbf{F}_{jk}\right)dW_j^1(t),\\ dy(t)&\simeq\sum_{j=0}^{\nu}\mathbf{C}_j\left(\bar{x},u(t)\right)x^{[j]}(t)dt+GdW^2(t). \quad (40)\end{aligned}$$

The ν-degree Carleman bilinear approximation of the stochastic differential system (1) around a given $\bar{x}\in I\!R^n$ is a system with state $X^\nu(t)\in I\!R^{n_\nu}$, with $n_\nu=n+n^2+\cdots+n^\nu$, and output $Y^\nu(t)\in I\!R^q$ that obey the equations

$$\begin{aligned}dX_k^\nu(t)&=\sum_{j=0}^{\nu}\mathbf{A}_j^k(\bar{x},u(t))X_j^\nu(t)dt\\ &\quad+\sum_{j=1}^{s}\left(\mathbf{B}_{jk}X_{k-1}^\nu(t)+\mathbf{F}_{jk}\right)dW_j^1(t)\\ dY^\nu(t)&=\sum_{j=0}^{\nu}\mathbf{C}_j(\bar{x},u(t))X_j^\nu(t)dt+GdW^2(t) \quad (41)\end{aligned}$$

where $X_k^\nu(t)\in I\!R^{n^k}$, $k=1,\ldots,\nu$, denotes the k-th block component of the state $X^\nu(t)$. Comparing the (32) and (41), it is clear that $X_k^\nu(t)$ is aimed to approximate $x^{[k]}(t)$, $k=1,\ldots,\nu$, while $Y^\nu(t)$ should approximate $y(t)$.

The (41) can be put in a compact matrix form as follows

$$\begin{aligned}dX^\nu(t)&=\mathbf{A}^\nu(\bar{x},u(t))X^\nu(t)dt+\mathbf{N}^\nu(\bar{x},u(t))dt\\ &\quad+\sum_{j=1}^{s}\left(\mathbf{B}_j^\nu X^\nu(t)+\mathbf{F}_j^\nu\right)dW_j^1(t),\\ dY^\nu(t)&=\mathbf{C}^\nu(\bar{x},u(t))X^\nu(t)dt\\ &\quad+\mathbf{D}^\nu(\bar{x},u(t))dt+GdW^2(t), \quad (42)\end{aligned}$$

with $X^\nu(0)=(x_0^T\ \cdots\ (x_0^{[\nu]})^T)^T$, $k=1,\ldots,\nu$, $Y^\nu(0)=0$, and

$$\mathbf{A}^\nu=\begin{bmatrix}\mathbf{A}_1^1&\cdots&\mathbf{A}_\nu^1\\ \vdots&\ddots&\vdots\\ \mathbf{A}_1^\nu&\cdots&\mathbf{A}_\nu^\nu\end{bmatrix},\quad \mathbf{N}^\nu=\begin{bmatrix}\mathbf{A}_0^1\\ \vdots\\ \mathbf{A}_0^\nu\end{bmatrix} \quad (43)$$

$$\mathbf{B}_j^\nu=\begin{bmatrix}0&0&\cdots&0\\ \mathbf{B}_{j2}&0&\cdots&0\\ 0&\ddots&\ddots&\vdots\\ 0&0&\mathbf{B}_{j\nu}&0\end{bmatrix},\quad \mathbf{F}_j^\nu=\begin{bmatrix}\mathbf{F}_{j1}\\ \mathbf{F}_{j2}\\ \vdots\\ \mathbf{F}_{j\nu}\end{bmatrix} \quad (44)$$

$$\mathbf{C}^\nu=[\mathbf{C}_1\ \cdots\ \mathbf{C}_\nu],\quad \mathbf{D}^\nu=\mathbf{C}_0 \quad (45)$$

III. The Filtering Algorithm

From the discussion in the previous section, it follows that the bilinear Carleman approximation (42) is an approximated generation model for the sequence $(x(t),y(t))$, produced by the exact model (1), and therefore they can be used for the construction of an approximate filter for system (1). From the approximation

$$x(t)\simeq\left[I_n\quad O_{n\times(n_\nu-n)}\right]X^\nu(t) \quad (46)$$

if an estimate $\tilde{X}^\nu(t)$ of the extended state $X^\nu(t)$ is known, then the following estimate of $x(t)$ is obtained

$$\tilde{x}(t)=\left[I_n\quad O_{n\times(n_\nu-n)}\right]\tilde{X}^\nu(t). \quad (47)$$

It is well known that the optimal estimate of $X^\nu(t)$ is provided by the conditional expectation w.r.t. all the Borel transformations of the measurements, whose computation in general can not be obtained through algorithms of finite dimension. However, recalling that $X^\nu(t)$ is generated by the bilinear system (42), the suboptimal filter presented in [9], specifically worked out for bilinear systems, can now be applied. In particular, the best affine estimator is used here. The output of such filter is the projection of $X(t)$ onto the space $L(Y_t^\nu)$ of all the affine transformations of the random variables $\{Y^\nu(\tau),t_0\leq\tau\leq t\}$. Let us denote $\hat{X}^\nu(t)=\Pi[X^\nu(t)\,|\,L(Y_t^\nu)]$ such projection (formally, the projection Π on the subspace $L(Y_t^\nu)$ is a random variable in $L(Y_t^\nu)$ such that the difference $X^\nu(t)-\Pi[X^\nu(t)|L(Y_t^\nu)]$ is *orthogonal* to $L(Y_t^\nu)$, i.e., is uncorrelated with all random variables in $L(Y_t^\nu)$). The estimate of $x(t)$ obtained from $\hat{X}^\nu(t)$ using (47), is denoted $\hat{x}^\nu(t)$, i.e.

$$\begin{aligned}\hat{x}^\nu(t)&=\left[I_n\quad O_{n\times(n_\nu-n)}\right]\hat{X}^\nu(t)\\ &=\left[I_n\quad O_{n\times(n_\nu-n)}\right]\Pi\left[X^\nu(t)\,|\,L\left(Y_t^\nu\right)\right]. \quad (48)\end{aligned}$$

In the filter equations the mean and covariance of $X^\nu(t)$ are required.

Lemma 3: Let $m_{X^\nu}(t)=I\!E\{X^\nu(t)\}$ and $\Psi_{X^\nu}(t)=\mathrm{Cov}(X^\nu(t))$ denote the mean value and the covariance matrix of $X^\nu(t)$, respectively. These obey the following equations:

$$\begin{aligned}\dot{m}_{X^\nu}(t)&=\mathbf{A}^\nu(\bar{x},u(t))m_{X^\nu}(t)+\mathbf{N}^\nu(\bar{x},u(t))\\ \dot{\Psi}_{X^\nu}(t)&=\mathbf{A}^\nu(\bar{x},u(t))\Psi_{X^\nu}(t)\\ &\quad+\Psi_{X^\nu}\mathbf{A}^{\nu T}(\bar{x},u(t))+Q(m_{X^\nu}(t),\Psi_{X^\nu}(t)) \quad (49)\end{aligned}$$

where

$$\begin{aligned}Q(m_{X^\nu},\Psi_{X^\nu})&=\sum_{i=1}^{s}\mathbf{B}_i^\nu\Psi_{X^\nu}\mathbf{B}_i^{\nu T}\\ &\quad+\sum_{i=1}^{s}\left(\mathbf{B}_i^\nu m_X^\nu+\mathbf{F}_i\right)\left(\mathbf{B}_i^\nu m_X^\nu+\mathbf{F}_i\right)^T \quad (50)\end{aligned}$$

with initial values $m_{X^\nu}(0)=(\zeta_1^T\ \cdots\ \zeta_\nu^T)^T$, $\Psi_{X^\nu}(0)=\mathrm{Cov}(X_0^\nu)$, where the vectors ζ_k are the moments defined in (2).

The proof of Lemma 3 can be found in [9].

Theorem 1: The optimal linear estimate $\hat{X}^\nu(t)$ of the process $X^\nu(t)$ given by (42), obeys the equation

$$\begin{aligned} d\hat{X}^\nu(t) = &\mathbf{A}^\nu(\bar{x}, u(t))\hat{X}^\nu(t)dt \\ &+ \mathbf{N}^\nu(\bar{x}, u(t))\,dt + P(t)\mathbf{C}^{\nu T}(\bar{x}, u(t))R^{-1} \\ &\cdot (dY^\nu(t) - (\mathbf{C}^\nu(\bar{x}, u(t))\hat{X}^\nu(t) \\ &+ \mathbf{D}^\nu(\bar{x}, u(t)))dt) \end{aligned} \tag{51}$$

with $R = GG^T$ and $P(t)$ is the error covariance matrix

$$P(t) = I\!E\{(X^\nu(t) - \hat{X}^\nu(t))(X^\nu(t) - \hat{X}^\nu(t))^T\} \tag{52}$$

evolving according to the following equation

$$\begin{aligned} \dot{P}(t) = &\mathbf{A}^\nu(\bar{x}, u(t))P(t) + P(t)\mathbf{A}^{\nu T}(\bar{x}, u(t)) \\ &+ Q(t) - P(t)\mathbf{C}^{\nu T}(\bar{x}, u(t))R^{-1}\mathbf{C}^\nu(\bar{x}, u(t))P(t) \end{aligned} \tag{53}$$

with $P(0) = \Psi_{X^\nu}(0)$, where $Q(t)$ in (53) is defined as $Q(t) = Q(m_{X^\nu}(t), \Psi_{X^\nu}(t))$, with $Q(\cdot, \cdot)$ defined by (50).

Proof: The proof is a straightforward consequence of ([9], Thm. 4.4). The equations are somewhat different (and shorter), because the state noise $W_1(t)$ and output noise $W_2(t)$ have been assumed independent in this paper. ■

Remark 1: The filter given in Theorem 1 provides the optimal affine estimate of $X^\nu(t)$ as a function of the observations $Y^\nu(t)$. From this, the estimate $\hat{x}^\nu(t)$ is computed using (48). However, the available measurement process is $y(t)$ and not $Y^\nu(t)$. Therefore, the differential $dY^\nu(t)$ in (51) should be replaced with $dy(t)$. Note that in the filter (51)–(53) the matrices $\mathbf{A}^\nu$, $\mathbf{N}^\nu$, $\mathbf{C}^\nu$ and $\mathbf{D}^\nu$ depend on $\bar{x}$ and $u(t)$ through the terms $\Phi_i(\bar{x}, u)$ and $H_i(\bar{x}, u)$, defined in (4) as the coefficients of the polynomial approximations of $\phi(x, u)$ and $h(x, u)$ around $\bar{x}$. Following a standard EKBF approach, these coefficients can be re-computed at each time t as a function of the current estimate $\hat{x}^\nu(t)$. Formally, this kind of filter is written replacing $\bar{x}$ with $\hat{x}^\nu(t)$ into the filter (51)–(53) as follows

$$\begin{aligned} d\hat{X}^\nu(t) = &\mathbf{A}^\nu(\hat{x}^\nu(t), u(t))\hat{X}^\nu(t)dt \\ &+ \mathbf{N}^\nu(\hat{x}^\nu(t), u(t))dt + P(t)\mathbf{C}^{\nu T}(\hat{x}^\nu(t), u(t))R^{-1} \\ &\cdot (dy(t) - (\mathbf{C}^\nu(\hat{x}^\nu(t), u(t))\hat{X}^\nu(t) + \mathbf{D}^\nu(\hat{x}^\nu(t), u(t)))dt) \end{aligned} \tag{54}$$

$$\begin{aligned} \dot{P}(t) = &\mathbf{A}^\nu(\hat{x}^\nu(t), u(t))P(t) + P(t)\mathbf{A}^{\nu T}(\hat{x}^\nu(t), u(t)) + Q(t) \\ &- P(t)\mathbf{C}^{\nu T}(\hat{x}^\nu(t), u(t))R^{-1}\mathbf{C}^\nu(\hat{x}^\nu(t), u(t))P(t) \end{aligned} \tag{55}$$

Remark 2: For $\nu = 1$ the (54)–(55) coincide with those of the EKBF. However, note that for $\nu = 2$ such equations *do not* coincide with those of the so called *second-order* EKBF (see [13]), in which the Riccati equations are exactly the same of the plain EKBF, and a second order Taylor approximation is used for the state process, where the second order state increments are substituted with the components of the error covariance matrix provided by the Riccati equation.

Remark 3: As discussed in Remark 1 of [15], the computational burden for real-time implementation can be reduced by eliminating the redundancies in the extended state vector X^ν. Recall that the Kronecker power $x^{[i]}$ has n^i components, of which only $\binom{n+i-1}{i}$ monomials are independent. It follows that X^ν has dimension $n + n^2 + \cdots n^\nu$, but only $\sum_{i=1}^{\nu}\binom{n+i-1}{i}$ (i.e., $\binom{n+\nu}{n} - 1$) components are independent. This allows to define a reduced-extended state that allows the derivation of filter equations with smaller size (see [15] for further details).

TABLE I
MEAN SQUARE ESTIMATION ERRORS FOR THE FILTER (51)–(53)

	$\nu = 0$	$\nu = 1$	$\nu = 2$	$\nu = 3$
$\mu_{x_1,\nu}$	0.5686	0.7983	0.4812	0.4787
$\mu_{x_2,\nu}$	0.4674	0.5893	0.3812	0.3811

IV. SIMULATION RESULTS

Numerical simulation results are here reported in order to show the effectiveness of the proposed algorithm. Consider the following nonlinear system:

$$\begin{aligned} dx_1(t) &= (-x_1(t) + x_1(t)x_2(t))dt + a\,dW^1(t), \\ dx_2(t) &= (-2x_2(t) - 2x_1(t)x_2(t))dt + b\,dW^1(t) \\ dy(t) &= (x_1(t) - x_1(t)x_2(k))dt + \gamma\,dW^2(t) \end{aligned} \tag{56}$$

with $a = b = 1$, $\gamma = 2$. The initial state x_0 is a Gaussian standard random vector (zero mean and identity covariance). Both filters (51)–(53) and (54)–(55) have been implemented for $\nu = 1, 2, 3$, using MATLAB©. The bilinearization point $\bar{x} = 0$ has been chosen for the filter (51)–(53). The computations needed for the filter derivation begin with the computation of the differential $dx^{[2]}(t) = [dx_1^2(t)\ \ d(x_1x_2)(t)\ \ d(x_1x_2)(t)\ \ dx_2^2(t)]$ (note the redundancy in the components of $dx^{[2]}(t)$). The application of the Itô formula (68) gives

$$\begin{aligned} dx_1^2(t) &= 2x_1(t)dx_1(t) + a^2dt \\ d(x_1x_2)(t) &= x_2(t)dx_1(t) + x_1(t)dx_2(t) + ab\,dt \\ dx_2^2(t) &= 2x_2(t)dx_2(t) + b^2dt \end{aligned} \tag{57}$$

from which, substituting the differentials (56), we have

$$\begin{aligned} dx_1^2(t) &= \left(-2x_1^2(t) + 2x_1^2(t)x_2(t) + a^2\right)dt + 2x_1(t)a\,dW^1(t) \\ d(x_1x_2)(t) &= \left(-3x_1(t)x_2(t) + x_1(t)x_2^2(t) - 2x_1^2(t)x_2(t) + ab\right)dt \\ &\quad + (ax_2(t) + bx_1(t))dW^1(t) \\ dx_2^2(t) &= \left(-4x_2^2(t) - 4x_1(t)x_2^2(t) + b^2\right)dt + 2x_2(t)b\,dW^1(t) \end{aligned} \tag{58}$$

Note that for the filter design with $\nu = 2$ and $\bar{x} = 0$ the third order terms $x_1^2x_2$ and $x_1x_2^2$ are neglected in the differential $dx^{[2]}(t)$. Due to lack of space, the differential $dx^{[3]}(t)$ is not reported.

The results displayed below are obtained simulating both system (56) and filter (51)–(53), in the time interval [0, 100] according to the Euler-Maruyama method [17] with integration step $\Delta = 0.1$. The performance of the filters are evaluated computing the mean of the squared estimation errors (MSE)

$$\mu_{x_i,\nu} = \frac{1}{N+1}\sum_{k=0}^{N}(x_i(t_k) - \hat{x}_i^\nu(t_k))^2 \tag{59}$$

where $t_k = k\Delta$, with $\Delta = 0.1$, and $N = 1\,000$. Let $\mu_{x_i,0}$ denote the sample mean square error of the trivial estimate $\hat{x}_i^0(t) \equiv 0$. The MSE $\mu_{x_i,\nu}$ averaged over 5 000 simulation runs are reported in Table I, where the improvement obtained by increasing the index ν can be recognized. Note that in this example the filter with $\nu = 1$ (i.e., the Kalman-Bucy filter based on the linear approximation of the process around $\bar{x} = 0$), provides worse performances than the trivial estimate $\hat{x}_i^0(t) \equiv 0$. It is interesting to note that in this example a considerable

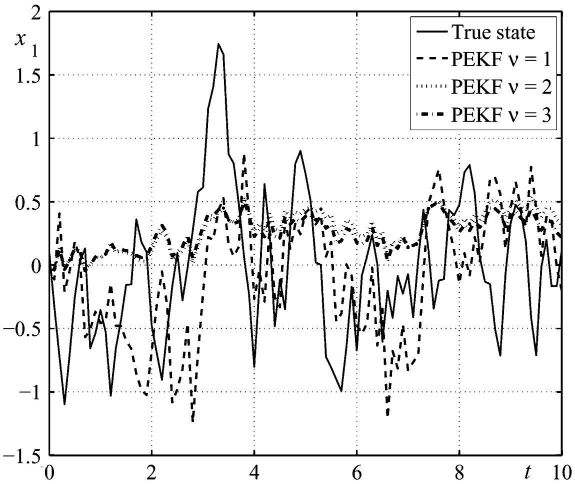

Fig. 1. True and estimated states: the first component.

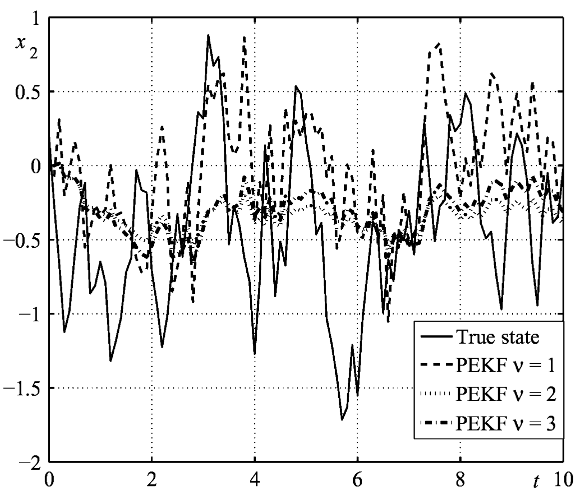

Fig. 2. True and estimated states: the second component.

TABLE II
MEAN SQUARE ESTIMATION ERRORS FOR THE FILTER (54)–(55)

	$\nu = 0$	$\nu = 1$	$\nu = 2$	$\nu = 3$
$\mu_{x_1,\nu}$	0.5686	0.7228	0.4723	0.4718
$\mu_{x_2,\nu}$	0.4674	0.5341	0.3721	0.3721

improvement is obtained when the second-order Carleman approximation is used for the filter derivation $(\nu = 2)$, while the third-order Carleman approximation $(\nu = 3)$ does not provide a significant improvement. The CPU times for the execution of a single run on a laptop with 2 GHz clock are

$$T_{\nu=1} = 0.079 \text{ s}, \quad T_{\nu=2} = 0.093 \text{ s}, \quad T_{\nu=3} = 0.22 \text{ s}. \tag{60}$$

Figs. 1 and 2 report the plots of the output of the last of the 5 000 simulation runs in the subinterval $[0, 10]$. Note that, due to the high measurement noise, in some instants the estimates are quite far form the true state. The mean square errors of the estimates produced by the filter (54)–(55), based on the system bilinearization around the current estimate $\hat{x}^{\nu}(t)$, are reported in Table II. The performances are somewhat better, especially for $\nu = 1$, but not dramatically different from those produced by the filter (51)–(53).

V. CONCLUSION

The problem of state estimation for stochastic nonlinear differential systems using bilinearized system approximation has been investigated in this paper. The filtering algorithm here proposed is based on two steps: first the nonlinear system is approximated by using a Carleman-like bilinearization approach, taking into account all the powers of the polynomial approximation of the drift and output functions, up to a chosen degree ν; next, the equations of the optimal linear filter for the approximating system are computed. This step is based on the paper [9], concerning suboptimal state estimate of stochastic bilinear systems. When the index $\nu = 1$, the proposed algorithm gives back the classical Extended Kalman-Bucy Filter, i.e., the Kalman-Bucy filter applied to the linear approximation of the original differential system.

APPENDIX

The proofs here reported exploits the two following properties of the ∇_x operator:

$$\nabla_x \otimes (f(x) \otimes g(x)) = (\nabla_x \otimes f) \otimes g(x) + C_{p,q}^T((\nabla_x \otimes g) \otimes f(x)) \tag{61}$$

for any differentiable $f : I\!R^n \mapsto I\!R^p$ and $g : I\!R^n \mapsto I\!R^q$, and

$$\begin{aligned} &\forall M \in I\!R^{r \times s}, \quad \forall N \in C^1(I\!R^n, I\!R^{s \times c}), \\ &\nabla_x \otimes (MN(x)) = M(\nabla_x \otimes N(x)). \end{aligned} \tag{62}$$

Proof of Lemma 1:

Proof: By finite induction. For $h = 1$ it is easy to compute $\nabla_x \otimes x = I_n$, so that (9) is true for $h = 1$. Assume that (9) is true for some $h - 1$, with $h > 1$, i.e.

$$\nabla_x \otimes x^{[h-1]} = U_n^{h-1}\left(I_n \otimes x^{[h-2]}\right). \tag{63}$$

Then

$$\begin{aligned} \nabla_x \otimes x^{[h]} &= \nabla_x\left(x \otimes x^{[h-1]}\right) \\ &= I_n \otimes x^{[h-1]} + C_{n^{h-1},n}^T\left(\left(\nabla_x \otimes x^{[h-1]}\right) \otimes x\right) \\ &= I_n \otimes x^{[h-1]} + C_{n^{h-1},n}^T\left(\left(U_n^{h-1}\left(I_n \otimes x^{[h-2]}\right)\right) \otimes x\right) \\ &= I_n \otimes x^{[h-1]} + C_{n^{h-1},n}^T\left(U_n^{h-1} \otimes I_n\right)\left(I_n \otimes x^{[h-1]}\right) \\ &= \left(I_{n^h} + C_{n^{h-1},n}^T\left(U_n^{h-1} \otimes I_n\right)\right)\left(I_n \otimes x^{[h-1]}\right). \end{aligned} \tag{64}$$

This proves (9), taking into account the definition of U_n^h in (11). The proof of (10) is obtained by expanding $\nabla_x^{[2]} \otimes x^{[h]}$, for $h \geq 2$, as follows

$$\begin{aligned} &\nabla_x^{[2]} \otimes x^{[h]} \\ &\quad = \nabla_x \otimes \left(\nabla_x \otimes x^{[h]}\right) \\ &\quad = \nabla_x \otimes \left(U_n^h\left(I_n \otimes x^{[h-1]}\right)\right) \\ &\quad = U_n^h\left(\nabla_x \otimes \left(I_n \otimes x^{[h-1]}\right)\right) \\ &\quad = U_n^h C_{n^{h-1},n}^L\left(\left(\nabla_x \otimes x^{[h-1]}\right) \otimes I_n\right) \\ &\quad = U_n^h C_{n^{h-1},n}^L\left(\left(U_n^{h-1}\left(I_n \otimes x^{[h-1]}\right)\right) \otimes I_n\right) \end{aligned} \tag{65}$$

$$\begin{aligned} &\nabla_x^{[2]} \otimes x^{[h]} \\ &\quad = U_n^h C_{n^{h-1},n}^L\left(\left(U_n^{h-1} C_{n^{h-2},n}^L\left(x^{[h-1]} \otimes I_n\right)\right) \otimes I_n\right) \\ &\quad = U_n^h C_{n^{h-1},n}^L\left(\left(U_n^{h-1} C_{n^{h-2},n}^L\right) \otimes I_n\right)\left(x^{[h-1]} \otimes I_{n^2}\right) \\ &\quad = U_n^h C_{n^{h-1},n}^L\left(\left(U_n^{h-1} C_{n^{h-2},n}^L\right) \otimes I_n\right) \\ &\qquad \times C_{n^2,n^{h-2}}^T\left(I_{n^2} \otimes x^{[h-1]}\right) = O_n^h\left(I_{n^2} \otimes x^{[h-1]}\right). \end{aligned} \tag{66}$$

■

In the derivation of the Carleman approximation of system (1) the Itô formula for the computation of stochastic differentials is needed (see [24]). Consider the stochastic process

$$dx_t = f(x_t)dt + g(x_t)dW_t = f(x_t)dt + \sum_{k=1}^{p} g_{:,k}(x_t)dW_{k,t} \tag{67}$$

where $x_t \in I\!R^n$ and W_t is a standard Wiener process in $I\!R^p$. The symbol $g_{:,k}(x_t)$ denotes the k-th column of matrix $g(x_t)$. Consider a transformation $z_t = r(t,x)$, where $r(\cdot,\cdot)$ is a scalar function, twice differentiable. The differential dz_t, computed according to the Itô formula, is as follows:

$$\begin{aligned} dz_t = & \left.\frac{\partial r}{\partial t}\right|_{(t,x_t)} dt + \left.\frac{\partial r}{\partial x}\right|_{(t,x_t)} dx_t \\ & + \frac{1}{2}\sum_{i,j} \frac{\partial^2 r}{\partial x_i \partial x_j} \left.\left(g_{i,:}g_{j,:}^T\right)\right|_{(t,x_t)} dt \end{aligned} \tag{68}$$

or, equivalently

$$dz_t = \left.\frac{\partial r}{\partial t}\right|_{(t,x_t)} dt + \left.\frac{\partial r}{\partial x}\right|_{(t,x_t)} dx_t + \frac{1}{2}\mathrm{tr}\left.\frac{\partial^2 r}{\partial x^2}(gg^T)\right|_{(t,x_t)} dt. \tag{69}$$

Using the Kronecker formalism the Itô differential can be written as

$$\begin{aligned} dz_t = & \left(\left.\frac{\partial r}{\partial t}\right|_{(t,x_t)} + \left.(\nabla_x \otimes r) f\right|_{(t,x_t)} \right. \\ & \left. + \frac{1}{2}\left.\left(\nabla_x^{[2]} \otimes r\right)\tilde{g}^2\right|_{(t,x_t)} \right) dt \\ & + (\nabla_x \otimes r)g|_{(t,x_t)} dW_t \end{aligned} \tag{70}$$

where

$$\tilde{g}^2 = \sum_{k=1}^{p} g_{:,k}^{[2]} \tag{71}$$

REFERENCES

[1] B. D. O. Anderson and J. B. Moore, *Optimal Filtering*. Englewood Cliffs, NJ: Prentice-Hall, 1979.

[2] Z. Bai, "Krylov subspace techniques for reduced-order modeling of large-scale dynamical systems," *Applied Numerical Mathematics*, vol. 43, pp. 9–44, 2002.

[3] C. Bruni, G. Di Pillo, and G. Koch, "Bilinear systems: An appealing class of *nearly linear* systems in theory and applications," *IEEE Trans. Automat. Contr.*, vol. 19, no. 4, pp. 334–348, 1974.

[4] R. Bucy and P. Joseph, *Filtering for Stochastic Processes with Applications to Guidance*. New York: Wiley, 1968.

[5] R. Bucy and E. Jonckheere, "Singular filtering problems," *Syst. Contr. Lett.*, no. 13, pp. 339–344, 1989.

[6] F. Carravetta, A. Germani, and M. Raimondi, "Polynomial filtering for linear discrete-time non-Gaussian systems," *SIAM J. Contr. Optim.*, vol. 34, no. 5, pp. 1666–1690, 1996.

[7] F. Carravetta, A. Germani, and M. Raimondi, "Polynomial filtering of discrete-time stochastic linear systems with multiplicative state noise," *IEEE Trans. Automat. Contr.*, vol. 42, pp. 1106–1126, 1997.

[8] F. Carravetta, A. Germani, and C. Manes, "Kalman-Bucy filtering for singular output-noise covariance," in *7th IEEE Med. Conf. on Control and Aut. (MED'99)*, Haifa, Israel, 1999, pp. 1320–1324.

[9] F. Carravetta, A. Germani, and M. K. Shuakayev, "A new suboptimal approach to the filtering problem for bilinear stochastic differential systems," *SIAM J. Contr. Optim.*, vol. 38, no. 4, pp. 1171–1203, 2000.

[10] C. K. Chui and G. Chen, *Kalman Filtering with Real-Time Applications*. New York: Springer-Verlag, 1987.

[11] G. B. Di Masi, M. Pratelli, and W. J. Runggaldier, "An approximation for the nonlinear filtering problem, with error bound," *Stochastics*, vol. 14, pp. 247–271, 1985.

[12] R. J. Elliot and R. Glowinski, "Approximations to solutions of the Zakai filtering equation," *Stoch. Anal. Appl.*, vol. 7, no. 2, pp. 145–168, 1989.

[13] A. Gelb, *Applied Optimal Estimation*. Cambridge, MA: MIT Press, 1984.

[14] A. Germani, C. Manes, and P. Palumbo, "Filtering of differential nonlinear systems via a Carleman approximation approach," in *44th IEEE Conf. on Decision and Control & European Control Conf. (CDC-ECC05)*, Seville, Spain, 2005, pp. 5917–5922.

[15] A. Germani, C. Manes, and P. Palumbo, "Polynomial extended Kalman filter," *IEEE Trans. Automat. Contr.*, vol. 50, no. 12, pp. 2059–2064, 2005.

[16] A. Germani and M. Piccioni, "Finite dimensional approximation for the equations of nonlinear filtering derived in mild form," *Applied Math. and Optim.*, vol. 16, pp. 51–72, 1987.

[17] D. J. Higham, "An algorithmic introduction to numerical simulation of stochastic differential equations," *SIAM Review*, vol. 43, no. 3, pp. 525–546, 2001.

[18] K. Ito, "Approximation of the Zakai equation for nonlinear filtering theory," *SIAM J. Contr. Optim.*, vol. 34, no. 2, pp. 620–634, 1996.

[19] K. Ito and K. Xiong, "Gaussian filters for nonlinear filtering problems," *IEEE Trans. Automat. Contr.*, vol. 45, pp. 910–927, 2000.

[20] A. H. Jazwinski, *Stochastic Processes and Filtering Theory*. New York: Academic, 1970.

[21] S. J. Julier and J. K. Uhlmann, "Unscented filtering and nonlinear estimation," *Proc. of IEEE*, vol. 92, pp. 401–422, 2004.

[22] K. Kowalski and W. H. Steeb, *Nonlinear Dynamical Systems and Carleman Linearization*. Singapore: World Scientific, 1991.

[23] H. J. Kushner, "Nonlinear filtering: The exact dynamical equations satisfied by the conditional models," *IEEE Trans. Automat. Contr.*, vol. 12, no. 3, pp. 262–267, 1967.

[24] R. S. Liptser and A. N. Shiryayev, *Statistics of Random Processes I and II*. Berlin: Springer, 1977.

[25] E. Pardoux, "Stochastic partial differential equations and filtering of diffusion processes," *Stochastics*, vol. 3, pp. 127–167, 1979.

[26] M. S. Arulampalam, S. Maskell, N. Gordon, and T. Clapp, "A tutorial on particle filters for online nonlinear/non-Gaussian bayesian tracking," *IEEE Trans. Sign. Proc.*, vol. 50, pp. 174–188, 2002.

[27] S. Sastry, *Nonlinear Systems: Analysis, Stability and Control*. New-York: Springer-Verlag, 1999.

[28] S. Svoronos, G. Stephanopoulos, and R. Aris, "Bilinear approximation of general nonlinear dynamic systems with linear inputs," *Int. J. Contr.*, vol. 31, no. 1, pp. 109–126, 1980.

[29] J. T.-H. Lo, "Global bilinearization of systems with control appearing linearly," *SIAM J. Control*, vol. 13, no. 4, pp. 879–885, 1975.

[30] W. S. Wong and S. T. Yau, "The estimation algebra of nonlinear filtering systems," in *Mathematical Control Theory*, J. Baillieul and J. C. Willems, Eds. : Springer Verlag, 1999, pp. 33–65.

[31] N. van de Wouw, H. Nijmeijer, and D. H. van Campen, "Statistical bilinearization in stochastic nonlinear dynamics," in *Proc. of the 2nd Int. Conf. on Control of Oscillations and Chaos*, 2000, vol. 3, pp. 394–399.

[32] M. Zakai, "On the optimal filtering of diffusion processes," *Z. Wahrschein. Verw. Geb.*, vol. 11, pp. 230–243, 1969.