Huang Y, Zhang Y, Wu Z, Li N, Chambers J.

A novel robust Student's t based Kalman filter.

IEEE Transactions on Aerospace and Electronic Systems 2017

DOI: http://dx.doi.org/10.1109/TAES.2017.2651684

A novel robust Student's t based Kalman filter

Yulong Huang, Yonggang Zhang, *Senior Member, IEEE,* Zhemin Wu, Ning Li, Jonathon Chambers, *Fellow, IEEE*

***Abstract*—A novel robust Student's t based Kalman filter is proposed by using the variational Bayesian approach, which provides a Gaussian approximation to the posterior distribution. Simulation results for a manoeuvring target tracking example illustrate that the proposed filter has smaller root mean square error and bias than existing filters.**

***Index Terms*—Kalman filtering, heay-tailed noises, Student's t distribution, Gaussian approximation, variational Bayesian**

I. INTRODUCTION

THE Kalman filter has been successfully applied in numerous practical applications, such as target tracking, navigation, positioning, control and signal processing due to its optimality, ease of implementation, and low computational complexity. The Kalman filter is optimal in terms of minimum mean square error for a linear state-space model with Gaussian process and measurement noises [1], [2]. However, in some engineering applications, such as tracking agile targets with measurement outliers from unreliable sensors, heavy-tailed non-Guassian process and measurement noises are present [3], [4], [5]. The performance of the conventional Kalman filter may break down in such engineering applications with heavy-tailed non-Gaussian process and measurement noises. A large number of robust filters have been proposed to solve the filtering problem of linear systems with heavy-tailed measurement noises, such as the Student's t mixture filter [6], outlier-robust Kalman filter [7], and Student's t and variational Bayesian (VB) based robust Kalman filters [8]–[11]. However, these robust filters may fail in the case of non-Gaussian heavy-tailed process noise since they all assume Gaussian process noise [3], [12].

To solve effectively and robustly the filtering problem of linear systems with both heavy-tailed process and measurement noises, a Huber-based Kalman filter (HKF) has been proposed by minimising a combined l_1 and l_2 norm [13]–[17]. The residual is bounded by using a Huber function, and the HKF can be deemed as a generalized maximum likelihood estimator [13]–[17]. However, the influence function of the HKF doesn't redescend, which results in limited estimation accuracy. To cope with heavy-tailed non-Gaussian noise induced by large outliers, many maximum correntropy criterion based Kalman filters (MCCKFs) have been proposed by maximising the correntropy of the predicted error and residual, and can be deemed as maximum a posterior estimators [18]–[22]. However, there is lack of theoretical basis to develop the estimation error covariance matrix, which degrades the estimation accuracy.

A reasonable approach to improve the estimation accuracy is better modelling the heavy-tailed non-Gaussian process and measurement noises. A linear Student's t based filter has been derived by assuming that both the process and measurement noises are Student's t distributions and the posterior probability density function (PDF) is approximated as a Student's t distribution to obtain a closed form solution for the filtering problem [3], [4]. However, this Student's t filter suffers from the following problems: (a) The derivation of the Student's t filter is based on an assumption that the Student's t PDFs of process and measurement noises have the same degrees of freedom (dof) parameter, which is seldom met in practical application because the process and measurement noises have different heavy-tailed characteristics, as shown in [3]; (b) The growth of the dof parameter must be prevented to preserve the heavy-tailed properties and closed Student's t-distributed form of the posterior PDF, which increases the bias of the state estimation, as discussed in [4]; (c) The Student's t approximation to the posterior PDF may be unreasonable, as will be illustrated in Section II. C.

In this paper, a new robust Student's t based Kalman filter is proposed to improve the estimation accuracy of the Student's t based filter with both heavy-tailed process and measurement noises. Firstly, to better model the linear systems with heavy-tailed process and measurement noises, the one-step predicted PDF and likelihood PDF are approximated as Student's t distributions with different dof parameters, and the PDF of the unknown scale matrix of the one-step predicted PDF is modelled as an inverse Wishart distribution. Secondly, motivated by the fact that the posterior PDF approaches to Gaussian in the case of moderate contaminated process and measurement noises, the posterior PDF is approximated as Gaussian, and an unknown scale matrix and auxiliary random variables are inferred based on the hierarchical Gaussian state-space model and a VB approach. Finally, the proposed robust Kalman filter and existing robust filters are applied to the problem of tracking an agile target that is observed in clutter. Simulation results show the proposed filter has smaller root mean square error (RMSE) and bias than existing robust filters.

This work was supported by the National Natural Science Foundation of China under Grant No. 61371173 and the Fundamental Research Funds for the Central Universities of Harbin Engineering University No. HEUCF160401 and the Engineering and Physical Sciences Research Council (EPSRC) of the UK grant no. EP/K014307/1.

Y. L. Huang, Y. G. Zhang and N. Li are with the Department of Automation, Harbin Engineering University, Harbin 150001, China (e-mail: heuedu@163.com; zhangyg@hrbeu.edu.cn; ningli@hrbeu.edu.cn).

Z. M. Wu is with the School of Electrical Engineering & Automation, Harbin Institute of Technology, Harbin 150080, China (e-mail: myemail-abc@163.com).

J. Chambers is with the School of Electrical and Electronic Engineering, Newcastle University, Newcastle upon Tyne, NE1 7RU, UK (e-mail: Jonathon.Chambers@newcastle.ac.uk).

II. Problem formulation

A. Review of the classical Kalman filter

Consider the following discrete-time linear stochastic system as shown by the state-space model

$$\mathbf{x}_k = \mathbf{F}_{k-1}\mathbf{x}_{k-1} + \mathbf{w}_{k-1} \qquad \text{(process equation)} \tag{1}$$

$$\mathbf{z}_k = \mathbf{H}_k\mathbf{x}_k + \mathbf{v}_k \qquad \text{(measurement equation)}, \tag{2}$$

where k is the discrete time index, $\mathbf{x}_k \in \mathbb{R}^n$ is the state vector, $\mathbf{z}_k \in \mathbb{R}^m$ is the measurement vector, $\mathbf{F}_k \in \mathbb{R}^{n\times n}$ is the state transition matrix, $\mathbf{H}_k \in \mathbb{R}^{m\times n}$ is the measurement matrix, $\mathbf{w}_k \in \mathbb{R}^n$ is the heavy-tailed process noise vector with zero mean and nominal covariance matrix $\mathbf{Q}_k$, and $\mathbf{v}_k \in \mathbb{R}^m$ is the heavy-tailed measurement noise vector with zero mean and nominal covariance matrix $\mathbf{R}_k$. Both $\mathbf{Q}_k$ and $\mathbf{R}_k$ are generally not accurate due to the existence of outliers. The initial state vector $\mathbf{x}_0$ is assumed to have a Gaussian distribution with mean vector $\hat{\mathbf{x}}_{0|0}$ and covariance matrix $\mathbf{P}_{0|0}$, i.e.

$$p(\mathbf{x}_0) = \mathrm{N}(\mathbf{x}_0; \hat{\mathbf{x}}_{0|0}, \mathbf{P}_{0|0}), \tag{3}$$

where $\mathrm{N}(\mathbf{x}; \mu, \mathbf{\Sigma})$ denotes the Gaussian PDF with mean vector μ and covariance matrix $\mathbf{\Sigma}$. Moreover, $\mathbf{x}_0$, $\mathbf{w}_k$ and $\mathbf{v}_k$ are assumed to be mutually uncorrelated in this work.

The Kalman filter is the most common method to infer the state vector $\mathbf{x}_k$ given the state-space model and measurements until time k. The recursive Kalman filter consists of a time update and a measurement update, which are given as follows:

Time update

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

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

Measurement update

$$\mathbf{K}_k = \mathbf{P}_{k|k-1}\mathbf{H}_k^T(\mathbf{H}_k\mathbf{P}_{k|k-1}\mathbf{H}_k^T + \mathbf{R}_k)^{-1} \tag{6}$$

$$\hat{\mathbf{x}}_{k|k} = \hat{\mathbf{x}}_{k|k-1} + \mathbf{K}_k(\mathbf{z}_k - \mathbf{H}_k\hat{\mathbf{x}}_{k|k-1}) \tag{7}$$

$$\mathbf{P}_{k|k} = \mathbf{P}_{k|k-1} - \mathbf{K}_k\mathbf{H}_k\mathbf{P}_{k|k-1}, \tag{8}$$

where $(\cdot)^{\mathrm{T}}$ denotes the transpose operation, $\hat{\mathbf{x}}_{k|k-1}$ and $\mathbf{P}_{k|k-1}$ are the predicted state vector and corresponding predicted error covariance matrix respectively, $\hat{\mathbf{x}}_{k|k}$ and $\mathbf{P}_{k|k}$ are the state estimation vector and corresponding estimation error covariance matrix respectively, and $\mathbf{K}_k$ is the Kalman gain.

The Kalman filter is a minimum mean square error (MMSE) estimator for a linear state-space model with Gaussian process and measurement noises. However, it is suboptimal for heavy-tailed non-Gaussian process and measurement noises since the required Gaussian assumptions are violated. Moreover, the Kalman filter may break down because a single outlier from any of the process and measurement noises can result in the filter's bias exceeding target bounds [17]. The Student's t filter discussed in the next subsection is intended to solve this problem.

B. Review of Student's t filter

The key idea of the Student's t filter is to approximate the posterior PDF by a Student's t PDF. To that end, the jointly predicted PDF $p(\mathbf{x}_k, \mathbf{z}_k|\mathbf{z}_{1:k-1})$ of state and measurement is assumed to be a Student's t PDF as follows [4]

$$p(\mathbf{x}_k, \mathbf{z}_k|\mathbf{z}_{1:k-1}) = \mathrm{St}(\begin{bmatrix}\mathbf{x}_k\\ \mathbf{z}_k\end{bmatrix}; \begin{bmatrix}\hat{\mathbf{x}}_{k|k-1}\\ \mathbf{H}_k\hat{\mathbf{x}}_{k|k-1}\end{bmatrix}, \begin{bmatrix}\mathbf{P}_{k|k-1} & \mathbf{P}_{k|k-1}\mathbf{H}_k^T\\ \mathbf{H}_k\mathbf{P}_{k|k-1} & \mathbf{S}_k\end{bmatrix}, \eta), \tag{9}$$

where $\hat{\mathbf{x}}_{k|k-1}$ and $\mathbf{P}_{k|k-1}$ are defined in (4)-(5), and the scale matrix $\mathbf{S}_k = \mathbf{H}_k\mathbf{P}_{k|k-1}\mathbf{H}_k^T + \mathbf{R}_k$, and η is the dof parameter of the jointly predicted PDF $p(\mathbf{x}_k, \mathbf{z}_k|\mathbf{z}_{1:k-1})$. According to the Bayesian rule and using (9), the posterior PDF $p(\mathbf{x}_k|\mathbf{z}_{1:k})$ can be updated as a Student's t distribution, i.e., [3]

$$p(\mathbf{x}_k|\mathbf{z}_{1:k}) = \mathrm{St}(\mathbf{x}_k; \hat{\mathbf{x}}'_{k|k}, \mathbf{P}'_{k|k}, \eta'), \tag{10}$$

where η' is the dof parameter of the posterior PDF $p(\mathbf{x}_k|\mathbf{z}_{1:k})$, and the updated parameters are given by [3]

$$\eta' = \eta + m \tag{11}$$

$$\hat{\mathbf{x}}'_{k|k} = \hat{\mathbf{x}}_{k|k-1} + \mathbf{P}_{k|k-1}\mathbf{H}_k^T\mathbf{S}_k^{-1}(\mathbf{z}_k - \mathbf{H}_k\hat{\mathbf{x}}_{k|k-1}) \tag{12}$$

$$\mathbf{P}'_{k|k} = \frac{\eta + \Delta_k^2}{\eta + m}(\mathbf{P}_{k|k-1} - \mathbf{P}_{k|k-1}\mathbf{H}_k^T\mathbf{S}_k^{-1}\mathbf{H}_k\mathbf{P}_{k|k-1}) \tag{13}$$

$$\Delta_k = \sqrt{(\mathbf{z}_k - \mathbf{H}_k\hat{\mathbf{x}}_{k|k-1})^T\mathbf{S}_k^{-1}(\mathbf{z}_k - \mathbf{H}_k\hat{\mathbf{x}}_{k|k-1})}, \tag{14}$$

where $\hat{\mathbf{x}}_{k|k-1}$ and $\mathbf{P}_{k|k-1}$ are formulated in (4)-(5).

It is seen from (11) that the dof parameter of the posterior PDF increases by m from time $k-1$ to time k. Thus, the Student's t approximation to the posterior PDF lacks strict closeness to the target distribution. Furthermore, the Student's t filter will degrade to a Kalman filter after a few steps with the increase of dof parameter, and the heavy-tailed properties will be lost. To retain the heavy-tailed properties and closed Student's t-distributed form of the posterior PDF, the moment matching approach has been used to obtain an approximate posterior PDF $\mathrm{St}(\mathbf{x}_k; \hat{\mathbf{x}}_{k|k}, \mathbf{P}_{k|k}, \eta)$ with dof parameter η, where the mean vector $\hat{\mathbf{x}}_{k|k}$ and scale matrix $\mathbf{P}_{k|k}$ are obtained by matching the first two moments, i.e., [3], [4]

$$\hat{\mathbf{x}}_{k|k} = \hat{\mathbf{x}}'_{k|k} \qquad \frac{\eta}{\eta-2}\mathbf{P}_{k|k} = \frac{\eta'}{\eta'-2}\mathbf{P}'_{k|k}. \tag{15}$$

The Student's t filter exhibits good robustness to heavy-tailed process and measurement noises, and has almost identical computational complexity to the Kalman filter.

C. Motivation of this work

Although the Student's t filter can effectively resist the influence of heavy-tailed process and measurement noises, it suffers from the following drawbacks.

(a) In the derivation of the Student's t filter, it is necessary to assume that the Student's t PDFs of process and measurement noises have the same dof parameter. However, this assumption is seldom met in practical application because the process and measurement noises normally have different heavy-tailed characteristics, which will degrade the approximation accuracy to the posterior PDF [3].

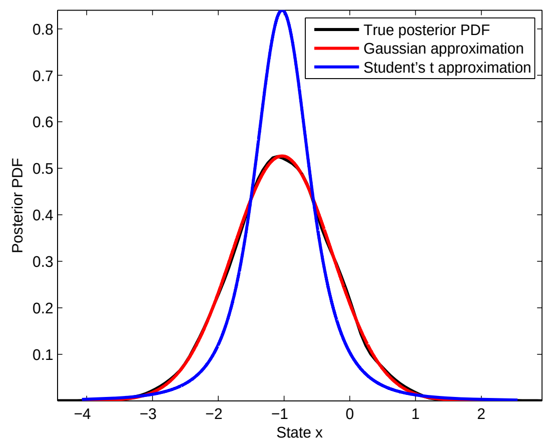

Fig. 1: True posterior distribution, Gaussian approximation and Student's t approximation of the one-dimensional numerical example

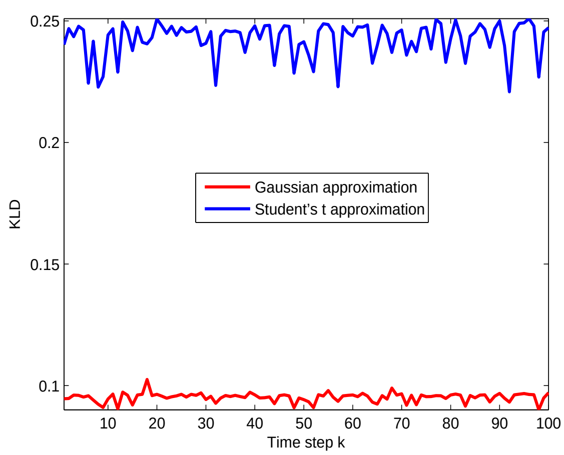

Fig. 2: KLD between the true posterior PDF and Gaussian and Student's t approximations of the one-dimensional numerical example

(b) In the Student's t filter, a moment matching approach is employed to preserve the heavy-tailed properties and closed Student's t-distributed form of the posterior PDF, as shown in (15). However, only the first two moments of the true posterior PDF are captured, and the higher order moments are lost. Thus, the Student's t filter may suffer from a significant bias since the moment matching approach imposes a further bias on the state estimation [4].

(c) The Gaussian approximation to the posterior PDF is more reasonable than the Student's t approximation with a fixed dof parameter in [3] for the case of moderate contaminated process and measurement noises. In order to illustrate this problem, a one-dimensional numerical example is shown. In this example, the state transition parameter $F_k = 0.5$, measurement parameter $H_k = 1$, and outlier corrupted process and measurement noises are generated according to [12]

$$w_k \sim \begin{cases} N(0,1) & \text{w.p. } 0.95 \\ N(0,100) & \text{w.p. } 0.05 \end{cases} \tag{16}$$

$$\mathbf{v}_k \sim \begin{cases} N(0,1) & \text{w.p. } 0.90 \\ N(0,100) & \text{w.p. } 0.10 \end{cases}, \tag{17}$$

where w.p. denotes "with probability". Definitions (16)-(17) imply that w_k and v_k are most frequently drawn from a Gaussian distribution with variances 1 and five percent of process noise values and ten percent of measurement noise values are generated from Gaussian distributions with variances 100. Process and measurement noises, which are generated according to (16)-(17), have heavier tails. In this example, the true posterior distribution is approximated by using the particle filter with 10000 particles [23]. The Gaussian approximation and Student's t approximation to the posterior PDF is obtained from the true posterior distribution with identical first two moments. As suggested in [3], the dof parameter of the Student's t approximation is set as $\eta = 3$. Fig. 1 shows the true posterior distribution, the Gaussian approximation and the Student's t approximation at time step $k = 100$. Fig. 2 shows the Kullback-Leibler divergence (KLD) between the true posterior PDF and Gaussian and Student's t approximations from time step $k = 1$ to $k = 100$. We can see from Fig. 1–Fig. 2 that the Gaussian approximation can match the true posterior PDF better than the Student's t approximation. Thus, the Student's t filter may exhibit unacceptable performance in the case of moderate contaminated process and measurement noises if the dof parameter is poorly chosen.

The problem discussed above represents the main motivation of this work. In this paper, to solve these problems, a new robust Student's t based Kalman filter will be derived, where the one-step predicted PDF and the likelihood PDF are modelled as Student's t PDFs with different dof parameters, and the posterior PDF is updated as a Gaussian distribution.

III. A Robust Student's t based Kalman filter

A. *Student's t based novel hierarchical Gaussian state-space model*

In order to derive an approximate closed form solution of the posterior PDF for a linear state-space model with heavy-tailed process and measurement noises, the one-step predicted PDF $p(\mathbf{x}_k|\mathbf{z}_{1:k-1})$ and the likelihood PDF $p(\mathbf{z}_k|\mathbf{x}_k)$ need to be modelled. Firstly, the heavy-tailed measurement noise is modelled as a Student's t distribution as follows

$$p(\mathbf{v}_k) = \mathrm{St}(\mathbf{v}_k; \mathbf{0}, \mathbf{R}_k, \nu), \tag{18}$$

where $\mathrm{St}(\mathbf{v}_k; \mathbf{0}, \mathbf{R}_k, \nu)$ denotes the Student's t PDF of the measurement noise with mean vector $\mathbf{0}$, scale matrix $\mathbf{R}_k$ (the nominal covariance matrix of measurement noise), and dof parameter ν. Using (2) and (18), the likelihood PDF $p(\mathbf{z}_k|\mathbf{x}_k)$ can be formulated as

$$p(\mathbf{z}_k|\mathbf{x}_k) = \mathrm{St}(\mathbf{z}_k; \mathbf{H}_k\mathbf{x}_k, \mathbf{R}_k, \nu). \tag{19}$$

Secondly, to model the heavy-tailed process noise, the one-step predicted PDF $p(\mathbf{x}_k|\mathbf{z}_{1:k-1})$ is approximated as a

Student's t distribution as follows

$$p(\mathbf{x}_k|\mathbf{z}_{1:k-1}) = \mathrm{St}(\mathbf{x}_k; \mu_k, \mathbf{\Sigma}_k, \omega), \tag{20}$$

where $\mathrm{St}(\mathbf{x}_k; \hat{\mathbf{x}}_{k|k-1}, \mathbf{\Sigma}_k, \omega)$ denotes the Student's t PDF of the state vector $\mathbf{x}_k$ given measurements $\mathbf{z}_{1:k-1}$ with mean vector μ_k, scale matrix $\mathbf{\Sigma}_k$, and dof parameter ω. According to the Bayesian theorem and using (1), $p(\mathbf{x}_k|\mathbf{z}_{1:k-1})$ can be computed as

$$\begin{aligned} p(\mathbf{x}_k|\mathbf{z}_{1:k-1}) &= \int p(\mathbf{x}_k|\mathbf{x}_{k-1})p(\mathbf{x}_{k-1}|\mathbf{z}_{1:k-1})d\mathbf{x}_{k-1} \\ &= \int p_{\mathbf{w}}(\mathbf{x}_k - \mathbf{F}_{k-1}\mathbf{x}_{k-1})p(\mathbf{x}_{k-1}|\mathbf{z}_{1:k-1})d\mathbf{x}_{k-1}, \end{aligned} \tag{21}$$

where $p_{\mathbf{w}}(\cdot)$ is the PDF of the process noise.

Considering that the process noise vector $\mathbf{w}_k$ has zero mean and nominal covariance matrix $\mathbf{Q}_k$ and using (21), the mean vector μ_k and nominal covariance matrix $\mathbf{\Delta}_{k|k-1}$ can be computed as

$$\mu_k = \mathbf{F}_{k-1}\hat{\mathbf{x}}_{k-1|k-1} = \hat{\mathbf{x}}_{k|k-1} \tag{22}$$

$$\mathbf{\Delta}_{k|k-1} = \mathbf{F}_{k-1}\mathbf{P}_{k-1|k-1}\mathbf{F}_{k-1}^T + \mathbf{Q}_{k-1} = \mathbf{P}_{k|k-1}. \tag{23}$$

One may choose the nominal covariance matrix $\mathbf{P}_{k|k-1}$ as the scale matrix $\mathbf{\Sigma}_k$ of the one-step predicted PDF $p(\mathbf{x}_k|\mathbf{z}_{1:k-1})$. However, $\mathbf{Q}_{k-1}$ is not accurate and the large process uncertainty induced by heavy-tailed process noise will be introduced into the scale matrix $\mathbf{\Sigma}_k$, and accurate $\mathbf{\Sigma}_k$ will be unknown, which will degrade the estimation performance.

To solve this problem, the unknown scale matrix $\mathbf{\Sigma}_k$ will be adaptively estimated based on the VB approach. To this end, a conjugate prior distribution needs to be firstly selected for the unknown scale matrix $\mathbf{\Sigma}_k$ since the conjugacy can guarantee that the posterior distribution is of the same functional form as the prior distribution. In Bayesian statistics, the inverse Wishart distribution is usually used as the conjugate prior for the covariance matrix of a Gaussian distribution with known mean [24]. The inverse Wishart PDF of a symmetric positive definite random matrix $\mathbf{B}$ of dimension $d \times d$ is formulated as $\mathrm{IW}(\mathbf{B}; \rho, \mathbf{\Psi}) = \frac{|\mathbf{\Psi}|^{\rho/2}|\mathbf{B}|^{-(\rho+d+1)/2}\exp\{-0.5\mathrm{tr}(\mathbf{\Psi}\mathbf{B}^{-1})\}}{2^{d\rho/2}\Gamma_d(\rho/2)}$, where ρ is the dof parameter, and $\mathbf{\Psi}$ is the inverse scale matrix that is a symmetric positive definite matrix of dimension $d \times d$, and $|\cdot|$ and $\mathrm{tr}(\cdot)$ denote the determinant and trace operations respectively, and $\Gamma_d(\cdot)$ is the d-variate gamma function [24]. If $\mathbf{B} \sim \mathrm{IW}(\mathbf{B}; \rho, \mathbf{\Psi})$, then $\mathrm{E}[\mathbf{B}^{-1}] = (\rho - d - 1)\mathbf{\Psi}^{-1}$ when $\rho > d + 1$ [24]. Since the scale matrix of a Student's t distribution is proportional to the covariance matrix of a Gaussian distribution in the hierarchical Gaussian form, as shown in equation (30), an inverse Wishart distribution is chosen as a prior distribution for the unknown scale matrix $\mathbf{\Sigma}_k$ as follows

$$p(\mathbf{\Sigma}_k) = \mathrm{IW}(\mathbf{\Sigma}_k; u_k, \mathbf{U}_k), \tag{24}$$

where $\mathrm{IW}(\mathbf{\Sigma}_k; u_k, \mathbf{U}_k)$ denotes the inverse Wishart PDF of $\mathbf{\Sigma}_k$ with dof parameter u_k and inverse scale matrix $\mathbf{U}_k$. To capture the prior information of $\mathbf{\Sigma}_k$, the mean value of $p(\mathbf{\Sigma}_k)$ is set as the nominal covariance matrix $\mathbf{P}_{k|k-1}$, i.e.

$$\frac{\mathbf{U}_k}{u_k - n - 1} = \mathbf{P}_{k|k-1}. \tag{25}$$

Let

$$u_k = n + \tau + 1, \tag{26}$$

where $\tau \geq 0$ is a tuning parameter. Using (26) in (25) results in

$$\mathbf{U}_k = \tau\mathbf{P}_{k|k-1}. \tag{27}$$

A new Student's t based state-space model thereby consists of (19)-(20), (24) and (26)-(27). However, the closed solution of the posterior PDF is unavailable because the Student's t PDF is not strictly closed. To solve this problem, the Student's t based state-space model needs to be transformed into a hierarchical Gaussian state-space model by introducing two auxiliary random variables.

Since the Student's t PDF can be viewed as an infinite mixture of Gaussian PDFs, the one-step predicted PDF $p(\mathbf{x}_k|\mathbf{z}_{1:k-1})$ and the likelihood PDF $p(\mathbf{z}_k|\mathbf{x}_k)$ can be rewritten as follows [12]

$$p(\mathbf{x}_k|\mathbf{z}_{1:k-1}) = \int \mathrm{N}(\mathbf{x}_k; \hat{\mathbf{x}}_{k|k-1}, \mathbf{\Sigma}_k/\xi_k)\mathrm{G}(\xi_k; \frac{\omega}{2}, \frac{\omega}{2})d\xi_k \tag{28}$$

$$p(\mathbf{z}_k|\mathbf{x}_k) = \int \mathrm{N}(\mathbf{z}_k; \mathbf{H}_k\mathbf{x}_k, \mathbf{R}_k/\lambda_k)\mathrm{G}(\lambda_k; \frac{\nu}{2}, \frac{\nu}{2})d\lambda_k, \tag{29}$$

where $\mathrm{G}(\cdot; \alpha, \beta)$ denotes the Gamma PDF with shape parameter α and rate parameter β, and ξ_k and λ_k are auxiliary random variables. According to (28)-(29), the one-step predicted PDF $p(\mathbf{x}_k|\mathbf{z}_{1:k-1})$ and the likelihood PDF $p(\mathbf{z}_k|\mathbf{x}_k)$ can be rewritten in the following hierarchical Gaussian forms

$$\begin{aligned} p(\mathbf{x}_k|\xi_k, \mathbf{z}_{1:k-1}) &= \mathrm{N}(\mathbf{x}_k; \hat{\mathbf{x}}_{k|k-1}, \mathbf{\Sigma}_k/\xi_k) \\ p(\xi_k) &= \mathrm{G}(\xi_k; \frac{\omega}{2}, \frac{\omega}{2}) \end{aligned} \tag{30}$$

$$\begin{aligned} p(\mathbf{z}_k|\mathbf{x}_k, \lambda_k) &= \mathrm{N}(\mathbf{z}_k; \mathbf{H}_k\mathbf{x}_k, \mathbf{R}_k/\lambda_k) \\ p(\lambda_k) &= \mathrm{G}(\lambda_k; \frac{\nu}{2}, \frac{\nu}{2}), \end{aligned} \tag{31}$$

where $p(\mathbf{\Sigma}_k)$ is given by (24) and (26)-(27).

Equations (22)-(24), (26)-(27) and (30)-(31) constitute a Student's t based hierarchical Gaussian state-space model. The problem of state estimation for a linear state space model with heavy-tailed process and measurement noises is transformed into the problem of state estimation for a Student's t based hierarchical Gaussian state-space model.

B. A new robust Student's t based Kalman filter

To estimate the state $\mathbf{x}_k$ of a hierarchical Gaussian state-space model formulated in (22)-(24), (26)-(27) and (30)-(31), we need to compute the joint posterior PDF $p(\mathbf{x}_k, \xi_k, \mathbf{\Sigma}_k, \lambda_k|\mathbf{z}_{1:k})$. For the hierarchical Gaussian state-space model, there is not an analytical solution for this posterior PDF. Thus, to obtain an approximate solution, the VB approach is used to look for a free form factored approximate PDF for $p(\mathbf{x}_k, \xi_k, \mathbf{\Sigma}_k, \lambda_k|\mathbf{z}_{1:k})$, i.e. [25], [26]

$$p(\mathbf{x}_k, \xi_k, \mathbf{\Sigma}_k, \lambda_k|\mathbf{z}_{1:k}) \approx q(\mathbf{x}_k)q(\xi_k)q(\mathbf{\Sigma}_k)q(\lambda_k), \tag{32}$$

where $q(\cdot)$ is the approximate posterior PDF. According to the standard VB approach, these approximate posterior PDFs can be obtained by minimizing the KLD between the approximate posterior PDF $q(\mathbf{x}_k)q(\xi_k)q(\mathbf{\Sigma}_k)q(\lambda_k)$ and true posterior PDF $p(\mathbf{x}_k, \xi_k, \mathbf{\Sigma}_k, \lambda_k|\mathbf{z}_{1:k})$, i.e. [25], [26]

$$\begin{aligned}&\{q(\mathbf{x}_k), q(\xi_k), q(\mathbf{\Sigma}_k), q(\lambda_k)\} = \arg\min \mathrm{KLD}\\&(q(\mathbf{x}_k)q(\xi_k)q(\mathbf{\Sigma}_k)q(\lambda_k)||p(\mathbf{x}_k, \xi_k, \mathbf{\Sigma}_k, \lambda_k|\mathbf{z}_{1:k})),\end{aligned} \quad (33)$$

where $\mathrm{KLD}(q(x)||p(x)) \triangleq \int q(x)\log\frac{q(x)}{p(x)}dx$ is the KLD. The optimal solution for (33) satisfies the following equation [12].

$$\log q(\phi) = \mathrm{E}_{\mathbf{\Theta}^{(-\phi)}}[\log p(\mathbf{\Theta}, \mathbf{z}_{1:k})] + c_\phi \quad (34)$$

$$\mathbf{\Theta} \triangleq \{\mathbf{x}_k, \xi_k, \mathbf{\Sigma}_k, \lambda_k\}, \quad (35)$$

where ϕ is an arbitrary element of $\mathbf{\Theta}$, and $\mathbf{\Theta}^{(-\phi)}$ is the set of all elements in $\mathbf{\Theta}$ except for ϕ, and $\mathrm{E}[\cdot]$ denotes the statistical expectation operation, and c_ϕ denotes the constant with respect to variable ϕ. Since the variational parameters of $q(\mathbf{x}_k)$, $q(\xi_k)$, $q(\mathbf{\Sigma}_k)$ and $q(\lambda_k)$ are coupled, we need to utilize fixed-point iterations to solve equation (34), where only one factor in (32) is updated while keeping other factors fixed [12], [25], [26].

1) Computations of approximate posterior PDFs: Using the conditional independence properties of the hierarchical Gaussian state-space model in (22)-(24), (26)-(27) and (30)-(31), the joint PDF $p(\mathbf{\Theta}, \mathbf{z}_{1:k})$ can be factored as

$$p(\mathbf{\Theta}, \mathbf{z}_{1:k}) = p(\mathbf{z}_k|\mathbf{x}_k, \lambda_k)p(\mathbf{x}_k|\xi_k, \mathbf{z}_{1:k-1})p(\xi_k)p(\mathbf{\Sigma}_k)p(\lambda_k). \quad (36)$$

Substituting (24) and (30)-(31) in (36) results in

$$\begin{aligned}&p(\mathbf{\Theta}, \mathbf{z}_{1:k}) = \mathrm{N}(\mathbf{z}_k; \mathbf{H}_k\mathbf{x}_k, \mathbf{R}_k/\lambda_k)\mathrm{N}(\mathbf{x}_k; \hat{\mathbf{x}}_{k|k-1}, \mathbf{\Sigma}_k/\xi_k) \times\\&\mathrm{G}(\xi_k; \frac{\omega}{2}, \frac{\omega}{2})\mathrm{IW}(\mathbf{\Sigma}_k; u_k, \mathbf{U}_k)\mathrm{G}(\lambda_k; \frac{\nu}{2}, \frac{\nu}{2}).\end{aligned} \quad (37)$$

Using (37), $\log p(\mathbf{\Theta}, \mathbf{z}_{1:k})$ can be formulated as

$$\begin{aligned}&\log p(\mathbf{\Theta}, \mathbf{z}_{1:k}) = (\frac{m+\nu}{2} - 1)\log\lambda_k - \frac{\nu}{2}\lambda_k -\\&\frac{\lambda_k}{2}(\mathbf{z}_k - \mathbf{H}_k\mathbf{x}_k)^T\mathbf{R}_k^{-1}(\mathbf{z}_k - \mathbf{H}_k\mathbf{x}_k) + (\frac{n+\omega}{2} - 1)\log\xi_k\\&-\frac{\omega}{2}\xi_k - \frac{\xi_k}{2}(\mathbf{x}_k - \hat{\mathbf{x}}_{k|k-1})^T\mathbf{\Sigma}_k^{-1}(\mathbf{x}_k - \hat{\mathbf{x}}_{k|k-1}) -\\&\frac{1}{2}\mathrm{tr}(\mathbf{U}_k\mathbf{\Sigma}_k^{-1}) - \frac{1}{2}(n + u_k + 2)\log|\mathbf{\Sigma}_k|.\end{aligned} \quad (38)$$

Let $\phi = \xi_k$ and using (38) in (34), we obtain

$$\begin{aligned}&\log q^{(i+1)}(\xi_k) = (\frac{n+\omega}{2} - 1)\log\xi_k -\\&0.5\left\{\omega + \mathrm{tr}(\mathbf{D}_k^{(i)}\mathrm{E}^{(i)}[\mathbf{\Sigma}_k^{-1}])\right\}\xi_k + c_\xi,\end{aligned} \quad (39)$$

where $q^{(i+1)}(\cdot)$ is the approximation of PDF $q(\cdot)$ at the $i+1$th iteration, and $\mathrm{E}^{(i)}[\rho]$ is the expectation of variable ρ at the ith iteration, and $\mathrm{tr}(\cdot)$ denotes the trace operation and $\mathbf{D}_k^{(i)}$ is given by

$$\mathbf{D}_k^{(i)} = \mathrm{E}^{(i)}[(\mathbf{x}_k - \hat{\mathbf{x}}_{k|k-1})(\mathbf{x}_k - \hat{\mathbf{x}}_{k|k-1})^T]. \quad (40)$$

Employing (39), $q^{(i+1)}(\xi_k)$ can be updated as a Gamma PDF with shape parameter α_k^{i+1} and rate parameter β_k^{i+1}, i.e.

$$q^{(i+1)}(\xi_k) = \mathrm{G}(\xi_k; \alpha_k^{i+1}, \beta_k^{i+1}), \quad (41)$$

where shape parameter α_k^{i+1} and rate parameter β_k^{i+1} are given by

$$\alpha_k^{i+1} = 0.5(n + \omega) \quad (42)$$

$$\beta_k^{i+1} = 0.5\left\{\omega + \mathrm{tr}(\mathbf{D}_k^{(i)}\mathrm{E}^{(i)}[\mathbf{\Sigma}_k^{-1}])\right\}. \quad (43)$$

Let $\phi = \lambda_k$ and using (38) in (34), we obtain

$$\begin{aligned}&\log q^{(i+1)}(\lambda_k) = (\frac{m+\nu}{2} - 1)\log\lambda_k -\\&0.5\left\{\nu + \mathrm{tr}(\mathbf{E}_k^{(i)}\mathbf{R}_k^{-1})\right\}\lambda_k + c_\lambda,\end{aligned} \quad (44)$$

where $\mathbf{E}_k^{(i)}$ is given by

$$\mathbf{E}_k^{(i)} = \mathrm{E}^{(i)}[(\mathbf{z}_k - \mathbf{H}_k\mathbf{x}_k)(\mathbf{z}_k - \mathbf{H}_k\mathbf{x}_k)^T]. \quad (45)$$

Using (44), $q^{(i+1)}(\lambda_k)$ can be updated as a Gamma PDF with shape parameter γ_k^{i+1} and rate parameter δ_k^{i+1}, i.e.

$$q^{(i+1)}(\lambda_k) = \mathrm{G}(\lambda_k; \gamma_k^{i+1}, \delta_k^{i+1}), \quad (46)$$

where shape parameter γ_k^{i+1} and rate parameter δ_k^{i+1} are given by

$$\gamma_k^{i+1} = 0.5(m + \nu) \quad (47)$$

$$\delta_k^{i+1} = 0.5\left\{\nu + \mathrm{tr}(\mathbf{E}_k^{(i)}\mathbf{R}_k^{-1})\right\}. \quad (48)$$

Let $\phi = \mathbf{\Sigma}_k$ and using (38) in (34), we obtain

$$\begin{aligned}&\log q^{(i+1)}(\mathbf{\Sigma}_k) = -\frac{1}{2}(n + u_k + 2)\log|\mathbf{\Sigma}_k| -\\&\frac{1}{2}\mathrm{tr}\left[(\mathbf{U}_k + \mathrm{E}^{(i+1)}[\xi_k]\mathbf{D}_k^{(i)})\mathbf{\Sigma}_k^{-1}\right] + c_{\mathbf{\Sigma}}.\end{aligned} \quad (49)$$

Exploiting (49), $q^{(i+1)}(\mathbf{\Sigma}_k)$ can be updated as an inverse Wishart PDF with dof parameter $\hat{u}_k^{(i+1)}$ and inverse scale matrix $\hat{\mathbf{U}}_k^{(i+1)}$, i.e.

$$q^{(i+1)}(\mathbf{\Sigma}_k) = \mathrm{IW}(\mathbf{\Sigma}_k; \hat{u}_k^{(i+1)}, \hat{\mathbf{U}}_k^{(i+1)}), \quad (50)$$

where the dof parameter $\hat{u}_k^{(i+1)}$ and inverse scale matrix $\hat{\mathbf{U}}_k^{(i+1)}$ are given by

$$\hat{u}_k^{(i+1)} = u_k + 1 \quad (51)$$

$$\hat{\mathbf{U}}_k^{(i+1)} = \mathbf{U}_k + \mathrm{E}^{(i+1)}[\xi_k]\mathbf{D}_k^{(i)}. \quad (52)$$

Let $\phi = \mathbf{x}_k$ and using (38) in (34), we obtain

$$\begin{aligned}&\log q^{(i+1)}(\mathbf{x}_k) = -0.5\mathrm{E}^{(i+1)}[\lambda_k](\mathbf{z}_k - \mathbf{H}_k\mathbf{x}_k)^T\mathbf{R}_k^{-1} \times\\&(\mathbf{z}_k - \mathbf{H}_k\mathbf{x}_k) - 0.5\mathrm{E}^{(i+1)}[\xi_k](\mathbf{x}_k - \hat{\mathbf{x}}_{k|k-1})^T\mathrm{E}^{(i+1)}[\mathbf{\Sigma}_k^{-1}] \times\\&(\mathbf{x}_k - \hat{\mathbf{x}}_{k|k-1}) + c_{\mathbf{x}}.\end{aligned} \quad (53)$$

Define the modified one-step predicted PDF $p^{(i+1)}(\mathbf{x}_k|\mathbf{z}_{1:k-1})$ and the modified likelihood PDF $p^{(i+1)}(\mathbf{z}_k|\mathbf{x}_k)$ as follows

$$p^{(i+1)}(\mathbf{x}_k|\mathbf{z}_{1:k-1}) = \mathrm{N}(\mathbf{x}_k; \hat{\mathbf{x}}_{k|k-1}, \tilde{\mathbf{P}}_{k|k-1}^{(i+1)}) \quad (54)$$

$$p^{(i+1)}(\mathbf{z}_k|\mathbf{x}_k) = \mathrm{N}(\mathbf{z}_k; \mathbf{H}_k\mathbf{x}_k, \tilde{\mathbf{R}}_k^{(i+1)}), \quad (55)$$

where the modified measurement noise covariance matrix $\tilde{\mathbf{R}}_k^{(i+1)}$ and modified predicted error covariance matrix $\tilde{\mathbf{P}}_{k|k-1}$ are given by

$$\tilde{\mathbf{R}}_k^{(i+1)} = \frac{\mathbf{R}_k}{\mathrm{E}^{(i+1)}[\lambda_k]} \quad \tilde{\mathbf{P}}_{k|k-1}^{(i+1)} = \frac{\left\{\mathrm{E}^{(i+1)}[\mathbf{\Sigma}_k^{-1}]\right\}^{-1}}{\mathrm{E}^{(i+1)}[\xi_k]}. \quad (56)$$

Using (54)-(56) in (53), we obtain

$$q^{(i+1)}(\mathbf{x}_k) = \frac{1}{c_k^{(i+1)}} p^{(i+1)}(\mathbf{z}_k|\mathbf{x}_k) p^{(i+1)}(\mathbf{x}_k|\mathbf{z}_{1:k-1}), \quad (57)$$

where the normalizing constant $c_k^{(i+1)}$ is given by

$$c_k^{(i+1)} = \int p^{(i+1)}(\mathbf{z}_k|\mathbf{x}_k) p^{(i+1)}(\mathbf{x}_k|\mathbf{z}_{1:k-1}) d\mathbf{x}_k. \quad (58)$$

Employing (54)-(58), $q^{(i+1)}(\mathbf{x}_k)$ can be updated as a Gaussian PDF with mean vector $\hat{\mathbf{x}}_{k|k}^{(i+1)}$ and covariance matrix $\mathbf{P}_{k|k}^{(i+1)}$, i.e.,

$$q^{(i+1)}(\mathbf{x}_k) = \mathrm{N}(\mathbf{x}_k; \hat{\mathbf{x}}_{k|k}^{(i+1)}, \mathbf{P}_{k|k}^{(i+1)}), \quad (59)$$

where the mean vector $\hat{\mathbf{x}}_{k|k}^{(i+1)}$ and covariance matrix $\mathbf{P}_{k|k}^{(i+1)}$ are given by

$$\hat{\mathbf{x}}_{k|k}^{(i+1)} = \hat{\mathbf{x}}_{k|k-1} + \mathbf{K}_k^{(i+1)}(\mathbf{z}_k - \mathbf{H}_k \hat{\mathbf{x}}_{k|k-1}) \quad (60)$$

$$\mathbf{P}_{k|k}^{(i+1)} = \tilde{\mathbf{P}}_{k|k-1}^{(i+1)} - \mathbf{K}_k^{(i+1)} \mathbf{H}_k \tilde{\mathbf{P}}_{k|k-1}^{(i+1)}, \quad (61)$$

where $\mathbf{K}_k^{(i+1)}$ denotes the modified Kalman gain given by

$$\mathbf{K}_k^{(i+1)} = \tilde{\mathbf{P}}_{k|k-1}^{(i+1)} \mathbf{H}_k^T (\mathbf{H}_k \tilde{\mathbf{P}}_{k|k-1}^{(i+1)} \mathbf{H}_k^T + \tilde{\mathbf{R}}_k^{(i+1)})^{-1}. \quad (62)$$

After fixed point iteration N, the approximate posterior PDFs $q(\mathbf{x}_k)$, $q(\xi_k)$, $q(\mathbf{\Sigma}_k)$ and $q(\lambda_k)$ can be updated as

$$q(\mathbf{x}_k) \approx q^{(N)}(\mathbf{x}_k) = \mathrm{N}(\mathbf{x}_k; \hat{x}_{k|k}^{(N)}, \mathbf{P}_{k|k}^{(N)}) = \mathrm{N}(\mathbf{x}_k; \hat{x}_{k|k}, \mathbf{P}_{k|k}) \quad (63)$$

$$q(\xi_k) \approx q^{(N)}(\xi_k) = \mathrm{G}(\xi_k; \alpha_k^N, \beta_k^N) \quad (64)$$

$$q(\mathbf{\Sigma}_k) \approx q^{(N)}(\mathbf{\Sigma}_k) = \mathrm{IW}(\mathbf{\Sigma}_k; \hat{u}_k^{(N)}, \hat{\mathbf{U}}_{k|k}^{(N)}) \quad (65)$$

$$q(\lambda_k) \approx q^{(N)}(\lambda_k) = \mathrm{G}(\lambda_k; \gamma_k^N, \delta_k^N). \quad (66)$$

2) Computation of expectations: Using (41), (46) and (50), we can compute the required expectations $\mathrm{E}^{(i+1)}[\xi_k]$, $\mathrm{E}^{(i+1)}[\lambda_k]$ and $\mathrm{E}^{(i+1)}[\mathbf{\Sigma}_k^{-1}]$ as follows

$$\mathrm{E}^{(i+1)}[\xi_k] = \frac{\alpha_k^{i+1}}{\beta_k^{i+1}} \quad \mathrm{E}^{(i+1)}[\lambda_k] = \frac{\gamma_k^{i+1}}{\delta_k^{i+1}}$$
$$\mathrm{E}^{(i+1)}[\mathbf{\Sigma}_k^{-1}] = (\hat{u}_k^{(i+1)} - n - 1)(\hat{\mathbf{U}}_k^{(i+1)})^{-1}. \quad (67)$$

Exploiting (59), the required expectations $\mathbf{D}_k^{(i+1)}$ and $\mathbf{E}_k^{(i+1)}$ are computed as

$$\begin{aligned}\mathbf{D}_k^{(i+1)} =& \mathrm{E}^{(i+1)}[(\mathbf{x}_k - \hat{\mathbf{x}}_{k|k-1})(\mathbf{x}_k - \hat{\mathbf{x}}_{k|k-1})^T] \\ =& \mathrm{E}^{(i+1)}[(\mathbf{x}_k - \hat{\mathbf{x}}_{k|k}^{(i+1)} + \hat{\mathbf{x}}_{k|k}^{(i+1)} - \hat{\mathbf{x}}_{k|k-1}) \times \\ & (\mathbf{x}_k - \hat{\mathbf{x}}_{k|k}^{(i+1)} + \hat{\mathbf{x}}_{k|k}^{(i+1)} - \hat{\mathbf{x}}_{k|k-1})^T] \\ =& \mathrm{E}^{(i+1)}[(\mathbf{x}_k - \hat{\mathbf{x}}_{k|k}^{(i+1)})(\mathbf{x}_k - \hat{\mathbf{x}}_{k|k}^{(i+1)})^T] + \\ & (\hat{\mathbf{x}}_{k|k}^{(i+1)} - \hat{\mathbf{x}}_{k|k-1})(\hat{\mathbf{x}}_{k|k}^{(i+1)} - \hat{\mathbf{x}}_{k|k-1})^T \\ =& \mathbf{P}_{k|k}^{(i+1)} + (\hat{\mathbf{x}}_{k|k}^{(i+1)} - \hat{\mathbf{x}}_{k|k-1})(\hat{\mathbf{x}}_{k|k}^{(i+1)} - \hat{\mathbf{x}}_{k|k-1})^T\end{aligned} \quad (68)$$

Algorithm 1: One time step of the proposed robust Kalman filter for linear state-space model with heavy-tailed process and measurement noises

---

Inputs: $\hat{\mathbf{x}}_{k-1|k-1}$, $\mathbf{P}_{k-1|k-1}$, $\mathbf{F}_{k-1}$, $\mathbf{H}_k$, $\mathbf{z}_k$, $\mathbf{Q}_{k-1}$, $\mathbf{R}_k$, m, n, ω, ν, τ, N

Time update:

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

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

Measurement update:

3. Initialization: $u_k = n + \tau + 1$, $\mathbf{U}_k = \tau \mathbf{P}_{k|k-1}$, $\hat{\mathbf{x}}_{k|k}^{(0)} = \hat{\mathbf{x}}_{k|k-1}$, $\mathbf{P}_{k|k}^{(0)} = \mathbf{P}_{k|k-1}$, $\mathrm{E}^{(0)}[\mathbf{\Sigma}_k^{-1}] = (u_k - n - 1)\mathbf{U}_k^{-1}$

for $i = 0 : N-1$

Update $q^{(i+1)}(\xi_k) = \mathrm{G}(\xi_k; \alpha_k^{i+1}, \beta_k^{i+1})$ given $q^{(i)}(\mathbf{x}_k)$ and $q^{(i)}(\mathbf{\Sigma}_k)$:

4. $\mathbf{D}_k^{(i)} = \mathbf{P}_{k|k}^{(i)} + (\hat{\mathbf{x}}_{k|k}^{(i)} - \hat{\mathbf{x}}_{k|k-1})(\hat{\mathbf{x}}_{k|k}^{(i)} - \hat{\mathbf{x}}_{k|k-1})^T$

5. $\alpha_k^{i+1} = 0.5(n + \omega)$, $\quad \beta_k^{i+1} = 0.5\left\{\omega + \mathrm{tr}(\mathbf{D}_k^{(i)}\mathrm{E}^{(i)}[\mathbf{\Sigma}_k^{-1}])\right\}$, $\mathrm{E}^{(i+1)}[\xi_k] = \alpha_k^{i+1}/\beta_k^{i+1}$

Update $q^{(i+1)}(\lambda_k) = \mathrm{G}(\lambda_k; \gamma_k^{i+1}, \delta_k^{i+1})$ given $q^{(i)}(\mathbf{x}_k)$:

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

7. $\gamma_k^{i+1} = 0.5(m + \nu)$, $\quad \delta_k^{i+1} = 0.5\left\{\nu + \mathrm{tr}(\mathbf{E}_k^{(i)}\mathbf{R}_k^{-1})\right\}$, $\mathrm{E}^{(i+1)}[\lambda_k] = \gamma_k^{i+1}/\delta_k^{i+1}$

Update $q^{(i+1)}(\mathbf{\Sigma}_k) = \mathrm{IW}(\mathbf{\Sigma}_k; \hat{u}_k^{(i+1)}, \hat{\mathbf{U}}_k^{(i+1)})$ given $q^{(i)}(\mathbf{x}_k)$ and $q^{(i+1)}(\xi_k)$:

8. $\hat{u}_k^{(i+1)} = u_k + 1$, $\quad \hat{\mathbf{U}}_k^{(i+1)} = \mathbf{U}_k + \mathrm{E}^{(i+1)}[\xi_k]\mathbf{D}_k^{(i)}$, $\mathrm{E}^{(i+1)}[\mathbf{\Sigma}_k^{-1}] = (\hat{u}_k^{(i+1)} - n - 1)(\hat{\mathbf{U}}_k^{(i+1)})^{-1}$

Update $q^{(i+1)}(\mathbf{x}_k) = \mathrm{N}(\mathbf{x}_k; \hat{\mathbf{x}}_{k|k}^{(i+1)}, \mathbf{P}_{k|k}^{(i+1)})$ given $q^{(i+1)}(\xi_k)$, $q^{(i+1)}(\lambda_k)$ and $q^{(i+1)}(\mathbf{\Sigma}_k)$:

9. $\tilde{\mathbf{R}}_k^{(i+1)} = \mathbf{R}_k/\mathrm{E}^{(i+1)}[\lambda_k]$, $\tilde{\mathbf{P}}_{k|k-1}^{(i+1)} = \left\{\mathrm{E}^{(i+1)}[\mathbf{\Sigma}_k^{-1}]\right\}^{-1}/\mathrm{E}^{(i+1)}[\xi_k]$

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

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

end for

13. $\hat{x}_{k|k} = \hat{x}_{k|k}^{(N)}$, $\quad \mathbf{P}_{k|k} = \mathbf{P}_{k|k}^{(N)}$

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

$$\begin{aligned}\mathbf{E}_k^{(i+1)} =& \mathrm{E}^{(i+1)}[(\mathbf{z}_k - \mathbf{H}_k\mathbf{x}_k)(\mathbf{z}_k - \mathbf{H}_k\mathbf{x}_k)^T] \\ =& \mathrm{E}^{(i+1)}[(\mathbf{z}_k - \mathbf{H}_k\hat{\mathbf{x}}_{k|k}^{(i+1)} + \mathbf{H}_k\hat{\mathbf{x}}_{k|k}^{(i+1)} - \mathbf{H}_k\mathbf{x}_k) \times \\ & (\mathbf{z}_k - \mathbf{H}_k\hat{\mathbf{x}}_{k|k}^{(i+1)} + \mathbf{H}_k\hat{\mathbf{x}}_{k|k}^{(i+1)} - \mathbf{H}_k\mathbf{x}_k)^T] \\ =& (\mathbf{z}_k - \mathbf{H}_k\hat{\mathbf{x}}_{k|k}^{(i+1)})(\mathbf{z}_k - \mathbf{H}_k\hat{\mathbf{x}}_{k|k}^{(i+1)})^T + \\ & \mathbf{H}_k\mathrm{E}^{(i+1)}[(\mathbf{x}_k - \hat{\mathbf{x}}_{k|k}^{(i+1)})(\mathbf{x}_k - \hat{\mathbf{x}}_{k|k}^{(i+1)})^T]\mathbf{H}_k^T \\ =& (\mathbf{z}_k - \mathbf{H}_k\hat{\mathbf{x}}_{k|k}^{(i+1)})(\mathbf{z}_k - \mathbf{H}_k\hat{\mathbf{x}}_{k|k}^{(i+1)})^T + \mathbf{H}_k\mathbf{P}_{k|k}^{(i+1)}\mathbf{H}_k^T.\end{aligned} \quad (69)$$

The proposed robust Kalman filter consists of (4)-(5), (26)-(27), (40)-(43), (45)-(48), (50)-(52), (56) and (59)-(69), and the implementation pseudocode for one time step of the proposed robust Kalman filter is shown in Algorithm 1.

Finally, we discuss the effect of the tuning parameter τ upon the proposed robust Kalman filter. Substituting (67) in (57), the modified predicted error covariance matrix $\tilde{\mathbf{P}}_{k|k-1}$ can be reformulated as

$$\tilde{\mathbf{P}}_{k|k-1}^{(i+1)}=\frac{\left\{(\hat{u}_k^{(i+1)}-n-1)(\hat{\mathbf{U}}_k^{(i+1)})^{-1}\right\}^{-1}}{\mathrm{E}^{(i+1)}[\xi_k]}=\frac{\hat{\mathbf{U}}_k^{(i+1)}}{(\hat{u}_k^{(i+1)}-n-1)\mathrm{E}^{(i+1)}[\xi_k]}. \quad (70)$$

Using (26)-(27) and (51)-(52) in (70) results in

$$\tilde{\mathbf{P}}_{k|k-1}^{(i+1)}=\frac{\tau\mathbf{P}_{k|k-1}/\mathrm{E}^{(i+1)}[\xi_k]+\mathbf{D}_k^{(i)}}{\tau+1}. \quad (71)$$

It is seen from (71) that τ can be deemed as a harmonic weight to balance the efficacy of $\mathbf{P}_{k|k-1}/\mathrm{E}^{(i+1)}[\xi_k]$ and $\mathbf{D}_k^{(i)}$. On the one hand, if τ is too large, the substantial prior uncertainties induced by the heavy-tailed process noise are introduced into the measurement update, which degrades the performance of the proposed filter. On the other hand, if τ is too small, a large quantity of information about the process model is lost, which also degrades the performance of the proposed filter. In this paper, the tuning parameter is chosen to lie within the range $\tau\in[2,\ 6]$, and the proposed filter with $\tau\in[2,\ 6]$ has essentially consistent estimation performance, as shown in the later simulation.

IV. Simulation

In this simulation, the superior performance of the proposed robust Kalman filter as compared with existing filters is illustrated in the problem of tracking an agile target that is observed in clutter. The target moves according to a constant velocity model in two-dimensional space and its position is observed. The linear state space can be formulated as [3]

$$\mathbf{x}_k=\mathbf{F}\mathbf{x}_{k-1}+\mathbf{w}_{k-1} \quad (72)$$

$$\mathbf{z}_k=\mathbf{H}\mathbf{x}_k+\mathbf{v}_k, \quad (73)$$

where the state $\mathbf{x}_k=[x_k\ y_k\ \dot{x}_k\ \dot{y}_k]$; x_k, y_k, $\dot{x}_k$ and $\dot{y}_k$ denote the cartesian coordinates and corresponding velocities; $\mathbf{F}$ and $\mathbf{H}$ denote the state transition matrix and observation matrix respectively, which are given by

$$\mathbf{F}=\begin{bmatrix}\mathbf{I}_2 & \Delta t\mathbf{I}_2\\ \mathbf{0} & \mathbf{I}_2\end{bmatrix}\quad \mathbf{H}=\begin{bmatrix}\mathbf{I}_2 & \mathbf{0}\end{bmatrix}, \quad (74)$$

where the parameter $\Delta t=1\mathrm{s}$ is the sampling interval and $\mathbf{I}_2$ is the two-dimensional identity matrix. Outlier corrupted process and measurement noises are generated according to [12]

$$\mathbf{w}_k\sim\begin{cases}N(\mathbf{0},\mathbf{Q}) & \text{w.p. } 0.95\\ N(\mathbf{0},100\mathbf{Q}) & \text{w.p. } 0.05\end{cases} \quad (75)$$

$$\mathbf{v}_k\sim\begin{cases}N(\mathbf{0},\mathbf{R}) & \text{w.p. } 0.90\\ N(\mathbf{0},100\mathbf{R}) & \text{w.p. } 0.10\end{cases}, \quad (76)$$

where $\mathbf{Q}$ and $\mathbf{R}$ are respectively nominal process and measurement noise covariance matrices

$$\mathbf{Q}=\begin{bmatrix}\frac{\Delta t^3}{3}\mathbf{I}_2 & \frac{\Delta t^2}{2}\mathbf{I}_2\\ \frac{\Delta t^2}{2}\mathbf{I}_2 & \Delta t\mathbf{I}_2\end{bmatrix}q \qquad \mathbf{R}=r\mathbf{I}_2, \quad (77)$$

where $q=1$ and $r=100\mathrm{m}^2$. Equations (75)-(76) imply that $\mathbf{w}_k$ and $\mathbf{v}_k$ are most frequently drawn from a Gaussian distribution with covariance matrix $\mathbf{Q}$ or $\mathbf{R}$ and five percent of process noise values and ten percent of measurement noise values are generated from Gaussian distributions with severely increased covariance matrices. Process and measurement noises, which are generated according to (75)-(76), have heavier tails.

TABLE I: AAVBs of the proposed filters and existing filters

Filters	$\mathrm{AAVB_{pos}}$ (m)	$\mathrm{AAVB_{vel}}$ (m/s)
KFTCM	0.948	0.303
HKF	0.728	0.272
MCCKF	0.742	0.277
STF	0.918	0.349
The proposed filter-fixed	0.607	0.257
The proposed filter-$\mathbf{\Sigma}$	0.561	0.243

TABLE II: Implementation times of the proposed filters and existing filters in single step run when $N=10$

Filters	Time (s)
KFTCM	2.7×10^{-5}
HKF	5.1×10^{-4}
MCCKF	1.0×10^{-4}
STF	3.9×10^{-5}
The proposed filter-fixed	6.5×10^{-4}
The proposed filter-$\mathbf{\Sigma}$	7.4×10^{-4}

In this simulation, the Kalman filter with true covariance matrices (KFTCM), the Huber based Kalman filter (HKF) [16], the maximum correntropy criterion Kalman filter (MCCKF) [22], the Student's t filter (STF) [3], the proposed filter with fixed $\mathbf{\Sigma}_k=\mathbf{P}_{k|k-1}$ (The proposed filter-fixed), and the proposed filter with estimated $\mathbf{\Sigma}_k$ (The proposed filter-$\mathbf{\Sigma}$) are tested. The tuning parameter and the number of iterations of the HKF are chosen as $\gamma=1.345$ [16] and $N=10$, the kernel size of the MCCKF is chosen as $\sigma=15$, and the dof parameter of the STF is chosen as $\nu=3$ [3]. In the proposed filter-fixed and the proposed filter-$\mathbf{\Sigma}$, the dof parameters, tuning parameter and the number of iterations are set as: $\omega=\nu=5$, $\tau=5$, $N=10$. The proposed filters and existing filters are coded with MATLAB and the simulations are run on a computer with Intel Core i7-3770 CPU at 3.40 GHz.

To compare the performance of existing filters and the proposed filter, the root mean square errors (RMSEs), the averaged RMSEs (ARMSEs) and the averaged absolute value of biases (AAVBs) of position and velocity are chosen as performance metrics. The RMSE, ARMSE and AAVB of position are respectively defined as follows

$$\mathrm{RMSE_{pos}}=\sqrt{\frac{1}{M}\sum_{s=1}^{M}((x_k^s-\hat{x}_k^s)^2+(y_k^s-\hat{y}_k^s)^2)} \quad (78)$$

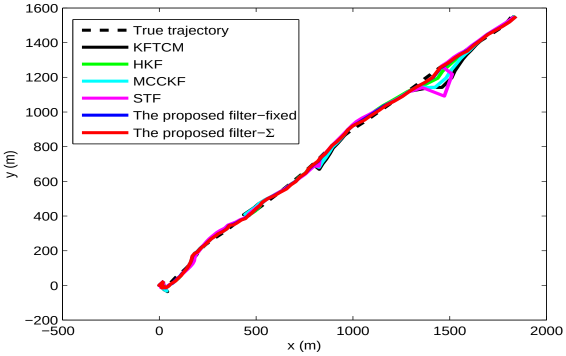

Fig. 3: True and estimated trajectories of the target

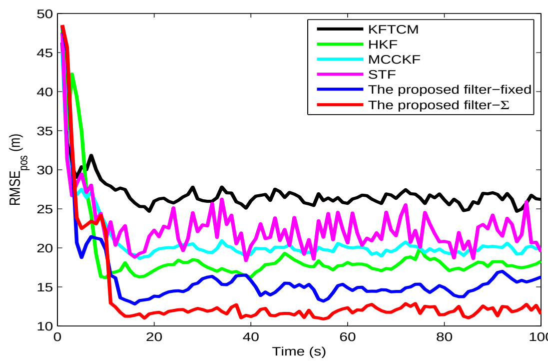

Fig. 4: RMSEs of the position from existing filters and the proposed filters

$$\mathrm{ARMSE}_{\mathrm{pos}} = \sqrt{\frac{1}{MT}\sum_{k=1}^{T}\sum_{s=1}^{M}((x_k^s - \hat{x}_k^s)^2 + (y_k^s - \hat{y}_k^s)^2)} \tag{79}$$

$$\mathrm{AAVB}_{\mathrm{pos}} = \frac{1}{T}\sum_{k=1}^{T}\left|\frac{1}{M}\sum_{s=1}^{M}(x_k^s - \hat{x}_k^s)\right| + \frac{1}{T}\sum_{k=1}^{T}\left|\frac{1}{M}\sum_{s=1}^{M}(y_k^s - \hat{y}_k^s)\right|, \tag{80}$$

where (x_k^s, y_k^s) and $(\hat{x}_k^s, \hat{y}_k^s)$ are the true and estimated positions at the s-th Monte Carlo run, and $M = 1000$ and $T = 100$s are respectively the total number of Monte Carlo runs and the simulation time. Similar to the RMSE, ARMSE and AAVB in position, we can also write formula for the RMSE, ARMSE and AAVB in velocity.

The true and estimated trajectories obtained from the existing filters and the proposed filter in a single Monte Carlo run are shown in Fig. 3. The RMSEs and AAVBs of position and velocity from the existing filters and the proposed filter are respectively shown in Fig. 4–Fig. 5 and Table I. The implementation times of the proposed filters and existing filters in single step run when $N = 10$ are shown in Table II. It is seen from Fig. 3 that the estimated trajectories from the proposed filter are closer to the true trajectory as compared with existing filters, particularly around coordinates $(1500, 1200)$, which is caused by the process and measurement outliers. We can see from Fig. 4–Fig. 5, Table I and Table II that the proposed filter has smaller RMSEs and AAVBs of position and velocity but slightly higher computational complexity than existing filters. Moreover, as expected, the proposed filter-$\boldsymbol{\Sigma}$ has higher estimation accuracy and smaller bias than the proposed filter-fixed.

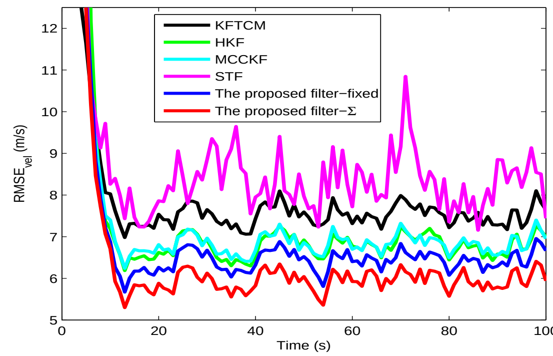

Fig. 5: RMSEs of the velocity from existing filters and the proposed filters

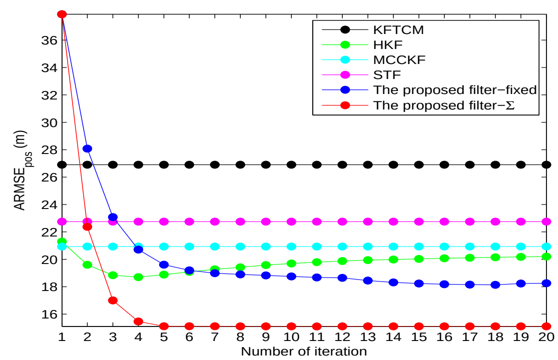

Fig. 6: ARMSEs of the position of the target when $N = 1, 2, \ldots, 20$

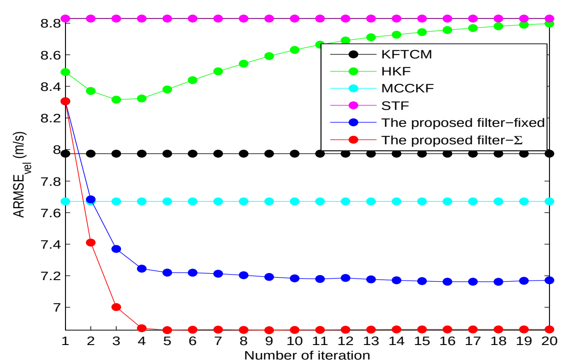

Fig. 7: ARMSEs of the velocity of the target when $N = 1, 2, \ldots, 20$

Fig. 6–Fig. 7 show the ARMSEs of position and velocity from the existing filters and the proposed filters with different

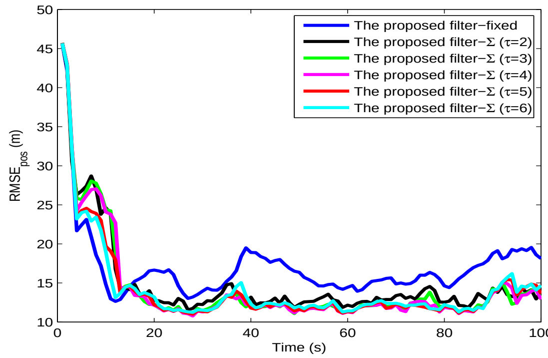

Fig. 8: RMSEs of the position from the proposed filters when $\tau = 2, 3, 4, 5, 6$

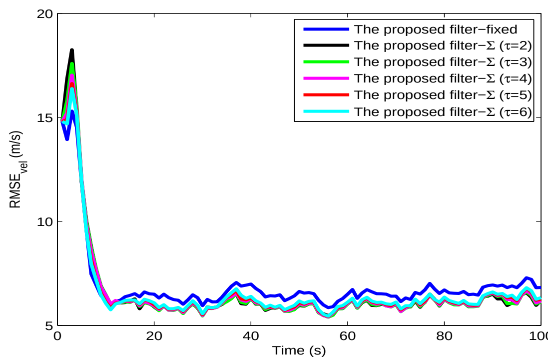

Fig. 9: RMSEs of the velocity from the proposed filters when $\tau = 2, 3, 4, 5, 6$

numbers of iterations $N = 1, 2, \ldots, 20$. It can be seen from Fig. 6–Fig. 7 that the proposed filter-fixed and the proposed filter-$\mathbf{\Sigma}$ respectively outperform existing filters for $N \geq 7$ and $N \geq 3$. The proposed filter-$\mathbf{\Sigma}$ has higher estimation accuracy than the proposed filter-fixed for $N \geq 2$.

Fig. 8–Fig. 9 show the RMSEs of position and velocity from the proposed filters with different tuning parameters $\tau = 2, 3, 4, 5, 6$. We can see from Fig. 8–Fig. 9 that the proposed filter-$\mathbf{\Sigma}$ with tuning parameters $\tau = 2, 3, 4, 5, 6$ has almost identical estimation accuracy, however, it outperforms the proposed filter-fixed in terms of RMSEs of position and velocity.

V. Conclusions

In this paper, a new robust Student's t based Kalman filter was proposed for linear systems with heavy-tailed process and measurement noises, which provided a Gaussian approximation to the posterior PDF. The one-step predicted PDF and likelihood PDF were modelled as Student's t distributions with different dof parameters, and the PDF of the unknown scale matrix was modelled as an inverse Wishart distribution. A Student's t based hierarchical Gaussian state-space model is presented by introducing auxiliary random variables, based on which approximate posterior PDFs of state, unknown scale matrix and auxiliary random variables can be obtained by performing structured VB inference. Simulation results illustrated that the proposed robust Kalman filter outperforms existing state-of-the-art filters in a manoeuvring target tracking example with moderate contaminated process and measurement noises.

References

[1] B. Jia and M. Xin. "Sparse-grid quadrature H_∞ filter for discrete-time systems with uncertain noise statistics," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 49, no. 3, pp. 1626–1636, July 2013.

[2] M-J. Yu. "INS/GPS integration system using adaptive filter for estimating measurement noise variance," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 48, no. 2, pp. 1786–1792, Apr 2012.

[3] M. Roth, E. Özkan, and F. Gustafsson, "A Student's filter for heavy-tailed process and measurement noise," in *2013 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, May 2013, pp. 5770–5774.

[4] Y. L. Huang, Y. G. Zhang, N. Li and J. Chambers, "Robust Student's t based nonlinear filter and smoother," accepted for publication in *IEEE Trans. Aero. Elec. Sys.*, vol. 52, no. 5, Oct. 2016.

[5] Y. L. Huang, Y. G. Zhang, N. Li, S. M. Naqvi and J. Chambers, "A robust Student's t based cubature filter," in *19th International Conference on Information Fusion (FUSION)*, July 2016, pp. 9–16.

[6] J. Loxam and T. Drummond, "Student mixture filter for robust, real-time visual tracking," *in Proceedings of 10th European Conference on Computer Vision: Part III*, 2008.

[7] 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.

[8] J. Ting, E. Theodorou, and S. Schaal, "Learning an outlier-robust Kalman filter," in Proceedings of 18th European Conference on Machine Learning Warsaw, Poland, 2007.

[9] H. Zhu, H. Leung, and Z. He, "A variational Bayesian approach to robust sensor fusion based on Student-t distribution," *Information Science*, vol. 221, no. 2013, pp. 201–214, Sep. 2012.

[10] H. Nurminen, T. Ardeshiri, R. Piché, and F. Gustafsson, "Robust inference for state-space models with skewed measurement noise," *IEEE Signal Processing Letters*, vol. 22, no. 11, pp. 2450–2454, Dec. 2015.

[11] Y. L. Huang, Y. G. Zhang, N. Li and J. Chambers, "A robust Gaussian approximate filter for nonlinear systems with heavy-tailed measurement noises," in *2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Mar 2016, pp. 4209–4213.

[12] Y. L. Huang, Y. G. 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 Processing Letters*, vol. 23, no. 4, pp. 468–472, 2016.

[13] B. Kovacevic, Z. Durovic, and S. Glavaski, "On robust Kalman filtering," *International Journal of Control*, vol. 56, no. 3, pp. 547C-562, 1992.

[14] G. Durgaprasad, and S. S. Thakur, "Robust dynamic state estimation of power systems based on M-estimation and realistic modeling of system dynamics," *IEEE Transactions on Power Systems*, vol. 13, no. 4, pp. 1331–1336, 1998.

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

[16] C. D. Karlgaard, and H. Schaub, "Huber-based divided difference filtering," *Journal of Guidance, Control, and Dynamics*, vol. 30, no. 3, pp. 885-C891, 2007.

[17] M. A. Gandhi, and L. Mili, "Robust Kalman filter based on a generalized maximum-likelihood-type estimator," *IEEE Transactions on Signal Processing*, vol. 58, no. 5, pp. 2509–2520, 2010.

[18] G. T. Cinar and J. C. Principe, "Hidden state estimation using the Correntropy Filter with fixed point update and adaptive kernel size," IEEE World Congress on Computational Intelligence, Brisbane, Australia, pp. 1-6, 2012.

[19] B. Chen and J. C. Principe, "Maximum correntropy estimation is a smoothed MAP estimation," *IEEE Signal Processing Letters*, vol. 19, no. 8, pp. 491–494, 2012.

[20] B. Chen, J. Wang, H. Zhao, N. Zheng, and J. C. Principe, "Convergence of a fixed-point algorithm under maximum correntropy criterion," *IEEE Signal Processing Letters*, vol. 22, no. 10, pp. 1723–1727, 2015.

[21] Y. D. Wang, W. Zheng, S. M. Sun, and L. Li, "Robust information filter based on maximum correntropy criterion," *Journal of Guidance, Control, and Dynamics*, pp. 1-6, 2016.

[22] R. Izanloo, S. A. Fakoorian, H. S. Yazdi, and D. Simon, "Kalman filtering based on the maximum correntropy criterion in the presence of non-Gaussian noise," in Proceedings of 50th Annual Conference on Information Science and Systems, 2016.

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

[24] A. O'Hagan, and J. J. Forster, *Kendall's Advanced Theory of Statistics: Bayesian Inference*. Arnold, 2004.

[25] C. M. Bishop, *Pattern Recognition and Machine Learning.* Springer, 2007.

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