ISTITUTO DI ANALISI DEI SISTEMI ED INFORMATICA
"Antonio Ruberti"
CONSIGLIO NAZIONALE DELLE RICERCHE

A. Germani, C. Manes, P. Palumbo

STATE AND MODE ESTIMATION OF STOCHASTIC SYSTEMS WITH SWITCHING MEASUREMENTS

R. 09-13 2009

Pasquale Palumbo – Istituto di Analisi dei Sistemi ed Informatica "A. Ruberti", Viale Manzoni 30, 00185 Roma, Italy. E-mail: pasquale.palumbo@iasi.cnr.it;.

Alfredo Germani – Dipartimento di Ingegneria Elettrica e dell'Informzione, Università degli Studi dell'Aquila, Loc. Campo di Pile, 67100 L'Aquila, Italy. E-mail: alfredo.germani@univaq.it.

Costanzo Manes – Dipartimento di Ingegneria Elettrica e dell'Informzione, Università degli Studi dell'Aquila, Loc. Campo di Pile, 67100 L'Aquila, Italy. E-mail: costanzo.manes@univaq.it.

This work was partially supported by IASI-CNR

ISSN: 1128–3378

Collana dei Rapporti dell'Istituto di Analisi dei Sistemi ed Informatica "Antonio Ruberti", 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 state estimation problem is here investigated for a class of stochastic linear switching-output systems, in which the output matrix switches in a finite set of possible values according to a not directly measured discrete Markov sequence. This note presents a real time algorithm, based on the optimal polynomial filtering approach, that achieves the simultaneous estimation of both the continuous system state and the switching parameter. The state and observation noises do not need to be Gaussian. It is shown that the optimal filter of degree one (best affine filter) does not solve the parameter estimation problem, due to a structural first-order unobservability property, and therefore the use of higher order filters becomes necessary. As an application of the proposed filter, the problem of the on-line simultaneous estimation of the transmitted signal and of the impulse response samples of a multipath fast-fading digital communication channel, is considered in this paper. Differently from other approaches, the polynomial filter solves the problem without the use of training sequences (preambles) in the transmitted data, so that the information flow through the channel is not interrupted.

Key words: Hybrid systems; Switching Systems; Kalman filter; Polynomial Filters.

1. Introduction

In the last few years, a large amount of scientific research in systems sciences has been devoted to hybrid systems, that is dynamical systems characterized by a finite (or countable) number of operating modes (see e.g. [2], [3], [31], [28]). In hybrid systems the switching among the different behavioral modes can be internally and/or externally driven, according to deterministic and/or stochastic rules. Each operating mode is indexed by means of a *discrete state*, while a *continuous state* is used for the description of the system dynamics within each discrete mode. Hybrid models are useful to describe complex systems in many application areas, and mostly in automotive [9], fault detection [47], [50], [43], and target tracking [10], [13] problems. Most authors indicate as *switching hybrid systems* those systems where the switching among the discrete modes depends also on spatial conditions on the continuous state. Conversely, the attribute *hybrid* is often omitted when the occurrence of the switching does not depend on the continuous state. Among switching systems, an important class is the class of jump Markov linear Systems (JMLS), where the switching sequence is a Markov chain, characterized by a matrix of discrete-states transition probabilities, and each operating mode is described by a linear model [18]. When each linear operating mode is a stochastic system forced by Gaussian noise, then we have Gauss-Markov jump linear systems. This kind of models are intensively used in the problem of tracking maneuvering objects observed through radar [10].

The problem of state observation/estimation in hybrid systems is currently widely investigated, in both deterministic and stochastic settings. For the case of deterministic hybrid systems, where the switching process is not stochastically characterized, see [46] and [8]. The state estimation problem for JMLS is investigated, among other papers, in [1], [12], [17], [21], [22], [27], [32] for the discrete-time case, and in [29], [33], [45], [24], [49] for the continuous-time case. All papers, with the exception of [17] and [49], deal explicitely with stochastic systems affected by additive Gaussian noises. The estimation problem for stochastic switching systems was first formulated in [1], where only the case of switching noise covariances was considered. The authors pointed out the complexity of the exact solution of the problem and proposed an approximate solution. A significant change in the state of the art of approximate Bayesian filtering for systems with Markov switching is given in [12], where the Interacting Multiple Model (IMM) was introduced. Such clever idea is at the basis of several subsequent papers, including the present one. In [22] the equations of the exact optimal filter are derived, that include the computation of the likelihood ratios of all the possible switching sequences. The authors point out that the number of admissible switching sequences grows geometrically with time, and therefore the real-time implementation of the exact filter is practically unfeasible. However, the filter presented in [22], together with some criteria and algorithms for pruning less probable switching sequences, is the basis for the construction of suboptimal filters of constant complexity over time [23]. Iterative estimation algorithms of JMLS over a fixed time interval (smoothers) have been analyzed in [27], [32]. For the same problem in [21] approximate estimators, based on stochastic sampling algorithms, have been investigated. The problem with smoothers is that they operate on a fixed time-interval, an can not be directly applied for real-time recursive estimation. In [6] a receding horizon strategy is devised for the estimation of switching systems without a stochastic characterization of the switching sequence. The best estimator among all the affine estimators is presented in [17], in the form of a filter of fixed dimension. No Gaussian assumption on the noise is required for the best affine filter, whose construction is based on a clever use of the characteristic function associated to the Markov jump parameter. The stationary version of the best affine filter has been presented in [19].

An important class of switching systems is the class of output-switching systems, i.e. systems with switching measurements, where the output matrix takes values on a finite set of possible values. First papers on the subject are [34] and [30]. In [4] an observer is derived for deterministic output-switching linear systems, without any stochastic characterization of the unknown switching sequence. The technique proposed in [4] provides an estimation only of the continuous state. The case in which the switching sequence is a Markov chain is considered in [39], where the computation of a Cramer-Rao type bound on the estimation error is investigated. In [48] the output-switching model is used in the problem of sensor failures detection, following the IMM estimation approach.

In this paper the state estimation problem for Markov output-switching systems affected by non Gaussian noises is investigated and solved using the polynomial filtering approach, introduced first in [20] and [14] to solve filtering problems for linear systems affected by non Gaussian noises. The same approach has been successfully used to solve filtering problems for bilinear systems [15], descriptor systems [25] and nonlinear systems [26]. The key idea in this paper is to define a bilinear representation for the Markov output-switching system, and to use it for the construction of a polynomial extended-system, where the degree is chosen by the user, whose state is made of the original discrete and continuous states and of some of their Kronecker products and powers. In this larger state-space the system has still the appearance of a bilinear system. Following [15], the standard Kalman filter equations applied to the polynomial extended system provide the optimal polynomial filter according to the minimum error variance criterion. Both the continuous and discrete states are estimated by the polynomial filter. A somewhat unexpected contribution of this paper is the observation that linear algorithms cannot estimate the discrete state of switching systems, because of a structural lack of observability of the first-order system representation. The use of higher order polynomial filters (at least of degree two) is therefore necessary if the estimate of the discrete state, i.e. of the system operating mode, is required. The optimal polynomial filter can be considered as an extension of the optimal affine filter (polynomial of degree one) presented in [17]. It is important to stress that the polynomial filtering approach does not require the Gaussianity assumption on the noises, whereas all the cited papers, with the only exception of [17], strongly rely on the Gaussian distribution of noises (e.g. for the computation of likelihood ratios or of probabilities of residuals). Moreover, whereas most approaches compute suboptimal estimates by pruning less probable switching sequences, according to some heuristics, the polynomial approach provides an estimate with a precise statistical meaning: it is the best estimate in a closed space of estimators (polynomial processing of measurements of a given degree). In addition, the Riccati equations provide the theoretical covariance of the estimation error, a precious information for assessing the quality of the estimate.

As a motivating application, the problem of the reconstruction of the transmitted data in a time-varying fast-fading digital communication channel has been formulated and solved as a problem of state estimation of a Markov output-switching system. Due to the random variability of the cannel, in this application the simultaneous estimation of both the transmitted data and the channel parameters has to be performed in real-time.

The paper is organized as follows. In Section 2 the model of the Markov output-switching system is presented in a standard form, and then is recasted into a particular bilinear model. In Section 3 the model presented in Section 2 is used for the construction of the polynomial filter of degree ν. Some observability and detectability properties of the filter are discussed in Section 4. In Section 5 the application of the filter to the digital-communication problem of simultaneous estimation of both the transmitted data and channel parameters is investigated.

Simulation results are presented in Section 6, and Conclusions are discussed in Section 7. An Appendix follows, where some formulas used throughout the paper are proved.

2. The stochastic switching-output system model

The switching-output model considered in this paper is the following:

$$\begin{aligned} x(k+1) &= \Lambda x(k) + N_f(k), \qquad x(0) = x_0, \qquad k \in \mathbb{Z}^+ \\ y(k) &= \Gamma_{\mu(k)} x(k) + N_g(k), \end{aligned} \tag{2.0.1}$$

where $x(k) \in \mathbb{R}^n$ is the continuous state, $y(k) \in \mathbb{R}^q$ is the measured output, and $\{N_f(k)\}$, $\{N_g(k)\}$ are the state and output noise sequences, respectively. The initial continuous state x_0 is a random vector. The output switching matrix $\Gamma_{\mu(k)}$ takes values in the finite set $\mathcal{A} = \{\Gamma_1, \ldots, \Gamma_m\}$, with $\Gamma_i \in \mathbb{R}^{q \times n}$, according to the switching parameter $\mu(k) \in \mathcal{S}$, where $\mathcal{S} = \{1, \ldots, m\}$. The sequence $\{\mu(k)\}$ is modeled as a Markov chain, described by the $m \times m$ probability transition matrix Π_μ:

$$\big[\Pi_\mu\big]_{ij} = P\big(\mu(k+1) = i | \mu(k) = j\big), \qquad \forall i, j \in \mathcal{S} = \{1, \ldots, m\} \tag{2.0.2}$$

and by its initial probability distribution $p_\mu \in \mathbb{R}^m$, i.e. $P\big(\mu(0) = i\big) = \big[p_\mu\big]_i$, for $i \in \mathcal{S}$. Here and in the following the notations $\big[A\big]_{ij}$, $\big[u\big]_i$ are adopted to denote the (i,j)-entries of a matrix A and the i-th element of a vector u, respectively. Throughout the paper the symbol $\otimes$ will denote the Kronecker product between matrices and the $M^{[i]}$ will denote the i-th Kronecker power of the matrix M (see [15], for a quick survey on the Kronecker product and its properties), and the symbol I_a will denote the identity matrix in $\mathbb{R}^{a \times a}$. The noises $\{N_f(k)\}$ and $\{N_g(k)\}$ are assumed to be stationary sequences of zero-mean mutually independent random vectors (white sequences), non Gaussian, in general. The assumption needed for the construction of a polynomial filter of degree ν is that the noises and the initial continuous state have finite and known moments up to the degree 2ν. Let

$$\zeta^i = \mathbb{E}\{x_0^{[i]}\}, \quad \xi_f^i = \mathbb{E}\{N_f^{[i]}(k)\}, \quad \xi_g^i = \mathbb{E}\{N_g^{[i]}(k)\}, \quad 0 \le i \le 2\nu, \tag{2.0.3}$$

(note that the zero-mean assumption on the noises implies $\xi_f^1 = 0$ and $\xi_g^1 = 0$). Moreover, the initial state x_0, the noises $\{N_f(k)\}$ and $\{N_g(k)\}$ and the Markov sequence $\{\mu(k)\}$ are assumed to be mutually independent.

The following theorem provides an equivalent representation for the sequence of the switching output matrices $\{\Gamma_{\mu(k)}\}$, that will be useful for the definition of polynomial filters. In the following the symbol $\mathcal{B}_m$ will denote the canonical basis of $\mathbb{R}^m$, i.e. $\mathcal{B}_m = \{e_1, \ldots, e_m\}$, where e_j is the j-th column of I_m.

Lemma 2.1. *Consider the output-switching system* (2.0.1). *Let* $\{\theta(k) \in \mathcal{B}_m,\ k \in \mathbb{Z}^+\}$ *be a stochastic sequence that obeys the following stochastic recursive equation:*

$$\theta(k+1) = V(k)\theta(k), \qquad \theta(0) = \theta_0, \tag{2.0.4}$$

where θ_0 *is a random variable taking values in* $\mathcal{B}_m$ *with distribution* p_θ*:*

$$P\big(\theta_0 = e_i\big) = \big[p_\theta\big]_i, \qquad i = 1, \ldots, m, \tag{2.0.5}$$

and $\{V(k)\}$ *is a sequence of* $m\times m$ *random matrices independent of* $\big(\theta_0, x_0, \{N_f(k)\}, \{N_g(k)\}\big)$*, such that each column, denoted* $V_j(k)$*,* $j=1,\dots,m$*, is a sequence of independent random vectors taking values in* $\mathcal{B}_m$*, with distribution* p_{V_j}*:*

$$P\big(V_j(k)=e_i\big)=[p_{V_j}]_i, \qquad i,j=1,\dots,m. \tag{2.0.6}$$

Then, the matrix sequence

$$\{C(\theta(k)\otimes I_n)\}, \qquad \textit{where} \quad C=\big[\Gamma_1\ \cdots\ \Gamma_m\big]\in\mathbb{R}^{q\times(nm)}, \tag{2.0.7}$$

is stochastically equivalent in distribution to $\{\Gamma_{\mu(k)}\}$ *if and only if* $p_\theta=p_\mu$ *and* $[\Pi_\mu]_{ij}=[p_{V_j}]_i$*,* $\forall i,j=1,\dots,m$*. Moreover* $\{C(\theta(k)\otimes I_n)\}$ *is independent of* $\big(x_0,\{N_f(k)\},\{N_g(k)\}\big)$*.*

Proof. With the given assumptions it is easy to show that $\{V(k)\}$ is independent of all $\theta(h)$, with $h\le k$. This is evident by writing $\theta(h)$ as an explicit function of θ_0 and $V(\tau)$, $\tau<h$:

$$\theta(h)=V(h-1)V(h-2)\cdots V(0)\theta_0, \tag{2.0.8}$$

as it easily follows from (2.0.4). Then, considering that, by assumption, $V(k)$ is independent of θ_0 and of all $V(\tau)$, for $\tau\ne k$, it is clear from (2.0.8) that $V(k)$ is independent of all $\theta(h)$, for $h\le k$. In order to prove that the sequence $\{\theta(k)\}$ is a Markov chain, consider the probabilities of the events $\big\{\theta(k+1)=e_i\big\}$, $i=1,\dots,m$, conditioned to the knowledge of all the values of $\theta(0),\dots,\theta(k)$. Thanks to the model (2.0.4) it is

$$P\big(\theta(k+1)=e_i\,|\,\theta(0),\dots,\theta(k)\big)=P\big(V(k)\theta(k)=e_i\,|\,\theta(0),\dots,\theta(k)\big). \tag{2.0.9}$$

The computation of the conditional probability (2.0.9) for $\theta(k)=e_j$ gives

$$\begin{aligned}P\big(\theta(k+1)=e_i\,|\,\theta(0),\dots,\theta(k)\big)\Big|_{\theta(k)=e_j}&=P\big(V(k)e_j=e_i\,|\,\theta(0),\dots,\theta(k)\big)\Big|_{\theta(k)=e_j}\\&=P\big(V(k)e_j=e_i\big),\end{aligned} \tag{2.0.10}$$

where the last passage is due to the independence of $V(k)$ of all $\theta(h)$ with $h\le k$. Note also that

$$P\big(V(k)e_j=e_i\big)=P\big(V(k)\theta(k)=e_i\,|\,\theta(k)=e_j\big)=P\big(\theta(k+1)=e_i\,|\,\theta(k)=e_j\big). \tag{2.0.11}$$

The comparison of (2.0.10) and (2.0.11) gives

$$P\big(\theta(k+1)=e_i\,|\,\theta(0),\dots,\theta(k)\big)\Big|_{\theta(k)=e_j}=P\big(\theta(k+1)=e_i\,|\,\theta(k)=e_j\big). \tag{2.0.12}$$

The identity (2.0.12) holds for any pair $i,j=1,\dots,m$, and therefore can be equivalently written as

$$P\big(\theta(k+1)|\theta(0),\dots,\theta(k)\big)=P\big(\theta(k+1)|\theta(k)\big), \tag{2.0.13}$$

which proves that the sequence $\{\theta(k)\}$ is a Markov chain.

By definition it is $[\Pi_\theta]_{ij}=P\big(\theta(k+1)=e_i\,|\,\theta(k)=e_j\big)$, and thanks to (2.0.11), it is

$$[\Pi_\theta]_{ij}=P\big(V(k)e_j=e_i\big)=P\big(V_j(k)=e_i\big)=[p_{V_j}]_i. \tag{2.0.14}$$

Consider now the sequence $\big\{C\big(\theta(k)\otimes I_n\big)\big\}$. It is such that

$$C\big(\theta(k)\otimes I_n\big)\Big|_{\theta(k)=e_i}=\big[\Gamma_1\ \cdots\ \Gamma_m\big]\begin{bmatrix}O_{(i-1)n\times n}\\ I_n\\ O_{(m-i)n\times n}\end{bmatrix}=\Gamma_i, \tag{2.0.15}$$

and therefore it is a Markov chain taking values in $\mathcal{A} = \{\Gamma_1, \ldots, \Gamma_m\}$ with the same transition probabilities matrix Π_θ and initial distribution p_θ of the sequence $\{\theta(k)\}$

$$\begin{aligned} &P\Big(C\big(\theta(k+1) \otimes I_n\big) = \Gamma_i \big| C\big(\theta(k) \otimes I_n\big) = \Gamma_j\Big) = P\big(\theta(k+1) = e_i | \theta(k) = e_j\big) = \big[\Pi_\theta\big]_{ij}, \\ &P\Big(C\big(\theta(0) \otimes I_n\big) = \Gamma_i\Big) = P\big(\theta(0) = e_i\big) = \big[p_\theta\big]_i. \end{aligned} \tag{2.0.16}$$

Note that, by definition, $\{\Gamma_{\mu(k)}\}$ is a Markov chain, with a transition probability matrix Π_μ and initial probability distribution p_μ:

$$\begin{aligned} &P\big(\Gamma_{\mu(k+1)} = \Gamma_i | \Gamma_{\mu(k)} = \Gamma_j\big) = P\big(\mu(k+1) = i | \mu(k) = j\big) = \big[\Pi_\mu\big]_{ij}, \\ &P\big(\Gamma_{\mu(0)} = \Gamma_i\big) = P\big(\mu(0) = i\big) = \big[p_\mu\big]_i. \end{aligned} \tag{2.0.17}$$

Comparing (2.0.17) and (2.0.16) it is evident that $\{C(\theta(k) \otimes I_n)\}$ has the same distribution of $\{\Gamma_{\mu(k)}\}$ if and only if $p_\theta = p_\mu$ and $\Pi_\theta = \Pi_\mu$.
The independence of $\{C(\theta(k) \otimes I_n)\}$ and $\big(x_0, \{N_f(k)\}, \{N_g(k)\}\big)$ trivially follows from the assumption of mutual independence of $\{V(k)\}$ and $\big(\theta_0, x_0, \{N_f(k)\}, \{N_g(k)\}\big)$. □

The definition of the Markov process $\{\theta(k)\}$ as in Lemma 2.1, such that $\Pi_\theta = \Pi_\mu$ and $p_\theta = p_\mu$, allows to replace the switching matrix sequence $\{\Gamma_{\mu(k)}\}$ with the sequence $\{C\big(\theta(k) \otimes I_n\}$ in the output equation of the switching-output system (2.0.1). From now on the symbol Π will be used instead of Π_θ or Π_μ. Note that, by the definition of $V(k)$, it follows that $\mathbb{E}\{V(k)\} = \Pi$.

Using $C\big(\theta(k) \otimes I_n\big)$ instead of $\Gamma_{\mu(k)}$ in the output equation in (2.0.1) gives the following identities

$$\begin{aligned} \Gamma_{\mu(k)} x(k) &= C\big(\theta(k) \otimes I_n\big) x(k) = C\big(\theta(k) \otimes I_n\big)\big(1 \otimes x(k)\big) \\ &= C\Big(\big(\theta(k) \cdot 1\big) \otimes \big(I_n \cdot x(k)\big)\Big) = C\big(\theta(k) \otimes x(k)\big), \end{aligned} \tag{2.0.18}$$

where the property $(AB) \otimes (CD) = (A \otimes C)(B \otimes D)$ of the Kronecker product has been used. For what follows, it is useful to define the zero-mean random sequence $\{\mathcal{V}(k)\}$ as

$$\mathcal{V}(k) = V(k) - \mathbb{E}\{V(k)\} = V(k) - \Pi. \tag{2.0.19}$$

The substitution $V(k) = \Pi + \mathcal{V}(k)$ in equation (2.0.4) gives:

$$\theta(k+1) = \Pi\theta(k) + \mathcal{V}(k)\theta(k). \tag{2.0.20}$$

Remark 1. *According to the above considerations, the following state space representation for the stochastic switching-output system* (2.0.1) *holds:*

$$\begin{aligned} x(k+1) &= \Lambda x(k) + N_f(k), \\ \theta(k+1) &= \Pi\theta(k) + \mathcal{V}(k)\theta(k), \\ y(k) &= C\big(\theta(k) \otimes x(k)\big) + N_g(k). \end{aligned} \tag{2.0.21}$$

where the zero-mean sequence $\{\mathcal{V}(k)\}$ *can be regarded as a white multiplicative noise, non-Gaussian because of its discrete probability distribution.*

3. The Polynomial Filter

According to the model equation (2.0.21), the state dynamics of the switching-output model is described by a bilinear system, i.e. a system forced by both additive and multiplicative white noises sequences, whilst the output is given by a quadratic transformation of the state plus additive white noise. In this Section it is shown how to represent such a nonlinear system through a standard bilinear model (linear state and output dynamics forced by multiplicative noise). The reason for finding a bilinear representation is that, as shown in [14], [15], [25] [26], this is the main step for the construction of a polynomial filter.

It is well known that the optimal solution of the minimum variance filtering problem is given by the state expectation conditioned by all the measurements up to the current time. This coincides with the projection $\boldsymbol{P}$ of the state onto the linear space $\mathcal{B}(Y_k)$ of all the Borel functions of the measurements:

$$\hat{x}(k) = \mathbb{E}\big[x(k)|Y_k\big] = \boldsymbol{P}\big[x(k)|\mathcal{B}(Y_k)\big], \tag{3.0.22}$$

with $Y_k = \begin{bmatrix} y(0) & \cdots & y(k) \end{bmatrix}^T$. In the linear Gaussian case the optimal estimate is an affine transformation of the measurements, recursively implemented by the Kalman Filter [7]. Unfortunately, in the non Gaussian case there is not a simple characterization of the conditional expectation, and therefore it is worthwhile to consider suboptimal estimates which have a simpler mathematical structure. The simplest suboptimal estimate is the optimal affine one. It consists in projecting the state $x(k)$ onto the subspace $L(Y_k) = \text{span}\{1, Y_k\}$ of all the affine transformations of the output [7]. For linear non-Gaussian systems the optimal affine estimate is achieved by the Kalman filter. Suboptimal estimates with better performances (smaller error variance) can be obtained by projecting the state onto subspaces larger than $L(Y_k)$, like subspaces of polynomial transformations of the measurements [14], [15]. In more detail, the subspace here considered is the class of polynomial transformations of the measurements of a chosen degree ν, defined as

$$L(Y_k^\nu) = \text{span}\left\{1, Y^\nu(0), \ \cdots, \ Y^\nu(k)\right\}, \tag{3.0.23}$$

where,

$$Y_k^\nu = \begin{bmatrix} Y^\nu(0) \\ \vdots \\ Y^\nu(k) \end{bmatrix}, \quad Y^\nu(h) = \begin{bmatrix} \mathcal{Y}_1(h) \\ \vdots \\ \mathcal{Y}_\nu(h) \end{bmatrix}, \quad \mathcal{Y}_i(h) = y^{[i]}(h), \quad \begin{matrix} i = 1, \dots, \nu, \\ h = 0, \dots, k. \end{matrix} \tag{3.0.24}$$

If the output sequence $\{y(k)\}$ is such that $\mathbb{E}\big[\|y^{[i]}(h)\|^2\big] < \infty$, for $i = 1, \dots, \nu$, then the sequence of the polynomial extended output $Y^\nu(k)$ has finite second order moments and $L(Y_k^\nu)$ is a Hilbert space. It follows that the optimal polynomial estimate of $x(k)$ can be computed as the projection

$$\hat{x}_\nu(k) = \boldsymbol{P}\big[x(k)|L(Y_k^\nu)\big]. \tag{3.0.25}$$

For the construction of a polynomial filter, the definition of a bilinear generation model for the sequence $Y^\nu(k)$ is required (polynomial extended model, see [14], [15]).

Differently from the general case discussed in [15], the bilinear state-transition model in (2.0.21) has the peculiarity that the state component $\theta(k)$ takes values in $\mathcal{B}_m$, the canonical basis in $\mathbb{R}^m$. From this, it follows that the Kronecker powers of $\theta(k)$ are linear functions of $\theta(k)$ (i.e., $\theta^{[h]}(k) = E_h\theta(k)$, as shown in Lemma 8.1 in the Appendix) an this property allows a different construction of the polynomial extended state, with a smaller dimension with respect to the general case of bilinear systems, presented in [15]. The definition of the polynomial extended state is the following

$$X^\nu(k) = \begin{bmatrix} X_0(k) \\ \vdots \\ X_\nu(k) \end{bmatrix}, \qquad \begin{array}{c} X_j(k) = \theta(k) \otimes x^{[j]}(k) \in \mathbb{R}^{m \cdot n^j}, \\ j = 0, 1, \ldots, \nu. \end{array} \tag{3.0.26}$$

In this section it is shown that, suitably exploiting the model (2.0.21) and the properties of the Kronecker products and powers, the sequences $\{X^\nu(k)\}$ and $\{Y^\nu(k)\}$ obey difference equations of the type:

$$\begin{aligned} X^\nu(k+1) &= \boldsymbol{A}^\nu X^\nu(k) + \mathcal{F}(k), \\ Y^\nu(k) &= \boldsymbol{C}^\nu X^\nu(k) + \mathcal{G}(k), \end{aligned} \tag{3.0.27}$$

with $\boldsymbol{A}^\nu$ and $\boldsymbol{C}^\nu$ suitably defined matrices (see Lemmas 3.1 and 3.2, and equations (3.0.46)), and

$$\begin{aligned} \mathcal{F}(k) &= \widetilde{\mathcal{F}}\big(k, X^\nu(k), N_f(k)\big), \\ \mathcal{G}(k) &= \widetilde{\mathcal{G}}\big(k, X^\nu(k), N_g(k)\big), \end{aligned} \tag{3.0.28}$$

suitably defined stochastic sequences, denoted extended noises (the block-components of $\mathcal{F}(k)$ and $\mathcal{G}(k)$, as indicated in (3.0.47), will be defined in (3.0.37) and (3.0.43), respectively). As it can be understood by the definitions, the functions $\widetilde{\mathcal{F}}$, $\widetilde{\mathcal{G}}$, are such that $X^\nu(k)$ multiplies the noises N_f, N_g and their powers up to order ν, in a way that $\mathcal{F}(k)$ and $\mathcal{G}(k)$ result to be uncorrelated stochastic sequences (see Lemma 3.4). Using the model (3.0.27) the original state variables $x(k)$ and $\theta(k)$ are obtained as the following linear functions of the extended state $X^\nu(k)$

$$x(k) = \mathcal{M}_n X^\nu(k), \qquad \theta(k) = \mathcal{T}_n X^\nu(k), \tag{3.0.29}$$

where

$$\begin{aligned} \mathcal{M}_n &= \big[O_{n\times m} \quad \mathcal{M} \quad O_{n\times m(n^2+\cdots+n^\nu)}\big], \\ \mathcal{M} &= \big[I_n \cdots I_n\big] \in \mathbb{R}^{n\times mn}, \\ \mathcal{T}_n &= \big[I_m \quad O_{m\times m(n+\cdots+n^\nu)}\big]. \end{aligned} \tag{3.0.30}$$

The Kalman filter applied to system (3.0.27) provides an estimate $\widehat{X}^\nu(k)$ that is the projection of $X^\nu(k)$ on the space $L(Y_k^\nu)$ defined in (3.0.23), i.e. $\widehat{X}^\nu(k) = \boldsymbol{P}\big[X^\nu(k)|L(Y_k^\nu)\big]$, and therefore coincides with the optimal ν-degree polynomial estimate of $X^\nu(k)$ (polynomial function of the original output $y(k)$). Given the identities (3.0.29), the polynomial estimates of the original state components $x(k)$ and $\theta(k)$ are given by

$$\begin{aligned} \hat{x}_\nu(k) &= \boldsymbol{P}\big[\mathcal{M}_n X^\nu(k)|L(Y_k^\nu)\big] = \mathcal{M}_n \boldsymbol{P}\big[X^\nu(k)|L(Y_k^\nu)\big] = \mathcal{M}_n \widehat{X}^\nu(k), \\ \hat{\theta}_\nu(k) &= \boldsymbol{P}\big[\mathcal{T}_n X^\nu(k)|L(Y_k^\nu)\big] = \mathcal{T}_n \boldsymbol{P}\big[X^\nu(k)|L(Y_k^\nu)\big] = \mathcal{T}_n \widehat{X}^\nu(k). \end{aligned} \tag{3.0.31}$$

Remark 2. *In general, the polynomial estimate $\hat{\theta}_\nu(k)$ of $\theta(k)$ does not belong to $\mathcal{B}_m$. Note that also the conditional expectation $\hat{\theta}(k) = \mathbb{E}\{\theta(k)|Y_k\}$, that is the minimum variance estimate, does not belong to $\mathcal{B}_m$. A strategy for exploiting $\hat{\theta}_\nu(k)$ for the estimation of the output-sensor mode $\Gamma_{\mu(k)}$ is the following:*

i) take as estimate $\tilde{\theta}(k)$ of $\theta(k)$ the element in $\mathcal{B}_m$ that is the closest one to $\hat{\theta}_\nu(k)$:

$$\tilde{\theta}(k) = e_{\hat{h}(k)}, \quad \textit{where} \quad \hat{h}(k) = \underset{h=1,\ldots,m}{\arg\max} \ \big(e_h^T \hat{\theta}_\nu(k)\big); \tag{3.0.32}$$

ii) compute $\widehat{\Gamma}_{\mu(k)}$ *as:*

$$\widehat{\Gamma}_{\mu(k)} = \Gamma_{\hat{h}(k)} = C\big(\tilde{\theta}(k) \otimes I_n\big). \tag{3.0.33}$$

It can be proved that this strategy applied to the conditional expectation $\hat{\theta}(k)$ *provides the Maximum A Posteriori estimate (MAP) of* $\Gamma_{\mu(k)}$*. This is true because* $\mathbb{E}\{\theta(k)|Y_k\}$ *coincides with the conditional (a posteriori) distribution of* $\theta(k)$*.*

The following two identities, proved in Lemma 8.2 in the Appendix, are important for proving the recurrence equation (3.0.27) and for the construction of the matrices $\boldsymbol{A}^\nu$ and $\boldsymbol{C}^\nu$ and of the functions $\widetilde{\mathcal{F}}$, $\widetilde{\mathcal{G}}$:

$$X_j^{[h]}(k) = \Theta_n^{h,j} X_{j\cdot h}(k), \qquad X_i(k) \otimes X_j(k) = \Xi_{i,j} X_{i+j}(k), \tag{3.0.34}$$

where $\Theta_n^{h,j}$ and $\Xi_{i,j}$ are constant matrices defined in Lemma 8.2. Thanks to the identities (3.0.34), the polynomial extended output $Y^\nu(k)$ can be expressed as a linear function of $X^\nu(k)$. This happens because $y(k)$ is a linear function of $X_1(k)$ (see eq.'s (2.0.21) and (3.0.26)), and thanks to the identity $X_1^{[i]}(k) = \Theta_n^{i,1} X_i(k)$, it follows that $Y^\nu(k)$ is a linear function of $X_i(k)$, $i = 1, \ldots, \nu$.

The following Lemmas show how the matrices and the noise sequences in (3.0.27) are obtained.

Lemma 3.1. *The components* $X_j(k)$ *of the extended state* $X^\nu(k)$ *as defined in* (3.0.26) *obey the equations:*

$$X_j(k+1) = \sum_{i=0}^{j} \boldsymbol{A}_{ji} X_i(k) + \mathcal{F}_j(k), \tag{3.0.35}$$

where:

$$\boldsymbol{A}_{ji} = \Pi \otimes J_i^j, \qquad J_i^j = M_i^j(n)\big(\Lambda^{[i]} \otimes \xi_f^{j-i}\big), \tag{3.0.36}$$

and the stochastic sequences $\mathcal{F}_j(k)$ *are defined as*

$$\mathcal{F}_j(k) = \sum_{i=0}^{j} S_i^j(k) X_i(k) \tag{3.0.37}$$

where

$$\begin{aligned} S_i^j(k) &= \big(\Pi \otimes L_i^j(k) + \mathcal{V}(k) \otimes J_i^j + \mathcal{V}(k) \otimes L_i^j(k)\big) \\ L_i^j(k) &= M_i^j(n)\big(\Lambda^{[i]} \otimes (N_f^{[j-i]}(k) - \xi_f^{j-i})\big), \end{aligned} \tag{3.0.38}$$

in which $M_i^j(n)$ *are the matrix coefficients of the Kronecker binomial power formula* (8.0.83)*, reported in the Appendix, and* $\xi_f^i = \mathbb{E}[N_f^{[i]}(k)]$ *are the state noise moments, defined in* (2.0.3)*.*

Proof. According to the Kronecker binomial power formula (8.0.83):

$$\begin{aligned} x^{[j]}(k+1) &= \big(\Lambda x(k) + N_f(k)\big)^{[j]} = \sum_{i=0}^{j} M_i^j(n)\big((\Lambda^{[i]} x^{[i]}(k)) \otimes N_f^{[j-i]}(k)\big) \\ &= \sum_{i=0}^{j} J_i^j x^{[i]}(k) + \sum_{i=0}^{j} L_i^j(k) x^{[i]}(k), \end{aligned} \tag{3.0.39}$$

with J_i^j and $L_i^j(k)$ as in the hypotheses. Then, taking into account the extended state components:

$$X_j(k+1) = \big((\Pi + \mathcal{V}(k))\theta(k)\big) \otimes x^{[j]}(k+1) = \sum_{i=0}^{j} \boldsymbol{A}_{ji} X_i(k) + \sum_{i=0}^{j} S_i^j(k) X_i(k), \tag{3.0.40}$$

with $\boldsymbol{A}_{ji}$ and $S_i^j(k)$ as in the hypotheses. □

Lemma 3.2. *The components $\mathcal{Y}_j(k)$ of the extended output $Y^\nu(k)$ as defined in* (3.0.24) *obey the equations:*

$$\mathcal{Y}_j(k) = \sum_{i=0}^{j} \boldsymbol{C}_{ji} X_i(k) + \mathcal{G}_j(k), \tag{3.0.41}$$

where

$$\boldsymbol{C}_{ji} = M_i^j(q)\big(C^{[i]} \otimes \xi_g^{j-i}\big)\Theta_n^{i,1}, \tag{3.0.42}$$

and the stochastic sequences $\mathcal{G}_j(k)$ are defined as

$$\mathcal{G}_j(k) = \sum_{i=0}^{j} T_i^j(k) X_i(k), \tag{3.0.43}$$

where

$$T_i^j(k) = M_i^j(q)\Big(C^{[i]} \otimes \big(N_g^{[j-i]} - \xi_g^{j-i}\big)\Big)\Theta_n^{i,1} \tag{3.0.44}$$

in which $M_i^j(q)$ are the matrix coefficients of the Kronecker binomial power formula (8.0.83) *and $\xi_g^i = \mathbb{E}[N_g^{[i]}(k)]$ are the output noise moments, defined in* (2.0.3).

Proof. Exploiting the binomial power, according to the Kronecker product properties:

$$\begin{aligned} \mathcal{Y}_j(k) &= \big(CX_1(k) + N_g(k)\big)^{[j]} = \sum_{i=0}^{j} M_i^j(q)\Big(\big(C^{[i]} X_1^{[i]}(k)\big) \otimes N_g^{[j-i]}(k)\Big) \\ &= \sum_{i=0}^{j} \boldsymbol{C}_{ji} X_i(k) + \sum_{i=0}^{j} T_i^j(k) X_i(k) \end{aligned} \tag{3.0.45}$$

with $\boldsymbol{C}_{ji}$ and $T_i^j(k)$ as in the hypotheses. □

Thanks to Lemmas 3.1 and 3.2 the following structure for matrices $\boldsymbol{A}^\nu$ an $\boldsymbol{C}^\nu$, $\nu \geq 1$, in equation (3.0.27) is understood

$$\boldsymbol{A}^\nu = \begin{bmatrix} \boldsymbol{A}_{0,0} & 0 & \dots & 0 \\ \boldsymbol{A}_{1,0} & \boldsymbol{A}_{1,1} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{A}_{\nu,0} & \boldsymbol{A}_{\nu,1} & \dots & \boldsymbol{A}_{\nu,\nu} \end{bmatrix} \qquad \boldsymbol{C}^\nu = \begin{bmatrix} \boldsymbol{C}_{1,0} & \boldsymbol{C}_{1,1} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{C}_{\nu,0} & \boldsymbol{C}_{\nu,1} & \dots & \boldsymbol{C}_{\nu,\nu} \end{bmatrix}, \tag{3.0.46}$$

where the matrices $\boldsymbol{A}_{j,i}$ and $\boldsymbol{C}_{j,i}$ are defined in equations (3.0.36) and (3.0.42), respectively. The extended noise sequences $\{\mathcal{F}(k)\}$ and $\{\mathcal{G}(k)\}$ in (3.0.27) are as follows

$$\mathcal{F}(k) = \begin{bmatrix} \mathcal{F}_0(k) \\ \vdots \\ \mathcal{F}_\nu(k) \end{bmatrix} \qquad \mathcal{G}(k) = \begin{bmatrix} \mathcal{G}_1(k) \\ \vdots \\ \mathcal{G}_\nu(k) \end{bmatrix}, \tag{3.0.47}$$

where the block components $\mathcal{F}_j(k) \in \mathbb{R}^{m \cdot n^j}$ and $\mathcal{G}_j(k) \in \mathbb{R}^{q^j}$ are defined in equations (3.0.37) and (3.0.43), respectively.

The following Lemma is useful for the computation of the covariances of the extended noise sequences. Here, the definition of the stack operator $\mathrm{st}(\cdot)$ and of its inverse $\mathrm{st}^{-1}_{r,s}(\cdot)$, given at the end of the Appendix, is used

Lemma 3.3. *Let $\{\xi_r(k),\ k \in \mathbb{Z}^+\}$, $r \in \mathbb{Z}^+$, $\xi_r \in \mathbb{R}^{n_r}$, be a class of random sequences, defined by:*

$$\xi_r(k) = \sum_{i=0}^{r} \Omega_i(k)\chi_i(k), \tag{3.0.48}$$

with $\{\Omega_i(k)\}$, $i \in \mathbb{Z}^+$, sequences of zero-mean uncorrelated random matrices and $\chi_i(k)$ suitably dimensioned random variables such that $\chi_i(k)$ is independent of $\Omega_j(h)$, $\forall i, j \in \mathbb{Z}^+$, $\forall h \geq k$. Then:

$$\mathbb{E}\big[\xi_r(k)\xi_s^T(h)\big] = 0, \quad \forall r, s \in \mathbb{Z}^+, \quad \forall h \neq k. \tag{3.0.49}$$

and the covariance matrix $\mathbb{E}[\xi_r(k)\xi_s^T(k)]$ is:

$$\mathrm{st}^{-1}_{n_r,n_s}\left(\sum_{i=0}^{r}\sum_{j=0}^{s}\mathbb{E}\big[\Omega_j(k)\otimes\Omega_i(k)\big]\mathbb{E}\big[\chi_j(k)\otimes\chi_i(k)\big]\right). \tag{3.0.50}$$

Moreover, let $\{\zeta_r(k)\}$, $r \in \mathbb{Z}^+$, be another class of random sequences, defined by:

$$\zeta_r(k) = \sum_{i=0}^{r} Z_i(k)\chi_i(k), \tag{3.0.51}$$

with $\{Z_i(k)\}$, $i \in \mathbb{Z}^+$, sequences of zero-mean uncorrelated random matrices, such that $Z_j(h)$ is independent of $\chi_i(k)$, $\forall i, j \in \mathbb{Z}^+$, $\forall h \geq k$ and independent of $\Omega_i(k)$, $\forall i, j \in I$, $\forall h, k \in \mathbb{Z}$. Then:

$$\mathbb{E}\big[\xi_r(k)\zeta_s^T(h)\big] = 0, \quad \forall r, s, k, h \in \mathbb{Z}^+. \tag{3.0.52}$$

Proof. Let $r, s \in \mathbb{Z}^+$. Then

$$\mathbb{E}\big[\xi_r(k)\xi_s^T(h)\big] = \sum_{i=0}^{r}\sum_{j=0}^{s}\mathbb{E}\big[\Omega_i(k)\chi_i(k)\chi_j^T(h)\Omega_j^T(h)\big]. \tag{3.0.53}$$

When $h > k$, thanks to the assumed uncorrelation

$$\mathbb{E}\big[\xi_r(k)\xi_s^T(h)\big] = \sum_{i=0}^{r}\sum_{j=0}^{s}\mathbb{E}\big[\Omega_i(k)\chi_i(k)\chi_j^T(h)\big]\mathbb{E}\big[\Omega_j^T(h)\big] = 0, \tag{3.0.54}$$

so that $\xi_r(h)$ and $\xi_s(k)$ result to be uncorrelated random variables. When $h = k$, equation (3.0.50) is obtained from (3.0.53) simply by applying the property (8.0.95) (see the Appendix) of the stak operator. Equation (3.0.52) is easily verified for any $h, k \in \mathbb{Z}^+$:

$$\begin{aligned}\mathbb{E}\big[\xi_r(k)\zeta_s^T(h)\big] &= \sum_{i=0}^{r}\sum_{j=0}^{s}\mathbb{E}\big[\Omega_i(k)\chi_i(k)\chi_j^T(h)Z_j^T(h)\big] \\ &= \sum_{i=0}^{r}\sum_{j=0}^{s}\mathbb{E}\big[\Omega_i(k)\chi_i(k)\chi_j^T(h)\big]\mathbb{E}\big[Z_j^T(h)\big] = 0.\end{aligned} \tag{3.0.55}$$

□

Lemma 3.4. *The extended noise sequences* $\{\mathcal{F}(k)\}$ *and* $\{\mathcal{G}(k)\}$ *defined in* (3.0.47) *are sequences of zero-mean, uncorrelated random variables, i.e.*

$$\mathbb{E}\big[\mathcal{F}_r(k)\mathcal{G}_s^T(h)\big] = 0, \quad \forall k,h \in \mathbb{Z}, \quad \forall r,s \in \{0,1,\dots,\nu\} \tag{3.0.56}$$

The components of the covariance matrices, defined as $\Psi^{\mathcal{F}}_{r,s}(k) = \mathbb{E}\big[\mathcal{F}_r(k)\mathcal{F}_s^T(k)\big]$ *and* $\Psi^{\mathcal{G}}_{r,s} = \mathbb{E}\big[\mathcal{G}_r(k)\mathcal{G}_s^T(k)\big]$*, are computed as follows:*

$$\Psi^{\mathcal{F}}_{r,s}(k) = \mathrm{st}^{-1}_{m\cdot n^r, m\cdot n^s}\left(\sum_{i=0}^{r}\sum_{j=0}^{s}\Phi^{S,s,r}_{j,i}\Xi_{j,i}\mathbb{E}\big[X_{i+j}(k)\big]\right), \tag{3.0.57}$$

$$\Psi^{\mathcal{G}}_{r,s}(k) = \mathrm{st}^{-1}_{q^r, q^s}\left(\sum_{i=0}^{r}\sum_{j=0}^{s}\Phi^{T,s,r}_{j,i}\Xi_{j,i}\mathbb{E}\big[X_{i+j}(k)\big]\right), \tag{3.0.58}$$

with $\Phi^{S,s,r}_{j,i} = \mathbb{E}\big[S^s_j(k)\otimes S^r_i(k)\big]$ *and* $\Phi^{T,s,r}_{j,i} = \mathbb{E}\big[T^s_j(k)\otimes T^r_i(k)\big]$.

Proof. According to Lemmas 3.1 and 3.2, $\{\mathcal{F}(k)\}$ and $\{G(k)\}$ are zero-mean sequences because $L^j_i(k)$, $\mathcal{V}(k)$ and $T^j_i(k)$ are zero-mean random matrices, independent of $X_i(k)$. Consequently, according to Lemma 3.3, both $\{\mathcal{F}(k)\}$ and $\{\mathcal{G}(k)\}$ are sequences of uncorrelated random vectors. Moreover, matrices $S^j_i(k)$ and $T^j_i(h)$ are independent $\forall h,k \in \mathbb{Z}^+$, as a consequence of the independence of the state and output noises sequences (i.e. $\{N_f(k)\}$ and $\{N_g(k)\}$ respectively). Then, according to Lemma 3.3, the extended state and output noises are uncorrelated random sequences. As for their covariance matrices, from (3.0.50):

$$\Psi^{\mathcal{F}}_{r,s}(k) = \mathrm{st}^{-1}_{m\cdot n^r, m\cdot n^s}\left(\sum_{i=0}^{r}\sum_{j=0}^{s}\mathbb{E}\big[S^s_j(k)\otimes S^r_i(k)\big]\mathbb{E}\big[X_j(k)\otimes X_i(k)\big]\right), \tag{3.0.59}$$

$$\Psi^{\mathcal{G}}_{r,s}(k) = \mathrm{st}^{-1}_{q^r, q^s}\left(\sum_{i=0}^{r}\sum_{j=0}^{s}\mathbb{E}\big[T^s_j(k)\otimes T^r_i(k)\big]\mathbb{E}\big[X_j(k)\otimes X_i(k)\big]\right), \tag{3.0.60}$$

from which equations (3.0.57) and (3.0.58) easily come by applying (3.0.34). □

Remark 3. *Note that in order to compute the noise covariance matrices* $\Psi^{\mathcal{F}}(k)$, $\Psi^{\mathcal{G}}(k)$*, whose components obey the equations* (3.0.57) *and* (3.0.58)*, respectively, the sequence of expectations* $\mathbb{E}\big[X_j(k)\big]$*, for* $j = 0\dots,2\nu$*, is needed. These are the components of the extended state* $\mathbb{E}\big[X^{2\nu}(k)\big]$*, and can be computed as the deterministic evolution of the system*

$$\mathbb{E}\big[X^{2\nu}(k+1)\big] = \boldsymbol{A}^{2\nu}\mathbb{E}\big[X^{2\nu}(k)\big], \tag{3.0.61}$$

properly initialized with $\mathbb{E}\big[X_j(0)\big] = \mathbb{E}\big[\theta(0)\otimes x_0^{[i]}\big] = \mathbb{E}\big[\theta(0)\big]\otimes\mathbb{E}\big[x_0^{[i]}\big] = p_\mu\otimes\zeta^i$ *(recall that* p_μ *is the probability distribution of* $\mu(0)$*, identical to the distribution of* $\theta(0)$*, and* ζ^i *are the moments of the initial state, defined in* (2.0.3)*.*

Theorem 3.5. *The recursive equations of the* ν*-degree polynomial filter for simultaneous state estimation and switching parameter identification are:*

$$\begin{aligned}\hat{x}_\nu(k) &= \mathcal{M}_n\widehat{X}^\nu(k),\\ \hat{\theta}_\nu(k) &= \mathcal{T}_n\widehat{X}^\nu(k),\\ \widehat{X}^\nu(k+1) &= \boldsymbol{A}^\nu\widehat{X}^\nu(k) + \mathcal{K}(k+1)\Big(Y^\nu(k+1) - \boldsymbol{C}^\nu\boldsymbol{A}^\nu\widehat{X}^\nu(k)\Big).\end{aligned} \tag{3.0.62}$$

The gain matrix $\mathcal{K}(k)$ is computed through the following Riccati equations:

$$\begin{aligned}\mathcal{P}_P(k+1) &= \boldsymbol{A}^{\nu}\mathcal{P}(k)\boldsymbol{A}^{\nu T} + \Psi^{\mathcal{F}}(k)\\ \mathcal{P}(k) &= \mathcal{P}_P(k) - \mathcal{K}(k)\boldsymbol{C}^{\nu}\mathcal{P}_P(k)\\ \mathcal{K}(k) &= \mathcal{P}_P(k)\boldsymbol{C}^{\nu T}\Big(\boldsymbol{C}^{\nu}\mathcal{P}_P(k)\boldsymbol{C}^{\nu T} + \Psi^{\mathcal{G}}(k)\Big)^{\dagger}\end{aligned} \tag{3.0.63}$$

The switching output matrix is estimated as discussed in Remark 2.

Proof. The estimates $\hat{x}_{\nu}(k)$ and $\hat{\theta}_{\nu}(k)$ are those reported in eq.s (3.0.31). The filter equations are those of the classical Kalman filter applied to the system (3.0.27), that has a multiplicative noise structure (see the expressions of $\mathcal{F}(k)$ and $\mathcal{G}(k)$ in Lemmas 3.1 and 3.2). The use of the Kalman algorithm on the polynomial extended system to achieve optimal polynomial filtering of systems with multiplicative noise has been already demonstrated in [15]. $\square$

Remark 4. *The covariances of the errors of the ν-degree polynomial state estimates can be computed from the Riccati sequence $\mathcal{P}(k)$ as follows*

$$\begin{aligned}\operatorname{cov}\Big(x(k) - \hat{x}_{\nu}(k)\Big) &= \mathcal{M}_n\mathcal{P}(k)\mathcal{M}_n^T,\\ \operatorname{cov}\Big(\theta(k) - \hat{\theta}_{\nu}(k)\Big) &= \mathcal{T}_n\mathcal{P}(k)\mathcal{T}_n^T,\end{aligned} \tag{3.0.64}$$

where the matrices $\mathcal{M}_n$ and $\mathcal{T}_n$ are defined in (3.0.30).

4. Affine vs. polynomial filtering

The best affine estimator is achieved choosing the degree $\nu = 1$ in the polynomial filter (3.0.62), (3.0.63). Taking into account the definitions of Lemmas 3.1 and 3.2, the system matrices for $\nu = 1$ are:

$$\boldsymbol{A}^1 = \begin{bmatrix}\Pi & 0_{m\times mn}\\ 0_{mn\times m} & \Pi\otimes\Lambda\end{bmatrix}, \quad \boldsymbol{C}^1 = \begin{bmatrix}0_{q\times m} & C\end{bmatrix}, \tag{4.0.65}$$

(note that $\boldsymbol{A}^1 \in \mathbb{R}^{m(n+1)\times m(n+1)}$). The rows of the observability matrix of the pair $(\boldsymbol{A}^1, \boldsymbol{C}^1)$ are

$$\boldsymbol{C}^1\big(\boldsymbol{A}^1\big)^k = \Big[0_{q\times m} \quad C\big(\Pi\otimes\Lambda\big)^{k-1}\Big]. \tag{4.0.66}$$

It comes out that the observability matrix is rank deficient (the first m columns are zero) and therefore the pair $(\boldsymbol{A}^1, \boldsymbol{C}^1)$ is structurally unobservable. Actually, the pair $(\boldsymbol{A}^1, \boldsymbol{C}^1)$ is not even detectable, as it can be seen by applying the the *Hautus Lemma* (see e.g. [42]) for $\lambda = 1$ (note that $\lambda = 1$ is always an eigenvalue of matrix Π and, therefore, of matrix $\boldsymbol{A}^1$):

$$\operatorname{rank}\begin{bmatrix}\boldsymbol{A}^1 - I_{m(n+1)}\\ \boldsymbol{C}^1\end{bmatrix} = \operatorname{rank}\begin{bmatrix}\Pi - I_m & 0_{m\times mn}\\ 0_{mn\times m} & \Pi\otimes\Lambda - I_{mn}\\ 0_{q\times m} & C\end{bmatrix} < m(n+1). \tag{4.0.67}$$

It can be shown that the optimal affine filter is not able to improve the *a priori* estimate of the Markov parameter (i.e. $\hat{\theta}_1(k) = \mathbb{E}[\theta(k)]$). On the other hand, the polynomial approach overcomes this structural undetectability. Consider for instance the case of $\nu = 2$ (second order filter):

$$\boldsymbol{A}^2 = \begin{bmatrix} \Pi & 0 & 0 \\ 0 & \Pi \otimes \Lambda & 0 \\ \Pi \otimes \xi_f^2 & 0 & \Pi \otimes \Lambda^{[2]} \end{bmatrix}, \qquad \boldsymbol{C}^2 = \begin{bmatrix} 0 & C & 0 \\ \Theta_n^{0,1} \otimes \xi_g^2 & 0 & C^{[2]}\Theta_n^{2,1} \end{bmatrix}. \tag{4.0.68}$$

Now, the first m columns of the observability matrix are not identically zero, so that the pair $(\boldsymbol{A}^2, \boldsymbol{C}^2)$ is not structurally unobservable.

5. The digital channel model

This section is devoted to illustrate, from a theoretical point of view, how the theory of polynomial estimation of stochastic switching-output systems may find application in a classical problem in the field of digital communications: the problem of the simultaneous signal estimation and parameter identification of a noisy fast-fading digital channel. In a great deal of literature, such problem is solved, under the Gaussian noises assumption, by using the Least Mean Squares [44], [37], the Recursive Mean Squares [16] or the Kalman Filter [35]. Each of these algorithms works in connection with a Viterbi Decoder, implementing the signal-channel deconvolution. More recently, other innovative filtering approaches, such as the Particle Filter [11], have been proposed to deal with this problem, and this demonstrates that, in the field of digital communications, it may be now of some interest to proceed in the comparative testing of new methods, exported from the control systems community. Of course, because of the nature of the present analysis, this is out of the scope of this work.

For the sake of simplicity, the case of real scalar observations is considered here (the extension of the approach to the complex framework, very popular in the communication community, is straightforward). The sequence $\{y(k)\}$ of received data from a multipath fading channel is usually modeled by using the following discrete-time convolution model [36]:

$$y(k) = \sum_{j=0}^{L-1} c(k,j)b(k-j) + N_g(k), \qquad k \in \mathbb{Z} \tag{5.0.69}$$

where $\{b(k)\}$ is the sequence of transmitted data, taking values on a given real alphabet $\mathcal{H}$ of size s. The symbols $c(k,j)$, $j = 0, \dots, L-1$, denote the samples (*taps*) of the time-varying impulse response of a channel of length L at time k. The integer L is said to be the *memory* of the channel. $\{N_g(k)\}$ is a zero-mean white noise sequence at the receiver. Let $\boldsymbol{b}(k) = \begin{bmatrix} b(k) & \cdots & b(k-L+1) \end{bmatrix}^T$ be the vector collecting the sequence of transmitted symbols in the time interval $(k-L, k]$. $\boldsymbol{b}(k)$ takes values in $\mathcal{H}^L$, the set of all s^L L-ples of symbols in $\mathcal{H}$. Let $m = s^L$, and let β_i, $i = 1, \dots, m$, denote all the L-ples in $\mathcal{H}^L$. Whether the sequence of elementary symbols $\{b(k)\}$ is made of independent random variables or it is a Markov chain, the sequence $\{\boldsymbol{b}(k)\}$ can be modeled as a Markov chain [36], whose probability transition matrix Π depends on the distribution of $\{b(k)\}$. The initial distribution of $\boldsymbol{b}(0)$ depends on the distribution of $b(k)$, for $k \in (-L, 0]$.

Denoting $x(k) = [c(k,0), \cdots, c(k, L-1)]^T \in \mathbb{R}^L$ the vector of the channel impulse-response samples, the measurement equation in (5.0.69) can be put in the more compact form:

$$y(k) = \boldsymbol{b}^T(k)x(k) + N_g(k). \tag{5.0.70}$$

According to [36], the random variability of the channel taps in $\{x(k)\}$ can be modeled as a discrete-time process

$$x(k+1) = \Lambda x(k) + N_f(k), \qquad x(0) = \boldsymbol{c}_0, \tag{5.0.71}$$

with $\{N_f(k)\}$ a zero-mean white sequence, independent of $\{N_g(k)\}$ and $\boldsymbol{c}_0$.

Note that system (5.0.71) endowed with the output equation (5.0.70), is a Markov switching-output system of the type (2.0.1), where the ouptut matrices Γ_i are the transmitted sequences β_i^T. Thus, the simultaneous estimation of $x(k)$ (the channel) and of $\boldsymbol{b}(k)$ (the transmitted signal) can be achieved by applying the polynomial filter of Theorem 3.5. Differently from other approaches studied in the specific literature, this kind of algorithm provides in real time both the signal and channel estimates without requiring the a priori knowledge of part of the signal (the preamble), required by most of the decoding approaches for the channel estimation.

As an example, consider a transmitted sequence $\{b(k)\}$ made of independent identically distributed random variables taking values on an alphabet $\mathcal{H}$ of two symbols, $\mathcal{H} = \{b_1, b_2\}$, where b_1 and b_2 are real numbers. Let

$$P\{b(k) = b_1\} = \rho, \qquad P\{b(k) = b_2\} = 1 - \rho, \tag{5.0.72}$$

for some $\rho \in (0,1)$. Let $b(k)$ be transmitted through a channel with memory $L = 2$. The possible sequences of symbols of length 2 makes up the alphabet $\mathcal{H}^2 = \{\beta_1, \beta_2, \beta_3, \beta_4\}$, where

$$\beta_1 = \begin{bmatrix} b_1 \\ b_1 \end{bmatrix}, \quad \beta_2 = \begin{bmatrix} b_1 \\ b_2 \end{bmatrix}, \quad \beta_3 = \begin{bmatrix} b_2 \\ b_1 \end{bmatrix}, \quad \beta_4 = \begin{bmatrix} b_2 \\ b_2 \end{bmatrix}. \tag{5.0.73}$$

The construction of the 4×4 matrix of transition probabilities, whose components are

$$[\Pi]_{ij} = P\{\boldsymbol{b}(k+1) = \beta_i \,|\, \boldsymbol{b}(k) = \beta_j\}, \tag{5.0.74}$$

proceeds by considering that, since $[\boldsymbol{b}(k+1)]_2 = [\boldsymbol{b}(k)]_1 = b(k)$, all transitions between states such that $[\beta_i]_2 \neq [\beta_j]_1$ are impossible. It follows

$$[\Pi]_{ij} = \begin{cases} 0 & \text{if } [\beta_i]_2 \neq [\beta_j]_1, \\ P\big\{b(k+1) = [\beta_i]_1\big\} & \text{if } [\beta_i]_2 = [\beta_j]_1. \end{cases} \tag{5.0.75}$$

Thus

$$\begin{aligned} &[\Pi]_{13} = [\Pi]_{14} = [\Pi]_{33} = [\Pi]_{34} = 0, \\ &[\Pi]_{21} = [\Pi]_{22} = [\Pi]_{41} = [\Pi]_{42} = 0, \\ &[\Pi]_{11} = [\Pi]_{12} = [\Pi]_{23} = [\Pi]_{24} = P\{b(k+1) = b_1\} = \rho, \\ &[\Pi]_{31} = [\Pi]_{32} = [\Pi]_{43} = [\Pi]_{44} = P\{b(k+1) = b_2\} = 1 - \rho. \end{aligned} \tag{5.0.76}$$

The initial probabilities are

$$\begin{aligned} &P\{\boldsymbol{b}(0) = \beta_1\} = P\{b(0) = b_1, b(-1) = b_1\} = \rho^2, \\ &P\{\boldsymbol{b}(0) = \beta_2\} = P\{\boldsymbol{b}(0) = \beta_3\} = \rho(1-\rho), \\ &P\{\boldsymbol{b}(0) = \beta_4\} = P\{b(0) = b_2, b(-1) = b_2\} = (1-\rho)^2. \end{aligned} \tag{5.0.77}$$

6. Simulation Results

In this section the polynomial filters of degree 1 and 2 (first order and second order polynomial filters) are numerically tested and compared in the problem of simultaneous data and channel estimation described in the previous section (as pointed out in the paper, the first order filter coincides with the standard Kalman filter).

The alphabet considered is $\mathcal{H} = \{1, 2\}$, and equiprobability of the symbols is assumed, so that $\rho = 0.5$. A channel of memory $L = 2$ is considered, so that the matrix of transition probabilities and the initial probability distribution are

$$\Pi = \begin{bmatrix} 0.5 & 0.5 & 0 & 0 \\ 0 & 0 & 0.5 & 0.5 \\ 0.5 & 0.5 & 0 & 0 \\ 0 & 0 & 0.5 & 0.5 \end{bmatrix}, \qquad p = \begin{bmatrix} 0.25 \\ 0.25 \\ 0.25 \\ 0.25 \end{bmatrix}. \tag{6.0.78}$$

The two taps $c(k, 1)$ and $c(k, 2)$ in the channel model (5.0.69) are assumed to evolve independently, and therefore in the model (5.0.71) the two components of the state noise $N_f(k)$ are assumed independent and the state transition matrix Λ is chosen diagonal, as follows:

$$\Lambda = \begin{bmatrix} 0.7 & 0 \\ 0 & 0.6 \end{bmatrix}. \tag{6.0.79}$$

In order to test the filtering algorithms in a strongly non Gaussian framework, both the state and output noises distributions are chosen asymmetric, with finite probability masses. The results of two simulations are reported, with two different choices of state and output noises distributions:

$$\begin{array}{lrr} \text{SIMULATION 1:} & P\{[N_f(k)]_1 = -0.5\} = 0.8, & P\{[N_f(k)]_1 = 2\} = 0.2, \\ & P\{[N_f(k)]_2 = -1/3\} = 0.9, & P\{[N_f(k)]_2 = 3\} = 0.1, \\ & P\{N_g(k) = -0.5\} = 0.8, & P\{N_g(k) = 2\} = 0.2. \end{array} \tag{6.0.80}$$

$$\begin{array}{lrr} \text{SIMULATION 2:} & P\{[N_f(k)]_1 = -0.5\} = 0.8, & P\{[N_f(k)]_1 = 2\} = 0.2, \\ & P\{[N_f(k)]_2 = -6/99\} = 0.99, & P\{[N_f(k)]_2 = 6\} = 0.01, \\ & P\{N_g(k) = -10/99\} = 0.99, & P\{N_g(k) = 10\} = 0.01. \end{array} \tag{6.0.81}$$

The noise distributions in Simulation 2 are chosen more asymmetric than in Simulation 1 in order to explore the improvement of the quadratic filter w.r.t. the linear one in the case of strongly non gaussian noises.

Note that all the needed moments ξ_f^i and ξ_g^i can be easily computed from the given discrete distributions.

From the estimate of the symbol $\boldsymbol{b}(k)$, the estimate of the second component $[\boldsymbol{b}(k)]_2$ is chosen as an estimate of the elementary symbol $b(k-1)$, thus achieving a one-step smoothing effect of the estimate. The estimation algorithms are evaluated in terms of the Bit Error Rate (BER).

The simulations performed confirm what has been discussed in Section 4: the polynomial filter of degree one (best affine estimator) does not allow to estimate the Markov parameter (the transmitted data, in this application). Indeed, in all the simulations performed the BER obtained using the first order filter is about 50%. Using the polynomial filter of degree 2, in the simulation 1 the BER is about 40 %, while in the simulation 2 the BER is about 33 %.

The results on the estimation of the taps of the time-varying channel (the continuous-state), are expressed in terms of theoretical error variances, provided by the Riccati matrices $\mathcal{P}(k)$ of equations (3.0.63), (the sample error variances obtained in the numerical simulations are in accordance with the theoretical ones). The obtained error variances are reported below, where

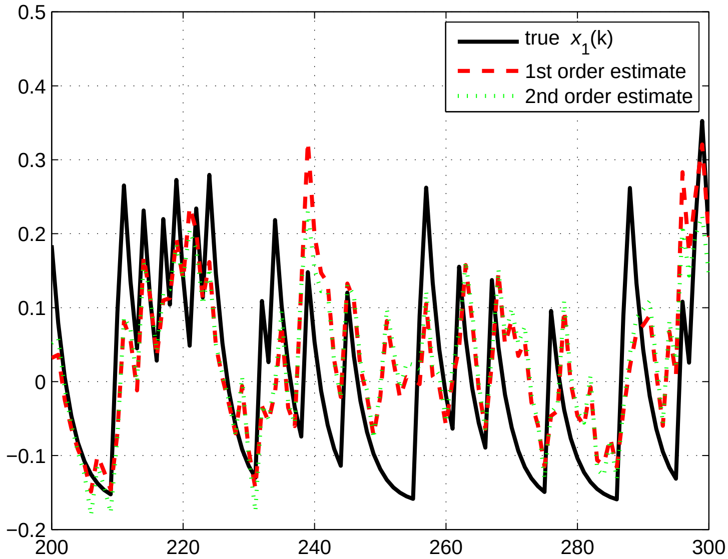

Figure 1: Simulation 1. True and estimated continuous state component $x_1(k)$

$\hat{x}_1(k)$ and $\hat{x}_2(k)$ denote the first and second degree estimates, respectively.

$$
\begin{array}{l}
\textsc{Simulation 1:}\\
\quad \text{Filter of degree 1:}\\
\qquad \mathbb{E}\{([x(k)-\hat{x}_1(k)]_1^2\} = 1.037\cdot 10^{-2}, \qquad \mathbb{E}\{([x(k)-\hat{x}_1(k)]_2^2\} = 9.971\cdot 10^{-3},\\
\quad \text{Filter of degree 2:}\\
\qquad \mathbb{E}\{([x(k)-\hat{x}_2(k)]_1^2\} = 1.006\cdot 10^{-2}, \qquad \mathbb{E}\{([x(k)-\hat{x}_2(k)]_2^2\} = 9.619\cdot 10^{-3},\\
\textsc{Simulation 2:}\\
\quad \text{Filter of degree 1:}\\
\qquad \mathbb{E}\{([x(k)-\hat{x}_1(k)]_1^2\} = 7.263\cdot 10^{-3}, \qquad \mathbb{E}\{([x(k)-\hat{x}_1(k)]_2^2\} = 4.679\cdot 10^{-3},\\
\quad \text{Filter of degree 2:}\\
\qquad \mathbb{E}\{([x(k)-\hat{x}_2(k)]_1^2\} = 5.895\cdot 10^{-3}, \qquad \mathbb{E}\{([x(k)-\hat{x}_2(k)]_2^2\} = 4.282\cdot 10^{-3},
\end{array}
\tag{6.0.82}
$$

In the Simulation 1 the improvement achieved by the second order filter w.r.t. the first order one is of 2.98 % for the first state component and of 3.53 % for the second component. Figures 1 an 2 show the estimation of the channel taps in a window of 100 steps.

In the Simulation 2 the improvement achieved by the second order filter w.r.t. the first order one is of 18.83 % for the first state component and of 8.47 % for the second component. Figures 3 an 4 show the estimation of the channel taps in a window of 100 steps.

7. Conclusions

The use of polynomial filters for the simultaneous estimation of both the continuous and discrete states in a class of stochastic switching systems has been addressed in this paper. The class considered is the set of stochastic linear systems with the output matrices switching in a finite set of possible values, according to a not directly measured discrete Markov sequence (Markov

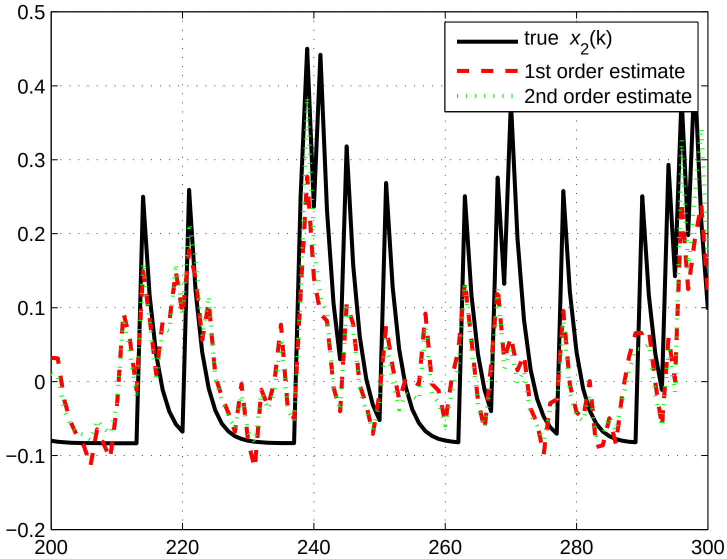

Figure 2: Simulation 1. True and estimated continuous state component $x_2(k)$

switching-output systems). Non Gaussian state and output noises affect the system. The first step for the application of the polynomial filtering approach consists in a suitable transformation of the switching system into a bilinear model, that allows for the construction of the polynomial filter. The construction of the filter of a given degree ν requires the knowlegde of the moments of the noises up to the degree 2ν. It is shown that the filter of degree one (best affine filter) does not solve the discrete-state estimation problem, and therefore the use of higher order filters becomes necessary. As a nonstandard application of the proposed theory, the problem of real time estimation of both the transmitted signal and the samples of the impulse response in a multipath fast-fading digital communication channel is addressed. Differently from other approaches, the use of a polynomial filter does not require the use of training sequences for the channel estimation which, of course, interrupt the data transfer. Simulation results shows that the improvement of the second order filter w.r.t. the first order filter is more evident in the presence of strongly asymmetric noises.

8. Appendix: some useful formulas

Kronecker binomial power formula (see Theorem B.6 in [15]): $\forall a, b \in \mathbb{R}^n$

$$(a+b)^{[h]} = \sum_{k=0}^{h} M_k^h(n)(a^{[k]} \otimes b^{[h-k]}); \tag{8.0.83}$$

where the matrices $M_k^h(n)$ are recursively defined as

$$\begin{aligned} &M_h^h(n) = I_{n^h}, \quad M_h^0(n) = I_{n^h}, \quad \forall h \geq 0, \\ &M_k^h(n) = M_k^{h-1}(n) \otimes I_n + (M_{k-1}^{h-1}(n) \otimes I_n)(I_{n^{k-1}} \otimes G_{h-k}(n)), \quad 0 < k < h., \end{aligned} \tag{8.0.84}$$

where $G_j(n)$ are commutation matrices such that $b^{[j]} \otimes a = G_j(n)(a \otimes b^{[j]})$.

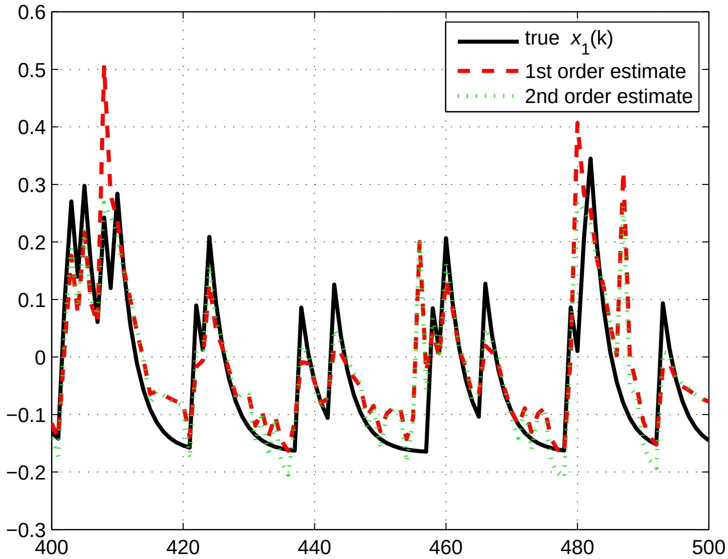

Figure 3: Simulation 2. True and estimated continuous state component $x_1(k)$

Lemma 8.1. *Consider $\theta(t) \in \mathcal{B}_m$ as defined in equation* (2.0.4)*. Let*

$$E_h = \left[e_1^{[h]} \;\cdots\; e_m^{[h]}\right] \in \mathbb{R}^{m^h \times m}, \qquad h \geq 0, \tag{8.0.85}$$

where the columns $e_i \in \mathcal{B}_m$. Then, $\forall k, h \in \mathbb{Z}^+$

$$\theta^{[h]}(k) = E_h\,\theta(k), \tag{8.0.86}$$

Proof. Identity (8.0.86) is verified by noting that the product $E_h e_i$ selects the i-th column of matrix E_h, and therefore $E_h e_i = e_i^{[h]}$. ∎

The following two Lemmas show the computation of the matrices used in equations (3.0.34).

Lemma 8.2. *Let $x \in R^n$, $\theta \in \mathcal{B}_m$, the natural basis of R^m, and let $X_j = \theta \otimes x^{[j]} \in \mathbb{R}^{mn^j}$. Then $\forall i, j, h \in \mathbb{Z}^+$:*

$$X_j^{[h]} = \Theta_{m,n}^{h,j} X_{jh}, \qquad X_i \otimes X_j = \Xi_{m,n}^{i,j} X_{i+j}, \tag{8.0.87}$$

with

$$\begin{aligned}\Theta_{m,n}^{h+1,j} = \left(\Theta_{m,n}^{h,j} \otimes I_{mn^j}\right)\left(I_m \otimes C_{mn^j,n^{jh}}^T\right)\left(E_2 \otimes I_{n^{j(h+1)}}\right),\\ \Theta_{m,n}^{0,j} = E_0 = [1\;1\;\cdots\;1].\end{aligned} \tag{8.0.88}$$

and

$$\Xi_{m,n}^{i,j} = \left(I_m \otimes C_{mn^j,n^i}^T\right)(E_2 \otimes I_{n^{i+j}}). \tag{8.0.89}$$

Proof. Equations (8.0.87) and (8.0.88) are proved by induction. Note first that they are obviously true for $h = 0$, since $E_0\big(\theta \otimes x^{[0]}\big) = [1\;1\;\cdots\;1]\theta = 1$. Now, assume that (8.0.87) is true for a given $h \in \mathbb{Z}^+$ and prove it for $h+1$.

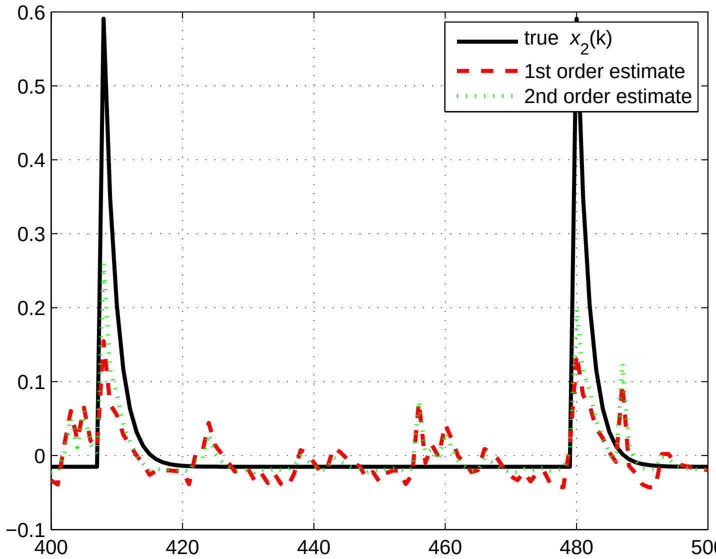

Figure 4: Simulation 2. True and estimated continuous state component $x_2(k)$

By repeatedly applying the Kronecker properties gives

$$\begin{aligned} X_j^{[k+1]} &= X_j^{[k]} \otimes X_j = \left(\Theta_{m,n}^{k,j} X_{jk}\right) \otimes \left(I_{nm^j} X_j\right) = \left(\Theta_{m,n}^{k,j} \otimes I_{mn^j}\right)\left(X_{jk} \otimes X_j\right) \\ &= \left(\Theta_{m,n}^{k,j} \otimes I_{mn^j}\right)\left(\theta \otimes x^{[jk]} \otimes \theta \otimes x^{[j]}\right) \\ &= \left(\Theta_{m,n}^{k,j} \otimes I_{mn^j}\right)\Big((I_m\theta) \otimes \left(x^{[jk]} \otimes (\theta \otimes x^{[j]})\right)\Big) \\ &= \left(\Theta_{m,n}^{k,j} \otimes I_{mn^j}\right)\Big((I_m\theta) \otimes \left(C_{mn^j,n^{jk}}^T\left((\theta \otimes x^{[j]}) \otimes x^{[jk]}\right)\right)\Big) \\ &= \left(\Theta_{m,n}^{k,j} \otimes I_{mn^j}\right)\left(I_m \otimes C_{mn^j,n^{jk}}^T\right)\left(\theta^{[2]} \otimes x^{[j(k+1)]}\right) \end{aligned} \tag{8.0.90}$$

From this, using equation (8.0.86) for $h = 2$, i.e. $\theta^{[2]} = E_2\theta$, gives

$$\begin{aligned} X_j^{[k+1]} &= \left(\Theta_{m,n}^{k,j} \otimes I_{mn^j}\right)\left(I_m \otimes C_{mn^j,n^{jk}}^T\right)\left((E_2\theta) \otimes (I_{n^{j(k+1)}} x^{[j(k+1)]})\right) \\ &= \left(\Theta_{m,n}^{k,j} \otimes I_{mn^j}\right)\left(I_m \otimes C_{mn^j,n^{jk}}^T\right)\left(E_2 \otimes I_{n^{j(k+1)}}\right)\left(\theta \otimes x^{[j(k+1)]}\right)) \\ &= \Theta_{m,n}^{k+1,j} X_{j(k+1)}, \end{aligned} \tag{8.0.91}$$

so that the first of (8.0.87) and (8.0.88) are proved. The second of (8.0.87) is obtained following similar computations:

$$\begin{aligned} X_i \otimes X_j &= \theta \otimes x^{[i]} \otimes \theta \otimes x^{[j]} = (I_m\theta) \otimes \left(x^{[i]} \otimes (\theta \otimes x^{[j]})\right) \\ &= (I_m\theta) \otimes \Big(C_{mn^j,n^i}^T\left((\theta \otimes x^{[j]}) \otimes x^{[i]}\right)\Big) \\ &= \left(I_m \otimes C_{mn^j,n^i}^T\right)\left(\theta^{[2]} \otimes x^{[i+j]}\right) = \left(I_m \otimes C_{mn^j,n^i}^T\right)(E_2 \otimes I_{n^{i+j}})X_{i+j}. \end{aligned} \tag{8.0.92}$$

□

The stack operator applied to a matrix A, denoted st(A), provides a column vector made of all the columns of matrix A one over the other. Thus, if $A \in \mathbb{R}^{r\times m}$, then st$(A) \in \mathbb{R}^{r\cdot m}$ is such that

$$[A]_{i,j} = [\mathrm{st}(A)]_{(j-1)r+i}. \tag{8.0.93}$$

The symbol $\mathrm{st}_{r,m}^{-1}$ denotes the inverse of the stack operator, where the subscripts specify the rows and columns of the resulting matrix, so that if $A \in \mathbb{R}^{r\times m}$, then $\mathrm{st}_{r,m}^{-1}\big(\mathrm{st}(A)\big) = A$. Of course, $\mathrm{st}_{r,m}^{-1}(v)$ makes sense if and only if $v \in \mathbb{R}^{r\cdot m}$. The following property of the stack operator is useful:

$$\mathrm{st}(A \cdot B \cdot C) = (C^T \otimes A)\,\mathrm{st}(B) \tag{8.0.94}$$

from which, as a corollary, $\mathrm{st}(ac^T) = c \otimes a$, where $a, c \in \mathbb{R}^m$. Moreover, if in (8.0.94) $B = ac^T$ and $A \in \mathbb{R}^{r\times m}$ and $C \in \mathbb{R}^{m\times s}$, it follows

$$A \cdot (a\,c^T) \cdot C = \mathrm{st}_{r,s}^{-1}\Big((C^T \otimes A)(c \otimes a)\Big) \in \mathbb{R}^{r\times s}. \tag{8.0.95}$$

The property (8.0.95) is used in Lemmas 3.3 and 3.4.

References

[1] G. A. Ackerson, K. S. Fu, “On State Estimation in Switching Environments,” *IEEE Trans. Automat. Contr.*, Vol. 15, No. 1, pp. 10–17, 1970.

[2] P.J. Antsaklis, “A brief introduction to the theory and applications of hybrid systems,” *Proc. of the IEEE, Special Issue on Hybrid Systems: Theory and Applications,* Vol. 88, No. 7, pp. 879–887, July 2000.

[3] P. Antsaklis and X. Koutsoukos, “Hybrid systems: Review and recent progress,” in *Software-Enabled Control,* T. Samad and G. Balas, Eds., pp. 272–298, IEEE Press, 2003.

[4] M. Babaali, M. Egerstedt, E.W. Kamen, “A direct algebraic approach to observer design under switching measurement equations,” *IEEE Trans. Automat. Contr.*, Vol. 49, No. 11, pp. 2044–2049, 2004.

[5] E. Baccarelli and R. Cusani, “Combined channel estimation and data detection using soft statistics for frequency-selective fast-fading digital links,” *IEEE Trans. Communications.*, Vol. 46, No. 4, pp. 424–427, 1998.

[6] A. Alessandri, M. Baglietto, and G. Battistelli, “Receding-Horizon Estimation for Switching Discrete-Time Linear Systems,” *IEEE Trans. Automat. Contr.*, Vol. 50, No. 11, pp. 1736–1748, 2005.

[7] A. V. Balakrishnan, *Kalman Filtering Theory*, New York: Optimization Software, 1984.

[8] A. Balluchi, L. Benvenuti, M.D. Di Benedetto, A.L. Sangiovanni-Vincentelli, “Design of observers for hybrid systems,” in *Hybrid Systems: Computation and Control,* Vol. 2289 of LNCS, Springer Verlag, pp. 76–89, 2002.

[9] A. Balluchi, L. Benvenuti, M. D. Di Benedetto, C. Pinello, A. L. Sangiovanni-Vincentelli, “Automotive Engine Control and Hybrid Systems: Challenge and Opportunities,” *Proc. of IEEE*, Vol. 88, No. 7, pp. 888–912, 2000.

[10] Y. Bar-Shalom, X.R. Li, and T. Kirubarajan, *Estimation with Applications to Tracking and Navigation,* JohnWiley & Sons, New York, 2001.

[11] T. Bertozzi, D. Le Ruyet, G. Rigal, H. Vu-Thien, "Joint data-channel estimation using the particle filtering on multipath fading channels," 2003. ICT 2003. proc. of 10th Int. Conf. on Telecommunications, ICT2003, Vol. 2 , pp. 1284–1289, Feb. 2003.

[12] H.A.P. Blom and Y. Bar-Shalom, "The interactive multiple model algorithm for systems with Markovian switching coefficients," *IEEE Trans. Automat. Contr.*, Vol. 33, No. 8, pp. 780–783, 1988.

[13] Y. Boers, H. Driessen, "Hybrid state estimation: A target tracking application," *Automatica,* Vol. 38, pp. 2153–2158, 2002.

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

[15] F. Carravetta, A. Germani and M. Raimondi, "Polynomial filtering of discrete-time stochastic linear systems with multiplicative state noise," *IEEE Trans. Automat. Contr.*,, Vol. 42, No. 8, pp. 1106–1126, 1997.

[16] P. Castoldi, R. Raheli and G. Marino, "Efficient trellis search algorithms for adaptive MLSE on fast Rayleigh fading channels," Proc. IEEE Globecom, Nov. 1994.

[17] O.L.V. Costa, "Linear Minimum Mean Square Error Estimation for Discrete-Time Markovian Jump Linear Systems," *IEEE Trans. Automat. Contr.*, Vol. 39, No. 8, pp. 1685–1689, 1994.

[18] O.L.V. Costa, M.D. Fragoso, R.P. Marques, *Discrete-Time Markov Jump Linear Systems,* Series: Probability and its Applications, Springer Verlag, 2005.

[19] O.L.V. Costa and S. Guerra, "Stationary filter for linear minimum mean square error estimator of discrete-time Markovian jump systems," *IEEE Trans. Automat. Contr.*, Vol. 47, No. 8, pp. 1351–1356, Aug. 2003.

[20] A. De Santis, A. Germani, M. Raimondi, "Optimal quadratic filtering of linear discrete-time non-Gaussian systems," *IEEE Trans. Automat. Contr.*, Vol. 40, No. 7, pp. 1274–1278, 1995.

[21] A. Doucet, A. Logothetis, V. Krishnamurthy, "Stochastic Sampling Algorithms for State Estimation of Jump Markov Linear Systems," *IEEE Trans. Automat. Contr.*, Vol. 45, No. 1, pp. 188–202, 2000.

[22] R. J. Elliot, F. Dufour, D. D. Sworder, "Exact Hybrid Filters in Discrete Time," *IEEE Trans. Automat. Contr.*, Vol. 41, No. 12, pp. 1807–1810, 1996.

[23] R.J. Elliott, F. Dufour, W.P. Malcolm, "State and Mode Estimation for Discrete-Time Jump Markov Systems," *SIAM J. Contr. Optim.*, Vol. 44, No. 3, pp. 1081–1104, 2005.

[24] M.D. Fragoso, N.C.S. Rocha, "Stationary Filter For Continuous-Time Markovian Jump Linear Systems," *SIAM J. Contr. Optim.*, Vol. 44, No. 3, pp. 801–815, 2005.

[25] A. Germani, C. Manes, P. Palumbo, "Polynomial filtering for stochastic nongaussian descriptor systems," *IEEE Trans. on Circuits and Systems-I: Regular papers,* Vol. 51, No. 8, pp. 1561–1576, 2004.

[26] A. Germani, C. Manes, P. Palumbo, "Polynomial Extended Kalman Filter," *IEEE Trans. Automat. Contr.*, Vol. 50, No. 12, pp. 2059–2064, dec. 2005.

[27] R.E. Helmick, W.D. Blair, S.A. Hoffman, "Fixed-interval smoothing for Markovian switching systems," *IEEE Trans. on Inf. Theory,* Vol. 41, No. 6, pp. 1845-1855, 1995.

[28] J. Hespanha, "Modeling and Analysis of Stochastic Hybrid Systems," *Proc. of IEE–Control Theory & Applications, Special Issue on Hybrid Systems,* Vol. 153, No. 5, pp. 520–535, 2007.

[29] J.L. Hibey, C.D. Charalambous, "Conditional Densities for Continuous-Time Nonlinear Hybrid Systems with Applications to Fault Detection," *IEEE Trans. Automat. Contr.*, Vol. 44, No. 11, pp. 2164–2169, 1999.

[30] A.G. Jaffer and S. Gupta, "Optimal sequential estimation of discrete processes with Markov interrupted observations," *IEEE Trans. Automat. Contr.*, Vol. 16, No. 5, pp. 471–475, Oct. 1971.

[31] D. Liberzon, *Switching in Systems and Control,* Systems & Control series: Foundations and Applications, Birkhauser, Boston, MA 2003.

[32] A. Logothetis, V. Krishnamurthy, "Expectation maximization algorithms for MAP estimation of Jump Markov Linear Systems," *IEEE Trans. and Signal Processing*, Vol. 47, No. 8, pp. 2139–2156, 1999.

[33] B.M. Miller, W.J. Runggaldier, "Kalman Filtering for Linear Systems with Coefficients Driven by a Hidden Markov Jump Process," *Syst. Contr. Lett.*, No. 31, pp. 93–102, 1997.

[34] N.E. Nahi, "Optimal recursive estimation with uncertain observations," *IEEE Trans. Inform. Theory,* Vol. 15, No. 7, pp. 457-.462, Jul. 1969.

[35] M.J. Omidi, S. Pasupathy and P.G. Gulak, "Joint data and Kalman estimation for Rayleigh fading channels," *J. of Wireless Pers. Com.*, Kluwers Publishers, 1998.

[36] J. G. Proakis, *Digital Communication*, New York: McGraw-Hill, 1989.

[37] R. Raheli, A. Polydoros and C.-K. Tzou, "Per-survivor processing: a general approach to MLSE in uncertain environment," *IEEE Trans. on Communications*, Vol. 43, No. 2/3/4, pp. 354–364, 1995.

[38] A.R. Rao, Y.F. Huang, "Recent developments in optimal bounding ellipsoidal parameter estimation," *Math. Comp. Simul.*, Vol. 32, pp. 515–526, 1990.

[39] I. Rapoport, Y. Oshman, "A CramerRao-Type Estimation Lower Bound for Systems With Measurement Faults," *IEEE Trans. Automat. Contr.*, Vol. 50, No. 9, pp. 1234–1245, Sept. 2005.

[40] G.S. Rodgers, *Matrix derivatives*, Marcel Dekker, New York, Basel, 1980

[41] S. P. Sing and R.W. Liu, "Existence of state equation representation of linear large-scale dynamical systems," *IEEE Trans. Circuit Theory.*, No. 20, pp. 239–246, 1973.

[42] E. Sontag *Mathematical Control Theory,* 2nd ed., series of Applied Mathematics, Springer Verlag, 1998.

[43] S. Tafazoli, X. Sun, " Hybrid System State Tracking and Fault Detection Using Particle Filters," *IEEE Trans. on Control Systems Technology,* Vol. 14, No. 6, pp. 1078–1087, Nov. 2006.

[44] C.-K. Tzou, R. Raheli and A. Polydoros, "Applications of per-survivor processing to mobile digital communications," *Proc. IEEE Globecom, Comm. Theory Mini-Conf.*, pp. 77–81, Nov. 1993.

[45] D.D. Sworder, J.E. Boyd, R.J. Elliot "Modal Estimation in Hybrid Systems," *J. of Mathematical Analysis and Applications* Vol. 245, pp. 225–247, 2000.

[46] R. Vidal, A. Chiuso, S. Soatto and S. Sastry, "Observability of linear hybrid systems," in *Hybrid Systems: Computation and Control,* Vol. 2623 of LNCS, Springer Verlag, pp. 526–539, 2003.

[47] W. Wang, D.H. Zhou, Z. Li, "Robust state estimation and fault diagnosis for uncertain hybrid systems," *Nonlinear Analysis,* Vol. 65, pp. 2193–2215, 2006.

[48] Y. Zhang and X.R. Li, "Detection and diagnosis of sensor and actuator failures using IMM estimator," *IEEE Trans. Aerospace Electron. Syst.*, Vol. 34, No. 4, pp. 1293-1313, Oct. 1998.

[49] Q. Zhang, "Hybrid Filtering for Linear Systems with Non-Gaussian Disturbances," *IEEE Trans. Automat. Contr.*, Vol. 45, No. 1, pp. 50–61, 2000.

[50] F. Zhao, X. Koutsoukos, H. Haussecker, J. Reich, P. Cheung, "Monitoring and fault diagnosis of hybrid systems," *IEEE Trans. on Systems, Man, and Cybernetics, Part B,* Vol. 35, No. 6, pp.. 1225–1240, Dec. 2005.