




# Robust Linear Filter with Parameter Estimation Under Student- $t$ Measurement Distribution

Zongyuan Wang<sup>1,2</sup> · Weidong Zhou<sup>2</sup> 

Received: 2 May 2018 / Revised: 23 October 2018 / Accepted: 24 October 2018 /

Published online: 3 November 2018

© Springer Science+Business Media, LLC, part of Springer Nature 2018

## Abstract

In this paper, we propose an approximate Kalman filter of measurements following Student's  $t$ -distribution by using the variational Bayes approach. This approach can decompose the estimation of multivariate parameters into a univariate estimation. The recursive formula for the approximate posterior densities of parameters and states is derived in detail. Then, the asymptotic Bayesian Cramer–Rao lower bounds are derived for the proposed filter. Numerical simulations verify both the performance of the proposed filter and the variance lower bounds under time-varying noise. The efficiency of the proposed filter is also demonstrated in a real application, namely an integrated strapdown inertial navigation system/Doppler velocity log shipborne test for navigation.

**Keywords** Kalman filter · Variational Bayes · Cramer–Rao lower bound · Parameter estimation

## 1 Introduction

The classical Kalman filter (KF) is a minimum mean-square-error (MSE) estimator for linear systems under Gaussian noise. Improved KFs are extensively studied to meet the demand of current, more complex systems. For instance, KF with noise correlation at one-epoch apart outperforms the classical KF when the measurement is dependent on the previous state [9]. In applications such as navigation, target tracking and location, noise plays an important role to KF, but its statistics are usually unknown and outliers may appear. Huang et al. [11] found that the KF is suboptimal for non-Gaussian pro-

---

✉ Weidong Zhou  
zhouweidong@hrbeu.edu.cn

Zongyuan Wang  
wangzongyuan@hrbeu.edu.cn

<sup>1</sup> College of Science, Harbin Engineering University, Harbin 150001, China

<sup>2</sup> College of Automation, Harbin Engineering University, Harbin 150001, China

cesses or measurement noise with outliers, as the required Gaussian assumptions are not satisfied. To overcome this problem, the contaminated Gaussian distribution [18] considers outliers for noise modeling, and recursion is based on the Bayesian framework with first-order approximation of the prior distribution of the state. Another approach to model the effect of outliers is to use the Student's  $t$ -distribution [19], which has longer tails than the Gaussian distribution assigning non-negligible probabilities to outliers and enabling filters to deal with outliers natively [1]. A linear filter based on the Student's  $t$ -distribution has been proposed assuming that the predicted state and measurement are both Student's  $t$ -distributions, and the posterior probability density function (PDF) is approximately a Student's  $t$ -distribution by adjusting matrix parameters and the degrees of freedom (DOF) [15]. However, this scheme is limited to a specific noise model with process and measurement  $t$ -distributions having the same number of DOF. Huang and Zhang [8] propose a robust stochastic cubature filter based on the Student's  $t$ -distribution by modeling a heavy-tailed process and measurement noises as  $t$ -distributions with varying DOF, thus extending the stochastic numerical integral based on the Gaussian assumption.

Recently, estimation using a Student- $t$  noise model has been combined with the expectation maximization algorithm [10] and variational Bayes (VB) approach [11, 24]. Expectation maximization iteratively determines the maximum likelihood estimates of parameters for models with some latent variables. The estimation accuracy is related to the number of observations available and often considered as a special case of the VB approach when the approximate density  $q(\theta)$  of parameter  $\theta$  satisfies  $q(\theta) = \delta(\theta - \theta^*)$ , with  $\theta^*$  being the true value and  $\delta$  representing a delta function [21]. In fact, the VB approach is an advanced Bayes estimator in the presence of latent variables [11, 24]. Compared with expectation maximization, in which only the single most probable value is estimated by maximum likelihood [4], the VB approach can provide estimation of a posterior distribution of parameters and the posterior distribution of latent variables [13], being an effective alternative to traditional Bayesian methods and expectation maximization [22]. Huang et al. [11] derive a hierarchical Gaussian estimation algorithm based on the Student's  $t$ -distribution in the presence of heavy-tailed process and measurement noise. They combine VB principles with the conventional KF, consider the coupling relation of state and noise variance, and estimate the state and parameters online by alternating recursion and updating. The VB approach has been successfully used in many other research fields like continuous-discrete stochastic dynamic systems [2] and nonlinear dynamical systems [3, 16].

The mean values of parameters are also important for determining Gaussian distribution or Student's  $t$ -distribution, and for applications such as navigation [7] and odometry [14], it is usually nonzero and time varying by factors such as sensor offset. Mean estimation has been exploited and theoretically developed using approaches such as maximum a posteriori with one-step smoothing [7] and the VB approach [25]. In [25], a Gaussian mixture measurement model is utilized but details on the derivation are not presented. From the perspective of Bayesian inference, the mean is treated as a random variable and coupled with the state and variance, and thus the posterior distribution deduced from the VB approach is different from the previous expressions without mean, as detailed in [11, 24].

On the other hand, lower bounds indicate performance limitations, and hence they can be used to determine whether performance requirements can be satisfied [20, 26]. Specifically, the Bayesian Cramer–Rao lower bounds (CRLBs), also called posterior CRLBs, are used for evaluating performance limitations for unknown random parameters with known prior distribution. Although several variations of the Bayesian CRLBs are available, lower bounds for linear approximate estimation with the VB approach have been seldom studied.

Motivated by the above-mentioned aspects, in this paper we use the Student’s  $t$ -distribution to model non-Gaussian measurement noise considering mean parameters. First, the VB approach estimates the distributions of the parameters and state of a linear system, where the conjugate prior PDFs of mean and scale matrix are modeled as Gaussian–Wishart distributions. Then, asymptotic Bayesian CRLBs for VB estimators of the linear filter under the Student- $t$  measurement noise are derived to establish the lower bound of error variance for multivariate linear estimation. Both the state and parameters are considered random, and the hybrid derivation rule of the CRLB is employed.

The rest of this paper is organized as follows. The problem formulation is detailed in Sect. 2, and the estimation of parameters and state under Student- $t$  measurement distribution is derived in Sect. 3. Then, the asymptotic Bayesian CRLBs for the estimated parameters and state are provided in Sect. 4. Simulation and comparison results on the stochastic resonator model and the integrated strapdown inertial navigation system (SINS)/Doppler velocity log (DVL) shipborne test are presented in Sect. 5. Finally, we draw conclusions in Sect. 6.

## 2 Problem Formulation

Consider the following state-space model, which is applicable to integrated navigation and target tracking [7]:

$$\begin{cases} \mathbf{x}_k = \mathbf{F}_{k-1}\mathbf{x}_{k-1} + \mathbf{G}_{k-1}\mathbf{w}_{k-1} \\ \mathbf{z}_k = \mathbf{H}_k\mathbf{x}_k + \mathbf{v}_k \end{cases}, \quad (1)$$

where  $\mathbf{x}_k \in \mathbb{R}^n$  is system state,  $\mathbf{z}_k \in \mathbb{R}^m$  is the measurement data,  $\mathbf{F}_k \in \mathbb{R}^{n \times n}$  is the state transition matrix,  $\mathbf{G}_k \in \mathbb{R}^{n \times p}$  is the system noise matrix,  $\mathbf{w}_k \in \mathbb{R}^p$  is the process noise with distribution  $p(\mathbf{w}_k) = N(\mathbf{w}_k | \mathbf{0}_{p \times 1}, \mathbf{Q}_k)$ , where  $N(\cdot | \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)$  denotes the Gaussian PDF with mean vector  $\boldsymbol{\mu}_k$  and covariance matrix  $\boldsymbol{\Sigma}_k$ , whose inverse is called the scale matrix. In addition,  $\mathbf{H}_k \in \mathbb{R}^{m \times n}$  is the measurement matrix, and all the matrices in Eq. (1) are assumed to be known. Initial state  $\mathbf{x}_0$  follows Gaussian distribution  $p(\mathbf{x}_0) = N(\mathbf{x}_0 | \hat{\mathbf{x}}_{0|0}, \mathbf{P}_{0|0})$ , and  $\mathbf{x}_0$ ,  $\mathbf{w}_k$ , and  $\mathbf{v}_k$  are mutually independent. We model measurement noise  $\mathbf{v}_k \in \mathbb{R}^m$  as a symmetrical Student’s  $t$ -distribution, whose PDF is given by [6]

$$p(\mathbf{v}_k) = S(\mathbf{v}_k | \boldsymbol{\mu}_k, \mathbf{A}_k, \nu_k) = \frac{\Gamma((\nu_k + m)/2)}{\Gamma(\nu_k/2)} \frac{|\mathbf{A}_k|^{1/2}}{(\pi \nu_k)^{m/2}} \left[ 1 + \frac{1}{\nu_k} (\mathbf{v}_k - \boldsymbol{\mu}_k)^T \mathbf{A}_k (\mathbf{v}_k - \boldsymbol{\mu}_k) \right]^{-(\nu_k + m)/2}, \quad (2)$$

where  $S(\cdot|\boldsymbol{\mu}_k, \boldsymbol{\Lambda}_k, \nu_k)$  denotes the PDF of the Student’s  $t$ -distribution with mean vector  $\boldsymbol{\mu}_k$ , scale matrix  $\boldsymbol{\Lambda}_k$ ,  $\nu_k$  denotes the DOF, and  $|\cdot|$  and  $\Gamma(\cdot)$  represent the determinant of square matrix and Gamma function, respectively. Note that  $(\boldsymbol{\Lambda}_k)^{-1}$  is the nominal covariance of the  $t$ -distribution.

As the Student’s  $t$ -distribution can be expressed as an infinite mixture of Gaussian distributions with common mean and variances scaled by the Gamma distribution, likelihood PDF  $p(\mathbf{z}_k|\mathbf{x}_k)$  is expressed as [8]

$$p(\mathbf{z}_k|\mathbf{x}_k) = \int_0^{+\infty} N\left(\mathbf{z}_k \mid \mathbf{H}_k \mathbf{x}_k + \boldsymbol{\mu}_k, (u_k \boldsymbol{\Lambda}_k)^{-1}\right) \times G(u_k | \nu_k/2, \nu_k/2) du_k, \tag{3}$$

where  $u_k$  is an introduced latent variable,  $G(\cdot|\alpha, \beta)$  is the Gamma distribution PDF with shape parameter  $\alpha$  and rate parameter  $\beta$ , and the PDF of  $u_k$  is given by

$$p(u_k) = G(u_k|\alpha, \beta) = \frac{\beta^\alpha}{\Gamma(\alpha)} (u_k)^{\alpha-1} \exp\{-u_k \beta\}. \tag{4}$$

Likelihood PDF  $p(\mathbf{z}_k|\mathbf{x}_k)$  can be rewritten in the hierarchical Gaussian form:

$$\begin{aligned} p(\mathbf{z}_k|\mathbf{x}_k, \boldsymbol{\mu}_k, u_k, \boldsymbol{\Lambda}_k) &= N\left(\mathbf{z}_k \mid \mathbf{H}_k \mathbf{x}_k + \boldsymbol{\mu}_k, (u_k \boldsymbol{\Lambda}_k)^{-1}\right), \\ p(u_k|\nu_k) &= G(u_k|\nu_k/2, \nu_k/2). \end{aligned} \tag{5}$$

Parameter estimation under the Student- $t$  measurement model can thus be formulated under the Bayesian inference framework. One-step predicted PDF  $p(\mathbf{x}_k|\mathbf{z}_{1:k-1})$  is assumed to follow a Gaussian distribution:

$$p(\mathbf{x}_k|\mathbf{z}_{1:k-1}) = N(\mathbf{x}_k | \hat{\mathbf{x}}_{k|k-1}, \mathbf{P}_{k|k-1}). \tag{6}$$

The prior distribution for DOF  $\nu_k$  is assumed to follow a Gamma distribution [11]:

$$p(\nu_k) = G(\nu_k | a_{k|k-1}, b_{k|k-1}). \tag{7}$$

As given in [23], the joint conjugate prior distribution for mean vector  $\boldsymbol{\mu}_k$  and scale matrix  $\boldsymbol{\Lambda}_k$  follows a Gaussian–Wishart distribution and is factorized as their constituting distributions:

$$p(\boldsymbol{\mu}_k, \boldsymbol{\Lambda}_k) = p(\boldsymbol{\mu}_k|\boldsymbol{\Lambda}_k)p(\boldsymbol{\Lambda}_k) = N\left(\boldsymbol{\mu}_k \mid \boldsymbol{\eta}_k, (\beta_{k|k-1} \boldsymbol{\Lambda}_k)^{-1}\right) W(\boldsymbol{\Lambda}_k | \lambda_{k|k-1}, \mathbf{U}_{k|k-1}), \tag{8}$$

where  $\beta_{k|k-1}$  is the precision factor,  $W(\cdot|\lambda_k, \mathbf{U}_k)$  denotes the Wishart distribution with  $\lambda_{k|k-1}$  DOF and scale matrix  $\mathbf{U}_{k|k-1}$  [6]. In the sequel, we investigate a VB approach to approximate the posterior densities of the measurement parameters and state for the system in Eq. (1) to minimize the Kullback–Leibler divergence.

### 3 Estimation Using VB Approach

VB is an iterative approach commonly studied in machine learning and statistics and approximates the posterior distribution as a factored distribution. The parameters of the approximated factored distribution are adjusted by minimizing the Kullback–Leibler divergence between the selected density and the true density. Approximate posterior density  $q(\theta_k)$  of parameter  $\theta_k$  is determined by [6]

$$q(\theta_k) = \exp\left(E_{q(\theta_{l \neq k})}[\ln p(\mathbf{X}, \Theta)]\right), \tag{9}$$

where  $E_{q(\theta_{l \neq k})}[\cdot]$  denotes the expectation with respect to each  $\theta_l$  such that  $l \neq k$ ,  $\ln(\cdot)$  denotes the natural logarithm, and  $q(\theta_k)$  is the variational posterior distribution (VPD) of  $\theta_k$ .

Given that VB requires the joint distribution of measurements and parameters, their joint PDF is given and factorized as

$$\begin{aligned} & p(\mathbf{x}_k, \boldsymbol{\mu}_k, u_k, \boldsymbol{\Lambda}_k, v_k | \mathbf{z}_{1:k}) \\ & \propto p(\mathbf{x}_k | \mathbf{z}_{1:k-1}) p(\boldsymbol{\mu}_k | \boldsymbol{\Lambda}_k) p(\boldsymbol{\Lambda}_k) p(u_k | v_k) p(v_k | \mathbf{z}_{1:k-1}) p(\mathbf{z}_k | \mathbf{x}_k, \boldsymbol{\mu}_k, u_k, \boldsymbol{\Lambda}_k). \end{aligned} \tag{10}$$

Considering Eqs. (5), (6), and (7), joint posterior PDF  $p(\mathbf{x}_k, \boldsymbol{\mu}_k, u_k, \boldsymbol{\Lambda}_k, v_k | \mathbf{z}_{1:k})$  in Eq. (10) has no analytical solution. Thus, a variational method is utilized. Let

$$p(\mathbf{x}_k, \boldsymbol{\mu}_k, u_k, \boldsymbol{\Lambda}_k, v_k | \mathbf{z}_{1:k}) \approx q(\mathbf{x}_k)q(\boldsymbol{\mu}_k)q(u_k)q(\boldsymbol{\Lambda}_k)q(v_k), \tag{11}$$

where  $q(\mathbf{x}_k)$ ,  $q(\boldsymbol{\mu}_k)$ ,  $q(u_k)$ ,  $q(\boldsymbol{\Lambda}_k)$ , and  $q(v_k)$  denote, respectively, the approximate posterior PDFs of state  $\mathbf{x}_k$ , mean  $\boldsymbol{\mu}_k$ , auxiliary parameter  $u_k$ , scale matrix  $\boldsymbol{\Lambda}_k$ , and DOF  $v_k$ . These VPDs are coupled to each other and cannot be directly derived from Eq. (11), but fixed-point iteration [11] can be used to solve them.

#### 3.1 State and Parameter Estimation Under Student-t Measurement Noise

The optimal solution in Eq. (9) and the PDF decomposition in Eq. (11) allow to calculate the approximate posterior PDF of  $\mathbf{x}_k$  as

$$\begin{aligned} \ln q_i(\mathbf{x}_k) = & -\frac{1}{2} E_{q_{i-1}(\boldsymbol{\mu})q_{i-1}(\boldsymbol{\Lambda})} \left[ \boldsymbol{\xi}_k^T \left( u_k^{(i-1)} \boldsymbol{\Lambda}_k \right) \boldsymbol{\xi}_k \right] \\ & - \frac{1}{2} (\mathbf{x}_k - \hat{\mathbf{x}}_{k|k-1})^T \mathbf{P}_{k|k-1}^{-1} (\mathbf{x}_k - \hat{\mathbf{x}}_{k|k-1}) + C_x, \end{aligned} \tag{12}$$

where  $\boldsymbol{\xi}_k = \mathbf{z}_k - \mathbf{H}_k \mathbf{x}_k - \boldsymbol{\mu}_k$ ,  $u_k^{(i-1)} = E_{q_{i-1}(u_k)}[u_k]$ , and  $q_{i-1}(\cdot)$  represents the posterior density at the  $(i - 1)$ -th iteration. Taking expectations with respect to the

parameters enclosed in square brackets in Eq. (12) and expressing the terms independent from  $\mathbf{x}_k$  as constant  $C_x$  results in the simplified expression:

$$\begin{aligned} \ln q_i(\mathbf{x}_k) &= \frac{1}{2} \boldsymbol{\zeta}_k^T \left( u_k^{(i-1)} E_{q_{i-1}(\Lambda_k)} [\Lambda_k] \right) \boldsymbol{\zeta}_k \\ &\quad - \frac{1}{2} (\mathbf{x}_k - \hat{\mathbf{x}}_{k|k-1})^T \mathbf{P}_{k|k-1}^{-1} (\mathbf{x}_k - \hat{\mathbf{x}}_{k|k-1}) + C, \end{aligned} \tag{13}$$

where  $\boldsymbol{\zeta}_k = \mathbf{z}_k - \mathbf{H}_k \mathbf{x}_k - E_{q_{i-1}(\boldsymbol{\mu}_k)}[\boldsymbol{\mu}_k]$ . Equation (13) is quadratic in  $\mathbf{x}_k$ , and hence  $q_i(\mathbf{x}_k)$  is Gaussian, i.e.,  $q_i(\mathbf{x}_k) = N(\mathbf{x}_k | \hat{\mathbf{x}}_{k|k}^{(i)}, \mathbf{P}_{k|k}^{(i)})$ , with  $\hat{\mathbf{x}}_{k|k}^{(i)}, \mathbf{P}_{k|k}^{(i)}$  being the first two moments of the distribution obtained as

$$\hat{\mathbf{x}}_{k|k}^{(i)} = \hat{\mathbf{x}}_{k|k-1} + \mathbf{K}_k \left( \mathbf{z}_k - \mathbf{H}_k \hat{\mathbf{x}}_{k|k-1} - E_{q_{i-1}(\boldsymbol{\mu}_k)}[\boldsymbol{\mu}_k] \right), \tag{14}$$

$$\mathbf{P}_{k|k}^{(i)} = \mathbf{P}_{k|k-1} - \mathbf{K}_k \mathbf{H}_k \mathbf{P}_{k|k-1}, \tag{15}$$

with Kalman gain

$$\mathbf{K}_k = \mathbf{P}_{k|k-1} \mathbf{H}_k^T \left( \left( u_k^{(i-1)} E_{q_{i-1}(\Lambda_k)} [\Lambda_k] \right)^{-1} + \mathbf{H}_k \mathbf{P}_{k|k-1} \mathbf{H}_k^T \right)^{-1}. \tag{16}$$

To derive the VPD of  $\Lambda_k$ , the logarithm of the joint VPD of  $(\boldsymbol{\mu}_k, \Lambda_k)$  is first computed as

$$\begin{aligned} \ln q_i(\boldsymbol{\mu}_k, \Lambda_k) &= E_{q_{i-1}(u_k) q_{i-1}(v_k) q_i(\mathbf{x}_k)} [\ln p(\mathbf{z}_k | \mathbf{x}_k, \boldsymbol{\mu}_k, u_k, \Lambda_k) + \ln p(\boldsymbol{\mu}_k | \Lambda_k) + \ln p(\Lambda_k)] + C \\ &= \frac{1}{2} E_{q_{i-1}(u_k)} [u_k^m] \ln |\Lambda_k| + \frac{1}{2} \ln |\beta_{k|k-1} u_k^{(i-1)} \Lambda_k| + \frac{\lambda_{k|k-1} - m - 1}{2} \ln |\Lambda_k| \\ &\quad - \frac{u_k^{(i-1)}}{2} E_{q_i(\mathbf{x}_k)} [\boldsymbol{\xi}_k^T \Lambda_k \boldsymbol{\xi}_k] - \frac{1}{2} \text{Tr} [(\mathbf{U}_{k|k-1})^{-1} \Lambda_k] \\ &\quad - \frac{1}{2} (\boldsymbol{\mu}_k - \boldsymbol{\eta}_{k|k-1})^T (\beta_{k|k-1} u_k^{(i-1)} \Lambda_k) (\boldsymbol{\mu}_k - \boldsymbol{\eta}_{k|k-1}) + C_{\boldsymbol{\mu}_k, \Lambda_k}. \end{aligned} \tag{17}$$

Then, approximate densities  $q_i(\boldsymbol{\mu}_k)$  and  $q_i(\Lambda_k)$  are derived, and by the conditional PDF of  $\Lambda_k$  in Eq. (8), the approximate logarithm of the posterior distribution of  $\boldsymbol{\mu}_k$  is given by

$$\begin{aligned} \ln q_i(\boldsymbol{\mu}_k | \Lambda_k) &= E_{q_{i-1}(u_k) q_{i-1}(v_k) q_i(\mathbf{x}_k)} [\ln p(\mathbf{z}_k | \mathbf{x}_k, \boldsymbol{\mu}_k, u_k, \Lambda_k) + \ln p(\boldsymbol{\mu}_k | \Lambda_k)] + C \\ &= - \frac{u_k^{(i-1)}}{2} E_{q_i(\mathbf{x}_k)} [\boldsymbol{\xi}_k^T \Lambda_k \boldsymbol{\xi}_k] + \frac{1}{2} E_{q_{i-1}(u_k)} [\ln |u_k \Lambda_k|] + \frac{1}{2} \ln |\beta_{k|k-1} \Lambda_k| \\ &\quad - \frac{1}{2} (\boldsymbol{\mu}_k - \boldsymbol{\eta}_{k|k-1})^T (\beta_{k|k-1} u_k^{(i-1)} \Lambda_k) (\boldsymbol{\mu}_k - \boldsymbol{\eta}_{k|k-1}) + C \\ &= - \frac{1}{2} \boldsymbol{\mu}_k^T \left( u_k^{(i-1)} + \beta_{k|k-1} u_k^{(i-1)} \right) \Lambda_k \boldsymbol{\mu}_k \\ &\quad + \boldsymbol{\mu}_k^T \Lambda_k \left[ u_k^{(i-1)} (\mathbf{z}_k - \mathbf{H}_k E_{q_i(\mathbf{x}_k)}[\mathbf{x}_k]) + \beta_{k|k-1} u_k^{(i-1)} \boldsymbol{\eta}_{k|k-1} \right] + C_{\boldsymbol{\mu}}, \end{aligned} \tag{18}$$

which is quadratic in  $\boldsymbol{\mu}_k$ , and  $q_i(\boldsymbol{\mu}_k | \boldsymbol{\Lambda}_k)$  is Gaussian:

$$q_i(\boldsymbol{\mu}_k | \boldsymbol{\Lambda}_k) = N\left(\boldsymbol{\mu}_k \mid \boldsymbol{\eta}_{k|k}^{(i)}, \left(\beta_{k|k}^{(i)} \boldsymbol{\Lambda}_k\right)^{-1}\right), \tag{19}$$

with parameters

$$\beta_{k|k}^{(i)} = \beta_{k|k-1} + u_k^{(i-1)}, \tag{20}$$

$$\boldsymbol{\eta}_{k|k}^{(i)} = \boldsymbol{\eta}_{k|k-1} + u_k^{(i-1)} \left(\beta_{k|k}^{(i)}\right)^{-1} \left(\mathbf{z}_k - \mathbf{H}_k \hat{\mathbf{x}}_{k|k}^{(i)} - \boldsymbol{\eta}_{k|k-1}\right). \tag{21}$$

Note that if  $u_k^{(i-1)} \equiv 1$  in Eqs. (20) and (21), these equations become a recursive formula for the mean parameter deduced by VB under Gaussian measurement distribution.

The posterior distribution of  $\ln q_i(\boldsymbol{\Lambda}_k)$  is obtained as the difference between Eqs. (17) and (18):

$$\begin{aligned} \ln q_i(\boldsymbol{\Lambda}_k) = & -\frac{m}{2} \ln |\beta_{k|k}| - \frac{1}{2} \ln |\boldsymbol{\Lambda}_k| - \frac{\lambda_{k|k-1} - m - 1}{2} \ln |\boldsymbol{\Lambda}_k| \\ & - \frac{1}{2} u_k^{(i-1)} \text{Tr} \left\{ \left[ E_{q_i(\mathbf{x}_k)} \left[ \boldsymbol{\xi}_k \boldsymbol{\xi}_k^T \right] + (\mathbf{U}_{k|k-1})^{-1} + \beta_{k|k-1} \right. \right. \\ & \left. \left. \cdot (\boldsymbol{\mu}_k - \boldsymbol{\eta}_{k|k-1})(\boldsymbol{\mu}_k - \boldsymbol{\eta}_{k|k-1})^T - \beta_{k|k}^{(i)} (\boldsymbol{\mu}_k - \boldsymbol{\eta}_{k|k}^{(i)})(\boldsymbol{\mu}_k - \boldsymbol{\eta}_{k|k}^{(i)})^T \right] \boldsymbol{\Lambda}_k \right\} + C_{\boldsymbol{\Lambda}}. \end{aligned} \tag{22}$$

By conjugate prior, the VPD of  $\boldsymbol{\Lambda}_k$  can be conveniently written as a Wishart distribution:

$$q_i(\boldsymbol{\Lambda}_k) = \mathbf{W}\left(\boldsymbol{\Lambda}_k \mid \lambda_{k|k}^{(i)}, \mathbf{U}_{k|k}^{(i)}\right), \tag{23}$$

where the hyperparameters are obtained by substituting Eqs. (20) and (21) into Eq. (22):

$$\lambda_{k|k}^{(i)} = \lambda_{k|k-1} + 1, \tag{24}$$

$$\begin{aligned} \left(\mathbf{U}_{k|k}^{(i)}\right)^{-1} = & \left(\mathbf{U}_{k|k-1}\right)^{-1} + u_k^{(i-1)} \mathbf{H}_k \mathbf{P}_{k|k}^{(i)} \mathbf{H}_k^T \\ & + u_k^{(i-1)} \left(1 - \left(\beta_{k|k}^{(i)}\right)^{-1} u_k^{(i-1)}\right) \left(\mathbf{z}_k - \mathbf{H}_k \hat{\mathbf{x}}_{k|k}^{(i)} - \boldsymbol{\eta}_{k|k-1}\right) \left(\mathbf{z}_k - \mathbf{H}_k \hat{\mathbf{x}}_{k|k}^{(i)} - \boldsymbol{\eta}_{k|k-1}\right)^T. \end{aligned} \tag{25}$$

Note that if  $u_k^{(i-1)}$  in the right-hand side of Eq. (25) equals one, this equation becomes a recursive form for the scale matrix of the corresponding Gaussian distribution.

VPD  $q_i(u_k)$  is obtained as

$$\begin{aligned} \ln q_i(u_k) &= E_{q_i(\mu_k) q_i(\Lambda_k) q_i(x_k) q_{i-1}(v_k)} [\ln p(z_k | \mathbf{x}_k, \mu_k, u_k, \Lambda_k) + \ln p(u_k | v_k)] + C_u \\ &= \left( \frac{E_{q_{i-1}(v_k)}[v_k]}{2} - 1 \right) \ln u_k - \frac{E_{q_{i-1}(v_k)}[v_k]}{2} u_k + \frac{m}{2} \ln u_k \\ &\quad + \frac{1}{2} \ln |\mathbf{A}_k| - \frac{1}{2} E_{q_i(\mu_k) q_i(\Lambda_k) q_i(x_k)} \left[ \xi_k^T (u_k \mathbf{A}_k) \xi_k^T \right] + C_u \\ &= \left( \frac{E_{q_{i-1}(v_k)}[v_k] - 2 + m}{2} \right) \ln u_k - \frac{E_{q_{i-1}(v_k)}[v_k]}{2} u_k \\ &\quad - \frac{1}{2} E_{q_i(\mu_k) q_i(\Lambda_k) q_i(x_k)} \left[ \xi_k^T (u_k \mathbf{A}_k) \xi_k^T \right] + C_u. \end{aligned} \tag{26}$$

By conjugate prior,  $q_i(u_k)$  is Gamma distribution  $q_i(u_k) = G(u_k | v_k^{1(i)}, v_k^{2(i)})$ , where

$$v_k^{1(i)} = \frac{1}{2} (E_{q_{i-1}(v_k)}[v_k] + m), \tag{27}$$

$$v_k^{2(i)} = \frac{1}{2} \left( E_{q_{i-1}(v_k)}[v_k] + E_{q_i(\mu_k) q_i(\Lambda_k) q_i(x_k)} \left[ \xi_k^T \mathbf{A}_k \xi_k^T \right] \right). \tag{28}$$

The computation of  $E_{q_i(\mu_k) q_i(\Lambda_k) q_i(x_k)} \left[ \xi_k^T \mathbf{A}_k \xi_k^T \right]$  is given by

$$\begin{aligned} &E_{q_i(\mu_k) q_i(\Lambda_k) q_i(x_k)} \left[ (z_k - \mathbf{H}_k \mathbf{x}_k - \mu_k)^T \mathbf{A}_k (z_k - \mathbf{H}_k \mathbf{x}_k - \mu_k) \right] \\ &= \text{Tr} \left[ E_{q_i(\Lambda_k)} [\mathbf{A}_k] \left( \mathbf{y}_k \mathbf{y}_k^T - \mathbf{y}_k \left( \hat{\mathbf{x}}_{k|k}^{(i)} \right)^T \mathbf{H}_k^T - \mathbf{H}_k \hat{\mathbf{x}}_{k|k}^{(i)} \mathbf{y}_k^T \right. \right. \\ &\quad \left. \left. + \mathbf{H}_k \left( D_{q_i(x_k)} [\mathbf{x}_k] + \hat{\mathbf{x}}_{k|k}^{(i)} \left( \hat{\mathbf{x}}_{k|k}^{(i)} \right)^T \right) \mathbf{H}_k^T \right) \right] + m \left( \beta_{k|k}^{(i)} \right)^{-1}, \end{aligned} \tag{29}$$

where  $\mathbf{y}_k = z_k - \boldsymbol{\eta}_{k|k}^{(i)}$  and  $D_{q_i(x_k)}[\cdot]$  denotes the variance of  $\mathbf{x}_k$ .

Next, we derive approximate density  $q_i(v_k)$  by using the conditional PDF of  $u_k$  in Eq. (5) and the prior distribution in Eq. (7):

$$\begin{aligned} \ln q_i(v_k) &= E_{q_i(u_k)} [\ln p(u_k | v_k) + \ln p(v_k)] + C_{v_k} \\ &= \frac{v_k}{2} \ln \frac{v_k}{2} - \ln \Gamma \left( \frac{v_k}{2} \right) + \left( \frac{v_k}{2} - 1 \right) E_{q_i(u_k)} [\ln u_k] - \frac{v_k}{2} E_{q_i(u_k)} [u_k] \\ &\quad + (a_{k|k-1} - 1) \ln v_k - b_{k|k-1} v_k + C_{v_k}. \end{aligned} \tag{30}$$

Using Stirling’s approximation  $\ln \Gamma \left( \frac{v_k}{2} \right) \approx \frac{v_k-1}{2} \ln \frac{v_k}{2} - \frac{v_k}{2}$  in Eq. (30), the variational estimator of DOF  $v_k$  follows Gamma distribution  $q_i(v_k) = G(v_k | a_{k|k}^{(i)}, b_{k|k}^{(i)})$ , where

$$a_{k|k}^{(i)} = a_{k|k-1} + \frac{1}{2}, \tag{31}$$

$$b_{k|k}^{(i)} = b_{k|k-1} + \frac{1}{2} u_k^{(i)} - \frac{1}{2} E_{q_i(u_k)} [\ln u_k] - \frac{1}{2}. \tag{32}$$



As indicated above, the approximate posterior distribution of each variable depends on the expected values of some of the others. The expectations of the parameters are given by

$$\begin{aligned}
 E_{q_i(\mathbf{A}_k)}[\mathbf{A}_k] &= \lambda_{k|k}^{(i)} \mathbf{U}_{k|k}^{(i)}, \\
 E_{q_i(\boldsymbol{\mu}_k)}[\boldsymbol{\mu}_k] &= \boldsymbol{\eta}_{k|k}^{(i)}, \\
 E_{q_i(u_k)}[u_k] &= u_k^{(i)} = v_k^{1(i)} / v_k^{2(i)}, \\
 E_{q_i(u_k)}[\ln u_k] &= \Psi\left(v_k^{1(i)}\right) - \ln v_k^{2(i)}, \\
 E[v_k] &= a_{k|k}^{(i)} / b_{k|k}^{(i)},
 \end{aligned} \tag{33}$$

where  $\Psi(\cdot)$  denotes the digamma function, i.e., the first derivative of the natural logarithm of the gamma function.

The selection of number of iterations  $N$  is discussed next, as this parameter determines the estimation accuracy and implementation time. Here, we use a stopping criterion based on the difference in values from two consecutive estimates:

$$\delta_k = \begin{cases} < 10^{-6} & \text{terminate process} \\ \text{otherwise} & \text{continue iterative process} \end{cases}, \tag{34}$$

where the difference is given by  $\delta_k \sim \left\| \hat{\mathbf{x}}_{k|k}^{(i)} - \hat{\mathbf{x}}_{k|k}^{(i-1)} \right\|^2$ .

### 3.2 Hyperparameter Update

State update is the same as that of the standard KF. Considering only the update of unknown parameters, similar to [16, 24], we introduce factor  $\rho$  that indicates the statistics of fluctuation. Uncertainty reduces as  $\rho$  approaches to zero more, and hence we considered a value of  $\rho$  not very small for consistency. The updating of noise parameters is expressed as

$$\begin{aligned}
 \lambda_{k|k-1} &= \rho \lambda_{k-1|k-1}, \\
 \beta_{k|k-1} &= \rho \beta_{k-1|k-1}, \\
 a_{k|k-1} &= \rho a_{k-1|k-1}, \\
 b_{k|k-1} &= \rho b_{k-1|k-1}, \\
 \boldsymbol{\eta}_{k|k-1} &= \rho \boldsymbol{\eta}_{k-1|k-1}, \\
 \mathbf{U}_{k|k-1} &= \frac{1}{\rho} \mathbf{U}_{k-1|k-1}.
 \end{aligned} \tag{35}$$

### 3.3 Filter Algorithm

We summarize the estimation process of the proposed filter in Algorithm 1 and illustrate the hierarchical Gaussian model in Fig. 1.

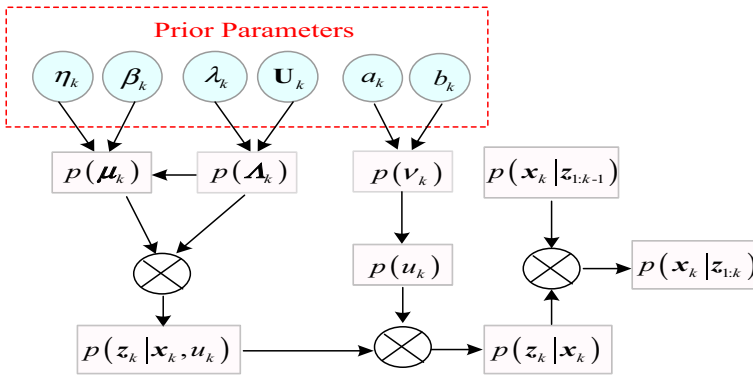


Fig. 1 Diagram of proposed Student-*t* hierarchical Gaussian state-space model

Note that if latent variable  $u_k \equiv 1$  and the recursive expression for  $v_k$  are omitted, the proposed algorithm degenerates into the estimation under Gaussian measurement noise. In addition, if the recursive expressions for  $\beta_{k|k}$  and  $\eta_k$  are omitted, the proposed algorithm degenerates into the estimation without mean under Student-*t* measurement noise [24].

### 4 Error Bounds of VB Estimator

The fundamental CRLB sets a lower limit on the MSE for any estimator  $\hat{\theta}(z)$ . Specifically, the Bayesian Cramer–Rao inequality [12, 20] shows that the MSE of any estimator  $\hat{\theta}(z)$  is lower bounded by

$$E_{p(z,\theta)} \left\{ \left( \hat{\theta}(z) - \theta \right) \left( \hat{\theta}(z) - \theta \right)^T \right\} \geq \mathbf{J}^{-1}, \tag{36}$$

where  $\theta$  is an  $r$ -dimensional estimated random parameter and  $\hat{\theta}(z)$  is an estimate of  $\theta$ . For square matrices  $\mathbf{A}$  and  $\mathbf{B}$ ,  $\mathbf{A} \geq \mathbf{B}$  indicates that  $\mathbf{A} - \mathbf{B}$  is a positive definite matrix.  $\mathbf{J}$  is the  $r \times r$  (Fisher) information matrix that is computed by expectation to joint density  $p(z, \theta)$  and can be written as

$$\mathbf{J} = E \left[ -\nabla_{\mathbf{Z}} \nabla_{\Theta}^T \log p_{z,\theta}(\mathbf{Z}, \Theta) \right], \tag{37}$$

where  $\nabla_{\Theta} = \left[ \frac{\partial}{\partial \theta_1}, \dots, \frac{\partial}{\partial \theta_r} \right]^T$ . As  $p(\mathbf{Z})$  is the integral of  $p_{z,\theta}(\mathbf{Z}, \Theta)$  over  $\Theta$ , dependency on  $\Theta$  is removed, and therefore we obtain the following alternative expression for the information matrix:

$$J = E \left[ -\nabla_{\boldsymbol{\theta}} \nabla_{\boldsymbol{\theta}}^T \log p_{\boldsymbol{\theta}|z}(\boldsymbol{\theta}|\mathbf{Z}) \right]. \tag{38}$$

The traditional Bayesian CRLB for sequential estimation requires a double integration over the parameter and every measurement at each iteration, thus being computationally intensive. A posterior CRLB and a conditional posterior CRLB were investigated in [20] and [26], respectively, for a general multidimensional discrete-time filtering problem, providing a recursive approach for calculating the sequential Bayesian CRLB.

We now introduce the VB-CRLB to set the lower bound on the performance of the VB estimator given past measurement  $\mathbf{Z}_k$  up to time  $k$ . As the Bayesian approach considers random state  $\mathbf{x}$  and parameter  $\boldsymbol{\theta}$ , the joint logarithm PDF for  $\mathbf{x}_k$  and  $\boldsymbol{\theta}_k$  is computed using VB and factorized as

$$\ln p(\boldsymbol{\theta}_k, \mathbf{x}_k|\mathbf{Z}_k) \approx \ln q(\boldsymbol{\theta}_k|\mathbf{Z}_k) + \ln q(\mathbf{x}_k|\boldsymbol{\theta}_k, \mathbf{Z}_k), \tag{39}$$

where  $q_i(\mathbf{x}_k|\mathbf{Z}_k)$  and  $q_i(\boldsymbol{\theta}_k|\mathbf{Z}_k)$  are not exact marginal distributions of joint PDF  $p(\mathbf{x}_k, \boldsymbol{\theta}_k|\mathbf{Z}_k)$  but factors whose product is an approximate joint PDF, as indicated in Eq. (11). Assume that parameter  $\boldsymbol{\theta}$  is decomposed into two parts as  $[\boldsymbol{\theta}_\alpha^T \boldsymbol{\theta}_\beta^T]^T$ , information matrix  $J$  can be decomposed as the corresponding block matrix:

$$J = \begin{bmatrix} J_{\alpha\alpha} & J_{\alpha\beta} \\ J_{\beta\alpha} & J_{\beta\beta} \end{bmatrix}. \tag{40}$$

For clarity, Student- $t$  nominal covariance  $(\boldsymbol{\Lambda}_k)^{-1}$  should be expressed in terms of Gaussian covariance  $\boldsymbol{\Sigma}_k$  by introducing multiplier  $1/u_k^{(i)}$ :

$$\boldsymbol{\Sigma}_k = (\boldsymbol{\Lambda}_k)^{-1} / u_k^{(i)}, \tag{41}$$

with expectation

$$\hat{\boldsymbol{\Sigma}}_k^{(i)} = E_{q_i(\boldsymbol{\Sigma}_k)}[\boldsymbol{\Sigma}_k] = (E_{q_i(\boldsymbol{\Lambda}_k)}[\boldsymbol{\Lambda}_k])^{-1} / u_k^{(i)}. \tag{42}$$

For parameters  $\boldsymbol{\mu}$  and  $\boldsymbol{\Sigma}$ , by Eqs. (19), (23), and (41), joint VPD  $q_i(\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k) = NW(\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k | \beta_{k|k}^{(i)}, \boldsymbol{\eta}_{k|k}^{(i)}, \lambda_{k|k}^{(i)}, \mathbf{U}_{k|k}^{(i)})$  is given by

$$q_i(\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k) \propto |(\boldsymbol{\Sigma}_k)^{-1}|^{(\lambda_{k|k}^{(i)} - m)/2} \exp \left\{ -\frac{1}{2} \left[ \left( u_k^{(i)} \right)^{-1} (\boldsymbol{\Sigma}_k)^{-1} \cdot \left( \left( \mathbf{U}_{k|k}^{(i)} \right)^{-1} + \beta_{k|k}^{(i)} \left( \boldsymbol{\mu}_k - \boldsymbol{\eta}_{k|k}^{(i)} \right) \left( \boldsymbol{\mu}_k - \boldsymbol{\eta}_{k|k}^{(i)} \right)^T \right) \right] \right\}. \tag{43}$$

Let parameters  $\theta_1 = \mu_k$  and  $\theta_2 = \Sigma_k$ . By Eqs. (38) and (40), we have

$$J_{11,k(i)} = -E_{q_i(\mu_k, \Sigma_k)} \left[ \nabla_{\mu_k^T} \left[ \nabla_{\mu_k} \ln q_i(\mu_k, \Sigma_k) \right] \right], \tag{44}$$

and omitting terms independent of  $\mu$ , entry  $J_{11,k(i)}$  can be rewritten as

$$J_{11,k(i)} = -E_{q_i(\mu_k, \Sigma_k)} \left[ \nabla_{\mu_k^T} \left[ \nabla_{\mu_k} \left( -\frac{1}{2} (\mu_k - \eta_{k|k}^{(i)})^T (\beta_{k|k}^{(i)} \Lambda_k) (\mu_k - \eta_{k|k}^{(i)}) \right) \right] \right], \tag{45}$$

which is simplified as

$$J_{11,k(i)} = E_{q_i(\mu_k | \Lambda_k)} \left[ E_{q_i(\Lambda_k)} \left[ \beta_{k|k}^{(i)} \Lambda_k \right] \right]. \tag{46}$$

Substituting  $E_{q_i(\Lambda_k)}[\Lambda_k] = \lambda_{k|k}^{(i)} \mathbf{U}_{k|k}^{(i)}$  into Eq. (46) yields

$$J_{11,k(i)} = \beta_{k|k}^{(i)} \lambda_{k|k}^{(i)} \mathbf{U}_{k|k}^{(i)}, \tag{47}$$

indicating that the derived asymptotic Bayesian CRLB for  $\mu_k$  is  $ABCLB_i(\mu_k) = (J_{11,k(i)})^{-1}$ .

Next, the first non-diagonal entry is computed by

$$J_{21,k(i)} = -E_{q_i(\mu_k, \Sigma_k)} \left[ \nabla_{\mu_k^T} \left[ \nabla_{\Sigma_k} (\ln q_i(\mu_k, \Sigma_k)) \right] \right], \tag{48}$$

which is simplified by omitting terms unrelated to  $\Sigma_k$  as

$$J_{21,k(i)} = -E_{q_i(\mu_k, \Sigma_k)} \left[ \nabla_{\mu_k^T} \left[ \nabla_{\Sigma_k} \left( \frac{(\lambda_{k|k}^{(i)} - m)}{2} \ln |\Sigma_k^{-1}| - \frac{1}{2} \text{Tr} \left( \left[ (\mathbf{U}_{k|k}^{(i)})^{-1} + \beta_{k|k}^{(i)} (\mu_k - \eta_{k|k}^{(i)}) (\mu_k - \eta_{k|k}^{(i)})^T \right] (u_k^{(i)})^{-1} \Sigma_k^{-1} \right) \right] \right] \right]. \tag{49}$$

Let us recall some rules on derivatives and traces of matrices that facilitate the calculation of derivatives of function  $\ln q_i(\mu_k, \Sigma_k)$  presented in Eq. (49). Let  $A$  and  $B$  be invertible matrices and  $X$  a column vector, then [5]

$$\begin{aligned} \frac{\partial}{\partial A} \ln |A^{-1}| &= -\frac{\partial}{\partial A} \ln |A| = -\frac{1}{|A|} \frac{\partial |A|}{\partial A} = -\left(A^{-1}\right)^T, \\ \text{Tr}(AB) &= \text{Tr}(BA), \\ \frac{\partial}{\partial A} \text{Tr}(BA^{-1}) &= -\left(A^{-1}BA^{-1}\right)^T. \end{aligned} \tag{50}$$

The derivative of a matrix product with respect to a vector is given by

$$\frac{\partial \mathbf{A}(\mathbf{X})\mathbf{B}(\mathbf{X})}{\partial \mathbf{X}} = \frac{\partial \mathbf{A}(\mathbf{X})}{\partial \mathbf{X}}\mathbf{B}(\mathbf{X}) + [\mathbf{I}_n \otimes \mathbf{A}(\mathbf{X})] \frac{\partial \mathbf{B}(\mathbf{X})}{\partial \mathbf{X}}, \tag{51}$$

where  $\otimes$  denotes the Kronecker product and  $\mathbf{I}_n$  denotes the identity matrix of dimension  $n$ .

Using Eqs. (50), (49) is computed as

$$\mathbf{J}_{21,k(i)} = \frac{1}{2} E_{q_i(\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)} \left[ \nabla_{\boldsymbol{\mu}_k^T} \left[ \mathbf{T}(\boldsymbol{\Sigma}_k) + \boldsymbol{\Sigma}_k^{-1} \cdot \left(u_k^{(i)}\right)^{-1} \beta_{k|k}^{(i)} \mathbf{S}(\boldsymbol{\mu}_k) \cdot \boldsymbol{\Sigma}_k^{-1} \right] \right], \tag{52}$$

where  $\mathbf{T}(\boldsymbol{\Sigma}_k) = (m - \lambda_{k|k}^{(i)}) (\boldsymbol{\Sigma}_k^{-1})^T + \left(u_k^{(i)}\right)^{-1} \boldsymbol{\Sigma}_k^{-1} \left(\mathbf{U}_{k|k}^{(i)}\right)^{-1} \boldsymbol{\Sigma}_k^{-1}$  and  $\mathbf{S}(\boldsymbol{\mu}_k) = (\boldsymbol{\mu}_k - \boldsymbol{\eta}_{k|k}^{(i)}) (\boldsymbol{\mu}_k - \boldsymbol{\eta}_{k|k}^{(i)})^T$  denote the statistics about  $\boldsymbol{\Sigma}_k$  and  $\boldsymbol{\mu}_k$ , respectively. Using the matrix rule in Eqs. (51), (52) is obtained as

$$\begin{aligned} \mathbf{J}_{21,k(i)} &= \frac{1}{2} E_{q_i(\boldsymbol{\Sigma}_k)} \left( \left(u_k^{(i)}\right)^{-1} \beta_{k|k}^{(i)} \left(\mathbf{I}_m \otimes \boldsymbol{\Sigma}_k^{-1}\right)_{mm \times mm} \right. \\ &\quad \cdot E_{q_i(\boldsymbol{\mu}_k | \boldsymbol{\Sigma}_k)} \left[ \left( (\text{vec}(\mathbf{I}_m))_{mm \times 1} \otimes \left(\boldsymbol{\mu}_k - \boldsymbol{\eta}_{k|k}^{(i)}\right)^T \right)_{1 \times m} \right. \\ &\quad \left. \left. + \left(\mathbf{I}_m \otimes \left(\boldsymbol{\mu}_k - \boldsymbol{\eta}_{k|k}^{(i)}\right)\right)_{mm \times m} \right] \cdot \boldsymbol{\Sigma}_k^{-1} \right), \end{aligned} \tag{53}$$

where subscript  $a \times b$  denotes the dimension of the corresponding matrix and  $\text{vec}(\mathbf{I}_m)$  is a column vector obtained by stacking the consecutive columns of  $\mathbf{I}_m$ . Note that  $E_{q_i(\boldsymbol{\mu}_k | \boldsymbol{\Sigma}_k)} \left[ \left(\boldsymbol{\mu}_k - \boldsymbol{\eta}_{k|k}^{(i)}\right)^T \right] = E_{q_i(\boldsymbol{\mu}_k | \boldsymbol{\Sigma}_k)} \left[ \left(\boldsymbol{\mu}_k - \boldsymbol{\eta}_{k|k}^{(i)}\right) \right] = 0$ , and  $\mathbf{J}_{12,k(i)}$  has a similar term. Hence, the non-diagonal entries are both zeros, i.e.,  $\mathbf{J}_{21,k(i)} = \mathbf{0}_{mm \times m}$  and  $\mathbf{J}_{12,k(i)} = \mathbf{0}_{m \times mm}$ . As a result, information matrix  $\mathbf{J}$  is a block diagonal matrix, and the asymptotic Bayesian CRLB is its inverse.

As entry  $\mathbf{J}_{22,k(i)}$  is mathematically intractable for a general matrix  $\boldsymbol{\Sigma}_k$ , we consider  $\boldsymbol{\Sigma}_k$  to be a diagonal matrix, and hence  $\boldsymbol{\Lambda}_k$  is also diagonal:

$$(\boldsymbol{\Lambda}_k)^{-1} = \text{diag}([\bar{\sigma}_{k,1}^2, \bar{\sigma}_{k,2}^2, \dots, \bar{\sigma}_{k,m}^2]), \tag{54}$$

where the prior distribution of element  $\bar{\sigma}_{k,j}^2$  follows inverse Gamma distribution  $p(\bar{\sigma}_{k,j}^2) = \text{IG}(\bar{\sigma}_{k,j}^2 | \bar{\kappa}_{k|k-1,j}, \bar{\gamma}_{k|k-1,j})$ , which leads to the VPD being inverse Gamma distribution

$$q_i(\bar{\sigma}_{k,j}^2) = \text{IG}(\bar{\sigma}_{k,j}^2 | \bar{\kappa}_{k|k,j}^{(i)}, \bar{\gamma}_{k|k,j}^{(i)}) \tag{55}$$

with parameters

$$\begin{aligned} \bar{\kappa}_{k|k,j}^{(i)} &= \bar{\kappa}_{k|k-1,j} + \frac{1}{2}, \\ \bar{\gamma}_{k|k,j}^{(i)} &= \bar{\gamma}_{k|k-1,j} + \frac{u_k^{(i-1)}}{2} \left[ \mathbf{H}_k \mathbf{P}_{k|k}^{(i)} \mathbf{H}_k^T \right]_{jj} + \frac{u_k^{(i-1)}}{2} \left[ 1 - \left( \beta_{k|k}^{(i)} \right)^{-1} \right] \\ &\quad \times \left[ \left( \mathbf{z}_k - \mathbf{H}_k \hat{\mathbf{x}}_{k|k}^{(i)} - \boldsymbol{\eta}_{k|k-1} \right) \left( \mathbf{z}_k - \mathbf{H}_k \hat{\mathbf{x}}_{k|k}^{(i)} - \boldsymbol{\eta}_{k|k-1} \right)^T \right]_{jj}, \end{aligned} \tag{56}$$

and predictions  $\bar{\kappa}_{k+1|k,j} = \rho \bar{\kappa}_{k|k,j}$  and  $\bar{\gamma}_{k+1|k,j} = \rho \bar{\gamma}_{k|k,j}$ . In this case, the mean of individual parameter  $\boldsymbol{\mu}_{k,j}$  follows distribution

$$q_i \left( \boldsymbol{\mu}_{k,j} \mid \bar{\sigma}_{k,j}^2 \right) = N \left( \boldsymbol{\mu}_{k,j} \mid \boldsymbol{\eta}_{k|k,j}^{(i)}, \left( \beta_{k|k}^{(i)} \right)^{-1} \bar{\sigma}_{k(i),j}^2 \right), \tag{57}$$

resulting in  $\mathbf{J}_{11,k(i),j} = \beta_{k|k}^{(i)} \left( E \left[ \bar{\sigma}_{k(i),j}^2 \right] \right)^{-1} = \beta_{k|k}^{(i)} \left( \bar{\kappa}_{k|k,j}^{(i)} - 1 \right) / \bar{\gamma}_{k|k,j}^{(i)}$ . The lower bound for the estimation variance associated with the mean of individual parameter  $\boldsymbol{\mu}_{k,j}$  is given by

$$ABCLB_i \left( \boldsymbol{\mu}_{k,j} \right) = \mathbf{J}_{11,k(i),j}^{-1} = \bar{\gamma}_{k|k,j}^{(i)} / \left( \beta_{k|k}^{(i)} \left( \bar{\kappa}_{k|k,j}^{(i)} - 1 \right) \right), \tag{58}$$

and the lower bound on the variance of the individual parameter is computed from

$$\mathbf{J}_{22,k(i)} = -E_{q_i \left( \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k \right)} \left[ \nabla_{\boldsymbol{\Sigma}_k} \left[ \nabla_{\boldsymbol{\Sigma}_k} \left( \ln q_i \left( \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k \right) \right) \right] \right]_{mm \times mm}, \tag{59}$$

with joint variational posterior distribution being

$$\begin{aligned} q_i \left( \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k \right) &\propto \left[ \prod_{j=1}^m \left( \sigma_{k,j}^2 \right)^{-\left( \kappa_{k|k,j}^{(i)} + \frac{3}{2} \right)} \right] \\ &\quad \times \exp \left\{ -\frac{1}{2u_k^{(i-1)}} \sum_{j=1}^m \frac{\beta_{k|k}^{(i)} \left( \boldsymbol{\mu}_k - \boldsymbol{\eta}_{k|k}^{(i)} \right)_j^2 + 2\bar{\gamma}_{k|k,j}^{(i)}}{\sigma_{k,j}^2} \right\}, \end{aligned} \tag{60}$$

where  $\sigma_{k,j}^2$  is the  $jj$ -th element of diagonal matrix  $\boldsymbol{\Sigma}_k$  distributed according to  $q_i \left( \sigma_{k,j}^2 \right) = \text{IG} \left( \sigma_{k,j}^2 \mid \kappa_{k|k,j}^{(i)}, \gamma_{k|k,j}^{(i)} \right)$ , with  $\kappa_{k|k,j}^{(i)} = \bar{\kappa}_{k|k,j}^{(i)}$  and  $\gamma_{k|k,j}^{(i)} = \bar{\gamma}_{k|k,j}^{(i)} / u_k^{(i-1)}$ .

As  $\mathbf{J}_{22,k(i)}$  is a block diagonal matrix involving  $m$  subblocks, the nonzero  $jj$ -th element of the  $j$ -th subblock is

$$\begin{aligned}
 \mathbf{J}_{22,k(i)}^{jj} &= E_{q_i(\boldsymbol{\mu}_{k,j}, \sigma_{k,j}^2)} \left[ \nabla_{\sigma_{k,j}^2} \left[ \nabla_{\sigma_{k,j}^2} \left( \left( \kappa_{k|k,j}^{(i)} + \frac{3}{2} \right) \ln \sigma_{k,j}^2 \right. \right. \right. \\
 &\quad \left. \left. \left. + \frac{1}{2u_k^{(i-1)}} \frac{\beta_{k|k}^{(i)} (\boldsymbol{\mu}_k - \boldsymbol{\eta}_{k|k}^{(i)})_j^2 + 2\bar{\gamma}_{k|k,j}^{(i)}}{\sigma_{k,j}^2} \right) \right] \right] \tag{61} \\
 &= E_{q_i(\sigma_{k,j}^2)} \left[ E_{q_i(\boldsymbol{\mu}_{k,j} | \sigma_{k,j}^2)} \left[ -\frac{2\kappa_{k|k,j}^{(i)} + 3}{2(\sigma_{k,j}^2)^2} + \frac{\left( \beta_{k|k}^{(i)} (\boldsymbol{\mu}_k - \boldsymbol{\eta}_{k|k}^{(i)})_j^2 + 2\bar{\gamma}_{k|k,j}^{(i)} \right)}{u_k^{(i-1)} (\sigma_{k,j}^2)^3} \right] \right] \\
 &= -\frac{1}{2} E_{q_i(\sigma_{k,j}^2)} \left[ \frac{2\kappa_{k|k,j}^{(i)} + 3}{(\sigma_{k,j}^2)^2} \right] + E_{q_i(\sigma_{k,j}^2)} \left[ E_{q(\boldsymbol{\mu}_{k,j} | \sigma_{k,j}^2)} \left[ \frac{\left( \beta_{k|k}^{(i)} (\boldsymbol{\mu}_k - \boldsymbol{\eta}_{k|k}^{(i)})_j^2 + 2\bar{\gamma}_{k|k,j}^{(i)} \right)}{u_k^{(i-1)} (\sigma_{k,j}^2)^3} \right] \right]. \tag{62}
 \end{aligned}$$

Using the conditional covariance of  $\boldsymbol{\mu}_{k,j}$  in Eq. (57), we obtain

$$E_{q(\boldsymbol{\mu}_{k,j} | \sigma_{k,j}^2)} \left[ \beta_{k|k}^{(i)} (\boldsymbol{\mu}_k - \boldsymbol{\eta}_{k|k}^{(i)})_j^2 \right] = \beta_{k|k}^{(i)} D[\boldsymbol{\mu}_{k,j}] = u_k^{(i-1)} \sigma_{k(i),j}^2, \tag{63}$$

which substituted into (62) simplifies  $\mathbf{J}_{22,k(i)}^{jj}$  as

$$\mathbf{J}_{22,k(i)}^{jj} = -\frac{1}{2} E_{q_i(\sigma_{k,j}^2)} \left[ \frac{2\kappa_{k|k,j}^{(i)} + 1}{(\sigma_{k,j}^2)^2} \right] + E_{q_i(\sigma_{k,j}^2)} \left[ \frac{2\gamma_{k|k,j}^{(i)}}{(\sigma_{k,j}^2)^3} \right]. \tag{64}$$

Note that if random variable  $\alpha$  follows inverse Gamma distribution  $IG(\alpha|c, d)$ , where  $c$  and  $d$  are shape and scale parameters, respectively, the following general result about expectation is obtained:

$$E_{IG(\alpha)} \left[ \frac{1}{\alpha^n} \right] = d^{-n} \Gamma(c+n) / \Gamma(c). \tag{65}$$

If  $\alpha = \sigma_{k,j}^2$ ,  $c = \kappa_{k|k,j}^{(i)}$ ,  $d = \gamma_{k|k,j}^{(i)}$ , and  $n = 2$  we obtain

$$\begin{aligned}
 E_{q_i(\sigma_{k,j}^2)} \left[ \frac{1}{(\sigma_{k,j}^2)^2} \right] &= \Gamma(\kappa_{k|k,j}^{(i)} + 2) / \Gamma(\kappa_{k|k,j}^{(i)}) / (\gamma_{k|k,j}^{(i)})^2 \\
 &= \kappa_{k|k,j}^{(i)} (\kappa_{k|k,j}^{(i)} + 1) / (\gamma_{k|k,j}^{(i)})^2. \tag{66}
 \end{aligned}$$

Similarly, for  $n = 3$  we obtain

$$E_{q_i(\sigma_{k,j}^2)} \left[ \frac{2\gamma_{k|k,j}^{(i)}}{(\sigma_{k,j}^2)^3} \right] = \frac{2\kappa_{k|k,j}^{(i)} (\kappa_{k|k,j}^{(i)} + 1) (\kappa_{k|k,j}^{(i)} + 2)}{(\gamma_{k|k,j}^{(i)})^2}. \tag{67}$$

Substituting Eqs. (66) and (67) into Eq. (64) and performing simple algebraic manipulation, we obtain

$$J_{22,k(i)}^{jj} = \kappa_{k|k,j}^{(i)} (\kappa_{k|k,j}^{(i)} + 1) (2\kappa_{k|k,j}^{(i)} + 7) / 2 (\gamma_{k|k,j}^{(i)})^2. \tag{68}$$

Hence, the asymptotic Bayesian CRLB on  $\sigma_{k,j}^2$  is  $ABCLB_i(\sigma_{k,j}^2) = (J_{22,k(i)}^{jj})^{-1}$ .

To determine the asymptotic Bayesian CRLB of parameter  $\nu$ , we use the logarithm of the VPD in Eq. (30). By Eq. (38), the VB information value of  $\nu$  is

$$VBI_{k(i)}(\nu_k) = -E_{q_i(\nu_k)} \left[ \frac{\partial^2 \ln q_i(\nu_k)}{\partial \nu_k^2} \right]. \tag{69}$$

Omitting the term not containing  $\nu_k$ , the first derivative of  $\ln q(\nu_k)$  is given by

$$\frac{\partial \ln q_i(\nu_k)}{\partial \nu_k} = \frac{1}{2} \ln \nu_k - \frac{\Gamma'(\nu_k/2)}{\Gamma(\nu_k/2)} + \frac{a_{k|k-1} - 1}{\nu_k}, \tag{70}$$

and its second derivative by

$$\frac{\partial^2 \ln q_i(\nu_k)}{\partial \nu_k^2} = \frac{1}{2} \frac{1}{\nu_k} + \frac{1 - a_{k|k-1}}{\nu_k^2} - \Psi' \left( \frac{\nu_k}{2} \right), \tag{71}$$

where  $\Psi'(\cdot)$  denotes the first derivative of the digamma function. Then,  $a_{k|k}^{(i)}$  from Eq. (31) is replaced into Eq. (71) for Eq. (69) to become

$$VBI_{k(i)}(\nu_k) = -\frac{1}{2} E_{q_i(\nu_k)} \left[ \frac{1}{\nu_k} \right] - E_{q_i(\nu_k)} \left[ \frac{3/2 - a_{k|k}^{(i)}}{\nu_k^2} \right] + E_{q_i(\nu_k)} \left[ \Psi' \left( \frac{\nu_k}{2} \right) \right]. \tag{72}$$

Using the expectation of the Gamma distribution, we obtain

$$E_{q_i(\nu_k)} \left[ \frac{1}{\nu_k} \right] = b_{k|k}^{(i)} / (a_{k|k}^{(i)} - 1) \quad E_{q_i(\nu_k)} \left[ \frac{1}{\nu_k^2} \right] = (b_{k|k}^{(i)})^2 / [(a_{k|k}^{(i)} - 1)(a_{k|k}^{(i)} - 2)], \tag{73}$$

which substituted into Eq. (72) results in

$$VBI_{k(i)}(\nu_k) = E_{q_i(\nu_k)} \left[ \Psi' \left( \frac{\nu_k}{2} \right) \right] - \frac{1}{2} \frac{b_{k|k}^{(i)}}{(a_{k|k}^{(i)} - 1)} - \frac{(b_{k|k}^{(i)})^2 (3/2 - a_{k|k}^{(i)})}{(a_{k|k}^{(i)} - 1)(a_{k|k}^{(i)} - 2)}. \tag{74}$$



Hence, the asymptotic Bayesian CRLB on  $v$  is  $ABCLB_i(v_k) = 1/VBI_{k(i)}(v_k)$ .

For parameter  $\mathbf{x}_k$ , when its posterior distribution  $q(\boldsymbol{\theta}|\mathbf{Z})$  becomes  $p(\boldsymbol{\theta}|\mathbf{Z})$ ,  $\hat{\mathbf{x}}_{k|k}^{(i)}$  asymptotically attains the asymptotic Bayesian CRLB corresponding to the following Fisher information matrix:

$$\mathbf{J}_{k(i)}^{\mathbf{x}\mathbf{x}} = -E_{q_i(\mathbf{x})q_{i-1}(\boldsymbol{\theta})} \left[ \frac{\partial^2 \ln q(\boldsymbol{\theta}_k, \mathbf{x}_k, \mathbf{z}_k)}{\partial \mathbf{x}_k \partial \mathbf{x}_k^T} \right] = -E_{q_i(\mathbf{x})q_{i-1}(\boldsymbol{\theta})} \left[ \frac{\partial^2 \ln q(\mathbf{x}_k|\boldsymbol{\theta}_k, \mathbf{z}_k)}{\partial \mathbf{x}_k \partial \mathbf{x}_k^T} \right], \tag{75}$$

which is only function of  $q_i(\boldsymbol{\theta}_k)$  given the quadratic nature of  $\ln q(\mathbf{x}_k|\boldsymbol{\theta}_k, \mathbf{z}_k)$  in  $\mathbf{x}_k$ , and by Eq. (15) we obtain

$$\mathbf{J}_{k(i)}^{\mathbf{x}\mathbf{x}} = E_{q_{i-1}(\boldsymbol{\theta})} \left[ \left( \mathbf{P}_{k|k}^{(i)}(\boldsymbol{\theta}_k) \right)^{-1} \right], \tag{76}$$

where  $\boldsymbol{\theta}_k$  includes  $u_k$  and  $\boldsymbol{\Lambda}_k$  as shown in Eq. (16). Then, we obtain  $ABCLB_i(\mathbf{x}_k) = \left( \mathbf{J}_{k(i)}^{\mathbf{x}\mathbf{x}} \right)^{-1}$ , which is identical to the derived covariance matrix of  $\mathbf{x}_k$ , indicating that the VB update for  $\mathbf{x}_k$  is an extension of the minimum MSE estimation when the estimate is intractable for model parameter inaccuracies.

## 5 Results and Discussion

We conducted a numerical simulation to test the performance of the proposed method and verify the derived asymptotic Bayesian CRLB. Then, SINS/DVL integrated navigation was tested to further evaluate the proposed filter algorithm.

### 5.1 Stochastic Resonator Model

For the first simulation, we used a randomly drifting stochastic resonator [17], which is a typical signal detection model widely used in the field of sensor measurements in areas such as navigation information acquisition, spread spectrum communication, and biological instruments. The state–state model is expressed as

$$\begin{bmatrix} \mathbf{x}_{k,1} \\ \mathbf{x}_{k,2} \\ \mathbf{x}_{k,3} \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos(\omega\Delta t) & \frac{\sin(\omega\Delta t)}{\omega} \\ 0 & -\omega \sin(\omega\Delta t) & \cos(\omega\Delta t) \end{bmatrix} \begin{bmatrix} \mathbf{x}_{k-1,1} \\ \mathbf{x}_{k-1,2} \\ \mathbf{x}_{k-1,3} \end{bmatrix} + \mathbf{w}_k, \tag{77}$$

and the measurement model as

$$\mathbf{z}_k = [0 \ 0 \ 1] \begin{bmatrix} \mathbf{x}_{k,1} \\ \mathbf{x}_{k,2} \\ \mathbf{x}_{k,3} \end{bmatrix} + \mathbf{v}_k, \tag{78}$$

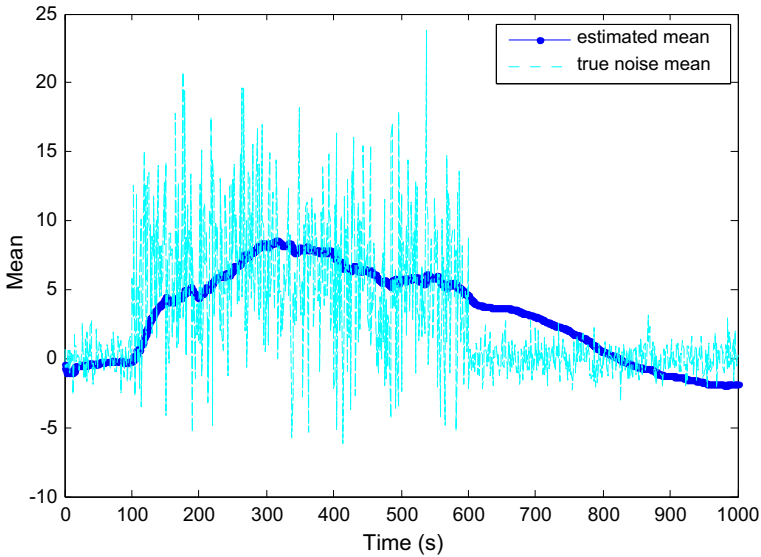


Fig. 2 True mean and estimated mean by applying the proposed filter to a stochastic resonator model

where  $\mathbf{w}_k = [\mathbf{w}_{k,1} \ \mathbf{w}_{k,2} \ \mathbf{w}_{k,3}]^T$ , measurement noise  $\mathbf{v}_k$  is an unknown random variable, sampling time  $\Delta t = 0.4$  s, and angular velocity  $\omega = 0.05$  rad/s. The initial state was assumed to follow Gaussian distribution  $\mathbf{x}_{0,i} \sim N(0, 1)$  for  $i = 1, 2, 3$ . In addition, process noise was considered as a Gaussian distribution with zero mean and covariance matrix  $\mathbf{Q}_0 = \text{diag}([0.01 \ 0.01 \ 0.01])$ , and the measurement noise was modeled as a Student's  $t$ -distribution.

We evaluated the robustness of the proposed algorithm considering Gaussian measurement noise with varying mean and variance over time. Specifically, for  $t \in [1 \text{ s}, 100 \text{ s}]$  and  $t \in [601 \text{ s}, 1000 \text{ s}]$ ,  $\mathbf{v}(t) \sim N(\mathbf{0}, \mathbf{R})$  with  $\mathbf{R} = 1$ , whereas for  $t \in [101 \text{ s}, 600 \text{ s}]$ ,  $\mathbf{v}(t) \sim N(6, 25 \times \mathbf{R})$ . Then, we performed state and parameter estimation for the mean and variance of the measurement noise using the proposed algorithm. We set the number of iterations for convergence to 10.

Figure 2 shows the estimated and true mean, indicating that the proposed filter can gradually track the true value from 100 to 600 s. Moreover, when the distribution changes at 600 s, the filter quickly adapts. Figure 3 depicts the variance estimate and compares it with that obtained from the VB-KF proposed in [17]. When the true variance changes, estimation using the proposed filter follows this change more quickly than VB-KF. However, when the variance remains unchanged from 100 to 600 s, the estimation of the proposed filter shows more intense oscillations than that of the VB-KF, indicating the lower consistency of the algorithm in this aspect compared to the VB-KF.

Figure 4 and Table 1, respectively, show the estimation error curves and root MSE (RMSE) of the states. For constant Gaussian noise, the state estimation results using the VB methods outperform those using the conventional KF, and the performance of

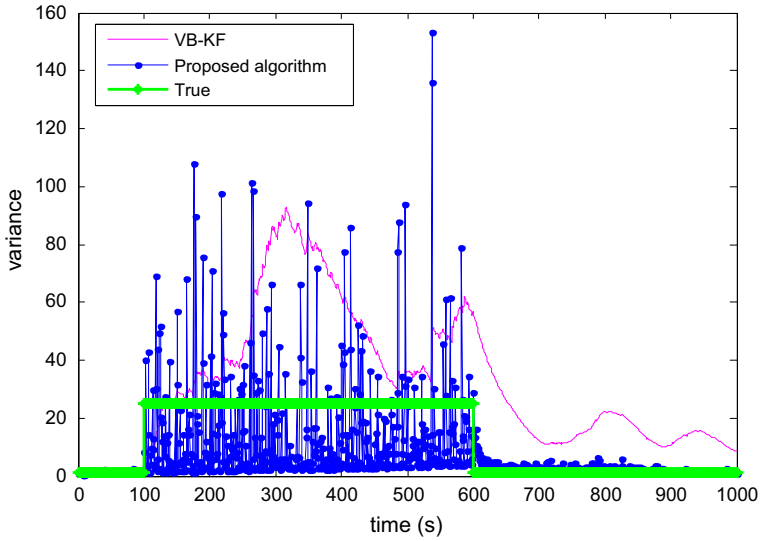


Fig. 3 True variance and estimated variance by applying the proposed filter to a stochastic resonator model

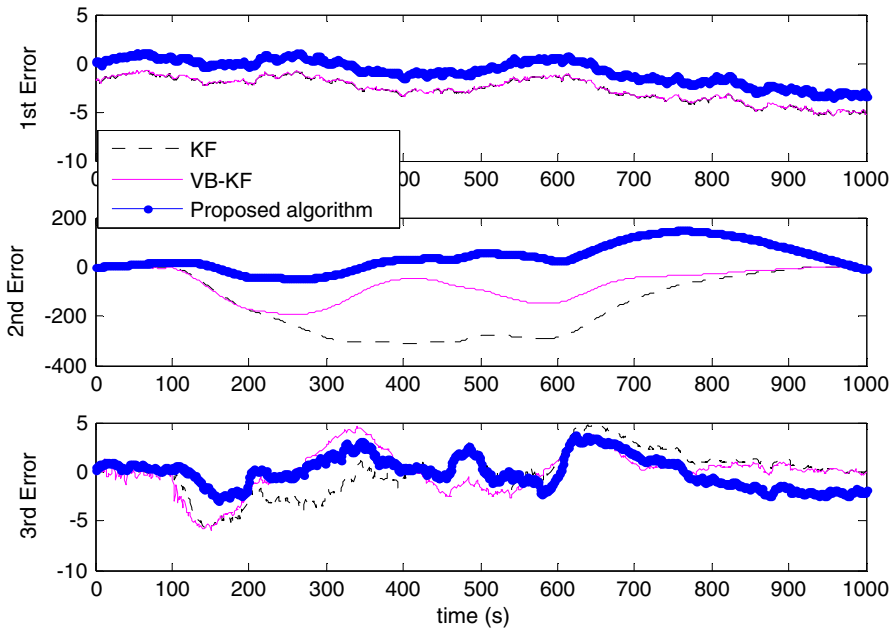


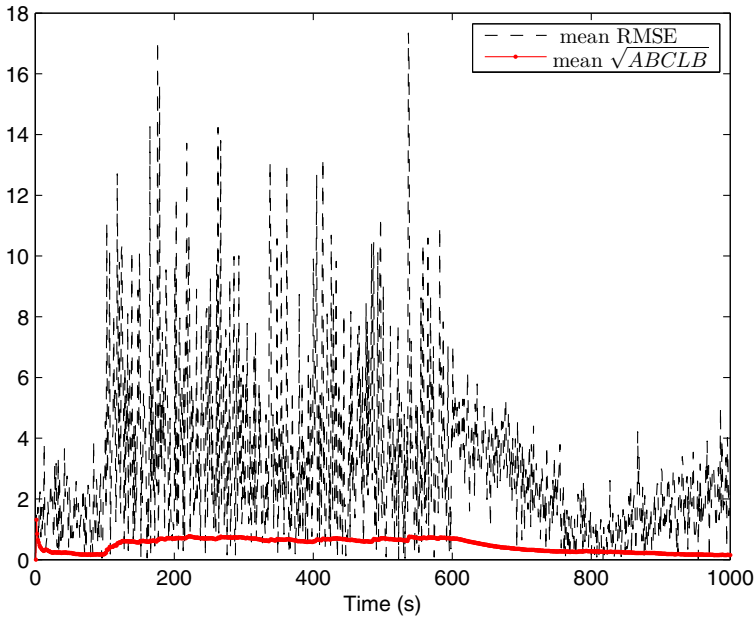
Fig. 4 State error from applying the proposed filter and similar methods to a stochastic resonator model

proposed filter is superior to that of VB-KF, which can be attributed to the inclusion of all noise parameters in the estimation of the proposed filter.

Figures 5 and 6 show the calculated RMSE of mean and variance, respectively, and their theoretical  $\sqrt{ABCLB}$ . The results verify that the asymptotic Bayesian CRLB

**Table 1** RMSE of states by applying the proposed filter and similar methods to a stochastic resonator model

Algorithm	1st RMSE	2nd RMSE	3rd RMSE
KF	2.9489	196.036	2.2248
Proposed algorithm	1.4577	69.0639	1.5864
VB-KF	2.9038	94.6053	2.0831

**Fig. 5** Mean and corresponding  $\sqrt{ABCLB}$  of simulations

is very close to the lower bounds of variance estimation for the posterior distributions of mean and variance under Gaussian and even non-Gaussian noise. Note that  $ABCLB$  provides the MSE performance bound of parameter estimation online.

## 5.2 Experiments on Navigation Data

We also evaluated the proposed filter on a SINS/DVL shipborne test for velocity estimation and compared it to the conventional KF and VB-KF. The experimental platform is composed of a SINS, a DVL, and a GPS receiver. The body angular rate and specific force were measured by gyroscopes and accelerometers, respectively, at a rate of 100 Hz, and GPS data were sampled at 1 Hz to provide accurate position and velocity information for the integrated SINS/DVL.

The filter was indirectly applied by considering navigation parameter errors as system state variables and using output correction to modify the SINS parameters. The dynamic model of the SINS/DVL navigation error is given by

$$\dot{X}(t) = A(t)X(t) + B(t)w(t), \quad (79)$$

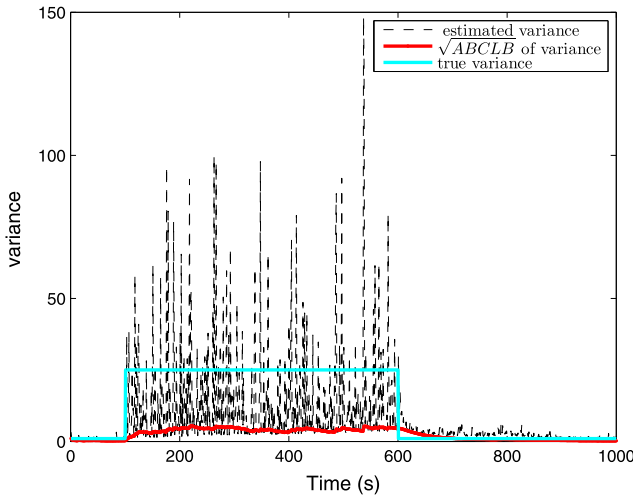


Fig. 6 Variance and corresponding  $\sqrt{ABCLB}$  of simulations

where state vector  $\mathbf{X}$  is composed of seven navigation error variables,  $\mathbf{A}$  is the state transition matrix of the system, and  $\mathbf{B}$  is the noise matrix, as detailed in [7]. Variable  $\mathbf{w}$  represents the Gaussian process noise with zero mean and covariance matrix  $\mathbf{Q}$ , representing the inertial sensor bias consisting of accelerometer and gyroscope biases, which are given by

$$\begin{aligned} \mathbf{X}(t) &= [\delta L \ \delta \lambda \ \delta V_e \ \delta V_n \ \varphi_e \ \varphi_n \ \varphi_u]^T \\ \mathbf{w}(t) &= [0 \ 0 \ w_{ax} \ w_{ay} \ w_{gx} \ w_{gy} \ w_{gz}]^T, \end{aligned} \tag{80}$$

where  $\delta L$  and  $\delta \lambda$  are the latitude and longitude errors, respectively,  $\delta V_e$  and  $\delta V_n$  are velocity errors, subscripts e, n, and u denote the east, north, and up components in the navigation frame,  $\varphi_e$ ,  $\varphi_n$ , and  $\varphi_u$  are the pitch, roll, and heading errors, respectively, and  $x$ ,  $y$ , and  $z$  denote the right, front, and up components in the body frame, respectively. The transformation from the body frame to the navigation frame is given by direction cosine matrix  $C_b^n$  in [7].

The measurement model utilizes a loosely coupled method. Hence, measurements are determined from the level velocity errors between the DVL and SINS, from which the measurement model is formulated as

$$\mathbf{z}(t) = \mathbf{H}\mathbf{x}(t) + \mathbf{v}(t) = \begin{bmatrix} v_{SINS}^e \\ v_{SINS}^n \end{bmatrix} - \begin{bmatrix} v_{DVL}^e \\ v_{DVL}^n \end{bmatrix}, \tag{81}$$

where  $\mathbf{z}$  is the measurement vector,  $\mathbf{H} = [\mathbf{0}_{2 \times 2} \ \mathbf{I}_{2 \times 2} \ \mathbf{0}_{2 \times 3}]$  is the measurement matrix, and  $\mathbf{v}$  is the measurement noise. The discretized form of dynamic Eq. (79) has the same form as the model in Eq. (1). The diagram of the SINS/DVL integrated navigation is shown in Fig. 7.

When measurements were available, we used the output of the filter to correct the solutions of the SINS. The voyage data from the sea vehicle were collected in the

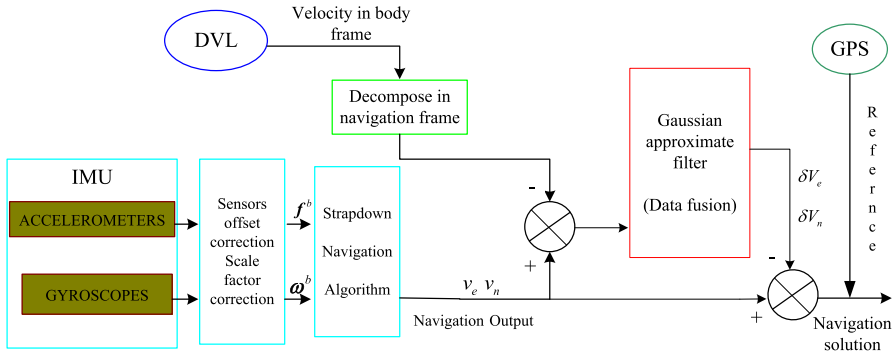


Fig. 7 Diagram of integrated navigation system for evaluating the proposed filter (IMU, inertial measurement unit)

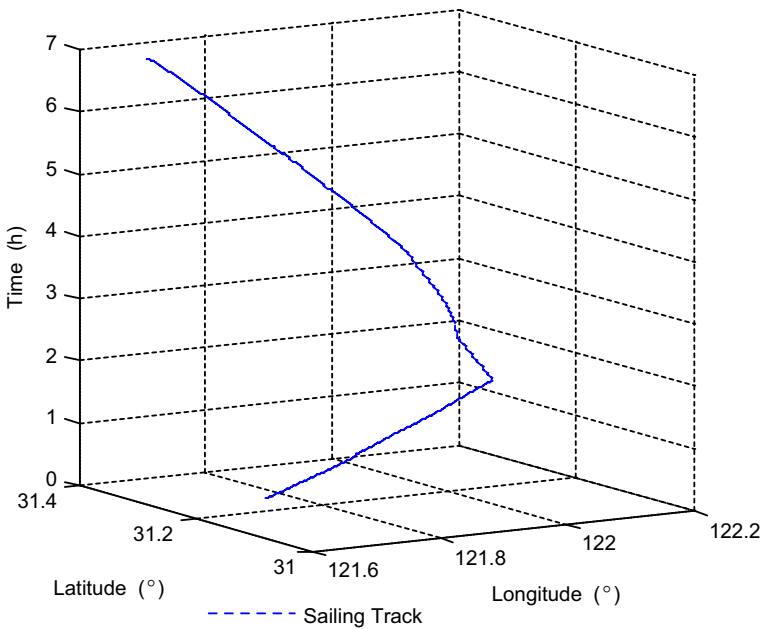


Fig. 8 Tracking of SINS of sea vehicle for filter experiments

East China Sea for a trajectory of a long endurance test over 7 h. The SINS trajectory and velocities from the DVL are shown in Figs. 8 and 9, respectively. Further, the performance parameters of the gyroscope and accelerometer are listed in Table 2.

We tested only the SINS for the first 2 h and then integrated the DVL and GPS data. The DVL provides velocities with  $\pm 1\%$  accuracy of the speed and updates at 10 Hz, which was the filter update rate when measurements were available. In addition, the sea status during hours 4–5 of the voyage was rough, leading to the oscillations in velocity measurements shown in Fig. 9, where non-Gaussian noise appears.

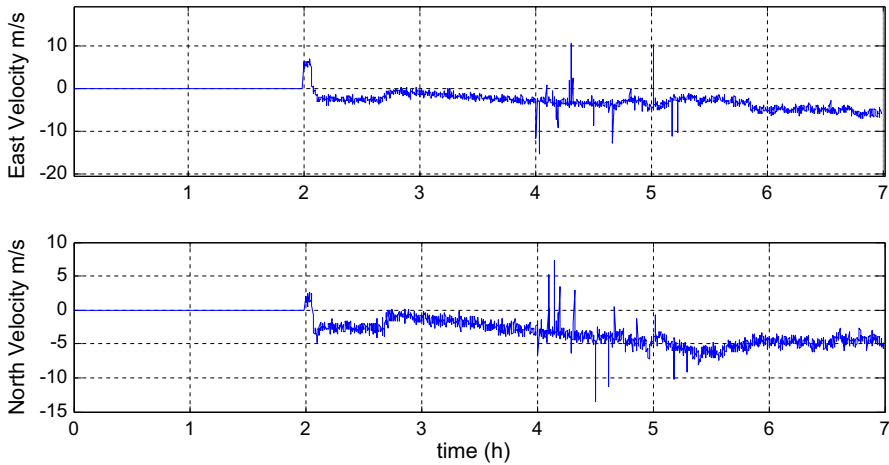


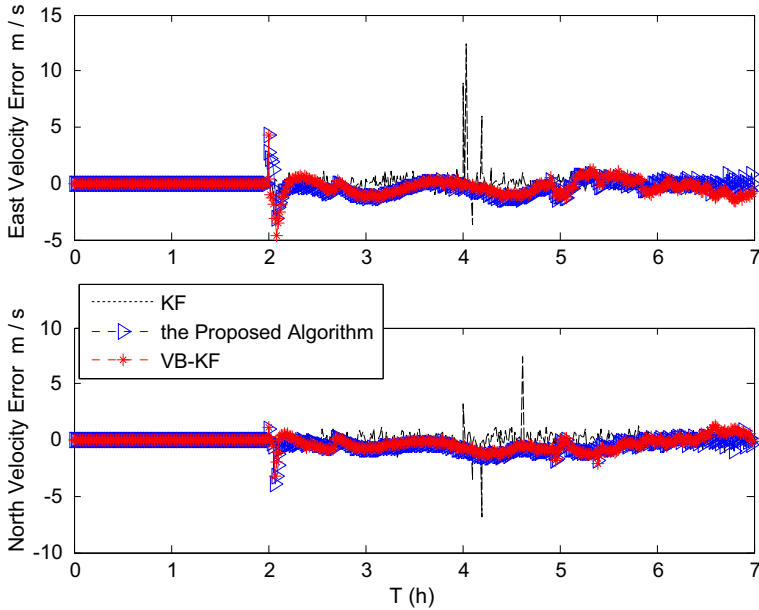
Fig. 9 Velocity from DVL after 2-h sailing for filter experiments

Table 2 Parameters of gyroscopes and accelerometers used to obtain data for testing the proposed filter

Parameter (Unit)	Value
Gyroscope constant drift (°/h)	0.05
Accelerometer zero offset (μg)	100

The initial latitude and longitude of the sea vehicle were 31.25°N and 121.76°E, respectively. The initial east and north velocity were 6 and −5.6 m/s (i.e., heading south). The iterative update of the VB-KF and proposed algorithm were set to  $N = 10$ . The initial filter parameters were selected as  $\hat{x}_{0|0} = \mathbf{0}_{7 \times 1}$ ,  $\mathbf{P}_{0|0} = \text{diag}\{(1000 \text{ m})^2 (1000 \text{ m})^2 (0.1 \text{ m/s})^2 (0.1 \text{ m/s})^2 (0.1^\circ)^2 (0.1^\circ)^2 (0.1^\circ)^2\}$ ,  $\boldsymbol{\mu}_0 = \mathbf{0}_{2 \times 1}$ ,  $\mathbf{R}_0 = \text{diag}\{(0.5 \text{ m/s})^2 (0.5 \text{ m/s})^2\}$ ,  $\mathbf{Q} = \text{diag}\{(100 \mu\text{g})^2 (100 \mu\text{g})^2 (0.01^\circ/\text{h})^2 (0.01^\circ/\text{h})^2 (0.01^\circ/\text{h})^2\}$ . For the proposed filter and VB-KF, the initial hyperparameters were set to  $\lambda_{0|0} = 0.1$ ,  $\mathbf{U}_{0|0} = \text{diag}\{(1 \text{ m/s})^2 (1 \text{ m/s})^2\}$ ,  $a_{0|0} = b_{0|0} = 0.12$ ,  $\boldsymbol{\eta}_{0|0} = [0.1 \text{ m/s } 0.1 \text{ m/s}]^T$ ,  $\beta_{0|0} = 2$ , and forgetting factor was set to  $\rho = 1 - \exp(-4)$ . The velocity errors compared with the other two methods are shown in Fig. 10, and the corresponding RMSE values and running times are listed in Table 3.

From Figs. 9 and 10 and Table 3, we can see that for Gaussian noise, the standard KF shows the best performance for velocity estimation. However, the RMSE in hours 4–5 shows that KF is sensitive to noise outliers and tends to diverge over some periods. In contrast, the two methods based on VB are more stable and show higher performances than KF under outliers. For estimation of east velocity, the proposed filter slightly outperforms VB-KF, whereas for north velocity, the VB-KF is slightly superior to the proposed filter. Regarding average running time (Table 3), the proposed algorithm takes approximately 2–3 times the running time of the VB-KF and 7–8 times that of KF for all the methods executed on a computer with Intel Core i3-4170 CPU at 3.70 GHz.



**Fig. 10** Velocity error comparison of evaluated methods applied on navigation data

**Table 3** State RMSE and average running time of evaluated methods applied on navigation data

Algorithm	RMSE of east velocity	RMSE of north velocity	Running time (s)
KF	0.7691	0.6289	$6.9 \times 10^{-5}$
Proposed algorithm	0.6259	0.6369	$5.5 \times 10^{-4}$
VB-KF	0.6903	0.6036	$1.9 \times 10^{-4}$

## 6 Conclusion

We propose a linear approximate filter with parameter estimation under Student- $t$  measurement model. Specifically, we derive a variational recursive formula to determine state and noise parameters. Then, the sequential variational CRLB is derived for the estimation variance of the proposed algorithm. Two numerical simulations, on a theoretical model and measurement data, were performed and demonstrate that under varying parameters of measurement noise, the proposed method shows better performance than comparison methods, and the derived CRLB agrees with the lower bound of the estimation variance.

**Acknowledgements** This work is supported by the National Natural Science Foundation of China under Grants 61773133 and the Fundamental Research Funds for the Central Universities under Grants HEUCF181110.



## References

1. G. Agamennoni, J.I. Nieto, E.M. Nebot, Approximate inference in state-space models with heavy-tailed noise. *IEEE Trans. Signal Process.* **60**(10), 5024–5037 (2012)
2. J. Ala-Luhtala, S. Särkkä, R. Piché, Gaussian filtering and variational approximations for Bayesian smoothing in continuous-discrete stochastic dynamic systems. *Signal Process.* **111**, 124–136 (2015)
3. T. Baldacchino, E.J. Cross, K. Worden, J. Rowson, Variational Bayesian mixture of experts models and sensitivity analysis for nonlinear dynamical systems. *Mech. Syst. Signal Process.* **66**, 178–200 (2016)
4. J.M. Bernardo, M.J. Bayarri, J.O. Berger, A.P. Dawid, D. Heckerman, A.F.M. Smith, M. West, The variational Bayesian EM algorithm for incomplete data: with application to scoring graphical model structures. *Bayesian Stat.* **7**, 453–464 (2003)
5. D.S. Bernstein, *Matrix Mathematics* (Princeton University Press, Princeton, 2009)
6. C.M. Bishop, *Pattern Recognition and Machine Learning*, 2nd edn. (Springer, Berlin, 2007)
7. W. Gao, J.C. Li, G.T. Zhou, Q. Li, Adaptive Kalman filtering with recursive noise estimator for integrated SINS/DVL systems. *J. Navig.* **68**(1), 142–161 (2015)
8. Y.L. Huang, Y.G. Zhang, Robust Student's  $t$ -based stochastic cubature filter for nonlinear systems with heavy-tailed process and measurement noises. *IEEE Access* **5**, 7964–7974 (2017)
9. Y.L. Huang, Y.G. Zhang, N. Li, Z. Shi, Design of Gaussian approximate filter and smoother for nonlinear systems with correlated noises at one epoch apart. *Circuits Syst. Signal Process.* **35**(11), 3981–4008 (2016)
10. Y.L. Huang, Y.G. Zhang, N. Li, S.M. Naqvi, J.A. Chambers, A robust and efficient system identification method for a state-space model with heavy-tailed process and measurement noises, in *19th International Conference on Information Fusion* (Heidelberg, 2016), pp. 441–448
11. Y.L. Huang, Y.G. Zhang, N. Li, Z.M. Wu, J.A. Chambers, A novel robust Student's  $t$ -based Kalman filter. *IEEE Trans. Aerosp. Electron. Syst.* **53**(3), 1545–1554 (2017)
12. E.G. Larsson, E.K. Larsson, The Cramér–Rao bound for continuous-time autoregressive parameter estimation with irregular sampling. *Circuits Syst. Signal Process.* **21**(6), 581–601 (2002)
13. N. Nasios, A.G. Bors, Variational learning for Gaussian mixture models. *IEEE Trans. Syst. Man Cybern. B Cybern.* **36**(4), 849–862 (2006)
14. E. Özkan, V. Smidl, S. Saha, C. Lundquist, F. Gustafsson, Marginalized adaptive particle filtering for nonlinear models with unknown time-varying noise parameters. *Automatica* **49**(6), 1566–1575 (2013)
15. M. Roth, E. Özkan, F. Gustafsson, A Student's filter for heavy-tailed process and measurement noise, in *Proceedings of the 2013 IEEE International Conference on Acoustics, Speech and Signal Processing* (2013), pp. 5770–5774
16. S. Särkkä, J. Hartikainen, Variational Bayesian adaptation of noise covariance in nonlinear Kalman filtering, in *Proceedings of International Workshop on Machine Learning for Signal Processing* (Southampton, 2013), pp. 1–6
17. S. Särkkä, A. Nummenmaa, Recursive noise adaptive Kalman filtering by variational Bayesian approximations. *IEEE Trans. Autom. Control* **54**(3), 596–600 (2009)
18. I.C. Schick, S.K. Mitter, Robust recursive estimation in the presence of heavy-tailed observation noise. *Ann. Stat.* **2**, 1045–1080 (1994)
19. A. Solin, S. Särkkä, State space methods for efficient inference in Student- $t$  process regression, in *Proceedings of the 18th International Conference on Artificial Intelligence and Statistics (AISTATS)* (San Diego, 2015), pp. 885–893
20. P. Tichavský, C.H. Muravchik, A. Nehorai, Posterior Cramér–Rao bounds for discrete-time nonlinear filtering. *IEEE Trans. Signal Process.* **46**(5), 1386–1396 (1998)
21. D.G. Tzikas, A.C. Likas, N.P. Galatsanos, The variational approximation for Bayesian inference. *IEEE Signal Process. Mag.* **25**(6), 131–146 (2008)
22. S.Y. Wang, C. Yin, S.K. Duan, L.D. Wang, A modified variational Bayesian noise adaptive Kalman filter. *Circuits Syst. Signal Process.* **36**(10), 4260–4277 (2017)
23. X. Wei, C. Li, The infinite Student's  $t$ -mixture for robust modeling. *Signal Process.* **92**(1), 224–234 (2012)
24. H. Zhu, H. Leung, Z. He, A variational Bayesian approach to robust sensor fusion based on Student- $t$  distribution. *Inf. Sci.* **221**, 201–214 (2013)
25. H. Zhu, H. Leung, Z.S. He, State estimation in unknown non-Gaussian measurement noise using variational Bayesian technique. *IEEE Trans. Aerosp. Electron. Syst.* **49**(4), 2601–2614 (2013)

26. L. Zuo, R.X. Niu, P.K. Varshney, Conditional posterior Cramér–Rao lower bounds for nonlinear sequential Bayesian estimation. *IEEE Trans. Signal Process.* **59**(1), 1–14 (2011)