Optimal State Estimation for Discrete-Time Markovian Jump Linear Systems, in the Presence of Delayed Output Observations

Ion Matei
ECE Department
University of Maryland
College Park, MD 20742
imatei@umd.edu

Nuno Martins
ECE Department
University of Maryland
College Park, MD 20742
nmartins@umd.edu

John S. Baras
ECE Department
University of Maryland
College Park, MD 20742
baras@umd.edu

***Abstract*— In this paper, we investigate the design of optimal state estimators for Markovian Jump Linear Systems. We consider that the state has two components: the first component is finite valued and is denoted as mode, while the second (continuous) component is in a finite dimensional Euclidean space. The continuous state is driven by a zero mean, white and Gaussian process noise. The observation output has two components: the first is the mode and the second is a linear combination of the continuous state observed and zero mean, white Gaussian noise. Both output components are affected by delays, not necessarily equal. Our paradigm is to design optimal estimators for the current state, given the current output observation. We provide a solution to this paradigm by giving a recursive estimator for the continuous state, in the minimum mean square sense, and a finitely parameterized recursive scheme for computing the probability mass function of the current mode conditioned on the observed output. We show that when the mode is observed with a greater delay then the continuous output component,the optimal estimator nonlinear in the observed outputs.**

I. Introduction

Markovian jump linear systems (MJLS) represents an important class of stochastic time-variant systems due to their ability to model random abrupt changes that occur in a linear plant structure. Linear plants with random time-delays [13] or more general networked control applications [12], where communication networks/channels are used to interconnect remote sensors, actuators and processors, were shown to be prone to MJLS modeling.

Motivated by a wide spectrum of applications, there has been active research in the analysis [3], [7] and in the design of controllers and estimators [6], [7], [9], [10] for Markovian jump linear systems.

We will consider in this paper a MJLS suitable for tracking problem which driven only by noise. The state of the system has two components: the first component is finite valued and is denoted as mode, while the second (continuous) component is in a finite dimensional Euclidean space. The continuous state is driven by some process process noise. The observation output has two components: the first is the mode and the second is a linear combination of the continuous state and some measurement noise.

Existing results solve the problem of state estimation for MJLS in the case of Gaussian noise for two main cases. In the first case, when the entire sequence of output observations up to the current time is considered known, the Minimum Mean Square Error (MMSE) estimator is derived from the Kalman filter for time varying systems [7], [10]. Off-line computation of the filter is inadvisable due to the mode path dependence of the filter's gain. An alternative estimator filter, whose gain depends only on the current mode and for which off-line computations are feasible, is given in [8]. In the second case, when only the continuous output observation is known without any observation of the mode, the optimal nonlinear filter is obtained by a bank of Kalman filters which requires exponentially increasing memory and computation with time [4]. To limit the computational requirements sub-optimal estimators have been proposed in the literature [1], [11], [2]. A linear MMSE estimator, for which the gain matrices can be calculated off-line , is described in [9].

It may be the case in some applications that the two components of the observation output do no reach the estimator computational block simultaneously, each of them being affected by delays, not necessarily equals. For example, the delayed mode observation setup could model networked systems which rely on acknowledgments as a way to deal with unreliable network links. These acknowledgments are not necessarily received instantaneously, instead they are delayed by one or more time-steps.

In this paper we address the problem of state estimation for discrete-time MJLS with Gaussian noise and arbitrary delays on the observation output components.

Notations and abbreviations: Consider a general random process Z_t. We denote by Z_0^t the history of the process from 0 up to time time t as $Z_0^t = \{Z_0, Z_1, ..., Z_t\}$. A realization of Z_0^t is referred by $z_0^t = \{z_0, z_1, ..., z_t\}$. Let $\{X_t | Y_0^{t-h_1} = y_0^{t-h_1}, M_0^{t-h_2} = m_0^{t-h_2}\}$ be a vector valued random process. We denote by $f_{X_t|Y_0^{t-h_1} M_0^{t-h_2}}$ its probability density function (p.d.f.). By $\mu^X_{t|(t-h_1,t-h_2)}$ and $\Sigma^X_{t|(t-h_1,t-h_2)}$ we will refer its mean and covariance matrix respectively. We will compactly write the sum $\sum_{m_0=1}^{s}\sum_{m_1=1}^{s}\cdots\sum_{m_t=1}^{s}$ as $\sum_{m_0^t}$. Assuming that x is a vector in $\mathbb{R}^n$, by the integral $\int f(x)dx$ we understand $\int\ldots\int f(x_1,\ldots,x_n)dx_1\ldots dx_n$, for some function f defined on $\mathbb{R}^n$ and values in $\mathbb{R}$.

Paper organization: This paper has four more sections besides the introduction. After the formulation of the problem in Section II, in Section III we introduce the main results. Two corollaries will present the formulas for the optimal state estimator (discrete and continuous components) in the mean square sense. In Section IV we provide the proofs of these corollaries together with some other supporting results. We end the paper with some conclusion and comments on our solution.

II. Problem formulation

In this section we formulate the problem for the MMSE state estimation for MJLS in the case of delayed output observations.

Definition 2.1: (Markovian jump linear system) Consider n, m, q and s to be given positive integers together with a transition probability matrix $P\in[0,1]^{s\times s}$ satisfying $\sum_{j=1}^{s}p_{ij}=1$, for each i in the set $\mathscr{S}=\{1,\ldots,s\}$, where p_{ij} is the (i,j) element of the matrix P. Consider also a given set of matrices $\{A_i\}_{i=1}^{s}$, $\{C_i\}_{i=1}^{s}$ with $A_i\in\mathbb{R}^{n\times n}$ and $C_i\in\mathbb{R}^{q\times n}$. In addition consider two independent random variable X_0 and M_0 taking values in $\mathbb{R}^n$ and $\mathscr{S}$, respectively. Given the vector valued random processes W_t and V_t taking values in $\mathbb{R}^n$ and $\mathbb{R}^q$ respectively, the following stochastic dynamic equations describe a discrete-time Markovian jump linear system:

$$X_{t+1}=A_{M_t}X_t+W_t \tag{1}$$

$$Y_t=C_{M_t}X_t+V_t. \tag{2}$$

The state of the system is represented by the doublet (X_t,M_t) where $X_t\in\mathbb{R}^n$ is the state continuous component and M_t is the discrete component. The process M_t is a Markovian jump process taking values in $\mathscr{S}$ with conditional probabilities given by $pr(M_{t+1}=j|M_t=i)=p_{ij}$. The observation output is given by the doublet (Y_t,M_t), where $Y_t\in\mathbb{R}^q$ is the continuous component. Throughout this paper we will consider W_t and V_t to be independent identically distributed (i.i.d.) Gaussian noises with zero means and identity covariance matrices. The initial condition vector X_0 has a Gaussian multivariate distribution with mean μ_{X_0} and covariance matrix Σ_{X_0} which, together with the Markovian process M_t and the noises W_t, V_t, are assumed independent for all time instants t.

For simplicity, throughout this paper we will differentiate among the different components of the MJLS state and observation output as following. We will refer to X_t as the state vector and to M_t as mode. If known, we will call Y_t as output observation and M_t and mode observation.

We can now proceed with the formulation of our problem of interest.

Problem 2.1: (*MMSE state and mode estimators for MJLS with delayed output and mode observations*) Consider a Markovian jump linear system as in *Definition 2.1.* Let h_1 and h_2 be two positive integers representing how long the output and the mode observations are delayed. Assuming that the state vector X_t and the mode M_t are not known, and that at the current time the data available consists in the output observations up to time $t-h_1$ ($Y_0^t=y_0^{t-h_1}$) and mode observations up to time $t-h_2$ ($M_0^{t-h_2}=m_0^{t-h_2}$) we want to derive the MMSE estimators for the state vector X_t and the mode indicator function $\mathbb{1}_{\{M_t=i\}}, i\in\mathscr{S}$. More precisely, considering the optimal solution of the MMSE estimators ([14]) we want to compute the following:
MMSE state estimator:

$$\hat{X}_t^{h_1,h_2}=E[X_t|Y_0^{t-h_1}=y_0^{t-h_1},M_0^{t-h_2}=m_0^{t-h_2}], \tag{3}$$

MMSE mode indicator function estimator:

$$\hat{\mathbb{1}}_{\{M_t=m_t\}}^{h_1,h_2}=E[\mathbb{1}_{\{M_t=m_t\}}|Y_0^{t-h_1}=y_0^{t-h_1},M_0^{t-h_2}=m_0^{t-h_2}], \tag{4}$$

where the indicator function $\mathbb{1}_{\{M_t=m_t\}}$ is one if $M_t=m_t$ and zero otherwise.

Remark 2.1: The problem formulated above encompasses a high degree of generality because the delays affecting the output and mode observations may take arbitrary values. Of course, their magnitude and the ordering relation between the delays will affect the form and the complexity of the estimators.

Remark 2.2: Obtaining an MMSE estimation of the mode indicator function allows us to replace any mode dependent function $g(M_t)$ by an estimation $\widehat{g(M_t)}=\sum_{i\in\mathscr{S}}g(i)\hat{\mathbb{1}}_{\{M_t=i\}}$. We are interested in an estimation of the indicator function rather than of the mode itself because the MMSE estimator of the mode can produce real values which may have limited usefulness;

Remark 2.3: Considering the definition of the indicator function, the MMSE mode indicator function estimator can be also written as: $\hat{\mathbb{1}}_{\{M_t=m_t\}}^{h_1,h_2}=pr(M_t=m_t|Y_0^{t-h_1}=y_0^{t-h_1},M_0^{t-h_2}=m_0^{t-h_2})$. Then we can also produce a marginal maximal a posteriori mode estimation expressed in terms of the indicator function: $\hat{M}_t^{h_1,h_2}=arg\max_{i\in\mathscr{S}}pr(M_t=m_t|Y_0^{t-h_1}=y_0^{t-h_1},M_0^{t-h_2}=m_0^{t-h_2})=arg\max_{m_t\in\mathscr{S}}\hat{\mathbb{1}}_{\{M_t=m_t\}}^{h_1,h_2}$.

III. Main result

In this section we present the solution for *Problem 2.1.* We introduce here two corollaries describing the formulas for computing the vectors state and mode indicator function estimators. The proofs of these corollaries are deferred for the next section. Let us first remind ourselves some properties of the Kalman filter for MJLS synthesized in the following theorem.

Theorem 3.1: Consider a discrete MJLS as in *Definition 2.1.* The random processes $\{X_t|Y_0^t=y_0^tM_0^t=m_0^t\}$, $\{X_t|Y_0^{t-1}=y_0^{t-1}M_0^{t-1}=m_0^{t-1}\}$ and $\{Y_t|Y_0^{t-1}=y_0^{t-1}M_0^t=m_0^t\}$ are Gaussian distributed with the means and covariance matrices calculated by the following recursive equations:

$$\Sigma^{X}_{t|(t,t)}{}^{-1} = \Sigma^{X}_{t|(t-1,t-1)}{}^{-1} + C^{T}_{m_t} C_{m_t} \quad (5)$$

$$\mu^{X}_{t|(t,t)} = \Sigma^{X}_{t|(t,t)} \left[C^{T}_{m_t} y_t + \Sigma^{X}_{t|(t-1,t-1)}{}^{-1} \mu^{X}_{t|(t-1,t-1)} \right] \quad (6)$$

$$\mu^{X}_{t|(t-1,t-1)} = A_{m_{t-1}} \mu^{X}_{t-1|(t-1,t-1)} \quad (7)$$

$$\Sigma^{X}_{t|(t-1,t-1)} = A_{m_{t-1}} \Sigma^{X}_{t-1|(t-1,t-1)} A^{T}_{m_{t-1}} + I_n \quad (8)$$

$$\mu^{Y}_{t|(t-1,t)} = C_{m_t} \mu^{X}_{t|(t-1,t-1)} \quad (9)$$

$$\Sigma^{Y}_{t|(t-1,t)} = C_{m_t} \Sigma^{X}_{t|(t-1,t-1)} C^{T}_{m_t} + I_q, \quad (10)$$

with initial conditions $\mu^{X}_{0|(-1,-1)} = \mu_{X_0}$ and $\Sigma^{X}_{0|(-1,-1)} = \Sigma_{X_0}$.

The above theorem contains in equations (5)-(8) an equivalent form of the Kalman filter for MJLS. Besides the Kalman filter equations we added the equations (9)-(10) as they will be used in what follows. Derivation of the Kalman filter equation can be found in [8], [7] for example.

Our main result consists in *Corollaries 3.1 and 3.2* which show the algorithmic steps necessary to compute the MMSE state and mode indicator estimators for MJLS when the observations of the outputs and modes are affected by some arbitrary (but fixed) delays.

Corollary 3.1: Given a MJLS as in *Definition 2.1* and two positive integers h_1 and h_2, the MMSE state estimator from the Problem *2.1* is given by the following formulas:

$$\hat{X}^{h_1,h_2}_{t} =$$

$$\begin{cases} \sum_{m^{t-1}_{t-h_2+1}} \prod_{k=1}^{h_2-1} p_{m_{t-k-1}m_{t-k}} \mu^{X}_{t|(t-h_1,t-1)} & 1 < h_2 \leq h_1 \\ \mu^{X}_{t|(t-h_1,t-1)} & 1 = h_2 \leq h_1 \\ \mu^{X}_{t|(t-h_1,t-1)} & 0 = h_2 < h_1 \\ \sum_{m^{t-h_1}_{t-h_2+1}} c_t(m^{t-h_1}_{t-h_2+1}) \mu^{X}_{t|(t-h_1,t-h_1)} & 0,1 = h_1 < h_2 \\ \sum_{m^{t-1}_{t-h_2+1}} \prod_{k=1}^{h_1-1} p_{m_{t-k-1}m_{t-k}} c_t(m^{t-h_1}_{t-h_2+1}) \mu^{X}_{t|(t-h_1,t-1)} & 1 < h_1 < h_2 \end{cases}$$

where $\mu^{X}_{t|(t-h_1,t-1)}$ (and $\mu^{X}_{t|(t-1,t-1)}$) is computed by the recurrence:

$$\mu^{X}_{t|(t-h_1,t-1)} = \left(\prod_{k=1}^{h_1} A_{m_{t-k}} \right) \mu^{X}_{t-h_1|(t-h_1,t-h_1)} \quad (11)$$

for each of the unknown mode paths represented by a term in the above sums. The means $\mu^{X}_{t-h_1|(t-h_1,t-h_1)}$ (or $\mu^{X}_{t|(t,t)}$) are calculated according to the Kalman filter in equations (5)-(8) and the coefficients $c_t(m^{t-h_1}_{t-h_2+1})$ are given by:

$$c_t\left(m^{t-h_1}_{t-h_2+1}\right) = \frac{\alpha(m^{t-h_1}_{t-h_2+1})}{\sum_{m^{t-h_1}_{t-h_2+1}} \alpha(m^{t-h_1}_{t-h_2+1})}, \quad (12)$$

where

$$\alpha(m^{t-h_1}_{t-h_2+1}) = \prod_{k=0}^{h_2-h_1-1} p_{m_{t-h_1-k-1}m_{t-h_1-k}} \times$$

$$\times f_{Y_{t-h_1-k}|Y_0^{t-h_1-k-1}M_0^{t-h_1-k}}(y_{t-h_1-k}|y_0^{t-h_1-k-1} m_0^{t-h_1-k})$$

and where $f_{Y_{t-h_1-k}|Y_0^{t-h_1-k-1}M_0^{t-h_1-k}}(y_{t-h_1-k}|y_0^{t-h_1-k-1} m_0^{t-h_1-k})$ is the Gaussian p.d.f. of the random process $\{Y_{t-h_1-k}|Y_0^{t-h_1-k-1} M_0^{t-h_1-k}\}$ whose mean and covariance matrix are computed according to equations (9)-(10) introduced in *Theorem 3.1*.

Corollary 3.2: Given a MJLS as in *Definition 2.1* and two positive integers h_1 and h_2, the MMSE mode indicator estimator from *Problem 2.1* is computed according to the next formulas:

$$\hat{\mathbb{1}}^{h_1,h_2}_{\{M_t=m_t\}} =$$

$$= \begin{cases} \prod_{k=1}^{h_2} p_{m_{t-k}m_{t-k+1}} & 0 < h_2 \leq h_1 \\ \sum_{m^{t-1}_{t-h_2+1}} c_t(m^{t}_{t-h_2+1}) & 0 = h_1 < h_2 \\ \sum_{m^{t-1}_{t-h_2+1}} \prod_{k=1}^{h_1} p_{m_{t-k}m_{t-k+1}} c_t(m^{t-h_1}_{t-h_2+1}) & 0 < h_1 < h_2 \end{cases}$$

where $c_t(m^{t-h_1}_{t-h_2+1})$ are computed according to (12).

These results can be regarded as a generalization of the estimation problem for MJLS. Since we assumed the delays to be fixed, the algorithms have a polynomial complexity. However the complexity increases exponentially with the values of the delays which is in accord with the results concerning the Kalman filter for MJLS with no mode observations [4]. We can observe that the ordering of the delays affecting the observation plays a major role in the number of operations of the algorithms and their form. When the delay h_2 affecting the modes observations is greater then the delay h_1 the algorithms become more complex. This is mainly due to the fact that the missing modes are indirectly observed through the output observations. We also notice that in this case the estimators become nonlinear in the outputs due to the coefficients $c_t(m^{t-h_1}_{t-h_2+1})$. In *Corollaries 3.1 and 3.2* we were not concern by the numerical efficiency of the algorithms. It can be noticed however the at the current time the algorithm uses information computed at previous steps indicating that an economy in memory space and computation power can be attained.*Corollaries 3.1 and 3.2* are a consequence of a set of results that will be detailed in the next section.

IV. Proof of the main result

In this section we introduce a series of theorems which will pave the road for proving the main results presented in Section III. In *Theorem 4.1* we characterize the statistical properties of the random process $\{X_t|Y_0^{t-h} = y_0^{t-h}, M_0^{t-1} = m_0^{t-1}\}$ where h is a positive integer. This result is related to the case of state estimation when the output observations are delayed but the modes are all known. In *Theorem 4.2* we analyze the mirrored case presented in *Theorem 4.1*, i.e. we characterize the statistical properties of the random process $\{X_t|Y_0^t = y_0^t, M_0^{t-h} = m_0^{t-h}\}$. From the later mentioned theorem we will derive the the MMSE state and mode indicator estimators when all the output observations are known but the mode observations are delayed. Finally in *Theorem 4.3* we examine the statistical properties of the random process $\{X_t|Y_0^{t-h_1} = y_0^{t-h_1}, M_0^{t-h_2} = m_0^{t-h_2}\}$ where h_1 and h_2 are some known arbitrary positive integer values. This last theorem will provide the formulas for the state and mode indicator MMSE estimators when the output and mode observations are arbitrarily delayed.

To simplify the proofs of *Theorems 4.1* and *4.2* we introduce the following corollary presenting properties of the p.d.f of a linear combination of Gaussian random vectors.

Corollary 4.1: Consider two Gaussian random vector V and X of dimension m and n respectively, with means $\mu_V = 0$ and μ_X and covariance matrices $\Sigma_V = I_m$ and Σ_X respectively. Let Y be a Gaussian random vector resulted from a linear combination of X and V, $Y = CX + V$ where C is a matrix of appropriate dimensions. Then the following holds:

$$\int_{\mathbb{R}^n} f_V(y - Cx) f_X(x) dx = f_Y(y), \tag{13}$$

where $f_Y(y)$ is the multivariate Gaussian p.d.f. of Y with parameters $\mu_Y = C\mu_Y$ and $\Sigma_Y = C\Sigma_X C^T + I_m$. Also,

$$f_V(y - Cx) \cdot f_X(x) = \tilde{f}_{\tilde{X}}(x) \cdot f_Y(y), \tag{14}$$

where $\tilde{f}_{\tilde{X}}(x)$ is a Gaussian p.d.f. with parameters $\mu_{\tilde{X}} = \Sigma_{\tilde{X}}(C^T y + \Sigma_X^{-1}\mu_X)$, $\Sigma_{\tilde{X}}^{-1} = \Sigma_X^{-1} + C^T C$ and, $f_Y(y)$ being defined in (13).

Theorem 4.1: Consider a discrete MJLS as in *Definition 2.1*. Let h be a known strictly positive integer value. Then the p.d.f. of the random process $\{X_t | Y_0^{t-h} = y_0^{t-h} M_0^{t-1} = m_0^{t-1}\}$ is Gaussian with mean computed by:

$$\mu_{t|t-h,t-1} = \left(\prod_{k=1}^{h} A_{m_{t-k}}\right) \mu^X_{t-h|(t-h,t-h)} \tag{15}$$

and covariance matrix given by the recurrence:

$$\Sigma^X_{t-k|(t-h,t-k)} = A_{m_{t-k-1}} \Sigma^X_{t-k-1|(t-h,t-k-1)} A^T_{m_{t-k-1}} + I_n \tag{16}$$

for $k \in \{h-1, h-2, ..., 1, 0\}$ and with initial covariance matrix $\Sigma^X_{t-h|t-h,t-h}$; $\mu^X_{t-h|(t-h,t-h)}$ and $\Sigma^X_{t-h|t-h,t-h}$ are calculated according to the Kalman filter described in equations (5)-(8).

Proof: The Gaussianity is shown by induction. Assume that for a k between $\{0, 1, ..., h-1\}$, $f_{X_{t-k-1}|Y_0^{t-h} M_0^{t-k-1}}$ is a Gaussian p.d.f. Then the $f_{X_{t-k}|Y_0^{t-h} M_0^{t-k}}$ ca be expressed as:

$$f_{X_{t-k}|Y_0^{t-h} M_0^{t-k}}(x_{t-k}|y_0^{t-h} m_0^{t-k}) =$$

$$= \int_{\mathbb{R}^n} f_{X_{t-k} X_{t-k-1}|Y_0^{t-h} M_0^{t-k}}(x_{t-k}, \tilde{x}_{t-k-1}|y_0^{t-h} m_0^{t-k}) d\tilde{x}_{t-k-1} =$$

$$= \int_{\mathbb{R}^n} f_{X_{t-k}|X_{t-k-1} M_{t-k-1}}(x_{t-k}|x_{t-k-1} m_{t-k-1}) \times$$

$$\times f_{X_{t-k-1}|Y_0^{t-h} M_0^{t-k-1}}(x_{t-k-1}|y_0^{t-h} m_0^{t-k-1}) dx_{t-k-1}$$

Using the (13) from *Corollary 4.1* we conclude that $f_{X_{t-k}|Y_0^{t-h} M_0^{t-k}}$ is Gaussian p.d.f. with mean given by

$$\mu^X_{t-k|(t-h,t-k)} = A_{m_{t-k-1}} \mu^X_{t-k-1|(t-h,t-k-1)}$$

and covariance matrix

$$\Sigma^X_{t-k|(t-h,t-k)} = A_{m_{t-k-1}} \Sigma^X_{t-k-1|(t-h,t-k-1)} A^T_{m_{t-k-1}} + I_n$$

Iterating over $k \in \{h-1, h-2, ..., 1, 0\}$ we obtain the equations (15) and (16). ∎

Theorem 4.2: Consider a discrete MJLS as in *Definition 2.1* and let h be a known positive integer value. Then the p.d.f. of the random process $\{X_t | Y_0^t = y_0^t M_0^{t-h} = m_0^{t-h}\}$ is a mixture of Gaussian probability densities. More precisely:

$$f_{X_t|Y_0^t M_0^{t-h}}(x|y_0^t m_0^{t-h}) = \sum_{m_{t-h+1}^t} c_t(m_{t-h+1}^t) f_{X_t|Y_0^t M_0^t}(x|y_0^t, m_0^t) \tag{17}$$

where $c_t(m_{t-h+1}^t) = f_{M_{t-h+1}^t|Y_0^t M_0^{t-h}}(m_{t-h+1}^t|y_0^t m_0^{t-h})$ are the (time varying) mixture coefficients and $f_{X_t|Y_0^t M_0^t}(x|y_0^t, m_0^t)$ is the gaussian p.d.f. of the process $\{X_t | Y_0^t = y_0^t, M_0^t = m_0^t\}$ whose statistics is computed according to the recursions (5)-(8). The coefficients $c_t(m_{t-h+1}^t)$ are computed by the following formula:

$$c_t(m_{t-h+1}^t) = \frac{\prod_{k=0}^{h-1} p_{m_{t-k-1} m_{t-k}} f_{Y_{t-k}|Y_0^{t-k-1} M_0^{t-k}}(y_{t-k}|y_0^{t-k-1} m_0^{t-k})}{\sum_{m_{t-h+1}^{t-k}} \prod_{k=0}^{h-1} p_{m_{t-k-1} m_{t-k}} f_{Y_{t-k}|Y_0^{t-k-1} M_0^{t-k}}(y_{t-k}|y_0^{t-k-1} m_0^{t-k})} \tag{18}$$

where $f_{Y_{t-k}|Y_0^{t-k-1} M_0^{t-k}}$ is the Gaussian p.d.f. of the process $\{Y_{t-k}|Y_0^{t-k-1} = y_0^{t-k-1}, M_0^{t-k} = m_0^{t-k}\}$ whose mean and covariance matrix are expressed in (9) and (10).

Proof: Using the law of marginal probabilities we get:

$$f_{X_t|Y_0^t M_0^{t-h}}(x|y_0^t m_0^{t-h}) = \sum_{m_{t-h+1}^t} f_{X_t M_{t-h+1}^t|Y_0^t M_0^{t-h}}(x, m_{t-h+1}^t|y_0^t m_0^{t-h}) =$$

$$= \sum_{m_{t-h+1}^t} f_{X_t|Y_0^t M_0^t}(x|y_0^t, m_0^t) f_{M_{t-h+1}^t|Y_0^t M_0^{t-h}}(m_{t-h+1}^t|y_0^t m_0^{t-h}) =$$

$$= \sum_{m_{t-h+1}^t} c_t(m_{t-h+1}^t) f_{X_t|Y_0^t M_0^t}(x|y_0^t, m_0^t)$$

Thus we obtained (17). All you are left to do is to compute coefficients of this linear combination. By applying the Bayes rule we get:

$$f_{M_{t-h+1}^t|Y_0^t M_0^{t-h}}(m_{t-h+1}^t|y_0^t m_0^{t-h}) = \frac{f_{Y_0^t M_0^t}(y_0^t m_0^t)}{\sum_{m_{t-h+1}^t} f_{Y_0^t M_0^t}(y_0^t m_0^t)} \tag{19}$$

The p.d.f. $f_{Y_0^t M_0^t}$ can be expressed recursively as:

$$f_{Y_0^t M_0^t}(y_0^t m_0^t) = \int_{\mathbb{R}^n} f_{X_t Y_0^t M_0^t}(x_t y_0^t m_0^t) dx_t =$$

$$= \int_{\mathbb{R}^n} f_{Y_t|X_t M_t}(y_t|x_t m_t) f_{X_t|Y_0^{t-1} M_0^{t-1}}(x_t|y_0^{t-1} m_0^{t-1}) dx_t$$

$$p_{m_{t-1} m_t} f_{Y_0^{t-1} M_0^{t-1}}(y_0^{t-1} m_0^{t-1}).$$

Applying (13) we obtain:

$$f_{Y_0^t M_0^t}(y_0^t m_0^t) = f_{Y_t|Y_0^{t-1} M_0^t}(y_t|y_0^{t-1} m_0^t) p_{m_{t-1} m_t} f_{Y_0^{t-1} M_0^{t-1}}(y_0^{t-1} m_0^{t-1})$$

Using this recursive expression we get:

$$f_{Y_0^t M_0^t}(y_0^t m_0^t) =$$

$$= \prod_{k=0}^{h-1} p_{m_{t-k-1}m_{t-k}} f_{Y_{t-k}|Y_0^{t-k-1}M_0^{t-k}}(y_{t-k}|y_0^{t-k-1}m_0^{t-k}) f_{Y_0^{t-h}M_0^{t-h}}(y_0^{t-h}m_0^{t-h})$$

By replacing the previous expression in (19) we obtain the coefficients $c_t(m_{t-h+1}^t)$ expressed in (18). We can conclude de proof by making the observations that the p.d.f. $f_{Y_{t-k}|Y_0^{t-k-1}M_0^{t-k}}$ is completely characterized in *Theorem 3.1*, equation (9) and (10). ∎

Theorem 4.3: Consider a discrete MJLS as in *Definition 2.1* and let h_1 and h_2 be two known non-negative integers. Then the p.d.f. of the random process $\{X_t|Y_0^{t-h_1} = y_0^{t-h_1} M_0^{t-h_2} = m_0^{t-h_2}\}$ is given by the following formula:

$$f_{X_t|Y_0^{t-h_1}M_0^{t-h_2}}(x_t|y_0^{t-h_1}m_0^{t-h_2}) =$$

Case 1: $h_1 \geq h_2,\ h2 > 1$

$$= \sum_{m_{t-h_2+1}^{t-1}} \prod_{k=1}^{h_2-1} p_{m_{t-k-1}m_{t-k}} f_{X_t|Y_0^{t-h_1}M_0^{t-1}}(x_t|y_0^{t-h_1}m_0^{t-1}) \qquad (20)$$

Case 2: $(h_1 \geq h_2,\ h_2 = 1)$ or $(h_1 > h_2,\ h_2 = 0)$

$$= f_{X_t|Y_0^{t-h_1}M_0^{t-1}}(x_t|y_0^{t-h_1}m_0^{t-1}) \qquad (21)$$

Case 3: $h_1 < h_2$, $(h_1 = 0$ or $h_1 = 1)$

$$= \sum_{m_{t-h_2+1}^{t-h_1}} c_t(m_{t-h_2+1}^{t-h_1}) f_{X_t|Y_0^{t-h_1}M_0^{t-h_1}}(x_t|y_0^{t-h_1}m_0^{t-h_1})$$

Case 4: $h_1 < h_2,\ h_1 > 1$

$$= \sum_{m_{t-h_2+1}^{t}} \prod_{k=1}^{h_1-1} p_{m_{t-k-1}m_{t-k}} c_t(m_{t-h_2+1}^{t-h_1}) f_{X_t|Y_0^{t-h_1}M_0^{t-1}}(x_t|y_0^{t-h_1}m_0^{t-1}) \qquad (22)$$

where the p.d.f. $f_{X_t|Y_0^{t-h_1}M_0^{t-1}}$ is characterized in *Theorem 4.1* and the coefficients $c_t(m_{t-h_2+1}^{t-h_1})$ are given by

$$c_t\left(m_{t-h_2+1}^{t-h_1}\right) = \frac{\alpha(m_{t-h_2+1}^{t-h_1})}{\sum_{m_{t-h_2+1}^{t-h_1}} \alpha(m_{t-h_2+1}^{t-h_1})} \qquad (23)$$

where

$$\alpha(m_{t-h_2+1}^{t-h_1}) = \prod_{k=0}^{h_2-h_1-1} p_{m_{t-h_1-k-1}m_{t-h_1-k}} \times$$

$$\times f_{Y_{t-h_1-k}|Y_0^{t-h_1-k-1}M_0^{t-h_1-k}}(y_{t-h_1-k}|y_0^{t-h_1-k-1}m_0^{t-h_1-k})$$

with $f_{Y_{t-h_1-k}|Y_0^{t-h_1-k-1}M_0^{t-h_1-k}}(y_{t-h_1-k}|y_0^{t-h_1-k-1}m_0^{t-h_1-k})$ the Gaussian p.d.f. of the random process $\{Y_{t-h_1-k}|Y_0^{t-h_1-k-1}M_0^{t-h_1-k}\}$ whose mean and covariance matrix are computed using recursions (9)-(10).

Proof: We will start the proof with Cases 1 and 4 since they are the most general. The rest of the cases are derived immediately from the fore-mentioned ones. Proof of Case 1 $h_1 \geq h_2,\ h2 > 1$:

$$f_{X_t|Y_0^{t-h_1}M_0^{t-h_2}}(x_t|y_0^{t-h_1}m_0^{t-h_2}) =$$

$$= \sum_{m_{t-h_2+1}^{t-1}} f_{X_tM_{t-h_2+1}^{t-1}|Y_0^{t-h_1}M_0^{t-h_2}}(x_t m_{t-h_2+1}^{t-1}|y_0^{t-h_1}m_0^{t-h_2}) =$$

$$\sum_{m_{t-h_2+1}^{t-1}} f_{X_t|Y_0^{t-h_1}M_0^{t-1}}(x_t|y_0^{t-h_1}m_0^{t-1}) f_{M_{t-h_2+1}^{t-1}|Y_0^{t-h_1}M_0^{t-h_2}}(m_{t-h_2+1}^{t-1}|y_0^{t-h_1}m_0^{t-h_2})$$

Observing that $\{M_{t-h_2+1}^{t-1}|Y_0^{t-h_1}M_0^{t-h_2}\} = \{M_{t-h_2+1}^{t-1}|M_0^{t-h_2}\}$ and that

$$f_{M_{t-h_2+1}^{t-1}|Y_0^{t-h_1}M_0^{t-h_2}}(m_{t-h_2+1}^{t-1}|y_0^{t-h_1}m_0^{t-h_2}) = \prod_{k=1}^{h_2-1} p_{m_{t-k-1}m_{t-k}}$$

we conclude the proof of this case. Proof of Case 4 $h_1 < h_2,\ h_1 > 1$:

$$f_{X_t|Y_0^{t-h_1}M_0^{t-h_2}}(x_t|y_0^{t-h_1}m_0^{t-h_2}) =$$

$$= \sum_{m_{t-h_2+1}^{t-1}} f_{X_tM_{t-h_2+1}^{t-1}|Y_0^{t-h_1}M_0^{t-h_2}}(x_t m_{t-h_2+1}^{t-1}|y_0^{t-h_1}m_0^{t-h_2}) =$$

$$\sum_{m_{t-h_2+1}^{t-1}} f_{X_t|Y_0^{t-h_1}M_0^{t-1}}(x_t|y_0^{t-h_1}m_0^{t-1}) \times$$

$$\times f_{M_{t-h_2+1}^{t-1}|Y_0^{t-h_1}M_0^{t-h_2}}(m_{t-h_2+1}^{t-1}|y_0^{t-h_1}m_0^{t-h_2}).$$

We separate the last p.d.f. in the above sum in two terms:

$$f_{M_{t-h_2+1}^{t-1}|Y_0^{t-h_1}M_0^{t-h_2}}(m_{t-h_2+1}^{t-1}|y_0^{t-h_1}m_0^{t-h_2}) =$$

$$= f_{M_{t-h_1+1}^{t-1}|Y_0^{t-h_1}M_0^{t-h_1}}(m_{t-h_1+1}^{t-1}|y_0^{t-h_1}m_0^{t-h_1}) \times$$

$$\times f_{M_{t-h_2+1}^{t-h_1}|Y_0^{t-h_1}M_0^{t-h_2}}(m_{t-h_2+1}^{t-h_1}|y_0^{t-h_1}m_0^{t-h_2})$$

From the Markovian property of the process M_t the first term is:

$$f_{M_{t-h_1+1}^{t-1}|Y_0^{t-h_1}M_0^{t-h_1}}(m_{t-h_1+1}^{t-1}|y_0^{t-h_1}m_0^{t-h_1}) =$$

$$= f_{M_{t-h_1+1}^{t-1}|M_0^{t-h_1}}(m_{t-h_1+1}^{t-1}|m_0^{t-h_1}) = \prod_{k=1}^{h_1-1} p_{m_{t-k-1}m_{t-k}}.$$

We can notice that $f_{M_{t-h_2+1}^{t-h_1}|Y_0^{t-h_1}M_0^{t-h_2}}$ is a shifted (by h_1) version of the coefficients c_t introduced in the *Theorem 4.2*. Thus:

$$f_{M_{t-h_2+1}^{t-h_1}|Y_0^{t-h_1}M_0^{t-h_2}}(m_{t-h_2+1}^{t-h_1}|y_0^{t-h_1}m_0^{t-h_2}) = c_t(m_{t-h_2+1}^{t-h_1})$$

where $c_t(m_{t-h_2+1}^{t-h_1})$ are given by (23).

Let us now address the particular cases 2,3. When $h_1 \geq h_2,\ h_2 = 1$, (21) follows trivially. If $h_1 > h_2,\ h_2 = 0$ we obtain (21) again from the fact that $\{X_t|Y_0^{t-h_1}M_0^t\} = \{X_t|Y_0^{t-h_1}M_0^{t-1}\}$ since we have at most $t-1$ output measurements and since X_t does not depend on M_t; in Case 3, for $h_1 < h_2$, $h_1 = 0$ we satisfy the conditions of *Theorem 4.2*. For $h_1 < h_2, h_1 = 1$ we follow the same lines as in the Case 4 with the difference that since $h_1 = 1$ there will be no products of probabilities multiplying the terms in the sum. ∎

Corollary 3.1 *Proof:* The proof follows from the linearity of the expectation operator and by applying the results about the p.d.f. $f_{X_t|Y_0^{t-h_1}M_0^{t-h_2}}$ detailed in *Theorem 4.3* together with the properties of $f_{X_t|Y_0^{t-h}M_0^{t-1}}$ and $f_{X_t|Y_0^{t}M_0^{t-h}}$ shown in *Theorems 4.1 and 4.2*. ■

Corollary 3.2 *Proof:* From the optimal estimator formula we have:

$$\hat{\mathbb{1}}^{h_1,h_2}_{\{M_t=m_t\}} = E[\mathbb{1}_{\{M_t=m_t\}}|Y_0^{t-h_1} = y_0^{t-h_1}, M_0^{t-h_2}] =$$
$$= f_{M_t|Y_0^{t-h_1}M_0^{t-h_2}}(m_t|y_0^{t-h_1}m_0^{t-h_2})$$

In the case $h_1 \geq h_2$ from the Markovian property of the process M_t and from the fact that $\{M_t|Y_0^{t-h_1}M_0^{t-h_2}\} = \{M_t|M_{t-h_2}\}$ we obtain:

$$\hat{\mathbb{1}}_{\{M_t=m_t|Y_0^{t-h_1}=y_0^{t-h_1}M_0^{t-h_2}=m_0^{t-h_2}\}} = \prod_{k=1}^{h_2-1} p_{m_{t-k-1}m_{t-k}}$$

In the case when $h_1 < h_2,\ h_1 \geq 1$ we have:

$$\hat{\mathbb{1}}_{\{M_t=m_t|Y_0^{t-h_1}=y_0^{t-h_1}M_0^{t-h_2}=m_0^{t-h_2}\}} =$$
$$= \sum_{m_{t-h_2+1}^{t-1}} f_{M_{t-h_2+1}^{t-1}|Y_0^{t-h_1}M_0^{t-h_2}}(m_{t-h_2+1}^{t}|y_0^{t-h_1}m_0^{t-h_2}) =$$
$$= \sum_{m_{t-h_2+1}^{t-1}} f_{M_{t-h_1+1}^{t}|Y_0^{t-h_1}M_0^{t-h_1}}(m_{t-h_1+1}^{t}|y_0^{t-h_1}m_0^{t-h_1})\times$$
$$\times f_{M_{t-h_2+1}^{t-h_1}|Y_0^{t-h_1}M_0^{t-h_2}}(m_{t-h_2+1}^{t-h_1}|y_0^{t-h_1}m_0^{t-h_2}) =$$
$$= \sum_{m_{t-h_2+1}^{t-h_1}} \left(\prod_{k=0}^{h_1-1} p_{m_{t-k-1}m_{t-k}}\right) c_t(m_{t-h_2+1}^{t-h_1})$$

where the last line was deduced from a similar analysis as in the proof of *Theorem 4.3*. When $h_1 < h_2$ and $h_1 = 0$ we obtain a formula as in the previous lines, with the difference that there will be no longer any product of probabilities. ■

V. Conclusions

In this paper we considered the problem of state estimation for MJLS when the two components of the observation output are affected by delays. We gave the formulas for the optimal estimators for both the continuous and discrete components of the state.

These formulas admit recursive implementation and have polynomial complexity and therefore are feasible for practical implementation. However the different ordering between the delays affects the complexity and structure of the estimators. An important observation is that when we have less mode observations then output observations the estimators become nonlinear in the outputs. Our problem setup can be viewed as a generalization of the state estimation problem for MJLS since represents the link between the main cases addressed in the literature: availability of mode and output observations and availability of only output observations.

References

[1] G.A. Ackerson and K.S. Fu, ”On the state estimation in switching environments”, *IEEE Transaction on Automatic Control*, 15:10-17, 1970.

[2] Y. Bar-Shalom, ”Tracking methods in a multi-target environment”, *IEEE Transaction on Automatic Control*, 23:618-626, 1978.

[3] W.P. Blair and D.D. Sworder, ”Feedback control of a class of linear discrete systems with jump parameters and quadratic cost criteria”, *International Journal of Control*, 21:833-844,1975.

[4] Y. Bar-Shalom and X.R. Li, ”Estimation and Tracking. Principles, Techniques and Software”, Artech House, 1993.

[5] Y. Ji, H.J. Chizeck, X. Feng and K.A. Loparo, ”Stability and control of dicrete-time jump linear systems”, *Control Theory Adv. Technol.*, vol. 7, no. 2, 1991, pages 247-270.

[6] H.J. Chizeck, A.S. Willsky and D. Castanon, ”Discrete-time Markovian-jump linear quadratic optimal control”, *Int. J. Control*, vol. 43, no. 1, 1986, pages 213-231.

[7] H.J. Chizeck and Y. Ji, ”Optimal quadratic control of jump linear systems with Gaussian noise in discrete-time”, *Proceedings of the 27th IEEE Conference on Decision and Control*, pages 1989-1992, 1988.

[8] O.L.V. Costa, M.D. Fragoso and R.P. Marques, ”Discrete-Time Markov Jump Linear Systems”, Springer, 2005.

[9] O.L.C. Costa, ”Linear Minimum Mean Square Error Estimation for Discrete-Time Markovian Jump Linear Systems”, *IEEE Transactions on Automatic Control*, vol. 39, no. 8, 1994.

[10] M.H.A. Davis and R.B. Vinter, ”Stochastic Modelling and Control”, Chapman and Hall, 1985.

[11] F. Dufour and R.J. Elliot, ”Adaptive control of linear systems with Markov perturbations”, *IEEE Transaction on Automatic Control*, 43:351-372,1997.

[12] J. P. Hespanha, P. Naghshtabrizi and Xu Yonggang , *“ A Survey of Recent Results in Networked Control Systems,”* Proceedings of the IEEE Volume 95, Issue 1, Jan. 2007 Page(s):138 - 162

[13] J. Nilsson, B. Bernhardsson and B. Wittenmark, *“Stochastic analysis and control of real-time systems with random time delays,”* Automatica, Volume 34, Issue 1, January 1998, Pages 57-64

[14] I.B. Rhodes, ”A Tutorial Introduction to Estimation and Filtering”, *IEEE Transactions on Automatic Control*, vol. AC-16, no. 6, 1971.

[15] L. Aggoun and R. Elliot, ”Measure Theory and Filtering”, Cambridge University Press, 2004.