State estimation in pairwise Markov models with improved robustness using unbiased FIR filtering

Frederic Lehmann, Wojciech Pieczynski

Version of Record: https://www.sciencedirect.com/science/article/pii/S0165168420301110
Manuscript_e931913c623afb56ffa70322b5afb091

State estimation in pairwise Markov models with improved robustness using unbiased FIR filtering

Frederic Lehmann*,[a], Wojciech Pieczynski[a]

[a] *Télécom SudParis, Institut Polytechnique de Paris, 9 rue Charles Fourier 91011 EVRY Cedex France.*

Abstract

We propose a novel estimation procedure for linear time-varying pairwise Markov models (PMM), that is robust to system parameter uncertainties occurring in real-world applications. In order to cope with mismodeling errors and ignorance of noise/initial state statistics, we solve a finite-horizon state estimation problem. The resulting unbiased finite impulse response filter for PMMs (PMM-UFIR) is first derived in batch form and then converted to a recursive Kalman-like form for the sake of complexity reduction. Closed forms for the error covariance matrix of the state estimate are also provided for analytical performance assessment.

Numerical results illustrate the effectiveness of the proposed estimation method over Gaussian processes, by showing that the PMM-UFIR is nearly as accurate as (resp. more robust than) optimal filtering under perfect (resp. uncertain) system parameters after tuning the horizon size.

Key words: Pairwise Markov models, optimal filtering, Kalman filter, unbiased finite impulse response filter, robustness.

1. Introduction

Many problems in signal processing can be described as discrete-time stochastic systems that generate observable outputs [1]. A problem of fundamental interest in estimation theory is the choice of a suitable signal model. The model should be flexible enough to deal with various application contexts, while remaining simple enough to lead to tractable estimators in some optimality sense. Linear Gaussian hidden Markov models (HMM) have had a great success in modeling signals as Markovian states, where each observation is a noisy function of the current state. This is mainly due to the seminal paper [2] deriving the optimal filter for HMMs in the minimum mean square error (MMSE) sense, the Kalman filter (HMM-KF). The HMM-KF is a standard approach in many engineering fields such as tracking, navigation [3], control [4],

*Corresponding author - Phone: (+33) 1 60 76 46 33. Fax: (+33) 1 60 76 44 33
Email address: e-mail: frederic.lehmann@it-sudparis.eu (Frederic Lehmann)

March 2, 2020

econometrics [5] communication theory [6], speech processing [7], graph signal processing [8] and machine learning [9], to name a few. Many important disciplines (such as engineering, physics, chemistry, biology, economics and data processing, etc) involve a noisy observation function, whose arguments include the previous observation and/or the previous state (see [10], [11], [12] and [13], respectively for examples in time series analysis, biochemistry, inertial navigation and system identification). The standard way to handle such situations within the HMM framework is via state-augmentation, i.e. by appending the previous observation to the state [5]. This is necessary in this context, since HMMs break the symmetry between the state and observation processes, letting the state account for the entire system dynamics, while the observation is merely a noisy function of the state. A more concise framework avoiding state augmentation is provided by the pairwise Markov model (PMM) framework, introduced in [14]. Interestingly, an optimal filter (in the MMSE sense) exists for linear Gaussian PMMs, which we shall refer to as the PMM Kalman filter (PMM-KF) [15]-[17]. We will see that an alternative derivation of the PMM-KF can be obtained by converting a Gaussian PMM to a state-augmented HMM (SA HMM). Note that the PMM-KF has found applications in signal and image restoration [18]-[20], multi-object tracking [21]-[23] and financial data analysis [24].

Regarding the HMM, the Kalman filter has also known pitfalls in real-world contexts, due to its infinite impulse response (IIR) filter structure [25]. In particular, conditioning the estimate on the distant past is responsible for potential numerical instability with respect to (wrt) roundoff errors [26], lack of robustness wrt imprecisely known dynamics [25] and dependence on initial state statistics, for which little more than a guess is usually available. A well-known solution to this problem is to build estimators with finite impulse response (FIR) structure [25], thus limiting the memory to a finite-horizon corresponding to the N most recent observations. Optimal (MMSE) FIR (HMM-OFIR) estimators ignoring the initial state but with prior knowledge of the noise statistics, were developed in [27] and [28], from a Bayesian and a receding-horizon perspective, respectively. For the case when noise statistics are also unavailable, a FIR estimator satisfying the unbiasedness condition (HMM-UFIR) has been derived in batch form in [29]-[31]. However, these FIR estimators have their own limitations, especially in terms of computational burden. Therefore, fast Kalman-like recursive versions of the OFIR and UFIR have been derived in [32] and [33]-[34], respectively. Furthemore, it is desirable to tune the horizon size in order to maximize the filter performance [35].

In this paper, we focus on state estimation in linear Gaussian PMMs for applications, where issues such as mismodeling and unknown (or uncertain) noise/initial state statistics matter [36]. Therefore, we propose

an estimator with FIR structure and whose gain matrix ignores the noise statistics. We derive the presented estimator as a UFIR filter adapted to the context of PMMs, which we shall refer to as the PMM-UFIR. However, the derivation of the PMM-UFIR is not as straightforward as its HMM counterpart. This comes from the fact that by construction, in a PMM, both the state and observation process are noise-driven and observation-driven [15]. Consequently, a linear Gaussian PMM lacks the direct input/output form that was exploited to derive the HMM-UFIR [33]. Reformulating the dependence of a state wrt to observations backward in time and taking observation feedback into account, the new PMM-UFIR processing the N most recent observations is obtained in batch form. In addition, we show that this FIR filter can also be transformed to a Kalman-like iterative form, for the sake of fast computation.

The main contributions of this work are:

- A state estimator for linear Gaussian PMMs agnostic to the noise and initial state statistics, in batch FIR form (batch PMM-UFIR);
- A fast implementation of the FIR estimator, processing each observation at a time in Kalman-like form (Kalman-like PMM-UFIR);
- An analysis of the estimation error, useful for theoretical performance evaluation
- Extensive numerical simulations showing that the PMM-UFIR outperforms its SA HMM counterpart over a wide range of applications.

Throughout the paper, bold letters indicate vectors and matrices while $\mathbf{0}_{m\times n}$ (resp. $\mathbf{I}_m$) is the $m \times n$ all-zero (resp. the $m \times m$ identity) matrix and diag($\mathbf{a}$) is the (block) diagonal matrix, whose diagonal entries are stored in $\mathbf{a}$ and whose off-diagonal entries are zero. The superscript ^{T} denotes the transpose of a matrix. $\mathcal{N}(\mathbf{m}, \mathbf{C})$ denotes a Gaussian distribution with mean $\mathbf{m}$ and covariance matrix $\mathbf{C}$. ID$(\mathbf{m}, \mathbf{C})$ stands for independently distributed with mean $\mathbf{m}$ and covariance matrix $\mathbf{C}$.

Let $\mathbf{y}_n$ be the n-th observation, a set of observations from time 0 up to time n is denoted by $\mathbf{y}_{0:n}$.

This paper is organized as follows. First, in Sec. 2 we formulate the state estimation problem under the linear Gaussian pairwise Markov model (PMM). In Sec. 3 we develop our unbiased FIR solution first in batch form, then in iterative Kalman-like form. Sec. 4, provides an analysis of the estimation error. Finally, in Sec. 5, the performances of the proposed algorithm are assessed for realistic applications.

2. Problem Formulation

We first introduce pairwise Markov models (PMMs) as a class of models with interesting features such as treating states and observations in a unified manner, that generalize hidden Markov models (HMMs). Considering the linear Gaussian case, we briefly recall the optimal infinite-horizon state estimator, referred to as the PMM-KF. Then, for the sake of finding an estimator in FIR form, we write the PMM over an horizon of N most recent time instants.

2.1. Pairwise Markov Model (PMM)

Let $\mathbf{x}_n \in \mathbb{R}^K$ and $\mathbf{y}_n \in \mathbb{R}^M$ denote the state and observation vectors at instant n, respectively. We consider a class of discrete time-varying pairwise Markov models (PMMs), where the pairwise process $\{\mathbf{x}_n, \mathbf{y}_n\}_{n\geq 0}$ is Markovian [14]. For the linear Gaussian case, a PMM can be described as the following discrete-time linear stochastic system

$$\begin{bmatrix} \mathbf{x}_n \\ \mathbf{y}_n \end{bmatrix} = \begin{bmatrix} \mathbf{A}_n^{(1)} & \mathbf{A}_n^{(2)} \\ \mathbf{A}_n^{(3)} & \mathbf{A}_n^{(4)} \end{bmatrix} \begin{bmatrix} \mathbf{x}_{n-1} \\ \mathbf{y}_{n-1} \end{bmatrix} + \begin{bmatrix} \mathbf{B}_n^{(1)} & \mathbf{B}_n^{(2)} \\ \mathbf{B}_n^{(3)} & \mathbf{B}_n^{(4)} \end{bmatrix} \begin{bmatrix} \mathbf{w}_n \\ \mathbf{v}_n \end{bmatrix}, \tag{1}$$

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 $\mathbf{z}_n = [\mathbf{w}_n^T, \mathbf{v}_n^T]^T$, $\forall n \geq 0$. The noise covariance is defined by $\mathbf{Q}_n = E\{\mathbf{w}_n \mathbf{w}_n^T\}$, $\mathbf{R}_n = E\{\mathbf{v}_n \mathbf{v}_n^T\}$ and $\mathbf{M}_n = E\{\mathbf{w}_n \mathbf{v}_n^T\}$. Using an infinite horizon approach, the predictive and posterior distributions

$$p(\mathbf{x}_n, \mathbf{y}_n | \mathbf{y}_{0:n-1}) = \mathcal{N}\left(\begin{bmatrix} \hat{\mathbf{x}}_n^- \\ \hat{\mathbf{y}}_n^- \end{bmatrix}, \begin{bmatrix} \mathbf{P}_n^- & \mathbf{\Sigma}_n^- \\ \mathbf{\Sigma}_n^{-T} & \mathbf{L}_n^- \end{bmatrix} \right)$$

$$p(\mathbf{x}_n | \mathbf{y}_{0:n}) = \mathcal{N}\left(\hat{\mathbf{x}}_n, \mathbf{P}_n\right).$$

are obtained through the Kalman-like recursions given by Algorithm 1, that we will refer to as the PMM-KF in the sequel. Although a demonstration is provided in [19] and [20] for particular pairwise Markov chains and trees, respectively, for the sake of completeness we provide a derivation for the considered general case in Appendix A.

Remark 2.1. *Note that a classical HMM of the form*

$$\begin{aligned} \mathbf{x}_n &= \mathbf{F}_n \mathbf{x}_{n-1} + \mathbf{B}_n \mathbf{w}_n \\ \mathbf{y}_n &= \mathbf{H}_n \mathbf{x}_n + \mathbf{D}_n \mathbf{v}_n \end{aligned} \tag{2}$$

is a particular PMM obtained by selecting

$$\begin{bmatrix} \mathbf{A}_n^{(1)} & \mathbf{A}_n^{(2)} \\ \mathbf{A}_n^{(3)} & \mathbf{A}_n^{(4)} \end{bmatrix} = \begin{bmatrix} \mathbf{F}_n & \mathbf{0}_{K\times M} \\ \mathbf{H}_n \mathbf{F}_n & \mathbf{0}_{M\times M} \end{bmatrix}, \begin{bmatrix} \mathbf{B}_n^{(1)} & \mathbf{B}_n^{(2)} \\ \mathbf{B}_n^{(3)} & \mathbf{B}_n^{(4)} \end{bmatrix} = \begin{bmatrix} \mathbf{B}_n & \mathbf{0} \\ \mathbf{H}_n \mathbf{B}_n & \mathbf{D}_n \end{bmatrix}.$$

In this case, the PMM-KF boils down to the classical Kalman filter (HMM-KF) for correlated process and measurement noise [37, p. 187].

Algorithm 1 PMM-KF algorithm

Require: $\mathbf{y}_n$, $\hat{\mathbf{x}}_0$, $\mathbf{P}_0$, $\mathbf{Q}_n$, $\mathbf{R}_n$, $\mathbf{M}_n$

for $n = 1, 2, \ldots$ **do**

$$\hat{\mathbf{x}}_n^- = \mathbf{A}_n^{(1)}\hat{\mathbf{x}}_{n-1} + \mathbf{A}_n^{(2)}\mathbf{y}_{n-1}$$

$$\hat{\mathbf{y}}_n^- = \mathbf{A}_n^{(3)}\hat{\mathbf{x}}_{n-1} + \mathbf{A}_n^{(4)}\mathbf{y}_{n-1}$$

$$\mathbf{P}_n^- = \mathbf{A}_n^{(1)}\mathbf{P}_{n-1}{\mathbf{A}_n^{(1)}}^T + \mathbf{B}_n^{(1)}\mathbf{Q}_n{\mathbf{B}_n^{(1)}}^T + \mathbf{B}_n^{(1)}\mathbf{M}_n{\mathbf{B}_n^{(2)}}^T + \mathbf{B}_n^{(2)}\mathbf{M}_n^T{\mathbf{B}_n^{(1)}}^T + \mathbf{B}_n^{(2)}\mathbf{R}_n{\mathbf{B}_n^{(2)}}^T$$

$$\mathbf{\Sigma}_n^- = \mathbf{A}_n^{(1)}\mathbf{P}_{n-1}{\mathbf{A}_n^{(3)}}^T + \mathbf{B}_n^{(1)}\mathbf{Q}_n{\mathbf{B}_n^{(3)}}^T + \mathbf{B}_n^{(1)}\mathbf{M}_n{\mathbf{B}_n^{(4)}}^T + \mathbf{B}_n^{(2)}\mathbf{M}_n^T{\mathbf{B}_n^{(3)}}^T + \mathbf{B}_n^{(2)}\mathbf{R}_n{\mathbf{B}_n^{(4)}}^T$$

$$\mathbf{L}_n^- = \mathbf{A}_n^{(3)}\mathbf{P}_{n-1}{\mathbf{A}_n^{(3)}}^T + \mathbf{B}_n^{(3)}\mathbf{Q}_n{\mathbf{B}_n^{(3)}}^T + \mathbf{B}_n^{(3)}\mathbf{M}_n{\mathbf{B}_n^{(4)}}^T + \mathbf{B}_n^{(4)}\mathbf{M}_n^T{\mathbf{B}_n^{(3)}}^T + \mathbf{B}_n^{(4)}\mathbf{R}_n{\mathbf{B}_n^{(4)}}^T$$

$$\mathbf{K}_n = \mathbf{\Sigma}_n^-(\mathbf{L}_n^-)^{-1}$$

$$\hat{\mathbf{x}}_n = \hat{\mathbf{x}}_n^- + \mathbf{K}_n(\mathbf{y}_n - \hat{\mathbf{y}}_n^-)$$

$$\mathbf{P}_n = \mathbf{P}_n^- - \mathbf{K}_n{\mathbf{\Sigma}_n^-}^T$$

return $\hat{\mathbf{x}}_n$, $\mathbf{P}_n$

end for

Remark 2.2. *Conversely, a PMM can be converted to the following state-augmented HMM (SA HMM) form*

$$\begin{cases} \begin{bmatrix}\mathbf{x}_n\\ \mathbf{y}_n\end{bmatrix} = \begin{bmatrix}\mathbf{A}_n^{(1)} & \mathbf{A}_n^{(2)}\\ \mathbf{A}_n^{(3)} & \mathbf{A}_n^{(4)}\end{bmatrix}\begin{bmatrix}\mathbf{x}_{n-1}\\ \mathbf{y}_{n-1}\end{bmatrix} + \begin{bmatrix}\mathbf{B}_n^{(1)} & \mathbf{B}_n^{(2)}\\ \mathbf{B}_n^{(3)} & \mathbf{B}_n^{(4)}\end{bmatrix}\begin{bmatrix}\mathbf{w}_n\\ \mathbf{v}_n\end{bmatrix} \\ \mathbf{y}_n = [\mathbf{0}_{M\times K}|\mathbf{I}_M]\begin{bmatrix}\mathbf{x}_n\\ \mathbf{y}_n\end{bmatrix} \end{cases} \tag{3}$$

Interestingly, it is straightforward to show that applying the standard HMM-KF [2] to the SA HMM (3), results in an estimation method (SA HMM-KF) that is formally equivalent to Algorithm 1. Although this idea is widely used in the literature [5], the SA HMM is somewhat counterintuitive (since the very name of HMM embodies the idea that the hidden states are not observed), therefore we argue that the original PMM is a more concise framework from which the optimal filter in the MMSE sense can be derived directly.

2.2. Extended pairwise Markov Model (e-PMM)

For the sake of finding an estimator in FIR form, a simple approach familiar in system identification [38], is to construct an extended model over the horizon $[m = n - N + 1, n]$ of N most recent time instants. Let us assume that the matrix $\mathbf{A}_n^{(1)}$ is invertible [39]. Starting from $\mathbf{x}_n$, we can compute the backward-in-time solution of the state in (1) for $i = m + 1, \ldots, n$ as

$$\mathbf{x}_{i-1} = \bar{\mathbf{A}}_{i:n}^{(1)}\mathbf{x}_n - \sum_{k=i}^{n}\bar{\mathbf{A}}_{i:k}^{(1)}(\mathbf{A}_k^{(2)}\mathbf{y}_{k-1} + \mathbf{B}_k^{(1)}\mathbf{w}_k + \mathbf{B}_k^{(2)}\mathbf{v}_k), \tag{4}$$

where the backward-in-time state transition matrix is defined as

$$\bar{\mathbf{A}}_{i:n}^{(1)} = \prod_{k=i}^{n} (\mathbf{A}_k^{(1)})^{-1}. \tag{5}$$

Consequently, we obtain the solution for the observation in (1) for $i = m+1, \ldots, n$ as

$$\begin{aligned} \mathbf{y}_i = {} & \mathbf{A}_i^{(4)} \mathbf{y}_{i-1} + \mathbf{B}_i^{(3)} \mathbf{w}_i + \mathbf{B}_i^{(4)} \mathbf{v}_i + \\ & \mathbf{A}_i^{(3)} \left[\bar{\mathbf{A}}_{i:n}^{(1)} \mathbf{x}_n - \sum_{k=i}^{n} \bar{\mathbf{A}}_{i:k}^{(1)} (\mathbf{A}_k^{(2)} \mathbf{y}_{k-1} + \mathbf{B}_k^{(1)} \mathbf{w}_k + \mathbf{B}_k^{(2)} \mathbf{v}_k) \right]. \end{aligned} \tag{6}$$

Let us introduce the extended state $\mathbf{X}_{k,l} \in \mathbb{R}^{(k-l+1)K}$ (resp. observation $\mathbf{Y}_{k,l} \in \mathbb{R}^{(k-l+1)M}$) vector over the horizon $[l, k]$

$$\begin{aligned} \mathbf{X}_{k,l} &= [\mathbf{x}_k^T, \mathbf{x}_{k-1}^T, \ldots, \mathbf{x}_l^T]^T \\ \mathbf{Y}_{k,l} &= [\mathbf{y}_k^T, \mathbf{y}_{k-1}^T, \ldots, \mathbf{y}_l^T]^T. \end{aligned} \tag{7}$$

Similarly, defining the extended noise vectors

$$\begin{aligned} \mathbf{W}_{k,l} &= [\mathbf{w}_k^T, \mathbf{w}_{k-1}^T, \ldots, \mathbf{w}_l^T]^T \\ \mathbf{V}_{k,l} &= [\mathbf{v}_k^T, \mathbf{v}_{k-1}^T, \ldots, \mathbf{v}_l^T]^T, \end{aligned} \tag{8}$$

the e-PMM is obtained by writing (4) and (6) in equivalent matrix form

$$\begin{cases} \mathbf{X}_{n-1,m} = \bar{\mathbf{A}}_{n,m+1}^{(1)} \mathbf{x}_n - \mathbf{A}_{n,m+1}^{(2)} \mathbf{Y}_{n-1,m} + \\ \qquad\qquad \mathbf{G}_{n,m+1}^{(1)} \mathbf{W}_{n,m+1} + \mathbf{G}_{n,m+1}^{(2)} \mathbf{V}_{n,m+1} \\ (\mathbf{J}_{n,m+1} + \mathbf{J}'_{n,m+1}) \mathbf{Y}_{n,m} = \mathbf{H}_{n,m+1} \mathbf{x}_n + \\ \qquad\qquad \mathbf{D}_{n,m+1}^{(3)} \mathbf{W}_{n,m+1} + \mathbf{D}_{n,m+1}^{(4)} \mathbf{V}_{n,m+1}, \end{cases} \tag{9}$$

where

$$\bar{\mathbf{A}}_{n,m+1}^{(1)} = [\bar{\mathbf{A}}_{n:n}^{(1)\,T}, \bar{\mathbf{A}}_{n-1:n}^{(1)\,T}, \ldots, \bar{\mathbf{A}}_{m+1:n}^{(1)\,T}]^T \tag{10}$$

$$\begin{aligned} & \mathbf{A}_{n,m+1}^{(2)} \\ & = \begin{bmatrix} \bar{\mathbf{A}}_{n:n}^{(1)} \mathbf{A}_n^{(2)} & \mathbf{0} & \ldots & \mathbf{0} \\ \bar{\mathbf{A}}_{n-1:n}^{(1)} \mathbf{A}_n^{(2)} & \bar{\mathbf{A}}_{n-1:n-1}^{(1)} \mathbf{A}_{n-1}^{(2)} & \ldots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \bar{\mathbf{A}}_{m+2:n}^{(1)} \mathbf{A}_n^{(2)} & \bar{\mathbf{A}}_{m+2:n-1}^{(1)} \mathbf{A}_{n-1}^{(2)} & \ldots & \mathbf{0} \\ \bar{\mathbf{A}}_{m+1:n}^{(1)} \mathbf{A}_n^{(2)} & \bar{\mathbf{A}}_{m+1:n-1}^{(1)} \mathbf{A}_{n-1}^{(2)} & \ldots & \bar{\mathbf{A}}_{m+1:m+1}^{(1)} \mathbf{A}_{m+1}^{(2)} \end{bmatrix}, \end{aligned} \tag{11}$$

$$\begin{aligned} & \mathbf{G}_{n,m+1}^{(1)} \\ & = - \begin{bmatrix} \bar{\mathbf{A}}_{n:n}^{(1)} \mathbf{B}_n^{(1)} & \mathbf{0} & \ldots & \mathbf{0} \\ \bar{\mathbf{A}}_{n-1:n}^{(1)} \mathbf{B}_n^{(1)} & \bar{\mathbf{A}}_{n-1:n-1}^{(1)} \mathbf{B}_{n-1}^{(1)} & \ldots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \bar{\mathbf{A}}_{m+2:n}^{(1)} \mathbf{B}_n^{(1)} & \bar{\mathbf{A}}_{m+2:n-1}^{(1)} \mathbf{B}_{n-1}^{(1)} & \ldots & \mathbf{0} \\ \bar{\mathbf{A}}_{m+1:n}^{(1)} \mathbf{B}_n^{(1)} & \bar{\mathbf{A}}_{m+1:n-1}^{(1)} \mathbf{B}_{n-1}^{(1)} & \ldots & \bar{\mathbf{A}}_{m+1:m+1}^{(1)} \mathbf{B}_{m+1}^{(1)} \end{bmatrix}, \end{aligned} \tag{12}$$

$$\mathbf{G}^{(2)}_{n,m+1}$$
$$= -\begin{bmatrix} \bar{\mathbf{A}}^{(1)}_{n:n}\mathbf{B}^{(2)}_{n} & \mathbf{0} & \dots & \mathbf{0} \\ \bar{\mathbf{A}}^{(1)}_{n-1:n}\mathbf{B}^{(2)}_{n} & \bar{\mathbf{A}}^{(1)}_{n-1:n-1}\mathbf{B}^{(2)}_{n-1} & \dots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \bar{\mathbf{A}}^{(1)}_{m+2:n}\mathbf{B}^{(2)}_{n} & \bar{\mathbf{A}}^{(1)}_{m+2:n-1}\mathbf{B}^{(2)}_{n-1} & \dots & \mathbf{0} \\ \bar{\mathbf{A}}^{(1)}_{m+1:n}\mathbf{B}^{(2)}_{n} & \bar{\mathbf{A}}^{(1)}_{m+1:n-1}\mathbf{B}^{(2)}_{n-1} & \dots & \bar{\mathbf{A}}^{(1)}_{m+1:m+1}\mathbf{B}^{(2)}_{m+1} \end{bmatrix}. \tag{13}$$

$\mathbf{J}_{n,m+1}$ is the upper bidiagonal block matrix expressed as

$$\mathbf{J}_{n,m+1} = \begin{bmatrix} \mathbf{I}_M & -\mathbf{A}^{(4)}_{n} & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_M & -\mathbf{A}^{(4)}_{n-1} & \dots & \mathbf{0} \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ \mathbf{0} & \dots & \mathbf{I}_M & -\mathbf{A}^{(4)}_{m+2} & \mathbf{0} \\ \mathbf{0} & \dots & \mathbf{0} & \mathbf{I}_M & -\mathbf{A}^{(4)}_{m+1} \end{bmatrix}. \tag{14}$$

Using the notation $\mathbf{A}^{(3)}_{n,m+1} = \mathrm{diag}(\mathbf{A}^{(3)}_{n}, \mathbf{A}^{(3)}_{n-1}, \dots, \mathbf{A}^{(3)}_{m+1})$ the remaining matrices are defined by

$$\begin{aligned} \mathbf{J}'_{n,m+1} &= [\mathbf{0}_{(N-1)M\times M} | \mathbf{A}^{(3)}_{n,m+1}\mathbf{A}^{(2)}_{n,m+1}] \\ \mathbf{H}_{n,m+1} &= \mathbf{A}^{(3)}_{n,m+1}\bar{\mathbf{A}}^{(1)}_{n,m+1} \\ \mathbf{D}^{(3)}_{n,m+1} &= \mathbf{A}^{(3)}_{n,m+1}\mathbf{G}^{(1)}_{n,m+1} + \mathrm{diag}(\mathbf{B}^{(3)}_{n}, \mathbf{B}^{(3)}_{n-1}, \dots, \mathbf{B}^{(3)}_{m+1}) \\ \mathbf{D}^{(4)}_{n,m+1} &= \mathbf{A}^{(3)}_{n,m+1}\mathbf{G}^{(2)}_{n,m+1} + \mathrm{diag}(\mathbf{B}^{(4)}_{n}, \mathbf{B}^{(4)}_{n-1}, \dots, \mathbf{B}^{(4)}_{m+1}). \end{aligned} \tag{15}$$

3. PMM Unbiased FIR Estimator

A FIR estimator has the form of a discrete-time convolution of the N most recent observations. In the absence of prior information on the initial state and noise statistics, an FIR estimator, satisfying the constraint that the expected value of the estimator equals that of the state (UFIR), is of interest. For a linear Gaussian HMM, the derivation of the HMM-UFIR relies on the fact that the extended model is in classical input/output form [34], i.e. $\mathbf{Y}_{n,m}$ is a linear combination of the desired state $\mathbf{x}_n$ and a linear anticausal transformation of noise terms. However, for a linear Gaussian PMM, the corresponding extended model obtained in (9) does not satisfy this property anymore. Thus, there is a need to find a new UFIR filter valid for PMMs (PMM-UFIR) in batch form. Then, we will show that the batch form also admits an iterative Kalman-like form. Importantly, the proposed PMM-UFIR applied to the PMM in (1) needs only the matrix $\mathbf{A}^{(1)}_{n}$ to be invertible. However, the UFIR in [34] applied to the equivalent SA HMM model in (3) (referred to as the SA HMM-UFIR) requires the whole PMM transition matrix to be invertible. Thus, from a theoretical point of view, the PMM-UFIR works under weaker hypotheses than its SA HMM-UFIR counterpart.

3.1. Batch UFIR Estimator

Let the batch UFIR estimator take the form of a linear combination of the transformed observations appearing in the second line of the e-PMM (9)

$$\hat{\mathbf{x}}_n = \bar{\mathcal{H}}_{n,m+1}(\mathbf{J}_{n,m+1} + \mathbf{J}'_{n,m+1})\mathbf{Y}_{n,m}, \tag{16}$$

where $\bar{\mathcal{H}}_{n,m+1}$ is a $K \times (N-1)M$-dimensional gain matrix of the filter. A UFIR filter [34] satisfies the constraint

$$E\{\hat{\mathbf{x}}_n\} = E\{\mathbf{x}_n\}, \quad \forall E\{\mathbf{x}_n\}, \tag{17}$$

which means that the expected value of the estimator is guaranteed to be that of the state. Using the fact that the noise terms are zero-mean in the second line of the e-PMM (9), (17) becomes

$$\bar{\mathcal{H}}_{n,m+1}\mathbf{H}_{n,m+1} = \mathbf{I}_K. \tag{18}$$

Moreover, the effect of noise and uncertainties on the estimator can be reduced by minimizing the Frobenius norm of the filter gain matrix $\bar{\mathcal{H}}_{n,m+1}$.

When $\text{rank}(\mathbf{H}_{n,m+1}) = K$ (which is analog to the observability condition for HMMs), minimizing the Frobenius norm of $\bar{\mathcal{H}}_{n,m+1}$ of under the constraint (18) admits a unique solution [40, pp. 85-86]

$$\bar{\mathcal{H}}_{n,m+1} = (\mathbf{H}^T_{n,m+1}\mathbf{H}_{n,m+1})^{-1}\mathbf{H}^T_{n,m+1}. \tag{19}$$

We also introduce the generalized noise power gain (GNPG) as

$$\mathbf{G}_n = (\mathbf{H}^T_{n,m+1}\mathbf{H}_{n,m+1})^{-1} = \bar{\mathcal{H}}_{n,m+1}\bar{\mathcal{H}}^T_{n,m+1}, \tag{20}$$

which is a quantity related to the noise reduction factor of the filter [33]-[34]. Importantly, note that unlike the Kalman filter, the PMM-UFIR performs state estimation without prior knowledge on the noise covariances and/or the initial condition. Only the horizon size N is needed.

Remark 3.1. *Considering the particular PMM in Remark 2.1 corresponding to the classical HMM (2), we obtain*

$$(\mathbf{J}_{n,m+1} + \mathbf{J}'_{n,m+1})\mathbf{Y}_{n,m} = \mathbf{Y}_{n,m+1}$$

$$\mathbf{H}_{n,m+1} = [(\mathbf{H}_n)^T, (\mathbf{H}_{n-1}\bar{\mathbf{F}}_{n:n})^T, \ldots, (\mathbf{H}_{m+1}\bar{\mathbf{F}}_{m+2:n})^T]^T.$$

It follows that (16) reduces to the classical batch HMM-UFIR estimate over the horizon $[m+1, n]$, unlike the horizon $[m, n]$ in [34]. Note that this slight difference in the horizon size comes from the fact that the

extended PMM does not reduce to the extended HMM when the PMM reduces to a classical HMM. In other words, the principle of the derivation of the PMM-UFIR is different from its HMM-UFIR counterpart.

3.2. Kalman-like UFIR Estimator

The PMM-UFIR in Sec. 3.1 has the drawbacks of any batch processor, i.e. a high computational complexity per time instant, especially when the horizon size N is large. For the sake of fast computation, we now seek a recursive form for (16), that processes each new observation one at a time, similarly to the Kalman filter.

As a starting point, we compute the batch estimator over the horizon $[m, s]$, by selecting $s = m + K$ to make sure that $\text{rank}(\mathbf{H}_{s,m+1}) = K$ is satisfied.

We now show that for any $n > s$, $\hat{\mathbf{x}}_n$ can be written as a linear combination of $\hat{\mathbf{x}}_{n-1}$ and the new observation $\mathbf{y}_n$. The prediction stage computes the predicted state and observation estimates as

$$\begin{aligned} \hat{\mathbf{x}}_n^- &= \mathbf{A}_n^{(1)}\hat{\mathbf{x}}_{n-1} + \mathbf{A}_n^{(2)}\mathbf{y}_{n-1} \\ \hat{\mathbf{y}}_n^- &= \mathbf{A}_n^{(3)}\hat{\mathbf{x}}_{n-1} + \mathbf{A}_n^{(4)}\mathbf{y}_{n-1}. \end{aligned} \tag{21}$$

Let us introduce the notation

$$\tilde{\mathbf{H}}_n = \mathbf{A}_n^{(3)}(\mathbf{A}_n^{(1)})^{-1}. \tag{22}$$

The update stage first updates the GNPG as

$$\mathbf{G}_n = [\tilde{\mathbf{H}}_n^T\tilde{\mathbf{H}}_n + (\mathbf{A}_n^{(1)}\mathbf{G}_{n-1}\mathbf{A}_n^{(1)T})^{-1}]^{-1} \tag{23}$$

then computes a bias correction gain (different from the Kalman gain) as

$$\mathbf{K}_n = \mathbf{G}_n\tilde{\mathbf{H}}_n^T, \tag{24}$$

and finally refines the state estimate as

$$\hat{\mathbf{x}}_n = \hat{\mathbf{x}}_n^- + \mathbf{K}_n(\mathbf{y}_n - \hat{\mathbf{y}}_n^-). \tag{25}$$

The demonstration is postponed to Appendix B.

The iterative Kalman-like procedure for computing the PMM-UFIR estimate is summarized in Algorithm 2.

Remark 3.2. *Considering again the particular PMM in Remark 2.1 corresponding to the classical HMM*

(2), we obtain

$$\hat{\mathbf{x}}_n^- = \mathbf{F}_n \hat{\mathbf{x}}_{n-1}$$

$$\hat{\mathbf{y}}_n^- = \mathbf{H}_n \mathbf{F}_n \hat{\mathbf{x}}_{n-1}$$

$$\mathbf{G}_n = [\mathbf{H}_n^T \mathbf{H}_n + (\mathbf{F}_n \mathbf{G}_{n-1} \mathbf{F}_n^T)^{-1}]^{-1}$$

$$\mathbf{K}_n = \mathbf{G}_n \mathbf{H}_n^T$$

$$\hat{\mathbf{x}}_n = \hat{\mathbf{x}}_n^- + \mathbf{K}_n(\mathbf{y}_n - \hat{\mathbf{y}}_n^-).$$

It follows that (25) reduces to the classical Kalman-like HMM-UFIR update rule [33]-[34].

Algorithm 2 Kalman-like PMM-UFIR algorithm

Require: $\mathbf{y}_n$, N

for $n = N,\, N+1,\, \ldots$ **do**

$m = n - N + 1$, $s = m + K$

$\mathbf{G}_s = (\mathbf{H}_{s,m+1}^T \mathbf{H}_{s,m+1})^{-1}$

$\hat{\mathbf{x}}_s = \bar{\mathcal{H}}_{s,m+1}(\mathbf{J}_{s,m+1} + \mathbf{J}'_{s,m+1})\mathbf{Y}_{s,m}$

for $l = s+1 : n$ **do**

$\hat{\mathbf{x}}_l^- = \mathbf{A}_l^{(1)} \hat{\mathbf{x}}_{l-1} + \mathbf{A}_l^{(2)} \mathbf{y}_{l-1}$

$\hat{\mathbf{y}}_l^- = \mathbf{A}_l^{(3)} \hat{\mathbf{x}}_{l-1} + \mathbf{A}_l^{(4)} \mathbf{y}_{l-1}$

$\mathbf{G}_l = [\tilde{\mathbf{H}}_l^T \tilde{\mathbf{H}}_l + (\mathbf{A}_l^{(1)} \mathbf{G}_{l-1} \mathbf{A}_l^{(1)T})^{-1}]^{-1}$

$\mathbf{K}_l = \mathbf{G}_l \tilde{\mathbf{H}}_l^T$

$\hat{\mathbf{x}}_l = \hat{\mathbf{x}}_l^- + \mathbf{K}_l(\mathbf{y}_l - \hat{\mathbf{y}}_l^-)$

end for

return $\hat{\mathbf{x}}_n$

end for

4. Performance Analysis

We analyse the performance of the proposed PMM-UFIR in terms of mean square error (MSE), based on the error covariance matrix. We will show that the error covariance matrix can be computed directly from the expression of the batch PMM-UFIR (see Sec. 4.1), but also in a computationally efficient iterative way using the Kalman-like form (see Sec. 4.2). Note that unlike the PMM-KF, the PMM-UFIR in iterative Kalman-like form does not need the error covariance matrix for the sake of state estimation. However, the resulting MSE will be instrumental in specifying the optimal horizon size (see Sec. 4.3) and we give indications of how this optimization can done in practice. Finally, we discuss complexity issues (see Sec. 4.4).

4.1. Direct Computation of the Error Covariance Matrix

Let us define the estimation error as $\boldsymbol{\epsilon}_n = \mathbf{x}_n - \hat{\mathbf{x}}_n$, then the corresponding error covariance matrix is

$$\mathbf{P}_n = E\{\boldsymbol{\epsilon}_n \boldsymbol{\epsilon}_n^T\}. \tag{26}$$

From the second equation in (9), the PMM-UFIR estimate (16) becomes

$$\hat{\mathbf{x}}_n = \mathbf{x}_n + \bar{\mathcal{H}}_{n,m+1}(\mathbf{D}^{(3)}_{n,m+1}\mathbf{W}_{n,m+1} + \mathbf{D}^{(4)}_{n,m+1}\mathbf{V}_{n,m+1}). \tag{27}$$

Consequently, we have

$$\begin{aligned}
\mathbf{P}_n = &\bar{\mathcal{H}}_{n,m+1} \\
&\times \Big[\mathbf{D}^{(3)}_{n,m+1}\,\mathrm{diag}(\mathbf{Q}_n, \mathbf{Q}_{n-1}, \ldots, \mathbf{Q}_{m+1}){\mathbf{D}^{(3)}_{n,m+1}}^T + \\
&\mathbf{D}^{(4)}_{n,m+1}\,\mathrm{diag}(\mathbf{R}_n, \mathbf{R}_{n-1}, \ldots, \mathbf{R}_{m+1}){\mathbf{D}^{(4)}_{n,m+1}}^T + \\
&\mathbf{D}^{(3)}_{n,m+1}\,\mathrm{diag}(\mathbf{M}_n, \mathbf{M}_{n-1}, \ldots, \mathbf{M}_{m+1}){\mathbf{D}^{(4)}_{n,m+1}}^T + \\
&\mathbf{D}^{(4)}_{n,m+1}\,\mathrm{diag}(\mathbf{M}_n, \mathbf{M}_{n-1}, \ldots, \mathbf{M}_{m+1})^T{\mathbf{D}^{(3)}_{n,m+1}}^T\Big] \\
&\times \bar{\mathcal{H}}^T_{n,m+1}.
\end{aligned} \tag{28}$$

The demonstration proceeds from (27) and the fact that the noise $\mathbf{z}_n = [\mathbf{w}_n^T, \mathbf{v}_n^T]^T$ is white and zero-mean in the PMM defined in Sec. 2.1.

4.2. Iterative Computation of the Error Covariance Matrix

For the sake of fast computation, we now seek a recursive form for (28). Starting from (1), we have the expression of $\mathbf{x}_n$ as

$$\mathbf{x}_n = \mathbf{A}^{(1)}_n \mathbf{x}_{n-1} + \mathbf{A}^{(2)}_n \mathbf{y}_{n-1} + \mathbf{B}^{(1)}_n \mathbf{w}_n + \mathbf{B}^{(2)}_n \mathbf{v}_n. \tag{29}$$

Then, combining (1) with (21) and (25), we have the expression of its PMM-UFIR estimate as

$$\begin{aligned}
\hat{\mathbf{x}}_n = &\mathbf{A}^{(1)}_n \hat{\mathbf{x}}_{n-1} + \mathbf{A}^{(2)}_n \mathbf{y}_{n-1} + \\
&\mathbf{K}_n(\mathbf{A}^{(3)}_n \mathbf{x}_{n-1} + \mathbf{A}^{(4)}_n \mathbf{y}_{n-1} + \mathbf{B}^{(3)}_n \mathbf{w}_n + \mathbf{B}^{(4)}_n \mathbf{v}_n \\
&- \mathbf{A}^{(3)}_n \hat{\mathbf{x}}_{n-1} - \mathbf{A}^{(4)}_n \mathbf{y}_{n-1}).
\end{aligned} \tag{30}$$

Substracting (30) from (29), the estimation error becomes

$$\begin{aligned}
\boldsymbol{\epsilon}_n = &(\mathbf{A}^{(1)}_n - \mathbf{K}_n \mathbf{A}^{(3)}_n)\boldsymbol{\epsilon}_{n-1} + \\
&(\mathbf{B}^{(1)}_n - \mathbf{K}_n \mathbf{B}^{(3)}_n)\mathbf{w}_n + (\mathbf{B}^{(2)}_n - \mathbf{K}_n \mathbf{B}^{(4)}_n)\mathbf{v}_n.
\end{aligned} \tag{31}$$

Now, using the fact that the noise $\mathbf{z}_n = [\mathbf{w}_n^T, \mathbf{v}_n^T]^T$ is white and zero-mean in the PMM defined in Sec. 2.1, we obtain the desired recursion

$$\begin{aligned}\mathbf{P}_n =&(\mathbf{A}_n^{(1)} - \mathbf{K}_n\mathbf{A}_n^{(3)})\mathbf{P}_{n-1}(\mathbf{A}_n^{(1)} - \mathbf{K}_n\mathbf{A}_n^{(3)})^T+\\ &(\mathbf{B}_n^{(1)} - \mathbf{K}_n\mathbf{B}_n^{(3)})\mathbf{Q}_n(\mathbf{B}_n^{(1)} - \mathbf{K}_n\mathbf{B}_n^{(3)})^T+\\ &(\mathbf{B}_n^{(2)} - \mathbf{K}_n\mathbf{B}_n^{(4)})\mathbf{R}_n(\mathbf{B}_n^{(2)} - \mathbf{K}_n\mathbf{B}_n^{(4)})^T+\\ &(\mathbf{B}_n^{(1)} - \mathbf{K}_n\mathbf{B}_n^{(3)})\mathbf{M}_n(\mathbf{B}_n^{(2)} - \mathbf{K}_n\mathbf{B}_n^{(4)})^T+\\ &(\mathbf{B}_n^{(2)} - \mathbf{K}_n\mathbf{B}_n^{(4)})\mathbf{M}_n^T(\mathbf{B}_n^{(1)} - \mathbf{K}_n\mathbf{B}_n^{(3)})^T.\end{aligned} \tag{32}$$

4.3. Optimal Horizon Selection

The PMM-UFIR has a degree of freedom that has not been exploited yet, namely the choice of the horizon size N [34]-[35]. Let us define the state estimation MSE for the PMM-UFIR over the horizon $[m, n]$ as trace$(\mathbf{P}_n)$. Obviously, this quantity depends on the horizon size since $N = n - m + 1$ (see Sec. 2.2). Therefore, a reasonable choice for tuning N to its optimum value N_{opt} consists in selecting

$$N_{\text{opt}} = \arg\min_N \text{trace}(\mathbf{P}_n). \tag{33}$$

In applications, we typically notice that $N < N_{\text{opt}}$ is the regime where the noise dominates the MSE (noise reduction is not optimal), while $N > N_{\text{opt}}$ corresponds to the regime where the estimate bias dominates [35]. This is a direct consequence of the fact that the estimator ignores the noise contributions over the horizon $[m + 1, n]$ (see Sec. 3.1).

Note that in practical applications, obtaining the optimal horizon size according to (33), would require:

- either an approximation of the of the noise/initial state covariance matrices
- or the availability of a test (training) sequence for the states, in order to compute the empirical estimator of the error covariance matrix as a function of the horizon size N [34]. In this case, the training sequence should be sufficiently large and smoothing of the objective function may also be desirable.

Alternatively, blind estimation of the optimal horizon size, relying on the measurements only, has been introduced in [35].

4.4. Complexity evaluation

We now focus on the worst-case time complexity, which is an important performance criterion. For all aforementioned algorithms, we evaluate the asymptotic computational complexity to generate a single

estimate at a given time instant n:

- the PMM-KF (see Algorithm 1) is $\mathcal{O}(M^3 + 3K^2M + 2KM^2 + 2K^3)$
- the batch PMM-UFIR (see Sec. 3.1) $\mathcal{O}(N^2KM^2)$
- for the Kalman-like PMM-UFIR (see Algorithm 2) is $\mathcal{O}(N(2K^2M + 3K^3))$.

As expected, the Kalman-like PMM-UFIR solves the computational inefficiency problem of the batch PMM-UFIR. Note that the complexity of the Kalman-like PMM-UFIR still grows linearly with the horizon size N, unlike the PMM-KF. Consequently, the price to be paid in order to get a robust estimate that is agnostic to initial state/noise covariances is an increase in computation time. This result for PMMs is not surprising, since it is in line with a similar complexity comparison for HMMs in [41].

5. Numerical Results

In this section, we first evaluate the performance of the proposed PMM-UFIR algorithm on synthetic data by avoiding state-augmentation in solving the following problems:

- the drift estimation problem in time series analysis/forecasting [5]
- the correlated measurement noise problem in tracking [37]
- estimation in a stochastic volatility model for financial time series.

In all simulations, the optimal horizon size, N_{opt}, is obtained offline through (33) using a grid search.

5.1. Random Walk With Drift

Many problems in time series analysis can be described by a simple random walk with drift model (see [5], [42], [43], [44] and [45], respectively for examples in econometrics, odometry, global temperature analysis, oil production forecasting and biology).

Let x_n be the stochastic drift at instant n, modeled by a stationary autoregressive process of order-1 (AR(1))

$$x_n = \rho x_{n-1} + \sqrt{1-\rho^2} w_n, \tag{34}$$

where $0 < \rho < 1$ sets the time correlation of the drift and $u_n \sim \mathcal{N}(0, Q)$ is the white Gaussian driving noise. The observed level of the time series of interest is modeled by a random walk with drift of the form

$$y_n = x_{n-1} + y_{n-1} + v_n, \tag{35}$$

Table 1: N_{opt} for the SA HMM-UFIR/PMM-UFIR applied to drift estimation with $Q = 1$ and $R = 1$.

ρ	0.80	0.81	0.82	0.83	0.84	0.85	0.86	0.87	0.88	0.89
N_{opt}	4	4	4	4	4	4	4	4	5	5
ρ	0.90	0.91	0.92	0.93	0.94	0.95	0.96	0.97	0.98	0.99
N_{opt}	5	5	5	6	6	6	7	8	10	13

where $v_n \sim \mathcal{N}(0, R)$ is white Gaussian noise.

The linear Gaussian PMM (1) corresponding to this problem has the following parameters

$$\begin{bmatrix} \mathbf{A}_n^{(1)} & \mathbf{A}_n^{(2)} \\ \mathbf{A}_n^{(3)} & \mathbf{A}_n^{(4)} \end{bmatrix} = \begin{bmatrix} \rho & 0 \\ 1 & 1 \end{bmatrix}, \begin{bmatrix} \mathbf{B}_n^{(1)} & \mathbf{B}_n^{(2)} \\ \mathbf{B}_n^{(3)} & \mathbf{B}_n^{(4)} \end{bmatrix} = \begin{bmatrix} \sqrt{1-\rho^2} & 0 \\ 0 & 1 \end{bmatrix}. \tag{36}$$

Assuming that w_n and v_n are independent, we also have $\mathbf{Q}_n = Q$, $\mathbf{R}_n = R$ and $\mathbf{M}_n = 0$, $\forall n \geq 0$. Note that both the state and the observation are one-dimensional (i.e. $K = M = 1$), so that as scalar version of the PMM-KF (Algorithm 1) and the PMM-UFIR (Algorithm 2) are in order for drift estimation. Setting $Q = 1$ and $R = 1$ and the range of the parameter $\rho \in [0.8, 0.99]$, the optimal horizon size for the PMM-UFIR is calculated using (33) and shown in Table I. As expected, the smaller the value of ρ, the lower the time-correlation of the drift and therefore the fewer the number of observations to build a meaningful PMM-UFIR estimate.

An SA HMM corresponding to the same problem, can be obtained by appending y_n to an augmented-state at time n, $\boldsymbol{\zeta}_n = [x_n, y_n]^T$, as proposed in [5]. The state-space representation (2) is obtained by selecting

$$\mathbf{F}_n = \begin{bmatrix} \rho & 0 \\ 1 & 1 \end{bmatrix}, \mathbf{B}_n = \begin{bmatrix} \sqrt{1-\rho^2} & 0 \\ 0 & 1 \end{bmatrix}, \mathbf{H}_n = [0, 1], \mathbf{D}_n = 0. \tag{37}$$

and $\mathbf{w}_n \sim \mathcal{N}\left([0, 0]^T, \text{diag}(Q, R)\right)$. The SA HMM-UFIR optimal horizon size was found to be identical to the one found for the PMM-UFIR.

Let us describe our simulation setup. The exact values of the true noise variances Q and R are used to simulate the trajectories of (34)-(35). However, in practical applications, the exact noise variances are difficult to determine exactly. Therefore, we assume that the SA HMM-KF and the PMM-KF algorithms have perfect knowledge of the initial condition, but employ p^2Q and q^2R instead of Q and R, with $p > 0$ and $q > 0$ in order to take noise covariance errors into account. The SA HMM-UFIR and the PMM-UFIR use the optimal horizon size in Table I.

We also run the SA HMM-KF/PMM-KF after expectation-maximization (EM) system parameter identification [43]. Note that for the PMM, EM estimates eight coefficients. EM is initialized with the correct

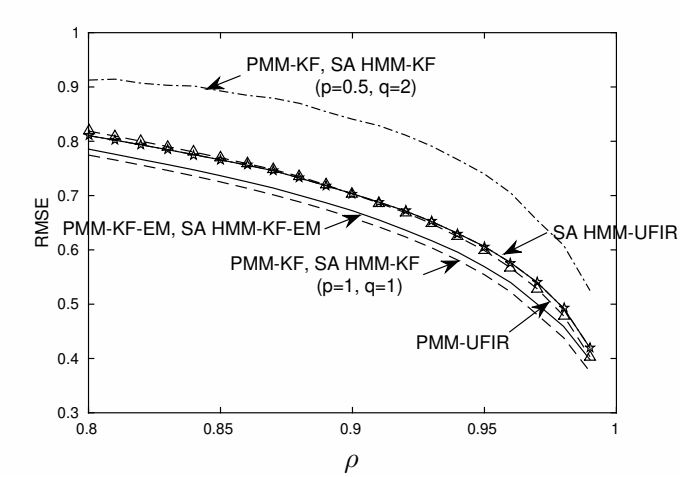

Figure 1: RMSE in drift estimation as a function of ρ, with $Q = 1$ and $R = 1$.

values of the unknown model parameters, so as to avoid being trapped in a local maximum of the log-likelihood. At each iteration, EM uses fixed-interval smoothing over the first 100 consecutive observations. A stopping criterion is met whenever the log-likelihood changes by less then 0.1 percent [43]. We refer to the Kalman algorithms using the system parameters obtained after EM convergence as the SA HMM-KF-EM and the PMM-KF-EM. Due to the very slow convergence of EM (tens of iterations), EM system identification is much longer than the subsequent SA HMM-KF-EM and the PMM-KF-EM over several thousands of observations. This method is also not robust, since any changes in the system parameters would require to run the computationally demanding EM again. Consequently, we merely consider the SA HMM-KF-EM and PMM-KF-EM as a reference method (i.e. the best Kalman filter under unknown system parameters), not as a real-time alternative to SA HMM-UFIR/PMM-UFIR.

Fig. 1 plots the root MSE (RMSE) for drift estimation using the aforementioned algorithms as a function of ρ. We observe that the PMM-KF reaches the same performances as the SA HMM-KF with the same complexity. Similarly, the PMM-UFIR has almost the same performances as its SA HMM counterpart at lower complexity. Also, by setting the horizon size to its optimal value, we observe that the RMSE of the PMM-UFIR is close the the RMSE of the PMM-KF with perfect knowledge of the noise variances (i.e. $p = q = 1$). However, the PMM-UFIR is insensitive to noise variance errors, while the SA HMM-KF and the PMM-KF are not, as exemplified by their deteriorated performances for $p = 0.5$ and $q = 2$.

In order to study the effect of the signal-to-noise ratio, we now set $Q = 1$, $\rho = 0.99$ and the range of the parameter $R \in [1, 10]$. The optimal horizon size for the PMM-UFIR is recalculated and shown in Table II, and is again the same as for the SA HMM-UFIR. As expected, the larger the value of R, the

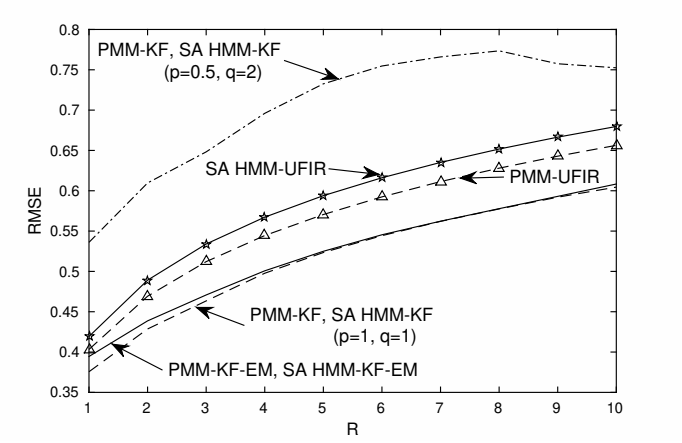

Figure 2: RMSE in drift estimation as a function of R, with $Q = 1$ and $\rho = 0.99$.

Table 2: N_{opt} for the SA HMM-UFIR/PMM-UFIR applied to drift estimation with $Q = 1$ and $\rho = 0.99$.

R	1	2	3	4	5	6	7	8	9	10
N_{opt}	13	18	22	25	28	31	33	36	38	40

higher the number of observations needed for noise averaging in the PMM-UFIR estimate. Fig. 2 plots the root MSE (RMSE) for drift estimation using the aforementioned algorithms as a function of R. Comparing all considered algorithms again, conclusions similar to the setting in Fig. 1 can be drawn. As expected, increasing the variance of the measurement noise R induces a moderate increase in the suboptimality of the PMM-UFIR wrt the SA HMM-KF/PMM-KF estimate with perfect knowledge of the noise variances ($p = 1$, $q = 1$). Note also, that the proposed PMM-UFIR outperforms the standard SA HMM-UFIR over the entire range of R.

5.2. Tracking with Colored Measurement Noise

Let us modify the discrete-time linear stochastic system (2), so that the white measurement noise is replaced by a Markovian measurement noise $\boldsymbol{\eta}_n$ [37], [46]

$$\begin{aligned} \mathbf{x}_n &= \mathbf{F}_n \mathbf{x}_{n-1} + \mathbf{B}_n \mathbf{w}_n \\ \boldsymbol{\eta}_n &= \boldsymbol{\Psi}_n \boldsymbol{\eta}_{n-1} + \mathbf{v}_n, \\ \mathbf{y}_n &= \mathbf{H}_n \mathbf{x}_n + \boldsymbol{\eta}_n, \end{aligned} \tag{38}$$

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 $\mathbf{z}_n = [\mathbf{w}_n^T, \mathbf{v}_n^T]^T$, $\forall n \geq 0$. The noise covariance is defined by $\mathbf{Q}_n = E\{\mathbf{w}_n \mathbf{w}_n^T\}$, $\mathbf{R}_n = E\{\mathbf{v}_n \mathbf{v}_n^T\}$ and $\mathbf{M}_n = \mathbf{0}$. The standard way to convert (38) back to a linear Gaussian HMM without state-augmentation, consists in

applying noise whitening by introducing the time-differenced measurement [46]

$$\begin{aligned}\mathbf{y}'_n &= \mathbf{y}_n - \mathbf{\Psi}_n \mathbf{y}_{n-1} \\ &= (\mathbf{H}_n \mathbf{F}_n - \mathbf{\Psi}_n \mathbf{H}_{n-1})\mathbf{x}_{n-1} + \mathbf{H}_n \mathbf{B}_n \mathbf{w}_n + \mathbf{v}_n,\end{aligned} \tag{39}$$

where the auxiliary measurement $\mathbf{y}'_n$ now depends on $\mathbf{x}_{n-1}$. Note that in order to take into account this lag by one time step in the time-differenced measurements, the HMM-KF needs to be generalized [46, Sec. IV and Appendix], [37, p. 191-192], [3, p. 329] and a similar generalization for the HMM-UFIR would also be needed. Therefore, in order to circumvent this problem, we introduce a general solution based on a PMM. Considering (39), it follows that (38) is equivalent to the linear Gaussian PMM (1) with the following parameters

$$\begin{aligned}\begin{bmatrix} \mathbf{A}_n^{(1)} & \mathbf{A}_n^{(2)} \\ \mathbf{A}_n^{(3)} & \mathbf{A}_n^{(4)} \end{bmatrix} &= \begin{bmatrix} \mathbf{F}_n & \mathbf{0} \\ \mathbf{H}_n \mathbf{F}_n - \mathbf{\Psi}_n \mathbf{H}_{n-1} & \mathbf{\Psi}_n \end{bmatrix}, \\ \begin{bmatrix} \mathbf{B}_n^{(1)} & \mathbf{B}_n^{(2)} \\ \mathbf{B}_n^{(3)} & \mathbf{B}_n^{(4)} \end{bmatrix} &= \begin{bmatrix} \mathbf{B}_n & \mathbf{0} \\ \mathbf{H}_n \mathbf{B}_n & \mathbf{I} \end{bmatrix},\end{aligned} \tag{40}$$

and the PMM white driving noise distribution is given by

$$\begin{bmatrix} \mathbf{w}_n \\ \mathbf{v}_n \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} \mathbf{0} \\ \mathbf{0} \end{bmatrix}, \begin{bmatrix} \mathbf{Q}_n & \mathbf{0} \\ \mathbf{0} & \mathbf{R}_n \end{bmatrix} \right).$$

Interestingly, it follows that both the PMM-KF in Algorithm 1 and the PMM-UFIR in Algorithm 2 are directly applicable on the raw measurements (i.e. without time-differencing), for the sake of state estimation.

Finding efficient estimation methods in the presence of colored measurement noise has been a long-standing problem in tracking [3] and navigation [47]. As an example, we consider the following discrete-time kinematic model for target tracking, where the state vector at instant n contains the position, velocity and acceleration, $\mathbf{x}_n = [p_n, v_n, a_n]^T$. Let us adopt a classical Wiener process acceleration model [3, p. 274], so that the first line in (38) is parameterized by

$$\mathbf{F}_n = \begin{bmatrix} 1 & T & \frac{T^2}{2} \\ 0 & 1 & T \\ 0 & 0 & 1 \end{bmatrix}, \mathbf{B}_n = \begin{bmatrix} \frac{T^2}{2} \\ T \\ 1 \end{bmatrix}, \mathbf{w}_n \sim \mathcal{N}(0, Q), \tag{41}$$

where T is the sampling period and the physical meaning of $\mathbf{w}_n$ is that of a scalar white acceleration increment. Assume that only position measurements are available [3, p. 277] under AR(1) colored noise at instant n, then the two last lines in (38) are parameterized by

$$\mathbf{\Psi}_n = \Psi, \mathbf{R}_n = R, \mathbf{H}_n = [1, 0, 0], \tag{42}$$

where $0 < \Psi < 1$ and $R > 0$ are the measurement noise coloration parameter and the measurement driving noise variance, respectively.

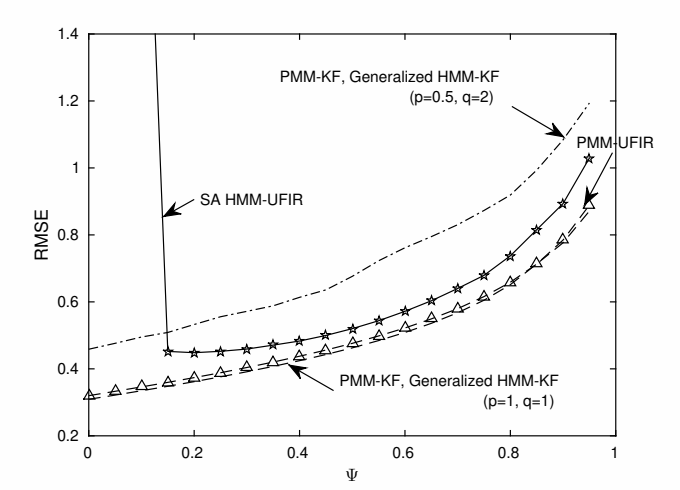

Figure 3: RMSE in tracking as a function of Ψ, with $T = 0.05$ s, $Q = 1$ $(m/s^2)^2$ and $R = 20^2$ m^2.

Table 3: N_{opt} for the PMM-UFIR applied to tracking with $T = 0.05$ s, $Q = 1$ $(m/s^2)^2$ and $R = 20^2$ m^2.

Ψ	0	0.05	0.1	0.15	0.2	0.25	0.3	0.35	0.4	0.45
N_{opt}	98	99	101	103	105	108	110	113	116	120
Ψ	0.5	0.55	0.6	0.65	0.7	0.75	0.8	0.85	0.9	0.95
N_{opt}	124	128	134	140	148	157	170	187	215	273

Table 4: N_{opt} for the PMM-UFIR applied to tracking with $T = 0.05$ s, $\Psi = 0.5$ and $R = 20^2$ m^2.

Q	1	2	3	4	5	6	7	8	9	10
N_{opt}	124	110	102	98	94	91	89	87	85	83

Let us describe our simulation setup for $T = 0.05$ s. The exact values of the true noise variances are selected as $Q = 1$ $(m/s^2)^2$ and $R = 20^2$ m^2 and used to simulate the trajectories of (38) using the parameters (41) and (42). Again, we assume that the generalized HMM-KF [46, Sec. IV and Appendix], [37, p. 191-192], [3, p. 329] and the PMM-KF algorithms have perfect knowledge of the initial condition, but employ p^2Q and q^2R instead of Q and R, with $p > 0$ and $q > 0$ in order to take noise covariance errors into account. The PMM-UFIR uses the optimal horizon size (33), reported in Table III as a function of the measurement noise coloration parameter Ψ. As expected, the higher the measurement noise correlation, the larger the horizon size needed to get a meaningful PMM-UFIR estimate. We also consider the SA HMM-UFIR with optimized horizon size for reference.

We adopt as a performance measure the RMSE of the dimensionless state vector $\tilde{\mathbf{x}}_n = [p_n, Tv_n, T^2a_n]^T/\sigma_m$ introduced in [48], where $\sigma_m^2 = R/(1-\Psi^2)$ is the measurement noise variance.

Fig. 3 plots the RMSE for the dimensionless state vector in the tracking problem as a function of Ψ, using the PMM-KF and the PMM-UFIR on raw measurements or the generalized HMM-KF on time-differenced

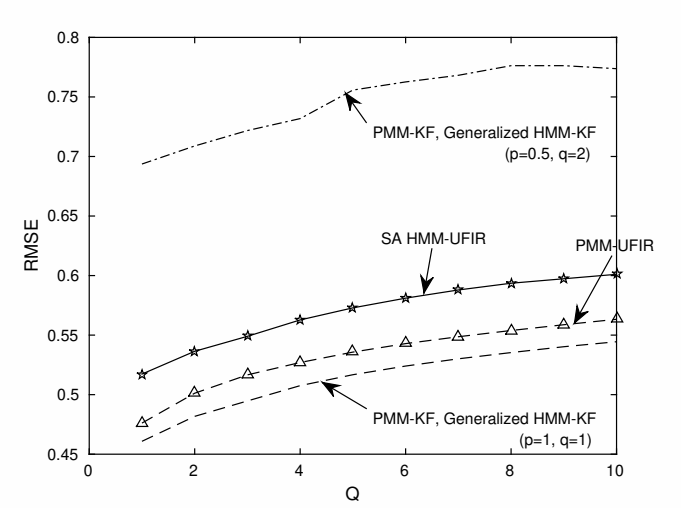

Figure 4: RMSE in tracking as a function of Q in $(m/s^2)^2$, with $T = 0.05$ s, $\Psi = 0.5$ and $R = 20^2$ m^2.

measurements. We observe that the PMM-KF reaches the same performances as the generalized HMM-KF, which is due to the fact that they both aim to minimize the MMSE criterion. Also, since the PMM-KF and the generalized HMM-KF are exempt of state-augmentation, both filters have the same dimensionality wrt the state and observation vectors, and thus have similar computational complexity. We note that by setting the horizon size to its optimal value, the RMSE of the PMM-UFIR is close the the RMSE of the PMM-KF with perfect knowledge of the noise variances (i.e. $p = q = 1$). Again, the PMM-UFIR is insensitive to noise variance errors, while the generalized HMM-KF and the PMM-KF are not, as exemplified by their deteriorated performances for $p = 0.5$ and $q = 2$. Interestingly, the proposed PMM-UFIR outperforms the SA HMM-UFIR with optimized horizon, over the entire range of Ψ. The SA HMM-UFIR even becomes numerically unstable for small values of Ψ, because the state transition matrix for in the SA HMM (given by the first line of (40)) becomes ill-conditioned.

In order to study the effect of the process noise variance, we now set $\Psi = 0.5$, $R = 20^2$ m^2 and the range of the parameter $Q \in [1, 10]$ $(m/s^2)^2$. The optimal horizon size for the PMM-UFIR is recalculated and shown in Table IV. The respective performance of all considered algorithms are shown in Fig. 4. Increasing the variance of the process noise Q induces a very slight increase in the suboptimality of the PMM-UFIR wrt the generalized HMM-KF/PMM-KF estimate with perfect knowledge of the noise variances ($p = 1$, $q = 1$). Again the proposed PMM-UFIR outperforms the standard SA HMM-UFIR over the entire range of Q.

5.3. Stochastic Volatility

Stochastic volatility models are widely used for analyzing financial time series. An example of such models is provided in [49]-[50] as

$$\begin{aligned} r_n &= \sigma \epsilon_n e^{h_n/2} \\ h_{n+1} &= \phi h_n + \eta_n, \end{aligned} \tag{43}$$

where r_n is the observed mean adjusted return on an asset and the hidden variable h_n is such that $V(r_n|h_n) = \sigma^2 e^{h_n}$, at time instant n. The disturbances (ϵ_n, η_n) are independently distributed according to [50]

$$\begin{bmatrix} \epsilon_n \\ \eta_n \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & \rho\sigma_\eta \\ \rho\sigma_\eta & \sigma_\eta^2 \end{bmatrix} \right). \tag{44}$$

By squaring the observations in (43) and taking logarithms,

$$\log(r_n^2) = \omega + h_n + \zeta_n,$$

where $\omega = \log(\sigma^2) + E[\log(\epsilon_n^2)]$ and $\zeta_n \sim \mathrm{ID}(0, \sigma_\zeta^2)$. Using (44), $\omega = \log(\sigma^2) - \log(2) - \gamma$ and $\sigma_\zeta^2 = \pi^2/2$, where γ is the Euler constant.

Let $s_n = \mathrm{sign}(r_n) \in \{-1, +1\}$ denote the sign of the n-th observation, we introduce the quantities $\mu^* = E[\eta_n | s_n = 1]$, and $\gamma^* = E[\eta_n \zeta_n | s_n = 1]$. From (44)

$$\begin{aligned} \mu^* &= \frac{\sqrt{2}}{\sqrt{\pi}} \rho \sigma_\eta \\ \gamma^* &= \frac{\sqrt{2}}{\sqrt{\pi}} \rho \sigma_\eta \left(\frac{\gamma}{2} + \frac{3}{2}\log(2) \right). \end{aligned}$$

Considering the state $\mathbf{x}_n = [h_n, \mu^*]^T$, and the transformed observations $y_n = \log(r_n^2) - \omega$, a HMM conditional on the signs of the original observations was obtained in [49, Eq. (6)] as

$$\begin{aligned} \mathbf{x}_n &= \mathbf{F}_n(s_{n-1})\mathbf{x}_{n-1} + \mathbf{B}_n w_n \\ y_n &= \mathbf{H}_n \mathbf{x}_n + \mathbf{D}_n v_n, \end{aligned} \tag{45}$$

parameterized by

$$\mathbf{F}_n(s_{n-1}) = \begin{bmatrix} \phi & s_{n-1} \\ 0 & 1 \end{bmatrix}, \mathbf{B}_n = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \mathbf{H}_n = [1, 0], \mathbf{D}_n = [1]. \tag{46}$$

and the non-Gaussian disturbance verifies

$$\begin{bmatrix} w_{n+1} \\ v_n \end{bmatrix} \Big| s_n \sim \mathrm{ID}\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma_\eta^2 - (\mu^*)^2 & \gamma^* s_n \\ \gamma^* s_n & \sigma_\zeta^2 \end{bmatrix} \right). \tag{47}$$

The best minimum mean square error linear estimator is obtained by applying the noise decorrelating Kalman filter in [3, p. 324-325], but needs accurate knowledge of the disturbance covariance. The HMM-UFIR on the contrary, relies only on the hypothesis that the non-Gaussian disturbance conditional on the original observation sign is zero-mean.

Table 5: N_{opt} for the HMM-UFIR applied to the stochastic volatility model with $\sigma = 1$, $\rho = -0.9$ and $\sigma_\eta = 2\sqrt{1-\phi^2}$.

ϕ	0.80	0.81	0.82	0.83	0.84	0.85	0.86	0.87	0.88	0.89
N_{opt}	16	15	13	10	11	11	11	11	13	13
ρ	0.90	0.91	0.92	0.93	0.94	0.95	0.96	0.97	0.98	0.99
N_{opt}	13	14	14	15	17	17	21	24	26	41

Table 6: N_{opt} for the PMM-UFIR applied to the stochastic volatility model with $\sigma = 1$, $\rho = -0.9$ and $\sigma_\eta = 2\sqrt{1-\phi^2}$.

ϕ	0.80	0.81	0.82	0.83	0.84	0.85	0.86	0.87	0.88	0.89
N_{opt}	54	52	49	45	43	42	40	36	30	29
ρ	0.90	0.91	0.92	0.93	0.94	0.95	0.96	0.97	0.98	0.99
N_{opt}	25	22	20	21	21	23	25	31	37	56

Interestingly, (43) is also naturally modeled by a PMM conditional on the signs of the original observations in [49, Eq. (7)]

$$\begin{bmatrix} \mathbf{x}_n \\ y_n \end{bmatrix} = \begin{bmatrix} \mathbf{A}_n^{(1)} & \mathbf{A}_n^{(2)} \\ \mathbf{A}_n^{(3)} & \mathbf{A}_n^{(4)} \end{bmatrix} (s_{n-1}) \begin{bmatrix} \mathbf{x}_{n-1} \\ y_{n-1} \end{bmatrix} + \begin{bmatrix} \mathbf{B}_n^{(1)} & \mathbf{B}_n^{(2)} \\ \mathbf{B}_n^{(3)} & \mathbf{B}_n^{(4)} \end{bmatrix} \begin{bmatrix} \tilde{w}_n \\ \tilde{v}_n \end{bmatrix}, \tag{48}$$

parameterized by

$$\begin{bmatrix} \mathbf{A}_n^{(1)} & \mathbf{A}_n^{(2)} \\ \mathbf{A}_n^{(3)} & \mathbf{A}_n^{(4)} \end{bmatrix} (s_{n-1}) = \left(\begin{array}{cc|c} \phi - \frac{\gamma^* s_{n-1}}{\sigma_\zeta^2} & s_{n-1} & \frac{\gamma^* s_{n-1}}{\sigma_\zeta^2} \\ 0 & 1 & 0 \\ \hline \phi - \frac{\gamma^* s_{n-1}}{\sigma_\zeta^2} & s_{n-1} & \frac{\gamma^* s_{n-1}}{\sigma_\zeta^2} \end{array}\right),$$
$$\begin{bmatrix} \mathbf{B}_n^{(1)} & \mathbf{B}_n^{(2)} \\ \mathbf{B}_n^{(3)} & \mathbf{B}_n^{(4)} \end{bmatrix} = \left(\begin{array}{c|c} 1 & 0 \\ 0 & 0 \\ \hline 1 & 1 \end{array}\right), \tag{49}$$

and the non-Gaussian disturbance verifies

$$\begin{bmatrix} \tilde{w}_n \\ \tilde{v}_n \end{bmatrix} \Big| s_{n-1} \sim \text{ID}\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma_\eta^2 - (\mu^*)^2 - (\frac{\gamma^*}{\sigma_\zeta})^2 & 0 \\ 0 & \sigma_\zeta^2 \end{bmatrix} \right). \tag{50}$$

The best minimum mean square error linear estimator is obtained by applying the PMM-KF, which again needs accurate knowledge of the disturbance covariance. The PMM-UFIR on the contrary, relies only on the hypothesis that the non-Gaussian disturbance conditional on the original observation sign is zero-mean. Note that the standard SA HMM-UFIR [34] is not even applicable, since the complete PMM system matrix in the first line of (49) is singular.

Let us describe our simulation setup. We simulate trajectories of (43), with $\sigma = 1$, $\rho = -0.9$. For any $\phi \in [0.8, 0.99]$, $\sigma_\eta = 2\sqrt{1-\phi^2}$ in order to normalize $E[h_n^2]$ to 4. The optimal horizon for the HMM-UFIR applied to (45) and the PMM-UFIR applied to (48) are given in Tab. V and VI, respectively.

Fig. 5 plots the RMSE for the hidden variable h_n in the stochastic volatility problem as a function of

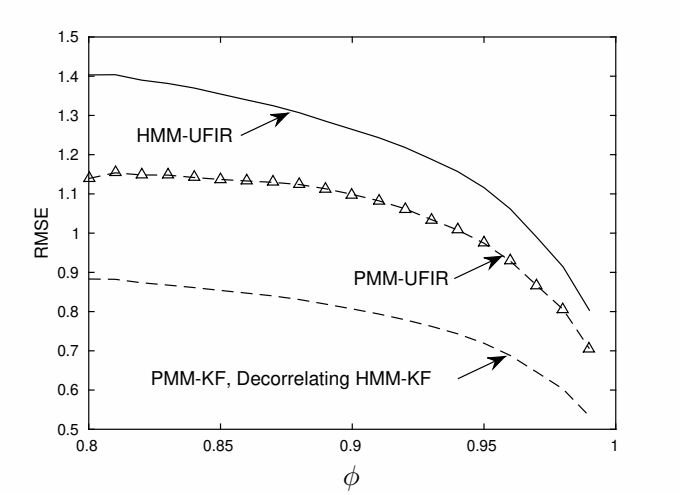

Figure 5: RMSE in the stochastic volatility problem as a function of ϕ, with $\sigma = 1$, $\rho = -0.9$ and $\sigma_\eta = 2\sqrt{1-\phi^2}$.

ϕ, for the decorrelating HMM-KF and the PMM-KF with perfect knowledge of the disturbance covariance. We observe that the PMM-KF reaches the same performances as the decorrelating HMM-KF, which is due to the fact that they both correspond to the best minimum mean square error linear estimator. In the absence of information about the disturbance covariance, the HMM-UFIR and the PMM-UFIR are interesting alternatives. The PMM-UFIR benefits from a higher optimal horizon and thus outperforms its HMM-UFIR couterpart over the entire range of ϕ, by allowing better noise averaging. Note also, that the performance degradation of the PMM-UFIR wrt the Kalman filters is not severe.

6. Conclusion

This paper considered the estimation problem in linear Gaussian pairwise Markov models (PMMs) for real-world applications in engineering, natural sciences or econometrics, where the initial state and/or the noise statistics are either unknown or difficult to estimate. The proposed solution has the form of a FIR estimator (PMM-UFIR). Since the implementation of the latter is typically too complex in batch form, we have derived a recursive form that is both computationally efficient and in intuitive Kalman-like form. The performance in terms of MSE of the proposed algorithm was analyzed and we also recovered the standard UFIR for HMMs as a special case (with a slight difference in the horizon size, though).

Numerical results gave valuable insights about the properties of the considered approach. In particular, the proposed UFIR is close to the MMSE-optimal Kalman filter (PMM-KF) in terms of estimation accuracy in a linear Gaussian PMM, provided that the horizon size is tuned to its optimal value. The considered examples have confirmed the improved robustness of the PMM-UFIR wrt the PMM-KF. Moreover, compar-

ing to the standard SA HMM-UFIR approach, our numerical results have demonstrated that the proposed PMM-UFIR not only has better performances over a wide range of applications, but also that it can be applicable in situations were the standard method fails completely.

Our results can be extended in several directions. For instance, the extended PMM used to derive the PMM-UFIR can be generalized to other structures depending on the retained matrix invertibility hypothesis. This may lead to other versions of the PMM-UFIR, that may perform even closer to the ideal PMM-KF in some applications, while retaining the inbuilt robustness properties. As another extension, tuning adaptively the horizon size to the operating conditions by controling the quality of estimation based on measurements only, would be worth investigating [51].

A. Derivation of the PMM-KF

The predictive distribution can be written as

$$\begin{aligned} p(\mathbf{x}_n, \mathbf{y}_n|\mathbf{y}_{0:n-1}) &= \int p(\mathbf{x}_n, \mathbf{y}_n, \mathbf{x}_{n-1}|\mathbf{y}_{0:n-1}) d\mathbf{x}_{n-1} \\ &= \int p(\mathbf{x}_n, \mathbf{y}_n|\mathbf{x}_{n-1}, \mathbf{y}_{n-1}) p(\mathbf{x}_{n-1}|\mathbf{y}_{0:n-1}) d\mathbf{x}_{n-1}. \end{aligned} \tag{A.1}$$

Now (1) implies,

$$p(\mathbf{x}_n, \mathbf{y}_n|\mathbf{x}_{n-1}, \mathbf{y}_{n-1}) = \\ \mathcal{N}\left(\begin{bmatrix} \mathbf{A}_n^{(1)} & \mathbf{A}_n^{(2)} \\ \mathbf{A}_n^{(3)} & \mathbf{A}_n^{(4)} \end{bmatrix} \begin{bmatrix} \mathbf{x}_{n-1} \\ \mathbf{y}_{n-1} \end{bmatrix}, \begin{bmatrix} \mathbf{B}_n^{(1)} & \mathbf{B}_n^{(2)} \\ \mathbf{B}_n^{(3)} & \mathbf{B}_n^{(4)} \end{bmatrix} \begin{bmatrix} \mathbf{Q}_n & \mathbf{M}_n \\ \mathbf{M}_n^T & \mathbf{R}_n \end{bmatrix} \begin{bmatrix} \mathbf{B}_n^{(1)} & \mathbf{B}_n^{(2)} \\ \mathbf{B}_n^{(3)} & \mathbf{B}_n^{(4)} \end{bmatrix}^T \right).$$

Since at instant $n-1$, the posterior distribution has the form

$$p(\mathbf{x}_{n-1}|\mathbf{y}_{0:n-1}) = \mathcal{N}\left(\hat{\mathbf{x}}_{n-1}, \mathbf{P}_{n-1}\right),$$

(A.1) is also a Gaussian distribution whose mean is given by

$$\begin{bmatrix} \hat{\mathbf{x}}_n^- \\ \hat{\mathbf{y}}_n^- \end{bmatrix} = \begin{bmatrix} \mathbf{A}_n^{(1)} & \mathbf{A}_n^{(2)} \\ \mathbf{A}_n^{(3)} & \mathbf{A}_n^{(4)} \end{bmatrix} \begin{bmatrix} E\{\mathbf{x}_{n-1}|\mathbf{y}_{0:n-1}\} \\ \mathbf{y}_{n-1} \end{bmatrix} = \begin{bmatrix} \mathbf{A}_n^{(1)} & \mathbf{A}_n^{(2)} \\ \mathbf{A}_n^{(3)} & \mathbf{A}_n^{(4)} \end{bmatrix} \begin{bmatrix} \hat{\mathbf{x}}_{n-1} \\ \mathbf{y}_{n-1} \end{bmatrix},$$

which is the desired result. Similarly, the covariance of the Gaussian (A.1) is obtained as

$$\begin{bmatrix} \mathbf{P}_n^- & \mathbf{\Sigma}_n^- \\ \mathbf{\Sigma}_n^{-T} & \mathbf{L}_n^- \end{bmatrix} = E\left\{ \begin{bmatrix} \mathbf{x}_n - \hat{\mathbf{x}}_n^- \\ \mathbf{y}_n - \hat{\mathbf{y}}_n^- \end{bmatrix} \begin{bmatrix} \mathbf{x}_n - \hat{\mathbf{x}}_n^- \\ \mathbf{y}_n - \hat{\mathbf{y}}_n^- \end{bmatrix}^T |\mathbf{y}_{0:n-1} \right\}.$$

Again (1) implies

$$\begin{bmatrix} \mathbf{x}_n - \hat{\mathbf{x}}_n^- \\ \mathbf{y}_n - \hat{\mathbf{y}}_n^- \end{bmatrix} = \begin{bmatrix} \mathbf{A}_n^{(1)} & \mathbf{A}_n^{(2)} \\ \mathbf{A}_n^{(3)} & \mathbf{A}_n^{(4)} \end{bmatrix} \begin{bmatrix} \mathbf{x}_{n-1} - \hat{\mathbf{x}}_{n-1} \\ \mathbf{0} \end{bmatrix} + \begin{bmatrix} \mathbf{B}_n^{(1)} & \mathbf{B}_n^{(2)} \\ \mathbf{B}_n^{(3)} & \mathbf{B}_n^{(4)} \end{bmatrix} \begin{bmatrix} \mathbf{w}_n \\ \mathbf{v}_n \end{bmatrix}.$$

Now the fact that $\mathbf{z}_n = [\mathbf{w}_n^T, \mathbf{v}_n^T]$ is white, yields

$$\begin{bmatrix} \mathbf{P}_n^- & \mathbf{\Sigma}_n^- \\ \mathbf{\Sigma}_n^{-T} & \mathbf{L}_n^- \end{bmatrix} = \begin{bmatrix} \mathbf{A}_n^{(1)} & \mathbf{A}_n^{(2)} \\ \mathbf{A}_n^{(3)} & \mathbf{A}_n^{(4)} \end{bmatrix} \begin{bmatrix} \mathbf{P}_{n-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \mathbf{A}_n^{(1)} & \mathbf{A}_n^{(2)} \\ \mathbf{A}_n^{(3)} & \mathbf{A}_n^{(4)} \end{bmatrix}^T \\ + \begin{bmatrix} \mathbf{B}_n^{(1)} & \mathbf{B}_n^{(2)} \\ \mathbf{B}_n^{(3)} & \mathbf{B}_n^{(4)} \end{bmatrix} \begin{bmatrix} \mathbf{Q}_n & \mathbf{M}_n \\ \mathbf{M}_n^T & \mathbf{R}_n \end{bmatrix} \begin{bmatrix} \mathbf{B}_n^{(1)} & \mathbf{B}_n^{(2)} \\ \mathbf{B}_n^{(3)} & \mathbf{B}_n^{(4)} \end{bmatrix}^T,$$

which is the result in Algorithm 1 after expansion.

Finally, obtaining the posterior distribution at instant n

$$p(\mathbf{x}_n|\mathbf{y}_{0:n}) = \mathcal{N}\left(\hat{\mathbf{x}}_n, \mathbf{P}_n\right),$$

with the mean and covariance in Algorithm 1, proceeds from a straightforward application of the Gaussian conditioning formula [3, p. 52-54] to (A.1).

B. Derivation of the PMM-UFIR in iterative Kalman-like form

Let us rewrite $\mathbf{H}_{n,m+1}$ in (15) in block matrix form using (10)

$$\mathbf{H}_{n,m+1}=\left[\begin{array}{c}\mathbf{A}_n^{(3)}\\ \hline \\ \mathbf{H}_{n-1,m+1}\end{array}\right](\mathbf{A}_n^{(1)})^{-1}, \tag{B.1}$$

it follows that the inverse of the GNPG (20) can be written as

$$\begin{aligned}\mathbf{G}_n^{-1}&=\mathbf{H}_{n,m+1}^T\mathbf{H}_{n,m+1}\\ &=(\mathbf{A}_n^{(1)})^{-T}[\mathbf{A}_n^{(3)T}\mathbf{A}_n^{(3)}+\mathbf{G}_{n-1}^{-1}](\mathbf{A}_n^{(1)})^{-1}.\end{aligned} \tag{B.2}$$

Taking the inverse of the previous expression proves the recursion (23).

Now, isolating the first line of $\mathbf{J}_{n,m+1}$ in (14) and $\mathbf{J}'_{n,m+1}$ in (15), we obtain

$$\begin{aligned}&(\mathbf{J}_{n,m+1}+\mathbf{J}'_{n,m+1})\mathbf{Y}_{n,m}=\\ &\left[\begin{array}{c}\mathbf{y}_n-\mathbf{A}_n^{(4)}\mathbf{y}_{n-1}+\mathbf{A}_n^{(3)}(\mathbf{A}_n^{(1)})^{-1}\mathbf{A}_n^{(2)}\mathbf{y}_{n-1}\\ \hline \\ (\mathbf{J}_{n-1,m+1}+\mathbf{J}'_{n-1,m+1})\mathbf{Y}_{n-1,m}+\mathbf{H}_{n-1,m+1}(\mathbf{A}_n^{(1)})^{-1}\mathbf{A}_n^{(2)}\mathbf{y}_{n-1}\end{array}\right]\end{aligned} \tag{B.3}$$

From the definition of the PMM-UFIR estimate in batch form (16) and the GNPG (20), we have

$$\mathbf{G}_n^{-1}\hat{\mathbf{x}}_n=\mathbf{H}_{n,m+1}^T(\mathbf{J}_{n,m+1}+\mathbf{J}'_{n,m+1})\mathbf{Y}_{n,m}. \tag{B.4}$$

Now, injecting (B.1) and (B.3) into (B.4) yields

$$\begin{aligned}&{\mathbf{G}_n}^{-1}\hat{\mathbf{x}}_n\\ &=(\mathbf{A}_n^{(1)})^{-T}\mathbf{A}_n^{(3)T}(\mathbf{y}_n-\mathbf{A}_n^{(4)}\mathbf{y}_{n-1}+\mathbf{A}_n^{(3)}(\mathbf{A}_n^{(1)})^{-1}\mathbf{A}_n^{(2)}\mathbf{y}_{n-1})+\\ &(\mathbf{A}_n^{(1)})^{-T}\mathbf{H}_{n-1,m+1}^T(\mathbf{J}_{n-1,m+1}+\mathbf{J}'_{n-1,m+1})\mathbf{Y}_{n-1,m}+\\ &(\mathbf{A}_n^{(1)})^{-T}\mathbf{H}_{n-1,m+1}^T\mathbf{H}_{n-1,m+1}(\mathbf{A}_n^{(1)})^{-1}\mathbf{A}_n^{(2)}\mathbf{y}_{n-1}.\end{aligned} \tag{B.5}$$

Using the fact that by definition $\mathbf{G}_{n-1}^{-1}\hat{\mathbf{x}}_{n-1}=\mathbf{H}_{n-1,m+1}^T(\mathbf{J}_{n-1,m+1}+\mathbf{J}'_{n-1,m+1})\mathbf{Y}_{n-1,m}$, (B.5) can be rearranged as

$$\begin{aligned}&{\mathbf{G}_n}^{-1}\hat{\mathbf{x}}_n\\ &=(\mathbf{A}_n^{(1)})^{-T}\mathbf{A}_n^{(3)T}(\mathbf{y}_n-\mathbf{A}_n^{(3)}\hat{\mathbf{x}}_{n-1}-\mathbf{A}_n^{(4)}\mathbf{y}_{n-1})+\\ &(\mathbf{A}_n^{(1)})^{-T}[\mathbf{A}_n^{(3)T}\mathbf{A}_n^{(3)}+\mathbf{G}_{n-1}^{-1}]\hat{\mathbf{x}}_{n-1}+\\ &(\mathbf{A}_n^{(1)})^{-T}[\mathbf{A}_n^{(3)T}\mathbf{A}_n^{(3)}+\mathbf{G}_{n-1}^{-1}](\mathbf{A}_n^{(1)})^{-1}\mathbf{A}_n^{(2)}\mathbf{y}_{n-1},\end{aligned} \tag{B.6}$$

which in turn becomes

$$\begin{aligned}&{\mathbf{G}_n}^{-1}\hat{\mathbf{x}}_n\\ &=(\mathbf{A}_n^{(1)})^{-T}\mathbf{A}_n^{(3)T}(\mathbf{y}_n-\mathbf{A}_n^{(3)}\hat{\mathbf{x}}_{n-1}-\mathbf{A}_n^{(4)}\mathbf{y}_{n-1})+\\ &\mathbf{G}_n^{-1}(\mathbf{A}_n^{(1)}\hat{\mathbf{x}}_{n-1}+\mathbf{A}_n^{(2)}\mathbf{y}_{n-1})\end{aligned} \tag{B.7}$$

using (B.2). Finally, multiplying both sides of (B.7) by $\mathbf{G}_n$ completes the proof of (21)-(25).

References

[1] S. M. Kay, Fundamentals of statistical signal processing: Estimation theory, Upper Saddle River, NJ, Prentice-Hall, 1993.

[2] R.E. Kalman, A new approach to linear filtering and prediction problems, Trans. of the ASME, Journal of Basic Engineering 82 (1960) 35-45.

[3] Y. Bar-Shalom, X.-R. Li, T. Kirubarajan, Estimation With Applications to Tracking and Navigation, New York, Wiley, 2001.

[4] R.F. Stengel, Stochastic Optimal Control: Theory and Application, New York, Wiley, 1986.

[5] J.D. Hamilton, Time Series Analysis, Princeton, NJ, Princeton University Press, 1994.

[6] Q. Guo, L. Ping, LMMSE turbo equalization based on factor graphs, IEEE J. Sel. Areas Commun. 26 (3) (2008) 311-319.

[7] D. Cherkassky, S. Gannot, New insights into the Kalman filter beamformer: Applications to speech and robustness, IEEE Sig. Proc. Lett. 23 (3) (2016) 376-380.

[8] V.N. Ioannidis, D. Romero, G.B. Giannakis, Inference of spatio-temporal functions over graphs via multikernel kriged Kalman filtering, IEEE Trans. Signal Process. 66 (12) (2018) 3228-3239.

[9] M. Weiss, M.S. Wiederoder, R.C. Paffenroth, E.C. Nallon, C.J. Bright, V.P. Schnee, S. McGraw, M. Polcha, J.R. Uzarski, Applications of the Kalman filter to chemical sensors for downstream machine learning, IEEE Sensors Journal 18 (13) (2018) 5455-5463.

[10] D.R. Cox, Statistical analysis of time series: some recent developments, Scand. J. Statist. 8 (2) (1981) 93-115.

[11] Y. Li, Hidden Markov models with states depending on observations, Pattern Recognition Lett. 26 (7) (2005) 977-984.

[12] S. Nassar, K.P. Schwarz, N. El-Sheimy, A. Noureldin, Modeling inertial sensor errors using autoregressive (AR) models, Journal of the Institute of Navigation 51 (4) (2004) 259-268.

[13] L. Ljung, System Identification - Theory For the User, 2nd ed, Upper Saddle River, NJ, Prentice-Hall, 1999.

[14] W. Pieczynski, Pairwise Markov chain, IEEE Trans. on Pattern Analysis and Machine Intelligence 25 (5) (2003) 634-639.

[15] W. Pieczynski, F. Desbouvries, Kalman filtering using pairwise Gaussian models, in: Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (2003), Hong-Kong.

[16] V. Nemesin, S. Derrode, Robust partial-learning in linear Gaussian systems, IEEE Trans. Autom. Control 60 (9) (2015) 2518-2523.

[17] M.V. Kulikova, Gradient-based parameter estimation in pairwise linear Gaussian system, IEEE Trans. Autom. Control 62 (3) (2017) 1511-1517.

[18] E. Monfrini, J. Lecomte, F. Desbouvries, W. Pieczynski, Image and signal restoration using pairwise Markov trees, in: IEEE Workshop on Statistical Signal Processing (2003), St. Louis, MO.

[19] D. Benboudjema, M. El Bouchraya Malainin, W. Pieczynski, Exact Kalman filtering in pairwise Gaussian switching systems, in: Proc. Applied Stochastic Models and Data Analysis (2011), Rome, Italy.

[20] F. Desbouvries, J. Lecomte, W. Pieczynski, Kalman filtering in pairwise Markov trees, Signal Processing 86 (5) (2006) 1049-1054.

[21] Y. Petetin, F. Desbouvries, Bayesian multi-object filtering for pairwise Markov chains, IEEE Trans. Signal Process. 61 (18) (2013) 4481-4490.

[22] F. Septier, S.K. Pang, S. Godsill, A. Carmi, Tracking of coordinated groups using marginalised MCMC-based particle algorithm, in: IEEE Aerospace conference (2009), Big Sky, MT.

[23] R. Mahler, Tracking targets with pairwise-Markov dynamics, in: 18th International Conference on Information Fusion (2015), Washington, DC.

[24] H. Qian, A flexible state space model and its applications, Journal of Time Series Analysis 35 (2) (2014) 79-88.

[25] A.H. Jazwinski, Stochastic Processes and Filtering Theory, New York, NJ, Academic Press, 1970.

[26] M. Verhaegen, P. Van Dooren, Numerical aspects of different Kalman filter implementations, IEEE Trans. Autom. Control 31 (10) (1986) 907-917.

[27] A.H. Jazwinski, Limited memory optimal filtering, IEEE Trans. Autom. Control 13 (5) (1968) 558-563.

[28] W.H. Kwon, P.S. Kim, S.H. Han, A receding horizon unbiased FIR filter for discrete-time state space models, Automatica 38 (3) (2002) 545-551.

[29] Y.S. Shmaliy, Unbiased FIR filtering for discrete-time polynomial state-space models, IEEE Trans. Signal Process. 57 (4) (2009) 1241-1249.

[30] L.J. Morales-Mendoza, H. Gamboa-Rosales, Y.S. Shmaliy, A new class of discrete orthogonal polynomials for blind fitting of finite data, Signal Processing 93 (7) (2013) 1785-1793.

[31] B.H.S. Asli, J. Flusser, New discrete orthogonal moments for signal analysis, Signal Processing 141 (2017) 57-73.

[32] S. Zhao, Y.S. Shmaliy, F. Liu, Fast computation of discrete optimal FIR estimates in white Gaussian noise, IEEE Signal Process. Lett. 22 (6) (2015) 718-722.

[33] Y.S. Shmaliy, An iterative Kalman-like algorithm ignoring noise and initial conditions, IEEE Trans. Signal Process. 59 (6) (2011) 2465-2473.

[34] Y.S. Shmaliy, S. Zhao, C.K. Ahn, Unbiased finite impulse response filtering: an iterative alternative to Kalman filtering ignoring noise and initial conditions, IEEE Control Systems Magazine 37 (5) (2017) 70-89.

[35] F. Ramirez-Echeverria, A. Sarr, Y.S. Shmaliy, Optimal memory for discrete-time FIR filters in state-space, IEEE Trans. Signal Process. 62 (3) (2014) 557-561.

[36] Y.S. Shmaliy, F. Lehmann, S. Zhao, C.K. Ahn, Comparing robustness of the Kalman, H_∞, and UFIR filters, IEEE Trans. Signal Process. 66 (13) (2018) 3447-3458.

[37] D. Simon, Optimal State Estimation: Kalman, H_∞ and Nonlinear Approaches, Hoboken, NJ, J. Wiley & Sons, 2006.

[38] D. Di Ruscio, A method for identification of combined deterministic stochastic systems, in: Applications of Computer Aided Time Series Modeling, Lecture Notes in Statistics 119, Eds. M. Aoki and A. M. Havenner, New York, NJ, Springer, 1997.

[39] A. Gelb, Applied Optimal Estimation, Cambridge, MA, MIT Press, 2001.

[40] D.J. Luenberger, Optimization by Vector Space Methods, New York, Wiley, 1969.

[41] S. Zhao, Y.S. Shmaliy, F. Liu, Fast Kalman-like optimal unbiased FIR filtering with applications, IEEE Trans. Signal Process. 64 (9) (2016) 2284-2297.

[42] M. Grewal, A. Andrews, How good is your gyro [Ask the Experts], IEEE Control Systems Magazine 30 (1) (2010) 12-86.

[43] R.H. Shumway, D.S. Stoffer, Times Series Analysis and Its Applications, with R examples, New York, NJ, Springer, 2017.

[44] A.D. Pwasong, S. A\P. Sathasivam, Forecasting performance of random walk with drift and feed forward neural network models, International Journal of Intelligent Systems and Applications 7 (9) (2015) 49-56.

[45] M. Auger-Méthé, M.A. Lewis, A.E. Derocher, Home ranges in moving habitats: polar bears and sea ice, Ecography 39 (1) (2016) 26-35.

[46] A.E. Bryson, L.J. Henrikson, Estimation using sampled data containing sequentially correlated noise, J. Spacecraft Rockets 5 (6) (1968) 662-665.

[47] J. Wendel, G.F. Trommer, An efficient method for considering time correlated noise in GPS/INS integration, in: Proc. National Technical Meeting of The Institute of Navigation (2004), San Diego, CA.

[48] R.J. Fitzgerald, Simple tracking filters: steady-state filtering and smoothing performance, IEEE Trans. on Aerospace and Electronic Systems 16 (6) (1980) 860-864.

[49] A.C. Harvey, N. Shephard, Estimation of an asymmetric stochastic volatility model for asset returns, Journal of Business & Economic Statistics 14 (4) (1996) 429-434.

[50] G. Sandmann, S.J. Koopman, Estimation of stochastic volatility models via Monte Carlo maximum likelohood, J. Econometrics 87 (1998) 271-301.

[51] J.M. Pak, C.K. Ahn, Y.S. Shmaliy, P. Shi, M.T. Lim, Switching extensible FIR filter bank for adaptive horizon state estimation with applications, IEEE Trans. Control Syst. Technol. 24 (3) (2016) 1052-1058.