Reduced-Dimension Filtering in Triplet Markov Models

Frederic Lehmann and Wojciech Pieczynski

***Abstract*—This article presents an optimal reduced-dimension Kalman filter for a family of triplet Markov models (TMMs). The problem is to estimate the state vector in the case when the auxiliary process in the TMM can be eliminated. Sufficient conditions for this elimination to be feasible are established and we give a selection of illustrative real-life TMM examples, where these conditions are satisfied. We subsequently show that the original TMM boils down to a pairwise Markov model (PMM) of second order. Then, we derive a new optimal Kalman filter applicable to any linear PMM of second order. Our numerical results confirm that the proposed estimator can provide substantial complexity reduction with either no or minor accuracy loss, depending on the use of model approximation.**

***Index Terms*—Kalman filter (KF), optimal filtering, pairwise Markov models (PMMs), reduced-dimension filtering (RDF), triplet Markov models (TMMs).**

I. Introduction

DISCRETE-TIME linear state-space models are ubiquitous in signal processing and control theory. They are used extensively in many application fields such as speech enhancement [1], image processing [2], tracking [3] econometrics [4], communications [5], bioinformatics [6], machine learning [7], and control [8].

State-space models are available in several forms. The hidden Markov model (HMM), modeling the unobserved process via a Markov chain, has received considerable attention due to the first apparition of an optimal filter in the minimum mean square error (MMSE) sense, namely the Kalman filter (HMM-KF) [9]. Pairwise Markov models (PMMs) nicely generalize HMMs, by assuming that the couple formed by the hidden and observed processes is Markovian [10], thus offering the possibility of improved modeling capability. Also, in many applications, adding an auxiliary (or latent) process is mandatory to capture the complete dynamics of the system. Such an auxiliary process is useful in well-known situations where a state-space model is subject to potential nonstationarity, parameter uncertainty [11], or error sources [13]. These situations can be taken into account using triplet Markov models (TMMs), which further extend PMMs by assuming that the triplet formed by the hidden, auxiliary, and observed processes is Markovian [12]. Note that the class of TMMs considered in the present article was shown to be efficient [13] for various applications such as tracking [14, Ch. 5], [15], speech processing with colored noise [16], and time series analysis using conditionally Gaussian models [4, Sec. 3.7.1]. What makes linear Gaussian PMMs and TMMs appealing is that optimal filtering (in the MMSE sense) is still feasible, and the corresponding filters shall be referred to as the PMM Kalman filter (PMM-KF) [10], [17]–[19] and the TMM Kalman filter (TMM-KF) [12], [13], respectively. Therefore, PMMs and TMMs gradually generalize the standard HMM while retaining the essential feature of fast recursive MMSE-optimal state estimation, processing each new observation one at a time.

We consider the problem of developing a reduced-dimension filter (RDF) to estimate only a subset of the hidden variables. This issue is relevant in situations where only a part of the hidden variables; (1) are considered as useful information, whereas the other state partition contains disturbances that contribute to exact modeling, but merely act as nuisance parameters; (2) are needed, whose choice is dictated by its subsequent use (in a control law definition for instance), so that a full-dimension filter is unattractive in terms of computational complexity. Several RDFs have been proposed in the HMM literature, by postulating an HMM-KF estimator structure for the retained part of the state while rederiving the Kalman gains to take into account coupling with the remaining part of the state. This design philosophy leads to a number of suboptimal filters. In [21], a weighted quadratic error criterion integrated over a time interval of interest is minimized, which involves the solution of a complicated two-point boundary value problem. In [22]–[24], the trace of the error covariance is minimized at each time instant (irrespective of all minimizations performed at earlier time instants), yielding modified HMM-KFs. Another line of thought assumes that the state can be projected on a low-dimensional subspace, so that the original state can be written as a linear combination of the corresponding basis vectors and the coefficients become the new reduced-dimension state variable [25]–[27]. However, selecting the subspace's dimension may be tricky [25] and the dynamics of the system may not be properly captured, since the selected subspace is usually invariant over a time window [28].

Manuscript received December 7, 2020; accepted January 6, 2021. Date of publication January 11, 2021; date of current version January 28, 2022. Recommended by Associate Editor D. Efimov. *(Corresponding author: Frederic Lehmann.)*

The authors are with the SAMOVAR, Télécom SudParis, Institut Polytechnique de Paris, 91120 Palaiseau, France (e-mail: frederic.lehmann@telecom-sudparis.eu; wojciech.pieczynski@telecom-sudparis.eu).

Digital Object Identifier 10.1109/TAC.2021.3050721

In this article, we consider the problem of RDF for a family of linear TMMs, which has not been addressed previously to the best of our knowledge. Note that in the considered TMM framework, the aforementioned unobserved and auxiliary variables have a natural interpretation as the desired part and the uninteresting part of the state, respectively. We show that such a TMM simplifies to a PMM of second order, if the auxiliary variable can be eliminated by writing it as a function of the unobserved and observed variables over the current and previous time instant. We provide sufficient conditions for this elimination to hold true and we show that many well-known applications, which can be recast in the TMM framework, actually satisfy these conditions. Then, we derive an MMSE-optimal state estimator having the dimension of the unobserved state, which compares favorably in terms of complexity with the TMM-KF, whose dimension is the sum of the dimensions of the unobserved and auxiliary variable. The main contributions of this article are as follows.

1) The transformation of a linear TMM to a PMM of second order, under conditions that hold true over a wide range of applications for which TMMs have been found relevant.
2) The design of a new MMSE-optimal state estimator applicable to any PMM of second order, which can thus be used as a reduced-dimension estimator in TMMs (i.e., having the dimension of the desired state), allowing complexity savings without loss of optimality.
3) A theoretical error analysis of the proposed estimator.

Note that the related work in [29] mainly focuses on reduced-order modeling (i.e., approximating a TMM by a first-order PMM), whereas estimation consists of the suboptimal application of the standard PMM-KF. In contrast, the present article falls under the umbrella of RDFs, since the original TMM is first transformed (under condition) to a second-order PMM, for which an optimal filter is then designed. Also, from a practical point of view, paper [29] is restricted to time-homogeneous TMMs and the corresponding state estimator usually needs batch processing. These restrictions are lifted in the present approach.

Throughout the article, bold letters indicate vectors and matrices, whereas $\mathbf{0}_{m\times n}$ (resp., $\mathbf{I}_m$) is the $m \times n$ all-zero (resp., the $m \times m$ identity) matrix and $\mathrm{diag}(\mathbf{a})$ is the (block) diagonal matrix, whose diagonal entries are stored in $\mathbf{a}$ and whose off-diagonal entries are zero. The Frobenius norm of matrix $\mathbf{M}$ is denoted by $||\mathbf{M}||_F$. $\mathcal{N}(\mathbf{m}, \mathbf{C})$ denotes a Gaussian distribution with mean $\mathbf{m}$ and covariance matrix $\mathbf{C}$. A sequence of observations from time m up to time n is denoted by $\mathbf{y}_{m:n}$.

This article is organized as follows. First, Section II defines the linear TMM system model. Then, its transformation to a PMM of second order is introduced in Section III. The rationale behind this approach is that the conditions under which this transformation is valid can be met in various applications as shown in Section IV. Based on this transformation, Section V investigates the design of an RDF for state estimation. Then, Section VI tackles the problem of performance analysis, by providing a closed-form expression of the error covariance. Numerical simulations for realistic applications show the simplicity and effectiveness of the proposed reduced-dimension state estimator in Section VII. Finally, Section VIII concludes this article.

II. System Model

We assume that the dynamics of a (possibly non-time-homogeneous) discrete-time linear state-space model can be captured by a particular TMM with two distinct noise sources of the form

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

where $\mathbf{x}_n \in \mathbb{R}^K$, $\mathbf{r}_n \in \mathbb{R}^L$, and $\mathbf{y}_n \in \mathbb{R}^M$ denote the state, the auxiliary, and the observation vector at instant n, respectively. Concrete applications are described in Section IV so that (1) is a generalization thereof. The initial triplet $\mathbf{t}_0$ is independent from the zero-mean white uncorrelated noise processes $\mathbf{w}_n \in \mathbb{R}^K$, $\mathbf{v}_n \in \mathbb{R}^M$, for all $n \geq 1$. Let us set $\mathbf{m}_0^{(\mathrm{t})} = E\{\mathbf{t}_0\}$ and $\mathbf{P}_0^{(\mathrm{tt})} = E\{(\mathbf{t}_0 - \mathbf{m}_0^{(\mathrm{t})})(\mathbf{t}_0 - \mathbf{m}_0^{(\mathrm{t})})^T\}$. The noise covariance is defined by

$$E\left\{\begin{bmatrix}\mathbf{w}_n\\ \mathbf{v}_n\end{bmatrix}\begin{bmatrix}\mathbf{w}_n\\ \mathbf{v}_n\end{bmatrix}^T\right\} = \begin{bmatrix}\mathbf{Q}_n & \mathbf{0}_{K\times M}\\ \mathbf{0}_{M\times K} & \mathbf{R}_n\end{bmatrix}. \tag{2}$$

The definition of the TMM given by (1) is not only simple, since it merely assumes the first-order Markovianity of the process $\{\mathbf{t}_n\}_{n\geq 0}$, but it also gives rise to potentially very rich models thanks to the introduction of the auxiliary variable, able to account for nonstationarity, parameter uncertainty, or error sources. Note that (1) is a nice generalization of the standard HMM for $\{[\mathbf{x}_n^T, \mathbf{y}_n^T]^T\}_{n\geq 0}$, which is recovered as the special case when $\mathbf{A}_n^{(12)}$, $\mathbf{A}_n^{(13)}$, $\mathbf{A}_n^{(32)}$, and $\mathbf{A}_n^{(33)}$ are set to all-zero matrices. Similarly, (1) boils down to a first-order PMM [10], [19], by selecting $\mathbf{A}_n^{(12)}$ and $\mathbf{A}_n^{(32)}$ as all-zero matrices. Therefore, the two aforementioned particular cases will be excluded from further consideration in the sequel.

In our setting, $\mathbf{x}_n$ (resp., $\mathbf{r}_n$) in (1) will be assimilated to the variables of special interest (resp., without interest) for the sake of further processing. This may need a straightforward linear change of variable in the original state-space formulation as in [20, p. 302] for the sake of variable reordering/combining.

III. Model Dimension Reduction

Model transformation will be instrumental in creating an RDF, whose order is the dimension of the desired state $\mathbf{x}_n$ in (1). A natural idea is to seek a reformulation of the TMM in the form of a PMM for the random process $\{[\mathbf{x}_n^T, \mathbf{y}_n^T]^T\}$, by eliminating the nonsignificant auxiliary variable $\mathbf{r}_n$ in (1). In Section III-A, we show that such a model dimension reduction is possible in equivalent form. Namely, we introduce sufficient conditions to transform the original TMM to a PMM of second order. Those conditions are further simplified under specific assumptions on the original TMM in Section III-B.

A. Sufficient Conditions for Existence of a Second-Order PMM

The original TMM can be collapsed to an equivalent substructure that can create the intended workable model where $\{\mathbf{x}_n\}$ is the only hidden random process. Let us extract $[\mathbf{x}_n^T, \mathbf{y}_n^T]^T$ from (1) as

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

where

$$\mathbf{r}_{n-1} = \mathbf{A}_{n-1}^{(22)}\mathbf{r}_{n-2} + \begin{bmatrix}\mathbf{A}_{n-1}^{(21)} & \mathbf{A}_{n-1}^{(23)}\end{bmatrix}\begin{bmatrix}\mathbf{x}_{n-2}\\ \mathbf{y}_{n-2}\end{bmatrix} + \begin{bmatrix}\mathbf{B}_{n-1}^{(21)} & \mathbf{B}_{n-1}^{(22)}\end{bmatrix}\begin{bmatrix}\mathbf{w}_{n-1}\\ \mathbf{v}_{n-1}\end{bmatrix}. \tag{4}$$

Proposition 3.1: The random process $\{[\mathbf{x}_n^T, \mathbf{y}_n^T]^T\}_{n\geq 2}$ is a PMM of second order driven by the noise process $\{[\mathbf{w}_n^T, \mathbf{v}_n^T]^T\}$ [see (6)], if one of the following conditions is satisfied for all $n \geq 2$:

$$(i)\ \begin{cases}\mathbf{B}_{n-1}^{(21)} = \mathbf{0}_{L\times K}, \mathbf{B}_{n-1}^{(22)} = \mathbf{0}_{L\times M}\\ \mathbf{A}_{n-1}^{(22)} = \mathbf{0}_{L\times L}\end{cases}$$

$$(ii)\ \begin{cases}\begin{bmatrix}\mathbf{B}_{n-1}^{(11)} & \mathbf{B}_{n-1}^{(12)}\\ \mathbf{B}_{n-1}^{(31)} & \mathbf{B}_{n-1}^{(32)}\end{bmatrix} \text{ is invertible}\\ \mathbf{A}_{n-1}^{(22)} - \mathbf{C}_{n-1}\mathbf{A}_{n-1}^{(12)} - \mathbf{D}_{n-1}\mathbf{A}_{n-1}^{(32)} = \mathbf{0}_{L\times L}\end{cases}$$

where $\mathbf{C}_{n-1} \in \mathbb{R}^{L\times K}$ and $\mathbf{D}_{n-1} \in \mathbb{R}^{L\times M}$ are defined by

$$\begin{bmatrix}\mathbf{C}_{n-1} & \mathbf{D}_{n-1}\end{bmatrix} = \begin{bmatrix}\mathbf{B}_{n-1}^{(21)} & \mathbf{B}_{n-1}^{(22)}\end{bmatrix}\begin{bmatrix}\mathbf{B}_{n-1}^{(11)} & \mathbf{B}_{n-1}^{(12)}\\ \mathbf{B}_{n-1}^{(31)} & \mathbf{B}_{n-1}^{(32)}\end{bmatrix}^{-1}. \tag{5}$$

The proof is postponed to Appendix A. Note that Condition (i) is the trivial way of transforming the original TMM to a PMM of second order, by letting $\mathbf{r}_n$ be a noise-free combination of $\mathbf{x}_{n-1}$ and $\mathbf{y}_{n-1}$, a situation encountered when (1) is a state-augmented model for instance [4]. On the contrary, Condition (ii) corresponds to situations where the auxiliary variable is noise driven, which can account for nonstationarity, parameter uncertainty, or error sources, as mentioned previously. Realistic applications for which Condition (ii) is met are determined in Section IV.

Corollary 3.2: If Condition (ii) is satisfied, (3) can be rewritten in the second-order PMM form for all $n \geq 2$ as

$$\begin{bmatrix}\mathbf{x}_n\\ \mathbf{y}_n\end{bmatrix} = \begin{bmatrix}\mathbf{A}_n^{(1)} & \mathbf{A}_n^{(2)}\\ \mathbf{A}_n^{(3)} & \mathbf{A}_n^{(4)}\end{bmatrix}\begin{bmatrix}\mathbf{x}_{n-1}\\ \mathbf{y}_{n-1}\end{bmatrix} + \begin{bmatrix}\tilde{\mathbf{A}}_n^{(1)} & \tilde{\mathbf{A}}_n^{(2)}\\ \tilde{\mathbf{A}}_n^{(3)} & \tilde{\mathbf{A}}_n^{(4)}\end{bmatrix}\begin{bmatrix}\mathbf{x}_{n-2}\\ \mathbf{y}_{n-2}\end{bmatrix} + \begin{bmatrix}\mathbf{B}_n^{(1)} & \mathbf{B}_n^{(2)}\\ \mathbf{B}_n^{(3)} & \mathbf{B}_n^{(4)}\end{bmatrix}\begin{bmatrix}\mathbf{w}_n\\ \mathbf{v}_n\end{bmatrix} \tag{6}$$

where

$$\begin{aligned}
&\mathbf{A}_n^{(1)} = \mathbf{A}_n^{(11)} + \mathbf{A}_n^{(12)}\mathbf{C}_{n-1}, \mathbf{A}_n^{(2)} = \mathbf{A}_n^{(13)} + \mathbf{A}_n^{(12)}\mathbf{D}_{n-1}\\
&\mathbf{A}_n^{(3)} = \mathbf{A}_n^{(31)} + \mathbf{A}_n^{(32)}\mathbf{C}_{n-1}, \mathbf{A}_n^{(4)} = \mathbf{A}_n^{(33)} + \mathbf{A}_n^{(32)}\mathbf{D}_{n-1}\\
&\tilde{\mathbf{A}}_n^{(1)} = \mathbf{A}_n^{(12)}\left(\mathbf{A}_{n-1}^{(21)} - \mathbf{C}_{n-1}\mathbf{A}_{n-1}^{(11)} - \mathbf{D}_{n-1}\mathbf{A}_{n-1}^{(31)}\right)\\
&\tilde{\mathbf{A}}_n^{(2)} = \mathbf{A}_n^{(12)}\left(\mathbf{A}_{n-1}^{(23)} - \mathbf{C}_{n-1}\mathbf{A}_{n-1}^{(13)} - \mathbf{D}_{n-1}\mathbf{A}_{n-1}^{(33)}\right)\\
&\tilde{\mathbf{A}}_n^{(3)} = \mathbf{A}_n^{(32)}\left(\mathbf{A}_{n-1}^{(21)} - \mathbf{C}_{n-1}\mathbf{A}_{n-1}^{(11)} - \mathbf{D}_{n-1}\mathbf{A}_{n-1}^{(31)}\right)\\
&\tilde{\mathbf{A}}_n^{(4)} = \mathbf{A}_n^{(32)}\left(\mathbf{A}_{n-1}^{(23)} - \mathbf{C}_{n-1}\mathbf{A}_{n-1}^{(13)} - \mathbf{D}_{n-1}\mathbf{A}_{n-1}^{(33)}\right)\\
&\mathbf{B}_n^{(1)} = \mathbf{B}_n^{(11)},\ \mathbf{B}_n^{(2)} = \mathbf{B}_n^{(12)},\ \mathbf{B}_n^{(3)} = \mathbf{B}_n^{(31)},\ \mathbf{B}_n^{(4)} = \mathbf{B}_n^{(32)}.
\end{aligned} \tag{7}$$

The reader is referred to Appendix A for the proof.

We consider the usual case where all submatrices in (7) can be precomputed offline, thus resulting only in incremental complexity increase if state inference is repeated over and over again for multiple observed datasets, as it is often the case in practice.

B. Existence of a Second-Order PMM Under Simplifying Assumptions

Now, we gradually introduce assumptions on the blocks of the partitioned matrix

$$\begin{bmatrix}\mathbf{B}_{n-1}^{(11)} & \mathbf{B}_{n-1}^{(12)}\\ \mathbf{B}_{n-1}^{(31)} & \mathbf{B}_{n-1}^{(32)}\end{bmatrix}$$

that may be verified in applications given in Section IV, so that Condition (ii) in Proposition 3.1 and the matrices in (5) admit more workable forms without inverses of block matrices.

Corollary 3.3: Assuming that $\mathbf{B}_{n-1}^{(32)}$ is nonsingular for all $n \geq 2$, Condition (ii) is equivalent to

$$\begin{cases}\mathbf{S}_{n-1} = \mathbf{B}_{n-1}^{(11)} - \mathbf{B}_{n-1}^{(12)}\mathbf{B}_{n-1}^{(32)-1}\mathbf{B}_{n-1}^{(31)} \text{ is invertible}\\ \mathbf{A}_{n-1}^{(22)} - \mathbf{C}_{n-1}\mathbf{A}_{n-1}^{(12)} - \mathbf{D}_{n-1}\mathbf{A}_{n-1}^{(32)} = \mathbf{0}_{L\times L}\end{cases}$$

where $\mathbf{S}_{n-1}$ is the Schur complement of $\mathbf{B}_{n-1}^{(32)}$ and where according to (5)

$$\begin{aligned}
\mathbf{C}_{n-1} &= \left[\mathbf{B}_{n-1}^{(21)} - \mathbf{B}_{n-1}^{(22)}\mathbf{B}_{n-1}^{(32)-1}\mathbf{B}_{n-1}^{(31)}\right]\mathbf{S}_{n-1}^{-1}\\
\mathbf{D}_{n-1} &= \left[\mathbf{B}_{n-1}^{(22)} - \left(\mathbf{B}_{n-1}^{(21)} - \mathbf{B}_{n-1}^{(22)}\mathbf{B}_{n-1}^{(32)-1}\mathbf{B}_{n-1}^{(31)}\right)\mathbf{S}_{n-1}^{-1}\mathbf{B}_{n-1}^{(12)}\right]\\
&\quad\times \mathbf{B}_{n-1}^{(32)-1}.
\end{aligned} \tag{8}$$

The demonstration proceeds from a straightforward application of the result [30, Eq. (2.3)]. We also obtain the following result as a special case.

Corollary 3.4: Assuming that $\mathbf{B}_{n-1}^{(32)}$ is nonsingular and $\mathbf{B}_{n-1}^{(12)} = \mathbf{0}_{K\times M}$ for all $n \geq 2$, Condition (ii) is equivalent to

$$\begin{cases}\mathbf{B}_{n-1}^{(11)} \text{ is invertible}\\ \mathbf{A}_{n-1}^{(22)} - \mathbf{C}_{n-1}\mathbf{A}_{n-1}^{(12)} - \mathbf{D}_{n-1}\mathbf{A}_{n-1}^{(32)} = \mathbf{0}_{L\times L}\end{cases}$$

where according to (5)

$$\mathbf{C}_{n-1} = \left[\mathbf{B}_{n-1}^{(21)} - \mathbf{B}_{n-1}^{(22)}\mathbf{B}_{n-1}^{(32)-1}\mathbf{B}_{n-1}^{(31)}\right]\mathbf{B}_{n-1}^{(11)-1}$$
$$\mathbf{D}_{n-1} = \mathbf{B}_{n-1}^{(22)}\mathbf{B}_{n-1}^{(32)-1}. \quad (9)$$

IV. Particular TMMs Amenable to a Second-Order PMM Form

In this section, we discuss the conditions, introduced in Section III, under which a TMM can be rewritten as a PMM of second order. First, Section IV-A shows that Condition (ii) is readily satisfied in many applications for which a TMM formulation has been proposed in the literature, which serves as an *a posteriori* justification of the model dimension reduction proposed in Section III. Also, standard RDFs along with their complexity orders are recalled, to serve as benchmarks against which we will measure the performance of our approach. Then, in Section IV-B, we briefly discuss how to handle TMMs that do not originally satisfy the conditions in Section III (the interested reader is referred to Section VII for illustrative examples).

A. Classical TMMs as Exact Second-Order PMMs

1) Colored Process Noise: In dynamical systems, process noise can be colored. Well-known examples include time-correlated velocity [3, p. 320] and acceleration [3, p. 321] in tracking, or the influence of error sources in GPS [31] and inertial sensors [32], [33].

One possible way to account for noise coloration is to modify a standard HMM, so that the white process noise is replaced by a Markovian process noise $\boldsymbol{\epsilon}_n$ [34, pp. 188–189]

$$\begin{aligned}\mathbf{x}_n &= \mathbf{F}_n\mathbf{x}_{n-1} + \mathbf{B}_n\boldsymbol{\epsilon}_n\\ \boldsymbol{\epsilon}_n &= \boldsymbol{\Theta}_n\boldsymbol{\epsilon}_{n-1} + \mathbf{w}_n\\ \mathbf{y}_n &= \mathbf{H}_n\mathbf{x}_n + \mathrm{v}_n\end{aligned} \quad (10)$$

where $\mathbf{B}_n$ is invertible, the initial state $\mathbf{x}_0$ (with distribution characterized by $\hat{\mathbf{x}}_0 = E\{\mathbf{x}_0\}$, $\mathbf{P}_0 = E\{(\mathbf{x}_0 - \hat{\mathbf{x}}_0)((\mathbf{x}_0 - \hat{\mathbf{x}}_0)^T\})$ is independent from the zero-mean white noise process $[\mathbf{w}_n^T, \mathbf{v}_n^T]^T$, for all $n \geq 1$ and $\boldsymbol{\epsilon}_0 = \mathbf{0}$. 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 $E\{\mathbf{w}_n\mathbf{v}_n^T\} = \mathbf{0}$. Equation (10) admits a TMM representation of the form (1) by selecting the auxiliary variable as $\mathrm{r}_n = \boldsymbol{\epsilon}_n$ and

$$\begin{bmatrix}\mathbf{A}_n^{(11)} & \mathbf{A}_n^{(12)} & \mathbf{A}_n^{(13)}\\ \mathbf{A}_n^{(21)} & \mathbf{A}_n^{(22)} & \mathbf{A}_n^{(23)}\\ \mathbf{A}_n^{(31)} & \mathbf{A}_n^{(32)} & \mathbf{A}_n^{(33)}\end{bmatrix} = \begin{bmatrix}\mathbf{F}_n & \mathbf{B}_n\boldsymbol{\Theta}_n & \mathbf{0}\\ \mathbf{0} & \boldsymbol{\Theta}_n & \mathbf{0}\\ \mathbf{H}_n\mathbf{F}_n & \mathbf{H}_n\mathbf{B}_n\boldsymbol{\Theta}_n & \mathbf{0}\end{bmatrix}$$
$$\begin{bmatrix}\mathbf{B}_n^{(11)} & \mathbf{B}_n^{(12)}\\ \mathbf{B}_n^{(21)} & \mathbf{B}_n^{(22)}\\ \mathbf{B}_n^{(31)} & \mathbf{B}_n^{(32)}\end{bmatrix} = \begin{bmatrix}\mathbf{B}_n & \mathbf{0}\\ \mathbf{I}_K & \mathbf{0}\\ \mathbf{H}_n\mathbf{B}_n & \mathbf{I}_M\end{bmatrix}. \quad (11)$$

Using Corollary 3.4, it can be easily checked that Condition (ii) in Proposition 3.1 is satisfied, so that (10) also admits a second-order PMM representation of the form (6), where from (7)

$$\begin{bmatrix}\mathbf{A}_n^{(1)} & \mathbf{A}_n^{(2)}\\ \mathbf{A}_n^{(3)} & \mathbf{A}_n^{(4)}\end{bmatrix} = \begin{bmatrix}\mathbf{F}_n + \mathbf{B}_n\boldsymbol{\Theta}_n\mathbf{B}_{n-1}^{-1} & \mathbf{0}\\ \mathbf{H}_n\left(\mathbf{F}_n + \mathbf{B}_n\boldsymbol{\Theta}_n\mathbf{B}_{n-1}^{-1}\right) & \mathbf{0}\end{bmatrix}$$
$$\begin{bmatrix}\tilde{\mathbf{A}}_n^{(1)} & \tilde{\mathbf{A}}_n^{(2)}\\ \tilde{\mathbf{A}}_n^{(3)} & \tilde{\mathbf{A}}_n^{(4)}\end{bmatrix} = \begin{bmatrix}-\mathbf{B}_n\boldsymbol{\Theta}_n\mathbf{B}_{n-1}^{-1}\mathbf{F}_{n-1} & \mathbf{0}\\ -\mathbf{H}_n\mathbf{B}_n\boldsymbol{\Theta}_n\mathbf{B}_{n-1}^{-1}\mathbf{F}_{n-1} & \mathbf{0}\end{bmatrix}. \quad (12)$$

Contrary to the method advocated in the present section, the colored process noise problem is usually handled without model dimension reduction [34, pp. 188–189]. An exception is the RDF in the form of a modified HMM-KF in [20, p. 297], obtained by recomputing the Kalman gain of the standard HMM-KF to account for process noise coloration, which in turn increases the complexity per time instant to $\mathcal{O}(M^3 + 2\,K^2M + 2KM^2 + 3\,K^3 + 4\,K^2L + KL^2 + 2\,L^3)$.

2) Colored Process and Measurement Noise: We now focus on situations where both the process and measurement noise are colored, which can arise in tracking [14, p. 75], [36] and navigation [32].

Let us again modify a standard HMM, so that the white process and measurement noises become jointly Markovian [14, p. 75]

$$\begin{aligned}\mathbf{x}_n &= \mathbf{F}_n\mathbf{x}_{n-1} + \mathbf{B}_n\boldsymbol{\epsilon}_n\\ \boldsymbol{\epsilon}_n &= \boldsymbol{\Theta}_n\boldsymbol{\epsilon}_{n-1} + \boldsymbol{\Phi}_n\boldsymbol{\eta}_{n-1} + \mathbf{w}_n\\ \boldsymbol{\eta}_n &= \boldsymbol{\Gamma}_n\boldsymbol{\epsilon}_{n-1} + \boldsymbol{\Psi}_n\boldsymbol{\eta}_{n-1} + \mathbf{v}_n\\ \mathbf{y}_n &= \mathbf{H}_n\mathbf{x}_n + \boldsymbol{\eta}_n\end{aligned} \quad (13)$$

where $\mathbf{B}_n$ is invertible, the initial state $\mathbf{x}_0$ (with distribution characterized by $\hat{\mathbf{x}}_0 = E\{\mathbf{x}_0\}$, $\mathbf{P}_0 = E\{(\mathbf{x}_0 - \hat{\mathbf{x}}_0)((\mathbf{x}_0 - \hat{\mathbf{x}}_0)^T\})$ is independent from the zero-mean white noise process $[\mathbf{w}_n^T, \mathbf{v}_n^T]^T$, for all $n \geq 1$ and $\boldsymbol{\epsilon}_0 = \mathbf{0}$, $\boldsymbol{\eta}_0 = \mathbf{0}$. 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 $E\{\mathbf{w}_n\mathbf{v}_n^T\} = \mathbf{0}$. Equation (13) admits a TMM representation of the form (1) by selecting the auxiliary variable as $\mathrm{r}_n = [\boldsymbol{\epsilon}_n^T, \boldsymbol{\eta}_n^T]^T$ and

$$\begin{bmatrix}\mathbf{A}_n^{(11)} & \mathbf{A}_n^{(12)} & \mathbf{A}_n^{(13)}\\ \mathbf{A}_n^{(21)} & \mathbf{A}_n^{(22)} & \mathbf{A}_n^{(23)}\\ \mathbf{A}_n^{(31)} & \mathbf{A}_n^{(32)} & \mathbf{A}_n^{(33)}\end{bmatrix} = \begin{bmatrix}\mathbf{F}_n & \mathbf{B}_n\begin{bmatrix}\boldsymbol{\Theta}_n & \boldsymbol{\Phi}_n\end{bmatrix} & \mathbf{0}\\ \mathbf{0} & \begin{bmatrix}\boldsymbol{\Theta}_n & \boldsymbol{\Phi}_n\\ \boldsymbol{\Gamma}_n & \boldsymbol{\Psi}_n\end{bmatrix} & \mathbf{0}\\ \mathbf{H}_n\mathbf{F}_n & \begin{bmatrix}\mathbf{H}_n\mathbf{B}_n\boldsymbol{\Theta}_n & \mathbf{H}_n\mathbf{B}_n\boldsymbol{\Phi}_n\\ +\boldsymbol{\Gamma}_n & +\boldsymbol{\Psi}_n\end{bmatrix} & \mathbf{0}\end{bmatrix}$$
$$\begin{bmatrix}\mathbf{B}_n^{(11)} & \mathbf{B}_n^{(12)}\\ \mathbf{B}_n^{(21)} & \mathbf{B}_n^{(22)}\\ \mathbf{B}_n^{(31)} & \mathbf{B}_n^{(32)}\end{bmatrix} = \left[\begin{array}{cc}\mathbf{B}_n & \mathbf{0}\\ \hline \mathbf{I}_K & \mathbf{0}\\ \mathbf{0} & \mathbf{I}_M\\ \hline \mathbf{H}_n\mathbf{B}_n & \mathbf{I}_M\end{array}\right]. \quad (14)$$

Using Corollary 3.4, it can be easily checked that Condition (ii) in Proposition 3.1 is satisfied, so that (13) also admits a second-order PMM representation of the form (6), where from

(7)

$$
\begin{bmatrix} \mathbf{A}_n^{(1)} & \mathbf{A}_n^{(2)} \\ \mathbf{A}_n^{(3)} & \mathbf{A}_n^{(4)} \end{bmatrix} =
$$

$$
\left[\begin{array}{cc} \begin{array}{c}\mathbf{F}_n \\ +\mathbf{B}_n\left(\boldsymbol{\Theta}_n\mathbf{B}_{n-1}^{-1} - \boldsymbol{\Phi}_n\mathbf{H}_{n-1}\right)\end{array} & \mathbf{B}_n\boldsymbol{\Phi}_n \\ \hline \begin{array}{c}\mathbf{H}_n\mathbf{F}_n \\ +\mathbf{H}_n\mathbf{B}_n\left(\boldsymbol{\Theta}_n\mathbf{B}_{n-1}^{-1} - \boldsymbol{\Phi}_n\mathbf{H}_{n-1}\right) \\ +\boldsymbol{\Gamma}_n\mathbf{B}_{n-1}^{-1} - \boldsymbol{\Psi}_n\mathbf{H}_{n-1}\end{array} & \mathbf{H}_n\mathbf{B}_n\boldsymbol{\Phi}_n + \boldsymbol{\Psi}_n \end{array}\right]
$$

$$
\begin{bmatrix} \tilde{\mathbf{A}}_n^{(1)} & \tilde{\mathbf{A}}_n^{(2)} \\ \tilde{\mathbf{A}}_n^{(3)} & \tilde{\mathbf{A}}_n^{(4)} \end{bmatrix} = \begin{bmatrix} -\mathbf{B}_n\boldsymbol{\Theta}_n\mathbf{B}_{n-1}^{-1}\mathbf{F}_{n-1} & \mathbf{0} \\ -\left(\mathbf{H}_n\mathbf{B}_n\boldsymbol{\Theta}_n + \boldsymbol{\Gamma}_n\right)\mathbf{B}_{n-1}^{-1}\mathbf{F}_{n-1} & \mathbf{0} \end{bmatrix}. \tag{15}
$$

Again contrary to the method advocated in the present section, the colored process and measurement noise problem is usually handled without model dimension reduction [36], [37]. An exception is the RDF in the form of a modified HMM-KF in [20, p. 301], obtained by recomputing the Kalman gain of the standard HMM-KF to account for process and measurement noise coloration, which in turn increases the complexity per time instant to $\mathcal{O}(M^3 + 3\,K^2M + 3KM^2 + 4\,K^3 + 2KLM + 5\,K^2L + 3KL^2 + L^2M + LM^2 + 2\,L^3)$. However, for the particular case with measurement noise coloration only, an optimal reduced-dimension HMM-KF using time-differenced measurements without complexity increase wrt the standard HMM-KF is available [35, Sec. IV and Appendix], [34, p. 191–192], [3, p. 329]. In this case, the proposed transformed model boils down to a first-order PMM (see [19]), so that the standard PMM-KF is also applicable.

3) Discrete Wiener Process Acceleration (DWPA) Model: We consider the following one-dimensional (1-D) kinematic model, where the state vector at instant n contains the position, velocity, and acceleration, $\bar{\mathbf{x}}_n = [p_n, v_n, a_n]^T$ [3, p. 274]:

$$
\begin{aligned} \bar{\mathbf{x}}_n &= \bar{\mathbf{F}}_n\bar{\mathbf{x}}_{n-1} + \bar{\mathbf{B}}_n\mathbf{w}_n \\ \mathbf{y}_n &= \bar{\mathbf{H}}_n\bar{\mathbf{x}}_n + \mathbf{v}_n \end{aligned} \tag{16}
$$

with parameters depending on the sampling period T

$$
\bar{\mathbf{F}}_n = \begin{bmatrix} 1 & T & \frac{T^2}{2} \\ 0 & 1 & T \\ 0 & 0 & 1 \end{bmatrix}, \bar{\mathbf{B}}_n = \begin{bmatrix} \frac{T^2}{2} & 0 \\ T & 1 \\ 1 & 0 \end{bmatrix}, \bar{\mathbf{H}}_n = [1, 0, 0] \tag{17}
$$

and the initial state $\bar{\mathbf{x}}_0$ (with distribution characterized by $\hat{\bar{\mathbf{x}}}_0 = E\{\bar{\mathbf{x}}_0\}$, $\bar{\mathbf{P}}_0 = E\{(\bar{\mathbf{x}}_0 - \hat{\bar{\mathbf{x}}}_0)((\bar{\mathbf{x}}_0 - \hat{\bar{\mathbf{x}}}_0)^T\})$ is independent from the zero-mean white noise process $[\mathbf{w}_n^T, \mathbf{v}_n^T]^T$, for all $n \geq 1$. The noise covariance is defined by $E\{\mathbf{w}_n\mathbf{w}_n^T\} = \mathrm{diag}([Q, 0])$, $E\{\mathbf{v}_n\mathbf{v}_n^T\} = R$, and $E\{\mathbf{w}_n\mathbf{v}_n^T\} = \mathbf{0}$. Equation (16) admits a TMM representation of the form (1) by selecting $\mathbf{x}_n = [p_n, v_n]^T$, the auxiliary variable as $\mathrm{r}_n = a_n$ and

$$
\begin{bmatrix} \mathbf{A}_n^{(11)} & \mathbf{A}_n^{(12)} & \mathbf{A}_n^{(13)} \\ \mathbf{A}_n^{(21)} & \mathbf{A}_n^{(22)} & \mathbf{A}_n^{(23)} \\ \mathbf{A}_n^{(31)} & \mathbf{A}_n^{(32)} & \mathbf{A}_n^{(33)} \end{bmatrix} = \left[\begin{array}{cc|c|c} 1 & T & \frac{T^2}{2} & 0 \\ 0 & 1 & T & 0 \\ \hline 0 & 0 & 1 & 0 \\ \hline 1 & T & \frac{T^2}{2} & 0 \end{array}\right]
$$

$$
\begin{bmatrix} \mathbf{B}_n^{(11)} & \mathbf{B}_n^{(12)} \\ \mathbf{B}_n^{(21)} & \mathbf{B}_n^{(22)} \\ \mathbf{B}_n^{(31)} & \mathbf{B}_n^{(32)} \end{bmatrix} = \left[\begin{array}{cc|c} \frac{T^2}{2} & 0 & 0 \\ T & 1 & 0 \\ \hline 1 & 0 & 0 \\ \hline \frac{T^2}{2} & 0 & 1 \end{array}\right]. \tag{18}
$$

Using Corollary 3.4, it can be easily checked that Condition (ii) in Proposition 3.1 is satisfied, so that (16) also admits a second-order PMM representation of the form (6), where from (7)

$$
\begin{bmatrix} \mathbf{A}_n^{(1)} & \mathbf{A}_n^{(2)} \\ \mathbf{A}_n^{(3)} & \mathbf{A}_n^{(4)} \end{bmatrix} = \left[\begin{array}{cc|c} 2 & T & 0 \\ \frac{2}{T} & 1 & 0 \\ \hline 2 & T & 0 \end{array}\right]
$$

$$
\begin{bmatrix} \tilde{\mathbf{A}}_n^{(1)} & \tilde{\mathbf{A}}_n^{(2)} \\ \tilde{\mathbf{A}}_n^{(3)} & \tilde{\mathbf{A}}_n^{(4)} \end{bmatrix} = \left[\begin{array}{cc|c} -1 & -T & 0 \\ -\frac{2}{T} & -2 & 0 \\ \hline -1 & -T & 0 \end{array}\right]. \tag{19}
$$

This result can be extended straightforwardly to 2-D and 3-D kinematic models in xy and xyz Cartesian coordinate system, respectively.

Contrary to the method advocated in the present section, the DWPA problem is usually handled without model dimension reduction [3, p. 274]. An exception is the RDF in the form of a modified HMM-KF in [23], obtained by recomputing the Kalman gain of the standard HMM-KF to account for the elimination of the auxiliary variable from the state space, which in turn increases the complexity per time instant to $\mathcal{O}(M^3 + 6\,K^2M + 4KM^2 + 32\,K^3 + 7KLM + 48\,K^2L + 24KL^2 + 2\,L^2M + 2LM^2 + 4\,L^3)$.

B. Approximating TMMs With PMMs of Second Order

We now consider TMMs for which Condition (i) or (ii) is not readily satisfied. The question then arises as to whether it is possible to depart from the original model to arrive at an approximate reduced-dimension PMM of second order. The resulting model must then be validated by simulation to check its ability to produce reasonable state estimates.

1) Model Identification: A first line of thought consists in finding the system parameters for an approximate model of the form (6) by maximizing the log-likelihood, given a set of observations following the truth-model in (1). This maximization would require an extension to second-order PMMs of the expectation maximization (EM) [17] or the gradient-based [18] approach available for first-order PMMs.

Another type of approximation consists in replacing the original TMM by a surrogate TMM that replaces $\mathbf{B}_{n-1}^{(11)}$ by $\widehat{\mathbf{B}}_{n-1}^{(11)}$ and $\mathbf{B}_{n-1}^{(21)}$ by $\widehat{\mathbf{B}}_{n-1}^{(21)}$ so that Condition (ii) is approximately satisfied while keeping mismodeling errors in (6) wrt the truth-model at an acceptable level. This technique is borrowed from process noise covariance tuning, which was introduced in the early 1970s [38, pp. 305–307] to compensate for HMM modeling uncertainties. There are several possibilities for such model perturbation. For simplicity, we restrict ourselves to the case where $\mathbf{B}_{n-1}^{(32)}$ is nonsingular and $\mathbf{B}_{n-1}^{(12)} = \mathbf{0}_{K\times M}$ for all $n \geq 2$, which is often verified in practice.

2) Closed-Form Model Perturbation: Without modifying $\mathbf{B}_{n-1}^{(21)}$, if we can replace $\mathbf{B}_{n-1}^{(11)}$ by $\widehat{\mathbf{B}}_{n-1}^{(11)}$ for all $n \geq 2$ defined as

$$\widehat{\mathbf{B}}_{n-1}^{(11)} = [\mathbf{A}_{n-1}^{(12)}][\mathbf{A}_{n-1}^{(22)} - \mathbf{B}_{n-1}^{(22)}\mathbf{B}_{n-1}^{(32)-1}\mathbf{A}_{n-1}^{(32)}]^{-1} \times [\mathbf{B}_{n-1}^{(21)} - \mathbf{B}_{n-1}^{(22)}\mathbf{B}_{n-1}^{(32)-1}\mathbf{B}_{n-1}^{(31)}] \tag{20}$$

where all matrices inside the brackets must be square (i.e., $K = L$) and invertible, then Condition (ii) is met according to Corollary 3.4.

3) Numerical Optimization-Based Model Perturbation: Another possibility is to minimize the cost function $||\mathbf{A}_{n-1}^{(22)} - \mathbf{C}_{n-1}\mathbf{A}_{n-1}^{(12)} - \mathbf{D}_{n-1}\mathbf{A}_{n-1}^{(32)}||_F$, replacing $\mathbf{B}_{n-1}^{(11)}$ by $\widehat{\mathbf{B}}_{n-1}^{(11)}$ and $\mathbf{B}_{n-1}^{(21)}$ by $\widehat{\mathbf{B}}_{n-1}^{(21)}$ in the expression of $\mathbf{C}_{n-1}$. Reparameterizing this expression using

$$\begin{aligned}\mathbf{X}_{n-1} &= \widehat{\mathbf{B}}_{n-1}^{(21)} - \mathbf{B}_{n-1}^{(22)}\mathbf{B}_{n-1}^{(32)-1}\mathbf{B}_{n-1}^{(31)}\\ \mathbf{Y}_{n-1} &= \widehat{\mathbf{B}}_{n-1}^{(11)-1}\\ \mathbf{\Lambda}_{n-1} &= \mathbf{A}_{n-1}^{(22)} - \mathbf{B}_{n-1}^{(22)}\mathbf{B}_{n-1}^{(32)-1}\mathbf{A}_{n-1}^{(32)}\end{aligned} \tag{21}$$

leads to the minimization of $||\mathbf{\Lambda}_{n-1} - \mathbf{X}_{n-1}\mathbf{Y}_{n-1}\mathbf{A}_{n-1}^{(12)}||_F$. Note that there is a one-to-one correspondence between the new matrix variables $\mathbf{X}_{n-1}$, $\mathbf{Y}_{n-1}$ and the desired matrices $\widehat{\mathbf{B}}_{n-1}^{(11)}$ $\widehat{\mathbf{B}}_{n-1}^{(21)}$, whereas $\mathbf{\Lambda}_{n-1}$ is a constant matrix. This task can then be performed via numerical optimization, for instance using a gradient descent since the partial derivatives wrt $\mathbf{X}_{n-1}$ and $\mathbf{Y}_{n-1}$ are computable at each time step $n \geq 2$. Note that the cost function will be able to approach zero up to the numerical precision of the computer only when there are sufficient degrees of freedom during optimization, which in turn sets an upper limit to the dimension of the auxiliary variable, i.e., $L < K$.

V. Optimal Linear RDF

Optimal filtering in the MMSE sense is feasible in the original model (1) by applying the TMM-KF [12], [13], 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. Therefore we derive the best linear MMSE RDF adapted to the transformed model (6). Let us recall that the proposed state estimator is applicable to any PMM of second order, irrespective of whether it is obtained from an original TMM by model dimension reduction as in Section III or not. We develop the state estimate as a linear combination of the measurements that minimizes the mean square error (MSE), using the classical prediction (see Section V-A) and update (see Section V-B) steps. Then, in the particular case of Gaussian second-order PMMs, we provide an interpretation as the minimum variance Bayes' estimate (see Section V-C). In the sequel, we consider a second-order PMM of the form (6) and the notation $\hat{\mathbf{x}}_{i|j}$ refers to an estimate of a random vector $\mathbf{x}$ at time instant i obtained as a weighted sum of all measurements up to time j. The corresponding error covariance is denoted by $\mathbf{P}_{i|j} = E\{(\mathbf{x}_i - \hat{\mathbf{x}}_{i|j})(\mathbf{x}_i - \hat{\mathbf{x}}_{i|j})^T\}$.

A. State Estimate Prediction

Assume that the updated state estimates $\hat{\mathbf{x}}_{n-2|n-1}$, $\hat{\mathbf{x}}_{n-1|n-1}$ at instant $n-1$ are available, we set the predicted estimates $\hat{\mathbf{x}}_{n|n-1}$, $\hat{\mathbf{y}}_{n|n-1}$ at instant n to

$$\begin{bmatrix}\hat{\mathbf{x}}_{n|n-1}\\ \hat{\mathbf{y}}_{n|n-1}\end{bmatrix} = \begin{bmatrix}\mathbf{A}_n^{(1)} & \mathbf{A}_n^{(2)}\\ \mathbf{A}_n^{(3)} & \mathbf{A}_n^{(4)}\end{bmatrix}\begin{bmatrix}\hat{\mathbf{x}}_{n-1|n-1}\\ \mathbf{y}_{n-1}\end{bmatrix} + \begin{bmatrix}\tilde{\mathbf{A}}_n^{(1)} & \tilde{\mathbf{A}}_n^{(2)}\\ \tilde{\mathbf{A}}_n^{(3)} & \tilde{\mathbf{A}}_n^{(4)}\end{bmatrix}\begin{bmatrix}\hat{\mathbf{x}}_{n-2|n-1}\\ \mathbf{y}_{n-2}\end{bmatrix} \tag{22}$$

so that the prediction error covariance has the form

$$E\left\{\begin{bmatrix}\mathbf{x}_n - \hat{\mathbf{x}}_{n|n-1}\\ \mathbf{y}_n - \hat{\mathbf{y}}_{n|n-1}\end{bmatrix}\begin{bmatrix}\mathbf{x}_n - \hat{\mathbf{x}}_{n|n-1}\\ \mathbf{y}_n - \hat{\mathbf{y}}_{n|n-1}\end{bmatrix}^T\right\} = \begin{bmatrix}\mathbf{P}_{n|n-1} & \mathbf{\Sigma}_{n|n-1}\\ \mathbf{\Sigma}_{n|n-1}^T & \mathbf{L}_{n|n-1}\end{bmatrix}. \tag{23}$$

Moreover, we also introduce the cross-covariance matrix $\mathbf{\Pi}_{n|n-1} = E\{(\mathbf{x}_{n-1} - \hat{\mathbf{x}}_{n-1|n-1})(\mathbf{y}_n - \hat{\mathbf{y}}_{n|n-1})^T\}$. All aforementioned (cross-)covariance matrices are obtained recursively as

$$\begin{aligned}\mathbf{P}_{n|n-1} &= \mathbf{A}_n^{(1)}\mathbf{P}_{n-1|n-1}\mathbf{A}_n^{(1)^T} + \mathbf{B}_n^{(1)}\mathbf{Q}_n\mathbf{B}_n^{(1)^T}\\ &\quad + \mathbf{B}_n^{(2)}\mathbf{R}_n\mathbf{B}_n^{(2)^T} + \mathbf{A}_n^{(1)}\mathbf{P}_{n-1,n-2|n-1}\tilde{\mathbf{A}}_n^{(1)T}\\ &\quad + \tilde{\mathbf{A}}_n^{(1)}\mathbf{P}_{n-1,n-2|n-1}^T\mathbf{A}_n^{(1)T} + \tilde{\mathbf{A}}_n^{(1)}\mathbf{P}_{n-2|n-1}\tilde{\mathbf{A}}_n^{(1)T}\\ \mathbf{\Sigma}_{n|n-1} &= \mathbf{A}_n^{(1)}\mathbf{P}_{n-1|n-1}\mathbf{A}_n^{(3)^T} + \mathbf{B}_n^{(1)}\mathbf{Q}_n\mathbf{B}_n^{(3)^T}\\ &\quad + \mathbf{B}_n^{(2)}\mathbf{R}_n\mathbf{B}_n^{(4)^T} + \mathbf{A}_n^{(1)}\mathbf{P}_{n-1,n-2|n-1}\tilde{\mathbf{A}}_n^{(3)T}\\ &\quad + \tilde{\mathbf{A}}_n^{(1)}\mathbf{P}_{n-1,n-2|n-1}^T\mathbf{A}_n^{(3)T} + \tilde{\mathbf{A}}_n^{(1)}\mathbf{P}_{n-2|n-1}\tilde{\mathbf{A}}_n^{(3)T}\\ \mathbf{L}_{n|n-1} &= \mathbf{A}_n^{(3)}\mathbf{P}_{n-1|n-1}\mathbf{A}_n^{(3)^T} + \mathbf{B}_n^{(3)}\mathbf{Q}_n\mathbf{B}_n^{(3)^T}\\ &\quad + \mathbf{B}_n^{(4)}\mathbf{R}_n\mathbf{B}_n^{(4)^T} + \mathbf{A}_n^{(3)}\mathbf{P}_{n-1,n-2|n-1}\tilde{\mathbf{A}}_n^{(3)T}\\ &\quad + \tilde{\mathbf{A}}_n^{(3)}\mathbf{P}_{n-1,n-2|n-1}^T\mathbf{A}_n^{(3)T} + \tilde{\mathbf{A}}_n^{(3)}\mathbf{P}_{n-2|n-1}\tilde{\mathbf{A}}_n^{(3)T}\\ \mathbf{\Pi}_{n|n-1} &= \mathbf{P}_{n-1|n-1}\mathbf{A}_n^{(3)^T} + \mathbf{P}_{n-1,n-2|n-1}\tilde{\mathbf{A}}_n^{(3)T}\end{aligned} \tag{24}$$

where we define

$$\mathbf{P}_{n-1,n-2|n-1} = E\{(\mathbf{x}_{n-1} - \hat{\mathbf{x}}_{n-1|n-1})(\mathbf{x}_{n-2} - \hat{\mathbf{x}}_{n-2|n-1})^T\}. \tag{25}$$

The demonstration is postponed to Appendix B.

B. State Estimate Update

Given the predicted estimates in (22), updated state estimates, obtained by linearly combining past measurements with the new measurement at instant n, have the form

$$\begin{aligned}\hat{\mathbf{x}}_{n|n} &= \hat{\mathbf{x}}_{n|n-1} + \mathbf{K}_n(\mathbf{y}_n - \hat{\mathbf{y}}_{n|n-1})\\ \hat{\mathbf{x}}_{n-1|n} &= \hat{\mathbf{x}}_{n-1|n-1} + \mathbf{J}_n(\mathbf{y}_n - \hat{\mathbf{y}}_{n|n-1})\end{aligned} \tag{26}$$

where $\mathbf{K}_n$ and $\mathbf{J}_n$ are weighting matrices to be optimized.

Let us now define the estimation errors during the update stage as $\boldsymbol{\epsilon}_{n|n} = \mathbf{x}_n - \hat{\mathbf{x}}_{n|n}$ and $\boldsymbol{\epsilon}_{n-1|n} = \mathbf{x}_{n-1} - \hat{\mathbf{x}}_{n-1|n}$; the corresponding error covariance matrices $\mathbf{P}_{n|n} = E\{\boldsymbol{\epsilon}_{n|n}\boldsymbol{\epsilon}_{n|n}^T\}$ and $\mathbf{P}_{n-1|n} = E\{\boldsymbol{\epsilon}_{n-1|n}\boldsymbol{\epsilon}_{n-1|n}^T\}$ can be written as

$$\mathbf{P}_{n|n} = \mathbf{P}_{n|n-1} - \mathbf{K}_n\boldsymbol{\Sigma}_{n|n-1}^T - \boldsymbol{\Sigma}_{n|n-1}\mathbf{K}_n^T + \mathbf{K}_n\mathbf{L}_{n|n-1}\mathbf{K}_n^T$$
$$\mathbf{P}_{n-1|n} = \mathbf{P}_{n-1|n-1} - \mathbf{J}_n\boldsymbol{\Pi}_{n|n-1}^T - \boldsymbol{\Pi}_{n|n-1}\mathbf{J}_n^T + \mathbf{J}_n\mathbf{L}_{n|n-1}\mathbf{J}_n^T. \quad (27)$$

We must also provide the expression of (25). It can be shown that it admits the recursive form

$$\begin{aligned}\mathbf{P}_{n,n-1|n} &= (\mathbf{A}_n^{(1)} - \mathbf{K}_n\mathbf{A}_n^{(3)})\mathbf{P}_{n-1|n-1} \\ &\quad + (\tilde{\mathbf{A}}_n^{(1)} - \mathbf{K}_n\tilde{\mathbf{A}}_n^{(3)})\mathbf{P}_{n-1,n-2|n-1}^T \\ &\quad + (\mathbf{K}_n\mathbf{L}_{n|n-1} - \boldsymbol{\Sigma}_{n|n-1})\mathbf{J}_n^T. \end{aligned} \quad (28)$$

The interested reader is referred to Appendix C for a demonstration.

Obtaining the best linear filter in the MMSE sense now consists in selecting $\mathbf{K}_n$ and $\mathbf{J}_n$ so as to minimize $\mathrm{trace}(\mathbf{P}_{n|n})$ and $\mathrm{trace}(\mathbf{P}_{n-1|n})$, respectively [39]. Finally, considering that

$$\frac{\partial\,\mathrm{trace}(\mathbf{P}_{n|n})}{\partial \mathbf{K}_n} = 2(\mathbf{K}_n\mathbf{L}_{n|n-1} - \boldsymbol{\Sigma}_{n|n-1})$$
$$\frac{\partial\,\mathrm{trace}(\mathbf{P}_{n-1|n})}{\partial \mathbf{J}_n} = 2(\mathbf{J}_n\mathbf{L}_{n|n-1} - \boldsymbol{\Pi}_{n|n-1}) \quad (29)$$

the solution of the matrix gain optimization is given by

$$\mathbf{K}_n^* = \boldsymbol{\Sigma}_{n|n-1}(\mathbf{L}_{n|n-1})^{-1}$$
$$\mathbf{J}_n^* = \boldsymbol{\Pi}_{n|n-1}(\mathbf{L}_{n|n-1})^{-1}. \quad (30)$$

C. Gaussian Interpretation for Second-Order PMMs

It is well-known that a second-order Markovian process such as (6) can be rewritten in the first-order Markovian form using state-augmentation [4]

$$\begin{bmatrix}\mathbf{x}_n \\ \mathbf{x}_{n-1} \\ \mathbf{y}_n\end{bmatrix} = \begin{bmatrix}\mathbf{A}_n^{(1)} & \tilde{\mathbf{A}}_n^{(1)} & \mathbf{A}_n^{(2)} \\ \mathbf{I}_K & \mathbf{0} & \mathbf{0} \\ \mathbf{A}_n^{(3)} & \tilde{\mathbf{A}}_n^{(3)} & \mathbf{A}_n^{(4)}\end{bmatrix}\begin{bmatrix}\mathbf{x}_{n-1} \\ \mathbf{x}_{n-2} \\ \mathbf{y}_{n-1}\end{bmatrix} + \begin{bmatrix}\tilde{\mathbf{A}}_n^{(2)} \\ \mathbf{0} \\ \tilde{\mathbf{A}}_n^{(4)}\end{bmatrix}\mathbf{y}_{n-2} + \begin{bmatrix}\mathbf{B}_n^{(11)} & \mathbf{B}_n^{(12)} \\ \mathbf{0} & \mathbf{0} \\ \mathbf{B}_n^{(31)} & \mathbf{B}_n^{(32)}\end{bmatrix}\begin{bmatrix}\mathbf{w}_n \\ \mathbf{v}_n\end{bmatrix}. \quad (31)$$

In the particular case where the random process defined by (31) is Gaussian (i.e., $[\mathbf{x}_0^T, \mathbf{y}_0^T, \mathbf{x}_1^T, \mathbf{y}_1^T]^T$ and $[\mathbf{w}_n^T, \mathbf{v}_n^T]^T$, for all $n \geq 2$ are Gaussian distributed), the PMM-KF applied to (31) in state-augmented form is equivalent to the updated estimates derived in Section V-B. Interestingly, since the PMM-KF was derived in a Bayesian setting for any first-order Gaussian PMM, we conclude that for the particular Gaussian PMM in (31), the proposed estimates maximize the posterior distribution

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

so that they coincide with the minimum-variance Bayes' estimates in this context, as expected.

Note however that for the sake of complexity reduction, our formulation derived from first principles regarding second-order PMMs is preferable over the direct use of the PMM-KF applied to (31), which computes and stores many unnecessary vector and matrix quantities relative to the previous time instant during the prediction step. Therefore, our formulation is not only more compact but is also obtained under less restrictive assumptions, that is, the noise only needs to be white and zero-mean.

VI. Performance Analysis

We analyze the performance of the RDF proposed in Section V in terms of MSE, based on a computationally efficient iterative form of the optimized error covariance matrices (see Section VI-A). Next, we summarize the proposed optimal RDF in simple algorithmic form in order to assess the computational complexity (see Section VI-B).

A. Iterative Computation of Optimal Error Covariance Matrices

Let us remind the definition of the estimation error at instant n (resp., instant $n-1$) during the update stage, $\boldsymbol{\epsilon}_{n|n} = \mathbf{x}_n - \hat{\mathbf{x}}_{n|n}$ (resp., $\boldsymbol{\epsilon}_{n-1|n} = \mathbf{x}_{n-1} - \hat{\mathbf{x}}_{n-1|n}$). We now focus on the covariance of these estimation errors, $\mathbf{P}_{n|n} = E\{\boldsymbol{\epsilon}_{n|n}\boldsymbol{\epsilon}_{n|n}^T\}$ and $\mathbf{P}_{n-1|n} = E\{\boldsymbol{\epsilon}_{n-1|n}\boldsymbol{\epsilon}_{n-1|n}^T\}$. The general expressions are given by (27), and we are interested in the particular case where the optimal matrix gains in (30) are selected. It follows that the optimal error covariance matrices can be computed in simple iterative form as

$$\mathbf{P}_{n|n} = \mathbf{P}_{n|n-1} - \mathbf{K}_n^*\boldsymbol{\Sigma}_{n|n-1}^T$$
$$\mathbf{P}_{n-1|n} = \mathbf{P}_{n-1|n-1} - \mathbf{J}_n^*\boldsymbol{\Pi}_{n|n-1}^T. \quad (32)$$

B. Complexity Evaluation

The most compact description of the iterative procedure for computing the proposed RDF estimates after matrix gain optimization is given by Algorithm 1, for the sake of complexity assessment.

Remark 6.1: Note that when $\tilde{\mathbf{A}}_n^{(1)}$, $\tilde{\mathbf{A}}_n^{(2)}$, $\tilde{\mathbf{A}}_n^{(3)}$, and $\tilde{\mathbf{A}}_n^{(4)}$ are all-zero matrices for all n, the proposed optimal RDF coincides with the PMM-KF (see [19]), which is expected since (6) boils down to a first-order PMM.

We are now ready to evaluate the asymptotic computational complexity to generate a single estimate at a given time instant n for all the aforementioned state estimators.

1) TMM-KF: $\mathcal{O}(M^3 + 3(K+L)^2M + 2(K+L)M^2 + 2(K+L)^3)$.

Algorithm 1: Proposed Optimal RDF.

Require: $\hat{\mathbf{x}}_{0|1}, \mathbf{P}_{0|1}, \hat{\mathbf{x}}_{1|1}, \mathbf{P}_{1|1}, \mathbf{P}_{1,0|1}, \mathbf{y}_n, \mathbf{Q}_n, \mathbf{R}_n$

for $n = 2, 3, \ldots$ **do**

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

$$\hat{\mathbf{y}}_{n|n-1} = \mathbf{A}_n^{(3)}\hat{\mathbf{x}}_{n-1|n-1} + \mathbf{A}_n^{(4)}\mathbf{y}_{n-1} + \tilde{\mathbf{A}}_n^{(3)}\hat{\mathbf{x}}_{n-2|n-1} + \tilde{\mathbf{A}}_n^{(4)}\mathbf{y}_{n-2}$$

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

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

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

$$\boldsymbol{\Pi}_{n|n-1} = \mathbf{P}_{n-1|n-1}\mathbf{A}_n^{(3)^T} + \mathbf{P}_{n-1,n-2|n-1}\tilde{\mathbf{A}}_n^{(3)T}$$

$$\mathbf{K}_n^* = \boldsymbol{\Sigma}_{n|n-1}(\mathbf{L}_{n|n-1})^{-1}$$

$$\mathbf{J}_n^* = \boldsymbol{\Pi}_{n|n-1}(\mathbf{L}_{n|n-1})^{-1}$$

$$\hat{\mathbf{x}}_{n|n} = \hat{\mathbf{x}}_{n|n-1} + \mathbf{K}_n^*(\mathbf{y}_n - \hat{\mathbf{y}}_{n|n-1})$$

$$\hat{\mathbf{x}}_{n-1|n} = \hat{\mathbf{x}}_{n-1|n-1} + \mathbf{J}_n^*(\mathbf{y}_n - \hat{\mathbf{y}}_{n|n-1})$$

$$\mathbf{P}_{n|n} = \mathbf{P}_{n|n-1} - \mathbf{K}_n^*\boldsymbol{\Sigma}_{n|n-1}^T$$

$$\mathbf{P}_{n-1|n} = \mathbf{P}_{n-1|n-1} - \mathbf{J}_n^*\boldsymbol{\Pi}_{n|n-1}^T$$

$$\mathbf{P}_{n,n-1|n} = (\mathbf{A}_n^{(1)} - \mathbf{K}_n^*\mathbf{A}_n^{(3)})\mathbf{P}_{n-1|n-1} + (\tilde{\mathbf{A}}_n^{(1)} - \mathbf{K}_n^*\tilde{\mathbf{A}}_n^{(3)})\mathbf{P}_{n-1,n-2|n-1}^T$$

return $\hat{\mathbf{x}}_{n|n}, \mathbf{P}_{n|n}$

end for

2) Proposed optimal RDF (see Algorithm 1): $\mathcal{O}(M^3 + 12\,K^2M + 5KM^2 + 6\,K^3)$.
3) PMM-KF: $\mathcal{O}(M^3 + 3\,K^2M + 2KM^2 + 2\,K^3)$.

We observe that the TMM-KF and the proposed RDF are both MMSE-optimal when (1) can be transformed to a PMM of second order. However, as we will see in the following section, the proposed RDF can have much smaller complexity. Also, in the particular case where (6) reduces to a PMM of first order, the proposed RDF simplifies to a PMM-KF, with further complexity reduction by roughly a factor of 3, when M is small wrt K.

VII. Numerical Results

This section is devoted to the performance assessment of the proposed RDF based on simulation results for the following applications.

1) Two-state tracking with colored process and measurement noise.
2) One-state tracking based on a full-dimension two-state kinematic model.
3) Finite-difference approximation to the heat equation.

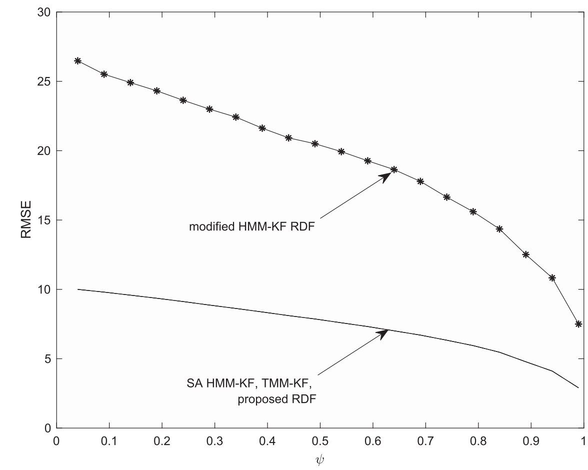

Fig. 1. RMSE for tracking with colored process and measurement noise as a function of ψ, with $T = 1$ s, $\theta = 0.99$, $Q = 10^2$ (m/s^2)^2, and $R = 0.1^2(1-\psi^2)$ m^2.

The first (resp., second and third) application is an example of exact (resp., approximate) conversion of a particular TMM to a PMM of second order. Moreover, the first and the second (resp., the third) applications exemplify the behavior of the proposed method in low (resp., high) dimensions.

A. Two-State Tracking With Colored Process and Measurement Noise

We first investigate a concrete case for Section IV-A2, where the TMM formulation admits a second-order PMM representation, since Condition (ii) is verified. We consider a 1-D discrete white noise acceleration model for tracking and navigation [3, p. 273], where the state vector at instant n contains the position and velocity, $\mathbf{x}_n = [p_n, v_n]^T$, with position-only measurements. It follows that (13) is parameterized by

$$\mathbf{F}_n = \begin{bmatrix} 1 & T \\ 0 & 1 \end{bmatrix}, \mathbf{B}_n = \begin{bmatrix} \frac{T^2}{2} & 0 \\ T & 1 \end{bmatrix}, \mathbf{Q}_n = \mathrm{diag}([Q, 0])$$

$$\begin{bmatrix} \boldsymbol{\Theta}_n & \boldsymbol{\Phi}_n \\ \boldsymbol{\Gamma}_n & \boldsymbol{\Psi}_n \end{bmatrix} = \begin{bmatrix} \theta\mathbf{I}_2 & \mathbf{0}_{2\times 1} \\ \mathbf{0}_{1\times 2} & \psi \end{bmatrix}, \mathbf{H}_n = [1, 0], \mathbf{R}_n = R \quad (33)$$

where $T = 1$ s, $\theta = 0.99$, and $0 < \psi < 1$ stand for the sampling period, the process, and the measurement noise coloration parameter, respectively. The noise components are zero-mean white Gaussian distributed with scalar variance parameters $Q = 10^2(\text{m/s}^2)^2$ and $R = 0.1^2(1-\psi^2)\text{m}^2$. While in simulations the initial state is set to $\mathbf{x}_0 = [1, 1]^T$, all filters work under the assumption that $\mathbf{x}_0 \sim \mathcal{N}(\hat{\mathbf{x}}_0, \mathbf{P}_0)$ is independent from the noise processes with $\hat{\mathbf{x}}_0 = [1, 1]^T$ and $\mathbf{P}_0 = 100^2\mathbf{I}_2$.

In Fig. 1, we choose as a performance measure the root MSE (RMSE) of the dimensionless state vector $\tilde{\mathbf{x}}_n = [p_n, Tv_n]^T/\sigma_m$ [40], where $\sigma_m^2 = 0.1^2$ is the steady-state measurement noise variance. Performance results are reported for two full-dimension estimators, that is, the HMM-KF in state-augmented form (SA HMM-KF) applied directly to (13) [37], and the TMM-KF applied to (1) with the parameters in (14).

Also, two RDFs are considered, that is, the modified HMM-KF in [20, p. 301] and the proposed RDF in Algorithm 1 applied to (6), with the parameters in (15). We initialize the proposed RDF with $\mathbf{P}_{1,0|1} = \mathbf{0}_{2\times 2}$, also replacing $\hat{\mathbf{x}}_{0|1}$ and $\mathbf{P}_{0|1}$ (resp., $\hat{\mathbf{x}}_{1|1}$ and $\mathbf{P}_{1|1}$) with $\hat{\mathbf{x}}_0$ and $\mathbf{P}_0$ (resp., with the filtered estimates of the HMM-KF ignoring noise coloration at instant $n = 1$). As shown by Fig. 1, this suboptimal initialization incurs an imperceptible performance loss, thus removing an unnecessary surge in complexity at instant $n = 1$ due to an optimal initialization using the TMM-KF.

Note that the proposed RDF not only outperforms the standard modified HMM-KF over the entire range of ψ, but also has a complexity that is about 26% lower. Furthermore, due to the equivalence of the proposed second-order PMM with the truth model, the proposed RDF is MMSE-optimal with a complexity reduction by about 40% wrt the competing SA HMM-KF and TMM-KF (note that for the sake of fair complexity comparison, we took into account the fact that $\mathbf{w}_n$ can be made single dimensional in the SA HMM-KF and TMM-KF). These results can be extended to tracking in four or six dimensions using xy or xyz coordinate systems.

As noted in [19, Remark 2.2], the original TMM in (1) could also be written as a state-augmented HMM (SA-HMM) to which the modified HMM-KF in [23] can be applied. Unfortunately, the resulting filter diverges, presumably due to a state unobservability issue. This fact can be seen as a further justification for seeking a new RDF in the context of TMMs.

B. One-State Tracking Based on a Full-Dimension Two-State Kinematic Model

We now consider a two-state kinematic model without noise coloration for $\bar{\mathbf{x}}_n = [p_n, v_n]^T$

$$\begin{aligned}\bar{\mathbf{x}}_n &= \bar{\mathbf{F}}_n\bar{\mathbf{x}}_{n-1} + \bar{\mathbf{B}}_n\mathbf{w}_n\\ \mathbf{y}_n &= \bar{\mathbf{H}}_n\bar{\mathbf{x}}_n + \mathbf{v}_n\end{aligned} \tag{34}$$

where

$$\bar{\mathbf{F}}_n = \begin{bmatrix}1 & T\\ 0 & 1\end{bmatrix}, \bar{\mathbf{B}}_n = \begin{bmatrix}\frac{T^2}{2}\\ T\end{bmatrix}, \bar{\mathbf{H}}_n = [1, 0] \tag{35}$$

with $T = 1$ s and $[\mathbf{w}_n^T, \mathbf{v}_n^T]^T \sim \mathcal{N}([0,0]^T, \mathrm{diag}([Q, R]))$ is white.

In order to study the impact of contrasted dynamics, we let $1^2 < Q < 10^2$ $(\mathrm{m/s^2})^2$, whereas $R = 10^2$ m^2. Again, while in simulations the initial state is set to $\bar{\mathbf{x}}_0 = [1,1]^T$, all filters work under the assumption that $\bar{\mathbf{x}}_0 \sim \mathcal{N}(\hat{\bar{\mathbf{x}}}_0, \bar{\mathbf{P}}_0)$ is independent from the noise processes with $\hat{\bar{\mathbf{x}}}_0 = [1,1]^T$ and $\bar{\mathbf{P}}_0 = 100^2\mathbf{I}_2$.

Only the position is estimated using an RDF, whereas the velocity component is ignored (i.e., $\mathbf{x}_n = p_n$ and $\mathbf{r}_n = v_n$); hence, (34) can be written as the TMM in (1) parameterized by

$$\begin{bmatrix}\mathbf{A}_n^{(11)} & \mathbf{A}_n^{(12)} & \mathbf{A}_n^{(13)}\\ \mathbf{A}_n^{(21)} & \mathbf{A}_n^{(22)} & \mathbf{A}_n^{(23)}\\ \mathbf{A}_n^{(31)} & \mathbf{A}_n^{(32)} & \mathbf{A}_n^{(33)}\end{bmatrix} = \begin{bmatrix}1 & T & 0\\ 0 & 1 & 0\\ 1 & T & 0\end{bmatrix}$$

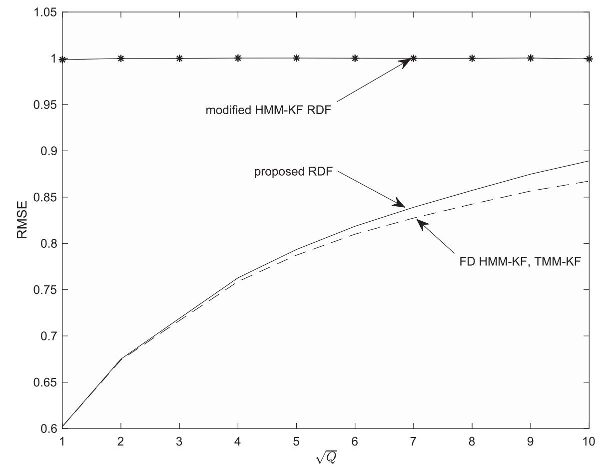

Fig. 2. RMSE for reduced-dimension one-state tracking without noise coloration as a function of $\sqrt{Q}$, with $T = 1$ s and $R = 10^2$ m^2.

$$\begin{bmatrix}\mathbf{B}_n^{(11)} & \mathbf{B}_n^{(12)}\\ \mathbf{B}_n^{(21)} & \mathbf{B}_n^{(22)}\\ \mathbf{B}_n^{(31)} & \mathbf{B}_n^{(32)}\end{bmatrix} = \begin{bmatrix}\frac{T^2}{2} & 0\\ T & 0\\ \frac{T^2}{2} & 1\end{bmatrix}. \tag{36}$$

In order to apply the proposed RDF, this TMM needs to be approximated by a second-order PMM, which can easily be done by replacing $\mathbf{B}_n^{(11)} = \frac{T^2}{2}$ with $\widehat{\mathbf{B}}_n^{(11)} = T^2$ as prescribed by the model perturbation (20). The increased value of $\mathbf{B}_n^{(11)} = \frac{T^2}{2}$ by a factor of two serves as artificial process noise, compensating for the fact that the truth model in reduced-dimension form is not originally second-order Markovian.

In Fig. 2, we choose as a performance measure the RMSE of the dimensionless state $\tilde{\mathbf{x}}_n = p_n/\sigma_m$ [40], where $\sigma_m^2 = R$ is the measurement noise variance. We compare two RDFs, that is, the modified HMM-KF in [23] and the proposed RDF. Similarly to Section VII-A, we initialize the proposed RDF with $\mathbf{P}_{1,0|1} = 0$, also replacing $\hat{\mathbf{x}}_{0|1}$ and $\mathbf{P}_{0|1}$ (resp., $\hat{\mathbf{x}}_{1|1}$ and $\mathbf{P}_{1|1}$) with $\hat{\bar{\mathbf{x}}}_0[1]$ and $\bar{\mathbf{P}}_0[1,1]$ (resp., with the filtered estimates at instant $n = 1$ of an HMM-KF with position-only estimation, assimilating the velocity to an additional source of white noise). Optimal performance results for the full-dimension HMM-KF (FD HMM-KF) applied directly to (34), and the TMM-KF applied to (1) with the parameters in (36) are also reported for reference.

Note that the proposed RDF not only outperforms the standard modified HMM-KF over the entire range of Q, but also has a complexity that is about five times lower. Similar results can be obtained for tracking in two or three dimensions using xy or xyz coordinate systems. We interpret these findings as support for our approach, as the RDFs use different approximations. Indeed, the standard method assumes the position to be first-order Markovian, then modifies the HMM-KF by letting the Kalman gain try to compensate for the fact that the velocity is ignored. In our approach, a small enough perturbation transforms the truth model to an approximate second-order PMM, in which exact filtering is feasible. We note that while our approximation is better, increasing Q still induces a moderate suboptimality increase wrt MMSE-optimal methods (FD HMM-KF, TMM-KF).

TABLE I
RMSE FOR THE HEAT CONDUCTION PROBLEM WITH $l = 10$ (m), $\alpha = 10^{-3}$ (m^2/s), $Q = 8 \times 10^{-7}$, AND $R = 0.1^2$

FD HMM-KF	modified HMM-KF	TMM-KF	Proposed RDF
0.025	0.24	0.025	0.031

This result is expected since increasing Q also increases the artificial process noise in the proposed method.

Again, as noted in [19, Remark 2.2], the original TMM in (1) could also be written as an SA-HMM, to which the modified HMM-KF in [23] can be applied. Unfortunately, the resulting filter diverges, presumably due to a state unobservability issue.

C. 1-D Heat Propagation

We consider the 1-D heat propagation partial differential equation defined over the trip $x \in [0, l]$ at time $t \geq 0$ as [41]

$$\frac{\partial u(t,x)}{\partial t} = \alpha \frac{\partial^2 u(t,x)}{\partial x^2} \tag{37}$$

where α (m^2/s) denotes the thermal diffusivity of the material. We assume the following initial and boundary conditions:

$$\begin{aligned} &u(0,x) = f(x) \\ &u(t,0) = 0, \quad u(t,l) = 0 \quad \forall t \geq 0 \end{aligned} \tag{38}$$

where f is a given function of x, but unknown to an external observer.

We set up a grid in the x, t plane with grid spacing $\Delta x = l/65$ (m) and $\Delta t = 1$ (s), at $x_i = i\Delta x$, $i = 1, 2, \ldots, 64$ and $t_n = n\Delta t$, $n = 1, 2, \ldots$. Letting u_n^i denote the solution of (37)–(38) obtained at $x_i = i\Delta x$ and $t_n = n\Delta t$, we define the full-dimension state by $\bar{\mathbf{x}}_n = [u_n^1, u_n^2, \ldots, u_n^{64}]^T$. The evolution in time of $\bar{\mathbf{x}}_n$ can be modeled using the finite-difference method [41], by replacing $\frac{\partial u(t,x)}{\partial t}|_{t_n, x_i}$ (resp., $\frac{\partial^2 u(t,x)}{\partial x^2}|_{t_n, x_i}$) by a forward difference in time (resp., a central difference in space)

$$\begin{aligned} \bar{\mathbf{x}}_n &= \bar{\mathbf{F}}_n \bar{\mathbf{x}}_{n-1} + \bar{\mathbf{w}}_n \\ \bar{\mathbf{F}}_n &= \begin{bmatrix} 1-2\gamma & \gamma & & 0 \\ \gamma & \ddots & \ddots & \\ & \ddots & \ddots & \gamma \\ 0 & & \gamma & 1-2\gamma \end{bmatrix} \end{aligned} \tag{39}$$

where $\gamma = \alpha\Delta t/(\Delta x)^2$ and $\bar{\mathbf{w}}_n$ is a discretization error affecting all spatial nodes. Furthermore, in a practical situation, the behavior of heat conduction would be restored by an external observer based on noisy temperature sensors, so that $\bar{\mathbf{w}}_n$ is unknown and modeled as a random noise vector. We assume that two such sensors directly observe the temperature at the locations x_{17} and x_{46}, thus generating a 2-D observation vector at instant n

$$\begin{aligned} \mathbf{y}_n &= \bar{\mathbf{H}}_n \bar{\mathbf{x}}_n + \mathbf{v}_n \\ \bar{\mathbf{H}}_n[1,i] &= \begin{cases} 1, \text{when} & i = 17 \\ 0, \text{otherwise} \end{cases} \\ \bar{\mathbf{H}}_n[2,i] &= \begin{cases} 1, \text{when} & i = 46 \\ 0, \text{otherwise} \end{cases} \end{aligned} \tag{40}$$

where $\mathbf{v}_n \sim \mathcal{N}([0,0]^T, R\mathbf{I}_2)$ is the white measurement noise.

We consider the case where the spatial nodes of interest are located on irregular grids, over which the finite-difference method may be neither flexible nor straightforward to apply [42]. For example, we will consider the situation where $\mathbf{r}_n = [u_n^2, u_n^4, \ldots, u_n^{30}, u^{34}, u^{36}, \ldots, u^{62}]$ is the 30-D vector of discarded variables, whereas $\mathbf{x}_n$ is the 34-D vector containing all other variables in $\bar{\mathbf{x}}_n$ corresponding to the desired spatial nodes $\{x'_1 = x_1, x'_2 = x_3, \ldots, x'_{34} = x_{64}\}$. We are now missing a suitable model for the discretization error $\bar{\mathbf{w}}_n$. While estimating the statistics of $\bar{\mathbf{w}}_n$ from observed data would in theory be possible, computing them in great detail would not only be computationally intensive but is also unlikely to pay off. Thus, we resort to simple heuristics instead. In order to comply with the dimension requirements in (1), we write the discretization error as $\bar{\mathbf{w}}_n = \bar{\mathbf{B}}_n \mathbf{w}_n$, where $\mathbf{w}_n$ is a discretization error affecting only the spatial nodes of interest in $\{x'_1, x'_2, \ldots, x'_{34}\}$, assimilated to a zero-mean non-Gaussian white noise parameterized by $\mathbf{Q}_n = E\{\mathbf{w}_n \mathbf{w}_n^T\}$. The matrix $\bar{\mathbf{B}}_n$ draws a one-to-one correspondence between the discretization error for any spatial node to the discretization error affecting the closest retained spatial node, i.e.,

$$\bar{\mathbf{B}}_n[i,j] = \begin{cases} 1, \text{when} & j = \arg\min(x_i - x'_j)^2 \\ 0, \text{otherwise} \end{cases} \tag{41}$$

for $i = 1, 2, 3, \ldots, 64$, and $j = 1, 2, 3, \ldots, 34$.

We now introduce a change of variable using the nonsingular matrix $\mathbf{T} = [\mathbf{S}_x^T, \mathbf{S}_r^T]^T$, where $\mathbf{S}_x$ (resp., $\mathbf{S}_r$) is the selection matrix defined by $\mathbf{x}_n = \mathbf{S}_x \bar{\mathbf{x}}_n$ (resp., $\mathbf{r}_n = \mathbf{S}_r \bar{\mathbf{x}}_n$). It follows that (39)–(40) can be converted to a TMM of the form (1) parameterized by

$$\begin{aligned} \left[\begin{array}{cc|c} \mathbf{A}_n^{(11)} & \mathbf{A}_n^{(12)} & \mathbf{A}_n^{(13)} \\ \mathbf{A}_n^{(21)} & \mathbf{A}_n^{(22)} & \mathbf{A}_n^{(23)} \\ \hline \mathbf{A}_n^{(31)} & \mathbf{A}_n^{(32)} & \mathbf{A}_n^{(33)} \end{array}\right] &= \left[\begin{array}{c|c} \mathbf{T}\bar{\mathbf{F}}_n\mathbf{T}^{-1} & \mathbf{0} \\ \hline \bar{\mathbf{H}}_n\bar{\mathbf{F}}_n\mathbf{T}^{-1} & \mathbf{0} \end{array}\right] \\ \left[\begin{array}{c|c} \mathbf{B}_n^{(11)} & \mathbf{B}_n^{(12)} \\ \mathbf{B}_n^{(21)} & \mathbf{B}_n^{(22)} \\ \hline \mathbf{B}_n^{(31)} & \mathbf{B}_n^{(32)} \end{array}\right] &= \left[\begin{array}{c|c} \mathbf{T}\bar{\mathbf{B}}_n & \mathbf{0} \\ \hline \bar{\mathbf{H}}_n\bar{\mathbf{B}}_n & \mathbf{I}_2 \end{array}\right]. \end{aligned} \tag{42}$$

In order to apply the proposed RDF, this TMM needs to be approximated by a second-order PMM, which can easily be done by replacing $\mathbf{B}_n^{(11)}$ with $\widehat{\mathbf{B}}_n^{(11)}$ and $\mathbf{B}_n^{(21)}$ with $\widehat{\mathbf{B}}_n^{(21)}$ as prescribed by the model perturbation in Section IV-B3.

Our simulations, running over a time window $t \in [0, 2.10^4]$ (s), adopt the setup with $l = 10$ (m), $\alpha = 10^{-3}$ (m^2/s), $R = 0.1^2$, and $f(x) = 100\sin(\pi\frac{x}{l})$. While in simulations the initial state is set to $\bar{\mathbf{x}}_0 = [f(x_1), f(x_2), \ldots, f(x_{64})]^T$, all filters work under the assumption that $\bar{\mathbf{x}}_0$ is independent from the noise processes with $E\{\bar{\mathbf{x}}_0\} = \hat{\bar{\mathbf{x}}}_0 = \mathbf{0}_{64\times 1}$ and $\mathrm{Cov}\{\bar{\mathbf{x}}_0\} = \bar{\mathbf{P}}_0 = 110^2\mathbf{I}_{64}$. For simplicity, we also set $\mathbf{Q}_n = Q\mathbf{I}_{34}$, where Q is set to 8×10^{-7} to account for the maximum discretization error. While it is understood that the external observer only has access to the inferred heat, the analytical heat conduction solution to (37)–(38) expressed as $u(t, x) = f(x)\exp(-\alpha t\pi^2/l^2)$ via Fourier analysis [41] is used as ground truth.

In Table I, we choose as a performance measure the RMSE of the dimensionless state $\tilde{\mathbf{x}}_n = \mathbf{x}_n/\sigma_m$, where $\sigma_m^2 = R$ is the measurement noise variance. We compare two RDFs, that is, the modified HMM-KF in [23] and the proposed RDF. Similarly to Section VII-B, we initialize the proposed RDF with $\mathbf{P}_{1,0|1} = \mathbf{0}_{34\times 34}$, also replacing $\hat{\mathbf{x}}_{0|1}$ and $\mathbf{P}_{0|1}$ (resp., $\hat{\mathbf{x}}_{1|1}$ and $\mathbf{P}_{1|1}$) with $\mathbf{S}_x\hat{\bar{\mathbf{x}}}_0$ and $\mathbf{S}_x\bar{\mathbf{P}}_0\mathbf{S}_x^T$ (resp., with the filtered estimates at instant $n = 1$ of an HMM-KF with $\mathbf{x}_n$-only estimation, assimilating $\mathbf{r}_n$ to an additional source of white noise). Performance results for the FD HMM-KF applied directly to (39)–(40), and the TMM-KF applied to (1) with the parameters in (42) are also reported for reference.

The proposed RDF outperforms the standard modified HMM-KF, but also has a complexity that is about 14 times lower. We interpret these findings in the following way for this example: perturbating the truth model to an approximate second-order PMM in which filtering is feasible is a better option than trying to modify the HMM-KF by letting the Kalman gain try to compensate for the fact that parts of the state variables are discarded. Also, the proposed RDF serves its purpose effectively by saving half the complexity comparing to full-dimension filters (TMM-KF and FD HMM-KF) while retaining a similar level of accuracy in spite of model approximation.

VIII. Conclusion

This article considered the problem of RDF in a family of linear TMMs. The proposed solution relies on the conversion of the original TMM to a lower dimensional second-order PMM, which may be equivalent or approximate. For the sake of linear state estimation, we also derived and analyzed a new MMSE-optimal filter for second-order PMMs.

Numerical results over several realistic applications showed that the proposed RDF can outperform existing RDFs from the literature based on HMMs. This fact demonstrates the interest of changing perspective wrt usual HMM modeling, in the sense that reformulating a problem in TMM form gives rise to an RDF that can be both more accurate and less computationally intensive than existing ones. Moreover, the proposed approach gives a unified view of RDF in linear systems, whereas existing methods have been developed in specific contexts.

Future work will consider the extension of RDF to the recent family of switching systems in which fast optimal filtering is feasible [43], [44].

Appendix A
Proof of Proposition 3.1 and Corollary 3.2

The random process $\{[\mathbf{x}_n^T, \mathbf{y}_n^T]^T\}_{n\geq 2}$ is a PMM of second order driven by the noise process $\{[\mathbf{w}_n^T, \mathbf{v}_n^T]^T\}$ if the contribution of the auxiliary variable $\mathbf{r}_{n-1}$ in (3) reduces to a linear combination of the vectors $[\mathbf{x}_{n-2}^T, \mathbf{y}_{n-2}^T]$ and $[\mathbf{x}_{n-1}^T, \mathbf{y}_{n-1}^T]$. Inspecting the expression of $\mathbf{r}_{n-1}$ in (4), we observe that two subcases need to be addressed separately.

In the first subcase, the noise contribution $[\mathbf{w}_{n-1}^T, \mathbf{v}_{n-1}^T]$ in (4) can be rewritten as a linear combination of $\mathbf{r}_{n-2}$, $[\mathbf{x}_{n-2}^T, \mathbf{y}_{n-2}^T]$ and $[\mathbf{x}_{n-1}^T, \mathbf{y}_{n-1}^T]$, by rewriting (3) as

$$\begin{bmatrix}\mathbf{w}_{n-1}\\ \mathbf{v}_{n-1}\end{bmatrix} = \begin{bmatrix}\mathbf{B}_{n-1}^{(11)} & \mathbf{B}_{n-1}^{(12)}\\ \mathbf{B}_{n-1}^{(31)} & \mathbf{B}_{n-1}^{(32)}\end{bmatrix}^{-1}$$
$$\times\left(\begin{bmatrix}\mathbf{x}_{n-1}\\ \mathbf{y}_{n-1}\end{bmatrix} - \begin{bmatrix}\mathbf{A}_{n-1}^{(11)} & \mathbf{A}_{n-1}^{(13)}\\ \mathbf{A}_{n-1}^{(31)} & \mathbf{A}_{n-1}^{(33)}\end{bmatrix}\begin{bmatrix}\mathbf{x}_{n-2}\\ \mathbf{y}_{n-2}\end{bmatrix} - \begin{bmatrix}\mathbf{A}_{n-1}^{(12)}\\ \mathbf{A}_{n-1}^{(32)}\end{bmatrix}\mathbf{r}_{n-2}\right)$$

which requires that

$$\begin{bmatrix}\mathbf{B}_{n-1}^{(11)} & \mathbf{B}_{n-1}^{(12)}\\ \mathbf{B}_{n-1}^{(31)} & \mathbf{B}_{n-1}^{(32)}\end{bmatrix} \text{ is invertible.}$$

Injecting this result into the expression of $\mathbf{r}_{n-1}$ in (4), $\{[\mathbf{x}_n^T, \mathbf{y}_n^T]^T\}_{n\geq 2}$ is a PMM of second order when the contribution of $\mathbf{r}_{n-2}$ cancels out, i.e., when Condition (ii) is satisfied. Then, developing (3) yields (6) and (7) in Corollary 3.2.

The second subcase corresponds to the fact that $[\mathbf{w}_{n-1}^T, \mathbf{v}_{n-1}^T]$ cannot be rewritten as a linear combination of $\mathbf{r}_{n-2}$, $[\mathbf{x}_{n-2}^T, \mathbf{y}_{n-2}^T]$ and $[\mathbf{x}_{n-1}^T, \mathbf{y}_{n-1}^T]$, which occurs when

$$\begin{bmatrix}\mathbf{B}_{n-1}^{(11)} & \mathbf{B}_{n-1}^{(12)}\\ \mathbf{B}_{n-1}^{(31)} & \mathbf{B}_{n-1}^{(32)}\end{bmatrix} \text{ is not invertible}$$

so that $\{[\mathbf{x}_n^T, \mathbf{y}_n^T]^T\}_{n\geq 2}$ is a PMM of second order when Condition (i) is satisfied.

Appendix B
Proof of the Prediction Error Covariance Recursion (24)

Combining (6) and (22) leads to the following prediction error for any second-order PMM:

$$\begin{bmatrix}\mathbf{x}_n - \hat{\mathbf{x}}_{n|n-1}\\ \mathbf{y}_n - \hat{\mathbf{y}}_{n|n-1}\end{bmatrix} = \begin{bmatrix}\mathbf{A}_n^{(1)} & \mathbf{A}_n^{(2)}\\ \mathbf{A}_n^{(3)} & \mathbf{A}_n^{(4)}\end{bmatrix}\begin{bmatrix}\mathbf{x}_{n-1} - \hat{\mathbf{x}}_{n-1|n-1}\\ \mathbf{0}\end{bmatrix}$$
$$+ \begin{bmatrix}\tilde{\mathbf{A}}_n^{(1)} & \tilde{\mathbf{A}}_n^{(2)}\\ \tilde{\mathbf{A}}_n^{(3)} & \tilde{\mathbf{A}}_n^{(4)}\end{bmatrix}\begin{bmatrix}\mathbf{x}_{n-2} - \hat{\mathbf{x}}_{n-2|n-1}\\ \mathbf{0}\end{bmatrix}$$
$$+ \begin{bmatrix}\mathbf{B}_n^{(11)} & \mathbf{B}_n^{(12)}\\ \mathbf{B}_n^{(31)} & \mathbf{B}_n^{(32)}\end{bmatrix}\begin{bmatrix}\mathbf{w}_n\\ \mathbf{v}_n\end{bmatrix} \tag{43}$$

so that the last equation in (24) readily follows from the fact that $[\mathbf{w}_n^T, \mathbf{v}_n^T]^T$ is a zero-mean white noise.

In the same way, injecting (43) into (23), the expression for the predicted error covariance is given as follows:

$$E\left\{\begin{bmatrix}\mathbf{x}_n - \hat{\mathbf{x}}_{n|n-1}\\ \mathbf{y}_n - \hat{\mathbf{y}}_{n|n-1}\end{bmatrix}\begin{bmatrix}\mathbf{x}_n - \hat{\mathbf{x}}_{n|n-1}\\ \mathbf{y}_n - \hat{\mathbf{y}}_{n|n-1}\end{bmatrix}^T\right\}$$
$$= \begin{bmatrix}\mathbf{A}_n^{(1)} & \mathbf{A}_n^{(2)}\\ \mathbf{A}_n^{(3)} & \mathbf{A}_n^{(4)}\end{bmatrix}\begin{bmatrix}\mathbf{P}_{n-1|n-1} & \mathbf{0}\\ \mathbf{0} & \mathbf{0}\end{bmatrix}\begin{bmatrix}\mathbf{A}_n^{(1)} & \mathbf{A}_n^{(2)}\\ \mathbf{A}_n^{(3)} & \mathbf{A}_n^{(4)}\end{bmatrix}^T$$
$$+ \begin{bmatrix}\tilde{\mathbf{A}}_n^{(1)} & \tilde{\mathbf{A}}_n^{(2)}\\ \tilde{\mathbf{A}}_n^{(3)} & \tilde{\mathbf{A}}_n^{(4)}\end{bmatrix}\begin{bmatrix}\mathbf{P}_{n-2|n-1} & \mathbf{0}\\ \mathbf{0} & \mathbf{0}\end{bmatrix}\begin{bmatrix}\tilde{\mathbf{A}}_n^{(1)} & \tilde{\mathbf{A}}_n^{(2)}\\ \tilde{\mathbf{A}}_n^{(3)} & \tilde{\mathbf{A}}_n^{(4)}\end{bmatrix}^T$$

$$+\begin{bmatrix}\mathbf{A}_n^{(1)} & \mathbf{A}_n^{(2)}\\ \mathbf{A}_n^{(3)} & \mathbf{A}_n^{(4)}\end{bmatrix}\begin{bmatrix}\mathbf{P}_{n-1,n-2|n-1} & \mathbf{0}\\ \mathbf{0} & \mathbf{0}\end{bmatrix}\begin{bmatrix}\tilde{\mathbf{A}}_n^{(1)} & \tilde{\mathbf{A}}_n^{(2)}\\ \tilde{\mathbf{A}}_n^{(3)} & \tilde{\mathbf{A}}_n^{(4)}\end{bmatrix}^T$$
$$+\begin{bmatrix}\tilde{\mathbf{A}}_n^{(1)} & \tilde{\mathbf{A}}_n^{(2)}\\ \tilde{\mathbf{A}}_n^{(3)} & \tilde{\mathbf{A}}_n^{(4)}\end{bmatrix}\begin{bmatrix}\mathbf{P}^T_{n-1,n-2|n-1} & \mathbf{0}\\ \mathbf{0} & \mathbf{0}\end{bmatrix}\begin{bmatrix}\mathbf{A}_n^{(1)} & \mathbf{A}_n^{(2)}\\ \mathbf{A}_n^{(3)} & \mathbf{A}_n^{(4)}\end{bmatrix}^T$$
$$+\begin{bmatrix}\mathbf{B}_n^{(11)} & \mathbf{B}_n^{(12)}\\ \mathbf{B}_n^{(31)} & \mathbf{B}_n^{(32)}\end{bmatrix}\begin{bmatrix}\mathbf{Q}_n & \mathbf{0}\\ \mathbf{0} & \mathbf{R}_n\end{bmatrix}\begin{bmatrix}\mathbf{B}_n^{(11)} & \mathbf{B}_n^{(12)}\\ \mathbf{B}_n^{(31)} & \mathbf{B}_n^{(32)}\end{bmatrix}^T. \quad (44)$$

Developing this expression yields the desired recursion for $\mathbf{P}_{n|n-1}$, $\mathbf{\Sigma}_{n|n-1}$, and $\mathbf{L}_{n|n-1}$.

Appendix C
Proof of the Update Error Covariance Recursions in (27) and (28)

Injecting (22) and the second line of (6) into (26) results in

$$\begin{aligned}\boldsymbol{\epsilon}_{n|n} &= (\mathbf{A}_n^{(1)} - \mathbf{K}_n\mathbf{A}_n^{(3)})\boldsymbol{\epsilon}_{n-1|n-1}\\ &\quad + (\tilde{\mathbf{A}}_n^{(1)} - \mathbf{K}_n\tilde{\mathbf{A}}_n^{(3)})\boldsymbol{\epsilon}_{n-2|n-1}\\ &\quad + (\mathbf{B}_n^{(1)} - \mathbf{K}_n\mathbf{B}_n^{(3)})\mathbf{w}_n + (\mathbf{B}_n^{(2)} - \mathbf{K}_n\mathbf{B}_n^{(4)})\mathbf{v}_n\\ \boldsymbol{\epsilon}_{n-1|n} &= (\mathbf{I}_K - \mathbf{J}_n\mathbf{A}_n^{(3)})\boldsymbol{\epsilon}_{n-1|n-1} - \mathbf{J}_n\tilde{\mathbf{A}}_n^{(3)}\boldsymbol{\epsilon}_{n-2|n-1}\\ &\quad - \mathbf{J}_n(\mathbf{B}_n^{(3)}\mathbf{w}_n + \mathbf{B}_n^{(4)}\mathbf{v}_n).\end{aligned}$$

The covariance of these estimation errors can be written as

$$\begin{aligned}\mathbf{P}_{n|n} &= (\mathbf{A}_n^{(1)} - \mathbf{K}_n\mathbf{A}_n^{(3)})\mathbf{P}_{n-1|n-1}(\mathbf{A}_n^{(1)} - \mathbf{K}_n\mathbf{A}_n^{(3)})^T\\ &\quad + (\tilde{\mathbf{A}}_n^{(1)} - \mathbf{K}_n\tilde{\mathbf{A}}_n^{(3)})\mathbf{P}_{n-2|n-1}(\tilde{\mathbf{A}}_n^{(1)} - \mathbf{K}_n\tilde{\mathbf{A}}_n^{(3)})^T\\ &\quad + (\mathbf{A}_n^{(1)} - \mathbf{K}_n\mathbf{A}_n^{(3)})\mathbf{P}_{n-1,n-2|n-1}(\tilde{\mathbf{A}}_n^{(1)} - \mathbf{K}_n\tilde{\mathbf{A}}_n^{(3)})^T\\ &\quad + (\tilde{\mathbf{A}}_n^{(1)} - \mathbf{K}_n\tilde{\mathbf{A}}_n^{(3)})\mathbf{P}^T_{n-1,n-2|n-1}(\mathbf{A}_n^{(1)} - \mathbf{K}_n\mathbf{A}_n^{(3)})^T\\ &\quad + (\mathbf{B}_n^{(1)} - \mathbf{K}_n\mathbf{B}_n^{(3)})\mathbf{Q}_n(\mathbf{B}_n^{(1)} - \mathbf{K}_n\mathbf{B}_n^{(3)})^T\\ &\quad + (\mathbf{B}_n^{(2)} - \mathbf{K}_n\mathbf{B}_n^{(4)})\mathbf{R}_n(\mathbf{B}_n^{(2)} - \mathbf{K}_n\mathbf{B}_n^{(4)})^T\\ \mathbf{P}_{n-1|n} &= (\mathbf{I}_K - \mathbf{J}_n\mathbf{A}_n^{(3)})\mathbf{P}_{n-1|n-1}(\mathbf{I}_K - \mathbf{J}_n\mathbf{A}_n^{(3)})^T\\ &\quad + \mathbf{J}_n\tilde{\mathbf{A}}_n^{(3)}\mathbf{P}_{n-2|n-1}(\mathbf{J}_n\tilde{\mathbf{A}}_n^{(3)})^T\\ &\quad - (\mathbf{I}_K - \mathbf{J}_n\mathbf{A}_n^{(3)})\mathbf{P}_{n-1,n-2|n-1}(\mathbf{J}_n\tilde{\mathbf{A}}_n^{(3)})^T\\ &\quad - \mathbf{J}_n\tilde{\mathbf{A}}_n^{(3)}\mathbf{P}^T_{n-1,n-2|n-1}(\mathbf{I}_K - \mathbf{J}_n\mathbf{A}_n^{(3)})^T\\ &\quad + \mathbf{J}_n(\mathbf{B}_n^{(3)}\mathbf{Q}_n\mathbf{B}_n^{(3)T} + \mathbf{B}_n^{(4)}\mathbf{R}_n\mathbf{B}_n^{(4)T})\mathbf{J}_n^T. \quad (45)\end{aligned}$$

Developing (45) and rearranging the terms using (24) completes the proof of (27).

Now, by definition $\mathbf{P}_{n,n-1|n} = E\{\boldsymbol{\epsilon}_{n|n}\boldsymbol{\epsilon}^T_{n-1|n}\}$, so that

$$\begin{aligned}\mathbf{P}_{n,n-1|n} &= (\mathbf{A}_n^{(1)} - \mathbf{K}_n\mathbf{A}_n^{(3)})\mathbf{P}_{n-1|n-1}\\ &\quad + (\tilde{\mathbf{A}}_n^{(1)} - \mathbf{K}_n\tilde{\mathbf{A}}_n^{(3)})\mathbf{P}^T_{n-1,n-2|n-1}\\ &\quad - \Big[(\mathbf{A}_n^{(1)} - \mathbf{K}_n\mathbf{A}_n^{(3)})(\mathbf{P}_{n-1|n-1}\mathbf{A}_n^{(3)T}\\ &\quad + \mathbf{P}_{n-1,n-2|n-1}\tilde{\mathbf{A}}_n^{(3)T}) + (\tilde{\mathbf{A}}_n^{(1)} - \mathbf{K}_n\tilde{\mathbf{A}}_n^{(3)})\\ &\quad \times (\mathbf{P}^T_{n-1,n-2|n-1}\mathbf{A}_n^{(3)T} + \mathbf{P}_{n-2|n-1}\tilde{\mathbf{A}}_n^{(3)T})\\ &\quad + (\mathbf{B}_n^{(1)} - \mathbf{K}_n\mathbf{B}_n^{(3)})\mathbf{Q}_n\mathbf{B}_n^{(3)T}\\ &\quad + (\mathbf{B}_n^{(2)} - \mathbf{K}_n\mathbf{B}_n^{(4)})\mathbf{R}_n\mathbf{B}_n^{(4)T}\Big]\mathbf{J}_n^T. \quad (46)\end{aligned}$$

Developing the bracket in (46) and rearranging the terms using (24) completes the proof of (28).

References

[1] T. Dietzen, S. Doclo, M. Moonen and T. van Waterschoot, "Integrated sidelobe cancellation and linear prediction Kalman filter for joint multi-microphone speech dereverberation, interfering speech cancellation, and noise reduction," *IEEE/ACM Trans. Audio, Speech, Lang. Process.*, vol. 28, pp. 740–754, Jan. 2020.

[2] R. L. Smith, A. A. A. Rahni, J. Jones, and K. Wells, "A Kalman-based approach with EM optimization for respiratory motion modeling in medical imaging," *IEEE Trans. Radiat. Plasma Med. Sci.*, vol. 3, no. 4, pp. 410–420, Jul. 2019.

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

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

[5] K. Muralidhar and D. Sreedhar, "Pilot design for vector state-scalar observation Kalman channel estimators in doubly-selective MIMO-OFDM systems," *IEEE Wireless Commun. Lett.*, vol. 2, no. 2, pp. 147–150, Apr. 2013.

[6] L. Qian, H. Wang, and E. R. Dougherty, "Inference of noisy nonlinear differential equation models for gene regulatory networks using genetic programming and Kalman filtering," *IEEE Trans. Signal Process.*, vol. 56, no. 7, pp. 3327–3339, Jul. 2008.

[7] M. Weiss *et al.*, "Applications of the Kalman filter to chemical sensors for downstream machine learning," *IEEE Sensors J.*, vol. 18, no. 13, pp. 5455–5463, Jul. 2018.

[8] J. M. Kanieski, R. V. Tambara, H. Pinheiro, R. Cardoso and H. A. Gründling, "Robust adaptive controller combined with a linear quadratic regulator based on Kalman filtering," *IEEE Trans. Autom. Control*, vol. 61, no. 5, pp. 1373–1378, May 2016.

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

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

[11] B. Ait-El-Fquih and I. Hoteit, "An efficient state-parameter filtering scheme combining ensemble Kalman and particle filters," *Monthly Weather Rev.*, vol. 146, pp. 871–887, Mar. 2018.

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

[13] 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.

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

[15] Y. Petetin and F. Desbouvries, "Exact Bayesian estimation in constrained triplet Markov chains," in *Proc. IEEE Int. Workshop Mach. Learn. Signal Process.*, Reims, France, Sep. 2014, pp. 1–16.

[16] 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.

[17] 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.

[18] 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.

[19] F. Lehmann and W. Pieczynski, "State estimation in pairwise Markov models with improved robustness using unbiased FIR filtering," Signal Process., vol. 172, Jul. 2020, Art. no. 107568.

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

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

[22] C. S. Sims, "An algorithm for estimating a portion of the state," *IEEE Trans. Autom. Control*, vol. AC-19, no. 4, pp. 391–393, Aug. 1974.

[23] C. E. Hutchinson, J. A. D'Appolito, and K. J. Roy, "Applications of minimum variance reduced-state estimators," *IEEE Trans. Aerosp. Electron. Syst.*, vol. AES-11, no. 5, pp. 785–794, Sep. 1975.

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

[25] M. A. Cane, A. Kaplan, R. N. Miller, B. Tang, E. C. Hackert, and A. J. Busalacchi, "Mapping tropical pacific sea level: Data assimilation via a reduced state space Kalman filter," *J. Geophys. Res.*, vol. 101, no. C10, pp. 22599–22617, Oct. 1996.

[26] A. Voutilainen, T. Pyhälahti, K. Y. Kallio, J. Poulliainen, H. Haario, and J. P. Kaipio, "A filtering approach for estimating lake water quality from remote sensing data," *Int. J. Appl. Earth Observ. Geoinf.*, vol. 9, no. 1, pp. 50–64, Feb. 2007.

[27] A. Solonen, T. Cui, J. Hakkarainen, and Y. Marzouk, "On dimension reduction in Gaussian filters," *Inverse Problems*, vol. 32, no. 4, Apr. 2016, Art. no. 045003.

[28] M. Fisher and E. Andersson, *Developments in 4D-Var and Kalman filtering*, Eur. Centre Medium Range Weather Forecasts, Reading, U.K., Tech. Memorandum 347, pp. 1–36, Sep. 2001.

[29] F. Lehmann and W. Pieczynski, "Suboptimal Kalman filtering in triplet Markov models using model order reduction," *IEEE Signal Process. Lett.*, vol. 27, pp. 1100–1104, Jun. 2020.

[30] T.-T. Lu and S.-H. Shiou, "Inverses of 2×2 block matrices," *Comput. Math. Appl.*, vol. 43, no. 1/2, pp. 119–129, Jan. 2002.

[31] J. Zhang *et al.*, "Southern California permanent GPS geodetic array: Error analysis of daily position estimates and site velocities," *J. Geophys. Res.*, vol. 102, no. B8, pp. 18035–18055, Aug. 1997.

[32] J. Wendel and G. F. Trommer, "An efficient method for considering time correlated noise in GPS/INS integration," in *Proc. Nat. Tech. Meeting Inst. Navig.*, San Diego, CA, USA, Jan. 2004, pp. 903–911.

[33] N. El-Sheimy, H. Hou, and X. Niu, "Analysis and modeling of inertial sensors using Allan variance," *IEEE Trans. Intrum. Meas.*, vol. 57, no. 1, pp. 140–149, Jan. 2008.

[34] D. Simon, *Optimal State Estimation: Kalman, H_∞ and Nonlinear Approaches*. Hoboken, NJ, USA: Wiley, 2006.

[35] A. E. Bryson and L. J. Henrikson, "Estimation using sampled data containing sequentially correlated noise," *J. Spacecraft Rockets*, vol. 5, no. 6, pp. 662–665, Jun. 1968.

[36] K. Lee and E. N. Johnson, "State estimation using Gaussian process regression for colored noise systems," in *Proc. IEEE Aerosp. Conf.*, Big Sky, MT, USA, 2017, pp. 1–8.

[37] G. Chen and C. K. Cui, "Design of near-optimal linear digital tracking filters with colored input," *J. Comput. Appl. Math.*, vol. 15, no. 3, pp. 353–370, Jul. 1986.

[38] A. H. Jazwinski, *Stochastic Processes and Filtering Theory*. New York, NY, USA: Academic, 1970.

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

[40] 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.

[41] K. W. Morton and D. F. Mayers, *Numerical Solution of Partial Differential Equations: An Introduction*. Cambridge, U.K.: Cambridge Univ. Press, 1994.

[42] A. E. P. Veldman and K. Rinzema, "Playing with nonuniform grids," *J. Eng. Math.*, vol. 26, no. 1, pp. 119–130, 1992.

[43] I. Gorynin, S. Derrode, E. Monfrini, and W. Pieczynski, "Fast filtering in switching approximations of nonlinear Markov systems with applications to stochastic volatility," *IEEE Trans. Autom. Control*, vol. 62, no. 2, pp. 853–862, Feb. 2017.

[44] F. Zheng, S. Derrode, and W. Pieczynski, "Parameter estimation in switching Markov systems and unsupervised smoothing," *IEEE Trans. Autom. Control*, vol. 64, no. 4, pp. 1761–1767, Apr. 2019.

Frederic Lehmann received the E.E. and M.S.E.E. degrees in microelectronics and signal processing from Ecole Nationale Superieure d'Electronique et de Radioelectricite de Grenoble, Grenoble, France, in 1998, and the Ph.D. degree in electrical engineering from the National Polytechnical Institute, Grenoble, France, in 2002.

From 1999 to 2002, he was a Research Engineer with ST Microelectronics. From 2003 to 2004, he was a Postdoctoral Researcher with the Laboratory for Analysis and Architecture of Systems, Centre national de la recherche scientifique, Toulouse, France. He is currently a Professor with Télécom SudParis, Institut Polytechnique de Paris, Palaiseau, France. His research interests include communication theory, nonlinear signal processing, and statistical image processing.

Wojciech Pieczynski received the Doctorat d'Etat degree in mathematical statistics from the Université Pierre et Marie Curie, Paris, France, in 1986.

He is currently a Professor with Télécom Sudparis, Institut Polytechnique de Paris, Palaiseau, France, and Head of the Communications, Image, and Information Processing Department. His research interests include mathematical statistics, stochastic processes, and statistical signal processing.