Huang Y, Zhang Y, Wu Z, Li N, Chambers J. A novel adaptive Kalman filter with inaccurate process and measurement noise covariance matrices. *IEEE Transactions on Automatic Control* 2017. DOI: 10.1109/TAC.2017.2730480

DOI link to article:

https://doi.org/10.1109/TAC.2017.2730480

Date deposited:

29/11/2017

A novel adaptive Kalman filter with inaccurate process and measurement noise covariance matrices

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

***Abstract*—In this paper, a novel variational Bayesian (VB) based adaptive Kalman filter (VBAKF) for linear Gaussian state-space models with inaccurate process and measurement noise covariance matrices is proposed. By choosing inverse Wishart priors, the state together with the predicted error and measurement noise covariance matrices are inferred based on the VB approach. Simulation results for a target tracking example illustrate that the proposed VBAKF has better robustness to resist the uncertainties of process and measurement noise covariance matrices than existing state-of-the-art filters.**

***Index Terms*—Adaptive filtering, variational Bayesian, Kalman filtering, time-varying noise covariance matrices, inverse Wishart distribution**

I. Introduction

THE Kalman filter is an optimal state estimator for linear Gaussian state-space models, and it has been widely used in many applications, such as navigation, target tracking and control. The performance of the Kalman filter depends largely on a priori knowledge of the noise statistics, and the use of wrong a priori statistics can result in substantial estimation errors or even filtering divergence [1]. However, in many applications, such as Global Positioning System (GPS) and Inertial Navigation System (INS) based integrated navigation systems, their noise statistics may be unknown and time-varying [2], [3], [4]. The adaptive Kalman filter (AKF) is the most common method to solve this problem, and it can be divided into correlation, covariance matching, maximum likelihood and Bayesian methods [1].

The Sage-Husa AKF (SHAKF) is a covariance matching method, which estimates the noise statistics recursively based on the maximum a posterior criterion [5], [6]. However, the convergence to the right noise covariance matrices is not guaranteed with SHAKF, which may lead to filtering divergence [1]. The Innovation-based AKF (IAKF) is a maximum likelihood method, which estimates the noise covariance matrices based on the fact that the innovation sequence of the Kalman filter is a white process [2]. However, the IAKF requires rather large windows of data to obtain reliable estimations of noise covariance matrices, which makes it impractical for rapidly varying noise covariance matrices [7]. The multiple model AKF (MMAKF) is an approximation of the Bayesian method, which can deal with the model uncertainty by operating a bank of Kalman filters with different models simultaneously [8]. However, the MMAKF suffers from substantial computational complexities [9].

The existing variational Bayesian (VB) based AKF (VBAKF) is also an approximation of the Bayesian method, which can estimate an inaccurate and slowly varying measurement noise covariance matrix (MNCM) by choosing appropriate conjugate prior distribution [9]–[12]. However, the performance of the existing VBAKF will degrade for an inaccurate process noise covariance matrix (PNCM) since it assumes accurate PNCM. Although the VB based Rauch-Tung-Striebel smoother can estimate unknown PNCM and MNCM simultaneously [13], [14], it can only estimate unknown and constant noise covariance matrices off-line. To the best of the knowledge of the authors, it is always a challenge to design a VBAKF for linear Gaussian state-space models with inaccurate PNCM and MNCM since the PNCM is difficult to be estimated directly with a rather small window of data.

In this paper, a novel VBAKF with inaccurate PNCM and MNCM is proposed. By choosing inverse Wishart priors for the predicted error covariance matrix (PECM) and MNCM, the state together with PECM and MNCM are inferred based on the VB approach. The proposed VBAKF and existing filters are applied to the problem of target tracking with inaccurate and slowly varying PNCM and MNCM. Simulation results show the proposed filter has smaller root mean square error (RMSE) than existing state-of-the-art filters.

II. Main results

A. Problem formulation

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

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

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

where (1) and (2) are respectively process and measurement equations, k is the discrete time index, $\mathbf{x}_k \in \mathbb{R}^n$ is the state vector, $\mathbf{z}_k \in \mathbb{R}^m$ is the measurement vector, $\mathbf{F}_k \in \mathbb{R}^{n\times n}$ is the state transition matrix, $\mathbf{H}_k \in \mathbb{R}^{m\times n}$ is the observation matrix; $\mathbf{w}_k \in \mathbb{R}^n$ and $\mathbf{v}_k \in \mathbb{R}^m$ are respectively Gaussian process and measurement noise vectors with zero mean vectors and covariance matrices $\mathbf{Q}_k$ and $\mathbf{R}_k$. The initial state vector $\mathbf{x}_0$ is assumed to have a Gaussian distribution with mean vector $\hat{\mathbf{x}}_{0|0}$ and covariance matrix $\mathbf{P}_{0|0}$. Moreover, $\mathbf{x}_0$, $\mathbf{w}_k$ and $\mathbf{v}_j$ are assumed to be mutually uncorrelated for any j and k.

The Kalman filter is frequently employed to estimate the state vector $\mathbf{x}_k$ given the state-space model and measurements $\mathbf{z}_{1:k}$, where $\mathbf{z}_{1:k} = \{\mathbf{z}_j\}_{j=1}^k$ denotes the measurements from time 1 to time k. The Kalman filter is optimal in terms of minimum mean square error (MMSE) for linear Gaussian state-space model (1)-(2) with accurate $\mathbf{Q}_k$ and $\mathbf{R}_k$. However, the use of wrong/inaccurate $\mathbf{Q}_k$ and $\mathbf{R}_k$ can result in substantial estimation errors or even filtering divergence [1]. Therefore, a novel VBAKF suitable for operation with inaccurate PNCM and MNCM will be proposed.

This work was supported by the National Natural Science Foundation of China under Grant Nos. 61371173 and 61633008 and the Natural Science Foundation of Heilongjiang Province Grant No. F2016008 and the Fundamental Research Founds for the Central University of Harbin Engineering University under Grant Nos. HEUCFP201705 and HEUCF041702 and the PhD Student Research and Innovation Fund of the Fundamental Research Founds for the Central Universities under Grant No. HEUGIP201706 and China Scholarship Council Foundation and the Engineering and Physical Sciences Research Council (EPSRC) of the UK Grant No. EP/K014307/1.
Corresponding author is Y. G. Zhang.

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

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

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

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TAC.2017.2730480, IEEE Transactions on Automatic Control

B. The choices of prior distributions

In the framework of the Kalman filter, the one-step predicted PDF $p(\mathbf{x}_k|\mathbf{z}_{1:k-1})$ and likelihood PDF $p(\mathbf{z}_k|\mathbf{x}_k)$ are Gaussian, i.e.

$$p(\mathbf{x}_k|\mathbf{z}_{1:k-1},\mathbf{P}_{k|k-1}) = \mathrm{N}(\mathbf{x}_k;\hat{\mathbf{x}}_{k|k-1},\mathbf{P}_{k|k-1}) \tag{3}$$

$$p(\mathbf{z}_k|\mathbf{x}_k,\mathbf{R}_k) = \mathrm{N}(\mathbf{z}_k;\mathbf{H}_k\mathbf{x}_k,\mathbf{R}_k) \tag{4}$$

where $\mathrm{N}(\cdot;\mu,\mathbf{\Sigma})$ denotes the Gaussian PDF with mean vector μ and covariance matrix $\mathbf{\Sigma}$, and $\hat{\mathbf{x}}_{k|k-1}$ and $\mathbf{P}_{k|k-1}$ are respectively the predicted state vector and corresponding PECM, and $\hat{\mathbf{x}}_{k|k-1}$ and $\mathbf{P}_{k|k-1}$ are given by

$$\hat{\mathbf{x}}_{k|k-1} = \mathbf{F}_{k-1}\hat{\mathbf{x}}_{k-1|k-1} \tag{5}$$

$$\mathbf{P}_{k|k-1} = \mathbf{F}_{k-1}\mathbf{P}_{k-1|k-1}\mathbf{F}_{k-1}^T + \mathbf{Q}_{k-1} \tag{6}$$

where $(\cdot)^T$ denotes the transpose operation, and $\hat{\mathbf{x}}_{k-1|k-1}$ and $\mathbf{P}_{k-1|k-1}$ are respectively the state estimation vector and corresponding estimation error covariance matrix at time $k-1$. Note that $\mathbf{P}_{k|k-1}$ obtained from (6) is inaccurate since the true PNCM $\mathbf{Q}_k$ is unavailable and an inaccurate PNCM is used.

Our aim is to infer $\mathbf{x}_k$ together with $\mathbf{P}_{k|k-1}$ and $\mathbf{R}_k$. To this end, the conjugate prior distributions need to be firstly selected for inaccurate PECM $\mathbf{P}_{k|k-1}$ and MNCM $\mathbf{R}_k$ since the conjugacy can guarantee that the posterior distribution is of the same functional form as the prior distribution. In Bayesian statistics, the inverse Wishart distribution is usually used as the conjugate prior for the covariance matrix of a Gaussian distribution with known mean [15]. The inverse Wishart PDF of a symmetric positive definite random matrix $\mathbf{B}$ of dimension $d\times d$ is formulated as $\mathrm{IW}(\mathbf{B};\lambda,\mathbf{\Psi}) = \frac{|\mathbf{\Psi}|^{\lambda/2}|\mathbf{B}|^{-(\lambda+d+1)/2}\exp\{-0.5\mathrm{tr}(\mathbf{\Psi}\mathbf{B}^{-1})\}}{2^{d\lambda/2}\Gamma_d(\lambda/2)}$, where λ is the degrees of freedom (dof) parameter, and $\mathbf{\Psi}$ is the inverse scale matrix that is a symmetric positive definite matrix of dimension $d\times d$, and $|\cdot|$ and $\mathrm{tr}(\cdot)$ denote the determinant and trace operations respectively, and $\Gamma_d(\cdot)$ is the d-variate gamma function [15]. If $\mathbf{B}\sim\mathrm{IW}(\mathbf{B};\lambda,\mathbf{\Psi})$, then $\mathrm{E}[\mathbf{B}^{-1}] = (\lambda-d-1)\mathbf{\Psi}^{-1}$ when $\lambda > d+1$ [15]. Since both $\mathbf{P}_{k|k-1}$ and $\mathbf{R}_k$ are the covariance matrices of Gaussian PDFs, their prior distributions $p(\mathbf{P}_{k|k-1}|\mathbf{z}_{1:k-1})$ and $p(\mathbf{R}_k|\mathbf{z}_{1:k-1})$ are chosen as inverse Wishart PDFs, i.e.,

$$p(\mathbf{P}_{k|k-1}|\mathbf{z}_{1:k-1}) = \mathrm{IW}(\mathbf{P}_{k|k-1};\hat{t}_{k|k-1},\hat{\mathbf{T}}_{k|k-1}) \tag{7}$$

$$p(\mathbf{R}_k|\mathbf{z}_{1:k-1}) = \mathrm{IW}(\mathbf{R}_k;\hat{u}_{k|k-1},\hat{\mathbf{U}}_{k|k-1}) \tag{8}$$

where $\mathrm{IW}(\cdot;\mu_k,\mathbf{\Sigma}_k)$ denotes the inverse Wishart PDF with dof parameter μ_k and inverse scale matrix $\mathbf{\Sigma}_k$, and $\hat{t}_{k|k-1}$ and $\hat{\mathbf{T}}_{k|k-1}$ are respectively the dof parameter and inverse scale matrix of $p(\mathbf{P}_{k|k-1}|\mathbf{z}_{1:k-1})$, and $\hat{u}_{k|k-1}$ and $\hat{\mathbf{U}}_{k|k-1}$ are respectively the dof parameter and inverse scale matrix of $p(\mathbf{R}_k|\mathbf{z}_{1:k-1})$. Next, the prior parameters $\hat{t}_{k|k-1}$, $\hat{\mathbf{T}}_{k|k-1}$, $\hat{u}_{k|k-1}$ and $\hat{\mathbf{U}}_{k|k-1}$ will be determined.

To capture the prior information of $\mathbf{P}_{k|k-1}$, the mean value of $\mathbf{P}_{k|k-1}$ is set as the nominal PECM $\tilde{\mathbf{P}}_{k|k-1}$, i.e.,

$$\frac{\hat{\mathbf{T}}_{k|k-1}}{\hat{t}_{k|k-1}-n-1} = \tilde{\mathbf{P}}_{k|k-1} = \mathbf{F}_{k-1}\mathbf{P}_{k-1|k-1}\mathbf{F}_{k-1}^T + \tilde{\mathbf{Q}}_{k-1} \tag{9}$$

where $\tilde{\mathbf{Q}}_{k-1}$ denotes the nominal PNCM and is an algorithm parameter of the proposed VBAKF. Let

$$\hat{t}_{k|k-1} = n+\tau+1 \tag{10}$$

where $\tau\geq 0$ is a tuning parameter. Using (10) in (9) yields

$$\hat{\mathbf{T}}_{k|k-1} = \tau\tilde{\mathbf{P}}_{k|k-1} \tag{11}$$

According to the Bayesian theorem, the prior distribution $p(\mathbf{R}_k|\mathbf{z}_{1:k-1})$ is formulated as

$$p(\mathbf{R}_k|\mathbf{z}_{1:k-1}) = \int p(\mathbf{R}_k|\mathbf{R}_{k-1})p(\mathbf{R}_{k-1}|\mathbf{z}_{1:k-1})d\mathbf{R}_{k-1} \tag{12}$$

where $p(\mathbf{R}_{k-1}|\mathbf{z}_{1:k-1})$ is the posterior PDF of MNCM $\mathbf{R}_{k-1}$.

Since the prior distribution $p(\mathbf{R}_{k-1}|\mathbf{z}_{1:k-2})$ of MNCM $\mathbf{R}_{k-1}$ is chosen as an inverse Wishart PDF in accordance with (7), the posterior PDF $p(\mathbf{R}_{k-1}|\mathbf{z}_{1:k-1})$ can be also updated as an inverse Wishart PDF, i.e.,

$$p(\mathbf{R}_{k-1}|\mathbf{z}_{1:k-1}) = \mathrm{IW}(\mathbf{R}_{k-1};\hat{u}_{k-1|k-1},\hat{\mathbf{U}}_{k-1|k-1}) \tag{13}$$

To guarantee $p(\mathbf{R}_k|\mathbf{z}_{1:k-1})$ is an inverse Wishart PDF formulated in (8), the forward predictive model $p(\mathbf{R}_k|\mathbf{R}_{k-1})$ needs to be determined. However, in practical application, the dynamical model $p(\mathbf{R}_k|\mathbf{R}_{k-1})$ is not known in detail. Considering that the MNCM is slowly varying in many practical applications, in this paper, we use similar heuristics as in [10], which just spreads previous approximate posteriors through a factor of ρ, and the prior parameters $\hat{u}_{k|k-1}$ and $\hat{\mathbf{U}}_{k|k-1}$ are given by

$$\hat{u}_{k|k-1} = \rho(\hat{u}_{k-1|k-1}-m-1)+m+1 \tag{14}$$

$$\hat{\mathbf{U}}_{k|k-1} = \rho\hat{\mathbf{U}}_{k-1|k-1} \tag{15}$$

where $\rho\in(0\ 1]$ is a forgetting factor which indicates the extent of the time-fluctuations.

In this paper, the initial MNCM $\mathbf{R}_0$ is also assumed to have an inverse Wishart PDF, i.e., $p(\mathbf{R}_0) = \mathrm{IW}(\mathbf{R}_0;\hat{u}_{0|0},\hat{\mathbf{U}}_{0|0})$. To capture the prior information of the initial MNCM, the mean value of $\mathbf{R}_0$ is set as the initial nominal MNCM $\tilde{\mathbf{R}}_0$, i.e.,

$$\frac{\hat{\mathbf{U}}_{0|0}}{\hat{u}_{0|0}-m-1} = \tilde{\mathbf{R}}_0 \tag{16}$$

where the initial nominal MNCM $\tilde{\mathbf{R}}_0$ is an algorithm parameter of the proposed VBAKF.

C. Variational approximation of posterior PDFs

To estimate $\mathbf{x}_k$ together with $\mathbf{P}_{k|k-1}$ and $\mathbf{R}_k$, the joint posterior PDF $p(\mathbf{x}_k,\mathbf{P}_{k|k-1},\mathbf{R}_k|\mathbf{z}_{1:k})$ needs to be computed. Since there is not an analytical solution for this joint posterior PDF, the VB approach is used to look for a free form factored approximate PDF for $p(\mathbf{x}_k,\mathbf{P}_{k|k-1},\mathbf{R}_k|\mathbf{z}_{1:k})$, i.e., [16], [17]

$$p(\mathbf{x}_k,\mathbf{P}_{k|k-1},\mathbf{R}_k|\mathbf{z}_{1:k}) \approx q(\mathbf{x}_k)q(\mathbf{P}_{k|k-1})q(\mathbf{R}_k) \tag{17}$$

where $q(\cdot)$ represents the approximate posterior PDF of $p(\cdot)$, and $q(\mathbf{x}_k)$, $q(\mathbf{P}_{k|k-1})$ and $q(\mathbf{R}_k)$ are given by minimizing the Kullback-Leibler divergence (KLD) between the factored approximate posterior PDF $q(\mathbf{x}_k)q(\mathbf{P}_{k|k-1})q(\mathbf{R}_k)$ and true joint posterior PDF $p(\mathbf{x}_k,\mathbf{P}_{k|k-1},\mathbf{R}_k|\mathbf{z}_{1:k})$, i.e. [16], [17]

$$\begin{aligned}&\{q(\mathbf{x}_k),q(\mathbf{P}_{k|k-1}),q(\mathbf{R}_k)\} = \arg\min\mathrm{KLD}\\&\big(q(\mathbf{x}_k)q(\mathbf{P}_{k|k-1})q(\mathbf{R}_k)||p(\mathbf{x}_k,\mathbf{P}_{k|k-1},\mathbf{R}_k|\mathbf{z}_{1:k})\big)\end{aligned} \tag{18}$$

where $\mathrm{KLD}(q(x)||p(x)) \triangleq \int q(x)\log\frac{q(x)}{p(x)}dx$ denotes the KLD between $q(x)$ and $p(x)$. The optimal solution for (18) satisfies the following equation [17].

$$\log q(\theta) = \mathrm{E}_{\mathbf{\Xi}^{(-\theta)}}[\log p(\mathbf{\Xi},\mathbf{z}_{1:k})] + c_\theta \tag{19}$$

$$\mathbf{\Xi} \triangleq \{\mathbf{x}_k,\mathbf{P}_{k|k-1},\mathbf{R}_k\} \tag{20}$$

where $\mathrm{E}[\cdot]$ represents the expectation operation, and $\log(\cdot)$ represents the logarithmic function, and θ is an arbitrary element of $\mathbf{\Xi}$, and $\mathbf{\Xi}^{(-\theta)}$ is the set of all elements in $\mathbf{\Xi}$ except for θ, and c_θ denotes the constant with respect to variable θ. Since the variational parameters of $q(\mathbf{x}_k)$, $q(\mathbf{P}_{k|k-1})$ and $q(\mathbf{R}_k)$ are coupled, we need to employ fixed-point iterations to solve (19), where the approximate posterior PDF $q(\theta)$ of the arbitrary element $\boldsymbol{\Xi}$ is updated as $q^{(i+1)}(\theta)$ at the $i+1$th iteration using the approximate posterior PDF $q^{(i)}(\boldsymbol{\Xi}^{(-\theta)})$ [16], [17]. The iterations converge to a local optimum of (19).

Remark 1: In the standard VB approach, the KLD is chosen as a distance measure between the factored approximate posterior PDF and true joint posterior PDF, and the optimal solution is obtained by minimizing the KLD. The VB approach can provide a closed form solution for the approximate posterior PDF and guarantee the local convergence of the fixed-point iterations. The alpha and tau divergences are generalized distance measures [18], [19], and in principle they can be also used as a distance measure between the factored approximate posterior PDF and true joint posterior PDF. However, the alpha or tau divergence based Bayesian inference approach may not provide a closed form solution for the approximate posterior PDF.

Using the conditional independence properties of the Gaussian-inverse-Wishart state-space model in (1)-(2), (3)-(4) and (7)-(8), the joint PDF $p(\boldsymbol{\Xi},\mathbf{z}_{1:k})$ can be factored as

$$p(\boldsymbol{\Xi},\mathbf{z}_{1:k})=p(\mathbf{z}_k|\mathbf{x}_k,\mathbf{R}_k)p(\mathbf{x}_k|\mathbf{z}_{1:k-1},\mathbf{P}_{k|k-1})\times p(\mathbf{P}_{k|k-1}|\mathbf{z}_{1:k-1})p(\mathbf{R}_k|\mathbf{z}_{1:k-1})p(\mathbf{z}_{1:k-1}) \tag{21}$$

Employing (3)-(4) and (7)-(8) in (21) obtains

$$p(\boldsymbol{\Xi},\mathbf{z}_{1:k})=\mathrm{N}(\mathbf{z}_k;\mathbf{H}_k\mathbf{x}_k,\mathbf{R}_k)\mathrm{N}(\mathbf{x}_k;\hat{\mathbf{x}}_{k|k-1},\mathbf{P}_{k|k-1})\times \mathrm{IW}(\mathbf{P}_{k|k-1};\hat{t}_{k|k-1},\hat{\mathbf{T}}_{k|k-1})\mathrm{IW}(\mathbf{R}_k;\hat{u}_{k|k-1},\hat{\mathbf{U}}_{k|k-1})\times p(\mathbf{z}_{1:k-1}) \tag{22}$$

Exploiting (22), $\log p(\boldsymbol{\Xi},\mathbf{z}_{1:k})$ is formulated as

$$\log p(\boldsymbol{\Xi},\mathbf{z}_{1:k})=-0.5(m+\hat{u}_{k|k-1}+2)\log|\mathbf{R}_k|-0.5(\mathbf{z}_k-\mathbf{H}_k\mathbf{x}_k)^T\mathbf{R}_k^{-1}(\mathbf{z}_k-\mathbf{H}_k\mathbf{x}_k)-0.5\mathrm{tr}(\hat{\mathbf{U}}_{k|k-1}\mathbf{R}_k^{-1}) -0.5(n+\hat{t}_{k|k-1}+2)\log|\mathbf{P}_{k|k-1}|-0.5(\mathbf{x}_k-\hat{\mathbf{x}}_{k|k-1})^T\times \mathbf{P}_{k|k-1}^{-1}(\mathbf{x}_k-\hat{\mathbf{x}}_{k|k-1})-0.5\mathrm{tr}(\hat{\mathbf{T}}_{k|k-1}\mathbf{P}_{k|k-1}^{-1})+c_{\boldsymbol{\Xi}} \tag{23}$$

Let $\theta=\mathbf{P}_{k|k-1}$ and using (23) in (19), we have

$$\begin{aligned}\log q^{(i+1)}(\mathbf{P}_{k|k-1})&=-0.5(m+\hat{u}_{k|k-1}+2)\mathrm{E}^{(i)}[\log|\mathbf{R}_k|] -0.5\mathrm{E}^{(i)}\left[(\mathbf{z}_k-\mathbf{H}_k\mathbf{x}_k)^T\mathbf{R}_k^{-1}(\mathbf{z}_k-\mathbf{H}_k\mathbf{x}_k)\right]-\\&0.5\mathrm{E}^{(i)}\left[\mathrm{tr}(\hat{\mathbf{U}}_{k|k-1}\mathbf{R}_k^{-1})\right]-0.5(n+\hat{t}_{k|k-1}+2)\times\\&\log|\mathbf{P}_{k|k-1}|-0.5\mathrm{tr}\left((\mathbf{A}_k^{(i)}+\hat{\mathbf{T}}_{k|k-1})\mathbf{P}_{k|k-1}^{-1}\right)+c_{\boldsymbol{\Xi}}\\&=-0.5(n+\hat{t}_{k|k-1}+2)\log|\mathbf{P}_{k|k-1}|-\\&0.5\mathrm{tr}\left((\mathbf{A}_k^{(i)}+\hat{\mathbf{T}}_{k|k-1})\mathbf{P}_{k|k-1}^{-1}\right)+c_{\mathbf{P}}\end{aligned} \tag{24}$$

where $q^{(i+1)}(\cdot)$ is the approximation of PDF $q(\cdot)$ at the $i+1$th iteration, and $\mathbf{A}_k^{(i)}$ is given by

$$\begin{aligned}\mathbf{A}_k^i=&\mathrm{E}^i[(\mathbf{x}_k-\hat{\mathbf{x}}_{k|k-1})(\mathbf{x}_k-\hat{\mathbf{x}}_{k|k-1})^T]\\=&\mathrm{E}^i[(\mathbf{x}_k-\hat{\mathbf{x}}_{k|k}^i+\hat{\mathbf{x}}_{k|k}^i-\hat{\mathbf{x}}_{k|k-1})\times\\&(\mathbf{x}_k-\hat{\mathbf{x}}_{k|k}^i+\hat{\mathbf{x}}_{k|k}^i-\hat{\mathbf{x}}_{k|k-1})^T]\\=&\mathrm{E}^i[(\mathbf{x}_k-\hat{\mathbf{x}}_{k|k}^i)(\mathbf{x}_k-\hat{\mathbf{x}}_{k|k}^i)^T]+\\&(\hat{\mathbf{x}}_{k|k}^i-\hat{\mathbf{x}}_{k|k-1})(\hat{\mathbf{x}}_{k|k}^i-\hat{\mathbf{x}}_{k|k-1})^T\\=&\mathbf{P}_{k|k}^i+(\hat{\mathbf{x}}_{k|k}^i-\hat{\mathbf{x}}_{k|k-1})(\hat{\mathbf{x}}_{k|k}^i-\hat{\mathbf{x}}_{k|k-1})^T\end{aligned} \tag{25}$$

where $\mathrm{E}^{(i)}[\rho]$ denotes the expectation of variable ρ at the ith iteration.

Exploiting (24), $q^{(i+1)}(\mathbf{P}_{k|k-1})$ can be updated as an inverse Wishart PDF with dof parameter $\hat{t}_k^{(i+1)}$ and inverse scale matrix $\hat{\mathbf{T}}_k^{(i+1)}$, i.e.

$$q^{(i+1)}(\mathbf{P}_{k|k-1})=\mathrm{IW}(\mathbf{P}_{k|k-1};\hat{t}_k^{(i+1)},\hat{\mathbf{T}}_k^{(i+1)}) \tag{26}$$

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

$$\hat{t}_k^{(i+1)}=\hat{t}_{k|k-1}+1 \tag{27}$$

$$\hat{\mathbf{T}}_k^{(i+1)}=\mathbf{A}_k^{(i)}+\hat{\mathbf{T}}_{k|k-1} \tag{28}$$

Let $\theta=\mathbf{R}_k$ and using (23) in (19), we have

$$\begin{aligned}\log q^{(i+1)}(\mathbf{R}_k)&=-0.5(m+\hat{u}_{k|k-1}+2)\log|\mathbf{R}_k|-\\&0.5\mathrm{tr}\left((\mathbf{B}_k^{(i)}+\hat{\mathbf{U}}_{k|k-1})\mathbf{R}_k^{-1}\right)-0.5(n+\hat{t}_{k|k-1}+2)\times\\&\mathrm{E}^{(i)}\left[\log|\mathbf{P}_{k|k-1}|\right]-0.5\mathrm{E}^{(i)}\left[\mathrm{tr}(\hat{\mathbf{T}}_{k|k-1}\mathbf{P}_{k|k-1}^{-1})\right]-\\&0.5\mathrm{E}^{(i)}\left[(\mathbf{x}_k-\hat{\mathbf{x}}_{k|k-1})^T\mathbf{P}_{k|k-1}^{-1}(\mathbf{x}_k-\hat{\mathbf{x}}_{k|k-1})\right]+c_{\boldsymbol{\Xi}}\\&=-0.5(m+\hat{u}_{k|k-1}+2)\log|\mathbf{R}_k|-0.5\mathrm{tr}\left((\mathbf{B}_k^{(i)}+\hat{\mathbf{U}}_{k|k-1}\right.\\&\left.)\,\mathbf{R}_k^{-1}\right)+c_{\mathbf{R}}\end{aligned} \tag{29}$$

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

$$\begin{aligned}\mathbf{B}_k^{(i)}=&\mathrm{E}^i[(\mathbf{z}_k-\mathbf{H}_k\mathbf{x}_k)(\mathbf{z}_k-\mathbf{H}_k\mathbf{x}_k)^T]\\=&\mathrm{E}^i[(\mathbf{z}_k-\mathbf{H}_k\hat{\mathbf{x}}_{k|k}^i+\mathbf{H}_k\hat{\mathbf{x}}_{k|k}^i-\mathbf{H}_k\mathbf{x}_k)\times\\&(\mathbf{z}_k-\mathbf{H}_k\hat{\mathbf{x}}_{k|k}^i+\mathbf{H}_k\hat{\mathbf{x}}_{k|k}^i-\mathbf{H}_k\mathbf{x}_k)^T]\\=&(\mathbf{z}_k-\mathbf{H}_k\hat{\mathbf{x}}_{k|k}^i)(\mathbf{z}_k-\mathbf{H}_k\hat{\mathbf{x}}_{k|k}^i)^T+\\&\mathbf{H}_k\mathrm{E}^i[(\mathbf{x}_k-\hat{\mathbf{x}}_{k|k}^i)(\mathbf{x}_k-\hat{\mathbf{x}}_{k|k}^i)^T]\mathbf{H}_k^T\\=&(\mathbf{z}_k-\mathbf{H}_k\hat{\mathbf{x}}_{k|k}^i)(\mathbf{z}_k-\mathbf{H}_k\hat{\mathbf{x}}_{k|k}^i)^T+\mathbf{H}_k\mathbf{P}_{k|k}^i\mathbf{H}_k^T\end{aligned} \tag{30}$$

Utilizing (29), $q^{(i+1)}(\mathbf{R}_k)$ can be updated as an inverse Wishart PDF with dof parameter $\hat{u}_k^{(i+1)}$ and inverse scale matrix $\hat{\mathbf{U}}_k^{(i+1)}$, i.e.

$$q^{(i+1)}(\mathbf{R}_k)=\mathrm{IW}(\mathbf{R}_k;\hat{u}_k^{(i+1)},\hat{\mathbf{U}}_k^{(i+1)}) \tag{31}$$

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

$$\hat{u}_k^{(i+1)}=\hat{u}_{k|k-1}+1 \tag{32}$$

$$\hat{\mathbf{U}}_k^{(i+1)}=\mathbf{B}_k^{(i)}+\hat{\mathbf{U}}_{k|k-1} \tag{33}$$

Let $\theta=\mathbf{x}_k$ and using (23) in (19) results in

$$\begin{aligned}\log q^{(i+1)}(\mathbf{x}_k)&=-0.5(m+\hat{u}_{k|k-1}+2)\mathrm{E}^{(i+1)}[\log|\mathbf{R}_k|]-\\&0.5\mathrm{E}^{(i+1)}\left[\mathrm{tr}(\hat{\mathbf{U}}_{k|k-1}\mathbf{R}_k^{-1})\right]-0.5(n+\hat{t}_{k|k-1}+2)\times\\&\mathrm{E}^{(i+1)}\left[\log|\mathbf{P}_{k|k-1}|\right]-0.5\mathrm{E}^{(i+1)}\left[\mathrm{tr}(\hat{\mathbf{T}}_{k|k-1}\mathbf{P}_{k|k-1}^{-1})\right]-\\&0.5(\mathbf{z}_k-\mathbf{H}_k\mathbf{x}_k)^T\mathrm{E}^{(i+1)}[\mathbf{R}_k^{-1}](\mathbf{z}_k-\mathbf{H}_k\mathbf{x}_k)-\\&0.5(\mathbf{x}_k-\hat{\mathbf{x}}_{k|k-1})^T\mathrm{E}^{(i+1)}[\mathbf{P}_{k|k-1}^{-1}](\mathbf{x}_k-\hat{\mathbf{x}}_{k|k-1})+c_{\boldsymbol{\Xi}}\\&=-0.5(\mathbf{z}_k-\mathbf{H}_k\mathbf{x}_k)^T\mathrm{E}^{(i+1)}[\mathbf{R}_k^{-1}](\mathbf{z}_k-\mathbf{H}_k\mathbf{x}_k)-\\&0.5(\mathbf{x}_k-\hat{\mathbf{x}}_{k|k-1})^T\mathrm{E}^{(i+1)}[\mathbf{P}_{k|k-1}^{-1}](\mathbf{x}_k-\hat{\mathbf{x}}_{k|k-1})+c_{\mathbf{x}}\end{aligned} \tag{34}$$

where $\mathrm{E}^{(i+1)}[\mathbf{R}_k^{-1}]$ and $\mathrm{E}^{(i+1)}[\mathbf{P}_{k|k-1}^{-1}]$ are given by

$$\mathrm{E}^{(i+1)}[\mathbf{R}_k^{-1}]=(\hat{u}_k^{(i+1)}-m-1)(\hat{\mathbf{U}}_k^{(i+1)})^{-1} \tag{35}$$

$$\mathrm{E}^{(i+1)}[\mathbf{P}_{k|k-1}^{-1}]=(\hat{t}_k^{(i+1)}-n-1)(\hat{\mathbf{T}}_k^{(i+1)})^{-1} \tag{36}$$

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

$$p^{(i+1)}(\mathbf{x}_k|\mathbf{z}_{1:k-1})=\mathrm{N}(\mathbf{x}_k;\hat{\mathbf{x}}_{k|k-1},\hat{\mathbf{P}}_{k|k-1}^{(i+1)}) \tag{37}$$

$$p^{(i+1)}(\mathbf{z}_k|\mathbf{x}_k)=\mathrm{N}(\mathbf{z}_k;\mathbf{H}_k\mathbf{x}_k,\hat{\mathbf{R}}_k^{(i+1)}) \tag{38}$$

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TAC.2017.2730480, IEEE Transactions on Automatic Control

where the modified PECM $\hat{\mathbf{P}}_{k|k-1}^{(i+1)}$ and MNCM $\hat{\mathbf{R}}_{k}^{(i+1)}$ are formulated as

$$\hat{\mathbf{P}}_{k|k-1}^{(i+1)}=\left\{\mathrm{E}^{(i+1)}[\mathbf{P}_{k|k-1}^{-1}]\right\}^{-1}\ \hat{\mathbf{R}}_{k}^{(i+1)}=\left\{\mathrm{E}^{(i+1)}[\mathbf{R}_{k}^{-1}]\right\}^{-1} \tag{39}$$

Employing (37)-(39) in (34) yields

$$q^{(i+1)}(\mathbf{x}_k)=\frac{1}{c_k^{(i+1)}}p^{(i+1)}(\mathbf{z}_k|\mathbf{x}_k)p^{(i+1)}(\mathbf{x}_k|\mathbf{z}_{1:k-1}) \tag{40}$$

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

$$c_k^{(i+1)}=\int p^{(i+1)}(\mathbf{z}_k|\mathbf{x}_k)p^{(i+1)}(\mathbf{x}_k|\mathbf{z}_{1:k-1})d\mathbf{x}_k \tag{41}$$

According to (37)-(41), $q^{(i+1)}(\mathbf{x}_k)$ can be updated as a Gaussian PDF with mean vector $\hat{\mathbf{x}}_{k|k}^{(i+1)}$ and covariance matrix $\mathbf{P}_{k|k}^{(i+1)}$, i.e.,

$$q^{(i+1)}(\mathbf{x}_k)=\mathrm{N}(\mathbf{x}_k;\hat{\mathbf{x}}_{k|k}^{(i+1)},\mathbf{P}_{k|k}^{(i+1)}) \tag{42}$$

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

$$\mathbf{K}_k^{(i+1)}=\hat{\mathbf{P}}_{k|k-1}^{(i+1)}\mathbf{H}_k^T(\mathbf{H}_k\hat{\mathbf{P}}_{k|k-1}^{(i+1)}\mathbf{H}_k^T+\hat{\mathbf{R}}_k^{(i+1)})^{-1} \tag{43}$$

$$\hat{\mathbf{x}}_{k|k}^{(i+1)}=\hat{\mathbf{x}}_{k|k-1}+\mathbf{K}_k^{(i+1)}(\mathbf{z}_k-\mathbf{H}_k\hat{\mathbf{x}}_{k|k-1}) \tag{44}$$

$$\mathbf{P}_{k|k}^{(i+1)}=\hat{\mathbf{P}}_{k|k-1}^{(i+1)}-\mathbf{K}_k^{(i+1)}\mathbf{H}_k\hat{\mathbf{P}}_{k|k-1}^{(i+1)} \tag{45}$$

After fixed point iteration N, the variational approximations of posterior PDFs are given by

$$q(\mathbf{x}_k)\approx q^{(N)}(\mathbf{x}_k)=\mathrm{N}(\mathbf{x}_k;\hat{\mathbf{x}}_{k|k}^{(N)},\mathbf{P}_{k|k}^{(N)})=\mathrm{N}(\mathbf{x}_k;\hat{\mathbf{x}}_{k|k},\mathbf{P}_{k|k}) \tag{46}$$

$$\begin{aligned}q(\mathbf{P}_{k|k-1})\approx q^{(N)}(\mathbf{P}_{k|k-1})&=\mathrm{IW}(\mathbf{P}_{k|k-1};\hat{t}_k^{(N)},\hat{\mathbf{T}}_k^{(N)})\\&=\mathrm{IW}(\mathbf{P}_{k|k-1};\hat{t}_{k|k},\hat{\mathbf{T}}_{k|k})\end{aligned} \tag{47}$$

$$\begin{aligned}q(\mathbf{R}_k)\approx q^{(N)}(\mathbf{R}_k)&=\mathrm{IW}(\mathbf{R}_k;\hat{u}_k^{(N)},\hat{\mathbf{U}}_k^{(N)})\\&=\mathrm{IW}(\mathbf{R}_k;\hat{u}_{k|k},\hat{\mathbf{U}}_{k|k})\end{aligned} \tag{48}$$

The proposed VBAKF operates recursively by combining time update (5), (9)-(11) and (14)-(16) with variational measurement update (25)-(28), (30)-(33), (35)-(36), (39) and (42)-(48), whose implementation pseudocode is shown in Algorithm 1.

Remark 2: In the standard Kalman filter, $\mathbf{P}_{k|k-1}$ is usually used to represent the covariance matrix of the predicted error based on the measurement information $\mathbf{z}_{1:k-1}$. However, in the proposed method, $\mathbf{P}_{k|k-1}$ is estimated using the measurement information $\mathbf{z}_{1:k-1}$ and $\mathbf{z}_k$ based on the VB approach. Thus, the estimation of PECM $\hat{\mathbf{P}}_{k|k-1}$ depends on not only previous measurement information $\mathbf{z}_{1:k-1}$ but also current measurement information $\mathbf{z}_k$.

Algorithm 1: One time step of the proposed VBAKF with inaccurate PNCM and MNCM.

Inputs: $\hat{\mathbf{x}}_{k-1|k-1}$, $\mathbf{P}_{k-1|k-1}$, $\hat{u}_{k-1|k-1}$, $\hat{\mathbf{U}}_{k-1|k-1}$, $\mathbf{F}_{k-1}$, $\mathbf{H}_k$, $\mathbf{z}_k$, $\tilde{\mathbf{Q}}_{k-1}$, m, n, τ, ρ, N

Time update:

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

Variational measurement update:

3. Initialization: $\hat{\mathbf{x}}_{k|k}^{(0)}=\hat{\mathbf{x}}_{k|k-1}$, $\mathbf{P}_{k|k}^{(0)}=\tilde{\mathbf{P}}_{k|k-1}$, $\hat{t}_{k|k-1}=n+\tau+1$, $\hat{\mathbf{T}}_{k|k-1}=\tau\tilde{\mathbf{P}}_{k|k-1}$, $\hat{u}_{k|k-1}=\rho(\hat{u}_{k-1|k-1}-m-1)+m+1$, $\hat{\mathbf{U}}_{k|k-1}=\rho\hat{\mathbf{U}}_{k-1|k-1}$

for $i=0:N-1$

Update $q^{(i+1)}(\mathbf{P}_{k|k-1})=\mathrm{IW}(\mathbf{P}_{k|k-1};\hat{t}_k^{(i+1)},\hat{\mathbf{T}}_k^{(i+1)})$ given $q^{(i)}(\mathbf{x}_k)$:

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

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

6. $\mathbf{B}_k^{(i)}=(\mathbf{z}_k-\mathbf{H}_k\hat{\mathbf{x}}_{k|k}^{(i)})(\mathbf{z}_k-\mathbf{H}_k\hat{\mathbf{x}}_{k|k}^{(i)})^T+\mathbf{H}_k\mathbf{P}_{k|k}^{(i)}\mathbf{H}_k^T$
7. $\hat{u}_k^{(i+1)}=\hat{u}_{k|k-1}+1$, $\hat{\mathbf{U}}_k^{(i+1)}=\mathbf{B}_k^{(i)}+\hat{\mathbf{U}}_{k|k-1}$

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

8. $\mathrm{E}^{(i+1)}[\mathbf{R}_k^{-1}]=(\hat{u}_k^{(i+1)}-m-1)(\hat{\mathbf{U}}_k^{(i+1)})^{-1}$, $\mathrm{E}^{(i+1)}[\mathbf{P}_{k|k-1}^{-1}]=(\hat{t}_k^{(i+1)}-n-1)(\hat{\mathbf{T}}_k^{(i+1)})^{-1}$
9. $\hat{\mathbf{P}}_{k|k-1}^{(i+1)}=\left\{\mathrm{E}^{(i+1)}[\mathbf{P}_{k|k-1}^{-1}]\right\}^{-1}$, $\hat{\mathbf{R}}_k^{(i+1)}=\left\{\mathrm{E}^{(i+1)}[\mathbf{R}_k^{-1}]\right\}^{-1}$
10. $\mathbf{K}_k^{(i+1)}=\hat{\mathbf{P}}_{k|k-1}^{(i+1)}\mathbf{H}_k^T(\mathbf{H}_k\hat{\mathbf{P}}_{k|k-1}^{(i+1)}\mathbf{H}_k^T+\hat{\mathbf{R}}_k^{(i+1)})^{-1}$
11. $\hat{\mathbf{x}}_{k|k}^{(i+1)}=\hat{\mathbf{x}}_{k|k-1}+\mathbf{K}_k^{(i+1)}(\mathbf{z}_k-\mathbf{H}_k\hat{\mathbf{x}}_{k|k-1})$
12. $\mathbf{P}_{k|k}^{(i+1)}=\hat{\mathbf{P}}_{k|k-1}^{(i+1)}-\mathbf{K}_k^{(i+1)}\mathbf{H}_k\hat{\mathbf{P}}_{k|k-1}^{(i+1)}$

end for

13. $\hat{\mathbf{x}}_{k|k}=\hat{\mathbf{x}}_{k|k}^{(N)}$, $\mathbf{P}_{k|k}=\mathbf{P}_{k|k}^{(N)}$, $\hat{t}_{k|k}=\hat{t}_k^{(N)}$, $\hat{\mathbf{T}}_{k|k}=\hat{\mathbf{T}}_k^{(N)}$, $\hat{u}_{k|k}=\hat{u}_k^{(N)}$, $\hat{\mathbf{U}}_{k|k}=\hat{\mathbf{U}}_k^{(N)}$

Outputs: $\hat{\mathbf{x}}_{k|k}$, $\mathbf{P}_{k|k}$, $\hat{t}_{k|k}$, $\hat{\mathbf{T}}_{k|k}$, $\hat{u}_{k|k}$, $\hat{\mathbf{U}}_{k|k}$

D. Parameter selection of the proposed VBAKF

To implement the proposed VBAKF, the tuning parameter τ, the forgetting factor ρ, the nominal PNCM $\tilde{\mathbf{Q}}_k$, and the initial nominal MNCM $\tilde{\mathbf{R}}_0$ need to be selected.

Firstly, we discuss the effect of the tuning parameter τ upon the proposed VBAKF. Substituting (36) in (39), the modified PECM $\hat{\mathbf{P}}_{k|k-1}^{(i+1)}$ can be reformulated as

$$\hat{\mathbf{P}}_{k|k-1}^{(i+1)}=\frac{\hat{\mathbf{T}}_k^{(i+1)}}{\hat{t}_k^{(i+1)}-n-1} \tag{49}$$

Using (10)-(11) and (27)-(28) in (49) yields

$$\hat{\mathbf{P}}_{k|k-1}^{(i+1)}=\frac{\hat{\mathbf{T}}_{k|k-1}+\mathbf{A}_k^{(i)}}{\hat{t}_{k|k-1}-n}=\frac{\tau\tilde{\mathbf{P}}_{k|k-1}+\mathbf{A}_k^{(i)}}{\tau+1} \tag{50}$$

It is seen from (50) that the tuning parameter τ can be deemed as a harmonic weight to balance the efficacy of $\tilde{\mathbf{P}}_{k|k-1}$ and $\mathbf{A}_k^{(i)}$. On the one hand, if τ is too large, the substantial prior uncertainties induced by the inaccurate nominal PNCM are introduced into the measurement update, which degrades the performance of the proposed VBAKF. On the other hand, if τ is too small, a large quantity of information about the process model is lost, which also degrades the performance of the proposed VBAKF. In this paper, the tuning parameter is selected to lie within the range $\tau\in[2,\ 6]$, and the proposed VBAKF with $\tau\in[2,\ 6]$ has essentially consistent estimation performance and higher estimation accuracy than existing filters, as shown in the later simulation.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TAC.2017.2730480, IEEE Transactions on Automatic Control

Secondly, we study the effect of the forgetting factor ρ upon the proposed VBAKF. Substituting (35) in (39), the modified MNCM $\hat{\mathbf{R}}_k^{(i+1)}$ is rewritten as

$$\hat{\mathbf{R}}_k^{(i+1)} = \frac{\hat{\mathbf{U}}_k^{(i+1)}}{\hat{u}_k^{(i+1)} - m - 1} \tag{51}$$

Using (14)-(15) and (32)-(33) in (51) results in

$$\hat{\mathbf{R}}_k^{(i+1)} = \frac{\hat{\mathbf{U}}_{k|k-1} + \mathbf{B}_k^{(i)}}{\hat{u}_{k|k-1} - m} = \frac{\rho\hat{\mathbf{U}}_{k-1|k-1} + \mathbf{B}_k^{(i)}}{\rho(\hat{u}_{k-1|k-1} - m - 1) + 1} \tag{52}$$

According to (48), the estimation of MNCM at time $k-1$ can be formulated

$$\hat{\mathbf{R}}_{k-1} = \frac{\hat{\mathbf{U}}_{k-1|k-1}}{\hat{u}_{k-1|k-1} - m - 1} \tag{53}$$

Substituting (53) in (52), we obtain

$$\hat{\mathbf{R}}_k^{(i+1)} = \frac{\rho(\hat{u}_{k-1|k-1} - m - 1)\hat{\mathbf{R}}_{k-1} + \mathbf{B}_k^{(i)}}{\rho(\hat{u}_{k-1|k-1} - m - 1) + 1} \tag{54}$$

Using (14), (32) and (48) yields

$$\hat{u}_{k|k} = \rho(\hat{u}_{k-1|k-1} - m - 1) + m + 2 \tag{55}$$

Solving equation (55) gives

$$\hat{u}_{k-1|k-1} - m - 1 = \rho^{k-1}(\hat{u}_{0|0} - m - 1) + (1 - \rho^{k-1})/(1 - \rho) \tag{56}$$

Utilizing (56) in (54) results in

$$\hat{\mathbf{R}}_k^{(i+1)} = \frac{w(\rho, k)\hat{\mathbf{R}}_{k-1} + \mathbf{B}_k^{(i)}}{w(\rho, k) + 1} \tag{57}$$

where $w(\rho, k)$ is given by

$$w(\rho, k) = \rho^k(\hat{u}_{0|0} - m - 1) + (\rho - \rho^k)/(1 - \rho) \tag{58}$$

Using (58) gives

$$\lim_{k \to +\infty} w(\rho, k) = \rho/(1 - \rho) \tag{59}$$

It is seen from (57)-(58) that $w(\rho, k)$ can be deemed as a harmonic weight to balance the efficacy of $\hat{\mathbf{R}}_{k-1}$ and $\mathbf{B}_k^{(i)}$. Moreover, we can see from (59) that $w(\rho, k)$ is a monotone increasing function of the forgetting factor ρ when $k \to +\infty$. Thus, the forgetting factor $\rho \in (0\ 1]$ can be used to adjust the efficacy of the previous estimation of MNCM $\hat{\mathbf{R}}_{k-1}$ upon the modified MNCM $\hat{\mathbf{R}}_k^{(i+1)}$. On the one hand, the smaller the forgetting factor ρ, the more the information from the previous estimation $\hat{\mathbf{R}}_{k-1}$ of MNCM is forgotten. On the other hand, the larger the forgetting factor ρ, the more the information from the previous estimation $\hat{\mathbf{R}}_{k-1}$ of MNCM is used. Considering that the MNCM is slowly varying in many practical applications, the forgetting factor is selected to lie within the range $\rho \in [0.9,\ 1]$, and the proposed VBAKF with $\rho \in [0.9,\ 1]$ has essentially consistent estimation performance and higher estimation accuracy than existing filters, as shown in the later simulation. Note that the forgetting factor $\rho = 1$ corresponds to stationary MNCM.

Thirdly, we discuss the effect of the nominal PNCM $\tilde{\mathbf{Q}}_k$ and the initial nominal MNCM $\tilde{\mathbf{R}}_0$ upon the proposed VBAKF. In the fixed-point iterations, the initial values $\hat{\mathbf{P}}_{k|k-1}^{(0)}$ and $\hat{\mathbf{R}}_k^{(0)}$ are set as

$$\begin{cases} \hat{\mathbf{P}}_{k|k-1}^{(0)} = \tilde{\mathbf{P}}_{k|k-1} = \mathbf{F}_{k-1}\mathbf{P}_{k-1|k-1}\mathbf{F}_{k-1}^T + \tilde{\mathbf{Q}}_{k-1} \\ \hat{\mathbf{R}}_k^{(0)} = \hat{\mathbf{R}}_{k-1} \end{cases} \tag{60}$$

Let

$$a_k = \frac{w(\rho, k)}{w(\rho, k) + 1} \qquad \mathbf{C}_k = \frac{\mathbf{B}_k^{(N-1)}}{w(\rho, k) + 1} \tag{61}$$

where $0 < a_k < 1$ and $\mathbf{C}_k \geq \mathbf{0}$. Substituting (61) in (57) results in

$$\hat{\mathbf{R}}_k = a_k\hat{\mathbf{R}}_{k-1} + \mathbf{C}_k \tag{62}$$

With $\hat{\mathbf{R}}_0 = \tilde{\mathbf{R}}_0$ and solving equation (62) obtains

$$\hat{\mathbf{R}}_{k-1} = \left(\prod_{i=1}^{k-1} a_i\right)\tilde{\mathbf{R}}_0 + \sum_{i=1}^{k-1}\left(\prod_{j=i+1}^{k-1} a_j\right)\mathbf{C}_i \tag{63}$$

It is seen from (60) and (63) that $\tilde{\mathbf{Q}}_k$ and $\tilde{\mathbf{R}}_0$ respectively have effects on the initial values $\hat{\mathbf{P}}_{k|k-1}^{(0)}$ and $\hat{\mathbf{R}}_k^{(0)}$. Moreover, we can see from (63) that the effect of $\tilde{\mathbf{R}}_0$ on $\hat{\mathbf{R}}_k^{(0)}$ is gradually reduced as k increases. To guarantee that $\hat{\mathbf{P}}_{k|k-1}^{(i)}$ and $\hat{\mathbf{R}}_k^{(i)}$ converge to true PECM $\mathbf{P}_{k|k-1}$ and MNCM $\mathbf{R}_k$, appropriate initial values $\hat{\mathbf{P}}_{k|k-1}^{(0)}$ and $\hat{\mathbf{R}}_k^{(0)}$ are required since the VB approach can only guarantee local convergence. To this end, the nominal PNCM $\tilde{\mathbf{Q}}_k$ needs to be near the true PNCM $\mathbf{Q}_k$ at each time, and the initial nominal MNCM $\tilde{\mathbf{R}}_0$ needs to be near the initial true MNCM $\mathbf{R}_0$. In this paper, the nominal PNCM and the initial nominal MNCM are respectively set as $\tilde{\mathbf{Q}}_k = \mathrm{diag}[\alpha_{1,k}, \ldots, \alpha_{i,k}, \ldots, \alpha_{n,k}]$ and $\tilde{\mathbf{R}}_0 = \mathrm{diag}[\beta_1, \ldots, \beta_j, \ldots, \beta_m]$, where $\alpha_{i,k} > 0$ and $\beta_j > 0$. The parameters $\alpha_{i,k}$ and β_j are selected based on engineering experience since the diagonal entries of the PNCM and MNCM can be approximately known in many practical applications.

Finally, we study the numerical stability of the proposed VBAKF with the selections of $\tilde{\mathbf{Q}}_k$ and $\tilde{\mathbf{R}}_0$. Using $\tilde{\mathbf{Q}}_k > \mathbf{0}$, $\tilde{\mathbf{R}}_0 > \mathbf{0}$, $0 < a_i < 1$ and $\mathbf{C}_k \geq \mathbf{0}$ in (60) and (63) gives

$$\tilde{\mathbf{P}}_{k|k-1} > \mathbf{0} \qquad \hat{\mathbf{R}}_{k-1} > \mathbf{0} \tag{64}$$

Exploiting (25) and (30) yields

$$\mathbf{A}_k^{(i)} \geq \mathbf{0} \qquad \mathbf{B}_k^{(i)} \geq \mathbf{0} \tag{65}$$

Employing (64)-(65) in (50) and (57) obtains

$$\hat{\mathbf{P}}_{k|k-1}^{(i+1)} > \mathbf{0} \qquad \hat{\mathbf{R}}_k^{(i+1)} > \mathbf{0} \tag{66}$$

It is seen from (66) that the modified PECM $\hat{\mathbf{P}}_{k|k-1}^{(i+1)}$ and MNCM $\hat{\mathbf{R}}_k^{(i+1)}$ are positive definite. Thus, the proposed VBAKF is numerically stable based on the selections of $\tilde{\mathbf{Q}}_k$ and $\tilde{\mathbf{R}}_0$.

Remark 3: The number of iterations N is an important parameter for the proposed filter since it determines the estimation accuracy and implementation time. As the number of iterations increases, the better estimation accuracy is achieved but the more implementation time is required. Generally, the higher dimensions of the state and measurement vectors, an increasing number of iterations is required since with the higher dimensions of the state and measurement vectors, the more inaccurate information involved in the PECM and MNCM needs to be estimated. In practical application, we suggest selecting sufficiently large value for the number of iterations to guarantee that the fixed-point iterations converges to a local optimum.

Remark 4: In this paper, the tuning parameter and the forgetting factor are selected to lie within the ranges $\tau \in [2,\ 6]$ and $\rho \in [0.9,\ 1]$ respectively based on the above discussions. The recommendations regarding parameter ranges are specific to the simulation study presented in the paper, and perhaps other parameter ranges are more appropriate in other situations. Fortunately, our experience has indicated that the proposed filter with the suggested parameter ranges exhibits good estimation performance in many contexts.

III. Simulations

The performance of the proposed VBAKF is illustrated in the problem of target tracking with slowly varying PNCM and MNCM. In this simulation scenario, the target moves according to the continuous white noise acceleration motion model in two dimensional Cartesian coordinates, and the target's positions are collected by a sensor. The state is defined as $\mathbf{x}_k \triangleq [x_k\ y_k\ \dot{x}_k\ \dot{y}_k]$, where x_k, y_k, $\dot{x}_k$ and $\dot{y}_k$ denote the cartesian coordinates and corresponding velocities [13],

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TAC.2017.2730480, IEEE Transactions on Automatic Control

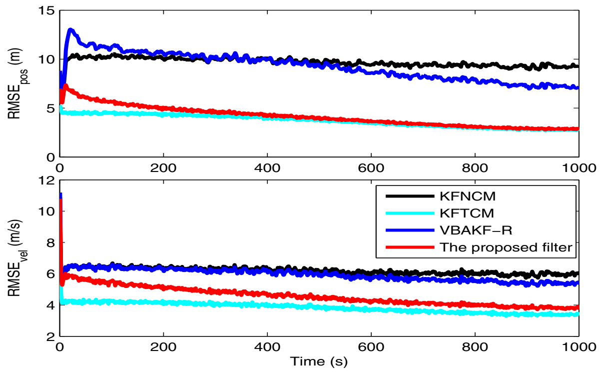

Fig. 1: RMSEs of the position and velocity.

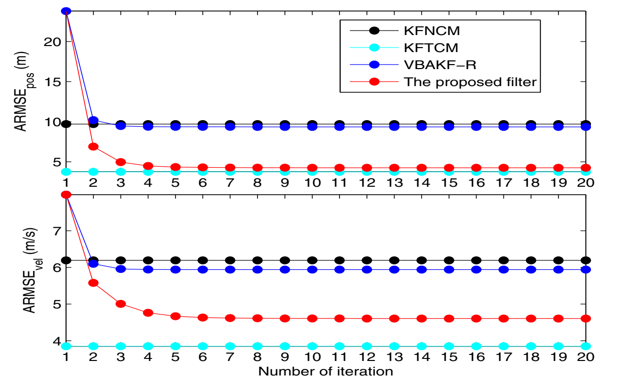

Fig. 3: ARMSEs of the position and velocity when $N = 1, 2, \ldots, 20$.

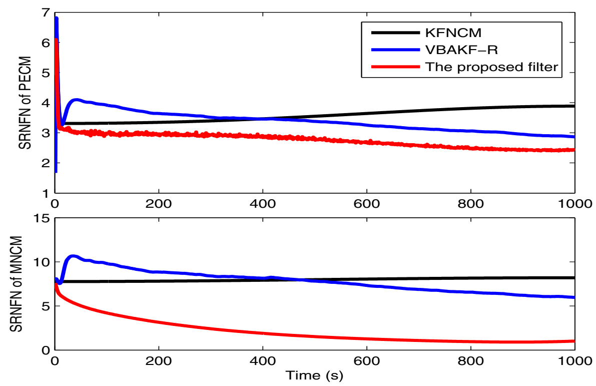

Fig. 2: SRNFNs of the PECM and MNCM.

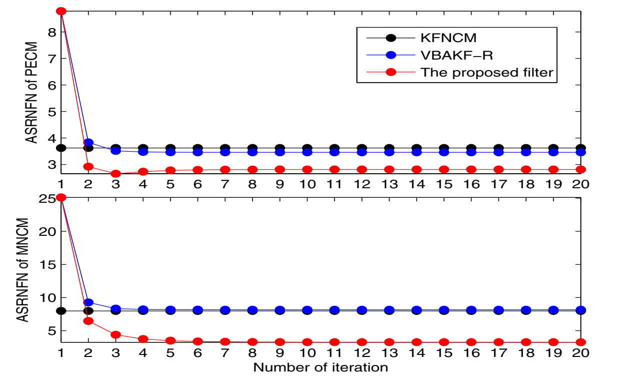

Fig. 4: ASRNFNs of the PECM and MNCM when $N = 1, 2, \ldots, 20$.

[20]. The state transition matrix $\mathbf{F}_{k-1}$ and observation matrix $\mathbf{H}_k$ are respectively given by

$$\mathbf{F}_{k-1} = \begin{bmatrix} \mathbf{I}_2 & \Delta t \mathbf{I}_2 \\ \mathbf{0} & \mathbf{I}_2 \end{bmatrix} \quad \mathbf{H}_k = \begin{bmatrix} \mathbf{I}_2 & \mathbf{0} \end{bmatrix} \tag{67}$$

where the parameter $\Delta t = 1$s is the sampling interval and $\mathbf{I}_2$ is the two-dimensional identity matrix. Similar to [13], the true PNCM and MNCM are given by

$$\begin{cases} \mathbf{Q}_k = \left[6.5 + 0.5\cos(\frac{\pi k}{T})\right] q \begin{bmatrix} \frac{\Delta t^3}{3}\mathbf{I}_2 & \frac{\Delta t^2}{2}\mathbf{I}_2 \\ \frac{\Delta t^2}{2}\mathbf{I}_2 & \Delta t \mathbf{I}_2 \end{bmatrix} \\ \mathbf{R}_k = \left[0.1 + 0.05\cos(\frac{\pi k}{T})\right] r \begin{bmatrix} 1 & 0.5 \\ 0.5 & 1 \end{bmatrix} \end{cases} \tag{68}$$

where $T = 1000$s denotes the simulation time, and $q = 1\mathrm{m}^2/\mathrm{s}^3$ and $r = 100\mathrm{m}^2$.

In this simulation, the nominal PNCM and MNCM are respectively selected as $\tilde{\mathbf{Q}}_k = \alpha \mathbf{I}_4$ and $\tilde{\mathbf{R}}_0 = \beta \mathbf{I}_2$, where $\mathbf{I}_4$ is the four-dimensional identity matrix. The Kalman filter with nominal covariance matrices $\tilde{\mathbf{Q}}_k$ and $\tilde{\mathbf{R}}_0$ (KFNCM), the Kalman filter with true covariance matrices $\mathbf{Q}_k$ and $\mathbf{R}_k$ (KFTCM), the existing IAKF [2], the existing SHAKF [6], the existing VBAKF for estimating only $\mathbf{R}_k$ (VBAKF-R) [10], [11], and the proposed VBAKF for estimating PECM and MNCM are tested. Note that the IAKF [2] and SHAKF [6] were often found filtering divergence, thus their simulation results are not shown in the following simulation. In the proposed VBAKF and existing KFNCM and VBAKF-R, the algorithm parameters are set as: parameter $\alpha = 1$, parameter $\beta = 100$, tuning parameter $\tau = 3$, forgetting factor $\rho = 1 - \exp(-4)$ and the number of iterations $N = 10$. All algorithms are coded with MATLAB and the simulations are run on a computer with Intel Core i7-3770 CPU at 3.40 GHz.

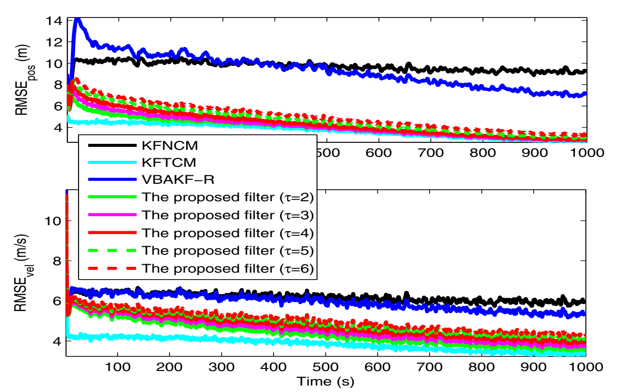

Fig. 5: RMSEs of the position and velocity when $\tau = 2, 3, 4, 5, 6$.

To evaluate the estimate accuracy of state, the RMSEs and the averaged RMSEs (ARMSEs) of position and velocity are chosen as performance metrics, which are defined as follows

$$\begin{cases} \mathrm{RMSE}_{\mathrm{pos}} \triangleq \sqrt{\frac{1}{M}\sum_{s=1}^{M}\left((x_k^s - \hat{x}_k^s)^2 + (y_k^s - \hat{y}_k^s)^2\right)} \\ \mathrm{ARMSE}_{\mathrm{pos}} \triangleq \sqrt{\frac{1}{MT}\sum_{k=1}^{T}\sum_{s=1}^{M}\left((x_k^s - \hat{x}_k^s)^2 + (y_k^s - \hat{y}_k^s)^2\right)} \end{cases} \tag{69}$$

where (x_k^s, y_k^s) and $(\hat{x}_k^s, \hat{y}_k^s)$ are the true and estimated positions at the s-th Monte Carlo run, and $M = 1000$ represents the total number of Monte Carlo runs. Similar to the RMSE and ARMSE in position,

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TAC.2017.2730480, IEEE Transactions on Automatic Control

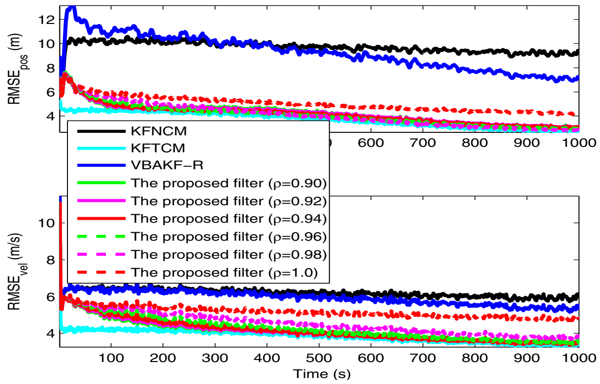

Fig. 6: RMSEs of the position and velocity when $\rho = 0.9, 0.92, 0.94, 0.96, 0.98, 1.0$.

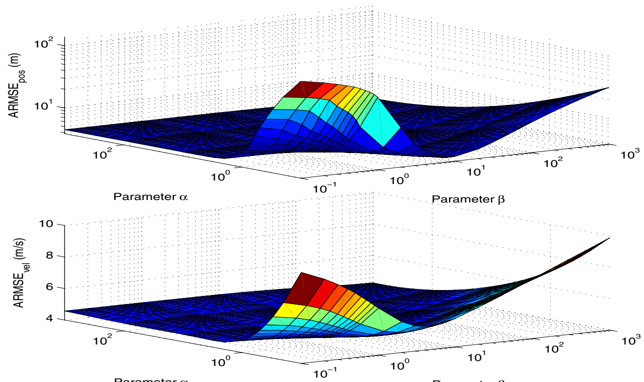

Fig. 7: ARMSEs of the position and velocity when parameters $(\alpha, \beta) \in [0.1,\ 1000] \times [0.1,\ 1000]$.

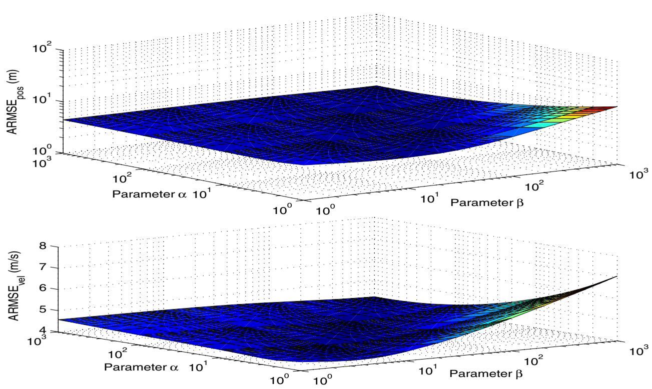

Fig. 8: ARMSEs of the position and velocity when parameters $(\alpha, \beta) \in [1,\ 1000] \times [1,\ 1000]$.

we can also write formula for the RMSE and ARMSE in velocity.

To evaluate the estimate accuracy of PECM and MNCM, the square root of normalized Frobenius norm (SRNFN) and averaged SRNFN (ASRNFN) are selected as error measures, which are defined as follows [13]

$$\begin{cases} \mathrm{SRNFN}_P \triangleq \left(\frac{1}{n^2 M} \sum\limits_{s=1}^{M} \|\hat{\mathbf{P}}_{k|k-1}^{s} - \mathbf{P}_{\mathrm{o},k|k-1}^{s}\|^2 \right)^{\frac{1}{4}} \\ \mathrm{ASRNFN}_P \triangleq \left(\frac{1}{n^2 MT} \sum\limits_{k=1}^{T} \sum\limits_{s=1}^{M} \|\hat{\mathbf{P}}_{k|k-1}^{s} - \mathbf{P}_{\mathrm{o},k|k-1}^{s}\|^2 \right)^{\frac{1}{4}} \end{cases} \tag{70}$$

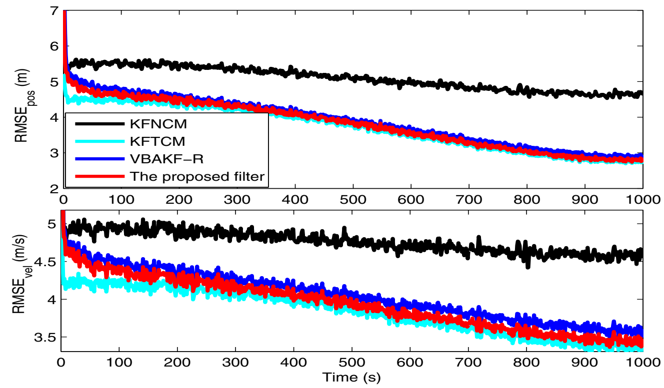

Fig. 9: RMSEs of the position and velocity when the nominal PNCM $\tilde{\mathbf{Q}}_k$ and the true PNCM $\mathbf{Q}_k$ are identical.

TABLE I: Steady-state ARMSEs over the last 100s from the existing filters and the proposed filter.

Filters	ARMSE$_{\mathrm{pos}}$ (m)	ARMSE$_{\mathrm{vel}}$ (m/s)
KFNCM	4.63	4.58
KFTCM	2.77	3.38
VBAKF-R	2.92	3.59
The proposed filter	2.81	3.45

where $\|\mathbf{D}\|^2 = \mathrm{tr}(\mathbf{D}\mathbf{D}^T)$, and $\hat{\mathbf{P}}_{k|k-1}^{s}$ denotes the estimated PECM at the s-th Monte Carlo run, and $\mathbf{P}_{\mathrm{o},k|k-1}^{s}$ represents the accurate PECM at the s-th Monte Carlo run provided by the KFTCM. Similar to the SRNFN and ASRNFN in PECM, we can also write formula for the SRNFN and ASRNFN in MNCM.

The RMSEs of position and velocity and the SRNFNs of PECM and MNCM from existing filters and the proposed filter are respectively shown in Fig. 1–Fig. 2. It is seen from Fig. 1 that the proposed filter has smaller RMSEs than existing KFNCM and VBAKF-R, and the RMSEs from the proposed filter are close to the RMSEs from KFTCM when $k > 600s$. The ARMSEs of position and velocity from the proposed filter are respectively reduced by 54.5% and 22.4% as compared with the existing VBAKF-R. We can see from Fig. 2 that the proposed filter has smaller SRNFNs than existing KFNCM and VBAKF-R. The ASRNFNs of PECM and MNCM from the proposed filter are respectively reduced by 18.7% and 60% as compared with existing VBAKF-R. Moreover, the implementation times of existing KFNCM, VBAKF-R and the proposed filter in a single step run are respectively 2.5×10^{-5}s, 3.8×10^{-4}s and 5.6×10^{-4}s. Thus, the proposed filter has better estimation accuracy but higher computational complexity than existing state-of-the-art filters.

Fig. 3–Fig. 4 show respectively the ARMSEs of position and velocity and the ASRNFNs of PECM and MNCM from the existing filters and the proposed filters when $N = 1, 2, \ldots, 20$. It can be seen from Fig. 3–Fig. 4 that the proposed filter has smaller ARMSEs and ASRNFNs than existing filters when $N \geq 2$, and the proposed filter converges when $N \geq 6$. Thus, the proposed filter exhibits satisfactory convergence speed with respect to the number of iterations.

Fig. 5 shows the RMSEs of position and velocity from the existing filters and the proposed filters when $\tau = 2, 3, 4, 5, 6$. We can see from Fig. 5 that the proposed filter with the tuning parameter $\tau = 2, 3, 4, 5, 6$ has essentially consistent estimation performance and higher estimation accuracy than existing filters.

Fig. 6 shows the RMSEs of position and velocity from the existing filters and the proposed filters when $\rho = 0.9, 0.92, 0.94, 0.96, 0.98, 1.0$. It can be seen from Fig. 6 that the proposed filter with $\rho = 0.9, 0.92, 0.94, 0.96, 0.98, 1.0$ has better estimation accuracy than existing filters, and the proposed filter with $\rho = 0.9, 0.92, 0.94, 0.96, 0.98$ has essentially consistent estimation performance. Moreover, the proposed filter with $\rho = 1.0$ has worse estimation accuracy than the proposed filter with $\rho = 0.9, 0.92, 0.94, 0.96, 0.98$, which is because $\rho = 1.0$ corresponds to stationary MNCM so that the estimation performance degrades when the MNCM is slowly varying.

Fig. 7–Fig. 8 show the ARMSEs of position and velocity from the proposed filter when parameters $(\alpha, \beta) \in [0.1,\ 1000] \times [0.1,\ 1000]$ and $(\alpha, \beta) \in [1,\ 1000] \times [1,\ 1000]$ respectively. It is seen from Fig. 7–Fig. 8 that the proposed filter exhibits good estimation performance only when parameters $(\alpha, \beta) \in [1,\ 1000] \times [1,\ 1000]$. Thus, the proposed filter may fail when the nominal PNCM and MNCM are too far away from the true PNCM and MNCM, which is induced by the fact that the VB approach can only guarantee local convergence so that the use of improper nominal PNCM and MNCM may result in error estimations even divergence.

Fig. 9 and Table I show respectively the RMSEs and steady-state ARMSEs over the last 100s of position and velocity from the existing filters and the proposed filter when the nominal PNCM $\tilde{\mathbf{Q}}_k$ and the true PNCM $\mathbf{Q}_k$ are identical. It is seen from Fig. 9 and Table I that the proposed filter has significantly smaller RMSEs and steady-state ARMSEs than the existing KFNCM and slightly smaller RMSEs and steady-state ARMSEs than the existing VBAKF-R, and the steady-state ARMSEs from the proposed filter are nearly identical to the steady-state ARMSEs from the KFTCM, which also indicates good performance of the proposed filter.

IV. Conclusions

In this paper, the authors focused on solving the filtering problem of linear Gaussian state-space models with inaccurate PNCM and MNCM. A novel VBAKF with inaccurate PNCM and MNCM was proposed, where the state together with PECM and MNCM were inferred by choosing inverse Wishart priors. Simulation results illustrated that the proposed VBAKF has better robustness to resist the uncertainties of PNCM and MNCM as compared with existing filters, which is induced by the fact that the proposed filter can iteratively find better estimates of PECM and MNCM.

Acknowledgment

The authors would like to thank Xiaodong Wang, at the electrical engineering department of Columbia University for providing comments and suggestions on the revised version of the manuscript.

References

[1] R. Mehra, "Approaches to adaptive filtering," *IEEE Trans. Automat. Control*, vol. 17, no. 5, pp. 693–698, Oct. 1972.
[2] A. H. Mohamed and K. P. Schwarz, "Adaptive Kalman filtering for INS/GPS," *J. Geod.*, vol. 73, no. 4, pp. 193–203, May 1999.
[3] C. Hide, T. Moore, and M. Smith, "Adaptive Kalman filtering algorithms for integrating GPS and low cost INS," in *Proc. Position Location Navigation Symp.*, Apr. 2004, pp. 227–233.
[4] M-J. Yu, "INS/GPS integration system using adaptive filter for estimating measurement noise variance," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 48, no. 2, pp. 1786–1792, Apr. 2012.
[5] A.P. Sage and G. W. Husa, "Adaptive filtering with unknown prior statistics," in *Proc. of joint automatic control conference*, 1969, Boulder, pp. 760–769.
[6] X. Gao, D. You and S. Katayama, "Seam tracking monitoring based on adaptive Kalman filter embedded elman neural network during high-power fiber laser welding," *IEEE Trans. Ind. Electron*, vol. 59, no. 11, pp. 4315–4325, Nov. 2012.
[7] M. Karasalo and X. M. Hu, "An optimization approach to adaptive Kalman filtering," *Automatica*, vol. 47, no. 8, pp. 1785–1793, Aug. 2011.
[8] X. R. Li and Y. Bar-Shalom, "A recursive multiple model approach to noise identification," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 30, no. 3, pp. 671–684, July 1994.
[9] S. Särkkä and A. Nummenmaa, "Recursive noise adaptive Kalman filtering by variational Bayesian approximations," *IEEE Trans. Automat. Control*, vol. 54, no. 3, pp. 596–600, Mar. 2009.
[10] S. Särkkä, and J. Hartikainen, "Variational Bayesian adaptation of noise covariance in nonlinear Kalman filtering," *arXiv: 1302.0681v1 [Stat.ME]*, Feb. 2013.
[11] G. Agamennoni, J.I. Nieto, and E.M. Nebot, "Approximate inference in state-space models with heavy-tailed noise," *IEEE Trans. Signal Process.*, vol. 60, no. 10, pp. 5024–5037, Oct. 2012.
[12] Y. L. Huang, Y. G. Zhang, N. Li and J. Chambers, "A robust Gaussian approximate filter for nonlinear systems with heavy-tailed measurement noises," in *2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Mar. 2016, pp. 4209–4213.
[13] T. Ardeshiri, E. Ozkan, U. Orguner, F. Gustafsson, "Approximate Bayesian smoothing with unknown process and measurement noise covariances," *IEEE Signal Process. Lett.*, vol. 22, no. 12, pp. 2450–2454, Dec. 2015.
[14] Y. L. Huang, Y. G. Zhang, N. Li and J. Chambers, "A robust Gaussian approximate fixed-interval smoother for nonlinear systems with heavy-tailed process and measurement noises," *IEEE Signal Process. Lett.*, vol. 23, no. 4, pp. 468–472, Apr. 2016.
[15] A. O'Hagan, and J. J. Forster, *Kendall's Advanced Theory of Statistics: Bayesian Inference*. Arnold, 2004.
[16] C. M. Bishop, *Pattern Recognition and Machine Learning.* Springer, 2007.
[17] D. Tzikas, A. Likas, and N. Galatsanos, "The variational approximation for Bayesian inference," *IEEE Signal Process. Mag.*, vol. 25, no. 6, pp. 131–146, Nov. 2008.
[18] M. Zorzi, "Rational approximations of spectral densities based on the Alpha divergence," *Math. Control Signals Syst.*, vol. 26, no. 2, pp. 259–278, June 2014.
[19] M. Zorzi,"Multivariate spectral estimation based on the concept of optimal prediction," *IEEE Trans. Automat. Control*, vol. 60, no. 6, pp. 1647–1652, June 2015.
[20] M. Roth, E. Özkan, and F. Gustafsson, "A Student's filter for heavy-tailed process and measurement noise," in *2013 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, May 2013, pp. 5770–5774.