This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TAC.2017.2692180, IEEE Transactions on Automatic Control

Linear minimum mean square filters for Markov jump linear systems

Eduardo F. Costa and Benoîte de Saporta

***Abstract*— This paper studies optimal mean square error estimation for discrete-time linear systems with observed Markov jump parameters. New linear estimators are introduced by considering a cluster information structure in the filter design. The set of filters constructed in this way can be ordered in a lattice according to the refines of clusters of the Markov chain, including the linear Markovian estimator at one end (with only one cluster) and the Kalman filter at the other end (with as many clusters as Markov states). The higher is the number of clusters, the heavier are pre-computations and smaller is the estimation error for embedded sequences of partitions, so that the cardinality and choice of the clusters allows for a trade-off between performance and computational requirements. In this paper we propose the estimator, give the formulas for pre-computation of gains, present some properties, and give an illustrative numerical example.**

I. Introduction

There is a vast number of applications benefiting from the nice properties of the Kalman filter (KF). Among these properties, the possibility of pre-computation of gains [1], [8] is of much relevance for applications. However, in some cases pre-computation is not possible or viable due to missing *a priori* relevant information. This is the case when using the KF to estimate the state of Markov jump linear systems (MJLS), since the parameters are not known prior to the current time instant k, in fact they depend on the Markov chain current state $\theta(k)$. Then, to use KF for MJLS one needs to do either online computation of the gains or offline pre-computation of a number of sample path dependent gains, a figure that grows exponentially with time. This drawback of KF for MJLS is one of the main motivations behind the emergence of other filters for this class of systems, see e.g. [2], [4], [6], [7]. Among them, the one that is closer to the KF in terms of structure and performance is the linear minimum mean square estimator (LMMSE), see for instance [3], [5]. Indeed, the LMMSE computation relies on coupled Riccati equations that are similar to the ones arising in Kalman filtering. Another similarity is that both are optimal in the mean square error sense, although under different constraints.

In this paper we study minimum mean square error linear estimation for discrete-time MJLS when the jump variable is observed. For this class of systems, a Markovian LMMSE has been introduced in [5]; we refer to that filter as the (standard) LMMSE. Here, we relax the markovianity constraint by allowing clustered information of the Markov chain to be considered when designing the filter gains. By clustered information, we mean that we have a partition of the state space of the Markov chain into several classes called clusters, and we observe the trajectory of classes the chain belongs to along time. The clusters may be chosen as they are considered as a design parameter, establishing a trade-off between complexity and performance that can be explored in practical systems aiming at the best feasible performance. At one extreme when only one cluster is taken into account our filter is equivalent to the LMMSE, and at another extreme with N clusters we retrieve the KF; intermediary number of clusters leads to filters with variable performances and computational burden (we are mainly concerned with the memory needed to store the gains and the CPU time to compute them). Reasonably enough, the higher is the number of clusters the smaller is the attained estimation error for some distributions of states in the prescribed number of clusters and higher is the number of gains to compute. Note that the error is decreasing along a sequence of partitions embedded in each other with an increasing number of clusters, as illustrated in Figure 1 given in Section V.

We start with a simple, precise formulation of the optimal estimation problem in Section II, with the estimator in the classical form of Luenberger observers. We then proceed in Section III to a

This work was supported by Inria associate team CDSS and ANR grant Piece, FAPESP, FAPESP/FAPs/INRIA/INS2i-CNRS under Grant 13/50759-3, and CNPq.

Eduardo F. Costa is with Univ. São Paulo - Instituto de Ciências Mathemáticas e de Computação, C.P. 668, 13560-970, São Carlos, SP, Brazil e-mail: efcosta@icmc.usp.br.

Benoîte de Saporta is with University of Montpellier, F-34095 Montpellier, France, CNRS, IMAG, UMR 5149, F-34095 Montpellier, France and Inria Bordeaux Sud Ouest, team CQFD, F-33400 Talence, France. e-mail: Benoite.de-Saporta@umontpellier.fr

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TAC.2017.2692180, IEEE Transactions on Automatic Control

constructive proof that evaluates the estimation error and uses the completion of squares method to obtain the optimal gains. Some remarks on how the proposed class of estimators includes both the KF and LMMSE, and on the number of gains and Riccati-like equations to be precomputed, are presented. Some variants of the studied optimization problem and how to extend optimality to general estimators are briefly discussed in Section IV. We have also included a numerical example in Section V comparing the estimation error computed via the proposed formula. The example makes clear that the performance is strongly dependent on the number of clusters and how the Markov states are distributed in the clusters.

II. Problem formulation

Consider the MJLS

$$\begin{aligned} x_{k+1} &= A_{\theta(k)}x_k + G_{\theta(k)}w_k \\ y_k &= L_{\theta(k)}x_k + H_{\theta(k)}w_k, \qquad k \geq 0, \end{aligned} \tag{1}$$

with initial condition x_0 such that $E[x_0] = \bar{x}$ and $E[x_0x_0'] = \Psi$. The variable $\theta(k)$ denotes the state of a Markov chain with finite state space $\{1, 2, \ldots, N\}$ and initial distribution $\pi_0 = [\Pr(\theta(0) = 1) \cdots \Pr(\theta(0) = N)]$. The noise sequence w is independent from x_0 and the Markov chain θ, $E[w_k] = 0$ and $E[w_kw_k']$ is the identity matrix for all k. We assume that $G_iH_i' = 0$ and $H_iH_i' > 0$, $1 \leq i \leq N$. In (1), $A_{\theta(k)} = A_i$ whenever $\theta(k) = i$, A_i belonging to a given set of matrices, and similarly for the other matrices G, L and H. We consider Luenberger observers[1]

$$\widehat{x}_{k+1} = A_{\theta(k)}\widehat{x}_k + M_k(y_k - L_{\theta(k)}\widehat{x}_k), \tag{2}$$

where matrix M_k is referred to as the filter gain, and the initial estimate is given by $\widehat{x}_0 = \bar{x}$. This produces an estimation error $\widetilde{x} = x - \widehat{x}$ satisfying

$$\begin{aligned} \widetilde{x}_{k+1} = (A_{\theta(k)} &- M_kL_{\theta(k)})\widetilde{x}_k \\ &+ (G_{\theta(k)} - M_kH_{\theta(k)})w_k, \end{aligned} \tag{3}$$

and $\widetilde{x}_0 = x_0 - \bar{x} \sim N(0, \Psi)$. As for the Markov chain, we consider a partition $S_1, \ldots, S_{N_C}$ for its state space, and employ the variable $\rho(k)$ to indicate the partition visited at time k, that is,

$$\rho(k) = \sum_{m=1}^{N_C} m \times \mathbb{1}_{\{\theta(k)\in S_m\}}.$$

We assume that the observations of ρ up to time k are available to calculate the filter gains. We also assume that the jump variable $\theta(k)$ is available, however we do not take into account its past values, that is, $\theta(0), \ldots, \theta(k-1)$ are not considered when calculating the gain (to avoid an excessive number of branches, as explained earlier). Moreover, the gain should not depend on future information, as it has to be implemented at every time instant k. Therefore, we impose that the filter matrices at time instant k are in the form

$$M_k = h_k(\rho(0), \ldots, \rho(k-1), \theta(k)), \tag{4}$$

for measurable functions h_k. Gains satisfying this constraint are referred to as *feasible gains*. We are interested in obtaining the minimum mean square state estimation at a given time $s \geq 0$, leading to the optimization problem

$$\min_{M_0,\ldots,M_s} E\{\|x_s - \widehat{x}_s\|^2|\mathcal{R}_s\}, \qquad \text{s.t. (4)}, \tag{5}$$

where $\mathcal{R}_s = \{\rho(0), \ldots, \rho(s), y(0), \ldots, y(s), \theta(s)\}$. We refer to filter (2) satisfying (5) as the *clustered information LMMSE*, or CLMMSE for short.

III. Clustered information LMMSE computation

Consider the MJLS in (1) and the filter in (2) with an arbitrary sequence of feasible gains $M = \{M_k, k \geq 0\}$. For each $k \geq 0$, $0 \leq i \leq N$ and $0 \leq \ell_m \leq N_C$, $0 \leq m \leq k-1$, we define

$$\begin{aligned} &X_{\ell_0,\ldots,\ell_{k-1},i,k}(M) \\ &\quad = E(\widetilde{x}_k\widetilde{x}_k'\mathbb{1}_{\{\rho(0)=\ell_0,\ldots,\rho(k-1)=\ell_{k-1},\theta(k)=i\}}). \end{aligned} \tag{6}$$

The variable X plays an important role in the derivation of the formula for the optimal filter because the optimal gains $M^\star = \{M_k^\star, 0 \leq k \leq s\}$ are such that $X_{\ell_0,\ldots,\ell_{k-1},i,k}(M^\star) = Y_{\ell_0,\ldots,\ell_{k-1},i,k}$ where Y is the solution of a Riccati-like equation, as we shall see in the next theorem. The physical interpretation for X is that, when divided by $\Pr(\rho(0) = \ell_0, \ldots, \rho(k-1) = \ell_{k-1}, \theta(k) = i)$ it gives the conditional error covariance matrix.

The optimal gain sequence can be (pre)computed based on the following sets of matrices. Let $Y_{i,0} = \pi_i(0)\Psi$ for each $1 \leq i \leq N$. Let us denote

$$\begin{aligned} p_{\ell_0,\ldots,\ell_{k-1},i,k} = \Pr(\rho(0) &= \ell_0, \ldots, \\ &\rho(k-1) = \ell_{k-1}, \theta(k) = i), \end{aligned}$$

$$\widetilde{S} = \{j \in S_{\ell_{k-1}} : p_{\ell_0,\ldots,\ell_{k-2},j,k-1} \neq 0\}.$$

for each $k \geq 1$, let $0 \leq m \leq k-1$ and compute for each $0 \leq i \leq N$ and $0 \leq \ell_m \leq M$ the variable $Y_{\ell_0,\ldots,\ell_{k-1},i,k}$ according to (7).

Theorem 3.1: Given the realization of $\theta(k)$, $k \leq s$, and the corresponding cluster observations $\rho(0), \ldots, \rho(s-1)$, the gains $M^\star = \{M_0^\star, \ldots, M_s^\star\}$ of the CLMMSE can be computed for each $k \leq s$

[1] General recursive linear estimators are briefly addressed in Section IV-D.

$$Y_{\ell_0,\ldots,\ell_{k-1},i,k} = \begin{cases} 0, & \text{if } p_{\ell_0,\ldots,\ell_{k-1},i,k} = 0, \\ \sum_{j\in\widetilde{S}} p_{ji}\big[A_j Y_{\ell_0,\ldots,\ell_{k-2},j,k-1}A_j' + p_{\ell_0,\ldots,\ell_{k-2},j,k-1}G_jG_j' + A_jY_{\ell_0,\ldots,\ell_{k-2},j,k-1}L_j'\big(L_j \\ \quad \times Y_{\ell_0,\ldots,\ell_{k-2},j,k-1}L_j' + p_{\ell_0,\ldots,\ell_{k-2},j,k-1}H_jH_j'\big)^{-1}L_jY_{\ell_0,\ldots,\ell_{k-2},j,k-1}A_j'\big], & \text{otherwise,} \end{cases} \tag{7}$$

as follows. If $\Pr(\rho(0),\ldots,\rho(k-1),\theta(k)) = 0$ then $M_k^\star = 0$, otherwise

$$\begin{aligned} M_k^\star &= A_{\theta(k)}Y_{\rho(0),\ldots,\rho(k-1),\theta(k),k}L_{\theta(k)}' \\ &\quad \cdot \big(L_{\theta(k)}Y_{\rho(0),\ldots,\rho(k-1),\theta(k),k}L_{\theta(k)}' \\ &\quad + \Pr(\rho(0),\ldots,\rho(k-1),\theta(k))H_{\theta(k)}H_{\theta(k)}'\big)^{-1}, \end{aligned} \tag{8}$$

where Y is given in (7). Moreover, for each $k \le s$,

$$X_{\rho(0),\ldots,\rho(k-1),\theta(k),k}(M^\star) = Y_{\rho(0),\ldots,\rho(k-1),\theta(k),k}, \tag{9}$$

and the variable X is optimal in the sense that, for any gain sequence $M = \{M_0,\ldots,M_s\}$,

$$\begin{aligned} &X_{\rho(0),\ldots,\rho(k-1),\theta(k),k}(M^\star) \\ &\le \quad X_{\rho(0),\ldots,\rho(k-1),\theta(k),k}(M), \quad 0 \le k \le s. \end{aligned} \tag{10}$$

Proof: In this proof we denote for brevity

$$\begin{aligned} A_{i,k}^F &= A_i - M_kL_i,\ G_{i,k}^F = G_i - M_kH_i, \\ \mathcal{R} &= \{\rho(0) = \ell_0,\ldots,\rho(k) = \ell_k\}, \\ \mathcal{R}^- &= \{\rho(0) = \ell_0,\ldots,\rho(k-1) = \ell_{k-1}\}. \end{aligned}$$

We start showing that (9) and (10) are true for the gains prescribed in (8). We proceed by induction in k. For the time instant $k = 0$ we have that the initial estimate is given by $\widehat{x}_0 = \bar{x}$, yielding $\widetilde{x}_0 = x_0 - \bar{x} \sim N(0,\Psi)$ (irrespectively of the filter gains), hence

$$\begin{aligned} X_{i,0}(M^\star) &= X_{i,0}(M) = E(\widetilde{x}_0\widetilde{x}_0'\mathbb{1}_{\{\theta(0)=i\}}) \\ &= E(\widetilde{x}_0\widetilde{x}_0')E(\mathbb{1}_{\{\theta(0)=i\}}) = \Psi\pi_i(0) = Y_{i,0}. \end{aligned} \tag{11}$$

By the induction hypothesis we assume that (9) and (10) are valid for some $0 \le k < s$. In order to complete the induction, we now consider the time instant $k+1$. For a filter with an arbitrary sequence of feasible gains, denoted by $M = \{M_0,\ldots,M_s\}$, and a given realization $\rho(0) = \ell_0,\ldots,\rho(k) = \ell_k, \theta(k+1) = i$ we have

$$\begin{aligned} &X_{\ell_0,\ldots,\ell_k,i,k+1}(M) \\ &\qquad = E(\widetilde{x}_{k+1}\widetilde{x}_{k+1}'\mathbb{1}_{\{\mathcal{R}\}}\mathbb{1}_{\{\theta(k+1)=i\}}). \end{aligned}$$

Note that the above quantity turns out to be zero whenever $p_{\ell_0,\ldots,\ell_k,i,k+1} = 0$, irrespectively of M, which makes (9) and (10) trivially true for $k+1$ in this case. Now, in case $p_{\ell_0,\ldots,\ell_k,i,k+1} \neq 0$ we write

$$\begin{aligned} &X_{\ell_0,\ldots,\ell_k,i,k+1}(M) \\ &= E(\widetilde{x}_{k+1}\widetilde{x}_{k+1}'\mathbb{1}_{\{\mathcal{R}\}}\mathbb{1}_{\{\theta(k+1)=i\}}) \\ &= E\Big(\big(A_{\theta(k),k}^F\widetilde{x}_k + G_{\theta(k),k}^Fw_k\big) \\ &\quad \times \big(A_{\theta(k),k}^F\widetilde{x}_k + G_{\theta(k),k}^Fw_k\big)'\mathbb{1}_{\{\mathcal{R}\}}\mathbb{1}_{\{\theta(k+1)=i\}}\Big) \end{aligned}$$

leading to

$$\begin{aligned} X_{\ell_0,\ldots,\ell_k,i,k+1}(M) &= E\Big([A_{\theta(k),k}^F\widetilde{x}_k\widetilde{x}_k'A_{\theta(k),k}^{F\prime} \\ &\quad + G_{\theta(k),k}^Fw_kw_k'G_{\theta(k),k}^{F\prime}]\mathbb{1}_{\{\mathcal{R}\}}\mathbb{1}_{\{\theta(k+1)=i\}}\Big) \\ &= E\Big(\sum_{j=1}^N[A_{\theta(k),k}^F\widetilde{x}_k\widetilde{x}_k'A_{\theta(k),k}^{F\prime} + G_{\theta(k),k}^Fw_k \\ &\quad \times w_k'G_{\theta(k),k}^{F\prime}]\mathbb{1}_{\{\mathcal{R}\}}\mathbb{1}_{\{\theta(k+1)=i\}}\mathbb{1}_{\{\theta(k)=j\}}\Big) \\ &= E\Big(\sum_{j\in S_{\ell_k}}[A_{j,k}^F\widetilde{x}_k\widetilde{x}_k'A_{j,k}^{F\prime} \\ &\quad + G_{j,k}^Fw_kw_k'G_{j,k}^{F\prime}]\mathbb{1}_{\{\mathcal{R}^-\}}\mathbb{1}_{\{\theta(k+1)=i\}}\mathbb{1}_{\{\theta(k)=j\}}\Big) \end{aligned} \tag{12}$$

where the last equality comes from the fact that $\Pr(\rho(k) = \ell_k, \theta(k) = j) = 0$ whenever j is not in the cluster S_{ℓ_k}, and $\Pr(\rho(k) = \ell_k, \theta(k) = j)) = \Pr(\theta(k) = j)$ otherwise. Resuming the above calculation, we have:

$$\begin{aligned} &X_{\ell_0,\ldots,\ell_k,i,k+1}(M) \\ &= \sum_{j\in S_{\ell_k}} \Pr(\mathcal{R},\theta(k) = j,\theta(k+1) = i) \\ &\quad \times\big[A_{j,k}^FE\big(\widetilde{x}_k\widetilde{x}_k'\big|\mathcal{R}^-,\theta(k) = j,\theta(k+1) = i\big)A_{j,k}^{F\prime} \\ &\quad + G_{j,k}^FE\big(w_kw_k'\big|\mathcal{R}^-,\theta(k) = j,\theta(k+1) = i\big)G_{j,k}^{F\prime}\big]. \end{aligned} \tag{13}$$

From basic properties of the Markov chain we have that $\rho(\ell)$ and $\theta(k+1)$ are conditionally independent given $\theta(k)$, for any $0 \le \ell \le k-1$. Moreover, from (1), (2) and (4) it can be shown that $\widetilde{x}_k$ and $\theta(k+1)$ are conditionally independent given $\theta(k)$, hence we may eliminate $\theta(k+1) = i$ from the first

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TAC.2017.2692180, IEEE Transactions on Automatic Control

conditional expectation in (13), yielding

$$\begin{aligned}
&X_{\ell_0,\ldots,\ell_k,i,k+1}(M)\\
&= \sum_{j\in S_{\ell_k}} \Pr(\theta(k+1)=i|\mathcal{R}^-,\theta(k)=j)\\
&\quad\times p_{\ell_0,\ldots,\ell_{k-1},j,k}\big[A^F_{j,k}E\big(\widetilde{x}_k\widetilde{x}_k'\big|\mathcal{R}^-,\theta(k)=j\big)\\
&\quad\times A^{F\prime}_{k,j}+G^F_{k,j}IG^{F\prime}_{k,j}\big]\\
&=\sum_{j\in\widetilde{S}} p_{ji}\big[A^F_{j,k}X_{\ell_0,\ldots,\ell_{k-1},j,k}(M)A^{F\prime}_{k,j}\\
&\qquad\qquad + p_{\ell_0,\ldots,\ell_{k-1},j,k}G^F_{k,j}G^{F\prime}_{k,j}\big],
\end{aligned} \tag{14}$$

where we denote $\widetilde{S}=\{j\in S_{\ell_k}: p_{\ell_0,\ldots,\ell_{k-1},j,k}\neq 0\}$. We now turn our attention to the optimality of M. Consider a feasible gain sequence in the form

$$\bar{M}=\{M_0^\star,\ldots,M_{k-1}^\star,M_k\},$$

where M_k is the variable to be determined; since $X_{\ell_0,\ldots,\ell_{k-1},j,k}(\bar{M})$ is a function of $M_0^\star,\ldots,M_{k-1}^\star$ only, we can use the induction hypothesis to write

$$X_{\ell_0,\ldots,\ell_{k-1},j,k}(\bar{M})=Y_{\ell_0,\ldots,\ell_{k-1},j,k}, \tag{15}$$

$$X_{\ell_0,\ldots,\ell_{k-1},j,k}(\bar{M})\leq X_{\ell_0,\ldots,\ell_k,j,k}(M). \tag{16}$$

Eq. (16) allows to write $A^F_{j,k}(X_{\ell_0,\ldots,\ell_{k-1},j,k}(\bar{M})-X_{\ell_0,\ldots,\ell_{k-1},j,k}(M))A^{F\prime}_{j,k}\leq 0$, irrespectively of M_k, and using (14) we evaluate

$$\begin{aligned}
&X_{\ell_0,\ldots,\ell_k,i,k+1}(\bar{M})-X_{\ell_0,\ldots,\ell_k,i,k+1}(M)\\
&=\sum_{j\in\widetilde{S}} p_{ji}\big[A^F_{j,k}(X_{\ell_0,\ldots,\ell_{k-1},j,k}(\bar{M})\\
&\qquad\qquad - X_{\ell_0,\ldots,\ell_{k-1},j,k}(M))A^{F\prime}_{j,k}\leq 0
\end{aligned} \tag{17}$$

Also, by plugging (15) into (14),

$$\begin{aligned}
&X_{\ell_0,\ldots,\ell_k,i,k+1}(\bar{M})\\
&\quad=\sum_{j\in\widetilde{S}} p_{ji}\big[A^F_{j,k}Y_{\ell_0,\ldots,\ell_{k-1},j,k}(\bar{M})A^{F\prime}_{j,k}\\
&\qquad + p_{\ell_0,\ldots,\ell_{k-1},j,k}G^F_{j,k}IG^{F\prime}_{j,k}\big],
\end{aligned} \tag{18}$$

and, by completing squares and denoting $\Phi=L_jY_{\ell_0,\ldots,\ell_{k-1},j,k}L_j'+\Pr(\rho(0)=\ell_0,\ldots,\rho(k-1)=\ell_{k-1},\theta(k)=j)H_jH_j'$ for brevity, we obtain

$$\begin{aligned}
&X_{\ell_0,\ldots,\ell_k,i,k+1}(\bar{M})=\sum_{j\in\widetilde{S}} p_{ji}\big[A_jY_{\ell_0,\ldots,\ell_{k-1},j,k}A_j'\\
&+p_{\ell_0,\ldots,\ell_{k-1},j,k}G_jG_j'\\
&+\big(M_k-A_jY_{\ell_0,\ldots,\ell_{k-1},j,k}L_j'\Phi^{-1}\big)\\
&\quad\times\Phi\big(M_k-A_jY_{\ell_0,\ldots,\ell_{k-1},i,k}L_j'\Phi^{-1}\big)'\\
&-A_jY_{\ell_0,\ldots,\ell_{k-1},j,k}L_j'\Phi^{-1}L_jY_{\ell_0,\ldots,\ell_{k-1},j,k}A_j'\big],
\end{aligned} \tag{19}$$

thus yielding that the minimal X is attained for

$$\begin{aligned}
M_k=M_k^\star&=g(\ell_0,\ldots,\ell_{k-1},j)\\
&=A_jY_{\ell_0,\ldots,\ell_{k-1},i,k}L_j'\Phi^{-1},
\end{aligned}$$

whenever $p_{\ell_0,\ldots,\ell_{k-1},j,k}\neq 0$, confirming the second equation in (8); the inverse always exists because we have assumed $H_iH_i'>0$. If j is such that $p_{\ell_0,\ldots,\ell_{k-1},j,k}=0$ then the gain M_k is immaterial for the error covariance, indeed we see from (18) that such gain is not accounted for, so that one can pick $M_k=0$, confirming the first equation in (8). Chosing the gain as above we get the gain sequence $M^\star=\{M_0^\star,\ldots,M_k^\star\}$ and

$$X_{\ell_0,\ldots,\ell_k,j,k+1}(M^\star)\leq X_{\ell_0,\ldots,\ell_k,j,k+1}(\bar{M}),$$

so that (16) produces

$$X_{\ell_0,\ldots,\ell_k,j,k+1}(M^\star)\leq X_{\ell_0,\ldots,\ell_k,j,k+1}(M),$$

which confirms (10) for the time instant $k+1$. Substituting $M_k=M_k^\star$ in (19) we get

$$\begin{aligned}
&X_{\ell_0,\ldots,\ell_k,i,k+1}(M^\star)=\sum_{j\in\widetilde{S}} p_{ji}\big[A_jY_{\ell_0,\ldots,\ell_{k-1},j,k}A_j'\\
&\quad+p_{\ell_0,\ldots,\ell_{k-1},j,k}G_jG_j'+A_jY_{\ell_0,\ldots,\ell_{k-1},j,k}L_j'\\
&\quad\times\Phi^{-1}L_jY_{\ell_0,\ldots,\ell_{k-1},j,k}A_j'\big]=Y_{\ell_0,\ldots,\ell_k,i,k+1},
\end{aligned}$$

which confirms (9) for $k+1$, thus completing the induction. It remains only to show the optimality of $M^\star$ in terms of (5). This follows directly from (10), in fact,

$$\begin{aligned}
E\{(\widetilde{x}_k^\star)'\widetilde{x}_k^\star\}&=\sum E((\widetilde{x}_k^\star)'\widetilde{x}_k^\star\mathbb{1}_{\{\mathcal{R}^-,\theta(k)=i\}})\\
&=\sum \operatorname{tr}(X_{\ell_0,\ldots,\ell_{k-1},i,k}(M^\star))\\
&\leq\sum \operatorname{tr}(X_{\ell_0,\ldots,\ell_{k-1},i,k}(M))\\
&=\sum E(\widetilde{x}_k'\widetilde{x}_k\mathbb{1}_{\{\mathcal{R}^-,\theta(k)=i\}})=E\{\widetilde{x}_k'\widetilde{x}_k\},
\end{aligned}$$

where all sums are in the indexes $0\leq i\leq N, 0\leq \ell_m\leq M, 0\leq m\leq k-1$ and we denote by $\widetilde{x}_k^\star$ the estimation error associated with the gain $M^\star$. ∎

IV. Properties of the CLMMSE

A. Number of matrices to be computed and stored

Memory availability of a controller is a limitation in many applications, making the number of gains involved in the implementation of the CLMMSE of much relevance. The number of gains is also determinant for the CPU time, a particularly important aspect when the gains are to be computed in an on-line fashion, as frequently found in the implementation of model predictive / receding horizon controllers. For each $0\leq k\leq s-1$, we compute NN_C^k matrices on the left hand side of (7), hence we have (up to) this number of recursive Riccati equations to solve. We also have the computation and storage of an equal number of gains. Then, to obtain the state estimate at time s, we have to store a total of $N(N_C^s-1)(N_C-1)^{-1}$

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TAC.2017.2692180, IEEE Transactions on Automatic Control

gains when $N_C \neq 1$, and sN gains otherwise. Regarding the number of matrix inverses, one may invert each Y given by (7) and store it at time step k for the forthcoming iterates, hence we have a total of (up to) $N(N_C^{(s-1)} - 1)(N_C - 1)^{-1}$ inverses.

B. Filtering in the entire interval $0 \leq k \leq s$

Note from (8) that, given a realization of the Markov chain $\theta(k)$, $k \geq 0$, the time instant s involved in the problem formulation (5) affects only the cardinality of the optimal gain sequence $M^\star$. More precisely, if $\{M_k^\star, 0 \leq k \leq s\}$ is the gain sequence attaining (5), and, if we replace s with $\ell \leq s$ in (5) and obtain the new optimal gain sequence $\{\widetilde{M}_k, 0 \leq k \leq \ell\}$ (considering the same Markov chain realization), then we have that $M_k^\star = \widetilde{M}_k$, $0 \leq k \leq \ell$. This is consistent with the sense of optimality in (10), and is in perfect harmony with the theory of both Kalman filter and the standard LMMSE. As a consequence, the provided clustered information LMMSE is also a solution for the multiobjective problem

$$\min_{M_0,\dots,M_s} \{E\{\|x_0 - \widehat{x}_0\|^2 | \mathcal{R}_0\}, \dots, E\{\|x_s - \widehat{x}_s\|^2 | \mathcal{R}_s\},$$

or for any linear combination of mean square errors writen in the form

$$\min_{M_0,\dots,M_s} \Big(\sum_{0\leq k\leq s} \alpha_k E\{\|x_k - \widehat{x}_k\|^2 | \mathcal{R}_k\} \Big).$$

C. Linking the Kalman filter and the LMMSE

It is simple to see that we retrieve the standard LMMSE when we consider only one partition $S_1 = \{1, \dots, N\}$. In this setup we have $p_{\ell_0,\dots,\ell_{k-1},i,k} = P(\theta(k) = i)$ and one can check by inspection that (7) and the LMMSE Riccati equation [5, Eq. 24] are identical. As for the Kalman filter, if we set $S_i = \{i\}$, $1 \leq i \leq N$, then

$$\begin{aligned} &p_{\ell_0,\dots,\ell_{k-1},i,k} \\ &\quad = \Pr(\theta(0) = \ell_0, \dots, \theta(k-1) = \ell_{k-1}, \theta(k) = i) \\ &\quad = \pi_{\ell_0}(0) p_{\ell_0,\ell_1} \cdots p_{\ell_{k-1},i}, \\ &\widetilde{S} = \{j \in S_{\ell_{k-1}} : p_{\ell_0,\dots,\ell_{k-2},j,k-1} \neq 0\} = \{\ell_{k-1}\}, \end{aligned}$$

if $p_{\ell_0,\dots,\ell_{k-2},\ell_{k-1},k-1} \neq 0$, and (7) is reduced to

$$\begin{aligned} Y_{\ell_0,\dots,\ell_{k-1},i,k} &= \big[A_j Y_{\ell_0,\dots,\ell_{k-2},j,k-1} A_j' \\ &\quad + p_{\ell_0,\dots,\ell_{k-2},j} G_j G_j' + A_j Y_{\ell_0,\dots,\ell_{k-2},j,k-1} L_j' \\ &\quad \times \big(L_j Y_{\ell_0,\dots,\ell_{k-2},j,k-1} L_j' + p_{\ell_0,\dots,\ell_{k-2},j,k-1} \\ &\quad \times H_j H_j'\big)^{-1} L_j Y_{\ell_0,\dots,\ell_{k-2},j,k-1} A_j'\big]. \end{aligned}$$

From Theorem 3.1 we have

$$\begin{aligned} Y_{\ell_0,\dots,\ell_{k-1},i,k} &= X_{\ell_0,\dots,\ell_{k-1},i,k}(M^\star) \\ &= E(\widetilde{x}_k^\star (\widetilde{x}_k^\star)' | \theta(0) = \ell_0, \dots, \theta(k-1) = \ell_{k-1}, \\ &\qquad \theta(k) = i) \pi_{\ell_0}(0)\, p_{\ell_0,\ell_1} \cdots p_{\ell_{k-1},i} \end{aligned}$$

so that, writing $Z_{\ell_0,\dots,\ell_{k-1},i} = E(\widetilde{x}_k^\star \widetilde{x}{}_k^{\star\prime} | \theta(0) = \ell_0, \dots, \theta(k-1) = \ell_{k-1}, \theta(k) = i)$, substituting in the above equation for Y and manipulating (cancelling the ps and πs) yields

$$\begin{aligned} Z_{\ell_0,\dots,\ell_{k-1},i,k} &= \big[A_j Z_{\ell_0,\dots,\ell_{k-2},j,k-1} A_j' + G_j G_j' \\ &\quad + A_j Z_{\ell_0,\dots,\ell_{k-2},j,k-1} L_j' \big(L_j Z_{\ell_0,\dots,\ell_{k-2},j,k-1} L_j' \\ &\quad + H_j H_j'\big)^{-1} L_j Z_{\ell_0,\dots,\ell_{k-2},j,k-1} A_j'\big], \end{aligned} \tag{20}$$

which is the usual Riccati difference equation appearing in Kalman filters. This means that $Y_{\ell_0,\dots,\ell_{k-1},i,k}$ is equal to the Kalman covariance matrix multiplied by the probability that the Markov chain visits $\ell_0, \dots, \ell_{k-1}, i$. In the gain formula (8), this probability is cancelled, yielding that the Kalman gain coincide with $M_k^\star$. Concluding, we have the Kalman filter and the markovian LMMSE in opposite "extremes" of the CLMMSE, and a lattice of estimators between them, depending on how the Markov states are arranged in clusters.

D. General LMMSE

Consider linear estimators of the general form

$$z_{k+1} = F_k z_k + \bar{G}_k y_k, \tag{21}$$

where matrices F_k and $\bar{B}_k$, $k \geq 0$, are the optimization variables replacing M_k in the problem (5); consider also that $F_k = f_k(\rho(0), \dots, \rho(k-1), \theta(k))$ and $\bar{G}_k = g_k(\rho(0), \dots, \rho(k-1), \theta(k))$, where f_k, g_k are measurable functions. It can be demonstrated that the optimal estimate satisfies

$$z_k = \widehat{x}_k^\star, \quad 0 \leq k \leq s, \text{ a.s.},$$

which is produced by setting $\bar{G}_k = M_k^\star$ and

$$F_k = A_{\theta(k)} - M_k^\star L_{\theta(k)}$$

where $M_k^\star$ is as in (8), thus retrieving the Luenberger observer form (2) and the solution given in Theorem 3.1. This is not surprising since the "innovation form" $F_k = A_{\theta(k)} - M_k L_{\theta(k)}$ for some M_k is necessary for some basic properties of an observer to be fulfilled, e.g. for $\widetilde{x}_k = z_k - x_k$ to remain zero a.s. for $k \geq 1$ in cases when $z_0 = x_0$ a.s. and there is no additive noise in the state ($G_i = 0$, $i = 1, 2, \dots, N$).

V. Illustrative example

We have applied the CLMMSE to the system given in [9]. The system data is reproduced next for ease of reference. The system is comprised of $N = 4$ Markov states. Denoting the elements of a matrix W by $[W]_{i,j}$, we have $[A_i]_{1,1} = 0$ and $[A_i]_{2,1} = 0.81$, $1 \leq i \leq 4$. Also, $[A_1]_{1,2} = -0.405$, $[A_1]_{2,2} = 0.81$, $[A_2]_{1,2} = -0.2673$,

$[A_2]_{2,2} = 1.134$, $[A_3]_{1,2} = -0.81$, $[A_3]_{2,2} = 0.972$, $[A_4]_{1,2} = -0.1863$, $[A_4]_{2,2} = 0.891$. $[G_i]_{1,1} = 0.5$ and $[G_i]_{j,k} = 0$, $1 \leq i \leq 4$, $1 \leq j \leq 2$, $1 \leq k \leq 2$. $[L_i]_{1,1} = 1$, $[L_i]_{1,2} = 0$ and $[H_i]_{1,1} = 1$, $1 \leq i \leq 4$. Also,

$$P = \begin{bmatrix} 0.3 & 0.2 & 0.1 & 0.4 \\ 0.3 & 0.2 & 0.3 & 0.2 \\ 0.1 & 0.1 & 0.5 & 0.3 \\ 0.2 & 0.2 & 0.1 & 0.5 \end{bmatrix}. \tag{22}$$

Every possible cluster configuration has been taken into account and we refer to configurations with clusters embedded in each other (that is, the clusters of one configuration form partitions of the other one) as *embedded configurations*. The aim was to estimate the state at time instant $s = 10$. The mean square error was calculated by means of (6) and (9), which lead to $E(\|\widetilde{x}_{10}\|^2) = \sum_{\ell_0,\ldots,\ell_9,i} \operatorname{trace}(Y_{\ell_0,\ldots,\ell_9,i,10})$; results have been confirmed by Monte Carlo simulation. Table I shows the obtained results in groups according to the number of clusters N_C. As expected, the standard LMMSE (with one cluster $\{1,2,3,4\}$) presented the largest estimation error and the Kalman filter (with four clusters $\{1\},\{2\},\{3\},\{4\}$) features the smallest one. The performance of other filters with "intermediary configurations" ($N_C = 2,3$) is similar to the LMMSE regarding the error, hence they are not much appealing in view of their higher complexity and storage requirements. Note that in this particular example the modes are similar to each other (only two elements of A change). Now, let us introduce a more relevant change in one

TABLE I
MEAN SQUARE ERROR, CPU TIME TO COMPUTE THE GAINS AND THE NUMBER OF GAINS FOR EVERY CLUSTER CONFIGURATION FOR THE MJLS WITH PARAMETERS GIVEN IN (22) AND ABOVE.

Clusters	$E(\|\widetilde{x}_{10}\|^2)$	CPU time	n. gains
{1,2,3,4}	0.6699	$2.24 \cdot 10^{-2}$	40
{1,2},{3,4}	0.6690	3.32	4,092
{1,3},{2,4}	0.6680	3.34	4,092
{1,4},{2,3}	0.6689	3.37	4,092
{1},{2,3,4}	0.6696	3.34	4,092
{2},{1,3,4}	0.6685	3.35	4,092
{3},{1,2,4}	0.6678	3.34	4,092
{4},{1,2,3}	0.6691	3.33	4,092
{1,2},{3},{4}	0.6675	$1.14 \cdot 10^3$	118,096
{1,3},{2},{4}	0.6675	$1.14 \cdot 10^3$	118,096
{1,4},{2},{3}	0.6672	$1.14 \cdot 10^3$	118,096
{1},{2,3},{4}	0.6687	$1.14 \cdot 10^3$	118,096
{1},{2,4},{3}	0.6674	$1.14 \cdot 10^3$	118,096
{1},{2},{3,4}	0.6682	$1.14 \cdot 10^3$	118,096
{1},{2},{3},{4}	0.6618	$1.35 \cdot 10^4$	1,398,100

mode by replacing A_4, G_4, L_4 and H_4 with $10A_4$, $10G_4$, $10L_4$ and $10H_4$, respectively. The results are displayed in Table II. We now can clearly distinguish two groups of filters, one with average errors around 10^6 and a second one around 10^4. Although the error is not necessarily reduced with more clusters (for instance we see the configuration $\{1,3\},\{2,4\}$ outperforming the configuration $\{1,4\},\{2\},\{3\}$), we have that it decreases for embedded configurations of clusters as illustrated in Figure 1. In this numerical example there is a tendency for better performance when $\theta = 4$ is isolated from other states, e.g. with $S_1 = \{4\}$ and $S_2 = \{1,2,3\}$ we have $E(\|\widetilde{x}_{10}\|^2) \approx 7,570$, while $S_1 = \{1,2\}$ and $S_2 = \{3,4\}$ lead to $E(\|\widetilde{x}_{10}\|^2) \approx 1.29 \cdot 10^6$. Moreover, considering that the Kalman filter yields $E(\|\widetilde{x}_{10}\|^2) \approx 5,150$ and is hard to implement (requiring storage of $N(N_C^s - 1)(N_C - 1)^{-1} = 1,398,100$ gains) we see that the filter with configurations $S_1 = \{4\}$, $S_2 = \{1,2,3\}$ and $S_1 = \{1,3\}$, $S_2 = \{2,4\}$ are quite competitive in this scenario ($4,092$ gains). In

TABLE II
MEAN SQUARE ERROR, CPU TIME TO COMPUTE THE GAINS AND THE NUMBER OF GAINS FOR EVERY CLUSTER CONFIGURATION FOR THE MJLS WITH THE MODIFIED DATA.

Clusters	$E(\|\widetilde{x}_{10}\|^2)$	CPU time	n. gains
{1,2,3,4}	$2.12 \cdot 10^6$	$1.86 \cdot 10^{-2}$	40
{1,2},{3,4}	$1.29 \cdot 10^6$	3.47	4,092
{1,3},{2,4}	$7.32 \cdot 10^3$	3.29	4,092
{1,4},{2,3}	$1.31 \cdot 10^4$	3.28	4,092
{1},{2,3,4}	$1.37 \cdot 10^6$	3.26	4,092
{2},{1,3,4}	$1.87 \cdot 10^6$	3.24	4,092
{3},{1,2,4}	$1.10 \cdot 10^4$	3.25	4,092
{4},{1,2,3}	$7.57 \cdot 10^3$	3.24	4,092
{1,2},{3},{4}	$6.42 \cdot 10^3$	$1.13 \cdot 10^3$	118,096
{1,3},{2},{4}	$6.14 \cdot 10^3$	$1.13 \cdot 10^3$	118,096
{1,4},{2},{3}	$9.99 \cdot 10^3$	$1.13 \cdot 10^3$	118,096
{1},{2,3},{4}	$6.64 \cdot 10^3$	$1.13 \cdot 10^3$	118,096
{1},{2,4},{3}	$6.23 \cdot 10^3$	$1.13 \cdot 10^3$	118,096
{1},{2},{3,4}	$1.17 \cdot 10^6$	$1.14 \cdot 10^3$	118,096
{1},{2},{3},{4}	$5.15 \cdot 10^3$	$1.34 \cdot 10^4$	1,398,100

this example, state 4 has a very different behavior from the other states. Therefore one can guess that the best cluster distributions will be those when state 4 is isolated. In other examples, one may have additional information from the nature of the system and its parameters to infer which cluster distributions may work best without necessarily testing all of them. Note also that although costly, the computations for every cluster can be made offline, so that one may choose only the best cluster distribution to filter the system in real-time.

VI. CONCLUDING REMARKS

We have explored the Markov state information structure in MJLS leading to new filters whose estimation error and complexity lie in between the

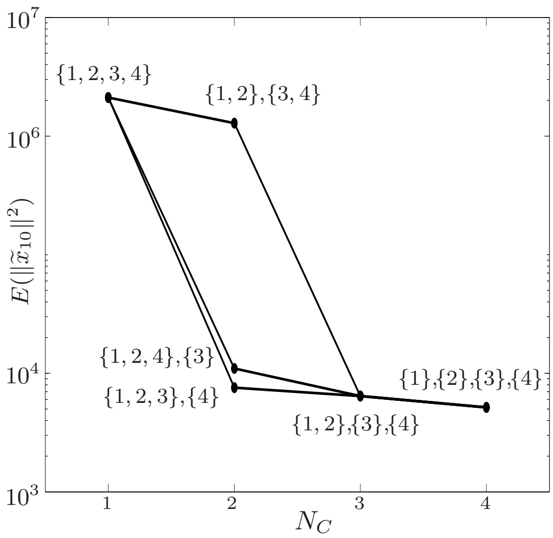

Fig. 1. Mean square error $E(\|\widetilde{x}_{10}\|^2)$ for some cluster configurations of the MJLS with the modified data. The lines are linking consecutively embedded configurations.

standard Kalman filter and the LMMSE, and establish a trade-off between performance and computational burden. This allows one to explore the computational resources at hands more deeply, seeking for the best possible estimates. As illustrated in the example, the new filters can provide competitive alternatives to the existing ones. We note that, for a given plant, the computational burden depends only on N_C, see Section IV-A. Then, a relevant question is how to arrange the Markov states in clusters so that the estimation error is minimal, which will be considered in future research.

References

[1] Anderson, B. D. O., and Moore, J. B. *Optimal Filtering*, first ed. Prentice-Hall, London, 1979.

[2] Costa, O. L., and Benites, G. R. Linear minimum mean square filter for discrete-time linear systems with Markov jumps and multiplicative noises. *Automatica 47*, 3 (2011), 466 – 476.

[3] Costa, O. L. V. Linear minimum mean square error estimation for discrete-time Markovian jump linear systems. *IEEE Transactions on Automatic Control 39*, 8 (1994), 1685–1689.

[4] Costa, O. L. V., Fragoso, M. D., and Marques, R. P. *Discrete-Time Markovian Jump Linear Systems*. Springer-Verlag, New York, 2005.

[5] Costa, O. L. V., and Tuesta, E. F. Finite horizon quadratic optimal control and a separation principle for Markovian jump linear systems. *IEEE Trans. Automat. Control 48* (2003).

[6] Fioravanti, A. R., Gonçalves, A. P., and Geromel, J. C. H2 filtering of discrete-time Markov jump linear systems through linear matrix inequalities. *International Journal of Control 81*, 8 (2008), 1221–1231.

[7] Gonçalves, A. P., Fioravanti, A. R., and Geromel, J. C. Markov jump linear systems and filtering through network transmitted measurements. *Signal Processing 90*, 10 (2010), 2842 – 2850.

[8] Miller, B. M., and Runggaldier, W. J. Kalman filtering for linear systems with coefficients driven by a hidden Markov jump process. *Systems & Control Letters 31* (1997), 93–102.

[9] Zhang, L., and Boukas, E.-K. Mode-dependent H_∞ filtering for discrete-time Markovian jump linear systems with partly unknown transition probabilities. *Automatica 45*, 6 (2009), 1462–1467.