2024 American Control Conference (ACC)
July 8-12, 2024. Toronto, Canada

A novel variational Bayesian adaptive Kalman filter for systems with unknown state-dependent noise covariance matrices

Nahid Uzzaman[1] and He Bai[1]

***Abstract*— We consider state estimation for a dynamical system that has unknown state-dependent dynamic process and measurement noise covariance matrices. When the noise covariances are state-dependent the typical Kalman filter (KF) fails to accurately estimate the states of the system. To estimate the states of such a system, we model the covariance matrices via the Wishart process and propose a novel variational Bayesian adaptive Kalman filter (VB-AKF). The proposed VB-AKF combines the variational Bayesian inference of the Wishart process with the KF. The resulting VB-AKF can estimate the states of the system together with the state-dependent dynamic process and measurement noise covariances. Through simulations, we show that the developed VB-AKF is effective and achieves satisfactory performance.**

I. Introduction

State estimation of dynamical systems is crucial for various applications such as autonomous vehicles, robotics, and process control. It allows for real-time tracking of system states based on sensor measurements and is commonly used for real-time decision-making, control, planning, and monitoring system efficiency and safety.

Several approaches and techniques have been developed for state estimation. Recursive Bayesian filters (BRF) use Bayesian probability theory to estimate the state of a system [1], [2]. They update the state estimate recursively as new sensor measurements become available. Recursive least squares estimation involves minimizing the least squares error between the predicted and measured states [3]. Machine learning techniques such as Artificial Neural Networks (ANNs) have also been employed for state estimation when dealing with complex and high-dimensional systems [4]. ANNs can learn to map sensor data to state estimates through training on historical data. Sequential Monte Carlo (SMC), also known as particle filter, combines aspects of Bayesian filtering and Monte Carlo techniques to estimate the states of a system in a sequential manner [5], [6].

The most popular approach for state estimation of a linear system is the Kalman filter (KF) [7]. The KF is a recursive algorithm that estimates the states of a linear system with Gaussian noise. It uses measurements from the sensors and predictions from the dynamic model to provide an optimal solution for the state estimation problem. Extended Kalman Filters (EKF) and Unscented Kalman Filters (UKF) extend this approach to nonlinear systems [8], [9].

[1]Mechanical and Aerospace Engineering, Oklahoma State University, Stillwater, OK, 74078 {nahid.uzzaman, he.bai}@okstate.edu. The work was partly supported by the National Science Foundation (NSF) under Grant No. 1925147.

The KF and its derivatives consider that the system has process noise as $w \sim \mathcal{N}(0, Q)$ and measurement noise $v \sim \mathcal{N}(0, R)$. The process noise covariance matrix Q and measurement noise covariance matrix R are assumed known and independent of the states. However, there are cases where these covariance matrices depend on the states, i.e, $Q_k = Q(x_k)$ and $R_k = R(x_k)$, where x_k is the system state at time step k. In the simulation section, we will present an example with a state-dependent sensor noise covariance.

Several previous works in the literature address the problem of state estimation of a system with state-dependent noise. Reference [10] presents an extended Kalman filter algorithm that can estimate the states when the observation noise covariance is state-dependent and a known function of the state. They show a target-localization example where the sensor bearing has a state-dependent noise. In [11], the authors present an extended Kalman smoothing framework with a generalized Gauss-Newton inference for systems with state-dependent noise covariance. They assume that both the process and measurement noise covariance matrices are known functions of the states. Reference [12] proposes a state-dependent sensor measurement model (SDSMM) that learns the expected measurements. The learned SDSMM is used in the EKF to solve a robot localization problem. The authors of [12] also present a new learning method in [13] where the SDSMM is learned from limited data and used in the Extended Kalman Particle Filter.

Although the above references estimate the states of the system while considering state-dependent noise covariance, they assume that the noise covariance matrices are known functions of the states. Although in [12] and [13], the learned SDSMMs are used to calculate the measurement noise covariance which is used in the EKF, they do not provide an explicit estimate for this covariance. Moreover, they do not provide an estimation for the unknown state-dependent process noise covariance.

To address this issue, we propose a variational Bayesian Adaptive Kalman filter (VB-AKF) that estimates the states of a dynamical system along with the state-dependent process and measurement noise covariance matrices. Our proposed VB-AKF combines the Kalman filtering with the Wishart process [14] and leverages the variational Bayesian inference to estimate state-dependent covariance matrices. The variational Wishart process (VWP) is a powerful tool to estimate the covariance matrices that are dependent on the inputs of the data [15], [16]. For our algorithm, the inputs are the system states.

The contribution of the paper includes 1) addressing

Authorized licensed use limited to: He Bai. Downloaded on December 05,2024 at 22:47:18 UTC from IEEE Xplore. Restrictions apply.

the problem of estimating the states along with the state-dependent process and measurement noise covariance matrices; 2) introducing a novel variational Bayesian Adaptive Kalman filter (VB-AKF); and 3) demonstrating the effectiveness of the proposed filter to estimate the state-dependent sensor noise covariance in simulations.

The rest of the paper is organized as follows. In Section II, the problem formulation is presented. The variational Wishart process (VWP) is reviewed in Section III. Section IV presents the developed VB-AKF algorithm. Numerical simulation results are provided in Section V. Section VI presents the concluding remarks.

II. Problem formulation

Consider the dynamical system

$$x_{k+1} = Ax_k + Bu_k + w_k \tag{1}$$

$$y_k = Cx_k + v_k \tag{2}$$

where w_k and v_k are state-dependent process and measurement noises satisfying $w_k \sim \mathcal{N}(0, Q(x_k))$ and $v_k \sim \mathcal{N}(0, R(x_k))$. Here, $\mathcal{N}(a, b)$ represents a normal distribution with mean a and covariance b. A typical approach for estimating the states when $Q(x_k)$ and $R(x_k)$ are known is using the KF [7]. Given the measurements y_k, the KF estimates the states x_k, assuming that $Q(x_k)$ and $R(x_k)$ are state-independent and available. Without the knowledge of $Q(x_k)$ and $R(x_k)$, the state estimates from the KF may not be accurate.

In this paper, we consider that $Q(x_k)$ and $R(x_k)$ are state-dependent and unknown. Our objective is to estimate the state-dependent process and measurement noise covariance matrices along with the states of the system, based on the measurements y_k and system matrices (A, B, C). We make use of the Wishart process to model the unknown noise covariance matrices and propose a variational Bayesian approach to the inference of the covariance matrices. Using the inferred covariance, we apply a Kalman filter to estimate the system state. We next briefly review the variational Wishart process in Section III and present our adaptive Kalman filter in Section IV.

III. Review of Variational Wishart Process (VWP)

Positive-definite covariance matrices can be estimated using the variational Wishart process (VWP) [15]. Let $Y_n \in \mathbb{R}^D$ represent a series of measurements at corresponding input locations $X_n \in \mathbb{R}^p$, where $n = 1, \cdots, N$. The conditional likelihood of Y_n is described by the multivariate Gaussian density as the following

$$Y_n | 0, \Sigma_n \sim \mathcal{N}\left(0, \Sigma(X_n)\right). \tag{3}$$

Here, $\Sigma_n = \Sigma(X_n) \in \mathbb{R}^{D\times D}$ is the covariance of the normal distribution at X_n. We define an independent and identically distributed (i.i.d.) collection of Gaussian processes (GPs) [17]:

$$f_{d,k} \sim \mathrm{GP}(0, k(.,.;\theta)), \qquad d \le D, k \le \nu \tag{4}$$

where θ represents the trainable parameters of the kernel function k and $\nu \ge D$ represents the degrees of freedom. Denote $f_{d,k}(X_n)$ by $F_{n,d,k}$, and define $F_n \in \mathbb{R}^{D\times\nu}$ as a matrix whose element at the (d,k)-th position is given by $F_{n,d,k}$. We then construct the Wishart distributed covariance Σ_n as [14]

$$\Sigma_n = A_s F_n F_n^\top A_s^\top, \tag{5}$$

where $A_s \in \mathbb{R}^{D\times D}$ is known as a symmetric scale matrix and $A_s A_s^\top$ is positive definite. The collection of constructed covariance matrices denoted as $\Sigma := (\Sigma_1, \Sigma_2, \ldots)$, is referred to as a Wishart Process. The log conditional likelihood of such a Wishart process is given by

$$\log p(Y_n|F_n) = -\frac{D}{2}\log(2\pi) - \frac{1}{2}\log|A_s F_n F_n^\top A_s^\top| - \frac{1}{2} Y_n^\top (A_s F_n F_n^\top A_s^\top)^{-1} Y_n. \tag{6}$$

For computational efficiency, sparse GP with M inducing points denoted by $Z := (Z_1, Z_2, ..., Z_M)$ is used, where Z and X are in the same space. To obtain sparse approximations of the Gaussian processes we may choose $M \ll N$. Let $U_{m,d,k} := f_{d,k}(Z_m)$ for $m \le M,\ d \le D$, and $k \le \nu$ and define $U_{d,k} := (U_{m,d,k},\ m \le M) \in \mathbb{R}^{M\times D\times\nu}$ and $F_{d,k} := (F_{n,d,k},\ n \le N) \in \mathbb{R}^{N\times D\times\nu}$. The joint distribution is given by

$$p(Y, F, U) = \prod_{n=1}^{N}[p(Y_n|F_n)] \prod_{d=1}^{D}\prod_{k=1}^{\nu}[p(F_{d,k}|U_{d,k})p(U_{d,k})], \tag{7}$$

where

$$p(F_{d,k}|U_{d,k}) = \mathcal{N}(K_{xz}K_{zz}^{-1}U_{d,k}, K_{xx} - K_{xz}K_{zz}^{-1}K_{xz}^\top), \tag{8}$$

$$p(U_{d,k}) = \mathcal{N}(0, K_{zz}), \tag{9}$$

in which $K_{xx} \in \mathbb{R}^{N\times N}$ has (n,n)-th element as $k(X_n, X_n; \theta)$, $K_{xz} \in \mathbb{R}^{N\times M}$ has (n,m)-th element as $k(X_n, Z_m; \theta)$, and $K_{zz} \in \mathbb{R}^{M\times M}$ has (m,m)-th element as $k(Z_m, Z_m; \theta)$.

A variational approximation to the posterior of $U_{d,k}$ is introduced as

$$q(F_{d,k}, U_{d,k}) = p(F_{d,k}|U_{d,k})q(U_{d,k}), \tag{10}$$

where $q(U_{d,k}) \sim \mathcal{N}(\mu_{d,k}, S_{d,k})$, with the variational parameters $\mu_{d,k} \in \mathbb{R}^M$ and a symmetric positive definite matrix $S_{d,k} \in \mathbb{R}^{M\times M}$. The marginal distribution $q(F_{d,k})$ is obtained as

$$q(F_{d,k}) = \int p(F_{d,k}|U_{d,k})q(U_{d,k})dU_{d,k} = \mathcal{N}(\tilde{K}\mu_{d,k}\ ,\ K_{xx} + \tilde{K}(S_{d,k} - K_{zz})\tilde{K}^\top) \tag{11}$$

where $\tilde{K} := K_{xz}K_{zz}^{-1}$. Now, we can lower-bound the measurement likelihood as

$$\log p(Y) \ge \sum_{n=1}^{N} E_{q(F_n)} \log(p(Y_n|F_n)) - \sum_{d=1}^{D}\sum_{k=1}^{\nu} \mathcal{KL}[q(U_{d,k}||p(U_{d,k})]. \tag{12}$$

Authorized licensed use limited to: He Bai. Downloaded on December 05,2024 at 22:47:18 UTC from IEEE Xplore. Restrictions apply.

The inequality (12), known as the evidence lower bound (ELBO) [18], is the objective function in the variational Bayesian inference. Here, $\mathcal{KL}$ is the Kullback-Leibler divergence. The inference of the parameters $\Theta := (Z, A_s, \mu, S, \theta)$ is performed by maximizing the ELBO in (12) via a stochastic gradient ascent algorithm.

IV. A Variational Bayesian KF

We propose a variational Bayesian adaptive Kalman filter (VB-AKF) that estimates the states x_k, process noise $Q(x_k)$ and measurement noise $R(x_k)$ of the dynamical system in (1)-(2) given the measurements $y_k \in \mathbb{R}$. We first model the state-dependent noise covariance matrices with the VWP. Consider two GPs f and g as

$$f \sim \mathcal{N}(0, k_f(.,., \theta_f)) \tag{13}$$
$$g \sim \mathcal{N}(0, k_g(.,., \theta_g)) \tag{14}$$

where θ_f and θ_g denote the parameters of the kernels k_f and k_g, respectively. Let $F_k = f(x_k)$ and $G_k = g(x_k)$. Then, we construct the Wishart distributed process noise covariance $Q(x_k)$ and measurement noise covariance $R(x_k)$ as

$$Q(x_k) = A_q F_k F_k^\top A_q^\top \tag{15}$$
$$R(x_k) = A_r G_k G_k^\top A_r^\top, \tag{16}$$

respectively. Here, A_q and A_r are the scale matrices. Similar to the VWP, we assume M inducing points $Z := (Z_1, Z_2, ..., Z_M)$ and consider $U_m = f(Z_m)$ and $V_m = g(Z_m)$, where $m \leq M$. Denote $U := (U_m,\ m \leq M)$, $V := (V_m,\ m \leq M)$, $F := (F_l,\ l \leq k)$, $G := (G_l,\ l \leq k)$, $x := (x_l,\ l \leq k)$, and $y := (y_l,\ l \leq k)$.

We next develop the joint probability as

$$\begin{aligned} &P(y, x, G, V, F, U) \\ =&P(x_0)P(U)P(V)\prod_{l=1}^{k} P(y_l|x_l, G_l)P(x_l|x_{l-1}, F_{l-1}) \\ &P(F_{l-1}|x_{l-1}, U)P(G_l|x_l, V). \end{aligned} \tag{17}$$

We introduce the variational distribution for $P(U)$, $P(V)$, and $P(x_0)$ as $q(U) \sim \mathcal{N}(\mu^q, S^q)$, $q(V) \sim \mathcal{N}(\mu^r, S^r)$ and $q(x_0) \sim \mathcal{N}(m_{x_0}, P_{x_0})$, respectively, and define

$$\begin{aligned} q(x, G, V, F, U) =&q(x_0)q(U)q(V)\prod_{l=1}^{k} P(x_l|x_{l-1}, F_{l-1}) \\ &P(G_l|x_l, V)P(F_{l-1}|x_{l-1}, U), \end{aligned} \tag{18}$$

which is the variational distribution for the joint distribution $p(x, G, V, F, U)$.

We now derive the ELBO as

$$\begin{aligned} &\log P(y) \geq \\ =&\sum_{l=1}^{k} \mathbb{E}_{q(G_l, x_l)} \log P(y_l|x_l, G_l) + \mathcal{KL}\left[q(U)||P(U)\right] \\ &+ \mathcal{KL}\left[q(V)||P(V)\right] + \mathcal{KL}\left[q(x_0)||P(x_0)\right]. \end{aligned} \tag{19}$$

The first term in (19) is known as the expectation of the log-likelihood of the data. From the Gaussian density, we get the log-likelihood as

$$\begin{aligned} \log P(y_k|x_k, G_k) &= -\frac{D}{2}\log(2\pi) - \frac{1}{2}\log|A_r G_k G_k^\top A_r^\top| \\ &- \frac{1}{2}(y_k - Cx_k)^\top (A_r G_k G_k^\top A_r^\top)^{-1}(y_k - Cx_k). \end{aligned} \tag{20}$$

The second, third, and fourth terms in (19) are the Kullback-Leibler divergences. Although x_0 and U do not appear explicitly in the first term of the ELBO, they are used in the calculation of the states x. The marginal distributions for F and G is obtained from the learned $q(U)$ and $q(V)$ as

$$q(F) = \int p(F|U)q(U)dU \tag{21}$$
$$q(G) = \int p(G|V)q(V)dV. \tag{22}$$

First, using $q(U)$ and x values from the past iterations (or the initial values for the first iteration), F is sampled from (21). Then using (15), $Q(x_k)$ is calculated. The initial state x_0 is sampled from $q(x_0)$. From x_0, states are propagated using (1) which provides x_l with $l = 1, \ldots, k$. Using $q(V)$ and x, G is sampled from (22). Thus, we obtain all the required variables for calculating the ELBO. Since the first term in (19) is an expectation with respect to G_l and x_l, we collect multiple samples of G_l and x_l and approximate the expectation based on the samples.

By maximizing the ELBO, we infer the trainable parameters to perform estimations on state x, covariance matrices $Q(x)$, and $R(x)$. The trainable parameters for the variational inference are $\Theta := (m_{x_0}, P_{x_0}, \mu^q, S^q, \mu^r, S^r, A_q, A_r)$, where $q(x_0) \sim \mathcal{N}(m_{x_0}, P_{x_0})$, $q(U) \sim \mathcal{N}(\mu^q, S^q)$ and $q(V) \sim \mathcal{N}(\mu^r, S^r)$, and A_q and A_r are the scale matrices used in the Wishart processes. From the trained $q(U)$ and $q(V)$ and the corresponding scale matrices A_q and A_r, we obtain the learned $Q(x)$ and $R(x)$ using (15) and (16), respectively. Once $Q(x)$ and $R(x)$ are learned, we estimate the system state using the Kalman filter equations [7] as

$$\hat{x}_{k+1|k} = Ax_k + Bu_k \tag{23}$$
$$\hat{P}_{k+1|k} = AP_kA^\top + A_qF_kF_k^\top A_q^\top \tag{24}$$
$$S_{k+1} = C\hat{P}_{k+1|k}C^\top + A_rG_{k+1}G_{k+1}^\top A_r^\top \tag{25}$$
$$K_k = \hat{P}_{k+1|k}C^\top S_{k+1}^{-1} \tag{26}$$
$$x_{k+1} = \hat{x}_{k+1|k} + K_k\left(y_{k+1} - C\hat{x}_{k+1|k}\right) \tag{27}$$
$$P_{k+1} = (\mathbf{I} - K_kC)\,\hat{P}_{k+1|k}. \tag{28}$$

The proposed VB-AKF algorithm is shown in Algorithm 1.

V. Numerical simulations

A. Target tracking problem

In this section, we apply Algorithm 1 for a target tracking example. We assume that the target is moving in a constant circular motion. In this example, the process has a fixed state-independent covariance while the measurement noise covariance R is state-dependent. The measurements are the range and bearing of the target from the sensor location.

Authorized licensed use limited to: He Bai. Downloaded on December 05,2024 at 22:47:18 UTC from IEEE Xplore. Restrictions apply.

Algorithm 1 Variational Bayesian learning of KF states, covariance $Q(x)$, and $R(x)$

Initialize the trainable parameters Θ
Provide data $\mathcal{D} = \{y_{1:k}\}$
while Number of iterations $\leq$ Maximum iterations **do**
    Sample x_0 from $q(X_0) \sim \mathcal{N}(m_{x_0}, P_{x_0})$
    while $l = 1 : k$ **do**
        Using x_{l-1} and $q(U)$ sample F_{l-1} from (21)
        Sample w_{l-1} from $\mathcal{N}(0, A_q F_{l-1} F_{l-1}^{\top} A_q^{\top})$
        Calculate x_l from $x_l = A x_{l-1} + Bu + w_{l-1}$
    end while
    Using x and $q(V)$ sample G from (22)
    Compute the log-likelihood in (20)
    Compute the ELBO in (19)
    Compute gradients of the ELBO
    Update Θ from gradient-ascent
end while
Sample x_0 from learned $q(X_0)$
while $l = 1 : k$ **do**
    Using x_{l-1} and $q(U)$ sample F_{l-1} from (21)
    Using $x_l = \hat{x}_{k+1|k}$ and $q(V)$ sample G_l from (22)
    Using equations (23)-(28) calculate x_l
end while

Specifically, the bearing measurement has state-dependent and the range has state-independent Gaussian noise.

The dynamics of the target moving in a circular motion are derived as

$$\begin{bmatrix} x_{k+1,1} \\ x_{k+1,2} \\ x_{k+1,3} \\ x_{k+1,4} \end{bmatrix} = \begin{bmatrix} 1 & 0 & \Delta t & 0 \\ 0 & 1 & 0 & \Delta t \\ 0 & 0 & \cos\Delta\theta & -\sin\Delta\theta \\ 0 & 0 & \sin\Delta\theta & \cos\Delta\theta \end{bmatrix} \begin{bmatrix} x_{k,1} \\ x_{k,2} \\ x_{k,3} \\ x_{k,4} \end{bmatrix} + w_k. \tag{29}$$

where $x_{k,1}$ and $x_{k,2}$ are the positions in a 2-D Cartesian plane, $x_{k,3}$ and $x_{k,4}$ are the 2-D velocities, and w_k is a state-independent Gaussian white noise with $w_k \sim \mathcal{N}(0, Q)$. In this example, we assume that Q is known and state-independent and focus on estimating $R(x)$.

The measurements are range y_1 and bearing y_2 given by

$$y_{k,1} = \sqrt{x_{k,1}^2 + x_{k,2}^2} + \epsilon_{k,1} \tag{30}$$

$$y_{k,2} = \arctan(x_{k,2}/x_{k,1}) + \epsilon_{k,2} \tag{31}$$

where $\epsilon_{k,1}$ and $\epsilon_{k,2}$ are sensor noises. $\epsilon_{k,1} \sim \mathcal{N}(0, r_1)$ is state-independent but $\epsilon_{k,2} \sim \mathcal{N}(0, r_2(x_k))$ is state-dependent. We use the following variance model for $r_2(x_k)$ [10]:

$$r_2(x) = \sigma_R(x) = K\frac{g(y_1)}{\cos^2(y_2)} = K\frac{g\left(\sqrt{x_{k,1}^2 + x_{k,2}^2}\right)}{\cos^2(\arctan(x_{k,2}/x_{k,1}))}, \tag{32}$$

where

$$g(y) = a_0 + a_1(a_2 - y)^2, \tag{33}$$

K is a constant and a_0, a_1, a_2 are some scalar parameters.

Converting the bearing-range measurements at time step k into the Cartesian coordinates yields

$$y_k = \begin{pmatrix} y_{k,1}\cos(y_{k,2}) & y_{k,1}\sin(y_{k,2}) \end{pmatrix}^T \tag{34}$$

$$= \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{pmatrix} x_k + \bar{\epsilon}_k \tag{35}$$

where $\bar{\epsilon}_k \in \mathbb{R}^2$ is a state-dependent noise. If r_1 and $r_2(x_k)$ are available, the covariance for $\bar{\epsilon}_k$ is approximated as

$$\mathrm{cov}(\bar{\epsilon}_k) = W \begin{pmatrix} r_1 & 0 \\ 0 & r_2(x_k) \end{pmatrix} W^T \tag{36}$$

where

$$W = \begin{pmatrix} \cos(y_{k,2}) & -y_{k,1}\sin(y_{k,2}) \\ \sin(y_{k,2}) & y_{k,1}\cos(y_{k,2}) \end{pmatrix}. \tag{37}$$

In our simulations, we assume no knowledge of r_1 and $r_2(x)$ and apply Algorithm 1 to estimate the covariance matrix for $\bar{\epsilon}_k$.

B. Simulation results

We first use the data for a target moving in a circular motion at a certain location to learn $R(x_k)$ and then test the learned $R(x_k)$ to estimate the states of the target when it is circling at a different location.

For training, the target is moving in a circular motion with the origin of the circle at $(4, 4)$ meters and the radius of the circle being 2 meters. The sensor is located at the origin of the 2-D plane, i.e., the sensor's location is $(0, 0)$ meters. The distribution for the initial state $q(x_0)$ is initialized as $q(x_0) \sim \mathcal{N}(y_1^1, y_1^2, 4, 4)$, where y_1^1, y_1^2 are the first measurement of the target position in the 2-D Cartesian plane. The prior for the initial state's distribution is chosen as $P(x_0) \sim \mathcal{N}(y_1^1, y_1^2, 0, 0)$. The true initial states of the target are $x_0 = (6, 4, 2, 2)$. For the kernel in (14) we use a squared exponential (SE) function. We set $K = 0.0001$ and $a_0 = a_1 = a_2 = 1$. To calculate the numerical expectation in (19), we use 5 samples of x_l and 10 samples of G_l. A learning rate of 0.01 is used with 20000 iterations. All the training data is utilized in each iteration. We use the following Q for our simulations

$$Q = \begin{pmatrix} p^2 & 0 & 0 & 0 \\ 0 & p^2 & 0 & 0 \\ 0 & 0 & p & 0 \\ 0 & 0 & 0 & p \end{pmatrix} \tag{38}$$

where $p = 0.001$. The data generation process for training and test data is shown in Algorithm 2.

For test, the target is moving in a circular motion with the center of the circle at $(5, 5)$ meters and the radius of the circle being 2 meter. The algorithm uses the previously learned $R(x)$ to perform state estimation from the new measurements.

We compare the state estimation performance of our developed VB-AKF with two other KFs. First, we consider the KF with the true measurement noise covariance (KF-TMC). In KF-TMC, the true covariance from (36) is used in the standard Kalman filter equations. Second, we consider

Authorized licensed use limited to: He Bai. Downloaded on December 05,2024 at 22:47:18 UTC from IEEE Xplore. Restrictions apply.

Algorithm 2 Data generation for target tracking problem

while $l = 1 : k$ **do**
  Sample w_{l-1} from $\mathcal{N}(0, Q))$
  Calculate x_l from (1)
  Sample $\epsilon_{1,l} \sim \mathcal{N}(0, 1)$
  Calculate $r_2(x_l) = \sigma_R(x_l)$ from (32)
  Sample $\epsilon_{2,l}$ from $\mathcal{N}(0, r_2(x_l))$
  Calculate $y_l = \begin{bmatrix} y_{1,l} \\ y_{2,l} \end{bmatrix} = \begin{bmatrix} \sqrt{x_{k,1}^2 + x_{k,2}^2} + \epsilon_{1,l} \\ \arctan(x_{k,2}/x_{k,1}) + \epsilon_{2,l} \end{bmatrix}$
  Convert $y_{1,l}, y_{2,l}$ into Cartesian position measurements $y_l = \begin{bmatrix} y_{1,l}\cos(y_{2,l}) \\ y_{1,l}\sin(y_{2,l}) \end{bmatrix}$
end while
Provide data $\mathcal{D} = \{y_{1:k}\}$

a nominal KF that also uses (36) as the measurement noise covariance. However, in the nominal KF the $r_2(x_k)$ has a fixed value, i.e. $r_2(x_k) = r_2$. We first choose $r_2 = 0.007$ for the nominal KF as this is the average value of the varying $r_2(x_k)$ for the test data. Later we will show how changing this value affects the nominal KF.

The 2-D circular motion of the target at test locations estimated by the KF-TMC, VB-AKF, and the nominal KF (with $r_2 = 0.007$) are shown in Figure 1. The individual state estimations are shown in Figure 2. Figure 3 shows the root mean square error (RMSE) of the estimated states by the three filters from 100 Monte-Carlo simulations. From the figures, we observe that the state estimation by the proposed VB-AKF is comparable to that of the KF-TMC and the nominal KF. However, for the nominal KF we assume that it has access to the covariance model in (36) with $r_2(x_k)$ being a fixed value. Moreover, we have provided the nominal KF with the average of the true $r_2(x_k)$ values.

We now change $r_2 = 0.007$ to $r_2 = 0.5$. The RMSEs for state estimation on the test data are shown in Figure 5. We observe that the RMSE for the nominal KF is worse than before. This shows that the nominal KF is sensitive to the choice of r_2. Also, in practice, the true covariance model (36) may not be available for the nominal KF, which will further degrade its performance. In contrast, our developed VB-AKF does not require any knowledge of the true covariance. It learns the covariance from raw measurements and performs state estimation based on that. Figure 4 shows the covariance values estimated by the VB-AKF on the test data locations.

VI. Conclusions

We present a VB-AKF algorithm that performs state estimation of a dynamical system when the system has an unknown state-dependent process and measurement noise. The presented VB-AKF combines the traditional Kalman filter with variational Bayesian inference of the Wishart process. From the measurements, the algorithm estimates the states together with the state-dependent process and measurement noise covariance matrices. We validate the proposed algorithm and demonstrate its effectiveness using a target-tracking simulation example. Ongoing work is focused on employing Kalman smoothing in the VWP learning to further improve the performance of the VB-AKF.

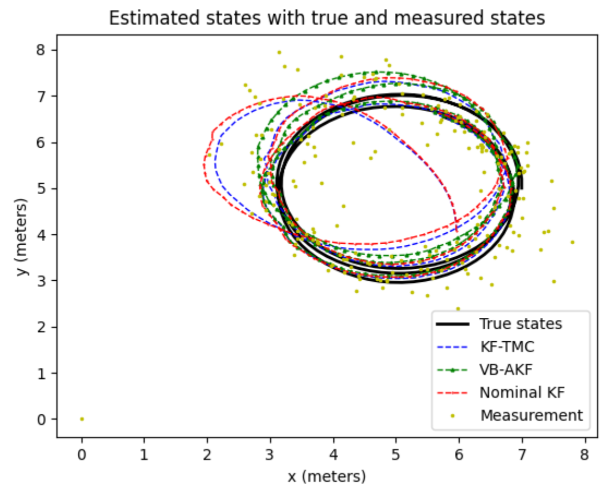

Fig. 1. Circular motion of the target in the 2-D plane for test data locations.

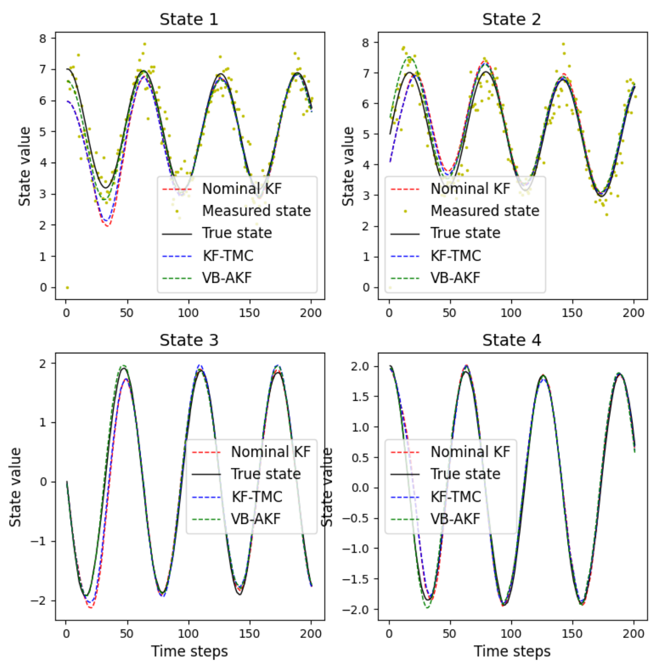

Fig. 2. States of the target estimated by the three filters for test data locations.

References

[1] M. Pelka and H. Hellbrück, "Introduction, discussion and evaluation of recursive bayesian filters for linear and nonlinear filtering problems in indoor localization," in *2016 International Conference on Indoor Positioning and Indoor Navigation (IPIN)*, 2016, pp. 1–8.
[2] G. Kurz, I. Gilitschenski, and U. D. Hanebeck, "Recursive bayesian filtering in circular state spaces," *IEEE Aerospace and Electronic Systems Magazine*, vol. 31, no. 3, pp. 70–87, 2016.
[3] M. B. Malik, "State-space recursive least-squares: Part i," *Signal processing*, vol. 84, no. 9, pp. 1709–1718, 2004.
[4] J. Wilson and L. Zorzetto, "A generalised approach to process state estimation using hybrid artificial neural network/mechanistic models," *Computers & chemical engineering*, vol. 21, no. 9, pp. 951–963, 1997.

Authorized licensed use limited to: He Bai. Downloaded on December 05,2024 at 22:47:18 UTC from IEEE Xplore. Restrictions apply.

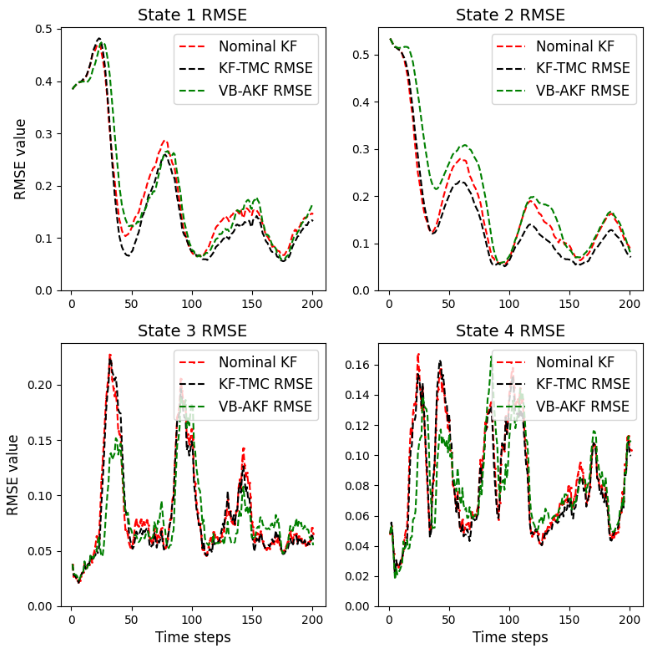

Fig. 3. State estimation RMSE. Nominal KF's r_2 is average of the true $r_2(x_k)$.

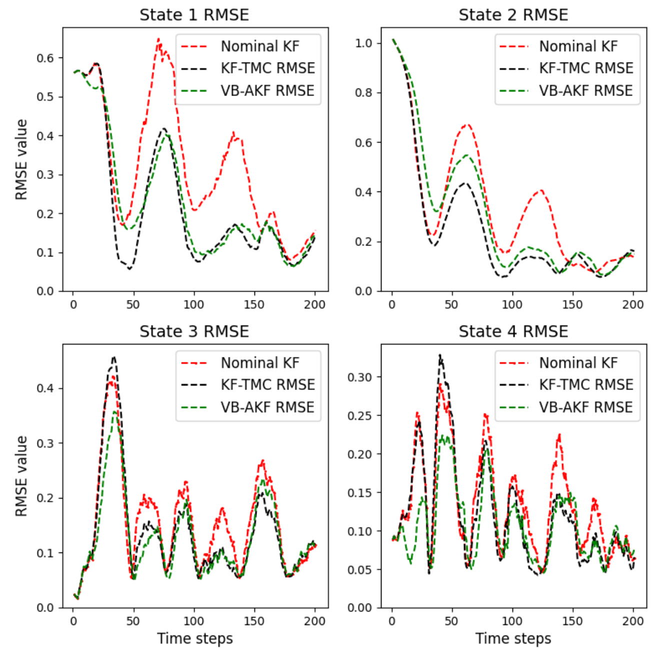

Fig. 5. State estimation RMSE. Nominal KF's $r_2 = 0.5$.

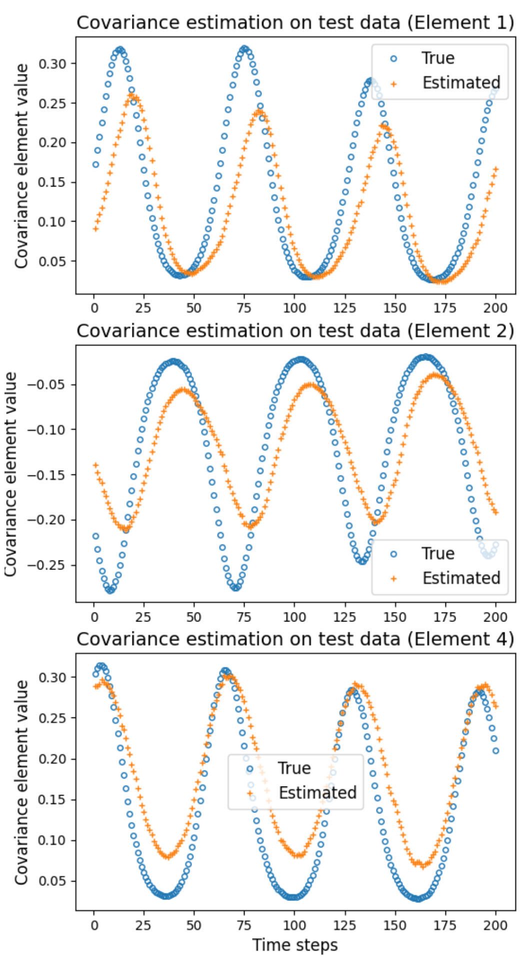

Fig. 4. Measurement noise covariance estimation on the test data locations.

[5] A. Doucet, N. De Freitas, and N. Gordon, "An introduction to sequential monte carlo methods," *Sequential Monte Carlo methods in practice*, pp. 3–14, 2001.

[6] O. Cappé, S. J. Godsill, and E. Moulines, "An overview of existing methods and recent advances in sequential monte carlo," *Proceedings of the IEEE*, vol. 95, no. 5, pp. 899–924, 2007.

[7] S. Särkkä and L. Svensson, *Bayesian filtering and smoothing*. Cambridge university press, 2023, vol. 17.

[8] M. I. Ribeiro, "Kalman and extended kalman filters: Concept, derivation and properties," *Institute for Systems and Robotics*, vol. 43, no. 46, pp. 3736–3741, 2004.

[9] E. A. Wan and R. Van Der Merwe, "The unscented kalman filter," *Kalman filtering and neural networks*, pp. 221–280, 2001.

[10] D. Spinello and D. J. Stilwell, "Nonlinear estimation with state-dependent gaussian observation noise," *IEEE Transactions on Automatic Control*, vol. 55, no. 6, pp. 1358–1366, 2010.

[11] A. Y. Aravkin and J. V. Burke, "Smoothing dynamic systems with state-dependent covariance matrices," in *53rd IEEE Conference on Decision and Control*. IEEE, 2014, pp. 3382–3387.

[12] T. Williams and Y. Sun, "Learning state-dependent, sensor measurement models for localization," in *2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2019, pp. 3090–3097.

[13] ——, "Learning state-dependent sensor measurement models with limited sensor measurements," in *2021 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2021, pp. 86–93.

[14] A. G. Wilson and Z. Ghahramani, "Generalised wishart processes," in *Proceedings of the Twenty-Seventh Conference on Uncertainty in Artificial Intelligence*, 2011, pp. 736–744.

[15] C. Heaukulani and M. van der Wilk, "Scalable bayesian dynamic covariance modeling with variational wishart and inverse wishart processes," *Advances in Neural Information Processing Systems*, vol. 32, 2019.

[16] N. Uzzaman and H. Bai, "Dynamic covariance prediction using variational wishart processes with uncertain inputs," in *2023 American Control Conference (ACC)*. IEEE, 2023, pp. 125–130.

[17] C. K. Williams and C. E. Rasmussen, *Gaussian processes for machine learning*. MIT press Cambridge, MA, 2006, vol. 2, no. 3.

[18] M. D. Hoffman and M. J. Johnson, "ELBO surgery: yet another way to carve up the variational evidence lower bound," in *Workshop in Advances in Approximate Bayesian Inference, NIPS*, vol. 1, no. 2, 2016.

Authorized licensed use limited to: He Bai. Downloaded on December 05,2024 at 22:47:18 UTC from IEEE Xplore. Restrictions apply.