



# Distributed adaptive Kalman filter based on variational Bayesian technique

Chen HU<sup>1†</sup>, Xiaoming HU<sup>2</sup>, Yiguang HONG<sup>3</sup>

1. Rocket Force University of Engineering, Xi'an Shannxi 710025, China;

2. Department of Mathematics, Royal Institute of Sweden (KTH), Sweden;

3. Institute of Systems Science and University of Chinese Academy of Sciences, Chinese Academy of Sciences, Beijing 100190, China

Received 31 August 2018; revised 23 October 2018; accepted 23 October 2018

## Abstract

In this paper, distributed Kalman filter design is studied for linear dynamics with unknown measurement noise variance, which modeled by Wishart distribution. To solve the problem in a multi-agent network, a distributed adaptive Kalman filter is proposed with the help of variational Bayesian, where the posterior distribution of joint state and noise variance is approximated by a free-form distribution. The convergence of the proposed algorithm is proved in two main steps: noise statistics is estimated, where each agent only use its local information in variational Bayesian expectation (VB-E) step, and state is estimated by a consensus algorithm in variational Bayesian maximum (VB-M) step. Finally, a distributed target tracking problem is investigated with simulations for illustration.

**Keywords:** Distributed Kalman filter, adaptive filter, multi-agent system, variational Bayesian

DOI <https://doi.org/10.1007/s11768-019-8183-9>

## 1 Introduction

In recent years, distributed estimation design of multi-agent systems has received more and more attention, because of its broad range of applications in engineering systems such as communication networks, sensor networks, and smart grids. Distributed Kalman filter (DKF) is one of the important distributed estimation methods, which plays a key role due to its power of

real-time estimation and non-stationary process tracking [1–7]. For example, R. Olfati-Saber [1] constructed DKF by putting the consensus terms in each Kalman filter update step, while S. Stanković et al. [2] considered the problem for missing observations and communication faults. Also, D. Casbeer et al. [3] and [4] applied consensus algorithms to an information filter, where the unbiasedness and convergence were shown, and W. Yang et al. [7] considered the problem of link acti-

<sup>†</sup>Corresponding author.

E-mail: [chenh628@hotmail.com](mailto:chenh628@hotmail.com). Tel.: +86-029-84741357.

This work was supported by the National Natural Science Foundation of China (Nos. 61733018, 61573344).

© 2019 South China University of Technology, Academy of Mathematics and Systems Science, CAS and Springer-Verlag GmbH Germany, part of Springer Nature

variation with power constraints. Moreover, G. Battistelli and L. Chisci [5] considered DKF with consensus on the inverse covariance matrix and the information vector, while Z. Zhou et al. [8] studied DKF for switching communication topologies. However, these works are based on a prior knowledge of measurement noise statistics.

For several decades, various centralized designs of adaptive filters have been proposed to solve the problem with unknown noise statistics, mainly based on the maximum likelihood or covariance matching methods [9–11], which are computationally complex or analytically intractable [12, 13]. On the other hand, variational Bayesian (VB) methods have been proposed for approximating posterior distribution, which have benefits of low computational cost and analytic tractability [14]. Therefore, S. Sarkka and A. Nummenmaa [13] used a VB method to approximate the joint posterior distribution of the dynamic state and measurement noise parameters in linear dynamics, where the variational distribution of noise is assumed as the product of the Inverse-Gamma distributions. Moreover, H. Zhu et al. [15] employed VB to estimate the joint distribution of state and non-Gaussian noises statistical, where the non-Gaussian noise was modeled by Gaussian mixture models. Additionally, P. Dong et al. [16] considered an adaptive VB filter based on the Wishart distribution, to estimate the full noise variance matrix with a fusion center. However, none of these algorithms is applicable to distributed design of multi-agent networks. To the best of our knowledge, the distributed adaptive VB Kalman filters have not yet been adequately investigated, and very few studies on adaptive design for distributed state estimation with unknown measurement noise variance, though many effective distributed Kalman filters have been developed.

In this paper, we investigate the distributed adaptive VB Kalman filter with unknown noise variance modeled by the Wishart distribution in a multi-agent network. With communication with neighbors, the posterior distribution of joint state and noise variance is approximated by a free-form distribution, where the approximation can be achieved by recursively performing VB-E and VB-M steps [14]. The main contribution of the paper can be summarized as follows. A distributed adaptive VB Kalman filter is proposed, which can be viewed as a distributed extension of the centralized one given in the linear case [16]. The analysis shows that VB-E step can be achieved only with the local information and the VB-M step can be achieved with a consensus algorithm, where the VB-M structure is consistent with some distributed

Kalman filters in [3, 4], and then the convergence of the proposed algorithm is proved.

The remainder of the paper is organized as follows. Section 2 presents problem formulation of this paper. Then Section 3 gives the design and analysis of the distributed adaptive VB Kalman filter, and Section 4 shows simulation examples. Finally, Section 5 provides concluding remarks.

**Notations**  $\mathcal{N}(x; \hat{x}, P)$  denotes the Gaussian distribution with mean  $\hat{x}$  and covariance  $P$ .  $\text{diag}\{\cdot\}$  represent the diagonalizations of block or scalar elements. An undirected graph of  $N$  nodes is denoted as  $\mathcal{G} = \{\mathcal{V}, \mathcal{E}\}$  with  $\mathcal{V} = 1, \dots, N$ . The neighbors of node  $i$  can be represented by the set  $\{j \in \mathcal{V} | (i, j) \in \mathcal{E}\} \triangleq \mathcal{N}_i$ , whose size is denoted as  $|\mathcal{N}_i|$ .

## 2 Problem formulation

Consider the following stochastic dynamics:

$$x_{k+1} = A_k x_k + B_k w_k, \quad (1)$$

with  $x_k \in \mathbb{R}^n$  as the state observed by a  $N$ -agent network  $\mathcal{G}$ ,  $A_k \in \mathbb{R}^{n \times n}$  as the system matrix, and  $w_k \in \mathbb{R}^n$  as the zero-mean Gaussian noise, with  $\mathbb{E}\{w_k w_k^T\} = Q_k$ . The measurement model is given as

$$y_{i,k} = C_{i,k} x_k + D_{i,k} v_{i,k}, \quad i = 1, \dots, N, \quad (2)$$

with  $y_{i,k} \in \mathbb{R}^{m_i}$  as the measurement of agent  $i$ ,  $C_{i,k}$  as the observation matrix of agent  $i$ , and  $v_{i,k} \sim \mathcal{N}(v_{i,k}; 0, R_{i,k})$  as the observation noise. Suppose that  $\mathbb{E}\{v_{i,k} v_{j,k}^T\} = 0, \forall i \neq j$  and  $\mathbb{E}\{v_{i,k} w_k^T\} = 0, \forall i \in \mathcal{V}$  at each  $k$ , and all  $y_{i,k}, \forall i \in \mathcal{V}$  are conditionally independent of given  $x_k$ .

In the Bayesian filter framework, the dynamics (1) and measurement mapping (2) can be written as state transition probability  $p(x_{k+1}|x_k)$  and measurement likelihood function  $p(y_{i,k}|x_k), i = 1, \dots, N$ . Hence, the joint likelihood function is the product of local likelihood functions,

$$p(Y_k|x_k) = \prod_{i=1}^N p(y_{i,k}|x_k), \quad (3)$$

where  $Y_k = [y_{1,k}^T, \dots, y_{N,k}^T]^T$ .

The following assumptions have been widely used in the DKF literature [1, 3–6].

**Assumption 1** The undirected graph  $\mathcal{G}$  is connected.

**Assumption 2** (Collective observability) The pair  $(A_k, C_k)$  is observable, where  $C_k = [C_{1,k}^T, \dots, C_{N,k}^T]^T$ .

Since  $y_{i,k}$  needs to be purely nondeterministic and  $D_{i,k}R_{i,k}D_{i,k}^T > 0$  is a sufficient condition for this, we assume  $D_{i,k} = I, \forall i$  without loss of generality.

Most Kalman filter designs are based on the assumption that  $R_{i,k}$  is known. However, this assumption may not hold in practice, and here we consider the case when  $R_{i,k}$  is unknown. To be strict, we assume that the measurement information matrix  $\Phi_{i,k} = R_{i,k}^{-1}$  follows the Wishart distribution  $\mathcal{W}(\Phi_{i,k}; v_{i,k}, V_{i,k})$  with parameters  $v_{i,k}$  and  $V_{i,k}$ . Note that the Wishart distribution is a conjugate prior for the unknown information matrix of a Gaussian model with a known mean, which is suitable to model the unknown noises [17].

Distributed Kalman filters [1–6] were proposed in (1) and (2) with known  $R_{i,k}$ . However, inaccurate  $R_{i,k}$  may enlarge the estimation error or even causing divergence. Therefore, we aim at the design of a distributed Kalman filter with unknown measurement noise to estimate the joint posterior distribution  $p(x_k, \Phi_k | Y_k)$  by a network of  $N$  agents, where  $\Phi_k = \text{diag}\{\Phi_{1,k}, \dots, \Phi_{N,k}\}$ .

**Remark 1** For comparison, let us define a centralized estimate  $\hat{x}_k^c$  associated with measurement  $Y_k$ . Then the centralized estimation consists of the prediction  $\bar{x}_k^c = E\{x_k | Y_{k-1}\}$  and the update  $\hat{x}_k^c = E\{x_k | Y_k\}$ . For linear dynamics (1) and measurement equation (2) the well-known Kalman filter gives a recursive solution as follows [18]:

Measurement update:

$$\begin{aligned} P_k^c &= \bar{P}_k^c - \bar{P}_k^c C_k^T (C_k \bar{P}_k^c C_k^T + R_k)^{-1} C_k \bar{P}_k^c \\ \hat{x}_k^c &= \bar{x}_k^c + P_k^c C_k^T R_k^{-1} (Y_k - C_k \bar{x}_k^c). \end{aligned}$$

Prediction:

$$\begin{aligned} \bar{x}_{k+1}^c &= A_k \hat{x}_k^c \\ \bar{P}_{k+1}^c &= A_k P_k^c A_k^T + B_k Q_k B_k^T, \end{aligned}$$

where  $R_k = \text{diag}\{R_{1,k}, \dots, R_{N,k}\}$ , and  $P_k^c$  and  $\bar{P}_k^c$  are estimate error covariance matrices of  $\hat{x}_k^c$  and  $\bar{x}_k^c$ , respectively.

Suppose that the posterior distribution of  $x_{k-1}$  and  $\Phi_{k-1}$  conditional on  $Y_{k-1}$  at time  $k - 1$  is approximated by

$$\begin{aligned} &p(x_{k-1}, \Phi_{k-1} | Y_{k-1}) \\ &= p(x_{k-1} | Y_{k-1}) p(\Phi_{k-1} | Y_{k-1}) \\ &= \mathcal{N}(x_{k-1}; \hat{x}_{k-1}, P_{k-1}) \mathcal{W}(\Phi_{k-1}; v_{k-1}, V_{k-1}). \end{aligned} \quad (4)$$

The first equation holds since  $x_{k-1}$  and  $\Phi_{k-1}$  are conditionally independent given  $Y_{k-1}$ . Note that the form of posterior distribution  $p(x_k, \Phi_k | Y_k)$  remains the same if the predictive distribution  $p(x_k, \Phi_k | Y_{k-1})$  is Gaussian-Wishart [19].

The predictive distribution of state  $x_k$  and  $\Phi_k$  are separable and independent, and can be given by the Chapman-Kolmogorov equation,

$$\begin{aligned} p(x_k | Y_{k-1}) &= \int p(x_k | x_{k-1}) p(x_{k-1} | Y_{k-1}) dx_{k-1} \\ &= \mathcal{N}(x_k; \bar{x}_k, \bar{P}_k), \end{aligned} \quad (5)$$

$$\begin{aligned} p(\Phi_k | Y_{k-1}) &= \int p(\Phi_k | \Phi_{k-1}) p(\Phi_{k-1} | Y_{k-1}) d\Phi_{k-1} \\ &= \mathcal{W}(\Phi_k; \bar{v}_k, \bar{V}_k). \end{aligned} \quad (6)$$

Prediction (5) for (1) results in a Gaussian distribution for predicted  $x_k$ , with parameters  $\bar{x}_k = A_k \hat{x}_{k-1}$  and  $\bar{P}_k = A_k P_{k-1} A_k^T + B_k Q_k B_k^T$  given by the standard Kalman filter prediction equations.

Similar to [13, 16], we take the dynamics of  $\Phi_{i,k}$  to ensure that the predictive distribution  $p(\Phi_k | Y_{k-1})$  is Wishart, with  $v_k = \mu v_{k-1}$  and  $V_k = \frac{1}{\mu} V_{k-1}$ . Clearly, the value  $\mu = 1$  indicates a stationary variance, while  $0 < \mu < 1$  indicates that the variance is time-varying [13].

Because  $x_k$  and  $\Phi_k$  are independent, the prediction for the joint of  $x_k$  and  $\Phi_k$  can be written as

$$\begin{aligned} p(x_k, \Phi_k | Y_{k-1}) &= p(x_k | Y_{k-1}) p(\Phi_k | Y_{k-1}) \\ &= \mathcal{N}(x_k | \bar{x}_k, \bar{P}_k) \mathcal{W}(\Phi_k | \bar{v}_k, \bar{V}_k) \\ &= \mathcal{N}(x_k | \bar{x}_k, \bar{P}_k) \prod_{i=1}^N \mathcal{W}(\Phi_i | \bar{v}_{i,k}, \bar{V}_{i,k}), \end{aligned} \quad (7)$$

where  $\bar{v}_{i,k} = \mu v_{i,k-1}$  and  $\bar{V}_{i,k} = \frac{1}{\mu} V_{i,k-1}$ . The third equation holds since the agents are independent with each other.

Based on the standard VB method [14, 20, 21], the joint posterior distribution  $p(x_k, \Phi_k | Y_k)$  can be approximated by a free-form distribution  $q(x_k, \Phi_k)$ ,

$$p(x_k, \Phi_k | Y_k) \approx q(x_k, \Phi_k) = q(x_k) \prod_{i=1}^N q(\Phi_{i,k}), \quad (8)$$

where  $q(x_k)$  and  $q(\Phi_{i,k})$  are the Gaussian and Wishart distributions, respectively. The VB approximation is achieved by minimizing the Kullback-Leibler (KL) divergence between the posterior distribution  $p(x_k, \Phi_k | Y_k)$

and the variational distribution  $q(x_k, \Phi_k)$ ,

$$\begin{aligned} & \text{KL}(q\|p) \\ &= \int q(x_k)q(\Phi_k) \log \frac{q(x_k)q(\Phi_k)}{p(x_k, \Phi_k|Y_k)} dx_k d\Phi_k \\ &= E_q [\log(q(x_k)q(\Phi_k))] - E_q [\log(p(x_k, \Phi_k|Y_k))] + \tilde{c} \\ &= -\mathcal{L}(q) + \tilde{c}, \end{aligned}$$

where  $\tilde{c}$  denotes the terms independent of the variational distribution  $q(x_k, \Phi_k)$ , and

$$\begin{aligned} & \mathcal{L}(q) \\ &= E_q [\log(p(x_k, \Phi_k|Y_k))] - E_q [\log(q(x_k)q(\Phi_k))] \end{aligned} \quad (9)$$

is called the evidence of lower bound (ELBO).

Therefore, our problem becomes

$$\max_{q(x), q(\Phi)} \mathcal{L}(q(x)q(\Phi)), \quad (10)$$

which is equivalent to the minimization of  $\text{KL}(q\|p)$ .

The VB approach updates each local variational distribution  $q(\Phi_{i,k})$  (VB-E step) and the global variational distribution  $q(x_k)$  (VB-M step) alternatively, with keeping the other distributions fixed:

$$\text{VB-E: } \hat{\Phi}_{i,k} = \arg \max_{\Phi_{i,k}} \mathcal{L}(x_k^*, \Phi_{i,k}), \quad \forall i \in \mathcal{V}, \quad (11)$$

$$\text{VB-M: } \hat{x}_k = \arg \max_{x_k} \mathcal{L}(x_k, \Phi_k^*). \quad (12)$$

Note that the estimate  $\hat{x}_k$  in (12) can be obtained by each agent if each agent knows the global likelihood function  $p(y_k|x_k)$  [13, 15, 16]. However, in our paper, each agent can only access to its local likelihood function  $p(y_{i,k}|x_k)$ , and thus, we have to propose a distributed adaptive VB Kalman filter to optimize  $\text{KL}(q\|p)$ .

### 3 Distributed adaptive VB Kalman filter

In this section, we discuss the VB approximation of the posterior distribution  $p(x_k, \Phi_k|Y_k)$  in a distributed way, and propose a distributed adaptive VB Kalman filter algorithm (DAVBKF).

Define a set of variables  $\{x_{i,k}, i = 1, \dots, N\}$  for each agent, and let  $s$  be the number of VB iterations. Then the approximation distributions  $q^{(s)}(\Phi_{i,k}; \hat{v}_{i,k}^{(s)}, \hat{V}_{i,k}^{(s)})$  and  $q^{(s)}(x_k; \hat{x}_k^{(s)}, P_k^{(s)})$  at the  $s$ th step can be given as follows:

$$\text{VB-E step: } \hat{v}_{i,k} = \bar{v}_{i,k} + N, \quad (13)$$

$$\begin{aligned} (\hat{V}_{i,k}^{(s)})^{-1} &= N[C_{i,k}P_k^{(s-1)}C_{i,k}^T \\ &+ (y_{i,k} - C_{i,k}\hat{x}_k^{(s-1)})(\cdot)^T] + \bar{V}_{i,k}^{-1}, \end{aligned} \quad (14)$$

$$\text{VB-M step: } (P_k^{(s)})^{-1} = \frac{1}{N} \sum_{i=1}^N (\check{P}_{i,k}^{(s)})^{-1}, \quad (15)$$

$$(P_k^{(s)})^{-1}\hat{x}_k^{(s)} = \frac{1}{N} \sum_{i=1}^N (\check{P}_{i,k}^{(s)})^{-1}\check{x}_{i,k}^{(s)}, \quad (16)$$

where  $\check{x}_{i,k}^{(s)}$  and  $\check{P}_{i,k}^{(s)}$  can be obtained at each agent locally as follows:

$$(\check{P}_{i,k}^{(s)})^{-1} = \bar{P}_{i,k}^{-1} + NC_{i,k}^T \check{\Phi}_{i,k}^{(s)} C_{i,k}, \quad (17)$$

$$(\check{P}_{i,k}^{(s)})^{-1}\check{x}_{i,k}^{(s)} = \bar{P}_{i,k}^{-1}\bar{x}_{i,k} + NC_{i,k}^T \check{\Phi}_{i,k}^{(s)} y_{i,k}, \quad (18)$$

and  $\check{\Phi}_{i,k}^{(s)} = E[\Phi_{i,k}^{(s)}] = \hat{v}_{i,k}^{(s)} \hat{V}_{i,k}^{(s)}$ .

The proposed VB algorithm updates  $q(\Phi_{i,k}), \forall i \in \mathcal{V}$  and  $q(x_k)$  alternatively while keeping the other fixed, to be discussed in Lemmas 1 and 2.

In (13) and (14), each agent only needs the local information in VB-E step. Nevertheless, in VB-M step, a global estimate is achieved by averaging local information of all agents. To compute  $(P_k^{(s)})^{-1}$  and  $(P_k^{(s)})^{-1}\hat{x}_k^{(s)}$  in a distributed way, denote

$$\omega_{i,k}^{(s)} = \bar{P}_{i,k}^{-1}\bar{x}_{i,k} + NC_{i,k}^T \check{\Phi}_{i,k}^{(s)} y_{i,k},$$

$$\Omega_{i,k}^{(s)} = \bar{P}_{i,k}^{-1} + NC_{i,k}^T \check{\Phi}_{i,k}^{(s)} C_{i,k}.$$

Based on the average consensus algorithm [22], each agent approximates  $(P_k^{(s)})^{-1}$  and  $(P_k^{(s)})^{-1}\hat{x}_k^{(s)}$  as follows:

$$\Omega_{i,k}^{(s,l)} = \Omega_{i,k}^{(s,l-1)} + \epsilon \sum_{j \in \mathcal{N}_i} (\Omega_{j,k}^{(s,l-1)} - \Omega_{i,k}^{(s,l-1)}), \quad l = 1, 2, \dots, \quad (19)$$

$$\omega_{i,k}^{(s,l)} = \omega_{i,k}^{(s,l-1)} + \epsilon \sum_{j \in \mathcal{N}_i} (\omega_{j,k}^{(s,l-1)} - \omega_{i,k}^{(s,l-1)}), \quad l = 1, 2, \dots, \quad (20)$$

where  $\epsilon$  is a consensus rate parameter within the range of  $(0, \frac{1}{\Delta})$  with  $\Delta = \max_i |\mathcal{N}_i|$  [22]. Then we give our DAVBKF in Table 1, with initial condition  $\bar{x}_{i,0} = E\{x_0\}$  and  $\bar{P}_{i,0} = P_0$ .

In the DAKFVB, the initialization of all local estimates is exactly the same mean of the initial state, which is not easy to satisfy in practice. In fact, a practical initialization can be obtained by letting  $\hat{x}_{i,0} = 0, \forall i \in \mathcal{V}$  and  $P_{i,0} = 0, \forall i \in \mathcal{V}$ , which means that no prior information available.

Table 1 DAVBKF at node  $i$  at time  $k$ .

**Ensure:**

At time  $k$ , a prior information  $\hat{x}_{i,k}^{(0)} = \bar{x}_{i,k}, P_{i,k}^{(0)} = \bar{P}_{i,k}$ ,

$\hat{v}_{i,k} = \bar{v}_{i,k} + N$ , and  $\bar{V}_{i,k}$ ;

**Measurement update:**

**For**  $s = 1$  to  $S$ :

**(VB-E)** Compute the variational parameter  $\hat{V}_{i,k}^{(s)}$ ,

$$(\hat{V}_{i,k}^{(s)})^{-1} = N[C_{i,k}P_k^{(s-1)}C_{i,k}^T + (y_{i,k} - C_{i,k}\hat{x}_k^{(s-1)})(\cdot)^T] + \bar{V}_{i,k}^{-1};$$

**(VB-M)** Perform the average consensus on  $\omega_{i,k}^{(s,0)}$  and  $\Omega_{i,k}^{(s,0)}$ ,

**For**  $l = 1$  to  $\bar{l}$ ,

$$\Omega_{i,k}^{(s,l)} = \Omega_{i,k}^{(s,l-1)} + \epsilon \sum_{j \in \mathcal{N}_i} (\Omega_{j,k}^{(s,l-1)} - \Omega_{i,k}^{(s,l-1)}),$$

$$\omega_{i,k}^{(s,l)} = \omega_{i,k}^{(s,l-1)} + \epsilon \sum_{j \in \mathcal{N}_i} (\omega_{j,k}^{(s,l-1)} - \omega_{i,k}^{(s,l-1)});$$

**End For**

$$P_{i,k}^{(s)} = (\Omega_{i,k}^{(s,\bar{l})})^{-1}, \hat{x}_{i,k}^{(s)} = P_{i,k}^{(s)}\omega_{i,k}^{(s,\bar{l})};$$

**End For**

Compute the estimate  $\hat{x}_{i,k}$  and covariance matrix  $P_{i,k}$ ,

$$P_{i,k} = (\Omega_{i,k}^{(S,\bar{l})})^{-1}, \hat{x}_{i,k} = P_{i,k}\omega_{i,k}^{(S,\bar{l})};$$

**Prediction:**

$$\bar{v}_{i,k+1} = \mu\hat{v}_{i,k}, \bar{V}_{i,k+1} = \frac{1}{\mu}\hat{V}_{i,k}^{(S)}$$

$$\bar{P}_{i,k+1} = A_kP_{i,k}^{-1}A_k^T + B_kQ_kB_k^T, \bar{x}_{i,k+1} = A_k\hat{x}_{i,k}.$$

**Remark 2** Note that, in the DAVBKF, the distributed realization of VB approach for  $q(x_{i,k})$  with fixed  $q(\Phi_{i,k})$  is consistent with distributed Kalman filters in [3, 4]. Namely, equations (19) and (20) perform measurement updating step of distributed Kalman filters in [3, 4] for given noise variance matrices.

Next, we give two lemmas about  $q(x_k)$  and  $q(\Phi_{i,k}), i = 1, \dots, N$ . The first lemma is for the solution of the variational distribution  $q(\Phi_{i,k})$ .

**Lemma 1** (VB-E) Given the global variational distribution  $q(x_k; \hat{x}_k, P_k)$ ,  $q(\Phi_{i,k}; \hat{v}_{i,k}, \hat{V}_{i,k})$  of each agent is obtained as follows:

$$\hat{v}_{i,k} = \bar{v}_{i,k} + N, \tag{21}$$

$$\hat{V}_{i,k}^{-1} = N[C_{i,k}P_kC_{i,k}^T + (y_{i,k} - C_{i,k}\hat{x}_k)(\cdot)^T] + \bar{V}_{i,k}^{-1}. \tag{22}$$

Its proof is in the appendix.

**Remark 3** S. Sarkka et al. [13, 16] presented adaptive VB Kalman filters. Note that [13] assumed that the measurement noise variance  $R_k$  was a diagonal matrix with a priori distribution of noise modeled as a product of Inverse-Gamma distribution, i.e.,  $R_k = \text{diag}\{\sigma_{k,1}, \dots, \sigma_{k,m}\}$  and  $p(R_k) = \prod_{t=1}^m \text{IG}(\sigma_{k,t}; \alpha_{k,t}, \beta_{k,t})$ , where  $\text{IG}(\sigma; \alpha, \beta)$  stands for the Inverse-Gamma dis-

tribution covered by the parameters  $\alpha$  and  $\beta$ . P. Dong et al. [16] presented an adaptive VB Kalman filter for nonlinear dynamics, with the noise variance matrix modeled as a Wishart distribution, which resulted in full noise variance matrix estimation. Clearly, VB-E step in Lemma 1 is different from the centralized one in [16], even though (21) and (22) can be obtained by each agent individually. With  $N = 1$  in Lemma 1, (21) and (22) will reduce to the VB-E step in [16].

According to the VB approach, an optimal solution of  $q(x_k; \hat{x}_k, P_k)$  can be achieved by giving  $q(\Phi_{i,k}; \hat{v}_{i,k}, \hat{V}_{i,k}), i = 1, \dots, N$ . However, in a distributed structure, it cannot be solved straightforwardly because each agent cannot access the joint likelihood function  $p(y_k|x_k) = \prod_{i=1}^N p(y_{i,k}|x_k)$  and  $q(\Phi_{i,k}; \hat{v}_{i,k}, \hat{V}_{i,k})$  of non-neighboring agent.

Since the agents are independent, the posterior can be written as  $p(x_k, \Phi_k|Y_k) = p(\Phi_k, x_k) \prod_{i=1}^N p(y_{i,k}|\Phi_{i,k}, x_k)$ . Decomposition of  $\mathcal{L}(q)$  can be obtained by substituting  $p(x_k, \Phi_k|Y_k)$  into (9). Namely,

$$\begin{aligned} \mathcal{L}(q) &= E_q[\log p(\Phi_k, x_k) \prod_{i=1}^N p(y_{i,k}|\Phi_{i,k}, x_k)] \\ &\quad - E_q[\log(q(x_k)q(\Phi_k))] \\ &= E_q[\sum_{i=1}^N \log(p(y_{i,k}|\Phi_{i,k}, x_k))] + E_q[\log \frac{p(\Phi_k, x_k)}{q(x_k)q(\Phi_k)}] \\ &= \sum_{i=1}^N \mathcal{L}_i(q), \end{aligned} \tag{23}$$

where

$$\mathcal{L}_i(q) = E_q[\log(p(y_{i,k}|\Phi_{i,k}, x_k))] + \frac{1}{N} E_q[\log \frac{p(\Phi_k, x_k)}{q(x_k)q(\Phi_k)}].$$

Denote  $q^*(x_k; \hat{x}_k, P_k)$  as an optimal variational distribution for  $\mathcal{L}(q)$ , and  $q_i^*(x_{i,k}, \hat{x}_{i,k}, \hat{P}_{i,k})$  as an optimal variational distribution for  $\mathcal{L}_i(q)$ . The next lemma gives the solution for global variational distribution  $q^*(x_k)$ , and the relationship with  $q_i^*(x_{i,k})$ .

**Lemma 2** (VB-M) Suppose that each agent gets the global optimal solution  $q(x_{k-1}; \hat{x}_{k-1}, P_{k-1})$  at time  $k - 1$ . Given  $q(\Phi_{i,k}; \hat{v}_{i,k}, \hat{V}_{i,k}), i = 1, \dots, N$ ,  $q(x_k; \hat{x}_k, P_k)$  can be achieved as follows:

$$(P_k)^{-1} = \frac{1}{N} \sum_{i=1}^N (\hat{P}_{i,k})^{-1}, \tag{25}$$



$$(P_k)^{-1}\hat{x}_k = \frac{1}{N} \sum_{i=1}^N (\check{P}_{i,k})^{-1}\check{x}_{i,k}, \tag{26}$$

where  $\check{x}_{i,k}$  and  $\check{P}_{i,k}$  can be obtained locally as follows:

$$(\check{P}_{i,k})^{-1} = \bar{P}_{i,k}^{-1} + NC_{i,k}^T \check{\Phi}_{i,k} C_{i,k}, \tag{27}$$

$$(\check{P}_{i,k})^{-1}\check{x}_{i,k} = \bar{P}_{i,k}^{-1}\bar{x}_{i,k} + NC_{i,k}^T \check{\Phi}_{i,k} y_{i,k}, \tag{28}$$

and  $\check{\Phi}_{i,k} = E[\Phi_{i,k}] = \hat{v}_{i,k} \hat{V}_{i,k}$ .

Its proof is in the appendix.

Then it is time to give our result for the proposed algorithm for an  $N$ -agent network.

**Theorem 1** Under Assumptions 1 and 2 for the proposed DAVBKF, if each agent is initialized by letting  $\bar{x}_{i,0} = E\{x_0\}$  and  $\bar{P}_{i,0} = \bar{P}_0, \forall i \in \mathcal{V}$ , then the estimates  $(\hat{x}_{i,k}^{(s,l)}, P_{i,k}^{(s,l)})$  converge to an estimate of centralized solution of (9) as  $\bar{l} \rightarrow \infty$  for all  $k$ .

**Proof** Suppose that, at time  $k - 1$ , the estimates  $q(x_{i,k-1}), \forall i \in \mathcal{V}$  of each agent achieve consensus, i.e.,  $\hat{x}_{i,k-1} = \hat{x}_{j,k-1} = \hat{x}_k$  and  $P_{i,k-1} = P_{j,k-1} = P_{k-1}, \forall i, j \in \mathcal{V}$ . In the prediction of the proposed algorithm, we have  $\bar{x}_{i,k} = A\hat{x}_{i,k-1}$  and  $\bar{P}_{i,k} = A_k P_{i,k-1} A_k^T + B_k Q_k B_k^T$ . Therefore,  $\bar{x}_{i,k} = \bar{x}_{j,k} = \bar{x}_k$  and  $\bar{P}_{i,k} = \bar{P}_{j,k} = \bar{P}_k, \forall i, j \in \mathcal{V}$ .

The VB approach for measurement update can be achieved by iteratively updating the VB-E and the VB-M steps in DAVBKF. Next, we show the optimality and stability of VB-M step in DAVBKF. At the  $s$ th iteration of VB, under Assumption 1, the consensus iterations (19) and (20) converge to the average of  $(\check{P}_{i,k})^{-1}$  and  $(\check{P}_{i,k})^{-1}\check{x}_{i,k}$  as  $\bar{l} \rightarrow \infty$  [22], respectively. Therefore,

$$\begin{aligned} (P_{i,k}^{(s)})^{-1} &= \lim_{l \rightarrow \infty} \Omega_{i,k}^{(l)} \\ &= \frac{1}{N} \sum_{i=1}^N (\bar{P}_{i,k})^{-1} + NC_{i,k}^T \check{\Phi}_{i,k}^{(s)} C_{i,k} \\ &= \bar{P}_k^{-1} + \sum_{i=1}^N C_{i,k}^T \check{\Phi}_{i,k}^{(s)} C_{i,k} \\ &= \bar{P}_k^{-1} + C_k^T \check{\Phi}_k^{(s)} C_k, \end{aligned} \tag{29}$$

and

$$\begin{aligned} (P_{i,k}^{(s)})^{-1}\hat{x}_{i,k}^{(s)} &= \lim_{l \rightarrow \infty} \omega_{i,k}^{(l)} \\ &= \frac{1}{N} \sum_{i=1}^N (\bar{P}_{i,k})^{-1}\bar{x}_{i,k} + NC_{i,k}^T \check{\Phi}_{i,k}^{(s)} y_{i,k} \\ &= \bar{P}_k^{-1}\bar{x}_k + \sum_{i=1}^N C_{i,k}^T \check{\Phi}_{i,k}^{(s)} y_{i,k} \\ &= \bar{P}_k^{-1}\bar{x}_k + C_k^T \check{\Phi}_k^{(s)} Y_k, \end{aligned} \tag{30}$$

where  $C_k = [C_{1,k}^T, \dots, C_{N,k}^T]^T, \check{\Phi}_k^{(s)} = [\check{\Phi}_{1,k}^{(s)\top}, \dots, \check{\Phi}_{N,k}^{(s)\top}]^T$ , and  $Y_k = [y_{1,k}, \dots, y_{N,k}]^T$ . Using the matrix inverse lemma to (29), we have

$$P_{i,k}^{(s)} = \bar{P}_k - \bar{P}_k C_k^T (C_k \bar{P}_k C_k^T + (\check{\Phi}_k^{(s)})^{-1})^{-1} C_k \bar{P}_k. \tag{31}$$

Multiplying both sides of (30) by  $P_{i,k}^{(s)}$  gives

$$\begin{aligned} \hat{x}_{i,k}^{(s)} &= \bar{x}_k - \bar{P}_k C_k^T (C_k \bar{P}_k C_k^T + (\check{\Phi}_k^{(s)})^{-1})^{-1} C_k \bar{x}_k \\ &\quad + P_{i,k}^{(s)} C_k^T \check{\Phi}_k^{(s)} Y_k. \end{aligned} \tag{32}$$

Similarly, we have

$$P_{i,k}^{(s)} C_k^T \check{\Phi}_k^{(s)} = \bar{P}_k C_k^T (C_k \bar{P}_k C_k^T + (\check{\Phi}_k^{(s)})^{-1})^{-1}. \tag{33}$$

Substituting (33) into (32), we obtain

$$\hat{x}_{i,k}^{(s)} = \bar{x}_k + \bar{P}_k C_k^T (C_k \bar{P}_k C_k^T + (\check{\Phi}_k^{(s)})^{-1})^{-1} (Y_k - C_k \bar{x}_k). \tag{34}$$

Note that (31) and (34) exactly are the measurement update equations of the standard Kalman filter with measurement noise variance  $(\check{\Phi}_k^{(s)})^{-1}$ . By Assumption 2, the optimality and stability of  $\hat{x}_{i,k}^{(s)}$  and  $P_{i,k}^{(s)}$  with given  $\check{\Phi}_k^{(s)}$  are guaranteed. Clearly,  $\hat{x}_{i,k} = \hat{x}_{j,k}$  and  $P_{i,k} = P_{j,k}, \forall i, j \in \mathcal{V}$  as  $\bar{l} \rightarrow \infty$  after the VB procedure, which satisfies the conditions of Lemma 2.

According to Theorem 2.1 [14], Lemma 1 and Lemma 2, proposed DAVBKF converges to a local maximum of  $\mathcal{L}(q)$  at time  $k$ . This convergence further guarantees that the local optimal centralized estimate will be reached at next step. Thus, DAVBKF with  $\bar{l} \rightarrow \infty$  at each time step can converge to an estimate of centralized solution with initial conditions.  $\square$

Theorem 1 shows that the global variational distribution  $q(x_k)$  can be achieved by averaging of a local optimal solution, and the updating of variational distributions  $q(\Phi_{i,k})$  only need local information of agent  $i$ . However, it is not easy to obtain the bound of estimate error, since the VB iteration can only converge to a local optimal solution, and it depend on the estimation at each instant.

**Remark 4** The VB approach guarantees the local convergence by iterations of the VB-E and VB-M steps [14]. The measurement update in DAVBKF consists of VB-E and VB-M steps, which leads to a fixed point iteration [23], while the prediction in DAVBKF is the same as the prediction step of standard Kalman filter. Moreover, with the collective observability assumption, the

dynamics (1) need not be observable by each agent in our analysis.

**Remark 5** In Theorem 1, for every  $k$ ,  $q(x_{i,k})$  of each agent should reach a consensus, as  $\bar{l} \rightarrow \infty$ . In practice, when  $0 < \bar{l} < \infty$ , it has been shown that  $P_{i,k}, \forall i \in \mathcal{V}, k = 1, 2, \dots$  is upper bounded by a positive definite matrix  $\hat{P}_{i,k}$  for given noise statistics in [3]. Recall Theorem 2.1 in [14], the iterations of VB-E (Lemma 1) and VB-M (Lemma 2) steps converge to a local maximum of  $\mathcal{L}(q)$ , which implies that both of  $P_{i,k}$  and  $\tilde{\Phi}_{i,k}$  are bounded. However, in this case,  $P_{i,k}$  will not represent the error covariance matrices. The simulation part illustrates that DAVBKF can converge fast even if  $\bar{l} = 1$ .

The centralized solution of VB procedure for measurement updating step can be easily obtained by letting  $N = 1$  in Theorem 1, which is mentioned in the following result.

**Corollary 1** Let  $s$  be the iterative number of VB. Assume there exists a center who can access  $C_k$  and  $y_k$ , then VB procedure for optimizing  $\mathcal{L}(q)$  in (9) can be achieved by the following updates:

**VB-E step:**  $\hat{v}_k = \bar{v}_k + 1, \tag{35}$

$$(\hat{V}_k^{(s)})^{-1} = [C_k P_k^{(s-1)} C_k^T + (y_k - C_k \hat{x}_k^{(s-1)}) (\cdot)^T + \bar{V}_k^{-1}], \tag{36}$$

**VB-M step:**  $(P_k^{(s)})^{-1} = \bar{P}_k^{-1} + C_k^T \tilde{\Phi}_k^{(s)} C_k, \tag{37}$

$$(P_k^{(s)})^{-1} \hat{x}_k^{(s)} = \bar{P}_k^{-1} \bar{x}_k + C_k^T \tilde{\Phi}_k^{(s)} y_k, \tag{38}$$

where  $\tilde{\Phi}_k^{(s)} = E[\Phi_k^{(s)}] = \hat{v}_k^{(s)} \hat{V}_k^{(s)}$ .

**Remark 6** Corollary 1 is consistent with the result in [16]. Also, the VB-M step in Corollary 1 is exactly the measurement updating step of the information form of the Kalman filter [18] with a given measurement noise variance.

### 4 Simulation

Theorem 1 considered the DAVBKF, with  $\bar{x}_{i,0} = E[x_0], \forall i \in \mathcal{V}$  and  $\bar{P}_{i,0} = P_0, \forall i \in \mathcal{V}$ . The VB algorithm alternatively updates each local variational distribution  $q(\Phi_{i,k})$  and the global variational distribution  $q(x_k)$ . In many simulation examples, the algorithm converges very fast, and only 2 or 3 iterations are good enough in practice.

Consider a target tracking problem, where the target

moves with a constant velocity, whose dynamics [24] is

$$x_{k+1} = \begin{bmatrix} 1 & 0 & T & 0 \\ 0 & 1 & 0 & T \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} x_k + G w_k, \tag{39}$$

where  $x_k = [d_x, d_y, \dot{d}_x, \dot{d}_y]^T$ ,  $d_x$  and  $d_y$  describe the target position, and  $\dot{d}_x$  and  $\dot{d}_y$  describe the target velocity.  $w_k \sim \mathcal{N}(0, Q_k)$ ,  $Q_k = \text{diag}\{w_1^2, w_2^2\}$  with  $w_1^2 = w_2^2 = 0.1$ .  $T = 1s$  is the sampling period, and

$$G = \begin{bmatrix} \frac{T^2}{2} & 0 & T & 0 \\ 0 & \frac{T^2}{2} & 0 & T \end{bmatrix}^T.$$

Suppose that the target position can be measured by a network of sensors, whose measurement equations are

$$y_{i,k} = C_i x_k + v_{i,k}, \quad i = 1, \dots, N.$$

The measurement matrices  $C_i, i = 1, \dots, N$  satisfy Assumption 2, of the following form

$$\begin{aligned} C_1 &= [1 \ 0 \ 0 \ 0], & C_2 &= [0 \ 1 \ 0 \ 0], \\ C_3 &= [0 \ 1 \ 0 \ 0], & C_4 &= [1 \ 0 \ 0 \ 0], \\ C_5 &= [0 \ 1 \ 0 \ 0], & C_6 &= [1 \ 0 \ 0 \ 0]. \end{aligned}$$

The topology of communication network is shown in Fig. 1.

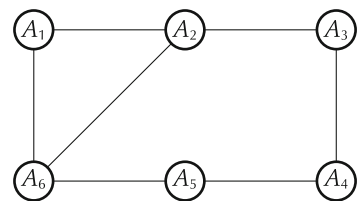


Fig. 1 Topology of the network.

The target trajectory is generated by (39) with the initial state  $x_0 = [0 \ 0 \ 2 \ 2]^T$ . The initial state of each agent is  $\bar{x}_{i,0} = x_0 + [10 \ 10 \ 5 \ 5]^T, \forall i \in \mathcal{V}$  and  $\bar{P}_{i,0} = \text{diag}\{10^2 \ 10^2 \ 5^2 \ 5^2\}, \forall i \in \mathcal{V}$ . The true measurement noise variance generated by  $R_{i,k} = 0.2 + 0.4(1 + \tanh(0.1T(k - T_f/4))), k = 1, 2, \dots$ , where  $T_f = 500$  is total simulation time. The fading parameter is  $\mu = 0.9$  for all agents. The initial parameters for  $q(\Phi_{i,0})$  are  $v_{i,0} = 1$  and  $V_{i,0} = 7.5, i \in \mathcal{V}$ . The number of recursive VB steps is  $S = 3$ .

Compare the performance of the DAVBKF with a centralized adaptive VB Kalman filter (CAVBKF), which is the linear counterpart the one given in [16] (see Corollary 1), by conducting numerical simulations, where 200 Monte Carlo trials for the CAVBKF and the DAVBKF are performed, respectively. In the CAVBKF, the measurement matrix is  $C = [C_1^T, \dots, C_6^T]^T$  and  $R_k = \text{diag}\{R_{1,k}, \dots, R_{6,k}\}$ .

Consider the mean square estimation errors (MSE), which has been widely used for the estimator performance,

$$\text{MSE}_k = \frac{1}{N} \sum_{i=1}^N \left[ \frac{1}{200} \sum_{j=1}^{200} (\hat{x}_{i,k}^{(j)} - x_k)^T (\hat{x}_{i,k}^{(j)} - x_k) \right].$$

In order to compare the estimation errors of the noise variances, we define the total estimate errors of noise variance as

$$\text{TEER}_k = \frac{1}{N} \sum_{i=1}^N \left[ \left( \frac{1}{200} \sum_{j=1}^{200} \text{tr}[(\hat{\Phi}_{i,k}^{(j)})^{-1}] \right) - \text{tr}(R_k) \right].$$

Figs.2 and 3 show the MSEs of position and velocity, respectively. DAVBKF-1 and DAVBKF-100 represent the numbers about consensus are  $\bar{l} = 1$  and  $\bar{l} = 100$  for the DAVBKF, respectively. It can be seen that the MSE of DAVBKF-1 is little larger than that of the MSE of DAVBKF-100. However, the DAVBKF performs better than the CAVBKF even when  $\bar{l} = 1$ . Fig. 4 shows the estimates of noise variance for different agent, and Fig. 5 shows the estimation errors of measurement noise variance between DAVBKF-100 and the CAVBKF, which demonstrates that the DAVBKF can outperform the CAVBKF.

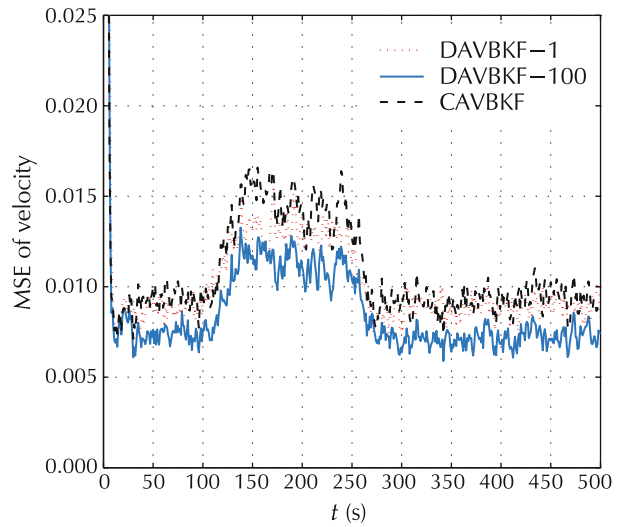


Fig. 3 MSE of velocity.

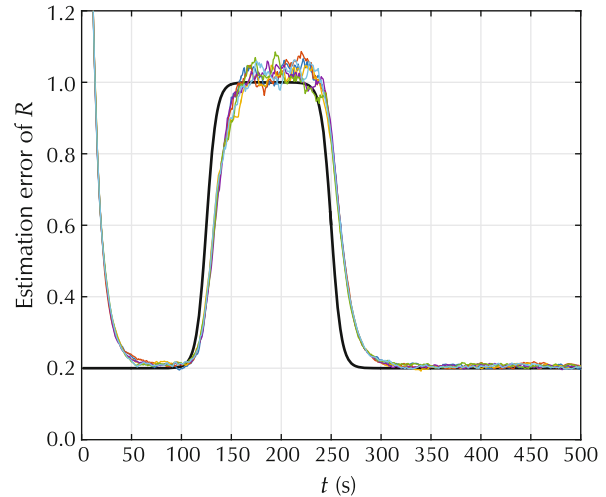


Fig. 4 Estimate of noise variance.

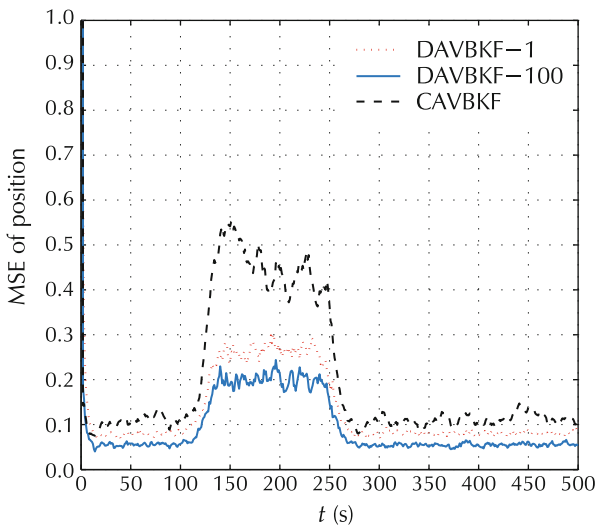


Fig. 2 MSE of position.

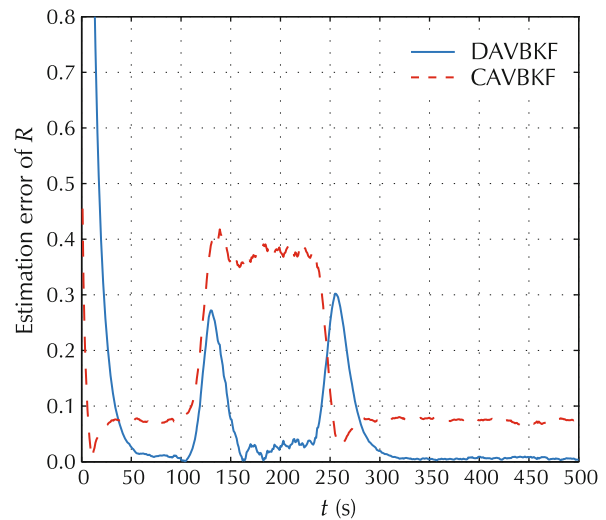


Fig. 5 Estimate errors of noise variance.



A reason why the DAVBKF performs better than the CAVBKF is that the global ELBO  $\mathcal{L}(q)$  over the whole network may be less than the summation of the local ELBO  $\mathcal{L}_i(q)$  over each agent [25], i.e.,  $\mathcal{L}(q^*) \leq \sum_{i=1}^N \mathcal{L}_i(q_i^*)$ , which shows that the DAVBKF may perform better than the CAVBKF.

## 5 Conclusions

In this paper, a distributed adaptive VB Kalman filter was investigated with unknown measurement noise variance modeled by the Wishart distribution. The posterior distribution of the joint state and noise variance were approximated by a free-form variational distribution. It was shown that the estimate of the noise variance only depended on the local information and the estimate could be achieved by a consensus algorithm based on VB. In the numerical simulations, the distributed tracking problem was studied, which verified and showed the effectiveness of the proposed algorithm.

## References

- [1] R. Olfati-Saber. Distributed Kalman filtering for sensor networks. *IEEE Conference on Decision and Control*, New York: IEEE, 2007: 5492 – 5498.
- [2] S. S. Stanković, M. S. Stanković, D. M. Stipanović. Consensus based overlapping decentralized estimation with missing observations and communication faults. *Automatica*, 2009, 45(6): 1397 – 1406.
- [3] D. W. Casbeer, R. Beard. Distributed information filtering using consensus filters. *Proceedings of the American Control Conference*, St. Louis: IEEE, 2009: 1882 – 1887.
- [4] A. T. Kamal, J. A. Farrell, A. K. Roy-Chowdhury. Information weighted consensus filters and their application in distributed camera networks. *IEEE Transactions on Automatic Control*, 2013, 58(12): 3112 – 3125.
- [5] G. Battistelli, L. Chisci. Kullback-Leibler average, consensus on probability densities, and distributed state estimation with guaranteed stability. *Automatica*, 2014, 50(3): 707 – 718.
- [6] G. Battistelli, L. Chisci. Stability of consensus extended Kalman filter for distributed state estimation. *Automatica*, 2016, 68: 169 – 178.
- [7] W. Yang, C. Yang, H. Shi, et al. Stochastic link activation for distributed filtering under sensor power constraint. *Automatica*, 2017, 75: 109 – 118.
- [8] Z. Zhou, H. Fang, Y. Hong. Distributed estimation for moving target based on state-consensus strategy. *IEEE Transactions on Automatic Control*, 2013, 58(8): 2096 – 2101.
- [9] R. Mehra. On the identification of variances and adaptive Kalman filtering. *IEEE Transactions on Automatic Control*, 1970, 15(2): 175 – 184.
- [10] A. Mohamed, K. Schwarz. Adaptive Kalman filtering for INS/GPS. *Journal of Geodesy*, 1999, 73(4): 193 – 203.
- [11] X. R. Li, Y. Bar-Shalom. A recursive multiple model approach to noise identification. *IEEE Transactions on Aerospace and Electronic Systems*, 1994, 30(3): 671 – 684.
- [12] R. Mehra. Approaches to adaptive filtering. *IEEE Transactions on Automatic Control*, 1972, 17(5): 693 – 698.
- [13] S. Sarkka, A. Nummenmaa. Recursive noise adaptive Kalman filtering by variational Bayesian approximations. *IEEE Transactions on Automatic Control*, 2009, 54(3): 596 – 600.
- [14] M. J. Beal. *Variational Algorithms for Approximate Bayesian Inference*. London: University of London, 2003.
- [15] H. Zhu, H. Leung, Z. He. State estimation in unknown nongaussian measurement noise using variational Bayesian technique. *IEEE Transactions on Aerospace and Electronic Systems*, 2013, 49(4): 2601 – 2614.
- [16] P. Dong, Z. Jing, H. Leung, et al. Variational Bayesian adaptive cubature information filter based on wishart distribution. *IEEE Transactions on Automatic Control*, 2007, 62(11): 6051 – 6057.
- [17] A. Gelman, J. B. Carlin, H. S. Stern, et al. *Bayesian Data Analysis*. Boca Raton: CRC Press, 2014.
- [18] B. D. Anderson, J. B. Moore. *Optimal filtering*. Englewood Cliffs: Prentice-Hall, 1979.
- [19] S. Särkkä. *Bayesian Filtering and Smoothing*. New York: Cambridge University Press, 2013.
- [20] V. Šmídl, A. Quinn. *The Variational Bayes Method in Signal Processing*. Berlin: Springer, 2006.
- [21] M. Yang. Variational Bayesian data analysis on manifold. *Control Theory & Technology*, 2018, 16(3): 212 – 220.
- [22] R. Olfati-Saber, J. A. Fax, R. M. Murray. Consensus and cooperation in networked multi-agent systems. *Proceedings of the IEEE*, 2007, 95(1): 215 – 233.
- [23] M.-A. Sato. Online model selection based on the variational Bayes. *Neural Computation*, 2001, 13(7): 1649 – 1681.
- [24] X. R. Li, V. P. Jilkov. Survey of maneuvering target tracking – Part I: Dynamic models. *IEEE Transactions on Aerospace and Electronic Systems*, 2003, 39(4): 1333 – 1364.
- [25] J. Hua, C. Li. Distributed variational Bayesian algorithms over sensor networks. *IEEE Transactions on Signal Processing*, 2016, 64(3): 783 – 798.
- [26] G. Casella, R. L. Berger. *Statistical Inference*. Duxbury: Duxbury Press, 2002.

## Appendix

**Proof of Lemma 1** Clearly, with  $q_i \triangleq q(x_k)q(\Phi_i)$ ,

$$\begin{aligned} \mathcal{L}(q(\Phi_{i,k})) &= E_{q_i} [\log(p(y_{i,k} | \Phi_{i,k}, x_k))] \\ &\quad + \frac{1}{N} E_{q_i} \left[ \log \frac{p(\Phi_{i,k}, x_k)}{q(x_k)q(\Phi_{i,k})} \right] + \bar{c} \\ &= \mathcal{L}_i(q(\Phi_{i,k})) + \bar{c}, \end{aligned}$$

where  $\bar{c}$  denotes a constant independent of  $\Phi_{i,k}$ , and  $\mathcal{L}_i(q)$  is defined in (24). For fixed  $q(x_k)$ , the variational distribution

$q(\Phi_{i,k})$  only related to the local measurement  $y_{i,k}$ , and thus, it can be solved individually by each agent.

Then the conditional distribution  $p(y_{i,k}|\Phi_{i,k}, x_k)$  is

$$\begin{aligned} p(y_{i,k}|\Phi_{i,k}, x_k) &= \frac{|\Phi_{i,k}|^{\frac{1}{2}}}{\sqrt{(2\pi)^{m_i}}} \exp\left\{-\frac{1}{2}(y_{i,k} - C_{i,k}x_k)^T \Phi_{i,k} (y_{i,k} - C_{i,k}x_k)\right\} \\ &= \exp\left\{-\frac{1}{2}\text{tr}((y_{i,k} - C_{i,k}x_k)(y_{i,k} - C_{i,k}x_k)^T \Phi_{i,k})\right. \\ &\quad \left. + \frac{1}{2} \log |\Phi_{i,k}| + \log \frac{1}{\sqrt{(2\pi)^{m_i}}}\right\}. \end{aligned}$$

Rewrite the following probability densities in the exponential family form [26]. Denote the natural parameter of  $q(\Phi_{i,k})$  is  $\hat{\eta}_{i,k}$ , and corresponding sufficient statistical and log-normalizer are  $u(\Phi_{i,k})$  and  $A_I(\hat{\eta})$ , respectively. Rewrite  $\mathcal{L}_i(q(\Phi_{i,k}))$  in the exponential family distribution form:

$$\begin{aligned} \mathcal{L}_i(q(\Phi_{i,k})) &= E_{q_i}[\hat{\eta}_{i,k}^T u(\Phi_{i,k})] + \frac{1}{N} E_{q_i}[(\hat{\eta}_{i,k} - \bar{\eta}_{i,k})^T u(\Phi_{i,k})] \\ &\quad - A_I(\bar{\eta}_{i,k}) + A_I(\hat{\eta}_{i,k}) + \bar{c}_1, \end{aligned} \tag{a1}$$

where  $\bar{c}_1$  is a constant independent of  $\Phi_{i,k}$ , and  $\bar{\eta}_{i,k}$  and  $A_I(\bar{\eta}_{i,k})$  are natural parameter and log-normalizer of  $\mathcal{W}(\Phi_{i,k}|\bar{v}_{i,k}, \bar{V}_{i,k})$ , respectively.  $\bar{\eta}_{i,k}$  and  $u(\Phi_{i,k})$  are given by<sup>1</sup>

$$\begin{aligned} \bar{\eta}_{i,k} &= \begin{bmatrix} -\frac{1}{2}(y_{i,k} - C_{i,k}x_k)(y_{i,k} - C_{i,k}x_k)^T \\ \frac{1}{2} \end{bmatrix}, \\ u(\Phi_{i,k}) &= \begin{bmatrix} \Phi_{i,k} \\ \log |\Phi_{i,k}| \end{bmatrix}. \end{aligned}$$

By taking the derivative of  $\mathcal{L}_i(q(\Phi_{i,k}))$  with respect to  $\hat{\eta}_{i,k}$ , we have

$$\nabla_{\hat{\eta}} \mathcal{L}_i(\hat{\eta}_{i,k}) = \nabla_{\hat{\eta}}^2 A_I(\hat{\eta}_{i,k})^T (E_{q_i}[\hat{\eta}_{i,k}] + \frac{1}{N}(\bar{\eta}_{i,k} - \hat{\eta}_{i,k}))$$

due to  $E[u(\Phi_{i,k})] = \nabla_{\hat{\eta}} A_I(\hat{\eta}_{i,k})$  [26]. By letting  $\nabla_{\hat{\eta}} \mathcal{L}_i(\hat{\eta}_{i,k}) = 0$ , we have the solution of natural parameters of  $q(\Phi_{i,k})$ ,

$$\hat{\eta}_{i,k} = N E_{q_i}[\bar{\eta}_{i,k}] + \bar{\eta}_{i,k}. \tag{a2}$$

Note that

$$\begin{aligned} E_{q_i(x_k)}[(y_{i,k} - C_{i,k}x_k)(y_{i,k} - C_{i,k}x_k)^T] &= E_{q_i(x_k)}[(y_{i,k} - C_{i,k}\hat{x}_k - C_{i,k}x_k + C_{i,k}\hat{x}_k)(\cdot)^T] \\ &= (y_{i,k} - C_{i,k}\hat{x}_k)(y_{i,k} - C_{i,k}\hat{x}_k)^T \\ &\quad + C_i E_{q_i(x_k)}(x_k - \hat{x}_k)(x_k - \hat{x}_k)^T C_{i,k}^T \\ &= (y_{i,k} - C_i\hat{x}_k)(y_{i,k} - C_{i,k}\hat{x}_k)^T + C_{i,k}P_k C_{i,k}^T. \end{aligned}$$

<sup>1</sup> Notice that we use  $\text{tr}(MN) = \text{vec}(M) \cdot \text{vec}(N)$ , and then vectorizing the matrix parameter when inserted into the exponential family form.

Then  $\hat{v}_{i,k}$  and  $\hat{V}_{i,k}$  are given by

$$\begin{aligned} \hat{v}_{i,k} &= \bar{v}_{i,k} + N, \\ \hat{V}_{i,k}^{-1} &= N[C_{i,k}P_k C_{i,k}^T + (y_{i,k} - C_{i,k}\hat{x}_k)(\cdot)^T] + \bar{V}_{i,k}^{-1}. \end{aligned}$$

□

**Proof of Lemma 2** Clearly,

$$p(y_{i,k}|\Phi_{i,k}, x_k) = \frac{p(x_k|y_{i,k}, \Phi_{i,k})p(y_{i,k}, \Phi_{i,k})}{p(\Phi_{i,k}, x_k|y_{i,k})},$$

and then

$$\begin{aligned} \mathcal{L}_i(q(x_k)) &= E_q[\log \frac{p(x_k|y_{i,k}, \Phi_{i,k})}{p(x_k)}] + \frac{1}{N} E_q[\log \frac{p(x_k)}{q(x_k)}] + \bar{c}_0, \end{aligned}$$

where  $\bar{c}_0$  denotes the terms independent of  $x_k$ .

Rewrite

$$p(x_k|y_{i,k}, \Phi_{i,k}) = h(x_k) \exp\{\theta^T(y_{i,k}, \Phi_{i,k})u(x_k) - A_g(\theta(y_{i,k}, \Phi_{i,k}))\}, \tag{a3}$$

$$p(x_k|y_{k-1}) = h(x_k) \exp\{\theta^T(y_{k-1})u(x_k) - A_g(\theta(y_{k-1}))\}, \tag{a4}$$

$$q(x_k) = h(x_k) \exp\{\theta_k^T u(x_k) - A_g(\theta_k)\}, \tag{a5}$$

where the natural parameter of a Gaussian distribution

$$p(x; \hat{x}, P) \text{ is } \theta = \begin{bmatrix} P^{-1}\hat{x} \\ -\frac{1}{2}P^{-1} \end{bmatrix} \text{ [26].}$$

Substitute equations (a3), (a4) and (a5) into  $\mathcal{L}_i(q)$ ,

$$\begin{aligned} \mathcal{L}_i(q(x_k)) &= E_q[(\theta(y_{i,k}, \Phi_{i,k}) - \theta(y_{k-1}))^T u(x_k) - A_g(\theta(y_{i,k}, \Phi_{i,k})) \\ &\quad + A_g(\theta(y_{k-1}))] + \frac{1}{N} E_q[(\theta(y_{k-1}) - \theta_k)^T u(x_k) \\ &\quad - A_g(\theta(y_{k-1})) + A_g(\theta_k)] + \bar{c}_1 \\ &= E_{q(\Phi_{i,k})}[(\theta(y_{i,k}, \Phi_{i,k}) - \theta(y_{k-1}))^T E_{q(x_k)}[u(x_k)] \\ &\quad + \frac{1}{N}\{E_{q(\Phi_{i,k})}[\theta(y_{k-1}) - \theta_k]^T E_{q(x_k)}[u(x_k)] \\ &\quad + A_g(\theta_k)\}] + \bar{c}_2 \end{aligned}$$

due to the exponential family property  $E_{q(x_k)}[u(x_k)] = \nabla_{\theta_k} \times A_g(\theta_k)$ , where  $\bar{c}_1$  and  $\bar{c}_2$  are the terms independent of  $x_k$ .

Taking the derivative of  $\mathcal{L}_i(q(x_k))$  with respect to  $\theta_k$  gives

$$\begin{aligned} \nabla_{\theta_k} \mathcal{L}_i(\theta_k) &= \nabla_{\theta_k}^2 A_g(\theta_k)^T (E_{q(\Phi_{i,k})}[\theta(y_{i,k}, \Phi_{i,k})] - \theta(y_{k-1})) \\ &\quad + \frac{1}{N}(\theta(y_{k-1}) - \theta_k). \end{aligned}$$

Then the global optimal solution can be found by letting

$$\sum_{i=1}^N \nabla_{\theta_k} \mathcal{L}_i(\theta_k) = 0, \text{ which yields}$$

$$\theta_k^* = -(N-1)\theta(y_{k-1}) + \sum_{i=1}^N E_{q(\Phi_{i,k})}[\theta(y_{i,k}, \Phi_{i,k})]. \tag{a6}$$

With a similar procedure, the local optimal solution can be given as

$$\theta_{i,k} = -(N - 1)\theta(y_{k-1}) + N E_{q(\Phi_{i,k})}[\theta(y_{i,k}, \Phi_{i,k})]. \quad (a7)$$

Combining (a6) and (a7) yields

$$\theta_k = \frac{1}{N} \sum_{i=1}^N \theta_{i,k}. \quad (a8)$$

Since  $p(x_{i,k}|y_{i,k}, \Phi_{i,k})$  and  $p(x_{i,k}|y_{i,k-1})$  are Gaussian distributions,  $\theta_{i,k} = \begin{bmatrix} (\check{P}_{i,k})^{-1} \check{x}_{i,k} \\ -\frac{1}{2}(\check{P}_{i,k})^{-1} \end{bmatrix}$  can be given as

$$(\check{P}_{i,k})^{-1} = \bar{P}_{i,k}^{-1} + N C_{i,k}^T \check{\Phi}_{i,k} C_{i,k}, \quad (a9)$$

$$(\check{P}_{i,k})^{-1} \check{x}_{i,k} = \bar{P}_{i,k}^{-1} \bar{x}_{i,k} + N C_{i,k}^T \check{\Phi}_{i,k} y_{i,k}, \quad (a10)$$

where  $\check{\Phi}_{i,k} = E_{q(\Phi_{i,k})}[\Phi_{i,k}] = \hat{v}_{i,k} \hat{V}_{i,k}$ . With (a8), (a9), and (a10), the conclusions follow.  $\square$



**Chen HU** received his B.Eng. degree from the Department of Automation, Xiamen University, Xiamen, China, in 2011 and the M.Sc. and Ph.D. degrees from the Rocket Force University, in 2014 and 2018, respectively. His research interests include distributed state estimation and machine learning. E-mail: chenauthor@163.com. — —



**Xiaoming HU** received the B.Sc. degree from the University of Science and Technology of China in 1983, and the M.Sc. and Ph.D. degrees from the Arizona State University, in 1986 and 1989, respectively. He served as a research assistant at the Institute of Automation, the Chinese Academy of Sciences, from 1983 to 1984. From 1989 to 1990 he was a Gustafsson Postdoctoral Fellow at the Royal Institute of Technology, Stockholm, where he is currently a professor of Optimization and Systems Theory. His main research interests are in nonlinear control systems, nonlinear observer design, sensing and active perception, motion planning, control of multi-agent systems, and mobile manipulation. E-mail: hu@kth.se.



**Yiguang HONG** received his B.Sc. and M.Sc. degrees from Peking University and his Ph.D. degree from Institute of Systems Science, Chinese Academy of Sciences (CAS). He is currently a professor in Academy of Mathematics and Systems Science, CAS. His research interests include multi-agent control, distributed optimization and game, nonlinear control, social networks, and machine learning. E-mail: yghong@iss.ac.cn.