Polynomial Filtering of Discrete-Time Stochastic Linear Systems with Multiplicative State Noise

Francesco Carravetta, Alfredo Germani, and Massimo Raimondi

***Abstract*— In this paper, the problem of finding an optimal polynomial state estimate for the class of stochastic linear models with a multiplicative state noise term is studied. For such models, a technique of state augmentation is used, leading to the definition of a general polynomial filter. The theory is developed for time-varying systems with nonstationary and non-Gaussian noises. Moreover, the steady-state polynomial filter for stationary systems is also studied. Numerical simulations show the high performances of the proposed method with respect to the classical linear filtering techniques.**

***Index Terms*— Kalman filter, Kronecker algebra, polynomial filter, stochastic bilinear systems, stochastic stability.**

I. Introduction

Systems with multiplicative state noise, also known in literature as bilinear stochastic systems (BLSS's), have been widely studied since the 1960's because, from an engineering point of view, they constitute a more adequate mathematical model for the analysis and control of some important physical processes. In particular, we stress that bilinear models are often derived from basic principles in chemistry, biology, ecology, economics, physics, and engineering [3]. Moreover, the well-known bilinear systems (BLS's) become BLSS's when the input is affected by additive noise.

In control engineering, BLS's are appealing for their better controllability with respect to the linear ones [2]. In this framework, considerable importance is devoted to control and stabilization problems, as shown in [5]–[11]. The problem of parameter estimation for BLS's and BLSS's was considered in [12]–[15].

The state estimation problem for BLSS's constitutes an important topic in all those cases in which the state itself is not available directly. In [4], the filtering problem for linear control systems is considered. In [16], the same problem, for a class of nonlinear systems including the bilinear ones, is studied, and a linear filter is obtained by considering the nonlinear term as an additive noise. BLSS's can be considered as linear systems whose dynamic matrices are a random process and vice versa [26]–[31]. In [17] and [18], following this interpretation, a linear filtering technique for the state estimation of BLSS's is proposed.

In this paper we consider the following class of BLSS's:

$$x(k+1) = A(k)x(k) + B\big(k, x(k), \xi'(k)\big) + \xi(k) \tag{1}$$

$$y(k) = C(k)x(k) + \eta(k) \tag{2}$$

where $x(k) \in \mathbb{R}^n$, $y(k) \in \mathbb{R}^m$, and $\xi'(k), \xi(k), \eta(k)$ are white sequences (not necessarily Gaussian) in $\mathbb{R}^p$, $\mathbb{R}^q$, and $\mathbb{R}^r$, respectively, A and C are matrices of suitable dimensions, whereas $B(k, \cdot, \cdot) : \mathbb{R}^n \times \mathbb{R}^p \rightarrow \mathbb{R}^n$ is a bilinear map. Moreover, we will assume the independence of $x(0)$, $\{\xi(k)\}$, $\{\xi'(k)\}$, and $\{\eta(k)\}$.

The problem we would like to face is the filtering of the state $x(k)$, given the measurement process $\{y(j), j \leq k\}$. It is well known that when $B = 0$, this problem is solved by the famous Kalman filter which yields the linear minimum variance optimal state estimate (actually optimal among all filters in the Gaussian case) [32]. The general case is, until now, unsolved. As mentioned above, a suboptimal solution can be obtained by substituting the stochastic forcing term in (1), namely

$$g(k) = B(k, x(k), \xi'(k)) + \xi(k) \tag{3}$$

by a process having the same first- and second-order properties. Indeed, it is readily proved that $\{g(k)\}$ is a white sequence so that the Kalman filter can be implemented in order to have the optimal linear estimate. Of course, the stochastic sequence given by (3) is not Gaussian so that the Kalman filter does not give the optimal estimate. Recently, the problem of finding nonlinear filters for non-Gaussian linear models has been considered. In particular, a quadratic filter is proposed in [19], and its extension to a more general polynomial case is considered in [20].

In this paper, we are able to define a filter for a BLSS such as (1) and (2), which is optimal in a class of polynomial transformations. We also stress that a Gaussian-noise setting is meaningful in the present case. The theory developed here includes, as a particular case, the one described in [20], which can be simply obtained by setting to zero the bilinear form in (1). It should also be noted that a converse point of view could be adopted in that a way of constructing a polynomial filter for BLSS's could be to compute all moments of the stochastic forcing term (3) and then using

Manuscript received August 11, 1995; revised October 29, 1996. Recommended by Associate Editor, E. Yaz. This work was supported in part by MURST and by Progetto Finalizzato Trasporti 2 under Grant 4.2.1.

F. Carravetta and M. Raimondi are with the Istituto di Analisi dei Sistemi e Informatica del CNR Viale Manzoni 30, 00185 Roma, Italy.

A. Germani is with the Dipartimento di Ingegneria Elettrica, Universitàdell'Aquila, 67100 Monteluco (L'Aquila), Italy, and Istituto di Analisi dei Sistemi e Informatica del CNR Viale Manzoni 30, 00185 Roma, Italy (e-mail: germani@iasi.rm.cnr.it).

Publisher Item Identifier S 0018-9286(97)05068-X.

the polynomial filter for linear non-Gaussian systems defined in [20]. However, this way is not convenient at all. Indeed, the computation of the moments of (3) requires the computation of the state moments. The application of the procedure described in [20] for state-moments computation leads, in this case, to a very cumbersome nonlinear equation, giving very hard implementation problems that are difficult to analyze as far as its convergence properties are concerned. In the general polynomial case, it is much more convenient to assume as a starting point for the development of the theory, the representation (just used in [17] and [18]) of the BLSS (1), (2) as a linear system with a stochastic dynamical matrix. In this framework, in order to obtain a self-contained general solution of the polynomial filtering problem for the class of the BLSS, here we will adopt just the basic strategy described in [20]. The resulting algorithm will be sufficiently general to include as a very particular case the polynomial filter for the linear non-Gaussian systems.

Roughly speaking, the method used here consists of defining a linear system whose state and output processes include Kronecker powers and products of the original state and output processes so that it is amenable to be treated with Kalman filtering theory. For this purpose, the main tool is the Kronecker algebra. Some important formulas about this subject are also deduced (e.g., the expression of the Kronecker power of a vector polynomial).

We stress that, in the present case, the existence of a stable solution for the polynomial filter is not guaranteed simply by the stability of the dynamic matrix as in the linear case.

The paper is organized as follows: in Section II, we recall some notions in estimation theory which are essential to better understanding the meaning of polynomial estimate. In this framework, we define the class of polynomial estimators and recursive algorithms which we will use later. In Section III, we make precise the problem statement, and Sections IV and V explain how to build up the augmented system. In Section VI, the way to implement the filter on the augmented system is described. In Section VII, we present the stationary case and the steady-state theory. Section VIII contains some remarks about the computer implementation of the algorithm. In Section IX, numerical simulations are presented showing the high performance of polynomial filtering with respect to the standard linear methods. Two appendixes are included: Appendix A, containing the proof of the main theorem of the paper defining the augmented system, and Appendix B, where the main definitions and properties about Kronecker algebra are reported together with some new results.

II. POLYNOMIAL ESTIMATES

Our aim is to improve the performance of standard linear filtering for the class of the BLSS (1), (2). For this purpose we will look for the optimal filter among the class of estimators constituted by all the fixed-degree causal polynomial transformations of the measurements. We now clarify this point by giving some definitions which will be useful in the following.

Let $(\Omega, \mathcal{F}, P)$ be a probability space. For any given sub-σ algebra $\mathcal{G}$ of $\mathcal{F}$ and integer p, let us denote by $L^p(\mathcal{G}, n)$ the Banach space of the n-dimensional $\mathcal{G}$-measurable random variables with finite pth moment as

$$L^p(\mathcal{G}, n) = \Big\{ X : \Omega \to \mathbb{R}^n,\ \mathcal{G} - \text{measurable}, \int_\Omega \|X(\omega)\|^p \, dP(\omega) < +\infty \Big\}$$

where $\|\cdot\|$ is the euclidean norm in $\mathbb{R}^n$. Moreover, when $\mathcal{G}$ is the σ-algebra generated by a random variable $Y : \Omega \to \mathbb{R}^m$, that is $\mathcal{G} = \sigma(Y)$, we will use the notation $L^p(Y, n)$ to indicate $L^p(\sigma(Y), n)$. Finally, if M is a closed subspace of $L^2(\mathcal{F}, n)$, we will use the symbol $\Pi(X/M)$ to indicate the orthogonal projection of $X \in L^2(\mathcal{F}, n)$ onto M.

As is well known, the optimal minimum variance estimate of a random variable $X \in L^2(\mathcal{F}, n)$ with respect to a random variable Y, that is $\Pi(X/L^2(Y, n))$, is given by the conditional expectation (C.E.) $E(X/Y)$. If X and Y are jointly Gaussian, then the C.E. is the following affine transformation of Y:

$$E(X/Y) = E(X) + E(X\tilde{Y}^T)\big(E(\tilde{Y}\tilde{Y}^T)\big)^{-1}\tilde{Y} \tag{4}$$

where $\tilde{Y} = Y - E(Y)$.

Moreover, defining

$$Y' = \begin{bmatrix} 1 \\ Y \end{bmatrix}, \qquad Y \in L^2(\mathcal{F}, m)$$

(4) also can be interpreted as the projection on the subspace

$$\mathcal{L}(Y', n) = \{Z : \Omega \to \mathbb{R}^n / \exists\, A \in \mathbb{R}^{n\times(m+1)} \text{ such that } Z = AY'\} \subset L^2(Y', n) = L^2(Y, n).$$

Unfortunately, in the non-Gaussian case, no simple characterization of the C.E. can be achieved. Consequently, it is worthwhile to consider suboptimal estimates which have a simpler mathematical structure that allows the treatment of real data. The simplest suboptimal estimate is the optimal affine one, that is $\Pi(X/\mathcal{L}(Y', n))$, which is still given by the right-hand side (RHS) of (4). In the following, such an estimate will be denoted with $\hat{X}$ and shortly called the optimal linear estimate. Suboptimal estimates comprised between the optimal linear and the C.E. can be considered by projecting onto subspaces, greater than $\mathcal{L}(Y', n)$, like subspaces of polynomial transformations of Y. We define the *ith degree space of the polynomial transformations of* $Y \in L^{2i}(\mathcal{F}, m)$ as the following (closed) subspace of $L^2(\mathcal{F}, n)$:

$$\mathcal{P}_i(Y, n) = \{Z \in L^2(\mathcal{F}, n) :\ Z = T_1 Y + T_2 Y^{[2]} + \cdots + T_i Y^{[i]} + b, T_j \in \mathbb{R}^{n\times m^j},\ b \in \mathbb{R}^n\}$$

where the symbol $Y^{[i]}$ denotes the Kronecker power (see Appendix B). By defining the vector

$$\mathcal{Y}_i = \begin{bmatrix} 1 \\ Y \\ \vdots \\ Y^{[i]} \end{bmatrix} \in L^2(\mathcal{F}, 1 + m + \cdots + m^i) \tag{5}$$

we have that

$$\mathcal{P}_i(Y, n) = \mathcal{L}(\mathcal{Y}_i, n).$$

We define the *ith-order polynomial estimate* as the random variable $\Pi(X/\mathcal{L}(\mathcal{Y}_i, n))$. Since

$$\mathcal{L}(\mathcal{Y}_1, n) \subset \mathcal{L}(\mathcal{Y}_2, n) \subset \cdots \subset \mathcal{L}(\mathcal{Y}_i, n)$$

the polynomial estimate improves (in terms of error variance) the linear one. Let $\mathcal{H}$ be the closure in $L^2(\mathcal{F}, n)$ of

$$\mathcal{P}(Y, n) \triangleq \bigcup_{i=0}^{\infty} \mathcal{L}(\mathcal{Y}_i, n)$$

since, in general, for $\mathcal{H} \neq L^2(\sigma(Y), n)$ we cannot assert that the polynomial estimate "approaches" the optimal one for increasing polynomial degrees. Nevertheless, the C.E. of X can be decomposed as

$$E(X/Y) = \Pi(X/\mathcal{H}) + \Pi(X/\mathcal{H}^{\perp})$$

where $\mathcal{H}^{\perp}$ is the orthogonal subspace of $\mathcal{H}$. From the previous relation we infer that the polynomial estimate can be considered as an approximation of the optimal one only when $\|\Pi(X/\mathcal{H}^{\perp})\|_{L^2}$ is suitably small. However, the polynomial estimate always yields an improvement with respect to the performance of a linear estimator. Moreover, we can calculate it by suitably modifying the space of observed random variables and using (4)

$$\Pi(X/\mathcal{P}_i(Y, n)) = E\left(X\overline{\mathcal{Y}}_i^T\right)E\left(\overline{\mathcal{Y}}_i\overline{\mathcal{Y}}_i^T\right)^{-1}\overline{\mathcal{Y}}_i + E(X) \tag{6}$$

where

$$\overline{\mathcal{Y}}_i \triangleq \mathcal{Y}_i - E(\mathcal{Y}_i).$$

Now, let us consider a *sequence* $\{X_i\}$ of random variables in $L^2(\mathcal{F}, n)$ and another $\{Y_i\}$ of observed ones in $L^2(\mathcal{F}, m)$. The problem of estimating X_k, given $\{Y_0, Y_1, \cdots, Y_k\}$, can be solved by defining the vector

$$Y_{e,k} = \begin{bmatrix} Y_0 \\ Y_1 \\ \vdots \\ Y_k \end{bmatrix}$$

and applying (4) with $\overline{Y} = Y_{e,k} - E(Y_{e,k})$ so that the optimal linear estimate of X_k is obtained. When the joint sequence $\{X_i, Y_i\}$ is Gaussian, (4) yields the optimal estimate $E(X_k/Y_0, \cdots, Y_k)$. Similarly, if the moments $E(Y_{e,k}^{[j]})$, $j = 1, \cdots, 2h$, $k = 0, 1, \cdots$, are finite and known, the hth-order polynomial estimate can be obtained by extending the vector $Y_{e,k}$ as in (5). However, such a method is highly inefficient, because it leads to a fast growth of the dimensions of involved matrices so that it does not result in being very useful from an application point of view. A more realistic approach should consist of searching for a recursive algorithm able to yield the above estimates. For this purpose, we give the following definition.

Definition 2.1: We say that the estimate $\hat{X}_k$ of X_k (not necessarily optimal) is recursive of Δ order if there exists a sequence of random variables $\{Z_k\}$ and transformations $\mathcal{R}_k$, T_k, such that the following equations hold:

$$Z_k = \mathcal{R}_k(Z_{k-1}, Y_k, Y_{k-1}, \cdots, Y_{k-\Delta}) \tag{7}$$

$$\hat{X}_k = T_k(Z_k). \tag{8}$$

As is well known, in the Gaussian case and when the sequences $\{X_i\}$, $\{Y_i\}$ are the state and output evolutions of a linear discrete-time dynamic system, the optimal estimator of the state satisfies a recursive equation as in (7), (8), with $\mathcal{R}_k$ given by a suitable linear transformation, $T_k = I$ (I identity matrix) and $\Delta = 0$ (the Kalman filter). The same equations give, in the non-Gaussian case, the optimal linear estimate.

In the next section, we will prove that when $\{X_k\}$ and $\{Y_k\}$ are the state and output processes, respectively, for a BLSS as in (1), (2), it is possible to find a structure (7), (8) where $\mathcal{R}_k$ has the form

$$\mathcal{R}_k(Z_{k-1}, Y_k, Y_{k-1}, \cdots, Y_{k-\Delta}) = \mathcal{R}'_k(Z_{k-1}, F_k(Y_k, Y_{k-1}, \cdots, Y_{k-\Delta}))$$

with $\mathcal{R}'_k$ *linear* and F_k *polynomial* such that (7) and (8) yield the sequence $\{\hat{X}_k\}$ of optimal estimates in a certain subclass of all the polynomial transformations of fixed and finite degree. In order to define more precisely this subclass of polynomial estimators, we need to give some preliminary definitions.

Consider the above-defined vector $Y_{e,k}$, and let $\nu > 0$ be a fixed integer; we define the subspace

$$\mathcal{P}_{\nu,\Delta}(Y_{e,k}, n) \subset \mathcal{P}_{\nu}(Y_{e,k}, n)$$

as

$$\mathcal{P}_{\nu,\Delta}(Y_{e,k}, n) = \left\{ \xi \in L^2(\mathcal{F}, n) : \xi = \sum_{0 \le h_1 + \cdots + h_\nu \le \nu} \cdot \sum_{\substack{0 \le i_1, \cdots, i_\nu \le k \\ \max |i_s - i_p| \le \Delta}} c_{h,l,i,j}\left(Y_{k-i_1}^{[h_1]} \otimes \cdots \otimes Y_{k-i_\nu}^{[h_\nu]}\right) \right\}$$

where the $c_{h,l,i,j}$'s are suitably dimensioned matrices. Since the subspace $\mathcal{P}_{\nu,\Delta}(Y_{e,k}, n)$ is finite dimensional, and therefore closed, we have that for any k, it is possible to orthoproject there the L^2 random variable X_k. Then, we can give the following definition.

Definition 2.2: The random variable $\hat{X}_k^{(\nu,\Delta)}$, given by

$$\hat{X}_k^{(\nu,\Delta)} = \Pi(X_k/\mathcal{P}_{\nu,\Delta}(Y_{e,k}, n))$$

is said to be the (ν, Δ)-order polynomial estimate of X_k.

The random variable $\hat{X}_k^{(\nu,\Delta)}$ represents the optimal estimate of X_k among all the ν-degree polynomials, including cross products between observations which lie in a time window of width Δ. Since

$$\mathcal{L}(Y_{e,k}, n) = \mathcal{P}_{1,0}(Y_{e,k}, n)$$

and

$$\mathcal{P}_{\nu,\Delta}(Y_{e,k},n) \subset \mathcal{P}_{\nu+1,\Delta}(Y_{e,k},n) \quad \nu = 1,2,\cdots \quad \Delta = 0,1,\cdots$$
$$\mathcal{P}_{\nu,\Delta}(Y_{e,k},n) \subset \mathcal{P}_{\nu,\Delta+1}(Y_{e,k},n) \quad \nu = 1,2,\cdots \quad \Delta = 0,1,\cdots$$

the result is that the estimate quality had to improve for increasing ν and/or Δ.

III. The Problem

The problem we are faced with is the filtering one for the following class of stochastic discrete-time bilinear systems:

$$x(k+1) = A(k)x(k) + B(k, x(k), \xi'(k)) + \xi(k) \qquad x(0) = \overline{x} \tag{9}$$
$$y(k) = C(k)x(k) + \eta(k) \tag{10}$$

where, for any k

$$x(k) \in \mathbb{R}^n, \quad y(k) \in \mathbb{R}^m, \quad \xi(k) \in \mathbb{R}^n \quad \xi'(k) \in \mathbb{R}^u, \quad \eta(k) \in \mathbb{R}^m.$$

Moreover, $A(k) \in \mathbb{R}^{n\times n}, C(k) \in \mathbb{R}^{m\times n}$, whereas $B(k,\cdot,\cdot)$ is a bilinear form in $\mathbb{R}^{n\times u}$. The random variable $\overline{x}$ (the initial condition) and the random sequences $\{\xi(k)\}$, $\{\xi'(k)\}$, $\{\eta(k)\}$ satisfy the following conditions for any $k \geq 0$.

1) There exists an integer $\nu \geq 1$ such that
$$E(\|\overline{x}\|^{2\nu}) < \infty, \qquad E(\|\xi(k)\|^{2\nu}) < \infty$$
$$E(\|\xi'(k)\|^{2\nu}) < \infty, \qquad E(\|\eta(k)\|^{2\nu}) < \infty.$$
2) The initial state $\overline{x}$ forms, together with the sequences $\{\xi(k)\}$, $\{\xi'(k)\}$, $\{\eta(k)\}$, a family of independent random variables.
3) All random sequences $\{\xi(k)\}$, $\{\xi'(k)\}$, $\{\eta(k)\}$ are white.

It should be noted that the vector $B(k, x(k), \xi'(k))$, in (9), due to the bilinearity hypothesis, can be written in the form

$$B(k, x(k), \xi'(k)) = \sum_{i=1}^{u} B_i(k)x(k)\xi'_i(k)$$

where $B_i(k) \in \mathbb{R}^{n\times n}$ is a suitable matrix and $\xi'_i(k)$ denotes the ith entry of the vector $\xi'(k)$. Then, system (9), (10) can be rewritten as

$$x(k+1) = \tilde{A}(k)x(k) + \xi(k), \qquad x(0) = \overline{x} \tag{11}$$
$$y(k) = C(k)x(k) + \eta(k) \tag{12}$$

where

$$\tilde{A}(k) = A(k) + \sum_{i=1}^{u} B_i(k)\xi'_i(k). \tag{13}$$

System (11), (12) is a linear system with a stochastic dynamic matrix. It is equivalent to the original bilinear system because it generates exactly the same state and output processes. Hence, in order to obtain a state estimate, we can consider this latter system in place of (9), (10).

Our goal is the determination of a discrete-time filter, that is a recursive algorithm in the form (7), (8), which gives at any time k the optimal polynomial state estimate of (ν, Δ)-order (see Definition 2.1) for the system (9), (10), given all the available observations at time k: $y(0), \cdots, y(k)$.

In the next sections, it will be shown how to obtain such a polynomial filter. Moreover, in the constant parameter case, conditions will be defined assuring the existence of a stationary polynomial filter.

The approach that follows goes along the same line as in [20], consisting essentially of the transpose of the originary problem to a linear filtering one, solvable by means of the Kalman filter. In order to define a polynomial estimator which also takes into account cross products between observations at different times, we need to introduce the following so-called "extended memory system."

IV. The Extended Memory System

Given the system (9), (10), and having chosen an integer $\Delta \geq 0$, let us define the following vectors:

$$x_e(k) = \begin{bmatrix} x(k) \\ y(k-1) \\ \vdots \\ y(k-\Delta) \end{bmatrix} \in \mathbb{R}^q, \qquad y_e(k) = \begin{bmatrix} y(k) \\ y(k-1) \\ \vdots \\ y(k-\Delta) \end{bmatrix} \in \mathbb{R}^p \tag{14}$$

with $q = n + m\Delta$ and $p = (\Delta+1)m$. Taking into account the equivalent equations (11), (12), we have that $x_e(k)$, $y_e(k)$ satisfy the following relations:

$$x_e(k+1) = \tilde{A}_e(k)x_e(k) + FN(k) \tag{15}$$
$$y_e(k) = C_e(k)x_e(k) + GN(k) \tag{16}$$

where

$$\tilde{A}_e(k) = \begin{bmatrix} \tilde{A}(k) & 0 & \cdots & \cdots & 0 \\ C(k) & 0 & \cdots & \cdots & 0 \\ 0 & I & \ddots & & \vdots \\ \vdots & & \ddots & \ddots & \vdots \\ 0 & \cdots & \cdots & I & 0 \end{bmatrix} \tag{17}$$

$$C_e(k) = \begin{bmatrix} C(k) & 0 & \cdots & 0 \\ 0 & I & \ddots & \\ \vdots & & \ddots & \vdots \\ 0 & & \cdots & I \end{bmatrix} \tag{18}$$

$$N(k) = \begin{bmatrix} \xi(k) \\ \eta(k) \end{bmatrix}, \qquad F = \begin{bmatrix} I & 0 \\ 0 & I \\ \vdots & 0 \\ \vdots & \vdots \\ 0 & 0 \end{bmatrix}, \qquad G = \begin{bmatrix} 0 & I \\ 0 & 0 \\ \vdots & \vdots \\ 0 & 0 \end{bmatrix}.$$

We call (15), (16) an *extended memory system*.

In the next section, which contains the main result of this paper, we will be able to derive the evolution of the Kronecker powers of the above-defined extended state and output.

V. The Augmented System

Let us consider the integer $\nu \geq 1$ for which Property 1) of Section III holds. We define the *augmented observation* as the vector $\mathcal{Y} \in \mathbb{R}^{\mu}$, $\mu = p + p^2 + \cdots + p^{\nu}$

$$\mathcal{Y}(k) = \begin{bmatrix} y_e(k) \\ y_e^{[2]}(k) \\ \vdots \\ y_e^{[\nu]}(k) \end{bmatrix}. \tag{19}$$

Moreover, we define the *augmented state* as the vector $\mathcal{X} \in \mathbb{R}^{\lambda}$, where $\lambda = q + q^2 + \cdots + q^{\nu}$

$$\mathcal{X}(k) = \begin{bmatrix} x_e(k) \\ x_e^{[2]}(k) \\ \vdots \\ x_e^{[\nu]}(k) \end{bmatrix}. \tag{20}$$

Now, for a bilinear system such as (9), (10), satisfying the Properties 1) and 2) of Section III, let us build up the extended memory system (15), (16), the augmented observations (19), and the augmented state (20). Let $I(l)$ and $I(\alpha, \beta)$ be the identity matrix in $\mathbb{R}^l$ and in $\mathbb{R}^{\alpha^\beta}$, respectively. Then, the following theorem holds.

Theorem 5.1: The processes $\{\mathcal{Y}(k)\}$ and $\{\mathcal{X}(k)\}$ defined in (19) and (20) satisfy the following equations:

$$\mathcal{X}(k+1) = \mathcal{A}(k)\mathcal{X}(k) + \mathcal{U}(k) + \mathcal{F}(k), \qquad \mathcal{X}(0) = \overline{\mathcal{X}} \tag{21}$$

$$\mathcal{Y}(k) = \mathcal{C}(k)\mathcal{X}(k) + \mathcal{V}(k) + \mathcal{G}(k) \tag{22}$$

where

$$\overline{\mathcal{X}} = \begin{bmatrix} \overline{x}_e \\ \overline{x}_e^{[2]} \\ \cdots \\ \overline{x}_e^{[\nu]} \end{bmatrix}, \qquad \mathcal{U}(k) = \begin{bmatrix} FE(N(k)) \\ F^{[2]}E(N^{[2]}(k)) \\ \vdots \\ F^{[\nu]}E(N^{[\nu]}(k)) \end{bmatrix}$$

$$\mathcal{V}(k) = \begin{bmatrix} GE(N(k)) \\ G^{[2]}E(N^{[2]}(k)) \\ \vdots \\ G^{[\nu]}E(N^{[\nu]}(k)) \end{bmatrix}$$

$$\mathcal{A}(k) = \begin{bmatrix} E(\tilde{A}_e(k)) & 0 & \cdots & 0 \\ H_{2,1}(k) & E(\tilde{A}_e^{[2]}(k)) & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots \\ H_{\nu,1}(k) & H_{\nu,2}(k) & \cdots & E(\tilde{A}_e^{[\nu]}(k)) \end{bmatrix}$$

$$\mathcal{C}(k) = \begin{bmatrix} C_e(k) & 0 & \cdots & 0 \\ L_{2,1}(k) & C_e^{[2]}(k) & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots \\ L_{\nu,1}(k) & L_{\nu,2}(k) & \cdots & C_e^{[\nu]}(k) \end{bmatrix}$$

$$H_{i,l}(k) = M_{i-l}^{i}(q)(F^{[i-l]} \otimes E(\tilde{A}_e^{[l]}(k))) \cdot \left(E(N^{[i-l]}(k)) \otimes I_{q,l}\right)$$

$$L_{i,l}(k) = M_{i-l}^{i}(p)(G^{[i-l]} \otimes C_e^{[l]}(k)) \cdot \left(E(N^{[i-l]}(k)) \otimes I_{q,l}\right).$$

Moreover, $\{\mathcal{F}(k)\}$, $\{\mathcal{G}(k)\}$ are zero mean uncorrelated sequences such that

$$E(\mathcal{F}(k)\mathcal{G}^T(j)) = 0, \qquad k \neq j \tag{23}$$

whose auto- and cross-covariance matrices

$$\mathcal{Q}(k) \triangleq E(\mathcal{F}(k)\mathcal{F}(k)^T), \qquad \mathcal{R}(k) \triangleq E(\mathcal{G}(k)\mathcal{G}(k)^T)$$
$$\mathcal{J}(k) \triangleq E(\mathcal{F}(k)\mathcal{G}(k)^T)$$

have the following block structure:

$$\mathcal{Q}(k) = \begin{bmatrix} \mathcal{Q}_{1,1}(k) & \mathcal{Q}_{1,2}(k) & \cdots & \mathcal{Q}_{1,\nu}(k) \\ \mathcal{Q}_{2,1}(k) & \mathcal{Q}_{2,2}(k) & \cdots & \mathcal{Q}_{2,\nu}(k) \\ \cdots & \cdots & \cdots & \cdots \\ \mathcal{Q}_{\nu,1}(k) & \mathcal{Q}_{\nu,2}(k) & \cdots & \mathcal{Q}_{\nu,\nu}(k) \end{bmatrix}$$

$$\mathcal{R}(k) = \begin{bmatrix} \mathcal{R}_{1,1}(k) & \mathcal{R}_{1,2}(k) & \cdots & \mathcal{R}_{1,\nu}(k) \\ \mathcal{R}_{2,1}(k) & \mathcal{R}_{2,2}(k) & \cdots & \mathcal{R}_{2,\nu}(k) \\ \cdots & \cdots & \cdots & \cdots \\ \mathcal{R}_{\nu,1}(k) & \mathcal{R}_{\nu,2}(k) & \cdots & \mathcal{R}_{\nu,\nu}(k) \end{bmatrix}$$

$$\mathcal{J}(k) = \begin{bmatrix} \mathcal{J}_{1,1}(k) & \mathcal{J}_{1,2}(k) & \cdots & \mathcal{J}_{1,\nu}(k) \\ \mathcal{J}_{2,1}(k) & \mathcal{J}_{2,2}(k) & \cdots & \mathcal{J}_{2,\nu}(k) \\ \cdots & \cdots & \cdots & \cdots \\ \mathcal{J}_{\nu,1}(k) & \mathcal{J}_{\nu,2}(k) & \cdots & \mathcal{J}_{\nu,\nu}(k) \end{bmatrix}$$

where, for $r, s = 1, \cdots, \nu$, $\mathcal{Q}_{r,s}(k)$, $\mathcal{R}_{r,s}(k)$, $\mathcal{J}_{r,s}(k)$ are $q^r \times q^s$, $p^r \times p^s$, $q^r \times p^s$ matrices, respectively, given by the following formulas:

$$\mathcal{Q}_{r,s}(k) = \mathcal{Q}_{r,s}^{(1)}(k) + \mathcal{Q}_{r,s}^{(2)}(k) \tag{24}$$

$$\begin{aligned} \mathcal{Q}_{r,s}^{(1)}(k) = \sum_{l=0}^{r-1}\sum_{j=0}^{s-1} & M_{r-l}^{r}(q)\mathrm{st}^{-1}(((I_{q,s-j} \otimes C_{q^r,q^j}^T) \\ & \cdot (F^{[r+s-l-j]} \otimes E(\tilde{A}_e^{[l+j]}(k))) \\ & \cdot (I_{n+m,s-j} \otimes C_{(n+m)^{r-l}q^l,q^j}) \\ & \cdot (I_{n+m,s-j} \otimes C_{(n+m)^{r-l}q^l,q^j}^T) \\ & \cdot ((E(N^{[s+r-j-l]}(k)) \\ & - E(N^{[s-j]}(k)) \otimes E(N^{[r-l]}(k))) \otimes I_{q,l+j}) \\ & \cdot C_{q^l,q^j})E(x_e^{[l+j]}(k)))M_{s-j}^{s}(q)^T \end{aligned} \tag{25}$$

$$\begin{aligned} \mathcal{Q}_{r,s}^{(2)}(k) = \sum_{l=1}^{r}\sum_{j=1}^{s} & M_{r-l}^{r}(q)\mathrm{st}^{-1}(((I_{q,s-j} \otimes C_{q^r,q^j}^T) \\ & \cdot (F^{[r+s-l-j]} \otimes (E(\tilde{A}_e^{[l+j]}(k)) \\ & - E(\tilde{A}_e^{[l]}(k)) \otimes E(\tilde{A}_e^{[j]}(k)))) \\ & \cdot (I_{n+m,s-j} \otimes C_{(n+m)^{r-l}q^l,q^j}) \\ & \cdot (I_{(n+m),s-j} \otimes C_{(n+m)^{r-l}q^l,q^j}^T) \\ & \cdot (E(N^{[s-j]}(k)) \otimes E(N^{[r-l]}(k)) \otimes I_{q,l+j}) \\ & \cdot C_{q^l,q^j})E(x_e^{[l+j]}(k)))M_{s-j}^{s}(q)^T \end{aligned} \tag{26}$$

$$\mathcal{R}_{r,s}(k) = \sum_{l=0}^{r-1}\sum_{j=0}^{s-1} M_{r-l}^{r}(p)\mathrm{st}^{-1}(((I_{p,s-j}\otimes C_{p^r,p^j}^{T}) \cdot (G^{[r+s-l-j]}\otimes C_e^{[l+j]}(k)) \cdot (I_{n+m,s-j}\otimes C_{(n+m)^{r-l}q^l,q^j}) \cdot (I_{n+m,s-j}\otimes C_{(n+m)^{r-l}q^l,q^j}^{T}) \cdot ((E(N^{[s+r-j-l]}(k)) - E(N^{[s-j]}(k))\otimes E(N^{[r-l]}(k)))\otimes I_{q,l+j}) \cdot C_{q^l,q^j})E(x_e^{[l+j]}(k)))M_{s-j}^{s}(p)^T \tag{27}$$

$$\mathcal{J}_{r,s}(k) = \sum_{l=0}^{r-1}\sum_{j=0}^{s-1} M_{r-l}^{r}(q)\mathrm{st}^{-1}(((I_{p,s-j}\otimes C_{q^r,p^j}^{T}) \cdot (G^{[s-j]}\otimes F^{[r-l]}\otimes E(\tilde{A}_e^{[l]}(k))\otimes C_e{}^{[j]}(k)) \cdot (I_{n+m,s-j}\otimes C_{(n+m)^{r-l}q^l,q^j}) \cdot (I_{n+m,s-j}\otimes C_{(n+m)^{r-l}q^l,q^j}^{T}) \cdot ((E(N^{[s+r-j-l]}(k)) - E(N^{[s-j]}(k))\otimes E(N^{[r-l]}(k)))\otimes I_{q,l+j}) \cdot C_{q^l,q^j})E(x_e^{[l+j]}(k)))M_{s-j}^{s}(p)^T. \tag{28}$$

Proof: See Appendix A. □

We call system (21), (22) an *augmented system*. It is a classical time-varying stochastic linear system. Its state and observation noises are zero mean uncorrelated sequences and are also mutually uncorrelated at different times. For these noises we are able to calculate their auto- and cross covariances. Hence, for the augmented system the optimal linear state estimate can be calculated by means of the Kalman filter equations. In order to proceed along this way, we first need to determine the quantities $E(\tilde{A}_e^{[i]}(k))$ and $E(x_e^{[i]}(k))$ for $i = 1, \cdots, 2\nu$, which appear in the augmented system matrices and in (23)–(28).

The matrices $E(\tilde{A}_e^{[i]}(k))$, $i = 1, \cdots, 2\nu$ can be recursively calculated from $E(\tilde{A}^{[i]}(k))$, as stated in the following theorem.

Theorem 5.2: Let, for $i = 1, \cdots, 2\nu$

$$\overline{Q}_i(k) \triangleq E(\tilde{A}_e^{[i]}(k))$$

and then we have

$$\overline{Q}_i(k) = \begin{bmatrix} Q_{1,i-1}(k) & 0 & 0 \\ C(k)\otimes \overline{Q}_{i-1}(k) & 0 & 0 \\ 0 & I(m\cdot(\Delta-1))\otimes\overline{Q}_{i-1}(k) & 0\end{bmatrix} \tag{29}$$

where $Q_{h,l}(k)$, $h \geq 1$, $l \geq 0$ are given by the following recursive equations:

$$Q_{h,l}(k) = (C_{q,n^h}^{T}\otimes I_{q,l-1})T_{h,l}(k)(C_{q,n^h}\otimes I_{q,l-1}) \tag{30}$$

$$T_{h,l}(k) = \begin{bmatrix} Q_{h+1,l-1}(k) & 0 & 0 \\ C(k)\otimes Q_{h,l-1}(k) & 0 & 0 \\ 0 & I(m\cdot(\Delta-1))\otimes Q_{h,l-1} & 0\end{bmatrix} \tag{31}$$

with initial conditions

$$\overline{Q}_1(k) = E(\tilde{A}_e(k)) = \begin{bmatrix} A(k)+\sum_{r=1}^{u}B_r(k)E(\xi_l^r(k)) & 0 & 0 \\ C(k) & 0 & 0 \\ 0 & I(m\cdot(\Delta-1)) & 0\end{bmatrix} \tag{32}$$

$$Q_{s,0}(k) = E(\tilde{A}^{[s]}(k)), \qquad s = 1, 2, \cdots, i. \tag{33}$$

Proof: First of all, note that the matrix $\tilde{A}_e(k)$, defined in (17), can be rewritten in the compact form

$$\tilde{A}_e(k) = \begin{bmatrix} \tilde{A}(k) & 0 & 0 \\ C(k) & 0 & 0 \\ 0 & I(m\cdot(\Delta-1)) & 0\end{bmatrix} \tag{34}$$

where the null blocks are suitably dimensioned, $\tilde{A}(k)$ is defined in (13), and $I(m\cdot(\Delta-1))$ denotes as usual the identity matrix in $\mathbb{R}^{m(\Delta-1)}$ (we conventionally assume that it vanishes for $\Delta = 1$). From (34), and for $i = 1, \cdots, 2\nu$, (35) follows, as shown at the bottom of the page. Moreover, for any pair of integers h, l, using Theorem B.3 and Property (93c), we have

$$\begin{aligned} E(\tilde{A}^{[h]}(k)\otimes\tilde{A}_e^{[l]}(k)) &= E(\tilde{A}^{[h]}(k)\otimes\tilde{A}_e(k)\otimes\tilde{A}_e^{[l-1]}(k)) \\ &= E((C_{q,n^h}^{T}(\tilde{A}_e(k)\otimes\tilde{A}^{[h]}(k))C_{q,n^h})\otimes\tilde{A}_e^{[l-1]}(k)) \\ &= E((C_{q,n^h}^{T}\otimes I_{q,l-1}) \cdot (((\tilde{A}_e(k)\otimes\tilde{A}^{[h]}(k))C_{q,n^h})\otimes\tilde{A}_e^{[l-1]}(k))) \\ &= (C_{q,n^h}^{T}\otimes I_{q,l-1}) \cdot E(\tilde{A}_e(k)\otimes\tilde{A}^{[h]}(k)\otimes\tilde{A}_e^{[l-1]}(k))(C_{q,n^h}\otimes I_{q,l-1}) \\ &= (C_{q,n^h}^{T}\otimes I_{q,l-1})T_{h,l}(k)(C_{q,n^h}\otimes I_{q,l-1}) \end{aligned} \tag{36}$$

where

$$T_{h,l}(k) \triangleq E(\tilde{A}_e(k)\otimes\tilde{A}^{[h]}(k)\otimes\tilde{A}_e^{[l-1]}(k)).$$

$$\begin{aligned} E(\tilde{A}_e^{[i]}(k)) = E(\tilde{A}_e(k)\otimes\tilde{A}_e^{[i-1]}(k)) &= \begin{bmatrix} E(\tilde{A}(k)\otimes\tilde{A}_e^{[i-1]}(k)) & 0 & 0 \\ E(C(k)\otimes\tilde{A}_e^{[i-1]}(k)) & 0 & 0 \\ 0 & I(m\cdot(\Delta-1))\otimes E(\tilde{A}_e^{[i-1]}(k)) & 0\end{bmatrix} \\ &= \begin{bmatrix} E(\tilde{A}(k)\otimes\tilde{A}_e^{[i-1]}(k)) & 0 & 0 \\ C(k)\otimes\overline{Q}_{i-1}(k) & 0 & 0 \\ 0 & I(m\cdot(\Delta-1))\otimes\overline{Q}_{i-1}(k) & 0\end{bmatrix} \end{aligned} \tag{35}$$

Taking into account (34), we have the resulting (37), as shown at the bottom of the page, where

$$Q_{h,l}(k) \triangleq E(\tilde{A}^{[h]}(k) \otimes \tilde{A}_e^{[l]}(k)). \quad (38)$$

Equation (37) substituted in (36) yields, by exploiting (38) and (35), (30) and (31). From (38) we have

$$Q_{1,i-1} = E(\tilde{A}(k) \otimes \tilde{A}_e^{[i-1]}(k))$$

and substituting this in (35), we obtain (29). Finally, note that from (13) and taking into account (34), the initial condition (32) follows. Moreover, from (29)–(31) we infer that to compute $Q_{1,i-1}(k)$ it is enough to know the matrices $Q_{s,0}(k)$, for $s = 1, 2, \cdots, i$, which are given by (33), as immediately follows from (38). □

Theorem 5.2 allows us to compute recursively the matrices $E(\tilde{A}_e^{[i]}(k))$ for $i = 1, \cdots, 2\nu$, from the initial conditions (32), (33). Condition (32) is immediately given from the data, whereas to obtain (33) we use the following result.

Theorem 5.3: The matrices $E(\tilde{A}^{[i]}(k))$ are given by the following formula:

$$E(\tilde{A}^{[i]}(k)) = \mathrm{st}^{-1}\Bigg(D_{n,n}^{(i)} \sum_{\substack{h_0,\cdots,h_u \geq 0\\ h_0+h_1+\cdots+h_u=i}} M_{h_0,\cdots,h_u}^i \cdot ((\mathrm{st}(A(k)))^{[h_0]} \otimes (\mathrm{st}(B_1(k)))^{[h_1]} \otimes \cdots \otimes \cdot (\mathrm{st}(B_u(k)))^{[h_u]}) E\Bigg(\prod_{j=1}^{u} \xi_j'^{\,h_j}(k)\Bigg)\Bigg). \quad (39)$$

Proof: By applying (106) and Corollary B.8, and by exploiting (13), we have

$$\begin{aligned} E(A(k) + \sum_{j=1}^{u} B_j(k)\xi_j'(k))^{[i]} &= \mathrm{st}^{-1}\Bigg(E\Bigg(\mathrm{st}\Bigg(A(k) + \sum_{j=1}^{u} B_j(k)\xi_j'(k)\Bigg)^{[i]}\Bigg)\Bigg) \\ &= \mathrm{st}^{-1}\Bigg(D_{n,n}^{(i)} E\Bigg(\Bigg(\mathrm{st}\Bigg(A(k) + \sum_{j=1}^{u} B_j(k)\xi_j'(k)\Bigg)\Bigg)^{[i]}\Bigg)\Bigg) \\ &= \mathrm{st}^{-1}\Bigg(D_{n,n}^{(i)} E\Bigg(\Bigg(\mathrm{st}(A(k)) + \sum_{j=1}^{u} \mathrm{st}(B_j(k))\xi_j'(k)\Bigg)^{[i]}\Bigg)\Bigg) \\ &= \mathrm{st}^{-1}\Bigg(D_{n,n}^{(i)} \sum_{\substack{h_0,\cdots,h_u \geq 0\\ h_0+h_1+\cdots+h_u=i}} M_{h_0,\cdots,h_u}^i \cdot ((\mathrm{st}(A(k)))^{[h_0]} \otimes (\mathrm{st}(B_1(k)))^{[h_1]} \otimes \cdots \otimes \cdot (\mathrm{st}(B_u(k)))^{[h_u]}) E\Bigg(\prod_{j=1}^{u} \xi_j'^{\,h_j}(k)\Bigg)\Bigg). \end{aligned}$$

□

As far as $E(x_e^{[i]}(k))$, $i = 1, \cdots, 2\nu$, the vectors appearing in the expressions of the augmented noise covariances, are concerned, the following theorem shows that their calculation is possible by means of a recursive algorithm.

Theorem 5.4: The vector of the expected values $\mathcal{M}_{2\nu}(k)$, defined as

$$\mathcal{M}_{2\nu}(k) = \begin{bmatrix} E(x_e(k)) \\ \vdots \\ E(x_e^{[2\nu]}(k)) \end{bmatrix}$$

satisfies the following recursive equation:

$$\begin{aligned} \mathcal{M}_{2\nu}(k+1) &= \mathcal{A}_{2\nu}(k)\mathcal{M}_{2\nu}(k) + \mathcal{U}_{2\nu}(k) \\ \mathcal{M}_{2\nu}(0) &= \overline{\mathcal{M}}_{2\nu} \end{aligned} \quad (40)$$

where

$$\overline{\mathcal{M}} = \begin{bmatrix} E(\overline{x}_e) \\ E(\overline{x}_e^{[2]}) \\ \cdots \\ E(\overline{x}_e^{[2\nu]}) \end{bmatrix}, \qquad \mathcal{U}_{2\nu}(k) = \begin{bmatrix} FE(N(k)) \\ F^{[2]}E(N^{[2]}(k)) \\ \vdots \\ F^{[2\nu]}E(N^{[2\nu]}(k)) \end{bmatrix}$$

$$\mathcal{A}_{2\nu}(k) = \begin{bmatrix} E(\tilde{A}_e(k)) & 0 & \cdots & 0 \\ H_{2,1}(k) & E(\tilde{A}_e^{[2]}(k)) & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots \\ H_{2\nu,1}(k) & H_{2\nu,2}(k) & \cdots & E(\tilde{A}_e^{[2\nu]}(k)) \end{bmatrix}$$

$$H_{i,l}(k) = M_{i-l}^i(q)(F^{[i-l]} \otimes E(\tilde{A}_e^{[l]}(k))) \cdot \Big(E(N^{[i-l]}(k)) \otimes I_{q,l}\Big).$$

Proof: See Appendix A. □

$$\begin{aligned} T_{h,l}(k) &= \begin{bmatrix} E(\tilde{A}^{[h+1]}(k) \otimes \tilde{A}_e^{[l-1]}(k)) & 0 & 0 \\ C(k) \otimes E(\tilde{A}^{[h]}(k) \otimes \tilde{A}_e^{[l-1]}(k)) & 0 & 0 \\ 0 & I_{m,\Delta-1} \otimes E(\tilde{A}^{[h]}(k) \otimes \tilde{A}_e^{[l-1]}(k)) & 0 \end{bmatrix} \\ &= \begin{bmatrix} Q_{h+1,l-1}(k) & 0 & 0 \\ C(k) \otimes Q_{h,l-1}(k) & 0 & 0 \\ 0 & I(m \cdot (\Delta-1)) \otimes Q_{h,l-1} & 0 \end{bmatrix} \end{aligned} \quad (37)$$

VI. Polynomial Filtering

Now we are able to apply the Kalman filter to system (21), (22). It should be highlighted that since the samples of the augmented state and output noises are in general correlated at the same time, the system needs to use the Kalman filter in a version given by [22], which takes into account this nonzero correlation. The equations to use are the following:

$$\hat{\mathcal{X}}(k) = \hat{\mathcal{X}}(k/k-1) + \mathcal{K}(k) \cdot \Big(\mathcal{Y}(k) - \mathcal{C}(k)\hat{\mathcal{X}}(k/k-1) - \mathcal{V}(k)\Big) \quad (41)$$

$$\hat{\mathcal{X}}(k+1/k) = (\mathcal{A}(k) - (\mathcal{A}(k)\mathcal{K}(k) + \mathcal{Z}(k))\mathcal{C}(k)) \cdot \hat{\mathcal{X}}(k/k-1) + (\mathcal{A}(k)\mathcal{K}(k) + \mathcal{Z}(k)) \cdot (\mathcal{Y}(k) - \mathcal{V}(k)) + \mathcal{U}(k) \quad (42)$$

$$\mathcal{Z}(k) = \mathcal{J}(k)\big(\mathcal{C}(k)\mathcal{P}(k/k-1)\mathcal{C}^T(k) + \mathcal{R}(k)\big)^{-1} \quad (43)$$

$$\mathcal{P}(k+1/k) = \mathcal{A}(k)\mathcal{P}(k)\mathcal{A}^T(k) + \mathcal{Q}(k) - \mathcal{Z}(k)\mathcal{J}^T(k) - \mathcal{A}(k)\mathcal{K}(k)\mathcal{J}^T(k) - \mathcal{J}(k)\mathcal{K}^T(k)\mathcal{A}^T(k) \quad (44)$$

$$\mathcal{P}(k) = \mathcal{P}(k/k-1) - \mathcal{K}(k)\mathcal{C}(k)\mathcal{P}(k/k-1) \quad (45)$$

$$\mathcal{K}(k) = \mathcal{P}(k/k-1)\mathcal{C}^T(k) \cdot \big(\mathcal{C}(k)\mathcal{P}(k/k-1)\mathcal{C}^T(k) + \mathcal{R}(k)\big)^{-1} \quad (46)$$

where $\mathcal{K}(k)$ is the filter gain, $\mathcal{P}(k)$, $\mathcal{P}(k/k-1)$ are the filtering and one-step prediction errors covariances, respectively, and the other symbols are defined as in Theorem 5.1. If the matrix $\mathcal{C}(k)\mathcal{P}(k/k-1)\mathcal{C}^T(k) + \mathcal{R}(k)$ is singular, it is possible to use the Moore–Penrose pseudo-inverse.

Equation (41) yields recursively the vector $\hat{\mathcal{X}}(k)$, that is the optimal linear estimate of $\mathcal{X}(k)$ with respect to the aggregate vector of all the augmented observations up to time k:

$$\mathcal{Y}_k \triangleq \begin{bmatrix} \mathcal{Y}(k) \\ \mathcal{Y}(k-1) \\ \vdots \\ \mathcal{Y}(0) \\ 1 \end{bmatrix}$$

(we remind readers that here the unit element allows us to reduce an *affine* estimation problem to a *strictly linear* one). From Definition (20) of $\mathcal{X}(k)$ and (14) of $x_e(k)$, it follows that the original state, $x(k)$, is the aggregate of the first n entries of the vector $\mathcal{X}(k)$. Since the optimal linear estimate with respect to $\mathcal{Y}_k$ is the L^2 projection of the random vector $\mathcal{X}(k)$ on the subspace linearly spanned by $\mathcal{Y}_k$, it follows that we can obtain the optimal linear estimate of $x(k)$ with respect to $\mathcal{Y}_k$, i.e., $\hat{x}(k)$, by extracting in $\hat{\mathcal{X}}(k)$ the first n entries

$$\hat{x}(k) = [I(n) \quad 0]\hat{\mathcal{X}}(k). \quad (47)$$

Equation (47) implies that the error covariance of the original state, namely $P(k)$, is given by

$$P(k) = [I(n) \quad 0]\mathcal{P}(k)\begin{bmatrix} I(n) \\ 0 \end{bmatrix} \quad (48)$$

where $\mathcal{P}(k)$ is given by (45) and hence is the $n \times n$ top left block of $\mathcal{P}(k)$. By remembering the structure of the extended observation (14) and of the augmented one (19), from Definition 2.2 we infer that $\hat{x}(k)$ is the (ν, Δ)-optimal polynomial estimate with respect to the originary measurements $\{y(0), y(1), \cdots, y(k)\}$. As in [20], we call a *polynomial filter* the whole set of operations constituted by the recursive equations (41), (42) and by the extraction of the first n entries in $\hat{\mathcal{X}}(k)$, resulting in an algorithm having the form (7), (8).

VII. Stationarity and Steady-State Behavior

Equations (41), (47) allow us the recursive calculation of the state polynomial estimate for the time-varying bilinear system (9), (10). However, in the time-varying case the result will be in general dependent on the initial conditions, whose statistics are often unknown. Moreover, the gain equations (43)–(46) need to be implemented simultaneously to the filter equations (41), (42).

Due to the high complexity of this filter, it assumes great importance from a practical point of view, to know when there exists the steady-state version of (44)–(46). Here we will limit ourselves to examining some important subclasses of bilinear systems for which we will be able to give necessary conditions under which a stationary behavior can be achieved.

First of all, let us consider the case when the system matrices $A(k)$, $C(k)$ and the bilinear form $B(k, x, \xi')$ of system (9), (10) are time independent: $A(k) \equiv A$, $C(k) \equiv C$, and $B(k, x, \xi') \equiv B(x, \xi')$. Moreover, let us assume the noises $\eta(k)$, $\xi(k)$, $\xi'(k)$ are weakly stationary sequences (that is, their moments are time invariant). This case is modeled by the following *stationary bilinear system*:

$$x(k+1) = Ax(k) + \sum_{i=1}^{u} B_i x(k)\xi_i'(k) + \xi(k), \qquad x(0) = \overline{x} \quad (49)$$

$$y(k) = Cx(k) + \eta(k) \quad (50)$$

which can be rewritten, as in the time-varying case, in the linear form with stochastic dynamic matrix

$$x(k+1) = \tilde{A}(k)x(k) + \xi(k), \qquad x(0) = \overline{x} \quad (51)$$

$$y(k) = Cx(k) + \eta(k) \quad (52)$$

where

$$\tilde{A}(k) = A + \sum_{i=1}^{u} B_i\xi_i'(k). \quad (53)$$

The corresponding augmented system is

$$\mathcal{X}(k+1) = \mathcal{A}\mathcal{X}(k) + \mathcal{U} + \mathcal{F}(k), \qquad \mathcal{X}(0) = \overline{\mathcal{X}} \quad (54)$$

$$\mathcal{Y}(k) = \mathcal{C}\mathcal{X}(k) + \mathcal{V} + \mathcal{G}(k). \quad (55)$$

As is well known, the Kalman filter implemented on a time-invariant system such as (54), (55), having second-order weakly stationary noises, admits a steady-state gain under the additional hypotheses of stabilizability and detectability [22]. Moreover, from Theorem 5.4, it follows that the extended state moments, $E(x_e^{[i]}(k))$, $i = 1, \cdots, 2\nu$, given by (40), converge if and only if the matrix $\mathcal{A}_{2\nu}(k)$ is asymptotically stable. By observing the structure of $\mathcal{A}_{2\nu}(k)$, we infer that it is asymptotically stable if and only if the eigenvalues of all the matrices $E(\tilde{A}_e^{[i]}(k))$, for $i = 1, \cdots, 2\nu$, belong to the unit circle of the complex plane. It also follows that the stability of the matrices $E(\tilde{A}_e^{[i]}(k))$, for $i = 1, \cdots, 2\nu$, implies the asymptotic stationarity of the augmented noises. Such a condition is then sufficient to assure the existence of a stationary filter. Now, the main problem is to give sufficient conditions for the stability of $E(\tilde{A}_e^{[i]}(k)),\ i = 1, \cdots, 2\nu$.

We will see that for a strictly bilinear system, even time-invariant and with stationary noises, the possibility to implement a stationary polynomial filter is not, in general, assured. Indeed, we are able to find a counterexample in a particular but important case, that is when the noises are Gaussian, as shown in the following theorem.

Theorem 7.1: For the matrices

$$E(\tilde{A}_e^{[l]}(k)) \tag{56}$$

with $\tilde{A}_e(k)$ given by

$$\tilde{A}_e(k) = A_e + B_e(k)$$

$$A_e = \begin{bmatrix} A & 0 & \cdots & \cdots & 0 \\ C & 0 & \cdots & \cdots & 0 \\ 0 & I & \ddots & & \vdots \\ \vdots & & \ddots & \ddots & \vdots \\ 0 & \cdots & \cdots & I & 0 \end{bmatrix} \tag{57}$$

$$B_e(k) = \begin{bmatrix} \sum_{i=1}^{u} B_i \xi_i'(k) & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & 0 \end{bmatrix} \tag{58}$$

and under the hypotheses that ξ' is Gaussian and $\mathrm{tr}(A) \neq 0$, $\mathrm{tr}(B_j) \neq 0$, for $j = 1, \cdots, u$, it results that there exists $\bar{l} < +\infty$ such that (56) is unstable for all $l > \bar{l}$.

Proof: Let us suppose, for sake of simplicity, that the entries of ξ' have unit variance and are mutually independent. By using Property (93h) and taking into account the structure of A_e (57) and that of $B_e(k)$ (58), we have

$$\begin{aligned} |\mathrm{tr}(E(\tilde{A}_e^{[l]}(k)))| &= |E(\mathrm{tr}(\tilde{A}_e^{[l]}(k)))| \\ &= |E((\mathrm{tr}(\tilde{A}_e(k)))^l)| = |E((\mathrm{tr}(A_e) + \mathrm{tr}(B_e(k)))^l)| \\ &= \left| E\left(\left(\mathrm{tr}(A) + \sum_{j=1}^{u} (\mathrm{tr}(B_j))\xi_j'(k) \right)^l \right) \right| \\ &= |E((\mathrm{tr}(A) + Y)^l)| \end{aligned}$$

where $Y = \sum_{j=1}^{u} \mathrm{tr}(B_j)\xi_j'(k)$. Hence

$$|\mathrm{tr}(E(\tilde{A}_e^{[l]}(k)))| = \left| \sum_{m=0}^{l} \binom{l}{m} (\mathrm{tr}(A))^{l-m} E(Y^m) \right|.$$

Since $\xi_j'(k)$, $j = 1, \cdots, u$ are Gaussian and independent, we have

$$E(Y^m) = \begin{cases} 0, & m \text{ odd,} \\ (m-1)!!(E(Y^2))^{m/2}, & \text{otherwise;} \end{cases}$$

and hence

$$|\mathrm{tr}(E(\tilde{A}_e^{[l]}(k)))| = \left| \sum_{m=0}^{[l/2]} \binom{l}{2m} (\mathrm{tr}(A))^{l-2m} E(Y^{2m}) \right|.$$

Note that all the terms of the summation in the right side have the same sign. Finally, we have

$$\begin{aligned} &\left| \sum_{m=0}^{[l/2]} \binom{l}{2m} (\mathrm{tr}(A))^{l-2m} E(Y^{2m}) \right| \\ &\quad \geq \min\{1, \mathrm{tr}(A)\}(l-2)!! \sum_{j=1}^{u} (\mathrm{tr}(B_j))^2 \end{aligned}$$

where the right side is obtained by calculating the sum for $m = [l/2]$ and taking into account that

$$E(Y^2) = \sum_{j=1}^{u} (\mathrm{tr}(B_j))^2.$$

Hence, we have $|\mathrm{tr}(E(\tilde{A}_e^{[l]}(k)))| \to +\infty$, for $l \to +\infty$, faster than $[l/2]!$. Since

$$\mathrm{tr}(E(\tilde{A}_e^{[l]}(k))) = \sum_{i=1}^{q^l} \lambda_i \leq q^l \max_i(\lambda_i)$$

where λ_i are the eigenvalues of $E(\tilde{A}_e^{[l]}(k))$, this implies the existence of at least one eigenvalue greater than one. □

The circumstance that the availability of the steady-state moments of any order is not assured for a bilinear system represents a limitation in designing stationary polynomial filters. In order to be more precise about this limitation, let us introduce the following definition.

Definition 7.2: For a stochastic bilinear system such as (49), (50), we define the stochastic stability degree ν_s as the maximum order for which the extended state moments $E(x_e^{[i]}(k)),\ \ i = 1, \cdots, \nu_s$ converge to a finite value for $k \to +\infty$, for any initial condition $E(x_e^{[i]}(0)) < +\infty$. We set $\nu_s = 0$ when the first moment is not convergent.

For a stochastic time-invariant linear system having finite noise moments of all orders, ν_s can assume only the values zero or $+\infty$; that is, if the dynamic matrix is stable (unstable), $\nu_s = +\infty (\nu_s = 0)$. This fact is a trivial reformulation of the theory developed in [20]. For a bilinear system such as (49), (50), it is possible to implement a stationary polynomial filter of (h, Δ) order, for any $\Delta \in \mathbb{N}$ and $h \leq [\nu_s/2]$ (here $[\cdot]$ denotes integer part). The determination of the stochastic stability degree is hence useful for stating in advance the maximum order for which the state polynomial estimate is computable by means of a stationary filter. For this purpose, some results, useful for the determination of the stochastic stability degree, can be found in [17], [18], [24], and [25]. Here we specialize the above-mentioned results in order to study the stochastic stability of the Kronecker powers, up to the νth degree of the extended state or the stationarity of the extended state moments, which is the same.

Lemma 7.3: The stochastic system

$$x(k+1) = A_s(k)x(k) \tag{59}$$

where $\{A_s(k)\}$ is a sequence of independent identically distributed stochastic matrices, has its lth moment asymptotically stable if

$$\lambda_M(E(A_s^{[l]^T}(k)A_s^{[l]}(k)) = \alpha < 1$$

where $\lambda_M(B)$ denotes the maximum eigenvalue of matrix B.

Proof: Taking the lth Kronecker power in (59) we have

$$x^{[l]}(k+1) = A_s^{[l]}(k)x^{[l]}(k). \tag{60}$$

From the hypotheses it follows that

$$E(A_s^{[l]^T}(k)A_s^{[l]}(k)) < I$$

hence

$$E(A_s^{[l]^T}(k)A_s^{[l]}(k)) - I = -P, \qquad P > 0.$$

The thesis follows by applying [18, Lemma 3.2] to (60). □

It is now possible to determine a sufficient condition for the stability of (40). In fact, the following theorem holds.

Theorem 7.4: If

$$E(\lambda_M(A_s^{[l]^T}(k)A_s^{[l]}(k))) < 1 \tag{61}$$

then $\forall\, j \leq l,\ E(A_s^{[j]}(k))$ is stable.

Proof: Observe that the function λ_M is convex on the set of symmetric nonnegative matrices. This easily follows by the property [21]:

$$\lambda_M(Q_1 + Q_2) \leq \lambda_M(Q_1) + \lambda_M(Q_2)$$

$\forall\, Q_1,\, Q_2 \geq 0\, Q_1 = Q_1^T,\, Q_2 = Q_2^T$; hence, using the Holder and Jensen inequality and (61), $\forall\, j \leq l$

$$\begin{aligned}\alpha &= \lambda_M(E(A_s^{[j]^T}(k)A_s^{[j]}(k))) \leq E(\lambda_M(A_s^{[j]^T}(k)A_s^{[j]}(k))) \\ &= E(\lambda_M(A_s{}^T(k)A_s(k))^{[j]}) = E((\lambda_M(A_s{}^T(k)A_s(k)))^j) \\ &= ((E((\lambda_M(A_s{}^T(k)A_s(k)))^j))^{1/j})^j \\ &\leq ((E(\lambda_M(A_s^{[l]^T}(k)A_s^{[l]}(k))))^{1/l})^j < 1\end{aligned}$$

which, using Lemma (7.3), proves the thesis. □

Corollary 7.5: A sufficient condition for the stability of (40) relative to (49), (50) is

$$E(\operatorname{tr}(\tilde{A}_e^T(k)\tilde{A}_e(k)))^{2\nu} < 1.$$

Proof: The thesis follows from the inequality:

$$E(\lambda_M(A_s^{[2\nu]^T}(k)A_s^{[2\nu]}(k))) \leq E(\operatorname{tr}(A_s^T(k)A_s(k)))^h$$

applying Theorem 7.4 with $A_s(k) = \tilde{A}_e(k)$ and taking into account of the block-triangular structure of $\mathcal{A}_{2\nu}$. □

VIII. Implementation Remarks

Some numerical simulations have been carried out on a Digital "alpha" workstation by implementing the polynomial filter equations in order to produce for any pair of integers $\nu > 0, \Delta \geq 0$, the (ν, Δ)-order optimal polynomial state estimate of a BLSS.

For this purpose, we have written a C-language program whose main part is devoted to the efficient implementation of the algorithms, described in Sections V and VI and Appendix B, for the computation of the filter parameters. By observing the formulas which define the augmented system parameters, in the statement of Theorem 5.1, it becomes evident that the computational effort of the whole polynomial filter algorithm quickly grows for increasing ν and/or Δ. Nevertheless, we point out that even low-order polynomial filters (quadratic or cubic filters) which do not require a particularly sophisticated implementation show very high performances with respect to the classical linear filter. Indeed, as shown in some numerical simulations of the polynomial filter for linear systems [20], the error variance of a cubic filter may be 80% smaller than the Kalman filter. As we will see later, these high performances are confirmed by low-order polynomial filters for a BLSS. In the case presented here, the second-degree polynomial filter yields a signal error variance of 54% less than linear filter. In the same case we have been able to compute the fourth-degree polynomial filter (indeed, a high-order filter, in that it requires a state space of dimension 30 for two-state variables of the system) which yields an improvement of 75% with respect to the linear filter. As shown in some pictures, the restoration of the noisy signal is very impressive.

We would like to stress that the high dimensionality of the filter is not by itself a true limitation for the implementation. In fact, by using an efficient implementation scheme for those data structures which appear in the formulas as matrices of prohibitive dimension, it is possible to overcome such difficulties. It should be noted that the computational effort is mostly due to the calculation of the augmented system parameters. In many cases that are relevant from an application point of view, that is, time-invariance of system parameters, stationarity of noises, polynomial degree less than the stochastic stability degree (see Section VII), we can separate the augmented system matrix computation from the filter equations (41)–(46) that do not show relevant computational troubles. In all of these cases, polynomial filtering is amenable to real-time applications. The numerical simulations presented here concern the filtering of time-invariant BLSS's with stationary noises and degree less than the stochastic stability degree so that the stationary polynomial filter is implemented using the steady-state gain, and the augmented matrices are calculated before filtering.

Among all the algorithms which are necessary for the computation of the augmented system parameters, the most burdensome are those involved in the computation of the extended state moments $E(x_e^{[i]}(k)),\ \ i = 1, \cdots, 2\nu$, that appear in the augmented noise covariance (24)–(28). These are obtained by running (40) until convergence is achieved. The dynamical matrix $\mathcal{A}_{2\nu}$ of (40) may be very large and exceed the available computer memory space. We think that for large degrees (i.e., three or more) many tricks can be conceived, when a larger computer memory is not available, in order to save memory space (for instance, to save and use only suitably small blocks of the matrix).

In order to calculate the matrix $\mathcal{A}_{2\nu}$ and the augmented noises covariances, the computation of the matrix moments $E(\tilde{A}_e^{[i]}(k)),\ i = 1, \cdots, 2\nu$ is needed. These are obtained by means of the algorithm defined by Theorem 5.2, which in turn requires the matrices $E(\tilde{A}^{[i]}(k))$ given by (39). The matrices $D_{n,n}^{(i)},\ i = 1, \cdots, 2\nu$, which appear in (39) (and are defined in Corollary B.8), are $n^{2i} \times n^{2i}$ dimensioned; that is, they may be too large. In our example, for $n = 2, \nu = 4, \Delta = 0$, they have 2^{16} entries! Nevertheless, this very high dimensionality is only apparent. In fact, by considering (104) we realize that $D_{n,n}^{(i)}$ can be viewed as an operator which simply permutes the entries of a vector (permutation matrix). A permutation matrix is a zero–one square matrix with one (and only one) unity on each row and column so that it can be simply implemented as a string of 2^{2i} integers, each one representing the column index of the unity in a row. Also note that the commutation matrices, given by B.6, which appears in many formulas, are permutation matrices.

Finally, the last kind of matrices widely used in the whole algorithm, which can easily grow toward huge dimensions, is the binomial matrix M_l^h, defined in Theorem B.6, and the generalization $M_{h_1, \cdots, h_p}^h$ defined in Theorem B.9. These are integer matrices with many null entries; for this reason we have implemented them as integer sparse matrices. In spite of this expediency, we have observed that the matrices $M_{h_0, \cdots, h_u}^i$ used in (39) can still exceed the computer memory. Anytime this happens we adopt the method, mentioned above, consisting of calculating only small blocks of the matrices and removing them after their utilization. Thus, we can avoid overcoming space memory availability, in spite of a growth of the CPU time. This method surely can be adopted for higher-polynomial degrees and system orders and always assures that the computation will be made with the same memory usage.

It should be underlined that, in the most important stationary case, all the above-mentioned expediences are useful, and sometimes necessary, in order to treat efficiently the major critical parts of the whole algorithm, even if they can produce a great growth of the CPU time needed for the filter parameter computations. However, they do not affect filter measurement processing.

IX. Simulation Example

The example of an application we are going to consider belongs to the class of the so-called *switching systems*, widely used in many research areas such as failure detection, speech recognition, and, more generally, in the modeling of physical systems affected by abrupt changes in the parameters [26]–[31]. In particular, we are interested in the class of systems described by the following partially observed equation defined on $(\Omega, \mathcal{F}, P)$, evolving in $\mathbb{R}^n$:

$$\begin{aligned} x(k+1) &= A(k)x(k) + \xi(k) \\ s(k) &= Cx(k) \\ y(k) &= s(k) + \eta(k) \end{aligned} \tag{62}$$

where $y(k) \in \mathbb{R}^m$; $\{\xi(k)\}$, $\{\eta(k)\}$ are white sequences and $\{A(k)\}$ is a white random matrix sequence taking values in the finite set $\{A_1, A_2, \cdots, A_q\}$ with probabilities $P(A_i) = p_i,\ i = 1, \cdots, q$. System (62) can be easily represented as a BLSS in the following way. Let $e_i,\ i = 1, \cdots, q$ be the canonical base in $\mathbb{R}^q$, and let us define the white sequence $\{\xi'(k)\}$ assuming values in $\{e_1, \cdots, e_q\}$ with $P(\xi'(k) = e_i) = p_i,\ i = 1, \cdots, q, k \geq 0$. Then

$$A(k) = \sum_{i=1}^{q} A_i \xi_i'(k). \tag{63}$$

From the above hypotheses, it follows that $\xi_q'(k) = 1 - \sum_{i=1}^{q-1} \xi_i'(k)$, and using this in (63) results in

$$A(k) = A_q + \sum_{i=1}^{q-1} (A_i - A_q)\xi_i'(k). \tag{64}$$

By combining (62) and (64), we obtain the BLSS (11), (13) with $A(k) = A_q$, $B_i(k) = A_i - A_q$, $u = q - 1$.

Now, in order to test the filter, let us consider the switching system (62), with $n = q = 2, m = 1$

$$\begin{aligned} A_1 &= \begin{bmatrix} 0.2 & 0.41 \\ -0.15 & -0.2 \end{bmatrix} \\ A_2 &= \begin{bmatrix} 0.8 & 0.59 \\ 0.15 & 0.4 \end{bmatrix} \\ C &= [0.7 \quad 0.3] \end{aligned}$$

and $P(A_1) = P(A_2) = 0.5$. Moreover, let the white random sequences $\{\xi(k)\}$, $\{\eta(k)\}$ be defined as

$$\begin{aligned} \xi(k) &= \begin{bmatrix} \xi_1(k) \\ \xi_2(k) \end{bmatrix} \\ \xi_1(k)(\omega) &= -0.48\chi_{F_1}(\omega) + 0.12\chi_{F_2}(\omega) \\ \xi_2(k)(\omega) &= 0.02\chi_{F_3}(\omega) - 0.18\chi_{F_4}(\omega) \\ \eta(k)(\omega) &= -0.168\chi_{H_1}(\omega) + 0.172\chi_{H_2}(\omega) + 1.17\chi_{H_3}(\omega) \end{aligned}$$

where χ_Q, $Q \in \mathcal{F}$ denotes the characteristic function of Q and the disjoint events (F_1, F_2), (F_3, F_4), and (H_1, H_2, H_3) have probability

$$\begin{aligned} &P(F_1) = 0.2 \qquad P(F_2) = 0.8 \\ &P(F_3) = 0.9 \qquad P(F_4) = 0.1 \\ &P(H_1) = 0.8 \qquad P(H_2) = 0.1 \qquad P(H_3) = 0.1. \end{aligned}$$

Following the above described procedure, such a switching system is represented by the following BLSS:

$$\begin{aligned} x(k+1) &= Ax(k) + Bx(k)\xi'(k) + \xi(k) \\ s(k) &= Cx(k) \\ y(k) &= s(k) + \eta(k) \end{aligned} \tag{65}$$

where $A = A_2$, $B = A_1 - A_2$, and $\{\xi'(k)\}$ is a white sequence defined as $\xi'(k) = \chi_G$ with $G \in \mathcal{F}$, $P(G) = 0.5$.

For this system, we have built the steady-state augmented system for the polynomial degrees $\nu = 1, 2, 3, 4$, with $\Delta = 0$ and the quadratic and cubic also with $\Delta = 1$. To each one of these augmented systems we have applied (43)–(46) in order to obtain the steady-state gains and error

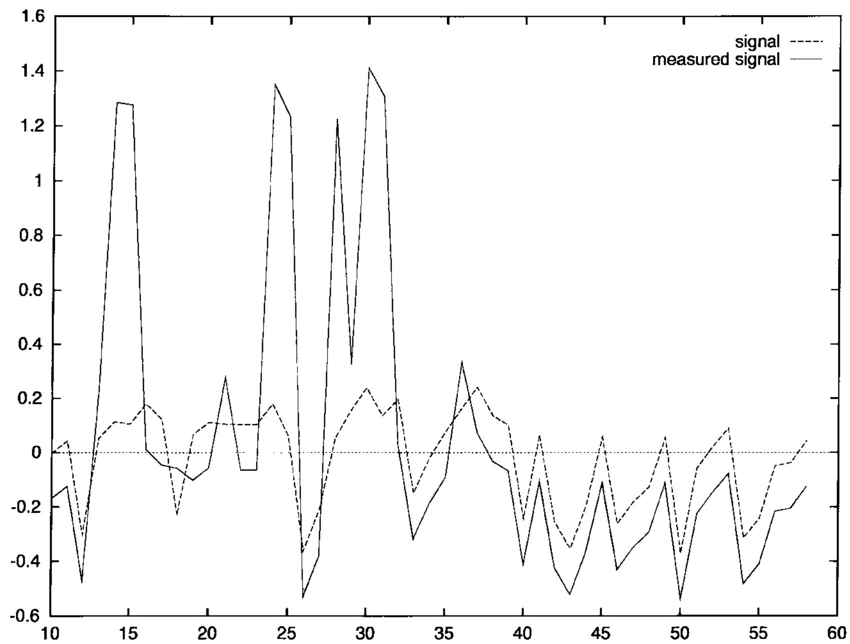

Fig. 1. True and measured signal.

covariances. Then, for all these cases, we have used the gains in the filter equations (41), (42), starting from initial condition $\hat{x}(k) = 0$. The corresponding estimates for the signal $s(k)$ are readily obtained by using the relation $\hat{s}(k) = C\hat{x}(k)$. Moreover, the signal error variance, namely P_s, is given by the relation $P_s = C^T PC$, where P is the steady-state value of the state error covariance given by (48). By denoting with $P^{(1,0)}$, $P^{(2,0)}$, $P^{(3,0)}$, $P^{(4,0)}$, the *a priori* state error covariances given by the 1, 2, 3, 4th-degree polynomial filters, respectively (all with $\Delta = 0$), and with $P^{(2,1)}$, $P^{(3,1)}$ the covariances relative to the $\Delta = 1$ quadratic and cubic cases, respectively, the obtained values are the following:

$$P^{(1,0)} = \begin{bmatrix} 0.068\,31 & 0.003\,77 \\ 0.003\,77 & 0.006\,76 \end{bmatrix}$$

$$P^{(2,0)} = \begin{bmatrix} 0.031\,48 & 0.000\,74 & \cdots \\ 0.000\,74 & 0.006\,45 & \cdots \\ \vdots & \vdots & \ddots \end{bmatrix}$$

$$P^{(3,0)} = \begin{bmatrix} 0.0027 & 0.001 & \cdots \\ 0.001 & 0.006\,34 & \cdots \\ \vdots & \vdots & \ddots \end{bmatrix}$$

$$P^{(4,0)} = \begin{bmatrix} 0.017\,72 & -0.000\,28 & \cdots \\ -0.000\,28 & 0.006\,32 & \cdots \\ \vdots & \vdots & \ddots \end{bmatrix}$$

$$P^{(2,1)} = \begin{bmatrix} 0.029\,86 & 0.000\,83 & \cdots \\ 0.000\,83 & 0.006\,44 & \cdots \\ \vdots & \vdots & \ddots \end{bmatrix}$$

$$P^{(3,1)} = \begin{bmatrix} 0.022\,08 & -0.000\,39 & \cdots \\ -0.000\,39 & 0.006\,32 & \cdots \\ \vdots & \vdots & \ddots \end{bmatrix}$$

where we have the 2 × 2 matrix blocks on the top left side because they contain in the main diagonal the steady-state estimate error variance of each component of the state. The corresponding values, $\sigma^{(1,0)}$, $\sigma^{(2,0)}$, $\sigma^{(3,0)}$, $\sigma^{(4,0)}$, $\sigma^{(2,1)}$, $\sigma^{(3,1)}$ for the signal error variances are

$$\begin{aligned} \sigma^{(1,0)} &= 0.035\,66 \\ \sigma^{(2,0)} &= 0.016\,31 \\ \sigma^{(2,1)} &= 0.015\,56 \\ \sigma^{(3,0)} &= 0.013\,84 \\ \sigma^{(3,1)} &= 0.011\,55 \\ \sigma^{(4,0)} &= 0.009\,13. \end{aligned}$$

As implied by the overall theory described in Section II, we can see that both signal-error variances and state-error variances of each component of the state decrease with the increasing of polynomial degree. In the $\nu = 4$, $\Delta = 0$ case, the signal-error variance is 75% less than in the linear filtering case. Also for the error-variance values relative to the quadratic and cubic cases with $\Delta = 1$, we observe, as expected, an improvement with respect to the same cases with $\Delta = 0$. However, in our experience, the contribution of the increasing Δ is less effective than the increasing of the polynomial degree.

In Table I, the sampled variances of the state and signal, obtained with a number $N = 100\,000$ iterations, are reported. As expected, these values are close to the above *a priori* variance values. In the same table are also reported the signal sampled variances for the Monte Carlo run of 60 iterations relative to Figs. 1–5. Fig. 1 displays the sample paths obtained for the observed and true signal, whereas Figs. 2–4 display the same path of the true signal with different polynomial estimates.

X. Conclusions

The (ν, Δ)-optimal polynomial filter for the BLSS (1), (2), given the (ν, Δ)-order polynomial estimate (see Definition

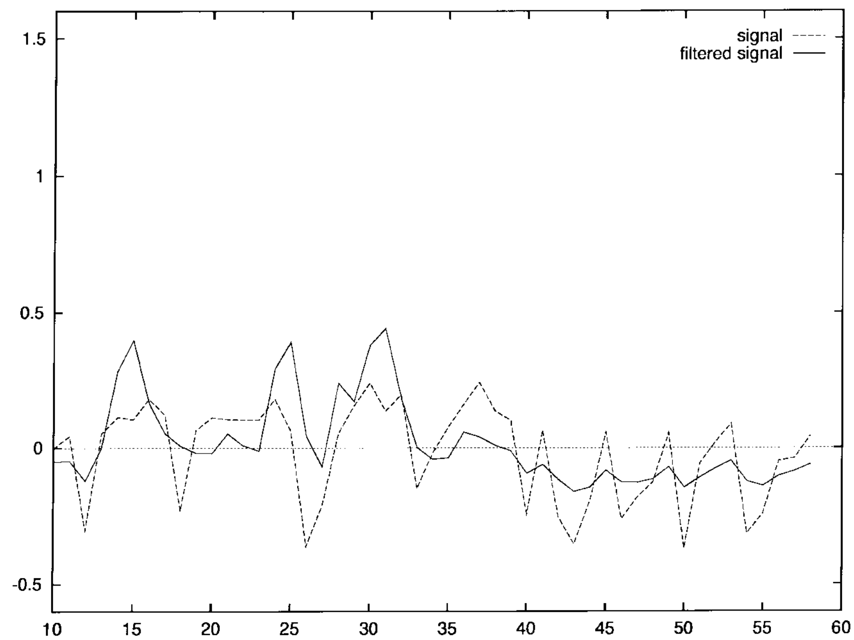

Fig. 2. True and filtered signal with $\nu = 1, \Delta = 0$.

TABLE I

	$\sigma^2_{(s-\hat{s})}$ $N = 60$	$\sigma^2_{(s-\hat{s})}$ $N = 10^5$	$\sigma^2_{(x-\hat{x})_1}$ $N = 10^5$	$\sigma^2_{(x-\hat{x})_2}$ $N = 10^5$
$\nu = 1,\ \Delta = 0$	0.01898	0.03531	0.06761	0.00675
$\nu = 2,\ \Delta = 0$	0.00483	0.01633	0.03143	0.00645
$\nu = 2,\ \Delta = 1$	0.00554	0.01571	0.03015	0.00645
$\nu = 3,\ \Delta = 0$	0.00342	0.01362	0.0265	0.00634
$\nu = 3,\ \Delta = 1$	0.00363	0.01137	0.0217	0.00633
$\nu = 4,\ \Delta = 0$	0.0029	0.0905	0.01757	0.00631

2.2) of the state by means of recursive equations in the form (7), (8), has been defined for any pair of integers $\nu \geq 1$, $\Delta \geq 0$. In particular, the polynomial filter equations are (41), (42), and (47). These need to use, at each step, only powers of the last Δ observations so that the computational burden remains constant over time. The polynomial filter is obtained by means of the following steps:

1) construction of the extended memory system (15), (16) (if $\Delta = 0$ this step is skipped);
2) construction of the augmented system;
3) application of the Kalman filter equations to the augmented system.

Equations (43) and (46) allow the computation of filter parameters. These need, in general, to be implemented simultaneously to the filter equations (41) and (43). Nevertheless, if the BLSS is time invariant, the noises are stationary sequences and the matrix $\mathcal{A}_{2\nu}$ (defined in the statement of Theorem 5.4) is asymptotically stable, then we can adopt the steady-state approximation of the Kalman filter, thus obtaining a great reduction of computational effort.

In Section VII, it is shown that the stability of the matrix $\mathcal{A}_{2\nu}$, or equivalently the stability of all the matrices $E(\tilde{A}_e^{[i]}(k))$, $i = 1, \cdots, 2\nu$, is not implied by the stability of $E(\tilde{A}_e(k))$ so that in general the steady-state polynomial filter can be implemented only up to a certain finite degree. Corollary 7.6 gives a sufficient condition for the stability of the matrix $\mathcal{A}_{2\nu}$.

Even if the computational burden of polynomial filtering grows when ν and/or Δ increase, many tricks (e.g., as in Section VIII) can be conceived in order to considerably reduce computer memory and CPU time utilization. Numerical simulations presented in Section IX show the high performance of polynomial filtering with respect to standard linear filtering. For a second-order BLSS, we have observed for the (4,0)-order filter, an error-variance reduction of 75%. It should be stressed that the (2, 0)-order (quadratic) filter also shows a high performance (54%). In this case, computer time for executing steps 1), 2), and 3), has been less than 1 s and practically all devoted to filter parameter computations.

We think that future research work on polynomial filtering should concern the following points:

1) reducing the computational burden of the algorithm in order to actually make very high-order filters implementable;
2) investigation of the possible convergence of polynomial estimators, with respect to ν and Δ increasing, toward C.E. and evaluation of the convergence error;
3) analysis of the influence of Δ values on the polynomial filter performance. We conjecture that, for a stable BLSS, this influence tends to vanish when Δ increases because the observations tend to be uncorrelated when their mutual distance in time grows;
4) extension of the polynomial filtering to the class of linear systems with a multiplicative state noise modeled as a Markov chain or, more in general, as a colored stochastic sequence.

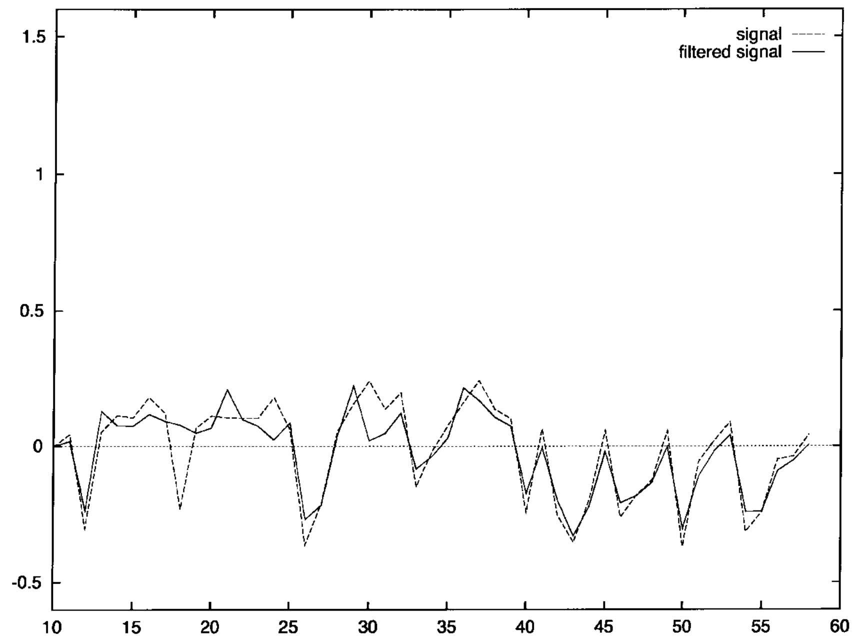

Fig. 3. True and filtered signal with $\nu = 2, \Delta = 0$.

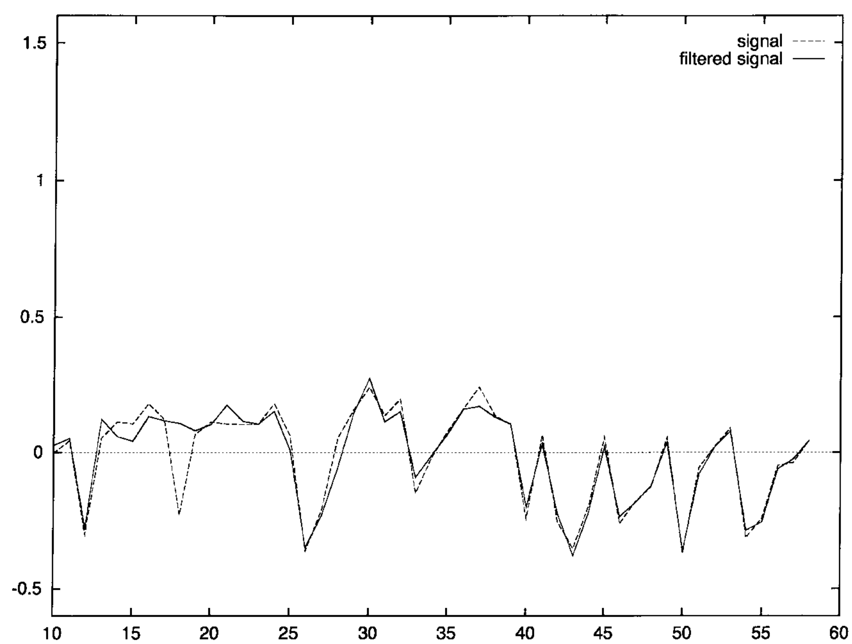

Fig. 4. True and filtered signal with $\nu = 4, \Delta = 0$.

To conclude, we say that this paper represents a first tentative attack upon nonlinear filtering problems via a polynomial algorithm. We feel that this could be a way of constructing suboptimal filters for a more general class of nonlinear systems.

APPENDIX A
AUGMENTED SYSTEM CONSTRUCTION

In this Appendix, the proof of Theorem 5.1, which defines the structure and the main properties of the augmented system, is reported. For this purpose we need to state some preliminary results (Lemma A.1 and Lemma A.2). In particular, Lemma A.2 will allow us to readily prove Theorem 5.4.

Let γ, γ', α, β denote positive integers, $\{\Gamma(k)\}$, $\{\Gamma'(k)\}$ sequences of random matrices in $\mathbb{R}^{\gamma\times\alpha}$ and $\mathbb{R}^{\gamma'\times\alpha}$, respectively, and $\{z(k)\}$, $\{\phi(k)\}$ sequences of random vectors in $\mathbb{R}^{\alpha}$ and $\mathbb{R}^{\beta}$, respectively. For any $k \geq 0$, let the following properties be satisfied.

1)
$$E(\|\Gamma(k)\|^{2\nu}) < \infty, \qquad E(\|\Gamma'(k)\|^{2\nu}) < \infty$$
$$E(\|z(k)\|^{2\nu}) < \infty, \qquad E(\|\phi(k)\|^{2\nu}) < \infty$$

where $\|\cdot\|$ denotes the Euclidean norm.

2) $\Gamma(k)$, $\phi(k)$, and the set $\{z(j), \phi(l), \Gamma'(l); j \le k, l < k\}$ are mutually independent.
3) $z(k)$, $\phi(k)$, and the set $\{\Gamma(k), \Gamma'(k)\}$ are mutually independent.

In the following the *binomial matrices* (101), (102) will be used often, which will be denoted as $M_i^j(l)$, highlighting the dependence from the dimension l of the vectors involved in the Kronecker power; moreover, the symbol $I_{i,j}$ will denote the identity matrix in $\mathbb{R}^{i^j}$.

In order to simplify the notations, let us introduce the following symbols:

$$\begin{aligned}\overline{\phi}_i(k) &\triangleq \phi^{[i]}(k) - E(\phi^{[i]}(k))\\ \overline{\Gamma}_i(k) &\triangleq \Gamma^{[i]}(k) - E(\Gamma^{[i]}(k))\\ \overline{\Gamma}'_i(k) &\triangleq \Gamma'^{[i]}(k) - E(\Gamma'^{[i]}(k)). \end{aligned} \tag{66}$$

Obviously for $i = 1, \cdots, 2\nu$ we have

$$E(\overline{\phi}_i(k)) = E(\overline{\Gamma}_i(k)) = E(\overline{\Gamma}'_i(k)) = 0. \tag{67}$$

With the above notations, it is now possible to state the following two lemmas.

Lemma A.1: Let $\{\Upsilon(k)\}$ be a sequence of stochastic matrices in $\mathbb{R}^{\theta\times\alpha}$ and Θ be a deterministic matrix in $\mathbb{R}^{\theta\times\beta}$. Moreover, let us define, for $i = 1, \cdots, 2\nu$, the following functions:

$$h_{i,\theta,\alpha}(k, \Upsilon, \Theta, z, \phi) = h^{(1)}_{i,\theta,\alpha}(k, \Upsilon, \Theta, z, \phi) + h^{(2)}_{i,\theta,\alpha}(k, \Upsilon, \Theta, z, \phi) \tag{68}$$

with

$$\begin{aligned}&h^{(1)}_{i,\theta,\alpha}(k, \Upsilon, \Theta, z, \phi)\\ &\quad= \sum_{l=0}^{i-1} M^i_{i-l}(\theta)(\Theta^{[i-l]} \otimes \Upsilon^{[l]}(k))\\ &\quad\quad\cdot ((\phi^{[i-l]}(k) - E(\phi^{[i-l]}(k))) \otimes I_{\alpha,l}) z^{[l]}(k)\end{aligned} \tag{69}$$

$$\begin{aligned}&h^{(2)}_{i,\theta,\alpha}(k, \Upsilon, \Theta, z, \phi)\\ &\quad= \sum_{l=1}^{i} M^i_{i-l}(\theta)(\Theta^{[i-l]} \otimes (\Upsilon^{[l]}(k) - E(\Upsilon^{[l]}(k))))\\ &\quad\quad\cdot (E(\phi^{[i-l]}(k)) \otimes I_{\alpha,l}) z^{[l]}(k).\end{aligned} \tag{70}$$

Then, for any couple Ψ, Ψ' of (deterministic) matrices in $\mathbb{R}^{\gamma\times\beta}$ and $\mathbb{R}^{\gamma'\times\beta}$, respectively, we have that

$$E(h_{i,\gamma,\alpha}(k, \Gamma, \Psi, z, \phi)) = 0, \qquad i = 1, \cdots, 2\nu \tag{71}$$

and furthermore, for $r, s = 1, \cdots, \nu$ we have that

$$\begin{aligned}&E(h^{(l)}_{r,\gamma,\alpha}(k, \Gamma, \Psi, z, \phi) \cdot h^{(m)}_{s,\gamma',\alpha}{}^T(j, \Gamma', \Psi', z, \phi))\\ &\quad= 0, \qquad l, m = 1, 2, \qquad j \ne k\end{aligned} \tag{72}$$

$$\begin{aligned}&E((h^{(1)}_{r,\gamma,\alpha}(k, \Gamma, \Psi, z, \phi) h^{(1)}_{s,\gamma',\alpha}{}^T(k, \Gamma', \Psi', z, \phi))\\ &\quad= T^{(1)}_{r,s,\gamma,\gamma',\alpha,\beta}(k, \Gamma, \Gamma', \Psi, \Psi', z, \phi)\end{aligned} \tag{73}$$

$$\begin{aligned}&E(h^{(1)}_{r,\gamma,\alpha}(k, \Gamma, \Psi, z, \phi) h^{(2)}_{s,\gamma',\alpha}{}^T(k, \Gamma', \Psi', z, \phi))\\ &\quad= 0\end{aligned} \tag{74}$$

$$\begin{aligned}&E(h^{(2)}_{r,\gamma,\alpha}(k, \Gamma, \Psi, z, \phi) h^{(1)}_{s,\gamma',\alpha}{}^T(k, \Gamma'(k), \Psi', z, \phi))\\ &\quad= 0\end{aligned} \tag{75}$$

$$\begin{aligned}&E(h^{(2)}_{r,\gamma,\alpha}(k, \Gamma, \Psi, z, \phi) h^{(2)}_{s,\gamma',\alpha}{}^T(k, \Gamma', \Psi', z, \phi))\\ &\quad= T^{(2)}_{r,s,\gamma,\gamma',\alpha,\beta}(k, \Gamma, \Gamma', \Psi, \Psi', z, \phi)\end{aligned} \tag{76}$$

where

$$\begin{aligned}&T^{(1)}_{r,s,\gamma,\gamma',\alpha,\beta}(k, \Gamma, \Gamma', \Psi, \Psi', z, \phi)\\ &\quad= \sum_{l=0}^{r-1} \sum_{j=0}^{s-1} M^r_{r-l}(\gamma) st^{-1}(((I_{\gamma',s-j} \otimes C^T_{\gamma^r,\gamma^j})\\ &\quad\quad\cdot (\Psi'^{[s-j]} \otimes \Psi^{[r-l]} \otimes E(\Gamma^{[l]}(k) \otimes \Gamma'^{[j]}(k)))\\ &\quad\quad\cdot (I_{\beta,s-j} \otimes C_{\beta^{r-l}\alpha^l,\alpha^j})(I_{\beta,s-j} \otimes C^T_{\beta^{r-l}\alpha^l,\alpha^j})\\ &\quad\quad\cdot ((E(\phi^{[s+r-j-l]}(k)) - E(\phi^{[s-j]}(k)) \otimes E(\phi^{[r-l]}(k)))\\ &\quad\quad\otimes I_{\alpha,l+j}) \cdot C_{\alpha^l,\alpha^j}) E(z^{[l+j]}(k))) M^s_{s-j}(\gamma')^T\end{aligned} \tag{77}$$

$$\begin{aligned}&T^{(2)}_{r,s,\gamma,\gamma',\alpha,\beta}(k, \Gamma, \Gamma', \Psi, \Psi', z, \phi)\\ &\quad= \sum_{l=1}^{r} \sum_{j=1}^{s} M^r_{r-l}(\gamma) \mathrm{st}^{-1}(((I_{\gamma',s-j} \otimes C^T_{\gamma^r,\gamma'^j})\\ &\quad\quad\cdot (\Psi'^{[s-j]} \otimes \Psi^{[r-l]} \otimes E((\Gamma^{[l]}(k) - E(\Gamma^{[l]}(k)))\\ &\quad\quad\otimes (\Gamma'^{[j]}(k) - E(\Gamma'^{[j]}(k)))))\\ &\quad\quad\cdot (I_{\beta,s-j} \otimes C_{\beta^{r-l}\alpha^l,\alpha^j})(I_{\beta,s-j} \otimes C^T_{\beta^{r-l}\alpha^l,\alpha^j})\\ &\quad\quad\cdot (E(\phi^{[s-j]}(k)) \otimes E(\phi^{[r-l]}(k)) \otimes I_{\alpha,l+j}) C_{\alpha^j,\alpha^l})\\ &\quad\quad\cdot E(z^{[l+j]}(k))) M^s_{s-j}(\gamma')^T.\end{aligned} \tag{78}$$

Proof: From 3) it follows that for any l, $\Gamma^{[l]}(k)$, $\phi^{[l]}(k)$, and $z^{[l]}(k)$ are mutually independent; hence taking into account (66), (67) we have

$$\begin{aligned}&E(h_{i,\gamma,\alpha}(k, \Gamma, \Psi, z, \phi))\\ &\quad= \sum_{l=0}^{i-1} M^i_{i-l}(\gamma)(\Psi^{[i-l]} \otimes E(\Gamma^{[l]}(k)))\\ &\quad\quad\cdot (E(\overline{\phi}_{i-l}(k)) \otimes I_{\alpha,l}) E(z^{[l]}(k))\\ &\quad\quad+ \sum_{l=1}^{i} M^i_{i-l}(\gamma)(\Psi^{[i-l]} \otimes E(\overline{\Gamma}_l(k)))\\ &\quad\quad\cdot (E(\phi^{[i-l]}(k)) \otimes I_{\alpha,l}) E(z^{[l]}(k))\\ &\quad= 0.\end{aligned}$$

As far as (72) is concerned, taking $j < k$, we have

$$\begin{aligned}&E(h^{(1)}_{r,\gamma,\alpha}(k, \Gamma, \Psi, z, \phi) h^{(1)}_{s,\gamma',\alpha}{}^T(j, \Gamma', \Psi', z, \phi))\\ &\quad= \sum_{i_1=0}^{r-1} \sum_{i_2=0}^{s-1} M^r_{r-i_1}(\gamma)(\Psi^{[r-i_1]} \otimes E(\Gamma^{[i_1]}(k)))\\ &\quad\quad\cdot (E(\overline{\phi}_{r-i_1}(k)) \otimes I_{\alpha,i_1})\\ &\quad\quad\cdot E(z^{[i_1]}(k) z^{[i_2]}(j)(\overline{\phi}_{s-i_2}(j) \otimes I_{\alpha,i_1})^T\\ &\quad\quad\cdot (\Psi'^{[s-i_2]} \otimes \Gamma'^{[i_2]}(j))^T) M^s_{s-i_2}(\gamma')^T\\ &\quad= 0\end{aligned}$$

where Condition 2) and (67) have been exploited.

Similarly, from 2) and (67) again, the result is

$$\begin{aligned}&E(h^{(1)}_{r,\gamma,\alpha}(k,\Gamma,\Psi,z,\phi)h^{(2)}_{s,\gamma',\alpha}{}^T(j,\Gamma',\Psi',z,\phi))\\&=\sum_{i_1=0}^{r-1}\sum_{i_2=1}^{s}M^r_{r-i_1}(\gamma)(\Psi^{[r-i_1]}\otimes E(\Gamma^{[i_1]}(k)))\\&\quad\cdot(E(\overline{\phi}_{r-i_1}(k))\otimes I_{\alpha,i_1})\\&\quad\cdot E(z^{[i_1]}(k)z^{[i_2]}(j)(E(\phi^{[s-i_2]}(j))\otimes I_{\alpha,i_1})^T\\&\quad\cdot(\Psi'^{[s-i_2]}\otimes\overline{\Gamma'}_{i_2}(j))^T)M^s_{s-i_2}(\gamma')^T\\&=0\end{aligned}$$

$$\begin{aligned}&E(h^{(2)}_{r,\gamma,\alpha}(k,\Gamma,\Psi,z,\phi)h^{(1)}_{s,\gamma',\alpha}{}^T(j,\Gamma',\Psi',z,\phi))\\&=\sum_{i_1=1}^{r}\sum_{i_2=0}^{s-1}M^r_{r-i_1}(\gamma)(\Psi^{[r-i_1]}\otimes E(\overline{\Gamma}_{i_1}(k)))\\&\quad\cdot(E(\phi^{[r-i_1]}(k))\otimes I_{\alpha,i_1})\\&\quad\cdot E(z^{[i_1]}(k)z^{[i_2]}(j)(\overline{\phi}_{s-i_2}(j)\otimes I_{\alpha,i_1})^T\\&\quad\cdot(\Psi'^{[s-i_2]}\otimes\Gamma'^{[i_2]}(j))^T)M^s_{s-i_2}(\gamma')^T\\&=0\end{aligned}$$

$$\begin{aligned}&E(h^{(2)}_{r,\gamma,\alpha}(k,\Gamma,\Psi,z,\phi)h^{(2)}_{s,\gamma',\alpha}{}^T(j,\Gamma',\Psi',z,\phi))\\&=\sum_{i_1=1}^{r}\sum_{i_2=1}^{s}M^r_{r-i_1}(\gamma)(\Psi^{[r-i_1]}\otimes E(\overline{\Gamma}_{i_1}(k)))\\&\quad\cdot(E(\phi^{[r-i_1]}(k))\otimes I_{\alpha,i_1})\\&\quad\cdot E(z^{[i_1]}(k)z^{[i_2]}(j)(E(\phi^{[s-i_2]}(j))\otimes I_{\alpha,i_1})^T\\&\quad\cdot(\Psi'^{[s-i_2]}\otimes\overline{\Gamma'}_{i_2}(j))^T)M^s_{s-i_2}(\gamma')^T\\&=0.\end{aligned}$$

In order to prove (73), consider (92); the result is

$$\begin{aligned}&E(h^{(1)}_{r,\gamma,\alpha}(k,\Gamma,\Psi,z,\phi)h^{(1)}_{s,\gamma',\alpha}{}^T(k,\Gamma',\Psi',z,\phi))\\&=\sum_{l=0}^{r-1}\sum_{j=0}^{s-1}M^r_{r-l}(\gamma)\mathrm{st}^{-1}(E(\mathrm{st}(P_{r,s}(k))))M^s_{s-j}(\gamma')\end{aligned}\tag{79}$$

where $P_{r,s}(k)$ is given by

$$\begin{aligned}P_{r,s}(k)=&(\Psi^{[r-l]}\otimes\Gamma^{[l]}(k))(\overline{\phi}_{r-l}(k)\otimes I_{\alpha,l})z^{[l]}(k)z^{[j]}(k)^T\\&\cdot(\overline{\phi}_{s-j}(k)\otimes I_{\alpha,j})^T(\Psi'^{[s-j]}\otimes\Gamma'^{[j]}(k))^T.\end{aligned}$$

By applying Properties (93c) and (93e) it follows that

$$\begin{aligned}\mathrm{st}(P_{r,s}(k))=&(((\Psi'^{[s-j]}\otimes\Gamma'^{[j]}(k))(\overline{\phi}_{s-j}(k)\otimes I_{\alpha,j}))\\&\otimes((\Psi^{[r-l]}\otimes\Gamma^{[l]}(k))(\overline{\phi}_{r-l}(k)\otimes I_{\alpha,l})))\\&\cdot z^{[l+j]}(k)\\=&((\Psi'^{[s-j]}\otimes\Gamma'^{[j]}(k)\otimes\Psi^{[r-l]}\otimes\Gamma^{[l]}(k))\\&\cdot(\overline{\phi}_{s-j}(k)\otimes I_{\alpha,j}\otimes\overline{\phi}_{r-l}(k)\otimes I_{\alpha,l}))\\&\cdot z^{[l+j]}(k).\end{aligned}\tag{80}$$

By applying Corollary B.4 we obtain

$$\begin{aligned}&\Psi'^{[s-j]}\otimes\Gamma'^{[j]}(k)\otimes\Psi^{[r-l]}\otimes\Gamma^{[l]}(k)\\&=(I_{\gamma',s-j}\otimes C^T_{\gamma^r,\gamma^j})(\Psi'^{[s-j]}\otimes\Psi^{[r-l]}\otimes\Gamma^{[l]}(k)\\&\quad\otimes\Gamma'^{[j]}(k))\,(I_{\beta,s-j}\otimes C_{\beta^{r-l}\alpha^l,\alpha^j})\end{aligned}\tag{81}$$

$$\begin{aligned}&\overline{\phi}_{s-j}\otimes I_{\alpha,j}\otimes\overline{\phi}_{r-l}(k)\otimes I_{\alpha,l}\\&=(I_{\beta,s-j}\otimes C^T_{\beta^{r-l}\alpha^l,\alpha^j})\\&\quad\cdot(\overline{\phi}_{s-j}(k)\otimes\overline{\phi}_{r-l}(k)\otimes I_{\alpha,l+j})C_{\alpha^l,\alpha^j}.\end{aligned}\tag{82}$$

By substituting (82) and (81) in (80) and then the result in (79), using Property 3) and taking into account (66), we obtain (73).

Equations (74) and (75) easily follow by applying Property 3):

$$\begin{aligned}&E(h^{(1)}_{r,\gamma,\alpha}(k,\Gamma,\Psi,z,\phi)h^{(2)}_{s,\gamma',\alpha}{}^T(k,\Gamma',\Psi',z,\phi))\\&=\sum_{i_1=0}^{r-1}\sum_{i_2=1}^{s}M^r_{r-i_1}(\gamma)E((\Psi^{[r-i_1]}\otimes\Gamma^{[i_1]}(k)))\\&\quad\cdot(E(\overline{\phi}_{r-i_1}(k))\otimes I_{\alpha,i_1})E(z^{[i_1]}(k)z^{[i_2]}(k))\\&\quad\cdot(E(\phi^{[s-i_2]}(k))\otimes I_{\alpha,i_1})^T\\&\quad\cdot(\Psi'^{[s-i_2]}\otimes\overline{\Gamma'}_{i_2}(k))^T)M^s_{s-i_2}(\gamma')^T\\&=0\end{aligned}$$

$$\begin{aligned}&E(h^{(2)}_{r,\gamma,\alpha}(k,\Gamma,\Psi,z,\phi)h^{(1)}_{s,\gamma',\alpha}{}^T(k,\Gamma',\Psi',z,\phi))\\&=\sum_{i_1=1}^{r}\sum_{i_2=0}^{s-1}M^r_{r-i_1}(\gamma)E((\Psi^{[r-i_1]}\otimes\overline{\Gamma}_{i_1}(k)))\\&\quad\cdot(E(\phi^{[r-i_1]}(k))\otimes I_{\alpha,i_1})E(z^{[i_1]}(k)z^{[i_2]}(k))\\&\quad\cdot(E(\overline{\phi}_{s-i_2}(k))\otimes I_{\alpha,i_1})^T\\&\quad\cdot(\Psi'^{[s-i_2]}\otimes\Gamma'^{[i_2]}(k))^T)M^s_{s-i_2}(\gamma')^T\\&=0.\end{aligned}$$

It remains to prove (76). For this purpose, note that $h^{(2)}_{i,\gamma,\alpha}(k,\Gamma,\Psi,z,\phi)$ is shown to be formally equal to $h^{(1)}_{i,\gamma,\alpha}(k,\Gamma,\Psi,z,\phi)$ with the substitution of $\Gamma^{[l]}(k)$ with $\overline{\Gamma}_l(k)$ and $\overline{\phi}_{i-l}(k)$ with $E(\phi^{[i-l]}(k))$. As a consequence, noting that with these substitutions, and taking into account (66), (78) becomes equal to (77), it follows that (76) holds true too. □

Lemma A.2: Let $w(k)\in\mathbb{R}^\gamma$ be the vector

$$w(k)=\Gamma(k)z(k)+\Psi\phi(k)\tag{83}$$

where $\Psi\in\mathbb{R}^{\gamma\times\beta}$ is a deterministic matrix and $1\le\nu'\le 2\nu$ an integer. Let us consider the augmented vectors

$$\mathcal{W}(k)=\begin{bmatrix}w(k)\\w^{[2]}(k)\\\vdots\\w^{[\nu']}(k)\end{bmatrix},\qquad\mathcal{Z}(k)=\begin{bmatrix}z(k)\\z^{[2]}(k)\\\vdots\\z^{[\nu']}(k)\end{bmatrix}$$

$$\mathcal{T}(k)=\begin{bmatrix}\Psi E(\phi(k))\\\Psi^{[2]}E(\phi^{[2]}(k))\\\vdots\\\Psi^{[\nu']}E(\phi^{[\nu']}(k))\end{bmatrix}$$

and the matrix

$$\mathcal{O} = \begin{bmatrix} E(\Gamma(k)) & 0 & \cdots & 0 \\ O_{2,1}(k) & E(\Gamma^{[2]}(k)) & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots \\ O_{\nu',1}(k) & O_{\nu',2}(k) & \cdots & E(\Gamma^{[\nu']}(k)) \end{bmatrix}$$

where

$$O_{i,l}(k) = M^i_{i-l}(\gamma)(\Psi^{[i-l]} \otimes E(\Gamma^{[l]}(k))) \cdot (E(\phi^{[i-l]}(k)) \otimes I_{\alpha,l}) \tag{84}$$

then there exists the following representation:

$$\mathcal{W}(k) = \mathcal{O}(k)\mathcal{Z}(k) + \mathcal{T}(k) + \mathcal{N}(k) \tag{85}$$

with $\mathcal{N}(k)$ defined as

$$\mathcal{N}(k) = \begin{bmatrix} h_{1,\gamma,\alpha}(k, \Gamma, \Psi, z, \phi) \\ h_{2,\gamma,\alpha}(k, \Gamma, \Psi, z, \phi) \\ \vdots \\ h_{\nu',\gamma,\alpha}(k, \Gamma, \Psi, z, \phi) \end{bmatrix} \tag{86}$$

where $h_{i,\gamma,\alpha}$ $i = 1, \cdots, \nu'$ are defined as (68).

Proof: Let us consider the ith Kronecker power of both sides of (83)

$$w^{[i]}(k) = (\Gamma(k)z(k) + \Psi\phi(k))^{[i]}.$$

Using Theorem B.6 and Property (93c) we have

$$\begin{aligned} w^{[i]}(k) &= (\Gamma(k)z(k))^{[i]} + \sum_{j=1}^{i} M^i_j(\gamma) \\ &\quad \cdot ((\Psi\phi(k))^{[j]} \otimes (\Gamma(k)z(k))^{[i-j]}) \\ &= \Gamma^{[i]}(k)z^{[i]}(k) + \sum_{j=1}^{i} M^i_j(\gamma) \\ &\quad \cdot (\Psi^{[j]} \otimes \Gamma^{[i-j]}(k))(\phi^{[j]}(k) \otimes z^{[i-j]}(k)) \\ &= \Gamma^{[i]}(k)z^{[i]}(k) + \sum_{l=0}^{i-1} M^i_{i-l}(\gamma) \\ &\quad \cdot (\Psi^{[i-l]} \otimes \Gamma^{[l]}(k))(\phi^{[i-l]}(k) \otimes I_{\alpha,l})z^{[l]}(k) \end{aligned}$$

and by adding and subtracting to $\phi^{[i-l]}(k)$, $l = 0, \cdots, i-1$, their expected values, we obtain

$$\begin{aligned} w^{[i]}(k) &= \Gamma^{[i]}(k)z^{[i]}(k) + \sum_{l=0}^{i-1} M^i_{i-l}(\gamma)(\Psi^{[i-l]} \otimes \Gamma^{[l]}(k)) \\ &\quad \cdot (E(\phi^{[i-l]}(k)) \otimes I_{\alpha,l})z^{[l]}(k) + \sum_{l=0}^{i-1} M^i_{i-l}(\gamma) \\ &\quad \cdot (\Psi^{[i-l]} \otimes \Gamma^{[l]}(k))(\overline{\phi}_{i-l}(k) \otimes I_{\alpha,l})z^{[l]}(k). \end{aligned}$$

By adding and subtracting to $\Gamma^{[i]}(k)$ and $\Gamma^{[l]}(k)$, $l = 0, \cdots, i-1$, their expected values, in the first two terms of the right side of the previous expression the result is

$$\begin{aligned} w^{[i]}(k) &= E(\Gamma^{[i]}(k))z^{[i]}(k) + \overline{\Gamma}_i(k)z^{[i]}(k) \\ &\quad + \sum_{l=0}^{i-1} M^i_{i-l}(\gamma)(\Psi^{[i-l]} \otimes E(\Gamma^{[l]}(k))) \\ &\quad \cdot (E(\phi^{[i-l]}(k)) \otimes I_{\alpha,l})z^{[l]}(k) \\ &\quad + \sum_{l=0}^{i-1} M^i_{i-l}(\gamma)(\Psi^{[i-l]} \otimes \overline{\Gamma}_l(k)) \\ &\quad \cdot (E(\phi^{[i-l]}(k)) \otimes I_{\alpha,l})z^{[l]}(k) \\ &\quad + \sum_{l=0}^{i-1} M^i_{i-l}(\gamma)(\Psi^{[i-l]} \otimes \Gamma^{[l]}(k)) \\ &\quad \cdot (\overline{\phi}_{i-l}(k) \otimes I_{\alpha,l})z^{[l]}(k) \end{aligned}$$

which can be rewritten as

$$\begin{aligned} w^{[i]}(k) &= E(\Gamma^{[i]}(k))z^{[i]}(k) \\ &\quad + \sum_{l=0}^{i-1} M^i_j(\gamma)(\Psi^{[i-l]} \otimes E(\Gamma^{[l]}(k))) \\ &\quad \cdot (E(\phi^{[i-l]}(k)) \otimes I_{\alpha,l})z^{[l]}(k) + \Psi^{[i]}E(\phi^{[i]}(k)) \\ &\quad + h_{i,\gamma,\alpha}(k, \Gamma, \Psi, z, \phi) \end{aligned} \tag{87}$$

where the $h_{i,\cdot,\cdot}$ are defined by (68)–(70). By aggregating in a vector the $w^{[i]}(k)$, $i = 1, \cdots, \nu'$ given by (87) and taking into account (66), we obtain (85). □

Proof of Theorem 5.1: Let us apply Lemmas A.1 and A.2 by setting

$$\Gamma(k) = \tilde{A}_e(k), \quad \Gamma'(k) = \tilde{A}_e(k), \quad \Psi = F, \quad \Psi' = F$$
$$\nu' = \nu, \quad z(k) = x_e(k), \quad \phi(k) = N(k)$$

and with these choices, from Conditions 1) and 2) of Section III, it follows that Properties 1)–3) are satisfied; moreover, we have $w(k) = x_e(k+1)$, $\gamma = \gamma' = q$, $\alpha = q$, $\beta = n+m$, and then (21) holds true, with $\mathcal{F}(k)$ given by

$$\mathcal{F}(k) = \begin{bmatrix} h_{1,q,q}(k, \tilde{A}_e, F, x_e, N) \\ h_{2,q,q}(k, \tilde{A}_e, F, x_e, N) \\ \vdots \\ h_{\nu,q,q}(k, \tilde{A}_e, F, x_e, N) \end{bmatrix}. \tag{88}$$

From (88), (72) it follows that the sequence $\{\mathcal{F}(k)\}$ is uncorrelated. Moreover, from (88) and (73)–(76) follows (24) with

$$\mathcal{Q}^{(1)}_{r,s}(k) = T^{(1)}_{r,s,q,q,q,n+m}(k, \tilde{A}_e, \tilde{A}_e, F, F, x_e, N)$$
$$\mathcal{Q}^{(2)}_{r,s}(k) = T^{(2)}_{r,s,q,q,q,n+m}(k, \tilde{A}_e, \tilde{A}_e, F, F, x_e, N)$$

from which (25) and (26) follow, taking into account (77), (78).

Now, let us apply Lemmas A.1 and A.2 by setting

$$\Gamma(k) = C_e(k), \quad \Gamma'(k) = C_e(k), \quad \Psi = G, \quad \Psi' = G$$
$$\nu' = \nu, \quad z(k) = x_e(k), \quad \phi(k) = N(k).$$

Then, Properties 1)–3) are again verified and we have $w(k) = y_e(k)$, $\gamma = \gamma' = p$, $\alpha = q$, $\beta = n+m$. Hence, (21) holds with $\mathcal{G}(k)$ given by

$$\mathcal{G}(k) = \begin{bmatrix} h_{1,p,q}(k, C_e, G, x_e, N) \\ h_{2,p,q}(k, C_e, G, x_e, N) \\ \vdots \\ h_{\nu,p,q}(k, C_e, G, x_e, N) \end{bmatrix}. \tag{89}$$

From (89) and (72) the uncorrelation of the sequence $\{\mathcal{G}(k)\}$ follows. Moreover, since $C_e(k) = E(C_e(k))$, from (78) we have that

$$T^{(2)}_{r,s,p,p,q,n+m}(k, C_e, C_e, G, G, x_e, N) = 0$$

and then from (88) and (89), (28) follows.

From (51), (52) and applying Lemmas A.1 and A.2 with

$$\Gamma(k) = \tilde{A}_e(k), \qquad \Gamma'(k) = \tilde{C}_e(k), \qquad \Psi = F, \qquad \Psi' = G$$
$$\nu' = \nu \qquad z(k) = x_e(k), \qquad \phi(k) = N(k)$$

(hence $\gamma = q$, $\gamma' = p$, $\alpha = q$, $\beta = n+m$) (23) follows. With the same assignments, from (31) we have

$$T^{(2)}_{r,s,q,p,q,n+m}(k, \tilde{A}_e, C_e, F, G, x_e, N) = 0$$

and then from (73)–(77), (28) follows, giving the cross-correlation matrix between augmented noises. □

Now, we can also prove Theorem 5.4.

Proof of Theorem 5.4: Let us apply Lemma A.2 by setting

$$\Gamma(k) = \tilde{A}_e(k), \qquad \Psi = F, \qquad \nu' = 2\nu$$
$$z(k) = x_e(k), \qquad \phi(k) = N(k).$$

These choices yield $w(k) = x_e(k+1)$ and $\gamma = p$, hence (85) has in this case the following form:

$$\mathcal{X}_{2\nu}(k+1) = \mathcal{A}_{2\nu}(k)\mathcal{X}_{2\nu}(k) + \mathcal{U}_{2\nu}(k) + \mathcal{F}_{2\nu}(k) \tag{90}$$

where

$$\mathcal{X}_{2\nu}(k) = \begin{bmatrix} x_e(k) \\ x_e^{[2]}(k) \\ \vdots \\ x_e^{[2\nu]}(k) \end{bmatrix}$$
$$\mathcal{F}_{2\nu}(k) = \begin{bmatrix} h_{1,q,q}(k, \tilde{A}_e, F, x_e, N) \\ h_{2,q,q}(k, \tilde{A}_e, F, x_e, N) \\ \vdots \\ h_{2\nu,q,q}(k, \tilde{A}_e, F, x_e, N) \end{bmatrix}.$$

By applying Lemma A.1, from (71) it follows that $E(h_{i,q,q}(k, \tilde{A}_e, F, x_e, N)) = 0$, $i = 1, \cdots, 2\nu$ and then $E(\mathcal{F}_{2\nu}(k)) = 0$. Hence, taking the expectations on both sides of (90), (40) follows. □

APPENDIX B
KRONECKER ALGEBRA

Throughout this paper, we have widely used Kronecker algebra [21]. Here, for the sake of completeness, we recall some definitions and properties and also give some new results on this subject.

Definition B.1: Let M and N be matrices of dimension $r \times s$ and $p \times q$, respectively. Then the Kronecker product $M \otimes N$ is defined as the $(r \cdot p) \times (s \cdot q)$ matrix

$$M \otimes N = \begin{bmatrix} m_{11}N & \cdots & m_{1s}N \\ \cdots & \cdots & \cdots \\ m_{r1}N & \cdots & m_{rs}N \end{bmatrix}$$

where the m_{ij} are the entries of M.

Of course, this kind of product is not commutative.

Definition B.2: Let M be the $r \times s$ matrix

$$M = [m_1 \quad m_2 \quad \cdots \quad m_s] \tag{91}$$

where m_i denotes the ith column of M, then the stack of M is the $r \cdot s$ vector

$$\text{st}(M) = [m_1^T \quad m_2 \quad \cdots \quad m_s]^T. \tag{92}$$

Observe that a vector such as in (92) can be reduced to a matrix M as in (91) by considering the inverse operation of the stack denoted by st^{-1}. With reference to the Kronecker product and the stack operation, the following properties hold [21]:

$$(A+B) \otimes (C+D) = A \otimes C + A \otimes D + B \otimes C + B \otimes D \tag{93a}$$
$$A \otimes (B \otimes C) = (A \otimes B) \otimes C \tag{93b}$$
$$(A \cdot C) \otimes (B \cdot D) = (A \otimes B) \cdot (C \otimes D) \tag{93c}$$
$$(A \otimes B)^T = A^T \otimes B^T \tag{93d}$$
$$\text{st}(A \cdot B \cdot C) = (C^T \otimes A) \cdot \text{st}(B) \tag{93e}$$
$$u \otimes v = \text{st}(v \cdot u^T) \tag{93f}$$
$$\text{tr}(A \otimes B) = \text{tr}(A) \cdot \text{tr}(B) \tag{93g}$$

where A, B, C, D are suitably dimensioned matrices, u, v are vectors, and $tr(M)$ denotes the trace of a square matrix M. The Kronecker power of the matrix M is defined as

$$M^{[0]} = 1$$
$$M^{[n]} = M \otimes M^{[n-1]} = M^{[n-1]} \otimes M, \qquad n > 0.$$

As an easy consequence of (93b) and (93g), it follows that

$$\text{tr}(A^{[h]}) = (\text{tr}(A))^h. \tag{93h}$$

It is easy to verify that for $u \in \mathbb{R}^r$, $v \in \mathbb{R}^s$, the ith entry of $u \otimes v$ is given by

$$(u \otimes v)_i = u_l \cdot v_m, \qquad l = \left[\frac{i-1}{s}\right] + 1, \qquad m = |i-1|_s + 1 \tag{94}$$

where $[\cdot]$ and $|\cdot|_s$ denote the integer part and s-modulo, respectively. Even if the Kronecker product is not commutative, the following property holds [20], [23].

Theorem B.3: For any given pair of matrices $A \in \mathbb{R}^{r\times s}$, $B \in \mathbb{R}^{n\times m}$, we have

$$B \otimes A = C_{r,n}^T (A \otimes B) C_{s,m} \tag{95}$$

where the commutation matrix $C_{u,v}$ is the $(u \cdot v) \times (u \cdot v)$ matrix such that its (h, l) entry is given by

$$\{C_{u,v}\}_{h,l} = \begin{cases} 1, & \text{if } l = (|h-1|_v)u + \left(\left[\dfrac{h-1}{v}\right] + 1\right); \\ 0, & \text{otherwise.} \end{cases} \tag{96}$$

Observe that $C_{1,1} = 1$, hence in the vector case when $a \in \mathbb{R}^r$ and $b \in \mathbb{R}^n$, (95) becomes

$$b \otimes a = C_{r,n}^T (a \otimes b). \tag{97}$$

Corollary B.4: For any given matrices A, B, C, D having dimensions $n_A \times m_A$, $n_B \times m_B$, $n_C \times m_C$, $n_D \times m_D$, respectively, denoted with $I(l)$, the identity matrix in $\mathbb{R}^l$, we have

$$\begin{aligned} A \otimes B \otimes C \otimes D = &(I(n_A) \otimes C_{n_C n_D, n_B}^T) \\ &\cdot (A \otimes C \otimes D \otimes B) \\ &\cdot (I(m_A) \otimes C_{m_C m_D, m_B}). \end{aligned}$$

Proof: By applying Properties (93b) and (93c) and Theorem B.3 we have

$$\begin{aligned} &A \otimes B \otimes C \otimes D \\ &\quad = (A \otimes (B \otimes (C \otimes D))) \\ &\quad = (A \otimes (C_{n_C n_D, n_B}^T (C \otimes D \otimes B) C_{m_C m_D, m_B})) \\ &\quad = (I(n_A) \otimes C_{n_C n_D, n_B}^T) \\ &\qquad \cdot (A \otimes ((C \otimes D \otimes B) C_{m_C m_D, m_B})) \\ &\quad = (I(n_A) \otimes C_{n_C n_D, n_B}^T)(A \otimes C \otimes D \otimes B) \\ &\qquad \cdot (I(m_A) \otimes C_{m_C m_D, m_B}). \end{aligned}$$

□

Moreover, let us recall the following recursive formula [20].

Lemma B.5: For any $a, b \in \mathbb{R}^n$ and for any $l = 1, 2 \cdots$, let G_l be the $n^{(l+1)} \times n^{(l+1)}$ matrix such that

$$b^{[l]} \otimes a = G_l (a \otimes b^{[l]}). \tag{98}$$

Then the sequence $\{G_l\}$ satisfies the following equations:

$$\begin{aligned} G_1 &= C_{n,n}^T \\ G_l &= (I_1 \otimes G_{l-1}) \cdot (G_1 \otimes I_{l-1}), \qquad l > 1 \end{aligned} \tag{99}$$

where I_r is the identity matrix in $\mathbb{R}^{n^r}$.

In [20] can be found the proof of a binomial formula for the Kronecker power, which generalizes the classical Newton one, as is asserted by the following theorem.

Theorem B.6: For any integer $h \geq 0$ the matrix coefficients of the following binomial power formula:

$$(a+b)^{[h]} = \sum_{k=0}^{h} M_k^h (a^{[k]} \otimes b^{[h-k]}) \tag{100}$$

constitute a set of matrices $\{M_0^h, \cdots, M_h^h\}$ such that for $1 \leq j \leq h-1$

$$M_h^h = M_0^h = I_h \tag{101}$$

$$M_j^h = (M_j^{h-1} \otimes I_1) + (M_{j-1}^{h-1} \otimes I_1)(I_{j-1} \otimes G_{h-j}) \tag{102}$$

where G_l and I_l are as in Lemma B.5.

Lemmas B.7 and B.9 and Corollary B.8 constitute new results about Kronecker algebra.

Lemma B.7: Given $A \in \mathbb{R}^{n\times m}$, $B \in \mathbb{R}^{r\times s}$, there exists a matrix $D_{n,m,r,s} \in \mathbb{R}^{(n\cdot m\cdot r\cdot s\cdot)^2}$ such that

$$\mathrm{st}(A \otimes B) = D_{n,m,r,s}(\mathrm{st}(A) \otimes \mathrm{st}(B))$$

where

$$D_{n,m,r,s} = (I(s \cdot m) \otimes C_{r,n}^T)(I(m) \otimes C_{n,r\cdot s}^T)$$

and $I(l)$ is the identity in $\mathbb{R}^l$, $\forall\, l \in \mathbb{N}$.

Proof: Let us express the vector $\mathrm{st}(A \otimes B)$ as

$$\begin{aligned} &\mathrm{st}(A \otimes B) \\ &\quad = [(a_1 \otimes b_1)^T \quad \cdots \quad (a_1 \otimes b_s)^T \quad \cdots \quad (a_m \otimes b_s)^T\,]^T \end{aligned} \tag{103}$$

where a_i, b_i are the ith column of A and B, respectively. Using Theorem B.3, (103) can be rewritten as

$$\begin{aligned} &\mathrm{st}(A \otimes B) \\ &\quad = [(b_1 \otimes a_1)^T C_{r,n} \quad \cdots \quad (b_s \otimes a_m)^T C_{r,n}\,]^T \\ &\quad = (I(s \cdot m) \otimes C_{r,n}^T) \begin{bmatrix} \mathrm{st}(B) \otimes a_1 \\ \vdots \\ \mathrm{st}(B) \otimes a_m \end{bmatrix} \\ &\quad = (I(s \cdot m) \otimes C_{r,n}^T)(I(m) \otimes C_{n,r\cdot s}^T) \\ &\qquad \cdot (\mathrm{st}(A) \otimes \mathrm{st}(B)) \end{aligned}$$

so that the proof is completed. □

Corollary B.8: Given a matrix $A \in \mathbb{R}^{n\times m}$, $\forall\, h \in \mathbb{N}$ there exists a matrix $D_{n,m}^{(h)} \in \mathbb{R}^{(n\cdot m)^h \times (n\cdot m)^h}$ such that

$$\mathrm{st}(A^{[h]}) = D_{n,m}^{(h)} \cdot (\mathrm{st}(A))^{[h]} \tag{104}$$

where

$$D_{n,m}^{(h)} = \begin{cases} I(n \cdot m), & \text{if } h = 1; \\ D_{n^{h-1}, m^{h-1}, n, m} \cdot (D_{n,m}^{(h-1)} \otimes I(n \cdot m)), & \text{if } h > 1. \end{cases} \tag{105}$$

Proof: Equation (104) is obviously true for $h = 1$. Let $h > 1$; by supposing (104) true for $h-1$ with $D_{n,m}^{(h-1)}$ as in (105), we obtain

$$\begin{aligned}\mathrm{st}(A^{[h]}) &= \mathrm{st}(A^{[h-1]} \otimes A)\\ &= D_{n^{h-1}, m^{h-1}, n, m}(\mathrm{st}(A^{[h-1]}) \otimes \mathrm{st}(A))\\ &= D_{n^{h-1}, m^{h-1}, n, m}((D_{n,m}^{(h-1)}(\mathrm{st}(A))^{[h-1]}) \otimes \mathrm{st}(A))\\ &= D_{n^{h-1}, m^{h-1}, n, m}(D_{n,m}^{(h-1)} \otimes I(n \cdot m))(\mathrm{st}(A))^{[h]}\end{aligned}$$

from which the thesis follows. □

We can also generalize formula (100) to the polynomial case. Obviously, given any polynomial $a_1 + \cdots + a_p$, $a_i \in \mathbb{R}^n$, $1 \le i \le p$, $p \in \mathbb{N}$, its hth Kronecker power admits a representation as

$$(a_1 + a_2 + \cdots + a_p)^{[h]} = \sum_{\substack{h_1, \cdots, h_p \ge 0\\ h_1 + \cdots + h_p = h}} M_{h_1, \cdots, h_p}^h (a_1^{[h_1]} \otimes a_2^{[h_2]} \otimes \cdots \otimes a_p^{[h_p]}) \tag{106}$$

where $M_{h_1, \cdots, h_p}^h$ are suitable matrices. We extend the definition of symbol $M_{l_1, \cdots, l_s}^l$, with $l > 0$ when at least one of the l_i's is negative, such as

$$M_{l_1, \cdots, l_s}^l = 0 \in \mathbb{R}^{n^l \times n^l}. \tag{107}$$

Moreover, we can prove the following statement.

Lemma B.9: The matrices $M_{h_1, \cdots, h_p}^h \in \mathbb{R}^{n^h \times n^h}$ in (106) satisfy the recursive formula

$$M_{h_1, \cdots, h_p}^h = I_1, \qquad \text{for} \quad h = 1 \tag{108}$$

$$\begin{aligned}M_{h_1, \cdots, h_p}^h = &\sum_{1 \le i \le p-1} (M_{h_1, \cdots, h_i - 1, \cdots, h_p}^{h-1} \otimes I_1)\\ &\cdot (I_{h_1 + \cdots + h_i - 1} \otimes G_{h_{i+1} + \cdots + h_p})\\ &+ (M_{h_1, \cdots, h_p - 1}^{h-1} \otimes I_1) \qquad \text{for} \quad h > 1.\end{aligned} \tag{109}$$

Proof: Equation (108) is obvious. In order to prove (109), let us consider the polynomial power

$$\begin{aligned}(a_1 + \cdots + a_p)^{[h]} &= (a_1 + \ldots + a_p)^{[h-1]} \otimes (a_1 + \cdots + a_p)\\ &= \left(\sum_{j_1 + \cdots + j_p = h-1} M_{j_1, \cdots, j_p}^{h-1}(a_1^{[j_1]} \otimes \cdots \otimes a_p^{[j_p]})\right)\\ &\quad \otimes \sum_{i=1}^{p} a_i\\ &= \sum_{i=1}^{p} \sum_{j_1 + \cdots + j_p = h-1}\\ &\quad \cdot ((M_{j_1, \cdots, j_p}^{h-1}(a_1^{[j_1]} \otimes \cdots \otimes a_p^{[j_p]})) \otimes a_i).\end{aligned}$$

Now, let us consider the term

$$(M_{j_1, \cdots, j_p}^{h-1}(a_1^{[j_1]} \otimes \cdots \otimes a_p^{[j_p]})) \otimes a_i.$$

If $i = p$, it is equal to

$$\begin{aligned}&(M_{j_1, \cdots, j_p}^{h-1} \otimes I_1)(a_1^{[j_1]} \otimes \cdots \otimes a_p^{[j_p]} \otimes a_p)\\ &\quad = (M_{j_1, \cdots, j_p}^{h-1} \otimes I_1)(a_1^{[j_1]} \otimes \cdots \otimes a_p^{[j_p+1]}).\end{aligned}$$

If $i \ne p$, then

$$\begin{aligned}&(M_{j_1, \cdots, j_p}^{h-1}(a_1^{[j_1]} \otimes \cdots \otimes a_p^{[j_p]})) \otimes a_i\\ &\quad = (M_{j_1, \cdots, j_p}^{h-1} \otimes I_1)((a_1^{[j_1]} \otimes \cdots \otimes a_i^{[j_i]})\\ &\qquad \otimes (G_{j_{i+1} + \cdots + j_p} \cdot (a_i \otimes a_{i+1}^{[j_{i+1}]} \otimes \cdots \otimes a_p^{[j_p]})))\\ &\quad = (M_{j_1, \cdots, j_p}^{h-1} \otimes I_1)(I_{j_1 + \cdots + j_i} \otimes G_{j_{i+1} + \cdots + j_p})\\ &\qquad \cdot (a_1^{[j_1]} \otimes \cdots \otimes a_i^{[j_i+1]} \otimes \cdots \otimes a_p^{[j_p]})\end{aligned}$$

then, taking into account (106) we can write

$$\begin{aligned}&\sum_{\substack{h_1, \cdots, h_p \ge 0\\ h_1 + \cdots + h_p = h}} M_{h_1, \cdots, h_p}^h (a_1^{[h_1]} \otimes \cdots \otimes a_p^{[h_p]})\\ &= \sum_{i=1}^{p-1} \sum_{j_1 + \cdots + j_p = h-1}\\ &\quad \cdot ((M_{j_1, \cdots, j_p}^{h-1} \otimes I_1)(I_{j_1 + \cdots + j_i} \otimes G_{j_{i+1} + \cdots + j_p})\\ &\quad \cdot (a_1^{[j_1]} \otimes \cdots \otimes a_i^{[j_i+1]} \otimes \cdots \otimes a_p^{[j_p]}))\\ &\quad + \sum_{j_1 + \cdots + j_p = h-1} (M_{j_1, \cdots, j_p}^{h-1} \otimes I_1)(a_1^{[j_1]} \otimes \cdots \otimes a_p^{[j_p+1]}).\end{aligned} \tag{110}$$

Now, by considering the generic term of the summation on the left-hand side of (110), that is $M_{h_1, \cdots, h_p}^{h-1}(a_1^{[h_1]} \otimes \cdots \otimes a_p^{[h_p]})$, we must look at the RHS for those terms which are characterized by the indexes $h_1, \cdots, h_p$. They are of the form

$$\begin{aligned}&(M_{j_1, \cdots, j_p}^{h-1} \otimes I_1)(I_{j_1 + \cdots + j_i} \otimes G_{j_{i+1} + \cdots + j_p})\\ &\quad \cdot (a_1^{[j_1]} \otimes \cdots \otimes a_i^{[j_i+1]} \otimes \cdots \otimes a_p^{[j_p]})\end{aligned}$$

with $j_1 = h_1$, $j_2 = h_2, \cdots, j_i = h_i - 1, \cdots, j_p = h_p$, for $i = 1, \cdots, p-1$, whenever $h_i \ne 0$, and

$$(M_{j_1, \cdots, j_p}^{h-1} \otimes I_1)(a_1^{[j_1]} \otimes \cdots \otimes a_p^{[j_p+1]})$$

with $j_1 = h_1$, $j_2 = h_2, \cdots$ $j_p = h_p - 1$. Then, taking into account (107), (109) is proved. □

REFERENCES

[1] R. R. Mohler and C. N. Shen, *Optimal Control of Nuclear Reactors.* New York: Academic, 1970.

[2] R. R. Mohler, *Bilinear Control Processes.* New York: Academic, 1970.

[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. AC-19, Aug. 1974.

[4] C. Bruni and G. Koch, "Suboptimal filtering for linear-in-control systems with small observation noise," in *Proc. IV IFAC Symp. Identification Syst. Parameter Estimation,* Tblisi, Sept. 1976.

[5] X. Yang, R. R. Mohler, and R. M. Burton, "Adaptive suboptimal filtering of bilinear systems," *Int. J. Contr.,* vol. 52, no. 1, 1990.

[6] R. Luesink and H. Nijmeijer "On the stabilization of bilinear systems via constant feedback," in *Linear Algebra and Its Applications.* New York: Elsevier, 1989.

[7] L. K. Chen, X. Yang, and R. R. Mohler, "Stability analysis of bilinear systems," *IEEE Trans. Automat. Contr.,* vol. 36, Nov. 1991.

[8] X. Yang and L. K. Chen, "Stability of discrete bilinear systems with time delayed feedback functions," *IEEE Trans. Automat. Contr.,* vol. 38, Jan. 1993.
[9] ——, "Stability of discrete bilinear systems with time-delayed feedback functions," *IEEE Trans. Automat. Contr.,* vol. 38, Jan. 1993.
[10] S. P. Banks and M. K. Yew, "On the optimal control of bilinear systems and its relation to lie algebras," *Int. J. Contr.,* vol. 43, no. 3, 1986.
[11] J. L. Willems and J. C. Willems, "Robust stabilization of uncertain systems," *SIAM J. Contr. Opt.,* vol. 21, no. 3, May 1983.
[12] M. S. Ahmed, "Parameter estimation in bilinear systems by instrumental variable method," *Int. J. Contr.,* vol. 44, no. 4, 1986.
[13] M. M. Gabr, "A recursive (on-line) identification of bilinear systems," *Int. J. Contr.,* vol. 44, no. 4, pp. 911–917, 1986.
[14] H. Dai, N. K. Sinha, and S. C. Puthenpura, "Robust combined estimation of states and parameters of bilinear systems," *Automatica,* vol. 25, no. 4, 1989.
[15] R. R. Mohler and W. J. Kolodzej, "An overview of stochastic bilinear control processes," *IEEE Trans. Syst., Man Cybern.,* vol. SMC-10, Dec. 1980.
[16] E. Yaz, "Robust, exponentially fast state estimator for some nonlinear stochastic systems," *Int. J. Syst. Sci.,* vol. 23, no. 4, 1992.
[17] ——, "Full and reduced order observer design for discrete stochastic bilinear systems," *IEEE Trans. Automat. Contr.,* vol. 37, Apr. 1992.
[18] W. L. De Koning, "Optimal estimation of linear discrete-time systems with stochastic parameters," *Automatica,* vol. 20, no. 1, 1984.
[19] A. De Santis, A. Germani, and M. Raimondi, "Optimal quadratic filtering of linear discrete time non-Gaussian systems," *IEEE Trans. Automat. Contr.,* vol. 40, June 1995.
[20] F. Carravetta, A. Germani, and M. Raimondi, "Polynomial filtering for linear discrete time non-Gaussian systems," *SIAM J. Contr. Opt.,* vol. 34, Sept. 1996.
[21] R. Bellman, *Introduction to Matrix Analysis.* New York: McGraw-Hill, 1970.
[22] A. V. Balakrishnan, *Kalman Filtering Theory.* New York: Optimization Software, 1984.
[23] G. S. Rodgers, *Matrix Derivatives.* New York: Marcel Dekker, 1980.
[24] X. Feng, K. A. Loparo, Y. Ji, and H. J. Chizeck, "Stochastic stability properties of jump linear systems," *IEEE Trans. Automat. Contr.,* vol. 37, Jan. 1992.
[25] Y. Ji and H. J. Chizeck, "Bounded sample path control of discrete time jump linear system," *IEEE Trans. Syst., Man Cybern.,* vol. 19, Mar. 1989.
[26] J. K. Tugnait, "Detection and estimation for abruptly changing systems," *Automatica,* vol. 18, no. 5, 1982.
[27] P. Andersson, "Adaptive forgetting in recursive identification through multiple models," *Int. J. Contr.,* vol. 42, no. 5, 1985.
[28] H. A. P. Blom and Y. Bar-Shalom, "Time-reversion of a hybrid state stochastic difference system with a jump-linear smoothing application," *IEEE Trans. Inform. Theory,* vol. 36, July 1990.
[29] R. E. Helmick and W. D. Blair, "Fixed interval smoothing for Markovian switching systems," *IEEE Trans. Inform. Theory,* vol. 41, Nov. 1995.
[30] A. S. Willsky, E. Y. Chow, S. B. Gershwin, C. S. Greene, P. K. Houpt, and A. L. Kurkjian, "Dynamic model-based techniques for the detection of incidents on freeways," *IEEE Trans. Automat. Contr.,* vol. 25, June 1980.
[31] A. S. Willsky, "A survey of design methods for failure detection in dynamic systems," *Automatica,* vol. 12, p. 601, 1976.
[32] R. E. Kalman, "A new approach to linear filtering and prediction problems," *ASME J. Basic Eng.,* series D, vol. 82, 1960.

Francesco Carravetta received the Laurea degree in electrical engineering from the University of Calabria, Italy, in 1988 and the Ph.D. degree in systems engineering from the University "La Sapienza," Rome, in 1995.

After graduating, he served as an Engineer on the staff of the Research Laboratory of Alcatel-Face Company, Pomezia, Italy. Since 1990, he has worked as a Researcher at the Institute of Systems Analysis and Computer Science (IASI) of the National Research Council (CNR) in Rome. His research interests include theory and applications of filtering, prediction, detection, identification, and control of stochastic dynamical systems.

Alfredo Germani received the Laurea degree in physics in 1972 and the Postgraduate Diploma in systems and control theory in 1974, both from the University "La Sapienza," Rome.

From 1975 to 1986 he worked as a Researcher at the Institute of Systems Analysis and Computer Science (IASI) of the National Research Council (CNR) in Rome. In 1978 and 1979, he was at the Department of System Science at the University of California, Los Angeles, as a Visiting Scholar. From 1981 to 1983, he was a Contract Professor of Systems and Identification at the University of Ancona, Italy. From 1986 to 1987, he was a Professor of Automatic Control at the University of Calabria, Italy. Since 1987, he has been a Full Professor of System Theory at the University of L'Aquila, Italy. He is now teaching courses on System Theory, Systems Identification, and Data Analysis and Nonlinear, Stochastic, and Optimal Control Theory. Since 1986, he has been a Coordinator of the System and Control Theory Research Group of the IASI-CNR, Rome. He has published more than 80 research papers in the fields of nonlinear systems, distributed systems, finitely additive white noise theory, approximation theory, nonlinear filtering, image processing, and mathematical modeling for biological processes.

Dr. Germani is member of the SIAM Society, the American Mathematical Society, and of the International Federation of Nonlinear Analysts.

Massimo Raimondi received the Laurea degree in 1990 and the Ph.D. in 1995, both in mathematics, from the University "La Sapienza," Rome.

He is currently a Teacher of Mathematics in a high school in Rome and has a research appointment at the Institute of Systems Analysis and Computer Science (IASI) of the National Research Council (CNR) in Rome. His interests include the probability theory and its applications.