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

A. Germani, C. Manes, P. Palumbo

A POLYNOMIAL APPROACH FOR SIMULTANEOUS CHANNEL ESTIMATION AND DATA DETECTION

R. 599 Ottobre 2003

Alfredo Germani – Dipartimento di Ingegneria Elettrica, Università degli Studi dell'Aquila, 67040 Monteluco (L'Aquila), Italy and Istituto di Analisi dei Sistemi ed Informatica del CNR, Viale Manzoni 30, 00185 Roma, Italy. Email: germani@ing.univaq.it.

Costanzo Manes – Dipartimento di Ingegneria Elettrica, Università degli Studi dell'Aquila, 67040 Monteluco (L'Aquila), Italy and Istituto di Analisi dei Sistemi ed Informatica del CNR, Viale Manzoni 30, 00185 Roma, Italy. Email: manes@ing.univaq.it.

Pasquale Palumbo – Istituto di Analisi dei Sistemi ed Informatica del CNR, Viale Manzoni 30, 00185 Roma, Italy. Email: palumbo@iasi.rm.cnr.it.

ISSN: 1128–3378

Collana dei Rapporti
dell'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

This work investigates the simultaneous estimate of both the impulse response samples and the transmitted signal in a digital communication channel. The minimum variance approach applied to a suitably defined extended state-space representation, which includes both the channel samples and the signal, is here adopted. Such a problem cannot be solved by the use of linear algorithms. It is shown how the proposed polynomial approach overcomes the previously mentioned drawback. Numerical simulations referred to a telecommunication framework support theoretical results.

Key words: Telecommunication systems, Polynomial filtering, Kalman Filtering, Markov Processes.

1. Introduction

The received signal of a time varying fast-fading digital communication system is usually modeled by the discrete convolution of the trasmitted data sequence with the channel finite impulse response samples, plus additive noise [8]. The problem here considered, very important in communication engineering, is the reconstruction of transmitted signals from noisy received data taking also into account the stochastic time evolution of the response samples characterizing the channel. The standard model for the channel of a fast-fading digital link is given by a Gauss-Markov process, described by a stochastic discrete-time linear system affected by an additive white Gaussian noise. Of course, in order to correctly reconstruct the transmitted signal, it is required the knowledge of the channel samples. In a great deal of literature, the channel estimate is performed by using the Least Mean Squares (LMS)[10], the Recursive Mean Squares (RMS) [5] or the Kalman Filter (KF) [7]. Each of these algorithms works in connection with a Viterbi Decoder (VD), implementing the signal deconvolution. More recently, a very interesting approach has been proposed, based on the Particle Filtering theory [3].

This paper presents a new approach, based on the use of polynomial filters [4]. The key idea is to define an extended state space including the transmitted data and the channel samples with some of their Kronecker products. In this larger state space a linear stochastic model driven by multiplicative noise is so defined, which is able to generate the received data together with its Kronecker powers up to degree ν. Finally to such an extended linear non Gaussian system the standard Kalman Filter is applied, which achieves the ν-degree optimal polynomial estimate according to the minimum variance criterion. The proposed methodology comes from a recent work in the filtering of switching systems (see [6] and references therein).

2. The digital channel model

For the sake of simplicity, here it will be considered the case of real scalar observations. The extension of the more useful complex framework is straightforward. Suppose that the received data sequence $\{y(k)\}$ is modeled by using the following discrete-time model [8]:

$$y(k) = \sum_{j=0}^{n-1} c(k,j)b(k-j) + N_g(k), \qquad k \in \mathbb{Z} \tag{2.1}$$

where $\{b(k)\}$ is the transmitted data sequence of independent identically distributed equiprobable symbols, taking values on a given real alphabet $\mathcal{A}$ of size S and $\{c(k,j),\ j = 0,\dots,n-1\}$ is the time-varying impulse response of a channel of length n. $\{N_g(k)\}$ is a zero-mean Gaussian white noise sequence, with finite and available covariance matrix Ψ_g. Let the "state transition sequence" of the channel $\{\mathbf{b}(k)\}$ be defined as:

$$\mathbf{b}^T(k) = \begin{bmatrix} b(k) & \cdots & b(k-n+1) \end{bmatrix}. \tag{2.2}$$

$\mathbf{b}(k)$ takes values in $\mathcal{A}^n$, the set of the $M = S^n$ distinct permutations of the elements of the alphabet $\mathcal{A}$. The symbols ζ_i, $i = 1,\dots,M$, denote the n-ples in $\mathcal{A}^n$. The sequence $\{\mathbf{b}(k)\}$ can be modeled as a first order Markov chain with Π its probability transition matrix, given by:

$$\big[\Pi\big]_{ij} = P\big\{\mathbf{b}(k+1) = \zeta_i | \mathbf{b}(k) = \zeta_j\big\}, \quad \zeta_i,\zeta_j \in \mathcal{A}^n, \tag{2.3}$$

for $i, j = 1, \ldots, M$, and $p = \{p_1, \ldots, p_M\}$ the initial probability distribution:

$$P\{\mathbf{b}(0) = \zeta_i\} = p_i, \qquad \zeta_i \in \mathcal{A}^n, \quad i = 1, \ldots, M. \tag{2.4}$$

Denoting $\mathbf{c}(k) = \big(c(k,0), \cdots, c(k, n-1)\big)^T$ the n-dimensional impulse-response samples vector of the channel, the measurement equation in (2.1) can be put in the more compact form:

$$y(k) = \mathbf{b}^T(k)\mathbf{c}(k) + N_g(k). \tag{2.5}$$

According to [8], $\{\mathbf{c}(k)\}$ can be modeled as a Gauss-Markov process, described by the following recursive equation:

$$\mathbf{c}(k+1) = \Lambda\mathbf{c}(k) + N_f(k), \qquad \mathbf{c}(0) = \mathbf{c}_0, \tag{2.6}$$

with $\{N_f(k)\}$ a zero-mean Gaussian white sequence, independent of $\{N_g(k)\}$, with finite and available covariance matrix Ψ_f and $\mathbf{c}_0$ a Gaussian random variable, independent of the noise sequences, with finite and available first and second order moments, $\bar{\mathbf{c}}_0$ and Ψ_0, respectively. By taking into account equations (2.6) and (2.5), it comes out the following stochastic linear system in the discrete time framework, endowed with a measurement equation, bilinear with respect to the pair $(\mathbf{c}, \mathbf{b})$:

$$\begin{aligned} \mathbf{c}(k+1) &= \Lambda\mathbf{c}(k) + N_f(k), \qquad \mathbf{c}(0) = \mathbf{c}_0, \\ y(k) &= \mathbf{b}^T(k)\mathbf{c}(k) + N_g(k). \end{aligned} \tag{2.7}$$

In order to give a simple state-space stochastic generation model of $\mathbf{b}(k)$, let us introduce a bijective map between the canonical basis of $I\!R^M$ and $\mathcal{A}^n$. Let $\mathcal{B}_M = \{e_i, \ i = 1, \ldots, M\}$ denote the canonical basis of $I\!R^M$, and consider the bijection $\mathcal{P} : \mathcal{B}_M \mapsto \mathcal{A}^n$ such that

$$\zeta_i = \mathcal{P}(e_i), \qquad i = 1, \ldots, M. \tag{2.8}$$

Defining the row vector

$$B^T = \big[\zeta_1^T \ \cdots \ \zeta_M^T\big], \tag{2.9}$$

an expression that transforms e_i in ζ_i is the following one:

$$\zeta_i^T = B^T\big(e_i \otimes I_n\big), \tag{2.10}$$

where the symbol $\otimes$ denotes the Kronecker product (see the Appendix for more details). Throughout the paper the symbol I_a denotes the identity matrix in $I\!R^{a\times a}$. The sequence $\mathbf{b}(k)$ can be generated by a Markov process $\{\theta(k)\}$ assuming values in $\mathcal{B}_M$, characterized by the transition matrix Π and initial probability distribution p defined in (2.3) and (2.4), as follows:

$$\mathbf{b}(k) = \mathcal{P}\big(\theta(k)\big), \qquad \mathbf{b}^T(k) = B^T\big(\theta(k) \otimes I_n\big). \tag{2.11}$$

Moreover $\theta(k)$ admits the following state-space representation:

$$\theta(k+1) = V(k)\theta(k), \qquad \theta(0) = \theta_0, \tag{2.12}$$

with $\theta_0 = \mathcal{P}^{-1}\big(\mathbf{b}(0)\big)$ and $\{V(k)\}$ a sequence of $M \times M$ random matrices where, because of the Markov property of $\{\mathbf{b}(k)\}$, each j-th column $\{V_j(k)\}$ is a sequence of independent random vectors assuming values in $\mathcal{B}_M$, with probabilities given by (2.3):

$$P\{V_j(k) = e_i\} = \big[\Pi\big]_{ij}, \qquad i, j = 1, \ldots, M. \tag{2.13}$$

By consequence, $V(k)$ is independent of $\theta(k)$ and $I\!E\big[V(k)\big] = \Pi$. It is useful in the sequel to refer to the parameter state equation (2.12) by using a zero-mean random matrix, that is, by naming $\mathcal{V}(k) = V(k) - \Pi$, equation (2.12) becomes:

$$\theta(k+1) = \Pi\theta(k) + \mathcal{V}(k)\theta(k), \qquad I\!E\big[\mathcal{V}(k)\big] = 0, \tag{2.14}$$

see [6] for more details.

Remark 2.1. According to the above considerations, the stochastic system (2.7) can be put in the following bilinear state space representation:

$$\begin{aligned} \mathbf{c}(k+1) &= \Lambda\mathbf{c}(k) + N_f(k+1), \\ \theta(k+1) &= \Pi\theta(k) + \mathcal{V}(k)\theta(k), \\ y(k) &= B^T\big(\theta(k) \otimes \mathbf{c}(k)\big) + N_g(k), \end{aligned} \tag{2.15}$$

in that, from (2.11), the output equation becomes:

$$\begin{aligned} B^T\big(\theta(k) \otimes I_n\big)\mathbf{c}(k) &= B^T\big(\theta(k) \otimes I_n\big)\big(1 \otimes \mathbf{c}(k)\big) \\ &= B^T\Big(\big(\theta(k) \cdot 1\big) \otimes \big(I_n \cdot \mathbf{c}(k)\big)\Big) = B^T\big(\theta(k) \otimes \mathbf{c}(k)\big). \end{aligned} \tag{2.16}$$

3. The polynomial filter

According to Remark 2.1, the dynamics of digital channel model is given by a linear system, forced by both additive and multiplicative white noises, whereas the output is given by a quadratic transformation of the state plus additive white Gaussian noise. In this Section it is shown how to represent such a nonlinear system with a suitable linear model, forced by multiplicative noise.

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

$$\hat{\mathbf{c}}(k) = I\!E\big[\mathbf{c}(k)|y(0), \ldots, y(k)\big] = \mathbf{P}\big[\mathbf{c}(k)|\mathcal{L}(Y_k)\big], \tag{3.1}$$

with $Y_k = \big[y(0) \;\; \cdots \;\; y(k)\big]^T$. In the linear Gaussian case the optimal filter is given by the Kalman Filter, which is a linear transformation of the measurements. Unfortunately, in the non Gaussian case, there is not a simple characterization of the conditional expectation, so that 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 onto the subspace $L(Y_k)$ of all the linear transformations of the output. For linear systems the optimal affine estimate is still achieved by the Kalman filter. Suboptimal estimates comprised between the optimal linear and the conditional expectation can be considered by projecting onto subspaces greater than $L(Y_k)$, like subspaces of polynomial transformations of the measurements [4]. More in details, the subspace here considered is the following Hilbert space of the polynomial transformations of the measurements, of a fixed degree ν:

$$\hat{\mathbf{c}}_\nu(k) = \mathbf{P}\big[\mathbf{c}(k)|L(Y_k^\nu)\big], \tag{3.2}$$

with $L(Y_k^\nu) = \text{span}\,\{Y^\nu(0),\ \cdots,\ Y^\nu(k)\}$, and

$$Y_k^\nu = \begin{bmatrix} Y^\nu(0) \\ \vdots \\ Y^\nu(k) \end{bmatrix}, \quad Y^\nu(h) = \begin{bmatrix} Y_1(h) \\ \vdots \\ Y_\nu(h) \end{bmatrix}, \quad \begin{array}{l} Y_i(h) = y^i(h), \\ h = 0, \dots, k, \end{array} \tag{3.3}$$

and the extra-assumption that $\mathbb{E}\big[|y^i(h)|^2\big] < \infty$, for $i = 1, \dots, \nu$. It will be shown that such a polynomial approach, together with the state space representation of the Markov chain, is the way to ensure the simultaneous estimate of both the channel samples and the transmitted data.

Before to give the main Theorem, we need to define the following vector:

$$X^\nu(k) = \begin{bmatrix} X_0(k) \\ \vdots \\ X_\nu(k) \end{bmatrix}, \qquad \begin{array}{l} X_j(k) = \theta(k) \otimes \mathbf{c}^{[j]}(k), \\ j = 0, \dots, \nu. \end{array} \tag{3.4}$$

Theorem 3.1. *The optimal ν-degree polynomial estimates of the components $\mathbf{c}(k)$ and $\theta(k)$ of the state of the bilinear system* (2.15) *are given by:*

$$\begin{aligned} \hat{\mathbf{c}}_\nu(k) &= \mathcal{M}_n \widehat{X}^\nu(k) = \mathcal{M}_n \mathbf{P}\big[X^\nu(k)|L(Y_k^\nu)\big], \\ \hat{\theta}_\nu(k) &= \mathcal{T}_n \widehat{X}^\nu(k) = \mathcal{T}_n \mathbf{P}\big[X^\nu(k)|L(Y_k^\nu)\big], \end{aligned} \tag{3.5}$$

with

$$\begin{aligned} \mathcal{M}_n &= \big[O_{n\times M} \quad \mathcal{M} \quad O_{n\times M(n^2+\cdots+n^\nu)}\big], \\ \mathcal{T}_n &= \big[I_M \quad O_{M\times M(n+\cdots+n^\nu)}\big], \quad \mathcal{M} = \big[I_n \cdots I_n\big]. \end{aligned} \tag{3.6}$$

Proof. According to the range of the parameter $\theta(k)$, it can be shown that the Kronecker powers and products in X_j defined by (3.4), are given by the following relations (see [6] for more details):

$$X_j^{[h]} = \Theta_n^{h,j} X_{jh}, \quad X_i \otimes X_j = \Xi_{i,j} X_{i+j}, \quad \forall i,j,h \in \mathbb{N}, \tag{3.7}$$

where $\Theta_n^{h,j}$ and $\Xi_{i,j}$ are defined by:

$$\begin{aligned} \Theta_n^{h+1,j} &= \big(\Theta_n^{h,j} \otimes I_{Mn^j}\big)\big(I_M \otimes C^T_{Mn^j,n^{jh}}\big)\big(E_2 \otimes I_{n^{j(h+1)}}\big), \\ \Theta_n^{0,j} &= [1 \cdots 1]. \end{aligned} \tag{3.8}$$

and

$$\Xi_{i,j} = \big(I_M \otimes C^T_{Mn^j,n^i}\big)\big(E_2 \otimes I_{n^{i+j}}\big) \tag{3.9}$$

with $E_2 = \big[e_1^{[2]} \ \cdots \ e_M^{[2]}\big]$ and $C_{a,b}$ suitably dimensioned commutation matrices for the Kronecker product, defined in Appendix. That means, taking into account that all the output Kronecker powers up to the ν-th degree of the measurement equation (2.15) depend of the powers $X_1^{[i]}(k) = \Theta_n^{i,1} X_i(k)$, the extended output equation is a linear transformation of the extended state $X^\nu(k)$, driven by a multiplicative noise. Moreover, both the state $\mathbf{c}(k)$ and the Markov parameter $\theta(k)$ are linearly dependent of the extended state $X^\nu(k)$ so that the polynomial minimum variance state estimate in (3.2) is:

$$\begin{aligned} \hat{\mathbf{c}}_\nu(k) &= \mathbf{P}\big[\mathbf{c}(k)|L(Y_k^\nu)\big] = \mathbf{P}\big[\mathcal{M}_n X^\nu(k)|L(Y_k^\nu)\big] \\ &= \mathcal{M}_n \mathbf{P}\big[X^\nu(k)|L(Y_k^\nu)\big] = \mathcal{M}_n \widehat{X}^\nu(k), \end{aligned} \tag{3.10}$$

Analogously the polynomial minimum variance state estimate of $\theta(k)$ in equation (3.5) comes. ∎

Remark 3.2. Note that the minimum variance estimate of $\theta(k)$, generally, does not belong to $\mathcal{B}_M$. That means that a possible strategy for the estimation of $\mathbf{b}(k)$ could be made in two steps: the first one consists in choosing the closest element in $\mathcal{B}_M$ to the estimate, namely $\tilde{\theta}(k)$:

$$\tilde{\theta}(k) = e_{\hat{i}(k)}, \qquad \hat{i}(k) = \mathrm{argmax}_{j=1,\dots,M}\hat{\theta}_j(k); \tag{3.11}$$

the second consists in calculating the corresponding $\hat{\mathbf{b}}(k)$ by using equation (2.11), that is:

$$\hat{\mathbf{b}}(k) = B^T\big(e_{\hat{i}(k)} \otimes I_n\big). \tag{3.12}$$

It is worthwhile to note that when $\hat{\theta}(k)$ is the minimum variance estimate (i.e. the conditional expectation), then $\tilde{\theta}(k)$ coincides with the Maximum Likelihood Estimate.

In order to obtain the projection in equations (3.5), it has to be shown that the sequences $\{X^\nu(k)\}$ and $\{Y^\nu(k)\}$ obey difference equations of the type:

$$\begin{aligned} X^\nu(k+1) &= \mathbf{A}^\nu(k)X^\nu(k) + \mathcal{F}(k), \\ Y^\nu(k) &= \mathbf{C}^\nu(k)X^\nu(k) + \mathcal{G}(k), \end{aligned} \tag{3.13}$$

with $\mathbf{A}^\nu(k)$ and $\mathbf{C}^\nu(k)$ suitably defined matrices and

$$\begin{aligned} \mathcal{F}(k) &= \widetilde{\mathcal{F}}\big(k, X^\nu(k), N_f(k)\big), \\ \mathcal{G}(k) &= \widetilde{\mathcal{G}}\big(k, X^\nu(k), N_g(k)\big), \end{aligned} \tag{3.14}$$

with $\widetilde{\mathcal{F}}$ and $\widetilde{\mathcal{G}}$ suitably defined functions where $X^\nu(k)$ 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 sequences. Such a problem is similar to the general one solved in [6] where the Markov parameter affected also the state equation. The following Lemmas show the way to achieve the matrices and the noise sequences in (3.13).

Lemma 3.3. *The iterative equation of the component $X_j(k)$ of the extended state $X^\nu(k)$ as defined in* (3.4) *is given by:*

$$X_j(k+1) = \sum_{i=0}^{j} \mathbf{A}_{ji}X_i(k) + \mathcal{F}_j(k), \tag{3.15}$$

where

$$\mathbf{A}_{ji} = \Pi \otimes J_i^j, \qquad \mathcal{F}_j(k) = \sum_{i=0}^{j} S_i^j(k)X_i(k), \tag{3.16}$$

and

$$S_i^j(k) = \big(\Pi \otimes L_i^j(k) + \mathcal{V}(k) \otimes J_i^j + \mathcal{V}(k) \otimes L_i^j(k)\big), \tag{3.17}$$

with:

$$\begin{aligned} J_i^j &= M_i^j(n)\big(\Lambda^{[i]} \otimes \xi_f^{j-i}\big), \\ L_i^j(k) &= M_i^j(n)\big(\Lambda^{[i]} \otimes (N_f^{[j-i]}(k) - \xi_f^{j-i})\big). \end{aligned} \tag{3.18}$$

$M_i^j(n)$ are the matrix coefficients coming from the Kronecker binomial formula (see Appendix), while $\xi_f^i = \mathbb{E}\big[N_f^{[i]}(k)\big]$ are the state noise moments, whose computation comes from the available first and second order moments, according to the Gaussian hypotheses.

Proof. According to the Kronecker binomial power formula, it comes:

$$\begin{aligned}\mathbf{c}^{[j]}(k+1) &= \big(\Lambda\mathbf{c}(k)+N_f(k)\big)^{[j]}\\ &= \sum_{i=0}^{j} M_i^j(n)\big((\Lambda^{[i]}\mathbf{c}^{[i]}(k))\otimes N_f^{[j-i]}(k)\big)\\ &= \sum_{i=0}^{j} J_i^j\mathbf{c}^{[i]}(k)+\sum_{i=0}^{j} L_i^j(k)\mathbf{c}^{[i]}(k),\end{aligned}\tag{3.19}$$

with J_i^j and $L_i^j(k)$ given by (3.18). Then, taking into account the extended state components:

$$\begin{aligned}X_j(k+1) &= \big((\Pi+\mathcal{V}(k))\theta(k)\big)\\ &\qquad\otimes\left(\sum_{i=0}^{j} J_i^j\mathbf{c}^{[i]}(k)+\sum_{i=0}^{j} L_i^j(k)\mathbf{c}^{[i]}(k)\right)\\ &= \sum_{i=0}^{j}\mathbf{A}_{ji}X_i(k)+\sum_{i=0}^{j} S_i^j(k)X_i(k),\end{aligned}\tag{3.20}$$

with $\mathbf{A}_{ji}$ and $S_i^j(k)$ as in (3.16) and (3.17). ∎

Analogously, the extended output equation in (3.13) is described by the following Lemma:

Lemma 3.4. *The iterative equation of the component $Y_j(k)$ of the extended output $Y^\nu(k)$ as defined in* (3.3) *is given by:*

$$Y_j(k)=\sum_{i=0}^{j}\mathbf{C}_{ji}X_i(k)+\mathcal{G}_j(k),\quad \mathcal{G}_j(k)=\sum_{i=0}^{j} T_i^j(k)X_i(k)\tag{3.21}$$

where

$$\mathbf{C}_{ji}=H_i^j\big(\Theta_n^{i,1}\otimes\xi_g^{j-i}\big),\qquad H_i^j=\binom{j}{i}(B^T)^{[i]},\tag{3.22}$$

and

$$T_i^j(k)=H_i^j\big(\Theta_n^{i,1}\otimes\big(N_g^{j-i}(k)-\xi_g^{j-i}\big)\big).\tag{3.23}$$

$\xi_g^i=I\!E\big[N_g^i(k)\big]$ *are the output noise moments, whose computation comes from the available first and second order moments, according to the Gaussian hypotheses.*

Proof. According to (2.15), applying the Kronecker binomial formula:

$$\begin{aligned}Y_j(k) &= \big(B^TX_1(k)+N_g(k)\big)^{[j]}\\ &= \sum_{i=0}^{j}\binom{j}{i}\Big(\big((B^T)^{[i]}X_1^{[i]}(k)\big)\otimes N_g^{j-i}(k)\Big)\\ &= \sum_{i=0}^{j}\binom{j}{i}(B^T)^{[i]}\Theta_n^{i,1}\big(X_i(k)\otimes N_g^{j-i}(k)\big)\\ &= \sum_{i=0}^{j}\mathbf{C}_{ji}X_i(k)+\sum_{i=0}^{j} T_i^j(k)X_i(k)\end{aligned}\tag{3.24}$$

with $\mathbf{C}_{ji}$ and $T_i^j(k)$ given by (3.22) and (3.23). ∎

The following Lemma is useful in order to find the statistics of the extended noise sequences.

Lemma 3.5. *Let $\{\xi_r(k)\}$, $r \in Q \subseteq I\!N$, be a class of random sequences, defined by:*

$$\xi_r(k) = \sum_{i=0}^{r} \Xi_i(k)\chi_i(k), \tag{3.25}$$

with $\{\Xi_i(k)\}$, $i \in Q$, sequences of zero-mean uncorrelated random matrices and $\chi_i(k)$ suitably dimensioned random variables such that $\chi_i(k)$ is independent of $\Xi_j(h)$, $\forall i, j \in Q$, $\forall h \geq k$. Then:

$$I\!E\big[\xi_r(k)\xi_s^T(h)\big] = 0, \qquad \forall r, s \in Q, \quad \forall h \neq k, \tag{3.26}$$

with

$$\begin{aligned} I\!E\big[\xi_r(k)\xi_s^T(k)\big] = \mathrm{st}^{-1}\Bigg(& \sum_{i=0}^{r}\sum_{j=0}^{s} I\!E\big[\Xi_j(k) \otimes \Xi_i(k)\big] \\ & \cdot I\!E\big[\chi_j(k) \otimes \chi_i(k)\big]\Bigg), \end{aligned} \tag{3.27}$$

where st^{-1} is the inverse of the stack of a matrix (see the Appendix for more details). Moreover, let $\{\zeta_r(k)\}$, $r \in I$, be a class of random sequences, defined by:

$$\zeta_r(k) = \sum_{i=0}^{r} Z_i(k)\chi_i(k), \tag{3.28}$$

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

$$I\!E\big[\xi_r(k)\zeta_s^T(h)\big] = 0, \qquad \forall r, s \in I, \quad \forall k, h \in I\!N. \tag{3.29}$$

Proof. Let $r, s \in I$ and $h > k$. Then $\xi_r(h)$ and $\xi_s(k)$ are uncorrelated random variables, in that:

$$\begin{aligned} I\!E\big[\xi_r(k)\xi_s^T(h)\big] &= \sum_{i=0}^{r}\sum_{j=0}^{s} I\!E\big[\Xi_i(k)\chi_i(k)\chi_j^T(h)\Xi_j^T(h)\big] \\ &= \sum_{i=0}^{r}\sum_{j=0}^{s} I\!E\big[\Xi_i(k)\chi_i(k)\chi_j^T(h)\big] I\!E\big[\Xi_j^T(h)\big] = 0. \end{aligned} \tag{3.30}$$

If $h = k$, equation (3.27) comes by applying the stack properties concerning the Kronecker product. Analogously to (3.30), equation (3.29) is easily verified for $h \neq k$. Moreover, if $h = k$:

$$\begin{aligned} I\!E\big[\xi_r(k)\zeta_s^T(k)\big] &= \sum_{i=0}^{r}\sum_{j=0}^{s} I\!E\big[\Xi_i(k)\chi_i(k)\chi_j^T(k)Z_j^T(k)\big] \\ &= \sum_{i=0}^{r}\sum_{j=0}^{s} I\!E\big[\Xi_i(k)\chi_i(k)\chi_j^T(k)\big] I\!E\big[Z_j^T(k)\big] = 0. \end{aligned} \tag{3.31}$$

∎

Lemma 3.6. *The extended noise sequences $\{\mathcal{F}(k)\}$ and $\{\mathcal{G}(k)\}$ are sequences of zero-mean, uncorrelated random variables and, moreover:*

$$\begin{aligned}\Psi^{\mathcal{F}}_{r,s} &= I\!E\big[\mathcal{F}_r(k)\mathcal{F}_s(k)^T\big]\\ &= \mathrm{st}^{-1}\left(\sum_{i=0}^{r}\sum_{j=0}^{s}\Phi^{S,s,r}_{j,i}\Xi_{j,i}I\!E\big[X_{i+j}(k)\big]\right)\end{aligned} \tag{3.32}$$

$$\begin{aligned}\Psi^{\mathcal{G}}_{r,s} &= I\!E\big[\mathcal{G}_j(k)\mathcal{G}_i(k)^T\big]\\ &= \mathrm{st}^{-1}\left(\sum_{i=0}^{r}\sum_{j=0}^{s}\Phi^{T,s,r}_{j,i}\Xi_{j,i}I\!E\big[X_{i+j}(k)\big]\right)\end{aligned} \tag{3.33}$$

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

$$I\!E\big[\mathcal{F}_r(k)\mathcal{G}_s(h)^T\big] = 0, \qquad \forall k,h\in Z\!\!\!Z. \tag{3.34}$$

Proof. According to (3.16), (3.17), (3.18) and (3.21), (3.22), (3.23), $\{\mathcal{F}(k)\}$ 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 $X_i(k)$. By consequence, according to Lemma 3.5, 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 $\forall h,k\in Z\!\!\!Z$, 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 3.5, the extended state and output noises are uncorrelated random sequences. As far as their covariances matrices are concerned, from (3.27):

$$\begin{aligned}I\!E\big[\mathcal{F}_r(k)\mathcal{F}^T_s(k)\big] = \mathrm{st}^{-1}\Bigg(&\sum_{i=0}^{r}\sum_{j=0}^{s}I\!E\big[S^s_j(k)\otimes S^r_i(k)\big]\\ &\cdot I\!E\big[X_j(k)\otimes X_i(k)\big]\Bigg),\end{aligned} \tag{3.35}$$

$$\begin{aligned}I\!E\big[\mathcal{G}_r(k)\mathcal{G}^T_s(k)\big] = \mathrm{st}^{-1}\Bigg(&\sum_{i=0}^{r}\sum_{j=0}^{s}I\!E\big[T^s_j(k)\otimes T^r_i(k)\big]\\ &\cdot I\!E\big[X_j(k)\otimes X_i(k)\big]\Bigg),\end{aligned} \tag{3.36}$$

from which equations (3.32) and (3.33) easily come by applying (3.7).

∎

Proposition 3.7. *According to Theorem* 3.1, *the ν-th degree polynomial filtering algorithm is the following:*

$$\begin{aligned}&\hat{\mathbf{c}}_\nu(k) = \mathcal{M}_n\widehat{X}^\nu(k), \qquad \hat{\theta}_\nu(k) = \mathcal{T}_n\widehat{X}^\nu(k),\\ &\widehat{X}^\nu(k+1) = \mathbf{A}^\nu\widehat{X}^\nu(k) + \mathcal{K}(k+1)\Big(Y^\nu(k+1)\\ &\qquad\qquad - \mathbf{C}^\nu\mathbf{A}^\nu\widehat{X}^\nu(k)\Big).\end{aligned} \tag{3.37}$$

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

$$\begin{aligned} \mathcal{P}_P(k+1) &= \mathbf{A}^\nu \mathcal{P}(k) \mathbf{A}^{\nu T} + \Psi^{\mathcal{F}} \\ \mathcal{P}(k) &= \mathcal{P}_P(k) - \mathcal{K}(k) \mathbf{C}^\nu \mathcal{P}_P(k) \\ \mathcal{K}(k) &= \mathcal{P}_P(k) \mathbf{C}^{\nu T} \Big(\mathbf{C}^\nu \mathcal{P}_P(k) \mathbf{C}^{\nu T} + \Psi^{\mathcal{G}} \Big)^\dagger \end{aligned} \tag{3.38}$$

where in (3.38) *the Moore-Penrose pseudoinverse has been used. According to Remark* 3.2*, the transmitted data estimate* $\hat{\mathbf{b}}(k)$ *is given by* (3.11)*.*

Proof. The filter equations are those of the classical Kalman filter [1] applied to the system (3.13), that has a multiplicative noise structure (see equations (3.15), (3.16) and (3.21), describing the components of $\mathcal{F}$ and of $\mathcal{G}$). The use of the Kalman algorithm on an extended system to achieve optimal polynomial filtering of system with multiplicative noise has been already demonstrated in [4]. Note that in order to obtain the noise covariance matrices $\Psi^{\mathcal{F}}$, $\Psi^{\mathcal{G}}$ of (3.38), the following 2ν order deterministic system has to be computed:

$$I\!E\big[X^\nu(k+1)\big] = \mathbf{A}^{2\nu} I\!E\big[X^\nu(k)\big], \tag{3.39}$$

which gives the evolution of the mean value of the extended state.

∎

Remark 3.8. Because $\mathbf{c}(k) - \hat{\mathbf{c}}(k) = \mathcal{M}_n\big(X^\nu(k) - \widehat{X}^\nu(k)\big)$, then:

$$I\!E\Big[\big(\mathbf{c}(k) - \hat{\mathbf{c}}_\nu(k)\big)\big(\mathbf{c}(k) - \hat{\mathbf{c}}_\nu(k)\big)^T\Big] = \mathcal{M}_n \mathcal{P}(k) \mathcal{M}_n^T. \tag{3.40}$$

4. Linear vs polynomial filtering

Keeping $\nu = 1$, according to Theorem 3.1, the best linear estimate is achieved. Taking into account the system matrices defined in Lemmas 3.3 and 3.4, it comes:

$$\mathbf{A}^1 = \begin{bmatrix} \Pi & O \\ O & \Pi \otimes \Lambda \end{bmatrix}, \qquad \mathbf{C}^1 = \begin{bmatrix} O & B^T \end{bmatrix}. \tag{4.1}$$

As a consequence, the pair $(\mathbf{A}^1, \mathbf{C}^1)$ is not observable, in that:

$$Q_1 = \begin{bmatrix} \mathbf{C}^1 \\ \mathbf{C}^1\mathbf{A}^1 \end{bmatrix} = \begin{bmatrix} O & B^T \\ O & B^T(\Pi \otimes \Lambda) \end{bmatrix}. \tag{4.2}$$

Actually, the pair $(\mathbf{A}^1, \mathbf{C}^1)$ is even not detectable, as it can be seen by applying the PBH test for $\lambda = 1$, which is a non-asymptotically-stable eigenvalue of matrix Π and, therefore, of matrix $\mathbf{A}^1$:

$$\begin{aligned} &\operatorname{rank} \begin{bmatrix} \mathbf{A}^1 - I_{M+nM} \\ \mathbf{C}^1 \end{bmatrix} \\ &= \operatorname{rank} \begin{bmatrix} \Pi - I_M & O \\ O & \Pi \otimes \Lambda - I_{nM} \\ O & B^T \end{bmatrix} < M + nM. \end{aligned} \tag{4.3}$$

It is possible to recognize that, in this case, the optimal linear filter is not able to improve the *a priori* estimate given by the mean value of the transmitted signal.

On the other hand, the polynomial approach overcomes this structural undetectability just starting with the second order filter, without any further assumption. In fact, according to (3.18) and (3.22):

$$\begin{aligned} \mathbf{A}^2 &= \begin{bmatrix} \Pi & O & O \\ O & \Pi \otimes \Lambda & O \\ \Pi \otimes \xi_f^2 & O & \Pi \otimes \Lambda^{[2]} \end{bmatrix}, \\ \mathbf{C}^2 &= \begin{bmatrix} O & B^T & O \\ \Theta_n^{0,1} \otimes \xi_g^2 & O & \left(B^T\right)^{[2]} \Theta_n^{2,1} \end{bmatrix}, \end{aligned} \tag{4.4}$$

so that the pair $(\mathbf{A}^2, \mathbf{C}^2)$ may well be observable.

5. Numerical simulations and conclusions

In this section are reported some of the numerical simulations produced in order to test the proposed algorithm. For the sake of brevity, here is reported only the simple case of a channel with impulse repsonse length $n = 2$, while the transmitted data sequence takes values on the real alphabet of two equiprobable symbols, $\{0, 1\}$, that means $M = 2^n = 4$. By naming:

$$\mathbf{b}_1 = \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \quad \mathbf{b}_2 = \begin{bmatrix} 0 \\ 1 \end{bmatrix}, \quad \mathbf{b}_3 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \quad \mathbf{b}_4 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \tag{5.1}$$

it comes that the transition probability matrix Π is given by:

$$\Pi = \begin{bmatrix} 0.5 & 0.5 & 0 & 0 \\ 0 & 0 & 0.5 & 0.5 \\ 0.5 & 0.5 & 0 & 0 \\ 0 & 0 & 0.5 & 0.5 \end{bmatrix}. \tag{5.2}$$

The channel dynamics is:

$$\Lambda = \begin{bmatrix} 0.3679 & 0 \\ 0 & 0.3679 \end{bmatrix}, \tag{5.3}$$

with the zero-mean Gaussian noises N_f and N_g endowed with the following covariances matrices:

$$\Psi_f^2 = \begin{bmatrix} 0.2253 & 0 \\ 0 & 0.1301 \end{bmatrix}, \qquad \Psi_g^2 = 0.01. \tag{5.4}$$

In order to simultaneously estimate both the channel samples and the transmitted signal in real time a second order filter has been implemented by using just real measurements instead of the more realistic bidimensional (complex) data. Figures 5.1 and 5.2 show the evolution of the first 100 simulation samples for the first and second taps of the channel together with their quadratic estimates. As far as what concern the transmitted signal, it comes that the percentage of success in this worst case (real data, only quadratic filtering, no use of training sequences) is greater than 35%. As a concluding remark, note that because the *a priori* estimate of the signal can trivially guarantee for our example the 25% of success, a working estimate has to ensure at least more than 25% of success. For real-time applications, this could be even enough to guarantee any pergentage of success by artificially introducing suitable redundancies in the transmitted signal. Moreover, from a computational point of view the complexity of the mathematical expression of the algorithm is only apparent, in that the estimate upgrade is achieved in real-time.

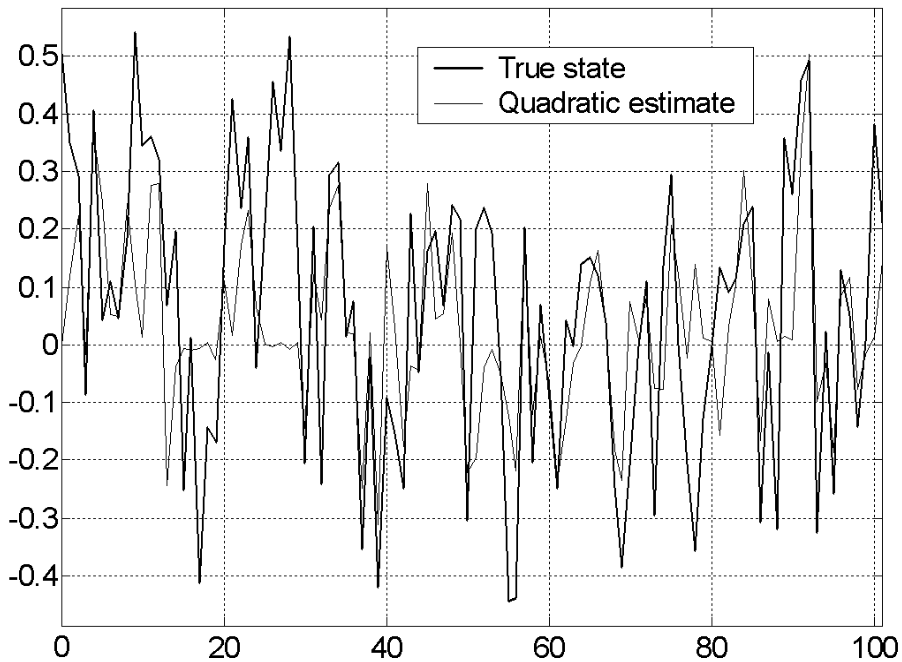

Fig. 5.1 – True and estimated first component.

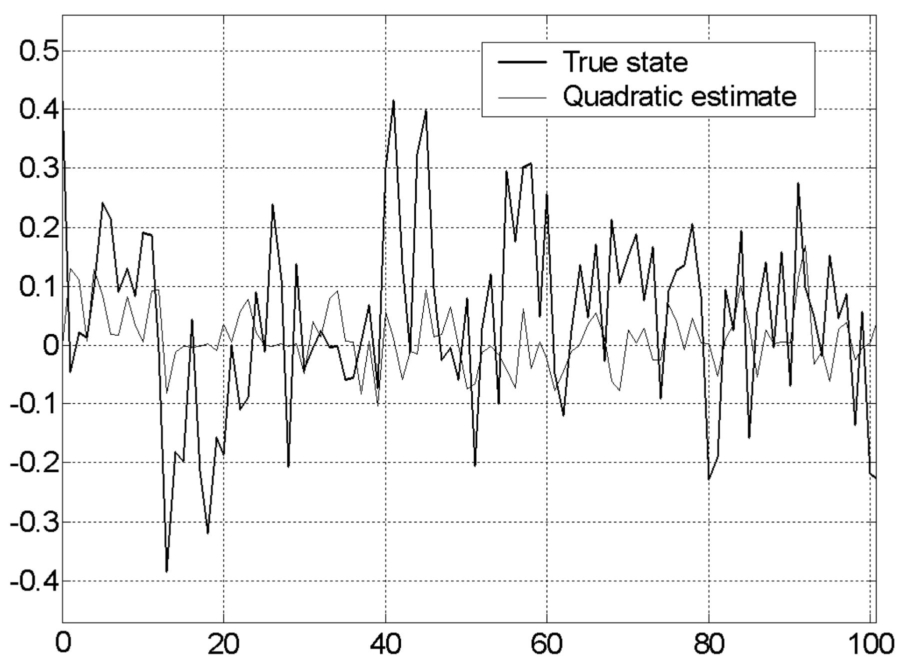

Fig. 5.2 – True and estimated second component.

Appendix

For the ease of the reader, in this Appendix are reported some useful results on the Kronecker algebra. The proofs and other further details can be found in [9].

Let M and N be matrices of dimensions $r \times s$ and $p \times q$ respectively, then the Kronecker product $M \otimes N$ is defined as the $(r \cdot p) \times (s \cdot q)$ matrix

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

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

Definition A.1. Let M be an $r \times s$ matrix:

$$M = [\, m_1 \quad m_2 \quad \dots \quad m_s \,], \tag{A.2}$$

where m_i denotes the i-th column of M. The stack of M is defined as the $r \cdot s$ vector:

$$\mathrm{st}(M) = \begin{bmatrix} m_1 \\ m_2 \\ \vdots \\ m_s \end{bmatrix}. \tag{A.3}$$

Observe that a vector as in (A.3) can be reduced to a matrix M as in (A.2), once it is known the number of the rows r of the original matrix, by considering the inverse operation of the stack denoted by st^{-1}. More generally, let m be a vector in $I\!R^{\mu}$, and r be a divisor of μ. Then the $r \times (\mu/r)$ matrix given by $M = \mathrm{st}^{-1}(m, r)$ is defined so that:

$$\mathrm{st}(M) = m. \tag{A.4}$$

In presence of vectors $m \in I\!R^{(\mu^2)}$, that is their length is given by a square, the notation $\mathrm{st}^{-1}(m)$ has to be considered as a short version of $\mathrm{st}^{-1}(m, \mu)$.

In case of vectors Kronecker products, it is easy to verify that, if $u \in I\!R^r$ and $v \in I\!R^s$, the i-th entry of $u \otimes v$ is given by

$$(u \otimes v)_i = u_l \cdot v_m; \quad l = \left[\frac{i-1}{s}\right] + 1, \quad m = |i-1|_s + 1, \tag{A.5}$$

where $[\cdot]$ and $|\cdot|_s$ denote integer part and s-modulo respectively. Moreover, the Kronecker power of M is defined as

$$\begin{aligned} M^{[0]} &= 1 \in I\!R, \\ M^{[l]} &= M \otimes M^{[l-1]} \quad l \geq 1. \end{aligned} \tag{A.6}$$

Some useful properties of the Kronecker product and stack operation are the following

$$\begin{aligned} (A+B) \otimes (C+D) =& A \otimes C + A \otimes D + B \otimes C + B \otimes D \\ A \otimes (B \otimes C) =& (A \otimes B) \otimes C \\ (A \cdot C) \otimes (B \cdot D) =& (A \otimes B) \cdot (C \otimes D) \\ (A \otimes B)^T =& A^T \otimes B^T \\ st(A \cdot B \cdot C) =& (C^T \otimes A) \cdot st(B) \\ u \otimes v =& st(v \cdot u^T) \\ tr(A \otimes B) =& tr(A) \cdot tr(B) \end{aligned} \tag{A.7}$$

Other useful properties can be found in [2].

According to its definition (A.1), the Kronecker product is not commutative. However, the following result holds:

Lemma A.2. *For any given pair of matrices $A \in I\!R^{r\times s}$, $B \in I\!R^{n\times m}$, it is:*

$$B \otimes A = C_{r,n}^T (A \otimes B) C_{s,m}, \tag{A.8}$$

where $C_{r,n}$, $C_{s,m}$ are defined so that, denoted $\{C_{u,v}\}_{h,l}$ their (h,l) entries:

$$\{C_{u,v}\}_{h,l} = \begin{cases} 1, & \text{if } l = (|h-1|_v)u + \left(\left[\dfrac{h-1}{v}\right] + 1\right); \\ 0, & \text{otherwise.} \end{cases} \tag{A.9}$$

Remark A.3. Observe that $C_{1,1} = 1$, hence in the vector case when $a \in I\!R^r$ and $b \in I\!R^n$, (A.8) becomes

$$b \otimes a = C_{r,n}^T (a \otimes b). \tag{A.10}$$

Moreover, in the vector case the commutation matrices satisfy also the following recursive formula.

Lemma A.4. *Let $a, b \in I\!R^n$ and $l \in I\!N$. Then*

$$b^{[l]} \otimes a = G_l(n)(a \otimes b^{[l]}), \tag{A.11}$$

with the sequence $\{G_l(n) = C_{n,n^l}^T\}$ given by the following recursive equations

$$\begin{aligned} G_1(n) =& C_{n,n}^T, \\ G_l(n) =& (I_{n,1} \otimes G_{l-1}(n)) \cdot (G_1(n) \otimes I_{n,l-1}), \qquad l > 1, \end{aligned} \tag{A.12}$$

where $I_{n,r}$ is the identity matrix in $I\!R^{n^r}$.

A binomial formula can be found for the Kronecker power, which generalizes the classical Newton one.

Lemma A.5. *Let $a, b \in I\!R^n$. For any integer $h \geq 0$ the matrix coefficients of the following binomial power formula:*

$$(a+b)^{[h]} = \sum_{k=0}^{h} M_k^h(n)(a^{[k]} \otimes b^{[h-k]}) \tag{A.13}$$

constitute a set of matrices $\{M_0^h(n), ..., M_h^h(n);\ M_k^h(n) \in I\!R^{n^h \times n^h}\}$ such that:

$$\begin{aligned} M_h^h(n) =& M_0^h(n) = I_{n,h}, \\ M_j^h(n) =& (M_j^{h-1}(n) \otimes I_{n,1}) + (M_{j-1}^{h-1}(n) \otimes I_{n,1}) \\ & \cdot (I_{n,j-1} \otimes G_{h-j}(n)), \quad 1 \leq j \leq h-1, \end{aligned} \tag{A.14}$$

where $G_l(n)$ and $I_{n,l}$ are as in Lemma A.4.

16.

References

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

[2] R. Bellman, *Introduction to matrix analysis*, McGraw-Hill, 1970.

[3] 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.

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

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

[6] A. Germani, C. Manes and P. Palumbo, "Polynomial linear filtering for stochastic systems with Markovian swiching systems," proc. of 42^{th} *Conf. on Decision and Control* (2003CDC), pp. 1392–1397, Maui, Hawaii, 2003.

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

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

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

[10] C.-K. Tzou, R. Raheli and A. Polydoros, "Applications of Per-Survivor Processing to mobile digital communications,"' Proc. IEEE Globecom, Nov. 1993.