Suboptimal Kalman Filtering in Triplet Markov Models Using Model Order Reduction

Frederic Lehmann and Wojciech Pieczynski

Abstract—**When the state space dimension increases, the computational burden can become a major challenge for optimal Kalman filtering in Gaussian triplet Markov models (TMMs). In this paper, we introduce a new model order reduction technique applicable to linear time-homogeneous Gaussian TMMs. Taking advantage of the lower state dimension of the resulting approximate model, a low-complexity suboptimal Kalman filter is obtained. The proposed estimator provides complexity reduction without significant accuracy loss and is shown to outperform two classical methods in the case of Markovian process noise.**

Index Terms—**Triplet Markov models, pairwise Markov models, Kalman filter, model order reduction, colored process noise.**

I. Introduction

Discrete-time Gaussian linear dynamical systems (GLDS) are ubiquitous in signal processing, due to their wide range of applications including tracking, control, econometrics, system identification, speech processing and machine learning, to name a few. A GLDS is often modeled as a hidden Markov model (HMM) via a double process $(\mathbf{X}, \mathbf{Y})$, where $\mathbf{X}$ is the Markovian hidden process, while $\mathbf{Y}$ is the observed one. Its popularity lies in the existence of the optimal filter in the minimum mean square error (MMSE) sense, the Kalman filter (HMM-KF) [1]. HMMs have been extended to pairwise Markov models (PMMs) [2], in which $(\mathbf{X}, \mathbf{Y})$ is Markovian, while $\mathbf{X}$ may not be (see [7] for the formulation used in this letter). Moreover, triplet Markov models (TMMs) are obtained by adding an auxiliary process $\mathbf{R}$, and by considering that $(\mathbf{X}, \mathbf{R}, \mathbf{Y})$ is Markovian [8]. $\mathbf{R}$, may have some physical meaning or not. For example in the first case, it can account for non-stationarity, parameter uncertainty [10] and error sources [9]. Interestingly, optimal filters (in the MMSE sense) have been derived for linear Gaussian PMMs and TMMs, that we shall refer to as the PMM Kalman filter (PMM-KF) [2]–[7] and the TMM Kalman filter (TMM-KF) [8]–[9], respectively. Note that the PMM-KF has recently found successful applications in tracking [7] and econometrics [11]–[12]. Also, the TMM-KF has been found suitable [9] for applications in tracking [13, ch. 5], [14], speech processing with colored noise [15] and for time series analysis using conditionally Gaussian models [16, Sec. 3.7.1].

In this letter, we deal with the order reduction problem. Assuming we have a mathematical description of the full order model for a GLDS considered as the "truth" model, we look for an approximate model with lower dimension for hidden variables. Reduced order models (ROMs) are of importance for several reasons: (i) simplify the understanding of the system [17, ch. 1]; (ii) reduce the sensitivity to uncertain parameters (e.g. spectral density of system noises) [18, ch. 7]; (iii) reduce the computational burden of state estimation via suboptimal Kalman filtering [19].

Regarding HMMs, ROM-based methods have been proposed in the literature, whose design criteria attempt to strike a balance between modeling accuracy and simplicity. A crude method consists in identifying and truncating a "less important" part of the state vector (possibly after coordinate transformation) and considering it as an additive noise term [20, pp. 301–302]-[21]. Also, partitioning the state vector enables estimation using separate low order filters [22]. In [19], a ROM is obtained by minimizing the mean weighted square error wrt the original full order model.

In this letter, we deal with model order reduction for a general TMM, that has not been adressed previously. Thus for $(\mathbf{X}, \mathbf{R}, \mathbf{Y})$ Markovian such that $(\mathbf{X}, \mathbf{Y})$ is not Markovian, the problem is to find a Markovian $(\mathbf{X}^*, \mathbf{Y}^*)$ such that PMM-KF applied to $(\mathbf{X}^*, \mathbf{Y}^*)$ give results as close as possible to the results obtained with TMM-KF applied to $(\mathbf{X}, \mathbf{R}, \mathbf{Y})$. Numerical experiments with application to tracking show the superiority of the proposed method over two classical approaches. Throughout the letter, $\mathcal{N}(\mathbf{m}, \mathbf{C})$ denotes a Gaussian distribution with mean $\mathbf{m}$ and covariance matrix $\mathbf{C}$. A set of observations from time m up to time n is denoted by $\mathbf{y}_{m:n}$.

II. Problem Formulation

A. System Model

We assume that a general time-homogeneous GLDS can be represented accurately by a full order (truth) model taken in the class of TMMs. Let $\mathbf{x}_n \in \mathbb{R}^K$, $\mathbf{r}_n \in \mathbb{R}^L$ and $\mathbf{y}_n \in \mathbb{R}^M$ denote the state, auxiliary and observation vector at instant n, respectively. Considering the random processes $\mathbf{X} = \{\mathbf{x}_n\}_{0 \leq n \leq N}$, $\mathbf{R} = \{\mathbf{R}_n\}_{0 \leq n \leq N}$ and $\mathbf{Y} = \{\mathbf{y}_n\}_{0 \leq n \leq N}$, the class of discrete-time TMMs is defined by the Markovianity of $(\mathbf{X}, \mathbf{R}, \mathbf{Y})$ [8]. It follows that a time-homogeneous GLDS modeled by a TMM is described using the discrete-time stochastic system

$$\underbrace{\begin{bmatrix} \mathbf{x}_n \\ \mathbf{r}_n \\ \mathbf{y}_n \end{bmatrix}}_{\mathbf{t}_n} = \underbrace{\begin{bmatrix} \mathbf{A}^{(\mathrm{t}11)} & \mathbf{A}^{(\mathrm{t}12)} & \mathbf{A}^{(\mathrm{t}13)} \\ \mathbf{A}^{(\mathrm{t}21)} & \mathbf{A}^{(\mathrm{t}22)} & \mathbf{A}^{(\mathrm{t}23)} \\ \mathbf{A}^{(\mathrm{t}31)} & \mathbf{A}^{(\mathrm{t}32)} & \mathbf{A}^{(\mathrm{t}33)} \end{bmatrix}}_{\mathbf{A}^{(\mathrm{t})}} \underbrace{\begin{bmatrix} \mathbf{x}_{n-1} \\ \mathbf{r}_{n-1} \\ \mathbf{y}_{n-1} \end{bmatrix}}_{\mathbf{t}_{n-1}} + \underbrace{\begin{bmatrix} \mathbf{B}^{(\mathrm{t}1)} \\ \mathbf{B}^{(\mathrm{t}2)} \\ \mathbf{B}^{(\mathrm{t}3)} \end{bmatrix}}_{\mathbf{B}^{(\mathrm{t})}} \mathbf{w}_n \tag{1}$$

Manuscript received March 31, 2020; revised June 10, 2020; accepted June 10, 2020. Date of publication June 15, 2020; date of current version July 14, 2020. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Wei Li. *(Corresponding author: Frederic Lehmann.)*

The authors are with the SAMOVAR, Télécom SudParis, Institut Polytechnique de Paris, 91011 Evry Cedex, France (e-mail: frederic.lehmann@telecom-sudparis.eu; wojciech.pieczynski@telecom-sudparis.eu).

Digital Object Identifier 10.1109/LSP.2020.3002420

where the initial triplet $\mathbf{t}_0 \sim \mathcal{N}(\mathbf{m}_0^{(\mathrm{t})}, \mathbf{P}_0^{(\mathrm{tt})})$ is independent from the zero-mean white Gaussian noise process $\mathbf{w}_n, \forall n \geq 1$. The noise covariance is defined by $\mathbf{Q}^{(\mathrm{t})} = E\{\mathbf{w}_n \mathbf{w}_n^T\}, \forall n \geq 1$. An optimal filter in the MMSE sense, that we refer to as the TMM-KF [9], exists for the model in (1). Note that $\forall n \geq 1$, the marginal distribution of $\mathbf{t}_n$ is $\mathcal{N}(\mathbf{m}_n^{(\mathrm{t})}, \mathbf{P}_n^{(\mathrm{tt})})$, where

$$\begin{cases} \mathbf{m}_n^{(\mathrm{t})} &= \mathbf{A}^{(\mathrm{t})}\mathbf{m}_{n-1}^{(\mathrm{t})} \\ \mathbf{P}_n^{(\mathrm{tt})} &= \mathbf{A}^{(\mathrm{t})}\mathbf{P}_{n-1}^{(\mathrm{tt})}\mathbf{A}^{(\mathrm{t})T} + \mathbf{B}^{(\mathrm{t})}\mathbf{Q}^{(\mathrm{t})}\mathbf{B}^{(\mathrm{t})T}, \end{cases} \quad (2)$$

since $\{\mathbf{w}_n\}_{1\leq n\leq N}$ is white and zero-mean.

Turning our attention to the process $\mathbf{Z} = \{\mathbf{z}_n\}_{0\leq n\leq N}$, where $\mathbf{z}_n = [\mathbf{x}_n^T, \mathbf{y}_n^T]^T$, it follows from (1) that $\mathbf{Z}$ is in general not Markovian and thus cannot be modeled by a PMM. The crux of the proposed model in Section III-A is that $p(\mathbf{z}_{n-1}, \mathbf{z}_n)$ is a marginal distribution of $p(\mathbf{t}_{n-1}, \mathbf{t}_n)$, $\forall n \geq 1$. Defining the selection matrix

$$\mathbf{C} = \begin{bmatrix} \mathbf{I}_K & \mathbf{0}_{K\times L} & \mathbf{0}_{K\times M} \\ \mathbf{0}_{M\times K} & \mathbf{0}_{M\times L} & \mathbf{I}_M \end{bmatrix},$$

then $\mathbf{z}_n = \mathbf{C}\mathbf{t}_n$ and using [23, p. 52] for $n = 1, \ldots, N$

$$p(\mathbf{z}_{n-1}, \mathbf{z}_n) = \mathcal{N}\left(\begin{bmatrix} \mathbf{m}_{n-1}^{(\mathrm{z})} \\ \mathbf{m}_n^{(\mathrm{z})} \end{bmatrix}, \begin{bmatrix} \mathbf{P}_{n-1}^{(\mathrm{zz})} & \mathbf{P}_{n-1,n}^{(\mathrm{zz})} \\ \mathbf{P}_{n-1,n}^{(\mathrm{zz})\,T} & \mathbf{P}_n^{(\mathrm{zz})} \end{bmatrix} \right), \quad (3)$$

where

$$\begin{cases} \mathbf{m}_n^{(\mathrm{z})} = E\{\mathbf{z}_n\} = \mathbf{C}\mathbf{m}_n^{(\mathrm{t})} \\ \mathbf{P}_n^{(\mathrm{zz})} = E\{(\mathbf{z}_n - \mathbf{m}_n)(\mathbf{z}_n - \mathbf{m}_n)^T\} = \mathbf{C}\mathbf{P}_n^{(\mathrm{tt})}\mathbf{C}^T \\ \mathbf{P}_{n-1,n}^{(\mathrm{zz})} = E\{(\mathbf{z}_{n-1} - \mathbf{m}_{n-1})(\mathbf{z}_n - \mathbf{m}_n)^T\} \\ \qquad = \mathbf{C}\mathbf{P}_{n-1}^{(\mathrm{tt})}\mathbf{A}^{(\mathrm{t})T}\mathbf{C}^T. \end{cases} \quad (4)$$

B. Truncation-Based Reduced Order Model

The essence of the TMM-KF is to jointly estimate $[\mathbf{x}_n^T, \mathbf{r}_n^T]^T$, though with a possibly prohibitive complexity order $\mathcal{O}(M^3 + 3(K+L)^2M + 2(K+L)M^2 + 2(K+L)^3)$ per time instant. Since in most applications, only the restoration of $\mathbf{X}$ is of interest, a natural idea to obtain a ROM for (1) is to build a PMM [2], under the approximate assumption that the couple $(\mathbf{X}, \mathbf{Y})$ is Markovian. A simple way to obtain such an approximate model consists in isolating the less important part of the state in the full order model (in our case the auxiliary process) as an additional noise source [20, pp. 301–302]–[21]. Let us extract $\mathbf{z}_n$ from (1) as

$$\mathbf{z}_n = \mathbf{C}\mathbf{A}^{(\mathrm{t})}\mathbf{C}^T\mathbf{z}_{n-1} + \mathbf{C}\mathbf{B}^{(\mathrm{t})}\mathbf{w}_n + \mathbf{D}^{(\mathrm{t})}\mathbf{t}_{n-1} \quad (5)$$

where

$$\mathbf{D}^{(\mathrm{t})} = \begin{bmatrix} \mathbf{0}_{K\times K} & \mathbf{A}^{(\mathrm{t}12)} & \mathbf{0}_{K\times M} \\ \mathbf{0}_{M\times K} & \mathbf{A}^{(\mathrm{t}32)} & \mathbf{0}_{M\times M} \end{bmatrix}.$$

Thus the truncation-based ROM considers the last term in (5) as a non-zero mean noise, whose coloration is ignored. Consequently, the process $\tilde{\mathbf{Z}} = \{\tilde{\mathbf{z}}_n\}_{0\leq n\leq N}$, defined as

$$\begin{cases} \tilde{\mathbf{z}}_0 = \mathbf{z}_0 \\ \tilde{\mathbf{z}}_n = \tilde{\mathbf{A}}\tilde{\mathbf{z}}_{n-1} + \tilde{\mathbf{b}}_n + \tilde{\mathbf{w}}_n, \quad \text{for } n = 1, \ldots, N \end{cases} \quad (6)$$

where $\{\tilde{\mathbf{w}}_n\}_{1\leq n\leq N}$ is white zero-mean Gaussian and

$$\begin{cases} \tilde{\mathbf{A}} = \mathbf{C}\mathbf{A}^{(\mathrm{t})}\mathbf{C}^T \\ \tilde{\mathbf{b}}_n = \mathbf{D}^{(\mathrm{t})}\mathbf{m}_{n-1}^{(\mathrm{t})} \\ E\{\tilde{\mathbf{w}}_n\tilde{\mathbf{w}}_n{}^T\} = \tilde{\mathbf{\Sigma}}_{n|n-1} = \mathbf{D}^{(\mathrm{t})}\mathbf{P}_{n-1}^{(\mathrm{tt})}\mathbf{D}^{(\mathrm{t})T} \\ \qquad + \mathbf{C}\mathbf{B}^{(\mathrm{t})}\mathbf{Q}^{(\mathrm{t})}\mathbf{B}^{(\mathrm{t})T}\mathbf{C}^T, \end{cases} \quad (7)$$

is the truncation-based ROM approximation of $\mathbf{Z}$.

III. Proposed Suboptimal Kalman Filter for TMMs

A. Proposed Reduced Order Model

The similarity between the random process of interest $\mathbf{Z}$ and any approximation $\mathbf{Z}^* = \{\mathbf{z}_n^*\}_{0\leq n\leq N}$ thereof can be assessed using well-known similarity measures from the literature, such as the Kullback-Leibler divergence between the marginal distributions [24] and the Kullback-Leibler divergence between the transition kernels [14]. Here, we seek a Markovian ROM $\mathbf{Z}^*$ for which both the marginal distributions and the transition kernels are equal to those of $\mathbf{Z}$. Thus any similarity measure comparing marginals and transition kernels (between $\mathbf{Z}^*$ and $\mathbf{Z}$) being null, is optimized.

Definition 3.1: The process $\mathbf{Z}^* = \{\mathbf{z}_n^*\}_{0\leq n\leq N}$ with probability distribution q where $\mathbf{z}_n^* = [\mathbf{x}_n^{*\,T}, \mathbf{y}_n^{*\,T}]^T$, is a time-pairwise Markov approximation of $\mathbf{Z}$ over a length-N window if

i) $\mathbf{Z}^*$ is Markovian;

ii)

$$q(\mathbf{z}_{n-1}^*, \mathbf{z}_n^*) = \mathcal{N}\left(\begin{bmatrix} \mathbf{m}_{n-1}^{(\mathrm{z})} \\ \mathbf{m}_n^{(\mathrm{z})} \end{bmatrix}, \begin{bmatrix} \mathbf{P}_{n-1}^{(\mathrm{zz})} & \mathbf{P}_{n-1,n}^{(\mathrm{zz})} \\ \mathbf{P}_{n-1,n}^{(\mathrm{zz})\,T} & \mathbf{P}_n^{(\mathrm{zz})} \end{bmatrix} \right),$$

for n = 1 ,..., N,

where (ii) means that the joint distribution $q(.,.)$ of $[\mathbf{z}_{n-1}^{*\,T}, \mathbf{z}_n^{*\,T}]^T$ is identical to (3), for $n = 1, \ldots, N$.

Using classical Gaussian conditioning $\mathbf{Z}^*$ can be chosen as

$$\begin{cases} \mathbf{z}_0^* = \mathbf{z}_0 \\ \mathbf{z}_n^* = \mathbf{A}_n^*\mathbf{z}_{n-1}^* + \mathbf{b}_n^* + \mathbf{w}_n^*, \quad \text{for } n = 1, \ldots, N \end{cases} \quad (8)$$

where $\{\mathbf{w}_n^*\}_{1\leq n\leq N}$ is zero-mean white Gaussian and

$$\begin{cases} \mathbf{A}_n^* = \mathbf{P}_{n-1,n}^{(\mathrm{zz})\,T}\mathbf{P}_{n-1}^{(\mathrm{zz})\,-1} \\ \mathbf{b}_n^* = \mathbf{m}_n^{(\mathrm{z})} - \mathbf{A}_n^*\mathbf{m}_{n-1}^{(\mathrm{z})} \\ E\{\mathbf{w}_n^*\mathbf{w}_n^{*\,T}\} = \mathbf{\Sigma}_{n|n-1}^* = \mathbf{P}_n^{(\mathrm{zz})} - \mathbf{A}_n^*\mathbf{P}_{n-1,n}^{(\mathrm{zz})}. \end{cases} \quad (9)$$

Injecting (4) into (9), the parameters of the proposed Markov approximation model can be rewritten as a function of the parameters of the original model in (1)

$$\begin{cases} \mathbf{A}_n^* = \mathbf{C}\mathbf{A}^{(\mathrm{t})}\mathbf{P}_{n-1}^{(\mathrm{tt})}\mathbf{C}^T(\mathbf{C}\mathbf{P}_{n-1}^{(\mathrm{tt})}\mathbf{C}^T)^{-1} \\ \mathbf{b}_n^* = (\mathbf{C}\mathbf{A}^{(\mathrm{t})} - \mathbf{A}_n^*\mathbf{C})\mathbf{m}_{n-1}^{(\mathrm{t})} \\ \mathbf{\Sigma}_{n|n-1}^* = (\mathbf{C}\mathbf{A}^{(\mathrm{t})} - \mathbf{A}_n^*\mathbf{C})\mathbf{P}_{n-1}^{(\mathrm{tt})}(\mathbf{C}\mathbf{A}^{(\mathrm{t})} - \mathbf{A}_n^*\mathbf{C})^T \\ \qquad + \mathbf{C}\mathbf{B}^{(\mathrm{t})}\mathbf{Q}^{(\mathrm{t})}\mathbf{B}^{(\mathrm{t})T}\mathbf{C}^T. \end{cases} \quad (10)$$

We note that the parameters of the transition kernels in (7) and (10) are similar.

We assume $\mathbf{P}_0^{(\mathrm{tt})}$ to have the same value over any length-N window, so that $\{\mathbf{P}_n^{(\mathrm{tt})}\}_{1\leq n\leq N}$ can be computed offline. Note that the original time-homogeneous full order TMM in (1) is approximated as a non-homogeneous reduced order PMM in (8)–(10). Thus model approximation, as summarized in Algorithm 1, is in order over any length-N window.

Algorithm 1: Model Approximation (Length-N Window).

Require: $\mathbf{A}^{(\mathrm{t})}$, $\mathbf{B}^{(\mathrm{t})}$, $\mathbf{Q}^{(\mathrm{t})}$, $\mathbf{m}_0^{(\mathrm{t})}$, $\{\mathbf{P}_n^{(\mathrm{tt})}\}_{0\le n\le N}$
$\mathbf{P}_0^{(\mathrm{zz})} = \mathbf{C}\mathbf{P}_0^{(\mathrm{tt})}\mathbf{C}^T$
for $n = 1, ..., N$ **do**
$\mathbf{m}_n^{(\mathrm{t})} = \mathbf{A}^{(\mathrm{t})}\mathbf{m}_{n-1}^{(\mathrm{t})}$
$\mathbf{m}_n^{(\mathrm{z})} = \mathbf{C}\mathbf{m}_n^{(\mathrm{t})}$
$\mathbf{P}_n^{(\mathrm{zz})} = \mathbf{C}\mathbf{P}_n^{(\mathrm{tt})}\mathbf{C}^T$
$\mathbf{P}_{n-1,n}^{(\mathrm{zz})} = \mathbf{C}\mathbf{P}_{n-1}^{(\mathrm{tt})}\mathbf{A}^{(\mathrm{t})T}\mathbf{C}^T$
$\mathbf{A}_n^* = \mathbf{P}_{n-1,n}^{(\mathrm{zz})\,T}\mathbf{P}_{n-1}^{(\mathrm{zz})\,-1}$
$\mathbf{b}_n^* = \mathbf{m}_n^{(\mathrm{z})} - \mathbf{A}_n^*\mathbf{m}_{n-1}^{(\mathrm{z})}$
$\mathbf{\Sigma}_{n|n-1}^* = \mathbf{P}_n^{(\mathrm{zz})} - \mathbf{A}_n^*\mathbf{P}_{n-1,n}^{(\mathrm{zz})}$
end for
return $\{\mathbf{A}_n^*, \mathbf{b}_n^*, \mathbf{\Sigma}_{n|n-1}^*\}_{1\le n\le N}$

Algorithm 2: PMM-KF Algorithm (Length-N Window).

Require: $\{\mathbf{y}_n\}_{1\le n\le N}$, $\hat{\mathbf{x}}_0$, $\mathbf{P}_0$, $\{\mathbf{A}_n^*, \mathbf{b}_n^*, \mathbf{\Sigma}_{n|n-1}^*\}_{1\le n\le N}$
for $n = 1, \ldots, N$ **do**
$\begin{bmatrix}\hat{\mathbf{x}}_n^-\\ \hat{\mathbf{y}}_n^-\end{bmatrix} = \mathbf{A}_n^*\begin{bmatrix}\hat{\mathbf{x}}_{n-1}\\ \mathbf{y}_{n-1}\end{bmatrix} + \mathbf{b}_n^*,$
$\begin{bmatrix}\mathbf{P}_n^- & \mathbf{\Sigma}_n^-\\ \mathbf{\Sigma}_n^{-T} & \mathbf{L}_n^-\end{bmatrix} = \mathbf{A}_n^*\begin{bmatrix}\mathbf{P}_{n-1} & \mathbf{0}\\ \mathbf{0} & \mathbf{0}\end{bmatrix}\mathbf{A}_n^{*T} + \mathbf{\Sigma}_{n|n-1}^*$
$\mathbf{K}_n = \mathbf{\Sigma}_n^-(\mathbf{L}_n^-)^{-1}$
$\hat{\mathbf{x}}_n = \hat{\mathbf{x}}_n^- + \mathbf{K}_n(\mathbf{y}_n - \hat{\mathbf{y}}_n^-)$
$\mathbf{P}_n = \mathbf{P}_n^- - \mathbf{K}_n\mathbf{\Sigma}_n^{-T}$
end for
return $\{\hat{\mathbf{x}}_n\}_{1\le n\le N}$

Again, assuming $\mathbf{A}^{(\mathrm{t})}$ to be time-invariant and $\mathbf{P}_0^{(\mathrm{tt})}$ to have the same value over any length-N window, all operations in Algorithm 1, except the computation of $\mathbf{m}_n^{(\mathrm{t})}$, $\mathbf{m}_n^{(\mathrm{z})}$ and $\mathbf{b}_n^*$, can be precomputed offline.

B. Proposed Approximate Bayesian Estimator

Since the TMM-KF over the full order model may have prohibitive complexity, we now introduce a reduced order state estimation procedure. We propose to apply the PMM-KF (summarized by Algorithm 2) without modification to the new ROM introduced in Section III-A. This strategy is suboptimal, since the hidden states and observations follow the truth model (1) instead of the proposed ROM. Therefore the filtering error taking this suboptimality into account, will be analyzed in Section III-C. Note that $\hat{\mathbf{x}}_n$ and $\mathbf{P}_0$ denote the state estimate at instant n and the initial state error covariance matrix, respectively. The interested reader is refered to [2], [7], [25] for the demonstration.

Remark 3.2: Note that the same state estimation procedure is applicable to the truncation-based ROM in Section II-B, since it is also a PMM. In order to do so, it suffices to replace the parameters in (9) by those given by (7).

C. Performance Analysis

We analyse the performance of the proposed approximate state estimation in Section III-B in terms of mean square error (MSE), based on the error covariance matrix. Let us define the estimation error as $\boldsymbol{\epsilon}_n = \mathbf{x}_n - \hat{\mathbf{x}}_n$, where $\mathbf{x}_n$ is extracted form truth model in (1) and $\hat{\mathbf{x}}_n$ is the estimate derived in Algorithm 2. After some algebra, we obtain the recursion

$$\boldsymbol{\epsilon}_n = [\mathbf{I}_K| - \mathbf{K}_n]\Big\{\mathbf{A}_n^*\tilde{\mathbf{C}}\boldsymbol{\epsilon}_{n-1} + \mathbf{C}\mathbf{B}^{(\mathrm{t})}\mathbf{w}_n + (\mathbf{C}\mathbf{A}^{(\mathrm{t})} - \mathbf{A}_n^*\mathbf{C})(\mathbf{t}_{n-1} - \mathbf{m}_{n-1}^{(\mathrm{t})})\Big\}$$

where $\tilde{\mathbf{C}} = [\mathbf{I}_K|\mathbf{0}_{K\times M}]^T$. It follows that the error covariance $\mathbf{P}_n^* = E\{\boldsymbol{\epsilon}_n\boldsymbol{\epsilon}_n^T\}$ admits the recursive expression

$$\mathbf{P}_n^* = [\mathbf{I}_K| - \mathbf{K}_n]\Big\{(\mathbf{A}_n^*\tilde{\mathbf{C}})\mathbf{P}_{n-1}^*(\mathbf{A}_n^*\tilde{\mathbf{C}})^T + \mathbf{\Sigma}_{n|n-1}^* + (\mathbf{A}_n^*\tilde{\mathbf{C}})\mathbf{P}_{n-1}^{(\mathrm{t}\epsilon)\,T}(\mathbf{C}\mathbf{A}^{(\mathrm{t})} - \mathbf{A}_n^*\mathbf{C})^T + (\mathbf{C}\mathbf{A}^{(\mathrm{t})} - \mathbf{A}_n^*\mathbf{C})\mathbf{P}_{n-1}^{(\mathrm{t}\epsilon)}(\mathbf{A}_n^*\tilde{\mathbf{C}})^T\Big\}[\mathbf{I}_K| - \mathbf{K}_n]^T, \quad (11)$$

where $\mathbf{P}_n^{(\mathrm{t}\epsilon)} = E[(\mathbf{t}_n - \mathbf{m}_n^{(\mathrm{t})})\boldsymbol{\epsilon}_n^T]$ is similarly obtained as

$$\mathbf{P}_n^{(\mathrm{t}\epsilon)} = \Big\{\mathbf{A}^{(\mathrm{t})}\Big(\mathbf{P}_{n-1}^{(\mathrm{tt})}(\mathbf{C}\mathbf{A}^{(\mathrm{t})} - \mathbf{A}_n^*\mathbf{C})^T + \mathbf{P}_{n-1}^{(\mathrm{t}\epsilon)}(\mathbf{A}_n^*\tilde{\mathbf{C}})^T\Big) + \mathbf{B}^{(\mathrm{t})}\mathbf{Q}^{(\mathrm{t})}\mathbf{B}^{(\mathrm{t})T}\mathbf{C}^T\Big\}[\mathbf{I}_K| - \mathbf{K}_n]^T.$$

Note that $\mathbf{P}_n^*$ is the exact error covariance of the PMM-KF estimate in Algorithm 2, while $\mathbf{P}_n$ in Algorithm 2 is the error covariance obtained by ignoring the suboptimality of the proposed ROM from Section III-A.

IV. Batch Processing Estimator

We now discuss the particular case when the spectral radius of $\mathbf{A}^{(\mathrm{t})}$ equals 1 (which is relevant in many important applications fields such as econometry [16] and tracking [23], among others). In this case, the values in $\mathbf{P}_n^{(\mathrm{tt})}$ grow unbounded when n becomes large (see (2)), thus rendering the matrix inversion in (9) ill-conditioned, which in turn sets a limit to the window length N for which Algorithm 1 can be applied in practice (see Section V). It follows that in such contexts, sequential processing is unfeasible and must be replaced by batch processing, described below.

During the i-th length-N window, the PMM-KF (Algorithm 2) processes the observations $\mathbf{y}_{iN+1:(i+1)N}$ and produces the state estimates $\hat{\mathbf{x}}_{iN+1:(i+1)N}$. It follows that initializing the mean TMM state at the beginning of the next window can be done by replacing $\mathbf{m}_0^{(\mathrm{t})}$ with $\mathbf{m}_{(i+1)N}^{(\mathrm{t})} = \mathbf{C}^T[\hat{\mathbf{x}}_{(i+1)N}^T, \mathbf{y}_{(i+1)N}^T]^T$. Also, reminding that the initial triplet error covariance over any length-N window is $\mathbf{P}_0^{(\mathrm{tt})}$, Algorithm 1 can be used to recompute the parameters $\{\mathbf{A}_n^*, \mathbf{b}_n^*, \mathbf{\Sigma}_{n|n-1}^*\}_{(i+1)N+1\le n\le(i+2)N}$ of the proposed reduced order PMM over the next length-N window. Finally, the initial state estimate over the next length-N window is obtained by replacing $\hat{\mathbf{x}}_0$ by $\mathbf{S}\mathbf{m}_{(i+1)N}^{(\mathrm{t})}$, where the selection matrix $\mathbf{S}$ is defined by $\mathbf{S} = [\mathbf{I}_K|\mathbf{0}_{K\times L}|\mathbf{0}_{K\times M}]$. Again, since the initial triplet error covariance, $\mathbf{P}_0^{(\mathrm{tt})}$, is the same for any window, the initial state error covariance matrix at the beginning of each window is set to $\mathbf{P}_0 = \mathbf{S}\mathbf{P}_0^{(\mathrm{tt})}\mathbf{S}^T$.

Batch processing over length-N windows is summarized in Fig. 1. Moreover, the asymptotic computational complexity to generate a single estimate at a given time instant n is $\mathcal{O}(M^3 +$

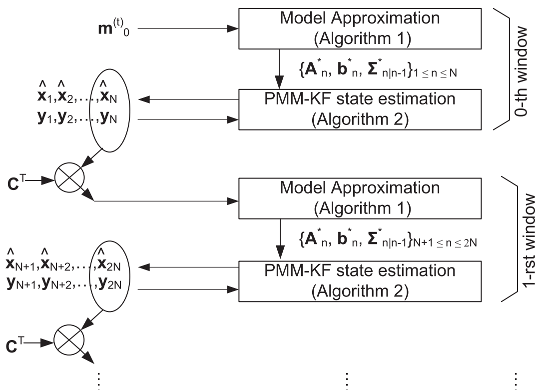

Fig. 1. Batch processing over length-N windows. For the sake of readability, only the parameters that change from one window to another are shown.

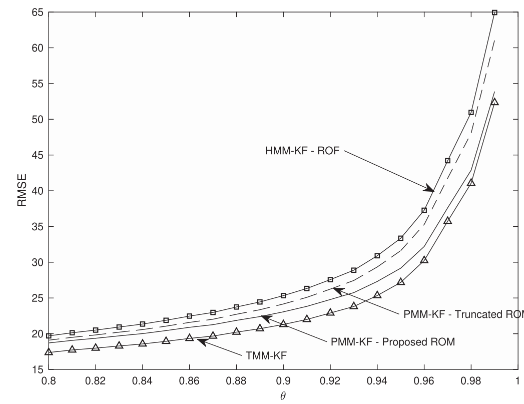

Fig. 2. Normalized RMSE as a function of θ, with $T = 1$ s, $Q = 10^2$ in $(\mathrm{m/s}^2)^2$ and $R = 0.1^2$ m^2.

$3K^2M + 2KM^2 + 2K^3$), which can be much less than the aforementioned complexity of the MMSE-optimal TMM-KF.

V. Numerical Results

The TMM formalism has been found suitable for applications with colored noise [9], [15]. Let us modify the standard HMM, so that the white process noise is replaced by a Markovian process noise $\mathbf{r}_n$ [20, p. 296]

$$\begin{aligned} \mathbf{x}_n &= \mathbf{F}\mathbf{x}_{n-1} + \mathbf{B}\mathbf{r}_n \\ \mathbf{r}_n &= \boldsymbol{\Theta}\mathbf{r}_{n-1} + \boldsymbol{\zeta}_n, \\ \mathbf{y}_n &= \mathbf{H}\mathbf{x}_n + \mathrm{v}_n, \end{aligned} \tag{12}$$

where the initial state $\mathbf{x}_0 \sim \mathcal{N}(\hat{\mathbf{x}}_0, \mathbf{P}_0)$ is independent from the zero-mean white Gaussian noise process $\mathbf{w}_n = [\boldsymbol{\zeta}_n^T, \mathbf{v}_n^T]^T$, $\forall n \geq 0$. The noise covariance is defined by $\mathbf{Q} = E\{\boldsymbol{\zeta}_n\boldsymbol{\zeta}_n^T\}$, $\mathbf{R} = E\{\mathbf{v}_n\mathbf{v}_n^T\}$ and $E\{\boldsymbol{\zeta}_n\mathbf{v}_n^T\} = \mathbf{0}$. In this system, $\mathbf{r}_n$ acts as an error source that we wish to eliminate from the state vector, corresponding to correlated process noise excited by the white driving noise $\boldsymbol{\zeta}_n$. (12) is straightfowardly converted to a Gaussian TMM of the form (1) by selecting the following parameters [9]

$$\mathbf{A}^{(\mathrm{t})} = \begin{bmatrix} \mathbf{F} & \mathbf{B}\boldsymbol{\Theta} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Theta} & \mathbf{0} \\ \mathbf{HF} & \mathbf{HB}\boldsymbol{\Theta} & \mathbf{0} \end{bmatrix}, \mathbf{B}^{(\mathrm{t})} = \begin{bmatrix} \mathbf{B} & \mathbf{0} \\ \mathbf{I} & \mathbf{0} \\ \mathbf{HB} & \mathbf{I} \end{bmatrix},$$

$$\mathbf{Q}^{(\mathrm{t})} = \begin{bmatrix} \mathbf{Q} & \mathbf{0} \\ \mathbf{0} & \mathbf{R} \end{bmatrix}. \tag{13}$$

Consider a classical two-state tracking problem, where $\mathbf{x}_n = [p_n, v_n]^T$ contains the position and velocity, with position measurements only [23, p. 273-277], so that (12) is parameterized by

$$\begin{aligned} &\mathbf{F} = \begin{bmatrix} 1 & T \\ 0 & 1 \end{bmatrix}, \mathbf{B} = \begin{bmatrix} \frac{T^2}{2} \\ T \end{bmatrix}, \mathbf{Q} = Q = 10^2 (\mathrm{m/s}^2)^2 \\ &\boldsymbol{\Theta} = \theta, \mathbf{R} = R = 0.1^2\,\mathrm{m}^2, \mathbf{H}_n = [1, 0], \end{aligned} \tag{14}$$

where $T = 1$ s and $0 < \theta < 1$ are the sampling period and the process noise coloration parameter, respectively. We choose as a performance measure the root MSE (RMSE) of the dimensionless state vector $\tilde{\mathbf{x}}_n = [p_n, T v_n]^T / \sqrt{R}$ [26].

Fig. 2 plots the RMSE for the dimensionless state vector as a function of θ, using the optimal TMM-KF, the batch PMM-KF applied to the proposed ROM (see Section IV), the PMM-KF applied to the truncated ROM (see Section II-B), and the classical reduced order filter (ROF) in [20, p. 297]. Note that since the spectral radius of $\mathbf{A}^{(\mathrm{t})}$ equals 1, the batch form of the proposed method over windows of length $N = 100$ is in order, while matrix inversion in Algorithm 1 was performed using INTLAB [27] to avoid numerical instability.

We observe that the proposed approach has performances close to the MMSE-optimal TMM-KF, while having lower complexity. When applying the PMM-KF (Algorithm 2), the truncated ROM approach has identical complexity wrt the proposed approach (see Rem. 3.2), but suffers from higher modeling error. Moreover the standard ROF with complexity order $\mathcal{O}(M^3 + 2\,K^2\,M + 2KM^2 + 3\,K^3 + 4\,K^2L + KL^2 + 2\,L^3)$, is an approximate HMM-KF having both higher complexity and worse accuracy compared to the proposed approach. We observe that the accuracy gain wrt the standard truncated ROM and ROF approaches increases with the noise coloration factor θ. A physical situation with a high value of θ in tracking, could correspond to a process noise related to a smooth acceleration.

VI. Conclusion

We introduded a reduced order model (ROM) for TMMs in the form of PMMs using probabilistic similarity criteria. An approximate Kalman filtering state estimator was developed for the ROM, enabling performances close to the MMSE optimal TMM-KF, at lower computational complexity. Efforts must be made to put this estimator in batch processing form, when the TMM system matrix is not asymptotically stable. The proposed method needs to strike a balance between complexity reduction (for which a large value of L is desirable) and filtering accuracy (which sets an upper limit to the value of L, combined with a careful selection of the discarded auxiliary variables). Simulation results confirm the improved performance of the proposed method compared to two classical ones.

REFERENCES

[1] R. E. Kalman, "A new approach to linear filtering and prediction problems," *Trans. ASME, J. Basic Eng.*, vol. 82, pp 35–45, Mar. 1960.

[2] W. Pieczynski and F. Desbouvries, "Kalman filtering using pairwise Gaussian models," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, 2003, pp. 57–60.

[3] V. Nemesin and S. Derrode, "Robust blind pairwise Kalman algorithms using QR decompositions," *IEEE Trans. Signal Process.*, vol. 61, no. 1, pp. 5–9, Jan. 2013.

[4] V. Nemesin and S. Derrode, "Robust partial-learning in linear Gaussian systems," *IEEE Trans. Autom. Control*, vol. 60, no. 9, pp. 2518–2523, Sep. 2015.

[5] M. V. Kulikova, "Gradient-based parameter estimation in pairwise linear Gaussian system," *IEEE Trans. Autom. Control*, vol. 62, no. 3, pp. 1511–1517, Mar. 2017.

[6] M. V. Kulikova and J. V. Tsyganova, "The UD-based approach for designing pairwise Kalman filtering algorithms," in *Proc. 20th World Congr. Int.-Federation Autom. Control*, 2017, pp. 1619–1624.

[7] F. Lehmann and W. Pieczynski, "State estimation in pairwise Markov models with improved robustness using unbiased FIR filtering," *Signal Process.*, no. 172, pp. 1–11, Jul. 2020.

[8] F. Desbouvries and W. Pieczynski, "Modèles de Markov triplet et filtrage de Kalman - Triplet Markov models and Kalman filtering," *Comptes Rendus de l'Académie des Sci. Mathématique, Série I*, vol. 336, no. 8, pp. 667–670, 2003.

[9] B. Ait-El-Fquih and F. Desbouvries, "Kalman filtering in triplet Markov chains," *IEEE Trans. Signal Process.*, vol. 54, no. 8, pp. 2957–2963, Aug. 2006.

[10] B. Ait-El-Fquih and I. Hoteit, "An efficient stateparameter filtering scheme combining ensemble Kalman and particle filters," *Monthly Weather Rev.*, vol. 146, no. 3, pp. 871–877, Mar. 2018.

[11] H. Qian, "A flexible state space model and its applications," *J. Time Ser. Anal.*, vol. 35, no. 2, pp. 79–88, Mar. 2014.

[12] M. V. Kulikova, J. V. Tsyganova, and G. Yu. Kulikov, "SVD-based state estimation of pairwise Markov models with application in econometrics," in *Proc. 22nd Int. Conf. Syst. Theory, Control Comput.*, 2018, pp. 188–193.

[13] C. K. Chui and G. Chen, *Kalman Filtering with Real-Time Applications*. Berlin, Germany: Springer, 1999.

[14] Y. Petetin and F. Desbouvries, "Exact Bayesian estimation in constrained triplet Markov chains," presented at the IEEE Int. Workshop Mach. Learn. Signal Process. (MLSP), Reims, France, 2014.

[15] J. D. Gibson, B. Koo, and S. D. Gray, "Filtering of colored noise for speech enhancement and coding," *IEEE Trans. Signal Process.*, vol. 39, no. 8, pp. 1732–1742, Aug. 1991.

[16] A. C. Harvey, *Forecasting*, *Structural Time Series Models and the Kalman Filter*. Cambridge, U.K.: Cambridge Univ. Press, 1989.

[17] L. Fortuna, G. Nunnari, and A. Gallo, *Model Order Reduction Techniques with Applications in Electrical Engineering*. Berlin, Germany: Springer, 1992.

[18] A. Gelb, *Applied Optimal Estimation*. Cambridge, MA, USA: MIT Press, 2001.

[19] C. S. Sims, "Reduced-order modelling and filtering," in *Control and Dynamic Systems*, New York, NY, USA: Academic, pp. 55–103, 1982.

[20] B. D. O. Anderson and J. B. Moore, *Optimal Filtering*. Englewood Cliffs, NJ, USA: Prentice-Hall, 1979.

[21] D. Simon, "Reduced order Kalman filtering without model reduction," *Control Intell. Syst.*, vol. 35, no. 2, pp. 169–174, Jan. 2007.

[22] E. E. Pentecost and A. R. Stubberud, "Synthesis of computationally efficient sequential linear estimators," *IEEE Trans. Aerosp. Electron. Syst.*, vol. AES-3, no. 2, pp. 242–249, Mar. 1967.

[23] Y. Bar-Shalom, X.-R. Li, and T. Kirubarajan, *Estimation With Applications to Tracking and Navigation*. Hoboken, NJ, USA: Wiley, 2001.

[24] A. R. Runnalls, "Kullback-Leibler approach to Gaussian mixture reduction," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 43, no. 3, pp. 989–999, Jul. 2007.

[25] D. Benboudjema, M. El Bouchraya Malainin, and W. Pieczynski, "Exact Kalman filtering in pairwise Gaussian switching systems," *Proc. Appl. Stochastic Models Data Anal.*, 2011, pp. 1–8.

[26] R. J. Fitzgerald, "Simple tracking filters: steady-state filtering and smoothing performance," *IEEE Trans. Aerosp. Electron. Syst.*, vol. AES-16, no. 6, pp. 860–864, Nov. 1980.

[27] S. Rump, *INTLAB-INTerval LABoratory. Developments in Reliable Computing*, T. Csendes, Ed. Norwell, MA, USA: Kluwer, pp. 77–104, 1999.