SIAM J. CONTROL OPTIM.
Vol. 38, No. 4, pp. 1171–1203


# A NEW SUBOPTIMAL APPROACH TO THE FILTERING PROBLEM FOR BILINEAR STOCHASTIC DIFFERENTIAL SYSTEMS*

FRANCESCO CARRAVETTA[†], ALFREDO GERMANI[†‡], AND MARAT K. SHUAKAYEV[§]

**Abstract.** The aim of this paper is to present a new approach to the filtering problem for the class of bilinear stochastic multivariable systems, consisting in searching for suboptimal state-estimates instead of the conditional statistics. As a first result, a finite-dimensional optimal linear filter for the considered class of systems is defined. Then, the more general problem of designing polynomial finite-dimensional filters is considered. The equations of a finite-dimensional filter are given, producing a state-estimate which is optimal in a class of polynomial transformations of the measurements with arbitrarily fixed degree. Numerical simulations show the effectiveness of the proposed filter.

**Key words.** square integrable martingales, wide-sense Wiener processes, stochastic bilinear systems, Kronecker algebra, Kalman–Bucy filtering, polynomial filtering, vector Ito formula

**AMS subject classifications.** 93E10, 93E11, 60H10

**PII.** S0363012997320912

## 1. Introduction.

Let us consider the class of nonlinear stochastic systems defined on some probability space, namely $(\Omega, \mathcal{F}, P)$, described by the Ito equations

$$dX(t) = A(t)X(t)dt + B^1(X(t), dW(t)), \tag{1.1}$$

$$dY(t) = C(t)\big(X(t)\big)dt + B^2(X(t), dW(t)), \tag{1.2}$$

where $X(t) \in \mathbf{R}^n$; $Y(t) \in \mathbf{R}^q$; $W(t) \in \mathbf{R}^p$ is a standard Wiener process with respect to some increasing family of $\sigma$-algebras, namely $\{\mathcal{F}_t\}$; $A(t), C(t)$ are matrices of proper dimensions; $B^1$ and $B^2$ are bilinear forms. System (1.1), (1.2) is commonly referred to in the literature as a *bilinear stochastic system* (BLSS) [4], [5], [6], [7], [8], [10].

The problem we are faced with consists in searching for finite-dimensional filters for the BLSS (1.1), (1.2). Indeed, for such a system even the *linear* optimal finite-dimensional filtering problem is still an interesting one.

With the name of finite-dimensional filter, we understand a stochastic differential equation in the form

$$dz(t) = f(z(t))dt + g(z(t))dY(t), \tag{1.3}$$

endowed with an output transformation

$$\hat{X}(t) = h(z(t)), \tag{1.4}$$

where $\{z(t),\ t > 0\}$ is some process taking values on a finite-dimensional linear space. We say that (1.3), (1.4) is a finite-dimensional optimal filter for system (1.1), (1.2) if

$$\hat{X}(t) = E(X(t)/\mathcal{F}_t^Y), \tag{1.5}$$

*Received by the editors May 8, 1997; accepted for publication (in revised form) September 21, 1999; published electronically April 18, 2000. This work was partially supported by the Progetto Finalizzato Trasporti 2 of the Italian National Research Council (CNR).

http://www.siam.org/journals/sicon/38-4/32091.html

[†]Istituto di Analisi dei Sistemi e Informatica del CNR, Viale Manzoni 30, 00185 Roma, Italy (carravetta@iasi.rm.cnr.it).

[‡]Dipartimento di Ingegneria Elettrica, Università dell'Aquila, 67100 Monteluco (L'Aquila), Italy (germani@ing.univaq.it).

[§]Department of Electrical Engineering, Kazakh National Technical University, 22 Satpaev Str., Almaty, Republic of Kazakhstan.

where we have denoted $\mathcal{F}_t^Y$ the $\sigma$-algebra generated by the observations $\{Y(s),\ 0 \leq s \leq t\}$.

As is well known, the optimal filter for system (1.1), (1.2) is an infinite-dimensional one. Nevertheless, from an application point of view, it becomes crucial to look for finite-dimensional approximations of the optimal filter.

In this paper we will derive, as an auxiliary result, the *optimal linear filtering* equations for a BLSS in the form of (1.1), (1.2) which will result in the finite-dimensional form (1.3), (1.4). We point out that in [3] the optimal linear filter is derived in the more general setting of linear stochastic equations driven by wide-sense Wiener (WSW) processes, resulting in a Kalman–Bucy scheme [1], [2]. Then, the optimal linear filter is defined for a scalar BLSS by representing the bilinear form as a WSW process. We will follow the same basic methodology in deriving the optimal linear filter for a *vector* BLSS.

Because of the infinite-dimensionality of the optimal filter for system (1.1), (1.2), it is of a great interest from an application point of view to search for finite-dimensional *suboptimal* filters showing a *better performance* with respect to the linear one.

This suboptimal approach has been recently developed for discrete-time systems in [9], [10], where a general *polynomial filter* of any arbitrarily fixed degree is defined for linear non-Gaussian systems [9] and bilinear systems [10]. The polynomial filter is able to produce, recursively, the optimal state-estimate in a class of polynomials of all the currently available measurements including the linear transformations. For this reason, in a non-Gaussian setting, it represents an improvement of the classical Kalman filtering. Indeed, many numerical simulations have shown that the improvement in performance may be very large especially when noise distributions are very far from Gaussianity.

In this paper we will propose this suboptimal approach for the filtering problem of continuous-time BLSSs. This will allow us to define a finite-dimensional filter in the form (1.3), (1.4), giving the optimal state-estimate in a suitably defined class of polynomial transformations of the measurements.

The program of the polynomial filtering methodology consists essentially in the following three steps.

(i) A class of polynomial estimators is defined.

(ii) The problem of finding the optimal filter for the BLSS in the above class of polynomial estimators is reduced to an optimal linear filtering problem for a suitable *augmented system*. The augmented system will result in a linear SDE with WSW diffusions. In particular, the state of the augmented system (augmented state) contains the original state, its Kronecker powers, and also Kronecker products with the observation process. The output of the augmented state (augmented observation) contains the original output process together with its Kronecker powers up to a fixed degree.

(iii) A Kalman–Bucy scheme is applied to the augmented system. This will give us the required polynomial filter.

The paper is organized as follows. Section 2 deals with point (i). In section 3, the overall setup of the problem is presented. Sections 4, 5, and 6 are concerned with some preliminary results. In particular, in section 4, a method for transforming a vector BLSS in a linear system with WSW diffusions is presented. In section 5 a vector Ito formula is defined by using the Kronecker formalism. In section 6, a general formula defining the stochastic differential of the Kronecker power of some process, solution of a bilinear SDE, is found. In section 7, point (ii) is treated. Finally, in section 8, the complete solution of the problem is presented, resulting in a system of equations which defines a polynomial filter (of an arbitrarily fixed degree) for a BLSS. In section 9, numerical simulations are presented for a linear and third degree polynomial filter applied to a second order BLSS. A comparision is made with respect to the extended Kalman filter, which shows an unstable behavior for the presented case. Two appendices are included in order to make the paper more readable.

**2. Suboptimal filtering.** This section is devoted to the definition of the class of estimators considered in this paper. First of all, let us recall some results of linear filtering [3].

Let $I$ be an interval (bounded or not) in the real line and consider a family $\{\xi_t,\ t \in I\}$ of $L^2$ random variables valued on some finite-dimensional euclidean space. For $t \in I$, let us define *the subspace* $\mathcal{L}_t(\xi) \subset L^2$ *linearly spanned by* $\{\xi_s,\ s \leq t\}$ as the $L^2$-closure of the set $\mathcal{L}'_t(\xi)$:

$$\mathcal{L}'_t(\xi) \triangleq \Big\{\lambda \in L^2 \ : \ \exists j \in \mathbf{N}, \exists t_1, \ldots, t_j \in I,\ t_1 \leq \cdots \leq t_j \leq t,$$

$$\exists \text{ matrices } M_{t_1}, \ldots, M_{t_j}, \exists \text{ a vector } b, \text{ such that } \lambda = \sum_{i=1}^{j} M_{t_i}\xi_{t_i} + b\Big\}.$$

Let $\Pi(\cdot/\mathcal{L}_t(\xi))$ denote the orthogonal projection operator onto $\mathcal{L}_t(\xi)$. Then, for any given $L^2$ random variable $\eta$ we can define the *optimal linear estimate of* $\eta$ *given* $\{\xi_s,\ s \leq t\}$ as $\Pi(\eta/\mathcal{L}_t(\xi))$. Now, suppose there exists an integer $\nu$ such that

$$E(\|\xi_t\|^{2\nu}) \leq +\infty \quad \forall\, t \in I.$$

Let us denote by $X^{[i]}$ the $i$th Kronecker power of a vector $X$. We can give the following definition.

Definition 2.1. *We call* $\nu$*th degree polynomial estimate of* $\eta$ *given* $\{\xi_s,\ s \leq t\}$ *the random variable* $\Pi(\eta/\mathcal{P}_t^{(\nu)}(\xi))$, *where*

$$\mathcal{P}_t^{(\nu)}(\xi) \triangleq \mathcal{L}_t(\xi^{(\nu)})$$

*and* $\xi^{(\nu)}$ *is the process*

$$\xi^{(\nu)} \triangleq \begin{bmatrix} \xi^{[\nu]} \\ \xi^{[\nu-1]} \\ \vdots \\ \xi \\ 1 \end{bmatrix}.$$

From Definition 2.1 we see that $\Pi(\eta/\mathcal{P}_t^{(\nu)})$ is the mean square optimal estimate of $\eta$ among all estimates, namely $\lambda$, that are either in the form

$$\lambda = \sum_{i,j=1}^{k} M_{i,j}\xi_{t_i}^{[j]} + b$$

for such a $k \in \mathbf{N}$, $t_1, \ldots, t_k \in I$, $t_1 \leq \cdots \leq t_k$ for such a vector $b$ and matrices $M_{i,j}$, $i, j = 1, \ldots, k$, or are mean square limits of these. $\Pi(\eta/\mathcal{P}_t^{(\nu)})$ includes the linear estimates and, moreover,

$$\mathcal{P}_t^{(\nu)}(\xi) \subset \mathcal{P}_t^{(\nu+1)}(\xi) \quad \forall\ \nu \geq 1,$$

so that, for the polynomial estimates $\hat{\eta}^{(\nu)} = \Pi(\eta/\mathcal{P}_t^{(\nu)}(\xi))$, $\hat{\eta}^{(\nu+1)} = \Pi(\eta/\mathcal{P}_t^{(\nu+1)}(\xi))$ one has

$$E(\|\eta - \hat{\eta}^{(\nu+1)}\|^2) \leq E(\|\eta - \hat{\eta}^{(\nu)}\|^2) \quad \forall\ \nu \geq 1.$$

That is, the estimation quality is not decreasing for increasing $\nu$.

Now, the aim of this paper can be expressed in a more precise manner as follows: for any given $\nu$ find a finite-dimensional filter in the form (1.3), (1.4) such that $\hat{X}(t)$ is the optimal $\nu$th degree polynomial estimate of the state of system (1.1), (1.2). Such a filter will be referred to in the following as a *$\nu$th degree polynomial filter*.

A crucial topic involved in the derivation of the polynomial filter is the linear estimation of stochastic processes generated by linear models driven by WSW processes, which we briefly describe below (see [3, Chap. 15], for a detailed discussion with proofs).

Let $\tilde{W}^{(i)}(t) \in \mathbf{R}^l$, $i = 1, \ldots, m$, be mutually uncorrelated WSW processes. Let us consider the linear stochastic system

$$\begin{aligned} dX(t) &= A(t)X(t)dt + \sum_{i=1}^{m} B_i(t) d\tilde{W}^{(i)}(t), \quad X(0) = \bar{X}, \\ dY(t) &= C(t)X(t)dt + \sum_{i=1}^{m} D_i(t) d\tilde{W}^{(i)}(t), \quad Y(0) = 0, \end{aligned} \tag{2.1}$$

where $t \in [0\ t_M]$, $X(t) \in \mathbf{R}^n$, $Y(t) \in \mathbf{R}^q$, $A(t), C(t), B_i(t), D_i(t)$, $i = 1, \ldots, m$, are suitably dimensioned matrices and $\bar{X}$ is a square integrable random vector. Model (2.1) can be interpreted as a continuous-time linear non-Gaussian system. We can consider the processes $X, Y$ evolving in suitable $L^2$ spaces of square integrable random vectors. Let us denote with $\hat{X}(t)$ the optimal linear estimate of $X(t)$, that is $\hat{X}(t) = \Pi(X(t)/\mathcal{L}_t(Y))$. Then the following system of equations can be easily derived from [3, Thm. 15.3]:

(2.2)

$$\begin{aligned} d\hat{X}(t) &= A(t)\hat{X}(t)dts \\ &\quad + \left( \sum_{i=1}^{m} B_i(t)D_i(t)^T + P(t)C(t)^T \right) R(t)^{-1}(dY(t) - C(t)\hat{X}(t)dt), \\ \frac{dP(t)}{dt} &= A(t)P(t) + P(t)A(t)^T + Q(t) \\ &\quad - \left( \sum_{i=1}^{m} B_i(t)D_i(t)^T + P(t)C(t)^T \right) R(t)^{-1} \left( \sum_{i=1}^{m} B_i(t)D_i(t)^T + P_t(t)C(t)^T \right)^T, \\ \hat{X}(0) &= E(\bar{X}), \quad P(0) = E\Big( (\bar{X} - E(\bar{X}))(\bar{X} - E(\bar{X}))^T \Big), \end{aligned}$$

where

$$R(t) \triangleq \sum_{i=1}^{m} D_i(t)D_i(t)^T; \quad Q(t) \triangleq \sum_{i=1}^{m} B_i(t)B_i(t)^T,$$

and $P(t)$ represents the filtering error covariance matrix. Note that in (2.2) the nonsingularity of the matrix function $R(t)$ over the time interval $[0\ t_M]$ is required.

As we will see in the next section, the BLSSs can be represented in the form (2.1). Then, (2.2) will allow us to obtain the optimal linear filter for a BLSS. This is a crucial point in the methodology here described. The way to derive the polynomial filter equations will consist indeed in reducing the original filtering problem to a linear one for a suitably defined BLSS.

**3. The system to be filtered.** Let $T = [0 \; t_{_M}]$, let $(\Omega, \mathcal{F}, P)$ be a probability triple, and let $\{\mathcal{F}_t\}$, $t \in T$, be a family of nondecreasing sub-$\sigma$-algebras of $\mathcal{F}$. Moreover let $(W(t), \mathcal{F}_t)$ be an $\mathbf{R}^p$-valued standard Wiener process and $\bar{X} \in \mathbf{R}^n$ an $\mathcal{F}_0$-measurable random variable, independent of $W$, such that

$$E(\|\bar{X}\|^{2\nu}) < +\infty$$

for some integer $\nu \geq 1$. For the random variable $\bar{X}$ we suppose the moments, namely $m_{\bar{X}}^{(i)}$,

$$m_{\bar{X}}^{(i)} \stackrel{\Delta}{=} E(\bar{X}^{[i]}), \quad i = 1, \dots, 2\nu, \tag{3.1}$$

are known. Let us consider the stochastic system

$$\begin{aligned} dX(t) &= A(t)X(t)dt + H(t)u(t)dt \\ &\quad + \sum_{k=1}^{p} \big(B_k X(t) + F_k\big) dW_k(t), \quad X(0) = \bar{X}, \end{aligned} \tag{3.2}$$

$$\begin{aligned} dY(t) &= C(t)X(t)dt \\ &\quad + \sum_{k=1}^{p} \big(D_k X(t) + G_k\big) dW_k(t), \quad Y(0) = 0, \end{aligned} \tag{3.3}$$

where $A(t) \in \mathbf{R}^{n\times n}$, $C(t) \in \mathbf{R}^{q\times n}$, $H(t) \in \mathbf{R}^{n\times m}$, $B_k \in \mathbf{R}^{n\times n}$, $F_k \in \mathbf{R}^n$, $D_k \in \mathbf{R}^{q\times n}$, $G_k \in \mathbf{R}^q$, for $k = 1, \dots, p$, $W_k(t)$ denotes the $k$th component of the standard Wiener process $W(t) \in \mathbf{R}^p$, and $u(t) \in \mathbf{R}^m$ is a deterministic input. Equation (3.2) is endowed with the initial condition $X(0) = \bar{X}$. In the following, we shall denote with $I_\alpha$, $\alpha = 0, 1, \dots$, the $\alpha \times \alpha$ identity matrix; we assume $I_0 = 1$. We make the following assumption on system (3.2), (3.3).

*Assumption* 3.1. There exists a $\bar{k}$, $1 \leq \bar{k} \leq p$, such that the matrix $D_{\bar{k}} D_{\bar{k}}^T$ is nonsingular.

*Remark* 3.2. Assumption 3.1 implies that we can assume, without loss of generality, that there exists a $\bar{k}$, $1 \leq \bar{k} \leq p$, such that

$$D_{\bar{k}} = [I_q \; 0]. \tag{3.4}$$

Indeed, let $\bar{k}$ be such that $D_{\bar{k}} D_{\bar{k}}^T$ is nonsingular, and define the matrix $T \in \mathbf{R}^{n\times n}$ as

$$T = \begin{bmatrix} D_{\bar{k}} \\ R \end{bmatrix},$$

where $R \in \mathbf{R}^{(n-q)\times n}$ is chosen such that the whole $T$ results in a nonsingular matrix. It is easy to verify that $D_{\bar{k}} T^{-1} = [I_q \; 0]$. Hence we can always modify system (3.2), (3.3) by using $T$ as a matrix performing a change of coordinates in the state space, and we can ensure that the representation (3.4) holds for at least one $\bar{k} \in \{1, \dots, p\}$.

The problem we are faced with consists in finding a finite-dimensional filter in the form of (1.3), (1.4), such that

$$\widehat{X}(t) = \Pi\Big(X(t)/\mathcal{P}_t^{(\nu)}(Y)\Big), \tag{3.5}$$

where the space $\mathcal{P}_t^{(\nu)}(Y)$ is given by Definition 2.1.

As above mentioned (see point (ii) in the introduction), we will prove that there exists an augmented linear system for which the optimal linear filtering problem is equivalent to the original polynomial filtering problem for system (3.2), (3.3). To this purpose, in the next two sections we state some preliminary results.

**4. Optimal linear filtering for BLSSs.** Before treating the more general polynomial case, in this section we limit ourselves in considering the optimal linear filtering problem for the BLSS (3.2), (3.3). The reason for considering this particular case in advance is twofold. First of all, as we will see later, the polynomial case reduces to the linear one once a suitable augmented system has been constructed. Moreover, the optimal (finite-dimensional) *linear* filtering problem for a BLSS is interesting by itself, in that it was up to now unsolved in the general case [3]. In this section, we give a solution of this problem, in that we will prove the existence of a *linear* stochastic system with WSW diffusions, which is equivalent to the original BLSS (3.2), (3.3). Indeed, a version of the classical Kalman–Bucy theory [3] solves the optimal linear filtering problem in this case.

Let $M \in \mathbf{R}^{\alpha\times\alpha}$ be a symmetric positive semidefinite matrix, such that $\operatorname{rank}(M) = \rho \le \alpha$. As is well known, there exists a full rank matrix $N \in \mathbf{R}^{\alpha\times\rho}$ such that $NN^T = M$. We will use the notation

$$M^{(\frac{1}{2})} \stackrel{\Delta}{=} N,$$

that is, a "rectangular square root" of the matrix $M$. Note that, by definition, the matrix ${M^{(1/2)}}^T M^{(1/2)}$ is nonsingular.

Let $\xi$ be a random vector; in the following we will use the notation $\operatorname{cov}(\xi,\xi) = E\big((\xi - E(\xi))(\xi - E(\xi))^T\big)$. Let us denote $m_{_X}(t) = E(X(t))$, $\Psi_X(t) = \operatorname{cov}(X(t), X(t))$, where $X$ is the state process of system (3.2), (3.3). Moreover, let us denote $\bar m_{_X} = E(\bar X)$ and $\bar\Psi_X = \operatorname{cov}(\bar X, \bar X)$, where $\bar X$ is the initial state vector of (3.2).

THEOREM 4.1. *Let us consider the system* (3.2), (3.3). *Suppose that the matrix* $\Psi_X(t)$ *is nonsingular for any* $t \in T$. *Let us consider, for* $k = 1, \dots, p$, *the integers* $\rho_k \le n$, $\sigma_k \le q$ *such that*

$$\left.\begin{aligned} \rho_k &\stackrel{\Delta}{=} \operatorname{rank}\Big\{B_k \cdot \Psi_X(t) \cdot {B_k}^T\Big\} \\ \sigma_k &\stackrel{\Delta}{=} \operatorname{rank}\Big\{D_k \cdot \Psi_X(t) \cdot {D_k}^T\Big\} \end{aligned}\right. \qquad \forall\ t \in T. \tag{4.1}$$

*Then there exists the representation*

$$dX(t) = A(t)X(t)dt + H(t)u(t) + \sum_{k=1}^{2p} \tilde B_k(t) d\tilde W_{k,1}(t), \qquad X(0) = \bar X, \tag{4.2}$$

$$dY(t) = C(t)X(t)dt + \sum_{k=1}^{2p} \tilde D_k(t) d\tilde W_{k,2}(t), \qquad Y(0) = 0, \tag{4.3}$$

*where, for $k = 1, \dots, p$: $\tilde{B}_k(t) \in \mathbf{R}^{n\times\rho_k}$ and $\tilde{D}_k(t) \in \mathbf{R}^{n\times\sigma_k}$ are given by*

$$\tilde{B}_k(t) \triangleq \Big(B_k \cdot \Psi_X(t) \cdot B_k^T\Big)^{(\frac{1}{2})}, \tag{4.4}$$

$$\tilde{D}_k(t) \triangleq \Big(D_k \cdot \Psi_X(t) \cdot {D_k}^T\Big)^{(\frac{1}{2})} \tag{4.5}$$

*for $k = p+1, \dots, 2p$:*

$$\tilde{B}_k(t) \triangleq B_{k-p}E\big(X(t)\big) + F_{k-p}, \tag{4.6}$$

$$\tilde{D}_k(t) \triangleq D_{k-p}E\big(X(t)\big) + G_{k-p}. \tag{4.7}$$

*For $i = 1, 2$, the set $\{\tilde{W}_{k,i},\ k = 1, \dots, 2p\}$ is a set of $2p$ mutually uncorrelated standard WSW processes. In particular, for $k = 1, \dots, p$, $\tilde{W}_{k,1}(t) \in \mathbf{R}^{\rho_k}$, $\tilde{W}_{k,2}(t) \in \mathbf{R}^{\sigma_k}$; for $k = p+1, \dots, 2p$:*

$$\tilde{W}_{k,1}(t) = \tilde{W}_{k,2}(t) = W_{k-p}(t). \tag{4.8}$$

*Proof.* For $k = 1, \dots, p$, let us define the processes $\tilde{W}_{k,1}, \tilde{W}_{k,2}$ as

$$\tilde{W}_{k,1}(t) = \int_0^t \big(\tilde{B}_k(\tau)^T\tilde{B}_k(\tau)\big)^{-1}\tilde{B}_k(\tau)^T B_k\big(X(\tau) - m_{X}(\tau)\big)dW_k(\tau), \tag{4.9}$$

$$\tilde{W}_{k,2}(t) = \int_0^t \big(\tilde{D}_k(\tau)^T\tilde{D}_k(\tau)\big)^{-1}\tilde{D}_k(\tau)^T D_k\big(X(\tau) - m_{X}(\tau)\big)dW_k(\tau), \tag{4.10}$$

where $\tilde{B}_k, \tilde{D}_k$ are given by (4.4), (4.5). Let us show that $\tilde{W}_{k,i}$, $i = 1, 2$, are standard WSW processes. As a matter of fact, using well-known properties of the Ito integral and (4.4), it results, for $s < t$:

$$\begin{aligned} &E\big(\tilde{W}_{k,1}(t)\tilde{W}_{k,1}(s)^T\big) \\ &\quad = \int_0^s \big(\tilde{B}_k(\tau)^T\tilde{B}_k(\tau)\big)^{-1}\tilde{B}_k(\tau)^T\Big(B_k\Psi_X(\tau)B_k^T\Big)\tilde{B}_k(\tau)\big(\tilde{B}_k(\tau)^T\tilde{B}_k(\tau)\big)^{-1}\,d\tau \\ &\quad = \int_0^s \big(\tilde{B}_k(\tau)^T\tilde{B}_k(\tau)\big)^{-1}\tilde{B}_k(\tau)^T\big(\tilde{B}_k(\tau)\tilde{B}_k(\tau)^T\big)\cdot\tilde{B}_k(\tau)\big(\tilde{B}_k(\tau)^T\tilde{B}_k(\tau)\big)^{-1}\,d\tau \\ &\quad = I_{\rho_k}\cdot s. \end{aligned}$$

Similarly, taking again an $s < t$, it can be proved that

$$E\big(\tilde{W}_{k,2}(t)\tilde{W}_{k,2}(s)^T\big) = I_{\sigma_k}\cdot s,$$

and hence, since the Wiener's process components $W_1, \dots, W_p$, are mutually independent, we have that, for $i = 1, 2$, $\{\tilde{W}_{k,i},\ k = 1, \dots, p\}$ is a family of mutually independent (vector) WSW processes with identity covariance.

Now let us show that, for $k = 1, \dots, p$ (almost surely),

$$\tilde{B}_k(t)d\tilde{W}_{k,1}(t) = B_k\big(X(t) - m_{X}(t)\big)dW_k(t), \tag{4.11}$$

$$\tilde{D}_k(t)d\tilde{W}_{k,2}(t) = D_k\big(X(t) - m_{X}(t)\big)dW_k(t). \tag{4.12}$$

From the hypotheses the symmetric positive-definite matrix $\Psi(t)^{1/2}$ is well defined. Hence, for any $y(t) \in \mathbf{R}^n$ we can define $\bar{y}(t) \in \mathbf{R}^n$ such that $y(t) = \Psi_x(t)\bar{y}(t)$. Next, let us consider the decomposition $\bar{y}(t) = \bar{y}_1(t) + \bar{y}_2(t)$, where

$$\bar{y}_1(t) \in \mathcal{R}\big(\Psi_X(t)^{1/2}B_k^T\big), \quad \bar{y}_2(t) \in \Big\{\mathcal{R}\big(\Psi_X(t)^{1/2}B_k^T\big)\Big\}^{\perp} = \mathcal{N}\big(B_k\Psi_X(t)^{1/2}\big), \tag{4.13}$$

where $\mathcal{N}(M)$, $\mathcal{R}(M)$ denote the null-space and the range, respectively, of a matrix $M$. Using (4.13) and choosing a $\bar{z}(t)$ such that $\bar{y}_1(t) = \Psi_X(t)^{1/2} B_k \bar{z}(t)$, we have

$$B_k y(t) = B_k \Psi_X(t)^{\frac{1}{2}} \bar{y}(t) = B_k \Psi_X(t)^{\frac{1}{2}} \bar{y}_1(t) = B_k \Psi_X(t) B_k^T \bar{z}(t) = \tilde{B}_k(t) \tilde{B}_k(t)^T \bar{z}(t),$$

where the definition of $\tilde{B}_k(t)$, given by (4.4) has been used. It follows that for any $y(t) \in \mathbf{R}^n$ there exists a $z(t) \in \mathbf{R}^{\rho_k}$ (indeed $z(t) = \tilde{B}_k(t)^T \bar{z}(t)$) such that

$$B_k y(t) = \tilde{B}_k(t) z(t) \quad \forall\, t \in T. \tag{4.14}$$

Then, for any $y(t)$ we have

$$\begin{aligned} \tilde{B}_k(t)\Big(\tilde{B}_k(t)^T \tilde{B}_k(t)\Big)^{-1} \tilde{B}_k(t)^T B_k y(t) &= \tilde{B}_k(t)\Big(\tilde{B}_k(t)^T \tilde{B}_k(t)\Big)^{-1} \tilde{B}_k(t)^T \tilde{B}_k(t) z(t) \\ &= \tilde{B}_k(t) z(t) = B_k y(t), \end{aligned}$$

from which, using the definition of $\tilde{W}_{k,1}$ given by (4.9), equality (4.11) follows. A similar argument can be used to prove (4.12).

Finally, by adding and subtracting the state-expectation $m_X(t)$, in the bilinear forms of (3.2), (3.3) and taking into account (4.11), (4.12), we obtain the representation (4.2), (4.3). The thesis follows as soon as it is proven that, for $i = 1, 2$, $\tilde{W}_{k',i}(t)$ ($p + 1 \leq k' \leq 2p$) is uncorrelated with $\tilde{W}_{k'',i}(t)$ ($1 \leq k'' \leq p$). As a matter of fact, from (4.8), for $p + 1 \leq k' \leq 2p$, $k'' \neq k' - p$,

$$E\big(\tilde{W}_{k'',1}(t) \tilde{W}_{k',1}(t)^T\big) = E\big(\tilde{W}_{k'',1}(t) W_{k'-p}(t)^T\big) = 0,$$

and, for $k'' = k' - p$,

$$\begin{aligned} &E\big(\tilde{W}_{k'',1}(t) \tilde{W}_{k',1}(t)^T\big) = E\big(\tilde{W}_{k'',1}(t) W_{k''}(t)^T\big) \\ &\quad = E\left( \int_0^t \big(\tilde{B}_{k''}(\tau)^T \tilde{B}_{k''}(\tau)\big)^{-1} \tilde{B}_{k''}(\tau)^T B_{k''} \big(X(\tau) - m_X(\tau)\big) dW_{k''}(\tau) \cdot \int_0^t dW_{k''}(\tau) \right) \\ &\quad = \int_0^t E\Big( \big(\tilde{B}_{k''}(\tau)^T \tilde{B}_{k''}(\tau)\big)^{-1} \tilde{B}_{k''}(\tau)^T B_{k''} \big(X(\tau) - m_X(\tau)\big)\Big) d\tau = 0. \end{aligned}$$

In the same way, it is possible to show that $E\big(\tilde{W}_{k'',2}(t) \tilde{W}_{k',2}(t)^T\big) = 0$ for $p + 1 \leq k' \leq 2p$. ◻

In the following theorem a sufficient condition will be given which guarantees the nonsingularity of $\Psi_X(t)$. Let us consider a time-invariant version of the BLSS given by (3.2), (3.3):

$$dX(t) = AX(t)dt + Hu(t)dt + \sum_{k=1}^{p} (B_k X(t) + F_k) dW_k(t), \quad X(t_0) = \bar{X}, \tag{4.15}$$

$$dY(t) = CX(t)dt + \sum_{k=1}^{p} (D_k X(t) + G_k) dW_k(t), \qquad Y(t_0) = 0, \tag{4.16}$$

where $t_0 \in \mathbf{R}$ is any "initial time." We suppose that system (4.15), (4.16) is well defined over the time interval $[t_0\ \infty)$.

THEOREM 4.2. *Let the matrix $\Psi_X(t_0)$ be nonsingular (or the pair $(A, F_k)$ of the state equation* (4.15) *be controllable for at least one $k = 1, \ldots, p$); then the state covariance matrix $\Psi_X(t)$ is nonsingular for any $t \geq t_0$, ($t > 0$).*

*Proof.* Let us denote $\tilde{X}(t) = X(t) - m_X(t)$. Taking the expectations of (4.15), we have

$$dm_X(t) = Am_X(t)dt + Hu(t)dt, \qquad m_X(0) = \bar{m}_X.$$

Subtracting this from (4.15) results in

$$d\tilde{X}(t) = A\tilde{X}(t)dt + \sum_{k=1}^{p} B_k\tilde{X}(t)dW_k(t) + \sum_{k=1}^{p}(B_k m_X(t) + F_k)dW_k(t), \quad \tilde{X}(t_0) = \bar{X} - \bar{m}_X$$

or

$$\begin{aligned}\tilde{X}(t) = e^{A(t-t_0)}\tilde{X}(t_0) + \sum_{k=1}^{p}\int_{t_0}^{t} e^{A(t-\tau)}B_k\tilde{X}(\tau)dW_k(\tau) \\ + \int_{t_0}^{t} e^{A(t-\tau)}(B_k m_X(\tau) + F_k)dW_k(\tau).\end{aligned} \tag{4.17}$$

From (4.17) the following equation is easily recognized to hold for $\Psi_X(t)$:

$$\begin{aligned}\Psi_X(t) = e^{A(t-t_0)}\Psi_X(t_0)e^{A^T(t-t_0)} + \sum_{k=1}^{p}\int_{t_0}^{t} e^{A(t-\tau)}B_k\Psi_X(\tau)B_k^T e^{A^T(t-\tau)}d\tau \\ + \sum_{k=1}^{p}\int_{t_0}^{t} e^{A(t-\tau)}(B_k m_X(\tau) + F_k)(B_k m_X(\tau) + F_k)^T e^{A^T(t-\tau)}d\tau.\end{aligned} \tag{4.18}$$

The thesis follows by noting that the three terms in the right-hand side of (4.18) are at least symmetric nonnegative definite and, in particular, the nonsingularity of $\phi_X(t_0)$ implies the positive definiteness of the first term, whereas the hypothesis of controllability of $(A, F_k)$ for some $k$ implies the positive definiteness of the term $\int e^{A(t-\tau)}F_kF_k^Te^{A^T(t-\tau)}d\tau$. ∎

*Remark* 4.3. Note that, when theorem 4.2 holds with $t_0 < 0$, it results that, for any finite time-interval $T \subset [t_0 \;\; +\infty)$, the state-covariance has the property $\Psi_X(t) > \alpha \cdot I \;\forall\, t \in T$ ($I$ denotes the identity) for some real number $\alpha > 0$, (it is *unifomly nonsingular in* $T$).

Now, we can state the following theorem, which defines the optimal linear filter for a BLSS.

THEOREM 4.4. *Let the time-invariant BLSS as defined in* (4.15), (4.16) *be given. Let the hypotheses of Theorem* 4.2 *be satisfied. Moreover, let us suppose that*

(H1) $\mathrm{rank}(D_k) = q$ *or* $\mathrm{rank}(G_k) = q$ *for some* $k$.

*Then, with reference to the notations of section* 2, *the optimal linear estimate of the state process* $X$, *namely* $\hat{X}$, *and the error covariance*

$$P(t) = E\big((X(t) - \hat{X}(t))(X(t) - \hat{X}(t))^T\big)$$

*satisfy the following system of equations:*

$$\frac{dm_X(t)}{dt} = Am_X(t) + Hu(t), \qquad m(0) = \bar{m}, \tag{4.19}$$

$$\frac{d\Psi_X(t)}{dt} = A\Psi_X(t) + \Psi_X(t)A^T + \sum_{k=1}^{p} B_k\Psi_X(t)B_k^T$$

$$+\sum_{k=1}^{p}(B_k m_X(t)+F_k)(B_k m_X(t)+F_k)^T, \qquad \Psi_X(0)=\bar{\Psi}_X, \tag{4.20}$$

$$\tilde{B}_k(t)=\begin{cases}\left(B_k\cdot\Psi_X(t)\cdot B_k^T\right)^{\left(\frac{1}{2}\right)}, & 1\le k\le p,\\ B_{k-p}m_X(t)+F_{k-p}, & p+1\le k\le 2p,\end{cases} \tag{4.21}$$

$$\tilde{D}_k(t)=\begin{cases}\left(D_k\cdot\Psi_X(t)\cdot D_k^T\right)^{\left(\frac{1}{2}\right)}, & 1\le k\le p,\\ D_{k-p}m_X(t)+G_{k-p}, & p+1\le k\le 2p,\end{cases} \tag{4.22}$$

$$R(t)=\sum_{i=1}^{2p}\tilde{D}_i(t)\tilde{D}_i(t)^T, \tag{4.23}$$

$$\frac{dP(t)}{dt}=AP(t)+P(t)A^T+R(t),$$

$$-\left(\sum_{i=1}^{2p}\tilde{B}_i(t)\tilde{D}_i(t)^T+P(t)C^T\right)R(t)^{-1}\left(\sum_{i=1}^{2p}\tilde{B}_i(t)\tilde{D}_i(t)^T+P(t)C^T\right)^T, \tag{4.24}$$

$$P(0)=\bar{\Psi}_X, \tag{4.25}$$

$$d\hat{X}(t)=A\hat{X}(t)dt+\left(\sum_{i=1}^{2p+1}\tilde{B}_i(t)\tilde{D}_i(t)^T+P(t)C^T\right)R(t)^{-1}(dY(t)-C\hat{X}(t)dt), \tag{4.26}$$

$$\hat{X}(0)=\bar{m}. \tag{4.27}$$

*Proof.* (4.19) readily derives by taking the expectations of both sides of (4.15). Moreover, (4.20) is easily obtained by differentiating (4.18). From Theorem 4.2 and Remark 4.3, $\Psi_X(t)$ is uniformly nonsingular in $T$. Then, we can apply Theorem 4.1 in order to put system (4.15), (4.16) in the form of a linear stochastic system with suitable WSW state and output diffusions, deriving from (4.2), (4.3). Note that such an equivalent system is a time-varying one even if it is derived from the time-invariant BLSS (4.15), (4.16). Now from (4.22), (4.23) it results that

$$R(t)\triangleq\sum_{k=1}^{p}D_k\Psi_X(t)D_k^T+\sum_{k=1}^{p}(D_k m_X(t)+G_k)(D_k m_X(t)+G_k)^T,$$

which is uniformly nonsingular in $T$, by the hypothesis (H1) (and possibly by the uniform nonsingularity of $\Psi_X(t)$). The thesis easily derives from an application of [3, Thm. 15.3] to the representation (4.2), (4.3). □

*Remark* 4.5. In the general case, when the BLSS is time-varying the uniform nonsingularity of $\Psi_X(t)$ cannot be guaranteed. Nevertheless, in all the cases of a nonsingular $\Psi_X(t)$, the equations of the optimal linear filter can be still derived using the representation given by Theorem 4.1. The resulting system of equations is formally similar to (4.19)–(4.27), but the constant parameters are replaced with the corresponding time-varying ones.

**5. The vector Ito formula in the Kronecker formalism.** In this section, by using a formalism derived from the Kronecker algebra, we present a new version of the Ito formula which has, with respect to the classical formulation, the advantage of being much more compact and will allow us to calculate, for a given stochastic process $\phi$, the stochastic differential of the process $\phi^{[h]}$, where $[h]$ is any integer Kronecker power.

Let $x \in \mathbf{R}^n$ and $F$ be any $C^2$ function in $\mathbf{R}^{m\times p}$; we introduce the matrix $(d/dx)\otimes F(x)$, having dimensions $m \times (n \cdot p)$, defined as

$$\frac{d}{dx} \otimes F(x) \triangleq \Big[\frac{\partial F(x)}{\partial x_1} \ \cdots \ \frac{\partial F(x)}{\partial x_n}\Big], \tag{5.1}$$

where the operator $d/dx$ is given by

$$\frac{d}{dx} \triangleq \Big[\frac{\partial}{\partial x_1} \ \cdots \ \frac{\partial}{\partial x_n}\Big]. \tag{5.2}$$

Note that in (5.1) the rules defining the Kroneker product between matrices (see Definition A.1) are formally satisfied, provided that the "multiplication" between the differential operator $\partial/\partial x_i$ and a matrix function $F(x)$ is conventionally defined as

$$\frac{\partial}{\partial x_i} \cdot F(x) = \frac{\partial F(x)}{\partial x_i},$$

where the right-hand side has the usual meaning. Similarly, we can define the operator:

$$\frac{d}{dx} \otimes \frac{d}{dx} \triangleq \Big[\frac{\partial^2}{\partial x_1^2} \ \frac{\partial^2}{\partial x_1 \partial x_2} \ \cdots \ \frac{\partial^2}{\partial x_n^2}\Big].$$

Also in this case the composition rule of the Kronecker product is satisfied, but the "multiplication" between the differential operators $\partial/\partial x_i$ and $\partial/\partial x_i$ had to be interpreted as resulting in the differential operator $\partial^2/\partial x_i \partial x_j$. In general, we will adopt the convention: the multiplication between a differential operator and a function $F$ results in a function (the derivative of $F$), whereas the multiplication between two differential operators results in a differential operator (the second order differential operator). Obviously, this convention could be generalized in order to give a precise meaning to the quantity

$$\frac{d^{[h]}}{dx^{[h]}} \otimes F(x)$$

for any integer $h \geq 0$. However, in this paper we are concerned at most with second-order derivatives.

It is easy to recognize that for any matrix, namely $M$, and for any pair of differentiable matrix functions having suitable dimensions, namely $V(x)$ and $W(x)$, it results that

$$\frac{d}{dx} \otimes (V(x) \otimes W(x)) = \Big(\frac{d}{dx} \otimes V(x)\Big) \otimes W(x) + V(x) \otimes \Big(\frac{d}{dx} \otimes W(x)\Big), \tag{5.3}$$

$$\frac{d}{dx} \otimes (MW(x)) = M\Big(\frac{d}{dx} \otimes W(x)\Big). \tag{5.4}$$

Moreover, the following "associative" property holds:

$$\frac{d}{dx} \otimes \frac{d}{dx} \otimes F(x) = \Big(\frac{d}{dx} \otimes \frac{d}{dx}\Big) \otimes F(x) = \frac{d}{dx} \otimes \Big(\frac{d}{dx} \otimes F(x)\Big).$$

Using the above notation, we can prove the following lemma, which will be very useful in the following sections.

LEMMA 5.1. *For any integer $h \geq 1$ and $x \in \mathbf{R}^n$, it results that*

$$\frac{d}{dx} \otimes x^{[h]} = U_n^h (I_n \otimes x^{[h-1]}), \tag{5.5}$$

*and for any $h > 1$,*

$$\frac{d}{dx} \otimes \frac{d}{dx} \otimes x^{[h]} = O_n^h (I_{n^2} \otimes x^{[h-2]}), \tag{5.6}$$

*where the matrices $C_{u,v}^T$, $u, v \in \mathbf{N}$, are the commutation matrices defined by Theorem* A.3 *and*

$$U_n^h \triangleq \left( \sum_{\tau=0}^{h-1} C_{n,n^{h-1-\tau}}^T \otimes I_{n^\tau} \right), \quad O_n^h \triangleq \sum_{\tau=0}^{h-1} \sum_{s=0}^{h-2} (C_{n,n^{h-1-\tau}}^T \otimes I_{n^\tau})(I_n \otimes C_{n,n^{h-2-s}}^T \otimes I_n).$$

*Proof.* According to the definition of the differential operator (5.1) and using (5.3), we have

$$\begin{aligned} Q^h \triangleq \frac{d}{dx} \otimes x^{[h]} &= \frac{d}{dx} \otimes \left( x \otimes x^{[h-1]} \right) = I_n \otimes x^{[h-1]} + x \otimes \left( \frac{d}{dx} \otimes x^{[h-1]} \right) \\ &= I_n \otimes x^{[h-1]} + x \otimes Q^{(h-1)}, \end{aligned} \tag{5.7}$$

from which, using Theorem A.3, we obtain

$$\frac{d}{dx} \otimes x^{[h]} = \sum_{\tau=0}^{h-1} x^{[h-1-\tau]} \otimes I_n \otimes x^{[\tau]} = \sum_{\tau=0}^{h-1} C_{n,n^{h-1-\tau}}^T \left( I_n \otimes x^{[h-1-\tau]} \right) \otimes x^{[\tau]},$$

from which (5.5) follows, taking into account the property (A.3c).

Similarly, by exploiting (5.3), (5.5), and (A.3c), it results that

$$\begin{aligned} &\frac{d}{dx} \otimes \frac{d}{dx} \otimes x^{[h]} \\ &\quad = \frac{d}{dx} \otimes \left( \left( \sum_{\tau=0}^{h-1} C_{n,n^{h-1-\tau}}^T \otimes I_{n^\tau} \right) \left( I_n \otimes x^{[h-1]} \right) \right) \\ &\quad = \sum_{\tau=0}^{h-1} \left( C_{n,n^{h-1-\tau}}^T \otimes I_{n^\tau} \right) \left( \frac{d}{dx} \otimes \left( I_n \otimes x^{[h-1]} \right) \right) \\ &\quad = \sum_{\tau=0}^{h-1} \left( C_{n,n^{h-1-\tau}}^T \otimes I_{n^\tau} \right) \left( I_n \otimes \left( \frac{d}{dx} \otimes x^{[h-1]} \right) \right) \\ &\quad = \sum_{\tau=0}^{h-1} \left( C_{n,n^{h-1-\tau}}^T \otimes I_{n^\tau} \right) \left( I_n \otimes \left( \left( \sum_{s=0}^{h-2} C_{n,n^{h-2-s}}^T \otimes I_{n^s} \right) \left( I_n \otimes x^{[h-2]} \right) \right) \right) \\ &\quad = \sum_{\tau=0}^{h-1} \sum_{s=0}^{h-2} \left( C_{n,n^{h-1-\tau}}^T \otimes I_{n^\tau} \right) \left( I_n \otimes \left( \left( C_{n,n^{h-2-s}}^T \otimes I_{n^s} \right) \left( I_n \otimes x^{[h-2]} \right) \right) \right) \\ &\quad = \sum_{\tau=0}^{h-1} \sum_{s=0}^{h-2} \left( C_{n,n^{h-1-\tau}}^T \otimes I_{n^\tau} \right) \left( \left( I_n \otimes \left( C_{n,n^{h-2-s}}^T \otimes I_{n^s} \right) \right) \left( I_n \otimes \left( I_n \otimes x^{[h-2]} \right) \right) \right), \end{aligned}$$

so that the proof is completed. □

Now, we are able to rewrite the vector valued version of the Ito formula in the Kronecker formalism.

THEOREM 5.2. *Let $(X_t, \mathcal{F}_t)$ be a vector-continuous semimartingale in $\mathbf{R}^n$ described by the Ito stochastic differential*

$$dX_t = d\beta_t + dM_t, \tag{5.8}$$

*where $(\beta_t, \mathcal{F}_t)$ is an almost surely continuous bounded variation process and $(M_t, \mathcal{F}_t)$ is a square integrable martingale. Let*

$$F : \mathbf{R}^n \to \mathbf{R}^p$$

*be a continuous function endowed with the first and second derivatives. Then the process $Z_t = F(X_t)$ is a square integrable semimartingale, whose differential is given by*

$$dZ_t = \left(\frac{d}{dx} \otimes F(x)\right)_{x=X_t} dX_t + \frac{1}{2}\left(\frac{d}{dx} \otimes \frac{d}{dx} \otimes F(x)\right)_{x=X_t} (dM_t)^{[2]}, \tag{5.9}$$

*with $(dM_t)^{[2]}$ denoting the associate quadratic variation process whose arguments are*

$$(dM_t)^{[2]} = \begin{bmatrix} d < M_1, M_1 >_t \\ d < M_1, M_2 >_t \\ \vdots \\ d < M_n, M_n >_t \end{bmatrix}, \tag{5.10}$$

*with obvious meaning of symbols* [11, 12].

*Proof.* Formula (5.10) can be directly verified by using the Ito formula in the scalar case (see for instance [11, Thm. 4.2.1]) and by taking into account the definition of the differential operator $d/dx$. □

**6. Stochastic differential for the Kronecker power of a BLSS solution.** Using the Ito formula, in the version given by Theorem 5.2, we can now prove the following theorem, which defines the stochastic differential for the power process of the solution of a bilinear SDE. This will be the fundamental tool in the derivation of the augmented system.

THEOREM 6.1. *Let $\phi(t) \in \mathbf{R}^d$ be the process defined by the following SDE:*

$$d\phi(t) = (\Gamma(t)\phi(t) + \gamma(t))dt + \sum_{k=1}^{p} (\Theta_k \phi(t) + \chi_k) dW_k(t), \tag{6.1}$$

*where $\Gamma(t), \Theta_k \in \mathbf{R}^{d\times d}$, $\gamma(t), \chi_k \in \mathbf{R}^d$. Then, defining*

$$\Phi_2 \triangleq \sum_{k=1}^{p} \Theta_k^{[2]}, \quad \Phi_1 \triangleq \sum_{k=1}^{p} (\Theta_k \otimes \chi_k + \chi_k \otimes \Theta_k), \quad \Phi_0 \triangleq \sum_{k=1}^{p} \chi_k^{[2]}, \tag{6.2}$$

*it results for $i \geq 2$ that*

$$\begin{aligned} d\phi^{[i]}(t) = {} & \left(\mathcal{M}_i^0(t)\phi^{[i]}(t) + \mathcal{M}_i^1(t)\phi^{[i-1]}(t) + \mathcal{M}_i^2 \phi^{[i-2]}(t)\right) dt \\ & + \sum_{k=1}^{p} \left(\mathcal{G}_{k,i}^0 \phi^{[i]}(t) + \mathcal{G}_{k,i}^1 \phi^{[i-1]}(t)\right) dW_k(t), \end{aligned}$$

*where*

$$
\begin{aligned}
\mathcal{M}_i^0(t) &= U_d^i(\Gamma(t)\otimes I_{d^{i-1}}) + \frac{1}{2}O_d^i(\Phi_2\otimes I_{d^{i-2}}),\\
\mathcal{M}_i^1(t) &= U_d^i(\gamma(t)\otimes I_{d^{i-1}}) + \frac{1}{2}O_d^i(\Phi_1\otimes I_{d^{i-2}}),\\
\mathcal{M}_i^2 &= \frac{1}{2}O_d^i\big(\Phi_0\otimes I_{d^{i-2}}\big),\\
\mathcal{G}_{k,i}^0 &= U_d^i(\Theta_k\otimes I_{d^{i-1}}),\\
\mathcal{G}_{k,i}^1 &= U_d^i(\chi_k\otimes I_{d^{i-1}}).
\end{aligned}
$$

*Proof.* By using property (A.3c) the following formula is easily recognized to hold for any $k = 0, 1, \ldots,$ $j = 1, 2, \ldots,$ $\psi \in \mathbf{R}^\sigma$, $M \in \mathbf{R}^{r\times\sigma^k}$:

$$
\left(I_r \otimes \psi^{[j]}\right) M\psi^{[k]} = (M\otimes I_{\sigma^j})\psi^{[k+j]}. \tag{6.3}
$$

Let us apply Theorem 5.2 for $X = \phi$, $F(\phi) = \phi^{[i]}$, $d\beta = (\Gamma\phi + \gamma)dt$, and $dM = d\Lambda$, where $\Lambda$ is the martingale:

$$
\Lambda(t) \triangleq \int_0^t \sum_{k=1}^{p} \big(\Theta_k(\tau)\phi(\tau) + \chi_k(\tau)\big)dW_k(\tau).
$$

Using formulas (5.5), (5.6), it results that (understanding time dependencies)

$$
d\phi^{[i]} = U_d^i\left(I_d\otimes\phi^{[i-1]}\right)(\Gamma\phi dt + \gamma dt + d\Lambda) + \frac{1}{2}O_d^i\left(I_{d^2}\otimes\phi^{[i-2]}\right)(d\Lambda)^{[2]}. \tag{6.4}
$$

By exploiting the definition (5.10) it results that

$$
(d\Lambda)^{[2]} = \left(\Phi^2\phi_{[2]} + \Phi_1\phi + \Phi_0\right)dt, \tag{6.5}
$$

where $\Phi_2$, $\Phi_1$, and $\Phi_0$ are given by (6.2). By substituting (6.5) in (6.4) and using formula (6.3), the thesis follows. $\square$

**7. The augmented system.** Let us return to consider the BLSS (3.2), (3.3). In this section, by means of a repeated application of Theorem 6.1, we will show that the process $(X, Y)$ and its powers up to a certain degree represent a solution of a suitably defined bilinear SDE. The latter will be next transformed into a linear system with WSW diffusions, generating the powers of the observation $Y$ up to the required degree (the augmented system).

Let $x \in \mathbf{R}^d$ and $h$ be a positive integer. We recall that the following relations hold, linking together the *reduced $h$th Kroneker power of* $x$ [3], [15], namely $x_{[h]}$ and the (ordinary) $h$th Kronecker power $x^{[h]}$:

$$
x^{[h]} = T_d^h x_{[h]}, \quad x_{[h]} = \tilde{T}_d^h x^{[h]}, \tag{7.1}
$$

where $T_d^h$ and $\tilde{T}_d^h$ are suitably dimensioned transformation matrices [3].

Let us define the process $Z$ as

$$
Z(t) \triangleq \begin{bmatrix} Y(t)\\ X(t)\end{bmatrix} \tag{7.2}
$$

and let $\delta = \dim(Z)$. Moreover, let us define the augmented process:

$$\mathcal{Z}(t) \triangleq \begin{bmatrix} Z(t) \\ Z_{[2]}(t) \\ \vdots \\ Z_{[\nu]}(t) \end{bmatrix}. \tag{7.3}$$

We can derive an SDE for the process $\mathcal{Z}$ in the following way. First, note that, from (3.2), (3.3), $Z$ satisfies the following SDE:

$$dZ(t) = \left(\tilde{A}(t)Z(t) + \alpha(t)\right) dt + \sum_{k=1}^{p} \left(B_k Z(t) + \tilde{\beta}_k\right) dW_k(t), \tag{7.4}$$

where

$$\tilde{A}(t) \triangleq \begin{bmatrix} 0 & C(t) \\ 0 & A(t) \end{bmatrix}; \quad \alpha(t) \triangleq \begin{bmatrix} 0 \\ Hu \end{bmatrix}; \quad B_k \triangleq \begin{bmatrix} 0 & D_k \\ 0 & B_k(t) \end{bmatrix}; \quad \beta_k = \begin{bmatrix} G_k \\ F_k \end{bmatrix}. \tag{7.5}$$

Next, by applying Theorem 6.1 to the process $Z$, it results for $i = 2, \ldots, \nu$ that

$$\begin{aligned} dZ^{[i]}(t) = {} & \left(L_i^0(t)Z^{[i]}(t) + L_i^1(t)Z^{[i-1]}(t) + L_i^2 Z^{[i-2]}(t)\right) dt \\ & + \sum_{k=1}^{p} \left(V_{k,i}^0 Z^{[i]}(t) + V_{k,i}^1 Z^{[i-1]}(t)\right) dW_k(t), \end{aligned} \tag{7.6}$$

where

$$L_i^0(t) = U_\delta^i \left(\tilde{A}(t) \otimes I_{\delta^{i-1}}\right) + \frac{1}{2} O_\delta^i \left(\Psi_2 \otimes I_{\delta^{i-2}}\right), \tag{7.7}$$

$$L_i^1(t) = U_\delta^{(i)}(\alpha(t) \otimes I_{\delta^{i-1}}) + \frac{1}{2} O_\delta^i \left(\Psi_1 \otimes I_{\delta^{i-2}}\right), \tag{7.8}$$

$$L_i^2 = \frac{1}{2} O_\delta^i \left(\Psi_0 \otimes I_{\delta^{i-2}}\right), \tag{7.9}$$

$$V_{k,i}^0 = U_\delta^i \left(\tilde{B}_k \otimes I_{\delta^{i-1}}\right), \tag{7.10}$$

$$V_{k,i}^1 = U_\delta^i(\beta_k \otimes I_{\delta^{i-1}}), \tag{7.11}$$

and $\Psi_2$, $\Psi_1$, and $\Psi_0$ are given by

$$\Psi_2 \triangleq \sum_{k=1}^{p} \tilde{B}_k^{[2]}, \quad \Psi_1 \triangleq \sum_{k=1}^{p} \left(\tilde{B}_k \otimes \beta_k + \beta_k \otimes \tilde{B}_k\right), \quad \Psi_0 \triangleq \sum_{k=1}^{p} \beta_k^{[2]}.$$

Observing that, from (7.1) we have

$$Z^{[i]} = T_\delta^i Z_{[i]}, \quad Z_{[i]} = \tilde{T}_\delta^i Z^{[i]},$$

and using (7.6), we can state the following proposition.

PROPOSITION 7.1. *The process $\mathcal{Z}$ defined in* (7.3) *satisfies the bilinear SDE*

$$d\mathcal{Z}(t) = (\mathcal{A}(t)\mathcal{Z}(t) + \mathcal{U}(t))dt + \sum_{k=1}^{p} \left(\mathcal{B}_k \mathcal{Z}(t) + \mathcal{V}_k\right) dW_k(t), \tag{7.12}$$

*where*

$$\mathcal{A}(t) = \begin{bmatrix} \tilde{A}(t) & 0 & \dots & & 0 \\ L_2^1(t) & \tilde{T}_\delta^2 L_2^0(t) T_\delta^2 & 0 & & \\ L_3^2 & \tilde{T}_\delta^3 L_3^1(t) T_\delta^2 & \tilde{T}_\delta^3 L_3^0(t) T_\delta^3 & & \vdots \\ & \ddots & \ddots & \ddots & \\ 0 & \dots & \tilde{T}_\delta^\nu L_\nu^2 T_\delta^{\nu-2} & \tilde{T}_\delta^\nu L_\nu^1(t) T_\delta^{\nu-1} & \tilde{T}_\delta^\nu L_\nu^0(t) T_\delta^\nu \end{bmatrix}, \tag{7.13}$$

$$\mathcal{B}_k = \begin{bmatrix} \tilde{B}_k & 0 & \dots & & 0 \\ V_{k,2}^1 & \tilde{T}_\delta^2 V_{k,2}^0 T_\delta^2 & & & \vdots \\ 0 & \tilde{T}_\delta^3 V_{k,3}^1 T_\delta^2 & \tilde{T}_\delta^3 V_{k,3}^0 T_\delta^3 & & \\ \vdots & & \ddots & \ddots & \\ 0 & \dots & & \tilde{T}_\delta^\nu V_{k,\nu}^1 T_\delta^{\nu-1} & \tilde{T}_\delta^\nu V_{k,\nu}^0 T_\delta^\nu \end{bmatrix}, \tag{7.14}$$

$$\mathcal{U}(t) = \begin{bmatrix} \alpha(t) \\ L_2^2 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \quad \mathcal{V}_k = \begin{bmatrix} \beta_k \\ 0 \\ \vdots \\ 0 \end{bmatrix}. \tag{7.15}$$

*The block matrices in* (7.13), (7.14) *are given by* (7.7)–(7.11) *and* (7.5)*, and the matrices* $\tilde{T}_\cdot^\cdot, T_\cdot^\cdot$*, are the reduction matrices defined in* (7.1)*.*

Now, we can use Theorem 4.1 in order to rewrite the bilinear SDE (7.12) in the form of a *linear* SDE with WSW diffusion term. The underlying hypothesis is that the covariance matrix of the process $\mathcal{Z}$ defined in (7.3), namely $\Phi_Z(t)$, is uniformly nonsingular over $T$. There are many ways to assure this, starting from some suitable, nonrestrictive hypothesis on the original system. As a matter of fact, since we are here concerned with a finite interval $T$, it is easy to recognize that the uniform nonsingularity of $\Phi_Z(t)$ is assured as soon as it is assumed that the covariance of the initial original state $X(0)$ is positive definite. Henceforth, we will understand the uniform nonsingularity in $T$ of $\Phi_Z(t)$.

PROPOSITION 7.2. *Let* $\rho_k$*,* $k = 1, \dots, p$*, be the ranks of the matrices* $\mathcal{B}_k$*, given in* (7.12)*. Then the process* $\mathcal{Z}$ *satisfies the SDE*

$$d\mathcal{Z}(t) = \big(\mathcal{A}(t)\mathcal{Z}(t) + \mathcal{U}(t)\big)dt + \sum_{k=1}^{2p} \tilde{\mathcal{B}}_k(t) d\tilde{W}_k(t), \tag{7.16}$$

*where* $\tilde{W}_k$*,* $k = 1, \dots, 2p$ *are independent standard WSW processes,* $\tilde{W}_k \in \mathbf{R}^{\rho_k}$ *for* $k = 1, \dots, p$*,* $\tilde{W}_k = W_k \in \mathbf{R}$*, for* $k = p+1, \dots, 2p$*, and*

$$\tilde{\mathcal{B}}_k(t) \triangleq \begin{cases} \big(\mathcal{B}_k \Psi_Z(t) \mathcal{B}_k^T\big)^{(\frac{1}{2})}, & 1 \le k \le p, \\ \mathcal{B}_{k-p} m_z(t) + \mathcal{V}_{k-p}, & p+1 \le k \le 2p, \end{cases} \tag{7.17}$$

*with* $m_z = E(\mathcal{Z})$*.*

In order to write down the equations of the *augmented system* we need to split out the vector SDE (7.12) into two SDEs: one for the *observed* components of $\mathcal{Z}$ and the other one for the remaining entries.

From the definition (7.2) we see that the components of the vector $\mathcal{Z}$ are of the form

$$X_1^{i_1} \cdots X_n^{i_n} \cdots Y_1^{j_1} \cdots Y_q^{j_q}, \tag{7.18}$$

where $X_l, Y_l$ denote the $l$th component of vectors $X, Y$, respectively, and $0 \leq i_l, j_r \leq \nu$ for $l = 1, \ldots, n$, $r = 1, \ldots, q$, $\sum_{l=1}^{n} i_l \leq \nu$, $\sum_{r=1}^{q} j_r \leq \nu$. The observed components are those of the form (7.18) with $i_1 = \cdots = i_n = 0$. Denote by $\mathcal{Y}$ the vector of all such components:

$$\mathcal{Y} \triangleq \begin{bmatrix} Y \\ Y_{[2]} \\ \vdots \\ Y_{[\nu]} \end{bmatrix}.$$

Moreover, let us denote by $\mathcal{E}_{\mathcal{Y}}$ the $(0,1)$-matrix such that

$$\mathcal{Y} = \mathcal{E}_{\mathcal{Y}} \mathcal{Z}. \tag{7.19}$$

It is easy to recognize that

$$\mathcal{E}_{\mathcal{Y}} = \begin{bmatrix} \mathcal{E}_{\mathcal{Y}}^1 & 0 & \ldots & 0 \\ 0 & \mathcal{E}_{\mathcal{Y}}^2 & \ddots & \\ \vdots & \ddots & \ddots & \\ 0 & \ldots & 0 & \mathcal{E}_{\mathcal{Y}}^{\nu} \end{bmatrix}, \tag{7.20}$$

where the diagonal blocks $\mathcal{E}_{\mathcal{Y}}^j$, $j = 1, \ldots, \nu$ are defined as

$$\mathcal{E}_{\mathcal{Y}}^j Z_{[j]} = Y_{[j]} \tag{7.21}$$

and have the expressions

$$\mathcal{E}_{\mathcal{Y}}^j = [I_q \;\; 0]^{[j]} T_\delta^j, \tag{7.22}$$

where $T_\delta^j$ is the expansion matrix defined in (7.1). Let us denote with $\mathcal{X}$ the aggregate vector of all the components in $\mathcal{Z}$ which are not components of $\mathcal{Y}$. Moreover, let us denote by $\mathcal{E}_{\mathcal{X}}$ the $(0,1)$-matrix such that

$$\mathcal{X} = \mathcal{E}_{\mathcal{X}} \mathcal{Z}. \tag{7.23}$$

A simple way to compute $\mathcal{E}_{\mathcal{X}}$ is just to remove from the identity matrix $I_{d_{\mathcal{Z}}}$ with $d_{\mathcal{Z}} = \dim(\mathcal{Z})$ (note that $I_{d_{\mathcal{Z}}}$ includes all the rows of $\mathcal{E}_{\mathcal{Y}}$) all those rows which are rows of $\mathcal{E}_{\mathcal{Y}}$.

From the above the aggregate matrix $\mathcal{I}$,

$$\mathcal{I} \triangleq \begin{bmatrix} \mathcal{E}_{\mathcal{Y}} \\ \mathcal{E}_{\mathcal{X}} \end{bmatrix} \tag{7.24}$$

results to be invertible. Let us consider the matrices $\mathcal{I}_1$, $\mathcal{I}_2$ such that

$$\mathcal{Z} = \mathcal{I}_1 \mathcal{Y} + \mathcal{I}_2 \mathcal{X}. \tag{7.25}$$

Note that, from (7.19), (7.23), and because of the invertibility of the matrix $\mathcal{I}$, it results that the matrices $\mathcal{I}_1, \mathcal{I}_2$ defined in (7.25) are obtained by means of a suitable partition of the matrix $\mathcal{I}^{-1} = [\mathcal{I}_1 \ \mathcal{I}_2]$.

Using (7.19), (7.23), (7.25), and (7.12), we can now state the following proposition.

PROPOSITION 7.3. *The processes* $\mathcal{X}$, $\mathcal{Y}$ *defined in* (7.23) *and* (7.19) *satisfy the pair of SDEs (augmented system)*

$$d\mathcal{X}(t) = \big(\mathcal{A}_1(t)\mathcal{Y}(t) + \mathcal{A}_2(t)\mathcal{X}(t) + \mathcal{U}_1(t)\big)dt + \sum_{k=1}^{2p} \mathcal{B}_k^1(t)d\tilde{W}_k(t), \tag{7.26}$$

$$d\mathcal{Y}(t) = \big(\mathcal{C}_1(t)\mathcal{Y}(t) + \mathcal{C}_2(t)\mathcal{X}(t) + \mathcal{U}_2(t)\big)dt + \sum_{k=1}^{2p} \mathcal{D}_k^1(t)d\tilde{W}_k(t), \tag{7.27}$$

*where*

$$\begin{aligned}
&\mathcal{A}_1(t) = \mathcal{E}_\mathcal{X}\mathcal{A}(t)\mathcal{I}_1, \quad \mathcal{A}_2(t) = \mathcal{E}_\mathcal{X}\mathcal{A}(t)\mathcal{I}_2, \quad \mathcal{U}_1(t) = \mathcal{E}_\mathcal{X}\mathcal{U}(t), \quad \mathcal{B}_k^1(t) = \mathcal{E}_\mathcal{X}\tilde{\mathcal{B}}_k(t),\\
&\mathcal{C}_1(t) = \mathcal{E}_\mathcal{Y}\mathcal{A}(t)\mathcal{I}_1, \quad \mathcal{C}_2(t) = \mathcal{E}_\mathcal{Y}\mathcal{A}(t)\mathcal{I}_2, \quad \mathcal{U}_2(t) = \mathcal{E}_\mathcal{Y}\mathcal{U}(t), \quad \mathcal{D}_k^1(t) = \mathcal{E}_\mathcal{Y}\tilde{\mathcal{B}}_k(t),
\end{aligned} \tag{7.28}$$

$\mathcal{A}$, $\tilde{\mathcal{B}}_k$, $\mathcal{U}$, *are the matrix coefficients of* (7.16)*, the matrices* $\mathcal{E}_\mathcal{X}$, $\mathcal{E}_\mathcal{Y}$, $\mathcal{I}_1$, $\mathcal{I}_2$, *are defined by means of* (7.19), (7.23), (7.25)*, and* $\{\tilde{W}_k,\ k = 1, \dots, 2p\}$ *is a set of mutually uncorrelated standard WSW processes.*

**8. Polynomial filter equations.** Proposition 7.3 states that the augmented observation process $\mathcal{Y}$ defined in (7.19) can be generated as the output process of the augmented representation (7.26), (7.27). This implies that the problem of finding the $\nu$th degree polynomial filter for the original system (7.26), (3.3) is now reduced to an *optimal linear filtering* problem for the linear system (7.26), (7.27). Indeed, by denoting with $\hat{\mathcal{X}}(t)$ the optimal linear estimate given $\{\mathcal{Y}_s,\ s \le t\}$ of the augmented state $\mathcal{X}(t)$, we have (see section 2)

$$\hat{\mathcal{X}}(t) = \Pi\Big(\mathcal{X}(t)/\mathcal{L}_t(\mathcal{Y})\Big).$$

On the other hand, from Definition 2.1 and taking into account the structure of the augmented observation $\mathcal{Y}$, it results that $\mathcal{L}_t(\mathcal{Y}) = \mathcal{P}_t^{(\nu)}(Y)$, where $Y$ is the original observation process given by (3.3). Hence we have

$$\hat{\mathcal{X}}(t) = \Pi\Big(\mathcal{X}(t)/\mathcal{P}_t^{(\nu)}(Y)\Big),$$

and, as we will see later, we can get $\hat{X}(t)$ (which is given by (3.5)) by extracting a suitable subvector in $\hat{\mathcal{X}}(t)$.

In [3] the optimal linear filter is defined for the class of linear stochastic systems whose noise terms are represented by WSW processes. System (7.26), (7.27) comes within this class of systems, and we can use here the same approach as in [3] in order to obtain the optimal linear filter with respect to the augmented observation process $\mathcal{Y}$ (and, hence the optimal $\nu$th degree polynomial filter with respect to the original observed process $Y$). In order to do this, first of all we state the following theorem, whose proof is given in Appendix B, showing the uniform nonsingularity in $T$ of the output-noise covariance of system (7.26), (7.27), namely

$$\mathcal{R}(t) \triangleq \sum_{k=1}^{2p} \mathcal{D}_k^1(t)\mathcal{D}_k^1(t)^T. \tag{8.1}$$

Indeed, the uniform nonsingularity of (8.1) is required, in order to apply the Kalman–Bucy scheme to system (7.26), (7.27).

THEOREM 8.1. *The noise covariance matrix function of the augmented measurement equation* (7.27)*, given by* (8.1)*, is uniformly nonsingular over* $T$*.*

*Proof.* See Appendix B. ∎

Now, we can prove the main theorem, defining the $\nu$th degree polynomial filter for system (3.2), (3.3). We remind readers that $\rho_k$ is the dimension of the WSW process $\tilde{W}_k$ when $k = 1, \ldots, p$, and for $k = p+1, \ldots, 2p$, $\tilde{W}_k = W_k \in \mathbf{R}$. Let us denote with $\gamma$ the dimension of the augmented process $\mathcal{Z}$. Moreover, we shall denote with $\operatorname{cov}(\chi, \eta)$ the cross-covariance between two random variables $\chi, \eta$. Finally, we shall denote with $M^\dagger$ the Moore–Penrose pseudoinverse of the square matrix $M$.

THEOREM 8.2. *The $\nu$th order polynomial filter for system* (3.2), (3.3) *is described by the following system of equations:*

$$\frac{dm_z(t)}{dt} = \mathcal{A}(t)m_z(t) + \mathcal{U}(t), \tag{8.2}$$

$$\bar{\mathcal{B}}_k(t) = \mathcal{B}_k m_z(t) + \mathcal{V}_k, \quad 1 \le k \le p, \tag{8.3}$$

$$\frac{d\Psi_Z(t)}{dt} = \mathcal{A}(t)\Psi_Z(t) + \Psi_Z(t)\mathcal{A}(t)^T + \sum_{k=1}^{p} \mathcal{B}_k \Psi_Z(t)\mathcal{B}_k^T + \sum_{k=1}^{p} \bar{\mathcal{B}}_k(t)\bar{\mathcal{B}}_k(t)^T, \tag{8.4}$$

$$\tilde{\mathcal{B}}_k(t) = \left(\mathcal{B}_k \Psi_Z(t)\mathcal{B}_k^T\right)^{(\frac{1}{2})}, \quad 1 \le k \le p, \tag{8.5}$$

$$\mathcal{J}(t) = \sum_{k=1}^{p} \mathcal{E}_\mathcal{X}\big(\tilde{\mathcal{B}}_k(t)\tilde{\mathcal{B}}_k(t)^T + \bar{\mathcal{B}}_k(t)\bar{\mathcal{B}}_k(t)^T\big)\mathcal{E}_\mathcal{Y}^T, \tag{8.6}$$

$$\mathcal{R}(t) = \sum_{k=1}^{p} \mathcal{E}_\mathcal{Y}\left(\tilde{\mathcal{B}}_k(t)\tilde{\mathcal{B}}_k(t)^T + \bar{\mathcal{B}}_k(t)\bar{\mathcal{B}}_k(t)^T\right)\mathcal{E}_\mathcal{Y}^T, \tag{8.7}$$

$$\mathcal{Q}(t) = \sum_{k=1}^{p} \mathcal{E}_\mathcal{X}\left(\tilde{\mathcal{B}}_k(t)\tilde{\mathcal{B}}_k(t)^T + \bar{\mathcal{B}}_k(t)\bar{\mathcal{B}}_k(t)^T\right)\mathcal{E}_\mathcal{X}^T, \tag{8.8}$$

$$\begin{aligned}\frac{d\mathcal{P}(t)}{dt} &= \mathcal{A}_2(t)\mathcal{P}(t) + \mathcal{P}(t)\mathcal{A}_2(t)^T + \mathcal{Q}(t)\\ &\qquad - \Big(\mathcal{J}(t) + \mathcal{P}(t)\mathcal{C}_2(t)^T\Big)\mathcal{R}(t)^{-1}\big(\mathcal{J}(t) + \mathcal{P}(t)\mathcal{C}_2(t)^T\big)^T,\end{aligned} \tag{8.9}$$

$$\begin{aligned}d\hat{\mathcal{X}}(t) &= \Big(\mathcal{A}_1(t)\mathcal{Y}(t) + \mathcal{A}_2(t)\hat{\mathcal{X}}(t) + \mathcal{U}_1(t)\Big)dt + \Big(\mathcal{J}(t) + \mathcal{P}(t)\mathcal{C}_2(t)^T\Big)\mathcal{R}(t)^{-1}\\ &\qquad \cdot\big(d\mathcal{Y}(t) - \big(\mathcal{C}_1(t)\mathcal{Y}(t) + \mathcal{C}_2(t)\hat{\mathcal{X}}(t) + \mathcal{U}_2(t)\big)dt\big),\end{aligned} \tag{8.10}$$

$$\hat{X}(t) = \mathcal{T}_\nu \hat{\mathcal{X}}(t), \tag{8.11}$$

*where $\mathcal{T}_\nu$ is the operator extracting the first $n$ entries of a vector, the matrices $\mathcal{A}(t)$, $\mathcal{U}(t)$, $\mathcal{A}_1(t), \mathcal{A}_2(t)$, $\mathcal{B}_1(t), \mathcal{B}_2(t), \mathcal{U}_1(t), \mathcal{U}_2(t)$ are defined in* (7.16) *and* (7.28)*, the matrices $\mathcal{B}_k$ are defined in* (7.14)*, $\rho_k = \operatorname{rank}(\mathcal{B}_k)$, and* (8.2), (8.4), (8.9), (8.10) *are endowed with the initial conditions*

$$\begin{aligned}
&m_z(0) = E(\mathcal{X}(0)),\\
&\Psi_Z(0) = \operatorname{cov}(\mathcal{X}(0), \mathcal{X}(0)),\\
&\hat{\mathcal{X}}(0) = E(\mathcal{X}(0)) + \operatorname{cov}(\mathcal{X}(0), \mathcal{Y}(0))\operatorname{cov}(\mathcal{Y}(0), \mathcal{Y}(0))^\dagger\big(\mathcal{Y}(0) - E(\mathcal{Y}(0))\big),\\
&\mathcal{P}(0) = \operatorname{cov}(\mathcal{X}(0), \mathcal{X}(0)) - \operatorname{cov}(\mathcal{X}(0), \mathcal{Y}(0))\operatorname{cov}(\mathcal{Y}(0), \mathcal{Y}(0))^\dagger\operatorname{cov}(\mathcal{X}(0), \mathcal{Y}(0))^T.
\end{aligned}$$

*Proof.* Equations (8.6)–(8.10) easily derive from an application of [3, Thm. 15.3] to the representation (7.26), (7.27). The augmented-state covariance $\Psi_Z(t)$, appearing in the definition of $\tilde{\mathcal{B}}_k$ given by (8.5) (see also (7.17)), satisfies the ODE (8.4). This can be readily proved in the same way of (4.20), but the time-varying BLSS (7.12) is considered now, and the semigroup generated by $\{\mathcal{A}(t),\ t \in T\}$ should be used.

The so-obtained estimate $\hat{\mathcal{X}}_t$ is the optimal one among all the linear transformation of the augmented observation process $\{\mathcal{Y}_s,\ s \le t\}$, and hence it is the $\nu$th degree polynomial estimate of the augmented state $\mathcal{X}_t$. In order to obtain the $\nu$th degree polynomial estimate of the state $X_t$ of the original system (3.2), (3.3), first of all note that, because $\hat{\mathcal{X}}_t$ is the $L^2$-projection of $\mathcal{X}_t$ onto the closed subspace linearly spanned by $\{\mathcal{Y}_s,\ s \le t\}$, we have that each entry of $\hat{\mathcal{X}}_t$ agrees with the $L^2$-projection (onto the same subspace) of the corresponding entry in $\mathcal{X}_t$. Now, by definition, $\mathcal{X}(t)$ includes the components of the original state $X_t$. From (7.2), (7.3), and by the definition of the extracting operator $\mathcal{E}_{\mathcal{X}}$, it results that these components are placed in the first $n$ entries of the vector $\mathcal{X}$. Hence, $\hat{X}(t)$ can be obtained simply by extracting the first $n$ components of $\hat{\mathcal{X}}_t$, that is (8.11). ∎

**9. Simulation example.** In order to test the algorithm described in the previous sections, the filtering problem for the following second-order system has been considered:

$$dX(t) = AX(t)dt + BX(t)dW(t) + UdN(t), \quad X(0) = 0, \tag{9.1}$$

$$dY(t) = CX(t)dt + DX(t)dV(t), \quad Y(0) = 0, \tag{9.2}$$

where

$$A = \begin{bmatrix} a_1 & 1 \\ 0 & a_2 \end{bmatrix}, \quad B = \begin{bmatrix} b_1 & 0 \\ 0 & b_2 \end{bmatrix}, \quad U = \begin{bmatrix} u_1 \\ u_2 \end{bmatrix},$$
$$C = [1 \quad 1], \qquad D = [g \quad 0],$$

and $W$, $N$, and $V$ are mutually independent scalar Wiener processes.

The well-known extended Kalman filter (EKF) was up to now the classical tool for the filtering of a nonlinear system in the form of (9.1), (9.2). However, nothing is known about the working conditions or the performances of the EFK. In the present case, for instance, the EKF does not work at all. Indeed the state-expectation is zero and hence the state process is expected to cross the zero. Since the term $D\hat{X}(t)\hat{X}(t)^T D^T$ ($\hat{X}$ denoting the EKF estimation) needs to be inverted in the EKF equations, we should expect a failure of the algorithm. This really happens in the simulations we carried out for several values of the parameters $a_1, a_2, b_1, b_2, u_1, u_2, g$. We have always observed a sudden and strong deviation to infinity. In order to improve the working conditions we have substituted the term $D\hat{X}(t)\hat{X}(t)^T D^T$ with $\epsilon + D\hat{X}(t)\hat{X}(t)^T D^T$, where the number $\epsilon$ has been chosen small enough. In these cases we have observed an improvement of the algorithm, in that for a small initial time-interval the EKF shows a good performance, even better than the third-degree polynomial filter below described (no theoretical argument is known about this). However, unavoidably, the EKF diverges in spite of the trick used, whereas the polynomial filters continue to work.

The linear, the second degree (quadratic), and the third degree (cubic) filters have been built up by using (8.2)–(8.11). We remind the reader that the matrices $\mathcal{A}$, $\mathcal{U}$, $\mathcal{B}_k$, and $\mathcal{V}_k$ that appear in the filter equations are the system-matrices of the augmented (7.12). These matrices can be obtained from the original system-matrices

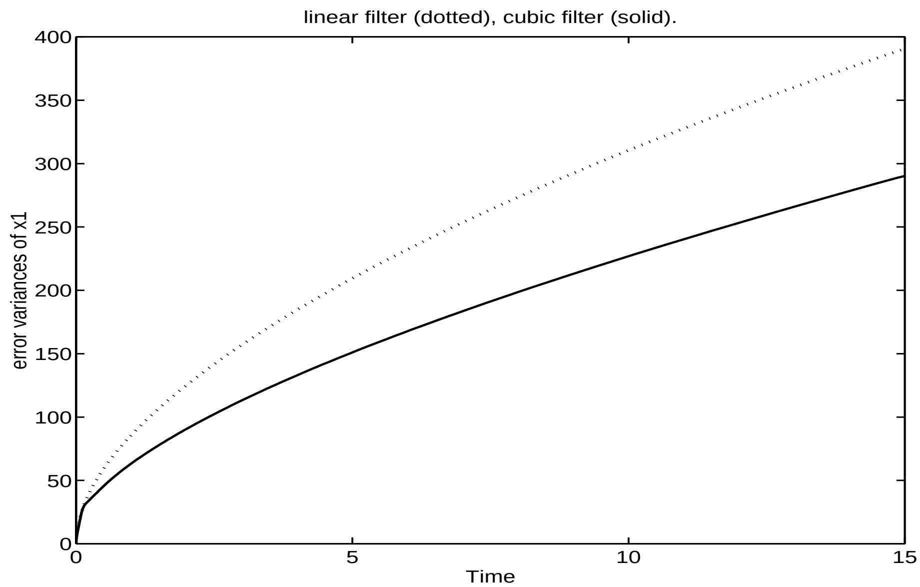


FIG. 9.1

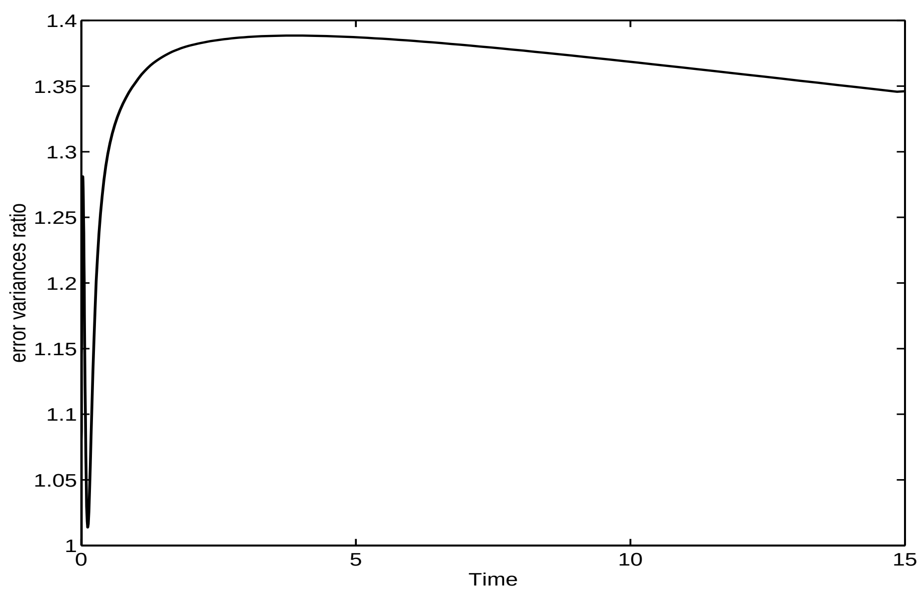


FIG. 9.2

by using the formulas given in section 7 for any polynomial degree. We show below our simulation results for the linear and cubic filters, with the following values of the parameters:

$$
\begin{aligned}
a_1 = -0.01, \quad a_2 = -0.5, \quad u_1 = 30, \quad u_2 = 2, \\
b_1 = 0.1, \quad b_2 = 0.1, \quad g = 0.1.
\end{aligned}
\tag{9.3}
$$

We do not show graphs related to the quadratic filter simulation because, in our case,

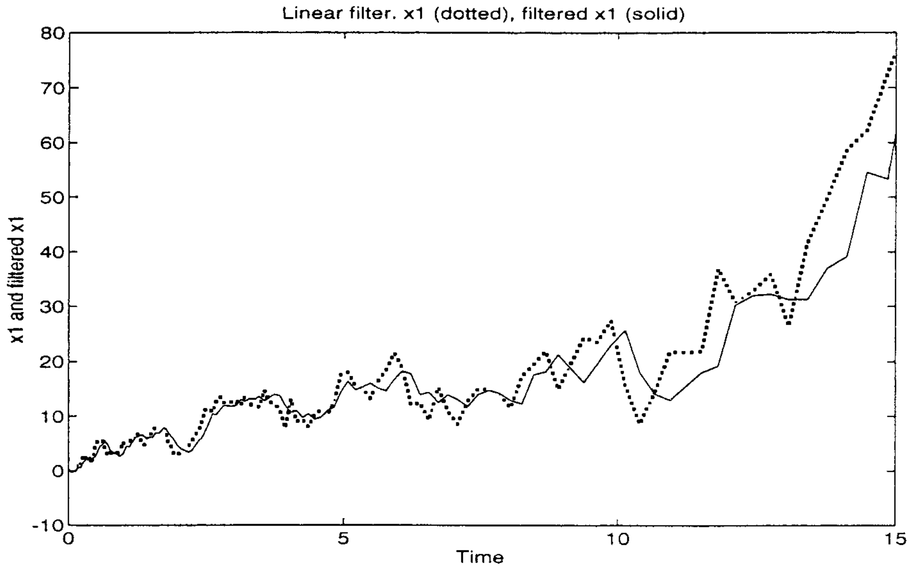


FIG. 9.3

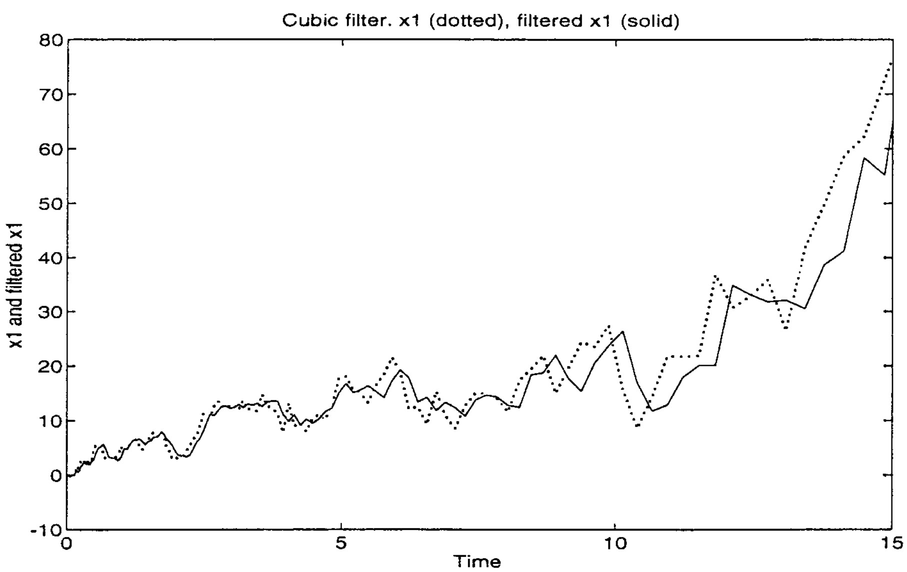


FIG. 9.4

the quadratic filter does not show any valuable improvement with respect to the linear case. Differently from the EKF, for the polynomial filters we are able to compute the a priori state-estimate error variances that are entries of the matrix $\mathcal{P}(t)$ given by (8.9), that is $\mathcal{P}_{1,1}(t)$, $\mathcal{P}_{2,2}(t)$ for the first and second state components, respectively. In our example these values are growing with time. The reason for this is that system (9.1), (9.2), with the values given by (9.3), is unstable. Nevertheless, as shown in Figure 9.1, the time-evolution of $\mathcal{P}_{1,1}(t)$ for the cubic filter (namely $P_C(t)$) is ever less than the $\mathcal{P}_{1,1}(t)$ for the linear filter (namely $P_L(t)$). In Figure 9.2 the evolution of the ratio $\rho(t) = P_L(t)/P_C(t)$ is shown. We can see that $\rho(t)$ stabilizes over the value $\bar{\rho} = 1.30$. Hence the improvement in the a priori performance of the cubic filter with respect to the linear one can be considered almost 30%.

The time-evolutions of filtered paths for the linear and cubic filters, compared with the corresponding true 1st component state path, are reported in Figures 9.3, 9.4. As we can see, even a visual comparison between the signal time-evolutions shows a valuable improvement in the estimation quality of the cubic filter with respect to the linear one. Several Monte Carlo runs have been carried out. For each one of these runs, the ratio, namely $\rho_S$, between the sampled error variances of the linear and cubic filters has been computed. We have chosen the paths with $\rho_S = 1.35$.

The simulation of the EKF confirms also in this case its unsatisfactory behavior. Indeed, after almost 0.01 time units the EKF estimate starts up and quickly goes to infinity.

All the simulations have been carried out using the standard functions of the Matlab software package for Windows. The computer was a PC, endowed with a 200 MHz Pentium processor.

**10. Conclusions.** Equations (8.2)–(8.11) define a finite-dimensional filter for the BLSS (3.2), (3.3) which is optimal in a class of polynomial estimates. Although the considered class does not include all the polynomials of the currently available measurements, it includes the linear estimates, and, moreover, it defines a nondecreasing sequence of spaces for increasing polynomial degree. This implies that the polynomial filter had to improve the estimation performance for increasing polynomial degree.

We underline that the proposed filter is finite-dimensional. Of course, it is always possible to approximate the optimal filter (for instance, by applying a finite-elements method to the Zakai equation, as shown in [13]) with an arbitrary approximation degree. However, the more accurately the approximation level is chosen, the heavier the computational burden of the algorithm is. The computational effort is prohibitive even for small approximation degrees. Moreover, it makes no sense, within this approach, to use a large approximation degree in order to make the filtering algorithm really implementable. Counterwise, our suboptimal approach allows us to get meaningful estimates also for small polynomial degrees, which do not present difficult implementation problems.

In section 4 we have presented the equations of the optimal linear filter for a BLSS. We highlight that this result is interesting by itself in that it was up to now known only for the scalar case. The main tool is given by Theorem 4.1, stating the existence of a linear representation for a general vector BLSS. The optimal linear filter is then obtained by an application of a classical Kalman–Bucy scheme. Nevertheless, in the framework of this paper, the main purpose of Theorem 4.1 remains its application to the bilinear SDE (7.12), which allows us to obtain the linear representation (7.16).

Theorem 8.1 states that the output noise covariance of the augmented system is uniformly nonsingular, as required by the Kalman–Bucy scheme, provided that the output noise covariance of the original system (3.2), (3.3) is nonsingular. The proof is presented in Appendix B.

In section 9 a numerical simulation is shown, where a second-order BLSS has been filtered using the polynomial filters up to the third degree. The EFK has been also simulated, however its performance is resulted to be unsatisfactory. The simulation results show that the estimation quality really improves as polynomial degree grows, and for the cubic filter we obtained an improvement valuable over 30% with respect to the linear filter.

We stress that, due to the well-known approximation capabilities of the polynomial functions, with the aim to define better and better implementable approximation schemes of the optimal filter, the use of polynomial estimators appears to be very promising.

**Appendix A. Kronecker algebra.**

Throughout this paper, we have widely used Kronecker algebra [14], [15]. Here, for the sake of completeness, we recall some definitions and properties on this subject.

DEFINITION A.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*

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

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

Of course this kind of product is not commutative.

DEFINITION A.2. *Let $M$ be the $r \times s$ matrix*

$$M = [\, m_1 \quad m_2 \quad \dots \quad m_s \,], \tag{A.1}$$

*where $m_i$ denotes the $i$th column of $M$, and then the stack of $M$ is the $r \cdot s$ vector*

$$st(M) = [\, m_1^T \quad m_2 \quad \dots \quad m_s \,]^T. \tag{A.2}$$

Observe that a vector as in (A.2) can be reduced to a matrix $M$ as in (A.1) 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 [15]:

$$\begin{aligned} (A+B)\otimes(C+D) &= A\otimes C + A\otimes D \\ &\quad + B\otimes C + B\otimes D, && \text{(A.3a)} \\ A\otimes(B\otimes C) &= (A\otimes B)\otimes C, && \text{(A.3b)} \\ (A\cdot C)\otimes(B\cdot D) &= (A\otimes B)\cdot(C\otimes D), && \text{(A.3c)} \\ (A\otimes B)^T &= A^T\otimes B^T, && \text{(A.3d)} \\ st(A\cdot B\cdot C) &= (C^T\otimes A)\cdot st(B), && \text{(A.3e)} \\ u\otimes v &= st(v\cdot u^T), && \text{(A.3f)} \\ tr(A\otimes B) &= tr(A)\cdot tr(B), && \text{(A.3g)} \end{aligned}$$

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

$$\begin{aligned} &M^{[0]} = 1, \\ &M^{[n]} = M \otimes M^{[n-1]} = M^{[n-1]} \otimes M, \quad n > 0. \end{aligned}$$

As an easy consequence of (A.3b) and (A.3g), it follows that

$$tr(A^{[h]}) = \big(tr(A)\big)^h. \tag{A.3h}$$

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

$$(u \otimes v)_i = u_l \cdot v_m; \quad l = \left[\frac{i-1}{s}\right] + 1, \quad m = |i-1|_s + 1, \tag{A.4}$$

where $[\cdot]$ and $|\cdot|_s$ denote integer part and $s$-modulo, respectively. Although the Kronecker product is not commutative, the following property holds [9, 15].

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

$$B \otimes A = C_{r,n}^T (A \otimes B) C_{s,m}, \tag{A.5}$$

*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*

$$\{C_{u,v}\}_{h,l} = \begin{cases} 1 & \textit{if } l = (|h-1|_v)u + \left(\left[\frac{h-1}{v}\right] + 1\right), \\ 0 & \textit{otherwise.} \end{cases} \tag{A.6}$$

Observe that $C_{1,1} = 1$; hence in the vector case when $a \in \mathbf{R}^r$ and $b \in \mathbf{R}^n$, (A.5) becomes

$$b \otimes a = C_{r,n}^T (a \otimes b). \tag{A.7}$$

COROLLARY A.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 $\mathbf{R}^l$, we have*

$$A \otimes B \otimes C \otimes D = \left(I(n_A) \otimes C_{n_C n_D, n_B}^T\right) (A \otimes C \otimes D \otimes B) \left(I(m_A) \otimes C_{m_C m_D, m_B}\right).$$

*Proof.* See [3]. □

**Appendix B. Proof of Theorem 8.1.** We need to state in advance some preliminary definitions and lemmas.

Let $\delta$ and $j$ be two positive integers.

DEFINITION B.1. *Let $r, s \in \{1, 2, \ldots, \delta^j\}$. The pair $(r, s)$ is said to be $(\delta, j)$-redundant ($(\delta, j)$-R for short) if $\forall\ x \in \mathbf{R}^\delta$, it results that $\left(x^{[j]}\right)_r = \left(x^{[j]}\right)_s$, where $\left(x^{[j]}\right)_l$ denotes the $l$th entry of the vector $x^{[j]}$. Otherwise, the pair $(r, s)$ is said to be $(\delta, j)$-nonredundant ($(\delta, j)$-NR for short).*

*Example* B.2. The pair $(2, 3)$ is $(2, 2)$-R; however, it is $(3, 2)$-NR. The pairs $(1, 1)$, $(2, 2), \ldots$ are $(\delta, j)$-NR for any $\delta$ and $j$.

*Remark* B.3. Let $x \in \mathbf{R}^\delta$. For some $s, r \in \{1, 2, \ldots, \delta^j\}$ let us consider the multi-indexes $s_1, \ldots, s_j$ and $r_1, \ldots, r_j$ in $\{1, \ldots, \delta\}$ defined by the identities

$$\left(x^{[j]}\right)_s = x_{s_1} x_{s_2} \cdots x_{s_j}, \quad \left(x^{[j]}\right)_r = x_{r_1} x_{r_2} \cdots x_{r_j}.$$

Then, we immediately realize that $(r, s)$ is $(\delta, j)$-R if and only if there exists a permutation of indexes transforming $s_1, \ldots, s_j$ in $r_1, \ldots, r_j$ (and vice versa).

*Remark* B.4. It is easy to verify that the $(\delta, j)$-R condition defines an equivalence relation in the set $\{1, 2, \ldots, \delta^j\}$. We shall denote with $\rho(s; \delta, j)$ the equivalence class generated by $s \in \{1, \ldots, \delta^j\}$ via the $(\delta, j)$-R relation

$$\rho(s; \delta, j) \triangleq \left\{ r \in \mathbf{N} \ :\ 1 \le r \le \delta^j,\ (s, r) \textit{ is } (\delta, j)\text{–R} \right\}. \tag{B.1}$$

We shall denote with $\delta_j$ the number of equivalence classes of the $(\delta, j)$-R relation, partitioning the set $\{1, 2, \ldots, \delta^j\}$. Moreover, we introduce the sets $\rho'(s;\delta,j)$, $\rho''(s;\delta,j) \subset \rho(s;\delta,j)$ defined as

$$\rho'(s;\delta,j) \triangleq \left\{ i \in \rho(s;\delta,j) \ \Big/ \left[\frac{i}{\delta^{j-1}}\right] = \left[\frac{s}{\delta^{j-1}}\right] \right\}, \tag{B.2}$$

$$\rho''(s;\delta,j) \triangleq \rho(s;\delta,j) \setminus \rho'(s;\delta,j), \tag{B.3}$$

where we have used in (B.2) the notation $[\ \cdot\ ]$ to indicate the integer part. The above defined sets have the following meaning. Let $x \in \mathbf{R}^\delta$ and note that

$$x^{[j]} = \begin{bmatrix} x_1 \cdot x^{[j-1]} \\ x_2 \cdot x^{[j-1]} \\ \vdots \\ x_\delta \cdot x^{[j-1]} \end{bmatrix}, \tag{B.4}$$

where every subvector $x_i x^{[j-1]}$ has dimension $\delta^{j-1}$. By setting $l = [s/\delta^{j-1}]$ and observing in (B.4) the structure of $x^{[j]}$, we realize that the set defined in (B.2) is composed with the integers $i$ such that $(i, s)$ is $(\delta, j)$-R and $\big(x^{[j]}\big)_i \in x_l x^{[j-1]}$. Counterwise, the set defined in (B.3) is composed with the integers $i$ such that $(i, s)$ is $(\delta, j)$-R and $\big(x^{[j]}\big)_i$ does not belong to $x_l x^{[j-1]}$. Let us denote by $|n_1|_{n_2}$ the remainder of the integer division $n_1/n_2$. Then, again from (B.4), it is easily recognized that

$$\big(x^{[j]}\big)_s = x_l \big(x^{[j-1]}\big)_r, \quad r \triangleq |s|_{\delta^{j-1}}. \tag{B.5}$$

*Remark* B.5. Note that the number $\delta_j$ agrees with the number of entries of $x_{[j]}$ for $x \in \mathbf{R}^\delta$.

LEMMA B.6. *Let $r, s \in \{1, \ldots, \delta^{j-1}\}$ such that $(r, s)$ is $(\delta, j-1)$-R. Then, for any $l = 0, 1, \ldots, \delta - 1$, the pair $(r + l\delta^{j-1}, s + l\delta^{j-1})$ is $(\delta, j)$-R. Counterwise, if $r, s \in \{1, \ldots, \delta^j\}$ are $(\delta, j)$-R and $r' = s'$ with*

$$r' \triangleq \left[\frac{r}{\delta^{j-1}}\right], \quad s' \triangleq \left[\frac{s}{\delta^{j-1}}\right],$$

*then, denoting $r'' = |r|_{\delta^{j-1}}$, $s'' = |s|_{\delta^{j-1}}$, it results that $(r'', s'')$ is $(\delta, j-1)$-R.*

*Proof.* From Definition B.1 it results that

$$\big(x^{[j-1]}\big)_r = \big(x^{[j-1]}\big)_s \quad \forall\, x \in \mathbf{R}^\delta. \tag{B.6}$$

From (B.4) we see that

$$\big(x^{[j]}\big)_{r+l\delta^{j-1}} = x_l \big(x^{[j-1]}\big)_r, \quad \big(x^{[j]}\big)_{s+l\delta^{j-1}} = x_l \big(x^{[j-1]}\big)_s,$$

and hence, from (B.6),

$$\big(x^{[j]}\big)_{r+l\delta^{j-1}} = \big(x^{[j]}\big)_{s+l\delta^{j-1}}.$$

Counterwise, if $r, s \in \{1, \ldots, \delta^j\}$ are $(\delta, j)$-R, then, taking into account (B.5), we have

$$\big(x^{[j]}\big)_r = x_{r'} \big(x^{[j-1]}\big)_{r''} = x_{s'} \big(x^{[j-1]}\big)_{s''} = \big(x^{[j]}\big)_s \quad \forall\, x \in \mathbf{R}^\delta. \tag{B.7}$$

Since, by hypothesis, $r' = s'$, (B.7) implies that $\left(x^{[j-1]}\right)_{r''} = \left(x^{[j-1]}\right)_{s''}$. ◻

Let $\mathcal{I} \subset \mathbf{N}$ and $n \in \mathbf{N}$. In the following, we will use the notation $\mathcal{I} - n$ to indicate the *translated* set:

$$\mathcal{I} - n = \{i \ / \ i \in \mathbf{N},\ \exists i' \in \mathcal{I},\ \text{such that } i = i' - n\}. \tag{B.8}$$

LEMMA B.7. *Suppose that*

$$\left[\frac{s}{\delta^{j-1}}\right] = l < \delta. \tag{B.9}$$

*Then, for any $q < \delta - l$ it results that*

$$\rho'(s; \delta, j) = \rho'(s + q\delta^{j-1}; \delta, j) - q\delta^{j-1},$$

*where $\rho'$ is the set defined in* (B.2).

*Proof.* It suffices to show that for any $r \in \{1, \dots, \delta^j\}$ such that $[r/\delta^{j-1}] = l$ and such that $(r, s)$ is $(\delta, j)$-NR ($(\delta, j)$-R), the pair $(r + q\delta^{j-1}, s + q\delta^{j-1})$ is $(\delta, j)$-NR ($(\delta, j)$-R).

Suppose first that $(r, s)$ is $(\delta, j)$-NR. Let $x \in \mathbf{R}^\delta$ and $z = x^{[j]}$. From the structure (B.4) of the vector $z$ and taking into account (B.9), we see that $z_s, z_r \in x_l \cdot x^{[j-1]}$. Hence since $(s, r)$ is $(\delta, j)$-NR, we have that, there exist integers $h_1, \dots, h_\delta$ and $h'_1, \dots, h'_\delta$, $h_1 + \cdots + h_\delta = h'_1 + \cdots + h'_\delta = j - 1$ such that it results that

$$\begin{aligned} z_s &= x_l \cdot x_1^{h_1} \cdots x_\delta^{h_\delta}, \\ z_r &= x_l \cdot x_1^{h'_1} \cdots x_\delta^{h'_\delta}. \end{aligned} \tag{B.10}$$

Since $z_r \neq z_s$ it follows that

$$x_1^{h_1} \cdots x_\delta^{h_\delta} \neq x_1^{h'_1} \cdots x_\delta^{h'_\delta}. \tag{B.11}$$

Again, looking at (B.4), we readily realize that

$$z_{s+q\delta^{j-1}} = x_{l+q} \cdot x_1^{h_1} \cdots x_\delta^{h_\delta}, \tag{B.12}$$

and

$$z_{r+q\delta^{j-1}} = x_{l+q} \cdot x_1^{h'_1} \cdots x_\delta^{h'_\delta}, \tag{B.13}$$

and hence, taking into account (B.11), it follows that $z_{s+q\delta^{j-1}} \neq z_{r+q\delta^{j-1}}$; that is $(s + q\delta^{j-1}, r + q\delta^{j-1})$ is $(\delta, j)$-NR.

Next, suppose that $(r, s)$ is $(\delta, j)$-R. Then $z_s, z_r \in x_l \cdot x^{[j-1]}$, $z_s = z_r$, and by (B.10) it follows that $h_i = h'_i$, $i = 1, \dots, \delta$. This in turn implies, taking into account (B.12), (B.13), that $z_{s+q\delta^{j-1}} = z_{r+q\delta^{j-1}}$; that is $(s + q\delta^{j-1}, r + q\delta^{j-1})$ is $(\delta, j)$-R. ◻

LEMMA B.8. *Let $(r, s)$ be a $(\delta, j)$-R pair such that*

$$\left[\frac{r}{\delta^{j-1}}\right] = l, \qquad \left[\frac{s}{\delta^{j-1}}\right] = m, \quad l < m < \delta. \tag{B.14}$$

*Then for any $q < \delta - l$ the pair $(r + q\delta^{j-1}, s + q\delta^{j-1})$ is $(\delta, j)$-NR.*

*Proof.* As in the proof of Lemma B.7 it is readily verified that, for some integers $h_1, \dots, h_\delta$ such that $h_1 + \cdots + h_\delta = j - 1$, it results that

$$z_r = x_l \cdot x_1^{h_1} \cdots x_l^{h_l} \cdots x_m^{h_m} \cdots x_\delta^{h_\delta}. \tag{B.15}$$

Since $z_s = z_r$, (B.15) implies that

$$z_s = x_m \cdot x_1^{h_1} \cdots x_l^{h_l+1} \cdots x_m^{h_m-1} \cdots x_\delta^{h_\delta}.$$

Hence we have

$$z_{r+q\delta^{j-1}} = x_{l+q} \cdot x_1^{h_1} \cdots x_l^{h_l} \cdots x_m^{h_m} \cdots x_\delta^{h_\delta} = x_{l+q} \frac{z_r}{x_l},$$
$$z_{s+q\delta^{j-1}} = x_{m+q} \cdot x_1^{h_1} \cdots x_l^{h_l+1} \cdots x_m^{h_m-1} \cdots x_\delta^{h_\delta} = x_{m+q} \frac{z_s}{x_m}.$$

From this, since $z_r = z_s$ and $l \neq m$, it follows that $z_{r+q\delta^{j-1}} \neq z_{s+q\delta^{j-1}}$. ∎

Let us consider the state and output processes $X, Y$, of system (3.2), (3.3). We remind the reader that $q$ and $\delta$ are the dimensions of the vectors $Y$ and $Z = [Y^T\ X^T]^T$, respectively. Note that the components of $Z^{[j]}$ can be divided into two groups: the one including monomials composed only with components of the vector $Y$, and the other one including the remaining monomials. We shall call the components belonging to the former group *the $Y$-monomials*.

Let us consider the extraction matrix $\mathcal{E}_\mathcal{Y}$ defined in (7.19), and recall that the diagonal blocks $\mathcal{E}_\mathcal{Y}^j$, $j = 1, \ldots, \nu$, appearing there are such that (7.21) holds. According to the above defined notation (see Remark B.5), we shall denote by $q_j$ the dimension of the vector $Y_{[j]}$. Finally, let us consider the reduction matrix $\tilde{T}_\delta^j$ defined in (7.1) and the matrix $U_\delta^j$ defined in (5.5). We can prove the following lemma.

LEMMA B.9. *There exists a full (row) rank matrix, namely $L_\delta^j$, having dimensions $q_j \times q\delta^{j-1}$, such that*

$$\mathcal{E}_\mathcal{Y}^j \tilde{T}_\delta^j U_\delta^j = [\, L_\delta^j \quad 0\,].$$

*Proof.* Using (7.21) and property (5.4) we have

$$\text{(B.16)}\qquad \begin{aligned} \mathcal{E}_\mathcal{Y}^j\Big(\frac{d}{dZ} \otimes Z_{[j]}\Big) &= \frac{d}{dZ} \otimes \mathcal{E}_\mathcal{Y}^j Z_{[j]} = \frac{d}{dZ} \otimes Y_{[j]} \\ &= \Big[\frac{\partial}{\partial Y} \quad \frac{\partial}{\partial X}\Big] \otimes Y_{[j]} = \Big[\frac{\partial}{\partial Y} \otimes Y_{[j]} \quad 0\Big]. \end{aligned}$$

On the other hand, by (7.1), (5.3), and using formula (5.5),

$$\text{(B.17)}\qquad \begin{aligned} \mathcal{E}_\mathcal{Y}^j\Big(\frac{d}{dZ} \otimes Z_{[j]}\Big) &= \mathcal{E}_\mathcal{Y}^j\Big(\frac{d}{dZ} \otimes \tilde{T}_\delta^j Z^{[j]}\Big) = \mathcal{E}_\mathcal{Y}^j \tilde{T}_\delta^j U_\delta^j \big(I_\delta \otimes Z^{[j-1]}\big) \\ &= \mathcal{E}_\mathcal{Y}^j \tilde{T}_\delta^j U_\delta^j \begin{bmatrix} I_q \otimes Z^{[j-1]} & 0 \\ 0 & I_{\delta-q} \otimes Z^{[j-1]} \end{bmatrix}. \end{aligned}$$

Using (B.16), (B.17), and defining $L_\delta^j$ as the matrix composed by the first $q\delta^{j-1}$ columns of $\mathcal{E}_\mathcal{Y}^j \tilde{T}_\delta^j U_\delta^j$, it results that

$$[\, \tfrac{\partial}{\partial Y} \otimes Y_{[j]} \quad 0\,] = [\, L_\delta^j \quad S\,] \begin{bmatrix} I_q \otimes Z^{[j-1]} & 0 \\ 0 & I_{\delta-q} \otimes Z^{[j-1]} \end{bmatrix},$$

from which it follows that $S = 0$ and

$$\text{(B.18)}\qquad \frac{d}{dY} \otimes Y_{[j]} = L_\delta^j \Big(I_q \otimes Z^{[j-1]}\Big).$$

Let $V = (d/dY) \otimes Y_{[j]}$. Note that the components of the matrix $V$ are either zero or they are monomials of $j-1$st degree. It results that $V$ has linearly independent rows (in the sense of linear independence of monomial functions). As a matter of fact, any row is different from zero, and there cannot exist two (nonzero) similar monomials on the same column, because $Y_{[j]}$ has not repeated entries. Hence $L_\delta^j$ necessarily has linearly independent rows. Indeed, suppose there exists $u \neq 0$ such that $u^T L_\delta^j = 0$; then we would have $u^T V = 0 \ \ \forall\, Y \in \mathbf{R}^q$, which is a contradiction. □

LEMMA B.10. *Let $s \in \{1, \dots, q\delta^{j-1}\}$ and denote with $\lambda_i$, $i = 1, \dots, q\delta^{j-1}$, the $i$th column of the matrix $L_\delta^j$. The following properties hold:*

(A) *$\forall\, i \in \{1, \dots, q\delta^{j-1}\}$, $\lambda_i$ has zero entries, but possibly one, nonnegative;*

(B) *the set $\{\lambda_i /\ i \in \rho'(s; \delta, j)\}$, with $\rho'(s; \delta, j)$ given by* (B.2), *is a set of linearly dependent vectors;*

(C) *if the $s$th component of $Z^{[j]}$ is not a $Y$-monomial, then $\lambda_s = 0$.*

*Proof.* Let us define $l$ and $r$ as

$$l \triangleq \left[\frac{s}{\delta^{j-1}}\right], \quad r \triangleq |s|_{\delta^{j-1}}. \tag{B.19}$$

Consider again the relation (B.18):

$$\frac{d}{dY} \otimes Y_{[j]} = \left[\, \tfrac{\partial}{\partial Y_1} Y_{[j]} \quad \cdots \quad \tfrac{\partial}{\partial Y_q} Y_{[j]} \,\right] = L_\delta^j \big(I_q \otimes Z^{[j-1]}\big). \tag{B.20}$$

From (B.20) it results that

$$\frac{\partial}{\partial Y_l} Y_{[j]} = \tilde{L}^{(l)} Z^{[j-1]}, \tag{B.21}$$

where

$$\tilde{L}^{(l)} \triangleq \left[\, \lambda_{(l-1)\delta^{j-1}+1} \quad \lambda_{(l-1)\delta^{j-1}+2} \quad \dots \quad \lambda_{l\delta^{j-1}} \,\right].$$

Now, from (B.21) we see that each component of $(\partial/\partial Y_l) Y_{[j]}$ either is equal to zero or is equal (unless an *integer positive* coefficient) to some component of $Z^{[j-1]}$. Let $h$ be the position of a nonzero entry of $(\partial/\partial Y_l) Y_{[j]}$, and let $r \in \{1, \dots, \delta^{j-1}\}$ be a position for which it appears (unless a coefficient, and possibly repeated) in $Z^{[j-1]}$. Then it results that the $h$th row of $\tilde{L}$ has, possibly, nonzero (hence positive) elements in the set $\rho(r; \delta, j-1)$. Indeed, this set of positions is determined by the position $(r)$ of the component to be extracted in $Z^{[j-1]}$, endowed with all its $(\delta, j-1)$-R positions.

Let $i \in \{1, \dots, l\delta^{j-1}\}$ such that $\lambda_{(l-1)\delta^{j-1}+i}$ has a nonzero component, namely the $h$th. Then $\big(\lambda_{(l-1)\delta^{j-1}+i}\big)_k = 0$ for $k = 1, \dots, q_j$ and $k \neq h$. As a matter of fact, if $\big(\lambda_{(l-1)\delta^{j-1}+i}\big)_k \neq 0$, and $k \neq h$, then some monomial, equal to the $i$th, would be taken in $Z^{[j-1]}$, and hence we would have two equal components in $(\partial/\partial Y_l) Y_{[j]}$, which is impossible because $Y_{[j]}$ has no redundancies. This proves part (A) of the lemma.

From the above it follows that all the columns $\{\lambda_{(l-1)\delta^{j-1}+i},\ i \in \rho(r; \delta, j-1)\}$ have zero entries, but possibly one, placed in the same position $h$ for any $i \in \rho(r; \delta, j-1)$. Hence, they constitute a set of linearly dependent vectors. Part (B) of the lemma follows as soon as it is noticed that, using Lemma B.6 and taking into account (B.19), it results that $\{\lambda_{(l-1)\delta^{j-1}+i},\ i \in \rho(r; \delta, j-1)\} = \{\lambda_i,\ i \in \rho'(s; \delta, j)\}$.

Finally, in order to prove part (C), note that, since $l \leq q$ (and hence, by recalling the structure of $Z$, given by (7.2), it results that $Z_l = Y_l$), we have that the $s$th component of $Z^{[j]}$ is of the form $Y_l Z_1^{h_1} \cdots Z_\delta^{h_\delta}$, where the powers $h_1, \dots, h_\delta$ are such that $h_1 + \cdots + h_\delta = j-1$, and it is not a $Y$-monomial by the hypothesis. Hence the monomial $Z_1^{h_1} \cdots Z_\delta^{h_\delta}$ is not a $Y$-monomial of $Z^{[j-1]}$, and then it cannot belong to the left-hand side of (B.21). This in turn implies, again by (B.21), that the $r$th column of $\tilde{L}^{(l)}$ (that is the $s$th column of $L_\delta^j$, because $l, r$ are defined by (B.19)) must be zero. □

Before proving Theorem 8.1, we need to give the following definition.

DEFINITION B.11. *We define the* $(\delta, j)$*-Kronecker space, namely* $\mathbf{K}(\delta, j)$*, as the following subspace of* $\mathbf{R}^{\delta^j}$:

$$\mathbf{K}(\delta, j) = \operatorname{span}\left(\left\{z \in \mathbf{R}^{\delta^j} \;\Big/\; \exists x \in \mathbf{R}^\delta \text{ such that } z = x^{[j]}\right\}\right).$$

*Remark* B.12. From Definition B.11 it follows that

$$\mathbf{K}(\delta, j) = \left\{z \in \mathbf{R}^{\delta^j} \;\Big/\; z_r = z_s \textit{ if } (r,s) \textit{ is } (\delta, j) - R\right\}.$$

*Proof of Theorem* 8.1. By exploiting the definition of $\mathcal{D}_k^1$ given in (7.28), and the definition of $\tilde{\mathcal{B}}_k$ given by (7.17), we can rewrite the matrix $\mathcal{R}(t)$, defined in (8.1), as

$$\text{(B.22)} \quad \mathcal{R}(t) = \sum_{k=1}^{p} \mathcal{E}_\mathcal{Y} \mathcal{B}_k \Phi_Z(t) \mathcal{B}_k^T \mathcal{E}_\mathcal{Y}^T + \sum_{k=1}^{p} \mathcal{E}_\mathcal{Y} \big(\mathcal{B}_k m_{z}(t) + \mathcal{V}_k\big)\big(\mathcal{B}_k m_{z}(t) + \mathcal{V}_k\big)^T \mathcal{E}_\mathcal{Y}^T.$$

We will prove the theorem by showing that the matrix $\mathcal{E}_\mathcal{Y}\mathcal{B}_k\Phi_Z\mathcal{B}_k^T\mathcal{E}_\mathcal{Y}^T$ is uniformly nonsingular for some $k = 1, \ldots, p$, or (which is the same because $\Phi_Z(t)$ is uniformly nonsingular over $T$) that $\mathcal{E}_\mathcal{Y}\mathcal{B}_k$ is a full (row) rank matrix for some $k$.

In order to verify this, first of all note that, from Assumption 3.1 and Remark 3.2, it results that there exists a $\bar{k}$ such that $\operatorname{rank}(D_{\bar{k}}) = q$ (we remind the reader that $q$ is the dimension of the original observation $Y$). Indeed, we have

$$\text{(B.23)} \qquad D_{\bar{k}} = [I_q \; 0].$$

For such a $\bar{k}$, let us show that

$$\text{(B.24)} \qquad \operatorname{rank}\big(\mathcal{E}_\mathcal{Y}\mathcal{B}_{\bar{k}}\big) = q + q_2 + \cdots + q_\nu;$$

that is, it is a full (row) rank matrix (remember that $q_i$ is the dimension of $Y_{[i]}$). From the definition of $\mathcal{B}_k$ and $\mathcal{E}_\mathcal{Y}$, given in (7.14) and (7.20), respectively, using (7.10) and taking into account the block triangular structure of $\mathcal{B}_k$, it results that condition (B.24) is equivalent to:

$$\text{(B.25)} \qquad \operatorname{rank}\big(\mathcal{E}_\mathcal{Y}^1 \tilde{B}_{\bar{k}}\big) = q,$$

$$\text{(B.26)} \qquad \operatorname{rank}\big(\mathcal{E}_\mathcal{Y}^j \tilde{T}_\delta^j U_\delta^j (\tilde{B}_{\bar{k}} \otimes I_{\delta^{j-1}}) T_\delta^j\big) = q_j \quad \forall\, j = 2, \ldots, \nu.$$

Now, from (7.22) we see that $\mathcal{E}_\mathcal{Y}^1 \in \mathbf{R}^{q \times \delta}$, $\mathcal{E}_\mathcal{Y}^1 = [I_q \; 0]$. Hence, by the definition of $\tilde{B}_k$ given in (7.5) and taking into account (B.23), it results that $\mathcal{E}_\mathcal{Y}^1 \tilde{B}_{\bar{k}} = [0 \; I_q \; 0]$, and hence condition (B.25) is verified.

It remains to prove (B.26). In order to do this, first note that, from the definition of $\tilde{B}_k$ given in (7.5) and taking into account (B.23), we can consider the following partition of the matrix $\tilde{B}_{\bar{k}} \otimes I_{\delta^{j-1}}$:

$$\text{(B.27)} \qquad \tilde{B}_{\bar{k}} \otimes I_{\delta^{j-1}} = \begin{bmatrix} M_1 \\ M_2 \end{bmatrix},$$

where $M_1$ has dimensions $q\delta^{j-1} \times \delta^j$ and has the following structure:

$$M_1 = [\,0 \quad I_{q\delta^{j-1}} \quad 0\,], \tag{B.28}$$

where the first null-block has dimensions $q\delta^{j-1} \times q\delta^{j-1}$. Using Lemma B.9 and (B.27), we have

$$\mathcal{E}^j_{\mathcal{Y}} \tilde{T}^j_\delta U^j_\delta (\tilde{B}_{\bar{k}} \otimes I_{\delta^{j-1}}) T^j_\delta = [\, L^j_\delta \quad 0\,] \begin{bmatrix} M_1 \\ M_2 \end{bmatrix} T^j_\delta = L^j_\delta M_1 T^j_\delta. \tag{B.29}$$

Now note that the range of the *expansion* matrix $T^j_\delta$ is equal to the Kronecker space $\mathbf{K}(\delta, j)$ (we remind the reader that $T^j_\delta$ performs the operation $Z^{[j]} = T^j_\delta Z_{[j]}$). Then by (B.29), we have that (B.26) is implied by the following condition: the operator $L^j_\delta M_1 : \mathbf{R}^{\delta^j} \to \mathbf{R}^{q_j}$, restricted to $\mathbf{K}(\delta, j)$ is surjective.

Let $y \in \mathbf{R}^{q_j}$, and we will prove that there exists a $z \in \mathbf{K}(\delta, j)$ such that $y = L^j_\delta M_1 z$. By Lemma B.9, $\text{rank}(L^j_\delta) = q_j$; then there exist $q_j$ indexes. $1 \le i_1, i_2, \dots, i_{q_j} \le q\delta^{j-1}$, such that the columns $\lambda_{i_1}, \lambda_{i_2}, \dots, \lambda_{i_{q_j}}$ ($\lambda_i$ denotes as usual the $i$th column of $L^j_\delta$) are linearly independent. For every $i_s$, $s = 1, \dots, q_j$, let us consider the sets $\rho'(i_s; \delta, j) \subset \rho(i_s; \delta, j)$ defined in (B.2). Let us define $\bar{\lambda}_{i_s}$ as

$$\bar{\lambda}_{i_s} \triangleq \sum_{i \in \rho'(i_s;\delta,j)} \lambda_i. \tag{B.30}$$

From Lemma B.10, parts (A) and (B), we have that the set $\{\bar{\lambda}_{i_s},\ s = 1, \dots, q_j\}$ is a set of linearly independent vectors, and hence there exist real numbers $\alpha_{i_1}, \alpha_{i_2}, \dots, \alpha_{i_{q_j}}$, such that

$$y = \alpha_{i_1} \bar{\lambda}_{i_1} + \cdots + \alpha_{i_{q_j}} \bar{\lambda}_{i_{q_j}}. \tag{B.31}$$

Now let us show that the elements of the set $\{i_s + q\delta^{j-1}\ \ s = 1, \dots, q_j\}$ are pairwise $(\delta, j)$-NR. To this purpose, for any pair $(i_r, i_s)$, $r \ne s$, $r, s = 1, \dots, q_j$, we can distinguish the following two cases.

(i)

$$\left[\frac{i_r}{\delta^{j-1}}\right] = \left[\frac{i_s}{\delta^{j-1}}\right].$$

In this case, since $\lambda_{i_r}$ and $\lambda_{i_s}$ are linearly independent, it follows that $(i_r, i_s)$ is $(\delta, j)$-NR. Indeed, if $(i_r, i_s)$ were $(\delta, j)$-R, then Lemma B.10, part (B) would imply that $\lambda_{i_r}$ and $\lambda_{i_s}$ are linearly dependent vectors. Hence, since $(i_r, i_s)$ is $(\delta, j)$-NR, Lemma B.7 implies that $(i_r + q\delta^{j-1}, i_s + q\delta^{j-1})$ is $(\delta, j)$-NR.

(ii)

$$i'_r \triangleq \left[\frac{i_r}{\delta^{j-1}}\right] \ne \left[\frac{i_s}{\delta^{j-1}}\right] \triangleq i'_s. \tag{B.32}$$

In this case, if $(i_r, i_s)$ is $(\delta, j)$-R then Lemma B.8 directly implies the same conclusion of (i). Else, if $(i_r, i_s)$ is $(\delta, j)$-NR, then we can show that $(i_r + q\delta^{j-1}, i_s + q\delta^{j-1})$ is again $(\delta, j)$-NR. For, let $h_1, \dots, h_\delta$, $h'_1, \dots, h'_\delta$ such that $h_1 + \cdots + h_\delta = h'_1 + \cdots + h'_\delta = j - 1$ and

$$\begin{aligned} \left(Z^{[j]}\right)_{i_r} &= Z_{i'_r} Z_1^{h_1} \cdots Z_\delta^{h_\delta}, \\ \left(Z^{[j]}\right)_{i_s} &= Z_{i'_s} Z_1^{h'_1} \cdots Z_\delta^{h'_\delta}, \end{aligned} \tag{B.33}$$

where $i'_r, i'_s$ are given by (B.32). Since $\lambda_{i_r}$ and $\lambda_{i_s}$ are linearly independent (hence nonzero), Lemma B.10, part (C) implies that both the monomials in (B.33) are $Y$-monomials. If $(i_r + q\delta^{j-1}, i_s + q\delta^{j-1})$ were $(\delta, j)$-R, we should have

$$Z_{i'_r+q} Z_1^{h_1} \cdots Z_\delta^{h_\delta} = Z_{i'_s+q} Z_1^{h'_1} \cdots Z_\delta^{h'_\delta},$$

which is possible if and only if

$$\text{(B.34)} \qquad \begin{aligned} &h_{i'_r+q} = h'_{i'_r+q} - 1, \quad h_{i'_s+q} = h'_{i'_s+q} + 1, \\ &h_i = h'_i \quad \forall\, i \neq i'_r + q, i'_s + q. \end{aligned}$$

Now $i'_r, i'_s \leq q$, and then we have that $Z_{i'_r+q}$ and $Z_{i'_s+q}$ are not components of the vector $Y$, hence condition (B.34) can be verified if and only if both the monomials in (B.33) are not $Y$-monomials, which is a contradiction.

Since the elements of the set $\{i_s + q\delta^{j-1},\ s = 1, \ldots q_j\}$ are pairwise $(\delta, j)$-NR, we have that

$$\text{(B.35)} \quad \rho(i_s + q\delta^{j-1}; \delta, j) \cap \rho(i_r + q\delta^{j-1}; \delta, j) = \emptyset \quad \forall\, r, s = 1, \ldots, q_j,\ r \neq s.$$

From (B.35) it results that the following vector $z \in \mathbf{R}^{\delta^{j-1}}$ is well defined:

$$\text{(B.36)} \qquad z_l = \begin{cases} \alpha_{i_s} & \text{if } l \in \rho(i_s + q\delta^{j-1}; \delta, j), \\ 0 & \text{otherwise.} \end{cases}$$

Noting that, by construction, $z \in \mathbf{K}(\delta, j)$, the theorem is proven as soon as it is shown that $y = L_\delta^j M_1 z$ with $y$ given by (B.31).

Let $z' = M_1 z$. By the structure of the matrix $M_1$ (B.28), it follows that

$$\text{(B.37)} \qquad z'_l = \begin{cases} \alpha_{i_s} & \text{if } l \in \rho(i_s + q\delta^{j-1}; \delta, j) - q\delta^{j-1}, \\ 0 & \text{otherwise,} \end{cases}$$

where the definition of translated set, given by (B.8), has been used. Observing (B.37), (B.31), and the definition of the $\bar{\lambda}_{i_s}$s (B.30), we see that the equality $y = L_\delta^j z'$, and hence the theorem is implied by the condition

$$\text{(B.38)} \qquad \sum_{i \in \rho'(i_s; \delta, j)} \lambda_i = \sum_{i \in \rho(i_s + q\delta^{j-1}; \delta, j) - q\delta^{j-1}} \lambda_i \quad \forall\, s = 1, \ldots, q_j.$$

Now, from Lemma B.7 we have $\rho'(i_s; \delta, j) = \rho'(i_s + q\delta^{j-1}; \delta, j) - q\delta^{j-1}$; moreover, by (B.3)

$$\rho(i_s + q\delta^{j-1}; \delta, j) = \rho'(i_s + q\delta^{j-1}; \delta, j) \cup \rho''(i_s + q\delta^{j-1}; \delta, j),$$

and hence (B.38) becomes

$$\sum_{i \in \rho''(i_s + q\delta^{j-1}; \delta, j) - q\delta^{j-1}} \lambda_i = 0 \quad \forall\, s = 1, \ldots, q_j,$$

which is implied by

$$\text{(B.39)} \qquad \lambda_i = 0 \quad \forall\, i \in \rho''(i_s + q\delta^{j-1}; \delta, j) - q\delta^{j-1}.$$

In order to prove (B.39), first note that by Lemma B.8, for any $i \in \rho''(i_s+q\delta^{j-1};\delta,j)-q\delta^{j-1}$, we must have that $(i,i_s)$ is $(\delta,j)$-NR and such that $(i+q\delta^{j-1}, i_s+q\delta^{j-1})$ is $(\delta,j)$-R. Now, let $h_1,\dots,h_\delta$, $h'_1,\dots,h'_\delta$ such that $h_1+\cdots+h_\delta = h'_1+\cdots+h'_\delta = j-1$ and

$$\text{(B.40)} \qquad \begin{aligned} \left(Z^{[j]}\right)_i &= Z_{i'} Z_1^{h_1}\cdots Z_\delta^{h_\delta},\\ \left(Z^{[j]}\right)_{i_s} &= Z_{i'_s} Z_1^{h'_1}\cdots Z_\delta^{h'_\delta}, \end{aligned}$$

with $i' = [i/\delta^{j-1}]$, $i'_s = [i_s/\delta^{j-1}]$. Since $(i+q\delta^{j-1}, i_s+q\delta^{j-1})$ is $(\delta,j)$-R, it results that

$$Z_{i'+q} Z_1^{h_1}\cdots Z_\delta^{h_\delta} = Z_{i'_s+q} Z_1^{h'_1}\cdots Z_\delta^{h'_\delta},$$

which in turn implies the following condition:

$$\text{(B.41)} \qquad \begin{aligned} &h_{i'+q} = h'_{i'+q}-1, \quad h_{i'_s+q} = h'_{i'_s+q}+1,\\ &h_i = h'_i \quad \forall\ i \neq i'+q, i'_s+q. \end{aligned}$$

Since $i', i'_s \le q$, $Z_{i'+q}$ and $Z_{i'_s+q}$ are not components of the vector $Y$. Hence, condition (B.41) implies that both the monomials in (B.40) are not $Y$-monomials. In particular, since $\left(Z^{[j]}\right)_i$ is not a $Y$-monomial, Lemma B.10, part (C) gives $\lambda_i = 0$, that is (B.39). $\square$

## REFERENCES


[1] R. E. Kalman, *A new approach to linear filtering and prediction problems*, J. Basic. Engrg., 1 (1960), pp. 35–45.

[2] R. E. Kalman and R. S. Bucy, *New results in linear filtering and prediction theory*, Trans. ASME Ser. D. J. Basic Engrg., 83 (1961), pp. 95–108.

[3] R. S. Liptser and A. N. Shiryayev, *Statistics of Random Processes*, Vol. 2, Springer-Verlag, New York, 1978.

[4] R. R. Mohler, *Bilinear Control Processes*, Academic Press, New York, 1970.

[5] C. Bruni, G. Di Pillo, and G. Koch, *Bilinear systems: An appealing class of "nearly linear" systems in theory and applications*, IEEE Trans. Automat. Control, 19 (1974), pp. 334–348.

[6] C. Bruni and G. Koch, *Suboptimal filtering for linear-in-control systems with small observation noise*, in IV International Federation of Automatic Control Symposium on Identification and System Parameter Estimation, Tblisi, SSSR, Sept. 1976.

[7] X. Yang, R. R. Mohler, and R. M. Burton, *Adaptive suboptimal filtering of bilinear systems*, Internat. J. Control, 52 (1990), pp. 135–158.

[8] E. Yaz, *Full and reduced order observer design for discrete stochastic bilinear systems*, IEEE Trans. Automat. Control, 37 (1992), pp. 503–505.

[9] F. Carravetta, A. Germani, and M. Raimondi, *Polynomial filtering for linear discrete-time non-Gaussian systems*, SIAM J. Control Optim., 34 (1996), pp. 1666–1690.

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

[11] G. Kallianpur, *Stochastic Filtering Theory*, Springer-Verlag, New York, 1980.

[12] V. Krishnan, *Nonlinear Filtering and Smoothing*, John Wiley, New York, 1984.

[13] A. Germani and M. Piccioni, *Semi-discretization of stochastic partial differential equations on* $\mathbf{R}^d$ *by a finite-element technique*, Stochastics, 23 (1988), pp. 131–148.

[14] R. Bellman, *Introduction to Matrix Analysis*, McGraw-Hill, New York, 1970.

[15] G. S. Rodgers, *Matrix Derivatives*, Lecture Notes in Statist. 2, Marcel Dekker, New York, 1980.