A Novel Kullback–Leibler Divergence Minimization-Based Adaptive Student's t-Filter

Yulong Huang , *Member, IEEE*, Yonggang Zhang , *Senior Member, IEEE*, and Jonathon A. Chambers , *Fellow, IEEE*

Abstract—In this paper, in order to improve the Student's t-matching accuracy, a novel Kullback-Leibler divergence (KLD) minimization-based matching method is firstly proposed by minimizing the upper bound of the KLD between the true Student's t-density and the approximate Student's t-density. To improve the Student's t-modelling accuracy, a novel KLD minimization-based adaptive method is then proposed to estimate the scale matrices of Student's t-distributions, in which the modified evidence lower bound is maximized. A novel KLD minimization-based adaptive Student's t-filter is derived via combining the proposed Student's t-matching technique and the adaptive method. A manoeuvring target tracking example is provided to demonstrate the effectiveness and potential of the proposed filter.

Index Terms—Adaptive filter, Kullback-Leibler divergence minimization, Student's t-distribution, upper bound minimization, lower bound maximization.

I. INTRODUCTION

THE non-Gaussian filtering problem of a state-space model (SSM) with heavy-tailed noises has been attracting more and more attention. The non-Gaussian heavy-tailed noises, which are often induced by unknown outliers, stochastic impulse interferences, and uncertain modelling errors, may be encountered in some engineering contexts, such as positioning, navigation and target tracking [1], [2]. Such non-Gaussian filtering is always a challenging problem since the non-Gaussian posterior probability density function (PDF) cannot be recursively formulated by a non-Gaussian PDF in a closed form [3].

As a popular non-Gaussian filter, the particle filter (PF) is able to solve the non-Gaussian filtering problem with heavy-tailed noises [4]. In the PF, the non-Gaussian heavy-tailed noises are modelled by non-Gaussian heavy-tailed distributions, and the posterior PDF is represented by a sufficiently large number of weighted particles, from which an approximate non-Gaussian filtering estimate can be achieved [4]–[6]. In theory, the PF can achieve an optimal filtering estimate for this non-Gaussian filtering problem when an infinite number of particles are utilized [4]. However, the PF is prone to particle degeneracy and dilution when a finite number of particles are employed [4]. To guarantee the filtering accuracy, a large number of particles are required by the PF, especially for a high-dimensional SSM, which results in substantial computational complexity [7]. As an alternative non-Gaussian filter, the Gaussian sum filter can also be employed to solve the non-Gaussian filtering problem with heavy-tailed noises [8]. In the Gaussian sum filter, a set of weighted Gaussian distributions are employed to model the non-Gaussian heavy-tailed noises, and then the non-Gaussian posterior PDF can be updated as a weighted sum of Gaussian PDFs that are achieved by running a group of sub Kalman filters (KFs) [8]–[10]. Unfortunately, the non-Gaussian heavy-tailed noises are very difficult to model based on a limited number of fixed Gaussian distributions since the true noise distributions are often unknown in engineering contexts, which degrades the filtering accuracy of the Gaussian sum filter dramatically.

In recent decades, a large number of computationally-efficient noise-robust Kalman filtering variants have been proposed to acquire a tradeoff between the computational complexity and the filtering accuracy. As a well-known outlier-robust technique, the M-estimate, which employs the influence function approach to resist the observation outliers, has been successfully employed to design an outlier-robust KF [11]. Many outlier-robust KFs have been derived on the basis of the M-estimate technique by selecting a robust cost function to restrain the one-step prediction error and estimated residual error [12]–[14]. As the most successful extension of the M-estimate to the KF setting, the Huber KF (HKF) chooses a weighted sum of l_1 and l_2 norms as a robust cost function and provides a generalized robust maximum likelihood estimate by minimizing the Huber cost function [15]. The maximum correntropy KF (MCKF) [16]–[18] is a novel noise-robust KF which has strong ability to suppress impulsive noise interferences in many practical applications. Essentially, the MCKF is also a variant of the M-estimator in the KF setting since the statistical correntropy is approximated by the sum of the Gaussian kernel functions, and the MCKF is derived by maximising the sum of the Gaussian kernel functions of the one-step prediction error and estimated residual error. The heavy-tailed features inherent in noises are ignored by the existing HKF and MCKF, which degrades the

filtering accuracy. In order to better exploit such heavy-tailed features, the Student's t-distribution has been employed to model heavy-tailed noises. Two noise-robust filters have been proposed based on Student's t-modelling: the robust Student's t-based KF (RSTKF) and the Student's t-filter (STF) [19]–[30]. For the STF, the predicted state and observation are jointly modelled as Student's t-distributed, and then the posterior filtering PDF is recursively approximated by a Student's t-PDF with fixed degrees of freedom (dof) parameter based on the Bayesian rule and moment matching method [19]–[21]. Nevertheless, for the RSTKF, the one-step prediction PDF and the observation likelihood PDF are modelled by Student's t-PDFs and formulated as Gaussian-Gamma mixture forms, based on which the posterior filtering PDF is approximated by a Gaussian PDF with adaptively selected covariance matrix using the variational Bayesian approach [28]–[30]. For a linear SSM with the moderately contaminated state and observation noises, the RSTKF has better filtering accuracy than the STF because the posterior filtering PDF can be better approximated by the Gaussian distribution as compared with the Student's t-distribution [28]. On the contrary, for a linear SSM with strongly contaminated state and observation noises, the STF has better filtering accuracy than the RSTKF since the posterior filtering PDF may be approximated by the Student's t-distribution better than the Gaussian distribution.

Although the STF is able to address such strongly heavy-tailed state and observation noises well, it suffers from two major drawbacks. Firstly, the Student's t-based measurement update to the posterior filtering PDF is not strictly closed due to the increase of the dof parameter in the measurement update. In order to keep a closed update form, the moment matching method is utilized to achieve a Student's t-approximation with a fixed dof parameter to the true posterior filtering PDF. However, the moment matching method can only capture the first and second moments of the true posterior filtering PDF, thus some higher-order moments and non-Gaussian information are lost. Secondly, in practical engineering applications, it is very difficult to select accurate scale matrices to model heavy-tailed state and observation noises well because such non-Gaussian heavy-tailed noises are often induced by unknown outliers, stochastic impulse interferences, and uncertain modelling errors. The above two problems will degrade the Student's t-approximation accuracy to the posterior filtering PDF, which then degrades dramatically the filtering accuracy of the existing STF.

In this paper, we aim to improve the filtering accuracy of the existing STF by improving the Student's t-matching accuracy and the Student's t-modelling accuracy. A novel Kullback-Leibler divergence (KLD) minimization-based matching method is first proposed to improve the Student's t-matching accuracy, in which the upper bound of the KLD between the true Student's t-PDF and the approximate Student's t-PDF is minimized. To improve the Student's t- modelling accuracy, a novel KLD minimization-based adaptive method is then proposed to estimate the scale matrices of the Student's t-distributions, in which the modified evidence lower bound is maximized. A novel KLD minimization-based adaptive Student's t-filter (ASTF) is derived via combining the proposed Student's t-matching technique and the adaptive method.

Simulation examples illustrate that the proposed ASTF has improved filtering accuracy over the existing HKF, MCKF, STF and RSTKF for strongly heavy-tailed state and observation noises.

The structure of this paper is as follows. In Section II, the problem formulation is presented. In Section III, a novel KLD minimization-based Student's t-matching method is proposed to improve the Student's t-matching accuracy. In Section IV, a novel KLD minimization-based ASTF is proposed to improve the filtering accuracy. In Section V, simulation results and comparisons with the existing HKF, MCKF, STF and RSTKF are given. Conclusions are drawn in Section VI.

II. PROBLEM FORMULATION

Consider a linear dynamical system described by a linear discrete-time SSM as follows

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

where 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 observation 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 observation matrix, and $\mathbf{w}_k \in \mathbb{R}^n$ and $\mathbf{v}_k \in \mathbb{R}^m$ are, respectively, state and observation noise vectors. In this paper, the initial state vector $\mathbf{x}_0$, the state noise vector $\mathbf{w}_k$, and the observation noise vector $\mathbf{v}_k$ are assumed to be mutually uncorrelated and have non-Gaussian heavy-tailed distributions, and they are modelled as Student's t-distributed as follows

$$\begin{cases} p(\mathbf{x}_0) = \mathrm{St}(\mathbf{x}_0; \hat{\mathbf{x}}_{0|0}, \mathbf{P}_{0|0}, \nu) \\ p(\mathbf{w}_k) = \mathrm{St}(\mathbf{w}_k; \mathbf{0}, \mathbf{Q}_k, \nu) \\ p(\mathbf{v}_k) = \mathrm{St}(\mathbf{v}_k; \mathbf{0}, \mathbf{R}_k, \nu) \end{cases} \tag{2}$$

where $\mathrm{St}(\cdot; \mu, \boldsymbol{\Sigma}, \omega)$ denotes the Student's t-PDF with mean vector μ, scale matrix $\boldsymbol{\Sigma}$ and dof parameter ω, and $\hat{\mathbf{x}}_{0|0}$ and $\mathbf{P}_{0|0}$ are, respectively, the initial state estimate and scale matrix of $\mathbf{x}_0$, and $\mathbf{Q}_k$ and $\mathbf{R}_k$ are, respectively, the scale matrices of $\mathbf{w}_k$ and $\mathbf{v}_k$, and ν is the common dof parameter of $\mathbf{x}_0$, $\mathbf{w}_k$ and $\mathbf{v}_k$.

The STF can be used to address the filtering problem of the above linear SSM with non-Gaussian heavy-tailed state and observation noises. Similar to the KF, the recursive STF is also composed of time and measurement updates. In the time update, the jointly predicted PDF $p(\mathbf{x}_k, \mathbf{z}_k | \mathbf{z}_{1:k-1})$ of state and observation vectors is approximated by a Student's t-PDF as follows [19], [20]

$$p(\mathbf{x}_k, \mathbf{z}_k | \mathbf{z}_{1:k-1}) = \mathrm{St}\left(\begin{bmatrix} \mathbf{x}_k \\ \mathbf{z}_k \end{bmatrix}; \begin{bmatrix} \hat{\mathbf{x}}_{k|k-1} \\ \mathbf{H}_k \hat{\mathbf{x}}_{k|k-1} \end{bmatrix}, \right.$$

$$\left. \times \begin{bmatrix} \mathbf{P}_{k|k-1} & \mathbf{P}_{k|k-1} \mathbf{H}_k^T \\ \mathbf{H}_k \mathbf{P}_{k|k-1} & \mathbf{H}_k \mathbf{P}_{k|k-1} \mathbf{H}_k^T + \mathbf{R}_k \end{bmatrix}, \nu \right) \tag{3}$$

where the predicted mean vector $\hat{\mathbf{x}}_{k|k-1}$ and the predicted scale matrix $\mathbf{P}_{k|k-1}$ are given by

$$\hat{\mathbf{x}}_{k|k-1} = \mathbf{F}_{k-1} \hat{\mathbf{x}}_{k-1|k-1} \tag{4}$$

$$\mathbf{P}_{k|k-1} = \mathbf{F}_{k-1} \mathbf{P}_{k-1|k-1} \mathbf{F}_{k-1}^T + \mathbf{Q}_k \tag{5}$$

where $\hat{\mathbf{x}}_{k-1|k-1}$ and $\mathbf{P}_{k-1|k-1}$ are, respectively, the mean vector and scale matrix of the posterior PDF at time $k-1$.

In the measurement update, the posterior PDF $p(\mathbf{x}_k|\mathbf{z}_{1:k})$ is updated as a Student's t-distribution based on the Bayesian rule using (3) as follows

$$p(\mathbf{x}_k|\mathbf{z}_{1:k}) = \mathrm{St}(\mathbf{x}_k; \hat{\mathbf{x}}'_{k|k}, \mathbf{P}'_{k|k}, \nu') \tag{6}$$

where $\hat{\mathbf{x}}'_{k|k}$, $\mathbf{P}'_{k|k}$ and ν' are, respectively, the mean vector, scale matrix and dof parameter of the posterior PDF $p(\mathbf{x}_k|\mathbf{z}_{1:k})$, which are given by

$$\mathbf{S}_k = \mathbf{H}_k \mathbf{P}_{k|k-1} \mathbf{H}_k^T + \mathbf{R}_k \tag{7}$$

$$\Delta_k = \sqrt{(\mathbf{z}_k - \mathbf{H}_k \hat{\mathbf{x}}_{k|k-1})^T \mathbf{S}_k^{-1} (\mathbf{z}_k - \mathbf{H}_k \hat{\mathbf{x}}_{k|k-1})} \tag{8}$$

$$\mathbf{K}_k = \mathbf{P}_{k|k-1} \mathbf{H}_k^T \mathbf{S}_k^{-1} \tag{9}$$

$$\hat{\mathbf{x}}'_{k|k} = \hat{\mathbf{x}}_{k|k-1} + \mathbf{K}_k (\mathbf{z}_k - \mathbf{H}_k \hat{\mathbf{x}}_{k|k-1}) \tag{10}$$

$$\mathbf{P}'_{k|k} = \frac{\nu + \Delta_k^2}{\nu + m} (\mathbf{P}_{k|k-1} - \mathbf{K}_k \mathbf{H}_k \mathbf{P}_{k|k-1}) \tag{11}$$

$$\nu' = \nu + m \tag{12}$$

The Student's t-approximation is not strictly closed for the posterior PDF due to the increase of the dof parameter in the measurement update. In order to keep a closed update form for the posterior PDF, the true posterior PDF needs to be approximated as an approximate posterior PDF with a fixed dof parameter ν, i.e.,

$$p(\mathbf{x}_k|\mathbf{z}_{1:k}) = \mathrm{St}(\mathbf{x}_k; \hat{\mathbf{x}}'_{k|k}, \mathbf{P}'_{k|k}, \nu') \approx \mathrm{St}(\mathbf{x}_k; \hat{\mathbf{x}}_{k|k}, \mathbf{P}_{k|k}, \nu) \tag{13}$$

where $\hat{\mathbf{x}}_{k|k}$ and $\mathbf{P}_{k|k}$ denote the mean vector and scale matrix of the approximate posterior PDF, respectively.

The existing STF employs the moment matching method to achieve $\hat{\mathbf{x}}_{k|k}$ and $\mathbf{P}_{k|k}$ via matching the first and second moments of the true posterior PDF $\mathrm{St}(\mathbf{x}_k; \hat{\mathbf{x}}'_{k|k}, \mathbf{P}'_{k|k}, \nu')$ and that of the approximate posterior PDF $\mathrm{St}(\mathbf{x}_k; \hat{\mathbf{x}}_{k|k}, \mathbf{P}_{k|k}, \nu)$ [19], [20], i.e.,

$$\hat{\mathbf{x}}_{k|k} = \hat{\mathbf{x}}'_{k|k}, \qquad \frac{\nu}{\nu-2} \mathbf{P}_{k|k} = \frac{\nu'}{\nu'-2} \mathbf{P}'_{k|k} \tag{14}$$

The existing STF is able to cope with the strongly heavy-tailed state and observation noises. Unfortunately, the existing STF suffers from these two major problems as follows.

- The moment matching method used in the existing STF cannot capture the higher order moments and some non-Gaussian information, which results in poor Student's t-matching accuracy.
- In practical engineering applications, it is very difficult to model the unknown non-Gaussian heavy-tailed state and observation noises accurately in advance, which leads to poor Student's t-modelling accuracy.

The poor Student's t-matching accuracy and Student's t-modelling accuracy will degrade the Student's t-approximation accuracy to the posterior PDF, and then the filtering accuracy of the existing STF will degrade, which represents the main research motivation of this paper. In this paper,

we aim to improve the filtering accuracy of the existing STF by addressing the above two problems. Next, a novel KLD minimization-based matching technique will be first proposed to improve the Student's t-matching accuracy, and a novel KLD minimization-based adaptive method will then be proposed to estimate the scale matrices of the Student's t-distributions to improve the Student's t- modelling accuracy, based on which a novel KLD minimization-based ASTF will be derived.

III. A NOVEL KLD MINIMIZATION-BASED STUDENT'S t-MATCHING METHOD

A. KLD Minimization-Based Student's t-Matching Method

The approximation problem in (13) is formulated as a general Student's t-matching problem as follows. The Student's t-matching aims to look for an optimal Student's t-PDF $\mathrm{St}(\mathbf{x}; \boldsymbol{\mu}_2, \boldsymbol{\Sigma}_2, \nu_2)$ with a fixed dof parameter ν_2 to approximate an arbitrary and known Student's t-PDF $\mathrm{St}(\mathbf{x}; \boldsymbol{\mu}_1, \boldsymbol{\Sigma}_1, \nu_1)$, where $\nu_1 > \nu_2$. To match more higher order moments and non-Gaussian information, the KLD criterion is used to select an optimal Student's t-PDF. That is to say, the mean vector $\boldsymbol{\mu}_2$ and the scale matrix $\boldsymbol{\Sigma}_2$ of the optimal Student's t-PDF is achieved by minimizing the KLD between the true Student's t-PDF and the optimal Student's t-approximation, i.e.,

$$\{\boldsymbol{\mu}_2, \boldsymbol{\Sigma}_2\} = \arg \min_{\{\boldsymbol{\mu}, \boldsymbol{\Sigma}\}} \mathrm{KLD}(p(\mathbf{x})||q(\mathbf{x})) \tag{15}$$

where $\mathrm{KLD}(\cdot||\cdot)$ denotes the KLD measure, and $p(\mathbf{x})$ and $q(\mathbf{x})$ denote the true Student's t-PDF and the approximate Student's t-PDF with a fixed dof parameter ν_2, respectively, given by

$$\begin{cases} p(\mathbf{x}) = \mathrm{St}(\mathbf{x}; \boldsymbol{\mu}_1, \boldsymbol{\Sigma}_1, \nu_1) \\ q(\mathbf{x}) = \mathrm{St}(\mathbf{x}; \boldsymbol{\mu}, \boldsymbol{\Sigma}, \nu_2) \end{cases} \tag{16}$$

According to the definition of the KLD, $\mathrm{KLD}(p(\mathbf{x})||q(\mathbf{x}))$ can be written as [31]

$$\begin{aligned} \mathrm{KLD}(p(\mathbf{x})||q(\mathbf{x})) &= \int p(\mathbf{x}) \log \frac{p(\mathbf{x})}{q(\mathbf{x})} d\mathbf{x} \\ &= \int p(\mathbf{x}) \log p(\mathbf{x}) d\mathbf{x} \\ &\quad - \int p(\mathbf{x}) \log q(\mathbf{x}) d\mathbf{x} \end{aligned} \tag{17}$$

Considering that the term $\int p(\mathbf{x}) \log p(\mathbf{x}) d\mathbf{x}$ is independent of $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ and using (16)–(17) in (15), the minimization problem (15) can be reformulated as

$$\{\boldsymbol{\mu}_2, \boldsymbol{\Sigma}_2\} = \arg \min_{\{\boldsymbol{\mu}, \boldsymbol{\Sigma}\}} J(\boldsymbol{\mu}, \boldsymbol{\Sigma}) \tag{18}$$

where the cost function $J(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ is given by

$$J(\boldsymbol{\mu}, \boldsymbol{\Sigma}) = - \int \mathrm{St}(\mathbf{x}; \boldsymbol{\mu}_1, \boldsymbol{\Sigma}_1, \nu_1) \log \mathrm{St}(\mathbf{x}; \boldsymbol{\mu}, \boldsymbol{\Sigma}, \nu_2) d\mathbf{x} \tag{19}$$

Using the definition of the Student's t-PDF, $\log \mathrm{St}(\mathbf{x}; \boldsymbol{\mu}, \boldsymbol{\Sigma}, \nu_2)$ is written as

$$\log \mathrm{St}(\mathbf{x}; \boldsymbol{\mu}, \boldsymbol{\Sigma}, \nu_2) = \log \Gamma\left(\frac{\nu_2 + n}{2}\right) - \log \Gamma\left(\frac{\nu_2}{2}\right)$$

$$- 0.5 \log |\boldsymbol{\Sigma}| - 0.5 n \log(\nu_2 \pi) - 0.5(\nu_2 + n)$$

$$\times \log\left[1 + \frac{1}{\nu_2}(\mathbf{x} - \mu)^{\mathrm{T}} \boldsymbol{\Sigma}^{-1}(\mathbf{x} - \mu)\right] \quad (20)$$

where n represents the dimension of random vector $\mathbf{x}$.

Employing (20) in (19) yields

$$J(\boldsymbol{\mu}, \boldsymbol{\Sigma}) = 0.5 \log |\boldsymbol{\Sigma}| + 0.5(\nu_2 + n)\int \mathrm{St}(\mathbf{x}; \boldsymbol{\mu}_1, \boldsymbol{\Sigma}_1, \nu_1)$$

$$\times \log\left[1 + \frac{1}{\nu_2}(\mathbf{x} - \mu)^{\mathrm{T}} \boldsymbol{\Sigma}^{-1}(\mathbf{x} - \mu)\right] d\mathbf{x} + c_{\{\boldsymbol{\mu}, \boldsymbol{\Sigma}\}} \quad (21)$$

where c_θ represents a constant independent of variable θ.

However, it is very difficult to solve the minimization problem (18) using (21) because the nonlinear integral in (21) cannot be analytically achieved. To address this problem, we propose to minimize the upper bound of the cost function $J(\boldsymbol{\mu}, \boldsymbol{\Sigma})$.

Considering that the natural logarithmic function $\log(\cdot)$ is a convex function and using Jensen's inequality, we have [32]

$$\int \mathrm{St}(\mathbf{x}; \boldsymbol{\mu}_1, \boldsymbol{\Sigma}_1, \nu_1) \log\left[1 + \frac{1}{\nu_2}(\mathbf{x} - \mu)^{\mathrm{T}} \boldsymbol{\Sigma}^{-1}(\mathbf{x} - \mu)\right] d\mathbf{x}$$

$$\leq \log f(\boldsymbol{\mu}, \boldsymbol{\Sigma}) \quad (22)$$

where $\log f(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ is the upper bound of the nonlinear integral, and the auxiliary function $f(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ is given by

$$f(\boldsymbol{\mu}, \boldsymbol{\Sigma}) = \int \mathrm{St}(\mathbf{x}; \boldsymbol{\mu}_1, \boldsymbol{\Sigma}_1, \nu_1)$$

$$\times \left[1 + \frac{1}{\nu_2}(\mathbf{x} - \mu)^{\mathrm{T}} \boldsymbol{\Sigma}^{-1}(\mathbf{x} - \mu)\right] d\mathbf{x}$$

$$= 1 + \frac{1}{\nu_2} \mathrm{tr}\left\{(\boldsymbol{\mu} - \boldsymbol{\mu}_1)(\boldsymbol{\mu} - \boldsymbol{\mu}_1)^{\mathrm{T}} \boldsymbol{\Sigma}^{-1}\right\}$$

$$+ \frac{1}{\nu_2} \mathrm{tr}(\tilde{\boldsymbol{\Sigma}}_1 \boldsymbol{\Sigma}^{-1}) \quad (23)$$

where $\tilde{\boldsymbol{\Sigma}}_1$ denotes the covariance matrix of the Student's t-PDF $p(\mathbf{x})$ given by

$$\tilde{\boldsymbol{\Sigma}}_1 = \frac{\nu_1}{\nu_1 - 2} \boldsymbol{\Sigma}_1 \quad (24)$$

Define a modified parameter α as follows

$$\int \mathrm{St}(\mathbf{x}; \boldsymbol{\mu}_1, \boldsymbol{\Sigma}_1, \nu_1) \log\left[1 + \frac{1}{\nu_2}(\mathbf{x} - \mu)^{\mathrm{T}} \boldsymbol{\Sigma}^{-1}(\mathbf{x} - \mu)\right] d\mathbf{x}$$

$$\leq \alpha \log f(\boldsymbol{\mu}, \boldsymbol{\Sigma}) \qquad \text{s.t. } 0 < \alpha \leq 1 \quad (25)$$

where $\alpha \log f(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ represents the modified upper bound of the nonlinear integral. Note that the modified parameter α can be used to reduce the difference between the nonlinear integral and the modified upper bound.

Exploiting (25) in (21) yields

$$J(\boldsymbol{\mu}, \boldsymbol{\Sigma}) \leq 0.5 \log |\boldsymbol{\Sigma}| + 0.5\alpha(\nu_2 + n) \log f(\boldsymbol{\mu}, \boldsymbol{\Sigma}) + c_{\{\boldsymbol{\mu}, \boldsymbol{\Sigma}\}} \quad (26)$$

We propose to achieve the approximate mean vector $\boldsymbol{\mu}_2$ and scale matrix $\boldsymbol{\Sigma}_2$ by minimizing the upper bound of the cost function $J(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, i.e.,

$$\{\boldsymbol{\mu}_2, \boldsymbol{\Sigma}_2\} \approx \arg \min_{\{\boldsymbol{\mu}, \boldsymbol{\Sigma}\}} J_u(\boldsymbol{\mu}, \boldsymbol{\Sigma}) \quad (27)$$

where $J_u(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ is given by

$$J_u(\boldsymbol{\mu}, \boldsymbol{\Sigma}) = 0.5 \log |\boldsymbol{\Sigma}| + 0.5\alpha(\nu_2 + n) \log f(\boldsymbol{\mu}, \boldsymbol{\Sigma}) + c_{\{\boldsymbol{\mu}, \boldsymbol{\Sigma}\}} \quad (28)$$

The first-order derivatives of $J_u(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ with respect to $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ are formulated as

$$\frac{\partial J_u(\boldsymbol{\mu}, \boldsymbol{\Sigma})}{\partial \boldsymbol{\mu}} = \frac{0.5\alpha(\nu_2 + n)}{\nu_2 f(\boldsymbol{\mu}, \boldsymbol{\Sigma})} \boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu} - \boldsymbol{\mu}_1) \quad (29)$$

$$\frac{\partial J_u(\boldsymbol{\mu}, \boldsymbol{\Sigma})}{\partial \boldsymbol{\Sigma}} = 0.5 \boldsymbol{\Sigma}^{-1} - \frac{0.5\alpha(\nu_2 + n)}{\nu_2 f(\boldsymbol{\mu}, \boldsymbol{\Sigma})} \boldsymbol{\Sigma}^{-1} \tilde{\boldsymbol{\Sigma}}_1 \boldsymbol{\Sigma}^{-1} \quad (30)$$

where the cross-term of $(\boldsymbol{\mu} - \boldsymbol{\mu}_1)$ and $\boldsymbol{\Sigma}$ in (30) is omitted for brevity since it is always zero when (31) holds.

The optimal mean vector $\boldsymbol{\mu}_2$ and scale matrix $\boldsymbol{\Sigma}_2$ need to satisfy the following equations

$$\frac{\partial J_u(\boldsymbol{\mu}, \boldsymbol{\Sigma})}{\partial \boldsymbol{\mu}}\Big|_{\{\boldsymbol{\mu} = \boldsymbol{\mu}_2, \boldsymbol{\Sigma} = \boldsymbol{\Sigma}_2\}} = \mathbf{0} \quad (31)$$

$$\frac{\partial J_u(\boldsymbol{\mu}, \boldsymbol{\Sigma})}{\partial \boldsymbol{\Sigma}}\Big|_{\{\boldsymbol{\mu} = \boldsymbol{\mu}_2, \boldsymbol{\Sigma} = \boldsymbol{\Sigma}_2\}} = \mathbf{0} \quad (32)$$

Substituting (29)–(30) in (31)–(32) yields

$$\frac{0.5\alpha(\nu_2 + n)}{\nu_2 f(\boldsymbol{\mu}_2, \boldsymbol{\Sigma}_2)} \boldsymbol{\Sigma}_2^{-1}(\boldsymbol{\mu}_2 - \boldsymbol{\mu}_1) = \mathbf{0} \quad (33)$$

$$0.5 \boldsymbol{\Sigma}_2^{-1} - \frac{0.5\alpha(\nu_2 + n)}{\nu_2 f(\boldsymbol{\mu}_2, \boldsymbol{\Sigma}_2)} \boldsymbol{\Sigma}_2^{-1} \tilde{\boldsymbol{\Sigma}}_1 \boldsymbol{\Sigma}_2^{-1} = \mathbf{0} \quad (34)$$

Solving equation (33) gives

$$\boldsymbol{\mu}_2 = \boldsymbol{\mu}_1 \quad (35)$$

Employing (23) and (35) in (34) results in

$$\boldsymbol{\Sigma}_2 = \frac{\alpha(\nu_2 + n)}{\nu_2 + \mathrm{tr}(\tilde{\boldsymbol{\Sigma}}_1 \boldsymbol{\Sigma}_2^{-1})} \tilde{\boldsymbol{\Sigma}}_1 \quad (36)$$

Equations (31)–(32) are only necessary conditions for guaranteeing that the optimal solutions in (35)–(36) are minimum solutions of the cost function in (27). To be more rigorous, the detailed verifications are provided in Appendix A, and the theoretical results show that if the dof parameters ν_1 and ν_2 and the modified parameter α satisfy the conditions $\nu_1 > 2$ and $\alpha \geq \frac{1}{\nu_2 + n}$, then the optimal solutions in (35)–(36) will be the minimum solutions of the cost function in (27).

It is not difficult to find that $\boldsymbol{\Sigma}_2 = \tilde{\boldsymbol{\Sigma}}_1$ is an analytical solution of equation (36) when the modified parameter is set as $\alpha = 1$. Unfortunately, such selection of the modified parameter will introduce substantial approximation errors in the minimization problem (27). To improve the matching accuracy, the modified parameter needs to satisfy the condition $\frac{1}{\nu_2 + n} < \alpha < 1$, and the fixed-point iteration method is used to solve an approximate solution of equation (36). Specifically, the scale matrix $\boldsymbol{\Sigma}_2^{(i+1)}$ at the $i + 1$th iteration is achieved based on the scale matrix $\boldsymbol{\Sigma}_2^{(i)}$

and modified parameter $\alpha^{(i)}$ at the ith iteration using (36), i.e.,

$$\boldsymbol{\Sigma}_2^{(i+1)} = \frac{\alpha^{(i)}(\nu_2 + n)}{\nu_2 + \mathrm{tr}(\tilde{\boldsymbol{\Sigma}}_1(\boldsymbol{\Sigma}_2^{(i)})^{-1})} \tilde{\boldsymbol{\Sigma}}_1 \qquad (37)$$

Since the optimal Student's t-approximation is the true Student's t-PDF $\mathrm{St}(\mathbf{x}; \boldsymbol{\mu}_1, \boldsymbol{\Sigma}_1, \nu_1)$ when the dof parameters ν_1 and ν_2 are identical, i.e., $\mathrm{St}(\mathbf{x}; \boldsymbol{\mu}_2, \boldsymbol{\Sigma}_2, \nu_2) = \mathrm{St}(\mathbf{x}; \boldsymbol{\mu}_1, \boldsymbol{\Sigma}_1, \nu_1)$, then the scale matrix $\boldsymbol{\Sigma}_2^{(i+1)}$ is equal to the scale matrix $\boldsymbol{\Sigma}_1$ when the dof parameter ν_2 is chosen as ν_1. Thus, the modified parameter $\alpha^{(i)}$ needs to satisfy the following constraint equation, i.e.,

$$\boldsymbol{\Sigma}_1 = \frac{\alpha^{(i)}(\nu_1 + n)}{\nu_1 + \mathrm{tr}(\tilde{\boldsymbol{\Sigma}}_1(\boldsymbol{\Sigma}_2^{(i)})^{-1})} \tilde{\boldsymbol{\Sigma}}_1 \qquad (38)$$

Using (24) and (38), the modified parameter $\alpha^{(i)}$ is formulated as

$$\alpha^{(i)} = \frac{(\nu_1 - 2)\left[\nu_1 + \mathrm{tr}(\tilde{\boldsymbol{\Sigma}}_1(\boldsymbol{\Sigma}_2^{(i)})^{-1})\right]}{\nu_1(\nu_1 + n)} \qquad (39)$$

Substituting (39) in (37) and using (24) yields

$$\boldsymbol{\Sigma}_2^{(i+1)} = \frac{(\nu_2 + n)}{(\nu_1 + n)} \frac{\left[\nu_1 + \mathrm{tr}(\tilde{\boldsymbol{\Sigma}}_1(\boldsymbol{\Sigma}_2^{(i)})^{-1})\right]}{\left[\nu_2 + \mathrm{tr}(\tilde{\boldsymbol{\Sigma}}_1(\boldsymbol{\Sigma}_2^{(i)})^{-1})\right]} \boldsymbol{\Sigma}_1 \qquad (40)$$

For the proposed fixed-point iteration, the recursive relation between $\boldsymbol{\Sigma}_2^{(i+1)}$ and $\boldsymbol{\Sigma}_2^{(i)}$ is given in (40). To implement the fixed-point iteration, a reasonable initial scale matrix $\boldsymbol{\Sigma}_2^{(0)}$ needs to be first selected. In this paper, the initial scale matrix $\boldsymbol{\Sigma}_2^{(0)}$ is chosen by matching the second-order moment of Student's t-PDFs $\mathrm{St}(\mathbf{x}; \boldsymbol{\mu}_2, \boldsymbol{\Sigma}_2^{(0)}, \nu_2)$ and $\mathrm{St}(\mathbf{x}; \boldsymbol{\mu}_1, \boldsymbol{\Sigma}_1, \nu_1)$, i.e.,

$$\frac{\nu_2}{\nu_2 - 2} \boldsymbol{\Sigma}_2^{(0)} = \frac{\nu_1}{\nu_1 - 2} \boldsymbol{\Sigma}_1 \qquad (41)$$

Using (24) and (41), the initial scale matrix $\boldsymbol{\Sigma}_2^{(0)}$ is calculated as

$$\boldsymbol{\Sigma}_2^{(0)} = \frac{\nu_2 - 2}{\nu_2} \tilde{\boldsymbol{\Sigma}}_1 \qquad (42)$$

The KLD minimization-based Student's t-matching algorithm is illustrated in Table I, where N_m and ϵ denote the maximum number of iterations and the iterative threshold, respectively.

B. Convergence Analysis of the Proposed KLD Minimization-Based Student's t-Matching Method

Define the following auxiliary function

$$\mathbf{g}(\boldsymbol{\Sigma}) = \frac{(\nu_2 + n)}{(\nu_1 + n)} \frac{\left[\nu_1 + \mathrm{tr}(\tilde{\boldsymbol{\Sigma}}_1 \boldsymbol{\Sigma}^{-1})\right]}{\left[\nu_2 + \mathrm{tr}(\tilde{\boldsymbol{\Sigma}}_1 \boldsymbol{\Sigma}^{-1})\right]} \boldsymbol{\Sigma}_1 \qquad (43)$$

Using (40) and (43), we have the following recursive relation between $\boldsymbol{\Sigma}_2^{(i+1)}$ and $\boldsymbol{\Sigma}_2^{(i)}$

$$\boldsymbol{\Sigma}_2^{(i+1)} = \mathbf{g}(\boldsymbol{\Sigma}_2^{(i)}) \qquad (44)$$

TABLE I
THE PROPOSED KLD MINIMIZATION-BASED STUDENT'S
t-MATCHING ALGORITHM

Inputs: $\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_1, \nu_1, \nu_2, N_m, \epsilon$.
1. Calculate mean vector: $\boldsymbol{\mu}_2 = \boldsymbol{\mu}_1$.
2. Calculate $\tilde{\boldsymbol{\Sigma}}_1 = \frac{\nu_1}{\nu_1 - 2} \boldsymbol{\Sigma}_1$.
3. Initialization: $\boldsymbol{\Sigma}_2^{(0)} = \frac{\nu_2 - 2}{\nu_2} \tilde{\boldsymbol{\Sigma}}_1$.
for $i = 0 : N_m - 1$
4. Calculate scale matrix: $\boldsymbol{\Sigma}_2^{(i+1)} = \frac{(\nu_2 + n)}{(\nu_1 + n)} \left[\frac{\nu_1 + \mathrm{tr}(\tilde{\boldsymbol{\Sigma}}_1(\boldsymbol{\Sigma}_2^{(i)})^{-1})}{\nu_2 + \mathrm{tr}(\tilde{\boldsymbol{\Sigma}}_1(\boldsymbol{\Sigma}_2^{(i)})^{-1})}\right] \boldsymbol{\Sigma}_1$
5. If $\|\boldsymbol{\Sigma}_2^{(i+1)} - \boldsymbol{\Sigma}_2^{(i+1)}\| \le \epsilon$, stop iteration.
end for
6. $\boldsymbol{\Sigma}_2 \approx \boldsymbol{\Sigma}_2^{(i+1)}$.
Outputs: $\boldsymbol{\mu}_2, \boldsymbol{\Sigma}_2, \nu_2$.

Considering that both the covariance matrix $\tilde{\boldsymbol{\Sigma}}_1$ and the scale matrix $\boldsymbol{\Sigma}$ are positive define matrices, we have

$$\mathrm{tr}(\tilde{\boldsymbol{\Sigma}}_1 \boldsymbol{\Sigma}^{-1}) > 0 \qquad (45)$$

Using (45) and the inequality $\nu_1 > \nu_2$ yields

$$1 < \frac{\left[\nu_1 + \mathrm{tr}(\tilde{\boldsymbol{\Sigma}}_1 \boldsymbol{\Sigma}^{-1})\right]}{\left[\nu_2 + \mathrm{tr}(\tilde{\boldsymbol{\Sigma}}_1 \boldsymbol{\Sigma}^{-1})\right]} < \frac{\nu_1}{\nu_2} \qquad (46)$$

Substituting (46) in (43) results in

$$\frac{(\nu_2 + n)}{(\nu_1 + n)} \boldsymbol{\Sigma}_1 < \mathbf{g}(\boldsymbol{\Sigma}) < \frac{(\nu_2 + n)\nu_1}{(\nu_1 + n)\nu_2} \boldsymbol{\Sigma}_1 \qquad (47)$$

Utilizing (43), the derivative of $\mathbf{g}(\boldsymbol{\Sigma})$ with respect to $\boldsymbol{\Sigma}$ can be formulated as

$$\frac{\partial \mathbf{g}(\boldsymbol{\Sigma})}{\partial \boldsymbol{\Sigma}} = \frac{(\nu_2 + n)}{(\nu_1 + n)} \frac{(\nu_1 - \nu_2)\boldsymbol{\Sigma}^{-1}\tilde{\boldsymbol{\Sigma}}_1\boldsymbol{\Sigma}^{-1}}{\left[\nu_2 + \mathrm{tr}(\tilde{\boldsymbol{\Sigma}}_1\boldsymbol{\Sigma}^{-1})\right]^2} \otimes \boldsymbol{\Sigma}_1 \qquad (48)$$

where $\otimes$ denotes the Kronecker product operation.

Since $\nu_1 - \nu_2 > 0$ and $\boldsymbol{\Sigma}^{-1}\tilde{\boldsymbol{\Sigma}}_1\boldsymbol{\Sigma}^{-1}$ and $\boldsymbol{\Sigma}_1$ are positive definite matrices, $\frac{\partial \mathbf{g}(\boldsymbol{\Sigma})}{\partial \boldsymbol{\Sigma}}$ is also a positive definite matrix, i.e.,

$$\frac{\partial \mathbf{g}(\boldsymbol{\Sigma})}{\partial \boldsymbol{\Sigma}} > \mathbf{0} \qquad (49)$$

It is observed from (47) and (49) that the auxiliary function is monotonically increasing and has upper and low bounds. Then, the matrix sequence $\{\boldsymbol{\Sigma}_2^{(i)}\}$ has a unique limit, i.e.,

$$\lim_{i \to +\infty} \boldsymbol{\Sigma}_2^{(i)} = \boldsymbol{\Sigma}_2 \qquad (50)$$

According to (50), the proposed KLD minimization-based Student's t-matching algorithm has a local convergence, and it will converge to a global optimal solution when a reasonable initial value is selected.

C. Numerical Validation

The effectiveness and superiority of the proposed KLD minimization-based Student's t-matching method will be demonstrated by a numerical simulation. The mean value, scale parameter and dof parameter of the known Student's t-PDF $p(\mathbf{x})$
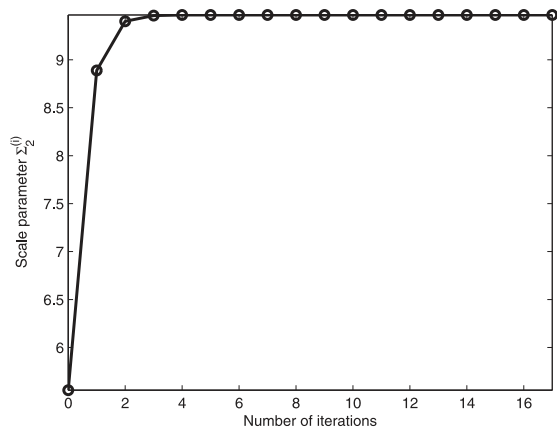
Fig. 1.   The estimate of scale parameter $\Sigma_2^{(i)}$ for different numbers of iterations.
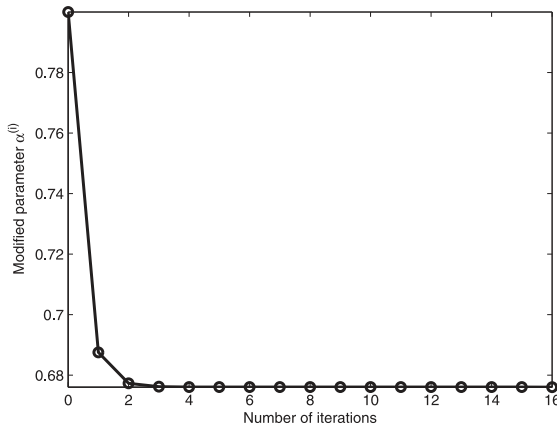

Fig. 2.   The modified parameter value $\alpha^{(i)}$ for different numbers of iterations.

are, respectively, set as $\mu_1 = 0, \Sigma_1 = 10$ and $\nu_1 = 5$, and the dof parameter of a Student's t-approximation is selected as $\nu_2 = 3$. According to the proposed KLD minimization-based Student's t-matching method, the optimal mean value of the Student's t-approximation is $\mu_2 = \mu_1 = 0$, and the scale parameter is iteratively calculated using the proposed algorithm in Table I, where the algorithm parameters are, respectively, selected as $N_m = 100$ and $\epsilon = 10^{-98}$.

Figs. 1 and 2 show the estimate of scale parameter $\Sigma_2^{(i)}$ and the estimate of modified parameter value $\alpha^{(i)}$ for different numbers of iterations, respectively. It can be observed from Fig. 1 that the estimate of scale parameter $\Sigma_2^{(i)}$ increases with the increase of the number of iterations, and the estimate of scale parameter $\Sigma_2^{(i)}$ converges to 9.468 when the number of iterations is greater than 4. We can observe from Fig. 2 that the modified parameter value $\alpha^{(i)}$ always satisfies the constraint $0 < \alpha^{(i)} < 1$ and decreases with the increase of the number of iterations, and the modified parameter value converges to 0.676 when the number of iterations is greater than 4.

The probability density curves of the true Student's t-distribution, the Student's t-approximation based on the moment matching method, and the Student's t-approximation based
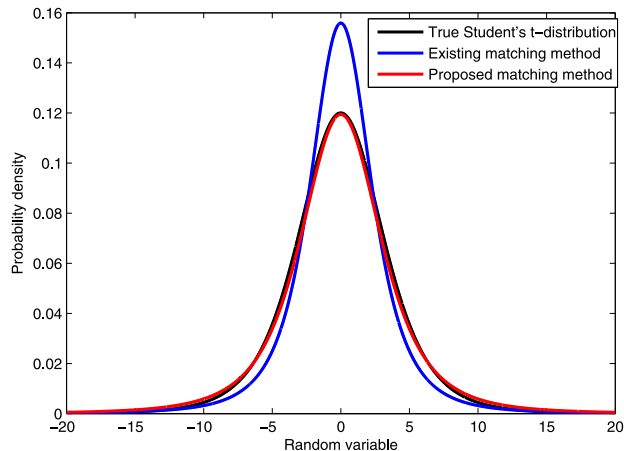

Fig. 3.   The probability density curves of the true Student's t-distribution and Student's t-approximations.

on the proposed KLD minimization-based Student's t-matching technique are shown in Fig. 3. It is observed from Fig. 3 that the Student's t-approximation based on the proposed KLD minimization-based Student's t-matching method is much closer to the true Student's t-distribution as compared with the Student's t-approximation based on the moment matching method. The KLD between the true Student's t-distribution and the Student's t-approximation based on the moment matching method is 0.029, and the KLD between the true Student's t-distribution and the Student's t-approximation based on the proposed KLD minimization-based Student's t-matching method is 0.011. As compared with the existing moment matching method, the improvement of the proposed method is 62.07%. Thus, the proposed KLD minimization-based matching method can fit the true Student's t-distribution much better than the existing moment matching method.

IV. A NOVEL KLD MINIMIZATION-BASED ASTF

In practical engineering applications, it is very difficult to select accurate scale matrices $\mathbf{Q}_k$ and $\mathbf{R}_k$ to model the non-Gaussian heavy-tailed state and observation noises in advance because such noises are often induced by unknown outliers, stochastic impulse interferences, and uncertain modelling errors. As a result, the STF exhibits poor filtering accuracy when inaccurate scale matrices are used to achieve recursive filtering estimates. Next, a novel KLD minimization-based ASTF is proposed to solve this problem.

A. KLD Minimization-Based Adaptive Estimation Scheme

To mitigate the effects of inaccurate scale matrices, the conventional idea is to jointly estimate $\mathbf{Q}_k$ and $\mathbf{R}_k$ using an adaptive method. Unfortunately, the scale matrix $\mathbf{Q}_k$ is very difficult to estimate directly when only one-step observation information is used. Building upon our previous work [1], [28], [29], [33], we propose to jointly estimate the state vector $\mathbf{x}_k$ together with the scale matrices $\mathbf{P}_{k|k-1}$ and $\mathbf{R}_k$, where the one-step prediction scale matrix $\mathbf{P}_{k|k-1}$ is derived from the state prediction equation

and state noise model, and treated as a random matrix. To this end, the joint posterior PDF $p(\mathbf{\Theta}_k|\mathbf{z}_{1:k})$ will be calculated based on the KLD minimization, where the parameter set $\mathbf{\Theta}_k$ is defined as follows

$$\mathbf{\Theta}_k \triangleq \{\mathbf{x}_k, \mathbf{P}_{k|k-1}, \mathbf{R}_k\} \tag{51}$$

The true joint posterior PDF $p(\mathbf{\Theta}_k|\mathbf{z}_{1:k})$ is approximated by a posterior PDF with free factored form $q(\mathbf{x}_k)\,q(\mathbf{P}_{k|k-1})q(\mathbf{R}_k)$, i.e.,

$$p(\mathbf{\Theta}_k|\mathbf{z}_{1:k}) \approx q(\mathbf{x}_k)q(\mathbf{P}_{k|k-1})q(\mathbf{R}_k) \tag{52}$$

and the approximate posterior PDFs $q(\mathbf{x}_k)$, $q(\mathbf{P}_{k|k-1})$ and $q(\mathbf{R}_k)$ are given by the KLD minimization as follows

$$\begin{aligned}&\{q(\mathbf{x}_k), q(\mathbf{P}_{k|k-1}), q(\mathbf{R}_k)\}\\&= \arg\min \mathrm{KLD}\left(q(\mathbf{x}_k)q(\mathbf{P}_{k|k-1})q(\mathbf{R}_k)||p(\mathbf{\Theta}_k|\mathbf{z}_{1:k})\right)\end{aligned} \tag{53}$$

Similar to the standard variational Bayesian approach, the KLD in (53) can be formulated as [34], [35]

$$\begin{aligned}&\mathrm{KLD}\left(q(\mathbf{x}_k)q(\mathbf{P}_{k|k-1})q(\mathbf{R}_k)||p(\mathbf{\Theta}_k|\mathbf{z}_{1:k})\right)\\&= \log p(\mathbf{z}_{1:k}) - L(q(\mathbf{x}_k), q(\mathbf{P}_{k|k-1}), q(\mathbf{R}_k))\end{aligned} \tag{54}$$

where $L(q(\mathbf{x}_k), q(\mathbf{P}_{k|k-1}), q(\mathbf{R}_k))$ denotes the evidence lower bound given by

$$\begin{aligned}L(q(\mathbf{x}_k), q(\mathbf{P}_{k|k-1}), q(\mathbf{R}_k)) &= \int q(\mathbf{x}_k)q(\mathbf{P}_{k|k-1})q(\mathbf{R}_k)\\&\times \log p(\mathbf{\Theta}_k, \mathbf{z}_{1:k})d\mathbf{x}_k d\mathbf{P}_{k|k-1}d\mathbf{R}_k\\&- \int q(\mathbf{x}_k)\log q(\mathbf{x}_k)d\mathbf{x}_k\\&- \int q(\mathbf{P}_{k|k-1})\log q(\mathbf{P}_{k|k-1})d\mathbf{P}_{k|k-1}\\&- \int q(\mathbf{R}_k)\log q(\mathbf{R}_k)d\mathbf{R}_k\end{aligned} \tag{55}$$

Considering that the log-likelihood function $\log p(\mathbf{z}_{1:k})$ is independent of the approximate posterior PDFs and using (54), the KLD minimization problem can be transformed into the evidence lower bound maximization problem as follows

$$\begin{aligned}&\{q(\mathbf{x}_k), q(\mathbf{P}_{k|k-1}), q(\mathbf{R}_k)\}\\&= \arg\max L(q(\mathbf{x}_k), q(\mathbf{P}_{k|k-1}), q(\mathbf{R}_k))\end{aligned} \tag{56}$$

Unfortunately, there are not conjugate prior distributions for the scale matrices of Student's *t*-distributions since the Student's *t*-distribution doesn't belong to an exponential family. As a result, the standard variational Bayesian approach cannot be employed to achieve the evidence lower bound maximization formulated in (55)–(56). To cope with this problem, we propose to achieve approximate posterior PDFs by maximizing the lower bound of $L(q(\mathbf{x}_k), q(\mathbf{P}_{k|k-1}), q(\mathbf{R}_k))$.

B. KLD Minimization-Based ASTF

Using (2)–(3), the one-step prediction PDF $p(\mathbf{x}_k|\mathbf{z}_{1:k-1})$ and the likelihood PDF $p(\mathbf{z}_k|\mathbf{x}_k)$ can be written as follows

$$p(\mathbf{x}_k|\mathbf{z}_{1:k-1}, \mathbf{P}_{k|k-1}) = \mathrm{St}(\mathbf{x}_k; \hat{\mathbf{x}}_{k|k-1}, \mathbf{P}_{k|k-1}, \nu) \tag{57}$$

$$p(\mathbf{z}_k|\mathbf{x}_k, \mathbf{R}_k) = \mathrm{St}(\mathbf{z}_k; \mathbf{H}_k\mathbf{x}_k, \mathbf{R}_k, \nu) \tag{58}$$

where the one-step predicted state $\hat{\mathbf{x}}_{k|k-1}$ is given in (4).

The prior distributions of the inaccurate scale matrices $\mathbf{P}_{k|k-1}$ and $\mathbf{R}_k$ are chosen as the inverse Wishart distributions, respectively, as follows

$$p(\mathbf{P}_{k|k-1}|\mathbf{z}_{1:k-1}) = \mathrm{IW}(\mathbf{P}_{k|k-1}; \hat{t}_{k|k-1}, \hat{\mathbf{T}}_{k|k-1}) \tag{59}$$

$$p(\mathbf{R}_k|\mathbf{z}_{1:k-1}) = \mathrm{IW}(\mathbf{R}_k; \hat{u}_{k|k-1}, \hat{\mathbf{U}}_{k|k-1}) \tag{60}$$

where $\mathrm{IW}(\cdot; \omega, \mathbf{\Sigma})$ denotes the inverse Wishart PDF with dof parameter ω and inverse scale matrix $\mathbf{\Sigma}$.

To fit the prior information of the scale matrices $\mathbf{P}_{k|k-1}$ and $\mathbf{R}_k$, the expectations of $\mathbf{P}_{k|k-1}^{-1}$ and $\mathbf{R}_k^{-1}$ are chosen as the inverses of their prior estimates, respectively, i.e.,

$$\begin{cases}\mathrm{E}[\mathbf{P}_{k|k-1}^{-1}] = \hat{t}_{k|k-1}\hat{\mathbf{T}}_{k|k-1}^{-1} = \tilde{\mathbf{P}}_{k|k-1}^{-1}\\\mathrm{E}[\mathbf{R}_k^{-1}] = \hat{u}_{k|k-1}\hat{\mathbf{U}}_{k|k-1}^{-1} = \tilde{\mathbf{R}}_k^{-1}\end{cases} \tag{61}$$

where $\tilde{\mathbf{P}}_{k|k-1}$ and $\tilde{\mathbf{R}}_k$ denote, respectively, the prior estimates of the scale matrices $\mathbf{P}_{k|k-1}$ and $\mathbf{R}_k$, and $\tilde{\mathbf{P}}_{k|k-1}$ is given by

$$\tilde{\mathbf{P}}_{k|k-1} = \mathbf{F}_{k-1}\mathbf{P}_{k-1|k-1}\mathbf{F}_{k-1}^T + \tilde{\mathbf{Q}}_k \tag{62}$$

where $\tilde{\mathbf{Q}}_k$ denotes the prior estimate of scale matrix $\mathbf{Q}_k$. Considering that in many practical contexts the state and observation models suffer only from external contaminations for a fraction of time, the prior estimates of scale matrices $\mathbf{Q}_k$ and $\mathbf{R}_k$ can be, respectively, set as the nominal state and observation noise covariance matrices, namely the covariance matrices of Gaussian state and observation noises without contamination.

Using (61) yields

$$\hat{\mathbf{T}}_{k|k-1} = \hat{t}_{k|k-1}\tilde{\mathbf{P}}_{k|k-1}, \qquad \hat{\mathbf{U}}_{k|k-1} = \hat{u}_{k|k-1}\tilde{\mathbf{R}}_k \tag{63}$$

1) Calculations of Posterior PDFs: We look for optimal posterior PDFs $q(\mathbf{P}_{k|k-1})$, $q(\mathbf{R}_k)$ and $q(\mathbf{x}_k)$ to maximize the lower bound of $L(q(\mathbf{x}_k), q(\mathbf{P}_{k|k-1}), q(\mathbf{R}_k))$. Next, the lower bound of $L(q(\mathbf{x}_k), q(\mathbf{P}_{k|k-1}), q(\mathbf{R}_k))$ will be achieved using Jensen's inequality, and the coordinate ascent approach will be employed to maximize the lower bound with respect to $q(\mathbf{P}_{k|k-1})$, $q(\mathbf{R}_k)$ and $q(\mathbf{x}_k)$.

Proposition 1: Using (55), Appendix B and Jensen's inequality, $q(\mathbf{P}_{k|k-1})$ is updated as an inverse Wishart PDF

$$q(\mathbf{P}_{k|k-1}) = \mathrm{IW}(\mathbf{P}_{k|k-1}; \hat{t}_{k|k}, \hat{\mathbf{T}}_{k|k}) \tag{64}$$

where the distribution parameters $\hat{t}_{k|k}$ and $\hat{\mathbf{T}}_{k|k}$ are formulated as

$$\hat{t}_{k|k} = \hat{t}_{k|k-1} + 1, \qquad \hat{\mathbf{T}}_{k|k} = \hat{\mathbf{T}}_{k|k-1} + \frac{\beta(\nu+n)}{\nu}\mathbf{A}_k \tag{65}$$

where β denotes a modified parameter, and $\mathbf{A}_k$ denotes the approximate one-step prediction error covariance matrix

given by

$$\mathbf{A}_k = \int (\mathbf{x}_k - \hat{\mathbf{x}}_{k|k-1})(\mathbf{x}_k - \hat{\mathbf{x}}_{k|k-1})^{\mathrm{T}} q(\mathbf{x}_k) d\mathbf{x}_k \qquad (66)$$

Proof: See Appendix C.   ∎

Proposition 2: Employing (55), Appendix B and Jensen's inequality, $q(\mathbf{R}_k)$ is updated as an inverse Wishart PDF

$$q(\mathbf{R}_k) = \mathrm{IW}(\mathbf{R}_k; \hat{u}_{k|k}, \hat{\mathbf{U}}_{k|k}) \qquad (67)$$

where the distribution parameters $\hat{u}_{k|k}$ and $\hat{\mathbf{U}}_{k|k}$ are formulated as

$$\hat{u}_{k|k} = \hat{u}_{k|k-1} + 1, \quad \hat{\mathbf{U}}_{k|k} = \hat{\mathbf{U}}_{k|k-1} + \frac{\gamma(\nu+m)}{\nu}\mathbf{B}_k \quad (68)$$

where γ denotes a modified parameter, and $\mathbf{B}_k$ denotes the approximate observation noise covariance matrix formulated as

$$\mathbf{B}_k = \int (\mathbf{z}_k - \mathbf{H}_k\mathbf{x}_k)(\mathbf{z}_k - \mathbf{H}_k\mathbf{x}_k)^{\mathrm{T}} q(\mathbf{x}_k) d\mathbf{x}_k \qquad (69)$$

Proof: See Appendix D.   ∎

Proposition 3: Exploiting (55), Appendix B and Jensen's inequality, $q(\mathbf{x}_k)$ is updated as a Student's t-PDF

$$q(\mathbf{x}_k) = \mathrm{St}(\mathbf{x}_k; \hat{\mathbf{x}}_{k|k}^*, \mathbf{P}_{k|k}^*, \nu^*) \qquad (70)$$

where the distribution parameters $\hat{\mathbf{x}}_{k|k}^*$, $\mathbf{P}_{k|k}^*$ and ν^* are formulated as

$$\mathbf{S}_k^* = \mathbf{H}_k\mathbf{P}_{k|k-1}^*\mathbf{H}_k^T + \mathbf{R}_k^* \qquad (71)$$

$$\Delta_k^* = \sqrt{(\mathbf{z}_k - \mathbf{H}_k\hat{\mathbf{x}}_{k|k-1})^T (\mathbf{S}_k^*)^{-1} (\mathbf{z}_k - \mathbf{H}_k\hat{\mathbf{x}}_{k|k-1})} \quad (72)$$

$$\mathbf{K}_k^* = \mathbf{P}_{k|k-1}^*\mathbf{H}_k^T (\mathbf{S}_k^*)^{-1} \qquad (73)$$

$$\hat{\mathbf{x}}_{k|k}^* = \hat{\mathbf{x}}_{k|k-1} + \mathbf{K}_k^*(\mathbf{z}_k - \mathbf{H}_k\hat{\mathbf{x}}_{k|k-1}) \qquad (74)$$

$$\mathbf{P}_{k|k}^* = \frac{\nu + (\Delta_k^*)^2}{\nu + m}(\mathbf{P}_{k|k-1}^* - \mathbf{K}_k^*\mathbf{H}_k\mathbf{P}_{k|k-1}^*) \qquad (75)$$

$$\nu^* = \nu + m \qquad (76)$$

where $\mathbf{P}_{k|k-1}^*$ and $\mathbf{R}_k^*$ denote the estimated scale matrices given by

$$\mathbf{P}_{k|k-1}^* = \left\{\mathrm{E}\left[\mathbf{P}_{k|k-1}^{-1}\right]\right\}^{-1}, \quad \mathbf{R}_k^* = \left\{\mathrm{E}\left[\mathbf{R}_k^{-1}\right]\right\}^{-1} \qquad (77)$$

Proof: See Appendix E.   ∎

2) Calculations of the Required Expectations: Employing (70), $\mathbf{A}_k$ and $\mathbf{B}_k$ in (66) and (69) are, respectively, calculated as

$$\mathbf{A}_k = \frac{\nu^*}{\nu^* - 2}\mathbf{P}_{k|k}^* + (\hat{\mathbf{x}}_{k|k}^* - \hat{\mathbf{x}}_{k|k-1})(\hat{\mathbf{x}}_{k|k}^* - \hat{\mathbf{x}}_{k|k-1})^{\mathrm{T}} \quad (78)$$

$$\mathbf{B}_k = \frac{\nu^*}{\nu^* - 2}\mathbf{H}_k\mathbf{P}_{k|k}^*\mathbf{H}_k^{\mathrm{T}} + (\mathbf{z}_k - \mathbf{H}_k\hat{\mathbf{x}}_{k|k}^*)(\mathbf{z}_k - \mathbf{H}_k\hat{\mathbf{x}}_{k|k}^*)^{\mathrm{T}}$$
$$(79)$$

Since $q(\mathbf{P}_{k|k-1})$ and $q(\mathbf{R}_k)$ are inverse Wishart PDFs as formulated in (64) and (67), $q(\mathbf{P}_{k|k-1}^{-1})$ and $q(\mathbf{R}_k^{-1})$ are Wishart PDFs given by

$$\begin{cases} q(\mathbf{P}_{k|k-1}^{-1}) = \mathrm{W}(\mathbf{P}_{k|k-1}^{-1}; \hat{t}_{k|k}, \hat{\mathbf{T}}_{k|k}^{-1}) \\ q(\mathbf{R}_k^{-1}) = \mathrm{W}(\mathbf{R}_k^{-1}; \hat{u}_{k|k}, \hat{\mathbf{U}}_{k|k}^{-1}) \end{cases} \qquad (80)$$

Using (80), $\mathrm{E}[\mathbf{P}_{k|k-1}^{-1}]$ and $\mathrm{E}[\mathbf{R}_k^{-1}]$ are calculated as

$$\mathrm{E}\left[\mathbf{P}_{k|k-1}^{-1}\right] = \hat{t}_{k|k}\hat{\mathbf{T}}_{k|k}^{-1}, \qquad \mathrm{E}\left[\mathbf{R}_k^{-1}\right] = \hat{u}_{k|k}\hat{\mathbf{U}}_{k|k}^{-1} \qquad (81)$$

Substituting (81) in (77) yields

$$\mathbf{P}_{k|k-1}^* = \hat{\mathbf{T}}_{k|k}/\hat{t}_{k|k}, \qquad \mathbf{R}_k^* = \hat{\mathbf{U}}_{k|k}/\hat{u}_{k|k} \qquad (82)$$

3) Selections of Modified Parameters β and γ: Substituting (63), (65) and (68) in (82) results in

$$\mathbf{P}_{k|k-1}^* = \frac{\hat{t}_{k|k-1}}{\hat{t}_{k|k-1}+1}\tilde{\mathbf{P}}_{k|k-1} + \frac{1}{\hat{t}_{k|k-1}+1}\frac{\beta(\nu+n)}{\nu}\mathbf{A}_k \quad (83)$$

$$\mathbf{R}_k^* = \frac{\hat{u}_{k|k-1}}{\hat{u}_{k|k-1}+1}\tilde{\mathbf{R}}_k + \frac{1}{\hat{u}_{k|k-1}+1}\frac{\gamma(\nu+m)}{\nu}\mathbf{B}_k \quad (84)$$

Let

$$\hat{t}_{k|k-1} = \frac{1}{\tau_k} - 1, \qquad \hat{u}_{k|k-1} = \frac{1}{\lambda_k} - 1 \qquad (85)$$

where $0 < \tau_k \leq 1$ and $0 < \lambda_k \leq 1$ denote the correction weights.

Substituting (85) in (83)–(84), we have

$$\mathbf{P}_{k|k-1}^* = (1-\tau_k)\tilde{\mathbf{P}}_{k|k-1} + \tau_k\frac{\beta(\nu+n)}{\nu}\mathbf{A}_k \qquad (86)$$

$$\mathbf{R}_k^* = (1-\lambda_k)\tilde{\mathbf{R}}_k + \lambda_k\frac{\gamma(\nu+m)}{\nu}\mathbf{B}_k \qquad (87)$$

Since $\mathbf{A}_k$ and $\mathbf{B}_k$ are, respectively, the approximate one-step prediction error covariance matrix and the approximate observation noise covariance matrix, $\frac{\nu^*-2}{\nu^*}\mathbf{A}_k$ and $\frac{\nu^*-2}{\nu^*}\mathbf{B}_k$ can be, respectively, deemed as the adaptive estimates of scale matrices $\mathbf{P}_{k|k-1}$ and $\mathbf{R}_k$. Normally, the estimate of the scale matrix is a weighted sum of prior estimate and adaptive estimate. Thus, the estimated scale matrices $\mathbf{P}_{k|k-1}^*$ and $\mathbf{R}_k^*$ are, respectively, the weighted sums of prior estimates $\tilde{\mathbf{P}}_{k|k-1}$ and $\tilde{\mathbf{R}}_k$ and adaptive estimates $\frac{\nu^*-2}{\nu^*}\mathbf{A}_k$ and $\frac{\nu^*-2}{\nu^*}\mathbf{B}_k$, i.e.,

$$\mathbf{P}_{k|k-1}^* = (1-\tau_k)\tilde{\mathbf{P}}_{k|k-1} + \tau_k\frac{\nu^*-2}{\nu^*}\mathbf{A}_k \qquad (88)$$

$$\mathbf{R}_k^* = (1-\lambda_k)\tilde{\mathbf{R}}_k + \lambda_k\frac{\nu^*-2}{\nu^*}\mathbf{B}_k \qquad (89)$$

By comparing (86)–(89), we have

$$\beta = \frac{\nu(\nu^*-2)}{(\nu+n)\nu^*}, \qquad \gamma = \frac{\nu(\nu^*-2)}{(\nu+m)\nu^*} \qquad (90)$$

We can observe from (90) that the modified parameters β and γ satisfy the constraints $0 < \beta < 1$ and $0 < \gamma < 1$. Thus, equations (88)–(89) can be employed to calculate the estimated scale matrices $\mathbf{P}_{k|k-1}^*$ and $\mathbf{R}_k^*$.

The calculations of posterior PDFs $q(\mathbf{x}_k)$, $q(\mathbf{P}_{k|k-1})$ and $q(\mathbf{R}_k)$ are interdependent and mutually coupled by the nonlinear equation. As a result, these posterior PDFs $q(\mathbf{x}_k)$, $q(\mathbf{P}_{k|k-1})$ and $q(\mathbf{R}_k)$ cannot be analytically achieved. In this paper, the fixed-point iteration method is used to achieve approximate solutions. That is to say, $q^{(i+1)}(\mathbf{P}_{k|k-1})$ and $q^{(i+1)}(\mathbf{R}_k)$ are firstly updated based on $q^{(i)}(\mathbf{x}_k)$, and then $q^{(i+1)}(\mathbf{x}_k)$ are updated based on

TABLE II
ONE TIME STEP OF THE PROPOSED ADAPTIVE STUDENT'S
t-FILTERING ALGORITHM

Inputs: $\hat{\mathbf{x}}_{k-1|k-1}$, $\mathbf{P}_{k-1|k-1}$, $\mathbf{F}_k$, $\mathbf{H}_k$, $\mathbf{z}_k$, $\tilde{\mathbf{Q}}_k$, $\tilde{\mathbf{R}}_k$, m, n, τ_k, λ_k, ν, N'_m, ϵ', N_m, ϵ.

Time update:

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

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

Measurement update:

3. Initialization: $\hat{\mathbf{x}}_{k|k}^{(0)} = \hat{\mathbf{x}}_{k|k-1}$, $\mathbf{P}_{k|k}^{(0)} = \tilde{\mathbf{P}}_{k|k-1}$, $\nu^{(0)} = \nu$.

for $i = 0 : N'_m - 1$

Update $q^{(i+1)}(\mathbf{P}_{k|k-1})$ based on $q^{(i)}(\mathbf{x}_k)$

4. $\mathbf{A}_k^{(i)} = \frac{\nu^{(i)}}{\nu^{(i)}-2}\mathbf{P}_{k|k}^{(i)} + (\hat{\mathbf{x}}_{k|k}^{(i)} - \hat{\mathbf{x}}_{k|k-1})(\hat{\mathbf{x}}_{k|k}^{(i)} - \hat{\mathbf{x}}_{k|k-1})^{\mathrm{T}}$

5. $\mathbf{P}_{k|k-1}^{(i+1)} = (1-\tau_k)\tilde{\mathbf{P}}_{k|k-1} + \tau_k\frac{\nu^{(i)}-2}{\nu^{(i)}}\mathbf{A}_k^{(i)}$

Update $q^{(i+1)}(\mathbf{R}_k)$ based on $q^{(i)}(\mathbf{x}_k)$

6. $\mathbf{B}_k^{(i)} = \frac{\nu^{(i)}}{\nu^{(i)}-2}\mathbf{H}_k\mathbf{P}_{k|k}^{(i)}\mathbf{H}_k^{\mathrm{T}} + (\mathbf{z}_k - \mathbf{H}_k\hat{\mathbf{x}}_{k|k}^{(i)})(\mathbf{z}_k - \mathbf{H}_k\hat{\mathbf{x}}_{k|k}^{(i)})^{\mathrm{T}}$

7. $\mathbf{R}_k^{(i+1)} = (1-\lambda_k)\tilde{\mathbf{R}}_k + \lambda_k\frac{\nu^{(i)}-2}{\nu^{(i)}}\mathbf{B}_k^{(i)}$

Update $q^{(i+1)}(\mathbf{x}_k)$ based on $q^{(i+1)}(\mathbf{P}_{k|k-1})$ and $q^{(i+1)}(\mathbf{R}_k)$

8. $\mathbf{S}_k^{(i+1)} = \mathbf{H}_k\mathbf{P}_{k|k-1}^{(i+1)}\mathbf{H}_k^T + \mathbf{R}_k^{(i+1)}$

9. $\Delta_k^{(i+1)} = \sqrt{(\mathbf{z}_k - \mathbf{H}_k\hat{\mathbf{x}}_{k|k-1})^T \left(\mathbf{S}_k^{(i+1)}\right)^{-1}(\mathbf{z}_k - \mathbf{H}_k\hat{\mathbf{x}}_{k|k-1})}$

10. $\mathbf{K}_k^{(i+1)} = \mathbf{P}_{k|k-1}^{(i+1)}\mathbf{H}_k^T\left(\mathbf{S}_k^{(i+1)}\right)^{-1}$

11. $\hat{\mathbf{x}}_{k|k}^{(i+1)} = \hat{\mathbf{x}}_{k|k-1} + \mathbf{K}_k^{(i+1)}(\mathbf{z}_k - \mathbf{H}_k\hat{\mathbf{x}}_{k|k-1})$

12. $\mathbf{P}_{k|k}^{(i+1)} = \frac{\nu + \left(\Delta_k^{(i+1)}\right)^2}{\nu+m}(\mathbf{P}_{k|k-1}^{(i+1)} - \mathbf{K}_k^{(i+1)}\mathbf{H}_k\mathbf{P}_{k|k-1}^{(i+1)})$

13. $\nu^{(i+1)} = \nu + m$

14. If $\frac{\|\hat{\mathbf{x}}_{k|k}^{(i+1)} - \hat{\mathbf{x}}_{k|k}^{(i)}\|}{\|\hat{\mathbf{x}}_{k|k}^{(i)}\|} \leq \epsilon'$, stop iteration.

end for

15. $\hat{\mathbf{x}}_{k|k}^* = \hat{\mathbf{x}}_{k|k}^{(i+1)}$, $\mathbf{P}_{k|k}^* = \mathbf{P}_{k|k}^{(i+1)}$, $\nu^* = \nu^{(i+1)}$.

16. $(\hat{\mathbf{x}}_{k|k}, \mathbf{P}_{k|k}, \nu) = \text{KLDMSTM}\left(\hat{\mathbf{x}}_{k|k}^*, \mathbf{P}_{k|k}^*, \nu^*, \nu, N_m, \epsilon\right)$.

Outputs: $\hat{\mathbf{x}}_{k|k}$, $\mathbf{P}_{k|k}$, ν.

$q^{(i+1)}(\mathbf{P}_{k|k-1})$ and $q^{(i+1)}(\mathbf{R}_k)$. Such fixed-point iterations are continued until convergence.

The proposed adaptive Student's t-filtering algorithm is composed of two parts, including the calculations of posterior PDFs based on the KLD-minimization and the KLD minimization-based Student's t-matching. The detailed implementation of the proposed ASTF algorithm is shown in Table II, where N'_m and ϵ' denote the maximum number of iterations and the iterative threshold, respectively, and the KLDMSTM denotes the proposed KLD minimization-based Student's t-matching algorithm in Section III.

Note that, as an extension, an adaptive nonlinear Student's t-filter and an adaptive nonlinear Student's t-smoother can be also derived for a nonlinear system with heavy-tailed state and observation noises based on the idea of the proposed adaptive

linear Student's t-filter. The existing Student's t-weighted integral rules [20], [36]–[40] can be used to implement the adaptive nonlinear Student's t-filter and the adaptive nonlinear Student's t-smoother.

V. SIMULATION STUDY

A. Simulation Description

In this section, a manoeuvring target tracking example is utilized to demonstrate the effectiveness and superiority of the proposed ASTF. The proposed ASTF is compared with the standard KF with true noise covariance matrices (KFTNCM), the Kalman filter with measurement validation gating (KF-G) [41], the existing HKF [15], the existing MCKF [16], the existing STF [19], and the existing RSTKF [28], where the KFTNCM employs the true noise covariance matrices to achieve the recursive filtering estimates. To better show the effectiveness and superiority, the proposed ASTF is also compared with the PF [4] and the optimal KF (OKF), where the PF uses the true noise distributions to obtain the optimal filtering estimates, and the OKF employs the instantaneous noise covariance matrices for all time points to achieve the optimal filtering estimates. Note that both the true noise distributions and the instantaneous noise covariance matrices are unavailable in practical applications with outlier interferences, and both the PF and the OKF are only used as optimal performance references. In this simulation, the two-sided acceptance region of the KF-G is set as 95%, and the tuning parameter of the existing HKF is selected as a common value of $\gamma = 1.345$ [15], and the kernel parameter of the existing MCKF is selected as $\sigma = 15$ to achieve a tradeoff between filtering accuracy and stability [16], the dof parameter of the existing STF is set as $\nu = 5$, and the tuning parameter and dof parameter of the existing RSTKF are, respectively, selected as $\tau = 4$ and $\nu = 5$, and 1000 particles are used in the PF. In the proposed ASTF, the correction weights are set as $\tau_k = \lambda_k = 0.15$, and the dof parameter is set as $\nu = 5$. Furthermore, to guarantee the convergence of the fixed-point iterations, the maximum number of iterations and the iterative threshold are, respectively, selected as $N_m = 100$ and $\epsilon = 10^{-98}$ in the proposed KLD minimization-based Student's t-matching algorithm, and the maximum number of iterations and the iterative threshold are, respectively, set as $N'_m = 50$ and $\epsilon' = 10^{-16}$ in the existing HKF, MCKF and RSTKF and the proposed ASTF. All filtering algorithms are coded with MATLAB and are executed on a computer with Intel Core i7-3770 CPU @ 3.40 GHz. The corresponding MATLAB codes of this paper have been made publicly available at https://www.researchgate.net/profile/Yulong_Huang3.

A problem of tracking an agile target is considered in this simulation. The horizontal positions and velocities are selected as state variables, and the real-time positions of the agile target are observed in clutter and used as measurements, based on which the linear SSM of the tracking problem can be formulated as (1). The resultant state transition matrix is $\mathbf{F}_k = \begin{bmatrix} \mathbf{I}_2 & T\mathbf{I}_2 \\ \mathbf{0} & \mathbf{I}_2 \end{bmatrix}$, and the observation matrix is $\mathbf{H}_k = \begin{bmatrix} \mathbf{I}_2 & \mathbf{0} \end{bmatrix}$, where $\mathbf{I}_2$ represents the 2-D identity matrix and the sampling interval is set as $T = 1$s. The outlier contaminated state and observation noises are

generated as follows [1]

$$\begin{cases} \mathbf{w}_k \sim \begin{cases} N(\mathbf{0}, \mathbf{Q}) & \text{w.p. } 0.95 \\ N(\mathbf{0}, U_1\mathbf{Q}) & \text{w.p. } 0.05 \end{cases} \\ \mathbf{v}_k \sim \begin{cases} N(\mathbf{0}, \mathbf{R}) & \text{w.p. } 0.95 \\ N(\mathbf{0}, U_2\mathbf{R}) & \text{w.p. } 0.05 \end{cases} \end{cases} \quad (91)$$

where $\mathbf{Q} = \begin{bmatrix} \frac{T^3}{3}\mathbf{I}_2 & \frac{T^2}{2}\mathbf{I}_2 \\ \frac{T^2}{2}\mathbf{I}_2 & T\mathbf{I}_2 \end{bmatrix}$ and $\mathbf{R} = 100\mathbf{I}_2$ are, respectively, the nominal state and observation noise covariance matrices. The conditions in (91) mean that the state and observation noise covariance matrices are, respectively, magnified by scalar factors U_1 and U_2 with a probability of 5%, where $U_1 \gg 1$ and $U_2 \gg 1$. Such a setup can simulate a scenario of state and observation outliers, which is often encountered in practical engineering applications, and the resultant state and observation noises have non-Gaussian heavy-tailed distributions.

In this simulation, the true initial state vector is $\mathbf{x}_0 = [0, 0, 10, 10]^T$, and the initial state estimation error covariance matrix is chosen as $\mathbf{P}_0 = \text{diag}([100, 100, 100, 100])$. To guarantee a fair comparison, 1000 Monte Carlo runs are performed, and in each Monte Carlo run, the initial state estimate is randomly selected from a Gaussian distribution, i.e., $\hat{\mathbf{x}}_{0|0} \sim N(\mathbf{x}_0, \mathbf{P}_0)$, and the initial state estimates and corresponding initial estimation error covariance matrices of all filtering algorithms are set as $\hat{\mathbf{x}}_{0|0}$ and $\mathbf{P}_0$. The simulation time is set as 100s in each Monte Carlo run. To compare the filtering accuracy of the proposed ASTF and the existing filters, the root mean square errors (RMSEs) and averaged RMSEs (ARMSEs) of position and velocity are used as performance metrics and they are, respectively, abbreviated as RMSE_{pos}, RMSE_{vel}, ARMSE_{pos} and ARMSE_{vel}, whose definitions are given in the literature [28].

B. Simulation Comparisons

To better compare the proposed ASTF and the existing filters, we consider two explanatory cases: the moderately contaminated state and observation noises and the strongly contaminated state and observation noises.

1) Case 1: In this case, we consider the moderately contaminated state and observation noises, and the magnified factors in (91) are set as $U_1 = 100$ and $U_2 = 100$. Such a setup can simulate a scenario of moderately heavy-tailed state and observation noises. The RMSEs of position and velocity from the proposed ASTF and the existing noise-robust filters for case 1 are illustrated in Fig. 4–Fig. 5. The run time in a single step and the ARMSEs of position and velocity for case 1 are listed in Table III.

It is seen from Fig. 4–Fig. 5 and Table III that the proposed ASTF has smaller RMSEs and ARMSEs of position and velocity than the existing KFTNCM, KF-G, HKF, MCCKF and STF, and the proposed ASTF has similar RMSEs and ARMSEs of position and velocity with the existing PF. It can be also seen from Fig. 4–Fig. 5 and Table III that the proposed ASTF has slightly larger RMSEs and ARMSEs of position and velocity than the existing RSTKF, which is because the Gaussian approximation to the posterior PDF is better than
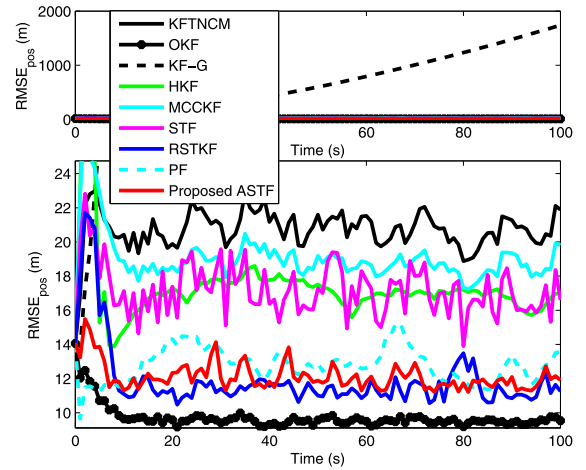

Fig. 4. RMSE_{pos} of the proposed ASTF and the existing noise-robust filters for case 1.
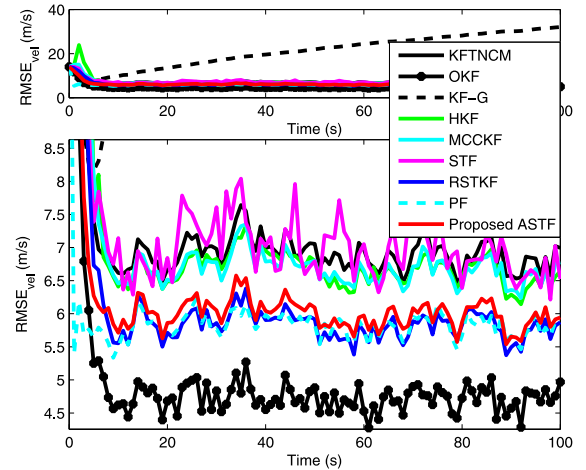

Fig. 5. RMSE_{vel} of the proposed ASTF and the existing noise-robust filters for case 1.

the Student's t-approximation for moderately heavy-tailed state and observation noises. Moreover, we can see from Table III that the proposed ASTF needs more implementation time than the existing KFTNCM, KF-G, HKF, MCKF, STF and RSTKF but significantly less implementation time than the existing PF. Thus, for moderately contaminated state and observation noises, the proposed ASTF has significantly better filtering accuracy than the existing KFTNCM, HKF, MCKF and STF and slightly worse filtering accuracy than the existing RSTKF and similar filtering accuracy with the existing PF with 1000 particles, and the proposed ASTF has higher computational burden than the existing KFTNCM, KF-G, HKF, MCKF, STF and RSTKF but significantly higher computational efficiency than the existing PF with 1000 particles.

2) Case 2: In this case, we consider the seriously contaminated state and observation noises, and the magnified factors in (91) are set as $U_1 = 1000$ and $U_2 = 1000$. Such a setup can simulate a scenario of strongly heavy-tailed state and observation noises. The RMSEs of position and velocity from the proposed

TABLE III
RUN TIME IN A SINGLE STEP AND ARMSES OF POSITION AND
VELOCITY FOR CASE 1

Filters	$\text{ARMSE}_{\text{pos}}$ (m)	$\text{ARMSE}_{\text{vel}}$ (m/s)	Time (ms)
KFTNCM	20.72	7.12	0.04
OKF	9.71	4.97	0.04
KF-G	689.45	21.18	0.15
HKF	17.29	7.17	0.72
MCKF	19.03	7.00	0.73
STF	17.18	7.23	0.06
RSTKF	11.92	6.13	3.26
PF	12.81	5.90	125.06
Proposed ASTF	12.12	6.20	3.88

TABLE IV
RUN TIME IN A SINGLE STEP AND ARMSES OF POSITION AND
VELOCITY FOR CASE 2

Filters	$\text{ARMSE}_{\text{pos}}$ (m)	$\text{ARMSE}_{\text{vel}}$ (m/s)	Time (ms)
KFTNCM	58.74	19.98	0.04
OKF	10.36	8.79	0.04
KF-G	2333.09	66.98	0.15
HKF	45.53	20.97	0.90
MCKF	39.42	19.85	0.86
STF	53.09	71.10	0.06
RSTKF	90.70	29.52	3.24
PF	43.54	15.51	125.06
Proposed ASTF	20.51	16.17	3.96

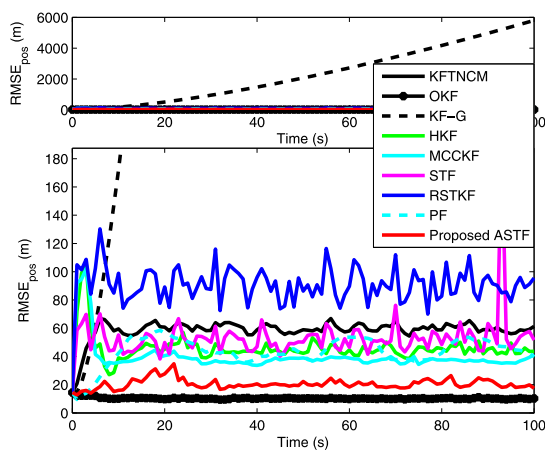
Fig. 6. RMSE_{pos} of the proposed ASTF and the existing noise-robust filters for case 2.

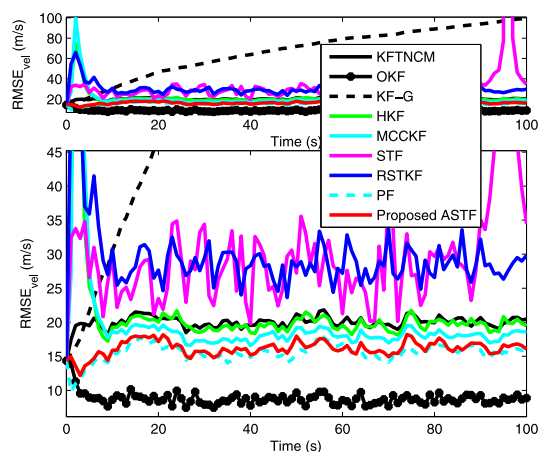Fig. 7. RMSE_{vel} of the proposed ASTF and the existing noise-robust filters for case 2.

ASTF and the existing noise-robust filters for case 2 are shown in Fig. 6–Fig. 7. The run time in a single step and the ARMSEs of position and velocity for case 2 are given in Table IV.

It is observed from Fig. 6–Fig. 7 and Table IV that the proposed ASTF has significantly smaller RMSEs and ARMSEs of position and velocity than the existing KFTNCM, KF-G,

HKF, MCKF, STF and RSTKF, and the proposed ASTF has significantly smaller RMSEs and ARMSEs of position but slightly larger RMSEs and ARMSEs of velocity than the existing PF. Note that, in the case of strongly heavy-tailed state and observation noises, the Student's t-approximation to the posterior PDF is better than the Gaussian approximation so the proposed ASTF has smaller RMSEs and ARMSEs of position and velocity than the existing RSTKF. We can also see from Table IV that the proposed ASTF needs greater implementation time than the existing filters except for the existing PF with 1000 particles. Thus, for strongly contaminated state and observation noises, the proposed ASTF has significantly better filtering accuracy but higher computational burden than the existing KFTNCM, KF-G, HKF, MCKF, STF and RSTKF, and the proposed ASTF has significantly better filtering accuracy of position and higher computational efficiency than the existing PF with 1000 particles.

In the cases 1 and 2, the KF-G is often found to diverge, which may be because some beneficial measurements are discarded based on misjudgments. In theory, the PF can achieve optimal filtering estimates when infinite particles are used. However, for the case of heavy-tailed non-Gaussian noises, the use of a limited number of particles is generally insufficient to capture the heavy tails of state and observation noises so the filtering accuracy degrades, especially for strongly heavy-tailed state and observation noises, as shown in Fig. 4–Fig. 7. Although the estimation accuracy of the PF can be further improved by increasing the number of particles, its computational complexity also increases dramatically.

VI. CONCLUSION

In this paper, a novel KLD minimization-based matching method was first proposed to improve the Student's t-matching accuracy, in which the upper bound of the KLD between the true Student's t-PDF and the approximate Student's t- PDF is minimized. To improve the Student's t-modelling accuracy, a novel KLD minimization-based adaptive method was then proposed to estimate the scale matrices of Student's t-distributions, in which the modified evidence lower bound is maximized. A novel KLD minimization-based ASTF was derived by combining the proposed Student's t-matching technique and an adaptive method. Simulation results showed that the proposed ASTF has

improved filtering accuracy but higher computational burden than the existing HKF, MCKF, STF and RSTKF for strongly heavy-tailed state and observation noises.

APPENDIX A
VERIFICATIONS OF MINIMUM SOLUTIONS (35)–(36)

In order to validate that (35)–(36) are minimum solutions, the joint Hessian matrix of the cost function $J_u(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ with respect to the mean vector $\boldsymbol{\mu}$ and the scale matrix $\boldsymbol{\Sigma}$ needs to be firstly calculated as follows

$$\mathbf{H}(\boldsymbol{\mu}, \boldsymbol{\Sigma}) = \begin{bmatrix} \frac{\partial^2 J_u(\boldsymbol{\mu}, \boldsymbol{\Sigma})}{\partial \boldsymbol{\mu} \partial \boldsymbol{\mu}^{\mathrm{T}}} & \frac{\partial^2 J_u(\boldsymbol{\mu}, \boldsymbol{\Sigma})}{\partial \boldsymbol{\mu} \partial \mathrm{vec}^{\mathrm{T}}(\boldsymbol{\Sigma})} \\ \left\{ \frac{\partial^2 J_u(\boldsymbol{\mu}, \boldsymbol{\Sigma})}{\partial \boldsymbol{\mu} \partial \mathrm{vec}^{\mathrm{T}}(\boldsymbol{\Sigma})} \right\}^{\mathrm{T}} & \frac{\partial^2 J_u(\boldsymbol{\mu}, \boldsymbol{\Sigma})}{\partial \mathrm{vec}(\boldsymbol{\Sigma}) \partial \mathrm{vec}^{\mathrm{T}}(\boldsymbol{\Sigma})} \end{bmatrix} \tag{92}$$

where $\mathrm{vec}(\cdot)$ represents the vectorization operation of a matrix.

Using (29)–(30), the sub-Hessian matrices of the cost function $J_u(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ with respect to the mean vector $\boldsymbol{\mu}$ and the scale matrix $\boldsymbol{\Sigma}$ are, respectively, calculated as

$$\frac{\partial^2 J_u(\boldsymbol{\mu}, \boldsymbol{\Sigma})}{\partial \boldsymbol{\mu} \partial \boldsymbol{\mu}^{\mathrm{T}}} = \frac{0.5\alpha(\nu_2 + n)\boldsymbol{\Sigma}^{-1}}{\nu_2 f(\boldsymbol{\mu}, \boldsymbol{\Sigma})} - \frac{0.5\alpha(\nu_2 + n)\boldsymbol{\Sigma}^{-1}}{\nu_2 [f(\boldsymbol{\mu}, \boldsymbol{\Sigma})]^2}$$
$$\times (\boldsymbol{\mu} - \boldsymbol{\mu}_1) \left[\frac{\partial f(\boldsymbol{\mu}, \boldsymbol{\Sigma})}{\partial \boldsymbol{\mu}} \right]^{\mathrm{T}} \tag{93}$$

$$\frac{\partial^2 J_u(\boldsymbol{\mu}, \boldsymbol{\Sigma})}{\partial \boldsymbol{\mu} \partial \mathrm{vec}^{\mathrm{T}}(\boldsymbol{\Sigma})} = \frac{0.25\alpha(\nu_2 + n)}{\nu_2 f(\boldsymbol{\mu}, \boldsymbol{\Sigma})} \left\{ \left[\boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu} - \boldsymbol{\mu}_1) \right]^{\mathrm{T}} \otimes \boldsymbol{\Sigma}^{-1} \right.$$
$$\left. + \boldsymbol{\Sigma}^{-1} \otimes \left[\boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu} - \boldsymbol{\mu}_1) \right]^{\mathrm{T}} \right\}$$
$$- \frac{0.5\alpha(\nu_2 + n)\boldsymbol{\Sigma}^{-1}}{\nu_2 [f(\boldsymbol{\mu}, \boldsymbol{\Sigma})]^2} (\boldsymbol{\mu} - \boldsymbol{\mu}_1) \frac{\partial f(\boldsymbol{\mu}, \boldsymbol{\Sigma})}{\partial \mathrm{vec}^{\mathrm{T}}(\boldsymbol{\Sigma})} \tag{94}$$

$$\frac{\partial^2 J_u(\boldsymbol{\mu}, \boldsymbol{\Sigma})}{\partial \mathrm{vec}(\boldsymbol{\Sigma}) \partial \mathrm{vec}^{\mathrm{T}}(\boldsymbol{\Sigma})} = -0.5\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{\Sigma}^{-1} + \frac{0.5\alpha(\nu_2 + n)}{\nu_2 f(\boldsymbol{\mu}, \boldsymbol{\Sigma})}$$
$$\times \left\{ \left[\boldsymbol{\Sigma}^{-1}\tilde{\boldsymbol{\Sigma}}_1\boldsymbol{\Sigma}^{-1} \right] \otimes \boldsymbol{\Sigma}^{-1} + \boldsymbol{\Sigma}^{-1} \otimes \left[\boldsymbol{\Sigma}^{-1}\tilde{\boldsymbol{\Sigma}}_1\boldsymbol{\Sigma}^{-1} \right] \right\}$$
$$+ \frac{0.25\alpha(\nu_2 + n)}{[\nu_2 f(\boldsymbol{\mu}, \boldsymbol{\Sigma})]^2} \left\{ \left[\boldsymbol{\Sigma}^{-1} \left[(\boldsymbol{\mu} - \boldsymbol{\mu}_1)(\boldsymbol{\mu} - \boldsymbol{\mu}_1)^{\mathrm{T}} + \tilde{\boldsymbol{\Sigma}}_1 \right] \boldsymbol{\Sigma}^{-1} \right] \right.$$
$$\otimes \left[\boldsymbol{\Sigma}^{-1}\tilde{\boldsymbol{\Sigma}}_1\boldsymbol{\Sigma}^{-1} \right] + \left[\boldsymbol{\Sigma}^{-1}\tilde{\boldsymbol{\Sigma}}_1\boldsymbol{\Sigma}^{-1} \right]$$
$$\left. \otimes \left[\boldsymbol{\Sigma}^{-1} \left[(\boldsymbol{\mu} - \boldsymbol{\mu}_1)(\boldsymbol{\mu} - \boldsymbol{\mu}_1)^{\mathrm{T}} + \tilde{\boldsymbol{\Sigma}}_1 \right] \boldsymbol{\Sigma}^{-1} \right] \right\} \tag{95}$$

where $\otimes$ denotes the Kronecker product operation.

Substituting $\boldsymbol{\mu} = \boldsymbol{\mu}_2$ and $\boldsymbol{\Sigma} = \boldsymbol{\Sigma}_2$ in (93)–(95) and using (35) yields

$$\left. \frac{\partial^2 J_u(\boldsymbol{\mu}, \boldsymbol{\Sigma})}{\partial \boldsymbol{\mu} \partial \boldsymbol{\mu}^{\mathrm{T}}} \right|_{\{\boldsymbol{\mu}=\boldsymbol{\mu}_2, \boldsymbol{\Sigma}=\boldsymbol{\Sigma}_2\}} = \frac{0.5\alpha(\nu_2 + n)\boldsymbol{\Sigma}_2^{-1}}{\nu_2 f(\boldsymbol{\mu}_2, \boldsymbol{\Sigma}_2)} \tag{96}$$

$$\left. \frac{\partial^2 J_u(\boldsymbol{\mu}, \boldsymbol{\Sigma})}{\partial \boldsymbol{\mu} \partial \mathrm{vec}^{\mathrm{T}}(\boldsymbol{\Sigma})} \right|_{\{\boldsymbol{\mu}=\boldsymbol{\mu}_2, \boldsymbol{\Sigma}=\boldsymbol{\Sigma}_2\}} = \mathbf{0} \tag{97}$$

$$\left. \frac{\partial^2 J_u(\boldsymbol{\mu}, \boldsymbol{\Sigma})}{\partial \mathrm{vec}(\boldsymbol{\Sigma}) \partial \mathrm{vec}^{\mathrm{T}}(\boldsymbol{\Sigma})} \right|_{\{\boldsymbol{\mu}=\boldsymbol{\mu}_2, \boldsymbol{\Sigma}=\boldsymbol{\Sigma}_2\}} = -0.5\boldsymbol{\Sigma}_2^{-1} \otimes \boldsymbol{\Sigma}_2^{-1}$$
$$+ \frac{0.5\alpha(\nu_2 + n)}{\nu_2 f(\boldsymbol{\mu}_2, \boldsymbol{\Sigma}_2)} \left\{ \left[\boldsymbol{\Sigma}_2^{-1}\tilde{\boldsymbol{\Sigma}}_1\boldsymbol{\Sigma}_2^{-1} \right] \otimes \boldsymbol{\Sigma}_2^{-1} + \boldsymbol{\Sigma}_2^{-1} \right.$$
$$\left. \otimes \left[\boldsymbol{\Sigma}_2^{-1}\tilde{\boldsymbol{\Sigma}}_1\boldsymbol{\Sigma}_2^{-1} \right] \right\} + \frac{0.5\alpha(\nu_2 + n)}{[\nu_2 f(\boldsymbol{\mu}_2, \boldsymbol{\Sigma}_2)]^2} \left[\boldsymbol{\Sigma}_2^{-1}\tilde{\boldsymbol{\Sigma}}_1\boldsymbol{\Sigma}_2^{-1} \right]$$
$$\otimes \left[\boldsymbol{\Sigma}_2^{-1}\tilde{\boldsymbol{\Sigma}}_1\boldsymbol{\Sigma}_2^{-1} \right] \tag{98}$$

Employing (23)–(24) and (35)–(36) in (96) and (98) results in

$$\left. \frac{\partial^2 J_u(\boldsymbol{\mu}, \boldsymbol{\Sigma})}{\partial \boldsymbol{\mu} \partial \boldsymbol{\mu}^{\mathrm{T}}} \right|_{\{\boldsymbol{\mu}=\boldsymbol{\mu}_2, \boldsymbol{\Sigma}=\boldsymbol{\Sigma}_2\}} = \frac{\nu_1 - 2}{2\nu_1} \boldsymbol{\Sigma}_1^{-1} \tag{99}$$

$$\left. \frac{\partial^2 J_u(\boldsymbol{\mu}, \boldsymbol{\Sigma})}{\partial \mathrm{vec}(\boldsymbol{\Sigma}) \partial \mathrm{vec}^{\mathrm{T}}(\boldsymbol{\Sigma})} \right|_{\{\boldsymbol{\mu}=\boldsymbol{\mu}_2, \boldsymbol{\Sigma}=\boldsymbol{\Sigma}_2\}} = 0.5 \left(1 - \frac{1}{\alpha(\nu_2 + n)} \right)$$
$$\times \boldsymbol{\Sigma}_2^{-1} \otimes \boldsymbol{\Sigma}_2^{-1} \tag{100}$$

Substituting (97) and (99)–(100) in (92), the value of Hessian matrix at $\boldsymbol{\mu} = \boldsymbol{\mu}_2$ and $\boldsymbol{\Sigma} = \boldsymbol{\Sigma}_2$ can be formulated as

$$\mathbf{H}(\boldsymbol{\mu}_2, \boldsymbol{\Sigma}_2) = \begin{bmatrix} \frac{\nu_1 - 2}{2\nu_1}\boldsymbol{\Sigma}_1^{-1} & \mathbf{0} \\ \mathbf{0} & 0.5(1 - \frac{1}{\alpha(\nu_2 + n)})\boldsymbol{\Sigma}_2^{-1} \otimes \boldsymbol{\Sigma}_2^{-1} \end{bmatrix} \tag{101}$$

It is observed from (101) that the Hessian matrix $\mathbf{H}(\boldsymbol{\mu}_2, \boldsymbol{\Sigma}_2)$ is positive semi-definite if the dof parameters ν_1 and ν_2 and the modified parameter α satisfy the following conditions

$$\nu_1 > 2, \quad \alpha \geq \frac{1}{\nu_2 + n} \tag{102}$$

and note that the dof parameter ν_1 can't be equal to 2, i.e., $\nu_1 \neq 2$, to guarantee that the covariance matrix $\tilde{\boldsymbol{\Sigma}}_1$ in (24) exists.

Thus, if the dof parameters ν_1 and ν_2 and the modified parameter α satisfy the conditions in (102), then the optimal solutions in (35)–(36) will be the minimum solutions of the cost function in (27).

APPENDIX B
CALCULATION OF $\log p(\boldsymbol{\Theta}_k, \mathbf{z}_{1:k})$

Employing (57)–(60), $\log p(\boldsymbol{\Theta}_k, \mathbf{z}_{1:k})$ can be formulated as

$$\log p(\boldsymbol{\Theta}_k, \mathbf{z}_{1:k}) = -0.5(\hat{t}_{k|k-1} + n + 2) \log |\mathbf{P}_{k|k-1}|$$
$$- 0.5(\nu + n) \log \left[1 + \frac{1}{\nu}(\mathbf{x}_k - \hat{\mathbf{x}}_{k|k-1})^{\mathrm{T}} \mathbf{P}_{k|k-1}^{-1} (\mathbf{x}_k - \hat{\mathbf{x}}_{k|k-1}) \right]$$
$$- 0.5\mathrm{tr}\left(\hat{\mathbf{T}}_{k|k-1} \mathbf{P}_{k|k-1}^{-1} \right) - 0.5(\hat{u}_{k|k-1} + m + 2) \log |\mathbf{R}_k|$$
$$- 0.5(\nu + m) \log \left[1 + \frac{1}{\nu}(\mathbf{z}_k - \mathbf{H}_k\mathbf{x}_k)^{\mathrm{T}} \mathbf{R}_k^{-1} (\mathbf{z}_k - \mathbf{H}_k\mathbf{x}_k) \right]$$
$$\times 0.5\mathrm{tr}\left(\hat{\mathbf{U}}_{k|k-1} \mathbf{R}_k^{-1} \right) + c_{\boldsymbol{\Theta}_k} \tag{103}$$

APPENDIX C
PROOF OF PROPOSITION 1

Substituting (103) in (55) and treating $q(\mathbf{x}_k)$ and $q(\mathbf{R}_k)$ as known posterior PDFs yields

$$L_1(q(\mathbf{P}_{k|k-1})) = -\int q(\mathbf{P}_{k|k-1}) \left[0.5(\hat{t}_{k|k-1} + n + 2) \right.$$
$$\times \log|\mathbf{P}_{k|k-1}| + 0.5\mathrm{tr}\left(\hat{\mathbf{T}}_{k|k-1} \mathbf{P}_{k|k-1}^{-1} \right) + 0.5(\nu + n)$$
$$\times Y_1(\mathbf{P}_{k|k-1}) \left. \right] d\mathbf{P}_{k|k-1} - \int q(\mathbf{P}_{k|k-1}) \log q(\mathbf{P}_{k|k-1}) d\mathbf{P}_{k|k-1}$$
$$+ c_{\mathbf{P}_{k|k-1}} \tag{104}$$

where $Y_1(\mathbf{P}_{k|k-1})$ is given by

$$Y_1(\mathbf{P}_{k|k-1}) = \int q(\mathbf{x}_k) \log \left[1 + \frac{1}{\nu}(\mathbf{x}_k - \hat{\mathbf{x}}_{k|k-1})^{\mathrm{T}} \mathbf{P}_{k|k-1}^{-1} \right.$$
$$\left. \times (\mathbf{x}_k - \hat{\mathbf{x}}_{k|k-1}) \right] d\mathbf{x}_k \tag{105}$$

Considering that the natural logarithmic function $\log(\cdot)$ is a convex function and using Jensen's inequality, we have [32]

$$Y_1(\mathbf{P}_{k|k-1}) \leq \log U_1(\mathbf{P}_{k|k-1}) \tag{106}$$

where $U_1(\mathbf{P}_{k|k-1})$ is given by

$$U_1(\mathbf{P}_{k|k-1}) = \int q(\mathbf{x}_k) \left[1 + \frac{1}{\nu}(\mathbf{x}_k - \hat{\mathbf{x}}_{k|k-1})^{\mathrm{T}} \mathbf{P}_{k|k-1}^{-1} \right.$$
$$\left. \times (\mathbf{x}_k - \hat{\mathbf{x}}_{k|k-1}) \right] d\mathbf{x}_k \tag{107}$$

Exploiting (107), $U_1(\mathbf{P}_{k|k-1})$ can be calculated as

$$U_1(\mathbf{P}_{k|k-1}) = 1 + \frac{1}{\nu}\mathrm{tr}\left(\mathbf{A}_k \mathbf{P}_{k|k-1}^{-1} \right) \tag{108}$$

where $\mathbf{A}_k$ is given by (66).

Since $\mathbf{A}_k$ and $\mathbf{P}_{k|k-1}$ are positive definite matrices, we have

$$\mathrm{tr}\left(\mathbf{A}_k \mathbf{P}_{k|k-1}^{-1} \right) > 0 \tag{109}$$

Considering that $\log(1 + t) < t$ for arbitrary $t > 0$ and using (108)–(109) results in

$$\log U_1(\mathbf{P}_{k|k-1}) = \log\left[1 + \frac{1}{\nu}\mathrm{tr}\left(\mathbf{A}_k \mathbf{P}_{k|k-1}^{-1} \right) \right]$$
$$< \frac{1}{\nu}\mathrm{tr}\left(\mathbf{A}_k \mathbf{P}_{k|k-1}^{-1} \right) \tag{110}$$

Define a modified parameter β as follows

$$\log U_1(\mathbf{P}_{k|k-1}) \leq \frac{\beta}{\nu}\mathrm{tr}\left(\mathbf{A}_k \mathbf{P}_{k|k-1}^{-1} \right) \quad \text{s.t. } 0 < \beta < 1 \tag{111}$$

where $\frac{\beta}{\nu}\mathrm{tr}(\mathbf{A}_k \mathbf{P}_{k|k-1}^{-1})$ denotes the modified upper bound of $\log U_1(\mathbf{P}_{k|k-1})$. The modified parameter β can be used to reduce the difference between $\log U_1(\mathbf{P}_{k|k-1})$ and the modified upper bound.

Substituting (106) and (111) in (104) yields

$$L_1(q(\mathbf{P}_{k|k-1})) \geq L_1'(q(\mathbf{P}_{k|k-1})) \tag{112}$$

where $L_1'(q(\mathbf{P}_{k|k-1}))$ is the lower bound of $L_1(q(\mathbf{P}_{k|k-1}))$ and is given by

$$L_1'(q(\mathbf{P}_{k|k-1})) = -\int q(\mathbf{P}_{k|k-1}) \left[0.5(\hat{t}_{k|k-1} + n + 2) \right.$$
$$\times \log|\mathbf{P}_{k|k-1}| + 0.5\mathrm{tr}\left\{ \left[\hat{\mathbf{T}}_{k|k-1} + \frac{\beta(\nu + n)}{\nu}\mathbf{A}_k \right] \mathbf{P}_{k|k-1}^{-1} \right\} \left. \right]$$
$$\times d\mathbf{P}_{k|k-1} - \int q(\mathbf{P}_{k|k-1}) \log q(\mathbf{P}_{k|k-1}) d\mathbf{P}_{k|k-1} + c_{\mathbf{P}_{k|k-1}} \tag{113}$$

We propose to achieve an approximate optimal $q(\mathbf{P}_{k|k-1})$ by maximizing the lower bound $L_1'(q(\mathbf{P}_{k|k-1}))$, i.e.,

$$q(\mathbf{P}_{k|k-1}) = \arg\max L_1'(q(\mathbf{P}_{k|k-1})) \tag{114}$$

Define an auxiliary PDF $\tilde{p}(\mathbf{P}_{k|k-1})$ as follows

$$\log \tilde{p}(\mathbf{P}_{k|k-1}) = -0.5(\hat{t}_{k|k-1} + n + 2)\log|\mathbf{P}_{k|k-1}|$$
$$- 0.5\mathrm{tr}\left\{ \left[\hat{\mathbf{T}}_{k|k-1} + \frac{\beta(\nu + n)}{\nu}\mathbf{A}_k \right] \mathbf{P}_{k|k-1}^{-1} \right\} + c_{\mathbf{P}_{k|k-1}} \tag{115}$$

Substituting (115) in (113) gives

$$L_1'(q(\mathbf{P}_{k|k-1})) = \int q(\mathbf{P}_{k|k-1}) \log \frac{\tilde{p}(\mathbf{P}_{k|k-1})}{q(\mathbf{P}_{k|k-1})}$$
$$= -\mathrm{KLD}(q(\mathbf{P}_{k|k-1})||\tilde{p}(\mathbf{P}_{k|k-1})) \leq 0 \tag{116}$$

where $L_1'(q(\mathbf{P}_{k|k-1}))$ achieves the maximum value 0 if and only if

$$q(\mathbf{P}_{k|k-1}) = \tilde{p}(\mathbf{P}_{k|k-1}) \tag{117}$$

Employing (117) in (115) yields

$$\log q(\mathbf{P}_{k|k-1}) = -0.5(\hat{t}_{k|k-1} + n + 2)\log|\mathbf{P}_{k|k-1}|$$
$$- 0.5\mathrm{tr}\left\{ \left[\hat{\mathbf{T}}_{k|k-1} + \frac{\beta(\nu + n)}{\nu}\mathbf{A}_k \right] \mathbf{P}_{k|k-1}^{-1} \right\} + c_{\mathbf{P}_{k|k-1}} \tag{118}$$

According to (118), we can obtain (64)–(65).

APPENDIX D
PROOF OF PROPOSITION 2

Substituting (103) in (55) and treating $q(\mathbf{x}_k)$ and $q(\mathbf{P}_{k|k-1})$ as known posterior PDFs results in

$$L_2(q(\mathbf{R}_k)) = -\int q(\mathbf{R}_k) \left[0.5(\hat{u}_{k|k-1} + m + 2)\log|\mathbf{R}_k| \right.$$
$$+ 0.5\mathrm{tr}\left(\hat{\mathbf{U}}_{k|k-1} \mathbf{R}_k^{-1} \right) + 0.5(\nu + m)Y_2(\mathbf{R}_k) \left. \right] d\mathbf{R}_k$$
$$- \int q(\mathbf{R}_k) \log q(\mathbf{R}_k) d\mathbf{R}_k + c_{\mathbf{R}_k} \tag{119}$$

where $Y_2(\mathbf{R}_k)$ is given by

$$Y_2(\mathbf{R}_k) = \int q(\mathbf{x}_k) \log \left[1 + \frac{1}{\nu}(\mathbf{z}_k - \mathbf{H}_k\mathbf{x}_k)^{\mathrm{T}} \mathbf{R}_k^{-1} \right.$$
$$\left. \times (\mathbf{z}_k - \mathbf{H}_k\mathbf{x}_k) \right] d\mathbf{x}_k \tag{120}$$

Considering that the natural logarithmic function $\log(\cdot)$ is a convex function and using Jensen's inequality yields [32]

$$Y_2(\mathbf{R}_k) \leq \log U_2(\mathbf{R}_k) \tag{121}$$

where $U_2(\mathbf{R}_k)$ is given by

$$U_2(\mathbf{R}_k) = \int q(\mathbf{x}_k) \left[1 + \frac{1}{\nu}(\mathbf{z}_k - \mathbf{H}_k\mathbf{x}_k)^{\mathrm{T}}\mathbf{R}_k^{-1} \right.$$
$$\left. \times (\mathbf{z}_k - \mathbf{H}_k\mathbf{x}_k) \right] d\mathbf{x}_k \tag{122}$$

Utilizing (122), $U_2(\mathbf{R}_k)$ is calculated as

$$U_2(\mathbf{R}_k) = 1 + \frac{1}{\nu}\mathrm{tr}\left(\mathbf{B}_k\mathbf{R}_k^{-1}\right) \tag{123}$$

where $\mathbf{B}_k$ is given by (69).

Considering that $\mathbf{B}_k$ and $\mathbf{R}_k$ are positive definite matrices, we obtain

$$\mathrm{tr}\left(\mathbf{B}_k\mathbf{R}_k^{-1}\right) > 0 \tag{124}$$

By the fact that $\log(1+t) < t$ for arbitrary $t > 0$ and employing (123)–(124) gives

$$\log U_2(\mathbf{R}_k) = \log\left[1 + \frac{1}{\nu}\mathrm{tr}\left(\mathbf{B}_k\mathbf{R}_k^{-1}\right)\right] < \frac{1}{\nu}\mathrm{tr}\left(\mathbf{B}_k\mathbf{R}_k^{-1}\right) \tag{125}$$

Define a modified parameter γ as follows

$$\log U_2(\mathbf{R}_k) \leq \frac{\gamma}{\nu}\mathrm{tr}\left(\mathbf{B}_k\mathbf{R}_k^{-1}\right) \qquad \text{s.t. } 0 < \gamma < 1 \tag{126}$$

where $\frac{\gamma}{\nu}\mathrm{tr}(\mathbf{B}_k\mathbf{R}_k^{-1})$ denotes the modified upper bound of $\log U_2(\mathbf{R}_k)$. The modified parameter γ can be used to reduce the difference between $\log U_2(\mathbf{R}_k)$ and the modified upper bound.

Substituting (121) and (126) in (119) results in

$$L_2(q(\mathbf{R}_k)) \geq L_2^{'}(q(\mathbf{R}_k)) \tag{127}$$

where $L_2^{'}(q(\mathbf{R}_k))$ represents the lower bound of $L_2(q(\mathbf{R}_k))$ and is given by

$$L_2^{'}(q(\mathbf{R}_k)) = -\int q(\mathbf{R}_k)\left[0.5(\hat{u}_{k|k-1} + m + 2)\log|\mathbf{R}_k|\right.$$
$$\left. + 0.5\mathrm{tr}\left\{\left[\hat{\mathbf{U}}_{k|k-1} + \frac{\gamma(\nu+m)}{\nu}\mathbf{B}_k\right]\mathbf{R}_k^{-1}\right\}\right]$$
$$\times d\mathbf{R}_k - \int q(\mathbf{R}_k)\log q(\mathbf{R}_k)d\mathbf{R}_k + c_{\mathbf{R}_k} \tag{128}$$

We propose to achieve an approximate optimal $q(\mathbf{R}_k)$ by maximizing the lower bound $L_2^{'}(q(\mathbf{R}_k))$, i.e.,

$$q(\mathbf{R}_k) = \arg\max L_2^{'}(q(\mathbf{R}_k)) \tag{129}$$

Define an auxiliary PDF $\tilde{p}(\mathbf{R}_k)$ as follows

$$\log\tilde{p}(\mathbf{R}_k) = -0.5(\hat{u}_{k|k-1} + m + 2)\log|\mathbf{R}_k|$$
$$- 0.5\mathrm{tr}\left\{\left[\hat{\mathbf{U}}_{k|k-1} + \frac{\gamma(\nu+m)}{\nu}\mathbf{B}_k\right]\mathbf{R}_k^{-1}\right\} + c_{\mathbf{R}_k} \tag{130}$$

Substituting (130) in (128) yields

$$L_2^{'}(q(\mathbf{R}_k)) = \int q(\mathbf{R}_k)\log\frac{\tilde{p}(\mathbf{R}_k)}{q(\mathbf{R}_k)}$$
$$= -\mathrm{KLD}\left(q(\mathbf{R}_k)||\tilde{p}(\mathbf{R}_k)\right) \leq 0 \tag{131}$$

where $L_2^{'}(q(\mathbf{R}_k))$ achieves the maximum 0 if and only if

$$q(\mathbf{R}_k) = \tilde{p}(\mathbf{R}_k) \tag{132}$$

Using (132) in (130) gives

$$\log q(\mathbf{R}_k) = -0.5(\hat{u}_{k|k-1} + m + 2)\log|\mathbf{R}_k|$$
$$-0.5\mathrm{tr}\left\{\left[\hat{\mathbf{U}}_{k|k-1} + \frac{\gamma(\nu+m)}{\nu}\mathbf{B}_k\right]\mathbf{R}_k^{-1}\right\} + c_{\mathbf{R}_k} \tag{133}$$

According to (133), we can obtain (67)–(68).

APPENDIX E
PROOF OF PROPOSITION 3

Substituting (103) in (55) and treating $q(\mathbf{P}_{k|k-1})$ and $q(\mathbf{R}_k)$ as known posterior PDFs results in

$$L_3(q(\mathbf{x}_k)) = -\int q(\mathbf{x}_k)\left[0.5(\nu+n)Y_3(\mathbf{x}_k) + 0.5(\nu+m)\right.$$
$$\left. \times Y_4(\mathbf{x}_k)\right]d\mathbf{x}_k - \int q(\mathbf{x}_k)\log q(\mathbf{x}_k)d\mathbf{x}_k + c_{\mathbf{x}_k} \tag{134}$$

where $Y_3(\mathbf{x}_k)$ and $Y_4(\mathbf{x}_k)$ are given by

$$Y_3(\mathbf{x}_k) = \int q(\mathbf{P}_{k|k-1})\log\left[1 + \frac{1}{\nu}(\mathbf{x}_k - \hat{\mathbf{x}}_{k|k-1})^{\mathrm{T}}\mathbf{P}_{k|k-1}^{-1}\right.$$
$$\left. \times (\mathbf{x}_k - \hat{\mathbf{x}}_{k|k-1})\right]d\mathbf{P}_{k|k-1} \tag{135}$$

$$Y_4(\mathbf{x}_k) = \int q(\mathbf{R}_k)\log\left[1 + \frac{1}{\nu}(\mathbf{z}_k - \mathbf{H}_k\mathbf{x}_k)^{\mathrm{T}}\mathbf{R}_k^{-1}\right.$$
$$\left. \times (\mathbf{z}_k - \mathbf{H}_k\mathbf{x}_k)\right]d\mathbf{R}_k \tag{136}$$

Since the natural logarithmic function $\log(\cdot)$ is a convex function and utilizing Jensen's inequality results in [32]

$$Y_3(\mathbf{x}_k) \leq \log\left[1 + \frac{1}{\nu}(\mathbf{x}_k - \hat{\mathbf{x}}_{k|k-1})^{\mathrm{T}}(\mathbf{P}_{k|k-1}^{*})^{-1}\right.$$
$$\left. \times (\mathbf{x}_k - \hat{\mathbf{x}}_{k|k-1})\right] \tag{137}$$

$$Y_4(\mathbf{x}_k) \leq \log\left[1 + \frac{1}{\nu}(\mathbf{z}_k - \mathbf{H}_k\mathbf{x}_k)^{\mathrm{T}}(\mathbf{R}_k^{*})^{-1}(\mathbf{z}_k - \mathbf{H}_k\mathbf{x}_k)\right] \tag{138}$$

where $\mathbf{P}_{k|k-1}^{*}$ and $\mathbf{R}_k^{*}$ are given by (77).

Substituting (137)–(138) in (134) yields

$$L_3(q(\mathbf{x}_k)) \geq L_3^{'}(q(\mathbf{x}_k)) \tag{139}$$

where $L'_3(q(\mathbf{x}_k))$ represents the lower bound of $L_3(q(\mathbf{x}_k))$ and is given by

$$
\begin{aligned}
L'_3(q(\mathbf{x}_k)) = &-\int q(\mathbf{x}_k)\left[0.5(\nu+n)\log\left[1+\frac{1}{\nu}(\mathbf{x}_k\right.\right.\\
&\left.-\hat{\mathbf{x}}_{k|k-1})^{\mathrm{T}}(\mathbf{P}^*_{k|k-1})^{-1}(\mathbf{x}_k-\hat{\mathbf{x}}_{k|k-1})\right]+0.5(\nu+m)\\
&\left.\times\log\left[1+\frac{1}{\nu}(\mathbf{z}_k-\mathbf{H}_k\mathbf{x}_k)^{\mathrm{T}}(\mathbf{R}^*_k)^{-1}(\mathbf{z}_k-\mathbf{H}_k\mathbf{x}_k)\right]\right]d\mathbf{x}_k\\
&-\int q(\mathbf{x}_k)\log q(\mathbf{x}_k)d\mathbf{x}_k+c_{\mathbf{x}_k}\quad(140)
\end{aligned}
$$

We propose to achieve an approximate optimal $q(\mathbf{x}_k)$ by maximizing the lower bound $L'_3(q(\mathbf{x}_k))$, i.e.,

$$
q(\mathbf{x}_k)=\arg\max L'_3(q(\mathbf{x}_k))\quad(141)
$$

Define an auxiliary PDF $\tilde{p}(\mathbf{x}_k)$ as follows

$$
\begin{aligned}
\log\tilde{p}(\mathbf{x}_k) = &-0.5(\nu+n)\log\left[1+\frac{1}{\nu}(\mathbf{x}_k-\hat{\mathbf{x}}_{k|k-1})^{\mathrm{T}}\right.\\
&\left.\times(\mathbf{P}^*_{k|k-1})^{-1}(\mathbf{x}_k-\hat{\mathbf{x}}_{k|k-1})\right]\\
&-0.5(\nu+m)\log\left[1+\frac{1}{\nu}(\mathbf{z}_k-\mathbf{H}_k\mathbf{x}_k)^{\mathrm{T}}\right.\\
&\left.\times(\mathbf{R}^*_k)^{-1}(\mathbf{z}_k-\mathbf{H}_k\mathbf{x}_k)\right]+c_{\mathbf{x}_k}\quad(142)
\end{aligned}
$$

Substituting (142) in (140) results in

$$
\begin{aligned}
L'_3(q(\mathbf{x}_k)) &= \int q(\mathbf{x}_k)\log\frac{\tilde{p}(\mathbf{x}_k)}{q(\mathbf{x}_k)}d\mathbf{x}_k\\
&= -\mathrm{KLD}\left(q(\mathbf{x}_k)||\tilde{p}(\mathbf{x}_k)\right)\leq 0\quad(143)
\end{aligned}
$$

where $L'_3(q(\mathbf{x}_k))$ achieves the maximum 0 if and only if

$$
q(\mathbf{x}_k)=\tilde{p}(\mathbf{x}_k)\quad(144)
$$

Employing (144) in (142) yields

$$
\begin{aligned}
\log q(\mathbf{x}_k) = &-0.5(\nu+n)\log\left[1+\frac{1}{\nu}(\mathbf{x}_k-\hat{\mathbf{x}}_{k|k-1})^{\mathrm{T}}\right.\\
&\left.\times(\mathbf{P}^*_{k|k-1})^{-1}(\mathbf{x}_k-\hat{\mathbf{x}}_{k|k-1})\right]\\
&-0.5(\nu+m)\log\left[1+\frac{1}{\nu}(\mathbf{z}_k-\mathbf{H}_k\mathbf{x}_k)^{\mathrm{T}}\right.\\
&\left.\times(\mathbf{R}^*_k)^{-1}(\mathbf{z}_k-\mathbf{H}_k\mathbf{x}_k)\right]+c_{\mathbf{x}_k}\quad(145)
\end{aligned}
$$

According to (145), we can obtain (70)–(76).

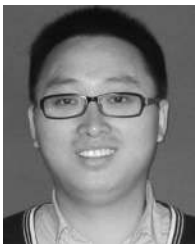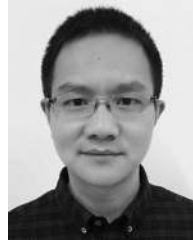## ACKNOWLEDGMENT

REFERENCES

[1] 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, Cybern. Syst.*, to be published, doi: 10.1109/TSMC.2017.2778269.

[2] S. Saha, "Noise robust online inference for linear dynamic systems," 2015, *arXiv:1504.05723.*

[3] Y. Huang, Y. Zhang, Y. Zhao, L. Mihaylova, and J. Chambers, "A novel robust Rauch-Tung-Striebel smoother based on slash and generalized hyperbolic skew Student's t-distributions," in *Proc. 21st Int. Conf. Inf. Fusion*, Jul. 2018, pp. 369–376.

[4] 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.

[5] N. Gordon and A. Smith, "Approximate non-Gaussian Bayesian estimation and modal consistency," *J. Roy. Statist. Soc.: Ser. B Methodol.*, vol. 55, no. 4, pp. 913–918, Apr. 1993.

[6] 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.

[7] J. Loxam and T. Drummond, "Student mixture filter for robust, realtime visual tracking," in *Proc. 10th Eur. Conf. Comput. Vision: Part III*, 2008, pp. 2–18.

[8] D. L. Alspach and H. Sorenson, "Non-linear Bayesian estimation using Gaussian sum approximation," *IEEE Trans. Autom. Control*, vol. 17, no. 4, pp. 439–448, Apr. 1972.

[9] I. Arasaratnam, S. Haykin, and R. Elliott, "Discrete-time nonlinear filtering algorithms using Gauss-Hermite quadrature," *Proc. IEEE*, vol. 95, no. 5, pp. 953–977, May 2007.

[10] I. Bilik and J. Tabrikian, "MMSE-based filtering in presence of non-Gaussian system and measurement noise," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 46, no. 3, pp. 1153–1170, Jul. 2010.

[11] P. J. Huber, *Robust Statistics* (International Encyclopedia of Statistical Science). Berlin, Germany: Springer, 2011.

[12] C. J. Masreliez and R. D. Martin, "Robust Bayesian estimation for the linear model and robustifying the Kalman filter," *IEEE Trans. Autom. Control*, vol. AC-22, no. 3, pp. 361–371, Jun. 1977.

[13] Z. M. Durovic and B. D. Kovachevic, "Robust estimation with unknown noise statistics," *IEEE Trans. Autom. Control*, vol. 44, no. 6, pp. 1292–1296, Jun. 1999.

[14] 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.

[15] C. D. Karlgaard and H. Schaub, "Huber-based divided difference filtering," *J. Guid., Control, Dyn.*, vol. 30, no. 3, pp. 885–891, 2007.

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

[17] X. Liu, B. Chen, B. Xu, Z. Wu, and P. Honeine, "Maximum correntropy unscented filter," *Int. J. Syst. Sci.*, vol. 48, no. 8, pp. 1607–1615, Aug. 2017.

[18] X. Liu, B. Chen, H. Zhao, J. Qin, and J. Cao, "Maximum correntropy Kalman filter with state constraints," *IEEE Access*, vol. 5, pp. 25846–25853, 2017.

[19] 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.

[20] 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.

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

[22] J. Ting, E. Theodorou, and S. Schaal, "Learning an outlier-robust Kalman filter," in *Proc. 18th Eur. Conf. Mach. Learn.*, Warsaw, Poland, 2007, pp. 748–756.

[23] R. Piché, S. Särkkä, and J. Hartikainen, "Recursive outlier-robust filtering and smoothing for nonlinear systems using the multivariate Student-t distribution," in *Proc. Int. Workshop Mach. Learn. Signal Process.*, Sep. 2012, pp. 1–6.

[24] 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.

[25] 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.

[26] 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.

[27] 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.

[28] 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.

[29] 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.

[30] Y. Huang, Y. Zhang, Y. Zhao, and J. Chambers, "A novel robust Gaussian-Student t mixture distribution based Kalman filter," *IEEE Trans. Signal Process.*, vol. 67, no. 13, pp. 3606–3620, Jul. 2019.

[31] C. M. Bishop, *Pattern Recognition and Machine Learning* (Information Science and Statistics). New York, NY, USA: Springer-Verlag, 2006.

[32] J. L. W. V. Jensen, "Sur les fonctions convexes et les inégalités entre les valeurs moyennes," *Acta Mathematica*, vol. 30, no. 1, pp. 175–193, Jan. 1906.

[33] 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.

[34] H. Nurminen, T. Ardeshiri, R. Piché, and F. Gustafsson, "Skew-t filter and smoother with improved covariance matrix approximation," *IEEE Trans. Signal Process.*, vol. 66, no. 21, pp. 5618–5633, Nov. 2018.

[35] D. Tzikas, A. Likas, and N. Galatsanos, "The variational approximation for Bayesian inference," *IEEE Signal Process. Mag.*, vol. 25, no. 6, pp. 131–146, Nov. 2008.

[36] Y. Huang, Y. Zhang, N. Li, S. M. Naqvi, and J. Chambers, "A robust Student's t based cubature filter," in *Proc. 19th Int. Conf. Inf. Fusion*, Jul. 2016, pp. 9–16.

[37] F. Tronarp, R. Hostettler, and S. Särkkä, "Sigma-point filtering for nonlinear systems with non-additive heavy-tailed noise," in *Proc. 19th Int. Conf. Inf. Fusion*, Jul. 2016, pp. 1859–1866.

[38] Y. Huang and Y. Zhang, "Robust Student's t-based stochastic cubature filter for nonlinear systems with heavy-tailed process and measurement noises," *IEEE Access*, vol. 5, pp. 7964–7974, 2017.

[39] O. Straka and J. Duník, "Stochastic integration Student's-t filter," in *Proc. 20th Int. Conf. Inf. Fusion*, Jul. 2017, pp. 1–8.

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

[41] Y. Bar-Shalom, X. Li, and T. Kirubarajan, *Estimation With Applications to Tracking and Navigation, Theory Algorithms and Software*. Hoboken, NJ, USA: Wiley, 2001.

Yonggang Zhang (S'06–M'07–SM'16) received the B.S. and M.S. degrees from the Department of Automation, Harbin Engineering University (HEU), Harbin, China, in 2002 and 2004, respectively, and the Ph.D. degree in electronic engineering from Cardiff University, Cardif, U.K., in 2007. He was a Postdoctoral Fellow with Loughborough University, U.K., from 2007 to 2008 in the area of adaptive signal processing. He is currently a Professor of Navigation, Guidance, and Control with HEU. 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.

Yulong Huang (M'19) received the B.S. degree in automation in 2012 and the Ph.D degree in control science and engineering in 2018 from the Department of Automation, Harbin Engineering University (HEU), Harbin, China. From November 2016 to November 2017, he was a Visiting Graduate Researcher with the Electrical Engineering Department, Columbia University, New York, NY, USA. He is currently an Associate Professor of Navigation, Guidance, and Control with HEU. From December 2019 to December 2021, he will be with the Department of Mechanical Engineering, City University of Hong Kong, Hong Kong, as a Hong Kong Scholar. His current research interests include signal processing, information fusion, and their applications in navigation technology, such as inertial navigation and integrated navigation.

Jonathon A. Chambers (S'83–M'90–SM'98–F'11) 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 currently a Professor of Signal and Information Processing and the Head of the School of Engineering, University of Leicester, Leicester, U.K. He is also an International Honorary Dean and a Guest Professor with Harbin Engineering University, Harbin, China, with support from the 1000 Talents Scheme. His research interests include adaptive signal processing and machine learning and their applications, and he has graduated 90 PhD researchers.

Dr. Chambers is a Fellow of the Royal Academy of Engineering, U.K., and the Institution of Electrical Engineers. He was the Technical Program Co-Chair for the 36th IEEE International Conference on Acoustics, Speech, and Signal Processing, ICASSP 2011, Prague, Czech Republic, and the Plenary Chair for ICASSP 2019, Brighton, U.K. He was on the IEEE Signal Processing Theory and Methods Technical Committee, the IEEE Signal Processing Society Awards and Conference Boards, and the Jack Kilby Medal Committee. He has was 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 and 2014. Since 2016, he has been an Area Editor for the Elsevier *Digital Signal Processing* journal.