Random Matrix Based Extended Target Tracking with Orientation: A New Model and Inference

Barkın Tuncer, *Student Member, IEEE,* Emre Özkan, *Member, IEEE,*

***Abstract*—In this study, we propose a novel extended target tracking algorithm which is capable of representing the extent of dynamic objects as an ellipsoid with a time-varying orientation angle. A diagonal positive semi-definite matrix is defined to model objects' extent within the random matrix framework where the diagonal elements have inverse-Gamma priors. The resulting measurement equation is non-linear in the state variables, and it is not possible to find a closed-form analytical expression for the true posterior because of the absence of conjugacy. We use the variational Bayes technique to perform approximate inference, where the Kullback-Leibler divergence between the true and the approximate posterior is minimized by performing fixed-point iterations. The update equations are easy to implement, and the algorithm can be used in real-time tracking applications. We illustrate the performance of the method in simulations and experiments with real data. The proposed method outperforms the state-of-the-art methods when compared with respect to accuracy and robustness.**

***Index Terms*—Target tracking, extended target tracking, orientation, variational Bayes,**

I. Introduction

EXTENDED target tracking (ETT) problem involves processing multiple measurements that belong to a single target at each scan. In contrast to conventional tracking algorithms, which rely on point target assumption, ETT algorithms aim at estimating the target extent, which can be defined as the target-specific region that generates multiple measurements. Previous studies in the ETT literature can be broadly categorized into four groups:

- Simple shape models
- Random matrix (RM) based models
- Random hyper-surface (RHS) based models
- Mixture models

A simple approach to ETT involves assuming a predefined shape for the extent/contour of the object such as a circle, a rectangle, or a line [1]–[3]. The most common approaches in the literature utilize RM models, where the target extent is represented by an ellipse [4]–[7]. Alternatively, RHS models are suggested in [8], [9]. More recently, Gaussian Process (GP) based models are proposed for extended target tracking [10]–[12]. Another fold of studies focuses on modeling the target extent with multiple ellipses [13]–[16].

RM models represent the elliptical extent of a target by an unknown positive semi-definite matrix (PSDM). In the Bayesian framework, inverse-Wishart (IW) distribution defines a conjugate prior for PSDMs. In RM based ETT models, the overall target state is composed of a Gaussian kinematic state vector and an IW distributed extent matrix. Several algorithms are proposed to approximate or compute the posterior of this augmented state. In [4], exact inference is performed by neglecting the measurement noise and exploiting the resulting conjugacy. This model is restrictive in the sense that the kinematic state vector has to be composed of the target's position and higher-order spatial components such as velocity and acceleration. Koch's RM model is later improved in [5] to account for the measurement noise in the updates at the expense of exact inference. The update equations in [5] are intuitive, but the approximations are difficult to quantify theoretically. This problem is later addressed by [6], where the variational Bayes technique is used to obtain approximate posteriors.

None of the aforementioned RM models is capable of tracking the heading angle of an extended target. They instead rely on a forgetting factor to forget the sufficient statistics of the unknown extent matrix in time to account for the changes in the orientation of the target. Such an approach is problematic as it aims to discard the information collected in the past and try to explain the change in the orientation as the change in the target shape. There are earlier studies that aim at estimating the orientation angle of elliptical objects [17]–[20]. In [20], the orientation of the target is estimated by using the information that is obtained from the trajectory of the target. The methods that are proposed in [17]–[19] express the unknown extent parametrically and perform inference using extended Kalman filters together with pseudo-measurements. In these approaches, an explicit nonlinear measurement equation is derived where the kinematic and shape parameters are related to measurements by multiplicative random variables. The inference in [19] involves second-order Taylor series approximation to approximate the pseudo-measurement covariance matrix. In [18], the authors improved the algorithm in [19] further and eliminated the need for computing Hessian matrices. Instead, they showed that the expectation and the covariance of the pseudo-measurements could be approximated from the original measurement covariance matrix. In [17], the predicted measurement covariance matrix approximation is calculated more precisely.

There are several drawbacks of the methods in [17]–[19]. The models used in these methods involve a multiplicative noise term, which introduces additional non-linearity in the problem, and it makes performing inference more difficult. The methods require a pseudo-measurement which must be constructed from the original measurements to update kinematic and extent states separately. The measurements collected at one time instant must be processed sequentially. Changing the order of the measurements causes minor changes in the performance [17]. The state variables corresponding to the semi-axes lengths, which are positive by definition, are distributed with Gaussian distributions whose support covers both positive and negative real line. In some cases, it can be challenging to reflect available information into the priors defined in [17], which may cause a collapse in the extent estimates in the subsequent measurement updates.

In this work, we propose a novel RM model that defines a Gaussian prior for the heading angle and an inverse Gamma

Authors are with the Department of Electrical and Electronics Engineering, Middle East Technical University, Ankara, Turkey e-mail: barkint@metu.edu.tr, emreo@metu.edu.tr.

prior for the extent parameters which guarantee positive semi-definiteness. It is not possible to find a closed-form expression for the resulting posterior hence we utilize the variational Bayes technique to perform approximate inference. The variational Bayes technique is successfully applied to complex filtering problems in the literature to obtain approximate posteriors [6], [21]–[25].

The contributions of this manuscript can be listed as follows.

- We provide a novel solution that can track the orientation of a target and estimate its extent jointly.
- The proposed solution utilizes appropriate priors, which are defined over non-negative real numbers, for the unknown extent parameters.
- The problem formulation does not rely on multiplicative noise terms or pseudo-measurements.
- The measurement update can be performed by processing multiple measurements as a batch. The update does not depend on the order of the measurements.
- The uncertainty in the orientation and shape parameters can be expressed separately.
- The inference is performed via the well-known variational Bayes technique.

The paper is organized as follows. In Section II, we formulate the problem of joint shape estimation and tracking of elliptical objects with time-varying orientation. In the subsequent sections, we present the inference method. The measurement update is derived in Section III. Time update is given in Section IV. Section III-A5 presents the closed form expressions for the expectations required for the variational measurement update. A closer look at a single measurement update and its comparison with the state-of-the-art extended Kalman filter (EKF) algorithm is given in Section V. Lastly, the results are presented and discussed in Section VI.

TABLE I: NOTATIONS

- Set of real matrices of size $m \times n$ is represented with $\mathbb{R}^{m\times n}$.
- Set of symmetric positive definite and semi-definite matrices of size $n \times n$ is represented with $\mathbb{S}^n_{++}$ and $\mathbb{S}^n_{+}$, respectively.
- $\mathcal{N}(\mathbf{x};\boldsymbol{\mu},\Sigma)$ represents the multivariate Gaussian distributions with mean vector $\boldsymbol{\mu} \in \mathbb{R}^{n_x}$ and covariance matrix $\Sigma \in \mathbb{S}^{n_x}_{++}$,
- $\mathcal{IG}(\sigma;\alpha,\beta)$ represents the inverse Gamma distribution over the scalar $\sigma \in \mathbb{R}^+$ with shape and scale parameters $\alpha \in \mathbb{R}^+$ and $\beta \in \mathbb{R}^+$ respectively,
$$\mathcal{IG}(\sigma;\alpha,\beta) = \frac{\beta^\alpha}{\Gamma(\alpha)}\sigma^{-\alpha-1}\exp\left(-\frac{\beta}{\sigma}\right),$$
- For a measurement number of $a \in \mathcal{Z}^+$, $\mathbf{Y}_k$ represents the measurement set $\{\mathbf{y}^1_k,\dots,\mathbf{y}^a_k\}$ at time k.
- For any number $a \in \mathcal{Z}^+$, $\mathbf{Z}_k$ represents the variable set $\{\mathbf{z}^1_k,\dots,\mathbf{z}^a_k\}$ at time k.
- $\mathbf{r}_k$ represent the vector $[r^1_k,\dots,r^a_k]^T$ with size $a \in \mathcal{Z}^+$.
- KL denotes the Kullback-Leibler divergence between two distributions $q(x)$ and $p(x)$,
$$\mathrm{KL}\big(q(x)||p(x)\big) \triangleq \int q(x)\log\left(\frac{q(x)}{p(x)}\right)dx.$$
- $\det(A)$ denotes the determinant of matrix A.
- c_ϕ is a generic constant that denotes the constant terms with respect to variable ϕ in an equation.
- $\mathrm{Tr}\left[A\right] = \sum_{i=1}^n a_{ii}$ where a_{ii} is the i^{th} diagonal element of $A \in \mathbb{R}^{n\times n}$.
- $\mathbb{E}_p$ denotes the expectation operator, and p emphasizes the underlying probability distribution(s).
- $h.o.t.$ stands for higher-order terms.

II. Problem Definition

Consider a single target from which multiple measurements are generated in a single scan. Assume that the state of the extended target consists of the kinematic state $\mathbf{x}_k \in \mathbb{R}^{n_x}$, the orientation angle $\theta_k \in \mathbb{R}$, and the diagonal positive definite target extent matrix $X_k \in \mathbb{R}^{n_y\times n_y}$, $X_k \triangleq \operatorname{diag}\left(\sigma^1_k,\sigma^2_k,\dots,\sigma^{n_y}_k\right)$, where n_x and n_y represent the dimensions of the kinematic target state and the measurements, respectively. Given $\mathbf{x}_k, X_k$ and θ_k, the measurements generated by the target are assumed to be independent and identically distributed,

$$p(\mathbf{y}^j_k|\mathbf{x}_k,X_k,\theta_k) \sim \mathcal{N}\big(\mathbf{y}^j_k; H\mathbf{x}_k, sT_{\theta_k}X_kT^T_{\theta_k}+R\big), \tag{1}$$

where

- $\mathbf{y}^j_k \in \mathbb{R}^{n_y}$ is the j$^{\text{th}}$ measurement at time k,
- $H \in \mathbb{R}^{n_y\times n_x}$ is the measurement matrix,
- $R \in \mathbb{R}^{n_y\times n_y}$ is the positive definite measurement noise covariance matrix,
- $s \in \mathbb{R}^+$ is the scaling parameter,
- $T_{\theta_k} \in \mathbb{R}^{n_y\times n_y}$ is the rotation matrix which performs a rotation around the center of the target by the orientation angle θ_k. T_{θ_k} satisfies the well known properties of the rotation matrices such as $T^{-1}_{\theta_k} = T^T_{\theta_k}$, and $\det(T_{\theta_k}) = 1$. In 2D, it is defined as,
$$T_{\theta_k} \triangleq \begin{bmatrix}\cos(\theta_k) & -\sin(\theta_k)\\ \sin(\theta_k) & \cos(\theta_k)\end{bmatrix}. \tag{2}$$

In the Bayesian filtering framework, we aim at estimating the unknown variables $\mathbf{x}_k, \theta_k$ and X_k given the measurements collected up to and including time k. To achieve this, we define appropriate priors for the unknowns, and try to compute their posteriors in a recursive manner. This is generally performed by repeating two recursive steps:

- **Time Update (Prediction):** At any time step k, the predictive distribution $p(\mathbf{x}_k,X_k,\theta_k|\mathbf{Y}_{1:k-1})$ is computed according to Chapman-Kolmogorov equation by using the posterior from the previous time step $k-1$, and the transition density induced by the system dynamics.
- **Measurement Update (Correction):** When the new measurements $\mathbf{Y}_k$ are available, the posterior distribution $p(\mathbf{x}_k,X_k,\theta_k|\mathbf{Y}_{1:k})$ is computed by using the Bayes' rule. In this step, the predictive distribution $p(\mathbf{x}_k,X_k,\theta_k|\mathbf{Y}_{1:k-1})$ is used as the prior.

Unfortunately, it is not possible to obtain a closed form expression for the posterior in our problem. Therefore, we will look for an approximate analytical solution using a variational approximation.

Before introducing the details of this approximation, we will first define the prior distributions of the unknown variables. The joint prior distribution of the kinematic state, the extent, and the orientation is specified as

$$\begin{aligned}p(\mathbf{x}_0,X_0,\theta_0) = \mathcal{N}\left(\mathbf{x}_0;\hat{\mathbf{x}}_0,P_0\right) \times \prod_{i=1}^{n_y}\mathcal{IG}\left(\sigma^i_0;\alpha^i_0,\beta^i_0\right)\\ \times\mathcal{N}\left(\theta_0;\hat{\theta}_0,\Theta_0\right),\end{aligned} \tag{3}$$

where $X_0 \triangleq \operatorname{diag}\left(\sigma^1_0,\sigma^2_0,\dots,\sigma^{n_y}_0\right)$, and $\mathcal{IG}\left(\sigma^i_0;\alpha^i_0,\beta^i_0\right)$ denotes the inverse Gamma distribution. Here, $\hat{\mathbf{x}}_0$ and P_0 are the prior mean and covariance matrix of the Gaussian kinematic state vector $\hat{\mathbf{x}}_0$, respectively. $\hat{\theta}_0$ and Θ_0 are the prior mean and covariance matrix of the orientation angle θ_0. Please see Table I for the complete list of the notations.

In the following sections, we will describe the *Measurement Update* and *Time Update* steps of the proposed method in detail.

III. Measurement Update

Suppose at time k, we have the following conditional predicted density for the kinematic, orientation and extent states:

$$p(\mathbf{x}_k, X_k, \theta_k|\mathbf{Y}_{1:k-1}) = \mathcal{N}(\mathbf{x}_k; \hat{\mathbf{x}}_{k|k-1}, P_{k|k-1}) \times \prod_{i=1}^{n_y} \mathcal{IG}(\sigma^i_{k|k-1}; \alpha^i_{k|k-1}, \beta^i_{k|k-1}) \times \mathcal{N}\left(\theta_k; \hat{\theta}_{k|k-1}, \Theta_{k|k-1}\right). \quad (4)$$

The left-hand side of the above expression is conditioned on the measurements up to and including time instant $k-1$. The predicted mean and covariance of the Gaussian state vector is represented by $\hat{\mathbf{x}}_{k|k-1}$ and $P_{k|k-1}$, respectively. $\alpha^i_{k|k-1}$ and $\beta^i_{k|k-1}$ are the shape and the scale variables for the i^{th} diagonal element of the inverse Gamma distributed extent state X_k.
When the measurements $\mathbf{Y}_k$ are available at time k, the posterior distribution can be computed using Bayes' rule

$$p(\mathbf{x}_k, X_k, \theta_k|\mathbf{Y}_{1:k}) = \frac{p(\mathbf{Y}_k|\mathbf{x}_k, X_k, \theta_k)p(\mathbf{x}_k, X_k, \theta_k|\mathbf{Y}_{1:k-1})}{p(\mathbf{Y}_k|\mathbf{Y}_{1:k-1})}. \quad (5)$$

By assuming conditional independence of the measurements at time k, the measurement likelihood can be factorized as

$$p(\mathbf{Y}_k|\mathbf{x}_k, X_k, \theta_k) = \prod_{j=1}^{m_k} p(\mathbf{y}^j_k|\mathbf{x}_k, X_k, \theta_k) = \prod_{j=1}^{m_k} \mathcal{N}(\mathbf{y}^j_k; H\mathbf{x}_k, sT_{\theta_k}X_kT^T_{\theta_k} + R). \quad (6)$$

In the following, we will describe a variational inference based approximation method to estimate the posterior distribution using the likelihood function in (6).

A. Variational Inference

An approximate analytical solution for the posterior density in (5) can be obtained as a product of factorized probability density functions (PDFs) using a variational approximation. Before we present the details, we need to define additional instrumental variables to address the absence of conjugacy caused by the additive measurement noise covariance term R in the likelihood. We will call these variables noise-free measurements [6], and denote them with $\mathbf{Z}_k = \{\mathbf{z}^j_k\}_{j=1}^{m_k}$. By using $\mathbf{Z}_k$, the measurement likelihood in (6) can be expressed as

$$\mathcal{N}(\mathbf{y}^j_k; H\mathbf{x}_k, sT_{\theta_k}X_kT^T_{\theta_k} + R) = \int \mathcal{N}(\mathbf{y}^j_k;\mathbf{z}^j_k, R)\mathcal{N}(\mathbf{z}^j_k; H\mathbf{x}_k, sT_{\theta_k}X_kT^T_{\theta_k})d\mathbf{z}^j_k \quad (7)$$

for a single measurement. One can interpret equation (7) as the marginalization of the following joint density

$$p(\mathbf{y}^j_k, \mathbf{z}^j_k|\mathbf{x}_k, X_k) = \mathcal{N}(\mathbf{y}^j_k; \mathbf{z}^j_k, R) \times \mathcal{N}(\mathbf{z}^j_k; H\mathbf{x}_k, sT_{\theta_k}X_kT^T_{\theta_k}). \quad (8)$$

Let us include the instrumental variable $\mathbf{Z}_k$ in the posterior. Later, it will be marginalized out to obtain the posterior of the states

$$p(\mathbf{x}_k, X_k, \theta_k, \mathbf{Z}_k|\mathbf{Y}_{1:k}) \approx q_{\mathbf{x}}(\mathbf{x}_k)q_X(X_k)q_\theta(\theta_k)q_{\mathbf{Z}}(\mathbf{Z}_k). \quad (9)$$

Here, $q_{\mathbf{Z}}(\mathbf{Z}_k)$ denotes the approximate density of the instrumental variable $\mathbf{Z}_k$. The idea of variational approximation is to seek factorized densities whose product minimizes the following cost function.

$$\hat{q}_{\mathbf{x}}, \hat{q}_X, \hat{q}_\theta, \hat{q}_{\mathbf{Z}} = \arg \min_{q_{\mathbf{x}}, q_X, q_\theta, q_{\mathbf{Z}}} \mathrm{KL}\big(q_{\mathbf{x}}(\mathbf{x}_k)q_X(X_k)q_\theta(\theta_k)q_{\mathbf{Z}}(\mathbf{Z}_k) || p(\mathbf{x}_k, X_k, \theta_k, \mathbf{Z}_k|\mathbf{Y}_{1:k})\big). \quad (10)$$

The solution of the optimization problem (10) satisfies the following set of equations [26, Ch. 10]:

$$\log \hat{q}_{\mathbf{x}}(\mathbf{x}_k) = \mathbb{E}_{\hat{q}_X, \hat{q}_\theta, \hat{q}_{\mathbf{Z}}}\left[\log p(\mathbf{x}_k, X_k, \theta_k, \mathbf{Z}_k, \mathbf{Y}_k|\mathbf{Y}_{1:k-1})\right] + c_{\mathbf{x}}, \quad (11a)$$

$$\log \hat{q}_X(X_k) = \mathbb{E}_{\hat{q}_{\mathbf{x}}, \hat{q}_\theta, \hat{q}_{\mathbf{Z}}}\left[\log p(\mathbf{x}_k, X_k, \theta_k, \mathbf{Z}_k, \mathbf{Y}_k|\mathbf{Y}_{1:k-1})\right] + c_X, \quad (11b)$$

$$\log \hat{q}_\theta(\theta_k) = \mathbb{E}_{\hat{q}_{\mathbf{x}}, \hat{q}_X, \hat{q}_{\mathbf{Z}}}\left[\log p(\mathbf{x}_k, X_k, \theta_k, \mathbf{Z}_k, \mathbf{Y}_k|\mathbf{Y}_{1:k-1})\right] + c_\theta, \quad (11c)$$

$$\log \hat{q}_{\mathbf{Z}}(\mathbf{Z}_k) = \mathbb{E}_{\hat{q}_{\mathbf{x}}, \hat{q}_X, \hat{q}_\theta}\left[\log p(\mathbf{x}_k, X_k, \theta_k, \mathbf{Z}_k, \mathbf{Y}_k|, \mathbf{Y}_{1:k-1})\right] + c_{\mathbf{Z}}. \quad (11d)$$

$c_{\mathbf{x}}$, c_X, c_θ and $c_{\mathbf{Z}}$ denote the constant terms with respect to the corresponding variables here, and in the sequel. The joint density $p(\mathbf{x}_k, X_k, \theta_k, \mathbf{Z}_k, \mathbf{Y}_k|\mathbf{Y}_{1:k-1})$ in (11) can be written explicitly as

$$\begin{aligned} &p(\mathbf{x}_k, X_k, \theta_k, \mathbf{Z}_k, \mathbf{Y}_k|, \mathbf{Y}_{1:k-1}) \\ =&p(\mathbf{Y}_k|\mathbf{Z}_k)p(\mathbf{Z}_k|\mathbf{x}_k, X_k, \theta_k)p(\mathbf{x}_k, X_k, \theta_k|\mathbf{Y}_{1:k-1}) \\ =&\left(\prod_{j=1}^{m_k} \mathcal{N}(\mathbf{y}^j_k; \mathbf{z}^j_k, R)\right)\left(\prod_{j=1}^{m_k} \mathcal{N}(\mathbf{z}^j_k; H\mathbf{x}_k, sT_{\theta_k}X_kT^T_{\theta_k})\right) \\ &\times \mathcal{N}(\mathbf{x}_k; \hat{\mathbf{x}}_{k|k-1}, P_{k|k-1}) \prod_{i=1}^{n_y} \mathcal{IG}(\sigma^i_{k|k-1}; \alpha^i_{k|k-1}, \beta^i_{k|k-1}) \\ &\times \mathcal{N}(\theta_k; \hat{\theta}_{k|k-1}, \Theta_{k|k-1}). \end{aligned} \quad (12)$$

The optimization problem (10) can be solved by fixed-point iterations [26, Ch. 10]. Each iteration is performed by updating only one factorized density in (9) while keeping all other densities fixed to their last estimated values. The update equations of the approximate densities in the $(\ell+1)^{\text{th}}$ iteration will be given in the following subsections. To simplify the notations, $p\big(\mathbf{x}, X, \theta_k, \mathbf{Z}_k, \mathbf{Y}_k|\mathbf{Y}_{1:k-1}\big)$ is denoted as $P_{\mathbf{x},X,\theta,\mathbf{Z},\mathbf{Y}}$ in the sequel.

1) Computation of $q^{(\ell+1)}_{\mathbf{x}}(\cdot)$: Substituting the previous estimates of the factorized densities into equation (11a) yields

$$\log q^{(\ell+1)}_{\mathbf{x}}\big(\mathbf{x}_k\big) = \mathbb{E}_{q^{(\ell)}_X, q^{(\ell)}_\theta, q^{(\ell)}_{\mathbf{Z}}}\left[\log P_{\mathbf{x},X,\theta,\mathbf{Z},\mathbf{Y}}\right] + c_{\mathbf{x}}. \quad (13)$$

The expectation above can be simplified as

$$\begin{aligned} &\mathbb{E}_{q^{(\ell)}_X, q^{(\ell)}_\theta, q^{(\ell)}_{\mathbf{Z}}}\left[\log P_{\mathbf{x},X,\theta,\mathbf{Z},\mathbf{Y}}\right] \\ &= \mathbb{E}_{q^{(\ell)}_X, q^{(\ell)}_\theta, q^{(\ell)}_{\mathbf{Z}}}\left[\log P\big(\mathbf{Z}_k|\mathbf{x}_k, X_k, \theta_k\big)\right] \\ &\quad + \log \mathcal{N}(\mathbf{x}_k; \hat{\mathbf{x}}_{k|k-1}, P_{k|k-1}) + c_{\mathbf{x}} \quad (14a) \\ &= \sum_{j=1}^{m_k} -0.5\,\mathrm{Tr}\left[\big(\overline{\mathbf{z}^j_k} - H\mathbf{x}_k\big)\big(\overline{\mathbf{z}^j_k} - H\mathbf{x}_k\big)^T \times \mathbb{E}_{q^{(\ell)}_{\mathbf{X}}, q^{(\ell)}_\theta}\left[(sT_{\theta_k}X_kT^T_{\theta_k})^{-1}\right]\right] \end{aligned}$$

$$+\log\mathcal{N}(\mathbf{x}_k;\hat{\mathbf{x}}_{k|k-1},P_{k|k-1})+c_{\mathbf{x}}$$
$$= -0.5\,\mathrm{Tr}\Bigg[m_k\big(\overline{\mathbf{z}_k}-H\mathbf{x}_k\big)\big(\overline{\mathbf{z}_k}-H\mathbf{x}_k\big)^T \times \mathbb{E}_{q_{\mathbf{X}}^{(\ell)},q_\theta^{(\ell)}}\left[(sT_{\theta_k}X_kT_{\theta_k}^T)^{-1}\right]\Bigg] + \log\mathcal{N}(\mathbf{x}_k;\hat{\mathbf{x}}_{k|k-1},P_{k|k-1})+c_{\mathbf{x}} \tag{14b}$$
$$= \log\mathcal{N}\Big(\overline{\mathbf{z}_k};H\mathbf{x}_k,\frac{\mathbb{E}_{q_{\mathbf{X}}^{(\ell)},q_\theta^{(\ell)}}\left[(sT_{\theta_k}X_kT_{\theta_k}^T)^{-1}\right]^{-1}}{m_k}\Big) + \log\mathcal{N}(\mathbf{x}_k;\hat{\mathbf{x}}_{k|k-1},P_{k|k-1})+c_{\mathbf{x}} \tag{14c}$$

It can be seen from (14c) that $q_{\mathbf{x}}^{(\ell+1)}(\mathbf{x}_k)$ is a Gaussian PDF with mean vector $\hat{\mathbf{x}}_{k|k}^{(\ell+1)}$ and covariance $P_{k|k}^{(\ell+1)}$

$$q_{\mathbf{x}}^{(\ell+1)}(\mathbf{x}_k)=\mathcal{N}(\mathbf{x}_k;\hat{\mathbf{x}}_{k|k}^{(\ell+1)},P_{k|k}^{(\ell+1)}), \tag{15}$$

where

$$\hat{\mathbf{x}}_{k|k}^{(\ell+1)} = P_{k|k}^{(\ell+1)}(P_{k|k-1}^{-1}\hat{\mathbf{x}}_{k|k-1} + m_kH^T\mathbb{E}_{q_{\mathbf{X}}^{(\ell)},q_\theta^{(\ell)}}\left[(sT_{\theta_k}X_kT_{\theta_k}^T)^{-1}\right]\bar{\mathbf{z}}_k), \tag{16a}$$
$$P_{k|k}^{(\ell+1)} = (P_{k|k-1}^{-1}+m_kH^T\mathbb{E}_{q_{\mathbf{X}}^{(\ell)},q_\theta^{(\ell)}}\left[(sT_{\theta_k}X_kT_{\theta_k}^T)^{-1}\right]H)^{-1}, \tag{16b}$$

and $\overline{\mathbf{z}}_k \triangleq \frac{1}{m_k}\sum_{j=1}^{m_k}\mathbb{E}_{q_{\mathbf{Z}}^{(\ell)}}[\mathbf{z}_k^j]$.

2) Computation of $q_X^{(\ell+1)}(\cdot)$: Substituting the factorized densities from the previous variational iteration into equation (11b) yields

$$\log q_X^{(\ell+1)}\big(X_k\big)=\mathbb{E}_{q_{\mathbf{x}}^{(\ell)},q_\theta^{(\ell)},q_{\mathbf{Z}}^{(\ell)}}\left[\log P_{\mathbf{x},X,\theta,\mathbf{Z},\mathbf{Y}}\right]+c_X \tag{17}$$

Substituting (12) into (17) and grouping the constant terms with respect to X_k results in

$$\mathbb{E}_{q_{\mathbf{x}}^{(\ell)},q_\theta^{(\ell)},q_{\mathbf{Z}}^{(\ell)}}\left[\log P_{\mathbf{x},X,\theta,\mathbf{Z},\mathbf{Y}}\right] = \mathbb{E}_{q_{\mathbf{x}}^{(\ell)},q_\theta^{(\ell)},q_{\mathbf{Z}}^{(\ell)}}\left[\log P\big(\mathbf{Z}_k|\mathbf{x}_k,X_k,\theta_k\big)\right] + \sum_{i=1}^{n_y}\log\mathcal{IG}(\sigma_{k|k-1}^i;\alpha_{k|k-1}^i,\beta_{k|k-1}^i)+c_X \tag{18a}$$
$$= \frac{-m_k}{2}\log|sX_k| - \frac{1}{2}\,\mathrm{Tr}\Bigg[\sum_{j=1}^{m_k}\mathbb{E}_{q_{\mathbf{x}}^{(\ell)},q_\theta^{(\ell)},q_{\mathbf{Z}}^{(\ell)}}\left[(\mathbf{z}_k^j-H\mathbf{x}_k)(\mathbf{z}_k^j-H\mathbf{x}_k)^T \times(sT_{\theta_k}X_kT_{\theta_k}^T)^{-1}\right]\Bigg] + \sum_{i=1}^{n_y}\log\mathcal{IG}(\sigma_{k|k-1}^i;\alpha_{k|k-1}^i,\beta_{k|k-1}^i)+c_X. \tag{18b}$$

Consequently, the approximate posterior density q_X follows an inverse-Gamma distribution

$$q_X^{(\ell+1)}(X_k)=\prod_{i=1}^{n_y}\mathcal{IG}(\sigma_{k|k}^{i,(\ell+1)};\alpha_{k|k}^{i,(\ell+1)},\beta_{k|k}^{i,(\ell+1)}), \tag{19}$$

where

$$\alpha_{k|k}^{i,(\ell+1)}=\alpha_{k|k-1}^i+0.5m_k \tag{20a}$$
$$\beta_{k|k}^{i,(\ell+1)}=\beta_{k|k-1}^i+\frac{1}{2s}\sum_{j=1}^{m_k}\mathbb{E}_{q_{\mathbf{x}},q_{\mathbf{Z}},q_\theta}[\widetilde{\mathbf{z}}_k^j(\widetilde{\mathbf{z}}_k^j)^T]_{ii}, \tag{20b}$$

and $\widetilde{\mathbf{z}}_k^j \triangleq T_{\theta_k}^T(\mathbf{z}_k^j-H\mathbf{x}_k)$.

3) Computation of $q_{\mathbf{Z}}^{(\ell+1)}(\cdot)$: Substituting the factorized densities from the previous variational iteration into equation (11c) yields

$$\log q_{\mathbf{Z}}^{(\ell+1)}\big(\mathbf{Z}_k\big)=\mathbb{E}_{q_{\mathbf{x}}^{(\ell)},q_X^{(\ell)},q_\theta^{(\ell)}}\left[\log P_{\mathbf{x},X,\theta,\mathbf{Z},\mathbf{Y}}\right]+c_{\mathbf{Z}}. \tag{21}$$

The expectation above can be expressed as

$$\mathbb{E}_{q_{\mathbf{x}}^{(\ell)},q_X^{(\ell)},q_\theta^{(\ell)}}\left[\log P_{\mathbf{x},X,\theta,\mathbf{Z},\mathbf{Y}}\right] = \mathbb{E}_{q_{\mathbf{x}}^{(\ell)},q_X^{(\ell)},q_\theta^{(\ell)}}\left[\log P\big(\mathbf{Z}_k|\mathbf{x}_k,X_k,\theta_k\big)\right] + \sum_{j=1}^{m_k}\log\mathcal{N}(\mathbf{y}_k^j;\mathbf{z}_k^j,R)+c_{\mathbf{Z}} \tag{22a}$$
$$= \sum_{j=1}^{m_k}\frac{-1}{2}\,\mathrm{Tr}\Bigg[(\mathbf{z}_k^j-H\overline{\mathbf{x}_k})(\mathbf{z}_k^j-H\overline{\mathbf{x}_k})^T \times\mathbb{E}_{q_{\mathbf{X}}^{(\ell)},q_\theta^{(\ell)}}\left[(sT_{\theta_k}X_kT_{\theta_k}^T)^{-1}\right]\Bigg] + \sum_{j=1}^{m_k}\log\mathcal{N}(\mathbf{y}_k^j;\mathbf{z}_k^j,R)+c_{\mathbf{Z}}. \tag{22b}$$

Update equations for the approximate posterior density $q_{\mathbf{Z}}$ in the $(\ell+1)^{\text{th}}$ iteration are given by

$$q_{\mathbf{Z}}^{(\ell+1)}(\mathbf{Z}_k)=\prod_{j=1}^{m_k}\mathcal{N}(\mathbf{z}_k^j;\hat{\mathbf{z}}_k^{j,(\ell+1)},\Sigma_k^{z,(\ell+1)}), \tag{23}$$

where

$$\hat{\mathbf{z}}_k^{j,(\ell+1)} = \Sigma_k^{z,(\ell+1)}\Big(\mathbb{E}_{q_{\mathbf{X}}^{(\ell)},q_\theta^{(\ell)}}\left[(sT_{\theta_k}X_kT_{\theta_k}^T)^{-1}\right]H\mathbb{E}_{q_{\mathbf{x}}^{(\ell)}}[\mathbf{x}_k] + R^{-1}\mathbf{y}_k^j\Big)$$
$$\Sigma_k^{z,(\ell+1)} = \Big(\mathbb{E}_{q_{\mathbf{X}}^{(\ell)},q_\theta^{(\ell)}}\left[(sT_{\theta_k}X_kT_{\theta_k}^T)^{-1}\right]+R^{-1}\Big)^{-1} \tag{24a}$$

4) Computation of $q_\theta^{(\ell+1)}(\cdot)$: The update equations for $q_\theta^{(\ell+1)}(\cdot)$ is obtained by substituting the factorized densities from the previous variational iteration into equation (11d)

$$\log q_\theta^{(\ell+1)}\big(\theta_k\big)=\mathbb{E}_{q_{\mathbf{x}}^{(\ell)},q_X^{(\ell)},q_{\mathbf{Z}}^{(\ell)}}\left[\log P_{\mathbf{x},X,\theta,\mathbf{Z},\mathbf{Y}}\right]+c_\theta. \tag{25}$$

Substituting (12) into (25) and grouping the constant terms with respect to θ_k results in

$$\mathbb{E}_{q_{\mathbf{x}}^{(\ell)},q_X^{(\ell)},q_{\mathbf{Z}}^{(\ell)}}\left[\log P_{\mathbf{x},X,\theta,\mathbf{Z},\mathbf{Y}}\right] = \mathbb{E}_{q_{\mathbf{x}}^{(\ell)},q_X^{(\ell)},q_{\mathbf{Z}}^{(\ell)}}\left[\log P\big(\mathbf{Z}_k|\mathbf{x}_k,X_k,\theta_k\big)\right] + \log\mathcal{N}(\theta_k;\hat{\theta}_{k|k-1},\Theta_{k|k-1})+c_\theta \tag{26a}$$
$$= \frac{-1}{2}\sum_{j=1}^{m_k}\mathbb{E}_{q_{\mathbf{x}}^{(\ell)},q_X^{(\ell)},q_{\mathbf{Z}}^{(\ell)}}\Bigg[\mathrm{Tr}\big[(\mathbf{z}_k^j-H\mathbf{x}_k)(\mathbf{z}_k^j-H\mathbf{x}_k)^T \times(sT_{\theta_k}X_kT_{\theta_k}^T)^{-1}\big]\Bigg] + \log\mathcal{N}(\theta_k;\hat{\theta}_{k|k-1},\Theta_{k|k-1})+c_\theta \tag{26b}$$

Unfortunately, it is not possible to obtain an exact compact form PDF for $q_\theta^{(\ell+1)}\big(\theta_k\big)$ because of the non-linearities involved in (26b). To address this issue, we will make a first order approximation of the non-linear function $f(\theta_k) \triangleq T_{\theta_k}^T(\mathbf{z}_k^j - H\mathbf{x}_k)$ using its Taylor series expansion around $\hat{\theta}_{k|k}^{(\ell)}$,

$$f(\theta_k) = f(\hat{\theta}_{k|k}^{(\ell)}) + \nabla f(\hat{\theta}_{k|k}^{(\ell)})(\theta_k - \hat{\theta}_{k|k}^{(\ell)}) + h.o.t., \tag{27}$$

$$\text{where} \quad \nabla f(\hat{\theta}_{k|k}^{(\ell)}) \triangleq \left.\frac{\partial f}{\partial \theta_k}\right|_{\theta_k = \hat{\theta}_{k|k}^{(\ell)}}.$$

By plugging in the first order approximation of $f(\theta_k)$ into (26b), the expectation term can be written as

$$\mathbb{E}_{q_{\mathbf{x}}^{(\ell)}, q_X^{(\ell)}, q_{\mathbf{Z}}^{(\ell)}}\left[(a - b\theta_k)^T (sX)^{-1}(a - b\theta_k)\right],$$

where

$$a \triangleq \left[f(\hat{\theta}_{k|k}^{(\ell)}) - \nabla f(\hat{\theta}_{k|k}^{(\ell)})\hat{\theta}_{k|k}^{(\ell)}\right], \tag{28}$$

$$b \triangleq -\nabla f(\hat{\theta}_{k|k}^{(\ell)}). \tag{29}$$

Through algebraic manipulations, $q_\theta^{(\ell+1)}(\theta_k)$ can be expressed as a Gaussian PDF with mean vector $\hat{\theta}_{k|k}^{(\ell+1)}$ and covariance $\Theta_{k|k}^{(\ell+1)}$,

$$q_\theta^{(\ell+1)}(\theta_k) = \mathcal{N}(\theta_k; \hat{\theta}_{k|k}^{(\ell+1)}, \Theta_{k|k}^{(\ell+1)}), \tag{30}$$

where

$$\hat{\theta}_{k|k}^{(\ell+1)} = \Theta_{k|k}^{(\ell+1)}\left(\Theta_{k|k-1}^{-1}\hat{\theta}_{k|k-1} + \delta\right), \tag{31a}$$

$$\Theta_{k|k}^{(\ell+1)} = \left(\Theta_{k|k-1}^{-1} + \Delta\right)^{-1}, \tag{31b}$$

$$\begin{aligned}\delta = \sum_{j=1}^{m_k} &\mathrm{Tr}\left[\overline{sX_k}^{-1}(T'_{\hat{\theta}_{k|k}^{(\ell)}})^T \overline{\left(\mathbf{z}_k^j - H\mathbf{x}_k\right)(\,\cdot\,)^T}(T'_{\hat{\theta}_{k|k}^{(\ell)}})\hat{\theta}_{k|k}^{(\ell)}\right] \\ &- \mathrm{Tr}\left[\overline{sX_k}^{-1} T_{\hat{\theta}_{k|k}^{(\ell)}}^T \overline{\left(\mathbf{z}_k^j - H\mathbf{x}_k\right)(\,\cdot\,)^T}(T'_{\hat{\theta}_{k|k}^{(\ell)}})\right],\end{aligned} \tag{31c}$$

$$\Delta = \sum_{j=1}^{m_k} \mathrm{Tr}\left[\overline{sX_k}^{-1}(T'_{\hat{\theta}_{k|k}^{(\ell)}})^T \overline{\left(\mathbf{z}_k^j - H\mathbf{x}_k\right)(\,\cdot\,)^T}(T'_{\hat{\theta}_{k|k}^{(\ell)}})\right], \tag{31d}$$

$$\text{and} \qquad T'_{\hat{\theta}_{k|k}^{(\ell)}} \triangleq \left.\frac{\partial T_{\theta_k}}{\partial \theta_k}\right|_{\theta_k = \hat{\theta}_{k|k}^{(\ell)}}.$$

The derivations of δ and Δ are given in Appendix A. By using the expressions derived so far, we can set up variational iterations to find the approximate posteriors $q_{\mathbf{x}}$, q_X, and $q_{\mathbf{Z}}$. $\mathbf{Z}_k$ can be marginalized out from the joint density, and an approximation for $p(\mathbf{x}_k, X_k, \theta_k|, \mathbf{Y}_{1:k})$ is obtained.

5) Expectation Calculations: The relevant expectations in the variational iterations can be computed by using the following set of equations:

$$\mathbb{E}_{q_{\mathbf{x}}^{(\ell)}}[\mathbf{x}_k] = \hat{\mathbf{x}}_{k|k}^{(\ell)}, \quad \mathbb{E}_{q_{\mathbf{Z}}^{(\ell)}}[\mathbf{z}_k^j] = \hat{\mathbf{z}}_k^{j,(\ell)},$$

$$\mathbb{E}_{q_X^{(\ell)}}[(sX_k)^{-1}] = \mathrm{diag}\left(\frac{s\beta^{1,\ell}}{\alpha^{1,\ell}}, \frac{s\beta^{2,\ell}}{\alpha^{2,\ell}}, \ldots, \frac{s\beta^{n_y,\ell}}{\alpha^{n_y,\ell}}\right)^{-1},$$

$$\begin{aligned}&\mathbb{E}_{q_x^{(\ell)}, q_{\mathbf{Z}}^{(\ell)}, q_\theta^{(\ell)}}\left[T_{\theta_k}^T(\mathbf{z}_k^j - H\mathbf{x}_k)(\mathbf{z}_k^j - H\mathbf{x}_k)^T T_{\theta_k}\right] \\ &= \mathbb{E}_{q_\theta^{(\ell)}}\left[T_{\theta_k}^T\left(\left(\hat{\mathbf{z}}_k^{j,(\ell)} - H\mathbf{x}_{k|k}^{(\ell)}\right)\left(\hat{\mathbf{z}}_k^{j,(\ell)} - H\mathbf{x}_{k|k}^{(\ell)}\right)^T \right.\right. \\ &\left.\left. + HP_{k|k}^{(\ell)}(H)^T + \Sigma_k^{z,\ell,(\ell)}\right)T_{\theta_k}\right],\end{aligned}$$

$$\begin{aligned}\overline{\left(\mathbf{z}_k^j - H\mathbf{x}_k\right)(\,\cdot\,)^T} &= \left(\hat{\mathbf{z}}_k^{j,(\ell)} - H\mathbf{x}_{k|k}^{(\ell)}\right)\left(\hat{\mathbf{z}}_k^{j,(\ell)} - H\mathbf{x}_{k|k}^{(\ell)}\right)^T \\ &+ HP_{k|k}^{(\ell)}H^T + \Sigma_k^{z,\ell}.\end{aligned}$$

The initial conditions for the quantities can be chosen as $\hat{\mathbf{z}}_k^{j,(0)} = \mathbf{y}_k^j$, $\Sigma_k^{z,\ell,(0)} = sX_{k|k-1}$, $\mathbf{x}_{k|k}^{(0)} = \hat{\mathbf{x}}_{k|k-1}$, $P_{k|k}^{(0)} = P_{k|k-1}$, $\alpha_{k|k}^{\ell,(0)} = \alpha_{k|k-1}$ and $\beta_{k|k}^{\ell,(0)} = \beta_{k|k-1}$.

6) Calculation of $\mathbb{E}_{q_{\mathbf{X}}^{(\ell)}, q_\theta^{(\ell)}}\left[(sT_{\theta_k}X_kT_{\theta_k}^T)^{-1}\right]$: This expectation can be calculated exactly, thanks to the factorized distributions.

Lemma 1: Given

$$M^{-1} = \begin{bmatrix} m_{11} & m_{12} \\ m_{21} & m_{22} \end{bmatrix},$$

and $q_\theta^{(\ell)}(\theta_k) = \mathcal{N}(\theta_k, \hat{\theta}_{k|k}^{(\ell)}, \Theta_{k|k}^{(\ell)})$, the entries of the matrix $\mathbb{E}_{q_\theta^{(\ell)}}\left[(T_{\theta_k}MT_{\theta_k}^T)^{-1}\right]$ can be computed as:

$$\begin{aligned}&\mathbb{E}_{q_\theta^{(\ell)}}\left[(sT_{\theta_k}MT_{\theta_k}^T)^{-1}\right]_{11} \\ &\qquad = \frac{m_{11}}{2}(1 + \cos(2\hat{\theta}_{k|k}^{(\ell)})\exp(-2\Theta_{k|k}^{(\ell)})) \\ &\qquad + \frac{m_{22}}{2}(1 - \cos(2\hat{\theta}_{k|k}^{(\ell)})\exp(-2\Theta_{k|k}^{(\ell)})) \\ &\qquad - \frac{(m_{12} + m_{21})}{2}(\sin(2\hat{\theta}_{k|k}^{(\ell)})\exp(-2\Theta_{k|k}^{(\ell)}))\end{aligned} \tag{32a}$$

$$\begin{aligned}&\mathbb{E}_{q_\theta^{(\ell)}}\left[(sT_{\theta_k}MT_{\theta_k}^T)^{-1}\right]_{12} \\ &\qquad = \frac{m_{12}}{2}(1 + \cos(2\hat{\theta}_{k|k}^{(\ell)})\exp(-2\Theta_{k|k}^{(\ell)})) \\ &\qquad - \frac{m_{21}}{2}(1 - \cos(2\hat{\theta}_{k|k}^{(\ell)})\exp(-2\Theta_{k|k}^{(\ell)})) \\ &\qquad + \frac{(m_{11} - m_{22})}{2}(\sin(2\hat{\theta}_{k|k}^{(\ell)})\exp(-2\Theta_{k|k}^{(\ell)}))\end{aligned} \tag{32b}$$

$$\begin{aligned}&\mathbb{E}_{q_\theta^{(\ell)}}\left[(sT_{\theta_k}MT_{\theta_k}^T)^{-1}\right]_{21} \\ &\qquad = \frac{m_{21}}{2}(1 + \cos(2\hat{\theta}_{k|k}^{(\ell)})\exp(-2\Theta_{k|k}^{(\ell)})) \\ &\qquad - \frac{m_{12}}{2}(1 - \cos(2\hat{\theta}_{k|k}^{(\ell)})\exp(-2\Theta_{k|k}^{(\ell)})) \\ &\qquad + \frac{(m_{11} - m_{22})}{2}(\sin(2\hat{\theta}_{k|k}^{(\ell)})\exp(-2\Theta_{k|k}^{(\ell)}))\end{aligned} \tag{32c}$$

$$\begin{aligned}&\mathbb{E}_{q_\theta^{(\ell)}}\left[(sT_{\theta_k}MT_{\theta_k}^T)^{-1}\right]_{22} \\ &\qquad = \frac{m_{22}}{2}(1 + \cos(2\hat{\theta}_{k|k}^{(\ell)})\exp(-2\Theta_{k|k}^{(\ell)})) \\ &\qquad + \frac{m_{11}}{2}(1 - \cos(2\hat{\theta}_{k|k}^{(\ell)})\exp(-2\Theta_{k|k}^{(\ell)})) \\ &\qquad + \frac{(m_{12} + m_{21})}{2}(\sin(2\hat{\theta}_{k|k}^{(\ell)})\exp(-2\Theta_{k|k}^{(\ell)}))\end{aligned} \tag{32d}$$

The proof is given in Appendix B.

Corollary 1:

$$\begin{aligned}&\mathbb{E}_{q_{\mathbf{X}}^{(\ell)}, q_\theta^{(\ell)}}\left[(sT_{\theta_k}X_kT_{\theta_k}^T)^{-1}\right] \\ &= (1 - \exp(-2\Theta_{k|k}^{(\ell)}))\frac{\mathrm{Tr}(\mathbb{E}_{q_{\mathbf{X}}^{(\ell)}}[(sX_k)^{-1}])}{2}\mathbb{I}_2 \\ &\quad + \exp(-2\Theta_{k|k}^{(\ell)})\left(T_{\hat{\theta}_{k|k}^{(\ell)}}\mathbb{E}_{q_{\mathbf{X}}^{(\ell)}}[(sX_k)^{-1}]T_{\hat{\theta}_{k|k}^{(\ell)}}^T\right),\end{aligned} \tag{33}$$

where $\mathbb{I}_2$ is 2x2 identity matrix. This expression is obtained from *Lemma* 1 by exploiting the fact that the matrix X_k is diagonal by definition.

IV. Time Update

Once the measurement update is performed, the sufficient statistics of the posterior density must be propagated in time in accordance with the target dynamics. An optimal time update step requires the solution to the following Chapman-Kolmogorov equation

$$p(\mathbf{x}_k^a, X_k|\mathbf{Y}_{1:k-1}) = \int p(\mathbf{x}_k^a, X_k|\mathbf{x}_{k-1}^a, X_{k-1}) \\ p(\mathbf{x}_{k-1}^a, X_{k-1}|\mathbf{Y}_{1:k-1}) d\mathbf{x}_{k-1}^a dX_{k-1}, \tag{34}$$

where $\mathbf{x}_k^a \triangleq \left[\mathbf{x}_k^T \, \theta_k\right]^T$. Unfortunately, it is not possible to obtain an exact compact form analytical expression for most extended target tracking models. Therefore various independence conditions are implied to perform time updates in the literature [4], [5], [7], [20]. For a detailed analysis of possible time update approaches, interested readers can refer to [20] and the references therein.

In the random matrix framework, it is possible to assume that the dynamical models of the kinematic state and the extent state are independent [5],

$$p(\mathbf{x}_k, X_k|\mathbf{x}_{k-1}, X_{k-1}) = p(\mathbf{x}_k|\mathbf{x}_{k-1})p(X_k|X_{k-1}). \tag{35}$$

Consequently, the time update of the kinematic state and the extent state can be decoupled for factorised posteriors. The time update of the kinematic state follows the Kalman filter prediction equations if the underlying dynamics are linear. Consider the following state space model which describes the dynamics of the augmented state vector $\mathbf{x}_k^a$,

$$\mathbf{x}_k^a = F\mathbf{x}_{k-1}^a + u_k, \quad u_k \sim \mathcal{N}(0, Q). \tag{36}$$

The prediction density $\mathcal{N}(\mathbf{x}_{k|k-1}^a; \hat{\mathbf{x}}_{k|k-1}^a, P_{k|k-1}^a)$ is obtained by updating the sufficient statistics (mean and covariance) of the Gaussian components in accordance with the system dynamics

$$\hat{\mathbf{x}}_{k|k-1}^a = F\hat{\mathbf{x}}_{k-1|k-1}^a, \tag{37a}$$

$$P_{k|k-1}^a = FP_{k-1|k-1}^a F^T + Q. \tag{37b}$$

where $P_k^a \triangleq \text{blkdiag}[(P_k, \ \ \Theta_k)]$.

In most tracking applications, the exact dynamics of the extent state is unknown. Even in the case where the dynamic equations of the extent states are available, the transition density induced by the known dynamics may not lead to a prediction update that results in the same family of probability distributions using (34). If the dynamics of the extent state is slowly varying but unknown, it is possible to obtain the maximum entropy prediction density of the extent states by utilizing a forgetting factor [27, Theorem 1]. In that case, the sufficient statistics of the inverse Gamma distribution is updated as

$$\alpha_{k|k-1}^i = \gamma_k \alpha_{k-1|k-1}^i, \tag{38a}$$

$$\beta_{k|k-1}^i = \gamma_k \beta_{k-1|k-1}^i, \quad \text{for} \quad i = 1, \ldots, n_y \tag{38b}$$

where γ is the forgetting factor. We prefer to use the maximum entropy prediction density in the time update. However, it is possible to perform alternative time updates within the proposed framework.

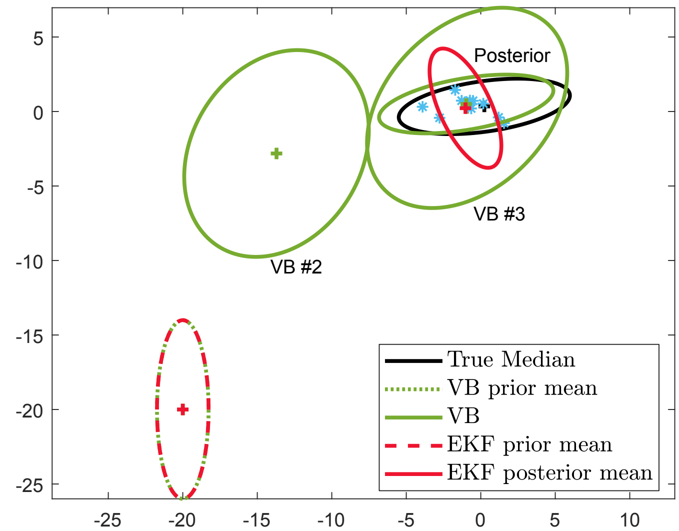

Fig. 1: A single measurement update for VB and the EKF approach. The prior and posterior mean shape estimates are represented by green dotted and solid lines for VB, respectively. The red dashed line indicates the prior mean shape estimate while the red solid line depicts the posterior mean estimate for EKF approach. The VB $\#i$ denotes the i^{th} variational iteration shape estimate mean of the VB algorithm.

V. A Closer Look to a Single Measurement Update

In this section, we investigate the proposed measurement update, here and after denoted as VB, in more detail and illustrate its capabilities in comparison with a state-of-the-art extended Kalman filter (EKF) algorithm [17]. For this purpose, we initiate the prior mean and covariance of both approaches the same; and we compare the posterior distribution of the extent states. Consider the example given in Figure 1, where the prior mean of the target's location is $[-20 \ -20]^T$. The measurements are shown with blue stars, and the posterior means of the VB and EKF updates are shown with the solid green and red lines, respectively. The median of the true posterior, which is computed by using 1 million Monte Carlo samples, is shown with the solid black line. The mean of the extent and kinematic state distributions at the end of each VB iteration is denoted by VB $\#i$ where $\#i$ stands for the i^{th} variational iteration. A total of 10 iterations are performed within the variational update. As shown in Figure 1, the posterior found by the VB algorithm is closer to the true posterior than the posterior computed by the EKF, thanks to the iterative nature of the VB updates. Unlike EKF, the VB algorithm performs multiple iterations in a single update and performs multiple linearizations during the iterations by taking all available measurements into account. The ability to compute the posterior iteratively is the key concept to explain the superior performance of VB in the experiments given in Section VI.

Lastly, we compare the average computation time of the algorithms. The simulations for the illustrative example are run in Matlab(R) R2019b on a standard laptop with an Intel(R) Core(TM) i7-6700HQ 2.60 GHz platform with 16 GB of RAM. We compare naive implementations of the algorithms without exploiting any code optimization methods. A single measurement update and a single variational iteration for VB takes 2.3×10^{-3} sec and 2.1×10^{-4} sec, respectively. On the other hand, it takes 8.6×10^{-4} sec to perform a measurement update for EKF. The relevant parameters of the illustrative example are given in the Appendix-C.

VI. Experimental Results

In this section, we evaluate the performance of the proposed method and compare it with relevant elliptical object tracking algorithms in the literature. The comparison is performed through both simulations and real data experiments. The alternative models are selected as the state-of-the-art EKF approach that is capable of tracking the orientation of elliptical objects [17] and the widely used RM based ETT model [5]. In the sequel, we denote these algorithms as Algorithm-1 and Algorithm-2, respectively. The simulation results are presented in Section VI-A, and the results of the real-data experiment are given in Section VI-B.

A. Simulations

In the simulations, we use the Gaussian Wasserstein (GW) distance [28], [29] and root-mean-square-error (RMSE) for performance evaluation and comparison,

$$\begin{aligned}&\mathrm{GW}(\mathbf{m}_a, X_a, \mathbf{m}_b, X_b)^2 \\ &\triangleq \underbrace{\|\mathbf{m}_a - \mathbf{m}_b\|_2^2}_{1^{\text{st}}\ Term} + \underbrace{\mathrm{Tr}[X_a + X_b - 2(X_a^{\frac{1}{2}} X_b X_a^{\frac{1}{2}})^{\frac{1}{2}}]}_{2^{\text{nd}}\ Term}.\end{aligned} \tag{39}$$

Here, $\mathbf{m}_a$, $\mathbf{m}_b$ and X_a, X_b stand for two different center locations and elliptic extent matrices, respectively. The first term in (39) corresponds to the error in the estimation of the object's center, and the second term corresponds to the error in extent estimation. We report both terms in (39) in addition to the overall GW distance to provide insight about the estimation performance of the algorithms in detail. Furthermore, we compare the RMSE of the orientation estimations which is defined by

$$\mathrm{RMSE}(\theta_{true}, \theta) = \sqrt{\frac{1}{N}\sum_{k=1}^{N}(\theta_{k,true} - \theta_k)^2}, \tag{40}$$

where N denotes the number of time steps in a single run.

TABLE II: Common parameter values for the simulations.

F	$\begin{bmatrix} 1 & 0 & T & 0 & 0 \\ 0 & 1 & 0 & T & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}$
H	$\begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \end{bmatrix}$
P_0	$I_{5\times 5}$
$\hat{\mathbf{x}}_0$	$[0 \quad 0 \quad 50 \quad 0 \quad 0]^T$
X_{true}	$\begin{bmatrix} 50 & 0 \\ 0 & 600 \end{bmatrix}$
R	$5 \times I_{2\times 2}$
Q	$\begin{bmatrix} \frac{T^3}{3}\sigma_x & 0 & \frac{T^2}{2}\sigma_x & 0 & 0 \\ 0 & \frac{T^3}{3}\sigma_y & 0 & \frac{T^2}{2}\sigma_y & 0 \\ \frac{T^2}{2}\sigma_x & 0 & T\sigma_x & 0 & 0 \\ 0 & \frac{T^2}{2}\sigma_y & 0 & T\sigma_y & 0 \\ 0 & 0 & 0 & 0 & \sigma_\theta \end{bmatrix}$
T	0.1
$[\sigma_x\ \sigma_y\ \sigma_\theta]$	$[1\ 1\ 10^{-2}]$

TABLE III: The GW distance values for Gaussian measurements.

	GW Distance 1st Term [m²]	GW Distance 2nd Term [m²]	GW Distance [m]
Algorithm-1	4.78	8.54	3.18
Algorithm-2	5.22	56.63	6.86
VB	**4.49**	**5.27**	**2.85**

TABLE IV: The GW distance values for uniformly distributed measurements.

	GW Distance 1st Term [m²]	GW Distance 2nd Term [m²]	GW Distance [m]
Algorithm-1	1.93	6.02	2.49
Algorithm-2	2.18	58.54	6.69
VB	**1.92**	**4.54**	**2.28**

1) Constant Velocity Model: In the first experiment, a dynamic object is simulated, which moves according to the nearly constant velocity model defined by the parameters given in Table II. In this simulation, the parameters of the motion model are fully provided to the tracking algorithms so that the error due to model-mismatch does not affect the estimation performance. Throughout the trajectory, the object generates an average of 10 measurements per scan. We investigate two different cases separately; in the first case, the measurements follow a Gaussian distribution, and in the second case, they follow a uniform distribution. All simulation experiments were performed 100 times with different realizations of the process noise, measurement noise, and measurement origin at each simulation. The presented numbers are the average of these 100 Monte Carlo (MC) runs. The algorithm specific initial shape variables for VB are set to $\alpha_0^{1,2} = [2\ 2]^T$ and $\beta_0^{1,2} = [100\ 100]^T$. The number of variational iterations is 10.

The shape variables are initialized for Algorithm-2 as $v_0 = 4$ and $V_0 = \mathrm{diag}([100,\ 100])$. The forgetting factor is set to $\gamma = 0.99$ for both VB and Algorithm-2. To be consistent with [17], we use the same notations for the parameters of Algorithm-1. The prior mean and covariance matrix of the shape variables of Algorithm-1 is selected to be $\hat{\mathbf{p}}_0 = [0\ \ 10\ \ 10]^T$ and $C_0^p = \mathrm{diag}([1,\ 20,\ 20])$. The vector $\hat{\mathbf{p}}_0$ consists of $[\theta,\ l_1,\ l_2]$ where, θ, l_1 and l_2 are the orientation and the semi-axis lengths, respectively. The process noise covariance matrix for the shape variables for Algorithm-1 is $C_p^w = \mathrm{diag}([10^{-2},\ 0.1,\ 0.1])$. The kinematic state transition matrix for Algorithm-1 is set to $A_r = F(1:4,\ 1:4)$, where F matrix is defined in Table II. The initial mean of the kinematic state vector is the same as VB, $\hat{\mathbf{r}}_0 = \hat{\mathbf{x}}_0(1:4)$. The state transition matrix for the shape variables is $A_p = I_{3\times 3}$. The initial values of the shape and kinematic variables are selected to make the prior means of the algorithms the same. The algorithmic specific parameters are hand-tuned to obtain the best performance of each algorithm. We report the average GW distance and the orientation RMSE for Gaussian and uniformly distributed measurements in Table III and Table IV, respectively. The proposed algorithm performs better in terms of estimating the extent of the target and outperforms the other algorithms in terms of GW distance. Additionally, VB shows a better performance in estimating the orientation of the target.

2) Experimental Trajectory: This experiment involves the scenario studied in [5], [7], [17], [18]. In this simulation, the trajectory composed of one $45°$ and two $90°$ turns pieced together with straight paths. The object of interest has an unknown but fixed semi-axes lengths, and its orientation varies

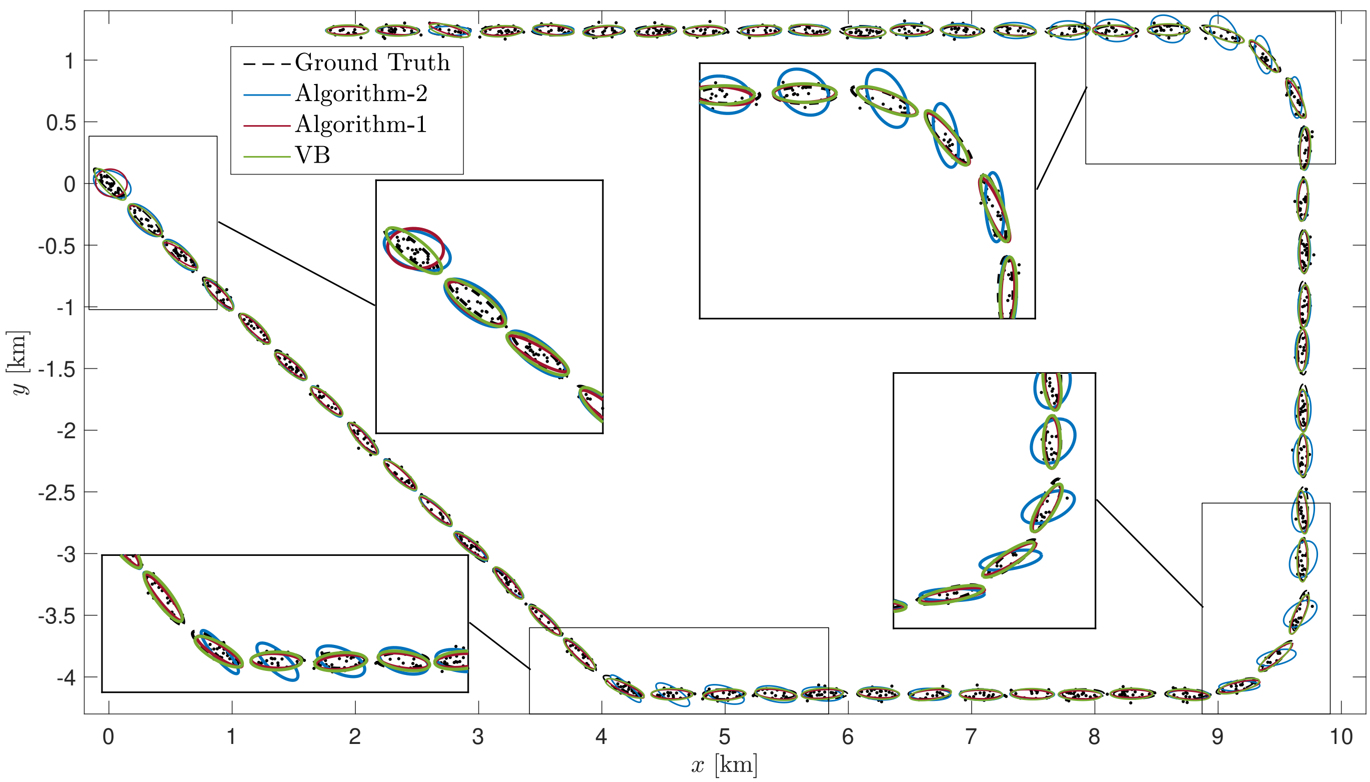

Fig. 2: An example MC run of the scenario in Section VI-A2

TABLE V: The heading angle RMSE values for Gaussian and uniform measurements.

	Gaussian Measurements Heading Angle RMSE [°]	Uniform Measurements Heading Angle RMSE [°]
Algorithm-1	4.46	4.93
Algorithm-2	59.98	60.15
VB	**3.93**	**4.00**

TABLE VI: The GW distance and heading angle RMSE values of the scenario in Section VI-A2 when the measurements are uniformly distributed.

	GW Dist. 1st Term [m^2]	GW Dist. 2nd Term [m^2]	GW Dist. [m]	Heading Angle RMSE [°]
Algorithm-2	284.05	1436.05	32.94	82.61
Algorithm-1	281.38	280.54	20.84	3.89
VB	**270.747**	**203.01**	**19.83**	**3.37**

TABLE VII: The GW distance and heading angle RMSE values of the scenario in Section VI-A2 when the measurements are generated from a Gaussian distribution.

	GW Dist. 1st Term [m^2]	GW Dist. 2nd Term [m^2]	GW Dist. [m]	Heading Angle RMSE [°]
Algorithm-2	826.32	1167.80	37.69	82.66
Algorithm-1	884.75	244.15	29.17	3.57
VB	**822.15**	**145.31**	**27.36**	**2.94**

in time. The object starts its motion from the origin with a speed of 50 $\mathrm{km/h}$, which is fixed throughout the trajectory. The measurements are generated from a uniform distribution, and the number of the measurements is drawn from a Poisson distribution with an average of 20 measurements per scan. In addition to simulations performed in [17], we will examine the performance of the algorithms with Gaussian distributed measurements. As in [17], the prior mean and covariance matrix of the shape variables are selected to be $\hat{\mathbf{p}}_0 = [\pi,\ 200,\ 90]^T$ and $C_0^p = \mathrm{diag}([1,\ 70^2,\ 70^2])$. The process noise covariance matrix for the shape variables and kinematics are $C_p^w = \mathrm{diag}([0.1,\ 1,\ 1])$ and $C_r^w = \mathrm{diag}([100,\ 100,\ 1,\ 1])$, respectively. The measurement noise covariance matrix is, $R = \mathrm{diag}([400,\ 400])$. In order to have a fair comparison, the prior mean values of the kinematic and shape variables for VB and Algorithm-2 are chosen to be the same as those of Algorithm-1. The shape variables for VB are selected to be $\alpha_0^{1,2} = [5\ 5]$ and $\beta_0^{1,2} = [400^2\ 180^2]$. The degrees of freedom is $v_0 = 7$ for Algorithm-2. The scale matrix is initialized as $V_0 = \mathrm{diag}([400^2,\ 180^2])$. The initial mean of the kinematic state is set to $\hat{\mathbf{x}}_0 = [100\ \ 100\ \ 5\ \ -8\ \ \pi]^T$ for VB. The number of the variational iterations is 10. The initial mean of the kinematic state for the Algorithm-1 and Algorithm-2 is selected to be $\hat{\mathbf{r}}_0 = \hat{\mathbf{x}}_0(1:4)$. We conducted 100 MC runs for each measurement distribution type. The GW distance and the orientation RMSE are presented in Table VI and Table VII. An example MC run is depicted in Fig 2. Algorithm-2 could not perform well during the turns because the method does not treat the orientation as a separate random variable and compensates the changes in the orientation by updating the extent estimate. However, VB and Algorithm-1 are able to overcome this problem. The results show that the proposed approach, VB, provides better orientation, center, and extent estimates.

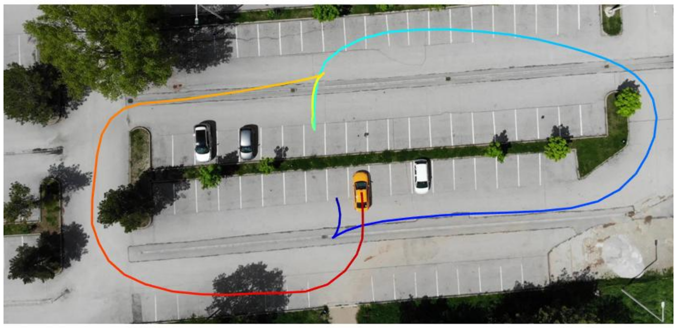

Fig. 3: The outline of the movement of the vehicle during the time-lapse. The vehicle starts from the dark blue colored parking spot; and follows the colored path until the red colored parking spot. In the figure, the last frame is shown.

B. Real Data Experiment

In this section, the algorithms' capabilities are illustrated with real data. The data is collected in an urban area of Ankara. The test scenario involves a commercial vehicle moving in a parking lot while a steady aerial camera captures images of the surveillance region every second, i.e., $T = 1s$. In the scenario, a long sampling time is intentionally chosen to minimize the computational power consumption, thereby prolonging the air-time of the aerial camera in possible real-time applications. The outline of the vehicle's trajectory is shown in Figure 3. The colored-line indicates the trajectory followed by the mid-point of the vehicle. The scenario starts while the vehicle is parked in the parking area, indicated by the dark blue color. The vehicle leaves the parking area and follows the path shown in blue until it is parked in the parking area, which is indicated by the green color. Then the vehicle performs a similar motion from the green-colored parking spot following the path to the red-colored parking spot.

Throughout the scenario, the captured images are processed for measurement extraction. Various feature extraction algorithms can be used to obtain measurements from the vehicle such as Harris corner detection [30], Scale Invariant Feature Transform (SIFT) [31], Speeded Up Robust Features (SURF) [32] or similar. In order to demonstrate that the algorithm can work with a wide range of feature extraction algorithms, we present a more general case, where the measurements are uniformly sampled from the vehicle's visible surface. The results obtained by using the features extracted by the Harris corner detector are also consistent with the results presented here, but they are not included in the manuscript because of the page limitations.

As part of the image processing step, a segmentation is performed in every frame in the HSV color band to separate the yellow vehicle from the background. Following that, a median filter is used to reduce the number of clutters. Finally, the pixels that belong to the vehicle are sampled uniformly to obtain the measurements. The initial position of the vehicle is extracted from the first frame. The initial velocity, on the other hand, is assumed to be unknown and assumed to be zero. Hence, the initial mean of the kinematic state vector is selected to be $\hat{\mathbf{x}}_0 = [450 \;\; 245 \;\; 0 \;\; 0 \;\; \frac{\pi}{2}]^T$. The initial parameters of the algorithms are selected to match the prior means of the corresponding distributions. For this purpose, the initial shape parameters for VB is selected to be $\alpha_0^{1,2} = [2 \; 2]^T$ and $\beta_0^{1,2} = [250 \; 1000]^T$. The prior mean and covariance matrix of the shape variables for Algorithm-1 is set to $\hat{\mathbf{p}}_0 = [0 \; 250^{0.5} \; 1000^{0.5}]^T$ and $C_0^p = \text{diag}([1, \; 100, \; 100])$, respectively. The degrees of freedom value and the initial scale matrix is set to $v_0 = 4$ and $V_0 = \text{diag}([250, \; 1000])$ for Algorithm-2, respectively. The process noise covariance matrix Q is given in Table II, where σ_x and σ_y are 4 and σ_θ is 0.1 for VB. The number of variational iterations is 10. The process noise covariance matrix for the shape variables for Algorithm-1 is set to $C_p^w = \text{diag}([0.1, \; 10^{-3}, \; 10^{-3}])$. Finally, the measurement noise covariance matrix is taken as $R = \text{diag}([1, \; 1])$. The parameters were optimized manually to obtain the best performances of the algorithms.

The extent estimates corresponding to the frames $\{1, 14, 24, 39, 45, 83\}$ are given in Figure 4. These snapshots were chosen for the sake of a clearer illustration of the differences between the algorithms' performances, starting from the initial frame. At the beginning of the scenario, the vehicle stays immobile, and the algorithms are able to estimate the vehicle's extent accurately (see: Frame 1). When the vehicle is moving in a straight path, such as in Frame 18 and Frame 39, the performance of all the algorithms are satisfactory. However, when the vehicle performs a maneuver, as in Frame 24, Frame 45 and Frame 83, VB shows superior performance in estimating the orientation of the vehicle. During the maneuvers, Algorithm-2 cannot estimate the extent accurately because it does not treat the heading angle as a separate random variable, and it tries to adapt to the changes in the orientation by updating the extent states. Algorithm-1 also struggles to find the correct orientation of the vehicle. However, VB can provide accurate estimates of the extent thanks to its iterative updates. Note that, VB and Algorithm-1 use the same process noise variance for the orientation. Since the vehicle is stable in the first couple of frames and the algorithms are able to estimate the extent accurately, increasing the variance values of the shape variables for Algorithm-1 does not improve the performance of estimating the extent further. Additionally, if the variance values are increased too much, the extent estimates of Algorithm-1 tend to collapse to zero. We encountered a similar problem while tuning Algorithm-1 in the simulation scenarios. We report one example of such behavior in a single measurement update in Appendix D for interested readers.

VII. Conclusion

ETT involves tracking objects that generate multiple measurements per scan. In most ETT applications, the orientation of the extended targets changes in time. In standard RM based ETT methods, this phenomenon is addressed by a forgetting factor which aims at forgetting the accumulated information. In this work, we proposed a novel approach for extended target tracking that is capable of simultaneously estimating the kinematic, extent, and orientation states of an extended target. We use the variational Bayes technique for inference and define appropriate priors for the unknown state variables that can accurately model the changes in the extended targets' orientation. The performance and capabilities of the algorithm are demonstrated through simulations and real data experiments. Experimental results on simulations and real data demonstrate that the proposed method significantly improves the tracking performance, as well as the accuracy in estimating the orientation and the shape of the object compared to the state-of-the-art methods.

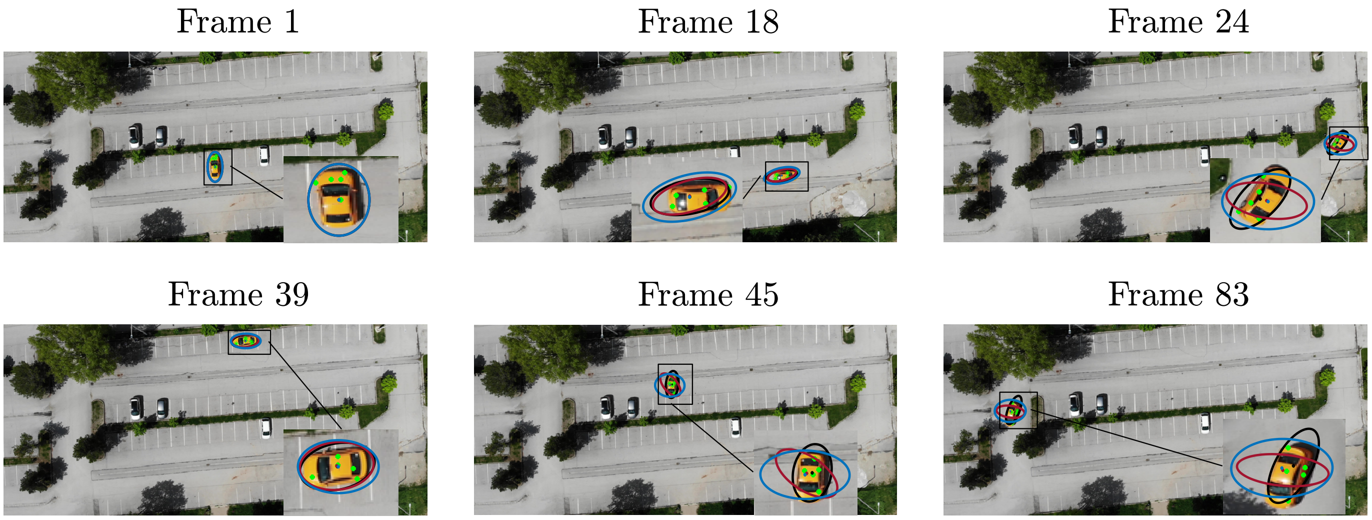

Fig. 4: A representative MC run of the real data experiment. The extent estimates of VB, Algorithm-1, and Algorithm-2 are shown in black, red and blue lines, respectively. The measurements are represented with green dots.

References

[1] M. Baum, V. Klumpp, and U. D. Hanebeck, "A novel Bayesian method for fitting a circle to noisy points," in *13th International Conference on Information Fusion (FUSION)*. IEEE, 2010, pp. 1–6.

[2] K. Granström, C. Lundquist, and U. Orguner, "Tracking rectangular and elliptical extended targets using laser measurements," in *14th International Conference on Information Fusion (FUSION)*, July 2011, pp. 1–8.

[3] K. Granström and C. Lundquist, "On the use of multiple measurement models for extended target tracking," in *16th International Conference on Information Fusion (FUSION)*, July 2013, pp. 1534–1541.

[4] J. W. Koch, "Bayesian approach to extended object and cluster tracking using random matrices," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 44, no. 3, pp. 1042–1059, July 2008.

[5] M. Feldmann, D. Franken, and W. Koch, "Tracking of extended objects and group targets using random matrices," *IEEE Transactions on Signal Processing*, vol. 59, no. 4, pp. 1409–1420, April 2011.

[6] U. Orguner, "A variational measurement update for extended target tracking with random matrices," *IEEE Transactions on Signal Processing*, vol. 60, no. 7, pp. 3827–3834, July 2012.

[7] J. Lan and X. R. Li, "Tracking of extended object or target group using random matrix – Part I: New model and approach," in *15th International Conference on Information Fusion (FUSION)*, Singapore, Jul. 2012, pp. 2177–2184.

[8] M. Baum and U. D. Hanebeck, "Random hypersurface models for extended object tracking," in *IEEE International Symposium on Signal Processing and Information Technology (ISSPIT)*, Dec 2009, pp. 178–183.

[9] M.Baum and U. D. Hanebeck, "Shape tracking of extended objects and group targets with star-convex rhms," in *14th International Conference on Information Fusion (FUSION)*, July 2011, pp. 1–8.

[10] N. Wahlström and E. Özkan, "Extended target tracking using Gaussian processes," *IEEE Transactions on Signal Processing*, vol. 63, no. 16, pp. 4165–4178, Aug 2015.

[11] E. Özkan, N. Wahlström, and S. J. Godsill, "Rao-Blackwellised particle filter for star-convex extended target tracking models," in *19th International Conference on Information Fusion (FUSION)*, July 2016, pp. 1193–1199.

[12] M. Kumru and E. Özkan, "3D extended object tracking using recursive Gaussian processes," in *2018 21st International Conference on Information Fusion (FUSION)*. IEEE, 2018, pp. 1–8.

[13] J. Lan and X. R. Li, "Tracking of maneuvering non-ellipsoidal extended object or target group using random matrix," *IEEE Transactions on Signal Processing*, vol. 62, no. 9, pp. 2450–2463, May 2014.

[14] K. Granström, "An extended target tracking model with multiple random matrices and unified kinematics," *CoRR*, vol. abs/1406.2135, 2014. [Online]. Available: http://arxiv.org/abs/1406.2135

[15] Q. Hu, H. Ji, and Y. Zhang, "Tracking of maneuvering non-ellipsoidal extended target with varying number of sub-objects," *Mechanical Systems and Signal Processing*, vol. 99, pp. 262 – 284, 2018. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0888327017303254

[16] S. F. Kara and E. Özkan, "Multi-ellipsoidal extended target tracking using sequential Monte Carlo," in *2018 21st International Conference on Information Fusion (FUSION)*. IEEE, 2018, pp. 1–8.

[17] S. Yang and M. Baum, "Tracking the orientation and axes lengths of an elliptical extended object," *IEEE Transactions on Signal Processing*, vol. 67, no. 18, pp. 4720–4729, Jul 2019.

[18] ——, "Extended Kalman filter for extended object tracking," in *2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2017, pp. 4386–4390.

[19] ——, "Second-order extended Kalman filter for extended object and group tracking," in *2016 19th International Conference on Information Fusion (FUSION)*. IEEE, 2016, pp. 1178–1184.

[20] K. Granström and U. Orguner, "New prediction for extended targets with random matrices," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 50, no. 2, pp. 1577–1589, Apr. 2014.

[21] V. Smidl and A. Quinn, "Variational Bayesian filtering," *IEEE Transactions on Signal Processing*, vol. 56, no. 10, pp. 5020–5030, 2008.

[22] S. Sarkka and A. Nummenmaa, "Recursive noise adaptive Kalman filtering by variational Bayesian approximations," *IEEE Transactions on Automatic control*, vol. 54, no. 3, pp. 596–600, 2009.

[23] H. Nurminen, T. Ardeshiri, R. Piché, and F. Gustafsson, "Skew-t filter and smoother with improved covariance matrix approximation," *IEEE Transactions on Signal Processing*, vol. 66, no. 21, pp. 5618–5633, 2018.

[24] H. Nurminen, T. Ardeshiri, R. Piche, and F. Gustafsson, "Robust inference for state-space models with skewed measurement noise," *IEEE Signal Processing Letters*, vol. 22, no. 11, pp. 1898–1902, 2015.

[25] Y. Huang, Y. Zhang, Z. Wu, N. Li, and J. Chambers, "A novel adaptive Kalman filter with inaccurate process and measurement noise covariance matrices," *IEEE Transactions on Automatic Control*, vol. 63, no. 2, pp. 594–601, 2017.

[26] C. M. Bishop, *Pattern Recognition and Machine Learning (Information Science and Statistics)*. Secaucus, NJ, USA: Springer-Verlag New York, Inc., 2006.

[27] E. Özkan, V. Šmídl, S. Saha, C. Lundquist, and F. Gustafsson, "Marginalized adaptive particle filtering for nonlinear models with unknown time-varying noise parameters," *Automatica*, vol. 49, no. 6, pp. 1566 – 1575, 2013. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0005109813001350

[28] C. R. Givens, R. M. Shortt *et al.*, "A class of Wasserstein metrics for probability distributions." *The Michigan Mathematical Journal*, vol. 31, no. 2, pp. 231–240, 1984.

[29] S. Yang, M. Baum, and K. Granström, "Metrics for performance evaluation of elliptic extended object tracking methods," in *2016 IEEE International Conference on Multisensor Fusion and Integration for Intelligent Systems (MFI)*. IEEE, 2016, pp. 523–528.

[30] C. G. Harris, M. Stephens *et al.*, "A combined corner and edge detector." in *Alvey vision conference*, vol. 15, no. 50. Citeseer, 1988, pp. 10–5244.

[31] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *International journal of computer vision*, vol. 60, no. 2, pp. 91–110, 2004.

[32] H. Bay, T. Tuytelaars, and L. Van Gool, "Surf: Speeded up robust features," in *European conference on computer vision*. Springer, 2006, pp. 404–417.

Appendix A
Calculation of δ and Δ

In Section III-A4, we introduced the update formulas for the orientation distribution as below.

$$\hat{\theta}_{k|k}^{(\ell+1)} = \Theta_{k|k}^{(\ell+1)}\big(\Theta_{k|k-1}^{-1}\hat{\theta}_{k|k-1} + \delta\big), \tag{41}$$

$$\Theta_{k|k}^{(\ell+1)} = \big(\Theta_{k|k-1}^{-1} + \Delta\big)^{-1}, \tag{42}$$

Here, the variables δ and Δ are derived from the expectation in (43).

$$\frac{-1}{2}\sum_{j=1}^{m_k}\mathbb{E}_{q_{\mathbf{x}}^{(\ell)},q_X^{(\ell)},q_{\mathbf{Z}}^{(\ell)}}\left[(a-b\theta_k)^T(sX)^{-1}(a-b\theta_k)\right], \tag{43}$$

where δ and Δ are denoted as

$$\delta \triangleq \sum_{j=1}^{m_k}\mathrm{Tr}\left[\mathbb{E}_{q_X^{(\ell)}}\left[(sX_k)^{-1}\right]\mathbb{E}_{q_{\mathbf{x}}^{(\ell)},q_{\mathbf{Z}}^{(\ell)}}\left[ab^T\right]\right], \tag{44}$$

$$\Delta \triangleq \sum_{j=1}^{m_k}\mathrm{Tr}\left[\mathbb{E}_{q_X^{(\ell)}}\left[(sX_k)^{-1}\right]\mathbb{E}_{q_{\mathbf{x}}^{(\ell)},q_{\mathbf{Z}}^{(\ell)}}\left[bb^T\right]\right], \tag{45}$$

and

$$a = (T_{\hat{\theta}_{k|k}^{(\ell)}})^T(\mathbf{z}_k^j - H\mathbf{x}_k) - (T'_{\hat{\theta}_{k|k}^{(\ell)}})^T(\mathbf{z}_k - H\mathbf{x}_k)\hat{\theta}_{k|k}^{(\ell)} \tag{46}$$

$$b = -(T'_{\hat{\theta}_{k|k}^{(\ell)}})(\mathbf{z}_k - H\mathbf{x}_k). \tag{47}$$

When the a and b variables are substituted into (44) and (45), we obtain the δ and Δ variables as

$$\begin{aligned}\delta = &\sum_{j=1}^{m_k}\mathrm{Tr}\left[\overline{sX_k}^{-1}(T'_{\hat{\theta}_{k|k}^{(\ell)}})^T\overline{\big(\mathbf{z}_k^j - H\mathbf{x}_k\big)\big(\,\cdot\,\big)^T}(T'_{\hat{\theta}_{k|k}^{(\ell)}})\hat{\theta}_{k|k}^{(\ell)}\right]\\ &-\mathrm{Tr}\left[\overline{sX_k}^{-1}T^T_{\hat{\theta}_{k|k}^{(\ell)}}\overline{\big(\mathbf{z}_k^j - H\mathbf{x}_k\big)\big(\,\cdot\,\big)^T}(T'_{\hat{\theta}_{k|k}^{(\ell)}})\right],\end{aligned} \tag{48}$$

$$\Delta = \sum_{j=1}^{m_k}\mathrm{Tr}\left[\overline{sX_k}^{-1}(T'_{\hat{\theta}_{k|k}^{(\ell)}})^T\overline{\big(\mathbf{z}_k^j - H\mathbf{x}_k\big)\big(\,\cdot\,\big)^T}(T'_{\hat{\theta}_{k|k}^{(\ell)}})\right]. \tag{49}$$

Appendix B
Proof of Lemma 1

In this section we will give the proof of *Lemma* 1 to calculate $\mathbb{E}_{q_\theta^{(\ell)}}\left[(sT_{\theta_k}MT_{\theta_k}^T)^{-1}\right]$. In the formulation, we first multiply the matrices inside the expectation. Then, the expectation of each entry of the resultant matrix is taken. First, notice that,

$$\big(T_{\theta_k}MT_{\theta_k}^T\big)^{-1} = T_{\theta_k}M^{-1}T_{\theta_k}^T. \tag{50}$$

Given

$$M^{-1} = \begin{bmatrix} m_{11} & m_{12}\\ m_{21} & m_{22}\end{bmatrix},$$

the expression whose expectation has to be taken becomes

$$\begin{aligned}\Lambda &= T_{\theta_k}M^{-1}T_{\theta_k}^T\\ &= \begin{bmatrix}\cos(\theta_k) & -\sin(\theta_k)\\ \sin(\theta_k) & \cos(\theta_k)\end{bmatrix}\begin{bmatrix} m_{11} & m_{12}\\ m_{21} & m_{22}\end{bmatrix}\\ &\quad\times\begin{bmatrix}\cos(\theta_k) & \sin(\theta_k)\\ -\sin(\theta_k) & \cos(\theta_k)\end{bmatrix}\end{aligned} \tag{51a}$$

$$\begin{aligned}\Lambda(1,1) &= m_{11}\cos^2(\theta_k) + m_{22}\sin^2(\theta_k)\\ &\quad - (m_{12}+m_{21})\cos(\theta_k)\sin(\theta_k)\end{aligned} \tag{51b}$$

$$\begin{aligned}\Lambda(1,2) &= m_{12}\cos^2(\theta_k) - m_{21}\sin^2(\theta_k)\\ &\quad + (m_{11}-m_{22})\cos(\theta_k)\sin(\theta_k)\end{aligned} \tag{51c}$$

$$\begin{aligned}\Lambda(2,1) &= m_{21}\cos^2(\theta_k) - m_{12}\sin^2(\theta_k)\\ &\quad + (m_{11}-m_{22})\cos(\theta_k)\sin(\theta_k)\end{aligned} \tag{51d}$$

$$\begin{aligned}\Lambda(2,2) &= m_{22}\cos^2(\theta_k) + m_{11}\sin^2(\theta_k)\\ &\quad + (m_{12}+m_{21})\cos(\theta_k)\sin(\theta_k)\end{aligned} \tag{51e}$$

Now the following trigonometric transformations are utilized and the expectations are taken with respect to the resulting expressions.

$$\cos^2(\theta_k) = \frac{1+\cos(2\theta_k)}{2} \tag{52a}$$

$$\sin^2(\theta_k) = \frac{1-\cos(2\theta_k)}{2} \tag{52b}$$

$$\cos(\theta_k)\sin(\theta_k) = \frac{\sin(2\theta_k)}{2} \tag{52c}$$

$$\mathbb{E}_{q_\theta^{(\ell)}}\left[\cos(2\theta_k)\right] = \cos(2\hat{\theta}_{k|k}^{(\ell)})\exp(-2\Theta_{k|k}^{(\ell)})) \tag{52d}$$

$$\mathbb{E}_{q_\theta^{(\ell)}}\left[\sin(2\theta_k)\right] = \sin(2\hat{\theta}_{k|k}^{(\ell)})\exp(-2\Theta_{k|k}^{(\ell)})) \tag{52e}$$

By substituting the expressions in (51) with the corresponding equalities given in (52) *Lemma* 1 is obtained.

Appendix C
The parameters of the experiment in section V

In this section, the parameters of the experiment given in Section V are summarized. The prior shape estimate is apart from the true center location of the target by $(20,\ 20)$ units in 2D coordinate frame. The extent of the ground truth object is parametrized as diag($[6,\ 0.5]$) with $0°$ orientation. The mean of the prior extent ellipse parameters for both methods are equal to the parameters of the ground truth extent. The prior shape parameters of the proposed algorithm are selected to be $\alpha_0^{1,2} = [101\ 101]^T$ and $\beta_0^{1,2} = [600\ 50]^T$. The prior mean of the orientation variable is taken as $\hat{\theta}_0 = \frac{\pi}{2}$. To have a reasonable comparison, the mean vector and the variance of the shape parameters for the EKF approach are selected to match with those of the VB algorithm. The prior kinematic state covariance matrix is $P_0 = \mathrm{diag}([300,\ 300,\ 1,\ 1,\ \frac{\pi}{2}])$. The measurement noise covariance matrix is $R = \mathrm{diag}([1,\ 1])$. In this experiment, the number of measurements is 10 and the measurements are generated according to a Gaussian distribution. However, the trials with the uniformly distributed measurements yields similar results.

Appendix D
An example for the collapsing extent estimates

Here we repeated the simulation in Section V with the following set of parameters:
For VB, the prior mean of the target's kinematic state is $\hat{\mathbf{x}}_0 = [-30\ \ -30\ \ 1\ \ 1\ \ \frac{\pi}{2}]^T$. The prior kinematic state covariance matrix is $P_0 = \mathrm{diag}([300,\ 300,\ 1,\ 1,\ \frac{\pi}{2}])$. The shape parameters for VB is $\alpha^{1,2} = [11\ \ 11]^T$ and $\beta^{1,2} = [600\ \ 50]^T$. For Algorithm-1, the prior mean vector and covariance matrix of the kinematic state is $\hat{\mathbf{r}}_0 = \hat{\mathbf{x}}_0(1:4)$ and $C_0^r = \mathrm{diag}([300,\ 300,\ 1,\ 1])$, respectively. The prior mean and variance of the extent parameters are the same for both algorithms. The number of measurements generated from

the target is 10. The measurement noise covariance matrix is $R = \text{diag}([1,\ 1])$.

A single measurement update of VB and Algorithm-1 is visualized in Figure 5. Blue stars represent the measurements, the solid green and red lines stand for the posterior means of the VB and EKF updates, respectively. The solid black line indicates the median of the true posterior, which is computed by using 1 million Monte Carlo samples.

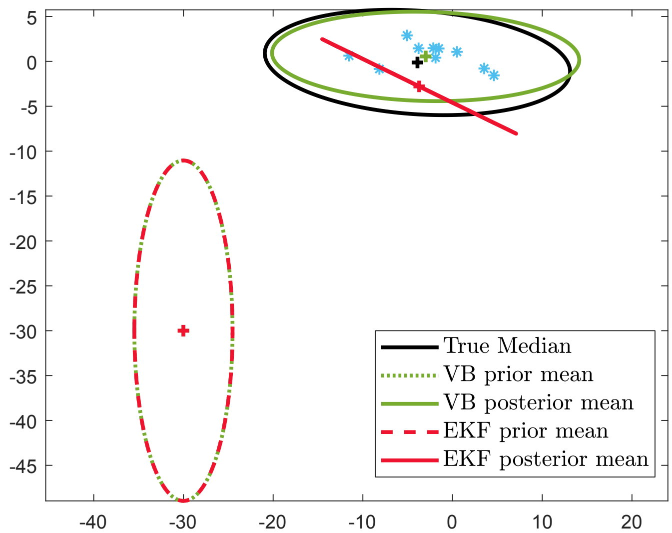

Fig. 5: The visualization of the collapsing behavior of Algorithm-1.