---

**Article:**
Zhu, F., Huang, Y., Xue, C. et al. (2 more authors) (2022) A sliding window variational outlier-robust Kalman filter based on student's t noise modelling. IEEE Transactions on Aerospace and Electronic Systems, 58 (5). pp. 4835-4849. ISSN 0018-9251

https://doi.org/10.1109/TAES.2022.3164012

---

# A Sliding Window Variational Outlier-Robust Kalman Filter based on Student's t Noise Modelling

Fengchi Zhu, Yulong Huang, *Member, IEEE,* Chao Xue, Lyudmila Mihaylova, *Senior Member, IEEE,* Jonathon Chambers, *Fellow, IEEE*

***Abstract*—Existing robust state estimation methods are generally unable to distinguish model uncertainties (state outliers) from measurement outliers as they only exploit the current measurement. In this paper, the measurements in a sliding window are therefore utilized to better distinguish them, and an adaptive method is embedded, leading to a sliding window variational outlier-robust Kalman filter based on Student's t noise modelling. Target tracking simulations and experiments show that the tracking accuracy and consistency of the proposed filter are superior to those of the existing state-of-the-art outlier-robust methods thanks to the improved ability to identify the outliers but at a cost of greater computational burden.**

***Index Terms*—Outlier-robust Kalman filter, outlier identification, sliding window, Student's t distribution, variational Bayesian, target tracking**

## I. Introduction

### A. Background

State estimation is of fundamental importance for many engineering applications such as navigation [1], target tracking [2], and power systems [3]. As one of the most common methods of state estimation, the Kalman filter (KF) can achieve a minimum mean square error (MMSE) estimate of the state in a linear state-space model under the assumption of Gaussian noises [4, pp. 130]. However, in some applications, such as target tracking, the model uncertainties (state outliers[1]) and measurement outliers may be induced by target manoeuvres and sensor failures, respectively [2]. The outliers lead to heavier tails in the probability density function (PDF) of the noise than in the Gaussian PDF, which violates the Gaussian hypothesis inherent in the conventional KF and thereby results in poor estimation performance.

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

F. Zhu, Y. Huang and C. Xue 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, Harbin 150001, China (e-mail: zfchiggins@163.com; heuedu@163.com; xc_adjuster@163.com).

L. Mihaylova is with the Department of Automatic Control and Systems Engineering, The University of Sheffield, Sheffield S10 2TG, U.K. (e-mail: L.S.Mihaylova@sheffield.ac.uk).

J. Chambers is with the College of Intelligent Systems Science and Engineering, Harbin Engineering University, Harbin 150001, China, and also with the School of Engineering, University of Leicester, Leicester LE1 7RH, Leicestershire, U.K. (e-mail: Jonathon.Chambers@le.ac.uk).

[1]The model uncertainties would lead to abnormally large state noises, which are also regarded as a form of outlier in [11], [28]. We call these state outliers in this paper corresponding to the measurement outliers, as in [16].

### B. Relevant works

To tackle simultaneously the state outliers and measurement outliers, some robust smoothers and filters have been proposed. The existing robust smoothers can be broadly classified into optimization-based methods [5]–[6] and Bayesian methods [7]–[9]. However, the smoothers cannot be carried out online since the computational burden increases gradually over time. For the robust filters, some popular approaches, classified as M-estimation-based KFs, aim to resist outliers by setting appropriate robust cost functions [10]–[14], such as the Huber-based KF (HKF) [10] and maximum correntropy KF (MCKF) [12]. Nevertheless, the ignorance of the randomness inherent in the state vector limits the estimation accuracy of these methods [15]. To overcome this shortcoming, robust KFs based on statistical similarity measure (SSMKFs) have been proposed in [16]–[17], where the state posterior PDF is calculated by maximizing the defined statistical similarity measure.

Different from the M-estimation-based KFs and SSMKFs, many robust filters model the non-Gaussian heavy-tailed noise with appropriate distributions to achieve robustness. By selecting the PDFs of state and measurement noises as heavy-tailed distributions, the particle filter (PF) [18]–[19] can produce robust state estimates. Nonetheless, the PF is not suitable for addressing the case of unknown noise distribution caused by state and measurement outliers, and it has heavy computational burden. On the other hand, by approximating the PDFs of the state and noises as Student's t, some Student's t filters have been proposed [20]–[22]. However, their accuracy is limited when the noises are slightly or moderately heavy-tailed distributed [15] or when the fixed scale matrices are improperly set in these filters. Beside the traditional counterparts, many variational-learning-based robust KFs model the PDFs of the state or noises as heavy-tailed distributions and learn the parameters of the heavy tails adaptively. For example, the robust Student's t-based KF (RSTKF) proposed in [15] models the one-step prediction PDF and measurement likelihood PDF as Student's t and obtains a Gaussian approximation of the state posterior PDF by employing the variational Bayesian (VB) method. Subsequent studies have extended the RSTKF to the cases of skewed or non-stationary heavy-tailed noises [23]–[26]. The current robust smoothers and filters against state and measurement outliers are compared and summarized in detail in Table I.

The existing M-estimation-based KF [12], SSMKF [16] and variational-learning-based robust KFs [15], [23]–[26] adjust the one-step prediction error covariance matrix (PECM) or

TABLE I: Outlines of different algorithms.

| Categories | Algorithms | Main ideas | Advantages | Deficiencies (Compared with the proposed SWRKF) |
|---|---|---|---|---|
| M-estimation-based KFs [10]–[14] | HKF [10] | Minimizing the Huber function | Enable the outliers that occur in partial dimensions to be tackled [17] | 1) A; 2) B; 3) Ignorance of the randomness inherent in the state vector |
| | MCKF [12] | Maximizing the correntropy | | |
| Student's t filters [20]–[22] | [20] | Student's t modelling and joint distribution approximation | Light computational burden | 1) A; 2) B;<br>2) Limited accuracy under slightly or moderately heavy-tailed noises [15] |
| | [21] | 1) Student's t modelling and joint distribution approximation<br>2) Unscented transformation | 1) Light computational burden<br>2) Enable the nonlinear system to be tackled | |
| SSMKFs [16]–[17] | SSMKF [16] | Maximizing the statistical similarity measure | C | 1) A; 2) B |
| | [17] | Maximizing the multiple statistical similarity measure | 1) C; 2) Enable the outliers that occur in partial dimensions to be tackled | |
| Variational-learning-based KFs [15], [23]–[27] | RSTKF [15] | Student's t noise modelling | Obtaining the outlier robustness with Student's t noise modelling based on variational learning | |
| | [23] | Gaussian-Student's t mixture distribution noise modelling | Enable the intermittent outliers to be tackled | |
| | [24] | Gaussian scale mixture distribution noise modelling | Enable the heavy-tailed and skewed noises to be tackled | |
| | [25] | Two Gaussian mixture distributions noise modelling | 1) Enable the intermittent outliers to be tackled<br>2) Enable the covariance matrices in two Gaussian mixture distributions to be estimated | 1) A;<br>2) Limited estimation accuracy of covariance matrices since the difficulty in distinguishing the outliers |
| | [26] | Gaussian-Gamma mixture distribution noise modelling | Enable the covariance matrices in Gaussian-Gamma mixture distributions to be estimated | |
| | SWVAKF [27] | 1) Gaussian noise modelling<br>2) Noise covariance matrices estimation using a sliding window | Enable the inaccurate slowly time-varying noise covariance matrices to be estimated | No particular resistance to outliers due to the same noise covariance matrices in the sliding window |
| | Proposed SWRKF | 1) Student's t noise modelling<br>2) Outliers identification using a sliding window | 1) Better outlier identification performance<br>2) Enable time-varying noise scale matrices to be estimated adaptively | Relative increase in computational burden compared with existing outlier-robust filters |
| Robust smoothers [5]–[9] | [5] | 1) Student's t noise modelling<br>2) Convex composite extension of the Gauss-Newton method | D | 1) Lack of local convergence guarantee<br>2) Limited to stationary noise because of the fixed noise parameters<br>3) Off-line implementation |
| | RGAS [8] | 1) Student's t noise modelling<br>2) variational learning | D | 1) Limited to stationary noise because of the fixed noise parameters<br>2) Off-line implementation |
| | [9] | 1) Generalized normal scale mixture distribution noise modelling<br>2) variational learning | 1) D<br>2) Enable the heavy-tailed and skewed noises to be tackled | |

Notations: A: Difficulty in distinguishing state outliers from measurement outliers; B: Sensitive to the nominal noise covariance matrices or noise scale matrices; C: Use of the randomness inherent in the state vector; D: Better outlier identification performance due to the use of all measurements.

state noise covariance matrix (SNCM) and measurement noise covariance matrix (MNCM) to resist the state outlier and measurement outlier, respectively, by using the measurement at the current step. However, the measurement at the current step alone does not have enough information to distinguish between the previous mentioned two sorts of outliers, which causes the inappropriate adjustment of the PECM or SNCM and MNCM. To solve this problem, we propose to exploit multiple measurements in a sliding window to better identify the two kinds of outliers, which will be explained in detail in Section II-B and II-C.

### C. Contributions and organization of this paper

In this paper, a sliding window variational outlier-robust Kalman filter based on Student's t noise modelling (SWRKF) is proposed. Compared with the preceding literature, the contributions of this paper are as follows.

- We propose to utilize the measurements over a period of time to better distinguish between the state and measurement outliers at previous steps, and implement this idea using the VB method.
- The sliding window adaptive method for estimating Gaussian covariance matrices [27] is embedded to achieve more accurate parameters of the Student's t distributions.
- Target tracking simulations and experiments demonstrate that the proposed filter has more superior tracking accuracy and consistency compared with the existing robust methods at a cost of greater computational burden.

The rest of this paper is organized as follows. In Section II, the issues with the existing methods are detailed to better motivate this paper. In Section III, the SWRKF is derived, and the outlier suppression mechanism and computational complexity are analyzed. In Section IV, the target tracking simulations are carried out to compare the proposed algorithm with the existing methods from many aspects. In Section V, two manoeuvring vehicle tracking experiments are performed to verify the practicality of the proposed algorithm in engineering. Section VI concludes this paper.

*Notation:* $\mathrm{N}(\cdot;\boldsymbol{\mu},\boldsymbol{\Sigma})$ denotes the normal distribution with mean vector $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$; $\mathrm{St}(\cdot;\boldsymbol{\mu},\boldsymbol{\Sigma},\nu)$ denotes the Student's t distribution with mean vector $\boldsymbol{\mu}$, scale matrix $\boldsymbol{\Sigma}$ and degree of freedom (dof) parameter $\nu$; $\mathrm{G}(\cdot;\alpha,\beta)$ denotes the Gamma distribution with shape parameter $\alpha$ and rate pa-

rameter $\beta$; $\mathrm{IW}(\cdot; u, \mathbf{U})$ denotes the inverse Wishart distribution with dof parameter $u$ and scale matrix $\mathbf{U}$; $\mathbf{z}_{i:j}$ represents the set made up of the measurements from step $i$ to $j$; $\mathrm{tr}(\cdot)$ and $\mathrm{E}[\cdot]$ denote the trace operation and mathematical expectation operation, respectively; the superscript $(i)$ represents the $i$-th iteration in the variational inference; and $\|\cdot\|$ denotes the Euclidean norm.

## II. Preliminaries

### A. State-space model

Generally, the state estimation problem with outlier interference in engineering applications is formulated as the following discrete-time linear state-space model

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

where $\mathbf{x}_k \in \mathbb{R}^n$ is the state vector to be estimated, $\mathbf{z}_k \in \mathbb{R}^m$ is the measurement vector, $\mathbf{F}_k$ and $\mathbf{H}_k$ are, respectively, the state transition and observation matrices, $\mathbf{w}_k$ and $\mathbf{v}_k$ are, respectively, the outlier-contaminated heavy-tailed state and measurement noises with zero mean and inaccurate nominal covariance matrix $\mathbf{\Sigma}_w$ and $\mathbf{\Sigma}_v$. It is assumed that the initial state $\mathbf{x}_0$ is a Gaussian vector of mean value $\hat{\mathbf{x}}_{0|0}$ and covariance matrix $\mathbf{P}_{0|0}$, and $\mathbf{x}_0$, $\mathbf{w}_k$ and $\mathbf{v}_k$ are mutually independent.

### B. Brief review of the existing methods

As a well-known method of state estimation, the conventional KF comprises time update and measurement update as follows [4].

*Time update:*

$$\hat{\mathbf{x}}_{k|k-1} = \mathbf{F}_k \hat{\mathbf{x}}_{k-1|k-1} \tag{2}$$

$$\mathbf{P}_{k|k-1} = \mathbf{F}_k \mathbf{P}_{k-1|k-1} \mathbf{F}_k^{\mathrm{T}} + \mathbf{\Sigma}_w \tag{3}$$

*Measurement update:*

$$\mathbf{K}_k = \mathbf{P}_{k|k-1} \mathbf{H}_k^{\mathrm{T}} (\mathbf{H}_k \mathbf{P}_{k|k-1} \mathbf{H}_k^{\mathrm{T}} + \mathbf{\Sigma}_v)^{-1} \tag{4}$$

$$\hat{\mathbf{x}}_{k|k} = \hat{\mathbf{x}}_{k|k-1} + \mathbf{K}_k (\mathbf{z}_k - \mathbf{H}_k \hat{\mathbf{x}}_{k|k-1}) \tag{5}$$

$$\mathbf{P}_{k|k} = (\mathbf{I}_n - \mathbf{K}_k \mathbf{H}_k) \mathbf{P}_{k|k-1} \tag{6}$$

where $\hat{\mathbf{x}}_{k|k-1}$ and $\hat{\mathbf{x}}_{k|k}$ represent the prior and posterior state estimates, respectively, and $\mathbf{P}_{k|k-1}$ and $\mathbf{P}_{k|k}$ denote the PECM and posterior estimation error covariance matrix, respectively, and $\mathbf{K}_k$ is the Kalman gain at time step $k$. It can be observed from (4)–(6) that the KF achieves a balance between the prior state estimate and measurement, whose degrees of trust are described by two significant variables $\mathbf{P}_{k|k-1}$ and $\mathbf{\Sigma}_v$, respectively. The MMSE estimate can be obtained by the classical KF if the state noise $\mathbf{w}_k$ and measurement noise $\mathbf{v}_k$ are Gaussian with zero mean and covariance matrices $\mathbf{\Sigma}_w$ and $\mathbf{\Sigma}_v$, respectively [4, pp. 130]. However, the true covariance matrix would be much larger than the nominal one $\mathbf{\Sigma}_w$ and/or $\mathbf{\Sigma}_v$ if the state and/or measurement outlier occur, which results in the PECM $\mathbf{P}_{k|k-1}$ and/or MNCM $\mathbf{\Sigma}_v$ used in (4) and (6) being much less than ideal values and degrades the performance of the traditional KF. Some existing state-of-the-art outlier-robust filters for this problem will be briefly reviewed next.

The existing M-estimation-based KF [10], [12], SSMKF [16] and variational-learning-based robust KFs [15], [23]–[26] modify the PECM or SNCM and MNCM to reduce the effects of state and measurement outliers, respectively. From [10], [12], [15]–[16] and [23]–[26], we have the following Proposition.

**Proposition 1.** *In the M-estimation-based KF [10], [12], SSMKF [16] and variational-learning-based robust KFs [15], [23]–[26], the modified PECM or SNCM and MNCM are determined only by* $\mathbf{\Xi}_k \triangleq \{\hat{\mathbf{x}}_{k-1|k-1}, \mathbf{P}_{k-1|k-1}, \mathbf{\Sigma}_w, \mathbf{\Sigma}_v, \mathbf{z}_k\}$ *beside the state-space model and algorithm parameters.*

*Proof.* Retracing the derivations in [10], [12], [15]–[16] and [23]–[26], the Proposition can be easily obtained. □

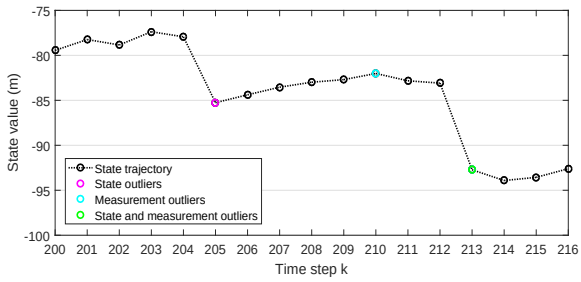
Among the five elements of $\mathbf{\Xi}_k$, only $\mathbf{z}_k$ is affected by the state or measurement outliers at step $k$. According to (1), we can formulate this effect as follows

$$\begin{aligned} \mathbf{z}_k &= \mathbf{H}_k \mathbf{F}_k \mathbf{x}_{k-1} + \mathbf{H}_k \mathbf{w}_k + \mathbf{v}_k \\ &= \hat{\mathbf{z}}_{k|k-1} - \mathbf{H}_k \mathbf{F}_k \tilde{\mathbf{x}}_{k-1|k-1} + \mathbf{H}_k \mathbf{w}_k + \mathbf{v}_k \end{aligned} \tag{7}$$

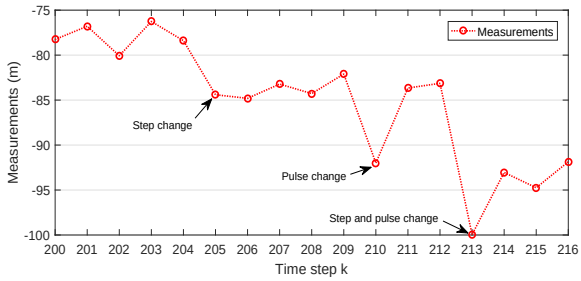
in which the predicted measurement $\hat{\mathbf{z}}_{k|k-1}$ is $\mathbf{H}_k \mathbf{F}_k \hat{\mathbf{x}}_{k-1|k-1}$ and the posterior estimation error $\tilde{\mathbf{x}}_{k-1|k-1}$ is $\hat{\mathbf{x}}_{k-1|k-1} - \mathbf{x}_{k-1}$. It can be seen from (7) that the effects of state and measurement outliers on the measurement $\mathbf{z}_k$ are similar, i.e., both the state outlier (large state noise $\mathbf{w}_k$) and measurement outlier (large measurement noise $\mathbf{v}_k$) are likely to cause the measurement $\mathbf{z}_k$ to deviate from the predicted measurement $\hat{\mathbf{z}}_{k|k-1}$. Thus, the information source $\mathbf{\Xi}_k$ for covariance matrix modification is similarly affected by the two sorts of outliers, which would inevitably lead to inappropriate modification of the PECM or SNCM and MNCM.

### C. Motivations of this paper

Motivated by this problem, it is our hypothesis that the state and measurement outliers could be better distinguished if the measurements over a period of time are used [28]. To confirm this, a simulation of the well-known one-dimensional Gaussian random walk model under the interference of outliers is performed, in which the measurement matrix is set as unity. The state trajectory and measurements over a period of time are sketched in Fig. 1, which shows that the states during steps 200–204 are around $-78m$, while the state outlier induced at step 205 pulls the states after that to around $-83m$, leading to the "step change" of the states and measurements. By contrast, the measurement outlier at step 210 only affects the measurement at the moment, not any states, causing the measurement at step 210 to be quite different from those near it, which is referred to as a "pulse change" of the measurements. In addition to the above cases, when the state outlier and measurement outlier arise simultaneously, the superposition of the "step change" and "pulse change" would be presented in the measurements, as shown at step 213 in Fig. 1(b). Therefore, it is possible to distinguish the outliers at previous steps based on measurements over a period of time. According to this idea, we can reduce the state estimation errors at previous steps, although the state outlier




(a) State trajectory




(b) Measurements

Fig. 1: State trajectory and measurements in the one-dimensional Gaussian random walk model with outliers, where the states affected by the state outlier and/or measurement outlier are highlighted.

and measurement outlier at the current step are still difficult to identify.

The above discussions motivate this work. In this paper, a sliding window variational outlier-robust Kalman filter based on Student's t noise modelling will be derived, in which the measurements in a sliding window are utilized to identify the state outliers and measurement outliers in this window.

## III. Proposed Method

The proposed method aims to identify the outliers in a sliding window and increase the corresponding noise covariance matrices to suppress them based on the VB method. We first model the heavy-tailed noises and next select the conjugate prior distribution for the noise parameters. The proposed method is then derived within the VB framework. After these derivations, we present the outlier suppression mechanism and computational complexity analysis.

### A. Noise modelling

To tackle the outliers, the heavy-tailed noises are modeled as Student's t. Assuming that the statistical characteristics of noises change slowly, we approximate the Student's t distributions of state and measurement noises in the sliding window $[k-L+1,k]$ to be the same [27]. Hence, the state transition distribution and measurement likelihood distribution at time $j \in [k-L+1,k]$ are expressed as follows

$$\begin{cases} p(\mathbf{x}_j|\mathbf{x}_{j-1},\mathbf{Q}_k) = \mathrm{St}(\mathbf{x}_j;\mathbf{F}_j\mathbf{x}_{j-1},\mathbf{Q}_k,\omega) \\ \quad = \int \mathrm{N}(\mathbf{x}_j;\mathbf{F}_j\mathbf{x}_{j-1},\mathbf{Q}_k/\xi_j)\mathrm{G}(\xi_j;\omega/2,\omega/2)d\xi_j \\ p(\mathbf{z}_j|\mathbf{x}_j,\mathbf{R}_k) = \mathrm{St}(\mathbf{z}_j;\mathbf{H}_j\mathbf{x}_j,\mathbf{R}_k,\nu) \\ \quad = \int \mathrm{N}(\mathbf{z}_j;\mathbf{H}_j\mathbf{x}_j,\mathbf{R}_k/\lambda_j)\mathrm{G}(\lambda_j;\nu/2,\nu/2)d\lambda_j \end{cases} \tag{8}$$

where $\mathbf{Q}_k$, $\mathbf{R}_k$, $\omega$ and $\nu$ are common scale matrices and dofs of state and measurement noises, respectively, and the Student's t distributions are written as the Gaussian hierarchical forms by introducing the auxiliary random variables (ARVs) $\xi_j$ and $\lambda_j$ since the Bayesian recursive update operations of the Student's t distribution are not closed [15].

In order to estimate the unknown noise scale matrices $\mathbf{Q}_k$ and $\mathbf{R}_k$ based on the VB approach, their conjugate prior distributions should be chosen. In probability theory, the inverse Wishart (IW) distribution is the conjugate distribution of the covariance matrix in a Gaussian distribution with determinate mean [30]. Furthermore, the scale matrices $\mathbf{Q}_k$ and $\mathbf{R}_k$ are proportional to the covariance matrices of the Gaussian distributions in the second and fourth formulas of (8). Consequently, we select IW distributions as the conjugate prior distributions of $\mathbf{Q}_k$ and $\mathbf{R}_k$ as follows [15]

$$\begin{cases} p(\mathbf{Q}_k|\mathbf{z}_{1:k-L}) = \mathrm{IW}(\mathbf{Q}_k;\hat{y}_{k|k-L},\hat{\mathbf{Y}}_{k|k-L}) \\ p(\mathbf{R}_k|\mathbf{z}_{1:k-L}) = \mathrm{IW}(\mathbf{R}_k;\hat{u}_{k|k-L},\hat{\mathbf{U}}_{k|k-L}) \end{cases} \tag{9}$$

where $\hat{y}_{k|k-L}$, $\hat{u}_{k|k-L}$, $\hat{\mathbf{Y}}_{k|k-L}$ and $\hat{\mathbf{U}}_{k|k-L}$ are the prior parameters of $\mathbf{Q}_k$ and $\mathbf{R}_k$, respectively. Different from [27] and [31], since multiple measurements $\mathbf{z}_{k-L+1:k}$ are utilized at time $k$, the prior parameters in (9) are spread from the posterior parameters at time $k-L$ by multiplying $L$ times the forgetting factor $\rho$ as follows

$$\begin{cases} \hat{y}_{k|k-L} = \rho^L \hat{y}_{k-L|k-L}, & \hat{\mathbf{Y}}_{k|k-L} = \rho^L \hat{\mathbf{Y}}_{k-L|k-L} \\ \hat{u}_{k|k-L} = \rho^L \hat{u}_{k-L|k-L}, & \hat{\mathbf{U}}_{k|k-L} = \rho^L \hat{\mathbf{U}}_{k-L|k-L} \end{cases} \tag{10}$$

where $\hat{y}_{k-L|k-L}$, $\hat{u}_{k-L|k-L}$, $\hat{\mathbf{Y}}_{k-L|k-L}$ and $\hat{\mathbf{U}}_{k-L|k-L}$ are the posterior dofs and scale matrices of $\mathbf{Q}_{k-L}$ and $\mathbf{R}_{k-L}$, and the forgetting factor is recommended to be set within $[0.9,1)$. By doing this, we retain the mean of the state noise scale matrix (SNSM) and measurement noise scale matrix (MNSM) and increase their uncertainty. The initial scale matrices are also modeled as IW distributed and their means are set as the nominal noise covariance matrices to utilize the prior information of the noises, i.e., [27]

$$\begin{cases} p(\mathbf{Q}_0) = \mathrm{IW}(\mathbf{Q}_0;\hat{y}_{0|0},\hat{\mathbf{Y}}_{0|0}), & \hat{\mathbf{Y}}_{0|0} = \hat{y}_{0|0}\boldsymbol{\Sigma}_w \\ p(\mathbf{R}_0) = \mathrm{IW}(\mathbf{R}_0;\hat{u}_{0|0},\hat{\mathbf{U}}_{0|0}), & \hat{\mathbf{U}}_{0|0} = \hat{u}_{0|0}\boldsymbol{\Sigma}_v \end{cases} \tag{11}$$

Based on the above model, we calculate the posterior joint distribution by employing the VB approach in the next subsection.

**Remark 1.** *In the proposed filter, the prior distributions of the scale matrices are spread from the previous posterior distributions, which leads to the form of exponential decay in the estimated scale matrices in (23). The proposed filter can adaptively estimate the scale matrices over time, and the nominal noise covariance matrices $\boldsymbol{\Sigma}_w$ and $\boldsymbol{\Sigma}_v$ will have little effect on the estimated scale matrices and modified noise covariance matrices, as shown in Fig. 8 in the simulation. Instead, in the existing RSTKF [15], the prior means of the scale matrices of the Student's t distributions are set as the nominal noise covariance matrices at each step, which makes its performance sensitive to the setting of the nominal noise covariance matrices. In fact, the existing M-estimation-based KF [10], [12] and SSMKF [16] also suffer from the same*

*problem because the nominal noise covariance matrices $\mathbf{\Sigma}_w$ and $\mathbf{\Sigma}_v$ have persistent and non-decaying effects on their modified PECM and MNCM.*

### *B. Variational inference*

In this subsection, the VB approach is applied to jointly estimate the state vectors, SNSM, MNSM and ARVs in the sliding window $[k-L+1,k]$ given measurements until time $k$. For brevity, we define the set $\mathbf{\Theta}_k = \{\mathbf{x}_{k-L:k}, \mathbf{Q}_k, \mathbf{R}_k, \xi_{k-L+1:k}, \lambda_{k-L+1:k}\}$. In the VB technique, the variables to be estimated are considered as pairwise independent, so that the posterior joint distribution is approximated as [29]

$$p(\mathbf{\Theta}_k|\mathbf{z}_{1:k}) \approx q(\mathbf{x}_{k-L:k})q(\mathbf{Q}_k) \\ q(\mathbf{R}_k)q(\xi_{k-L+1:k})q(\lambda_{k-L+1:k}) \tag{12}$$

in which every component $q(\boldsymbol{\theta}_k)$ is calculated as

$$\log q(\boldsymbol{\theta}_k) = \mathrm{E}_{(\mathbf{\Theta}_k-\boldsymbol{\theta}_k)}\left[\log p(\mathbf{\Theta}_k, \mathbf{z}_{1:k})\right] + c_{\boldsymbol{\theta}_k} \tag{13}$$

where $\boldsymbol{\theta}_k$ denotes an element of the set $\mathbf{\Theta}_k$, and $\mathrm{E}_x[\cdot]$ represents the mathematical expectation operation with respect to $x$, and the joint distribution in (13) is expressed as

$$\begin{aligned} p(\mathbf{\Theta}_k, \mathbf{z}_{1:k}) = &\prod_{j=k-L+1}^{k} [\mathrm{N}(\mathbf{z}_j; \mathbf{H}_j\mathbf{x}_j, \mathbf{R}_k/\lambda_j) \\ &\times \mathbf{G}(\lambda_j; \nu/2, \nu/2)\mathrm{N}(\mathbf{x}_j; \mathbf{F}_j\mathbf{x}_{j-1}, \mathbf{Q}_k/\xi_j)\mathrm{G}(\xi_j; \omega/2, \omega/2)] \\ &\times \mathrm{IW}(\mathbf{R}_k; \hat{u}_{k|k-L}, \hat{\mathbf{U}}_{k|k-L})\mathrm{IW}(\mathbf{Q}_k; \hat{y}_{k|k-L}, \hat{\mathbf{Y}}_{k|k-L}) \\ &\times \mathrm{N}(\mathbf{x}_{k-L}; \hat{\mathbf{x}}_{k-L|k-L}, \mathbf{P}_{k-L|k-L}) \end{aligned} \tag{14}$$

It is observed from (13) that the calculations of each of the components $q(\boldsymbol{\theta}_k)$ in (12) are mutually coupled. As a result, the fixed-point iteration is used to obtain their approximate solutions. That is, the components in (12) are solved iteratively in turn alternatively, fixing the others while solving one component. The iterative process is described in detail in Propositions 2–4.

**Proposition 2.** *Setting $\boldsymbol{\theta}_k = \mathbf{x}_{k-L:k}$ in (13), the iterative posterior distribution of the state trajectory $q^{(i+1)}(\mathbf{x}_{k-L:k})$ is obtained by running the conventional KF and Kalman smoother (KS) whose initial distribution, modified state and measurement noise covariance matrices are given as*

$$\begin{cases} p(\mathbf{x}_{k-L}|\mathbf{z}_{1:k-L}) = \mathrm{N}(\mathbf{x}_{k-L}; \hat{\mathbf{x}}_{k-L|k-L}, \mathbf{P}_{k-L|k-L}) \\ \tilde{\mathbf{Q}}_{j|k}^{(i)} = \dfrac{\left\{\mathrm{E}^{(i)}[\mathbf{Q}_k^{-1}]\right\}^{-1}}{\mathrm{E}^{(i)}[\xi_j]}, \quad \tilde{\mathbf{R}}_{j|k}^{(i)} = \dfrac{\left\{\mathrm{E}^{(i)}[\mathbf{R}_k^{-1}]\right\}^{-1}}{\mathrm{E}^{(i)}[\lambda_j]} \end{cases} \tag{15}$$

*where $j \in [k-L+1, k]$.*

*Proof.* See Appendix A. □

**Proposition 3.** *Let $\boldsymbol{\theta}_k = \mathbf{Q}_k$ and $\boldsymbol{\theta}_k = \mathbf{R}_k$ in (13) respectively, and their iterative posterior distributions $q^{(i+1)}(\mathbf{Q}_k)$ and $q^{(i+1)}(\mathbf{R}_k)$ are obtained as IW distributions, i.e.,*

$$\begin{cases} q^{(i+1)}(\mathbf{Q}_k) = \mathrm{IW}(\mathbf{Q}_k; \hat{y}_{k|k}^{(i+1)}, \hat{\mathbf{Y}}_{k|k}^{(i+1)}) \\ q^{(i+1)}(\mathbf{R}_k) = \mathrm{IW}(\mathbf{R}_k; \hat{u}_{k|k}^{(i+1)}, \hat{\mathbf{U}}_{k|k}^{(i+1)}) \end{cases} \tag{16}$$

*in which the parameters are updated by*

$$\begin{cases} \hat{y}_{k|k}^{(i+1)} = \hat{y}_{k|k-L} + L \\ \hat{\mathbf{Y}}_{k|k}^{(i+1)} = \hat{\mathbf{Y}}_{k|k-L} + \sum\limits_{j=k-L+1}^{k} \mathbf{A}_j^{(i+1)}\mathrm{E}^{(i)}[\xi_j] \\ \hat{u}_{k|k}^{(i+1)} = \hat{u}_{k|k-L} + L \\ \hat{\mathbf{U}}_{k|k}^{(i+1)} = \hat{\mathbf{U}}_{k|k-L} + \sum\limits_{j=k-L+1}^{k} \mathbf{B}_j^{(i+1)}\mathrm{E}^{(i)}[\lambda_j] \end{cases} \tag{17}$$

*where the auxiliary matrices $\mathbf{A}_j^{(i+1)}$ and $\mathbf{B}_j^{(i+1)}$ are given by*

$$\begin{cases} \mathbf{A}_j^{(i+1)} = \mathrm{E}^{(i+1)}[(\mathbf{x}_j - \mathbf{F}_j\mathbf{x}_{j-1})(\mathbf{x}_j - \mathbf{F}_j\mathbf{x}_{j-1})^{\mathrm{T}}] \\ \mathbf{B}_j^{(i+1)} = \mathrm{E}^{(i+1)}[(\mathbf{z}_j - \mathbf{H}_j\mathbf{x}_j)(\mathbf{z}_j - \mathbf{H}_j\mathbf{x}_j)^{\mathrm{T}}] \end{cases} \tag{18}$$

*Proof.* See Appendix B. □

**Proposition 4.** *Using $\boldsymbol{\theta}_k = \xi_{k-L+1:k}$ and $\boldsymbol{\theta}_k = \lambda_{k-L+1:k}$ in (13) respectively, the iterative posterior distributions of the auxiliary random variables $\xi_{k-L+1:k}$ and $\lambda_{k-L+1:k}$ are updated as the following Gamma distributions*

$$\begin{cases} q^{(i+1)}(\xi_j) = \mathrm{G}(\xi_j; \hat{a}_{j|k}^{(i+1)}, \hat{b}_{j|k}^{(i+1)}) \\ q^{(i+1)}(\lambda_j) = \mathrm{G}(\lambda_j; \hat{c}_{j|k}^{(i+1)}, \hat{d}_{j|k}^{(i+1)}) \end{cases} \tag{19}$$

*in which the shape parameters and rate parameters are calculated as*

$$\begin{cases} \hat{a}_{j|k}^{(i+1)} = 0.5(\omega + n) \\ \hat{b}_{j|k}^{(i+1)} = 0.5\left(\omega + \mathrm{tr}(\mathbf{A}_j^{(i+1)}\mathrm{E}^{(i)}[\mathbf{Q}_k^{-1}])\right) \\ \hat{c}_{j|k}^{(i+1)} = 0.5(\nu + m) \\ \hat{d}_{j|k}^{(i+1)} = 0.5\left(\nu + \mathrm{tr}(\mathbf{B}_j^{(i+1)}\mathrm{E}^{(i)}[\mathbf{R}_k^{-1}])\right) \end{cases} \tag{20}$$

*Proof.* See Appendix C. □

The expectations used in Propositions 2–4 are computed as

$$\begin{cases} \mathbf{A}_j^{(i+1)} = (\hat{\mathbf{x}}_{j|k}^{(i+1)} - \mathbf{F}_j\hat{\mathbf{x}}_{j-1|k}^{(i+1)})(\hat{\mathbf{x}}_{j|k}^{(i+1)} - \mathbf{F}_j\hat{\mathbf{x}}_{j-1|k}^{(i+1)})^{\mathrm{T}} \\ \quad +\mathbf{F}_j\mathbf{P}_{j-1|k}^{(i+1)}\mathbf{F}_j^{\mathrm{T}} - [\mathbf{P}_{j-1,j|k}^{(i+1)}]^{\mathrm{T}}\mathbf{F}_j^{\mathrm{T}} - \mathbf{F}_j\mathbf{P}_{j-1,j|k}^{(i+1)} + \mathbf{P}_{j|k}^{(i+1)} \\ \mathbf{B}_j^{(i+1)} = \mathbf{H}_j\mathbf{P}_{j|k}^{(i+1)}\mathbf{H}_j^{\mathrm{T}} + (\mathbf{z}_j - \mathbf{H}_j\hat{\mathbf{x}}_{j|k}^{(i+1)})(\mathbf{z}_j - \mathbf{H}_j\hat{\mathbf{x}}_{j|k}^{(i+1)})^{\mathrm{T}} \\ \mathrm{E}^{(i)}[\mathbf{Q}_k^{-1}] = \hat{y}_{k|k}^{(i)}\{\hat{\mathbf{Y}}_{k|k}^{(i)}\}^{-1}, \quad \mathrm{E}^{(i)}[\mathbf{R}_k^{-1}] = \hat{u}_{k|k}^{(i)}\{\hat{\mathbf{U}}_{k|k}^{(i)}\}^{-1} \\ \mathrm{E}^{(i)}[\xi_j] = \hat{a}_{j|k}^{(i)}/\hat{b}_{j|k}^{(i)}, \quad \mathrm{E}^{(i)}[\lambda_j] = \hat{c}_{j|k}^{(i)}/\hat{d}_{j|k}^{(i)} \end{cases} \tag{21}$$

in which $\hat{\mathbf{x}}_{j|k}^{(i+1)}$ and $\mathbf{P}_{j|k}^{(i+1)}$ represent the smoothing estimate and the corresponding error covariance matrix at time $j$ in the $(i+1)$–th iteration, and $\mathbf{P}_{j-1,j|k}^{(i+1)}$ represents the smoothing errors cross-covariance matrix at times $j-1$ and $j$ calculated as [32]

$$\mathbf{P}_{j-1,j|k}^{(i+1)} = \mathbf{G}_{j-1}^{(i+1)}\mathbf{P}_{j|k}^{(i+1)} \tag{22}$$

and $\mathbf{G}_{j-1}^{(i+1)}$ is the smoothing gain at time $j-1$.

The posterior distributions of the state trajectory and other variables are updated in turn at each iteration. The implementation pseudo codes of the proposed filter are displayed in Table II, where $\epsilon$ represent the iteration threshold and $N_m$ the maximum number of iterations. Prior to the fixed-point iteration, the initial expectations of SNSM, MNSM and ARVs should be precisely selected, so that the posterior distributions would converge with fewer iterations. In this paper, the initial expectations of SNSM, MNSM and ARVs are set as the

posterior expectations at the previous time, as shown in steps 1–3 in Table II.

**Remark 2.** *Although a linear state-space model is used for the derivations in this paper, the proposed SWRKF can be easily expanded to nonlinear systems by exploiting the existing analytical and statistical linearization methods, such as the first order Taylor expansion, unscented transformation and cubature transformation [33].*

**Remark 3.** *The sliding window adaptive method for estimating Gaussian noise covariance matrices [27] is used in this paper to estimate the scale matrices of the Student's t distributions. Despite this, the sliding window variational adaptive Kalman filter (SWVAKF) [27] and the proposed SWRKF are different both in terms of design ideas and algorithm implementations. On one hand, in the design ideas, the existing SWVAKF aims to address the state estimation problem under Gaussian noises with unknown and/or slowly time-varying covariance matrices, while the proposed SWRKF is intended to tackle the non-Gaussian heavy-tailed noises contaminated by the outliers. Accordingly, the noises in the sliding window are modelled as Gaussian with the same covariance matrices in the SWVAKF instead of Student's t in this paper. On the other hand, in the algorithm implementations, the proposed SWRKF adds the updates of ARVs as compared with the SWVAKF, which enables the noise covariance matrices in the sliding window to be adjusted discriminately to counter individual outliers, as shown on lines 14–16 in Table II and discussed in Section III-C. Besides, the initial expectations of the ARVs in the proposed SWRKF are set as the posterior expectations at the previous time to accelerate convergence as shown on line 3 in Table II, which is not covered in the SWVAKF.*

### C. Mechanism of outlier suppression

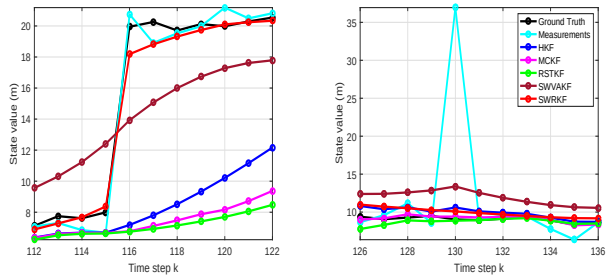
The proposed SWRKF identifies the state outliers and measurement outliers by auxiliary matrices $\mathbf{A}_j^{(i)}$ and $\mathbf{B}_j^{(i)}$, respectively, whereby the state and measurement noise covariance matrices $\tilde{\mathbf{Q}}_{j|k}^{(i)}$ and $\tilde{\mathbf{R}}_{j|k}^{(i)}$ are modified accordingly to suppress the effects of two kinds of outliers. From Proposition 2, the modified state and measurement noise covariance matrices $\tilde{\mathbf{Q}}_{j|k}^{(i)}$ and $\tilde{\mathbf{R}}_{j|k}^{(i)}$ are obtained by dividing the estimated SNSMs $\left\{\mathrm{E}^{(i)}[\mathbf{Q}_k^{-1}]\right\}^{-1}$ and $\left\{\mathrm{E}^{(i)}[\mathbf{R}_k^{-1}]\right\}^{-1}$ (simplified to $\hat{\mathbf{Q}}_k^{(i)}$ and $\hat{\mathbf{R}}_k^{(i)}$) by the estimated ARVs $\mathrm{E}^{(i)}[\xi_j]$ and $\mathrm{E}^{(i)}[\lambda_j]$. Utilizing (10), (17), (20) and (21), the estimated SNSMs and ARVs are formulated as

$$\begin{cases} \hat{\mathbf{Q}}_k^{(i)} = \frac{\mathbf{M}_{k-L+1}^{k(i)} + \rho^L \mathbf{M}_{k-2L+1}^{k-L(i)} + \rho^{2L} \mathbf{M}_{k-3L+1}^{k-2L(i)} + \cdots}{(1+\rho^L+\rho^{2L}+\cdots)L} \\ \hat{\mathbf{R}}_k^{(i)} = \frac{\mathbf{N}_{k-L+1}^{k(i)} + \rho^L \mathbf{N}_{k-2L+1}^{k-L(i)} + \rho^{2L} \mathbf{N}_{k-3L+1}^{k-2L(i)} + \cdots}{(1+\rho^L+\rho^{2L}+\cdots)L} \\ \mathrm{E}^{(i)}[\xi_j] = \frac{\omega+n}{\omega+\mathrm{tr}\left(\mathbf{A}_j^{(i)}\mathrm{E}^{(i-1)}[\mathbf{Q}_k^{-1}]\right)} \\ \mathrm{E}^{(i)}[\lambda_j] = \frac{\nu+m}{\nu+\mathrm{tr}\left(\mathbf{B}_j^{(i)}\mathrm{E}^{(i-1)}[\mathbf{R}_k^{-1}]\right)} \end{cases} \tag{23}$$

where $\mathbf{M}_p^{q(i)} \triangleq \sum_{j=p}^{q} \mathbf{A}_j^{(i)} \mathrm{E}^{(i-1)}[\xi_j]$ and $\mathbf{N}_p^{q(i)} \triangleq \sum_{j=p}^{q} \mathbf{B}_j^{(i)} \mathrm{E}^{(i-1)}[\lambda_j]$. It is found from (23) that the estimated SNSM $\hat{\mathbf{Q}}_k^{(i)}$ is a weighted average of the terms $\mathbf{A}_j^{(i)}\mathrm{E}^{(i-1)}[\xi_j]$ in the form of exponential decay, where the terms in the earlier sliding window have smaller weights. The monomial $\mathbf{A}_j^{(i)}\mathrm{E}^{(i-1)}[\xi_j]$ has limited effects on the estimated SNSM $\hat{\mathbf{Q}}_k^{(i)}$ since $\hat{\mathbf{Q}}_k^{(i)}$ is a weighted average of several terms. So the auxiliary matrix $\mathbf{A}_j^{(i)}$ mainly affects the estimated ARV $\mathrm{E}^{(i)}[\xi_j]$. A large $\mathbf{A}_j^{(i)}$ would cause small $\mathrm{E}^{(i)}[\xi_j]$ and finally result in large $\tilde{\mathbf{Q}}_{j|k}^{(i)}$, which weakens the effects of the state

TABLE II: The implementation pseudo codes of the proposed SWRKF for each step.

**Inputs**: $\{\hat{\mathbf{x}}_{j|j}, \mathbf{P}_{j|j}, \hat{y}_{j|j}, \hat{\mathbf{Y}}_{j|j}, \hat{u}_{j|j}, \hat{\mathbf{U}}_{j|j}, \mathbf{z}_{j+1} | j \in [k-L, k-1]\}$, $\{\hat{\xi}_{j|k-1}, \hat{\lambda}_{j|k-1} | j \in [k-L+1, k-1]\}$, $\mathbf{F}_k$, $\mathbf{H}_k$, $\omega$, $\nu$, $\rho$, $L$, $m$, $n$, $\epsilon$, $\delta$, $N_m$.
**Initialization**:
1. $\mathrm{E}^{(0)}[\mathbf{Q}_k^{-1}] = \hat{y}_{k-1|k-1}(\hat{\mathbf{Y}}_{k-1|k-1})^{-1}$,
2. $\mathrm{E}^{(0)}[\mathbf{R}_k^{-1}] = \hat{u}_{k-1|k-1}(\hat{\mathbf{U}}_{k-1|k-1})^{-1}$.
*for* $j = k-L+1 : k-1$
3. $\mathrm{E}^{(0)}[\xi_j] = \hat{\xi}_{j|k-1}, \mathrm{E}^{(0)}[\lambda_j] = \hat{\lambda}_{j|k-1}$ \\ Accelerate the convergence
*end for*
\\ Enable the scale matrices $\mathbf{Q}_k$ and $\mathbf{R}_k$ to be estimated adaptively
4. Calculate $\hat{y}_{k|k-L}$, $\hat{\mathbf{Y}}_{k|k-L}$, $\hat{u}_{k|k-L}$ and $\hat{\mathbf{U}}_{k|k-L}$ using (10)
**Iteration**:
*for* $i = 0 : N_m - 1$
\\ Update $q^{(i+1)}(\mathbf{x}_{k-L:k})$ given $q^{(i+1)}(\mathbf{Q}_k)$, $q^{(i+1)}(\mathbf{R}_k)$,
\\ $q^{(i+1)}(\xi_{k-L+1:k})$ and $q^{(i+1)}(\lambda_{k-L+1:k})$
5. $\hat{\mathbf{x}}_{k-L|k-L}^{(i+1)} = \hat{\mathbf{x}}_{k-L|k-L}$, $\hat{\mathbf{P}}_{k-L|k-L}^{(i+1)} = \hat{\mathbf{P}}_{k-L|k-L}$
*for* $j = k-L+1 : k$
\\ Run the conventional KF with modified SNCM $\frac{\mathrm{E}^{(i)}[\mathbf{Q}_k^{-1}]^{-1}}{\mathrm{E}^{(i)}[\xi_j]}$ and
\\ MNCM $\frac{\mathrm{E}^{(i)}[\mathbf{R}_k^{-1}]^{-1}}{\mathrm{E}^{(i)}[\lambda_j]}$
6. $\hat{\mathbf{x}}_{j|j-1}^{(i+1)} = \mathbf{F}_j \hat{\mathbf{x}}_{j-1|j-1}^{(i+1)}$
7. $\mathbf{P}_{j|j-1}^{(i+1)} = \mathbf{F}_j \mathbf{P}_{j-1|j-1}^{(i+1)} \mathbf{F}_j^{\mathrm{T}} + \frac{\mathrm{E}^{(i)}[\mathbf{Q}_k^{-1}]^{-1}}{\mathrm{E}^{(i)}[\xi_j]}$
8. $\mathbf{K}_j^{(i+1)} = \mathbf{P}_{j|j-1}^{(i+1)} \mathbf{H}_j^T (\mathbf{H}_j \mathbf{P}_{j|j-1}^{(i+1)} \mathbf{H}_j^{\mathrm{T}} + \frac{\mathrm{E}^{(i)}[\mathbf{R}_k^{-1}]^{-1}}{\mathrm{E}^{(i)}[\lambda_j]})^{-1}$
9. $\hat{\mathbf{x}}_{j|j}^{(i+1)} = \hat{\mathbf{x}}_{j|j-1}^{(i+1)} + \mathbf{K}_j^{(i+1)}(\mathbf{z}_j - \mathbf{H}_j \hat{\mathbf{x}}_{j|j-1}^{(i+1)})$ \\ $\mathbf{z}_{k-L+1:k}$ are used
10. $\mathbf{P}_{j|j}^{(i+1)} = (\mathbf{I}_n - \mathbf{K}_j^{(i+1)} \mathbf{H}_j) \mathbf{P}_{j|j-1}^{(i+1)}$
*end for*
*for* $j = k : (-1) : k-L+1$
\\ Run the conventional KS with modified SNCM $\frac{\mathrm{E}^{(i)}[\mathbf{Q}_k^{-1}]^{-1}}{\mathrm{E}^{(i)}[\xi_j]}$
11. $\mathbf{G}_{j-1}^{(i+1)} = \mathbf{P}_{j-1|j-1}^{(i+1)} \mathbf{F}_j^{\mathrm{T}} [\mathbf{P}_{j|j-1}^{(i+1)}]^{-1}$
12. $\hat{\mathbf{x}}_{j-1|k}^{(i+1)} = \hat{\mathbf{x}}_{j-1|j-1}^{(i+1)} + \mathbf{G}_{j-1}^{(i+1)}(\hat{\mathbf{x}}_{j|k}^{(i+1)} - \hat{\mathbf{x}}_{j|j-1}^{(i+1)})$
13. $\mathbf{P}_{j-1|k}^{(i+1)} = \mathbf{P}_{j-1|j-1}^{(i+1)} + \mathbf{G}_{j-1}^{(i+1)}(\mathbf{P}_{j|k}^{(i+1)} - \mathbf{P}_{j|j-1}^{(i+1)})[\mathbf{G}_{j-1}^{(i+1)}]^{\mathrm{T}}$
*end for*
\\ Update $q^{(i+1)}(\xi_{k-L+1:k})$ and $q^{(i+1)}(\lambda_{k-L+1:k})$ given
\\ $q^{(i+1)}(\mathbf{x}_{k-L:k})$, which enables the outliers in the sliding window
\\ to be resisted
*for* $j = k : (-1) : k-L+1$
14. Calculate $\mathbf{A}_j^{(i+1)}$ and $\mathbf{B}_j^{(i+1)}$ using (21) \\ $\mathbf{z}_{k-L+1:k}$ are used
15. Calculate $\hat{a}_{j|k}^{(i+1)}$, $\hat{b}_{j|k}^{(i+1)}$, $\hat{c}_{j|k}^{(i+1)}$ and $\hat{d}_{j|k}^{(i+1)}$ using (20)
16. Calculate $\mathrm{E}^{(i+1)}[\xi_j]$ and $\mathrm{E}^{(i+1)}[\lambda_j]$ using (21)
*end for*
\\ Update $q^{(i+1)}(\mathbf{Q}_k)$ and $q^{(i+1)}(\mathbf{R}_k)$ given $q^{(i+1)}(\mathbf{x}_{k-L:k})$
17. Calculate $\hat{y}_{k|k}^{(i+1)}$, $\hat{\mathbf{Y}}_{k|k}^{(i+1)}$, $\hat{u}_{k|k}^{(i+1)}$ and $\hat{\mathbf{U}}_{k|k}^{(i+1)}$ using (17)
18. Calculate $\mathrm{E}^{(i+1)}[\mathbf{Q}_k^{-1}]$ and $\mathrm{E}^{(i+1)}[\mathbf{R}_k^{-1}]$ using (21).
19. *if* $\left\|\hat{\mathbf{x}}_{k|k}^{(i+1)} - \hat{\mathbf{x}}_{k|k}^{(i)}\right\| \Big/ \left\|\hat{\mathbf{x}}_{k|k}^{(i)}\right\| \leq \epsilon$, terminate iterations, *end if*
*end for*
**Data saving**:
*for* $j = k-L+2 : k$
20. $\hat{\xi}_{j|k} = \mathrm{E}^{(i+1)}[\xi_j]$, $\hat{\lambda}_{j|k} = \mathrm{E}^{(i+1)}[\lambda_j]$
*end for*
21. $\hat{\mathbf{x}}_{k|k} = \hat{\mathbf{x}}_{k|k}^{(i+1)}$, $\mathbf{P}_{k|k} = \mathbf{P}_{k|k}^{(i+1)}$, $\hat{y}_{k|k} = \hat{y}_{k|k}^{(i+1)}$, $\hat{\mathbf{Y}}_{k|k} = \hat{\mathbf{Y}}_{k|k}^{(i+1)}$,
22. $\hat{u}_{k|k} = \hat{u}_{k|k}^{(i+1)}$, $\hat{\mathbf{U}}_{k|k} = \hat{\mathbf{U}}_{k|k}^{(i+1)}$.
**Outputs**: $\{\hat{\mathbf{x}}_{j|j}, \mathbf{P}_{j|j}, \hat{y}_{j|j}, \hat{\mathbf{Y}}_{j|j}, \hat{u}_{j|j}, \hat{\mathbf{U}}_{j|j} | j \in [k-L+1, k]\}$, $\{\hat{\xi}_{j|k}, \hat{\lambda}_{j|k} | j \in [k-L+2, k]\}$




(a) (b)

Fig. 2: Two intervals in a Gaussian random walk example, where the state and measurement noises are produced by (27) with $\mathbf{\Sigma}_w = 0.1m^2$, $\mathbf{\Sigma}_v = 1m^2$, $U_1 = 500$, $U_2 = 100$ and $p_1 = p_2 = 0.9$. In (a) and (b), the state outlier and measurement outlier occur at steps 116 and 130, respectively. The parameters of different filters are set according to Table IV, and the forgetting factor and window length of the SWVAKF are set to be the same as those of SWRKF. The curves labeled HKF, MCKF and RSTKF denote their posterior filtering estimates, while the curves labeled SWVAKF and SWRKF represent their posterior smoothing estimates $\hat{\mathbf{x}}_{k-L:k|k}^{(N)}$.

outliers. Likewise, we can deduce that a large $\mathbf{B}_j^{(i)}$ would lead to small $\mathrm{E}^{(i)}[\lambda_j]$ and large $\tilde{\mathbf{R}}_{j|k}^{(i)}$.

Next, we explain how large $\mathbf{A}_j^{(i)}$ or $\mathbf{B}_j^{(i)}$ is obtained when the state or measurement outlier occurs in the proposed filter, respectively. Firstly, it is assumed that the filter runs to step $k$ and a state outlier occurs at step $k-b$ ($0 < b < L$). The state outlier $\mathbf{w}_{k-b}$ divides the ground truth and measurements in the sliding window into two stages (before and after step $k-b$, for example, the levels of states and measurements before and after step 116 are markedly different in the case in Fig. 2 (a)). Due to the step change of the measurements, the difference $\hat{\mathbf{x}}_{k-b|k}^{(i)} - \mathbf{F}_{k-b}\hat{\mathbf{x}}_{k-b-1|k}^{(i)}$ would be relatively larger than usual, which results in relatively larger $\mathbf{A}_{k-b}^{(i)}$ and $\tilde{\mathbf{Q}}_{k-b|k}^{(i)}$ according to (21) at the $i$-th iteration. At the $i+1$-th iteration, the Kalman smoothing process in the proposed method would correct the estimated state trajectory after step $k-b$ to close to the ground truth $\mathbf{x}_{k-b:k}$. Instead, the estimated state trajectory before step $k-b$ would be less affected by the later stage in the Kalman smoothing process because of the large $\tilde{\mathbf{Q}}_{k-b|k}^{(i)}$ and small smoothing gain $\mathbf{G}_{k-b-1}^{(i+1)}$ according to lines 7 and 11 in Table II. Then, the estimated $\hat{\mathbf{x}}_{k-L:k-b-1}^{(i+1)}$ would still be close to the ground truth $\mathbf{x}_{k-L:k-b-1}$. Therefore, the difference $\hat{\mathbf{x}}_{k-b|k}^{(i+1)} - \mathbf{F}_{k-b}\hat{\mathbf{x}}_{k-b-1|k}^{(i+1)}$ would get much larger, contributing to much larger $\mathbf{A}_{k-b}^{(i+1)}$ in the $i+1$-th iteration.

Secondly, we consider the case where a measurement outlier occurs at step $k-b$, which affects the measurement $\mathbf{z}_{k-b}$ only (see step 130 in Fig. 2 (b) for instance). Although the state estimate $\hat{\mathbf{x}}_{k-b|k-b}^{(i)}$ may be skewed by the measurement outlier $\mathbf{v}_{k-b}$ when the filter runs to step $k-b$, the Kalman smoothing process would correct the state estimate $\hat{\mathbf{x}}_{k-b|k}^{(i)}$ to be close to the ground truth $\mathbf{x}_{k-b}$ when the filter runs to step $k$. Hence, the relatively large residual $\mathbf{z}_{k-b} - \mathbf{H}_{k-b}\hat{\mathbf{x}}_{k-b|k}^{(i)}$ would be obtained, which results in relatively large $\mathbf{B}_{k-b}^{(i)}$ and $\tilde{\mathbf{R}}_{k-b|k}^{(i)}$. In the $i+1$-th iteration, the relatively large $\tilde{\mathbf{R}}_{k-b|k}^{(i)}$ further resists the abnormal measurement $\mathbf{z}_{k-b}$ in the Kalman filtering process and brings about larger residual and larger $\mathbf{B}_{k-b}^{(i+1)}$, which can be confirmed at step 130 in Fig. 2 (b).

In summary, when a state outlier or measurement outlier occurs at step $k-b$ ($0 < b < L$), the proposed filter would get large $\mathbf{A}_{k-b}^{(i)}$ or $\mathbf{B}_{k-b}^{(i)}$ and increase the $\tilde{\mathbf{Q}}_{k-b|k}^{(i)}$ or $\tilde{\mathbf{R}}_{k-b|k}^{(i)}$ to suppress the state or measurement outlier. Since the outliers before step $k$ are suppressed, the proposed filter can generate more accurate state estimate and its error covariance matrix $\{\hat{\mathbf{x}}_{k-1|k}^{(i)}, \mathbf{P}_{k-1|k}^{(i)}\}$ at step $k-1$ than the filters that do not identify the outliers before step $k$, which improves its estimation accuracy at the current step $k$.

**Remark 4.** *In the existing M-estimation-based KF [10], [12], SSMKF [16] and variational-learning-based robust KFs [15], [23]–[26], the PECM $\mathbf{P}_{k|k-1}$ or SNCM $\mathbf{Q}_k$ and MNCM $\mathbf{R}_k$ are adjusted on the basis of $\mathbf{\Xi}_k$. However, as discussed in Section II-B, there is not enough information in $\mathbf{\Xi}_k$ to distinguish the state outlier and measurement outlier, which causes the difficulty in adjusting the PECM or SNCM and MNCM correctly at each step (including before step $k$). Hence, the state estimate and its error covariance matrix $\{\hat{\mathbf{x}}_{k-1|k-1}^{(i)}, \mathbf{P}_{k-1|k-1}^{(i)}\}$ given by these are less accurate than $\{\hat{\mathbf{x}}_{k-1|k}^{(i)}, \mathbf{P}_{k-1|k}^{(i)}\}$ given by the proposed filter.*

**Remark 5.** *The conclusion in this subsection can be confirmed by Fig. 2, especially the case of the state outlier. In Fig. 2 (a), at steps 117–122, the existing HKF [10], MCKF [12] and RSTKF [15] cannot identify the state outlier at step 116. At the same time, these filters cannot correctly decide whether to trust the measurement or the model, which results in large errors after the state outlier. On the contrary, the proposed filter can identify the state outlier after that and enhance the SNCM at step 116, which resists the state outlier and decreases the estimation errors after the state outlier.*

**Remark 6.** *As discussed in Remark 3, the noise covariance matrices in the sliding window are assumed to be the same in the SWVAKF [27], which leads to its inability to enhance some covariance matrices larger than others to resist the outliers. Therefore, the performance of the SWVAKF is easy to degrade by outliers, as verified in Fig. 2. In Fig. 2(a), the smoothing estimates of the SWVAKF do not approach the step change of the state trajectory at step 116, leading to its large errors after step 116. In Fig. 2(b), the SWVAKF is misled by the measurement outlier at step 130 and gives a skewed estimate even 6 steps after the outlier.*

### *D. Computational complexity analysis*

The computational complexity in terms of the floating point operations is analysed for the proposed SWRKF and existing RSTKF [15]. According to Table II and [15], the

TABLE III: Computational complexity of the proposed SWRKF

| Steps | Computational complexity |
|---|---|
| 1–4 | $\mathcal{O}(m^3)+\mathcal{O}(n^3)+2m^2+2n^2+2$ |
| 5–10 | $NL[2\mathcal{O}(m^3)+2m^2n+3mn^2+3n^3+\mathcal{O}(n^3)$ $+m^2+2mn+2n^2]$ |
| 11–13 | $NL[4n^3+\mathcal{O}(n^3)+n^2]$ |
| 14–16 | $NL[m^3+m^2n+mn^2+5n^3+m^2+2mn+3n^2+6]$ |
| 17–19 | $N[L(m^2+n^2)+\mathcal{O}(m^3)+\mathcal{O}(n^3)+m^2+n^2]$ |

computational complexity of the two filters are respectively

$$\begin{aligned}\mathrm{S}_{\mathrm{SWRKF}} &= N\left\{L\left[m^3+2\mathcal{O}(m^3)+3m^2n+4mn^2+12n^3\right.\right.\\ &\left.+2\mathcal{O}(n^3)+3m^2+4mn+7n^2+6\right]+\mathcal{O}(m^3)+\mathcal{O}(n^3)\\ &\left.+m^2+n^2\right\}+\mathcal{O}(m^3)+\mathcal{O}(n^3)+2m^2+2n^2+2 \qquad (24)\end{aligned}$$

$$\begin{aligned}\mathrm{S}_{\mathrm{RSTKF}} &= N\left[m^3+\mathcal{O}(m^3)+3m^2n+4mn^2+2n^3\right.\\ &\left.+2\mathcal{O}(n^3)+2m^2+4mn+4n^2+6\right]+2n^3+\mathcal{O}(n^3)+3n^2\\ &\qquad (25)\end{aligned}$$

where $N$ and $L$ represent the number of iterations and window length, respectively, and the computational complexity of the proposed SWRKF is decomposed as Table III.

It can be found from (24) that the computational complexity of the proposed SWRKF increases linearly with increasing $N$ and $L$. As shown in the simulation later, the proposed SWRKF exhibits superior estimation performance with small $N$ and $L$. Hence, the number of iterations $N$ and the window length $L$ are, respectively, recommended to be set as 2 and within $[10, 20]$ to achieve a tradeoff between computational efficiency and estimation accuracy. In addition, the proposed SWRKF has greater spatial complexity resulting from the additional storages of $\{\hat{\mathbf{x}}_{j|j}, \mathbf{P}_{j|j}, \hat{y}_{j|j}, \hat{\mathbf{Y}}_{j|j}, \hat{u}_{j|j}, \hat{\mathbf{U}}_{j|j}, \mathbf{z}_{j+1}|j \in [k-L, k-1]\}$ and $\{\hat{\xi}_{j|k-1}, \hat{\lambda}_{j|k-1}|j \in [k-L+1, k-1]\}$.

## IV. Simulation Study

The merits of the proposed SWRKF are illustrated by the simulations of manoeuvring target tracking in this section. The proposed filter will be compared with the typical robust filters, i.e. HKF [10], MCKF [12], RSTKF [15] and Student's t-based PF (STPF) [19] when necessary. We also compare the proposed filter with the off-line robust Gaussian approximate smoother (RGAS) [8] which is used as a performance reference. The SSMKF [16] is omitted since the RSTKF is equivalent to the SSMKF with logarithmic similarity function.

In the simulations of manoeuvring target tracking, we define the state vector as $\mathbf{x}_k = [x_k,\ y_k,\ \dot{x}_k,\ \dot{y}_k]^{\mathrm{T}}$, where $x_k$ and $y_k$ represent respectively the abscissa and ordinate of the target and $\dot{x}_k$ and $\dot{y}_k$ denote the corresponding velocities. The kinematic and measurement models are described as (1) in which the state transition and measurement matrices are

$$\mathbf{F}_k = \begin{bmatrix}\mathbf{I}_2 & \Delta t\mathbf{I}_2\\ \mathbf{0} & \mathbf{I}_2\end{bmatrix}, \qquad \mathbf{H}_k = \begin{bmatrix}\mathbf{I}_2 & \mathbf{0}\end{bmatrix} \qquad (26)$$

TABLE IV: Parameter Settings for different algorithms

| Algorithms | Parameter settings |
|---|---|
| HKF | Tuning parameter $\gamma = 1.345$ |
| MCKF | Kernel size $\sigma = 5$ |
| RSTKF | D.f. of Student's t noises $\omega = \nu = 5$<br>D.f. of IW distributed scale matrices $\tau_w = \tau_v = 5$ |
| TPF | Particle numbers $N_p = 1000$ |
| STPF | D.f. of Student's t noises $\omega = \nu = 5$<br>Particle numbers $N_p = 1000$ |
| RGAS | D.f. of Student's t noises $\omega = \nu = 5$<br>Prior parameters of scale matrices $t_0 = 6$, $\mathbf{T}_0 = \mathbf{\Sigma}_w$, $u_0 = 4$, $\mathbf{U}_0 = \mathbf{\Sigma}_v$ |
| SWRKF | D.f. of Student's t noises $\omega = \nu = 5$<br>D.f. of IW distributed scale matrices $\hat{y}_{0\|0} = \hat{u}_{0\|0} = 5$<br>Window length $L = 10$<br>Forgetting factor $\rho = 1 - e^{-4}$ |

and the outlier-contaminated noises are generated according to

$$\begin{cases}\mathbf{w}_k \sim \begin{cases}N(\mathbf{0}, \mathbf{\Sigma}_w) & \text{w.p. } 0.9\\ N(\mathbf{0}, 100\mathbf{\Sigma}_w) & \text{w.p. } 0.1\end{cases}\\ \mathbf{v}_k \sim \begin{cases}N(\mathbf{0}, \mathbf{\Sigma}_v) & \text{w.p. } 0.9\\ N(\mathbf{0}, 100\mathbf{\Sigma}_v) & \text{w.p. } 0.1\end{cases}\end{cases} \qquad (27)$$

where $\Delta t = 1s$, the nominal covariance matrices are chosen as $\mathbf{\Sigma}_w = 0.5\begin{bmatrix}\frac{T^3}{3}\mathbf{I}_2 & \frac{T^2}{2}\mathbf{I}_2\\ \frac{T^2}{2}\mathbf{I}_2 & T\mathbf{I}_2\end{bmatrix}$ and $\mathbf{\Sigma}_v = 100\mathbf{I}_2 m^2$, and the abbreviation w.p. stands for with probability. The initial state of the target is generated randomly from the normal distribution with mean $\mathbf{x}_0 = [0,\ 0,\ 10,\ 10]^{\mathrm{T}}$ and covariance matrix $\mathbf{P}_0 = \mathrm{diag}([10000,\ 10000,\ 100,\ 100])$ and 5000 subsequent points are simulated. Besides the methods mentioned above, the KF with true noise covariance matrices (TKF) and KF with nominal noise covariance matrices (NKF) and PF with true Gaussian mixture distribution (TPF) act as references in the simulation. The parameters of different methods are all selected according to the recommendations in the literature, as shown in detail in Table IV. For STPF, the scale matrices are selected as the nominal covariance matrices, i.e. $\mathbf{\Sigma}_w$ and $\mathbf{\Sigma}_v$. The maximum numbers of iterations for HKF, MCKF, RSTKF and RGAS are set as $N_m = 50$ while that for SWRKF is set as $N_m = 2$. Besides, the iteration threshold $\epsilon = 10^{-8}$. The MATLAB codes of this simulation can be open access from the link https://www.researchgate.net/profile/Yulong-Huang-4.

The root mean square error (RMSE) and normalized estimation error squared (NEES) [34] are used to evaluate the estimation accuracy and consistency. The RMSEs and NEESs of different algorithms computed based on 1000 Monte Carlo runs are shown in Figs. 3–4, which are smoothed using a 20-point moving average for convenient comparison. The corresponding average RMSEs (ARMSEs), average NEESs (ANEESs) and single step run time are listed in Table V.

The STPF method is omitted here because it diverges in this scenario. As can be seen from Figs. 3–4 and Table V, in terms of accuracy and consistency, the NKF performs the worst and the proposed filter performs the best except for the off-line smoother RGAS and the unavailable TPF and TKF. Compared with the existing HKF, MCKF and RSTKF, the

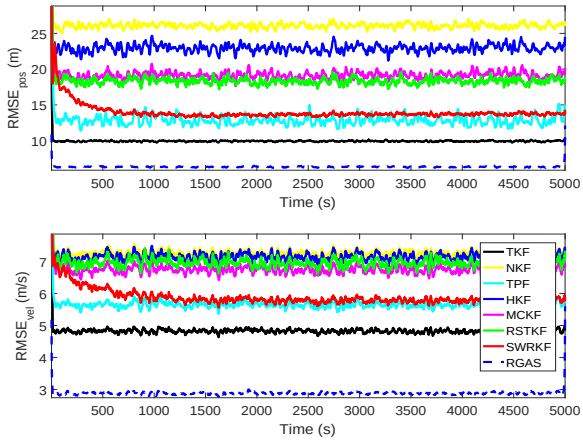



Fig. 3: RMSEs of different algorithms.

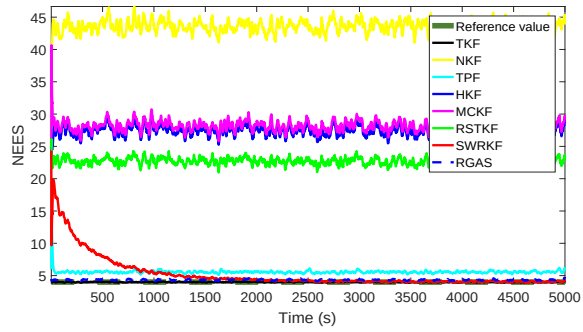



Fig. 4: NEESs of different algorithms (The closer the NEES is to state dimension 4, the better the consistency).

TABLE V: ARMSEs, ANEESs and Single step run time of different algorithms

| Algorithms | $\text{ARMSE}_{\text{pos}}$ (m) | $\text{ARMSE}_{\text{vel}}$ (m/s) | ANEES | Time (ms) |
|---|---|---|---|---|
| TKF | 9.94 | 4.83 | 4.00 | 0.010 |
| NKF | 25.93 | 7.26 | 43.38 | 0.010 |
| TPF | 12.70 | 5.64 | 5.56 | 22.277 |
| HKF | 22.87 | 7.17 | 27.53 | 0.193 |
| MCKF | 19.10 | 6.77 | 28.25 | 0.129 |
| RSTKF | 18.18 | 6.95 | 22.73 | 0.271 |
| SWRKF | 13.90 | 5.88 | 5.24 | 0.964 |
| RGAS | 6.35 | 2.88 | 4.34 | N/A |

proposed filter has smaller (A)RMSE and (A)NEES closer to the state dimension 4 but heavier computational burden.

To reveal the reason for the superiority of the proposed filter, we study the ability of each algorithm to identify outliers next. In this paper, we determine that the noise $\mathbf{w}_k$ or $\mathbf{v}_k$ is identified as an outlier by the robust filter or smoother if the following inequality holds

$$\sum_i \text{eig}_i(\tilde{\mathbf{Q}}_k^* - d\boldsymbol{\Sigma}_w) > 0 \quad \text{or} \quad \sum_i \text{eig}_i(\tilde{\mathbf{R}}_k^* - d\boldsymbol{\Sigma}_v) > 0 \tag{28}$$

where $\text{eig}_i(\mathbf{A})$ denotes the $i$-th eigenvalue of matrix $\mathbf{A}$, and

| | Noises identified as outliers by the robust filter | Noises not identified as outliers by the robust filter |
|---|---|---|
| Outliers noises | *TP* | *FN* |
| Non-outliers noises | *FP* | *TN* |

Fig. 5: Confusion matrix of outliers identification results.

TABLE VI: Precision, recall and F-score of different algorithms (s.o. denotes the state outliers and m.o. the measurement outliers).

| Algorithms | Outlier type | Precision (%) | Recall (%) | F-score (%) |
|---|---|---|---|---|
| HKF | s.o. | 9.99 | 0.21 | 0.40 |
| MCKF | s.o. | 9.54 | 0.16 | 0.32 |
| RSTKF | s.o. | 10.14 | 22.42 | 13.96 |
| SWRKF | s.o. | 10.61 | 72.91 | 18.48 |
| RGAS | s.o. | 14.04 | 87.27 | 22.78 |
| HKF | m.o. | 99.66 | 29.78 | 45.81 |
| MCKF | m.o. | 97.45 | 44.35 | 60.92 |
| RSTKF | m.o. | 97.34 | 71.03 | 82.11 |
| SWRKF | m.o. | 99.89 | 79.48 | 88.51 |
| RGAS | m.o. | 100 | 79.91 | 88.82 |

$\tilde{\mathbf{Q}}_k^*$ and $\tilde{\mathbf{R}}_k^*$ denote the modified state and measurement noise covariance matrices of the HKF [10], MCKF [12], RSTKF [15], the proposed SWRKF and RGAS [8], and the factor $d = 7$. Specially, the modified state and measurement noise covariance matrices of the proposed filter are defined as

$$\tilde{\mathbf{Q}}_k^{SWRKF} = \tilde{\mathbf{Q}}_{k|k+b}, \quad \tilde{\mathbf{R}}_k^{SWRKF} = \tilde{\mathbf{R}}_{k|k+b} \tag{29}$$

to highlight the effect of the sliding window, where $\tilde{\mathbf{Q}}_{k|k+b}$ and $\tilde{\mathbf{R}}_{k|k+b}$ are defined in (15) and the delay step $b$ is set as 5. As such, all the noises can be classified by four categories defined in the confusion matrix in Fig. 5 [35], whose numbers are respectively denoted as $TP$, $FN$, $FP$ and $TN$. The identification performance is evaluated by precision, recall and F-score [35], where precision and recall are defined as $P = \frac{TP}{TP+FP}$ and $R = \frac{TP}{TP+FN}$, respectively, and F-score is defined as their harmonic mean $F = \frac{2 \cdot P \cdot R}{P+R}$ to combine them. The precision and recall are generally contradictory. Only when both are high (i.e., F-score is high) the performance of outlier identification is good. These indexes in the simulation environment above are displayed in Table VI. It is evident from Table VI that the RGAS has the largest F-score both in the aspect of state outliers and measurement outliers, since all the measurements are utilized to identify the outliers. Except for the RGAS, the proposed SWRKF has larger precision, recall and F-score than others, which mainly benefits from its better mechanism of outlier identification. As discussed in Section III-C, the better ability to identify the outliers brings about the superior estimation performance as shown in Figs. 3–4.

To illustrate the effect of the maximum number of iterations $N_m$ and window length $L$ on the proposed algorithm, some further simulations are carried out. To this end, we change $N_m$ and $L$ from 1 to 20. The relationship between ARMSEs

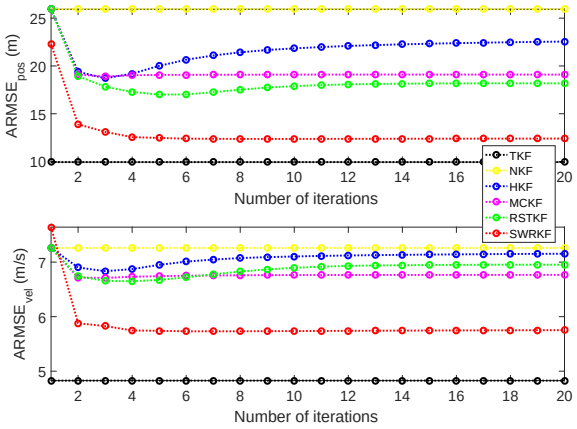



Fig. 6: ARMSEs for different numbers of iterations.

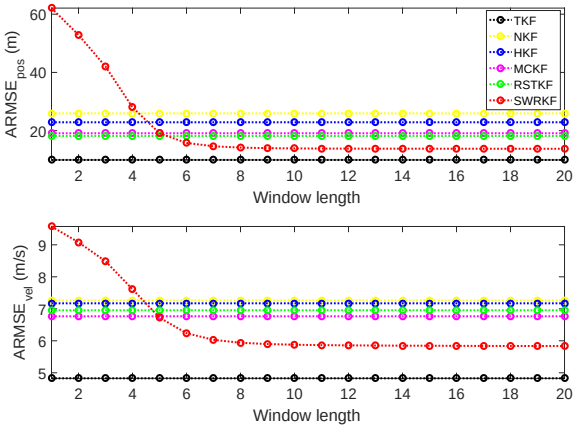



Fig. 7: ARMSEs for different window length.

and the number of iterations for each filter is sketched in Fig. 6. It is observed that the proposed filter performs better than the existing methods when $N_m \geq 2$ and its ARMSE improves a little as $N_m$ continues to increase. We attribute it to the appropriate selections of the initial values of scale matrices and ARVs at the iteration, which makes it converge with fewer iterations. The ARMSEs for different window length are sketched in Fig. 7 and one can observe that the proposed filter outperforms the existing methods when $L \geq 6$ and the ARMSE tends to plateau when $L \geq 8$, because the proposed SWRKF requires sufficient measurements over a period of time to identify outliers. In this paper, the maximum number of iterations $N_m$ and window length $L$ are respectively set as 2 and 10 to provide a trade off between the estimation accuracy and computational burden.

In practical applications, the noise covariance matrices are often unknown. Therefore, the algorithms are finally tested with inaccurate nominal noise covariance matrices, which are set as $\mathbf{\Sigma}'_w = \alpha\mathbf{\Sigma}_w$ and $\mathbf{\Sigma}'_v = \beta\mathbf{\Sigma}_v$ with different values of $\alpha$ and $\beta$. The ARMSEs of different filters are illustrated in Fig. 8. It shows that the estimation accuracy of the existing HKF, MCKF and RSTKF is obviously worse when $\alpha, \beta \neq 1$. On the contrary, the proposed filter outperforms the existing methods

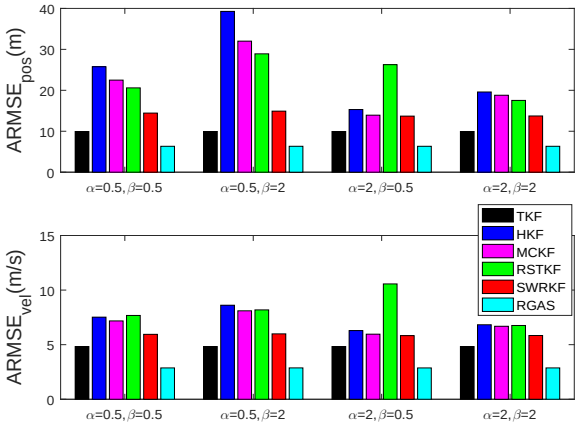



Fig. 8: ARMSEs for inaccurate nominal SNSMs and MNSMs.

in each case and is minimally affected by inaccurate nominal noise covariance matrices due to the adaptive estimates of the scale matrices, as discussed in Remark 1.

## V. Experimental Study

In this section, the proposed SWRKF and the existing NKF, HKF [10], MCKF [12], RSTKF [15] and RGAS [8] are verified by two experiments of manoeuvring vehicle tracking.

### A. Case 1

The data set of *Case 1* is provided by the Udacity course self-driving car [36], in which an autonomous vehicle is tracked by the lidar and radar sensors. The kinematics model of the vehicle is the same as that in the simulation, and the nominal SNCM is set as $\mathbf{\Sigma}_w = \kappa \begin{bmatrix} \frac{T^3}{3}\mathbf{I}_2 & \frac{T^2}{2}\mathbf{I}_2 \\ \frac{T^2}{2}\mathbf{I}_2 & T\mathbf{I}_2 \end{bmatrix}$, where the state noise intensity factor $\kappa$ is set as 0.5. The lidar measures the position of the vehicle, i.e. $\mathbf{z}_k^L = [x_k \quad y_k]^{\mathrm{T}}$, while the radar provides the range, angle and range rate of the vehicle, i.e. $\mathbf{z}_k^R = [\rho_k \quad \varphi_k \quad \dot{\rho}_k]$. The measurement models of lidar and radar are respectively expressed as

$$\mathbf{z}_k^L = \begin{bmatrix} \mathbf{I}_2 & \mathbf{0} \end{bmatrix} \mathbf{x}_k + \mathbf{v}_k^L \tag{30}$$

$$\mathbf{z}_k^R = \begin{bmatrix} \sqrt{x_k^2 + y_k^2} \\ \arctan\left(\frac{x_k}{y_k}\right) \\ \frac{x_k\dot{x}_k + y_k\dot{y}_k}{\sqrt{x_k^2+y_k^2}} \end{bmatrix} + \mathbf{v}_k^R \tag{31}$$

where the nominal covariance matrices of $\mathbf{v}_k^L$ and $\mathbf{v}_k^R$ are $\mathbf{\Sigma}_v^L = \mathrm{diag}[0.0025,\ 0.0025]m^2$ and $\mathbf{\Sigma}_v^R = \mathrm{diag}[0.09m^2,\ 0.05rad^2,\ 0.09m^2/s^2]$, respectively. The vehicle is observed by lidar and radar sensors in turn. The total experiment time is $61.2s$, and the discretization time $T = 0.05s$. All the algorithms are expanded to a non-linear system by using the first order Taylor expansion [33], and their parameters are set as Table IV.

The ground truth and the estimated trajectories are illustrated in Fig. 9. It can be seen from Fig. 9 that the vehicle undergoes frequent turns, which induces the periodic state

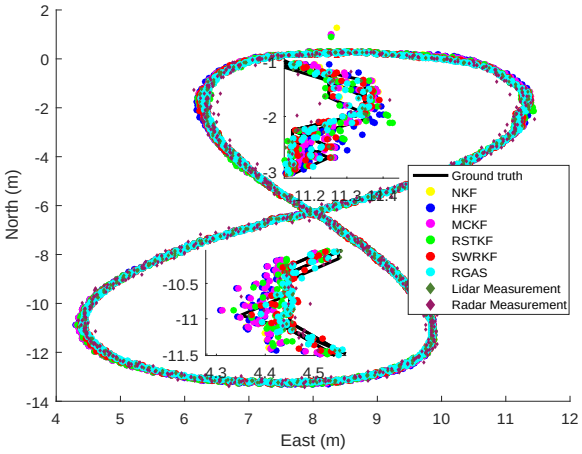



Fig. 9: Estimated trajectories and ground truth.

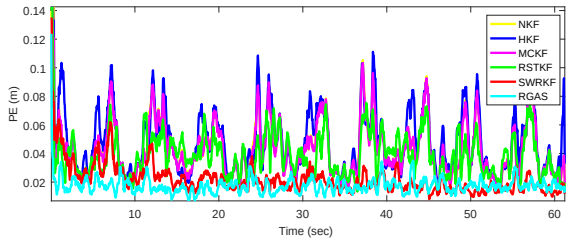



Fig. 10: PEs of different algorithms (smoothed using a 10-point moving average).

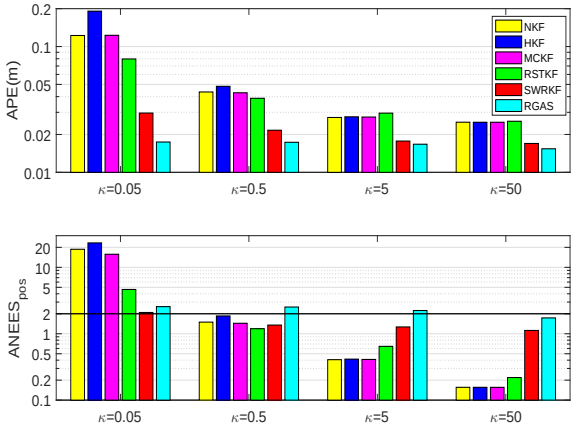



Fig. 11: Average PEs (APEs) and ANEESs for different noiminal SNCMs (The closer the $\text{ANEES}_{\text{pos}}$ is to position dimension 2, the better the consistency).

outliers in the kinematic model of uniform linear motion (26). The position errors (PEs, defined in (32)) of these algorithms are shown in Fig. 10, where the periodic PEs are caused by the periodic state outliers. In Fig. 10, the RGAS has the best estimation accuracy due to the use of all measurements. Except for the outline RGAS, the proposed SWRKF outperforms other filters in terms of estimation accuracy due to the better ability

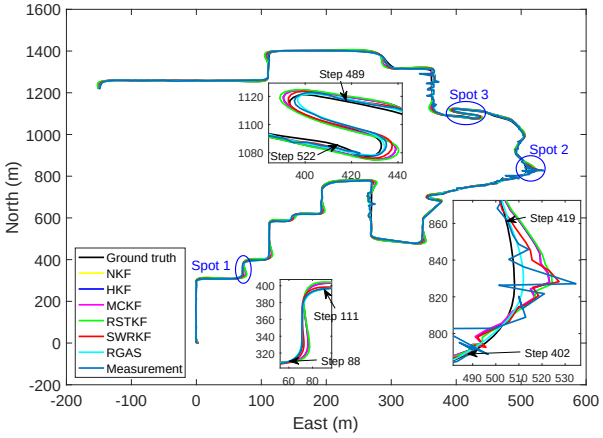



Fig. 12: True trajectory and GPS measurements.

to identify the state outliers.

$$\text{PE}_k = \sqrt{(x_k - \hat{x}_k)^2 + (y_k - \hat{y}_k)^2} \tag{32}$$

To verify the adaptive ability of the proposed SWRKF to estimate the scale matrices, we test the compared methods with different nominal SNCMs in Fig. 11, in which the state noise intensity factor $\kappa$ is set as $\{0.05, 0.5, 5, 50\}$. Note that the reference value of the ANEES for the position dimensions is 2. It can be observed from Fig. 11 that the existing NKF, HKF, MCKF and RSTKF can be overconfident or overconservative if the parameter $\kappa$ is set too small or too large. The value of $\kappa$ has great influence on the accuracy and consistency of the existing NKF, HKF, MCKF and RSTKF. On the contrary, the performance of the proposed SWRKF is better when the accuracy and consistency are considered together and less sensitive to $\kappa$ on account of the adaptive estimation of the SNSM.

### *B. Case 2*

In this subsection, we carry out a manoeuvring vehicle tracking experiment with real world data. A real vehicle is equipped with an inertial measurement unit (IMU) and two GPS receivers. One non-differential GPS receiver provides the measurements for the tracking. In addition, high-precision real-time kinematic GPS information is integrated with IMU information by a NovAtel synchronized position attitude navigation (SPAN) position system, which provides the reference position of the vehicle. Note that the reference position information is compressed in time to synchronize with the GPS measurement information whose frequency is 1Hz. The whole trajectory runs for 3177m with an average speed of 4.3m/s. Furthermore, the state space model, nominal MNCM and the parameter settings of different algorithms are consistent with the simulation, and the nominal SNCM is the same as *Case 1*. We plot the true trajectory and GPS measurements in Fig. 12 to obtain an intuitive sense of the distribution of outliers.

It is clear from Fig. 12 that the vehicle makes several turns during the experiment, especially in the early part of

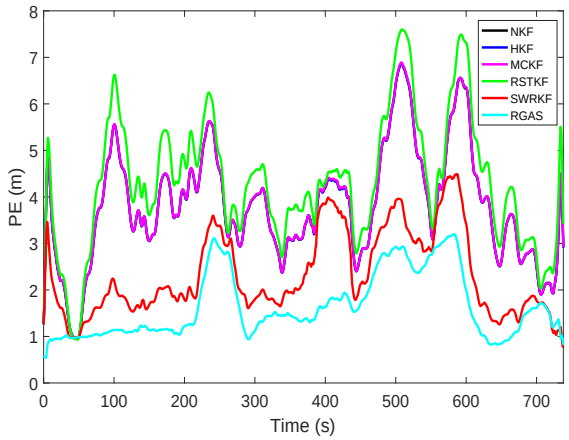



Fig. 13: PEs of different algorithms (smoothed using a 50-point moving average).

TABLE VII: APEs within different intervals.

| Intervals | NKF | HKF | MCKF | RSTKF | SWRKF | RGAS |
|---|---|---|---|---|---|---|
| Spot 1 $[88, 111]$ | 5.08 | 5.08 | 5.09 | 6.06 | 2.14 | 1.08 |
| Spot 2 $[402, 419]$ | 7.31 | 7.28 | 7.34 | 7.60 | 6.54 | 2.45 |
| Spot 3 $[489, 522]$ | 6.91 | 6.91 | 6.94 | 7.60 | 3.84 | 2.80 |
| Entirety $[0, 738]$ | 3.80 | 3.80 | 3.81 | 4.29 | 2.34 | 1.65 |

the trajectory, which gives rise to the state outliers in the kinematic model of uniform linear motion. Besides, it can be seen from the zoomed graph that the GPS measurement outliers are distributed around the true trajectory, which may be induced by the vehicle being occasionally obscured by trees or buildings. Next, we mark three typical spots with ellipses in Fig. 12 to make detailed explanations. At spots 1 and 3, the vehicle is experiencing diversions, which will induce state outliers. Nonetheless, the measurement quality is satisfactory. The performance of the proposed SWRKF is second only to the RGAS, which cannot be executed online. We attribute this performance to its excellent ability to identify the state outliers caused by the diversion accurately and rapidly when the measurements are accurate. The proposed filter thereby increases the corresponding SNCMs to facilitate the turn of the estimated trajectory, as shown in spots 1 and 3 in Fig. 12. However, the other algorithms take a long time to re-track the true trajectory but are left behind again in the next diversion, which results in their poor tracking performance.

At spot 2, the vehicle suffers from severe measurement outliers and slight state outliers, where the former are indicated by the significant deviation of the measurement curve from the real trajectory with drastic fluctuations. In this case, both the model predictions and measurements are not reliable. As a consequence, even the RGAS shows an unsatisfying performance which rationalizes the poor performance of other algorithms. Nevertheless, the proposed algorithm still tracks the true trajectory rapidly in a short time once the measurements become normal. The APEs at the three spots are listed in Table VII, which shows the same results as the above analyses.

The PE curves and APEs of all algorithms over the entire interval are shown in Fig. 13 and Table VII, respectively. It can be seen from Fig. 13 and Table VII that the RGAS certainly has the best estimation accuracy. Besides, the proposed SWRKF has the second best estimation accuracy among all compared algorithms, which verifies the superiority of the proposed SWRKF as compared with the existing cutting-edge outlier-robust filters.

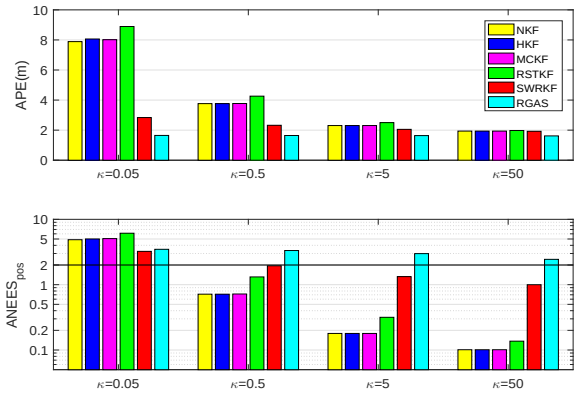



Fig. 14: APEs and ANEESs for different nominal SNCMs.

In the experiment, the setting of the nominal SNCM usually depends on engineering experience. So we test the dependence of the algorithms performance on the factor $\kappa$ in Fig. 14. As shown in Fig. 14, due to the adaptive estimation of the SNSM, the proposed SWRKF has better estimation accuracy and consistency than the existing outlier-robust filters in each case. However, the performance of the existing NKF, HKF, MCKF and RSTKF are greatly dependent on the selection of $\kappa$, which demonstrates the benefits of adaptive behavior in the proposed SWRKF. Note that the RGAS even has worse consistency than the proposed SWRKF when $\kappa$ is set as 0.05 or 0.5. This is because the assumption of stationary noise in the RGAS does not hold in the scenario of this experiment where the noise statistics are time-varying due to turns and continuous GPS block, as shown in Fig. 12. The proposed SWRKF, by contrast, is better suited to this scenario since it is able to estimate the time-varying SNSM by the use of the sliding window method.

## VI. Conclusions

In this paper, a novel SWRKF was developed. By modeling the state and measurement noises in a sliding window as Student's t distributed with random scale matrices, the states were jointly estimated with the scale matrices and ARVs in the sliding window using the VB method. The proposed SWRKF was validated by the simulations and experiments of manoeuvring target tracking, which show that the proposed SWRKF enjoys better estimation accuracy and consistency in response to state and measurement outliers on account of the superior performance of outlier identification but has higher computational complexity than the existing outlier-robust algorithms.

## APPENDICES

### A. Proof of Proposition 2

Using $\boldsymbol{\theta}_k = \mathbf{x}_{k-L:k}$ and (14) in (13), the iterative logarithmic posterior distribution $\log q^{(i+1)}(\mathbf{x}_{k-L:k})$ is formulated as

$$\begin{aligned}\log q^{(i+1)}(\mathbf{x}_{k-L:k}) &= -0.5(\mathbf{x}_{k-L} - \hat{\mathbf{x}}_{k-L|k-L})^{\mathrm{T}}\mathbf{P}_{k-L|k-L}^{-1} \\ &\times (\mathbf{x}_{k-L} - \hat{\mathbf{x}}_{k-L|k-L}) + \sum_{j=k-L+1}^{k}\Big[-0.5\mathrm{E}^{(i)}[\lambda_j] \\ &\times (\mathbf{z}_j - \mathbf{H}_j\mathbf{x}_j)^{\mathrm{T}}\mathrm{E}^{(i)}[\mathbf{R}_k^{-1}](\mathbf{z}_j - \mathbf{H}_j\mathbf{x}_j) - 0.5\mathrm{E}^{(i)}[\xi_j] \\ &\times (\mathbf{x}_j - \mathbf{F}_j\mathbf{x}_{j-1})^{\mathrm{T}}\mathrm{E}^{(i)}[\mathbf{Q}_k^{-1}](\mathbf{x}_j - \mathbf{F}_j\mathbf{x}_{j-1})\Big] + c_{\mathbf{x}_{k-L:k}}\end{aligned} \tag{33}$$

Taking exponents on both sides of (33) and utilizing the Gaussian PDF yield

$$\begin{aligned}q^{(i+1)}(\mathbf{x}_{k-L:k}) &\propto \mathrm{N}(\mathbf{x}_{k-L}; \hat{\mathbf{x}}_{k-L|k-L}, \mathbf{P}_{k-L|k-L}) \\ &\times \prod_{j=k-L+1}^{k}\Big[\mathrm{N}(\mathbf{z}_j; \mathbf{H}_j\mathbf{x}_j, \tilde{\mathbf{R}}_{j|k}^{(i)})\mathrm{N}(\mathbf{x}_j; \mathbf{F}_j\mathbf{x}_{j-1}, \tilde{\mathbf{Q}}_{j|k}^{(i)})\Big]\end{aligned} \tag{34}$$

Hence, Proposition 2 is confirmed.

### B. Proof of Proposition 3

Employing (14) in (13) and setting $\boldsymbol{\theta}_k = \mathbf{Q}_k$, we have

$$\begin{aligned}\log q^{(i+1)}(\mathbf{Q}_k) &= -0.5(\hat{y}_{k|k-L} + 2n + 3)\log|\mathbf{Q}_k| - \\ 0.5\mathrm{tr}&\left[\left(\sum_{j=k-L+1}^{k}\mathbf{A}_j^{(i+1)}\mathrm{E}^{(i)}[\xi_j] + \hat{\mathbf{Y}}_{k|k-L}\right)\mathbf{Q}_k^{-1}\right] + c_{\mathbf{Q}_k}\end{aligned} \tag{35}$$

Likewise, one can get $\log q^{(i+1)}(\mathbf{R}_k)$ by setting $\boldsymbol{\theta}_k = \mathbf{R}_k$. According to the IW PDF, Proposition 3 can be easily justified.

### C. Proof of Proposition 4

Utilizing $\boldsymbol{\theta}_k = \xi_{k-L+1:k}$ and (14) in (13), the iterative PDF of $\xi_{k-L+1:k}$ can be expressed as

$$\begin{aligned}\log q^{(i+1)}(\xi_{k-L+1:k}) &= \sum_{j=k-L+1}^{k}\Big\{[0.5(n+\omega) - 1]\log\xi_j \\ &-0.5\left[\omega + \mathrm{tr}(\mathbf{A}_j^{(i+1)}\mathrm{E}^{(i)}[\mathbf{Q}_k])\right]\xi_j\Big\} + c_{\xi_{j-L+1:k}}\end{aligned} \tag{36}$$

According to (36), the marginal PDF of $\xi_j$ is formulated as

$$q^{(i+1)}(\xi_j) = \mathrm{G}(\xi_j; \hat{a}_{j|k}^{(i+1)}, \hat{b}_{j|k}^{(i+1)}) \tag{37}$$

Similarly, the marginal PDF of $\lambda_j$ can be updated as Gamma. Therefore, Proposition 4 is verified.

## REFERENCES


[1] A. H. Mohamed and K. P. Schwarz, "Adaptive Kalman filtering for INS/GPS," in *J. Geod.*, vol. 73, no. 4, pp. 193–203, May 1999.

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

[3] J. Zhao and L. Mili, "A theoretical framework of robust H-infinity unscented Kalman filter and its application to power system dynamic state estimation," *IEEE Trans. Signal Process.*, vol. 67, no. 10, pp. 2734–2746, May 2019.

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

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

[6] G. Wang, Y. Zhang, and X. Wang, "Maximum correntropy Rauch-Tung-Striebel smoother for nonlinear and non-Gaussian systems," *IEEE Transactions on Automatic Control*, to be published, doi: 10.1109/TAC.2020.2997315.

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

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

[9] Y. Huang, Y. Zhang, Y. Zhao, L. Mihaylova, and J. Chambers, "Robust Rauch-Tung-Striebel Smoothing Framework for Heavy-Tailed and/or Skew Noises," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 56, no. 1, pp. 415–441, Feb. 2020.

[10] P. J. Huber, *Robust Statistics*. International Encyclopedia of Statistical Science. Springer, Berlin, Heidelberg, 2011.

[11] M. 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, 2010.

[12] B. Chen, X. Liu, H. Zhao, and J. C. Principe, "Maximum correntropy Kalman filter," *Automatica*, vol. 76, no. 2, pp. 70–77, Feb. 2017.

[13] B. Chen, L. Dang, Y. Gu, N. Zheng, and J. C. Principe, "Minimum error entropy Kalman filter," *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, pp. 1–11, 2019.

[14] S. Fakoorian, A. Santamaria-Navarro, B. T. Lopez, D. Simon and A. Agha-Mohammadi, "Towards Robust State Estimation by Boosting the Maximum Correntropy Criterion Kalman Filter With Adaptive Behaviors," *IEEE Robotics and Automation Letters*, vol. 6, no. 3, pp. 5469–5476, Jul. 2021.

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

[16] Y. Huang, Y. Zhang, Y. Zhao, P. Shi, and J. A. 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, Jun. 2021.

[17] Y. Huang, M. Bai and Y. Zhang, "A novel multiple-outlier-robust Kalman filter," *Front. Inform. Technol. Electron. Eng.*, to be published, 2021.

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

[19] S. Li, H. Wang, and T. Chai, "A t-distribution based particle filter for target tracking," in *Proc. Amer. Control Conf.*, pp. 2191–2196, Jun. 2006.

[20] M. Roth, E. Özkan, and F. Gustafsson, "A Student's t filter for heavy-tailed process and measurement noise," in *Proc. IEEE Int. Conf. Acoust., Speech*, May 2013, pp. 5770–5774.

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

[22] 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.

[23] Y. Huang, Y. Zhang, Y. Zhao and J. 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, Jul. 2019.

[24] Y. Huang, Y. Zhang, P. Shi, Z. Wu, J. Qian and J. Chambers, "Robust Kalman Filters Based on Gaussian Scale Mixture Distributions With

Application to Target Tracking," *IEEE Transactions on Systems, Man and Cybernetics: Systems*, vol. 49, no. 10, pp. 2082–2096, Oct. 2019.
[25] H. Zhu, G. Zhang, Y. Li and H. Leung, "A novel robust Kalman filter with unknown non-stationary heavy-tailed noise," *Automatica*, vol, 127, 2021, 109511.
[26] 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 Autom. Control*, vol. 67, no. 1, pp. 374–381, Jan. 2022.
[27] Y. Huang, F. Zhu, G. Jia and Y. Zhang, "A Slide Window Variational Adaptive Kalman Filter," *IEEE Trans. Circuits Syst. II, Exp. Briefs*, vol. 67, no. 12, pp. 3552–3556, Dec. 2020.
[28] A. Fisch, I. Eckley and P. Fearnhead, "Innovative and additive outlier robust Kalman filtering with a robust particle filter," *IEEE Transactions on Signal Processing*, to be published, 2021.
[29] D. Tzikas, A. Likas, and N. Galatsanos, "The variational approximation for Bayesian inference," *IEEE Signal Processs Mag.*, vol. 25, no. 6, pp. 131–146, Nov. 2008.
[30] C. M. Bishop, *Pattern Recognition and Machine Learning.* Springer, 2007.
[31] Y. Huang, Y. Zhang, Z. Wu, N. Li, and J. Chambers, "A novel adaptive Kalman filter with inaccurate process and measurement noise covariance matrices," *IEEE Trans. Automat. Control*, vol. 63, no. 2, pp. 594–601, Feb. 2018.
[32] J. Kokkala, A. Solin, and S. Särkkä, "Expectation maximization based parameter estimation by sigma-point and particle smoothing," in *Proc. 17th Int. Conf. Inf. Fusion, Salamanca*, Spain, Jul. 2014, pp. 1–8.
[33] S. Särkkä, *Bayesian Filtering and Smoothing*. New York, NY, USA: Cambridge Univ. Press, 2013.
[34] Y. Bar-Shalom, X. Li, and T. Kirubarajan, *Estimation with applications to tracking and navigation: theory algorithms and software*. John Wiley & Sons, 2004.
[35] C. Goutte and E. Gaussier, "A Probabilistic Interpretation of Precision, Recall and F-Score, with Implication for Evaluation," in *27th European Conference on Information Retrieval Research*, Mar. 2005, pp. 345–359.
[36] *Udacitys Self-Driving Car Simulator*, Udacity, Mountain View, CA, USA, 2017. [Online]. Available: https://github.com/udacity/self-drivingcar-sim