

# An Adaptive PHD Filter for Tracking with Unknown Sensor Characteristics

Tohid Ardeshiri, Emre Özkan

Division of Automatic Control, Department of Electrical Engineering, Linköping University  
581 83 Linköping, Sweden, email: {tohid,emre}@isy.liu.se

**Abstract**—In multi-target tracking, the discrepancy between the nominal and the true values of the model parameters might result in poor performance. In this paper, an adaptive Probability Hypothesis Density (PHD) filter is proposed which accounts for sensor parameter uncertainty. Variational Bayes technique is used for approximate inference which provides analytic expressions for the PHD recursions analogous to the Gaussian mixture implementation of the PHD filter. The proposed method is evaluated in a multi-target tracking scenario. The improvement in the performance is shown in simulations.

**Keywords**—variational Bayes, adaptive filtering, sensor calibration, probability hypothesis density filter, robust filtering, multiple target tracking

## I. INTRODUCTION

Multi-target tracking (MTT) problem involves resolving the uncertainty in a surveillance region by estimating the number of targets and their trajectories based on the measurements collected by a number of sensors which suffer from possible missed detections and false alarms. Among the vast number of studies proposed for MTT, Mahler's PHD filter [1], [2] has drawn significant interest as it reformulates the problem using Random Finite Sets (RFS). PHD filter approximates the multi-target posterior distribution by propagating its first order moment. It has been implemented by different inference techniques like sequential Monte Carlo (SMC) [3], point-mass [4], or the Gaussian mixture approximation [5]. The Gaussian mixture implementation of the PHD filter (GMPHD) has been the most common choice among the others, where Kalman filters (KF) are used for updating the individual components of the mixture under linear Gaussian model assumptions. Having KFs as its building blocks, the performance of the GMPHD heavily relies on the performance of the KFs. One deficiency of the KF is its lack of robustness to parameter uncertainties. In this study, we aim to address this problem by replacing the KF with a more sophisticated approximate inference technique which accounts for the uncertainty in the sensor noise covariance. The sensitivity of the KFs to model uncertainties has been known for long, and early studies addressing the issue dates back as early as 70's [6]. One common approach in the noise adaptive filtering involves monitoring the innovation sequence, which is the difference between the predicted and the actual measurements [7]. In a Bayesian framework, one can define prior on unknown noise parameters and then try to compute the posterior. In this study, the variational approximation is used to compute the approximate posterior for which analytical solution does not exist. Variational inference based techniques have been used for state estimation and filtering in a number of recent studies. In [8] the noise covariance is modeled as a diagonal matrix whose entries are assumed to be distributed as inverse Gamma. This result is extended and used in an interactive multiple model (IMM) framework for jump Markov linear systems in [9]. In [10], the Inverse Wishart distribution is used to model the measurement noise covariance matrix for robust estimation within the variational Bayes framework. It is also shown in [10] that the mean square error of the state estimate can be

reduced by using the proposed variational Bayes measurement update. In another recent paper [11], the generalization of [10] for nonlinear systems is given.

In this contribution, a novel implementation of the PHD filter which accounts for the unknown sensor noise covariance using appropriate priors and approximate inference technique is presented. The variational Bayes approximation is used in the updates for the first time within the PHD filtering framework. In order to employ variational Bayes approximation, the recursions needed for the measurement update are derived explicitly. Moreover, the predictive likelihood of the measurements are derived which enables the use of the model introduced in [10] in target tracking applications. The resulting algorithm is easy to implement and computationally efficient. The algorithm is able to estimate jointly the state and the measurement noise covariance over the surveillance region which makes it most suitable for the applications where the sensor noise covariance is non-uniform over the region of interest such as acoustic sensors or thresholded detections dependent on the signal strength. In the current settings, each target learns the sensor characteristics independent of each other. The results are promising for a future work where the approximate posterior for the unknown sensor covariance can be learned from multiple targets jointly. There are some similarities as well as differences to the extended target tracking framework in [12], [13], [14], [15], [16]. As pointed out in the extended target tracking literature, such as [17], [12], tracking of point targets with an unknown measurement error is a limiting case of their proposed extended target tracker. Many studies have focused on the extended target tracking problem, but the unknown sensor characteristics for point targets is in general overlooked.

The rest of this paper is organized as follows: In Section II, the standard assumptions and expressions for the PHD filter are given. In Section III the assumptions and expressions for a PHD filter where sensor noise covariance is unknown are provided. In Section IV the analytical expressions for a PHD filter with unknown measurement noise covariance is derived using variational approximation. The proposed PHD filter is evaluated in Section V in a numerical simulation. The conclusions are drawn in Section VI.

## II. THE PHD RECURSIONS

The PHD filter approximates the multi-target Bayes filter by propagating its first-order moment. The filter is recursive and the update is done in two steps, namely the prediction and the correction steps. At each update, the PHD, which is the first-order moment approximation of the multi-target posterior, is propagated in time and corrected with the available measurements. Suppose we have the PHD  $D_{k|k}(\mathbf{x}|Z^{(k)})$  for the target state  $\mathbf{x}$  at time step  $k$ . The predicted PHD at time

step  $k + 1$  becomes [2]

$$D_{k+1|k}(\mathbf{x}|Z^{(k)}) = b_{k+1|k}(\mathbf{x}) + \int F_{k+1|k}(\mathbf{x}|\xi)D_{k|k}(\xi|Z^{(k)}) d\xi, \quad (1)$$

where

$$F_{k+1|k}(\mathbf{x}|\xi) \triangleq p_{S,k+1|k}(\xi)f_{k+1|k}(\mathbf{x}|\xi), \quad (2)$$

and  $f_{k+1|k}(\mathbf{x}|\xi)$  is the Markovian single target transition density,  $p_{S,k+1|k}(\xi)$  is the probability that a target with state  $\xi$  at time step  $k$  survives in time step  $k + 1$ ,  $b_{k+1|k}(\mathbf{x})$  is the target birth intensity, and  $Z^{(k)}$ , which is a subset of a measurement space  $\mathcal{Z}$ , is the measurement set at time  $k$ . The predicted expected number of targets is

$$N_{k+1|k} = \int D_{k+1|k}(\mathbf{x}|Z^{(k)}) d\mathbf{x}. \quad (3)$$

The PHD corrector step is given as

$$D_{k+1|k+1}(\mathbf{x}|Z^{(k+1)}) \cong L_{Z_{k+1}}(\mathbf{x})D_{k+1|k}(\mathbf{x}|Z^{(k)}) \quad (4)$$

where for a measurement set  $Z^{(k+1)}$  the PHD pseudo likelihood  $L_{Z_{k+1}}(\mathbf{x})$  is

$$L_{Z_{k+1}}(\mathbf{x}) \triangleq 1 - p_D(\mathbf{x}) + p_D(\mathbf{x}) \sum_{z \in Z^{(k+1)}} \frac{L_z(\mathbf{x})}{\lambda_c c(z) + \int p_D(\xi)L_z(\xi)D_{k+1|k}(\xi|Z^{(k)}) d\xi} \quad (5)$$

where,  $L_z(\mathbf{x}) = f_{k+1}(z|\mathbf{x})$  is the single target likelihood function and  $p_D(\mathbf{x})$  is the probability of detection. The sensor collects an average number of  $\lambda_c$  Poisson-distributed false alarms and the spatial distribution of the clutter is governed by the probability density  $c(z)$ . Posterior to the measurement update, the expected number of targets is

$$N_{k+1|k+1} = \int D_{k+1|k+1}(\mathbf{x}|Z^{(k+1)}) d\mathbf{x}. \quad (6)$$

In the following section the PHD filter recursion for the case where the state is augmented with the unknown sensor covariance will be presented.

### III. PHD FILTERING WITH UNKNOWN SENSOR NOISE COVARIANCE

In many applications, the quality of the measurements depends on the physical distance between the target and the sensor position. For example, in [4], the acoustic sensors provide good quality measurements for the direction of arrival estimates when the targets are close to the microphones whereas the quality of these measurements deteriorates as the targets move further away from the microphones because of the change in the received signal levels. Similar behavior can also be observed in radar tracking. In such scenarios, the measurement noise covariance is non-uniform over the surveillance region. The spatially varying covariance matrix can be estimated by inspecting the target trajectories and the target originated measurements. In this paper, the aim is to jointly estimate the sensor noise covariance and the target states by augmenting the state in the PHD filter. The target state augmentation approach is adopted in other PHD filtering studies such as [18], [19] and [20] where target state is augmented by the detection probability. Also in [13], [12] the target state is augmented by the target extent for the extended

target tracking. The augmented state is defined as follows

$$\check{\mathbf{x}} = [\mathbf{x}, \mathbf{x}^*], \quad (7)$$

where  $\mathbf{x} \in \chi^{(\mathbf{x})}$  is the target kinematic state,  $\chi^{(\mathbf{x})} \subset \mathbb{R}^{n_x}$  is the target's kinematics state space,  $\mathbf{x}^* \in \chi^{(\mathbf{x}^*)}$  augmented part of the state which will account for the unknown sensor noise covariance and  $\chi^{(\mathbf{x}^*)} \subset \mathbb{S}_{++}^{n_z}$  is the space of the unknown noise covariances,  $\check{\mathbf{x}} \in \chi^{(\mathbf{x})} \times \chi^{(\mathbf{x}^*)}$  and  $\times$  denoted the Cartesian product. Here, the conditions which are required to hold for the augmented target state in the prediction step of the PHD recursion will be emphasized;

- 1) The augmented target states evolve independently of each other. The birth of the new targets and the survival of the targets in the scene are independent.
- 2) The single target transition density  $f_{k+1|k}(\mathbf{x}, \mathbf{x}^*|\xi, \xi^*)$  is Markovian. It is further assumed that the state transition for the augmented state  $[\mathbf{x}, \mathbf{x}^*]$  can be factorized as in
$$f_{k+1|k}(\mathbf{x}, \mathbf{x}^*|\xi, \xi^*) = f_{k+1|k}(\mathbf{x}|\xi)f_{k+1|k}(\mathbf{x}^*|\xi^*). \quad (8)$$
- 3) The probability that a target with augmented state  $[\xi, \xi^*]$  at time step  $k$  survives at time step  $k + 1$  is  $p_{S,k+1|k}(\xi, \xi^*)$  which is abbreviated by  $p_S(\xi, \xi^*)$ .
- 4) The target birth intensity at time step  $k + 1$  is  $b_{k+1|k}(\mathbf{x}, \mathbf{x}^*)$ .

Under the aforementioned conditions the PHD time update becomes

$$D_{k+1|k}(\mathbf{x}, \mathbf{x}^*|Z^{(k)}) = b_{k+1|k}(\mathbf{x}, \mathbf{x}^*) + \int F_{k+1|k}(\mathbf{x}, \mathbf{x}^*|\xi, \xi^*)D_{k|k}(\xi, \xi^*|Z^{(k)}) d\xi d\xi^*, \quad (9)$$

where

$$F_{k+1|k}(\mathbf{x}, \mathbf{x}^*|\xi, \xi^*) \triangleq p_S(\xi, \xi^*)f_{k+1|k}(\mathbf{x}|\xi)f_{k+1|k}(\mathbf{x}^*|\xi^*). \quad (10)$$

The correction step of the PHD filter can be written as

$$D_{k+1|k+1}(\mathbf{x}, \mathbf{x}^*|Z^{(k+1)}) \cong L_{Z_{k+1}}(\mathbf{x}, \mathbf{x}^*)D_{k+1|k}(\mathbf{x}, \mathbf{x}^*|Z^{(k)}) \quad (11)$$

where

$$L_{Z_{k+1}}(\mathbf{x}, \mathbf{x}^*) \triangleq 1 - p_D(\mathbf{x}, \mathbf{x}^*) + \frac{\sum_{z \in Z^{(k+1)}} p_D(\mathbf{x}, \mathbf{x}^*)L_z(\mathbf{x}, \mathbf{x}^*)}{\lambda_c c(z) + \int p_D(\xi, \xi^*)L_z(\xi, \xi^*)D_{k+1|k}(\xi, \xi^*|Z^{(k)}) d\xi d\xi^*} \quad (12)$$

and

$$L_z(\mathbf{x}, \mathbf{x}^*) = f_{k+1}(z|\mathbf{x}, \mathbf{x}^*). \quad (13)$$

The following assumptions are made for the correction step;

- 1) Each target generates at most one measurement and each measurement is generated by at most one target.
- 2) The probability of detection is  $p_D(\mathbf{x}, \mathbf{x}^*) = p_{D,k+1}(\mathbf{x})$ .
- 3) The sensor collects an average number of  $\lambda_c$  Poisson-distributed false alarms and the spatial distribution of the clutter is governed by the probability density  $c(z)$ .
- 4) The predicted multi-target RFS is Poisson.

#### IV. IMPLEMENTATION

The PHD filter can be implemented using different inference techniques such as SMC and Gaussian mixture approximation. In this section, an analytical mixture implementation where  $D_{k|k}(\mathbf{x}, \mathbf{x}^*|Z^{(k)})$  is represented by a mixture will be studied. The mixture is given by

$$D_{k|k}(\mathbf{x}, \mathbf{x}^*|Z^{(k)}) = \sum_{i=1}^{n_{k|k}} w_{k|k}^{(i)} \mathcal{N}(\mathbf{x}; m_{k|k}^{(i)}, P_{k|k}^{(i)}) \mathcal{IW}(\mathbf{x}^*; v_{k|k}^{(i)}, \Psi_{k|k}^{(i)}), \quad (14)$$

where,

- $n_{k|k}$  is the number of components in the mixture;
- $w_{k|k}^{(i)}$  represents a positive weight;
- The measurement likelihood and the target motion models are assumed to follow a linear and Gaussian model i.e.,

$$f_{k+1|k}(\mathbf{x}|\xi) = \mathcal{N}(\mathbf{x}; A_k \xi, Q_k), \quad (15)$$

$$f_{k+1}(z|\mathbf{x}, \mathbf{x}^*) = \mathcal{N}(z; C_{k+1} \mathbf{x}, \mathbf{x}^*); \quad (16)$$

- The birth intensity is

$$b_{k+1|k}(\mathbf{x}, \mathbf{x}^*) = \sum_{j=1}^{n_\gamma} w_\gamma^{(j)} \mathcal{N}(\mathbf{x}; m_\gamma^{(j)}, P_\gamma^{(j)}) \times \mathcal{IW}_{n_z}(\mathbf{x}^*; \nu_\gamma^{(j)}, \Psi_\gamma^{(j)}); \quad (17)$$

- $\{z \in \mathbb{R}^{n_z}\}$  are the measurements;
- $A_k \in \mathbb{R}^{n_x \times n_x}$  and  $C_k \in \mathbb{R}^{n_z \times n_x}$  are assumed to be known state transition matrix and measurement matrix, respectively;
- $Q_k$  is the prediction noise covariance which is assumed to be known;
- Probabilities of survival and detection are assumed to be constant;  $p_S(\mathbf{x}, \mathbf{x}^*) = p_S$  and  $p_D(\mathbf{x}, \mathbf{x}^*) = p_D$ .
- The clutter is Poisson and independent of target-originated measurements and has the intensity

$$\kappa_{k+1}(z) = \lambda_c \mathcal{U}(z), \quad (18)$$

where  $\mathcal{U}(\cdot)$  is the uniform density over the surveillance area and  $\lambda_c$  is the expected number of clutter returns over the surveillance area.

The aim is to obtain an analytical approximation for the PHD  $D_{k+1|k+1}(\mathbf{x}, \mathbf{x}^*|Z^{(k+1)})$  where the exact analytical solution does not exist. To that end, an approximate recursive filter which will propagate the sufficient statistics of the approximate distributions through the time update and measurement update equations as in the GMPHD filter will be derived. The expressions for the PHD predictor and PHD corrector will be given in sections IV-A and IV-B, respectively.

##### A. PHD prediction

The expression (2) for  $F_{k+1|k}(\mathbf{x}, \mathbf{x}^*|\xi, \xi)$  can be written as

$$F_{k+1|k}(\mathbf{x}, \mathbf{x}^*|\xi, \xi) = p_S f_{k+1|k}(\mathbf{x}|\xi) f_{k+1|k}(\mathbf{x}^*|\xi) = f_{k+1|k}(\mathbf{x}^*|\xi) F_{k+1|k}(\mathbf{x}|\xi), \quad (19)$$

where  $F_{k+1|k}(\mathbf{x}|\xi)$  is defined in (2). The resulting PHD predictor is in the form,

$$\begin{aligned} D_{k+1|k}(\mathbf{x}, \mathbf{x}^*|Z^{(k)}) &= b_{k+1|k}(\mathbf{x}, \mathbf{x}^*) \\ &+ \int f_{k+1|k}(\mathbf{x}^*|\xi) F_{k+1|k}(\mathbf{x}|\xi) D_{k|k}(\xi, \xi|Z^{(k)}) d\xi d\xi^* \\ &= b_{k+1|k}(\mathbf{x}, \mathbf{x}^*) + \sum_{i=1}^{n_{k|k}} w_{k|k}^{(i)} \\ &\times \left\{ \int F_{k+1|k}(\mathbf{x}|\xi) \mathcal{N}(\xi; m_{k|k}^{(i)}, P_{k|k}^{(i)}) d\xi \right\} \\ &\times \left\{ \int f_{k+1|k}(\mathbf{x}^*|\xi) \mathcal{IW}(\xi; v_{k|k}^{(i)}, \Psi_{k|k}^{(i)}) d\xi^* \right\}, \end{aligned} \quad (20)$$

therefore the prediction step of the PHD for the target state and the sensor noise covariance can be done independently yielding the mixture

$$D_{k+1|k}(\mathbf{x}, \mathbf{x}^*|Z^{(k)}) \approx \sum_{i=1}^{n_{k+1|k}} w_{k+1|k}^{(i)} \mathcal{N}(\mathbf{x}; m_{k+1|k}^{(i)}, P_{k+1|k}^{(i)}) \mathcal{IW}(\mathbf{x}^*; v_{k+1|k}^{(i)}, \Psi_{k+1|k}^{(i)}). \quad (21)$$

The expression for the PHD time update of a Gaussian component representing the target state follows standard Kalman filter time update equations,

$$m_{k+1|k}^{(i)} = A_k m_{k|k}^{(i)} \quad (22)$$

$$P_{k+1|k}^{(i)} = A_k P_{k|k}^{(i)} A_k^T + Q_k. \quad (23)$$

If the sufficient statistics are assumed to be slowly varying, exponential forgetting strategy [21] can be used to account for possible changes in the parameters in time. In our recent contribution, it is shown that using the exponential forgetting factor will produce maximum entropy distribution in the time update for the processes which are slowly varying with unknown dynamics but bounded by a Kullback-Leibler (KL) distance constraint [22]. Forgetting factor is applied to the sufficient statistics of the inverse Wishart and Gaussian distribution as follows.

$$\Psi_{k+1|k}^{(i)} = \lambda_\Psi \Psi_{k|k}^{(i)}, \quad (24)$$

$$\nu_{k+1|k}^{(i)} = \lambda_\nu \nu_{k|k}^{(i)} \quad (25)$$

where the forgetting factor  $\lambda$  is a scalar real number which is  $0 < \lambda < 1$ . The exponential forgetting weighs the effects of the past measurements on the sufficient statistics. The use of this operation corresponds to the application of an exponential window with effective length  $h = \frac{1}{1-\lambda}$ . The statistics roughly relies on the measurements within the last  $h$  time instances.

##### B. PHD correction

The PHD corrector step equation (4) can be split in two parts

$$\begin{aligned} D_{k+1|k+1}(\mathbf{x}, \mathbf{x}^*|Z^{(k+1)}) \\ = D_{k+1|k+1}^\emptyset(\mathbf{x}, \mathbf{x}^*|Z^{(k+1)}) + D_{k+1|k+1}^Z(\mathbf{x}, \mathbf{x}^*|Z^{(k+1)}), \end{aligned} \quad (26)$$

where

$$D_{k+1|k+1}^\emptyset(\mathbf{x}, \mathbf{x}^*|Z^{(k+1)}) = [1 - p_D] D_{k+1|k}(\mathbf{x}, \mathbf{x}^*|Z^{(k)}) \quad (27)$$

and

$$\begin{aligned}
& D_{k+1|k+1}^Z(\mathbf{x}, \mathbf{x}^* | Z^{(k+1)}) \\
&= \sum_{z \in Z^{(k+1)}} \frac{p_D L_z(\mathbf{x}, \mathbf{x}^*) D_{k+1|k}(\mathbf{x}, \mathbf{x}^* | Z^{(k)})}{\lambda_c c(z) + \int p_D L_z(\xi, \xi) D_{k+1|k}(\xi, \xi | Z^{(k)}) d\xi d\xi^*}. \quad (28)
\end{aligned}$$

$D_{k+1|k+1}^\theta(\mathbf{x}, \mathbf{x}^* | Z^{(k+1)})$  represents the components which account for no detection cases while  $D_{k+1|k+1}^Z(\mathbf{x}, \mathbf{x}^* | Z^{(k+1)})$  represents the components updated by the received measurements at time  $k+1$ .

The numerator of the rational expression in (28) can be written as

$$\begin{aligned}
& p_D L_z(\mathbf{x}, \mathbf{x}^*) D_{k+1|k}(\mathbf{x}, \mathbf{x}^* | Z^{(k)}) \\
&= \sum_{i=1}^{n_{k+1|k}} w_{k+1|k}^{(i)} p_D \mathcal{N}(z; C_{k+1} \mathbf{x}, \mathbf{x}^*) \quad (29) \\
&\quad \times \mathcal{N}(\mathbf{x}; m_{k+1|k}^{(i)}, P_{k+1|k}^{(i)}) \mathcal{IW}(\mathbf{x}^*; \nu_{k+1|k}^{(i)}, \Psi_{k+1|k}^{(i)}).
\end{aligned}$$

The integral in the denominator of (28) can be written as

$$\begin{aligned}
& \sum_{i=1}^{n_{k+1|k}} w_{k+1|k}^{(i)} p_D \int \mathcal{N}(z; C_{k+1} \xi, \xi^*) \quad (30) \\
&\quad \times \mathcal{N}(\xi; m_{k+1|k}^{(i)}, P_{k+1|k}^{(i)}) \mathcal{IW}(\xi^*; \nu_{k+1|k}^{(i)}, \Psi_{k+1|k}^{(i)}) d\xi d\xi^*.
\end{aligned}$$

Unfortunately, the expressions in equations (29) and (30) do not have closed forms. Instead, an approximation of (29) which has the form

$$\begin{aligned}
& \sum_{i=1}^{n_{k+1|k}} w^{(i)} \mathcal{N}(\mathbf{x}, m_{k+1|k+1}^{(i)}, P_{k+1|k+1}^{(i)}) \times \\
&\quad \mathcal{IW}(\mathbf{x}^*; \nu_{k+1|k+1}^{(i)}, \Psi_{k+1|k+1}^{(i)})
\end{aligned}$$

is sought to have an approximate but analytical recursive PHD update. Before continuing with the derivations, the following lemma is introduced.

*Lemma 1:* Let  $x$ ,  $x^*$  and  $z$  be three random variables with the factorized joint density as

$$p(x, x^*, z) = p(z|x, x^*) p(x) p(x^*), \quad (31)$$

where  $x \in \mathbb{R}^{n_x}$ ,  $z \in \mathbb{R}^{n_z}$ ,  $x^* \in \mathbb{R}^{n_x \times n_x}$  and

$$p(z|x, x^*) = \mathcal{N}(z; Cx, x^*), \quad (32)$$

$$p(x) = \mathcal{N}(x; m_{k+1|k}, P_{k+1|k}), \quad (33)$$

$$p(x^*) = \mathcal{IW}(x^*; \nu_{k+1|k}, \Psi_{k+1|k}). \quad (34)$$

Then,

(a) The posterior  $p(x, x^* | z)$  can be approximated via minimizing the KL distance by the product of the two densities  $\hat{q}(x)$  and  $\hat{q}(x^*)$  where

$$\hat{q}(x) = \mathcal{N}(x; m_{k+1|k+1}, P_{k+1|k+1}), \quad (35)$$

$$\hat{q}(x^*) = \mathcal{IW}(x^*; \nu_{k+1|k+1}, \Psi_{k+1|k+1}) \quad (36)$$

and the parameters  $m_{k+1|k+1}$ ,  $P_{k+1|k+1}$ ,  $\nu_{k+1|k+1}$ ,  $\Psi_{k+1|k+1}$  are the final values of the parameters, obtained by repeating

the iterations given below.

$$\begin{aligned}
\Delta^{(j)} &= C P_{k+1|k+1}^{(j)} C^T \\
&\quad + (z - C m_{k+1|k+1}^{(j)})(z - C m_{k+1|k+1}^{(j)})^T, \quad (37)
\end{aligned}$$

$$\Omega^{(j)} = (\nu_{k+1|k+1}^{(j)} - n_z - 1) (\Psi_{k+1|k+1}^{(j)})^{-1}, \quad (38)$$

$$P_{k+1|k+1}^{(j+1)} = ((P_{k+1|k}^{(j)})^{-1} + C^T \Omega^{(j)} C)^{-1}, \quad (39)$$

$$m_{k+1|k+1}^{(j+1)} = P_{k+1|k+1}^{(j+1)} ((P_{k+1|k}^{(j)})^{-1} m_{k+1|k} + C^T \Omega^{(j)} z), \quad (40)$$

$$\nu_{k+1|k+1}^{(j+1)} = \nu_{k+1|k} + 1, \quad (41)$$

$$\Psi_{k+1|k+1}^{(j+1)} = \Psi_{k+1|k} + \Delta^{(j)} \quad (42)$$

where  $a^{(j)}$  denotes the value of the variable  $a$  at the  $j^{\text{th}}$  iteration.

(b) The predictive likelihood  $p(z) = \int p(x, x^*, z) dx dx^*$  can be approximated via minimizing the KL distance by the variational lower bound given by

$$\begin{aligned}
\hat{q}(z) &= \exp \left\{ -\frac{1}{2} n_z \log(\pi) + \frac{1}{2} n_x \log(2\pi) \right. \\
&\quad + \log \mathcal{N}(m_{k+1|k+1}; m_{k+1|k}, P_{k+1|k}) - \frac{1}{2} \text{tr}(P_{k+1|k}^{-1} P_{k+1|k+1}) \\
&\quad + \frac{1}{2} \log |P_{k+1|k+1}| + \frac{1}{2} n_x + \frac{1}{2} (\nu_{k+1|k} - n_z - 1) \log |\Psi_{k+1|k}| \\
&\quad - \log \Gamma_{n_z} \left[ \frac{1}{2} (\nu_{k+1|k} - n_z - 1) \right] \\
&\quad - \frac{1}{2} (\nu_{k+1|k+1} - n_z - 1) \log |\Psi_{k+1|k+1}| \\
&\quad \left. + \log \Gamma_{n_z} \left[ \frac{1}{2} (\nu_{k+1|k+1} - n_z - 1) \right] \right\}. \quad (43)
\end{aligned}$$

*Proof:* For the proof see appendix A.  $\blacksquare$

*Corollary 1:* The approximate densities  $\hat{q}(x)$ ,  $\hat{q}(x^*)$  and  $\hat{q}(z)$  provide an approximation of the joint density  $p(x, x^*, z)$  as follows,

$$p(x, x^*, z) \approx \hat{q}(x) \hat{q}(x^*) \hat{q}(z), \quad (44)$$

therefore,

$$\begin{aligned}
& \mathcal{N}(z; Cx, x^*) \mathcal{N}(x; m_{k+1|k}, P_{k+1|k}) \mathcal{IW}(x^*; \nu_{k+1|k}, \Psi_{k+1|k}) \approx \\
& \mathcal{N}(x; m_{k+1|k+1}, P_{k+1|k+1}) \mathcal{IW}(x^*; \nu_{k+1|k+1}, \Psi_{k+1|k+1}) \hat{q}(z). \quad (45)
\end{aligned}$$

Using Lemma 1 and Corollary 1 the PHD posterior can be approximated by

$$\begin{aligned}
& D_{k+1|k+1}(\mathbf{x}, \mathbf{x}^* | Z^{(k+1)}) \approx D_{k+1|k+1}^\theta(\mathbf{x}, \mathbf{x}^* | Z^{(k+1)}) \\
&\quad + \sum_{z \in Z^{(k+1)}} \frac{\sum_{i=1}^{n_{k+1|k}} p_D w_{k+1|k} \hat{q}^{(i)}(\mathbf{x}, \mathbf{x}^*, z)}{\lambda_c c(z) + \sum_{i=1}^{n_{k+1|k}} w_{k+1|k} p_D \hat{q}^{(i)}(z)}, \quad (46)
\end{aligned}$$

where

$$\begin{aligned}
& \hat{q}^{(i)}(\mathbf{x}, \mathbf{x}^*, z) = \mathcal{N}(x; m_{k+1|k+1}^{(i)}, P_{k+1|k+1}^{(i)}) \\
&\quad \times \mathcal{IW}(x^*; \nu_{k+1|k+1}^{(i)}, \Psi_{k+1|k+1}^{(i)}) \hat{q}^{(i)}(z). \quad (47)
\end{aligned}$$

The superscript on  $\hat{q}^{(i)}(z)$  indicates that (43) is evaluated using

parameters with superscript  $(i)$  such as  $m_{k+1|k+1}^{(i)}$ ,  $P_{k+1|k+1}^{(i)}$ ,  $\Psi_{k+1|k}^{(i)}$  et cetera. The pseudo-code for the proposed filter is given in Table I. In order to implement the proposed algorithm, a mixture reduction and state extraction step are needed which will be discussed in the following subsection.

### C. Mixture reduction and state extraction

The mixture implementation of the PHD filter results in an exponential increase in the number of components, which needs to be reduced whenever necessary. In the implementation of the proposed PHD filter, the intensity is represented by a mixture of Gaussian Inverse Wishart densities for which there are existing mixture reduction algorithms such as [23]. In this implementation, the multi-target tracking flowchart proposed in [24, Section II] is used which consists of two reduction algorithms; one for computational feasibility and one for state extraction. Pruning is used to keep the number of components less than a predefined bound to maintain the computations feasibility. The reduction algorithm for the state extraction first prunes some components, then the remaining components are reduced by merging using the algorithm proposed in [5, Table II]. The state extraction is done by first rounding the cardinality to find the number of peaks to be extracted and then extracting the peaks by finding the components with highest weight in the intensity.

## V. NUMERICAL SIMULATION

The performance of the algorithm is illustrated in a multi-target tracking scenario where four targets appear in a two dimensional surveillance region. Nearly constant velocity model is used in the generation of the measurements and in the filter. The state vector consists of the position and the velocity,  $\mathbf{x} = [p_x, \dot{p}_x, p_y, \dot{p}_y]^T$ . A sensor collects noisy measurements of the target's Cartesian positions corrupted by white Gaussian noise where target originated measurements are generated according to  $f(z|\mathbf{x}) = \mathcal{N}(z; C_{k+1}\mathbf{x}, R_{k+1})$ . The parameters of the target motion model and the measurement equation are

$$A_k = \begin{bmatrix} I_2 & \tau I_2 \\ 0_2 & I_2 \end{bmatrix}, \quad Q_k = \sigma_\nu^2 \begin{bmatrix} \frac{\tau^4}{4} I_2 & \frac{\tau^3}{2} I_2 \\ \frac{\tau^3}{2} I_2 & \tau^2 I_2 \end{bmatrix},$$

$$C_{k+1} = [I_2 \quad 0_2], \quad R_{k+1} = \sigma_e^2 I_2,$$

$$\tau = 1s, \quad \sigma_\nu = 0.3m/s^2,$$

where,  $I_2$  and  $0_2$  are  $2 \times 2$  identity and zero matrix, respectively. Clutter is generated uniformly over the surveillance region with a rate of 5 per scan. The surveillance area is roughly  $1500m \times 1500m$ . The measurements are generated according to a known probability of detection  $p_D = 0.98$ . The mixture reduction and state extraction is done according to Section IV-C where 125 components are kept to represent the PHD and the rest are pruned. The number of iterations in the variational update is set to 5. In the filter the survival probability of the targets is set to  $p_S = 0.99$ .

A scenario where a sensor performs worse than its nominal performance is simulated. The nominal value of the sensor's standard deviation is  $\sigma_e = 10m$ , but a failure causes the sensor to provide measurements with standard deviation of  $\sigma_{e,TRUE} = 20m$ . The results of GMPHD and the proposed adaptive PHD algorithm with the abbreviation VBPHD are compared. In the VBPHD, the initial parameters of the inverse Wishart components in the birth density are chosen as  $\nu_\gamma = 20$  and  $\Psi_\gamma = R_k \times (\nu_\gamma - 6)$  which yields the expected value of the measurement covariance to coincide with the nominal noise covariance of the GMPHD. The exponential forgetting factor used in the prediction update is set to  $\lambda = 0.99$ . The target

Table I. PSEUDO CODE FOR THE ADAPTIVE PHD FILTER

**Require:**  $\{w_{k|k}^{(i)}, m_{k|k}^{(i)}, P_{k|k}^{(i)}, \nu_{k|k}^{(i)}, \Psi_{k|k}^{(i)}\}_{i=1}^{n_{k|k}}$ ,  
**PHD predictor:** prediction for the target birth  
 $i = 0$ ,  
**for**  $j = 1 : n_{\gamma, k+1}$  **do**  
 $i = i + 1$ ,  
 $w_{k+1|k}^{(i)} = w_{\gamma, k+1}^{(j)}$ ,  $m_{k+1|k}^{(i)} = m_{\gamma, k+1}^{(j)}$ ,  $P_{k+1|k}^{(i)} = P_{\gamma, k+1}^{(j)}$ ,  
 $\nu_{k+1|k}^{(i)} = \nu_{\gamma, k+1}^{(j)}$ ,  $\Psi_{k+1|k}^{(i)} = \Psi_{\gamma, k+1}^{(j)}$ ,  
**end for**  
**PHD predictor:** prediction for the existing targets  
**for**  $j = 1 : n_{k|k}$  **do**  
 $i = i + 1$ ,  
 $w_{k+1|k}^{(i)} = p_S w_{k|k}^{(j)}$ ,  $m_{k+1|k}^{(i)} = A_k m_{k|k}^{(j)}$ ,  
 $P_{k+1|k}^{(i)} = Q_k + A_k P_{k|k}^{(j)} A_k^T$ ,  
 $\nu_{k+1|k}^{(i)} = \lambda_\nu \nu_{k|k}^{(j)}$ ,  $\Psi_{k+1|k}^{(i)} = \lambda_\Psi \Psi_{k|k}^{(j)}$ ,  
**end for**  
 $n_{k+1|k} = i$ ,  
**PHD corrector:** measurement update  
**for**  $j = 1 : n_{k+1|k}$  **do**  
 $i = i + 1$ ,  
 $w_{k+1|k+1}^{(i)} = (1 - p_D) w_{k+1|k}^{(j)}$ ,  $m_{k+1|k+1}^{(i)} = m_{k+1|k}^{(j)}$ ,  
 $P_{k+1|k+1}^{(i)} = P_{k+1|k}^{(j)}$ ,  $\nu_{k+1|k+1}^{(i)} = \nu_{k+1|k}^{(j)}$ ,  
 $\Psi_{k+1|k+1}^{(i)} = \Psi_{k+1|k}^{(j)}$ ,  
**end for**  
 $l = 0$ ,  
**for**  $z \in Z^{(k+1)}$  **do**  
 $l = l + 1$ ,  
**for**  $j = 1 : n_{k+1|k}$  **do**  
 $\nu = \nu_{k+1|k}^{(j)}$ ,  $\Psi = \Psi_{k+1|k}^{(j)}$ ,  $m = m_{k+1|k}^{(j)}$ ,  $P = P_{k+1|k}^{(j)}$ ,  
 $\Lambda = P^{-1}$ ,  
**repeat**  
 $\Omega = (\nu - n_z - 1)\Psi^{-1}$ ,  
 $\Delta = C_{k+1} P C_{k+1}^T + (z - C_{k+1} m)(z - C_{k+1} m)^T$ ,  
 $P = (\Lambda + C_{k+1}^T \Omega C_{k+1})^{-1}$ ,  
 $m = P(\Lambda m_{k+1|k} + C_{k+1}^T \Omega z)$ ,  
 $\Psi = \Psi_{k+1|k} + \Delta$ ,  
 $\nu = \nu_{k+1|k} + 1$ ,  
**until converged**  
 $m_{k+1|k+1}^{(ln_{k+1|k}+j)} = m$ ,  $P_{k+1|k+1}^{(ln_{k+1|k}+j)} = P$ ,  
 $\nu_{k+1|k+1}^{(ln_{k+1|k}+j)} = \nu$ ,  $\Psi_{k+1|k+1}^{(ln_{k+1|k}+j)} = \Psi$ ,  
Calculate  $q^{(j)}(z)$  using (43) and parameters of the  $j^{th}$  component,  
 $w_{k+1|k+1}^{(ln_{k+1|k}+j)} = p_D w_{k+1|k}^{(j)} q^{(j)}(z)$ ,  
**end for**  
 $w_{k+1|k+1}^{(ln_{k+1|k}+j)} = \frac{w_{k+1|k+1}^{(ln_{k+1|k}+j)}}{\kappa_{k+1}(z) + \sum_{i=1}^{n_{k+1|k}} w_{k+1|k+1}^{(ln_{k+1|k}+i)}}$  for  $j = 1, \dots, n_{k+1|k}$ ,  
**end for**  
 $n_{k+1|k+1} = (l + 1)n_{k+1|k}$ ,  
**return**  $\{w_{k+1|k+1}^{(i)}, m_{k+1|k+1}^{(i)}, P_{k+1|k+1}^{(i)}, \nu_{k+1|k+1}^{(i)}, \Psi_{k+1|k+1}^{(i)}\}_{i=1}^{n_{k+1|k+1}}$ .

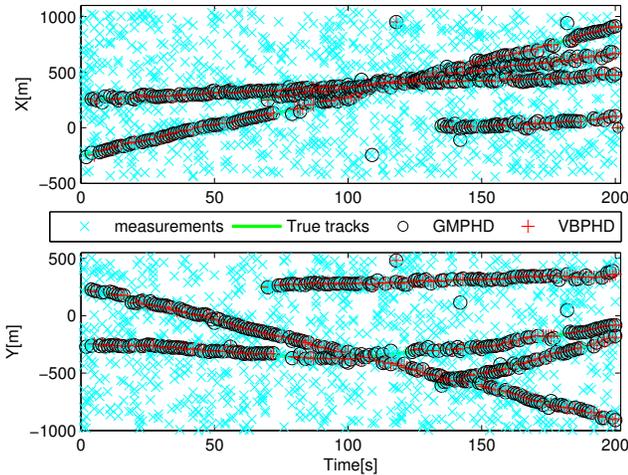


Figure 1. The adaptive PHD filter and the GMPHD filter in an MTT scenario: The filter's estimates and measurements in  $X$  and  $Y$  coordinate are plotted versus time.

birth is assumed to be partially uniform and is implemented according to [25].

In Figure 1, the measurements, the estimated trajectories of the targets by GMPHD and VBPHD and the true target trajectories are plotted. The evaluation is made with respect to the optimal subpattern assignment (OSPA) metric [26] and is shown in Figure 2. The parameters of the OSPA metric are chosen as  $p = 1$  and  $c = 60$ . The average OSPA for the VBPHD is 63 while the average OSPA for the GMPHD is 96. With its ability to adapt to the unknown noise covariance, VBPHD outperforms GMPHD. As it can be seen from Figure 2, in the second half of the simulation the difference between OSPA values is more significant than the first half which illustrates the result of the online learning of the noise parameters. In Figure 3, the estimated standard deviations of the individual components having the top three highest weights and top three highest degrees of freedom in the posterior PHD are plotted. As it can be seen in Figure 3, the components learn the measurement covariance in time, and the estimate of the measurement noise standard deviation converges to the true value. The choice of the components with highest weight is due to the fact that such components are more likely to present the true target and are more likely to have been updated by a target oriented measurement. On the other hand the components with the largest degrees of freedom are the components which have survived the longest and have collected more information about the statistics of the noise.

## VI. CONCLUSION

In this work, an analytical implementation of the PHD filter for the applications where the sensor noise covariance can be assumed as unknown and can vary spatially over the surveillance region is proposed. Variational Bayes approximation technique is used to approximate the posteriors which do not have closed form expressions. In the simulation results, the proposed method outperforms the GMPHD by using its ability to adapt to unknown sensor noise covariance. In the current settings, each component of the intensity learns the sensor characteristics independently. The results are promising for a future work where the approximate posterior for the unknown sensor covariance can be learned from multiple targets jointly.

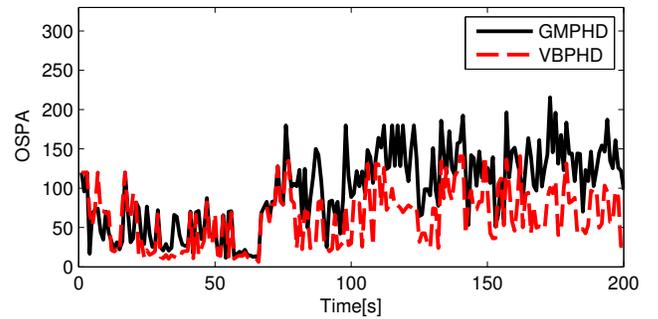


Figure 2. The OSPA metric is plotted versus time for the adaptive PHD and the GMPHD filter. Average OSPA for VBPHD is 63 while the average OSPA for the GMPHD is 96.

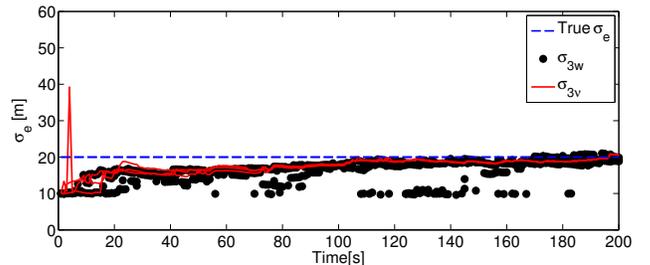


Figure 3. Estimated measurement noise standard deviations of the components having the top three highest weights ( $\sigma_{3w}$ ) in VBPHD is plotted in black together with the true value for the standard deviation of sensor noise in the  $X$ -dimension. The red curve is the estimated measurement noise standard deviations of the components having the top three highest degrees of freedom ( $\sigma_{3v}$ ) in VBPHD.

## ACKNOWLEDGMENT

The authors would like to thank the frame project grant Extended Target Tracking (621-2010-4301), funded by the Swedish Research Council (VR) for financial support. We also thank Umüt Orguner for proof reading the paper.

## REFERENCES

- [1] R. Mahler, "Multitarget Bayes filtering via first-order multitarget moments," *Aerospace and Electronic Systems, IEEE Transactions on*, vol. 39, no. 4, pp. 1152–1178, October 2003.
- [2] R. P. S. Mahler, *Statistical Multisource-Multitarget Information Fusion*. Norwood, MA, USA: Artech House, Inc., 2007.
- [3] B.-N. Vo, S. Singh, and A. Doucet, "Sequential Monte Carlo methods for multitarget filtering with random finite sets," *Aerospace and Electronic Systems, IEEE Transactions on*, vol. 41, no. 4, pp. 1224–1245, October 2005.
- [4] E. Özkan, M. Guldogan, U. Orguner, and F. Gustafsson, "Ground multiple target tracking with a network of acoustic sensor arrays using PHD and CPHD filters," in *Information Fusion (FUSION), 2011 Proceedings of the 14th International Conference on*, July 2011, pp. 1–8.
- [5] B.-N. Vo and W.-K. Ma, "The Gaussian mixture probability hypothesis density filter," *Signal Processing, IEEE Transactions on*, vol. 54, no. 11, pp. 4091–4104, November 2006.
- [6] R. Mehra, "On the identification of variances and adaptive Kalman filtering," *Automatic Control, IEEE Transactions on*, vol. 15, no. 2, pp. 175 – 184, April 1970.
- [7] —, "Approaches to adaptive filtering," *Automatic Control, IEEE Transactions on*, vol. 17, no. 5, pp. 693 – 698, October 1972.
- [8] S. Särkkä and A. Nummenmaa, "Recursive noise adaptive Kalman filtering by variational Bayesian approximations," *Automatic Control, IEEE Transactions on*, vol. 54, no. 3, pp. 596 –600, March 2009.

- [9] W. Li and Y. Jia, "State estimation for jump Markov linear systems by variational Bayesian approximation," *Control Theory Applications, IET*, vol. 6, no. 2, pp. 319–326, 2012.
- [10] G. Agamennoni, J. Nieto, and E. Nebot, "Approximate inference in state-space models with heavy-tailed noise," *Signal Processing, IEEE Transactions on*, vol. 60, no. 10, pp. 5024–5037, October 2012.
- [11] R. Piche, S. Särkkä, and J. Hartikainen, "Recursive outlier-robust filtering and smoothing for nonlinear systems using the multivariate student-t distribution," in *Machine Learning for Signal Processing (MLSP), 2012 IEEE International Workshop on*, September 2012, pp. 1–6.
- [12] C. Lundquist, K. Granström, and U. Orguner, "An extended target CPHD filter and a Gamma Gaussian inverse Wishart implementation," *Selected Topics in Signal Processing, IEEE Journal of*, to appear in 2013.
- [13] K. Granström and U. Orguner, "A PHD filter for tracking multiple extended targets using random matrices," *Signal Processing, IEEE Transactions on*, vol. 60, no. 11, pp. 5657–5671, November 2012.
- [14] M. Feldmann, D. Franken, and W. Koch, "Tracking of extended objects and group targets using random matrices," *Signal Processing, IEEE Transactions on*, vol. 59, no. 4, pp. 1409–1420, April 2011.
- [15] M. Feldmann and W. Koch, "Comments on Bayesian approach to extended object and cluster tracking using random matrices," *Aerospace and Electronic Systems, IEEE Transactions on*, vol. 48, no. 2, pp. 1687–1693, April 2012.
- [16] U. Orguner, "A variational measurement update for extended target tracking with random matrices," *Signal Processing, IEEE Transactions on*, vol. 60, no. 7, pp. 3827–3834, July 2012.
- [17] J. W. Koch, "Bayesian approach to extended object and cluster tracking using random matrices," *Aerospace and Electronic Systems, IEEE Transactions on*, vol. 44, no. 3, pp. 1042–1059, July 2008.
- [18] R. Mahler, B.-T. Vo, and B.-N. Vo, "CPHD filtering with unknown clutter rate and detection profile," *Signal Processing, IEEE Transactions on*, vol. 59, no. 8, pp. 3497–3513, August 2011.
- [19] R. Mahler and A. El-Fallah, "CPHD and PHD filters for unknown backgrounds, part III: tractable multitarget filtering in dynamic clutter," *Signal and Data Processing of Small Targets, SPIE Proc.*, vol. 7698, 2010.
- [20] —, "CPHD filtering with unknown probability of detection," *Signal Processing, Sensor Fusion, and Target Recognition XIX, SPIE Proc.*, vol. 7697, 2010.
- [21] R. Kulhavy and M. B. Zarrop, "On a general concept of forgetting," *International Journal of Control*, vol. 58, no. 4, pp. 905–924, 1993.
- [22] E. Özkan, V. Smidl, S. Saha, C. Lundquist, and F. Gustafsson, "Marginalized adaptive particle filtering for non-linear models with unknown time-varying noise parameters," to appear in *Automatica*.
- [23] K. Granström and U. Orguner, "On the reduction of Gaussian inverse Wishart mixtures," in *Information Fusion (FUSION), 2012 15th International Conference on*, July 2012, pp. 2162–2169.
- [24] T. Ardeshiri, U. Orguner, C. Lundquist, and T. Schön, "On mixture reduction for multiple target tracking," in *Information Fusion (FUSION), 2012 15th International Conference on*, July 2012, pp. 692–699.
- [25] M. Beard, B.-T. Vo, B.-N. Vo, and S. Arulampalam, "Gaussian mixture PHD and CPHD filtering with partially uniform target birth," in *Information Fusion (FUSION), 2012 15th International Conference on*, July 2012, pp. 535–541.
- [26] D. Schuhmacher, B.-T. Vo, and B.-N. Vo, "A consistent metric for performance evaluation of multi-object filters," *Signal Processing, IEEE Transactions on*, vol. 56, no. 8, pp. 3447–3457, August 2008.
- [27] C. M. Bishop, *Pattern Recognition and Machine Learning*. Springer, 2007.
- [28] D. Tzikas, A. Likas, and N. Galatsanos, "The variational approximation for Bayesian inference," *IEEE Signal Process. Mag.*, vol. 25, no. 6, pp. 131–146, November 2008.
- [29] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge University Press, March 2004.
- [30] M. J. Beal, "Variational algorithms for approximate Bayesian inference," Ph.D. dissertation, Gatsby Computational Neuroscience Unit, University College London, 2003.
- [31] A. Gupta and D. Nagar, *Matrix Variate Distributions*, ser. Chapman & Hall/CRC Monographs and Surveys in Pure and Applied Mathematics. Chapman & Hall, 2000.

## APPENDIX A

The well-known technique of variational inference [27, Chapter 10][28] will be used to find the factorized approximate density

$\hat{q}(x)\hat{q}^*(x)$  which minimizes the KL distance between the true posterior and its approximation. A similar approach is followed in [16],

$$\hat{q}(x)\hat{q}^*(x) = \operatorname{argmin}_{q(x)q^*(x)} \operatorname{KL}(q(x)q^*(x) || p(x, \hat{x}|z)). \quad (48)$$

The minimization is done with respect to parameters of  $q(x)$  and  $q^*(x)$ . The solution to the optimization problem is obtained iteratively by optimizing with respect to only one of the multiplicative factors and fixing the other to its last estimated value. It can be shown that the factorized solutions for  $\hat{q}(x)$  and  $\hat{q}^*(x)$  minimizing the KL distance satisfy the following equations [27, Chapter 10].

$$\log \hat{q}(x) = \mathbb{E}_{\hat{q}^*(x)} [\log p(x, \hat{x}, z)] + c_x, \quad (49a)$$

$$\log \hat{q}^*(x) = \mathbb{E}_{\hat{q}(x)} [\log p(x, \hat{x}, z)] + c_{x^*}, \quad (49b)$$

where  $c_x$  and  $c_{x^*}$  are constants. The solutions for  $\hat{q}(x)$  and  $\hat{q}^*(x)$  can be reached by alternating between the two equations. Because of the convexity of the problem, the convergence is guaranteed for the iterations [29], [30].

In the derivations, for the sake of convenience all normalization constants are absorbed in the term  $\operatorname{const}$  and whenever needed the normalization constant is instated by inspection. The  $(j+1)^{th}$  iterates of  $\hat{q}(x)$  and  $\hat{q}^*(x)$ , denoted as  $\hat{q}^{(j+1)}(x)$  and  $\hat{q}^{(j+1)}(x^*)$  respectively.

From the definitions given in (31), (34) and (49a)  $\hat{q}^{(j+1)}(x)$  can be determined as follows,

$$\begin{aligned} \log \hat{q}^{(j+1)}(x) &= \mathbb{E}_{\hat{q}^{(j)}(x^*)} [\log p(z|x, \hat{x})] \\ &\quad + \log \mathcal{N}(x; m_{k+1|k}, P_{k+1|k}) + \operatorname{const} \\ &= -0.5 \operatorname{tr} \left( \mathbb{E}_{\hat{q}^{(j)}(x^*)} [(\hat{x}^*)^{-1}] (z - Cx)(z - Cx)^T \right) \\ &\quad + \log \mathcal{N}(x; m_{k+1|k}, P_{k+1|k}) + \operatorname{const} \\ &= \log \mathcal{N}(z; Cx, \mathbb{E}_{\hat{q}^{(j)}(x^*)} [(\hat{x}^*)^{-1}]) \\ &\quad + \log \mathcal{N}(x; m_{k+1|k}, P_{k+1|k}) + \operatorname{const}. \end{aligned}$$

Hence,  $\hat{q}^{(j+1)}(x)$  is given as

$$\hat{q}^{(j+1)}(x) = \mathcal{N}(x; m_{k+1|k+1}^{(j+1)}, P_{k+1|k+1}^{(j+1)}), \quad (50)$$

where the definition of the Inverse Wishart density given by [31] is used and

$$m_{k+1|k+1}^{(j+1)} = P_{k+1|k+1}^{(j+1)} \times ((P_{k+1|k})^{-1} m_{k+1|k} + C^T \mathbb{E}_{\hat{q}^{(j)}(x^*)} [(\hat{x}^*)^{-1}] z), \quad (51a)$$

$$P_{k+1|k+1}^{(j+1)} = ((P_{k+1|k})^{-1} + C^T \mathbb{E}_{\hat{q}^{(j)}(x^*)} [(\hat{x}^*)^{-1}] C)^{-1}. \quad (51b)$$

Similarly,  $\hat{q}^{(j+1)}(x^*)$  can be determined using (49b)

$$\begin{aligned} \log \hat{q}^{(j+1)}(x^*) &= \mathbb{E}_{\hat{q}^{(j)}(x)} [\log p(z|x, \hat{x})] \\ &\quad + \log \mathcal{IW}(\hat{x}^*; \nu_{k+1|k}, \Psi_{k+1|k}) + \operatorname{const} \\ &= -0.5 \log |\hat{x}^*| \\ &\quad - 0.5 \operatorname{tr} \left( (\hat{x}^*)^{-1} \mathbb{E}_{\hat{q}^{(j)}(x)} [(z - Cx)(z - Cx)^T] \right) \\ &\quad + \log \mathcal{IW}(\hat{x}^*; \nu_{k+1|k}, \Psi_{k+1|k}) + \operatorname{const}, \end{aligned}$$

which gives

$$\hat{q}^{(j+1)}(x^*) = \mathcal{IW}(\hat{x}^*; \nu_{k+1|k+1}^{(j+1)}, \Psi_{k+1|k+1}^{(j+1)}), \quad (52)$$

where

$$\nu_{k+1|k+1}^{(j+1)} = \nu_{k+1|k} + 1, \quad (53a)$$

$$\Psi_{k+1|k+1}^{(j+1)} = \Psi_{k+1|k} + \mathbb{E}_{\hat{q}^{(j)}(x)} [(z - Cx)(z - Cx)^T]. \quad (53b)$$

Now the expected values  $\mathbb{E}_{\hat{q}^{(j)}(\hat{x})}[(\hat{x}^*)^{-1}]$  and  $\mathbb{E}_{\hat{q}^{(j)}(x)}[(z - Cx)(z - Cx)^T]$  can be calculated and plugged into equations (51) and (53b), respectively. Since  $(\hat{x}^*)^{-1}$  is distributed according to a Wishart distribution,

$$\mathbb{E}_{\hat{q}^{(j)}(\hat{x}^*)}[(\hat{x}^*)^{-1}] = (\nu_{k+1|k+1}^{(j)} - n_z - 1)(\Psi_{k+1|k+1}^{(j)})^{-1}, \quad (54)$$

and since  $x$  is distributed according to a Gaussian distribution,

$$\mathbb{E}_{\hat{q}^{(j)}(x)} [(z - Cx)(z - Cx)^T] = CP_{k+1|k+1}^{(j)}C^T + (z - Cm_{k+1|k+1}^{(j)})(z - Cm_{k+1|k+1}^{(j)})^T. \quad (55)$$

For the proof of part (b) of the lemma, the logarithm of the predictive likelihood is decomposed as follows [27]

$$\log p(z) = \mathcal{L}(q(x)q(\hat{x}^*)) + \text{KL}(q(x)q(\hat{x}^*)||p(x, \hat{x}^*|z)), \quad (56)$$

where

$$\mathcal{L}(q(x)q(\hat{x}^*)) = \int q(x)q(\hat{x}^*) \log \frac{p(x, \hat{x}^*, z)}{q(x)q(\hat{x}^*)} dx d\hat{x}^* \quad (57)$$

and

$$\text{KL}(q(x)q(\hat{x}^*)||p(x, \hat{x}^*|z)) = \int q(x)q(\hat{x}^*) \log \frac{q(x)q(\hat{x}^*)}{p(x, \hat{x}^*|z)} dx d\hat{x}^*. \quad (58)$$

Since the second term on the right hand side is minimized by the variational inference, an approximation of the predictive likelihood  $p(z)$  can be obtained from

$$p(z) \approx \exp(\mathcal{L}(q(x)q(\hat{x}^*))) \quad (59)$$

using the final estimates provided by the iterations given in Lemma 1. The first steps of the derivations are given in (61) which after standard but tedious derivations simplifies to

$$\begin{aligned} \hat{q}(z) = \exp & \left\{ -\frac{1}{2}n_z \log(\pi) + \frac{1}{2}n_x \log(2\pi) \right. \\ & + \log \mathcal{N}(m_{k+1|k+1}; m_{k+1|k}, P_{k+1|k}) - \frac{1}{2} \text{tr}(P_{k+1|k}^{-1}P_{k+1|k+1}) \\ & + \frac{1}{2} \log |P_{k+1|k+1}| + \frac{1}{2}n_x + \frac{1}{2}(\nu_{k+1|k} - n_z - 1) \log |\Psi_{k+1|k}| \\ & - \log \Gamma_{n_z} \left[ \frac{1}{2}(\nu_{k+1|k} - n_z - 1) \right] \\ & - \frac{1}{2}(\nu_{k+1|k+1} - n_z - 1) \log |\Psi_{k+1|k+1}| \\ & \left. + \log \Gamma_{n_z} \left[ \frac{1}{2}(\nu_{k+1|k+1} - n_z - 1) \right] \right\}. \end{aligned} \quad (60)$$

$$\begin{aligned} \mathcal{L}(\hat{q}(x)\hat{q}(\hat{x}^*)) = & \mathbb{E}_{\hat{q}(x)\hat{q}(\hat{x}^*)} \left[ -\frac{1}{2}n_z \log(2\pi) - \frac{1}{2} \log |\hat{x}^*| \right. \\ & + \text{tr} \left( -\frac{1}{2}(\hat{x}^*)^{-1}(z - Cx)(z - Cx)^T \right) \\ & - \frac{1}{2}n_x \log(2\pi) - \frac{1}{2} \log |P_{k+1|k}| \\ & + \text{tr} \left( -\frac{1}{2}P_{k+1|k}^{-1}(x - m_{k+1|k})(x - m_{k+1|k})^T \right) \\ & + \frac{1}{2}n_x \log(2\pi) + \frac{1}{2} \log |P_{k+1|k+1}| \\ & + \text{tr} \left( \frac{1}{2}P_{k+1|k+1}^{-1}(x - m_{k+1|k+1})(x - m_{k+1|k+1})^T \right) \\ & - \frac{1}{2}(\nu_{k+1|k} - n_z - 1)n_z \log 2 \\ & + \frac{1}{2}(\nu_{k+1|k} - n_z - 1) \log |\Psi_{k+1|k}| \\ & - \log \Gamma_{n_z} \left[ \frac{1}{2}(\nu_{k+1|k} - n_z - 1) \right] - \frac{1}{2}\nu_{k+1|k} \log |\hat{x}^*| \\ & + \text{tr} \left( -\frac{1}{2}(\hat{x}^*)^{-1}\Psi_{k+1|k} \right) \\ & + \frac{1}{2}(\nu_{k+1|k+1} - n_z - 1)n_z \log 2 \\ & - \frac{1}{2}(\nu_{k+1|k+1} - n_z - 1) \log |\Psi_{k+1|k+1}| \\ & + \log \Gamma_{n_z} \left[ \frac{1}{2}(\nu_{k+1|k+1} - n_z - 1) \right] + \frac{1}{2}\nu_{k+1|k+1} \log |\hat{x}^*| \\ & - \text{tr} \left( -\frac{1}{2}(\hat{x}^*)^{-1}\Psi_{k+1|k+1} \right) \quad (61) \\ = & -\frac{1}{2}n_z \log(2\pi) - \frac{1}{2} \mathbb{E}[\log |\hat{x}^*|] \\ & + \text{tr} \left( -\frac{1}{2} \mathbb{E}[(\hat{x}^*)^{-1}] \mathbb{E}[(z - Cx)(z - Cx)^T] \right) \\ & - \frac{1}{2}n_x \log(2\pi) - \frac{1}{2} \log |P_{k+1|k}| \\ & + \text{tr} \left( -\frac{1}{2}P_{k+1|k}^{-1} \mathbb{E}[(x - m_{k+1|k})(x - m_{k+1|k})^T] \right) \\ & + \frac{1}{2}n_x \log(2\pi) + \frac{1}{2} \log |P_{k+1|k+1}| \\ & + \text{tr} \left( \frac{1}{2}P_{k+1|k+1}^{-1} \mathbb{E}[(x - m_{k+1|k+1})(x - m_{k+1|k+1})^T] \right) \\ & - \frac{1}{2}(\nu_{k+1|k} - n_z - 1)n_z \log 2 \\ & + \frac{1}{2}(\nu_{k+1|k} - n_z - 1) \log |\Psi_{k+1|k}| \\ & - \log \Gamma_{n_z} \left[ \frac{1}{2}(\nu_{k+1|k} - n_z - 1) \right] - \frac{1}{2}\nu_{k+1|k} \mathbb{E} \log |\hat{x}^*| \\ & + \text{tr} \left( -\frac{1}{2} \mathbb{E}[(\hat{x}^*)^{-1}] \Psi_{k+1|k} \right) \\ & + \frac{1}{2}(\nu_{k+1|k+1} - n_z - 1)n_z \log 2 \\ & - \frac{1}{2}(\nu_{k+1|k+1} - n_z - 1) \log |\Psi_{k+1|k+1}| \\ & + \log \Gamma_{n_z} \left[ \frac{1}{2}(\nu_{k+1|k+1} - n_z - 1) \right] + \frac{1}{2}\nu_{k+1|k+1} \mathbb{E} \log |\hat{x}^*| \\ & - \text{tr} \left( -\frac{1}{2} \mathbb{E}[(\hat{x}^*)^{-1}] \Psi_{k+1|k+1} \right) \end{aligned}$$