Adaptive IMM Smoothing Algorithms for Jumping Markov System With Mismatched Measurement Noise Covariance Matrix

HONG XU, Member, IEEE
Xidian University, Hangzhou, China

QIN PAN, Student Member, IEEE
HENG XU, Student Member, IEEE
YINGHUI QUAN, Senior Member, IEEE
Xidian University, Xi'an, China

In this article, the adaptive online state smoothing problem is studied for a Markov jump system where the measurement noise covariance matrix (MNCM) is unknown. To address this problem, two adaptive interacting multiple model online smoothing algorithms are proposed to jointly estimate the target state and the unknown MNCM. Specifically, the article quantitatively examines the impact of noise covariance matrix mismatch on state estimation and theoretically demonstrates that the joint posterior distribution of the target state and the MNCM cannot be analytically obtained using the original variational Bayesian (VB) approach. To overcome this limitation, an approximate VB method is introduced, which utilizes the approximated state distribution obtained through the moment-matching method to update the MNCM. In addition, the convergence criterion of the proposed adaptive smoothing algorithms is designed. Finally, the estimation consistency of MNCM is analyzed. A maneuvering target tracking simulation example is presented to evaluate the effectiveness and applicability of the proposed adaptive algorithms.

Manuscript received 4 September 2023; revised 4 January 2024 and 3 April 2024; accepted 17 April 2024. Date of publication 23 April 2024; date of current version 9 August 2024.

DOI. No. 10.1109/TAES.2024.3392552

Refereeing of this contribution was handled by M. Kok.

This work was supported in part by the Chinese Postdoctoral Science Foundation under Grant 2022M722503, in part by the Fundamental Research Funds for the Central Universities, in part by the Innovation Fund of Xidian University, and in part by the National Natural Science Foundation of China under Grant 62301408 and Grant 62331019.

Authors' addresses: Hong Xu is with the Hangzhou Research Institute of Xidian University, Hangzhou 311231, China, and also with the Xi'an Key Laboratory of Advanced Remote Sensing, School of Electronic Engineering, Xidian University, Xi'an 710071, China, E-mail: (xuhong@xidian.edu.cn); Qin Pan, Heng Xu and Yinghui Quan are with the Department of remote sensing science and technology, School of Electronic Engineering, Xidian University, also with the Xi'an Key Laboratory of Advanced Remote Sensing, School of Electronic Engineering, Xidian University, and also with the Key Laboratory of Collaborative Intelligence Systems, Ministry of Education, School of Electronic Engineering, Xidian University, Xi'an 710071, China, E-mail: (qinp@stu.xidian.edu.cn; hengx@stu.xidian.edu.cn;yhquan@mail.xidian.edu.cn). *(Hong Xu and Qin Pan contributed equally to this work.) (Corresponding authors: Qin Pan; Yinghui Quan.)*

I. INTRODUCTION

State estimation for a jumping Markov system (JMS) is concerned with estimating the hidden system state from noisy measurements with hybrid system models, and it has been widely used in practical applications, such as target tracking, control systems, and fault detection, to name but a few [1], [2], [3]. The JMS is represented by a finite set of dynamics, which are controlled by a Markov chain. This representation captures the system's ability to transition between different modes or behaviors. However, finding the optimal Bayesian estimator for JMS is not feasible due to the NP-hard nature of the problem [4]. Consequently, researchers commonly employ approximate methods or heuristic approaches to address the estimation problem effectively, among which the interacting multiple model (IMM) is considered a preferable algorithm because it achieves the tradeoff between estimation accuracy and algorithm complexity [5], [6], [7]. Conventional IMM-based estimators operate within a model-driven framework, assuming a priori knowledge of the measurement noise covariance matrix (MNCM). In real-world applications, the MNCM is usually time-varying and difficult to know a priori due to nonideal factors [8], [9]. Under such circumstances, the performance of the IMM may deteriorate, leading to a decrease in accuracy or even rendering the estimator invalid.

A. Related Works

To improve the estimation performance in the presence of unknown MNCM, adaptive state estimation methods have been introduced to jointly estimate the state and the unknown parameters of MNCM. Typical methods to address this challenge include covariance matching [10], [11], maximum likelihood [12], [13], and the Bayesian approach [14], [15], etc. It should be emphasized that the Bayesian method is considered to be the optimal approach due to its ability to estimate the joint posterior distribution of the state and the unknown parameters. However, in general cases, calculating the joint posterior probability density distribution using the Bayesian theorem is computationally challenging, necessitating the use of approximate inference methods [15], [16].

In recent years, there has been significant interest in variational Bayesian (VB) [17], [18]-based adaptive state estimation methods, primarily because of their capacity to offer an efficient Bayesian probability inference framework [15], [16]. It has been extended to consider the nonlinear systems [19], [20], nonGaussian conditions [21],

[22], multitarget tracking [23], etc. For the adaptive state estimation for the JMS, a VB-based IMM algorithm in the presence of the unknown process and measurement noise covariance is introduced in [24]. In [25], it further considers the problem in the presence of missing measurements. In [26], the VB and maximum correntropy-based IMM algorithm is developed to solve the nonlinear filtering problem with unknown non-Gaussian noise. A distributed filtering problem for a JMS over sensor networks is investigated [27], in which a VB-based robust IMM approach is introduced to identify and exclude measurement outliers automatically. It should be emphasized that the above methods only consider the adaptive state filtering problem in the presence of the MNCM. In our previous work [28], the online adaptive state smoothing problem with a single-state space model in the presence of unknown MNCM was studied. However, the algorithm is not suitable for the JMS, since it only uses a single-motion model. To the best of our knowledge, the online state smoothing problem for the JMS in the presence of the MNCM has not been reported, which motivates our research. To address this problem, this article aims to improve the state estimation accuracy of JMS by using multiple time measurements based on the fixed-lag smoothing (FLS) [29] framework.

FLS achieves improved estimation performance by utilizing multiple time measurements at the cost of a certain time delay [30]. In most cases, employing a small smoothing lag allows FLS to achieve estimation performance comparable to that of fixed-interval smoothing (FIS). Due to this fact, the FLS methods have been widely applied to various state smoothing problems [31], [32], [29]. In the context of online state smoothing for JMS, an effective approach is to integrate the IMM algorithm with the FLS method. In [33], a suboptimal method suitable for FLS with JMS is introduced by using a heuristic hypothesis-pruning approach. In [34], the one-step FLS problem for JMS is examined. To further improve the state estimation performance, a state-augmentation IMM (AS-IMM) algorithm [35] is developed to form a fixed-lag IMM smoother, and it has been extended to consider nonlinear systems [36]. In [37], a Rauch–Tung–Striebel (RTS)-type IMM (RTSt-IMM) algorithm was proposed to implement FLS, involving forward filtering followed by backward smoothing. It should be pointed out that AS-IMM and RTSt-IMM are both model-based methods, and it assumes that the MNCM is known in advance. As mentioned earlier, model-driven state estimation methods are not suitable for scenarios where the MNCM is unknown or time-varying. Therefore, the AS-IMM and RTSt-IMM algorithms are also sensitive to noise models.

B. Contributions of This Work

Based on the aforementioned works, it is clear that the existing AS-IMM and RTSt-IMM methods are both model-based approaches. Therefore, in the presence of unknown MNCM, the state estimation performance of these algorithms may degrade or fail. The existing VB adaptive state filtering method applied to JMS cannot be directly applied to FLS problems for several reasons: on one hand, FLS necessitates the estimation of target state variables and MNCM across multiple time instants, rendering the probabilistic graphical model for variational optimization more complex with a greater number of variables to be estimated, thus increasing the difficulty of solving the problem (as detailed in Theorem 2 and Theorem 3). On the other hand, for FLS problems using the RTSt method, joint estimation of states and MNCM requires knowledge of the Markov transition model of MNCM, which is actually unknown (as discussed in detail in Section III-D).

To better solve the online state smoothing in the presence of unknown MNCM, we propose adaptive IMM smoothing algorithms by the VB framework. The main contributions of this article are given as follows.

1) We examine the quantitative impact of noise covariance matrix mismatch on state estimation. To be more specific, the difference in error covariance matrix under mismatched noise covariance matrices is analyzed in a quantitative manner.
2) We demonstrate that the VB optimization method for approximating the joint posterior distribution of the JMS state and the MNCM is nonconjugate. To address this, we introduce an efficient and effective approximate method. By utilizing the approximated state distribution (a single Gaussian computed using the moment matching method) to update the MNCM, the posterior distribution of the state and MNCM can be effectively calculated. Building upon this method, we propose two adaptive IMM smoothing algorithms: one based on the AS-IMM algorithm and the other based on the RTSt-IMM algorithm.
3) We discuss the consistency of the proposed methods for estimating MNCM and design simulation experiments to validate the algorithms' performance.

C. Paper Structure and Notations

The rest of this article is organized as follows. Section II summarizes the background and provides a quantitative analysis of the impact of noise covariance matrix mismatch on state estimation. Section III presents the approximate method, two adaptive IMM smoothing algorithms and the estimation consistency of MNCM in detail, which are the main contributions of this article. Section IV presents the advantages of the proposed algorithms through comparative simulation. Finally, Section V concludes this article.

NOTATIONS $(\cdot)^{\mathrm{T}}$ denotes transpose, $\mathrm{tr}(\cdot)$ denotes the trace of a square matrix, $|\cdot|$ is the modulus of a matrix, $\|\cdot\|$ denote any matrix norm, $\circ$ is the Hadamard product, $\mathbb{E}[\cdot]$ stands for the statistical expectation operation, $\mathrm{KL}[q(\cdot)\|p(\cdot)]$ represents the Kullback–Leibler divergence (KLD) between $q(\cdot)$ and $p(\cdot)$, $\mathrm{diag}(\cdot)$ represents the vector formed by collecting the diagonal entries of the matrix argument. $\mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ denotes a Gaussian distribution with a mean $\boldsymbol{\mu}$ and a covariance matrix $\boldsymbol{\Sigma}$, $\mathcal{IW}(v, \mathbf{V})$ denotes an inverse-Wishart distribution with v degrees of freedom (DOFs) and a scale

matrix $\mathbf{V}$. The function outer$(\cdot)$ is utilized to represent the outer product operation. Specifically, outer$(\mathbf{x})$ is defined as $\mathbf{x}\mathbf{x}^{\mathrm{T}}$.

II. PROBLEM FORMULATION

A. JMS Models for State Estimation

For the target state estimation problem, this article considers a linear JMS containing M models, where the state-space model of the jth model is defined as follows:

$$\mathbf{x}_k = \mathbf{F}_{k-1}^{j}\mathbf{x}_{k-1} + \mathbf{v}_{k-1}^{j}, \quad \mathbf{v}_{k-1}^{j} \overset{i.i.d.}{\sim} \mathcal{N}\left(\mathbf{0}, \mathbf{Q}_{k-1}^{j}\right) \tag{1}$$

$$\mathbf{z}_k = \mathbf{H}_{k}^{j}\mathbf{x}_k + \mathbf{n}_k, \quad \mathbf{n}_k \overset{i.i.d.}{\sim} \mathcal{N}\left(\mathbf{0}, \mathbf{R}_k\right). \tag{2}$$

Here, $k \in \mathbb{N}$ represents the discrete time index, $\mathbf{x}_k \in \mathbb{R}^{n_x}$ and $\mathbf{z}_k \in \mathbb{R}^{n_z}$ represent the state vector and measurement vector, respectively. $\mathbf{F}_{k-1}^{j} \in \mathbb{R}^{n_x \times n_x}$ and $\mathbf{H}_{k}^{j} \in \mathbb{R}^{n_z \times n_x}$ indicate the state transition matrix and the measurement matrix, individually. The process noise vector $\mathbf{v}_{k-1}^{j}$ and measurement noise vector $\mathbf{n}_k$ are uncorrelated Gaussian sequences with covariance matrices $\mathbf{Q}_{k-1}^{j}$ and $\mathbf{R}_k$. The prior probability $P(m_0^j)$ of the jth model and the Markov transition probability from model i to model j are given by

$$P\left(m_0^j\right) = \mu_0^j; \quad P\left(m_k^j \left| m_{k-1}^i \right.\right) = \pi_{ij} \tag{3}$$

where m_k^j denotes the jth model at time index k. It should be noted that the definition of JMS models in (1) and (2) imply an assumption that the MNCMs of different motion models are consistent, i.e., $\mathbf{R}_k^i = \mathbf{R}_k^j = \mathbf{R}_k \ \forall i \neq j \in [1, M]$. The reasons lie in the following: 1) The JMS models defined in the article mainly represent the diversity of the maneuvering motion for the target and 2) the MNCM is mainly determined by the signal-to-noise ratio, the detection accuracy of the sensor, and nonideal factors, which can be considered uncorrelated to the target motion. In this article, we focus on the slowly time-varying MNCM [15], [38] with the assumption that $\mathbf{R}_k$ is unknown.

The online state smoothing estimation for JMS models can be achieved by the AS-IMM and RTSt-IMM algorithms. To get a better insight, we briefly introduce the above algorithms. The basic idea in the AS-IMM algorithm is to augment the system state in the smoothing window to form an augmented JMS model. Define the augmented state $\mathbf{X}_k$ as $\mathbf{X}_k = [\mathbf{x}_k^{\mathrm{T}} \quad \mathbf{x}_{k-1}^{\mathrm{T}} \quad \cdots \quad \mathbf{x}_{k-L}^{\mathrm{T}}]^{\mathrm{T}}$, the augmented JMS model can be expressed by

$$\begin{aligned}\mathbf{X}_k &= \mathbf{F}_{k-1}^{a(j)}\mathbf{X}_{k-1} + \Gamma\mathbf{v}_{k-1} \\ &= \begin{bmatrix} \mathbf{F}_{k-1}^{j} & \mathbf{0} & \mathbf{0} \\ \mathbf{I}_{n_x} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{E}_{n_x(L-1)} & \mathbf{0} \end{bmatrix}\mathbf{X}_{k-1} + \begin{bmatrix} \mathbf{I}_{n_x} \\ \mathbf{0} \\ \mathbf{0} \end{bmatrix}\mathbf{v}_{k-1}\end{aligned} \tag{4}$$

$$\mathbf{z}_k = \mathbf{H}_k^{a(j)}\mathbf{X}_k + \mathbf{n}_k = \mathbf{H}_k^{j}\begin{bmatrix}\mathbf{I}_{n_x} & \mathbf{0} & \mathbf{0}\end{bmatrix}\mathbf{X}_k + \mathbf{n}_k \tag{5}$$

where $j \in [1, M]$. The superscript a stands for augmentation, indicating $\mathbf{F}_{k-1}^{a(j)} \in \mathbb{R}^{[n_z(L+1)]\times[n_x(L+1)]}$ as the augmented form of $\mathbf{F}_{k-1}^{j}$, and this notation is consistently extended throughout the entire article. $\mathbf{I}_{n_x} \in \mathbb{R}^{n_z \times n_x}$ represents the identity matrix and $\mathbf{E}_{n_x(L-1)} \in \mathbb{R}^{[n_z(L-1)]\times[n_x(L-1)]}$ is composed of $L-1$ copies of the n_x-dimensional identity matrix. The $\mathbf{0}$ denotes the zero matrix of appropriate dimensions in (4) and (5). Theoretically, for a given model m_k^j, it is observed that the conditional posterior probability density distribution of the augmented state vector denoted as $p(\mathbf{X}_k|m_k^j, \mathbf{z}_{1:k})$, is constituted by a Gaussian mixture distribution. This distribution comprises $M \times L$ components within the smoothing interval $[k-L, k]$. To reduce the computational complexity, in the AS-IMM, $p(\mathbf{X}_k|m_k^j, \mathbf{z}_{1:k})$ is approximated to be a Gaussian. Thus, the posterior distribution of the augmented state $\mathbf{X}_k$, i.e., $p(\mathbf{X}_k|\mathbf{z}_{1:k})$ can be approximated by

$$p\left(\mathbf{X}_k \left| \mathbf{z}_{1:k} \right.\right) \approx \sum_{j=1}^{M} p\left(\mathbf{X}_k \left| m_k^j, \mathbf{z}_{1:k} \right.\right) P\left(m_k^j \left| \mathbf{z}_{1:k} \right.\right) \tag{6}$$

which can be calculated by the traditional IMM algorithm with an augmented JMS model. Clearly, (6) indicates that there is no mode jump within the smoothing lag [37]. For more details about the recursion cycle of the AS-IMM algorithm, see [35].

Unlike the AS-IMM algorithm, RTSt-IMM proceeds in the RTS-type smoothing form, in which forward filtering and backward smoothing are realized by standard IMM filter and smoothing estimation combining possible hypotheses, respectively, in which the backward smoothing posterior distribution can be given by

$$\begin{aligned}&p\left(\mathbf{x}_{k-L} \left| \mathbf{z}_{1:k} \right.\right) \\ &= \sum_{j=1}^{M} p\left(\mathbf{x}_{k-L} \left| m_{k-L}^j, \mathbf{z}_{1:k} \right.\right) P\left(m_{k-L}^j \left| \mathbf{z}_{1:k} \right.\right).\end{aligned} \tag{7}$$

To efficiently calculate (7), several approximation techniques are introduced in [37]. For detailed implementation steps about the RTSt-IMM, see [37].

B. State Estimation Performance Analysis Under Mismatched Noise Covariance Matrix

In this part, we show the influence of mismatched noise covariance on the state estimation performance, in which the mismatch may be caused by the process noise covariance matrix (PNCM), the MNCM, or both. To facilitate our discussion, we give a specific analysis with a single-state transition model. The reason lies in the fact that the quantitative analysis for the JMS is nonanalytical, thus it is natural to consider the single-state model and the relevant conclusions can be extended to the JMS.

Consider a linear time-invariant (LTI) system with a stationary noise process. Based on the following conditions, we quantitatively calculate the performance of KF under the mismatch noise model.

CONDITION 1 The initial state, denoted as $\mathbf{x}_0$, is characterized by a distribution $\mathbf{x}_0 \sim \mathcal{N}(\hat{\mathbf{x}}_{0|0}, \mathbf{P}_0)$. In this context, it is assumed that $\hat{\mathbf{x}}_{0|0}$, along with its error covariance, $\mathbf{P}_{0|0}(\mathbf{P}_{0|0} \succ 0)$, is known in advance.

CONDITION 2 $(\mathbf{F}, \mathbf{I}_{n_x})$ is uniformly and completely controllable, and $(\mathbf{F}, \mathbf{H})$ is uniformly and completely observable.

It has been noted in [39] that if the Condition 2 holds, then the error covariance $\mathbf{P}_{k|k}$ will converge to a bounded stable positive-definite matrix, i.e., $\lim_{k\to\infty} \mathbf{P}_{k|k} = \mathbf{P}$ for any positive-definite initial covariance matrix $\mathbf{P}_{0|0}(\mathbf{P}_{0|0} \succ 0)$. The same result holds for residual covariance $\mathbf{S}_k$ and filter gain matrix $\mathbf{K}_k$, i.e., $\lim_{k\to\infty} \mathbf{S}_k = \mathbf{S}$ and $\lim_{k\to\infty} \mathbf{K}_k = \mathbf{K}$. In addition, the computed matrix parameters do not depend on the measurements themselves because of the LTI assumption, therefore the state covariance output matrix may not reflect the actual mean-squared error of the state. In this case, we have the following theorem.

THEOREM 1 For the LTI systems, if the Conditions 1 and 2 hold, then under bounded mismatched noise covariance, KF is an unbiased estimator but not an effective estimator, and its real error covariance matrix will increase and is bounded. At the time index k, the increment of the covariance matrix is $\sum_{i=0}^{k-1} \mathbf{M}^i\mathbf{B}(\mathbf{M}^i)^{\mathrm{T}}$, where $\mathbf{M} = [\mathbf{I}_{n_x} - (\mathbf{K} + \Delta\mathbf{K})\mathbf{H}]\mathbf{F}$, $\mathbf{B} = \Delta\mathbf{K}\mathbf{S}\Delta\mathbf{K}^{\mathrm{T}}$, and $\Delta\mathbf{K}$ denotes the filter gain difference between KF and matched KF (MKF) caused by the mismatched noise covariance.

PROOF See Appendix A. □

REMARK 1 Theorem 1 indicates that the KF is still an unbiased but not an effective estimator with mismatched noise covariance, and its real error covariance matrix will increase and the increment of the error covariance matrix is bounded. In addition, as the suboptimal smoothing estimators, AS-IMM and RTSt-IMM are both based on the KF, the result in Theorem 1 provides some useful guidelines for AS-IMM and RTSt-IMM.

III. ADAPTIVE IMM SMOOTHING ALGORITHMS

This section presents our main results about the adaptive IMM smoothing algorithms. In Section III-A, we introduce the VB models for adaptive state estimation, in which both the AS-IMM and RTSt-IMM algorithms will be discussed. In Section III-B, we introduce a novel approximate method by which the system state and unknown MNCM can be jointly estimated. In Section III-C, we demonstrate an adaptive IMM smoothing algorithm when the augmented state method is used. In Section III-D, we give another algorithm when the RTSt method is used. Finally, we give the consistency analysis of the proposed algorithm to estimate MNCM in Section III-E.

A. VB Optimization Models for Adaptive FLS With JMS

From the Bayesian perspective, the joint posterior distribution containing the state vector and MNCM should be calculated to form an adaptive state estimator. To this end, we give the VB optimization models both with augmented-state and RTSt methods in this part.

For the FLS estimation using an augmented JMS model in the presence of mismatched MNCM, the adaptive estimator must calculate the joint posterior distribution $p(\Theta_k|\mathbf{z}_{1:k})$, where $\Theta_k = \{\mathbf{X}_k, \mathbf{I}_k, \mathbf{R}_k\}$ represents a collective set encompassing the state vector, the latent model indicator, and the MNCM. $\mathbf{I}_k$ satisfies $I_k^j \in \{0 \quad 1\}$, $j \in [1 \quad M]$ and $\sum_{j=1}^{M} I_k^j = 1$. The fundamental insight in VB is to minimize the KLD between the real joint posterior distribution and an approximate distribution. Suppose that the approximate distribution of $p(\Theta_k|\mathbf{z}_{1:k})$ is $q(\Theta_k)$, VB method can be described by the following optimization problem, given by

$$q^* = \arg\min_{q\in\phi} \mathrm{KL}\left[q(\Theta_k) \| p(\Theta_k|\mathbf{z}_{1:k})\right] \tag{8}$$

where ϕ indicates the potential approximate set. It has been proved in [40], that the optimization problem in (8) can be solved in a closed form. However, for the optimization problem in (8), we have the following theorem.

THEOREM 2 For the augmented JMS model defined in (4) and (5), the VB optimization problem to approximate $p(\Theta_k|\mathbf{z}_{1:k})$ defined in (8) is not a conditionally conjugate model.

PROOF See Appendix B. □

Now let us discuss the VB optimization model for adaptive state estimation using the RTSt method mentioned in Section II-A. In this case, the adaptive estimator must calculate the joint posterior distribution $p(\Phi_{k-L:k}|\mathbf{z}_{1:k})$ with a forward filtering and backward smoothing structure, where

$$\Phi_{k-L:k} = \{\mathbf{x}_{k-L:k}, \mathbf{I}_{k-L:k}, \mathbf{R}_{k-L:k}\}.$$

In regard to the VB optimization model for approximating $p(\Phi_{k-L:k}|\mathbf{z}_{1:k})$, we have the following.

THEOREM 3 For the FLS estimation using an RTSt method with JMS in the presence of mismatched MNCM, the VB optimization problem to approximate $p(\Phi_{k-L:k}|\mathbf{z}_{1:k})$ is not a conditionally conjugate model.

PROOF See Appendix C. □

Theorems 2 and 3 indicate that $q(\Theta_k)$ and $q(\Phi_{k-L:k})$ cannot be computed analytically by the traditional VB method. For solving this problem, we propose an effective approximate method in Section III-B.

B. Main Idea of the Approximate Method

In Section III-A, we theoretically demonstrate that the joint posterior distributions of the system state and unknown MNCM cannot be analytically computed using the traditional VB method. To overcome these challenges, this article proposes an approximate method for solving the joint posterior distribution. The method stems from the following two aspects.

1) The VB method employs fixed-point iteration (FPI) [40], [41] to estimate the posterior distribution of each component. Therefore, when estimating the posterior distribution of MNCM given a fixed posterior distribution of the state, the problem can be simplified by utilizing the approximated posterior

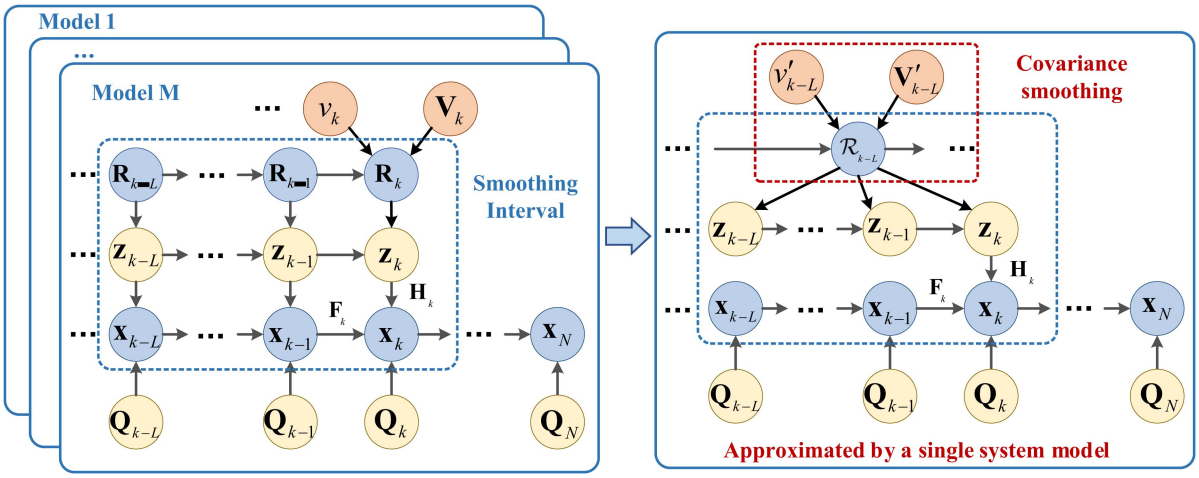

Fig. 1. Probabilistic graphical model for the adaptive IMM smoothing algorithms. In this article, the posterior distribution of MNCM is calculated by utilizing the approximated posterior distribution of the state through the moment-matching method, thereby transforming the multimodel VB inference problem into a single-model problem. For the RTSt-IMM method, $\mathcal{R}_{k-L}$ represents a smoothing MNCM in the smoothing interval. And for the AS-IMM method, $\mathcal{R}_{k-L}$ can be replaced by $\mathbf{R}_k$.

distribution of the state obtained through moment-matching, which corresponds to a single Gaussian distribution.

2) The smoothing estimator using RTSt-IMM can mitigate the challenges posed by the unknown Markov evolution model of MNCM by incorporating approximate smoothing statistics of MNCM within the smoothing interval.

To this end, we illustrate the probabilistic graphical model for the adaptive IMM smoothing algorithms in Fig. 1. Following the traditional VB inference, the proposed algorithms work in a FPI fashion. To be more specific, when fixing $q(\mathbf{R}_{k-L:k})$, $q(\mathbf{x}_{k-L:k})$ and $q(\mathbf{I}_{k-L:k})$ can be analytically calculated by traditional IMM smoothing algorithms. In what follows, by moment-matching approximation, $q(\mathbf{x}_{k-L:k})$ can be approximated as a single Gaussian distribution, i.e., $\tilde{q}(\mathbf{x}_{k-L:k})$. Thus, the problem of updating $q(\mathbf{R}_{k-L:k})$ given a fixed $\tilde{q}(\mathbf{x}_{k-L:k})$ degenerates into the smoothing estimation problem of $q(\mathbf{R}_{k-L:k})$ in a non-JMS model. We summarize the aforementioned FPI procedures in Fig. 2. It is clear that the proposed algorithms operate in a similar manner compared to the VB-AKF [15], with the difference being that we additionally introduce an approximate treatment of the state distribution to simplify the VB inference problem. The implementation process of FPI both in the augmented-state method and RTSt method will be discussed in detail in the following sections.

C. Adaptive IMM Smoothing Algorithm With Augmented-State Method

In this part, we give the steps to implement the algorithm of the adaptive IMM smoothing algorithm with the augmented-state method. For simplicity of notation, we note the adaptive IMM smoothing algorithm as VB-AS-IMM. From Section III-B, VB-AS-IMM works in an FPI manner.

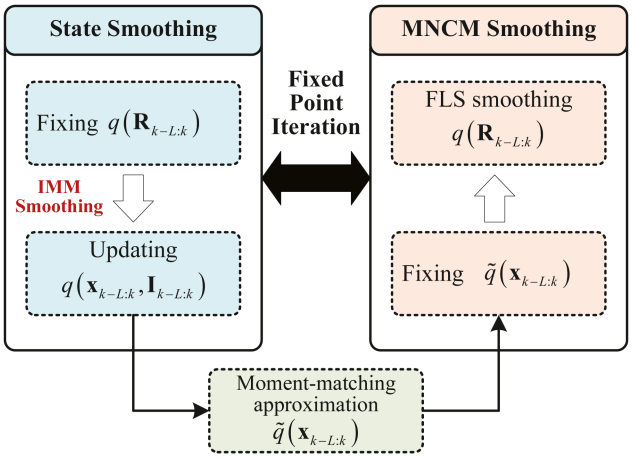

Fig. 2. Block diagram of FPI for the adaptive IMM smoothing algorithms.

For the remainder of this part, we only present one iteration procedure of joint estimation and omit the iteration subscript for simplicity. The implementation steps of the algorithm are as follows.

1) *Fixing $q(\mathbf{R}_k)$ to Calculate $q(\mathbf{I}_k)$ and $q(\mathbf{X}_k)$:* Given $q(\mathbf{R}_k)$, $\mathbf{R}_k$ can be set as $\hat{\mathbf{R}}_k = \mathbb{E}_{p(\mathbf{R}_k|\mathbf{z}_{1:k})}[\mathbf{R}_k^{-1}]^{-1}$ and then $q(\mathbf{I}_k)$ and $q(\mathbf{X}_k)$ can be calculated by the AS-IMM algorithm. The specific implementation process is as follows.

Reinitialization: For the augmented-state JMS, the initial state estimation and its error covariance matrix at time index k can be reinitialized by

$$\hat{\tilde{\mathbf{X}}}^j_{k-1|k-1} = \sum_{i=1}^{M} \mu^{i|j}_{k-1|k-1} \hat{\mathbf{X}}^i_{k-1|k-1} \tag{9}$$

$$\tilde{\mathbf{P}}^j_{k-1|k-1} = \sum_{i=1}^{M} \mu^{i|j}_{k-1|k-1} \left[\mathbf{P}^i_{k-1|k-1}\right.$$

$$+\text{outer}\left(\Delta\hat{\mathbf{X}}^{i}_{k-1|k-1}\right)\Big] \tag{10}$$

where

$$\Delta\hat{\mathbf{X}}^{i}_{k-1|k-1}=\hat{\mathbf{X}}^{i}_{k-1|k-1}-\hat{\hat{\mathbf{X}}}^{j}_{k-1|k-1}. \tag{11}$$

Here, $\hat{\mathbf{X}}^{i}_{k-1|k-1}$ and $\mathbf{P}^{i}_{k-1|k-1}$ represent the estimated state of the ith model at time index $k-1$ and the corresponding covariance matrix, respectively. $\mu^{i|j}_{k-1|k-1}$ represents the mixing probability.

Calculating $q(\mathbf{X}_k)$: Given the reinitialized augmented-state and covariance matrix, $q(\mathbf{X}_k)$ can be calculated by model-matched filtering step which is given in [35]. From (6), it is clear that $q(\mathbf{X}_k)$ can be approximated as a Gaussian mixture, formulated by

$$q\left(\mathbf{X}_k\right)=\sum_{j=1}^{M}\mu^{j}_{k}\mathcal{N}\left(\mathbf{X}_k\left|\hat{\mathbf{X}}^{j}_{k|k},\mathbf{P}^{j}_{k|k}\right.\right) \tag{12}$$

where μ^{j}_{k}, $\hat{\mathbf{X}}^{j}_{k|k}$, and $\mathbf{P}^{j}_{k|k}$ denote the model probability, state estimation, and error covariance matrix of the jth augmented model.

Calculating $q(\mathbf{I}_k)$: Equipped with (12), $q(\mathbf{I}_k)$ can be represented by the model probability, i.e.,

$$q\left(\mathbf{I}_k\right)=\prod_{j=1}^{M}\left[\mu^{j}_{k}\right]^{I_j}. \tag{13}$$

2) *Moment-Matching Approximation of $q(\mathbf{X}_k)$:* The moment-matching method is usually used to approximate a Gaussian mixture distribution to a single Gaussian. Thus, $q(\mathbf{X}_k)$ can be approximated by

$$\tilde{q}\left(\mathbf{X}_k\right)\approx\mathcal{N}\left(\mathbf{X}_k\left|\hat{\mathbf{X}}_{k|k},\mathbf{P}_{k|k}\right.\right) \tag{14}$$

where the augmented state $\hat{\mathbf{X}}_{k|k}$ and corresponding error covariance matrix $\mathbf{P}_{k|k}$ are calculated by moment-matching

$$\hat{\mathbf{X}}_{k|k}=\sum_{j=1}^{M}\mu^{j}_{k}\hat{\mathbf{X}}^{j}_{k|k} \tag{15}$$

$$\mathbf{P}_{k|k}=\sum_{j=1}^{M}\mu^{j}_{k}\left[\mathbf{P}^{j}_{k|k}+\text{outer}\left(\hat{\mathbf{X}}^{j}_{k|k}-\hat{\mathbf{X}}_{k|k}\right)\right]. \tag{16}$$

Therefore, the smoothing state and its error covariance matrix can be given by $\hat{\mathbf{x}}_{k-L|k}=\hat{\mathbf{X}}^{(L)}_{k|k}$, $\mathbf{P}_{k-L|k}=\mathbf{P}^{(L,L)}_{k|k}$, the superscript L denotes the Lth block of the vector or matrix.

3) *Fixing $\tilde{q}(\mathbf{x}_k)$ to Calculate $q(\mathbf{R}_k)$:* Equipped with $\tilde{q}(\mathbf{X}_k)$, $q(\mathbf{R}_k)$ can be calculated by

$$\begin{aligned}&\log q\left(\mathbf{R}_k\right)\\&=\mathbb{E}\left[\log p\left(\mathbf{X}_k,\mathbf{R}_k,\mathbf{z}_k\left|\mathbf{z}_{1:k-1}\right.\right)\right]+C_{\mathbf{R}_k}\\&=\mathbb{E}\left[\log p\left(\mathbf{z}_k\left|\mathbf{X}_k,\mathbf{R}_k\right.\right)\right]+\log p\left(\mathbf{R}_k\left|\mathbf{z}_{1:k-1}\right.\right)+C_{\mathbf{R}_k}\end{aligned}$$

where the expectation is taken with respect to $\tilde{q}(\mathbf{X}_k)$ given in (14). $C_{\mathbf{R}_k}$ stands for a constant variable with respect to $\mathbf{R}_k$. It is obvious that when $\tilde{q}(\mathbf{X}_k)$ is used to update $q(\mathbf{R}_k)$, the corresponding VB optimization problem degrades into the FLS estimation for MNCM with a non-JMS. Thus, $q(\mathbf{R}_k)$ obeys an inverse-Wishart distribution, i.e.,

$$q\left(\mathbf{R}_k\right)=\mathcal{IW}\left(\mathbf{R}_k\mid\hat{v}_{k|k},\hat{\mathbf{V}}_{k|k}\right) \tag{17}$$

where $\hat{v}_{k|k}$ and $\hat{\mathbf{V}}_{k|k}$ are calculated as follows:

$$\hat{v}_{k|k}=\hat{v}_{k|k-1}+1 \tag{18}$$

$$\hat{\mathbf{V}}_{k|k}=\hat{\mathbf{V}}_{k|k-1}+\mathbf{A}_k \tag{19}$$

$$\mathbf{A}_k=\text{outer}\left(\mathbf{z}_k-\left(\mathbf{H}^{a}_{k}\right)^{\mathrm{T}}\hat{\mathbf{X}}_{k|k}\right)+\left(\mathbf{H}^{a}_{k}\right)^{\mathrm{T}}\mathbf{P}_{k|k}\mathbf{H}^{a}_{k} \tag{20}$$

where $\hat{v}_{k|k-1}$ and $\hat{\mathbf{V}}_{k|k-1}$ can be calculated by (57) and (58).

Since the evidence lower bound (ELBO) of the nonconjugate variational model is nonanalytic, we use a more efficient criterion for convergence here. Intuitively, $q^{(n)}(\mathbf{X}_k)$ tends to be stable when $q^{(n)}(\mathbf{R}_k)$ converges. Consequently, the convergence of the VB-AS-IMM can be determined by measuring the distance between $\mathbf{T}^{(n)}=\mathbb{E}_{q^{(n)}(\mathbf{R}_k)}[\mathbf{R}_k]$ and $\mathbf{T}^{(n-1)}=\mathbb{E}_{q^{(n-1)}(\mathbf{R}_k)}[\mathbf{R}_k]$. In more detail, the Euclidean matrix distance is used to evaluate the distance between them, given by

$$D^{(n)}_{E}=\sqrt{\text{tr}\left[\text{outer}\left(\mathbf{T}^{(n)}-\mathbf{T}^{(n-1)}\right)\right]}. \tag{21}$$

In the setting, with a predefined threshold $D_{\mathcal{L}}$ (a small constant with $D_{\mathcal{L}}>0$), the convergence of the VB-AS-IMM can be realized when $D^{(n)}_{E}<D_{\mathcal{L}}$ is met.

D. Adaptive IMM Smoothing Algorithm With RTSt Method

In this part, we give the implementation steps of the adaptive IMM smoothing algorithm with the RTSt method. For simplicity of notation, we note the adaptive IMM smoothing algorithm with the RTSt method as VB-RTSt-IMM. It should be noted that the Markov transition distribution, i.e., $p(\mathbf{R}_i|\mathbf{R}_{i-1})$ is needed to be known when the RTSt method is considered. However, as demonstrated in [15] and [38], it is actually unknown in practice. Thus, the joint likelihood distribution is nonanalytic. To overcome this problem, an interval MNCM $\mathcal{R}_{k-L}$ is introduced for slowly time-varying MNCM [28], which represents a smoothed statistic in the smoothing interval. In this article, we follow this idea to simplify our approach. The implementation steps in one iteration of the algorithm are as follows.

1) *Fixing $q(\mathcal{R}_{k-L})$ to Calculate $q(\mathbf{x}_{k-L:k})$ and $q(\mathbf{I}_{k-L:k})$:* In order to satisfy the conjugation, the prior distribution of $\mathcal{R}_{k-L}$ is chosen as an inverse-Wishart distribution, given by

$$p_0\left(\mathcal{R}_{k-L}\right)=\mathcal{IW}\left(\mathcal{R}_{k-L}\mid\hat{\hat{v}}_{k-L|k-L-1},\hat{\hat{\mathbf{V}}}_{k-L|k-L-1}\right)$$

where $\hat{\hat{v}}_{k-L|k-L-1}$ and $\hat{\hat{\mathbf{V}}}_{k-L|k-L-1}$ represent predicted parameters within the smoothing interval $i\in[k-L,k]$, calculated by

$$\hat{\hat{v}}_{k-L|k-L-1}=\lambda\hat{\hat{v}}_{k-L-1|k-L-1}+(1-\lambda)n_z \tag{22}$$

$$\hat{\hat{\mathbf{V}}}_{k-L|k-L-1}=\lambda\hat{\hat{\mathbf{V}}}_{k-L-1|k-L-1}. \tag{23}$$

Here, λ is a discount factor usually in the range (0.95, 1].

Given $q(\mathcal{R}_{k-L})$, $\mathcal{R}_{k-L}$ can be set as

$$\bar{\mathcal{R}}_{k-L}=\mathbb{E}_{p_0(\mathcal{R}_{k-L})}\left[\mathcal{R}^{-1}_{k-L}\right]^{-1}.$$

Then, $q(\mathbf{x}_{k-L:k})$ and $q(\mathbf{I}_{k-L:k})$ can be calculated by the RTSt-IMM algorithm. The specific implementation process is as follows.

Forward IMM Filtering: Given $\bar{\mathcal{R}}_{k-L}$, it is clear that $\{\hat{\mathbf{x}}^j_{t|t}, \mathbf{P}^j_{t|t}, \boldsymbol{\mu}^j_i\}_{j\in\mathbf{M}}$ and $q_f(\mathbf{I}_t)$ can be calculated through forward IMM filtering [7], [37] for $t = k-L, \ldots, k-1, k$. In this follows, the Gaussian approximation of the state filtering posterior distribution $q_f(\mathbf{x}_t)$ can be realized by

$$q_f(\mathbf{x}_t) \approx \mathcal{N}\left(\mathbf{x}_t \mid \hat{\mathbf{x}}_{t|t}, \mathbf{P}_{t|t}\right). \tag{24}$$

Backward Smoothing: Equipped with the forward filtering parameters, RTSt-IMM works backward to achieve state smoothing. For $t = k-1, \ldots, k-L$, we first calculate the conditional probability distribution given by

$$p\left(\mathbf{x}_{t+1} \left| m^j_t, \mathbf{z}_{1:k}\right.\right) = \sum_{j=1}^{M} \tilde{\mu}^i_{t+1|k} \mathcal{N}\left(\mathbf{x}_{t+1} \left| \tilde{\mathbf{x}}^j_{t+1|k}, \tilde{\mathbf{P}}^j_{t+1|k}\right.\right)$$

where $\tilde{\mu}^i_{t+1|k}$, $\tilde{\mathbf{x}}^j_{t+1|k}$, and $\tilde{\mathbf{P}}^j_{t+1|k}$ can be calculated according to [37]. It follows that the smoothing probability distribution $p(\mathbf{x}_t \mid m^j_t, \mathbf{z}_{1:k})$ can be formulated by

$$p\left(\mathbf{x}_t \left| m^j_t, \mathbf{z}_{1:k}\right.\right) = \sum_{j=1}^{M} \mu^i_{t|k} \mathcal{N}\left(\mathbf{x}_t \left| \hat{\mathbf{x}}^j_{t|k}, \mathbf{P}^j_{t|k}\right.\right) \tag{25}$$

where $\mu^j_{t|k}$, $\hat{\mathbf{x}}^j_{t|k}$, and $\mathbf{P}^j_{t|k}$ can be calculated by the RTS smoothing.

2) *Moment-Matching Approximation of $q(\mathbf{x}_{k-L:k})$:* As described in Section III-C, we use the moment-matching method to approximate the state posterior distribution. Thus, $q(\mathbf{x}_{k-L:k})$ can be approximated by

$$\tilde{q}\left(\mathbf{x}_{k-L:k}\right) = \prod_{t=k-L}^{k} \mathcal{N}\left(\mathbf{x}_t \left| \hat{\mathbf{x}}_{t|k}, \mathbf{P}_{t|k}\right.\right) \tag{26}$$

where $\hat{\mathbf{x}}_{t|k}$ and its covariance matrix $\mathbf{P}_{t|k}$ at time index t, $t \in [k-L, k]$ can be calculated by

$$\hat{\mathbf{x}}_{t|k} = \sum_{j\in M} \mu^j_{t|k} \hat{\mathbf{x}}^j_{t|k} \tag{27}$$

$$\mathbf{P}_{t|k} = \sum_{j\in M} \mu^j_{t|k} \left[\mathbf{P}^j_{t|k} + \text{outer}\left(\hat{\mathbf{x}}^j_{t|k} - \hat{\mathbf{x}}_{t|k}\right)\right]. \tag{28}$$

3) *Fixing $\tilde{q}(\mathbf{x}_{k-L:k})$ to Calculate $q(\mathcal{R}_{k-L})$:* Equipped with $\tilde{q}(\mathbf{x}_{k-L:k})$, $q(\mathcal{R}_{k-L})$ can be calculated by (29) shown at the bottom of this page, where the expectation is taken with respect to $\tilde{q}(\mathbf{x}_{k-L:k})$. $\mathrm{C}_{\mathcal{R}_{k-L}}$ stands for the constant variable with respect to $\mathcal{R}_{k-L}$. $\mathbf{A}_t$ satisfies

$$\begin{aligned}\mathbf{A}_t &= \mathbb{E}_{q(\mathbf{x}_t)}\left[\text{outer}\left(\mathbf{z}_t - \mathbf{H}^1_t \mathbf{x}_t\right)\right] \\ &= \text{outer}\left(\mathbf{z}_t - \mathbf{H}^1_t \hat{\mathbf{x}}_{t|k}\right) + \mathbf{H}^1_t \mathbf{P}_{t|k} [\mathbf{H}^1_t]^{\mathrm{T}}\end{aligned} \tag{30}$$

where $\mathbf{H}^1_t$ denotes the measurement matrix of the 1th model. It can be seen from (29) that $q(\mathcal{R}_{k-L})$ satisfies

$$q\left(\mathcal{R}_{k-L}\right) = \mathcal{IW}\left(\mathcal{R}_{k-L} \mid \hat{\hat{v}}_{k-L|k-L}, \hat{\hat{\mathbf{V}}}_{k-L|k-L}\right) \tag{31}$$

where

$$\begin{cases}\hat{\hat{v}}_{k-L|k-L} = \hat{\hat{v}}_{k-L|k-L-1} + L + 1 \\ \hat{\hat{\mathbf{V}}}_{k-L|k-L} = \hat{\hat{\mathbf{V}}}_{k-L|k-L-1} + \sum_{t=k-L}^{k} \mathbf{A}_t.\end{cases} \tag{32}$$

It is important to note that the convergence criterion of the VB-RTSt-IMM can be designed with reference to VB-AS-IMM in Section III-C. Compared (32) with (17), we can find that the VB-AS-IMM only uses the measurement in time index k to update $q(\mathbf{R}_k)$ while the VB-RTSt-IMM utilizes the measurements in the whole smoothing interval to update $q(\mathcal{R}_{k-L})$. Thus, VB-RTSt-IMM reaches a faster convergence.

E. Estimation Consistency of the MNCM

In this part, we show the estimation consistency of MNCM. To facilitate our discussion, we take VB-AS-IMM into account since the update process of MNCM can be recursively integrated. From (17) to (20), the estimation of the MNCM in the VB-AS-IMM can be expressed in the following recursive form at the time index k, formulated by:

$$\begin{aligned}\hat{\mathbf{R}}_k &= \mathbb{E}\left[p\left(\mathbf{R}_k \mid \mathbf{z}_{1:k}\right)\right] \\ &= \frac{\hat{\mathbf{V}}_{k|k}}{\hat{v}_{k|k} - n_z - 1} \\ &= \frac{\hat{\mathbf{R}}_{k-1}\left(\hat{v}_{k-1|k-1} - n_z - 1\right) + \mathbf{A}_k}{\hat{v}_{k-1|k} - n_z} \\ &= \alpha_k \hat{\mathbf{R}}_{k-1} + (1-\alpha_k)\mathbf{A}_k\end{aligned} \tag{33}$$

where

$$\alpha_k = \frac{\hat{v}_{k-1|k-1} - n_z - 1}{\hat{v}_{k-1|k-1} - n_z}. \tag{34}$$

Here, $0 < \alpha_k < 1$. Similarly, the MNCM at time index $k-1$ can be calculated by

$$\hat{\mathbf{R}}_{k-1} = \alpha_{k-1}\hat{\mathbf{R}}_{k-2} + \left(1-\alpha_{k-1}\right)\mathbf{A}_{k-1}. \tag{35}$$

$$\begin{aligned}\log q\left(\mathcal{R}_{k-L}\right) &= \mathbb{E}\left[\log p\left(\mathbf{x}_{k-L:k}, \mathbf{z}_{k-L:k}, \mathcal{R}_{k-L} \left|\mathbf{z}_{1:k-L-1}\right.\right)\right] + \mathrm{C}_{\mathcal{R}_{k-L}} \\ &= \sum_{t=k-L}^{k} \mathbb{E}\left[\log p\left(\mathbf{z}_t \left|\mathbf{x}_t, \mathcal{R}_{k-L}\right.\right)\right] + \log p_0\left(\mathcal{R}_{k-L}\right) + \mathrm{C}_{\mathcal{R}_{k-L}} \\ &= -\frac{1}{2}\left(L + \hat{\hat{v}}_{k-L|k-L-1} + n_z + 2\right)\log\left|\mathcal{R}_{k-L}\right| - \frac{1}{2}\mathrm{tr}\left[\left(\hat{\hat{\mathbf{V}}}_{k-L|k-L-1} + \sum_{t=k-L}^{k}\mathbf{A}_t\right)\mathcal{R}^{-1}_{k-L}\right] + \mathrm{C}_{\mathcal{R}_{k-L}}\end{aligned} \tag{29}$$

Substituting (35) into (33) and proceeding the above procedure in the same fashion, we finally give the recursive MNCM estimation at time index k, given by

$$\begin{aligned}\hat{\mathbf{R}}_k &= \prod_{i=1}^{k}\alpha_i\hat{\mathbf{R}}_0 + \sum_{i=1}^{k}\prod_{j=i+1}^{k}\alpha_j\left(1-\alpha_i\right)\mathbf{A}_i \\ &= \frac{\hat{v}_{0|0}-n_z-1}{\hat{v}_{k-1|k-1}-n_z}\hat{\mathbf{R}}_0 + \sum_{i=1}^{k}\prod_{j=i+1}^{k}\alpha_j\left(1-\alpha_i\right)\mathbf{A}_i \\ &= \frac{1-\alpha_k}{1-\alpha_0}\hat{\mathbf{R}}_0 + \left(1-\alpha_k\right)\sum_{i=1}^{k}\mathbf{A}_i. \end{aligned} \tag{36}$$

Further, for LTI discrete systems, the expectations of the MNCM satisfy

$$\begin{aligned}&\lim_{k\to\infty}\mathbb{E}_{\mathbf{v}_k}\left[\hat{\mathbf{R}}_k\right] \\ &= \lim_{k\to\infty}\mathbb{E}_{\mathbf{v}_k}\left[\left(1-\alpha_k\right)\sum_{i=1}^{k}\mathbf{A}_i\right] \\ &= \lim_{k\to\infty}\frac{1}{\hat{v}_{0|0}+k-n_z-1}\sum_{i=1}^{k}\mathbf{M}\mathbb{E}_{\mathbf{v}_k}\left[\mathbf{v}_k\mathbf{v}_k^{\mathrm{T}}\right]\mathbf{M}^{\mathrm{T}} \\ &\quad + \lim_{k\to\infty}\frac{k}{\hat{v}_{0|0}+k-n_z-1}\mathbf{H}\mathbf{P}_{\infty}\mathbf{H}^{\mathrm{T}} \\ &= \mathbf{M}\mathbf{C}_v\mathbf{M}^{\mathrm{T}} + \mathbf{H}\mathbf{P}_{\infty}\mathbf{H}^{\mathrm{T}}. \end{aligned} \tag{37}$$

It is easy to know from (36) that $\hat{\mathbf{R}}_k$ is bounded. It can be shown from (37) that the mean value of MNCM is convergent. In addition, the influence of the forgetting factor is not considered in (37). If the forgetting factor is considered, (36) is modified as

$$\hat{\mathbf{R}}_k = \prod_{i=1}^{k}\gamma_i\hat{\mathbf{R}}_0 + \sum_{i=1}^{k}\prod_{j=i+1}^{k}\gamma_j\left(1-\gamma_i\right)\mathbf{A}_i \tag{38}$$

where $\gamma_i = \frac{\lambda(\hat{v}_{i-1|i-1}-n_z-1)}{\lambda(\hat{v}_{i-1|i-1}-n_z-1)+1}$. From (38), it is clear that the forgetting factor has less weight on the historical information, thus speeding up the convergence of the algorithm. From (37) and (38), the VB-AS-IMM realizes online estimation of MNCM and the estimation of MNCM is consistent. From (32), it is clear that the VB-RTSt-IMM utilizes the measurements in the whole smoothing interval to update $q(\mathcal{R}_{k-L})$ in which it smooths the $\mathbf{A}_t, t\in[k-L,k]$. Thus, the result in (37) and (38) can be extended to VB-RTSt-IMM.

IV. SIMULATIONS

In this section, a maneuvering target tracking example is carried out to verify the effectiveness of the proposed IMM smoothing algorithms. We compare the performance of VB-AS-IMM and VB-RTSt-IMM with model matched IMM (MIMM), model matched AS-IMM (MAS-IMM), IMM, and AS-IMM. In the following cases, we use the MAS-IMM as the performance benchmark.

The simulation scene is set as the radar maneuvering target tracking in 2-D space. The state vector of the system is $\mathbf{x}_k = [x_k \quad \dot{x}_k \quad y_k \quad \dot{y}_k]^{\mathrm{T}}$. There are three kinds of motion models in the simulation scene, namely, the constant velocity (CV) model, the right coordinated turn (CT1) model and the left coordinated turn (CT2) model. The nonaugmented state equation of the CV model is

$$\mathbf{x}_k = \mathbf{F}_{\mathrm{CV}}\mathbf{x}_{k-1} + \mathbf{Q}_{\mathrm{CV}},\ \mathbf{Q}_{CV} = \sigma_1^2\mathbf{Q}_1. \tag{39}$$

The nonaugmented state equation of the CT1 and CT2 models is

$$\mathbf{x}_k = \mathbf{F}_{\mathrm{CT}}\mathbf{x}_{k-1} + \mathbf{Q}_{\mathrm{CT}},\ \mathbf{Q}_{\mathrm{CT}} = \sigma_2^2\mathbf{Q}_2 \tag{40}$$

where the definitions of $\mathbf{F}_{\mathrm{CV}}$, $\mathbf{Q}_1$, $\mathbf{F}_{\mathrm{CT}}$, and $\mathbf{Q}_2$ can be referred to [42]. The initial state vector $\mathbf{x}_0$ is set as $\mathbf{x}_0 = [100\ \mathrm{km}, 20\ \mathrm{m/s}, 100\ \mathrm{km}, 0\ \mathrm{m/s}]^{\mathrm{T}}$ and initial state covariance $\mathbf{P}_0$ is set as $\mathbf{P}_0 = \mathrm{diag}(500, 50, 500, 50)$. The radar scanning interval is $T = 5s$. The total tracking time is 400 frames. The process noise is set to $\sigma_1^2 = 10^{-4}$ and $\sigma_2^2 = 10^{-4}$. The target makes a CV motion in frames $1\sim 80$, $161\sim 240$, $321\sim 400$, a right coordinated turn motion in frames $81\sim 160$ with an angular velocity of $-0.45^\circ/s$, and a left coordinated turn motion in frames $241\sim 320$ with an angular velocity of $0.45^\circ/s$.

In the simulation, the radial range and azimuth angle of the target are measured, so the measurement equation is formulated by

$$\mathbf{z}_k = \begin{bmatrix}\sqrt{x_k^2+y_k^2}\\ \arctan\left(\frac{y_k}{x_k}\right)\end{bmatrix} + \mathbf{n}_k \tag{41}$$

where x_k and y_k are the X-axis and Y-axis coordinates of the target. The MNCM is $\mathbf{R}_k = \mathrm{diag}([\,\delta_{k,r}^2 \quad \delta_{k,\theta}^2\,])$. The standard deviations of radar ranging and angle measuring errors are $\delta_{k,r}$ and $\delta_{k,\theta}$, respectively. In the simulation, $\delta_{k,r}$ and $\delta_{k,\theta}$ are set to 60 m and 0.2°, respectively. Since the measurement equation is nonlinear, the unbiased converted measurement (UCM) [43] method is used for coordinate transformation in all algorithms. The simulation scenario is shown in Fig. 3.

We take root mean square error (RMSE) of position and RMSE of velocity as the performance index of state estimation, which are defined in [28]. Moreover, the estimation accuracy of measurement noise is evaluated by the averaged KLD (AKLD), defined by

$$\mathrm{AKLD} \triangleq \frac{1}{N}\sum_{s=1}^{N}\mathrm{KL}\left[\hat{p}_s\left(\mathbf{n}_k\right)\|p_s\left(\mathbf{n}_k\right)\right] \tag{42}$$

where $\hat{p}_s(\cdot)$ and $p_s(\cdot)$ represent the estimated distribution and the true distribution of $\mathbf{n}_k$ in the sth run, respectively. $N = 100$ denotes the number of Monte Carlo runs. The convergence threshold is set to $D_{\mathcal{L}} = 10^{-3}$. Finally, the model probability value is used as the identification index of the motion model.

A. Performance of Two Adaptive Algorithms

First, we compare the performance of two proposed adaptive IMM smoothing algorithms. The initial MNCM of the two adaptive smoothing algorithms is 10 times larger

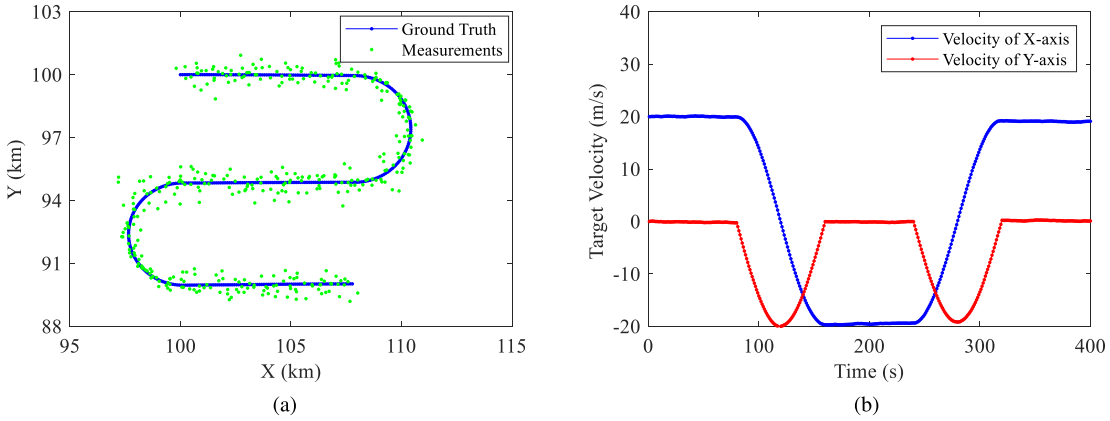

Fig. 3. Simulation scenario. (a) Real moving track of the target and the measurement plot of the radar. (b) Target velocity.

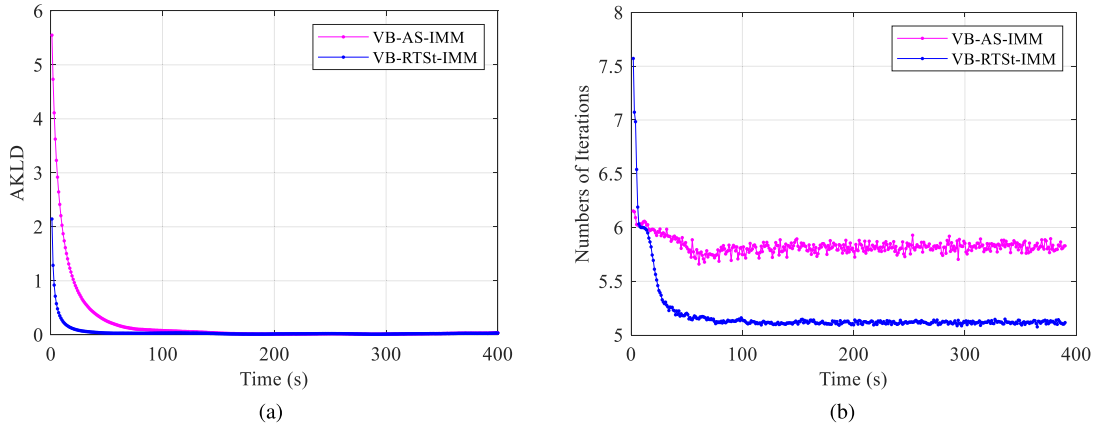

Fig. 4. AKLD and number of iterations of two adaptive smoothing algorithms. (a) AKLD of VB-AS-IMM and VB-RTSt-IMM. (b) Number of iterations of VB-AS-IMM and VB-RTSt-IMM.

TABLE I
Averaged RMSE of the Proposed Algorithms

Algorithms	RMSE of Position (m)	RMSE of Velocity (m/s)
MAS-IMM	90.71	1.06
VB-AS-IMM	91.13	1.08
VB-RTSt-IMM	90.72	1.06

than the true MNCM because the simulation considers that the $\delta_{k,r}$ and $\delta_{k,\theta}$ prior information of the MNCM is not accurate. The smoothing lag is set to $L = 10$.

Fig. 4(a) shows the AKLD of the algorithms. In Fig. 4(a), the AKLD of VB-AS-IMM and VB-RTSt-IMM algorithms gradually decrease over time and finally converge to 0, which indicates that the algorithms can learn the MNCM online. Fig. 4(b) shows the convergence speed comparison of the two algorithms. In Fig. 4(b), the number of iterations of VB-AS-IMM and VB-RTSt-IMM algorithms will gradually stabilize.

VB-RTSt-IMM can estimate the unknown interval MNCM using all the data in the smoothing interval, resulting in a faster convergence speed compared with the VB-AS-IMM algorithm. Table I shows the average RMSE of the two algorithms. The data presented in Table I demonstrate that the average RMSE associated with the two adaptive algorithms closely approximates that of the MAS-IMM. This observation suggests that, in scenarios where the prior knowledge of the MNCM is inaccurate, the proposed algorithms are capable of achieving near-optimal state estimation performance through the process of online learning.

B. Performance Comparison With Other Algorithms

In this section, we illustrate the performance of VB-AS-IMM and VB-RTSt-IMM compared with IMM, IMM2, AS-IMM, AS-IMM2, MIMM, and MAS-IMM. In IMM, AS-IMM, VB-AS-IMM, and VB-RTSt-IMM, the initial MNCM $\hat{\mathbf{R}}_0$ is set as $\hat{\mathbf{R}}_0 = 10\mathbf{R}_0$. Unlike the IMM and AS-IMM algorithms, the IMM2 and AS-IMM2 algorithms consider the scenario of multiple measurement models, where initial MNCM $\hat{\mathbf{R}}_0$ in each measurement model is set as $\hat{\mathbf{R}}_0 = b_i\mathbf{R}_0 (i = 1, \ldots, B)$, b_i denotes mismatch degree coefficient of the ith measurement model and B represents the number of measurement models. The other parameters of VB-AS-IMM and VB-RTSt-IMM are consistent with Section IV-A.

Figs. 5 and 6, respectively, illustrate the RMSE and the motion model probabilities associated with each algorithm. For the IMM2 and AS-IMM2 algorithms, we consider four measurement models, i.e., $B = 4$, and set $[b_1, b_2, b_3, b_4] =$

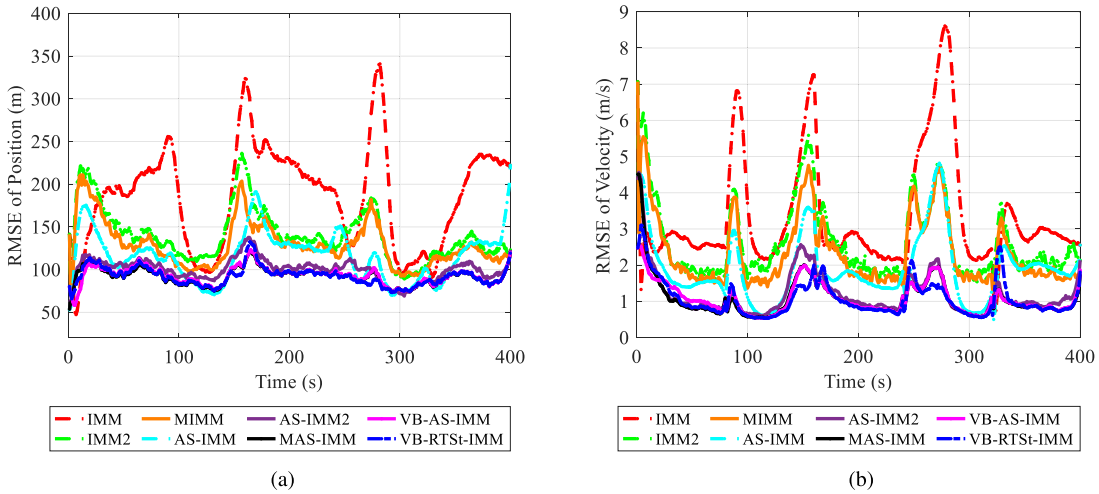

Fig. 5. RMSE of different algorithms (IMM: containing three motion models and a single measurement model; IMM2: containing three motion models and four measurement models). (a) RMSE of Position. (b) RMSE of Velocity.

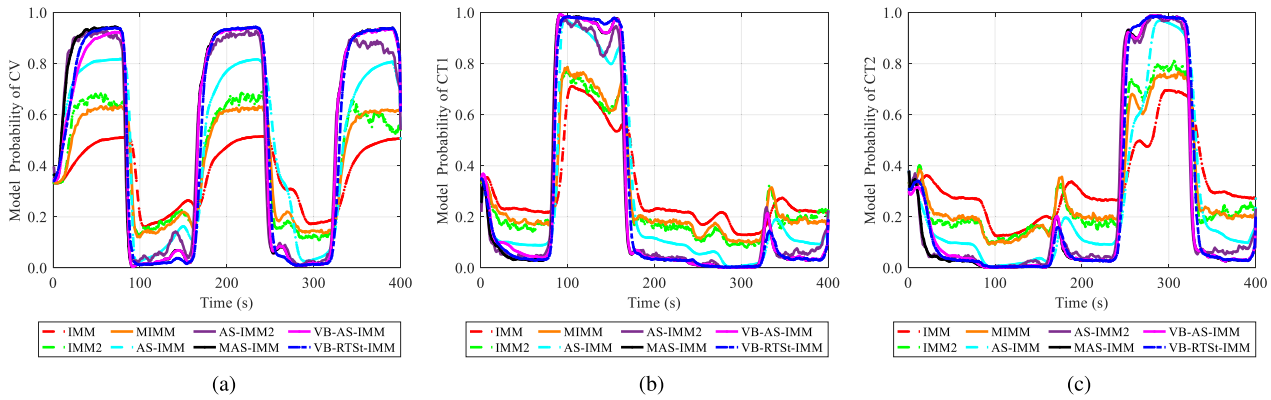

Fig. 6. Model probability of different algorithms (IMM: containing three motion models and a single measurement model; IMM2: containing three motion models and four measurement models). (a) CV model. (b) CT1 model. (c) CT2 model.

[0.1, 0.5, 5, 10]. From Fig. 5, we can see that the RMSE of traditional filtering algorithms, namely, IMM, IMM2, and MIMM, will increase rapidly during model switching. In Fig. 6, during the switching of the motion model, the model probability estimates provided by the IMM, IMM2, and MIMM algorithms exhibit high inaccuracy. This inaccuracy, primarily attributed to the presence of model competition, consequently leads to an increased RMSE. The performance of AS-IMM and AS-IMM2 algorithms is better than that of filtering algorithms, but it is worse than that of MAS-IMM algorithm. Owing to model mismatch in the AS-IMM and AS-IMM2 algorithms, its estimated model probability is inferior to that of MAS-IMM. This discrepancy consequently leads to higher RMSE in the AS-IMM and AS-IMM2 algorithms compared to MAS-IMM. The model probability and RMSE associated with both VB-AS-IMM and VB-RTSt-IMM algorithms closely approximate those observed in MAS-IMM.

Table II shows the averaged RMSE and execution time of the different algorithms, where the execution time is measured to assess the computational complexity of the algorithms in the MATLAB R2021a. For computation, 12th Gen Intel Core i5-12500H CPU (2.50 GHz) and 16 GB RAM are used. The VB-RTSt-IMM algorithm improves the RMSE of position by 50.70%, 33.83%, 20.15%, and 7.24% compared to the IMM, IMM2, AS-IMM, and AS-IMM2 algorithms, respectively. Similarly, it improves the RMSE of velocity by 68.66%, 60.23%, 43.55%, and 11.76% compared to the IMM, IMM2, AS-IMM, and AS-IMM2

TABLE II
Averaged RMSE and Execution Time of the Different Algorithms (IMM: Containing Three Motion Models and a Single Measurement Model; IMM2: Containing Three Motion Models and Four Measurement Models)

Algorithms	RMSE of Position (m)	RMSE of Velocity (m/s)	Execution Time (s)
IMM	184.91	3.35	0.029
IMM2	137.77	2.64	0.124
MIMM	127.42	2.33	0.029
AS-IMM	114.17	1.86	0.399
AS-IMM2	98.27	1.19	2.282
MAS-IMM	91.21	1.05	0.400
VB-AS-IMM	91.65	1.07	1.038
VB-RTSt-IMM	91.16	1.05	1.605

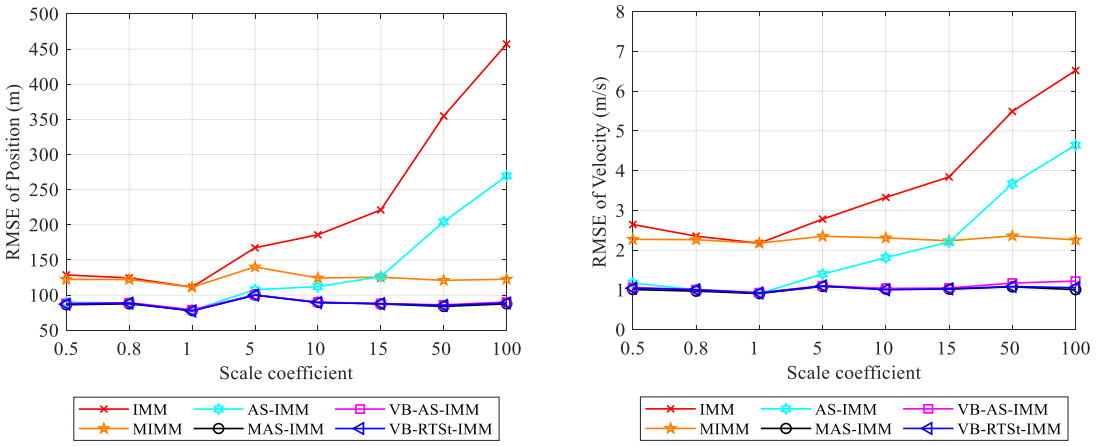

Fig. 7. RMSE of different algorithms under different scale coefficients. (a) RMSE of Position. (b) RMSE of Velocity.

algorithms, respectively. The computation times of various algorithms indicate that, compared to traditional methods, the proposed algorithm does not exhibit a significant increase in computational complexity, yet it enhances estimation accuracy. In addition, an analysis of the average iteration numbers at convergence for the proposed algorithm [as depicted in Fig. 4(b)] reveals its advantages of fewer iterations and faster convergence. For the IMM2 and AS-IMM2 algorithms, further consideration of additional measurement models leads to a reduction in RMSE; however, this comes at the cost of sacrificing time resources. The proposed algorithm, in contrast, eliminates the need for considering multiple measurement models, saving time resources. In addition, its computational complexity is manageable, rendering it suitable for online estimation.

C. Performance Comparison Under Different Mismatch Degrees

Finally, to assess the performance of VB-AS-IMM and VB-RTSt-IMM with varying degrees of mismatch. We set the initial $\hat{\mathbf{R}}_0$ to be $\hat{\mathbf{R}}_0 = s\mathbf{R}_0$, where s denotes a mismatch degree coefficient. The other parameters of VB-AS-IMM and VB-RTSt-IMM are consistent with Section IV-A.

Fig. 7 illustrates the RMSE of various algorithms under different scale coefficients. In Fig. 7, we set s as [0.5, 0.8, 1, 10, 15, 50, 100]. Fig. 7(a) and (b), respectively, represent the RMSE of position and velocity for different algorithms. Both of the aforementioned results indicate that as the mismatch level of MNCM increases in the scenario, the traditional algorithms (IMM, MIMM, AS-IMM) show a growing tendency of higher errors in estimating the target state. In contrast, the proposed algorithms (VB-AS-IMM and VB-RTSt-IMM) demonstrate the ability to maintain a stable estimation accuracy, which shows that the proposed algorithms can effectively learn the statistical characteristics of the noise covariance matrix online under different levels of mismatch.

V. CONCLUSION

In this article, we consider the online adaptive state smoothing problem of the JMS in the presence of unknown MNCM. First, we examine the quantitative impact of noise covariance matrix mismatch on state estimation and prove that the VB optimization model of JMS is nonconjugate compared with the single state-space model. Next, by introducing an efficient and effective approximate method, the VB-AS-IMM and VB-RTSt-IMM algorithms are proposed. The simulation results show that. 1) The VB-AS-IMM and VB-RTSt-IMM algorithms are superior to the existing IMM and AS-IMM algorithms in state estimation and motion model estimation due to their online adjustment and learning capabilities. 2) The VB-RTSt-IMM algorithm, which possesses smooth processing for MNCM, converges faster than the VB-AS-IMM algorithm. Finally, we theoretically show that the online estimation of MNCM both for the VB-AS-IMM and VB-RTSt-IMM is consistent.

APPENDIX A
PROOF OF THEOREM 1

At time index $k = 1$, the state estimate of KF $\hat{\mathbf{x}}_{\mathrm{KF},1}$ and MKF $\hat{\mathbf{x}}_{\mathrm{MKF},1}$ can be given by

$$\hat{\mathbf{x}}_{\mathrm{KF},1} = (\mathbf{K}+\Delta\mathbf{K})\,\mathbf{z}_1 + \left(\mathbf{I}_{n_x} - (\mathbf{K}+\Delta\mathbf{K})\,\mathbf{H}\right)\mathbf{F}\hat{\mathbf{x}}_{0|0} \quad (43)$$

$$\hat{\mathbf{x}}_{\mathrm{MKF},1} = \mathbf{K}\mathbf{z}_1 + \left(\mathbf{I}_{n_x} - \mathbf{K}\mathbf{H}\right)\mathbf{F}\hat{\mathbf{x}}_{0|0}\,. \quad (44)$$

Then, the state estimate difference between KF and MKF is $\Delta\hat{\mathbf{x}}_1 = \hat{\mathbf{x}}_{\mathrm{MKF},1} - \hat{\mathbf{x}}_{\mathrm{KF},1} = -\Delta\mathbf{K}(\mathbf{z}_1 - \mathbf{H}\mathbf{F}\hat{\mathbf{x}}_{0|0}) = -\Delta\mathbf{K}\mathbf{v}_1$. At time index $k = 2$, the state estimate difference $\Delta\hat{\mathbf{x}}_2$ is formulated by

$$\begin{aligned}\Delta\hat{\mathbf{x}}_2 &= -\Delta\mathbf{K}\mathbf{v}_{\mathrm{KF},2} + \left(\mathbf{I}_{n_x} - \mathbf{K}\mathbf{H}\right)\mathbf{F}\Delta\hat{\mathbf{x}}_1 \\ &= -\Delta\mathbf{K}\mathbf{v}_2 - \left[\mathbf{I}_{n_x} - (\mathbf{K}+\Delta\mathbf{K})\,\mathbf{H}\right]\mathbf{F}\Delta\mathbf{K}\mathbf{v}_1 \end{aligned} \quad (45)$$

where $\mathbf{v}_{\mathrm{KF},2}$ and $\mathbf{v}_2$ represent the residual vector of KF and MKF at time index $k = 2$, respectively. Proceeding the above procedure in the same fashion, the state estimate error

$\Delta\hat{\mathbf{x}}_k (k >= 2)$ is calculated by

$$\Delta\hat{\mathbf{x}}_k = -\sum_{i=1}^{k} [\mathbf{M}]^{(k-i)} \Delta\mathbf{K}\mathbf{v}_i \tag{46}$$

where $\mathbf{M} = [\mathbf{I}_{n_x} - (\mathbf{K} + \Delta\mathbf{K})\mathbf{H}]\mathbf{F}$. Taking the expectation with (46), we have

$$\mathbb{E}\left[\hat{\mathbf{x}}_{\mathrm{KF},k}\right] = \mathbf{x}_k - \sum_{i=1}^{k} [\mathbf{M}]^{(k-i)} \Delta\mathbf{K}\mathbb{E}\left[\mathbf{v}_i\right] = \mathbf{x}_k. \tag{47}$$

From (47), it is clear that KF is still an unbiased estimator under the mismatched noise covariance. In what follows, we analyze the influence of the mismatched noise covariance on the real state error covariance matrix of KF. The second-order central moment of $\hat{\mathbf{x}}_{\mathrm{KF},k}$ is calculated by

$$\begin{aligned}
&\mathbb{E}\left[\mathrm{outer}\left(\hat{\mathbf{x}}_{\mathrm{KF},k} - \mathbf{x}_k\right)\right] \\
&= \mathbb{E}\left[\mathrm{outer}\left(\hat{\mathbf{x}}_{\mathrm{MKF},k} - \Delta\hat{\mathbf{x}}_k - \mathbf{x}_k\right)\right] \\
&= \mathbb{E}\left[\tilde{\mathbf{x}}_{\mathrm{MKF},k}\tilde{\mathbf{x}}_{\mathrm{MKF},k}^{\mathrm{T}}\right] - \mathbb{E}\left[\tilde{\mathbf{x}}_{\mathrm{MKF},k}\Delta\hat{\mathbf{x}}_k^{\mathrm{T}}\right] \\
&\quad - \mathbb{E}\left[\Delta\hat{\mathbf{x}}_k\tilde{\mathbf{x}}_{\mathrm{MKF},k}^{\mathrm{T}}\right] + \mathbb{E}\left[\Delta\hat{\mathbf{x}}_k\Delta\hat{\mathbf{x}}_k^{\mathrm{T}}\right] \\
&= \mathbf{P}_{\mathrm{MKF},k} + \mathbb{E}\left[\Delta\hat{\mathbf{x}}_k\Delta\hat{\mathbf{x}}_k^{\mathrm{T}}\right]
\end{aligned} \tag{48}$$

where $\tilde{\mathbf{x}}_{\mathrm{MKF},k} = \hat{\mathbf{x}}_{\mathrm{MKF},k} - \mathbf{x}_k$ and $\mathbb{E}[\tilde{\mathbf{x}}_{\mathrm{MKF},k}\Delta\hat{\mathbf{x}}_k^{\mathrm{T}}] = 0$ can be given by following the orthogonal projection [44]. The second term on the right-hand side of (48) can be formulated as

$$\begin{aligned}
\mathbb{E}\left[\Delta\hat{\mathbf{x}}_k\Delta\hat{\mathbf{x}}_k^{\mathrm{T}}\right] &= \mathbb{E}\left[\mathrm{outer}\left(\sum_{i=1}^{k}\mathbf{M}^{(k-i)}\Delta\mathbf{K}\mathbf{v}_i\right)\right] \\
&= \sum_{i=1}^{k}\mathbf{M}^{(k-i)}\Delta\mathbf{K}\mathbf{S}\Delta\mathbf{K}^{\mathrm{T}}\left[\mathbf{M}^{(k-i)}\right]^{\mathrm{T}}.
\end{aligned} \tag{49}$$

Then, we have

$$\begin{aligned}
&\mathbb{E}\left[\mathrm{outer}\left(\hat{\mathbf{x}}_{\mathrm{KF},k} - \mathbf{x}_k\right)\right] \\
&= \mathbf{P}_{\mathrm{MKF},k} + \sum_{i=1}^{k}\mathbf{M}^{(k-i)}\Delta\mathbf{K}\mathbf{S}\Delta\mathbf{K}^{\mathrm{T}}\left[\mathbf{M}^{(k-i)}\right]^{\mathrm{T}}.
\end{aligned} \tag{50}$$

It is obvious that the second term in (50) is positive definite, and it can be simplified as

$$\sum_{i=1}^{k}\mathbf{M}^{(k-i)}\Delta\mathbf{K}\mathbf{S}\Delta\mathbf{K}^{\mathrm{T}}\left[\mathbf{M}^{(k-i)}\right]^{\mathrm{T}} = \sum_{i=0}^{k-1}\mathbf{M}^{i}\mathbf{B}\left(\mathbf{M}^{i}\right)^{\mathrm{T}}$$

where $\mathbf{B} = \Delta\mathbf{K}\mathbf{S}\Delta\mathbf{K}^{\mathrm{T}}$. It can be seen from (50) that under the mismatched noise covariance, the real error covariance matrix of KF increases a positive-definite term $\sum_{i=0}^{k-1}\mathbf{M}^{i}\mathbf{B}(\mathbf{M}^{i})^{\mathrm{T}}$ compared with the MKF.

Since $\mathbf{M}$ denotes the system matrix of KF, then with the Conditions 1 and 2, we have $\lambda_0 = \rho(\mathbf{M}) < 1$ and $\lim_{i\to\infty}\mathbf{M}^{i} = \mathbf{0}$, where $\rho(\mathbf{M})$ denotes the spectral radius of $\mathbf{M}$. In the following, we have:

$$\begin{aligned}
&\lim_{k\to\infty}\left\|\sum_{i=0}^{k-1}\mathbf{M}^{i}\mathbf{B}\left(\mathbf{M}^{i}\right)^{\mathrm{T}}\right\| \\
&\le \lim_{k\to\infty}\sum_{i=0}^{k-1}\left\|\mathbf{M}^{i}\right\|^{2}\|\mathbf{B}\| \\
&\le \lim_{k\to\infty}\sum_{i=0}^{k-1}\left[(\lambda_0+\varepsilon)^2\right]^{i}\|\mathbf{B}\| = \frac{\|\mathbf{B}\|}{1-(\lambda_0+\varepsilon)^2}
\end{aligned} \tag{51}$$

where $\varepsilon > 0$ and $\lambda_0+\varepsilon < 1$. It can be seen from (51) that matrix series $\sum_{i=0}^{\infty}\mathbf{M}^{i}\mathbf{B}(\mathbf{M}^{i})^{\mathrm{T}}$ is absolutely convergent, so it is convergent. To sum up, we obtain Theorem 1.

APPENDIX B
PROOF OF THEOREM 2

According to the mean-field assumption, the posterior distribution $q(\Theta_k)$ can be factorized in the following form:

$$q\left(\Theta_k\right) \approx q\left(\mathbf{X}_k\right)q\left(\mathbf{R}_k\right)q\left(\mathbf{I}_k\right). \tag{52}$$

If the VB optimization problem defined in (8) is a conditionally conjugate model, then $q(\mathbf{X}_k)$, $q(\mathbf{R}_k)$, and $q(\mathbf{I}_k)$ can be calculated by the FPI, i.e.,

$$\ln q\left(\mathbf{I}_k\right) = \mathbb{E}_{\mathbf{X}_k,\mathbf{R}_k}\left[\ln p\left(\Theta_k, \mathbf{z}_k\,|\mathbf{z}_{1:k-1}\right)\right] + \mathrm{C}_{\mathbf{I}_k} \tag{53}$$

$$\ln q\left(\mathbf{X}_k\right) = \mathbb{E}_{\mathbf{I}_k,\mathbf{R}_k}\left[\ln p\left(\Theta_k, \mathbf{z}_k\,|\mathbf{z}_{1:k-1}\right)\right] + \mathrm{C}_{\mathbf{X}_k} \tag{54}$$

$$\ln q\left(\mathbf{R}_k\right) = \mathbb{E}_{\mathbf{X}_k,\mathbf{I}_k}\left[\ln p\left(\Theta_k, \mathbf{z}_k\,|\mathbf{z}_{1:k-1}\right)\right] + \mathrm{C}_{\mathbf{R}_k} \tag{55}$$

where $\mathrm{C}_{\mathbf{I}_k}$, $\mathrm{C}_{\mathbf{X}_k}$, and $\mathrm{C}_{\mathbf{R}_k}$, respectively, denote the constant items with respect to $\mathbf{I}_k$, $\mathbf{X}_k$, and $\mathbf{R}_k$. The log of the joint likelihood in (53) can be given by

$$\begin{aligned}
&\ln p\left(\Theta_k, \mathbf{z}_k\,|\mathbf{z}_{1:k-1}\right) \\
&= \ln p\left(\mathbf{z}_k\,|\Theta_k\right) + \ln p\left(\mathbf{X}_k\,|\mathbf{I}_k, \mathbf{z}_{1:k-1}\right) \\
&\quad + \ln p\left(\mathbf{I}_k\,|\mathbf{z}_{1:k-1}\right) + \ln p\left(\mathbf{R}_k\,|\mathbf{z}_{1:k-1}\right).
\end{aligned} \tag{56}$$

Given the prior distribution of the augmented state $\mathbf{X}_{k-1}$, i.e.,

$$\begin{aligned}
&p\left(\mathbf{X}_{k-1}\,|\mathbf{z}_{1:k-1}\right) \\
&= \sum_{j=1}^{M} p\left(\mathbf{X}_{k-1}\left|m_{k-1}^{j}, \mathbf{z}_{1:k-1}\right.\right) P\left(m_{k-1}^{j}\,|\mathbf{z}_{1:k-1}\right).
\end{aligned}$$

Then, $p(\mathbf{X}_k|\mathbf{I}_k, \mathbf{z}_{1:k-1})$ and $p(\mathbf{I}_k|\mathbf{z}_{1:k-1})$ in (56) can be obtained through model reinitialization and model condition prediction by the AS-IMM, given by

$$p\left(\mathbf{X}_k\,|\mathbf{I}_k, \mathbf{z}_{1:k-1}\right) = \prod_{j=1}^{M}\mathcal{N}\left(\mathbf{X}_k\left|\hat{\mathbf{X}}_{k|k-1}^{j}, \mathbf{P}_{k|k-1}^{j}\right.\right)^{I_k^j}$$

$$p\left(\mathbf{I}_k\,|\mathbf{z}_{1:k-1}\right) = \prod_{j=1}^{M}\left(\mu_{k|k-1}^{j}\right)^{I_k^j}.$$

In order to satisfy the conjugation, the prior distribution of the MNCM is taken as an inverse-Wishart distribution, i.e.,

$$p\left(\mathbf{R}_{k-1}\,|\mathbf{z}_{1:k-1}\right) = \mathcal{IW}\left(\hat{\upsilon}_{k-1}, \hat{\mathbf{V}}_k\right).$$

In what follows, it is straightforward to give the prediction distribution $p(\mathbf{R}_k|\mathbf{z}_{1:k-1})$ by the Beta-Bartlett model [45], i.e.,

$$p\left(\mathbf{R}_k\,|\mathbf{z}_{1:k-1}\right) = \mathcal{IW}\left(\mathbf{R}_k\left|\hat{\upsilon}_{k|k-1}, \hat{\mathbf{V}}_{k|k-1}\right.\right)$$

where

$$\hat{v}_{k|k-1} = \lambda \hat{v}_{k-1} + (1-\lambda)\left(n_z - 1\right) \tag{57}$$

$$\hat{\mathbf{V}}_{k|k-1} = \lambda \hat{\mathbf{V}}_{k-1} \tag{58}$$

where λ is a forgetting factor.

With the result in (56), (54) can be written in the form

$$\begin{aligned}\ln q\left(\mathbf{X}_k\right) = & \sum_{j=1}^{M} \mu_k^j \ln\left(\mathcal{N}\left(\mathbf{z}_k \left| \mathbf{H}_k^{a(j)}\mathbf{X}_k, \mathbb{E}\left[\mathbf{R}_k^{-1}\right]^{-1}\right.\right)\right) \\ & + \sum_{j=1}^{M} \mu_k^j \ln\left(\mathcal{N}\left(\mathbf{X}_k \left| \hat{\mathbf{X}}_{k|k-1}^j, \mathbf{P}_{k|k-1}^j\right.\right)\right) \\ = & \sum_{j=1}^{M} \mu_k^j \ln \mathcal{N}\left(\mathbf{X}_k \left| \hat{\mathbf{X}}_{k|k}^j, \mathbf{P}_{k|k}^j\right.\right)\end{aligned} \tag{59}$$

where

$$\begin{aligned}\mathbf{P}_{k|k}^j &= \left(\left[\mathbf{P}_{k|k-1}^j\right]^{-1} + \left[\mathbf{H}_k^{a(j)}\right]^{\mathrm{T}} \mathbb{E}\left[\mathbf{R}_k^{-1}\right] \mathbf{H}_k^{a(j)}\right)^{-1} \\ \hat{\mathbf{X}}_{k|k}^j &= \mathbf{P}_{k|k}^j \left[\left(\mathbf{P}_{k|k-1}^j\right)^{-1} \hat{\mathbf{X}}_{k|k-1}^j \right. \\ &\quad \left. + \left(\mathbf{H}_k^{a(j)}\right)^{\mathrm{T}} \mathbb{E}\left[\mathbf{R}_k^{-1}\right] \mathbf{z}_k\right].\end{aligned}$$

From (59), it is obvious that $\hat{\mathbf{X}}_{k|k}^j$ and $\mathbf{P}_{k|k}^j$ can be calculated by the traditional AS-IMM with

$$\bar{\mathbf{R}}_k = \mathbb{E}_{q(\mathbf{R}_k|\mathbf{z}_{1:k-1})}\left[\mathbf{R}_k^{-1}\right]^{-1}$$

as the MNCM. We can further see form (59) that $q(\mathbf{X}_k)$ is not a Gaussian mixture. Similarly, it can be proved that $q(\mathbf{I}_k)$ and $q(\mathbf{R}_k)$ also do not satisfy the conjugation condition. To sum up, we obtain Theorem 2.

APPENDIX C
PROOF OF THEOREM 3

Similar to the proof process in Theorem 2, we first calculated the joint likelihood distribution, given by

$$\begin{aligned}&\ln p\left(\mathbf{x}_{k-L:k}, \mathbf{z}_{k-L:k}, \mathbf{I}_{k-L:k}, \mathcal{R}_{k-L} \left| \mathbf{z}_{1:k-L-1}\right.\right) \\ &= \sum_{i=k-L}^{k} \sum_{j=1}^{M} I_i^j \ln p\left(\mathbf{z}_i \left| \mathbf{x}_i, \mathcal{R}_{k-L}\right.\right) \\ &\quad + \sum_{i=k-L+1}^{k} \sum_{j=1}^{M} I_i^j \ln p\left(\mathbf{x}_i \left| \mathbf{x}_{i-1}\right.\right) \\ &\quad + \sum_{j=1}^{M} I_{k-L}^j \ln p\left(\mathbf{x}_{k-L} \left| \mathbf{z}_{1:k-L-1}\right.\right) \\ &\quad + p_0\left(\mathcal{R}_{k-L}\right) + p\left(\mathbf{I}_{k-L} \left| \mathbf{z}_{1:k-L-1}\right.\right)\end{aligned} \tag{60}$$

in which we use a smoothed MNCM $\mathcal{R}_{k-L}$ (denotes an approximated smoothing MNCM within interval $[k-L \quad k]$), defined in [28], to replace $\mathbf{R}_{k-L:k}$ for simplicity. To simplify our discussion, we consider a special case when the smoothing lag is zero, i.e., $L = 0$. In this case, the forward filtering process to approximate $p(\Phi_{k-L:k}|\mathbf{z}_{1:k})$ is reduced to be a filtering problem with an unaugmented JMS model. By a similar analysis to Appendix B, we can see that the forward filtering problem is not a conditionally conjugate model. Thus, we have Theorem 3. □

REFERENCES

[1] R. Visina, Y. Bar-Shalom, and P. Willett, "Multiple-model estimators for tracking sharply maneuvering ground targets," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 54, no. 3, pp. 1404–1414, Jun. 2018.

[2] H. Lin, J. Lam, M. Z. Q. Chen, Z. Shu, and Z.-G. Wu, "Interacting multiple model estimator for networked control systems: Stability, convergence, and performance," *IEEE Trans. Autom. Control*, vol. 64, no. 3, pp. 928–943, Mar. 2019.

[3] S. Dong, Z.-G. Wu, P. Shi, H. R. Karimi, and H. Su, "Networked fault detection for Markov jump nonlinear systems," *IEEE Trans. Fuzzy Syst.*, vol. 26, no. 6, pp. 3368–3378, Dec. 2018.

[4] X. R. Li and V. P. Jilkov, "Survey of maneuvering target tracking. Part V. Multiple-model methods," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 41, no. 4, pp. 1255–1321, Oct. 2005.

[5] X. Rong Li, "Multiple-model estimation with variable structure. II. Model-set adaptation," *IEEE Trans. Autom. Control*, vol. 45, no. 11, pp. 2047–2060, Nov. 2000.

[6] H. Blom and Y Bar-Shalom, "The interacting multiple model algorithm for systems with Markovian switching coefficients," *IEEE Trans. Autom. Control*, vol. AC-33, no. 8, pp. 780–783, Aug. 1988.

[7] Y. Bar-Shalom, S. Challa, and H. Blom, "IMM estimator versus optimal estimator for hybrid systems," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 41, no. 3, pp. 986–991, Jul. 2005.

[8] J. L. Maryak, J. C. Spall, and B. D. Heydon, "Use of the Kalman filter for inference in state-space models with unknown noise distributions," *IEEE Trans. Autom. Control*, vol. 49, no. 1, pp. 87–90, Jan. 2004.

[9] X. Yu and J. Li, "Adaptive Kalman filtering for recursive both additive noise and multiplicative noise," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 58, no. 3, pp. 1634–1649, Jun. 2022.

[10] M. R. Rajamani and J. B. Rawlings, "Estimation of the disturbance structure from data using semidefinite programming and optimal weighting," *Automatica*, vol. 45, no. 1, pp. 142–148, 2009.

[11] R. M. Watson, J. N. Gross, C. N. Taylor, and R. C. Leishman, "Enabling robust state estimation through measurement error covariance adaptation," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 56, no. 3, pp. 2026–2040, Jun. 2020.

[12] K. A. Myers and B. D. Tapley, "Adaptive sequential estimation with unknown noise statistics," *IEEE Trans. Autom. Control*, vol. AC-21, no. 4, pp. 520–523, Aug. 1976.

[13] A. H. Mohamed and K. P. Schwarz, "Adaptive Kalman filtering for INS/GPS," *J. Geodesy*, vol. 73, pp. 193–203, 1999.

[14] R. Dehghannasiri, M. S. Esfahani, and E. R. Dougherty, "Intrinsically Bayesian robust Kalman filter: An innovation process approach," *IEEE Trans. Signal Process.*, vol. 65, no. 10, pp. 2531–2546, May 2017.

[15] S. Sarkka and A. Nummenmaa, "Recursive noise adaptive Kalman filtering by variational Bayesian approximations," *IEEE Trans. Autom. Control*, vol. 54, no. 3, pp. 596–600, Mar. 2009.

[16] H. Cui, L. Mihaylova, X. Wang, and S. Gao, "Uncertainty-aware variational inference for target tracking," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 59, no. 1, pp. 258–273, Feb. 2023.

[17] D. G. Tzikas, A. C. Likas, and N. P. Galatsanos, "The variational approximation for Bayesian inference," *IEEE Signal Process. Mag.*, vol. 25, no. 6, pp. 131–146, Nov. 2008.

[18] C. Zhang, J. Bütepage, H. Kjellström, and S. Mandt, "Advances in variational inference," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 41, no. 8, pp. 2008–2026, Aug. 2019.

[19] P. Dong, Z. Jing, H. Leung, and K. Shen, "Variational Bayesian adaptive cubature information filter based on Wishart distribution," *IEEE Trans. Autom. Control*, vol. 62, no. 11, pp. 6051–6057, Nov. 2017.

[20] A. H. Chughtai, A. Majal, M. Tahir, and M. Uppal, "Variational-based nonlinear Bayesian filtering with biased observations," *IEEE Trans. Signal Process.*, vol. 70, pp. 5295–5307, 2022.
[21] H. Zhu, G. Zhang, Y. Li, and H. Leung, "An adaptive Kalman filter with inaccurate noise covariances in the presence of outliers," *IEEE Trans. Autom. Control*, vol. 67, no. 1, pp. 374–381, Jan. 2022.
[22] G. Zhang, J. Lan, L. Zhang, F. He, and S. Li, "Filtering in in pairwise Markov model with Student's t non-stationary noise with application to target tracking," *IEEE Trans. Signal Process.*, vol. 69, pp. 1627–1641, 2021.
[23] P. Dong, Z. Jing, H. Leung, K. Shen, and M. Li, "The labeled multi-Bernoulli filter for multitarget tracking with glint noise," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 55, no. 5, pp. 2253–2268, Oct. 2019.
[24] G. Wang, X. Wang, and Y. Zhang, "Variational Bayesian IMM-filter for JMSs with unknown noise covariances," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 56, no. 2, pp. 1652–1661, Apr. 2020.
[25] C. Lu, W. Feng, W. Li, Y. Zhang, and Y. Guo, "An adaptive IMM filter for jump Markov systems with inaccurate noise covariances in the presence of missing measurements," *Digit. Signal Process.*, vol. 127, 2022, Art. no. 103529.
[26] X. Wei, B. Hua, Y. Wu, and Z. Chen, "Robust interacting multiple model cubature Kalman filter for nonlinear filtering with unknown non-Gaussian noise," *Digit. Signal Process.*, vol. 136, 2023, Art. no. 103982.
[27] H. Li, L. Yan, and Y. Xia, "Distributed multiple model filtering for Markov jump systems with measurement outliers," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 59, no. 3, pp. 2823–2837, Jun. 2023.
[28] H. Xu, K. Duan, H. Yuan, W. Xie, and Y. Wang, "Adaptive fixed-lag smoothing algorithms based on the variational Bayesian method," *IEEE Trans. Autom. Control*, vol. 66, no. 10, pp. 4881–4887, Oct. 2021.
[29] F. Papi, M. Bocquel, M. Podt, and Y. Boers, "Fixed-lag smoothing for Bayes optimal knowledge exploitation in target tracking," *IEEE Trans. Signal Process.*, vol. 62, no. 12, pp. 3143–3152, Jun. 2014.
[30] B. D. O. Anderson and J. B. Moore, *Optimal Filtering*. Englewood Cliffs, NJ, USA: Prentice-Hall, 1979.
[31] T. Ogle and W. Blair, "Fixed-lag alpha-beta filter for target trajectory smoothing," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 40, no. 4, pp. 1417–1421, Oct. 2004.
[32] B. Chen and J. Tugnait, "Multisensor tracking of a maneuvering target in clutter using IMMPDA fixed-lag smoothing," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 36, no. 3, pp. 983–991, Jul. 2000.
[33] V. J. Mathews and J. K. Tugnait, "Detection and estimation with fixed lag for abruptly changing systems," *IEEE Trans. Aerosp. Electron. Syst.*, vol. AES-19, no. 5, pp. 730–739, Sep. 1983.
[34] R. Helmick, W. Blair, and S. Hoffman, "One-step fixed-lag smoothers for Markovian switching systems," *IEEE Trans. Autom. Control*, vol. 41, no. 7, pp. 1051–1056, Jul. 1996.
[35] B. Chen and J. Tugnait, "Interacting multiple model fixed-lag smoothing algorithm for Markovian switching systems," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 36, no. 1, pp. 243–250, Jan. 2000.
[36] M. R. Morelande and B. Ristic, "Smoothed state estimation for nonlinear Markovian switching systems," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 44, no. 4, pp. 1309–1325, Oct. 2008.
[37] N. Nadarajah, R. Tharmarasa, M. McDonald, and T. Kirubarajan, "IMM forward filtering and backward smoothing for maneuvering target tracking," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 48, no. 3, pp. 2673–2678, Jul. 2012.
[38] G. Agamennoni, J. I. Nieto, and E. M. Nebot, "Approximate inference in state-space models with heavy-tailed noise," *IEEE Trans. Signal Process.*, vol. 60, no. 10, pp. 5024–5037, Oct. 2012.
[39] A. H. Jazwinski, *Jazwinski Stochastic Processes and Filtering Theory*. New York, NY, USA: Academic Press, 1970.
[40] M. Sato, "Online model selection based on the variational," *Bayes Neural Computation*, vol. 13, no. 7, pp. 1649–1681, 2001.
[41] M. J. Wainwright and M. I. Jordan, "Graphical models, exponential families, and variational inference," *Foundations Trends Mach. Learn.*, vol. 1, no. 1/2, pp. 1–305, 2008.
[42] X. Rong Li and V. Jilkov, "Survey of maneuvering target tracking. Part I. Dynamic models," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 39, no. 4, pp. 1333–1364, Oct. 2003.
[43] M. Longbin, S. Xiaoquan, Z. Yiyu, S. Z. Kang, and Y. Bar-Shalom, "Unbiased converted measurements for tracking," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 34, no. 3, pp. 1023–1027, Jul. 1998.
[44] I. Rhodes, "A tutorial introduction to estimation and filtering," *IEEE Trans. Autom. Control*, vol. 16, no. 6, pp. 688–706, Dec. 1971.
[45] C. M. Carvalho and M. West, "Dynamic matrix-variate graphical models," *Bayesian Anal.*, vol. 2, no. 1, pp. 69–97, 2007.

Hong Xu (Member, IEEE) received the B.S. degree in information engineering and the M.S. degree in information and communication engineering from Wuhan Early Warning Academy, Wuhan, China, in 2013 and 2016, respectively, and the Ph.D. degree in information and communication engineering from the Naval University of Engineering, Wuhan, China, in 2020.

He is currently an Associate Professor with the Hangzhou Institute of Technology, Xidian University, Hangzhou, China. His research interests include radar signal processing and target tracking.

Qin Pan (Student Member, IEEE) is currently working toward the Ph.D. degree in electronic science and technology with the Department of Remote Sensing Science and Technology, School of Electronic Engineering, Xidian University, Xi'an, China.

His research interests include radar signal processing and target tracking.

Heng Xu (Student Member, IEEE) is currently working toward the Ph.D. degree in electronic science and technology with the Department of Remote Sensing Science and Technology, School of Electronic Engineering, Xidian University, Xi'an, China.

His research interests include radar signal processing and electronic countermeasures.

Yinghui Quan (Senior Member, IEEE) received the B.S. and Ph.D. degrees in electrical engineering from Xidian University, Xi'an, China, in 2004 and 2012, respectively.

He is currently a Full Professor with the Department of Remote Sensing Science and Technology, School of Electronic Engineering, Xidian University. His research interests include agile radar and radar remote sensing.