



ARTICLE

Skew t Distribution-Based Nonlinear Filter with Asymmetric Measurement Noise Using Variational Bayesian Inference

Chen Xu¹, Yawen Mao², Hongtian Chen^{3,*}, Hongfeng Tao¹ and Fei Liu¹

¹Key Laboratory of Advanced Process Control for Light Industry (Ministry of Education), Jiangnan University, Wuxi, 214122, China

²School of Science, Jiangnan University, Wuxi, 214122, China

³Department of Chemical and Materials Engineering, University of Alberta, Edmonton, AB T6G 2G6, Canada

*Corresponding Author: Hongtian Chen. Email: chtbaylor@163.com

Received: 30 August 2021 Accepted: 20 October 2021

ABSTRACT

This paper is focused on the state estimation problem for nonlinear systems with unknown statistics of measurement noise. Based on the cubature Kalman filter, we propose a new nonlinear filtering algorithm that employs a skew t distribution to characterize the asymmetry of the measurement noise. The system states and the statistics of skew t noise distribution, including the shape matrix, the scale matrix, and the degree of freedom (DOF) are estimated jointly by employing variational Bayesian (VB) inference. The proposed method is validated in a target tracking example. Results of the simulation indicate that the proposed nonlinear filter can perform satisfactorily in the presence of unknown statistics of measurement noise and outperform than the existing state-of-the-art nonlinear filters.

KEYWORDS

Nonlinear filter; asymmetric measurement noise; skew t distribution; unknown noise statistics; variational Bayesian inference

1 Introduction

State estimation serves as an important role in various fields, such as control, signal processing, fault detection and diagnosis, and many more [1–7]. Due to its effectiveness and optimality, the Kalman filter (KF) is the state estimation method of the most widespread used for linear systems with Gaussian noise distribution [8–10]. Limited by the assumption of linear, many nonlinear filters have been presented [11–13], the most famous of which is the extended Kalman filter (EKF) [14]. To solve the error caused by linearization in EKF, the cubature Kalman filter (CKF) and unscented Kalman filter (UKF) were developed by using sigma points to approximate the posterior distribution [15,16]. Among them, CKF is proven to have better estimation performance in high-dimensional nonlinear estimation. However, the above state estimation methods assume Gaussian noise distribution and their noise statistics are completely known, which is not available in practice.



To deal with the unknown noise statistics, various adaptive and robust filters were designed for joint state estimation [17–19]. For example, a recursive state estimation method was presented with unknown Gaussian noise covariance for linear systems [20]. Further, an adaptive variational Bayesian (VB)-based filter was designed for estimating the covariance of process noise and measurement noise by selecting inverse Wishart priors [21]. Combining with maximum correntropy criterion, an adaptive and robust filter was developed by estimating the Gaussian measurement noise covariance [22]. However, the above Gaussian-based estimation methods are unsuitable for the heavy-tailed noise which caused by outliers or impulse interferences. In the case where both the measurement noise and the process noise are Student's t distributed noise, the Student's t filter was first proposed in [23]. By minimizing the Kullback-Leibler divergence, an adaptive t -filter was developed to estimate the scale matrix of Student's t distribution [24]. For nonlinear system, a recursive outlier-robust nonlinear filter was proposed for Student's t distributed noise in [25] and a robust Gaussian approximate filter was presented with unknown statistics of Student's t noise distribution in [26].

Due to the complex environment, not only the noise distribution with heavy-tailed characteristics but also the asymmetry of noise distribution should be considered. As shown in Fig. 1, skew t distribution obtains better fitting performance than the Gaussian distribution and Student's t distribution, which are symmetric distributions. Thus, several estimation methods were presented for the skew t distribution, which has both skewness and heavy-tails [27–29]. For example, a skew t variational Bayesian filter was designed for measurement noise with heavy-tails and skewness in [30] and the estimation accuracy was improved by covariance matrix approximation in [31]. In [32], a robust filter was developed to estimate the skew t distribution, consisting of the scale matrix and degree of freedom (DOF). Moreover, some other filtering algorithms that can describe asymmetric noise distribution also have been proposed in [33,34]. Unfortunately, the above skew t distribution-based methods are all in linear systems and cannot be applied to nonlinear systems.

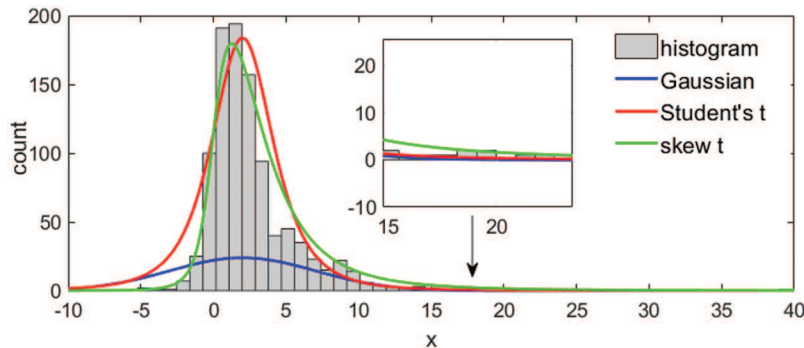


Figure 1: Skew t distribution has better fitting than symmetric distributions

In this paper, a new skew t cubature Kalman filter (STCKF) is proposed for nonlinear system with heavy-tailed and skewed measurement noise. The skew t distribution is adopted to describe the measurement noise and the prior distributions of the shape matrix, scale matrix and DOF are chosen as Gaussian, inverse Wishart and Gamma distributions, respectively. The unknown statistics including shape matrix, scale matrix and DOF are inferred with the VB approach and the posterior of states is also simultaneously obtained. The results of simulation demonstrate that

the proposed STCKF has better estimation accuracy as compared with the CKF and Student's t distribution-based CKF.

The paper is structured as follows: [Section 2](#) describes the problem studied in this paper. [Section 3](#) proposes a skew t cubature Kalman filter based on VB inference. In [Section 4](#), an example of target tracking is presented to verify the estimation performance of the proposed STCKF. The conclusions of this paper are given in [Section 5](#).

Notations: \mathbb{R}^K denotes the K -dimensional Euclidean space, $E[\cdot]$ and $\text{tr}(\cdot)$ represent the expectation operator and the trace operator, I is the identity matrix with appropriate dimension, A^T is the transpose of matrix A , $\text{diag}(\cdot)$ is the diagonal matrix, $N(x, P)$ is a Gaussian distribution with mean vector x and covariance matrix P , $N_+(\mu, \Sigma)$ represents the truncated Gaussian distribution in positive orthant, and its location is μ and scale matrix is Σ .

2 Problem Formulation

Consider the nonlinear state-space model

$$x_n = f(x_{n-1}) + w_{n-1}, \tag{1}$$

$$z_n = h(x_n) + v_n, \tag{2}$$

where n is the discrete time index, $f(\cdot)$ and $h(\cdot)$ are the nonlinear state function and measurement function, $z_n \in \mathbb{R}^d$ is the measurement vector, $x_n \in \mathbb{R}^m$ is the state vector. The process noise w_n is Gaussian white noise and $w_n \sim N(0, Q_n)$. The measurement noise v_n is the heavy-tailed and asymmetric noise. The initial state vector x_0 is assumed to have a Gaussian distribution,

$$p(x_0) = N(x_0; x_{0|0}, P_{0|0}). \tag{3}$$

The skew t distribution is used to describe the heavy-tailed and asymmetric of noise, therefore, the measurement noise v_n :

$$p(v_n) = \text{ST}(v_n; 0, \Delta_n, R_n, \nu_n). \tag{4}$$

where $\text{ST}(v_n; 0, \Delta_n, R_n, \nu_n)$ denotes the skew t distribution with location 0, shape matrix Δ_n , scale matrix R_n and DOF ν_n . Specifically, the definition of skew t distribution can be seen in [30,32].

[Fig. 2](#) shows the different Δ of distributions $\text{ST}(x; 0, 1, \Delta, 3)$. As shown, with the decreasing of Δ , the skew t distribution will deteriorate to the Student's t distribution. Therefore, the skew t distribution has both heavy-tails and asymmetric properties with suitable values.

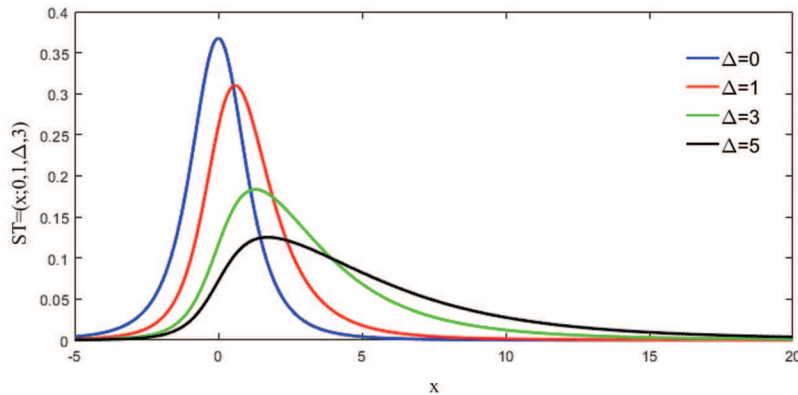


Figure 2: Different Δ for $\text{ST}(x; 0, 1, \Delta, 3)$

In the following, based on the CKF, we will design a nonlinear filter under nonlinear model (1)–(2) with the measurement noise followed by skew t distribution. Specifically, the statistics of skew t distribution including the shape matrix, the scale matrix and the DOF are unknown and need to be estimated together with the system states by using VB inference.

3 Proposed Skew t Cubature Kalman Filter Using VB Inference

3.1 Prior Distributions Update

Similar to CKF, the predicted distribution of system state x_n is

$$p(x_n|z_{1:n-1}) = \mathbf{N}(x_n; x_{n|n-1}, P_{n|n-1}), \quad (5)$$

where $x_{n|n-1}$ and $P_{n|n-1}$ can be approximated by the CKF here. More specifically, the cubature points are obtained by

$$\bar{\chi}_{i,n-1} = x_{n-1|n-1} + \sqrt{P_{n-1|n-1}} \xi_i, \quad (6)$$

where $i = 1, \dots, 2m$, $\xi_i = \sqrt{m} \times [1]_i$, and $[1]_i$ denotes the i -th element of the following set:

$$\left[\begin{array}{c} \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix} \begin{pmatrix} 0 \\ 1 \\ \vdots \\ 0 \end{pmatrix} \cdots \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 1 \end{pmatrix} \begin{pmatrix} -1 \\ 0 \\ \vdots \\ 0 \end{pmatrix} \begin{pmatrix} 0 \\ -1 \\ \vdots \\ 0 \end{pmatrix} \cdots \begin{pmatrix} 0 \\ 0 \\ \vdots \\ -1 \end{pmatrix} \end{array} \right].$$

The predicted cubature points are

$$\chi_{i,n-1} = f(\bar{\chi}_{i,n-1}), \quad (7)$$

Hence, the predicted state and the corresponding error covariance are given by

$$x_{n|n-1} = \frac{1}{2m} \sum_{i=1}^{2m} \chi_{i,n-1} \quad (8)$$

$$P_{n|n-1} = \frac{1}{2m} \sum_{i=1}^{2m} \chi_{i,n-1} \chi_{i,n-1}^T - x_{n|n-1} x_{n|n-1}^T + Q_{n-1}. \quad (9)$$

In Bayesian theory, the conjugate prior needs ensure posterior distribution have the same functional form with prior distribution. Therefore, to infer shape matrix Δ_n , scale matrix R_n , and DOF ν_n , the conjugate priors of Δ_n , R_n , and ν_n need to be selected. The prior distribution of Δ_n is selected as Gaussian distribution [33]:

$$p(\Delta_n|z_{1:n-1}) = \prod_{j=1}^d \mathbf{N}(\Delta_{n,j}; \mu_{n|n-1,j}, \sigma_{n|n-1,j}), \quad (10)$$

where $\mu_{n|n-1,j}$ is the predicted mean of j -th dimension and $\sigma_{n|n-1,j}$ is the corresponding variance, respectively. The prior distribution of R_n is selected as inverse Wishart distribution [21]:

$$p(R_n|y_{1:n-1}) = \mathbf{IW}(R_n; c_{n|n-1}, D_{n|n-1}), \quad (11)$$

where $D_{n|n-1}$ and $c_{n|n-1}$ are the inverse scale matrix and DOF, respectively. The prior distribution of v_n is selected as Gamma distribution [26]:

$$p(v_n|z_{1:n-1}) = G(v_n; a_{n|n-1}, b_{n|n-1}), \quad (12)$$

where $a_{n|n-1}$ and $b_{n|n-1}$ are the shape parameter and the rate parameter, respectively.

To obtain (10)–(12), the dynamic model $p(\Delta_n|\Delta_{n-1})$, $p(R_n|R_{n-1})$ and $p(v_n|v_{n-1})$ need to be specified. In practice, the variation of the measurement noise parameters is slow, and we can use a forgetting factor $\rho \in (0, 1]$ to describe the predicted distribution in this paper [21]:

$$\mu_{n|n-1,j} = \rho\mu_{n-1|n-1,j}, \quad \sigma_{n|n-1,j} = \rho\sigma_{n-1|n-1,j}, \quad (13)$$

$$c_{n|n-1} = \rho(c_{n-1|n-1} - d - 1) + d + 1, \quad D_{n|n-1} = \rho D_{n-1|n-1}, \quad (14)$$

$$a_{n|n-1} = \rho a_{n-1|n-1}, \quad b_{n|n-1} = \rho b_{n-1|n-1}. \quad (15)$$

Because of the skewed t distribution does not have a strictly closed form, the state posterior distribution will be difficult to obtain. With the introduction of two hidden variables u_n and Λ_n , the likelihood $p(z_n|x_n)$ can be rewritten by the following hierarchical Gaussian model [30]:

$$p(z_n|x_n, \Delta_n, u_n, \Lambda_n, R_n) = \mathbf{N}(h(x_n) + \Delta_n u_n, \Lambda_n^{-1} R_n), \quad (16)$$

$$p(u_n|\Lambda_n) = \mathbf{N}_+(0, \Lambda_n^{-1}), \quad (17)$$

$$p(\Lambda_n|v_n) = \mathbf{G}\left(\frac{v_n}{2}, \frac{v_n}{2}\right). \quad (18)$$

3.2 Posterior Estimation

In order to estimate x_n from (5), (10)–(12), and (16)–(18), the joint posterior $p(x_n, \Delta_n, u_n, \Lambda_n, R_n, v_n|z_{1:n})$ needs to be computed. According to Bayes' theorem,

$$p(\mathcal{H}|z_{1:n}) = \frac{p(z_n|\mathcal{H})p(\mathcal{H}|z_{1:n-1})}{p(z_n|z_{1:n-1})}, \quad (19)$$

where $\mathcal{H} = \{x_n, \Delta_n, u_n, \Lambda_n, R_n, v_n\}$. Due to inter-coupled parameters, it is infeasible to infer the posterior $p(\mathcal{H}|z_{1:n})$ analytically. From (19), the logarithmic marginal likelihood $\log p(z_n|z_{1:n-1})$ can be derived as [35]

$$\log p(z_n|z_{1:n-1}) = \text{KLD}(q(\mathcal{H})||p(\mathcal{H}|z_{1:n})) + \mathcal{L}(q(\mathcal{H})), \quad (20)$$

where $\text{KLD}(\cdot)$ and \mathcal{L} are the Kullback-Liebler divergence and the lower bound of $\log p(z_n|z_{1:n-1})$, respectively. Due to the non-negativity of the KLD, we can obtain the true posterior by minimizing the KLD between $q(\mathcal{H})$ and $p(\mathcal{H}|z_{1:n})$ [35,36]. Thus, the fix-point iteration of VB inference is utilized to approximate $p(x_n, \Delta_n, u_n, \Lambda_n, R_n, v_n|z_{1:n})$ by means of the product of some individual distributions [21,37], i.e.,

$$p(x_n, \Delta_n, u_n, \Lambda_n, R_n, v_n|z_{1:n}) \approx q(x_n)q(\Delta_n)q(u_n)q(\Lambda_n)q(R_n)q(v_n), \quad (21)$$

where $q(\cdot)$ represents the approximate posterior of $p(\cdot)$. An analytical solution of $q(x_n)$, $q(\Delta_n)$, $q(u_n)$, $q(\Lambda_n)$, $q(R_n)$, and $q(v_n)$ can be obtained by [21]

$$\log q(\varphi) = \mathbb{E}_{\mathcal{H}^{(-\varphi)}}[\log p(z_n, \mathcal{H}|z_{1:n-1})] + c_\varphi \quad (22)$$

where φ is an element of \mathcal{H} , $\mathcal{H}^{(-\varphi)}$ is all elements in \mathcal{H} except for φ , and c_φ is the constant on the variable φ .

Based on Bayesian theory, we can obtain the joint posterior distribution as follows:

$$\begin{aligned} p(z_n, \mathcal{H}|z_{1:n-1}) &= p(z_n|\mathcal{H})p(\mathcal{H}|z_{1:n-1}) \\ &= p(z_n|x_n, \Delta_n, u_n, \Lambda_n, R_n)p(x_n|z_{1:n-1})p(\Delta_n|z_{1:n-1})p(u_n|\Lambda_n)p(\Lambda_n|v_n) \\ &\quad \times p(R_n|c_{n|n-1}, D_{n|n-1})p(v_n|a_{n|n-1}, b_{n|n-1}). \end{aligned} \quad (23)$$

Substituting (5), (10)–(12), and (16)–(18) into (23) results in

$$\begin{aligned} p(z_n, \mathcal{H}|z_{1:n-1}) &= \mathbf{N}(z_n; h(x_n) + \Delta_n u_n, \Lambda_n^{-1} R_n) \mathbf{N}(x_n; x_{n|n-1}, P_{n|n-1}) \prod_{j=1}^d \mathbf{N}(\Delta_{n,j}; \mu_{n|n-1,j}, \sigma_{n|n-1,j}) \\ &\quad \times \mathbf{N}_+(u_n; 0, \Lambda_n^{-1}) \mathbf{G}(\Lambda_n; \frac{v_n}{2}, \frac{v_n}{2}) \mathbf{IW}(R_n; c_{n|n-1}, D_{n|n-1}) \mathbf{G}(v_n; a_{n|n-1}, b_{n|n-1}). \end{aligned} \quad (24)$$

When $\varphi = x_n$, the posterior distribution $q(x_n)$ is calculated as:

$$q(x_n) = \mathbf{N}(x_n; x_{n|n}, P_{n|n}), \quad (25)$$

where $x_{n|n}$ and $P_{n|n}$ are the estimate of state and the corresponding covariance respectively, and can be obtained by

$$\chi_{i,n|n-1} = x_{n|n-1} + \sqrt{P_{n|n-1}} \xi_i, \quad (26)$$

$$\bar{z}_n = \frac{1}{2m} \sum_{i=1}^{2m} h(\chi_{i,n|n-1}), \quad (27)$$

$$P_{xz,n} = \frac{1}{2m} \sum_{i=1}^{2m} (\chi_{i,n|n-1} - x_{n|n-1}) (h(\chi_{i,n|n-1}) - \bar{z}_n)^T, \quad (28)$$

$$P_{zz,n} = \frac{1}{2m} \sum_{i=1}^{2m} (h(\chi_{i,n|n-1}) - \bar{z}_n) (h(\chi_{i,n|n-1}) - \bar{z}_n)^T + \tilde{R}_n, \quad (29)$$

$$K_x = P_{xz,n} P_{zz,n}^{-1}, \quad (30)$$

$$x_{n|n} = x_{n|n-1} + K_x (z_n - \bar{z}_n - \mathbf{E}[\Delta_n] \mathbf{E}[u_n]), \quad (31)$$

$$P_{n|n} = P_{n|n-1} - K_x P_{zz,n} K_x, \quad (32)$$

where $\tilde{R}_n = \frac{\{\mathbf{E}[R_n^{-1}]\}^{-1}}{\mathbf{E}[\Lambda_n]}$. The derivation of (25)–(32) can be seen in [Appendix A](#).

When $\varphi = \Delta_n$, the posterior distribution $q(\Delta_n)$ is calculated as

$$q(\Delta_n) = \prod_{j=1}^d \mathbf{N}(\Delta_{n,j}; \mu_{n|n,j}, \sigma_{n|n,j}), \quad (33)$$

where the mean $\mu_{n|n,j}$ and variance $\sigma_{n|n,j}$ are given by

$$K_{\Delta,j} = \sigma_{n|n-1,j} \mathbf{E}[u_{n,j}] (\sigma_{n|n-1,j} (\mathbf{E}[u_{n,j}])^2 + \tilde{R}_{n,j})^{-1}, \quad (34)$$

$$\mu_{n|n,j} = \mu_{n|n-1,j} + K_{\Delta,j} (\varepsilon_{n,j} - \mathbf{E}[u_{n,j}] \mu_{n|n-1,j}), \quad (35)$$

$$\sigma_{n|n,j} = \sigma_{n|n-1,j} - \sigma_{n|n-1,j} \mathbf{E}[u_{n,j}] K_{\Delta,j}, \quad (36)$$

where $\varepsilon_n = \mathbf{E}[z_n - h(x_n)]$. The derivation of (33)–(36) can be seen in [Appendix B](#).

When $\varphi = u_n$, the posterior distribution $q(u_n)$ is calculated as

$$q(u_n) = \mathbf{N}_+(u_n; u_{n|n}, U_{n|n}), \quad (37)$$

where the location $u_{n|n}$ and covariance $U_{n|n}$ are obtained by

$$K_u = \mathbf{E}[\Delta_n] \left((\mathbf{E}[\Delta_n])^2 + \tilde{R}_n \right)^{-1}, \quad (38)$$

$$u_{n|n} = K_u \varepsilon_n, \quad (39)$$

$$U_{n|n} = (\mathbf{I} - K_u \mathbf{E}[\Delta_n]) (\mathbf{E}[\Lambda_n])^{-1}. \quad (40)$$

The derivation of (36)–(40) can be seen in [Appendix B](#).

When $\varphi = \Lambda_n$, the posterior distribution $q(\Lambda_n)$ is calculated as

$$q(\Lambda_n) = \mathbf{G}(\Lambda_n; \alpha_n, \beta_n), \quad (41)$$

where the shape parameter α_n and the rate parameter β_n are given by

$$\alpha_n = \frac{1}{2} (\mathbf{E}[v_n] + m), \quad (42)$$

$$\beta_n = \frac{1}{2} (\Psi_n + \mathbf{E}[v_n]), \quad (43)$$

where ψ_n is an auxiliary parameter and the derivation of (41)–(43) can be seen in [Appendix C](#).

When $\varphi = R_n$, the posterior distribution $q(R_n)$ is calculated as

$$q(R_n) = \mathbf{IW}(R_n; c_{n|n}, D_{n|n}), \quad (44)$$

where the DOF $c_{n|n}$ and inverse scale matrix $D_{n|n}$ are obtained by

$$c_{n|n} = c_{n|n-1} + 1, \quad (45)$$

$$D_{n|n} = \Upsilon_n + D_{n|n-1}, \quad (46)$$

where Υ_n is an auxiliary parameter and the derivation of (44)–(46) can be seen in [Appendix C](#).

When $\varphi = v_n$, the posterior distribution $q(v_n)$ is calculated as

$$q(v_n) = \mathbf{G}(v_n; a_{n|n}, b_{n|n}), \quad (47)$$

where the parameters a_n and b_n are given by

$$a_{n|n} = a_{n|n-1} + \frac{1}{2}, \quad (48)$$

$$b_{n|n} = b_{n|n-1} - \frac{1}{2} - \frac{1}{2} \mathbf{E}[\log \Lambda_n] + \frac{1}{2} \mathbf{E}[\Lambda_n]. \quad (49)$$

The derivation of (47)–(49) can be seen in [Appendix D](#).

Using (25), (33), (36), (41), (44), and (47), the following expectations are required:

$$E[\Lambda_n] = \alpha_n / \beta_n, \quad (50)$$

$$E[R_n^{-1}] = (c_{n|n} - M - 1)D_{n|n}^{-1}, \quad (51)$$

$$E[v_n] = a_{n|n} / b_{n|n}, \quad (52)$$

$$E[\Delta_n] = \text{diag}(\mu_{n|n,1}, \dots, \mu_{n|n,d}), \quad (53)$$

$$E[\log \Lambda_n] = \psi(\alpha_n) - \log \beta_n, \quad (54)$$

where $\psi(\cdot)$ denotes the digamma function. The computations of $E[u_n]$ and $E[u_n u_n^T]$ can be found in [38].

After taking \mathbb{N} fix-point iteration steps, the approximations of posterior distributions are updated as

$$q(x_n) \approx \mathbf{N}(x_n; x_{n|n}^{(\mathbb{N})}, P_{n|n}^{(\mathbb{N})}) = \mathbf{N}(x_n; x_{n|n}, P_{n|n}), \quad (55)$$

$$q(\Delta_n) \approx \prod_{j=1}^d \mathbf{N}(\Delta_{n,j}; \mu_{n|n,j}^{(\mathbb{N})}, \sigma_{n|n,j}^{(\mathbb{N})}) = \prod_{j=1}^d \mathbf{N}(\Delta_{n,j}; \mu_{n|n,j}, \sigma_{n|n,j}), \quad (56)$$

$$q(u_n) \approx \mathbf{N}_+(u_n; u_{n|n}^{(\mathbb{N})}, U_{n|n}^{(\mathbb{N})}) = \mathbf{N}_+(x_n; u_{n|n}, U_{n|n}), \quad (57)$$

$$q(\Lambda_n) \approx \mathbf{G}(\Lambda_n; \alpha_n^{(\mathbb{N})}, \beta_n^{(\mathbb{N})}) = \mathbf{G}(\Lambda_n; \alpha_n, \beta_n), \quad (58)$$

$$q(R_n) \approx \mathbf{IW}(R_n; c_{n|n}^{(\mathbb{N})}, D_{n|n}^{(\mathbb{N})}) = \mathbf{IW}(R_n; c_{n|n}, D_{n|n}), \quad (59)$$

$$q(v_n) \approx \mathbf{G}(v_n; a_{n|n}^{(\mathbb{N})}, b_{n|n}^{(\mathbb{N})}) = \mathbf{G}(v_n; a_{n|n}, b_{n|n}). \quad (60)$$

Combining prediction steps (5) and (10)–(12) with measurement updates (25), (33), (36), (41), (44) and (47), the proposed STCKF can be realized recursively. To implement the proposed filter, the initial shape matrix Δ_0 , the initial scale matrix R_0 , the initial DOF ν_0 , and the forgetting factor ρ need to be determined. Generally, Δ_0 , R_0 , and ν_0 can be approximately achieved from prior knowledge. The forgetting factor ρ can determine how much information from the previous estimation. That is, choosing a small ρ means forgetting more information and vice versa. In the proposed STCKF, the selection of \mathbb{N} is a trade-off. Increasing \mathbb{N} will obtain better estimation accuracy but more time-consuming.

4 Target Tracking Simulation

To validate the estimation performance of the proposed STCKF, a target tracking simulation is introduced to perform an evaluation of the results obtained. The STCKF is compared with the CKF [16], the Student's t distribution based-CKF (T-CKF) [25], and the robust Student's t distribution-based CKF (RT-CKF) [26].

In this paper, a typical air traffic control scenario is considered, in which the aircraft performs maneuvering turns on the horizontal plane at a constant but unknown turning rate Ω .

The kinematics of rotational motion can be described by the following nonlinear state-space model [16]:

$$x_n = \begin{bmatrix} 1 & \frac{\sin \Omega T}{\Omega} & 0 - (\frac{1 - \cos \Omega T}{\Omega}) & 0 \\ 0 & \cos \Omega T & 0 & -\sin \Omega T & 0 \\ 0 & (\frac{1 - \cos \Omega T}{\Omega}) & 1 & \frac{\sin \Omega T}{\Omega} & 0 \\ 0 & \sin \Omega T & 0 & \cos \Omega T & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix} x_{n-1} + w_n, \quad (61)$$

$$z_n = \begin{bmatrix} \sqrt{\xi_n^2 + \eta_n^2} \\ \arctan \frac{\eta_n}{\xi_n} \end{bmatrix} + v_n, \quad (62)$$

where the state $x = (\xi \ \dot{\xi} \ \eta \ \dot{\eta} \ \Omega)^T$, the position and velocity of target in x and y directions are (ξ, η) and $(\dot{\xi}, \dot{\eta})$, T is the time interval and Ω is the tuning rate. The process noise covariance is $Q_n = \text{diag}(q_1 M, q_1 M, q_2 T)$, where

$$M = \begin{bmatrix} T^3 & T^2 \\ \frac{3}{T^2} & \frac{2}{T} \end{bmatrix}.$$

The associated parameters are set as: $q_1 = 0.1 \text{ m}^2\text{s}^{-3}$, $q_2 = 1.75 \times 10^{-4} \text{ s}^{-3}$, $\Omega = -3^\circ \text{ s}^{-1}$, $T = 1 \text{ s}$. The true initial state and the corresponding covariance are $x_0 = (1000, 300, 1000, 0, -3)^T$ and $P_0 = \text{diag}(100, 10, 100, 10, 100)$, respectively.

In this simulation, we consider three cases for measurement noise:

Case 1: Gaussian distribution, that is, $v_n \sim N(0, R_n)$ with noise covariance $R_n = \text{diag}(100, 10)$,

Case 2: Contaminated Gaussian distribution (mixture Gaussian distribution). The mixture Gaussian distributed noise is generated according to [33]

$$v_n \sim \begin{cases} N(0, R_n) & w.p. \ 1 - p_c \\ N(0, 100R_n) & w.p. \ p_c, \end{cases} \quad (63)$$

where $R_n = \text{diag}(100, 10)$. Eq. (63) means that v_n is drawn from $N(0, R_n)$ with probability $1 - p_c$ and $N(0, 100R_n)$ with probability p_c . In this paper, $p_c = 0.1$.

Case 3: Contaminated skew t distribution (mixture skew t distribution).

According to [32], the measurement noise v_n is generated by

$$v_n \sim \begin{cases} ST(0, R_n, \Delta_n, v_n) & w.p. \ 1 - p_c \\ ST(0, 10R_n, \Delta_n, v_n) & w.p. \ p_c, \end{cases} \quad (64)$$

where $R_n = \text{diag}(100, 10)$, $\Delta_n = \text{diag}(5, 5)$, and $v_n = 4$.

In this paper, the root mean square error (RMSE) is adopted to test the filtering performance, and its formula is

$$\text{RMSE}_n = \sqrt{\frac{1}{M} \sum_{j=1}^M \|x_{n|n}^j - x_n^j\|^2}, \quad (65)$$

where $x_{n|n}^j$ and x_n^j are the estimated and true values, respectively, at the j -th Monte-Carlo run. The simulated inputs are $a_0 = 4$, $b_0 = 1$, $c_0 = 3$, $D_0 = \text{diag}(1, 1)$, $\mu_0 = \text{diag}(1, 1)$, $\sigma_0 = \text{diag}(1, 1)$, $\rho = 1 - \exp(-5)$ and $N = 5$.

Figs. 3–5 show that the RMSEs of position, velocity, and tuning rate based on 20 Monte-Carlo runs for three cases. From Fig. 3, the CKF, T-CKF, RT-CKF and the proposed STCKF almost have the same estimation accuracy under Gaussian distribution noise. However, the methods based on the non-Gaussian distribution outperform the methods based on the Gaussian distribution when the measurement noise no longer satisfies the Gaussian distribution. As shown in Fig. 4, the non-Gaussian filters (T-CKF, RT-CKF and STCKF) perform better than the CKF. From Fig. 5, the proposed STCKF obtains the best accuracy in the case of asymmetric noise distribution and unknown noise statistics. As also observed in Table 1, the filter based on the skewed t distribution and the filters based on the Student's t distribution perform better than the filter based on the Gaussian distribution in Cases 2 and 3, and the proposed STCKF is significantly better than other filters for the asymmetric noise distribution.

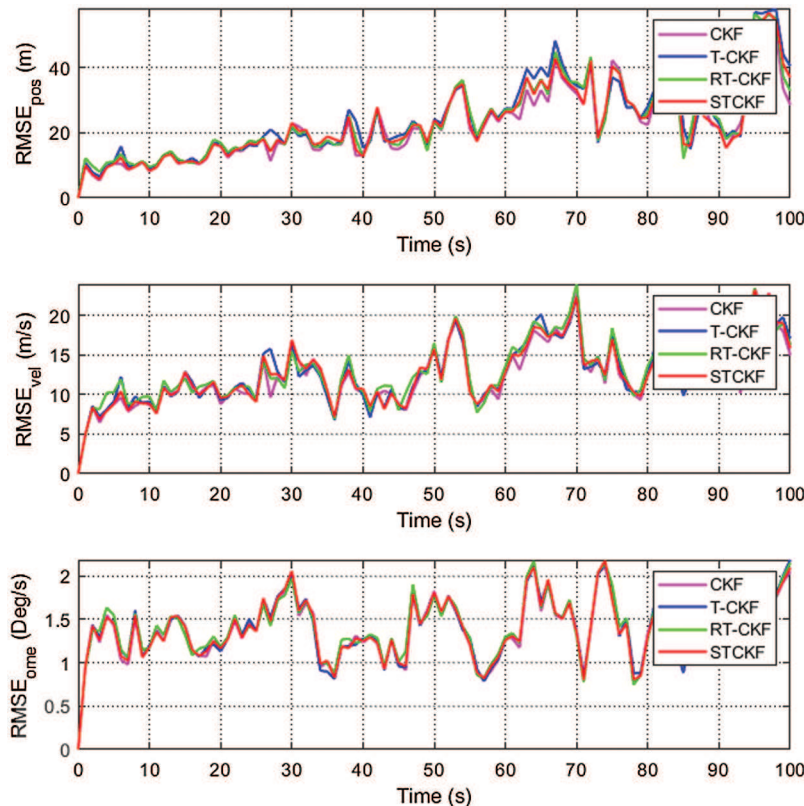


Figure 3: RMSEs of the position, velocity, and tuning rate by different filters under Case 1

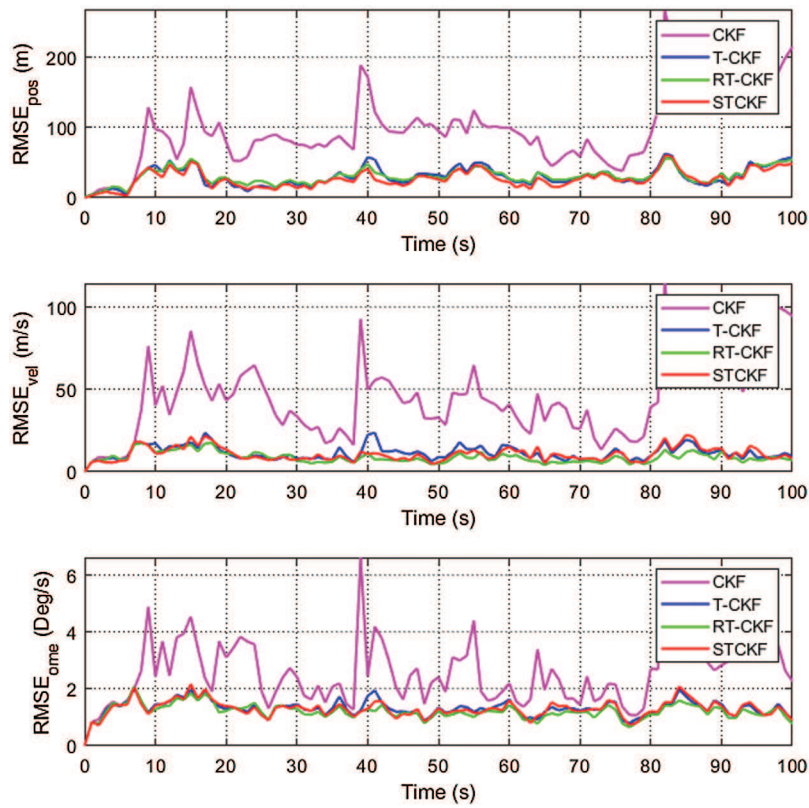


Figure 4: RMSEs of the position, velocity, and tuning rate by different filters under Case 2

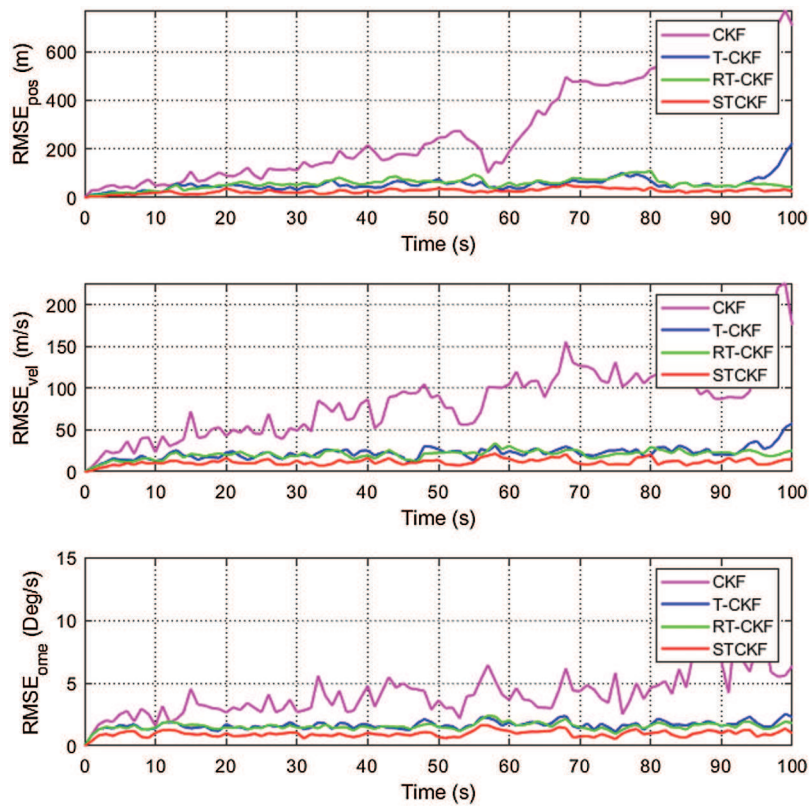


Figure 5: RMSEs of the position, velocity, and tuning rate by different filters under Case 3

Table 1: Average RMSEs by different filters

| | | CKF | T-CKF | RT-CKF | STCKF |
|-------------|--------|---------|--------|--------|--------|
| Position | Case 1 | 23.392 | 24.127 | 23.656 | 23.142 |
| | Case 2 | 98.756 | 30.794 | 30.161 | 27.920 |
| | Case 3 | 275.719 | 56.467 | 56.823 | 26.715 |
| Velocity | Case 1 | 12.922 | 12.887 | 12.889 | 12.830 |
| | Case 2 | 45.373 | 11.346 | 10.680 | 9.756 |
| | Case 3 | 83.887 | 22.471 | 20.631 | 11.623 |
| Tuning rate | Case 1 | 0.024 | 0.024 | 0.024 | 0.024 |
| | Case 2 | 0.046 | 0.023 | 0.021 | 0.021 |
| | Case 3 | 0.073 | 0.028 | 0.027 | 0.016 |

5 Conclusion

In this work, we consider the joint estimation problem of system states and unknown noise statistics for nonlinear discrete-time systems. Combining with the properties of skew t distribution, a hierarchical nonlinear Gaussian model is developed. Based on this model, a skew t cubature Kalman filter is proposed, in which the states, shape matrix, scale matrix and DOF are simultaneously estimated by using VB approach. The results of simulation show that the proposed filter in this paper has better estimation accuracy than the conventional CKF and the Student's t distribution-based CKF under heavy-tailed and skewed measurement noise. It should be noted that the proposed method in this paper can only realize state estimation of asymmetric measurement noise. How to extend the proposed method to handle asymmetric process and measurement noise is still an open problem.

Funding Statement: This work was supported in part by National Natural Science Foundation of China under Grants 62103167 and 61833007, and in part by the Natural Science Foundation of Jiangsu Province under Grant BK20210451.

Conflicts of Interest: The authors declare that they have no conflicts of interest to report regarding the present study.

References

1. Frueh, C. (2016). Modeling impacts on space situational awareness phd filter tracking. *Computer Modeling in Engineering & Sciences*, 111(2), 171–201. DOI 10.3970/cmcs.2016.111.171.
2. Zhao, S., Shmaliy, Y. S., Ahn, C. K., Liu, F. (2017). Adaptive-horizon iterative ufir filtering algorithm with applications. *IEEE Transactions on Industrial Electronics*, 65(8), 6393–6402. DOI 10.1109/TIE.41.
3. Chen, H., Jiang, B., Ding, S. X., Huang, B. (2020). Data-driven fault diagnosis for traction systems in high-speed trains: A survey, challenges, and perspectives. *IEEE Transactions on Intelligent Transportation Systems*, Early Access. DOI 10.1109/TITS.6979.
4. Jiang, Q., Fu, X., Yan, S., Li, R., Du, W. et al. (2021). Neural network aided approximation and parameter inference of non-markovian models of gene expression. *Nature Communications*, 12(1), 1–12. DOI 10.1038/s41467-021-22919-1.
5. Chen, H., Jiang, B. (2019). A review of fault detection and diagnosis for the traction system in high-speed trains. *IEEE Transactions on Intelligent Transportation Systems*, 21(2), 450–465. DOI 10.1109/TITS.6979.

6. Hou, X., Qiao, G. (2020). Observability analysis in parameters estimation of an uncooperative space target. *Computer Modeling in Engineering & Sciences*, 122(1), 175–206. DOI 10.32604/cmcs.2020.08452.
7. Zhao, S., Huang, B., Liu, F. (2016). Linear optimal unbiased filter for time-variant systems without apriori information on initial conditions. *IEEE Transactions on Automatic Control*, 62(2), 882–887. DOI 10.1109/TAC.2016.2557999.
8. Simon, D. (2006). *Optimal state estimation: Kalman, H infinity, and nonlinear approaches*. Hoboken, New Jersey: John Wiley & Sons.
9. Geng, H., Wang, Z., Cheng, Y., Alsaadi, F. E., Dobaie, A. M. (2019). State estimation under non-Gaussian lévy and time-correlated additive sensor noises: A modified tobit kalman filtering approach. *Signal Processing*, 154, 120–128. DOI 10.1016/j.sigpro.2018.08.005.
10. Zhao, S., Huang, B. (2020). Trial-and-error or avoiding a guess? Initialization of the kalman filter. *Automatica*, 121, 109184. DOI 10.1016/j.automatica.2020.109184.
11. Myers, M., Jorge, A., Yuhas, D., Walker, D. (2012). An adaptive extended kalman filter incorporating state model uncertainty for localizing a high heat flux point source using an ultrasonic sensor array. *Computer Modeling in Engineering & Sciences*, 83(3), 221–248. DOI 10.3970/cmcs.2012.083.221.
12. Wang, H., Haynes, R., Huang, H., Dong, L., Atluri, S. N. (2015). The use of high-performance fatigue mechanics and the extended kalman/particle filters, for diagnostics and prognostics of aircraft structures. *Computer Modeling in Engineering & Sciences*, 105(1), 1–24. DOI 10.3970/cmcs.2015.105.001.
13. Geng, H., Haile, M. A., Fang, H. (2021). Ssue: Simultaneous state and uncertainty estimation for dynamical systems. *International Journal of Robust and Nonlinear Control*, 31(4), 1068–1083. DOI 10.1002/rnc.5344.
14. Welch, G., Bishop, G. (2006). An introduction to the Kalman filter. *Technical Report*. University of North Carolina, Chapel Hill, North Carolina, USA.
15. Julier, S., Uhlmann, J., Durrant-Whyte, H. F. (2000). A new method for the nonlinear transformation of means and covariances in filters and estimators. *IEEE Transactions on Automatic Control*, 45(3), 477–482. DOI 10.1109/9.847726.
16. Arasaratnam, I., Haykin, S. (2009). Cubature kalman filters. *IEEE Transactions on Automatic Control*, 54(6), 1254–1269. DOI 10.1109/TAC.2009.2019800.
17. Xu, C., Zhao, S., Ma, Y., Huang, B., Liu, F. et al. (2021). Sensor fault estimation in a probabilistic framework for industrial processes and its applications. *IEEE Transactions on Industrial Informatics*, Early Access. DOI 10.1109/TII.2021.3063838.
18. Stojanovic, V., He, S., Zhang, B. (2020). State and parameter joint estimation of linear stochastic systems in presence of faults and non-Gaussian noises. *International Journal of Robust and Nonlinear Control*, 30(16), 6683–6700. DOI 10.1002/rnc.5131.
19. Beelen, H., Bergveld, H. J., Donkers, M. (2020). Joint estimation of battery parameters and state of charge using an extended kalman filter: A single-parameter tuning approach. *IEEE Transactions on Control Systems Technology*, 29(3), 1087–1101. DOI 10.1109/TCST.2020.2992523.
20. Sarkka, S., Nummenmaa, A. (2009). Recursive noise adaptive kalman filtering by variational Bayesian approximations. *IEEE Transactions on Automatic Control*, 54(3), 596–600. DOI 10.1109/TAC.2008.2008348.
21. Huang, Y., Zhang, Y., Wu, Z., Li, N., Chambers, J. (2017). A novel adaptive kalman filter with inaccurate process and measurement noise covariance matrices. *IEEE Transactions on Automatic Control*, 63(2), 594–601. DOI 10.1109/TAC.9.
22. He, J., Sun, C., Zhang, B., Wang, P. (2020). Variational Bayesian-based maximum correntropy cubature kalman filter with both adaptivity and robustness. *IEEE Sensors Journal*, 21(2), 1982–1992. DOI 10.1109/JSEN.7361.
23. Roth, M., Özkan, E., Gustafsson, F. (2013). A student's t filter for heavy tailed process and measurement noise. *IEEE International Conference on Acoustics, Speech and Signal Processing*, pp. 5770–5774. Vancouver, Canada.
24. Huang, Y., Zhang, Y., Chambers, J. A. (2019). A novel kullback–leibler divergence minimization-based adaptive Student's t-filter. *IEEE Transactions on Signal Processing*, 67(20), 5417–5432. DOI 10.1109/TSP.78.

25. Piche, R., Sarkka, S., Hartikainen, J. (2012). Recursive outlier-robust filtering and smoothing for nonlinear systems using the multivariate Student-t distribution. *IEEE International Workshop on Machine Learning for Signal Processing*, pp. 1–6. Santander, Spain.
26. Huang, Y., Zhang, Y., Li, N., Chambers, J. (2016). A robust Gaussian approximate filter for nonlinear systems with heavy tailed measurement noises. *IEEE International Conference on Acoustics, Speech and Signal Processing*, pp. 4209–4213. Shanghai, China.
27. Naveau, P., Genton, M. G., Shen, X. (2005). A skewed kalman filter. *Journal of Multivariate Analysis*, *94*(2), 382–400. DOI 10.1016/j.jmva.2004.06.002.
28. Kim, H. M., Ryu, D., Mallick, B. K., Genton, M. G. (2014). Mixtures of skewed kalman filters. *Journal of Multivariate Analysis*, *123*, 228–251. DOI 10.1016/j.jmva.2013.09.002.
29. Lu, C., Zhang, Y., Ge, Q. (2020). Kalman filter based on multiple scaled multivariate skew normal variance mean mixture distributions with application to target tracking. *IEEE Transactions on Circuits and Systems II: Express Briefs*, *68*(2), 802–806. DOI 10.1109/TCSII.8920.
30. Nurminen, H., Ardeshiri, T., Piche, R., Gustafsson, F. (2015). Robust inference for state-space models with skewed measurement noise. *IEEE Signal Processing Letters*, *22*(11), 1898–1902. DOI 10.1109/LSP.2015.2437456.
31. Nurminen, H., Ardeshiri, T., Piché, R., Gustafsson, F. (2018). Skew-*t* filter and smoother with improved covariance matrix approximation. *IEEE Transactions on Signal Processing*, *66*(21), 5618–5633. DOI 10.1109/TSP.2018.2865434.
32. Xu, C., Zhao, S., Ma, Y., Huang, B., Liu, F. (2019). Robust filter design for asymmetric measurement noise using variational Bayesian inference. *IET Control Theory & Applications*, *13*(11), 1656–1664. DOI 10.1049/iet-cta.2018.6016.
33. Huang, Y., Zhang, Y., Shi, P., Wu, Z., Qian, J. et al. (2017). 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. DOI 10.1109/TSMC.2017.2778269.
34. Li, S., Feng, X., He, R., Pan, F. (2021). Joint parameter and state estimation for stochastic uncertain system with multivariate skew t noises. *Chinese Journal of Aeronautics*, Early Access. DOI 10.1016/j.cja.2021.04.032.
35. Ma, Y., Huang, B. (2017). Bayesian learning for dynamic feature extraction with application in soft sensing. *IEEE Transactions on Industrial Electronics*, *64*(9), 7171–7180. DOI 10.1109/TIE.2017.2688970.
36. Zhao, S., Shmaliy, Y. S., Ahn, C. K., Zhao, C. (2019). Probabilistic monitoring of correlated sensors for nonlinear processes in state space. *IEEE Transactions on Industrial Electronics*, *67*(3), 2294–2303. DOI 10.1109/TIE.41.
37. Xu, C., Zhao, S., Liu, F. (2019). Sensor fault detection and diagnosis in the presence of outliers. *Neurocomputing*, *349*, 156–163. DOI 10.1016/j.neucom.2019.01.025.
38. Barr, D. R., Sherrill, E. T. (1999). Mean and variance of truncated normal distributions. *The American Statistician*, *53*(4), 357–361. DOI 10.2307/2686057.

Appendix A

Substituting $\varphi = x_n$ into (22) yields

$$\begin{aligned} \log q(x_n) = & -\frac{1}{2}(z_n - h(x_n) - E[\Delta_n]E[u_n])^T E[R_n^{-1}]E[\Lambda_n](z_n - h(x_n) - E[\Delta_n]E[u_n]) \\ & - \frac{1}{2}(x_n - x_{n|n-1})^T P_{n|n-1}^{-1} \times (x_n - x_{n|n-1}) + c_x. \end{aligned} \quad (66)$$

Defining the modified likelihood distribution $p(z_n|x_n)$ as

$$p(z_n|x_n) = N(h(x_n) + E[\Delta_n]E[u_n], \tilde{R}_n), \quad (67)$$

and using (5) and (67) in (66), we have

$$q(x_n) \propto N(z_n; h(x_n) + E[\Delta_n]E[u_n], \tilde{R}_n)N(x_n; x_{n|n-1}, P_{n|n-1}). \quad (68)$$

According to (5) and (66)–(68), (25)–(32) can be obtained.

Appendix B

Substituting $\varphi = \Delta_n$ into (22) yields

$$\begin{aligned} \log q(\Delta_n) = & -\frac{1}{2}(\varepsilon_n - \Delta_n E[u_n])^T E[R_n^{-1}]E[\Lambda_n](\varepsilon_n - \Delta_n E[u_n]) \\ & - \sum_{j=1}^d \frac{1}{2}(\Delta_{n,j} - \mu_{n|n-1,j})^T \sigma_{n|n-1,j}^{-1}(\Delta_{n,j} - \mu_{n|n-1,j}) + c_\Delta, \end{aligned} \quad (69)$$

where the auxiliary parameter ε_n is given by

$$\varepsilon_n = E[z_n - h(x_n)] = z_n - h(x_{n|n}). \quad (70)$$

Substituting $\varphi = u_n$ into (22), we have

$$\log q(u_n) = -\frac{1}{2}(\varepsilon_n - E[\Delta_n]u_n)^T E[R_n^{-1}]E[\Lambda_n](\varepsilon_n - E[\Delta_n]u_n) - \frac{1}{2}(u_n^T E[\Lambda_n]u_n) + c_u. \quad (71)$$

Similar to (66)–(68), (33)–(40) can be obtained.

Appendix C

Substituting $\varphi = \Lambda_n$ into (22), we have

$$\log q(\Lambda_n) = -\frac{1}{2}\Lambda_n \Psi_n + \left(\frac{\langle \nu_n \rangle + d}{2} - 1 \right) \log \Lambda_n - \frac{\langle \nu_n \rangle}{2} \Lambda_n + c_\Lambda, \quad (72)$$

where

$$\begin{aligned} \Psi_n = & \text{tr}(E[R_n^{-1}]E[(z_n - h(x_n))(z_n - h(x_n))^T]) - \text{tr}(E[R_n^{-1}]E[\Delta_n]E[u_n]\varepsilon_n^T) - \text{tr}(E[\Delta_n]E[R_n^{-1}]\varepsilon_n E[u_n]^T) \\ & + \text{tr}((E[\Delta_n]E[R_n^{-1}]E[\Delta_n] + I)E[u_n]u_n^T). \end{aligned} \quad (73)$$

Substituting $\varphi = R_n$ into (22) yields

$$\log q(R_n) = -\frac{1}{2}(c_{n|n-1} + d + 2) \log |R_n| - \frac{1}{2} \text{tr}((\Upsilon_n + D_{n|n-1})R_n^{-1}) + c_R. \quad (74)$$

where

$$\begin{aligned} \Upsilon_n = & \text{E}[\Lambda_n] \text{E}[(z_n - h(x_n))(z_n - h(x_n))^T] + \text{E}[\Lambda_n] \text{E}[\Delta_n \Delta_n^T] \text{E}[u_n u_n^T] - \text{E}[\Lambda_n] \text{E}[\Delta_n] \text{E}[u_n] \varepsilon_n^T \\ & - \text{E}[\Lambda_n] \text{E}[\Delta_n] \varepsilon_n \text{E}[u_n]^T. \end{aligned} \quad (75)$$

According to (72) and (74), (41)–(46) can be obtained.

Appendix D

Substituting $\varphi = v_n$ into (22) yields

$$\log q(v_n) = \frac{v_n}{2} \log \frac{v_n}{2} - \log \Gamma\left(\frac{v_n}{2}\right) + \left(\frac{v_n}{2} - 1\right) \text{E}[\log \Lambda_n] - \frac{v_n}{2} \text{E}[\Lambda_n] + (a_{n|n-1} - 1) \log v_n - b_{n|n-1} v_n + c_v. \quad (76)$$

Using Stirling's approximation: $\log \Gamma\left(\frac{v_n}{2}\right) \approx \frac{v_n-1}{2} \log\left(\frac{v_n}{2}\right) - \frac{v_n}{2}$ [26], $\log q(v_n)$ can be rewritten as

$$\log q(v_n) = (a_{n|n-1} + \frac{1}{2} - 1) \log v_n - (b_{n|n-1} - \frac{1}{2} - \frac{1}{2} \text{E}[\log \Lambda_n] + \frac{1}{2} \text{E}[\Lambda_n]) v_n + c_v. \quad (77)$$

According to (77), (41)–(43) can be obtained.