



ISTITUTO DI ANALISI DEI SISTEMI ED INFORMATICA
CONSIGLIO NAZIONALE DELLE RICERCHE

F. Carravetta, A. Germani, M.K. Shuakayev

**A NEW SUBOPTIMAL APPROACH
TO THE FILTERING PROBLEM FOR
CUBIC-SENSOR-LIKE NONLINEAR SYSTEMS**

R. 489 Dicembre 1998

Francesco Carravetta - Istituto di Analisi dei Sistemi ed Informatica del CNR, Viale Manzoni 30, 00185 Roma, Italy. Email: *carravetta@iasi.rm.cnr.it*.

Alfredo Germani - Dipartimento di Ingegneria Elettrica, Università dell'Aquila, 67100 Monteluco (L'Aquila), Italy, and Istituto di Analisi dei Sistemi ed Informatica del CNR, Viale Manzoni 30, 00185 Roma, Italy. Email: *germani@iasi.rm.cnr.it*.

Marat Kapacevic Shuakayev - Department of Electrical Engineering, Kazakh National Technical University, 22, Satpaev Str., Almaty, Republic of Kazakhstan.

Istituto di Analisi dei Sistemi ed Informatica, CNR
viale Manzoni 30
00185 ROMA, Italy

tel. ++39-06-77161

fax ++39-06-7716461

email: iasi@iasi.rm.cnr.it

URL: <http://www.iasi.rm.cnr.it>

Abstract

The filtering problem for a nice class of nonlinear stochastic multivariable systems is here considered. The state equation is an Ito bilinear stochastic differential equation, whereas the output process is given by a linear transformation of the state powers, corrupted by multiplicative noise. This class includes bilinear stochastic systems and, moreover, it is a generalization of the well known “cubic sensor”, for which the nonexistence of finite-dimensional filters was proved. The aim of this paper is to present a new approach consisting in searching for suboptimal state-estimates instead of the conditional statistics. As a first result, a finite-dimensional optimal linear filter for the subclass of the bilinear stochastic systems is defined. Next, 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.

Key words: square integrable martingales, wide-sense Wiener processes, stochastic bilinear systems, cubic sensor, Kronecker algebra, Kalman-Bucy filtering, polynomial filtering, vector Ito formula

1. Introduction

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

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

$$dY(t) = C(t)(X(t))^{[\mu]}dt + B^2(X(t), dW(t)), \quad (1.2)$$

where $X(t) \in \mathbb{R}^n$; $Y(t) \in \mathbb{R}^q$; $\mu \in \mathbb{N}$; $X^{[\mu]}$ denotes the Kronecker power (see appendix A) of the vector X ; $W(t) \in \mathbb{R}^p$, is a standard Wiener process with respect to some increasing family of σ -algebras, namely $\{\mathcal{F}_t\}$; $A(t), C(t)$ are matrices of proper dimensions; B^1, B^2 are bilinear forms.

We will call system (1.1), (1.2) a *cubic-sensor-like system*. This name reminds that, in the scalar case, for $B^1 = B^2 = dW$ and $\mu = 3$, system (1.1), (1.2) becomes the well known *cubic sensor* [1-4].

The problem we are faced with consists in searching for finite-dimensional filters for the cubic-sensor-like system (1.1), (1.2). We stress that, the class of systems described by (1.1), (1.2), includes also as a particular case, for $\mu = 1$, the well known bilinear stochastic systems (BLSS) [5-9], for which the 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

$$dz_t = f(z_t)dt + g(z_t)dY_t, \quad (1.3)$$

endowed with an output transformation:

$$\hat{X}_t = h(z_t), \quad (1.4)$$

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

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

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

As shown in [3], for the cubic sensor an optimal finite dimensional filter in the form (1.3) does not exist. Indeed the optimal filter is an infinite-dimensional one. Of course, from an application point of view, it becomes crucial to look for finite-dimensional approximations of the optimal filter.

This problem is considered in [4] where, an approximation of the Zakai equation for the unnormalized conditional density is built up under the hypothesis of high signal to noise ratio. However, this method cannot be easily generalizable to the vector case and, in any case, it does not yield the state-estimate directly through a finite-dimensional stochastic differential equation as (1.3), (1.4). In the same paper, a finite-dimensional filter, strongly based on the same assumption of low noise, is defined by using the classical Kushner's equation and an error bound is also given. However, the error bound becomes not meaningful when the signal-to-noise ratio (SNR) decreases. Indeed, the case of low SNR is the most meaningful from an engineering point of view.

In this paper we will derive, as an auxiliary result, the *optimal linear filtering* equations for a BLSS with linear observations corrupted by multiplicative measurement noise, which will result in the finite-dimensional form (1.3), (1.4). We point out that in [12] 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. 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.

Given the situation above described, that is nonexistence or unknowledge of the finite-dimensional optimal filter for system (1.1), (1.2), it is of a great interest from an application point of view to search for *suboptimal* filters showing a *better performance* with respect to the linear one.

This suboptimal approach has been recently developed for discrete-time systems in [10-11] where a general *polynomial filter* of any arbitrarily fixed degree is defined for linear non-Gaussian systems [10] and bilinear systems [11]. 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 noises distributions are very far from Gaussianity.

In this paper we will adopt an extension of this suboptimal approach to continuous time processes which will allow us to define a finite-dimensional filter in the form (1.3), (1.4), producing 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)* definition of a class of polynomial estimators;
- ii)* the problem of finding the optimal filter for the cubic-sensor-like system 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)* Application of a Kalman-Bucy scheme to the augmented system. This will give us the required polynomial filter.

The paper is organized as follows.

§2 deals with point *i)*. In §3 the overall setup of the problem is presented. §4,5,6 are concerned with some preliminary results. In particular, in §4 a method for transforming a vector BLSS in a linear system with WSW diffusions is presented. In §5 a vector Ito's formula is defined, by using the Kronecker formalism. In §6 a general formula, defining the stochastic differential of the Kronecker power of some process, solution of a bilinear SDE, is found. In §7 point *ii)* is treated. Finally, in §8 the complete solution of the problem is presented, resulting in a system of equations which define a polynomial filter (of an arbitrarily fixed degree) for the cubic-sensor-like system. In §9 the above theory is specialized for the classical scalar cubic sensor. Two Appendices are included in order to make more readable the paper.

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 about linear filtering [12] which will be widely used in the following.

Let I 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 $\mathcal{L}'_t(\xi)$:

$$\mathcal{L}'_t(\xi) \triangleq \left\{ \lambda \in L^2 : \exists j \in \mathbb{N}, \exists t_1, \dots, t_j \in I, t_1 \leq \dots \leq t_j \leq t, \right. \\ \left. \exists \text{ matrices } M_{t_1}, \dots, M_{t_j}, \exists \text{ a vector } b, \text{ such that } \lambda = \sum_{i=1}^j M_{t_i} \xi_{t_i} + b \right\}.$$

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 η we can define the *optimal linear estimate of η given $\{\xi_s, s \leq t\}$* as $\Pi(\eta/\mathcal{L}_t(\xi))$. Now, suppose there exists an integer ν such that

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

then, we can give the following definition.

Definition 2.1. We call ν -th degree polynomial estimate of η 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 η among all estimates, namely λ , that are either in the form:

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

for such a $k \in \mathbb{N}$, $t_1, \dots, t_k \in I$, $t_1 \leq \dots \leq t_k$, for such a vector b and matrices $M_{i,j}$, $i, j = 1, \dots, k$, or are mean square limits of these. It 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: $\hat{\eta}^{(\nu)} = \Pi(\eta/\mathcal{P}_t^{(\nu)}(\xi))$, $\hat{\eta}^{(\nu+1)} = \Pi(\eta/\mathcal{P}_t^{(\nu+1)}(\xi))$ it results

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

that is, the estimation quality is not decreasing for increasing ν .

Now, the aim of this paper can be expressed in a more precise manner as follows: for any given ν find a finite-dimensional filter in the form (1.3), (1.4), such that $\hat{X}(t)$ agrees with the optimal ν -th degree polynomial estimate of the state of the cubic-sensor-like system (1.1), (1.2). Such a filter will be referred in the following as ν -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 [12, ch. 15] for a detailed discussion with proofs).

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

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

where $t \in [0, t_M]$, $X(t) \in \mathbb{R}^n$, $Y(t) \in \mathbb{R}^q$, $A(t), C(t), B_i(t), D_i(t)$, $i = 1, \dots, m$, are suitably dimensioned matrices, \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 [12, Th. 15.3]

$$\begin{aligned} d\hat{X}(t) &= A(t)\hat{X}(t)dt \\ &+ \left(\sum_{i=1}^m B_i(t)D_i(t)^T + P(t)C(t)^T \right) \left(\sum_{i=1}^m D_i(t)D_i(t)^T \right)^{-1} (dY(t) - C(t)\hat{X}(t)dt), \\ \frac{dP(t)}{dt} &= A(t)P(t) + P(t)A(t)^T + \sum_{i=1}^m D_i(t)D_i(t)^T \\ &- \left(\sum_{i=1}^m B_i(t)D_i(t)^T + P(t)C(t)^T \right) \left(\sum_{i=1}^m D_i(t)D_i(t)^T \right)^{-1} \left(\sum_{i=1}^m B_i(t)D_i(t)^T + P(t)C(t)^T \right)^T, \\ \hat{X}(0) &= E(\bar{X}), \quad P(0) = E\left((\bar{X} - E(\bar{X}))(\bar{X} - E(\bar{X}))^T \right), \end{aligned} \tag{2.2}$$

where $P(t)$ represents the filtering error covariance matrix. Note that in equations (2.2) the nonsingularity of the matrix function $\sum_i D_i(t)D_i(t)^T$ over the time interval $[0, t_M]$ is required.

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

3. The system to be filtered

Let $T = [0, t_M]$, (Ω, \mathcal{F}, P) be a probability triple and $\{\mathcal{F}_t\}$, $t \in T$, be a family of non decreasing sub σ -algebras of \mathcal{F} . Moreover let $(W(t), \mathcal{F}_t)$ be an \mathbb{R}^p -valued standard Wiener process, $\mu \geq 1$ a given integer, and $\bar{X} \in \mathbb{R}^n$ an \mathcal{F}_0 -measurable random variable such that there exists an integer $\nu \geq 1$ for which:

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

For the random variable \bar{X} we suppose to know the moments $m_{\bar{X}}^{(i)}$:

$$m_{\bar{X}}^{(i)} \triangleq E(\bar{X}^{[i]}), \quad i = 1, \dots, 2\nu\mu. \quad (3.1)$$

Let us consider the stochastic system:

$$dX(t) = A(t)X(t)dt + H(t)u(t)dt + \sum_{k=1}^p (B_k X(t) + F_k) dW_k(t), \quad (3.2)$$

$$dY(t) = C(t)(X(t))^{[\mu]} dt + \sum_{k=1}^p (D_k X(t) + G_k) dW_k(t), \quad (3.3)$$

where $A(t) \in \mathbb{R}^{n \times n}$, $C(t) \in \mathbb{R}^{q \times n^\mu}$, $H(t) \in \mathbb{R}^{n \times m}$, $B_k \in \mathbb{R}^{n \times n}$, $F_k \in \mathbb{R}^n$, $D_k \in \mathbb{R}^{q \times n}$, $G_k \in \mathbb{R}^q$, for $k = 1, \dots, p$, $W_k(t)$ denotes the k -th component of the standard Wiener process $W(t) \in \mathbb{R}^p$, $u(t) \in \mathbb{R}^m$ is a deterministic input. Eq. (3.2) is endowed with the initial condition $X(0) = \bar{X}$. In the following we shall denote with I_α , $\alpha = 0, 1, \dots$, the $\alpha \times \alpha$ identity matrix; we conventionally 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

$$D_{\bar{k}} = [I_q \ 0]. \quad (3.4)$$

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

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

where $R \in \mathbb{R}^{(n-q) \times n}$ 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 assure 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 (1.3), (1.4), such that:

$$\hat{X}(t) = \Pi \left(X(t) / \mathcal{P}_t^{(\nu)}(Y) \right), \quad (3.5)$$

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 a BLSS for which the optimal linear filtering problem is equivalent to the original polynomial filtering problem for system (3.2), (3.3). At this purpose, in the next two sections we state some preliminary results. The first one concerns the definition of the optimal linear filter for a BLSS.

4. Optimal linear filtering for BLSS's

Let us consider system (3.2), (3.3), with $\mu = 1$ (hence, a BLSS). Moreover let $\Psi_X(t)$ the covariance matrix of the state process: $\Psi_X(t) = E((X(t) - E(X(t)))(X(t) - E(X(t)))^T)$. The problem of finding a finite-dimensional *optimal linear* filter for the BLSS (3.2), (3.3) was up to now unsolved in the general case [12]. 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 [19], solves the optimal linear filtering problem in this case.

Let $M \in \mathbb{R}^{\alpha \times \alpha}$ a symmetric positive semidefinite matrix, such that $\text{rank}(M) = \rho \leq \alpha$. As well known, there exists a full rank matrix $N \in \mathbb{R}^{\alpha \times \rho}$ such that $NN^T = M$. We will use the following notation:

$$M^{(\frac{1}{2})} \triangleq N,$$

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.

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

$$\begin{aligned} \rho_k &\triangleq \text{rank} \left\{ B_k \cdot \Psi_X(t) \cdot B_k^T \right\}, \\ \sigma_k &\triangleq \text{rank} \left\{ D_k \cdot \Psi_X(t) \cdot D_k^T \right\}, \end{aligned} \quad \forall t \in T. \quad (4.1)$$

Then there exists the following representation:

$$dX(t) = A(t)X(t)dt + H(t)u(t) + \sum_{k=1}^{2p} \tilde{B}_k(t)d\tilde{W}_{k,1}(t), \quad X_0 = \bar{X} \quad (4.2)$$

$$dY(t) = C(t)X(t)dt + \sum_{k=1}^{2p} \tilde{D}_k(t)d\tilde{W}_{k,2}(t), \quad (4.3)$$

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

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

$$\tilde{D}_k(t) \triangleq \left(D_k \cdot \Psi_X(t) \cdot D_k^T \right)^{(\frac{1}{2})}, \quad (4.5)$$

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

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

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

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 $k = 1, \dots, p$, $\tilde{W}_{k,1}(t) \in \mathbb{R}^{\rho_k}$, $\tilde{W}_{k,2}(t) \in \mathbb{R}^{\sigma_k}$; for $k = p+1, \dots, 2p$:

$$\tilde{W}_{k,1}(t) = \tilde{W}_{k,2}(t) = W_{k-p}(t), \quad (4.9)$$

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

$$\tilde{W}_{k,1}(t) = \int_0^t (\tilde{B}_k(\tau)^T \tilde{B}_k(\tau))^{-1} \tilde{B}_k(\tau)^T B_k (X(\tau) - E(X(\tau))) dW_k(\tau), \quad (4.10)$$

$$\tilde{W}_{k,2}(t) = \int_0^t (\tilde{D}_k(\tau)^T \tilde{D}_k(\tau))^{-1} \tilde{D}_k(\tau)^T D_k (X(\tau) - E(X(\tau))) dW_k(\tau), \quad (4.11)$$

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 $s < t$:

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

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

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

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):

$$\tilde{B}_k(t) d\tilde{W}_{k,1}(t) = B_k (X(t) - E(X(t))) dW_k(t), \quad (4.12)$$

$$\tilde{D}_k(t) d\tilde{W}_{k,2}(t) = D_k (X(t) - E(X(t))) dW_k(t). \quad (4.13)$$

From the hypotheses it results well defined the symmetric positive-definite matrix $\Psi(t)^{1/2}$. Hence, for any $y(t) \in \mathbb{R}^n$ we can define $\bar{y}(t) \in \mathbb{R}^n$ 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

$$\bar{y}_1(t) \in \mathcal{R}(\Psi_X(t)^{1/2} B_k^T), \quad \bar{y}_2(t) \in \left\{ \mathcal{R}(\Psi_X(t)^{1/2} B_k^T) \right\}^- = \mathcal{N}(B_k \Psi_X(t)^{1/2}), \quad (4.14)$$

where $\mathcal{N}(M)$, $\mathcal{R}(M)$ denote the null-space and the range respectively of a matrix M . Using (4.14) and choosing a $\bar{z}(t)$ such that $\bar{y}_1(t) = \Psi_X(t)^{1/2} B_k \bar{z}(t)$, we have

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

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

$$B_k y(t) = \tilde{B}_k(t) z(t), \quad \forall t \in T. \quad (4.15)$$

10.

Equality (4.15) implies that, for any $y(t)$ we have

$$\begin{aligned}\tilde{B}_k(t)\left(\tilde{B}_k(t)^T\tilde{B}_k(t)\right)^{-1}\tilde{B}_k(t)^TB_ky(t) &= \tilde{B}_k(t)\left(\tilde{B}_k(t)^T\tilde{B}_k(t)\right)^{-1}\tilde{B}_k(t)^TB_kz(t) \\ &= \tilde{B}_k(t)z(t) = B_ky(t),\end{aligned}$$

from which, using the definition of $\tilde{W}_{k,1}$ given by (4.10) equality (4.12) follows. A similar argument can be used to prove (4.13).

Finally, by adding and subtracting the state-expectation $E(X(t))$, in the bilinear forms of (3.2), (3.3) and taking into account of (4.12), (4.13), 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 (4.9), for $p + 1 \leq k' \leq 2p$, $k'' \neq k' - p$:

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

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

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

Similarly, it is possible to show that $E(\tilde{W}_{k'',2}(t)\tilde{W}_{k',2}(t)^T) = 0$ for $p + 1 \leq k' \leq 2p$. ■

Remark 4.2. Although Theorem 4.1 is stated for a BLSS, nevertheless the hypothesis $\mu = 1$ in (3.3) is not essential. Indeed, the Theorem gives a way to transform a bilinear form of X and of a Wiener process into a linear form of a WSW process. Hence, for $\mu > 1$ we would obtain the same representation (4.2), (4.3), where in (4.3) we should replace $X(t)$ with $X(t)^{[\mu]}$.

In the following theorem it will be given a sufficient condition which guarantees the non singularity of $\Phi_X(t)$. Let us consider a time-invariant version of the BLSS given by (3.2), (3.3) with $\mu = 1$:

$$dX(t) = AX(t)dt + Hu(t)dt + \sum_{k=1}^p B_k X(t)dW_k(t) + FdN(t), \quad X(t_0) = \bar{X} \quad (4.16)$$

$$dY(t) = CX(t)dt + \sum_{k=1}^p D_k X(t)dW_k(t) + GdN(t), \quad (4.17)$$

where $t_0 \in \mathbb{R}$ is an “initial time”, and we have introduced another Wiener process $(N(t), \mathcal{F}_t)$ uncorrelated with W . We suppose that the system (4.16), (4.17) is well defined over the time interval $[t_0, \infty)$.

Theorem 4.3. *Let the matrix $\Psi_X(t_0)$ be nonsingular (or the pair (A, F) of the state equation (4.16) be controllable), then the state-covariance matrix $\Psi_X(t)$ is non singular for any $t \geq t_0$, ($t > 0$).*

Proof. The following equation is easily recognized to hold for $\Psi_X(t)$:

$$\begin{aligned} \Psi_X(t) &= e^{A(t-t_0)}\Psi_X(t_0)e^{A^T(t-t_0)} + \sum_{k=1}^p \int_{t_0}^t e^{A(t-\tau)}B_k E(X(\tau)X(\tau)^T)B_k^T e^{A^T(t-\tau)}d\tau \\ &\quad + \int_{t_0}^t e^{A(t-\tau)}FF^T e^{A^T(t-\tau)}d\tau. \end{aligned} \quad (4.18)$$

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 $\phi_X(t_0)$ and the hypothesis of controllability imply the positive definiteness of the first and the third term respectively. ■

Remark 4.4. Note that, when theorem 4.3 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 *uniformly nonsingular in T*).

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

Theorem 4.5. *Let be given the time-invariant BLSS as defined in equations (4.16), (4.17). Let the hypotheses of theorem 4.3 be satisfied. Moreover, let us suppose that*

$$\text{rank}(D_k) = q \quad \text{for such a } k, \text{ or } \text{rank}(G) = q. \quad (4.28)$$

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

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

satisfy the following system of equations:

$$\frac{dm(t)}{dt} = Am(t) + Hu(t), \quad m(0) = \bar{m}, \quad (4.19)$$

$$\begin{aligned} \frac{d\Psi_X(t)}{dt} &= A\Psi_X(t) + \Psi_X(t)A^T + \sum_{k=1}^p B_k \Psi_X(t)B_k^T + FF^T \\ &\quad + \sum_{k=1}^p B_k m(t)m(t)^T B_k^T, \quad \Psi_X(0) = \bar{\Psi}_X, \end{aligned} \quad (4.20)$$

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

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

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

$$\begin{aligned} \frac{dP_t}{dt} &= AP(t) + P(t)A^T + R(t) \\ &\quad - \left(\sum_{i=1}^{2p+1} \tilde{B}_i(t)\tilde{D}_i(t)^T + P(t)C^T \right) R(t)^{-1} \left(\sum_{i=1}^{2p+1} \tilde{B}_i(t)\tilde{D}_i(t)^T + P(t)C^T \right)^T, \end{aligned} \quad (4.24)$$

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

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

$$\hat{X}(0) = \bar{m}, \quad (4.27)$$

Proof. (4.19) readily derives by taking the expectations of both sides of (4.16). Moreover (4.20) is easily obtained by differentiating eq. (4.18). From Theorem 4.3 and Remark 4.4 $\Psi_X(t)$ is uniformly nonsingular in T . Then, we can apply theorem 4.1 in order to put system (4.16), (4.17), in the form of a linear stochastic system with suitable WSW state and output diffusions, deriving from eq.s (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 (4.16), (4.17). Now from (4.22), (4.23) it results:

$$R(t) \triangleq \sum_{k=1}^p D_k \Psi_X(t) D_k^T + \sum_{k=1}^p D_k m(t) m(t)^T D_k^T + GG^T$$

which is uniformly nonsingular in T , by the hypothesis (4.28) (and possibly by the uniform nonsingularity of $\Psi_X(t)$). The thesis easily derives from an application of [12, Th.15.3] to the representation (4.2), (4.3). ■

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 ϕ , the stochastic differential of the process $\phi^{[h]}$, where $[h]$ is any integer Kronecker power.

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

$$\frac{d}{dx} \otimes F(x) \triangleq \left[\frac{\partial F(x)}{\partial x_1} \quad \dots \quad \frac{\partial F(x)}{\partial x_n} \right], \quad (5.1)$$

where the operator d/dx is given by

$$\frac{d}{dx} \triangleq \left[\frac{\partial}{\partial x_1} \quad \dots \quad \frac{\partial}{\partial x_n} \right]. \quad (5.2)$$

Note that in (5.1) the rules defining the Kronecker 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:

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

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, namely $V(x)$ and $W(x)$, having suitable dimensions, it results

$$\frac{d}{dx} \otimes (V(x) \otimes W(x)) = \left(\frac{d}{dx} \otimes V(x) \right) \otimes W(x) + V(x) \otimes \left(\frac{d}{dx} \otimes W(x) \right). \quad (5.3)$$

$$\frac{d}{dx} \otimes (MW(x)) = M \left(\frac{d}{dx} \otimes W(x) \right). \quad (5.4)$$

Moreover, the following “associative” property holds:

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

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 $x \in \mathbb{R}^n$, it results*

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

and for any $h > 1$:

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

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

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

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

Proof. According to definition (5.1) and using (3.3), we have

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

that is

$$Q^h = I_n \otimes x^{[h-1]} + x \otimes Q^{(h-1)}, \quad (5.8)$$

from which, using theorem A.3, we obtain

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

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

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

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

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 \mathbb{R}^n described by the Ito's stochastic differential:*

$$dX_t = d\beta_t + dM_t, \quad (5.9)$$

where (β_t, \mathcal{F}_t) is an a.s. continuous bounded variation process and (M_t, \mathcal{F}_t) is a square integrable martingale. Let

$$F : \mathbb{R}^n \rightarrow \mathbb{R}^p,$$

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

$$dZ_t = \left(\frac{d}{dx} \otimes F(x) \right)_{x=X_t} dX_t + \frac{1}{2} \left(\frac{d}{dx} \otimes \frac{d}{dx} \otimes F(x) \right)_{x=X_t} (dM_t)^{[2]}, \quad (5.10)$$

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

$$(dM_t)^{[2]} = \begin{bmatrix} d \langle M_1, M_1 \rangle_t \\ d \langle M_1, M_2 \rangle_t \\ \vdots \\ d \langle M_n, M_n \rangle_t \end{bmatrix}, \quad (5.11)$$

with obvious meaning of symbols [12-14].

Proof. Formula (5.11) can be directly verified by using Ito formula in the scalar case (see for instance [13, Th. 4.2.1]) and 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 $\phi(t) \in \mathbb{R}^d$ the process defined by the following SDE:*

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

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

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

it results, for $i \geq 2$:

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

where

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

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

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

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

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

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

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

By exploiting the definition (5.11) it results

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

where Φ_2, Φ_1, Φ_0 are given by (6.2). By substituting (6.5) in (6.4) and using formula (6.3), the thesis follows. ■

7. The augmented system

Let us return to consider the cubic-sensor-like system (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, represents a solution of 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).

First of all, we define the process X_e as

$$X_e(t) \triangleq \begin{bmatrix} X(t) \\ X^{[2]}(t) \\ \vdots \\ X^{[\nu]}(t) \end{bmatrix}, \quad (7.1)$$

and call it the *extended state process*. By using Theorem 6.1 for $\phi = X(t)$, we can derive a SDE for the processes $X^{[i]}$, $i = 2, \dots, \nu$:

$$\begin{aligned} dX^{[i]}(t) &= (M_i^0(t)X^{[i]}(t) + M_i^1(t)X^{[i-1]}(t) + M_i^2X^{[i-2]}(t))dt \\ &+ \sum_{k=1}^p (G_{k,i}^0X^{[i]}(t) + G_i^1X^{[i-1]}(t))dW_k(t), \end{aligned} \quad (7.2)$$

where

$$\begin{aligned} M_i^0(t) &= U_n^i(A(t) \otimes I_{n^{i-1}}) + \frac{1}{2}O_n^i(\Psi_2 \otimes I_{n^{i-2}}) \\ M_i^1(t) &= U_n^i((H(t)u(t)) \otimes I_{n^{i-1}}) + \frac{1}{2}O_n^i(\Psi_1 \otimes I_{n^{i-2}}) \\ M_i^2 &= \frac{1}{2}O_n^i(\Psi_0 \otimes I_{n^{i-2}}) \\ G_{k,i}^0 &= U_n^i(B_k \otimes I_{n^{i-1}}) \\ G_{k,i}^1 &= U_n^i(F_k \otimes I_{n^{i-1}}), \end{aligned} \quad (7.3)$$

and Ψ_2, Ψ_1, Ψ_0 are given by

$$\Psi_2 \triangleq \sum_{k=1}^p B_k^{[2]}, \quad \Psi_1 \triangleq \sum_{k=1}^p (B_k \otimes F_k + F_k \otimes B_k), \quad \Psi_0 \triangleq \sum_{k=1}^p F_k^{[2]}.$$

By aggregating eq (3.2) and eq. (7.2) for $i = 2, \dots, \nu$, rewriting eq (3.3) so to enhance the dependance of X_e , we can write the equations of an *extended system* as stated in the following proposition.

Proposition 7.1. *The processes X_e and Y defined in (7.1), (3.3) are the state and output processes respectively, of the following bilinear stochastic differential system (extended system):*

$$dX_e(t) = (A_e(t)X_e(t) + \alpha(t))dt + \sum_{k=1}^p (B_{e,k}X_e(t) + \beta_k)dW_k(t)$$

$$dY(t) = C_e(t)X_e(t)dt + \sum_{k=1}^p (D_{e,k}X_e(t) + G_k)dW_k(t)$$

where

$$A_e(t) = \begin{bmatrix} A(t) & 0 & \dots & 0 \\ M_2^1(t) & M_2^0(t) & \ddots & \\ M_3^2 & M_3^1(t) & M_3^0(t) & \vdots \\ \vdots & \ddots & \ddots & \ddots & 0 \\ 0 & \dots & M_\mu^2 & M_\mu^1(t) & M_\mu^0(t) \end{bmatrix};$$

$$B_{e,k} = \begin{bmatrix} B_k & 0 & \dots & 0 \\ G_{k,2}^1 & G_{k,2}^0 & & \vdots \\ \vdots & \ddots & \ddots & \\ 0 & \dots & G_{k,\mu}^1 & G_{k,\mu}^0 \end{bmatrix}; \quad \alpha(t) = \begin{bmatrix} H(t)u(t) \\ M_2^2 \\ 0 \\ \vdots \\ 0 \end{bmatrix}; \quad \beta_k = \begin{bmatrix} F_k \\ 0 \\ \vdots \\ 0 \end{bmatrix};$$

$$C_e(t) = [0 \quad \dots \quad 0 \quad C(t)]; \quad D_{e,k} = [D_k \quad 0 \quad \dots \quad 0]. \quad (7.4)$$

The block-entries of matrices $A_e(t)$, $B_{e,k}$ are defined in (7.3).

The definition of the extended system allows us to write the observed process Y as the output of a bilinear system instead of the ordinary nonlinear system (3.2), (3.3). However, the extended state process, as is defined in (7.1), collects all the products between powered components of the ordinary state X (up to the μ -th degree). This implies the presence of redundant scalar equations in the vector extended system defined in Proposition 3.1. This in turn implies the following:

- a) The dimension of the extended system is uselessly great;
- b) some non-degeneracy conditions involved in the filtering problem are not satisfied (this point will be more clear later).

Nevertheless it is possible to overcome these troubles by writing down the equations of a new *reduced extended system*, as explained below.

Let $x \in \mathbb{R}^d$ and h a positive integer. We recall that, the following relations hold, linking together the *reduced h -th Kronecker power of x* [11], [16], namely $x_{[h]}$ and the (ordinary) h -th Kronecker power $x^{[h]}$:

$$x^{[h]} = T_d^h x_{[h]}; \quad x_{[h]} = \tilde{T}_d^h x^{[h]}, \quad (7.5)$$

where T_d^h and \tilde{T}_d^h are suitably dimensioned transformation matrices [11].

Now, let us define the vector $X_e^{(r)}(t)$ as:

$$X_e^{(r)}(t) \triangleq \begin{bmatrix} X(t) \\ X_{[2]}(t) \\ \vdots \\ X_{[\mu]}(t) \end{bmatrix}, \quad (7.6)$$

18.

then the following relations are easily recognized to hold:

$$X_e^{(r)}(t) = \tilde{R}X_e(t); \quad X_e(t) = RX_e^{(r)}(t), \quad (7.7)$$

where

$$\tilde{R} \triangleq \begin{bmatrix} I_n & 0 & \dots & 0 \\ 0 & \tilde{T}_n^2 & & \vdots \\ \vdots & & \ddots & \\ & \dots & & \tilde{T}_n^\mu \end{bmatrix}; \quad R \triangleq \begin{bmatrix} I_n & 0 & \dots & 0 \\ 0 & T_n^2 & & \vdots \\ \vdots & & \ddots & \\ & \dots & & T_n^\mu \end{bmatrix}.$$

The above defined process $X_e^{(r)}$ agree with the extended state process X_e , but the redundancy contained therein is eliminated. Using relations (7.7) and the extended system equations we obtain the following reduced extended system

$$\begin{aligned} dX_e^{(r)}(t) &= (\tilde{R}A_e(t)R)X_e^{(r)}(t) + \tilde{R}\alpha(t)dt + \sum_{k=1}^p ((\tilde{R}B_{e,k}R)X_e^{(r)}(t) + \tilde{R}\beta_k)dW_k(t) \\ dY(t) &= (C_e(t)R)X_e^{(r)}(t)dt + \sum_{k=1}^p (D_{e,k}X_e^{(r)}(t) + G_k)dW_k(t). \end{aligned} \quad (7.8)$$

Now, let us define the process Z :

$$Z(t) \triangleq \begin{bmatrix} Y(t) \\ X_e^{(r)}(t) \end{bmatrix}, \quad (7.9)$$

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

$$\mathcal{Z}(t) \triangleq \begin{bmatrix} Z(t) \\ Z_{[2]}(t) \\ \vdots \\ Z_{[\nu]}(t) \end{bmatrix}. \quad (7.10)$$

We can derive a SDE for the process \mathcal{Z} in the following way. First of all note that, from (7.8), Z satisfies the following SDE:

$$dZ(t) = (\tilde{A}(t)Z(t) + \tilde{\alpha}(t))dt + \sum_{k=1}^p (\tilde{B}_k Z(t) + \tilde{\beta}_k)dW_k(t), \quad (7.11)$$

where

$$\tilde{A}(t) \triangleq \begin{bmatrix} 0 & C_e(t)R \\ 0 & \tilde{R}A_e(t)R \end{bmatrix}; \quad \tilde{\alpha}(t) \triangleq \begin{bmatrix} 0 \\ \tilde{R}\alpha(t) \end{bmatrix}; \quad \tilde{B}_k \triangleq \begin{bmatrix} 0 & D_{e,k} \\ 0 & \tilde{R}B_{e,k}R \end{bmatrix}; \quad \tilde{\beta}_k = \begin{bmatrix} G_k \\ \tilde{R}\beta_k \end{bmatrix}. \quad (7.12)$$

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

$$dZ^{[i]}(t) = (L_i^0(t)Z^{[i]}(t) + L_i^1(t)Z^{[i-1]}(t) + L_i^2(t)Z^{[i-2]}(t))dt + \sum_{k=1}^p (V_{k,i}^0 Z^{[i]}(t) + V_{k,i}^1 Z^{[i-1]}(t))dW_k(t), \quad (7.13)$$

where

$$L_i^0(t) = U_\delta^i(\tilde{A}(t) \otimes I_{\delta^{i-1}}) + \frac{1}{2}O_\delta^i(\tilde{\Psi}_2 \otimes I_{\delta^{i-2}}) \quad (7.14)$$

$$L_i^1(t) = U_\delta^i(\tilde{\alpha}(t) \otimes I_{\delta^{i-1}}) + \frac{1}{2}O_\delta^i(\tilde{\Psi}_1 \otimes I_{\delta^{i-2}}) \quad (7.15)$$

$$L_i^2 = \frac{1}{2}O_\delta^i(\tilde{\Psi}_0 \otimes I_{\delta^{i-2}}) \quad (7.16)$$

$$V_{k,i}^0 = U_\delta^i(\tilde{B}_k \otimes I_{\delta^{i-1}}) \quad (7.17)$$

$$V_{k,i}^1 = U_\delta^i(\tilde{\beta}_k \otimes I_{\delta^{i-1}}), \quad (7.18)$$

and $\tilde{\Psi}_2, \tilde{\Psi}_1, \tilde{\Psi}_0$ are given by

$$\tilde{\Psi}_2 \triangleq \sum_{k=1}^p \tilde{B}_k^{[2]}, \quad \tilde{\Psi}_1 \triangleq \sum_{k=1}^p (\tilde{B}_k \otimes \tilde{\beta}_k + \tilde{\beta}_k \otimes \tilde{B}_k), \quad \tilde{\Psi}_0 \triangleq \sum_{k=1}^p \tilde{\beta}_k^{[2]}.$$

Observing that, from (7.5) we have

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

and using (7.13), we can state the following proposition.

Proposition 7.2. *The process \mathcal{Z} defined in (7.10) satisfies the following bilinear SDE,*

$$d\mathcal{Z}(t) = (\mathcal{A}(t)\mathcal{Z}(t) + \mathcal{U}(t))dt + \sum_{k=1}^p (\mathcal{B}_k \mathcal{Z}(t) + \mathcal{V}_k) dW_k(t), \quad (7.19)$$

where

$$\mathcal{A}(t) = \begin{bmatrix} \tilde{A}(t) & 0 & \dots & 0 \\ L_2^1(t) & \tilde{T}_\delta^2 L_2^0(t) T_\delta^2 & 0 & \\ L_3^2 & \tilde{T}_\delta^3 L_3^1(t) T_\delta^2 & \tilde{T}_\delta^3 L_3^0(t) T_\delta^3 & \vdots \\ & \ddots & \ddots & \ddots \\ 0 & \dots & \tilde{T}_\delta^\nu L_\nu^2 T_\delta^{\nu-2} & \tilde{T}_\delta^\nu L_\nu^1(t) T_\delta^{\nu-1} & \tilde{T}_\delta^\nu L_\nu^0(t) T_\delta^\nu \end{bmatrix}, \quad \mathcal{U}(t) = \begin{bmatrix} \tilde{\alpha}(t) \\ L_2^2 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \quad (7.20)$$

$$\mathcal{B}_k = \begin{bmatrix} \tilde{B}_k & 0 & \dots & 0 \\ V_{k,2}^1 & \tilde{T}_\delta^2 V_{k,2}^0 T_\delta^2 & & \vdots \\ 0 & \tilde{T}_\delta^3 V_{k,3}^1 T_\delta^2 & \tilde{T}_\delta^3 V_{k,3}^0 T_\delta^3 & \\ \vdots & & \ddots & \ddots \\ 0 & \dots & \tilde{T}_\delta^\nu V_{k,\nu}^1 T_\delta^{\nu-1} & \tilde{T}_\delta^\nu V_{k,\nu}^0 T_\delta^\nu \end{bmatrix}, \quad \mathcal{V}_k = \begin{bmatrix} \tilde{\beta}_k \\ 0 \\ \vdots \\ 0 \end{bmatrix}. \quad (7.21)$$

The block matrices in (7.20), (7.21) are given by (7.14)-(7.18) and (7.12), the matrices \tilde{T}_\cdot, T_\cdot , are the reduction matrices defined in (7.5).

Now, we can use Theorem 4.1 in order to rewrite the bilinear SDE (7.19) 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 (7.10), namely $\Phi_{\mathcal{Z}}(t)$, is uniformly nonsingular over T . There are many ways to assure this, starting from some suitable, non restrictive, hypothesis on the ordinary 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 $\Phi_{\mathcal{Z}}(t)$ 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 $\Phi_{\mathcal{Z}}(t)$.

Proposition 7.3. *Let ρ_k , $k = 1, \dots, p$, be the ranks of the matrices \mathcal{B}_k , given in (7.19). Then the process \mathcal{Z} satisfies the following SDE,*

$$d\mathcal{Z}(t) = (\mathcal{A}(t)\mathcal{Z}(t) + \mathcal{U}(t))dt + \sum_{k=1}^{2p} \tilde{\mathcal{B}}_k(t)d\tilde{W}_k(t), \quad (7.22)$$

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

$$\tilde{\mathcal{B}}_k(t) \triangleq \begin{cases} (\mathcal{B}_k \Phi_{\mathcal{Z}}(t) \mathcal{B}_k^T)^{(\frac{1}{2})}, & 1 \leq k \leq p \\ \mathcal{B}_{k-p} m_{\mathcal{Z}}(t) + \mathcal{V}_{k-p}, & p+1 \leq k \leq 2p \end{cases} \quad (7.23)$$

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

In order to write down the equations of the *augmented system* we need to split out the vector SDE (7.19) into two SDE's: one for the *observed* components of \mathcal{Z} and the other one for the remaining entries.

From the definition (7.9) we see that the components of the vector \mathcal{Z} are of the form:

$$X_1^{i_1} \dots X_n^{i_n} \dots Y_1^{j_1} \dots Y_q^{j_q}, \quad (7.24)$$

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 $l = 1, \dots, n$, $r = 1, \dots, q$, $\sum_{l=1}^n i_l \leq \nu$, $\sum_{r=1}^q j_r \leq \nu$. The observed components are those of the form (7.24) with $i_1 = \dots = i_n = 0$. 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

$$\mathcal{Y} = \mathcal{E}_{\mathcal{Y}} \mathcal{Z}. \quad (7.25)$$

It is easy to recognize that:

$$\mathcal{E}_{\mathcal{Y}} = \begin{bmatrix} \mathcal{E}_{\mathcal{Y}}^1 & 0 & \dots & 0 \\ 0 & \mathcal{E}_{\mathcal{Y}}^2 & \ddots & \\ \vdots & \ddots & \ddots & \\ 0 & \dots & 0 & \mathcal{E}_{\mathcal{Y}}^{\nu} \end{bmatrix}, \quad (7.26)$$

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

$$\mathcal{E}_{\mathcal{Y}}^j Z_{[j]} = Y_{[j]}, \quad (7.27)$$

and have the expressions

$$\mathcal{E}_{\mathcal{Y}}^j = [I_q \ 0]^{[j]} T_{\delta}^j, \quad (7.28)$$

where T_δ^j is the expansion matrix defined in (7.5). Let us denote with \mathcal{X} the aggregate vector of all the components in \mathcal{Z} which are not components of \mathcal{Y} . Then it results well defined the $(0, 1)$ -matrix $\mathcal{E}_\mathcal{X}$ such that

$$\mathcal{X} = \mathcal{E}_\mathcal{X} \mathcal{Z}. \quad (7.29)$$

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} :

$$\mathcal{I} \triangleq \begin{bmatrix} \mathcal{E}_\mathcal{Y} \\ \mathcal{E}_\mathcal{X} \end{bmatrix}, \quad (7.30)$$

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

$$\mathcal{Z} = \mathcal{I}_1 \mathcal{Y} + \mathcal{I}_2 \mathcal{X}. \quad (7.31)$$

Note that, from (7.25), (7.29), and because of the invertibility of the matrix \mathcal{I} , it results that the matrices $\mathcal{I}_1, \mathcal{I}_2$ defined in (7.31) are obtained by means of a suitable partition of the matrix $\mathcal{I}^{-1} = [\mathcal{I}_1 \ \mathcal{I}_2]$.

Using (7.25), (7.29), (7.31) and eq. (7.19) we can now state the following proposition.

Proposition 7.4. *The processes \mathcal{X}, \mathcal{Y} defined in (7.29) and (7.25) satisfy the following pair of SDE's (augmented system):*

$$d\mathcal{X}(t) = (\mathcal{A}_1(t)\mathcal{Y}(t) + \mathcal{A}_2(t)\mathcal{X}(t) + \mathcal{U}_1(t))dt + \sum_{k=1}^{2p} \mathcal{B}_k^1(t) d\tilde{W}_k(t), \quad (7.32)$$

$$d\mathcal{Y}(t) = (\mathcal{C}_1(t)\mathcal{Y}(t) + \mathcal{C}_2(t)\mathcal{X}(t) + \mathcal{U}_2(t))dt + \sum_{k=1}^{2p} \mathcal{D}_k^1(t) d\tilde{W}_k(t), \quad (7.33)$$

where

$$\begin{aligned} \mathcal{A}_1(t) &= \mathcal{E}_\mathcal{X} \mathcal{A}(t) \mathcal{I}_1, & \mathcal{A}_2(t) &= \mathcal{E}_\mathcal{X} \mathcal{A}(t) \mathcal{I}_2, & \mathcal{U}_1(t) &= \mathcal{E}_\mathcal{X} \mathcal{U}(t), & \mathcal{B}_k^1(t) &= \mathcal{E}_\mathcal{X} \tilde{\mathcal{B}}_k(t), \\ \mathcal{C}_1(t) &= \mathcal{E}_\mathcal{Y} \mathcal{A}(t) \mathcal{I}_1, & \mathcal{C}_2(t) &= \mathcal{E}_\mathcal{Y} \mathcal{A}(t) \mathcal{I}_2, & \mathcal{U}_2(t) &= \mathcal{E}_\mathcal{Y} \mathcal{U}(t), & \mathcal{D}_k^1(t) &= \mathcal{E}_\mathcal{Y} \tilde{\mathcal{B}}_k(t), \end{aligned} \quad (7.34)$$

$\mathcal{A}, \tilde{\mathcal{B}}_k, \mathcal{U}$, are the matrix coefficients of eq. (7.22), the matrices $\mathcal{E}_\mathcal{X}, \mathcal{E}_\mathcal{Y}, \mathcal{I}_1, \mathcal{I}_2$, are defined by means of eq.s (7.25), (7.29), (7.31), and $\{\tilde{W}_k, k = 1, \dots, 2p\}$ is a set of mutually uncorrelated standard WSW processes.

8. Polynomial filter equations

Proposition 7.4 states that the augmented observation process \mathcal{Y} defined in (7.25), can be generated as the output process of the augmented representation (7.32), (7.33). This implies that the problem of finding the ν -th degree polynomial filter for the original system (3.2), (3.3) is now reduced to an *optimal linear filtering* problem for the linear system (7.32), (7.33). Indeed, by denoting with $\hat{\mathcal{X}}(t)$ the optimal linear estimate given $\{\mathcal{Y}_s, s \leq t\}$ of the augmented state $\mathcal{X}(t)$, we have (see §2):

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

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

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

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 [12] the optimal linear filter is defined for the class of linear stochastic systems whose noise terms are represented by WSW processes. System (7.32), (7.33), comes within this class of systems, and we can use here the same approach as in [12] in order to obtain the optimal linear filter with respect to the augmented observation process \mathcal{Y} (and, hence the optimal ν -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 (7.32), (7.33), namely

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

Indeed, the uniform nonsingularity of (8.1) is required, in order to apply the Kalman-Bucy scheme to system (7.32), (7.33).

Theorem 8.1. *The noise covariance matrix function of the augmented measurement equation (7.33), given by (8.1), is uniformly nonsingular over T .*

Proof. See Appendix B. ■

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

Theorem 8.2. *The ν -th order polynomial filter for system (3.2), (3.3) is described by the following system of equations:*

$$\frac{dm_{\mathcal{Z}}(t)}{dt} = \mathcal{A}(t)m_{\mathcal{Z}}(t) + \mathcal{U}(t), \quad (8.2)$$

$$\bar{\mathcal{B}}_k(t) = \mathcal{B}_k m_{\mathcal{Z}}(t) + \mathcal{V}_k \quad 1 \leq k \leq p, \quad (8.3)$$

$$\Omega_p(t) = \sum_{k=1}^p \left(\tilde{\mathcal{B}}_k(t)^{[2]st} (I_{\rho_k}) + \bar{\mathcal{B}}_k(t)^{[2]} \right), \quad (8.4)$$

$$\mathcal{H}_1(t) = U_\gamma^2(\mathcal{A}(t) \otimes I_\gamma), \quad \mathcal{H}_2(t) = U_\gamma^2(\mathcal{U}(t) \otimes I_\gamma), \quad \mathcal{H}_3(t) = \frac{1}{2} O_\gamma^2 \Omega_p(t), \quad (8.5)$$

$$\frac{d\Gamma(t)}{dt} = \mathcal{H}_1(t)\Gamma(t) + \mathcal{H}_2(t)m_z(t) + \mathcal{H}_3(t), \quad (8.6)$$

$$\tilde{\mathcal{B}}_k(t) = \left(\mathcal{B}_k st^{-1} (\Gamma(t) - m_z(t)^{[2]}) \mathcal{B}_k^T \right)^{\left(\frac{1}{2}\right)} \quad 1 \leq k \leq p, \quad (8.7)$$

$$\mathcal{J}(t) = \sum_{k=1}^p \mathcal{E}_X (\tilde{\mathcal{B}}_k(t) \tilde{\mathcal{B}}_k(t)^T + \bar{\mathcal{B}}_k(t) \bar{\mathcal{B}}_k(t)^T) \mathcal{E}_Y^T, \quad (8.8)$$

$$\mathcal{R}(t) = \sum_{k=1}^p \mathcal{E}_Y (\tilde{\mathcal{B}}_k(t) \tilde{\mathcal{B}}_k(t)^T + \bar{\mathcal{B}}_k(t) \bar{\mathcal{B}}_k(t)^T) \mathcal{E}_Y^T, \quad (8.9)$$

$$\mathcal{Q}(t) = \sum_{k=1}^p \mathcal{E}_X (\tilde{\mathcal{B}}_k(t) \tilde{\mathcal{B}}_k(t)^T + \bar{\mathcal{B}}_k(t) \bar{\mathcal{B}}_k(t)^T) \mathcal{E}_X^T, \quad (8.10)$$

$$\begin{aligned} \frac{d\mathcal{P}(t)}{dt} &= \mathcal{A}_2(t)\mathcal{P}(t) + \mathcal{P}(t)\mathcal{A}_2(t)^T + \mathcal{Q}(t), \\ &\quad - (\mathcal{J}(t) + \mathcal{P}(t)\mathcal{C}_2(t)^T)\mathcal{R}(t)^{-1}(\mathcal{J}(t) + \mathcal{P}(t)\mathcal{C}_2(t)^T)^T, \end{aligned} \quad (8.11)$$

$$\begin{aligned} d\hat{\mathcal{X}}(t) &= (\mathcal{A}_1(t)\mathcal{Y}(t) + \mathcal{A}_2(t)\hat{\mathcal{X}}(t) + \mathcal{U}^1(t))dt + (\mathcal{J}(t) + \mathcal{P}(t)\mathcal{C}_2(t)^T)\mathcal{R}(t)^{-1} \\ &\quad \cdot (d\mathcal{Y}(t) - (\mathcal{C}_1(t)\mathcal{Y}(t) + \mathcal{C}_2(t)\hat{\mathcal{X}}(t) + \mathcal{U}^2(t))dt), \end{aligned} \quad (8.12)$$

$$\hat{\mathcal{X}}(t) = \mathcal{T}_\nu \hat{\mathcal{X}}(t), \quad (8.13)$$

where \mathcal{T}_ν 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 (7.22) and (7.34), the matrices \mathcal{B}_k are defined in (7.21), $\rho_k = \text{rank}(\mathcal{B}_k)$, and eqns (8.2), (8.4), (8.11), (8.12), are endowed with the initial conditions:

$$\begin{aligned} m_z(0) &= E(\mathcal{X}(0)), \\ \Gamma(0) &= E(\mathcal{X}(0)^{[2]}), \\ \hat{\mathcal{X}}(0) &= E(\mathcal{X}(0)) + \text{cov}(\mathcal{X}(0), \mathcal{Y}(0)) \text{cov}^\dagger(\mathcal{Y}(0), \mathcal{Y}(0)) (\mathcal{Y}(0) - E(\mathcal{Y}(0))), \\ \mathcal{P}(0) &= \text{cov}(\mathcal{X}(0), \mathcal{X}(0)) - \text{cov}(\mathcal{X}(0), \mathcal{Y}(0)) \text{cov}^\dagger(\mathcal{Y}(0), \mathcal{Y}(0)) \text{cov}^T(\mathcal{X}(0), \mathcal{Y}(0)). \end{aligned}$$

Proof. Eqns. (8.7)-(8.11), easily derives from an application of [12, Th. 15.3] to the representation (4.2), (4.3). In particular eq. (8.7) immediately derives from (7.23), as soon as it is noticed that $\Phi_Z(t) = st^{-1}(\Gamma(t) - m_z(t)^{[2]})$, where

$$\Gamma(t) \triangleq E(\mathcal{Z}(t)^{[2]}) = st \left(E(\mathcal{Z}(t)\mathcal{Z}(t)^T) \right). \quad (8.14)$$

Taking the expectations of both sides of (7.22) we obtain eq. (8.2). In order to obtain an ODE for the vector function $\Gamma(t)$, let us apply the vector Ito formula, as is given by Theorem 5.2, to the eq. (7.22), by setting

$$X_t = \mathcal{Z}(t), \quad \beta_t = \int_0^t (\mathcal{A}(\tau)\mathcal{Z}(\tau) + \mathcal{U}(\tau))d\tau, \quad M_t = \int_0^t \sum_{k=1}^{2p} \tilde{\mathcal{B}}_k(\tau) d\tilde{W}_k(\tau),$$

24.

and using Lemma 3.2 for $h = 2$ and $n = \gamma$, it results (time dependencies are skipped for convenience):

$$\begin{aligned} d\mathcal{Z}^{[2]} &= U_\delta^2(I_\gamma \otimes \mathcal{Z})d\mathcal{Z} + \frac{1}{2}O_\gamma^2(dM_t)^{[2]} \\ &= \left(U_\delta^2(I_\gamma \otimes \mathcal{Z})\mathcal{A}\mathcal{Z} + U_\delta^2(I_\gamma \otimes \mathcal{Z})\mathcal{U} \right) dt + \frac{1}{2}O_\gamma^2(dM_t)^{[2]} + \sum_{k=1}^{2p} U_\delta^2(I_\gamma \otimes \mathcal{Z})\tilde{\mathcal{B}}_k d\tilde{W}_k. \end{aligned} \quad (8.15)$$

Now, taking into account (7.23) and the definition of the associate quadratic variation process, given by (5.11), we have:

$$\begin{aligned} (dM_t)^{[2]} &= \sum_{k=1}^p \tilde{\mathcal{B}}_k^{[2]}(d\tilde{W}_k)^{[2]} + \sum_{k=1}^p (\mathcal{B}_k m_z + \mathcal{V}_k)^{[2]}(dW_k)^{[2]} \\ &= \sum_{k=1}^p \tilde{\mathcal{B}}_k^{[2]} st(d \langle \tilde{W}_k, \tilde{W}_k^T \rangle) + \sum_{k=1}^p (\mathcal{B}_k m_z + \mathcal{V}_k)^{[2]} dt \\ &= \Omega_p dt, \end{aligned} \quad (8.16)$$

where Ω_p has the expression (8.4). Moreover, using formula (6.3) with suitable substitution of symbols, the following equalities are recognized to hold:

$$\begin{aligned} (I_\gamma \otimes \mathcal{Z})\mathcal{A}\mathcal{Z} &= (\mathcal{A} \otimes I_\gamma)\mathcal{Z}^{[2]} \\ (I_\gamma \otimes \mathcal{Z})\mathcal{U} &= (\mathcal{U} \otimes I_\gamma)\mathcal{Z}. \end{aligned} \quad (8.17)$$

By substituting (8.17) and (8.16) in (8.15), and taking the expectation we obtain the ODE (8.6).

The so obtained estimate $\hat{\mathcal{X}}_t$ is the optimal one among all the linear transformation of the augmented observation process $\{\mathcal{Y}_s, s \leq t\}$ and hence it is the ν -th degree polynomial estimate of the augmented state \mathcal{X}_t . In order to obtain the analogous 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 $\{\mathcal{Y}_s, s \leq t\}$, we have that each entry of $\hat{\mathcal{X}}_t$ agree 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 (7.6), (7.9), (7.10) and by the definition of the extracting operator $\mathcal{E}_{\mathcal{X}_t}$, 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 eq. (8.13). ■

9. An example of application: the scalar cubic sensor

As a simple example of application, in this section we derive the equations of the quadratic filter (that is the filter producing the polynomial state-estimate in the sense of Definition 2.1 with $\nu = 2$) for the scalar cubic sensor. As was pointed out in the introduction, for this class of systems a finite-dimensional optimal filter does not exist. Nevertheless, it is possible to use the methodology described in the previous sections in order to obtain optimal polynomial state-estimates of an arbitrarily fixed degree (of course, provided the existence of the moments of the involved processes up to a suitable order). The equations of the cubic sensor are the following:

$$dx(t) = dW(t), \quad (9.1)$$

$$dy(t) = x^3(t)dt + dV(t), \quad (9.2)$$

where $t \in T$, all the processes x, y, W, V are \mathbb{R} -valued, W, V are independent standard Wiener processes, and $x(0)$ and $y(0)$ are standard Gaussian random variables. We will show that the quadratic filter for system (9.1), (9.2), is well defined, in that the output covariance $\mathcal{R}(t)$ appearing in the polynomial filter equations (eq.(8.9)), results to be uniformly nonsingular in T . Note that, this is not a consequence of Theorem 8.1. Indeed, Theorem 8.1 holds under the Assumption 3.1 which is not verified by eq. (9.2). Nevertheless, Theorem 8.1 gives only a sufficient condition for the uniform nonsingularity of $\mathcal{R}(t)$.

In order to compute the augmented system equations, let us consider the cubic sensor equations (9.1), (9.2), and compute, using the standard scalar Ito formula, the differential of the squared output process $y(t)^2$:

$$dy(t)^2 = 2y(t)x(t)^3 dt + 2y(t)dV(t) + dt. \quad (9.3)$$

In order to compute the stochastic differential of the product $y(t)x(t)^3$, let us define the process

$$z(t) \triangleq \begin{bmatrix} x(t) \\ y(t) \end{bmatrix},$$

and apply Theorem 5.2 with

$$X_t = z(t); \quad \beta_t = A_z z(t)^{[3]}; \quad M_t = \begin{bmatrix} W(t) \\ V(t) \end{bmatrix}; \quad Z_t = F(z(t)) = y(t)x(t)^3,$$

where A_z is some suitably defined matrix. It results

$$d(y(t)x(t)^3) = x(t)^6 dt + 3y(t)x(t)dt + x(t)^3 dV(t) + 3x(t)^2 y(t)dW(t). \quad (9.4)$$

In a similar way, let us calculate the differentials $dx(t)^i$, $i = 2, \dots, 6$ and $d(y(t)x(t)^2)$, $d(y(t)x(t))$. It results

$$\begin{aligned} dx^6(t) &= 6x(t)^5 dW(t) + 12x(t)^4 dt, \\ dx^5(t) &= 5x(t)^4 dW(t) + 10x(t)^3 dt, \\ dx^4(t) &= 4x(t)^3 dW(t) + 6x(t)^2 dt, \\ dx^3(t) &= 3x(t)^2 dW(t) + 3x(t)dt, \\ dx^2(t) &= 2x(t)dW(t) + dt, \\ d(y(t)x(t)^2) &= x(t)^5 dt + y(t)dt + x(t)^2 dV(t) + 2y(t)x(t)dW(t), \\ d(y(t)x(t)) &= x(t)^4 dt + y(t)dt + x(t)^2 dV(t) + y(t)dW(t). \end{aligned} \quad (9.5)$$

Now let us define the augmented observation process $\{\mathcal{Y}(t)\}$ and the augmented state process $\{\mathcal{X}(t)\}$ as

$$\mathcal{Y}(t) \triangleq [y(t) \ y(t)^2]^T;$$

$$\mathcal{X}(t) \triangleq [y(t) \ x(t) \ x(t)^2 \ x(t)^3 \ x(t)^4 \ x(t)^5 \ x(t)^6 \ y(t)x(t) \ y(t)x(t)^2 \ y(t)x(t)^3]^T;$$

then, from (9.1)-(9.5), it follows that the processes $\{\mathcal{X}(t)\}$, $\{\mathcal{Y}(t)\}$ can be represented as the state and output processes of the following stochastic augmented bilinear model:

$$\begin{aligned} d\mathcal{X}(t) &= \mathcal{A}\mathcal{X}(t)dt + b_1dt + \mathcal{B}_1\mathcal{X}(t)dW(t) + \mathcal{B}_2\mathcal{X}(t)dV(t) + b_2dW(t), \\ d\mathcal{Y}(t) &= \mathcal{C}\mathcal{X}(t)dt + b_3dt + \mathcal{B}_3\mathcal{X}(t)dV(t) + b_4dV(t) \end{aligned} \quad (9.6)$$

where b_1, b_2, b_3, b_4 are suitably defined constant vectors and $\mathcal{A}, \mathcal{B}_1, \mathcal{B}_2, \mathcal{B}_3, \mathcal{C}$ are suitably defined constant matrices. Now, we can apply Theorem 4.1 in order to put the bilinear system (9.6) in the form of a linear stochastic system with WSW diffusions:

$$\begin{aligned} d\mathcal{X}(t) &= \mathcal{A}\mathcal{X}(t)dt + b_1dt + \tilde{\mathcal{B}}_1d\tilde{W}(t) + \mathcal{B}_1E(\mathcal{X}(t))dW(t) + \tilde{\mathcal{B}}_2d\tilde{V}(t) \\ &\quad + \mathcal{B}_2E(\mathcal{X}(t))dV(t) + b_2dW(t), \end{aligned} \quad (9.7)$$

$$d\mathcal{Y}(t) = \mathcal{C}\mathcal{X}(t)dt + b_3dt + \tilde{\mathcal{B}}_3d\tilde{V}(t) + \mathcal{B}_3E(\mathcal{X}(t))dV(t) + b_4dV(t), \quad (9.8)$$

where, for $i = 1, 2, 3$:

$$\tilde{\mathcal{B}}_i = \left(\mathcal{B}_i \Psi_{\mathcal{X}}(t) \mathcal{B}_i^T \right)^{\left(\frac{1}{2}\right)},$$

where $\Psi_{\mathcal{X}}(t)$ is the covariance of $\mathcal{X}(t)$.

In order to show the existence of the quadratic filter we have only to show that the covariance of the output noise of eq. (9.8), namely $\xi(t)$:

$$\xi(t) = \tilde{\mathcal{B}}_3d\tilde{V}(t) + \left(\mathcal{B}_3E(\mathcal{X}(t)) + b_4 \right) dV(t),$$

is uniformly nonsingular over T . To this purpose, first of all note that:

$$\mathcal{B}_3 = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}; \quad b_4 = \begin{bmatrix} 1 \\ 0 \end{bmatrix};$$

from which one has:

$$\mathcal{B}_3 \Psi_{\mathcal{X}}(t) \mathcal{B}_3^T = \begin{bmatrix} 0 & 0 \\ 0 & E(y(t) - E(y(t)))^2 \end{bmatrix}, \quad \mathcal{B}_3 E(\mathcal{X}(t)) + b_4 = \begin{bmatrix} 1 \\ E(y(t)) \end{bmatrix}.$$

Hence, the covariance of $\xi(t)$, namely $\Psi_{\xi}(t)$, is given by

$$\Psi_{\xi}(t) = \begin{bmatrix} 1 & E(y(t)) \\ E(y(t)) & E(y(t)^2) \end{bmatrix},$$

whose determinant is equal to $E(y(t) - E(y(t)))^2$, that is the variance of the process $y(t)$. From the system equations (9.1), (9.2) and using (9.3)-(9.5), we can easily see that $E(y(t)) = 0$ and $E(y(t)^2)$ is an increasing function of t . Hence, since $E(y(0)^2) = 1$ by the hypothesis, we have that the variance of the process $y(t)$ is greater than one, for all $t \in T$. This proves the uniform nonsingularity of $\Psi_{\xi}(t)$ in T and hence the existence of a finite-dimensional quadratic filter for scalar cubic sensor.

10. Conclusions

Eqns. (8.2)-(8.13) define a finite-dimensional filter for the cubic-sensor-like system (3.2), (3.3) which is optimal in the class of all the estimates which can be written either as a finite linear combination of Kronecker powers of the currently available measurements, or as a mean square limit of these. We have called *polynomial estimates* this kind of estimates. Even if the considered class does not includes all the polynomials, however it includes the linear estimates and, moreover, it defines a not decreasing 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; indeed, this is an important feature because, for the considered class of systems, the optimal filter is *necessarily* infinite-dimensional (as shown in [3]). 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 [15]) with an arbitrary approximation degree. However, the more accurate 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 has no sense, within this approach, to use a large approximation degree in order to make really implementable the filtering algorithm. Counterwise, our suboptimal approach allows to get meaningful estimates also for small polynomial degrees, which does not present difficult implementation problems.

As an auxiliary result we have obtained in §2, 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 (7.19), which allows us to obtain the linear representation (7.22).

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

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 [16]. 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

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

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

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

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

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

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 [16]:

$$(A + B) \otimes (C + D) = A \otimes C + A \otimes D + B \otimes C + B \otimes D, \quad (\text{A.3a})$$

$$A \otimes (B \otimes C) = (A \otimes B) \otimes C, \quad (\text{A.3b})$$

$$(A \cdot C) \otimes (B \cdot D) = (A \otimes B) \cdot (C \otimes D), \quad (\text{A.3c})$$

$$(A \otimes B)^T = A^T \otimes B^T, \quad (\text{A.3d})$$

$$st(A \cdot B \cdot C) = (C^T \otimes A) \cdot st(B), \quad (\text{A.3e})$$

$$u \otimes v = st(v \cdot u^T), \quad (\text{A.3f})$$

$$tr(A \otimes B) = tr(A) \cdot tr(B), \quad (\text{A.3g})$$

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

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

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

$$tr(A^{[h]}) = (tr(A))^h. \quad (\text{A.3h})$$

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

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

where $[\cdot]$ and $|\cdot|_s$ denote integer part and s -modulo respectively. Even if the Kronecker product is not commutative, the following property holds [10], [11], [17].

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

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

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:

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

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

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

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 \mathbb{R}^l we have

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

Proof. See [11]. ■

APPENDIX B

In this Appendix, the proof of theorem 8.1 is presented. We need to state in advance some preliminary definitions and lemmas.

Let δ and j two positive integers.

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

Remark B.3. Let $x \in \mathbb{R}^\delta$. For some $s, r \in \{1, 2, \dots, \delta^j\}$ let us consider the multiindexes : s_1, \dots, s_j and r_1, \dots, r_j in $\{1, \dots, \delta\}$ defined by the identities:

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

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

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

$$\rho(s; \delta, j) \triangleq \left\{ r \in \mathbb{N} : 1 \leq r \leq \delta^j, (s, r) \text{ is } (\delta, j)\text{-R} \right\}. \quad (\text{B.1})$$

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

$$\rho'(s; \delta, j) \triangleq \left\{ i \in \rho(s; \delta, j) \mid \left\lfloor \frac{i}{\delta^{j-1}} \right\rfloor = \left\lfloor \frac{s}{\delta^{j-1}} \right\rfloor \right\}, \quad (\text{B.2})$$

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

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

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

where every subvector $x_i x^{[j-1]}$ has dimension δ^{j-1} . By setting $l = \lfloor s/\delta^{j-1} \rfloor$ 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 (δ, j) -R and $(x^{[j]})_i \in x_l x^{[j-1]}$. Counterwise, the set defined in (B.3) is composed with the integers i such that (i, s) is (δ, j) -R and $(x^{[j]})_i$ 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:

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

Remark B.5. Note that the number δ_j , agree with the number of entries of $x_{[j]}$, for $x \in \mathbb{R}^\delta$.

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

$$r' \triangleq \left\lfloor \frac{r}{\delta^{j-1}} \right\rfloor, \quad s' \triangleq \left\lfloor \frac{s}{\delta^{j-1}} \right\rfloor.$$

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

Proof. From the Definition B.1 it results:

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

From (B.4) we see that:

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

and hence, from (B.6),

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

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

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

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

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

$$\mathcal{I} - n = \{i / i \in \mathbb{N}, \exists i' \in \mathcal{I}, \text{ such that } i = i' - n\}. \quad (\text{B.8})$$

Lemma B.7. Suppose that

$$\left\lfloor \frac{s}{\delta^{j-1}} \right\rfloor = l < \delta, \quad (\text{B.9})$$

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

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

where ρ' is the set defined in (B.2)

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

Suppose first that (r, s) is (δ, j) -NR. Let $x \in \mathbb{R}^\delta$ and $z = x^{[j]}$. From the structure (B.4) of the vector z and taking into account of (B.9), we see that: $z_s, z_r \in x_l \cdot x^{[j-1]}$. Hence, since (s, r) is

(δ, j) -NR, we have that, there exist integers h_1, \dots, h_δ and h'_1, \dots, h'_δ , $h_1 + \dots + h_\delta = h'_1 + \dots + h'_\delta = j - 1$ such that it results

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

Since $z_r \neq z_s$ it follows that

$$x_1^{h_1} \cdots x_\delta^{h_\delta} \neq x_1^{h'_1} \cdots x_\delta^{h'_\delta}. \quad (\text{B.11})$$

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

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

and

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

and hence, taking into account of (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 (δ, j) -NR.

Next, suppose that (r, s) is (δ, 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 of (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 (δ, j) -R. ■

Lemma B.8. *Let (r, s) be a (δ, j) -R pair, such that*

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

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

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

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

Since $z_s = z_r$, eq. (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:

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

From which, 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 output process Y of system (3.2), (3.3), and the extended state process X_e , defined in (7.1). We remind that q and δ are the dimensions of the vectors Y and $Z = [Y^T X_e^T]^T$ 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}_y defined in (7.25), and recall that the diagonal blocks \mathcal{E}_y^j , $j = 1, \dots, \nu$, appearing there, are such that eq. (7.27) 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}_δ^j defined in (7.5), and the matrix U_δ^j defined in (5.5). We can prove the following Lemma.

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

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

Proof. Using (7.27) and property (5.4) we have:

$$\begin{aligned} \mathcal{E}_y^j \left(\frac{d}{dZ} \otimes Z_{[j]} \right) &= \frac{d}{dZ} \otimes \mathcal{E}_y^j Z_{[j]} = \frac{d}{dZ} \otimes Y_{[j]} \\ &= \begin{bmatrix} \frac{\partial}{\partial Y} & \frac{\partial}{\partial X_e} \end{bmatrix} \otimes Y_{[j]} = \begin{bmatrix} \frac{\partial}{\partial Y} \otimes Y_{[j]} & 0 \end{bmatrix}. \end{aligned} \quad (\text{B.16})$$

On the other hand, by (7.5), (5.3), and using formula (5.5):

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

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

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

from which it follows that $S = 0$ and

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

Let $V = (d/dY) \otimes Y_{[j]}$. Note that the components of the matrix V are either zero or they are monomials of $j-1$ -th 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 cannot exist two (nonzero) similar monomials on the same column, because $Y_{[j]}$ has not repeated entries. Hence, L_δ^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 $u^T V = 0, \forall Y \in \mathbb{R}^q$, 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_δ^j . The following properties holds:*

- 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

$$l \triangleq \left\lfloor \frac{s}{\delta^{j-1}} \right\rfloor, \quad r \triangleq |s|_{\delta^{j-1}}. \quad (\text{B.19})$$

Consider again the relation (B.18):

$$\frac{d}{dY} \otimes Y_{[j]} = \begin{bmatrix} \frac{\partial}{\partial Y_1} Y_{[j]} & \dots & \frac{\partial}{\partial Y_q} Y_{[j]} \end{bmatrix} = L_\delta^j (I_q \otimes Z^{[j-1]}). \quad (\text{B.20})$$

From (B.20) it results

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

where

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

Now, from (B.21) we see that each component of $(\partial/\partial Y_l)Y_{[j]}$ is either equal to zero or it 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 $(\lambda_{(l-1)\delta^{j-1}+i})_k = 0$ for $k = 1, \dots, q_j$ and $k \neq h$. As a matter of fact, if $(\lambda_{(l-1)\delta^{j-1}+i})_k \neq 0$, 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 the 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. The part B) of the Theorem follows as soon as it is noticed that, using Lemma B.6 and taking into account of (B.19), it results $\{\lambda_{(l-1)\delta^{j-1}+i}, i \in \rho(r; \delta, j-1)\} = \{\lambda_i, i \in \rho'(s; \delta, j)\}$.

Finally, in order to prove part C) note that, since $l \leq q$ (and hence, by recalling the structure of Z , given by (7.9), it results $Z_l = Y_l$) we have that the s -th component of $Z^{[j]}$ is in the form: $Y_l Z_1^{h_1} \dots Z_\delta^{h_\delta}$, where the powers h_1, \dots, h_δ are such that $h_1 + \dots + h_\delta = j-1$, and it is not a Y -monomial by the hypothesis. Hence the monomial $Z_1^{h_1} \dots Z_\delta^{h_\delta}$ 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_δ^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 (δ, j) -Kronecker space, namely $K(\delta, j)$ as the following subspace of \mathbb{R}^{δ^j} :

$$K(\delta, j) = \text{span} \left(\left\{ z \in \mathbb{R}^{\delta^j} \mid \exists x \in \mathbb{R}^\delta \text{ such that } z = x^{[j]} \right\} \right).$$

Proof of Theorem 8.1. By exploiting the definition of \mathcal{D}_k^1 given in (7.34), and the definition of $\tilde{\mathcal{B}}_k$, given by (7.23), we can rewrite the matrix $\mathcal{R}(t)$, defined in (8.1), as

$$\mathcal{R}(t) = \sum_{k=1}^p \mathcal{E}_y \mathcal{B}_k \Phi_Z(t) \mathcal{B}_k^T \mathcal{E}_y^T + \sum_{k=1}^p \mathcal{E}_y (\mathcal{B}_k m_Z(t) + \mathcal{V}_k) (\mathcal{B}_k m_Z(t) + \mathcal{V}_k)^T \mathcal{E}_y^T. \quad (\text{B.22})$$

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

In order to verify this, first of all note that, from the Assumption 3.1, Remark 3.2, and the definition of the matrix $D_{e,k}$ given in (7.4), it results that there exists a \bar{k} such that $\text{rank}(D_{e,\bar{k}}) = q$, (we remind reader that q is the dimension of the ordinary observation Y). More exactly, from (7.4), we have:

$$D_{e,\bar{k}} = [I_q \ 0]. \quad (\text{B.23})$$

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

$$\text{rank}(\mathcal{E}_Y \mathcal{B}_{\bar{k}}) = q + q_2 + \dots + q_\nu, \quad (\text{B.24})$$

that is, it is a full (row) rank matrix (remind that q_i is the dimension of $Y_{[i]}$). From the definition of \mathcal{B}_k and \mathcal{E}_Y , given in (7.21) and (7.26) respectively, using (7.17) and taking into account the block triangular structure of \mathcal{B}_k , it results that condition (B.24) is equivalent to:

$$\text{rank}(\mathcal{E}_Y^1 \tilde{B}_{\bar{k}}) = q, \quad (\text{B.25})$$

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

Now, from (7.28) we see that $\mathcal{E}_Y^1 \in \mathbb{R}^{q \times \delta}$, $\mathcal{E}_Y^1 = [I_q \ 0]$. Hence, by the definition of $\tilde{B}_{\bar{k}}$, given in (7.12), and taking into account of (B.23), it results $\mathcal{E}_Y^1 \tilde{B}_{\bar{k}} = [0 \ I_q \ 0]$, and hence condition (B.25) is verified.

It remains to prove (B.26). In order to do this, first of all note that, from the definition of $\tilde{B}_{\bar{k}}$ given in (7.12), and taking into account of (B.23), we can consider the following partition of the matrix $\tilde{B}_{\bar{k}} \otimes I_{\delta^{j-1}}$:

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

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

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

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

$$\mathcal{E}_Y^j \tilde{T}_\delta^j U_\delta^j (\tilde{B}_{\bar{k}} \otimes I_{\delta^{j-1}}) T_\delta^j = [L_\delta^j \ 0] \begin{bmatrix} M_1 \\ M_2 \end{bmatrix} T_\delta^j = L_\delta^j M_1 T_\delta^j. \quad (\text{B.29})$$

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

Let $y \in \mathbb{R}^{q_j}$, we will prove that there exists a $z \in K(\delta, j)$ such that $y = L_\delta^j M_1 z$. For, by Lemma B.9, $\text{rank}(L_\delta^j) = q_j$, then there exist q_j indexes: $1 \leq i_1, i_2, \dots, i_{q_j} \leq q\delta^{j-1}$, such that the columns $\lambda_{i_1}, \lambda_{i_2}, \dots, \lambda_{i_{q_j}}$ (λ_i denotes as usual the i -th column of L_δ^j) 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

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

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

$$y = \alpha_{i_1} \bar{\lambda}_{i_1} + \dots + \alpha_{i_{q_j}} \bar{\lambda}_{i_{q_j}}. \quad (\text{B.31})$$

Now, let us show that the elements of the set $\{i_s + q\delta^{j-1} \mid s = 1, \dots, q_j\}$ are pairwise (δ, j) -NR. At this purpose, for any pair (i_r, i_s) , $r \neq s$, $r, s = 1, \dots, q_j$, 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 λ_{i_r} and λ_{i_s} are linearly independent, it follows that (i_r, i_s) is (δ, j) -NR. Indeed, if (i_r, i_s) were (δ, j) -R, then Lemma B.10 part *B)* would imply that λ_{i_r} and λ_{i_s} are linearly dependent vectors. Hence, since (i_r, i_s) is (δ, j) -NR, Lemma B.7 implies that $(i_r + q\delta^{j-1}, i_s + q\delta^{j-1})$ is (δ, j) -NR.

ii)

$$i'_r \triangleq \left[\frac{i_r}{\delta^{j-1}} \right] \neq \left[\frac{i_s}{\delta^{j-1}} \right] \triangleq i'_s. \quad (\text{B.32})$$

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

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

where i'_r, i'_s are given by (B.32). Since λ_{i_r} and λ_{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 (δ, j) -R we should have

$$Z_{i_r+q} Z_1^{h_1} \dots Z_\delta^{h_\delta} = Z_{i_s+q} Z_1^{h'_1} \dots Z_\delta^{h'_\delta},$$

which is possible if and only if

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

Now, $i'_r, i'_s \leq q$, 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, a contradiction.

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

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

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

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

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

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

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

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

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

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)

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

hence, (B.38) becomes:

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

which is implied by

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

In order to prove (B.39), first of all 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 (δ, j) -NR and such that $(i + q\delta^{j-1}, i_s + q\delta^{j-1})$ is (δ, j) -R. Now, let $h_1, \dots, h_\delta, h'_1, \dots, h'_\delta$ such that $h_1 + \dots + h_\delta = h'_1 + \dots + h'_\delta = j - 1$ and

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

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 (δ, j) -R it results

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

which in turn implies the following condition

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

Since $i', i'_s \leq q$, $Z_{i'+q}$ 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).

References

- [1] R.S. Bucy: "Bayes theorem and digital realizations for nonlinear systems". *J. Astron. Sci.*, Vol. 17, 1969, pp. 80-94.
- [2] R.S. Bucy and J. Pages: "A priori error bounds for the cubic sensor problem", *IEEE Trans. Autom. Control*, Vol. 23, No.1, 1978, pp. 88-91.
- [3] M. Hazewinkel, S.I. Marcus and H.J. Sussmann: "Nonexistence of finite-dimensional filters for conditional statistics of the cubic sensor problem", *System & Control Letters* 3, December 1983.
- [4] Y. Steinberg, B. Z. Bobrowsky and Z. Schuss, "On the optimal filtering problem for the cubic sensor", *Circ. Sys. Sign. Proc.*, Vol 7, No. 3, 1988, p. 381-408.
- [5] R.R. Mohler: "Bilinear Control Processes", Academic Press, New York, 1970.
- [6] C. Bruni, G. Di Pillo and G. Koch: "Bilinear Systems: An Appealing Class of "Nearly Linear" Systems in Theory and Applications", *IEEE Trans. Aut. Contr.*, vol. AC-19, no. 4, Aug. 1974.
- [7] C. Bruni and G. Koch: "Suboptimal Filtering for Linear-in-Control Systems with Small Observation Noise", *IV IFAC Symposium on Identification and System Parameter Estimation*, Tblisi, Sept. 1976.
- [8] X. Yang, R.R. Mohler and R.M. Burton: "Adaptive Suboptimal Filtering of Bilinear Systems", *Int. J. Cont.*, vol 52, no. 1, 1990.
- [9] E. Yaz: "Full and Reduced Order Observer Design for Discrete Stochastic Bilinear Systems": *IEEE Trans. Aut. Contr.*, vol. 37, no. 4, april 1992.
- [10] F. Carravetta, A. Germani and M. Raimondi: "Polynomial Filtering for Linear Discrete-Time non-Gaussian Systems", *SIAM J. Cont. & Opt.*, vol 34, no. 5, September 1996.
- [11] F. Carravetta, A. Germani and M. Raimondi: "Polynomial Filtering of Discrete-Time Stochastic Linear Systems with Multiplicative State Noise", *IEEE Trans. Aut. Contr.*, vol 42, No. 8, August 1997.
- [12] R.S. Liptser, A.N. Shiriyayev. "Statistics of Random Processes", Vol 2, Springer Verlag, New York, 1978.
- [13] G. Kallianpur. "Stochastic Filtering Theory", Springer-Verlag, New York, 1980.
- [14] V. Krishnan. "Nonlinear Filtering and Smoothing", John Wiley & Sons, New York, 1984.
- [15] A. Germani and M. Piccioni. "Semi-discretization of Stochastic Partial Differential Equations on \mathbb{R}^d by a Finite-element Technique", *Stochastics*, Vol 23, 1988, pp. 131-148.
- [16] R. Bellman: "Introduction to Matrix Analysis", McGraw-Hill, New York 1970.
- [17] G.S. Rodgers, "Matrix Derivatives", Marcel Dekker, Inc., New York and Basel, 1980.
- [18] R. E. Kalman, "A new approach to linear filtering and prediction problems.", *J. Basic. Eng.*, Vol 1, 1960, pp. 35-45.
- [19] R. E. Kalman and R. S. Bucy, "New results in linear filtering and prediction theory.", *Trans. ASME Ser. D. J. Basic Eng.*, Vol 83, 1961, pp. 95-108.