Proceedings Paper:

Huang, Y., Zhang, Y., Zhao, Y. et al. (2 more authors) (2018) A Novel Robust Rauch-Tung-Striebel Smoother Based on Slash and Generalized Hyperbolic Skew Student's T-Distributions. In: Proceedings of the International Conference on Information Fusion. International Conference on Information Fusion, 10-13 Jul 2018, Cambridge, United Kingdom. IEEE . ISBN 978-0-9964527-6-2

10.23919/ICIF.2018.8455256

A Novel Robust Rauch-Tung-Striebel Smoother Based on Slash and Generalized Hyperbolic Skew Student's T-Distributions

Yulong Huang[a], Yonggang Zhang[a], Yuxin Zhao[a], Lyudmila Mihaylova[b], Jonathon Chambers[a,c]

[a]Department of Automation, Harbin Engineering University, Harbin, China
[b]Department of Automatic Control and Systems Engineering, University of Sheffield, Sheffield, UK
[c]Department of Engineering, University of Leicester, Leicester, UK
heuedu@163.com, {zhangyg, zhaoyuxin}@hrbeu.edu.cn
L.S.Mihaylova@sheffield.ac.uk, Jonathon.Chambers@le.ac.uk

***Abstract*—In this paper, a novel robust Rauch-Tung-Striebel smoother is proposed based on the Slash and generalized hyperbolic skew Student's t-distributions. A novel hierarchical Gaussian state-space model is constructed by formulating the Slash distribution as a Gaussian scale mixture form and formulating the generalized hyperbolic skew Student's t-distribution as a Gaussian variance-mean mixture form, based on which the state trajectory, mixing parameters and unknown noise parameters are jointly inferred using the variational Bayesian approach. The posterior probability density functions of mixing parameters of the Slash and generalized hyperbolic skew Student's t-distributions are, respectively, approximated as truncated Gamma and generalized inverse Gaussian. Simulation results illustrate that the proposed robust Rauch-Tung-Striebel smoother has better estimation accuracy than existing state-of-the-art smoothers.**

***Index Terms*—State estimation, Rauch-Tung-Striebel smoother, heavy-tailed and/or skew noise, Slash distribution, generalized hyperbolic skew Student's t-distribution, variational Bayesian**

I. Introduction

As a smoothing extension of the Kalman filter, the Rauch-Tung-Striebel (RTS) smoother has been widely used in a range of applications, including target tracking, navigation, positioning, and signal processing [1], [2]. It employs the Kalman filter as its building block, and it is an optimal estimator in terms of minimum mean square error for a linear state-space model with Gaussian state and measurement noises. However, in some engineering applications, the state and measurement noises may have heavy-tailed and/or skew distributions, such as in manoeuvring target tracking [3]–[5], integrated navigation [6], and cooperative localization of autonomous underwater vehicles [7], [8], which are often induced by the impulsive interferences, outliers and modelling artifacts [9]. The performance of the conventional RTS smoother degrades considerably for such engineering applications with heavy-tailed and/or skew non-Gaussian noises [10], [11]. Generally, it is difficult to derive an analytical non-Gaussian smoother since there is no general mathematical formulation for non-Gaussian noises nor an analytical and closed form solution for non-Gaussian posterior probability density function (PDF).

Recently, Student's t and Skew t-distributions based smoothing algorithms have been proposed [11]–[14] to solve a class of non-Gaussian smoothing problems, in which the state noise may have heavy-tailed distribution and the measurement noise may have heavy-tailed and/or skew distribution. A robust and trend-following Student's t-RTS (RTF-ST-RTS) smoother has been proposed by modelling the state and measurement noises as Student's t-distributed and utilizing the convex composite extension of the Gauss-Newton method to find an approximate maximum a posteriori estimate of the state trajectory [15]. A Student's t-smoother has been proposed based on Student's t modelling of the state and measurement noises and Student's t approximations of the posterior PDFs [16], [17]. To further improve the performance of Student's t-distribution based smoothers, the variational Bayesian (VB) and Student's t-based RTS (VB-ST-RTS) smoother has been proposed, in which the VB approach is utilized to jointly infer the state trajectory, auxiliary random variables, and unknown noise parameters so that the models of the noise terms can be more accurate [11]. A Skew t-RTS smoother has also been proposed by modelling the measurement noise as the Skew t-distribution [13]. Unfortunately, the above smoothers all cannot solve the smoothing problem of a linear state-space model with a heavy-tailed state noise and a heavy-tailed and skew measurement noise, which may be encountered in manoeuvring target tracking, integrated navigation, and cooperative localization of autonomous underwater vehicles. A combined method of the VB-ST-RTS smoother for heavy-tailed state and measurement noises [11] and the Skew t-RTS smoother [13] for heavy-tailed and skew measurement noise may address such a smoothing problem. However, the performance of the combined method is very sensitive to the distribution parameters of the Skew

This work was supported by the National Natural Science Foundation of China under Grant No. 61773133 and the Natural Science Foundation of Heilongjiang Province Grant No. F2016008 and the PhD Student Research and Innovation Fund of the Fundamental Research Founds for the Central Universities under Grant No. HEUGIP201706 and 1000 Talents Funding. The authors gratefully acknowledge the support of K. C. Wong Education Foundation.

t-distribution, which are difficult to be determined in practical engineering applications. Moreover, the combined method still cannot solve the smoothing problem of a linear state-space model with heavy-tailed and skew state and measurement noises.

In this paper, we focus on the smoothing problem of a linear state-space model with heavy-tailed and/or skew noises. We propose to model heavy-tailed noises using a Gaussian scale mixture distribution and model heavy-tailed and skew noises using a Gaussian variance-mean mixture distribution. As an example, the state and measurement noises are, respectively, assumed to have a heavy-tailed distribution and a heavy-tailed and skew distribution, and the heavy-tailed state noise is modelled by a Slash distribution which is a special Gaussian scale mixture distribution, and the heavy-tailed and skew measurement noise is modelled by a generalized hyperbolic (GH) skew Student's t-distribution which is a special Gaussian variance-mean mixture distribution, so that a novel robust RTS smoother is thereby proposed. By formulating the Slash distribution as a Gaussian scale mixture form and formulating the GH skew Student's t-distribution as a Gaussian variance-mean mixture form, a novel hierarchical Gaussian state-space model is achieved. The state trajectory, mixing parameters and unknown noise parameters are jointly inferred based on the constructed hierarchical Gaussian state-space model using the VB approach. The posterior PDFs of mixing parameters of the Slash and GH skew Student's t-distributions are, respectively, approximated by truncated Gamma and generalized inverse Gaussian. Simulation results of a manoeuvring target tracking example show that the proposed robust RTS smoother has better estimation accuracy than existing state-of-the-art smoothers [11], [13], [15], [17], [18].

The remainder of this paper is organized as follows. Section II presents notations and brief descriptions about Slash and GH skew Student's t-distributions. Section III proposes a novel robust RTS smoother. In Section IV, the proposed robust RTS smoother is applied to a manoeuvring target tracking example and simulation results are given. Finally, conclusions are summarised in Section V.

II. Preliminaries

A. Notations

Throughout this paper, we denote $\mathbf{y}_{i:j} \triangleq \{\mathbf{y}_k | i \leq k \leq j\}$; $\mathrm{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ denotes the multivariate Gaussian distribution with mean vector $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$, and $\mathrm{g}(\mathbf{x}; \boldsymbol{\mu}, \boldsymbol{\Sigma})$ denotes the PDF of $\mathbf{x} \sim \mathrm{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$; $\mathrm{ST}(\cdot; \boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\Delta}, \nu)$ denotes the Skew-t PDF with location parameter $\boldsymbol{\mu}$, scale matrix $\boldsymbol{\Sigma}$, shape parameter $\boldsymbol{\Delta}$ and degrees of freedom (dof) parameter ν; $\mathrm{IW}(\cdot; \nu, \boldsymbol{\Sigma})$ denotes the inverse-Wishart PDF with dof parameter ν and inverse scale matrix $\boldsymbol{\Sigma}$; $\mathrm{N}_+(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ denotes the truncated Gaussian distribution with the closed positive orthant as support, location parameter $\boldsymbol{\mu}$ and squared-scale matrix $\boldsymbol{\Sigma}$; $\mathrm{G}(\cdot; a, b)$ denotes the Gamma PDF with shape parameter a and rate parameter b; $\mathrm{Be}(\cdot; a, b)$ denotes the Beta PDF with shape parameters a and b; $\mathrm{IG}(\cdot; a, b)$ denotes the inverse-Gamma PDF with shape parameter a and scale parameter b; $\mathrm{GIG}(\cdot; a, b, p)$ denotes the Generalized Inverse Gaussian (GIG) PDF with shape parameters a, b and p; $K_\rho(\cdot)$ denotes a modified Bessel function of the second kind with the order ρ; $\log$ denotes the natural logarithm; $\exp$ denotes the natural exponential; $\mathbf{I}_n$ denotes the $n \times n$ identity matrix; $|\mathbf{A}|$ denotes the determinant of a square matrix $\mathbf{A}$; the superscript "-1" denotes the inverse operation of a matrix; the superscript "T" denotes the transpose operation of a vector or matrix; $\mathrm{E}_x[\cdot]$ is the expectation operator with respect to the distribution of x; $\bigcup$ denotes the union operation; and $\mathrm{tr}(\cdot)$ denotes the trace operation of a matrix.

TABLE I: Tail behaviours of Gaussian distribution, Student's t-distribution, Slash distribution, Skew normal distribution, Skew t-distribution and GH skew Student's t-distribution for a scalar case.

Distributions	Tail behaviours	Conditions
Gaussian	$c\exp(-0.5x^2)$	$x \to \pm\infty$
Student's t	$c\|x\|^{-\nu-1}$	$x \to \pm\infty$
Slash	$c\|x\|^{-2\nu-1}$	$x \to \pm\infty$
Skew normal	$c\exp(-0.5x^2)$	$c \to 1$ as $\beta x \to +\infty$ $c \to 0$ as $\beta x \to -\infty$
Skew t-distribution	$c\|x\|^{-\nu-1}$	$x \to \pm\infty$
GH skew Student's t	$c\|x\|^{-\frac{\nu}{2}-1}$ $c\|x\|^{-\frac{\nu}{2}-1}\exp(-2\|\beta\|\|x\|)$	$\beta x \to +\infty$ $\beta x \to -\infty$

B. Slash and GH skew Student's t-distributions

In engineering practice, many types of noises, which are induced by impulsive interferences, outliers and modelling artifacts, are naturally non-Gaussian. The non-Gaussian noises often have heavy-tailed and/or skew distributions. The heavy-tailed noise can be modelled by a Slash distribution, and the heavy-tailed and skew noise can be modelled by a GH skew Student's t-distribution. The Slash distribution is a heavy-tailed non-Gaussian distribution, and the GH skew Student's t-distribution is a heavy-tailed and skew non-Gaussian distribution. A random vector $\mathbf{X}$ obeys a Slash distribution and a random vector $\mathbf{Z}$ follows a GH skew Student's t-distribution if their PDFs can be, respectively, formulated as [9], [19], [20]

$$\begin{cases} p(\mathbf{x}) = \int_0^1 \mathrm{g}(\mathbf{x}; \boldsymbol{\mu}, \boldsymbol{\Sigma}/y)\mathrm{Be}(y; \nu, 1)dy \\ p(\mathbf{z}) = \int_0^{+\infty} \mathrm{g}(\mathbf{z}; \boldsymbol{\mu} + y\boldsymbol{\beta}, y\boldsymbol{\Sigma})\mathrm{IG}(y; \frac{\nu}{2}, \frac{\nu}{2})dy \end{cases} \quad (1)$$

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 $\boldsymbol{\beta}$ is a shape parameter. The shape parameter $\boldsymbol{\beta}$ dominates the symmetry and skewness of a GH skew Student's t-distribution. The GH skew Student's t-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 an arbitrary element of $\boldsymbol{\beta}$.

The Slash distribution has heavier tails than the Gaussian distribution, and the GH skew Student's t-distribution has both heavier tails and higher skewness than the Gaussian

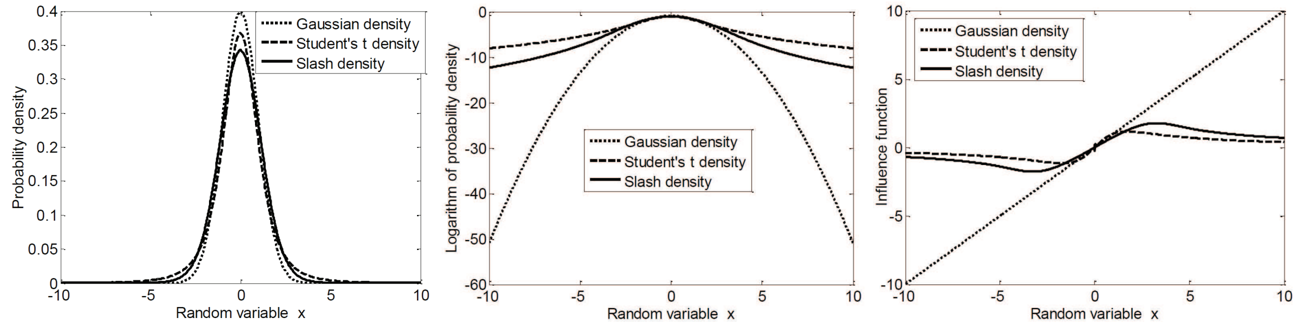

Fig. 1: Gaussian, Student's t, Slash densities, corresponding log plots, and influence functions for a scalar case.

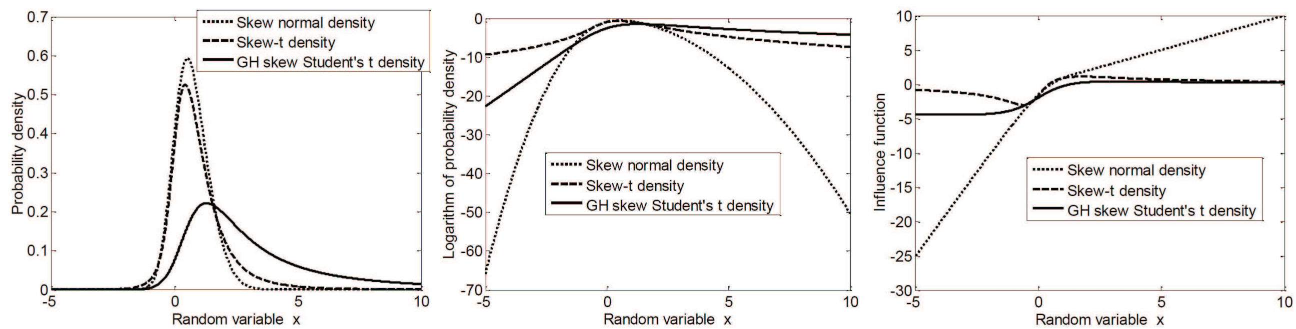

Fig. 2: Skew normal, Skew-t and GH skew Student's t densities, corresponding log plots, and influence functions for a scalar case.

distribution. The tail behaviours of the Gaussian distribution, Student's t-distribution, Slash distribution, Skew normal distribution [21], Skew t-distribution [22] and GH skew Student's t-distribution for a scalar case are listed in Table I. Also, comparisons of the probability densities and influence functions of these distributions are shown in Fig.1–Fig.2, where the parameters are selected as $\mu = 0$, $\Sigma = 1$, $\nu = 3$, and $\beta = 2$. It is seen from Table I that the tail behaviours of the Student's t-distribution, Slash distribution, and Skew t-distribution are only determined by the dof parameter ν, but the tail behaviour of the GH skew Student's t-distribution is determined by both the dof parameter ν and shape parameter β. It can be seen from Table I and Fig. 1 that the Student's t-distribution has heavier tails than the Slash distribution, and the influence function of the Student's t-distribution redescends faster than that of the Slash distribution. Thus, the Slash distribution may be more suitable for fitting and modelling moderately heavy-tailed data as compared with the Student's t-distribution. We can see from Table I and Fig. 2 that the Skew t-distribution has two moderately heavy tails, and the GH skew Student's t-distribution has one heavy tail and one slightly heavy tail, but they both have heavier tails than the Skew normal distribution. We can also observe from Fig. 2 that the influence function of the GH skew Student's t-distribution redescends slightly faster than that of the Skew t-distribution when $x > 0$ and significantly slower than that of the Skew t-distribution when $x < 0$. Therefore, as compared with the Skew t-distribution, the GH skew Student's t-distribution may be more suitable for modelling substantially skew and heavy-tailed data. Moreover, for the problem of designing a robust RTS smoother, the GH skew Student's t-distribution is easier to handle as compared with the Skew t-distribution since it can be written as the Gaussian variance-mean mixture form in (1). Next, a novel robust RTS smoother will be proposed based on the Slash and GH skew Student's t-distributions using the VB approach, where the heavy-tailed state noise is modelled by a Slash distribution and the heavy-tailed and skew measurement noise is modelled by a GH skew Student's t-distribution.

III. A Novel Robust RTS Smoother

A. Novel Hierarchical Gaussian State-Space Model

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

$$\begin{cases} \mathbf{x}_k = \mathbf{F}_k \mathbf{x}_{k-1} + \mathbf{w}_{k-1} & \text{(state equation)} \\ \mathbf{z}_k = \mathbf{H}_k \mathbf{x}_k + \mathbf{v}_k & \text{(measurement equation)} \end{cases} \tag{2}$$

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 obtain a smoothing estimate of state trajectory $\mathbf{x}_{0:T}$ based on a linear state-space model and all available measurements $\mathbf{z}_{1:T}$ from time sample 1 to time sample T. Note that the filtering estimate of state vector $\mathbf{x}_k$ is only based on available measurements $\mathbf{z}_{1:k}$ from time sample 1 to time sample k, but the smoothing estimate of state vector $\mathbf{x}_k$ is based on all available measurements $\mathbf{z}_{1:T}$ from time sample 1 to time sample T.

In this paper, the state and measurement noises are, respectively, assumed to have heavy-tailed distribution and heavy-tailed and skew distribution, which are, respectively, modelled by the stationary Slash distributed and the stationary GH skew Student's t-distributed as

$$\begin{cases} p(\mathbf{w}_{k-1}) = \int_0^1 \mathrm{g}(\mathbf{w}_{k-1}; \mathbf{0}, \mathbf{Q}/\xi_k)\mathrm{Be}(\xi_k; \omega, 1)d\xi_k \\ p(\mathbf{v}_k) = \int_0^{+\infty} \mathrm{g}(\mathbf{v}_k; \lambda_k\boldsymbol{\beta}, \lambda_k\mathbf{R})\mathrm{IG}(\lambda_k; \frac{\nu}{2}, \frac{\nu}{2})d\lambda_k \end{cases} \tag{3}$$

where $\mathbf{Q}$, $\mathbf{R}$, ω, ν, ξ_k and λ_k are, respectively, the scale matrices, dof parameters and mixing parameters of the state and measurement noises, and $\boldsymbol{\beta}$ is the shape parameter of the measurement noise.

In this paper, the scale matrices $\mathbf{Q}$ and $\mathbf{R}$ and shape parameter $\boldsymbol{\beta}$ are unknown, whose joint prior PDF is defined over a limited support and assumed to be a constant, i.e.,

$$p(\mathbf{Q}, \mathbf{R}, \boldsymbol{\beta}) = c \tag{4}$$

and they will be jointly estimated using the VB approach.

Exploiting (2)–(3), 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 formulated as

$$\begin{cases} p(\mathbf{x}_k|\mathbf{x}_{k-1}) = \int_0^1 \mathrm{g}(\mathbf{x}_k; \mathbf{F}_k\mathbf{x}_{k-1}, \mathbf{Q}/\xi_k)\mathrm{Be}(\xi_k; \omega, 1)d\xi_k \\ p(\mathbf{z}_k|\mathbf{x}_k) = \int_0^{+\infty} \mathrm{g}(\mathbf{z}_k; \mathbf{H}_k\mathbf{x}_k + \lambda_k\boldsymbol{\beta}, \lambda_k\mathbf{R})\mathrm{IG}(\lambda_k; \frac{\nu}{2}, \frac{\nu}{2})d\lambda_k \end{cases} \tag{5}$$

According to (5), the state transition PDF and the likelihood PDF can be, respectively, written in the following hierarchical Gaussian forms

$$\begin{cases} p(\mathbf{x}_k|\mathbf{x}_{k-1}, \xi_k) = \mathrm{g}(\mathbf{x}_k; \mathbf{F}_k\mathbf{x}_{k-1}, \mathbf{Q}/\xi_k) \\ p(\xi_k) = \mathrm{Be}(\xi_k; \omega, 1), \qquad \text{s.t. } 0 < \xi_k < 1 \\ p(\mathbf{z}_k|\mathbf{x}_k, \lambda_k) = \mathrm{g}(\mathbf{z}_k; \mathbf{H}_k\mathbf{x}_k + \lambda_k\boldsymbol{\beta}, \lambda_k\mathbf{R}) \\ p(\lambda_k) = \mathrm{IG}(\lambda_k; \frac{\nu}{2}, \frac{\nu}{2}), \qquad \text{s.t. } \lambda_k > 0 \end{cases} \tag{6}$$

Equations (4) and (6) constitute a novel hierarchical Gaussian state-space model based on Slash and GH skew Student's t-distributions. The smoothing estimation problem for a linear state-space model with heavy-tailed state noise and heavy-tailed and skew measurement noise is transformed into the smoothing estimation problem for a hierarchical Gaussian state-space model formulated in (4) and (6). Next, we propose to jointly estimate the state trajectory, mixing parameters, scale matrices and shape parameter, i.e., $\boldsymbol{\Theta} = \{\mathbf{x}_{0:T}, \xi_{1:T}, \lambda_{1:T}, \mathbf{Q}, \mathbf{R}, \boldsymbol{\beta}\}$, based on the constructed hierarchical Gaussian state-space model using the VB approach.

B. Joint Estimates of State Trajectory, Mixing Parameters and Unknown Noise Parameters

To jointly infer state trajectory, mixing parameters and unknown noise parameters, the joint posterior PDF $p(\boldsymbol{\Theta}|\mathbf{z}_{1:T})$ needs to be calculated. Unfortunately, optimal solution of the joint posterior PDF is unavailable for hierarchical Gaussian state-space model (4) and (6) since beta, inverse-Gamma, inverse-Wishart PDFs don't have corresponding closed forms. 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 [23]

$$p(\boldsymbol{\Theta}|\mathbf{z}_{1:T}) \approx q(\mathbf{x}_{0:T})q(\xi_{1:T})q(\lambda_{1:T})q(\mathbf{Q})q(\mathbf{R})q(\boldsymbol{\beta}) \tag{7}$$

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 $\boldsymbol{\Theta}$. The approximate posterior PDF $q(\theta)$ satisfies the equation as follows [23], [24]

$$\log q(\theta) = \mathrm{E}_{\boldsymbol{\Theta}^{(-\theta)}}[\log p(\boldsymbol{\Theta}, \mathbf{z}_{1:k})] + c_\theta \tag{8}$$

where $\boldsymbol{\Theta}^{(-\theta)}$ is a subset of $\boldsymbol{\Theta}$ and it has all elements in $\boldsymbol{\Theta}$ except for θ, i.e., $\{\theta\} \bigcup \boldsymbol{\Theta}^{(-\theta)} = \boldsymbol{\Theta}$, 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 (8). To address this problem, the fixed-point iteration is employed to achieve an approximation of $q(\theta)$ by iteratively solving (8), 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 $q^{(i)}(\boldsymbol{\Theta}^{(-\theta)})$ to calculate the expectation in (8).

1) Variational Approximations of Posterior PDFs: Using (4) and (6), the joint PDF $p(\boldsymbol{\Theta}, \mathbf{z}_{1:T})$ is formulated as

$$p(\boldsymbol{\Theta}, \mathbf{z}_{1:T}) = c\mathrm{g}(\mathbf{x}_0; \hat{\mathbf{x}}_{0|0}, \mathbf{P}_{0|0}) \prod_{k=1}^{T} [\mathrm{g}(\mathbf{z}_k; \mathbf{H}_k\mathbf{x}_k + \lambda_k\boldsymbol{\beta}, \lambda_k\mathbf{R})\mathrm{g}(\mathbf{x}_k; \mathbf{F}_k\mathbf{x}_{k-1}, \mathbf{Q}/\xi_k)\mathrm{IG}(\lambda_k; \frac{\nu}{2}, \frac{\nu}{2})\mathrm{Be}(\xi_k; \omega, 1) \tag{9}$$

Let $\theta = \mathbf{x}_{0:T}$ and employing (9) in (8), $q^{(i+1)}(\mathbf{x}_{0:T})$ can be updated as Gaussian, i.e.,

$$q^{(i+1)}(\mathbf{x}_{0:T}) = \mathrm{g}(\mathbf{x}_{0:T}; \hat{\mathbf{x}}^{(i+1)}_{0:T|T}, \mathbf{P}^{(i+1)}_{0:T|T}) \tag{10}$$

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 with modified state and measurement noise covariance matrices $\tilde{\mathbf{Q}}^{(i+1)}_k$ and $\tilde{\mathbf{R}}^{(i+1)}_k$ and modified mean vector of measurement noise $\tilde{\mathbf{r}}^{(i+1)}_k$, which are given by

$$\begin{cases} \tilde{\mathbf{Q}}^{(i+1)}_k = \frac{\left\{\mathrm{E}^{(i)}_{\mathbf{Q}}[\mathbf{Q}^{-1}]\right\}^{-1}}{\mathrm{E}^{(i)}_{\xi_k}[\xi_k]}, \quad \tilde{\mathbf{R}}^{(i+1)}_k = \frac{\left\{\mathrm{E}^{(i)}_{\mathbf{R}}[\mathbf{R}^{-1}]\right\}^{-1}}{\mathrm{E}^{(i)}_{\lambda_k}[1/\lambda_k]} \\ \tilde{\mathbf{r}}^{(i+1)}_k = \mathrm{E}^{(i)}_{\boldsymbol{\beta}}[\beta]/\mathrm{E}^{(i)}_{\lambda_k}[1/\lambda_k] \end{cases} \tag{11}$$

where $\mathrm{E}^{(i)}_x[\cdot]$ denotes the expectation with respect to the approximate posterior PDF $q^{(i)}(x)$ at the i-th iteration.

Let $\theta = \xi_{1:T}$ and using (9) in (8), $q^{(i+1)}(\xi_k)$ can be updated as truncated Gamma PDF, and let $\theta = \lambda_{1:T}$ and employing (9) in (8), $q^{(i+1)}(\lambda_k)$ can be updated as GIG PDF, i.e.,

$$\begin{cases} q^{(i+1)}(\xi_k) = c_k\mathrm{G}(\xi_k; \alpha^{(i+1)}_k, \beta^{(i+1)}_k) \\ q^{(i+1)}(\lambda_k) = \mathrm{GIG}(\lambda_k; \eta^{(i+1)}_k, \varphi^{(i+1)}_k, \varrho^{(i+1)}_k) \end{cases} \tag{12}$$

where c_k is a normalizing constant formulated as

$$c_k = 1/\int_0^1 \mathrm{G}(\xi_k; \alpha_k^{(i+1)}, \beta_k^{(i+1)}) d\xi_k \tag{13}$$

and the shape and rate parameters are given by

$$\begin{cases} \alpha_k^{(i+1)} = 0.5n + \omega, & \beta_k^{(i+1)} = 0.5\Delta_0^{(i+1)} \\ \eta_k^{(i+1)} = \Delta_2^{(i+1)}, & \varphi_k^{(i+1)} = \Delta_1^{(i+1)} + \nu \\ \varrho_k^{(i+1)} = 0.5(m - \nu) \end{cases} \tag{14}$$

and the auxiliary parameters are given by

$$\begin{cases} \Delta_0^{(i+1)} = \mathrm{tr}\left(\mathbf{A}_k^{(i+1)} \mathrm{E}_{\mathbf{Q}}^{(i)}[\mathbf{Q}^{-1}]\right) \\ \Delta_1^{(i+1)} = \mathrm{tr}\left(\mathbf{B}_k^{(i+1)} \mathrm{E}_{\mathbf{R}}^{(i)}[\mathbf{R}^{-1}]\right) \\ \Delta_2^{(i+1)} = \mathrm{tr}\left(\mathrm{E}_{\boldsymbol{\beta}}^{(i)}[\boldsymbol{\beta}\boldsymbol{\beta}^{\mathrm{T}}] \mathrm{E}_{\mathbf{R}}^{(i)}[\mathbf{R}^{-1}]\right) \\ \mathbf{A}_k^{(i+1)} = \mathrm{E}_{\mathbf{x}_{k-1:k}}^{(i+1)}[(\mathbf{x}_k - \mathbf{F}_k\mathbf{x}_{k-1})(\mathbf{x}_k - \mathbf{F}_k\mathbf{x}_{k-1})^{\mathrm{T}}] \\ \mathbf{B}_k^{(i+1)} = \mathrm{E}_{\mathbf{x}_k}^{(i+1)}[(\mathbf{z}_k - \mathbf{H}_k\mathbf{x}_k)(\mathbf{z}_k - \mathbf{H}_k\mathbf{x}_k)^{\mathrm{T}}] \\ \mathbf{b}_k^{(i+1)} = \mathrm{E}_{\mathbf{x}_k}^{(i+1)}[\mathbf{z}_k - \mathbf{H}_k\mathbf{x}_k] \end{cases} \tag{15}$$

Let $\theta = \boldsymbol{\beta}$ and using (9) in (8), $q^{(i+1)}(\boldsymbol{\beta})$ is updated as Gaussian, i.e.,

$$q^{(i+1)}(\boldsymbol{\beta}) = \mathrm{g}(\boldsymbol{\beta}; \hat{\boldsymbol{\beta}}^{(i+1)}, \mathbf{P}_{\boldsymbol{\beta}}^{(i+1)}) \tag{16}$$

where the mean vector $\hat{\boldsymbol{\beta}}^{(i+1)}$ and covariance matrix $\mathbf{P}_{\boldsymbol{\beta}}^{(i+1)}$ are given by

$$\begin{cases} \hat{\boldsymbol{\beta}}^{(i+1)} = \left(\mathbf{D}^{(i+1)}\right)^{-1} \mathbf{d}^{(i+1)} \\ \mathbf{P}_{\boldsymbol{\beta}}^{(i+1)} = \left(\mathbf{D}^{(i+1)}\right)^{-1} \\ \mathbf{d}^{(i+1)} = \mathrm{E}_{\mathbf{R}}^{(i)}[\mathbf{R}^{-1}] \sum_{k=1}^{T} \mathbf{b}_k^{(i+1)} \\ \mathbf{D}^{(i+1)} = \mathrm{E}_{\mathbf{R}}^{(i)}[\mathbf{R}^{-1}] \sum_{k=1}^{T} \mathrm{E}_{\lambda_k}^{(i+1)}[\lambda_k] \end{cases} \tag{17}$$

Let $\theta = \mathbf{Q}$ and exploiting (9) in (8), $q^{(i+1)}(\mathbf{Q})$ is updated as inverse-Wishart PDF, and let $\theta = \mathbf{R}$ and employing (9) in (8), $q^{(i+1)}(\mathbf{R})$ is updated as inverse-Wishart PDF, i.e.,

$$\begin{cases} q^{(i+1)}(\mathbf{Q}) = \mathrm{IW}(\mathbf{Q}; t^{(i+1)}, \mathbf{T}^{(i+1)}) \\ q^{(i+1)}(\mathbf{R}) = \mathrm{IW}(\mathbf{R}; u^{(i+1)}, \mathbf{U}^{(i+1)}) \end{cases} \tag{18}$$

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

$$\begin{cases} t^{(i+1)} = T - n - 1, & \mathbf{T}^{(i+1)} = \mathbf{E}^{(i+1)} \\ u^{(i+1)} = T - m - 1, & \mathbf{U}^{(i+1)} = \mathbf{F}^{(i+1)} \end{cases} \tag{19}$$

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

$$\begin{cases} \mathbf{E}^{(i+1)} = \sum_{k=1}^{T} \mathrm{E}_{\xi_k}^{(i+1)}[\xi_k] \mathbf{A}_k^{(i+1)} \\ \mathbf{F}^{(i+1)} = \sum_{k=1}^{T} \left\{ \mathrm{E}_{\lambda_k}^{(i+1)}[1/\lambda_k] \mathbf{B}_k^{(i+1)} - \mathrm{E}_{\boldsymbol{\beta}}^{(i+1)}[\boldsymbol{\beta}] \left(\mathbf{b}_k^{(i+1)}\right)^{\mathrm{T}} \right. \\ \left. -\mathbf{b}_k^{(i+1)} \left(\mathrm{E}_{\boldsymbol{\beta}}^{(i+1)}[\boldsymbol{\beta}]\right)^{\mathrm{T}} + \mathrm{E}_{\lambda_k}^{(i+1)}[\lambda_k] \mathrm{E}_{\boldsymbol{\beta}}^{(i+1)}[\boldsymbol{\beta}\boldsymbol{\beta}^{\mathrm{T}}] \right\} \end{cases} \tag{20}$$

After fixed-point iteration N, the posterior PDF of the state trajectory is approximated as

$$p(\mathbf{x}_{0:T}|\mathbf{z}_{1:T}) \approx \mathrm{g}(\mathbf{x}_{0:T}; \hat{\mathbf{x}}_{0:T|T}^{(N)}, \mathbf{P}_{0:T|T}^{(N)}) \tag{21}$$

2) Calculation of Expectations: Using (12), (16) and (18), the required expectations are calculated as follows

$$\begin{cases} \mathrm{E}_{\xi_k}^{(i+1)}[\xi_k] = c_k \int_0^1 \xi_k \mathrm{G}(\xi_k; \alpha_k^{(i+1)}, \beta_k^{(i+1)}) d\xi_k \\ \mathrm{E}_{\lambda_k}^{(i+1)}[\lambda_k] = \dfrac{\sqrt{\varphi_k^{(i+1)}} K_{\varrho_k^{(i+1)}+1}(\sqrt{\eta_k^{(i+1)}\varphi_k^{(i+1)}})}{\sqrt{\eta_k^{(i+1)}} K_{\varrho_k^{(i+1)}}(\sqrt{\eta_k^{(i+1)}\varphi_k^{(i+1)}})} \\ \mathrm{E}_{\lambda_k}^{(i+1)}[\frac{1}{\lambda_k}] = \dfrac{\sqrt{\eta_k^{(i+1)}} K_{\varrho_k^{(i+1)}+1}(\sqrt{\eta_k^{(i+1)}\varphi_k^{(i+1)}})}{\sqrt{\varphi_k^{(i+1)}} K_{\varrho_k^{(i+1)}}(\sqrt{\eta_k^{(i+1)}\varphi_k^{(i+1)}})} - \\ \frac{2\varrho_k^{(i+1)}}{\varphi_k^{(i+1)}} \\ \mathrm{E}_{\boldsymbol{\beta}}^{(i+1)}[\boldsymbol{\beta}] = \hat{\boldsymbol{\beta}}^{(i+1)} \\ \mathrm{E}_{\boldsymbol{\beta}}^{(i+1)}[\boldsymbol{\beta}\boldsymbol{\beta}^{\mathrm{T}}] = \mathbf{P}_{\boldsymbol{\beta}}^{(i+1)} + \hat{\boldsymbol{\beta}}^{(i+1)} \left(\hat{\boldsymbol{\beta}}^{(i+1)}\right)^{\mathrm{T}} \\ \mathrm{E}_{\mathbf{Q}}^{(i+1)}[\mathbf{Q}^{-1}] = t^{(i+1)} \left(\mathbf{T}^{(i+1)}\right)^{-1} \\ \mathrm{E}_{\mathbf{R}}^{(i+1)}[\mathbf{R}^{-1}] = u^{(i+1)} \left(\mathbf{U}^{(i+1)}\right)^{-1} \end{cases} \tag{22}$$

where the integrals in (13) and (22) are calculated using the rectangular integration method with step size length 0.01.

Employing (10), $\mathbf{A}_k^{(i+1)}$, $\mathbf{B}_k^{(i+1)}$ and $\mathbf{b}_k^{(i+1)}$ are, respectively, calculated as follows [11]

$$\begin{cases} \mathbf{A}_k^{(i+1)} = (\hat{\mathbf{x}}_{k|T}^{(i+1)} - \mathbf{F}_k\hat{\mathbf{x}}_{k-1|T}^{(i+1)})(\hat{\mathbf{x}}_{k|T}^{(i+1)} - \mathbf{F}_k\hat{\mathbf{x}}_{k-1|T}^{(i+1)})^{\mathrm{T}} + \\ \mathbf{P}_{k|T}^{(i+1)} - \left(\mathbf{F}_k\mathbf{G}_{k-1}^{(i+1)}\mathbf{P}_{k|T}^{(i+1)}\right)^{\mathrm{T}} - \mathbf{F}_k\mathbf{G}_{k-1}^{(i+1)}\mathbf{P}_{k|T}^{(i+1)} + \\ \mathbf{F}_k\mathbf{P}_{k-1|T}^{(i+1)}\mathbf{F}_k^{\mathrm{T}} \\ \mathbf{B}_k^{(i+1)} = (\mathbf{z}_k - \mathbf{H}_k\hat{\mathbf{x}}_{k|T}^{(i+1)})(\mathbf{z}_k - \mathbf{H}_k\hat{\mathbf{x}}_{k|T}^{(i+1)})^{\mathrm{T}} + \\ \mathbf{H}_k\mathbf{P}_{k|T}^{(i+1)}\mathbf{H}_k^{\mathrm{T}} \\ \mathbf{b}_k^{(i+1)} = \mathbf{z}_k - \mathbf{H}_k\hat{\mathbf{x}}_{k|T}^{(i+1)} \end{cases} \tag{23}$$

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

The proposed robust RTS smoother is composed of variational approximations of posterior PDFs in (10)-(21) and calculations of expectations in (22)-(23). The implementation pseudo-code for the proposed robust RTS smoother is summarized in Table II, where $\boldsymbol{\Sigma}_w$ and $\boldsymbol{\Sigma}_v$ denote the nominal state and measurement noise covariance matrices, respectively, and σ denotes the initial variance of shape parameter. The proposed robust RTS smoother can be easily extended to a nonlinear case by employing standard Gaussian approximate smoother and modifying calculation of expectations in (23).

IV. Performance Validation

A problem of tracking an agile target is used to demonstrate the efficiency and superiority of the proposed robust RTS smoother. The agile target runs in a plane with a constant velocity. A constant velocity model is employed to track the agile target, and the position of the target is observed online in clutter. By choosing Cartesian coordinates and corresponding velocities as the state variables, the discrete-time linear state-space model can be formulated as (2), and state transition matrix and measurement matrix are, respectively, given by [10]

$$\mathbf{F}_k = \begin{bmatrix} \mathbf{I}_2 & \Delta t\mathbf{I}_2 \\ \mathbf{0} & \mathbf{I}_2 \end{bmatrix}, \qquad \mathbf{H}_k = \begin{bmatrix} \mathbf{I}_2 & \mathbf{0} \end{bmatrix} \tag{24}$$

TABLE III: Parameter selections of existing state-of-the-art smoothers and the proposed smoother.

Smoothers	Parameter selections
Standard RTS-true	$\mathbf{q}=\mathbf{q}_t$, $\mathbf{r}=\mathbf{r}_t$, $\mathbf{Q}=\mathbf{Q}_t$, $\mathbf{R}=\mathbf{R}_t$
Standard RTS-nominal	$\mathbf{q}=\mathbf{0}$, $\mathbf{r}=\mathbf{0}$, $\mathbf{Q}=\mathbf{\Sigma}_w$, $\mathbf{R}=\mathbf{\Sigma}_v$
Adaptive RTS	$\mathbf{q}=\mathbf{0}$, $\mathbf{r}=\mathbf{0}$, $\nu_0=6$, $\mu_0=4$, $\mathbf{V}_0=\mathbf{\Sigma}_w$, $\mathbf{M}_0=\mathbf{\Sigma}_v$, $N=10$
VB-ST-RTS	$\mathbf{q}=\mathbf{0}$, $\mathbf{r}=\mathbf{0}$, $t_0=6$, $\mathbf{T}_0=\mathbf{\Sigma}_w$, $u_0=4$, $\mathbf{U}_0=\mathbf{\Sigma}_v$, $a_0=c_0=5$, $b_0=d_0=1$, $N=10$
Skew t-RTS	$\mathbf{q}=\mathbf{0}$, $\mathbf{Q}=\mathbf{\Sigma}_w$, $\mathbf{\Delta}=\mathbf{\Omega}$, $\nu=\eta$, $N=10$
Student's t	$\mathbf{q}=\mathbf{0}$, $\mathbf{r}=\mathbf{0}$, $\mathbf{Q}=\mathbf{\Sigma}_w$, $\mathbf{R}=\mathbf{\Sigma}_v$, $\nu_1=3$, $\nu_2=3$
RTF-ST-RTS	$\mathbf{q}=\mathbf{0}$, $\mathbf{r}=\mathbf{0}$, $\mathbf{Q}=\mathbf{\Sigma}_w$, $\mathbf{R}=\mathbf{\Sigma}_v$, $\nu_1=5$, $\nu_2=5$, $N=10$
The proposed smoother	$\sigma=10^{-8}$, $\omega=1$, $\nu=5$, $N=10$

TABLE II: Implementation pseudo-code for the proposed robust RTS smoother.

Inputs: $\mathbf{z}_{1:T}$, $\hat{\mathbf{x}}_{0|0}$, $\mathbf{P}_{0|0}$, $\{\mathbf{F}_k,\mathbf{H}_k|1\leq k\leq T\}$, $\mathbf{\Sigma}_w$, $\mathbf{\Sigma}_v$, σ, ω, ν, N.

1. Initialization: $\left\{\mathrm{E}^{(0)}_{\xi_k}[\xi_k]=\mathrm{E}^{(0)}_{\lambda_k}[\lambda_k]=\mathrm{E}^{(0)}_{\lambda_k}[\frac{1}{\lambda_k}]=1|1\leq k\leq T\right\}$, $\mathrm{E}^{(0)}_{\mathbf{Q}}[\mathbf{Q}^{-1}]=\mathbf{\Sigma}_w^{-1}$, $\mathrm{E}^{(0)}_{\mathbf{R}}[\mathbf{R}^{-1}]=\mathbf{\Sigma}_v^{-1}$, $\mathrm{E}^{(0)}_{\boldsymbol{\beta}}[\boldsymbol{\beta}\boldsymbol{\beta}^{\mathrm{T}}]=\sigma\mathbf{I}_m$, $\mathrm{E}^{(0)}_{\boldsymbol{\beta}}[\boldsymbol{\beta}]=\mathbf{0}$.

for $i=0:N-1$

2. Calculate the modified noise covariance matrices $\tilde{\mathbf{Q}}^{(i+1)}_k$ and $\tilde{\mathbf{R}}^{(i+1)}_k$ from time sample 1 to time sample T using (11).

3. Calculate the modified mean vector of measurement noise $\tilde{\mathbf{r}}^{(i+1)}_k$ from time sample 1 to time sample T using (11).

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

5. Calculate $\mathbf{A}^{(i+1)}_k$, $\mathbf{B}^{(i+1)}_k$ and $\mathbf{b}^{(i+1)}_k$ using (23).

6. Calculate $\Delta^{(i+1)}_0$, $\Delta^{(i+1)}_1$, and $\Delta^{(i+1)}_2$ using (15).

7. Update $q^{(i+1)}(\xi_k)$ and $q^{(i+1)}(\lambda_k)$ using (12)-(15).

8. Calculate $\mathrm{E}^{(i+1)}_{\xi_k}[\xi_k]$, $\mathrm{E}^{(i+1)}_{\lambda_k}[\lambda_k]$ and $\mathrm{E}^{(i+1)}_{\lambda_k}[\frac{1}{\lambda_k}]$ using (22).

9. Update $q^{(i+1)}(\boldsymbol{\beta})$ using (16)-(17).

10. Calculate $\mathrm{E}^{(i+1)}_{\boldsymbol{\beta}}[\boldsymbol{\beta}]$ and $\mathrm{E}^{(i+1)}_{\boldsymbol{\beta}}[\boldsymbol{\beta}\boldsymbol{\beta}^{\mathrm{T}}]$ using (22).

11. Update $q^{(i+1)}(\mathbf{Q})$ and $q^{(i+1)}(\mathbf{R})$ using (18)-(20).

12. Calculate $\mathrm{E}^{(i+1)}_{\mathbf{Q}}[\mathbf{Q}^{-1}]$ and $\mathrm{E}^{(i+1)}_{\mathbf{R}}[\mathbf{R}^{-1}]$ using (22).

end

13. $\{\hat{\mathbf{x}}_{k|T}=\hat{\mathbf{x}}^{(N)}_{k|T},\mathbf{P}_{k|T}=\mathbf{P}^{(N)}_{k|T}|0\leq k\leq T\}$.

Outputs: $\{\hat{\mathbf{x}}_{k|T},\mathbf{P}_{k|T}|0\leq k\leq T\}$.

where the state vector $\mathbf{x}_k\triangleq[x_k\ y_k\ \dot{x}_k\ \dot{y}_k]$, x_k, y_k, $\dot{x}_k$ and $\dot{y}_k$ denote the Cartesian coordinates and corresponding velocities, respectively, and the parameter $\Delta t=1$s denotes the sampling interval.

Outlier contaminated state noise, which has a heavy-tailed and symmetric distribution, is generated in terms of [25]– [27]

$$\mathbf{w}_k\sim\left\{\begin{array}{ll}\mathrm{N}(\mathbf{0},\mathbf{\Sigma}_w) & \text{w.p. } 0.90\\ \mathrm{N}(\mathbf{0},500\mathbf{\Sigma}_w) & \text{w.p. } 0.10\end{array}\right. \tag{25}$$

where w.p. denotes "with probability", and the nominal state noise covariance matrix $\mathbf{\Sigma}_w$ is given by

$$\mathbf{\Sigma}_w=q\left[\begin{array}{cc}\frac{\Delta t^3}{3}\mathbf{I}_2 & \frac{\Delta t^2}{2}\mathbf{I}_2\\ \frac{\Delta t^2}{2}\mathbf{I}_2 & \Delta t\mathbf{I}_2\end{array}\right] \tag{26}$$

where the noise parameter $q=0.1\mathrm{m}^2/\mathrm{s}^3$.

Outlier corrupted measurement noise is assumed to have a Skew t-distribution, which is produced in terms of [13]

$$\left\{\begin{array}{l}\mathbf{v}_k\sim\mathrm{N}(\mathbf{\Omega}\mathbf{u}_k,\mathbf{\Lambda}_k^{-1}\mathbf{\Sigma}_v)\\ \mathbf{u}_k\sim\mathrm{N}_+(\mathbf{0},\mathbf{\Lambda}_k^{-1})\\ [\mathbf{\Lambda}_k]_{ii}\sim\mathrm{G}(\frac{\eta}{2},\frac{\eta}{2})\end{array}\right. \tag{27}$$

and the corresponding PDF is formulated as

$$p(\mathbf{v}_k)=\mathrm{ST}(\mathbf{v}_k;\mathbf{0},\mathbf{\Sigma}_v,\mathbf{\Omega},\eta) \tag{28}$$

where the nominal measurement noise covariance matrix $\mathbf{\Sigma}_v=100\mathbf{I}_2$, and $\mathbf{\Omega}=10\mathbf{I}_2$ with shape parameters as diagonal elements, and $\mathbf{\Lambda}_k$ is a 2×2 diagonal matrix whose random diagonal elements $[\mathbf{\Lambda}_k]_{ii}$ are independent and identically distributed, and $\mathbf{u}_k$ is an auxiliary random vector, and $\eta=2$ is a dof parameter.

In this simulation, the proposed robust RTS smoother is compared with the existing state-of-the-art smoothers, including Standard RTS smoother, Adaptive RTS smoother [18], VB-ST-RTS smoother [11], Skew t-RTS smoother [13], Student's t-smoother [17], and RTF-ST-RTS smoother [15]. To better show the advantages of the proposed robust RTS smoother, both the standard RTS smoother with nominal noise mean vectors and covariance matrices and the standard RTS smoother with true noise mean vectors and covariance matrices are compared with the proposed method. For convenience, the two standard RTS smoothers mentioned above are, respectively, abbreviated as "standard RTS-nominal smoother" and "standard RTS-true smoother". The parameter selections of existing state-of-the-art smoothers and the proposed smoother are listed in Table III, where $\mathbf{q}_t=\mathbf{0}$, $\mathbf{r}_t=[14.0\ \ 14.0]^{\mathrm{T}}$, $\mathbf{Q}_t=50.9\mathbf{\Sigma}_w$ and $\mathbf{R}_t=\mathrm{diag}([2400\ \ 2400])$ denote the true mean vectors and

TABLE IV: ARMSEs and implementation times in a single Monte Carlo run of existing state-of-the-art smoothers and the proposed smoother when $N = 10$.

Smoothers	$\mathrm{ARMSE_{pos}}$ (m)	$\mathrm{ARMSE_{vel}}$ (m/s)	Time (s)
Standard RTS-true	19.222	4.212	0.010
Standard RTS-nominal	28.161	4.370	0.010
Adaptive RTS	27.690	4.207	0.139
VB-ST-RTS	16.866	3.474	0.196
Skew t-RTS	25.878	5.173	0.747
Student's t	167.417	14.193	0.013
RTF-ST-RTS	21.144	4.289	1.094
The proposed smoother	14.901	3.386	0.915

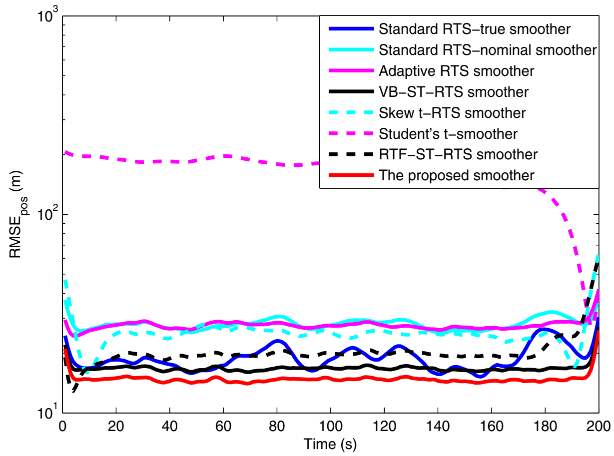

Fig. 3: RMSEs of the position when $N = 10$.

covariance matrices of state and measurement noises, respectively, and $\mathbf{r}_t$ and $\mathbf{R}_t$ are obtained based on random sampling method using 10^5 samples. The parameters of existing state-of-the-art smoothers are chosen as suggested by the original authors. The true initial state vector and the initial estimation error covariance matrix are, respectively, set as $\mathbf{x}_0 = [10\ 10\ 10\ 10]^{\mathrm{T}}$ and $\mathbf{P}_{0|0} = \mathrm{diag}([1000\ 1000\ 1000\ 1000])$, and the initial state estimate $\hat{\mathbf{x}}_{0|0}$ is randomly chosen from a Gaussian distribution $N(\mathbf{x}_0, \mathbf{P}_{0|0})$. The simulation time is set as 200s, and 1000 independent Monte Carlo runs are executed. All smoothing algorithms are coded with MATLAB and simulations are run on a computer with Intel Core i7-6900K CPU @ 3.20 GHz.

To compare the estimation accuracy of existing smoothers and the proposed smoother, the root mean square errors (RMSEs) and the averaged RMSEs (ARMSEs) are selected as performance metrics. The RMSE and ARMSE of position are defined as follows [10]

$$\begin{cases} \mathrm{RMSE_{pos}} = \sqrt{\frac{1}{L}\sum_{s=1}^{L}\left((x_k^s - \hat{x}_{k|T}^s)^2 + (y_k^s - \hat{y}_{k|T}^s)^2\right)} \\ \mathrm{ARMSE_{pos}} = \sqrt{\frac{1}{LT}\sum_{k=1}^{T}\sum_{s=1}^{L}\left((x_k^s - \hat{x}_{k|T}^s)^2 + (y_k^s - \hat{y}_{k|T}^s)^2\right)} \end{cases} \tag{29}$$

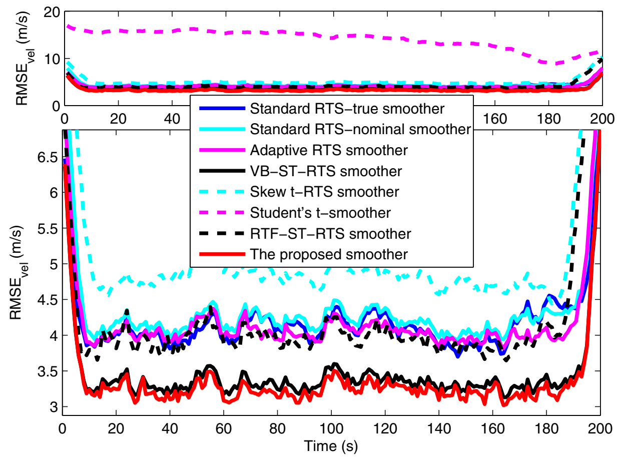

Fig. 4: RMSEs of the velocity when $N = 10$.

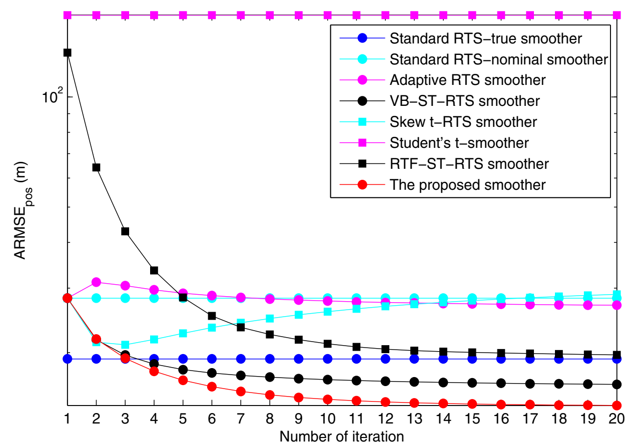

Fig. 5: ARMSEs of the position when $N = 1, 2, \ldots, 20$.

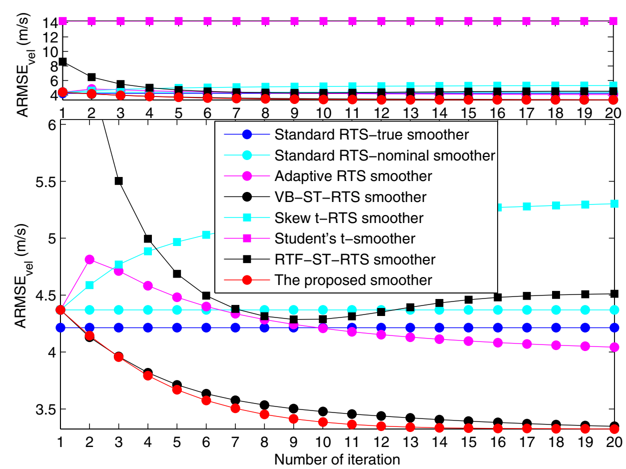

Fig. 6: ARMSEs of the velocity when $N = 1, 2, \ldots, 20$.

where $\mathrm{RMSE_{pos}}$ denotes the RMSE of position, and (x_k^s, y_k^s) and $(\hat{x}_{k|T}^s, \hat{y}_{k|T}^s)$ are, respectively, the true and estimated positions at the s-th Monte Carlo run, and $T = 200$ and $L = 1000$

denote the simulation steps and the total number of Monte Carlo runs, respectively. Similar to the RMSE and ARMSE in position, we can also define the RMSE and ARMSE in velocity, which are, respectively, formulated as $\mathrm{RMSE}_{\mathrm{vel}}$ and $\mathrm{ARMSE}_{\mathrm{vel}}$.

Fig.3 – Fig.4 and Table IV, respectively, show the RMSEs and ARMSEs of position and velocity and the implementation times in a single Monte Carlo run of the proposed robust RTS smoother and existing smoothers when $N = 10$. The ARMSEs of position and velocity from the proposed robust RTS smoother and existing smoothers when $N = 1, 2, \ldots, 20$ are, respectively, shown in Fig.5 – Fig.6. It is observed from Fig.3 – Fig.4 and Table IV that the proposed robust RTS smoother has smaller RMSEs and ARMSEs than existing smoothers. Also, we can observe from Table IV that the proposed robust RTS smoother needs more implementation times than existing smoothers except for the existing RTF-ST-RTS smoother. Moreover, it can be observed from Fig.5 – Fig.6 that the proposed robust RTS smoother has smaller ARMSEs than existing smoothers when $N \geq 4$. Thus, the proposed robust RTS smoother has better estimation accuracy but higher computational complexity than existing state-of-the-art smoothers.

V. Conclusions

In this paper, a novel robust RTS smoother was proposed by modelling the state noise as Slash distributed and modelling the measurement noise as GH skew Student's t-distributed. A novel hierarchical Gaussian state-space model was constructed by formulating the Slash distribution as a Gaussian scale mixture form and formulating the GH skew Student's t-distribution as a Gaussian variance-mean mixture form, based on which the state trajectory, mixing parameters and unknown noise parameters were jointly inferred using the VB approach. The posterior PDFs of mixing parameters of the Slash and GH skew Student's t-distributions were, respectively, approximated as truncated Gamma and GIG. Simulation results demonstrated that the proposed robust RTS smoother has better estimation accuracy but higher computational complexity than existing state-of-the-art smoothers.

References

[1] D. Simon, *Optimal state estimation: Kalman, H infinity, and nonlinear approaches*. John Wiley & Sons, 2006.

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

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

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

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

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

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

[8] Y. L. Huang, Y. G. Zhang, B. Xu, Z. M. Wu, and J. Chambers, "A new adaptive extended Kalman filter for cooperative localization," *IEEE Trans. Aerosp. Electron. Syst.*, DOI: 10.1109/TAES.2017.2756763.

[9] S. Saha, "Noise robust online inference for linear dynamic systems," [Online]. Available: http://arxiv.org/pdf/1504.05723, 2015.

[10] Y. L. Huang, Y. G. Zhang, Z. M. Wu, N. Li, and J. Chambers, "A novel robust Student's t based Kalman filter," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 53, no. 1, pp. 1545–1554, Feb. 2017.

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

[12] R. Piché, S. Särkkä, and J. Hartikainen, "Recursive outlier-robust filtering and smoothing for nonlinear systems using the multivariate Student-t distribution," In Proc. of *IEEE International Workshop on Machine Learning for Signal Processing (MLSP)*, Sept. 2012.

[13] H. Nurminen, T. Ardeshiri, R. Piché, and F. Gustafsson, "Robust inference for state-space models with skew measurement noise," *IEEE Signal Process. Lett.*, 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, 2017.

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

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

[17] M. Roth, T. Ardeshiri, E. Özkan, and F. Gustafsson, "Robust Bayesian filtering and smoothing using Student's t distribution," arXiv preprint arXiv:1703.02428, 2017.

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

[19] K. Aas and I. H. Haff, "The generalized hyperbolic skew Student's t-distribution," *Journal of Financial Econometrics*, vol. 4, no. 2, pp. 275–309, Mar. 2006.

[20] 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 Trans. Syst., Man, Cybern.: Syst.*, DOI: 10.1109/TSMC.2017.2778269.

[21] A. Azzalini and A. Capitanio, "Statistical applications of the multivariate Skew normal distribution," *J. R. Stat. Soc. B*, vol. 61, no. 3, pp. 579–602, Mar. 1999.

[22] A. Azzalini and A. Capitanio, "Distributions generated by perturbation of symmetry with emphasis on a multivariate skew t-distribution," *J. R. Stat. Soc. B*, vol. 65, no. 2, pp. 367–389, Apr. 2003.

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

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

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

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

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