Distributed Sequential Bayesian Estimation of a Diffusive Source in Wireless Sensor Networks

Tong Zhao, *Student Member, IEEE*, and Arye Nehorai, *Fellow, IEEE*

***Abstract*—We develop an efficient distributed sequential Bayesian estimation method for applications relating to diffusive sources—localizing a diffusive source, determining its space-time concentration distribution, and predicting its cloud envelope evolution using wireless sensor networks. Potential applications include security, environmental and industrial monitoring, as well as pollution control. We first derive the physical model of the substance dispersion by solving the diffusion equations under different environment scenarios and then integrate the physical model into the distributed processing technologies. We propose a distributed sequential Bayesian estimation method in which the state belief is transmitted in the wireless sensor networks and updated using the measurements from the new sensor node. We propose two belief representation methods: a Gaussian density approximation and a new LPG function (linear combination of polynomial Gaussian density functions) approximation. These approximations are suitable for the distributed processing in wireless sensor networks and are applicable to different sensor network situations. We implement the idea of information-driven sensor collaboration and select the next sensor node according to certain criterions, which provides an optimal subset and an optimal order of incorporating the measurements into our belief update, reduces response time, and saves energy consumption of the sensor network. Numerical examples demonstrate the effectiveness and efficiency of the proposed methods.**

***Index Terms*—Diffusive source, distributed estimation, sensor node scheduling, sequential Bayesian method, wireless sensor networks.**

I. Introduction

RECENTLY, wireless sensor networks have been the object of intensive interest in the research community [1], [2]. Inexpensive, smart nodes with multiple on-board sensors, networked through wireless lines as well as the Internet and deployed in large numbers, interact intelligently with the physical world. These systems are aimed at various applications: environmental, medical, food-safety, and habitant monitoring; assessing the "health" of machines, aerospace vehicles, and civil engineering structures; energy management; inventory control; home and building automation; homeland security; and military initiatives. One important set of applications involves monitoring diffusion phenomena. In this paper, we address the issue of developing efficient distributed parameter estimation methods and applying them to diffusive source monitoring, i.e., localizing a diffusive source, determining its space-time concentration distribution, and predicting its cloud envelope evolution, using wireless sensor networks. Examples of diffusive sources include biological agents, toxic chemicals, explosives, hazardous materials, pollutants, drugs, and temperature fields.

In our previous paper, we proposed model-based integrated biochemical sensor array processing methods for detecting and estimating dispersion in realistic environments, in which the underlying physical and statistical models were coupled with information processing to obtain the best performance [3]–[7]. However, these methods use centralized processing approaches, i.e., they assume all measured data from the sensors are transmitted to a fusion center for the processing; they cannot be directly used in our current problem due to the strict power and resource constraints of the wireless sensor networks. In a typical wireless sensor network, each sensor node operates unattended with limited battery power and limited signal-processing capability; each sensor node communicates wirelessly with other nodes in its radio communication range. These wireless communications consume a significant part of the available energy, and, hence, dominate the life of a wireless sensor network. Therefore, in this paper the essential point is that we propose a distributed parameter estimation method based on a sequential Bayesian approach, which is suitable for wireless sensor networks, and apply it to the estimation of diffusive sources. In this method, we apply information-driven dynamic collaborative information processing between sensor nodes [8], [9]; hence, we decrease the total wireless communications while maintaining the processing performance above a required threshold. The proposed method works efficiently with low computation complexity and fits the limited processing capability at each node.

We first derive the physical models for the spatial and temporal concentration distribution of the dispersed substance from a diffusion source. We obtain these models by solving the diffusion equations under different initial conditions and environmental scenarios. Our methods can be applied to different source types, such as instantaneous sources, continuous sources, or distributed sources; they can also be applied to different environments, such as infinite medium, semiinfinite medium, multilayer medium, a cylinder, or a sphere. We also consider the effects of external forces, e.g., wind, temperature, and air turbulence.

We then transform the physical models to parametric statistical measurement models and develop a distributed parameter estimation method based on the sequential Bayesian approach. The main idea is that the posterior density function, also known

Manuscript received November 24, 2005; revised June 7, 2006. This work was supported by the Department of Defense under the Air Force Office of Scientific Research MURI Grant FA9550-05-1-0443, AFOSR Grant FA9550-05-1-0018, and by the National Science Foundation Grant CCR-0330342. The associate editor coordinating the review of this paper and approving it for publication was Dr. Lang Tong.

The authors are with the Department of Electrical and Systems Engineering, Washington University, Saint Louis, MO 63130 USA (e-mail: tzhao3@ese.wustl.edu; nehorai@ese.wustl.edu).

Digital Object Identifier 10.1109/TSP.2006.889975

as the belief, of the parameters of interest is updated incrementally when new measurements are obtained, until a desired performance threshold is satisfied. The estimates of the parameters are calculated from the obtained belief according to certain criteria. Here we assume that the related parameters are stationary. In contrast with the ordinary sequential Bayesian methods in which the belief is updated sequentially in the time domain [10], in our method the belief is transmitted in the sensor network and updated incrementally in the space domain. Since the information is transmitted only between sensor nodes and their neighbors, a fusion center is not needed; hence, we realize a fully distributed estimation.

The derived dispersion models yield statistical measurement models that are highly nonlinear and non-Gaussian, especially when nuisance parameters appear. Therefore, existent recursive Bayesian estimation methods, such as the decentralized Kalman filters or Extended Kalman filters [11], [12] cannot be directly applied. Some possible alternative methods, such as grid-based methods and sequential Monte Carlo (or particle filtering) methods [13], [14], are also inapplicable because they require large data transmissions between sensor nodes. Even though some distributed particle filters are proposed [15], [16], they still have limitations, e.g., the effectiveness of these algorithms depends on some strict assumptions, and the computation complexity of these algorithms are very high [17]. Since decreasing power consumption is always an important issue in wireless sensor networks, in our method we propose to represent and transmit the belief parametrically, i.e., we approximate the belief by a family of parameterizable probability distributions; hence, we need to transmit only a limited number of parameters between sensor nodes, which decreases the communication requirements dramatically. In this paper, we propose two parametric belief representation methods: Gaussian density approximation and a new LPG function approximation. The distribution family of LPG functions is a natural generalization of the Gaussian density; hence, it provides a more accurate approximation that leads to fewer data transmission hops while maintaining the desired estimation performance. LPG function approximation approaches have the potential advantage to be used in other nonlinear, non-Gaussian recursive Bayesian estimation problems. Note that even though we focus on the estimation of diffusive sources, the proposed distributed processing methods, including belief approximation approaches and node selection schemes also provide a general framework for other applications in wireless sensor networks.

In Section II, we derive the physical and statistical measurement models of the diffused substance distribution. In Section III, we present the proposed distributed sequential Bayesian estimation method, including the distributed estimation algorithms, belief representation methods, and the sensor selection strategies. In Section IV, numerical examples are used to demonstrate the performance of the proposed signal processing methods. Conclusions and directions for future work are given in Section V.

II. Physical and Statistical Measurement Models

In this section, we first derive computational physical models describing the space-time substance dispersion mechanisms under various environmental scenarios; then we transform the obtained dispersion model to a statistical measurement model. The following distributed parameter estimation methods are developed based on this statistical model.

A. Physical Models for Dispersion

We model the transport of a substance from a diffusive source by solving the diffusion equation, which describes the dispersion of particles from a region of high concentration to regions of lower concentration due to random motion. Various initial and boundary conditions and environmental effects are considered according to different scenarios. We first compute the concentration for a stationary impulse point source to find the Green function. In this step, we consider different types of boundary conditions (permeable and impermeable boundaries) and environmental effects (external forces and turbulence). Then we extend the results to a continuous source by integrating the source release rate with the Green function.

1) Stationary Impulse Source: Let $c(\boldsymbol{r},t)$ denote the concentration of the diffused substance at a position $\boldsymbol{r}=(x,y,z)$ and time t. For a source-free volume and space-invariant diffusivity κ, when we ignore the effects of external forces, the concentration of a dispersed substance follows the diffusion equation [18]

$$\begin{aligned}\frac{\partial c(\boldsymbol{r},t)}{\partial t} &= \kappa\nabla^2 c(\boldsymbol{r},t)\\ &= \kappa\left[\frac{\partial^2 c(\boldsymbol{r},t)}{\partial x^2}+\frac{\partial^2 c(\boldsymbol{r},t)}{\partial y^2}+\frac{\partial^2 c(\boldsymbol{r},t)}{\partial z^2}\right]. \quad (1)\end{aligned}$$

The appropriate boundary conditions and initial conditions are applied to solve this differential equation. For the simplification of presentation, we consider only the diffusion equation in the z dimension, i.e.,

$$\frac{\partial c_z(z,t)}{\partial t}=\kappa\frac{\partial^2 c_z(z,t)}{\partial z^2}. \quad (2)$$

This equation can be easily extended to the three-dimension cases.

Initial conditions: In order to calculate the Green function, we assume there is an impulse point source of a unit mass release rate stationary at $z=z_0$, i.e., the initial condition is

$$c_z(z,t)=\delta(z-z_0) \quad \text{at} \quad t=t_{\mathrm{I}} \quad (3)$$

where t_{I} is the initial release time.

Boundary conditions: For the boundary conditions, we use a homogeneous semiinfinite medium as an example. The derivation process is also applicable to other environments straightforwardly. A homogeneous semiinfinite medium in the region $z>0$ can be used to model the diffusion in air above the ground. For a generally permeable boundary condition, a reasonable assumption is that the rate of loss of the diffusing substance is proportional to the actual concentration in the surface at any time [18], i.e., the boundary condition at the surface is

$$-\kappa\frac{\partial c_z}{\partial z}=-\alpha c_z \quad \text{at} \quad z=0 \quad (4)$$

where α is a constant of proportionality.

Solutions: Solving the diffusion equation (2) under the above boundary and initial conditions, we obtain

$$c_z(z,t) = \frac{1}{2\sqrt{\pi\kappa(t-t_{\mathrm{I}})}} \times \left\{\exp\left[-\frac{(z-z_0)^2}{4\kappa(t-t_{\mathrm{I}})}\right] + \exp\left[-\frac{(z+z_0)^2}{4\kappa(t-t_{\mathrm{I}})}\right]\right\} - \frac{\alpha}{\kappa}\exp\left[\frac{\alpha(z+z_0)+\alpha^2(t-t_{\mathrm{I}})}{\kappa}\right] \times \operatorname{erfc}\left\{\frac{z+z_0}{2\sqrt{\kappa(t-t_{\mathrm{I}})}} + \frac{\alpha}{\kappa}\sqrt{\kappa(t-t_{\mathrm{I}})}\right\} \quad (5)$$

where "erfc" is the error-function complement. The first term on the right-hand side (RHS) of (5) is the solution for an impermeable boundary, which can also be interpreted as a superposition of contributions from the actual source and a mirror-image source of equal magnitude at $z = -z_0$; the second term represents the loss effect of the diffusing substance.

External forces: In the presence of external forces such as wind, flow, and gravity, the diffusion equation (2) takes the form

$$\frac{\partial c_z}{\partial t} = \kappa\frac{\partial^2 c_z}{\partial z^2} - v_z\frac{\partial c_z}{\partial z} \quad (6)$$

where v_z can represent the wind speed in the z direction [19]. The second term on the RHS is called *advection*. We can reduce the above problem to the ordinary diffusion equation without the advection term by applying the following transformation [19]

$$c_z^* = c_z \exp\left\{\frac{-v_z}{2\kappa}(z-z_0) + \frac{v_z^2(t-t_{\mathrm{I}})}{4\kappa}\right\} \quad (7)$$

to the differential equation (6). Hence, we can apply the same procedures as aforementioned to solve this type of problem (please refer to [7] for more discussion about other environmental effects).

2) Continuous Sources: A solution for a continuous source, from which the diffusion substance is liberated continuously at a certain rate, is deduced from the corresponding Green function by integrating it with the source release rate function.

Point source: Suppose that we have a stationary point source continuously releasing a substance at a mass rate $\mu(t)$. Let $c_{\mathrm{Green}}(\boldsymbol{r},t)$ denote the Green's function of the impulse source case. Then the concentration of a continuous point source is obtained using the following integral:

$$c(\boldsymbol{r},t) = \int_{t_{\mathrm{I}}}^{t} \mu(\tau)c_{\mathrm{Green}}(\boldsymbol{r},t-\tau)\mathrm{d}\tau. \quad (8)$$

Using the three-dimensional (3-D) solution in a homogeneous semi-infinite medium under the impermeable boundary condition as an example and supposing $\mu(t) = \mu_0$, we obtain the concentration distribution

$$c(\boldsymbol{r},t) = \mu_0 \int_{t_{\mathrm{I}}}^{t} \frac{1}{8\left[\pi\kappa(t-\tau)\right]^{3/2}} \times \left\{\exp\left[-\frac{|\boldsymbol{r}-\boldsymbol{r}_0|^2}{4\kappa(t-\tau)}\right] + \exp\left[-\frac{|\boldsymbol{r}-\boldsymbol{r}_1|^2}{4\kappa(t-\tau)}\right]\right\}\mathrm{d}\tau \quad (9)$$

where $\boldsymbol{r}_0 = [x_0, y_0, z_0]^{\mathrm{T}}$ is the source location; and $\boldsymbol{r}_1 = [x_0, y_0, -z_0]^{\mathrm{T}}$.

Distributed source: These results can be extended also to a distributed source. Let $\mu(\boldsymbol{r},t)$ denote the distributed mass release rate, and let S_0 denote the area of the distributed source; then

$$c(\boldsymbol{r},t) = \int_{S_0}\int_{t_{\mathrm{I}}}^{t} \mu(\tau,\boldsymbol{\varsigma})c_{\mathrm{Green}}(\boldsymbol{r}-\boldsymbol{\varsigma},t-\tau)\mathrm{d}\tau\mathrm{d}\boldsymbol{\varsigma}. \quad (10)$$

B. Measurement Models

To model the measurements, we suppose a spatially distributed wireless sensor network has been deployed. Each sensor node in this network is located at a known position and can measure the substance concentration from a diffusion source. Assuming the physical models derived above represent the underlying dispersion mechanism, we obtain a measurement model for a sensor at a position $\{\boldsymbol{r}_i\}$ and taking measurements at time $\{t_j\}$ as

$$y(\boldsymbol{r}_i,t_j) = c(\boldsymbol{r}_i,t_j) + b + e(\boldsymbol{r}_i,t_j), \quad e(\boldsymbol{r}_i,t_j) \sim \mathcal{N}(0,\sigma^2) \quad (11)$$

where $c(\boldsymbol{r}_i,t_j)$ is the concentration of interests; b is a bias term, representing the sensor's response to foreign substances and assumed to be a unknown constant; and $e(\boldsymbol{r}_i,t_j)$ is the sensor's noise, assumed to be Gaussian distributed, independent in time and space (see [3] for a further discussion of this model). Assuming for simplicity that the source's substance-releasing rate is time invariant, and denoting $y_{ij} = y(\boldsymbol{r}_i,t_j)$, $e_{ij} = e(\boldsymbol{r}_i,t_j)$, and $\mu a_{ij}(\boldsymbol{\theta}) = c(\boldsymbol{r}_i,t_j)$, we rewrite (11) as

$$y_{ij} = A_{ij}(\boldsymbol{\theta})\boldsymbol{x} + e_{ij} \quad (12)$$

where $A_{ij}(\boldsymbol{\theta}) = [a_{ij}(\boldsymbol{\theta}), 1]$ and $\boldsymbol{x} = [\mu, b]^{\mathrm{T}}$. Here $\boldsymbol{\theta}$ represents the source and medium parameters. Assuming a stationary source in a semiinfinite medium with impermeable boundary whose concentration distribution is as in (9), we have $\boldsymbol{\theta} = [x_0, y_0, z_0, t_{\mathrm{I}}, \kappa]^{\mathrm{T}}$.

1) Multiple Sources: If there exist multiple sources, the concentration of interests can be written as

$$c(\boldsymbol{r}_i,t_j) = \sum_{m=1}^{M} c_m(\boldsymbol{r}_i,t_j) = \sum_{m=1}^{M} \mu_m a_{m,ij}(\boldsymbol{\theta}_m)$$

where M is the number of sources, and $\boldsymbol{\theta}_m$ includes the parameters related to the m-th source. In this case, the form of the measurement model for y_{ij} is the same as (12), however, here $A_{ij}(\boldsymbol{\theta}) = [a_{1,ij}(\boldsymbol{\theta}_1), \ldots, a_{M,ij}(\boldsymbol{\theta}_M), 1]$ and $\boldsymbol{x} = [\mu_1, \ldots, \mu_M, b]^{\mathrm{T}}$.

2) Nuisance Parameters: In the above parametric statistical measurement model, the unknown parameters include $\boldsymbol{\theta}$, $\boldsymbol{x}$, and σ^2. However, in practice only parts of them are of interest; others are nuisance parameters. For example, under the scenario of a stationary source in a semiinfinite medium with impermeable boundary, if we want to estimate the substance dispersion, $(x_0, y_0, z_0, \mu, t_{\mathrm{I}}, \kappa)$ are the parameters of interest; whereas if we just want to localize the diffusion source, only (x_0, y_0) are parameters of interest. For these parameters, some of them can be

measured or estimated through other methods, thus are removed from the unknown parameter list. For example, the diffusivity κ, bias term b, and noise variance σ^2 can be measured during the calibration, which is processed before the estimation phase.

For the remaining nuisance parameters, e.g., source release rate μ, we can remove them by integration. We first assume an *a priori* probability density function (pdf) for μ, which is a uniform distribution between μ_{low} and μ_{high}, where μ_{low} and μ_{high} are the possible lowest and highest values of μ. For a localization problem, we denote the unknown parameter vector as $\boldsymbol{\theta} = [x_0, y_0]^{\mathrm{T}}$. Consequently, according to the measurement model in (12), the marginal pdf of y_{ij} given $\boldsymbol{\theta}$ is

$$\begin{aligned} p(y_{ij}|\boldsymbol{\theta}) &= \int p(y_{ij}|\boldsymbol{\theta},\mu)p(\mu)\mathrm{d}\mu \\ &= \frac{1}{a_{ij}(\boldsymbol{\theta})\Delta}\left[\Phi\left(\frac{y_{ij}-\mu_{\text{high}}a_{ij}(\boldsymbol{\theta})-b}{\sigma}\right)\right. \\ &\qquad \left.-\Phi\left(\frac{y_{ij}-\mu_{\text{low}}a_{ij}(\boldsymbol{\theta})-b}{\sigma}\right)\right] \end{aligned} \tag{13}$$

where $\Delta = \mu_{\text{high}} - \mu_{\text{low}}$; $\Phi(\cdot)$ is the standard normal cumulative distribution function

$$\Phi(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} \exp\left[-\frac{z^2}{2}\right] \mathrm{d}z. \tag{14}$$

III. Distributed Sequential Bayesian Estimation

In this section, we first develop the distributed sequential Bayesian estimation algorithm and apply it to the estimation of a diffusion source using wireless sensor networks. In our algorithm, how to approximate the state belief appropriately is critical for efficient estimation and decreasing the communication requirement. We then propose two parametric belief representation methods: Gaussian approximation and LPG function approximation. Finally, we discuss the sensor node selection strategies.

A. Distributed Sequential Bayesian Estimation Algorithm

1) Estimation Scheme: The proposed scheme of estimating a diffusion source using wireless sensor networks has two phases: the measuring phase and the estimation phase. In the **measuring phase**, when the alarm sensor detects a possible release from a diffusion source, it then activates the sensor nodes around the potential target. The activated sensor nodes take the measurements of the substance concentration at specific time samples t_j, $j = 1, \ldots, N$, and then return to sleeping status. The determination of the number of samples depends on how quickly the task requires the sensor network to respond. Here we assume the sensor network is synchronized such that the sampling times t_j are known at each sensor node. In the **estimation phase**, we process a distributed sequential (or incremental) Bayesian algorithm, activated by an initial sensor node, to estimate the diffusion source, e.g., to localize the source position. We will find that the estimation method itself provides a general framework for the stationary parameter related applications in wireless sensor networks. It is independent on the sensing models that are used.

List of Notations

- $i \in \{1, \ldots, K\}$, the sensor node index; K is the total number of nodes that are awakened.
- $j \in \{1, \ldots, N\}$, the time index; N is the total number of measurements in each sensor node.
- $\boldsymbol{y}_i \triangleq [y_{i1}, y_{i2}, \ldots, y_{iN}]^{\mathrm{T}}$, the collection of all measurements at sensor node i.
- $\boldsymbol{y}_{1:i} \triangleq \{\boldsymbol{y}_1, \boldsymbol{y}_2, \ldots, \boldsymbol{y}_i\}$, the measurement sequence up to sensor node i.

2) Distributed Sequential Bayesian Estimation: In wireless sensor networks, because of their energy limitation and failure tolerance properties, the usually centralized Bayesian estimation methods cannot be directly applied. In the following, we propose a distributed sequential Bayesian estimation method. "Sequential" here means that the messages are transmitted in the sensor network, and the state belief and Bayesian estimate are updated spatially according to the previous information and the measurements at the current sensor node. To derive the sequential Bayesian estimation [here we use the minimum-mean-squared error (mmse) estimation], we extend the basic Bayesian estimation such that it can incrementally combine measurements over space.

Suppose the current sensor node i obtains the belief passed from the sensor node $i-1$ as $p(\boldsymbol{\theta}|\boldsymbol{y}_{1:i-1})$, and the measurements at the current sensor node i are the $\boldsymbol{y}_i$. We assume that the following conditional independence assumption is satisfied:

A1: Conditioned on $\boldsymbol{\theta}$, the measurements at the current sensor node $\boldsymbol{y}_i$ are independent of the measurements at the previous sensor node $\boldsymbol{y}_{1:i-1}$, i.e.,

$$p(\boldsymbol{y}_i|\boldsymbol{\theta}, \boldsymbol{y}_{1:i-1}) = p(\boldsymbol{y}_i|\boldsymbol{\theta}). \tag{15}$$

Note that in our diffusive source estimation problem, we assume the noise e_{ij} in the measurement model (12) is independent in space, which guarantees the assumption *A1* is satisfied. Then, we can update the belief to $p(\boldsymbol{\theta}|\boldsymbol{y}_{1:i})$ using $\boldsymbol{y}_i$ by applying the Bayesian rule as

$$p(\boldsymbol{\theta}|\boldsymbol{y}_{1:i}) = \frac{p(\boldsymbol{y}_i|\boldsymbol{\theta})p(\boldsymbol{\theta}|\boldsymbol{y}_{1:i-1})}{\int p(\boldsymbol{y}_i|\boldsymbol{\theta})p(\boldsymbol{\theta}|\boldsymbol{y}_{1:i-1})\mathrm{d}\boldsymbol{\theta}}. \tag{16}$$

At the first sensor node where $i = 1$, we have

$$p(\boldsymbol{\theta}|\boldsymbol{y}_1) = \frac{p(\boldsymbol{y}_1|\boldsymbol{\theta})\pi(\boldsymbol{\theta})}{\int p(\boldsymbol{y}_1|\boldsymbol{\theta})\pi(\boldsymbol{\theta})\mathrm{d}\boldsymbol{\theta}} \tag{17}$$

where $p(\boldsymbol{y}_i|\boldsymbol{\theta})$ is the likelihood function of the current observation at the sensor node i

$$p(\boldsymbol{y}_i|\boldsymbol{\theta}) = \prod_{j=1}^{N} p(y_{ij}|\boldsymbol{\theta}) \tag{18}$$

and $\pi(\boldsymbol{\theta})$ is the prior pdf for the location $\boldsymbol{\theta}$ which is obtained from the prior information. In practice, if we do not have prior knowledge about $\boldsymbol{\theta}$ we assume the prior pdf is Gaussian with a large variance. Therefore, the current mmse estimate at the sensor node i can be calculated as

$$\widehat{\boldsymbol{\theta}} = \mathrm{E}[\boldsymbol{\theta}|\boldsymbol{y}_{1:i}] = \frac{\int \boldsymbol{\theta} p(\boldsymbol{y}_i|\boldsymbol{\theta})p(\boldsymbol{\theta}|\boldsymbol{y}_{1:i-1})\mathrm{d}\boldsymbol{\theta}}{\int p(\boldsymbol{y}_i|\boldsymbol{\theta})p(\boldsymbol{\theta}|\boldsymbol{y}_{1:i-1})\mathrm{d}\boldsymbol{\theta}} \tag{19}$$

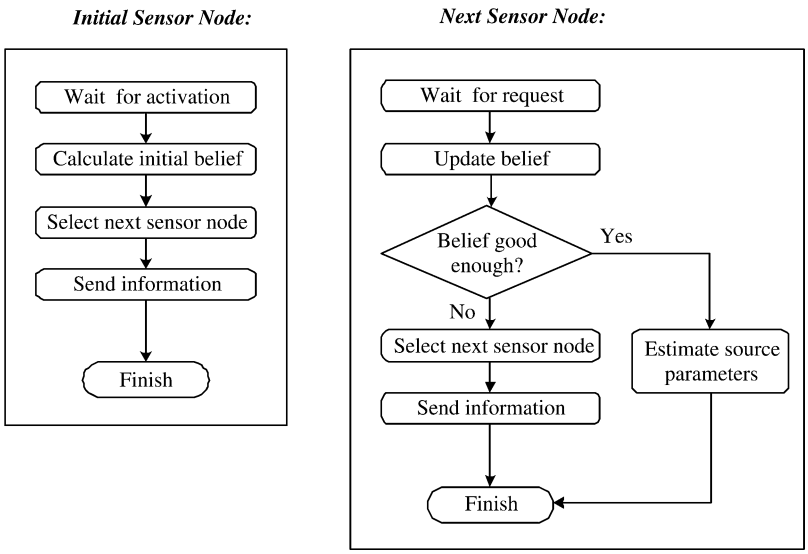


Fig. 1. Flowchart of the distributed sequential Bayesian estimation algorithm.

and the covariance matrix of the current estimate is

$$\mathrm{Cov}[\boldsymbol{\theta}|\boldsymbol{y}_{1:i}] = \mathrm{E}\left[(\boldsymbol{\theta}-\widehat{\boldsymbol{\theta}})(\boldsymbol{\theta}-\widehat{\boldsymbol{\theta}})^{\mathrm{T}}|\boldsymbol{y}_{1:i}\right] = \frac{\int(\boldsymbol{\theta}-\widehat{\boldsymbol{\theta}})(\boldsymbol{\theta}-\widehat{\boldsymbol{\theta}})^{\mathrm{T}}p(\boldsymbol{y}_i|\boldsymbol{\theta})p(\boldsymbol{\theta}|\boldsymbol{y}_{1:i-1})\mathrm{d}\boldsymbol{\theta}}{\int p(\boldsymbol{y}_i|\boldsymbol{\theta})p(\boldsymbol{\theta}|\boldsymbol{y}_{1:i-1})\mathrm{d}\boldsymbol{\theta}} \quad (20)$$

which is used as the performance measure.

3) Algorithm: In (16) we observe that the current belief is a product of the previous belief and the current likelihood function (up to a normalization coefficient), which provides us with a sequential algorithm to update the belief. Hence, we obtain the distributed sequential Bayesian estimation algorithm shown in Fig. 1 and summarized as follows:

The initial sensor node $i = 1$ processes the following steps:

- *Belief initialization:* Calculate the initial belief according to the prior parameter pdf and its own measurements, using (17).
- *Sensor node selection:* Select a sensor node from its neighbor according to certain utility information criteria (the details will be discussed in Section III-C).
- *Data transmission:* Wake up the selected sensor node, transmit the current belief to it, and return to sleeping status.

The next sensor nodes $i > 1$ perform the following steps.

- *Data receiving*: Be activated by the previous sensor node and receive the transmitted data.
- *Belief update:* Calculate a new belief according to the received previous belief and its own measurements, using (16).
- *Belief quality test:* Test the quality of the updated belief according to some measure such as the trace or determinant of the covariance matrix (20). If the belief is "good enough," it will calculate the mmse estimate of the source parameters according to (19), and the estimation phase is terminated.
- *Sensor node selection:* Select a sensor node from its neighbor according to certain utility information criteria (the details will be discussed in Section III-C).
- *Data transmission:* Wake up the selected sensor node, transmit the current belief to it, and return to sleeping status.

4) Discussion: In this algorithm the messages are transmitted only between a sensor node and its neighbor; hence, we realize a fully distributed estimation. However, power and resource constraints in wireless sensor networks create some difficulties of implementation:

- In the algorithm, we need to transmit the current belief to the next sensor node. Because of the nonlinear and non-Gaussian characteristics of the measurement model, we cannot obtain an analytical form of the belief density. And directly transmitting the grid-samples of the belief would require significant power consumption. Therefore, we need to represent the belief in a proper way.
- In the steps of belief update, mmse calculation, and belief quality evaluation, high dimensional multiple integration calculation is required. Due to the nonlinearity and non-Gaussian nature of the measurement model, we can apply only numerical algorithms, which may lead to high computation cost and exceed the limited processing capability of the sensor nodes.

B. Belief Approximation

To overcome the difficulties in the derived distributed estimation algorithm, and consider the fact that the power constraint is critical in wireless sensor networks, we propose to approximate the state belief $p(\boldsymbol{\theta}|\boldsymbol{y}_{1:i})$ by a family of parameterizable distributions. Therefore, in the belief update we will need to transmit only a finite number of parameters, which decreases the communication burden significantly, with consequent decrease in power consumption and increase in sensor network longevity. We propose two parametric distribution approximation methods: a Gaussian approximation and a new approach—a linear combination of Gaussian multiplied by polynomials (LPG) approximation. In both of these methods, we consider the accuracy of the approximation as

well as decreasing the computation complexity to fit the limited processing capability of the sensor nodes.

1) Gaussian Approximation:

a) Approximation algorithm: For the continuous posterior distribution function, there exists an asymptotic posterior normality property [20], which is presented as follows.

Asymptotic posterior normality. Consider $p_n(\boldsymbol{\theta})$ as the posterior probability density function of a random variable $\boldsymbol{\theta}$, $\boldsymbol{\theta} \in \Re^k$, given the observation $y_{1:n}$, i.e., $p_n(\boldsymbol{\theta}) = p(\boldsymbol{\theta}|y_{1:n})$; define $\boldsymbol{\phi}_n = \Sigma_n^{-1/2}(\boldsymbol{\theta} - \boldsymbol{m}_n)$, where $p_n(\boldsymbol{\theta})$ reaches its local maximum at $\boldsymbol{m}_n$, i.e.,

$$\nabla \log p_n(\boldsymbol{\theta})|_{\boldsymbol{\theta}=\boldsymbol{m}_n} = 0$$

and $\Sigma_n = [-\nabla^2 \log p_n(\boldsymbol{\theta})|_{\boldsymbol{\theta}=\boldsymbol{m}_n}]^{-1}$. Let (c1), (c2), and (c3) represent the conditions "steepness," "smoothness," and "concentration," respectively (see [20, p. 289] for the definitions of these conditions). Then, given (c1) and (c2), (c3) is a necessary and sufficient condition for $\boldsymbol{\phi}_n$ to converge in distribution to a standard multivariate normal random variable $\boldsymbol{\phi}$, whose probability density function is $p(\boldsymbol{\phi}) = (2\pi)^{-k/2} \exp\{-(1/2)\boldsymbol{\phi}^{\mathrm{T}}\boldsymbol{\phi}\}$.

This proposition tells us that for large n, the posterior function p_n becomes highly peaked and behaves like the multivariate normal density kernel $\exp\{-(1/2)(\boldsymbol{\theta} - \boldsymbol{m}_n)^{\mathrm{T}}\Sigma_n^{-1}(\boldsymbol{\theta} - \boldsymbol{m}_n)\}$. Therefore, in our problem, for the belief $p(\boldsymbol{\theta}|\boldsymbol{y}_{1:i})$ that (c1) to (c3) are satisfied, we propose to approximate it by a multivariate Gaussian distribution with mean $\mathrm{E}[\boldsymbol{\theta}|\boldsymbol{y}_{1:i}]$ and covariance $\mathrm{Cov}[\boldsymbol{\theta}|\boldsymbol{y}_{1:i}]$, i.e.,

$$p(\boldsymbol{\theta}|\boldsymbol{y}_{1:i}) \approx \mathcal{N}\left(\mathrm{E}[\boldsymbol{\theta}|\boldsymbol{y}_{1:i}], \mathrm{Cov}[\boldsymbol{\theta}|\boldsymbol{y}_{1:i}]\right) \tag{21}$$

where

$$\begin{aligned}
&\mathrm{E}[\boldsymbol{\theta}|\boldsymbol{y}_{1:i}] \\
&\quad = \int \boldsymbol{\theta} p(\boldsymbol{\theta}|\boldsymbol{y}_{1:i})\mathrm{d}\boldsymbol{\theta} \\
&\quad = \frac{\int \boldsymbol{\theta}\left[\prod_{j=1}^{N} p(y_{ij}|\boldsymbol{\theta})\right] p(\boldsymbol{\theta}|\boldsymbol{y}_{1:i-1})\mathrm{d}\boldsymbol{\theta}}{\int \left[\prod_{j=1}^{N} p(y_{ij}|\boldsymbol{\theta})\right] p(\boldsymbol{\theta}|\boldsymbol{y}_{1:i-1})\mathrm{d}\boldsymbol{\theta}} \qquad (22a) \\
&\mathrm{Cov}[\boldsymbol{\theta}|\boldsymbol{y}_{1:i}] \\
&\quad = \int (\boldsymbol{\theta} - \mathrm{E}[\boldsymbol{\theta}|\boldsymbol{y}_{1:i}])(\boldsymbol{\theta} - \mathrm{E}[\boldsymbol{\theta}|\boldsymbol{y}_{1:i}])^{\mathrm{T}} p(\boldsymbol{\theta}|\boldsymbol{y}_{1:i})\mathrm{d}\boldsymbol{\theta}. \qquad (22b)
\end{aligned}$$

To calculate the mean and covariance using above equations requires multiple integrations which introduces significant computation cost. In order to decrease the processing capability demand, we apply the Laplace approximation method to calculate the integration in (22a) and (22b).

b) Laplace's method: Laplace's method, a family of asymptotic methods used to approximate integrals, is presented as a potential candidate for the toolbox of techniques used for knowledge acquisition and probabilistic inference in belief networks with continuous variables [20], [21]. This technique approximates posterior moments and marginal posterior distributions with reasonable accuracy in many cases. Let $g(\boldsymbol{\theta})$ be a scalar function of parameters of interest $\boldsymbol{\theta}$. Consider the expectation of the posterior $\mathrm{E}[g(\boldsymbol{\theta})|\boldsymbol{y}]$ of the form

$$\mathrm{E}\left[g(\boldsymbol{\theta})|\boldsymbol{y}\right] = \frac{\int g(\boldsymbol{\theta}) p(\boldsymbol{y}|\boldsymbol{\theta}) p(\boldsymbol{\theta})\mathrm{d}\boldsymbol{\theta}}{\int p(\boldsymbol{y}|\boldsymbol{\theta}) p(\boldsymbol{\theta})\mathrm{d}\boldsymbol{\theta}} \tag{23}$$

where $\boldsymbol{y} = [y_1, y_2, \ldots, y_{\mathrm{N}}]^{\mathrm{T}}$ is the observation vector. By applying the Laplace method, we obtain the approximation for $\mathrm{E}[g(\boldsymbol{\theta})|\boldsymbol{y}]$ as

$$\widehat{\mathrm{E}}\left[g(\boldsymbol{\theta})|\boldsymbol{y}\right] = \left(\frac{\det\Sigma_{\boldsymbol{\theta}_a^*}}{\det\Sigma_{\boldsymbol{\theta}_b^*}}\right)^{\frac{1}{2}} \exp\left\{-N\left[h_a\left(\boldsymbol{\theta}_a^*\right) - h_b\left(\boldsymbol{\theta}_b^*\right)\right]\right\} \tag{24}$$

where

- the functions $h_a(\boldsymbol{\theta})$ and $h_b(\boldsymbol{\theta})$ are defined by

$$h_a(\boldsymbol{\theta}) = -N^{-1}[\log g(\boldsymbol{\theta}) + \log p(\boldsymbol{\theta}) + \log p(\boldsymbol{y}|\boldsymbol{\theta})] \tag{25a}$$
$$h_b(\boldsymbol{\theta}) = -N^{-1}[\log p(\boldsymbol{\theta}) + \log p(\boldsymbol{y}|\boldsymbol{\theta})]. \tag{25b}$$

- $\boldsymbol{\theta}_a^*$ and $\boldsymbol{\theta}_b^*$ are, respectively, the minimizers for $h_a(\boldsymbol{\theta})$ and $h_b(\boldsymbol{\theta})$, i.e.

$$\boldsymbol{\theta}_a^* = \arg\max\{-h_a(\boldsymbol{\theta})\} \tag{26a}$$
$$\boldsymbol{\theta}_b^* = \arg\max\{-h_b(\boldsymbol{\theta})\}. \tag{26b}$$

- $\Sigma_{\boldsymbol{\theta}_a^*}$ and $\Sigma_{\boldsymbol{\theta}_b^*}$ are the inverse of the Hessian of $h_a(\boldsymbol{\theta})$ and $h_b(\boldsymbol{\theta})$ evaluated at $\boldsymbol{\theta}_a^*$ and $\boldsymbol{\theta}_b^*$, respectively

$$\left[\Sigma_{\boldsymbol{\theta}_a^*}\right]^{-1} = \nabla^2 h_a(\boldsymbol{\theta}_a^*), \quad \left[\Sigma_{\boldsymbol{\theta}_b^*}\right]^{-1} = \nabla^2 h_b\left(\boldsymbol{\theta}_b^*\right). \tag{27}$$

In this case, the accuracy of the approximation can be shown as

$$\mathrm{E}\left[g(\boldsymbol{\theta})|\boldsymbol{y}\right] = \widehat{\mathrm{E}}\left[g(\boldsymbol{\theta})|\boldsymbol{y}\right]\left(1 + O(N^{-2})\right). \tag{28}$$

c) Application to diffusive source localization: To localize a diffusive source, the parameter vector of interest is $\boldsymbol{\theta} = [x_0, y_0]^{\mathrm{T}}$. Thus, by using the Gaussian approximation, in order to update the belief, we need only to transmit total 5 scalars (2 for mean and 3 for covariance matrix) to the next sensor node, which decreases the communications required between sensor nodes dramatically.

In order to apply the above Laplace approximation to our localization problem, we rewrite the mean and covariance of the posterior in (22a) and (22b) as

$$\mathrm{E}[\boldsymbol{\theta}|\boldsymbol{y}_{1:i}] = \begin{bmatrix} \mathrm{E}[x_0|\boldsymbol{y}_{1:i}] \\ \mathrm{E}[y_0|\boldsymbol{y}_{1:i}] \end{bmatrix} \tag{29a}$$

$$\mathrm{Cov}[\boldsymbol{\theta}|\boldsymbol{y}_{1:i}] = \begin{bmatrix} \mathrm{Var}[x_0|\boldsymbol{y}_{1:i}] & \mathrm{Cov}[x_0, y_0|\boldsymbol{y}_{1:i}] \\ \mathrm{Cov}[y_0, x_0|\boldsymbol{y}_{1:i}] & \mathrm{Var}[y_0|\boldsymbol{y}_{1:i}] \end{bmatrix}. \tag{29b}$$

Therefore, when we define

$$g_1(\boldsymbol{\theta}) = x_0, \quad g_2(\boldsymbol{\theta}) = y_0 \tag{30}$$

and replace $p(\boldsymbol{\theta})$ by $p(\boldsymbol{\theta}|\boldsymbol{y}_{1:i-1})$ and $p(\boldsymbol{y}|\boldsymbol{\theta})$ by $p(\boldsymbol{y}_i|\boldsymbol{\theta})$, we can directly use the above Laplace's method to approximate the posterior mean and variance in (22a) and (22b). Here, we approximate $\mathrm{Cov}[g_1(\boldsymbol{\theta}), g_2(\boldsymbol{\theta})|\boldsymbol{y}]$ using the expression (omitting $\boldsymbol{\theta}$):

$$\widehat{\mathrm{Cov}}[g_1, g_2|\boldsymbol{y}] = \widehat{\mathrm{E}}[g_1 g_2|\boldsymbol{y}] - \widehat{\mathrm{E}}[g_1|\boldsymbol{y}]\widehat{\mathrm{E}}[g_2|\boldsymbol{y}]. \tag{31}$$

Note that the functions in this algorithm are related only to the local measurements $\boldsymbol{y}_i$; hence, the algorithm can be completely implemented locally. No distributed processing is needed.

d) Discussion: This Gaussian approximation method has the advantage that its belief representation is very simple that leads to the substantially low amount of wireless transmission between sensor nodes and low computation complexity at each sensor node. However, the approximation accuracy is not very good, especially at the beginning of the process when the number of collected measurements is small. This low accuracy may cause the process of data transmission between sensor nodes to be relatively long and lead to a high total communication power consumption. Therefore, the Gaussian approximation is suitable only for the cases when the belief density has Gaussian-like distribution or the processing capacity at the sensor node is very limited.

2) LPG Function Approximation: We propose to use a parameterizable distribution family of LPG functions to approximate the belief. The PG functions represent Gaussian density functions multiplied by polynomials, and LPG denotes a linear combination of PG functions.

a) LPG functions: PG or LPG functions were first presented by Paldi and Nehorai in [22] to enable exact analytical solutions for nonlinear filtering problems by approximating the posterior pdf using LPG functions. In our problem, using LPG functions to approximate the belief in wireless sensor networks has these advantages.

- This approximation family is a natural generalization of the above Gaussian approximation in Section III-B-1; thus, it can provide higher approximation accuracy: this family of LPG functions has the property that it can approximate any pdf with an arbitrarily small error, provided that the polynomial degree is sufficiently large.
- Any moments for this family can be evaluated analytically: for PG functions the calculation of moments is equivalent to the calculation of the moments of the corresponding Gaussian functions, for which the analytical results can be derived.
- The PG and LPG families are closed under multiplication and integration operations. In some cases, this point enables the analytical computation of the posterior density functions.

We also note that another typical nonlinear Bayesian estimation method using Gaussian sum [23], [24] is a special case of the implementation of LPG function approximation.

b) Approximation procedure: The procedure of using an LPG function to approximate the belief is performed by calculating the optimal parameters for the LPG function (including the mean and covariance matrix of the Gaussian density, the polynomial coefficients, and the linear combination coefficients) to minimize the error between the real and the approximated belief function. However, if we directly calculate these optimal parameters using some numerical methods, e.g., a least-squares (LS) method, the resulting computation cost is very high due to the large number of unknown parameters. Therefore, we propose an efficient approximation algorithm in which we separate the whole approximation procedure into two phases: local approximation and global approximation.

TABLE I
THE MULTIVARIATE HERMITE POLYNOMIALS

Order	Polynomial
0	$\mathrm{He}_{00} = 1$
1	$\mathrm{He}_{10} = z_1$
2	$\mathrm{He}_{20} = z_1^2 - w_{11}$
	$\mathrm{He}_{11} = z_1 z_2 - w_{12}$
3	$\mathrm{He}_{30} = z_1^3 - 3w_{11} z_1$
	$\mathrm{He}_{21} = z_1^2 z_2 - 2w_{12} z_1 - w_{11} z_2$
4	$\mathrm{He}_{40} = z_1^4 - 6w_{11} z_1^2 + 3w_{11}$
	$\mathrm{He}_{31} = z_1^3 z_2 - 3w_{12} z_1^2 - 3w_{11} z_1 z_2 + 3w_{11} w_{12}$
	$\mathrm{He}_{22} = z_1^2 z_2^2 - w_{22} z_1^2 - w_{11} z_1 z_2 - 4w_{12} z_1 z_2 + w_{11} w_{22} + 2w_{12}^2$

That is, we first approximate the belief at several important peaks (peaks whose contribution to the density cannot be ignored) using PG functions, then we create an optimal linear combination of these local expansions to represent the global belief function.

Local approximations: In local approximations, first we localize the important peaks (local maximums) of the belief density. We can apply numerical methods, such as the efficient quasi-Newton methods, to find them. Another advantage of using the quasi-Newton methods is that they also give an estimate of the Hessian at the peak. We then use an orthogonal family of PG functions to approximate the density around each peak. For such a family we use the Hermite functions [25], which (up to normalization factors) are given by a Gaussian density multiplied by Hermite polynomials.

Denote $f_l(\boldsymbol{\theta}|\boldsymbol{y}_{1:i})$ as the local approximation of the belief around l-th peak, and apply the orthogonal expansions in terms of multivariate Hermite polynomials to the local approximations, then we have

$$f_l(\boldsymbol{\theta}|\boldsymbol{y}_{1:i}) = \phi(\boldsymbol{\theta};\boldsymbol{\mu},\Sigma) \sum_{0\le\nu_1+\ldots+\nu_n\le k} C_{\boldsymbol{\nu}} \cdot \mathrm{He}_{\boldsymbol{\nu}}(\boldsymbol{\theta};\boldsymbol{\mu},\Sigma) \quad (32)$$

where $\boldsymbol{\nu} = [\nu_1, \nu_2, \ldots, \nu_n]^{\mathrm{T}}$ is a nonnegative integer vector, n is the dimension of $\boldsymbol{\theta}$; k is the bounded degree of the polynomials. In (32), the summation is over every combination of $\{\nu_1, \ldots, \nu_n\}$ whose summation is less or equal k. In (32), $\phi(\boldsymbol{\theta};\boldsymbol{\mu},\Sigma)$ is the multinormal probability density function with mean vector $\boldsymbol{\mu}$ and covariance matrix Σ; $\mathrm{He}_{\boldsymbol{\nu}}(\boldsymbol{\theta};\boldsymbol{\mu},\Sigma)$ denotes the multivariate Hermite polynomials defined by

$$\mathrm{He}_{\boldsymbol{\nu}}(\boldsymbol{\theta};\boldsymbol{\mu},\Sigma) = \frac{(-1)^{\Sigma\nu_i}}{\phi(\boldsymbol{\theta};\boldsymbol{\mu},\Sigma)} \frac{\partial^{\Sigma\nu_i}}{\partial\boldsymbol{\theta}^{\boldsymbol{\nu}}} \phi(\boldsymbol{\theta};\boldsymbol{\mu},\Sigma) \quad (33)$$

where $\boldsymbol{\theta}^{\boldsymbol{\nu}} = \theta_1^{\nu_1} \cdots \theta_n^{\nu_n}$ [25]. For a localization problem, $n = 2$. We derive the multivariate Hermite polynomials for the bivariate case up to order four which are given in Table I. Here we denote $W = \Sigma^{-1}$ and $\boldsymbol{z} = W(\boldsymbol{\theta} - \boldsymbol{\mu})$.

We observe that the approximation in (32) is represented by the parameters $\boldsymbol{\mu}$, Σ, and $C_{\boldsymbol{\nu}}$, which are to be determined. For the multivariate Gaussian density $\phi(\boldsymbol{\theta};\boldsymbol{\mu},\Sigma)$, we use the position of the peak as its mean, denoted by $\boldsymbol{\mu}_l$, and let its associated

covariance matrix Σ_l be equal to the minus of the inverse of the Hessian (at the peak) of the logarithm of the belief

$$\Sigma_l = (-H_l)^{-1} = \left[-\nabla^2 L_i(\boldsymbol{\mu}_l)\right]^{-1} \tag{34}$$

where $L_i(\boldsymbol{\theta}) = \log p(\boldsymbol{\theta}|\boldsymbol{y}_{1:i})$ and H_l denotes the Hessian matrix at the peak l. This is a locally best Gaussian approximation in the sense that its Hessian at the peak is the same as that of the true density.

After we obtain the Gaussian function, we can calculate the expansion coefficients $C_{\boldsymbol{\nu}}$ using least squares methods. For efficient processing, we first create a mesh of points $\{\boldsymbol{\theta}_{j,k}\}$ densely sampled around the peak. The distribution of these samples is determined by the estimated Hessian at the peak. Then we find a set of coefficients $C_{\boldsymbol{\nu}}$ that minimizes the square of the errors between the precise belief density function $p_l(\boldsymbol{\theta}|\boldsymbol{y}_{1:i})$ and local approximation $f_l(\boldsymbol{\theta}|\boldsymbol{y}_{1:i})$ at these samples, i.e.,

$$C_{\boldsymbol{\nu}} = \arg\min \sum_{j,k} \left|p_l(\boldsymbol{\theta}_{j,k}|\boldsymbol{y}_{1:i}) - f_l(\boldsymbol{\theta}_{j,k}|\boldsymbol{y}_{1:i})\right|^2. \tag{35}$$

Global approximations: After we calculate the local expansions for all the peaks, we merge them to obtain a global approximation of the belief based on an optimal linear combinations of these local expansions. Then the final global approximation is

$$f(\boldsymbol{\theta}|\boldsymbol{y}_{1:i}) = \sum_l \lambda_l f_l(\boldsymbol{\theta}|\boldsymbol{y}_{1:i}) \tag{36}$$

where the combination coefficients λ_l are determined using LS methods on a new global mesh of measurement points. Since the form of each local expansion is known, we can apply the linear LS to calculate the coefficients, with low computation cost.

c) Discussion: Validity of approximation: In the local approximation, since we use only the truncated expansion in terms of Hermite polynomials, it is possible that the resulting series approximation is not a valid density function, because some values of the approximation may be negative. One way to compensate for this disadvantage is to approximate the square root of the density function. Then the final approximation is its square (which is still a PG function, since PG functions are closed for multiplication) and is guaranteed to be nonnegative.

Computation complexity: Considering the limited signal processing capability at the sensor nodes, we should always try to reduce the processing complexity of the above method. An efficient way is to reduce the total number of parameters (which also decreases the required communication amount) by keeping the number of PG functions and the total degrees of their associated multivariate polynomials as small as possible while still maintaining an acceptable global density function approximation accuracy. The possible approaches to solve this problem include: i) when meeting a predetermined quality measure of local approximation, try to reduce by one the total degree bound of its polynomial and repeat the approximation process until we get the minimal total degree bound that meets the approximation quality measure; ii) if two or more local peaks are clustered together, try to merge them into a single peak (at their weighted center of gravity—according to their importance) and check if the minimal total degree needed for this single peak gives fewer parameters than the total number of parameters needed in the former approximation (using a PG function of minimal total degree for each peak in the cluster); iii) try to ignore in our local approximation the less important peaks one by one (starting from the weakest one) as long as we meet the quality measure tolerance.

Moment calculation: Since the obtained approximation is a linear combination of Gaussian density multiplied by polynomials, calculating the mmse estimation (which is the mean of the posterior function) and accuracy measure (which is the covariance matrix of the posterior function) is equivalent to the computation of the moments of corresponding multinormal density functions. The moments of the multivariate normal can be calculated analytically [26] as follows. Suppose $\boldsymbol{x} = [x_1, \ldots, x_n]^{\mathrm{T}}$ is an n-dimensional normally distributed random variable with mean $\boldsymbol{\mu}$ and covariance $\Sigma = (\sigma_{ij})$. We consider such a form of moment $E[x_{\alpha_1} \ldots x_{\alpha_r}]$, where $\alpha_1, \ldots, \alpha_r$ lie in $\{1, \ldots, n\}$. The $\{\alpha_i\}$ need not all be distinct. Then we have

$$E\left[x_{\alpha_1} \ldots x_{\alpha_r}\right] = \sum_{l+2k=r} \sum_m \mu_{a_1} \ldots \mu_{a_l} \sigma_{b_1 b_2} \ldots \sigma_{b_{2k-1} b_{2k}} \tag{37}$$

where the first summation $\sum_{l+2k=r}$ sums over all combinations of indices l and k for which $l + 2k = r$; the second summation $\sum_m$ sums over all $m = r!/(l!2^k k!)$ permutations $(a_1, \ldots, a_l, b_1, b_2, \ldots, b_{2k})$ of $(\alpha_1, \ldots, \alpha_r)$, giving distinct terms allowing for the symmetry of Σ.

Comparison with Gaussian approximation: When we compare LPG approximation with Gaussian approximation, we observe that at a single sensor node, the LPG approximation uses more expensive computation to achieve a higher representation accuracy. In the Gaussian approximation where we apply Laplace's method to calculate the mean and variance, the most computation power is spent on the two minimization problems in (26). Whereas in the LPG approximation, the most computation power is used to localize modes (or important peaks) of the belief. The computation cost of each localization is nearly the same as one minimization problem in the Gaussian approximation. Hence, it depends on how many modes or how complex the belief is to determine how much more computation power is required for the LPG approximation than the Gaussian approximation in a single sensor node. However, according to the asymptotic posterior normality property, the number of the modes in the belief, and, hence, also the computation complexity of the LPG approximation will decrease along the estimation process. We also consider that since the LPG approximation provides more accurate representation, the corresponding distributed estimation requires less number of sensor nodes to achieve an acceptable performance. Therefore, the total required computation power for the LPG approximation is not substantial higher than the Gaussian approximation. Another advantage of the LPG approximation is that since it needs less number of sensor nodes, it saves the communication power and shortens the processing time.

C. Optimal Sensor Node Selection

The above distributed sequential Bayesian estimation algorithm can be summarized as: given the current belief, we wish to update the belief incrementally by incorporating the measurements of other nearby sensor node, and then estimate the parameters of interest based on the obtained belief. However, not all available sensor nodes in the network provide useful information that improves the estimate; furthermore, some information may be redundant. Therefore, it is useful to select an optimal subset and an optimal order of incorporating these measurements into our belief update. Usually this selection process provides a faster reduction in estimation uncertainty compared with blind or nearest-neighbor sensor selection schemes, and incurs a lower communication burden for meeting a given estimation performance requirement.

In our distributed estimation method, we propose the following four different information utility measures, fitting different scenarios, for selecting the next sensor node optimally. They are mutual information measure, posterior Cramér-Rao bound, Mahalanobis distance, and covariance-based measure.

1) Mutual Information: To use the mutual information as an information utility measure, we select a sensor node l which will maximize the mutual information $I(\boldsymbol{\Theta}; \boldsymbol{Y}_l | \boldsymbol{Y}_{1:i} = \boldsymbol{y}_{1:i})$

$$NODE_{i+1} = \arg\max_{l \in \mathcal{S}} I(\boldsymbol{\Theta}; \boldsymbol{Y}_l | \boldsymbol{Y}_{1:i} = \boldsymbol{y}_{1:i}) \tag{38}$$

where $\mathcal{S}$ is the collection of sensors that are in the neighborhood of current sensor node i, and $NODE_i$ represents the index of the ith selected node. This sensor node selection criterion is equivalent to selecting the node whose measurement $\boldsymbol{y}_l$, when conditioned on the current measurement history $\boldsymbol{y}_{1:i}$, would provide the greatest amount of new information about the target location $\boldsymbol{\theta}$.

2) Posterior Cramér–Rao Bounds: For random parameters, as in our distributed sequential Bayesian estimation, a lower bound that is analogous to the Cramér-Rao bound (CRB) in a nonrandom parameter estimation exists and is derived in [27], usually referred to as posterior CRB (PCRB).

Let $\boldsymbol{x}$ be a vector of measurements, $\boldsymbol{\theta}$ is an n-dimensional estimated random parameter, $p(\boldsymbol{x}, \boldsymbol{\theta})$ is the joint probability density of the pair $(\boldsymbol{x}, \boldsymbol{\theta})$, and let $\widehat{\boldsymbol{\theta}} = \psi(\boldsymbol{x})$ be an estimate of $\boldsymbol{\theta}$. Then, the PCRB on the mean-square estimation error satisfies

$$P \triangleq \mathrm{E}\left\{ [\psi(\boldsymbol{x}) - \boldsymbol{\theta}] [\psi(\boldsymbol{x}) - \boldsymbol{\theta}]^{\mathrm{T}} \right\} \geq F^{-1} \tag{39}$$

where F is the $n \times n$ Fisher information matrix with the elements

$$F_{ij} = \mathrm{E}\left[-\frac{\partial^2 \log p(\boldsymbol{x}, \boldsymbol{\theta})}{\partial \theta_i \partial \theta_j} \right], \quad i, j = 1, \ldots, n \tag{40}$$

provided that the derivatives and expectations in (39) and (40) exist. The inequality in (39) means that the difference $P - F^{-1}$ is a nonnegative definite matrix. From this property we know that PCRB is a lower bound on the error covariance matrix; hence, we can use PCRB as a measure of the information utility, which is equivalent to using the Fisher information matrix as a measure.

According to the definition of PCRB in (39), in our estimation problem we will select the next node by maximizing a information utility measure based on the following Fisher information matrix (usually, we will use its trace or determinant as the information utility function)

$$F_{i+1} = \mathrm{E}\left[-\frac{\partial^2 \log p(\boldsymbol{y}_{1:i+1}, \boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^{\mathrm{T}}} \right]. \tag{41}$$

Upon applying the independence assumption (15), we can rewrite the joint pdf $p(\boldsymbol{y}_{1:i+1}, \boldsymbol{\theta})$ as

$$p(\boldsymbol{y}_{1:i+1}, \boldsymbol{\theta}) = p(\boldsymbol{y}_{i+1} | \boldsymbol{\theta}) p(\boldsymbol{y}_{1:i}, \boldsymbol{\theta}). \tag{42}$$

Inserting this equation into the Fisher information matrix (41), we obtain the following recursive equation for the sequence of the posterior Fisher information matrices $\{F_i, i = 1, 2, \ldots\}$

$$F_{i+1} = F_{\mathrm{D}} + F_i \tag{43}$$

where

$$F_{\mathrm{D}} = \mathrm{E}\left[-\frac{\partial^2 \log p(\boldsymbol{y}_{i+1} | \boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^{\mathrm{T}}} \right] \tag{44}$$

represents the information obtained from the measurements at the next sensor node $i + 1$. Therefore, we can obtain our information utility measure based only on F_{D}.

We note that both of the above two information utility measures are expectations over all possible measurement values $\boldsymbol{y}_l \in \Re^N$ and, hence, can be computed before $\boldsymbol{y}_l$ is actually observed. Therefore, it is unnecessary for the current sensor to request the measurements from its neighboring nodes, thereby saving the energy of the sensor networks. However, to obtain the expectation, we need to compute a multidimensional integration numerically, which leads to high computation complexity at the current sensor node. Hence, these two measures are suitable for the cases when the dimension of the measurement vector $\boldsymbol{y}_i$ is low or sensor nodes have relatively high processing capabilities.

3) Mahalanobis Distance: At the current sensor node i, the utility function for selecting the next sensor node $i+1$ is defined as the negative of the Mahalanobis distance, i.e.,

$$-(\boldsymbol{r}_{i+1} - \widehat{\boldsymbol{\theta}}_i)^{\mathrm{T}} \Sigma_i^{-1} (\boldsymbol{r}_{i+1} - \widehat{\boldsymbol{\theta}}_i) \tag{45}$$

where $\boldsymbol{r}_{i+1}$ is the position of the next sensor node; $\widehat{\boldsymbol{\theta}}_i$ is the current estimate of the source position given by (22a) and Σ_i is covariance matrix of the current posterior density function $p(\boldsymbol{\theta} | \boldsymbol{y}_{1:i})$, as shown in (22b). Intuitively, this criterion means that

the closer a sensor lies to the longer axis of the uncertainty ellipsoid, the more information it would provide for the estimation.

This utility function is easy to calculate and works well when the current belief can be well approximated by a Gaussian distribution and the sensor measurement is a function of the distance between sensors and sources only. For a complex environmental scenario, for example, when there exist many obstacles in the environment, the above utility function may not be appropriate.

4) Covariance-Based Measure: In this case we derive utility measures based on the covariance Σ of the posterior distribution $p(\boldsymbol{\theta}|\boldsymbol{y}_{1:i+1})$, i.e., the covariance of the mmse estimate when we obtain the measurements from the next sensor $i+1$, using the formula in (20). We note that to calculate this covariance matrix, we need to obtain measurements of the sensor nodes in the neighborhood of the current node. In order to avoid requiring the measurements from each neighboring node, which leads to great communication cost, we propose predicting the covariance matrix $\mathrm{Cov}[\boldsymbol{\theta}|\boldsymbol{y}_{1:i+1}]$ according to the current belief and the prior knowledge about the sensor positions and sensing model.

First we apply the sensing model in (11) to predict the measurements $\boldsymbol{y}_{i+1}$ at the next sensor using the current mmse estimate as the source location. The next sensor positions and the sampling times are known and the bias term and noise variance can be predetermined in a calibration phase. The other unknown parameters, e.g., μ, are estimated by maximizing the likelihood function at the current sensor node. Then, we insert the predicted measurements in (13) to obtain the marginal probability density of the measurement $\boldsymbol{y}_{i+1}$ given $\boldsymbol{\theta}$. Hence, we can calculate the covariance matrix $\mathrm{Cov}[\boldsymbol{\theta}|\boldsymbol{y}_{1:i+1}]$ using (20). Even though this prediction is not very accurate, it carries decreased communication requirements, and its computation complexity is very much lower than the mutual information measure and the PCRB-based measure.

IV. Numerical Examples

In this section, we use numerical examples to illustrate the performance of our proposed distributed sequential Bayesian estimation algorithm. We will study the effects of the maximum communication range on estimation accuracy; compare the efficiency of different belief representation methods, namely the Gaussian approximation and the LPG function approximation; and consider the influence of the diffusion models as well as sensor selection criteria on the performance of the estimation algorithms.

In these examples, $N = 300$ wireless sensor nodes are randomly deployed in a square area with dimension 100×100 m. We define a pair of neighboring sensor nodes whose distance is less than the maximum communication range. For the diffusion model, we consider the environment as a homogeneous semi-infinite medium with a impermeable boundary, which fits a dispersion in air above the ground. We use a scenario of a stationary continuous point source located at the boundary surface, i.e., $z_0 = 0$. The other parameters μ, κ, and t_1 are taken to be 1 kg/s, 40 m^2/s, and 0 s, respectively.

A. Distributed Localization With Gaussian Approximation

In this numerical example, we implement the proposed distributed sequential Bayesian estimation method in Section III-A to localize a dispersion source whose position is $\boldsymbol{r}_0 = (73, 38)$. Here we approximate the belief by a Gaussian density function, and we use the Mahalanobis distance as the criterion to select the next sensor node. We assume the prior pdf of the source location is Gaussian, the mean of which is the position of the alarm sensor that first detects a possible release. We also assume for simplicity that there is no external force in the environment. The bias term in (11) is $b = 10^{-5}\ \mathrm{kg/m^3}$ and the noise standard deviation is $\sigma = 4 \times 10^{-6}\ \mathrm{kg/m^3}$. We take 10 temporal measurements at each sensor with the time sampling interval of 5 seconds. In Fig. 2 we show the message transmission route when applying our algorithm where the position of the initial sensor node is $\boldsymbol{r} = (44.396, 52.042)$; in Fig. 3 and Fig. 4 we describe the estimation bias $\sqrt{||\boldsymbol{\theta} - \widehat{\boldsymbol{\theta}}||}$ and the determinant of the covariance matrix Σ with respect to the number of used sensor nodes under different maximum sensor node communication ranges.

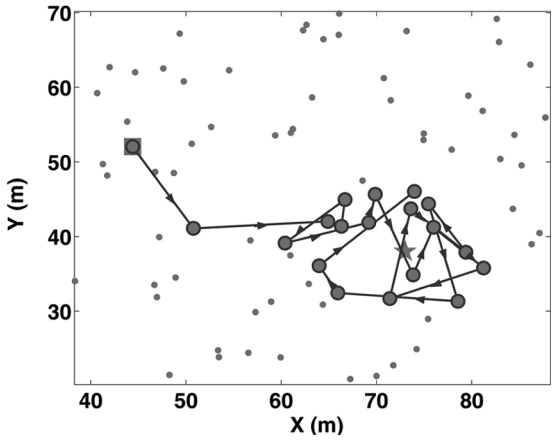


Fig. 2. Message transmission routing for localizing a diffusion source in a square area with $N = 300$ sensor nodes randomly placed. "Star" denotes the source position; "Square" denotes the initial sensor node; "Dot" denotes the deployed sensor nodes; "Circle" denotes the activated sensor nodes.

In Fig. 2 the solid line connects the sensor nodes that are selected (according to the sensor node selection criterion) to estimate the source position. The direction of the message transmission is shown by the arrows. We observe that our algorithm processes well: even though the initial sensor node is located far from the source, the messages still transmit to the area near the source and the data measured by the sensors in this area are collected to make the estimation. According to the dispersion model in (9), the data measured by the sensor nodes close the source have a high signal-to-noise ratio (SNR) and, thus, provide an important benefit in the estimation. Therefore, our distributed estimation algorithm provides a relatively optimal path to collect the information, which decreases the communication cost while maintaining the required estimation accuracy.

From Fig. 3 and Fig. 4 we find that the communication range (which is related to the wireless communication power) has an important effect on the estimation process. As shown in Fig. 4, when we set the determinant of the covariance $\det(\Sigma) = 10^{-6}$ as a threshold of the estimation performance and when we use the communication $\mathrm{radius\ range} = 15$ meters, the number of sensor nodes that are passed is 14, fewer than in the

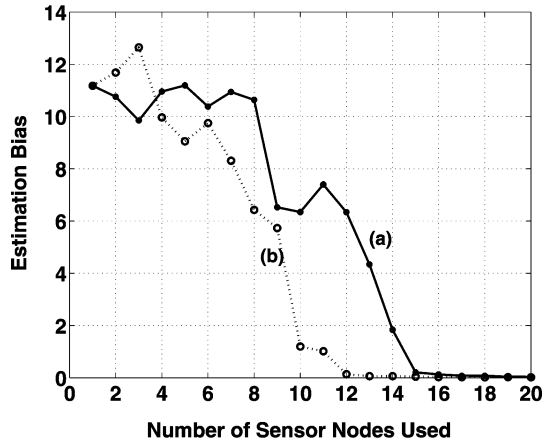


Fig. 3. Estimation bias versus number of sensor nodes used under different communication ranges. (a) Communication $\mathrm{range} = 10$ m. (b) Communication $\mathrm{range} = 15$ m.

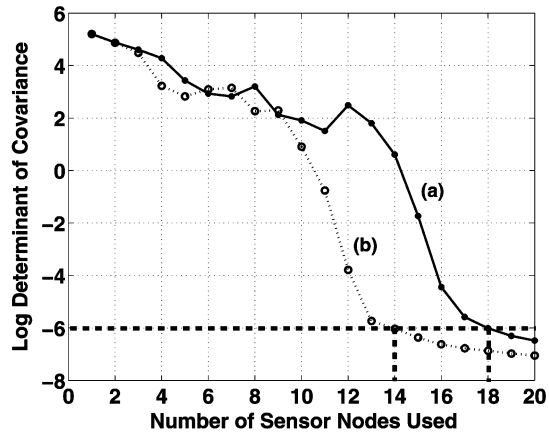


Fig. 4. Log determinant of the estimation covariance versus number of sensor nodes used under different communication ranges: (a) Communication $\mathrm{range} = 10$ m. (b) Communication $\mathrm{range} = 15$ m.

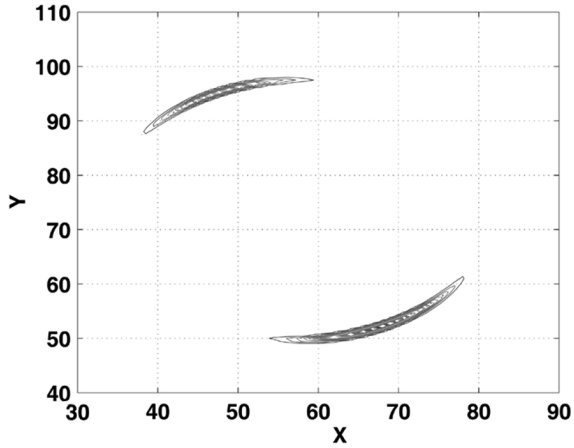


Fig. 5. The real posterior density function (belief) when we collect the measurements from three sensor nodes.

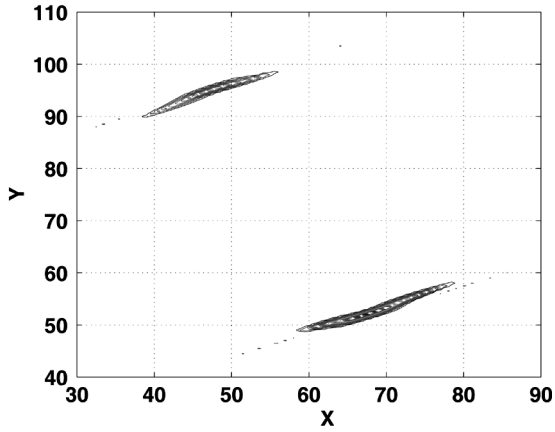


Fig. 6. The approximated posterior density function (belief) when we collect the measurements from three sensor nodes, using LPG function.

case when the communication range is 10 m. This result is reasonable since a long communication range means that the current sensor node can communicate with more other nodes in its neighborhood, and, thus, has advantages in finding a better path to collect the information; therefore, it leads to a lower number of used sensor nodes. However, the larger communication range means a higher power consumption; but a decrease in the number of wireless transmission hops. Therefore, in practice, we need to design this parameter carefully to find an optimal trade-off setting.

B. Distributed Localization With LPG Function Approximation

In this section, we illustrate the approximation performance of the LPG function and compare the estimation performance when we use the LPG approximation with the results when we use Gaussian approximation. In these examples, the diffusive source is located at $\boldsymbol{r}_0 = (62, 50)$. The bias term is $b = 10^{-5}\ \mathrm{kg/m^3}$ and the noise standard deviation is $\sigma = 8 \times 10^{-6}\ \mathrm{kg/m^3}$. We take five measurements at each sensor, and the time sampling interval is fixed at 5 s. We still use the Mahalanobis distance as the criterion to select the next sensor node.

In the first example, we implement the approximation of the belief density using LPG function. In Fig. 5 we show the true belief density (the belief without any approximation) after we collect the measurements of three sensor nodes. We find that the current belief includes two peaks, which is difficult to be approximated using only the Gaussian or Gaussian-sum density function. In Fig. 6, we draw the approximated belief density when we use the LPG function with the Hermite polynomials up to fifth-order to represent it. We observe that the approximation result is close to the true one. If we want to obtain a more accurate belief representation, or the belief density is very complex, we may need to use higher order Hermite polynomials, which would lead to high processing cost at each sensor node. Thus, there is a tradeoff between the computation complexity and the approximation accuracy. Note that when we apply the LPG function approximation, the order of the Hermite polynomials that need to be used will decrease when we collect the data from more sensor nodes.

In the next example, we compare the estimation performance of the proposed algorithm using the LPG function approximation with the Gaussian approximation. Namely, we illustrate the estimation bias and the log determinant of the estimation covariance matrix with respect to the number of used sensor

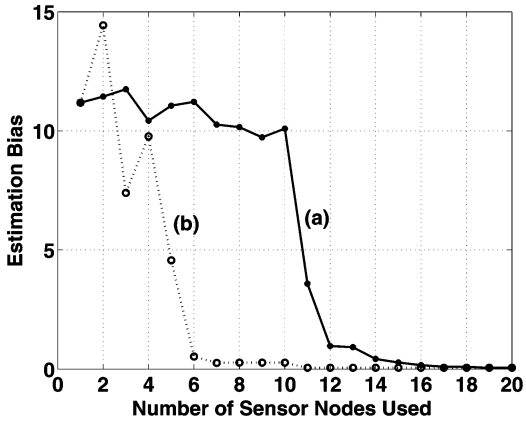


Fig. 7. Estimation bias versus number of sensor nodes used under different belief representation. (a) Gaussian density approximation. (b) LPG function approximation, for the examples in Section IV-B.

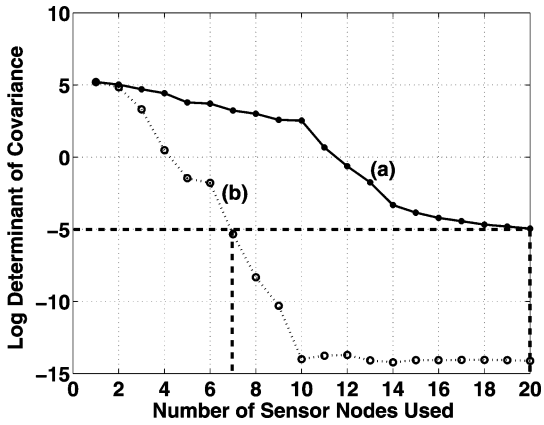


Fig. 8. Log determinant of the estimation covariance versus number of sensor nodes used under different belief representation. (a) Gaussian density approximation. (b) LPG function approximation, for the examples in Section IV-B.

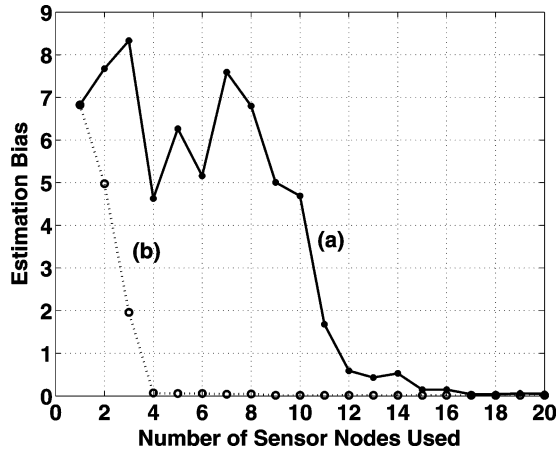


Fig. 9. Estimation bias versus number of sensor nodes used for different sensor node selection measure. (a) Mahalanobis distance. (b) Covariance-based measure, for the examples in Section IV-C.

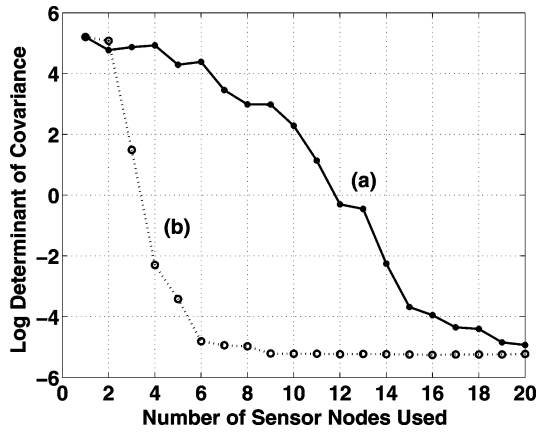


Fig. 10. Log determinant of the estimation covariance versus number of sensor nodes for different sensor node selection measure. (a) Mahalanobis distance. (b) Covariance-based measure, for the examples in Section IV-C.

nodes in Fig. 7 and Fig. 8, respectively. We observe that the distributed estimation using the LPG function approximation converges to the true value much faster than the Gaussian approximation, since the LPG function approximation provides us with a more accurate belief representation. In Fig. 8 we can see that when we use the LPG approximation, we need to use only seven sensor nodes to reach the performance threshold $\det(\Sigma) = 10^{-5}$, while for Gaussian approximation, we need 20 nodes. Even though the LPG function approximation needs to transmit more parameters between two sensor nodes, it decreases the data transmission hops and the number of sensors that are required to be activated; thus it can provide very low total sensor network power consumption (especially when the form of the belief is very complex, e.g., highly nonlinear and non-Gaussian) and a fast response speed, which is critical for some security applications.

C. Sensor Node Selection

In Section III-C, we proposed four different sensor node selection criteria. For the nonlinear and non-Gaussian measurement model, the mutual information-based measure and the posterior CRB provide a better characterization of the usefulness of the sensor data. However, the computational complexity to obtain these two measures is very high; hence, it requires significant processing capability of the sensor node, especially when the dimension of the measurement vector at each sensor node is large. Therefore, in this section we compare the performances of only two other utility measures: the Mahalanobis distance and the covariance-based measure.

In the examples in Section IV-A and IV-B, we used the Mahalanobis distance as the utility information measure and found that it worked well. However, in those examples we assumed a simple environment, i.e., a homogeneous semiinfinite medium without external force. In the following example, we assume there exists wind in the medium with the speed $\boldsymbol{v} = (-20, -10)$ meter/s. We implement the proposed distributed estimation algorithm of Section III-A with two different sensor selection criteria: the Mahalanobis distance and the covariance-based measure. We compare their performance in Fig. 9 and Fig. 10, and observe that under the current situation, the covariance-based measure performs better than the Mahalanobis distance, i.e., when we apply the covariance-based measure, the estimate of the source position approaches the true value much faster than using the Mahalanobis distance as a

measure. In addition, to reach a required performance threshold, we need to wake up fewer sensor nodes. As we presented in Section III-C-4, since we use only the prediction value of the covariance, the sensor node selection is not optimal; however, we save the required communication amount, and at the same time we incorporate the information of the dispersion model in the selection of the next sensor node. Hence, this selection criterion provides better performance than the Mahalanobis distance measure in a complex scenario.

V. Conclusion

In this paper, we addressed the problem of deriving efficient distributed estimation methods and apply them to the estimation of a diffusive source in wireless sensor networks. To develop our approaches, we first derived the substance dispersion model based on solving the diffusion equation under different environmental effects, and then we integrated the obtained physical model into the signal processing technologies and proposed a distributed sequential Bayesian estimation method. Actually, except for the applications concerned with a diffusive source, the proposed distributed estimation method provide a general framework for the other applications with stationary parameters in wireless sensor networks. In our algorithm, we dynamically update the belief density function in the space domain, which originates from the nonlinear recursive Bayesian filtering, i.e., we transmit the state belief in the sensor networks and update the belief using the measurements from the new sensor node. In this way, we realize a distributed estimation. We also implemented the idea of information-driven sensor collaboration for the sensor node selection, which provides an optimal subset and an optimal order of incorporating the measurements into our belief update, reduces the response time and saves on energy consumption of the sensor network. The belief representation is another important issue in our method because of the strict power and computation capability restrictions in wireless sensor networks. We propose two belief representation methods: Gaussian density approximation and a new LPG function approximation. Both of these approximations are suitable for wireless sensor networks and are applicable to different sensor network situations.

In future research, we will extend our proposed distributed processing methods to more complex environments (e.g., urban environments). In these cases, we will need to use numerical methods to solve the diffusion equation and obtain the physical model. Theoretically, the proposed algorithm can be generalized to include numerical solutions of the dispersion model. However, the involvement of the numerical results may influence the computation complexity and required communication amount; therefore, we need to investigate it carefully. Another research direction is to extend our methods to applications involving both stationary and dynamic state models in diffusion sources. The approaches in [2] provide some ideas on this topic. However, because of the specificities of the investigated diffusion model and the considered types of sensors (e.g., biochemical sensors), we must study new and practically fitting methods. We also plan to investigate new belief representations and sensor node selection methods by considering the tradeoff between signal processing performance, required communication amount, and computation complexity.

Acknowledgment

The authors would like to thank Dr. A. Dogandžić and Dr. E. Paldi for their helpful comments and discussions. We also thank anonymous reviewers for their careful checking and constructive comments.

References

[1] S. Kumar, F. Zhao, and D. Shepherd, Eds., "Special issue on collaborative signal and information processing in microsensor networks," *IEEE Signal Process. Mag.*, vol. 19, no. 2, Mar. 2002.

[2] F. Zhao and L. J. Guibas, *Wireless Sensor Networks: An Information Processing Approach*. San Francisco, CA: Morgan Kaufmann, 2004.

[3] A. Nehorai, B. Porat, and E. Paldi, "Detection and localization of vapor-emitting sources," *IEEE Trans. Signal Process.*, vol. SP-43, pp. 243–253, Jan. 1995.

[4] B. Porat and A. Nehorai, "Localizing vapor-emitting sources by moving sensors," *IEEE Trans. Signal Process.*, vol. SP-44, pp. 1018–1021, Apr. 1996.

[5] A. Jeremić and A. Nehorai, "Design of chemical sensor arrays for monitoring disposal sites on the ocean floor," *IEEE J. Ocean. Eng.*, vol. 23, pp. 334–343, Oct. 1998.

[6] ——, "Landmine detection and localization using chemical sensor array processing," *IEEE Trans. Signal Process.*, vol. SP-48, pp. 1295–1265, May 2000.

[7] T. Zhao and A. Nehorai, "Detecting and estimating biochemical dispersion of a moving source in a semi-infinite medium," *IEEE Trans. Signal Process.*, vol. 54, no. 6, pt. 1, pp. 2213–2225, Jun. 2006.

[8] F. Zhao, J. Shin, and J. Reich, "Information-driven dynamic sensor collaboration," *IEEE Signal Process. Mag.*, vol. 19, no. 2, pp. 61–72, March 2002.

[9] M. Chu, H. Haussecker, and F. Zhao, "Scalable information-driven sensor querying and routing for ad hoc heterogeneous sensor networks," *Int. J. High-Perform. Comput. Applicat.*, vol. 16, no. 3, pp. 90–110, 2002.

[10] S. M. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory*. Englewood Cliffs, NJ: Prentice-Hall, 1993.

[11] A. Mutambara, *Decentralized Estimation and Control for Multisensor Systems*. Boca Raton, FL: CRC, 1998.

[12] B. D. Anderson and J. B. Moore, *Optimal Filtering*. Englewood Cliffs, NJ: Prentice-Hall, 1979.

[13] R. S. Busy and K. D. Senne, "Digital synthesis of nonlinear filters," *Automatica*, vol. 7, pp. 287–298, 1971.

[14] A. Doucet, N. de Freitas, and N. J. Gordon, Eds., *Sequential Monte Carlo Methods in Practice*, ser. Series: Statistics for Eng. Inf. Sci.. New York: Springer-Verlag, 2001.

[15] M. Coates, "Distributed particle filters for sensor networks," in *3rd Int. Symp. Inf. Process. Sensor Netw.*, Apr. 26–27, 2004, pp. 93–107.

[16] M. Rosencrantz, G. Gordon, and S. Thrun, "Decentralized sensor fusion with distributed particle filters," in *Prof. Conf. Uncertainty in Artif. Intell.*, Acapulco, Mexico, Aug. 2003.

[17] X. Sheng, Y.-H. Hu, and P. Ramanathan, "Distributed particle filter with GMM approximation for multiple targets localization and tracking in wireless sensor network," in *Proc. 4th Int. Symp. Inf. Process. Sensor Netw.*, Apr. 15, 2005, pp. 181–188.

[18] J. Crank, *The Mathematics of Diffusion*, 2nd ed. Oxford: Oxford Univ. Press, 1975.

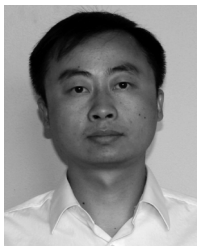
[19] W. Jost, *Diffusion in Solids, Liquids, Gases*. New York: Academic, 1952.

[20] J. M. Bernardo and A. F. M. Smith, *Bayesian Theory*. New York: Wiley, 1994.

[21] A. Azevedo-Filho and R. D. Shachter, "Laplace's method approximations for probabilistic inference in belief networks with continuous variables," in *Proc. 10th Conf. Uncertainty in Artif. Intell.*, San Mateo, CA, 1994, pp. 28–36.

[22] E. Paldi and A. Nehorai, "Exact nonlinear filtering via Gauss transform eigenfunctions," in *Proc. Int. Conf. Acoust., Speech, Signal Process.*, Apr. 1991, vol. 5, pp. 3389–3392.

[23] H. W. Sorenson and D. L. Alspach, "Recursive Bayesian estimation using Gaussian sums," *Automatica*, vol. 7, pp. 465–479, 1971.
[24] D. L. Alspach and H. W. Sorenson, "Nonlinear Bayesian estimation using Gaussian sum approximations," *IEEE Trans. Autom. Control*, vol. 17, no. 4, Aug. 1972.
[25] C. S. Withers, "A simple expression for the multivariate Hermite polynomials," *Stat. Probab. Lett.*, vol. 47, pp. 165–169, 2000.
[26] ——, "The moments of the multivariate normal," *Bull. Austral. Math. Soc.*, vol. 32, pp. 103–107, 1985.
[27] P. Tichavský, C. H. Muravchik, and A. Nehorai, "Posterior Cramér-Rao bounds for discrete-time nonlinear filtering," *IEEE Trans. Signal Process.*, vol. 46, no. 5, May 1998.



Tong Zhao (S'02) received the B.Eng. and M.Sc. degrees in electrical engineering from the University of Science and Technology of China, Hefei, China, in 1997 and 2000, respectively.

Currently, he is pursuing the D.Sc. degree with the Department of Electrical and Systems Engineering, Washington University, Saint Louis. His research interests are statistical signal processing, including detection and estimation theory, distributed signal processing, sequential Bayesian methods, Monte Carlo methods, etc., and their applications in radar, communications, biochemical sensor, sensor arrays, and wireless sensor networks.

Arye Nehorai (S'80–M'83–SM'90–F'94) received the B.Sc. and M.Sc. degrees in electrical engineering from the Technion, Israel, and the Ph.D. degree in electrical engineering from Stanford University, Stanford, CA.

From 1985 to 1995, he was a faculty member with the Department of Electrical Engineering, Yale University, New Haven, CT. In 1995, he joined the Department of Electrical Engineering and Computer Science, The University of Illinois at Chicago (UIC), as a Full Professor. From 2000 to 2001, he was Chair of the department's Electrical and Computer Engineering (ECE) Division, which then became a new department. In 2001, he was named University Scholar of the University of Illinois. In 2006, he became Chairman of the Department of Electrical and Systems Engineering, Washington University, St. Louis.

Dr. Nehorai is the inaugural holder of the Eugene and Martha Lohman Professorship and the Director of the Center for Sensor Signal and Information Processing (CSSIP) at WUSTL since 2006. He was Editor-in-Chief of the IEEE TRANSACTIONS ON SIGNAL PROCESSING during 2000 to 2002. During 2003 to 2005, he was Vice President (Publications) of the IEEE Signal Processing Society, Chair of the Publications Board, member of the Board of Governors, and member of the Executive Committee of this Society. He is the founding editor of the special columns on Leadership Reflections in the IEEE SIGNAL PROCESSING MAGAZINE. He was corecipient of the IEEE SPS 1989 Senior Award for Best Paper with P. Stoica, coauthor of the 2003 Young Author Best Paper Award, and corecipient of the 2004 Magazine Paper Award with A. Dogandzic. He was elected Distinguished Lecturer of the IEEE SPS for the term 2004 to 2005. He is the Principal Investigator of the new multidisciplinary university research initiative (MURI) project entitled Adaptive Waveform Diversity for Full Spectral Dominance. He has been a Fellow of the Royal Statistical Society since 1996.