



Article Variational Bayesian Algorithms for Maneuvering Target Tracking with Nonlinear Measurements in Sensor Networks

Yumei Hu^{1,*}, Quan Pan^{2,3,*}, Bao Deng¹, Zhen Guo⁴, Menghua Li¹ and Lifeng Chen⁵

- ¹ Xi'an Aeronautics Computing Technique Research Institute, AVIC, Xi'an 710069, China; dengbao15@sina.com (B.D.); lmhua1@126.com (M.L.)
- ² School of Automation, Northwestern Polytechnical University, Xi'an 710072, China
- ³ Key Laboratory of Information Fusion Technology, Ministry of Education, Xi'an 710072, China
- ⁴ System Design Institute of Hubei Aerospace Technology Academy, Wuhan 430040, China; guozhennpu@126.com
- ⁵ Department of Precision Instrument, Tsinghua University, Beijing 100084, China; chenlf22@mails.tsinghua.edu.cn
- * Correspondence: hym_henu@163.com (Y.H.); quanpan@nwpu.edu.cn (Q.P.)

Abstract: The variational Bayesian method solves nonlinear estimation problems by iteratively computing the integral of the marginal density. Many researchers have demonstrated the fact its performance depends on the linear approximation in the computation of the variational density in the iteration and the degree of nonlinearity of the underlying scenario. In this paper, two methods for computing the variational density, namely, the natural gradient method and the simultaneous perturbation stochastic method, are used to implement a variational Bayesian Kalman filter for maneuvering target tracking using Doppler measurements. The latter are collected from a set of sensors subject to single-hop network constraints. We propose a distributed fusion variational Bayesian Kalman filter for a networked maneuvering target tracking scenario and both of the evidence lower bound and the posterior Cramér–Rao lower bound of the proposed methods are presented. The simulation results are compared with centralized fusion in terms of posterior Cramér–Rao lower bounds, root-mean-squared errors and the 3σ bound.

Keywords: distributed fusion; nonlinear estimation; variational Bayesian optimization; natural gradient; simultaneous perturbation stochastic approximation; Kullback–Leibler divergence

1. Introduction

There has been much interest in recent years in the use of sensor networks as opposed to single sensors in the fields of target tracking [1], intelligent transportation [2], environmental monitoring [3], spacecraft navigation [4], etc. Multi-sensor nodes can provide greater spatial coverage and, by cooperation, effectively complement the limitations of a single sensor, potentially resulting in improving estimation and fusion performance. Along with the rapid development of sensor network technologies, the two aspects of fusion architecture and state estimation optimization have been studied in the past few years [5–7].

Sensor network fusion can be classified into three categories, i.e., centralized, decentralized, and distributed architectures [6,8]. Schematically, examples of different fusion architectures are shown in Figure 1. The topology of sensor networks is, typically, described as an undirected graph where nodes can exchange measurements (or estimation) from neighbors via a bidirectional edge or a directed graph where nodes only can deliver measurements (or estimation) in a fixed direction. Broadcast communications can be single-hop or multi-hop [9]. Here, we only consider the sensor network with undirected topology and single-hop communications. The yellow circles and blue circles describe fusion centers and sensor nodes, respectively.



Citation: Hu, Y; Pan, Q; Deng, B; Guo, Z.; Li, M.; Chen, L. Variational Bayesian Algorithms for Maneuvering Target Tracking with Nonlinear Measurements in Sensor Networks. *Entropy* 2023, 25, 1235. https://doi.org/10.3390/e25081235

Academic Editor: Éloi Bossé

Received: 15 June 2023 Revised: 22 July 2023 Accepted: 14 August 2023 Published: 18 August 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https:// creativecommons.org/licenses/by/ 4.0/).



Figure 1. The architectures of sensor networks. (a) Centralized sensor network architectures. (b) Decentralized sensor network architectures. (c) Distributed sensor network architectures.

In the centralized architecture, information fusion takes place after the local sensor measurements are delivered into the fusion center. This architecture can make full use of measurements if the communication bandwidth is high enough to accommodate transmission of all sensor measurements to the fusion center, leading to theoretically optimal fusion [10]. As a result, it is generally utilized as a benchmark for performance comparison and evaluation of the other fusion architectures that will be discussed here. However, there are several inherent problems that need to be taken into account in considering a centralized fusion architecture. One problem is the measurement delay resulting from communication or sensor sampling rate [11–13]. Another is that the centralized architecture will, typically, result in a trade-off between timely fusion accuracy performance, and the requirements of communication bandwidth and computational costs because of the broadcast and processing of measurements from all sensors [6,14]. In addition, this architecture is also generally more sensitive to outliers than other fusion architectures [15], such as the following decentralized one.

In decentralized architecture, sensors are partitioned into several clusters, each of them with a fusion center [16,17] where measurements from neighboring nodes are collected to yield a local fused estimation. The decentralized architecture is distinguished from the centralized one by the multiple fusion centers, which can lead to a spread of computational costs across multiple devices and better robustness [6].

In the distributed fusion architecture, each node individually provides a local estimation by using measurements collected from itself and its neighbors. It could be considered as a special form of the decentralized one. Compared with the other two forms, it has the following advantages: (1) Enhanced scalability and feasibility; it allows relatively easy scaling of the network up or down by adding or subtracting nodes depending on practical applications [18]. (2) Increased robustness and fault tolerance to node failures, especially in harsh situations, such as an underwater environment. (3) Reduced bandwidth requirement; it mitigates communication bottlenecks that might arise, for example, in target detection and tracking systems. (4) Reduced computational cost of the three architectures because some processing such as inverse operations generally needed in estimation are performed in their own individual fusion centers. A significant amount of literature on distributed fusion architectures has been published. Ref. [10] is an early paper describing the basic principles of distributed fusion architecture in target tracking systems. A comprehensive review of the characteristics, advantages and estimation solutions for distributed low-cost sensor networks is presented in [6]. For distributed robust filtering, a variational Bayesian (VB) algorithm with a conjugate-exponential model is proposed in [19]. As mentioned in [20], the problem of distributed detection and tracking over a Doppler-shift sensor network is studied. Consensus methods are used in [21,22] for the problem of uncertain noise statistics of distributed sensor networks. However, distributed fusion architecture also has some negative issues, for example, lack of awareness of the sensor network as a whole.

In the context of fusion architectures, state estimation is also essential for target tracking in sensor networks, especially for a nonlinear system. Exact solutions for the posterior probability density function (PDF) for nonlinear systems are mostly unavailable, resulting in a considerable amount of work on approximations of the posterior. A great number of nonlinear filters, such as the extended Kalman filter (EKF) [23], the unscented Kalman filter (UKF) [24], the Gauss–Hermite Kalman filter (GHKF) [25], the central difference Kalman filter (CDKF) [25], the cubature Kalman filter (CKF) [26] and their variants [27,28], have been proposed under a Gaussian noise assumption both on measurements and system. For non-Gaussian noise models, stochastic estimation methods, such as Markov Chain Monte Carlo [29], and sequential Monte Carlo (also named particle filters (PF)) [30] and variants [31,32] have been the focus of much attention over the last two decades. Since these methods need a large number of particles to ensure nonlinear estimation accuracy, Doucet and his colleagues introduced the Rao–Blackwellized particle filter (RBPF) [33] to marginalize out the linear variables to be solved by an optimal filter, and focused the sampled particles on the remaining variables (nonlinear part). For more information about nonlinear estimation problems, we refer the reader to the review articles [34–36] where more comprehensive interpretations have been provided.

In terms of optimization techniques, gradient-based optimization is often used in nonlinear filters to improve the performance of nonlinear estimation. As shown in [37], gradient descent is adopted for smartphone orientation estimation, yielding a quaternionbased Kalman filter algorithm to estimate exercise motion. In [38], a gradient descent iterative nonlinear Kalman is proposed for the problem of random missing outputs. For problems where the objective function, such as the variational evidence lower bound (ELBO) and the Kullback–Leibler Divergence (KLD), need to be maximized or minimized, the gradient method generally requires linearization of the objective function. For nonlinear estimation optimization problems, the gradient can be written as a mean of the samples of interest, yielding a stochastic gradient optimization method [39]. In [40], the authors present a modified particle filter where stochastic gradient is used to minimize the criterion function. Compared with the ordinary gradient, natural gradient (NG) has the advantage of theoretical connections from information geometry. According to Amari's works [41,42], it has a steepest direction in Riemannian space. As shown in [41], NG comes with a theoretical guarantee of asymptotic optimality and can be used to produce a Fisher-efficient iterative estimator on a statistical manifold. In [43], the NG method for the nonlinear estimation problem is proved to be asymptotically optimal in the sense of the Cramér–Rao bound. The earliest interest in NG can be found in [44,45]. Simultaneous perturbation stochastic approximation (SPSA) [46] is an alternative optimization method which can be use for stochastic search. It is a gradient approximation statistic optimization method, which does not require the linearization of the objective function [46]. In this method, a group of samples of objective functions is sampled to obtain a two-side differential function, so that it can be used to approximate the otherwise intractable gradient of the objective function to be estimated. An application of SPSA for detection of the center of a thermal updraft is presented in [47], resulting in an adaptive autonomous soaring algorithm for multiple unmanned aerial vehicles.

In terms of the iterative optimization methods, among the most common approaches are expectation-maximization (EM) [48] and VB [49,50]. EM realizes estimation optimization by establishing a feedback loop, including an expectation step (E-step) and a maximization step (M-step). In the E-step, the conditional expectation of the likelihood function is calculated according to a given prior and measurement. In the M-step, the conditional expectation is maximized to obtain the estimation of the variables. While the state and parameters are estimated and optimized alternately in the iteration cycle, EM still has problems for large scale models and data. VB differs from EM in that the parameters are stochastic [50], resulting in VB being capable of making a joint distribution of state and parameters, and then being especially suitable for dealing with high-dimensional and large-scale problems via mean field theory. Recent advances in the variational iterative framework can be found in [19,51], in which a unified VB approach is provided for the joint estimation of system state and parameters in a target tracking system.

This paper considers the problems of estimation optimization and fusion in networked target tracking systems, and aims to derive a nonlinear estimation, optimization and fusion approach by utilizing VB with the optimization methods of NG and SPSA, achieving a high performance in accuracy. The key contributions of this paper are as follows:

- 1. Development of a distributed variational fusion framework by utilizing variational mean field theory to approximately partition the joint posterior distribution into several solvable variational distributions under the assumption of independent measurements.
- 2. Presentation of a novel deterministic nonlinear estimation optimization method to maximize the distributed ELBO using NG, based on linearization to approximate the posterior distributions closely, producing the DVBKF-NG algorithm which yields closed form nonlinear state estimation with the associated covariance for the sensor network.
- 3. Presentation of a novel stochastic estimation optimization for the VB framework by using SPSA and deriving the stochastic gradient estimation of ELBO, thus producing an iterative filter, i.e., DVBKF-SPSA.
- 4. Demonstrate performance metrics: distributed ELBO and the posterior Cramér–Rao lower bound (PCRLB) for these algorithms over different iterations.

The rest of the paper is organized as follows. In Section 2, we introduce the general nonlinear estimation problem in target tracking over a sensor network. A distributed variational Bayesian estimation optimization framework is proposed in Section 3, involving partitioning of the joint measurement likelihood. Under the optimization framework, two distributed iterative nonlinear variational Bayesian Kalman filtering algorithms (DVBKF-NG and DVBKF-SPSA) are presented in Section 4, using NG and SPSA, respectively. In Section 5, we consider two kinds of metrics, i.e., ELBO and PCRLB, to evaluate the performance of the proposed methods. In Section 6, we give an example to verify the proposed methods for maneuvering target tracking over a sensor network. Finally, conclusions are drawn and future work is discussed in Section 7.

2. Problem Formulation

Consider a maneuvering target travels in the area covered by a fully connected sensor network with *N* nodes; the measurements $\mathcal{Z}_k = [\mathcal{Z}_{1,k}, \cdots, \mathcal{Z}_{n,k}, \cdots, \mathcal{Z}_{N,k}]^T$ with noise covariance \mathcal{R}_k at time index *k* are collected from the set of sensors, where $\mathcal{Z}_{n,k} = [z_{n,k}, z_{n,k}^1, \cdots, z_{n,k}^l, \cdots, z_{n,k}^L]^T$ denotes the measurements from sensor *n* and its neighbors, where $z_{n,k} \in \mathbb{R}^d$ is the measurement of the *n*th sensor and $z_{n,k}^l$ is the measurement from the *l*th neighbor of the *n*th sensor, the sensor index $n \in \{1, 2, \cdots, N\}$, $l \in \{1, 2, \cdots, L\}$ and satisfies $L \leq N - 1$, *d* denotes the dimension of $z_{n,k}$. We assume that only the single-hop communication between two sensors is considered in this paper. The system model and the measurement model are given as follows, respectively,

$$x_k = f_{k|k-1}(x_{k-1}) + \omega_{k-1} \tag{1}$$

$$z_{n,k} = h_{n,k}(x_k) + v_{n,k}, \ l \in \{1, 2, \cdots, L\}$$
(2)

where $x_k \in \mathbb{R}^m$ is the system state which may include position, velocity and other related quantities, and *m* is the dimension of x_k . Both the system process ω_k and the measurement noise $v_{n,k}$ of sensor *n* are assumed to be mutually independent and zero-mean Gaussian noises with covariances of Q_{k-1} and $R_{n,k}$, respectively. And $f_{k|k-1}(\cdot)$ and $h_{n,k}(\cdot)$ denote the state transfer function and measurement function, respectively.

What we want is to estimate the system state based on the collection of sensor measurements, which is the process of inferring the system state of interest using measurements Z_k , usually, under the Minimum Mean Square Error (MMSE) criterion, i.e.,

$$x_{k|k} \triangleq \mathbb{E}_p[x_k|\mathcal{Z}_k] = \int x_k p(x_k|\mathcal{Z}_k) dx_k$$
(3)

$$P_{k|k} \triangleq \mathbb{E}_p\left[\left(x_k - x_{k|k}\right)\left(x_k - x_{k|k}\right)^{\mathrm{T}} | \mathcal{Z}_k\right] = \int \left(x_k - x_{k|k}\right)\left(x_k - x_{k|k}\right)^{\mathrm{T}} p(x_k | \mathcal{Z}_k) dx_k \quad (4)$$

where $x_{k|k}$ and $P_{k|k}$ are the state estimation and the associated error covariance, and $\mathbb{E}_p[\cdot] \triangleq \mathbb{E}_{p(x_k|z_k)}[\cdot]$. The posterior distribution $p(x_k|\mathcal{Z}_k)$ is expressed as

$$p(x_k|\mathcal{Z}_k) = \frac{p(\mathcal{Z}_k|x_k)p(x_k|\mathcal{Z}_{k-1})}{\int p(x_k,\mathcal{Z}_k)dx_k}$$
(5)

where $p(x_k|\mathcal{Z}_{k-1})$ is the a priori distribution and $p(\mathcal{Z}_k|x_k)$ is the likelihood of measurement \mathcal{Z}_k with given x_k . Under Gaussian assumptions, the predicted density $p(x_k|\mathcal{Z}_{k-1})$ is given by

$$p(x_k|\mathcal{Z}_{k-1}) \sim \mathcal{N}\Big(x_k|f_{k|k-1}(x_{k-1|k-1}), F_{k|k-1}P_{k-1|k-1}F_{k|k-1}^{\mathrm{T}} + Q_{k-1}\Big)$$
(6)

where $F_{k|k-1} = \frac{\partial f_k(x)}{\partial x}|_{x=x_{k-1|k-1}}$. For nonlinear systems, computing the integral $\int p(x_k, z_k) dx_k$ in the denominator of (5) is in general difficult and the optimal approximation is needed. In this paper, we are interested to use the variational Bayesian method to approximate the a posteriori distribution and thus solve the nonlinear estimation problem of maneuvering target tracking in the described sensor network.

3. Distributed Variational Bayesian Estimation Optimization Framework

Given measurement \mathcal{Z}_k collected from sensors, the principle of VB is given as

$$\log p(\mathcal{Z}_k) = \mathcal{L}(\psi_k) + \mathbb{D}_{\mathrm{KL}}[q(x_k|\psi_k)||p(x_k|\mathcal{Z}_k)]$$
(7)

and

$$\mathcal{L}(\psi_k) = \int q(x_k | \psi_k) \log \frac{p(\mathcal{Z}_k, x_k)}{q(x_k | \psi_k)} dx_k$$
(8)

$$\mathbb{D}_{\mathrm{KL}}[q(x_k|\psi_k)||p(x_k|\mathcal{Z}_k)] = -\int q(x_k|\psi_k)\log\frac{p(x_k|\mathcal{Z}_k)}{q(x_k|\psi_k)}dx_k \tag{9}$$

where $q(x_k|\psi_k)$ is a variational distribution with parameter ψ_k , $\mathcal{L}(\psi_k)$, and $\mathbb{D}_{\text{KL}}[q(x_k|\psi_k)||p(x_k|\mathcal{Z}_k)]$ are the ELBO and the KLD between $q(x_k|\psi_k)$ and $p(x_k|\mathcal{Z}_k)$, respectively. What we wish to do is to approximate the $p(x_k|\mathcal{Z}_k)$ as closely as possible by minimizing the KLD. Actually, the approximation can be considered as a moving variational distribution along a chosen search direction iteratively to the position of the posterior distribution in a statistical manifold. However, it is difficult to seek out an appropriate method to achieve the approximation, because the variational distribution is usually underparameterized and not sufficiently flexible to capture the true posterior [52]. As a result, the ELBO is considered to be maximized since maximizing ELBO is equivalent to minimizing KLD from (7), i.e.,

$$\log p(\mathcal{Z}_k) \ge \mathcal{L}(\psi_k) = \mathbb{E}_q[\log p(\mathcal{Z}_k|x_k)] - \mathbb{D}_{\mathrm{KL}}[q(x_k|\psi_k)||p(x_k)].$$
(10)

We assume that measurements from all sensors are mutually independent; the term $\mathbb{E}_q[\log p(\mathcal{Z}_k|x_k)]$ in (10) can be partitioned in distributed fusion architecture as follows

$$\mathbb{E}_{q}[\log p(\mathcal{Z}_{k}|x_{k})] = \sum_{n=1}^{N} \mathbb{E}_{q}[\log p(\mathcal{Z}_{n,k}|x_{k})]$$
(11)

where the measurements $Z_{n,k}$ with associated noise variance $\mathcal{R}_{n,k}$ are collected from sensor *n* and its neighbors, where

$$\mathcal{Z}_{n,k} = \left[z_{n,k}^{\mathrm{T}}, (z_{n,k}^{1})^{\mathrm{T}}, \cdots, (z_{n,k}^{l})^{\mathrm{T}}, \cdots, (z_{n,k}^{L})^{\mathrm{T}} \right]^{\mathrm{T}} \\
p(\mathcal{Z}_{n,k} \mid x_{k}) = \mathcal{N}(\mathcal{Z}_{n,k}; \mathcal{H}_{n,k}(x_{k}), \mathcal{R}_{n,k}) \\
\mathcal{H}_{n,k}(\cdot) = \left[h_{n,k}^{\mathrm{T}}(\cdot), \cdots, (h_{n,k}^{L}(\cdot))^{\mathrm{T}} \right]^{\mathrm{T}} \\
\mathcal{R}_{n,k} = \operatorname{Cov}\left(\mathcal{V}_{n,k}, \mathcal{V}_{n,k}^{\mathrm{T}} \right) = \operatorname{blkdiag}(R_{n,k}, R_{n,k}^{1}, \cdots, R_{n,k}^{l}, \cdots, R_{n,k}^{L})$$
(12)

where $z_{n,k}$ denotes the measurement of sensor n, and $z_{n,k}^l$ denotes the measurement from the *l*th neighbor of sensor n, $\mathcal{V}_{n,k} = \left[v_{n,k}^{\mathsf{T}}, (v_{n,k}^1)^{\mathsf{T}}, \cdots, (v_{n,k}^l)^{\mathsf{T}}, \cdots, (v_{n,k}^L)^{\mathsf{T}}\right]^{\mathsf{T}}$ is the measurement noise of sensor n and its neighbors, and $\mathcal{R}_{n,k}$ and $\mathcal{H}_{n,k}(x_k)$ are the associated noise covariance and measurement matrix, respectively, where $R_{n,k}^l = \operatorname{Cov}\left(v_{n,k}^l, (v_{n,k}^l)^{\mathsf{T}}\right)^{\mathsf{T}}$ denotes the variance of the noise $v_{n,k}^l$ from the *l*th neighbor of sensor n. The math symbol blkdiag(\cdot) denotes a block diagonal matrix created by aligning input matrices.

For the (i + 1)th variational iteration, we can take the *i*th variational distribution $q(x_k | \psi_k^i)$ as the prior; the optimized ELBO can be given as

$$\mathcal{L}(\psi_k^*) = \sum_{n=1}^N \mathbb{E}_q[\log p(\mathcal{Z}_{n,k}|x_k)] - \mathbb{D}_{\mathrm{KL}}\Big[q(x_k|\psi_k)||q(x_k|\psi_k^i)\Big].$$
(13)

We note the definition $\psi_k \triangleq (x_{k|k}, P_{k|k})$ in this paper, and wish to update state estimation $x_{k|k}$ and the associated error covariance $P_{k|k}$ in each iteration by maximizing the above ELBO. In the following section, we present two distributed iterative variational Bayesian Kalman filters for maneuvering target tracking by using NG and SPSA, respectively.

4. Distributed Iterative Variational Bayesian Kalman Filters over Sensor Network

In this section, with the assumption that the measurements are mutually independent, we present alternative distribution variational Bayesian Kalman filtering algorithms (DVBKF) via NG and SPSA.

4.1. NG-Based DVBKF

NG calculated by using information geometry (generally KLD linearization) has a steepest direction in Riemannian space. According to Amari's works [41,42], we present the NG of objective function $\mathcal{L}(\psi_k)$ with respect to parameter ψ_k as follows. Set $\Delta \psi_k \triangleq \psi_k - \psi_k^i \to 0$, and (13) can be rewritten as

$$\psi_{k}^{*} = \underset{\Delta\psi_{k}\to0}{\arg\max} \left\{ \sum_{n=1}^{N} \nabla_{\psi_{k}} \mathbb{E}_{q} [\log p(\mathcal{Z}_{n,k}|x_{k})] \Delta\psi_{k} - (\Delta\psi_{k})^{\mathrm{T}} F_{\psi_{k}^{i}} \Delta\psi_{k} \right\}$$
(14)

where $F_{\psi_{L}^{i}}$ is Fisher information and presented as

$$F_{\psi_k^i} \approx \nabla_{\psi_k^i}^2 \mathbb{D}_{\mathrm{KL}} \Big[q(x_k | \psi_k) \| q\left(x_k | \psi_k^i\right) \Big].$$
(15)

The proof can be seen in [53]. After computing the partial derivative of the right side of (14) and setting it equal to 0, the NG of the ELBO at the *i*th iteration in distributed architecture is expressed as

$$\widetilde{\nabla}_{\psi_k^i} = F_{\psi_k^i}^{-1} \sum_{n=1}^N \nabla_{\psi_k^i} \mathbb{E}_q[\log p(\mathcal{Z}_{n,k}|x_k)].$$
(16)

The NG $\nabla_{\psi_k^i}$ is the direction in which the increase of ELBO is greatest [54]; that means it has the greatest descent or ascent at each iteration in statistical manifold space, and can move the variational distribution to approximate the posterior fastest. At this point, the optimal ψ_k is presented as

$$\psi_k^{i+1} = \psi_k^i + \widetilde{\nabla}_{\psi_k^i} \tag{17}$$

It is observed that (17) is a general expression of the iterative update of parameter ψ_k . In terms of the update of $x_{k|k}^{i+1}$ and $P_{k|k}^{i+1}$, the special forms of Fisher information matrices $(F_{x_{k|k}^i}^{-1} \text{ and } F_{P_{k|k}^i}^{-1})$ and the gradients of log-likelihood expectation $(\nabla_{x_{k|k}^i} \mathbb{E}_q[\log p(\mathcal{Z}_{n,k}|x_k)])$ and $\nabla_{P_{k|k}^i} \mathbb{E}_q[\log p(\mathcal{Z}_{n,k}|x_k)]$ with respect to $x_{k|k}$ and $P_{k|k}$ need to be analyzed.

With the Gaussian system assumption, the KLD between two Gaussian distributions $q_1 \sim \mathbb{N}(\xi_1; \mu_1, C_1)$ and $q_2 \sim \mathbb{N}(\xi_2; \mu_2, C_2)$ with the same dimension *d* is given as

$$\mathbb{D}_{\mathrm{KL}}[q_1 \| q_2] = \frac{1}{2} \Big\{ \ln(|C_2| |C_1|^{-1}) + \operatorname{tr} \Big(C_2^{-1} C_1 \Big) + (\mu_2 - \mu_1)^{\mathrm{T}} C_2^{-1} (\mu_2 - \mu_1) - d \Big\}$$
(18)

Combining with (15), the Fisher information matrices with respect to $x_{k|k}^{l}$ and $P_{k|k}^{l}$ are presented as

$$F_{x_{k|k}^{i}} = (P_{k|k}^{i})^{-1}$$
(19)

$$F_{P_{k|k}^{i}} \approx \frac{1}{2} (P_{k|k}^{i})^{-1} \otimes (P_{k|k}^{i})^{-1}.$$
⁽²⁰⁾

The gradient of log-likelihood expectation with respect to $x_{k|k}^i$ and $P_{k|k}^i$ of sensor *n* are presented as

$$\nabla_{x_{k|k}^{i}} \mathbb{E}_{q}[\log p(\mathcal{Z}_{n,k}|x_{k})] \approx H_{n,x_{k|k}^{i}}^{\mathrm{T}} \mathcal{R}_{n,k}^{-1} \Big(\mathcal{Z}_{n,k} - \mathcal{H}_{n,k}\Big(x_{k|k}^{i}\Big)\Big)$$
(21)

$$\nabla_{P_{k|k}^{i}} \mathbb{E}_{q}[\log p(\mathcal{Z}_{n,k}|x_{k})] \approx -\frac{1}{2} H_{n,x_{k|k}^{i}}^{\mathrm{T}} \mathcal{R}_{n,k}^{-1} H_{n,x_{k|k}^{i}}$$
(22)

where $H_{n,x_{k|k}^i} = \frac{\partial \mathcal{H}_{n,k}(x_k)}{\partial x_{k|k}}\Big|_{x_{k|k}=x_{k|k}^i}$ denotes the Jacobian matrix of the measurement matrix of the *n*th sensor.

Recalling the iterative optimization forms in (16) and (17), the distributed iterative state estimation $x_{k|k}^{i+1}$ and the associated covariance $P_{k|k}^{i+1}$ in DVBKF-NG are, respectively, given by

$$x_{k|k}^{i+1} = x_{k|k}^{i} + P_{k|k}^{i} \sum_{n=1}^{N} H_{n,x_{k|k}^{i}}^{\mathrm{T}} \mathcal{R}_{n,k}^{-1} \Big(\mathcal{Z}_{n,k} - \mathcal{H}_{n,k} \Big(x_{k|k}^{i} \Big) \Big)$$
(23)

$$P_{k|k}^{i+1} = P_{k|k}^{i} \left(I - \sum_{n=1}^{N} H_{n,x_{k|k}^{i}}^{\mathrm{T}} \mathcal{R}_{n,k}^{-1} H_{n,x_{k|k}^{i}} P_{k|k}^{i} \right).$$
(24)

The iterative optimization process of DVBKF-NG is summarized in Algorithm 1. We make the following remarks:

- 1. The update of $x_{k|k}^i$ is preconditioned by $P_{k|k'}^i$ which produces an adaptive movement to the posterior PDF along the direction of NG.
- 2. Clearly, Equation (24) shows that $P_{n,k|k}^{i+1} \leq P_{n,k|k}^{i}$, which means that estimation error covariance decreases gradually at each iteration. Additionally, since $H_{n,x_{k|k}}^{T} \mathcal{R}_{n,k}^{-1} H_{n,x_{k|k}^{i}}$ is the expectation of the Hessian, $P_{n,k|k}^{i}$ has a quadratic convergence.
- 3. To make the algorithm adaptive, the relative estimation error e_r can be used for judging the iteration termination.

$$e_r = \left| \frac{x_{k|k}^{i+1} - x_{k|k}^i}{x_{k|k}^i} \right| \le \epsilon \tag{25}$$

where ϵ is a small positive number which can be chosen according to practical scenarios.

Algorithm 1 The DVBKF-NG algorithm

- 1: Initialize state estimation $x_{1|1}$, estimation error covariance $P_{1|1}$, the number of iterations Iter:
- 2: Compute one-step predicted state

 $x_{k|k-1} = f_{k|k-1} \left(x_{k-1|k-1} \right);$ 3: Compute predicted state error covariance

- $P_{k|k-1} = \vec{F}_{k|k-1} P_{k-1|k-1} F_{k|k-1}^{\mathrm{T}} + Q_k;$
- 4: Let i = 1, and $x_{k|k}^1 = x_{k|k-1}$, $P_{k|k}^1 = P_{k|k-1}$.
- 5: **for** each iteration i = 1: Iter and $e_r \leq \epsilon$ **do**
- Compute Fisher information matrices with respect to $x_{k|k}^{i+1}$ and $P_{k|k}^{i+1}$ 6: by (19) and (20).
- Compute the gradients of log-likelihood expectation of each sensor with respect to 7: $x_{k|k}^{i+1}$ and $P_{k|k}^{i+1}$ by (21) and (22), respectively.
- Compute the iterative state estimation $x_{k|k}^{i+1}$ and the associated error covariance $P_{k|k}^{i+1}$ 8: by (23) and (24), respectively.
- Compute relative error e_r by (25). 9:

10: end for

11: Output $x_{k|k} = x_{k|k}^{i+1}$, $P_{k|k} = P_{k|k}^{i+1}$.

4.2. SPSA-Based DVBKF

SPSA is a statistical optimization method for gradient approximation, which does not require the full knowledge of the objective function being minimized (or maximized) and parameters being optimized [46]. In this method, a group of samples of objective function $\mathcal{L}(\psi_k)$ is sampled as $\mathcal{Y}(\psi_k) = \mathcal{L}(\psi_k) + \zeta$ to obtain a two-side differential function, where ζ is a random infinitesimal perturbation. Then, the two-side infinitesimal function is computed by

$$d\mathcal{Y}(\psi_k^i) = \mathcal{Y}(\psi_k + c_{\psi_k^i} \Delta_{\psi_k^i}) - \mathcal{Y}(\psi_k - c_{\psi_k^i} \Delta_{\psi_k^i}).$$
(26)

where $\Delta_{\psi_i^i}$ is random perturbation vector with a Gaussian distribution form and $c_{\psi_i^i}$ is a small positive number that decreases with *i*. As a result, the estimation of the gradient of objective function $\mathcal{L}(\psi_k)$ can be obtained by the two-side differential. The *m*th component of the gradient estimator at the *i*th iteration is given as follows,

$$(\hat{g}(\psi_{k}^{i}))_{m} = \frac{\mathcal{Y}(\psi_{k} + c_{\psi_{k}^{i}}\Delta_{\psi_{k}^{i}}) - \mathcal{Y}(\psi_{k} - c_{\psi_{k}^{i}}\Delta_{\psi_{k}^{i}})}{2c_{\psi_{k}^{i}}(\Delta_{\psi_{k}^{i}})_{m}}$$
(27)

where $m \in \{1, 2, \dots, M\}$, *M* is the dimension of ψ_k . The gradient estimation of $\mathcal{L}(\psi_k)$ at the *i*th iteration is presented as

$$\hat{\mathcal{G}}\left(\psi_{k}^{i}\right) = \left[(\hat{g}(\psi_{k}^{i}))_{1} \left(\hat{g}(\psi_{k}^{i})\right)_{2} \cdots \left(\hat{g}(\psi_{k}^{i})\right)_{M}\right]^{\mathrm{T}} = \frac{d\mathcal{Y}(\psi_{k}^{i})}{2c_{\psi_{k}^{i}}}\Lambda_{\psi_{k}^{i}}$$
(28)

where $\Lambda_{\psi_k^i} = \left[(\Delta_{\psi_k^i})_1^{-1} (\Delta_{\psi_k^i})_2^{-1} \cdots (\Delta_{\psi_k^i})_M^{-1} \right]^T$, $\Lambda_{\psi_k^i}$ is random perturbation vector with multivariate Gaussian distribution form and $(\Delta_{\psi_k^i})_m$ denotes the *m*th element of $\Lambda_{\psi_k^i}$. At this point, the ψ_k^{i+1} can be updated by

$$\psi_k^{i+1} = \psi_k^i + a_{\psi_k^i} \hat{\mathcal{G}}\left(\psi_k^i\right) \tag{29}$$

where $a_{\psi_{L}^{i}}$ is a weighted factor.

Taking the distributed variational ELBO $\mathcal{L}(\psi_k)$ in (13) as the objective function, the two-side infinitesimal perturbations of $\mathcal{L}(\psi_k)$ with respect to $x_{k|k}^i$ and $P_{k|k}^i$ are given, respectively, as

$$\mathcal{L}\left(x_{k|k} \pm c_{x_{k|k}^{i}} \Delta_{x_{k|k}^{i}}\right) = \sum_{n=1}^{N} \mathbb{E}_{q}\left[\log p\left(\mathcal{Z}_{n,k} | x_{k|k}^{i} \pm c_{x_{k|k}^{i}} \Delta_{x_{k|k}^{i}}\right)\right] - \mathbb{D}_{\mathrm{KL}}\left[\mathcal{N}\left(x_{k} | x_{k|k}^{i} \pm c_{x_{k|k}^{i}} \Delta_{x_{k|k}^{i}}, P_{k|k}\right) ||\mathcal{N}\left(x_{k} | x_{k|k}^{i}, P_{k|k}^{i}\right)\right]$$
(30)

$$\mathcal{L}\left(P_{k|k} \pm c_{P_{k|k}^{i}} \Delta_{P_{k|k}^{i}}\right) = -\frac{1 \pm c_{P_{k|k}^{i}}}{2} \sum_{n=1}^{N} H_{n,x_{k|k}^{i}}^{\mathrm{T}} \mathcal{R}_{k}^{-1} H_{n,x_{k|k}^{i}} \Delta_{P_{k|k}^{i}} \\ -\mathbb{D}_{\mathrm{KL}}\left[\mathcal{N}\left(x_{k}|x_{k|k}^{i}, P_{k|k} \pm c_{P_{k|k}^{i}} \Delta_{P_{k|k}^{i}}\right) ||\mathcal{N}\left(x_{k}|x_{k|k}^{i}, P_{k|k}^{i}\right)\right]$$
(31)

where the parameters $c_{x_{k|k}^i}$ and $c_{P_{k|k}^i}$ are the special forms of $c_{\psi_k^i}$ with respect to $x_{k|k}^i$ and $P_{k|k'}^i$ respectively.

Now, we formulate the random perturbation factors of $\Delta_{x_{k|k}^i}$ and $\Delta_{P_{k|k}^i}$ by sampling from the Gaussian distribution $x_k^s \sim \mathcal{N}(x_{k|k'}^i, P_{k|k}^i)$; the mean and covariance of the samples are $\overline{x}_k^s = \frac{1}{S} \sum_{s=1}^S x_k^s$ and $\delta^2 = \frac{1}{S-1} \sum_{s=1}^S (x_k^s - \overline{x}_k^s) (x_k^s - \overline{x}_k^s)^T$, where $s \in \{1, 2, \dots, S\}$. Randomly choose a sample x_k^s ; the random perturbation of state $\Delta_{x_{k|k}^i}$ is given by

$$\Delta_{x_{k|k}^{i}} = x_{k|k}^{i} - x_{k}^{s}.$$
(32)

The associated random perturbation of covariance is given as

$$\Delta_{P^i_{k|k}} = \delta^2 - P^i_{k|k}. \tag{33}$$

Recall (26), the two-side differential $d\mathcal{Y}(\psi_k^i)$ with respect to $x_{k|k}^i$ and $P_{k|k}^i$ is written as

$$d\mathcal{Y}\left(x_{k|k}^{i}\right) = \mathcal{L}\left(x_{k|k}^{i} + c_{x_{k|k}^{i}}\Delta_{x_{k|k}^{i}}\right) - \mathcal{L}\left(x_{k|k}^{i} - c_{x_{k|k}^{i}}\Delta_{x_{k|k}^{i}}\right) + \zeta_{x}$$
(34)

$$d\mathcal{Y}\left(P_{k|k}^{i}\right) = \mathcal{L}\left(P_{k|k}^{i} + c_{P_{k|k}^{i}}\Delta_{P_{k|k}^{i}}\right) - \mathcal{L}\left(P_{k|k}^{i} - c_{P_{k|k}^{i}}\Delta_{P_{k|k}^{i}}\right) + \zeta_{P}.$$
(35)

where ζ_x and ζ_P are random infinitesimal perturbations.

It follows that the estimation of the gradient of $\mathcal{L}(\psi_k)$ with respect to $x_{k|k}^i$ and $P_{k|k}^i$ at the *i*th iteration is given by

$$\hat{\mathcal{G}}\left(x_{k|k}^{i}\right) = \frac{d\mathcal{Y}\left(x_{k|k}^{i}\right)}{2c_{x_{k|k}^{i}}\Delta_{x_{k|k}^{i}}}\Lambda_{x_{k|k}^{i}}$$
(36)

$$\hat{\mathcal{G}}(P_{k|k}^{i}) = \frac{d\mathcal{Y}\left(P_{k|k}^{i}\right)}{2c_{P_{k|k}^{i}}\Delta_{P_{k|k}^{i}}}\Lambda_{P_{k|k}^{i}}$$
(37)

where

$$\Lambda_{x_{k|k}^{i}} = \left[(\Delta_{x_{k|k}^{i}})_{1}^{-1} (\Delta_{x_{k|k}^{i}})_{2}^{-1} \cdots (\Delta_{x_{k|k}^{i}})_{M}^{-1} \right]^{\mathrm{T}}$$

$$\Lambda_{P_{k|k}^{i}} = \mathrm{diag} \left[(\Delta_{P_{k|k}^{i}})_{1,1}^{-1} (\Delta_{P_{k|k}^{i}})_{2,2}^{-1} \cdots (\Delta_{P_{k|k}^{i}})_{M,M}^{-1} \right]$$

and $(\Delta_{P_{k|k}^{i}})_{m,m}$ denotes the element in the *m*th row and *m*th column of $\Delta_{P_{k|k}^{i}}$. Therefore, state estimation and the associated covariance are updated by

$$x_{k|k}^{i+1} = x_{k|k}^{i} + a_{x_{k|k}^{i}} \hat{\mathcal{G}}\left(x_{k|k}^{i}\right)$$
(38)

$$P_{k|k}^{i+1} = P_{k|k}^{i} + a_{P_{k|k}^{i}} \hat{\mathcal{G}}(P_{k|k}^{i})$$
(39)

where the parameters of $a_{x_{k|k}^i}$ and $a_{P_{k|k}^i}$ are weighted factors. Set the same iteration termination as (25); the DVBKF-SPSA algorithm is summarized in Algorithm 2.

Algorithm 2 The iterative optimization process in the DVBKF-SPSA algorithm

- 1: Initialize state estimation $x_{1|1}$, estimation error covariance $P_{1|1}$, the number of iterations Iter;
- 2: Compute one-step predicted state

$$x_{k|k-1} = f_{k|k-1}(x_{k-1|k-1});$$

- 3: Compute predicted state error covariance $P_{k|k-1} = F_{k|k-1}P_{k-1|k-1}F_{k|k-1}^{T} + Q_k$;
- 4: Let i = 1, and $x_{k|k}^1 = x_{k|k-1}$, $P_{k|k}^1 = P_{k|k-1}$.
- 5: **for** each iteration i = 1: Iter and $e \le \epsilon$ **do**
- 6: Sampling from the Gaussian distribution $x_k^s \sim \mathcal{N}(x_{k|k}^i, P_{k|k}^i).$
- 7: Compute stochastic perturbations $\Delta_{x_{k|k}^{i}}$ and $\Delta_{P_{k|k}^{i}}$ by (32) and (33), respectively.
- 8: Compute the two-side infinitesimal perturbations of $\mathcal{L}(\psi_k)$ with respect to $x_{k|k}^i$ and $P_{k|k}^i$ by (30) and (31), respectively.
- 9: Compute the estimation of gradient $\hat{\mathcal{G}}(x_{k|k}^i)$ and $\hat{\mathcal{G}}(P_{k|k}^i)$ by (36) and (37), respectively.
- 10: Update distributed iterative state estimation $x_{k|k}^{i+1}$ and the associated error covariance $P_{k|k}^{i+1}$ by (38) and (39), respectively.
- 11: Compute relative error e_r by (25).
- 12: end for
- 13: Output $x_{k|k} = x_{k|k}^{i+1}$, $P_{k|k} = P_{k|k}^{i+1}$.

5. Performance Evaluation

It is often of interest to know how closely the posterior distribution is approximated and how accurately a variable can be estimated. In this section, we present two metrics for the proposed algorithm performance evaluation: One is variational ELBO which is monotonically increasing in iteration index to measure the convergence of variational iteration. The other is PCRLB which provides a lower bound on the mean square error of system state estimation [27]. In this section, we present the general forms of ELBO and PCRLB both of DVBKF-NG and DVBKF-SPSA.

5.1. Performance in ELBO

From (10) and (11), the iterative ELBO of distributed architecture can be rewritten as

$$\mathcal{L}\left(\psi_{k}^{i}\right) = \sum_{n=1}^{N} \mathbb{E}_{q}\left[\log p(\mathcal{Z}_{n,k}|x_{k})\right] + \mathbb{E}_{q}\left[\log p(x_{k})\right] - \mathbb{E}_{q}\left[\log q\left(x_{k}|\psi_{k}^{i}\right)\right]$$
(40)

After computing the log-likelihood expectations in (40) under the Gaussian assumption, we present the iterative ELBO of DVBKF (see the derivation in Appendix A) over the sensor network, as follows

$$\mathcal{L}\left(\psi_{k}^{i}\right) = -\frac{1}{2} \left\{ \sum_{n=1}^{N} \operatorname{tr}\left(\mathcal{R}_{n,k}^{-1}\left(\left(\mathcal{Z}_{n,k} - H_{n,x_{k|k}^{i}}x_{k|k}^{i}\right)\left(\mathcal{Z}_{n,k} - H_{n,x_{k|k}^{i}}x_{k|k}^{i}\right)^{\mathrm{T}} + H_{n,x_{k|k}^{i}}P_{k|k}^{i}H_{n,x_{k|k}^{i}}^{\mathrm{T}}\right)\right) + \operatorname{tr}\left(P_{k|k-1}^{-1}\left(\left(x_{k|k}^{i} - x_{k|k-1}\right)\left(x_{k|k}^{i} - x_{k|k-1}\right)^{\mathrm{T}} + P_{k|k}^{i}\right)\right) + \log\left(|P_{k|k-1}||P_{k|k}^{i}|^{-1}\prod_{n=1}^{N}|\mathcal{R}_{n,k}|\right) + \sum_{n=1}^{N} D_{n,z}\log(2\pi) - D_{x}\right\}$$

$$(41)$$

in which $D_{n,z}$ and D_x are the dimensions of $\mathcal{Z}_{n,k}$ and x_k , and $H_{n,x_{k|k}^i} = \frac{\partial \mathcal{H}_{n,k}(x_k)}{\partial x_k}|_{x_k = x_{k|k}^i}$.

5.2. Performance in PCRLB

The PCRLB provides a theoretical lower bound for the estimation problem under a distributed Bayesian framework. It has the following defined form [55]

$$P_{k+1|k+1} \triangleq \mathbb{E}_{q(x_k|\psi_k)} \left[\left(x_k - x_{k|k} \right) \left(x_k - x_{k|k} \right)^{\mathrm{T}} \right] \ge J_{k+1}^{-1}$$

$$\tag{42}$$

where J_{k+1} is the posterior Fisher information matrix, recursively computed by

$$J_{k+1} = D_k^{22} - D_k^{21} \left(J_k + D_k^{11} \right)^{-1} D_k^{12}.$$
(43)

The terms in (43) can be expressed by

$$D_{k}^{11} = \mathbb{E} \left[-\nabla_{x_{k}} \nabla_{x_{k}}^{T} \log p(x_{k+1}|x_{k}) \right]$$

$$D_{k}^{12} = \mathbb{E} \left[-\nabla_{x_{k}} \nabla_{x_{k+1}}^{T} \log p(x_{k+1}|x_{k}) \right]$$

$$D_{k}^{21} = \mathbb{E} \left[-\nabla_{x_{k+1}} \nabla_{x_{k}}^{T} \log p(x_{k+1}|x_{k}) \right] = \left[D_{k}^{12} \right]^{T}$$

$$D_{k}^{22} = \mathbb{E} \left[-\nabla_{x_{k+1}} \nabla_{x_{k+1}}^{T} \log p(x_{k+1}|x_{k}) \right] + \mathbb{E} \left[-\nabla_{x_{k+1}} \nabla_{x_{k+1}}^{T} \log p(\mathcal{Z}_{k+1}|x_{k+1}) \right].$$
(44)

From (44), we can know that the terms of D_k^{11} , D_k^{12} and D_k^{21} are related only to the system model and irrelated to the fusion architecture of the sensor network. With the assumption of independent measurements, $\log p(\mathcal{Z}_{k+1}|x_{k+1}) = \sum_{n=1}^{N} \log p(\mathcal{Z}_{n,k+1}|x_{k+1})$.

$$D_{k}^{22} = \mathbb{E}\Big[-\nabla_{x_{k+1}}\nabla_{x_{k+1}}^{\mathrm{T}}\log p(x_{k+1}|x_{k})\Big] + \sum_{n=1}^{N}\mathbb{E}\Big[-\nabla_{x_{k+1}}\nabla_{x_{k+1}}^{\mathrm{T}}\log p(\mathcal{Z}_{n,k+1}|x_{k+1})\Big].$$
(45)

After computing the gradients in (44) and (45) by linearizing $h_k(x_k)$ and $f_{k|k-1}(x_{k-1|k-1})$, the PCRLB (see the derivation in Appendix B) of DVBKF has the following form

$$P_{k|k}^{-1} = \left(Q_{k-1} + F_{k|k-1}P_{k-1|k-1}F_{k|k-1}^{\mathrm{T}}\right)^{-1} + \sum_{n=1}^{N} \left(H_{n,x_{k|k}^{i}}\right)^{1} \mathcal{R}_{n,k}^{-1}H_{n,x_{k|k}^{i}}.$$
 (46)

5.3. Remarks

1. From (40) and (41), it is observed that the ELBO values both of DVBKF-NG and DVBKF-SPSA are related to measurements, prior, variational distribution and the number of iterations. Generally, it is assumed that the prior of a given objective system is known. Therefore, choosing an appropriate form of parameterized variational distribution, increasing the number of iterations and providing more measurement are of the essence to maximize the ELBO and lead a close approximation of the posterior. However, the balance between computation cost and accuracy should be considered according to practical applications.

- The PCRLB is an important metric to evaluate the accuracy of estimation algorithms. On the one hand, it is determined by models of dynamic systems and measurement systems. On the other hand, similarly to the ELBO, the PCRLB values both of DVBKF-NG and DVBKF-SPSA are still related to the number of iterations and the amount of measurement.
- 3. The two metrics are inextricably linked with each other: The former is the means and the latter is the goal. ELBO maximization means approximating the posterior distribution closely by an iterative variational distribution, which can lead the PCRLB to a lower trend.

6. Numerical Simulation

In this section, we present a scenario of 2-D maneuvering target tracking over a sensor network with Doppler-only measurement to illustrate the performance of DVBKF-NG and DVBKF-SPSA. We also present the NG-based and SPSA-based optimizations for centralized fusion architecture, named CVBKF-NG and CVBKF-SPSA, which can be utilized, respectively, as benchmarks corresponding to DVBKF-NG and DVBKF-SPSA for performance comparison.

From the system observability theory, we can know that target state with Doppler-only measurement only can be observable after collecting measurements from at least three Doppler sensors with different fixed locations. Thus, the measurements from neighbors along with the measurements of local sensors are used to estimate variables in distribution fusion architecture.

6.1. Performance Metrics

The estimation performance of the proposed algorithms is measured by root-meansquared error (RMSE), 3σ rule and the mean running overhead with 1000 Monte Carlo simulations. The RMSE in range R_k^{RMSE} , RMSE in radical velocity V_k^{RMSE} , 3σ in range $R_k^{3\sigma}$, 3σ in radical velocity $V_k^{3\sigma}$ and mean running time T_{mean} are given, respectively, as follows

$$R_{k}^{\text{RMSE}} = \sqrt{\frac{1}{M} \sum_{m=1}^{M} ||x_{k}^{p} - x_{k|k}^{p}||^{2}}$$

$$V_{k}^{\text{RMSE}} = \sqrt{\frac{1}{M} \sum_{m=1}^{M} ||x_{k}^{v} - x_{k|k}^{v}||^{2}}$$

$$R_{k}^{3\sigma} = 3\sqrt{\frac{1}{M} \sum_{m=1}^{M} (||x_{k}^{p}|| - ||x_{k|k}^{p}||)^{2}}$$

$$V_{k}^{3\sigma} = 3\sqrt{\frac{1}{M} \sum_{m=1}^{M} (||x_{k}^{v}|| - ||x_{k|k}^{v}||)^{2}}$$

$$T_{\text{mean}} = \frac{1}{MK} \sum_{m=1}^{M} \sum_{k=1}^{K} t_{m,k}$$

where $x_k = [x_k^p x_k^v]^T$, $x_k^p = [x y]^T$ and $x_k^v = [\dot{x} \dot{y}]^T$ denote the true position vector and velocity vector of the target, where $x_{k|k}^p$ and $x_{k|k}^v$ denote the associated estimation, respectively, $\|\cdot\|$ denotes the Euclidean norm and $t_{m,k}$ is the running time at the *k*th estimation in the *m*th Monte Carlo simulation. The values of the parameters in the proposed algorithms are given in Table 1.

Table 1. The values of parameters.

Parameter	a	b	с	α	γ	$c_{x_{k k}^{i}}$	$c_{P^i_{k k}}$	$a_{x^i_{k k}}$	$a_{P^i_{k k}}$
Value	0.01	20	100	1	0.16666701	$\frac{c}{i\gamma}$	$0.01c_{x^i_{k k}}$	$\frac{a}{(i+b)^{\alpha}}$	$0.001a_{x^i_{k k}}$

6.2. Simulation Setup

In this scenario, we consider a sensor network which consists of 20 Doppler-only sensors located in a square of 120 m × 120 m randomly, as shown in Figure 2. The communication capability of each sensor node is 50 m. The target is maneuvering with dynamic multiple models F_k^j , $j \in \{1, 2\}$ given as follows.

$$F_{k}^{j} = \begin{bmatrix} 1 & 0 & \sin(\theta_{j}T)/\theta_{j} & (\cos(\theta_{j}T) - 1)/\theta_{j} \\ 0 & 1 & (1 - \cos(\theta_{j}T))/\theta_{j} & \sin(\theta_{j}T)/\theta_{j} \\ 0 & 0 & \cos(\theta_{j}T) & -\sin(\theta_{j}T) \\ 0 & 0 & \sin(\theta_{j}T) & \cos(\theta_{j}T) \end{bmatrix}$$
(47)

$$\begin{cases} j = 1, \quad k \in \{ [1, 14), [18, 30), [34, 49), [53, 71), [75, 90) \}; \\ j = 2, \quad k \in \{ [14, 18), [30, 34), [49, 53), [71, 75) \}. \end{cases}$$
(48)

where the turn rates are $\theta_1 = -9.8N_1/v_k$ and $\theta_2 = -\theta_1$, $N_1 = 0.2$ is the overloads of target maneuvering and the scan period is T = 0.2 s. In this manuscript, the target state is represented as $x_k = [x_k^p \ x_k^v]^T$, where $x_k^p = [x \ y]^T$ and $x_k^v = [\dot{x} \ \dot{y}]^T$ denote the true position and velocity of the target. Both measurements z_k and x_k are defined in a macro sense, not as one state or one measurement. In the simulation, the state of velocity is updated in the coordinates. The initial state estimation is $x_{1|1} = [-40 \text{ m} - 40 \text{ m} 3.5 \text{ m/s} 0 \text{ m/s}]^T$ and its estimated error covariance is given as $P_{1|1} = \text{diag}([\delta_1^2 \ \delta_1^2 \ \delta_2^2 \ \delta_2^2])$, where $\delta_1 = \sqrt{1.5} \text{ m}$ and $\delta_2 = \sqrt{0.03} \text{ m/s}$. System noise covariance is given as $Q_k = \text{diag}([q_1^2 \ q_1^2 \ q_2^2 \ q_2^2])$, where $q_1 = 0.01 \text{ m}$ and $q_2 = 0.01 \text{ m/s}$.

Sensor measurements are described by a nonlinear equation of Doppler shift between the target and each of the sensors, and measurement noise covariances are time-varying because of target motion. The measurement function and measurement noise covariances are given by [56,57]

$$h_k^{(j)}(x_k) = \frac{2\left(x_k^p - S^n\right)||x_k^v||}{\|x_k^p - S^n\|}$$
(49)

$$R_{n,k}^j = \frac{3}{\pi^2 T_d^2 \mathbf{R}_{\mathrm{SNR}}} \tag{50}$$

Equations (49) and (50) represent the measurement model of the sensors, where (49) is the mapping from state to measurement and represents the Doppler shift between the target and each of the sensors. Since the radial velocity of the target is related to Doppler shift, we update it by using the velocity in the coordinates. Equation (50) represents the associated time-varying measurement variance because of target motion, where $T_d = 1 \,\mu s$ is the pulse Doppler waveform width, R_{SNR} is the signal-to-noise ratio (SNR) and $R_{SNR} = \frac{P_e}{P_n}$. For a given transmitted waveform with unit energy, the energy of the received signal $P_e = \frac{P_i GA_e \delta}{(4\pi)^2 r^4}$, where *r* is radar radius, *G* and δ are the radar antenna and the cross-section area, respectively, and P_t and A_e are the echo power and the effective receiving area of radar antenna, respectively. The relationship between measurement noise covariance and the range from radar to target is presented in Appendix C for a Gaussian noise $P_n = \frac{kT_s}{2}$, where $T_s = 290 \,\mathrm{K}$ and $k = 1.3806 \times 10^{-23} \,\mathrm{J/K}$ are the temperature in degrees Kelvin and the Boltzmann constant, respectively.

Figure 3a shows the number of neighbors of each sensor node. It is clearly the 4th sensor has the most neighbors, numbering 12, then the 1st, 6th, 12th and 14th sensors,

and the 17th sensor has the least number of neighbors. Besides the proposed iterative algorithms, we use the interacting multiple model (IMM) method to achieve the model transfer in simulation. In distributed architecture, the sensor with best estimation accuracy is chosen for output. From Figure 3b, we can observe that sensor 4 has the largest number of outputs, then sensors 6, 14, 12 and 1. To a certain degree, this reflects that multi-sensor measurements have the advantage of improving accuracy.



Figure 2. Sensor network scenario.



Figure 3. (a) The number of the neighbors of each sensor. (b) The output sensor against scan index.

6.3. Simulation Results and Analysis

Figure 4a,b shows the PCRLB and RMSE curves in range and radical velocity, respectively. From the view of fusion architecture, centralized fusion architecture has lower RMSE curves and PCRLB curves than the distributed one. Namely, CEKF, CVBKF-NG and CVBKF-SPSA are, respectively, better than the associated distributed DEKF, DVBKF-NG and DVBKF-SPSA, since the state estimation is updated by using the measurements from the sensor and its neighbor in distributed architecture. In contrast, the measurements collected from all sensors are used to update state estimation in centralized fusion architecture.

From the view of optimization methods, we can observe that the RMSE curves of DVBKF-NG and DVBKF-SPSA are better than DEKF which does not use any optimization methods. Besides iterative linearization, random perturbation sampling which can capture more information from nonlinear measurement is used in DVBKF-SPSA. As a result, the

measurement is utilized effectively to improve the accuracy performance. As shown in Figure 4, the RMSE curves obtained by using SPSA are lower than those obtained by NG optimization both in range and radical velocity. However, from Figure 4b, it is also found that the proposed algorithms and comparison algorithms in distributed architecture are more sensitive than those in the centralized one when the dynamic model transfers. The comparisons of the PCRLB and RMSE of estimated position and velocity in coordinates are given in Figures 5 and 6, respectively.



Figure 4. The comparison of PCRLB and RMSE. (a) The RMSE of range. (b) The RMSE of velocity.



Figure 5. The comparison of the PCRLB and RMSE of estimated position in coordinates. (**a**) The RMSE on x-axis. (**b**) The RMSE on y-axis.



Figure 6. The comparison of the PCRLB and RMSE of estimated velocity in coordinates. (**a**) The RMSE on x-axis. (**b**) The RMSE on y-axis.

More quantitative RMSE comparison in centralized architecture and distributed architecture can be seen in Tables 2 and 3, from which it can be seen that the RMSE curves of the proposed algorithms DVBKF-NG and DVBKF-SPSA are close to those in the centralized architecture. It is also clear that the PCRLB in distributed architecture is slightly bigger than

that in centralized architecture, because only a part of the sensors are used in distributed architecture. However the computational cost in distributed architecture is much smaller than that in the centralized one. Tables 4 and 5 give the computational cost comparison of the algorithms in centralized architecture and distributed architecture mentioned above.

Table 2. The comparison of mean RMSE in centralized architecture.

Algorithm	CEKF	CVBKF-NG	CVBKF-SPSA	PCRLB
Position (m)	0.6285	0.4636	0.4212	0.3411
Velocity (m/s)	0.0762	0.0695	0.0667	0.0627

Table 3. The comparison of mean RMSE in distributed architecture.

Algorithm	DEKF	DVBKF-NG	DVBKF-SPSA	PCRLB
Position (m)	0.6524	0.5373	0.4388	0.3277
Velocity (m/s)	0.0957	0.0837	0.0816	0.0560

Table 4. The comparison of computational cost in centralized architecture.

Algorithm	CEKF	CVBKF-NG	CVBKF-SPSA
Time (s)	0.0725	0.1827	0.3231

Table 5. The comparison of computational cost in distributed architecture.

Algorithm	DEKF	DVBKF-NG	DVBKF-SPSA
Time (s)	0.0614	0.1047	0.2102

The 3σ bound is another evaluation of target tracking accuracy. A solid line indicates the estimation error of associated algorithm. A dashed line indicates the 3σ error of the algorithm which presented as a solid line with the same color. Figure 7 presents the comparison of the 3σ bound. It is seen that the algorithms with NG or SPSA have smaller 3σ bound in range than CEKF and DEKF. The radical velocities 3σ of the proposed algorithms with NG and SPSA are slightly bigger than that of CEKF at some scans in Figure 7b. But we can observe that the radical velocities 3σ of the proposed algorithms are robust for the maneuvering target tracking both in centralized fusion architecture and distributed fusion architecture. The comparisons of 3σ error of estimated position and velocity in coordinates are given in Figures 8 and 9, respectively. A solid line indicates the estimation error. A dashed line indicates the 3σ error of the algorithm which presented as a solid line with the same color. The quantitative 3σ comparison is given in Tables 6 and 7.

As mentioned above, the minimization of KLD between variational distribution and posterior distribution is equivalent to maximization variational ELBO. To observe the changes in ELBO and KLD with iteration clearly, we present Figures 10 and 11 to illustrate the normalized ELBO and KLD of the proposed DVBKF-NG and DVBKF-SPSA, respectively. In Figure 10, each line denotes the ELBO in one scan with 200 iterations, and each line in Figure 11 denotes KLD in one scan with 200 iterations. The results in Figures 10 and 11 verify our standpoints by the following fact: the ELBO curves increase with the number of iterations, corresponding to a decrease in the KLD.



Figure 7. The 3σ comparison. (a) The 3σ of range. (b) The 3σ of velocity.



Figure 8. The 3σ comparison of estimated position. (a) The 3σ on x-axis. (b) The 3σ on y-axis.



Figure 9. The 3σ comparison of estimated velocity. (a) The 3σ on x-axis. (b) The 3σ on y-axis.

Table 6. The comparison of 3σ centralized architecture.

Algorithm	CEKF	CVBKF-NG	CVBKF-SPSA
Position (m)	1.3833	1.0086	0.9196
Velocity (m/s)	0.1537	0.1550	0.1474

Table 7. The comparison of 3σ distributed architecture.

Algorithm	DEKF	DVBKF-NG	DVBKF-SPSA
Position (m)	1.4430	1.2050	0.9606
Velocity (m/s)	0.1835	0.1684	0.1714



Figure 10. ELBO. (a) DVBKF-NG. (b) DVBKF-SPSA.



Figure 11. KLD. (a) DVBKF-NG. (b) DVBKF-SPSA.

7. Discussion

In this paper, we address the problem of improving the accuracy for maneuvering target tracking in sensor networks. Two kinds of optimization methods, NG and SPSA, are introduced to maximize the distributed ELBO where the joint likelihood is partitioned approximately into several simple marginal likelihoods by variational mean field, formulating the algorithms of DVBKF-NG and DVBKF-SPSA. Moreover, the performance metrics, both of ELBO and PCRLB, are presented over different iterative indexes. In addition, a maneuvering target tracking scenario over a sensor network is given to verify the performance of the proposed algorithms. From the view of fusion architecture, centralized fusion architecture has lower RMSE curves and PCRLB curves than the distributed one. From the view of optimization methods, the simulation results show that the RMSE curves of the proposed algorithms are better than those which do not use any optimization methods.

For future work, we plan to adopt VB for the robust estimation of sensor networks. For example, outliers always lead to heavy-tailed and asymmetric distributions which are apt to lead to large estimation errors. It is expected for novel methods to mitigate the adverse influence and VB in which an unsolvable distribution can be approximated by a parameterized distribution is an appropriate method at this point. We also plan to develop VB for multiple passive sensor placement. For example, in a bearings-only target tracking system, tracking accuracy is highly dependent on the locations of the bearings-only sensors. Therefore, it is desirable to schedule the sensors' moving trajectories in a way to achieve a minimized tracking error at a future time.

Author Contributions: Conceptualization, Y.H. and Q.P.; methodology, Y.H. and B.D.; software, Y.H. and Z.G.; validation, B.D., Z.G. and M.L.; writing—original draft preparation, Y.H. and B.D.; writing—review and editing, Y.H., Q.P., L.C. and Z.G.; supervision, Q.P.; project administration, Q.P.; funding acquisition, Q.P. and Y.H. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the National Natural Science Foundation of China (Grant No. 427 61790552).

Institutional Review Board Statement: Not Applicable.

Data Availability Statement: No new data were created or analyzed in this study. Data sharing is not applicable to this article.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A. Derivation of ELBO of DVBKF

With Gaussian assumption, the log-likelihood expectations in (40) are computed as (A1), in which $\partial \mathcal{H}_{-1}(x_{1})$

$$H_{n,x_{k+1|k+1}^i} = \frac{\partial \mathcal{H}_{n,k}(x_k)}{\partial x_k}|_{x_k = x_{k+1|k+1}^i}$$

The expectation of log-a priori distribution and the expectation of log-variational distribution are given as (A2) and (A3), respectively. As a result, the distributed ELBO $\mathcal{L}(\psi_k^i)$ in (40) is calculated as (A4).

$$\begin{split} \mathbb{E}_{q}[\log p(\mathcal{Z}_{n,k}|x_{k})] &\approx -\frac{1}{2}\log\left((2\pi)^{D_{n,k}}|\mathcal{R}_{n,k}|\right) \\ &\quad -\frac{1}{2}\mathrm{tr}\left(\mathcal{R}_{n,k}^{-1}\mathbb{E}_{q}\left[\left(\mathcal{Z}_{n,k}-H_{n,x_{k|k}^{i}}(x_{k|k}^{i}+\bar{x}_{k|k})\right)\left(\mathcal{Z}_{n,k}-H_{n,x_{k|k}^{i}}(x_{k|k}^{i}+\bar{x}_{k|k})\right)^{\mathrm{T}}\right]\right) \\ &\quad = -\frac{1}{2}\log\left((2\pi)^{D_{n,k}}|\mathcal{R}_{n,k}|\right) \\ &\quad -\frac{1}{2}\mathrm{tr}\left(\mathcal{R}_{n,k}^{-1}\left((\mathcal{Z}_{n,k}-H_{n,x_{k|k}^{i}}x_{k|k}^{i})(\mathcal{Z}_{n,k}-H_{n,x_{k|k}^{i}}x_{k|k}^{i})^{\mathrm{T}}+H_{n,x_{k|k}^{i}}P_{k|k}^{i}H_{n,x_{k|k}^{i}}^{\mathrm{T}}\right)\right) \\ &\quad \mathbb{E}_{q}[\log p(x_{k})] \approx -\frac{1}{2}\log\left((2\pi)^{D_{x}}|\mathcal{P}_{k|k-1}|\right) - \frac{1}{2}\mathrm{tr}\left(\mathcal{P}_{k|k-1}^{-1}\mathbb{E}_{q}\left[\left(x_{k}-x_{k|k-1}\right)\left(x_{k}-x_{k|k-1}\right)^{\mathrm{T}}\right]\right) \\ &\quad = -\frac{1}{2}\log\left((2\pi)^{D_{x}}|\mathcal{P}_{k|k-1}|\right) - \frac{1}{2}\mathrm{tr}\left(\mathcal{P}_{k|k-1}^{-1}\mathbb{E}_{q}\left[\left(x_{k|k}^{i}+\bar{x}_{k}\right)\left(x_{k|k}^{i}+\bar{x}_{k}\right)^{\mathrm{T}}\right. \right) \\ &\quad -\left(x_{k|k}^{i}+\bar{x}_{k}\right)x_{k|k-1}^{\mathrm{T}}-x_{k|k-1}\left(x_{k|k}^{i}+\bar{x}_{k}\right)^{\mathrm{T}}+x_{k|k-1}x_{k|k-1}^{\mathrm{T}}\right]\right) \\ &\quad = -\frac{1}{2}\log\left((2\pi)^{D_{x}}|\mathcal{P}_{k|k-1}|\right) - \frac{1}{2}\mathrm{tr}\left(\mathcal{P}_{k|k-1}^{-1}\mathcal{P}_{k|k}^{i}\right) \\ &\quad -\frac{1}{2}\mathrm{tr}\left(\mathcal{P}_{k|k-1}^{-1}\left(\left(x_{k|k}^{i}-x_{k|k-1}\right)\left(x_{k|k}^{i}-x_{k|k-1}\right)^{\mathrm{T}}\right)\right) \\ &\quad = -\frac{1}{2}\log\left((2\pi)^{D_{x}}|\mathcal{P}_{k|k}|\right) - \frac{1}{2}\mathrm{tr}\left((\mathcal{P}_{k|k}^{i})^{-1}\mathbb{E}_{q}\left[\left(x_{k}-x_{k|k}^{i}\right)\left(x_{k}-x_{k|k}^{i}\right)^{\mathrm{T}}\right]\right) \\ &\quad = -\frac{1}{2}\log\left((2\pi)^{D_{x}}|\mathcal{P}_{k|k}|\right) - \frac{1}{2}\mathrm{tr}\left((\mathcal{P}_{k|k}^{i})^{-1}\mathbb{E}_{q}\left[\left(x_{k}-x_{k|k}^{i}\right)\left(x_{k}-x_{k|k}^{i}\right)^{\mathrm{T}}\right)\right) \\ &\quad = -\frac{1}{2}\log\left((2\pi)^{D_{x}}|\mathcal{P}_{k|k}^{i}|\right) - \frac{D_{x}}{2}. \end{split}$$

$$\mathcal{L}\left(\psi_{k}^{i}\right) = -\frac{1}{2}\left\{\sum_{n=1}^{N}\mathrm{tr}\left(\mathcal{R}_{n,k}^{-1}\left((\mathcal{Z}_{n,k}-H_{n,x_{k|k}^{i}}x_{k|k}^{i})(\mathcal{Z}_{n,k}-H_{n,x_{k|k}^{i}}x_{k|k}^{i})^{\mathrm{T}}+H_{n,x_{k|k}^{i}}\mathcal{P}_{k|k}^{i}H_{n,x_{k|k}^{i}}\right)\right)$$

$$+ \operatorname{tr}\left(P_{k|k-1}\left((x_{k|k}^{i} - x_{k|k-1})(x_{k|k}^{i} - x_{k|k-1})^{2} + P_{k|k}^{i}\right)\right) + \log\left(|P_{k|k-1}||P_{k|k}^{i}|^{-1}\prod_{n=1}^{N}|\mathcal{R}_{n,k}|\right) + \sum_{n=1}^{N}D_{n,2}\log(2\pi) - D_{x}\bigg\}.$$
(A4)

Appendix B. Derivation of the PCRLB of DVBKF

After the linearization of $h_k(x_k)$ and $f_k(x_{k-1|k-1})$, the above equations can be written as

$$D_{k}^{11} = F_{k+1|k}^{T} Q_{k}^{-1} F_{k+1|k}$$

$$D_{k}^{12} = -F_{k+1|k}^{T} Q_{k}^{-1}$$

$$D_{k}^{21} = -Q_{k}^{-1} F_{k+1|k}$$

$$D_{k}^{22} = Q_{k}^{-1} + \sum_{n=1}^{N} \left(H_{n,x_{k+1|k+1}^{i}} \right)^{T} \mathcal{R}_{k+1}^{-1} H_{n,x_{k+1|k+1}^{i}}.$$
(A5)

According to the matrix inversion lemma, the Fisher information in Equation (43) can be rewritten as

$$J_{k+1} = \left(Q_k + F_{k+1|k}J_k^{-1}F_{k+1|k}^{\mathrm{T}}\right)^{-1} + \sum_{n=1}^N \left(H_{n,x_{k+1|k+1}^i}\right)^{\mathrm{T}}\mathcal{R}_{k+1}^{-1}H_{n,x_{k+1|k+1}^i}.$$
 (A6)

Therefore it yields the distributed PRLB,

$$P_{k+1|k+1}^{-1} = \left(Q_k + F_{k+1|k}P_{k|k}F_{k+1|k}^{\mathrm{T}}\right)^{-1} + \sum_{n=1}^{N} \left(H_{n,x_{k+1|k+1}^i}\right)^{\mathrm{T}} \mathcal{R}_{k+1}^{-1} H_{n,x_{k+1|k+1}^i}.$$
 (A7)

Appendix C. The Relationship between Measurement Noise Covariance and the Range from Radar to Target

The standard deviation of the time-varying measurement noise of radar is $\sigma \ge \frac{\sqrt{3}\Delta f}{\pi\sqrt{R_{SNR}}}$, which is given as Equation (18.46) in reference [56]; the Δf is the frequency domain resolution and $\Delta f = \frac{1}{T_d}$. The time-varying measurement noise can be rewritten as follows

$$R = \sigma^2 \ge \frac{3\Delta f^2}{\pi^2 R_{SNR}} = \frac{3}{\pi^2 T_d^2 R_{SNR}}.$$
 (A8)

Assume a radar transmit power is P_t ; the power density is written as $S_1 = \frac{P_t}{4\pi r^4}$. Define the radar antenna gain *G*; the power density can be rewritten as

$$S_2 = S_1 G = \frac{P_t G}{4\pi r^2}.$$
 (A9)

Set the cross-section area as δ ; the received power density S_r is calculated by

$$S_r = S_2 \frac{\delta}{4\pi r^2} = \frac{P_t G \delta}{\left(4\pi r^2\right)^2}.$$
(A10)

Assume the effective receiving area of radar antenna A_e ; the echo power is given as

$$P_e = A_e S_r = \frac{P_t G A_e \delta}{(4\pi)^2 r^4} \propto \frac{1}{r^4}.$$
(A11)

since signal-to-noise ratio is defined as $R_{SNR} = 10 \log_{10} \left(\frac{P_e}{P_n}\right)$, where P_n is noise power. For Gaussian noise $P_n = \frac{kT_s}{2}$, where *k* and T_s are Boltzman constant and Kelvin temperature, respectively, it yields R_{SNR} ,

$$R_{SNR} = 10 \log_{10} \left(\frac{P_t G A_\ell \delta}{8\pi^2 k T_s r^4} \right). \tag{A12}$$

References

- 1. Gu, D. A game theory approach to target tracking in sensor networks. *IEEE Trans. Syst. Man Cybern. B Cybern.* 2011, 41, 1–13. [CrossRef] [PubMed]
- Hu, X.; Yang, L.; Xiong, W. A novel wireless sensor network frame for urban transportation. *IEEE Internet Things J.* 2015, 2, 586–595. [CrossRef]

- 3. Silva, B.; Fisher, R.M.; Kumar, A.; Hancke, G.P. Experimental link quality characterization of wireless sensor networks for underground monitoring. *IEEE Trans. Ind. Inform.* **2015**, *11*, 1099–1110. [CrossRef]
- 4. Vu, T.T.; Rahmani, A.R. Distributed consensus-based Kalman filter estimation and control of formation flying spacecraft: Simulation and validation. *Biulleten Eksp. Biol. I Meditsiny* **2013**, *37*, 7–12.
- Cetin, M.; Chen, L.; Fisher, J.W., III; Ihler, A.T.; Moses, R.L.; Wainwright, M.J.; Willsky, A.S. Distributed fusion in sensor networks. *IEEE Signal Process. Mag.* 2006, 23, 42–55. [CrossRef]
- 6. He, S.; Shin, H.S.; Xu, S.; Tsourdos, A. Distributed estimation over a low-cost sensor network: A review of state-of-the-art. *IEEE Trans. Syst. Man Cybern. B Cybern.* 2020, 54, 21–43. [CrossRef]
- Yang, X.; Zhang, W.A.; Liu, A.; Yu, L. Linear fusion estimation for range-only target tracking with nonlinear transformation. *IEEE Trans. Ind. Inform.* 2020, 16, 6403–6412. [CrossRef]
- 8. Li, T.; Fan, H.; García, J.; Corchado, J.M. Second-order statistics analysis and comparison between arithmetic and geometric average fusion: Application to multi-sensor target tracking. *Inf. Fusion* **2019**, *52*, 233–243. [CrossRef]
- 9. Lin, Y.; Chen, B.; Varshney, P.K. Decision fusion rules in multi-hop wireless sensor networks. *IEEE Trans. Aerosp. Electron. Syst.* **2005**, *41*, 475–488.
- Liggins, M.E., II; Chong, C.Y.; Kadar, I.; Alford, M.G.; Vannicola, V.; Thomopoulos, S. Distributed fusion architectures and algorithms for target tracking. *Proc. IEEE* 1997, *85*, 95–107. [CrossRef]
- 11. García-Ligero, M.J.; Hermoso-Carazo, A.; Linares-Pérez, J. Distributed and centralized fusion estimation from multiple sensors with Markovian delays. *Appl. Math. Comput.* 2012, 219, 2932–2948. [CrossRef]
- Hounkpevi, F.O.; Yaz, E.E. Minimum variance generalized state estimators for multiple sensors with different delay rates. *Signal Process.* 2007, 87, 602–613. [CrossRef]
- 13. Linares-Pérez, J.; Hermoso-Carazo, A.; Caballero-éguila, R.; Jiménez-López, J.D. Least-squares linear filtering using observations coming from multiple sensors with one- or two-step random delay. *Signal Process.* **2009**, *89*, 2045–2052. [CrossRef]
- 14. Hu, Z.; Hu, Y.; Jin, Y.; Zheng, S. Measurement bootstrapping Kalman filter. *Opt.—J. Light Electron. Opt.* **2016**, *127*, 2094–2101. [CrossRef]
- 15. Bhuvana, V.P.; Preissl, C.; Tonello, A.M.; Huemer, M. Multi-sensor information filtering with information-based sensor selection and outlier rejection. *IEEE Sens. J.* 2018, 18, 2442–2454. [CrossRef]
- Alshamaa, D.; Mourad-Chehade, F.; Honeine, P. Decentralized kernel-based localization in wireless sensor networks using belief functions. *IEEE Sens. J.* 2019, 19, 4149–4159. [CrossRef]
- 17. Yang, X.; Zhang, W.A.; Yu, L. A bank of decentralized extended information filters for target tracking in event-triggered WSNs. *IEEE Trans. Syst. Man Cybern. Syst.* 2020, *50*, 381–390. [CrossRef]
- Stamatescu, G.; Stamatescu, I.; Dragana, C.; Popescu, D. Large scale heterogeneous monitoring system with decentralized sensor fusion. In Proceedings of the 2015 IEEE 8th International Conference on Intelligent Data Acquisition and Advanced Computing Systems: Technology and Applications (IDAACS), Warsaw, Poland, 24–26 September 2015; IEEE: Piscataway, NJ, USA, 2015; pp. 2–5.
- Hua, J.; Li, C. Distributed variational Bayesian algorithms over sensor networks. *IEEE Trans. Signal Process.* 2016, 64, 783–798. [CrossRef]
- Guldogan, M.B. Consensus Bernoulli filter for distributed detection and tracking using multi-static doppler shifts. *IEEE Signal Process. Lett.* 2014, 21, 672–676. [CrossRef]
- 21. Yu, Y. Consensus-based distributed mixture Kalman filter for maneuvering target tracking in wireless sensor networks. *IEEE Trans. Veh. Technol.* **2016**, *65*, 8669–8681. [CrossRef]
- 22. Zhang, H.; Zhou, X.; Wang, Z.; Yan, H.; Sun, J. Adaptive consensus-based distributed target tracking with dynamic cluster in sensor networks. *IEEE Trans. Cybern.* 2019, *49*, 1580–1591. [CrossRef] [PubMed]
- Gustafsson, F.; Hendeby, G. Some relations between extended and unscented Kalman filters. *IEEE Trans. Signal Process.* 2012, 60, 545–555. [CrossRef]
- 24. Julier, S.; Uhlmann, J.; Durrantwhyte, H.F. A new method for nonlinear transformation of means and covariances in filters and estimates. *IEEE Trans. Autom. Control* 2000, 45, 477–482. [CrossRef]
- 25. Ito, K.; Xiong, K. Gaussian filters for nonlinear filtering problems. IEEE Trans. Autom. Control 2000, 45, 910–927. [CrossRef]
- 26. Arasaratnam, I.; Haykin, S. Cubature Kalman filters. IEEE Trans. Autom. Control 2009, 54, 1254–1269. [CrossRef]
- Hu, X.; Bao, M.; Zhang, X.; Guan, L.; Hu, Y. Generalized iterated Kalman filter and its performance evaluation. *IEEE Trans. Signal Process.* 2015, 63, 3204–3217. [CrossRef]
- 28. Khamseh, H.B.; Ghorbani, S.; Janabi-Sharifi, F. Unscented Kalman filter state estimation for manipulating unmanned aerial vehicles. *Aerosp. Sci. Technol.* **2019**, *92*, 446–463. [CrossRef]
- Andrieu, C.; Doucet, A.; Holenstein, R. Particle Markov Chain Monte Carlo methods. J. R. Stat. Soc. Ser. B (Stat. Methodol.) 2010, 72, 269–342. [CrossRef]
- Cappe, O.; Godsill, S.J.; Moulines, E. An overview of existing methods and recent advances in sequential Monte Carlo. *Proc.* IEEE 2007, 95, 899–924. [CrossRef]
- Merwe, R.V.D.; Doucet, A.; Freitas, N.D.; Wan, E. The unscented particle filter. In Proceedings of the Advances in Neural Information Processing Systems 13 (NIPS 2000), Denver, CO, USA, 1 January 2000; pp. 563–569.

- 32. Aleardi, M.; Salusti, A. Markov Chain Monte Carlo algorithms for target-oriented and interval-oriented amplitude versus angle inversions with non-parametric priors and non-linear forward modellings. *Geophys. Prospect.* **2020**, *68*, 735–760. [CrossRef]
- Doucet, A.; Freitas, N.D.; Murphy, K.; Russell, S. Rao-Blackwellised particle filtering for dynamic Bayesian networks. In Proceedings of the Sixth Annual Conference on Uncertainty in Artificial Intelligence, Cambridge, MA, USA, 27–29 July 1990; pp. 176–183.
- 34. Lan, J.; Li, X.R. Nonlinear estimation based on conversion-sample optimization. Automatica 2020, 121, 109160. [CrossRef]
- 35. Patwardhan, S.C.; Narasimhan, S.; Jagadeesan, P.; Gopaluni, B.; Shah, S.L. Nonlinear Bayesian state estimation: A review of recent developments. *Control Eng. Pract.* 2012, 20, 933–953. [CrossRef]
- Jouin, M.; Gouriveau, R.; Hissel, D.; Péra, M.C.; Zerhouni, N. Particle filter-based prognostics: Review, discussion and perspectives. Mech. Syst. Signal Process. 2016, 72, 2–31. [CrossRef]
- Yean, S.; Lee, B.S.; Yeo, K.C.; Vun, C.H.; Oh, L.H. Smartphone orientation estimation algorithm combining Kalman Filter with gradient descent. *IEEE J. Biomed. Health Inform.* 2017, 22, 1421–1433. [CrossRef]
- Chen, J.; Zhu, Q.; Liu, Y. Modified Kalman filtering based multi-step-length gradient iterative algorithm for ARX models with random missing outputs. *Automatica* 2020, 118, 1093–1940. [CrossRef]
- 39. Hoffman, M.D.; Blei, D.M.; Wang, C.; Paisley, J. Stochastic variational inference. J. Mach. Learn. Res. 2013, 14, 1303–1347.
- 40. Chen, J.; Li, J.; Liu, Y. Gradient iterative algorithm for dual-rate nonlinear systems based on a novel particle filter. *J. Frankl. Inst.* **2017**, *354*, 4425–4437. [CrossRef]
- 41. Amari, S.I. Natural gradient works efficiently in learning. Neural Comput. 1998, 10, 251–276. [CrossRef]
- 42. Amari, S.I. Information Geometry and Its Applications; Springer: Berlin/Heidelberg, Germany, 2016.
- Schmitt, L.; Fichter, W. Globally valid posterior Cramér-Rao bound for tThree-dimensional bearings-only filtering. *IEEE Trans.* Aerosp. Electron. Syst. 2019, 55, 2036–2044. [CrossRef]
- 44. Ollivier, Y. Online natural gradient as a Kalman filter. Electron. J. Stat. 2018, 12, 2930–2961. [CrossRef]
- 45. Ollivier, Y. The extended Kalman filter is a natural gradient descent in trajectory space. arXiv 2019, arXiv:1901.00696v1.
- Spall, J.C. An overview of the simultaneous perturbation method for efficient optimization. *Johns Hopkins APL Tech. Dig.* 1998, 19, 482–492.
- Antal, C.; Granichin, O.; Levi, S. Adaptive autonomous soaring of multiple UAVs using simultaneous perturbation stochastic approximation. In Proceedings of the 49th IEEE Conference on Decision and Control (CDC), Atlanta, GA, USA, 15–17 December 2010; IEEE: Piscataway, NJ, USA, 2010; pp. 3656–3661.
- 48. Moon, T.K. The expectation-maximization algorithm. IEEE Signal Process. Mag. 1996, 13, 47–60. [CrossRef]
- Beal, M.J. Variational Algorithms for Approximate Bayesian Inference. Ph.D. Thesis, Cambridge University, Cambridge, UK, 2003.
 Bishop, C.M. *Pattern Pecognition and Machine Learning*; Springer: New York, NY, USA, 2006.
- 51. Lan, H.; Sun, S.; Wang, Z.; Pan, Q.; Zhang, Z. Joint target detection and tracking in multipath environment: A variational Bayesian approach. *IEEE Trans. Aerosp. Electron. Syst.* 2020, 56, 2136–2156. [CrossRef]
- 52. Zhang, C.; Bütepage, J.; Kjellström, H.; Mandt, S. Advances in variational inference. *IEEE Trans. Pattern Anal. Mach. Intell.* **2019**, *41*, 2008–2026. [CrossRef] [PubMed]
- Hu, Y.; Wang, X.; Pan, Q.; Hu, Z.; Moran, B. Variational Bayesian Kalman filter using natural gradient. *Chin. J. Aeronaut.* 2021, 35, 1–10. [CrossRef]
- 54. Absil, P.A.; Mahony, R.; Sepulchre, R. *Optimization Algorithms on Matrix Manifolds*; Princeton University Press: Princeton, NJ, USA, 2009.
- Tichavsky, P.; Muravchik, C.H.; Nehorai, A. Posterior Cramér-Rao bounds for discrete-time nonlinear filtering. *IEEE Trans. Signal Process.* 1998, 46, 1386–1396. [CrossRef]
- 56. Richards, M.A.; Scheer, J.; Holm, W.A.; Melvin, W.L. Principles of Modern Radar; Citeseer: University Park, PA, USA, 2010.
- Cheng, Y.; Wang, X.; Caelli, T.; Li, X.; Moran, B. On information resolution of radar systems. *IEEE Trans. Aerosp. Electron. Syst.* 2012, 48, 3084–3102. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.