An Adaptive Kalman Filter With Inaccurate Noise Covariances in the Presence of Outliers

Hao Zhu, Guorui Zhang, Yongfu Li, and Henry Leung

***Abstract*—In this article, a novel variational Bayesian (VB) adaptive Kalman filter with inaccurate nominal process and measurement noise covariances (PMNC) in the presence of outliers is proposed. The probability density functions of state transition and measurement likelihood are modeled as Gaussian–Gamma mixture distributions. The VB inference is used to perform the state and PMNC simultaneously. Simulations show that the effectiveness of the proposed method with inaccurate noise covariances in the presence of outliers environments.**

***Index Terms*—Adaptive Kalman filter (KF), Gaussian–Gamma mixture (GGM) distribution, outliers, variational Bayesian (VB).**

I. Introduction

Estimating hidden dynamical system states from noisy measurements has widely been used in military and civilian systems [1], [2]. The Kalman filter (KF) achieves an optimal estimate of state with the linear Gaussian system and many extended filters are used for nonlinear systems [3].

The main problem of KF and its extensions is that the statistical knowledge of process and measurement noises (PMN) should be known in advance and PMN are approximated as Gaussian distributions. In practice, PMN may be non-Gaussian noises and the covariances are unknown. For example, heavy-tailed noises (HTN) have been recognized in the radar systems [4]. In order to estimate the states with HTN, many computation heavy filters have been developed, such as particle filter [3] and Gaussian sum filter [5], [6]. The robust outlier adaptive KF is a commonly used online approach to solve the problem of state estimation with HTN [7]–[13]. The challenge of the adaptive KF is to design an appropriate objective function or modeling method for the state estimation with HTN to guarantee the required accuracy while preserving low computation cost. A suboptimal recursive solution is developed to perform the filtering problem for liner discrete-time systems driven by non-Gaussian noises [11], [12]. A maximum correntropy Kalman filter [14], [15], which uses maximum correntropy criterion in the Kalman framework, is proposed for HTN. However, the aforementioned adaptive KFs assume that the nominal covariances of PMN are known. Recently, although a variational Bayesian (VB) approach has been developed to perform the states and unknown covariances with Gaussian noise in [16]–[18], it may fail when the dynamical system has a strongly HTN.

Many Student's t (ST) distribution-based adaptive filters have been presented to estimate the state of dynamical systems with HTN. They can be roughly categorized into the following two classes: the ST filter (STF) [19]–[25] and the robust ST-based Kalman filter (RSTKF) [26]–[28]. In STF, an ST distribution is used to approximate the probability density function (PDF) of state posterior. However, the STF also assumes that the nominal covariances of PMN are known. The RSTKF is first proposed in [26] and exhibits an excellent estimation performance for moderately heavy-tailed PMN, in which the ST distribution is approximated as a Gaussian–Gamma (GG) distribution. However, in practice, the characteristic of outliers is unknown. It is a challenge to select the parameters of the heavy-tailed PMN in advance or estimate the parameters of the unknown heavy-tailed PMN.

In this article, a novel VB adaptive KF with Gaussian–Gamma mixture (GGM) distribution (VBAKF-GGM) is proposed to estimate the state with inaccurate PMN covariances in the presence of outliers. The proposed GGM distributions, which are developed in a hierarchical Gaussian form [29], are developed to approximate the PDFs of state transition and measurement likelihood. A new hierarchical linear state-space model (SSM) is then formulated. The VB technique is used to simultaneous estimate the state and parameter in the hierarchical SSM. The proposed VBAKF-GGM has advantage over the RSTKF that the accurate prior information can be extracted via adaptive learning. The effectiveness of the proposed VBAKF-GGM filter is verified by computer simulations. The contributions of this article are three-fold.

1) Compared with the GG distribution in RSTKF [26], [27], a novel GGM distribution is developed here to model the HTN. This mixture of Gamma distribution models the scale parameter for HTN based on the accurate prior, which is extracted from the set of rough prior by adaptive learning the mixing probabilities.
2) While most estimation frameworks involve state and unknown parameters, the proposed framework considers estimation of the state, together with unknown parameters and accurate prior information using the VB method.
3) A novel adaptive filter incorporating inaccurate PMN covariances is proposed in the presence of outliers. Extensive simulations demonstrate the effectiveness of the proposed VBAKF-GGM filter, which is improved in comparison with existing filters in the presence of outliers.

II. Problem Formulation

A linear SSM with heavy-tailed PMN is described as

$$\mathbf{x}_k = \mathbf{F}\mathbf{x}_{k-1}+\mathbf{w}_{k-1} \tag{1}$$

Manuscript received July 2, 2020; revised October 22, 2020; accepted January 20, 2021. Date of publication February 2, 2021; date of current version December 29, 2021. This work was supported in part by the National Natural Science Foundation of China under Grant 62073052, Grant U1964202, and Grant 61773082 and in part by the Key Project of Crossing and Emerging Area of CQUPT under Grant A2018-02. Recommended by Associate Editor S. K. Spurgeon. *(Corresponding author: Hao Zhu.)*

Hao Zhu, Guorui Zhang, and Yongfu Li are with the Key Laboratory of Intelligent Air-Ground Cooperative Control for Universities in Chongqing, Department of Automation, Chongqing University of Posts and Telecommunications, Chongqing 400065, China (e-mail: zhuhao@cqupt.edu.cn; zhanggr2019@163.com; laf1212@163.com).

Henry Leung is with the Department of Electrical and Computer Engineering, University of Calgary, Calgary, AB T2N 1N4, Canada (e-mail: leungh@ucalgary.ca).

Color versions of one or more figures in this article are available at https://doi.org/10.1109/TAC.2021.3056343.

Digital Object Identifier 10.1109/TAC.2021.3056343

$$\mathbf{z}_k = \mathbf{C}\mathbf{x}_k + \mathbf{v}_k \tag{2}$$

where $\mathbf{x}_k \in \mathbb{R}^n$, $\mathbf{z}_k \in \mathbb{R}^m$, $\mathbf{F} \in \mathbb{R}^{n\times n}$, and $\mathbf{C} \in \mathbb{R}^{m\times n}$ are the state, measurement state transition, and measurement function, respectively. $k = 1, 2, \ldots, N$, N is the total sampling number. $\mathbf{w}_k$ and $\mathbf{v}_k$ are considered to be independent of initial state $\mathbf{x}_0$. In this article, $\mathbf{w}_k$ and $\mathbf{v}_k$ are the unknown HTN, which are usually given as

$$p(\mathbf{w}_k) \sim \delta_k^w \mathcal{N}(\mathbf{0}, \mathbf{Q}_k) + (1-\delta_k^w)\tilde{p}(\mathbf{w}_k) \tag{3}$$

$$p(\mathbf{v}_k) \sim \delta_k^v \mathcal{N}(\mathbf{0}, \mathbf{R}_k) + (1-\delta_k^v)\tilde{p}(\mathbf{v}_k) \tag{4}$$

where $\mathcal{N}(\mathbf{0}, \mathbf{Q}_k)$ is referred to the Gaussian PDF, which has mean zero and covariance $\mathbf{Q}_k$, δ_k^w, and δ_k^v are the weights, $\tilde{p}(\mathbf{w}_k)$ and $\tilde{p}(\mathbf{v}_k)$ denote the PDFs of outliers, respectively. $\mathbf{Q}_k$ and $\mathbf{R}_k$ are nominal covariances of PMN, respectively. They are usually assumed to be known in the current robust adaptive filters. But in practice, $\mathbf{Q}_k$ and $\mathbf{R}_k$ are unknown and maybe vary in time.

In this article, our aim is to estimate the state $\mathbf{x}_k$ with inaccurate nominal covariances of PMN. The PDFs of state transition and measurement likelihood are given as

$$p(\mathbf{z}_k | \mathbf{x}_k, \mathbf{R}_k, \lambda_k) = \mathcal{N}(\mathbf{z}_k; \mathbf{C}\mathbf{x}_k, \mathbf{R}_k/\lambda_k) \tag{5}$$

$$p(\mathbf{x}_k | \mathbf{x}_{k-1}, \mathbf{Q}_k, \varepsilon_k) = \mathcal{N}(\mathbf{x}_k; \mathbf{F}\mathbf{x}_{k-1}, \mathbf{Q}_k/\varepsilon_k) \tag{6}$$

where λ_k and ε_k are scale parameters. If $\lambda_k = 1$, it means that the measurement likelihood is formulated as a Gaussian with a nominal covariance. If $0 < \lambda_k < 1$, the measurement likelihood is formulated as a Gaussian with a larger covariance, which is used to model outliers of measurement noise. If $\varepsilon_k = 1$, it means that the density function of the state transition is generated by a Gaussian with a nominal covariance. If $0 < \varepsilon_k < 1$, the PDF of the state transition is generated by a Gaussian with a larger covariance, which is used to model outliers of process noise. In this article, mixture of Gamma distributions are proposed to model the characteristics of λ_k and ε_k, respectively, i.e.,

$$p(\lambda_k) = \sum_{i=1}^{I} \pi_{k,i} \mathrm{G}(\lambda_k; a_{0,i}, b_{0,i}) \tag{7}$$

$$p(\varepsilon_k) = \sum_{j=1}^{J} \gamma_{k,j} \mathrm{G}(\varepsilon_k; e_{0,j}, f_{0,j}) \tag{8}$$

where $\mathrm{G}(.; a_{0,i}, b_{0,i})$ stands for the Gamma distribution, which has shape parameter $a_{0,i}$ and rate parameter $b_{0,i}$, I and J denote the numbers of Gamma mixture components, and $\pi_{k,i}$ and $\gamma_{k,j}$ stand for the mixing coefficient of component i for parameter λ_k and mixing coefficient of component j for parameter ε_k, respectively. Let $\mathbf{a}_0 = [a_{0,1}, a_{0,2}, \ldots, a_{0,I}]$, $\mathbf{b}_0 = [b_{0,1}, b_{0,2}, \ldots, b_{0,I}]$, $\mathbf{e}_0 = [e_{0,1}, e_{0,2}, \ldots, e_{0,J}]$, and $\mathbf{f}_0 = [f_{0,1}, f_{0,2}, \ldots, f_{0,J}]$. For convenience, we introduce indicator variables $\mathbf{y}_k = [y_{k,1}, y_{k,2}, \ldots, y_{k,I}]$ and $\boldsymbol{\rho}_k = [\rho_{k,1}, \rho_{k,2}, \ldots, \rho_{k,J}]$, where $y_{k,i}$ and $\rho_{k,j}$ are binary variables. If $y_{k,i} = 1$, it means that λ_k is generated by the ith component. If $\rho_{k,j} = 1$, it means that ε_k is generated by the jth component. We obtain

$$p(\lambda_k | \mathbf{y}_k) = \prod_{i=1}^{I} (\mathrm{G}(\lambda_k; a_{0,i}, b_{0,i}))^{y_{k,i}} \tag{9}$$

$$p(\mathbf{y}_k | \boldsymbol{\pi}_k) = \prod_{i=1}^{I} (\pi_{k,i})^{y_{k,i}} \tag{10}$$

$$p(\varepsilon_k | \boldsymbol{\rho}_k) = \prod_{j=1}^{J} (\mathrm{G}(\varepsilon_k; e_{0,j}, f_{0,j}))^{\rho_{k,j}} \tag{11}$$

$$p(\boldsymbol{\rho}_k | \boldsymbol{\gamma}_k) = \prod_{j=1}^{J} (\gamma_{k,j})^{\rho_{k,j}} \tag{12}$$

where $\boldsymbol{\pi}_k = \{\pi_{k,i}\}_{i=1}^{J}$ and $\boldsymbol{\gamma}_k = \{\gamma_{k,j}\}_{j=1}^{J}$. Therefore, the PDFs of state transition and measurement likelihood are modeled as GGM distributions, respectively. Unlike the GG formulation in ST distribution, the proposed GGM has a set of rough prior information from the shape parameters $\mathbf{a}_0$ and $\mathbf{e}_0$, and rate parameters $\mathbf{b}_0$ and $\mathbf{f}_0$. Then, the accurate prior information can be selected from the rough prior information set $\mathbf{a}_0$, $\mathbf{b}_0$, $\mathbf{e}_0$, and $\mathbf{f}_0$ by adaptive learning the mixing probabilities $\boldsymbol{\pi}_k$ and $\boldsymbol{\gamma}_k$.

Then, suppose that, the conjugate priors for $\boldsymbol{\pi}_k$ and $\boldsymbol{\gamma}_k$ are Dirichlet distributions, i.e.,

$$p(\boldsymbol{\pi}_k) = \mathrm{Dir}(\boldsymbol{\pi}_k; \boldsymbol{\alpha}_0), p(\boldsymbol{\gamma}_k) = \mathrm{Dir}(\boldsymbol{\gamma}_k; \boldsymbol{\beta}_0) \tag{13}$$

where $\mathrm{Dir}(.; \boldsymbol{\alpha}_0)$ denotes the Dirichlet distribution with parameter vector $\boldsymbol{\alpha}_0$, $\boldsymbol{\alpha}_0 = [\alpha_{0,1}, \alpha_{0,2}, \ldots, \alpha_{0,I}]$, and $\boldsymbol{\beta}_0 = [\beta_{0,1}, \beta_{0,2}, \ldots, \beta_{0,J}]$. The conjugate priors for $\mathbf{Q}_k$ and $\mathbf{R}_k$ are supposed to follow inverse Wishart distributions [29], i.e.,

$$p(\mathbf{Q}_k) = \mathrm{IW}(\mathbf{Q}_k; \mathrm{u}_{k|k-1}, \mathbf{U}_{k|k-1}) \tag{14}$$

$$p(\mathbf{R}_k) = \mathrm{IW}(\mathbf{R}_k; \mathrm{t}_{k|k-1}, \mathbf{T}_{k|k-1}) \tag{15}$$

where $\mathrm{IW}(\cdot; \mathrm{u}, \mathbf{U})$ are the inverse-Wishart distribution with u and $\mathbf{U}$. A similar heuristics in [16] is proposed to model the slowly time-varying covariances of PMN, i.e.,

$$\mathrm{t}_{k|k-1} = \tau(\mathrm{t}_{k-1|k-1} - m - 1) + m + 1 \tag{16}$$

$$\mathbf{T}_{k|k-1} = \tau\mathbf{T}_{k-1|k-1} \tag{17}$$

$$\mathrm{u}_{k|k-1} = \tau(\mathrm{u}_{k-1|k-1} - n - 1) + n + 1 \tag{18}$$

$$\mathbf{U}_{k|k-1} = \tau\mathbf{U}_{k-1|k-1} \tag{19}$$

where τ is a forgetting factor and $\tau \in (0, 1]$ [16]. Therefore, initially, we have

$$\frac{\mathbf{T}_{0|0}}{\mathrm{t}_{0|0} - m - 1} = \bar{\mathbf{R}}_0 \tag{20}$$

$$\frac{\mathbf{U}_{0|0}}{\mathrm{u}_{0|0} - n - 1} = \bar{\mathbf{Q}}_0 \tag{21}$$

where $\bar{\mathbf{R}}_0$ and $\bar{\mathbf{Q}}_0$ denote the initial inaccurate nominal measurement and process noise covariances, respectively.

Given the initial inaccurate nominal covariances $\bar{\mathbf{R}}_0$ and $\bar{\mathbf{Q}}_0$, the hierarchical SSM is developed, and the proposed VBAKF-GGM filter is illustrated in Fig. 1.

III. VB Adaptive KF With GGM Distribution

A. VB Technique

From the Bayes' rule, we have

$$p(\Phi_k | \mathbf{z}_{1:k}, \Psi_k) = \frac{p(\mathbf{z}_{1:k} | \Phi_k) p(\Phi_k | \Psi_k)}{p(\mathbf{z}_{1:k} | \Psi_k)} \tag{22}$$

where $\Phi_k = \{\mathbf{x}_k, \mathbf{x}_{k-1}, \mathbf{Q}_k, \varepsilon_k, \mathbf{R}_k, \lambda_k, \mathbf{y}_k, \boldsymbol{\pi}_k, \boldsymbol{\alpha}_k, \boldsymbol{\rho}_k, \boldsymbol{\gamma}_k, \boldsymbol{\beta}_k\}$ and $\Psi_k = \{\mathrm{u}_{k|k-1}, \mathbf{U}_{k|k-1}, \mathrm{t}_{k|k-1}, \mathbf{T}_{k|k-1}, \mathbf{a}_0, \mathbf{b}_0, \mathbf{e}_0, \mathbf{f}_0\}$. A distribution $q(\Phi_k)$ is introduced and is proposed to approximate $p(\Phi_k|\mathbf{z}_{1:k}, \Psi_k)$ in the VB framework. Furthermore, a mean field approach for $q(\Phi_k)$ is used, i.e.,

$$\begin{aligned} q(\Phi_k) = \ & q(\mathbf{x}_k, \mathbf{x}_{k-1}) q(\mathbf{Q}_k) q(\varepsilon_k) q(\mathbf{R}_k) q(\lambda_k) q(\mathbf{y}_k) q(\boldsymbol{\pi}_k) \\ & \times q(\boldsymbol{\alpha}_k) q(\boldsymbol{\rho}_k) q(\boldsymbol{\gamma}_k) q(\boldsymbol{\beta}_k). \end{aligned} \tag{23}$$

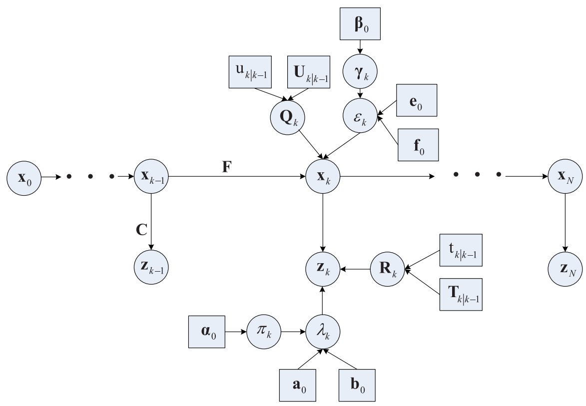

Fig. 1. Proposed VBAKF-GGM.

From [30], we have

$$\log q(\Delta) = \mathrm{E}_{\Phi_k^{(-\Delta)}}[\log p(\mathbf{z}_{1:k}|\Phi_k)p(\Phi_k|\Psi_k)] \quad (24)$$

where $\mathrm{E}[.]$ is referred to the expectation operation, Δ denotes a member of Φ_k, and $\Phi_k{}^{(-\Delta)}$ denotes the set of all members in Φ_k except for Δ. As the $q(\mathbf{x}_k, \mathbf{x}_{k-1})$, $q(\mathbf{Q}_k)$, $q(\varepsilon_k)$, $q(\mathbf{R}_k)$, $q(\lambda_k)$, $q(\mathbf{y}_k)$, $q(\boldsymbol{\pi}_k)$, $q(\boldsymbol{\alpha}_k)$, $q(\boldsymbol{\rho}_k)$, $q(\boldsymbol{\gamma}_k)$, and $q(\boldsymbol{\beta}_k)$ are coupled. In this article, $q(\Delta)$ is calculated as $q^{(l+1)}(\Delta)$ at the $l+1$ iteration by $q^{(l)}(\Phi_k{}^{(-\Delta)})$. Then, we obtain

$$\begin{aligned}
& p(\mathbf{z}_{1:k}|\Phi_k)p(\Phi_k|\Psi_k) \\
&= \mathcal{N}(\mathbf{z}_k;\mathbf{C}\mathbf{x}_k, \mathbf{R}_k/\lambda_k)\prod_{i=1}^{I}(\mathrm{G}(\lambda_k;a_{k,i}, b_{k,i}))^{y_{k,i}} \\
&\quad \times \prod_{i=1}^{I}(\pi_{k,i})^{y_{k,i}}\mathrm{Dir}(\boldsymbol{\pi}_k;\boldsymbol{\alpha}_0)\mathcal{N}(\mathbf{x}_k;\mathbf{F}\mathbf{x}_{k-1}, \mathbf{Q}_k/\varepsilon_k) \\
&\quad \times \prod_{j=1}^{J}(\mathrm{G}(\varepsilon_k;e_{k,j}, f_{k,j}))^{\rho_{k,j}}\prod_{j=1}^{J}(\gamma_{k,j})^{\rho_{k,j}}\mathrm{Dir}(\boldsymbol{\gamma}_k;\boldsymbol{\beta}_0) \\
&\quad \times \mathrm{IW}(\mathbf{Q}_k;\mathrm{u}_{k|k-1}, \mathbf{U}_{k|k-1})\mathrm{IW}(\mathbf{R}_k;\mathrm{t}_{k|k-1}, \mathbf{T}_{k|k-1}) \\
&\quad \times \mathcal{N}(\mathbf{x}_{k-1};\hat{\mathbf{x}}_{k-1|k-1}, \mathbf{P}_{k-1|k-1})p(\mathbf{z}_{1:k-1}).
\end{aligned} \quad (25)$$

Proposition 1: Set $\Delta = \lambda_k$ and employing (25) in (24), $q^{(l+1)}(\lambda_k)$ is considered as a Gamma distribution, and let $\Delta = \varepsilon_k$ and employing (25) in (24), $q^{(l+1)}(\varepsilon_k)$ is also considered as a Gamma distribution, i.e.,

$$q^{(l+1)}(\lambda_k) = \mathrm{G}(\lambda_k; c_k^{(l+1)}, d_k^{(l+1)}) \quad (26)$$

$$q^{(l+1)}(\varepsilon_k) = \mathrm{G}(\varepsilon_k; g_k^{(l+1)}, h_k^{(l+1)}) \quad (27)$$

where $c_k^{(l+1)}$, $d_k^{(l+1)}$, $g_k^{(l+1)}$, and $h_k^{(l+1)}$ are

$$c_k^{(l+1)} = 0.5n + \sum_{i=1}^{I}\mathrm{E}^{(l)}[y_{k,i}]a_{0,i} \quad (28)$$

$$d_k^{(l+1)} = 0.5\mathrm{Tr}\left(\mathbf{A}_k^{(l)}\mathrm{E}^{(l)}[\mathbf{R}_k{}^{-1}]\right) + \sum_{i=1}^{I}\mathrm{E}^{(l)}[y_{k,i}]b_{0,i} \quad (29)$$

$$g_k^{(l+1)} = 0.5\,m + \sum_{j=1}^{J}\mathrm{E}^{(l)}[\rho_{k,j}]e_{0,j} \quad (30)$$

$$h_k^{(l+1)} = 0.5\mathrm{Tr}\left(\mathbf{B}_k^{(l)}\mathrm{E}^{(l)}[\mathbf{Q}_k{}^{-1}]\right) + \sum_{j=1}^{J}\mathrm{E}^{(l)}[\rho_{k,j}]f_{0,j} \quad (31)$$

where $\mathrm{Tr}(.)$ stands for the trace operation, $\mathbf{A}_k^{(l)}$ and $\mathbf{B}_k^{(l)}$ are

$$\mathbf{A}_k^{(l)} = \mathrm{E}^{(l)}[(\mathbf{z}_k - \mathbf{C}\mathbf{x}_k)(\mathbf{z}_k - \mathbf{C}\mathbf{x}_k)^{\mathrm{T}}] \quad (32)$$

$$\mathbf{B}_k^{(l)} = \mathrm{E}^{(l)}[(\mathbf{x}_k - \mathbf{F}\mathbf{x}_{k-1})(\mathbf{x}_k - \mathbf{F}\mathbf{x}_{k-1})^{\mathrm{T}}]. \quad (33)$$

Proof: See Appendix A ■

Proposition 2: Set $\Delta = y_{k,i}$ and using (25) in (24), $q^{(l+1)}(y_{k,i})$ is a multinomial PDF, and let $\Delta = \rho_{k,j}$ and using (25) in (24), $q^{(l+1)}(\rho_{k,j})$ is a multinomial PDF, i.e.,

$$q^{(l+1)}(y_{k,i}) = \mathrm{Mult}(y_{k,i}; \varpi_{k,i}^{(l+1)}) \quad (34)$$

$$q^{(l+1)}(\rho_{k,j}) = \mathrm{Mult}(\rho_{k,j}; \omega_{k,j}^{(l+1)}) \quad (35)$$

where $\mathrm{Mult}(.)$ denotes a multinomial distribution, the parameters $\varpi_{k,i}^{(l+1)}$ and $\omega_{k,j}^{(l+1)}$ are

$$\varpi_{k,i}^{(l+1)} = \frac{\tilde{\varpi}_{k,i}^{(l+1)}}{\sum_{i=1}^{I}\tilde{\varpi}_{k,i}^{(l+1)}}, \omega_{k,j}^{(l+1)} = \frac{\tilde{\omega}_{k,j}^{(l+1)}}{\sum_{j=1}^{J}\tilde{\omega}_{k,j}^{(l+1)}} \quad (36)$$

where the parameters $\tilde{\varpi}_{k,i}^{(l+1)}$ and $\tilde{\omega}_{k,j}^{(l+1)}$ are given by

$$\begin{aligned}
\tilde{\varpi}_{k,i}^{(l+1)} = \; & \exp\{a_{0,i}b_{0,i} + (a_{0,i}-1)\mathrm{E}^{(l+1)}[\log\lambda_k] \\
& - b_{0,i}\mathrm{E}^{(l+1)}[\lambda_k] - \log\Gamma(a_{0,i}) + \mathrm{E}^{(l)}[\log\pi_{k,i}]\}
\end{aligned} \quad (37)$$

$$\begin{aligned}
\tilde{\omega}_{k,j}^{(l+1)} = \; & \exp\{e_{0,j}f_{0,j} + (e_{0,j}-1)\mathrm{E}^{(l+1)}[\log\varepsilon_k] \\
& - f_{0,j}\mathrm{E}^{(l+1)}[\varepsilon_k] - \log\Gamma(e_{0,j}) + \mathrm{E}^{(l)}[\log\gamma_{k,j}]\}
\end{aligned} \quad (38)$$

where $\Gamma(.)$ denotes the gamma function [29].

Proof: See Appendix B ■

Proposition 3: Set $\Delta = \pi_{k,i}$ and substituting (25) in (24), $q^{(l+1)}(\pi_{k,i})$ is approximated by a Dirichlet distribution, and let $\Delta = \gamma_{k,j}$ and substituting (25) in (24), $q^{(l+1)}(\gamma_{k,j})$ is approximated by a Dirichlet distribution, i.e.,

$$q^{(l+1)}(\pi_{k,i}) = \mathrm{Dir}(\pi_{k,i}; \alpha_{k,i}^{(l+1)}) \quad (39)$$

$$q^{(l+1)}(\gamma_{k,j}) = \mathrm{Dir}(\gamma_{k,j}; \beta_{k,j}^{(l+1)}) \quad (40)$$

where the parameters $\alpha_{k,i}^{(l+1)}$ and $\beta_{k,j}^{(l+1)}$ are given by

$$\alpha_{k,i}^{(l+1)} = \alpha_{0,i} + \mathrm{E}^{(l+1)}[y_{k,i}] \quad (41)$$

$$\beta_{k,j}^{(l+1)} = \beta_{0,j} + \mathrm{E}^{(l+1)}[\rho_{k,j}]. \quad (42)$$

Proof: See Appendix C ■

Proposition 4: Set $\Delta = \mathbf{R}_k$ and employing (25) in (24), $q^{(l+1)}(\mathbf{R}_k)$ is an inverse Wishart, and let $\Delta = \mathbf{Q}_k$ and employing (25) in (24), $q^{(l+1)}(\mathbf{Q}_k)$ is an inverse Wishart, i.e.,

$$q^{(l+1)}(\mathbf{R}_k) = \mathrm{IW}(\mathbf{R}_k; \mathrm{t}_{k|k}^{(l+1)}, \mathbf{T}_{k|k}^{(l+1)}) \quad (43)$$

$$q^{(l+1)}(\mathbf{Q}_k) = \mathrm{IW}(\mathbf{Q}_k; \mathrm{u}_{k|k}^{(l+1)}, \mathbf{U}_{k|k}^{(l+1)}) \quad (44)$$

where the parameters $\mathrm{t}_{k|k}^{(l+1)}$, $\mathbf{T}_{k|k}^{(l+1)}$, $\mathrm{u}_{k|k}^{(l+1)}$, and $\mathbf{U}_{k|k}^{(l+1)}$ are given by

$$\mathrm{t}_{k|k}^{(l+1)} = \mathrm{t}_{k|k-1} + 1 \quad (45)$$

$$\mathbf{T}_{k|k}^{(l+1)} = \mathbf{T}_{k|k-1} + \mathrm{E}^{(l+1)}[\lambda_k]\mathbf{A}_k^{(l)} \quad (46)$$

$$\mathrm{u}_{k|k}^{(l+1)} = \mathrm{u}_{k|k-1} + 1 \quad (47)$$

$$\mathbf{U}_{k|k}^{(l+1)} = \mathbf{U}_{k|k-1} + \mathrm{E}^{(l+1)}[\varepsilon_k]\mathbf{B}_k^{(l)}. \quad (48)$$

Proof: See Appendix D ■

The hidden state sufficient statistics can be obtained by the variational Kalman smoother [31]. In this article, the only difference between variational Kalman smoother and the standard Kalman smoothing algorithm is the use of expectations $\tilde{\mathbf{Q}}_k^{(l+1)}$ and $\tilde{\mathbf{R}}_k^{(l+1)}$ instead of point estimates. The $\tilde{\mathbf{Q}}_k^{(l+1)}$ and $\tilde{\mathbf{R}}_k^{(l+1)}$ are calculated as follows:

$$\tilde{\mathbf{Q}}_k^{(l+1)} = \frac{(\mathrm{E}^{(l+1)}[\mathbf{Q}_k{}^{-1}])^{-1}}{\mathrm{E}^{(l+1)}[\varepsilon_k]}, \tilde{\mathbf{R}}_k^{(l+1)} = \frac{(\mathrm{E}^{(l+1)}[\mathbf{R}_k{}^{-1}])^{-1}}{\mathrm{E}^{(l+1)}[\lambda_k]}. \quad (49)$$

The $q^{(l+1)}(\mathbf{x}_k)$ is a Gaussian distribution

$$q^{(l+1)}(\mathbf{x}_k) = \mathcal{N}(\mathbf{x}_k; \hat{\mathbf{x}}_{k|k}^{(l+1)}, \mathbf{P}_{k|k}^{(l+1)}) \quad (50)$$

where $\hat{\mathbf{x}}_{k|k}^{(l+1)}$ and $\mathbf{P}_{k|k}^{(l+1)}$ are [31]

$$\begin{cases} \hat{\mathbf{x}}_{k|k-1}^{(l+1)} = \mathbf{F}\hat{\mathbf{x}}_{k-1|k}^{(l)} \\ \mathbf{P}_{k|k-1}^{(l+1)} = \mathbf{F}\mathbf{P}_{k-1|k}^{(l)}\mathbf{F}^{\mathrm{T}} + \tilde{\mathbf{Q}}_k^{(l+1)} \\ \mathbf{K}_k^{(l+1)} = \mathbf{P}_{k|k-1}^{(l+1)}\mathbf{C}^{\mathrm{T}}(\mathbf{C}\mathbf{P}_{k|k-1}^{(l+1)}\mathbf{C}^{\mathrm{T}} + \tilde{\mathbf{R}}_k^{(l+1)})^{-1} \\ \hat{\mathbf{x}}_{k|k}^{(l+1)} = \hat{\mathbf{x}}_{k|k-1}^{(l+1)} + \mathbf{K}_k^{(l+1)}(\mathbf{z}_k - \mathbf{C}\hat{\mathbf{x}}_{k|k-1}^{(l+1)}) \\ \mathbf{P}_{k|k}^{(l+1)} = \mathbf{P}_{k|k-1}^{(l+1)} - \mathbf{K}_k^{(l+1)}\mathbf{C}\mathbf{P}_{k|k-1}^{(l+1)} \\ \mathbf{Kb}_{k-1}^{(l+1)} = \mathbf{P}_{k-1|k}^{(l)}\mathbf{F}^{\mathrm{T}}(\mathbf{P}_{k|k-1}^{(l+1)})^{-1} \\ \hat{\mathbf{x}}_{k-1|k}^{(l+1)} = \hat{\mathbf{x}}_{k-1|k}^{(l)} + \mathbf{Kb}_{k-1}^{(l+1)}(\hat{\mathbf{x}}_{k|k}^{(l+1)} - \mathbf{F}\hat{\mathbf{x}}_{k-1|k}^{(l)}) \\ \mathbf{P}_{k-1|k}^{(l+1)} = \mathbf{Kb}_{k-1}^{(l+1)}(\mathbf{P}_{k|k}^{(l+1)} - \mathbf{P}_{k|k-1}^{(l+1)})(\mathbf{Kb}_{k-1}^{(l+1)})^{\mathrm{T}} \\ \qquad + \mathbf{P}_{k-1|k}^{(l)} \end{cases} \quad (51)$$

where $\hat{\mathbf{x}}_{k|k-1}^{(l+1)}$ and $\mathbf{P}_{k|k-1}^{(l+1)}$ stand for the modified predicted state estimate and covariance estimate at time k for $l+1$th iteration, respectively, $\mathbf{K}_k^{(l+1)}$ and $\mathbf{Kb}_{k-1}^{(l+1)}$ stand for the Kalman gain at time k in variational KF and at time $k-1$ in variational Kalman smoother for $l+1$th iteration, respectively, $\hat{\mathbf{x}}_{k-1|k}^{(l+1)}$ and $\mathbf{P}_{k-1|k}^{(l+1)}$ stand for the state estimate and covariance estimate at time $k-1$ for $l+1$th iteration in variational Kalman smoother, respectively. The procedure is initialized using $\hat{\mathbf{x}}_{k-1|k}^{(0)} = \hat{\mathbf{x}}_{k-1|k-1}$ and $\mathbf{P}_{k-1|k}^{(0)} = \mathbf{P}_{k-1|k-1}$.

The required expectations is calculated as follows. The $\mathbf{A}_k^{(l)}$ and $\mathbf{B}_k^{(l)}$ in (32) and (33) are calculated as

$$\mathbf{A}_k^{(l)} = (\mathbf{z}_k - \mathbf{C}\hat{\mathbf{x}}_{k|k}^{(l)})(\mathbf{z}_k - \mathbf{C}\hat{\mathbf{x}}_{k|k}^{(l)})^{\mathrm{T}} + \mathbf{C}\mathbf{P}_{k|k}^{(l)}\mathbf{C}^{\mathrm{T}} \quad (52)$$

$$\mathbf{B}_k^{(l)} = (\hat{\mathbf{x}}_{k|k}^{(l)} - \mathbf{F}\hat{\mathbf{x}}_{k-1|k}^{(l)})(\hat{\mathbf{x}}_{k|k}^{(l)} - \mathbf{F}\hat{\mathbf{x}}_{k-1|k}^{(l)})^{\mathrm{T}} + \mathbf{P}_{k|k}^{(l)} + \mathbf{F}\mathbf{P}_{k-1|k}^{(l)}\mathbf{F}^{\mathrm{T}} - \mathbf{F}\mathbf{M}_{k,k-1}^{(l)} - (\mathbf{F}\mathbf{M}_{k,k-1}^{(l)})^{\mathrm{T}} \quad (53)$$

where $\mathbf{M}_{k,k-1}^{(l)}$ denotes the cross-covariance between the hidden states at times k and $k-1$. $\mathbf{M}_{k,k-1}^{(l)}$ is calculated as $\mathbf{M}_{k,k-1}^{(l)} = \mathbf{Kb}_{k-1}^{(l)}\mathbf{P}_{k|k}^{(l)}$ [32].

Since $q^{(l+1)}(\lambda_k)$ and $q^{(l+1)}(\varepsilon_k)$ are Gamma distributions, then $\mathrm{E}^{(l+1)}[\lambda_k]$, $\mathrm{E}^{(l+1)}[\varepsilon_k]$, $\mathrm{E}^{(l+1)}[\log\lambda_k]$, and $\mathrm{E}^{(l+1)}[\log\varepsilon_k]$ are calculated as

$$\mathrm{E}^{(l+1)}[\lambda_k] = \frac{c_k^{(l+1)}}{d_k^{(l+1)}}, \mathrm{E}^{(l+1)}[\varepsilon_k] = \frac{g_k^{(l+1)}}{h_k^{(l+1)}} \quad (54)$$

$$\mathrm{E}^{(l+1)}[\log\lambda_k] = \psi(c_k^{(l+1)}) - \log d_k^{(l+1)} \quad (55)$$

$$\mathrm{E}^{(l+1)}[\log\varepsilon_k] = \psi(g_k^{(l+1)}) - \log h_k^{(l+1)} \quad (56)$$

where $\psi(.)$ is the digamma function [29].

Employing (34) and (35), $\mathrm{E}^{(l+1)}[y_{k,i}]$ and $\mathrm{E}^{(l+1)}[\rho_{k,j}]$ are computed as follows:

$$\mathrm{E}^{(l+1)}[y_{k,i}] = \varpi_{k,i}^{(l+1)}, \mathrm{E}^{(l+1)}[\rho_{k,j}] = \omega_{k,j}^{(l+1)}. \quad (57)$$

Employing (39) and (40), $\mathrm{E}^{(l+1)}[\pi_{k,i}]$ and $\mathrm{E}^{(l+1)}[\log\pi_{k,i}]$, $\mathrm{E}^{(l+1)}[\gamma_{k,j}]$, and $\mathrm{E}^{(l+1)}[\log\gamma_{k,j}]$ are calculated as

$$\mathrm{E}^{(l+1)}[\pi_{k,i}] = \alpha_{k,i}^{(l+1)} / \sum_{i=1}^{I}\alpha_{k,i}^{(l+1)} \quad (58)$$

$$\mathrm{E}^{(l+1)}[\log\pi_{k,i}] = \psi(\alpha_{k,i}^{(l+1)}) - \psi\left(\sum_{i=1}^{I}\alpha_{k,i}^{(l+1)}\right) \quad (59)$$

$$\mathrm{E}^{(l+1)}[\gamma_{k,j}] = \beta_{k,j}^{(l+1)} / \sum_{j=1}^{J}\beta_{k,j}^{(l+1)} \quad (60)$$

$$\mathrm{E}^{(l+1)}[\log\gamma_{k,j}] = \psi(\beta_{k,j}^{(l+1)}) - \psi\left(\sum_{i=1}^{I}\beta_{k,j}^{(l+1)}\right). \quad (61)$$

Employing (43) and (44), $\mathrm{E}^{(l+1)}[\mathbf{R}_k{}^{-1}]$ and $\mathrm{E}^{(l+1)}[\mathbf{Q}_k{}^{-1}]$ are computed as

$$\mathrm{E}^{(l+1)}[\mathbf{R}_k{}^{-1}] = (\mathrm{t}_{k|k}^{(l+1)} - m - 1)(\mathbf{T}_{k|k}^{(l+1)})^{-1} \quad (62)$$

$$\mathrm{E}^{(l+1)}[\mathbf{Q}_k{}^{-1}] = (\mathrm{u}_{k|k}^{(l+1)} - n - 1)(\mathbf{U}_{k|k}^{(l+1)})^{-1}. \quad (63)$$

The proposed VBAKF-GGM filter is given in Algorithm 1, where η and L are the threshold and the iteration number, respectively.

B. Discussion on the Proposed VBAKF-GGM

In this section, some parameters and features in the proposed VBAKF-GGM are analyzed.

Remark 1: If the numbers of Gamma mixture components are chosen as $I = 1$ and $J = 1$, and the priors of scale parameters are selected as $\mathbf{a}_0 = \mathbf{b}_0$ and $\mathbf{e}_0 = \mathbf{f}_0$, the proposed VBAKF-GGM is then formulated as a hierarchical GG form. Therefore, the proposed VBAKF-GGM has almost the same formulation as the RSTKF [26], [27] when $I = 1$, $J = 1$, $\mathbf{a}_0 = \mathbf{b}_0$, and $\mathbf{e}_0 = \mathbf{f}_0$. It can be seen that the RSTKF is a special case of the proposed VBAKF-GGM. In the proposed VBAKF-GGM, GGM distributions are proposed to model the parameters λ_k and ε_k. The GGM has prior information in the shape parameters $\mathbf{a}_0$ and $\mathbf{e}_0$, and rate parameters $\mathbf{b}_0$ and $\mathbf{f}_0$. From (7) and (8), we obtain $p(\lambda_k) = \sum_{i=1}^{I}\pi_{k,i}\lambda_{k,i}$ and $p(\varepsilon_k) = \sum_{j=1}^{J}\gamma_{k,j}\varepsilon_{k,j}$, where $\lambda_{k,i}$ and $\varepsilon_{k,j}$ denote the ith component for λ_k and jth component for ε_k, $0 < \lambda_k \leq 1$, and $0 < \varepsilon_k \leq 1$. In the proposed VBAKF-GGM, the numbers of Gamma mixture components are chosen as $I = J = 4$, and the initial $\mathrm{E}^{(0)}[\lambda_{k,i}]$ and $\mathrm{E}^{(0)}[\varepsilon_{k,j}]$ are chosen as $\mathrm{E}^{(0)}[\lambda_{k,1}] = \mathrm{E}^{(0)}[\varepsilon_{k,1}] = 1$, $\mathrm{E}^{(0)}[\lambda_{k,2}] = \mathrm{E}^{(0)}[\varepsilon_{k,2}] = 0.1$, $\mathrm{E}^{(0)}[\lambda_{k,3}] = \mathrm{E}^{(0)}[\varepsilon_{k,3}] = 0.01$, and $\mathrm{E}^{(0)}[\lambda_{k,4}] = \mathrm{E}^{(0)}[\varepsilon_{k,4}] = 0.001$. As $\mathrm{E}^{(0)}[\lambda_{k,i}] = \frac{a_{0,i}}{b_{0,i}}$ and $\mathrm{E}^{(0)}[\varepsilon_{k,j}] = \frac{e_{0,j}}{f_{0,j}}$, the priors are then selected as $\mathbf{a}_0 = \mathbf{e}_0 = [10, 10, 10, 10]$ and $\mathbf{b}_0 = \mathbf{f}_0 = [10, 100, 1000, 10\,000]$ for rough prior information in this article. The accurate prior information can be selected from the rough prior information by adaptive learning the mixing probabilities $\boldsymbol{\pi}_k$ and $\boldsymbol{\gamma}_k$.

Remark 2: The parameter τ on the proposed VBAKF-GGM is analyzed. Using (62) and (63) in (49), we obtain

$$\tilde{\mathbf{Q}}_k^{(l+1)} = \frac{\mathbf{U}_{k|k}^{(l+1)}}{\mathrm{E}^{(l+1)}[\varepsilon_k](\mathrm{u}_{k|k}^{(l+1)} - n - 1)} \quad (64)$$

$$\tilde{\mathbf{R}}_k^{(l+1)} = \frac{\mathbf{T}_{k|k}^{(l+1)}}{\mathrm{E}^{(l+1)}[\lambda_k](\mathrm{t}_{k|k}^{(l+1)} - m - 1)}. \quad (65)$$

Algorithm 1: Proposed VBAKF-GGM Filter.

Require:
$\hat{\mathbf{x}}_{k-1|k-1}, \mathbf{P}_{k-1|k-1}, \mathbf{F}, \mathbf{C}, \mathrm{t}_{k-1|k-1}, \mathbf{T}_{k-1|k-1}, \mathbf{u}_{k-1|k-1}, \mathbf{U}_{k-1|k-1}, \mathbf{z}_k, \mathbf{a}_0, \mathbf{b}_0, \mathbf{e}_0, \mathbf{f}_0, \boldsymbol{\alpha}_0, \boldsymbol{\beta}_0, \tau, I, J, \eta$, and L

1: Initial $\mathrm{t}_{k|k-1}$ and $\mathbf{T}_{k|k-1}$ as in (16)–(17), $\mathrm{u}_{k|k-1}$ and $\mathbf{U}_{k|k-1}$ as in (18)–(19), $\mathrm{E}^{(0)}[\lambda_k]=1$, $\mathrm{E}^{(0)}[\varepsilon_k]=1$, $\mathrm{E}^{(0)}[\log\lambda_k]=0$, $\mathrm{E}^{(0)}[\log\varepsilon_k]=0$, $\mathrm{E}^{(0)}[\mathbf{y}_k]=[1,0,\ldots,0]$, $\mathrm{E}^{(0)}[\boldsymbol{\rho}_k]=[1,0,\ldots,0]$, $\mathrm{E}^{(0)}[\pi_{k,i}]=1/I$, and $\mathrm{E}^{(0)}[\gamma_{k,j}]=1/J$
2: Calculate $\mathrm{E}^{(0)}[\log\pi_k]$ and $\mathrm{E}^{(0)}[\log\gamma_k]$ as $\mathrm{E}^{(0)}[\log\pi_k]=\log(\mathrm{E}^{(0)}[\pi_k])$ and $\mathrm{E}^{(0)}[\log\gamma_k]=\log(\mathrm{E}^{(0)}[\gamma_k])$
3: Calculate $\mathrm{E}^{(0)}[\mathbf{R}_k{}^{-1}]$ and $\mathrm{E}^{(0)}[\mathbf{Q}_k{}^{-1}]$ as $\mathrm{E}^{(0)}[\mathbf{R}_k{}^{-1}]=(\mathrm{t}_{k|k-1}-m-1)(\mathbf{T}_{k|k-1})^{-1}$ and $\mathrm{E}^{(0)}[\mathbf{Q}_k{}^{-1}]=(\mathrm{u}_{k|k-1}-n-1)(\mathbf{U}_{k|k-1})^{-1}$
4: Initial $\hat{\mathbf{x}}^{(0)}_{k-1|k}=\hat{\mathbf{x}}_{k-1|k-1}$, $\mathbf{P}^{(0)}_{k-1|k}=\mathbf{P}_{k-1|k-1}$, $\mathbf{M}^{(0)}_{k,k-1}=0$, $\hat{\mathbf{x}}^{(0)}_{k|k}=\mathbf{F}\hat{\mathbf{x}}_{k-1|k-1}$, and $\mathbf{P}^{(0)}_{k|k}=\mathbf{F}\mathbf{P}_{k-1|k-1}\mathbf{F}^{\mathrm{T}}+(\mathrm{E}^{(0)}[\mathbf{Q}_k{}^{-1}])^{-1}$
5: Calculate $\mathbf{A}^{(0)}_k$ and $\mathbf{B}^{(0)}_k$ as in (52)–(53)
6: **for** $l=0:L-1$ **do**
7: Update $q^{(l+1)}(\lambda_k)$ and $q^{(l+1)}(\varepsilon_k)$ as Gamma distributions, calculate $\mathrm{E}^{(l+1)}[\lambda_k]$, $\mathrm{E}^{(l+1)}[\varepsilon_k]$, $\mathrm{E}^{(l+1)}[\log\lambda_k]$, and $\mathrm{E}^{(l+1)}[\log\varepsilon_k]$ as in (54)–(56)
8: Update $q^{(l+1)}(y_{k,i})$ and $q^{(l+1)}(\rho_{k,j})$ as multinomial distributions, calculate $\mathrm{E}^{(l+1)}[y_{k,i}]$ and $\mathrm{E}^{(l+1)}[\rho_{k,j}]$ as in (57)
9: Update $q^{(l+1)}(\pi_{k,i})$ and $q^{(l+1)}(\gamma_{k,j})$ as Dirichlet distributions, calculate $\mathrm{E}^{(l+1)}[\pi_{k,i}]$ and $\mathrm{E}^{(l+1)}[\log\pi_{k,i}]$, $\mathrm{E}^{(l+1)}[\gamma_{k,j}]$, and $\mathrm{E}^{(l+1)}[\log\gamma_{k,j}]$ as in (58)–(61)
10: Update $q^{(l+1)}(\mathbf{R}_k{}^{-1})$ and $q^{(l+1)}(\mathbf{Q}_k{}^{-1})$ as inverse Wishart distributions, calculate $\mathrm{E}^{(l+1)}[\mathbf{R}_k{}^{-1}]$ and $\mathrm{E}^{(l+1)}[\mathbf{Q}_k{}^{-1}]$ as in (62)–(63)
11: Calculate $\tilde{\mathbf{Q}}^{(l+1)}_k$ and $\tilde{\mathbf{R}}^{(l+1)}_k$ as in (49)
12: Update state as in (51)
13: If $\frac{\|\hat{\mathbf{x}}^{(l+1)}_{k|k}-\hat{\mathbf{x}}^{(l)}_{k|k}\|}{\|\hat{\mathbf{x}}^{(l)}_{k|k}\|}\leq\eta$, stop iteration
14: **end for**
15: $\hat{\mathbf{x}}_{k|k}=\hat{\mathbf{x}}^{(L)}_{k|k}$, $\mathbf{P}_{k|k}=\mathbf{P}^{(L)}_{k|k}$, $\mathrm{t}_{k|k}=\mathrm{t}^{(L)}_{k|k}$, $\mathbf{T}_{k|k}=\mathbf{T}^{(L)}_{k|k}$, $\mathrm{u}_{k|k}=\mathrm{u}^{(L)}_{k|k}$, and $\mathbf{U}_{k|k}=\mathbf{U}^{(L)}_{k|k}$

Ensure:
$\hat{\mathbf{x}}_{k|k}, \mathbf{P}_{k|k}, \mathrm{t}_{k|k}, \mathbf{T}_{k|k}, \mathrm{u}_{k|k}$, and $\mathbf{U}_{k|k}$

Using (47), (48), (18), and (19) in (64), and using (45), (46), (16), and (17) in (65), we have

$$\begin{aligned}\tilde{\mathbf{Q}}^{(l+1)}_k &= \frac{\mathbf{U}_{k|k-1}+\mathrm{E}^{(l+1)}[\varepsilon_k]\mathbf{B}^{(l)}_k}{\mathrm{E}^{(l+1)}[\varepsilon_k](\mathrm{u}_{k|k-1}-n)}\\ &= \frac{\tau(\mathrm{u}_{k-1|k-1}-n-1)\mathrm{E}^{(L)}[\mathbf{Q}_{k-1}]/\mathrm{E}^{(l+1)}[\varepsilon_k]+\mathbf{B}^{(l)}_k}{\tau(\mathrm{u}_{k-1|k-1}-n-1)+1}\end{aligned} \tag{66}$$

$$\begin{aligned}\tilde{\mathbf{R}}^{(l+1)}_k &= \frac{\mathbf{T}_{k|k-1}+\mathrm{E}^{(l+1)}[\lambda_k]\mathbf{A}^{(l)}_k}{\mathrm{E}^{(l+1)}[\lambda_k](\mathrm{t}_{k|k-1}-m)}\\ &= \frac{\tau(\mathrm{t}_{k-1|k-1}-m-1)\mathrm{E}^{(L)}[\mathbf{R}_{k-1}]/\mathrm{E}^{(l+1)}[\lambda_k]+\mathbf{A}^{(l)}_k}{\tau(\mathrm{t}_{k-1|k-1}-m-1)+1}.\end{aligned} \tag{67}$$

According to (66) and (67), the τ is a weight for $\mathrm{E}^{(L)}[\mathbf{Q}_{k-1}]/\mathrm{E}^{(l+1)}[\varepsilon_k]$ and $\mathbf{B}^{(l)}_k$, and a weight for $\mathrm{E}^{(L)}[\mathbf{R}_{k-1}]/\mathrm{E}^{(l+1)}[\lambda_k]$ and $\mathbf{A}^{(l)}_k$. Considering the covariances of PMN are slowly time varying, the parameter τ is chosen as $\tau\in[0.9,1]$ [16].

Remark 3: Let $\mathrm{t}_{0|0}>m+1$ and $\mathrm{u}_{0|0}>n+1$. As parameter $\tau\in(0,1]$, from (16) and (18), we have

$$\mathrm{t}_{1|0}>m+1, \mathrm{u}_{1|0}>n+1. \tag{68}$$

Using $\bar{\mathbf{R}}_0>0$ and $\bar{\mathbf{Q}}_0>0$, we have

$$\mathrm{T}_{1|0}>0, \mathrm{U}_{1|0}>0. \tag{69}$$

Using $\mathbf{A}^{(l)}_1>0$ and $0<\mathrm{E}^{(l+1)}[\lambda_1]\leq 1$ in (46), and $\mathbf{B}^{(l)}_k>0$ and $0<\mathrm{E}^{(l+1)}[\varepsilon_1]\leq 1$ in (48), we have

$$\begin{aligned}&\mathrm{t}^{(l+1)}_{1|1}>\mathrm{m}+1, \mathbf{T}^{(l+1)}_{1|1}>0\\ &\mathrm{u}^{(l+1)}_{1|1}>n+1, \mathbf{U}^{(l+1)}_{1|1}>0.\end{aligned} \tag{70}$$

Then, we have $\mathrm{t}_{1|1}=\mathrm{t}^{(L)}_{1|1}>m+1$, $\mathbf{T}_{1|1}=\mathbf{T}^{(L)}_{1|1}>0$, $\mathrm{u}_{1|1}=\mathrm{u}^{(L)}_{1|1}>n+1$, and $\mathbf{U}_{1|1}=\mathbf{U}^{(L)}_{1|1}>0$. According to mathematical induction method, we have

$$\begin{aligned}&\mathrm{t}^{(l+1)}_{\mathrm{k|k}}>\mathrm{m}+1, \mathbf{T}^{(l+1)}_{\mathrm{k|k}}>0\\ &\mathrm{u}^{(l+1)}_{k|k}>n+1, \mathbf{U}^{(l+1)}_{k|k}>0.\end{aligned} \tag{71}$$

Utilizing (71) in (62) and (63), we have

$$\mathrm{E}^{(l+1)}[\mathbf{R}_k{}^{-1}]>0, \mathrm{E}^{(l+1)}[\mathbf{Q}_k{}^{-1}]>0. \tag{72}$$

From (72), the expectations of $\mathbf{R}_k{}^{-1}$ and $\mathbf{Q}_k{}^{-1}$ at every time k in every lth iteration are positive definite. Then, from (49), we have

$$\tilde{\mathbf{R}}^{(l+1)}_k>0, \tilde{\mathbf{Q}}^{(l+1)}_k>0. \tag{73}$$

It is summarized that the PMN covariances estimates in the proposed VBAKF-GGM filter are positive definite when $\mathrm{t}_{0|0}>m+1$, $\mathrm{u}_{0|0}>n+1$, $\bar{\mathbf{Q}}_0>0$, and $\bar{\mathbf{R}}_0>0$.

Remark 4: As the fixed-point method has a local convergence in the VB framework, the fixed-point iteration approach [16], [18], [26], [27] is widely used to simultaneous estimate the state and parameters.

IV. Simulations

The effectiveness of the proposed VBAKF-GGM for estimating the dynamical state with inaccurate nominal covariances of PMN under outlier environment is verified by computer simulations. The SSM with heavy-tailed PMN is formulated as

$$\mathbf{x}_k=\begin{bmatrix}\mathbf{I}_2 & t\mathbf{I}_2\\ \mathbf{0} & \mathbf{I}_2\end{bmatrix}\mathbf{x}_{k-1}+\mathbf{w}_{k-1} \tag{74}$$

$$\mathbf{z}_k=\begin{bmatrix}\mathbf{I}_2 & \mathbf{0}\end{bmatrix}\mathbf{x}_k+\mathbf{v}_k \tag{75}$$

where the state $\mathbf{x}_k=[x_{1\,k}, x_{2\,k}, \dot{x}_{1\,k}, \dot{x}_{2\,k}]$, $x_{1\,k}$, and $x_{2\,k}$ are, respectively, the x-axis position and y-axis position, $\dot{x}_{1\,k}$ and $\dot{x}_{2\,k}$ are, respectively, the x-axis velocity and y-axis velocity, $t=1\,\mathrm{s}$ stands for the sampling interval. The heavy-tailed PMN $\mathbf{w}_k$ and $\mathbf{v}_k$ are given as

$$\begin{aligned}p(\mathbf{w}_k)&=\Omega\mathcal{N}(\mathbf{0},\mathbf{Q}_k)+(1-\Omega)\mathcal{N}(\mathbf{0},\delta\mathbf{Q}_k)\\ p(\mathbf{v}_k)&=\Omega\mathcal{N}(\mathbf{0},\mathbf{R}_k)+(1-\Omega)\mathcal{N}(\mathbf{0},\delta\mathbf{R}_k).\end{aligned} \tag{76}$$

It means that $\mathbf{w}_k$ and $\mathbf{v}_k$ are generated by Gaussian distributions with nominal covariances $\mathbf{Q}_k$ and $\mathbf{R}_k$ with probability Ω, respectively, and are generated by Gaussian distributions with increased covariances $\delta\mathbf{Q}_k$ and $\delta\mathbf{R}_k$ with probability $1-\Omega$, respectively, where δ denotes the magnified factor. Similar to [18], the true $\mathbf{Q}_k$ and $\mathbf{R}_k$ are defined as

$$\mathbf{Q}_k=[6.5+0.5\cos(\pi k/N)]\,q\begin{bmatrix}\frac{\Delta t^3}{3}\mathbf{I}_2 & \frac{\Delta t^2}{2}\mathbf{I}_2\\ \frac{\Delta t^2}{2}\mathbf{I}_2 & \Delta t\mathbf{I}_2\end{bmatrix} \tag{77}$$

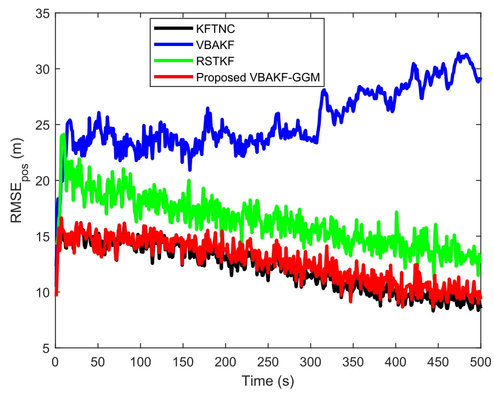

Fig. 2. $\mathrm{RMSE}_{\mathrm{pos}}$ by the proposed VBAKF-GGM and other filters.

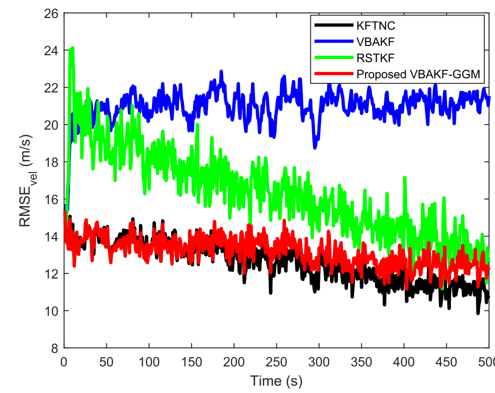

Fig. 3. $\mathrm{RMSE}_{\mathrm{vel}}$ by proposed VBAKF-GGM and other filters.

$$\mathbf{R}_k = [0.1 + 0.05\cos(\pi k/N)]\, r \begin{bmatrix} 1 & 0.5 \\ 0.5 & 1 \end{bmatrix} \tag{78}$$

where $N = 500\,\mathrm{s}$, $q = 1\,\mathrm{m}^2/\mathrm{s}^3$, and $r = 100\,\mathrm{m}^2$.

In this article, the inaccurate nominal covariances of PMN are set as $\bar{\mathbf{Q}}_k = \nu \times \mathbf{I}_4$ and $\bar{\mathbf{R}}_0 = \vartheta \times \mathbf{I}_2$. The proposed VBAKF-GGM is compared with KF with true noise covariances (KFTNC), VB adaptive KF (VBAKF) [18], and the existing RSTKF with adaptive selections of distribution parameters [27]. In the proposed VBAKF-GGM, we set $\tau = 1 - \exp(-4)$, $\nu = 1$, $\vartheta = 100$, $\mathrm{t}_{0|0} = 5$, $\mathrm{u}_{0|0} = 10$, $\mathbf{a}_0 = \mathbf{e}_0 = [10, 10, 10, 10]$, $\mathbf{b}_0 = \mathbf{f}_0 = [10, 100, 1000, 10\,000]$, $\eta = 10^{-10}$, and $L = 20$. The proposed VBAKF-GGM and the compared filters are performed by MATLAB with 2.70 GHz i7-7500.

The root mean square error (RMSE) and average RMSE (ARMSE) are chosen as performance metrics. $\mathrm{RMSE}_{\mathrm{pos}}$ and $\mathrm{RMSE}_{\mathrm{vel}}$ denote the RMSE on position and velocity, respectively. $\mathrm{ARMSE}_{\mathrm{pos}}$ and $\mathrm{ARMSE}_{\mathrm{vel}}$ denote the ARMSE on position and velocity, respectively, i.e.,

$$\mathrm{RMSE}_{\mathrm{pos}} = \sqrt{\frac{1}{M_C}\sum\nolimits_{i=1}^{M_C} (x_{1\,k}^i - \hat{x}_{1\,k}^i)^2 + (x_{2\,k}^i - \hat{x}_{2\,k}^i)^2} \tag{79}$$

$$\mathrm{ARMSE}_{\mathrm{pos}} = \sqrt{\frac{1}{M_C N}\sum\nolimits_{k=1}^{N}\sum\nolimits_{i=1}^{M_C} (x_{1\,k}^i - \hat{x}_{1\,k}^i)^2 + (x_{2\,k}^i - \hat{x}_{2\,k}^i)^2} \tag{80}$$

where $(x_{1\,k}^i, x_{2\,k}^i)$ and $(\hat{x}_{1\,k}^i, \hat{x}_{2\,k}^i)$ stand for the true position and estimated position at the ith run, respectively, and M_C is the number of runs.

First, the probability and magnified factor in (76) are set as $\Omega = 0.9$ and $\delta = 100$. $\mathrm{RMSE}_{\mathrm{pos}}$ and $\mathrm{RMSE}_{\mathrm{vel}}$ with $M_C = 1000$ of KFTNC, VBAKF, RSTKF, and the proposed VBAKF-GGM are given in Figs. 2 and 3, respectively. It is shown that the $\mathrm{RMSE}_{\mathrm{pos}}$ and $\mathrm{RMSE}_{\mathrm{vel}}$ of the proposed VBAKF-GGM are better than those of VBAKF and RSTKF, and are close to those of KFTNC. Compared with the RSTKF, the $\mathrm{ARMSE}_{\mathrm{pos}}$ and $\mathrm{ARMSE}_{\mathrm{vel}}$ of the proposed VBAKF-GGM are reduced by 22.03% and 19.04%. In a single step run, the KFTNC, VBAKF, RSTKF, and the proposed VBAKF-GGM need 0.02 ms, 0.78 ms, and 0.58 ms, and 1.03 ms, respectively. Thus, the proposed VBAKF-GGM outperforms the VBAKF and RSTKF, and the computation time of the proposed VBAKF-GGM is more than that of VBAKF and RSTKF.

The ARMSEs of the proposed VBAKF-GGM and other filters with different L are shown in Fig. 4. From Fig. 4, it is observed that the proposed VBAKF-GGM guarantees convergence for $L > 6$. Then, the $\mathrm{ARMSE}_{\mathrm{pos}}$ and $\mathrm{ARMSE}_{\mathrm{vel}}$ of the proposed VBAKF-GGM filter with

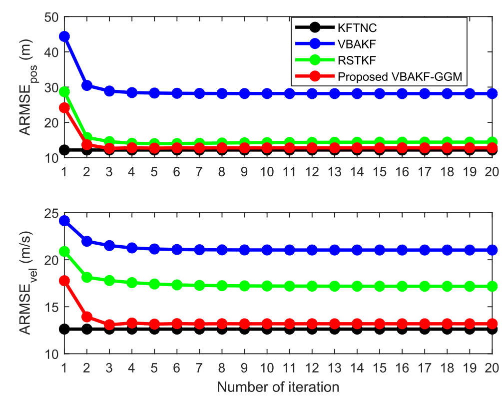

Fig. 4. $\mathrm{ARMSE}_{\mathrm{pos}}$ and $\mathrm{ARMSE}_{\mathrm{vel}}$ by the proposed VBAKF-GGM with $L = 1, 2, \ldots, 20$.

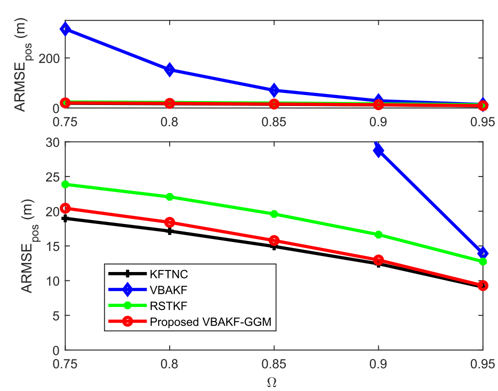

Fig. 5. $\mathrm{ARMSE}_{\mathrm{pos}}$ by proposed VBAKF-GGM and other filters with $\delta = 100$ and $\Omega \in [0.75, 0.95]$.

varying δ and Ω are investigated. Figs. 5 and 6 show the $\mathrm{ARMSE}_{\mathrm{pos}}$ and $\mathrm{ARMSE}_{\mathrm{vel}}$ by the KFTNC, VBAKF, RSTKF, and the proposed VBAKF-GGM filter when $\delta = 100$ and $\Omega \in [0.75, 0.95]$, respectively. The proposed VBAKF-GGM has consistently outperforms the VBAKF and RSTKF with varying Ω. Figs. 7 and 8 show the $\mathrm{ARMSE}_{\mathrm{pos}}$ and $\mathrm{ARMSE}_{\mathrm{vel}}$ by the KFTNC, VBAKF, RSTKF, and the proposed VBAKF-GGM filter when $\Omega = 0.9$ and $\delta \in [100, 500]$. It is shown from Figs. 7 and 8 that the performance of proposed VBAKF-GGM is also better than those of VBAKF and RSTKF in all cases. As the

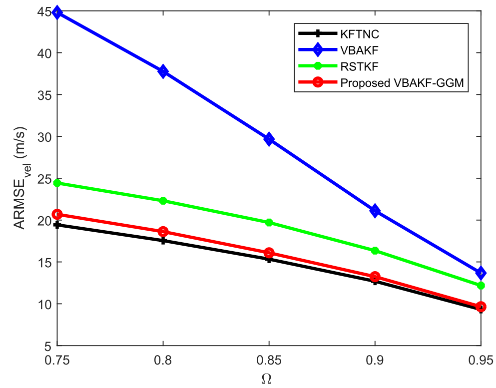

Fig. 6. $\mathrm{ARMSE}_{\mathrm{vel}}$ by proposed VBAKF-GGM and other filters with $\delta = 100$ and $\Omega \in [0.75, 0.95]$.

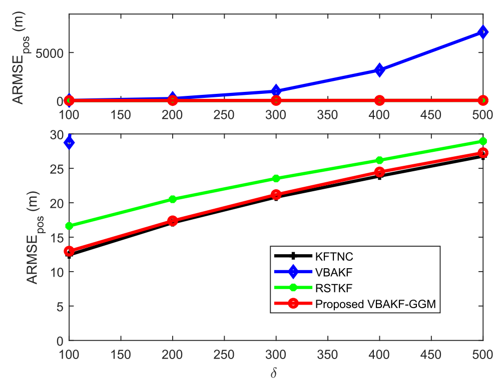

Fig. 7. $\mathrm{ARMSE}_{\mathrm{pos}}$ by proposed VBAKF-GGM and other filters with $\Omega = 0.9$ and $\delta \in [100, 500]$.

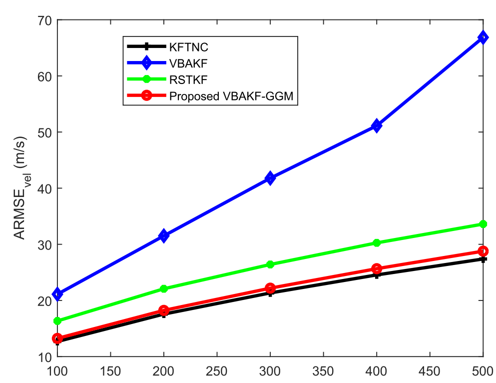

Fig. 8. $\mathrm{ARMSE}_{\mathrm{vel}}$ by proposed VBAKF-GGM and other filters with $\Omega = 0.9$ and $\delta \in [100, 500]$.

VBAKF is designed for Gaussian systems, the VBAKF fails in strongly HTN.

V. Conclusion

In the article, we develop a VBAKF-GGM to address the problem of state estimation with inaccurate nominal PMN covariances under outlier environment. In the proposed VBAKF-GGM, GGM distributions are developed to approximate the heavy-tailed PMN covariances. The accurate prior information for heavy-tailed PMN covariances is extracted by the proposed GGM distribution. Compared with the state-of-the-art filters, the developed VBAKF-GGM is demonstrated to have an improved performance.

Appendix

A. Proof of Proposition 1

Using (25) gives

$$\begin{aligned}
&\log p(\mathbf{z}_{1:k}|\Phi_k)p(\Phi_k|\Psi_k) \\
&\propto \tfrac{n}{2}\log\lambda_k - \tfrac{1}{2}\log|\mathbf{R}_k| - \tfrac{\lambda_k}{2}(\mathbf{z}_k - \mathbf{C}\mathbf{x}_k)^{\mathrm{T}}\mathbf{R}_k{}^{-1}(\mathbf{z}_k - \mathbf{C}\mathbf{x}_k) \\
&+ \prod_{i=1}^{I} y_{k,i}\left[a_{0,i}b_{0,i} + (a_{0,i}-1)\log\lambda_k - b_{0,i}\lambda_k\right] \\
&+ \prod_{i=1}^{I} y_{k,i}\log\pi_{k,i} + \tfrac{m}{2}\log\varepsilon_k - \tfrac{1}{2}\log|\mathbf{Q}_k| \\
&- \tfrac{\varepsilon_k}{2}(\mathbf{x}_k - \mathbf{F}\mathbf{x}_{k-1})^{\mathrm{T}}\mathbf{Q}_k{}^{-1}(\mathbf{x}_k - \mathbf{F}\mathbf{x}_{k-1}) + \prod_{j=1}^{J}\rho_{k,j}\log\gamma_{k,j} \\
&+ \prod_{j=1}^{J}\rho_{k,j}\left[e_{0,j}f_{0,j} + (e_{0,j}-1)\log\varepsilon_k - f_{0,j}\varepsilon_k\right] \\
&- \tfrac{n+\mathrm{u}_{k|k-1}+1}{2}\log|\mathbf{Q}_k| - \tfrac{1}{2}\mathrm{Tr}(\mathbf{U}_{k|k-1}\mathbf{Q}_k{}^{-1}) \\
&- \tfrac{m+\mathrm{t}_{k|k-1}+1}{2}\log|\mathbf{R}_k| - \tfrac{1}{2}\mathrm{Tr}(\mathbf{T}_{k|k-1}\mathbf{R}_k{}^{-1}) \\
&+ \log\mathrm{Dir}(\boldsymbol{\gamma}_k;\boldsymbol{\beta}_0) + \log\mathrm{Dir}(\boldsymbol{\pi}_k;\boldsymbol{\alpha}_0) - \tfrac{1}{2}\log\left|\mathbf{P}_{k-1|k-1}\right| \\
&- \tfrac{1}{2}(\mathbf{x}_{k-1} - \mathbf{F}\hat{\mathbf{x}}_{k-1|k-1})^{\mathrm{T}}\mathbf{P}_{k-1|k-1}{}^{-1}(\mathbf{x}_{k-1} - \mathbf{F}\hat{\mathbf{x}}_{k-1|k-1}).
\end{aligned} \tag{81}$$

Employing $\Delta = \lambda_k$ and $\Delta = \varepsilon_k$ in (24) and exploiting (81) yields

$$\begin{aligned}
q^{(l+1)}(\lambda_k) &\propto \tfrac{n}{2}\log\lambda_k - \tfrac{\lambda_k}{2}\mathrm{Tr}\left(\mathbf{A}_k^{(l)}\mathrm{E}^{(l)}[\mathbf{R}_k{}^{-1}]\right) \\
&+ \prod_{i=1}^{I}\mathrm{E}^{(l)}[y_{k,i}]\left[(a_{0,i}-1)\log\lambda_k - b_{0,i}\lambda_k\right]
\end{aligned} \tag{82}$$

$$\begin{aligned}
q^{(l+1)}(\varepsilon_k) &\propto \tfrac{m}{2}\log\varepsilon_k - \tfrac{\varepsilon_k}{2}\mathrm{Tr}\left(\mathbf{B}_k^{(l)}\mathrm{E}^{(l)}[\mathbf{Q}_k{}^{-1}]\right) \\
&+ \prod_{j=1}^{J}\mathrm{E}^{(l)}[\rho_{k,j}]\left[(e_{0,i}-1)\log\varepsilon_k - f_{0,i}\varepsilon_k\right].
\end{aligned} \tag{83}$$

Exploiting (28)–(31) in (82) and (83), we have

$$\log q^{(l+1)}(\lambda_k) \propto (c_k^{(l+1)} - 1)\log\lambda_k - d_k^{(l+1)}\lambda_k \tag{84}$$

$$\log q^{(l+1)}(\varepsilon_k) \propto (g_k^{(l+1)} - 1)\log\varepsilon_k - h_k^{(l+1)}\varepsilon_k. \tag{85}$$

According to (84) and (85), we can obtain (26) and (27).

B. Proof of Proposition 2

Using $\Delta = y_{k,i}$ and $\Delta = \rho_{k,j}$ in (24) and (81), we have

$$\begin{aligned}
\log q^{(l+1)}(y_{k,i}) &\propto y_{k,i}\left\{a_{0,i}b_{0,i} + (a_{0,i}-1)\mathrm{E}^{(l+1)}[\log\lambda_k]\right. \\
&\left. - b_{0,i}\mathrm{E}^{(l+1)}[\lambda_k] - \log\Gamma(a_{0,i}) + \mathrm{E}^{(l)}[\log\pi_{k,i}]\right\}
\end{aligned} \tag{86}$$

$$\begin{aligned}
\log q^{(l+1)}(\rho_{k,j}) &\propto \rho_{k,j}\left\{e_{0,j}f_{0,j} + (e_{0,j}-1)\mathrm{E}^{(l+1)}[\log\varepsilon_k]\right. \\
&\left. - f_{0,j}\mathrm{E}^{(l+1)}[\varepsilon_k] - \log\Gamma(e_{0,j}) + \mathrm{E}^{(l)}[\log\gamma_{k,j}]\right\}.
\end{aligned} \tag{87}$$

Exploiting (37) and (38) in (86) and (87), we have

$$\log q^{(l+1)}(y_{k,i}) \propto y_{k,i}\tilde{\varpi}_{k,i}^{(l+1)} \tag{88}$$

$$\log q^{(l+1)}(\rho_{k,j}) \propto \rho_{k,j}\tilde{\omega}_{k,j}^{(l+1)}. \tag{89}$$

According to (88) and (89), we can obtain (34) and (35).

C. Proof of Proposition 3

Utilizing $\Delta = \pi_{k,i}$ and $\Delta = \gamma_{k,j}$ in (24) and exploiting (81), we have

$$\log q^{(l+1)}(\pi_{k,i}) \propto \mathrm{E}^{(l+1)}[y_{k,i}] \log \pi_{k,i} + \alpha_{0,i} \log \pi_{k,i} \tag{90}$$

$$\log q^{(l+1)}(\gamma_{k,j}) \propto \mathrm{E}^{(l+1)}[\rho_{k,j}] \log \gamma_{k,j} + \beta_{0,j} \log \gamma_{k,j}. \tag{91}$$

Exploiting (41) and (42) in (90) and (91), we have

$$\log q^{(l+1)}(\pi_{k,i}) \propto \alpha_{k,i}^{(l+1)} \log \pi_{k,i} \tag{92}$$

$$\log q^{(l+1)}(\gamma_{k,j}) \propto \beta_{k,i}^{(l+1)} \log \gamma_{k,j}. \tag{93}$$

According to (92) and (93), we can obtain (39) and (40).

D. Proof of Proposition 4

Utilizing $\Delta = \mathbf{R}_k$ and $\Delta = \mathbf{Q}_k$ in (24) and using (81) yields

$$\begin{aligned}\log q^{(l+1)}(\mathbf{R}_k) &\propto -\tfrac{1}{2}\log|\mathbf{R}_k| - \tfrac{\mathrm{E}^{(l+1)}[\lambda_k]}{2}\mathrm{Tr}\left(\mathbf{A}_k^{(l)}\mathbf{R}_k{}^{-1}\right)\\ &-\tfrac{m+\mathrm{t}_{k|k-1}+1}{2}\log|\mathbf{R}_k| - \tfrac{1}{2}\mathrm{Tr}(\mathbf{T}_{k|k-1}\,\mathbf{R}_k{}^{-1})\end{aligned} \tag{94}$$

$$\begin{aligned}\log q^{(l+1)}(\mathbf{Q}_k) &\propto -\tfrac{1}{2}\log|\mathbf{Q}_k| - \tfrac{\mathrm{E}^{(l+1)}[\varepsilon_k]}{2}\mathrm{Tr}\left(\mathbf{B}_k^{(l)}\mathbf{Q}_k{}^{-1}\right)\\ &-\tfrac{n+\mathrm{u}_{k|k-1}+1}{2}\log|\mathbf{Q}_k| - \tfrac{1}{2}\mathrm{Tr}(\mathbf{U}_{k|k-1}\,\mathbf{Q}_k{}^{-1}).\end{aligned} \tag{95}$$

Exploiting (45)–(48) in (94) and (95), we have

$$\log q^{(l+1)}(\mathbf{R}_k) \propto -\frac{m+\mathrm{t}_{k|k}^{(l+1)}+1}{2}\log|\mathbf{R}_k| - \frac{1}{2}\mathrm{Tr}(\mathbf{T}_{k|k}^{(l+1)}\mathbf{R}_k{}^{-1}) \tag{96}$$

$$\log q^{(l+1)}(\mathbf{Q}_k) \propto -\frac{n+\mathrm{u}_{k|k}^{(l+1)}+1}{2}\log|\mathbf{Q}_k| - \frac{1}{2}\mathrm{Tr}(\mathbf{U}_{k|k}^{(l+1)}\mathbf{Q}_k{}^{-1}). \tag{97}$$

According to (96)–(97), we can obtain (43) and (44).

References

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

[2] H. Zhu, K. Yuen, L. Mihaylova, and H. Leung, "Overview of environment perception for intelligent vehicles," *IEEE Trans. Intell. Transp. Syst.*, vol. 18, no. 10, pp. 2584–2601, Oct. 2017.

[3] M. S. Arulampalam, S. Maskell, N. Gordon, and T. Clapp, "A tutorial on particle filters for online nonlinear/non-Gaussian Bayesian tracking," *IEEE Trans. Signal Process.*, vol. 50, no. 2, pp. 174–188, Feb. 2002.

[4] W. R. Wu, "Target racking with glint noise," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 29, no. 1, pp. 174–185, Jan. 1993.

[5] H. Zhu, H. Leung, and Z. He, "State estimation in unknown non-Gaussian measurement noise using variational Bayesian technique," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 49, no. 4, pp. 2601–2614, Oct. 2013.

[6] I. Bilik and J. Tabrikian, "MMSE-based filtering in presence of non-Gaussian system and measurement noise," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 46, no. 3, pp. 1153–1170, Jul. 2010.

[7] R. Mehra, "Approaches to adaptive filtering," *IEEE Trans. Autom. Control*, vol. AC-17, no. 5, pp. 693–698, Oct. 1972.

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

[9] B. J. Odelson, M. R. Rajamani, and J. B. Rawlings, "A new autocovariance least-squares method for estimating noise covariances," *Automatica*, vol. 42, no. 2, pp. 303–308, 2006.

[10] R. Kashyap, "Maximum likelihood identification of stochastic linear systems," *IEEE Trans. Autom. Control*, vol. AC-15, no. 1, pp. 25–34, Feb. 1970.

[11] S. Battilotti, F. Cacace, M. d'Angelo, and A. Germani, "The polynomial approach to the LQ non-Gaussian regulator problem through output injection," *IEEE Trans. Autom. Control*, vol. 64, no. 2, pp. 538–552, Feb. 2019.

[12] F. Cacace, F. Conte, M. d' Angelo, and A. Germani, "Feedback polynomial filtering and control of non-Gaussian linear time-varying systems," *Syst. Control Lett.*, vol. 123, pp. 108–115, 2018.

[13] F. Cacace, F. Conte, M. d' Angelo, and A. Germani, "Filtering of systems with nonlinear measurements with an application to target tracking," *Int. J. Robust Nonlinear Control*, vol. 29, no. 14, pp. 4956–4970, 2019.

[14] B. Chen, X. Liu, H. Zhao, and J. C. Principe, "Maximum correntropy Kalman filter," *Automatica*, vol. 76, pp. 70–77, 2017.

[15] R. Izanloo, S. A. Fakoorian, H. S. Yazdi, and D. Simon, "Kalman filtering based on the maximum correntropy criterion in the presence of non-Gaussian noise," in *Proc. Annu. Conf. Inf. Sci. Syst.*, 2016, pp. 500–505.

[16] S. Särkkä 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.

[17] 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.

[18] Y. Huang, Y. Zhang, Z. Wu, N. Li, and J. A. Chambers, "A novel adaptive Kalman filter with inaccurate process and measurement noise covariance matrices," *IEEE Trans. Autom. Control*, vol. 63, no. 2, pp. 594–601, 2018.

[19] M. Roth, E. Özkan, and F. Gustafsson, "A Student's t filter for heavy tailed process and measurement noise," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process.*, 2013, pp. 5770–5774.

[20] R. Piché, S. Särkkä, and J. Hartikainen, "Recursive outlier-robust filtering and smoothing for nonlinear systems using the multivariate Student-t distribution," in *Proc. IEEE Int. Workshop Mach. Learn. Signal Process.*, 2012, pp. 1–6.

[21] H. Nurminen, T. Ardeshiri, R. Piché, and F. Gustafsson, "Robust inference for state-space models with skewed measurement noise," *IEEE Signal Process. Lett.*, vol. 22, no. 11, pp. 1898–1902, Nov. 2015.

[22] H. Zhu, H. Leung, and Z. He, "A variational Bayesian approach to robust sensor fusion based on Student-t distribution," *Inf. Sci.*, vol. 221, pp. 201–214, 2013.

[23] Y. Huang, Y. Zhang, and J. A. Chambers, "A novel Kullback-Leilber divergence minimization-based adaptive Student's t-filter," *IEEE Trans. Signal Process.*, vol. 67, no. 20, pp. 5417–5432, Oct. 2019.

[24] P. Dong, Z. Jing, H. Leung, K. Shen, and J. Wang, "Student-t mixture labeled multi-Bernoulli filter for multi-target tracking with heavy-tailed noise," *Signal Process.*, vol. 152, pp. 331–339, 2018.

[25] Y. Huang, Y. Zhang, N. Li, and J. A. Chambers, "Robust Student's t-based nonlinear filter and smoother," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 52, no. 5, pp. 2586–2596, Oct. 2016.

[26] Y. Huang, Y. Zhang, N. Li, Z. Wu, and J. A. Chambers, "A novel robust Student's t-based Kalman filter," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 53, no. 3, pp. 1545–1554, Jun. 2017.

[27] Y. Huang, Y. Zhang, B. Xu, Z. Wu, and J. A. Chambers, "A new outlier-robust Student's t based Gaussian approximate filter for cooperative localization," *IEEE/ASME Trans. Mechatronics*, vol. 22, no. 5, pp. 2380–2386, Oct. 2017.

[28] Y. Huang, Y. Zhang, Y. Zhao, and J. A. Chambers, "A novel robust Gaussian-Student's t mixture distribution based Kalman filter," *IEEE Trans. Signal Process.*, vol. 67, no. 13, pp. 3606–3620, Jul. 2019.

[29] M. J. Beal, "Variational algorthims for approximate Bayesian inference," Ph.D. dissertation, Gatsby Comput. Neurosci. Unit, University of London, London, U.K., 2003.

[30] C. M. Bishop, *Pattern Recognition and Machine Learning*. New York, NY, USA: Springer, 2006.

[31] M. J. Cassidy and W. D. Penny, "Bayesian nonstationary autoregressive models for biomedical signal analysis," *IEEE Trans. Biomed. Eng.*, vol. 49, no. 10, pp. 1142–1152, Oct. 2002.

[32] Y. Huang, Y. Zhang, N. Li, and J. Chambers, "A robust Gaussian approximate fixed-interval smoother for nonlinear systems with heavy-tailed process and measurement noises," *IEEE Signal Process. Lett.*, vol. 23, no. 4, pp. 468–472, Apr. 2016.