SHORT PAPER

A New Adaptive Kalman Filter with Inaccurate Noise Statistics

Dingjie Xu[1] · Zhemin Wu[1] · Yulong Huang[2]

Abstract

In this paper, a new adaptive Kalman filter is proposed for a linear Gaussian state-space model with inaccurate noise statistics based on the variational Bayesian (VB) approach. Both the prior joint probability density function (PDF) of the one-step prediction and corresponding prediction error covariance matrix and the joint PDF of the mean vector and covariance matrix of measurement noise are selected as Normal-inverse-Wishart (NIW), from which a new NIW-based hierarchical Gaussian state-space model is constructed. The state vector, the one-step prediction and corresponding prediction error covariance matrix, and the mean vector and covariance matrix of measurement noise are jointly estimated based on the constructed hierarchical Gaussian state-space model using the VB approach. Simulation results show that the proposed filter has better estimation accuracy than existing state-of-the-art adaptive Kalman filters.

Keywords Adaptive filtering · Kalman filter · Variational Bayesian · Normal-inverse-Wishart distribution · Inaccurate noise statistics

✉ Zhemin Wu
   myemailabc@163.com

   Dingjie Xu
   xdj1966@hit.edu.cn

   Yulong Huang
   heuedu@163.com

[1] Harbin Institute of Technology, No. 2, Yikuang Street, Nangang District, Harbin 150080, People's Republic of China

[2] Harbin Engineering University, No. 145, Nantong Street, Nangang District, Harbin 150001, People's Republic of China

Birkhäuser

1 Introduction

Kalman filtering has been extensively applied in many science and engineering disciplines, such as target tracking, control, signal processing, communication, navigation and robotics [16]. The Kalman filter is an optimal state estimator in terms of minimum mean square error for a linear state-space model with Gaussian state and measurement noises. A priori knowledge of the noise statistics (mean vectors and covariance matrices of state and measurement noises) are required to implement the Kalman filter. The performance of the Kalman filter relies heavily on the prior noise statistics, and the estimation accuracy of the Kalman filter degrades dramatically when the inaccurate or wrong prior noise statistics are used [13]. However, in a range of practical applications, the noise statistics may be unknown or even time-varying. For example, in the strap-down inertial navigation system (SINS)- and global positioning system (GPS)-based integrated navigation system, the state and measurement noises are, respectively, induced by the measurement errors of inertial measurement units (IMUs) and GPS, and the biases and observation accuracy of IMUs and GPS may vary with the change of the environment, which results in unknown and time-varying mean vectors and covariance matrices of state and measurement noises [3–5,14,19].

To address the filtering problem of a linear Gaussian state-space model with inaccurate noise statistics, a large number of adaptive Kalman filters (AKFs) have been proposed based on different methods. The Sage-Husa AKF (SHAKF) has been proposed based on the covariance matching method, in which the noise statistics are recursively estimated using the maximum a posterior criterion [2,17]. Although the SHAKF can estimate the mean vectors and covariance matrices of state and measurement noises simultaneously, both the convergence to the right noise covariance matrices and the positive definiteness of the estimated noise covariance matrices are not guaranteed by the SHAKF [6]. As a result, the SHAKF is often found to halt its operation due to filtering divergence or numerical instability. The IAKF has been proposed based on a maximum likelihood rule, and it estimates the noise covariance matrices using the innovation sequence that is a white process in the Kalman filtering framework [14]. In general, a large windows of innovations are required by the IAKF to achieve reliable and accurate estimations of noise covariance matrices so that the performance of IAKF may degrade for rapidly varying noise covariance matrices [11]. Also, the positive definiteness of the estimated noise covariance matrices is not guaranteed by the IAKF, which may result in numerical instability. The multiple-model AKF (MMAKF) has been proposed based on the Bayesian rule, in which the state and measurement noises are modeled by Gaussian distributions with different noise statistics, and a bank of Kalman filters are operated using different noise statistics, and the state estimate of the MMAKF is a weighted sum of the estimates of these Kalman filters [12]. Unfortunately, in practical applications, a large number of Gaussian distributions are selected to model the state and measurement noises to guarantee the performance of the MMAKF, which leads to substantial computational complexities [18].

Recently, the expectation-maximization-based AKF (EMAKF) and the variational Bayesian-based AKF (VBAKF) have been proposed to deal with the filtering problem of a Gaussian state-space model with inaccurate noise covariance matrices, in

which the prediction error covariance matrix and measurement noise covariance matrix instead of the state and measurement noise covariance matrices are jointly estimated [6,7]. For the EMAKF, the approximate maximum likelihood estimates of the prediction error covariance matrix and measurement noise covariance matrix are obtained based on the expectation-maximization (EM) algorithm [7]. On the other hand, for the VBAKF, the state vector, the prediction error covariance matrix and the measurement noise covariance matrix are jointly inferred using the variational Bayesian (VB) approach, and the posterior probability density functions (PDFs) of the prediction error covariance matrix and measurement noise covariance matrix are approximated by inverse-Wishart (IW) [6]. Although the EMAKF and VBAKF exhibit satisfactory performance for the case of time-varying and inaccurate noise covariance matrices, their estimation accuracy degrades for the case of time-varying and inaccurate mean vectors and covariance matrices of state and measurement noises.

In this paper, a new AKF is proposed for a linear Gaussian state-space model with inaccurate noise statistics. Both the prior joint PDF of the one-step prediction and corresponding prediction error covariance matrix and the joint PDF of the mean vector and covariance matrix of measurement noise are modeled by Normal-inverse-Wishart (NIW) distributions, from which a new NIW-based hierarchical Gaussian state-space model is constructed. The state vector, the one-step prediction and corresponding prediction error covariance matrix, and the mean vector and covariance matrix of measurement noise are jointly estimated based on the VB approach. The proposed AKF and existing state-of-the-art AKFs are applied to a problem of target tracking with inaccurate noise statistics. Simulation results demonstrate that the proposed AKF has better estimation accuracy than existing state-of-the-art AKFs.

The remainder of this paper is organized as follows. Section 2 gives the notations that are used in this paper. Section 3 presents the problem formulation. In Sect. 4, a new NIW-based hierarchical Gaussian state-space model is constructed, based on which a new AKF is proposed using the VB approach. In Sect. 5, the proposed AKF and existing state-of-the-art AKFs are compared by a target tracking example, and simulation results are given. Concluding remarks are given in Sect. 6.

2 Notations

Throughout this paper, we denote $z_{i:j} \triangleq \{z_k | i \leq k \leq j\}$; $N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ and $N(\cdot; \boldsymbol{\mu}, \boldsymbol{\Sigma})$ denote, respectively, the multivariate Gaussian distribution and Gaussian PDF with mean vector $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$; $NiW(\cdot, \cdot; \boldsymbol{\mu}, \gamma, \delta, \boldsymbol{\Lambda})$ denotes the NIW PDF with mean vector $\boldsymbol{\mu}$, priori confidence parameter γ, degrees of freedom (dof) parameter δ, and scale matrix $\boldsymbol{\Lambda}$; $IW(\cdot; \mu, \boldsymbol{\Sigma})$ denotes the IW PDF with dof parameter μ and inverse scale matrix $\boldsymbol{\Sigma}$; $E_x[\cdot]$ denotes the expectation operator with respect to the PDF of random variable x; log denotes the natural logarithm; $\boldsymbol{I}_n$ denotes the $n \times n$ identity matrix; $\boldsymbol{0}_{n \times 1}$ denotes n dimensional zero vector; the superscript "−1" denotes the inverse operation of a matrix; the superscript "T" denotes the transpose operation of a vector or matrix; and $|\cdot|$ and $tr(\cdot)$ denote the determinant and trace operations of a matrix, respectively.

3 Problem Formulation

Consider the following discrete-time linear stochastic system as shown by the state-space model [8]

$$x_k = F_{k-1}x_{k-1} + w_{k-1}, \tag{1}$$

$$z_k = H_k x_k + v_k, \tag{2}$$

where (1) and (2) are, respectively, the state and measurement equations, k is the discrete-time index, $x_k \in \mathbb{R}^n$ is the state vector, $z_k \in \mathbb{R}^m$ is the measurement vector, $F_k \in \mathbb{R}^{n \times n}$ is the state transition matrix, $H_k \in \mathbb{R}^{m \times n}$ is the observation matrix, and $w_k \in \mathbb{R}^n$ and $v_k \in \mathbb{R}^m$ are, respectively, the state and measurement noise vectors. The state and measurement noises are assumed to have stationary Gaussian distributions, i.e., $w_k \sim N(q, Q)$ and $v_k \sim N(r, R)$, where q and Q are, respectively, the mean vector and covariance matrix of state noise, and r and R are, respectively, the mean vector and covariance matrix of measurement noise. The initial state vector x_0 is assumed to have a Gaussian distribution with mean vector $\hat{x}_{0|0}$ and covariance matrix $P_{0|0}$. Moreover, x_0, w_k and v_j are assumed to be mutually uncorrelated for any time samples j and k.

For the linear Gaussian state-space model formulated in (1)–(2), the Kalman filter is often employed to estimate unknown state vector x_k based on the available measurement information $z_{1:k}$, model parameters $\{F_{k-1}, H_k\}$, and noise statistics $\{q, Q, r, R\}$. The recursive Kalman filter is composed of time update and measurement update, which is formulated as follows: [16]

Time update

$$\hat{x}_{k|k-1} = F_{k-1}\hat{x}_{k-1|k-1} + q, \tag{3}$$

$$P_{k|k-1} = F_{k-1}P_{k-1|k-1}F_{k-1}^{\mathrm{T}} + Q, \tag{4}$$

Measurement update

$$K_k = P_{k|k-1}H_k^{\mathrm{T}}(H_k P_{k|k-1}H_k^{\mathrm{T}} + R)^{-1}, \tag{5}$$

$$\hat{x}_{k|k} = \hat{x}_{k|k-1} + K_k \left(z_k - H_k\hat{x}_{k|k-1} - r\right), \tag{6}$$

$$P_{k|k} = (I_n - K_k H_k) P_{k|k-1}, \tag{7}$$

where $\hat{x}_{k|k-1}$ and $P_{k|k-1}$ denote, respectively, the one-step prediction and corresponding prediction error covariance matrix, and K_k denotes the Kalman gain, and $\hat{x}_{k|k}$ and $P_{k|k}$ are, respectively, the state estimate vector and corresponding estimation error covariance matrix.

Kalman filter is a minimum mean square error state estimator for a linear Gaussian state-space model with accurate noise statistics $\{q, Q, r, R\}$. The performance of Kalman filter degrades severely when inaccurate mean vectors and covariance matrices of state and measurement noises are used. Unfortunately, accurate noise statistics may be unavailable in some engineering applications, such as target tracking. For example, the inaccurate noise statistics may be induced by severe maneuvering when

an agile target is tracked. Therefore, there is a great demand for a new AKF suitable for operation with inaccurate noise statistics. Next, a new AKF will be proposed based on NIW distribution using the VB approach.

4 Main Results

4.1 A New NIW-Based Hierarchical Gaussian State-Space Model

By the fact that the measurement noise vector v_k has a Gaussian distribution with mean vector r and covariance matrix R and using measurement model (2), the likelihood PDF is formulated as

$$p(z_k|x_k, r, R) = N(z_k; H_k x_k + r, R). \tag{8}$$

In the Kalman filtering framework, the one-step prediction PDF is updated as Gaussian, i.e.,

$$p(x_k|\hat{x}_{k|k-1}, P_{k|k-1}, z_{1:k-1}) = N(x_k; \hat{x}_{k|k-1}, P_{k|k-1}). \tag{9}$$

To resist the uncertainties of noise statistics, the state vector x_k, the one-step prediction $\hat{x}_{k|k-1}$, the prediction error covariance matrix $P_{k|k-1}$, the mean vector r and covariance matrix R of measurement noise will be jointly inferred. To this end, the joint conjugate priori distributions for $\{\hat{x}_{k|k-1}, P_{k|k-1}\}$ and $\{r, R\}$ need to be firstly selected since the conjugacy can guarantee that the posterior distribution is of the same functional form as the priori distribution. In Bayesian statistics, a NIW distribution is usually used as the joint conjugate priori for the mean vector and covariance matrix of a Gaussian distribution [15]. Since both $\{\hat{x}_{k|k-1}, P_{k|k-1}\}$ and $\{r, R\}$ are the mean vectors and covariance matrices of Gaussian PDFs in (8)–(9), their joint prior distributions $p(\hat{x}_{k|k-1}, P_{k|k-1}|z_{1:k-1})$ and $p(r, R|z_{1:k-1})$ are chosen as NIW PDFs, i.e.,

$$p(\hat{x}_{k|k-1}, P_{k|k-1}|z_{1:k-1}) = NiW(\hat{x}_{k|k-1}, P_{k|k-1}; u_k, \alpha_k, \omega_k, \Sigma_k), \tag{10}$$

$$p(r, R|z_{1:k-1}) = NiW(r, R; \lambda_k, \beta_k, \nu_k, \Delta_k), \tag{11}$$

where u_k, α_k, ω_k and Σ_k are, respectively, the mean vector, priori confidence parameter, dof parameter and scale matrix of $p(\hat{x}_{k|k-1}, P_{k|k-1}|z_{1:k-1})$, and $\lambda_k, \beta_k, \nu_k$ and Δ_k are, respectively, the mean vector, priori confidence parameter, dof parameter and scale matrix of $p(r, R|z_{1:k-1})$, and the NIW PDF $NiW(a, A; \mu, \gamma, \delta, \Lambda)$ can be written as [1]

$$NiW(a, A; \mu, \gamma, \delta, \Lambda) = N(a; \mu, A/\gamma)IW(A; \delta, \Lambda). \tag{12}$$

Utilizing (12) in (10)–(11) gives

$$p(\hat{x}_{k|k-1}, P_{k|k-1}|z_{1:k-1}) = N(\hat{x}_{k|k-1}; u_k, P_{k|k-1}/\alpha_k) \times IW(P_{k|k-1}; \omega_k, \Sigma_k) \tag{13}$$

$$p(r, R|z_{1:k-1}) = N(r; \lambda_k, R/\beta_k)IW(R; \nu_k, \Delta_k). \tag{14}$$

Employing (13)–(14), the joint priori PDFs for $\left\{\hat{x}_{k|k-1}, \quad P_{k|k-1}\right\}$ and $\{r, R\}$ can be rewritten as the following hierarchical Gaussian forms

$$p(\hat{x}_{k|k-1}|P_{k|k-1}, z_{1:k-1}) = \mathrm{N}(\hat{x}_{k|k-1}; u_k, P_{k|k-1}/\alpha_k), \tag{15}$$

$$p(P_{k|k-1}|z_{1:k-1}) = \mathrm{IW}(P_{k|k-1}; \omega_k, \Sigma_k), \tag{16}$$

$$p(r|R, z_{1:k-1}) = \mathrm{N}(r; \lambda_k, R/\beta_k), \tag{17}$$

$$p(R|z_{1:k-1}) = \mathrm{IW}(R; \nu_k, \Delta_k). \tag{18}$$

To capture the priori information of $\left\{\hat{x}_{k|k-1}, P_{k|k-1}\right\}$, the mean values of $\hat{x}_{k|k-1}$ and $P_{k|k-1}$ are, respectively, set as the nominal one-step prediction $\hat{x}_{k|k-1}^*$ and the nominal prediction error covariance matrix $P_{k|k-1}^*$, i.e.,

$$u_k = \hat{x}_{k|k-1}^* = F_{k-1}\hat{x}_{k-1|k-1} + q^*, \tag{19}$$

$$\frac{\Sigma_k}{\omega_k - n - 1} = P_{k|k-1}^* = F_{k-1}P_{k-1|k-1}F_{k-1}^{\mathrm{T}} + Q^*, \tag{20}$$

where q^* and Q^* are, respectively, the nominal mean vector and covariance matrix of state noise. Let

$$\omega_k = \tau + n + 1. \tag{21}$$

where τ is a tuning parameter satisfying $\tau > 0$.

Substituting (21) in (20) yields

$$\Sigma_k = \tau P_{k|k-1}^*. \tag{22}$$

To capture the priori information of $\{r, R\}$, the mean value of r is set as the nominal mean vector r^*, and the dof parameter ν_k and scale matrix Δ_k are set as previous estimates, i.e.,

$$\lambda_k = r^*, \quad \nu_k = \hat{\nu}_{k-1|k-1}, \quad \Delta_k = \hat{\Delta}_{k-1|k-1}. \tag{23}$$

where $\hat{\nu}_{k-1|k-1}$ and $\hat{\Delta}_{k-1|k-1}$ denote, respectively, the estimates of the dof parameter and scale matrix of R at time $k-1$, and the initial dof parameter ν_0 and scale matrix Δ_0 satisfy $\frac{\Delta_0}{\nu_0 - m - 1} = R^*$ and $\nu_0 > m + 1$ with R^* denoting the nominal measurement noise covariance matrix.

Equations (8)–(9) and (15)–(23) constitute a new NIW-based hierarchical Gaussian state-space model, whose diagram is illustrated in Fig. 1. Next, the state vector, the one-step prediction and corresponding prediction error covariance matrix, and the mean vector and covariance matrix of measurement noise, i.e., $\Psi_k \triangleq \{x_k, \hat{x}_{k|k-1}, P_{k|k-1}, r, R\}$, will be jointly estimated based on the constructed hierarchical Gaussian state-space model using the VB approach, from which a new AKF with inaccurate noise statistics will be proposed.

4.2 Variational Approximations of Posterior PDFs

To infer the state vector, the one-step prediction and corresponding prediction error covariance matrix, and the mean vector and covariance matrix of measurement noise
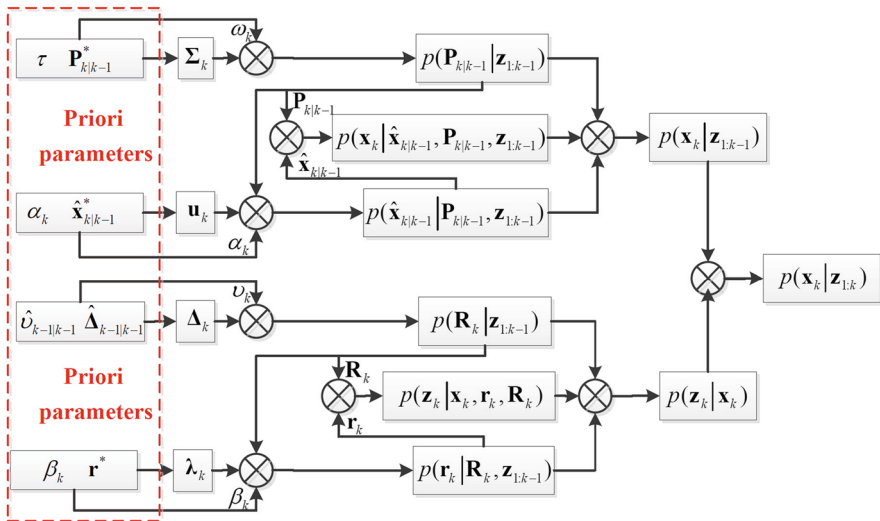
Fig. 1 Diagram of the constructed NIW-based hierarchical Gaussian state-space model

simultaneously, the joint posterior PDF $p(\boldsymbol{\Psi}_k|z_{1:k})$ needs to be calculated since it provides full description for parameter set $\boldsymbol{\Psi}_k$. Unfortunately, the optimal solution for the joint posterior PDF is unavailable based on the constructed hierarchical Gaussian state-space model because the IW PDF has not a closed form. In the paper, the VB approach is employed to obtain a freeform factored approximate solution for the joint posterior PDF $p(\boldsymbol{\Psi}_k|z_{1:k})$, i.e., [9,10]

$$p(\boldsymbol{\Psi}_k|z_{1:k}) \approx q(\boldsymbol{x}_k)q(\hat{\boldsymbol{x}}_{k|k-1})q(\boldsymbol{P}_{k|k-1})q(\boldsymbol{r})q(\boldsymbol{R}), \tag{24}$$

where $q(\cdot)$ denotes the approximate posterior PDF.

Based on the VB approach, these approximate posterior PDFs are achieved by minimizing the Kullback–Leibler divergence (KLD) between approximate posterior PDF and true posterior PDF, i.e., [9,10]

$$\begin{aligned} \left\{ q(\mathbf{x}_k), q(\hat{\mathbf{x}}_{k|k-1}), q(\mathbf{P}_{k|k-1}), q(\mathbf{r}), q(\mathbf{R}) \right\} = \arg\min \\ \text{KLD}\left(q(\mathbf{x}_k)q(\hat{\mathbf{x}}_{k|k-1})q(\mathbf{P}_{k|k-1})q(\mathbf{r})q(\mathbf{R}) \| p(_k|\mathbf{z}_{1:k}) \right), \end{aligned} \tag{25}$$

where $\text{KLD}\left(q(\boldsymbol{x})\|p(\boldsymbol{x})\right) \triangleq \int q(\boldsymbol{x}) \log \frac{q(\boldsymbol{x})}{p(\boldsymbol{x})} d\boldsymbol{x}$ denotes the KLD operation between PDFs $q(\boldsymbol{x})$ and $p(\boldsymbol{x})$.

The optimal solution for (25) can be formulated as [9,10]

$$\log q(\boldsymbol{\theta}) = \text{E}_{\boldsymbol{\Psi}_k^{(-\boldsymbol{\theta})}}[\log p(\boldsymbol{\Psi}_k, z_{1:k})] + c_{\boldsymbol{\theta}}, \tag{26}$$

where $\boldsymbol{\theta}$ is an arbitrary element of $\boldsymbol{\Psi}_k$, and $\boldsymbol{\Psi}_k^{(-\boldsymbol{\theta})}$ is the set of all elements in $\boldsymbol{\Psi}_k$ except for $\boldsymbol{\theta}$, and $c_{\boldsymbol{\theta}}$ denotes the constant with respect to variable $\boldsymbol{\theta}$.

However, an analytical solution for (26) is unavailable since the variational parameters are mutually coupled. The fixed-point iteration approach is often employed to obtain an approximate solution for (26). That is to say, the approximate posterior PDF $q(\boldsymbol{\theta})$ of an arbitrary element $\boldsymbol{\theta}$ is updated as $q^{(i+1)}(\boldsymbol{\theta})$ at the $i+1$th iteration by using the approximate posterior PDFs $q^{(i+1)}(\boldsymbol{\gamma})$ and $q^{(i)}(\boldsymbol{\Psi}_k^{(-\{\boldsymbol{\theta},\boldsymbol{\gamma}\})})$ to calculate the expectation in (26), where $\boldsymbol{\gamma}$ is composed of the elements that have been updated at the $i+1$th iteration. The iterations converge to a local optimum of Eq. (26).

According to the Bayesian theorem and using the conditional independence properties of the constructed NIW-based hierarchical Gaussian state-space model, the joint PDF $p(\boldsymbol{\Psi}_k, z_{1:k})$ can be formulated as

$$p(\boldsymbol{\Psi}_k, z_{1:k}) = p(z_k|\boldsymbol{x}_k, \boldsymbol{r}, \boldsymbol{R}) p(\boldsymbol{x}_k|z_{1:k-1}, \hat{\boldsymbol{x}}_{k|k-1}, \boldsymbol{P}_{k|k-1}) p(\hat{\boldsymbol{x}}_{k|k-1}|\boldsymbol{P}_{k|k-1}, z_{1:k-1})$$
$$p(\boldsymbol{P}_{k|k-1}|z_{1:k-1}) p(\boldsymbol{r}|\boldsymbol{R}, z_{1:k-1}) p(\boldsymbol{R}|z_{1:k-1}) p(z_{1:k-1}). \tag{27}$$

Substituting (8)–(9) and (15)–(18) in (27) yields

$$p(\boldsymbol{\Psi}_k, z_{1:k}) = \mathrm{N}(z_k; \boldsymbol{H}_k \boldsymbol{x}_k + \boldsymbol{r}, \boldsymbol{R}) \mathrm{N}(\boldsymbol{x}_k; \hat{\boldsymbol{x}}_{k|k-1}, \boldsymbol{P}_{k|k-1}) \mathrm{N}(\hat{\boldsymbol{x}}_{k|k-1}; \boldsymbol{u}_k, \boldsymbol{P}_{k|k-1}/\alpha_k)$$
$$\mathrm{IW}(\boldsymbol{P}_{k|k-1}; \omega_k, \boldsymbol{\Sigma}_k) \mathrm{N}(\boldsymbol{r}; \lambda_k, \boldsymbol{R}/\beta_k) \mathrm{IW}(\boldsymbol{R}; \nu_k, \boldsymbol{\Delta}_k) p(z_{1:k-1}). \tag{28}$$

Let $\boldsymbol{\theta} = \hat{\boldsymbol{x}}_{k|k-1}$ and $\boldsymbol{\theta} = \boldsymbol{r}$ and using (28) in (26), $q^{(i+1)}(\hat{\boldsymbol{x}}_{k|k-1})$ and $q^{(i+1)}(\boldsymbol{r})$ are updated as Gaussian PDFs, i.e.,

$$q^{(i+1)}(\hat{\boldsymbol{x}}_{k|k-1}) = \mathrm{N}(\hat{\boldsymbol{x}}_{k|k-1}; \hat{\boldsymbol{u}}_k^{(i+1)}, \hat{\boldsymbol{U}}_k^{(i+1)}), \tag{29}$$

$$q^{(i+1)}(\boldsymbol{r}) = \mathrm{N}(\boldsymbol{r}; \hat{\boldsymbol{\lambda}}_k^{(i+1)}, \hat{\boldsymbol{\Omega}}_k^{(i+1)}), \tag{30}$$

where the mean vectors $\hat{\boldsymbol{u}}_k^{(i+1)}$ and $\hat{\boldsymbol{\lambda}}_k^{(i+1)}$ and covariance matrices $\hat{\boldsymbol{U}}_k^{(i+1)}$ and $\hat{\boldsymbol{\Omega}}_k^{(i+1)}$ are given by

$$\hat{\boldsymbol{u}}_k^{(i+1)} = \frac{\alpha_k \boldsymbol{u}_k + \hat{\boldsymbol{x}}_{k|k}^{(i)}}{\alpha_k + 1}, \qquad \hat{\boldsymbol{U}}_k^{(i+1)} = \frac{\bar{\boldsymbol{P}}_{k|k-1}^{(i)}}{\alpha_k + 1}, \tag{31}$$

$$\hat{\boldsymbol{\lambda}}_k^{(i+1)} = \frac{\beta_k \lambda_k + \left(z_k - \boldsymbol{H}_k \hat{\boldsymbol{x}}_{k|k}^{(i)}\right)}{\beta_k + 1}, \quad \hat{\boldsymbol{\Omega}}_k^{(i+1)} = \frac{\bar{\boldsymbol{R}}_k^{(i)}}{\beta_k + 1}, \tag{32}$$

and the modified prediction error covariance matrix $\bar{\boldsymbol{P}}_{k|k-1}^{(i)}$ and modified measurement noise covariance matrix $\bar{\boldsymbol{R}}_k^{(i)}$ are formulated as

$$\bar{\boldsymbol{P}}_{k|k-1}^{(i)} = \left\{\mathrm{E}^{(i)}[\boldsymbol{P}_{k|k-1}^{-1}]\right\}^{-1}, \quad \bar{\boldsymbol{R}}_k^{(i)} = \left\{\mathrm{E}^{(i)}\left[\boldsymbol{R}^{-1}\right]\right\}^{-1}, \tag{33}$$

where the proofs of (29)–(33) are given in "Appendix A".

Let $\boldsymbol{\theta} = \boldsymbol{P}_{k|k-1}$ and $\boldsymbol{\theta} = \boldsymbol{R}$ and exploiting (28) in (26), $q^{(i+1)}(\boldsymbol{P}_{k|k-1})$ and $q^{(i+1)}(\boldsymbol{R})$ are updated as IW PDFs, i.e.,

$$q^{(i+1)}(\boldsymbol{P}_{k|k-1}) = \mathrm{IW}(\boldsymbol{P}_{k|k-1}; \hat{\omega}_k^{(i+1)}, \hat{\boldsymbol{\Sigma}}_k^{(i+1)}), \tag{34}$$

$$q^{(i+1)}(\boldsymbol{R}) = \mathrm{IW}(\boldsymbol{R}; \hat{v}_k^{(i+1)}, \hat{\boldsymbol{\Delta}}_k^{(i+1)}), \tag{35}$$

where the dof parameters $\hat{\omega}_k^{(i+1)}$ and $\hat{v}_k^{(i+1)}$ and inverse scale matrices $\hat{\boldsymbol{\Sigma}}_k^{(i+1)}$ and $\hat{\boldsymbol{\Delta}}_k^{(i+1)}$ are given by

$$\hat{\omega}_k^{(i+1)} = \omega_k + 2, \tag{36}$$

$$\hat{\boldsymbol{\Sigma}}_k^{(i+1)} = \boldsymbol{A}_k^{(i+1)} + \boldsymbol{B}_k^{(i+1)} + \boldsymbol{\Sigma}_k, \tag{37}$$

$$\hat{v}_k^{(i+1)} = v_k + 2, \tag{38}$$

$$\hat{\boldsymbol{\Delta}}_k^{(i+1)} = \boldsymbol{C}_k^{(i+1)} + \boldsymbol{D}_k^{(i+1)} + \boldsymbol{\Delta}_k, \tag{39}$$

and the auxiliary parameters $\boldsymbol{A}_k^{(i+1)}, \boldsymbol{B}_k^{(i+1)}, \boldsymbol{C}_k^{(i+1)}$ and $\boldsymbol{D}_k^{(i+1)}$ are, respectively, given by

$$\boldsymbol{A}_k^{(i+1)} = \mathrm{E}^{(i)}\left[\left(\boldsymbol{x}_k - \hat{\boldsymbol{x}}_{k|k-1}\right)\left(\boldsymbol{x}_k - \hat{\boldsymbol{x}}_{k|k-1}\right)^{\mathrm{T}}\right], \tag{40}$$

$$\boldsymbol{B}_k^{(i+1)} = \alpha_k \mathrm{E}^{(i)}\left[\left(\hat{\boldsymbol{x}}_{k|k-1} - \boldsymbol{u}_k\right)\left(\hat{\boldsymbol{x}}_{k|k-1} - \boldsymbol{u}_k\right)^{\mathrm{T}}\right], \tag{41}$$

$$\boldsymbol{C}_k^{(i+1)} = \mathrm{E}^{(i)}\left[\left(\boldsymbol{z}_k - \boldsymbol{H}_k\boldsymbol{x}_k - \boldsymbol{r}\right)\left(\boldsymbol{z}_k - \boldsymbol{H}_k\boldsymbol{x}_k - \boldsymbol{r}\right)^{\mathrm{T}}\right], \tag{42}$$

$$\boldsymbol{D}_k^{(i+1)} = \beta_k \mathrm{E}^{(i)}\left[\left(\boldsymbol{r} - \boldsymbol{\lambda}_k\right)\left(\boldsymbol{r} - \boldsymbol{\lambda}_k\right)^{\mathrm{T}}\right], \tag{43}$$

where the proofs of (34)–(43) are given in "Appendix B".

Let $\boldsymbol{\theta} = \boldsymbol{x}_k$ and using (28) in (26), $q^{(i+1)}(\boldsymbol{x}_k)$ is updated as a Gaussian PDF, i.e.,

$$q^{(i+1)}(\boldsymbol{x}_k) = \mathrm{N}(\boldsymbol{x}_k; \hat{\boldsymbol{x}}_{k|k}^{(i+1)}, \boldsymbol{P}_{k|k}^{(i+1)}), \tag{44}$$

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

$$\boldsymbol{K}_k^{(i+1)} = \bar{\boldsymbol{P}}_{k|k-1}^{(i+1)} \boldsymbol{H}_k^{\mathrm{T}} (\boldsymbol{H}_k \bar{\boldsymbol{P}}_{k|k-1}^{(i+1)} \boldsymbol{H}_k^{\mathrm{T}} + \bar{\boldsymbol{R}}_k^{(i+1)})^{-1}, \tag{45}$$

$$\hat{\boldsymbol{x}}_{k|k}^{(i+1)} = \hat{\boldsymbol{u}}_k^{(i+1)} + \boldsymbol{K}_k^{(i+1)} (\boldsymbol{z}_k - \hat{\boldsymbol{\lambda}}_k^{(i+1)} - \boldsymbol{H}_k \hat{\boldsymbol{u}}_k^{(i+1)}), \tag{46}$$

$$\boldsymbol{P}_{k|k}^{(i+1)} = \left(\boldsymbol{I}_n - \boldsymbol{K}_k^{(i+1)} \boldsymbol{H}_k\right) \bar{\boldsymbol{P}}_{k|k-1}^{(i+1)}, \tag{47}$$

where $\boldsymbol{K}_k^{(i+1)}$ denotes the modified Kalman gain, and the proofs of (44)–(47) are given in "Appendix C".

After fixed-point iteration N, the approximate posterior PDFs of the state vector, the one-step prediction and corresponding prediction error covariance matrix, and the

mean vector and covariance matrix of measurement noise are, respectively, updated as

$$q(\boldsymbol{x}_k) \approx \mathrm{N}(\boldsymbol{x}_k; \hat{\boldsymbol{x}}_{k|k}^{(N)}, \boldsymbol{P}_{k|k}^{(N)}) = \mathrm{N}(\boldsymbol{x}_k; \hat{\boldsymbol{x}}_{k|k}, \boldsymbol{P}_{k|k}), \tag{48}$$

$$q(\hat{\boldsymbol{x}}_{k|k-1}) \approx \mathrm{N}(\hat{\boldsymbol{x}}_{k|k-1}; \hat{\boldsymbol{u}}_k^{(N)}, \hat{\boldsymbol{U}}_k^{(N)}) = \mathrm{N}(\hat{\boldsymbol{x}}_{k|k-1}; \hat{\boldsymbol{u}}_{k|k}, \hat{\boldsymbol{U}}_{k|k}), \tag{49}$$

$$q(\boldsymbol{P}_{k|k-1}) \approx \mathrm{IW}(\boldsymbol{P}_{k|k-1}; \hat{\omega}_k^{(N)}, \hat{\boldsymbol{\Sigma}}_k^{(N)}) = \mathrm{IW}(\boldsymbol{P}_{k|k-1}; \hat{\omega}_{k|k}, \hat{\boldsymbol{\Sigma}}_{k|k}), \tag{50}$$

$$q(\boldsymbol{r}) \approx \mathrm{N}(\boldsymbol{r}; \hat{\boldsymbol{\lambda}}_k^{(N)}, \hat{\boldsymbol{\Omega}}_k^{(N)}) = \mathrm{N}(\boldsymbol{r}; \hat{\boldsymbol{\lambda}}_{k|k}, \hat{\boldsymbol{\Omega}}_{k|k}), \tag{51}$$

$$q(\boldsymbol{R}) \approx \mathrm{IW}(\boldsymbol{R}; \hat{v}_k^{(N)}, \hat{\boldsymbol{\Delta}}_k^{(N)}) = \mathrm{IW}(\boldsymbol{R}; \hat{v}_{k|k}, \hat{\boldsymbol{\Delta}}_{k|k}). \tag{52}$$

To update the approximate posterior PDFs, we need to calculate the expectations in (33) and (40)–(43). Using (29)–(30), (34)–(35) and (44), the required expectations in (33) and (40)–(43) are calculated as follows

$$\bar{\boldsymbol{P}}_{k|k-1}^{(i+1)} = \hat{\boldsymbol{\Sigma}}_k^{(i+1)} \bigg/ \left(\hat{\omega}_k^{(i+1)} - n - 1 \right), \tag{53}$$

$$\bar{\boldsymbol{R}}_k^{(i+1)} = \hat{\boldsymbol{\Delta}}_k^{(i+1)} \bigg/ \left(\hat{v}_k^{(i+1)} - m - 1 \right), \tag{54}$$

$$\boldsymbol{A}_k^{(i+1)} = \boldsymbol{P}_{k|k}^{(i)} + \hat{\boldsymbol{U}}_k^{(i+1)} + \left(\hat{\boldsymbol{x}}_{k|k}^{(i)} - \hat{\boldsymbol{u}}_k^{(i+1)} \right) \left(\hat{\boldsymbol{x}}_{k|k}^{(i)} - \hat{\boldsymbol{u}}_k^{(i+1)} \right)^{\mathrm{T}}, \tag{55}$$

$$\boldsymbol{B}_k^{(i+1)} = \alpha_k \hat{\boldsymbol{U}}_k^{(i+1)} + \alpha_k \left(\hat{\boldsymbol{u}}_k^{(i+1)} - \boldsymbol{u}_k \right) \left(\hat{\boldsymbol{u}}_k^{(i+1)} - \boldsymbol{u}_k \right)^{\mathrm{T}}, \tag{56}$$

$$\boldsymbol{C}_k^{(i+1)} = \hat{\boldsymbol{\Omega}}_k^{(i+1)} + \left(\boldsymbol{z}_k - \boldsymbol{H}_k \hat{\boldsymbol{x}}_{k|k}^{(i)} - \hat{\boldsymbol{\lambda}}_k^{(i+1)} \right) \left(\boldsymbol{z}_k - \boldsymbol{H}_k \hat{\boldsymbol{x}}_{k|k}^{(i)} - \hat{\boldsymbol{\lambda}}_k^{(i+1)} \right)^{\mathrm{T}}$$
$$+ \boldsymbol{H}_k \boldsymbol{P}_{k|k}^{(i)} \boldsymbol{H}_k^{\mathrm{T}}, \tag{57}$$

$$\boldsymbol{D}_k^{(i+1)} = \beta_k \hat{\boldsymbol{\Omega}}_k^{(i+1)} + \beta_k \left(\hat{\boldsymbol{\lambda}}_k^{(i+1)} - \boldsymbol{\lambda}_k \right) \left(\hat{\boldsymbol{\lambda}}_k^{(i+1)} - \boldsymbol{\lambda}_k \right)^{\mathrm{T}}, \tag{58}$$

where the proofs of (55)–(58) are given in "Appendix D".

The proposed AKF operates recursively by combining the variational approximations of posterior PDFs in (29)–(32), (34)–(39) and (44)–(52) and the calculations of expectations in (53)–(58). The implementation pseudo-code for the one-time step of the proposed AKF is given in Table 1.

4.3 Parameter Selection of the Proposed AKF

To implement the proposed AKF, the nominal mean vectors $\boldsymbol{q}^*$ and $\boldsymbol{r}^*$ and covariance matrices $\boldsymbol{Q}^*$ and $\boldsymbol{R}^*$ of state and measurement noises, the priori confidence parameters α_k and β_k, the tuning parameter τ, and the initial dof parameter v_0 require to be selected in advance.

Firstly, we derive the specific forms of the modified one-step prediction $\hat{\boldsymbol{u}}_k^{(i+1)}$, the modified measurement noise mean vector $\hat{\boldsymbol{\lambda}}_k^{(i+1)}$, the modified prediction error

Table 1 One-time step of the proposed AKF

Inputs: $\hat{x}_{k-1|k-1}, P_{k-1|k-1}, \hat{v}_{k-1|k-1}, \hat{\Delta}_{k-1|k-1}, F_{k-1}, H_k, z_k, q^*, Q^*, r^*, \tau, \alpha_k, \beta_k, N$.

1. Calculate $\hat{x}^*_{k|k-1}, P^*_{k|k-1}, u_k, \Sigma_k, \lambda_k$ and Δ_k using (19)–(23).

2. Initialization: $\bar{P}^{(0)}_{k|k-1} = P^*_{k|k-1}, \bar{R}^{(0)}_k = \frac{\hat{\Delta}_{k-1|k-1}}{\hat{v}_{k-1|k-1}-m-1}, \hat{x}^{(0)}_{k|k} = \hat{x}^*_{k|k-1}, P^{(0)}_{k|k} = P^*_{k|k-1}$.

for $i = 0 : N - 1$

3. Update $q^{(i+1)}(\hat{x}_{k-1})$ and $q^{(i+1)}(r)$ using (29)–(32).

4. Calculate $A^{(i+1)}_k, B^{(i+1)}_k, C^{(i+1)}_k$ and $D^{(i+1)}_k$ using (55)–(58).

5. Update $q^{(i+1)}(P_{k|k-1})$ and $q^{(i+1)}(R)$ using (34)–(39).

6. Calculate $\bar{P}^{(i)}_{k|k-1}$ and $\bar{R}^{(i)}_k$ using (53)–(54).

7. Update $q^{(i+1)}(x_k)$ using (44)–(47).

end for

8. $\hat{x}_{k|k} = \hat{x}^{(N)}_{k|k}, \quad P_{k|k} = P^{(N)}_{k|k}, \quad \hat{v}_{k|k} = \hat{v}^{(N)}_k, \quad \hat{\Delta}_{k|k} = \hat{\Delta}^{(N)}_k$.

Outputs: $\hat{x}_{k|k}, P_{k|k}, \hat{v}_{k|k}, \hat{\Delta}_{k|k}$.

---

covariance matrix $\bar{P}^{(i+1)}_{k|k-1}$ and the modified measurement noise covariance matrix $\bar{R}^{(i+1)}_k$ at the $i + 1$th iteration. Substituting (19) and (23) in (31)–(32) gives

$$\hat{u}^{(i+1)}_k = \frac{\alpha_k \left(F_{k-1}\hat{x}_{k-1|k-1} + q^* \right) + \hat{x}^{(i)}_{k|k}}{\alpha_k + 1}, \tag{59}$$

$$\hat{\lambda}^{(i+1)}_k = \frac{\beta_k r^* + \left(z_k - H_k \hat{x}^{(i)}_{k|k} \right)}{\beta_k + 1}. \tag{60}$$

Using (36)–(37) and (53) in (33), the modified prediction error covariance matrix $\bar{P}^{(i+1)}_{k|k-1}$ can be rewritten as

$$\bar{P}^{(i+1)}_{k|k-1} = \frac{A^{(i+1)}_k + B^{(i+1)}_k + \Sigma_k}{\omega_k - n + 1}, \tag{61}$$

where $A^{(i+1)}_k$ and $B^{(i+1)}_k$ are, respectively, given by (55)–(56).

Substituting (20)–(22) in (61) yields

$$\bar{P}^{(i+1)}_{k|k-1} = \frac{\tau \left(F_{k-1} P_{k-1|k-1} F^T_{k-1} + Q^* \right) + A^{(i+1)}_k + B^{(i+1)}_k}{\tau + 2}. \tag{62}$$

Exploiting (38)–(39) and (52), we have

$$\hat{v}_{k|k} = v_k + 2, \tag{63}$$

$$\hat{\Delta}_{k|k} = C^{(N)}_k + D^{(N)}_k + \Delta_k. \tag{64}$$

Substituting (23) in (63)–(64) results in

$$\hat{v}_{k|k} = \hat{v}_{k-1|k-1} + 2, \tag{65}$$

$$\hat{\boldsymbol{\Delta}}_{k|k} = \hat{\boldsymbol{\Delta}}_{k-1|k-1} + \boldsymbol{C}_k^{(N)} + \boldsymbol{D}_k^{(N)}. \tag{66}$$

According to the recursive relation in (65)–(66), $\hat{v}_{k-1|k-1}$ and $\hat{\boldsymbol{\Delta}}_{k-1|k-1}$ can be calculated as

$$\hat{v}_{k-1|k-1} = v_0 + 2(k-1), \tag{67}$$

$$\hat{\boldsymbol{\Delta}}_{k-1|k-1} = (v_0 - m - 1)\boldsymbol{R}^* + \sum_{j=1}^{k-1} \left[\boldsymbol{C}_j^{(N)} + \boldsymbol{D}_j^{(N)} \right], \tag{68}$$

where $\hat{v}_{0|0} = v_0$ and $\hat{\boldsymbol{\Delta}}_{0|0} = \boldsymbol{\Delta}_0 = (v_0 - m - 1)\boldsymbol{R}^*$ are used in (67)–(68).

Employing (23) and (67)–(68) in (38)–(39) yields

$$\hat{v}_k^{(i+1)} = v_0 + 2k, \tag{69}$$

$$\hat{\boldsymbol{\Delta}}_k^{(i+1)} = (v_0 - m - 1)\boldsymbol{R}^* + \boldsymbol{C}_k^{(i+1)} + \boldsymbol{D}_k^{(i+1)} + \sum_{j=1}^{k-1} \left[\boldsymbol{C}_j^{(N)} + \boldsymbol{D}_j^{(N)} \right]. \tag{70}$$

Substituting (69)–(70) in (54), $\bar{\boldsymbol{R}}_k^{(i+1)}$ can be reformulated as

$$\bar{\boldsymbol{R}}_k^{(i+1)} = \left\{ (v_0 - m - 1)\boldsymbol{R}^* + \boldsymbol{C}_k^{(i+1)} + \boldsymbol{D}_k^{(i+1)} + \sum_{j=1}^{k-1} \left[\boldsymbol{C}_j^{(N)} + \boldsymbol{D}_j^{(N)} \right] \right\} /$$

$$[(v_0 - m - 1) + 2k]. \tag{71}$$

Next, we discuss the effects of parameters $\boldsymbol{q}^*$, $\boldsymbol{r}^*$, $\boldsymbol{Q}^*$, $\boldsymbol{R}^*$, α_k, β_k, τ and v_0 upon the proposed AKF. It is observed from (59)–(60), (62) and (71) that the modified one-step prediction $\hat{\boldsymbol{u}}_k^{(i+1)}$ is a weighted sum of priori information $\left(\boldsymbol{F}_{k-1}\hat{\boldsymbol{x}}_{k-1|k-1} + \boldsymbol{q}^* \right)$ and innovation $\hat{\boldsymbol{x}}_{k|k}^{(i)}$ with weights α_k and 1, respectively; the modified measurement noise mean vector $\hat{\boldsymbol{\lambda}}_k^{(i+1)}$ is a weighted sum of priori information $\boldsymbol{r}^*$ and innovation $\left(\boldsymbol{z}_k - \boldsymbol{H}_k\hat{\boldsymbol{x}}_{k|k}^{(i)} \right)$ with weights β_k and 1, respectively; the modified prediction error covariance matrix $\bar{\boldsymbol{P}}_{k|k-1}^{(i+1)}$ is a weighted sum of priori information $\left(\boldsymbol{F}_{k-1}\boldsymbol{P}_{k-1|k-1}\boldsymbol{F}_{k-1}^{\mathrm{T}} + \boldsymbol{Q}^* \right)$ and innovation $\left(\boldsymbol{A}_k^{(i+1)} + \boldsymbol{B}_k^{(i+1)} \right)$ with weights τ and 2, respectively; and the modified measurement noise covariance matrix $\bar{\boldsymbol{R}}_k^{(i+1)}$ is a weighted sum of priori information $\boldsymbol{R}^*$ and innovation $\left\{ \boldsymbol{C}_k^{(i+1)} + \boldsymbol{D}_k^{(i+1)} + \sum_{j=1}^{k-1} \left[\boldsymbol{C}_j^{(N)} + \boldsymbol{D}_j^{(N)} \right] \right\}$ with weights $v_0 - m - 1$ and $2k$, respectively. Thus, the parameters $\boldsymbol{q}^*$, $\boldsymbol{r}^*$, $\boldsymbol{Q}^*$ and $\boldsymbol{R}^*$ dominate the accuracy of priori

information, and the parameters α_k, β_k, τ and ν_0 dominate the confidence level to priori information.

Since the VB approach only guarantees the local convergence of variational iterations, accurate priori information is necessary. To this end, the nominal parameters $\boldsymbol{q}^*$, $\boldsymbol{r}^*$, $\boldsymbol{Q}^*$ and $\boldsymbol{R}^*$ require to be near the true values $\boldsymbol{q}$, $\boldsymbol{r}$, $\boldsymbol{Q}$ and $\boldsymbol{R}$, respectively. In this paper, the nominal mean vectors and covariance matrices of state and measurement noises are, respectively, suggested to be selected as $\boldsymbol{q}^* = [q_1, \ldots, q_i, \ldots, q_n]^T$, $\boldsymbol{r}^* = [r_1, \ldots, r_j, \ldots, r_m]^T$, $\boldsymbol{Q}^* = \text{diag}[Q_1, \ldots, Q_i, \ldots, Q_n]$ and $\boldsymbol{R}^* = \text{diag}[R_1, \ldots, R_j, \ldots, R_m]$, where $Q_i > 0$ and $R_j > 0$. The explicit selections of the parameters q_i, r_j, Q_i and R_j depend on practical application, and the approximated values of the parameters are available in many practical application. The selections of the parameters α_k, β_k, τ and ν_0 rely heavily on the selections of the nominal parameters $\boldsymbol{q}^*$, $\boldsymbol{r}^*$, $\boldsymbol{Q}^*$ and $\boldsymbol{R}^*$ because they determine the accuracy of priori information.

Finally, we study the numerical stability of the proposed AKF with the selected parameters. Since the nominal covariance matrices of state and measurement noises are, respectively, set as $\boldsymbol{Q}^* = \text{diag}[Q_1, \ldots, Q_i, \ldots, Q_n]$ and $\boldsymbol{R}^* = \text{diag}[R_1, \ldots, R_j, \ldots, R_m]$ with $Q_i > 0$ and $R_j > 0$, both $\boldsymbol{Q}^*$ and $\boldsymbol{R}^*$ are positive matrices, i.e.,

$$\boldsymbol{Q}^* > \boldsymbol{0}, \qquad \boldsymbol{R}^* > \boldsymbol{0}. \tag{72}$$

According to (40)–(43), the auxiliary parameters $\boldsymbol{A}_k^{(i+1)}$, $\boldsymbol{B}_k^{(i+1)}$, $\boldsymbol{C}_k^{(i+1)}$ and $\boldsymbol{D}_k^{(i+1)}$ are all positive semi-definite matrices, i.e.,

$$\boldsymbol{A}_k^{(i+1)} \geq \boldsymbol{0}, \quad \boldsymbol{B}_k^{(i+1)} \geq \boldsymbol{0}, \quad \boldsymbol{C}_k^{(i+1)} \geq \boldsymbol{0}, \quad \boldsymbol{D}_k^{(i+1)} \geq \boldsymbol{0}. \tag{73}$$

Utilizing $\boldsymbol{P}_{k-1|k-1} > \boldsymbol{0}$, $\tau > 0$, $\nu_0 - m - 1 > 0$ and (72)–(73) in (62) and (71) obtains

$$\bar{\boldsymbol{P}}_{k|k-1}^{(i+1)} > \boldsymbol{0}, \qquad \bar{\boldsymbol{R}}_k^{(i+1)} > \boldsymbol{0}. \tag{74}$$

We can see from (74) that both the modified prediction error covariance matrix $\bar{\boldsymbol{P}}_{k|k-1}^{(i+1)}$ and the modified measurement noise covariance matrix $\bar{\boldsymbol{R}}_k^{(i+1)}$ are positive definite, based on which the modified innovation covariance matrix $(\boldsymbol{H}_k \bar{\boldsymbol{P}}_{k|k-1}^{(i+1)} \boldsymbol{H}_k^T + \bar{\boldsymbol{R}}_k^{(i+1)})$ in (45) is also positive definite. Thus, the proposed AKF has numerical stability using the selected parameters.

Remark 1 The estimation accuracy and computational complexity of the proposed filter are determined by the number of iterations N. The more number of iterations is used, the better estimation accuracy can be obtained but the higher computational complexity is needed. Normally, the number of iterations is a tradeoff of the estimation accuracy and computational complexity. In practical engineering applications, the selection of the number of iterations relies on the users' requirements. If the users require good estimation accuracy, then a large N should be selected. Conversely, if the users require good real-time performance, then an appropriately large N should be chosen, which implies that the computational complexity can be improved by sacrificing estimation accuracy.

5 Simulation Study

The proposed AKF and existing state-of-the-art adaptive Kalman filters are tested and compared in a maneuvring target tracking example. The target moves with a constant velocity in a plane, whose positions are observed in clutter. The target is tracked using a constant velocity model, and the noise corrupted positions are used for measurement vectors. The cartesian coordinates and corresponding velocities are selected as a state vector, i.e., $\boldsymbol{x}_k \triangleq [x_k \; y_k \; \dot{x}_k \; \dot{y}_k]$, where x_k, y_k, $\dot{x}_k$ and $\dot{y}_k$ denote the cartesian coordinates and corresponding velocities, respectively. The discrete-time linear state-space model is given by (1)–(2), and the state transition matrix $\boldsymbol{F}_k$ and measurement matrix $\boldsymbol{H}_k$ are given by [8]

$$\boldsymbol{F}_k = \begin{bmatrix} \boldsymbol{I}_2 & \Delta t\, \boldsymbol{I}_2 \\ \boldsymbol{0} & \boldsymbol{I}_2 \end{bmatrix}, \qquad \boldsymbol{H}_k = \begin{bmatrix} \boldsymbol{I}_2 & \boldsymbol{0} \end{bmatrix}, \tag{75}$$

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

The state and measurement noises are assumed to have stationary Gaussian distributions, i.e., $\boldsymbol{w}_k \sim \mathrm{N}(\boldsymbol{q}, \boldsymbol{Q})$ and $\boldsymbol{v}_k \sim \mathrm{N}(\boldsymbol{r}, \boldsymbol{R})$, where the true mean vectors and covariance matrices are given by

$$\boldsymbol{q} = [1 \; 1 \; 1 \; 1]^{\mathrm{T}}, \quad \boldsymbol{Q} = 3 \begin{bmatrix} \frac{\Delta t^3}{3} \boldsymbol{I}_2 & \frac{\Delta t^2}{2} \boldsymbol{I}_2 \\ \frac{\Delta t^2}{2} \boldsymbol{I}_2 & \Delta t\, \boldsymbol{I}_2 \end{bmatrix}, \tag{76}$$

$$\boldsymbol{q} = [1 \; 1]^{\mathrm{T}}, \quad \boldsymbol{R} = \begin{bmatrix} 10 & 5 \\ 5 & 10 \end{bmatrix}. \tag{77}$$

In this simulation, the nominal mean vectors and covariance matrices of state and measurement noises are, respectively, set as $\boldsymbol{q}^* = \boldsymbol{0}_{4\times 1}$, $\boldsymbol{r}^* = \boldsymbol{0}_{2\times 1}$, $\boldsymbol{Q}^* = \boldsymbol{I}_4$ and $\boldsymbol{R}^* = 100\boldsymbol{I}_2$. To show the effectiveness and superiority of the proposed AKF, we compare the performance of the Kalman filter with nominal noise statistics (KFNNS), the Kalman filter with true noise statistics (KFTNS), the existing SHAKF [2], the existing VBAKF for estimating $\boldsymbol{P}_{k|k-1}$ and $\boldsymbol{R}$ (VBAKF-PR) [6], and the proposed AKF with inaccurate noise statistics. Note that, in this simulation, the existing SHAKF often diverges, so its simulation results are not given in the following analysis and comparison. The algorithm parameters of the existing VBAKF-PR are selected as: the tuning parameter $\tau = 3$ and the forgetting factor $\rho = 1$. The algorithm parameters of the proposed AKF are set as: the tuning parameter $\tau = 3$, the priori confidence parameters $\alpha_k = 0.5$ and $\beta_k = 3$, and the initial dof parameter $\nu_0 = 4$. Moreover, in the existing VBAKF-PR, the mean vectors of state and measurement noises are, respectively, set as $\boldsymbol{q}^*$ and $\boldsymbol{r}^*$ due to its inability to estimate unknown mean vectors. The true initial state vector $\boldsymbol{x}_0 = [100, 100, 10, 10]^{\mathrm{T}}$, and the initial estimation error covariance matrix $\boldsymbol{P}_{0|0} = 10\boldsymbol{I}_4$, and the initial state estimate $\hat{\boldsymbol{x}}_{0|0}$ is chosen from $N(\boldsymbol{x}_0, \boldsymbol{P}_{0|0})$ randomly. The number of measurements is 200, and the number of iteration is set as $N = 10$, and 1000 independent Monte Carlo runs are performed. All Kalman filtering algorithms are coded with MATLAB and the used computer has an Intel Core i7-6500U CPU at 2.50 GHz.

The root mean square errors (RMSEs) and the averaged root mean square errors (ARMSEs) of position and velocity are utilized to evaluate the estimation accuracy of state vector. We define the RMSE and ARMSE of position as follows

$$
\left\{
\begin{aligned}
\text{RMSE}_{\text{pos}} &= \sqrt{\frac{1}{M} \sum_{j=1}^{M} \left((x_k^j - \hat{x}_{k|k}^j)^2 + (y_k^j - \hat{y}_{k|k}^j)^2 \right)} \\
\text{ARMSE}_{\text{pos}} &= \sqrt{\frac{1}{MT} \sum_{k=1}^{T} \sum_{j=1}^{M} \left((x_k^j - \hat{x}_{k|k}^j)^2 + (y_k^j - \hat{y}_{k|k}^j)^2 \right)}
\end{aligned}
\right. , \tag{78}
$$

where (x_k^j, y_k^j) and $(\hat{x}_{k|k}^j, \hat{y}_{k|k}^j)$ denote, respectively, the true position and the filtering estimate of position at the j-th Monte Carlo run, and $T = 200$ s denotes the simulation times, and $M = 1000$ denotes the total number of Monte Carlo runs. Similarly, we can also formulate the RMSE and ARMSE of velocity by RMSE_{vel} and $\text{ARMSE}_{\text{vel}}$.

The square root of normalized Euclidean norm (SRNEN) and averaged SRNEN (ASRNEN) are employed to evaluate the estimation accuracy of the one-step prediction and measurement noise mean vector. We define the SRNEN and ASRNEN of the one-step prediction as follows

$$
\left\{
\begin{aligned}
\text{SRNEN}_x &\triangleq \left(\frac{1}{nM} \sum_{j=1}^{M} \| \hat{\boldsymbol{x}}_{k|k-1}^j - \hat{\boldsymbol{x}}_{\text{o},k|k-1}^j \|_2^2 \right)^{\frac{1}{4}} \\
\text{ASRNEN}_x &\triangleq \left(\frac{1}{nMT} \sum_{k=1}^{T} \sum_{j=1}^{M} \| \hat{\boldsymbol{x}}_{k|k-1}^j - \hat{\boldsymbol{x}}_{\text{o},k|k-1}^j \|_2^2 \right)^{\frac{1}{4}}
\end{aligned}
\right. , \tag{79}
$$

where $\| \cdot \|_2$ denotes the Euclidean norm, and $\hat{\boldsymbol{x}}_{k|k-1}^j$ denotes the estimate of one-step prediction at the j-th Monte Carlo run, and $\hat{\boldsymbol{x}}_{\text{o},k|k-1}^j$ denotes the accurate one-step prediction at the j-th Monte Carlo run provided by the KFTNS. Similarly, we can also formulate the SRNEN and ASRNEN of the measurement noise mean vector by SRNEN_r and ASRNEN_r.

The square root of normalized Frobenius norm (SRNFN) and averaged SRNFN (ASRNFN) are used to evaluate the estimate accuracy of the prediction error and measurement noise covariance matrices. The SRNFN and ASRNFN of the prediction error covariance matrix are defined as follows [6]

$$
\left\{
\begin{aligned}
\text{SRNFN}_P &\triangleq \left(\frac{1}{n^2 M} \sum_{j=1}^{M} \| \hat{\boldsymbol{P}}_{k|k-1}^j - \boldsymbol{P}_{\text{o},k|k-1}^j \|_F^2 \right)^{\frac{1}{4}} \\
\text{ASRNFN}_P &\triangleq \left(\frac{1}{n^2 MT} \sum_{k=1}^{T} \sum_{j=1}^{M} \| \hat{\boldsymbol{P}}_{k|k-1}^j - \boldsymbol{P}_{\text{o},k|k-1}^j \|_F^2 \right)^{\frac{1}{4}}
\end{aligned}
\right. , \tag{80}
$$

where $\| \cdot \|_F$ denotes the Frobenius norm, and $\hat{\boldsymbol{P}}_{k|k-1}^j$ denotes the estimate of prediction error covariance matrix at the j-th Monte Carlo run, and $\boldsymbol{P}_{\text{o},k|k-1}^j$ denotes the accurate
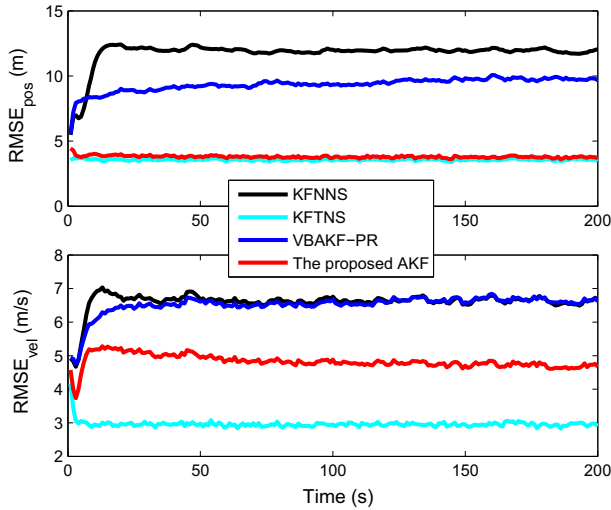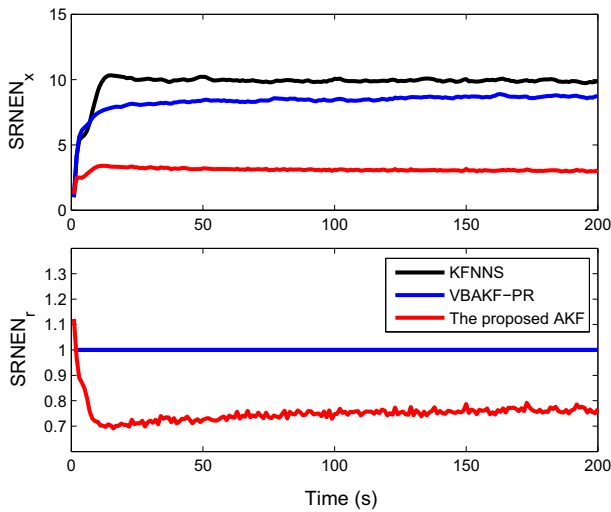
Fig. 2 RMESs of position and velocity

Fig. 3 SRNENs of $\hat{x}_{k|k-1}$ and r

prediction error covariance matrix at the j-th Monte Carlo run provided by the KFTNS. Similarly, we can also formulate the SRNFN and ASRNFN of the measurement noise covariance matrix by SRNFN_R and ASRNFN_R.

Figures 2, 3 and 4 show, respectively, the RMSEs of position and velocity, the SRNENs of the one-step prediction $\hat{x}_{k|k-1}$ and measurement noise mean vector r, and the SRNFNs of the prediction error covariance matrix $P_{k|k-1}$ and measurement noise covariance matrix R, where the black and blue lines coincide in the bottom subfigure of Fig. 3. It is seen from Fig. 2 that the proposed AKF has smaller RMSEs
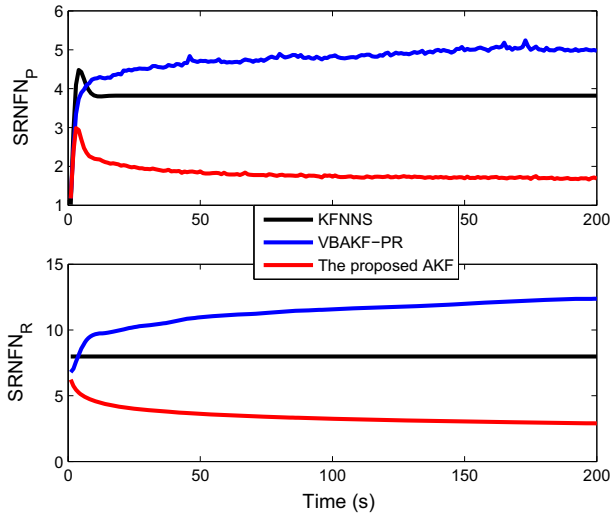
Fig. 4 SRNFNs of $\boldsymbol{P}_{k|k-1}$ and $\boldsymbol{R}$
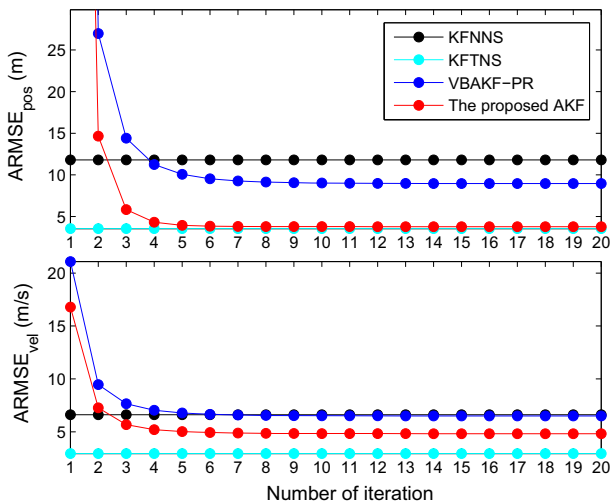
Fig. 5 ARMESs of position and velocity for different numbers of iteration $N = 1, 2, \ldots, 20$

of position and velocity than the existing KFNNS and VBAKF-PR, and the proposed AKF can achieve almost the same RMSEs of position as the optimal KFTNS. We can also see from Figs. 3 and 4 that the proposed AKF has smaller SRNENs and SRNFNs than the existing KFNNS and VBAKF-PR. Furthermore, the implementation times of the proposed AKF and existing KFNNS and VBAKF-PR in a single step run are, respectively, 0.63 ms, 0.02 ms and 0.46 ms. Thus, the proposed AKF can better estimate the statistical parameters $\left\{ \hat{\boldsymbol{x}}_{k|k-1}, \boldsymbol{r}, \boldsymbol{P}_{k|k-1}, \boldsymbol{R} \right\}$ but has higher computational
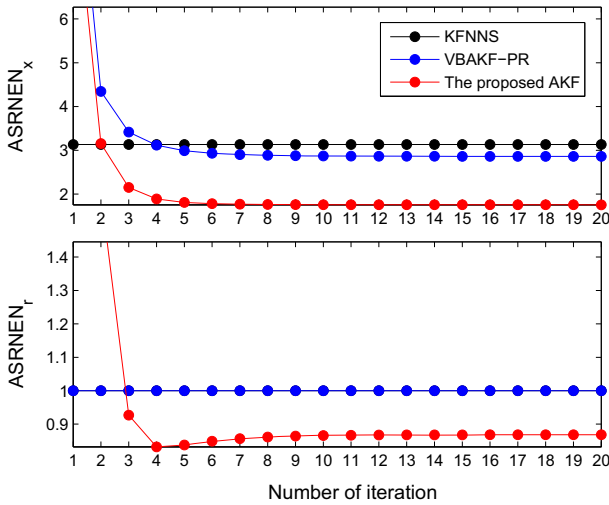
Fig. 6 ASRNENs of $\hat{x}_{k|k-1}$ and r for different numbers of iteration $N = 1, 2, \ldots, 20$
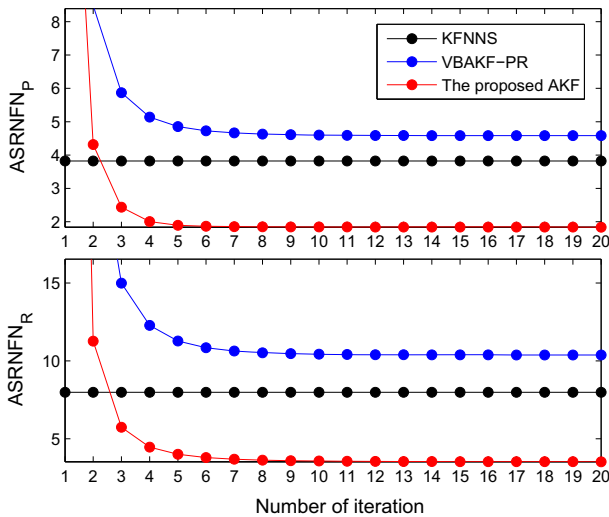

Fig. 7 ASRNFNs of $P_{k|k-1}$ and R for different numbers of iteration $N = 1, 2, \ldots, 20$

complexity as compared with the existing KFNNS and VBAKF-PR, which results in improved state estimation accuracy.

Figures 5, 6 and 7 show, respectively, the ARMSEs of position and velocity, the ASRNENs of $\hat{x}_{k|k-1}$ and r, and the ASRNFNs of $P_{k|k-1}$ and R when different numbers of iteration $N = 1, 2, \ldots, 20$ are selected, where the black and blue lines coincide in the bottom subfigure of Fig. 6. It can be observed from Figs. 5, 6 and 7 that the proposed AKF has smaller ARMSEs, ASRNENs and ASRNFNs than the existing KFNNS and VBAKF-PR when $N \geq 3$, and the proposed AKF has almost the same
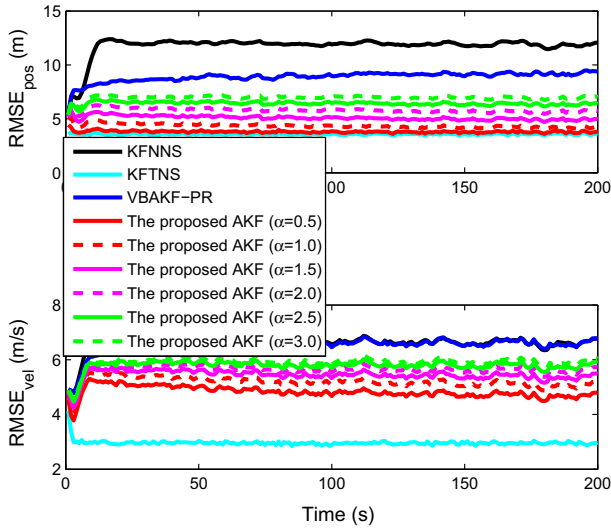
Fig. 8 RMESs of position and velocity for different priori confidence parameters $\alpha_k = 0.5, 1.0, 1.5, 2.0, 2.5, 3.0$
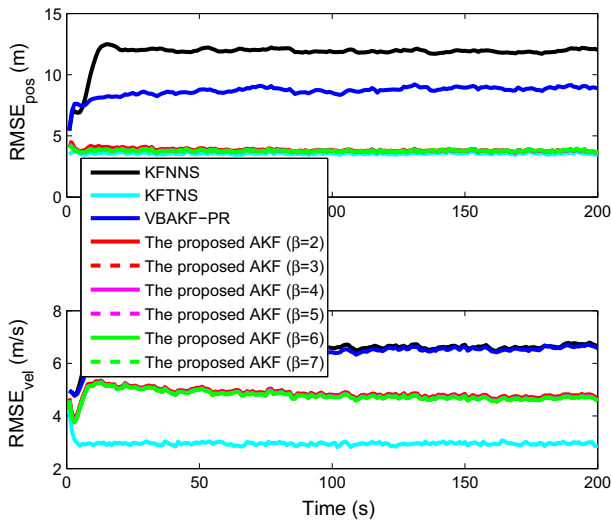

Fig. 9 RMESs of position and velocity for different priori confidence parameters $\beta_k = 2, 3, 4, 5, 6, 7$

ARMSEs of position as the optimal KFTNS when $N \geq 5$. Thus, the proposed AKF has better accuracy for jointly estimating position, velocity and statistical parameters $\{\hat{x}_{k|k-1}, r, P_{k|k-1}, R\}$ when $N \geq 3$. Moreover, we can observe from Figs. 5, 6 and 7 that the ARMSEs, ASRNENs and ASRNFNs of the proposed AKF all converge when $N \geq 10$. Therefore, the number of iteration $N = 10$ is enough to guarantee the performance of the proposed AKF.
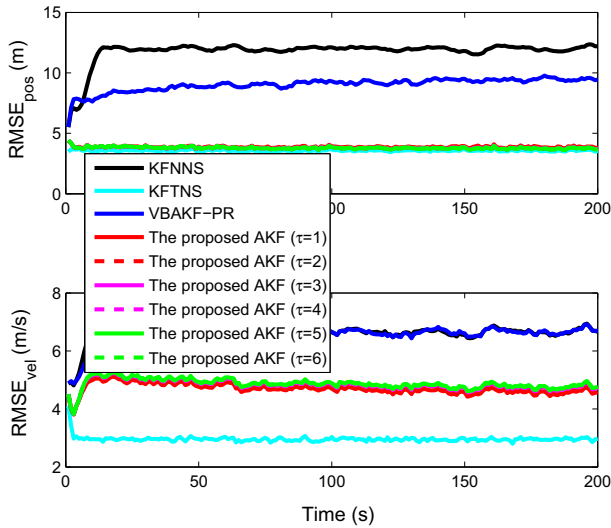
Fig. 10 RMESs of position and velocity for different tuning parameters $\tau = 1, 2, 3, 4, 5, 6$
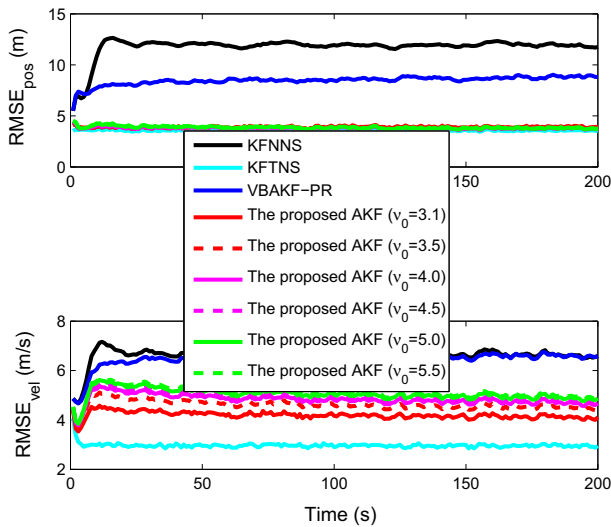

Fig. 11 RMESs of position and velocity for different initial dof parameters $\nu_0 = 3.1, 3.5, 4.0, 4.5, 5.0, 5.5$

Figures 8, 9, 10 and 11 show the RMSEs of position and velocity when different priori confidence parameters $\alpha_k = 0.5, 1.0, 1.5, 2.0, 2.5, 3.0$ and $\beta_k = 2, 3, 4, 5, 6, 7$, different tuning parameters $\tau = 1, 2, 3, 4, 5, 6$, and different initial dof parameters $\nu_0 = 3.1, 3.5, 4.0, 4.5, 5.0, 5.5$ are, respectively, selected. It is seen from Fig. 8 that the proposed AKFs with selected priori confidence parameters α_k have better estimation accuracy than existing KFNNS and VBAKF-PR, and the proposed AKF with $\alpha_k = 0.5$ achieves the best estimation accuracy. It can be seen from Figs. 9 and 10 that the

proposed AKFs with selected priori confidence parameters β_k and tuning parameters τ have almost consistent estimation accuracy and achieve better estimation accuracy than the existing KFNNS and VBAKF-PR. We can see from Fig. 11 that the proposed AKFs with selected initial dof parameters have almost consistent estimation accuracy of position and better estimation accuracy as compared with the existing KFNNS and VBAKF-PR.

6 Conclusion

In this paper, a new AKF was proposed for a linear Gaussian state-space model with inaccurate noise statistics based on the VB approach. To resist the uncertainties of noise statistics, the state vector, the one-step prediction and corresponding prediction error covariance matrix, and the mean vector and covariance matrix of measurement noise are jointly estimated based on the VB approach. Both the prior joint PDFs of the one-step prediction and corresponding prediction error covariance matrix and the joint PDF of the mean vector and covariance matrix of measurement noise are selected as NIW distributions, based on which the posterior PDFs of these unknown parameters are, respectively, approximated by Gaussian and IW distributions using the VB approach. Simulation results of a target tracking example illustrated that the proposed AKF has better estimation accuracy but higher computational complexity than existing state-of-the-art AKFs, which is induced by the fact that the proposed AKF can better estimate the one-step prediction and corresponding prediction error covariance matrix and the mean vector and covariance matrix of measurement noise as compared with existing state-of-the-art AKFs.

Appendices

A. Proofs of (29)–(33)

Substituting $\boldsymbol{\theta} = \hat{\boldsymbol{x}}_{k|k-1}$, $\boldsymbol{\theta} = \boldsymbol{r}$ and (28) in (26), $\log q^{(i+1)}(\hat{\boldsymbol{x}}_{k|k-1})$ and $\log q^{(i+1)}(\boldsymbol{r})$ are written as

$$
\begin{aligned}
\log q^{(i+1)}(\hat{\boldsymbol{x}}_{k|k-1}) = &-\frac{1}{2}\mathrm{E}^{(i)}\left[\left(\boldsymbol{x}_k - \hat{\boldsymbol{x}}_{k|k-1}\right)^{\mathrm{T}} \boldsymbol{P}_{k|k-1}^{-1} \left(\boldsymbol{x}_k - \hat{\boldsymbol{x}}_{k|k-1}\right)\right] \\
&-\frac{1}{2}\alpha_k \left(\hat{\boldsymbol{x}}_{k|k-1} - \boldsymbol{u}_k\right)^{\mathrm{T}} \mathrm{E}^{(i)}\left[\boldsymbol{P}_{k|k-1}^{-1}\right]\left(\hat{\boldsymbol{x}}_{k|k-1} - \boldsymbol{u}_k\right) + c_{\hat{\boldsymbol{x}}_{k|k-1}},
\end{aligned}
\tag{81}
$$

$$
\begin{aligned}
\log q^{(i+1)}(\boldsymbol{r}) = &-\frac{1}{2}\mathrm{E}^{(i)}\left[(\boldsymbol{z}_k - \boldsymbol{H}_k\boldsymbol{x}_k - \boldsymbol{r})^{\mathrm{T}} \boldsymbol{R}^{-1} (\boldsymbol{z}_k - \boldsymbol{H}_k\boldsymbol{x}_k - \boldsymbol{r})\right] \\
&-\frac{1}{2}\beta_k (\boldsymbol{r} - \boldsymbol{\lambda}_k)^{\mathrm{T}} \mathrm{E}^{(i)}\left[\boldsymbol{R}^{-1}\right](\boldsymbol{r} - \boldsymbol{\lambda}_k) + c_{\boldsymbol{r}}.
\end{aligned}
\tag{82}
$$

The first expectations in (81)–(82) are calculated as follows

$$E^{(i)} \left[\left(\boldsymbol{x}_k - \hat{\boldsymbol{x}}_{k|k-1} \right)^{\mathrm{T}} \boldsymbol{P}_{k|k-1}^{-1} \left(\boldsymbol{x}_k - \hat{\boldsymbol{x}}_{k|k-1} \right) \right]$$

$$= \mathrm{tr} \left\{ E^{(i)} \left[\left(\boldsymbol{x}_k - \hat{\boldsymbol{x}}_{k|k-1} \right) \left(\boldsymbol{x}_k - \hat{\boldsymbol{x}}_{k|k-1} \right)^{\mathrm{T}} \right] E^{(i)} \left[\boldsymbol{P}_{k|k-1}^{-1} \right] \right\}$$

$$= \mathrm{tr} \left\{ E^{(i)} \left[\left(\boldsymbol{x}_k - \hat{\boldsymbol{x}}_{k|k}^{(i)} + \hat{\boldsymbol{x}}_{k|k}^{(i)} - \hat{\boldsymbol{x}}_{k|k-1} \right) \left(\boldsymbol{x}_k - \hat{\boldsymbol{x}}_{k|k}^{(i)} + \quad \hat{\boldsymbol{x}}_{k|k}^{(i)} - \hat{\boldsymbol{x}}_{k|k-1} \right)^{\mathrm{T}} \right] \times \right.$$

$$E^{(i)} \left[\boldsymbol{P}_{k|k-1}^{-1} \right] \right\}$$

$$= \left(\hat{\boldsymbol{x}}_{k|k}^{(i)} - \hat{\boldsymbol{x}}_{k|k-1} \right)^{\mathrm{T}} E^{(i)} \left[\boldsymbol{P}_{k|k-1}^{-1} \right] \left(\hat{\boldsymbol{x}}_{k|k}^{(i)} - \hat{\boldsymbol{x}}_{k|k-1} \right) + c_{\hat{\boldsymbol{x}}_{k|k-1}}, \tag{83}$$

$$E^{(i)} \left[\left(\boldsymbol{z}_k - \boldsymbol{H}_k \boldsymbol{x}_k - \boldsymbol{r} \right)^{\mathrm{T}} \boldsymbol{R}^{-1} \left(\boldsymbol{z}_k - \boldsymbol{H}_k \boldsymbol{x}_k - \boldsymbol{r} \right) \right]$$

$$= \mathrm{tr} \{ E^{(i)} \left[\left(\boldsymbol{z}_k - \boldsymbol{H}_k \boldsymbol{x}_k - \boldsymbol{r} \right) \left(\boldsymbol{z}_k - \boldsymbol{H}_k \boldsymbol{x}_k - \boldsymbol{r} \right)^{\mathrm{T}} \right] E^{(i)} \left[\boldsymbol{R}^{-1} \right] \}$$

$$= \mathrm{tr} \{ E^{(i)} [(\boldsymbol{z}_k - \boldsymbol{H}_k \hat{\boldsymbol{x}}_{k|k}^{(i)} - \boldsymbol{r} - \boldsymbol{H}_k (\boldsymbol{x}_k - \hat{\boldsymbol{x}}_{k|k}^{(i)})) (\boldsymbol{z}_k - \boldsymbol{H}_k \hat{\boldsymbol{x}}_{k|k}^{(i)} - \boldsymbol{r} -$$

$$\boldsymbol{H}_k (\boldsymbol{x}_k - \hat{\boldsymbol{x}}_{k|k}^{(i)}))^{\mathrm{T}}] E^{(i)} \left[\boldsymbol{R}^{-1} \right] \}$$

$$= \left(\boldsymbol{z}_k - \boldsymbol{H}_k \hat{\boldsymbol{x}}_{k|k}^{(i)} - \boldsymbol{r} \right)^{\mathrm{T}} E^{(i)} \left[\boldsymbol{R}^{-1} \right] \left(\boldsymbol{z}_k - \boldsymbol{H}_k \hat{\boldsymbol{x}}_{k|k}^{(i)} - \boldsymbol{r} \right) + c_{\boldsymbol{r}}. \tag{84}$$

Substituting (83)–(84) in (81)–(82) and using (33) yields

$$\log q^{(i+1)}(\hat{\boldsymbol{x}}_{k|k-1}) = -\frac{1}{2} \left(\hat{\boldsymbol{x}}_{k|k}^{(i)} - \hat{\boldsymbol{x}}_{k|k-1} \right)^{\mathrm{T}} \left[\bar{\boldsymbol{P}}_{k|k-1}^{(i)} \right]^{-1} \left(\hat{\boldsymbol{x}}_{k|k}^{(i)} - \hat{\boldsymbol{x}}_{k|k-1} \right)$$

$$- \frac{1}{2} \alpha_k \left(\hat{\boldsymbol{x}}_{k|k-1} - \boldsymbol{u}_k \right)^{\mathrm{T}} \left[\bar{\boldsymbol{P}}_{k|k-1}^{(i)} \right]^{-1} \left(\hat{\boldsymbol{x}}_{k|k-1} - \boldsymbol{u}_k \right) + c_{\hat{\boldsymbol{x}}_{k|k-1}}, \tag{85}$$

$$\log q^{(i+1)}(\boldsymbol{r}) = -\frac{1}{2} \left(\boldsymbol{z}_k - \boldsymbol{H}_k \hat{\boldsymbol{x}}_{k|k}^{(i)} - \boldsymbol{r} \right)^{\mathrm{T}} \left[\bar{\boldsymbol{R}}_k^{(i)} \right]^{-1} \left(\boldsymbol{z}_k - \boldsymbol{H}_k \hat{\boldsymbol{x}}_{k|k}^{(i)} - \boldsymbol{r} \right)$$

$$- \frac{1}{2} \beta_k \left(\boldsymbol{r} - \boldsymbol{\lambda}_k \right)^{\mathrm{T}} \left[\bar{\boldsymbol{R}}_k^{(i)} \right]^{-1} \left(\boldsymbol{r} - \boldsymbol{\lambda}_k \right) + c_{\boldsymbol{r}}. \tag{86}$$

Utilizing (85)–(86), we can obtain (29)–(32).

B. Proofs of (34)–(43)

Exploiting $\boldsymbol{\theta} = \boldsymbol{P}_{k|k-1}, \boldsymbol{\theta} = \boldsymbol{R}$ and (28) in (26), $\log q^{(i+1)}(\boldsymbol{P}_{k|k-1})$ and $\log q^{(i+1)}(\boldsymbol{R})$ are calculated as

$$\log q^{(i+1)}(\boldsymbol{P}_{k|k-1}) = -\frac{1}{2}(\omega_k + n + 3)\log\left|\boldsymbol{P}_{k|k-1}\right|$$

$$-\frac{1}{2}\mathrm{E}^{(i)}\left[\left(\boldsymbol{x}_k - \hat{\boldsymbol{x}}_{k|k-1}\right)^{\mathrm{T}}\boldsymbol{P}_{k|k-1}^{-1}\left(\boldsymbol{x}_k - \hat{\boldsymbol{x}}_{k|k-1}\right)\right]$$

$$-\frac{1}{2}\alpha_k\mathrm{E}^{(i)}\left[\left(\hat{\boldsymbol{x}}_{k|k-1} - \boldsymbol{u}_k\right)^{\mathrm{T}}\boldsymbol{P}_{k|k-1}^{-1}(\hat{\boldsymbol{x}}_{k|k-1} - \boldsymbol{u}_k)\right]$$

$$-\frac{1}{2}\mathrm{tr}\left\{\boldsymbol{\Sigma}_k\boldsymbol{P}_{k|k-1}^{-1}\right\} + c_{\boldsymbol{P}_{k|k-1}}, \tag{87}$$

$$\log q^{(i+1)}(\boldsymbol{R}) = -\frac{1}{2}(\nu_k + m + 3)\log|\boldsymbol{R}|$$

$$-\frac{1}{2}\mathrm{E}^{(i)}\left[(\boldsymbol{z}_k - \boldsymbol{H}_k\boldsymbol{x}_k - \boldsymbol{r})^{\mathrm{T}}\boldsymbol{R}^{-1}\left(\boldsymbol{z}_k - \boldsymbol{H}_k\boldsymbol{x}_k - \boldsymbol{r}\right)\right]$$

$$-\frac{1}{2}\beta_k\mathrm{E}^{(i)}\left[(\boldsymbol{r} - \boldsymbol{\lambda}_k)^{\mathrm{T}}\boldsymbol{R}^{-1}\left(\boldsymbol{r} - \boldsymbol{\lambda}_k\right)\right] - \frac{1}{2}\mathrm{tr}\left\{\boldsymbol{\Delta}_k\boldsymbol{R}^{-1}\right\} + c_{\boldsymbol{R}}. \tag{88}$$

Substituting (40)–(43) in (87)–(88) gives

$$\log q^{(i+1)}(\boldsymbol{P}_{k|k-1}) = -\frac{1}{2}(\omega_k + n + 3)\log\left|\boldsymbol{P}_{k|k-1}\right|$$

$$-\frac{1}{2}\mathrm{tr}\left\{\left(\boldsymbol{A}_k^{(i+1)} + \boldsymbol{B}_k^{(i+1)} + \boldsymbol{\Sigma}_k\right)\boldsymbol{P}_{k|k-1}^{-1}\right\} + c_{\boldsymbol{P}_{k|k-1}}, \tag{89}$$

$$\log q^{(i+1)}(\boldsymbol{R}) = -\frac{1}{2}(\nu_k + m + 3)\log|\boldsymbol{R}|$$

$$-\frac{1}{2}\mathrm{tr}\left\{\left(\boldsymbol{C}_k^{(i+1)} + \boldsymbol{D}_k^{(i+1)} + \boldsymbol{\Delta}_k\right)\boldsymbol{R}^{-1}\right\} + c_{\boldsymbol{R}}. \tag{90}$$

According to (89)–(90), we can obtain (34)–(39).

C. Proofs of (44)–(47)

Using $\boldsymbol{\theta} = \boldsymbol{x}_k$ and (28) in (26), $\log q^{(i+1)}(\boldsymbol{x}_k)$ can be formulated as

$$\log q^{(i+1)}(\boldsymbol{x}_k) = -\frac{1}{2}\mathrm{tr}\left\{\mathrm{E}^{(i+1)}\left[(\boldsymbol{z}_k - \boldsymbol{H}_k\boldsymbol{x}_k - \boldsymbol{r})\times \ (\boldsymbol{z}_k - \boldsymbol{H}_k\boldsymbol{x}_k - \boldsymbol{r})^{\mathrm{T}}\right]\times\mathrm{E}^{(i+1)}\left[\boldsymbol{R}^{-1}\right]\right\}$$

$$-\frac{1}{2}\mathrm{tr}\left\{\mathrm{E}^{(i+1)}\left[\left(\boldsymbol{x}_k - \hat{\boldsymbol{x}}_{k|k-1}\right)\left(\boldsymbol{x}_k - \hat{\boldsymbol{x}}_{k|k-1}\right)^{\mathrm{T}}\right]\mathrm{E}^{(i+1)}\left[\boldsymbol{P}_{k|k-1}^{-1}\right]\right\} + c_{\boldsymbol{x}_k}, \tag{91}$$

where the first and third expectations are calculated as

$$\mathrm{E}^{(i+1)}\left[\left(z_k - H_k x_k - r\right)\left(z_k - H_k x_k - r\right)^{\mathrm{T}}\right]$$

$$= \mathrm{E}^{(i+1)}\left[\left(z_k - H_k x_k - \hat{\boldsymbol{\lambda}}_k^{(i+1)} + \hat{\boldsymbol{\lambda}}_k^{(i+1)} - r\right)\left(z_k - H_k x_k - \hat{\boldsymbol{\lambda}}_k^{(i+1)} + \hat{\boldsymbol{\lambda}}_k^{(i+1)} - r\right)^{\mathrm{T}}\right]$$

$$= \left(z_k - \hat{\boldsymbol{\lambda}}_k^{(i+1)} - H_k x_k\right)\left(z_k - \hat{\boldsymbol{\lambda}}_k^{(i+1)} - H_k x_k\right)^{\mathrm{T}} + \hat{\boldsymbol{\Omega}}_k^{(i+1)}, \tag{92}$$

$$\mathrm{E}^{(i+1)}\left[\left(x_k - \hat{x}_{k|k-1}\right)\left(x_k - \hat{x}_{k|k-1}\right)^{\mathrm{T}}\right]$$

$$= \mathrm{E}^{(i+1)}\left[\left(x_k - \hat{u}_k^{(i+1)} + \hat{u}_k^{(i+1)} - \hat{x}_{k|k-1}\right)\left(x_k - \hat{u}_k^{(i+1)} + \hat{u}_k^{(i+1)} - \hat{x}_{k|k-1}\right)^{\mathrm{T}}\right]$$

$$= \left(x_k - \hat{u}_k^{(i+1)}\right)\left(x_k - \hat{u}_k^{(i+1)}\right)^{\mathrm{T}} + \hat{U}_k^{(i+1)}. \tag{93}$$

Employing (31) and (92)–(93) in (91) yields

$$\log q^{(i+1)}(x_k) = -\frac{1}{2}\left(z_k - \hat{\boldsymbol{\lambda}}_k^{(i+1)} - H_k x_k\right)^{\mathrm{T}}\left[\bar{R}_k^{(i)}\right]^{-1}\left(z_k - \hat{\boldsymbol{\lambda}}_k^{(i+1)} - H_k x_k\right)$$

$$- \frac{1}{2}\left(x_k - \hat{u}_k^{(i+1)}\right)^{\mathrm{T}} \times \left[\bar{P}_{k|k-1}^{(i)}\right]^{-1}\left(x_k - \hat{u}_k^{(i+1)}\right) + c_{x_k}, \tag{94}$$

According to (94), we can obtain (44)–(47), where (45)–(47) is given by the measurement update of the Kalman filter.

D. Proofs of (55)–(58)

Exploiting (29)–(30) and (44), the auxiliary parameters $A_k^{(i+1)}$, $B_k^{(i+1)}$, $C_k^{(i+1)}$ and $D_k^{(i+1)}$ in (40)–(43) can be, respectively, calculated as

$$A_k^{(i+1)} = \mathrm{E}^{(i)}\left[\left(x_k - \hat{x}_{k|k}^{(i)} + \hat{x}_{k|k}^{(i)} - \hat{u}_k^{(i+1)} + \hat{u}_k^{(i+1)} - \hat{x}_{k|k-1}\right)\left(x_k - \hat{x}_{k|k}^{(i)}\right.\right.$$

$$\left.\left. + \hat{x}_{k|k}^{(i)} - \hat{u}_k^{(i+1)} + \hat{u}_k^{(i+1)} - \hat{x}_{k|k-1}\right)^{\mathrm{T}}\right]$$

$$= P_{k|k}^{(i)} + \left(\hat{x}_{k|k}^{(i)} - \hat{u}_k^{(i+1)}\right)\left(\hat{x}_{k|k}^{(i)} - \hat{u}_k^{(i+1)}\right)^{\mathrm{T}} + \hat{U}_k^{(i+1)}, \tag{95}$$

$$B_k^{(i+1)} = \alpha_k \mathrm{E}^{(i+1)}\left[\left(\hat{x}_{k|k-1} - \hat{u}_k^{(i+1)} + \hat{u}_k^{(i+1)} - u_k\right)\right.$$

$$\left. \times \left(\hat{x}_{k|k-1} - \hat{u}_k^{(i+1)} + \hat{u}_k^{(i+1)} - u_k\right)^{\mathrm{T}}\right]$$

$$= \alpha_k \hat{U}_k^{(i+1)} + \alpha_k\left(\hat{u}_k^{(i+1)} - u_k\right)\left(\hat{u}_k^{(i+1)} - u_k\right)^{\mathrm{T}}, \tag{96}$$

$$C_k^{(i+1)} = \mathrm{E}^{(i)}\left[\left(z_k - H_k \hat{x}_{k|k}^{(i)} - \hat{\boldsymbol{\lambda}}_k^{(i+1)} - H_k(x_k - \hat{x}_{k|k}^{(i)}) + \hat{\boldsymbol{\lambda}}_k^{(i+1)} - r\right)\right.$$

$$\left. \times \left(z_k - H_k \hat{x}_{k|k}^{(i)} - \hat{\boldsymbol{\lambda}}_k^{(i+1)} - H_k(x_k - \hat{x}_{k|k}^{(i)}) + \hat{\boldsymbol{\lambda}}_k^{(i+1)} - r\right)^{\mathrm{T}}\right]$$

$$= H_k P_{k|k}^{(i)} H_k^{\mathrm{T}} + \hat{\boldsymbol{\Omega}}_k^{(i+1)} + \left(z_k - H_k \hat{x}_{k|k}^{(i)} - \hat{\boldsymbol{\lambda}}_k^{(i+1)}\right)\left(z_k - H_k \hat{x}_{k|k}^{(i)} - \hat{\boldsymbol{\lambda}}_k^{(i+1)}\right)^{\mathrm{T}}, \tag{97}$$

$$D_k^{(i+1)} = \beta_k \mathrm{E}^{(i+1)}\left[\left(r - \hat{\boldsymbol{\lambda}}_k^{(i+1)} + \hat{\boldsymbol{\lambda}}_k^{(i+1)} - \lambda_k\right)\left(r - \hat{\boldsymbol{\lambda}}_k^{(i+1)} + \hat{\boldsymbol{\lambda}}_k^{(i+1)} - \lambda_k\right)^{\mathrm{T}}\right]$$

$$= \beta_k \hat{\boldsymbol{\Omega}}_k^{(i+1)} + \beta_k\left(\hat{\boldsymbol{\lambda}}_k^{(i+1)} - \lambda_k\right)\left(\hat{\boldsymbol{\lambda}}_k^{(i+1)} - \lambda_k\right)^{\mathrm{T}}. \tag{98}$$

According to (95)–(98), we can obtain (55)–(58).

References

1. C.M. Bishop, *Pattern Recognition and Machine Learning* (Springer, Berlin, 2007)
2. X. Gao, D. You, S. Katayama, Seam tracking monitoring based on adaptive Kalman filter embedded elman neural network during high-power fiber laser welding. IEEE Trans. Ind. Electron. **59**(11), 4315–4325 (2012)
3. C. Hide, T. Moore, M. Smith, Adaptive Kalman filtering algorithms for integrating GPS and low cost INS, in *Proceedings of Position Location Navigation Symposium*. IEEE, Monterey, CA, USA, pp. 227–233 (2004)
4. Y.L. Huang, Y.G. Zhang, B. Xu, Z.M. Wu, J. Chambers, A new outlier-robust Student's t based Gaussian approximate filter for cooperative localization. IEEE/AMSE Trans. Mech. **22**(5), 2380–2386 (2017)
5. Y.L. Huang, Y.G. Zhang, A new process uncertainty robust Student's t based Kalman filter for SINS/GPS integration. IEEE Access **5**(7), 14391–14404 (2017)
6. Y.L. Huang, Y.G. Zhang, Z.M. Wu, N. Li, J. Chambers, A novel adaptive Kalman filter with inaccurate process and measurement noise covariance matrices. IEEE Trans. Autom. Control **63**(2), 594–601 (2018)
7. Y.L. Huang, Y.G. Zhang, B. Xu, Z.M. Wu, J. Chambers, A new adaptive extended Kalman filter for cooperative localization. IEEE Trans. Aerosp. Electron. Syst. **54**(1), 353–368 (2018)
8. Y.L. Huang, Y.G. Zhang, Z.M. Wu, N. Li, J. Chambers, A novel robust Student's t based Kalman filter. IEEE Trans. Aerosp. Electron. Syst. **53**(3), 1545–1554 (2017)
9. Y.L. Huang, Y.G. Zhang, N. Li, J. Chambers, A robust Gaussian approximate filter for nonlinear systems with heavy-tailed measurement noises, in *2016 IEEE International Conference on Acoustics, Speech and Signal Processing* (ICASSP) (IEEE, Shanghai, China), pp. 4209–4213 (2016)
10. Y.L. Huang, Y.G. Zhang, N. Li, J. Chambers, A robust Gaussian approximate fixed-interval smoother for nonlinear systems with heavy-tailed process and measurement noises. IEEE Signal Proc. Lett. **23**(4), 468–472 (2016)
11. M. Karasalo, X.M. Hu, An optimization approach to adaptive Kalman filtering. Automatica **47**(8), 1785–1793 (2011)
12. X.R. Li, Y. Bar-Shalom, A recursive multiple model approach to noise identification. IEEE Trans. Aerosp. Electron. Syst. **30**(3), 671–684 (1994)
13. R. Mehra, Approaches to adaptive filtering. IEEE Trans. Autom. Control **17**(5), 693–698 (1972)
14. A. Mohamed, K.P. Schwarz, Adaptive Kalman filtering for INS/GPS. J. Geod. **73**(4), 193–203 (1999)
15. K.P. Murphy, Conjugate Bayesian analysis of the Gaussian distribution. Technical report (2007)
16. D. Simon, *Optimal State Estimation: Kalman, H Infinity, and Nonlinear Approaches* (Wiley, New York, 2006)
17. A.P. Sage, G.W. Husa, Adaptive filtering with unknown prior statistics, in *Proceedings of Joint Automatic Control Conference*. Boulder, CO, pp. 760–769 (1969)
18. S. Särkkä, A. Nummenmaa, Recursive noise adaptive Kalman filtering by variational Bayesian approximations. IEEE Trans. Autom. Control. **54**(3), 596–600 (2009)
19. M.J. Yu, INS/GPS integration system using adaptive filter for estimating measurement noise variance. IEEE Trans. Aerosp. Electron. Syst. **48**(2), 1786–1792 (2012)