An Outlier-Robust Kalman Filter With Adaptive Selection of Elliptically Contoured Distributions

Chao Xue, Yulong Huang, *Member, IEEE,* Fengchi Zhu, Yonggang Zhang, *Senior Member, IEEE*
Jonathon Chambers, *Fellow, IEEE*

***Abstract*—In this paper, elliptically contoured (EC) distributions are used to model outlier-contaminated measurement noises. Exploiting a heuristic approach to introduce an unknown parameter, we present an analytical update form of the joint posterior probability density function of the state vector and auxiliary random variable, from which a novel robust EC distributions-based Kalman filtering framework is first derived. To illustrate the effectiveness of the proposed framework, the convergence, robustness, optimality and computational complexity analyses of the proposed method are then given. In addition, to cope with complex noise environments, the interaction multiple model is employed to achieve the adaptive selection of EC distributions such that well-behaved estimation performance can be obtained for different noise cases. Simulation results demonstrate the validity and superiority of the proposed algorithm.**

***Index Terms*—Kalman filter, non-Gaussian heavy-tailed noise, fixed-point iteration, elliptically contoured distribution, interaction multiple model**

I. Introduction

A. Background

THE Kalman filter (KF) has been widely used in many practical situations, such as navigation, localization and target tracking [1]. The KF can provide an optimal estimate in the sense of minimum variance when both the state and measurement noises are Gaussian distributed in the case of a linear state space model. However, the measurement noise often exhibits a heavy-tailed characteristic due to unknown measurement outliers in practical engineering applications [2]. In this case, the Gaussian assumption inherent in the KF is violated, which generally results in poor performance of the KF for such non-Gaussian heavy-tailed measurement noises.

B. Relevant works

To address the state estimation problem of a state-space model with non-Gaussian heavy-tailed measurement noise, many robust KFs have been presented. For example, the Huber-based KF and the Entropy-based KF are two typical robust M-estimate based KFs, where the former is derived by minimizing a mixture of l_1 and l_2 norms that is named as the Huber cost function [3]–[6], and the latter is designed by maximizing a correntropy cost function [7] or minimizing an error entropy cost function [8]. These two M-estimators, however, do not exploit the heavy-tailed features of the noises, which degrades their accuracy. To improve this point, recently, a statistical similarity measure based KF (SSMKF) has been proposed which maximizes the statistical similarity measure to obtain a Gaussian approximation of the non-Gaussian posterior PDF [9].

As a representative heavy-tailed distribution, the Student's t-distribution has been successfully used to design many robust Kalman filters. For example, in the existing outlier-robust KF based on the variational Bayesian approach (ORKF-VB), the measurement noise is modelled as Student's t-distributed, and the state vector and the auxiliary random variable are jointly inferred based on the VB approach [10]. Unfortunately, the existing ORKF-VB exhibits poor estimation performance in the case of large prior uncertainty in the state noise which may be induced by severe manoeuvres of an agile target [11]. Another application of the Student's t-distribution is the existing robust Student's t-based KF (RSTKF) [12] where both the one-step prediction PDF and the measurement noise PDF are modelled as Student's t-distributed, and the approximate joint inference solution is obtained by using the VB approach or the KLD minimization method [12]–[14]. Subsequent study has extended the RSTKF to skewed noises [15]. Recently, an interesting work about a pairwise Markov model (PMM) based Student's t filter has been developed where the independence assumptions are relaxed by using a PMM instead of a Hidden Markov model [16]. Nevertheless, the problems inherent in the existing Student's t filter are still inevitable. In the existing Student's t filter, the prior state PDF and the likelihood PDF of the Student's t filter share the same degree of freedom (dof) parameter, which indicates that the prior state PDF and likelihood PDF require to be contaminated by the same degree of outliers and thereby could be well captured by the Student's t distributions with the same dof parameter [12]. This condition may seldom be satisfied in practice and thus this problem is inevitable in the existing Student's t filter. Except for the Student's t-distribution, the Laplace distribution has also been used for modeling non-Gaussian heavy-tailed noises in the case of uncertain parameters [17]–[19]. However, lack of tuning parameters for the tails of the Laplace distribution

This work was supported by the National Natural Science Foundation of China under Grant Nos. 61903097 and 62173105, and the Fundamental Research Funds for the Central Universities under Grant No. 3072021CFT0401. (Chao Xue, Yulong Huang and Fengchi Zhu are co-first authors.)(Corresponding author: Yulong Huang.)

C. Xue, Y. Huang, F. Zhu and Y. Zhang are with the College of Intelligent Systems Science and Engineering, Harbin Engineering University, Harbin 150001, China, and also with the Engineering Research Center of navigation instruments, Ministry of Education (e-mail: xc_adjuster@163.com; heuedu@163.com; zfchiggins@163.com; zhangyg@hrbeu.edu.cn).

J. Chambers is with the College of Intelligent Systems Science and Engineering, Harbin Engineering University, and also with the School of Engineering, University of Leicester, Leicester, UK (e-mail: Jonathon.Chambers@le.ac.uk).

results in limited estimation performance in non-stationary noise environments.

For non-stationary environments with outliers, some mixture densities based works have been put forward. For example, [20] has proposed a mixture of two Gaussian distributions based robust KF, where a nominal covariance with a high probability and an adaptive larger covariance with a low probability are simultaneously used to accommodate unknown non-stationary heavy-tailed noises. However, the use of the Bernoulli distribution for the selection of distributions means that the algorithm can only switch between two distributions and cannot accommodate more complex cases. In [21], a novel robust KF has been proposed with the state and measurement noises modeled as Gaussian-Gamma mixture distributed, but it needs to learn too many unknown parameters from limited measurement data.

Multiple model methods are also effective in the case of non-Gaussian noises. The Gaussian sum filter can address the non-Gaussian filtering problem by modeling the heavy-tailed noises as a set of weighted Gaussian distributions [23]. However, exponential growing model order leads to its impracticality. Based on this, the switching KF [24] and the Gaussian mixture KF [25] utilize the expectation-maximization method to solve the exponential model order growth problem. The interaction multiple model (IMM) runs parallel sub-filters and mixes the results initially to achieve model interaction, which leads to a computationally efficient solution [26]. The above methods are all based on multiple Gaussian distributions. In fact, it is difficult for the Gaussian mixture distributions with limited number to model the time-varying non-Gaussian heavy-tailed noises. Therefore, [27] has applied the IMM on the dof parameter selections of the Student's t filter and derived a Student's t IMM filter. However, fusing multiple Student's t distributions into one Student's t is required to obtain a closed-form solution, which is a severe approximation and may cause accuracy degradation [28]. Hence, [28] has employed the maximum Versoria criterion (MVC) to mitigate this problem. Nonetheless, the Student's t IMM filter will inherit the defects of the Student's t filter which requires the state and measurement outliers to occur concurrently and have the same magnitude, which limits their performance.

C. Motivations for the paper

As a family of generalized distribution, the EC distribution includes Gaussian, Student's t and many other useful distributions as its special cases. Additionally, due to their flexible density generator functions, the EC distributions can theoretically present arbitrary heavy-tailed features if suitable density generator functions are selected, and thus are widely used in robust statistics. Hence, the EC distribution is superior to the Student's t distribution in the case of time-varying and intricate noises. Thus, we aim to utilize the EC distribution to derive a generalized outlier-robust filtering framework.

Nevertheless, only deriving a closed-form solution of the framework as in [15] cannot deal with the time-varying and intricate environment in a real sense, because the sub-filter still employs fixed dof parameters. Although the VB-based methods have speciality for dealing with this problem by three typical steps, i.e., constructing auxiliary parameters, seeking conjugate distributions for the auxiliary parameters and jointly inferring these parameters by VB methods based on the mean-field hypothesis, this process may encounter the following problems

- VB-based methods are required to construct more parameters to deal with more complex scenarios. In this case, they need to seek more appropriate conjugate distributions to promote the joint inference process. However, finding appropriate conjugate distributions might be difficult, and the derivations would be complicated.
- In VB-based methods, all learned parameters are updated together in the same iteration, which may make the parameter updating process unstable. Once there is a deviation in the learning of one specific parameter, it will affect the learning process of other parameters, resulting in the overall performance deterioration of the algorithm.

Motivated by the above considerations, we firstly develop an EC-based robust Kalman filtering framework based on a heuristic idea without using the VB method. Then the IMM approach[1] is exploited to achieve an adaptive selection of different EC distributions to accommodate time-varying and intricate noise environments.

D. Contributions and organization of paper

The proposed robust Kalman filtering framework with the measurement noise modeled as EC distributed (RECKF) provides a new fixed prior covariance-based iterative method, which turns out to be more accurate in the case of only measurement outliers and computationally efficient, as will be detailed in the later simulation study. Furthermore, the proposed IMM-based RECKF is quite generalized, since it can be used for switching not only different types of EC distributions, but also different parameters of a fixed EC distribution, such as a Student's t-distribution with different dof parameters. This paper is an extension of our previous work [29]. In contrast with [29], the contributions of this paper are summarized as follows.

- The analyses of the convergence, optimality, robustness and computational complexity of the proposed RECKF are specified to illustrate the effectiveness.
- An adaptive switching method of EC distributions is presented by exploiting the IMM approach, such that the well-behaved estimation performance can be obtained for different noise cases.
- A series of simulation comparisons under time-varying and complex environments are carried out to verify the validity and superiority of the proposed algorithm.

The remainder of this paper is structured in the following manner. The problem statement is presented in Section II. The proposed outlier-robust KF is developed in Section III. The

[1]In contrast with learning all switching parameters to achieve mode switching in VB-based methods, the IMM is more stable because each sub-filter is executed independently and the model fusion is only performed in the initial mixing procedure. Thus, the IMM is expected to have better performance in complex scenarios where the parameter learning process could be susceptible.

Acronyms and Nomenclatures

Notations	Definitions
$\mathrm{EC}(\cdot;\boldsymbol{\mu},\boldsymbol{\Sigma},g(\cdot))$	Multivariate EC PDF with mean vector $\boldsymbol{\mu}$, shape matrix $\boldsymbol{\Sigma}$ and density generator function $g(\cdot)$
$\mathrm{N}(\cdot;\boldsymbol{\mu},\mathbf{P})$	Multivariate Gaussian PDF with mean vector $\boldsymbol{\mu}$ and covariance matrix $\mathbf{P}$
$(\cdot)^{-1}$, $\lvert\cdot\rvert$	Inverse operation and determinant operation
$\mathrm{E}\{\cdot\}$	Mathematical expectation operation
$\mathrm{tr}\{\cdot\}$, $(\cdot)^T$	Trace operation and transpose operation
$\lVert\cdot\rVert_F$	Frobenius norm operation
$\mathbf{I}_n, \mathbf{0}_n$	Identity matrix and null matrix of dimension $n\times n$

theoretical analyses of the new filtering algorithm are provided in Section IV. The IMM-based RECKF is derived in Section V. Simulation results and comparisons are conducted in Section VI. This paper is concluded in Section VII.

II. Problem Statement

A. Linear discrete-time state space model

The state-space model (SSM) for a linear discrete-time system is shown as follows

$$\begin{cases}\mathbf{x}_k = \mathbf{F}_k\mathbf{x}_{k-1}+\mathbf{w}_k\\ \mathbf{z}_k = \mathbf{H}_k\mathbf{x}_k+\mathbf{v}_k\end{cases} \tag{1}$$

where $\mathbf{x}_k\in\mathbb{R}^n$ denotes the state vector, and $\mathbf{z}_k\in\mathbb{R}^m$ represents the measurement vector, and $\mathbf{F}_k\in\mathbb{R}^{n\times n}$ and $\mathbf{H}_k\in\mathbb{R}^{m\times n}$ are, respectively, the state transition and measurement matrices, and $\mathbf{w}_k$ and $\mathbf{v}_k$ are the state and measurement noise vectors, respectively. In this paper, the state noise is assumed to be zero mean Gaussian distributed with covariance matrix $\mathbf{Q}_k$, i.e., $\mathbf{w}_k\sim\mathrm{N}(\mathbf{w}_k;\mathbf{0},\mathbf{Q}_k)$, but the measurement noise is assumed to be non-Gaussian heavy-tailed distributed, which is normally caused by modeling errors or outliers. To model such noise, the EC distribution will be introduced in detail in the next subsection.

B. EC distribution

The PDF of an EC distributed random variable $\mathbf{x}$ is written as [22]–[31]

$$\begin{aligned}p_{EC}(\mathbf{x}) &= \mathrm{EC}(\mathbf{x};\boldsymbol{\mu},\boldsymbol{\Sigma},g(\cdot))\\ &= c_g|\boldsymbol{\Sigma}|^{-0.5}\exp\left\{g((\mathbf{x}-\boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu}))\right\}\end{aligned} \tag{2}$$

where μ and $\boldsymbol{\Sigma}$ denote the mean vector and shape matrix, respectively, and $g(\cdot)$ denotes a density generator function that determines the form of the EC distribution. For instance, if $g(t)=-0.5\frac{\nu+p}{\nu+t}$, the EC distribution becomes the Student's t distribution, where p denotes the dimension of the random vector $\mathbf{x}$, and ν denotes the dof parameter of the Student's t distribution. As shown in (2), the EC PDF relies on the quadratic form $(\mathbf{x}-\mu)^T\boldsymbol{\Sigma}^{-1}(\mathbf{x}-\mu)$ completely. Therefore, if $\boldsymbol{\Sigma}$ is a positive definite matrix, the region with constant PDF will form an elliptical area in volume, which explains why the distribution with the PDF in (2) is called an EC distribution. To better model the measurement noise contaminated by outliers, the selection of the density generator function $g(t)$ should fulfil several conditions as follows

$\ell_1: g(t)$ is the third order differentiable on $[0,+\infty)$.
$\ell_2: g(t)$ is monotonically decreasing on $[0,+\infty)$.
$\ell_3: \dot{g}(t)$ is monotonically increasing on $[0,+\infty)$.

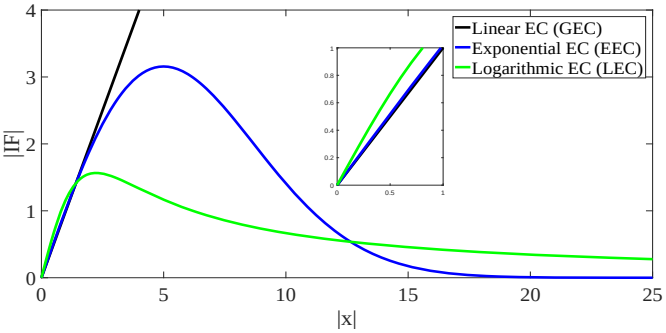


Fig. 1: Influence curves of different EC distributions.

TABLE I: Particular EC distributions and parameter settings.

EC distributions	$g(t)$	Parameter settings
GEC (Gaussian)	$-0.5t$	\
LEC (Student's t)	$-0.5(\nu+p)\log(1+\frac{t}{\nu})$	$p=2,\nu=5$
EEC	$\sigma^2\exp(\frac{p-t}{2\sigma^2})$	$p=2,\sigma=5$

$\ell_4: \ddot{g}(t)$ is monotonically decreasing on $[0,+\infty)$.
$\ell_5: \dot{g}(p)=-0.5$.
$\ell_6: \lim_{t\to+\infty} g(t)/t > -0.5$.

Based on these conditions, Fig. 1 and Table I give three typical EC distributions and their parameter settings, where $|\mathrm{IF}|$ denotes the value of the influence function, and $|\mathbf{x}|$ denotes the norm of the random vector $\mathbf{x}$[2].

Remark 1. *It can be seen from Fig. 1 that the influence curve of the EEC distribution is very close to the GEC distribution when* $\mathbf{x}$ *is small, indicating that the EEC distribution might behave similarly to the Gaussian distribution when there are no outliers. In addition, it can be also observed that the influence curve of the EEC distribution goes down quickly, indicating that it has the strongest ability to suppress large outliers. Similarly, we can basically conclude that the LEC distribution performs well in the case of small outliers while poorly suppressing large outliers relative to the EEC distribution.*

Among these conditions, ℓ_2 guarantees the PDF $p(\mathbf{x})$ won't increase as $\|\mathbf{x}\|_2$ increases which is consistent with a real situation, i.e., $p(\mathbf{x})$ should be smaller when $\mathbf{x}$ is further away from its mean; ℓ_6 guarantees that the selected EC distributions have heavier tails than Gaussian. ℓ_3,ℓ_4,ℓ_5 are required in subsequent robust Kalman filter design. Next, the EC distributions will be used to model the heavy-tailed measurement noise, and the problems encountered in the filter design will be detailed.

C. Problem of filter design based on the EC distribution

The likelihood PDF and the prior PDF of the state vector are modeled as Gaussian and EC distributed, respectively, i.e.,

$$p(\mathbf{z}_k|\mathbf{x}_k) = \mathrm{EC}(\mathbf{z}_k;\mathbf{H}_k\mathbf{x}_k,\mathbf{R}_k,g(\cdot)) \tag{3}$$

$$p(\mathbf{x}_k|\mathbf{z}_{1:k-1}) = \mathrm{N}(\mathbf{x}_k;\hat{\mathbf{x}}_{k|k-1},\mathbf{P}_{k|k-1}) \tag{4}$$

[2]Note that $\mu=0,\boldsymbol{\Sigma}=\mathbf{I}$ for succinctness here.

where the prior mean vector $\hat{\mathbf{x}}_{k|k-1}$ and the prior error covariance matrix $\mathbf{P}_{k|k-1}$ are the same as the time-update results of a traditional KF.

According to Bayes' rule and using (3)-(4), the posterior PDF $p(\mathbf{x}_k|\mathbf{z}_{1:k})$ is calculated as

$$p(\mathbf{x}_k|\mathbf{z}_{1:k}) \propto \mathrm{EC}(\mathbf{z}_k; \mathbf{H}_k\mathbf{x}_k, \mathbf{R}_k, g(\cdot))\mathrm{N}(\mathbf{x}_k; \hat{\mathbf{x}}_{k|k-1}, \mathbf{P}_{k|k-1}) \tag{5}$$

Unfortunately, an analytical and closed-form solution for the posterior PDF of the state vector is not available because the general EC distribution and Gaussian distribution are not conjugate. To address this problem, a heuristic idea will be proposed to achieve an approximate solution of the posterior PDF in the next section, from which a robust Kalman filter based on the EC distribution (RECKF) will be derived.

III. A robust Kalman filtering framework based on EC distributions

To address the above problem, in this section, the likelihood PDF in (3) is formulated as a hierarchical Gaussian form first and an auxiliary random variable is introduced. Then, the state vector and auxiliary random variable are inferred jointly based on an equivalent transformation for the prior PDF of the state vector. Afterwards, an approximate method is proposed to obtain the posterior PDFs of the state vector and auxiliary random variable. Finally, the proposed RECKF is derived.

A. A hierarchical Gaussian form for the likelihood PDF

Inspired by the idea of the ORKF-VB [10] that the Student's t-distribution can be formulated as a Gaussian-Gamma hierarchical form and then the joint posterior PDF can be approximately calculated using the VB approach, an intuitive idea is to represent an EC distribution as a Gaussian scale mixture form which will be shown next. Before presenting our idea, an important Lemma is given first.

Lemma 1. *If a random vector* $\mathbf{y}$ *obeys an EC distribution, i.e.,* $p(\mathbf{y}) = \mathrm{EC}(\mathbf{y}; \mu, \mathbf{\Sigma}, g(\cdot))$*, then* $p(\mathbf{y})$ *can be expressed as a Gaussian scale mixture form as follows [32]*

$$p(\mathbf{y}) = \mathrm{EC}(\mathbf{y}; \boldsymbol{\mu}, \mathbf{\Sigma}, g(\cdot)) = \int \mathrm{N}(\mathbf{y}; \boldsymbol{\mu}, \mathbf{\Sigma}/\lambda) P_g(\lambda) d\lambda \tag{6}$$

where λ *and* $P_g(\lambda)$ *are, respectively, an auxiliary random variable and an unknown PDF of* λ *relying on* $g(\cdot)$*.*

Employing **Lemma 1**, the likelihood PDF is formulated as the following hierarchical Gaussian form

$$p(\mathbf{z}_k|\mathbf{x}_k, \lambda_k) = \mathrm{N}(\mathbf{z}_k; \mathbf{H}_k\mathbf{x}_k, \mathbf{R}_k/\lambda_k), \;\; p(\lambda_k) = P_g(\lambda_k) \tag{7}$$

where $p(\mathbf{z}_k|\mathbf{x}_k, \lambda_k)$ denotes the conditional likelihood PDF and will be used in the following analytical joint update.

B. Analytical joint update

According to Bayes' rule and using (4) and (7), the joint posterior PDF $p(\mathbf{x}_k, \lambda_k|\mathbf{z}_{1:k})$ can be written as

$$p(\mathbf{x}_k, \lambda_k|\mathbf{z}_{1:k}) \propto p(\mathbf{x}_k, \mathbf{z}_k|\mathbf{z}_{1:k-1}, \lambda_k) p(\lambda_k) \tag{8}$$

To obtain the posterior PDF of the state vector, we need to marginalize λ_k on the joint posterior PDF in (8), i.e.,

$$p(\mathbf{x}_k|\mathbf{z}_{1:k}) \propto \int \mathrm{N}(\begin{bmatrix}\mathbf{x}_k \\ \mathbf{z}_k\end{bmatrix}; \begin{bmatrix}\hat{\mathbf{x}}_{k|k-1} \\ \mathbf{H}_k\hat{\mathbf{x}}_{k|k-1}\end{bmatrix}, \overline{\mathbf{P}}_{k|k-1}) P_g(\lambda_k) d\lambda_k \tag{9}$$

where $\overline{\mathbf{P}}_{k|k-1} = \begin{bmatrix}\mathbf{P}_{k|k-1} & \mathbf{P}^{\mathbf{xz}}_{k|k-1} \\ (\mathbf{P}^{\mathbf{xz}}_{k|k-1})^T & \mathbf{P}^{\mathbf{zz}}_{k|k-1}\end{bmatrix}$ denotes the joint prior covariance matrix, and $\mathbf{P}^{\mathbf{xz}}_{k|k-1} = \mathbf{P}_{k|k-1}\mathbf{H}_k^T$ and $\mathbf{P}^{\mathbf{zz}}_{k|k-1} = \mathbf{H}_k\mathbf{P}_{k|k-1}\mathbf{H}_k^T + \mathbf{R}_k/\lambda_k$ denote the cross-covariance matrix and innovation covariance matrix, respectively.

Remark 2. *Although the integral in (9) is similar to that in (6), they are essentially different because the shape matrix* $\mathbf{\Sigma}$ *is entirely scaled by the auxiliary random variable* λ*, while* $\overline{\mathbf{P}}_{k|k-1}$ *is only partially scaled, i.e.,* $\mathbf{R}_k$ *in* $\mathbf{P}^{\mathbf{zz}}_{k|k-1}$*. Thus, to make the integral (9) analytically solvable, an equivalent transformation for (4) will be introduced next.*

Motivated by the fact that if the prior covariance $\mathbf{P}_{k|k-1}$ in (9) is scaled by λ_k, then the joint covariance matrix $\overline{\mathbf{P}}_{k|k-1}$ will be also scaled by λ_k, which makes the integral in (9) able to be analytically calculated by employing (6). To this end, we introduce an equivalent transformation for the prior PDF in (4) as follows

$$\begin{cases} p_{\theta_k}(\mathbf{x}_k|\mathbf{z}_{1:k-1}, \lambda_k) \approx \mathrm{N}(\mathbf{x}_k; \hat{\mathbf{x}}_{k|k-1}, \frac{\theta_k}{\lambda_k}\mathbf{P}_{k|k-1}) \\ s.t. \qquad \theta_k = \mathrm{E}\{\lambda_k\} \end{cases} \tag{10}$$

where $p_{\theta_k}(\cdot)$ represents the PDF depending on the parameter θ_k. Note that the prior PDF in (10) is approximately identical to that in (4). The introduction of the undetermined parameter θ_k is only to simplify the derivations.

Employing (3) and (10), (8) is reformulated as

$$p_{\theta_k}(\mathbf{x}_k, \lambda_k|\mathbf{z}_{1:k}) \propto \mathrm{N}(\begin{bmatrix}\mathbf{x}_k \\ \mathbf{z}_k\end{bmatrix}; \begin{bmatrix}\hat{\mathbf{x}}_{k|k-1} \\ \mathbf{H}_k\hat{\mathbf{x}}_{k|k-1}\end{bmatrix}, \frac{\breve{\mathbf{P}}_{k|k-1}}{\lambda_k}) P_g(\lambda_k) \tag{11}$$

where $\breve{\mathbf{P}}_{k|k-1} = \begin{bmatrix}\theta_k\mathbf{P}_{k|k-1} & \theta_k\mathbf{P}^{\mathbf{xz}}_{k|k-1} \\ \theta_k(\mathbf{P}^{\mathbf{xz}}_{k|k-1})^T & \breve{\mathbf{P}}^{\mathbf{zz}}_{k|k-1}\end{bmatrix}$ and $\breve{\mathbf{P}}^{\mathbf{zz}}_{k|k-1} = \mathbf{H}_k\theta_k\mathbf{P}_{k|k-1}\mathbf{H}_k^T + \mathbf{R}_k$ denote the modified joint prior covariance matrix and the modified innovation covariance matrix, respectively, and the measurement likelihood can be calculated as follows

$$p_{\theta_k}(\mathbf{z}_k|\mathbf{z}_{1:k-1}) = \int\int p_{\theta_k}(\mathbf{x}_k, \mathbf{z}_k, \lambda_k|\mathbf{z}_{1:k-1}) d\mathbf{x}_k d\lambda_k \tag{12}$$

Theorem 1. *The joint posterior PDF can be analytically calculated as follows*

$$\begin{aligned} &p_{\theta_k}(\mathbf{x}_k, \lambda_k|\mathbf{z}_{1:k}) = p_{\theta_k}(\mathbf{x}_k|\mathbf{z}_{1:k}, \lambda_k) p(\lambda_k|\mathbf{z}_{1:k}) \\ &= \mathrm{N}(\mathbf{x}_k; \hat{\mathbf{x}}'_{k|k}(\theta_k), \mathbf{P}'_{k|k}(\theta_k)/\lambda_k) p(\lambda_k|\mathbf{z}_{1:k}) \end{aligned} \tag{13}$$

where the mean vector $\hat{\mathbf{x}}'_{k|k}(\theta_k)$ *and the covariance matrix* $\mathbf{P}'_{k|k}(\theta_k)$ *are, respectively, calculated as*

$$\begin{cases} \hat{\mathbf{x}}'_{k|k}(\theta_k) = \hat{\mathbf{x}}_{k|k-1} + \mathbf{K}_k(\theta_k)(\mathbf{z}_k - \hat{\mathbf{z}}_{k|k-1}) \\ \mathbf{P}'_{k|k}(\theta_k) = (\mathbf{I}_n - \mathbf{K}_k(\theta_k)\mathbf{H}_k)\theta_k\mathbf{P}_{k|k-1} \\ \hat{\mathbf{z}}_{k|k-1} = \mathbf{H}_k\hat{\mathbf{x}}_{k|k-1} \\ \mathbf{P}^{\mathbf{zz}}_{k|k-1}(\theta_k) = \mathbf{H}_k\theta_k\mathbf{P}_{k|k-1}\mathbf{H}_k^T + \mathbf{R}_k \\ \mathbf{K}_k(\theta_k) = \theta_k\mathbf{P}_{k|k-1}\mathbf{H}_k^T(\mathbf{P}^{\mathbf{zz}}_{k|k-1}(\theta_k))^{-1} \end{cases} \tag{14}$$

where $\mathbf{K}_k(\theta_k)$ *denotes the Kalman gain, and* $\hat{\mathbf{z}}_{k|k-1}$ *denotes*

the predicted measurement vector, and $\mathbf{P}^{\mathbf{zz}}_{k|k-1}(\theta_k)$ *denotes the modified innovation covariance matrix, and* $p_{\theta_k}(\mathbf{x}_k|\mathbf{z}_{1:k},\lambda_k)$ *is defined as the modified posterior PDF of the state vector, and* $p(\lambda_k|\mathbf{z}_{1:k})$ *denotes the posterior PDF of* λ_k.

Proof. See Appendix A. □

Remark 3. *Note that* $\hat{\mathbf{x}}'_{k|k}(\theta_k)$ *and* $\mathbf{P}'_{k|k}(\theta_k)/\lambda_k$ *in (14) are not the state estimate and covariance we need exactly due to* λ_k *and* θ_k *in* $p_{\theta_k}(\mathbf{x}_k|\mathbf{z}_{1:k},\lambda_k)$. *Next, the marginal posterior PDF of the state vector will be derived by using* ***Theorem 1***.

C. An approximate solution for the posterior PDF

In this part, the analytical solution for the posterior PDF of the state vector is derived. Before that, we first provide the marginal posterior PDF of the auxiliary random variable.

Proposition 1. *The marginal posterior PDF of the auxiliary random variable* $p(\lambda_k|\mathbf{z}_{1:k})$ *in (13) can be formulated as follows*

$$p(\lambda_k|\mathbf{z}_{1:k}) = \frac{\mathrm{N}(\mathbf{z}_k;\hat{\mathbf{z}}_{k|k-1},\mathbf{P}^{\mathbf{zz}}_{k|k-1}(\theta_k)/\lambda_k)P_g(\lambda_k)}{p_{\theta_k}(\mathbf{z}_k|\mathbf{z}_{1:k-1})} \tag{15}$$

Proof. See Appendix B. □

Equation (15) shows that the posterior PDF of the auxiliary random variable λ_k is uncorrelated with the posterior state PDF, which means that it can be updated only using the measurement data without the participation of the posterior state PDF. We will show this operation can improve the stability of posterior updates but at the cost of a little accuracy degradation in Section IV-A. Employing (15), the posterior state PDF can be calculated by integrating λ_k from (13) as

$$p_{\theta_k}(\mathbf{x}_k|\mathbf{z}_{1:k}) = \int \mathrm{N}(\mathbf{x}_k;\hat{\mathbf{x}}'_{k|k}(\theta_k),\frac{\mathbf{P}'_{k|k}(\theta_k)}{\lambda_k})p_{\theta_k}(\lambda_k|\mathbf{z}_{1:k})d\lambda_k \tag{16}$$

Following (10), $p(\lambda_k|\mathbf{z}_{1:k})$ is approximated as a Dirac delta function at the posterior mean of λ_k, i.e.,

$$p(\lambda_k|\mathbf{z}_{1:k}) \approx \delta(\lambda_k - \mathrm{E}\{\lambda_k|\mathbf{z}_{1:k}\}) \tag{17}$$

where

$$\mathrm{E}(\lambda_k|\mathbf{z}_{1:k}) = \int_0^{+\infty} \lambda_k p(\lambda_k|\mathbf{z}_{1:k})d\lambda_k \tag{18}$$

Remark 4. *(17) is a crucial approximation in this paper, which makes it possible for us to deal with the unknown density associated with* λ_k. *It is seen from (16) that although the update form of the posterior PDF of* λ_k *has been obtained, the posterior PDFs of* λ_k *and* $\mathbf{x}_k$ *are still unavailable due to the unknown PDF* $P_g(\lambda_k)$. *Therefore, an approximation for* $p(\lambda_k|\mathbf{z}_{1:k})$ *is inevitable. This paper uses the simplest approximation, i.e., the Dirac delta function, which is similar to the approximation in [15]. However, the form of the auxiliary parameters in [15] is very complex due to the heavy coupling with other parameters and the state vector (refer to (27) and (28) in [15]), which increases the error of approximating the complicated density into a Delta dirac function. Nevertheless, the form of* λ_k *in this paper is much simpler as (15) shows. Besides, the detailed rationality of this approximation will be discussed in Section IV-C.*

Using (17), (16) can be approximated as

$$p_{\theta_k}(\mathbf{x}_k|\mathbf{z}_{1:k}) \approx \mathrm{N}(\mathbf{x}_k;\hat{\mathbf{x}}'_{k|k}(\theta_k),\mathbf{P}'_{k|k}(\theta_k)/\mathrm{E}(\lambda_k)) \tag{19}$$

It is clear to see from (17) and (19) that both the posterior PDFs of the state vector and auxiliary random variable only rely on the posterior expectation of λ_k after the approximation in (17). Hence, the unknown posterior expectation of λ_k needs to be determined to obtain θ_k, whereby the required posterior PDFs can be achieved. Next, an approximate estimate of the expectation of λ_k will be derived based on a fixed-point iteration method.

D. Determination of the unknown expectation of λ_k

We still use $\theta_k = \mathrm{E}\{\lambda_k|\mathbf{z}_{1:k}\}$ to simplify the derivations. It can be observed from (10) that the unknown parameter θ_k is identical to the expectation of λ_k, i.e.,

$$\theta_k \approx \int_0^{+\infty} \lambda_k p(\lambda_k|\mathbf{z}_{1:k})d\lambda_k \tag{20}$$

Next, θ_k will be determined with the help of $p_{\theta_k}(\mathbf{z}_k|\mathbf{z}_{1:k-1})$.

Proposition 2. *The measurement likelihood PDF in (12) can be written as EC distributed, i.e.,*

$$p_{\theta_k}(\mathbf{z}_k|\mathbf{z}_{1:k-1}) = \mathrm{EC}(\mathbf{z}_k;\hat{\mathbf{z}}_{k|k-1},\mathbf{P}^{\mathbf{zz}}_{k|k-1}(\theta_k),g(\cdot)) \tag{21}$$

Proof. The result can be obtained by integrating $\mathbf{x}_k$ into (12) using the Gaussian integral formula and **Lemma 1**. □

By exploiting (21), an ingenious trick will be introduced to attain the solution for θ_k (i.e., $\mathrm{E}\{\lambda_k|\mathbf{z}_{1:k}\}$).

Proposition 3. θ_k *satisfies the following equation*

$$\theta_k = -2\dot{g}(\mathrm{tr}\{\tilde{\mathbf{P}}^{\mathbf{zz}}_{k|k-1}(\mathbf{P}^{\mathbf{zz}}_{k|k-1}(\theta_k))^{-1}\}) \tag{22}$$

where $\tilde{\mathbf{P}}^{\mathbf{zz}}_{k|k-1}$ *denotes the second-order moment sample of the innovation, which satisfies*

$$\tilde{\mathbf{P}}^{\mathbf{zz}}_{k|k-1} = (\mathbf{z}_k - \hat{\mathbf{z}}_{k|k-1})(\mathbf{z}_k - \hat{\mathbf{z}}_{k|k-1})^T \tag{23}$$

Proof. See Appendix C. □

It can be observed that (22) is a nonlinear equation due to the form of $\mathbf{P}^{\mathbf{zz}}_{k|k-1}(\theta_k)$ in (14). Thus, it is very difficult to derive an analytical solution. To solve this problem, we utilize a fixed-point iteration approach to obtain an approximate solution for the unknown parameter θ_k. That is to say, the unknown parameter θ_k is updated as $\theta_k^{(i+1)}$ at the $(i+1)-th$ iteration by employing $\theta_k = \theta_k^{(i)}$ in (22) to calculate the auxiliary function $\dot{g}(\cdot)$ approximately, i.e.,

$$\theta_k^{(i+1)} = -2\dot{g}(\mathrm{tr}\{\tilde{\mathbf{P}}^{\mathbf{zz}}_{k|k-1}(\mathbf{P}^{\mathbf{zz}}_{k|k-1}(\theta_k^{(i)}))^{-1}\}) \tag{24}$$

where

$$\mathbf{P}^{\mathbf{zz}}_{k|k-1}(\theta_k^{(i)}) = \mathbf{H}_k\theta_k^{(i)}\mathbf{P}_{k|k-1}\mathbf{H}_k^T + \mathbf{R}_k \tag{25}$$

To correspond to the situation without outliers which means no corrections to the MNCM, the unknown parameter θ_k is initialized as 1, i.e,

$$\theta_k^{(0)} = 1 \tag{26}$$

Algorithm 1: One-time step of the proposed RECKF.

Inputs: $\hat{\mathbf{x}}_{k-1|k-1}, \mathbf{P}_{k-1|k-1}, \mathbf{F}_k, \mathbf{H}_k, \mathbf{z}_k, \mathbf{Q}_k, \mathbf{R}_k, \epsilon, N_{max}$.

Time-update of the state vector:

1. $\hat{\mathbf{x}}_{k|k-1} = \mathbf{F}_k \hat{\mathbf{x}}_{k-1|k-1}$
2. $\mathbf{P}_{k|k-1} = \mathbf{F}_k \mathbf{P}_{k-1|k-1} \mathbf{F}_k^T + \mathbf{Q}_k$
3. $\hat{\mathbf{z}}_{k|k-1} = \mathbf{H}_k \hat{\mathbf{x}}_{k|k-1}$
4. $\tilde{\mathbf{P}}^{\mathbf{zz}}_{k|k-1} = (\mathbf{z}_k - \hat{\mathbf{z}}_{k|k-1})(\mathbf{z}_k - \hat{\mathbf{z}}_{k|k-1})^T$

Fixed-point iteration of the unknown parameter θ_k:

5. Initialize $\theta_k^{(0)} = 1$

for $i = 0 : N_{max} - 1$

6. $\mathbf{P}^{\mathbf{zz}}_{k|k-1}(\theta_k^{(i)}) = \mathbf{H}_k \theta_k^{(i)} \mathbf{P}_{k|k-1} \mathbf{H}_k^T + \mathbf{R}_k$
7. $\theta_k^{(i+1)} = -2\dot{g}\left(tr\{\tilde{\mathbf{P}}^{\mathbf{zz}}_{k|k-1}(\mathbf{P}^{\mathbf{zz}}_{k|k-1}(\theta_k^{(i)}))^{-1}\}\right)$
8. If $(|\theta_k^{(i+1)} - \theta_k^{(i)}|)/\theta_k^{(i)} \leq \epsilon$, terminate the iteration

end for

Analytical measurement update of the state vector:

9. $\hat{\theta}_k = \theta_k^{(N_{min})}$
10. $\mathbf{K}_k(\hat{\theta}_k) = \hat{\theta}_k \mathbf{P}_{k|k-1} \mathbf{H}_k^T (\mathbf{P}^{\mathbf{zz}}_{k|k-1}(\hat{\theta}_k))^{-1}$
11. $\hat{\mathbf{x}}_{k|k} = \hat{\mathbf{x}}_{k|k-1} + \mathbf{K}_k(\hat{\theta}_k)(\mathbf{z}_k - \hat{\mathbf{z}}_{k|k-1})$
12. $\mathbf{P}_{k|k} = (\mathbf{I}_n - \mathbf{K}_k(\hat{\theta}_k)\mathbf{H}_k)\mathbf{P}_{k|k-1}$

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

After the fixed-point iteration converges, the estimate of the unknown parameter θ_k is given by $\hat{\theta}_k = \theta_k^{(N_{min})}$, where $\hat{\theta}_k$ denotes the approximate estimate for the unknown parameter θ_k, and N_{min} denotes the minimum number of fixed-point iterations to guarantee convergence. The convergence will be illustrated in the next section. Substituting $\theta = \hat{\theta}_k$ into (14), the required posterior state estimate and covariance can be approximated as

$$\hat{\mathbf{x}}_{k|k} = \hat{\mathbf{x}}'_{k|k}(\hat{\theta}_k), \quad \mathbf{P}_{k|k} = (\mathbf{I}_n - \mathbf{K}_k(\hat{\theta}_k)\mathbf{H}_k)\mathbf{P}_{k|k-1} \tag{27}$$

The proposed RECKF is composed of the fixed-point iteration in (24)-(26) and the analytical update in (27). The implementation pseudo code of the proposed RECKF is listed in **Algorithm 1**, where ϵ and N_{max}, respectively, represent the iterative threshold and the maximum number of iterations. We can see from step 6-step 7 that the iterative process of the proposed algorithm is self-related, thus we will denote the iteration of the proposed algorithm as self-iteration.

IV. Theoretical analyses of the proposed RECKF

Up to now, we have completed the analytical joint update and obtained the analytical solution for the posterior PDFs. In addition, the proposed RECKF is established as shown in Algorithm 1. In this section, four properties will be analyzed to prove the reliability of the RECKF. The convergence analysis of the iterative process is presented in Section IV-A, and the robustness analysis of the proposed RECKF is specified in Section IV-B, and the rationality of the approximation in (17) is discussed in Section IV-C, and the computational complexity analysis is specified in Section IV-D.

Before the theoretical analyses of the proposed RECKF, two useful assumptions are given first.

Assumption 1. *If there is a measurement outlier at time k, the following approximated equality holds*

$$\mathrm{tr}\left\{\tilde{\mathbf{P}}^{\mathbf{zz}}_{k|k-1}\left(\mathbf{H}_k \mathbf{P}_{k|k-1} \mathbf{H}_k^T + \mathbf{R}_k\right)^{-1}\right\} \approx m \tag{28}$$

Assumption 2. *If there is a measurement outlier at time k, the following inequality holds*

$$\mathrm{tr}\left\{\tilde{\mathbf{P}}^{\mathbf{zz}}_{k|k-1}\left(\mathbf{H}_k \mathbf{P}_{k|k-1} \mathbf{H}_k^T + \mathbf{R}_k\right)^{-1}\right\} > m \tag{29}$$

In general, if there are no measurement outliers at time k, the second-order moment sample of the innovation $\tilde{\mathbf{P}}^{\mathbf{zz}}_{k|k-1}$ (named instantaneous innovation covariance matrix) is close to the theoretical innovation covariance matrix $\left(\mathbf{H}_k \mathbf{P}_{k|k-1} \mathbf{H}_k^T + \mathbf{R}_k\right)$. However, if there is a measurement outlier at time k, the difference between the measurement and the predicted measurement will enlarge, which can increase the instantaneous innovation covariance matrix. Thus, Assumption 1 and 2 are reasonable in theory. In fact, we find that (28) and (29) hold for more than 97% of time steps in our simulation.

A. Convergence analysis of the iterative process

In this part, the convergence of the proposed fixed-point iteration approach will be proved through confirming the boundedness and monotonicity of the iterative sequence $\{\theta_k^{(i)}\}_{i=1}^{+\infty}$.

Theorem 2. *The proposed fixed-point iteration method can achieve global convergence.*

Proof. See Appendix D. □

Theorem 2 reveals that the iterative process can achieve a unique solution regardless of the initial value. Nevertheless, we should highlight that the global convergence point is not guaranteed to be the global optimal point because the global convergence point is just a point that the algorithm 'thinks' is best. Thus, a numerical simulation will be carried out next to verify the quality of the convergent value of the RECKF.

B. Robustness analysis of the proposed RECKF

In this part, the robustness of the proposed RECKF is proved first. Then we carry out a numerical simulation to compare the robustness of the proposed RECKF and the VB-based method (represented by the ORKF-VB [10]). To prove the robustness of the proposed RECKF, a useful theorem is given first.

Theorem 3. *If there are measurement outliers, the iterative sequence fulfills the following constraint*

$$0 < \theta_k^{(i+1)} < \theta_k^{(i)} < 1, \quad for \quad \forall i > 0 \tag{30}$$

Proof. See Appendix E. □

Substituting $\theta_k = \theta_k^{(i+1)}$ and $\theta_k = \theta_k^{(i)}$ into (14) and employing **Theorem 3**, we can easily obtain $\left\|\mathbf{K}_k(\theta_k^{(i+1)})\right\|_F < \left\|\mathbf{K}_k(\theta_k^{(i)})\right\|_F$ in the presence of outliers. This result illustrates that the proposed filter can identify the measurement outliers adaptively, and then reduce the level of confidence on the measurement innovation and thereby confirms the robustness of the proposed RECKF.

Next, the difference between the ORKF-VB and the proposed RECKF is illustrated in detail. To be clearer, we select the LEC (i.e., Student's t) distribution to represent the proposed RECKF to be consistent with [10]. It is not hard to observe that both the ORKF-VB (refer to Algorithm 1 in

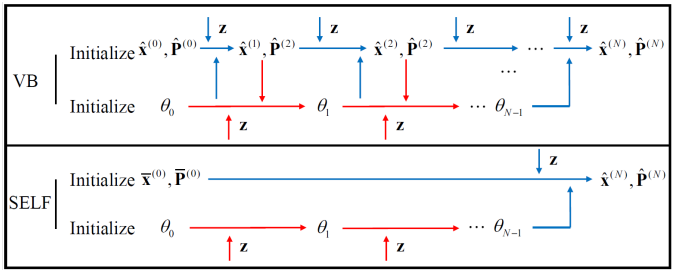


Fig. 2: Iterative mechanisms of the ORKF-VB and the proposed algorithm (SELF).

[10]) and the proposed RECKF (refer to step 10 in Algorithm 1) put less confidence in the possible measurement outlier by enlarging the measurement noise covariance matrix (MNCM) (i.e., decreasing the scalar divided by the nominal MNCM), whereby the robustness can be achieved. The scalar estimates of the ORKF-VB and the proposed algorithm are given as follows [10] (time subscripts are omitted)

$$\theta_{VB}^{(i+1)} = f(\mathrm{tr}\{\underbrace{[(\mathbf{z}-\mathbf{H}\hat{\mathbf{x}}^{(i)})(\mathbf{z}-\mathbf{H}\hat{\mathbf{x}}^{(i)})^T + \mathbf{H}\hat{\mathbf{P}}^{(i)}\mathbf{H}^T]}_{\mathbf{D}^{(i)}}\mathbf{R}^{-1}\}) \tag{31a}$$

$$\theta_{Self}^{(i+1)} = f(\mathrm{tr}\{\underbrace{[(\mathbf{z}-\mathbf{H}\bar{\mathbf{x}})(\mathbf{z}-\mathbf{H}\bar{\mathbf{x}})^T]}_{\tilde{\mathbf{P}}_{\mathbf{z}}}\underbrace{[\theta_{Self}^{(i)}\mathbf{H}\bar{\mathbf{P}}\mathbf{H}^T + \mathbf{R}]^{-1}}_{=:\theta_{Self}^{(i)}\mathbf{P}_{\mathbf{z}}^{(i)}}\}) \tag{31b}$$

where $f(t) = \frac{\nu+p}{\nu+t}$, and $\mathbf{D}^i$ denotes the second-moment order sample of the residual, and $\bar{\mathbf{x}}$ and $\hat{\mathbf{x}}^{(i)}$, $\hat{\mathbf{P}}^{(i)}$ denote the prior state estimate and the posterior state estimate and covariance at the i-th iteration, respectively, where the detailed expression can be found in [10]. It is worth noting that the only difference in algorithm execution of these two methods lies in their iteration process. Specifically, the ORKF-VB adjusts the scalar estimate according to the difference between $\mathbf{D}^i$ and the nominal MNCM $\mathbf{R}$, as shown in (31a), while (31b) shows that the proposed method executes this adjustment based on the difference between the instantaneous innovation covariance matrix, i.e, $\tilde{\mathbf{P}}_{\mathbf{z}}$, and the theoretical innovation covariance matrix (which will be $\theta_{Self}^{(i)}\mathbf{P}_{\mathbf{z}}^{(i)}$ in our case). The iterative mechanisms of the ORKF-VB and the proposed RECKF are shown in Fig. 2, where the blue lines refer to the KF process and the red lines denote the update process of $\theta^{(i)}$.

Fig. 2 shows that the ORKF-VB requires loop iterations between the posterior $\hat{\mathbf{x}}^{(i)}$, $\hat{\mathbf{P}}^{(i)}$ and $\theta^{(i)}$, whose purpose is to improve the estimation accuracy of θ by utilizing more accurate posterior state information. In this setting, the updates of the state estimate and auxiliary variable in the ORKF-VB are tightly coupled, leading to a better estimation of θ but more computation burden. In contrast, the proposed algorithm performs self-iterations[3] of the auxiliary variable θ, which is decoupled from the posterior state and covariance updates and gives rise to the computational efficiency of the proposed algorithm. Although the proposed algorithm does not take advantage of posterior information, it can bring about other benefits, as explained next. Note that in the remainder, we represent the iterative approaches in the ORKF-VB and the proposed algorithm as the coupling-iteration and the self-iteration, respectively. The advantage of the proposed self-iteration is illustrated in detail next.

[3]In the self-iteration, we mainly refer that the update process of θ is directly self-relevant rather than requiring the intervention of posterior information.

Proposition 4. *Under Assumption 2 and the condition that the initial values of the coupling-iteration and the self-iteration are equal, the following equality holds*

$$\theta_{VB}^{(1)} = f\left(\mathrm{tr}\left\{\frac{\mathbf{R}(\mathbf{P}_{\mathbf{z}}^{(0)})^{-1}\tilde{\mathbf{P}}_{\mathbf{z}}(\mathbf{P}_{\mathbf{z}}^{(0)})^{-1} + \mathbf{R}(\mathbf{P}_{\mathbf{z}}^{(0)})^{-1} + (\theta^{(0)})\mathbf{I}_m}{(\theta^{(0)})^2}\right\}\right) \tag{32a}$$

$$\theta_{Self}^{(1)} = f\left(\mathrm{tr}\left\{\tilde{\mathbf{P}}_{\mathbf{z}}(\mathbf{P}_{\mathbf{z}}^{(0)})^{-1}/(\theta^0)^2\right\}\right) \tag{32b}$$

where $\theta_{VB}^0 = \theta_{Self}^0 = \theta^0$.

Proof. Proposition 4 can be proved by substituting $\theta_{VB}^0 = \theta_{Self}^0 = \theta^0$ into (31a) and (31b). □

Employing Assumption 2, we can easily obtain the following two results by subtracting (32a) from (32b)

- The first value obtained by the self-iteration is smaller than that by the coupling-iteration in the case of a measurement outlier, i.e., $\theta_{VB}^{(1)} \geq \theta_{Self}^{(1)}$.
- The first value obtained by the self-iteration is larger than that by the coupling-iteration in the case of no measurement outliers, i.e., $\theta_{VB}^{(1)} \leq \theta_{Self}^{(1)}$.

With these results, the following corollary can be attained.

Corollary 1. *Under the condition of the same iterative number, the self-iteration can achieve a smaller θ than the coupling-iteration in the case of measurement outliers, and a bigger θ than the coupling-iteration in the case of no measurement outliers.*

Proof. Due to the monotonicity of both iterations, $\theta_{VB}^{(i+1)} \leq \theta_{VB}^{(i)}$ and $\theta_{Self}^{(i+1)} \leq \theta_{Self}^{(i)}$ will hold in the case of measurement outliers. Assuming that $\theta_{Self}^{(i)} \leq \theta_{VB}^{(i)}$ holds for $i = j$ and setting $\theta_{VB}^{(i)} = \theta_{Self}^{(i)}$, we can obtain similar results with Proposition 4, i.e, $\theta_{Self}^{(i+1)} \leq \theta_{VB}^{(i+1)}$. □

Corollary 1 indicates that the self-iteration adopts a more conservative strategy than the coupling-iteration when there are measurement outliers, which is beneficial to the filter in the case of outliers. We carry out a numerical simulation to make this corollary more convincing. The values obtained by the coupling-iteration and the self-iteration are compared under the same initial conditions, i.e., the same $\mathbf{z}$, $\bar{\mathbf{x}}$, $\bar{\mathbf{P}}$, $\mathbf{R}$, $\mathbf{H}$ and $\theta_{VB}^{(0)} = \theta_{Self}^{(0)} = 1$ in (31b). In addition, the measurement at the current time is contaminated by Gaussian noise with enlarged covariance matrix, i.e., $\mathbf{v} \sim \mathrm{N}(\mathbf{0}, U\mathbf{R})$. 1000 Monte Carlo runs are executed and both iterations are performed 50 times. The results are shown in Fig. 3, where the reciprocals of baselines are the true MNCMs. It can be seen from Fig. 3 that the values from the self-iteration are generally less than the coupling-iteration, except for the case without measurement outliers, which is consistent with our previous analyses. This mechanism helps self-iteration further approach the baselines and achieve better robustness. Moreover, although the self-iteration may suffer from accuracy degradation due to overconfidence

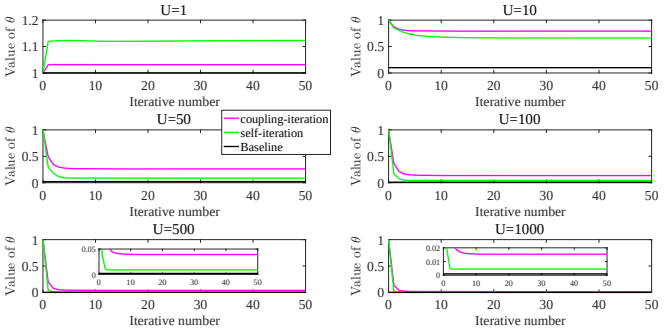


Fig. 3: Iteration results of the coupling-iteration and the self-iteration under different measurement outlier intensities.

in the outlier-free scenarios, this accuracy degradation is trivial compared with the improvement brought about in the case of measurement outliers, as will be shown in Section VI-A.

C. The rationality of the approximation in (17)

The rationality of the approximation (17) is demonstrated in both the outlier case and outlier-free case. Firstly, the rationality in the outlier case is illustrated in **Proposition 5**.

Proposition 5. *As the iteration proceeds, the variance of the auxiliary random variable λ_k reduces gradually in the presence of measurement outliers, i.e.,*

$$Var^{(i+1)}(\lambda_k) < Var^{(i)}(\lambda_k) \quad for \quad \forall i \geq 0 \tag{33}$$

where $Var^{(i)}(\lambda_k)$ denotes the variance of λ_k at the $(i)-th$ iteration.

Proof. See Appendix F. □

It is worth noting that when there are outliers, $Var(\lambda_k)$ will decrease, indicating that λ_k will tend to its mean. In the other case, the likelihood PDF $p(\mathbf{z}_k|\mathbf{z}_{1:k-1})$ should degrade into Gaussian, which means $\lambda_k = 1$ holds for the conditional likelihood PDF. Therefore, using condition ℓ_5 and Assumption 1, we can obtain $\hat{\theta}_k \approx 1$ according to the mathematical induction method, which indicates no corrections for the measurement noise and thus is consistent with the theoretical case. Hence, the approximation in (17) is rational whether there are measurement outliers or not.

D. Computational complexity analysis

In this part, the computational complexity of the proposed algorithm is compared with the existing ORKF-VB algorithm which has relatively low complexity. The computational complexities include the floating point operations required by the multiplications of matrices and vectors, the multiplications of matrices, and the inverse of a positive definite matrix. The computational complexities of the existing ORKF-VB and the proposed filter in one-time step can be, respectively, formulated as follows

$$\begin{aligned} fl_{VB} &= (N+2)n^3 + N[4mn^2 + m^3 \\ &+ 3nm^2 + 2O(m^3) + 3mn + m^2] \end{aligned} \tag{34}$$

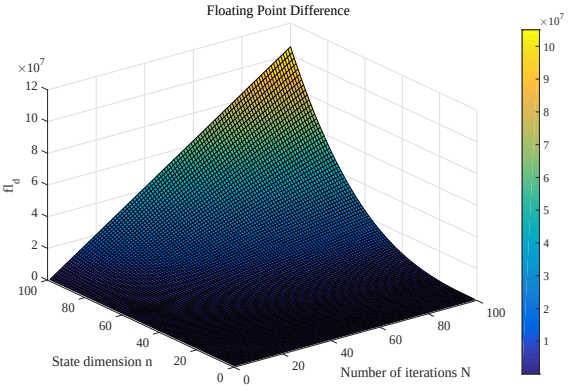


Fig. 4: Floating point difference fl_d.

$$\begin{aligned} fl_{new} &= 3n^3 + N[mn^2 + m^3 + nm^2 \\ &+ O(m^3)] + 2mn^2 + nm^2 + O(m^3) \end{aligned} \tag{35}$$

where fl_{VB} and fl_{new} denote the total floating point operations of the existing ORKF-VB and the proposed filter, respectively. Subtracting (35) from (34) yields

$$\begin{aligned} fl_{VB} - fl_{new} &= (N-1)n^3 + (3N-2)mn^2 \\ &+ NO(m^3) + (2N-1)nm^2 \end{aligned} \tag{36}$$

Considering that the number of iterations N is always larger than 1, i.e., $N \geq 1$, the existing ORKF-VB always has larger computational complexity than the proposed algorithm, especially in the scenario where the dimension of the state vector is large, such as in an integrated navigation system. Additionally, with the number of iterations increasing, the existing ORKF-VB becomes more computationally burdensome than the proposed algorithm. Therefore, $fl_{VB} \geq fl_{new}$ holds.

Fig. 4 illustrates the difference between the floating point operations fl_{VB} and fl_{new}, where the dimension of the measurement vector is set as $m = 2$, and the x-axis is the number of iterations N, the y-axis is the state dimension n, and the z-axis is the floating point difference fl_d that is defined as $fl_d = fl_{VB} - fl_{new}$. It is clear to observe from Fig. 4 that the floating point difference fl_d gradually increases as the state dimension n and the number of iterations N enlarges from 1 to 100, which is consistent with the above analysis.

V. An IMM-based adaptive selection of EC distributions

So far, the effectiveness and the advantages of the RECKF have been proved. Thus, as mentioned in Section II-B, under the premise that the true heavy-tailed feature has been known, the proposed filter can deal with the measurement noise with an arbitrary thick tail. Nevertheless, two significant factors make the premise unattainable, as listed below.

- The extent of the thick tail of the measurement noise is usually unknown in practice which leads to difficulty in the selection of the EC distributions.
- In engineering applications, outliers are often time-varying and unpredictable, so it is unreasonable to select a fixed EC distribution to model the non-stationary measurement noise.

Therefore, how to circumvent the prior knowledge of outliers and find a general method that can accommodate all cases (even in an outlier-free scenario) is worthy of further investigation. In this paper, to address the above problems, multiple EC distributions with different extent of thick tails are considered to deal with the unknown measurement noise features under the practical environment. Then, to adapt to the time-varying and non-stationary outliers, the IMM approach is exploited to adaptively select EC distributions, from which an IMM-based RECKF will be derived.

A. Special SSM design with measurement outliers

Based on (1), we consider the following jump Markov linear discrete-time SSM

$$\begin{cases} \mathbf{x}_k = \mathbf{F}_k \mathbf{x}_{k-1} + \mathbf{w}_k \\ \mathbf{z}_k = \mathbf{H}_k \mathbf{x}_k + \mathbf{v}(\xi_k) \end{cases} \tag{37}$$

where $\xi_k \in \{1, 2, ..., N\}$ denotes the mode state. In this paper, $\mathbf{v}(\xi_k)$ is assumed to obey an EC distribution selected for the mode state ξ_k.

The time behavior of the mode state $\xi_k \in \{1, 2, ..., N\}$ is modeled as a homogeneous (time invariant) Markov chain with transition probability matrix (TPM) $\mathrm{H} = \left[\mathrm{H}_{ij} \triangleq P\{\xi_k = i | \xi_{k-1} = j\}\right]$. The TPM denotes the probability of the model j transferring to the model i.

B. Derivation of the IMM-based RECKF

In this subsection, the IMM-based RECKF will be derived within four steps.

Step 1: Mixing

Firstly, the TPM is employed to make a prediction of the probability for mode i, i.e.,

$$P\{\xi_k = i | \mathbf{z}_{1:k-1}\} = \sum_{j=1}^{N} \mathrm{H}_{ij} P\{\xi_{k-1} = j | \mathbf{z}_{1:k-1}\} \tag{38}$$

where $P\{\xi_k = i | \mathbf{z}_{1:k-1}\}$ is defined as the mixed prior probability matched to $\xi_k = i$, and N denotes the number of models.

Define $p(\mathbf{x}_{k-1} | \xi_k = i, \mathbf{z}_{1:k-1})$ as the mixed prior PDF of the state vector matched to $\xi_k = i$, which will be used in the time-update of the IMM-based RECKF for the state vector. The mixed prior PDF of $\mathbf{x}_{k-1}$ can be formulated as follows

$$p(\mathbf{x}_{k-1} | \xi_k = i, \mathbf{z}_{1:k-1}) = \sum_{j=1}^{N} \mathrm{H}_{ij} \frac{p(\mathbf{x}_{k-1} | \xi_{k-1} = j, \mathbf{z}_{1:k-1})}{P\{\xi_k = i | \mathbf{z}_{1:k-1}\}} \tag{39}$$

In that the $(i) - th$ model at time k could be transformed from any model at time $k-1$, the mixed prior PDF will participate in the time-update of the RECKF at time k instead of the posterior PDF matched to time $k-1$. The mixed prior PDF is approximated as

$$p(\mathbf{x}_{k-1} | \xi_k = i, \mathbf{z}_{1:k-1}) \approx \mathrm{N}(\mathbf{x}_{k-1}; \hat{\mathbf{x}}^{0i}_{k-1|k-1}, \mathbf{P}^{0i}_{k-1|k-1}) \tag{40}$$

where the merged mean $\hat{\mathbf{x}}^{0i}_{k-1|k-1}$ and covariance $\mathbf{P}^{0i}_{k-1|k-1}$ are obtained by moment matching.

Step 2: Parallel RECKFs

In this step, N parallel RECKFs are performed using the corresponding merged mean $\hat{\mathbf{x}}^{0i}_{k-1|k-1}$ and covariance $\mathbf{P}^{0i}_{k-1|k-1}$ as inputs, and each parallel RECKF obtains a different result by employing **Algorithm 1** which will be used in the fourth step.

Step 3: Posterior probability calculation

The posterior probability for mode i is a necessity before all estimates and covariances are fused in probability to obtain a comprehensive result. Employing the Bayes' rule, which can be calculated as

$$P\{\xi_k = i | \mathbf{z}_{1:k}\} = \frac{P\{\xi_k = i | \mathbf{z}_{1:k-1}\} p(\mathbf{z}_k | \xi_k = i, \mathbf{z}_{1:k-1})}{p(\mathbf{z}_k | \mathbf{z}_{1:k-1})} \tag{41}$$

where $p(\mathbf{z}_k | \xi_k = i, \mathbf{z}_{1:k-1})$ denotes the measurement likelihood PDF for mode i, and $p(\mathbf{z}_k | \mathbf{z}_{1:k-1})$ is written as

$$p(\mathbf{z}_k | \mathbf{z}_{1:k-1}) = \sum_{i=1}^{N} P\{\xi_k = i | \mathbf{z}_{1:k-1}\} p(\mathbf{z}_k | \xi_k = i, \mathbf{z}_{1:k-1}) \tag{42}$$

Based on (21), the measurement likelihood PDF for mode i can be formulated as

$$p(\mathbf{z}_k | \xi_k = i, \mathbf{z}_{1:k-1}) = c_g^i \sqrt{|\mathbf{P}^{i,\mathbf{zz}}_{k|k-1}(\theta_k)|} \exp(g^i(\Delta_k^i)) \tag{43}$$

where $\Delta_k^i = (\mathbf{z}_k - \hat{\mathbf{z}}^i_{k|k-1})^T (\mathbf{P}^{i,\mathbf{zz}}_{k|k-1}(\theta_k))^{-1} (\mathbf{z}_k - \hat{\mathbf{z}}^i_{k|k-1})$ and $\mathbf{P}^{i,\mathbf{zz}}_{k|k-1}(\theta_k)$, $\hat{\mathbf{z}}^i_{k|k-1}$ and c_g^i denote the innovation covariance matrix, predicted measurement vector and normalized constant of the $(i) - th$ model, respectively. Exploiting (38), (42) and (43), the posterior probability for mode i can be obtained.

Remark 5. *Note that the measurement likelihood PDF for mode i is related with parameter θ_k. In this paper, $\hat{\theta}_k$ is chosen as a reasonable approximation for θ_k because it would be difficult to track changes in parameter θ_k at every moment to match the likelihood of each mode, such approximation results in inevitable but acceptable errors.*

Remark 6. *The core of the proposed IMM-based RECKF is the weight calculation of respective RECKFs, which depends on the likelihood calculation in (43). This creates a problem because the constant coefficient c_g of an EC distribution is unknown. Although the coefficients of some special analytic EC distributions like the Gaussian or Student's t are known, many more are unknown, and some are not even densities (EEC distribution). In this case, numerical integration can be used to calculate the normalized coefficients of EC distributions. In short, based on the property that the integral of the density in (21) is 1, a massive number of samples of the theoretical innovation covariance matrix and the second-order moment of the innovation are utilized to calculate c_g.*

Step 4: Overall estimate and covariance calculation

Employing (41), the overall posterior estimate and covariance can be combined as follows,

$$\hat{\mathbf{x}}_{k|k} = \sum_{i=1}^{N} \hat{\mathbf{x}}^i_{k|k} P\{\xi_k = i | \mathbf{z}_{1:k}\} \tag{44}$$

$$\mathbf{P}_{k|k} = \sum_{i=1}^{N} P\{\xi_k = i | \mathbf{z}_{1:k}\} \left[\mathbf{P}^i_{k|k} + (\hat{\mathbf{x}}^i_{k|k} - \hat{\mathbf{x}}_{k|k})(\hat{\mathbf{x}}^i_{k|k} - \hat{\mathbf{x}}_{k|k})^T\right]$$

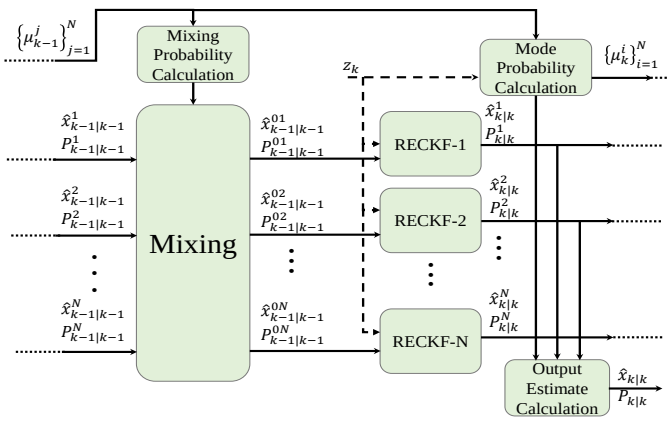


Fig. 5: The scheme of the IMM-based RECKF.

Algorithm 2: One-time step of the IMM-based RECKF.

Inputs: $\left\{\hat{\mathbf{x}}^i_{k-1|k-1}, \mathbf{P}^i_{k-1|k-1}, \mu^i_{k-1}\right\}_{i=1}^{N}, \mathbf{F}_k, \mathbf{H}_k, \mathbf{z}_k, \mathbf{Q}_k, \mathbf{R}_k$.

Mixing:

Calculate:

for $i = 1 : N$

1. the mixed prior probability $P\{\xi_k|\mathbf{z}_{1:k-1}\}$ in (38)

3. the merged mean and covariance $\hat{\mathbf{x}}^{0i}_{k-1|k-1}, \mathbf{P}^{0i}_{k-1|k-1}$ in (40)

Parallel RECKFs:

4. the posterior estimate and covariance $\hat{\mathbf{x}}^i_{k|k}, \mathbf{P}^i_{k|k}$ in (44)

Posterior probability:

5. the measurement likelihood PDF $p(\mathbf{z}_k|\xi_k = i, \mathbf{z}_{1:k-1})$ in (43)

end for

6. the overall measurement likelihood PDF $p(\mathbf{z}_k|\mathbf{z}_{1:k-1})$ in (42)

7. the posterior probability $P\{\xi_k|\mathbf{z}_{1:k}\}$ in (41)

Fusion

7. the overall posterior information $\hat{\mathbf{x}}_{k|k}, \mathbf{P}_{k|k}$ in (44)

Outputs: $\hat{\mathbf{x}}_{k|k}$ and $\mathbf{P}_{k|k}$, $\left\{\hat{\mathbf{x}}^i_{k|k}, \mathbf{P}^i_{k|k}, \mu^i_k\right\}_{i=1}^{N}$

where $\hat{\mathbf{x}}^i_{k|k}$ and $\mathbf{P}^i_{k|k}$ denote the estimate state and estimate error covariance matrix of the $(i) - th$ model, respectively. The scheme of the proposed IMM-based RECKF algorithm is drawn in Fig. 5 where $\mu^i_k = P\{\xi_k = i|\mathbf{z}_{1:k}\}$.

The implementation pseudo code of the proposed RECKF is listed in **Algorithm 2**. In summary, the IMM-based RECKF performs well for dealing with unknown and time-varying measurement noise by comprehensively considering every model in the model set. That is to say, instead of picking the most suitable model at every moment, it uses every model in the model set, but with different probability. Additionally, the IMM-based RECKF executes parallel RECKFs in time, and determines the probability for each mode state according to its likelihood probability. The likelihood probability of a mode state in essence represents the matching magnitude of the measurement (produced by the real model) to the model corresponding to the current mode. Advantage and disadvantage discussions are given next.

Remark 7. *For the non-interactive RECKF, the first advantage lies in its low computational complexity, as analyzed in Section IV-D. The underlying reason is that the proposed RECKF adjusts the covariance matrix by changing a scalar parameter θ. The proposed algorithm is more effective when the measurement outliers affect the whole dimension of the measurement, but when the measurement outliers have different effects on different dimensions, the performance of the RECKF algorithm will degrade, as will be shown in the simulation.*

For the IMM-based RECKF, the advantage is that it provides a feasible method to adaptively select the EC distributions in time-varying and unknown environments. Adaptively adjusting the weights of the pre-selected distributions makes the proposed IMM-based RECKF accommodate the environments quickly. However, when the environment is constant or changing slowly, the performance of the proposed IMM-based RECKF will degrade. For example, in an outliers-free scenario, the traditional KF is optimal, and the pre-selected linear EC distribution will play a decisive role in all selected distributions. Nevertheless, the weight of the linear EC distribution is not exactly 1 because other selected distributions will be assigned some weights as well. That is to say, the IMM-based RECKF is essentially a suboptimal filter.

C. Discussions about the existing IMM filters

Finally, some discussions are conducted to reveal the differences among the proposed IMM-based RECKF, the Gaussian IMM filter [26] and the Student's t IMM filter [27]. On one hand, the Gaussian IMM filter is limited to its nominal MNCM settings. Its estimation accuracy will deteriorate significantly with a relatively adventurous nominal MNCM. To be more specific, when the true MNCM is larger than any of its nominal MNCM settings, large errors will be induced. Although a conservative nominal MNCM setting (i.e., selecting a model with a very large nominal MNCM) can alleviate this problem, poor performance can be achieved in the case of a mild environment. Therefore, the performance of the Gaussian IMM filter depends heavily on the accuracy of the prior knowledge of the environment, which will be shown in Section VI-B.

On the other hand, each sub-filter of the Student's t IMM filter is a Student's t filter, which avoids the adventurous MNCM setting problem to some extent because the Student's t distribution has heavier tails than Gaussian with a large covariance matrix. Hence, each sub-filter can resist outliers itself rather than relying heavily on the nominal MNCM settings. However, the moment matching strategy adopted to ensure the posterior analytical and multiple Student's t fusion[4] brings about unavoidable estimation errors. Although [28] can address the latter problem, the former problem is still unsolvable. Additionally, the inherent problem of the Student's t distribution mentioned in Section I-B limits its accuracy in the case of only measurement outliers. Thus, the proposed IMM-based RECKF has theoretical advantages over the Gaussian IMM filter and the existing Student's t IMM filter, as will be shown in Section VI-B. In contrast, the proposed RECKF is essentially a Gaussian approximation filter whose posterior state density is Gaussian. This feature enables the proposed IMM-based RECKF to avoid the problem of large approximation error induced by multiple Student's fusion [28].

[4]In [27], fusing multiple Student's t distributions into one Student's t distribution is required in calculating mixed prior PDF, similarly to (40). According to [28], these approximation errors are very large. Nonetheless, the error of approximating multiple Gaussian distributions as one Gaussian distribution will remain bounded [35].

Remark 8. *Although the proposed RECKF can theoretically perform better than some VB-based methods in the case of measurement outliers as illustrated in Section IV-E, it will be difficult to extend it to address state uncertainty like a VB-based robust algorithm. As a matter of fact, most VB-based algorithms identify the state outlier by using the difference between the posterior state estimate and the prior state estimate. However, due to abandoning using the posterior information, the proposed RECKF cannot resist state outliers, which means the proposed algorithm can only apply in some scenarios where the state uncertainty is mild, for example, where the vehicle is equipped with an INS or an odometry [36]. Nevertheless, we would like to emphasize that being rid of the ability to deal with state outliers can bring about better ability to address measurement outliers because the IMM technique can be utilized to improve the performance in complex environments. In some existing VB-based methods which can resist the state and measurement outliers simultaneously, the IMM approach cannot be used to improve performance in complex environments because their measurement likelihood PDF is not analytical (not Student's t or Gaussian). Overall, the proposed algorithm gives up the ability to deal with state outliers, but obtains better suppression of measurement outliers in return.*

VI. Simulations

In this section, a target tracking example is utilized to demonstrate the effectiveness and superiority of the proposed IMM-based RECKF. The state vector is composed of the positions and velocities, and the state transition matrix and measurement matrix are, respectively, set as $\mathbf{F}_k = \begin{bmatrix} \mathbf{I}_2 & \Delta t\mathbf{I}_2 \\ \mathbf{0}_2 & \mathbf{I}_2 \end{bmatrix}$ and $\mathbf{H}_k = \begin{bmatrix} \mathbf{I}_2 & \mathbf{0}_2 \end{bmatrix}$, where $\Delta t = 1s$ denotes the sampling interval. The Gaussian distributed state and non-Gaussian heavy-tailed distributed measurement noises are generated as

$$\begin{cases} \mathbf{w}_k \sim \mathrm{N}(\mathbf{0}, \mathbf{Q}) \\ \mathbf{v}_k \sim \begin{cases} \mathrm{N}(\mathbf{0}, \mathbf{R}) & w.p. \quad p \\ \mathrm{N}(\mathbf{0}, U\mathbf{R}) & w.p. \quad 1-p \end{cases} \end{cases} \tag{45}$$

where the notation $w.p.$ denotes "with probability", and U, p denote the scale factors and normal probabilities of state and measurement noises, respectively, and $\mathbf{Q}$ and $\mathbf{R}$ denote the state nominal and MNCMs, respectively, and they are selected as $\mathbf{Q} = \begin{bmatrix} \frac{\Delta t^3}{3}\mathbf{I}_2 & \frac{\Delta t^2}{2}\mathbf{I}_2 \\ \frac{\Delta t^2}{2}\mathbf{I}_2 & \Delta t\mathbf{I}_2 \end{bmatrix}$ and $\mathbf{R} = \mathbf{I}_2$, respectively. Some measurement sensors like lidar, can achieve such high measurement accuracy. Note that high accuracy measurement data are useful for some measurement-based learning algorithms. The simulation is segregated into two cases to illustrate the validity and advantages of the proposed algorithm. 1000 Monte Carlo runs are executed and the simulation time is chosen as $200s$ and $1000s$ for two cases, respectively, at each Monte Carlo run. The noise parameters setting for two cases are listed in Table II, and all outlier parameters are selected empirically because it's often difficult to quantify outliers, however, it makes intuitive sense that '100' is relatively smaller than '100000'. The root mean square errors (RMSEs), average RMSEs (ARMSEs) of position and velocity, normalized estimation squared (NEES) and average NEES (ANEES) are selected as performance metrics to compare the estimation accuracy and consistency, whose definition can be found in [12, 34].

TABLE II: Parameter settings of noises.

Case	Periods		Noise parameters
Case 1	\		U=1000, p=0.9
Case 2	Period 1 (1∼200s)		U=100, p=0.75
	Period 2 (201∼400s)		U=1000, p=0.85
	Period 3 (401∼600s)		U=1, p=1
	Period 4 (601∼800s)	601∼605s	U=100000, p=1
		606∼700s	U=100, p=0.85
		701∼705s	U=100000, p=1
		706∼800s	U=100, p=0.85
	Period 5 (801∼1000s)		U=500, p=0.75

TABLE III: Parameter settings of all filters in case 1.

Filters	Parameter settings
ORKF-VB	dof parameter $\nu = 5$
HKF	Tuning parameter $\gamma = 1.345$
MCKF	Kernel size $\sigma = 5$
RSTKF	Tuning parameter $\tau = 5$
SSMKF	Similarity function sqrt

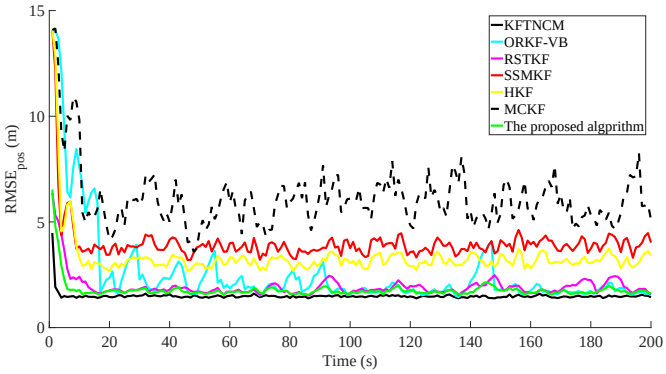


Fig. 6: RMSEs of the position and velocity for **Algorithm 1**.

A. Case 1: Non-interactive RECKF

1) Validity verification: In this case, our purpose is to demonstrate the validity of the proposed RECKF, and when the proposed RECKF is superior to the VB-based method. The LEC distribution is selected to implement the proposed RECKF algorithm. We conduct the comparisons among the proposed algorithm, the existing ORKF-VB [10], HKF [4], MCKF [7], RSTKF [12], SSMKF [9], and standard KF with true noise covariance matrices (KFTNCM) which is used as the accuracy reference. The parameter settings for all filters in this case are listed in Table III. Note that some compared algorithms can handle both state outliers and measurement outliers. Therefore, to be fair, the parameter adjustment function associated with the state error covariance in these algorithms is shielded and all filters adopt the accurate state noise covariance matrix to implement the filtering work.

The simulation results for case 1 are shown in Figs. 6 and Table IV. It can be seen from Fig. 6 that the proposed algorith-

TABLE IV: ARMSEs of position (AP) and velocity (AV) and ANEESs of all filters for **Algorithm 1**.

Filters	AP(m)	AV(m/s)	ANEESs	Times (0.1ms)
KFTNCM	1.4936	1.5300	4.0030	0.0978
ORKF-VB	2.5387	1.6578	7.1436	1.2533
HKF	3.3222	2.1451	10.9128	1.2942
MCKF	6.0876	2.7448	23.6304	1.3462
RSTKF	1.9395	1.6000	4.6417	9.0704
SSMKF	3.9826	2.5259	12.7442	1.0419
RECKF	1.7756	1.5888	4.5415	1.0403

TABLE V: The percentage gain of accuracy of all algorithms in the case of multi-outliers for **Algorithm 1**.

Filters	AP-1	AP-2	PgAP	AV-1	AV-2	PgAV
KFTNCM	1.4936	1.3635	**8.71%**	1.5300	1.4863	**2.86%**
ORKF-VB	2.5387	2.3715	**6.59%**	1.6578	1.6505	**0.44%**
HKF	3.3222	2.4812	**25.31%**	2.1451	1.8162	**15.33%**
MCKF	6.0876	4.4073	**27.60%**	2.7448	2.1748	**20.77%**
RSTKF	1.9395	2.3378	**-20.54%**	1.6000	1.6249	**-1.56%**
SSMKF	3.9826	3.3905	**14.87%**	2.5259	2.2864	**9.48%**
RECKF	1.7756	1.9276	**-8.56%**	1.5888	1.6015	**-0.80%**

m performs better than the existing RSTKF and ORKF-VB, and the position estimation accuracy is improved by 30.06% and 4.16% with respect to ORKF-VB, 46.55% and 25.93% with respect to HKF, 70.83% and 42.12% with respect to MCKF, 8.45% and 0.70% with respect to RSTKF, together with 55.42% and 37.10% with respect to SSMKF, respectively. It can also be seen from Table IV that HKF and MCKF performs relatively poorly, because these two algorithms do not take the statistics of the state vector into account. Additionally, it is worthy noting that although the accuracy of the proposed algorithm in velocity estimation is not outstanding compared to ORKF-VB and RSTKF, the running times are reduced by 17.00% and 88.53%, respectively, which is in concert with our previous analysis on the computational complexity. Besides, the caveat is that the consistency[5] of the proposed algorithm is also the best except for KFTNCM.

2) Multidimensional outliers verification: Next, the simulation results in the case of multidimensional outliers, i.e., the measurement outliers have different effects on different dimensions, are shown in Table V, where AP-1 denotes the AP in Table V, and AP-2 denotes the AP in the case of multidimensional outliers, and PgAP denotes the percentage gain of AP, and AV-1, AV-2 and PgAV are similarly defined. For simplicity, these effects are simulated by only contaminating the first dimension of the measurement with small outliers, where the definition of 'small' can be found in Table III.

From the PgAP and PgAV in Table V, it is shown that both HKF and MCKF have significant improvements in the case of multidimensional outliers because the robust M-estimate based KFs take into account the fact that the outliers may have different influence on different dimensions, which results in their advantage in the case of multidimensional outliers. However, the actual estimation accuracy (AP-2 and AV-2) of the robust M-estimate based KFs is limited by their inherent defect that the randomness of the estimate is ignored. Note that the proposed RECKF still has the best accuracy, which is due to its efficient iteration, as discussed in Section IV-B.

3) Robustness comparisons with VB-based methods: We can observe from Fig. 6 that VB-based methods suffer from instability occasionally. Therefore, the iterative process of a typical time segment of the ORKF-VB and the proposed RECKF will be singled out to analyze the potential instability of the VB-based methods. The $T(\theta) - \theta$ curves of two algorithms during this time segment are plotted in Fig. 7-d, where $T(\theta) = f(\theta) - \theta$. Note that the purpose of Fig. 7-d is to show how many local convergence points exist currently.

Fig. 7-a shows the position estimation errors of the two algorithms. It is obvious that the ORKF-VB has poor estimation performance during $91s - 93s$. Besides, Fig. 7-b plots the relationship of the convergent θ estimates among the RECKF, the ORKF-VB and true θ, In 7-b, the black dots denote the theoretical θs which is 0.001 for a measurement outlier and 1 for a normal measurement in this case. We can see from Fig. 7-b that the proposed RECKF can well capture the variation of θ during this time segment. However, the ORKF-VB misjudged the quality of the measurement at $91s$ and was cheated by this measurement outlier. We attribute this result to a large state uncertainty caused by consecutive measurement outliers at previous steps, i.e., $89 \sim 90s$. It is easy to imagine that the state uncertainty will increase dramatically if the consecutive measurements do not provide useful information, which is manifested by a large state error covariance matrix. As such, the algorithm will tend to trust the measurement at the next step at the first iteration regardless of a measurement outlier or a normal measurement, because the nominal MNCM is small relative to the large state error covariance matrix[6]. With this result, Fig. 7-c and 7-d can explain the reason why the ORKF-VB was cheated by the measurement outlier while the RECKF was not. Fig. 7-c gives the iteration values of two iterations and Fig. 7-d shows the $T(\theta) - \theta$ curves of the self-iteration and the coupling iteration at $91s$. As explained previously, both the first iteration values obtained by the self-iteration and the coupling iteration are close to 1 due to a large prior state uncertainty (For the self-iteration, refer to (32b)). Nevertheless, note from Fig. 7-c and 7-d that the RECKF converged unimpeded to a unique solution due to the monotonicity of the self-iteration, which is consistent with Section IV-A. However, the coupling-iteration was stuck at the local convergent point which is far away from the ideal convergent value considering that the scale factor is set as $U = 1000$ (i.e., the optimal θ is 0.001). This explains why the ORKF-VB performed worse than the RECKF at this segment. Further, the more serious the parameter coupling of the algorithm is, the more complex the coupling-iteration may be, and the algorithm may have more local optimal points, which is a disadvantage of VB-based methods.

More interestingly, the state uncertainty will be greatly reduced due to the incorrect measurement update, leading to the overconfidence of the algorithm so that it mistakenly

[5]An estimate is called consistent if its theoretical covariance is not smaller than the actual covariance, and then its NEES is not greater than the state dimension. The closer the NEES gets to the state dimension, the better consistency is achieved.

[6]The KF essentially balances the prior estimate and measurement by the magnitude of prior uncertainty and MNCM.

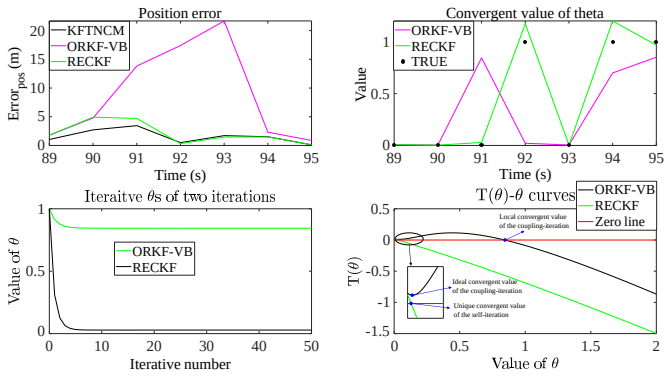
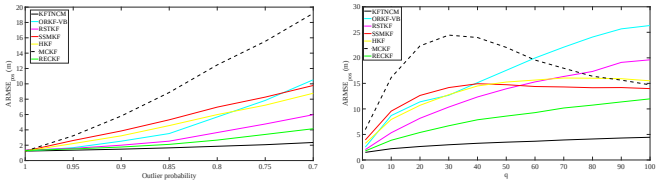


Fig. 7: RMSEs of the position and velocity for **Algorithm 1**.



(a) Position ARMSEs under different outlier probabilities.

(b) Position ARMSEs under different state noise covariance matrices.

Fig. 8: Position ARMSEs under different parameters.

regards the next good measurement (at $92s$) as an outlier (because the estimated state has deviated far from the true state), which further leads to the poor performance of the ORKF-VB. This will happen almost every time the outlier is incorrectly taken as a good measurement.

Overall, the proposed RECKF is expected to show its advantages in the case of large state uncertainty, which can be caused by consecutive measurement outliers (large outlier probability) or a large state noise covariance. To verify this point, the position ARMSEs of all algorithms under different outlier probabilities and different state noise covariance matrices are given in Fig. 8-a and 8-b. It is clear that estimation accuracy of the VB-based methods indeed deteriorate in the case of large outlier probability or large state uncertainty, which proves the previous conjecture that the estimation accuracy of VB-based methods will degrade when the measurement outlier probability or the state uncertainty is large.

B. Case 2: IMM-based RECKF

In this case, we focus on the overall performance of the existing filters in predetermined time-varying environments during five periods, which are listed in Table II. Two periods of five time steps were continuously interfered by large measurement outliers to simulate the anti-interference performance of the algorithm in the case of system failure. However, in order to be more consistent with the actual situation, these two periods are not included in the calculations of the ARMSEs and ANEESs. The comparisons are split into two parts, where one part compares the proposed IMM-based RECKF with some existing non-interactive algorithms, and the other part aims to compare the proposed IMM-based RECKF with the existing typical interactive algorithms. The GEC distribution,

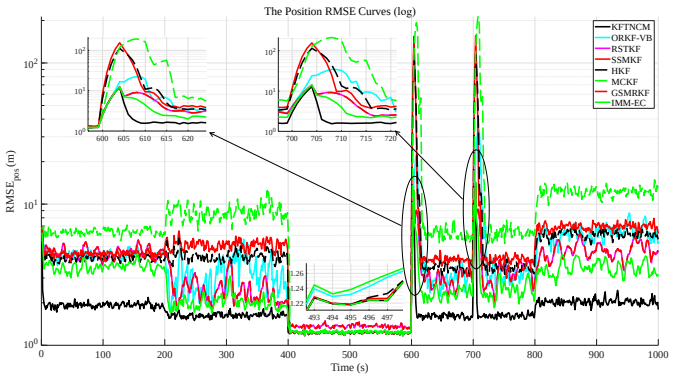


Fig. 9: Position RMSEs for case 2 in Part 1.

TABLE VI: ARMSEs of position (AP) and velocity (AV) and ANEESs of all filters for **Algorithm 2**.

Filters	AP(m)	AV(m/s)	ANEESs	Times (0.1ms)
KFTNCM	1.694	1.582	3.997	0.25
ORKF-VB	3.666	1.896	8.681	1.37
HKF	3.867	2.357	15.052	1.80
MCKF	7.013	3.299	39.211	1.58
RSTKF (GSMRKF)	3.249	1.769	6.320	10.53
SSMKF	4.383	2.733	17.529	1.18
RECKF	2.577	1.731	5.503	2.48

LEC distribution and EEC distribution are selected to implement the proposed IMM-based RECKF for the following reasons.

- The GEC distribution plays a dominant rule when there are no measurement outliers.
- The LEC distribution is used to deal with light measurement outliers and improve the robustness and accuracy of the system.
- The EEC distribution is an insurance in the case of system failure where large outliers may be induced.

1) Part 1: In this part, the performance of the proposed IMM-based RECKF (abbreviated as IMM-EC) is compared with the existing ORKF-VB, HKF, MCKF, RSTKF, SSMKF, KFTNCM, together with the GSMRKF [15]. Note that the position RMSE curves are shown in Fig. 9 in the logarithm form to be clearer, and all ARMSEs, ANEESs and execution time are listed in Table VI.

It can be seen from Fig. 9 that the proposed IMM-EC always has the best estimation accuracy except for Period 3. The IMM-EC is inferior to HKF and MCKF during this period, the underlying reason has been elaborated in Section VI. The performance of the RSTKF and the GSMRKF is similar, which is reasonable because the GSMRKF is essentially an extended version of the RSTKF to deal with skewed noises. Moreover, benefiting from the strong robustness of the EEC distribution, the proposed IMM-EC retains stability and accuracy in the case of consecutive large measurement outliers, which shows the usefulness of the EEC distribution.

Additionally, in terms of the overall performance without the large outliers during two segments, the estimation accuracy of position and velocity is improved by 29.71% and 8.70% with respect to ORKF-VB, 33.36% and 26.57% with

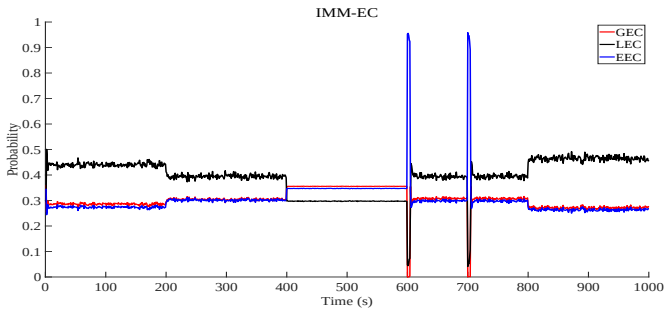


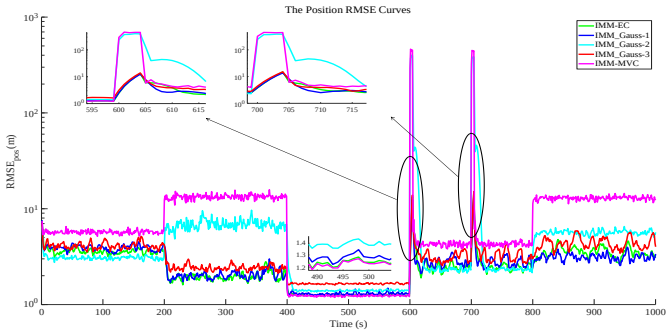
Fig. 10: The weights of three interactive RECKFs.

respect to HKF, 63.25% and 47.53% with respect to MCKF, 21.30% and 2.15% with respect to RSTKF and GSMRKF, together with 41.66% and 36.66% with respect to SSMKF, respectively. Besides, the consistency of the proposed IMM-based algorithm is also the best except for KFTNCM, which is shown in Table VI. It is worth noting that the number of iterations is set as 50 for all the comparison algorithms to guarantee the convergence of these algorithms except for the proposed IMM-based RECKF. The number of iterations of the proposed algorithm is set as 2 because with the number of iterations increasing, the proposed algorithm will bear a relatively huge computational burden due to the interactions among the pre-selected RECKFs. Fortunately, reducing the number of iterations still results in a satisfactory accuracy. Thus, the proposed algorithm can achieve a satisfactory result at the cost of a slightly larger computational complexity. The reason why the proposed algorithm has priority in such an intricate noise environment can be found in Fig. 10.

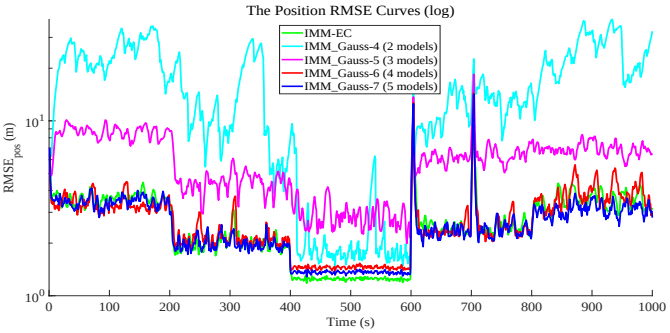
The weight curves in Fig. 10 denote the average weights of respective EC distributions-based RECKFs with respect to Monte Carlo runs. According to Table II, small outliers appear during Periods 1,2 and 4. Theoretically, the RECKF based on the LEC distribution performs best in this period, and as can be seen from Fig. 10, the RECKF based on the LEC distribution has the highest weight in these periods, and it turns out that the interaction works successfully, which indicates that the approximation error mentioned in **Remark 5** is acceptable. Note that, nonetheless, there are no outliers during Period 3, the weight of the RECKF based on the GEC distribution is not exactly 1, but the highest. Moreover, the high weight of the RECKF based on the EEC distribution during the in Period 4 verifies the validity of EEC distribution in the case of large outliers.

2) Part 2: In this part, the proposed IMM-EC is compared with the existing Gaussian IMM filter with different model numbers and different model parameters and the Student't IMM filter, represented by the IMM-MVC algorithm [28]. The parameter settings in this part are listed in Table VII, where M denotes the magnification of the MNCM used in each model relative to nominal MNCM. Note that the parameters chosen by IMM-Gauss-1, 2 and 3 are accurate, adventurous and conservative, and the parameters chosen by IMM-Gauss-4, 5, 6 and 7 are all conservative. The results are shown in Fig. 11-a and 11-b.

Notice from Fig. 11 that the IMM-MVC has poor performance in our simulation because it is specially designed for simultaneous state and measurement outliers. In the IMM-MVC, the state noise and the measurement noise require to be contaminated by the same degree of outliers as mentioned in Section I-B. However, only measurement outliers exist in our simulation, leading to the poor performance of the IMM-MVC.



(a) Comparison with the IMM-MVC and IMM-Gauss with different parameters.



(b) Comparison with the IMM-Gauss with different model numbers.

Fig. 11: Comparison with the IMM-MVC and IMM-Gauss.

Besides, we can observe from Fig. 11-a that the IMM-Gauss-1 with accurate parameters has the best performance overall. The IMM-Gauss-2 with adventurous parameters performs well when the model parameters are close to the true parameters, i.e., Periods 1 and 4. Nonetheless, it has poor performance during Periods 2, 4 and 5 where the model parameters are much smaller than the true parameters. On the other hand, the IMM-Gauss-3 does not have large errors like IMM-Gauss-2, which benefits from its conservative parameter settings, but this also leads to its poor accuracy as a whole.

Next, some discussions are conducted to verify the influence of the model number on the Gaussian IMM filter. Conservative parameters are set for all four Gaussian IMM filters. We can observe from Fig. 11-b that the Gaussian IMM filter cannot obtain an equivalent accuracy under the same model number with the proposed IMM-EC. It is also worth noting that the Gaussian IMM filter with four or more models will consume more computation than the IMM-EC. In conclusion, the proposed IMM-EC has better performance than the existing Gaussian IMM filter and existing Student's t IMM filter in the case of complex environments.

It is also worth noting that the advantage of the EEC distribution not only lies in its ability to resist extremely large measurement outliers, but also helps in an outlier-free

TABLE VII: Parameter settings of all filters in case 2.

Filters		Parameter settings
IMM-Gauss (different parameters)	IMM-Gauss-1	M_{11}=1, M_{12}=500,M_{13}=100000
	IMM-Gauss-2	M_{21}=0.1, M_{22}=10,M_{23}=50
	IMM-Gauss-3	M_{31}=10, M_{32}=3000,M_{33}=1000000
IMM-Gauss (different model numbers)	IMM-Gauss-4	M_{41}=0.1, M_{42}=1000000
	IMM-Gauss-5	M_{51}=0.1, M_{52}=3000, M_{52}=1000000
	IMM-Gauss-6	M_{61}=0.1, M_{62}=20 M_{63}=4500, M_{64}=1000000
	IMM-Gauss-7	M_{71}=0.1, M_{72}=10,M_{73}=750 M_{74}=50000, M_{75}=1000000
IMM-MVC		dof1=3, dof2=10, dof3=100

case. Due to the similarity of influence functions of the EEC distribution and the Gaussian distribution, they have similar performance when there are no measurement outliers, as can be seen from the similar weights of the EEC curve and the GEC curve during $400 \sim 600s$ in Fig. 10. Thus, the EEC distribution can also improve the performance of the proposed algorithm in an outlier-free scenario.

VII. Conclusion

In this paper, we firstly proposed a novel outlier-robust Kalman filtering framework with measurement noise modeled as EC distributed. Then some properties of the proposed algorithm are specified. Additionally, a method using IMM for adaptively selecting EC distributions to accommodate time-varying and intricate noise environment is given. Simulation results have demonstrated that the proposed non-interactive RECKF has slightly better accuracy but less computations than the existing algorithms, and the proposed IMM-based RECKF has a significant accuracy advantage but slightly heavier computational burden than the existing algorithms in the case of time-varying and intricate noise environments.

VIII. Appendices

A. Proof of Theorem 1

The conditional measurement likelihood PDF can be calculated by employing the Gaussian integral formula [33] as

$$p_{\theta_k}(\mathbf{z}_k|\mathbf{z}_{1:k-1},\lambda_k)=\mathrm{N}(\mathbf{z}_k;\hat{\mathbf{z}}_{k|k-1},\mathbf{P}^{\mathbf{zz}}_{k|k-1}(\theta_k)/\lambda_k) \tag{46}$$

where $\hat{\mathbf{z}}_{k|k-1}$ and $\mathbf{P}^{\mathbf{zz}}_{k|k-1}(\theta_k)$ are calculated as (14).

Therefore, using (7), (10) and (46), the modified posterior PDF of the state vector can be analytically calculated as (14).

B. Proof of Proposition 1

Using Bayes' rule, we can easily attain

$$\begin{aligned}&p_(\mathbf{z}_k|\mathbf{x}_k,\lambda_k)p_{\theta_k}(\mathbf{x}_k|\mathbf{z}_{1:k-1},\lambda_k)\\&=p_{\theta_k}(\mathbf{z}_k|\mathbf{z}_{1:k-1},\lambda_k)p_{\theta_k}(\mathbf{x}_k|\mathbf{z}_{1:k},\lambda_k)\end{aligned} \tag{47}$$

Utilizing (7), (10), (14) and (46), the modified posterior PDF can be formulated as

$$\begin{aligned}&p_{\theta_k}(\mathbf{x}_k,\lambda_k|\mathbf{z}_{1:k})\\&=\frac{\mathrm{N}(\mathbf{x}_k;\hat{\mathbf{x}}'_{k|k}(\theta_k),\frac{\mathbf{P}'_{k|k}(\theta_k)}{\lambda_k})\mathrm{N}(\mathbf{z}_k;\hat{\mathbf{z}}_{k|k-1},\frac{\mathbf{P}^{\mathbf{zz}}_{k|k-1}(\theta_k)}{\lambda_k})P_g(\lambda_k)}{p_{\theta_k}(\mathbf{z}_k|\mathbf{z}_{1:k-1})}\end{aligned} \tag{48}$$

By integrating $\mathbf{x}_k$ into (48), the marginal posterior PDF of λ_k can be calculated as (15).

C. Proof of Proposition 3

Taking the derivative of both (12) and (21) with respect to $\mathbf{z}_k$ and using (18), we can obtain

$$\begin{aligned}&-\mathrm{E}(\lambda_k)\underline{(\mathbf{P}^{\mathbf{zz}}_{k|k-1}(\theta_k))^{-1}(\mathbf{z}_k-\hat{\mathbf{z}}_{k|k-1})}\\&=2\dot{g}(\Delta_k)\underline{(\mathbf{P}^{\mathbf{zz}}_{k|k-1}(\theta_k))^{-1}(\mathbf{z}_k-\hat{\mathbf{z}}_{k|k-1})}\end{aligned} \tag{49}$$

where $\Delta_k=(\mathbf{z}_k-\hat{\mathbf{z}}_{k|k-1})^T(\mathbf{P}^{\mathbf{zz}}_{k|k-1}(\theta_k))^{-1}(\mathbf{z}_k-\hat{\mathbf{z}}_{k|k-1})$. The underlined expression is an m by 1 non-zero vector and the first term on the left-hand side is the expectation of λ_k. Therefore, using (20), (22) can be obtained.

D. Proof of Theorem 2

It is observed from (14) and (23) that both the matrices $\mathbf{P}^{\mathbf{zz}}_{k|k-1}(\theta_k)$ and $\tilde{\mathbf{P}}^{\mathbf{zz}}_{k|k-1}$ are positive semi-definite when $\theta_k>0$. Also, it is observed from ℓ_2 that for all $t\in[0,+\infty)$, $\dot{g}(t)<0$, which causes $\theta^{(i)}_k>0$ for arbitrary i. Moreover, under condition ℓ_3, for all $t\in[0,+\infty)$, $\dot{g}(t)>\dot{g}(0)$, thus $-2\dot{g}(t)<-2\dot{g}(0)$, so $\theta^{(i)}_k<-2\dot{g}(0)$, therefore, the iterative sequence $\{\theta^{(i)}_k\}^{+\infty}_{i=1}$ is bounded, i.e., $0<\theta^{(i)}_k<-2\dot{g}(0)$. Then, we will study the monotonicity of the iterative sequence.

Define $f(\theta_k)=-2\dot{g}(tr\{\tilde{\mathbf{P}}^{\mathbf{zz}}_{k|k-1}(\mathbf{P}^{\mathbf{zz}}_{k|k-1}(\theta_k))^{-1}\})$. Then taking the derivative of $f(\theta_k)$ with respect to θ_k, we can attain

$$\begin{aligned}&\dot{f}(\theta_k)=2\ddot{g}(\mathrm{tr}\{(\tilde{\mathbf{P}}^{\mathbf{zz}}_{k|k-1}(\mathbf{P}^{\mathbf{zz}}_{k|k-1}(\theta_k))^{-1})\}\tilde{\mathbf{P}}^{\mathbf{zz}}_{k|k-1}\\&\times(\mathbf{P}^{\mathbf{zz}}_{k|k-1}(\theta_k))^{-1}\left(\mathbf{H}_k\mathbf{P}_{k|k-1}\mathbf{H}_k^T\right)(\mathbf{P}^{\mathbf{zz}}_{k|k-1}(\theta_k))^{-1}\end{aligned} \tag{50}$$

Since the matrices $\tilde{\mathbf{P}}^{\mathbf{zz}}_{k|k-1},\mathbf{H}_k\mathbf{P}_{k|k-1}\mathbf{H}_k^T,\mathbf{P}^{\mathbf{zz}}_{k|k-1}(\theta_k)$ are all positive semi-definite symmetric matrices and $\ddot{g}(t)>0$, we can obtain $\dot{f}(\theta_k)>0$, which means that $f(\theta_k)$ is a monotonically increasing function. That is to say, if $\theta^{(1)}_k>\theta^{(0)}_k$, the iteration converges to the upper bound and vice versa. As a result, the iterative sequence $\{\theta^{(i)}_k\}^{+\infty}_{i=1}$ is able to converge to a unique limit, i.e.,

$$\lim_{i\to+\infty}\theta^{(i)}_k=\hat{\theta}_k \tag{51}$$

E. Proof of Theorem 3

Employing (24) and (26), we can obtain

$$\theta^{(1)}_k=-2\dot{g}(\mathrm{tr}\{\tilde{\mathbf{P}}^{\mathbf{zz}}_{k|k-1}\left(\mathbf{P}^{\mathbf{zz}}_{k|k-1}(\theta^{(0)}_k)\right)^{-1}\}) \tag{52}$$

Using the condition ℓ_3 and **Assumption 2**, we can obtain $\theta^{(1)}_k<\theta^{(0)}_k=1$. If $\theta^{(i+1)}_k<\theta^{(i)}_k$ holds when $i=j$, i.e., $\theta^{(j+1)}_k<\theta^{(j)}_k$. Then we can obtain

$$\mathrm{tr}\{\tilde{\mathbf{P}}^{\mathbf{zz}}_{k|k-1}(\mathbf{P}^{\mathbf{zz}}_{k|k-1}(\theta^{(j+1)}_k))^{-1}\}>\mathrm{tr}\{\tilde{\mathbf{P}}^{\mathbf{zz}}_{k|k-1}(\mathbf{P}^{\mathbf{zz}}_{k|k-1}(\theta^{(j)}_k))^{-1}\} \tag{53}$$

Then, exploiting (24), (53) and condition ℓ_3, we can obtain $\theta^{(j+2)}_k<\theta^{(j+1)}_k$. According to the mathematical induction method, $\theta^{(i+1)}_k<\theta^{(i)}_k$ holds for an arbitrary $i\geq 0$. Thus, the inequality (30) holds for an arbitrary $i\geq 0$.

F. Proof of Proposition 5

The variance of λ_k is calculated as

$$Var(\lambda_k)=\int_0^{+\infty}\lambda_k^2 p_{\theta_k}(\lambda_k|\mathbf{z}_{1:k})d\lambda_k-[\mathrm{E}(\lambda_k)]^2 \tag{54}$$

Define

$$\tilde{\mathbf{P}}_1(\theta_k) = (\mathbf{P}^{\mathbf{zz}}_{k|k-1}(\theta_k))^{-1}\tilde{\mathbf{P}}^{\mathbf{zz}}_{k|k-1}(\mathbf{P}^{\mathbf{zz}}_{k|k-1}(\theta_k))^{-1} \quad (55)$$

$$\tilde{\mathbf{P}}_2(\theta_k) = (\mathbf{z}_k - \hat{\mathbf{z}}_{k|k-1})^T(\mathbf{P}^{\mathbf{zz}}_{k|k-1}(\theta_k))^{-1}(\mathbf{z}_k - \hat{\mathbf{z}}_{k|k-1}) \quad (56)$$

Similarly, taking the derivatives of (12) and (21) with respect to $\mathbf{z}_k^T$ and using (54), we can obtain

$$\begin{aligned}&(Var(\lambda_k) + [\mathrm{E}(\lambda_k)]^2)\tilde{\mathbf{P}}_1(\theta_k) - (\mathrm{E}(\lambda_k))(\mathbf{P}^{\mathbf{zz}}_{k|k-1}(\theta_k))^{-1}\\ &= 4\dot{g}^2(\tilde{\mathbf{P}}_2(\theta_k))\tilde{\mathbf{P}}_1(\theta_k) + 4\ddot{g}(\tilde{\mathbf{P}}_2(\theta_k))\tilde{\mathbf{P}}_1(\theta_k)\\ &+ 2\dot{g}(\tilde{\mathbf{P}}_2(\theta_k))(\mathbf{P}^{\mathbf{zz}}_{k|k-1}(\theta_k))^{-1} \quad (57)\end{aligned}$$

Utilizing (18) and (22), (57) can be reformulated as follows

$$Var(\lambda_k)\tilde{\mathbf{P}}_1(\theta_k) = 4\ddot{g}(\tilde{\mathbf{P}}_2(\theta_k))\tilde{\mathbf{P}}_1(\theta_k) \quad (58)$$

where $\tilde{\mathbf{P}}_1(\theta_k)$ is an $m \times m$ non-zero matrix. Therefore, (58) can transform into $Var(\lambda_k) = 4\ddot{g}(\tilde{\mathbf{P}}_2(\theta_k))$. Employing **Theorem 3** and condition ℓ_4, we can obtain (33).

REFERENCES

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

[2] G. Agamennoni, J. I. Nieto and E. M. Nebot, "Approximate Inference in State-Space Models With Heavy-Tailed Noise," *IEEE Transactions on Signal Processing*, vol. 60, no. 10, pp. 5024–5037, Oct. 2012.

[3] C. D. Karlgaard, "Robust rendezvous navigation in elliptical orbit", *Journal of Guidance Control & Dynamics,* vol. 29, no. 2, pp. 495–499, 2006.

[4] C. D. Karlgaard, and H. Schaub, "Huber-based divided difference filtering", *Journal of Guidance Control & Dynamics,* vol. 30, no. 3, pp. 885–891, 2007.

[5] M. A. Gandhi and L. Mili, "Robust Kalman Filter Based on a Generalized Maximum-Likelihood-Type Estimator," *IEEE Transactions on Signal Processing*, vol. 58, no. 5, pp. 2509–2520, May 2010.

[6] L. Chang, B. Hu, G. Chang, and A. Li, "Multiple outliers suppression derivative-free filter based on unscented transformation", *Journal of Guidance Control & Dynamics,* vol. 35, no. 6, pp. 1902–1907, 2015.

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

[8] B. Chen, L. Dang, Y. Gu, N. Zheng and J. C. Príncipe, "Minimum Error Entropy Kalman Filter," *IEEE Transactions on Systems, Man, and Cybernetics: Systems,* vol. 51, no. 9, pp. 5819–5829, 2021.

[9] Y. Huang, Y. Zhang, Y. Zhao, P. Shi and J. Chambers, "A Novel Outlier-Robust Kalman Filtering Framework based on Statistical Similarity Measure," *IEEE Transactions on Automatic Control,* vol. 66, no. 6, pp. 2677–2692, 2021.

[10] R. Piché, S. Särkkä and J. Hartikainen, "Recursive outlier-robust filtering and smoothing for nonlinear systems using the multivariate student-t distribution," in *2012 IEEE International Workshop on Machine Learning for Signal Processing,* 2012, pp. 1–6.

[11] X. Rong Li and V. P. Jilkov, "Survey of maneuvering target tracking. Part I. Dynamic models," *IEEE Transactions on Aerospace and Electronic Systems,* vol. 39, no. 4, pp. 1333–1364, Oct. 2003.

[12] 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, June 2017.

[13] Y. Huang, Y. Zhang, Y. Zhao and J. A. Chambers, "A Novel Robust Gaussian-Student's t Mixture Distribution Based Kalman Filter," *IEEE Transactions on Signal Processing,* vol. 67, no. 13, pp. 3606–3620, July 2019.

[14] Y. Huang, Y. Zhang and J. A. Chambers, "A Novel Kullback-Leibler Divergence Minimization-Based Adaptive Student's t-Filter," *IEEE Transactions on Signal Processing,* vol. 67, no. 20, pp. 5417–5432, Oct. 2019.

[15] 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.

[16] G. Zhang, J. Lan, Le Zhang, F. He and S. Li, "Filtering in Pairwise Markov Model With Student's t Non-Stationary Noise With Application to Target Tracking," *IEEE Transactions on Signal Processing,* vol. 69, pp. 1627–1641, 2021.

[17] G. Wang, C. Yang and X. Ma, "A novel robust nonlinear Kalman filter based on multivariate Laplace distribution," *IEEE Transactions on Circuits and Systems II: Express Briefs,* vol. 68, no. 7, pp. 2705–2709, 2021.

[18] J. Neri, P. Depalle and R. Badea, "Laplace State Space Filter with Exact Inference and Moment Matching," in *ICASSP 2020 - 2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP),* 2020, pp. 5880–5884.

[19] J. Neri, P. Depalle and R. Badeau, "Approximate Inference and Learning of State Space Models with Laplace Noise," *IEEE Transactions on Signal Processing,* vol. 69, pp. 3176–3189, 2021.

[20] H. Zhu, G. Zhang, Y. Li, H. Leung, "A novel robust Kalman filter with unknown non-stationary heavy-tailed noise", *Automatica,* vol. 127, 2021.

[21] H. Zhu, G. Zhang, Y. Li and H. Leung, "An adaptive Kalman filter with inaccurate noise covariances in the presence of outliers," *IEEE Transactions on Automatic Control,* to be published, doi: 10.1109/TAC.2021.3056343, 2021.

[22] S. Cambanis, S. Huang, and G. Simons, "On the theory of elliptically contoured distributions", *Journal of Multivariate Analysis,* vol. 11, no. 3, pp. 368–385, Sep. 1981.

[23] D. Alspach and H. Sorenson, "Nonlinear Bayesian estimation using Gaussian sum approximations," *IEEE Transactions on Automatic Control*, vol. 17, no. 4, pp. 439-448, 1972.

[24] K. P. Murphy, "Switching Kalman Filters", *International Journal of Control*, 1999.

[25] I. Bilik and J. Tabrikian, "MMSE-Based Filtering in Presence of Non-Gaussian System and Measurement Noise," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 46, no. 3, pp. 1153-1170, Jul. 2010.

[26] H. A. P. Blom and Y. Bar-Shalom, "The interacting multiple model algorithm for systems with Markovian switching coefficients," *IEEE Transactions on Automatic Control*, vol. 33, no. 8, pp. 780-783, Aug. 1988.

[27] Q. Li, Y. Ben, and J. Tan et al., "Robust selection of the degrees of freedom in the Student's t distribution through Multiple Model Adaptive Estimation," *Signal Processing*, vol. 153, pp. 255-265, 2018.

[28] C. Shen, and L. Mihaylova, "A flexible robust Student's t-based multi-model approach with maximum Versoria criterion," *Signal Processing*, vol. 182, May. 2021.

[29] C. Xue, Y. Huang et al., "Robust Elliptically Contoured Distribution Based Kalman Filter for INS/DVL Integrated Navigation," in *Proceedings of the 40th Chinese Control Conference*, 2021, pp. 3570–3575.

[30] K. T. Fang, S. Kotz, and K. W. Ng, Symmetric Multivariate and Related Distributions. *Chapman and Hall/CRC,* Nov. 1989.

[31] T. W. Anderson, An Introduction to Multivariate Statistical Analysis, *3rd ed. Wiley-Interscience,* 2003.

[32] K. C, "Estimation and decision for linear systems with elliptical random processes," *IEEE Transactions on Automatic Control*, vol. 18, no. 5, pp. 499–505, Oct. 1973.

[33] B. N. Vo and W. K. Ma, "The Gaussian Mixture Probability Hypothesis Density Filter," *IEEE Transactions on Signal Processing*, vol. 54, no. 11, pp. 4091–4104, Nov. 2006.

[34] Y. Bar-Shalom, R. X. Li, and T. Kirubarajan, "Estimation with Applications to Tracking and Navigation, Theory Algorithms and Software", *John Wiley & Sons,* 2001.

[35] X. Boyen and D. Koller. Tractable inference for complex stochastic processes. arXiv preprint arXiv: 1301.7362, 2013.

[36] J. Wang, T. Zhang, B. Jin, Y. Zhu and J. Tong, "Student's t-Based Robust Kalman Filter for a SINS/USBL Integration Navigation Strategy," *IEEE Sensors Journal*, vol. 20, no. 10, pp. 5540–5553, May. 2020.