Preprints of the 21st IFAC World Congress (Virtual)
Berlin, Germany, July 12-17, 2020

A Novel Robust Kalman Filter With Non-stationary Heavy-tailed Measurement Noise ⋆

Guangle Jia * Yulong Huang *,** Mingming Bai * Yonggang zhang *

* *College of Automation, Harbin Engineering University, Harbin 150001, China (e-mail: jiaguangle2017@163.com; heuedu@163.com; mingmingbai@hrbeu.edu.cn; zhangyg@hrbeu.edu.cn)*
** *Department of Mechanical Engineering, City University of Hong Kong, Hong Kong, 999077, China*

Abstract: A novel robust Kalman filter based on Gaussian-Student's t mixture (GSTM) distribution is proposed to address the filtering problem of a linear system with non-stationary heavy-tailed measurement noise. The mixing probability is recursively estimated by using its previous estimates as prior information, and the state vector, the auxiliary parameter, the Bernoulli random variable and the mixing probability are jointly estimated utilizing the variational Bayesian method. The excellent performance of the proposed robust Kalman filter, compared with the existing state-of-the-art filters, is illustrated by a target tracking simulation results under the case of non-stationary heavy-tailed measurement noise.

Keywords: Robust Kalman filter; Gaussian-Student's t mixture; non-stationary heavy-tailed measurement noise; variational Bayesian

1. INTRODUCTION

The Kalman filter (KF) is a optimal estimator in terms of minimum mean square error (MMSE) under the case of a Gaussian distributed measurement noise. However, the measurement noise may have a non-Gaussian heavy-tailed distribution when measurement is contaminated by outliers from unreliable sensors in many actual application (Roth et al., 2013), (Huang et al., 2016). And then the estimation accuracy of the KF may degrade dramatically when the measurement noise has a non-Gaussian heavy-tailed distribution. Lots of filtering algorithms have been derived to address the scenario with non-Gaussian heavy-tailed measurement noise, such as the Huber-based Kalman filter (HKF) (Gandhi and Mili, 2010), (Durgaprasad and Thakur, 1998), (Karlgaard and Schaub, 2007), the maximum correntropy criterion based Kalman filter (MCCKF) (Cinar and Prncipe, 2012), (Chen and Principe, 2012), (Izanloo et al., 2016), and the robust Student's t-based Kalman filter (Huang et al., 2017a), (Huang et al., 2019a), (Huang et al., 2017b). However, the above filters is no longer suitable for the scenario with non-stationary heavy-tailed measurement noise (NHMN), because these filters are specifically designed for stationary heavy-tailed measurement noise.

Recently, a novel robust Gaussian-Student's t mixture (GSTM) distribution based KF (RGSTMDKF) has been derived for the scenario with non-stationary heavy-tailed process and measurement noises, in which the likelihood probability density function (PDF) is modelled as a weighted sum of a Gaussian distribution and a Student's t distribution under the case of NHMN, and the mixing probability of the measurement likelihood PDF is estimated based on fixed prior information (Huang et al., 2019b). However, the prior for the mixing probability may be slowly time-varying under the case of NHMN. Then, the performance of the existing GSTM based KF may degrade since it employs unreliable fixed prior information. Our idea is that the mixing probability can be recursively estimated by using its previous estimates as prior information, and then the state vector together with the auxiliary parameter, the Bernoulli random variable and the mixing probability can be estimated using the variational Bayesian (VB) method.

In this paper, a novel robust KF is therefore derived for the scenario with NHMN. Firstly, by introducing a Bernoulli random variable, the conditional likelihood PDF can be rewritten as an exponential multiplication for Gaussian PDFs. Secondly, a novel robust KF is derived by utilizing the previous estimates to provide the prior information for the mixing probability, and then the state vector together with the auxiliary parameter, the Bernoulli random variable and the mixing probability are jointly inferred using the variational Bayesian (VB) method. Finally, the excellent performance of the proposed filter, compared with the existing filters, is illustrated in a target tracking simulation of a linear stochastic system with NHMN.

⋆ This work was supported in part by the National Natural Science Foundation of China under Grants 61903097 and 61773133, in part by the Fundamental Research Funds for the Central Universities under Grants 3072019CFJ0411 and GK204026025901. Corresponding author is Y. L. Huang.

2. MAIN RESULTS

Consider the following linear state-space model with NHMN

$$\mathbf{x}_k^n = \mathbf{F}_{k-1}\mathbf{x}_{k-1}^n + \mathbf{w}_{k-1} \tag{1}$$

$$\mathbf{z}_k^n = \mathbf{H}_k\mathbf{x}_k^n + \mathbf{v}_k \tag{2}$$

where $\mathbf{F}_k \in \mathbb{R}^{n\times n}$ denotes the state transition matrix, and $\mathbf{H}_k \in \mathbb{R}^{m\times n}$ denotes the observation matrix, and $\mathbf{x}_k^n \in \mathbb{R}^n$ is the state vector, and $\mathbf{z}_k^n \in \mathbb{R}^m$ is the measurement vector, and $\mathbf{w}_k \in \mathbb{R}^n$ represents the process noise vector that has a Gaussian distribution with zero means and covariance matrix $\mathbf{Q}_k$, and $\mathbf{v}_k \in \mathbb{R}^m$ represents the measurement noise vector that has a non-stationary heavy-tailed distribution. The initial state vector $\mathbf{x}_0^n \in \mathbb{R}^n$ is assumed that it has a Gaussian distribution i. e., $p(\mathbf{x}_0^n) = \mathrm{N}(\mathbf{x}_k^n; \hat{\mathbf{x}}_{0|0}^n, \mathbf{P}_{0|0})$. Moreover, $\mathbf{x}_0^n$, $\mathbf{w}_k$, $\mathbf{v}_k$ are mutually independent in this paper.

In this paper, the PDF of NHMN is modelled as a GSTM distribution formulated as follows (Huang et al., 2019b)

$$p(\mathbf{v}_k) = \tau_k\mathrm{N}(\mathbf{v}_k; \mathbf{0}, \mathbf{R}_k) + (1-\tau_k)\mathrm{St}(\mathbf{v}_k; \mathbf{0}, \mathbf{R}_k, \nu) \tag{3}$$

where $\mathrm{N}(\mathbf{x}^n; \mu, \mathbf{\Sigma})$ is the Gaussian PDF with mean vector μ and covariance matrix $\mathbf{\Sigma}$, and $\mathrm{St}(\mathbf{x}^n; \mu, \mathbf{\Sigma}, \nu)$ denotes the Student's t PDF with mean vector μ, scale matrix $\mathbf{\Sigma}$ and the degrees of freedom (dof) parameter ν, and τ_k denotes the mixing probability at time k.

Using (3), the likelihood PDF conditioned on the mixing probability $p(\mathbf{z}_k^n|\mathbf{x}_k^n, \tau_k)$ can be formulated as

$$\begin{aligned} p(\mathbf{z}_k^n|\mathbf{x}_k^n, \tau_k) &= p_{\mathbf{v}_k}(\mathbf{z}_k^n - \mathbf{H}_k\mathbf{x}_k^n) \\ &= \tau_k\mathrm{N}(\mathbf{z}_k^n - \mathbf{H}_k\mathbf{x}_k^n; \mathbf{0}, \mathbf{R}_k) \\ &\quad +(1-\tau_k)\mathrm{St}(\mathbf{z}_k^n - \mathbf{H}_k\mathbf{x}_k^n; \mathbf{0}, \mathbf{R}_k, \nu) \\ &= \tau_k\mathrm{N}(\mathbf{z}_k^n; \mathbf{H}_k\mathbf{x}_k^n, \mathbf{R}_k) \\ &\quad +(1-\tau_k)\mathrm{St}(\mathbf{z}_k^n; \mathbf{H}_k\mathbf{x}_k^n, \mathbf{R}_k, \nu) \end{aligned} \tag{4}$$

After introducing a Bernoulli random variable ξ_k, the likelihood PDF $p(\mathbf{z}_k^n|\mathbf{x}_k^n, \tau_k)$ can be rewritten as the following form (Huang et al., 2019b)

$$\begin{aligned} p(\mathbf{z}_k^n|\mathbf{x}_k^n, \tau_k) &= \sum_{\xi_k=0}^{1} p(\xi_k|\tau_k)p(\mathbf{z}_k^n|\mathbf{x}_k^n, \xi_k) \\ &= \sum_{\xi_k=0}^{1} \Big\{\tau_k^{\xi_k}[\mathrm{N}(\mathbf{z}_k^n; \mathbf{H}_k\mathbf{x}_k^n, \mathbf{R}_k)]^{\xi_k}(1-\tau_k)^{(1-\xi_k)} \\ &\quad \times [\mathrm{St}(\mathbf{z}_k^n; \mathbf{H}_k\mathbf{x}_k^n, \mathbf{R}_k, \nu)]^{(1-\xi_k)}\Big\} \end{aligned} \tag{5}$$

where the probability mass function (PMF) of the Bernoulli random variable ξ_k and the conditional likelihood PDF $p(\mathbf{z}_k^n|\mathbf{x}_k^n, \xi_k)$ are, respectively, given by

$$p(\xi_k|\tau_k) = \tau_k^{\xi_k}(1-\tau_k)^{(1-\xi_k)} \tag{6}$$

$$\begin{aligned} p(\mathbf{z}_k^n|\mathbf{x}_k^n, \xi_k) &= [\mathrm{N}(\mathbf{z}_k^n; \mathbf{H}_k\mathbf{x}_k^n, \mathbf{R}_k)]^{\xi_k} \\ &\quad \times[\mathrm{St}(\mathbf{z}_k^n; \mathbf{H}_k\mathbf{x}_k^n, \mathbf{R}_k, \nu)]^{(1-\xi_k)} \end{aligned} \tag{7}$$

where the Bernoulli random variable ξ_k takes the value of 1 with probability τ_k.

The mixing probability of the measurement likelihood PDF has been estimated based on a fixed prior information using the VB approach (Huang et al., 2019b). However, the prior information of the mixing probability may be slowly time-varying under the case of NHMN. The estimation accuracy of the existing filter may break down by using the fixed prior information under the case of NHMN. To address scenario with NHMN, the mixing probability is recursively estimated by using its previous estimates as its prior information under the case of NHMN in this paper.

2.1 Choices of prior PDFs

Firstly, in the scenario with NHMN, the one-step prediction PDF $p(\mathbf{x}_k^n|\mathbf{z}_{1:k-1}^n)$ is formulated as

$$p(\mathbf{x}_k^n|\mathbf{z}_{1:k-1}^n) = \mathrm{N}(\mathbf{x}_k^n; \hat{\mathbf{x}}_{k|k-1}^n, \mathbf{P}_{k|k-1}) \tag{8}$$

where $\hat{\mathbf{x}}_{k|k-1}^n$ and $\mathbf{P}_{k|k-1}$ can be derived by the traditional KF.

Secondly, utilizing the property of Student's t distribution, the exponential multiplication PDF $p(\mathbf{z}_k^n|\mathbf{x}_k^n, \xi_k)$ can be rewritten a hierarchical Gaussian form as follows

$$\begin{aligned} p(\mathbf{z}_k^n|\mathbf{x}_k^n, \lambda_k, \xi_k) &= [\mathrm{N}(\mathbf{z}_k^n; \mathbf{H}_k\mathbf{x}_k^n, \mathbf{R}_k)]^{\xi_k} \\ &\quad \times[\mathrm{N}(\mathbf{z}_k^n; \mathbf{H}_k\mathbf{x}_k^n, \mathbf{R}_k/\lambda_k)]^{(1-\xi_k)} \end{aligned} \tag{9}$$

$$p(\lambda_k) = \mathrm{G}(\lambda_k; \frac{\nu_k}{2}, \frac{\nu_k}{2}) \tag{10}$$

Thirdly, according to (6), (8) and (10), the prior PDF $p(\mathbf{x}_k^n, \lambda_k, \xi_k, \tau_k|\mathbf{z}_{1:k-1}^n)$ need to computer from

$$\begin{aligned} p(\mathbf{\Xi}|\mathbf{z}_{1:k-1}^n) &= p(\mathbf{x}_k^n|\mathbf{z}_{1:k-1}^n)p(\lambda_k)p(\xi_k|\tau_k)p(\tau_k|\mathbf{z}_{1:k-1}^n) \\ &= \mathrm{N}(\mathbf{x}_k^n; \hat{\mathbf{x}}_{k|k-1}^n, \mathbf{P}_{k|k-1})\mathrm{G}(\lambda_k; \frac{\nu_k}{2}, \frac{\nu_k}{2})\tau_k^{\xi_k}(1-\tau_k)^{(1-\xi_k)} \\ &\quad \times p(\tau_k|\mathbf{z}_{1:k-1}^n) \end{aligned} \tag{11}$$

$$\mathbf{\Xi} \triangleq \{\mathbf{x}_k^n, \lambda_k, \xi_k, \tau_k\} \tag{12}$$

where $p(\tau_k|\mathbf{z}_{1:k-1}^n)$ describes the prior information for τ_k.

To guarantee that the prior PDF for the mixing probability τ_k has conjugacy, the Beta distribution has been selected as the conjugate prior for the mixing probability τ_k (Huang et al., 2019b). In this paper, the prior distribution $p(\tau_k|\mathbf{z}_{1:k-1}^n)$ is chosen by

$$p(\tau_k|\mathbf{z}_{1:k-1}^n) = \mathrm{Be}(\tau_k; \hat{\alpha}_{k|k-1}, \hat{\beta}_{k|k-1}) \tag{13}$$

where $\hat{\alpha}_{k|k-1}$ and $\hat{\beta}_{k|k-1}$ are, respectively, the prior parameters of the Beta distribution $p(\tau_k|\mathbf{z}_{1:k-1}^n)$. According to Bayes' theorem, the prior PDF $p(\tau_k|\mathbf{z}_{1:k-1}^n)$ is given by

$$p(\tau_k|\mathbf{z}_{1:k-1}^n) = \int p(\tau_k|\tau_{k-1})p(\tau_{k-1}|\mathbf{z}_{1:k-1}^n)\, d\tau_{k-1} \tag{14}$$

where $p(\tau_{k-1}|\mathbf{z}_{1:k-1}^n)$ represents the posterior distribution of τ_{k-1}, and this distribution can be selected as a Beta distribution as follow

$$p(\tau_{k-1}|\mathbf{z}_{1:k-1}^n) = \mathrm{Be}(\tau_{k-1}; \hat{\alpha}_{k-1}, \hat{\beta}_{k-1}) \tag{15}$$

where $\hat{\alpha}_{k-1}$ and $\hat{\beta}_{k-1}$, respectively, represent the posterior parameters of τ_{k-1}. However, the details for spread process of the mixing probability in (14) is unknown, the accurate PDF $p(\tau_k|\mathbf{z}_{1:k-1}^n)$ is unavailable. In this paper, the prior information for the mixing probability is provided by spreading the previous information through a forgetting factor $\rho \in (0, 1]$, and $\hat{\alpha}_{k|k-1}$ and $\hat{\beta}_{k|k-1}$ become

$$\begin{cases} \hat{\alpha}_{k|k-1} = \rho\hat{\alpha}_{k-1} \\ \hat{\beta}_{k|k-1} = \rho\hat{\beta}_{k-1} \end{cases} \tag{16}$$

In this paper, it is assumed that the initial prior distribution for the mixing probability is a Beta distribution, i.e., $p(\tau_0) = \text{Be}(\tau_0; \hat{\alpha}_{0|0}, \hat{\beta}_{0|0})$. Utilizing the property of the Beta distribution, the initial shape parameters $\hat{\alpha}_{0|0}$ and $\hat{\beta}_{0|0}$ need to satisfy the following equation

$$\mathbb{E}[\tau_0] = \frac{\hat{\alpha}_{0|0}}{\hat{\alpha}_{0|0} + \hat{\beta}_{0|0}} = \hat{\tau}_0 \tag{17}$$

where $\hat{\tau}_0$ denotes the initial nominal mixing probability.

2.2 Updating Approximations Posterior Distributions

In this paper, to estimate $\mathbf{x}_k^n$ together with λ_k, ξ_k, and τ_k, the joint posterior distribution $p(\boldsymbol{\Xi}|\mathbf{z}_{1:k}^n)$ needs to be firstly computed. Since the analytical solution for the true joint posterior PDF $p(\boldsymbol{\Xi}|\mathbf{z}_{1:k}^n)$ is not obtainable, the factored approximate posterior PDF $P(\mathbf{x}_k^n)P(\lambda_k)P(\xi_k)P(\tau_k)$ is used to approximate this true joint posterior PDF by employing the VB method, i.e., (Bishop, 2007), (Tzikas et al., 2008)

$$p(\boldsymbol{\Xi}|\mathbf{z}_{1:k}^n) \approx P(\mathbf{x}_k^n)P(\lambda_k)P(\xi_k)P(\tau_k) \tag{18}$$

where the approximate posterior PDFs can be determined by

$$\ln P(\vartheta) = \mathbb{E}_{\boldsymbol{\Xi}^{-\vartheta}}\left[\ln p(\boldsymbol{\Xi}, \mathbf{z}_{1:k}^n)\right] + c_\vartheta \tag{19}$$

where ϑ represents an arbitrary element of $\boldsymbol{\Xi}$, and $\mathbb{E}_{\boldsymbol{\Xi}^{-\vartheta}}[\cdot]$ represents the expectation operation for the remaining elements of $\boldsymbol{\Xi}$ except for ϑ, and c_ϑ represents the constant with respect to ϑ. To solve $P(\mathbf{x}_k^n)$, $P(\lambda_k)$, $P(\xi_k)$ and $P(\tau_k)$, we need to use the fixed-point iterative method, in which only one element of $\boldsymbol{\Xi}$ is updated while others keeping fixed (Bishop, 2007), (Tzikas et al., 2008).

Exploiting (9), (11) and (13), the PDF $p(\boldsymbol{\Xi}, \mathbf{z}_{1:k}^n)$ is given by

$$\begin{aligned} &p(\boldsymbol{\Xi}, \mathbf{z}_{1:k}^n) = p(\mathbf{z}_k^n|\mathbf{x}_k^n, \lambda_k, \xi_k)p(\mathbf{x}_k^n|\mathbf{z}_{1:k-1}^n)p(\lambda_k)p(\xi_k|\tau_k) \\ &\times p(\tau_k|\mathbf{z}_{1:k-1}^n)p(\mathbf{z}_{1:k-1}^n) \\ &= \text{N}(\mathbf{x}_k^n; \hat{\mathbf{x}}_{k|k-1}^n, \mathbf{P}_{k|k-1})[\text{N}(\mathbf{z}_k^n; \mathbf{H}_k\mathbf{x}_k^n, \mathbf{R}_k)]^{\xi_k} \\ &\times[\text{N}(\mathbf{z}_k^n; \mathbf{H}_k\mathbf{x}_k^n, \mathbf{R}_k/\lambda_k)]^{(1-\xi_k)}\text{G}(\lambda_k; \frac{\nu_k}{2}, \frac{\nu_k}{2})\tau_k^{\xi_k} \\ &\times(1-\tau_k)^{(1-\xi_k)}\text{Be}(\tau_k; \hat{\alpha}_{k|k-1}, \hat{\beta}_{k|k-1})p(\mathbf{z}_{1:k-1}^n) \end{aligned} \tag{20}$$

Using (20), $\ln p(\boldsymbol{\Xi}, \mathbf{z}_{1:k}^n)$ can be formulated as

$$\begin{aligned} &\ln p(\boldsymbol{\Xi}, \mathbf{z}_{1:k}^n) = -0.5[\xi_k + \lambda_k(1-\xi_k)](\mathbf{z}_k^n - \mathbf{H}_k\mathbf{x}_k^n)^T\mathbf{R}_k^{-1} \\ &\times(\mathbf{z}_k^n - \mathbf{H}_k\mathbf{x}_k^n) - 0.5(\mathbf{x}_k^n - \hat{\mathbf{x}}_{k|k-1}^n)^T\mathbf{P}_{k|k-1}^{-1}(\mathbf{x}_k^n - \hat{\mathbf{x}}_{k|k-1}^n) \\ &+[0.5m(1-\xi_k) + 0.5\nu_k - 1]\ln\lambda_k - 0.5\nu_k\lambda_k + (\xi_k \\ &+\hat{\alpha}_{k|k-1} - 1\big)\ln\tau_k + (\hat{\beta}_{k|k-1} - \xi_k)\ln(1-\tau_k) + c_{\boldsymbol{\Xi}} \end{aligned} \tag{21}$$

where $c_{\boldsymbol{\Xi}}$ denotes a constant with respect to any element of $\boldsymbol{\Xi}$.

Let $\vartheta = \mathbf{x}_k^n$ and substituting (21) into (19), then, we get

$$\begin{aligned} &\ln P^{(l+1)}(\mathbf{x}_k^n) = -0.5(\mathbb{E}^{(l)}[\xi_k] + \mathbb{E}^{(l)}[\lambda_k]\mathbb{E}^{(l)}[1-\xi_k]) \\ &\times(\mathbf{z}_k^n - \mathbf{H}_k\mathbf{x}_k^n)^T\mathbf{R}_k^{-1}(\mathbf{z}_k^n - \mathbf{H}_k\mathbf{x}_k^n) - 0.5(\mathbf{x}_k^n - \hat{\mathbf{x}}_{k|k-1}^n)^T \\ &\times\mathbf{P}_{k|k-1}^{-1}(\mathbf{x}_k^n - \hat{\mathbf{x}}_{k|k-1}^n) + c_{\mathbf{x}^n} \end{aligned} \tag{22}$$

where $P^{(l)}(\cdot)$ and $\mathbb{E}^{(l)}[\vartheta]$, which are, the approximated distribution for $P(\cdot)$ and the expectation operation for ϑ at the l-th iteration, respectively.

The correctional measurement noise covariance matrix $\hat{\mathbf{R}}_k^{(l+1)}$ is formulated as

$$\hat{\mathbf{R}}_k^{(l+1)} = \frac{\mathbf{R}_k}{\mathbb{E}^{(l)}[\xi_k] + \mathbb{E}^{(l)}[\lambda_k]\mathbb{E}^{(l)}[1-\xi_k]} \tag{23}$$

where the expectations $\mathbb{E}^{(l)}[\lambda_k]$, $\mathbb{E}^{(l)}[\xi_k]$ and $\mathbb{E}^{(l)}[1-\xi_k]$ can be, respectively, obtained through (40) and (42)-(43) at the l-th iteration.

Utilizing (23), $P^{(l+1)}(\mathbf{x}_k^n)$ can be updated as a Gaussian distribution, i.e.,

$$P^{(l+1)}(\mathbf{x}_k^n) = \text{N}(\mathbf{x}_k^n; \hat{\mathbf{x}}_{k|k}^{n(l+1)}, \mathbf{P}_{k|k}^{(l+1)}) \tag{24}$$

where $\hat{\mathbf{x}}_{k|k}^{n(l+1)}$ and $\mathbf{P}_{k|k}^{(l+1)}$, which are, respectively, calculated by the following equations

$$\mathbf{K}_k^{h(l+1)} = \mathbf{P}_{k|k-1}\mathbf{H}_k^T\left(\mathbf{H}_k\mathbf{P}_{k|k-1}\mathbf{H}_k^T + \hat{\mathbf{R}}_k^{(l+1)}\right)^{-1} \tag{25}$$

$$\hat{\mathbf{x}}_{k|k}^{n(l+1)} = \hat{\mathbf{x}}_{k|k-1}^n + \mathbf{K}_k^{h(l+1)}\left(\mathbf{z}_k^n - \mathbf{H}_k\hat{\mathbf{x}}_{k|k-1}^n\right) \tag{26}$$

$$\mathbf{P}_{k|k}^{(l+1)} = \hat{\mathbf{P}}_{k|k-1}^{(l+1)} - \mathbf{K}_k^{h(l+1)}\mathbf{H}_k\mathbf{P}_{k|k-1} \tag{27}$$

Let $\vartheta = \lambda_k$ and substituting (21) into (19), then, we get

$$\begin{aligned} &\ln P^{(l+1)}(\lambda_k) = -0.5(\mathbb{E}^{(l)}[1-\xi_k]\text{tr}(\mathbf{B}_k^{(l+1)}\mathbf{R}_k^{-1}))\lambda_k \\ &+(0.5m\mathbb{E}^{(l)}[1-\xi_k] + 0.5\nu_k - 1)\ln\lambda_k + c_\lambda \end{aligned} \tag{28}$$

where $\mathbf{B}_k^{(l+1)}$ can be obtained by

$$\begin{aligned} &\mathbf{B}_k^{(l+1)} = \mathbb{E}^{(l+1)}[(\mathbf{z}_k^n - \mathbf{H}_k\mathbf{x}_k^n)(\mathbf{z}_k^n - \mathbf{H}_k\mathbf{x}_k^n)^T|\mathbf{z}_{1:k-1}^n] \\ &= (\mathbf{z}_k^n - \mathbf{H}_k\hat{\mathbf{x}}_{k|k}^{n(l+1)})(\mathbf{z}_k^n - \mathbf{H}_k\hat{\mathbf{x}}_{k|k}^{n(l+1)})^T + \mathbf{H}_k\mathbf{P}_{k|k}^{(l+1)}\mathbf{H}_k^T \end{aligned} \tag{29}$$

Utilizing (28), $P^{(l+1)}(\lambda_k)$ can be updated as a Gamma distribution as follows

$$P^{(l+1)}(\lambda_k) = \text{G}(\lambda_k; \hat{\phi}_k^{(l+1)}, \hat{\varphi}_k^{(l+1)}) \tag{30}$$

where the shape parameters $\hat{\phi}_k^{(l+1)}$ and $\hat{\varphi}_k^{(l+1)}$ can be, respectively, obtained by

$$\hat{\phi}_k^{(l+1)} = 0.5m\mathbb{E}^{(l)}[1-\xi_k] + 0.5\nu_k \tag{31}$$

$$\hat{\varphi}_k^{(l+1)} = 0.5\mathbb{E}^{(l)}[1-\xi_k]\text{tr}(\mathbf{B}_k^{(l+1)}\mathbf{R}_k^{-1}) + 0.5\nu_k \tag{32}$$

Let $\vartheta = \xi_k$ and substituting (21) into (19), then, we get

$$\begin{aligned} &\ln P^{(l+1)}(\xi_k) = -0.5(\text{tr}(\mathbf{B}_k^{(l+1)}\mathbf{R}_k^{-1}) + \mathbb{E}^{(l)}[\ln\tau_k])\xi_k \\ &+(0.5m\mathbb{E}^{(l+1)}[\ln\lambda_k] - 0.5\mathbb{E}^{(l+1)}[\lambda_k]\text{tr}(\mathbf{B}_k^{(l+1)}\mathbf{R}_k^{-1}) \\ &+\mathbb{E}^{(l)}[\ln(1-\tau_k)]\Big)(1-\xi_k) + c_\xi \end{aligned} \tag{33}$$

where $\mathbb{E}^{(l)}[\ln\tau_k]$ and $\mathbb{E}^{(l)}[\ln(1-\tau_k)]$ are, respectively, calculated through (44) and (45) at the l-th iteration, and the expectations $\mathbb{E}^{(l+1)}[\lambda_k]$ and $\mathbb{E}^{(l+1)}[\ln\lambda_k]$ can be, respectively, obtained through (40) and (41) at the $(l+1)$-th iteration.

Exploiting (33), $P^{(l+1)}(\xi_k)$ is updated as a Bernoulli distribution, then the PDFs $\Pr^{(l+1)}(\xi_k=1)$ and $\Pr^{(l+1)}(\xi_k=0)$ are, respectively, calculated by the following equations

$$\Pr^{(l+1)}(\xi_k=1)=\Lambda^{(l+1)}\exp\Big\{\mathrm{tr}(\mathbf{B}_k^{(l+1)}\mathbf{R}_k^{-1}) +\mathbb{E}^{(l)}[\ln\tau_k]\Big\} \tag{34}$$

$$\Pr^{(l+1)}(\xi_k=0)=\Lambda^{(l+1)}\exp\Big\{0.5m\mathbb{E}^{(l+1)}[\ln\lambda_k] -0.5\mathbb{E}^{(l+1)}[\lambda_k]\mathrm{tr}(\mathbf{B}_k^{(l+1)}\mathbf{R}_k^{-1})+\mathbb{E}^{(l)}[\ln(1-\tau_k)]\Big\} \tag{35}$$

where $\Lambda^{(l+1)}$ represents a constant with respect to the Bernoulli random variable ξ_k.

Let $\vartheta=\tau_k$ and substituting (21) into (19), then, we get

$$\ln P^{(l+1)}(\tau_k)=(\hat{\alpha}_{k|k-1}+\mathbb{E}^{(l+1)}[\xi_k]-1)\ln\tau_k +(\hat{\beta}_{k|k-1}+\mathbb{E}^{(l+1)}[1-\xi_k]-1)\ln(1-\tau_k)+c_\tau \tag{36}$$

where $\mathbb{E}^{(l+1)}[\xi_k]$ and $\mathbb{E}^{(l+1)}[1-\xi_k]$ can be, respectively, obtained through (42) and (43) at the $(l+1)$-th iteration.

Utilizing (36), $P^{(l+1)}(\tau_k)$ can be updated as a Beta distribution, i.e.,

$$P^{(l+1)}(\tau_k)=\mathrm{Be}(\tau_k;\hat{\alpha}_k^{(l+1)},\hat{\beta}_k^{(l+1)}) \tag{37}$$

where the shape parameters $\hat{\alpha}_k^{(l+1)}$ and $\hat{\beta}_k^{(l+1)}$ are, respectively, obtained by

$$\hat{\alpha}_k^{(l+1)}=\hat{\alpha}_{k|k-1}+\mathbb{E}^{(l+1)}[\xi_k] \tag{38}$$

$$\hat{\beta}_k^{(l+1)}=\hat{\beta}_{k|k-1}+\mathbb{E}^{(l+1)}[1-\xi_k] \tag{39}$$

The expectations $\mathbb{E}^{(l+1)}[\lambda_k]$, $\mathbb{E}^{(l+1)}[\ln\lambda_k]$, $\mathbb{E}^{(l+1)}[\xi_k]$, $\mathbb{E}^{(l+1)}[1-\xi_k]$, $\mathbb{E}^{(l+1)}[\ln\tau_k]$ and $\mathbb{E}^{(l+1)}[\ln(1-\tau_k)]$ can be, respectively, formulated as follows

$$\mathbb{E}^{(l+1)}[\lambda_k]=\frac{\hat{\phi}_k^{(l+1)}}{\hat{\varphi}_k^{(l+1)}} \tag{40}$$

$$\mathbb{E}^{(l+1)}[\ln\lambda_k]=\psi(\hat{\phi}_k^{(l+1)})-\ln\hat{\varphi}_k^{(l+1)} \tag{41}$$

$$\mathbb{E}^{(l+1)}[\xi_k]=\frac{\Pr^{(l+1)}(\xi_k=1)}{\Pr^{(l+1)}(\xi_k=1)+\Pr^{(l+1)}(\xi_k=0)} \tag{42}$$

$$\mathbb{E}^{(l+1)}[1-\xi_k]=1-\mathbb{E}^{(l+1)}[\xi_k] \tag{43}$$

$$\mathbb{E}^{(l+1)}[\ln\tau_k]=\psi(\alpha_k^{(l+1)})-\psi(\alpha_k^{(l+1)}+\beta_k^{(l+1)}) \tag{44}$$

$$\mathbb{E}^{(l+1)}[\ln(1-\tau_k)]=\psi(\beta_k^{(l+1)})-\psi(\alpha_k^{(l+1)}+\beta_k^{(l+1)}) \tag{45}$$

where ψ denotes a digamma function in (Zhu et al., 2013). The proposed filter consists of (16), (23), (25)-(27), (29), (31)-(32), (34)-(35) and (38)-(45), and its implementation details are shown in Algorithm 1.

Algorithm 1: The proposed filtering algorithm with NHMN.

Inputs: $\hat{\mathbf{x}}_{k-1|k-1}^n$, $\mathbf{P}_{k-1|k-1}$, $\mathbf{F}_{k-1}$, $\mathbf{H}_k$, $\mathbf{z}_k^n$, $\mathbf{Q}_{k-1}$, $\mathbf{R}_k$, $\hat{\alpha}_{k-1}$, $\hat{\beta}_{k-1}$, ν_k, n, m, ρ, δ, N_m

Time update:

1. $\hat{\mathbf{x}}_{k|k-1}^n$ and $\mathbf{P}_{k|k-1}$ can be derived by the traditional KF,
2. The prior shape parameters $\hat{\alpha}_{k|k-1}$ and $\hat{\beta}_{k|k-1}$ are obtained by (16),

Measurement update:

3. Initialization: $\hat{\mathbf{x}}_{k|k}^{(0)}=\hat{\mathbf{x}}_{k|k-1}^n$, $\mathbf{P}_{k|k}^{(0)}=\mathbf{P}_{k|k-1}$, $\mathbb{E}^{(0)}[\lambda_k]=1$,
$\mathbb{E}^{(0)}[\ln\lambda_k]=0$, $\hat{\alpha}_k^{(0)}=\hat{\alpha}_{k|k-1}$, $\hat{\beta}_k^{(0)}=\hat{\beta}_{k|k-1}$,
$\mathbb{E}^{(0)}[\xi_k]=\hat{\alpha}_k^{(0)}/(\hat{\alpha}_k^{(0)}+\hat{\beta}_k^{(0)})$,
$\mathbb{E}^{(0)}[\ln\tau_k]=\psi(\hat{\alpha}_k^{(0)})-\psi(\hat{\alpha}_k^{(0)}+\hat{\beta}_k^{(0)})$,
$\mathbb{E}^{(0)}[\ln(1-\tau_k)]=\psi(\hat{\beta}_k^{(0)})-\psi(\hat{\alpha}_k^{(0)}+\hat{\beta}_k^{(0)})$,

for $l=0:N_m-1$

Update $P^{(l+1)}(\mathbf{x}_k^n)=\mathrm{N}(\mathbf{x}_k^n;\hat{\mathbf{x}}_{k|k}^{n(l+1)},\mathbf{P}_{k|k}^{(l+1)})$ given $\mathbb{E}^{(l)}[\lambda_k]$, $\mathbb{E}^{(l)}[\xi_k]$ and $\mathbb{E}^{(l)}[1-\xi_k]$:

4. Calculate $\hat{\mathbf{x}}_{k|k}^{n(l+1)}$ and $\mathbf{P}_{k|k}^{(l+1)}$ using (23) and (25)-(27)

Update $P^{(l+1)}(\lambda_k)=\mathrm{G}(\lambda_k;\hat{\phi}_k^{(l+1)},\hat{\varphi}_k^{(l+1)})$ given $P^{(l+1)}(\mathbf{x}_k^n)$ and $\mathbb{E}^{(l)}[1-\xi_k]$:

5. Calculate $\hat{\phi}_k^{(l+1)}$ and $\hat{\varphi}_k^{(l+1)}$ using (29) and (31)-(32)

Update $P^{(l+1)}(\xi_k)$ as a Bernoulli distribution given $P^{(l+1)}(\mathbf{x}_k^n)$, $P^{(l+1)}(\lambda_k)$, $\mathbb{E}^{(l)}[\ln\tau_k]$ and $\mathbb{E}^{(l)}[\ln(1-\tau_k)]$:

6. Calculate $\Pr^{(l+1)}(\xi_k=1)$ and $\Pr^{(l+1)}(\xi_k=0)$ using (29), (34)-(35), and (40)-(41)

Update $P^{(l+1)}(\tau_k)=\mathrm{Be}(\tau_k;\hat{\alpha}_k^{(l+1)},\hat{\beta}_k^{(l+1)})$ given $P^{(l+1)}(\xi_k)$:

7. Calculate $\hat{\alpha}_k^{(l+1)}$ and $\hat{\beta}_k^{(l+1)}$ using (38)-(39) and (42)-(43),
8. Calculate the expectations $\mathbb{E}^{(l+1)}[\lambda_k]$, $\mathbb{E}^{(l+1)}[\ln\lambda_k]$, $\mathbb{E}^{(l+1)}[\xi_k]$, $\mathbb{E}^{(l+1)}[1-\xi_k]$, $\mathbb{E}^{(l+1)}[\ln\tau_k]$ and $\mathbb{E}^{(l+1)}[\ln(1-\tau_k)]$ using (40)-(45)
9. If $\|\mathbf{x}_{k|k}^{n(l+1)}-\mathbf{x}_{k|k}^{n(l)}\|/\|\mathbf{x}_{k|k}^{n(l)}\|\leq\delta$, stop iteration,

end for

$\hat{\mathbf{x}}_{k|k}^n=\hat{\mathbf{x}}_{k|k}^{n(l)}$, $\mathbf{P}_{k|k}=\mathbf{P}_{k|k}^{(l)}$, $\hat{\alpha}_k=\hat{\alpha}_k^{(l)}$, $\hat{\beta}_k=\hat{\beta}_k^{(l)}$,

Outputs: $\hat{\mathbf{x}}_{k|k}^n$, $\mathbf{P}_{k|k}$, $\hat{\alpha}_k$, $\hat{\beta}_k$

3. SIMULATIONS

The excellent performance of the proposed robust KF, compared with the existing state-of-the-art filtersis illustrated in a target tracking simulation of a linear stochastic system with NHMN. The linear stochastic system is given by

$$\mathbf{x}_k^n=\begin{bmatrix}\mathbf{I}_2 & \Delta t\mathbf{I}_2\\ \mathbf{0} & \mathbf{I}_2\end{bmatrix}\mathbf{x}_{k-1}^n+\mathbf{w}_{k-1} \tag{46}$$

$$\mathbf{z}_k^n=[\,\mathbf{I}_2\quad\mathbf{0}\,]\,\mathbf{x}_k^n+\mathbf{v}_k \tag{47}$$

where the state vector is $\mathbf{x}_k^n=[\iota_k\quad\kappa_k\quad\dot{\iota}_k\quad\dot{\kappa}_k]^T$, ι_k and κ_k, respectively, represent the positions in the X axial and Y axial of the cartesian coordinates, and $\dot{\iota}_k$ and $\dot{\kappa}_k$, respectively, represent the corresponding velocities (Huang et al., 2017a). The process noise vector $\mathbf{w}_k$ has a Gaussian distribution with zero mean vector and covariance matrix $\mathbf{Q}_k$ formulated as follow

$$\mathbf{Q}_k=q\begin{bmatrix}\frac{\Delta t^3}{3}\mathbf{I}_2 & \frac{\Delta t^2}{2}\mathbf{I}_2\\ \frac{\Delta t^2}{2}\mathbf{I}_2 & \Delta t\mathbf{I}_2\end{bmatrix} \tag{48}$$

where $\Delta t=1$s denotes the sampling interval.

Similar to (Huang et al., 2019b), the NHMN $\mathbf{v}_k$ is simulated according to

$$\begin{cases} \mathbf{v}_k \sim \mathrm{N}(\mathbf{0}, \mathbf{R}_k), & 1 \leq k \leq 100 \\ \mathbf{v}_k \sim \begin{cases} \mathrm{N}(\mathbf{0}, \mathbf{R}_k) & \text{w.p. } 0.99 \\ \mathrm{N}(\mathbf{0}, 100\mathbf{R}_k) & \text{w.p. } 0.01 \end{cases}, & 101 \leq k \leq 200 \\ \mathbf{v}_k \sim \begin{cases} \mathrm{N}(\mathbf{0}, \mathbf{R}_k) & \text{w.p. } 0.95 \\ \mathrm{N}(\mathbf{0}, 100\mathbf{R}_k) & \text{w.p. } 0.05 \end{cases}, & 201 \leq k \leq 300 \\ \mathbf{v}_k \sim \mathrm{N}(\mathbf{0}, \mathbf{R}_k), & 301 \leq k \leq 400 \end{cases} \tag{49}$$

where the covariance matrix $\mathbf{R}_k$ is formulated as follows

$$\mathbf{R}_k = \begin{bmatrix} r & 0 \\ 0 & r \end{bmatrix} \tag{50}$$

In this simulation, the KF with the nominal covariance matrices (KFNCM), the KF with the true instantaneous covariance matrices (KFTICM), the HKF (Karlgaard and Schaub, 2007), the RSTKF with the true one-step prediction error covariance matrix (Huang et al., 2017b), the RGSTMDKF with the true one-step prediction error covariance matrix (Huang et al., 2019b) and the proposed filter are compered. Note that the true instantaneous covariance matrices cannot be obtained in practical engineering applications, resluting that the KFTICM cannot be applied in above applications. The parameters are set as $q = 1\mathrm{m}^2/\mathrm{s}^3$, $r = 100\mathrm{m}^2$ in this simulation, and the total of simulation time is set as $T = 400s$. The filtering parameters in this simulation are given in Table I. To test the estimation accuracy of the different filters, the root mean square errors (RMSEs) and averaged RMSEs (ARMSEs) of position and velocity are selected as performance measures. As an example, the RMSE and ARMSE of position is defined as (Huang et al., 2017a):

$$\begin{cases} \mathrm{RMSE}_{\mathrm{pos}} \triangleq \sqrt{\dfrac{1}{M}\sum_{s=1}^{M}\left((a_k^i - \hat{a}_k^i)^2 + (b_k^i - \hat{b}_k^i)^2\right)} \\ \mathrm{ARMSE}_{\mathrm{pos}} \triangleq \sqrt{\dfrac{1}{MT}\sum_{k=1}^{T}\sum_{i=1}^{M}\left((a_k^i - \hat{a}_k^i)^2 + (b_k^i - \hat{b}_k^i)^2\right)} \end{cases} \tag{51}$$

where (a_k^i, b_k^i) and $(\hat{a}_k^i, \hat{b}_k^i)$ are, respectively, the true and estimated positions at the i-th Monte Carlo run. Similarly, the RMSE and ARMSE of velocity can be also defined.

Table 1. The filtering parameters

Index	Parameters
Initial prior parameter α_0	5
Initial prior parameter β_0	5
Dof parameter ν_k	5
The forgetting factor ρ	0.99
The maximum number of iteration N_m	50
Determine convergence parameter δ	10^{-16}

Fig. 1 shows the RMSEs of position and velocity from different filters over 1000 Monte Carlo runs. Since KF is an optimal state estimator in terms of MMSE for the linear stochastic system with a Gaussian distributed measurement noise, the performance of the KFTICM is optimal in this simulation. Note that the RMSEs of RSTKF is not a constant bias compared with the RMSEs of the KFTICM. From these figures, the proposed filter is closer to the KFTICM than the existing filters, which is illustrated that the proposed robust filter has better

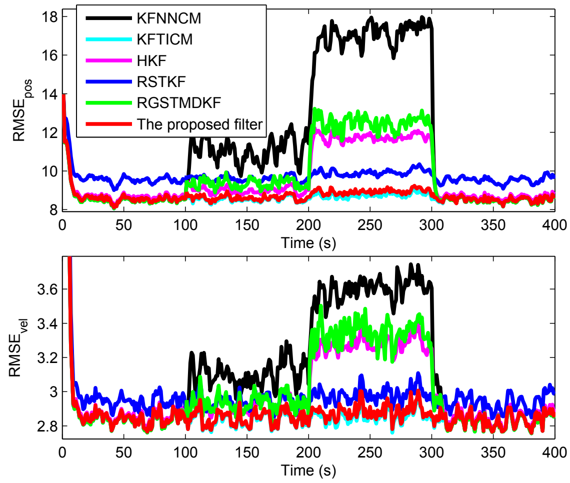

Fig. 1. $\mathrm{RMSE}_{\mathrm{pos}}$ and $\mathrm{RMSE}_{\mathrm{vel}}$ with different filters.

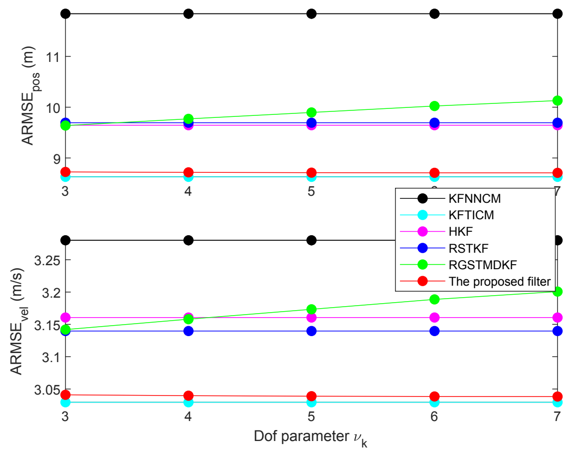

Fig. 2. $\mathrm{ARMSE}_{\mathrm{pos}}$ and $\mathrm{ARMSE}_{\mathrm{vel}}$ from different filters with different dof parameters.

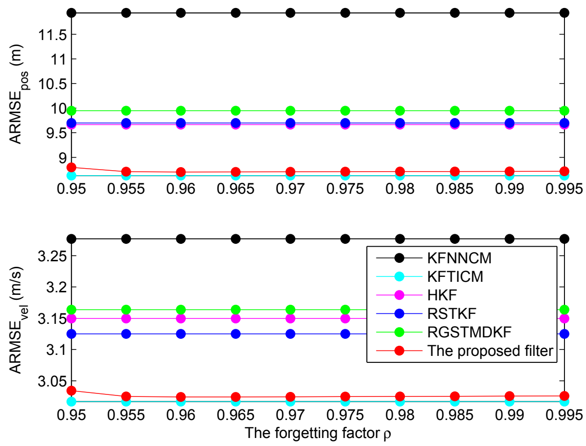

Fig. 3. $\mathrm{ARMSE}_{\mathrm{pos}}$ and $\mathrm{ARMSE}_{\mathrm{vel}}$ from different filters with different forgetting factors.

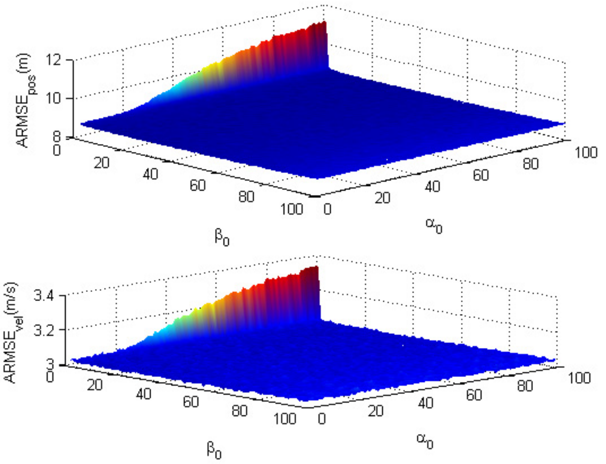

Fig. 4. ARMSE$_{pos}$ and ARMSE$_{vel}$ from different filters with different initial prior parameters.

performance than the existing filters in the scenario with NHMN.

The ARMSE$_{pos}$ and ARMSE$_{vel}$ from different filters with different dof parameters over 1000 Monte Carlo runs are shown in Fig. 2. From Fig. 2, the proposed filter with different dof parameters has the essentially consistent estimation error and is smaller than the existing filters, which is illustrated that the proposed filter with different dof parameters has better estimation performance than the existing filters in the scenario with NHMN.

The forgetting factor is selected as $\rho = 0.95 : 0.005 : 0.99$, and the values of other parameters are the same as those in Table I. Fig. 3 shows the ARMSE$_{pos}$ and ARMSE$_{vel}$ from different filters with different forgetting factors over 1000 Monte Carlo runs. As shown in Fig. 3, the proposed filter has smaller ARMSEs than the existing filters, and the ARMSEs of the proposed filter almost remain the same when the forgetting factor is selected as $\rho = 0.95 : 0.005 : 0.99$ in the scenario with NHMN.

The ARMSE$_{pos}$ and ARMSE$_{vel}$ from the proposed filter with initial prior parameters $(\alpha_0, \beta_0) \in [1, 50] \times [1, 50]$ over 1000 Monte Carlo runs are shown in Fig. 4. From Fig. 5, the estimation error from the proposed filter with the initial prior parameter $\beta_0 \in [1, 5]$ may increase. This is owing to the estimation of the mixing probability tending to be the value of 1, which is no longer suitable for the scenario with NHMN, under the case of the initial prior parameter β_0 selected as a very small value. Thus, in practical application, it is suggest that selecting a large value for β_0 to guarantee the proposed filter exhibits a good performance, such as the proposed filter with initial prior parameters $(\alpha_0, \beta_0) \in [1, 50] \times [5, 50]$ exhibiting a good estimation accuracy as shown in Fig. 4.

4. CONCLUSION

In this paper, a novel robust KF was derived for the linear system with NHMN, where the mixing probability is recursively estimated by using its previous estimates as prior information. Compared with the existing filters, the excellent performance of the proposed robust KF is illustrated in the target tracking simulation results under the case of NHMN.

REFERENCES

Bishop, C.M. (2007). A variational bayesian approach to robust sensor fusion based on student's t distribution. *Patt. Recogni. and Machi. Learni.*

Chen, B. and Principe, J.C. (2012). Maximum correntropy estimation is a smoothed map estimation. *IEEE Signal Processing Letters*, 19(8), 491–494.

Cinar, G.T. and Prncipe, J.C. (2012). Hidden state estimation using the correntropy filter with fixed point update and adaptive kernel size. *The 2012 International Joint Conference on Neural Networks (IJCNN)*, 1–6.

Durgaprasad, G. and Thakur, S.S. (1998). Robust dynamic state estimation of power systems based on m-estimation and realistic modeling of system dynamics. *IEEE Transactions on Power Systems*, 13(4), 1331–1336.

Gandhi, M.A. and Mili, L. (2010). Robust kalman filter based on a generalized maximum-likelihood-type estimator. *IEEE Transactions on Signal Processing*, 58(5), 2509–2520.

Huang, Y., Zhang, Y., Li, N., Mohsen.Naqvi, S., and Chambers, J. (2016). A robust student's t based cubature filter. *2016 19th International Conference on Information Fusion (FUSION)*, 9–16.

Huang, Y., Zhang, Y., Li, N., Wu, Z., and Chambers, J.A. (2017a). A novel robust student's t-based kalman filter. *IEEE Transactions on Aerospace and Electronic Systems*, 53(3), 1545–1554.

Huang, Y., Zhang, Y., Shi, P., Wu, Z., Qian, J., and Chambers, J.A. (2019a). Robust kalman filters based on gaussian scale mixture distributions with application to target tracking. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, 49(10), 2082–2096.

Huang, Y., Zhang, Y., Xu, B., Wu, Z., and Chambers, J.A. (2017b). A new outlier-robust studnet's t based gaussian approxiamte filter for cooperative localization. *IEEE/ASME Trans. on Mech.*, 22(5), 2380–2386.

Huang, Y., Zhang, Y., Zhao, Y., and Chambers, J.A. (2019b). A novel robust gaussianstudent's t mixture distribution based kalman filter. *IEEE Transactions on Signal Processing*, 67(13), 3606–3620.

Izanloo, R., Fakoorian, S.A., Yazdi, H.S., and Simon, D. (2016). Kalman filtering based on the maximum correntropy criterion in the presence of non-gaussian noise. *2016 Annual Conference on Information Science and Systems (CISS)*, 500–505.

Karlgaard, C.D. and Schaub, H. (2007). Huber-based divided difference filtering. *Journal of Guidance, Control, and Dynamics*, 30(3), 885–891.

Roth, M., zkan, E., and Gustafsson, F. (2013). A student's t filter for heavy tailed process and measurement noise. *2013 IEEE International Conference on Acoustics, Speech and Signal Processing*, 5770–5774.

Tzikas, D.G., Likas, A.C., and Galatsanos, N.P. (2008). The variational approximation for bayesian inference. *IEEE Signal Processing Magazine*, 25(6), 131–146.

Zhu, H., Leung, H., and He, Z. (2013). A variational bayesian approach to robust sensor fusion based on student's t distribution. *Inf. Sci.*, 221, 201–214.