Linear minimum mean square filters for Markov jump linear systems
Eduardo Costa, Benoîte De Saporta

To cite this version:
Eduardo Costa, Benoîte De Saporta. Linear minimum mean square filters for Markov jump linear systems. IEEE Transactions on Automatic Control, Institute of Electrical and Electronics Engineers, 2017, 62 (7), pp.3567-3572. <10.1109/TAC.2017.2692180>. <hal-01251334>

HAL Id: hal-01251334
https://hal.archives-ouvertes.fr/hal-01251334
Submitted on 6 Jan 2016

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L’archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d’enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.
Linear minimum mean square filters for Markov jump linear systems

Eduardo F. Costa and Benoîte de Saporta

January 6, 2016

Abstract

New linear minimum mean square estimators are introduced in this paper 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 hand (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, so that the cluster cardinality allows for a trade-off between performance and computational burden. In this paper we propose the estimator, give the formulas for pre-computation of gains, present some properties, and give an illustrative numerical example.

1 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, 7] 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, 5, 6]. Among them, the one that is closer to the KF in terms of structure and performance is the linear minimum mean square estimator (LMMSE) first introduced for MJLS in [3]. Indeed, as we shall see later, the LMMSE computation relies on coupled Riccati equations that are quite 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. The main dissimilarity with KF lies in the fact that, instead of having path dependent gains, the LMMSE has sets of \( N \) gains, \( N \) being the cardinality of the Markov state space (which we assume finite in this paper), that are precomputed based on the system matrices and the initial distribution of \( \theta \). During application, after observing \( \theta(k) \) one picks the corresponding gain from the precomputed sets of gains. In this way, one obtains the best estimate for the system state \( x_k \) among all estimators that are linear and Markovian, these being precisely the constraints we mentioned before.

However, there is currently no intermediary solution between the KF and the LMMSE in literature. In this paper we provide a “lattice” of filters bridging the KF to the LMMSE by relaxing the markovianity constraint and allowing clustered information of the Markov chain to be considered when designing the filter gains. By clustered information, we mean that we have a

\*Eduardo F. Costa is with Univ. São Paulo - Instituto de Ciências Matemáticas e de Computação, C.P. 668, 13560-970, São Carlos, SP, Brazil 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 work was supported by Inria associate team CDSS and ANR grant Piece, FAPESP and CNPq.
We assume that the observations of the filter we do not take into account its past values, that is, also assume that the jump variable \( k \) for all \( \theta(k) \) is independent from time instant \( k \). Therefore, we impose that the filter matrices at time instant \( k \) are in the form \( M_k \) and initial condition \( x_{\theta(0)} = \bar{x}_0 \). This produces an estimation error \( \tilde{x}_{k+1} = (A_{\theta(k)} - M_k L_{\theta(k)})\bar{x}_k + (G_{\theta(k)} - M_k H_{\theta(k)})w_k \), and \( \bar{x}_0 = x_0 - \bar{x} \sim N(0, \Psi) \).

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 being 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 \( \rho \) up to time \( k \) are available to calculate the filter gains. We also assume that the jump variable \( \theta \) at time \( k \) is available, however in the clustered information filter we do not take into account its past values, that is, \( \theta(0), \ldots, \theta(k-1) \) are not taken into account 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

\[
M_k = h_k(\rho(0), \ldots, \rho(k-1), \theta(k)), \quad k \geq 0
\]

1. General recursive linear estimators are briefly addressed in Section 4.

2 Problem formulation

Consider the MJLS

\[
x_{k+1} = A_{\theta(k)} x_k + G_{\theta(k)} w_k
\]

\[
y_k = L_{\theta(k)} x_k + H_{\theta(k)} w_k, \quad k \geq 0.
\]

with initial condition \( x_0 \sim N(\bar{x}, \Psi) \) where \( N(\bar{x}, \Psi) \) is the normal distribution with mean \( \bar{x} \) and covariance matrix \( \Psi \). 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 \( \theta \), \( E[w_k] = 0 \) and \( E[w_k w'_k] \) is the identity matrix for all \( k \). We assume that \( G_i H_i^T = 0 \) and \( H_i H_i^T > 0, 1 \leq i \leq N \). We shall consider Luenberger observers in the form

\[
\hat{x}_{k+1} = A_{\theta(k)} \hat{x}_k + M_k (y_k - L_{\theta(k)} \hat{x}_k),
\]

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

\[
\tilde{x}_{k+1} = (A_{\theta(k)} - M_k L_{\theta(k)})\tilde{x}_k + (G_{\theta(k)} - M_k H_{\theta(k)})w_k.
\]

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 being visited at time \( k \), that is,

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

$$
\min_{R_0, \ldots, R_s} E[\|x_s - \hat{x}_s\|^2 | R_s], \quad \text{s.t. } (3),
$$

where we write $R_s = \{\rho(0), \ldots, \rho(s), y(0), \ldots, y(s), \theta(s)\}$ for ease of notation. We refer to the filter satisfying (3) as the clustered information LMMSE, or CLMMSE for short.

### 3 Clustered information LMMSE computation

Consider the MJLS in (1) and the filter in (3) with an arbitrary sequence of feasible gains $M = \{M_k, k \geq 0\}$. For each $k \geq 0$, $0 \leq i \leq N$ and $0 \leq \ell_m \leq M$, $0 \leq m \leq k - 1$, we define

$$
X_{\ell_0, \ldots, \ell_{k-1}, i, k}(M) = E(\bar{x}_k \bar{x}_k^\prime | \rho(0) = \ell_0, \ldots, \rho(k-1) = \ell_{k-1}, \theta(k) = i).
$$

The variable $X$ plays an important role in the derivation of the formula for the optimal filter because the optimal gains $M^* = \{M^*_k, 0 \leq k \leq s\}$ are such that $X_{\ell_0, \ldots, \ell_{k-1}, i, k}(M^*) = Y_{\ell_0, \ldots, \ell_{k-1}, i, k}$ 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, e.g. it coincides with the filtering Riccati equation of the Kalman filter when $N = N_C$ (as many clusters as Markov states).

**Example 3.1** We illustrate the notation introduced in (6) on a simple example. Consider the MJLS (1) with $A_i = G_i = 1, 1 \leq i \leq 3, \Psi = 1, \bar{x} = 0$, the clusters $S_1 = \{1, 2\}, S_2 = \{3\}$, initial distribution $\pi_0 = [0.5, 0.3, 0.2]$ and probability matrix

$$
P = \begin{bmatrix}
0.5 & 0.4 & 0.1 \\
1 & 0 & 0 \\
0.5 & 0 & 0.5
\end{bmatrix}.
$$

Consider trivial gains $M_k = 0$, so that $\bar{x}_k = x_k$ and $E(\bar{x}_k \bar{x}_k^\prime) = E(x_k x_k^\prime) + 1$. Moreover, the estimation error and the Markov state are independent because all modes are identical, then $X_{\ell_0, \ldots, \ell_{k-1}, i, k}(M) = E(\bar{x}_k \bar{x}_k^\prime) \Pr(\rho(0) = \ell_0, \ldots, \rho(k-1) = \ell_{k-1}, \theta(k) = i)$. For instance, for $\ell_0 = 1, i = 2$ and $k = 1$ we have

$$
X_{1,2,1} = E(\bar{x}_1 \bar{x}_1) \Pr(\rho(0) = 1, \theta(1) = 2) = 2 \Pr(\theta(1) = 2 | \rho(0) = 1) \Pr(\rho(0) = 1)
= 2 \Pr(\rho(0) = 1) \left( \sum_{j \in S_{\ell_0}} \Pr(\theta(1) = 2 | \rho(0) = 1, \theta(0) = j) \Pr(\theta(0) = j | \rho(0) = 1) \right)
= 2 \times 0.8 \times (0.4 \times (0.5/0.8) + 0 \times (0.3/0.8)) = 0.4.
$$

Similarly, $X_{1,1,1} = 1.1, X_{1,3,1} = 0.1, X_{2,1,1} = 0.2, X_{2,2,1} = 0$ and $X_{2,3,1} = 0.2$. Note from (6) that summing $X$ in the indexes corresponding to $\ell$ and $i$ we obtain $E(\bar{x}_k \bar{x}_k^\prime)$, e.g. from the above we have $E(\bar{x}_1 \bar{x}_1^\prime) = 2$.

### Coupled Riccati of the CLMMSE

the optimal gain sequence can be (pre-)computed based on the following sets of matrices. Let $Y_{\ell,0} = \pi_i(0) \Psi$ for each $1 \leq i \leq N$. For each $k \geq 1$, let $0 \leq m \leq k - 1$ and compute for each $0 \leq i \leq N$ and $0 \leq \ell_m \leq M$,

$$
Y_{\ell_0, \ldots, \ell_{k-1}, i, k} = \begin{cases}
0, & \text{if } \ell_0, \ldots, \ell_{k-1}, i, k = 0, \\
\sum_{j \in S} p_{\ell,0, \ldots, \ell_{k-2}, j, k-1} [A_j Y_{\ell_0, \ldots, \ell_{k-2}, j, k-1} A_j^\prime + p_{\ell_0, \ldots, \ell_{k-2}, j, k-1} G_j G_j^\prime] + p_{\ell_0, \ldots, \ell_{k-2}, j, k-1} L_j Y_{\ell_0, \ldots, \ell_{k-2}, j, k-1} L_j^\prime, & \text{otherwise},
\end{cases}
$$

where

$$
p_{\ell_0, \ldots, \ell_{k-1}, i, k} = \begin{cases}
\rho_{\ell_0, \ldots, \ell_{k-1}, i, k}, & \text{if } \ell_0, \ldots, \ell_{k-1}, i, k = 0, \\
\sum_{j \in S} p_{\ell,0, \ldots, \ell_{k-2}, j, k-1} [A_j Y_{\ell_0, \ldots, \ell_{k-2}, j, k-1} A_j^\prime + p_{\ell_0, \ldots, \ell_{k-2}, j, k-1} G_j G_j^\prime] + p_{\ell_0, \ldots, \ell_{k-2}, j, k-1} L_j Y_{\ell_0, \ldots, \ell_{k-2}, j, k-1} L_j^\prime, & \text{otherwise},
\end{cases}
$$

and

$$
P_{\ell_0, \ldots, \ell_{k-1}, i, k} = \begin{cases}
\rho_{\ell_0, \ldots, \ell_{k-1}, i, k}, & \text{if } \ell_0, \ldots, \ell_{k-1}, i, k = 0, \\
\sum_{j \in S} p_{\ell,0, \ldots, \ell_{k-2}, j, k-1} [A_j Y_{\ell_0, \ldots, \ell_{k-2}, j, k-1} A_j^\prime + p_{\ell_0, \ldots, \ell_{k-2}, j, k-1} G_j G_j^\prime] + p_{\ell_0, \ldots, \ell_{k-2}, j, k-1} L_j Y_{\ell_0, \ldots, \ell_{k-2}, j, k-1} L_j^\prime, & \text{otherwise}.
\end{cases}
$$
where we denote
\[ p_{\ell_0, \ldots, \ell_{k-1}, i, k} = \Pr(\rho(0) = \ell_0, \ldots, \rho(k - 1) = \ell_{k-1}, \theta(k) = i), \]
\[ \hat{S} = \{ j \in S_{k-1} : p_{\ell_0, \ldots, \ell_{k-2}, j, k-1} \neq 0 \}. \]

**Theorem 3.2** Given the realization of \( \theta(k), k \leq s \), and the corresponding cluster observations \( \rho(0), \ldots, \rho(s - 1) \), the gains \( M^* = \{M^*_0, \ldots, M^*_s\} \) of the LMMSE can be computed for each \( k \leq s \) as
\[
M^*_k = \begin{cases} 
0, & \Pr(\rho(0), \ldots, \rho(k-1), \theta(k), k) = (9), \theta(k), k) \text{ is valid for some } k \leq s, \\
A_{\theta(k)}(\rho(0), \ldots, \rho(k-1), \theta(k), k) \mathcal{L}_{\theta(k)}(\rho(0), \ldots, \rho(k-1), \theta(k), k) \mathcal{L}_{\theta(k)}^{-1}, & \text{otherwise}, 
\end{cases}
\]
where \( Y \) is given in (7). Moreover, the conditional second moment of the estimation error is given, for each \( k \leq s \), by
\[
X(\rho(0), \ldots, \rho(k-1), \theta(k), k) = Y(\rho(0), \ldots, \rho(k-1), \theta(k), k),
\]
and it is optimal in the sense that, for any gain sequence \( M = \{M_0, \ldots, M_s\} \),
\[
X(\rho(0), \ldots, \rho(k-1), \theta(k), k) \leq X(\rho(0), \ldots, \rho(k-1), \theta(k), k) \text{, } 0 \leq k \leq s.
\]

**Proof** 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 \( k \), the initial estimate is given by \( \tilde{x}_0 = \bar{x} \), yielding \( \tilde{x}_0 = x_0 - \bar{x} \sim N(0, \Psi) \) (irrespective of the filter gains), hence
\[
X(\rho(0), \ldots, \rho(k-1), \theta(k), k) = E(\tilde{x}_0 \tilde{x}_0^\top) \Psi \Psi^{-1} = Y(\rho(0), \ldots, \rho(k-1), \theta(k), k).
\]
By the induction hypothesis we assume that (9) and (10) are valid for some \( 0 \leq k < s \). In order to complete the induction, we now consider the time instant \( k + 1 \). For a filter with an arbitrary sequence of casable 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
\[
X(\ell_0, \ldots, \ell_k, i, k+1) = E(\tilde{x}_{k+1} \tilde{x}_{k+1}^\top | \rho(0) = \ell_0, \ldots, \rho(k) = \ell_k, \theta(k+1) = i).
\]
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
\[
X(\ell_0, \ldots, \ell_k, i, k+1) = E(\tilde{x}_{k+1} \tilde{x}_{k+1}^\top | \rho(0) = \ell_0, \ldots, \rho(k) = \ell_k, \theta(k+1) = i) = E\left( \left( (A_{\theta(k)} - M_k \mathcal{L}_{\theta(k)}) \tilde{x}_k + (G_{\theta(k)} - M_k \mathcal{H}_{\theta(k)}) w_k \right) \right.
\]
\[
\times \left( (A_{\theta(k)} - M_k \mathcal{L}_{\theta(k)}) \tilde{x}_k + (G_{\theta(k)} - M_k \mathcal{H}_{\theta(k)}) w_k \right) \right)^\top \Psi \left( \rho(0) = \ell_0, \ldots, \rho(k) = \ell_k, \theta(k+1) = i \right)
\]
\[
= E\left( \left( (A_{\theta(k)} - M_k \mathcal{L}_{\theta(k)}) \tilde{x}_k, \tilde{x}_k^\top (A_{\theta(k)} - M_k \mathcal{L}_{\theta(k)}) \right) \right.
\]
\[
+ (G_{\theta(k)} - M_k \mathcal{H}_{\theta(k)}) w_k w_k^\top (G_{\theta(k)} - M_k \mathcal{H}_{\theta(k)}) \left( \rho(0) = \ell_0, \ldots, \rho(k) = \ell_k, \theta(k+1) = i \right)
\]
\[
= E\left( \left( (A_j - M_k \mathcal{L}_j) \tilde{x}_k, \tilde{x}_k^\top (A_j - M_k \mathcal{L}_j) \right) \right.
\]
\[
+ (G_j - M_k \mathcal{H}_j) w_k w_k^\top (G_j - M_k \mathcal{H}_j) \left( \rho(0) = \ell_0, \ldots, \rho(k) = \ell_k, \theta(k+1) = i, \theta(k) = j \right)
\]
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 \( S_{\ell_k} \), and \( \Pr(\rho(k) = \ell_k, \theta(k) = j) = \Pr(\theta(k) = j) \) otherwise. Resuming the above
Thus making clear that the minimal notation, we have:

\begin{equation}
X_{\ell_0,\ldots,\ell_k,i,k+1}(M) = \sum_{j \in S_\ell_k} \Pr(\mathcal{R})[(A_j - M_kL_j)E(\bar{x}_k\bar{x}_k^i)\mathcal{R})(A_j - M_kL_j]'\]
+ (G_j - M_kH_j)E(w_kw_k'^j)\mathcal{R})(G_j - M_kH_j)'\]
\end{equation}

From basic properties of the Markov chain we have that \(\rho(\ell)\) and \(\theta(\ell+1)\) are conditionally independent given \(\theta(\ell)\), for any \(0 \leq \ell \leq k - 1\). Moreover, from \([1], [3] \text{ and } [4]\) it can be shown that \(\bar{x}_k\) and \(\theta(\ell+1)\) are conditionally independent given \(\theta(\ell)\), hence we may eliminate \(\theta(\ell+1) = i\) from the first conditional expectation in \([13]\), leading to

\begin{equation}
X_{\ell_0,\ldots,\ell_k,i,k+1}(M) = \sum_{j \in S_\ell_k} \Pr(\theta(\ell+1) = i|\rho(0) = \ell_0,\ldots,\rho(k-1) = \ell_{k-1},\theta(k) = j)\rho_{\ell_0,\ldots,\ell_k-1,i,j,k}
\times [(A_j - M_kL_j)E(\bar{x}_k\bar{x}_k^i)\rho(0) = \ell_0,\ldots,\rho(\ell-1) = \ell_{\ell-1},\theta(\ell) = j)(A_j - M_kL_j)'
+ (G_j - M_kH_j)I(G_j - M_kH_j)']
= \sum_{j \in S} p_{ji}[(A_j - M_kL_j)X_{\ell_0,\ldots,\ell_k-1,j,k}(M)(A_j - M_kL_j)'
+ p_{\ell_0,\ldots,\ell_k-1,j,k}(G_j - M_kH_j)(G_j - M_kH_j)'] ,
\end{equation}

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

\(M = \{M_0^*,\ldots,M_{k-1}^*,M_k\}\),

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

\begin{equation}
X_{\ell_0,\ldots,\ell_k-1,j,k}(M) = Y_{\ell_0,\ldots,\ell_k-1,j,k},
\end{equation}

\begin{equation}
X_{\ell_0,\ldots,\ell_k-1,j,k}(M) \leq X_{\ell_0,\ldots,\ell_k-1,j,k}(M).
\end{equation}

Eq. \([16]\) allows to write \((A_j - M_kL_j)(X_{\ell_0,\ldots,\ell_k-1,j,k}(M) - X_{\ell_0,\ldots,\ell_k-1,j,k}(M))(A_j - M_kL_j)' \leq 0\), irrespectively of \(M_k\), and using \([14]\) we evaluate

\begin{equation}
X_{\ell_0,\ldots,\ell_k-1,i,k+1}(M) - X_{\ell_0,\ldots,\ell_k-1,i,k+1}(M) = \sum_{j \in S} p_{ji}[(A_j - M_kL_j)(X_{\ell_0,\ldots,\ell_k-1,j,k}(M) - X_{\ell_0,\ldots,\ell_k-1,j,k}(M))(A_j - M_kL_j)' \leq 0
\end{equation}

Also, by plugging \([15]\) into \([14]\),

\begin{equation}
X_{\ell_0,\ldots,\ell_k-1,i,k+1}(M) = \sum_{j \in S} p_{ji}[(A_j - M_kL_j)Y_{\ell_0,\ldots,\ell_k-1,j,k}(M)(A_j - M_kL_j)'
+ p_{\ell_0,\ldots,\ell_k-1,j,k}(G_j - M_kH_j)I(G_j - M_kH_j)'] ,
\end{equation}

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{equation}
X_{\ell_0,\ldots,\ell_k-1,i,k+1}(M) = \sum_{j \in S} p_{ji} [A_jY_{\ell_0,\ldots,\ell_k-1,j,k}A_j' + p_{\ell_0,\ldots,\ell_k-1,j,k}G_jG_j' + (M_k - A_jY_{\ell_0,\ldots,\ell_k-1,j,k}L_j'\Phi^{-1}\Phi(M_k - A_jY_{\ell_0,\ldots,\ell_k-1,i,k}L_j'\Phi^{-1})'
- 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'] ,
\end{equation}

thus making clear that the minimal \(X\) is attained by setting

\(M_k = M_k^* = g(\ell_0,\ldots,\ell_k-1,j) = A_jY_{\ell_0,\ldots,\ell_k-1,i,k}L_j'\Phi^{-1}
= A_jY_{\ell_0,\ldots,\ell_k-1,i,k}L_j' (L_jY_{\ell_0,\ldots,\ell_k-1,j,k}L_j + p_{\ell_0,\ldots,\ell_k-1,j,k}H_jH_j')^{-1} ,\)
whenever \( p_{\ell_0,\ldots,\ell_{k-1},k} \neq 0 \), confirming the second equation in (8); the inverse always exists because we have assumed \( H_iH_j^* > 0 \). 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 \( M^* = \{ M^*_0, \ldots, M^*_k \} \) and

\[
X_{t_0,\ldots,t_{k-1},k+1}(M^*) \leq X_{t_0,\ldots,t_{k-1},k+1}(M),
\]

so that (16) produces

\[
X_{t_0,\ldots,t_{k-1},k+1}(M^*) \leq X_{t_0,\ldots,t_{k-1},k+1}(M),
\]

which confirms (10) for the time instant \( k+1 \). Substituting \( M_k = M_k^* \) in (19) we get after some manipulations

\[
X_{t_0,\ldots,t_{k-1},i,k+1}(M^*) = \sum_{j \in S} p_{ji} \left[ A_j Y_{t_0,\ldots,t_{k-1},j,k} A_j^* + p_{t_0,\ldots,t_{k-1},j,k} G_j G_j^* \right. \\
+ A_j Y_{t_0,\ldots,t_{k-1},j,k} L_j^* H_j^* L_j Y_{t_0,\ldots,t_{k-1},j,k} A_j^* \left. \right] = Y_{t_0,\ldots,t_{k-1},i,k+1},
\]

which confirms (10) for \( k+1 \), thus completing the induction. It remains only to show the optimality of \( M^* \) in terms of (6). This follows directly from (10), in fact,

\[
E(\tilde{x}_k^2 | \tilde{x}_k) = \sum_{i=0} E(\tilde{x}_k^2 | \rho(0)=t_0,\ldots,\rho(k-1)=t_{k-1},\theta(k)=1) \\
= \sum_{i=0} \text{tr}(X_{t_0,\ldots,t_{k-1},i,k}(M^*)) \\
\leq \sum_{i=0} \text{tr}(X_{t_0,\ldots,t_{k-1},i,k}(M)) \\
= \sum_{i=0} E(\tilde{x}_k^2 | \rho(0)=t_0,\ldots,\rho(k-1)=t_{k-1},\theta(k)=1) = E(\tilde{x}_k^2 | \tilde{x}_k),
\]

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 \( \tilde{x}_k^2 \) the estimation error associated with the gain \( M^* \). \( \square \)

4 Properties of the CLMMSE

4.1 Number of matrices to be computed and stored

For each \( 0 \leq k \leq s-1 \), we compute \( NN^k \) 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 \( N(N^2_{-1})/(N_{-1}) \) gains when \( N_{-1} \neq 1 \), 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) \( N(M^*_{s-1})/(M_{s-1}) \) inverses.

4.2 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 \( M^* \). More precisely, if \( \{ M^*_k, 0 \leq k \leq s \} \) is the gain sequence attaining (5), and, if we replace \( s \) with \( \ell \leq s \) in (5), and obtain the new optimal gain sequence \( \{ \tilde{M}_k, 0 \leq k \leq \ell \} \) (considering the same Markov chain realization), then we have that \( M^*_k = \tilde{M}_k, 0 \leq k \leq \min(\ell,s) \). This is consistent with the sense of optimality in (10), and 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

\[
\min_{M_{t_0,\ldots,t_{k-1}},k} \{ E\{||x_0 - \hat{x}_0||^2|\mathcal{R}_0\}, \ldots, E\{||x_s - \hat{x}_s||^2|\mathcal{R}_s\} \},
\]
or for any linear combination of mean square errors written in the form

\[
\min_{M_0, \ldots, M_s} \left( \sum_{0 \leq k \leq s} \alpha_k E\{\|x_k - \hat{x}_k\|^2 | R_k\} \right).
\]

### 4.3 Linking the Kalman filter and the standard LMMSE

It is simple to see that we retrieve the standard LMMSE when we consider only one partition \(S_1 = \{1, \ldots, N\}\). In fact, in this setup we have \(p_{\ell_0, \ldots, \ell_{k-1}, i, k} = P(\theta(k) = i)\) and one can check by inspection that \(7\) and the LMMSE Riccati equation \([3, \text{Eq. XX}]\) are identical. As for the Kalman filter, if we set \(S_1 = \{i\}, 1 \leq i \leq N,\) then

\[
p_{\ell_0, \ldots, \ell_{k-1}, i, k} = \Pr(\theta(0) = \ell_0, \ldots, \rho(k-1) = \ell_{k-1}, \theta(k) = i)
\]

\[
= \Pr(\theta(0) = \ell_0, \ldots, \theta(k-1) = \ell_{k-1}, \theta(k) = i) = \pi_{\ell_0}(0)p_{\ell_0, \ell_1} \cdots p_{\ell_{k-1}, i},
\]

\(\tilde{S} = \{j \in S_{\ell_{k-1}} : p_{\ell_0, \ldots, \ell_{k-2}, j, k-1} \neq 0\} = \{\ell_{k-1}\},\)

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

\[
Y_{\ell_0, \ldots, \ell_{k-1}, i, k} = [A_j Y_{\ell_0, \ldots, \ell_{k-2}, j, k-1} A_j' + p_{\ell_0, \ldots, \ell_{k-2}, j} G_j G_j'] + A_j Y_{\ell_0, \ldots, \ell_{k-2}, j, k-1} L_j^* \{L_j Y_{\ell_0, \ldots, \ell_{k-2}, j, k-1} R_j^{-1} L_j Y_{\ell_0, \ldots, \ell_{k-2}, j, k-1} A_j'\]  

From Theorem 3.2 we have

\[
Y_{\ell_0, \ldots, \ell_{k-1}, i, k} = X_{\ell_0, \ldots, \ell_{k-1}, i}(\mathbf{M}^*)
\]

\[
eq E(\tilde{x}^*_k \tilde{x}_k | \theta(0) = \ell_0, \ldots, \theta(k-1) = \ell_{k-1}, \theta(k) = i) \pi_{\ell_0}(0)p_{\ell_0, \ell_1} \cdots p_{\ell_{k-1}, i}
\]

so that, writing \(Z_{\ell_0, \ldots, \ell_{k-1}, i} = E(\tilde{x}^*_k \tilde{x}_k | \theta(0) = \ell_0, \ldots, \theta(k-1) = \ell_{k-1}, \theta(k) = i)\), substituting in the above equation for \(Y\) and manipulating (cancelling the ps and \(\pi s)\) yields

\[
Z_{\ell_0, \ldots, \ell_{k-1}, i, k} = [A_j Z_{\ell_0, \ldots, \ell_{k-2}, j, k-1} A_j' + G_j G_j'] + A_j Z_{\ell_0, \ldots, \ell_{k-2}, j, k-1} L_j^* \{L_j Z_{\ell_0, \ldots, \ell_{k-2}, j, k-1} R_j L_j^* \}^{-1} L_j Z_{\ell_0, \ldots, \ell_{k-2}, j, k-1} A_j'.
\]

which is the usual Riccati difference equation appearing in Kalman filters. This means that \(Y_{\ell_0, \ldots, \ell_{k-1}, i, k}\) is equal to the Kalman covariance matrix multiplied by the probability that the Markov chain visits \(\ell_0, \ldots, \ell_{k-1}, i\). In the gain formula \([5]\), this probability is cancelled, yielding that the Kalman gain coincide with \(M_k^*\). 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.

### 4.4 General LMMSE

Consider linear estimators of the general form

\[
z_{k+1} = F_k z_k + G_k y_k,
\]

where matrices \(F_k\) and \(B_k, k \geq 0,\) are the optimization variables replacing \(M_k\) in the problem \([5]\), consider also that \(F_k = f_k(\rho(0), \ldots, \rho(k-1), \theta(k))\) and \(G_k = g_k(\rho(0), \ldots, \rho(k-1), \theta(k))\), where \(f_k, g_k\) are measurable functions. It can be demonstrated that the optimal estimate satisfies

\[
z_k = \tilde{x}_k, \quad 0 \leq k \leq s, \text{ a.s.,}
\]

which is produced by setting \(\bar{G}_k = M_k^*\) and

\[
F_k = A_{\theta(k)} - M_k^* L_{\theta(k)}
\]
where $M_\star^k$ is as in (8), thus retrieving the Luenberger observer form (3) and the solution given in Theorem 3.2. This is not surprising since the “innovation form” $F_k = A_{\theta(k)} - M_kL_{\theta(k)}$ for some $M_k$ is necessary for some basic properties of an observer to be fulfilled, e.g. for $\hat{x}_k = z_k - x_k$ 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 ($G_i = 0$, $i = 1, 2, \ldots, N$).

5 Illustrative example

We have applied the CLMMSE to the system given in [?]. The system data is reproduced below for ease of reference.

\[
A_1 = \begin{bmatrix} 0 & 0.405 \\ 0.810.81 \end{bmatrix}, \quad A_2 = \begin{bmatrix} 0 & 0.2673 \\ 0.811.134 \end{bmatrix}, \quad A_3 = \begin{bmatrix} 0 & 0.81 \\ 0.810.972 \end{bmatrix}, \quad A_4 = \begin{bmatrix} 0 & 0.1863 \\ 0.810.891 \end{bmatrix},
\]

\[
G_i = \begin{bmatrix} 0.5 & 0 \\ 0 & 0 \end{bmatrix}, \quad L_i = \begin{bmatrix} 1 & 0 \end{bmatrix}, \quad H_i = 1, i = 1, \ldots, 4,
\]

\[
P = \begin{bmatrix} 0.3 & 0.2 & 0.1 & 0.4 \\ 0.3 & 0.2 & 0.3 & 0.2 \\ 0.1 & 0.1 & 0.5 & 0.3 \\ 0.2 & 0.2 & 0.1 & 0.5 \end{bmatrix}.
\] (22)

Every possible cluster configuration has been taken into account, and 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 $E(||\hat{x}_{10}||^2) = \sum_{i=0}^{\infty} \text{trace}(Y_{\theta_{10}, i, 1})$; results have been confirmed by Monte Carlo simulation. Figure 1 shows the obtained results in groups according to the number of clusters $N_C$. As expected, the standard LMMSE (with $N_C = 1$) presented the largest estimation error and the Kalman filter ($N_C = 4$) 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. Note that in this particular example, the modes are similar to each other (only two parameters of $A$ change).

![Figure 1: Mean square error $E(||\hat{x}_{10}||^2)$ versus number of clusters $N_C$ for every cluster configuration for the system with data given in (22).](image)

Now let us introduce a more relevant change in one 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 Figure 2. We now can clearly distinguish two groups of filters, one with average errors around 106 and a second one around 104. In this setup, there
is a tendency for better performance when $\theta = 4$ is isolated from other states, e.g. with $S_1 = \{1, 2, 3\}$ and $S_2 = \{4\}$ we have $E(\|\tilde{x}_{10}\|^2) \simeq 7,566.69$, while $S_1 = \{1, 2\}$ and $S_2 = \{3, 4\}$ lead to $E(\|\tilde{x}_{10}\|^2) \simeq 1,286,306.18$. Moreover, considering that the (hard to implement) Kalman filter yields $E(\|\tilde{x}_{10}\|^2) \simeq 5,152.95$ we see that the filter with configuration $S_1 = \{1, 2, 3\}$ and $S_2 = \{4\}$ is quite competitive in this scenario.

![Figure 2: Mean square error $E(\|\tilde{x}_{10}\|^2)$ versus number of clusters $N_C$ for every cluster configuration for the system with data given in [23].](image)

6 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 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 the formulas given in Section 4.1. Then, a relevant question is how to select the Markov states to form each cluster so that the estimation error is minimized, which will be considered in future research.

References


