Robust Kalman Filters Based on Gaussian Scale Mixture Distributions with Application to Target Tracking

Yulong Huang, Yonggang Zhang, *Senior Member, IEEE,* Peng Shi, *Fellow, IEEE,* Zhemin Wu, Junhui Qian, Jonathon Chambers, *Fellow, IEEE*

***Abstract*—This paper proposes a new robust Kalman filtering framework for a linear system with non-Gaussian heavy-tailed and/or skewed state and measurement noises, where the Gaussian scale mixture (GSM) distributions are utilized to model the one-step prediction and likelihood probability density functions. The state vector, mixing parameters, scale matrices and shape parameters are simultaneously inferred utilizing the standard variational Bayesian approach. As the implementations of the proposed method, several solutions corresponding to some special GSM distributions are derived. The proposed robust Kalman filters are tested in a manoeuvring target tracking example. Simulation results show that the proposed robust Kalman filters have better estimation accuracy and smaller biases as compared with the existing state-of-the-art Kalman filters.**

***Index Terms*—State estimation, heavy-tailed noise, skewed noise, Gaussian scale mixture distribution, variational Bayesian, Kalman filter, target tracking**

I. Introduction

FILTERING in the context of state-space models is estimating the current state vector based on the noisy measurements from the initial time to the current time. The Kalman filter is an optimal state estimator and provides an unbiased minimum-variance estimate for a linear system with Gaussian state and measurement noises, which has been widely applied in many fields, such as robotics, target tracking, navigation, positioning, control and signal processing [1]–[5]. However, in some engineering applications, the state and measurement noises may have heavy-tailed and/or skewed non-Gaussian distributions. For example, for the problem of tracking an agile target which is observed in clutter, the heavy-tailed state noise and skewed measurement noise may be respectively induced by severe manoeuvering and measurement outliers from unreliable sensors [6]–[11]. The conventional Kalman filter and its some improved variants [12]–[14] suffer from performance degradation for such engineering applications with heavy-tailed and/or skewed state and measurement noises [9]–[11].

To solve the state estimation problem given a state-space model with heavy-tailed or skewed measurement noise, many robust Kalman filters have been proposed by employing the Student's t or skew-t distribution to model the measurement noise, such as the Student's t mixture filter [15], the Student's t based outlier robust Kalman filter [16]–[21], and the skew-t Kalman filter [8], [9]. However, the performance of these robust Kalman filters degrades dramatically for the state-space models with heavy-tailed state noise since they are all based on a Gaussian state noise model.

To solve the state estimation problem given a state-space model with heavy-tailed state and measurement noises, the Huber-based Kalman filter (HKF) and the maximum-correntropy-based Kalman filter (MCKF) have been presented, which are essentially generalized maximum likelihood estimators [22], [25]. The HKF is derived by minimizing a weighted combination of l_1 and l_2 norms of the prediction error and residual [22], [23], [24], and the MCKF is developed by maximizing the correntropy of the prediction error and residual [25]. Both the HKF and MCKF are able to suppress the increased estimation errors which are induced by heavy-tailed noises so that the negative effects are mitigated. Unfortunately, the characteristic of heavy tail inherent in state and measurement noises is not exploited in the designs of HKF and MCKF, which leads to limited estimation accuracy. To achieve better estimation performance, a reasonable scheme is to improve the modelling of the heavy-tailed non-Gaussian probability density function (PDF). The Student's t distribution has a heavier tail than the Gaussian distribution when the degrees of freedom (dof) parameter is less than infinity so that it can better model a heavy-tailed non-Gaussian PDF as compared with the Gaussian distribution. Based on this

This work was supported by the National Natural Science Foundation of China under Grant Nos. 61773133, 61633008, 61773131 and U1509217, the Natural Science Foundation of Heilongjiang Province Grant No. F2016008, the Fundamental Research Founds for the Central University of Harbin Engineering University under Grant Nos. HEUCFP201705 and HEUCF041702, the PhD Student Research and Innovation Fund of the Fundamental Research Founds for the Central Universities under Grant No. HEUGIP201706, China Scholarship Council Foundation, the Engineering and Physical Sciences Research Council (EPSRC) of the UK Grant No. EP/K014307/1, Australian Research Council (DP170102644), and 111 Project (B17048, B17017). Corresponding author is Y. G. Zhang.

Y. L. Huang and Y. G. Zhang are with the Department of Automation, Harbin Engineering University, Harbin 150001, China (e-mail: heuedu@163.com; zhangyg@hrbeu.edu.cn).

P. Shi is with the School of Electrical and Electronic Engineering, University of Adelaide, Adelaide, SA 5005, Australia, and also with the College of Engineering and Science, Victoria University, Melbourne, VIC 8001, Australia (e-mail: peng.shi@adelaide.edu.au).

Z. M. Wu is with the School of Electrical Engineering & Automation, Harbin Institute of Technology, Harbin 150080, China (e-mail: myemailabc@163.com).

J. H. Qian is with the Department of Electrical Engineering, University of Electronic Science and Technology of China, Chengdu, 611731, China (email: junhuiq123@163.com).

J. Chambers is with the Department of Automation, Harbin Engineering University, and also with the School of Electrical and Electronic Engineering, Newcastle University, Newcastle upon Tyne, NE1 7RU, UK (e-mail: Jonathon.Chambers@newcastle.ac.uk).

idea, both the Student's t based filter (STF) and the robust Student's t based Kalman filter (RSTKF) have been proposed, in which the Student's t distributions are employed to model the heavy-tailed one-step prediction and likelihood PDFs [6], [11], [26]–[30]. For the STF, the posterior PDF is directly approximated as Student's t based on Bayes' rule, and the moment matching approach is utilized to prevent the growth of the dof parameters so that the one-step prediction PDF of state and measurement is jointly Student's t with a common dof parameter [6]. On the other hand, for the RSTKF, both the one-step prediction and likelihood PDFs are formulated as the hierarchical Gaussian forms, based on which a Gaussian approximation to the posterior PDF is achieved using the standard variational Bayesian (VB) approach [11].

In engineering practice, skewed state noise or skewed measurement noise may be induced by impulsive interference or outliers [8], [9], [31]. Unfortunately, the existing STF and RSTKF suffer from performance degradation for a non-Gaussian system with skewed noises due to the use of the symmetric Student's t distribution. Furthermore, for some engineering applications, there may be some non-Gaussian distributions that can better model the heavy-tailed PDF as compared with the Student's t distribution. As a result, the estimation performance can be further improved. Therefore, it is necessary to propose a new robust Kalman filtering framework which is able to deal with both skewed noises and heavy-tailed noises and thereby being suitable for more non-Gaussian distributions.

This paper proposes a new robust Kalman filtering framework for a linear system with non-Gaussian heavy-tailed and/or skewed state and measurement noises, in which the one-step prediction and likelihood PDFs are modelled as Gaussian scale mixture (GSM) distributions. The GSM distributions are formulated as hierarchical Gaussian forms given the prior PDFs of mixing parameters, and the prior distributions of the scale matrices and shape parameters are respectively selected as inverse-Wishart and Gaussian PDFs, based on which the state vector, mixing parameters, scale matrices and shape parameters are simultaneously inferred using the standard VB approach. As the implementations of the proposed method, several solutions corresponding to some special GSM distributions are derived, including the Pearson type VII distribution, the slash distribution, the variance gamma distribution, the generalized hyperbolic (GH) skew Student's t distribution, and the GH variance gamma distribution. Simulation results show that the proposed robust Kalman filters have better estimation accuracy and smaller biases but higher computational complexities than the existing state-of-the-art Kalman filters for the case of heavy-tailed state noise and skewed measurement noise.

The remainder of this paper is organized as follows. In Section II, notations and a brief description about GSM distribution are given. In Section III, a new robust Kalman filtering framework is proposed by employing the GSM distributions to model the one-step prediction and likelihood PDFs, and several particular solutions for some special GSM distributions are derived. In Section IV, the proposed robust Kalman filters are tested in a manoeuvring target tracking example and simulation results are given. Concluding remarks are provided in Section V.

II. Preliminaries

A. Notations

Throughout this paper, we denote $\mathbf{z}_{i:j} \triangleq \{\mathbf{z}_k | i \leq k \leq j\}$; $\mathrm{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ and $\mathrm{N}(\cdot; \boldsymbol{\mu}, \boldsymbol{\Sigma})$ denote respectively the multivariate Gaussian distribution and Gaussian PDF with mean vector $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$; $\mathrm{Be}(\cdot; a, b)$ denotes the Beta PDF with shape parameters a and b; $\mathrm{G}(\cdot; a, b)$ and $\mathrm{IG}(\cdot; a, b)$ denote respectively the Gamma PDF and inverse-Gamma PDF with shape parameter a and scale parameter b; $\mathrm{IW}(\cdot; \mu, \boldsymbol{\Sigma})$ denotes the inverse-Wishart PDF with dof parameter μ and inverse scale matrix $\boldsymbol{\Sigma}$; $\mathrm{N}_+(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ denotes the truncated Gaussian distribution with the closed positive orthant as support, location parameter $\boldsymbol{\mu}$ and squared-scale matrix $\boldsymbol{\Sigma}$; $\mathrm{E}_x[\cdot]$ is the expectation operator with respect to the PDF of x; $\delta(\cdot)$ denotes the Dirac delta function; $\mathbf{I}_n$ denotes the $n \times n$ identity matrix; $\log$ denotes the natural logarithm; the superscript "-1" denotes the inverse operation of a matrix; the superscript "T" denotes the transpose operation of a vector or matrix; $\bigcup$ denotes the union operation, and $|\cdot|$ and $\mathrm{tr}(\cdot)$ denote the determinant and trace operations of a matrix respectively.

B. GSM Distribution

In engineering practice, many types of non-Gaussian noise are induced by impulsive interferences or outliers, which often have heavy-tailed and/or skewed distributions. Such non-Gaussian noises can be modelled by a GSM distribution [32], [33]. To the best our knowledge, many popular non-Gaussian distributions are special cases of the GSM distribution, such as the Cauchy distribution, Student's t distribution, Pearson type-VII distribution, slash distribution, Laplace distribution, variance gamma distribution, GH skew Student's t distribution, and GH variance gamma distribution [31].

A random vector $\mathbf{x}$ has a GSM distribution if its PDF can be expressed as follows [33]

$$p(\mathbf{x}) = \int_0^{+\infty} \mathrm{N}(\mathbf{x}; \boldsymbol{\mu} + y\boldsymbol{\beta}, \boldsymbol{\Sigma}/\kappa(y))\pi(y)dy, \tag{1}$$

where $\boldsymbol{\mu}$ is a mean vector, $\boldsymbol{\Sigma}$ is a scale matrix, $y > 0$ is a mixing parameter, $\kappa(\cdot)$ is a positive scale function, $\pi(\cdot)$ is a mixing density defined on $\mathbb{R}^+$, and $\boldsymbol{\beta}$ is a shape parameter.

The shape parameter $\boldsymbol{\beta}$ dominates the symmetry of a GSM distribution, and a GSM distribution is symmetric when $\boldsymbol{\beta} = \mathbf{0}$ and non-symmetric when $\boldsymbol{\beta} \neq \mathbf{0}$. Exemplary GSM distributions and their parameters are listed in Table I. As an example, we provide the specific form for Pearson type-VII distribution. According to (1) and Table I, the PDF of Pearson type-VII distribution can be formulated as

$$\begin{aligned} &\mathrm{PV}(\mathbf{x}; \boldsymbol{\mu}, \boldsymbol{\Sigma}, \nu, \delta) = \int_0^{+\infty} \mathrm{N}(\mathbf{x}; \boldsymbol{\mu}, \boldsymbol{\Sigma}/y)\mathrm{G}(y; \frac{\nu}{2}, \frac{\delta}{2})dy, \\ &\text{s.t. } y > 0,\ \nu > 0,\ \delta > 0, \end{aligned} \tag{2}$$

where $\mathrm{PV}(\cdot; \boldsymbol{\mu}, \boldsymbol{\Sigma}, \nu, \delta)$ denotes the Pearson type-VII PDF with mean vector $\boldsymbol{\mu}$, scale matrix $\boldsymbol{\Sigma}$, and dof parameters ν and δ. Note that, the Pearson type-VII distribution becomes a

TABLE I: Exemplary GSM distributions and their parameters.

GSM distributions	Shape parameter	Scale function	Mixing density	Constraints
Pearson type-VII distribution	$\boldsymbol{\beta} = \mathbf{0}$	$\kappa(y) = y$	$\pi(y) = \mathrm{G}(y; \frac{\nu}{2}, \frac{\delta}{2})$	$y > 0$, $\nu > 0$, $\delta > 0$
Slash distribution	$\boldsymbol{\beta} = \mathbf{0}$	$\kappa(y) = y$	$\pi(y) = \mathrm{Be}(y; \nu, 1)$	$0 < y < 1$, $\nu > 0$
Variance gamma distribution	$\boldsymbol{\beta} = \mathbf{0}$	$\kappa(y) = y$	$\pi(y) = \mathrm{IG}(y; \frac{\nu}{2}, \frac{\nu}{2})$	$y > 0$, $\nu > 0$
GH skew Student's t distribution	$\boldsymbol{\beta} \neq \mathbf{0}$	$\kappa(y) = 1/y$	$\pi(y) = \mathrm{IG}(y; \frac{\nu}{2}, \frac{\nu}{2})$	$y > 0$, $\nu > 0$
GH variance gamma distribution	$\boldsymbol{\beta} \neq \mathbf{0}$	$\kappa(y) = 1/y$	$\pi(y) = \mathrm{G}(y; \frac{\nu}{2}, \frac{\nu}{2})$	$y > 0$, $\nu > 0$

Student's t distribution when $\nu = \delta$ and a Cauchy distribution when $\nu = \delta = 1$ respectively, and the variance gamma distribution degrades into a Laplace distribution when $\nu = 2$.

III. Robust Kalman Filtering Framework Based on GSM Distribution

A. Problem Statement

Consider a linear system that is represented by a discrete-time linear state-space model as follows

$$\mathbf{x}_k = \mathbf{F}_k \mathbf{x}_{k-1} + \mathbf{w}_{k-1}, \tag{3}$$

$$\mathbf{z}_k = \mathbf{H}_k \mathbf{x}_k + \mathbf{v}_k, \tag{4}$$

where (3) and (4) are respectively state and measurement equations, k represents the discrete time index, $\mathbf{x}_k \in \mathbb{R}^n$ is the state vector, $\mathbf{z}_k \in \mathbb{R}^m$ is the measurement vector, $\mathbf{F}_k \in \mathbb{R}^{n\times n}$ and $\mathbf{H}_k \in \mathbb{R}^{m\times n}$ are respectively known state transition and measurement matrices, and $\mathbf{w}_k \in \mathbb{R}^n$ and $\mathbf{v}_k \in \mathbb{R}^m$ are respectively state and measurement noise vectors. The initial state vector $\mathbf{x}_0$ has a Gaussian distribution, i.e., $\mathbf{x}_0 \sim \mathrm{N}(\hat{\mathbf{x}}_{0|0}, \mathbf{P}_{0|0})$, where $\hat{\mathbf{x}}_{0|0}$ and $\mathbf{P}_{0|0}$ denote the initial state estimate and the initial estimate error covariance matrix respectively. Moreover, $\mathbf{x}_0$, $\mathbf{w}_k$ and $\mathbf{v}_j$ are assumed to be mutually independent for any k and j.

Kalman filter is a minimum mean square error (MMSE) state estimator for linear state-space model (3)-(4) with Gaussian state and measurement noises. Unfortunately, in many engineering applications, the state and measurement noises may have heavy-tailed and/or skewed non-Gaussian distributions, which are often induced by impulsive interferences or outliers. The conventional Kalman filter exhibits poor estimation performance for such linear state-space model with heavy-tailed and/or skewed non-Gaussian state and measurement noises. Next, to solve this problem, a new robust Kalman filtering framework will be proposed by employing the GSM distributions to model the heavy-tailed and/or skewed non-Gaussian state and measurement noises.

B. A New Hierarchical Gaussian State-space Model based on GSM Distribution

The state and measurement noises have heavy-tailed and/or skewed distributions and are modelled as GSM distributed as follows

$$p(\mathbf{w}_{k-1}) = \int_0^{+\infty} \mathrm{N}(\mathbf{w}_{k-1}; \xi_k \boldsymbol{\beta}_1, \mathbf{Q}_{k-1}/\kappa_1(\xi_k)) \pi_1(\xi_k) d\xi_k, \tag{5}$$

$$p(\mathbf{v}_k) = \int_0^{+\infty} \mathrm{N}(\mathbf{v}_k; \lambda_k \boldsymbol{\beta}_2, \mathbf{R}_k/\kappa_2(\lambda_k)) \pi_2(\lambda_k) d\lambda_k, \tag{6}$$

where ξ_k, $\kappa_1(\cdot)$, $\pi_1(\cdot)$, $\boldsymbol{\beta}_1$ and $\mathbf{Q}_{k-1}$ are respectively the mixing parameter, positive scale function, mixing density, shape parameter and scale matrix of state noise, and λ_k, $\kappa_2(\cdot)$, $\pi_2(\cdot)$, $\boldsymbol{\beta}_2$ and $\mathbf{R}_k$ are respectively the mixing parameter, positive scale function, mixing density, shape parameter and scale matrix of measurement noise.

According to (3)-(6), the one-step prediction PDF $p(\mathbf{x}_k|\mathbf{z}_{1:k-1})$ and likelihood PDF $p(\mathbf{z}_k|\mathbf{x}_k)$ are formulated as

$$p(\mathbf{x}_k|\mathbf{z}_{1:k-1}) = \int_0^{+\infty} \mathrm{N}\left(\mathbf{x}_k; \mathbf{F}_k \hat{\mathbf{x}}_{k-1|k-1} + \xi_k \boldsymbol{\beta}_1, \mathbf{F}_k \mathbf{P}_{k-1|k-1} \mathbf{F}_k^{\mathrm{T}} + \mathbf{Q}_{k-1}/\kappa_1(\xi_k)\right) \pi_1(\xi_k) d\xi_k, \tag{7}$$

$$p(\mathbf{z}_k|\mathbf{x}_k) = \int_0^{+\infty} \mathrm{N}(\mathbf{z}_k; \mathbf{H}_k \mathbf{x}_k + \lambda_k \boldsymbol{\beta}_2, \mathbf{R}_k/\kappa_2(\lambda_k)) \times \pi_2(\lambda_k) d\lambda_k, \tag{8}$$

where the derivation of (7) is given in Appendix B.

It can be seen from (7) that the one-step prediction PDF $p(\mathbf{x}_k|\mathbf{z}_{1:k-1})$ is not a GSM distribution, which makes an analytical estimate of mixing parameter ξ_k unavailable. In this paper, to address this problem, the one-step prediction PDF $p(\mathbf{x}_k|\mathbf{z}_{1:k-1})$ is modelled as a GSM distribution, i.e.,

$$p(\mathbf{x}_k|\mathbf{z}_{1:k-1}) = \int_0^{+\infty} \mathrm{N}\left(\mathbf{x}_k; \mathbf{F}_k \hat{\mathbf{x}}_{k-1|k-1} + \xi_k \boldsymbol{\beta}_1, \boldsymbol{\Sigma}_k/\kappa_1(\xi_k)\right) \pi_1(\xi_k) d\xi_k, \tag{9}$$

where $\boldsymbol{\Sigma}_k$ denotes the scale matrix of the one-step prediction PDF.

According to (8)-(9), the one-step prediction PDF $p(\mathbf{x}_k|\mathbf{z}_{1:k-1})$ and likelihood PDF $p(\mathbf{z}_k|\mathbf{x}_k)$ can be written as hierarchical Gaussian forms as follows

$$p(\mathbf{x}_k|\mathbf{z}_{1:k-1}, \xi_k, \boldsymbol{\Sigma}_k, \boldsymbol{\beta}_1) = \mathrm{N}(\mathbf{x}_k; \mathbf{F}_k \hat{\mathbf{x}}_{k-1|k-1} + \xi_k \boldsymbol{\beta}_1, \boldsymbol{\Sigma}_k/\kappa_1(\xi_k)), \tag{10}$$

$$p(\mathbf{z}_k|\mathbf{x}_k, \lambda_k, \mathbf{R}_k, \boldsymbol{\beta}_2) = \mathrm{N}(\mathbf{z}_k; \mathbf{H}_k \mathbf{x}_k + \lambda_k \boldsymbol{\beta}_2, \mathbf{R}_k/\kappa_2(\lambda_k)), \tag{11}$$

where the prior PDFs of mixing parameters ξ_k and λ_k are given by

$$p(\xi_k) = \pi_1(\xi_k), \qquad p(\lambda_k) = \pi_2(\lambda_k). \tag{12}$$

The scale matrices $\mathbf{\Sigma}_k$ and $\mathbf{R}_k$ and shape parameters $\boldsymbol{\beta}_1$ and $\boldsymbol{\beta}_2$ are assumed to be inaccurate, and will be jointly inferred utilizing the standard VB approach in this paper. The prior distributions of the scale matrices and shape parameters are chosen as inverse-Wishart and Gaussian PDFs respectively, i.e.,

$$p(\mathbf{\Sigma}_k) = \mathrm{IW}(\mathbf{\Sigma}_k; t_k, \mathbf{T}_k),\ p(\mathbf{R}_k) = \mathrm{IW}(\mathbf{R}_k; u_k, \mathbf{U}_k), \tag{13}$$

$$p(\boldsymbol{\beta}_1) = \mathrm{N}(\boldsymbol{\beta}_1; \bar{\boldsymbol{\beta}}_1, \sigma_1\mathbf{I}_n),\ \ p(\boldsymbol{\beta}_2) = \mathrm{N}(\boldsymbol{\beta}_2; \bar{\boldsymbol{\beta}}_2, \sigma_2\mathbf{I}_m), \tag{14}$$

where t_k, u_k, $\mathbf{T}_k$ and $\mathbf{U}_k$ are the dof parameters and inverse scale matrices of prior distributions $p(\mathbf{\Sigma}_k)$ and $p(\mathbf{R}_k)$ respectively, and $\bar{\boldsymbol{\beta}}_1$ and $\bar{\boldsymbol{\beta}}_2$ are the nominal shape parameters of the state and measurement noises respectively, and σ_1 and σ_2 are instrumental parameters that are employed to dominate the confidence values for nominal shape parameters.

Since the nominal prediction error and measurement noise covariance matrices $\bar{\mathbf{P}}_{k|k-1}$ and $\bar{\mathbf{R}}_k$ contain a large amount of prior information about $\mathbf{\Sigma}_k$ and $\mathbf{R}_k$, the mean values of $\mathbf{\Sigma}_k$ and $\mathbf{R}_k$ are selected as $\bar{\mathbf{P}}_{k|k-1}$ and $\bar{\mathbf{R}}_k$ respectively, i.e.,

$$\frac{\mathbf{T}_k}{t_k - n - 1} = \bar{\mathbf{P}}_{k|k-1}, \qquad \frac{\mathbf{U}_k}{u_k - m - 1} = \bar{\mathbf{R}}_{\mathbf{k}}, \tag{15}$$

where the nominal prediction error covariance matrix $\bar{\mathbf{P}}_{k|k-1}$ is given by

$$\bar{\mathbf{P}}_{k|k-1} = \mathbf{F}_k\mathbf{P}_{k-1|k-1}\mathbf{F}_k^{\mathrm{T}} + \bar{\mathbf{Q}}_{k-1}, \tag{16}$$

and $\bar{\mathbf{Q}}_{k-1}$ denotes the nominal state noise covariance matrix.

Equations (10)-(16) form a new hierarchical Gaussian state-space model based on GSM distribution. Next, a new robust Kalman filtering framework will be proposed, in which the state vector, mixing parameters, scale matrices and shape parameters, i.e., $\mathbf{\Theta}_k \triangleq \{\mathbf{x}_k, \xi_k, \lambda_k, \boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \mathbf{\Sigma}_k, \mathbf{R}_k\}$, will be jointly estimated based on the constructed hierarchical Gaussian state-space model exploiting the standard VB approach.

C. The Proposed Robust Kalman Filtering Framework

In order to estimate the state vector, mixing parameters, scale matrices and shape parameters simultaneously, the joint posterior PDF $p(\mathbf{\Theta}_k|\mathbf{z}_{1:k})$ needs to be calculated. Unfortunately, the optimal solution of the joint posterior PDF is unavailable for the hierarchical Gaussian state-space model (10)-(16) because the Gamma and inverse-Wishart PDFs are not closed [34]. In this paper, the standard VB approach is employed to achieve an approximate solution for $p(\mathbf{\Theta}_k|\mathbf{z}_{1:k})$ as follows [35], [36]

$$p(\mathbf{\Theta}_k|\mathbf{z}_{1:k}) \approx q(\mathbf{x}_k)q(\xi_k)q(\lambda_k)q(\boldsymbol{\beta}_1)q(\boldsymbol{\beta}_2)q(\mathbf{\Sigma}_k)q(\mathbf{R}_k), \tag{17}$$

where $q(\cdot)$ denotes a free form factored approximation of true posterior PDF $p(\cdot)$, and the approximate posterior PDF satisfies the equation as follows [36]

$$\log q(\boldsymbol{\phi}) = \mathrm{E}_{\mathbf{\Theta}_k^{(-\boldsymbol{\phi})}}[\log p(\mathbf{\Theta}_k, \mathbf{z}_{1:k})] + c_{\boldsymbol{\phi}}, \tag{18}$$

where $\boldsymbol{\phi} \in \mathbf{\Theta}_k$ is an arbitrary element of the set $\mathbf{\Theta}_k$, and $\mathbf{\Theta}_k^{(-\boldsymbol{\phi})}$ is a subset of $\mathbf{\Theta}_k$, which has all elements in $\mathbf{\Theta}_k$ except for $\boldsymbol{\phi}$, i.e., $\{\boldsymbol{\phi}\} \bigcup \mathbf{\Theta}_k^{(-\boldsymbol{\phi})} = \mathbf{\Theta}_k$, and $c_{\boldsymbol{\phi}}$ denotes a constant value relative to the variable $\boldsymbol{\phi}$.

However, it is not possible to solve (18) analytically due to mutually dependent and coupled variational parameters. The fixed-point iteration is often employed to solve (18) and a local optimum solution can be achieved, in which the posterior PDF of $\boldsymbol{\phi}$ is approximated as $q^{(i+1)}(\boldsymbol{\phi})$ by using $q^{(i)}(\mathbf{\Theta}_k^{(-\boldsymbol{\phi})})$ to calculate the required expectations in (18) [36].

1) Variational Approximations of Posterior PDFs: Using (10)-(14), the joint PDF $p(\mathbf{\Theta}_k, \mathbf{z}_{1:k})$ can be formulated as

$$\begin{aligned} &p(\mathbf{\Theta}_k, \mathbf{z}_{1:k}) = \mathrm{N}(\mathbf{z}_k; \mathbf{H}_k\mathbf{x}_k + \lambda_k\boldsymbol{\beta}_2, \mathbf{R}_k/\kappa_2(\lambda_k)) \times \\ &\mathrm{N}(\mathbf{x}_k; \mathbf{F}_k\hat{\mathbf{x}}_{k-1|k-1} + \xi_k\boldsymbol{\beta}_1, \mathbf{\Sigma}_k/\kappa_1(\xi_k))\pi_1(\xi_k)\pi_2(\lambda_k) \times \\ &\mathrm{IW}(\mathbf{\Sigma}_k; t_k, \mathbf{T}_k)\mathrm{IW}(\mathbf{R}_k; u_k, \mathbf{U}_k)\mathrm{N}(\boldsymbol{\beta}_1; \bar{\boldsymbol{\beta}}_1, \sigma_1\mathbf{I}_n) \times \\ &\mathrm{N}(\boldsymbol{\beta}_2; \bar{\boldsymbol{\beta}}_2, \sigma_2\mathbf{I}_m)p(\mathbf{z}_{1:k-1}). \end{aligned} \tag{19}$$

Let $\boldsymbol{\phi} = \mathbf{x}_k$ and using (19) in (18), $q^{(i+1)}(\mathbf{x}_k)$ is updated as Gaussian, i.e.,

$$q^{(i+1)}(\mathbf{x}_k) = \mathrm{N}(\mathbf{x}_k; \hat{\mathbf{x}}_{k|k}^{(i+1)}, \mathbf{P}_{k|k}^{(i+1)}), \tag{20}$$

where $\hat{\mathbf{x}}_{k|k}^{(i+1)}$ and $\mathbf{P}_{k|k}^{(i+1)}$ are given by

$$\hat{\mathbf{x}}_{k|k-1}^{(i+1)} = \mathbf{F}_k\hat{\mathbf{x}}_{k-1|k-1} + \tilde{\mathbf{q}}_{k-1}^{(i)}, \tag{21}$$

$$\mathbf{K}_k^{(i+1)} = \tilde{\mathbf{P}}_{k|k-1}^{(i)}\mathbf{H}_k^T\left(\mathbf{H}_k\tilde{\mathbf{P}}_{k|k-1}^{(i)}\mathbf{H}_k^T + \tilde{\mathbf{R}}_k^{(i)}\right)^{-1}, \tag{22}$$

$$\hat{\mathbf{x}}_{k|k}^{(i+1)} = \hat{\mathbf{x}}_{k|k-1}^{(i+1)} + \mathbf{K}_k^{(i+1)}\left(\mathbf{z}_k - \mathbf{H}_k\hat{\mathbf{x}}_{k|k-1}^{(i+1)} - \tilde{\mathbf{r}}_k^{(i)}\right), \tag{23}$$

$$\mathbf{P}_{k|k}^{(i+1)} = \left(\mathbf{I}_n - \mathbf{K}_k^{(i+1)}\mathbf{H}_k\right)\tilde{\mathbf{P}}_{k|k-1}^{(i)}, \tag{24}$$

where $\tilde{\mathbf{q}}_{k-1}^{(i)}$ and $\tilde{\mathbf{r}}_k^{(i)}$ denote the modified mean vectors of the state and measurement noises respectively, and $\tilde{\mathbf{P}}_{k|k-1}^{(i)}$ and $\tilde{\mathbf{R}}_k^{(i)}$ denote the modified prediction error and measurement noise covariance matrices respectively, which are given by

$$\tilde{\mathbf{q}}_{k-1}^{(i)} = \mathrm{E}^{(i)}[\xi_k]\mathrm{E}^{(i)}[\boldsymbol{\beta}_1], \qquad \tilde{\mathbf{r}}_k^{(i)} = \mathrm{E}^{(i)}[\lambda_k]\mathrm{E}^{(i)}[\boldsymbol{\beta}_2], \tag{25}$$

$$\tilde{\mathbf{P}}_{k|k-1}^{(i)} = \frac{\left\{\mathrm{E}^{(i)}[\mathbf{\Sigma}_k^{-1}]\right\}^{-1}}{\mathrm{E}^{(i)}[\kappa_1(\xi_k)]}, \quad \tilde{\mathbf{R}}_k^{(i)} = \frac{\left\{\mathrm{E}^{(i)}[\mathbf{R}_k^{-1}]\right\}^{-1}}{\mathrm{E}^{(i)}[\kappa_2(\lambda_k)]}, \tag{26}$$

where the derivations of (20)-(26) are given in Appendix C.

Let $\boldsymbol{\phi} = \xi_k$ and $\boldsymbol{\phi} = \lambda_k$ and utilizing (19) in (18), $\log q^{(i+1)}(\xi_k)$ and $\log q^{(i+1)}(\lambda_k)$ can be calculated as

$$\begin{aligned} &\log q^{(i+1)}(\xi_k) = -\frac{1}{2}\kappa_1(\xi_k)\mathrm{tr}\left\{\mathbf{A}_k^{i+1}\mathrm{E}^{(i)}[\mathbf{\Sigma}_k^{-1}]\right\} + \\ &\xi_k\kappa_1(\xi_k)\left\{\mathrm{E}^{(i)}[\boldsymbol{\beta}_1]\right\}^{\mathrm{T}}\mathrm{E}^{(i)}[\mathbf{\Sigma}_k^{-1}]\mathbf{a}_k^{i+1} - \frac{1}{2}\xi_k^2\kappa_1(\xi_k) \times \\ &\mathrm{tr}\left\{\mathrm{E}^{(i)}[\boldsymbol{\beta}_1\boldsymbol{\beta}_1^{\mathrm{T}}]\mathrm{E}^{(i)}[\mathbf{\Sigma}_k^{-1}]\right\} + \frac{n}{2}\log\kappa_1(\xi_k) + \log\pi_1(\xi_k), \end{aligned} \tag{27}$$

$$\begin{aligned} &\log q^{(i+1)}(\lambda_k) = -\frac{1}{2}\kappa_2(\lambda_k)\mathrm{tr}\left\{\mathbf{B}_k^{i+1}\mathrm{E}^{(i)}[\mathbf{R}_k^{-1}]\right\} + \\ &\lambda_k\kappa_2(\lambda_k)\left\{\mathrm{E}^{(i)}[\boldsymbol{\beta}_2]\right\}^{\mathrm{T}}\mathrm{E}^{(i)}[\mathbf{R}_k^{-1}]\mathbf{b}_k^{i+1} - \frac{1}{2}\lambda_k^2\kappa_2(\lambda_k) \times \\ &\mathrm{tr}\left\{\mathrm{E}^{(i)}[\boldsymbol{\beta}_2\boldsymbol{\beta}_2^{\mathrm{T}}]\mathrm{E}^{(i)}[\mathbf{R}_k^{-1}]\right\} + \frac{m}{2}\log\kappa_2(\lambda_k) + \log\pi_2(\lambda_k), \end{aligned} \tag{28}$$

where the auxiliary parameters $\mathbf{A}_k^{i+1}$, $\mathbf{a}_k^{i+1}$, $\mathbf{B}_k^{i+1}$ and $\mathbf{b}_k^{i+1}$ are respectively given by

$$\mathbf{A}_k^{i+1} = \mathrm{E}^{(i+1)}\left[(\mathbf{x}_k - \mathbf{F}_k\hat{\mathbf{x}}_{k-1|k-1})(\mathbf{x}_k - \mathbf{F}_k\hat{\mathbf{x}}_{k-1|k-1})^{\mathrm{T}}\right], \tag{29}$$

$$\mathbf{a}_k^{i+1} = \mathrm{E}^{(i+1)}\left[\mathbf{x}_k - \mathbf{F}_k\hat{\mathbf{x}}_{k-1|k-1}\right], \tag{30}$$

$$\mathbf{B}_k^{i+1} = \mathrm{E}^{(i+1)}\left[(\mathbf{z}_k - \mathbf{H}_k\mathbf{x}_k)(\mathbf{z}_k - \mathbf{H}_k\mathbf{x}_k)^{\mathrm{T}}\right], \tag{31}$$

$$\mathbf{b}_k^{i+1} = \mathrm{E}^{(i+1)}\left[\mathbf{z}_k - \mathbf{H}_k\mathbf{x}_k\right]. \tag{32}$$

It can be seen from (27)-(28) that both $q^{(i+1)}(\xi_k)$ and $q^{(i+1)}(\lambda_k)$ can't be analytically updated for a general case. In this paper, to address this problem, $q^{(i+1)}(\xi_k)$ and $q^{(i+1)}(\lambda_k)$ are approximated as point distributions, i.e.,

$$q^{(i+1)}(\xi_k) \approx \delta(\xi_k - \xi_k^{(i+1)}), \quad q^{(i+1)}(\lambda_k) \approx \delta(\lambda_k - \lambda_k^{(i+1)}), \tag{33}$$

where $\xi_k^{(i+1)}$ and $\lambda_k^{(i+1)}$ are respectively the maximum a posterior (MAP) estimates of $q^{(i+1)}(\xi_k)$ and $q^{(i+1)}(\lambda_k)$, i.e.,

$$\xi_k^{(i+1)} = \underset{\xi_k>0}{\operatorname{argmax}} \log q^{(i+1)}(\xi_k), \tag{34}$$

$$\lambda_k^{(i+1)} = \underset{\lambda_k>0}{\operatorname{argmax}} \log q^{(i+1)}(\lambda_k). \tag{35}$$

Let $\boldsymbol{\phi} = \boldsymbol{\beta}_1$ and $\boldsymbol{\phi} = \boldsymbol{\beta}_2$ and exploiting (19) in (18), $q^{(i+1)}(\boldsymbol{\beta}_1)$ and $q^{(i+1)}(\boldsymbol{\beta}_2)$ are updated as Gaussian, i.e.,

$$q^{(i+1)}(\boldsymbol{\beta}_1) = \mathrm{N}(\boldsymbol{\beta}_1; \boldsymbol{\beta}_1^{(i+1)}, \mathbf{P}_{\boldsymbol{\beta}_1}^{(i+1)}), \tag{36}$$

$$q^{(i+1)}(\boldsymbol{\beta}_2) = \mathrm{N}(\boldsymbol{\beta}_2; \boldsymbol{\beta}_2^{(i+1)}, \mathbf{P}_{\boldsymbol{\beta}_2}^{(i+1)}), \tag{37}$$

where the mean vectors $\boldsymbol{\beta}_1^{(i+1)}$ and $\boldsymbol{\beta}_2^{(i+1)}$ and covariance matrices $\mathbf{P}_{\boldsymbol{\beta}_1}^{(i+1)}$ and $\mathbf{P}_{\boldsymbol{\beta}_2}^{(i+1)}$ are respectively given by

$$\mathbf{W}_{\boldsymbol{\beta}_1}^{(i+1)} = \sigma_1\mathrm{E}^{(i+1)}[\xi_k]\left[\sigma_1\left(\mathrm{E}^{(i+1)}[\xi_k]\right)^2\mathbf{I}_n + \bar{\mathbf{P}}_{k|k-1}^{(i+1)}\right]^{-1}, \tag{38}$$

$$\boldsymbol{\beta}_1^{(i+1)} = \bar{\boldsymbol{\beta}}_1 + \mathbf{W}_{\boldsymbol{\beta}_1}^{(i+1)}\left(\mathbf{a}_k^{(i+1)} - \mathrm{E}^{(i+1)}[\xi_k]\bar{\boldsymbol{\beta}}_1\right), \tag{39}$$

$$\mathbf{P}_{\boldsymbol{\beta}_1}^{(i+1)} = \sigma_1\mathbf{I}_n - \sigma_1\mathrm{E}^{(i+1)}[\xi_k]\mathbf{W}_{\boldsymbol{\beta}_1}^{(i+1)}, \tag{40}$$

$$\mathbf{W}_{\boldsymbol{\beta}_2}^{(i+1)} = \sigma_2\mathrm{E}^{(i+1)}[\lambda_k]\left[\sigma_2\left(\mathrm{E}^{(i+1)}[\lambda_k]\right)^2\mathbf{I}_m + \bar{\mathbf{R}}_k^{(i+1)}\right]^{-1}, \tag{41}$$

$$\boldsymbol{\beta}_2^{(i+1)} = \bar{\boldsymbol{\beta}}_2 + \mathbf{W}_{\boldsymbol{\beta}_2}^{(i+1)}\left(\mathbf{b}_k^{(i+1)} - \mathrm{E}^{(i+1)}[\lambda_k]\bar{\boldsymbol{\beta}}_2\right), \tag{42}$$

$$\mathbf{P}_{\boldsymbol{\beta}_2}^{(i+1)} = \sigma_2\mathbf{I}_m - \sigma_2\mathrm{E}^{(i+1)}[\lambda_k]\mathbf{W}_{\boldsymbol{\beta}_2}^{(i+1)}, \tag{43}$$

where the modified prediction error and measurement noise covariance matrices $\bar{\mathbf{P}}_{k|k-1}^{(i+1)}$ and $\bar{\mathbf{R}}_k^{(i+1)}$ are given by

$$\bar{\mathbf{P}}_{k|k-1}^{(i+1)} = \frac{\left\{\mathrm{E}^{(i)}[\boldsymbol{\Sigma}_k^{-1}]\right\}^{-1}}{\mathrm{E}^{(i+1)}[\kappa_1(\xi_k)]}, \quad \bar{\mathbf{R}}_k^{(i+1)} = \frac{\left\{\mathrm{E}^{(i)}[\mathbf{R}_k^{-1}]\right\}^{-1}}{\mathrm{E}^{(i+1)}[\kappa_2(\lambda_k)]}, \tag{44}$$

where the derivations of (36)-(44) are given in Appendix D.

Let $\boldsymbol{\phi} = \boldsymbol{\Sigma}_k$ and $\boldsymbol{\phi} = \mathbf{R}_k$ and employing (19) in (18), $q^{(i+1)}(\boldsymbol{\Sigma}_k)$ and $q^{(i+1)}(\mathbf{R}_k)$ are updated as inverse-Wishart, i.e.,

$$q^{(i+1)}(\boldsymbol{\Sigma}_k) = \mathrm{IW}(\boldsymbol{\Sigma}_k; t_k^{(i+1)}, \mathbf{T}_k^{(i+1)}), \tag{45}$$

$$q^{(i+1)}(\mathbf{R}_k) = \mathrm{IW}(\mathbf{R}_k; u_k^{(i+1)}, \mathbf{U}_k^{(i+1)}), \tag{46}$$

where the dof parameters $t_k^{(i+1)}$ and $u_k^{(i+1)}$ and inverse scale matrices $\mathbf{T}_k^{(i+1)}$ and $\mathbf{U}_k^{(i+1)}$ are respectively given by

$$t_k^{(i+1)} = t_k + 1, \qquad u_k^{(i+1)} = u_k + 1, \tag{47}$$

$$\mathbf{T}_k^{(i+1)} = \mathbf{T}_k + \mathbf{C}_k^{(i+1)}, \qquad \mathbf{U}_k^{(i+1)} = \mathbf{U}_k + \mathbf{D}_k^{(i+1)}, \tag{48}$$

$$\mathbf{C}_k^{(i+1)} = \mathrm{E}^{(i+1)}[\kappa_1(\xi_k)(\mathbf{x}_k - \mathbf{F}_k\hat{\mathbf{x}}_{k-1|k-1} - \xi_k\boldsymbol{\beta}_1) \times (\mathbf{x}_k - \mathbf{F}_k\hat{\mathbf{x}}_{k-1|k-1} - \xi_k\boldsymbol{\beta}_1)^{\mathrm{T}}], \tag{49}$$

$$\mathbf{D}_k^{(i+1)} = \mathrm{E}^{(i+1)}[\kappa_2(\lambda_k)(\mathbf{z}_k - \mathbf{H}_k\mathbf{x}_k - \lambda_k\boldsymbol{\beta}_2) \times (\mathbf{z}_k - \mathbf{H}_k\mathbf{x}_k - \lambda_k\boldsymbol{\beta}_2)^{\mathrm{T}}], \tag{50}$$

where the derivations of (45)-(50) are given in Appendix E.

After fixed-point iteration N, the approximate posterior PDFs of the state vector, mixing parameters, shape parameters and scale matrices are respectively updated as

$$q(\mathbf{x}_k) \approx \mathrm{N}(\mathbf{x}_k; \hat{\mathbf{x}}_{k|k}^{(N)}, \mathbf{P}_{k|k}^{(N)}) = \mathrm{N}(\mathbf{x}_k; \hat{\mathbf{x}}_{k|k}, \mathbf{P}_{k|k}), \tag{51}$$

$$q(\xi_k) \approx \delta(\xi_k - \xi_k^N), \qquad q(\lambda_k) \approx \delta(\lambda_k - \lambda_k^N), \tag{52}$$

$$q(\boldsymbol{\beta}_1) \approx \mathrm{N}(\boldsymbol{\beta}_1; \boldsymbol{\beta}_1^{(N)}, \mathbf{P}_{\boldsymbol{\beta}_1}^{(N)}), \tag{53}$$

$$q(\boldsymbol{\beta}_2) \approx \mathrm{N}(\boldsymbol{\beta}_2; \boldsymbol{\beta}_2^{(N)}, \mathbf{P}_{\boldsymbol{\beta}_2}^{(N)}), \tag{54}$$

$$q(\boldsymbol{\Sigma}_k) \approx \mathrm{IW}(\boldsymbol{\Sigma}_k; t_k^{(N)}, \mathbf{T}_k^{(N)}), \tag{55}$$

$$q(\mathbf{R}_k) \approx \mathrm{IW}(\mathbf{R}_k; u_k^{(N)}, \mathbf{U}_k^{(N)}). \tag{56}$$

2) Calculation of Expectations: Exploiting (33), (36)-(37) and (45)-(46), the required expectations can be given by

$$\mathrm{E}^{(i+1)}[\xi_k] = \xi_k^{(i+1)}, \qquad \mathrm{E}^{(i+1)}[\kappa_1(\xi_k)] = \kappa_1(\xi_k^{(i+1)}), \tag{57}$$

$$\mathrm{E}^{(i+1)}[\lambda_k] = \lambda_k^{(i+1)}, \quad \mathrm{E}^{(i+1)}[\kappa_2(\lambda_k)] = \kappa_2(\lambda_k^{(i+1)}), \tag{58}$$

$$\mathrm{E}^{(i+1)}[\boldsymbol{\beta}_1] = \boldsymbol{\beta}_1^{(i+1)}, \qquad \mathrm{E}^{(i+1)}[\boldsymbol{\beta}_2] = \boldsymbol{\beta}_2^{(i+1)}, \tag{59}$$

$$\mathrm{E}^{(i+1)}[\boldsymbol{\beta}_1\boldsymbol{\beta}_1^{\mathrm{T}}] = \mathbf{P}_{\boldsymbol{\beta}_1}^{(i+1)} + \boldsymbol{\beta}_1^{(i+1)}\left(\boldsymbol{\beta}_1^{(i+1)}\right)^{\mathrm{T}}, \tag{60}$$

$$\mathrm{E}^{(i+1)}[\boldsymbol{\beta}_2\boldsymbol{\beta}_2^{\mathrm{T}}] = \mathbf{P}_{\boldsymbol{\beta}_2}^{(i+1)} + \boldsymbol{\beta}_2^{(i+1)}\left(\boldsymbol{\beta}_2^{(i+1)}\right)^{\mathrm{T}}, \tag{61}$$

$$\mathrm{E}^{(i+1)}[\boldsymbol{\Sigma}_k^{-1}] = \left(t_k^{(i+1)} - n - 1\right)\left(\mathbf{T}_k^{(i+1)}\right)^{-1}, \tag{62}$$

$$\mathrm{E}^{(i+1)}[\mathbf{R}_k^{-1}] = \left(u_k^{(i+1)} - m - 1\right)\left(\mathbf{U}_k^{(i+1)}\right)^{-1}. \tag{63}$$

Utilizing (20), (33) and (36)-(37), we can calculate the auxiliary parameters $\mathbf{A}_k^{i+1}$, $\mathbf{a}_k^{i+1}$, $\mathbf{B}_k^{i+1}$, $\mathbf{b}_k^{i+1}$, $\mathbf{C}_k^{i+1}$ and $\mathbf{D}_k^{i+1}$ as follows

$$\mathbf{A}_k^{i+1} = \mathbf{P}_{k|k}^{(i+1)} + \left(\hat{\mathbf{x}}_{k|k}^{(i+1)} - \mathbf{F}_k\hat{\mathbf{x}}_{k-1|k-1}\right) \times \left(\hat{\mathbf{x}}_{k|k}^{(i+1)} - \mathbf{F}_k\hat{\mathbf{x}}_{k-1|k-1}\right)^{\mathrm{T}}, \tag{64}$$

$$\mathbf{a}_k^{i+1} = \hat{\mathbf{x}}_{k|k}^{(i+1)} - \mathbf{F}_k\hat{\mathbf{x}}_{k-1|k-1}, \tag{65}$$

$$\mathbf{B}_k^{i+1} = \left(\mathbf{z}_k - \mathbf{H}_k\hat{\mathbf{x}}_{k|k}^{(i+1)}\right)\left(\mathbf{z}_k - \mathbf{H}_k\hat{\mathbf{x}}_{k|k}^{(i+1)}\right)^{\mathrm{T}} + \mathbf{H}_k\mathbf{P}_{k|k}^{(i+1)}\mathbf{H}_k^{\mathrm{T}}, \tag{66}$$

TABLE II: The implementation pseudo-code for the proposed PTV-GHSST-KF at one time step.

Inputs: $\mathbf{z}_k$, $\hat{\mathbf{x}}_{k-1|k-1}$, $\mathbf{P}_{k-1|k-1}$, $\mathbf{F}_k$, $\mathbf{H}_k$, $\bar{\mathbf{Q}}_{k-1}$, $\bar{\mathbf{R}}_k$, $\bar{\boldsymbol{\beta}}_1$, $\bar{\boldsymbol{\beta}}_2$, σ_1, σ_2, t_k, u_k, ν_1, δ_1, ν_2, N.

1. Calculate $\bar{\mathbf{P}}_{k|k-1}$ using (16).

2. Calculate $\mathbf{T}_k$ and $\mathbf{U}_k$ using (15).

3. Initialization: $\mathrm{E}^{(0)}[\xi_k]=1$, $\mathrm{E}^{(0)}[\lambda_k]=1$, $\mathrm{E}^{(0)}[\kappa_1(\xi_k)]=1$, $\mathrm{E}^{(0)}[\kappa_2(\lambda_k)]=1$, $\mathrm{E}^{(0)}[\boldsymbol{\beta}_1]=\bar{\boldsymbol{\beta}}_1$, $\mathrm{E}^{(0)}[\boldsymbol{\beta}_2]=\bar{\boldsymbol{\beta}}_2$, $\mathrm{E}^{(0)}[\boldsymbol{\beta}_1\boldsymbol{\beta}_1^{\mathrm{T}}]=\sigma_1\mathbf{I}_n+\bar{\boldsymbol{\beta}}_1\bar{\boldsymbol{\beta}}_1^{\mathrm{T}}$, $\mathrm{E}^{(0)}[\boldsymbol{\beta}_2\boldsymbol{\beta}_2^{\mathrm{T}}]=\sigma_2\mathbf{I}_m+\bar{\boldsymbol{\beta}}_2\bar{\boldsymbol{\beta}}_2^{\mathrm{T}}$, $\mathrm{E}^{(0)}[\boldsymbol{\Sigma}_k^{-1}]=(t_k-n-1)\mathbf{T}_k^{-1}$, $\mathrm{E}^{(0)}[\mathbf{R}_k^{-1}]=(u_k-m-1)\mathbf{U}_k^{-1}$.

for $i=0:N-1$

4. Calculate $\tilde{\mathbf{q}}_{k-1}^{(i)}$, $\tilde{\mathbf{r}}_k^{(i)}$, $\tilde{\mathbf{P}}_{k|k-1}^{(i)}$, $\tilde{\mathbf{R}}_k^{(i)}$ using (25)-(26).

5. Update $q^{(i+1)}(\mathbf{x}_k)$ using (20)-(24).

6. Calculate $\mathbf{A}_k^{i+1}$, $\mathbf{a}_k^{i+1}$, $\mathbf{B}_k^{i+1}$ and $\mathbf{b}_k^{i+1}$ using (64)-(67).

7. Calculate η_1 and η_3 using (114)-(115).

8. Calculate $\xi_k^{(i+1)}$ and $\lambda_k^{(i+1)}$ using (124) and (137).

9. Update $q^{(i+1)}(\xi_k)$ and $q^{(i+1)}(\lambda_k)$ using (33).

10. Calculate $\mathrm{E}^{(i+1)}[\xi_k]$, $\mathrm{E}^{(i+1)}[\kappa_1(\xi_k)]$, $\mathrm{E}^{(i+1)}[\lambda_k]$ and $\mathrm{E}^{(i+1)}[\kappa_2(\lambda_k)]$ using (57)-(58).

11. Calculate $\bar{\mathbf{P}}_{k|k-1}^{(i+1)}$ and $\bar{\mathbf{R}}_k^{(i+1)}$ using (44).

12. Update $q^{(i+1)}(\boldsymbol{\beta}_1)$ and $q^{(i+1)}(\boldsymbol{\beta}_2)$ using (36)-(43).

13. Calculate $\mathrm{E}^{(i+1)}[\boldsymbol{\beta}_1]$, $\mathrm{E}^{(i+1)}[\boldsymbol{\beta}_2]$, $\mathrm{E}^{(i+1)}[\boldsymbol{\beta}_1\boldsymbol{\beta}_1^{\mathrm{T}}]$, $\mathrm{E}^{(i+1)}[\boldsymbol{\beta}_2\boldsymbol{\beta}_2^{\mathrm{T}}]$, $\mathbf{C}_k^{i+1}$ and $\mathbf{D}_k^{i+1}$ using (59)-(61) and (68)-(69).

14. Update $q^{(i+1)}(\boldsymbol{\Sigma}_k)$ and $q^{(i+1)}(\mathbf{R}_k)$ using (45)-(48).

15. Calculate $\mathrm{E}^{(i+1)}[\boldsymbol{\Sigma}_k^{-1}]$ and $\mathrm{E}^{(i+1)}[\mathbf{R}_k^{-1}]$ using (62)-(63).

end

16. $\hat{\mathbf{x}}_{k|k}=\hat{\mathbf{x}}_{k|k}^{(N)}$, $\mathbf{P}_{k|k}=\mathbf{P}_{k|k}^{(N)}$.

Outputs: $\hat{\mathbf{x}}_{k|k}$ and $\mathbf{P}_{k|k}$.

$$\mathbf{b}_k^{i+1}=\mathbf{z}_k-\mathbf{H}_k\hat{\mathbf{x}}_{k|k}^{(i+1)}, \tag{67}$$

$$\mathbf{C}_k^{(i+1)}=\kappa_1(\xi_k^{(i+1)})\left[\mathbf{P}_{k|k}^{(i+1)}+\left(\mathbf{a}_k^{i+1}-\xi_k^{(i+1)}\boldsymbol{\beta}_1^{(i+1)}\right)\times\left(\mathbf{a}_k^{i+1}-\xi_k^{(i+1)}\boldsymbol{\beta}_1^{(i+1)}\right)^{\mathrm{T}}+\left(\xi_k^{(i+1)}\right)^2\mathbf{P}_{\boldsymbol{\beta}_1}^{(i+1)}\right], \tag{68}$$

$$\mathbf{D}_k^{(i+1)}=\kappa_2(\lambda_k^{(i+1)})\left[\mathbf{H}_k\mathbf{P}_{k|k}^{(i+1)}\mathbf{H}_k^{\mathrm{T}}+\left(\mathbf{b}_k^{i+1}-\lambda_k^{(i+1)}\boldsymbol{\beta}_2^{(i+1)}\right)\left(\mathbf{b}_k^{i+1}-\lambda_k^{(i+1)}\boldsymbol{\beta}_2^{(i+1)}\right)^{\mathrm{T}}+\left(\lambda_k^{(i+1)}\right)^2\mathbf{P}_{\boldsymbol{\beta}_2}^{(i+1)}\right], \tag{69}$$

where the derivations of (64)-(69) are given in Appendix F.

The proposed robust Kalman filtering framework is composed of the variational approximations of posterior PDFs in (20)-(56) and the calculations of expectations in (57)-(69). It is seen from (27)-(28) and (33)-(35) that the proposed robust Kalman filtering framework depends on the scale functions $\kappa_1(\cdot)$ and $\kappa_2(\cdot)$ and mixing densities $\pi_1(\cdot)$ and $\pi_2(\cdot)$. Different robust Kalman filters can be obtained when different scale functions and mixing densities are utilized. Thus, the explicit expressions for the scale functions and mixing densities are essential to implement the proposed robust Kalman filtering framework. To this end, several particular solutions corresponding to some particular GSM distributions are derived in Appendix G. In order to illustrate how to implement the proposed robust Kalman filtering framework, an implementation example is presented. The Pearson type-VII distribution and GH skew Student's t distribution are respectively utilized to model the one-step prediction and likelihood PDFs, and the Pearson type-VII and GH skew Student's t based Kalman filter (PTV-GHSST-KF) is obtained. The implementation pseudo-code for the proposed PTV-GHSST-KF at one time step is given in Table II, where ν_1 and δ_1 are the dof parameters of the Pearson type-VII distribution, and ν_2 is the dof parameter for the GH skew Student's t distribution, and $\kappa_1(y)=y$ and $\kappa_2(y)=1/y$. Note that, for Pearson type-VII distribution, the prior parameters $\bar{\boldsymbol{\beta}}_1=\mathbf{0}$ and $\sigma_1=0$.

To implement the proposed robust Kalman filtering framework, the dof parameters t_k and u_k, the nominal state and measurement noise covariance matrices $\bar{\mathbf{Q}}_{k-1}$ and $\bar{\mathbf{R}}_k$, the nominal shape parameters $\bar{\boldsymbol{\beta}}_1$ and $\bar{\boldsymbol{\beta}}_2$, and the confidence parameters σ_1 and σ_2 require to be selected. Generally, in practical applications, the nominal parameters $\bar{\mathbf{Q}}_{k-1}$, $\bar{\mathbf{R}}_k$, $\bar{\boldsymbol{\beta}}_1$ and $\bar{\boldsymbol{\beta}}_2$ are respectively selected as normal noise covariance matrices and shape parameters that can be approximately obtained based on engineering experience and simulation/experiment study. Next, we discuss how to choose the dof parameters t_k and u_k and confidence parameters σ_1 and σ_2.

Firstly, we derive the specific forms of the modified prediction error and measurement noise covariance matrices $\tilde{\mathbf{P}}_{k|k-1}^{(i+1)}$ and $\tilde{\mathbf{R}}_k^{(i+1)}$ and the estimated shape parameters $\boldsymbol{\beta}_1^{(i+1)}$ and $\boldsymbol{\beta}_2^{(i+1)}$ at the $i+1$th iteration. Substituting (15)-(16), (47)-(48), (57) and (62)-(63) in (26) yields

$$\tilde{\mathbf{P}}_{k|k-1}^{(i+1)}=\left\{\left[(t_k-n-1)\left(\mathbf{F}_k\mathbf{P}_{k-1|k-1}\mathbf{F}_k^{\mathrm{T}}+\bar{\mathbf{Q}}_{k-1}\right)+\mathbf{C}_k^{(i+1)}\right]/\kappa_1(\xi_k^{(i+1)})\right\}/[(t_k-n-1)+1], \tag{70}$$

$$\tilde{\mathbf{R}}_k^{(i+1)}=\frac{\left[(u_k-m-1)\bar{\mathbf{R}}_k+\mathbf{D}_k^{(i+1)}\right]/\kappa_2(\lambda_k^{(i+1)})}{(u_k-m-1)+1}. \tag{71}$$

Utilizing (38)-(39), (41)-(42) and (57)-(58), $\boldsymbol{\beta}_1^{(i+1)}$ and $\boldsymbol{\beta}_2^{(i+1)}$ can be rewritten as

$$\boldsymbol{\beta}_1^{(i+1)}=\left[\left(\xi_k^{(i+1)}\right)^2\left(\bar{\mathbf{P}}_{k|k-1}^{(i+1)}\right)^{-1}+\frac{1}{\sigma_1}\mathbf{I}_n\right]^{-1}\times\left[\frac{1}{\sigma_1}\mathbf{I}_n\bar{\boldsymbol{\beta}}_1+\left(\xi_k^{(i+1)}\right)^2\left(\bar{\mathbf{P}}_{k|k-1}^{(i+1)}\right)^{-1}\frac{\mathbf{a}_k^{(i+1)}}{\xi_k^{(i+1)}}\right], \tag{72}$$

$$\boldsymbol{\beta}_2^{(i+1)}=\left[\left(\lambda_k^{(i+1)}\right)^2\left(\bar{\mathbf{R}}_k^{(i+1)}\right)^{-1}+\frac{1}{\sigma_2}\mathbf{I}_m\right]^{-1}\times\left[\frac{1}{\sigma_2}\mathbf{I}_m\bar{\boldsymbol{\beta}}_2+\left(\lambda_k^{(i+1)}\right)^2\left(\bar{\mathbf{R}}_k^{(i+1)}\right)^{-1}\frac{\mathbf{b}_k^{(i+1)}}{\lambda_k^{(i+1)}}\right]. \tag{73}$$

It is observed from (70)-(73) that the modified prediction error covariance matrix $\tilde{\mathbf{P}}_{k|k-1}^{(i+1)}$ is a weighted sum of priori information $\frac{\left(\mathbf{F}_k\mathbf{P}_{k-1|k-1}\mathbf{F}_k^{\mathrm{T}}+\bar{\mathbf{Q}}_{k-1}\right)}{\kappa_1(\xi_k^{(i+1)})}$ and innovation $\frac{\mathbf{C}_k^{(i+1)}}{\kappa_1(\xi_k^{(i+1)})}$ with weights (t_k-n-1) and 1 respectively; the modified

measurement noise covariance matrix $\tilde{\mathbf{R}}_k^{(i+1)}$ is a weighted sum of priori information $\frac{\bar{\mathbf{R}}_k}{\kappa_2(\lambda_k^{(i+1)})}$ and innovation $\frac{\mathbf{D}_k^{(i+1)}}{\kappa_2(\lambda_k^{(i+1)})}$ with weights $(u_k - m - 1)$ and 1 respectively; the estimated shape parameter $\boldsymbol{\beta}_1^{(i+1)}$ is a weighted sum of priori information $\bar{\boldsymbol{\beta}}_1$ and innovation $\frac{\mathbf{a}_k^{(i+1)}}{\xi_k^{(i+1)}}$ with weight matrices $\frac{1}{\sigma_1}\mathbf{I}_n$ and $\left(\xi_k^{(i+1)}\right)^2 \left(\bar{\mathbf{P}}_{k|k-1}^{(i+1)}\right)^{-1}$ respectively; and the estimated shape parameter $\boldsymbol{\beta}_2^{(i+1)}$ is a weighted sum of priori information $\bar{\boldsymbol{\beta}}_2$ and innovation $\frac{\mathbf{b}_k^{(i+1)}}{\lambda_k^{(i+1)}}$ with weight matrices $\frac{1}{\sigma_2}\mathbf{I}_m$ and $\left(\lambda_k^{(i+1)}\right)^2 \left(\bar{\mathbf{R}}_k^{(i+1)}\right)^{-1}$ respectively. The dof parameters t_k and u_k and confidence parameters σ_1 and σ_2 can be respectively used to adjust the effects of the nominal parameters $\bar{\mathbf{Q}}_{k-1}$, $\bar{\mathbf{R}}_k$, $\bar{\boldsymbol{\beta}}_1$ and $\bar{\boldsymbol{\beta}}_2$ on the modified prediction error and measurement noise covariance matrices $\tilde{\mathbf{P}}_{k|k-1}^{(i+1)}$ and $\tilde{\mathbf{R}}_k^{(i+1)}$ and the estimated shape parameters $\boldsymbol{\beta}_1^{(i+1)}$ and $\boldsymbol{\beta}_2^{(i+1)}$. In general, if the nominal noise covariance matrices $\bar{\mathbf{Q}}_{k-1}$ and $\bar{\mathbf{R}}_k$ are close to true noise covariance matrices and the state and measurement noises have slightly heavy-tailed and/or skewed distributions, the values of dof parameters t_k and u_k will need to increase properly and vice versa; and if the nominal shape parameters $\bar{\boldsymbol{\beta}}_1$ and $\bar{\boldsymbol{\beta}}_2$ near the true shape parameters, the confidence parameters σ_1 and σ_2 will require to reduce properly and vice versa. The explicit selections of parameters t_k, u_k, σ_1 and σ_2 depend on practical application scenarios.

IV. Simulation Study

The proposed robust Kalman filters and existing state-of-the-art Kalman filters are tested and compared in a manoeuvring target tracking example. The target moves with a constant velocity in a plane, whose positions are observed in clutter. The target is tracked using a constant velocity model, and the noise corrupted positions are used for measurement vectors. The cartesian coordinates and corresponding velocities are selected as a state vector, i.e., $\mathbf{x}_k \triangleq [x_k \ y_k \ \dot{x}_k \ \dot{y}_k]$, where x_k, y_k, $\dot{x}_k$ and $\dot{y}_k$ denote the cartesian coordinates and corresponding velocities respectively. The discrete-time linear state-space model is given by (3)-(4), and the state transition matrix $\mathbf{F}_k$ and measurement matrix $\mathbf{H}_k$ are given by [11]

$$\mathbf{F}_k = \begin{bmatrix} \mathbf{I}_2 & \Delta t \mathbf{I}_2 \\ \mathbf{0} & \mathbf{I}_2 \end{bmatrix}, \qquad \mathbf{H}_k = \begin{bmatrix} \mathbf{I}_2 & \mathbf{0} \end{bmatrix}, \tag{74}$$

where the sampling interval $\Delta t = 1\text{s}$.

In this simulation, an agile target is tracked using the noise corrupted positions observed in clutter. For such problem of maneuvering target tracking, the target may be lost due to severe manoeuvering, which may induce heavy-tailed state noise [6]. Moreover, the significant variations of radar reflections may result in position outliers when the target maneuvers severely, which may induce heavy-tailed or skewed measurement noise [37]. Outlier contaminated state and measurement noises are produced according to [8], [11]

$$\mathbf{w}_k \sim \begin{cases} \mathrm{N}(\mathbf{0}, \bar{\mathbf{Q}}) & \text{w.p. } 0.9 \\ \mathrm{N}(\mathbf{0}, 100\bar{\mathbf{Q}}) & \text{w.p. } 0.1 \end{cases}, \tag{75}$$

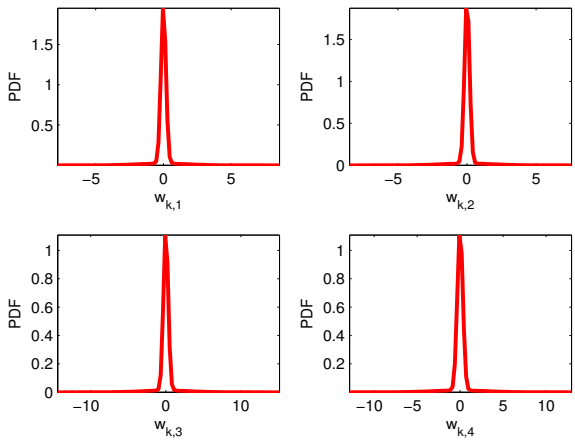


Fig. 1: Probability density curves of state noises.

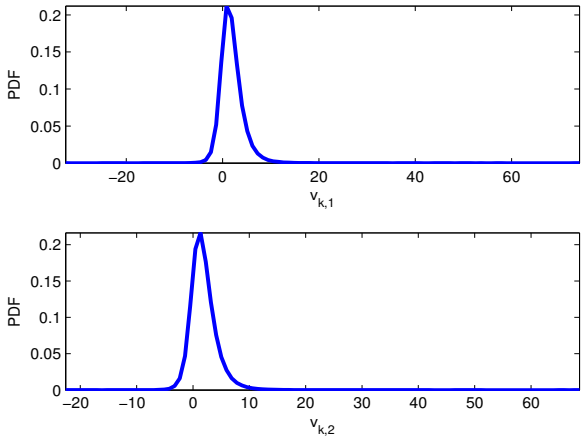


Fig. 2: Probability density curves of measurement noises.

$$\begin{cases} \mathbf{v}_k \sim \mathrm{N}(\boldsymbol{\Omega}\mathbf{u}_k, \boldsymbol{\Lambda}_k^{-1}\bar{\mathbf{R}}) \\ \mathbf{u}_k \sim \mathrm{N}_+(\mathbf{0}, \boldsymbol{\Lambda}_k^{-1}) \\ [\boldsymbol{\Lambda}_k]_{ii} \sim \mathrm{G}(\frac{\eta}{2}, \frac{\eta}{2}) \end{cases}, \tag{76}$$

where w.p. denotes "with probability", and $\bar{\mathbf{Q}} = q \begin{bmatrix} \frac{\Delta t^3}{3}\mathbf{I}_2 & \frac{\Delta t^2}{2}\mathbf{I}_2 \\ \frac{\Delta t^2}{2}\mathbf{I}_2 & \Delta t \mathbf{I}_2 \end{bmatrix}$ and $\bar{\mathbf{R}} = 10\mathbf{I}_2$ denote the nominal state and measurement noise covariance matrices respectively, and noise parameter $q = 0.1\text{m}^2/\text{s}^3$, and $\boldsymbol{\Omega} = 5\mathbf{I}_2$ with shape parameters as diagonal elements, and $\boldsymbol{\Lambda}_k$ is a 2×2 diagonal matrix whose random diagonal elements $[\boldsymbol{\Lambda}_k]_{ii}$ are independent and identically distributed, and $\mathbf{u}_k$ is an auxiliary random vector, and $\eta = 5$ is a dof parameter. Equation (75) indicates that the state noises are most frequently generated from a Gaussian distribution with the nominal state noise covariance matrix $\bar{\mathbf{Q}}$, and ten percent of state noise values are drawn from a Gaussian distribution with severely increased covariance matrix. The probability density curves of the state and measurement noises, which are generated in terms of (75)-(76), are shown in Fig. 1–Fig. 2 respectively. It is seen from Fig. 1–Fig. 2 that the state noise has a heavy-tailed and symmetric distribution and the measurement noise has a skewed distribution.

TABLE III: Implementation times in a single step run, ARMSEs and AAVBs when $N = 10$.

Filters	Standard KF	HKF	RSTKF	ST-GHSST-KF	SL-GHSST-KF
Time (ms)	0.021	0.420	0.072	0.159	0.158
$\mathrm{ARMSE}_{\mathrm{pos}}$ (m)	5.020	6.204	4.213	2.659	2.794
$\mathrm{ARMSE}_{\mathrm{vel}}$ (m/s)	2.730	2.897	2.640	2.316	2.394
$\mathrm{AAVB}_{\mathrm{pos}}$ (m)	3.831	3.743	3.722	0.801	0.589
$\mathrm{AAVB}_{\mathrm{vel}}$ (m/s)	0.115	0.101	0.113	0.078	0.089
Filters	VG-GHSST-KF	ST-GHVG-KF	SL-GHVG-KF	VG-GHVG-KF	
Time (ms)	0.162	0.159	0.159	0.163	
$\mathrm{ARMSE}_{\mathrm{pos}}$ (m)	3.098	2.695	2.808	3.189	
$\mathrm{ARMSE}_{\mathrm{vel}}$ (m/s)	2.369	2.298	2.374	2.384	
$\mathrm{AAVB}_{\mathrm{pos}}$ (m)	0.981	1.169	0.884	1.481	
$\mathrm{AAVB}_{\mathrm{vel}}$ (m/s)	0.081	0.080	0.096	0.085	

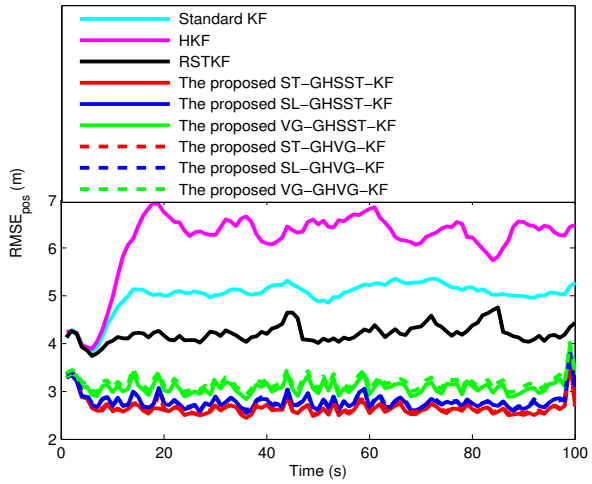


Fig. 3: RMSEs of the position when $N = 10$.

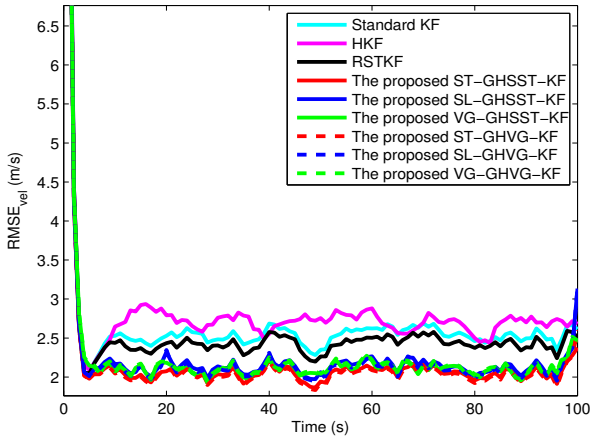


Fig. 4: RMSEs of the velocity when $N = 10$.

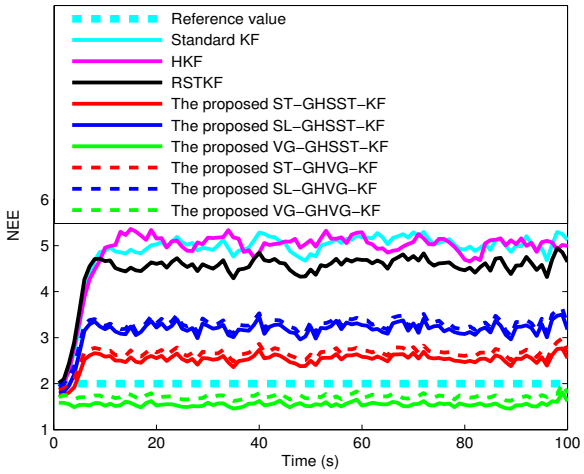


Fig. 5: NEEs when $N = 10$.

To handle the heavy-tailed state noise, we utilize the S-tudent's t distribution (a special case of the Pearson type VII distribution when dof parameters $\nu_1 = \delta_1$), slash distribution and variance gamma distribution to model the one-step prediction PDF, and the dof parameters of these GSM distributions are respectively selected as $\nu_1 = 5$, $\nu_1 = 1$ and $\nu_1 = 0.5$. On the other hand, to address the skewed measurement noise, the GH skew Student's t distribution and GH variance gamma distribution are employed to model measurement noise, and corresponding dof parameters are both set as $\nu_2 = 5$. Exploiting the five GSM distributions, six robust Kalman filters can be derived based on the proposed robust Kalman filtering framework, including the Student's t and GH skew Student's t based Kalman filter (ST-GHSST-KF), the slash and GH skew Student's t based Kalman filter (SL-GHSST-KF), the variance gamma and GH skew Student's t based Kalman filter (VG-GHSST-KF), the Student's t and GH variance gamma based Kalman filter (ST-GHVG-KF), the slash and GH variance gamma based Kalman filter (SL-GHVG-KF), and the variance gamma and GH variance gamma based Kalman filter (VG-GHVG-KF).

In this simulation, the existing standard Kalman filter (KF), the existing HKF [23], [24], the existing RSTKF [11], and the proposed robust Kalman filters are tested and compared. In the existing RSTKF, both the one-step prediction and likelihood PDFs are modelled as Student's t distributions, and

the state vector, auxiliary parameters and scale matrices are jointly inferred based on the constructed hierarchical Gaussian state-space model using the VB approach, from which the posterior PDF of state vector is approximated as Gaussian [11]. The existing RSTKF is a filtering estimation reference for a linear system with non-Gaussian heavy-tailed state and measurement noises since it can achieve the best estimation performance as compared with existing robust Kalman filters [11]. In the existing HKF, the tuning parameter is selected as $\gamma = 1.345$ [22]. In the existing RSTKF, the prior parameters are chosen as: $\omega = \nu = 5$ and $\tau = 5$ [11]. In the proposed robust Kalman filters, the prior parameters are set as: $\bar{\boldsymbol{\beta}}_1 = [0,0,0,0]^{\mathrm{T}}$, $\bar{\boldsymbol{\beta}}_2 = [2,2]^{\mathrm{T}}$, $\sigma_1 = 0$, $\sigma_2 = 0.01$, $t_k = 10$, and $u_k = 8$. The true initial state vector $\mathbf{x}_0 = [0,0,10,10]^{\mathrm{T}}$, and the initial estimation error covariance matrix $\mathbf{P}_{0|0} = \mathrm{diag}([100\ 100\ 100\ 100])$, and the initial state estimate $\hat{\mathbf{x}}_{0|0}$ is chosen from $N(\mathbf{x}_0, \mathbf{P}_{0|0})$ randomly. The number of measurements is 100, and the number of iteration is set as $N = 10$, and 1000 independent Monte Carlo runs are performed. All Kalman filters are coded with MATLAB and the used computer has an Intel Core i7-6500U CPU at 2.50 GHz.

The root mean square errors (RMSEs), the averaged root mean square errors (ARMSEs) and the averaged absolute value of biases (AAVBs) of position and velocity and the normalized estimation error (NEE) are utilized to evaluate the performance. We define the RMSE, ARMSE and AAVB of position and the NEE as follows [9], [11]

$$\mathrm{RMSE}_{\mathrm{pos}} = \sqrt{\frac{1}{M}\sum_{j=1}^{M}\left((x_k^j - \hat{x}_{k|k}^j)^2 + (y_k^j - \hat{y}_{k|k}^j)^2\right)}, \tag{77}$$

$$\mathrm{ARMSE}_{\mathrm{pos}} = \sqrt{\frac{1}{MT}\sum_{k=1}^{T}\sum_{j=1}^{M}\left((x_k^j - \hat{x}_{k|k}^j)^2 + (y_k^j - \hat{y}_{k|k}^j)^2\right)}, \tag{78}$$

$$\mathrm{AAVB}_{\mathrm{pos}} = \frac{1}{T}\sum_{k=1}^{T}\left|\frac{1}{M}\sum_{j=1}^{M}(x_k^j - \hat{x}_{k|k}^j)\right| + \frac{1}{T}\sum_{k=1}^{T}\left|\frac{1}{M}\sum_{j=1}^{M}(y_k^j - \hat{y}_{k|k}^j)\right|, \tag{79}$$

$$\mathrm{NEE} = \sqrt{\frac{1}{M}\sum_{j=1}^{M}\left(\mathbf{x}_k^j - \hat{\mathbf{x}}_{k|k}^j\right)^{\mathrm{T}}\left(\mathbf{P}_{k|k}^j\right)^{-1}\left(\mathbf{x}_k^j - \hat{\mathbf{x}}_{k|k}^j\right)}, \tag{80}$$

where (x_k^j, y_k^j) and $(\hat{x}_{k|k}^j, \hat{y}_{k|k}^j)$ are respectively the true position and the filtering estimate of position at the j-th Monte Carlo run, $\mathbf{x}_k^j$, $\hat{\mathbf{x}}_{k|k}^j$ and $\mathbf{P}_{k|k}^j$ are respectively the true state vector, the filtering estimate of state vector and corresponding estimation error covariance matrix at the j-th Monte Carlo run, $T = 100$s denotes the simulation time, and $M = 1000$ denotes the total number of Monte Carlo run. We can also formulate the RMSE, ARMSE and AAVB of velocity in a similar manner. The performance metrics AAVB and NEE are often employed to evaluate the bias of state estimate and the approximation accuracy of the posterior covariance matrix

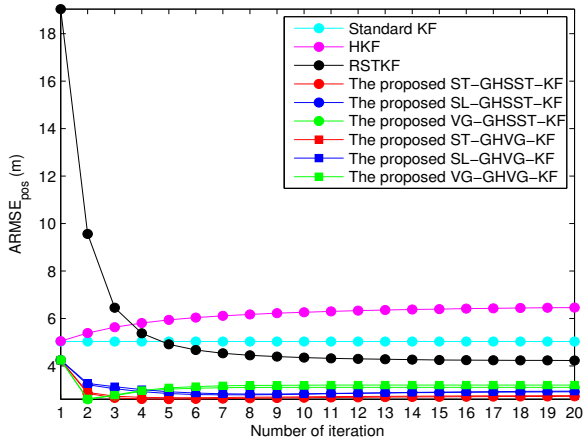


Fig. 6: ARMSEs of the position when $N = 1, 2, \ldots, 20$.

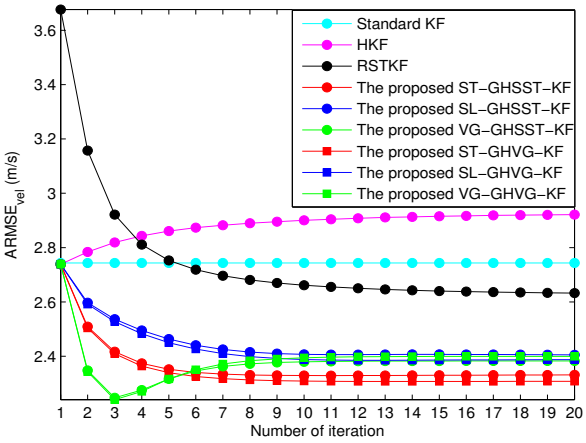


Fig. 7: ARMSEs of the velocity when $N = 1, 2, \ldots, 20$.

respectively. That is to say, if the AAVB is 0, then the state estimate is unbiased, and if the NEE is the square root of the state dimensionality 2, then the posterior covariance matrix is accurate [9].

Fig. 3–Fig. 5 respectively show the RMSEs of position and velocity and NEEs from different Kalman filters when the number of iteration $N = 10$. Also, the implementation times in a single step run, ARMSEs and AAVBs of different Kalman filters are given in Table III. We can observe from Fig. 3–Fig. 4 and Table III that the proposed robust Kalman filters have smaller RMSEs than the existing Kalman filters but greater implementation times as compared with the existing standard Kalman filter and RSTKF. We can also observe from Table III that the estimates of position and velocity from the existing Kalman filters and the proposed robust Kalman filters are all biased, especially the estimates of position, which is induced by the heavy-tailed state noises and skewed measurement noises. Fortunately, the AAVBs of position and velocity from the proposed robust Kalman filters are smaller than that from the existing Kalman filters. Furthermore, it is observed from Fig. 5 that the NEEs of the proposed robust Kalman filters

are closer to the reference value as compared with that of the existing Kalman filters. Thus, the proposed robust Kalman filters have better estimation accuracy and smaller biases but greater implementation times than the existing Kalman filters.

The ARMSEs of position and velocity from different Kalman filters when the number of iteration $N = 1, 2, \ldots, 20$ are respectively shown in Fig. 6–Fig. 7. It can be observed from Fig. 6–Fig. 7 that the proposed robust Kalman filters have better estimation accuracy than the existing Kalman filters when the number of iteration $N \geq 2$. We can also observe from Fig. 6–Fig. 7 that the ARMSEs from the existing HKF and RSTKF and the proposed robust Kalman filters are all convergent when $N \geq 9$. Thus, the number of iteration $N = 10$ is sufficient to achieve a local optimum.

The existing standard Kalman filter has poor estimation performance because it is based on Gaussian state and measurement noise models so that it is sensitive to state and measurement outliers. The HKF exhibits poor estimation performance since the characteristics of heavy tail and skewness inherent in state and measurement noises are not exploited in the designs of HKF. The RSTKF suffers from performance degradation since it is specially designed for symmetric heavy-tailed state and measurement noises and unsuitable for skewed measurement noise. The proposed robust Kalman filters achieve better estimation accuracy and smaller biases as compared with the existing Kalman filters, which is induced by the fact that the symmetric heavy-tailed GSM distributions and skewed GSM distributions are respectively employed to model the heavy-tailed one-step prediction PDF and skewed measurement noise in the proposed robust Kalman filters.

V. Conclusions

A new robust Kalman filtering framework for a linear system with non-Gaussian heavy-tailed and/or skewed state and measurement noises was proposed in this paper, where the GSM distributions are employed to model the one-step prediction and likelihood PDFs. The GSM distributions were formulated as hierarchical Gaussian forms given the prior PDFs of mixing parameters, and the prior distributions of the scale matrices and shape parameters were respectively selected as inverse-Wishart and Gaussian PDFs, based on which the state vector, mixing parameters, scale matrices and shape parameters were jointly inferred using the standard VB approach. As the implementations of the proposed method, several solutions corresponding to some special GSM distributions were derived. Simulation results have shown that the proposed robust Kalman filters have better estimation accuracy and smaller biases but higher computational complexities than the existing state-of-the-art Kalman filters.

Appendices

A. Gaussian Integral Formula

If $\mathbf{\Phi}$, $\mathbf{d}$, $\mathbf{\Sigma}$, $\boldsymbol{\mu}$ and $\mathbf{P}$ have appropriate dimensions and $\mathbf{\Sigma}$ and $\mathbf{P}$ are positive definite, it can be obtained that [38]

$$\int \mathrm{N}(\mathbf{x}; \mathbf{\Phi}\boldsymbol{\lambda} + \mathbf{d}, \mathbf{\Sigma})\mathrm{N}(\boldsymbol{\lambda}; \boldsymbol{\mu}, \mathbf{P})d\boldsymbol{\lambda} = \mathrm{N}(\mathbf{x}; \mathbf{\Phi}\boldsymbol{\mu} + \mathbf{d}, \mathbf{\Phi}\mathbf{P}\mathbf{\Phi}^T + \mathbf{\Sigma}). \quad (81)$$

B. Derivation of (7)

According to the Chapman-Kolmogorov equation, the one-step prediction PDF $p(\mathbf{x}_k|\mathbf{z}_{1:k-1})$ is formulated as

$$p(\mathbf{x}_k|\mathbf{z}_{1:k-1}) = \int p(\mathbf{x}_k|\mathbf{x}_{k-1})p(\mathbf{x}_{k-1}|\mathbf{z}_{1:k-1})d\mathbf{x}_{k-1}, \quad (82)$$

where $p(\mathbf{x}_{k-1}|\mathbf{z}_{1:k-1})$ denotes the posterior filtering PDF at time $k-1$ given by

$$p(\mathbf{x}_{k-1}|\mathbf{z}_{1:k-1}) = \mathrm{N}(\mathbf{x}_{k-1}; \hat{\mathbf{x}}_{k-1|k-1}, \mathbf{P}_{k-1|k-1}), \quad (83)$$

where $\hat{\mathbf{x}}_{k-1|k-1}$ and $\mathbf{P}_{k-1|k-1}$ are state estimate and corresponding estimate error covariance matrix at time $k-1$ respectively.

Using (3) and (5), the state transition PDF $p(\mathbf{x}_k|\mathbf{x}_{k-1})$ is written as

$$p(\mathbf{x}_k|\mathbf{x}_{k-1}) = \int_0^{+\infty} \mathrm{N}(\mathbf{x}_k; \mathbf{F}_k\mathbf{x}_{k-1} + \xi_k\boldsymbol{\beta}_1, \mathbf{Q}_{k-1}/\kappa_1(\xi_k)) \pi_1(\xi_k)d\xi_k. \quad (84)$$

Substituting (83)-(84) in (82) yields

$$p(\mathbf{x}_k|\mathbf{z}_{1:k-1}) = \int_0^{+\infty} g(\mathbf{x}_k, \xi_k)\pi_1(\xi_k)d\xi_k, \quad (85)$$

where $g(\mathbf{x}_k, \xi_k)$ is given by

$$g(\mathbf{x}_k, \xi_k) = \int \mathrm{N}\left(\mathbf{x}_k; \mathbf{F}_k\mathbf{x}_{k-1} + \xi_k\boldsymbol{\beta}_1, \mathbf{Q}/\kappa_1(\xi_k)\right) \times \mathrm{N}(\mathbf{x}_{k-1}; \hat{\mathbf{x}}_{k-1|k-1}, \mathbf{P}_{k-1|k-1})d\mathbf{x}_{k-1}. \quad (86)$$

According to the Gaussian integral formula in Appendix A, $g(\mathbf{x}_k, \xi_k)$ is calculated as

$$g(\mathbf{x}_k, \xi_k) = \mathrm{N}\left(\mathbf{x}_k; \mathbf{F}_k\hat{\mathbf{x}}_{k-1|k-1} + \xi_k\boldsymbol{\beta}_1, \mathbf{F}_k\mathbf{P}_{k-1|k-1}\mathbf{F}_k^{\mathrm{T}} + \mathbf{Q}_{k-1}/\kappa_1(\xi_k)\right). \quad (87)$$

Utilizing (87) in (85), we can obtain (7).

C. Derivations of (20)-(26)

Using $\phi = \mathbf{x}_k$ and (19) in (18), we obtain

$$\log q^{(i+1)}(\mathbf{x}_k) = -\frac{1}{2}\mathrm{E}^{(i)}[\kappa_1(\xi_k)]\mathrm{tr}\{(\mathbf{x}_k - \mathbf{F}_k\hat{\mathbf{x}}_{k-1|k-1} - \mathrm{E}^{(i)}[\xi_k]\mathrm{E}^{(i)}[\boldsymbol{\beta}_1])(\mathbf{x}_k - \mathbf{F}_k\hat{\mathbf{x}}_{k-1|k-1} - \mathrm{E}^{(i)}[\xi_k]\mathrm{E}^{(i)}[\boldsymbol{\beta}_1])^{\mathrm{T}} \times \mathrm{E}^{(i)}[\mathbf{\Sigma}_k^{-1}]\} - \frac{1}{2}\mathrm{E}^{(i)}[\kappa_2(\lambda_k)]\mathrm{tr}\{(\mathbf{z}_k - \mathbf{H}_k\mathbf{x}_k - \mathrm{E}^{(i)}[\lambda_k] \times \mathrm{E}^{(i)}[\boldsymbol{\beta}_2])(\mathbf{z}_k - \mathbf{H}_k\mathbf{x}_k - \mathrm{E}^{(i)}[\lambda_k]\mathrm{E}^{(i)}[\boldsymbol{\beta}_2])^{\mathrm{T}}\mathrm{E}^{(i)}[\mathbf{R}_k^{-1}]\} + c_{\mathbf{x}_k}. \quad (88)$$

Substituting (25)-(26) in (88) yields

$$\log q^{(i+1)}(\mathbf{x}_k) = -\frac{1}{2}(\mathbf{x}_k - \mathbf{F}_k\hat{\mathbf{x}}_{k-1|k-1} - \tilde{\mathbf{q}}_{k-1}^{(i)})^{\mathrm{T}} \times \left(\tilde{\mathbf{P}}_{k|k-1}^{(i)}\right)^{-1}(\mathbf{x}_k - \mathbf{F}_k\hat{\mathbf{x}}_{k-1|k-1} - \tilde{\mathbf{q}}_{k-1}^{(i)}) - \frac{1}{2}(\mathbf{z}_k - \mathbf{H}_k\mathbf{x}_k - \tilde{\mathbf{r}}_k^{(i)})^{\mathrm{T}}\left(\tilde{\mathbf{R}}_k^{(i)}\right)^{-1}(\mathbf{z}_k - \mathbf{H}_k\mathbf{x}_k - \tilde{\mathbf{r}}_k^{(i)}) + c_{\mathbf{x}_k}. \quad (89)$$

Define the modified one-step prediction PDF $\tilde{p}(\mathbf{x}_k|\mathbf{z}_{1:k-1})$ and the modified likelihood PDF $\tilde{p}(\mathbf{z}_k|\mathbf{x}_k)$ as follows

$$\tilde{p}(\mathbf{x}_k|\mathbf{z}_{1:k-1}) = \mathrm{N}(\mathbf{x}_k; \mathbf{F}_k\hat{\mathbf{x}}_{k-1|k-1} + \tilde{\mathbf{q}}_{k-1}^{(i)}, \tilde{\mathbf{P}}_{k|k-1}^{(i)}), \quad (90)$$

$$\tilde{p}(\mathbf{z}_k|\mathbf{x}_k) = \mathrm{N}(\mathbf{z}_k; \mathbf{H}_k\mathbf{x}_k + \tilde{\mathbf{r}}_k^{(i)}; \tilde{\mathbf{R}}_k^{(i)}). \tag{91}$$

Employing (90)-(91) in (89) gives

$$q^{(i+1)}(\mathbf{x}_k) \propto \tilde{p}(\mathbf{x}_k|\mathbf{z}_{1:k-1})\tilde{p}(\mathbf{z}_k|\mathbf{x}_k). \tag{92}$$

According to (90)-(92) and using Bayes' rule [39], we can obtain (20)-(24), where (21)-(24) are given by the measurement update of the Kalman filter.

D. Derivations of (36)-(44)

Substituting $\boldsymbol{\phi} = \boldsymbol{\beta}_1$, $\boldsymbol{\phi} = \boldsymbol{\beta}_2$ and (19) in (18), $\log q^{(i+1)}(\boldsymbol{\beta}_1)$ and $\log q^{(i+1)}(\boldsymbol{\beta}_2)$ can be calculated as

$$\begin{aligned}\log q^{(i+1)}(\boldsymbol{\beta}_1) = &-\frac{1}{2}(\boldsymbol{\beta}_1 - \bar{\boldsymbol{\beta}}_1)^{\mathrm{T}}(\sigma_1\mathbf{I}_n)^{-1}(\boldsymbol{\beta}_1 - \bar{\boldsymbol{\beta}}_1) - \\ &\frac{1}{2}\left[\mathbf{a}_k^{(i+1)} - \mathrm{E}^{(i+1)}[\xi_k]\boldsymbol{\beta}_1\right]^{\mathrm{T}}\mathrm{E}^{(i+1)}[\kappa_1(\xi_k)]\mathrm{E}^{(i)}[\boldsymbol{\Sigma}_k^{-1}] \times \\ &\left[\mathbf{a}_k^{(i+1)} - \mathrm{E}^{(i+1)}[\xi_k]\boldsymbol{\beta}_1\right] + c_{\boldsymbol{\beta}_1},\end{aligned} \tag{93}$$

$$\begin{aligned}\log q^{(i+1)}(\boldsymbol{\beta}_2) = &-\frac{1}{2}(\boldsymbol{\beta}_2 - \bar{\boldsymbol{\beta}}_2)^{\mathrm{T}}(\sigma_2\mathbf{I}_m)^{-1}(\boldsymbol{\beta}_2 - \bar{\boldsymbol{\beta}}_2) - \\ &\frac{1}{2}\left[\mathbf{b}_k^{(i+1)} - \mathrm{E}^{(i+1)}[\lambda_k]\boldsymbol{\beta}_2\right]^{\mathrm{T}}\mathrm{E}^{(i+1)}[\kappa_2(\lambda_k)]\mathrm{E}^{(i)}[\mathbf{R}_k^{-1}] \times \\ &\left[\mathbf{b}_k^{(i+1)} - \mathrm{E}^{(i+1)}[\lambda_k]\boldsymbol{\beta}_2\right] + c_{\boldsymbol{\beta}_2}.\end{aligned} \tag{94}$$

Employing (44), (93)-(94) can be reformulated as

$$\begin{aligned}\log q^{(i+1)}(\boldsymbol{\beta}_1) = &-\frac{1}{2}(\boldsymbol{\beta}_1 - \bar{\boldsymbol{\beta}}_1)^{\mathrm{T}}(\sigma_1\mathbf{I}_n)^{-1}(\boldsymbol{\beta}_1 - \bar{\boldsymbol{\beta}}_1) - \\ &\frac{1}{2}\left[\mathbf{a}_k^{(i+1)} - \mathrm{E}^{(i+1)}[\xi_k]\boldsymbol{\beta}_1\right]^{\mathrm{T}}\left(\bar{\mathbf{P}}_{k|k-1}^{(i+1)}\right)^{-1} \times \\ &\left[\mathbf{a}_k^{(i+1)} - \mathrm{E}^{(i+1)}[\xi_k]\boldsymbol{\beta}_1\right] + c_{\boldsymbol{\beta}_1},\end{aligned} \tag{95}$$

$$\begin{aligned}\log q^{(i+1)}(\boldsymbol{\beta}_2) = &-\frac{1}{2}(\boldsymbol{\beta}_2 - \bar{\boldsymbol{\beta}}_2)^{\mathrm{T}}(\sigma_2\mathbf{I}_m)^{-1}(\boldsymbol{\beta}_2 - \bar{\boldsymbol{\beta}}_2) - \\ &\frac{1}{2}\left[\mathbf{b}_k^{(i+1)} - \mathrm{E}^{(i+1)}[\lambda_k]\boldsymbol{\beta}_2\right]^{\mathrm{T}}\left(\bar{\mathbf{R}}_k^{(i+1)}\right)^{-1} \times \\ &\left[\mathbf{b}_k^{(i+1)} - \mathrm{E}^{(i+1)}[\lambda_k]\boldsymbol{\beta}_2\right] + c_{\boldsymbol{\beta}_2}.\end{aligned} \tag{96}$$

Define the modified likelihood PDFs of shape parameters $\tilde{p}(\mathbf{z}_k|\boldsymbol{\beta}_1)$ and $\tilde{p}(\mathbf{z}_k|\boldsymbol{\beta}_2)$ as follows

$$\tilde{p}(\mathbf{a}_k^{(i+1)}|\boldsymbol{\beta}_1) = \mathrm{N}(\mathbf{a}_k^{(i+1)}; \mathrm{E}^{(i+1)}[\xi_k]\boldsymbol{\beta}_1, \bar{\mathbf{P}}_{k|k-1}^{(i+1)}), \tag{97}$$

$$\tilde{p}(\mathbf{b}_k^{(i+1)}|\boldsymbol{\beta}_2) = \mathrm{N}(\mathbf{b}_k^{(i+1)}; \mathrm{E}^{(i+1)}[\lambda_k]\boldsymbol{\beta}_2, \bar{\mathbf{R}}_k^{(i+1)}). \tag{98}$$

Substituting (14) and (97)-(98) in (95)-(96) gives

$$q^{(i+1)}(\boldsymbol{\beta}_1) \propto p(\boldsymbol{\beta}_1)\tilde{p}(\mathbf{a}_k^{(i+1)}|\boldsymbol{\beta}_1), \tag{99}$$

$$q^{(i+1)}(\boldsymbol{\beta}_2) \propto p(\boldsymbol{\beta}_2)\tilde{p}(\mathbf{b}_k^{(i+1)}|\boldsymbol{\beta}_2). \tag{100}$$

According to (14) and (97)-(100) and using Bayes' rule [39], we can obtain (36)-(43), where (38)-(43) are given by the measurement update of the Kalman filter.

E. Derivations of (45)-(50)

Substituting $\boldsymbol{\phi} = \boldsymbol{\Sigma}_k$, $\boldsymbol{\phi} = \mathbf{R}_k$ and (19) in (18), $\log q^{(i+1)}(\boldsymbol{\Sigma}_k)$ and $\log q^{(i+1)}(\mathbf{R}_k)$ can be formulated as

$$\begin{aligned}\log q^{(i+1)}(\boldsymbol{\Sigma}_k) = &-\frac{1}{2}(t_k + n + 2)\log|\boldsymbol{\Sigma}_k| - \frac{1}{2}\mathrm{tr}\left\{\mathbf{T}_k\boldsymbol{\Sigma}_k^{-1}\right\} \\ &-\frac{1}{2}\mathrm{tr}\Big\{\mathrm{E}^{(i+1)}[\kappa_1(\xi_k)(\mathbf{x}_k - \mathbf{F}_k\hat{\mathbf{x}}_{k-1|k-1} - \xi_k\boldsymbol{\beta}_1)\times \\ &(\mathbf{x}_k - \mathbf{F}_k\hat{\mathbf{x}}_{k-1|k-1} - \xi_k\boldsymbol{\beta}_1)^{\mathrm{T}}]\boldsymbol{\Sigma}_k^{-1}\Big\} + c_{\boldsymbol{\Sigma}_k},\end{aligned} \tag{101}$$

$$\begin{aligned}\log q^{(i+1)}(\mathbf{R}_k) = &-\frac{1}{2}(u_k + m + 2)\log|\mathbf{R}_k| - \frac{1}{2}\mathrm{tr}\{\mathbf{U}_k\mathbf{R}_k^{-1} \\ &\} - \frac{1}{2}\mathrm{tr}\Big\{\mathrm{E}^{(i+1)}[\kappa_2(\lambda_k)(\mathbf{z}_k - \mathbf{H}_k\mathbf{x}_k - \lambda_k\boldsymbol{\beta}_2)(\mathbf{z}_k - \mathbf{H}_k\mathbf{x}_k - \\ &\lambda_k\boldsymbol{\beta}_2)^{\mathrm{T}}]\mathbf{R}_k^{-1}\Big\} + c_{\mathbf{R}_k}.\end{aligned} \tag{102}$$

Utilizing (49)-(50) in (101)-(102) gives

$$\begin{aligned}\log q^{(i+1)}(\boldsymbol{\Sigma}_k) = &-\frac{1}{2}(t_k + n + 2)\log|\boldsymbol{\Sigma}_k| - \\ &\frac{1}{2}\mathrm{tr}\left\{[\mathbf{T}_k + \mathbf{C}_k^{(i+1)}]\boldsymbol{\Sigma}_k^{-1}\right\} + c_{\boldsymbol{\Sigma}_k},\end{aligned} \tag{103}$$

$$\begin{aligned}\log q^{(i+1)}(\mathbf{R}_k) = &-\frac{1}{2}(u_k + m + 2)\log|\mathbf{R}_k| - \\ &\frac{1}{2}\mathrm{tr}\left\{[\mathbf{U}_k + \mathbf{D}_k^{(i+1)}]\mathbf{R}_k^{-1}\right\} + c_{\mathbf{R}_k}.\end{aligned} \tag{104}$$

According to (103)-(104), we can obtain (45)-(48).

F. Derivations of (64)-(69)

Using (20) and (29)-(30), $\mathbf{A}_k^{i+1}$ and $\mathbf{a}_k^{i+1}$ are calculated as

$$\begin{aligned}\mathbf{A}_k^{i+1} = &\mathrm{E}^{(i+1)}\left[(\mathbf{x}_k - \hat{\mathbf{x}}_{k|k}^{(i+1)} + \hat{\mathbf{x}}_{k|k}^{(i+1)} - \mathbf{F}_k\hat{\mathbf{x}}_{k-1|k-1})\times\right. \\ &\left.(\mathbf{x}_k - \hat{\mathbf{x}}_{k|k}^{(i+1)} + \hat{\mathbf{x}}_{k|k}^{(i+1)} - \mathbf{F}_k\hat{\mathbf{x}}_{k-1|k-1})^{\mathrm{T}}\right] \\ = &\mathrm{E}^{(i+1)}\left[(\mathbf{x}_k - \hat{\mathbf{x}}_{k|k}^{(i+1)})(\mathbf{x}_k - \hat{\mathbf{x}}_{k|k}^{(i+1)})^{\mathrm{T}}\right] + \\ &\left(\hat{\mathbf{x}}_{k|k}^{(i+1)} - \mathbf{F}_k\hat{\mathbf{x}}_{k-1|k-1}\right)\left(\hat{\mathbf{x}}_{k|k}^{(i+1)} - \mathbf{F}_k\hat{\mathbf{x}}_{k-1|k-1}\right)^{\mathrm{T}} \\ = &\mathbf{P}_{k|k}^{(i+1)} + \left(\hat{\mathbf{x}}_{k|k}^{(i+1)} - \mathbf{F}_k\hat{\mathbf{x}}_{k-1|k-1}\right)\times \\ &\left(\hat{\mathbf{x}}_{k|k}^{(i+1)} - \mathbf{F}_k\hat{\mathbf{x}}_{k-1|k-1}\right)^{\mathrm{T}},\end{aligned} \tag{105}$$

$$\begin{aligned}\mathbf{a}_k^{i+1} = &\mathrm{E}^{(i+1)}\left[\mathbf{x}_k - \mathbf{F}_k\hat{\mathbf{x}}_{k-1|k-1}\right] = \mathrm{E}^{(i+1)}\left[\mathbf{x}_k\right] - \\ &\mathbf{F}_k\hat{\mathbf{x}}_{k-1|k-1} = \hat{\mathbf{x}}_{k|k}^{(i+1)} - \mathbf{F}_k\hat{\mathbf{x}}_{k-1|k-1}.\end{aligned} \tag{106}$$

Exploiting (20) and (31)-(32), $\mathbf{B}_k^{i+1}$ and $\mathbf{b}_k^{i+1}$ can be calculated as

$$\begin{aligned}\mathbf{B}_k^{i+1} = &\mathrm{E}^{(i+1)}\Big\{[\mathbf{z}_k - \mathbf{H}_k\hat{\mathbf{x}}_{k|k}^{(i+1)} + \mathbf{H}_k(\hat{\mathbf{x}}_{k|k}^{(i+1)} - \mathbf{x}_k)]\times \\ &[\mathbf{z}_k - \mathbf{H}_k\hat{\mathbf{x}}_{k|k}^{(i+1)} + \mathbf{H}_k(\hat{\mathbf{x}}_{k|k}^{(i+1)} - \mathbf{x}_k)]^{\mathrm{T}}\Big\} \\ = &(\mathbf{z}_k - \mathbf{H}_k\hat{\mathbf{x}}_{k|k}^{(i+1)})(\mathbf{z}_k - \mathbf{H}_k\hat{\mathbf{x}}_{k|k}^{(i+1)})^{\mathrm{T}} + \\ &\mathbf{H}_k\mathrm{E}^{(i+1)}\left[(\mathbf{x}_k - \hat{\mathbf{x}}_{k|k}^{(i+1)})(\mathbf{x}_k - \hat{\mathbf{x}}_{k|k}^{(i+1)})^{\mathrm{T}}\right]\mathbf{H}_k^{\mathrm{T}} \\ = &\left(\mathbf{z}_k - \mathbf{H}_k\hat{\mathbf{x}}_{k|k}^{(i+1)}\right)\left(\mathbf{z}_k - \mathbf{H}_k\hat{\mathbf{x}}_{k|k}^{(i+1)}\right)^{\mathrm{T}} + \\ &\mathbf{H}_k\mathbf{P}_{k|k}^{(i+1)}\mathbf{H}_k^{\mathrm{T}},\end{aligned} \tag{107}$$

$$\mathbf{b}_k^{i+1}=\mathrm{E}^{(i+1)}\left[\mathbf{z}_k-\mathbf{H}_k\mathbf{x}_k\right]=\mathbf{z}_k-\mathbf{H}_k\mathrm{E}^{(i+1)}\left[\mathbf{x}_k\right]$$
$$=\mathbf{z}_k-\mathbf{H}_k\hat{\mathbf{x}}_{k|k}^{(i+1)}. \tag{108}$$

Substituting (33) in (49)-(50) gives

$$\mathbf{C}_k^{(i+1)}=\kappa_1(\xi_k^{(i+1)})\mathrm{E}^{(i+1)}[(\mathbf{x}_k-\mathbf{F}_k\hat{\mathbf{x}}_{k-1|k-1}-\xi_k^{(i+1)}\boldsymbol{\beta}_1)$$
$$(\mathbf{x}_k-\mathbf{F}_k\hat{\mathbf{x}}_{k-1|k-1}-\xi_k^{(i+1)}\boldsymbol{\beta}_1)^{\mathrm{T}}], \tag{109}$$

$$\mathbf{D}_k^{(i+1)}=\kappa_2(\lambda_k^{(i+1)})\mathrm{E}^{(i+1)}[(\mathbf{z}_k-\mathbf{H}_k\mathbf{x}_k-\lambda_k^{(i+1)}\boldsymbol{\beta}_2)\times$$
$$(\mathbf{z}_k-\mathbf{H}_k\mathbf{x}_k-\lambda_k^{(i+1)}\boldsymbol{\beta}_2)^{\mathrm{T}}]. \tag{110}$$

Employing (20), (36)-(37), (65) and (67) in (109)-(110), $\mathbf{C}_k^{i+1}$ and $\mathbf{D}_k^{i+1}$ are recalculated as

$$\mathbf{C}_k^{(i+1)}=\kappa_1(\xi_k^{(i+1)})\mathrm{E}^{(i+1)}\Big[(\mathbf{x}_k-\hat{\mathbf{x}}_{k|k}^{(i+1)}+\hat{\mathbf{x}}_{k|k}^{(i+1)}-$$
$$\mathbf{F}_k\hat{\mathbf{x}}_{k-1|k-1}-\xi_k^{(i+1)}\boldsymbol{\beta}_1^{(i+1)}+\xi_k^{(i+1)}\boldsymbol{\beta}_1^{(i+1)}-\xi_k^{(i+1)}\boldsymbol{\beta}_1)\times$$
$$(\mathbf{x}_k-\hat{\mathbf{x}}_{k|k}^{(i+1)}+\hat{\mathbf{x}}_{k|k}^{(i+1)}-\mathbf{F}_k\hat{\mathbf{x}}_{k-1|k-1}-\xi_k^{(i+1)}\boldsymbol{\beta}_1^{(i+1)}+$$
$$\xi_k^{(i+1)}\boldsymbol{\beta}_1^{(i+1)}-\xi_k^{(i+1)}\boldsymbol{\beta}_1)^{\mathrm{T}}\Big]=\kappa_1(\xi_k^{(i+1)})\Big[\mathbf{P}_{k|k}^{(i+1)}+$$
$$\left(\mathbf{a}_k^{i+1}-\xi_k^{(i+1)}\boldsymbol{\beta}_1^{(i+1)}\right)\left(\mathbf{a}_k^{i+1}-\xi_k^{(i+1)}\boldsymbol{\beta}_1^{(i+1)}\right)^{\mathrm{T}}+$$
$$\left(\xi_k^{(i+1)}\right)^2\mathbf{P}_{\boldsymbol{\beta}_1}^{(i+1)}\Big], \tag{111}$$

$$\mathbf{D}_k^{(i+1)}=\kappa_2(\lambda_k^{(i+1)})\mathrm{E}^{(i+1)}\Big[(\mathbf{z}_k-\mathbf{H}_k\hat{\mathbf{x}}_{k|k}^{(i+1)}-\lambda_k^{(i+1)}\times$$
$$\boldsymbol{\beta}_2^{(i+1)}+\mathbf{H}_k\hat{\mathbf{x}}_{k|k}^{(i+1)}-\mathbf{H}_k\mathbf{x}_k+\lambda_k^{(i+1)}\boldsymbol{\beta}_2^{(i+1)}-\lambda_k^{(i+1)}\boldsymbol{\beta}_2)\times$$
$$(\mathbf{z}_k-\mathbf{H}_k\hat{\mathbf{x}}_{k|k}^{(i+1)}-\lambda_k^{(i+1)}\boldsymbol{\beta}_2^{(i+1)}+\mathbf{H}_k\hat{\mathbf{x}}_{k|k}^{(i+1)}-\mathbf{H}_k\mathbf{x}_k+$$
$$\lambda_k^{(i+1)}\boldsymbol{\beta}_2^{(i+1)}-\lambda_k^{(i+1)}\boldsymbol{\beta}_2)^{\mathrm{T}}\Big]=\kappa_2(\lambda_k^{(i+1)})\Big[\mathbf{H}_k\mathbf{P}_{k|k}^{(i+1)}\mathbf{H}_k^{\mathrm{T}}+$$
$$\left(\mathbf{b}_k^{i+1}-\lambda_k^{(i+1)}\boldsymbol{\beta}_2^{(i+1)}\right)\left(\mathbf{b}_k^{i+1}-\lambda_k^{(i+1)}\boldsymbol{\beta}_2^{(i+1)}\right)^{\mathrm{T}}+$$
$$\left(\lambda_k^{(i+1)}\right)^2\mathbf{P}_{\boldsymbol{\beta}_2}^{(i+1)}\Big]. \tag{112}$$

According to (105)-(108) and (111)-(112), we can obtain (64)-(69).

G. Special Cases

Exploiting (27)-(28), $\log q^{(i+1)}(\xi_k)$ and $\log q^{(i+1)}(\lambda_k)$ can be written as the unified form as follows

$$J(y)=-\frac{\eta_1}{2}\kappa(y)+\eta_2 y\kappa(y)-\frac{\eta_3}{2}y^2\kappa(y)+\frac{s}{2}\log\kappa(y)+$$
$$\log\pi(y), \tag{113}$$

where $J(y)$ turns into $\log q^{(i+1)}(\xi_k)$ and $\log q^{(i+1)}(\lambda_k)$ if the following equations hold respectively

$$\eta_1=\mathrm{tr}\left\{\mathbf{A}_k^{i+1}\mathrm{E}^{(i)}[\boldsymbol{\Sigma}_k^{-1}]\right\},\ \ \eta_2=\left\{\mathrm{E}^{(i)}[\boldsymbol{\beta}_1]\right\}^{\mathrm{T}}\times$$
$$\mathrm{E}^{(i)}[\boldsymbol{\Sigma}_k^{-1}]\mathbf{a}_k^{i+1},\ \ \eta_3=\mathrm{tr}\left\{\mathrm{E}^{(i)}[\boldsymbol{\beta}_1\boldsymbol{\beta}_1^{\mathrm{T}}]\mathrm{E}^{(i)}[\boldsymbol{\Sigma}_k^{-1}]\right\},$$
$$y=\xi_k,\ \ s=n,\ \ \kappa(\cdot)=\kappa_1(\cdot),\ \ \pi(\cdot)=\pi_1(\cdot), \tag{114}$$

$$\eta_1=\mathrm{tr}\left\{\mathbf{B}_k^{i+1}\mathrm{E}^{(i)}[\mathbf{R}_k^{-1}]\right\},\ \ \eta_2=\left\{\mathrm{E}^{(i)}[\boldsymbol{\beta}_2]\right\}^{\mathrm{T}}\times$$
$$\mathrm{E}^{(i)}[\mathbf{R}_k^{-1}]\mathbf{b}_k^{i+1},\ \ \eta_3=\mathrm{tr}\left\{\mathrm{E}^{(i)}[\boldsymbol{\beta}_2\boldsymbol{\beta}_2^{\mathrm{T}}]\mathrm{E}^{(i)}[\mathbf{R}_k^{-1}]\right\},$$
$$y=\lambda_k,\ \ s=m,\ \ \kappa(\cdot)=\kappa_2(\cdot),\ \ \pi(\cdot)=\pi_2(\cdot). \tag{115}$$

According to (15)-(16), (48), (60)-(64), (66), (68)-(69) and (114)-(115) and using $\mathbf{P}_{k-1|k-1}>\mathbf{0}$ and $\bar{\mathbf{R}}_k>\mathbf{0}$, we can obtain

$$\eta_1>0,\qquad\qquad \eta_3\geq 0, \tag{116}$$

where $\eta_3=0$ if $\bar{\boldsymbol{\beta}}_1=\mathbf{0}$ and $\sigma_1=0$ or $\bar{\boldsymbol{\beta}}_2=\mathbf{0}$ and $\sigma_2=0$, and the proof of (116) is given in Appendix H. Next, several particular solutions will be derived when certain GSM distributions are employed.

1) Pearson type-VII distribution: We can see from Table I that $\boldsymbol{\beta}=\mathbf{0}$, $\kappa(y)=y$, $\pi(y)=\mathrm{G}(y;\frac{\nu}{2},\frac{\delta}{2})$ and $y>0$ when the Pearson type-VII distribution is utilized. Since $\boldsymbol{\beta}=\mathbf{0}$, the prior parameters of the shape parameters satisfy

$$\bar{\boldsymbol{\beta}}_1=\mathbf{0},\qquad \bar{\boldsymbol{\beta}}_2=\mathbf{0},\qquad \sigma_1=\sigma_2=0, \tag{117}$$

Substituting (117) in (38)-(43), we have

$$\boldsymbol{\beta}_1^{(i+1)}=\mathbf{0},\qquad\qquad \mathbf{P}_{\boldsymbol{\beta}_1}^{(i+1)}=\mathbf{0}, \tag{118}$$

$$\boldsymbol{\beta}_2^{(i+1)}=\mathbf{0},\qquad\qquad \mathbf{P}_{\boldsymbol{\beta}_2}^{(i+1)}=\mathbf{0}. \tag{119}$$

Utilizing (118)-(119) in (114)-(115) gives

$$\eta_2=\eta_3=0. \tag{120}$$

Employing (120), $\kappa(y)=y$ and $\pi(y)=\mathrm{G}(y;\frac{\nu}{2},\frac{\delta}{2})$ in (113) yields

$$J(y)=-\frac{\eta_1+\delta}{2}y+\frac{s+\nu-2}{2}\log y+c_y. \tag{121}$$

Using (121), the maximum point y_* satisfies the following equations

$$-(\eta_1+\delta)y_*+s+\nu-2=0, \tag{122}$$

$$\text{s.t.}\ \ y_*>0,\qquad s+\nu-2>0, \tag{123}$$

where s.t. represents to "subject to".

Solving equations (122)-(123) and using (116), we obtain

$$y_*=\frac{s+\nu-2}{\eta_1+\delta},\qquad \text{s.t.}\quad s+\nu>2. \tag{124}$$

Substituting (114)-(115) in (124), we can obtain MAP estimates $\xi_k^{(i+1)}$ and $\lambda_k^{(i+1)}$ respectively for the Pearson type-VII distribution.

2) Slash distribution: It is seen from Table I that $\boldsymbol{\beta}=\mathbf{0}$, $\kappa(y)=y$, $\pi(y)=\mathrm{Be}(y;\nu,1)$ and $0<y<1$ for the Slash distribution. Since $\boldsymbol{\beta}=\mathbf{0}$, we can obtain (117)-(120). Substituting (120), $\kappa(y)=y$ and $\pi(y)=\mathrm{Be}(y;\nu,1)$ in (113) yields

$$J(y)=-\frac{\eta_1}{2}y+\frac{s+2\nu-2}{2}\log y+c_y. \tag{125}$$

According to (125), the maximum point y_* satisfies the following equations

$$-\eta_1 y_*+s+2\nu-2=0, \tag{126}$$

$$\text{s.t.}\ \ 0<y_*<1,\quad s+2\nu-2>0. \tag{127}$$

Solving equations (126)-(127) and using (116), we have

$$y_*=\frac{s+2\nu-2}{\eta_1},\qquad \text{s.t.}\quad 2<s+2\nu<\eta_1+2. \tag{128}$$

Exploiting (114)-(115) in (128), we can obtain MAP estimates $\xi_k^{(i+1)}$ and $\lambda_k^{(i+1)}$ respectively for the Slash distribution.

3) Variance gamma distribution: It can be seen from Table I that, for the variance gamma distribution, $\boldsymbol{\beta} = \mathbf{0}$, $\kappa(y) = y$, $\pi(y) = \mathrm{IG}(y; \frac{\nu}{2}, \frac{\nu}{2})$ and $y > 0$. Since $\boldsymbol{\beta} = \mathbf{0}$, we can obtain (117)-(120). Substituting (120), $\kappa(y) = y$ and $\pi(y) = \mathrm{IG}(y; \frac{\nu}{2}, \frac{\nu}{2})$ in (113) yields

$$J(y) = -\frac{\eta_1}{2}y + \frac{s-\nu-2}{2}\log y - \frac{\nu}{2y} + c_y. \tag{129}$$

Employing (129), the maximum point y_* satisfies the following equations

$$\eta_1 y_*^2 - (s-\nu-2)y_* - \nu = 0, \tag{130}$$

$$\text{s.t. } y_* > 0, \quad s - \nu - 2 + \frac{2\nu}{y_*} > 0. \tag{131}$$

Solving equations (130)-(131) and using (116), we obtain

$$y_* = \frac{s-\nu-2+\sqrt{(s-\nu-2)^2+4\nu\eta_1}}{2\eta_1}, \tag{132}$$

Substituting (114)-(115) in (132), we can obtain MAP estimates $\xi_k^{(i+1)}$ and $\lambda_k^{(i+1)}$ respectively for the variance gamma distribution.

4) GH skew Student's t distribution: It is seen from Table I that $\boldsymbol{\beta} \neq \mathbf{0}$, $\kappa(y) = 1/y$, $\pi(y) = \mathrm{IG}(y; \frac{\nu}{2}, \frac{\nu}{2})$ and $y > 0$ when the GH skew Student's t distribution is used. Using $\boldsymbol{\beta} \neq \mathbf{0}$ and (116), we have

$$\eta_1 > 0, \qquad \eta_3 > 0. \tag{133}$$

Substituting $\kappa(y) = 1/y$ and $\pi(y) = \mathrm{IG}(y; \frac{\nu}{2}, \frac{\nu}{2})$ in (113) yields

$$J(y) = -\frac{\eta_1+\nu}{2y} - \frac{\eta_3}{2}y - \frac{s+\nu+2}{2}\log y + c_y. \tag{134}$$

Exploiting (134), the maximum point y_* satisfies the following equations

$$\eta_3 y_*^2 + (s+\nu+2)y_* - (\eta_1+\nu) = 0, \tag{135}$$

$$\text{s.t. } y_* > 0, \quad s + \nu + 2 - \frac{2(\eta_1+\nu)}{y_*} < 0. \tag{136}$$

Solving equations (135)-(136) and using (133), we obtain

$$y_* = \frac{-(s+\nu+2)+\sqrt{(s+\nu+2)^2+4\eta_3(\eta_1+\nu)}}{2\eta_3}. \tag{137}$$

Substituting (114)-(115) in (137), we can obtain MAP estimates $\xi_k^{(i+1)}$ and $\lambda_k^{(i+1)}$ respectively for the GH skew Student's t distribution.

5) GH variance gamma distribution: We can see from Table I that $\boldsymbol{\beta} \neq \mathbf{0}$, $\kappa(y) = 1/y$, $\pi(y) = \mathrm{G}(y; \frac{\nu}{2}, \frac{\nu}{2})$ and $y > 0$ for the GH variance gamma distribution. Since $\boldsymbol{\beta} \neq \mathbf{0}$, we can obtain (133). Substituting $\kappa(y) = 1/y$ and $\pi(y) = \mathrm{G}(y; \frac{\nu}{2}, \frac{\nu}{2})$ in (113) gives

$$J(y) = -\frac{\eta_1}{2y} - \frac{\eta_3+\nu}{2}y - \frac{s-\nu+2}{2}\log y + c_y. \tag{138}$$

Using (138), the maximum point y_* satisfies the following equations

$$(\eta_3+\nu)y_*^2 + (s-\nu+2)y_* - \eta_1 = 0, \tag{139}$$

$$\text{s.t. } y_* > 0, \quad s - \nu + 2 - \frac{2\eta_1}{y_*} < 0. \tag{140}$$

Solving equations (139)-(140) and using (133), we have

$$y_* = \frac{-(s-\nu+2)+\sqrt{(s-\nu+2)^2+4\eta_1(\eta_3+\nu)}}{2(\eta_3+\nu)}. \tag{141}$$

Exploiting (114)-(115) in (141), we can obtain MAP estimates $\xi_k^{(i+1)}$ and $\lambda_k^{(i+1)}$ respectively for the GH variance gamma distribution.

H. Proof of (116)

Considering that the estimate error covariance matrix $\mathbf{P}_{k-1|k-1}$ and the nominal measurement noise covariance matrix $\bar{\mathbf{R}}_k$ are positive-definite and utilizing (15)-(16), we obtain

$$\mathbf{T}_k > \mathbf{0}, \qquad \mathbf{U}_k > \mathbf{0}. \tag{142}$$

According to (68)-(69) and using $\kappa_1(\xi_k^{(i+1)}) > 0$ and $\kappa_2(\lambda_k^{(i+1)}) > 0$ yields

$$\mathbf{C}_k^{(i+1)} > \mathbf{0}, \qquad \mathbf{D}_k^{(i+1)} > \mathbf{0}. \tag{143}$$

Substituting (142)-(143) in (48) gives

$$\mathbf{T}_k^{(i+1)} > \mathbf{0}, \qquad \mathbf{U}_k^{(i+1)} > \mathbf{0}. \tag{144}$$

Exploiting (144) in (62)-(63), we have

$$\mathrm{E}^{(i+1)}[\boldsymbol{\Sigma}_k^{-1}] > \mathbf{0}, \qquad \mathrm{E}^{(i+1)}[\mathbf{R}_k^{-1}] > \mathbf{0}. \tag{145}$$

Using (64) and (66) results in

$$\mathbf{A}_k^{i+1} \geq \mathbf{0}, \qquad \mathbf{B}_k^{i+1} \geq \mathbf{0}. \tag{146}$$

Employing (145), $\mathrm{E}^{(i+1)}[\boldsymbol{\Sigma}_k^{-1}]$ and $\mathrm{E}^{(i+1)}[\mathbf{R}_k^{-1}]$ can be factored as

$$\mathrm{E}^{(i+1)}[\boldsymbol{\Sigma}_k^{-1}] = \mathbf{L}_1\mathbf{L}_1^{\mathrm{T}}, \qquad \mathrm{E}^{(i+1)}[\mathbf{R}_k^{-1}] = \mathbf{L}_2\mathbf{L}_2^{\mathrm{T}}, \tag{147}$$

where $\mathbf{L}_1$ and $\mathbf{L}_2$ are invertible lower triangular matrices.

Substituting (147) in (114)-(115) yields

$$\eta_1 = \mathrm{tr}\left\{\mathbf{L}_1^{\mathrm{T}}\mathbf{A}_k^{i+1}\mathbf{L}_1\right\}, \quad \eta_3 = \mathrm{tr}\left\{\mathbf{L}_1^{\mathrm{T}}\mathrm{E}^{(i)}[\boldsymbol{\beta}_1\boldsymbol{\beta}_1^{\mathrm{T}}]\mathbf{L}_1\right\}, \tag{148}$$

$$\eta_1 = \mathrm{tr}\left\{\mathbf{L}_2^{\mathrm{T}}\mathbf{B}_k^{i+1}\mathbf{L}_2\right\}, \quad \eta_3 = \mathrm{tr}\left\{\mathbf{L}_2^{\mathrm{T}}\mathrm{E}^{(i)}[\boldsymbol{\beta}_2\boldsymbol{\beta}_2^{\mathrm{T}}]\mathbf{L}_2\right\}. \tag{149}$$

Since $\mathbf{L}_1$ and $\mathbf{L}_2$ are invertible matrices and $\mathbf{A}_k^{i+1}$ and $\mathbf{B}_k^{i+1}$ are nonzero positive-semidefinite matrices, $\mathbf{L}_1^{\mathrm{T}}\mathbf{A}_k^{i+1}\mathbf{L}_1$ and $\mathbf{L}_2^{\mathrm{T}}\mathbf{B}_k^{i+1}\mathbf{L}_2$ are also nonzero positive-semidefinite matrices, i.e,

$$\mathbf{L}_1^{\mathrm{T}}\mathbf{A}_k^{i+1}\mathbf{L}_1 \geq \mathbf{0}, \qquad \mathbf{L}_1^{\mathrm{T}}\mathbf{A}_k^{i+1}\mathbf{L}_1 \neq \mathbf{0}, \tag{150}$$

$$\mathbf{L}_2^{\mathrm{T}}\mathbf{B}_k^{i+1}\mathbf{L}_2 \geq \mathbf{0}, \qquad \mathbf{L}_2^{\mathrm{T}}\mathbf{B}_k^{i+1}\mathbf{L}_2 \neq \mathbf{0}. \tag{151}$$

Utilizing (60)-(61), we have

$$\mathrm{E}^{(i)}[\boldsymbol{\beta}_1\boldsymbol{\beta}_1^{\mathrm{T}}] \geq \mathbf{0}, \qquad \mathrm{E}^{(i)}[\boldsymbol{\beta}_2\boldsymbol{\beta}_2^{\mathrm{T}}] \geq \mathbf{0}, \tag{152}$$

where $\mathrm{E}^{(i)}[\boldsymbol{\beta}_1\boldsymbol{\beta}_1^{\mathrm{T}}] = \mathbf{0}$ and $\mathrm{E}^{(i)}[\boldsymbol{\beta}_2\boldsymbol{\beta}_2^{\mathrm{T}}] = \mathbf{0}$ if

$$\bar{\boldsymbol{\beta}}_1 = \mathbf{0}, \qquad \bar{\boldsymbol{\beta}}_2 = \mathbf{0}, \qquad \sigma_1 = \sigma_2 = 0. \tag{153}$$

Substituting (150)-(153) in (148)-(149), we can obtain (116).

REFERENCES

[1] N. Sadeghzadeh-Nokhodberiz and J. Poshtan, "Distributed interacting multiple filters for fault diagnosis of navigation sensors in a robotic system," *IEEE Trans. Syst., Man, Cybern.: Syst.*, vol. 47, no. 7, pp. 1383–1393, July 2017.

[2] W. Liu, P. Shi and J. S. Pan, "State estimation for discrete-time Markov jump linear systems with time-correlated and mode-Dependent measurement noise," *Automatica*, vol. 85, no. 11, pp. 9–21, Nov. 2017.

[3] M. Gupta, S. Kumar, L. Behera, and V. K. Subramanian, "A novel vision-based tracking algorithm for a human-following mobile robot," *IEEE Trans. Syst., Man, Cybern.: Syst.*, vol. 47, no. 7, pp. 1415–1427, July 2017.

[4] S. Zhao, Y. S. Shmaliy, P. Shi and C. K. Ahn, "Fusion Kalman/UFIR filter for state estimation with uncertain parameters and noise statistics," *IEEE Trans. Ind. Electron.*, vol. 64, no. 4, pp. 3075–3083, Apr. 2017.

[5] L. Y. Wang, Z. Liu, C. L. P. Chen, Y. Zhang, S. Lee, and X. Chen, "A UKF-based predictable SVR learning controller for biped walking," *IEEE Trans. Syst., Man, Cybern.: Syst.*, vol. 43, no. 6, pp. 1440–1450, June 2013.

[6] M. Roth, E. Özkan, and F. Gustafsson, "A Student's t filter for heavy-tailed process and measurement noise," in *2013 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, May 2013, pp. 5770–5774.

[7] Y. L. Huang, Y. G. 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.

[8] 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. 2450–2454, Nov. 2015.

[9] H. Nurminen, T. Ardeshiri, R. Piché, and F. Gustafsson, "Skew-t filter and smoother with improved covariance matrix approximation," [Online]. Available: arXiv preprint arXiv:1608.07435.

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

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

[12] Z. R. Xing, Y. Q. Xia, L. P. Yan, K. F. Lu, and Q. H. Gong, "Multisensor distributed weighted Kalman filter fusion with network delays, stochastic uncertainties, autocorrelated, and cross-correlated noises," *IEEE Trans. Syst., Man, Cybern.: Syst.*, doi: 10.1109/TSMC.2016.2633283, 2016.

[13] W. Y. Li, G. L. Wei, D. R. Ding, Y. R. Liu, and F. E. Alsaadi, "A new look at boundedness of error covariance of Kalman filtering," *IEEE Trans. Syst., Man, Cybern.: Syst.*, doi: 10.1109/TSMC.2016.2598845, 2017.

[14] Q. B. Ge, T. Shao, Q. M. Yang, X. F. Shen, and C. L. Wen, "Multisensor nonlinear fusion methods based on adaptive ensemble fifth-degree iterated cubature information filter for biomechatronics," *IEEE Trans. Syst., Man, Cybern.: Syst.*, vol. 46, no. 7, pp. 912-925, July 2016.

[15] J. Loxam and T. Drummond, "Student mixture filter for robust, real-time visual tracking," in *Proceedings of 10th European Conference on Computer Vision: Part III*, 2008.

[16] J. Ting, E. Theodorou, and S. Schaal, "Learning an outlier-robust Kalman filter," in *Proceedings of 18th European Conference on Machine Learning Warsaw*, Poland, 2007.

[17] R. Piché, S. Särkkä, and J. Hartikainen, "Recursive outlier-robust filtering and smoothing for nonlinear systems using the multivariate Student-t distribution," in *Proceedings of MLSP*, Sep. 2012.

[18] 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.

[19] H. Zhu, H. Leung, and Z. He, "A variational Bayesian approach to robust sensor fusion based on Student-t distribution," *Inform. Sciences*, vol. 221, no. 2013, pp. 201–214, Sep. 2012.

[20] G. Agamennoni and E.M. Nebot, "Robust estimation in non-linear state-space models with state-dependent noise," *IEEE Trans. Signal Process.*, vol. 62, no. 8, pp. 2165–2175, Apr. 2014.

[21] Y. L. Huang, Y. G. Zhang, N. Li, and J. Chambers, "A robust Gaussian approximate filter for nonlinear systems with heavy-tailed measurement noises," in *2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Mar. 2016.

[22] C. D. Karlgaard and H. Schaub, "Huber-based divided difference filtering," *J. Guid., Control, Dyn.*, vol. 30, no. 3, pp. 885–891, June 2007.

[23] L. B. Chang, B. Q. Hu, G. B. Chang, A. Li, "Multiple outliers suppression derivative-free filter based on unscented transformation," *J. Guid., Control, Dyn.*, vol. 35, no. 6, pp. 1902–1906, Dec. 2012.

[24] L. B. Chang, B. Q. Hu, G. B. Chang, A. Li, "Robust derivative-free Kalman filter based on Huber's M-estimation methodology," *J. Process. Control*, vol. 23, no. 10, pp. 1555–1561, Nov. 2013.

[25] 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 *Proceedings of 50th Annual Conference on Information Science and Systems*, 2016.

[26] Y. L. Huang, Y. G. Zhang, N. Li, S. M. Naqvi, and J. Chambers, "A robust Student's t based cubature filter," in *19th International Conference on Information Fusion (FUSION)*, July 2016, pp. 9–16.

[27] Y. L. Huang and Y. G. Zhang, "Design of high-degree Student's t-based cubature filters," *Circ. Syst. Signal Process.*, DOI: https://doi.org/10.1007/s00034-017-0662-y, 2017.

[28] Y. L. Huang and Y. G. Zhang, "Robust Student's t based stochastic cubature filter for nonlinear systems with heavy-tailed process and measurement noises," *IEEE Access*, vol. 5, no. 5, pp. 7964–7974, May 2017.

[29] Y. L. Huang and Y. G. Zhang, "A new process uncertainty robust Student's t based Kalman filter for SINS/GPS integration," *IEEE Access*, vol. 5, no. 7, pp. 14391–14404, July 2017.

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

[31] S. Saha, "Noise robust online inference for linear dynamic systems," [Online]. Available: http://arxiv.org/pdf/1504.05723.

[32] M. West, "On scale mixtures of normal distributions," *Biometrika*, vol. 74, no. 3, pp. 646–648, Sep. 1987.

[33] S. B. Choy and J. S. Chan, "Scale mixtures distributions in statistical modelling," *Australian & New Zealand Journal of Statistics*, vol. 50, no. 2, pp. 135–146, May 2008.

[34] A. O'Hagan and J. J. Forster, *Kendall's Advanced Theory of Statistics: Bayesian Inference*. Arnold, 2004.

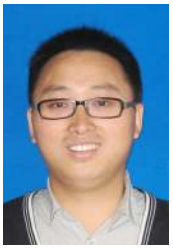
[35] Y. L. Huang, Y. G. Zhang, Z. M. Wu, N. Li, and J. Chambers, "A novel adaptive Kalman filter with inaccurate process and measurement noise covariance matrices," *IEEE Trans. Autom. Control*, DOI: 10.1109/TAC.2017.2730480, 2017.

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

[37] I. Bilik and J. Tabrikian, "Maneuvering target tracking in the presence of glint using the nonlinear gaussian mixture Kalman filter," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 46, no. 1, pp. 246–262, Jan. 2010.

[38] Y. L. Huang, Y. G. Zhang, N. Li, and Z. Shi, "Design of Gaussian approximate filter and smoother for nonlinear systems with correlated noises at one epoch apart," *Circ. Syst. Signal Process.*, vol. 35, no. 11, pp. 3981–4008, Nov. 2016.

[39] Y. G. Zhang and Y. L. Huang, "Gaussian approximate filter for stochastic dynamic systems with randomly delayed measurements and colored measurement noises," *Sci. China Inform. Sci.*, vol. 59, no. 9, pp. 1–18, Sep. 2016.



Yulong Huang received the B.S. degree from the Department of Automation, Harbin Engineering University, Harbin, China, in 2012, and is currently working towards a Ph.D degree in control science and engineering. Since Nov. 2016, he has been a visiting graduate researcher at the electrical engineering department of Columbia University, New York, USA. His current research interests include signal processing, information fusion and their applications in navigation technology, such as inertial navigation and integrated navigation.

Yonggang Zhang (S'06-M'07-SM'16) received the B.S. and M.S. degrees from the Department of Automation, Harbin Engineering University, Harbin, China, in 2002 and 2004, respectively. He received his Ph.D. degree in Electronic Engineering from Cardiff University, UK in 2007 and worked as a Post-Doctoral Fellow at Loughborough University, UK from 2007 to 2008 in the area of adaptive signal processing. Currently, he is a Professor of navigation, guidance, and control in Harbin Engineering University (HEU) in China. His current research interests include signal processing, information fusion and their applications in navigation technology, such as fiber optical gyroscope, inertial navigation and integrated navigation.

Peng Shi (M'95-SM'98-F'15) received the PhD degree in Electrical Engineering from the University of Newcastle, Australia in 1994; the PhD degree in Mathematics from the University of South Australia in 1998. He was awarded the Doctor of Science degree from the University of Glamorgan, Wales in 2006; and the Doctor of Engineering degree from the University of Adelaide, Australia in 2015.

Dr Shi is now a professor at the University of Adelaide. His research interests include system and control theory, computational intelligence, and operational research. He is a co-recipient of the Andrew Sage Best Transactions Paper Award from IEEE Systems, Man and Cybernetics Society in 2016. He has served on the editorial board of a number of journals, including Automatica, IEEE Transactions on Automatic Control; IEEE Transactions on Fuzzy Systems; IEEE Transactions on Cybernetics; IEEE Transactions on Circuits and Systems; IEEE Access; and IEEE Control Systems Letters. He now serves as an IEEE Distinguished Lecturer; and Australian Research Council College of Expert Member. He is a Fellow of the Institution of Engineering and Technology.

Zhemin Wu received the B.S. and M.S. degrees from the Department of Automation, Harbin Engineering University, Harbin, China, in 2012 and 2015, respectively. She is currently working towards a Ph.D degree at the School of Electrical Engineering and Automation, Harbin Institute of Technology, Harbin, China. Her current research interests include signal processing and its application in navigation technology, such as GPS, inertial navigation and integrated navigation.

Junhui Qian is currently working toward the Ph.D. degree in signal and information processing at the University of Electronic Science and Technology of China, Chengdu, China. Since Nov. 2016, he has been a visiting graduate researcher at the electrical engineering department of Columbia University, New York, USA. His research interests are in signal processing for MIMO radar and communication systems.

Jonathon A. Chambers (S'83-M'90-SM'98-F'11) received the Ph.D. and D.Sc. degrees in signal processing from the Imperial College of Science, Technology and Medicine (Imperial College London), London, U.K., in 1990 and 2014, respectively. From 1991 to 1994, he was a Research Scientist with the Schlumberger Cambridge Research Center, Cambridge, U.K. In 1994, he returned to Imperial College London as a Lecturer in signal processing and was promoted to Reader (Associate Professor) in 1998. From 2001 to 2004, he was the Director of the Center for Digital Signal Processing and a Professor of signal processing with the Division of Engineering, King's College London. From 2004 to 2007, he was a Cardiff Professorial Research Fellow with the School of Engineering, Cardiff University, Cardiff, U.K. Between 2007-2014, he led the Advanced Signal Processing Group, within the School of Electronic, Electrical and Systems Engineering and is now a Visiting Professor. In 2015, he joined the School of Electrical and Electronic Engineering, Newcastle University, where he is a Professor of signal and information processing and heads the ComS2IP group. He is also a Guest Professor at Harbin Engineering University, China. He is co-author of the books *Recurrent Neural Networks for Prediction: Learning Algorithms, Architectures and Stability*(New York, NY, USA:Wiley, 2001) and *EEG Signal Processing*(New York, NY, USA: Wiley, 2007). He has advised more than 60 researchers through to Ph.D. graduation and published more than 400 conference proceedings and journal articles, many of which are in IEEE journals. His research interests include adaptive and blind signal processing and their applications.

Dr. Chambers is a Fellow of the Royal Academy of Engineering, U.K., and the Institution of Electrical Engineers. He was the Technical Program Chair of the 15th International Conference on Digital Signal Processing and the 2009 IEEE Workshop on Statistical Signal Processing, both held in Cardiff, U.K., and a Technical Program Cochair for the 36th IEEE International Conference on Acoustics, Speech, and Signal Processing, Prague, Czech Republic. He received the first QinetiQ Visiting Fellowship in 2007 for his outstanding contributions to adaptive signal processing and his contributions to QinetiQ, as a result of his successful industrial collaboration with the international defense systems company QinetiQ. He has served on the IEEE Signal Processing Theory and Methods Technical Committee for six years and the IEEE Signal Processing Society Awards Board for three years. He is currently a member of the IEEE Signal Processing Conference Board and the European Signal Processing Society Best Paper Awards Selection Panel. He has also served as an Associate Editor for the IEEE TRANSACTIONS ONSIGNALPROCESSING for three terms over the periods 1997-1999, 2004-2007, and since 2011 as a Senior Area Editor.