



**VICTORIA UNIVERSITY**  
MELBOURNE AUSTRALIA

## *A Novel Outlier-Robust Kalman Filtering Framework Based on Statistical Similarity Measure*

This is the Published version of the following publication

Huang, Y, Zhang, Y, Zhao, Y, Shi, Peng and Chambers, JA (2021) A Novel Outlier-Robust Kalman Filtering Framework Based on Statistical Similarity Measure. IEEE Transactions on Automatic Control, 66 (6). pp. 2677-2692. ISSN 0018-9286

The publisher's official version can be found at  
<https://ieeexplore.ieee.org/document/9146725>  
Note that access to this version may require subscription.

Downloaded from VU Research Repository <https://vuir.vu.edu.au/44679/>

# A Novel Outlier-Robust Kalman Filtering Framework Based on Statistical Similarity Measure

Yulong Huang<sup>1</sup>, Member, IEEE, Yonggang Zhang<sup>2</sup>, Senior Member, IEEE,  
Yuxin Zhao<sup>1</sup>, Senior Member, IEEE, Peng Shi<sup>3</sup>, Fellow, IEEE,  
and Jonathon A. Chambers<sup>4</sup>, Fellow, IEEE

**Abstract**—In this article, a statistical similarity measure is introduced to quantify the similarity between two random vectors. The measure is, then, employed to develop a novel outlier-robust Kalman filtering framework. The approximation errors and the stability of the proposed filter are analyzed and discussed. To implement the filter, a fixed-point iterative algorithm and a separate iterative algorithm are given, and their local convergent conditions are also provided, and their comparisons have been made. In addition, selection of the similarity function is considered, and four exemplary similarity functions are established, from which the relations between our new method and existing outlier-robust Kalman filters are revealed. Simulation examples are used to illustrate the effectiveness and potential of the new filtering scheme.

**Index Terms**—Heavy-tailed noise, Kalman filter, outliers, statistical similarity measure, separate iterative algorithm.

## I. INTRODUCTION

THE Kalman filter is best-known as an optimal recursive state estimator in the sense of minimum variance for a linear system with Gaussian noises. In view of this, the Kalman filter has been widely used in positioning, navigation, target tracking, and signal processing [1]. The estimation accuracy of the Kalman filter degrades dramatically for a linear system with

Manuscript received July 3, 2019; revised November 28, 2019 and May 21, 2020; accepted July 19, 2020. Date of publication July 23, 2020; date of current version May 27, 2021. This work was supported in part by the National Natural Science Foundation of China under Grant 61903097 and Grant 61773133 and in part by the Fundamental Research Funds for the Central Universities under Grant 3072020CF0404 and Grant 3072020GIP0409. Recommended by Associate Editor D. Efimov. (Corresponding author: Yonggang Zhang.)

Yulong Huang, Yonggang Zhang, and Yuxin Zhao are with the College of Intelligent Systems Science and Engineering, the Engineering Research Center of Navigation Instruments, Harbin Engineering University, Harbin 150001, China (e-mail: heuedu@163.com; zhangyg@hrbeu.edu.cn; zhaoyuxin@hrbeu.edu.cn).

Peng Shi is with the School of Electrical and Electronic Engineering, The University of Adelaide, Adelaide, SA 5005, Australia (e-mail: peng.shi@adelaide.edu.au).

Jonathon A. Chambers is with the Department of Engineering, University of Leicester, Leicester LE1 7RH, U.K. (e-mail: jonathon.chambers@newcastle.ac.uk).

Color versions of one or more of the figures in this article are available online at <https://ieeexplore.ieee.org>.

Digital Object Identifier 10.1109/TAC.2020.3011443

non-Gaussian heavy-tailed noises, which are often induced by state and measurement outliers from external interference or unreliable sensors [2].

For such non-Gaussian filtering problems, the particle filter (PF) can achieve an approximate state estimate by modeling the noises as non-Gaussian heavy-tailed distributed and approximating the posterior probability density function (pdf) as a set of weighted random samples using the sequential Monte Carlo sampling technique [3]–[5]. The Gaussian sum filter (GSF) can be also used to address the non-Gaussian filtering problem by running a group of Kalman filters, in which a finite sum of Gaussian distributions are employed to model non-Gaussian noises, and the posterior pdf can be updated as a weighted sum of Gaussian pdfs [6]. Moreover, the interacting multiple model (IMM) filter is a promising approach to address model uncertainty, in which several subfilters are performed in parallel based on the preselected model set and the corresponding model transition probability matrix, and the subfilters interact with each other by fusing the state estimates and corresponding estimation error covariance matrices based on the recursive estimates of the mode probabilities [7]. Normally, the performances of the PF, GSF, and IMM filter rely heavily on the preselected distributions to model non-Gaussian state and measurement noises. Unfortunately, it is very difficult to select accurate distributions to model unknown and time-varying non-Gaussian noises, which are often induced by outliers so that the estimation accuracy of the PF, GSF, and IMM filter degrades significantly when inaccurate or even wrong noise distributions are used. The contribution of this article is, therefore, to provide a unified theoretical framework to solve the non-Gaussian filtering problem for a linear state-space model with unknown non-Gaussian heavy-tailed noises.

A large number of outlier-robust Kalman filters have been proposed to achieve a tradeoff between estimation accuracy and computational complexity. As a classic robust regression technique, the M-estimator is robust to measurement outliers and has been successfully extended to the Kalman filter setting [8]. By employing the influence function approach on the prediction error and residual error, many outlier-robust Kalman filters have been proposed based on the M-estimate method [9]. The Huber Kalman filter (HKF) is the most famous extension of the M-estimator to the Kalman filter setting, which utilizes a combined

$l_1$  and  $l_2$  norm as a robust cost function, and a generalized robust maximum likelihood estimate is achieved by minimizing the Huber cost function [10]. The maximum correntropy Kalman filter (MCKF) is an alternative method to handle state and measurement outliers. The sum of the Gaussian kernel functions of the prediction error and residual error are selected as the robust cost function, and many MCKFs have been proposed by maximizing such robust cost functions [11]–[13]. Motivated by the fact that the outlier contaminated state and measurement noises often have non-Gaussian heavy-tailed distributions, many outlier-robust filters have been proposed by modeling the state and measurement noises as Student's  $t$  distributed [14]–[23]. These robust filters can be divided into two categories: Student's  $t$  filter and robust Student's  $t$  based Kalman filter (RSTKF). In the Student's  $t$  filter, the posterior pdf is approximated as a Student's  $t$  pdf with fixed degrees of freedom (DOF) parameter based on the Bayesian rule [14]–[16], nevertheless, in the RSTKF, the posterior pdf is approximated as a Gaussian pdf based on the variational Bayesian (VB) approach [22], [23]. For a linear system with moderately heavy-tailed state and measurement noises, the RSTKF can achieve better estimation accuracy than the Student's  $t$  filter but at the cost of higher computational complexity. Also, the adaptive Kalman filter based on the VB approach can to some extent address the heavy-tailed state and measurement noises induced by outliers based on the adaptive modifications of the one-step prediction error covariance matrix and measurement noise covariance matrix [24].

Although the HKF, MCKF, and RSTKF can all achieve better estimation accuracy than the standard Kalman filter for a linear system with outlier corrupted state and measurement noises, their interrelationships have as yet not been revealed. Except for the Gaussian scale mixture distributions, the design method of the existing RSTKF cannot be generalized to a general non-Gaussian heavy-tailed distribution, which limits its further improvement. In addition, it is worth studying whether the estimation accuracy of the existing HKF, MCKF, and RSTKF can be further improved. An advanced outlier-robust Kalman filtering framework is, therefore, required to reveal the interrelationships between the existing HKF, MCKF, and RSTKF and further improve the estimation accuracy of the existing HKF, MCKF, and RSTKF.

In this article, a statistical similarity measure (SSM) is introduced to quantify the similarity between two random vectors. The measure is then used to develop a novel outlier-robust Kalman filtering framework, in which lower bounds of the SSM between the state vector and the predicted state vector and that between the measurement vector and the predicted measurement vector are maximized, and the posterior pdf is approximated as Gaussian. To illustrate the effectiveness of the proposed framework, the effects of the approximation errors on the proposed framework are analyzed in detail, and the numerical and filtering stabilities of the proposed framework are also discussed. To implement the proposed framework, the fixed-point iterative algorithm and the separate iterative algorithm are proposed, and their local convergence conditions are also provided and compared. Furthermore, the selections of the similarity functions are presented, and four exemplary similarity functions are provided,

from which the relations between the proposed method and the existing outlier-robust Kalman filters are revealed. Simulation results of a manoeuvring target tracking example illustrate that by selecting appropriate similarity functions, the proposed filters have improved estimation accuracy but higher computational complexities than the existing state-of-the-art filters.

The remainder of this article is organized as follows. In Section II, a novel SSM is proposed to quantify the similarity between two random vectors. In Section III, a novel outlier-robust Kalman filtering framework based on the SSM is proposed, and its error analyses and stability discussions are provided. In Section IV, two novel iterative algorithms for proposed outlier-robust Kalman filtering framework are proposed, and their local convergence conditions are provided. In Section V, we present the selections of the similarity functions. In Section VI, simulation results of a manoeuvring target tracking example and comparisons with existing filters are given. Conclusions are drawn in Section VII.

## II. PROPOSED SSM

In this article, a novel SSM is proposed to quantify the similarity between two random vectors. The proposed SSM  $s(\mathbf{x}, \mathbf{y})$  for random vectors  $\mathbf{x}$  and  $\mathbf{y}$  is defined as follows:

$$s(\mathbf{x}, \mathbf{y}) = \mathbb{E} [f(\|\mathbf{x} - \mathbf{y}\|^2)] = \iint f(\|\mathbf{x} - \mathbf{y}\|^2) p(\mathbf{x}, \mathbf{y}) d\mathbf{x} d\mathbf{y} \quad (1)$$

where  $\mathbb{E}[\cdot]$  denotes the expectation operation,  $\|\cdot\|$  denotes the Euclidean norm, and  $p(\mathbf{x}, \mathbf{y})$  denotes the joint pdf of random vectors  $\mathbf{x}$  and  $\mathbf{y}$ . In this article, the scalar function  $f(\cdot)$  is named as the similarity function, and it must satisfy the following three conditions.

- 1) *Condition 1:*  $f(\cdot)$  is a continuous function defined on  $[0, +\infty)$ .
- 2) *Condition 2:*  $f(\cdot)$  is a strictly monotonically decreasing function:  $\dot{f}(t) < 0$  for  $t \in [0, +\infty)$ .
- 3) *Condition 3:* The second derivative of  $f(\cdot)$  is nonnegative:  $\ddot{f}(t) \geq 0$  for  $t \in [0, +\infty)$ .

The monotonically decreasing property of the similarity function  $f(\cdot)$  can guarantee that the proposed SSM  $s(\mathbf{x}, \mathbf{y})$  is increasing as the distance between  $\mathbf{x}$  and  $\mathbf{y}$  decreases. As a result, the proposed SSM conforms to the usual definition of a similarity measure that is in some sense the inverse of distance metrics. The higher similarity between random vectors  $\mathbf{x}$  and  $\mathbf{y}$ , the larger SSM becomes. The proposed SSM has some basic properties as follows.

- 1) *Property 1:* The proposed SSM is symmetric:  $s(\mathbf{x}, \mathbf{y}) = s(\mathbf{y}, \mathbf{x})$ .
- 2) *Property 2:* The proposed SSM achieves the maximum value if and only if the random vectors  $\mathbf{x}$  and  $\mathbf{y}$  are identical.
- 3) *Property 3:* The proposed SSM includes all the even order moments of the random vector  $\mathbf{x} - \mathbf{y}$ :  $s(\mathbf{x}, \mathbf{y}) = \sum_{l=0}^{+\infty} \frac{f^{(l)}(0)}{l!} \mathbb{E}[\|\mathbf{x} - \mathbf{y}\|^{2l}]$ , if the Taylor expansion of the similarity function  $f(t)$  exists when  $t \geq 0$ .

Property 1 can be easily verified using the definition of the proposed SSM in (1), and Property 3 can be directly derived by

exploiting the Taylor series expansion of the similarity function  $f(t)$  at  $t = 0$ . Next, we will prove Property 2 in the following Theorem 1.

*Theorem 1:* If the similarity function  $f(\cdot)$  satisfies Condition 2, then Property 2 holds.

*Proof:* See Appendix A.  $\blacksquare$

Property 2 guarantees that the proposed SSM has a maximum point, i.e.,  $\mathbf{x} = \mathbf{y}$ . It is noteworthy that the proposed SSM  $s(\mathbf{x}, \mathbf{y})$  is an upper bound of  $f(\mathbb{E}[\|\mathbf{x} - \mathbf{y}\|^2])$ , and they have the same maximum point, i.e.,  $\mathbf{x} = \mathbf{y}$ .

*Remark 1:* The proposed SSM is a generalized similarity measure between two random vectors and encompasses existing similarity measures. For example, the proposed SSM  $s(\mathbf{x}, \mathbf{y})$  is the negative mean squared error (MSE) between random vectors  $\mathbf{x}$  and  $\mathbf{y}$  when the similarity function is chosen as  $f(t) = -t$ . The proposed SSM  $s(\mathbf{x}, \mathbf{y})$  is the correntropy between random vectors  $\mathbf{x}$  and  $\mathbf{y}$  when the similarity function is selected as  $f(t) = \exp(-\frac{1}{2\sigma^2}t)$  [11], [13]. More interestingly, the generalized correntropy [25] between random vectors  $\mathbf{x}$  and  $\mathbf{y}$  is also a special case of the proposed SSM  $s(\mathbf{x}, \mathbf{y})$  when the similarity function is chosen as  $f(t) = \frac{\alpha}{2\beta\Gamma(1/\alpha)} \exp(-t^{\frac{\alpha}{2}}\beta^{-\alpha})$  and the shape parameter satisfies the constraint  $0 < \alpha \leq 2$ .

*Remark 2:* For the proposed SSM, except for the first two conditions, the similarity function  $f(\cdot)$  has to satisfy the third condition. The third condition not only can facilitate the design of an outlier-robust Kalman filter but also can guarantee robustness to outliers and local convergence of fixed point iterations, as shown in Sections III, IV, and V.

The proposed SSM can be used in Bayesian inference. Different SSMs are achieved when different similarity functions  $f(\cdot)$  are selected, based on which different state estimates can be obtained by maximizing the corresponding SSMs. Next, we will propose an outlier-robust Kalman filtering framework based on the proposed SSM.

### III. PROPOSED OUTLIER-ROBUST KALMAN FILTERING FRAMEWORK BASED ON SSM

Consider a linear dynamical system described by a linear discrete-time state-space model as follows:

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

where  $k$  is the discrete time index,  $\mathbf{x}_k \in \mathbb{R}^n$  is the state vector,  $\mathbf{z}_k \in \mathbb{R}^m$  is the measurement vector,  $\mathbf{F}_k \in \mathbb{R}^{n \times n}$  and  $\mathbf{H}_k \in \mathbb{R}^{m \times n}$  are, respectively, the known state transition matrix and measurement matrix, and  $\mathbf{w}_k \in \mathbb{R}^n$  and  $\mathbf{v}_k \in \mathbb{R}^m$  are, respectively, state and measurement noise vectors. In this article, the state and measurement noises are assumed to have non-Gaussian distributions that are, respectively, induced by state and measurement outliers.

#### A. Design of Outlier-Robust Kalman Filtering Framework

Similar to the standard Kalman filter, the proposed outlier-robust Kalman filtering framework is also composed of time and measurement updates. In the time update, the one-step predicted state vector  $\hat{\mathbf{x}}_{k|k-1}$  and corresponding nominal prediction error

covariance matrix  $\mathbf{P}_{k|k-1}$  are calculated as follows:

$$\begin{cases} \hat{\mathbf{x}}_{k|k-1} = \mathbf{F}_k \hat{\mathbf{x}}_{k-1|k-1} \\ \mathbf{P}_{k|k-1} = \mathbf{F}_k \mathbf{P}_{k-1|k-1} \mathbf{F}_k^T + \mathbf{Q}_k \end{cases} \quad (3)$$

where  $\hat{\mathbf{x}}_{k-1|k-1}$  and  $\mathbf{P}_{k-1|k-1}$  are, respectively, the state estimate and corresponding estimation error covariance matrix at time  $k-1$ , and  $\mathbf{Q}_k$  denotes the nominal state noise covariance matrix. Note that  $\mathbf{P}_{k|k-1}$  is called the nominal prediction error covariance matrix since the used nominal state noise covariance matrix  $\mathbf{Q}_k$  is inaccurate in the presence of state outliers.

In the measurement update, we aim to achieve an approximate posterior pdf  $q^*(\mathbf{x}_k) \approx p(\mathbf{x}_k | z_{1:k})$  through maximizing the sum of the SSM between  $\mathbf{S}_{k|k-1}^{-1} \mathbf{x}_k$  and  $\mathbf{S}_{k|k-1}^{-1} \hat{\mathbf{x}}_{k|k-1}$  and the SSM between  $\mathbf{S}_{\mathbf{R}_k}^{-1} \mathbf{z}_k$  and  $\mathbf{S}_{\mathbf{R}_k}^{-1} \mathbf{H}_k \mathbf{x}_k$ , i.e.,

$$q^*(\mathbf{x}_k) = \arg \max_{q(\mathbf{x}_k)} \left\{ s \left( \mathbf{S}_{k|k-1}^{-1} \mathbf{x}_k, \mathbf{S}_{k|k-1}^{-1} \hat{\mathbf{x}}_{k|k-1} \right) + s \left( \mathbf{S}_{\mathbf{R}_k}^{-1} \mathbf{z}_k, \mathbf{S}_{\mathbf{R}_k}^{-1} \mathbf{H}_k \mathbf{x}_k \right) \right\} \quad (4)$$

where  $\mathbf{S}_{k|k-1}$  and  $\mathbf{S}_{\mathbf{R}_k}$  are, respectively, the square roots of the predicted error covariance matrix  $\mathbf{P}_{k|k-1}$  and the nominal measurement noise covariance matrix  $\mathbf{R}_k$ , i.e.,

$$\mathbf{P}_{k|k-1} = \mathbf{S}_{k|k-1} \mathbf{S}_{k|k-1}^T \quad \mathbf{R}_k = \mathbf{S}_{\mathbf{R}_k} \mathbf{S}_{\mathbf{R}_k}^T \quad (5)$$

where  $\mathbf{R}_k$  denotes the nominal measurement noise covariance matrix. Note that the 1-step statistical similarity measure based cost function in (4) is sufficient for designing an outlier-robust Kalman filtering framework in this article, in which the one-step predicted state vector  $\hat{\mathbf{x}}_{k|k-1}$  and measurement vector  $\mathbf{z}_k$  at the current time are used to construct the cost function. The idea of this article can be extended to derive an outlier-robust Kalman smoothing framework based on a multiple-steps SSM cost function by resorting to some standard techniques for designing a maximum a posterior estimator in a Bayesian framework [26].

Considering that the one-step predicted state vector  $\hat{\mathbf{x}}_{k|k-1}$  and measurement vector  $\mathbf{z}_k$  are known and deterministic quantities in the measurement update of the Kalman filter, the maximization problem in (4) can be reformulated as

$$q^*(\mathbf{x}_k) = \arg \max_{q(\mathbf{x}_k)} \left\{ \int f_x(\|\mathbf{S}_{k|k-1}^{-1}(\mathbf{x}_k - \hat{\mathbf{x}}_{k|k-1})\|^2) q(\mathbf{x}_k) d\mathbf{x}_k + \int f_z(\|\mathbf{S}_{\mathbf{R}_k}^{-1}(\mathbf{z}_k - \mathbf{H}_k \mathbf{x}_k)\|^2) q(\mathbf{x}_k) d\mathbf{x}_k \right\} \quad (6)$$

where  $f_x(\cdot)$  and  $f_z(\cdot)$  denote the similarity functions of state and measurement models, respectively.

It is very difficult to achieve an optimal solution for the maximization problem in (6) since both the explicit form of posterior pdf  $q(\mathbf{x}_k)$  and closed form solutions for the integrals are unavailable. To solve this problem, we propose to achieve an approximate  $q^*(\mathbf{x}_k)$  by approximating the posterior pdf  $q(\mathbf{x}_k)$  as Gaussian and maximizing the lower bound of the cost function.

First, we propose to approximate the posterior pdf  $q(\mathbf{x}_k)$  as a Gaussian pdf, i.e.,

$$q(\mathbf{x}_k) \approx \mathcal{N}(\mathbf{x}_k; \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k) \quad (7)$$

where  $\boldsymbol{\mu}_k$  and  $\boldsymbol{\Sigma}_k$  are, respectively, the mean vector and covariance matrix of the posterior pdf  $q(\mathbf{x}_k)$ .

Substituting (7) in (6), the maximization problem with respect to the posterior pdf  $q(\mathbf{x}_k)$  in (6) is approximately transformed into a maximization problem with respect to the posterior mean vector and covariance matrix, i.e.,

$$\begin{aligned} \{\boldsymbol{\mu}_k^*, \boldsymbol{\Sigma}_k^*\} \approx \arg \max_{\{\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k\}} & \left\{ \int f_x(\|\mathbf{S}_{k|k-1}^{-1}(\mathbf{x}_k - \hat{\mathbf{x}}_{k|k-1})\|^2) \right. \\ & \times \mathcal{N}(\mathbf{x}_k; \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k) d\mathbf{x}_k + \int f_z(\|\mathbf{S}_{\mathbf{R}_k}^{-1}(\mathbf{z}_k - \mathbf{H}_k \mathbf{x}_k)\|^2) \mathcal{N} \\ & \left. \times (\mathbf{x}_k; \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k) d\mathbf{x}_k \right\}, \quad \text{s.t. } \boldsymbol{\Sigma}_k > \mathbf{0} \end{aligned} \quad (8)$$

where  $\boldsymbol{\mu}_k^*$  and  $\boldsymbol{\Sigma}_k^*$  denote the optimal posterior mean vector and covariance matrix. Next, the cost function in (8) will be approximated as its lower bound, and then an approximate solution can be obtained.

*Theorem 2:* If the similarity functions satisfy Condition 3, i.e.,  $\ddot{f}_x(t) \geq 0$  and  $\ddot{f}_z(t) \geq 0$  for  $t \in [0, +\infty)$ , the maximization problem in (8) can be transformed as follows:

$$\{\boldsymbol{\mu}_k^*, \boldsymbol{\Sigma}_k^*\} \approx \arg \max_{\{\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k\}} J_1(\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k) \quad \text{s.t. } \boldsymbol{\Sigma}_k > \mathbf{0} \quad (9)$$

where  $J_1(\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)$  is the lower bound of the cost function in (8) and formulated as

$$J_1(\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k) = f_x\left(\text{tr}(\mathbf{A}_k \mathbf{P}_{k|k-1}^{-1})\right) + f_z\left(\text{tr}(\mathbf{B}_k \mathbf{R}_k^{-1})\right) \quad (10)$$

where  $\text{tr}(\cdot)$  denotes the trace operation of a matrix, and  $\mathbf{A}_k$  and  $\mathbf{B}_k$  are, respectively, given by

$$\begin{aligned} \mathbf{A}_k &= \int (\mathbf{x}_k - \hat{\mathbf{x}}_{k|k-1})(\mathbf{x}_k - \hat{\mathbf{x}}_{k|k-1})^T \mathcal{N}(\mathbf{x}_k; \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k) d\mathbf{x}_k \\ &= \boldsymbol{\Sigma}_k + (\boldsymbol{\mu}_k - \hat{\mathbf{x}}_{k|k-1})(\boldsymbol{\mu}_k - \hat{\mathbf{x}}_{k|k-1})^T \end{aligned} \quad (11)$$

$$\begin{aligned} \mathbf{B}_k &= \int (\mathbf{z}_k - \mathbf{H}_k \mathbf{x}_k)(\mathbf{z}_k - \mathbf{H}_k \mathbf{x}_k)^T \mathcal{N}(\mathbf{x}_k; \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k) d\mathbf{x}_k \\ &= (\mathbf{z}_k - \mathbf{H}_k \boldsymbol{\mu}_k)(\mathbf{z}_k - \mathbf{H}_k \boldsymbol{\mu}_k)^T + \mathbf{H}_k \boldsymbol{\Sigma}_k \mathbf{H}_k^T. \end{aligned} \quad (12)$$

*Proof:* See Appendix B.  $\blacksquare$

Define four auxiliary variables  $\xi_k$ ,  $\lambda_k$ ,  $\tilde{\xi}_k$ , and  $\tilde{\lambda}_k$  as follows:

$$\begin{cases} \xi_k \triangleq -2\dot{f}_x\left(\text{tr}(\mathbf{A}_k \mathbf{P}_{k|k-1}^{-1})\right), & \lambda_k \triangleq -2\dot{f}_z\left(\text{tr}(\mathbf{B}_k \mathbf{R}_k^{-1})\right) \\ \tilde{\xi}_k \triangleq 2\ddot{f}_x\left(\text{tr}(\mathbf{A}_k \mathbf{P}_{k|k-1}^{-1})\right), & \tilde{\lambda}_k \triangleq 2\ddot{f}_z\left(\text{tr}(\mathbf{B}_k \mathbf{R}_k^{-1})\right) \end{cases} \quad (13)$$

and three auxiliary matrices as follows:

$$\begin{cases} \boldsymbol{\Delta}_{\boldsymbol{\mu}_k}(\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k) = \frac{\partial J_1(\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)}{\partial \boldsymbol{\mu}_k} \\ \boldsymbol{\Delta}_{\boldsymbol{\Sigma}_k}(\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k) = \frac{\partial J_1(\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)}{\partial \boldsymbol{\Sigma}_k} \\ \boldsymbol{\Theta}_{\boldsymbol{\mu}_k}(\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k) = \frac{\partial^2 J_1(\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)}{\partial \boldsymbol{\mu}_k \partial \boldsymbol{\mu}_k^T} \end{cases} \quad (14)$$

where  $\boldsymbol{\Delta}_{\boldsymbol{\mu}_k}(\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)$  and  $\boldsymbol{\Delta}_{\boldsymbol{\Sigma}_k}(\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)$  denote the Jacobian matrices of the approximate cost function  $J_1(\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)$  with respect to the posterior mean vector and covariance matrix, respectively, and  $\boldsymbol{\Theta}_{\boldsymbol{\mu}_k}(\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)$  denotes the Hessian matrix of the approximate cost function  $J_1(\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)$  with respect to the posterior mean vector.

*Theorem 3:* The optimal solution  $\boldsymbol{\mu}_k^*$  of the approximate cost function  $J_1(\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)$  is formulated as follows:

$$\boldsymbol{\mu}_k^* = \hat{\mathbf{x}}_{k|k-1} + \tilde{\mathbf{K}}_k^*(\mathbf{z}_k - \mathbf{H}_k \hat{\mathbf{x}}_{k|k-1}) \quad (15)$$

$$\tilde{\mathbf{K}}_k^* = \tilde{\mathbf{P}}_{k|k-1}^* \mathbf{H}_k^T \left( \mathbf{H}_k \tilde{\mathbf{P}}_{k|k-1}^* \mathbf{H}_k^T + \tilde{\mathbf{R}}_k^* \right)^{-1} \quad (16)$$

where  $\tilde{\mathbf{P}}_{k|k-1}^*$  and  $\tilde{\mathbf{R}}_k^*$  are, respectively, the modified one-step prediction error covariance matrix and measurement noise covariance matrix given by

$$\tilde{\mathbf{P}}_{k|k-1}^* = \mathbf{P}_{k|k-1} / \xi_k^*, \quad \tilde{\mathbf{R}}_k^* = \mathbf{R}_k / \lambda_k^* \quad (17)$$

and the auxiliary parameters  $\xi_k^*$  and  $\lambda_k^*$  are given by

$$\xi_k^* = -2\dot{f}_x\left(\text{tr}(\mathbf{A}_k^* \mathbf{P}_{k|k-1}^{-1})\right), \quad \lambda_k^* = -2\dot{f}_z\left(\text{tr}(\mathbf{B}_k^* \mathbf{R}_k^{-1})\right) \quad (18)$$

and the auxiliary parameters  $\mathbf{A}_k^*$  and  $\mathbf{B}_k^*$  are given by

$$\mathbf{A}_k^* = \boldsymbol{\Sigma}_k^* + (\boldsymbol{\mu}_k^* - \hat{\mathbf{x}}_{k|k-1})(\boldsymbol{\mu}_k^* - \hat{\mathbf{x}}_{k|k-1})^T \quad (19)$$

$$\mathbf{B}_k^* = (\mathbf{z}_k - \mathbf{H}_k \boldsymbol{\mu}_k^*)(\mathbf{z}_k - \mathbf{H}_k \boldsymbol{\mu}_k^*)^T + \mathbf{H}_k \boldsymbol{\Sigma}_k^* \mathbf{H}_k^T. \quad (20)$$

*Proof:* See Appendix C.  $\blacksquare$

Next, we will further confirm the extreme point  $\boldsymbol{\mu}_k^*$  in (15) is a maximum point or a minimum point, and present the monotonicity of the approximate cost function  $J_1(\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)$  with respect to the posterior covariance matrix  $\boldsymbol{\Sigma}_k$ .

*Theorem 4:* If the similarity functions satisfy Condition 2 and Condition 3 and the following inequalities hold:

$$-(\xi_k^*)^2 + 2\tilde{\xi}_k^* < 0, \quad -(\lambda_k^*)^2 + 2\tilde{\lambda}_k^* < 0 \quad (21)$$

then both the Hessian matrix  $\boldsymbol{\Theta}_{\boldsymbol{\mu}_k}(\boldsymbol{\mu}_k^*, \boldsymbol{\Sigma}_k^*)$  and the Jacobian matrix  $\boldsymbol{\Delta}_{\boldsymbol{\Sigma}_k}(\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)$  are negative definite, i.e.,

$$\boldsymbol{\Theta}_{\boldsymbol{\mu}_k}(\boldsymbol{\mu}_k^*, \boldsymbol{\Sigma}_k^*) < \mathbf{0}, \quad \boldsymbol{\Delta}_{\boldsymbol{\Sigma}_k}(\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k) < \mathbf{0}. \quad (22)$$

*Proof:* See Appendix D.  $\blacksquare$

Theorem 4 implies that the extreme point  $\boldsymbol{\mu}_k^*$  in (15) is a maximum point of the approximate cost function  $J_1(\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)$ . It can be also observed from Theorem 4 that the approximate cost function  $J_1(\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)$  is monotonically decreasing with respect to the posterior covariance matrix  $\boldsymbol{\Sigma}_k$ , and then the approximate cost function  $J_1(\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)$  can achieve a unique optimal solution  $\boldsymbol{\Sigma}_k^*$  at the lower bound of the posterior covariance matrix. Next, we will determine the optimal posterior covariance matrix  $\boldsymbol{\Sigma}_k^*$ .

To obtain the maximum point  $\boldsymbol{\Sigma}_k^*$ , we need to find a reasonable constraint to apply upon  $\boldsymbol{\Sigma}_k$ . Motivated by the fact that the covariance matrix of the posterior pdf is the negative inverse of the Hessian matrix of the least squares cost function in the traditional maximum *a posteriori* estimation framework, we propose a heuristic assumption that  $\boldsymbol{\Sigma}_k$  is not less than the negative inverse of the Hessian matrix  $\boldsymbol{\Theta}_{\boldsymbol{\mu}_k}(\boldsymbol{\mu}_k^*, \boldsymbol{\Sigma}_k^*)$ , i.e.,

$$\boldsymbol{\Sigma}_k \geq -\boldsymbol{\Theta}_{\boldsymbol{\mu}_k}^{-1}(\boldsymbol{\mu}_k^*, \boldsymbol{\Sigma}_k^*). \quad (23)$$

It is worth noting that the well-known M-estimate employs a similar way to deal with the posterior covariance matrix, which is set as the negative inverse of the Hessian matrix of the robust cost function [8]–[10]. Since the cost function  $J(\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)$  is monotonically decreasing with respect to  $\boldsymbol{\Sigma}_k$ , the optimal covariance matrix should be the negative inverse of the Hessian matrix, i.e.,

$-\Theta_{\mu_k}^{-1}(\mu_k^*, \Sigma_k^*)$ . Unfortunately, the Hessian matrix can easily lose its negative definiteness during the iterative computation, and then the filter was often found to halt its operation due to the numerical problem. To keep the negative definiteness, the positive semidefinite term  $\mathbf{D}_1(\mu_k^*, \Sigma_k^*)$  is mandatorily subtracted from the Hessian matrix  $\Theta_{\mu_k}(\mu_k^*, \Sigma_k^*)$ . That is to say, the lower bound of the posterior covariance matrix is further reduced, i.e.,

$$\Sigma_k \geq -\Theta_{\mu_k}^{-1}(\mu_k^*, \Sigma_k^*) \geq -[\Theta_{\mu_k}(\mu_k^*, \Sigma_k^*) - \mathbf{D}_1(\mu_k^*, \Sigma_k^*)]^{-1}. \quad (24)$$

Exploiting (24) and (60), the maximum point  $\Sigma_k^*$  can be formulated as

$$\Sigma_k^* = \left( \xi_k^* \mathbf{P}_{k|k-1}^{-1} + \lambda_k^* \mathbf{H}_k^T \mathbf{R}_k^{-1} \mathbf{H}_k \right)^{-1}. \quad (25)$$

Employing the matrix inversion lemma [1, pp. 11–12] and (16), (17), and (25) gives

$$\Sigma_k^* = \tilde{\mathbf{P}}_{k|k-1}^* - \tilde{\mathbf{K}}_k^* \mathbf{H}_k \tilde{\mathbf{P}}_{k|k-1}^*. \quad (26)$$

## B. Error Analyses of the Proposed Framework

In Section III-A, three approximations are employed to derive an analytical solution for the original maximization problem in (6), which are listed as follows.

- 1) Approximation 1: The posterior pdf  $q(\mathbf{x}_k)$  is approximated as a Gaussian pdf in (7).
- 2) Approximation 2: The original cost function is approximated as its lower bound by Theorem 2.
- 3) Approximation 3: The original Hessian matrix is approximated as  $\Theta_{\mu_k}(\mu_k^*, \Sigma_k^*) - \mathbf{D}_1(\mu_k^*, \Sigma_k^*)$ .

First, we discuss the reasonability of Approximation 1. The outlier-contaminated state and measurement noises often have non-Gaussian distributions, and then the true posterior pdf  $p(\mathbf{x}_k | \mathbf{z}_{1:k})$  also has a non-Gaussian distribution [14], [16]. Unfortunately, there is often not a mathematical formulation for a general non-Gaussian distribution. As a result, it is not possible to achieve an optimal solution of the maximization problem in (6) for a general non-Gaussian linear system. Motivated by the fact that the Gaussian approximation to the posterior pdf has been widely accepted in designing a cost-effective filter for a linear system with non-Gaussian noises, the posterior pdf is also approximated as a Gaussian pdf in this article, as shown in (7), based on which an approximately analytical solution can be obtained. Although such an approximation may impose an error on the posterior pdf, it often exhibits good estimation accuracy with reasonable computational complexity. Thus, the Gaussian approximation to the posterior pdf can provide a tradeoff between estimation accuracy and computational complexity.

Second, we analyze the effects of Approximation 2 on the optimal solution. To this end, the relation between the maximization problem in (8) and the maximization problem in (9)–(12) will be further revealed. Define four auxiliary variables as follows:

$$\begin{cases} Y_{1k} = \|\mathbf{S}_{k|k-1}^{-1}(\mathbf{x}_k - \hat{\mathbf{x}}_{k|k-1})\|^2, & Y_{1k}^* = \text{tr}(\mathbf{A}_k^* \mathbf{P}_{k|k-1}^{-1}) \\ Y_{2k} = \|\mathbf{S}_{\mathbf{R}_k}^{-1}(\mathbf{z}_k - \mathbf{H}_k \mathbf{x}_k)\|^2, & Y_{2k}^* = \text{tr}(\mathbf{B}_k^* \mathbf{R}_k^{-1}) \end{cases} \quad (27)$$

where  $Y_{1k}^*$  and  $Y_{2k}^*$  are, respectively, the expectations of  $Y_{1k}$  and  $Y_{2k}$  with respect to the approximate posterior pdf  $q^*(\mathbf{x}_k) = \mathcal{N}(\mathbf{x}_k; \mu_k^*, \Sigma_k^*)$ .

Taking the first-order Taylor series expansions of the similarity functions  $f_x(t)$  and  $f_z(t)$  at  $t = Y_{1k}^*$  and  $t = Y_{2k}^*$ , respectively, we have

$$\begin{cases} f_x(t) = f_x(Y_{1k}^*) + \dot{f}_x(Y_{1k}^*)(t - Y_{1k}^*) + o(t - Y_{1k}^*) \\ f_z(t) = f_z(Y_{2k}^*) + \dot{f}_z(Y_{2k}^*)(t - Y_{2k}^*) + o(t - Y_{2k}^*) \end{cases} \quad (28)$$

where  $o(t - Y_{1k}^*)$  and  $o(t - Y_{2k}^*)$  denote the first-order remainder terms of the similarity functions  $f_x(t)$  and  $f_z(t)$  at  $t = Y_{1k}^*$  and  $t = Y_{2k}^*$ , respectively.

Using (28) yields

$$\begin{cases} f_x(Y_{1k}) \approx f_x(Y_{1k}^*) + \dot{f}_x(Y_{1k}^*)(Y_{1k} - Y_{1k}^*) \\ f_z(Y_{2k}) \approx f_z(Y_{2k}^*) + \dot{f}_z(Y_{2k}^*)(Y_{2k} - Y_{2k}^*) \end{cases} \quad (29)$$

*Proposition 1:* The maximization problem in Theorem 2 and the maximization problem in (8) with the first-order Taylor approximations (27) and (29) have the same optimal solution  $\{\mu_k^*, \Sigma_k^*\}$ .

*Proof:* See Appendix E. ■

Proposition 1 means that the lower bound of the original cost function and the first-order approximation of the original cost function have the same maximum solution. Thus, the effects of Approximation 2 on the optimal solution are determined by the higher order approximation errors  $o(Y_{1k} - Y_{1k}^*)$  and  $o(Y_{2k} - Y_{2k}^*)$ . Since  $Y_{1k}^*$  and  $Y_{2k}^*$  are the mean values of the random variables  $Y_{1k}$  and  $Y_{2k}$ , the difference values  $Y_{1k} - Y_{1k}^*$  and  $Y_{2k} - Y_{2k}^*$  depend heavily on the variances of the random variables  $Y_{1k}$  and  $Y_{2k}$ . That is to say, the higher variances of the random variables  $Y_{1k}$  and  $Y_{2k}$ , the larger fluctuation ranges of the difference values  $Y_{1k} - Y_{1k}^*$  and  $Y_{2k} - Y_{2k}^*$  will be generated and then result in larger higher order approximation errors, and vice versa. Thus, the approximation accuracy of Theorem 2 is mainly dominated by the variances of the random variables  $Y_{1k}$  and  $Y_{2k}$ . Therefore, next, we will study the variances of the random variables  $Y_{1k}$  and  $Y_{2k}$ .

*Proposition 2:* The variances of the random variables  $Y_{1k}$  and  $Y_{2k}$  satisfy the following upper bound constraints:

$$\begin{cases} \text{Var}[Y_{1k}] \leq (n^2 + 2n - 1) \left[ \text{tr}(\Sigma_k \mathbf{P}_{k|k-1}^{-1}) \right]^2 \\ \quad + \left\| 2\Sigma_k^{\frac{\text{T}}{2}} \mathbf{P}_{k|k-1}^{-1} (\mu_k - \hat{\mathbf{x}}_{k|k-1}) \right\|^2 \\ \text{Var}[Y_{2k}] \leq (n^2 + 2n - 1) \left[ \text{tr}(\mathbf{H}_k \Sigma_k \mathbf{H}_k^T \mathbf{R}_k^{-1}) \right]^2 \\ \quad + \left\| 2\Sigma_k^{\frac{\text{T}}{2}} \mathbf{H}_k^T \mathbf{R}_k^{-1} (\mathbf{z}_k - \mathbf{H}_k \mu_k) \right\|^2 \end{cases} \quad (30)$$

where  $\text{Var}[\cdot]$  denotes the variance operation.

*Proof:* See Appendix F. ■

It is observed from Proposition 2 that both the variances of the random variables  $Y_{1k}$  and  $Y_{2k}$  depend on the state dimension  $n$  and the posterior covariance matrix  $\Sigma_k$ . It can be observed from (30) that the higher the state dimension, the larger will be the variances of the random variables  $Y_{1k}$  and  $Y_{2k}$ . Such a result is consistent with the intuition that the higher the state dimension, the larger the errors will be when the original cost function is approximated by its lower bound in Theorem 2.

TABLE I  
EXEMPLARY SIMILARITY FUNCTIONS  $f(\cdot)$  AND THEIR FIRST AND SECOND DERIVATIVES

$f(t)$	$\dot{f}(t)$	$\ddot{f}(t)$
$-0.5t$	$-0.5$	$0$
$\sigma^2 \exp(\frac{p-t}{2\sigma^2})$	$-0.5 \exp(\frac{p-t}{2\sigma^2})$	$\frac{1}{4\sigma^2} \exp(\frac{p-t}{2\sigma^2})$
$-0.5(\nu + p) \log(1 + \frac{t}{\nu})$	$-0.5 \frac{\nu+p}{\nu+t}$	$0.5 \frac{\nu+p}{(\nu+t)^2}$
$-\sqrt{(\omega+p)(\omega+t)}$	$-0.5 \sqrt{\frac{\omega+p}{\omega+t}}$	$0.25 \frac{\sqrt{\omega+p}}{\sqrt[3]{\omega+t}}$

We can also observe from (30) that the larger the posterior covariance matrix  $\Sigma_k$ , the larger will be the variances of the random variables  $Y_{1k}$  and  $Y_{2k}$ . This result is also consistent with the fact that the approximation errors are mainly dominated by the randomness of state vector  $\mathbf{x}_k$  (i.e., the covariance matrix) when the original cost function is approximated by its lower bound based on Jensen's inequality. In conclusion, the effects of Approximation 2 on the optimal solution relies mainly on the state dimension and the posterior covariance matrix of the state vector. Fortunately, the posterior uncertainty is gradually reduced as the filter converges, which can to some extent mitigate the effects of Approximation 2 on the optimal solution. More importantly, both the errors  $Y_{1k} - Y_{1k}^*$  and  $Y_{2k} - Y_{2k}^*$  can be deemed as small terms as the filter converges, and the higher order derivatives of the similarity functions  $f_x(t)$  and  $f_z(t)$  are significantly smaller than the first-order derivatives for the exemplary similarity functions in Table I. As a result, the higher order approximation errors  $o(Y_{1k} - Y_{1k}^*)$  and  $o(Y_{2k} - Y_{2k}^*)$  are significantly smaller than the first-order terms for the exemplary similarity functions, which contributes to the effectiveness and reasonability of Approximation 2.

Finally, we discuss the effects of Approximation 3 on the optimal solution. The original Hessian matrix is forcibly reduced by subtracting a positive semidefinite matrix, and the resultant posterior covariance matrix is less than the optimal value. Although such approximation imposes an error on the posterior covariance matrix, it is beneficial to guarantee the positive definiteness of the posterior covariance matrix and then improve the numerical and filtering stabilities, as shown in Sections III-C and IV-A. Moreover, the reduction of the posterior covariance matrix is also beneficial to mitigate the effects of Approximation 2 on the optimal solution.

### C. Stability Discussions of Proposed Framework

In order to employ the proposed algorithm in engineering applications, it is necessary to guarantee the numerical and filtering stabilities of the proposed method. Next, these will be discussed.

It is seen from (18) that if the similarity functions  $f_x(\cdot)$  and  $f_z(\cdot)$  satisfy Condition 2, i.e.,  $\dot{f}_x(t) < 0$  and  $\dot{f}_z(t) < 0$  for  $t \in [0, +\infty)$ , then both the auxiliary parameters  $\xi_k^*$  and  $\lambda_k^*$  are greater than zero, i.e.,

$$\xi_k^* > 0, \quad \lambda_k^* > 0. \quad (31)$$

Using (17), (25) and (31) yields

$$\tilde{\mathbf{P}}_{k|k-1}^* > \mathbf{0}, \quad \tilde{\mathbf{R}}_k^* > \mathbf{0}, \quad \Sigma_k^* > \mathbf{0}. \quad (32)$$

We can see from (32) that if  $\dot{f}_x(t) < 0$  and  $\dot{f}_z(t) < 0$  for  $t \in [0, +\infty)$ , then the modified one-step prediction error and measurement noise covariance matrices and the posterior covariance matrix are all positive definite. Thus, the proposed filter is numerically stable if the similarity function similarity functions  $f_x(\cdot)$  and  $f_z(\cdot)$  satisfy Condition 2.

In this article, the filtering stability means that the state estimation error  $\tilde{\mathbf{x}}_{k|k}$  is bounded in the sense of mean square, i.e.,  $E\{\|\tilde{\mathbf{x}}_{k|k}\|^2\} < +\infty$  [27]. According to the theoretical analysis of the filtering stability in [27], if the modified one-step prediction error and measurement noise covariance matrices and the posterior covariance matrix satisfy the following constraints:

$$\begin{cases} \tilde{\mathbf{P}}_{k|k-1}^* \leq q_{\max} \mathbf{I}_n, & \tilde{\mathbf{R}}_k^* \geq r_{\min} \mathbf{I}_m \\ p_{\min} \mathbf{I}_n \leq \Sigma_k^* \leq p_{\max} \mathbf{I}_n \end{cases} \quad (33)$$

where  $q_{\max}$ ,  $r_{\min}$ ,  $p_{\min}$ , and  $p_{\max}$  are all positive real numbers, then the state estimation error  $\tilde{\mathbf{x}}_{k|k}$  of the proposed filter is bounded in the sense of mean square, i.e.,  $E\{\|\tilde{\mathbf{x}}_{k|k}\|^2\} < +\infty$ . Note that the other conditions for guaranteeing the filtering stability in [27] hold naturally for linear systems.

*Proposition 3:* If there exists positive real numbers  $\xi_{\min}$ ,  $\xi_{\max}$ ,  $\lambda_{\min}$ , and  $\lambda_{\max}$  such that the following constraints are fulfilled:

$$\xi_{\min} \leq \xi_k^* \leq \xi_{\max}, \quad \lambda_{\min} \leq \lambda_k^* \leq \lambda_{\max} \quad (34)$$

then, the inequalities in (33) hold.

*Proof:* See Appendix G. ■

Proposition 3 demonstrates that if both the auxiliary parameters  $\xi_k^*$  and  $\lambda_k^*$  have lower and upper bounds, then the proposed outlier-robust Kalman filtering framework is always stable, which will impose additional constraints on the similarity functions  $f_x(\cdot)$  and  $f_z(\cdot)$ .

*Remark 3:* The proposed outlier-robust Kalman filtering framework is related to some existing advanced Kalman filtering algorithms. The standard Kalman filter and existing RSTKF [22] are special cases of the proposed outlier-robust Kalman filtering framework. The proposed outlier-robust Kalman filter becomes the standard Kalman filter when the similarity functions are, respectively, chosen as  $f_x(t) = -0.5t$  and  $f_z(t) = -0.5t$ , and the proposed outlier-robust Kalman filter becomes the existing RSTKF when the similarity functions are, respectively, selected as  $f_x(t) = -0.5(\nu + n) \log(1 + \frac{t}{\nu})$  and  $f_z(t) = -0.5(\nu + m) \log(1 + \frac{t}{\nu})$ .

## IV. NOVEL ITERATIVE ALGORITHMS FOR PROPOSED OUTLIER-ROBUST KALMAN FILTERING FRAMEWORK

### A. Fixed-Point Iterative Algorithm

In general, it is very difficult to find analytical solutions for  $\mu_k^*$  and  $\Sigma_k^*$  through solving (15)–(20) and (26) directly when the similarity functions  $f_x(\cdot)$  and  $f_z(\cdot)$  are nonlinear functions. To solve this problem, we employ fixed-point iterations to solve these equations approximately. The detailed implementation of the proposed outlier-robust Kalman filtering framework based on fixed-point iterations is listed in Algorithm 1, where  $\epsilon$  denotes the iteration threshold,  $N_m$  denotes the maximum number of iterations,  $i^*$  denotes the cyclic variable at the end of the loop,

**Algorithm 1:** One Time Step of The Proposed Outlier-Robust Kalman Filtering Framework Based On Fixed-Point Iterations.

**Inputs:**  $\hat{\mathbf{x}}_{k-1|k-1}$ ,  $\mathbf{P}_{k-1|k-1}$ ,  $\mathbf{F}_k$ ,  $\mathbf{H}_k$ ,  $\mathbf{z}_k$ ,  $\mathbf{Q}_k$ ,  $\mathbf{R}_k$ ,  $f_x(\cdot)$ ,  $f_z(\cdot)$ ,  $\epsilon$ ,  $N_m$ ,  $\delta$ .

**Time update:**

- 1:  $\hat{\mathbf{x}}_{k|k-1} = \mathbf{F}_k \hat{\mathbf{x}}_{k-1|k-1}$
- 2:  $\mathbf{P}_{k|k-1} = \mathbf{F}_k \mathbf{P}_{k-1|k-1} \mathbf{F}_k^T + \mathbf{Q}_k$

**Iterative measurement update:**

- 3: Initialization:  $\xi_k^{(0)} = 1$ ,  $\lambda_k^{(0)} = 1$

**for**  $i = 0 : N_m - 1$

Calculate  $\tilde{\mathbf{P}}_{k|k-1}^{(i+1)}$  and  $\tilde{\mathbf{R}}_k^{(i+1)}$

- 4:  $\tilde{\mathbf{P}}_{k|k-1}^{(i+1)} = \mathbf{P}_{k|k-1} / \xi_k^{(i)}$ ,  $\tilde{\mathbf{R}}_k^{(i+1)} = \mathbf{R}_k / \lambda_k^{(i)}$

Evaluate  $\boldsymbol{\mu}_k^{(i+1)}$  and  $\boldsymbol{\Sigma}_k^{(i+1)}$

- 5:  $\tilde{\mathbf{K}}_k^{(i+1)} = \tilde{\mathbf{P}}_{k|k-1}^{(i+1)} \mathbf{H}_k^T (\mathbf{H}_k \tilde{\mathbf{P}}_{k|k-1}^{(i+1)} \mathbf{H}_k^T + \tilde{\mathbf{R}}_k^{(i+1)})^{-1}$

- 6:  $\boldsymbol{\mu}_k^{(i+1)} = \hat{\mathbf{x}}_{k|k-1} + \tilde{\mathbf{K}}_k^{(i+1)} (\mathbf{z}_k - \mathbf{H}_k \hat{\mathbf{x}}_{k|k-1})$

- 7:  $\boldsymbol{\Sigma}_k^{(i+1)} = \tilde{\mathbf{P}}_{k|k-1}^{(i+1)} - \tilde{\mathbf{K}}_k^{(i+1)} \mathbf{H}_k \tilde{\mathbf{P}}_{k|k-1}^{(i+1)}$

Check the convergence of iteration

- 8: If  $\frac{\|\boldsymbol{\mu}_k^{(i+1)} - \boldsymbol{\mu}_k^{(i)}\|}{\|\boldsymbol{\mu}_k^{(i)}\|} \leq \epsilon$ , stop iteration.

Calculate  $\mathbf{A}_k^{(i+1)}$  and  $\mathbf{B}_k^{(i+1)}$

- 9:  $\mathbf{A}_k^{(i+1)} = \boldsymbol{\Sigma}_k^{(i+1)} + (\boldsymbol{\mu}_k^{(i+1)} - \hat{\mathbf{x}}_{k|k-1})(\boldsymbol{\mu}_k^{(i+1)} - \hat{\mathbf{x}}_{k|k-1})^T$
- 10:  $\mathbf{B}_k^{(i+1)} = (\mathbf{z}_k - \mathbf{H}_k \boldsymbol{\mu}_k^{(i+1)})(\mathbf{z}_k - \mathbf{H}_k \boldsymbol{\mu}_k^{(i+1)})^T + \mathbf{H}_k \boldsymbol{\Sigma}_k^{(i+1)} \mathbf{H}_k^T$

Evaluate  $\xi_k^{(i+1)}$  and  $\lambda_k^{(i+1)}$

- 11:  $\xi_k^{(i+1)} = -2\dot{f}_x(\text{tr}(\mathbf{A}_k^{(i+1)} \mathbf{P}_{k|k-1}^{-1}))$

- 12:  $\lambda_k^{(i+1)} = -2\dot{f}_z(\text{tr}(\mathbf{B}_k^{(i+1)} \mathbf{R}_k^{-1}))$

Check the constraints in (36)

- 13: If  $\xi_k^{(i+1)} < \delta$ , then  $\xi_k^{(i+1)} = \delta$ .

- 14: If  $\lambda_k^{(i+1)} < \delta$ , then  $\lambda_k^{(i+1)} = \delta$ .

**end for**

- 13:  $\hat{\mathbf{x}}_{k|k} = \boldsymbol{\mu}_k^{(i^*)}$ ,  $\mathbf{P}_{k|k} = \boldsymbol{\Sigma}_k^{(i^*)}$

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

and  $\delta$  denotes the lower bounds of the auxiliary parameters  $\xi_k$  and  $\lambda_k$  that is beneficial to guarantee the filtering stability, as discussed in Section V.

The premise of applying the proposed outlier-robust filtering algorithm to practical engineering is to guarantee the convergence of fixed-point iterations. To this end, the relation between the fixed-point iterative algorithm and the existing nonlinear optimization algorithm is first revealed.

*Proposition 4:* The fixed-point iterative algorithm is identical to the existing Newton's method with the modified Hessian matrix as follows:

$$\tilde{\Theta}_{\boldsymbol{\mu}_k}(\boldsymbol{\mu}_k^{(i)}, \boldsymbol{\Sigma}_k^{(i)}) = \Theta_{\boldsymbol{\mu}_k}(\boldsymbol{\mu}_k^{(i)}, \boldsymbol{\Sigma}_k^{(i)}) - \mathbf{D}_1(\boldsymbol{\mu}_k^{(i)}, \boldsymbol{\Sigma}_k^{(i)}). \quad (35)$$

*Proof:* See Appendix H. ■

Proposition 4 means that the fixed-point iterative algorithm is an improved version of the existing Newton's method. The

positive semidefinite term  $\mathbf{D}_1(\boldsymbol{\mu}_k^{(i)}, \boldsymbol{\Sigma}_k^{(i)})$  is subtracted from the original Hessian matrix so that the negative definiteness can be preserved. As a result, the fixed-point iterative algorithm has better numerical stability than the standard Newton's method. Normally, to guarantee the local convergence of the Newton's method, the Hessian matrix needs to satisfy the Lipschitz condition, and the initial value is sufficiently close to the optimal value. Next, we will provide the convergence conditions of the fixed-point iterative algorithm.

*Theorem 5:* If the initial mean vector  $\boldsymbol{\mu}_k^{(0)}$  is sufficiently close to the optimal mean vector  $\boldsymbol{\mu}_k^*$  and there are positive and bounded real numbers  $\alpha_1$  and  $\alpha_2$  making the following inequalities hold:

$$\begin{cases} 0 \leq \ddot{f}_x(t^2)t \leq \alpha_1 \\ 0 \leq \ddot{f}_z(t^2)t \leq \alpha_2 \end{cases}, \quad \forall t \geq 0 \quad (36)$$

then, the fixed-point iterative algorithm has local convergence.

*Proof:* See Appendix I. ■

Generally, the initial mean vector  $\boldsymbol{\mu}_k^{(0)}$  is selected as the one-step predicted state vector  $\hat{\mathbf{x}}_{k|k-1}$ , i.e.,  $\boldsymbol{\mu}_k^{(0)} = \hat{\mathbf{x}}_{k|k-1}$ , since  $\hat{\mathbf{x}}_{k|k-1}$  is the only available information for the mean vector before the iterative measurement update. It is seen from Theorem 5 that the initial value  $\boldsymbol{\mu}_k^{(0)}$ , i.e.,  $\hat{\mathbf{x}}_{k|k-1}$ , needs to be sufficiently close to the optimal value  $\boldsymbol{\mu}_k^*$  to guarantee the local convergence of the fixed-point iterative algorithm. Unfortunately, the one-step predicted state vector  $\hat{\mathbf{x}}_{k|k-1}$  may be far away from the optimal value  $\boldsymbol{\mu}_k^*$  when the linear system suffers from large process uncertainty or a state outlier. As a result, the fixed-point iterative algorithm may not converge to a local optimum when the linear system suffers from large process uncertainty or a state outlier, which will degrade the filtering accuracy of the proposed outlier-robust filter dramatically.

## B. Novel Separate Iterative Algorithm

In order to motivate the proposed separate iterative algorithm, we first present the problems that exist in the fixed-point iterative algorithm. It can be seen from Algorithm 1 that the Kalman gain  $\tilde{\mathbf{K}}_k^{(i+1)}$  plays an important role in the iterative measurement update because it can adjust the weights of one-step predicted state vector  $\hat{\mathbf{x}}_{k|k-1}$  and measurement innovation  $\mathbf{z}_k - \mathbf{H}_k \hat{\mathbf{x}}_{k|k-1}$  adaptively. Next, we will discuss the behavior of the Kalman gain  $\tilde{\mathbf{K}}_k^{(i+1)}$  during the iterative measurement update.

Using the 4th and 5th equations in Algorithm 1, the Kalman gain  $\tilde{\mathbf{K}}_k^{(i+1)}$  can be rewritten as

$$\tilde{\mathbf{K}}_k^{(i+1)} = \mathbf{P}_{k|k-1} \mathbf{H}_k^T \left( \mathbf{H}_k \mathbf{P}_{k|k-1} \mathbf{H}_k^T + \frac{\xi_k^{(i)}}{\lambda_k^{(i)}} \mathbf{R}_k \right)^{-1}. \quad (37)$$

It is observed from (37) that the behavior of the Kalman gain  $\tilde{\mathbf{K}}_k^{(i+1)}$  depends only on the ratio of auxiliary parameters, i.e.,  $\frac{\xi_k^{(i)}}{\lambda_k^{(i)}}$ , during the iterative measurement update. It can be also observed from 9th–12th equations in Algorithm 1 that the ratio of auxiliary parameters  $\frac{\xi_k^{(i)}}{\lambda_k^{(i)}}$  relies on not only the similarity functions  $f_x(\cdot)$  and  $f_z(\cdot)$  but also the parameters  $\mathbf{A}_k^{(i)}$  and  $\mathbf{B}_k^{(i)}$ , and the parameters  $\mathbf{A}_k^{(i)}$  and  $\mathbf{B}_k^{(i)}$  are adaptively adjusted by



**Algorithm 2:** One Time Step of The Proposed Outlier-Robust Kalman Filtering Framework Based on The Proposed Separate Iterative Algorithm.

**Inputs:**  $\hat{\mathbf{x}}_{k-1|k-1}$ ,  $\mathbf{P}_{k-1|k-1}$ ,  $\mathbf{F}_k$ ,  $\mathbf{H}_k$ ,  $\mathbf{z}_k$ ,  $\mathbf{Q}_k$ ,  $\mathbf{R}_k$ ,  $f_x(\cdot)$ ,  $f_z(\cdot)$ ,

$\epsilon_1$ ,  $\epsilon_2$ ,  $N_1$ ,  $N_2$ ,  $\delta$ .

**Time update:**

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

$$2: \mathbf{P}_{k|k-1} = \mathbf{F}_k \mathbf{P}_{k-1|k-1} \mathbf{F}_k^T + \mathbf{Q}_k$$

**Iterative measurement update:**

$$3: \text{Initialization: } \xi_k^{(0)(0)} = 1, \lambda_k^{(0)(0)} = 1$$

**for**  $i = 0 : N_1 - 1$

$$\text{Calculate } \tilde{\mathbf{P}}_{k|k-1}^{(i+1)(0)} \text{ and } \tilde{\mathbf{R}}_k^{(i+1)(0)}$$

$$4: \tilde{\mathbf{P}}_{k|k-1}^{(i+1)(0)} = \mathbf{P}_{k|k-1} / \xi_k^{(0)(0)}, \quad \tilde{\mathbf{R}}_k^{(i+1)(0)} = \mathbf{R}_k / \lambda_k^{(i)(0)}$$

Evaluate  $\boldsymbol{\mu}_k^{(i+1)(0)}$  and  $\boldsymbol{\Sigma}_k^{(i+1)(0)}$

$$5: \tilde{\mathbf{K}}_k^{(i+1)(0)} = \tilde{\mathbf{P}}_{k|k-1}^{(i+1)(0)} \mathbf{H}_k^T (\mathbf{H}_k \tilde{\mathbf{P}}_{k|k-1}^{(i+1)(0)} \mathbf{H}_k^T + \tilde{\mathbf{R}}_k^{(i+1)(0)})^{-1}$$

$$6: \boldsymbol{\mu}_k^{(i+1)(0)} = \hat{\mathbf{x}}_{k|k-1} + \tilde{\mathbf{K}}_k^{(i+1)(0)} (\mathbf{z}_k - \mathbf{H}_k \hat{\mathbf{x}}_{k|k-1})$$

$$7: \boldsymbol{\Sigma}_k^{(i+1)(0)} = \tilde{\mathbf{P}}_{k|k-1}^{(i+1)(0)} - \tilde{\mathbf{K}}_k^{(i+1)(0)} \mathbf{H}_k \tilde{\mathbf{P}}_{k|k-1}^{(i+1)(0)}$$

Calculate  $\mathbf{B}_k^{(i+1)(0)}$

$$8: \mathbf{B}_k^{(i+1)(0)} = (\mathbf{z}_k - \mathbf{H}_k \boldsymbol{\mu}_k^{(i+1)(0)}) (\mathbf{z}_k - \mathbf{H}_k \boldsymbol{\mu}_k^{(i+1)(0)})^T + \mathbf{H}_k \boldsymbol{\Sigma}_k^{(i+1)(0)} \mathbf{H}_k^T$$

Evaluate  $\lambda_k^{(i+1)(0)}$

$$9: \lambda_k^{(i+1)(0)} = -2 \dot{f}_z(\text{tr}(\mathbf{B}_k^{(i+1)(0)} \mathbf{R}_k^{-1}))$$

Check the constraints in (36)

$$10: \text{If } \lambda_k^{(i+1)(0)} < \delta, \text{ then } \lambda_k^{(i+1)(0)} = \delta.$$

Check the convergence of iteration

$$11: \text{If } \frac{|\lambda_k^{(i+1)(0)} - \lambda_k^{(i)(0)}|}{|\lambda_k^{(i)(0)}|} \leq \epsilon_1, \text{ stop iteration.}$$

**end for**

**for**  $j = 0 : N_2 - 1$

$$\text{Calculate } \tilde{\mathbf{P}}_{k|k-1}^{(i_1^*)(j+1)} \text{ and } \tilde{\mathbf{R}}_k^{(i_1^*)(j+1)}$$

$$12: \tilde{\mathbf{P}}_{k|k-1}^{(i_1^*)(j+1)} = \mathbf{P}_{k|k-1} / \xi_k^{(0)(j)},$$

$$\tilde{\mathbf{R}}_k^{(i_1^*)(j+1)} = \mathbf{R}_k / \lambda_k^{(i_1^*)(0)}$$

Evaluate  $\boldsymbol{\mu}_k^{(i_1^*)(j+1)}$  and  $\boldsymbol{\Sigma}_k^{(i_1^*)(j+1)}$

$$13: \tilde{\mathbf{K}}_k^{(i_1^*)(j+1)} = \tilde{\mathbf{P}}_{k|k-1}^{(i_1^*)(j+1)} \mathbf{H}_k^T (\mathbf{H}_k \tilde{\mathbf{P}}_{k|k-1}^{(i_1^*)(j+1)} \mathbf{H}_k^T + \tilde{\mathbf{R}}_k^{(i_1^*)(j+1)})^{-1}$$

$$14: \boldsymbol{\mu}_k^{(i_1^*)(j+1)} = \hat{\mathbf{x}}_{k|k-1} + \tilde{\mathbf{K}}_k^{(i_1^*)(j+1)} (\mathbf{z}_k - \mathbf{H}_k \hat{\mathbf{x}}_{k|k-1})$$

$$15: \boldsymbol{\Sigma}_k^{(i_1^*)(j+1)} = \tilde{\mathbf{P}}_{k|k-1}^{(i_1^*)(j+1)} - \tilde{\mathbf{K}}_k^{(i_1^*)(j+1)} \mathbf{H}_k \tilde{\mathbf{P}}_{k|k-1}^{(i_1^*)(j+1)}$$

Calculate  $\mathbf{A}_k^{(0)(j+1)}$

$$16: \mathbf{A}_k^{(0)(j+1)} = (\boldsymbol{\mu}_k^{(i_1^*)(j+1)} - \hat{\mathbf{x}}_{k|k-1}) (\boldsymbol{\mu}_k^{(i_1^*)(j+1)} - \hat{\mathbf{x}}_{k|k-1})^T + \boldsymbol{\Sigma}_k^{(i_1^*)(j+1)}$$

Evaluate  $\xi_k^{(0)(j+1)}$

$$17: \xi_k^{(0)(j+1)} = -2 \dot{f}_x(\text{tr}(\mathbf{A}_k^{(0)(j+1)} \mathbf{P}_{k|k-1}^{-1}))$$

Check the constraints in (36)

$$18: \text{If } \xi_k^{(0)(j+1)} < \delta, \text{ then } \xi_k^{(0)(j+1)} = \delta.$$

Check the convergence of iteration

Continued

$$19: \text{If } \frac{|\xi_k^{(0)(j+1)} - \xi_k^{(0)(j)}|}{|\xi_k^{(0)(j)}|} \leq \epsilon_2, \text{ stop iteration.}$$

**end for**

$$20: \hat{\mathbf{x}}_{k|k} = \boldsymbol{\mu}_k^{(i_1^*)(j_2^*)}, \mathbf{P}_{k|k} = \boldsymbol{\Sigma}_k^{(i_1^*)(j_2^*)}$$

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

the iterative posterior mean vector  $\boldsymbol{\mu}_k^{(i)}$  and covariance matrix  $\boldsymbol{\Sigma}_k^{(i)}$  at the same time. As a result, the behavior of the Kalman gain  $\tilde{\mathbf{K}}_k^{(i+1)}$  is likely to be indefinite and depends heavily on the relative size of state and measurement outliers. The indefinite behavior of the Kalman gain  $\tilde{\mathbf{K}}_k^{(i+1)}$  is easy to cause the iterative posterior mean vector  $\boldsymbol{\mu}_k^{(i)}$  and covariance matrix  $\boldsymbol{\Sigma}_k^{(i)}$  not to converge to local optimums.

In this article, we propose a heuristic idea that the iterations of auxiliary parameters are separated to guarantee the definite behavior of the Kalman gain. That is to say, the auxiliary parameter  $\lambda_k^{(i)(0)}$  is first iterated with fixed auxiliary parameter  $\xi_k^{(0)(0)} = 1$  until convergence, i.e.,  $\lim_{i \rightarrow +\infty} \lambda_k^{(i)(0)} = \bar{\lambda}_k$ , and then the auxiliary parameter  $\xi_k^{(0)(j)}$  is iterated with fixed auxiliary parameter  $\bar{\lambda}_k$  until convergence, i.e.,  $\lim_{j \rightarrow +\infty} \xi_k^{(0)(j)} = \bar{\xi}_k$ , where  $\bar{\xi}_k$  and  $\bar{\lambda}_k$  denote the local optimums of the auxiliary parameters. The detailed implementation of the proposed outlier-robust Kalman filtering framework based on the proposed separate iterative algorithm is listed in Algorithm 2, where  $\epsilon_1$  and  $\epsilon_2$  denote the iteration thresholds of  $\xi_k$  and  $\lambda_k$ , respectively, and  $N_1$  and  $N_2$  denotes the maximum numbers of iterations of  $\xi_k$  and  $\lambda_k$ , respectively, and  $i_1^*$  and  $j_2^*$  denote the cyclic variables at the end of the first and second loops, respectively.

*Theorem 6:* If the similarity functions  $f_z(\cdot)$  and  $f_x(\cdot)$  satisfy Condition 3 and  $\dot{f}_z(0)$  and  $\dot{f}_x(0)$  have lower bounds, then the iterative auxiliary parameters  $\lambda_k^{(i)(0)}$  and  $\xi_k^{(0)(j)}$  will converge to local optimums  $\bar{\lambda}_k$  and  $\bar{\xi}_k$ , respectively, i.e.,

$$\lim_{i \rightarrow +\infty} \lambda_k^{(i)(0)} = \bar{\lambda}_k, \quad \lim_{j \rightarrow +\infty} \xi_k^{(0)(j)} = \bar{\xi}_k. \quad (38)$$

*Proof:* See Appendix J.  $\blacksquare$

We can see from Theorem 6 that if the similarity functions  $f_x(\cdot)$  and  $f_z(\cdot)$  satisfy Condition 3 and  $\dot{f}_z(0)$  and  $\dot{f}_x(0)$  have lower bounds, the iterative auxiliary parameters  $\lambda_k^{(i)(0)}$  and  $\xi_k^{(0)(j)}$  will converge to local optima, and then the corresponding posterior mean vector and covariance matrix also converge to local optima, which guarantees the local convergence of the proposed separate iterative algorithm. As compared with the fixed-point iteration algorithm, the proposed algorithm does not require initial mean vector to be sufficiently close to the optimal mean vector. Thus, the local convergence conditions of the fixed-point iterative algorithm are more harsh than those of the proposed separate iterative algorithm.

## V. SELECTIONS OF THE SIMILARITY FUNCTIONS

It is seen from Algorithms 1 and 2 that the similarity functions  $f_x(\cdot)$  and  $f_z(\cdot)$  are necessary to implement the proposed

outlier-robust framework. Next, we will provide the selections of the similarity functions to guarantee that the proposed framework is identical to the standard Kalman filter for the case of Gaussian noises and robust to state and measurement outliers.

First, the relationship between the approximate one-step prediction error covariance matrix  $\mathbf{A}_k^*$  and the true one-step prediction error covariance matrix  $\mathbf{P}_{k|k-1}$  and that between the approximate measurement noise covariance matrix  $\mathbf{B}_k^*$  and the true measurement noise covariance matrix  $\mathbf{R}_k$  are studied for a linear state-space model with Gaussian state and measurement noises.

*Proposition 5:* For a linear state-space model with Gaussian state and measurement noises,  $\mathbf{A}_k^*$  and  $\mathbf{B}_k^*$  can be, respectively, approximated as the one-step prediction error covariance matrix and measurement noise covariance matrix, i.e.,

$$\mathbf{A}_k^* \approx \mathbf{P}_{k|k-1}, \quad \mathbf{B}_k^* \approx \mathbf{R}_k. \quad (39)$$

*Proof:* See Appendix K. ■

Second, we study the conditions of the similarity functions  $f_x(\cdot)$  and  $f_z(\cdot)$  to guarantee that the proposed framework is identical to the standard Kalman filter for the case of Gaussian noises. Substituting (39) in (18), we have

$$\xi_k^* = -2\dot{f}_x(n), \quad \lambda_k^* = -2\dot{f}_z(m) \quad (40)$$

where  $n$  and  $m$  are, respectively, the dimensions of the state vector and the measurement vector.

It is seen from Algorithms 1 and 2 that the proposed robust filter is identical to the standard Kalman filter when the modified parameters are unity, i.e.,  $\xi_k^* = \lambda_k^* = 1$ . Therefore, in order to guarantee that the proposed filter is identical to the standard Kalman filter when there are no state and measurement outliers, the similarity functions  $f_x(\cdot)$  and  $f_z(\cdot)$  need to satisfy the following:

$$\dot{f}_x(n) = -0.5, \quad \dot{f}_z(m) = -0.5. \quad (41)$$

Finally, we discuss the conditions of the similarity functions  $f_x(\cdot)$  and  $f_z(\cdot)$  to guarantee that the proposed framework is robust to outliers. Define two auxiliary matrices as follows:

$$\Psi_{k1} = \mathbf{A}_k^* - \mathbf{P}_{k|k-1}, \quad \Psi_{k2} = \mathbf{B}_k^* - \mathbf{R}_k. \quad (42)$$

In general, if there are state and measurement outliers, the approximate one-step prediction error covariance matrix  $\mathbf{A}_k^*$  is not less than the nominal one-step prediction error covariance matrix  $\mathbf{P}_{k|k-1}$ , and the approximate measurement noise covariance matrix  $\mathbf{B}_k^*$  is also not less than the nominal measurement noise covariance matrix  $\mathbf{R}_k$ , i.e.,

$$\Psi_{k1} \geq \mathbf{0}, \quad \Psi_{k2} \geq \mathbf{0}. \quad (43)$$

Furthermore, the auxiliary matrices  $\Psi_{k1}$  and  $\Psi_{k2}$  depend on the magnitudes of the state and measurement outliers, respectively, and the larger the state and measurement outliers, the larger auxiliary matrices  $\Psi_{k1}$  and  $\Psi_{k2}$  will be generated.

*Proposition 6:* For a linear state-space model with outlier-contaminated state and measurement noises, if the similarity

functions  $f_x(\cdot)$  and  $f_z(\cdot)$  satisfy the following:

$$\begin{cases} \dot{f}_x(t) < 0 & \ddot{f}_x(t) \geq 0 & \dot{f}_x(n) = -0.5 & t \in [0, +\infty) \\ \dot{f}_z(t) < 0 & \ddot{f}_z(t) \geq 0 & \dot{f}_z(m) = -0.5 & t \in [0, +\infty) \end{cases} \quad (44)$$

then, the modified auxiliary parameters  $\xi_k^*$  and  $\lambda_k^*$  satisfy the following:

$$0 < \xi_k^* \leq 1, \quad 0 < \lambda_k^* \leq 1 \quad (45)$$

and the larger state and measurement outliers, the smaller modified auxiliary parameters  $\xi_k^*$  and  $\lambda_k^*$  will be obtained.

*Proof:* See Appendix L. ■

Employing (45) in (17) results in

$$\tilde{\mathbf{P}}_{k|k-1}^* - \mathbf{P}_{k|k-1} \geq \mathbf{0}, \quad \tilde{\mathbf{R}}_k^* - \mathbf{R}_k \geq \mathbf{0}. \quad (46)$$

It is observed from (46) that the modified prediction error covariance matrix and the modified measurement noise covariance matrix are, respectively, not less than the nominal prediction error covariance matrix and the nominal measurement noise covariance matrix when there are, respectively, state and measurement outliers. Moreover, according to Proposition 6 and (17), the larger the state and measurement outliers, the smaller the modified auxiliary parameters  $\xi_k^*$  and  $\lambda_k^*$  will be, and the larger the modified prediction error covariance matrix and modified measurement noise covariance matrix will become.

Using (16) and (17), the Kalman gain  $\tilde{\mathbf{K}}_k^*$  can be reformulated as

$$\tilde{\mathbf{K}}_k^* = \mathbf{P}_{k|k-1} \mathbf{H}_k^T \left( \mathbf{H}_k \mathbf{P}_{k|k-1} \mathbf{H}_k^T + \frac{\xi_k^*}{\lambda_k^*} \mathbf{R}_k \right)^{-1}. \quad (47)$$

According to Proposition 6 and (47), the Kalman gain  $\tilde{\mathbf{K}}_k^*$  depends heavily on the relative magnitudes of the state and measurement outliers. Specifically, the Kalman gain is increased if the state outlier has larger magnitude than the measurement outlier, and vice versa. As a result, the negative effects of outliers on the proposed Kalman filtering framework can be resisted through adjusting the Kalman gain adaptively.

As an example, some exemplary similarity functions  $f(\cdot)$  are listed in Table I, where  $p$  denotes the dimension of the state vector for  $f(\cdot) = f_x(\cdot)$  and  $p$  denotes the dimension of the measurement vector for  $f(\cdot) = f_z(\cdot)$ , and  $\sigma$  and  $\nu$  are, respectively, named as the kernel size and the DOF parameter to be consistent with the existing MCKF [12] and RSTKF [22], and  $\omega$  is also named as the DOF parameter.

It is easy to verify that the exemplary similarity functions listed in Table I satisfy the conditions of Theorems 1 and 2 and Proposition 6. Theorem 3 and Propositions 1, 2, 4, and 5 do not need the exemplary similarity functions to satisfy the additional conditions. Next, we will further confirm whether the exemplary similarity functions satisfy the conditions of Theorems 4–6 and Proposition 3.

*Corollary 1:* If the kernel size  $\sigma$  and the DOF parameter  $\nu$  satisfy the following constraints:

$$\sigma^2 \exp\left(\frac{n - Y_{1k}^*}{2\sigma^2}\right) > 1, \quad \sigma^2 \exp\left(\frac{m - Y_{2k}^*}{2\sigma^2}\right) > 1, \quad \nu > 2 - p \quad (48)$$

then, the exemplary similarity functions in Table I satisfy (21) in Theorem 4.

It is worth noting that there is not a constraint on the DOF parameter  $\omega$  to guarantee that Theorem 4 holds. It is seen from Corollary 1, the constraint on the kernel size  $\sigma$  depends on the auxiliary parameters  $Y_{1k}^*$  and  $Y_{2k}^*$ . As a result, the constraint on the kernel size  $\sigma$  may change for different application scenarios. To address this problem, a reasonable scheme is choosing a sufficiently large kernel size  $\sigma$  so that the constraint on the kernel size always holds.

It is seen from (45) that  $\xi_k^*$  and  $\lambda_k^*$  have positive upper bounds  $\xi_{\max} = \lambda_{\max} = 1$ . Since the second derivatives of the similarity functions are nonnegative, the minimum values of the negative derivatives of the similarity functions  $-\dot{f}_x(t)$  and  $-\dot{f}_z(t)$  are achieved at  $t = +\infty$ . It can be seen from Table I that the negative derivatives of the exemplary similarity functions approach 0 as  $t$  tends to  $+\infty$ . As a result, the modified auxiliary parameters  $\xi_k^*$  and  $\lambda_k^*$  do not have positive lower bounds, and then Proposition 3 does not hold, which may lead to filtering instability. To address this problem, we can impose a very small lower bound  $\delta$  on the modified auxiliary parameters  $\xi_k^*$  and  $\lambda_k^*$  to guarantee filtering stability, as shown in the 13th and 14th equations of Algorithm 1 and the 10th and 18th equations of Algorithm 2.

*Corollary 2:* For the exemplary similarity functions, if the kernel size  $\sigma \rightarrow 0$  and the DOF parameters  $\nu \rightarrow 0$  and  $\omega \rightarrow 0$ , then there exists positive and bounded real numbers  $\alpha_1$  and  $\alpha_2$  making (36) in Theorem 5 hold.

Finally, we discuss the conditions of Theorem 6. It is easy to verify that if the kernel size  $\sigma \rightarrow 0$  and the DOF parameters  $\nu \rightarrow 0$  and  $\omega \rightarrow 0$ , then  $\dot{f}(0)$  has a lower bound, which guarantees the local convergence of the proposed separate iterative algorithm. Also, according to Theorem 5 and Corollary 2, if the kernel size  $\sigma \rightarrow 0$  and the DOF parameters  $\nu \rightarrow 0$  and  $\omega \rightarrow 0$ , then the fixed-point iterative algorithm has local convergence when the initial mean vector  $\mu_k^{(0)}$  is sufficiently close to the optimal mean vector  $\mu_k^*$ . Thus, for the exemplary similarity functions, the convergence conditions of the proposed separate iterative algorithm is easier to be satisfied as compared with that of the fixed-point iterative algorithm.

*Remark 4:* It is observed from Table I that the derivatives of the exemplary similarity functions are all  $-0.5$  when the kernel size  $\sigma$  and the DOF parameters  $\nu$  and  $\omega$  tend to infinity, i.e.,  $\sigma \rightarrow +\infty$ ,  $\nu \rightarrow +\infty$  and  $\omega \rightarrow +\infty$ . As a result, the proposed outlier-robust Kalman filters based on the exemplary similarity functions all reduce to the standard Kalman filter when the parameters  $\sigma$ ,  $\nu$ , and  $\omega$  tend to infinity.

*Remark 5:* The state vector is deemed as a random quantity in the proposed outlier-robust Kalman filter but the state vector is treated as a deterministic quantity in the existing M-estimator. It is observed from Algorithms 1 and 2 that the posterior covariance matrix is employed to calculate the modified parameters during the iterative measurement update in the proposed filter but the posterior covariance matrix is independent of the minimization of the robust cost function in the existing M-estimator. As compared with the existing M-estimator, the proposed outlier-robust Kalman filter considers the randomness inherent in the state

vector by exploiting the posterior covariance matrix during the iterative measurement update so that the state estimation accuracy can be further improved, as will be shown in the simulation study.

*Remark 6:* Different from the existing MCKF, the correntropy is approximated as its lower-bound by using Jensen's inequality in the proposed outlier-robust Kalman filter when the similarity functions are, respectively, set as  $f_x(t) = \sigma^2 \exp(\frac{p-t}{2\sigma^2})$  and  $f_z(t) = \sigma^2 \exp(\frac{p-t}{2\sigma^2})$ , and the robust state estimate is obtained by maximizing the lower-bound of the correntropy. Thus, the proposed outlier-robust based Kalman filter is an improved version of the existing MCKF when the similarity functions are, respectively, set as  $f_x(t) = \sigma^2 \exp(\frac{p-t}{2\sigma^2})$  and  $f_z(t) = \sigma^2 \exp(\frac{p-t}{2\sigma^2})$ , and the improved state estimation accuracy can be achieved, as will be illustrated in the simulation study.

## VI. SIMULATION STUDY

### A. Simulation Setup and Description

We consider a problem of tracking an agile target whose positions are measured in clutter, and the horizontal positions and corresponding velocities are chosen as elements of the state vector. The state transition matrix and measurement matrix are, respectively

$$\mathbf{F}_k = \begin{bmatrix} \mathbf{I}_2 & T\mathbf{I}_2 \\ \mathbf{0} & \mathbf{I}_2 \end{bmatrix}$$

and  $\mathbf{H}_k = [\mathbf{I}_2 \quad \mathbf{0}]$ , where  $T = 1\text{s}$  and  $\mathbf{I}_2$  denote the sampling interval and two-dimensional identity matrix, respectively. The outlier contaminated state and measurement noises are generated according to [28]

$$\begin{cases} \mathbf{w}_k \sim \begin{cases} N(\mathbf{0}, \mathbf{Q}) & \text{w.p. } 0.95 \\ N(\mathbf{0}, 100\mathbf{Q}) & \text{w.p. } 0.05 \end{cases} \\ \mathbf{v}_k \sim \begin{cases} N(\mathbf{0}, \mathbf{R}) & \text{w.p. } 0.95 \\ N(\mathbf{0}, 500\mathbf{R}) & \text{w.p. } 0.05 \end{cases} \end{cases} \quad (49)$$

where the nominal state and measurement noise covariance matrices are, respectively, selected as

$$\mathbf{Q} = \begin{bmatrix} \frac{T^3}{3}\mathbf{I}_2 & \frac{T^2}{2}\mathbf{I}_2 \\ \frac{T^2}{2}\mathbf{I}_2 & T\mathbf{I}_2 \end{bmatrix}$$

and  $\mathbf{R} = 100\mathbf{I}_2$ . The true initial state vector is chosen as  $\mathbf{x}_0 = [0, 0, 10, 10]^T$ , and the initial estimation error variance is set as  $\mathbf{P}_0 = \text{diag}([10000, 10000, 100, 100])$ , and the initial state estimate is randomly selected from a Gaussian distribution, i.e.,  $\hat{\mathbf{x}}_{0|0} \sim N(\mathbf{x}_0, \mathbf{P}_0)$ .

As an example, the similarity functions  $f_x(\cdot)$  and  $f_z(\cdot)$  are, respectively, selected as exponential, logarithmic, and square-root functions as in Table I, and the separate iterative algorithm is employed to implement the proposed outlier-robust Kalman filtering framework. Then, three outlier-robust Kalman filters can be obtained including SSMKF-exp-S, SSMKF-log-S, and SSMKF-sqrt-S, where SSMKF-exp-S denotes the exponential similarity function and the separate iterative algorithm based Kalman filter, and the explanations of the other two abbreviations are similar to the SSMKF-exp-S.

The proposed outlier-robust Kalman filters are compared with the standard Kalman filter with true noise covariance matrices (KFTNCM), the HKF [10], the MCKF [12], the RSTKF [22], the IMM filter [7], and the PF [3], where the true noise covariance matrices are used to obtain filtering estimates in the KFTNCM. The tuning parameter of the existing HKF is set as a common value of  $\gamma = 1.345$  [10], and the kernel size of the proposed SSMKF-exp-S and the existing MCKF is selected as  $\sigma = 10$  to achieve a tradeoff between estimation accuracy and stability [12], and the DOF parameter of the proposed SSMKF-log-S and the existing RSTKF is set as  $\nu = 10$ , and the DOF parameter of the proposed SSMKF-sqrt-S is set as  $\omega = 5$ . To guarantee the convergence of the iterations, the iteration threshold and the maximum number of iterations are, respectively, set as  $\epsilon = 10^{-16}$  and  $N_m = 50$  in the proposed filters and the existing outlier-robust Kalman filters. To guarantee the filtering stability of the proposed filters, the lower bounds of the auxiliary parameters are set as  $\delta = 10^{-8}$ . To better show the advantages of the proposed filters, two IMM filters and three particle filters are performed. In the first IMM filter (IMM-1), the true instantaneous values of state and measurement noise covariance matrices are used, and the four corresponding noise models are, respectively,  $\{\mathbf{Q}, \mathbf{R}\}$ ,  $\{\mathbf{Q}, 500\mathbf{R}\}$ ,  $\{100\mathbf{Q}, \mathbf{R}\}$ , and  $\{100\mathbf{Q}, 500\mathbf{R}\}$ , and the model transition probability matrix of the first IMM filter is set as  $\mathbf{\Pi}_1$ , where  $\mathbf{\Pi}_1(i, i) = 0.85$  and  $\mathbf{\Pi}_1(i, j) = 0.05 (i \neq j)$ . In the second IMM filter (IMM-2), the inaccurate instantaneous state and measurement noise covariance matrices are employed, and the four corresponding noise models are, respectively,  $\{\mathbf{Q}, \mathbf{R}\}$ ,  $\{\mathbf{Q}, 100\mathbf{R}\}$ ,  $\{1000\mathbf{Q}, \mathbf{R}\}$ , and  $\{1000\mathbf{Q}, 100\mathbf{R}\}$ , and the model transition probability matrix of the first IMM filter is selected as  $\mathbf{\Pi}_2$ , where all elements of  $\mathbf{\Pi}_2$  are 0.25. The initial model probability vectors of the IMM-1 and IMM-2 are both chosen as  $[0.25, 0.25, 0.25, 0.25]$ . In the first PF (PF-1) and the third PF (PF-3), the true Gaussian mixture pdfs of state and measurement noises given in (49) are used, and the particle numbers are, respectively, selected as 1000 and 500 in the PF-1 and PF-3. In the second PF (PF-2), the inaccurate Gaussian mixture pdfs of state and measurement noises are employed, where the used state and measurement noise pdfs are, respectively,  $p(\mathbf{w}_k) = 0.98N(\mathbf{w}_k; \mathbf{0}, \mathbf{Q}) + 0.02N(\mathbf{w}_k; \mathbf{0}, 1000\mathbf{Q})$  and  $p(\mathbf{v}_k) = 0.98N(\mathbf{v}_k; \mathbf{0}, \mathbf{R}) + 0.02N(\mathbf{v}_k; \mathbf{0}, 100\mathbf{R})$ . Note that the IMM-1 and PF-1 are only used as filtering references since the true instantaneous values of state and measurement noise covariance matrices and the true state and measurement noise pdfs are all unavailable in the presence of random and unknown state and measurement outliers. All filtering algorithms are coded with MATLAB and are executed on a computer with Intel Core i7-6900K CPU @ 3.20 GHz. The MATLAB codes of this article will be open access and can be freely downloaded from the link [https://www.researchgate.net/profile/Yulong\\_Huang3](https://www.researchgate.net/profile/Yulong_Huang3).

In this simulation, the simulation time is set as 200 s, and the total number of Monte Carlo runs is selected as 1000. The root-mean-square errors (RMSEs) and averaged RMSEs (ARMSEs) of position and velocity are chosen as performance metrics to compare the estimation accuracy, whose definitions are given in the literature [22]. To better exhibit the RMSEs of position and

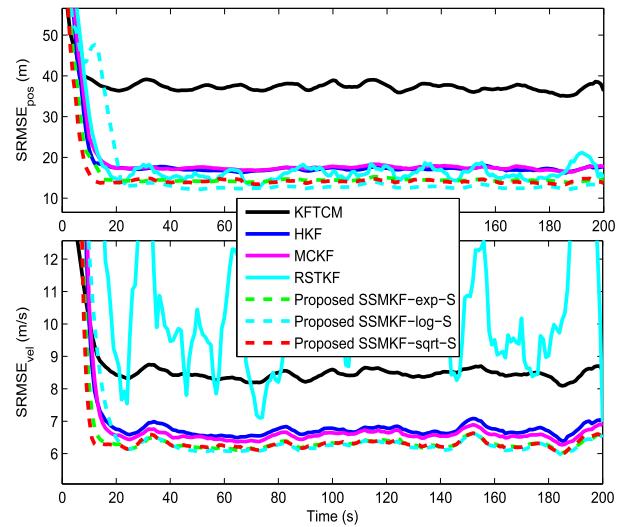


Fig. 1. RMSEs of the proposed SSMKF-exp-S, SSMKF-log-S, and SSMKF-sqrt-S and the existing outlier-robust Kalman filters.

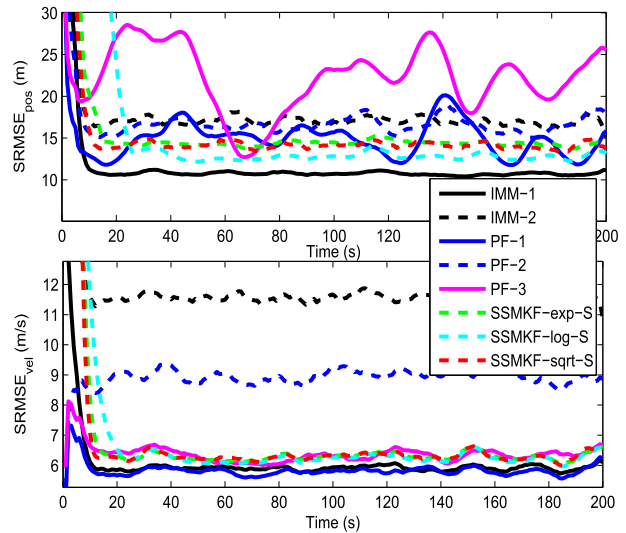


Fig. 2. RMSEs of the proposed SSMKF-exp-S, SSMKF-log-S, and SSMKF-sqrt-S and the existing IMM filters and PFs.

velocity of all filters in Figs. 1 and 2, the RMSEs are smoothed using a moving average method with span of 10 s.

## B. Simulation Results and Comparisons

The RMSEs and ARMSEs (40–200s) of position and velocity and single step run time from the proposed SSMKF-exp-S, SSMKF-log-S and SSMKF-sqrt-S and the existing outlier-robust Kalman filters are, respectively, illustrated in Fig. 1 and Table II. It can be seen from Fig. 1 and Table II that the RMSEs and ARMSEs of position and velocity from the proposed SSMKF-exp-S, SSMKF-log-S, and SSMKF-sqrt-S are all smaller than those from the existing KFTNCM, HKF, MCKF, and RSTKF. We can also see from Table II that the proposed SSMKFs all require more run time than the existing outlier-robust Kalman filters. As compared with the best  $ARMSE_{pos}$

TABLE II  
SINGLE STEP RUN TIME AND ARMSEs OVER 40–200 s

Filters	ARMSE <sub>pos</sub> (m)	ARMSE <sub>vel</sub> (m/s)	Time (ms)
KFTNCM	37.17	8.44	0.019
HKF	17.11	6.71	0.445
MCKF	17.43	6.56	0.465
RSTKF	16.12	10.48	0.505
IMM-1	10.73	5.91	0.4337
IMM-2	17.04	11.53	0.4337
PF-1	15.00	5.80	80.365
PF-2	16.65	8.95	80.365
PF-3	21.54	6.30	40.183
SSMKF-exp-S	14.38	6.28	0.563
SSMKF-log-S	12.75	6.22	1.359
SSMKF-sqrt-S	14.11	6.29	1.220

from the existing RSTKF and the best ARMSE<sub>vel</sub> from the existing MCKF, the ARMSEs of position and velocity from the proposed SSMKF-log-S improve 20.91% and 5.18%, respectively. Thus, the proposed SSMKF-exp-S, SSMKF-log-S, and SSMKF-sqrt-S all have better estimation accuracy but higher computational complexities than the existing KFTNCM, HKF, MCKF, and RSTKF.

Fig. 2 and Table II, respectively, show the RMSEs and ARMSEs (40–200 s) of position and velocity and single step run time from the proposed SSMKF-exp-S, SSMKF-log-S, and SSMKF-sqrt-S and the existing IMM filters and PFs. It is observed from Fig. 2 and Table II that the proposed SSMKF-exp-S, SSMKF-log-S, and SSMKF-sqrt-S all have smaller RMSEs and ARMSEs of position and velocity than the existing IMM-2 (inaccurate noise models), PF-2 (inaccurate noise pdfs and 1000 particles), and PF-3 (accurate noise pdfs and 500 particles). As compared with the best ARMSE<sub>pos</sub> from the PF-2 and the best ARMSE<sub>vel</sub> from the PF-3, the ARMSEs of position and velocity from the proposed SSMKF-log-S improve 23.42% and 1.27%, respectively. It can be also observed from Fig. 2 and Table II that the RMSEs and ARMSEs of position of the proposed SSMKF-log-S are close to those of the IMM-1 (filtering reference), and the RMSEs and ARMSEs of velocity of the proposed SSMKF-log-S are close to those of the PF-1 (filtering reference), and the proposed filters all have smaller RMSEs and ARMSEs of position than the PF-1. The reason that PF-1 exhibits poor estimation accuracy of position may be because the heavy-tailed features of posterior pdfs are easily lost during the particle filtering process when a limited number of particles are used. Furthermore, we can observe from Table II that the proposed SSMKFs have slightly greater run time than the IMM filters but significantly less run time than the PFs. Although the computational time of the PF can be significantly reduced if it is implemented in a parallel fashion, it still requires accurate knowledge of the probability distributions of the state and measurement noises. Thus, the proposed SSMKF-exp-S, SSMKF-log-S, and SSMKF-sqrt-S all have better estimation accuracy than the IMM-2, PF-2, and PF-3, and slightly higher computational complexities than the IMM filters but significantly lower computational complexities than the standard PFs.

## VII. CONCLUSION

In this article, a SSM was proposed to quantify the similarity between two random vectors. The measure was then employed to develop a novel outlier-robust Kalman filtering framework. Some theoretical analyses and discussions about the approximate errors and the numerical and filtering stabilities were provided to illustrate the effectiveness of the proposed framework. The fixed-point iterative algorithm and the separate iterative algorithm were proposed to implement the proposed framework, and their local convergence conditions were also provided and compared. In addition, the selections of the similarity functions were presented, and four exemplary similarity functions were provided, from which the relations between the proposed method and the existing outlier-robust Kalman filters were revealed. Simulation results illustrated that by selecting appropriate similarity functions, the proposed filters can achieve improved estimation accuracy but have higher computational complexities than the existing outlier-robust Kalman filters. Also, as compared with the existing IMM filter and PF, the proposed filters are more suitable for addressing the filtering problem of a linear system with outlier-contaminated state and measurement noises.

## APPENDIX

### A. Proof of Theorem 1

Using  $\dot{f}(t) < 0$  and  $\|\mathbf{x} - \mathbf{y}\|^2 \geq 0$  yields

$$f(\|\mathbf{x} - \mathbf{y}\|^2) \leq f(0). \quad (50)$$

Substituting (50) in (1) gives

$$s(\mathbf{x}, \mathbf{y}) = \mathbb{E} [f(\|\mathbf{x} - \mathbf{y}\|^2)] \leq \mathbb{E} [f(0)] = f(0). \quad (51)$$

Considering that the inequality (51) holds for arbitrary random vectors  $\mathbf{x}$  and  $\mathbf{y}$ , we have

$$\max s(\mathbf{x}, \mathbf{y}) = f(0). \quad (52)$$

It is evident that the SSM  $s(\mathbf{x}, \mathbf{y})$  is identical to  $f(0)$  when  $\mathbf{x} = \mathbf{y}$ , and then  $\mathbf{x} = \mathbf{y}$  is a maximum point of the SSM  $s(\mathbf{x}, \mathbf{y})$ .

### B. Proof of Theorem 2

Since  $\ddot{f}_x(t) \geq 0$  and  $\ddot{f}_z(t) \geq 0$  for  $t \in [0, +\infty)$ ,  $f_x(\cdot)$ , and  $f_z(\cdot)$  are convex functions. Using Jensen's inequality, we have

$$\int f_x(\|\mathbf{S}_{k|k-1}^{-1}(\mathbf{x}_k - \hat{\mathbf{x}}_{k|k-1})\|^2) \mathcal{N}(\mathbf{x}_k; \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k) d\mathbf{x}_k \geq \int f_x \left( \int \|\mathbf{S}_{k|k-1}^{-1}(\mathbf{x}_k - \hat{\mathbf{x}}_{k|k-1})\|^2 \mathcal{N}(\mathbf{x}_k; \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k) d\mathbf{x}_k \right) \quad (53)$$

$$\int f_z(\|\mathbf{S}_{\mathbf{R}_k}^{-1}(\mathbf{z}_k - \mathbf{H}_k \mathbf{x}_k)\|^2) \mathcal{N}(\mathbf{x}_k; \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k) d\mathbf{x}_k \geq \int f_z \left( \int \|\mathbf{S}_{\mathbf{R}_k}^{-1}(\mathbf{z}_k - \mathbf{H}_k \mathbf{x}_k)\|^2 \mathcal{N}(\mathbf{x}_k; \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k) d\mathbf{x}_k \right) \quad (54)$$

where the equalities hold if and only if the similarity functions  $f_x(\cdot)$  and  $f_z(\cdot)$  are linear or the covariance matrix of  $q(\mathbf{x}_k)$  is zero.

Substituting (53) and (54) into (8), the maximization problem with respect to the posterior mean vector and covariance matrix can be approximated as (9)–(12).

### C. Proof of Theorem 3

Using (10)–(13), the Jacobian matrix  $\Delta_{\mu_k}(\mu_k, \Sigma_k)$  is calculated as

$$\Delta_{\mu_k}(\mu_k, \Sigma_k) = -\xi_k \mathbf{P}_{k|k-1}^{-1}(\mu_k - \hat{\mathbf{x}}_{k|k-1}) + \lambda_k \mathbf{H}_k^T \mathbf{R}_k^{-1} \times (\mathbf{z}_k - \mathbf{H}_k \mu_k). \quad (55)$$

According to the maximum criterion, the maximum point  $\mu_k^*$  satisfies the following:

$$\Delta_{\mu_k}(\mu_k^*, \Sigma_k^*) = \mathbf{0}. \quad (56)$$

Utilizing (18)–(20) and (55) and (56) yields

$$-\xi_k^* \mathbf{P}_{k|k-1}^{-1}(\mu_k^* - \hat{\mathbf{x}}_{k|k-1}) + \lambda_k^* \mathbf{H}_k^T \mathbf{R}_k^{-1}(\mathbf{z}_k - \mathbf{H}_k \mu_k^*) = \mathbf{0}. \quad (57)$$

Solving (57), we can obtain the maximum point  $\mu_k^*$  as follows:

$$\mu_k^* = \left( \tilde{\mathbf{P}}_{k|k-1}^{*-1} + \mathbf{H}_k^T \tilde{\mathbf{R}}_k^{*-1} \mathbf{H}_k \right)^{-1} \left( \tilde{\mathbf{P}}_{k|k-1}^{*-1} \hat{\mathbf{x}}_{k|k-1} + \mathbf{H}_k^T \tilde{\mathbf{R}}_k^{*-1} \mathbf{z}_k \right) \quad (58)$$

Substituting (17) in (58) and employing the matrix inversion lemma [1, pp. 11–12], we can obtain (15)–(16).

### D. Proof of Theorem 4

Using (10)–(13), the Jacobian matrix  $\Delta_{\Sigma_k}(\mu_k, \Sigma_k)$  and the Hessian matrix  $\Theta_{\mu_k}(\mu_k, \Sigma_k)$  can be calculated as

$$\Delta_{\Sigma_k}(\mu_k, \Sigma_k) = -0.5\xi_k \mathbf{P}_{k|k-1}^{-1} - 0.5\lambda_k \mathbf{H}_k^T \mathbf{R}_k^{-1} \mathbf{H}_k \quad (59)$$

$$\Theta_{\mu_k}(\mu_k, \Sigma_k) = -\xi_k \mathbf{P}_{k|k-1}^{-1} - \lambda_k \mathbf{H}_k^T \mathbf{R}_k^{-1} \mathbf{H}_k + \mathbf{D}_1(\mu_k, \Sigma_k) \quad (60)$$

where  $\mathbf{D}_1(\mu_k, \Sigma_k)$  denotes the second-order term given by

$$\mathbf{D}_1(\mu_k, \Sigma_k) = 2\tilde{\xi}_k \mathbf{P}_{k|k-1}^{-1}(\mu_k - \hat{\mathbf{x}}_{k|k-1})(\mu_k - \hat{\mathbf{x}}_{k|k-1})^T \times \mathbf{P}_{k|k-1}^{-1} + 2\tilde{\lambda}_k \mathbf{H}_k^T \mathbf{R}_k^{-1}(\mathbf{z}_k - \mathbf{H}_k \mu_k)(\mathbf{z}_k - \mathbf{H}_k \mu_k)^T \mathbf{R}_k^{-1} \mathbf{H}_k. \quad (61)$$

Utilizing  $\xi_k > 0$  and  $\lambda_k > 0$  in (59), we can obtain  $\Delta_{\Sigma_k}(\mu_k, \Sigma_k) < \mathbf{0}$ , and then the maximum point  $\Sigma_k^*$  can be given by (26). Substituting (17)–(20) in (60) yields

$$\Theta_{\mu_k}(\mu_k^*, \Sigma_k^*) = -\xi_k^* \mathbf{P}_{k|k-1}^{-1} - \lambda_k^* \mathbf{H}_k^T \mathbf{R}_k^{-1} \mathbf{H}_k + \mathbf{D}_1(\mu_k^*, \Sigma_k^*). \quad (62)$$

Employing (15) and (16) in (61)  $\mathbf{D}_1(\mu_k^*, \Sigma_k^*)$  is calculated as

$$\mathbf{D}_1(\mu_k^*, \Sigma_k^*) = 2\tilde{\xi}_k^* \mathbf{P}_{k|k-1}^{-1} \tilde{\mathbf{K}}_k^* \tilde{\mathbf{P}}_{k|k-1}^{zz} (\tilde{\mathbf{K}}_k^*)^T \mathbf{P}_{k|k-1}^{-1} + 2\tilde{\lambda}_k^* \mathbf{H}_k^T \mathbf{R}_k^{-1} (\mathbf{I}_m - \mathbf{H}_k \tilde{\mathbf{K}}_k^*) \tilde{\mathbf{P}}_{k|k-1}^{zz} (\mathbf{I}_m - \mathbf{H}_k \tilde{\mathbf{K}}_k^*)^T \mathbf{R}_k^{-1} \mathbf{H}_k \quad (63)$$

where  $\tilde{\mathbf{P}}_{k|k-1}^{zz}$  denotes an approximate innovation matrix given by

$$\tilde{\mathbf{P}}_{k|k-1}^{zz} = (\mathbf{z}_k - \mathbf{H}_k \hat{\mathbf{x}}_{k|k-1})(\mathbf{z}_k - \mathbf{H}_k \hat{\mathbf{x}}_{k|k-1})^T. \quad (64)$$

It is seen from (63) that the second-order term  $\mathbf{D}_1(\mu_k^*, \Sigma_k^*)$  depends on the real-time measurement  $\mathbf{z}_k$ . As a result, it is very difficult to compare the term  $-\xi_k^* \mathbf{P}_{k|k-1}^{-1} - \lambda_k^* \mathbf{H}_k^T \mathbf{R}_k^{-1} \mathbf{H}_k$  and the second-order term  $\mathbf{D}_1(\mu_k^*, \Sigma_k^*)$ . To solve this problem, we propose a reasonable approximation as follows:

$$\tilde{\mathbf{P}}_{k|k-1}^{zz} \approx \tilde{\mathbf{P}}_{k|k-1}^{zz*} = \mathbf{H}_k \tilde{\mathbf{P}}_{k|k-1}^* \mathbf{H}_k^T + \tilde{\mathbf{R}}_k^* \quad (65)$$

where  $\tilde{\mathbf{P}}_{k|k-1}^{zz*}$  denotes the modified innovation matrix.

Exploiting (16), (26) and (65) yields

$$\begin{cases} \tilde{\mathbf{K}}_k^* \tilde{\mathbf{P}}_{k|k-1}^{zz} (\tilde{\mathbf{K}}_k^*)^T \approx \tilde{\mathbf{P}}_{k|k-1}^* - \Sigma_k^* < \tilde{\mathbf{P}}_{k|k-1}^* \\ (\mathbf{I}_m - \mathbf{H}_k \tilde{\mathbf{K}}_k^*) \tilde{\mathbf{P}}_{k|k-1}^{zz} (\mathbf{I}_m - \mathbf{H}_k \tilde{\mathbf{K}}_k^*)^T \approx \tilde{\mathbf{R}}_k^* \times \\ \left( \mathbf{H}_k \tilde{\mathbf{P}}_{k|k-1}^* \mathbf{H}_k^T + \tilde{\mathbf{R}}_k^* \right)^{-1} \tilde{\mathbf{R}}_k^* < \tilde{\mathbf{R}}_k^*. \end{cases} \quad (66)$$

Substituting (63) in (62) and using (17) and (66) obtains

$$\Theta_{\mu_k}(\mu_k^*, \Sigma_k^*) = (-\xi_k^* + 2\tilde{\xi}_k^*/\xi_k^*) \mathbf{P}_{k|k-1}^{-1} - (\lambda_k^* - 2\tilde{\lambda}_k^*/\lambda_k^*) \times \mathbf{H}_k^T \mathbf{R}_k^{-1} \mathbf{H}_k. \quad (67)$$

Employing (21) in (67), we can obtain  $\Theta_{\mu_k}(\mu_k^*, \Sigma_k^*) < \mathbf{0}$ .

### E. Proof of Proposition 1

Using (29), the cost function in (8) can be approximated as

$$\tilde{J}_1(\mu_k, \Sigma_k) = \int [f_x(Y_{1k}^*) Y_{1k} + f_z(Y_{2k}^*) Y_{2k}] N(\mathbf{x}_k; \mu_k, \Sigma_k) d\mathbf{x}_k + c_{\{\mu_k, \Sigma_k\}}. \quad (68)$$

Substituting (27) in (68) and using (11)–(13) and (18) yields

$$\tilde{J}_1(\mu_k, \Sigma_k) = -0.5\xi_k^* \text{tr}(\mathbf{A}_k \mathbf{P}_{k|k-1}^{-1}) - 0.5\lambda_k^* \text{tr}(\mathbf{B}_k \mathbf{R}_k^{-1}) + c_{\{\mu_k, \Sigma_k\}}. \quad (69)$$

Employing (69), the Jacobian matrices of  $\tilde{J}_1(\mu_k, \Sigma_k)$  with respect to  $\mu_k$  and  $\Sigma_k$  and the Hessian matrix of  $\tilde{J}_1(\mu_k, \Sigma_k)$  with respect to  $\mu_k$  are, respectively, formulated as

$$\begin{cases} \frac{\partial \tilde{J}_1(\mu_k, \Sigma_k)}{\partial \mu_k} = -\xi_k^* \mathbf{P}_{k|k-1}^{-1}(\mu_k - \hat{\mathbf{x}}_{k|k-1}) \\ + \lambda_k^* \mathbf{H}_k^T \mathbf{R}_k^{-1}(\mathbf{z}_k - \mathbf{H}_k \mu_k) \\ \frac{\partial \tilde{J}_1(\mu_k, \Sigma_k)}{\partial \Sigma_k} = -0.5\xi_k^* \mathbf{P}_{k|k-1}^{-1} - 0.5\lambda_k^* \mathbf{H}_k^T \mathbf{R}_k^{-1} \mathbf{H}_k \\ \frac{\partial^2 \tilde{J}_1(\mu_k, \Sigma_k)}{\partial \mu_k \partial \mu_k^T} = -\xi_k^* \mathbf{P}_{k|k-1}^{-1} - \lambda_k^* \mathbf{H}_k^T \mathbf{R}_k^{-1} \mathbf{H}_k. \end{cases} \quad (70)$$

According to the maximum criterion and utilizing (70), the maximum points  $\mu_k^*$  and  $\Sigma_k^*$  can be formulated as (15) and (26). Thus, the maximization problem in Theorem 2 and the maximization problem in (8) with the first-order Taylor approximations (27) and (29) have the same optimal solution.

### F. Proof of Proposition 2

Using (27) yields

$$\begin{cases} Y_{1k} = \|\mathbf{S}_{k|k-1}^{-1} \Sigma_k^{\frac{1}{2}} \boldsymbol{\tau}_x\|^2 + 2\mathbf{a}_x^T \mathbf{S}_{k|k-1}^{-1} \Sigma_k^{\frac{1}{2}} \boldsymbol{\tau}_x + \|\mathbf{a}_x\|^2 \\ \boldsymbol{\tau}_x = \Sigma_k^{-\frac{1}{2}} (\mathbf{x}_k - \boldsymbol{\mu}_k), \quad \mathbf{a}_x = \mathbf{S}_{k|k-1}^{-1} (\boldsymbol{\mu}_k - \hat{\mathbf{x}}_{k|k-1}). \end{cases} \quad (71)$$

Since the posterior mean vector and covariance matrix of the Gaussian distributed random vector  $\mathbf{x}_k$  are, respectively,  $\boldsymbol{\mu}_k$  and  $\Sigma_k$ , the random vector  $\boldsymbol{\tau}_x$  has a standard Gaussian distribution, i.e.,  $\boldsymbol{\tau}_x \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_n)$ . Employing (71) and  $\boldsymbol{\tau}_x \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_n)$ , the variance of  $Y_{1k}$  is calculated as

$$\begin{aligned} \text{Var}[Y_{1k}] &= \mathbb{E}[\|\mathbf{S}_{k|k-1}^{-1} \Sigma_k^{\frac{1}{2}} \boldsymbol{\tau}_x\|^4] - \left[ \text{tr}(\Sigma_k \mathbf{P}_{k|k-1}^{-1}) \right]^2 \\ &\quad + \left\| 2\Sigma_k^{\frac{T}{2}} \mathbf{P}_{k|k-1}^{-1} (\boldsymbol{\mu}_k - \hat{\mathbf{x}}_{k|k-1}) \right\|^2 \end{aligned} \quad (72)$$

where note that the cross variance between  $\|\mathbf{S}_{k|k-1}^{-1} \Sigma_k^{\frac{1}{2}} \boldsymbol{\tau}_x\|^2$  and  $2\mathbf{a}_x^T \mathbf{S}_{k|k-1}^{-1} \Sigma_k^{\frac{1}{2}} \boldsymbol{\tau}_x$  is zero since the odd origin moments of  $\boldsymbol{\tau}_x$  are all zeros. According to the compatibility of matrix and vector norms, we have

$$\mathbb{E}[\|\mathbf{S}_{k|k-1}^{-1} \Sigma_k^{\frac{1}{2}} \boldsymbol{\tau}_x\|^4] \leq \mathbb{E}[\|\boldsymbol{\tau}_x\|^4] \left[ \text{tr}(\Sigma_k \mathbf{P}_{k|k-1}^{-1}) \right]^2. \quad (73)$$

Considering that the random vector  $\boldsymbol{\tau}_x$  has a standard Gaussian distribution, then the random variable  $\|\boldsymbol{\tau}_x\|^2$  is a sum of the squares of  $n$  independent Gaussian random variables and has a chi-square distribution with the DOF parameter  $n$ , i.e.,  $\|\boldsymbol{\tau}_x\|^2 \sim \chi^2(n)$ . According the property of the chi-square distribution, the second-order origin moment of  $\|\boldsymbol{\tau}_x\|^2$  can be calculated as  $\mathbb{E}[\|\boldsymbol{\tau}_x\|^4] = n^2 + 2n$ . Employing  $\mathbb{E}[\|\boldsymbol{\tau}_x\|^4] = n^2 + 2n$  and (72)–(73), we can obtain the upper bound constraint of  $\text{Var}[Y_{1k}]$  in (30). Similarly, we can also derive the upper bound constraint of  $\text{Var}[Y_{2k}]$  in (30).

### G. Proof of Proposition 3

Using (34) in (17) and (25) yields

$$\tilde{\mathbf{P}}_{k|k-1}^* \leq \mathbf{P}_{k|k-1} / \xi_{\min}, \quad \tilde{\mathbf{R}}_k^* \geq \mathbf{R}_k / \lambda_{\max} \quad (74)$$

$$\begin{cases} \Sigma_k^* \geq \left( \xi_{\max} \mathbf{P}_{k|k-1}^{-1} + \lambda_{\max} \mathbf{H}_k^T \mathbf{R}_k^{-1} \mathbf{H}_k \right)^{-1} \\ \Sigma_k^* \leq \left( \xi_{\min} \mathbf{P}_{k|k-1}^{-1} + \lambda_{\min} \mathbf{H}_k^T \mathbf{R}_k^{-1} \mathbf{H}_k \right)^{-1}. \end{cases} \quad (75)$$

Choosing  $q_{\max}$  and  $p_{\max}$  as the maximum eigenvalues of  $\mathbf{P}_{k|k-1} / \xi_{\min}$  and  $(\xi_{\min} \mathbf{P}_{k|k-1}^{-1} + \lambda_{\min} \mathbf{H}_k^T \mathbf{R}_k^{-1} \mathbf{H}_k)^{-1}$  and selecting  $r_{\min}$  and  $p_{\min}$  as the minimum eigenvalues of  $\mathbf{R}_k / \lambda_{\max}$  and  $(\xi_{\max} \mathbf{P}_{k|k-1}^{-1} + \lambda_{\max} \mathbf{H}_k^T \mathbf{R}_k^{-1} \mathbf{H}_k)^{-1}$ , we can obtain (33).

### H. Proof of Proposition 4

Using (35) and (60), the modified Hessian matrix is calculated as

$$\tilde{\Theta}_{\boldsymbol{\mu}_k}(\boldsymbol{\mu}_k^{(i)}, \Sigma_k^{(i)}) = -\xi_k^{(i)} \mathbf{P}_{k|k-1}^{-1} - \lambda_k^{(i)} \mathbf{H}_k^T \mathbf{R}_k^{-1} \mathbf{H}_k. \quad (76)$$

According to the Newton's iterative scheme, we have

$$\begin{cases} \boldsymbol{\mu}_k^{(i+1)} = \boldsymbol{\mu}_k^{(i)} - \left[ \tilde{\Theta}_{\boldsymbol{\mu}_k}(\boldsymbol{\mu}_k^{(i)}, \Sigma_k^{(i)}) \right]^{-1} \Delta_{\boldsymbol{\mu}_k}(\boldsymbol{\mu}_k^{(i)}, \Sigma_k^{(i)}) \\ \Sigma_k^{(i+1)} = - \left[ \tilde{\Theta}_{\boldsymbol{\mu}_k}(\boldsymbol{\mu}_k^{(i)}, \Sigma_k^{(i)}) \right]^{-1}. \end{cases} \quad (77)$$

Employing (55) and (76) in (77) results in

$$\begin{cases} \boldsymbol{\mu}_k^{(i+1)} = \Sigma_k^{(i+1)} \left( \xi_k^{(i)} \mathbf{P}_{k|k-1}^{-1} \hat{\mathbf{x}}_{k|k-1} + \lambda_k^{(i)} \mathbf{H}_k^T \mathbf{R}_k^{-1} \mathbf{z}_k \right) \\ \Sigma_k^{(i+1)} = \left( \xi_k^{(i)} \mathbf{P}_{k|k-1}^{-1} + \lambda_k^{(i)} \mathbf{H}_k^T \mathbf{R}_k^{-1} \mathbf{H}_k \right)^{-1}. \end{cases} \quad (78)$$

By utilizing the matrix inversion lemma [1, pp. 11–12],  $\boldsymbol{\mu}_k^{(i+1)}$  and  $\Sigma_k^{(i+1)}$  in (78) can be written as the 6th and 7th equations in Algorithm 1. Thus, we can obtain Proposition 4.

### I. Proof of Theorem 5

Let  $\Delta_{\tilde{\Theta}}(\boldsymbol{\mu}_k, \Sigma_k) = \frac{\partial \tilde{\Theta}_{\boldsymbol{\mu}_k}(\boldsymbol{\mu}_k, \Sigma_k)}{\partial \boldsymbol{\mu}_k}$ . Using (11)–(13), (35) and (60) yields

$$\begin{aligned} \Delta_{\tilde{\Theta}}(\boldsymbol{\mu}_k, \Sigma_k) &= 2\tilde{\xi}_k \left[ \mathbf{P}_{k|k-1}^{-1} (\boldsymbol{\mu}_k - \hat{\mathbf{x}}_{k|k-1}) \right] \otimes \mathbf{P}_{k|k-1}^{-1} \\ &\quad - 2\tilde{\lambda}_k \left[ \mathbf{H}_k^T \mathbf{R}_k^{-1} (\mathbf{z}_k - \mathbf{H}_k \boldsymbol{\mu}_k) \right] \otimes \left[ \mathbf{H}_k^T \mathbf{R}_k^{-1} \mathbf{H}_k \right] \end{aligned} \quad (79)$$

where  $\otimes$  denotes the Kronecker product.

Taking the norm on both sides of (79) and utilizing the properties of matrix norms results in

$$\begin{aligned} \|\Delta_{\tilde{\Theta}}(\boldsymbol{\mu}_k, \Sigma_k)\|_F &\leq 2\beta_1 \tilde{\xi}_k \left\| \mathbf{S}_{k|k-1}^{-1} (\boldsymbol{\mu}_k - \hat{\mathbf{x}}_{k|k-1}) \right\|_F \\ &\quad + 2\beta_2 \tilde{\lambda}_k \left\| \mathbf{S}_{\mathbf{R}_k}^{-1} (\mathbf{z}_k - \mathbf{H}_k \boldsymbol{\mu}_k) \right\|_F \end{aligned} \quad (80)$$

where  $\|\cdot\|$  denotes the Frobenius norm, and  $\beta_1$  and  $\beta_2$  are, respectively, given by

$$\begin{cases} \beta_1 = \left\| \mathbf{S}_{k|k-1}^{-T} \right\|_F \left\| \mathbf{P}_{k|k-1}^{-1} \right\|_F \\ \beta_2 = \left\| \mathbf{H}_k^T \mathbf{S}_{\mathbf{R}_k}^{-T} \right\|_F \left\| \mathbf{H}_k^T \mathbf{R}_k^{-1} \mathbf{H}_k \right\|_F. \end{cases} \quad (81)$$

Employing (5) and (11)–(12), we have

$$\begin{cases} \left\| \mathbf{S}_{k|k-1}^{-1} (\boldsymbol{\mu}_k - \hat{\mathbf{x}}_{k|k-1}) \right\|_F \leq \sqrt{\text{tr}(\mathbf{A}_k \mathbf{P}_{k|k-1}^{-1})} \\ \left\| \mathbf{S}_{\mathbf{R}_k}^{-1} (\mathbf{z}_k - \mathbf{H}_k \boldsymbol{\mu}_k) \right\|_F \leq \sqrt{\text{tr}(\mathbf{B}_k \mathbf{R}_k^{-1})}. \end{cases} \quad (82)$$

Exploiting (13), (36) and (81)–(82) in (80) gives

$$\|\Delta_{\tilde{\Theta}}(\boldsymbol{\mu}_k, \Sigma_k)\|_F \leq 4\alpha_1 \beta_1 + 4\alpha_2 \beta_2. \quad (83)$$

Define an auxiliary function as follows:

$$\varphi(\tau) = \tilde{\Theta}_{\boldsymbol{\mu}_k}(\boldsymbol{\mu}_k^2 + \tau(\boldsymbol{\mu}_k^1 - \boldsymbol{\mu}_k^2), \Sigma_k), \quad \text{s.t.}, \tau \in [0, 1] \quad (84)$$

where  $\boldsymbol{\mu}_k^1$  and  $\boldsymbol{\mu}_k^2$  are arbitrary two posterior mean vectors.

Taking the first-order derivative of  $\varphi(\tau)$  obtains

$$\dot{\varphi}(\tau) = \Delta_{\tilde{\Theta}}(\boldsymbol{\mu}_k^2 + \tau(\boldsymbol{\mu}_k^1 - \boldsymbol{\mu}_k^2), \Sigma_k) (\boldsymbol{\mu}_k^1 - \boldsymbol{\mu}_k^2). \quad (85)$$

According to the Lagrange mean value theorem, there is a variable  $\theta \in [0, 1]$  such that the following is fulfilled:

$$\varphi(1) - \varphi(0) = \dot{\varphi}(\theta)(1 - 0). \quad (86)$$

Substituting (84)–(85) in (86) yields

$$\begin{aligned} & \tilde{\Theta}_{\mu_k}(\mu_k^1, \Sigma_k) - \tilde{\Theta}_{\mu_k}(\mu_k^2, \Sigma_k) \\ &= \Delta_{\tilde{\Theta}}(\mu_k^2 + \theta(\mu_k^1 - \mu_k^2), \Sigma_k)(\mu_k^1 - \mu_k^2). \end{aligned} \quad (87)$$

Taking the norm on both sides of (87) and using (83), we have

$$\left\| \tilde{\Theta}_{\mu_k}(\mu_k^1, \Sigma_k) - \tilde{\Theta}_{\mu_k}(\mu_k^2, \Sigma_k) \right\|_F \leq \beta \|\mu_k^1 - \mu_k^2\| \quad (88)$$

where  $\beta$  is given by

$$\beta = 4\alpha_1\beta_1 + 4\alpha_2\beta_2. \quad (89)$$

It is seen from (88) and (89) that the modified Hessian matrix  $\tilde{\Theta}_{\mu_k}(\mu_k, \Sigma_k)$  satisfies the Lipschitz condition. Thus, if the initial mean vector  $\mu_k^{(0)}$  is sufficiently close to the optimal mean vector  $\mu_k^*$ , then the fixed-point iterative algorithm has local convergence.

### J. Proof of Theorem 6

First, we prove the convergence of  $\lambda_k^{(i)(0)}$ . To this end, we consider the two cases:  $\lambda_k^{(1)(0)} \geq \lambda_k^{(0)(0)}$  and  $\lambda_k^{(1)(0)} \leq \lambda_k^{(0)(0)}$ .

*Case 1:* We first assume  $\lambda_k^{(i)(0)} \geq \lambda_k^{(i-1)(0)}$ . Then, using the 4th–8th equations in Algorithm 2 yields

$$\begin{cases} \mathbf{B}_k^{(i+1)(0)} = (\mathbf{I}_m - \mathbf{H}_k \tilde{\mathbf{K}}_k^{(i+1)(0)}) (\mathbf{z}_k - \mathbf{H}_k \hat{\mathbf{x}}_{k|k-1}) \\ \quad \times (\mathbf{z}_k - \mathbf{H}_k \hat{\mathbf{x}}_{k|k-1})^T (\mathbf{I}_m - \mathbf{H}_k \tilde{\mathbf{K}}_k^{(i+1)(0)})^T \\ \mathbf{I}_m - \mathbf{H}_k \tilde{\mathbf{K}}_k^{(i+1)(0)} = \mathbf{R}_k (\mathbf{H}_k \lambda_k^{(i)(0)} \mathbf{P}_{k|k-1} \mathbf{H}_k^T + \mathbf{R}_k)^{-1}. \end{cases} \quad (90)$$

Employing (90) and  $\lambda_k^{(i)(0)} \geq \lambda_k^{(i-1)(0)}$  results in

$$\text{tr}(\mathbf{B}_k^{(i)(0)} \mathbf{R}_k^{-1}) \geq \text{tr}(\mathbf{B}_k^{(i+1)(0)} \mathbf{R}_k^{-1}). \quad (91)$$

Since  $\dot{f}_z(t) \geq 0$ ,  $-\dot{f}_z(t)$  is a monotonically decreasing function. Utilizing (91) and the 9th equation in Algorithm 2 obtains

$$\lambda_k^{(i+1)(0)} \leq \lambda_k^{(i)(0)}. \quad (92)$$

According to the mathematical induction method, we have

$$\lambda_k^{(0)(0)} \leq \lambda_k^{(1)(0)} \leq \dots \leq \lambda_k^{(i)(0)} \leq \dots \leq \lambda_k^{(\infty)(0)}. \quad (93)$$

Considering that  $-\dot{f}_z(t)$  is a monotonically decreasing function and  $\text{tr}(\mathbf{B}_k^{(i)(0)} \mathbf{R}_k^{-1}) \geq 0$ , we obtain

$$\lambda_k^{(i)(0)} = -2\dot{f}_z \left( \text{tr}(\mathbf{B}_k^{(i)(0)} \mathbf{R}_k^{-1}) \right) \leq -2\dot{f}_z(0). \quad (94)$$

It is seen from (93) and (94) that  $\{\lambda_k^{(i)(0)}\}$  is a monotonically increasing sequence with an upper bound  $-2\dot{f}_z(0)$ . Thus, the sequence  $\{\lambda_k^{(i)(0)}\}$  converges when  $\lambda_k^{(1)(0)} \geq \lambda_k^{(0)(0)}$  and  $\dot{f}_z(0)$  has lower bound.

*Case 2:* Similar to the Case 1, if  $\lambda_k^{(1)(0)} \leq \lambda_k^{(0)(0)}$ , we have

$$\lambda_k^{(0)(0)} \geq \lambda_k^{(1)(0)} \geq \dots \geq \lambda_k^{(i)(0)} \geq \dots \geq \lambda_k^{(\infty)(0)} > 0. \quad (95)$$

It can be seen from (95) that  $\{\lambda_k^{(i)(0)}\}$  is a monotonically decreasing sequence with a lower bound 0. Thus, the sequence  $\{\lambda_k^{(i)(0)}\}$  also converges when  $\lambda_k^{(1)(0)} \leq \lambda_k^{(0)(0)}$ . Above all, the sequence  $\{\lambda_k^{(i)(0)}\}$  converges if the similarity function  $f_z(\cdot)$

satisfies Condition 3 and  $\dot{f}_z(0)$  has lower bound. Similarly, we can also prove that the sequence  $\xi_k^{(0)(j)}$  converges if the similarity function  $f_x(\cdot)$  satisfies Condition 3 and  $\dot{f}_x(0)$  has lower bound.

### K. Proof of Proposition 5

Using the Kalman measurement update in (19) and (20) yields

$$\begin{cases} \mathbf{A}_k^* = \mathbf{P}_{k|k} + \mathbf{K}_k \bar{\mathbf{P}}_{k|k-1}^{zz} \mathbf{K}_k^T \\ \mathbf{B}_k^* = (\mathbf{I}_m - \mathbf{H}_k \mathbf{K}_k) \bar{\mathbf{P}}_{k|k-1}^{zz} (\mathbf{I}_m - \mathbf{H}_k \mathbf{K}_k)^T + \\ \mathbf{H}_k \mathbf{P}_{k|k} \mathbf{H}_k^T. \end{cases} \quad (96)$$

Considering that  $\bar{\mathbf{P}}_{k|k-1}^{zz} \approx \mathbf{H}_k \mathbf{P}_{k|k-1} \mathbf{H}_k^T + \mathbf{R}_k$  for the case of Gaussian noises and employing (96) results in

$$\begin{cases} \mathbf{A}_k^* \approx \mathbf{P}_{k|k} + \mathbf{K}_k (\mathbf{H}_k \mathbf{P}_{k|k-1} \mathbf{H}_k^T + \mathbf{R}_k) \mathbf{K}_k^T \\ \mathbf{B}_k^* \approx \mathbf{R}_k - \mathbf{H}_k \mathbf{P}_{k|k-1} \mathbf{H}_k^T + \mathbf{H}_k \mathbf{K}_k \times \\ (\mathbf{H}_k \mathbf{P}_{k|k-1} \mathbf{H}_k^T + \mathbf{R}_k) \mathbf{K}_k^T \mathbf{H}_k^T + \mathbf{H}_k \mathbf{P}_{k|k} \mathbf{H}_k^T. \end{cases} \quad (97)$$

Substituting the Kalman measurement update in (97), we can obtain (39).

### L. Proof of Proposition 6

Using (42) and (43) yields

$$\begin{cases} \eta_1 = \text{tr}(\mathbf{A}_k^* \mathbf{P}_{k|k-1}^{-1}) = \text{tr}(\Psi_{k1} \mathbf{P}_{k|k-1}^{-1}) + n \geq n \\ \eta_2 = \text{tr}(\mathbf{B}_k^* \mathbf{R}_k^{-1}) = \text{tr}(\Psi_{k2} \mathbf{R}_k^{-1}) + m \geq m. \end{cases} \quad (98)$$

Substituting (98) in (18) results in

$$\xi_k^* = -2\dot{f}_x(\eta_1), \quad \lambda_k^* = -2\dot{f}_z(\eta_2). \quad (99)$$

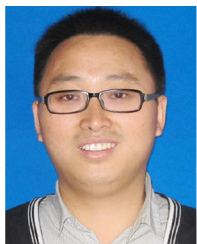
Employing (44) in (99), we can obtain (45). Moreover, the larger state and measurement outliers, the larger auxiliary matrices  $\Psi_{k1}$  and  $\Psi_{k2}$  are generated. Then, the larger  $\eta_1$  and  $\eta_2$  are obtained, based on which the smaller modified auxiliary parameters  $\xi_k^*$  and  $\lambda_k^*$  are achieved by using (44).

## REFERENCES

- [1] D. Simon, *Optimal State Estimation: Kalman, H infinity, and Nonlinear Approaches*. Hoboken, NJ, USA: Wiley, 2006.
- [2] 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.
- [3] S. Arulampalam, S. Maskell, N. Gordon, and T. Clapp, "A tutorial on particle filters for on-line non-linear/non-Gaussian Bayesian tracking," *IEEE Trans. Signal Process.*, vol. 50, no. 2, pp. 174–189, Feb. 2002.
- [4] N. Gordon and A. Smith, "Approximate non-Gaussian Bayesian estimation and modal consistency," *J. Roy. Statistical Soc. B.*, vol. 55, no. 4, pp. 913–918, Apr. 1993.
- [5] S. Li, H. Wang, and T. Chai, "A t-distribution based particle filter for target tracking," in *Proc. Amer. Control Conf.*, Jun. 2006, pp. 2191–2196.
- [6] D. L. Alspach and H. Sorenson, "Non-linear Bayesian estimation using Gaussian sum approximation," *IEEE Trans. Autom. Control*, vol. AC-17, no. 4, pp. 439–448, Apr. 1972.
- [7] H. A. P. Blom and Y. Bar-Shalom, "The interacting multiple model algorithm for systems with Markovian switching coefficients," *IEEE Trans. Autom. Control*, vol. 33, no. 8, pp. 780–783, Aug. 1988.
- [8] P. J. Huber, "Robust statistics" in, *International Encyclopedia of Statistical Science*. Berlin, Germany: Springer, 2011.
- [9] M. A. Gandhi and L. Mili, "Robust Kalman filter based on a generalized maximum-likelihood-type estimator," *IEEE Trans. Signal Process.*, vol. 58, no. 5, pp. 2509–2520, May 2010.



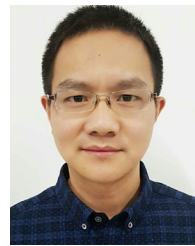
- [10] C. D. Karlgaard, and H. Schaub, "Huber-based divided difference filtering," *J. Guid., Control, Dyn.*, vol. 30, no. 3, pp. 885–891, 2007.
- [11] G. T. Cinar and J. C. Principe, "Hidden state estimation using the Correntropy Filter with fixed point update and adaptive kernel size," in *Proc. IEEE World Congr. Comput. Intell.*, Brisbane, Australia, 2012, pp. 1–6.
- [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] Y. Wang, W. Zheng, S. Sun, and L. Li, "Robust information filter based on maximum correntropy criterion," *J. Guid., Control, Dyn.*, vol. 39, no. 5, pp. 1126–1131, May 2016.
- [14] 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 Signal Process.*, May 2013, pp. 5770–5774.
- [15] Y. Huang, Y. Zhang, N. Li, and J. Chambers, "Robust Student's t based nonlinear filter and smoother," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 52, no. 5, pp. 2586–2596, Oct. 2016.
- [16] M. Roth, T. Ardeschiri, E. Özkan, and F. Gustafsson, "Robust Bayesian filtering and smoothing using Student's t distribution," *arXiv:1703.02428*.
- [17] J. Ting, E. Theodorou, and S. Schaal, "Learning an outlier-robust Kalman filter," in *Proc. 18th Euro. Conf. Mach. Learn. Warsaw*, Poland, 2007, pp. 748–756.
- [18] R. Piché, S. Särkkä, and J. Hartikainen, "Recursive outlier-robust filtering and smoothing for nonlinear systems using the multivariate Student-t distribution," in *Proc. IEEE Int. Workshop Mach. Learn. Signal Process.*, Sep. 2012, pp. 1–6.
- [19] H. Zhu, H. Leung, and Z. He, "A variational Bayesian approach to robust sensor fusion based on Student-t distribution," *Inf. Sci.*, vol. 221, no. 2013, pp. 201–214, Sep. 2012.
- [20] G. Agamennoni, J. I. Nieto, and E. M. Nebot, "Approximate inference in state-space models with heavy-tailed noise," *IEEE Trans. Signal Process.*, vol. 60, no. 10, pp. 5024–5037, Oct. 2012.
- [21] S. S. Khalid, N. U. Rehman, S. Abrar, and L. Mihaylova, "Robust Bayesian filtering using Bayesian model averaging and restricted variational Bayes," in *Proc. 21st Int. Conf. Inf. Fusion*, Jul. 2018, pp. 361–368.
- [22] 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.
- [23] Y. Huang, Y. Zhang, B. Xu, Z. Wu, and J. Chambers, "A new outlier-robust Student's t based Gaussian approximate filter for cooperative localization," *IEEE/ASME Trans. Mechatronics*, vol. 22, no. 5, pp. 2380–2386, Oct. 2017.
- [24] 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. Autom. Control*, vol. 63, no. 2, pp. 594–601, Feb. 2018.
- [25] B. Chen, L. Xing, H. Zhao, N. Zheng, and J. C. Principe, "Generalized correntropy for robust adaptive filtering," *IEEE Trans. Signal Process.*, vol. 64, no. 13, pp. 3376–3387, Jul. 2016.
- [26] D. Shi, T. Chen, and L. Shi, "Event-triggered maximum likelihood state estimation," *Automatica*, vol. 50, no. 1, pp. 247–254, Jan. 2014.
- [27] K. Xiong, H. Zhang, and C. W. Chan, "Performance evaluation of UKF-based nonlinear filtering," *Automatica*, vol. 42, no. 2, pp. 261–270, Feb. 2006.
- [28] 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 Trans. Syst., Man, Cybernet., Syst.*, vol. 49, no. 10, pp. 2082–2096, Oct. 2019.



**Yulong Huang** (Member, IEEE) received the B.S. degree in automation and the Ph.D. degree in control science and engineering from the College of Automation, Harbin Engineering University, Harbin, China, in 2012 and 2018, respectively.

He is currently an Associate Professor of navigation, guidance, and control with Harbin Engineering University. His current research interests include signal processing, information fusion and their applications. Dr. Huang is the

recipient of Excellent Doctoral Thesis from the Chinese Association of Automation in 2019.



**Yonggang Zhang** (Senior Member, IEEE) received the B.S. and M.S. degrees in automation from the College of Automation, Harbin Engineering University, Harbin, China, in 2002 and 2004, respectively, and the Ph.D. degree in electronic engineering from Cardiff University, Cardiff, U.K., in 2007.

He is currently a Professor of navigation, guidance, and control with Harbin Engineering University. His current research interests include signal processing, information fusion and their applications in navigation technology, such as fiber optical gyroscope, inertial navigation, and integrated navigation.



**Yuxin Zhao** (Senior Member, IEEE) received the Ph.D. degree in navigation, guidance and control from the Harbin Engineering University, Harbin, China, in 2005.

He is currently the Head and a Professor with the College of Intelligent Systems Science and Engineering, Harbin Engineering University. His research interests include artificial intelligence, filtering theory, marine navigation system and intelligent transportation system.



**Peng Shi** (Fellow, IEEE) received the Ph.D. degree in electrical engineering from the University of Newcastle, NSW, Australia, in 1994, the D.Sc. degree from the University of Glamorgan, Wales, U.K., in 2006, and the D.Eng. degree from the University of Adelaide, Adelaide, Australia, in 2015.

He is currently a Professor with the University of Adelaide. His research interests include system and control theory, intelligent systems, and operational research.

He is (was) in the editorial board of a number of journals, including *Automatica*; IEEE TRANSACTIONS ON AUTOMATIC CONTROL. He is an IEEE Systems, Man, and Cybernetics Society Distinguished Lecturer, a Fellow of the Institution of Engineering and Technology and the Institute of Engineers, Australia.



**Jonathon A. Chambers** (Fellow, IEEE) received the Ph.D. and D.Sc. degrees in signal processing from the Imperial College of Science, Technology, and Medicine (Imperial College London), London, U.K., in 1990 and 2014, respectively.

He is Emeritus Professor of signal and information processing and Former Head of the School of Engineering, University of Leicester, Leicester, United Kingdom. His research interests include adaptive signal processing and machine learning and their applications.

Dr. Chambers is a Fellow of the Royal Academy of Engineering, U.K., and the Institution of Electrical Engineers. He has served as an Associate Editor for the IEEE TRANSACTIONS ON SIGNAL PROCESSING for three terms over the periods 1997–1999, 2004–2007, and as a Senior Area Editor between 2011–2014. Since 2016, he has been an Area Editor for the Elsevier *DSP journal*.