Adaptive Kalman Filter for Linear Systems with Additive and Multiplicative Noises

Xingkai Yu and Jianxun Li

Abstract—In this paper, a novel adaptive Kalman filter is proposed to solve the adaptive Kalman filtering problem for linear systems with dynamic multiplicative and additive noises. The noise covariances are considered unknown and a variational Bayesian (VB) approach is proposed. We use inverse Wishart priors for noise covariances and Student's t-distribution to represent the likelihood functions. Estimation algorithms for recursive noise covariance matrices and the dynamic state are proposed following VB process. A target tracking example is provided to validate the effectiveness of the proposed algorithm.

Index Terms—Multiplicative noise, additive noise, adaptive Kalman filter, Variational Bayesian.

I. INTRODUCTION

The Kalman filter is an optimal state estimator in the sense of minimum variance for linear systems with Gaussian measurement and process noises [1]. It is widely used in communication, signal processing and aerospace engineering due to its real-time, optimal, and stable characteristics. The usefulness of the Kalman filter is limited by *priori* information: the noise statistics and system model parameters. Also, the statistics property of the noises are usually assumed to be known and time-invariant. The unknown *priori* information may lead to lose efficiency. On the other hand, in many practical applications, e.g., navigation and positioning systems [2]–[5], the prior information on noise statistics and model parameters may be unknown, part-known and even time-varying and missing. The classical methods to solve such a problem are so-called the adaptive Kalman filter, and can be further classified by four branches including maximum likelihood, Bayesian, covariance matching and correlation approaches [2]. The typical implements of adaptive Kalman filters include *innovation-based adaptive filter* [6], *interacting multiple model filter* [6], [7] and the *Sage-Husa adaptive filter* [8]. In particular, *innovation-based adaptive filter* uses the innovation sequence and a maximum likelihood criterion to estimate the noise covariance matrices. *Interacting multiple model filter* is a Bayesian approach where other methods can be seen as its approximations. *Sage-Husa adaptive filter* uses the maximum-a-posterior criterion to estimate the noise statistics recursively and it is a covariance matching method.

This work was jointly supported by National Natural Science Foundation under Grant 61673265; Special research projects for civil aircraft under Grant MJ-2017-S-38; Shanghai Commercial Aircraft System Engineering Joint Research Fund; CEMEE under Grant 2019K0302A.

X. Yu is with the Department of Precision Instrument, Tsinghua University, Beijing 100084, China, e-mail: xkyu2007@mail.tsinghua.edu.cn.

J. Li is with the Department of Automation, Shanghai Jiao Tong University, Shanghai 200240, China, e-mail: lijx@sjtu.edu.cn.

This class of adaptive Kalman filter may suffer from issues of the convergence, practical applications, and computation burden [9], [10].

More recently, variational Bayesian (VB) approach has received much attention since it can be used to perform approximate posterior inference and to estimate uncertain hidden parameters or state variables [11], [12] in broad application including machine learning, visual tracking, signal processing and etc [13]–[16]. In fact, the existing VB approaches are also seen as an approximation of the Bayesian inference, which are proposed to estimate the additive measurement noise covariance matrix (AMNCM) by using an appropriate conjugate prior probability density function (PDF) or distribution [9], [17], [18]. Due to its strong adaptability, a large number of research results are obtained. In particular, a theoretical framework for VB-based filter is presented in [11], [19] and a variant of filters are proposed including VB-based extended Kalman filter [20], unscented Kalman filter [21], cubature information filter [22] and particle filter [23], [24]. Existing filters usually model additive Gaussian process and measurement noises with Student's t-distribution in a heavy-tailed form and model AMNCM or additive process noise covariance matrix (APNCM) with Wishart distribution, inverse Wishart distribution or inverse Gamma distribution. However, to the best of our knowledge, the study on the *multiplicative* noise problem under a VB approach is still open and how to design an adaptive Kalman filter is a challenge. Motivated by the fact that multiplicative noise can be used to describe a wider class of physical systems, we will address the estimation problem for systems with unknown multiplicative measurement noise covariance matrix (MMNCM) using a VB approach.

Compared with additive noise, the mean of multiplicative noise is always non-zero and this results in additional difficulties for state signals [25], [26], [26]–[29]. In engineering practice, the mean and covariance of the multiplicative noise are usually difficult to be obtained. In this paper, we present a recursive adaptive Kalman filter to estimate the unknown noise covariances of multiplicative and additive noises (including MMNCM, AMNCM and APNCM). First, we use inverse Wishart priors for process noise covariance and observation noise covariance, and the one-step predicted error covariance and choose Gaussian priors or non-Gaussian Student's t priors for likelihood functions. Then, the dynamic states together with noise covariances are introduced in a recursive way based on a VB-based approach. For comparison, the proposed filter and existing state-of-the-art filters are used to a target tracking example with unknown noise covariances. Simulation results

demonstrate the superiority of the proposed filter than these existing filters.

The main contributions of this paper are two-fold. First, compared with the filter for linear systems with additive noises, we propose an algorithm that is applicable to address both additive and multiplicative noises. To solve the technical issue raised by the multiplicative noise, a variational Bayesian approach is proposed. Moreover, the proposed approach is able to approach the multiplicative and additive measurement noise covariances as a whole, while the algorithms proposed in [25], [27], [28] often operate in a separate way. Second, in contrast with [25], [27], [28], where the covariance of the multiplicative noise is assumed to be fixed and known, we focus on the situation that the covariances of both additive and multiplicative noises are time-varying and unknown. Towards this end, a novel adaptive Kalman filter is presented to jointly approach the covariances of multiplicative and additive noises based on projection formula and a VB-based approach.

The reminder parts of this paper are organized as follows. Section 1 introduces the background. Sections II and III state the considered problem and the proposed algorithms respectively. Then, Section IV evaluates the performance analysis and Section V gives a numerical simulation example. Finally, Section VI concludes the paper.

Notation. $E[\cdot]$ and $\log[\cdot]$ denote the expectation operation and the logarithmic function, respectively. $\text{tr}(\cdot)$ denotes trace operation. $(p||q)$ denotes the distance between p and q . $\|\cdot\|$ denotes 2-Euclid-norm of a matrix or vector. $|\cdot|$ denotes determinant of a matrix or vector.

II. PROBLEM FORMULATION

We consider the following discrete-time linear state-space model

$$\mathbf{X}_k = \mathbf{F}_{k-1}\mathbf{X}_{k-1} + \mathbf{W}_{k-1} \quad (1)$$

$$\mathbf{Y}_k = m_k \mathbf{H}_k \mathbf{X}_k + \mathbf{V}_k, \quad (2)$$

where k is the time index, $\mathbf{X}_k \in \mathbb{R}^n$ is the state vector, $\mathbf{Y}_k \in \mathbb{R}^m$ is the measured vector, $\mathbf{W}_k \in \mathbb{R}^n$ and $\mathbf{V}_k \in \mathbb{R}^m$ are mutually uncorrelated zero-mean Gaussian additive noises with covariance matrices $\mathbf{Q}_k$ and $\mathbf{R}_k$, $\mathbf{F}_k \in \mathbb{R}^{n \times n}$ and $\mathbf{H}_k \in \mathbb{R}^{m \times n}$ are the state transition matrix and the observation equation matrix, respectively, and $m_k \in \mathbb{R}$ is the Gaussian multiplicative noise with mean $\bar{m}_k$ and covariance σ_k . Assume the initial state $\mathbf{X}_0$ be Gaussian with mean $\hat{\mathbf{X}}_{0|0}$ and covariance $\mathbf{P}_{0|0}$. Moreover, $\bar{m}_k$, $\hat{\mathbf{X}}_{0|0}$ and $\mathbf{P}_{0|0}$ are given.

The standard Kalman filter employs measurements $\mathbf{z}_{1:k}$ to estimate the state in measurement update step. The Kalman filter is optimal for systems without multiplicative noise, while Kalman filter for model (1)-(2) is optimal in terms of linear minimum variance with given $\mathbf{Q}_k$, $\mathbf{R}_k$ and σ_k . However, in general situation, $\mathbf{Q}_k$, $\mathbf{R}_k$ and σ_k are unknown or inaccurate. Hence a novel VB-based filter is proposed in this paper that is suitable for unknown covariances of both multiplicative noise and additive noise.

Related work. In [30], [31], multiplicative noise m_k is considered as a Gaussian process with known mean and covariance matrices. In [27], [28], [32], m_k is also considered

as Gaussian process but time-correlated. In our previous work [25], [33], we use the Monte Carlo sampling and variational approach to solve the issue of unknown and non-Gaussian multiplicative noise. However, the proposed method is approximation one with enough number of particles and the guarantee of sampling accuracy. The existence of multiplicative noise will lead to the complexity of sampling, and then lead to inaccurate estimation. Therefore, this paper will adopt a more accurate iterative method. For simplicity, we take Gaussian distribution example and the cases for other fixed distributed can be similarly solved.

Before proceeding on, we first introduce several definitions.

Definition 1: According to Mellin-Barnes integral representation in the complex plane, Meijer G-function is defined as:

$$G_{p,q}^{m,n} \left(\begin{matrix} a_1, \dots, a_p \\ b_1, \dots, b_q \end{matrix}; z \right) = \frac{1}{2\pi i} \int_L \frac{\prod_{l=1}^m \Gamma(b_l - s) \prod_{l=1}^n \Gamma(1 - a_l + s)}{\prod_{l=n+1}^p \Gamma(a_{l+1} - s)} z^s ds,$$

where a_i and b_i are real or complex parameters; m ($0 \leq m \leq q$) and n ($0 \leq n \leq p$) are integers; L is the integration path; i denotes the complex plane; Γ denotes the gamma function and is defined as:

$$\Gamma(z) = \int_0^\infty x^{z-1} e^{-x} dx.$$

Definition 2: The Student's t-distribution of variable t is defined as:

$$p(t; \nu) = \frac{\Gamma(\frac{\nu+1}{2})}{\sqrt{\nu\pi}\Gamma(\frac{\nu}{2})} \left(1 + \frac{t^2}{\nu}\right)^{-\frac{\nu+1}{2}},$$

where Γ is the gamma function and ν is the number of degrees of freedom (dof).

Definition 3: The Cauchy distribution of variable x is defined as:

$$p(x; x_0, \gamma) = \frac{1}{\pi\gamma \left[1 + \left(\frac{x-x_0}{\gamma}\right)^2\right]} = \frac{1}{\pi\gamma} \left[\frac{\gamma^2}{(x-x_0)^2 + \gamma^2} \right],$$

where x_0 is the location parameter and γ is the scale parameter.

Definition 4: For a $d \times d$ symmetric positive definite random matrix $\mathbf{B}$, the inverse Wishart distribution is defined as:

$$\text{IW}(\mathbf{B}; \lambda, \Psi) = \frac{|\Psi|^{\lambda/2} |\mathbf{B}|^{-(\lambda+d+1)/2} \exp\{-0.5 \text{tr}(\Psi \mathbf{B}^{-1})\}}{2^{d\lambda/2} \Gamma_d(\lambda/2)},$$

where λ is the dof parameter, Ψ is a $d \times d$ dimensional symmetric positive definite inverse scale matrix, and $\Gamma_d(\cdot)$ is the d -variate gamma function [34].

Definition 5: For discrete probability distributions p and q defined on the same probability space $\mathcal{X}$, the Kullback-Leibler divergence (KLD) from q to p is defined to be:

$$\text{KLD}(p||q) = \sum_{x \in \mathcal{X}} p(x) \log \left(\frac{p(x)}{q(x)} \right).$$

Consider the linear system (1) and (2). The objective of this paper is to calculate state $\mathbf{x}_k$ together with $\mathbf{Q}_k$, $\mathbf{R}_k$, and σ_k using an iterative approach. In other words, the objective is to compute the joint posterior distribution $p(\mathbf{x}_k, \mathbf{Q}_k, \sigma_k, \mathbf{R}_k | \mathbf{y}_{1:k})$.

III. PROPOSED ALGORITHMS

This section will propose the filters from the two aspects: one-step prediction and measurement update. The measurement update step contain choices of prior distributions, choice of the likelihood function, choices of covariance matrices and variational approximations of posterior PDFs.

A. One-step prediction

Since the multiplicative noise doesn't affect the state equation (1), so the one-step prediction $p(\mathbf{x}_k|\mathbf{z}_{1:k-1})$ is Gaussian:

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

where $\hat{\mathbf{x}}_{k|k-1}$ and $\mathbf{P}_{k|k-1}$ are respectively the one-step predicted state vector and corresponding one-step predicted error covariance matrix and $\mathbf{N}(\cdot; \mu, \Sigma)$ denotes the Gaussian distribution with mean vector μ and covariance matrix Σ . Then, it follows that

$$\hat{\mathbf{x}}_{k|k-1} = \mathbf{F}_{k-1} \hat{\mathbf{x}}_{k-1|k-1}, \quad (4)$$

$$\mathbf{P}_{k|k-1} = \mathbf{F}_{k-1} \mathbf{P}_{k-1|k-1} \mathbf{F}_{k-1}^T + \mathbf{Q}_{k-1}, \quad (5)$$

$\hat{\mathbf{x}}_{k-1|k-1}$ and $\mathbf{P}_{k-1|k-1}$ denote the $k-1$ filtering state and corresponding estimation error covariance matrix, respectively. Since $\mathbf{Q}_{k-1}$ is unknown and is a linear correlation with $\mathbf{P}_{k|k-1}$. So we choose to estimate $\mathbf{P}_{k|k-1}$ instead of $\mathbf{Q}_{k-1}$.

Remark 1: From (5), $\mathbf{P}_{k|k-1}$ is a linear function of $\mathbf{Q}_{k-1}$ and therefore estimation of $\mathbf{Q}_{k-1}$ can be replaced by the estimation of $\mathbf{P}_{k|k-1}$.

Remark 2: The one-step prediction is assumed to be Gaussian no matter with or without multiplicative noise [28]. For other distributions, for example, in the view of Student's t-likelihood function described in the following, (3) can be updated by

$$p(\mathbf{x}_k|\mathbf{z}_{1:k-1}, \mathbf{P}_{k|k-1}) = \text{St}(\mathbf{x}_k; \hat{\mathbf{x}}_{k|k-1}, \mathbf{P}_{k|k-1}).$$

$\hat{\mathbf{x}}_{k|k-1}$ and $\mathbf{P}_{k|k-1}$ can be obtained by Student's t-distribution ((4)-(5) is the special case).

B. Measurement update

Since the presence of multiplicative noise in measurement equation affect the distribution of likelihood function and then affect the derivation of measurement update step, we will first discuss the choice of the likelihood, then discuss the approximation of noise covariances, and finally induce the estimation results in the following.

1) Characterization and selection of likelihood function:

We first consider the likelihood PDF $p(\mathbf{z}_k|\mathbf{x}_k)$. From (2), we can see that m_k changes the distribution of $p(\mathbf{z}_k|\mathbf{x}_k)$, because the PDFs of products of independent random Gaussian variables are not necessarily Gaussian functions, but shown to be Meijer G-functions in general [35], [36]. This means that the PDF of $m_k \times \mathbf{x}_k$ in (2) is a Meijer G-function, and therefore the likelihood $p(\mathbf{z}_k|\mathbf{x}_k)$ is non-Gaussian.

Remark 3: The convolution and the product of Gaussian probability density functions (PDFs) are also Gaussian functions, while this fact is limited to different PDFs of the same random variable [37].

The integrations in Definition 1 for the general solution are usually not analytically tractable. Also, an explicit solution for the shape parameters (a_i, b_j) of Meijer G-function cannot be obtained analytically and the majority of the established special distributions can be used to represent the Meijer G-function. On the other hand, it is well known that an infinite mixture of Gaussian functions can approach arbitrary distribution [38]. According to Definition 2, a Student's t-distribution can be obtained by an infinite number of Gaussian mixture and therefore can be used to represent arbitrary distribution since an infinite mixture of Gaussian functions can deduce arbitrary distribution (approximation of a Meijer G-function) [38]:

$$\text{St}(\mathbf{x}; \mu, \lambda, \nu) = \frac{\Gamma(\nu/2+1/2)}{\Gamma(\nu/2)} \left(\frac{\lambda}{\pi\nu}\right)^{1/2} \left[1 + \frac{\lambda(\mathbf{x}-\mu)^2}{\nu}\right]^{-\nu/2-1/2}. \quad (6)$$

Here, μ is the mean, λ is the covariance, and ν is the dof. In particular, when $\nu = 1$, the Student's t-distribution reduces to the Cauchy distribution (Definition 3). In addition, and as $\nu \rightarrow \infty$, it becomes a Gaussian distribution $\mathbf{N}(\mathbf{x}; \mu, \lambda^{-1})$ with mean μ and covariance λ .

Then, we use a Student's t-distribution to formulate the likelihood PDF $p(\mathbf{y}_k|\mathbf{x}_k)$ [39]:

$$\begin{aligned} p(\mathbf{y}_k|\mathbf{x}_k) &= \text{St}(\mathbf{y}_k; m_k \mathbf{H}_k \mathbf{x}_k, \mathbf{R}_k^M, \nu) \\ &= \int \mathbf{N}(\mathbf{y}_k; m_k \mathbf{H}_k \mathbf{x}_k, \mathbf{R}_k^M / \lambda_k) \mathbf{G}(\lambda_k; \frac{\nu}{2}, \frac{\nu}{2}) d\lambda_k, \end{aligned} \quad (7)$$

where $\mathbf{G}(\cdot; \alpha, \beta)$ is the Gamma function (Definition 1) with an auxiliary stochastic variable λ_k , the shape parameter α and the rate parameter β . First define $\mathbf{S}_k \triangleq \mathbf{E}[\mathbf{x}_k \mathbf{x}_k^T]$. Then, μ and λ in (6) for likelihood (2) are organized as

$$\begin{aligned} \mu &= m_k \mathbf{H}_k \mathbf{x}_k, \\ \lambda &= \mathbf{R}_k^M = \bar{m}_k^2 \mathbf{H}_k \mathbf{P}_{k|k-1} \mathbf{H}_k^T + \sigma_k \mathbf{H}_k \mathbf{S}_k \mathbf{H}_k^T + \mathbf{R}_k. \end{aligned}$$

According to (7), $p(\mathbf{y}_k|\mathbf{x}_k)$ can be reformulated as the hierarchical Gaussian distribution form:

$$\begin{aligned} p(\mathbf{y}_k|\mathbf{x}_k, \lambda_k) &= \mathbf{N}(\mathbf{y}_k; m_k \mathbf{H}_k \mathbf{x}_k, \mathbf{R}_k^M / \lambda_k), \\ p(\lambda_k) &= \mathbf{G}(\lambda_k; \frac{\nu}{2}, \frac{\nu}{2}). \end{aligned} \quad (8)$$

From (8), we notice that the auxiliary random variable λ_k adjusts the likelihood PDF of $p(\mathbf{z}_k|\mathbf{x}_k)$. When $\lambda_k = 1$, (8) is simplified into Gaussian likelihood:

$$p(\mathbf{z}_k|\mathbf{x}_k, \mathbf{R}_k^M) = \mathbf{N}(\mathbf{z}_k; m_k \mathbf{H}_k \mathbf{x}_k, \mathbf{R}_k^M), \quad (9)$$

where

$$\mathbf{R}_k^M = \bar{m}_k^2 \mathbf{H}_k \mathbf{P}_{k|k-1} \mathbf{H}_k^T + \sigma_k \mathbf{H}_k \mathbf{S}_k \mathbf{H}_k^T + \mathbf{R}_k.$$

This means that $p(\mathbf{z}_k|\mathbf{x}_k)$ is always a Gaussian cluster or Gaussian truncated λ -PDF [40].

2) *Choices of Covariance Matrices* : Note that objective is to estimate state $\mathbf{x}_k$ together with $\mathbf{P}_{k|k-1}$, $\mathbf{R}_k$ and σ_k , where $\mathbf{R}_k$ and σ_k are AMNCM and MMNCM, respectively. Generally speaking, there are two classes of solutions including direct method and indirect one. The direct method is to compute multiplicative and additive measurement noise covariance matrices $p(\sigma_k, \mathbf{R}_k|\mathbf{y}_{1:k})$ separately while indirect one is to compute $p(\sigma_k, \mathbf{R}_k|\mathbf{y}_{1:k})$ as a whole measurement noise

covariance matrix $p(\mathbf{R}_k^M | \mathbf{y}_{1:k})$. Without loss of generality, we only introduce the indirect method.

In engineering practice, as shown in (2), we cannot distinguish between additive noise and multiplicative noise, since both are regarded as observation noise. Therefore, we choose to estimate them as a whole with a unified covariance matrix $\mathbf{R}_k^M$. We next specify how to infer state $\mathbf{x}_k$ together with $\mathbf{P}_{k|k-1}$, $\mathbf{R}_k^M$.

In order to guarantee that the posterior PDF is consistent with the prior PDF, we need to select the conjugate prior distributions with inaccurate PECM $\mathbf{P}_{k|k-1}$, MMNCM σ_k and AMNCM $\mathbf{R}_k$. In Bayesian analysis [41], we usually choose the inverse Wishart distribution as the conjugate prior for covariance matrix [34].

Remark 4: Both inverse Gamma (IG) distribution and inverse Wishart (IW) distribution can be chosen as the conjugate prior for the covariance matrix [5], [9]. IG is the special case of IW.

Since σ_k , $\mathbf{P}_{k|k-1}$ and $\mathbf{R}_k^M$ are the covariance variable/matrix of Gaussian PDFs, their prior PDFs $p(\mathbf{P}_{k|k-1} | \mathbf{y}_{1:k-1})$ and $p(\mathbf{R}_k^M | \mathbf{y}_{1:k-1})$ are chosen as inverse Wishart PDFs (Definition 4):

$$p(\mathbf{P}_{k|k-1} | \mathbf{y}_{1:k-1}) = \text{IW}(\mathbf{P}_{k|k-1}; \hat{\sigma}_{k|k-1}, \hat{\mathbf{O}}_{k|k-1}), \quad (10)$$

$$p(\mathbf{R}_k^M | \mathbf{y}_{1:k-1}) = \text{IW}(\mathbf{R}_k^M; \hat{u}_{k|k-1}, \hat{\mathbf{U}}_{k|k-1}), \quad (11)$$

where $\text{IW}(\cdot; \mu_k, \Sigma_k)$ is the inverse Wishart distribution with inverse scale matrix Σ_k and dof parameter μ_k . In addition, $\hat{\sigma}_{k|k-1}$, $\hat{\mathbf{O}}_{k|k-1}$ and $\hat{u}_{k|k-1}$, $\hat{\mathbf{U}}_{k|k-1}$ denote the corresponding dof parameters and inverse scale matrices, respectively. We next specify and choose the prior parameters $\hat{\sigma}_{k|k-1}$, $\hat{\mathbf{O}}_{k|k-1}$, $\hat{u}_{k|k-1}$, and $\hat{\mathbf{U}}_{k|k-1}$.

According to [5], the prior $\hat{\sigma}_{k|k-1}$ can be chosen as:

$$\hat{\sigma}_{k|k-1} = n + \tau + 1, \quad (12)$$

where n is the dimension of states and $\tau \geq 0$ is a tuning parameter. To get the prior $\mathbf{P}_{k|k-1}$, $\mathbf{P}_{k|k-1}$ should be set as the nominal one-step prediction of error covariance matrix (PECM) $\tilde{\mathbf{P}}_{k|k-1}$

$$\tilde{\mathbf{P}}_{k|k-1} = \frac{\hat{\mathbf{O}}_{k|k-1}}{\hat{\sigma}_{k|k-1} - n - 1} = \mathbf{F}_{k-1} \mathbf{P}_{k-1|k-1} \mathbf{F}_{k-1}^T + \tilde{\mathbf{Q}}_{k-1}, \quad (13)$$

where $\tilde{\mathbf{Q}}_{k-1}$ is the nominal APNCM. Then,

$$\hat{\mathbf{O}}_{k|k-1} = \tau \tilde{\mathbf{P}}_{k|k-1}. \quad (14)$$

For $\mathbf{R}_k^M$, we choose [5]:

$$\hat{u}_{k|k-1} = \rho (\hat{u}_{k-1|k-1} - m - 1) + m + 1, \quad (15)$$

$$\hat{\mathbf{U}}_{k|k-1} = \rho \hat{\mathbf{U}}_{k-1|k-1}, \quad (16)$$

where m is the dimension of measurements and $\rho \in (0, 1]$ denotes a time-fluctuations forgetting factor. The initial value is

$$\frac{\hat{\mathbf{U}}_{0|0}}{\hat{u}_{0|0} - m - 1} = \tilde{\mathbf{R}}_0^M. \quad (17)$$

3) *Variational Approximations of Posterior PDFs:* Our objective here is to obtain the joint posterior distribution $p(\mathbf{x}_k, \mathbf{P}_{k|k-1}, \mathbf{R}_k^M | \mathbf{y}_{1:k})$. Since the solution for this joint posterior update is not analytically tractable, motivated by [12, 19, 38] a VB-based approach ([12], [19], [38]) is used to obtain an approximate distribution for $p(\mathbf{x}_k, \mathbf{P}_{k|k-1}, \mathbf{R}_k^M | \mathbf{y}_{1:k})$ as follows:

$$p(\mathbf{x}_k, \mathbf{P}_{k|k-1}, \mathbf{R}_k^M | \mathbf{y}_{1:k}) \approx q(\mathbf{x}_k)q(\mathbf{P}_{k|k-1})q(\mathbf{R}_k^M),$$

where $q(\cdot)$ is the approximation distribution of $p(\cdot)$.

$q(\cdot)$ can be calculated by minimizing the KLD [12], [19], [38] as

$$\arg \min \text{KLD} [q(\mathbf{x}_k)q(\mathbf{P}_{k|k-1})q(\mathbf{R}_k^M) || p(\mathbf{x}_k, \mathbf{P}_{k|k-1}, \mathbf{R}_k^M | \mathbf{y}_{1:k})], \quad (18)$$

where $\text{KLD}[\cdot]$ is defined in Definition 5. According to [12], the solution for (18) is :

$$\log q(\theta) = \mathbb{E}_{\Omega(-\theta)} [\log p(\Omega, \mathbf{y}_{1:k})] + C_\theta, \quad (19)$$

$$\Omega \triangleq \{\mathbf{x}_k, \mathbf{P}_{k|k-1}, \mathbf{R}_k^M\},$$

where, θ is one element of Ω , and $-\theta$ represents the complementary set of θ in Ω , and C_θ denotes the constant. Since (19) cannot be solved directly, we apply the fixed-point iteration method to solve it, where $q(\theta)$ is updated after the $(i+1)$ th iteration step result $q^{i+1}(\theta)$ [12], [19], [38].

We next specify likelihood function (8) to obtain the proposed filter (Algorithm 1 in Appendix), and then specify likelihood function as its special Gaussian form (9) to obtain another proposed filter (Algorithm 2 in Appendix).

Based on the conditional independence properties, the joint PDF $p(\Omega, \mathbf{y}_{1:k})$ can be obtained by

$$p(\Omega, \mathbf{y}_{1:k}) = p(\mathbf{y}_{1:k-1})p(\lambda_k)p(\mathbf{R}_k^M | \mathbf{y}_{1:k-1})p(\mathbf{P}_{k|k-1} | \mathbf{y}_{1:k-1}) \\ \times p(\mathbf{y}_k | \mathbf{x}_k, \mathbf{R}_k^M, \lambda_k)p(\mathbf{x}_k | \mathbf{y}_{1:k-1}, \mathbf{P}_{k|k-1}). \quad (20)$$

While the joint PDF of the direct method can be factored as

$$p(\Omega, \mathbf{y}_{1:k}) = p(\sigma_k | \mathbf{y}_{1:k-1})p(\lambda_k)p(\mathbf{y}_{1:k-1}) \\ \times p(\mathbf{y}_k | \mathbf{x}_k, \mathbf{R}_k^M, \lambda_k)p(\mathbf{x}_k | \mathbf{y}_{1:k-1})p(\mathbf{R}_k^M | \sigma_k).$$

Substituting (3), (8) and (10-11) into (20) yields

$$p(\Omega, \mathbf{y}_{1:k}) = \text{N}(\mathbf{y}_k; m_k \mathbf{H}_k \mathbf{x}_k, \mathbf{R}_k^M / \lambda_k) \text{G}(\lambda_k; \frac{v}{2}, \frac{v}{2}) \\ \times \text{N}(\mathbf{x}_k; \hat{\mathbf{x}}_{k|k-1}, \mathbf{P}_{k|k-1}) \text{IW}(\mathbf{P}_{k|k-1}; \hat{\sigma}_{k|k-1}, \hat{\mathbf{O}}_{k|k-1}) \\ \times \text{IW}(\mathbf{R}_k^M; \hat{u}_{k|k-1}, \hat{\mathbf{U}}_{k|k-1}) p(\mathbf{y}_{1:k-1}). \quad (21)$$

It then follows that $\log p(\Omega, \mathbf{y}_{1:k})$ is formulated as

$$\log p(\Omega, \mathbf{y}_{1:k}) = \left(\frac{m+v}{2} - 1 \right) \log \lambda_k - \frac{v}{2} \lambda_k \\ - 0.5(m + \hat{u}_{k|k-1} + 2) \log |\mathbf{R}_k^M| \\ - 0.5 \lambda_k (\mathbf{y}_k - m_k \mathbf{H}_k \mathbf{x}_k)^T (\mathbf{R}_k^M)^{-1} (\mathbf{y}_k - m_k \mathbf{H}_k \mathbf{x}_k) \\ - 0.5 \text{tr}(\hat{\mathbf{U}}_{k|k-1} (\mathbf{R}_k^M)^{-1}) - 0.5(n + \hat{\sigma}_{k|k-1} + 2) \log |\mathbf{P}_{k|k-1}| \\ - 0.5(\mathbf{x}_k - \hat{\mathbf{x}}_{k|k-1})^T \times \mathbf{P}_{k|k-1}^{-1} (\mathbf{x}_k - \hat{\mathbf{x}}_{k|k-1}) \\ - 0.5 \text{tr}(\hat{\mathbf{O}}_{k|k-1} \mathbf{P}_{k|k-1}^{-1}) + C_\Omega. \quad (22)$$

By using $\theta = \lambda_k$ and (22) in (19), we obtain

$$\log q^{(i+1)}(\lambda_k) = \left(\frac{m+v}{2} - 1\right) \log \lambda_k - 0.5 \left\{ v + \text{tr} \left(\mathbf{E}_k^{(i)} (\mathbf{R}_k^M)^{-1} \right) \right\} \lambda_k + C_\lambda, \quad (23)$$

where $\mathbf{E}_k^{(i)}$ can be computed by

$$\mathbf{E}_k^{(i)} = \mathbf{E}^{(i)}[(\mathbf{y}_k - \bar{m}_k \mathbf{H}_k \mathbf{x}_k)(\mathbf{y}_k - \bar{m}_k \mathbf{H}_k \mathbf{x}_k)^T]. \quad (24)$$

Using (23), $q^{(i+1)}(\lambda_k)$ is updated and computed by a Gamma distribution with rate parameter δ_k^{i+1} and shape parameter γ_k^{i+1} are given by

$$\gamma_k^{i+1} = 0.5(m+v), \quad (25)$$

$$\delta_k^{i+1} = 0.5 \left\{ \text{tr} \left(\mathbf{E}_k^{(i)} [\mathbf{R}_k^M]^{-1} \right) + v \right\}. \quad (26)$$

Letting $\theta = \mathbf{P}_{k|k-1}$, and employing (21) in (19), we obtain

$$\begin{aligned} \log q^{(i+1)}(\mathbf{P}_{k|k-1}) &= -0.5(m + \hat{u}_{k|k-1} + 2) \mathbf{E}^{(i)}[\log |\mathbf{R}_k^M|] \\ &\quad - 0.5 \mathbf{E}^{(i)}[(\mathbf{y}_k - m_k \mathbf{H}_k \mathbf{x}_k)^T (\mathbf{R}_k^M)^{-1} (\mathbf{y}_k - m_k \mathbf{H}_k \mathbf{x}_k)] \\ &\quad - 0.5 \mathbf{E}^{(i)}[\text{tr}(\hat{\mathbf{U}}_{k|k-1} (\mathbf{R}_k^M)^{-1})] - 0.5(n + \hat{o}_{k|k-1} + 2) \\ &\quad \times \log |\mathbf{P}_{k|k-1}| - 0.5 \text{tr}((\mathbf{A}_k^{(i)} + \hat{\mathbf{T}}_{k|k-1}) \mathbf{P}_{k|k-1}^{-1}) + C_\Omega \\ &= -0.5(n + \hat{t}_{k|k-1} + 2) \log |\mathbf{P}_{k|k-1}| \\ &\quad - 0.5 \text{tr}((\mathbf{A}_k^{(i)} + \hat{\mathbf{O}}_{k|k-1}) \mathbf{P}_{k|k-1}^{-1}) + C_P, \end{aligned} \quad (27)$$

where $\mathbf{A}_k^{(i)}$ is defined as

$$\begin{aligned} \mathbf{A}_k^{(i)} &= \mathbf{E}^{(i)}[(\mathbf{x}_k - \hat{\mathbf{x}}_{k|k-1})(\mathbf{x}_k - \hat{\mathbf{x}}_{k|k-1})^T] \\ &= \mathbf{P}_{k|k}^i + (\hat{\mathbf{x}}_{k|k}^i - \hat{\mathbf{x}}_{k|k-1})(\hat{\mathbf{x}}_{k|k}^i - \hat{\mathbf{x}}_{k|k-1})^T. \end{aligned} \quad (28)$$

Using (27), $q^{(i+1)}(\mathbf{P}_{k|k-1})$ is updated and computed by an inverse Wishart distribution with inverse scale matrix $\hat{\mathbf{O}}_k^{(i+1)}$ and dof parameter $\hat{o}_k^{(i+1)}$, i.e.,

$$q^{(i+1)}(\mathbf{P}_{k|k-1}) = \text{IW}(\mathbf{P}_{k|k-1}; \hat{o}_k^{(i+1)}, \hat{\mathbf{O}}_k^{(i+1)}),$$

where shape parameters are given by

$$\hat{t}_k^{(i+1)} = \hat{o}_{k|k-1} + 1, \quad (29)$$

$$\hat{\mathbf{O}}_k^{(i+1)} = \mathbf{A}_k^{(i)} + \hat{\mathbf{O}}_{k|k-1}. \quad (30)$$

Letting $\theta = \mathbf{R}_k^M$ and using (22) in (19), we obtain

$$\begin{aligned} \log q^{(i+1)}(\mathbf{R}_k^M) &= -0.5(m + \hat{u}_{k|k-1} + 2) \log |\mathbf{R}_k^M| \\ &\quad - 0.5 \text{tr} \left((\mathbf{B}_k^{(i)} \mathbf{E}^{(i+1)}[\lambda_k] + \hat{\mathbf{U}}_{k|k-1}) (\mathbf{R}_k^M)^{-1} \right) + C_R, \end{aligned} \quad (31)$$

where $\mathbf{B}_k^{(i)}$ can be obtained by

$$\begin{aligned} \mathbf{B}_k^{(i)} &= \mathbf{E}^{(i)}[(\mathbf{y}_k - m_k \mathbf{H}_k \mathbf{x}_k)(\mathbf{y}_k - m_k \mathbf{H}_k \mathbf{x}_k)^T] \\ &= \mathbf{E}^{(i)}[(\mathbf{y}_k - \bar{m}_k \mathbf{H}_k \hat{\mathbf{x}}_{k|k}^i + \bar{m}_k \mathbf{H}_k \hat{\mathbf{x}}_{k|k}^i - m_k \mathbf{H}_k \mathbf{x}_k) \\ &\quad \times (\mathbf{y}_k - \bar{m}_k \mathbf{H}_k \hat{\mathbf{x}}_{k|k}^i + \bar{m}_k \mathbf{H}_k \hat{\mathbf{x}}_{k|k}^i - m_k \mathbf{H}_k \mathbf{x}_k)^T] \\ &= (\mathbf{y}_k - \bar{m}_k \mathbf{H}_k \hat{\mathbf{x}}_{k|k}^i)(\mathbf{y}_k - \bar{m}_k \mathbf{H}_k \hat{\mathbf{x}}_{k|k}^i)^T \\ &\quad + \bar{m}_k^2 \mathbf{H}_k \mathbf{E}^{(i)}[(\mathbf{x}_k - \hat{\mathbf{x}}_{k|k}^i)(\mathbf{x}_k - \hat{\mathbf{x}}_{k|k}^i)^T] \mathbf{H}_k^T \\ &= (\mathbf{y}_k - \bar{m}_k \mathbf{H}_k \hat{\mathbf{x}}_{k|k}^i)(\mathbf{y}_k - \bar{m}_k \mathbf{H}_k \hat{\mathbf{x}}_{k|k}^i)^T \\ &\quad + \bar{m}_k^2 \mathbf{H}_k \mathbf{P}_{k|k}^i \mathbf{H}_k^T. \end{aligned} \quad (32)$$

Employing (31), an inverse Wishart distribution with dof parameter $\hat{u}_k^{(i+1)}$ and inverse scale matrix $\hat{\mathbf{U}}_k^{(i+1)}$ is used to update $q^{(i+1)}(\mathbf{R}_k^M)$ as:

$$q^{(i+1)}(\mathbf{R}_k^M) = \text{IW}(\mathbf{R}_k^M; \hat{u}_k^{(i+1)}, \hat{\mathbf{U}}_k^{(i+1)}), \quad (33)$$

where

$$\hat{u}_k^{(i+1)} = \hat{u}_{k|k-1} + 1, \quad (34)$$

$$\hat{\mathbf{U}}_k^{(i+1)} = \mathbf{B}_k^{(i)} \mathbf{E}^{(i+1)}[\lambda_k] + \hat{\mathbf{U}}_{k|k-1}. \quad (35)$$

Letting $\theta = \mathbf{x}_k$ and using (22) in (19), it has

$$\begin{aligned} \log q^{(i+1)}(\mathbf{x}_k) &= -0.5 \mathbf{E}^{(i+1)}[\lambda_k] (\mathbf{y}_k - \bar{m}_k \mathbf{H}_k \mathbf{x}_k)^T \\ &\quad \times \mathbf{E}^{(i+1)}[(\mathbf{R}_k^M)^{-1}] (\mathbf{y}_k - \bar{m}_k \mathbf{H}_k \mathbf{x}_k) - 0.5(\mathbf{x}_k - \hat{\mathbf{x}}_{k|k-1})^T \\ &\quad \times \mathbf{E}^{(i+1)}[\mathbf{P}_{k|k-1}^{-1}] (\mathbf{x}_k - \hat{\mathbf{x}}_{k|k-1}) + C_\Omega, \end{aligned} \quad (36)$$

where

$$\mathbf{E}^{(i+1)}[\lambda_k] = \frac{\gamma_k^{i+1}}{\delta_k^{i+1}}, \quad \mathbf{E}^{(i+1)}[(\mathbf{R}_k^M)^{-1}] \quad \text{and} \quad \mathbf{E}^{(i+1)}[\mathbf{P}_{k|k-1}^{-1}]$$

can be computed by

$$\mathbf{E}^{(i+1)}[\lambda_k] = \gamma_k^{i+1} / \delta_k^{i+1}, \quad (37)$$

$$\mathbf{E}^{(i+1)}[(\mathbf{R}_k^M)^{-1}] = (\hat{u}_k^{(i+1)} - n - 1)(\hat{\mathbf{U}}_k^{(i+1)})^{-1}, \quad (38)$$

$$\mathbf{E}^{(i+1)}[\mathbf{P}_{k|k-1}^{-1}] = (\hat{o}_k^{(i+1)} - n - 1)(\hat{\mathbf{O}}_k^{(i+1)})^{-1}. \quad (39)$$

Define the updated one-step prediction

$$p^{(i+1)}(\mathbf{x}_k | \mathbf{y}_{1:k-1}) = \text{N}(\mathbf{x}_k; \hat{\mathbf{x}}_{k|k-1}, \hat{\mathbf{P}}_{k|k-1}^{(i+1)}), \quad (40)$$

$$p^{(i+1)}(\mathbf{y}_k | \mathbf{x}_k) = \text{N}(\mathbf{y}_k; \bar{m}_k \mathbf{H}_k \mathbf{x}_k, (\hat{\mathbf{R}}_k^M)^{(i+1)}), \quad (41)$$

where $\hat{\mathbf{P}}_{k|k-1}^{(i+1)}$ and $(\hat{\mathbf{R}}_k^M)^{(i+1)}$ are organized as

$$\hat{\mathbf{P}}_{k|k-1}^{(i+1)} = \left\{ \mathbf{E}^{(i+1)}[\mathbf{P}_{k|k-1}^{-1}] \right\}^{-1} (\hat{\mathbf{R}}_k^M)^{(i+1)} = \frac{\{\mathbf{E}^{(i+1)}[(\mathbf{R}_k^M)^{-1}]\}^{-1}}{\mathbf{E}^{(i+1)}[\lambda_k]}. \quad (42)$$

Using (40)-(42) in (36) yields

$$q^{(i+1)}(\mathbf{x}_k) = \frac{p^{(i+1)}(\mathbf{y}_k | \mathbf{x}_k) p^{(i+1)}(\mathbf{x}_k | \mathbf{y}_{1:k-1})}{\int p^{(i+1)}(\mathbf{y}_k | \mathbf{x}_k) p^{(i+1)}(\mathbf{x}_k | \mathbf{y}_{1:k-1}) d\mathbf{x}_k}. \quad (43)$$

According to (40)-(43), we use a Gaussian PDF with mean vector $\mathbf{x}_k$ and covariance $\mathbf{P}_{k|k}^{(i+1)}$ to compute $q^{(i+1)}(\mathbf{x}_k)$:

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

$\hat{\mathbf{x}}_{k|k}^{(i+1)}$, $\mathbf{P}_{k|k}^{(i+1)}$ and gain matrix $\mathbf{K}_k^{(i+1)}$ are updated through N iterations, where the $i+1$ th iteration is given by

$$\mathbf{K}_k^{(i+1)} = \bar{m}_k \tilde{\mathbf{P}}_{k|k-1}^{(i+1)} \mathbf{H}_k^T [(\bar{m}_k)^2 \mathbf{H}_k \tilde{\mathbf{P}}_{k|k-1}^{(i+1)} \mathbf{H}_k^T + (\tilde{\mathbf{R}}_k^M)^{(i+1)}]^{-1}, \quad (44)$$

$$\hat{\mathbf{x}}_{k|k}^{(i+1)} = \hat{\mathbf{x}}_{k|k-1} + \mathbf{K}_k^{(i+1)} (\mathbf{y}_k - \bar{m}_k \mathbf{H}_k \hat{\mathbf{x}}_{k|k-1}), \quad (45)$$

$$\mathbf{P}_{k|k}^{(i+1)} = \tilde{\mathbf{P}}_{k|k-1}^{(i+1)} - \bar{m}_k \mathbf{K}_k^{(i+1)} \mathbf{H}_k \tilde{\mathbf{P}}_{k|k-1}^{(i+1)}. \quad (46)$$

After fixed-point iteration N , the approximate posterior PDFs of $q(\mathbf{x}_k)$, $q(\mathbf{P}_{k|k-1})$, $q(\mathbf{R}_k^M)$ and $q(\lambda_k)$ can be updated as

$$\begin{aligned} q(\mathbf{x}_k) &\approx q^{(N)}(\mathbf{x}_k) = \mathcal{N}(\mathbf{x}_k; \hat{\mathbf{x}}_{k|k}^{(N)}, \mathbf{P}_{k|k}^{(N)}) = \mathcal{N}(\mathbf{x}_k; \hat{\mathbf{x}}_{k|k}, \mathbf{P}_{k|k}), \\ q(\mathbf{P}_{k|k-1}) &\approx q^{(N)}(\mathbf{P}_{k|k-1}) = \text{IW}(\mathbf{P}_{k|k-1}; \hat{\sigma}_{k|k}^{(N)}, \hat{\mathbf{O}}_k^{(N)}) \\ &= \text{IW}(\mathbf{P}_{k|k-1}; \hat{\sigma}_{k|k}, \hat{\mathbf{O}}_{k|k}), \\ q(\mathbf{R}_k^M) &\approx q^{(N)}(\mathbf{R}_k^M) = \text{IW}(\mathbf{R}_k^M; \hat{u}_k^{(N)}, \hat{\mathbf{U}}_k^{(N)}) \\ &= \text{IW}(\mathbf{R}_k^M; \hat{u}_{k|k}, \hat{\mathbf{U}}_{k|k}), \\ q(\lambda_k) &\approx q^{(N)}(\lambda_k) = \mathbf{G}(\lambda_k; \gamma_k^N, \delta_k^N). \end{aligned} \quad (47)$$

To sum up, proposed VB-based filter with unknown covariance of multiplicative and additive noises consists of (4), (12)-(17), (24)-(26), (31-35), (37)-(42) and (44)-(46), whose implementation pseudo code is shown in Algorithm 1 in Appendix.

We also consider a special likelihood function $p(\mathbf{y}_k|\mathbf{x}_k)$ when (9) is Gaussian ($\nu \rightarrow \infty$). The algorithm can be obtained similarly to that of non-Gaussian case and then organized as Algorithm 2 in Appendix.

IV. PERFORMANCE ANALYSIS

We analyze the stability of the proposed filter because the effect of the parameters τ , ρ , $\tilde{\mathbf{Q}}_k$ and $\tilde{\mathbf{R}}_0^M$.

First, we discuss the influence of τ on the proposed filter. Substituting (39) into (42) yields

$$\hat{\mathbf{P}}_{k|k-1}^{(i+1)} = \frac{\hat{\mathbf{O}}_k^{(i+1)}}{\hat{\sigma}_k^{(i+1)} - n - 1}. \quad (48)$$

According to (12), (14) and (29)-(30), (48) can be rewritten as

$$\hat{\mathbf{P}}_{k|k-1}^{(i+1)} = \frac{\hat{\mathbf{O}}_{k|k-1} + \mathbf{A}_k^{(i)}}{\hat{t}_{k|k-1} - n} = \frac{\tau \tilde{\mathbf{P}}_{k|k-1} + \mathbf{A}_k^{(i)}}{\tau + 1}. \quad (49)$$

From (49), we know that τ is a tradeoff parameter to balance $\tilde{\mathbf{P}}_{k|k-1}$ and $\mathbf{A}_k^{(i)}$. The larger τ corresponds to much prior uncertainties about the nominal APNCM, and leads to worse performance of filter. The smaller τ means that much information is lost, thus it also leads the worse accuracy of filter. We have experimental and observational evidence concerning the choice, τ can be within the range $\tau \in [10, 800]$, since this scope can accommodate the effects of multiplicative noise.

Remark 5: The proposed filter selects τ more flexibly than additive noise filter does [5], [9].

Second, we discuss the effect of ρ . Substituting (38) into (42) yields

$$(\hat{\mathbf{R}}_k^M)^{(i+1)} = \frac{\hat{\mathbf{U}}_k^{(i+1)}}{\hat{u}_k^{(i+1)} - m - 1}. \quad (50)$$

Based on (15)-(16) and (34-35), (50) can be reorganized as

$$(\hat{\mathbf{R}}_k^M)^{(i+1)} = \frac{\mathbf{B}_k^{(i)} + \hat{\mathbf{U}}_{k|k-1}}{\hat{u}_{k|k-1} - m} = \frac{\mathbf{B}_k^{(i)} + \rho \hat{\mathbf{U}}_{k-1|k-1}}{1 + \rho(\hat{u}_{k-1|k-1} - m - 1)}. \quad (51)$$

Based on (47), $(\hat{\mathbf{R}}_{k-1}^M)^{(i+1)}$ can be calculated by

$$(\hat{\mathbf{R}}_{k-1}^M)^{(i+1)} = \frac{\hat{\mathbf{U}}_{k-1|k-1}}{\hat{u}_{k-1|k-1} - m - 1}. \quad (52)$$

Combining (52) with (51) yields

$$(\hat{\mathbf{R}}_k^M)^{(i+1)} = \frac{\rho(\hat{u}_{k-1|k-1} - m - 1)(\hat{\mathbf{R}}_{k-1}^M)^{(i)} + \mathbf{B}_k^{(i)}}{\rho(\hat{u}_{k-1|k-1} - m - 1) + 1}. \quad (53)$$

Based on (15), (34) and (47), we obtain

$$\hat{u}_{k|k} = \rho(\hat{u}_{k-1|k-1} - m - 1) + m + 2. \quad (54)$$

Then, (54) can be rewritten as

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

Set $\eta(k, \rho) = \rho(\hat{u}_{k-1|k-1} - m - 1)$ and (53) can be rewritten as

$$(\hat{\mathbf{R}}_k^M)^{(i+1)} = \frac{\eta(k, \rho)(\hat{\mathbf{R}}_{k-1}^M)^{(i)} + \mathbf{B}_k^{(i)}}{\eta(k, \rho) + 1}. \quad (55)$$

Then, $\eta(k, \rho)$ can be reorganized as

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

This yields

$$\lim_{k \rightarrow +\infty} \eta(k, \rho) = \rho/(1 - \rho). \quad (57)$$

From (55)-(57), we similarly obtain that $\eta(k, \rho)$ is a tradeoff parameter to balance the weight of $\hat{\mathbf{R}}_{k-1}^M$ and $\mathbf{B}_k^{(i)}$. In addition, $\eta(k, \rho)$ is monotonically increasing with respect to ρ . Therefore, $\rho \in (0, 1]$ can be directly employed to adjust $\hat{\mathbf{R}}_{k-1}^M$ upon the updated $(\hat{\mathbf{R}}_k^M)^{(i+1)}$. The larger ρ is, the more information $\hat{\mathbf{R}}_{k-1}^M$ inherits, and vice versa. The forgetting factor is within the range $\rho \in [0.4, 1]$ experimentally [5], [9].

Third, we discuss the effect of $\tilde{\mathbf{Q}}_k$ and $\tilde{\mathbf{R}}_0^M$. The initial values are

$$\begin{cases} \hat{\mathbf{P}}_{k|k-1}^{(0)} = \tilde{\mathbf{P}}_{k|k-1} = \mathbf{F}_{k-1} \mathbf{P}_{k-1|k-1} \mathbf{F}_{k-1}^T + \tilde{\mathbf{Q}}_{k-1}, \\ (\hat{\mathbf{R}}_k^M)^{(0)} = \hat{\mathbf{R}}_{k-1}^M. \end{cases} \quad (58)$$

Set $\zeta_k = \frac{\eta(k, \rho)}{\eta(k, \rho) + 1}$, $\mathbf{C}_k = \frac{\mathbf{B}_k^{(N-1)}}{\eta(k, \rho) + 1}$, where $0 < \zeta(k) < 1$ and $\mathbf{C}_k \geq \mathbf{0}$, (55) can be rewritten as

$$\hat{\mathbf{R}}_k^M = \zeta_k \hat{\mathbf{R}}_{k-1}^M + \mathbf{C}_k. \quad (59)$$

Set $\hat{\mathbf{R}}_0^M = \tilde{\mathbf{R}}_0^M$, (59) can be reformulated as

$$\hat{\mathbf{R}}_{k-1}^M = \sum_{i=1}^{k-1} \left(\prod_{j=i+1}^{k-1} \zeta_j \right) \mathbf{C}_i + \left(\prod_{i=1}^{k-1} \zeta_i \right) \tilde{\mathbf{R}}_0^M. \quad (60)$$

From (58) and (60), we know that the initial APNCM $\tilde{\mathbf{Q}}_k$ and MNCM $\tilde{\mathbf{R}}_0^M$ affect $\hat{\mathbf{P}}_{k|k-1}^{(0)}$ and $(\hat{\mathbf{R}}_k^M)^{(0)}$. When k increases, $\tilde{\mathbf{R}}_0^M$ proportionally affects on $\hat{\mathbf{R}}_k^M$. To ensure that $\hat{\mathbf{P}}_{k|k-1}^{(i)}$ and $(\hat{\mathbf{R}}_k^M)^{(i)}$ converge, we need to choose appropriate $\hat{\mathbf{P}}_{k|k-1}^{(0)}$ and $(\hat{\mathbf{R}}_k^M)^{(0)}$ since the VB inference can only guarantee local convergence. Therefore, these parameters need to be selected near the true value and we rely on the experimental experience, that we selects $\tilde{\mathbf{Q}}_k$ and $\tilde{\mathbf{R}}_0^M$ as $\tilde{\mathbf{Q}}_k = \text{diag}[\alpha_{1,k}, \dots, \alpha_{i,k}, \dots, \alpha_{n,k}]$ and $\tilde{\mathbf{R}}_0^M = \text{diag}[\beta_1, \dots, \beta_j, \dots, \beta_m]$, where $\alpha_{i,k} > 0$ and $\beta_j > 0$.

Finally, we discuss the influence of initial $\tilde{\mathbf{Q}}_k$ and $\tilde{\mathbf{R}}_0^M$ on stability of the proposed filter. Based on (58) and (60), $\mathbf{Q}_k > \mathbf{0}$, $\tilde{\mathbf{R}}_0^M > \mathbf{0}$, $0 < \zeta(k) < 1$ and $\mathbf{C}_k \geq \mathbf{0}$, we obtain

$$\hat{\mathbf{P}}_{k|k-1} > \mathbf{0}, \quad \hat{\mathbf{R}}_{k-1}^M > \mathbf{0}. \quad (61)$$

Note that (28) and (32) yields

$$\mathbf{A}_k^{(i)} \geq \mathbf{0}, \quad \mathbf{B}_k^{(i)} \geq \mathbf{0}. \quad (62)$$

Then, substituting (61)-(62) into (48) and (55) yields

$$\hat{\mathbf{p}}_{k|k-1}^{(i+1)} > \mathbf{0}, \quad (\hat{\mathbf{R}}_k^M)^{(i+1)} > \mathbf{0}.$$

This means that PECM $\hat{\mathbf{p}}_{k|k-1}^{(i+1)}$ and MNCM $(\hat{\mathbf{R}}_k^M)^{(i+1)}$ are positive definite and indicates that the proposed filter is stable.

Remark 6: The higher the number of iterations N , the more run time will need and the more estimation accuracy. We need to balance precision and time complexity according to the dimensions of noises. Different from system with additive noises, we suggest that system with multiplicative noise should choose a larger N to guarantee that algorithms converge to a local optimum. Meanwhile, the parameters discussed above have wider selection range and more flexible selection degree as compared with system with additive noises, since the corresponding system changes more dramatically.

V. SIMULATIONS

In this section, we verify the proposed adaptive filter by a target tracking example. In 2-D space, the target moves with a constant velocity and we can measure its position. Process equation and observation equation are

$$\begin{aligned} \mathbf{x}_k &= \mathbf{F}_k \mathbf{x}_{k-1} + \mathbf{w}_{k-1}, \\ \mathbf{y}_k &= m_k \mathbf{H}_k \mathbf{x}_k + \mathbf{v}_k, \end{aligned}$$

where the state has four dimension $\mathbf{x}_k = [x_k \ y_k \ \dot{x}_k \ \dot{y}_k]^T$; Components x_k , y_k , $\dot{x}_k$ and $\dot{y}_k$ denote the positions and corresponding velocities.

The parameters are set as

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

where $\Delta t = 1s$ is the sampling interval and $\mathbf{I}_2$ is a 2×2 identity matrix. Similar to [5], [42], the true additive process noise covariance is slowly time-varying and additive observation noise covariance is also slowly time-varying, which are given by

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

where $T = 1000s$ denotes the step size and $q = 1m^2/s^3$ and $r = 100m^2$. Multiplicative noise m_k is set as a Gaussian process with mean $\bar{m}_k$ and variance: $\sigma_k = [0.1 + 0.05 \cos(\frac{\pi k}{T})] r$.

In addition, the nominal APNCM and MNCM are respectively set as $\tilde{\mathbf{Q}}_k = \alpha \mathbf{I}_4$ and $\tilde{\mathbf{R}}_0^M = \beta \mathbf{I}_2$, where $\mathbf{I}_4$ is the 4×4 identity matrix. Filter without/multiplicative noise

Fig. 1. RMSEs of the positions and velocities.

and nominal covariance matrices ($\tilde{\mathbf{Q}}_k$ and $\tilde{\mathbf{R}}_0^M$) (denoted as KFNCM/MKFNCM) and Kalman filter without/with multiplicative noise and true covariance matrices ($\mathbf{Q}_k$ and $\mathbf{R}_k^M$) (denoted as KFTCM/MKFTCM) [28], the existing VBAKF [5], the existing particle filter (PF) [43], VBPF [33], and the proposed VB-based filter (λ_k in Algorithm 1 is set to be 1) are tested. For existing KFNCM/MKFTCM and VBAKF and the presented filter, we set $\rho = 1 - \exp(-4)$, $\tau = 200$, $\alpha = 1$, $\beta = 100$, and $N = 100$. The simulation is performed in a computer with Intel Core i7-5600 CPU at 2.60 GHz and MATLAB R2019a software.

Remark 7: The existing state-of-the-art filters either for time-invariant multiplicative noise [27], [28], [30]–[32] or only for additive time-varying noise covariance [5], [9], [10], [25], [33]. Therefore, we compare the results with particle filters [43] to demonstrate the superiority of the presented filters. The particle filters are used with 1000 particles.

To evaluate the estimation accuracy of the state and the predicted error covariance matrix (PECM) and measurement noise covariance matrix (MNCM), the RMSEs of positions and velocities and the square roots of normalized Frobenius norm (SRNFN) are chosen as performance metrics and are defined by

$$\begin{aligned} \text{RMSE} &\triangleq \sqrt{\frac{1}{M} \sum_{s=1}^M ((x_k^s - \hat{x}_k^s)^2 + (y_k^s - \hat{y}_k^s)^2)}, \\ \text{SRNFN} &\triangleq \left(\frac{1}{n^2 M} \sum_{s=1}^M \|\hat{\mathbf{P}}_{k|k-1}^s - \mathbf{P}_{k|k-1}^s\|^2 \right)^{\frac{1}{4}}, \end{aligned}$$

where $M = 1000$ denotes the total number of Monte Carlo runs, s denotes s th Monte Carlo run, (x_k^s, y_k^s) and $(\hat{x}_k^s, \hat{y}_k^s)$ are the true and estimated variables (position or velocity), and $\mathbf{P}_{k|k-1}^s$ denotes the true PECM (or MNCM) at the s th run given by MKFTCM.

The RMSEs and the SRNFNs from existing filters and the proposed filter are respectively shown in Figs. 1 and 2 with multiplicative mean $\bar{m} = 20$. It can be seen from Fig. 1 that the proposed filter converge and has relatively smaller and

Fig. 2. SRNFNs of PECM and MNMCM.

Fig. 5. RMSEs of the positions and velocities.

Fig. 3. RMSEs of the positions and velocities.

Fig. 6. SRNFNs of the PECM and MNMCM.

Fig. 4. SRNFNs of the PECM and MNMCM.

Fig. 7. RMSEs of the positions and velocities when $\tau = 2, 10, 100, 200, 600, 800$.

stable RMSEs than those of KFNCM, VBAKF, and KFTCM. It can be also seen from the figures, the existing filters diverge eventually, while the proposed filter has robust convergence.

Figs. 3 and 4 show the results with multiplicative mean $\bar{m} = 3$. Figs. 5 and 6 demonstrate the results with multiplicative

Fig. 8. RMSEs of the positions and velocities when $\rho = 0.4, 0.6, 0.8, 0.9, 0.98, 1$.

Fig. 9. RMSEs of the positions and velocities using nonlinear filters.

mean $\bar{m} = 0.2$. We can see from the figures, that although the multiplicative disturbance enlarges or weakens the true signal, it cannot affect the proposed filter but has a huge impact on the existing filters.

Fig. 7 demonstrates the RMSEs of positions and velocities from the existing filters and the proposed filters when $\tau = 2, 10, 100, 200, 600, 800$. As we can see from Fig. 7, the presented adaptive filter has wider tuning range than existing filters (all existing filters diverge eventually) and has better estimation and essentially consistent estimation performance than existing filters do.

Fig. 8 shows results of the RMSEs when $\rho = 0.4, 0.6, 0.8, 0.9, 0.98, 1$. From Fig. 8, the proposed adaptive filter is more flexible and has wider selection range within $\rho \in [0.4, 1]$ than existing filters within $\rho \in [0.9, 1]$ and has higher estimation accuracy and essentially consistent estimation performance than those existing filters, which diverge eventually.

Fig. 9 draws the RMSEs from the proposed filters, the existing particle filter (PF) [43] and VBPF [33] with 1000

particles, and $\tau = 400, \rho = 0.98$. We can easily get that the presented adaptive filter can estimate the adaptive noise covariance and can also achieve better performance than these state-of-the-art nonlinear filters when system are affected by multiplicative noise.

VI. CONCLUSIONS

In this paper, we mainly focus on estimation problem of linear state-space systems with unknown additive and multiplicative noise covariances. A novel VB-based adaptive filter is proposed, where the states together with noise covariances were estimated, approached and inferred by choosing inverse Wishart priors as the unknown covariance and Student's-t distribution as the likelihood function. Simulation results illustrate that the proposed VB-based adaptive filter is more robust to eliminate strong noise (multiplicative) effects and to resist the uncertainties and disturbance with unknown noise covariances.

REFERENCES

- [1] R. Mehra, "Approaches to adaptive filtering," *IEEE Transactions on Automatic Control*, vol. 17, no. 5, pp. 693–698, October 1972.
- [2] A. H. Mohamed and K. P. Schwarz, "Adaptive kalman filtering for ins/gps," *Journal of Geodesy*, vol. 73, no. 4, pp. 193–203.
- [3] M. Yu, "Ins/gps integration system using adaptive filter for estimating measurement noise variance," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 48, no. 2, pp. 1786–1792, APRIL 2012.
- [4] Y. Huang, Y. Zhang, B. Xu, Z. Wu, and J. Chambers, "A new outlier-robust student's t based gaussian approximate filter for cooperative localization," *IEEE/ASME Transactions on Mechatronics*, vol. 22, no. 5, pp. 2380–2386, 2017.
- [5] Y. Huang, Y. Zhang, Z. Wu, N. Li, and J. Chambers, "A novel adaptive kalman filter with inaccurate process and measurement noise covariance matrices," *IEEE Transactions on Automatic Control*, vol. 63, no. 2, pp. 594–601, 2017.
- [6] M. Karasol and X. Hu, "An optimization approach to adaptive kalman filtering," *Automatica*, vol. 47, no. 8, pp. 1785–1793, 2011.
- [7] X. R. Li and Y. Bar-Shalom, "A recursive multiple model approach to noise identification," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 30, no. 3, pp. 671–684, 1994.
- [8] 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 Transactions on Industrial Electronics*, vol. 59, no. 11, pp. 4315–4325, Nov 2012.
- [9] S. Särkkä and A. Nummenmaa, "Recursive noise adaptive kalman filtering by variational bayesian approximations," *IEEE Transactions on Automatic control*, vol. 54, no. 3, pp. 596–600, 2009.
- [10] I. S. Mbalawata, S. Särkkä, M. Vihola, and H. Haario, "Adaptive metropolis algorithm using variational bayesian adaptive kalman filter," *Computational Statistics & Data Analysis*, vol. 83, pp. 101–115, 2015.
- [11] V. Smidl and A. Quinn, "Variational bayesian filtering," *IEEE Transactions on Signal Processing*, vol. 56, no. 10, pp. 5020–5030, Oct 2008.
- [12] D. G. Tzikas, A. C. Likas, and N. P. Galatsanos, "The variational approximation for bayesian inference," *IEEE Signal Processing Magazine*, vol. 25, no. 6, pp. 131–146, 2008.
- [13] S. Ji, B. Krishnapuram, and L. Carin, "Variational bayes for continuous hidden markov models and its application to active learning," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 28, no. 4, pp. 522–532, 2006.
- [14] J. Vermaak, N. D. Lawrence, and P. Pérez, "Variational inference for visual tracking," in *2003 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2003. Proceedings.*, vol. 1. IEEE, 2003, pp. 1–1.
- [15] R. Ksantini, D. Ziou, B. Colin, and F. Dubeau, "Weighted pseudometric discriminatory power improvement using a bayesian logistic regression model based on a variational method," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 30, no. 2, pp. 253–266, Feb 2008.

- [16] M. Shiga and H. Mamitsuka, "A variational bayesian framework for clustering with multiple graphs," *IEEE Transactions on Knowledge and Data Engineering*, vol. 24, no. 4, pp. 577–590, April 2012.
- [17] S. Särkkä and J. Hartikainen, "Non-linear noise adaptive kalman filtering via variational bayes," in *2013 IEEE International Workshop on Machine Learning for Signal Processing (MLSP)*. IEEE, 2013, pp. 1–6.
- [18] G. Agamennoni, J. I. Nieto, and E. M. Nebot, "Approximate inference in state-space models with heavy-tailed noise," *IEEE Transactions on Signal Processing*, vol. 60, no. 10, pp. 5024–5037, 2012.
- [19] V. Šmídl and A. Quinn, *The variational Bayes method in signal processing*. Springer Science & Business Media, 2006.
- [20] X. Yu, J. Li, and J. Xu, "Nonlinear filtering in unknown measurement noise and target tracking system by variational bayesian inference," *Aerospace Science and Technology*, vol. 84, pp. 37–55, 2019.
- [21] K. Li, L. Chang, and B. Hu, "A variational bayesian-based unscented kalman filter with both adaptivity and robustness," *IEEE Sensors Journal*, vol. 16, no. 18, pp. 6966–6976, Sep. 2016.
- [22] P. Dong, Z. Jing, H. Leung, and K. Shen, "Variational bayesian adaptive cubature information filter based on wishart distribution," *IEEE Transactions on Automatic Control*, vol. 62, no. 11, pp. 6051–6057, 2017.
- [23] D. Xu, C. Shen, and F. Shen, "A robust particle filtering algorithm with non-gaussian measurement noise using student-t distribution," *IEEE Signal Processing Letters*, vol. 21, no. 1, pp. 30–34.
- [24] B. Ait-El-Fquih and I. Hoteit, "A variational bayesian multiple particle filtering scheme for large-dimensional systems," *IEEE Transactions on Signal Processing*, vol. 64, no. 20, pp. 5409–5422, 2016.
- [25] X. Yu, G. Jin, and J. Li, "Target tracking algorithm for system with gaussian/non-gaussian multiplicative noise," *IEEE Transactions on Vehicular Technology*, vol. 69, no. 1, pp. 90–100, Jan 2020.
- [26] X. Song, J. H. Park, and X. Yan, "Linear estimation for measurement-delay systems with periodic coefficients and multiplicative noise," *IEEE Transactions on Automatic Control*, vol. 62, no. 8, pp. 4124–4130, 2016.
- [27] W. Liu, "Optimal filtering for discrete-time linear systems with time-correlated multiplicative measurement noises," *IEEE Transactions on Automatic Control*, vol. 61, no. 7, pp. 1972–1978, 2015.
- [28] L. Zhang and X.-D. Zhang, "An optimal filtering algorithm for systems with multiplicative/additive noises," *IEEE Signal Processing Letters*, vol. 14, no. 7, pp. 469–472, 2007.
- [29] E. Gershon, E. Fridman, and U. Shaked, "Predictor-based control of systems with state multiplicative noise," *IEEE Transactions on Automatic Control*, vol. 62, no. 2, pp. 914–920, Feb 2017.
- [30] P. Rajasekaran, N. Satyanarayana, and M. Srinath, "Optimum linear estimation of stochastic signals in the presence of multiplicative noise," *IEEE Transactions on Aerospace & Electronic Systems*, vol. AES-7, no. 3, pp. 462–468, 1971.
- [31] J. Tugnait, "Stability of optimum linear estimators of stochastic signals in white multiplicative noise," *IEEE Transactions on Automatic Control*, vol. 26, no. 3, pp. 757–761, 1981.
- [32] X. Song and J. H. Park, "Linear optimal estimation for discrete-time measurement delay systems with multichannel multiplicative noise," *IEEE Transactions on Circuits and Systems II: Express Briefs*, vol. 64, no. 2, pp. 156–160, 2016.
- [33] X. Yu, J. Li, and J. Xu, "Estimation algorithm for system with non-gaussian multiplicative/additive noises based on variational bayesian inference," *International Journal of Adaptive Control and Signal Processing*, vol. 33, no. 4, pp. 586–608, 2019.
- [34] Ankerst and D. Pauler, "Kendall's advanced theory of statistics, bayesian inference," *Publications of the American Statistical Association*, vol. 100, no. 472, pp. 1465–1466, 2004.
- [35] M. D. Springer and W. Thompson, "The distribution of products of independent random variables," *SIAM Journal on Applied Mathematics*, vol. 14, no. 3, pp. 511–526, 1966.
- [36] —, "The distribution of products of beta, gamma and gaussian random variables," *SIAM Journal on Applied Mathematics*, vol. 18, no. 4, pp. 721–737, 1970.
- [37] P. Bromiley, "Products and convolutions of gaussian probability density functions," *Tina-Vision Memo*, vol. 3, no. 4, p. 1, 2003.
- [38] C. M. Bishop, *Pattern recognition and machine learning*. springer, 2006.
- [39] Y. Huang, Y. Zhang, N. Li, Z. Wu, and J. A. Chambers, "A novel robust student's-t-based kalman filter," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 53, no. 3, pp. 1545–1554, 2017.
- [40] X. Gu and W. Zhu, "Probability-weighted nonlinear stochastic optimal control strategy of quasi-integrable hamiltonian systems with uncertain parameters," *International Journal of Robust and Nonlinear Control*, vol. 25, no. 16, pp. 3086–3098, 2015.
- [41] A. Gelmann, J. B. Carlin, H. S. Stern, and D. B. Rubin, "Bayesian data analysis," vol. 46, no. 3, pp. 363–364, 2004.
- [42] T. Ardeschiri, E. Özkan, U. Orguner, and F. Gustafsson, "Approximate bayesian smoothing with unknown process and measurement noise covariances," *IEEE Signal Processing Letters*, vol. 22, no. 12, pp. 2450–2454, 2015.
- [43] M. S. Arulampalam, S. Maskell, N. Gordon, and T. Clapp, "A tutorial on particle filters for online nonlinear/non-gaussian bayesian tracking," *IEEE Transactions on signal processing*, vol. 50, no. 2, pp. 174–188, 2002.

APPENDIX

Algorithm 1: Proposed VB-based filter with unknown additive and multiplicative noise covariances for non-Gaussian likelihood function

- 1 **Input:** $\hat{\mathbf{x}}_{k-1|k-1}, \mathbf{P}_{k-1|k-1}, \hat{u}_{k-1|k-1}, \hat{\mathbf{U}}_{k-1|k-1}, \mathbf{F}_{k-1}, \mathbf{H}_k, \mathbf{y}_k, \bar{m}_k, \tilde{\mathbf{Q}}_{k-1}, m, n, \tau, \rho, N$
 - 2 **Output:** $\hat{\mathbf{x}}_{k|k}, \mathbf{P}_{k|k}, \gamma_k, \delta_k, \hat{t}_{k|k}, \hat{\mathbf{O}}_{k|k}, \hat{o}_{k|k}, \hat{\mathbf{U}}_{k|k}$
Time update: $\hat{\mathbf{x}}_{k|k-1} = \mathbf{F}_{k-1} \hat{\mathbf{x}}_{k-1|k-1},$
 $\tilde{\mathbf{P}}_{k|k-1} = \mathbf{F}_{k-1} \mathbf{P}_{k-1|k-1} \mathbf{F}_{k-1}^T + \tilde{\mathbf{Q}}_{k-1};$
 - 3 **Variational measurement update:**
 - 4 **Initialization:** $\hat{\mathbf{x}}_{k|k}^{(0)} = \hat{\mathbf{x}}_{k|k-1}, \mathbf{P}_{k|k}^{(0)} = \tilde{\mathbf{P}}_{k|k-1},$
 $\mathbf{E}^{(0)}[(\mathbf{R}_k^M)^{-1}] = (o_k - n - 1) \mathbf{O}_k^{-1},$
 $\hat{t}_{k|k-1} = n + \tau + 1, \hat{\mathbf{O}}_{k|k-1} = \tau \tilde{\mathbf{P}}_{k|k-1},$
 $\hat{u}_{k|k-1} = \rho(\hat{u}_{k-1|k-1} - m - 1) + m + 1,$
 $\hat{\mathbf{U}}_{k|k-1} = \rho \hat{\mathbf{U}}_{k-1|k-1};$
 - 5 **For i=0:N-1**
 - 6 **Update**
 - 7 $q^{(i+1)}(\lambda_k) = G(\lambda_k; \gamma_k^{i+1}, \delta_k^{i+1})$ **given** $q^{(i)}(\mathbf{x}_k) :$
 - 8 $\mathbf{E}_k^{(i)} = (\mathbf{y}_k - \bar{m}_k \mathbf{H}_k \hat{\mathbf{x}}_{k|k}^{(i)})(\mathbf{y}_k - \bar{m}_k \mathbf{H}_k \hat{\mathbf{x}}_{k|k}^{(i)})^T + \mathbf{H}_k \mathbf{P}_{k|k}^{(i)} \mathbf{H}_k^T,$
 - 9 $\gamma_k^{i+1} = 0.5(m + v), \delta_k^{i+1} = 0.5\{v + \text{tr}(\mathbf{E}_k^{(i)} [\mathbf{R}_k^M]^{-1})\},$
 $\mathbf{E}^{(i+1)}[\lambda_k] = \gamma_k^{i+1} / \delta_k^{i+1};$
 - 10 **Update**
 - 11 $q^{(i+1)}(\mathbf{P}_{k|k-1}) = \text{IW}(\mathbf{P}_{k|k-1}; \hat{o}_k^{(i+1)}, \hat{\mathbf{O}}_k^{(i+1)})$ **given** $q^{(i)}(\mathbf{x}_k) :$
 $\mathbf{A}_k^{(i)} = \mathbf{P}_{k|k}^{(i)} + (\hat{\mathbf{x}}_{k|k}^{(i)} - \hat{\mathbf{x}}_{k|k-1})(\hat{\mathbf{x}}_{k|k}^{(i)} - \hat{\mathbf{x}}_{k|k-1})^T,$
 - 12 $\hat{o}_k^{(i+1)} = \hat{o}_{k|k-1} + 1, \hat{\mathbf{O}}_k^{(i+1)} = \mathbf{A}_k^{(i)} + \hat{\mathbf{O}}_{k|k-1};$
 - 13 **Update**
 - 14 $q^{(i+1)}(\mathbf{R}_k^M) = \text{IW}(\mathbf{R}_k^M; \hat{u}_k^{(i+1)}, \hat{\mathbf{U}}_k^{(i+1)})$ **given** $q^{(i)}(\mathbf{x}_k)$
and $q^{(i+1)}(\lambda_k) :$
 $\mathbf{B}_k^{(i)} = (\mathbf{y}_k - \bar{m}_k \mathbf{H}_k \hat{\mathbf{x}}_{k|k}^{(i)})(\mathbf{y}_k - \bar{m}_k \mathbf{H}_k \hat{\mathbf{x}}_{k|k}^{(i)})^T + (\bar{m}_k)^2 \mathbf{H}_k \mathbf{P}_{k|k}^{(i)} \mathbf{H}_k^T,$
 $\hat{u}_k^{(i+1)} = \hat{u}_{k|k-1} + 1,$
 $\hat{\mathbf{U}}_k^{(i+1)} = \mathbf{B}_k^{(i)} \mathbf{E}^{(i+1)}[\lambda_k] + \hat{\mathbf{U}}_{k|k-1}$
 - 15 **Update** $q^{(i+1)}(\mathbf{x}_k) = \text{N}(\mathbf{x}_k; \hat{\mathbf{x}}_{k|k}^{(i+1)}, \mathbf{P}_{k|k}^{(i+1)})$ **given**
 $q^{(i+1)}(\lambda_k), q^{(i+1)}(\mathbf{P}_{k|k-1})$ **and** $q^{(i+1)}(\mathbf{R}_k^M) :$
 - 16 $\mathbf{E}^{(i+1)}[(\mathbf{R}_k^M)^{-1}] = (\hat{u}_k^{(i+1)} - m - 1)(\hat{\mathbf{U}}_k^{(i+1)})^{-1}$
 - 17 $\mathbf{E}^{(i+1)}[\mathbf{P}_{k|k-1}^{-1}] = (\hat{o}_k^{(i+1)} - n - 1)(\hat{\mathbf{O}}_k^{(i+1)})^{-1}$
 - 18 $\tilde{\mathbf{P}}_{k|k-1}^{(i+1)} = \{\mathbf{E}^{(i+1)}[\mathbf{P}_{k|k-1}^{-1}]\}^{-1}$
 - 19 $(\tilde{\mathbf{R}}_k^M)^{(i+1)} = \{\mathbf{E}^{(i+1)}[(\mathbf{R}_k^M)^{-1}]\}^{-1} / \mathbf{E}^{(i+1)}[\lambda_k]$
 - 20 $\mathbf{K}_k^{(i+1)} = \bar{m}_k \tilde{\mathbf{P}}_{k|k-1}^{(i+1)} \mathbf{H}_k^T [(\bar{m}_k)^2 \mathbf{H}_k \tilde{\mathbf{P}}_{k|k-1}^{(i+1)} \mathbf{H}_k^T + (\tilde{\mathbf{R}}_k^M)^{(i+1)}]^{-1}$
 - 21 $\hat{\mathbf{x}}_{k|k}^{(i+1)} = \hat{\mathbf{x}}_{k|k-1} + \mathbf{K}_k^{(i+1)} (\mathbf{y}_k - \bar{m}_k \mathbf{H}_k \hat{\mathbf{x}}_{k|k-1})$
 - 22 $\mathbf{P}_{k|k}^{(i+1)} = \tilde{\mathbf{P}}_{k|k-1}^{(i+1)} - \bar{m}_k \mathbf{K}_k^{(i+1)} \mathbf{H}_k \tilde{\mathbf{P}}_{k|k-1}^{(i+1)}$
 - 23 **End for**
 - 24 $\hat{\mathbf{x}}_{k|k} = \hat{\mathbf{x}}_{k|k}^{(N)}, \mathbf{P}_{k|k} = \mathbf{P}_{k|k}^{(N)}, \gamma_k = \gamma_k^N, \delta_k = \delta_k^N,$
 $\hat{o}_{k|k} = \hat{o}_k^{(N)}, \hat{\mathbf{O}}_{k|k} = \hat{\mathbf{O}}_k^{(N)}, \hat{u}_{k|k} = \hat{u}_k^{(N)}, \hat{\mathbf{U}}_{k|k} = \hat{\mathbf{U}}_k^{(N)}$
-

Algorithm 2: Proposed VB-based filter with unknown additive and multiplicative noise covariances for Gaussian likelihood function

- 1 **Input:** $\hat{\mathbf{x}}_{k-1|k-1}, \mathbf{P}_{k-1|k-1}, \hat{u}_{k-1|k-1}, \hat{\mathbf{U}}_{k-1|k-1}, \mathbf{F}_{k-1}, \mathbf{H}_k, \mathbf{y}_k, \bar{m}_k, \tilde{\mathbf{Q}}_{k-1}, m, n, \tau, \rho, N$
 - 2 **Output:** $\hat{\mathbf{x}}_{k|k}, \mathbf{P}_{k|k}, \hat{o}_{k|k}, \hat{\mathbf{O}}_{k|k}, \hat{u}_{k|k}, \hat{\mathbf{U}}_{k|k}$
 - 3 **Time update:** $\hat{\mathbf{x}}_{k|k-1} = \mathbf{F}_{k-1} \hat{\mathbf{x}}_{k-1|k-1},$
 $\tilde{\mathbf{P}}_{k|k-1} = \mathbf{F}_{k-1} \mathbf{P}_{k-1|k-1} \mathbf{F}_{k-1}^T + \tilde{\mathbf{Q}}_{k-1}$
 - 4 **Variational measurement update:**
 - 5 **Initialization:** $\hat{\mathbf{x}}_{k|k}^{(0)} = \hat{\mathbf{x}}_{k|k-1}, \mathbf{P}_{k|k}^{(0)} = \tilde{\mathbf{P}}_{k|k-1},$
 $\hat{o}_{k|k-1} = n + \tau + 1, \hat{\mathbf{O}}_{k|k-1} = \tau \tilde{\mathbf{P}}_{k|k-1},$
 $\hat{u}_{k|k-1} = \rho(\hat{u}_{k-1|k-1} - m - 1) + m + 1,$
 $\hat{\mathbf{U}}_{k|k-1} = \rho \hat{\mathbf{U}}_{k-1|k-1}$
 - 6 **For i=0:N-1**
 - 7 **Update** $q^{(i+1)}(\mathbf{P}_{k|k-1}) = \text{IW}(\mathbf{P}_{k|k-1}; \hat{o}_k^{(i+1)}, \hat{\mathbf{O}}_k^{(i+1)})$
given $q^{(i)}(\mathbf{x}_k) :$
 - 8 $\mathbf{A}_k^{(i)} = \mathbf{P}_{k|k}^{(i)} + (\hat{\mathbf{x}}_{k|k}^{(i)} - \hat{\mathbf{x}}_{k|k-1})(\hat{\mathbf{x}}_{k|k}^{(i)} - \hat{\mathbf{x}}_{k|k-1})^T$
 - 9 $\hat{o}_k^{(i+1)} = \hat{o}_{k|k-1} + 1, \hat{\mathbf{O}}_k^{(i+1)} = \mathbf{A}_k^{(i)} + \hat{\mathbf{O}}_{k|k-1}$
 - 10 **Update**
 $q^{(i+1)}(\mathbf{R}_k^M) = \text{IW}(\mathbf{R}_k^M; \hat{u}_k^{(i+1)}, \hat{\mathbf{U}}_k^{(i+1)})$ **given** $q^{(i)}(\mathbf{x}_k) :$
 $\mathbf{B}_k^{(i)} = (\mathbf{y}_k - \bar{m}_k \mathbf{H}_k \hat{\mathbf{x}}_{k|k}^{(i)})(\mathbf{y}_k - \bar{m}_k \mathbf{H}_k \hat{\mathbf{x}}_{k|k}^{(i)})^T + (\bar{m}_k)^2 \mathbf{H}_k \mathbf{P}_{k|k}^{(i)} \mathbf{H}_k^T$
 - 11 $\hat{u}_k^{(i+1)} = \hat{u}_{k|k-1} + 1, \hat{\mathbf{U}}_k^{(i+1)} = \mathbf{B}_k^{(i)} + \hat{\mathbf{U}}_{k|k-1}$
 - 12 **Update** $q^{(i+1)}(\mathbf{x}_k) = \text{N}(\mathbf{x}_k; \hat{\mathbf{x}}_{k|k}^{(i+1)}, \mathbf{P}_{k|k}^{(i+1)})$ **given**
 $q^{(i+1)}(\mathbf{P}_{k|k-1})$ **and** $q^{(i+1)}(\mathbf{R}_k^M) :$
 - 13 $\mathbf{E}^{(i+1)}[(\mathbf{R}_k^M)^{-1}] = (\hat{u}_k^{(i+1)} - m - 1)(\hat{\mathbf{U}}_k^{(i+1)})^{-1}$
 - 14 $\mathbf{E}^{(i+1)}[\mathbf{P}_{k|k-1}^{-1}] = (\hat{o}_k^{(i+1)} - n - 1)(\hat{\mathbf{O}}_k^{(i+1)})^{-1}$
 - 15 $\hat{\mathbf{P}}_{k|k-1}^{(i+1)} = \{\mathbf{E}^{(i+1)}[\mathbf{P}_{k|k-1}^{-1}]\}^{-1}$
 - 16 $(\hat{\mathbf{R}}_k^M)^{(i+1)} = \{\mathbf{E}^{(i+1)}[(\mathbf{R}_k^M)^{-1}]\}^{-1}$
 - 17 $\mathbf{K}_k^{(i+1)} = \bar{m}_k \hat{\mathbf{P}}_{k|k-1}^{(i+1)} \mathbf{H}_k^T [(\bar{m}_k)^2 \mathbf{H}_k \hat{\mathbf{P}}_{k|k-1}^{(i+1)} \mathbf{H}_k^T + (\hat{\mathbf{R}}_k^M)^{(i+1)}]^{-1}$
 - 18 $\hat{\mathbf{x}}_{k|k}^{(i+1)} = \hat{\mathbf{x}}_{k|k-1} + \mathbf{K}_k^{(i+1)} (\mathbf{y}_k - \bar{m}_k \mathbf{H}_k \hat{\mathbf{x}}_{k|k-1})$
 - 19 $\mathbf{P}_{k|k}^{(i+1)} = \hat{\mathbf{P}}_{k|k-1}^{(i+1)} - \bar{m}_k \mathbf{K}_k^{(i+1)} \mathbf{H}_k \hat{\mathbf{P}}_{k|k-1}^{(i+1)}$
 - 20 **End For**
 - 21 $\hat{\mathbf{x}}_{k|k} = \hat{\mathbf{x}}_{k|k}^{(N)}, \mathbf{P}_{k|k} = \mathbf{P}_{k|k}^{(N)}, \hat{o}_{k|k} = \hat{o}_k^{(N)}, \hat{\mathbf{O}}_{k|k} = \hat{\mathbf{O}}_k^{(N)}$
 $\hat{u}_{k|k} = \hat{u}_k^{(N)}, \hat{\mathbf{U}}_{k|k} = \hat{\mathbf{U}}_k^{(N)}$
-