SCIENCE CHINA
Information Sciences

• RESEARCH PAPER •
Special Focus on Multi-source Information Fusion

November 2020, Vol. 63 210202:1–210202:14
https://doi.org/10.1007/s11432-020-3000-1

Fully distributed variational Bayesian non-linear filter with unknown measurement noise in sensor networks

Yu LIU[1,2], Jun LIU[1*], Congan XU[1], Gang LI[2] & You HE[1]

[1]*Research Institute of Information Fusion, Naval Aviation University, Yantai* 264001*, China;*
[2]*Department of Electronic Engineering, Tsinghua University, Beijing* 100084*, China*

Received 13 May 2020/Revised 18 June 2020/Accepted 9 July 2020/Published online 22 October 2020

Abstract In practical applications, the measurement noise statistics is usually unknown or may change over time. However, most existing distributed filtering algorithms for sensor networks are constructed based on exact knowledge of measurement noise statistics. Therefore, under situations with measurement uncertainty, the existing algorithms may result in deteriorated performance. To solve such problems, a distributed adaptive cubature information filter based on variational Bayesian (VB-DACIF) is proposed here. Firstly, the predicted estimates of interest from inclusive neighbours are fused by minimizing the weighted Kullback-Leibler average, in which the cubature rule is utilized to tackle system nonlinearity. Then, the free form variational Bayesian approximation is applied to recursively update both the local estimate and the precision matrices of sensing nodes. Finally, the posterior Cramér-Rao lower bound is exploited to evaluate performance of the proposed VB-DACIF. Simulation results with a maneuvering target tracking scenario validates the feasibility and superiority of the proposed VB-DACIF.

Keywords sensor networks, distributed state estimation, cubature Kalman filter, variational Bayesian inference, Kullback-Leibler divergence

Citation Liu Y, Liu J, Xu C A, et al. Fully distributed variational Bayesian non-linear filter with unknown measurement noise in sensor networks. Sci China Inf Sci, 2020, 63(11): 210202, https://doi.org/10.1007/s11432-020-3000-1

1 Introduction

Distributed state estimation (DSE) with networked sensors has attracted a great amount of attention in recent years. In the distributed paradigm, neighboring nodes iteratively communicate local state related information with each other, and the entire network is finally expected to reach a global decision (to some degree) [1–4]. Unlike the centralized schemes [5,6], where measurement information from different sensors is integrated in the so-called fusion center, DSE algorithms accomplish estimation tasks in a distributed manner, and no fusion center or local observability of each node is required. Due to its easy scalability for adding new nodes or removing existing nodes and robust to node failures [7,8], DSE algorithms have been extensively researched in fields such as environmental monitoring, intelligent transportation and target tracking [9,10].

The well-known Kalman consensus filter [1,11] performs an average consensus only on prior estimates and works well when the target of interest can be jointly observed by the inclusive neighbors. The consensus on measurements (CM) [2,12] approach performs consensus on measurement information obtained by neighboring nodes, but it requires sufficient consensus iterations to ensure filtering stability. On the other hand, from an information-theoretic perspective, the consensus on information (CI) approach proposed

* Corresponding author (email: 18615042187@163.com)

© Science China Press and Springer-Verlag GmbH Germany, part of Springer Nature 2020

in [13] guarantees filtering stability for any number (even a single one) of consensus iterations, but the estimation accuracy may deteriorate since a conservative fusion rule is adopted. To take advantage of the complementary benefits of both approaches, the so-called hybrid CMCI (HCMCI) was proposed in [2, 3]. However, all these approaches require multiple communication iterations between consecutive sampling intervals to achieve an acceptable performance, and hence may result in heavy burden on computation and communication. To alleviate this issue, the diffusion-based Kalman filter was proposed in [14], but it requires local joint detectability at each node which is generally impractical as discussed in [2, 15]. Recently, an innovative distributed hybrid information fusion (DHIF) scheme was proposed in [4], which only requires single communication iteration between neighboring nodes and is robust against the existence of nodes with no sensing abilities. Although the above-mentioned algorithms are designed for linear systems, it can be further extended to nonlinear systems by directly applying the extended Kalman filter (EKF) as presented in [16, 17]. Other efforts to handle nonlinearities for distributed filtering include the unscented Kalman filter (UKF) based algorithms [18, 19], and the particle filter (PF) based algorithms [20, 21]. However, the UKF needs to adjust the scaling parameter and may halt its operation in the case of negative weights, while the PF suffers from huge amount of computation. The cubature Kalman filter (CKF) presented in [22] may be a remedy to avoid parameter tuning and negative weights in UKF, and is more computationally efficient compared with the PF-based algorithms. More about the CKF-based DSE algorithms can be found in [23–25].

The aforementioned algorithms are derived with full knowledge of measurement noise statistics. However, it is unrealistic to accurately model the measurement noise statistics for sensor nodes deployed in a changeable environment. Adaptive filtering is the most common approach to address this issue of parameter uncertainty, in which the noise statistics and dynamic state are simultaneously estimated from the obtained measurements [26]. As one of the most general noise adaptive filtering techniques, the Bayesian approach, such as state augmentation [27], interacting multiple models [28], and particle filters [29], has attracted increasing popularity in tackling state estimation problems. In [30], an adaptive Kalman filter based on the free form variational Bayesian (VB) approximation was proposed to jointly estimate the dynamic state and measurement noise covariance, which benefits from lower computational load and analytical tractability. This work was further extended to the nonlinear filtering problems in [31], in which the unknown measurement noise covariance is modelled by an inverse Wishart distribution that takes noise correlations into account. In [32], a VB adaptive cubature information filter (CIF) based on Wishart distribution was presented from the perspective of information filtering. However, these VB-based algorithms are carried out in a centralized manner. Recently, a VB consensus CKF (VB-CCKF) was proposed in [33] to handle the measurement uncertainty in a distributed style. However, it bears inherent drawbacks of the consensus-based approaches that require multiple communication iterations between neighboring nodes.

Motivated by the studies mentioned above, a distributed adaptive cubature information filter based on VB (VB-DACIF) is proposed for nonlinear filtering with measurement uncertainty. In the proposed VB-DACIF, prior estimates from the neighboring nodes are fused before measurement update by minimizing the weighted Kullback-Leibler divergence (KLD), which is effective to assimilate information with possible unknown correlated errors [34]. Then, the free form VB approximation combined with the DHIF scheme is used to recursively approximate the unknown measurement noise covariance and update the posterior estimate for each node. During each cycle of VB recursion, only a single communication iteration is required, which mitigates the computational burden for the entire network compared with the consensus-based schemes. The main contributions are summarized as follows: (1) an extension of DHIF to nonlinear systems with measurement uncertainty is proposed from an information-theoretic point of view; (2) with resort to the VB approximation, a distributed adaptive cubature information filter based on KLD is derived; (3) the estimation performance of the proposed VB-DACIF is evaluated by the posterior Cramér-Rao lower bound (PCRLB).

The rest of this paper is arranged as follows. Section 2 formulates the nonlinear DSE problem with measurement uncertainty, and briefly reviews some relative preliminaries. Derivation of the proposed VB-DACIF is detailed in Section 3. The PCRLB analysis is presented in Section 4. A maneuvering target

tracking scenario is considered in Section 5. Finally, some concluded remarks are given in Section 6.

2 Problem formulation

To be clear, some notations are defined in advance. $\mathbb{R}^n$ denotes the Euclidean space of n-dimension. The inverse and transpose of arbitrary matrix $\boldsymbol{A}$ are denoted by $\boldsymbol{A}^{-1}$ and $\boldsymbol{A}^{\mathrm{T}}$, respectively. The trace of $\boldsymbol{A}$ is written as $\mathrm{tr}(\boldsymbol{A})$. $\boldsymbol{I}_n$ represents the n-dimensional identity matrix. $\mathrm{E}\{\cdot\}$ denotes the statistical expectation operator.

2.1 System model

Consider a sensor network composed of two types of nodes: communication nodes and sensor nodes. The former is able to process local information and communicate with its neighboring nodes, while the latter is also capable of sensing information from the surveillance area. The communication topology between nodes is denoted by an undirected graph $\mathcal{G} = (\mathcal{S}, \mathcal{C}, \mathcal{E})$. Here, $\mathcal{S}$ and $\mathcal{C}$ represent the set of sensor nodes and communication nodes, respectively. $\mathcal{N} = \mathcal{S} \cup \mathcal{C}$ is the set of all nodes in the network, and $\mathcal{E} \subseteq \mathcal{N} \times \mathcal{N}$ is the edge set such that $(i, j) \in \mathcal{S}$ if node i is able to communicate with node j. $\mathcal{N}_i = \{j | (i, j) \in \mathcal{E}, \forall j \neq i\}$ represents the immediate neighbors of node i. Further, $\mathcal{J}_i = \mathcal{N}_i \cup \{i\}$ denotes the inclusive neighbors of node i.

Consider the following discrete-time nonlinear dynamic system:

$$\boldsymbol{x}_k = \boldsymbol{f}(\boldsymbol{x}_{k-1}) + \boldsymbol{w}_{k-1}, \tag{1}$$

$$\boldsymbol{z}_{i,k} = \boldsymbol{h}_i(\boldsymbol{x}_k) + \boldsymbol{v}_{i,k},\ i \in \mathcal{S}, \tag{2}$$

where $\boldsymbol{x}_k \in \mathbb{R}^n$ denotes the state vector and $\boldsymbol{z}_{i,k} \in \mathbb{R}^m$ is the measurement of node i. $\boldsymbol{f}(\cdot)$ is the nonlinear state transition function. $\boldsymbol{h}(\cdot)$ denotes the measurement function. $\boldsymbol{w}_{k-1} \sim N(\boldsymbol{0}, \boldsymbol{Q}_{k-1})$ is the Gaussian process noise with zero-mean and covariance. $\boldsymbol{v}_{i,k} \sim N(\boldsymbol{0}, \boldsymbol{R}_{i,k})$ is the zero-mean Gaussian measurement noise with unknown covariance $\boldsymbol{R}_{i,k}$. To facilitate the proposed VB-DACIF in the information form, the precision matrix $\boldsymbol{\Lambda}_{i,k} = \boldsymbol{R}_{i,k}^{-1}$, i.e., inverse of the measurement noise covariance, is used in the later derivation. Since Wishart distribution is a conjugate prior for the unknown precision matrix of a Gaussian distribution with known mean [35], the precision matrix of each sensor node can be modeled by

$$W(\boldsymbol{\Lambda}_{i,k};\, v_{i,k}, \boldsymbol{V}_{i,k}) = c\, |\boldsymbol{\Lambda}_{i,k}|^{(v_{i,k}-m-1)/2} \exp\left(-\frac{1}{2}\mathrm{Tr}\left(\boldsymbol{V}_{i,k}^{-1}\boldsymbol{\Lambda}_{i,k}\right)\right), \tag{3}$$

where $v_{i,k}$ is the degree of freedom, $\boldsymbol{V}_{i,k}$ is the positive scale matrix, and c is the normalization factor independent of $\boldsymbol{\Lambda}_{i,k}$.

2.2 Information fusion based on Kullback-Leibler divergence

From the perspective of information theory, the KLD can be used to measure the difference between two probability density functions (PDFs) $p(\boldsymbol{x})$ and $q(\boldsymbol{x})$. The KLD is defined as follows:

$$\mathrm{KLD}(q(\boldsymbol{x})\,\|p(\boldsymbol{x})\,) = \int q(\boldsymbol{x}) \log\left(\frac{q(\boldsymbol{x})}{p(\boldsymbol{x})}\right) \mathrm{d}\boldsymbol{x}. \tag{4}$$

The weighted Kullback-Leibler average (KLA) over a set of PDFs $\{p^i(\boldsymbol{x})\}$ with relative weights w_i is defined by

$$\bar{p}^i(\boldsymbol{x}) = \arg\min_{p(\boldsymbol{x})} \sum_{i\in\mathcal{N}} w_i \mathrm{KLD}\left(q(\boldsymbol{x})\,\|p(\boldsymbol{x})\,\right), \tag{5}$$

where $w_i \geqslant 0$ satisfies $\sum_{i\in\mathcal{N}} w_i = 1$. It can be seen from (5) that minimizes the sum of information gains from the initial set of PDFs $\{p^i(\boldsymbol{x})\}$. Such a choice best describes the current knowledge of the estimated state and may yield an information gain as small as possible [36]. Therefore, it is rational to exploit (5)

for information fusion. It has been shown in [34] that the weighted KLA defined by (5) is consistent with the normalized weighted geometric mean of the initial PDFs, which can be formulated as

$$\bar{p}^i(\boldsymbol{x}) = \frac{\prod_{i\in\mathcal{N}}[p^i(\boldsymbol{x})]^{w_i}}{\int\prod_{i\in\mathcal{N}}[p^i(\boldsymbol{x})]^{w_i}\mathrm{d}\boldsymbol{x}}. \tag{6}$$

It can be seen that Eq. (6) provides a more explicit expression than (5), and thus can be utilized to integrate the given set of PDFs $\{p^i(\boldsymbol{x})\}$ for a more accurate solution. Further, if the given PDFs in (6) are Gaussian with $p^i(\boldsymbol{x}) = N(\boldsymbol{x};\boldsymbol{\mu}_i,\boldsymbol{P}_i)$, then the obtained $\bar{p}$ is also Gaussian and takes the form $\bar{p}(\boldsymbol{x}) = N(\boldsymbol{x};\bar{\boldsymbol{\mu}},\bar{\boldsymbol{P}})$ with

$$\bar{\boldsymbol{P}}^{-1} = \sum_{i\in\mathcal{N}} w_i\boldsymbol{P}_i^{-1}, \quad \bar{\boldsymbol{P}}^{-1}\bar{\boldsymbol{\mu}} = \sum_{i\in\mathcal{N}} w_i\boldsymbol{P}_i^{-1}\boldsymbol{\mu}_i. \tag{7}$$

The proof is available in [34] and hence omitted here. By writing (7) in information form, one has

$$\bar{\boldsymbol{Y}} = \sum_{i\in\mathcal{N}} w_i\boldsymbol{Y}_i, \quad \bar{\boldsymbol{y}} = \sum_{i\in\mathcal{N}} w_i\boldsymbol{y}_i, \tag{8}$$

where $\boldsymbol{Y}_i = \boldsymbol{P}_i^{-1}$, $\boldsymbol{y}_i = \boldsymbol{P}_i^{-1}\boldsymbol{\mu}_i$. It is worth mentioning that the fusion rule (6) coincides with the so-called generalized covariance intersection (GCI) which was discussed in [37–39] as a generalization of the covariance intersection originally designed for Gaussian PDFs [40]. It is well known that GCI provides a fusion strategy that is robust against the unknown correlations between diverse information sources.

2.3 Variational Bayesian approximation

The joint distribution of $\boldsymbol{x}_k$ and $\boldsymbol{\Lambda}_k = \{\boldsymbol{\Lambda}_{i,k}|i\in\mathcal{S}\}$ can be approximated by the VB approximation [30,35] as follows:

$$p(\boldsymbol{x}_k,\boldsymbol{\Lambda}_k|\boldsymbol{Z}_k) \approx q(\boldsymbol{x}_k)q(\boldsymbol{\Lambda}_k), \tag{9}$$

where $\boldsymbol{Z}_k = \{\boldsymbol{z}_1,\boldsymbol{z}_2,\ldots,\boldsymbol{z}_k\}$ denotes the collective measurements of all sensor nodes up to time instant k, with $\boldsymbol{z}_k = \{\boldsymbol{z}_{i,k}|i\in\mathcal{S}\}$. $q(\boldsymbol{x}_k)$ and $q(\boldsymbol{\Lambda}_k)$ are the unknown approximated posterior PDFs, which can be obtained by minimizing the KLD between true posterior distribution and the approximation

$$(q(\boldsymbol{x}_k),q(\boldsymbol{\Lambda}_k)) = \underset{q(\boldsymbol{x}_k)q(\boldsymbol{\Lambda}_k)}{\arg\min}\ \mathrm{KLD}\left(q(\boldsymbol{x}_k)q(\boldsymbol{\Lambda}_k)\|p(\boldsymbol{x}_k,\boldsymbol{\Lambda}_k|\boldsymbol{Z}_k)\right). \tag{10}$$

The optimal solution to (10) is achieved with the following equations [32]:

$$\ln\hat{q}(\boldsymbol{x}_k) = \int \ln p(\boldsymbol{z}_k,\boldsymbol{x}_k,\boldsymbol{\Lambda}_k|\boldsymbol{Z}_{k-1})q(\boldsymbol{\Lambda}_k)\mathrm{d}\boldsymbol{\Lambda}_k + c_1, \tag{11}$$

$$\ln\hat{q}(\boldsymbol{\Lambda}_k) = \int \ln p(\boldsymbol{z}_k,\boldsymbol{x}_k,\boldsymbol{\Lambda}_k|\boldsymbol{Z}_{k-1})q(\boldsymbol{x}_k)\mathrm{d}\boldsymbol{x}_k + c_2, \tag{12}$$

where $\hat{q}(\boldsymbol{x}_k)$ and $\hat{q}(\boldsymbol{\Lambda}_k)$ are the desired PDFs. c_1 and c_2 are terms independent of $\boldsymbol{x}_k$ and $\boldsymbol{\Lambda}_k$, respectively. Since $q(\boldsymbol{x}_k)$ and $q(\boldsymbol{\Lambda}_k)$ are coupled with each other, it is intractable to obtain analytical solutions to (11) and (12). However, a fix-point iteration technique can be exploited to approximate the expected sufficient statistics, which will be discussed in detail later.

3 Distributed adaptive nonlinear filter based on weighted KLA and VB approximation

In actual applications, the measurement noise statistics of each sensor node is usually unknown or may change over time. Therefore, it is not practical to determine the measurement noise statistics ahead of time, which will make the conditions for traditional algorithms violated. To address state estimation problems with uncertain measurement noise, the VB approach is combined with the weighted KLA-based fusion rule to simultaneously estimate the state of interest and the noise statistics. With the third-degree

cubature quadrature rule [22] embedded into the filtering framework, a novel VB-DACIF for nonlinear system is proposed, which can be derived as follows.

Since only one iteration is allowed for each node $i \in \mathcal{N}$ to communicate with its neighbors, it is impossible for all nodes to obtain the global measurements $\boldsymbol{Z}_k$. In such a case, at time instant k, the available information sources are only the prior estimates and current measurements of node i and its neighbors. The goal is to achieve estimates of $\boldsymbol{x}_k$ and $\boldsymbol{\Lambda}_k$ as accurately as possible with available information from the inclusive neighbors.

Assume at time instant $k-1$, the local posterior PDF of $\boldsymbol{x}_{k-1}$ for node i follows Gaussian distribution

$$p^i(\boldsymbol{x}_{k-1}) = N(\boldsymbol{x}_{k-1}; \hat{\boldsymbol{x}}_{i,k-1|k-1}, \boldsymbol{P}_{i,k-1|k-1}) \tag{13}$$

and the local posterior PDF of $\boldsymbol{\Lambda}_{i,k-1}$ for node i is subject to Wishart distribution, i.e.,

$$p^i(\boldsymbol{\Lambda}_{i,k-1}) = W(\boldsymbol{\Lambda}_{i,k-1}; v_{i,k-1}, \boldsymbol{V}_{i,k-1}). \tag{14}$$

Based on the prediction step of CKF, $\boldsymbol{x}_k$ is predicted at node i by

$$p^i(\boldsymbol{x}_{k|k-1}) = N(\boldsymbol{x}_k; \hat{\boldsymbol{x}}_{i,k|k-1}, \boldsymbol{P}_{i,k|k-1}), \tag{15}$$

where

$$\hat{\boldsymbol{x}}_{i,k|k-1} = \frac{1}{2n}\sum_{s=1}^{2n} \boldsymbol{f}(\boldsymbol{\chi}_{i,s,k-1|k-1}), \tag{16}$$

$$\boldsymbol{P}_{i,k|k-1} = \frac{1}{2n}\sum_{s=1}^{2n} \boldsymbol{f}(\boldsymbol{\chi}_{i,s,k-1|k-1})\boldsymbol{f}^{\mathrm{T}}(\boldsymbol{\chi}_{i,s,k-1|k-1}) - \hat{\boldsymbol{x}}_{i,k|k-1}\hat{\boldsymbol{x}}_{i,k|k-1}^{\mathrm{T}} + \boldsymbol{Q}_{k-1}. \tag{17}$$

Here, $\boldsymbol{\chi}_{i,s,k-1|k-1} = \sqrt{\boldsymbol{P}_{i,k-1|k-1}}\boldsymbol{\xi}_s + \hat{\boldsymbol{x}}_{i,k-1|k-1}$, $\boldsymbol{\xi}_s$ is the s-th column vector of the n-dimensional identity matrix [22, 41].

The dynamic of $\boldsymbol{\Lambda}_{i,k}$ is modelled to ensure that the prediction of precision matrix is subject to Wishart distribution as well. Similar to [32], the predicted PDF of $\boldsymbol{\Lambda}_{i,k}$ can be written as

$$p^i(\boldsymbol{\Lambda}_{i,k|k-1}) = W(\boldsymbol{\Lambda}_{i,k}; v_{i,k|k-1}, \boldsymbol{V}_{i,k|k-1}), \tag{18}$$

where

$$v_{i,k|k-1} = \rho v_{i,k-1}, \quad \boldsymbol{V}_{i,k|k-1} = \boldsymbol{B}\boldsymbol{V}_{i,k-1}\boldsymbol{B}^{\mathrm{T}}, \tag{19}$$

and ρ is the discount parameter satisfying $0 < \rho \leqslant 1$. The parameter $\rho = 1$ corresponds to stationary variances, which means no decay of information. Smaller values of ρ increase their assumed fluctuations of time with more decay of information. To maintain the expected measurement precision matrix unchanged at the predicted step, a rational choice for $\boldsymbol{B}$ is $\boldsymbol{B} = \boldsymbol{I}_n/\sqrt{\rho}$.

To better describe the prior distribution of state $\boldsymbol{x}_k$, the KLA over the predicted PDFs from the inclusive neighbors $j \in \mathcal{J}_i$ is fused based on (5). Then, the fused PDF $\hat{\boldsymbol{x}}_{i,k|k-1}$ is Gaussian distributed as follows:

$$\bar{p}^i(\boldsymbol{x}_{k|k-1}) = N(\boldsymbol{x}_{k|k-1}; \bar{\boldsymbol{x}}_{i,k|k-1}, \bar{\boldsymbol{P}}_{i,k|k-1}), \tag{20}$$

$$\bar{\boldsymbol{P}}_{i,k|k-1}^{-1} = \sum_{j\in\mathcal{J}_i} w_{i,j,k}\boldsymbol{P}_{j,k|k-1}^{-1}, \tag{21}$$

$$\bar{\boldsymbol{P}}_{i,k|k-1}^{-1}\bar{\boldsymbol{x}}_{i,k|k-1} = \sum_{i\in\mathcal{J}_i} w_{i,j,k}\boldsymbol{P}_{j,k|k-1}^{-1}\hat{\boldsymbol{x}}_{j,k|k-1}, \tag{22}$$

where $w_{i,j,k} > 0$, $\sum_{i\in\mathcal{J}_i} w_{i,j,k} = 1$ is the weight that node i assigns to the information received from node j. Similar to (8), the corresponding information form can be written as

$$\bar{\boldsymbol{Y}}_{i,k|k-1} = \sum_{j\in\mathcal{J}_i} w_{i,j,k}\boldsymbol{Y}_{j,k|k-1}, \tag{23}$$

$$\bar{\boldsymbol{y}}_{i,k|k-1} = \sum_{i\in\mathcal{J}_i} w_{i,j,k}\hat{\boldsymbol{y}}_{j,k|k-1}, \tag{24}$$

where

$$\boldsymbol{Y}_{j,k|k-1} = \boldsymbol{P}_{j,k|k-1}^{-1}, \quad \hat{\boldsymbol{y}}_{j,k|k-1} = \boldsymbol{P}_{j,k|k-1}^{-1}\hat{\boldsymbol{x}}_{j,k|k-1}. \tag{25}$$

Such a fusion strategy provides a consistent estimate robust to unknown correlations among the set of predicted PDFs $\{p^j(\boldsymbol{x}_{k|k-1}), j \in \mathcal{J}_i\}$. However, the optimality is sacrificed to guarantee the consistency. Consequently, the KLA fusion may be too conservative especially when $\{p^j(\boldsymbol{x}_{k|k-1}), j \in \mathcal{J}_i\}$ are less correlated. It is possible that some nodes produce estimates of very low confidence. If such estimates are assigned with relatively high weight $w_{i,j,k}$, the fused estimate will become less confident. This is especially challenging for sensor networks where only one communication iteration is allowed between neighboring nodes. Therefore, the selection of $w_{i,j,k}$ is of critical importance for (23) and (24) to offer satisfactory results. The optimal $w_{i,j,k}$ can be determined by minimizing the trace of $\bar{\boldsymbol{Y}}_{i,k|k-1}^{-1}$, which can be regarded as a semidefinite programming problem (SDP) [42]. The selection of $w_{i,j,k}$ is not the focus of this paper, and the interested reader can refer to [4] and references therein for more details. For the sake of computational efficiency, in this note, a suboptimal solution presented in [43] is adopted, where $w_{i,j,k}$ is computed by

$$w_{i,j,k} = \frac{1/\mathrm{tr}(\boldsymbol{P}_{j,k|k-1})}{\sum_{j\in\mathcal{J}_i} 1/\mathrm{tr}(\boldsymbol{P}_{j,k|k-1})}. \tag{26}$$

Since the prediction steps of $\boldsymbol{x}_k$ and $\boldsymbol{\Lambda}_{i,k}$ are separable and independent, similar to (11), the VB marginal for $\boldsymbol{x}_k$ at node i can be rewritten as

$$\ln \hat{q}^i(\boldsymbol{x}_k) = \mathop{\mathrm{E}}_{\hat{q}(\boldsymbol{\Lambda}_{i,k})}\left\{\ln p(\breve{\boldsymbol{z}}_{i,k}|\boldsymbol{x}_k, \breve{\boldsymbol{\Lambda}}_{i,k})\bar{p}^i(\boldsymbol{x}_{k|k-1})\right\} + c_1, \tag{27}$$

where $\breve{\boldsymbol{\Lambda}}_{i,k} = \{\boldsymbol{\Lambda}_{j,k}, j \in \mathcal{J}_i\}$ and $\breve{\boldsymbol{z}}_{i,k} = \{\boldsymbol{z}_{j,k}, j \in \mathcal{J}_i\}$. Substituting (20) and (19) into (27), one has

$$\begin{aligned}
\ln \hat{q}^i(\boldsymbol{x}_k) &= -\frac{1}{2}\mathop{\mathrm{E}}_{\hat{q}(\boldsymbol{\Lambda}_{j,k})}\left\{\sum_{j\in\mathcal{J}_i}(\boldsymbol{z}_{j,k} - \boldsymbol{h}_j(\boldsymbol{x}_k))^{\mathrm{T}}\boldsymbol{\Lambda}_{j,k}(\boldsymbol{z}_{j,k} - \boldsymbol{h}_j(\boldsymbol{x}_k))\right\} \\
&\quad -\frac{1}{2}(\boldsymbol{x}_k - \bar{\boldsymbol{x}}_{i,k|k-1})^{\mathrm{T}}\bar{\boldsymbol{P}}_{i,k|k-1}^{-1}(\boldsymbol{x}_k - \bar{\boldsymbol{x}}_{i,k|k-1}) + c_1 \\
&= -\frac{1}{2}\sum_{j\in\mathcal{J}_i}\mathop{\mathrm{E}}_{\hat{q}(\boldsymbol{\Lambda}_{j,k})}\left\{\mathrm{tr}\left((\boldsymbol{z}_{j,k} - \boldsymbol{h}_j(\boldsymbol{x}_k))^{\mathrm{T}}\boldsymbol{\Lambda}_{j,k}(\boldsymbol{z}_{j,k} - \boldsymbol{h}_j(\boldsymbol{x}_k))\right)\right\} \\
&\quad -\frac{1}{2}\mathrm{tr}\left((\boldsymbol{x}_k - \bar{\boldsymbol{x}}_{i,k|k-1})^{\mathrm{T}}\bar{\boldsymbol{P}}_{i,k|k-1}^{-1}(\boldsymbol{x}_k - \bar{\boldsymbol{x}}_{i,k|k-1})\right) + c_1 \\
&= -\frac{1}{2}\sum_{j\in\mathcal{J}_i}\mathrm{tr}\left(\mathop{\mathrm{E}}_{\hat{q}(\boldsymbol{\Lambda}_{j,k})}\{\boldsymbol{\Lambda}_{j,k}\}(\boldsymbol{z}_{j,k} - \boldsymbol{h}_j(\boldsymbol{x}_k))(\boldsymbol{z}_{j,k} - \boldsymbol{h}_j(\boldsymbol{x}_k))^{\mathrm{T}}\right) \\
&\quad -\frac{1}{2}\mathrm{tr}\left(\bar{\boldsymbol{P}}_{i,k|k-1}^{-1}(\boldsymbol{x}_k - \bar{\boldsymbol{x}}_{i,k|k-1})(\boldsymbol{x}_k - \bar{\boldsymbol{x}}_{i,k|k-1})^{\mathrm{T}}\right) + c_1.
\end{aligned} \tag{28}$$

Here, for matrices $\boldsymbol{A}$, $\boldsymbol{B}$, and $\boldsymbol{C}$, $\mathrm{tr}(\boldsymbol{ABC}) = \mathrm{tr}(\boldsymbol{BCA})$ is exploited to obtain the result.

It can be observed from (28) that $\hat{q}^i(\boldsymbol{x}_k)$ is also Gaussian with the PDF given by

$$\hat{q}^i(\boldsymbol{x}_k) = N(\boldsymbol{x}_k; \bar{\boldsymbol{x}}_{i,k|k}, \boldsymbol{P}_{i,k|k}). \tag{29}$$

Substituting (29) into (28) and matching terms in both sides of (28), one has

$$\boldsymbol{Y}_{i,k|k} = \bar{\boldsymbol{Y}}_{i,k|k-1} + \sum_{j\in\mathcal{J}_i}\underbrace{\boldsymbol{\mathcal{H}}_{j,k}^{\mathrm{T}}\hat{\boldsymbol{\Lambda}}_{j,k}\boldsymbol{\mathcal{H}}_{j,k}}_{\boldsymbol{u}_{j,k}}, \tag{30}$$

$$\boldsymbol{y}_{i,k|k} = \bar{\boldsymbol{y}}_{i,k|k-1} + \sum_{j\in\mathcal{J}_i}\underbrace{\boldsymbol{\mathcal{H}}_{j,k}^{\mathrm{T}}\hat{\boldsymbol{\Lambda}}_{j,k}(\tilde{\boldsymbol{z}}_{j,k} + \boldsymbol{\mathcal{H}}_{j,k}\boldsymbol{x}_{j,k|k-1})}_{\boldsymbol{U}_{j,k}}, \tag{31}$$

where

$$\boldsymbol{P}_{i,k|k} = \boldsymbol{Y}_{i,k|k}^{-1}, \quad \boldsymbol{x}_{i,k|k} = \boldsymbol{Y}_{i,k|k}^{-1}\boldsymbol{y}_{i,k|k}, \tag{32}$$

$$\hat{\boldsymbol{\Lambda}}_{j,k} = \underset{\hat{q}(\boldsymbol{\Lambda}_{j,k})}{\mathrm{E}}\{\boldsymbol{\Lambda}_{j,k}\} = v_{j,k}\boldsymbol{V}_{j,k}, \quad \tilde{\boldsymbol{z}}_{j,k} = \boldsymbol{z}_{j,k} - \hat{\boldsymbol{z}}_{j,k|k-1}, \tag{33}$$

and the predicted measurement $\hat{\boldsymbol{z}}_{j,k}$ can be computed by

$$\hat{\boldsymbol{z}}_{j,k|k-1} = \frac{1}{2n}\sum_{s=1}^{2n}\boldsymbol{h}_j(\boldsymbol{\chi}_{j,s,k|k-1}). \tag{34}$$

Here, $\boldsymbol{\chi}_{j,s,k|k-1} = \sqrt{\boldsymbol{P}_{j,k|k-1}}\boldsymbol{\xi}_s + \hat{\boldsymbol{x}}_{j,k|k-1}$. By taking advantage of the statistical linear error propagation methodology [6, 44], the pseudo-measurement matrix can be calculated by

$$\boldsymbol{\mathcal{H}}_{j,k} = \boldsymbol{P}_{j,xz,k|k-1}^{\mathrm{T}}\boldsymbol{P}_{j,k|k-1}^{-1}, \tag{35}$$

where

$$\boldsymbol{P}_{j,xz,k|k-1} = \frac{1}{2n}\sum_{s=1}^{2n}\boldsymbol{\chi}_{j,s,k|k-1}\boldsymbol{h}_j^{\mathrm{T}}(\boldsymbol{\chi}_{j,s,k|k-1}) - \hat{\boldsymbol{x}}_{j,k|k-1}\hat{\boldsymbol{z}}_{j,k|k-1}^{\mathrm{T}}. \tag{36}$$

Note that the update steps (30) and (31) are almost the same as that in [18] except for the cubature rule and $\hat{\boldsymbol{\Lambda}}_{j,k}$.

Similar to (12), the VB marginal for $\boldsymbol{\Lambda}_{j,k}$ can be rewritten as

$$\ln\hat{q}^i(\boldsymbol{\Lambda}_{i,k}) = \underset{\hat{q}^i(\boldsymbol{x}_k)}{\mathrm{E}}\{\ln p(\boldsymbol{z}_{i,k}|\boldsymbol{x}_k, \boldsymbol{\Lambda}_{i,k})p^i(\boldsymbol{\Lambda}_{i,k|k-1})\} + c_2. \tag{37}$$

By exploiting (19) in (37), one has

$$\begin{aligned}
\ln\hat{q}^i(\boldsymbol{\Lambda}_{i,k}) =\ & \frac{v_{i,k|k-1}-m}{2}\ln|\boldsymbol{\Lambda}_{i,k}| - \frac{1}{2}\mathrm{tr}\left(\boldsymbol{V}_{i,k|k-1}^{-1}\boldsymbol{\Lambda}_{i,k}\right) \\
& -\frac{1}{2}\underset{\hat{q}^i(\boldsymbol{x}_k)}{\mathrm{E}}\left\{(\boldsymbol{z}_{i,k}-\boldsymbol{h}_i(\boldsymbol{x}_k))^{\mathrm{T}}\boldsymbol{\Lambda}_{i,k}(\boldsymbol{z}_{i,k}-\boldsymbol{h}_i(\boldsymbol{x}_k))\right\} + c_2 \\
=\ & \frac{v_{i,k|k-1}-m}{2}\ln|\boldsymbol{\Lambda}_{i,k}| - \frac{1}{2}\mathrm{tr}\left(\boldsymbol{V}_{i,k|k-1}^{-1}\boldsymbol{\Lambda}_{i,k}\right) \\
& -\frac{1}{2}\underset{\hat{q}^i(\boldsymbol{x}_k)}{\mathrm{E}}\left\{\mathrm{tr}\left((\boldsymbol{z}_{i,k}-\boldsymbol{h}_i(\boldsymbol{x}_k))^{\mathrm{T}}\boldsymbol{\Lambda}_{i,k}(\boldsymbol{z}_{i,k}-\boldsymbol{h}_i(\boldsymbol{x}_k))\right)\right\} + c_2 \\
=\ & \frac{v_{i,k|k-1}-m}{2}\ln|\boldsymbol{\Lambda}_{i,k}| - \frac{1}{2}\mathrm{tr}\Big(\boldsymbol{V}_{i,k|k-1}^{-1}\boldsymbol{\Lambda}_{i,k} \\
& + \underset{\hat{q}^i(\boldsymbol{x}_k)}{\mathrm{E}}\left\{(\boldsymbol{z}_{i,k}-\boldsymbol{h}_i(\boldsymbol{x}_k))(\boldsymbol{z}_{i,k}-\boldsymbol{h}_i(\boldsymbol{x}_k))^{\mathrm{T}}\boldsymbol{\Lambda}_{i,k}\right\}\Big) + c_2.
\end{aligned} \tag{38}$$

It can be observed from (38) that $\hat{q}^i(\boldsymbol{\Lambda}_{i,k})$ also follows Wishart distribution given by

$$\hat{q}^i(\boldsymbol{\Lambda}_{i,k}) = W(\boldsymbol{\Lambda}_{i,k}; v_{i,k}, \boldsymbol{V}_{i,k}). \tag{39}$$

Substituting (39) into (38) and matching terms in both sides of (38), one has

$$v_{i,k} = vv_{i,k|k-1} + 1, \tag{40}$$

$$\boldsymbol{V}_{i,k}^{-1} = \boldsymbol{V}_{i,k|k-1}^{-1} + \int(\boldsymbol{z}_{i,k}-\boldsymbol{h}_i(\boldsymbol{x}_k))(\boldsymbol{z}_{i,k}-\boldsymbol{h}_i(\boldsymbol{x}_k))^{\mathrm{T}}N(\boldsymbol{x}_k; \hat{\boldsymbol{x}}_{i,k|k}, \boldsymbol{P}_{i,k|k})\mathrm{d}\boldsymbol{x}. \tag{41}$$

The integral in (41) can be approximated by exploiting the cubature quadrature rule as follows:

$$\begin{aligned}
\boldsymbol{P}_{i,zz} &= \int(\boldsymbol{z}_{i,k}-\boldsymbol{h}_i(\boldsymbol{x}_k))(\boldsymbol{z}_{i,k}-\boldsymbol{h}_i(\boldsymbol{x}_k))^{\mathrm{T}}N(\boldsymbol{x}_k; \hat{\boldsymbol{x}}_{i,k|k}, \boldsymbol{P}_{i,k|k})\mathrm{d}\boldsymbol{x} \\
&= \frac{1}{2n}\sum_{s=1}^{2n}(\boldsymbol{z}_{i,k}-\boldsymbol{h}_i(\boldsymbol{\chi}_{i,s,k}))(\boldsymbol{z}_{i,k}-\boldsymbol{h}_i(\boldsymbol{\chi}_{i,s,k}))^{\mathrm{T}},
\end{aligned} \tag{42}$$

where $\boldsymbol{\chi}_{i,s,k} = \sqrt{\boldsymbol{P}_{i,k|k}}\boldsymbol{\xi}_s + \boldsymbol{x}_{i,k|k}$.

The above steps give one cycle of the proposed VB-DACIF, which can operate recursively with new measurements available. Details of the proposed VB-DACIF are summarized in Algorithm 1.

Algorithm 1 VB-DACIF implemented at node i

Input: The posterior estimates $\hat{\boldsymbol{x}}_{i,k-1|k-1}$, $\boldsymbol{P}_{i,k-1|k-1}$, $v_{i,k-1}$, and $\boldsymbol{V}_{i,k-1}$ at time instant $k-1$.

Time update:

Compute the predicted $\hat{\boldsymbol{x}}_{i,k|k-1}$ and $\boldsymbol{P}_{i,k|k-1}$ via Eqs. (16) and (17);

Compute the predicted $v_{i,k|k-1}$, $\boldsymbol{V}_{i,k|k-1}$ via Eq. (19);

Fuse the predicted estimates obtained from $j \in \mathcal{J}_i$ via Eqs. (20)–(25);

Recursive measurement update:

Initialization: set $\boldsymbol{Y}^0_{i,k|k} = \bar{\boldsymbol{Y}}_{i,k|k-1}$, $\hat{\boldsymbol{y}}^0_{i,k|k} = \bar{\boldsymbol{y}}_{i,k|k-1}$, $v_{i,k} = v_{i,k|k-1} + 1$, $\boldsymbol{V}^0_{i,k} = \boldsymbol{V}_{i,k|k-1}$;

VB approximation;

for $l = 1 : L$ **do**

Compute the local measurement information contribution: if $i \in \mathcal{S}$, $\hat{\boldsymbol{\Lambda}}^l_{i,k} = v_{i,k}\boldsymbol{V}^{l-1}_{i,k}$, $\boldsymbol{u}^l_{i,k} = \boldsymbol{\mathcal{H}}^{\mathrm{T}}_{i,k}\hat{\boldsymbol{\Lambda}}_{i,k}(\tilde{\boldsymbol{z}}_{i,k} - \boldsymbol{\mathcal{H}}_{i,k}\boldsymbol{x}_{i,k|k-1})$, $\boldsymbol{U}^l_{i,k} = \boldsymbol{\mathcal{H}}^{\mathrm{T}}_{i,k}\hat{\boldsymbol{\Lambda}}_{i,k}\boldsymbol{\mathcal{H}}_{i,k}$; if $i \in \mathcal{C}$, $\boldsymbol{u}^l_{i,k} = \boldsymbol{0}$, $\boldsymbol{U}^l_{i,k} = \boldsymbol{0}$;

Update the local posterior estimate

$$\boldsymbol{Y}^l_{i,k|k} = \bar{\boldsymbol{Y}}_{i,k|k-1} + \sum_{j\in\mathcal{J}_i}\boldsymbol{U}^l_{j,k}, \quad \boldsymbol{y}^l_{i,k|k} = \bar{\boldsymbol{y}}_{i,k|k-1} + \sum_{j\in\mathcal{J}_i}\boldsymbol{u}^l_{j,k};$$
$$\boldsymbol{P}^l_{i,k|k} = (\boldsymbol{Y}^l_{i,k|k})^{-1}, \quad \hat{\boldsymbol{x}}^l_{i,k|k} = (\boldsymbol{Y}^l_{i,k|k})^{-1}\boldsymbol{y}^l_{i,k|k};$$

Generate cubature points $\boldsymbol{\chi}^l_{i,s,k} = \sqrt{\boldsymbol{P}^l_{i,k|k}}\boldsymbol{\xi}_s + \boldsymbol{x}^l_{i,k|k}$ and update the scale matrix by

$$(\boldsymbol{V}^l_{i,k})^{-1} = (\boldsymbol{V}^{l-1}_{i,k})^{-1} + \frac{1}{2n}\sum_{s=1}^{2n}(\boldsymbol{z}_{i,k} - \boldsymbol{h}_i(\boldsymbol{\chi}^l_{i,s,k}))(\boldsymbol{z}_{i,k} - \boldsymbol{h}_i(\boldsymbol{\chi}^l_{i,s,k}))^{\mathrm{T}};$$

end for

Set $\hat{\boldsymbol{y}}_{i,k|k} = \hat{\boldsymbol{y}}^L_{i,k|k}$, $\boldsymbol{Y}_{i,k|k} = \boldsymbol{Y}^L_{i,k|k}$, $\hat{\boldsymbol{x}}_{i,k|k} = (\boldsymbol{Y}^L_{i,k|k})^{-1}\hat{\boldsymbol{y}}^L_{i,k|k}$, $\boldsymbol{P}_{i,k|k} = (\boldsymbol{Y}^L_{i,k|k})^{-1}$, $\boldsymbol{V}_{i,k} = \boldsymbol{V}^L_{i,k}$;

Output: The updated estimates $\hat{\boldsymbol{x}}_{i,k|k}$, $\boldsymbol{P}_{i,k|k}$, $v_{i,k}$, and $\boldsymbol{V}_{i,k}$ at time instant k.

4 Analysis of the Cramér-Rao lower bound

The posterior Cramér-Rao lower bound (PCRLB), defined as the inverse of the Fisher information matrix (FIM), can be utilized to indicate the best possible performance that an algorithm is able to achieve [45, 46]. In [46], a framework to derive the filtering PCRLB has been proposed for discrete-time nonlinear dynamic systems. Here, PCRLB is used to evaluate the performance of the proposed VB-DACIF.

Proposition 1. The Fisher information submatrix $\boldsymbol{J}(\boldsymbol{x}_k)$ for estimating the state $\boldsymbol{x}_k$ follows the recursion

$$\boldsymbol{J}(\boldsymbol{x}_{k+1}) = \left[\boldsymbol{Q}_k + \boldsymbol{F}_k\boldsymbol{J}^{-1}(\boldsymbol{x}_k)\boldsymbol{F}^{\mathrm{T}}_k\right]^{-1} + \sum_{i\in\mathcal{S}}\boldsymbol{H}^{\mathrm{T}}_{i,k+1}\boldsymbol{R}^{-1}_{i,k+1}\boldsymbol{H}_{i,k+1}, \tag{43}$$

where $\boldsymbol{F}_k = \frac{\partial \boldsymbol{f}(\boldsymbol{x}_k)}{\partial \boldsymbol{x}_k}$ and $\boldsymbol{H}_{i,k+1} = \frac{\partial \boldsymbol{H}(\boldsymbol{x}_{k+1})}{\partial \boldsymbol{x}_{k+1}}$ are Jacobian matrices with respect to $\boldsymbol{x}_k$ and $\boldsymbol{x}_{k+1}$, respectively.

Proof. According to [46], $\boldsymbol{J}(\boldsymbol{x}_k)$ can be recursively computed by

$$\boldsymbol{J}(\boldsymbol{x}_{k+1}) = \boldsymbol{D}^{22}_k - \boldsymbol{D}^{21}_k\left[\boldsymbol{J}(\boldsymbol{x}_k) + \boldsymbol{D}^{11}_k\right]^{-1}\boldsymbol{D}^{12}_k \tag{44}$$

with

$$\boldsymbol{D}^{11}_k = \mathrm{E}\{-\Delta^{\boldsymbol{x}_k}_{\boldsymbol{x}_k}\log p(\boldsymbol{x}_{k+1}|\boldsymbol{x}_k)\}, \tag{45}$$
$$\boldsymbol{D}^{12}_k = (\boldsymbol{D}^{21}_k)^{\mathrm{T}} = \mathrm{E}\{-\Delta^{\boldsymbol{x}_k}_{\boldsymbol{x}_{k+1}}\log p(\boldsymbol{x}_{k+1}|\boldsymbol{x}_k)\}, \tag{46}$$
$$\boldsymbol{D}^{22}_k = \mathrm{E}\{-\Delta^{\boldsymbol{x}_{k+1}}_{\boldsymbol{x}_{k+1}}\log P(\boldsymbol{x}_{k+1}|\boldsymbol{x}_k)\} + \sum_{i\in\mathcal{S}}\mathrm{E}\{-\Delta^{\boldsymbol{x}_{k+1}}_{\boldsymbol{x}_{k+1}}\log p(\boldsymbol{z}_{i,k+1}|\boldsymbol{x}_{k+1})\}, \tag{47}$$

where $\Delta^{\boldsymbol{y}}_{\boldsymbol{x}} = \bigtriangledown_{\boldsymbol{x}}\bigtriangledown^{\mathrm{T}}_{\boldsymbol{y}}$ and $\bigtriangledown$ is the gradient operator. Since $\boldsymbol{w}_k$ and $\boldsymbol{v}_{i,k}$ are both Gaussian, one has

$$-\log p(\boldsymbol{x}_{k+1}|\boldsymbol{x}_k) = \frac{1}{2}\left[\boldsymbol{x}_{k+1} - \boldsymbol{f}(\boldsymbol{x}_k)\right]^{\mathrm{T}}\boldsymbol{Q}^{-1}_k\left[\boldsymbol{x}_{k+1} - \boldsymbol{f}(\boldsymbol{x}_k)\right] + c_3, \tag{48}$$

$$-\log p(\boldsymbol{z}_{i,k+1}|\boldsymbol{x}_{k+1}) = \frac{1}{2}\left[\boldsymbol{z}_{i,k+1} - \boldsymbol{h}_i(\boldsymbol{x}_{k+1})\right]^{\mathrm{T}}\boldsymbol{R}^{-1}_{i,k}\left[\boldsymbol{z}_{i,k+1} - \boldsymbol{h}_i(\boldsymbol{x}_{k+1})\right] + c_4. \tag{49}$$

Here, c_3 and c_4 are constants independent of $\boldsymbol{x}_k$ and $\boldsymbol{x}_{k+1}$. By exploiting (48) and (49) in (45)–(47), one has

$$\begin{cases} \boldsymbol{D}_k^{11} = \boldsymbol{F}_k^{\mathrm{T}} \boldsymbol{Q}_k^{-1} \boldsymbol{F}_k; \\ \boldsymbol{D}_k^{12} = (\boldsymbol{D}_k^{21})^{\mathrm{T}} = -\boldsymbol{F}_k^{\mathrm{T}} \boldsymbol{Q}_k^{-1}; \\ \boldsymbol{D}_k^{22} = \boldsymbol{Q}_k^{-1} + \sum\limits_{i \in \mathcal{S}} \boldsymbol{H}_{i,k+1}^{\mathrm{T}} \boldsymbol{R}_{i,k+1}^{-1} \boldsymbol{H}_{i,k+1}. \end{cases} \tag{50}$$

By substituting (50) into (44) and exploiting the matrix inversion lemma [47], the desired result in (43) is obtained.

5 Simulations

Performance of the proposed VB-DACIF is validated in this section by considering a typical air-traffic control scenario. The target of interest moves in a horizontal plane and conducts a maneuvering turn at a constant turn rate Ω, where Ω is known. The dynamic of target is modeled by the following equation [22]:

$$\boldsymbol{x}_k = \begin{bmatrix} 1 & \dfrac{\sin \Omega T}{\Omega} & 0 & \dfrac{\cos \Omega T - 1}{\Omega} & 0 \\ 0 & \cos \Omega T & 0 & -\sin \Omega & 0 \\ 0 & \dfrac{1 - \cos \Omega T}{\Omega} & 1 & \dfrac{\sin \Omega T}{\Omega} & 0 \\ 0 & \sin \Omega T & 0 & \cos \Omega T & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix} \boldsymbol{x}_{k-1} + \boldsymbol{w}_{k-1}, \tag{51}$$

where $\boldsymbol{x}_k = [x_k, \dot{x}_k, y_k, \dot{y}_k, \Omega]^{\mathrm{T}}$ is the target state, $\boldsymbol{w}_{k-1} \sim N(\boldsymbol{0}, \boldsymbol{Q}_{k-1})$ is the Gaussian process noise, in which the covariance is given by

$$\boldsymbol{Q}_{k-1} = 0.1 \begin{bmatrix} \dfrac{T^3}{3} & \dfrac{T^2}{2} & 0 & 0 & 0 \\ \dfrac{T^2}{2} & T & 0 & 0 & 0 \\ 0 & 0 & \dfrac{T^3}{3} & \dfrac{T^2}{2} & 0 \\ 0 & 0 & \dfrac{T^2}{2} & T & 0 \\ 0 & 0 & 0 & 0 & 10^{-5} T \end{bmatrix},$$

and $T = 1$ s denotes the sampling interval between consecutive time instants. The measurement of each node is comprised of range and bearing of the interested target and is modelled by

$$\boldsymbol{z}_{i,k} = \begin{bmatrix} \sqrt{(x_k - x_i)^2 + (y_k - y_i)^2} \\ \arctan\left(\dfrac{y_k - y_i}{x_k - x_i}\right) \end{bmatrix} + \boldsymbol{v}_{i,k}, \tag{52}$$

where (x_i, y_i) is the position of the i-th sensor node, $\boldsymbol{v}_{i,k} \sim N(\boldsymbol{0}, \boldsymbol{R}_{i,k})$ denotes the Gaussian measurement noise. The considered sensor network consists of 4 sensor nodes and 8 communication nodes, the communication topology of which is depicted by Figure 1. Define a measurement noise covariance

$$\boldsymbol{R}_0 = \begin{bmatrix} \sigma_r^2 & \sigma_{r\theta} \\ \sigma_{\theta r} & \sigma_\theta^2 \end{bmatrix},$$

where $\sigma_r = 10$ m, $\sigma_\theta = 0.2°$, $\sigma_{r\theta} = \sigma_{\theta r} = 0.034$ m°, which is the same as that in [32]. The true measurement noise covariances are modelled by

$$\boldsymbol{R}_{i,k} = \begin{cases} \boldsymbol{R}_0, & \text{if } k < k_0; \\ \begin{bmatrix} (\alpha \sigma_r)^2 & \sigma_{r\theta} \\ \sigma_{\theta r} & (\alpha \sigma_\theta)^2 \end{bmatrix}, & \text{else}, \end{cases} \tag{53}$$

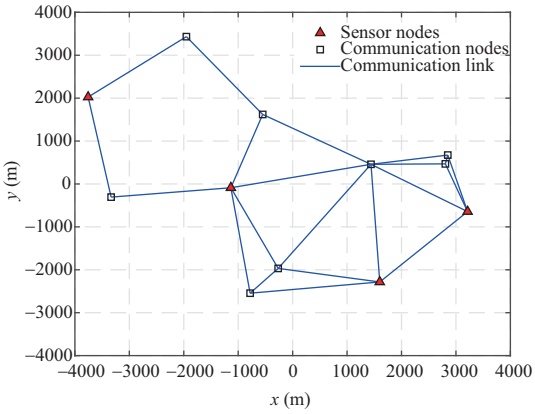

Figure 1 (Color online) Communication topology of the sensor network.

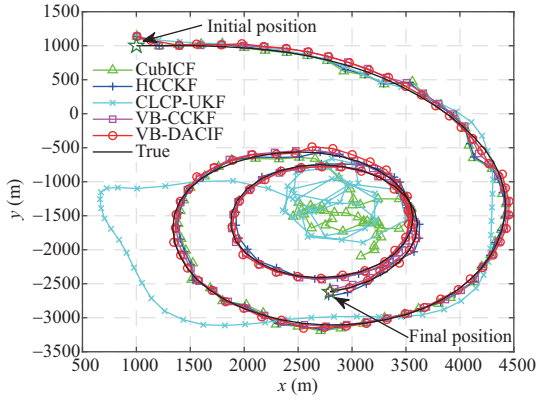

Figure 2 (Color online) True and estimated trajectories of the maneuvering target.

where the time instants k_0 for the sensor nodes are 10, 16, 20, and 25, respectively, and the corresponding scale parameters α are, respectively, 10, 8, 5, and 9.

The true trajectory is generated according to (51) with initial state $\boldsymbol{x}_0 = [1000\ \text{m},\ 300\ \text{m/s},\ 1000\ \text{m},\ 0\ \text{m/s},\ -3^\circ/\text{s}]^{\mathrm{T}}$. Initial estimate for each sensor is randomly chosen from $N(\boldsymbol{x}_0, \boldsymbol{P}_0)$ with $\boldsymbol{P}_0 = \mathrm{diag}(100\ \text{m}^2,\ 25\ \text{m}^2/\text{s}^2,\ 100\ \text{m}^2,\ 25\ \text{m}^2/\text{s}^2,\ 100\ \text{mrad}^2/\text{s}^2)$. The total time span is $K = 100$. For a fair comparison, a total number of $M = 100$ independent Monte Carlo experiments are conducted.

The proposed VB-DACIF is compared with the CubICF [17], HCCKF [24], and CLCP-UKF [19], which are not conscious of the measurement noise covariance change at k_0. The suggested scaling parameter $\kappa = 3 - n$ for CLCP-UKF is negative in the considered scenario, which often causes the algorithm to fail to operate and hence $\kappa = 1$ is used here to avoid divergence [41]. Since the proposed VB-DACIF only requires a single communication iteration, VB-CCKF presented in [33] with only one consensus iteration is chosen for comparison. Similar to [32], initial parameters for both VB-CCKF and VB-DACIF are $v_0 = 1$ and $\boldsymbol{V}_0 = 2\boldsymbol{R}_0^{-1}$. The discount factor is set to $\rho = 0.9$ for all VB filters, unless stated otherwise. PCRLB is exploited as the benchmark to check the effectiveness of all the algorithms under consideration. The root mean square error (RMSE) [22] and disagreement of estimates (DoE) [7] are introduced to evaluate the estimation performance of different algorithms.

The true and estimated tracks by different algorithms in a certain simulation run are shown in Figure 2. It can be seen that CLCP-UKF and CubICF can hardly track the target, while other algorithms have satisfactory tracking results. Figures 3–5 show the RMSE in position, velocity and turn rate averaged over all simulation runs, in which the results of VB-CCKF and VB-DACIF are obtained with only one VB iteration. As is presented in Figure 3, CLCP-UKF diverges after a short period of time. CubICF and HCCKF have good results before $k = 50$, but result in large errors when $k > 60$ due to unawareness of change of the measurement noise covariance. Compared with the above algorithms, VB-CCKF has relatively small RMSE, but the value is much larger than that of VB-DACIF, which is closer to that of the PCRLB. As for velocity and turn rate shown in Figures 4 and 5, the results are similar to that in Figure 3. It should be noted that before divergence, CubICF and HCCKF have relatively smaller RMSE than VB-CCKF. The reason may be that VB-CCKF does not have accurate estimation of noise covariance since only one consensus iteration and VB iteration are conducted. However, this phenomenon does not occur when using the proposed VB-DACIF. During the simulation time, VB-DACIF always has much lower RMSE in position, velocity or turn rate, and is closer to the PCRLB, which indicates relatively higher estimation accuracy.

Disagreement of estimates from different sensors are shown in Figure 6. It can be seen that estimates from different sensors obtained by CLCP-UKF are quite far away from each other, which finally leads to divergence of DoE. CubICF, VB-CCKF, and HCCKF have relatively lower DoE compared with CLCP-UKF, but have much larger DoE than VB-DACIF, which indicates that each node achieves better

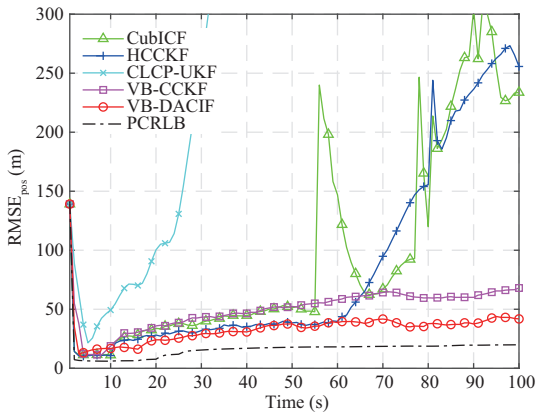

Figure 3 (Color online) Position RMSE for different algorithms over time.

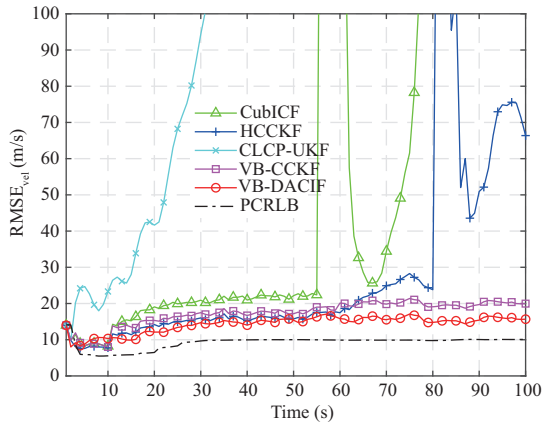

Figure 4 (Color online) Velocity RMSE for different algorithms over time.

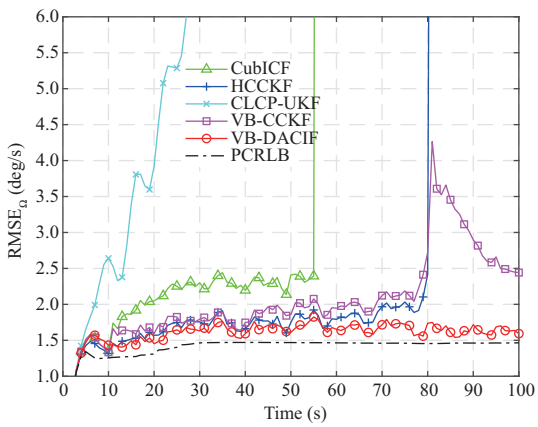

Figure 5 (Color online) Turn rate RMSE for different algorithms over time.

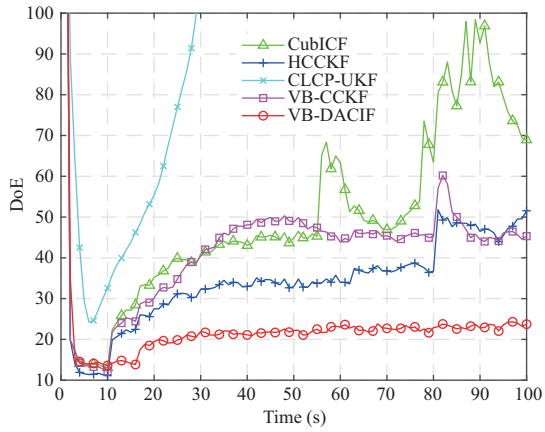

Figure 6 (Color online) DoEs for different algorithms.

Table 1 Averaged RMSEs for different algorithms

	Position (m)	Velocity (m/s)	Turn rate (deg/s)
CubICF	131.473	206.281	18.004
HCCKF	120.344	41.010	7.200
CLCP-UKF	123.210	908.755	59.644
VB-CCKF	64.694	34.689	3.281
VB-DACIF	**36.025**	**14.470**	**1.627**

consensus on estimates with exploitation of VB-DACIF.

Table 1 gives RMSEs in position, velocity and turn rate estimated by different algorithms with only one VB iteration. The result is similar to that shown in Figures 3–5. It can be seen that the VB-based algorithms exhibit smaller RMSEs than traditional distributed algorithms, which indicates that estimates obtained by VB-based algorithms are of higher accuracy. In addition, the proposed VB-DACIF shows almost half the RMSEs of VB-CCKF, which suggests that the performance of the proposed VB-DACIF is greatly improved by incorporating the diffusion strategy.

To further compare the VB-based algorithms, Tables 2–4 present the averaged position, velocity and turn rate RMSEs with varying VB iterations, and the averaged DoEs with varying VB iterations are shown in Table 5. It can be observed that the proposed VB-DACIF always has lower RMSE and DoE than the existing VB-CCKF regardless of VB iterations. Furthermore, for the proposed VB-DACIF,

Table 2 Averaged position RMSEs (m) with different VB iterations

Algorithm	$L=1$	$L=2$	$L=3$	$L=4$	$L=5$
VB-CCKF	64.694	62.015	63.198	64.271	64.871
VB-DACIF	**36.025**	**35.739**	**37.874**	**38.244**	**38.486**

Table 3 Averaged velocity RMSEs (m/s) with different VB iterations

Algorithm	$L=1$	$L=2$	$L=3$	$L=4$	$L=5$
VB-CCKF	34.689	34.490	34.388	35.356	36.170
VB-DACIF	**14.470**	**14.386**	**14.214**	**14.392**	**14.645**

Table 4 Averaged turn rate RMSEs (deg/s) with different VB iterations

Algorithm	$L=1$	$L=2$	$L=3$	$L=4$	$L=5$
VB-CCKF	3.281	3.214	3.121	3.289	3.312
VB-DACIF	**1.627**	**1.622**	**1.617**	**1.620**	**1.621**

Table 5 DoEs with different VB iterations

Algorithm	$L=1$	$L=2$	$L=3$	$L=4$	$L=5$
VB-CCKF	40.725	36.120	31.026	30.579	30.192
VB-DACIF	**21.066**	**20.310**	**20.086**	**19.984**	**19.927**

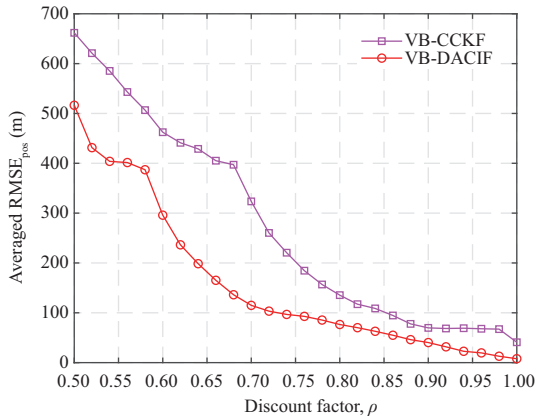

Figure 7 (Color online) Comparison of averaged position RMSE with different ρ.

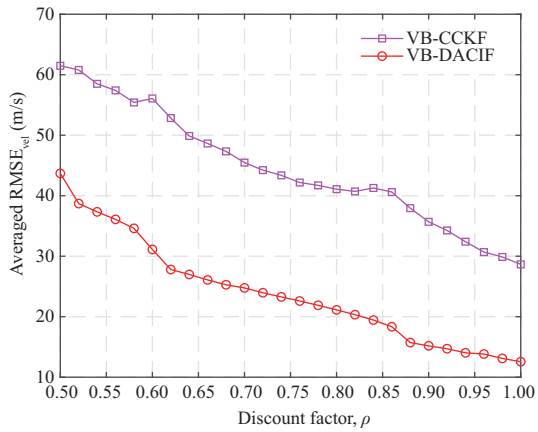

Figure 8 (Color online) Comparison of averaged velocity RMSE with different ρ.

the position RMSE no longer decreases when $L=2$, while the velocity and turn rate RMSEs no longer decrease when $L=3$. As the number of VB iterations increases, the obtained DoE decreases all the time. Therefore, $L=3$ is chosen for later discussion and comparison.

To investigate the influence of discount factor ρ on the estimation performance of VB-based algorithms, ρ is set to vary within [0.5, 1], at the increment of 0.02 in the simulations below. The resultant averaged position, velocity and turn rate RMSEs with different ρ are presented in Figures 7–9. It can be seen that both algorithms achieve better performance with a larger ρ. From the perspective of RMSE, a satisfactory estimation performance can be fulfilled when ρ falls within [0.9, 1]. In addition, for a certain ρ, the proposed VB-DACIF always gets lower RMSEs than VB-CCKF, which is similar to the above analysis with $\rho=0.9$. The averaged DoEs with different ρ are shown in Figure 10. With increase of discount factor ρ, a lower DoE is obtained by both algorithms, which indicates that the estimate obtained by each node becomes closer to each other and a better consensus on estimate is achieved. Compared with the existing VB-CCKF, the proposed VB-DACIF has relatively lower averaged DoE, which is about half of that of VB-CCKF.

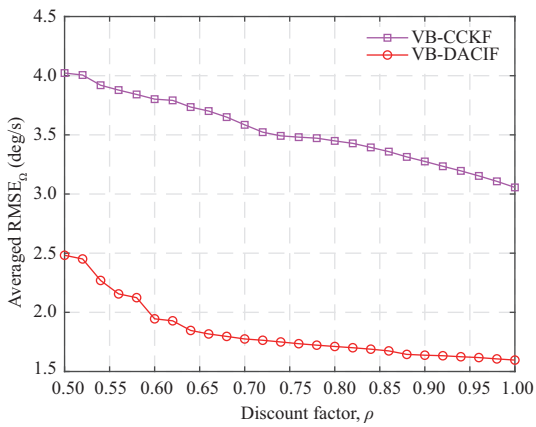

Figure 9 (Color online) Comparison of averaged turn rate RMSE with different ρ.

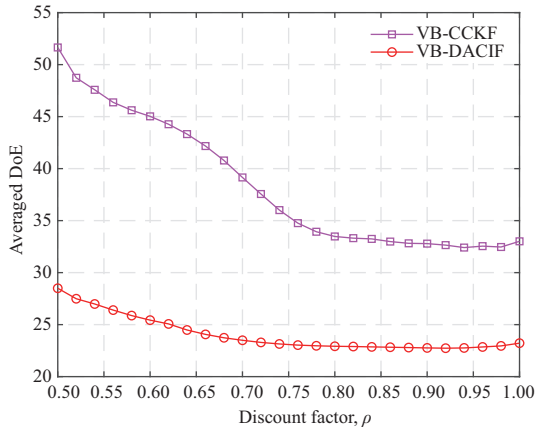

Figure 10 (Color online) Comparison of averaged DoE with different ρ.

6 Conclusion

To tackle the nonlinear DSE problems with measurement uncertainty, a VB-DACIF is proposed. For each node in the sensor network, according to KLD-based fusion strategy, KLA over the local predicted PDFs received from its inclusive neighbors are integrated to better characterize the state of interest. Then, the free form VB approximation is exploited to jointly estimate the state and measurement noise statistics. During the filtering procedure, nonlinearity of the system is handled by CKF. The estimation performance is further evaluated by the posterior Cramér-Rao lower bound. Simulations with a maneuvering target tracking scenario indicate that the proposed VB-DACIF outperforms VB-CCKF and other existing distributed estimation algorithms. In future research, the problem of distributed nonlinear filtering with unknown non-Gaussian process or measurement noise will be further investigated, and the maximum correntropy criterion may be a feasible solution.

Acknowledgements This work was supported by National Natural Science Foundation of China (Grant Nos. 61790550, 91538201, 61531020, 61671463). The authors give their sincere thanks to the anonymous reviewers for their constructive comments of the manuscripts.

References

1 Olfati-Saber R, Fax J A, Murray R M. Consensus and cooperation in networked multi-agent systems. Proc IEEE, 2007, 95: 215–233
2 Kamal A T, Farrell J A, Roy-Chowdhury A K. Information weighted consensus filters and their application in distributed camera networks. IEEE Trans Autom Control, 2013, 58: 3112–3125
3 Battistelli G, Chisci L, Mugnai G, et al. Consensus-based linear and nonlinear filtering. IEEE Trans Autom Control, 2015, 60: 1410–1415
4 Wang S C, Ren W. On the convergence conditions of distributed dynamic state estimation using sensor networks: a unified framework. IEEE Trans Control Syst Technol, 2018, 26: 1300–1316
5 Chandra K P B, Gu D W, Postlethwaite I. Square root cubature information filter. IEEE Sens J, 2013, 13: 750–758
6 Lee D J. Nonlinear estimation and multiple sensor fusion using unscented information filtering. IEEE Signal Process Lett, 2008, 15: 861–864
7 Olfati-Saber R, Murray R M. Consensus problems in networks of agents with switching topology and time-delays. IEEE Trans Autom Control, 2004, 49: 1520–1533
8 Talebi S P, Werner S. Distributed Kalman filtering and control through embedded average consensus information fusion. IEEE Trans Autom Control, 2019, 64: 4396–4403
9 Song B, Kamal A T, Soto C, et al. Tracking and activity recognition through consensus in distributed camera networks. IEEE Trans Image Process, 2010, 19: 2564–2579
10 Kwon C, Hwang I. Sensing-based distributed state estimation for cooperative multiagent systems. IEEE Trans Autom Control, 2019, 64: 2368–2382
11 Olfati-Saber R. Kalman-consensus filter: optimality, stability, and performance. In: Proceedings of the 48th IEEE Conference on Decision and Control (CDC), Shanghai, 2009. 7036–7042
12 Li W L, Jia Y M. Distributed consensus filtering for discrete-time nonlinear systems with non-Gaussian noise. Signal

Process, 2012, 92: 2464–2470
13 Hu C, Lin H S, Li Z H, et al. Kullback-Leibler divergence based distributed cubature Kalman filter and its application in cooperative space object tracking. Entropy, 2018, 20: 116
14 Cattivelli F S, Lopes C G, Sayed A H. Diffusion strategies for distributed Kalman filtering: formulation and performance analysis. In: Proceedings of the 1st IAPR Workshop on Cognitive Information Processing, 2008. 36–41
15 Hu J W, Xie L H, Zhang C S. Diffusion Kalman filtering based on covariance intersection. IEEE Trans Signal Process, 2012, 60: 891–902
16 Kamal A T, Bappy J H, Farrell J A, et al. Distributed multi-target tracking and data association in vision networks. IEEE Trans Pattern Anal Mach Intell, 2016, 38: 1397–1410
17 Jia B, Pham K D, Blasch E, et al. Cooperative space object tracking using space-based optical sensors via consensus-based filters. IEEE Trans Aerosp Electron Syst, 2016, 52: 1908–1936
18 Wang S C, Lyu Y, Ren W. Unscented-transformation-based distributed nonlinear state estimation: algorithm, analysis, and experiments. IEEE Trans Control Syst Technol, 2019, 27: 2016–2029
19 Battistelli G, Chisci L, Fantacci C. Parallel consensus on likelihoods and priors for networked nonlinear filtering. IEEE Signal Process Lett, 2014, 21: 787–791
20 Mohammadi A, Asif A. Distributed consensus + innovation particle filtering for bearing/range tracking with communication constraints. IEEE Trans Signal Process, 2015, 63: 620–635
21 Hlinka O, Hlawatsch F, Djuric P M. Consensus-based distributed particle filtering with distributed proposal adaptation. IEEE Trans Signal Process, 2014, 62: 3029–3041
22 Arasaratnam I, Haykin S. Cubature Kalman filters. IEEE Trans Autom Control, 2009, 54: 1254–1269
23 He S M, Shin H, Xu S Y, et al. Distributed estimation over a low-cost sensor network: a review of state-of-the-art. Inf Fusion, 2020, 54: 21–43
24 Chen Q, Yin C, Zhou J, et al. Hybrid consensus-based cubature Kalman filtering for distributed state estimation in sensor networks. IEEE Sens J, 2018, 18: 4561–4569
25 Chen Q, Wang W C, Yin C, et al. Distributed cubature information filtering based on weighted average consensus. Neurocomputing, 2017, 243: 115–124
26 Mehra R. Approaches to adaptive filtering. IEEE Trans Autom Control, 1972, 17: 693–698
27 Maybeck P S. Stochastic Models, Estimation, and Control. Orlando: Academic Press, 1982
28 Li X R, Bar-Shalom Y. Recursive multiple model approach to noise identification. IEEE Trans Aerosp Electron Syst, 1994, 30: 671–684
29 Storvik G. Particle filters for state-space models with the presence of unknown static parameters. IEEE Trans Signal Process, 2002, 50: 281–289
30 Sarkka S, Nummenmaa A. Recursive noise adaptive kalman filtering by variational Bayesian approximations. IEEE Trans Autom Control, 2009, 54: 596–600
31 Sarkka S, Hartikainen J. Non-linear noise adaptive Kalman filtering via variational Bayes. In: Proceedings of IEEE International Workshop on Machine Learning for Signal Processing (MLSP), 2013
32 Dong P, Jing Z L, Leung H, et al. Variational Bayesian adaptive cubature information filter based on wishart distribution. IEEE Trans Autom Control, 2017, 62: 6051–6057
33 Shen K, Jing Z L, Dong P. A consensus nonlinear filter with measurement uncertainty in distributed sensor networks. IEEE Signal Process Lett, 2017, 24: 1631–1635
34 Battistelli G, Chisci L. Kullback-Leibler average, consensus on probability densities, and distributed state estimation with guaranteed stability. Automatica, 2014, 50: 707–718
35 Bishop C M. Pattern Recognition and Machine Learning. Berlin: Springer, 2006
36 Akaike H. Information theory and an extension of the maximum likelihood principle. Berlin: Springer, 1998
37 Hurley M B. An information theoretic justification for covariance intersection and its generalization. In: Proceedings of the 5th International Conference on Information Fusion, 2002. 505–511
38 Julier S, Uhlmann J K. General decentralized data fusion with covariance intersection. In: Handbook of Multisensor Data Fusion. Boca Raton: CRC Press, 2017. 339–364
39 Wang B L, Yi W, Hoseinnezhad R, et al. Distributed fusion with multi-bernoulli filter based on generalized covariance intersection. IEEE Trans Signal Process, 2017, 65: 242–255
40 Julier S J, Uhlmann J K. A non-divergent estimation algorithm in the presence of unknown correlations. In: Proceedings of American Control Conference, 1997, 4: 2369–2373
41 Jia B, Xin M, Cheng Y. High-degree cubature Kalman filter. Automatica, 2013, 49: 510–518
42 Boyd S, Vandenberghe L. Convex Optimization. Cambridge: Cambridge University Press, 2004
43 Niehsen W. Information fusion based on fast covariance intersection filtering. In: Proceedings of the 5th International Conference on Information Fusion, 2002, 2: 901–904
44 Liu G L, Worgotter F, Markelic I. Square-root sigma-point information filtering. IEEE Trans Autom Control, 2012, 57: 2945–2950
45 Leong P H, Arulampalam S, Lamahewa T A, et al. A Gaussian-sum based cubature Kalman filter for bearings-only tracking. IEEE Trans Aerosp Electron Syst, 2013, 49: 1161–1176
46 Tichavsky P, Muravchik C H, Nehorai A. Posterior Cramer-Rao bounds for discrete-time nonlinear filtering. IEEE Trans Signal Process, 1998, 46: 1386–1396
47 Golub G H, van Loan C F. Matrix Computations. 4th ed. Baltimore: Johns Hopkins University Press, 2012