Digit. Signal Process. 153 (2024) 104648

Contents lists available at ScienceDirect

Digital Signal Processing

journal homepage: www.elsevier.com/locate/dsp

# Passive underwater tracking with unknown measurement noise statistics using variational Bayesian approximation

Shreya Das [a,*], Kundan Kumar [b], Shovan Bhaumik [a]

[a] Department of Electrical Engineering, Indian Institute of Technology Patna, 801106, Bihar, India
[b] Department of Electrical Engineering and Automation, Aalto University, Espoo, 02150, Finland

A R T I C L E I N F O

*Keywords:*
Bearings-only tracking
Target motion analysis
Gaussian filters
Adaptive filtering
Normal inverse Wishart
Variational Bayesian approximation

A B S T R A C T

This paper considers a bearings-only tracking problem with unknown measurement noise statistics. It is assumed that the measurement noise follows a Gaussian probability density function where the mean and the covariance of the noise are unknown. Here, an adaptive nonlinear filtering technique is proposed, where the joint distribution of the measurement noise mean and its covariance are considered to follow a normal inverse Wishart (NIW) distribution. Using the variational Bayesian (VB) approximation, joint distribution of the target state, the measurement noise mean and covariance is factorized as the product of their individual probability density function (pdf). Minimizing the Kullback-Leibler divergence (KLD) between the factorized and true joint pdfs, probability distributions of the noise mean, covariance and the target states are evaluated. The estimation of states with the proposed VB based method is compared with the maximum a posteriori (MAP) and the maximum likelihood estimation (MLE) based adaptive filtering. Deterministic sigma points are used to realize the filtering algorithms. The proposed adaptive filter with VB approximation is found to be more accurate compared to their corresponding MAP-MLE based counterparts.

## 1. Introduction

Bearings-only tracking (BOT) involves tracking a target using only noisy line of sight (LOS) measurements, commonly referred to as the bearing angle measurement [1,2]. As only bearing angle measurements are used, the measuring device, sonar, works without emitting any signal. Such tracking is known as passive tracking and the reason for it is not to disclose ownship position to the enemy during war [3]. Further, when the measurement is angle, the measurement equation becomes nonlinear. Under such a restrictive sensing environment and nonlinear measurement relation, achieving a good performance becomes a paramount challenge [4]. So, the BOT attracts researchers over many decades [5–10] and still an optimal solution is far to achieve.

The angle measurements are corrupted with sensor noise which was mostly assumed as Gaussian with zero mean and known covariance in previous literature. With such assumption, initially, the extended Kalman filter (EKF) [11,12] and its variants [13–15] which linearize the measurement equation with a first-order Taylor series approximation, are implemented for target motion analysis (TMA). However, the EKF exhibits limitations such as poor estimation accuracy and it is prone to high track divergence [16,17]. To enhance estimation accuracy, several other nonlinear filtering algorithms have been developed. These include the cubature Kalman filter (CKF) [2], the unscented Kalman filter (UKF) [18], the Gauss-Hermite filter (GHF) [19] and their respective variants [20–22]. These filters approximate the probability density functions (pdfs) of the state using deterministic sample points and their associated weights, so they are commonly referred to as deterministic sample point filters [23].

Gaussian filters perform well when we have prior knowledge of the measurement noise statistics, including its mean and covariance. However, in many practical applications, the statistics of measurement noise are often unknown. In such cases, the traditional filters mentioned above cannot be implemented. In order to estimate states, the noise statistics need to be estimated first. This kind of estimation is known as adaptive estimation [24,25]. Various adaptive estimation techniques have been proposed in the literature. One of them is to employ the maximum a posteriori (MAP) method [26,27]. This method estimates both the mean and the covariance of the process and measurement noise, in addition to the target's state [26,27]. However, this technique may not always guarantee positive definiteness of the estimated noise covariance [27,28] and sometimes does not converge to the true covariance value

* Corresponding author.
*E-mail address:* shreya_2121ee15@iitp.ac.in (S. Das).

https://doi.org/10.1016/j.dsp.2024.104648

Available online 18 June 2024

[28,29]. Another adaptive filtering technique is based on maximum likelihood estimation (MLE) [30–32] which estimates the measurement noise covariance alongside the target state. There are two variations of MLE based adaptive filtering: (i) using an innovation based sequence [33] (ii) using a residual based sequence [34]. The residual and the innovation based adaptive UKF were proposed in [25] which was analogous to the adaptive Kalman filter in [34]. The positive definiteness of the estimated noise covariance matrix is guaranteed in the residual adaptive method based on MLE [34]. Nonetheless, the MLE based adaptive filtering required large data windows to yield reliable estimates of the noise covariance matrix. This limitation renders it unfeasible for situations involving time varying noise covariance matrix [35]. The application of adaptive filtering technique on underwater BOT problem is quite limited except a few notable publications [8,36,37].

In recent years, the variational Bayesian (VB) approach has been studied extensively to estimate the unknown noise parameters for a linear system [35,38–42]. In [38], only the measurement noise covariance and in [35,39] both the process and the measurement noise covariance matrices are considered to be unknown. The noise covariance matrices are assumed to be following inverse Wishart distribution and are jointly estimated with the state vector. Instead of using the fixed point iteration as in [35,38], an approximation of backward smoothing of the posterior distribution of a sliding window (SW) state vectors is proposed in [39]. In [40], the measurement noise is estimated along with state assuming the joint distribution follows Gaussian generalized inverse Gaussian distribution (GGIG). In [41,42], the mean and covariance of both process and measurement noise are considered to be unknown. Robust VB based adaptive filters are also formulated to handle high initialization uncertainty [43,44], and noise mismatch [45–48].

For nonlinear system, the VB based adaptive UKF [49], adaptive CKF [50,51] and adaptive interpolatory CKF [44,52], and adaptive cubature information filter (CIF) [53] are developed in the literature. But in all these works, the noises mean are considered to be known and zero. In this paper, we develop an adaptive estimation method for a nonlinear system whose measurement noise mean and covariance are unknown. We applied our method for an underwater bearings-only tracking problem where both the measurement noise mean and covariance are considered to be unknown.

To jointly estimate the measurement noise mean and covariance with the target state, their joint probability density function (pdf) is considered to follow the normal inverse Wishart (NIW) distribution. The joint distribution is factorized using VB approximation and the approximated individual distributions are obtained by minimizing the Kullback-Leibler divergence (KLD) between the approximated factored pdfs and the true joint pdf. The estimation is realized with deterministic sample points. We also optimized the likelihood to obtain the appropriate initialization of the tuning parameters, needed during the estimation.

The proposed VB based adaptive estimation is applied to an underwater BOT problem and the results are compared with the MAP-MLE technique in two engagement scenarios in terms of estimation accuracy and execution time. In MAP-MLE method, measurement noise mean is estimated using the MAP method and its covariance is estimated using the residual based MLE method. Estimation accuracy is evaluated in terms of root mean square error (RMSE), bias norm, and average normalized estimation error squared (ANEES), percentage of track divergence, and relative execution time. We have compared the proposed VB technique with MAP-MLE, MAP-VB, MAP-GGIG, and MAP-SW, where in the measurement noise mean is estimated using MAP and the covariance is estimated using MLE [34], VB [35,38], GGIG [40] and SW [39] techniques, respectively. It has been observed that the VB based adaptive filtering technique provides better accuracy than the other adaptive filters.

## 2. Problem formulation

### 2.1. System model

The state vector of the target is assumed to follow a nearly straight line path with nearly constant velocity in the discrete time domain and it is expressed as $\mathcal{X}_k^t = \begin{bmatrix} x_k^t & y_k^t & \dot{x}_k^t & \dot{y}_k^t \end{bmatrix}^T$, where $x_k^t$ and $y_k^t$ are the target's position along $x$ and $y$ axis, respectively, and $\dot{x}_k^t$ and $\dot{y}_k^t$ are the target's velocity along $x$ and $y$ axis, respectively, at the $k$th time instant. Similarly, the observer's state vector can be expressed as $\mathcal{X}_k^o = \begin{bmatrix} x_k^o & y_k^o & \dot{x}_k^o & \dot{y}_k^o \end{bmatrix}^T$, where $x_k^o$ and $y_k^o$ are the observer's position along $x$ and $y$ axis, respectively, and $\dot{x}_k^o$ and $\dot{y}_k^o$ are the observer's velocity along $x$ and $y$ axis, respectively. The dynamic model of the target can be expressed as [1,2]

$$\mathcal{X}_k = F\mathcal{X}_{k-1} + \omega_{k-1} - \mho_{k-1,k}, \tag{1}$$

where $\mathcal{X}_k = \mathcal{X}_k^t - \mathcal{X}_k^o = \begin{bmatrix} x_k & y_k & \dot{x}_k & \dot{y}_k \end{bmatrix}^T$ is the relative state vector of the target from the observer, $F$ is the state transition matrix evaluated as,

$$F = \begin{bmatrix} I_{2\times2} & \Delta I_{2\times2} \\ 0_{2\times2} & I_{2\times2} \end{bmatrix}, \tag{2}$$

where $\Delta$ is the sampling time, $\mho_{k-1,k}$ is a vector of inputs evaluated as,

$$\mho_{k-1,k} = \begin{bmatrix} x_k^o - x_{k-1}^o - \Delta\dot{x}_{k-1}^o \\ y_k^o - y_{k-1}^o - \Delta\dot{y}_{k-1}^o \\ \dot{x}_k^o - \dot{x}_{k-1}^o \\ \dot{y}_k^o - \dot{y}_{k-1}^o \end{bmatrix}, \tag{3}$$

and $\omega_{k-1} \sim \mathcal{N}(0, Q)$ is the process noise which is assumed to be Gaussian with zero mean and covariance, $Q$ given as,

$$Q = \begin{bmatrix} \dfrac{\Delta^3}{3} I_{2\times2} & \dfrac{\Delta^2}{2} I_{2\times2} \\ \dfrac{\Delta^2}{2} I_{2\times2} & \Delta I_{2\times2} \end{bmatrix} \bar{q}, \tag{4}$$

where $\bar{q}$ is the intensity of process noise.

### 2.2. Measurement model

The bearing measurement of the target with respect to the true north is measured using sensors mounted to ownship. The measurement model is expressed as:

$$\mathcal{Y}_k = h(\mathcal{X}_k) + \nu_{\theta_k}, \tag{5}$$

where $h(\mathcal{X}_k) = \tan^{-1}(x_k/y_k)$ and $\nu_{\theta_k}$ is the measurement noise, assumed to be Gaussian, *i.e.*, $\nu_{\theta_k} \sim \mathcal{N}(r', R'_k)$. However, the mean $(r')$ and covariance $(R'_k)$ of the noise $\nu_{\theta_k}$ is unknown. The objective of the work is to estimate $r'$ and $R'_k$ along with the states of the system.

## 3. Adaptive filtering using variational Bayesian approximation

### 3.1. Normal inverse Wishart (NIW) distribution

In this work, we propose a VB based adaptive filtering to estimate the target state when both the measurement noise mean and covariance are unknown. The VB approximation can be used to estimate the measurement noise statistics by choosing an appropriate conjugate prior distribution. When the mean of a normal distribution is known, the inverse Wishart (IW) distribution is used as the conjugate prior for the covariance of normal distribution [54]. But if the mean as well as the covariance of a normal distribution are unknown, then in general, the normal inverse Wishart (NIW) distribution is preferred as the joint conjugate prior [55].

The inverse Wishart distribution of a random symmetric positive definite matrix, $\mathcal{B} \in \mathbb{R}^{n\times n}$ can be written as

$$IW(\mathcal{B};\lambda,\psi) = \frac{|\psi|^{\lambda/2}|\mathcal{B}|^{-(\lambda+n+1)/2}\exp\{-0.5tr(\psi\mathcal{B}^{-1})\}}{2^{n\lambda/2}\Gamma_n(\lambda/2)}, \tag{6}$$

where $\lambda$ is the degrees of freedom, $\psi \in \mathbb{R}^{n\times n}$ is a symmetric positive definite matrix and is known as the scale matrix, $|\cdot|$, $tr(\cdot)$ are the determinant and trace of a matrix, respectively, and $\Gamma_n(\cdot)$ is the $n$ variate gamma function [54]. If $\mathcal{B} \sim IW(\mathcal{B};\lambda,\psi)$, then a property of the IW distribution is that the mean of the distribution is evaluated as, $E[\mathcal{B}] = \psi/(\lambda - n - 1)$, such that $\lambda > n + 1$ [54]. The NIW probability density function of $\mathcal{X}$ and $\mathcal{B}$ can be expressed as

$$NIW(\mathcal{X},\mathcal{B};\mu_{\mathcal{X}},\alpha,\lambda,\psi) = \mathcal{N}\left(\mathcal{X};\mu_{\mathcal{X}},\alpha\mathcal{B}\right)IW(\mathcal{B};\lambda,\psi), \tag{7}$$

where $\mu_{\mathcal{X}} = E[\mathcal{X}]$ and $\alpha > 0$ is named as the confidence parameter. With marginalization the pdf of a NIW distributed random vector $\mathcal{X}$ can be written as

$$NIW(\mathcal{X};\mu_{\mathcal{X}},\alpha,\lambda,\psi) = \int \mathcal{N}(\mathcal{X};\mu_{\mathcal{X}},\alpha\mathcal{B})IW(\mathcal{B};\lambda,\psi)d\mathcal{B}. \tag{8}$$

### 3.2. Adaptive filtering

#### 3.2.1. Process update

The prior state follows Gaussian distribution with mean, $\hat{\mathcal{X}}_{k|k-1}$ and covariance, $P_{k|k-1}$, *i.e.*,

$$p(\mathcal{X}_k|\mathcal{Y}_{1:k-1}) = \mathcal{N}(\mathcal{X}_k;\hat{\mathcal{X}}_{k|k-1},P_{k|k-1}), \tag{9}$$

As the process noise statistics are known, *i.e.*, $\mathcal{N}(0,Q)$ and the process is linear in our case, so the process update will follow the Kalman filter equations:

$$\hat{\mathcal{X}}_{k|k-1} = F\hat{\mathcal{X}}_{k-1|k-1} - \mho_{k-1,k}, \tag{10}$$

$$P_{k|k-1} = FP_{k-1|k-1}F^T + Q. \tag{11}$$

#### 3.2.2. Measurement update

Let us assume the variables, $r'$ and $R'_k$ represent the measurement noise mean and covariance matrix at any step, $k$. As we assume mean of the noise is time invariant, we omit the subscript $k$ while denoting it. Accumulated measurements up to time step, $k$ is denoted as $\mathcal{Y}_{1:k} = \{\mathcal{Y}_1,\mathcal{Y}_2,\cdots,\mathcal{Y}_k\}$. Using the formula of conditional probability, we can factorize the joint density as follows,

$$\begin{aligned}p(\mathcal{X}_k,R'_k,r',\mathcal{Y}_{1:k}) &= p(\mathcal{X}_k|R'_k,r',\mathcal{Y}_{1:k})p(R'_k,r',\mathcal{Y}_{1:k})\\ &= p(\mathcal{X}_k|R'_k,r',\mathcal{Y}_{1:k})p(r'|R'_k,\mathcal{Y}_{1:k})p(R'_k,\mathcal{Y}_{1:k})\\ &= p(\mathcal{X}_k|R'_k,r',\mathcal{Y}_{1:k})p(r'|R'_k,\mathcal{Y}_{1:k})p(R'_k|\mathcal{Y}_{1:k})p(\mathcal{Y}_{1:k}).\end{aligned}$$

Using the formula of conditional probability, we can also write $p(\mathcal{X}_k,R'_k,r',\mathcal{Y}_{1:k}) = p(\mathcal{X}_k,R'_k,r'|\mathcal{Y}_{1:k})p(\mathcal{Y}_{1:k})$, and finally we obtain

$$p(\mathcal{X}_k,R'_k,r'|\mathcal{Y}_{1:k}) = p(\mathcal{X}_k|R'_k,r',\mathcal{Y}_{1:k})p(r'|R'_k,\mathcal{Y}_{1:k})p(R'_k|\mathcal{Y}_{1:k}). \tag{12}$$

However, it would be difficult to evaluate each joint pdfs and in this respect, we approximately factorize it as follows,

$$p(\mathcal{X}_k,r',R'_k|\mathcal{Y}_{1:k}) \approx q(\mathcal{X}_k)q(r')q(R'_k), \tag{13}$$

where $q(\cdot)$ represents approximate pdf of $p(\cdot)$, where

$$\log q(\theta) = E_{-\theta}[\log p(\mathcal{X}_k,r',R'_k,\mathcal{Y}_{1:k})] + c_\theta, \tag{14}$$

where the $E_{-\theta}$ represents the expectation on the variables other than $\theta$, and $c_\theta$ is a constant. By minimizing the Kullback-Leibler divergence (KLD) between the approximate pdf $q(\mathcal{X}_k)q(r')q(R'_k)$ and the true pdf $p(\mathcal{X}_k,r',R'_k|\mathcal{Y}_{1:k})$, we will receive (14); see [56, pp. 450] and [57].

As $R'_k$ is the covariance of a Gaussian pdf, we consider the distribution of it, *i.e.*, $p(R'_k|\mathcal{Y}_{1:k})$ follows an IW distribution. So, we can write

$$p(R'_k|\mathcal{Y}_{1:k}) = IW(R'_k;\hat{u}'_k,\hat{U}'_k), \tag{15}$$

where $\hat{u}'_k > m+1$, $m$ is the dimension of measurement, and estimated value of $R'_k$ is the mean of the IW distribution given by

$$E[R'_k] = \frac{\hat{U}'_k}{\hat{u}'_k - m - 1}, \tag{16}$$

where the degree of freedom, $\hat{u}'_k > m+1$, as $m$ is the dimension of measurement, so $m > 0$, therefore, $\hat{u}'_k > 0$, and the scale parameter, $\hat{U}'_k > 0$. Consequently, the mean, $E[R'_k]$ is also positive, ensuring that the estimated measurement noise covariance is positive definite by nature. As $r'$ is the mean of a Gaussian pdf, we consider the prior distribution of it, *i.e.*, $p(r'|\mathcal{Y}_{1:k})$ follows a normal distribution,

$$p(r'|\mathcal{Y}_{1:k}) = \mathcal{N}(r';\hat{r}_k,\hat{\alpha}_kR'_k), \tag{17}$$

where $\hat{r}_k = E[r']$ and $\hat{\alpha}_k$ is the confidence parameter.

**Remark.** Please note that $\hat{r}_k$ is an estimate of the mean of measurement noise. Although we assume the mean of the measurement noise is not time varying, we shall receive estimates which are not the same in all the time steps. To represent that, we use subscript, $k$ while writing the estimate of the mean of the measurement noise.

In our problem described above, the measurement is bearing angle and hence the measurement equation is nonlinear which makes the pdf of states non Gaussian. However, we approximate them as Gaussian and represent them with a few deterministic sample points and their corresponding weights. These points are alternatively known as sigma points or support points and can be generated using the unscented transform [58,59], or cubature rule [60,20] or Gauss-Hermite rule of integration [19,61]. To read more about the various approaches for generating sample points and weights interested readers may read [23]. Any Gaussian distribution of mean, $\hat{\mathcal{X}}_{k|k-1}$ and covariance, $P_{k|k-1}$ is approximated with the points

$$\mathrm{X}_{j,k|k-1} = \hat{\mathcal{X}}_{k|k-1} + S_{k|k-1}\xi_j, \tag{18}$$

and their weights $w_j$, where $S_{k|k-1}$ is the Cholesky factorization of $P_{k|k-1}$ and $\xi_j$ are the points to represent standard normal distribution. The predicted measurements at each sigma point are evaluated as

$$\mathrm{Y}_{j,k|k-1} = h(\mathrm{X}_{j,k|k-1}). \tag{19}$$

**Theorem 1.** *Recursive estimates of the degree of freedom,* $\hat{u}'_k$ *and the estimated scale parameter,* $\hat{U}'_k$ *are given by*

$$\hat{u}'_k = \hat{u}'_{k-1} + 2, \tag{20}$$

*and*

$$\hat{U}'_k = \hat{U}'_{k-1} + B_k + D_k, \tag{21}$$

*respectively, where*

$$B_k = E_{\mathcal{X}_k,r'}[\tilde{\mathcal{Y}}_k\tilde{\mathcal{Y}}_k^T], \tag{22}$$

*and*

$$D_k = \frac{1}{\hat{\alpha}_{k-1}}E_{r'}[\tilde{r}\tilde{r}^T], \tag{23}$$

*where* $\tilde{\mathcal{Y}}_k = (\mathcal{Y}_k - h(\mathcal{X}_k) - r')$ *and* $\tilde{r} = (r' - \hat{r}_{k-1})$.

**Proof.** The proof is provided in Appendix A. □

**Lemma 1.** *With the help of deterministic sample points, and their corresponding weights,* $B_k$ *of* (22) *can be manipulated as*

$$B_k = (\mathcal{Y}_k - \sum_{j=1}^{N_s} w_j \Upsilon_{j,k|k-1} - \hat{r}_{k-1})(\mathcal{Y}_k - \sum_{j=1}^{N_s} w_j \Upsilon_{j,k|k-1} - \hat{r}_{k-1})^T$$
$$+ \sum_{j=1}^{N_s} w_j [\Upsilon_{j,k|k-1} - h(\hat{\mathcal{X}}_{k|k-1})][\Upsilon_{j,k|k-1} - h(\hat{\mathcal{X}}_{k|k-1})]^T + \hat{\alpha}_{k-1}\hat{R}_k. \tag{24}$$

**Proof.** The proof is provided in Appendix B. □

**Lemma 2.** *$D_k$ of (23) can be manipulated as*

$$D_k = \frac{\hat{\alpha}_k}{\hat{\alpha}_{k-1}}\hat{R}_k + \frac{1}{\hat{\alpha}_{k-1}}(\hat{r}_k - \hat{r}_{k-1})(\hat{r}_k - \hat{r}_{k-1})^T. \tag{25}$$

**Proof.** The proof is provided in Appendix C. □

**Theorem 2.** *The covariance of the measurement likelihood, $\hat{R}_k$ can be evaluated as:*

$$\hat{R}_k = (\hat{U}'_{k-1} + (\mathcal{Y}_k - \sum_{j=1}^{N_s} w_j \Upsilon_{j,k|k-1} - \hat{r}_{k-1})(\mathcal{Y}_k - \sum_{j=1}^{N_s} w_j \Upsilon_{j,k|k-1} - \hat{r}_{k-1})^T$$
$$+ \sum_{j=1}^{N_s} w_j [\Upsilon_{j,k|k-1} - h(\hat{\mathcal{X}}_{k|k-1})][\Upsilon_{j,k|k-1} - h(\hat{\mathcal{X}}_{k|k-1})]^T + \hat{\alpha}_{k-1}\hat{R}_k$$
$$+ \frac{\hat{\alpha}_k \hat{R}_k}{\hat{\alpha}_{k-1}} + \frac{1}{\hat{\alpha}_{k-1}}(\hat{r}_k - \hat{r}_{k-1})(\hat{r}_k - \hat{r}_{k-1})^T)(\hat{u}'_k - m - 1)^{-1}. \tag{26}$$

**Proof.** The proof is provided in Appendix D. □

**Theorem 3.** *The estimated mean, $\hat{r}_k$ and the estimated confidence parameter, $\hat{\alpha}_k$ of the measurement noise mean can be evaluated with the help of deterministic sample points as*

$$\hat{r}_k = \frac{\hat{r}_{k-1} + \hat{\alpha}_{k-1}(\mathcal{Y}_k - \sum_{j=1}^{N_s} w_j \Upsilon_{j,k|k-1})}{\hat{\alpha}_{k-1} + 1}, \tag{27}$$

*and*

$$\hat{\alpha}_k = \frac{\hat{\alpha}_{k-1}}{\hat{\alpha}_{k-1} + 1}, \tag{28}$$

*respectively.*

**Proof.** The proof is provided in Appendix E. □

**Theorem 4.** *The expressions of the mean and the covariance of Gaussian approximated posterior pdf are*

$$\hat{\mathcal{X}}_{k|k} = \hat{\mathcal{X}}_{k|k-1} + K_k(\mathcal{Y}_k - h(\hat{\mathcal{X}}_{k|k-1}) - \hat{r}_k), \tag{29}$$

*and*

$$P_{k|k} = P_{k|k-1} - K_k P_{\mathcal{Y}\mathcal{Y}} K_k^T, \tag{30}$$

*where $K_k = P_{\mathcal{X}\mathcal{Y}} P_{\mathcal{Y}\mathcal{Y}}^{-1}$ is the Kalman gain, where*

$$P_{\mathcal{X}\mathcal{Y}} = \sum_{j=1}^{N_s} w_j (\mathrm{X}_{j,k|k-1} - \hat{\mathcal{X}}_{k|k-1})(\Upsilon_{j,k|k-1} - h(\hat{\mathcal{X}}_{k|k-1}))^T, \tag{31}$$

*and*

$$P_{\mathcal{Y}\mathcal{Y}} = \hat{R}_k + \sum_{j=1}^{N_s} w_j (\Upsilon_{j,k|k-1} - h(\hat{\mathcal{X}}_{k|k-1}))(\Upsilon_{j,k|k-1} - h(\hat{\mathcal{X}}_{k|k-1}))^T. \tag{32}$$

**Proof.** The proof is provided in Appendix F. □

Interestingly, the filter proposed with the help of above theorems is unbiased and it is proved in the following lemma.

**Lemma 3.** *Estimation of VB adaptive filter is unbiased* i.e., $E[\hat{\mathcal{X}}_{k|k} - \mathcal{X}_k] = 0$.

**Proof.** The proof is provided in Appendix G. □

### 3.3. Fixed point iteration

We can see in (24), that $B_k$ is a function of $\hat{R}_k$ and (25), $D_k$ is a function of $\hat{R}_k$, which is used to evaluate $\hat{U}'_k$; as in (21), which is again used to evaluate $\hat{R}_k$, making $\hat{R}_k$ a function of $\hat{R}_k$; see (26). As evaluation of $\hat{R}_k$ depends on itself, fixed point iteration (FPI) method is needed to obtain it. The FPI is performed during each measurement update step where (24) and (25) are iterated as,

$$B_k^{i+1} = (\mathcal{Y}_k - \sum_{j=1}^{N_s} w_j \Upsilon_{j,k|k-1} - \hat{r}_{k-1})(\mathcal{Y}_k - \sum_{j=1}^{N_s} w_j \Upsilon_{j,k|k-1} - \hat{r}_{k-1})^T$$
$$+ \sum_{j=1}^{N_s} w_j [\Upsilon_{j,k|k-1} - h(\hat{\mathcal{X}}_{k|k-1})][\Upsilon_{j,k|k-1} - h(\hat{\mathcal{X}}_{k|k-1})]^T + \hat{\alpha}_{k-1}\hat{R}_k^i, \tag{33}$$

and

$$D_k^{i+1} = \frac{\hat{\alpha}_k}{\hat{\alpha}_{k-1}}\hat{R}_k^i + \frac{1}{\hat{\alpha}_{k-1}}(\hat{r}_k - \hat{r}_{k-1})(\hat{r}_k - \hat{r}_{k-1})^T, \tag{34}$$

respectively. Using these, $\hat{U}_k^{i+1}$ in (21), is evaluated for every $(i+1)$th iteration which is further used to calculate $\hat{R}_k^{i+1}$. Using $\hat{R}_k^{i+1}$, measurement error covariance, $P_{\mathcal{Y}\mathcal{Y}}^{i+1}$, the Kalman gain, $K_k^{i+1}$ and estimated state, $\hat{\mathcal{X}}_{k|k}^{i+1}$ and its error covariance, $P_{k|k}^{i+1}$ are evaluated. The loop continues till $|\hat{\mathcal{X}}_{k|k}^{i+1} - \hat{\mathcal{X}}_{k|k}^{i}| < \zeta$, a user defined threshold. In our work, we chose the value of $\zeta$ to be $10^{-3}$.

The measurement noise covariance is obtained using fixed point iteration method. As the fixed point iteration is a natural gradient method [29], it ensures an asymptotic convergence [62,35] for a limited initial error. If the fixed point iteration converges, the estimated measurement noise covariance will also converge. However, for a large initialization error, the estimated measurement noise covariance might not converge, as the VB approach guarantees only local convergence [35]. From the experiments, we have observed that the posterior estimate value converges quickly generally within a few iteration (4 or 5 iterations). Further, in earlier literature [63, pp. 158], [64,65], it has been argued that the fixed point iteration method is numerically stable if and only if iterated sequence converges to a fixed point. So, we can say that the fixed point iteration method is numerically stable. The algorithm of FPI is mentioned in Algorithm 1.

During initialization, $\hat{u}'_0$, is used to evaluate $\hat{U}'_0 = (\hat{u}'_0 - m - 1)\hat{R}_0$ which is further used to evaluate $\hat{U}_1^{i+1}$ and $\hat{\alpha}_0$ is used to evaluate $D_1^{i+1}$. So, proper initialization of $\hat{u}'_0$, and $\hat{\alpha}_0$ is essential as it has a significant impact on the estimation accuracy. The confidence parameters, $\hat{\alpha}_0$ and $\hat{u}'_0$ are initialized by maximizing the likelihood *i.e.*

$$(\hat{u}'_0, \hat{\alpha}_0) = \arg\max_{(\hat{u}'_0, \hat{\alpha}_0)} L_k, \tag{35}$$

where

$$L_k = \frac{1}{\sqrt{2\pi P_{\mathcal{Y}\mathcal{Y}}^i}} \exp(-\frac{1}{2}\frac{(\mathcal{Y}_k - \hat{\mathcal{Y}}_{k|k-1}^i)^2}{P_{\mathcal{Y}\mathcal{Y}}^i}). \tag{36}$$

It has been observed that fair accuracy is achieved with $\hat{\alpha}_0$ and $\hat{u}'_0$ values ranging from 1 to 20 and 3 to 23, respectively. So the optimization problem is solved by varying $\hat{\alpha}_0$ and $\hat{u}'_0$ in this range and choosing a set of values where the cost function becomes maximum.

**Algorithm 1** Subroutine for fixed point iteration.

$[\hat{\mathcal{X}}_{k|k},\ P_{k|k},\ \hat{R}_k,\ \hat{U}_k] = FPI[\hat{\mathcal{X}}_{k|k-1},\ P_{k|k-1},\ \mathcal{Y}_k,\ \hat{R}_{k-1},\ \hat{r}_k,\ \hat{r}_{k-1},\ \hat{U}'_{k-1},\ \hat{u}'_k,\ \hat{\alpha}_k,\ \hat{\alpha}_{k-1}]$

Initialize: $N = 0$, $i = 0$, $\hat{R}_k^0 = \hat{R}_{k-1}$.
**while** N = 0 **do**
    Evaluate: $\hat{U}_k^{i+1}$ from Eqn. (21) using $B_k^{i+1}$ and $D_k^{i+1}$ from (33) and (34), respectively.
    Evaluate: $\hat{R}_k^{i+1} = \hat{U}_k^{i+1}/(\hat{u}'_k - m - 1)$.
    Calculate measurement error covariance, $P_{\mathcal{YY}}^{i+1}$ using (32) replacing $\hat{R}_k$ by $\hat{R}_k^{i+1}$.
    Evaluate Kalman gain, $K_k^{i+1} = P_{\mathcal{XY}}(P_{\mathcal{YY}}^{i+1})^{-1}$.
    Evaluate posterior mean, $\hat{\mathcal{X}}_{k|k}^{i+1}$ using $K_k^{i+1}$ in (29).
    Evaluate posterior error covariance, $P_{k|k}^{i+1}$ using $K_k^{i+1}$ and $P_{\mathcal{YY}}^{i+1}$ in (30).
    **if** $|\hat{\mathcal{X}}_{k|k}^{i+1} - \hat{\mathcal{X}}_{k|k}^{i}| < \zeta$ **then**
        $N = i + 1$.
    **else**
        $i = i + 1$.
    **end if**
**end while**
$\hat{\mathcal{X}}_{k|k} = \hat{\mathcal{X}}_{k|k}^N, P_{k|k} = P_{k|k}^N, \hat{U}'_k = \hat{U}_k^N,\ \hat{R}_k = \hat{R}_k^N$.

**Algorithm 2** VB approach for unknown noise covariance.

Inputs: $\hat{\mathcal{X}}_{0|0}, P_{0|0}, \hat{r}_0, \hat{R}_0, \hat{\alpha}_0, \hat{u}'_0, \hat{U}'_0 = (\hat{u}'_0 - m - 1)\hat{R}_0$.
**for** k = 1 to $TT$ **do**
    Time Update:
    $\hat{\mathcal{X}}_{k|k-1} = F\hat{\mathcal{X}}_{k-1|k-1} - \mho_{k-1,k}$
    $P_{k|k-1} = FP_{k-1|k-1}F^T + Q_k$
    Measurement Update:
    Calculate: $S_{k|k-1} = chol(P_{k|k-1})$.
    Evaluate sigma points: $\mathrm{X}_{j,k|k-1} = S_{k|k-1}\xi_j + \hat{\mathcal{X}}_{k|k-1}$, where $j = 1, 2, ..., N_s$.
    Predicted measurement at each point, $\Upsilon_{k|k-1} = h(\mathrm{X}_{j,k|k-1})$.
    Predicted measurement estimation: $h(\hat{\mathcal{X}}_{k|k-1}) = \sum_{j=1}^{N_s} w_j \Upsilon_{j,k|k-1}$.
    Calculate $P_{\mathcal{XY}}$ using (31).
    Evaluate: $\hat{r}_k$, $\hat{\alpha}_k$ and $\hat{u}'_k$ from (27), (28), and (20), respectively.
    Determine: $[\hat{\mathcal{X}}_{k|k},\ P_{k|k},\ \hat{R}_k,\ \hat{U}'_k]$=FPI$[\hat{\mathcal{X}}_{k|k-1},\ P_{k|k-1},\ \mathcal{Y}_k,\ \hat{R}_{k-1},\ \hat{r}_k,\ \hat{r}_{k-1},\ \hat{U}'_{k-1},\ \hat{u}'_k,\ \hat{\alpha}_k,\ \hat{\alpha}_{k-1}]$.
**end for**

## 4. Simulation results

We have implemented the proposed estimation method in two engagement scenarios [2,59] as shown in Fig. 1a and 1b, respectively. In both the scenarios, the target moves in a nearly straight line motion with constant velocity. The ownship maneuvers from 13th min to 17th min in the Scenario I and at 15th min in Scenario II. As we discussed earlier, the ownship maneuver is required in order to make the system observable so that the estimators start converging. Among the two scenarios, tracking is more difficult in Scenario II because in this engagement the system's nonlinearity is high [2,59]. The simulation parameters of both the scenarios are provided in Table 1. The sampling time for both the scenarios is considered to be 5 s, *i.e.,* $\Delta = 5$ s, and the simulation is performed for a total time period of 30 min. The process noise intensity, $\bar{q}$ is considered to be $1.944 \times 10^{-6}$ km$^2$/min$^3$ for both the scenarios.

As we discussed earlier, our objective is to track the target from bearings-only measurements. We assume the measurements are corrupted with a Gaussian noise whose mean and covariance are unknown. Further, we consider the noise covariance could be time invariant or time varying in nature. Case I considers time invariant standard deviation, $\sigma_\theta$ which is $1.5^o$ for Scenario I, and $2^o$ for Scenario II. Case II represents time varying standard deviation $\sigma_{\theta_k}$, and we assume it is highest at the farthest distance and decreases linearly with range. The maximum and the minimum values for $\sigma_{\theta_k}$ in Case II are considered to

**Table 1**
Parameters of the scenarios.

| Parameters | Scenario I | Scenario II |
|---|---|---|
| Initial range ($r$) | 5 km | 10 km |
| Target speed ($s$) | 4 knots | 15 knots |
| Target course | $-140^o$ | $-135.4^o$ |
| Observer speed | 5 knots | 5 knots |
| Observer initial course | $140^o$ | $-80^o$ |
| Observer final course | $20^o$ | $146^o$ |
| Observer maneuver | From $13^{th}$ to $17^{th}$ min | $15^{th}$ min |
| Initial range S.D. ($\sigma_r$) | 2 km | 4 km |
| Initial target speed S.D. ($\sigma_s$) | 2 knots | 2 knots |
| Initial course S.D. ($\sigma_c$) | $\pi/\sqrt{12}$ | $\pi/\sqrt{12}$ |
| Measurement noise mean ($r$) | $0.1^o$ | $0.1^o$ |

be $4^o$ and $1.5^o$, respectively. The true value of the measurement noise mean is considered to be $0.1^o$ for all the cases.

In this study, we implemented the EKF-VB, CKF-VB, UKF-VB, and GHF-VB filters on both the scenarios and both time varying and time invariant noise covariance cases. The initial values of the measurement noise mean and covariance used in all the adaptive filtering are considered to be half of their initial true values. The results obtained from them are compared to their respective MAP-MLE adaptive filters. In the MAP-MLE adaptive filters, the mean of the measurement noise is estimated using MAP [26] method and the covariance of the measurement noise is estimated using MLE [66], *i.e.*, $\hat{r}_k = \frac{1}{k}\sum_{i=1}^{k}(\mathcal{Y}_k - h(\hat{\mathcal{X}}_{k|k-1}))$, and $\hat{R}_k = \hat{C}_v + \sum_{j=1}^{N_s} w_j[\Upsilon_{j,k|k-1} - h(\hat{\mathcal{X}}_{k|k-1})][\Upsilon_{j,k|k-1} - h(\hat{\mathcal{X}}_{k|k-1})]^T$.

Figs. 2a and 2b show the true and the estimated measurement noise means and its standard deviations for Scenario I and II, respectively for a single representative run. From the plots, we can see that both the MAP-MLE and the VB adaptive filters are able to estimate the measurement noise mean and its standard deviation. As a single run cannot be used for performance comparison of the two adaptive techniques, we rely on performance metrics such as RMSE, ANEES, bias norm, and track loss % for the purpose.

### 4.1. Performance comparison of tracking filters

#### 4.1.1. Scenario I

For Scenario I, Figs. 3a, 3b show the RMSE of position and velocity, respectively, for time invariant noise covariance, $R$ (Case I), and 4a, 4b show RMSE of position and velocity, respectively, for time varying $R_k$ (Case II). All the RMSE plots are evaluated excluding the diverged tracks, as in [8] for 500 Monte Carlo runs. A track is said to be diverging when the terminal estimation error in range is beyond a certain value which is referred to as the track bound. Here, we are considering a track bound of 200 m. From the figures, we see that all the adaptive filters have higher RMSE than the nonadaptive filters because nonadaptive filters know the noise statistics. Further, we observed, that for unknown noise statistics, VB filters show lower RMSE than the filters with the MAP-MLE technique. Among all the VB adaptive filters, the GHF-VB shows better performance than the rest and the others are comparable to each other. In Fig. 3, we can see that the RMSE of EKF-VB is slightly lower than the VB based deterministic sample point filters. The reason behind it is that the RMSE results are assessed excluding the diverged tracks. Please note that the track loss percentage of EKF-VB is much higher (see Table 2) and if we include them, the RMSE of it will be much higher than the rest of the filters. As expected, the CRLB remains the lowest.

The bias norm plots for Scenario I are shown in 5a and 6a for Case I and II, respectively. The bias norm plots are also evaluated as in [8] for 500 Monte Carlo runs, excluding the diverged tracks, considering a track bound of 200 m. All the plots converge to zero at the end of the simulation signifying zero bias. The ANEES plots, evaluated from 500 Monte Carlo (MC) runs excluding diverged tracks with a track bound

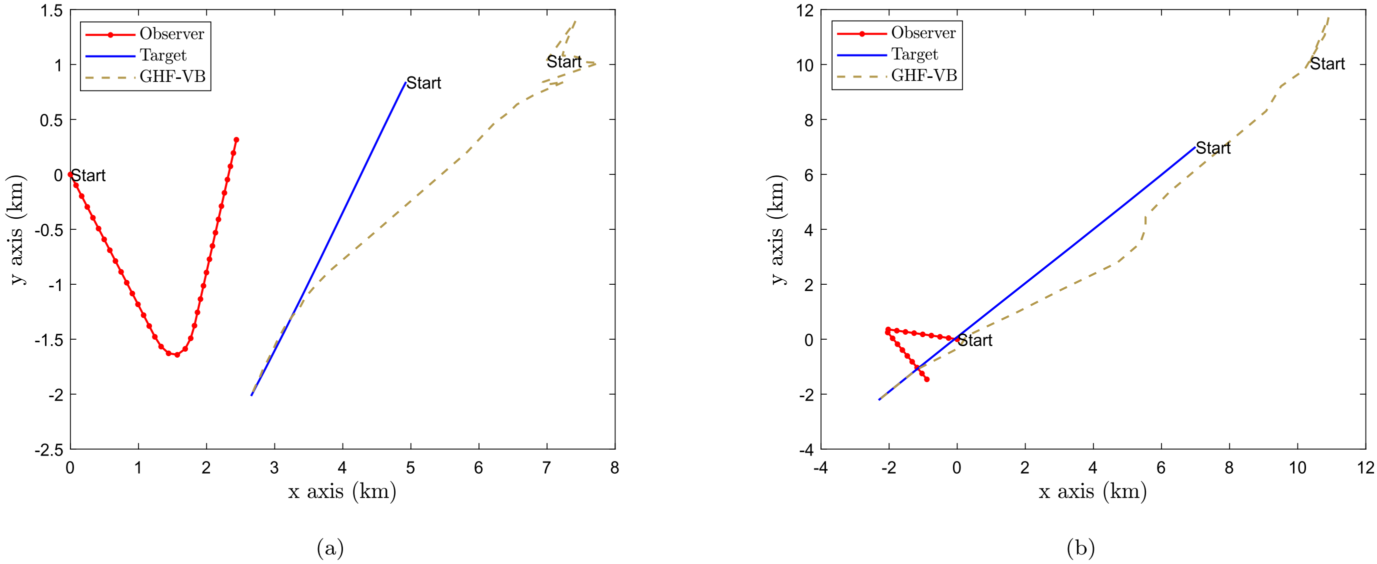



**Fig. 1.** Tracking scenario for (a) moderately nonlinear scenario (Scenario I), and (b) highly nonlinear scenario (Scenario II).

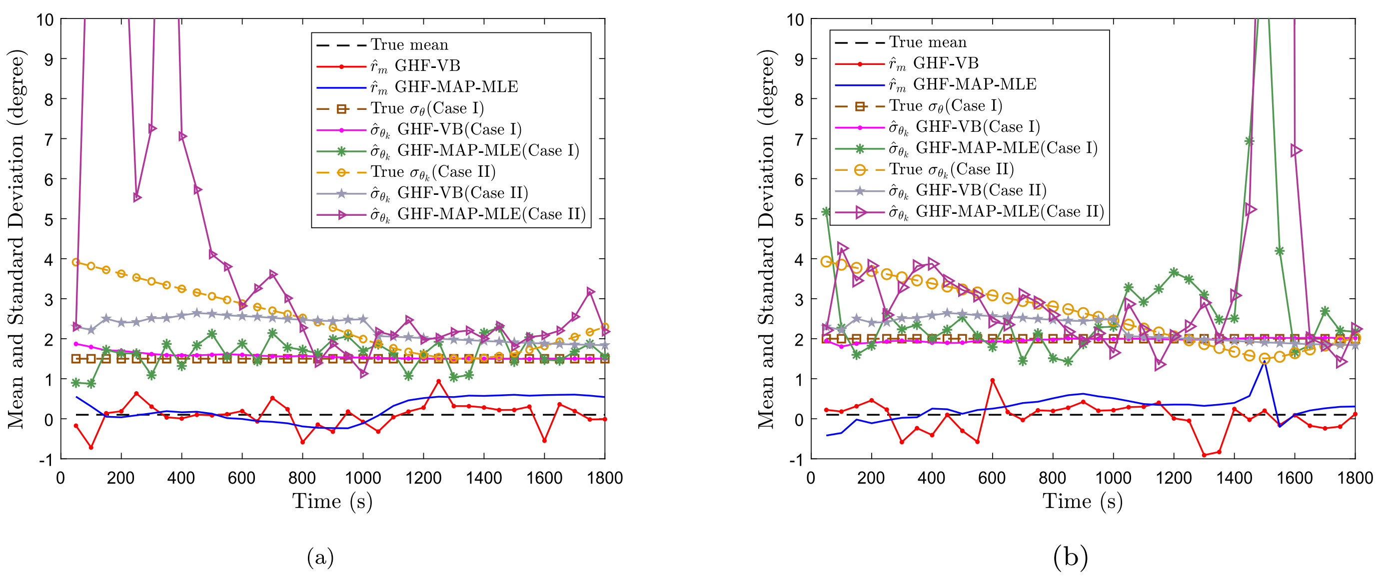



**Fig. 2.** True and estimated measurement noise mean and standard deviation for (a) Scenario I, and (b) Scenario II.

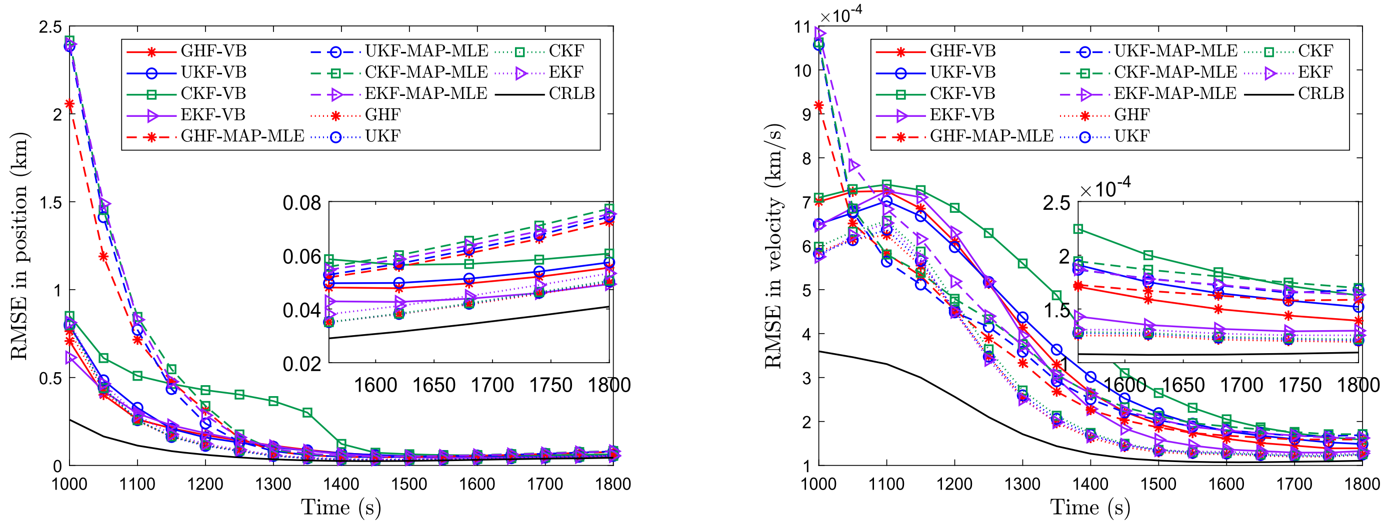
of 200 m, are shown in 5b and 6b for Case I and II, respectively. The ANEES plots of the VB adaptive filters are almost similar to the plots of the nonadaptive filters after 1600 s, which are within or close to the 95% bound in both cases which signifies consistent estimation. However, the MAP-MLE adaptive filters have much higher ANEES values which indicates that the estimation technique is optimistic which is not desirable.

The percentage of track loss for Scenario I is shown in Table 2 for both time varying and time invariant measurement noise covariance. The track loss % is evaluated from 10,000 MC runs, considering a track bound of 200 m. From the table, we can see that the VB adaptive filters show much less track loss % compared to the MAP-MLE adaptive filters at the expense of slightly higher execution time. Among all the VB filters, GHF-VB shows the lowest track loss % than the rest. The track loss % of all the filters in time varying $R_k$ (Case II) is slightly higher than the time invariant $R$ (Case I).

The VB based adaptive filtering method for unknown measurement noise statistics is compared to some of the existing adaptive filtering methods found in the literature. The non adaptive GHF, GHF-VB, GHF-

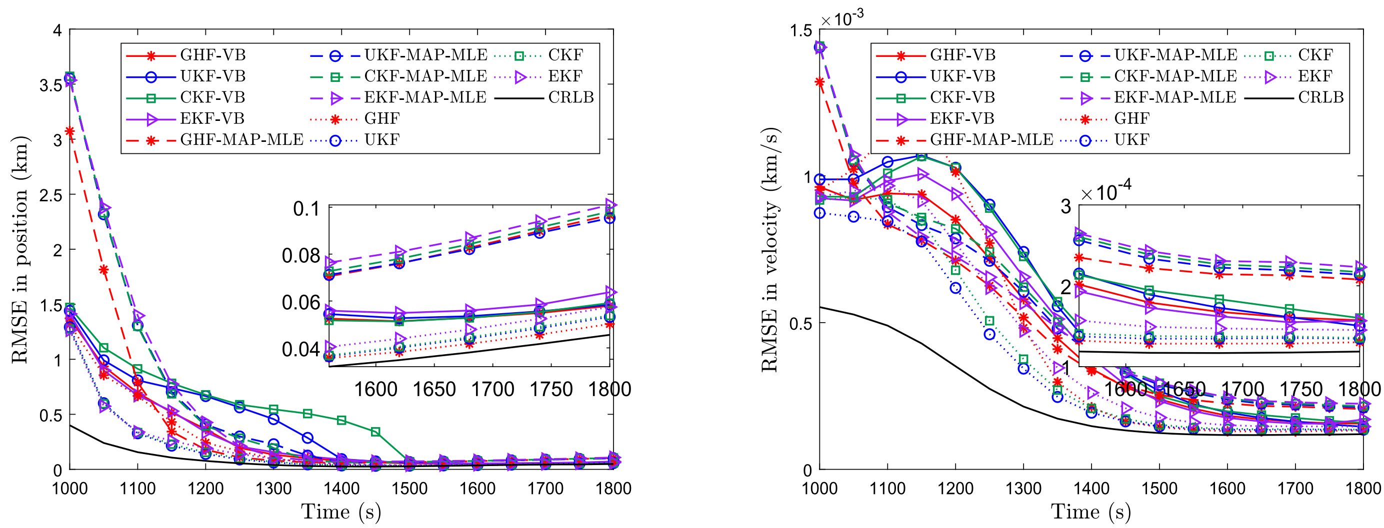
**Table 2**
The track loss %, relative execution time, and the number of flop counts.

| Filter | Track loss % | | | | Rel. | Flop |
|---|---|---|---|---|---|---|
| | Scenario I | | Scenario II | | exe. | count |
| | Case I | Case II | Case I | Case II | time | ($\times 10^4$) |
| EKF | 16.35 | 16.85 | 19.55 | 21.79 | 1 | 1.271 |
| EKF-MAP-MLE | 40.36 | 46.85 | 56.13 | 70.97 | 1.13 | 1.293 |
| EKF-VB | 18.57 | 19.58 | 20.47 | 29.62 | 1.48 | 1.678 |
| CKF | 1.48 | 2.21 | 4.56 | 11.24 | 2.59 | 1.311 |
| CKF-MAP-MLE | 34.68 | 42.67 | 52.48 | 63.48 | 2.69 | 1.329 |
| CKF-VB | 9.48 | 10.47 | 17.53 | 24.32 | 3.99 | 1.933 |
| UKF | 1.46 | 2.13 | 4.35 | 10.29 | 2.85 | 1.356 |
| UKF-MAP-MLE | 34.66 | 38.53 | 49.44 | 63.09 | 3.12 | 1.335 |
| UKF-VB | 8.33 | 8.99 | 16.45 | 23.12 | 4.62 | 1.975 |
| GHF | 1.44 | 2.05 | 4.10 | 9.44 | 3.76 | 1.683 |
| GHF-MAP-MLE | 34.47 | 37.92 | 39.78 | 54.08 | 4.64 | 1.731 |
| GHF-VB | 7.43 | 8.86 | 15.24 | 19.83 | 13.55 | 5.021 |




(a) (b)

**Fig. 3.** RMSE in (a) position, and (b) velocity for time invariant measurement noise covariance, Scenario I.




(a) (b)

**Fig. 4.** RMSE in (a) position, and (b) velocity for time varying measurement noise covariance, Scenario I.

MAP-MLE, GHF-MAP-GGIG, and GHF-MAP-SW are compared with each other. The RMSE plots for position and velocity are shown in Fig. 7a and 7b, respectively for Case I and 8a, and 8b, respectively for Case II. The RMSE is evaluated for 500 Monte Carlo runs excluding the diverged tracks, considering a track bound of 1 km. It can be seen that the GHF-VB has the lowest RMSE among all the other adaptive techniques. The track loss for different adaptive filtering techniques, along with the non adaptive GHF, is listed in Table 3. The track loss percentage is evaluated for 10,000 Monte Carlo runs, considering a track bound of 1 km. From the table, it can be observed that the GHF-VB has the lowest track loss percentage than the other adaptive filtering techniques for both cases.

#### 4.1.2. Scenario II

For Scenario II, Figs. 9a, 9b show the RMSE of position and velocity, respectively, for time invariant $R$ (Case I), and 10a, 10b show RMSE of position and velocity, respectively, for time varying $R_k$ (Case II). All the RMSE plots are evaluated excluding the diverged tracks, considering a track bound of 200 m for 500 MC runs. From the figures we see that the MAP-MLE adaptive filters have higher RMSE than the nonadaptive filters and the VB filters have similar or slightly higher RMSE compared to that of the nonadaptive filters. For nonadaptive filters, the exact noise statistics are assumed to be known, so they display the lowest RMSE. Among all the VB filters, the GHF-VB shows lower RMSE than the rest, while others are comparable to each other. In Figs. 9 and 10, we can see that the RMSE of EKF-VB is slightly lower than the CKF-VB and UKF-VB for the same reason mentioned in Sub-subsection 4.1.1.

The bias norm plots, excluding the diverged tracks, for Case I and II are shown in 11a and 12a, respectively. In both the figures, we can observe that the bias norm plots converge to zero at the end of the simulation signifying unbiased estimation. The ANEES plots, excluding diverged tracks, for Case I and II are shown in 11b and 12b, respectively. The ANEES plots of the VB adaptive filters are almost similar to the ANEES of their nonadaptive counterparts, and they are within or near the 95% probability region which signifies a consistent estimation. However, the ANEES of the MAP-MLE filters are much higher which indicates inferiority in estimation compared to the VB filters. The % of track loss, evaluated for 10,000 MC runs (with the same track bound), for Scenario II, are listed in Table 2. As Scenario II is highly nonlinear,

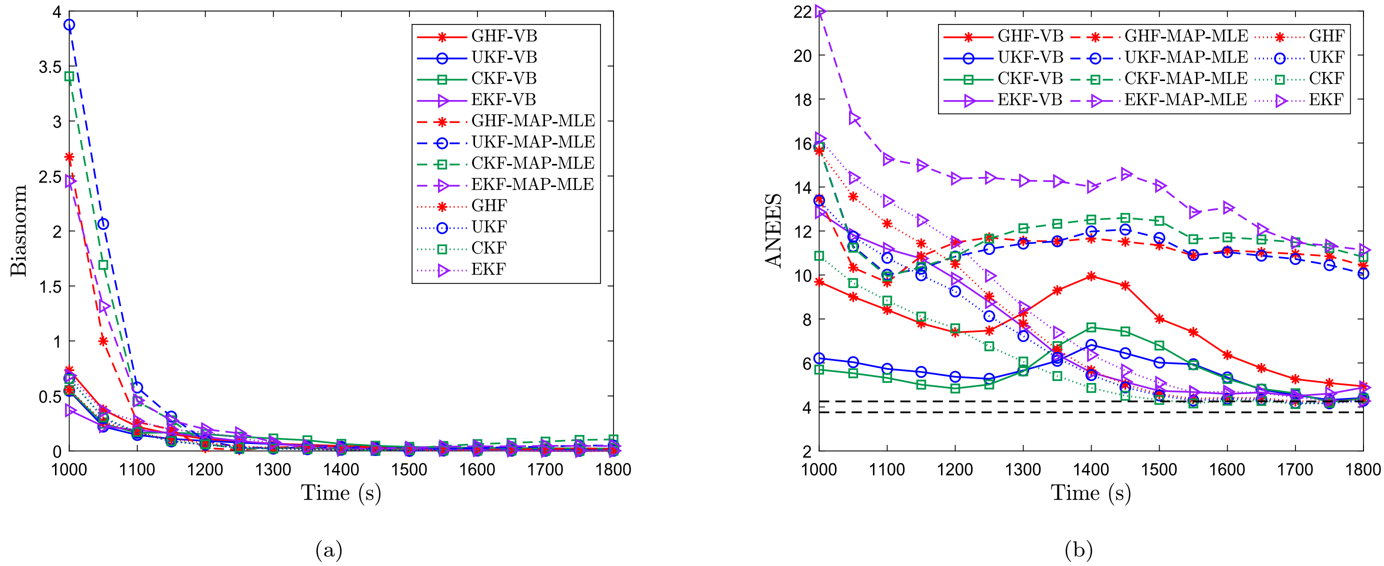



**Fig. 5.** (a) Bias norm, and (b) ANEES plots for time invariant measurement noise covariance, Scenario I.

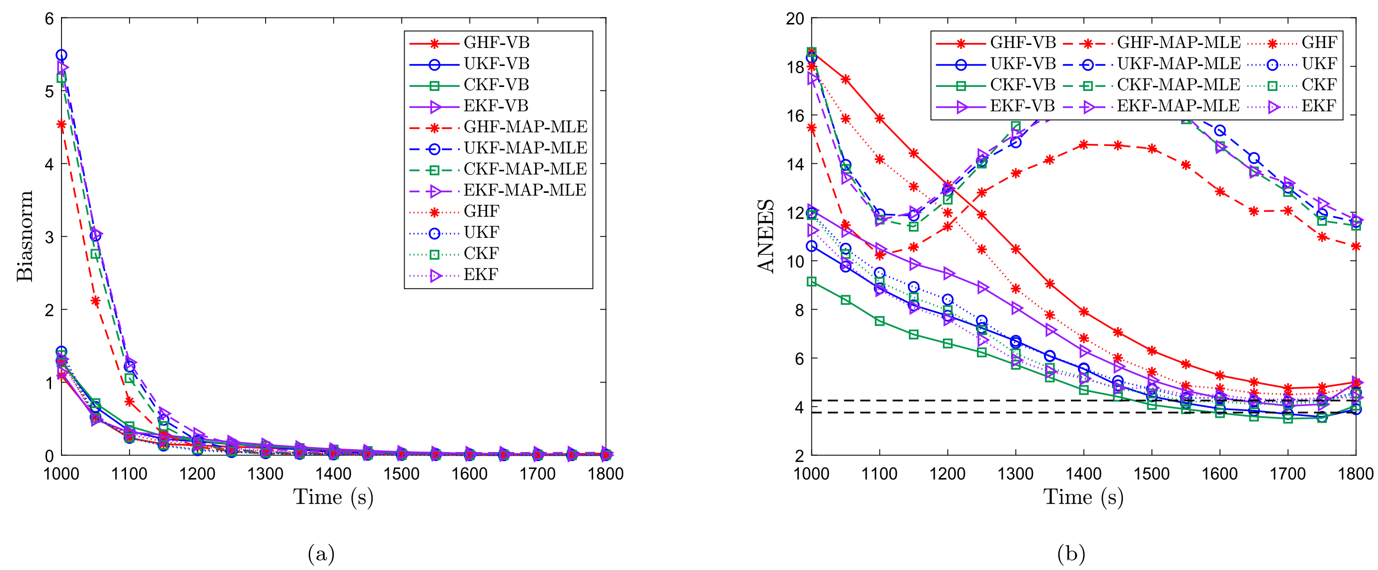



**Fig. 6.** (a) Bias norm, and (b) ANEES plots for time varying measurement noise covariance, Scenario I.

it is more difficult to estimate compared to Scenario I which was moderately nonlinear. So the track loss % of all the filters in Scenario II is much higher than in Scenario I. The VB filters have lower track loss % compared to the MAP-MLE filters. Among all the VB filters, the GHF-VB has the lowest track loss %, although the execution time of it is quite high.

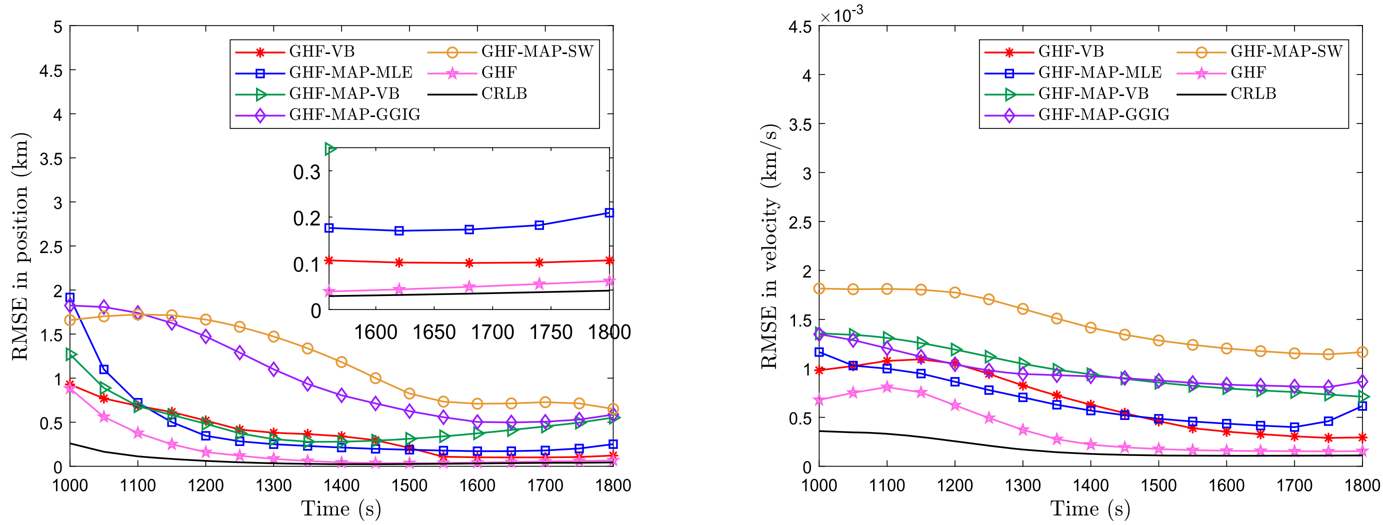
A comparison with the non adaptive GHF filter with GHF-VB, MAP-MLE, MAP-GGIG, and MAP-SW methods is done for Scenario II. The RMSE in position and velocity for the same are depicted in Figs. 13a and 13b for Case I, and Figs. 14a and 14a for Case II. These plots are generated based on 500 Monte Carlo runs, excluding diverged tracks, with a track bound of 1 km. From the results, it can be seen that GHF-VB exhibits the lowest RMSE compared to all other adaptive techniques. The track loss for different adaptive filtering techniques, including non-adaptive GHF, is summarized in Table 3, with the track loss percentage evaluated over 10,000 Monte Carlo runs, considering a track bound of 1 km. Notably, GHF-VB demonstrates the lowest track loss percentage among all adaptive filtering techniques for both cases.

### 4.2. Computational budget

The computation budget of any estimation technique can be evaluated using flop counts. The number of floating point operations termed flops can be counted by counting the number of addition, subtraction, multiplication, or division operations of two floating points. It is to be noted that while evaluating computation complexity the index fetching, storing of values, input, output, and initialization are neglected. The flop count for $A_{m\times n}+B_{m\times n}$ is $mn$, $A_{m\times n}\times B_{n\times p}$ is $(2n-1)pm$, $chol(A_{m\times m})$ is $m^3/3+2m^2$ and $exp(\cdot)$ is decomposed till third order Taylor series to evaluate the flop count. The computation complexity of the VB approach as shown in Algorithm 2 is given by

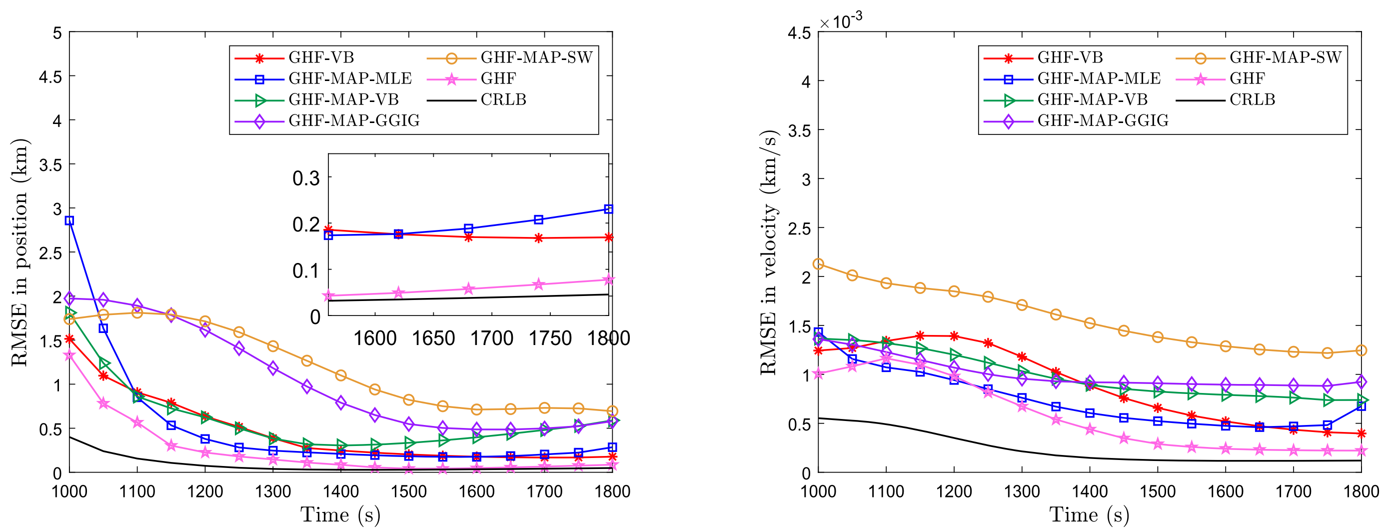
$$C(n,N_s,N_{FPI}) = 4n^6 - 4n^5 + n^4 + \frac{n^3}{3} + n^2(5+2N_s) + 3nN_s + 3N_s + 11 + N_{FPI}[2n^2 + 4n + 8N_s + 22], \tag{37}$$

where the system dimension, $n=4$ in our work, $N_s$ is the total number of sigma points, and $N_{FPI}$ is the total number of iterations that have to be performed at each time step. If we include the initialization of $\hat{\alpha}_0$ and $\hat{u}'_0$, as mentioned in Section 3.3, while calculating the total flop




(a) (b)

**Fig. 7.** RMSE in (a) position, and (b) velocity for time invariant measurement noise covariance, Scenario I.




(a) (b)

**Fig. 8.** RMSE in (a) position, and (b) velocity for time varying measurement noise covariance, Scenario I.

counts then the number of flops required to execute a VB Gaussian filter will be $C + N_i(2n^2 + 4n + 8N_s + 22)$, where $N_i$ is the number of loops executed during initialization.

As we are comparing the VB adaptive filtering with the MAP-MLE technique, it is interesting to evaluate the computation complexity of the MAP-MLE method as well. For the MAP-MLE approach of adaptive filtering, the number of flop counts is as follows

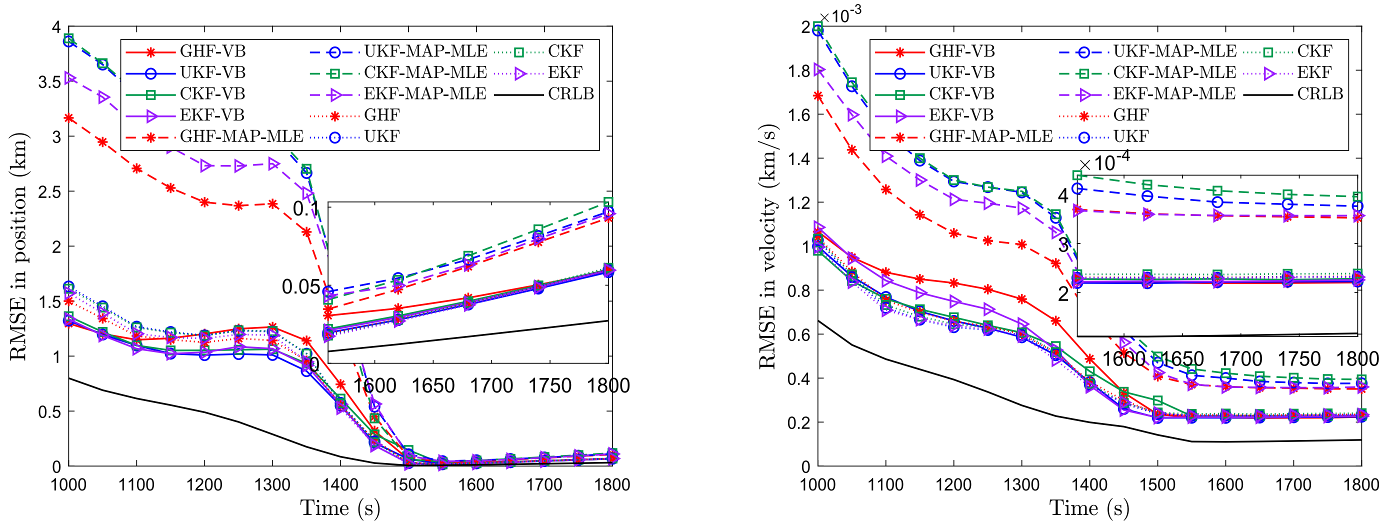
$$C(n, N_s, L, k) = 4n^6 - 4n^5 + n^4 + \frac{n^3}{3} + n^2(7 + 2N_s) + n(3N_s + 4) + 3k + 3L + 11N_s + 8, \quad (38)$$

where $k$ represents the $k$th time step and $L$ represents the window length.

The number of flop counts and the relative execution time for all the filters are listed in Table 2. In the flop count calculation, we have considered the window length in MAP-MLE, $L = 50$. As it is observed that the FPI algorithm generally converges within 6 to 7 iterations, $N_{FPI}$ is considered to be 7. From the table, we can see that the number of flops for the VB filters is slightly higher than the MAP-MLE. The relative execution time of all the filters is noted, *w.r.t.,* the time taken by the EKF. The execution time of the VB filters is a few times more than that of the MAP-MLE and nonadaptive filters. Among the VB filters, both the execution time and the flop counts of GHF-VB are highest because it requires 81 support points as we implemented a third order Gauss-Hermite approximation.

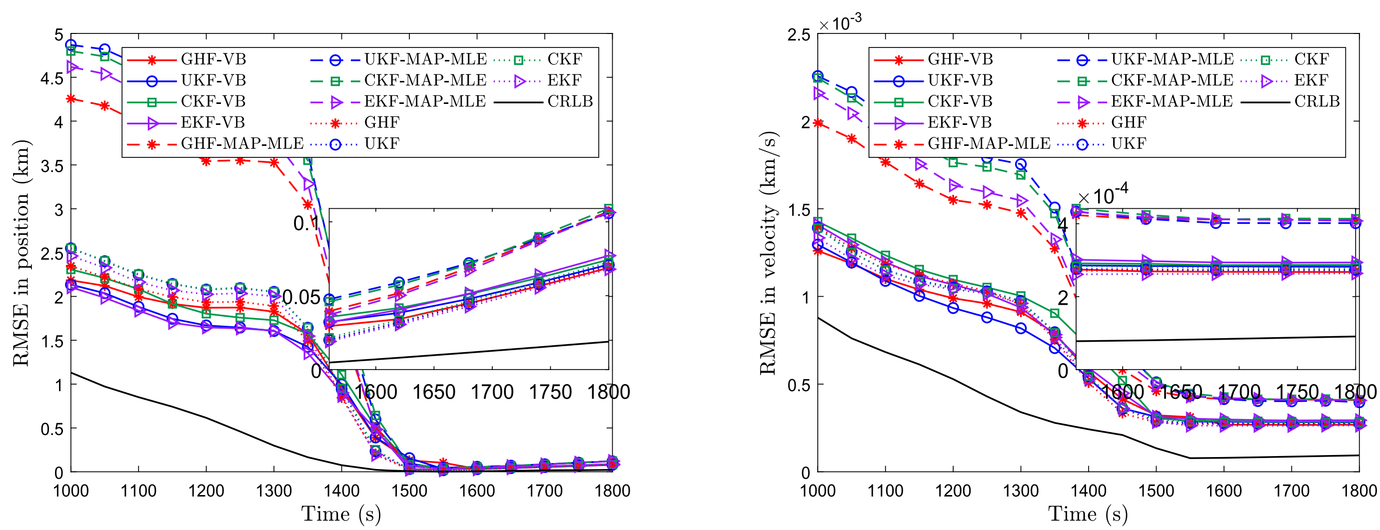
## 5. Discussion and conclusion

This paper presents a variational Bayesian (VB) based adaptive filtering framework for Gaussian filters to estimate the target trajectory using bearings-only measurements of unknown noise statistics. The proposed adaptive estimation is based on the assumption that the joint distribution of the measurement noise mean and covariance follow normal inverse Wishart distribution. The VB estimation for nonlinear system is realized using a set of deterministic sample points generated with various approaches. The VB Gaussian filters such as EKF-VB, CKF-VB, UKF-VB, and GHF-VB are implemented to track an underwater target for both time invariant and time varying unknown measurement noise




(a) (b)

**Fig. 9.** RMSE in (a) position, and (b) velocity for time invariant measurement noise covariance, Scenario II.




(a) (b)

**Fig. 10.** RMSE in (a) position, and (b) velocity for time varying measurement noise covariance, Scenario II.

**Table 3**
The % of track loss comparison of the proposed filter with the methods available in the literature for the track bound 1 km.

| | Scenario I | | Scenario II | |
|---|---|---|---|---|
| Filter | Case I | Case II | Case I | Case II |
| GHF | 0.7 | 1.6 | 2.2 | 2.4 |
| GHF-VB | 2 | 2.6 | 10.4 | 7.2 |
| GHF-MAP-MLE | 19.2 | 21.8 | 22 | 24.2 |
| GHF-MAP-VB | 16.8 | 19.19 | 22.6 | 24.6 |
| GHF-MAP-GGIG | 32.8 | 34.8 | 45.2 | 47.3 |
| GHF-MAP-SW | 31.8 | 32.8 | 36.4 | 38.6 |

covariance. The results are compared with the MAP-MLE based adaptive filters and with the nonlinear filters of known noise statistics for two engagement scenarios. Simulation results showed that the proposed VB filters estimate well without knowing the noise statistics and they outperformed the adaptive filters based on MAP-MLE at the expense of slight higher computational cost.

## Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## Data availability

No data was used for the research described in the article.

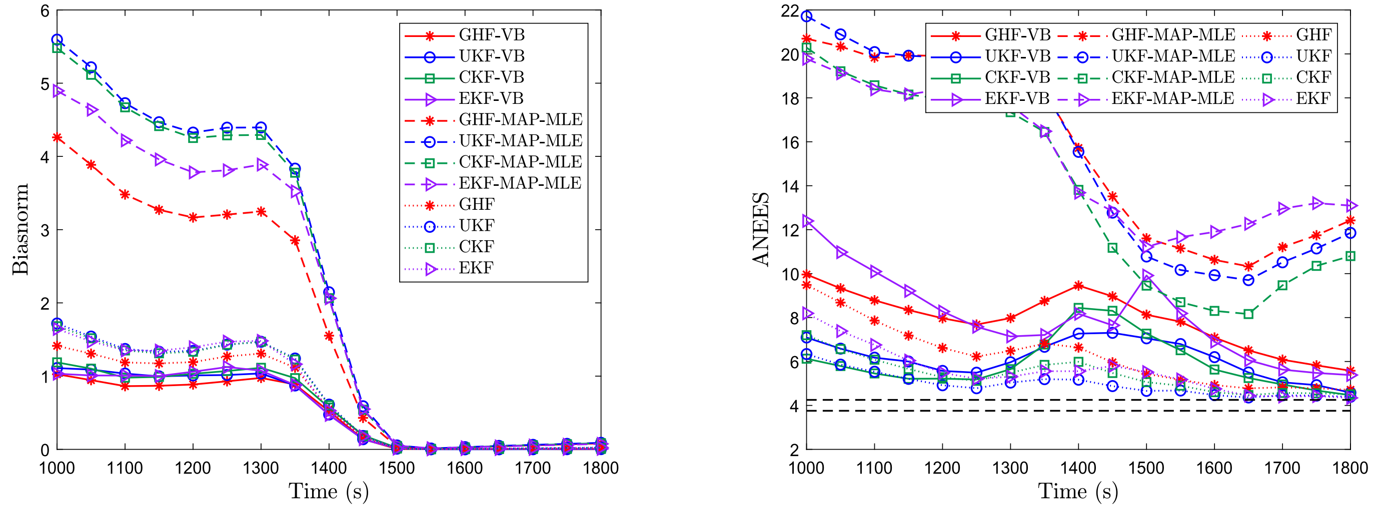
## Appendix A. Proof for Theorem 1

From (14), when $\theta = R'_k$,

$$\log q(R'_k) = E_{\mathcal{X}_k, r'}[\log p(\mathcal{X}_k, r', R'_k, \mathcal{Y}_{1:k-1}, \mathcal{Y}_k)] + c. \tag{A.1}$$

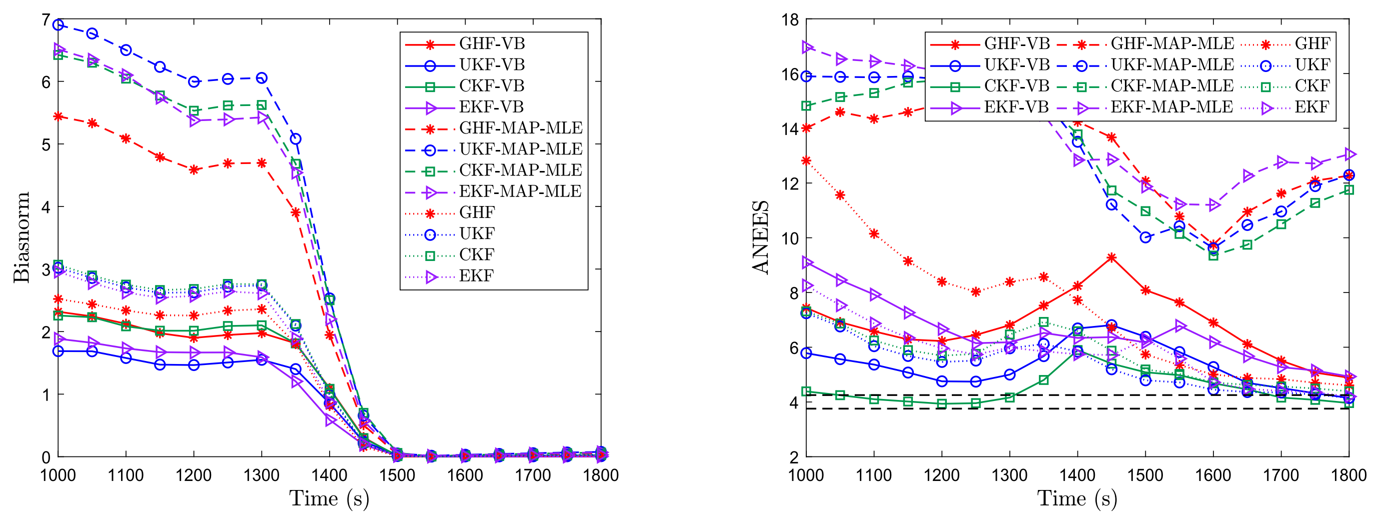
Using factorization with conditional probability,

$$\log q(R'_k) = E_{\mathcal{X}_k, r'} \log[p(\mathcal{Y}_k|\mathcal{X}_k, r', R'_k, \mathcal{Y}_{1:k-1})p(\mathcal{X}_k|r', R'_k, \mathcal{Y}_{1:k-1}) \\ p(r'|R'_k, \mathcal{Y}_{1:k-1})p(R'_k|\mathcal{Y}_{1:k-1})p(\mathcal{Y}_{1:k-1})] + c$$




(a) (b)

**Fig. 11.** (a) Bias norm, and (b) ANEES plots for time invariant measurement noise covariance, Scenario II.




(a) (b)

**Fig. 12.** (a) Bias norm, and (b) ANEES plots for time varying measurement noise covariance, Scenario II.

$$
\begin{aligned}
&= E_{\mathcal{X}_k, r'} \log[\mathcal{N}(\mathcal{Y}_k; h(\mathcal{X}_k) + r', R'_k)\mathcal{N}(\mathcal{X}_k; \hat{\mathcal{X}}_{k|k-1}, P_{k|k-1}) \\
&\quad \mathcal{N}(r'; \hat{r}_{k-1}, \hat{\alpha}_{k-1}R'_k) IW(R'_k; \hat{u}'_{k-1}, \hat{U}'_{k-1})] + c \\
&= -\frac{\hat{u}'_{k-1} + m + 3}{2} \log|R'_k| - \frac{1}{2} tr(\hat{U}'_{k-1} R'^{-1}_k) - E_{\mathcal{X}_k}[\frac{1}{2}\tilde{\mathcal{X}}_k^T P^{-1}_{k|k-1} \tilde{\mathcal{X}}_k] \\
&- E_{r'}[\frac{1}{2\hat{\alpha}_{k-1}} \tilde{r}^T (R'_k)^{-1} \tilde{r}] - E_{\mathcal{X}_k, r'}[\frac{1}{2}\tilde{\mathcal{Y}}_k^T R'^{-1}_k \tilde{\mathcal{Y}}_k] + c_{R_k},
\end{aligned}
\tag{A.2}
$$

where $\tilde{\mathcal{X}}_k = \mathcal{X}_k - \hat{\mathcal{X}}_{k|k-1}$ and $c_{R_k}$ is a constant. As $\tilde{\mathcal{Y}}_k^T R'^{-1}_k \tilde{\mathcal{Y}}_k$ is a scalar, using cyclic property of trace, we write,

$$
\tilde{\mathcal{Y}}_k^T R'^{-1}_k \tilde{\mathcal{Y}}_k = tr[\tilde{\mathcal{Y}}_k^T R'^{-1}_k \tilde{\mathcal{Y}}_k] = tr[\tilde{\mathcal{Y}}_k \tilde{\mathcal{Y}}_k^T R'^{-1}_k]. \tag{A.3}
$$

Similarly,

$$
\tilde{r}^T (R'_k)^{-1} \tilde{r} = tr[\tilde{r}\tilde{r}^T (R'_k)^{-1}], \tag{A.4}
$$

and

$$
\tilde{\mathcal{X}}_k^T P^{-1}_{k|k-1} \tilde{\mathcal{X}}_k = tr[\tilde{\mathcal{X}}_k \tilde{\mathcal{X}}_k^T P^{-1}_{k|k-1}]. \tag{A.5}
$$

As the expectation of (A.5) is a constant we can write using (A.2),

$$
\begin{aligned}
E_{\mathcal{X}_k, r'}[\log p(\mathcal{X}_k, r', R'_k, \mathcal{Y}_{1:k})] &= -\frac{\hat{u}'_{k-1} + m + 3}{2} \log|R'_k| - \frac{1}{2} tr(\hat{U}'_{k-1} R'^{-1}_k) \\
&- \frac{1}{2} tr[(B_k + D_k) R'^{-1}_k] + c_{R_k} \\
&= -\frac{\hat{u}'_{k-1} + m + 3}{2} \log|R'_k| - \frac{1}{2} tr[(\hat{U}'_{k-1} + B_k + D_k) R'^{-1}_k] + c_{R_k},
\end{aligned}
\tag{A.6}
$$

where $B_k = E_{\mathcal{X}_k, r'}[\tilde{\mathcal{Y}}_k \tilde{\mathcal{Y}}_k^T]$, and $D_k = \frac{1}{\hat{\alpha}_{k-1}} E_{r'}[\tilde{r}\tilde{r}^T]$. Comparing $q(R'_k) = IW(R'_k; \hat{u}'_k, \hat{U}'_k)$ with (A.6) we receive (20) and (21).

## Appendix B. Proof of Lemma 1

Adding and subtracting $h(\hat{\mathcal{X}}_{k|k-1})$, and $\hat{r}_{k-1}$ and rearranging (22),

$$
\begin{aligned}
B_k &= E_{\mathcal{X}_k, r'}[(\mathcal{Y}_k - h(\mathcal{X}_k) - r' - h(\hat{\mathcal{X}}_{k|k-1}) + h(\hat{\mathcal{X}}_{k|k-1}) - \hat{r}_{k-1} + \hat{r}_{k-1}) \\
&\quad (\mathcal{Y}_k - h(\mathcal{X}_k) - r' - h(\hat{\mathcal{X}}_{k|k-1}) + h(\hat{\mathcal{X}}_{k|k-1}) - \hat{r}_{k-1} + \hat{r}_{k-1})^T] \\
&= (\mathcal{Y}_k - h(\hat{\mathcal{X}}_{k|k-1}) - \hat{r}_{k-1})(\mathcal{Y}_k - h(\hat{\mathcal{X}}_{k|k-1}) - \hat{r}_{k-1})^T
\end{aligned}
$$

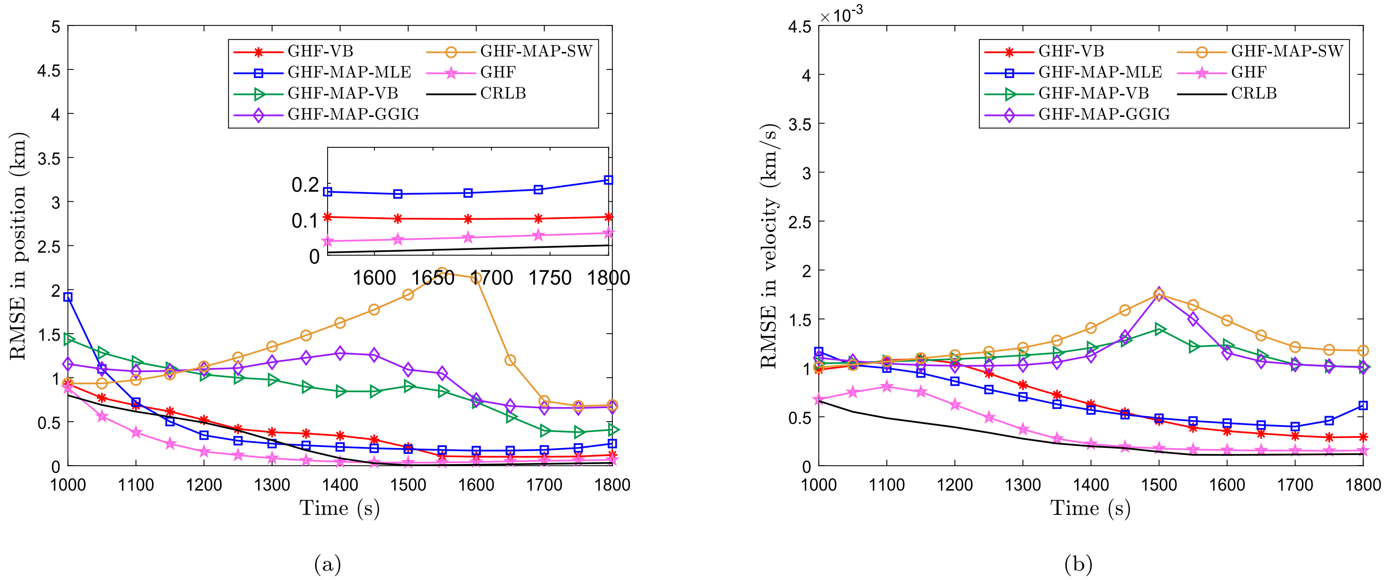



**Fig. 13.** RMSE in (a) position, and (b) velocity for time invariant measurement noise covariance, Scenario II.

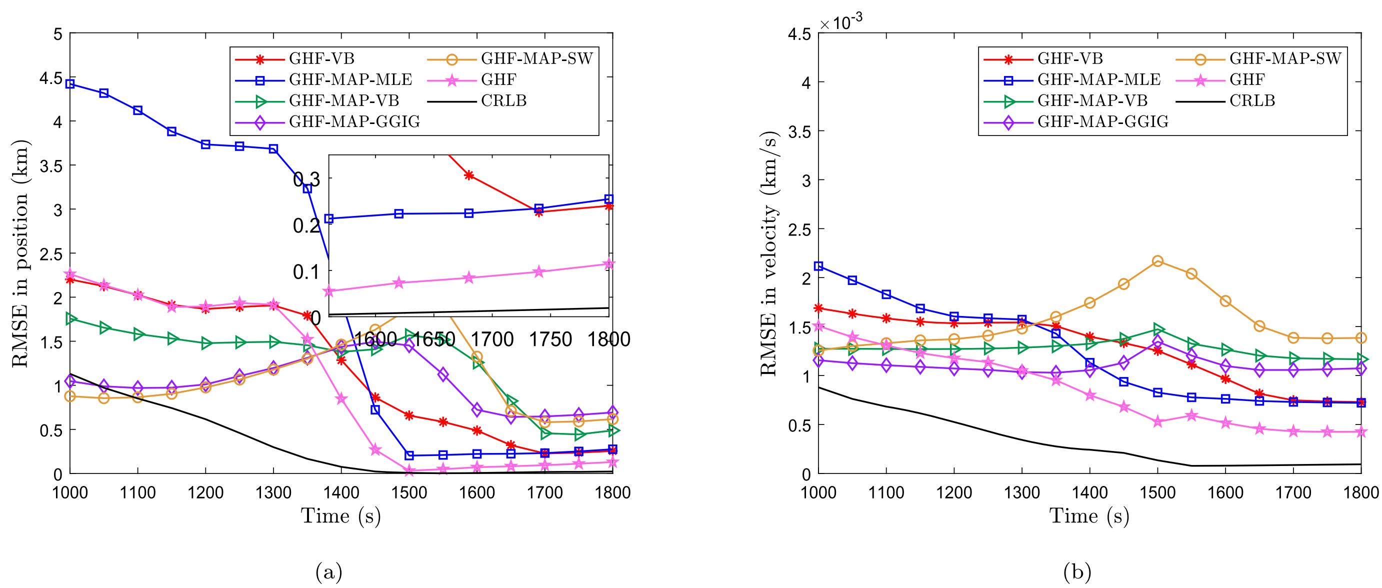



**Fig. 14.** RMSE in (a) position, and (b) velocity for time varying measurement noise covariance, Scenario II.

$$
\begin{aligned}
&+(\mathcal{Y}_k - h(\hat{\mathcal{X}}_{k|k-1}) - \hat{r}_{k-1})E_{\mathcal{X}_k,r'}[(h(\hat{\mathcal{X}}_{k|k-1}) - h(\mathcal{X}_k) + \hat{r}_{k-1} - r')^T]\\
&+E_{\mathcal{X}_k,r'}[(h(\hat{\mathcal{X}}_{k|k-1}) - h(\mathcal{X}_k) + \hat{r}_{k-1} - r')](\mathcal{Y}_k - h(\hat{\mathcal{X}}_{k|k-1}) - \hat{r}_{k-1})^T\\
&+E_{\mathcal{X}_k,r'}[(h(\hat{\mathcal{X}}_{k|k-1}) - h(\mathcal{X}_k) + \hat{r}_{k-1} - r')(h(\hat{\mathcal{X}}_{k|k-1}) - h(\mathcal{X}_k)\\
&+\hat{r}_{k-1} - r')^T]\\
&=(\mathcal{Y}_k - h(\hat{\mathcal{X}}_{k|k-1}) - \hat{r}_{k-1})(\mathcal{Y}_k - h(\hat{\mathcal{X}}_{k|k-1}) - \hat{r}_{k-1})^T\\
&+E_{\mathcal{X}_k,r'}[(h(\hat{\mathcal{X}}_{k|k-1}) - h(\mathcal{X}_k) + \hat{r}_{k-1} - r')(h(\hat{\mathcal{X}}_{k|k-1}) - h(\mathcal{X}_k)\\
&+\hat{r}_{k-1} - r')^T].
\end{aligned}
\tag{B.1}
$$

Now, the second term of the above equation can be written as,

$$
\begin{aligned}
&E_{\mathcal{X}_k,r'}[(h(\hat{\mathcal{X}}_{k|k-1}) - h(\mathcal{X}_k) + \hat{r}_{k-1} - r')(h(\hat{\mathcal{X}}_{k|k-1}) - h(\mathcal{X}_k) + \hat{r}_{k-1} - r')^T]\\
&=E_{\mathcal{X}_k}[(h(\hat{\mathcal{X}}_{k|k-1}) - h(\mathcal{X}_k))(h(\hat{\mathcal{X}}_{k|k-1}) - h(\mathcal{X}_k))^T]\\
&+E_{\mathcal{X}_k,r'}[(h(\hat{\mathcal{X}}_{k|k-1}) - h(\mathcal{X}_k))(\hat{r}_{k-1} - r')^T]\\
&+E_{\mathcal{X}_k,r'}[(\hat{r}_{k-1} - r')(h(\hat{\mathcal{X}}_{k|k-1}) - h(\mathcal{X}_k)^T] + E_{r'}[\tilde{r}\tilde{r}^T]\\
&=\sum_{j=1}^{N_s} w_j[\Upsilon_{j,k|k-1} - h(\hat{\mathcal{X}}_{k|k-1})][\Upsilon_{j,k|k-1} - h(\hat{\mathcal{X}}_{k|k-1})]^T + \hat{\alpha}_{k-1}\hat{R}_k,
\end{aligned}
\tag{B.2}
$$

where, $\hat{\alpha}_{k-1}$ is the confidence parameter and its recursion is provided in Theorem 3. Substituting (B.2) in (B.1) we receive (24).

## Appendix C. Proof of Lemma 2

Adding and subtracting $\hat{r}_k$ in (23) we have,

$$
\begin{aligned}
D_k &= \frac{1}{\hat{\alpha}_{k-1}} E_{r'}[(r' - \hat{r}_{k-1} - \hat{r}_k + \hat{r}_k)(r' - \hat{r}_{k-1} - \hat{r}_k + \hat{r}_k)^T]\\
&= \frac{1}{\hat{\alpha}_{k-1}} E_{r'}[(r' - \hat{r}_k)(r' - \hat{r}_k)^T + (r' - \hat{r}_k)(\hat{r}_k - \hat{r}_{k-1})^T\\
&+(\hat{r}_k - \hat{r}_{k-1})(r' - \hat{r}_k)^T + (\hat{r}_k - \hat{r}_{k-1})(\hat{r}_k - \hat{r}_{k-1})^T]\\
&= \frac{\hat{\alpha}_k}{\hat{\alpha}_{k-1}}\hat{R}_k + \frac{1}{\hat{\alpha}_{k-1}}(\hat{r}_k - \hat{r}_{k-1})(\hat{r}_k - \hat{r}_{k-1})^T.
\end{aligned}
\tag{C.1}
$$

## Appendix D. Proof for Theorem 2

Using (14), for $\theta = \mathcal{X}_k$ we have,

$$\log q(\mathcal{X}_k) = E_{R'_k, r'} \log[\mathcal{N}(\mathcal{Y}_k; h(\mathcal{X}_k) + r', R'_k)\mathcal{N}(\mathcal{X}_k; \hat{\mathcal{X}}_{k|k-1}, P_{k|k-1}) \mathcal{N}(r'; \hat{r}_{k-1}, \hat{\alpha}_{k-1} R'_{k-1}) IW(R'_{k-1}; \hat{u}'_{k-1}, \hat{U}'_{k-1})] + c = -\frac{1}{2} E_{R'_k, r'}[\tilde{\mathcal{Y}}_k^T R'^{-1}_k \tilde{\mathcal{Y}}_k] - \frac{1}{2}\tilde{\mathcal{X}}^T_{k|k-1} P^{-1}_{k|k-1} \tilde{\mathcal{X}}_{k|k-1} + c_{\mathcal{X}_k}, \quad (D.1)$$

where $\tilde{\mathcal{X}}_{k|k-1} = \mathcal{X}_k - \hat{\mathcal{X}}_{k|k-1}$. Again using trace and cyclic property we can write,

$$E_{R'_k, r'}[\tilde{\mathcal{Y}}_k^T R'^{-1}_k \tilde{\mathcal{Y}}_k] = tr(E_{r'}[\tilde{\mathcal{Y}}_k \tilde{\mathcal{Y}}_k^T] E[R'^{-1}_k]) = tr(E_{r'}[(\mathcal{Y}_k - h(\mathcal{X}_k) - r' + \hat{r}_k - \hat{r}_k)(\mathcal{Y}_k - h(\mathcal{X}_k) - r' + \hat{r}_k - \hat{r}_k)^T] E[R'^{-1}_k]) = tr((\mathcal{Y}_k - h(\mathcal{X}_k) - \hat{r}_k)(\mathcal{Y}_k - h(\mathcal{X}_k) - \hat{r}_k)^T + (\mathcal{Y}_k - h(\mathcal{X}_k) - \hat{r}_k) E[(\hat{r}_k - r')^T] + E[(\hat{r}_k - r')](\mathcal{Y}_k - h(\mathcal{X}_k) - \hat{r}_k)^T + E[(\hat{r}_k - r')(\hat{r}_k - r')^T] E[R'^{-1}_k])) = tr[((\mathcal{Y}_k - h(\mathcal{X}_k) - \hat{r}_k)(\mathcal{Y}_k - h(\mathcal{X}_k) - \hat{r}_k)^T + \hat{\alpha}_k R'_k) E[R'^{-1}_k]]. \quad (D.2)$$

As $tr(A + B) = tr(A) + tr(B)$ we can write,

$$E_{R'_k, r'}[\tilde{\mathcal{Y}}_k^T R'^{-1}_k \tilde{\mathcal{Y}}_k] = tr[(\mathcal{Y}_k - h(\mathcal{X}_k) - \hat{r}_k)(\mathcal{Y}_k - h(\mathcal{X}_k) - \hat{r}_k)^T E[R'^{-1}_k]] + tr[\hat{\alpha}_k R'_k E[R'^{-1}_k]]. \quad (D.3)$$

Using cyclic property of trace and removing trace as the content within trace becomes scalar, we have

$$\log q(\mathcal{X}_k) = -\frac{1}{2}(\mathcal{Y}_k - h(\mathcal{X}_k) - \hat{r}_k)^T E[R'^{-1}_k](\mathcal{Y}_k - h(\mathcal{X}_k) - \hat{r}_k) - \frac{1}{2}\tilde{\mathcal{X}}^T_{k|k-1} P^{-1}_{k|k-1}\tilde{\mathcal{X}}_{k|k-1} + c_{\mathcal{X}_k}, \quad (D.4)$$

or,

$$q(\mathcal{X}_k) = \frac{1}{c_k} p(\mathcal{Y}_k|\mathcal{X}_k) p(\mathcal{X}_k|\mathcal{Y}_{1:k-1}), \quad (D.5)$$

where the normalizing constant, $c_k = \int p(\mathcal{Y}_k|\mathcal{X}_k)p(\mathcal{X}_k|\mathcal{Y}_{1:k-1})d\mathcal{X}_k$. As we know, $p(\mathcal{Y}_k|\mathcal{X}_k) = \mathcal{N}(\mathcal{Y}_k; h(\mathcal{X}_k) + \hat{r}_k, \hat{R}_k)$ we can write,

$$\hat{R}_k = E[R'^{-1}_k]^{-1} = E[R'_k] = \frac{\hat{U}'_k}{\hat{u}'_k - m - 1}. \quad (D.6)$$

Using (21), (24) and (25) in (D.6) can arrive at (26).

## Appendix E. Proof for Theorem 3

Evaluating the solution of (14), when $\theta = r'$ and following (A.2),

$$\log q(r') = E_{\mathcal{X}_k, R'_k} \log[\mathcal{N}(\mathcal{Y}_k; h(\mathcal{X}_k) + r', R'_k)\mathcal{N}(\mathcal{X}_k; \hat{\mathcal{X}}_{k|k-1}, P_{k|k-1}) \mathcal{N}(r'; \hat{r}_{k-1}, \hat{\alpha}_{k-1} R'_k) IW(R'_k; \hat{u}'_{k-1}, \hat{U}'_{k-1})] + c_r. \quad (E.1)$$

Using (A.5), we can write

$$\log q(r') = E_{\mathcal{X}_k, R'_k} \log[\mathcal{N}(\mathcal{Y}_k; h(\mathcal{X}_k) + r', R'_k)\mathcal{N}(r'; \hat{r}_{k-1}, \hat{\alpha}_{k-1} R'_k)] + c_r = -\frac{1}{2} E_{\mathcal{X}_k, R'_k}[\tilde{\mathcal{Y}}_k^T R'^{-1}_k \tilde{\mathcal{Y}}_k] - \frac{1}{2\hat{\alpha}_{k-1}} \tilde{r}^T E_{R'_k}[(R'_k)^{-1}]\tilde{r} + c_r, \quad (E.2)$$

where $\tilde{\mathcal{Y}}_k = (\mathcal{Y}_k - h(\mathcal{X}_k) - r')$ and $\tilde{r} = (r' - \hat{r}_{k-1})$. Using (A.3) the expectation in the first term of (E.2) can be evaluated as

$$E_{\mathcal{X}_k, R'_k}[\tilde{\mathcal{Y}}_k^T R'^{-1}_k \tilde{\mathcal{Y}}_k] = tr\{E_{\mathcal{X}_k}[\tilde{\mathcal{Y}}_k \tilde{\mathcal{Y}}_k^T] E_{R'_k}[R'^{-1}_k]\} = tr\{E_{\mathcal{X}_k}[(\mathcal{Y}_k - h(\mathcal{X}_k) - r' + h(\hat{\mathcal{X}}_{k|k-1}) - h(\hat{\mathcal{X}}_{k|k-1}))(\mathcal{Y}_k - h(\mathcal{X}_k) - r' + h(\hat{\mathcal{X}}_{k|k-1}) - h(\hat{\mathcal{X}}_{k|k-1}))^T] E_{R'_k}[R'^{-1}_k]\}$$
$$= tr\{(\mathcal{Y}_k - h(\hat{\mathcal{X}}_{k|k-1}) - r')(\mathcal{Y}_k - h(\hat{\mathcal{X}}_{k|k-1}) - r')^T - E_{\mathcal{X}_k}[(\mathcal{Y}_k - h(\hat{\mathcal{X}}_{k|k-1}) - r')(h(\mathcal{X}_k) - h(\hat{\mathcal{X}}_{k|k-1}))^T] - E_{\mathcal{X}_k}[(h(\mathcal{X}_k) - h(\hat{\mathcal{X}}_{k|k-1}))(\mathcal{Y}_k - h(\hat{\mathcal{X}}_{k|k-1}) - r')^T] + E_{\mathcal{X}_k}[(h(\mathcal{X}_k) - h(\hat{\mathcal{X}}_{k|k-1}))(h(\mathcal{X}_k) - h(\hat{\mathcal{X}}_{k|k-1}))^T] E[R'^{-1}_k]\} = tr\{[(\mathcal{Y}_k - h(\hat{\mathcal{X}}_{k|k-1}) - r')(\mathcal{Y}_k - h(\hat{\mathcal{X}}_{k|k-1}) - r')^T] E[R'^{-1}_k]\} + c. \quad (E.3)$$

Again from (E.3) using cyclic property of trace we can rewrite the above equation as

$$E_{\mathcal{X}_k, R'_k}[\tilde{\mathcal{Y}}_k^T R'^{-1}_k \tilde{\mathcal{Y}}_k] = tr\{(\mathcal{Y}_k - h(\hat{\mathcal{X}}_{k|k-1}) - r')^T E[R'^{-1}_k](\mathcal{Y}_k - h(\hat{\mathcal{X}}_{k|k-1}) - r')\} + c. \quad (E.4)$$

We can remove the trace from the right hand side as the expression is a scalar. So, (E.2) becomes

$$\log q(r') = -\frac{1}{2}(\mathcal{Y}_k - h(\hat{\mathcal{X}}_{k|k-1}) - r')^T E[R'^{-1}_k](\mathcal{Y}_k - h(\hat{\mathcal{X}}_{k|k-1}) - r') - \frac{1}{2\hat{\alpha}_{k-1}}\tilde{r}^T E[R'^{-1}_k]\tilde{r} + c_r. \quad (E.5)$$

As we know the product of two Gaussian pdfs will be another Gaussian pdf with a resultant mean, $(m_1\sigma_2^2 + m_2\sigma_1^2)/(\sigma_1^2 + \sigma_2^2)$ and covariance, $(\sigma_1^2\sigma_2^2)/(\sigma_1^2 + \sigma_2^2)$, where $(m_1, m_2)$ and $(\sigma_1^2, \sigma_2^2)$ are the mean and covariance of the two Gaussian pdfs, from (E.5) we receive

$$\hat{r}_k = \frac{\hat{r}_{k-1}(E[R'^{-1}_k])^{-1} + (\mathcal{Y}_k - h(\hat{\mathcal{X}}_{k|k-1}))(\hat{\alpha}^{-1}_{k-1} E[R'^{-1}_k])^{-1}}{(E[R'^{-1}_k])^{-1} + (\hat{\alpha}^{-1}_{k-1} E[R'^{-1}_k])^{-1}} = \frac{(E[R'^{-1}_k])^{-1}(\hat{r}_{k-1} + (\mathcal{Y}_k - h(\hat{\mathcal{X}}_{k|k-1}))\hat{\alpha}_{k-1})}{(E[R'^{-1}_k])^{-1}(1 + \hat{\alpha}_{k-1})} = \frac{\hat{r}_{k-1} + (\mathcal{Y}_k - h(\hat{\mathcal{X}}_{k|k-1}))\hat{\alpha}_{k-1}}{(1 + \hat{\alpha}_{k-1})}. \quad (E.6)$$

Evaluating $\hat{\mathcal{X}}_{k|k-1}$ with the help of deterministic sample points we receive (27).

Similarly from (E.5) we get the estimated confidence parameter,

$$\hat{\alpha}_k \hat{R}_k = \frac{(E[R'^{-1}_k])^{-1}(E[R'^{-1}_k]\hat{\alpha}^{-1}_{k-1})^{-1}}{(E[R'^{-1}_k])^{-1}(1 + \hat{\alpha}_{k-1})} = \frac{(E[R'^{-1}_k])^{-1}\hat{\alpha}_{k-1}}{1 + \hat{\alpha}_{k-1}}. \quad (E.7)$$

As, we know $\hat{R}_k = (E[R'^{-1}_k])^{-1}$ from (D.6), so,

$$\hat{\alpha}_k = \frac{\hat{\alpha}_{k-1}}{\hat{\alpha}_{k-1} + 1}. \quad (E.8)$$

## Appendix F. Proof for Theorem 4

The Eqn. (D.4) can be considered as a merit function such that if we maximize it, the likelihood $p(\mathcal{Y}_k|\mathcal{X}_k)$ and the prediction probability density, $p(\mathcal{X}_k|\mathcal{Y}_{1:k-1})$ will be maximized thus we shall receive an optimal or suboptimal (for nonlinear system) posterior estimate. The first derivative of (D.4) becomes,

$$\frac{d \log q(\mathcal{X}_k)}{d\mathcal{X}_k} = -P^{-1}_{k|k-1}(\mathcal{X}_k - \hat{\mathcal{X}}_{k|k-1}) + \nabla_{\mathcal{X}_k} h^T(\mathcal{X}_k)\hat{R}^{-1}_k(\mathcal{Y}_k - h(\mathcal{X}_k) - \hat{r}_k), \quad (F.1)$$

where $\nabla_{\mathcal{X}_k} = \left[\frac{\partial}{\partial x_1} \quad \frac{\partial}{\partial x_2} \quad \cdots \quad \frac{\partial}{\partial x_n}\right]^T$. Let us say, $\Delta_{\mathcal{X}_k} = \mathcal{X}_k - \hat{\mathcal{X}}_{k|k-1}$ and $\Delta_{h(\mathcal{X}_k)} = h(\mathcal{X}_k) - E[h(\mathcal{X}_k)]$. Using statistical linearization, as in [67, pp. 172], $h(\mathcal{X}_k) = A\mathcal{X}_k + b$, where

$$A = E[\Delta_{\mathcal{X}_k}\Delta^T_{h(\mathcal{X}_k)}]^T P^{-1}_{k|k-1}, \quad (F.2)$$

and $b = E[h(\mathcal{X}_k)] - A\hat{\mathcal{X}}_{k|k-1}$. Therefore,

$$\nabla_{\mathcal{X}_k} h^T(\mathcal{X}_k) = [\nabla_{\mathcal{X}_k} \Delta_{\mathcal{X}_k}^T](P_{k|k-1}^{-1})^T E[\Delta_{\mathcal{X}_k} \Delta_{h(\mathcal{X}_k)}^T]$$
$$= (P_{k|k-1}^{-1})^T E[\Delta_{\mathcal{X}_k} \Delta_{h(\mathcal{X}_k)}^T]. \tag{F.3}$$

Using these in (F.1), we have

$$\frac{d \log q(\mathcal{X}_k)}{d\mathcal{X}_k} = -P_{k|k-1}^{-1}\Delta_{\mathcal{X}_k} + (P_{k|k-1}^{-1})^T E[\Delta_{\mathcal{X}_k} \Delta_{h(\mathcal{X}_k)}^T]\hat{R}_k^{-1}$$
$$(\mathcal{Y}_k - E[\Delta_{\mathcal{X}_k} \Delta_{h(\mathcal{X}_k)}^T]^T P_{k|k-1}^{-1}\mathcal{X}_k - E[h(\mathcal{X}_k)]$$
$$+ E[\Delta_{\mathcal{X}_k} \Delta_{h(\mathcal{X}_k)}^T]^T P_{k|k-1}^{-1}\hat{\mathcal{X}}_{k|k-1} - \hat{r}_k)$$
$$= -P_{k|k-1}^{-1}\Delta_{\mathcal{X}_k} + (P_{k|k-1}^{-1})^T E[\Delta_{\mathcal{X}_k} \Delta_{h(\mathcal{X}_k)}^T]\hat{R}_k^{-1}$$
$$(\mathcal{Y}_k - E[\Delta_{\mathcal{X}_k} \Delta_{h(\mathcal{X}_k)}^T]^T P_{k|k-1}^{-1}\Delta_{\mathcal{X}_k} - E[h(\mathcal{X}_k)] - \hat{r}_k)$$
$$= -P_{k|k-1}^{-1}\Delta_{\mathcal{X}_k} + (P_{k|k-1}^{-1})^T E[\Delta_{\mathcal{X}_k} \Delta_{h(\mathcal{X}_k)}^T]\hat{R}_k^{-1}(\mathcal{Y}_k - h(\hat{\mathcal{X}}_{k|k-1}) - \hat{r}_k)$$
$$- (P_{k|k-1}^{-1})^T E[\Delta_{\mathcal{X}_k} \Delta_{h(\mathcal{X}_k)}^T]\hat{R}_k^{-1} E[\Delta_{\mathcal{X}_k} \Delta_{h(\mathcal{X}_k)}^T]^T P_{k|k-1}^{-1}\Delta_{\mathcal{X}_k}. \tag{F.4}$$

To find the maxima of $\log q(\mathcal{X}_k)$ we have to equate the above equation to zero,

$$(P_{k|k-1}^{-1} + (P_{k|k-1}^{-1})^T E[\Delta_{\mathcal{X}_k} \Delta_{h(\mathcal{X}_k)}^T]\hat{R}_k^{-1} E[\Delta_{\mathcal{X}_k} \Delta_{h(\mathcal{X}_k)}^T]^T P_{k|k-1}^{-1})\Delta_{\mathcal{X}_k}$$
$$= (P_{k|k-1}^{-1})^T E[\Delta_{\mathcal{X}_k} \Delta_{h(\mathcal{X}_k)}^T]\hat{R}_k^{-1}(\mathcal{Y}_k - h(\hat{\mathcal{X}}_{k|k-1}) - \hat{r}_k). \tag{F.5}$$

Substituting $\Delta_{\mathcal{X}_k} = \mathcal{X}_k - \hat{\mathcal{X}}_{k|k-1}$

$$\text{or, } \mathcal{X}_k = \hat{\mathcal{X}}_{k|k-1}$$
$$+ (P_{k|k-1}^{-1} + (P_{k|k-1}^{-1})^T E[\Delta_{\mathcal{X}_k} \Delta_{h(\mathcal{X}_k)}^T]\hat{R}_k^{-1} E[\Delta_{\mathcal{X}_k} \Delta_{h(\mathcal{X}_k)}^T]^T P_{k|k-1}^{-1})^{-1}$$
$$(P_{k|k-1}^{-1})^T E[\Delta_{\mathcal{X}_k} \Delta_{h(\mathcal{X}_k)}^T]\hat{R}_k^{-1}(\mathcal{Y}_k - h(\hat{\mathcal{X}}_{k|k-1}) - \hat{r}_k). \tag{F.6}$$

Taking expectation on both the sides we could write,

$$\hat{\mathcal{X}}_{k|k} = \hat{\mathcal{X}}_{k|k-1} + K_k(\mathcal{Y}_k - h(\hat{\mathcal{X}}_{k|k-1}) - \hat{r}_k), \tag{F.7}$$

where

$$K_k = (P_{k|k-1}^{-1} + (P_{k|k-1}^{-1})^T E[\Delta_{\mathcal{X}_k} \Delta_{h(\mathcal{X}_k)}^T]\hat{R}_k^{-1} E[\Delta_{\mathcal{X}_k} \Delta_{h(\mathcal{X}_k)}^T]^T P_{k|k-1}^{-1})^{-1}$$
$$(P_{k|k-1}^{-1})^T E[\Delta_{\mathcal{X}_k} \Delta_{h(\mathcal{X}_k)}^T]\hat{R}_k^{-1}. \tag{F.8}$$

Further, using (F.2), the expression of $K_k$, the Kalman gain, reduces as,

$$K_k = (P_{k|k-1}^{-1} + A^T \hat{R}_k^{-1} A)^{-1} A^T \hat{R}_k^{-1},$$
$$\text{or, } K_k^{-1} = \hat{R}_k (A^T)^{-1} P_{k|k-1}^{-1} + \hat{R}_k (A^T)^{-1} A^T \hat{R}_k^{-1} A,$$
$$\text{or, } K_k = P_{k|k-1} A^T (\hat{R}_k + A P_{k|k-1} A^T)^{-1}. \tag{F.9}$$

Considering the first factor of the above equation

$$P_{k|k-1}A^T = P_{k|k-1}(E[\Delta_{\mathcal{X}_k} \Delta_{h(\mathcal{X}_k)}^T]^T P_{k|k-1}^{-1})^T$$
$$= P_{k|k-1}(P_{k|k-1}^{-1})^T E[\Delta_{\mathcal{X}_k} \Delta_{h(\mathcal{X}_k)}^T]$$
$$= E[\Delta_{\mathcal{X}_k} \Delta_{h(\mathcal{X}_k)}^T] = P_{\mathcal{X}\mathcal{Y}}. \tag{F.10}$$

For Gaussian filters, the expression of $P_{\mathcal{X}\mathcal{Y}}$ can be written using support points and weights as

$$P_{\mathcal{X}\mathcal{Y}} = \sum_{j=1}^{N_s} w_j (\mathrm{X}_{j,k|k-1} - \hat{\mathcal{X}}_{k|k-1})(\Upsilon_{j,k|k-1} - h(\hat{\mathcal{X}}_{k|k-1}))^T. \tag{F.11}$$

The second factor of (F.9) can be evaluated as,

$$\hat{R}_k + A P_{k|k-1} A^T = \hat{R}_k + E[(A\mathcal{X}_k - A\hat{\mathcal{X}}_{k|k-1})(A\mathcal{X}_k - A\hat{\mathcal{X}}_{k|k-1})^T]$$
$$= \hat{R}_k + E[(h(\mathcal{X}_k) - b - A\hat{\mathcal{X}}_{k|k-1})(h(\mathcal{X}_k) - b - A\hat{\mathcal{X}}_{k|k-1})^T]$$
$$= \hat{R}_k + E[(h(\mathcal{X}_k) - h(\hat{\mathcal{X}}_{k|k-1}))(h(\mathcal{X}_k) - h(\hat{\mathcal{X}}_{k|k-1}))^T]$$
$$= \hat{R}_k + \sum_{j=1}^{N_s} w_j (\Upsilon_{j,k|k-1} - h(\hat{\mathcal{X}}_{k|k-1}))(\Upsilon_{j,k|k-1} - h(\hat{\mathcal{X}}_{k|k-1}))^T = P_{\mathcal{Y}\mathcal{Y}}. \tag{F.12}$$

The first statement of the theorem is proved. The posterior error covariance,

$$P_{k|k} = E[(\mathcal{X}_k - \hat{\mathcal{X}}_{k|k})(\mathcal{X}_k - \hat{\mathcal{X}}_{k|k})^T]$$
$$= E[(\mathcal{X}_k - (\hat{\mathcal{X}}_{k|k-1} + K_k(\mathcal{Y}_k - h(\hat{\mathcal{X}}_{k|k-1}) - \hat{r}_k)))(\mathcal{X}_k - (\hat{\mathcal{X}}_{k|k-1}$$
$$+ K_k(\mathcal{Y}_k - h(\hat{\mathcal{X}}_{k|k-1}) - \hat{r}_k)))^T]$$
$$= E[(\mathcal{X}_k - \hat{\mathcal{X}}_{k|k-1})(\mathcal{X}_k - \hat{\mathcal{X}}_{k|k-1})^T] - K_k E[(\mathcal{Y}_k - h(\hat{\mathcal{X}}_{k|k-1}) - \hat{r}_k)$$
$$(\mathcal{Y}_k - h(\hat{\mathcal{X}}_{k|k-1}) - \hat{r}_k)^T]K_k^T + E[(\mathcal{X}_k - \hat{\mathcal{X}}_{k|k-1})(\mathcal{Y}_k - h(\hat{\mathcal{X}}_{k|k-1}) - \hat{r}_k)^T]K_k^T$$
$$+ K_k E[(\mathcal{Y}_k - h(\hat{\mathcal{X}}_{k|k-1}) - \hat{r}_k)(\mathcal{X}_k - \hat{\mathcal{X}}_{k|k-1})^T]$$
$$= P_{k|k-1} + K_k E[(h(\mathcal{X}_k) + v_{\theta_k} - h(\hat{\mathcal{X}}_{k|k-1}) - \hat{r}_k)(h(\mathcal{X}_k)$$
$$+ v_{\theta_k} - h(\hat{\mathcal{X}}_{k|k-1}) - \hat{r}_k)^T]K_k^T - E[(\mathcal{X}_k - \hat{\mathcal{X}}_{k|k-1})(h(\mathcal{X}_k) + v_{\theta_k}$$
$$- h(\hat{\mathcal{X}}_{k|k-1}) - \hat{r}_k)^T]K_k^T - K_k E[(h(\mathcal{X}_k) + v_{\theta_k} - h(\hat{\mathcal{X}}_{k|k-1}) - \hat{r}_k)(\mathcal{X}_k$$
$$- \hat{\mathcal{X}}_{k|k-1})^T]$$
$$= P_{k|k-1} + K_k(E[(h(\mathcal{X}_k) - h(\hat{\mathcal{X}}_{k|k-1}))(h(\mathcal{X}_k) - h(\hat{\mathcal{X}}_{k|k-1}))^T]$$
$$+ E[(v_{\theta_k} - \hat{r}_k)(v_{\theta_k} - \hat{r}_k)^T])K_k^T - E[(\mathcal{X}_k - \hat{\mathcal{X}}_{k|k-1})(h(\mathcal{X}_k)$$
$$- h(\hat{\mathcal{X}}_{k|k-1}))^T]K_k^T - K_k E[(h(\mathcal{X}_k) - h(\hat{\mathcal{X}}_{k|k-1}))(\mathcal{X}_k - \hat{\mathcal{X}}_{k|k-1})^T]$$
$$= P_{k|k-1} + K_k(\sum_{j=1}^{N_s} w_j(\Upsilon_{j,k|k-1} - h(\hat{\mathcal{X}}_{k|k-1}))(\Upsilon_{j,k|k-1} - h(\hat{\mathcal{X}}_{k|k-1}))^T$$
$$+ \hat{R}_k)K_k^T - P_{\mathcal{X}\mathcal{Y}}K_k^T - K_k P_{\mathcal{Y}\mathcal{X}}$$
$$= P_{k|k-1} + K_k P_{\mathcal{Y}\mathcal{Y}} K_k^T - P_{\mathcal{X}\mathcal{Y}}K_k^T - K_k P_{\mathcal{Y}\mathcal{X}}$$
$$= P_{k|k-1} + P_{\mathcal{X}\mathcal{Y}}(P_{\mathcal{X}\mathcal{Y}}P_{\mathcal{Y}\mathcal{Y}}^{-1})^T - P_{\mathcal{X}\mathcal{Y}}(P_{\mathcal{Y}\mathcal{Y}}^{-1})^T P_{\mathcal{X}\mathcal{Y}}^T - P_{\mathcal{X}\mathcal{Y}}P_{\mathcal{Y}\mathcal{Y}}^{-1}P_{\mathcal{Y}\mathcal{X}}$$
$$= P_{k|k-1} + P_{\mathcal{X}\mathcal{Y}}P_{\mathcal{Y}\mathcal{Y}}^{-1}P_{\mathcal{Y}\mathcal{X}} - P_{\mathcal{X}\mathcal{Y}}P_{\mathcal{Y}\mathcal{Y}}^{-1}P_{\mathcal{Y}\mathcal{Y}}(P_{\mathcal{X}\mathcal{Y}}P_{\mathcal{Y}\mathcal{Y}}^{-1})^T - P_{\mathcal{X}\mathcal{Y}}P_{\mathcal{Y}\mathcal{Y}}^{-1}P_{\mathcal{Y}\mathcal{X}}$$
$$= P_{k|k-1} - K_k P_{\mathcal{Y}\mathcal{Y}} K_k^T. \tag{F.13}$$

## Appendix G. Proof of Lemma 3

$$\hat{\mathcal{X}}_{k|k} - \mathcal{X}_k = \hat{\mathcal{X}}_{k|k-1} + K_k(\mathcal{Y}_k - h(\hat{\mathcal{X}}_{k|k-1}) - \hat{r}_k)$$
$$- (F\mathcal{X}_{k-1} + \omega_{k-1} - \mho_{k-1,k})$$
$$= F\hat{\mathcal{X}}_{k-1|k-1} - \mho_{k-1,k} + K_k(\mathcal{Y}_k - h(\hat{\mathcal{X}}_{k|k-1}) - \hat{r}_k)$$
$$- (F\mathcal{X}_{k-1} + \omega_{k-1} - \mho_{k-1,k})$$
$$= F(\hat{\mathcal{X}}_{k-1|k-1} - \mathcal{X}_{k-1}) - \omega_{k-1} + K_k(\mathcal{Y}_k - h(\hat{\mathcal{X}}_{k|k-1}) - \hat{r}_k). \tag{G.1}$$

Taking expectations on both sides and using (27) we get,

$$E[\hat{\mathcal{X}}_{k|k} - \mathcal{X}_k] = F(\hat{\mathcal{X}}_{k-1|k-1} - E[\mathcal{X}_{k-1}]) - E[\omega_{k-1}]$$
$$+ K_k(E[\mathcal{Y}_k] - h(\hat{\mathcal{X}}_{k|k-1}) - E[\hat{r}_k])$$
$$= K_k(h(\hat{\mathcal{X}}_{k|k-1}) + r - h(\hat{\mathcal{X}}_{k|k-1}) - E[\frac{\hat{r}_{k-1} + \hat{\alpha}_{k-1}(\mathcal{Y}_k - h(\hat{\mathcal{X}}_{k|k-1}))}{\hat{\alpha}_{k-1} + 1}])$$
$$= K_k(r - \frac{\hat{r}_{k-1} + \hat{\alpha}_{k-1}(E[\mathcal{Y}_k] - h(\hat{\mathcal{X}}_{k|k-1}))}{\hat{\alpha}_{k-1} + 1})$$
$$= K_k(r - \frac{\hat{r}_{k-1} + \hat{\alpha}_{k-1}(r)}{\hat{\alpha}_{k-1} + 1})$$
$$= K_k \frac{\hat{\alpha}_{k-1}r + r - \hat{r}_{k-1} - \hat{\alpha}_{k-1}r}{\hat{\alpha}_{k-1} + 1},$$
$$\text{or, } E[\hat{\mathcal{X}}_{k|k} - \mathcal{X}_k] = K_k \frac{r - \hat{r}_{k-1}}{\hat{\alpha}_{k-1} + 1}. \tag{G.2}$$

As the approximate mean, $\hat{r}_{k-1}$ approaches the true mean, $r$ using fixed point iteration (G.2) equates to zero.

## References


[1] B. Ristic, S. Arulampalam, N. Gordon, Beyond the Kalman Filter: Particle Filters for Tracking Applications, Artech House, 2003.
[2] P.H. Leong, S. Arulampalam, T.A. Lamahewa, T.D. Abhayapala, A Gaussian-sum based cubature Kalman filter for bearings-only tracking, IEEE Trans. Aerosp. Electron. Syst. 49 (2) (2013) 1161–1176.
[3] R. Radhakrishnan, A.K. Singh, S. Bhaumik, N.K. Tomar, Quadrature filters for underwater passive bearings-only target tracking, in: 2015 Sensor Signal Processing for Defence (SSPD), IEEE, 2015, pp. 1–5.
[4] R. Radhakrishnan, S. Bhaumik, N.K. Tomar, Gaussian sum shifted Rayleigh filter for underwater bearings-only target tracking problems, IEEE J. Ocean. Eng. 44 (2) (2018) 492–501.
[5] M. Mallick, A. Sinha, J. Liu, Enhancements to bearing-only filtering, in: 2017 20th International Conference on Information Fusion (Fusion), 2017, pp. 1–8.
[6] J. Annabattula, S.K. Rao, A.S.D. Murthy, K. Srikanth, R.P. Das, Multi-sensor submarine surveillance system using MGBEKF, Indian J. Sci. Technol. 8 (35) (2015) 1–5.
[7] R. Karlsson, F. Gustafsson, Recursive Bayesian estimation: bearings-only applications, IEE Proc. Radar Sonar Navig. 152 (5) (2005) 305–313.
[8] K. Kumar, S. Bhaumik, S. Arulampalam, Tracking an underwater object with unknown sensor noise covariance using orthogonal polynomial filters, Sensors 22 (13) (2022) 4970.
[9] A. Farina, Target tracking with bearings–only measurements, Signal Process. 78 (1) (1999) 61–78.
[10] S. Sadhu, S. Mondal, M. Srinivasan, T.K. Ghoshal, Sigma point Kalman filter for bearing only tracking, Signal Process. 86 (12) (2006) 3769–3777.
[11] Y. Bar-Shalom, X.R. Li, T. Kirubarajan, Estimation with Applications to Tracking and Navigation: Theory Algorithms and Software, John Wiley & Sons, 2001.
[12] A. Toloei, S. Niazi, State estimation for target tracking problems with nonlinear Kalman filter algorithms, Int. J. Comput. Appl. 98 (17) (2014).
[13] T. Song, J. Speyer, A stochastic analysis of a modified gain extended Kalman filter with applications to estimation with bearings only measurements, IEEE Trans. Autom. Control 30 (10) (1985) 940–949.
[14] S.K. Rao, Modified gain extended Kalman filter with application to bearings-only passive manoeuvring target tracking, IEE Proc. Radar Sonar Navig. 152 (4) (2005) 239–244.
[15] A. Jawahar, S.K. Rao, Modified polar extended Kalman filter (MP-EKF) for bearings-only target tracking, Indian J. Sci. Technol. 9 (26) (2016) 1–5.
[16] V.J. Aidala, Kalman filter behavior in bearings-only tracking applications, IEEE Trans. Aerosp. Electron. Syst. AES-15 (1) (1979) 29–39, https://doi.org/10.1109/TAES.1979.308793.
[17] S. Nardone, A. Lindgren, K. Gong, Fundamental properties and performance of conventional bearings-only target motion analysis, IEEE Trans. Autom. Control 29 (9) (1984) 775–787.
[18] Y. Xu, L. Liping, Single observer bearings-only tracking with the unscented Kalman filter, in: 2004 International Conference on Communications, Circuits and Systems (IEEE Cat. No. 04EX914), vol. 2, IEEE, 2004, pp. 901–905.
[19] G. Chalasani, S. Bhaumik, Bearing only tracking using Gauss-Hermite filter, in: 2012 7th IEEE Conference on Industrial Electronics and Applications (ICIEA), IEEE, 2012, pp. 1549–1554.
[20] S. Bhaumik, Swati, Cubature quadrature Kalman filter, IET Signal Process. 7 (7) (2013) 533–541.
[21] D. Wang, H. Hua, H. Cao, Algorithm of modified polar coordinates UKF for bearings-only target tracking, in: 2010 2nd International Conference on Future Computer and Communication, vol. 3, IEEE, 2010, pp. V3–557.
[22] R. Radhakrishnan, A.K. Singh, S. Bhaumik, N.K. Tomar, Multiple sparse-grid Gauss–Hermite filtering, Appl. Math. Model. 40 (7–8) (2016) 4441–4450.
[23] S. Bhaumik, P. Date, Nonlinear Estimation: Methods and Applications with Deterministic Sample Points, CRC Press, 2019.
[24] A. Dey, S. Sadhu, T.K. Ghoshal, Adaptive Gauss–Hermite filter for non-linear systems with unknown measurement noise covariance, IET Sci. Meas. Technol. 9 (8) (2015) 1007–1015.
[25] M. Das, S. Sadhu, T.K. Ghoshal, An adaptive sigma point filter for nonlinear filtering problems, International Journal of Electrical, Electronics and Computer Engineering 2 (2) (2013) 13–19.
[26] L. Zhao, X. Wang, An adaptive UKF with noise statistic estimator, in: 2009 4th IEEE Conference on Industrial Electronics and Applications, IEEE, 2009, pp. 614–618.
[27] L. Zhao, J. Wang, T. Yu, H. Jian, T. Liu, Design of adaptive robust square-root cubature Kalman filter with noise statistic estimator, Appl. Math. Comput. 256 (2015) 352–367.
[28] L.L. Scharf, D.L. Alspach, Non-linear state estimation in observation noise of unknown covariance, Int. J. Control 27 (2) (1978) 293–304.
[29] M.A. Sato, Online model selection based on the variational Bayes, Neural Comput. 13 (7) (2001) 1649–1681.
[30] R. Mehra, S. Seereeram, D. Bayard, F. Hadaegh, Adaptive Kalman filtering, failure detection and identification for spacecraft attitude estimation, in: Proceedings of International Conference on Control Applications, IEEE, 1995, pp. 176–181.
[31] T.G. Lee, Centralized Kalman filter with adaptive measurement fusion: its application to a GPS/SDINS integration system with an additional sensor, Int. J. Control. Autom. Syst. 1 (4) (2003) 444–452.
[32] D.J. Jwo, M.Y. Chen, C.H. Tseng, T.S. Cho, Adaptive and nonlinear Kalman filtering for GPS navigation processing, in: Kalman Filter: Recent Advances and Applications, vol. 19, 2009.
[33] A. Mohamed, K. Schwarz, Adaptive Kalman filtering for INS/GPS, J. Geod. 73 (4) (1999) 193–203.
[34] A. Almagbile, J. Wang, W. Ding, Evaluating the performances of adaptive Kalman filter methods in GPS/INS integration, J. Glob. Position. Syst. 9 (1) (2010) 33–40.
[35] Y. Huang, Y. Zhang, Z. Wu, N. Li, J. Chambers, A novel adaptive Kalman filter with inaccurate process and measurement noise covariance matrices, IEEE Trans. Autom. Control 63 (2) (2017) 594–601.
[36] B. Ristic, X. Wang, S. Arulampalam, Target motion analysis with unknown measurement noise variance, in: 2017 20th International Conference on Information Fusion (Fusion), IEEE, 2017, pp. 1–8.
[37] H. Jiang, Y. Cai, Adaptive fifth-degree cubature information filter for multi-sensor bearings-only tracking, Sensors 18 (10) (2018) 3241.
[38] S. Särkkä, A. Nummenmaa, Recursive noise adaptive Kalman filtering by variational Bayesian approximations, IEEE Trans. Autom. Control 54 (3) (2009) 596–600.
[39] Y. Huang, F. Zhu, G. Jia, Y. Zhang, A slide window variational adaptive Kalman filter, IEEE Trans. Circuits Syst. II, Express Briefs 67 (12) (2020) 3552–3556.
[40] C. Xue, Y. Huang, C. Zhao, X. Li, L. Mihaylova, Y. Li, J.A. Chambers, A Gaussian-generalized-inverse-Gaussian joint-distribution-based adaptive MSCKF for visual-inertial odometry navigation, IEEE Trans. Aerosp. Electron. Syst. 59 (3) (2022) 2307–2328.
[41] H. Xu, K. Duan, H. Yuan, W. Xie, Y. Wang, Black box variational inference to adaptive Kalman filter with unknown process noise covariance matrix, Signal Process. 169 (2020) 107413.
[42] C. Sun, Y. Zhang, G. Wang, W. Gao, A new variational Bayesian adaptive extended Kalman filter for cooperative navigation, Sensors 18 (8) (2018) 2538.
[43] Y. Huang, Y. Zhang, J.A. Chambers, A novel Kullback–Leibler divergence minimization-based adaptive student's t-filter, IEEE Trans. Signal Process. 67 (20) (2019) 5417–5432.
[44] Y. Huang, Y. Zhang, P. Shi, J. Chambers, Variational adaptive Kalman filter with Gaussian-inverse-Wishart mixture distribution, IEEE Trans. Autom. Control 66 (4) (2020) 1786–1793.
[45] Y. Huang, Y. Zhang, Y. Zhao, P. Shi, J.A. Chambers, A novel outlier-robust Kalman filtering framework based on statistical similarity measure, IEEE Trans. Autom. Control 66 (6) (2020) 2677–2692.
[46] H. Zhu, G. Zhang, Y. Li, H. Leung, An adaptive Kalman filter with inaccurate noise covariances in the presence of outliers, IEEE Trans. Autom. Control 67 (1) (2021) 374–381.
[47] M. Bai, Y. Huang, Y. Zhang, J. Chambers, Statistical similarity measure-based adaptive outlier-robust state estimator with applications, IEEE Trans. Autom. Control 67 (8) (2022) 4354–4361.
[48] Y. Huang, Y. Zhang, N. Li, J. Chambers, A robust Gaussian approximate filter for nonlinear systems with heavy tailed measurement noises, in: 2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), IEEE, 2016, pp. 4209–4213.
[49] S. Särkkä, 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.
[50] Z. Zhang, P. Dong, H. Tuo, G. Liu, H. Jia, Robust variational Bayesian adaptive cubature Kalman filtering algorithm for simultaneous localization and mapping with heavy-tailed noise, J. Shanghai Jiaotong Univ. 25 (1) (2020) 76–87.
[51] X. Dong, L. Chisci, Y. Cai, An adaptive variational Bayesian filter for nonlinear multi-sensor systems with unknown noise statistics, Signal Process. 179 (2021) 107837.
[52] J. Wang, T. Zhang, X. Xu, Y. Li, A variational Bayesian based strong tracking interpolatory cubature Kalman filter for maneuvering target tracking, IEEE Access 6 (2018) 52544–52560.
[53] H. Lin, C. Hu, Variational inference based distributed noise adaptive Bayesian filter, Signal Process. 178 (2021) 107775.
[54] A. O'Hagan, J.J. Forster, Kendall's Advanced Theory of Statistics, Volume 2B: Bayesian Inference, vol. 2, University of Southampton Institutional Repository, 2004.
[55] K.P. Murphy, Conjugate Bayesian analysis of the Gaussian distribution, Tech. Rep., University of British Columbia, Vancouver, BC, Canada, 2007.
[56] C.M. Bishop, N.M. Nasrabadi, Pattern Recognition and Machine Learning, vol. 4, Springer, 2006.
[57] D.G. Tzikas, A.C. Likas, N.P. Galatsanos, The variational approximation for Bayesian inference, IEEE Signal Process. Mag. 25 (6) (2008) 131–146.
[58] S.J. Julier, J.K. Uhlmann, Unscented filtering and nonlinear estimation, Proc. IEEE 92 (3) (2004) 401–422.
[59] R. Radhakrishnan, S. Bhaumik, N.K. Tomar, Gaussian sum shifted Rayleigh filter for underwater bearings-only target tracking problems, IEEE J. Ocean. Eng. 44 (2) (2018) 492–501.
[60] I. Arasaratnam, S. Haykin, Cubature Kalman filters, IEEE Trans. Autom. Control 54 (6) (2009) 1254–1269.
[61] K. Ito, K. Xiong, Gaussian filters for nonlinear filtering problems, IEEE Trans. Autom. Control 45 (5) (2000) 910–927.
[62] S.I. Amari, Natural gradient works efficiently in learning, Neural Comput. 10 (2) (1998) 251–276.

[63] V. Berinde, Iterative Approximation of Fixed Points, vol. 1912, Springer, 2007.
[64] V. Berinde, On the stability of some fixed point procedures, Buletinul Ştiinţific al Universitatii Baia Mare, Seria B, Fascicola Matematică-Informatică (2002) 7–14.
[65] F. Akutsah, O.K. Narain, K. Afassinou, A.A. Mebawondu, An iterative scheme for fixed point problems, Adv. Math. Sci. J. 10 (5) (2021) 2295–2316.
[66] J. Wang, Stochastic modeling for real-time kinematic GPS/GLONASS positioning, J. Inst. Navig. 46 (4) (1999) 297–305.
[67] S. Särkkä, L. Svensson, Bayesian Filtering and Smoothing, 2nd edition, vol. 17, Cambridge University Press, 2023.

**Shreya Das** received B.Sc. degree in Physics, B.Tech, and M.Tech degree in Electrical Engineering from the University of Calcutta in 2015, 2018, and 2020 respectively. Presently, she is pursuing Ph.D. degree with the Department of Electrical Engineering, Indian Institute of Technology Patna. Her topics of interests include state estimation and bearings-only tracking.

**Kundan Kumar** received B.Tech degree in Electrical Engineering from the Muzaffarpur Institute of Technology, Muzaffarpur and PhD degree in Electrical Engineering from Indian Institute of Technology Patna in 2016 and 2023, respectively. Currently, he is working as a Postdoctoral Researcher at Aalto University, Finland. His topics of interests include Bayesian filtering and smoothing, target tracking and estimation in cyber-physical systems.

**Shovan Bhaumik** received the B.Sc. degree in Physics from the University of Calcutta, in 1999, and the B.Tech. degree in Electronics and Instrumentation Engineering and the M.E. and Ph.D. degrees in Electrical Engineering from Jadavpur University, in 2002, 2004, and 2009, respectively. From 2007 to 2009, he was a Research Engineer with the GE Global Research, GE India Technology Center, Bengaluru, India. He is currently an Associate Professor with the Department of Electrical Engineering, Indian Institute of Technology Patna. His research activities are supported by the Naval Research Board, Department of Science and Technology, Ministry of Electronics and Information Technology, and Defence Research and Development Organization, India. He has authored a book Nonlinear Estimation Methods and Applications with Deterministic Sample Points (CRC Press). His main research interests include nonlinear estimation, statistical signal processing, and underwater target tracking. He was a recipient of Young Faculty Research Fellowship Award from the Ministry of Electronics and Information Technology, Government of India. Currently he is serving as an Editorial Board member of Digital Signal Processing journal.