
Article:
Huang, Y., Zhang, Y., Zhao, Y. et al. (2 more authors) (2019) Robust Rauch-Tung-Striebel smoothing framework for heavy-tailed and/or skew noises. IEEE Transactions on Aerospace and Electronic Systems. ISSN 0018-9251

https://doi.org/10.1109/TAES.2019.2914520

Robust Rauch-Tung-Striebel Smoothing Framework for Heavy-tailed and/or Skew Noises

Yulong Huang, Yonggang Zhang, *Senior Member, IEEE,* Yuxin Zhao, *Senior Member, IEEE,* Lyudmila Mihaylova, *Senior Member, IEEE,* Jonathon Chambers, *Fellow, IEEE*

Abstract

A novel robust Rauch-Tung-Striebel smoothing framework is proposed based on a generalized Gaussian scale mixture (GGScM) distribution for a linear state-space model with heavy-tailed and/or skew noises. The state trajectory, mixing parameters and unknown distribution parameters are jointly inferred using the variational Bayesian approach. As such, a major contribution of this work is unifying results within the GGScM distribution framework. Simulation and experimental results demonstrate that the proposed smoother has better accuracy than existing smoothers.

Index Terms

Rauch-Tung-Striebel smoother, heavy-tailed noise, heavy-tailed and skew noise, generalized Gaussian scale mixture distribution, variational Bayesian methods

This work was supported by the National Natural Science Foundation of China under Grant Nos. 61773133 and 61633008 and the PhD Student Research and Innovation Fund of the Fundamental Research Founds of the Central Universities under Grant No. HEUGIP201706 and 1000 Talents Funding. Corresponding author is Y. G. Zhang.

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

L. Mihaylova is with the Department of Automatic Control and Systems Engineering, University of Sheffield, Sheffield, UK (e-mail: L.S.Mihaylova@sheffield.ac.uk).

J. Chambers is with the Department of Automation, Harbin Engineering University, and also with the Department of Engineering, University of Leicester, Leicester, UK (e-mail: Jonathon.Chambers@le.ac.uk).

I. INTRODUCTION

The Rauch-Tung-Striebel (RTS) smoother is an off-line state estimator and employs the Kalman filter as its building block. The RTS smoother has better estimation accuracy than the Kalman filter, and it has been extensively used in a large number of applications, such as positioning, navigation, target tracking, and signal processing [1]. The RTS smoother is an optimal state estimator in terms of minimum mean square error for a linear state-space model with Gaussian state and measurement noises, but the estimation accuracy of the RTS smoother degrades dramatically for a state-space model with non-Gaussian noises [2]. In general, it is very difficult to derive an analytical non-Gaussian smoother for a state-space model with non-Gaussian noises because there is no general mathematical formulation for non-Gaussian noises nor a closed form for a non-Gaussian posterior probability density function (PDF). The particle filter and smoother can provide approximate solutions for non-Gaussian state estimation by approximating the posterior PDF as a set of weighted random samples based on the sequential Monte Carlo sampling technique [3], [4]. Unfortunately, the particle filter and smoother suffer from particle degeneracy and substantial computational complexities in high-dimensional state estimation [5].

A class of non-Gaussian smoothing problems has been attracting more and more attention, in which the state and/or measurement noises of the state-space model may have heavy-tailed and/or skew distributions. Recently, a large number of robust smoothers have been proposed to solve such non-Gaussian smoothing problems, in which the state and/or measurement noises are modelled by Student's t-distribution or skew t-distribution [6]–[14]. These robust smoothers can be divided into three categories: robust and trend-following Student's t RTS (RTF-ST-RTS) smoother [7], [8], Student's t-smoother [10], [11] and robust Student's t or skew t-based RTS smoother [6], [9], [12]–[14]. The RTF-ST-RTS smoother is a robust maximum a posteriori (MAP) estimator, in which the convex composite extension of the Gauss-Newton method is utilized to find an approximate MAP estimate of the state trajectory [8]. The Student's t-smoother and robust Student's t or skew t-based RTS smoother are, respectively, the smoothing extensions of the Student's t-filter [10], [11], [15] and robust Student's t or skew t-based Kalman filter [2], [13], [14], [16]–[19]. For the Student's t-filter and smoother, the posterior PDF of the state vector is approximated by a Student's t-distribution with fixed degrees of freedom (dof) parameter using

the Bayesian rule [10], and many advanced Student's weighted integral rules have been also proposed to implement the nonlinear Student's t filter and smoother [20]–[24]. On the contrary, for the robust Student's t or skew t-based Kalman filter and RTS smoother, the posterior PDF of the state vector is approximated by a Gaussian distribution, where the Student's t-distribution or skew t-distribution is formulated as a hierarchical Gaussian form conditioned on an auxiliary random variable, and the state vector and auxiliary random variable are jointly inferred based on the variational Bayesian (VB) approach [6], [14]. These filtering and smoothing algorithms have been used in many applications, including manoeuvring target tracking [15], [25], [26], visual tracking [5], and cooperative navigation and localization of autonomous underwater vehicles (AUVs) [27], [28], [29], in which the heavy-tailed and/or skew distributions of state and/or measurement noises are often induced by the impulsive interferences, outliers and modelling artifacts. Furthermore, the adaptive Kalman filter and RTS smoother based on the VB approach can to some extent address heavy-tailed state and measurement noises by adaptively modifying the one-step prediction error covariance and noise covariance matrices [30], [31].

To better model heavy-tailed and/or skew noises, Gaussian scale mixture (GScM) distribution based Kalman filters (GScM-KFs) have been proposed using the VB approach [32], which can achieve better estimation accuracy than Student's t or skew t-based filters. Unfortunately, the GScM distribution only covers limited non-Gaussian heavy-tailed and/or skew distributions since it employs a fixed skew function and three fixed mixing densities so that modelling errors may exist in some applications. More importantly, the existing GScM-KFs suffer from some drawbacks as follows. Firstly, in GScM-KFs, the unknown mixing parameters are approximated by their maximum a posterior estimates that are very crude Bayesian approximations since they only include the mode information of the posterior PDF [33]. As a result, the estimation accuracy may degrade. Secondly, in GScM-KFs, the dof parameter of the GScM distribution requires to be selected beforehand in terms of simulation or engineering experience, which reduces engineering practicality. Furthermore, so far, a general framework of robust RTS smoother based on a generalized non-Guassian heavy-tailed and/or skew distribution has not been proposed, which can further improve the estimation accuracy of robust Kalman filters. Although the existing GScM-KFs can be extended to robust RTS smoothers by modelling state and measurement noises as GScM distributed, the resultant robust RTS smoothers will inherit the drawbacks of existing GScM-KFs.

Advanced approaches are needed that are able to improve the estimation accuracy of non-Gaussian smoothers in two aspects: 1) Non-Gaussian noises must be modelled in a better way as compared with the GScM distribution. 2) Efficient approaches are required for the joint estimation of state trajectory, mixing parameters and unknown parameters of modelling distributions.

In this paper, a generalized Gaussian scale mixture (GGScM) distribution is proposed to better model the heavy-tailed and/or skew noises, for which the existing GScM distribution is a special case. The GGScM distribution can be formulated as a hierarchical Gaussian form conditioned on a random mixing parameter that follows a continuous probability distribution with a positive orthant as support. Different GGScM distributions can be achieved when the mixing parameter is sampled from carefully chosen distributions. As such, a major contribution of this work is to provide unified results on the basis of the exemplary GGScM distribution framework.

A new hierarchical Gaussian state-space model is constructed based on the GGScM distribution, from which the state trajectory, mixing parameters and unknown distribution parameters are jointly inferred using the VB approach. A novel robust RTS smoothing framework is proposed based on the GGScM distribution for a linear state-space model with heavy-tailed and/or skew state and measurement noises. The posterior PDFs of state trajectory, shape parameters, scale matrices and dof parameters are, respectively, approximated by Gaussian, Gaussian, inverse-Wishart and Gamma distributions, and the posterior PDFs of mixing parameters are approximated by a weighted set of particles using the Monte Carlo approach. The existing adaptive RTS smoother [31] is a special case of the proposed robust RTS smoothing framework when the state and measurement noises are modelled by a Gaussian distribution. The proposed robust RTS smoothing framework degrades into the existing robust VB and Student's t based RTS (VB-ST-RTS) smoother [6] when the state and measurement noises are modelled by a Gaussian-Gamma mixture distribution that is a special form of the GGScM distribution.

To illustrate the proposed robust RTS smoothing framework, several particular solutions corresponding to the exemplary GGScM distributions are derived, in which both the analytical update and Monte Carlo update of posterior PDFs of mixing parameters are provided. The robustness analyses of the proposed robust RTS smoothers based on exemplary GGScM distributions are provided to reveal the advantages of the proposed method. Moreover, a new Kullback-Leibler divergence (KLD)-based scheme is proposed to facilitate the selection of GGScM distributions in practical applications. The proposed method has the potential to be used in some applications,

such as the cooperative navigation and localization of AUVs, the radio signal based indoor localization, and the inertial navigation system and global positioning system based integrated navigation, which may suffer from heavy-tailed and/or skew noises. The proposed robust RTS smoothers and existing state-of-the-art smoothers are compared by two representative examples: stochastic volatility model and cooperative localization of an AUV. Simulation and experimental results show that the proposed robust RTS smoothers have better estimation accuracy but higher computational complexities than existing state-of-the-art smoothers.

The remainder of this paper is organized as follows. In Section II, the notations that are used in this paper are given. In Section III, a new GGScM distribution is proposed, and brief descriptions and comparisons of the exemplary GGScM distributions are provided. In Section IV, a novel robust RTS smoothing framework based on the GGScM distribution is proposed using the VB approach. In Section V, robust RTS smoothers based on the exemplary GGScM distributions are derived, and robustness analyses are provided, and a new KLD-based selection scheme for GGScM distributions is proposed. In Section VI, simulation and experimental comparisons between the proposed robust RTS smoothers and existing state-of-the-art smoothers are given. Concluding remarks are drawn in Section VII.

II. NOTATIONS

Throughout this paper, we denote $\mathbf{y}_{i:j} \triangleq \{\mathbf{y}_k | i \leq k \leq j\}$, and i, j and k denote the time samples i, j and k, respectively; $\mathrm{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ denotes the multivariate Gaussian distribution with mean vector $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$; $\mathrm{g}(\mathbf{x}; \boldsymbol{\mu}, \boldsymbol{\Sigma})$ denotes the PDF of $\mathbf{x} \sim \mathrm{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$; $\mathrm{IW}(\cdot; \nu, \boldsymbol{\Sigma})$ denotes the inverse-Wishart PDF with dof parameter ν and inverse scale matrix $\boldsymbol{\Sigma}$; $\mathrm{W}(\cdot; \nu, \boldsymbol{\Sigma})$ denotes the Wishart PDF with dof parameter ν and scale matrix $\boldsymbol{\Sigma}$; $\mathrm{G}(\cdot; a, b)$ denotes the Gamma PDF with shape parameter a and rate parameter b; $\mathrm{IG}(\cdot; a, b)$ denotes the inverse-Gamma PDF with shape parameter a and scale parameter b; $\mathrm{Ex}(\cdot; \lambda)$ denotes the exponential distribution with rate parameter λ; $\mathrm{IEx}(\cdot; \lambda)$ denotes the inverse exponential distribution with scale parameter λ; $\mathrm{Be}(\cdot; a, b)$ denotes the Beta PDF with shape parameters a and b; $\mathrm{GIG}(\cdot; a, b, p)$ denotes the generalized inverse Gaussian PDF with shape parameters a, b and p; $\log$ denotes the natural logarithm; $\exp$ denotes the natural exponential; $\mathbf{I}_n$ denotes the $n \times n$ identity matrix; $\mathrm{E}_x[\cdot]$ is the expectation operator with respect to the distribution of x; the superscript "-1" denotes the inverse operation of a matrix; the superscript "T" denotes the transpose operation of a vector or

matrix; $\mathrm{tr}(\cdot)$ denotes the trace operation of a matrix; $\delta(\cdot)$ denotes the Dirac delta function; $\bigcup$ denotes the union operation; $K_\rho(\cdot)$ denotes a modified Bessel function of the second kind with the order ρ; and $\psi(\cdot)$ denotes the digamma function.

III. GGScM Distribution

In practical engineering application, many types of non-Gaussian noises, which are induced by impulsive interferences, outliers and modelling artifacts, often have heavy-tailed and/or skew distributions. In this paper, we propose a GGScM distribution which is able to model such non-Gaussian noises. A random vector $\mathbf{x}$ follows a GGScM distribution if its PDF can be formulated as

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

where $y > 0$ is the mixing parameter, and $\boldsymbol{\mu}$, $\boldsymbol{\Sigma}$ and ν are, respectively, the location parameter, scale matrix and dof parameter, and $s(\cdot)$ and $\kappa(\cdot)$ are, respectively, positive skew and scale functions, i.e., $s(y) > 0$ and $\kappa(y) > 0$ for $\forall y > 0$, and $\pi(\cdot;\nu)$ is the mixing density with a dof parameter ν defined on $(0,\ +\infty)$, and $\boldsymbol{\beta}$ is a shape parameter. The shape parameter $\boldsymbol{\beta}$ dominates the symmetry and skewness of a GGScM distribution. The GGScM distribution is symmetric when $\boldsymbol{\beta} = \mathbf{0}$ and non-symmetric when $\boldsymbol{\beta} \neq \mathbf{0}$, and it is positive skew when $\boldsymbol{\beta}_i > 0$ and negative skew when $\boldsymbol{\beta}_i < 0$, where $\boldsymbol{\beta}_i$ is the i-th element of $\boldsymbol{\beta}$. The GGScM distribution can be divided into two categories: symmetric GGScM distribution and skew GGScM distribution. The GGScM distribution degrades into a GScM distribution when $\boldsymbol{\beta} = \mathbf{0}$ [34], and the GGScM distribution becomes a Gaussian distribution when $\boldsymbol{\beta} = \mathbf{0}$, $\kappa(y) = y$ and $\pi(y;\nu) = \delta(y-1)$.

Next, we compare the tail behaviours of the proposed GGScM distribution and Gaussian distribution. Without loss of generality, a one-dimensional GGScM distribution and a one-dimensional Gaussian distribution are considered, in which both the location parameter of the GGScM distribution and the mean value of the Gaussian distribution are set as $\mu = 0$. The GGScM and Gaussian distributions are, respectively, formulated as

$$\begin{cases} p_{GGScM}(x) = \int_0^{+\infty} \mathrm{g}(x; \beta/s(y), \Sigma/\kappa(y))\pi(y;\nu)dy \\ p_G(x) = \mathrm{g}(x; \beta/s(y), \Sigma) \end{cases} \tag{2}$$

where $p_{GGScM}(\cdot)$ and $p_G(\cdot)$ denote the GGScM and Gaussian PDFs, respectively. Using (2), the logarithm of the GGScM PDF can be written as

$$\log p_{GGScM}(x) = \log \mathrm{E}_y[\mathrm{g}(x; \beta/s(y), \Sigma/\kappa(y))] \tag{3}$$

Considering the natural logarithm function $\log$ which is a convex function and employing Jensen's inequality, we have

$$\log \mathrm{E}_y[\mathrm{g}(x;\beta/s(y),\Sigma/\kappa(y))] \geq \mathrm{E}_y[\log \mathrm{g}(x;\beta/s(y),\Sigma/\kappa(y))] \tag{4}$$

Exploiting (3)-(4) yields

$$\log p_{GGScM}(x) \geq \mathrm{E}_y[\log \mathrm{g}(x;\beta/s(y),\Sigma/\kappa(y))] \tag{5}$$

According to the definition of the Gaussian PDF and using (2), the expectation $\mathrm{E}_y[\log \mathrm{g}(x;\beta/s(y),\Sigma/\kappa(y))]$ and $\log p_G(x)$ can be formulated as

$$\mathrm{E}_y[\log \mathrm{g}(x;\beta/s(y),\Sigma/\kappa(y))] = -0.5\log 2\pi\Sigma + 0.5\mathrm{E}_y[\log\kappa(y)] - \frac{0.5\mathrm{E}_y[\kappa(y)]}{\Sigma}x^2 + \frac{\beta\mathrm{E}_y[\kappa(y)/s(y)]}{\Sigma}x - \frac{0.5\beta^2\mathrm{E}_y[\kappa(y)/s^2(y)]}{\Sigma} \tag{6}$$

$$\log p_G(x) = -0.5\log 2\pi\Sigma - \frac{0.5}{\Sigma}x^2 \tag{7}$$

Subtracting (7) from (6) gives

$$\mathrm{E}_y[\log \mathrm{g}(x;\beta/s(y),\Sigma/\kappa(y))] - \log p_G(x) = 0.5\mathrm{E}_y[\log\kappa(y)] + \frac{0.5(1-\mathrm{E}_y[\kappa(y)])}{\Sigma}x^2 + \frac{\beta\mathrm{E}_y[\kappa(y)/s(y)]}{\Sigma}x - \frac{0.5\beta^2\mathrm{E}_y[\kappa(y)/s^2(y)]}{\Sigma} \tag{8}$$

Taking the limit operation on both sides of the equation (8) yields

$$\lim_{x\to\pm\infty}\{\mathrm{E}_y[\log \mathrm{g}(x;\beta/s(y),\Sigma/\kappa(y))] - \log p_G(x)\} = \lim_{x\to\pm\infty}\left\{\frac{0.5(1-\mathrm{E}_y[\kappa(y)])}{\Sigma}x^2 + \frac{\mathrm{E}_y[\kappa(y)/s(y)]}{\Sigma}\beta x\right\} \tag{9}$$

If $\mathrm{E}_y[\kappa(y)] \leq 1$ and $\beta x \geq 0$, then equation (9) can be rewritten as

$$\lim_{x\to\pm\infty}\{\mathrm{E}_y[\log \mathrm{g}(x;\beta/s(y),\Sigma/\kappa(y))] - \log p_G(x)\} \geq 0 \tag{10}$$

Substituting (5) in (10), we have

$$\lim_{x\to\pm\infty}\{\log p_{GGScM}(x) - \log p_G(x)\} \geq 0 \quad \text{if } \mathrm{E}_y[\kappa(y)] \leq 1, \quad \beta x \geq 0 \tag{11}$$

Employing (11) yields

$$\lim_{x\to\pm\infty}\frac{p_{GGScM}(x)}{p_G(x)} \geq 1 \quad \text{if } \mathrm{E}_y[\kappa(y)] \leq 1, \quad \beta x \geq 0 \tag{12}$$

TABLE I: Exemplary symmetric GGScM distributions and their parameters.

Symmetric GGScM distributions	Shape parameter	Scale function	Mixing density	Constraints
GGM	$\boldsymbol{\beta} = 0$	$\kappa(y) = y$	$\pi(y;\nu) = \mathrm{G}(y; \frac{\nu}{2}, \frac{\nu}{2})$	$y > 0$, $\nu > 0$
GEM	$\boldsymbol{\beta} = 0$	$\kappa(y) = y$	$\pi(y;\nu) = \mathrm{Ex}(y;\nu)$	$y > 0$, $\nu > 0$
GBM	$\boldsymbol{\beta} = 0$	$\kappa(y) = y$	$\pi(y;\nu) = \mathrm{Be}(y;\nu,1)$	$0 < y < 1$, $\nu > 0$
GIGM	$\boldsymbol{\beta} = 0$	$\kappa(y) = y$	$\pi(y;\nu) = \mathrm{IG}(y; \frac{\nu}{2}, \frac{\nu}{2})$	$y > 0$, $\nu > 0$
GIEM	$\boldsymbol{\beta} = 0$	$\kappa(y) = y$	$\pi(y;\nu) = \mathrm{IEx}(y;\nu)$	$y > 0$, $\nu > 0$

It is observed from (12) that the proposed GGScM distribution has heavier tails than the Gaussian distribution when $\mathrm{E}_y[\kappa(y)] \leq 1$ and $\beta x \geq 0$. The condition $\beta x \geq 0$ implies that the range of x depends on the shape parameter β, and β and x have the same sign. This means that if the shape parameter is positive, i.e., $\beta > 0$, then x has to be positive to satisfy the condition $\beta x \geq 0$, and vice versa. Thus, the tail behaviour of the proposed GGScM distribution is determined by both the dof parameter ν and the shape parameter β.

A. Symmetric GGScM distribution

The symmetric GGScM distribution is a special case of the GGScM distribution when the shape parameter $\boldsymbol{\beta} = \mathbf{0}$. The symmetric GGScM distribution has heavier tails than the Gaussian distribution, which makes it more suitable for modelling heavy-tailed noises as compared with the Gaussian distribution. Different symmetric GGScM distributions can be achieved when different scale functions and mixing densities are selected. For example, five symmetric GGScM distributions can be obtained when the scale function is set as $\kappa(y) = y$ and mixing densities are, respectively, chosen as Gamma, exponential, Beta, inverse Gamma, and inverse exponential distributions. In this paper, the five exemplary symmetric GGScM distributions are, respectively, named Gaussian-Gamma mixture (GGM) distribution, Gaussian-exponential mixture (GEM) distribution, Gaussian-Beta mixture (GBM) distribution, Gaussian-inverse-Gamma mixture (GIGM) distribution, and Gaussian-inverse-exponential mixture (GIEM) distribution. Five exemplary symmetric GGScM distributions and their parameters are listed in Table I. Substituting $\boldsymbol{\beta} = \mathbf{0}$ and $\kappa(y) = y$ in (1), the PDFs of the five exemplary symmetric GGScM distributions can be

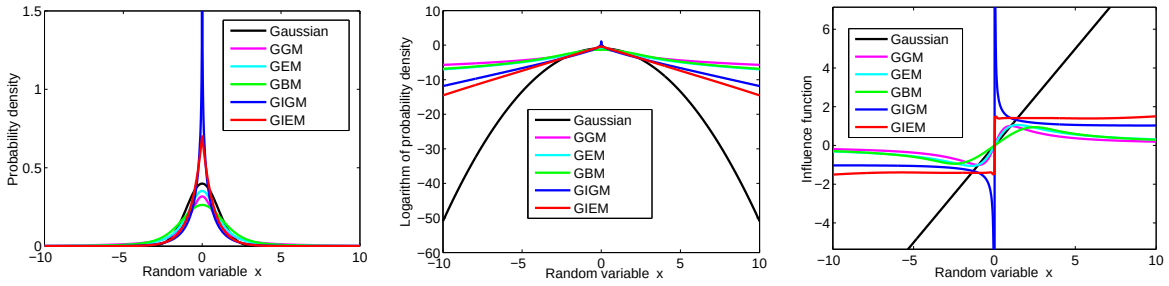


Fig. 1: Gaussian and exemplary symmetric GGScM densities, corresponding log plots, and influence functions for a scalar case.

formulated as a general form

$$p(\mathbf{x}) = \int_0^{+\infty} \mathrm{g}(\mathbf{x}; \boldsymbol{\mu}, \boldsymbol{\Sigma}/y)\pi(y;\nu)dy \tag{13}$$

The existing Student's t, slash, and variance-gamma distributions [34] can be formulated as (13) with mixing densities $\pi(y) = \mathrm{G}(y; \frac{\nu}{2}, \frac{\nu}{2})$, $\pi(y) = \mathrm{Be}(y; \nu, 1)$ and $\pi(y) = \mathrm{IG}(y; \frac{\nu}{2}, \frac{\nu}{2})$, respectively. It is seen from Table I and equation (13) that the GGM, GBM, GIGM distributions are, respectively, identical to the existing Student's t, slash, variance-gamma distributions.

Comparisons of the Gaussian and exemplary symmetric GGScM densities and corresponding influence functions are shown in Fig. 1, where the localization parameter, scale parameter and dof parameter are, respectively, selected as $\mu = 0$, $\Sigma = 1$ and $\nu = 1$. It is seen from Fig. 1 that the exemplary symmetric GGScM distributions all have heavier tails than the Gaussian distribution, and they have different tail behaviours. We can also see from Fig. 1 that the exemplary symmetric GGScM distributions all have limited influence functions. As a result, the heavy-tailed noises have limited influences on the state estimator, which can resist negative effects of heavy-tailed noises.

B. Skew GGScM distribution

The skew GGScM distribution is non-symmetric, and its shape parameter $\boldsymbol{\beta} \neq \mathbf{0}$. The skew GGScM distribution has both heavier tails and higher skewness than the Gaussian distribution, which makes it more suitable for modelling heavy-tailed and skew noises as compared with the Gaussian distribution. By choosing different skew and scale functions and mixing densities, different skew GGScM distributions can be obtained. For instance, five skew GGScM distributions

TABLE II: Exemplary skew GGScM distributions and their parameters.

Skew GGScM distributions	Shape parameter	Skew and scale functions	Mixing density	Constraints
SGGM	$\boldsymbol{\beta} \neq 0$	$s(y) = \kappa(y) = y$	$\pi(y;\nu) = \mathrm{G}(y; \frac{\nu}{2}, \frac{\nu}{2})$	$y > 0$, $\nu > 0$
SGEM	$\boldsymbol{\beta} \neq 0$	$s(y) = \kappa(y) = y$	$\pi(y;\nu) = \mathrm{Ex}(y;\nu)$	$y > 0$, $\nu > 0$
SGBM	$\boldsymbol{\beta} \neq 0$	$s(y) = \kappa(y) = y$	$\pi(y;\nu) = \mathrm{Be}(y;\nu,1)$	$0 < y < 1$, $\nu > 0$
SGIGM	$\boldsymbol{\beta} \neq 0$	$s(y) = \kappa(y) = y$	$\pi(y;\nu) = \mathrm{IG}(y; \frac{\nu}{2}, \frac{\nu}{2})$	$y > 0$, $\nu > 0$
SGIEM	$\boldsymbol{\beta} \neq 0$	$s(y) = \kappa(y) = y$	$\pi(y;\nu) = \mathrm{IEx}(y;\nu)$	$y > 0$, $\nu > 0$

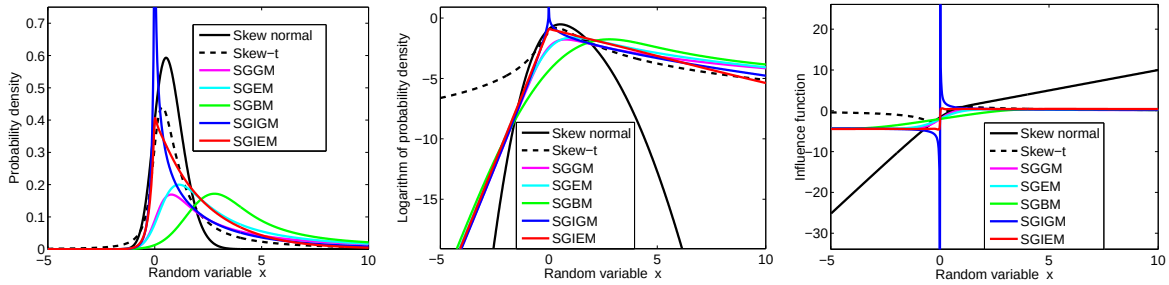


Fig. 2: Skew normal, skew t and exemplary skew GGScM densities, corresponding log plots, and influence functions for a scalar case.

can be achieved when the skew and scale functions are set as $s(y) = \kappa(y) = y$ and mixing densities are, respectively, selected as Gamma, exponential, Beta, inverse Gamma, and inverse exponential distributions, and the corresponding skew GGScM distributions are, respectively, named skew Gaussian-Gamma mixture (SGGM) distribution, skew Gaussian-exponential mixture (SGEM) distribution, skew Gaussian-Beta mixture (SGBM) distribution, skew Gaussian-inverse-Gamma mixture (SGIGM) distribution, and skew Gaussian-inverse-exponential mixture (SGIEM) distribution. Five exemplary skew GGScM distributions and their parameters are listed in Table II. Employing $s(y) = \kappa(y) = y$ in (1), the PDFs of the five exemplary skew GGScM distributions can be written in a unified form

$$p(\mathbf{x}) = \int_0^{+\infty} \mathrm{g}(\mathbf{x}; \boldsymbol{\mu} + \boldsymbol{\beta}/y, \boldsymbol{\Sigma}/y)\pi(y;\nu)dy \tag{14}$$

Using $y = 1/z$, (14) can be rewritten as

$$\begin{aligned} p(\mathbf{x}) &= \int_0^{+\infty} \mathrm{g}(\mathbf{x}; \boldsymbol{\mu} + z\boldsymbol{\beta}, z\boldsymbol{\Sigma}) \frac{1}{z^2} \pi(\frac{1}{z}; \nu) dz \\ &= \int_0^{+\infty} \mathrm{g}(\mathbf{x}; \boldsymbol{\mu} + y\boldsymbol{\beta}, y\boldsymbol{\Sigma}) \frac{1}{y^2} \pi(\frac{1}{y}; \nu) dy \end{aligned} \tag{15}$$

Utilizing (15), the PDFs of SGGM and SGIGM distributions can be, respectively, reformulated as

$$\begin{cases} p(\mathbf{x}) = \int_0^{+\infty} \mathrm{g}(\mathbf{x}; \boldsymbol{\mu} + y\boldsymbol{\beta}, y\boldsymbol{\Sigma}) \mathrm{IG}(y; \frac{\nu}{2}, \frac{\nu}{2}) dy \\ p(\mathbf{x}) = \int_0^{+\infty} \mathrm{g}(\mathbf{x}; \boldsymbol{\mu} + y\boldsymbol{\beta}, y\boldsymbol{\Sigma}) \mathrm{G}(y; \frac{\nu}{2}, \frac{\nu}{2}) dy \end{cases} \tag{16}$$

We can observe from (16) that the SGGM and SGIGM distributions are, respectively, identical to the existing generalized hyperbolic skew Student's t-distribution and generalized hyperbolic variance-gamma distribution [34].

Comparisons of the skew normal, skew t, and exemplary skew GGScM densities and corresponding influence functions are shown in Fig. 2, where the localization parameter, scale parameter, dof parameter and shape parameter are, respectively, selected as $\mu = 0$, $\Sigma = 1$, $\nu = 1$ and $\beta = 2$. It is observed from Fig. 2 that the skew t-distribution and the exemplary skew GGScM distributions all have heavier tails than the skew normal distribution, and the exemplary skew GGScM distributions have different tail behaviours and skew properties. We can also observe from Fig. 2 that the exemplary skew GGScM distributions all have limited influence functions. As a result, the heavy-tailed and skew noises have limited influences on the state estimator, which is able to resist negative influences of heavy-tailed and skew noises.

IV. A Novel Robust RTS Smoothing Framework

A. Novel hierarchical Gaussian state-space model

Consider the following discrete-time linear stochastic system as represented by a linear state-space model

$$\begin{cases} \mathbf{x}_k = \mathbf{F}_k \mathbf{x}_{k-1} + \mathbf{w}_{k-1} & \text{(state equation)} \\ \mathbf{z}_k = \mathbf{H}_k \mathbf{x}_k + \mathbf{v}_k & \text{(measurement equation)} \end{cases} \tag{17}$$

where $k = 1, \ldots, T$ is 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, the known state transition matrix and measurement matrix, 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 $\mathbf{x}_0$ is assumed to have 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 $\mathbf{P}_{0|0}$, respectively, denote the initial state estimate and the initial estimation error covariance matrix. Moreover, $\mathbf{x}_0$, $\mathbf{w}_k$ and $\mathbf{v}_k$ are assumed to be mutually independent. Our aim is to estimate the state trajectory $\mathbf{x}_{0:T}$ based on the linear state-space model and available measurements $\mathbf{z}_{1:T}$ from time sample 1 to time sample T.

In this paper, the state and measurement noises are assumed to have heavy-tailed and/or skew distributions, which are modelled as GGScM distributed as

$$\begin{cases} p(\mathbf{w}_{k-1}) = \int_0^{+\infty} \mathrm{g}(\mathbf{w}_{k-1}; \boldsymbol{\beta}_1/s_1(\xi_k), \mathbf{Q}/\kappa_1(\xi_k))\pi_1(\xi_k;\omega)d\xi_k \\ p(\mathbf{v}_k) = \int_0^{+\infty} \mathrm{g}(\mathbf{v}_k; \boldsymbol{\beta}_2/s_2(\lambda_k), \mathbf{R}/\kappa_2(\lambda_k))\pi_2(\lambda_k;\nu)d\lambda_k \end{cases} \tag{18}$$

where $\mathbf{Q}$, $\mathbf{R}$, $\boldsymbol{\beta}_1$, $\boldsymbol{\beta}_2$, ξ_k, λ_k, ω, ν, $s_1(\cdot)$, $s_2(\cdot)$, $\kappa_1(\cdot)$, $\kappa_2(\cdot)$, $\pi_1(\cdot;\cdot)$ and $\pi_2(\cdot;\cdot)$ are, respectively, the scale matrices, shape parameters, mixing parameters, dof parameters, positive skew functions, positive scale functions and mixing densities of the state and measurement noises.

The scale matrices $\mathbf{Q}$ and $\mathbf{R}$, shape parameters $\boldsymbol{\beta}_1$ and $\boldsymbol{\beta}_2$, and dof parameters ω and ν are unknown, whose joint prior PDF is defined over a limited support and assumed to be a constant, i.e.,

$$p(\mathbf{Q}, \mathbf{R}, \boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \omega, \nu) = c \tag{19}$$

and they will be jointly estimated using the VB approach. Equation (19) means that the joint prior PDF of the scale matrices, shape parameters and dof parameters is uninformative. Such a prior model is reasonable since the scale matrices, shape parameters and dof parameters often have no prior information available in practical applications. Note that, in order to estimate the scale matrices, shape parameters and dof parameters accurately, a large amount of measurement data is required when there is no prior information available.

Exploiting (17)-(18), the state transition PDF $p(\mathbf{x}_k|\mathbf{x}_{k-1})$ and the likelihood PDF $p(\mathbf{z}_k|\mathbf{x}_k)$ can be expressed as

$$p(\mathbf{x}_k|\mathbf{x}_{k-1}) = \int_0^{+\infty} \mathrm{g}(\mathbf{x}_k; \mathbf{F}_k\mathbf{x}_{k-1} + \boldsymbol{\beta}_1/s_1(\xi_k), \mathbf{Q}/\kappa_1(\xi_k))\pi_1(\xi_k;\omega)d\xi_k \tag{20}$$

$$p(\mathbf{z}_k|\mathbf{x}_k) = \int_0^{+\infty} \mathrm{g}(\mathbf{z}_k; \mathbf{H}_k\mathbf{x}_k + \boldsymbol{\beta}_2/s_2(\lambda_k), \mathbf{R}/\kappa_2(\lambda_k))\pi_2(\lambda_k;\nu)d\lambda_k \tag{21}$$

According to (20)-(21), the state transition PDF and the likelihood PDF can be, respectively, written in the following hierarchical Gaussian forms

$$\begin{cases} p(\mathbf{x}_k|\mathbf{x}_{k-1},\xi_k) = \mathrm{g}(\mathbf{x}_k;\mathbf{F}_k\mathbf{x}_{k-1}+\boldsymbol{\beta}_1/s_1(\xi_k),\mathbf{Q}/\kappa_1(\xi_k)) \\ p(\xi_k) = \pi_1(\xi_k;\omega), \qquad \text{s.t. } \xi_k > 0 \end{cases} \tag{22}$$

$$\begin{cases} p(\mathbf{z}_k|\mathbf{x}_k,\lambda_k) = \mathrm{g}(\mathbf{z}_k;\mathbf{H}_k\mathbf{x}_k+\boldsymbol{\beta}_2/s_2(\lambda_k),\mathbf{R}/\kappa_2(\lambda_k)) \\ p(\lambda_k) = \pi_2(\lambda_k;\nu), \qquad \text{s.t. } \lambda_k > 0 \end{cases} \tag{23}$$

Equations (19) and (22)-(23) constitute a novel hierarchical Gaussian state-space model based on the proposed GGScM distribution. The fixed interval smoothing estimation problem for a linear state-space model with heavy-tailed and/or skew state and measurement noises is transformed into the fixed interval smoothing estimation problem for a hierarchical Gaussian state-space model formulated in (19) and (22)-(23). Next, we propose to jointly estimate the state trajectory, mixing parameters, scale matrices, shape parameters and dof parameters, i.e., $\boldsymbol{\Theta} = \{\mathbf{x}_{0:T}, \xi_{1:T}, \lambda_{1:T}, \mathbf{Q}, \mathbf{R}, \boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \omega, \nu\}$, based on the constructed hierarchical Gaussian state-space model using the VB approach.

B. Joint estimates of state trajectory, mixing parameters and unknown distribution parameters

To jointly infer state trajectory, mixing parameters and unknown distribution parameters, the joint posterior PDF $p(\boldsymbol{\Theta}|\mathbf{z}_{1:T})$ needs to be calculated. Unfortunately, the optimal solution of the joint posterior PDF is unavailable for hierarchical Gaussian state-space model (19) and (22)-(23) since the marginal posterior PDF of the state trajectory $\mathbf{x}_{0:T}$ does not have a closed-form expression due to the model lacking conjugacy properties. In this paper, the standard VB approach is utilized to achieve an approximation to the true joint posterior PDF $p(\boldsymbol{\Theta}|\mathbf{z}_{1:T})$ as follows

$$p(\boldsymbol{\Theta}|\mathbf{z}_{1:T}) \approx q(\boldsymbol{\Theta}) = q(\mathbf{x}_{0:T})q(\xi_{1:T})q(\lambda_{1:T})q(\mathbf{Q})q(\mathbf{R})q(\boldsymbol{\beta}_1)q(\boldsymbol{\beta}_2)q(\omega)q(\nu) \tag{24}$$

where $q(\theta)$ denotes a free form factored approximation of the true posterior PDF $p(\theta)$, and $\theta \in \boldsymbol{\Theta}$ is an arbitrary element of the set $\boldsymbol{\Theta}$.

In the standard VB approach, the KLD is used as a measure to evaluate the difference between the approximate joint posterior PDF $q(\boldsymbol{\Theta})$ and the true joint posterior PDF $p(\boldsymbol{\Theta}|\mathbf{z}_{1:T})$, and the approximate posterior PDF $q(\theta)$ is achieved by minimizing the KLD between $q(\boldsymbol{\Theta})$ and $p(\boldsymbol{\Theta}|\mathbf{z}_{1:T})$, i.e., $q(\theta) = \arg\min \mathrm{KLD}(q(\boldsymbol{\Theta})||p(\boldsymbol{\Theta}|\mathbf{z}_{1:T}))$. However, it is not possible to minimize

$\mathrm{KLD}(q(\mathbf{\Theta})||p(\mathbf{\Theta}|\mathbf{z}_{1:T}))$ directly because the true joint posterior PDF $p(\mathbf{\Theta}|\mathbf{z}_{1:T})$ is not available. In the standard VB approach, the minimization of $\mathrm{KLD}(q(\mathbf{\Theta})||p(\mathbf{\Theta}|\mathbf{z}_{1:T}))$ is transformed into the maximization of the lower bound of the log-likelihood $F(q(\mathbf{\Theta})) = \int q(\mathbf{\Theta}) \log \frac{p(\mathbf{\Theta},\mathbf{z}_{1:T})}{q(\mathbf{\Theta})}$, and the approximate posterior PDF $q(\theta)$ satisfies the following equation [33], [30]

$$\log q(\theta) = \mathrm{E}_{\mathbf{\Theta}^{(-\theta)}}[\log p(\mathbf{\Theta}, \mathbf{z}_{1:k})] + c_\theta \tag{25}$$

where $\mathbf{\Theta}^{(-\theta)}$ is a subset of $\mathbf{\Theta}$ and it has all elements in $\mathbf{\Theta}$ except for θ, i.e., $\{\theta\} \bigcup \mathbf{\Theta}^{(-\theta)} = \mathbf{\Theta}$, and c_θ denotes a constant value with respect to variable θ.

Due to the mutual dependence and coupling, it is not possible to achieve an analytic solution of $q(\theta)$ using (25). To address this problem, a fixed-point iteration is employed to achieve an approximation of $q(\theta)$ by iteratively solving (25), and a local optimum approximation can be obtained. That is to say, at the $i+1$-th iteration, for an arbitrary element θ, its approximate posterior PDF $q(\theta)$ is updated as $q^{(i+1)}(\theta)$ by using $q^{(i)}(\mathbf{\Theta}^{(-\theta)})$ to calculate the expectation in (25).

1) Variational approximations of posterior PDFs: Using (19) and (22)-(23), the joint PDF $p(\mathbf{\Theta}, \mathbf{z}_{1:T})$ can be formulated as

$$p(\mathbf{\Theta}, \mathbf{z}_{1:T}) = c\mathrm{g}(\mathbf{x}_0; \hat{\mathbf{x}}_{0|0}, \mathbf{P}_{0|0}) \prod_{k=1}^{T} [\mathrm{g}(\mathbf{x}_k; \mathbf{F}_k\mathbf{x}_{k-1} + \boldsymbol{\beta}_1/s_1(\xi_k), \mathbf{Q}/\kappa_1(\xi_k))\mathrm{g}(\mathbf{z}_k; \mathbf{H}_k\mathbf{x}_k +$$
$$\boldsymbol{\beta}_2/s_2(\lambda_k), \mathbf{R}/\kappa_2(\lambda_k))\pi_1(\xi_k;\omega)\pi_2(\lambda_k;\nu)] \tag{26}$$

Let $\theta = \mathbf{x}_{0:T}$ and utilizing (26) in (25), $q^{(i+1)}(\mathbf{x}_{0:T})$ can be updated as Gaussian, i.e.,

$$q^{(i+1)}(\mathbf{x}_{0:T}) = \mathrm{g}(\mathbf{x}_{0:T}; \hat{\mathbf{x}}^{(i+1)}_{0:T|T}, \mathbf{P}^{(i+1)}_{0:T|T}) \tag{27}$$

where the smoothing estimate $\hat{\mathbf{x}}^{(i+1)}_{0:T|T}$ and corresponding estimation error covariance matrix $\mathbf{P}^{(i+1)}_{0:T|T}$ are obtained using the standard RTS smoother [36] with modified mean vectors $\tilde{\mathbf{q}}^{(i)}_k$ and $\tilde{\mathbf{r}}^{(i)}_k$ and covariance matrices $\tilde{\mathbf{Q}}^{(i)}_k$ and $\tilde{\mathbf{R}}^{(i)}_k$ for the state and measurement noises, which are, respectively, given by

$$\begin{cases} \tilde{\mathbf{q}}^{(i)}_k = \frac{\mathrm{E}^{(i)}[\frac{\kappa_1(\xi_k)}{s_1(\xi_k)}]\mathrm{E}^{(i)}[\boldsymbol{\beta}_1]}{\mathrm{E}^{(i)}[\kappa_1(\xi_k)]}, & \tilde{\mathbf{Q}}^{(i)}_k = \frac{\left\{\mathrm{E}^{(i)}[\mathbf{Q}^{-1}]\right\}^{-1}}{\mathrm{E}^{(i)}[\kappa_1(\xi_k)]} \\ \tilde{\mathbf{r}}^{(i)}_k = \frac{\mathrm{E}^{(i)}[\frac{\kappa_2(\lambda_k)}{s_2(\lambda_k)}]\mathrm{E}^{(i)}[\boldsymbol{\beta}_2]}{\mathrm{E}^{(i)}[\kappa_2(\lambda_k)]}, & \tilde{\mathbf{R}}^{(i)}_k = \frac{\left\{\mathrm{E}^{(i)}[\mathbf{R}^{-1}]\right\}^{-1}}{\mathrm{E}^{(i)}[\kappa_2(\lambda_k)]} \end{cases} \tag{28}$$

where the derivations of (27)-(28) are given in Appendix A.

Let $\theta = \boldsymbol{\beta}_1$ and using (26) in (25), $q^{(i+1)}(\boldsymbol{\beta}_1)$ is updated as Gaussian, and let $\theta = \boldsymbol{\beta}_2$ and employing (26) in (25), and $q^{(i+1)}(\boldsymbol{\beta}_2)$ is updated as Gaussian, i.e.,

$$\begin{cases} q^{(i+1)}(\boldsymbol{\beta}_1) = \mathrm{g}(\boldsymbol{\beta}_1; \hat{\boldsymbol{\beta}}_1^{(i+1)}, \mathbf{P}_{\boldsymbol{\beta}_1}^{(i+1)}) \\ q^{(i+1)}(\boldsymbol{\beta}_2) = \mathrm{g}(\boldsymbol{\beta}_2; \hat{\boldsymbol{\beta}}_2^{(i+1)}, \mathbf{P}_{\boldsymbol{\beta}_2}^{(i+1)}) \end{cases} \tag{29}$$

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

$$\begin{cases} \hat{\boldsymbol{\beta}}_1^{(i+1)} = \frac{\sum_{k=1}^{T} \mathrm{E}^{(i)}[\frac{\kappa_1(\xi_k)}{s_1(\xi_k)}]\mathbf{a}_k^{(i+1)}}{\sum_{k=1}^{T} \mathrm{E}^{(i)}[\frac{\kappa_1(\xi_k)}{s_1^2(\xi_k)}]} \\ \mathbf{P}_{\boldsymbol{\beta}_1}^{(i+1)} = \left\{ \mathrm{E}^{(i)}[\mathbf{Q}^{-1}] \sum_{k=1}^{T} \mathrm{E}^{(i)}[\frac{\kappa_1(\xi_k)}{s_1^2(\xi_k)}] \right\}^{-1} \end{cases} \tag{30}$$

$$\begin{cases} \hat{\boldsymbol{\beta}}_2^{(i+1)} = \frac{\sum_{k=1}^{T} \mathrm{E}^{(i)}[\frac{\kappa_2(\lambda_k)}{s_2(\lambda_k)}]\mathbf{b}_k^{(i+1)}}{\sum_{k=1}^{T} \mathrm{E}^{(i)}[\frac{\kappa_2(\lambda_k)}{s_2^2(\lambda_k)}]} \\ \mathbf{P}_{\boldsymbol{\beta}_2}^{(i+1)} = \left\{ \mathrm{E}^{(i)}[\mathbf{R}^{-1}] \sum_{k=1}^{T} \mathrm{E}^{(i)}[\frac{\kappa_2(\lambda_k)}{s_2^2(\lambda_k)}] \right\}^{-1} \end{cases} \tag{31}$$

and $\mathbf{a}_k^{(i+1)}$ and $\mathbf{b}_k^{(i+1)}$ are, respectively, given by

$$\begin{cases} \mathbf{a}_k^{(i+1)} = \mathrm{E}^{(i+1)}[\mathbf{x}_k - \mathbf{F}_k \mathbf{x}_{k-1}] \\ \mathbf{b}_k^{(i+1)} = \mathrm{E}^{(i+1)}[\mathbf{z}_k - \mathbf{H}_k \mathbf{x}_k] \end{cases} \tag{32}$$

where the derivations of (29)-(32) are given in Appendix B.

Let $\theta = \xi_{1:T}$ and using (26) in (25), $\log q^{(i+1)}(\xi_{1:T})$ is calculated as (33), and let $\theta = \lambda_{1:T}$ and employing (26) in (25), $\log q^{(i+1)}(\lambda_{1:T})$ is calculated as (34), i.e.,

$$\begin{aligned} \log q^{(i+1)}(\xi_{1:T}) = 0.5n \sum_{k=1}^{T} \log \kappa_1(\xi_k) - 0.5 \sum_{k=1}^{T} \kappa_1(\xi_k) \mathrm{tr} \big\{ \mathrm{E}^{(i+1)}[(\mathbf{x}_k - \mathbf{F}_k \mathbf{x}_{k-1} - \boldsymbol{\beta}_1 / s_1(\xi_k)) \\ (\mathbf{x}_k - \mathbf{F}_k \mathbf{x}_{k-1} - \boldsymbol{\beta}_1 / s_1(\xi_k))^{\mathrm{T}}] \mathrm{E}^{(i)}[\mathbf{Q}^{-1}] \big\} + \sum_{k=1}^{T} \mathrm{E}^{(i)}[\log \pi_1(\xi_k; \omega)] + c_{\xi_{1:T}} \end{aligned} \tag{33}$$

$$\begin{aligned} \log q^{(i+1)}(\lambda_{1:T}) = 0.5m \sum_{k=1}^{T} \log \kappa_2(\lambda_k) - 0.5 \sum_{k=1}^{T} \kappa_2(\lambda_k) \mathrm{tr} \big\{ \mathrm{E}^{(i+1)}[(\mathbf{z}_k - \mathbf{H}_k \mathbf{x}_k - \boldsymbol{\beta}_2 / s_2(\lambda_k)) \\ (\mathbf{z}_k - \mathbf{H}_k \mathbf{x}_k - \boldsymbol{\beta}_2 / s_2(\lambda_k))^{\mathrm{T}}] \mathrm{E}^{(i)}[\mathbf{R}^{-1}] \big\} + \sum_{k=1}^{T} \mathrm{E}^{(i)}[\log \pi_2(\lambda_k; \nu)] + c_{\lambda_{1:T}} \end{aligned} \tag{34}$$

Using (33)-(34), $\log q^{(i+1)}(\xi_k)$ and $\log q^{(i+1)}(\lambda_k)$ can be formulated as

$$\log q^{(i+1)}(\xi_k) = 0.5n \log \kappa_1(\xi_k) - 0.5\Delta_{1,k}^{(i+1)}\kappa_1(\xi_k) + \Delta_{2,k}^{(i+1)}\frac{\kappa_1(\xi_k)}{s_1(\xi_k)} - 0.5\Delta_{3,k}^{(i+1)}\frac{\kappa_1(\xi_k)}{s_1^2(\xi_k)} + \mathrm{E}^{(i)}[\log \pi_1(\xi_k;\omega)] + c_{\xi_k} \tag{35}$$

$$\log q^{(i+1)}(\lambda_k) = 0.5m \log \kappa_2(\lambda_k) - 0.5\bar{\Delta}_{1,k}^{(i+1)}\kappa_2(\lambda_k) + \bar{\Delta}_{2,k}^{(i+1)}\frac{\kappa_2(\lambda_k)}{s_2(\lambda_k)} - 0.5\bar{\Delta}_{3,k}^{(i+1)}\frac{\kappa_2(\lambda_k)}{s_2^2(\lambda_k)} + \mathrm{E}^{(i)}[\log \pi_2(\lambda_k;\nu)] + c_{\lambda_k} \tag{36}$$

where the parameters $\Delta_{1,k}^{(i+1)}$, $\bar{\Delta}_{1,k}^{(i+1)}$, $\Delta_{2,k}^{(i+1)}$, $\bar{\Delta}_{2,k}^{(i+1)}$, $\Delta_{3,k}^{(i+1)}$ and $\bar{\Delta}_{3,k}^{(i+1)}$ are, respectively, given by

$$\begin{cases} \Delta_{1,k}^{(i+1)} = \mathrm{tr}\left\{\mathbf{A}_k^{(i+1)}\mathrm{E}^{(i)}[\mathbf{Q}^{-1}]\right\} \\ \bar{\Delta}_{1,k}^{(i+1)} = \mathrm{tr}\left\{\mathbf{B}_k^{(i+1)}\mathrm{E}^{(i)}[\mathbf{R}^{-1}]\right\} \\ \Delta_{2,k}^{(i+1)} = \left(\mathbf{a}_k^{(i+1)}\right)^{\mathrm{T}}\mathrm{E}^{(i)}[\mathbf{Q}^{-1}]\mathrm{E}^{(i+1)}[\boldsymbol{\beta}_1] \\ \bar{\Delta}_{2,k}^{(i+1)} = \left(\mathbf{b}_k^{(i+1)}\right)^{\mathrm{T}}\mathrm{E}^{(i)}[\mathbf{R}^{-1}]\mathrm{E}^{(i+1)}[\boldsymbol{\beta}_2] \\ \Delta_{3,k}^{(i+1)} = \mathrm{tr}\left\{\mathrm{E}^{(i+1)}[\boldsymbol{\beta}_1\boldsymbol{\beta}_1^{\mathrm{T}}]\mathrm{E}^{(i)}[\mathbf{Q}^{-1}]\right\} \\ \bar{\Delta}_{3,k}^{(i+1)} = \mathrm{tr}\left\{\mathrm{E}^{(i+1)}[\boldsymbol{\beta}_2\boldsymbol{\beta}_2^{\mathrm{T}}]\mathrm{E}^{(i)}[\mathbf{R}^{-1}]\right\} \end{cases} \tag{37}$$

and $\mathbf{A}_k^{(i+1)}$ and $\mathbf{B}_k^{(i+1)}$ are, respectively, given by

$$\begin{cases} \mathbf{A}_k^{(i+1)} = \mathrm{E}^{(i+1)}[(\mathbf{x}_k - \mathbf{F}_k\mathbf{x}_{k-1})(\mathbf{x}_k - \mathbf{F}_k\mathbf{x}_{k-1})^{\mathrm{T}}] \\ \mathbf{B}_k^{(i+1)} = \mathrm{E}^{(i+1)}[(\mathbf{z}_k - \mathbf{H}_k\mathbf{x}_k)(\mathbf{z}_k - \mathbf{H}_k\mathbf{x}_k)^{\mathrm{T}}] \end{cases} \tag{38}$$

In this paper, the prior distributions $\pi_1(\xi_k;\omega)$ and $\pi_2(\lambda_k;\nu)$ are selected so that the following equations hold

$$\begin{cases} \mathrm{E}^{(i)}[\log \pi_1(\xi_k;\omega)] = \log \pi_1(\xi_k;\mathrm{E}^{(i)}[\omega]) \\ \mathrm{E}^{(i)}[\log \pi_2(\lambda_k;\nu)] = \log \pi_2(\lambda_k;\mathrm{E}^{(i)}[\nu]) \end{cases} \tag{39}$$

Substituting (39) in (35)-(36), $q^{(i+1)}(\xi_k)$ and $q^{(i+1)}(\lambda_k)$ can be rewritten as

$$\begin{cases} q^{(i+1)}(\xi_k) = c_1c_2\exp(f_1(\xi_k))\pi_1(\xi_k;\mathrm{E}^{(i)}[\omega]) \\ q^{(i+1)}(\lambda_k) = \bar{c}_1\bar{c}_2\exp(f_2(\lambda_k))\pi_2(\lambda_k;\mathrm{E}^{(i)}[\nu]) \end{cases} \tag{40}$$

where c_1, c_2, $\bar{c}_1$, and $\bar{c}_2$ are normalizing constants, and nonlinear functions $f_1(\xi_k)$ and $f_2(\lambda_k)$ are given by

$$f_1(\xi_k) = 0.5n \log \kappa_1(\xi_k) - 0.5\Delta_{1,k}^{(i+1)}\kappa_1(\xi_k) + \Delta_{2,k}^{(i+1)}\frac{\kappa_1(\xi_k)}{s_1(\xi_k)} - 0.5\Delta_{3,k}^{(i+1)}\frac{\kappa_1(\xi_k)}{s_1^2(\xi_k)} \tag{41}$$

$$f_2(\lambda_k) = 0.5m \log \kappa_2(\lambda_k) - 0.5\bar{\Delta}_{1,k}^{(i+1)}\kappa_2(\lambda_k) + \bar{\Delta}_{2,k}^{(i+1)}\frac{\kappa_2(\lambda_k)}{s_2(\lambda_k)} - 0.5\bar{\Delta}_{3,k}^{(i+1)}\frac{\kappa_2(\lambda_k)}{s_2^2(\lambda_k)} \tag{42}$$

Define the likelihood PDFs $l_1(\xi_k)$ and $l_2(\lambda_k)$ as follows

$$l_1(\xi_k) = c_2 \exp(f_1(\xi_k)), \qquad l_2(\lambda_k) = \bar{c}_2 \exp(f_2(\lambda_k)) \tag{43}$$

Employing (43) in (40) yields

$$\begin{cases} q^{(i+1)}(\xi_k) = c_1 l_1(\xi_k)\pi_1(\xi_k; \mathrm{E}^{(i)}[\omega]) \\ q^{(i+1)}(\lambda_k) = \bar{c}_1 l_2(\lambda_k)\pi_2(\lambda_k; \mathrm{E}^{(i)}[\nu]) \end{cases} \tag{44}$$

It is observed from (44) that $q^{(i+1)}(\xi_k)$ can be deemed as a posterior PDF of ξ_k with prior PDF $\pi_1(\xi_k; \mathrm{E}^{(i)}[\omega])$ and likelihood PDF $l_1(\xi_k)$, and $q^{(i+1)}(\lambda_k)$ can be deemed as a posterior PDF of λ_k with prior PDF $\pi_2(\lambda_k; \mathrm{E}^{(i)}[\nu])$ and likelihood PDF $l_2(\lambda_k)$. Thus, the Monte Carlo approach can be used to obtain the approximations of posterior PDFs $q^{(i+1)}(\xi_k)$ and $q^{(i+1)}(\lambda_k)$.

Draw M random samples $\left\{\xi_k^j\right\}_{j=1}^M$ and $\left\{\lambda_k^j\right\}_{j=1}^M$ from the prior PDFs $\pi_1(\xi_k; \mathrm{E}^{(i)}[\omega])$ and $\pi_2(\lambda_k; \mathrm{E}^{(i)}[\nu])$, respectively, and the prior PDFs can be, respectively, approximated as

$$\begin{cases} \pi_1(\xi_k; \mathrm{E}^{(i)}[\omega]) \approx \frac{1}{M}\sum\limits_{j=1}^{M}\delta(\xi_k - \xi_k^j) \\ \pi_2(\lambda_k; \mathrm{E}^{(i)}[\nu]) \approx \frac{1}{M}\sum\limits_{j=1}^{M}\delta(\lambda_k - \lambda_k^j) \end{cases} \tag{45}$$

where M denotes the number of random samples.

Utilizing (43)-(45), the posterior PDFs $q^{(i+1)}(\xi_k)$ and $q^{(i+1)}(\lambda_k)$ can be, respectively, approximated by a set of weighted particles, i.e.,

$$\begin{cases} q^{(i+1)}(\xi_k) \approx \sum\limits_{j=1}^{M} w_{\xi,k}^{(i+1)j}\delta(\xi_k - \xi_k^j) \\ q^{(i+1)}(\lambda_k) \approx \sum\limits_{j=1}^{M} w_{\lambda,k}^{(i+1)j}\delta(\lambda_k - \lambda_k^j) \end{cases} \tag{46}$$

where the weights $w_{\xi,k}^{(i+1)j}$ and $w_{\lambda,k}^{(i+1)j}$ are, respectively, given by

$$\begin{cases} w_{\xi,k}^{(i+1)j} = \exp(f_1(\xi_k^j)) \Big/ \sum\limits_{l=1}^{M}\exp(f_1(\xi_k^l)) \\ w_{\lambda,k}^{(i+1)j} = \exp(f_2(\lambda_k^j)) \Big/ \sum\limits_{l=1}^{M}\exp(f_2(\lambda_k^l)) \end{cases} \tag{47}$$

Let $\theta = \mathbf{Q}$ and exploiting (26) in (25), $q^{(i+1)}(\mathbf{Q})$ is updated as an inverse-Wishart PDF, and let $\theta = \mathbf{R}$ and utilizing (26) in (25), $q^{(i+1)}(\mathbf{R})$ is updated as an inverse-Wishart PDF, i.e.,

$$\begin{cases} q^{(i+1)}(\mathbf{Q}) = \mathrm{IW}(\mathbf{Q}; t^{(i+1)}, \mathbf{T}^{(i+1)}) \\ q^{(i+1)}(\mathbf{R}) = \mathrm{IW}(\mathbf{R}; u^{(i+1)}, \mathbf{U}^{(i+1)}) \end{cases} \tag{48}$$

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

$$\begin{cases} t^{(i+1)} = T - n - 1, & \mathbf{T}^{(i+1)} = \mathbf{E}^{(i+1)} \\ u^{(i+1)} = T - m - 1, & \mathbf{U}^{(i+1)} = \mathbf{F}^{(i+1)} \end{cases} \tag{49}$$

where $\mathbf{E}^{(i+1)}$ and $\mathbf{F}^{(i+1)}$ are, respectively, calculated as

$$\begin{aligned} \mathbf{E}^{(i+1)} = \sum_{k=1}^{T} \Bigg\{ & \mathrm{E}^{(i+1)}[\kappa_1(\xi_k)]\mathbf{A}_k^{(i+1)} - \mathrm{E}^{(i+1)}[\frac{\kappa_1(\xi_k)}{s_1(\xi_k)}]\mathrm{E}^{(i+1)}[\boldsymbol{\beta}_1] \left(\mathbf{a}_k^{(i+1)}\right)^{\mathrm{T}} - \\ & \mathrm{E}^{(i+1)}[\frac{\kappa_1(\xi_k)}{s_1(\xi_k)}]\mathbf{a}_k^{(i+1)}\mathrm{E}^{(i+1)}[\boldsymbol{\beta}_1]^{\mathrm{T}} + \mathrm{E}^{(i+1)}[\frac{\kappa_1(\xi_k)}{s_1^2(\xi_k)}]\mathrm{E}^{(i+1)}[\boldsymbol{\beta}_1\boldsymbol{\beta}_1^{\mathrm{T}}] \Bigg\} \qquad \text{s.t. } \mathbf{E}^{i+1} > \mathbf{0} \end{aligned} \tag{50}$$

$$\begin{aligned} \mathbf{F}^{(i+1)} = \sum_{k=1}^{T} \Bigg\{ & \mathrm{E}^{(i+1)}[\kappa_2(\lambda_k)]\mathbf{B}_k^{(i+1)} - \mathrm{E}^{(i+1)}[\frac{\kappa_2(\lambda_k)}{s_2(\lambda_k)}]\mathrm{E}^{(i+1)}[\boldsymbol{\beta}_2] \left(\mathbf{b}_k^{(i+1)}\right)^{\mathrm{T}} - \\ & \mathrm{E}^{(i+1)}[\frac{\kappa_2(\lambda_k)}{s_2(\lambda_k)}]\mathbf{b}_k^{(i+1)}\mathrm{E}^{(i+1)}[\boldsymbol{\beta}_2]^{\mathrm{T}} + \mathrm{E}^{(i+1)}[\frac{\kappa_2(\lambda_k)}{s_2^2(\lambda_k)}]\mathrm{E}^{(i+1)}[\boldsymbol{\beta}_2\boldsymbol{\beta}_2^{\mathrm{T}}] \Bigg\} \qquad \text{s.t. } \mathbf{F}^{i+1} > \mathbf{0} \end{aligned} \tag{51}$$

where the derivations of (48)-(51) are given in Appendix C.

Let $\theta = \omega$ and $\theta = \nu$, respectively, and utilizing (26) in (25) results in

$$\log q^{(i+1)}(\omega) = \sum_{k=1}^{T} \mathrm{E}^{(i+1)}[\log \pi_1(\xi_k; \omega)] + c_\omega \tag{52}$$

$$\log q^{(i+1)}(\nu) = \sum_{k=1}^{T} \mathrm{E}^{(i+1)}[\log \pi_2(\lambda_k; \nu)] + c_\nu \tag{53}$$

In the paper, the prior distributions $\pi_1(\xi_k; \omega)$ and $\pi_2(\lambda_k; \nu)$ are chosen such that the posterior PDFs can be updated as Gamma, i.e.,

$$\begin{cases} q^{(i+1)}(\omega) = \mathrm{G}(\omega; c^{(i+1)}, d^{(i+1)}) \\ q^{(i+1)}(\nu) = \mathrm{G}(\nu; a^{(i+1)}, b^{(i+1)}) \end{cases} \tag{54}$$

where the shape parameters $c^{(i+1)}$ and $a^{(i+1)}$ and rate parameters $d^{(i+1)}$ and $b^{(i+1)}$ will be given in the next section.

The smoothing estimates of the state trajectory, mixing parameters and unknown distribution parameters are approximated as

$$\begin{cases} \hat{\mathbf{x}}_{k|T} \approx \hat{\mathbf{x}}_{k|T}^{(N)}, & \mathbf{P}_{k|T} = \mathbf{P}_{k|T}^{(N)}, \quad 0 \le k \le T \\ \hat{\xi}_{k|T} \approx \sum_{j=1}^{M} w_{\xi,k}^{(N)j} \xi_k^j, & 1 \le k \le T \\ \hat{\lambda}_{k|T} \approx \sum_{j=1}^{M} w_{\lambda,k}^{(N)j} \lambda_k^j, & 1 \le k \le T \\ \hat{\boldsymbol{\beta}}_1 \approx \hat{\boldsymbol{\beta}}_1^{(N)}, & \hat{\boldsymbol{\beta}}_2 = \hat{\boldsymbol{\beta}}_2^{(N)} \\ \hat{\mathbf{Q}} \approx \mathbf{T}^{(N)}/t^{(N)}, & \hat{\mathbf{R}} \approx \mathbf{U}^{(N)}/u^{(N)} \\ \hat{\omega} \approx c^{(N)}/d^{(N)}, & \hat{\nu} \approx a^{(N)}/b^{(N)} \end{cases} \tag{55}$$

where N is the minimum number of iterations to guarantee convergence.

2) Calculation of expectations: Using (46), the expectations of nonlinear functions of mixing parameters can be approximated by the Monte Carlo approach as follows

$$\begin{cases} \mathrm{E}^{(i+1)}[\kappa_1(\xi_k)] \approx \sum_{j=1}^{M} w_{\xi,k}^{(i+1)j} \kappa_1(\xi_k^j) \\ \mathrm{E}^{(i+1)}[\frac{\kappa_1(\xi_k)}{s_1(\xi_k)}] \approx \sum_{j=1}^{M} w_{\xi,k}^{(i+1)j} \frac{\kappa_1(\xi_k^j)}{s_1(\xi_k^j)} \\ \mathrm{E}^{(i+1)}[\frac{\kappa_1(\xi_k)}{s_1^2(\xi_k)}] \approx \sum_{j=1}^{M} w_{\xi,k}^{(i+1)j} \frac{\kappa_1(\xi_k^j)}{s_1^2(\xi_k^j)} \end{cases} \tag{56}$$

$$\begin{cases} \mathrm{E}^{(i+1)}[\kappa_1(\lambda_k)] \approx \sum_{j=1}^{M} w_{\lambda,k}^{(i+1)j} \kappa_1(\lambda_k^j) \\ \mathrm{E}^{(i+1)}[\frac{\kappa_2(\lambda_k)}{s_2(\lambda_k)}] \approx \sum_{j=1}^{M} w_{\lambda,k}^{(i+1)j} \frac{\kappa_2(\lambda_k^j)}{s_2(\lambda_k^j)} \\ \mathrm{E}^{(i+1)}[\frac{\kappa_2(\lambda_k)}{s_2^2(\lambda_k)}] \approx \sum_{j=1}^{M} w_{\lambda,k}^{(i+1)j} \frac{\kappa_2(\lambda_k^j)}{s_2^2(\lambda_k^j)} \end{cases} \tag{57}$$

Employing (29), the first and second order moments of the shape parameters are calculated as follows

$$\begin{cases} \mathrm{E}^{(i+1)}[\boldsymbol{\beta}_1] = \hat{\boldsymbol{\beta}}_1^{(i+1)}, & \mathrm{E}^{(i+1)}[\boldsymbol{\beta}_2] = \hat{\boldsymbol{\beta}}_2^{(i+1)} \\ \mathrm{E}^{(i+1)}[\boldsymbol{\beta}_1 \boldsymbol{\beta}_1^{\mathrm{T}}] = \mathbf{P}_{\boldsymbol{\beta}_1}^{(i+1)} + \hat{\boldsymbol{\beta}}_1^{(i+1)} \left(\hat{\boldsymbol{\beta}}_1^{(i+1)}\right)^{\mathrm{T}} \\ \mathrm{E}^{(i+1)}[\boldsymbol{\beta}_2 \boldsymbol{\beta}_2^{\mathrm{T}}] = \mathbf{P}_{\boldsymbol{\beta}_2}^{(i+1)} + \hat{\boldsymbol{\beta}}_2^{(i+1)} \left(\hat{\boldsymbol{\beta}}_2^{(i+1)}\right)^{\mathrm{T}} \end{cases} \tag{58}$$

According to (48), the posterior PDFs of scale matrices $\mathbf{Q}^{-1}$ and $\mathbf{R}^{-1}$ are updated as Wishart, i.e.,

$$\begin{cases} q^{(i+1)}(\mathbf{Q}^{-1}) = \mathrm{W}(\mathbf{Q}^{-1}; t^{(i+1)}, \left(\mathbf{T}^{(i+1)}\right)^{-1}) \\ q^{(i+1)}(\mathbf{R}^{-1}) = \mathrm{W}(\mathbf{R}^{-1}; u^{(i+1)}, \left(\mathbf{U}^{(i+1)}\right)^{-1}) \end{cases} \tag{59}$$

Utilizing (59), the expectations of scale matrices $\mathbf{Q}^{-1}$ and $\mathbf{R}^{-1}$ are calculated as

$$\begin{cases} \mathrm{E}^{(i+1)}[\mathbf{Q}^{-1}] = t^{(i+1)} \left(\mathbf{T}^{(i+1)}\right)^{-1} \\ \mathrm{E}^{(i+1)}[\mathbf{R}^{-1}] = u^{(i+1)} \left(\mathbf{U}^{(i+1)}\right)^{-1} \end{cases} \tag{60}$$

Exploiting (54) yields

$$\mathrm{E}^{(i)}[\omega] = c^{(i+1)}/d^{(i+1)}, \qquad \mathrm{E}^{(i)}[\nu] = a^{(i+1)}/b^{(i+1)} \tag{61}$$

Substituting (27) in (32) and (38), $\mathbf{a}_k^{(i+1)}$, $\mathbf{b}_k^{(i+1)}$, $\mathbf{A}_k^{(i+1)}$ and $\mathbf{B}_k^{(i+1)}$ are, respectively, calculated as follows

$$\begin{cases} \mathbf{a}_k^{(i+1)} = \hat{\mathbf{x}}_{k|T}^{(i+1)} - \mathbf{F}_k \hat{\mathbf{x}}_{k-1|T}^{(i+1)}, \qquad \mathbf{b}_k^{(i+1)} = \mathbf{z}_k - \mathbf{H}_k \hat{\mathbf{x}}_{k|T}^{(i+1)} \\ \mathbf{B}_k^{(i+1)} = (\mathbf{z}_k - \mathbf{H}_k \hat{\mathbf{x}}_{k|T}^{(i+1)})(\mathbf{z}_k - \mathbf{H}_k \hat{\mathbf{x}}_{k|T}^{(i+1)})^{\mathrm{T}} + \mathbf{H}_k \mathbf{P}_{k|T}^{(i+1)} \mathbf{H}_k^{\mathrm{T}} \\ \mathbf{A}_k^{(i+1)} = (\hat{\mathbf{x}}_{k|T}^{(i+1)} - \mathbf{F}_k \hat{\mathbf{x}}_{k-1|T}^{(i+1)})(\hat{\mathbf{x}}_{k|T}^{(i+1)} - \mathbf{F}_k \hat{\mathbf{x}}_{k-1|T}^{(i+1)})^{\mathrm{T}} + \\ \mathbf{P}_{k|T}^{(i+1)} - \left(\mathbf{F}_k \mathbf{G}_{k-1}^{(i+1)} \mathbf{P}_{k|T}^{(i+1)}\right)^{\mathrm{T}} - \mathbf{F}_k \mathbf{G}_{k-1}^{(i+1)} \mathbf{P}_{k|T}^{(i+1)} + \mathbf{F}_k \mathbf{P}_{k-1|T}^{(i+1)} \mathbf{F}_k^{\mathrm{T}} \end{cases} \tag{62}$$

where $\mathbf{G}_{k-1}^{(i+1)}$ denotes the RTS smoothing gain at the $i+1$th iteration.

The proposed robust RTS smoothing framework is composed of variational approximations of posterior PDFs in (27)-(55) and calculations of expectations in (56)-(62). To implement the proposed robust RTS smoothing framework, the skew functions $s_1(\cdot)$ and $s_2(\cdot)$, the scale functions $\kappa_1(\cdot)$ and $\kappa_2(\cdot)$, and the mixing densities $\pi_1(\cdot;\cdot)$ and $\pi_2(\cdot;\cdot)$ require to be firstly selected. Next, to illustrate how to implement the proposed robust RTS smoothing framework, several special solutions based on the exemplary symmetric and skew GGScM distributions will be derived.

V. Robust RTS Smoothers Based on Exemplary GGScM Distributions

In this section, we will derive several particular solutions when the GGM, GEM, GBM, GIGM, GIEM, SGGM, SGEM, SGBM, SGIGM and SGIEM distributions are employed to model state and measurement noises.

A. Updates of $q^{(i+1)}(\mathbf{x}_{0:T})$, $q^{(i+1)}(\boldsymbol{\beta}_1)$, $q^{(i+1)}(\boldsymbol{\beta}_2)$, $q^{(i+1)}(\mathbf{Q})$ and $q^{(i+1)}(\mathbf{R})$

1) Exemplary symmetric GGScM distributions: It is observed from Table I that, for the exemplary symmetric GGScM distributions, $\boldsymbol{\beta}_1 = \mathbf{0}$, $\boldsymbol{\beta}_2 = \mathbf{0}$, $\kappa_1(\xi_k) = \xi_k$ and $\kappa_2(\lambda_k) = \lambda_k$. Then, the exemplary symmetric GGScM distributions have the same update forms for posterior PDFs $q^{(i+1)}(\mathbf{x}_{0:T})$, $q^{(i+1)}(\boldsymbol{\beta}_1)$, $q^{(i+1)}(\boldsymbol{\beta}_2)$, $q^{(i+1)}(\mathbf{Q})$ and $q^{(i+1)}(\mathbf{R})$. Since $\boldsymbol{\beta}_1 = \mathbf{0}$ and $\boldsymbol{\beta}_2 = \mathbf{0}$,

for the exemplary symmetric GGScM distributions, both the estimates of shape parameters and corresponding estimation error covariance matrices are zeros, i.e.,

$$\begin{cases} \hat{\boldsymbol{\beta}}_1^{(i+1)} = \mathbf{0}, & \mathbf{P}_{\boldsymbol{\beta}_1}^{(i+1)} = \mathbf{0} \\ \hat{\boldsymbol{\beta}}_2^{(i+1)} = \mathbf{0}, & \mathbf{P}_{\boldsymbol{\beta}_2}^{(i+1)} = \mathbf{0} \end{cases} \tag{63}$$

Substituting (63) in (58) gives

$$\begin{cases} \mathrm{E}^{(i+1)}[\boldsymbol{\beta}_1] = \mathbf{0}, & \mathrm{E}^{(i+1)}[\boldsymbol{\beta}_1\boldsymbol{\beta}_1^{\mathrm{T}}] = \mathbf{0} \\ \mathrm{E}^{(i+1)}[\boldsymbol{\beta}_2] = \mathbf{0}, & \mathrm{E}^{(i+1)}[\boldsymbol{\beta}_2\boldsymbol{\beta}_2^{\mathrm{T}}] = \mathbf{0} \end{cases} \tag{64}$$

Employing $\kappa_1(\xi_k) = \xi_k$, $\kappa_2(\lambda_k) = \lambda_k$ and (64) in (28) and (50)-(51) results in

$$\begin{cases} \tilde{\mathbf{q}}_k^{(i)} = \mathbf{0}, & \tilde{\mathbf{Q}}_k^{(i)} = \frac{\left\{\mathrm{E}^{(i)}[\mathbf{Q}^{-1}]\right\}^{-1}}{\mathrm{E}^{(i)}[\xi_k]} \\ \tilde{\mathbf{r}}_k^{(i)} = \mathbf{0}, & \tilde{\mathbf{R}}_k^{(i)} = \frac{\left\{\mathrm{E}^{(i)}[\mathbf{R}^{-1}]\right\}^{-1}}{\mathrm{E}^{(i)}[\lambda_k]} \end{cases} \tag{65}$$

$$\begin{cases} \mathbf{E}^{(i+1)} = \sum\limits_{k=1}^{T} \mathrm{E}^{(i+1)}[\xi_k]\mathbf{A}_k^{(i+1)} \\ \mathbf{F}^{(i+1)} = \sum\limits_{k=1}^{T} \mathrm{E}^{(i+1)}[\lambda_k]\mathbf{B}_k^{(i+1)} \end{cases} \tag{66}$$

For the exemplary symmetric GGScM distributions, the posterior PDFs $q^{(i+1)}(\mathbf{x}_{0:T})$, $q^{(i+1)}(\boldsymbol{\beta}_1)$, $q^{(i+1)}(\boldsymbol{\beta}_2)$, $q^{(i+1)}(\mathbf{Q})$ and $q^{(i+1)}(\mathbf{R})$ are, respectively, updated as Gaussian, Gaussian and inverse-Wishart distributions by (27), (29) and (48), where the required parameters are given (49), (63) and (65)-(66).

2) Exemplary skew GGScM distributions: It is seen from Table II that, for the exemplary skew GGScM distributions, $\boldsymbol{\beta}_1 \neq \mathbf{0}$, $\boldsymbol{\beta}_2 \neq \mathbf{0}$, $s_1(\xi_k) = \kappa_1(\xi_k) = \xi_k$ and $s_2(\lambda_k) = \kappa_2(\lambda_k) = \lambda_k$. Then, the exemplary skew GGScM distributions also have the same update forms for posterior PDFs $q^{(i+1)}(\mathbf{x}_{0:T})$, $q^{(i+1)}(\boldsymbol{\beta}_1)$, $q^{(i+1)}(\boldsymbol{\beta}_2)$, $q^{(i+1)}(\mathbf{Q})$ and $q^{(i+1)}(\mathbf{R})$. Exploiting $s_1(\xi_k) = \kappa_1(\xi_k) = \xi_k$ and $s_2(\lambda_k) = \kappa_2(\lambda_k) = \lambda_k$ in (28), (30)-(31) and (50)-(51) yields

$$\begin{cases} \tilde{\mathbf{q}}_k^{(i)} = \frac{\mathrm{E}^{(i)}[\boldsymbol{\beta}_1]}{\mathrm{E}^{(i)}[\xi_k]}, & \tilde{\mathbf{Q}}_k^{(i)} = \frac{\left\{\mathrm{E}^{(i)}[\mathbf{Q}^{-1}]\right\}^{-1}}{\mathrm{E}^{(i)}[\xi_k]} \\ \tilde{\mathbf{r}}_k^{(i)} = \frac{\mathrm{E}^{(i)}[\boldsymbol{\beta}_2]}{\mathrm{E}^{(i)}[\lambda_k]}, & \tilde{\mathbf{R}}_k^{(i)} = \frac{\left\{\mathrm{E}^{(i)}[\mathbf{R}^{-1}]\right\}^{-1}}{\mathrm{E}^{(i)}[\lambda_k]} \end{cases} \tag{67}$$

$$\begin{cases} \hat{\boldsymbol{\beta}}_1^{(i+1)} = \frac{\sum\limits_{k=1}^{T} \mathbf{a}_k^{(i+1)}}{\sum\limits_{k=1}^{T} \mathrm{E}^{(i+1)}[\frac{1}{\xi_k}]}, & \hat{\boldsymbol{\beta}}_2^{(i+1)} = \frac{\sum\limits_{k=1}^{T} \mathbf{b}_k^{(i+1)}}{\sum\limits_{k=1}^{T} \mathrm{E}^{(i+1)}[\frac{1}{\lambda_k}]} \\ \mathbf{P}_{\boldsymbol{\beta}_1}^{(i+1)} = \left\{\mathrm{E}^{(i)}[\mathbf{Q}^{-1}] \sum\limits_{k=1}^{T} \mathrm{E}^{(i+1)}[\frac{1}{\xi_k}]\right\}^{-1} \\ \mathbf{P}_{\boldsymbol{\beta}_2}^{(i+1)} = \left\{\mathrm{E}^{(i)}[\mathbf{R}^{-1}] \sum\limits_{k=1}^{T} \mathrm{E}^{(i+1)}[\frac{1}{\lambda_k}]\right\}^{-1} \end{cases} \tag{68}$$

$$\begin{cases} \mathbf{E}^{(i+1)} = \sum\limits_{k=1}^{T} \Big\{ \mathrm{E}^{(i+1)}[\xi_k]\mathbf{A}_k^{(i+1)} - \mathrm{E}^{(i+1)}[\boldsymbol{\beta}_1] \left(\mathbf{a}_k^{(i+1)}\right)^{\mathrm{T}} - \\ \mathbf{a}_k^{(i+1)}\mathrm{E}^{(i+1)}[\boldsymbol{\beta}_1]^{\mathrm{T}} + \mathrm{E}^{(i+1)}[\frac{1}{\xi_k}]\mathrm{E}^{(i+1)}[\boldsymbol{\beta}_1\boldsymbol{\beta}_1^{\mathrm{T}}] \Big\} \\ \mathbf{F}^{(i+1)} = \sum\limits_{k=1}^{T} \Big\{ \mathrm{E}^{(i+1)}[\lambda_k]\mathbf{B}_k^{(i+1)} - \mathrm{E}^{(i+1)}[\boldsymbol{\beta}_2] \left(\mathbf{b}_k^{(i+1)}\right)^{\mathrm{T}} - \\ \mathbf{b}_k^{(i+1)}\mathrm{E}^{(i+1)}[\boldsymbol{\beta}_2]^{\mathrm{T}} + \mathrm{E}^{(i+1)}[\frac{1}{\lambda_k}]\mathrm{E}^{(i+1)}[\boldsymbol{\beta}_2\boldsymbol{\beta}_2^{\mathrm{T}}] \Big\} \end{cases} \tag{69}$$

For the exemplary skew GGScM distributions, the posterior PDFs $q^{(i+1)}(\mathbf{x}_{0:T})$, $q^{(i+1)}(\boldsymbol{\beta}_1)$, $q^{(i+1)}(\boldsymbol{\beta}_2)$, $q^{(i+1)}(\mathbf{Q})$ and $q^{(i+1)}(\mathbf{R})$ are, respectively, updated as Gaussian, Gaussian and inverse-Wishart distributions by (27), (29) and (48), where the required parameters are given (49) and (67)-(69).

B. Analytical updates of $q^{(i+1)}(\xi_k)$ and $q^{(i+1)}(\lambda_k)$

For the exemplary symmetric GGScM distributions, substituting (64) in (37) results in

$$\Delta_{2,k}^{(i+1)} = \bar{\Delta}_{2,k}^{(i+1)} = \Delta_{3,k}^{(i+1)} = \bar{\Delta}_{3,k}^{(i+1)} = 0 \tag{70}$$

As an example, next, we derive the analytical updates of posterior PDFs $q^{(i+1)}(\xi_k)$ and $q^{(i+1)}(\lambda_k)$ when state and measurement noises are modelled by the GIGM distribution. It is observed from Table I that, for the GIGM distribution, $\pi_1(\xi_k;\omega) = \mathrm{IG}(\xi_k; \frac{\omega}{2}, \frac{\omega}{2})$, $\pi_2(\lambda_k;\nu) = \mathrm{IG}(\lambda_k; \frac{\nu}{2}, \frac{\nu}{2})$, $\kappa_1(\xi_k) = \xi_k$, $\kappa_2(\lambda_k) = \lambda_k$, $\xi_k > 0$ and $\lambda_k > 0$. Substituting prior distributions and scale functions in (35)-(36) and using (70) yields

$$\log q^{(i+1)}(\xi_k) = \big(0.5n - 0.5\mathrm{E}^{(i)}[\omega] - 1\big)\log\xi_k - 0.5\Delta_{1,k}^{(i+1)}\xi_k - 0.5\mathrm{E}^{(i)}[\omega]/\xi_k + c_{\xi_k} \tag{71}$$

$$\log q^{(i+1)}(\lambda_k) = \big(0.5m - 0.5\mathrm{E}^{(i)}[\nu] - 1\big)\log\lambda_k - 0.5\bar{\Delta}_{1,k}^{(i+1)}\lambda_k - 0.5\mathrm{E}^{(i)}[\nu]/\lambda_k + c_{\lambda_k} \tag{72}$$

Exploiting (71)-(72), $q^{(i+1)}(\xi_k)$ and $q^{(i+1)}(\lambda_k)$ are updated as generalized inverse Gaussian PDFs, i.e.,

$$\begin{cases} q^{(i+1)}(\xi_k) = \mathrm{GIG}(\xi_k; \alpha_k^{(i+1)}, \beta_k^{(i+1)}, \rho_k^{(i+1)}) \\ q^{(i+1)}(\lambda_k) = \mathrm{GIG}(\lambda_k; \eta_k^{(i+1)}, \varphi_k^{(i+1)}, \varrho_k^{(i+1)}) \end{cases} \tag{73}$$

where shape parameters are given by

$$\begin{cases} \alpha_k^{(i+1)} = \Delta_{1,k}^{(i+1)}, & \beta_k^{(i+1)} = \mathrm{E}^{(i)}[\omega] \\ \rho_k^{(i+1)} = 0.5n - 0.5\mathrm{E}^{(i)}[\omega] & \\ \eta_k^{(i+1)} = \bar{\Delta}_{1,k}^{(i+1)}, & \varphi_k^{(i+1)} = \mathrm{E}^{(i)}[\nu] \\ \varrho_k^{(i+1)} = 0.5m - 0.5\mathrm{E}^{(i)}[\nu] & \end{cases} \tag{74}$$

Similarly, the analytical updates of posterior PDFs $q^{(i+1)}(\xi_k)$ and $q^{(i+1)}(\lambda_k)$ for other exemplary GGScM distributions can be also obtained, and the results are given in Table III, where c_3, $\bar{c}_3$, c_4 and $\bar{c}_4$ are all normalizing constants.

For the GGM and GEM distributions, the posterior PDFs $q^{(i+1)}(\xi_k)$ and $q^{(i+1)}(\lambda_k)$ of the mixing parameters are updated as Gamma PDFs, and then the required expectations of the mixing parameters are analytically calculated as [6]

$$\begin{cases} \mathrm{E}^{(i+1)}[\xi_k] = \alpha_k^{(i+1)}/\beta_k^{(i+1)} \\ \mathrm{E}^{(i+1)}[\log \xi_k] = \psi(\alpha_k^{(i+1)}) - \log \beta_k^{(i+1)} \\ \mathrm{E}^{(i+1)}[\lambda_k] = \eta_k^{(i+1)}/\varphi_k^{(i+1)} \\ \mathrm{E}^{(i+1)}[\log \lambda_k] = \psi(\eta_k^{(i+1)}) - \log \varphi_k^{(i+1)} \end{cases} \tag{75}$$

For the GBM distribution, the posterior PDFs $q^{(i+1)}(\xi_k)$ and $q^{(i+1)}(\lambda_k)$ of the mixing parameters are updated as truncated Gamma PDFs, and then the required expectations of the mixing parameters are calculated as

$$\begin{cases} \mathrm{E}^{(i+1)}[\xi_k] = c_3 \int_0^1 \xi_k \mathrm{G}(\xi_k; \alpha_k^{(i+1)}, \beta_k^{(i+1)}) d\xi_k \\ \mathrm{E}^{(i+1)}[\log \xi_k] = c_3 \int_0^1 \log \xi_k \mathrm{G}(\xi_k; \alpha_k^{(i+1)}, \beta_k^{(i+1)}) d\xi_k \\ \mathrm{E}^{(i+1)}[\lambda_k] = \bar{c}_3 \int_0^1 \lambda_k \mathrm{G}(\lambda_k; \eta_k^{(i+1)}, \varphi_k^{(i+1)}) d\lambda_k \\ \mathrm{E}^{(i+1)}[\log \lambda_k] = \bar{c}_3 \int_0^1 \log \lambda_k \mathrm{G}(\lambda_k; \eta_k^{(i+1)}, \varphi_k^{(i+1)}) d\lambda_k \end{cases} \tag{76}$$

For the GIGM, GIEM, SGGM, SGEM, SGIGM and SGIEM distributions, the posterior PDFs $q^{(i+1)}(\xi_k)$ and $q^{(i+1)}(\lambda_k)$ of the mixing parameters are updated as GIG PDFs, and then the required expectations of the mixing parameters are analytically calculated as [35]

$$\begin{cases} \mathrm{E}^{(i+1)}[\xi_k] = \dfrac{\sqrt{\beta_k^{(i+1)}} K_{\rho_k^{(i+1)}+1}(\sqrt{\alpha_k^{(i+1)}\beta_k^{(i+1)}})}{\sqrt{\alpha_k^{(i+1)}} K_{\rho_k^{(i+1)}}(\sqrt{\alpha_k^{(i+1)}\beta_k^{(i+1)}})} \\ \mathrm{E}^{(i+1)}[\frac{1}{\xi_k}] = \dfrac{\sqrt{\alpha_k^{(i+1)}} K_{\rho_k^{(i+1)}+1}(\sqrt{\alpha_k^{(i+1)}\beta_k^{(i+1)}})}{\sqrt{\beta_k^{(i+1)}} K_{\rho_k^{(i+1)}}(\sqrt{\alpha_k^{(i+1)}\beta_k^{(i+1)}})} - \dfrac{2\rho_k^{(i+1)}}{\beta_k^{(i+1)}} \\ \mathrm{E}^{(i+1)}[\log \xi_k] = \log \dfrac{\sqrt{\beta_k^{(i+1)}}}{\sqrt{\alpha_k^{(i+1)}}} + \dfrac{\partial \log K_\rho(\sqrt{\alpha_k^{(i+1)}\beta_k^{(i+1)}})}{\partial \rho}\Big|_{\rho=\rho_k^{(i+1)}} \\ \mathrm{E}^{(i+1)}[\lambda_k] = \dfrac{\sqrt{\varphi_k^{(i+1)}} K_{\varrho_k^{(i+1)}+1}(\sqrt{\eta_k^{(i+1)}\varphi_k^{(i+1)}})}{\sqrt{\eta_k^{(i+1)}} K_{\varrho_k^{(i+1)}}(\sqrt{\eta_k^{(i+1)}\varphi_k^{(i+1)}})} \\ \mathrm{E}^{(i+1)}[\frac{1}{\lambda_k}] = \dfrac{\sqrt{\eta_k^{(i+1)}} K_{\varrho_k^{(i+1)}+1}(\sqrt{\eta_k^{(i+1)}\varphi_k^{(i+1)}})}{\sqrt{\varphi_k^{(i+1)}} K_{\varrho_k^{(i+1)}}(\sqrt{\eta_k^{(i+1)}\varphi_k^{(i+1)}})} - \dfrac{2\varrho_k^{(i+1)}}{\varphi_k^{(i+1)}} \\ \mathrm{E}^{(i+1)}[\log \lambda_k] = \log \dfrac{\sqrt{\varphi_k^{(i+1)}}}{\sqrt{\eta_k^{(i+1)}}} + \dfrac{\partial \log K_\rho(\sqrt{\eta_k^{(i+1)}\varphi_k^{(i+1)}})}{\partial \rho}\Big|_{\rho=\varrho_k^{(i+1)}} \end{cases} \tag{77}$$

where the partial derivatives in (77) are calculated using a numerical approach.

For the SGBM distribution, the posterior PDFs $q^{(i+1)}(\xi_k)$ and $q^{(i+1)}(\lambda_k)$ of the mixing parameters are updated as truncated GIG PDFs, and then the required expectations of the mixing parameters are calculated as

$$\begin{cases} \mathrm{E}^{(i+1)}[\xi_k] = c_4 \int_0^1 \xi_k \mathrm{GIG}(\xi_k; \alpha_k^{(i+1)}, \beta_k^{(i+1)}, \rho_k^{(i+1)}) d\xi_k \\ \mathrm{E}^{(i+1)}[\frac{1}{\xi_k}] = c_4 \int_0^1 \frac{1}{\xi_k} \mathrm{GIG}(\xi_k; \alpha_k^{(i+1)}, \beta_k^{(i+1)}, \rho_k^{(i+1)}) d\xi_k \\ \mathrm{E}^{(i+1)}[\log \xi_k] = c_4 \int_0^1 \log \xi_k \mathrm{GIG}(\xi_k; \alpha_k^{(i+1)}, \beta_k^{(i+1)}, \rho_k^{(i+1)}) d\xi_k \\ \mathrm{E}^{(i+1)}[\lambda_k] = \bar{c}_4 \int_0^1 \lambda_k \mathrm{GIG}(\lambda_k; \eta_k^{(i+1)}, \varphi_k^{(i+1)}, \varrho_k^{(i+1)}) d\lambda_k \\ \mathrm{E}^{(i+1)}[\frac{1}{\lambda_k}] = \bar{c}_4 \int_0^1 \frac{1}{\lambda_k} \mathrm{GIG}(\lambda_k; \eta_k^{(i+1)}, \varphi_k^{(i+1)}, \varrho_k^{(i+1)}) d\lambda_k \\ \mathrm{E}^{(i+1)}[\log \lambda_k] = \bar{c}_4 \int_0^1 \log \lambda_k \mathrm{GIG}(\lambda_k; \eta_k^{(i+1)}, \varphi_k^{(i+1)}, \varrho_k^{(i+1)}) d\lambda_k \end{cases} \tag{78}$$

C. Monte Carlo updates of $q^{(i+1)}(\xi_k)$ and $q^{(i+1)}(\lambda_k)$

For the exemplary symmetric GGScM distributions, employing $\kappa_1(\xi_k) = \xi_k$, $\kappa_2(\lambda_k) = \lambda_k$ and (70) in (41)-(42), we obtain

$$\begin{cases} f_1(\xi_k) = 0.5n \log \xi_k - 0.5\xi_k \Delta_{1,k}^{(i+1)} \\ f_2(\lambda_k) = 0.5m \log \lambda_k - 0.5\lambda_k \bar{\Delta}_{1,k}^{(i+1)} \end{cases} \tag{79}$$

For the exemplary skew GGScM distributions, exploiting $s_1(\xi_k) = \kappa_1(\xi_k) = \xi_k$ and $s_2(\lambda_k) = \kappa_2(\lambda_k) = \lambda_k$ in (41)-(42) yields

$$\begin{cases} f_1(\xi_k) = 0.5n \log \xi_k - 0.5\xi_k \Delta_{1,k}^{(i+1)} + \Delta_{2,k}^{(i+1)} - \frac{0.5\Delta_{3,k}^{(i+1)}}{\xi_k} \\ f_2(\lambda_k) = 0.5m \log \lambda_k - 0.5\lambda_k \bar{\Delta}_{1,k}^{(i+1)} + \bar{\Delta}_{2,k}^{(i+1)} - \frac{0.5\bar{\Delta}_{3,k}^{(i+1)}}{\lambda_k} \end{cases} \tag{80}$$

We can see from Tables I and II that, for the exemplary symmetric and skew GGScM distributions, the prior distribution can be, respectively, selected as Gamma, exponential, Beta, inverse Gamma and inverse exponential distributions. According to Appendix D, we can see that equation (39) holds when the prior distributions are, respectively, chosen as Gamma, exponential, Beta, inverse Gamma, and inverse exponential distributions. Then, $q^{(i+1)}(\xi_k)$ and $q^{(i+1)}(\lambda_k)$ can be updated by (46)-(47), where the samples of mixing samples $\left\{\xi_k^j\right\}_{j=1}^M$ and $\left\{\lambda_k^j\right\}_{j=1}^M$ are randomly drawn from the prior distributions with the dof parameters $\mathrm{E}^{(i)}[\omega]$ and $\mathrm{E}^{(i)}[\nu]$, and $f_1(\xi_k)$ and $f_2(\lambda_k)$ are given by (79) and (80) for the exemplary symmetric and skew GGScM distributions, respectively.

TABLE III: Analytical updates of posterior PDFs $q^{(i+1)}(\xi_k)$ and $q^{(i+1)}(\lambda_k)$ for exemplary GGScM distributions.

GGScM distributions	$q^{(i+1)}(\xi_k)$ and $q^{(i+1)}(\lambda_k)$	Shape parameters and rate parameters
GGM	$q^{(i+1)}(\xi_k) = \mathrm{G}(\xi_k; \alpha_k^{(i+1)}, \beta_k^{(i+1)})$	$\alpha_k^{(i+1)} = 0.5n + 0.5\mathrm{E}^{(i)}[\omega]$, $\beta_k^{(i+1)} = 0.5\Delta_{1,k}^{(i+1)} + 0.5\mathrm{E}^{(i)}[\omega]$
	$q^{(i+1)}(\lambda_k) = \mathrm{G}(\lambda_k; \eta_k^{(i+1)}, \varphi_k^{(i+1)})$	$\eta_k^{(i+1)} = 0.5m + 0.5\mathrm{E}^{(i)}[\nu]$, $\varphi_k^{(i+1)} = 0.5\bar{\Delta}_{1,k}^{(i+1)} + 0.5\mathrm{E}^{(i)}[\nu]$
GEM	$q^{(i+1)}(\xi_k) = \mathrm{G}(\xi_k; \alpha_k^{(i+1)}, \beta_k^{(i+1)})$	$\alpha_k^{(i+1)} = 0.5n + 1$, $\beta_k^{(i+1)} = 0.5\Delta_{1,k}^{(i+1)} + \mathrm{E}^{(i)}[\omega]$
	$q^{(i+1)}(\lambda_k) = \mathrm{G}(\lambda_k; \eta_k^{(i+1)}, \varphi_k^{(i+1)})$	$\eta_k^{(i+1)} = 0.5m + 1$, $\varphi_k^{(i+1)} = 0.5\bar{\Delta}_{1,k}^{(i+1)} + \mathrm{E}^{(i)}[\nu]$
GBM	$q^{(i+1)}(\xi_k) = c_3\mathrm{G}(\xi_k; \alpha_k^{(i+1)}, \beta_k^{(i+1)})$	$\alpha_k^{(i+1)} = 0.5n + \mathrm{E}^{(i)}[\omega]$, $\beta_k^{(i+1)} = 0.5\Delta_{1,k}^{(i+1)}$, $\xi_k \in (0,1)$
	$q^{(i+1)}(\lambda_k) = \bar{c}_3\mathrm{G}(\lambda_k; \eta_k^{(i+1)}, \varphi_k^{(i+1)})$	$\eta_k^{(i+1)} = 0.5m + \mathrm{E}^{(i)}[\nu]$, $\varphi_k^{(i+1)} = 0.5\bar{\Delta}_{1,k}^{(i+1)}$, $\lambda_k \in (0,1)$
GIGM	$q^{(i+1)}(\xi_k) = \mathrm{GIG}(\xi_k; \alpha_k^{(i+1)}, \beta_k^{(i+1)}, \rho_k^{(i+1)})$	$\alpha_k^{(i+1)} = \Delta_{1,k}^{(i+1)}$, $\beta_k^{(i+1)} = \mathrm{E}^{(i)}[\omega]$, $\rho_k^{(i+1)} = 0.5n - 0.5\mathrm{E}^{(i)}[\omega]$
	$q^{(i+1)}(\lambda_k) = \mathrm{GIG}(\lambda_k; \eta_k^{(i+1)}, \varphi_k^{(i+1)}, \varrho_k^{(i+1)})$	$\eta_k^{(i+1)} = \bar{\Delta}_{1,k}^{(i+1)}$, $\varphi_k^{(i+1)} = \mathrm{E}^{(i)}[\nu]$, $\varrho_k^{(i+1)} = 0.5m - 0.5\mathrm{E}^{(i)}[\nu]$
GIEM	$q^{(i+1)}(\xi_k) = \mathrm{GIG}(\xi_k; \alpha_k^{(i+1)}, \beta_k^{(i+1)}, \rho_k^{(i+1)})$	$\alpha_k^{(i+1)} = \Delta_{1,k}^{(i+1)}$, $\beta_k^{(i+1)} = 2\mathrm{E}^{(i)}[\omega]$, $\rho_k^{(i+1)} = 0.5n - 1$
	$q^{(i+1)}(\lambda_k) = \mathrm{GIG}(\lambda_k; \eta_k^{(i+1)}, \varphi_k^{(i+1)}, \varrho_k^{(i+1)})$	$\eta_k^{(i+1)} = \bar{\Delta}_{1,k}^{(i+1)}$, $\varphi_k^{(i+1)} = 2\mathrm{E}^{(i)}[\nu]$, $\varrho_k^{(i+1)} = 0.5m - 1$
SGGM	$q^{(i+1)}(\xi_k) = \mathrm{GIG}(\xi_k; \alpha_k^{(i+1)}, \beta_k^{(i+1)}, \rho_k^{(i+1)})$	$\alpha_k^{(i+1)} = \Delta_{1,k}^{(i+1)} + \mathrm{E}^{(i)}[\omega]$, $\beta_k^{(i+1)} = \Delta_{3,k}^{(i+1)}$, $\rho_k^{(i+1)} = 0.5(n + \mathrm{E}^{(i)}[\omega])$
	$q^{(i+1)}(\lambda_k) = \mathrm{GIG}(\lambda_k; \eta_k^{(i+1)}, \varphi_k^{(i+1)}, \varrho_k^{(i+1)})$	$\eta_k^{(i+1)} = \bar{\Delta}_{1,k}^{(i+1)} + \mathrm{E}^{(i)}[\nu]$, $\varphi_k^{(i+1)} = \bar{\Delta}_{3,k}^{(i+1)}$, $\varrho_k^{(i+1)} = 0.5(m + \mathrm{E}^{(i)}[\nu])$
SGEM	$q^{(i+1)}(\xi_k) = \mathrm{GIG}(\xi_k; \alpha_k^{(i+1)}, \beta_k^{(i+1)}, \rho_k^{(i+1)})$	$\alpha_k^{(i+1)} = \Delta_{1,k}^{(i+1)} + 2\mathrm{E}^{(i)}[\omega]$, $\beta_k^{(i+1)} = \Delta_{3,k}^{(i+1)}$, $\rho_k^{(i+1)} = 0.5n + 1$
	$q^{(i+1)}(\lambda_k) = \mathrm{GIG}(\lambda_k; \eta_k^{(i+1)}, \varphi_k^{(i+1)}, \varrho_k^{(i+1)})$	$\eta_k^{(i+1)} = \bar{\Delta}_{1,k}^{(i+1)} + 2\mathrm{E}^{(i)}[\nu]$, $\varphi_k^{(i+1)} = \bar{\Delta}_{3,k}^{(i+1)}$, $\varrho_k^{(i+1)} = 0.5m + 1$
SGBM	$q^{(i+1)}(\xi_k) = c_4\mathrm{GIG}(\xi_k; \alpha_k^{(i+1)}, \beta_k^{(i+1)}, \rho_k^{(i+1)})$	$\alpha_k^{(i+1)} = \Delta_{1,k}^{(i+1)}$, $\beta_k^{(i+1)} = \Delta_{3,k}^{(i+1)}$, $\rho_k^{(i+1)} = 0.5n + \mathrm{E}^{(i)}[\omega]$, $\xi_k \in (0,1)$
	$q^{(i+1)}(\lambda_k) = \bar{c}_4\mathrm{GIG}(\lambda_k; \eta_k^{(i+1)}, \varphi_k^{(i+1)}, \varrho_k^{(i+1)})$	$\eta_k^{(i+1)} = \bar{\Delta}_{1,k}^{(i+1)}$, $\varphi_k^{(i+1)} = \bar{\Delta}_{3,k}^{(i+1)}$, $\varrho_k^{(i+1)} = 0.5m + \mathrm{E}^{(i)}[\nu]$, $\lambda_k \in (0,1)$
SGIGM	$q^{(i+1)}(\xi_k) = \mathrm{GIG}(\xi_k; \alpha_k^{(i+1)}, \beta_k^{(i+1)}, \rho_k^{(i+1)})$	$\alpha_k^{(i+1)} = \Delta_{1,k}^{(i+1)}$, $\beta_k^{(i+1)} = \Delta_{3,k}^{(i+1)} + \mathrm{E}^{(i)}[\omega]$, $\rho_k^{(i+1)} = 0.5(n - \mathrm{E}^{(i)}[\omega])$
	$q^{(i+1)}(\lambda_k) = \mathrm{GIG}(\lambda_k; \eta_k^{(i+1)}, \varphi_k^{(i+1)}, \varrho_k^{(i+1)})$	$\eta_k^{(i+1)} = \bar{\Delta}_{1,k}^{(i+1)}$, $\varphi_k^{(i+1)} = \bar{\Delta}_{3,k}^{(i+1)} + \mathrm{E}^{(i)}[\nu]$, $\varrho_k^{(i+1)} = 0.5(m - \mathrm{E}^{(i)}[\nu])$
SGIEM	$q^{(i+1)}(\xi_k) = \mathrm{GIG}(\xi_k; \alpha_k^{(i+1)}, \beta_k^{(i+1)}, \rho_k^{(i+1)})$	$\alpha_k^{(i+1)} = \Delta_{1,k}^{(i+1)}$, $\beta_k^{(i+1)} = \Delta_{3,k}^{(i+1)} + 2\mathrm{E}^{(i)}[\omega]$, $\rho_k^{(i+1)} = 0.5n - 1$
	$q^{(i+1)}(\lambda_k) = \mathrm{GIG}(\lambda_k; \eta_k^{(i+1)}, \varphi_k^{(i+1)}, \varrho_k^{(i+1)})$	$\eta_k^{(i+1)} = \bar{\Delta}_{1,k}^{(i+1)}$, $\varphi_k^{(i+1)} = \bar{\Delta}_{3,k}^{(i+1)} + 2\mathrm{E}^{(i)}[\nu]$, $\varrho_k^{(i+1)} = 0.5m - 1$

For the Monte Carlo updates of posterior PDFs $q^{(i+1)}(\xi_k)$ and $q^{(i+1)}(\lambda_k)$, the required expec-

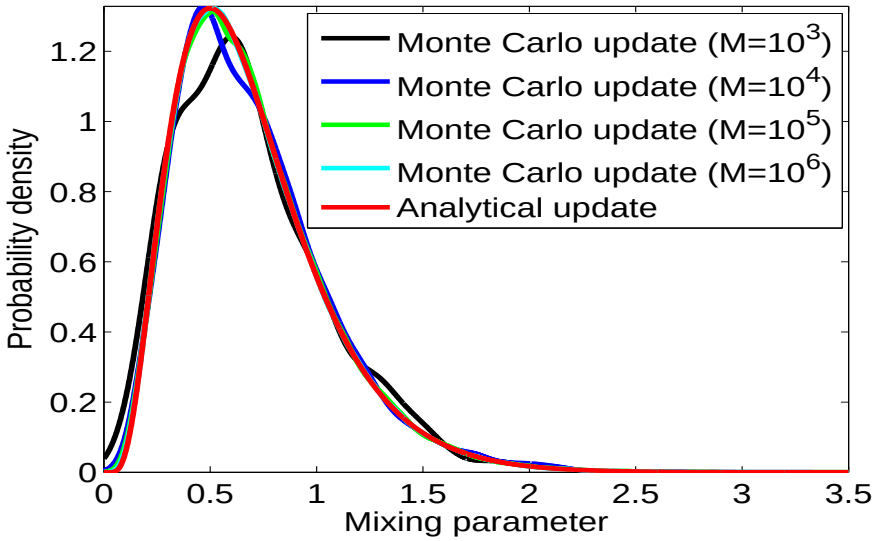


Fig. 3: Monte Carlo update and analytical update of $q^{(i+1)}(\xi_k)$ and $q^{(i+1)}(\lambda_k)$ for the SGGM distribution when $M \in \{10^3, 10^4, 10^5, 10^6\}$.

tations of mixing parameters can be approximated as

$$\begin{cases} \mathrm{E}^{(i+1)}[\xi_k] \approx \sum\limits_{j=1}^{M} w_{\xi,k}^{(i+1)j} \xi_k^j, \quad \mathrm{E}^{(i+1)}[\lambda_k] \approx \sum\limits_{j=1}^{M} w_{\lambda,k}^{(i+1)j} \lambda_k^j \\ \mathrm{E}^{(i+1)}[\frac{1}{\xi_k}] \approx \sum\limits_{j=1}^{M} w_{\xi,k}^{(i+1)j} \frac{1}{\xi_k^j}, \quad \mathrm{E}^{(i+1)}[\frac{1}{\lambda_k}] \approx \sum\limits_{j=1}^{M} w_{\lambda,k}^{(i+1)j} \frac{1}{\lambda_k^j} \\ \mathrm{E}^{(i+1)}[\log \xi_k] \approx \sum\limits_{j=1}^{M} w_{\xi,k}^{(i+1)j} \log \xi_k^j \\ \mathrm{E}^{(i+1)}[\log \lambda_k] \approx \sum\limits_{j=1}^{M} w_{\lambda,k}^{(i+1)j} \log \lambda_k^j \end{cases} \tag{81}$$

To illustrate the effectiveness of the proposed Monte Carlo update method, the Monte Carlo update and analytical update of the posterior PDFs $q^{(i+1)}(\xi_k)$ and $q^{(i+1)}(\lambda_k)$ are compared. As an example, the results of the SGGM distribution are given, where the parameters are set as $n = m = 1$, $\Delta_{1,k}^{(i+1)} = \bar{\Delta}_{1,k}^{(i+1)} = 5$, $\Delta_{2,k}^{(i+1)} = \bar{\Delta}_{2,k}^{(i+1)} = 2$, $\Delta_{3,k}^{(i+1)} = \bar{\Delta}_{3,k}^{(i+1)} = 0.5$, and $\mathrm{E}^{(i)}[\omega] = \mathrm{E}^{(i)}[\nu] = 5$. Based on such parameter selections, the mixing parameters ξ_k and λ_k have the same posterior PDFs, i.e., $q^{(i+1)}(\xi_k) = q^{(i+1)}(\lambda_k)$, and only the results of one of the mixing parameters will be shown in the next numerical comparisons.

The Monte Carlo update and analytical update of $q^{(i+1)}(\xi_k)$ and $q^{(i+1)}(\lambda_k)$ for the SGGM distribution when $M = 10^3$, $M = 10^4$, $M = 10^5$ and $M = 10^6$ are shown in Fig. 3. It is

TABLE IV: Shape parameters and rate parameters of $q^{(i+1)}(\omega)$ and $q^{(i+1)}(\nu)$ for exemplary GGScM distributions.

GGScM distributions	Shape parameters and rate parameters
GGM and SGGM	$c^{(i+1)} = 0.5T + 1$, $d^{(i+1)} = -0.5T - 0.5\sum_{k=1}^{T}(\mathrm{E}^{(i+1)}[\log \xi_k] - \mathrm{E}^{(i+1)}[\xi_k])$
	$a^{(i+1)} = 0.5T + 1$, $b^{(i+1)} = -0.5T - 0.5\sum_{k=1}^{T}(\mathrm{E}^{(i+1)}[\log \lambda_k] - \mathrm{E}^{(i+1)}[\lambda_k])$
GEM and SGEM	$c^{(i+1)} = T + 1$, $d^{(i+1)} = \sum_{k=1}^{T}\mathrm{E}^{(i+1)}[\xi_k]$
	$a^{(i+1)} = T + 1$, $b^{(i+1)} = \sum_{k=1}^{T}\mathrm{E}^{(i+1)}[\lambda_k]$
GBM and SGBM	$c^{(i+1)} = T + 1$, $d^{(i+1)} = -\sum_{k=1}^{T}\mathrm{E}^{(i+1)}[\log \xi_k]$
	$a^{(i+1)} = T + 1$, $b^{(i+1)} = -\sum_{k=1}^{T}\mathrm{E}^{(i+1)}[\log \lambda_k]$
GIGM and SGIGM	$c^{(i+1)} = 0.5T + 1$, $d^{(i+1)} = -0.5T + 0.5\sum_{k=1}^{T}(\mathrm{E}^{(i+1)}[\log \xi_k] + \mathrm{E}^{(i+1)}[\frac{1}{\xi_k}])$
	$a^{(i+1)} = 0.5T + 1$, $b^{(i+1)} = -0.5T + 0.5\sum_{k=1}^{T}(\mathrm{E}^{(i+1)}[\log \lambda_k] + \mathrm{E}^{(i+1)}[\frac{1}{\lambda_k}])$
GIEM and SGIEM	$c^{(i+1)} = T + 1$, $d^{(i+1)} = \sum_{k=1}^{T}\mathrm{E}^{(i+1)}[\frac{1}{\xi_k}]$
	$a^{(i+1)} = T + 1$, $b^{(i+1)} = \sum_{k=1}^{T}\mathrm{E}^{(i+1)}[\frac{1}{\lambda_k}]$

observed from Fig. 3 that the Monte Carlo update has almost the same outlines as the analytical update when $M = 10^3$, and the Monte Carlo update approaches the analytical update as the number of random samples increases, and the Monte Carlo update and analytical update are almost identical when $M = 10^5$ and $M = 10^6$. The KLDs between Monte Carlo update and analytical update when $M = 10^3$, $M = 10^4$, $M = 10^5$ and $M = 10^6$ are, respectively, 2.4×10^{-2}, 4.0×10^{-3}, 9.6×10^{-4} and 2.6×10^{-4}. The KLDs between Monte Carlo update and analytical update reduce gradually as the number of random samples raises. Thus, the posterior PDFs of mixing parameters are well updated based on the Monte Carlo approach with $M = 10^3$ random samples, and the accuracy of the Monte Carlo update can be further improved by increasing the number of random samples.

D. Updates of $q^{(i+1)}(\omega)$ and $q^{(i+1)}(\nu)$

As an example, next, we derive the updates of $q^{(i+1)}(\omega)$ and $q^{(i+1)}(\nu)$ when state and measurement noises are modelled by the GIGM distribution. Substituting $\pi_1(\xi_k;\omega) = \mathrm{IG}(\xi_k;\frac{\omega}{2},\frac{\omega}{2})$ and $\pi_2(\lambda_k;\nu) = \mathrm{IG}(\lambda_k;\frac{\nu}{2},\frac{\nu}{2})$ in (52)-(53) and using Stirling's approximation: $\log\Gamma(0.5y) \approx$

TABLE V: The proposed robust RTS smoother based on the GGM and SGEM distributions.

Inputs: $\hat{\mathbf{x}}_{0|0}$, $\mathbf{P}_{0|0}$, $\{\mathbf{F}_k, \mathbf{H}_k, \mathbf{z}_k | 1 \leq k \leq T\}$, $\bar{\mathbf{Q}}$, $\bar{\mathbf{R}}$, $\bar{\omega}$, $\bar{\nu}$, ϵ, M, N_m.

1. Initialization: $\mathrm{E}^{(0)}[\xi_k] = 1$, $\mathrm{E}^{(0)}[\lambda_k] = 1$, $\mathrm{E}^{(0)}[\boldsymbol{\beta}_1] = \mathbf{0}$, $\mathrm{E}^{(0)}[\boldsymbol{\beta}_2] = \mathbf{0}$, $\mathrm{E}^{(0)}[\mathbf{Q}^{-1}] = \bar{\mathbf{Q}}^{-1}$, $\mathrm{E}^{(0)}[\mathbf{R}^{-1}] = \bar{\mathbf{R}}^{-1}$, $\mathrm{E}^{(0)}[\omega] = \bar{\omega}$, $\mathrm{E}^{(0)}[\nu] = \bar{\nu}$.

for $i = 0 : N - 1$

Update $q^{(i+1)}(\mathbf{x}_{0:T})$ by (27):

2. Calculate the modified mean vector $\tilde{\mathbf{q}}_k^{(i)}$ and covariance matrix $\tilde{\mathbf{Q}}_k^{(i)}$ of state noise from time sample 1 to time sample T using (65).

3. Calculate the modified mean vector $\tilde{\mathbf{r}}_k^{(i)}$ and covariance matrix $\tilde{\mathbf{R}}_k^{(i)}$ of measurement noise from time 1 to T using (67).

4. Calculate $\{\hat{\mathbf{x}}_{k|T}^{(i+1)}, \mathbf{P}_{k|T}^{(i+1)}, \mathbf{G}_{k-1}^{(i+1)} | 0 \leq k \leq T\}$ by running standard RTS smoother [36] with inputs $\mathbf{z}_{1:T}$, $\hat{\mathbf{x}}_{0|0}$, $\mathbf{P}_{0|0}$, and $\left\{\mathbf{F}_k, \mathbf{H}_k, \tilde{\mathbf{q}}_k^{(i)}, \tilde{\mathbf{r}}_k^{(i)}, \tilde{\mathbf{Q}}_k^{(i)}, \tilde{\mathbf{R}}_k^{(i)} | 1 \leq k \leq T\right\}$.

5. Calculate $\mathbf{A}_k^{(i+1)}$, $\mathbf{B}_k^{(i+1)}$, $\mathbf{a}_k^{(i+1)}$ and $\mathbf{b}_k^{(i+1)}$ using (62).

Update $q^{(i+1)}(\boldsymbol{\beta}_1)$ and $q^{(i+1)}(\boldsymbol{\beta}_2)$ by (29):

6. Calculate the mean vector $\hat{\boldsymbol{\beta}}_1^{(i+1)}$ and covariance matrix $\mathbf{P}_{\boldsymbol{\beta}_1}^{(i+1)}$ using (63).

7. Calculate the mean vector $\hat{\boldsymbol{\beta}}_2^{(i+1)}$ and covariance matrix $\mathbf{P}_{\boldsymbol{\beta}_2}^{(i+1)}$ using (68).

8. Calculate $\mathrm{E}^{(i+1)}[\boldsymbol{\beta}_1]$, $\mathrm{E}^{(i+1)}[\boldsymbol{\beta}_1\boldsymbol{\beta}_1^{\mathrm{T}}]$, $\mathrm{E}^{(i+1)}[\boldsymbol{\beta}_2]$ and $\mathrm{E}^{(i+1)}[\boldsymbol{\beta}_2\boldsymbol{\beta}_2^{\mathrm{T}}]$ using (64) and (58), respectively.

Update $q^{(i+1)}(\xi_k)$ and $q^{(i+1)}(\lambda_k)$ by analytical method in Table VI or Monte Carlo approach in Table VII.

Update $q^{(i+1)}(\mathbf{Q})$ and $q^{(i+1)}(\mathbf{R})$ by (48):

9. Calculate $\mathbf{E}^{(i+1)}$ and $\mathbf{F}^{(i+1)}$ using (66) and (69), respectively.

10. Calculate the dof parameters $t^{(i+1)}$ and $u^{(i+1)}$ and inverse scale matrices $\mathbf{T}^{(i+1)}$ and $\mathbf{U}^{(i+1)}$ using (49).

11. Calculate $\mathrm{E}^{(i+1)}[\mathbf{Q}^{-1}]$ and $\mathrm{E}^{(i+1)}[\mathbf{R}^{-1}]$ using (60).

Update $q^{(i+1)}(\omega)$ and $q^{(i+1)}(\nu)$ by (54):

12. Calculate the shape parameter $c^{(i+1)}$ and rate parameter $d^{(i+1)}$ of GGM distribution in Table IV.

13. Calculate the shape parameter $a^{(i+1)}$ and rate parameter $b^{(i+1)}$ of SGEM distribution in Table IV.

14. Calculate $\mathrm{E}^{(i+1)}[\omega]$ and $\mathrm{E}^{(i+1)}[\nu]$ using (61).

15. If $\frac{\left\|\hat{\mathbf{x}}_{0:T|T}^{(i+1)} - \hat{\mathbf{x}}_{0:T|T}^{(i)}\right\|}{\left\|\hat{\mathbf{x}}_{0:T|T}^{(i)}\right\|} <= \epsilon$, stop iteration.

end

16. Calculate the smoothing estimates of the state trajectory, mixing parameters and unknown distribution parameters using (53).

Outputs: $\{\hat{\mathbf{x}}_{k|T}, \mathbf{P}_{k|T} | 0 \leq k \leq T\}$, $\{\hat{\xi}_{k|T}, \hat{\lambda}_{k|T} | 1 \leq k \leq T\}$, $\hat{\boldsymbol{\beta}}_1$, $\hat{\boldsymbol{\beta}}_2$, $\hat{\mathbf{Q}}$, $\hat{\mathbf{R}}$, $\hat{\omega}$, $\hat{\nu}$.

$(0.5y - 0.5)\log(0.5y)\ -0.5y$ yields

$$\begin{cases} \log q^{(i+1)}(\omega) \approx 0.5T \log \omega - \{-0.5T + \\ 0.5 \sum_{k=1}^{T} (\mathrm{E}^{(i+1)}[\log \xi_k] + \mathrm{E}^{(i+1)}[\frac{1}{\xi_k}]) \Big\} \omega + c_\omega \\ \log q^{(i+1)}(\nu) \approx 0.5T \log \nu - \{-0.5T + \\ 0.5 \sum_{k=1}^{T} (\mathrm{E}^{(i+1)}[\log \lambda_k] + \mathrm{E}^{(i+1)}[\frac{1}{\lambda_k}]) \Big\} \nu + c_\nu \end{cases} \tag{82}$$

TABLE VI: Analytical update of mixing parameters based on the GGM and SGEM distributions.

Update $q^{(i+1)}(\xi_k)$ and $q^{(i+1)}(\lambda_k)$ by Table III.
1. Calculate the parameters $\Delta_{1,k}^{(i+1)}$, $\bar{\Delta}_{1,k}^{(i+1)}$ and $\bar{\Delta}_{3,k}^{(i+1)}$ using (37).
2. Calculate the parameter $\Delta_{3,k}^{(i+1)}$ using (70).
3. Calculate the shape parameter $\alpha_k^{(i+1)}$ and rate parameter $\beta_k^{(i+1)}$ for GGM distribution in Table III.
4. Calculate the shape parameters $\eta_k^{(i+1)}$, $\varphi_k^{(i+1)}$ and $\varrho_k^{(i+1)}$ for SGEM distribution in Table III.
5. Calculate the expectations $\mathrm{E}^{(i+1)}[\xi_k]$ and $\mathrm{E}^{(i+1)}[\log \xi_k]$ using (75).
6. Calculate the expectations $\mathrm{E}^{(i+1)}[\lambda_k]$, $\mathrm{E}^{(i+1)}[\frac{1}{\lambda_k}]$ and $\mathrm{E}^{(i+1)}[\log \lambda_k]$ using (77).

TABLE VII: Monte Carlo update of mixing parameters based on the GGM and SGEM distributions.

Update $q^{(i+1)}(\xi_k)$ and $q^{(i+1)}(\lambda_k)$ by (46)-(47).
1. Calculate the parameters $\Delta_{1,k}^{(i+1)}$, $\bar{\Delta}_{1,k}^{(i+1)}$, $\bar{\Delta}_{2,k}^{(i+1)}$ and $\bar{\Delta}_{3,k}^{(i+1)}$
using (37).
2. Calculate the parameters $\Delta_{2,k}^{(i+1)}$ and $\Delta_{3,k}^{(i+1)}$ using (70).
3. Draw M random samples $\left\{\xi_k^j\right\}_{j=1}^M$ from the prior PDF $\mathrm{G}\left(\xi_k; \frac{\mathrm{E}^{(i)}[\omega]}{2}, \frac{\mathrm{E}^{(i)}[\omega]}{2}\right)$ using (45).
4. Draw M random samples $\left\{\lambda_k^j\right\}_{j=1}^M$ from the prior PDF $\mathrm{Ex}(\lambda_k; \mathrm{E}^{(i)}[\nu])$ using (45).
5. Calculate the weights $\left\{w_{\xi,k}^{(i+1)j}\right\}_{j=1}^M$ using (47) and (79).
6. Calculate the weights $\left\{w_{\lambda,k}^{(i+1)j}\right\}_{j=1}^M$ using (47) and (80).
7. Calculate $\mathrm{E}^{(i+1)}[\xi_k]$, $\mathrm{E}^{(i+1)}[\log \xi_k]$, $\mathrm{E}^{(i+1)}[\lambda_k]$, $\mathrm{E}^{(i+1)}[\frac{1}{\lambda_k}]$ and $\mathrm{E}^{(i+1)}[\log \lambda_k]$ using (81).

According to (82), $q^{(i+1)}(\omega)$ and $q^{(i+1)}(\nu)$ can be updated as Gamma through (54), where the shape and rate parameters $c^{(i+1)}$, $d^{(i+1)}$, $a^{(i+1)}$ and $b^{(i+1)}$ are, respectively, given by

$$\begin{cases} c^{(i+1)} = 0.5T + 1, \qquad a^{(i+1)} = 0.5T + 1 \\ d^{(i+1)} = -0.5T + 0.5 \sum_{k=1}^{T} (\mathrm{E}^{(i+1)}[\log \xi_k] + \mathrm{E}^{(i+1)}[\frac{1}{\xi_k}]) \\ b^{(i+1)} = -0.5T + 0.5 \sum_{k=1}^{T} (\mathrm{E}^{(i+1)}[\log \lambda_k] + \mathrm{E}^{(i+1)}[\frac{1}{\lambda_k}]) \end{cases} \tag{83}$$

Similarly, for other exemplary GGScM distributions, the posterior PDFs $q^{(i+1)}(\omega)$ and $q^{(i+1)}(\nu)$ can be also updated as Gamma through (54), and the shape parameters and rate parameters are

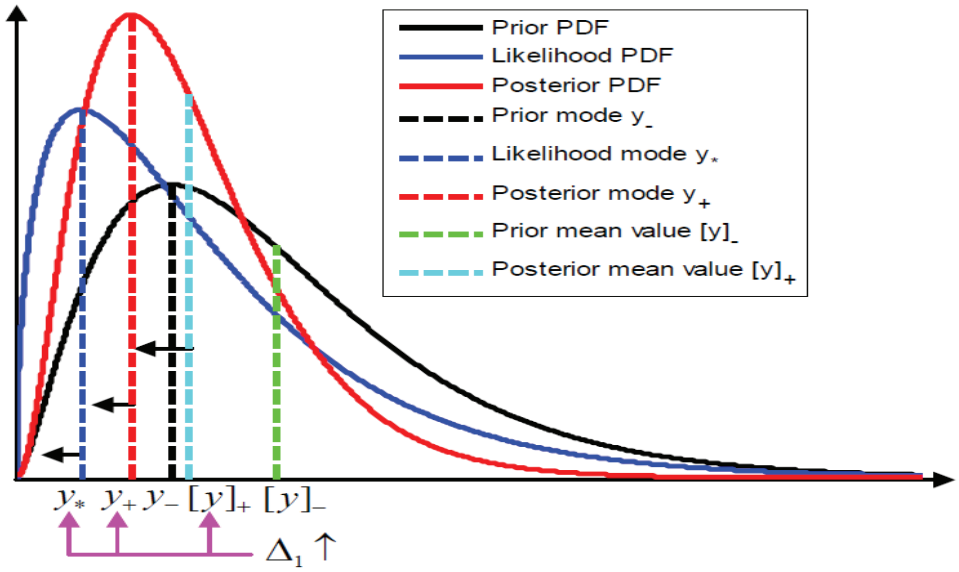


Fig. 4: Diagram of prior, likelihood and posterior PDFs for the exemplary GGScM distributions.

given in Table IV.

The proposed robust RTS smoothers based on exemplary GGScM distributions are composed of the updates of posterior PDFs of state trajectory, mixing parameters and unknown distribution parameters and the calculations of expectations. Different robust RTS smoothers can be derived when different exemplary GGScM distributions are employed to model state and measurement noises. As an example, the implementation pseudo-code for the proposed robust RTS smoother based on the GGM and SGEM distributions is illustrated in Table V, where the state and measurement noises are, respectively, modelled by the GGM and SGEM distributions, and $\bar{\mathbf{Q}}$ and $\bar{\mathbf{R}}$ denote the nominal state and measurement noise covariance matrices, respectively, and $\bar{\omega}$ and $\bar{\nu}$ denote the initial dof parameters of the GGM and SGEM distributions, respectively, and ϵ denotes the iteration threshold, and N_m denotes the maximum number of iterations. In Table V, the posterior PDFs of mixing parameters can be updated using the analytical method or Monte Carlo approach, which are, respectively, shown in Tables VI and VII.

E. Robustness analyses

Generally, in the RTS smoothing framework, the covariance matrices of state and measurement residuals will automatically increase when the state and measurement noises have abnormal and

TABLE VIII: The modes and mean values of posterior PDF $q(y)$ for the exemplary GGScM distributions.

GGScM distributions	Mode y_+	Mean value $[y]_+$	Parameters
GGM	$(\alpha-1)/\beta$	α/β	$\alpha=0.5s+0.5\nu$, $\beta=0.5\Delta_1+0.5\nu$
GEM	$(\alpha-1)/\beta$	α/β	$\alpha=0.5s+1$, $\beta=0.5\Delta_1+\nu$
GBM	$(\alpha-1)/\beta$	\	$\alpha=0.5s+\nu$, $\beta=0.5\Delta_1$
GIGM	$\frac{\beta}{\sqrt{(\rho-1)^2+\alpha\beta}+(1-\rho)}$	$\frac{\sqrt{\beta}K_{\rho+1}(\sqrt{\alpha\beta})}{\sqrt{\alpha}K_\rho(\sqrt{\alpha\beta})}$	$\alpha=\Delta_1$, $\beta=\nu$, $\rho=0.5s-0.5\nu$
GIEM	$\frac{\beta}{\sqrt{(\rho-1)^2+\alpha\beta}+(1-\rho)}$	$\frac{\sqrt{\beta}K_{\rho+1}(\sqrt{\alpha\beta})}{\sqrt{\alpha}K_\rho(\sqrt{\alpha\beta})}$	$\alpha=\Delta_1$, $\beta=2\nu$, $\rho=0.5s-1$
SGGM	$\frac{\beta}{\sqrt{(\rho-1)^2+\alpha\beta}+(1-\rho)}$	$\frac{\sqrt{\beta}K_{\rho+1}(\sqrt{\alpha\beta})}{\sqrt{\alpha}K_\rho(\sqrt{\alpha\beta})}$	$\alpha=\Delta_1+\nu$, $\beta=\Delta_3$, $\rho=0.5s+0.5\nu$
SGEM	$\frac{\beta}{\sqrt{(\rho-1)^2+\alpha\beta}+(1-\rho)}$	$\frac{\sqrt{\beta}K_{\rho+1}(\sqrt{\alpha\beta})}{\sqrt{\alpha}K_\rho(\sqrt{\alpha\beta})}$	$\alpha=\Delta_1+2\nu$, $\beta=\Delta_3$, $\rho=0.5s+1$
SGBM	$\frac{\beta}{\sqrt{(\rho-1)^2+\alpha\beta}+(1-\rho)}$	\	$\alpha=\Delta_1$, $\beta=\Delta_3$, $\rho=0.5s+\nu$
SGIGM	$\frac{\beta}{\sqrt{(\rho-1)^2+\alpha\beta}+(1-\rho)}$	$\frac{\sqrt{\beta}K_{\rho+1}(\sqrt{\alpha\beta})}{\sqrt{\alpha}K_\rho(\sqrt{\alpha\beta})}$	$\alpha=\Delta_1$, $\beta=\Delta_3+\nu$, $\rho=0.5s-0.5\nu$
SGIEM	$\frac{\beta}{\sqrt{(\rho-1)^2+\alpha\beta}+(1-\rho)}$	$\frac{\sqrt{\beta}K_{\rho+1}(\sqrt{\alpha\beta})}{\sqrt{\alpha}K_\rho(\sqrt{\alpha\beta})}$	$\alpha=\Delta_1$, $\beta=\Delta_3+2\nu$, $\rho=0.5s-1$

infrequent values which are induced by the heavy-tailed and/or skew state and measurement noises. That is to say, for the proposed robust RTS smoothing framework, the parameters $\mathbf{A}_k^{(i+1)}$ and $\mathbf{B}_k^{(i+1)}$ will increase when the abnormal and infrequent noise values appear. Then, according to (37), the parameters $\Delta_{1,k}^{(i+1)}$ and $\bar{\Delta}_{1,k}^{(i+1)}$ also increase. For the proposed robust RTS smoothers based on the exemplary GGScM distributions, the modified mean vectors and covariance matrices of non-Gaussian state and measurement noises are time-varying and adaptively adjusted based on the estimates of mixing parameters ξ_k and λ_k, as is shown in (65) and (67). To accommodate the heavy-tailed and/or skew state and measurement noises, the posterior mean values of mixing parameters ξ_k and λ_k need to decrease adaptively when the parameters $\Delta_{1,k}^{(i+1)}$ and $\bar{\Delta}_{1,k}^{(i+1)}$ raise. To this end, next, we will analyse the behaviours of the posterior mean values of mixing parameters ξ_k and λ_k in terms of the parameters $\Delta_{1,k}^{(i+1)}$ and $\bar{\Delta}_{1,k}^{(i+1)}$.

Before presenting the analyses, four important inequalities are firstly given as follows

$$\begin{cases} \Delta_{1,k}^{(i+1)} > 0, & \bar{\Delta}_{1,k}^{(i+1)} > 0 \\ \Delta_{3,k}^{(i+1)} \geq 0, & \bar{\Delta}_{3,k}^{(i+1)} \geq 0 \end{cases} \tag{84}$$

where $\Delta_{3,k}^{(i+1)} = 0$ and $\bar{\Delta}_{3,k}^{(i+1)} = 0$ if and only if $\mathrm{E}^{(i+1)}[\boldsymbol{\beta}_1\boldsymbol{\beta}_1^{\mathrm{T}}] = \mathbf{0}$ and $\mathrm{E}^{(i+1)}[\boldsymbol{\beta}_2\boldsymbol{\beta}_2^{\mathrm{T}}] = \mathbf{0}$. The

proof of (84) is given in Appendix F.

Employing (80) in (43), the likelihood PDFs of the exemplary GGScM distributions can be written as

$$\begin{cases} l_1(\xi_k) = c_2 \exp(0.5n \log \xi_k - 0.5\xi_k \Delta_{1,k}^{(i+1)} + \Delta_{2,k}^{(i+1)} - 0.5\Delta_{3,k}^{(i+1)}/\xi_k) \\ l_2(\lambda_k) = \bar{c}_2 \exp(0.5m \log \lambda_k - 0.5\lambda_k \bar{\Delta}_{1,k}^{(i+1)} + \bar{\Delta}_{2,k}^{(i+1)} - 0.5\bar{\Delta}_{3,k}^{(i+1)}/\lambda_k) \end{cases} \tag{85}$$

where $\Delta_{3,k}^{(i+1)} = \bar{\Delta}_{3,k}^{(i+1)} = 0$ for the exemplary symmetric GGScM distributions.

According to (85), the likelihood PDFs of the exemplary GGScM distributions can be expressed as a unified form

$$l(y) = c \exp(0.5s \log y - 0.5\Delta_1 y + \Delta_2 - 0.5\Delta_3/y) \quad \text{s.t. } y > 0, \;\; \Delta_1 > 0, \;\; \Delta_3 \geq 0 \tag{86}$$

where $l(y)$ becomes $l_1(\xi_k)$ and $l_2(\lambda_k)$ when the following equations hold

$$\begin{cases} y = \xi_k, c = c_2, s = n, \Delta_1 = \Delta_{1,k}^{(i+1)}, \Delta_2 = \Delta_{2,k}^{(i+1)}, \Delta_3 = \Delta_{3,k}^{(i+1)} \\ y = \lambda_k, c = \bar{c}_2, s = m, \Delta_1 = \bar{\Delta}_{1,k}^{(i+1)}, \Delta_2 = \bar{\Delta}_{2,k}^{(i+1)}, \Delta_3 = \bar{\Delta}_{3,k}^{(i+1)} \end{cases} \tag{87}$$

Using (40) and (86), the posterior PDFs can be also expressed as a unified form

$$q(y) = cl(y)\pi(y;\nu), \quad \text{s.t. } y > 0, \;\; \Delta_1 > 0, \;\; \Delta_3 \geq 0 \tag{88}$$

where c denotes a normalizing constant, and $\pi(y;\nu)$ denotes a prior PDF, and $q(y)$ becomes $q^{(i+1)}(\xi_k)$ and $q^{(i+1)}(\lambda_k)$ when the following equations hold

$$\begin{cases} c = c_1, \quad l(y) = l_1(\xi_k), \quad \pi(y;\nu) = \pi_1(\xi_k; \mathrm{E}^{(i)}[\omega]) \\ c = \bar{c}_1, \quad l(y) = l_2(\lambda_k), \quad \pi(y;\nu) = \pi_2(\lambda_k; \mathrm{E}^{(i)}[\nu]) \end{cases} \tag{89}$$

Employing (86), we obtain

$$\log l(y) = 0.5s \log y - 0.5\Delta_1 y - 0.5\Delta_3/y + c_y \quad \text{s.t. } y > 0, \;\; \Delta_1 > 0, \;\; \Delta_3 \geq 0 \tag{90}$$

It is observed from (90) that the likelihood PDF $l(y)$ is a generalized inverse Gaussian PDF with shape parameters Δ_1, Δ_3 and $0.5s + 1$, i.e.,

$$l(y) = \mathrm{GIG}(y; \Delta_1, \Delta_3, 0.5s + 1) \tag{91}$$

Then, according to (91), it is apparent that the likelihood PDF $l(y)$ is unimodal and has a unique mode y_* that is given by

$$y_* = \frac{s}{2\Delta_1} + \sqrt{\left(\frac{s}{2\Delta_1}\right)^2 + \frac{\Delta_3}{\Delta_1}} \tag{92}$$

The modes and mean values of posterior PDF $q(y)$ for the exemplary GGScM distributions are listed in Table VIII, where y_+ and $[y]_+$, respectively, denote the mode and mean value of the posterior PDF. The modes and mean values are obtained using the analytical posterior PDFs of mixing parameters, and the modes and mean values of the GBM and SGBM distributions are unavailable since the posterior PDFs of mixing parameters for the GBM and SGBM distributions are, respectively, truncated Gamma and generalized inverse Gaussian distributions. It is seen from (92) that the mode y_* of the likelihood PDF reduces with the increase of parameter Δ_1, which makes the likelihood PDF move left. We can see from Table VIII that the mode y_+ of the posterior PDF decreases with the increase of parameter Δ_1. Then, the posterior PDF with a fixed prior PDF also moves left, and the mean value $[y]_+$ of the posterior PDF decreases. Thus, the mean value $[y]_+$ of the posterior PDF reduces as the parameter Δ_1 raises. The diagram of the prior PDF $\pi(y;\nu)$, likelihood PDF $l(y)$ and posterior PDF $q(y)$ for the exemplary GGScM distributions is shown in Fig. 4, where y_- and $[y]_-$, respectively, denote the mode and mean value of the prior PDF.

Based on the above discussions, both the posterior mean values $\mathrm{E}^{(i+1)}[\xi_k]$ and $\mathrm{E}^{(i+1)}[\lambda_k]$ of the mixing parameters ξ_k and λ_k reduce with the increases of $\Delta_{1,k}^{(i+1)}$ and $\bar{\Delta}_{1,k}^{(i+1)}$. Then, according to (67), the modified mean vectors $\tilde{\mathbf{q}}_k^{(i)}$ and $\tilde{\mathbf{r}}_k^{(i)}$ and covariance matrices $\tilde{\mathbf{Q}}_k^{(i)}$ and $\tilde{\mathbf{R}}_k^{(i)}$ of the state and measurement noises all raise with the increases of the parameters $\Delta_{1,k}^{(i+1)}$ and $\bar{\Delta}_{1,k}^{(i+1)}$, which accommodates the heavy-tailed and/or skew state and measurement noises.

F. A KLD-based selection scheme for GGScM distributions

In practical applications, two GGScM distributions need to be selected to model the state and measurement noises. The optimal selection of the two GGScM distributions can minimize the KLD between the approximate joint PDF $q_{\gamma_1,\gamma_2}(\mathbf{w}_{k-1})q_{\gamma_1,\gamma_2}(\mathbf{v}_k)$ and the true joint PDF $p(\mathbf{w}_{k-1},\mathbf{v}_k)$ of the state and measurement noises, i.e.,

$$\{\gamma_{1,o},\gamma_{2,o}\} = \arg\min_{\gamma_1,\gamma_2} \mathrm{KLD}\{q_{\gamma_1,\gamma_2}(\mathbf{w}_{k-1})q_{\gamma_1,\gamma_2}(\mathbf{v}_k) \parallel p(\mathbf{w}_{k-1},\mathbf{v}_k)\} \tag{93}$$

where γ_1 and γ_2 represent the GGScM distributions for modelling the state and measurement noises, respectively, and $\{\gamma_{1,o},\gamma_{2,o}\}$ is the optimal selection of $\{\gamma_1,\gamma_2\}$, and $q_{\gamma_1,\gamma_2}(\mathbf{w}_{k-1})$ and $q_{\gamma_1,\gamma_2}(\mathbf{v}_k)$ denote, respectively, the GGScM approximations of state and measurement noises based on the selection of GGScM distributions $\{\gamma_1,\gamma_2\}$.

Employing (18) and (55), the approximate PDFs $q_{\gamma_1,\gamma_2}(\mathbf{w}_{k-1})$ and $q_{\gamma_1,\gamma_2}(\mathbf{v}_k)$ of state and measurement noises are formulated as

$$\begin{cases} q_{\gamma_1,\gamma_2}(\mathbf{w}_{k-1}) = \int_0^{+\infty} \mathrm{g}\left(\mathbf{w}_{k-1}; \hat{\boldsymbol{\beta}}_{1,\gamma_1,\gamma_2}/s_{1,\gamma_1}(\xi_k),\ \hat{\mathbf{Q}}_{\gamma_1,\gamma_2}/\kappa_{1,\gamma_1}(\xi_k)\right) \pi_{1,\gamma_1}(\xi_k; \hat{\omega}_{\gamma_1,\gamma_2}) d\xi_k \\ q_{\gamma_1,\gamma_2}(\mathbf{v}_k) = \int_0^{+\infty} \mathrm{g}\left(\mathbf{v}_k; \hat{\boldsymbol{\beta}}_{2,\gamma_1,\gamma_2}/s_{2,\gamma_2}(\lambda_k),\ \hat{\mathbf{R}}_{\gamma_1,\gamma_2}/\kappa_{2,\gamma_2}(\lambda_k)\right) \pi_{2,\gamma_2}(\lambda_k; \hat{\nu}_{\gamma_1,\gamma_2}) d\lambda_k \end{cases} \tag{94}$$

where $s_{1,\gamma_1}(\cdot)$, $s_{2,\gamma_2}(\cdot)$, $\kappa_{1,\gamma_1}(\cdot)$, $\kappa_{2,\gamma_2}(\cdot)$, $\pi_{1,\gamma_1}(\cdot)$ and $\pi_{2,\gamma_2}(\cdot)$ denote the skew functions, scale functions and mixing densities of GGScM distributions γ_1 and γ_2, respectively, and $\hat{\boldsymbol{\beta}}_{1,\gamma_1,\gamma_2}$, $\hat{\boldsymbol{\beta}}_{2,\gamma_1,\gamma_2}$, $\hat{\mathbf{Q}}_{\gamma_1,\gamma_2}$, $\hat{\mathbf{R}}_{\gamma_1,\gamma_2}$, $\hat{\omega}_{\gamma_1,\gamma_2}$ and $\hat{\nu}_{\gamma_1,\gamma_2}$ denote the approximate estimates of shape parameters, scale matrices and dof parameters of GGScM distributions γ_1 and γ_2, respectively.

Since the state noise $\mathbf{w}_{k-1}$ is independent of the measurement noise $\mathbf{v}_k$, the joint noise PDF $p(\mathbf{w}_{k-1}, \mathbf{v}_k)$ can be formulated as

$$p(\mathbf{w}_{k-1}, \mathbf{v}_k) = p(\mathbf{w}_{k-1}) p(\mathbf{v}_k) \tag{95}$$

Employing (95), the KLD in (93) can be reformulated as

$$\begin{aligned} &\mathrm{KLD}\left\{q_{\gamma_1,\gamma_2}(\mathbf{w}_{k-1}) q_{\gamma_1,\gamma_2}(\mathbf{v}_k) \parallel p(\mathbf{w}_{k-1}, \mathbf{v}_k)\right\} = \int q_{\gamma_1,\gamma_2}(\mathbf{w}_{k-1}) \log \frac{q_{\gamma_1,\gamma_2}(\mathbf{w}_{k-1})}{p(\mathbf{w}_{k-1})} d\mathbf{w}_{k-1} + \\ &\int q_{\gamma_1,\gamma_2}(\mathbf{v}_k) \log \frac{q_{\gamma_1,\gamma_2}(\mathbf{v}_k)}{p(\mathbf{v}_k)} d\mathbf{v}_k = \mathrm{KLD}\left\{q_{\gamma_1,\gamma_2}(\mathbf{w}_{k-1}) \parallel p(\mathbf{w}_{k-1})\right\} + \mathrm{KLD}\left\{q_{\gamma_1,\gamma_2}(\mathbf{v}_k) \parallel p(\mathbf{v}_k)\right\} \end{aligned} \tag{96}$$

Substituting (96) in (93) yields

$$\mathrm{GGScM}_{opt} = \arg\min_{\gamma_1,\gamma_2} \left(\mathrm{KLD}\left\{q_{\gamma_1,\gamma_2}(\mathbf{w}_{k-1}) \parallel p(\mathbf{w}_{k-1})\right\} + \mathrm{KLD}\left\{q_{\gamma_1,\gamma_2}(\mathbf{v}_k) \parallel p(\mathbf{v}_k)\right\}\right) \tag{97}$$

In practical applications, it is very difficult to obtain the optimal selection $\{\gamma_{1,o}, \gamma_{2,o}\}$ of GGScM distributions to model state and measurement noises since the optimization problem in (97) can not be solved analytically. Fortunately, a reasonable selection of GGScM distributions can be obtained by selecting two GGScM distributions from exemplary GGScM distributions in Tables I and II based on off-line analysis, in which the two GGScM distributions have minimum KLD between the approximate joint PDF and the true joint PDF of the state and measurement noises as compared with other GGScM distributions.

Remark 1: In practical applications, some useful prior information is required to select a reasonable GGScM distribution to model a heavy-tailed and/or skew noise. For the proposed KLD-based selection scheme, explicit PDFs or some sample values of noises are required to calculate the KLDs between the approximate PDFs of noises and the true PDFs of noises.

Remark 2: The existing adaptive RTS smoother [31] is a special case of the proposed robust RTS smoothing framework when the state and measurement noises are modelled by a Gaussian distribution. The derivations of the robust RTS smoother based on the Gaussian distribution are given in Appendix G. The existing VB-ST-RTS smoother is also a special case of the proposed robust RTS smoothers since the latter can degrade to the former when the state and measurement noises are modelled by a GGM distribution and the posterior PDFs of mixing parameters are updated using the analytical method.

Remark 3: A robust Kalman filtering framework for a linear state-space model with heavy-tailed and/or skew noises can be also derived based on the proposed GGScM distribution using the variational Bayesian approach. Moreover, the idea of the proposed method can be extended to design a nonlinear filter and a nonlinear smoother for a nonlinear state-space model with heavy-tailed and/or skew noises, and existing Gaussian weighted integral rules can be employed to implement the nonlinear filter and smoother.

VI. Simulation and Experimental Study

In this section, the superior performance of the proposed robust RTS smoothers as compared with existing state-of-the-art RTS smoothers is demonstrated by two representative examples: stochastic volatility model and cooperative localization of an AUV. In the two examples, the proposed robust RTS smoothers are compared with the existing state-of-the-art smoothers including the standard RTS smoother with true mean vectors and covariance matrices of state and measurement noises [36], the skew t-RTS smoother [14], and the VB-ST-RTS smoother [6]. All smoothing algorithms are coded with MATLAB and are executed on a computer with Intel Core i7-6900K CPU @ 3.20 GHz.

A. Stochastic volatility model

The nonlinear discrete-time stochastic volatility model is formulated as [34]

$$\begin{cases} x_k = \gamma_0 + \gamma_1 x_{k-1} + w_{k-1} \\ y_k = \varepsilon_k \exp(x_k/2) \end{cases} \tag{98}$$

where x_k and y_k denote, respectively, the latent log-volatility and obtained asset return at time sample k, and the initial log-volatility x_0, state noise w_k and multiplicative measurement noise ε_k are Gaussian distributed, i.e., $x_0 \sim \mathrm{N}(0, \frac{\Sigma_w}{1-\gamma_1^2})$, $w_k \sim \mathrm{N}(0, \Sigma_w)$ and $\varepsilon_k \sim \mathrm{N}(0, \Sigma_v)$. In this

simulation, the leverage effect is not considered, and then w_k and ε_j are mutually uncorrelated for any time samples k and j. The model parameters are set as $\gamma_0 = 0$, $\gamma_1 = 0.9$, $\Sigma_w = 0.5$ and $\Sigma_v = 1$.

By performing the natural logarithm operation on both sides of measurement equation in (98), the above nonlinear measurement model can be equivalently transformed into a linear measurement model as follows

$$z_k = x_k + v_k, \qquad \text{s.t. } z_k = \log(y_k^2), \qquad v_k = \log(\varepsilon_k^2) \tag{99}$$

According to (98)-(99) and using $\gamma_0 = 0$, the stochastic volatility model can be also written as a linear state-space model in (17), where $F_k = \gamma_1$ and $H_k = 1$. Utilizing the transformation theorem, the PDF of the measurement noise v_k is calculated as

$$p(v_k) = \frac{1}{\sqrt{2\pi}} \exp\left(-0.5 \exp(v_k) + 0.5 v_k\right) \tag{100}$$

The measurement noise v_k has a heavy-tailed and negative skew distribution. The true mean value and variance of the measurement noise v_k are, respectively, $r_t = -1.26$ and $R_t = 4.82$, where r_t and R_t are calculated in the interval $[-20\ 5]$ using the rectangular integration method with step width 0.001.

In this paper, outlier corrupted state noise is produced in terms of

$$w_k \sim \begin{cases} \mathrm{N}(0, \Sigma_w) & \text{w.p. } 0.95 \\ \mathrm{N}(0, 100\Sigma_w) & \text{w.p. } 0.05 \end{cases} \tag{101}$$

where w.p. denotes "with probability". The state noise that is generated according to (101) has a heavy-tailed and symmetric distribution, and its PDF can be formulated as follows

$$p(w_k) = 0.95\mathrm{N}(w_k; 0, \Sigma_w) + 0.05\mathrm{N}(w_k; 0, 100\Sigma_w) \tag{102}$$

and its true mean value and variance are, respectively, $q_t = 0$ and $Q_t = 5.95\Sigma_w$.

To address the heavy-tailed state noise, the GGM, GEM, GBM, GIGM and GIEM distributions are utilized to model state noise, and to handle the heavy-tailed and skew measurement noise, the SGGM, SGEM, SGBM, SGIGM and SGIEM distributions are employed to model the measurement noise. Under the proposed robust RTS smoothing framework, twenty-five different robust RTS smoothers can be obtained based on the five symmetric GGScM distributions and the five skew GGScM distributions. As an example, we select five robust RTS smoothers to

demonstrate the efficiency and superiority of the proposed robust RTS smoothing framework, including the GGM and SGGM-based RTS (GGM-SGGM-RTS) smoother, the GEM and SGEM-based RTS (GEM-SGEM-RTS) smoother, the GBM and SGBM-based RTS (GBM-SGBM-RTS) smoother, the GIGM and SGIGM-based RTS (GIGM-SGIGM-RTS) smoother, and the GIEM and SGIEM-based RTS (GIEM-SGIEM-RTS) smoother. The initial dof parameters of the proposed GGM-SGGM-RTS and GIGM-SGIGM-RTS smoothers are set as $\bar{\omega} = \bar{\nu} = 5$, and the initial dof parameters of the proposed GEM-SGEM-RTS, GBM-SGBM-RTS and GIEM-SGIEM-RTS smoothers are set as $\bar{\omega} = \bar{\nu} = 1$.

In this simulation, the true initial state variable $x_0 = 0$, and the initial error variance $P_0 = \frac{\Sigma_w}{1-\gamma_1^2}$, and the initial state estimate $\hat{x}_{0|0}$ is randomly drawn from $N(x_0, P_0)$, and the simulation time is set as 5000s. The parameters of the standard RTS smoother are selected as $q = q_t$, $r = r_t$, $Q = Q_t$, $R = R_t$, and the parameters of the skew t-RTS smoother are chosen as $q = q_t$, $Q = Q_t$, $R = 1.2$, $\Delta = -1.6$, $\nu = 10$, and the parameters of the VB-ST-RTS smoother are set as $q = 0$, $r = 0$, $a_0 = c_0 = 5$, $b_0 = d_0 = 1$, $t_0 = u_0 = 3$, $T_0 = \Sigma_w$, $U_0 = \Sigma_v$, and the parameters of the proposed robust RTS smoothers are selected as $q = 0$, $\bar{Q} = \Sigma_w$, $\bar{R} = \Sigma_v$, $\epsilon = 10^{-8}$, $M = 1000$, $N_m = 100$. Note that the parameters of the skew t-distribution are selected off-line by minimizing the KLD between the skew t-PDF and the PDF of the measurement noise.

To compare the estimation accuracy of existing smoothers and the proposed smoothers, the root mean square error (RMSE) and the averaged RMSE (ARMSE) of the state variable are selected as performance metrics defined as follows

$$\begin{cases} \mathrm{RMSE}(k) = \sqrt{\frac{1}{M_c}\sum\limits_{s=1}^{M_c}\left(x_k^s - \hat{x}_{k|T}^s\right)^2} \\ \mathrm{ARMSE} = \frac{1}{T}\sum\limits_{k=1}^{T}\sqrt{\frac{1}{M_c}\sum\limits_{s=1}^{M_c}\left(x_k^s - \hat{x}_{k|T}^s\right)^2} \end{cases} \tag{103}$$

where $\mathrm{RMSE}(k)$ denotes the RMSE at time sample k, and x_k^s and $\hat{x}_{k|T}^s$ are, respectively, the true and estimated state variable at the s-th Monte Carlo run, and M_c and T, respectively, represent the total number of Monte Carlo runs and the simulation steps that are set as $M_c = 1000$ and $T = 5000$. To clearly exhibit the RMSEs of existing RTS smoothers and the proposed robust RTS smoothers in Fig. 5 and Fig. 8, the RMSEs of all RTS smoohters are smoothed using a moving average method with span of 100s.

The RMSEs, ARMSEs and implementation times of the proposed robust RTS smoothers based

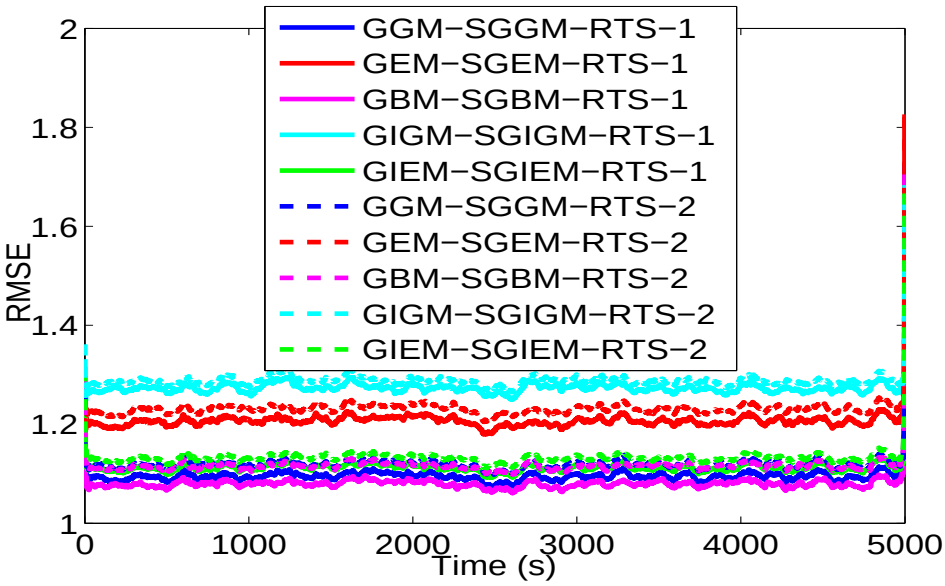


Fig. 5: RMSEs of the proposed robust RTS smoothers based on different GGScM distributions.

TABLE IX: ARMSEs and implementation times of the proposed robust RTS smoothers based on different GGScM distributions.

Smoothers	GGM-SGGM-RTS-1	GEM-SGEM-RTS-1	GBM-SGBM-RTS-1	GIGM-SGIGM-RTS-1	GIEM-SGIEM-RTS-1
ARMSE	1.10	1.21	1.08	1.27	1.11
Time (s)	15.28	13.25	101.81	18.53	17.56
Smoothers	GGM-SGGM-RTS-2	GEM-SGEM-RTS-2	GBM-SGBM-RTS-2	GIGM-SGIGM-RTS-2	GIEM-SGIEM-RTS-2
ARMSE	1.12	1.23	1.11	1.29	1.13
Time (s)	79.94	52.55	141.91	65.73	74.88

on different GGScM distributions are, respectively, shown in Fig. 5 and Table IX, where "1" and "2" represent that the posterior PDFs of mixing parameters in the proposed robust RTS smoothers are updated using analytical method and Monte Carlo approach, respectively. It is seen from Fig. 5 and Table IX that the proposed robust RTS-1 smoothers have slightly better estimation accuracy than the proposed robust RTS-2 smoothers, which is induced by the fact that the mixing parameters can be better updated by the analytical method as compared with the Motel Carlo approach. We can see from Fig. 5 and Table IX that the proposed GGM-SGGM-RTS-1

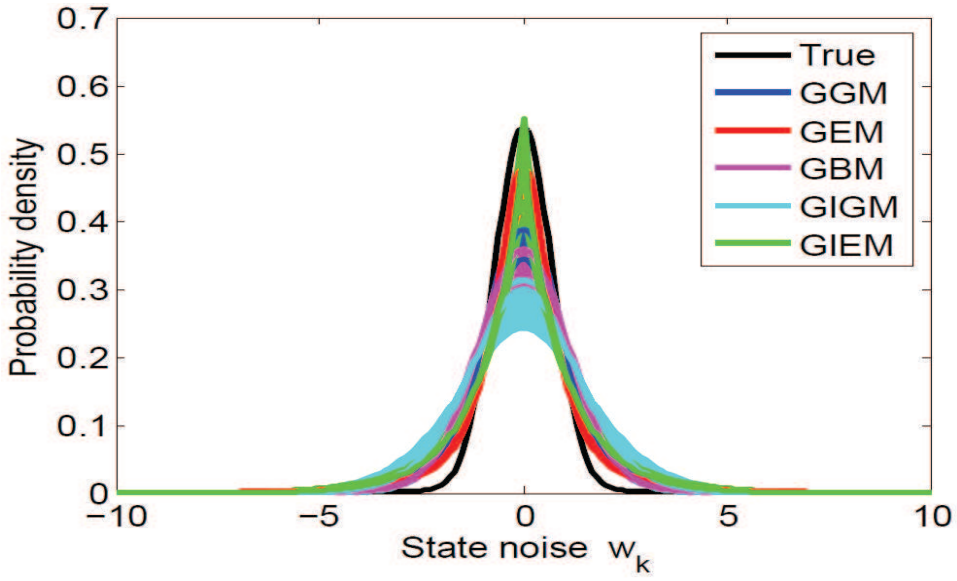


Fig. 6: True PDF and different GGScM approximations of the state noise in 1000 Monte Carlo runs.

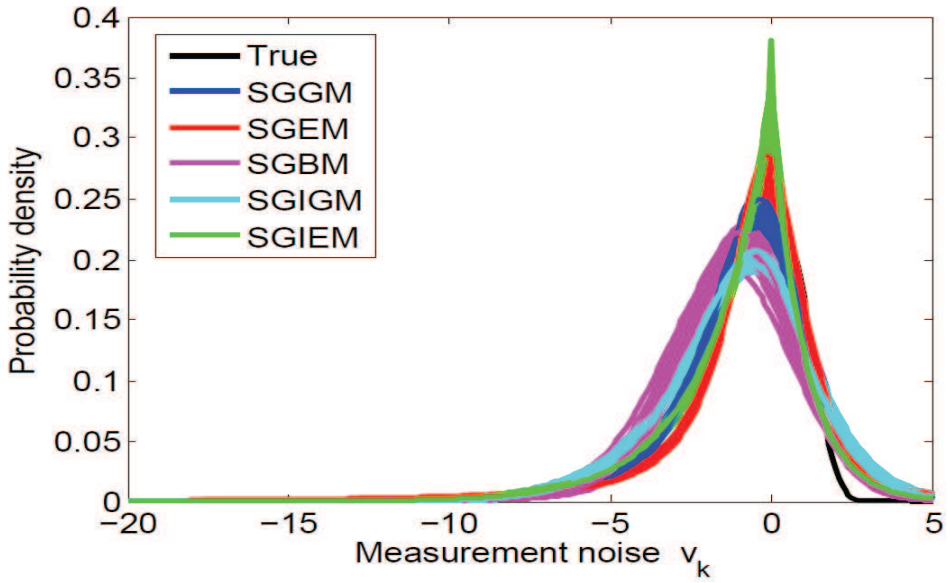


Fig. 7: True PDF and different GGScM approximations of the measurement noise in 1000 Monte Carlo runs.

smoother has better estimation accuracy than other proposed robust RTS smoothers except for the proposed GBM-SGBM-RTS-1 smoother, and the proposed GGM-SGGM-RTS-1 smoother

TABLE X: Averaged KLDs between the true PDFs and different GGScM approximations of the state and measurement noises.

PDFs	GGM	GEM	GBM	GIGM	GIEM
KLD_w	0.21	0.14	0.26	0.44	0.28
PDFs	SGGM	SGEM	SGBM	SGIGM	SGIEM
KLD_v	0.37	0.54	0.29	0.55	0.36
PDFs	GGM-SGGM	GEM-SGEM	GBM-SGBM	GIGM-SGIGM	GIEM-SGIEM
KLD_{wv}	0.58	0.68	0.55	0.99	0.64

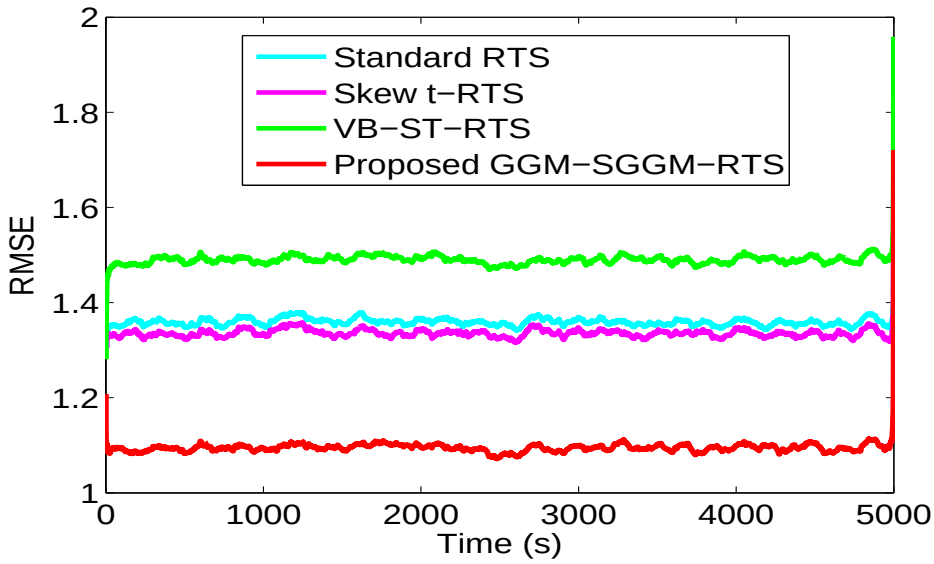


Fig. 8: RMSEs of existing RTS smoothers and the proposed GBM-SGBM-RTS smoother.

TABLE XI: ARMSEs and implementation times of existing RTS smoothers and the proposed GGM-SGGM-RTS smoother.

Smoothers	Standard RTS	Skew t-RTS	VB-ST-RTS	GGM-SGGM-RTS
ARMSE	1.36	1.34	1.49	1.10
Time (s)	0.28	13.43	5.63	15.28

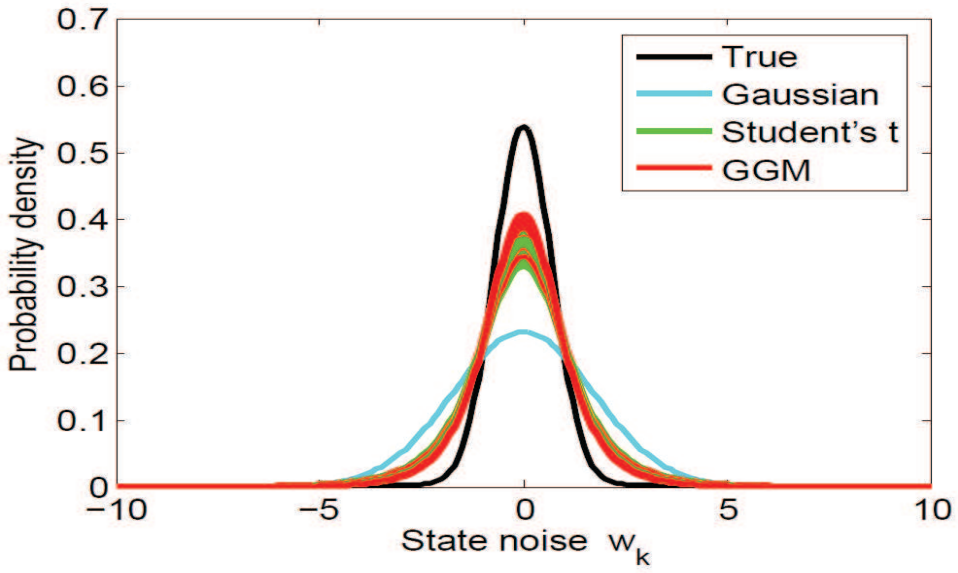


Fig. 9: True PDF, Gaussian, Student's t and GGM approximations of the state noise in 1000 Monte Carlo runs.

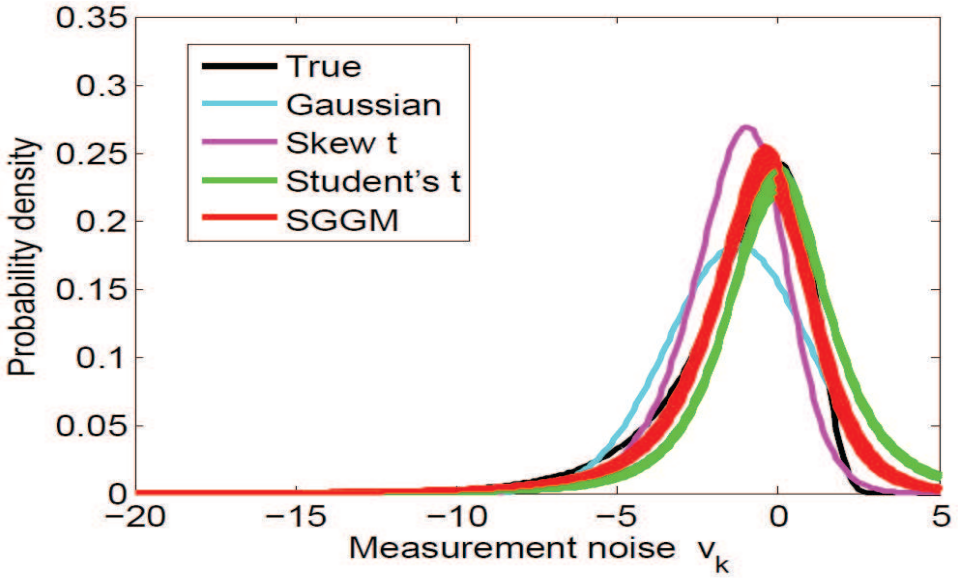


Fig. 10: True PDF, Gaussian, Student's t, skew t and SGGM approximations of the measurement noise in 1000 Monte Carlo runs.

has slightly worse estimation accuracy but significantly lower computational complexity than the proposed GBM-SGBM-RTS-1 smoother.

TABLE XII: Averaged KLDs between the true PDFs and different approximations of the state and measurement noises.

PDFs	Gaussian	Gaussian	Student's t	GGM
KLD_w	0.69	0.69	0.23	0.21
PDFs	Gaussian	Skew t	Student's t	SGGM
KLD_v	0.55	0.11	1.29	0.37
PDFs	Gaussian-Gaussian	Gaussian-skew t	Student's t-Student's t	GGM-SGGM
KLD_{wv}	1.24	0.80	1.52	0.58

Fig. 6 and Fig. 7, respectively, show the true PDFs and different GGScM approximations of the state and measurement noises in 1000 Monte Carlo runs, where the parameters of the GGScM distributions are adaptively learned based on the proposed method. The averaged KLDs between the true PDFs and different GGScM approximations of the state and measurement noises are listed in Table X, where KLD_w denotes the averaged KLD between the true PDF and approximate PDF of the state noise, and KLD_v denotes the averaged KLD between the true PDF and approximate PDF of the measurement noise, and $\mathrm{KLD}_{wv} = \mathrm{KLD}_w + \mathrm{KLD}_v$ denotes the averaged KLD between the true joint PDF and approximate joint PDF of the state and measurement noises. It is observed from Fig. 6, Fig. 7 and Table X that the GGM-SGGM distribution can better jointly model the state and measurement noises than the GEM-SGEM, GIGM-SGIGM and GIEM-SGIEM distributions but slightly worse than the GBM-SGBM distribution. We can also observe from Tables IX and X that with better joint modelling of the state and measurement noises, better estimation accuracy can be achieved by the proposed robust RTS smoothers. Based on the above discussions, the GGM and SGGM distributions will be, respectively, employed to model the state and measurement noises due to their good modelling accuracy and reasonable computational complexity, and the proposed GGM-SGGM-RTS-1 smoother will be compared with existing state-of-the-art RTS smoothers.

The RMSEs, ARMSEs and implementation times of existing RTS smoothers and the proposed GGM-SGGM-RTS smoother are, respectively, shown in Fig. 8 and Table XI. It is seen from Fig. 8 and Table XI that the proposed GGM-SGGM-RTS smoother has better estimation accuracy

but higher computational complexity than existing state-of-the-art RTS smoothers. Fig. 9 and Fig. 10, respectively, show the true PDFs and different approximations of state and measurement noises in 1000 Monte Carlo runs, and the averaged KLDs between the true PDFs and different approximations of state and measurement noises are listed in Table XII. It can be seen from Fig. 9, Fig. 10 and Table XII that the GGM-SGGM distribution can better jointly model state and measurement noises than the Gaussian-Gaussian, Gaussian-skew t, and Student's t-Student's t distributions, which results in improved estimation accuracy as compared with existing state-of-the-art RTS smoothers.

B. Cooperative localization experiments

A master-slave cooperative localization experiment is used to illustrate the performance of the proposed robust RTS smoothers, in which the master AUV assists the slave AUV using acoustic range measurement. In the experiment, two survey vessels were used to serve as surface leaders, and one survey vessel was employed to act as a surrogate AUV, and the three survey vessels were all equipped with acoustic modem ATM-885 so that the surface leaders and the AUV can broadcast information mutually. High-accuracy differential GPSs were installed in the two surface leaders, and the AUV was equipped with a low-cost dead-reckoning system that is composed by a Doppler velocity log (DVL) and a self-made compass. In order to provide a benchmark for cooperative localization, the AUV was also equipped with a high-accuracy differential GPS which can collect true positions of the AUV. The employed sensors and computer in the test are illustrated in Fig. 11, and their performance parameters are listed in Table XIII, where RMS represents root mean square. In the experiment, in order to improve the observability of the cooperative localization of an AUV, the AUV was always located between the two surface leaders, and only one surface leader communicated with the AUV at every time, where the acoustic data packets were sent from the two surface leaders to the AUV every 5s and staggered in time.

Based on the above experimental descriptions, the discrete-time state-space model of the cooperative localization of an AUV is formulated as [27]

$$\mathbf{x}_k = \mathbf{F}\mathbf{x}_{k-1} + \mathbf{u}_k + \mathbf{w}_{k-1} \tag{104}$$

$$z_k = \sqrt{(x_k - x_k^r)^2 + (y_k - y_k^r)^2} + v_k \tag{105}$$

TABLE XIII: The performance parameters of employed sensors.

Sensors	Index	Parameters
Acoustic modem	Working range	Up to 8000m
	Error rate	Less than 10^{-7}
GPS	Position accuracy	1.8m (RMS)
	Data update rate	10Hz
DVL	Working range	$-150\text{m/s} - 200\text{m/s}$
	Measurement accuracy	$0.1\% - 0.3\%$
Compass	Heading accuracy	0.3°

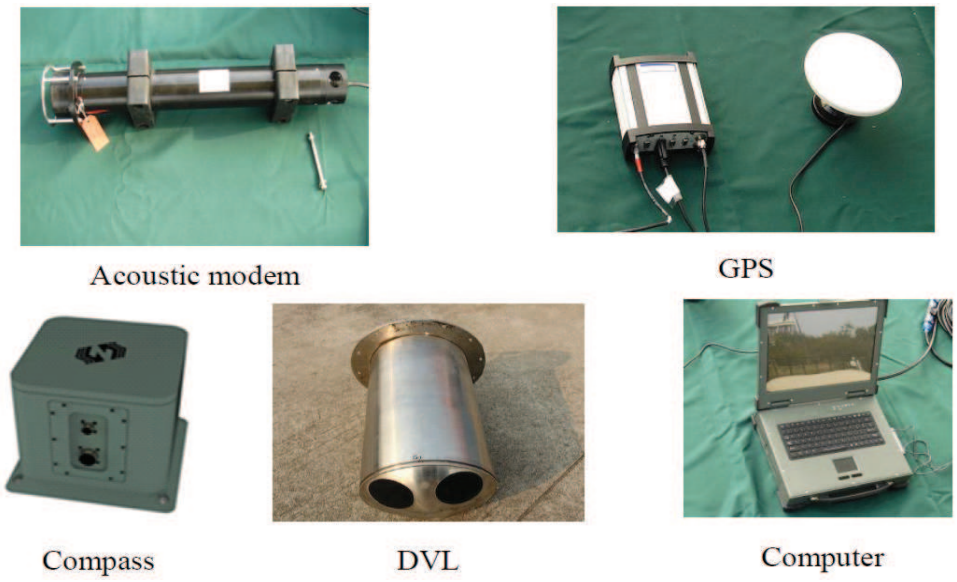


Fig. 11: The employed sensors and computer in the test.

where $\mathbf{x}_k \triangleq [x_k,\ y_k]^T$ denotes the state vector; and z_k denotes the range measurement at time sample k; and $\mathbf{F} = \mathbf{I}_2$ denotes the state transition matrix; and $\mathbf{u}_k = [\Delta t(\hat{v}_k \cos\hat{\theta}_k + \hat{\omega}_k \sin\hat{\theta}_k),\ \Delta t(\hat{v}_k \sin\hat{\theta}_k - \hat{\omega}_k \cos\hat{\theta}_k)]^T$ denotes the control input; and $\mathbf{w}_k = [w_{x,k}, w_{y,k}]^T$ denotes the state noise vector at time sample k; and v_k denotes the measurement noise at time sample k; and (x_k, y_k) and (x_k^r, y_k^r) denote, respectively, the east and north positions of the AUV and leader at time sample k; and $\hat{\omega}_k$ and $\hat{v}_k$ denote, respectively, the starboard and forward velocities

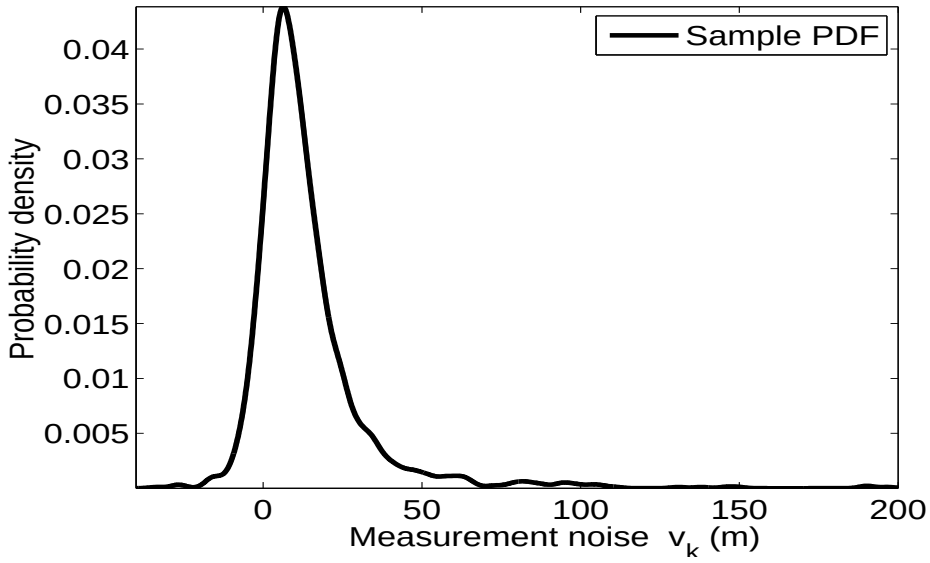


Fig. 12: Approximate probability density curve of the measurement noise.

TABLE XIV: ALEs and implementation times of the proposed robust RTS smoothers based on different skew GGScM distributions.

Smoothers	SGGM-RTS-1	SGEM-RTS-1	SGBM-RTS-1	SGIGM-RTS-1	SGIEM-RTS-1
ALE(m)	13.7	13.8	10.7	12.5	11.8
Time (s)	24.14	17.69	30.09	14.92	17.09
Smoothers	SGGM-RTS-2	SGEM-RTS-2	SGBM-RTS-2	SGIGM-RTS-2	SGIEM-RTS-2
ALE(m)	13.7	13.8	11.1	12.8	12.5
Time (s)	101.84	101.62	102.12	101.84	101.90

measured by the DVL in the body framework at time sample k; and $\hat{\theta}_k$ denotes the heading angle provided by the self-made compass at time sample k; and Δt denotes the sampling time. In the experiment, the sampling time is set as $\Delta t = 1$s, and the experimental time is $T = 1600$s.

The sample values of measurement noise is approximately calculated as [27]

$$\hat{v}_k = z_k - \sqrt{(\hat{x}_k - x_k^r)^2 + (\hat{y}_k - y_k^r)^2} \tag{106}$$

where $(\hat{x}_k, \hat{y}_k)$ represents the true position of the AUV at time sample k that is provided by

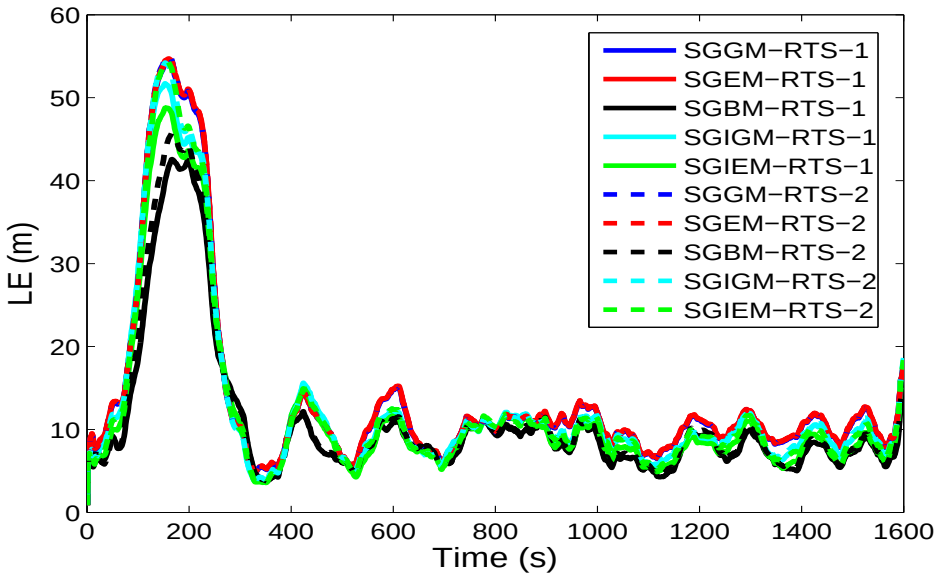


Fig. 13: LEs of the proposed robust RTS smoothers based on different skew GGScM distributions.

TABLE XV: Averaged KLDs between the sample PDF and different skew GGScM approximations of measurement noise.

PDFs	SGGM	SGEM	SGBM	SGIGM	SGIEM
KLD_w	0.27	0.29	0.08	0.17	0.10

a high-accuracy differential GPS. Employing these sample values, an approximate probability density curve of measurement noise is obtained off-line, as is shown in Fig. 12. It can be observed from Fig. 12 that the sample values of measurement noise has a heavy-tailed and skew distribution. The cooperative localization of an AUV has a heavy-tailed and skew measurement noise, and can be used to illustrate the performance of the proposed robust RTS smoothers.

In the experiment, the initial state estimate $\hat{\mathbf{x}}_{0|0}$ is provided by a high-accuracy differential GPS which is installed in the surrogate AUV, and the initial state estimation error covariance matrix is set as $\mathbf{P}_{0|0} = (1.8\mathrm{m})^2\mathbf{I}_2$ according to the RMS of the employed GPS. The parameters of the standard RTS smoother are selected as $\mathbf{q} = \mathbf{q}_t$, $r = r_t$, $\mathbf{Q} = \mathbf{Q}_t$, $R = R_t$, the parameters of the skew t-RTS smoother are chosen as $\mathbf{q} = \mathbf{q}_t$, $\mathbf{Q} = \mathbf{Q}_t$, $R = 100\mathrm{m}^2$, $\Delta = 5\mathrm{m}$, $\nu = 5$, and the parameters of the VB-ST-RTS smoother are set as $\mathbf{q} = \mathbf{q}_t$, $\mathbf{Q} = \mathbf{Q}_t$, $a_0 = 5$, $b_0 = 1$,

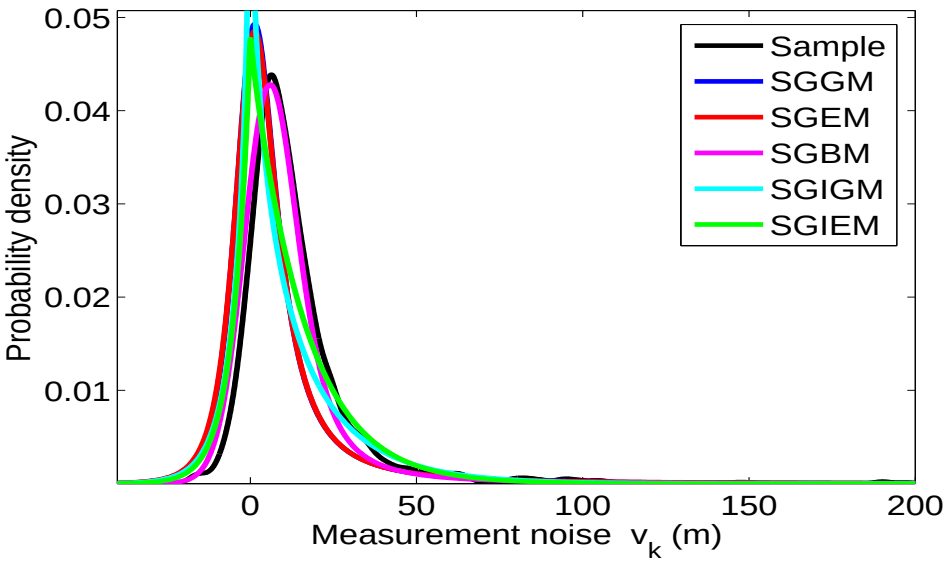


Fig. 14: Sample PDF and different skew GGScM approximations of the measurement noise.

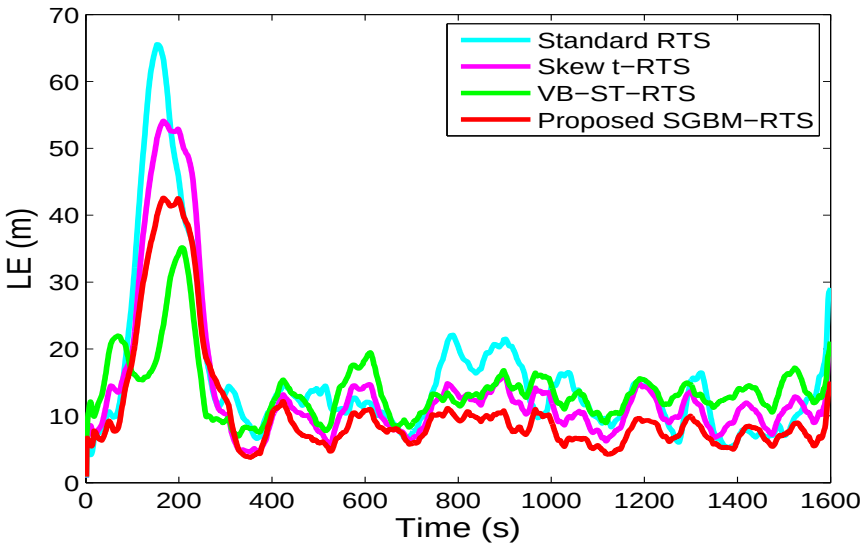


Fig. 15: LEs of existing RTS smoothers and the proposed SGBM-RTS smoother.

$u_0 = 3$, $U_0 = 100\mathrm{m}^2$, and the parameters of the proposed robust RTS smoothers are selected as $\mathbf{q} = \mathbf{q}_t$, $\mathbf{Q} = \mathbf{Q}_t$, $\bar{R} = 100\mathrm{m}^2$, $\epsilon = 10^{-8}$, $M = 1000$, $N_m = 100$, where the true mean vectors and covariance matrices of state and measurement noises are approximately calculated as $\mathbf{q}_t \approx [0\mathrm{m} \;\; 0\mathrm{m}]^{\mathrm{T}}$, $r_t \approx 14.1\mathrm{m}$, $\mathbf{Q}_t \approx \mathrm{diag}([(2.9\mathrm{m})^2 \;\; (2.3\mathrm{m})^2])$, and $R_t \approx (24.8\mathrm{m})^2$ by using

TABLE XVI: ALEs (0s-1600s), steady state ALEs (350s-1600s) and implementation times of existing RTS smoothers and the proposed SGBM-RTS smoother.

Smoothers	Standard RTS	Skew t-RTS	VB-ST-RTS	SGBM-RTS
ALE (m) (0s-1600s)	15.5	13.8	13.9	10.7
Steady state ALE (m) (350s-1600s)	12.2	10.6	12.9	7.8
Time (s)	0.33	5.19	13.39	30.09

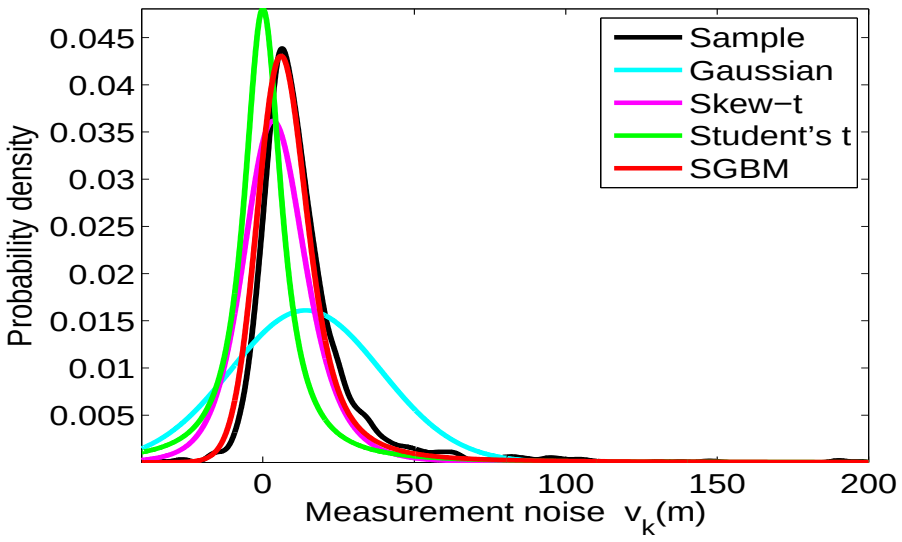


Fig. 16: Sample PDF and Gaussian, Student's t, skew t and SGBM approximations of the measurement noise.

TABLE XVII: Averaged KLDs between the sample PDF and Gaussian, Student's t, skew t and SGBM approximations of the measurement noise.

PDFs	Gaussian	Student's t	Skew t	SGBM
KLD_v	0.54	0.55	0.26	0.08

the obtained sample values of state and measurement noises.

To compare the estimation accuracy of the proposed robust RTS smoothers and the existing state-of-the-art RTS smoothers, the localization error (LE) and averaged LE (ALE) are chosen

as performance metrics, which are defined as follows [27]

$$\begin{cases} \mathrm{LE}(k) = \sqrt{(\hat{x}_k - \hat{x}_{k|k})^2 + (\hat{y}_k - \hat{y}_{k|k})^2} \\ \mathrm{ALE} = \frac{1}{T}\sum\limits_{k=1}^{T} \sqrt{(\hat{x}_k - \hat{x}_{k|k})^2 + (\hat{y}_k - \hat{y}_{k|k})^2} \end{cases} \tag{107}$$

where $(\hat{x}_k, \hat{y}_k)$ represents the reference position of the AUV at time sample k, and $(\hat{x}_{k|k}, \hat{y}_{k|k})$ represents the position estimate at time sample k, and $T = 1600\mathrm{s}$ denotes the experimental time. To better exhibit the LEs of all RTS smoothers in Fig. 13 and Fig. 15, the LEs are smoothed using a moving average method with span of 50s.

The LEs, ALEs and implementation times of the proposed robust RTS smoothers based on different skew GGScM distributions are, respectively, shown in Fig. 13 and Table XIV. It is observed from Fig. 13 and Table XIV that the proposed robust RTS-1 smoothers and the proposed robust RTS-2 smoothers have almost identical estimation accuracy, but the proposed robust RTS-2 smoothers have higher computational complexities than the proposed robust RTS-1 smoothers. Thus, the mixing parameters can be accurately updated by the Monte Carlo approach with 1000 random samples. We can also observe from Fig. 13 and Table XIV that the proposed SGBM-RTS-1 smoother has better estimation accuracy than other proposed robust RTS smoothers, and the proposed SGBM-RTS-1 smoother has reasonable computational complexity.

The sample PDF and different skew GGScM approximations of measurement noise are illustrated in Fig. 14, and the averaged KLDs between the sample PDF and different skew GGScM approximations of measurement noise are listed in Table XV. It can be observed from Fig. 14 and Table XV that the SGBM distribution can better model measurement noise than the SGGM, SGEM, SGIGM and SGIEM distributions. Also, we can observe from Table XIV and XV that with better modelling of measurement noise, better estimation accuracy can be achieved by the proposed robust RTS smoothers. Based on the above discussions, the SGBM distribution will be utilized to model measurement noise, and the proposed SGBM-RTS-1 smoother will be compared with existing state-of-the-art RTS smoothers.

The LEs, ALEs (0s-1600s), steady state ALEs (350s-1600s) and implementation times of existing RTS smoothers and the proposed SGBM-RTS smoother are, respectively, shown in Fig. 15 and Table XVI. It is seen from Fig. 15 and Table XVI that the proposed SGBM-RTS smoother has better estimation accuracy but higher computational complexity than existing RTS smoothers. The sample PDF and Gaussian, Student's t, skew t and SGBM approximations of measurement

noise are illustrated in Fig. 16, and averaged KLDs between the sample PDF and Gaussian, Student's t, skew t and SGBM approximations of measurement noise are listed in Table XVII. It can be seen from Fig. 16 and Table XVII that the SGBM distribution can better model heavy-tailed and skew measurement noise than the Gaussian, Student's t, skew t-distributions, which leads to improved estimation accuracy.

VII. Conclusions

This paper proposed a GGScM distribution which can be formulated as a hierarchical Gaussian form conditioned on a random mixing parameter that follows a continuous probability distribution with a positive orthant as support, for which the existing GScM distribution is a special case of the proposed GGScM distribution. As such, a major contribution of this work is to provide a unified family of RTS smoothers on the basis of the exemplary GGScM distribution framework for a linear state-space model with heavy-tailed and/or skew state and measurement noises. The state trajectory, mixing parameters and unknown distribution parameters were jointly inferred using the VB approach. Both the existing adaptive RTS smoother [31] and VB-ST-RTS smoother [6] are special cases of the proposed robust RTS smoothing framework when the state and measurement noises are, respectively, modelled by a Gaussian distribution and a GGM distribution.

Several particular solutions corresponding to the exemplary GGScM distributions were derived, and the robustness analyses were provided to reveal the advantages of the proposed method. Moreover, a new KLD-based scheme was proposed to facilitate the selection of GGScM distributions in practical applications. The proposed robust RTS smoothers and existing state-of-the-art smoothers were compared by two representative examples: stochastic volatility model and cooperative localization of an AUV. Simulation and experimental results showed that the proposed robust RTS smoothers have better estimation accuracy but higher computational complexities than existing state-of-the-art smoothers.

VIII. APPENDICES

A. Derivations of (27)-(28)

Exploiting (26), we have

$$\log p(\boldsymbol{\Theta}, \mathbf{z}_{1:T}) = \log \mathrm{g}(\mathbf{x}_0; \hat{\mathbf{x}}_{0|0}, \mathbf{P}_{0|0}) + 0.5n \sum_{k=1}^{T} \log \kappa_1(\xi_k) - 0.5T \log |\mathbf{Q}| -$$
$$0.5 \sum_{k=1}^{T} \kappa_1(\xi_k)(\mathbf{x}_k - \mathbf{F}_k \mathbf{x}_{k-1} - \boldsymbol{\beta}_1/s_1(\xi_k))^{\mathrm{T}} \mathbf{Q}^{-1} (\mathbf{x}_k - \mathbf{F}_k \mathbf{x}_{k-1} - \boldsymbol{\beta}_1/s_1(\xi_k)) +$$
$$0.5m \sum_{k=1}^{T} \log \kappa_2(\lambda_k) - 0.5T \log |\mathbf{R}| - 0.5 \sum_{k=1}^{T} \kappa_2(\lambda_k)(\mathbf{z}_k - \mathbf{H}_k \mathbf{x}_k - \boldsymbol{\beta}_2/s_2(\lambda_k))^{\mathrm{T}} \times$$
$$\mathbf{R}^{-1}(\mathbf{z}_k - \mathbf{H}_k \mathbf{x}_k - \boldsymbol{\beta}_2/s_2(\lambda_k)) + \sum_{k=1}^{T} \log \pi_1(\xi_k; \omega) + \sum_{k=1}^{T} \log \pi_2(\lambda_k; \nu) + c_{\boldsymbol{\Theta}} \tag{108}$$

Substituting $\theta = \mathbf{x}_{0:T}$ in (25) and using (108) gives

$$\log q^{(i+1)}(\mathbf{x}_{0:T}) = \log \mathrm{g}(\mathbf{x}_0; \hat{\mathbf{x}}_{0|0}, \mathbf{P}_{0|0}) - 0.5 \sum_{k=1}^{T} \mathrm{E}^{(i)}[\kappa_1(\xi_k)(\mathbf{x}_k - \mathbf{F}_k \mathbf{x}_{k-1} - \boldsymbol{\beta}_1/s_1(\xi_k))^{\mathrm{T}} \mathbf{Q}^{-1} \times$$
$$(\mathbf{x}_k - \mathbf{F}_k \mathbf{x}_{k-1} - \boldsymbol{\beta}_1/s_1(\xi_k))] - 0.5 \sum_{k=1}^{T} \mathrm{E}^{(i)}[\kappa_2(\lambda_k)(\mathbf{z}_k - \mathbf{H}_k \mathbf{x}_k - \boldsymbol{\beta}_2/s_2(\lambda_k))^{\mathrm{T}} \mathbf{R}^{-1} \times$$
$$(\mathbf{z}_k - \mathbf{H}_k \mathbf{x}_k - \boldsymbol{\beta}_2/s_2(\lambda_k))] + c_{\mathbf{x}_{0:T}} \tag{109}$$

The expectations in (109) are, respectively, calculated as

$$\mathrm{E}^{(i)}[\kappa_1(\xi_k)(\mathbf{x}_k - \mathbf{F}_k \mathbf{x}_{k-1} - \boldsymbol{\beta}_1/s_1(\xi_k))^{\mathrm{T}} \mathbf{Q}^{-1} (\mathbf{x}_k - \mathbf{F}_k \mathbf{x}_{k-1} - \boldsymbol{\beta}_1/s_1(\xi_k))] = (\mathbf{x}_k - \mathbf{F}_k \mathbf{x}_{k-1})^{\mathrm{T}} \times$$
$$\left(\tilde{\mathbf{Q}}_k^{(i)}\right)^{-1} (\mathbf{x}_k - \mathbf{F}_k \mathbf{x}_{k-1}) - \left(\tilde{\mathbf{q}}_k^{(i)}\right)^{\mathrm{T}} \left(\tilde{\mathbf{Q}}_k^{(i)}\right)^{-1} (\mathbf{x}_k - \mathbf{F}_k \mathbf{x}_{k-1}) - (\mathbf{x}_k - \mathbf{F}_k \mathbf{x}_{k-1})^{\mathrm{T}} \left(\tilde{\mathbf{Q}}_k^{(i)}\right)^{-1} \tilde{\mathbf{q}}_k^{(i)}$$
$$+ c_{\mathbf{x}_{0:T}} \tag{110}$$

$$\mathrm{E}^{(i)}[\kappa_2(\lambda_k)(\mathbf{z}_k - \mathbf{H}_k \mathbf{x}_k - \boldsymbol{\beta}_2/s_2(\lambda_k))^{\mathrm{T}} \mathbf{R}^{-1} (\mathbf{z}_k - \mathbf{H}_k \mathbf{x}_k - \boldsymbol{\beta}_2/s_2(\lambda_k))] = (\mathbf{z}_k - \mathbf{H}_k \mathbf{x}_k)^{\mathrm{T}} \times$$
$$\left(\tilde{\mathbf{R}}_k^{(i)}\right)^{-1} (\mathbf{z}_k - \mathbf{H}_k \mathbf{x}_k) - \left(\tilde{\mathbf{r}}_k^{(i)}\right)^{\mathrm{T}} \left(\tilde{\mathbf{R}}_k^{(i)}\right)^{-1} (\mathbf{z}_k - \mathbf{H}_k \mathbf{x}_k) - (\mathbf{z}_k - \mathbf{H}_k \mathbf{x}_k)^{\mathrm{T}} \left(\tilde{\mathbf{R}}_k^{(i)}\right)^{-1} \tilde{\mathbf{r}}_k^{(i)} +$$
$$c_{\mathbf{x}_{0:T}} \tag{111}$$

where the modified mean vectors $\tilde{\mathbf{q}}_k^{(i)}$ and $\tilde{\mathbf{r}}_k^{(i)}$ and covariance matrices $\tilde{\mathbf{Q}}_k^{(i)}$ and $\tilde{\mathbf{R}}_k^{(i)}$ of state and measurement noises are given in (17).

Employing (110)-(111) in (109), $\log q^{(i+1)}(\mathbf{x}_{0:T})$ is reformulated as

$$\log q^{(i+1)}(\mathbf{x}_{0:T}) = \log \mathrm{g}(\mathbf{x}_0; \hat{\mathbf{x}}_{0|0}, \mathbf{P}_{0|0}) + \sum_{k=1}^{T} \log \mathrm{g}(\mathbf{x}_k; \mathbf{F}_k \mathbf{x}_{k-1} + \tilde{\mathbf{q}}_k^{(i)}, \tilde{\mathbf{Q}}_k^{(i)}) + \sum_{k=1}^{T} \log \mathrm{g}(\mathbf{z}_k; \mathbf{H}_k \mathbf{x}_k + \tilde{\mathbf{r}}_k^{(i)}, \tilde{\mathbf{R}}_k^{(i)}) + c_{\mathbf{x}_{0:T}} \tag{112}$$

Exploiting (112), $q^{(i+1)}(\mathbf{x}_{0:T})$ is calculated as

$$q^{(i+1)}(\mathbf{x}_{0:T}) = \mathrm{g}(\mathbf{x}_0; \hat{\mathbf{x}}_{0|0}, \mathbf{P}_{0|0}) \prod_{k=1}^{T} \mathrm{g}(\mathbf{x}_k; \mathbf{F}_k \mathbf{x}_{k-1} + \tilde{\mathbf{q}}_k^{(i)}, \tilde{\mathbf{Q}}_k^{(i)}) \mathrm{g}(\mathbf{z}_k; \mathbf{H}_k \mathbf{x}_k + \tilde{\mathbf{r}}_k^{(i)}, \tilde{\mathbf{R}}_k^{(i)}) \tag{113}$$

It is observed from (113) that the posterior PDF $q^{(i+1)}(\mathbf{x}_{0:T})$ can be deemed as a smoothing PDF for the linear state-space model with Gaussian state and measurement noises, i.e., $\mathbf{w}_k \sim \mathrm{N}(\tilde{\mathbf{q}}_k^{(i)}, \tilde{\mathbf{Q}}_k^{(i)})$ and $\mathbf{v}_k \sim \mathrm{N}(\tilde{\mathbf{r}}_k^{(i)}, \tilde{\mathbf{R}}_k^{(i)})$. Then, the posterior PDF $q^{(i+1)}(\mathbf{x}_{0:T})$ can be updated as Gaussian in (27) by the standard RTS smoother [36].

B. Derivations of (29)-(32)

Substituting $\theta = \boldsymbol{\beta}_1$ and $\theta = \boldsymbol{\beta}_2$ in (25) and using (108) yields

$$\begin{cases} \log q^{(i+1)}(\boldsymbol{\beta}_1) = -0.5 \sum_{k=1}^{T} \mathrm{E}^{(i)}[\kappa_1(\xi_k)(\mathbf{x}_k - \mathbf{F}_k \mathbf{x}_{k-1} - \boldsymbol{\beta}_1 / s_1(\xi_k))^{\mathrm{T}} \mathbf{Q}^{-1} \times \\ (\mathbf{x}_k - \mathbf{F}_k \mathbf{x}_{k-1} - \boldsymbol{\beta}_1 / s_1(\xi_k))] + c_{\boldsymbol{\beta}_1} \\ \log q^{(i+1)}(\boldsymbol{\beta}_2) = -0.5 \sum_{k=1}^{T} \mathrm{E}^{(i)}[\kappa_2(\lambda_k)(\mathbf{z}_k - \mathbf{H}_k \mathbf{x}_k - \boldsymbol{\beta}_2 / s_2(\lambda_k))^{\mathrm{T}} \mathbf{R}^{-1} \times \\ (\mathbf{z}_k - \mathbf{H}_k \mathbf{x}_k - \boldsymbol{\beta}_2 / s_2(\lambda_k))] + c_{\boldsymbol{\beta}_2} \end{cases} \tag{114}$$

Employing (30)-(32) in (114), $\log q^{(i+1)}(\boldsymbol{\beta}_1)$ and $\log q^{(i+1)}(\boldsymbol{\beta}_2)$ can be rewritten as

$$\begin{cases} \log q^{(i+1)}(\boldsymbol{\beta}_1) = -0.5 \left(\boldsymbol{\beta}_1 - \hat{\boldsymbol{\beta}}_1^{(i+1)}\right)^{\mathrm{T}} \left(\mathbf{P}_{\boldsymbol{\beta}_1}^{(i+1)}\right)^{-1} \left(\boldsymbol{\beta}_1 - \hat{\boldsymbol{\beta}}_1^{(i+1)}\right) + c_{\boldsymbol{\beta}_1} \\ \log q^{(i+1)}(\boldsymbol{\beta}_2) = -0.5 \left(\boldsymbol{\beta}_2 - \hat{\boldsymbol{\beta}}_2^{(i+1)}\right)^{\mathrm{T}} \left(\mathbf{P}_{\boldsymbol{\beta}_2}^{(i+1)}\right)^{-1} \left(\boldsymbol{\beta}_2 - \hat{\boldsymbol{\beta}}_2^{(i+1)}\right) + c_{\boldsymbol{\beta}_2} \end{cases} \tag{115}$$

where $\hat{\boldsymbol{\beta}}_1^{(i+1)}$, $\hat{\boldsymbol{\beta}}_2^{(i+1)}$, $\mathbf{P}_{\boldsymbol{\beta}_1}^{(i+1)}$ and $\mathbf{P}_{\boldsymbol{\beta}_2}^{(i+1)}$ are, respectively, given by (30)-(31). According to (115), we can achieve (29).

C. Derivations of (48)-(51)

Substituting $\theta = \mathbf{Q}$ and $\theta = \mathbf{R}$ in (25), respectively, and using (108) yields

$$\begin{cases} \log q^{(i+1)}(\mathbf{Q}) = -0.5T \log |\mathbf{Q}| - 0.5 \mathrm{tr}\left\{\mathbf{E}^{i+1} \mathbf{Q}^{-1}\right\} \\ \log q^{(i+1)}(\mathbf{R}) = -0.5T \log |\mathbf{R}| - 0.5 \mathrm{tr}\left\{\mathbf{F}^{i+1} \mathbf{R}^{-1}\right\} \end{cases} \tag{116}$$

where $\mathbf{E}^{i+1}$ and $\mathbf{F}^{i+1}$ are given by

$$\begin{cases} \mathbf{E}^{i+1} = \sum\limits_{k=1}^{T} \mathrm{E}^{(i+1)}[\kappa_1(\xi_k)(\mathbf{x}_k - \mathbf{F}_k\mathbf{x}_{k-1} - \boldsymbol{\beta}_1/s_1(\xi_k))(\mathbf{x}_k - \mathbf{F}_k\mathbf{x}_{k-1} - \boldsymbol{\beta}_1/s_1(\xi_k))^{\mathrm{T}}] \\ \text{s.t. } \mathbf{E}^{i+1} > \mathbf{0} \\ \mathbf{F}^{i+1} = \sum\limits_{k=1}^{T} \mathrm{E}^{(i+1)}[\kappa_2(\lambda_k)(\mathbf{z}_k - \mathbf{H}_k\mathbf{x}_k - \boldsymbol{\beta}_2/s_2(\lambda_k))(\mathbf{z}_k - \mathbf{H}_k\mathbf{x}_k - \boldsymbol{\beta}_2/s_2(\lambda_k))^{\mathrm{T}}] \\ \text{s.t. } \mathbf{F}^{i+1} > \mathbf{0} \end{cases} \tag{117}$$

According to (116)-(117), we can obtain (48)-(51).

D. Proof of (39) for different prior distributions

As an example, firstly, we prove that (39) holds for the Gamma distribution. If the prior distributions are selected as a Gamma distribution, i.e., $\pi_1(\xi_k;\omega) = \mathrm{G}(\xi_k; \frac{\omega}{2}, \frac{\omega}{2})$ and $\pi_2(\lambda_k;\nu) = \mathrm{G}(\lambda_k; \frac{\nu}{2}, \frac{\nu}{2})$, we have

$$\begin{cases} \mathrm{E}^{(i)}[\log \pi_1(\xi_k;\omega)] = \left(\frac{\mathrm{E}^{(i)}[\omega]}{2} - 1\right)\log \xi_k - \frac{\mathrm{E}^{(i)}[\omega]}{2}\xi_k + c_{\xi_k} \\ \mathrm{E}^{(i)}[\log \pi_2(\lambda_k;\nu)] = \left(\frac{\mathrm{E}^{(i)}[\nu]}{2} - 1\right)\log \lambda_k - \frac{\mathrm{E}^{(i)}[\nu]}{2}\lambda_k + c_{\lambda_k} \end{cases} \tag{118}$$

Using (118) yields

$$\begin{cases} \mathrm{E}^{(i)}[\log \mathrm{G}(\xi_k; \frac{\omega}{2}, \frac{\omega}{2})] = \log \mathrm{G}\left(\xi_k; \frac{\mathrm{E}^{(i)}[\omega]}{2}, \frac{\mathrm{E}^{(i)}[\omega]}{2}\right) \\ \mathrm{E}^{(i)}[\log \mathrm{G}(\lambda_k; \frac{\nu}{2}, \frac{\nu}{2})] = \log \mathrm{G}\left(\lambda_k; \frac{\mathrm{E}^{(i)}[\nu]}{2}, \frac{\mathrm{E}^{(i)}[\nu]}{2}\right) \end{cases} \tag{119}$$

It is seen from (119) that (39) holds when the prior distributions are chosen as a Gamma distribution. Similarly, we can prove that (39) also holds when the prior distributions are, respectively, selected as the exponential, Beta, inverse Gamma, and inverse exponential distributions.

E. Proof of Proposition 1

Proposition 1: If $\mathbf{C}$ is a positive semi-definite matrix and $\mathbf{D}$ is a positive definite matrix, then we have

$$\mathrm{tr}(\mathbf{CD}) \geq 0 \tag{120}$$

where $\mathrm{tr}(\mathbf{CD}) = 0$ if and only if $\mathbf{C}$ is a zero matrix, i.e., $\mathbf{C} = 0$.

Proof: Since $\mathbf{D}$ is a positive definite matrix, $\mathbf{D}$ can be factored by the Cholesky Decomposition, i.e.,

$$\mathbf{D} = \mathbf{L}\mathbf{L}^T \tag{121}$$

where $\mathbf{L}$ is an invertible lower triangular matrix.

Exploiting (121), we obtain

$$\mathrm{tr}(\mathbf{CD}) = \mathrm{tr}(\mathbf{CLL}^T) = \mathrm{tr}(\mathbf{L}^T\mathbf{CL}) \tag{122}$$

Considering that $\mathbf{L}$ is an invertible matrix and $\mathbf{C}$ is a positive semi-definite matrix, then $\mathbf{L}^T\mathbf{CL}$ is also a positive semi-definite matrix, i.e.,

$$\mathbf{L}^T\mathbf{CL} \geq \mathbf{0} \tag{123}$$

where $\mathbf{L}^T\mathbf{CL} = \mathbf{0}$ if and only if $\mathbf{C} = 0$.

Utilizing (123) in (122), we can obtain (120).

F. Proof of (84)

Using (49)-(51) yields

$$\mathbf{T}^{i+1} > \mathbf{0}, \qquad \mathbf{U}^{i+1} > \mathbf{0} \tag{124}$$

Employing (124) in (60), we have

$$\mathrm{E}^{(i+1)}[\mathbf{Q}^{-1}] > \mathbf{0}, \qquad \mathrm{E}^{(i+1)}[\mathbf{R}^{-1}] > \mathbf{0} \tag{125}$$

Since $\mathbf{A}_k^{(i+1)}$ and $\mathbf{B}_k^{(i+1)}$ are, respectively, the covariance matrices of the residuals of state and measurement vectors, $\mathbf{A}_k^{(i+1)}$ and $\mathbf{B}_k^{(i+1)}$ are nonzero positive semi-definite matrices, i.e.,

$$\begin{cases} \mathbf{A}_k^{(i+1)} \geq \mathbf{0}, & \mathbf{A}_k^{(i+1)} \neq \mathbf{0} \\ \mathbf{B}_k^{(i+1)} \geq \mathbf{0}, & \mathbf{B}_k^{(i+1)} \neq \mathbf{0} \end{cases} \tag{126}$$

Exploiting (126), $\mathrm{E}^{(i)}[\boldsymbol{\beta}_1\boldsymbol{\beta}_1^{\mathrm{T}}] \geq \mathbf{0}$ and $\mathrm{E}^{(i)}[\boldsymbol{\beta}_2\boldsymbol{\beta}_2^{\mathrm{T}}] \geq \mathbf{0}$ in (37) and using Proposition 1, we can obtain (84).

G. Robust RTS smoother based on Gaussian distribution

If state and measurement noises are modelled by a Gaussian distribution, then the PDFs of state and measurement noises can be formulated as (18) with $\boldsymbol{\beta}_1 = \mathbf{0}$, $\boldsymbol{\beta}_2 = \mathbf{0}$, $\kappa_1(\xi_k) = \xi_k$, $\kappa_2(\lambda_k) = \lambda_k$, $\pi(\xi_k;\omega) = \delta(\xi_k - 1)$ and $\pi(\lambda_k;\nu) = \delta(\lambda_k - 1)$.

Since $\boldsymbol{\beta}_1 = \mathbf{0}$, $\boldsymbol{\beta}_2 = \mathbf{0}$, $\kappa_1(\xi_k) = \xi_k$ and $\kappa_2(\lambda_k) = \lambda_k$, we can obtain (63)-(66), (70) and (79). Employing (79) in (43), the likelihood PDFs $l_1(\xi_k)$ and $l_2(\lambda_k)$ are rewritten as

$$\begin{cases} l_1(\xi_k) = c_2 \exp(0.5n\log\xi_k - 0.5\xi_k\Delta_{1,k}^{(i+1)}) \\ l_2(\lambda_k) = \bar{c}_2 \exp(0.5m\log\lambda_k - 0.5\lambda_k\bar{\Delta}_{1,k}^{(i+1)}) \end{cases} \tag{127}$$

Substituting (127) in (44) and using $\pi(\xi_k;\omega)=\delta(\xi_k-1)$, $\pi(\lambda_k;\nu)=\delta(\lambda_k-1)$ yields

$$\begin{cases} q^{(i+1)}(\xi_k)=c_1c_2\exp(0.5n\log\xi_k-0.5\xi_k\Delta_{1,k}^{(i+1)})\delta(\xi_k-1) \\ q^{(i+1)}(\lambda_k)=\bar{c}_1\bar{c}_2\exp(0.5m\log\lambda_k-0.5\lambda_k\bar{\Delta}_{1,k}^{(i+1)})\delta(\lambda_k-1) \end{cases} \tag{128}$$

Since the integrals of $q^{(i+1)}(\xi_k)$ and $q^{(i+1)}(\lambda_k)$ over the entire space are equal to one, we have

$$\begin{cases} \int c_1c_2\exp(0.5n\log\xi_k-0.5\xi_k\Delta_{1,k}^{(i+1)})\delta(\xi_k-1)d\xi_k=1 \\ \int \bar{c}_1\bar{c}_2\exp(0.5m\log\lambda_k-0.5\lambda_k\bar{\Delta}_{1,k}^{(i+1)})\delta(\lambda_k-1)d\lambda_k=1 \end{cases} \tag{129}$$

Exploiting (129) gives

$$c_1c_2=\exp(0.5\Delta_{1,k}^{(i+1)}),\qquad \bar{c}_1\bar{c}_2=\exp(0.5\bar{\Delta}_{1,k}^{(i+1)}) \tag{130}$$

Utilizing (128) and (130), the expectations of mixing parameters are calculated as

$$\begin{cases} \mathrm{E}^{(i+1)}[\xi_k]=c_1c_2\exp(-0.5\Delta_{1,k}^{(i+1)})=1 \\ \mathrm{E}^{(i+1)}[\lambda_k]=\bar{c}_1\bar{c}_2\exp(-0.5\bar{\Delta}_{1,k}^{(i+1)})=1 \end{cases} \tag{131}$$

Substituting (131) in (65)-(66) results in

$$\begin{cases} \tilde{\mathbf{q}}_k^{(i)}=\mathbf{0}, & \tilde{\mathbf{Q}}_k^{(i)}=\left\{\mathrm{E}^{(i)}[\mathbf{Q}^{-1}]\right\}^{-1} \\ \tilde{\mathbf{r}}_k^{(i)}=\mathbf{0}, & \tilde{\mathbf{R}}_k^{(i)}=\left\{\mathrm{E}^{(i)}[\mathbf{R}^{-1}]\right\}^{-1} \end{cases} \tag{132}$$

$$\begin{cases} \mathbf{E}^{(i+1)}=\sum_{k=1}^{T}\mathbf{A}_k^{(i+1)} \\ \mathbf{F}^{(i+1)}=\sum_{k=1}^{T}\mathbf{B}_k^{(i+1)} \end{cases} \tag{133}$$

where $\mathrm{E}^{(i+1)}[\mathbf{Q}^{-1}]$, $\mathrm{E}^{(i+1)}[\mathbf{R}^{-1}]$, $\mathbf{A}_k^{(i+1)}$ and $\mathbf{B}_k^{(i+1)}$ are, respectively, given by (60) and (62).

REFERENCES

[1] D. Simon, *Optimal State Estimation: Kalman, H infinity, and Nonlinear Approaches*. John Wiley & Sons, 2006.

[2] Y. L. Huang, Y. G. Zhang, Z. M. Wu, N. Li, and J. Chambers, "A novel robust Student's t based Kalman filter," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 53, no. 1, Feb. 2017.

[3] S. Arulampalam, S. Maskell, N. Gordon, and T. Clapp, "A tutorial on particle filters for on-line non-linear/non-Gaussian Bayesian tracking," *IEEE Transactions on Signal Processing*, vol. 50, no. 2, pp. 174–189, Feb. 2002.

[4] S. J. Godsill, A. Doucet, and M. West, "Monte Carlo smoothing for nonlinear time series," *Journal of the American Statistical Association*, vol. 99, no. 465, pp. 156–168, 2004.

[5] J. Loxam and T. Drummond, "Student mixture filter for robust, realtime visual tracking," In Proceedings of *10th European Conference on Computer Vision: Part III*, 2008.

[6] 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 Processing Letters*, vol. 23, no. 4, pp. 468–472, Apr. 2016.

[7] A. Y. Aravkin, J. V. Burke, and G. Pillonetto, "Robust and trend-following Kalman smoothers using Student's t," In *IFAC Proceedings*, vol. 45, no. 16, pp. 1215–1220, July 2012.

[8] A. Y. Aravkin, J. V. Burke, and G. Pillonetto, "Robust and trend-following Student's t Kalman smoothers," *SIAM Journal on Control and Optimization*, vol. 52, no. 5, pp. 2891–2916, Sep. 2014.

[9] 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.

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

[11] M. Roth, T. Ardeshiri, E. Özkan, and F. Gustafsson, "Robust Bayesian filtering and smoothing using Student's t distribution," arXiv preprint arXiv:1703.02428.

[12] G. Agamennoni and E.M. Nebot, "Robust estimation in non-linear state-space models with state-dependent noise," *IEEE Transactions on Signal Processing*, vol. 62, no. 8, Apr. 2014.

[13] H. Nurminen, T. Ardeshiri, R. Piché, and F. Gustafsson, "Robust inference for state-space models with skewed measurement noise," *IEEE Signal Processing Letters*, vol. 22, no. 11, pp. 2450–2454, Dec. 2015.

[14] 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.

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

[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] G. Agamennoni, J.I. Nieto, and E.M. Nebot, "Approximate inference in state-space models with heavy-tailed noise," *IEEE Transactions on Signal Processing*, vol. 60, no. 10, pp. 5024–5037, Oct. 2012.

[18] 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 Proceedings of *2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Mar 2016.

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

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

[21] F. Tronarp, R. Hostettler, and S. Särkkä, "Sigma-point filtering for nonlinear systems with non-additive heavy-tailed noise," In Proceedings of *2016 19th International Conference on Information Fusion (FUSION)*, July 2016, pp. 1859-1866.

[22] 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, pp. 7964–7974, May 2017.

[23] O. Straka and J. Duník, "Stochastic integration Student's-t filter," In Proceedings of *2017 20th International Conference on Information Fusion (FUSION)*, July 2017, pp. 1–8.

[24] Y. L. Huang and Y. G. Zhang, "Design of high-degree Student's t-based cubature filters," *Circuits, Systems, and Signal Processing*, vol. 37, no. 5, pp. 2206–2225, May 2018.

[25] I. Bilik and J. Tabrikian, "Manoeuvering target tracking in the presence of glint using the nonlinear Gaussian mixture Kalman filter," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 46, no. 1, pp. 246–262, Jan. 2010.

[26] M. Roth, "Kalman filters for nonlinear systems and heavy-tailed noise," Licentiate thesis, Linköping University, Linköping, Sweden, Sep. 2013.

[27] Y. L. Huang, Y. G. Zhang, B. Xu, Z. M. Wu, and J. Chambers, "A new adaptive extended Kalman filter for cooperative localization," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 54, no. 1, pp. 353–368, Feb. 2018.

[28] 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 Transactions on Mechatronics*, vol. 22, no. 5, pp. 2380–2386, Oct. 2017.

[29] Q. Li, Y. Y. Ben, S. M. Naqvi, J. A. Neasham, and J. Chambers, "Robust Student's t-based cooperative navigation for autonomous underwater vehicles," to appear in *IEEE Transactions on Instrumentation and Measurement*, 2018.

[30] 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 Transactions on Automatic Control*, vol. 63, no. 2, pp. 594–601, Feb. 2018.

[31] T. Ardeshiri, E. Ozkan, U. Orguner, and F. Gustafsson, "Approximate Bayesian smoothing with unknown process and measurement noise covariances," *IEEE Signal Processing Letters*, vol. 22, no. 12, pp. 2450–2454, Dec. 2015.

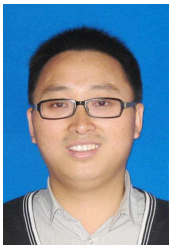
[32] Y. L. Huang, Y. G. Zhang, P. Shi, Z. M. Wu, J. H. Qian, and J. Chambers, "Robust Kalman filters based on Gaussian scale mixture distributions with application to target tracking," *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, 2017.

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

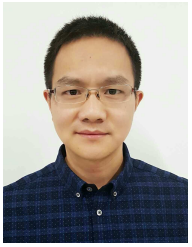
[34] S. Saha, "Noise robust online inference for linear dynamic systems," [Online]. Available: http://arxiv.org/pdf/1504.05723

[35] B. Jorgensen, *Statistical properties of the generalized inverse Gaussian distribution*. Springer Science & Business Media, 2012.

[36] S. Särkkä and J. Hartikainen, "On Gaussian optimal smoothing of non-linear state space models," *IEEE Transactions on Automatic Control*, vol. 55, no. 8, pp. 1938–1941, Aug. 2010.



Yulong Huang received the B.S. degree in Automation from the Department of Automation, Harbin Engineering University, Harbin, China, in 2012. He received the Ph.D degree in control science and engineering from the Department of Automation, Harbin Engineering University, Harbin, China, in 2018. From Nov. 2016 to Nov. 2017, he was a visiting graduate researcher at the Electrical Engineering Department of Columbia University, New York, USA. Currently, he is an Associate 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 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.

Yuxin Zhao received the Ph.D. degree in navigation, guidance and control from Harbin Engineering University (HEU) in 2005, and completed post-doctorial research in Control Science and engineering from Harbin Institute of Technology (HIT) in 2008. In 2004, he was awarded scholarship of the State Administration of Foreign Experts Affairs for visiting Levin Institute of State University of New York (SUNY), and from 2012 to 2013, he visited the Centre for Transport Studies (CTS) at Imperial College London as a research scholar.

He is currently the head and professor at College of Automation, Harbin Engineering University (HEU) in China. His research interests include Artificial Intelligencefiltering theory, marine navigation system and intelligent transportation system. He has published more than 100 papers, including more than 40 international journal papers in these areas.

Dr. Zhao has served the academic communities in various capacities including the fellow of IET, the senior member of IEEE, the member of Royal Institute of Navigation, the fellow of China Navigation Institute.

Lyudmila Mihaylova (M'98-SM'2008) is a Professor of Signal Processing and Control at the Department of Automatic Control and Systems Engineering at the University of Sheffield, United Kingdom. Her research is in the areas of machine learning and autonomous systems with various applications such as navigation, surveillance and sensor network systems. She has given a number of talks and tutorials, including Fusion 2017 (Xi'an, China), the plenary talk for the IEEE Sensor Data Fusion 2015 (Germany), invited talks at IPAMI Traffic Workshop 2016 (USA) and others. Prof. Mihaylova is an Associate Editor of the IEEE Transactions on Aerospace and Electronic Systems and of the Elsevier Signal Processing Journal. She was elected in March 2016 as a president of the International Society of Information Fusion (ISIF). She is on the board of Directors of ISIF and a Senior IEEE member. She was the general vice-chair for the International Conference on Information Fusion 2018 (Cambridge, UK), of the IET Data Fusion & Target Tracking 2014 and 2012 Conferences, Program co-chair for the 19th International Conference on Information Fusion 2016 and others.

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 at Loughborough University and is now a Visiting Professor. In 2015, he joined the School of Electrical and Electronic Engineering, Newcastle University, where he was Professor of signal and information processing and led the ComS2IP group and is now a Visiting Professor. In 2017 he became the Head of the Department of Engineering at the University of Leicester. He is also an International Honorary Dean and Guest Professor at Harbin Engineering University, China with support from the 1000 Talents Scheme. 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 almost 90 researchers through to Ph.D. graduation and published nearly 600 conference papers 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 as a Senior Area Editor between 2011-2014.