A Robust Generalized t Distribution Based Kalman Filter

Mingming Bai, Chengjiao Sun, Yonggang Zhang, *Senior Member, IEEE*

Abstract—Since the Gaussian-inverse Wishart hierarchical form has similar properties to Student's t distribution, we name it generalized t distribution in this paper. Based on this, a robust generalized t distribution based Kalman filter (GTKF) is proposed for state-space models that are eroded by state and measurement outliers. Different from the existing algorithms, the state transition and measurement likelihood densities are directly modeled as generalized t distributions by employing the one-step smoothing strategy. An analytical closed-form solution can be obtained through the variational inference approach. Moreover, two variants of the proposed GTKF are also presented to apply to different engineering scenarios. Simulation and experimental examples demonstrate that the proposed GTKFs yield improved robustness over the existing algorithms.

Index Terms—Roust estimation, heavy-tailed noise, Student's t distribution, outliers, variational inference

I. INTRODUCTION

ESTIMATING latent state from a series of noisy measurements is a fundamental task in plenty of engineering fields such as robotics [1], multi-target tracking [2] and navigation [3]. For linear Gaussian state-space models, the Kalman filter (KF) presents an optimal estimation and has been extensively applied due to its simple structure and real-time operation [4]. Outliers are widespread in engineering applications. For example, in the cooperative localization (CL) for autonomous underwater vehicles (AUVs), the acoustic ranging outliers may be induced by sound ray bending and multipath effect of underwater acoustic channel [5], [6]. The velocity outliers often occur when the Doppler velocity log (DVL) is misaligned with the body framework [7]. Other outliers can also be caused by sensor failures, model mismatches or power surges [8]. The performance of KF degrades dramatically given such non-Gaussian noises.

To counteract the impacts of outlier interferences, a batch of M-estimators have been developed by embedding different error criterions into the maximum likelihood framework. The Huber KF (HKF) is proposed by minimizing the Huber cost function [9]. The maximum correntropy KF (MCKF) is proposed by maximizing the correntropy cost function [10]. By minimizing an error entropy cost function, the minimum error entropy KF (MEEKF) is also developed to suppress the negative effects of outliers. Although present better robustness than the classical KF, the M-estimators discard the statistical

Corresponding author is Y. Zhang.
M. Bai and Y. Zhang are with the College of Intelligent Systems Science and Engineering, Harbin Engineering University, Harbin, 150001, China; C. Sun is with the School of Computer Engineering, Hubei University of Arts and Science, Xiangyang, 441053, China. (e-mail: mingming.bai@hrbeu.edu.cn; jiao1952@126.com; zhangyg@hrbeu.edu.cn).

property contained in the non-Gaussian noises, which degrades their estimation performance [11]. Motivated by this, some Student's t distribution based robust filters are presented to fully consider the tail behaviors of the non-Gaussian noises [12]-[15]. The Student's t filter (STF) is derived based on the Bayesian rule by formulating the joint predicted probability density function (PDF) of the state and measurement as a multivariate Student's t distribution [12], [13]. More advanced, the Student's t based KF (STKF) is proposed by formulating the Student's t distributed predicted and measurement likelihood PDFs as Gaussian-Gamma hierarchical (GGH) forms, and the state variable and modeling parameters are jointly estimated through the variational Bayesian (VB) principle [11]. The STKF is further improved by estimating the degrees of freedom (Dof) parameters online [14], [15]. However, upon addressing the state outliers, the exiting STKFs model the predicted PDF rather than the state transition PDF as a Student's t distribution, which leads to poor robustness to the state outliers. Taking such indirectly modeling manner because the state transition PDF is non-conjugate under the GGH formulation. In [16], a robust inference technology is proposed by factorizing the measurement likelihood PDF as a Gaussian-inverse Wishart hierarchical (GIWH) form, whereas it performs poorly when the system is eroded by state outliers since this inference technology is derived based on a Gaussian state noise assumption. The GIWH form has similar properties to Student's t distribution, which we name generalized t in this paper [16]. The generalized t distribution is also employed to design adaptive KF [17].

In this paper, a robust generalized t distribution based KF (RTKF) is proposed. An analytical closed-form solution can be obtained by utilizing the VB principle. The main contributions of this paper are summarized as follows: 1) Different from the existing STKFs, we propose to directly model the state transition and measurement likelihood PDFs as generalized t distributions by employing the one-step smoothing strategy, which makes the proposed algorithms achieve better robustness to the state outliers. 2) Two variants, namely, the low computational complexity GTKF (LCC-GTKF) and nonlinear GTKF(NGTKF), are also presented to apply to different engineering scenarios. 3) Simulation and experimental examples are conducted to demonstrate the superiority and usefulness of the proposed algorithms.

The remainder of this paper is organized as follows. Section II is devoted to providing the preliminaries and problem formulations. The GTKF are developed in Section III. Variants of GTKF and discussions are presented in Section IV. Simulation analysis and experimental test are given in Section V.

Conclusions are drawn in Section VI.

Notations: We use $\mathbf{I}_h$ to represent the $h-$dimensional identity matrix, and use $\Gamma_m(\cdot)$ to represent the $m-$dimensional Gamma function, and use $p(\cdot)$ to represent the PDF. We denote $\mathbf{E}[\cdot]$ the mathematical expectation, and $|\cdot|$ the determinant operation, and $\log(\cdot)$ the logarithmic operation, and $\exp(\cdot)$ the exponential operation. We denote $N(\cdot; \boldsymbol{\mu}, \mathbf{V})$ the Gaussian PDF with mean vector $\boldsymbol{\mu}$ and covariance matrix $\mathbf{V}$, $G(\cdot; \theta, \vartheta)$ the Gamma PDF with shape parameter θ and rate parameter ϑ, $IW(\cdot; \boldsymbol{\Delta}, \sigma)$ the inverse Wishart PDF with scale matrix $\boldsymbol{\Delta}$ and Dof σ, $St(\cdot; \mathbf{y}, \mathbf{Y}, \eta)$ the Student's t PDF with mean $\mathbf{y}$, scale matrix matrix $\mathbf{Y}$ and Dof η. The symbol $D_{KL}(\cdot||\cdot)$ denotes the Kullback-Leilber divergence (KLD), and superscript $[l+1]$ denotes the $[l+1]^{th}$ iteration.

II. PRELIMINARIES AND PROBLEM FORMULATIONS

A. Linear State-Space Models

A linear and non-Gaussian discrete-time state-space model is described as follows

$$\begin{cases} \mathbf{x}_t = \mathbf{A}_t \mathbf{x}_{t-1} + \mathbf{v}_{t-1} \\ \mathbf{z}_t = \mathbf{C}_t \mathbf{x}_t + \mathbf{n}_t \end{cases} \quad (1)$$

where t denotes the discrete time index. $\mathbf{x}_t \in \mathbb{R}^p$ is the latent state and $\mathbf{z}_t \in \mathbb{R}^d$ is the measurement vector. $\mathbf{A}_t \in \mathbb{R}^{p \times p}$ and $\mathbf{C}_t \in \mathbb{R}^{p \times d}$ are, respectively, the known state transition and measurement matrices. $\mathbf{v}_t$ and $\mathbf{n}_t$ are, respectively, the outliers-contaminated state and measurement noises, which can be modeled by the following mixture forms

$$p(\mathbf{v}_t) = (1 - \alpha_t^v) N(\mathbf{v}_t; \mathbf{0}, \mathbf{Q}_0) + \alpha_t^v \tilde{p}(\mathbf{v}_t) \quad (2)$$

$$p(\mathbf{n}_t) = (1 - \alpha_t^n) N(\mathbf{n}_t; \mathbf{0}, \mathbf{R}_0) + \alpha_t^n \tilde{p}(\mathbf{n}_t) \quad (3)$$

where α_t^v and α_t^n are outlier probabilities. $\mathbf{Q}_0$ and $\mathbf{R}_0$ are, respectively, the nominal state noise covariance matrix (SNCM) and measurement noise covariance matrix (MNCM). We denote $\mathbf{Q}_t$ and $\mathbf{R}_t$ the unknown real SNCM and MNCM, and $\tilde{p}(\cdot)$ is the PDF of outliers. The measurement outliers are mostly caused by failing observations [9], and the state outliers are usually induced by unexpected dynamics, modeling errors and even sensor malfunctions [11], [14], such as the severe manoeuvering for a kinematic evolution model in the target tracking applications [18], and the measurement anomalies of DVL and compass for a dead-reckoning (DR) model in the CL for AUVs [14].

B. Student's t based Kalman filters

To cope with the heavy-tailed state and measurement noises, the STKFs propose to model the predicted and measurement likelihood PDFs as Student's t distributions [11], [14]

$$p(\mathbf{x}_t | \mathbf{z}_{1:t-1}) = St(\mathbf{x}_t; \hat{\mathbf{x}}_t^-, \boldsymbol{\Sigma}_t, \omega_x) \quad (4)$$

$$p(\mathbf{z}_t | \mathbf{x}_t) = St(\mathbf{z}_t; \mathbf{C}_t \mathbf{x}_t, \mathbf{R}_0, \omega_z) \quad (5)$$

where $\boldsymbol{\Sigma}_t$ is the predicted scale matrix. Employing the GGH property of Student's t distribution, the above densities are rewritten as

$$p(\mathbf{x}_t | \mathbf{z}_{1:t-1}) = \int N\left(\mathbf{x}_t; \hat{\mathbf{x}}_t^-, \frac{\boldsymbol{\Sigma}_t}{\xi_t}\right) G\left(\xi_t; \frac{\omega_x}{2}, \frac{\omega_x}{2}\right) d\xi_t \quad (6)$$

$$p(\mathbf{z}_t | \mathbf{x}_t) = \int N\left(\mathbf{z}_t; \mathbf{C}_t \mathbf{x}_t, \frac{\mathbf{R}_t}{\lambda_t}\right) G\left(\lambda_t; \frac{\omega_z}{2}, \frac{\omega_z}{2}\right) d\lambda_t \quad (7)$$

To further suppress the affects of state outliers, meanwhile, the predicted scale matrix $\boldsymbol{\Sigma}_t$ is estimated online by selecting an IW distribution as the conjugate prior [11], [14]

$$p(\boldsymbol{\Sigma}_t) = IW(\boldsymbol{\Sigma}_t; \boldsymbol{\Phi}_0, \phi_0) \quad (8)$$

The posterior PDF is assumed to be Gaussian distributed, i.e. $p(\mathbf{x}_{t-1} | \mathbf{z}_{1:t-1}) = N(\mathbf{x}_{t-1}; \hat{\mathbf{x}}_{t-1}, \mathbf{P}_{t-1})$. Following the VB principle, the intractable joint posterior PDF $p(\mathbf{x}_t, \boldsymbol{\Sigma}_t, \xi_t, \lambda_t | \mathbf{z}_{1:t})$ is approximated by a factorable product of several independent PDFs, i.e.,

$$p(\mathbf{x}_t, \boldsymbol{\Sigma}_t, \xi_t, \lambda_t | \mathbf{z}_{1:t}) \approx q(\mathbf{x}_t) q(\boldsymbol{\Sigma}_t) q(\xi_t) q(\lambda_t) \quad (9)$$

where $q(\cdot)$ is the approximate posterior PDF, which can be iteratively computed by minimizing the KLD between the approximate posterior PDF and real posterior PDF [11], [14]:

$$\{q(\mathbf{x}_t), q(\boldsymbol{\Sigma}_t), q(\xi_t), q(\lambda_t)\} = \min$$
$$D_{KL}(q(\mathbf{x}_t) q(\boldsymbol{\Sigma}_t) q(\xi_t) q(\lambda_t) || p(\mathbf{x}_t, \boldsymbol{\Sigma}_t, \xi_t, \lambda_t | \mathbf{z}_{1:t})) \quad (10)$$

where $D_{KL}(q(a)||p(a)) \triangleq \int q(a) \log \frac{q(a)}{p(a)} da$.

Embedding the modified predicted covariance matrix and MNCM into the KF framework yields the STKFs [11], [14].

C. Motivation of This Work

Although the existing STKFs can effectively suppress the influence of measurement outliers, they exhibit poor robustness to the state outliers due to the significant approximate errors in dealing with the heavy-tailed state noise. The detailed motivation of this work is presented as follows.

1) Large non-conjugate approximation errors. In view of (2), the heavy-tailed state noise is described as a Student's t distribution [12]. Naturally, the state transition PDF can be formulated by

$$p(\mathbf{x}_t | \mathbf{x}_{t-1}) = \int N\left(\mathbf{x}_t; \mathbf{A}_t \mathbf{x}_{t-1}, \frac{\mathbf{Q}_0}{\gamma_t}\right) G\left(\gamma_t; \frac{\omega_x}{2}, \frac{\omega_x}{2}\right) d\gamma_t \quad (11)$$

Using the Chapman-Kolmogorov equation and exploiting the Gaussian assumption for posterior density in STKFs, the predicted PDF should be formulated as

$$p(\mathbf{x}_t | \mathbf{z}_{1:t-1}) = \int p(\mathbf{x}_t | \mathbf{x}_{t-1}) p(\mathbf{x}_{t-1} | \mathbf{z}_{1:t-1}) d\mathbf{x}_{t-1}$$
$$= \int N(\mathbf{x}_t; \mathbf{A}_t \hat{\mathbf{x}}_{t-1}, \bar{\mathbf{P}}_t(\gamma_t)) G(\gamma_k; \frac{\omega_x}{2}, \frac{\omega_x}{2}) d\gamma_t \quad (12)$$

where the scale matrix $\bar{\mathbf{P}}_t(\gamma_t)$ is computed by $\bar{\mathbf{P}}_t(\gamma_t) = \mathbf{A}_t \mathbf{P}_{t-1} \mathbf{A}_t^T + \frac{\mathbf{Q}_0}{\gamma_t}$. It is obvious that the integral in (12) is analytically intractable, and the variational update will be non-conjugate under the GGH forms. Hence, an approximation is made in STKFs to acquire a conjugate update [11], [14], i.e.,

$$\bar{\mathbf{P}}_t(\gamma_t) \approx (\mathbf{A}_t \mathbf{P}_{t-1} \mathbf{A}_t^T + \mathbf{Q}_0)/\gamma_t \quad (13)$$

Using (13) in (12) and using $\boldsymbol{\Sigma}_t = \mathbf{A}_t \mathbf{P}_{t-1} \mathbf{A}_t^{\mathrm{T}} + \mathbf{Q}_0$ yield the approximate equation in (4). That is, it is the predicted PDF rather than the state transition PDF that is modeled as a student's t distribution. This *indirectly* modeling manner leads to poor robustness of the STKFs to the state outliers.

2) The Gaussian-inverse Wishart density follows a generalized t distribution. If the PDF of a random vector $\mathbf{y} \in \mathbb{R}^m$ can be formulated as the following GIWH form

$$p(\mathbf{y}) = \int \mathrm{N}(\mathbf{y}; \boldsymbol{\mu}_y, \boldsymbol{\Lambda}_y) \mathrm{IW}(\boldsymbol{\Lambda}_y; \mathbf{F}, f) d\boldsymbol{\Lambda}_y \qquad (14)$$

Then the PDF $p(\mathbf{y})$ follows a generalized t distribution [16], [17], i.e.,

$$p(\mathbf{y}) = \frac{\Gamma_m\left(\frac{f+1}{2}\right)}{\sqrt{|\pi\mathbf{F}|}\Gamma_m\left(\frac{f}{2}\right)} \left[1 + (\mathbf{y} - \boldsymbol{\mu}_y)^{\mathrm{T}} \mathbf{F}^{-1} (\mathbf{y} - \boldsymbol{\mu}_y)\right]^{-\frac{f+1}{2}} \qquad (15)$$

Similar to the standard Student's t distribution, the generalized t distribution is also a sub-exponential distribution and has much heavier tails than the Gaussian. Particularly, if we set the scale matrix $\mathbf{F} = f\boldsymbol{\Delta}$ and dimension $m > 1$, the generalized t distribution exhibits heavier tail than the standard Student's t distribution $\mathrm{St}(\mathbf{y}; \boldsymbol{\mu}_y, \boldsymbol{\Delta}, f)$ since the exponent of the former is lager than that of the later. When dimension $m = 1$, the generalized t distribution degrades to the standard Student's t distribution. This setting for scale matrix $\mathbf{F}$ is commonly used in the initialization of VB, see Algorithm 1.

The aforementioned problem represents the main motivation of this work. To achieve better robustness to the state and measurement outliers, a GTKF is proposed later, where the state transition and measurement likelihood PDFs are *directly* modeled as generalized t distributions. The analytical closed-form solution are derived by the VB principle.

III. PROPOSED GTKF

A. Hierarchical Gaussian State-Space Model

Exploiting (14) and (15), the heavy-tailed state and measurement noises can be modeled as generalized t distributions:

$$p(\mathbf{v}_t) = \int \mathrm{N}(\mathbf{v}_t; \mathbf{0}, \mathbf{Q}_t) \mathrm{IW}(\mathbf{Q}_t; \mathbf{U}_0, u_0) d\mathbf{Q}_t \qquad (16)$$

$$p(\mathbf{n}_t) = \int \mathrm{N}(\mathbf{n}_t; \mathbf{0}, \mathbf{R}_t) \mathrm{IW}(\mathbf{R}_t; \mathbf{V}_0, v_0) d\mathbf{R}_t \qquad (17)$$

Using (1) in the above equations, the state transition and measurement likelihood PDFs are, respectively, formulated by

$$p(\mathbf{x}_t|\mathbf{x}_{t-1}) = \int \mathrm{N}(\mathbf{x}_t; \mathbf{A}_t\mathbf{x}_{t-1}, \mathbf{Q}_{t-1}) \mathrm{IW}(\mathbf{Q}_{t-1}; \mathbf{U}_0, u_0) d\mathbf{Q}_{t-1} \qquad (18)$$

$$p(\mathbf{z}_t|\mathbf{x}_t) = \int \mathrm{N}(\mathbf{z}_t; \mathbf{C}_t\mathbf{x}_t, \mathbf{R}_t) \mathrm{IW}(\mathbf{R}_t; \mathbf{V}_0, v_0) d\mathbf{R}_t \qquad (19)$$

As mentioned earlier, the scale matrices for (18) and (19) are initialized as $\mathbf{U}_0 = u_0\mathbf{Q}_0$ and $\mathbf{V}_0 = v_0\mathbf{R}_0$, respectively.

The Gaussian assumption for posterior density in STKFs [11] is inherited, namely

$$p(\mathbf{x}_{t-1}|\mathbf{z}_{1:t-1}) = \mathrm{N}(\mathbf{x}_{t-1}; \hat{\mathbf{x}}_{t-1}, \mathbf{P}_{t-1}) \qquad (20)$$
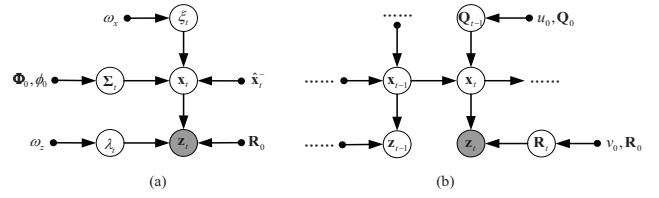

Fig. 1: Graphical models comparisons. In (a), the state outliers are indirectly addressed by estimating the predicted scale matrix, while (b) directly estimates the SNCM on the Markov nets. (a) Graphical model for the existing STKFs. (b) Graphical model for the proposed algorithm.

Summarizing (18)-(20) we obtain the hierarchical Gaussian state-space model

$$\begin{cases} p(\mathbf{x}_t|\mathbf{x}_{t-1}, \mathbf{Q}_{t-1}) = \mathrm{N}(\mathbf{x}_t; \mathbf{A}_t\mathbf{x}_{t-1}, \mathbf{Q}_{t-1}) \\ p(\mathbf{z}_t|\mathbf{x}_t, \mathbf{R}_t) = \mathrm{N}(\mathbf{z}_t; \mathbf{C}_t\mathbf{x}_t, \mathbf{R}_t) \\ p(\mathbf{x}_{t-1}|\mathbf{z}_{1:t-1}) = \mathrm{N}(\mathbf{x}_{t-1}; \hat{\mathbf{x}}_{t-1}, \mathbf{P}_{t-1}) \\ p(\mathbf{Q}_{t-1}|\mathbf{z}_{1:t-1}) = \mathrm{IW}(\mathbf{Q}_{t-1}; \mathbf{U}_0, u_0) \\ p(\mathbf{R}_t|\mathbf{z}_{1:t-1}) = \mathrm{IW}(\mathbf{R}_t; \mathbf{V}_0, v_0) \end{cases} \qquad (21)$$

It is worth noting that, unlike the existing algorithms where the predicted PDF is modeled by a heavy-tailed distribution, in the proposed hierarchical Gaussian probabilistic model, we directly model the state transition PDF and measurement likelihood PDF as generalized t distributions. The graphical models comparisons are depicted in Fig. 1.

B. Proposed GTKF

To estimate the latent state $\mathbf{x}_t$ given such outliers eroded state and measurement noises, the joint posterior PDF $p(\boldsymbol{\Theta}_t|\mathbf{z}_{1:t})$ should be computed firstly, where $\boldsymbol{\Theta}_t \triangleq \{\mathbf{x}_t, \mathbf{x}_{t-1}, \mathbf{Q}_{t-1}, \mathbf{R}_t\}$. According to the VB principle, the real but unknown PDF $p(\boldsymbol{\Theta}_t|\mathbf{z}_{1:t})$ can be approximated by a factorable density, i.e.,

$$q(\boldsymbol{\Theta}_t) = q(\mathbf{x}_t, \mathbf{x}_{t-1})q(\mathbf{Q}_{t-1})q(\mathbf{R}_t) \qquad (22)$$

where each $q(\cdot)$ can be analytically updated by solving the following equation [19]

$$q(\boldsymbol{\tau}_t) \propto \exp\left\{\mathbf{E}_{\boldsymbol{\Theta}_t^{(-\boldsymbol{\tau}_t)}}\left[\log p(\boldsymbol{\Theta}_t, \mathbf{z}_t|\mathbf{z}_{1:t-1})\right]\right\} \qquad (23)$$

where $\boldsymbol{\tau}_t \in \boldsymbol{\Theta}_t$, and $\boldsymbol{\Theta}_t^{(-\boldsymbol{\tau}_t)}$ is the complementary set satisfying $\boldsymbol{\tau}_t \bigcup \boldsymbol{\Theta}_t^{(-\boldsymbol{\gamma}_t)} = \boldsymbol{\Theta}_t$. The joint PDF $p(\boldsymbol{\Theta}_t, \mathbf{z}_t|\mathbf{z}_{1:t-1})$ in (23) can be factorized as

$$p(\boldsymbol{\Theta}_t, \mathbf{z}_t|\mathbf{z}_{1:t-1}) = p(\mathbf{z}_t|\mathbf{x}_t, \mathbf{R}_t)p(\mathbf{x}_t|\mathbf{x}_{t-1}, \mathbf{Q}_{t-1})p(\mathbf{x}_{t-1}|\mathbf{z}_{1:t-1})$$
$$\times p(\mathbf{Q}_{t-1}|\mathbf{z}_{1:t-1})p(\mathbf{R}_t|\mathbf{z}_{1:t-1}) \qquad (24)$$

The fixed-point iterative strategy is utilized to solve (23). The update process of the approximate PDFs in $[l+1]^{th}$ iteration are given in the following subsections.

1) Update of $q^{[l+1]}(\mathbf{Q}_{t-1})$ and $q^{[l+1]}(\mathbf{R}_t)$: Letting $\boldsymbol{\tau}_t = \mathbf{Q}_{t-1}$ and substituting (21), (24) in (23), the PDF $q^{[l+1]}(\mathbf{Q}_{t-1})$ can be updated as a IW distribution. Similarly, letting $\boldsymbol{\tau}_t = \mathbf{R}_t$ and substituting (21), (24) in (23), the PDF $q^{[l+1]}(\mathbf{R}_t)$ can also be updated as a IW distribution. Namely

$$\begin{cases} q^{[l+1]}(\mathbf{Q}_{t-1}) = \mathrm{IW}(\mathbf{Q}_{t-1}; \mathbf{U}_t^{[l+1]}, u_t^{[l+1]}) \\ q^{[l+1]}(\mathbf{R}_t) = \mathrm{IW}(\mathbf{R}_t; \mathbf{V}_t^{[l+1]}, v_t^{[l+1]}) \end{cases} \qquad (25)$$

where the updated Dofs and scale matrices for $\mathbf{Q}_{t-1}$ and $\mathbf{R}_t$ are, respectively, computed by

$$\begin{cases} u_t^{[l+1]} = u_0 + 1 \\ \mathbf{U}_t^{[l+1]} = \mathbf{U}_0 + \mathbf{D}_t^{[l]} \end{cases}, \quad \begin{cases} v_t^{[l+1]} = v_0 + 1 \\ \mathbf{V}_t^{[l+1]} = \mathbf{V}_0 + \mathbf{H}_t^{[l]} \end{cases} \quad (26)$$

where the correction matrices $\mathbf{D}_t^{[l]}$ and $\mathbf{H}_t^{[l]}$ are expressed as

$$\begin{cases} \mathbf{D}_t^{[l]} = \mathbf{E}^{[l]} \left[(\mathbf{x}_t - \mathbf{A}_t \mathbf{x}_{t-1})(\mathbf{x}_t - \mathbf{A}_t \mathbf{x}_{t-1})^{\mathrm{T}} \right] \\ \mathbf{H}_t^{[l]} = \mathbf{E}^{[l]} \left[(\mathbf{z}_t - \mathbf{C}_t \mathbf{x}_t)(\mathbf{z}_t - \mathbf{C}_t \mathbf{x}_t)^{\mathrm{T}} \right] \end{cases} \quad (27)$$

Further exploiting (27), the mathematical expectations are calculated by

$$\mathbf{D}_t^{[l]} = \left(\hat{\mathbf{x}}_t^{[l]} - \mathbf{A}_t \hat{\mathbf{x}}_{t-1}^{s,[l]} \right) \left(\hat{\mathbf{x}}_t^{[l]} - \mathbf{A}_t \hat{\mathbf{x}}_{t-1}^{s,[l]} \right)^{\mathrm{T}} + \mathbf{P}_t^{[l]}$$
$$+ \mathbf{A}_t \mathbf{P}_{t-1}^{s,[l]} \mathbf{A}_t^{\mathrm{T}} - \mathbf{A}_t \mathbf{G}_t^{[l]} \mathbf{P}_t^{[l]} - \left(\mathbf{A}_t \mathbf{G}_t^{[l]} \mathbf{P}_t^{[l]} \right)^{\mathrm{T}} \quad (28)$$

$$\mathbf{H}_t^{[l]} = \left(\mathbf{z}_t - \mathbf{C}_t \hat{\mathbf{x}}_t^{[l]} \right) \left(\mathbf{z}_t - \mathbf{C}_t \hat{\mathbf{x}}_t^{[l]} \right)^{\mathrm{T}} + \mathbf{C}_t \mathbf{P}_t^{[l]} \mathbf{C}_t^{\mathrm{T}} \quad (29)$$

where $\mathbf{G}_t^{[l]}$ is the smoothing gain. $\hat{\mathbf{x}}_t^{[l]}$ and $\mathbf{P}_t^{[l]}$ are the filtering estimate and error covariance matrix at time t. $\hat{\mathbf{x}}_{t-1}^{s,[l]}$ and $\mathbf{P}_{t-1}^{s,[l]}$ are the smoothing estimate and error covariance matrix at time $t-1$. Their calculations will be given later.

Exploiting (25)-(27), the required expected sufficient statistics (ESS) of $\mathbf{Q}_{t-1}$ and $\mathbf{R}_t$ are calculated by

$$\hat{\mathbf{Q}}_{t-1}^{[l+1]} = \mathbf{U}_t^{[l+1]}/u_t^{[l+1]}, \quad \hat{\mathbf{R}}_t^{[l+1]} = \mathbf{V}_t^{[l+1]}/v_t^{[l+1]} \quad (30)$$

2) Update of $q^{[l+1]}(\mathbf{x}_t, \mathbf{x}_{t-1})$: Employing the multiplication theorem [19], the PDF $q^{[l+1]}(\mathbf{x}_t, \mathbf{x}_{t-1})$ can be factorized as

$$q^{[l+1]}(\mathbf{x}_t, \mathbf{x}_{t-1}) = q^{[l+1]}(\mathbf{x}_{t-1}|\mathbf{x}_t) q^{[l+1]}(\mathbf{x}_t)$$
$$= \mathrm{N}(\mathbf{x}_{t-1}; \mathbf{g}^{[l+1]}(\mathbf{x}_t), \mathbf{L}_{t-1}^{[l+1]}) \mathrm{N}(\mathbf{x}_t; \hat{\mathbf{x}}_t^{[l+1]}, \mathbf{P}_t^{[l+1]}) \quad (31)$$

where the intermediate variables for $q^{[l+1]}(\mathbf{x}_{t-1}|\mathbf{x}_t)$ are calculated by

$$\begin{cases} \hat{\mathbf{x}}_t^{-,[l+1]} = \mathbf{A}_t \hat{\mathbf{x}}_{t-1}^{s,[l]} \\ \mathbf{P}_t^{-,[l+1]} = \mathbf{A}_t \mathbf{P}_{t-1}^{s,[l]} \mathbf{A}_t^{\mathrm{T}} + \hat{\mathbf{Q}}_{t-1}^{[l+1]} \\ \mathbf{g}^{[l+1]}(\mathbf{x}_t) = \hat{\mathbf{x}}_{t-1}^{s,[l]} + \mathbf{G}_t^{[l+1]} \left(\mathbf{x}_t - \hat{\mathbf{x}}_t^{-,[l+1]} \right) \\ \mathbf{L}_{t-1}^{[l+1]} = \mathbf{P}_{t-1}^{s,[l]} - \mathbf{G}_t^{[l+1]} \mathbf{P}_t^{-,[l+1]} \left(\mathbf{G}_t^{[l+1]} \right)^{\mathrm{T}} \\ \mathbf{G}_t^{[l+1]} = \mathbf{P}_{t-1}^{s,[l]} \mathbf{A}_t^{\mathrm{T}} (\mathbf{P}_t^{-,[l+1]})^{-1} \end{cases} \quad (32)$$

The approximate posterior PDF $q^{[l+1]}(\mathbf{x}_t)$ can be computed through the KF measurement update process:

$$\begin{cases} \hat{\mathbf{x}}_t^{[l+1]} = \hat{\mathbf{x}}_t^{-,[l+1]} + \mathbf{K}_t^{[l+1]} \left(\mathbf{z}_t - \mathbf{C}_t \hat{\mathbf{x}}_t^{-,[l+1]} \right) \\ \mathbf{K}_t^{[l+1]} = \mathbf{P}_t^{-,[l+1]} \mathbf{C}_t^{\mathrm{T}} \left[\mathbf{C}_t \mathbf{P}_t^{-,[l+1]} \mathbf{C}_t^{\mathrm{T}} + \hat{\mathbf{R}}_t^{[l+1]} \right]^{-1} \\ \mathbf{P}_t^{[l+1]} = (\mathbf{I}_p - \mathbf{K}_t^{[l+1]} \mathbf{C}_t) \mathbf{P}_t^{-,[l+1]} \end{cases} \quad (33)$$

Marginalizing $q^{[l+1]}(\mathbf{x}_t, \mathbf{x}_{t-1})$ over the current state $\mathbf{x}_t$ yields $q^{[l+1]}(\mathbf{x}_{t-1}) = \mathrm{N}(\mathbf{x}_{t-1}; \hat{\mathbf{x}}_{t-1}^{s,[l+1]}, \mathbf{P}_{t-1}^{s,[l+1]})$, where the smoothing estimate and error covariance are computed by

$$\begin{cases} \hat{\mathbf{x}}_{t-1}^{s,[l+1]} = \hat{\mathbf{x}}_{t-1}^{s,[l]} + \mathbf{G}_t^{[l+1]} \left(\hat{\mathbf{x}}_t^{[l+1]} - \hat{\mathbf{x}}_t^{-,[l+1]} \right) \\ \mathbf{P}_{t-1}^{s,[l+1]} = \mathbf{P}_{t-1}^{s,[l]} + \mathbf{G}_t^{[l+1]} \left(\mathbf{P}_t^{[l+1]} - \mathbf{P}_t^{-,[l+1]} \right) \left(\mathbf{G}_t^{[l+1]} \right)^{\mathrm{T}} \end{cases} \quad (34)$$

Algorithm 1: One Cycle of the Proposed GTKF

Input: $\hat{\mathbf{x}}_{t-1}$, $\mathbf{P}_{t-1}$, $\mathbf{z}_t$, $\mathbf{Q}_0$, $\mathbf{R}_0$, u_0, v_0, θ, L
Output: $\hat{\mathbf{x}}_t = \hat{\mathbf{x}}_t^{[L]}$, $\mathbf{P}_t = \mathbf{P}_t^{[L]}$
Initialization: $\mathbf{U}_0 = u_0 \mathbf{Q}_0$, $\mathbf{V}_0 = v_0 \mathbf{R}_0$, $\hat{\mathbf{x}}_{t-1}^{s,[0]} = \hat{\mathbf{x}}_{t-1}$,
$\mathbf{P}_{t-1}^{s,[0]} = \mathbf{P}_{t-1}$, $\mathbf{P}_t^{-,[0]} = \mathbf{A}_t \mathbf{P}_{t-1} \mathbf{A}_t^{\mathrm{T}} + \mathbf{Q}_0$
Variational inference:
for $l = 1, 2, \cdots, L$ **do**
    1. Update $q^{[l]}(\mathbf{x}_t)$ by (33);
    2. Update $q^{[l]}(\mathbf{x}_{t-1})$ by (32) and (34);
    3. Calculate $\mathbf{D}_t^{[l]}$ and $\mathbf{H}_t^{[l]}$ by (28) and (29);
    4. Update $q^{[l]}(\mathbf{Q}_{t-1})$, $q^{[l]}(\mathbf{R}_t)$ by (25)-(26) and (30);
    if $\|\hat{\mathbf{x}}_t^{[l]} - \hat{\mathbf{x}}_t^{[l-1]}\|/\|\hat{\mathbf{x}}_{k|k}^{[l]}\| \le \theta$ **then**
      | **break**;
    end
end

One cycle of the proposed algorithm is summarized in Algorithm 1, where θ is the termination threshold and L is the maximal iteration number.

Remark 1. *Substituting (26) in (30) and using the initialization condition in Algorithm 1, the ESS of $\mathbf{Q}_{t-1}$ can be recalculated as $\hat{\mathbf{Q}}_{t-1}^{[l+1]} = \frac{u_0}{u_0+1} \mathbf{Q}_0 + \frac{1}{u_0+1} \mathbf{D}_t^{[l]}$. Obviously, the tuning parameter u_0 balances the weights between the prior nominal SNCM $\mathbf{Q}_0$ and the correction matrix $\mathbf{D}_t^{[l]}$. The larger the tuning parameter u_0 is, the more prior information is introduced, and vice versa. The influence of the tuning parameter v_0 can be analysed by the same way. Detailed guideline for parameter selections see Section V. C.*

IV. VARIANTS AND DISCUSSIONS

A. Low computational complexity GTKF

According to (31)-(32) we know that, the filtering PDF $q(\mathbf{x}_{t-1}|\mathbf{z}_{1:t-1})$ is substituted by the smoothing PDF $q^{[l]}(\mathbf{x}_{t-1}|\mathbf{z}_{1:t})$ in the proposed GTKF. The smoother can provide higher estimation accuracy than filter while more computational resources are needed. One can also obtain a LCC-GTKF by giving up such substitution. This means that the PDF $q^{[l+1]}(\mathbf{x}_t, \mathbf{x}_{t-1})$ in (31) is factorized as

$$q^{[l+1]}(\mathbf{x}_t, \mathbf{x}_{t-1}) = q^{[l+1]}(\mathbf{x}_t|\mathbf{x}_{t-1}) q(\mathbf{x}_{t-1}|\mathbf{z}_{1:t-1})$$
$$= \mathrm{N}(\mathbf{x}_t; \mathbf{A}_t \mathbf{x}_{t-1}, \hat{\mathbf{Q}}_{t-1}^{[l+1]}) \mathrm{N}(\mathbf{x}_{t-1}; \hat{\mathbf{x}}_{t-1}, \mathbf{P}_{t-1}) \quad (35)$$

and the calculation of correction matrix $\mathbf{D}_t^{[l]}$ is simplified as

$$\mathbf{D}_t^{[l]} = \left(\hat{\mathbf{x}}_t^{[l]} - \mathbf{A}_t \hat{\mathbf{x}}_{t-1} \right) \left(\hat{\mathbf{x}}_t^{[l]} - \mathbf{A}_t \hat{\mathbf{x}}_{t-1} \right)^{\mathrm{T}} + \mathbf{P}_t^{[l]} + \mathbf{A}_t \mathbf{P}_{t-1} \mathbf{A}_t^{\mathrm{T}} \quad (36)$$

However, the estimation accuracy of such LCC-GTKF is also decreased. The implementary VB process of the proposed LCC-GTKF is presented in Algorithm 2.

B. Nonlinear GTKF

The NGTKF is easily to extend. Consider a nonlinear state-space model as follows

$$\begin{cases} \mathbf{x}_t = \mathbf{a}_t(\mathbf{x}_{t-1}) + \mathbf{m}_t + \mathbf{v}_{t-1} \\ \mathbf{z}_t = \mathbf{c}_t(\mathbf{x}_t) + \mathbf{n}_t \end{cases} \quad (37)$$

Algorithm 2: VB Process of the Proposed LCC-GTKF

$\hat{\mathbf{x}}_t^- = \mathbf{A}_t \hat{\mathbf{x}}_{t-1}$
for $l = 1, 2, \cdots, L$ **do**

  1. Update $q^{[l]}(\mathbf{x}_t)$:
    $\mathbf{P}_t^{-,[l]} = \mathbf{A}_t \mathbf{P}_{t-1} \mathbf{A}_t^T + \hat{\mathbf{Q}}_{t-1}^{[l-1]}$;
    $\mathbf{K}_t^{[l]} = \mathbf{P}_t^{-,[l]} \mathbf{C}_t^T \left[\mathbf{C}_t \mathbf{P}_t^{-,[l]} \mathbf{C}_t^T + \hat{\mathbf{R}}_t^{[l-1]} \right]^{-1}$;
    $\hat{\mathbf{x}}_t^{[l]} = \hat{\mathbf{x}}_t^- + \mathbf{K}_t^{[l]} \left(\mathbf{z}_t - \mathbf{C}_t \hat{\mathbf{x}}_t^- \right)$;
    $\mathbf{P}_t^{[l]} = (\mathbf{I}_p - \mathbf{K}_t^{[l]} \mathbf{C}_t) \mathbf{P}_t^{-,[l]}$;
  2. Calculate $\mathbf{D}_t^{[l]}$ and $\mathbf{H}_t^{[l]}$:
    $\mathbf{D}_t^{[l]} =$
    $\left(\hat{\mathbf{x}}_t^{[l]} - \mathbf{A}_t \hat{\mathbf{x}}_{t-1} \right) \left(\hat{\mathbf{x}}_t^{[l]} - \mathbf{A}_t \hat{\mathbf{x}}_{t-1} \right)^T + \mathbf{P}_t^{[l]} + \mathbf{A}_t \mathbf{P}_{t-1} \mathbf{A}_t^T$;
    $\mathbf{H}_t^{[l]} = \left(\mathbf{z}_t - \mathbf{C}_t \hat{\mathbf{x}}_t^{[l]} \right) \left(\mathbf{z}_t - \mathbf{C}_t \hat{\mathbf{x}}_t^{[l]} \right)^T + \mathbf{C}_t \mathbf{P}_t^{[l]} \mathbf{C}_t^T$;
  3. Update $q^{[l]}(\mathbf{Q}_{t-1})$, $q^{[l]}(\mathbf{R}_t)$ by (25)-(26) and (30);
  if $\|\hat{\mathbf{x}}_t^{[l]} - \hat{\mathbf{x}}_t^{[l-1]}\| / \|\hat{\mathbf{x}}_{k|k}^{[l]}\| \le \theta$ **then**
    | **break**;
  end

end

---

where $\mathbf{m}_t$ is the known control input. $\mathbf{a}_t(\cdot)$ and $\mathbf{c}_t(\cdot)$ are state transition function and measurement function, respectively. The CL process for AUVs investigated in Section V can be described by this nonlinear model.

To reduce the linearization errors, we propose to linearize $\mathbf{a}_t(\cdot)$ and $\mathbf{c}_t(\cdot)$ at the smoothing estimate $\hat{\mathbf{x}}_{t-1}^{s,[l]}$ and filtering estimate $\hat{\mathbf{x}}_t^{[l]}$, respectively. Since $\hat{\mathbf{x}}_{t-1}^{s,[l]}$ and $\hat{\mathbf{x}}_t^{[l]}$ are, respectively, the nearest approximations of $\mathbf{x}_{t-1}$ and $\mathbf{x}_t$ at the current time. Employing the first-order Taylor expansion, the resultant linear state-space model can be formulated as

$$\begin{cases} \mathbf{x}_t \approx \mathbf{A}_t \mathbf{x}_{t-1} + \mathbf{a}_t(\hat{\mathbf{x}}_{t-1}^{s,[l]}) - \mathbf{A}_t \hat{\mathbf{x}}_{t-1}^{s,[l]} + \mathbf{m}_t + \mathbf{v}_{t-1} \\ \mathbf{z}_t \approx \mathbf{c}_t \mathbf{x}_t + \mathbf{c}_t(\hat{\mathbf{x}}_t^{[l]}) - \mathbf{c}_t \hat{\mathbf{x}}_t^{[l]} + \mathbf{n}_t \end{cases} \quad (38)$$

where for the sake of convenience, we still use $\mathbf{A}_t$ and $\mathbf{C}_t$ to denote the Jacobian matrices of $\mathbf{a}_t(\cdot)$ and $\mathbf{c}_t(\cdot)$, that is,

$$\mathbf{A}_t = \left. \frac{\partial \mathbf{a}_t(\mathbf{x}_{t-1})}{\partial \mathbf{x}_{t-1}} \right|_{\mathbf{x}_{t-1} = \hat{\mathbf{x}}_{t-1}^{s,[l]}}, \quad \mathbf{C}_t = \left. \frac{\partial \mathbf{c}_t(\mathbf{x}_t)}{\partial \mathbf{x}_t} \right|_{\mathbf{x}_t = \hat{\mathbf{x}}_t^{[l]}} \quad (39)$$

Then we define a pair of pseudo state and pseudo measurement as follows

$$\begin{cases} \tilde{\mathbf{x}}_t = \mathbf{x}_t - \mathbf{a}_t(\hat{\mathbf{x}}_{t-1}^{s,[l]}) + \mathbf{A}_t \hat{\mathbf{x}}_{t-1}^{s,[l]} - \mathbf{m}_t \\ \tilde{\mathbf{z}}_t = \mathbf{z}_t - \mathbf{c}_t(\hat{\mathbf{x}}_t^{[l]}) + \mathbf{c}_t \hat{\mathbf{x}}_t^{[l]} \end{cases} \quad (40)$$

Using (40) in (38), the resultant linear state-space model has a similar form with (1). Exploiting (38)-(40), the predicted estimate $\hat{\mathbf{x}}_t^{-,[l+1]}$, posterior estimate $\hat{\mathbf{x}}_t^{[l+1]}$, correction matrices in the proposed GTKF should be recalculated as follows

$$\begin{cases} \hat{\mathbf{x}}_t^{-,[l+1]} = \mathbf{a}_t(\hat{\mathbf{x}}_{t-1}^{s,[l]}) + \mathbf{m}_t \\ \hat{\mathbf{x}}_t^{[l+1]} = \hat{\mathbf{x}}_t^{-,[l+1]} + \mathbf{K}_t^{[l+1]} \left(\tilde{\mathbf{z}}_t - \mathbf{C}_t \hat{\mathbf{x}}_t^{-,[l+1]} \right) \\ \mathbf{D}_t^{[l]} = \mathbf{E}^{[l]} \left[(\tilde{\mathbf{x}}_t - \mathbf{A}_t \mathbf{x}_{t-1})(\tilde{\mathbf{x}}_t - \mathbf{A}_t \mathbf{x}_{t-1})^T \right] \\ \mathbf{H}_t^{[l]} = \mathbf{E}^{[l]} \left[(\tilde{\mathbf{z}}_t - \mathbf{C}_t \mathbf{x}_t)(\tilde{\mathbf{z}}_t - \mathbf{C}_t \mathbf{x}_t)^T \right] \end{cases} \quad (41)$$

Incorporating the above modifications into Algorithm 1, we obtain the implementation process of the proposed GTKF.

TABLE I: Parameter settings for the compared algorithms.

Filters	Parameter settings
HKF [9]	Tuning parameter $\alpha = 1.345$
MCKF [10]	Kernel size $\gamma = 15$
STKF [11]	Tuning parameter $\tau_p = 5$
STKF-Dof [14]	Prior parameters $a_0 = c_0 = 5$, $b_0 = d_0 = 1$, $\tau_p = \tau_r = 5$
GTKF	$u_0 = v_0 = 5$

C. Superiority Analyses

Since the the scale matrices for (18) and (19) are initialized as $\mathbf{U}_0 = u_0 \mathbf{Q}_0$ and $\mathbf{V}_0 = v_0 \mathbf{R}_0$, the densities in (18) and (19) will degrade to $\mathrm{St}\left(\mathbf{x}_t; \mathbf{A}_t \mathbf{x}_{t-1}, \mathbf{Q}_0, u_0\right)$ and $\mathrm{St}\left(\mathbf{z}_t; \mathbf{C}_t \mathbf{x}_t, \mathbf{R}_0, v_0\right)$ respectively when $p = 1$ and $d = 1$, and exhibit heavier tails than the corresponding standard Student's t distributions when $p > 1$ and $d > 1$. This provides advantages for the proposed GTKFs to address the complicated outlier interferences.

Besides, the proposed GTKFs directly model the state transition PDF and measurement likelihood PDF as generalized t distributions by employing the one-step smoothing strategy, which avoids unnecessary approximation errors as in the existing STKFs. Therefore, the proposed GTKFs achieve better estimation accuracy than the existing robust filters.

V. SIMULATIONS AND EXPERIMENTAL VERIFICATION

To validate the superiority, in this section, we first apply the proposed GTKF in a simulation of agile target tracking. Then a univariate growth model is used to validate the effectiveness in nonlinear systems. Finally, an experiment on the CL for AUVs is conducted to further verify the usefulness. The KF with real noise covariances (KFRNC), HKF [20], MCKF [10], STKF [11], and the STKF with adaptive Dof (STKF-Dof) [14] are also applied to sketch a fair comparative picture. The parameter settings for these compared algorithms are listed in TABLE I. The root mean square errors (RMSEs) and averaged RMSEs (ARMSEs) are taken as the performance indices in the simulations [11]. Moreover, the localization error (LE), averaged LE (ALE) and absolute value of biases (AVB) are utilized in the experimental example to evaluate the localization accuracy, which are, respectively, defined as [21]

$$\begin{cases} \mathrm{LE}(t) = \sqrt{(P_{x,t} - \hat{P}_{x,t})^2 + (P_{y,t} - \hat{P}_{y,t})^2} \\ \mathrm{ALE}(t) = \frac{1}{N_s} \sum_{t=1}^{N_s} \mathrm{LE}(t) \\ \mathrm{AVB}(t) = |\, P_{x,t} - \hat{P}_{x,t} \,| + |\, P_{y,t} - \hat{P}_{y,t} \,| \end{cases} \quad (42)$$

where $(P_{x,t}, P_{y,t})$ and $(\hat{P}_{x,t}, \hat{P}_{y,t})$ represent the benchmark position and estimated position of AUV, respectively. N_s represents the number of data samples.

The termination threshold and maximal iteration number are, respectively, set as $\theta = 10^{-16}$ and $L = 50$. All the algorithms are coded by MATLAB with an Intel Core i7-10875H CPU @ 2.30 GHz.

A. Simulation of Agile Target Tracking

In this simulation, we assume a two-dimensional constant velocity (CV) model to describe the dynamic process of tracking an agile target. The state and measurement noises are heavy-tailed because of the target maneuver and glint measurements. The CV model has a similar form with (1) by setting

$$\mathbf{A}_t = \begin{bmatrix} \mathbf{I}_2 & \Delta\mathbf{I}_2 \\ \mathbf{0} & \mathbf{I}_2 \end{bmatrix} \qquad \mathbf{C}_t = \begin{bmatrix} \mathbf{I}_2 & \mathbf{0} \end{bmatrix} \tag{43}$$

where $\Delta = 0.5s$ is the sample period. $\mathbf{x}_k = [p_k^x, p_k^y, v_k^x, v_k^y]$ is the unknown state with components being the positions and velocities along the X and Y coordinates, respectively.

The heavy-tailed state and measurement noises are produced according to the following Gaussian mixture manners [11]

$$p(\mathbf{v}_t) = (1 - \alpha_t^v)\,\mathrm{N}(\mathbf{v}_t; \mathbf{0}, \mathbf{Q}_0) + \alpha_t^v \mathrm{N}(\mathbf{v}_t; \mathbf{0}, \mathrm{U}_1\mathbf{Q}_0) \tag{44}$$

$$p(\mathbf{n}_t) = (1 - \alpha_t^n)\,\mathrm{N}(\mathbf{n}_t; \mathbf{0}, \mathbf{R}_0) + \alpha_t^n \mathrm{N}(\mathbf{n}_t; \mathbf{0}, \mathrm{U}_2\mathbf{R}_0) \tag{45}$$

where the nominal SNCM and MNCM are given by

$$\mathbf{Q}_0 = \begin{bmatrix} \frac{\Delta^3}{3}\mathbf{I}_2 & \frac{\Delta^2}{2}\mathbf{I}_2 \\ \frac{\Delta^2}{2}\mathbf{I}_2 & \Delta\mathbf{I}_2 \end{bmatrix} \qquad \mathbf{R}_0 = \begin{bmatrix} 100\,m^2 & 0\,m^2 \\ 0\,m^2 & 100\,m^2 \end{bmatrix} \tag{46}$$

with the outlier probabilities being set as $\alpha_t^v = \alpha_t^n = 0.1$.

The initial state vector $\hat{\mathbf{x}}_0$ follows a Gaussian distribution with mean vector $\mathbf{x}_0 = [0\,m,\ 0\,m,\ 10\,m/s,\ 10\,m/s]^{\mathrm{T}}$ and error covariance matrix $\mathbf{P}_0 = \mathrm{diag}([100\,m^2,\ 100\,m^2,\ 100\,m^2/s^2,\ 100\,m^2/s^2])$. The total simulation steps are 100, and 1000 Monte Carlo (MC) runs are carried out. To better demonstrate the superiority of the proposed GTKF, we consider three noise scenarios in this example.

Case (i): In this case, we consider the moderately contaminated state and measurement noises with the magnified coefficients being set as $\mathrm{U}_1 = \mathrm{U}_2 = 100$. The RMSEs from the compared filters are sketched in Fig. 2, where "RMSE_{pos}" and "RMSE_{vel}" denote the RMSEs of position and velocity, respectively. As shown in Fig. 2, the existing STKF presents smaller estimation deviations than other existing filters in such moderate case. But the proposed GTKF outperforms the STKF benefiting from the new design. The proposed LCC-GTKF presents decreased estimation accuracy because the smoothing estimate of the state at last time is replaced by the filtering estimate. The STKF-Dof performs poorly due to the biased estimation of the Dof parameters.

Case (ii): Then we consider the case with worse state outlier interferences, where the magnified coefficients are, respectively, set as $\mathrm{U}_1 = 500$, $\mathrm{U}_2 = 100$. The RMSEs are sketched in Fig. 3. It is visible that all the existing robust algorithms including the STKF suffers from significant accuracy divergence in this case. The proposed GTKF and LCC-GTKF exhibit satisfactory performance because the state transition PDF is *directly* modeled as a generalized t distribution in our design. The influence of state outliers can thus be effectively resisted under this framework. To be clear, Fig. 4 shows the ARMSEs by the KFRNC, STKF, the proposed GTKF and LCC-GTKF when $U_1 \in [100, 1000]$ and $\mathrm{U}_2 = 100$. As shown in Fig. 4, the varying intensity of state outliers has only small
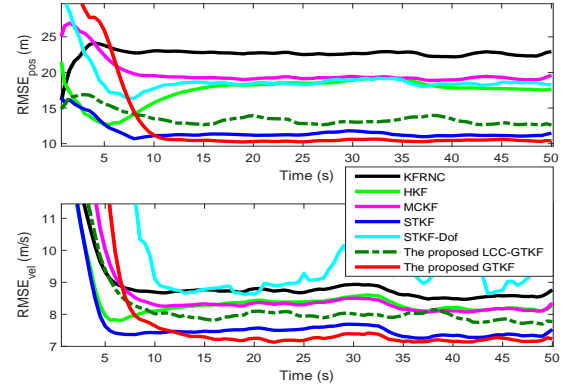

Fig. 2: RMSE comparisons in case (i).

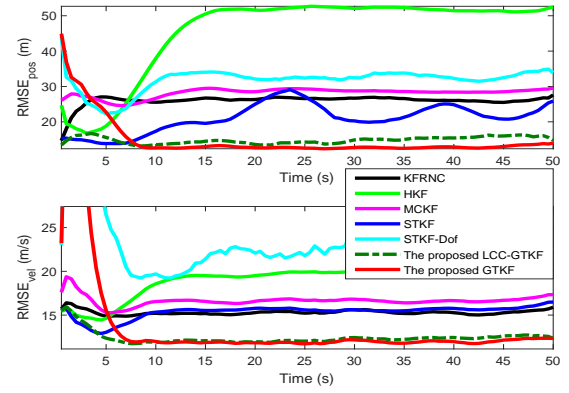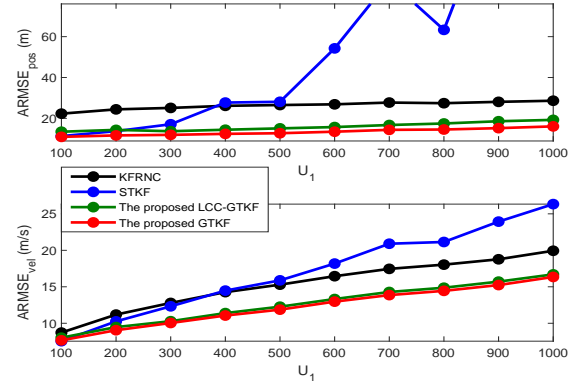Fig. 3: RMSE comparisons in case (ii).

Fig. 4: ARMSE comparisons by the KFRNC, STKF, the proposed GTKF and LCC-GTKF when $U_1 \in [100, 1000]$ and $\mathrm{U}_2 = 100$.

influence on the performance of the proposed GTKF and LCC-GTKF.

Case (iii): Finally, we investigate the case where the state and measurement noises are seriously contaminated, where the magnified coefficients are set as $\mathrm{U}_1 = \mathrm{U}_2 = 1000$. As show in Fig. 5, the STKF cannot converge in such severe scenario. It is interesting that the MCKF can present some robustness against the severe outlier interferences, but it is still
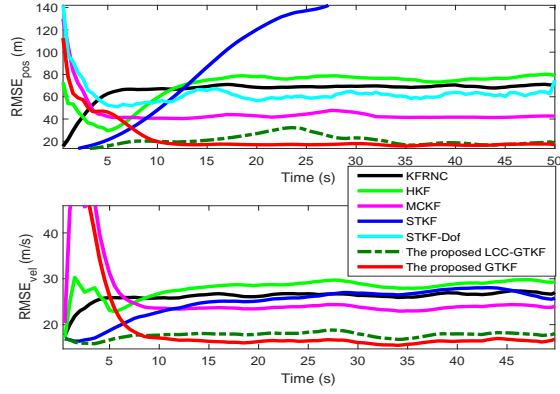
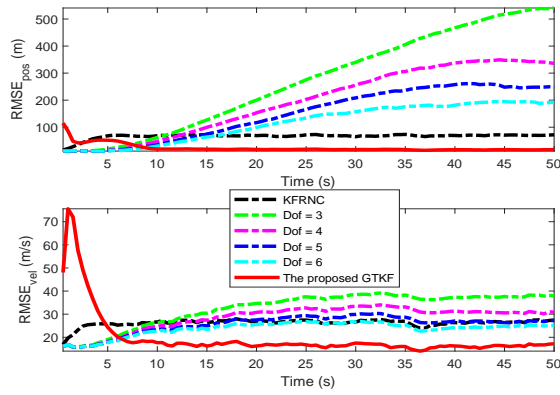This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TAES.2022.3160984, IEEE Transactions on Aerospace and Electronic Systems

7

Fig. 5: RMSE comparisons in case (iii).

Fig. 7: RMSE comparisons from the compared algorithms.

Fig. 6: Performance comparison between the STKF with different Dofs and the proposed GTKF.

inferior in estimation accuracy. The proposed GTKF and LCC-GTKF achieve the best performance because the employed generalized t distributions have heavier tails than the Student's t distribution when dimensions $p > 1$ and $d > 1$, and also benefits from the *directly* modeling manner. In STKF, the values of Dof determine the tail behaviors of a Student's t distribution. A performance comparison between the STKF with different Dofs and the proposed GTKF are depicted in Fig. 6. It is observed that the accuracy degradation of STKF caused by the non-conjugate approximation errors cannot be easily counteracted by changing the value of Dof. The ARMSEs comparisons and computational costs for a single step run are listed in TABLE II, where "Pos." and "Vel." denote the ARMSEs of position and velocity, respectively. The seriously divergent results are omitted here. We can see that the proposed GTKF achieves the best performance in all the three cases. The proposed LCC-GTKF requires quite lower computational cost than the proposed GTKF but slightly worse estimation accuracy is presented.

B. Nonlinear Univariate Growth Model

In this part, the nonlinear univariate growth model is employed to further validate the effectiveness of the proposed GTKF. The NGTKF proposed in Section IV. B is applied in
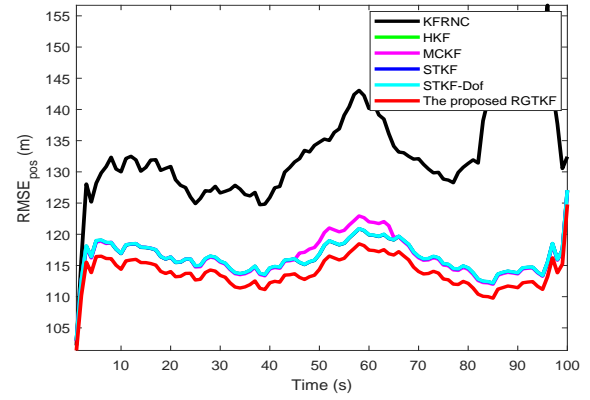
this example, and the first-order Taylor approximation is used for other compared algorithms to address the nonlinearity. The state-space model is given by [22]

$$\begin{cases} x_t = 0.5x_{t-1} + 15\frac{x_{t-1}}{1+x_{t-1}^2} + v_{t-1} \\ z_t = \frac{x_t^2}{20} + n_t \end{cases} \tag{47}$$

where the outliers eroded state and measurement noises v_t and n_t are produced by a similar way as (44)–(45) with the nominal noise variances being set as $Q_0 = R_0 = 100$, and the magnified coefficients being set as $U_1 = U_2 = 1000$. The number of total simulation steps is 100, and 1000 MC runs are executed. A slide window with span of 10 is employed to smooth the RMSE curves.

The RMSEs from the compared algorithms are depicted in Fig. 7. It is visible that the existing robust filters exhibit some robustness against the outlier interferences, but the proposed NGTKF presents the best estimation performance. The existing filters can hardly work well in the situations with strong nonlinearity and severe outlier interferences. The proposed NGTKF outperforms the existing algorithms mainly because of two reasons: 1) The linearization errors are greatly reduced by carefully selecting Taylor expansion points for the nonlinear state transition and measurement functions, as given in (39). 2) The negative impact of outlier interferences is effectively suppressed by directly modeling the state transition and measurement likelihood PDFs as generalized t distributions.

C. Experiment on CL for AUVs

In a master-slave CL framework, the leaders assist the slave AUVs by sharing their accurate position information with them. The leader usually works near the surface and is equipped with a GPS receiver or a high accuracy integrated navigation system to obtain accurate position in real time. Correspondingly, the slave AUV is equipped with a low-cost DR system composed of a DVL, a low-cost compass and a pressure sensor [5], [6]. The DVL is used to perceive the forward velocity $\eta_{f,t}$ and starboard velocity $\eta_{s,t}$ in body frame. The compass is used to perceive the azimuth ψ_t in navigation frame. The depth h_t can be precisely measured by the pressure sensor [5]. Besides, an acoustic modem (AM) is equipped on each AUV to achieve the underwater acoustic communication.

TABLE II: ARMSEs comparisons and computational times for a single step run.

Cases	Filters	KFRNC	HKF	MCKF	STKF	STKF-Dof	LCC-GTKF	GTKF
Case (i)	Pos. (m)	22.80	18.91	19.41	11.46	19.21	13.11	10.50
	Vel. (m/s)	8.97	8.63	8.54	7.69	9.50	8.03	7.37
Case (ii)	Pos. (m)	26.29	51.90	28.69	22.35	32.94	15.23	12.84
	Vel. (m/s)	15.33	19.75	16.67	15.68	22.81	12.36	11.99
Case (iii)	Pos. (m)	69.35	76.77	42.95	-	61.87	19.81	16.97
	Vel. (m/s)	26.63	28.98	23.82	26.99	-	17.89	16.32
Run time (ms)		0.01	0.56	0.41	1.40	2.16	0.85	1.59

1) Principles and Models: Given the position $(P_{x,t-1}, P_{y,t-1})$ at last time and the information perceived by the sensors, the slave AUVs can roughly locate themselves by using the following DR model [14]

$$\begin{cases} P_{x,t} = P_{x,t-1} + T_s(\eta_{f,t}\cos\psi_t + \eta_{s,t}\sin\psi_t) + v_{x,t-1} \\ P_{y,t} = P_{y,t-1} + T_s(\eta_{f,t}\sin\psi_t - \eta_{s,t}\cos\psi_t) + v_{y,t-1} \end{cases}$$
(48)

where the three-dimensional (3D) state $\mathbf{x}_t = [P_{x,t}, P_{y,t}, h_t]^{\mathrm{T}}$ is simplified as a 2D state $\mathbf{x}_t = [P_{x,t}, P_{y,t}]^{\mathrm{T}}$ here because the depth h_t can be precisely measured by the pressure sensor. Besides, T_s represents the sampling period, and $v_{x,t}$ and $v_{y,t}$ are, respectively, the eastward and northward positioning noises.

However, the LE of slave AUV will accumulate unboundedly over time due to the measurement errors [21]. To overcome this problem, the relative distance between the leader AUV and slave AUV is introduced to correct the divergent LE, which is calculated by

$$d_t = \sqrt{(P_{x,t} - P_{x,t}^r)^2 + (P_{y,t} - P_{y,t}^r)^2 + (h_t - h_t^r)^2} \quad (49)$$

where $(P_{x,t}^r, P_{y,t}^r, h_t^r)$ represents the accurate position of leader AUV at time t. Considering that the depth h_t of slave AUV can be precisely measured by the pressure sensor and using (49), the 2D acoustic ranging equation is formulated as follows

$$\begin{aligned} z_t &= \sqrt{d_t^2 - (h_t - h_t^r)^2} + n_t \\ &= \sqrt{(P_{x,t} - P_{x,t}^r)^2 + (P_{y,t} - P_{y,t}^r)^2} + n_t \end{aligned} \quad (50)$$

where z_t denotes the 2D relative distance measurement, which can be determined by the time of arrival (TOA) method, and n_t denotes the measurement noise.

According to the DR model and the acoustic ranging model given in (48) and (50), the discrete-time state-space model for a master-slave CL framework can be summarized as follows

$$\begin{cases} \mathbf{x}_t = \mathbf{A}_t\mathbf{x}_{t-1} + \mathbf{m}_t + \mathbf{v}_{t-1} \\ z_t = \mathbf{g}_t(\mathbf{x}_t) + n_t \end{cases} \quad (51)$$

where the state transition matrix $\mathbf{A}_t = \mathbf{I}_2$; control input $\mathbf{m}_t = [T_s(\eta_{f,t}\cos\psi_t + \eta_{s,t}\sin\psi_t), T_s(\eta_{f,t}\sin\psi_t - \eta_{s,t}\cos\psi_t)]^{\mathrm{T}}$; state noise $\mathbf{v}_t = [v_{x,t}, v_{y,t}]^{\mathrm{T}}$, nonlinear measurement function $\mathbf{g}_t(\mathbf{x}_t) = \sqrt{(P_{x,t} - P_{x,t}^r)^2 + (P_{y,t} - P_{y,t}^r)^2}$.

As described in (48), the state noise $\mathbf{v}_t$ is used to model the measurement errors of DVL and compass. The state outliers are easily to appear when the sensors suffer from measurement anomalies or the lost lock effect of DVL occurs [7], [14]. The acoustic ranging outliers in (50) are mainly induced by

TABLE III: Performance specifications of the equipments.

Sensors	Types	Indices	Parameters
GPS	OEMV-2RT-2	Position precision	1.8 m (RMS)
		Output frequency	10 Hz
AM	ATM-885	Working range	Up to 8000 m
		Error rate	Less than 10^{-7}
DVL	DS-99	Working range	-150~200 m/s
		Detection precision	0.1%~0.3%
Compass		Heading precision	$0.3°$

sound ray bending and multipath effect of underwater acoustic channel [6], [21].

2) Experimental verification: In this experiment, three survey vessels are used to build up a master-slave CL framework, where one vessel acts as the AUV and other two serve as the leaders. The GPS receivers are equipped on the leaders to obtain accurate positions. A low-cost DVL and a self-made compass are equipped on the AUV to locate itself according to (48). The GPS receiver is also equipped on the AUV to provide benchmarks. All vessels communicate with each other by the equipped AM. A sketch map of the underwater communication process is depicted in Fig. 8 (a). As shown, the slave AUV sends a request signal firstly, and then the leader, which received the request signal, sends back a data packet including its accurate position and the arrival time of the request signal. Finally, the slave AUV can calculate the relative distance based on the received information [14]. The experimental configurations are given in Fig. 8 (b), and the real trajectories of all vessels are depicted in Fig. 8 (c). Performance indices of the equipments in Fig. 8 (b) are summarized in TABLE III, where "RMS" denotes root mean square. As shown in Fig. 8 (c), the AUV travels between the two leaders and only communicates with one leader at any moment. About 1742 data samples are collected in the experiment with a sampling frequency at 1 Hz. In practical applications, the real noise covariances are usually unknown. Considering the sensor performance indices and statistical analysis of the off-line data, we set the nominal SNCM and MNCM as $\mathbf{Q}_0 = \mathrm{diag}[(0.5m)^2 \ (0.5m)^2]$ and $\mathbf{R}_0 = (\sqrt{20}m)^2$. The initial state is read from the GPS data and set the initial error covariance as $\mathbf{P}_0 = \mathbf{I}_2$.

Fig. 9 plots the LEs from different algorithms. As shown, the proposed NGTKF achieves the lowest estimation deviations over the compared filters. Large deviations appear in the LEs of extended KF (EKF) since it is specially designed for Gaussian noises. The STKF performs poorly since the state outliers are *indirectly* treated by modeling the predicted PDF
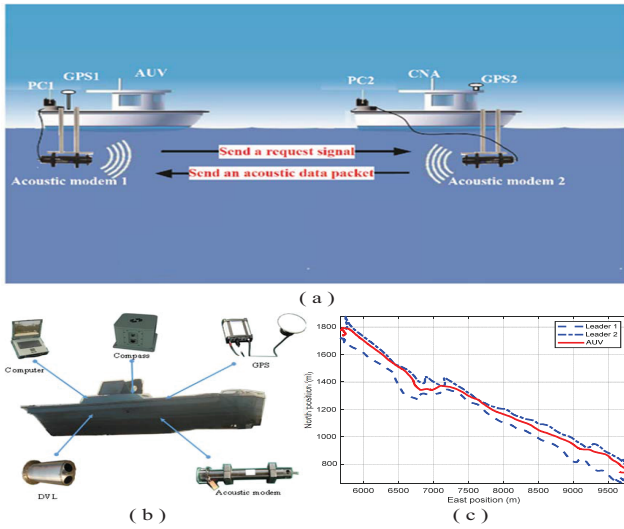
Fig. 8: Experimental configurations. (a) A sketch map of the underwater communication process. (b) The experimental setup of the employed vessel (c) The real trajectories of the AUV and leaders.
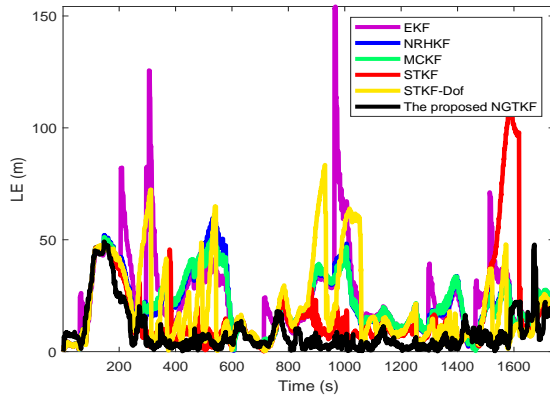

Fig. 9: LE comparisons by the compared algorithms.

TABLE IV: ALEs and computation times for a single step run.

Filters	EKF	NRHKF	MCKF	STKF	STKF-Dof	NGTKF
ALE (m)	24.68	21.80	21.58	16.44	19.67	8.63
Time (ms)	0.02	0.16	0.03	0.75	0.82	0.48

as a heavy-tailed distribution. For a Student's t distribution, its modeling accuracy is very sensitive to the value of Dof parameter. Hence the inaccurate estimation of Dof aggravates the estimation degradation of STKF-Dof dramatically. Although the HKF and MCKF present some robustness in this experiment, their estimation accuracy is still unsatisfactory since these M-estimators neglect the randomness of the state variable.

In Fig. 10, we depict the box plots to compare the AVBs from all filters statistically. In a box plot, the lowest point of the lower whisker and the highest point of the higher whisker denote the minimum and maximum, respectively. The box is drawn from the first quartile to the third quartile with a
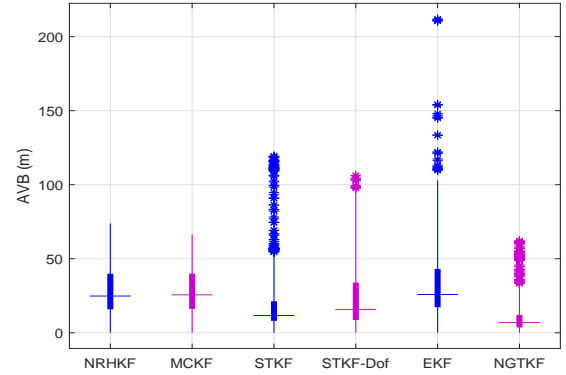

Fig. 10: Box plots of AVBs from different filters.

horizontal line drawn in the middle to denote the median. Asterisks beyond the whiskers denote the samples different significantly from the rest of the dataset. The average level, spread and skewness of each AVB are clearly displayed in Fig. 10. As shown, the proposed NGTKF achieves the smallest maximum and median among all algorithms, which verifies that the NGTKF possesses better accuracy both in extreme and mild environments. The shortest spread indicates that the proposed NGTKF has more stable performance. The positive skewness of the AVBs from the existing algorithms indicates they are easily to provide large deviations in complicated situations. The ALEs and computation times for the applied algorithms in a single step run are compared in TABLE IV. It is seen that compared with the existing nonlinear versions of the STKF and STKF-Dof, the proposed NGTKF has been improved by 47.50% and 56.12% in accuracy but only moderate computational cost is required.

The ALEs along with different iteration numbers are shown in Fig. 11. One can see that the proposed NGTKF tends to converge after only seven iterations. To provide a guideline for parameter selections, the ALE of the NGTKF along with different tuning parameters u_0 and v_0 are also depicted in Fig. 12. It is visible that there is an optimal point for u_0 lie in $(5, 8)$ when $v_0 = 1$. Similarly, there is also an optimal point for v_0 lie in $(0.5, 1.5)$ when $u_0 = 1$. One can select a couple of appropriate parameters for their engineering applications by trial-and-error method. However, it's interesting to see that the fluctuations of ALE is quite small with varying tuning parameters u_0 and v_0. That is, these tuning parameters can be easily determined in applications.

VI. CONCLUSION

This paper proposes a GTKF, where the state transition and measurement likelihood PDFs are modeled as generalized t distributions by resorting to the one-step smoothing strategy. Two variants of the GTKF are also presented to apply to different engineering scenarios. Simulation and experimental examples verify that the proposed GTKFs yield improved robustness over the existing algorithms. Essentially, this paper provides a new framework to address the non-Gaussian noises.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TAES.2022.3160984, IEEE Transactions on Aerospace and Electronic Systems
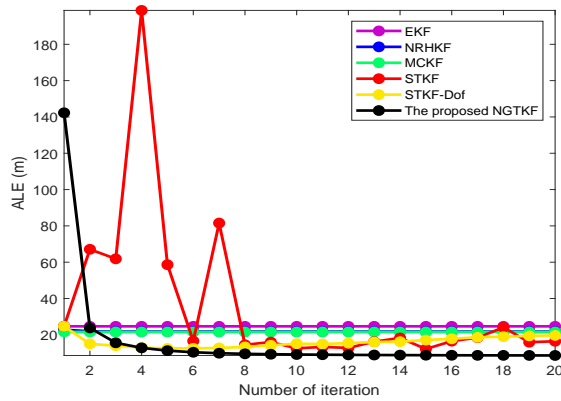
10

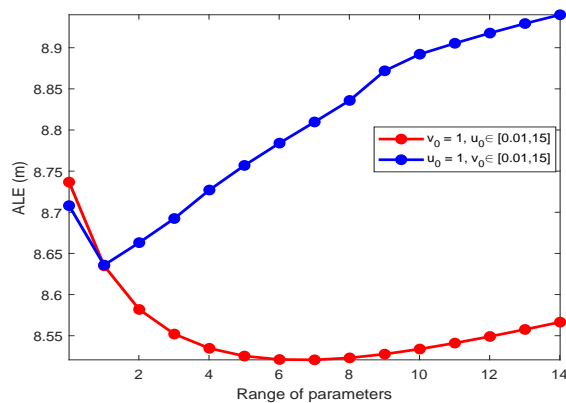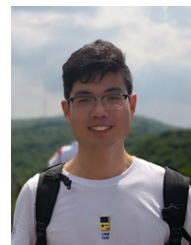Fig. 11: The ALEs with different iteration numbers.

Fig. 12: The ALEs with different tuning parameters.

Other non-Gaussian filters can be easily extended using this framework by selecting appropriate distributions.

References

[1] R. M. Watson, J. N. Gross, C. N. Taylor, and R. C. Leishman, "Robust incremental state estimation through covariance adaptation," *IEEE Robotics and Automation Letters*, vol. 5, no. 2, pp. 3737–3744, 2020.

[2] K. Granstrom, M. Fatemi, and L. Svensson, "Poisson multi-bernoulli mixture conjugate prior for multiple extended target filtering," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 56, no. 1, pp. 208–225, 2020.

[3] M. Bai, Y. Huang, Y. Zhang, and G. Jia, "A novel progressive Gaussian approximate filter for tightly coupled GNSS/INS integration," *IEEE Transactions on Instrumentation and Measurement*, vol. 69, no. 6, pp. 3493–3505, 2020.

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

[5] J. Kim, "Cooperative localization and unknown currents estimation using multiple autonomous underwater vehicles," *IEEE Robotics and Automation Letters*, vol. 5, no. 2, pp. 2365–2371, 2020.

[6] A. Bahr, J. Leonard, M. Fallon, "Cooperative Localization for Autonomous Underwater Vehicles," *The International Journal of Robotics Research*, vol. 28, no. 6, pp. 714-728, May 2009.

[7] J. J. Leonard and A. Bahr, "Autonomous underwater vehicle navigation," in *Springer Handbook of Ocean Engineering*. Springer, 2016, pp. 341–358.

[8] P. J. Huber, *Robust Statistics*. John Wiley & Sons, 2009.

[9] C. D. Karlgaard and H. Schaub, "Huber-based divided difference filtering," *Journal of guidance, control, and dynamics*, vol. 30, no. 3, pp. 885–891, 2007.

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

[11] Y. Huang, Y. Zhang, N. Li, Z. Wu, and J. A. Chambers, "A novel robust Student's t-based Kalman filter," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 53, no. 3, pp. 1545–1554, 2017.

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

[13] R. Piché, S. Särkkä, and J. Hartikainen, "Recursive outlier-robust filtering and smoothing for nonlinear systems using the multivariate Student's t distribution," in *2012 IEEE International Workshop on Machine Learning for Signal Processing*. IEEE, 2012, pp. 1–6.

[14] Y. Huang, Y. Zhang, B. Xu, Z. 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, 2017.

[15] H. Zhu, H. Leung, and Z. He, "A variational bayesian approach to robust sensor fusion based on Student's t distribution," *Information Sciences*, vol. 221, pp. 201–214, 2013.

[16] 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, 2012.

[17] Y. Huang, Y. Zhang, P. Shi, and J. Chambers, "Variational adaptive Kalman filter with Gaussian-inverse-Wishart mixture distribution," *IEEE Transactions on Automatic Control*, vol. 66, no. 4, pp. 1786–1793, 2021.

[18] Y. Huang, Y. Zhang, P. Shi, Z. Wu, J. Qian and J. A. Chambers, "Robust Kalman Filters Based on Gaussian Scale Mixture Distributions With Application to Target Tracking," *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 49, no. 10, pp. 2082-2096, Oct. 2019.

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

[20] C. D. Karlgaard, "Nonlinear regression Huber–Kalman filtering and fixed-interval smoothing," *Journal of guidance, control, and dynamics*, vol. 38, no. 2, pp. 322–330, 2014.

[21] M. Bai, Y. Huang, Y. Zhang, and F. Chen, "A novel heavy-tailed mixture distribution based robust Kalman filter for cooperative localization," *IEEE Transactions on Industrial Informatics*, vol. 17, no. 5, pp. 3671–3681, 2021.

[22] Y. Wu, D. Hu, M. Wu and X. Hu, "A Numerical-Integration Perspective on Gaussian Filters," *IEEE Transactions on Signal Processing*, vol. 54, no. 8, pp. 2910–2921, Aug. 2006.
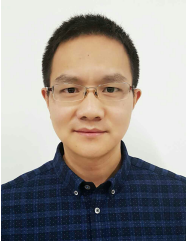
Mingming Bai received the B.S. degree in measurement and control technology and instrumentation from the College of Automation, China University of Geosciences (CUG), Wuhan, China, in 2016. He is currently pursuing the Ph.D. degree in control science and engineering with the Harbin Engineering University (HEU), Harbin, China. He is also a visiting graduate researcher at the Department of Electrical and Computer Engineering in McGill University, Montreal, Canada, since 2021. His current research interests include target tracking, machine learning, statistical signal processing, estimation theory, and multisensor fusion.

Chengjiao Sun received the Ph.D. degree from the College of Automation, Harbin Engineering University, Harbin, China, in 2019. Currently, she is a lecturer in Hubei University of Arts and Science. Her current research interests include state estimation, information fusion and their applications in inertial navigation and cooperative navigation.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TAES.2022.3160984, IEEE Transactions on Aerospace and Electronic Systems

11

Yonggang Zhang (S'06-M'07-SM'16) received the B.S. and M.S. degrees in automation from the College 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.