A Spatial Sampling Scheme Based on Innovations Diffusion in Sensor Networks*

Zhi Quan, William J. Kaiser, and Ali H. Sayed
Electrical Engineering Department
University of California
Los Angeles, CA 90095 USA
Emails: {quan, kaiser, sayed}@ee.ucla.edu

ABSTRACT

This paper considers an estimation network of many distributed sensors with a certain correlation structure. Due to limited communication resources, the network selects only a subset of sensor measurements for estimation as long as the resulting fidelity is tolerable. We present a distributed sampling and estimation framework based on *innovations diffusion*, within which the sensor selection and estimation are accomplished through local computation and communications between sensor nodes. In order to achieve energy efficiency, the proposed algorithm uses a greedy heuristics to select a nearly minimum number of active sensors in order to ensure the desired fidelity for each estimation period. Extensive simulations illustrate the effectiveness of the proposed sampling scheme.

Categories and Subject Descriptors

H.4 [**Information Systems Applications**]: Communications Applications; C.2.4 [**Communication Networks**]: Distributed Systems; G.1.3 [**Numerical Linear Algebra**]: Linear Systems—*Signal Processing*

General Terms

Algorithm, Theory

Keywords

Distributed processing, sampling, estimation, innovations, mean-squared error, sensor networks

1. INTRODUCTION

A large class of wireless sensor networks (WSNs) are developed to estimate an underlying physical phenomenon over time and space in a noisy environment. The network consolidates data collected from sensor nodes to reconstruct the state of nature, e.g., estimating a field variable given the sensor observations in applications such as environmental monitoring, military surveillance, and space exploration [1]. The key issues are the fidelity at which the field variable can be estimated and the cost of operating the sensor network.

Because wireless sensor devices are usually battery-powered and the replacement is very difficult if not impossible, energy efficiency is critical for sensor networks and has a direct influence on the system lifetime. For senor networks with a dense deployment, it is necessary to select a group of sensors that are more informative for data fusion and to set other nodes inactive (or sleeping) in order to achieve energy efficiency. Although there might not always be a direct relation between energy efficiency and the number of active sensors, reducing the number of active sensors generally leads to energy efficiency. The goal of this paper is to find a nearly minimum number of sensor nodes that can achieve the desired estimation fidelity.

1.1 Linear Estimation

Consider a network of N sensor nodes estimating an unknown deterministic parameter $\boldsymbol{\theta} \in \mathcal{R}^m$. The observation of each sensor is distorted by a matrix $\mathbf{H}_i \in \mathcal{R}^{n_i \times m}$ and corrupted by additive noise, i.e.,

$$\mathbf{y}_i = \mathbf{H}_i\boldsymbol{\theta} + \mathbf{v}_i \qquad i = 1, 2, \ldots, N \tag{1}$$

Equation (1) can be written compactly as $\mathbf{y} = \mathbf{H}\boldsymbol{\theta} + \mathbf{v}$, where $\mathbf{y} = \mathrm{col}\{\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_N\}$, $\mathbf{H} = \mathrm{col}\{\mathbf{H}_1, \mathbf{H}_2, \ldots, \mathbf{H}_N\}$, and $\mathbf{v} = \mathrm{col}\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_N\}$. The measurement noise $\mathbf{v}$ is zero mean and has covariance matrix $\mathbf{C}$. Linear models of the form (1) are widely used in many sensing applications due to their mathematical tractability. For example, the power of an electromagnetic signal decays proportionally to the channel gain. In sensor target tracking problems [2], the state variables (such as position and velocity) evolve linearly with time and motivate a linear model.

At the beginning of each estimation period, the network selects a subset of the sensor measurements, denoted by $\mathcal{A}_n = \{i_1, i_2, \ldots, i_n\}$, for estimation as long as a desired estimation fidelity can be guaranteed. For a given $\mathcal{A}_n$, we use the notations $\mathbf{H}_{\mathcal{A}_n} = \mathrm{col}\{\mathbf{H}_{i_1}, \mathbf{H}_{i_1}, \ldots, \mathbf{H}_{i_n}\}$ and $\mathbf{y}_{\mathcal{A}_n} = \mathrm{col}\{\mathbf{y}_{i_1}, \mathbf{y}_{i_2}, \ldots, \mathbf{y}_{i_n}\}$. Let $\mathbf{C}_{\mathcal{A}_n}$ be the partial matrix selected from $\mathbf{C}$ with rows and columns corresponding to $\{i_1, i_2, \ldots, i_n\}$. With the data model in (1), the linear minimum-variance-unbiased-estimator (m.v.u.e.) [3] of $\boldsymbol{\theta}$ using $\mathbf{y}_{\mathcal{A}_n}$ is

*This work was partially supported by NSF grants ECS-0401188 and ECS-0601266.

IPSN'07, April 25-27, 2007, Cambridge, Massachusetts, USA.

given by

$$\hat{\boldsymbol{\theta}}\left(\mathbf{y}_{\mathcal{A}_n}\right) = \left(\mathbf{H}_{\mathcal{A}_n}^T \mathbf{C}_{\mathcal{A}_n}^{-1} \mathbf{H}_{\mathcal{A}_n}\right)^{-1} \mathbf{H}_{\mathcal{A}_n}^T \mathbf{C}_{\mathcal{A}_n}^{-1} \mathbf{y}_{\mathcal{A}_n} \quad (2)$$

and the resulting minimum mean-squared error (MMSE) is

$$\mathbf{D}\left(\mathcal{A}_n\right) = \left(\mathbf{H}_{\mathcal{A}_n}^T \mathbf{C}_{\mathcal{A}_n}^{-1} \mathbf{H}_{\mathcal{A}_n}\right)^{-1} \quad (3)$$

The MMSE of the estimator should be less than or equal to some desired distortion D_0, i.e.,

$$\begin{aligned} D_{\mathcal{A}_n} &= \mathrm{Tr}\left[\mathbf{D}\left(\mathcal{A}_n\right)\right] \\ &= \mathrm{E}\left(\left[\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}(\mathbf{y}_{\mathcal{A}_n})\right]^T \left[\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}(\mathbf{y}_{\mathcal{A}_n})\right]\right) \\ &\leq D_0 \end{aligned} \quad (4)$$

In this paper, the objective is to develop a method to select a subset of the sensors to estimate the unknown parameter $\boldsymbol{\theta}$ in a distributed way. The method will require each sensor to locally compute the m.v.u.e. of $\boldsymbol{\theta}$ under a certain distortion constraint by communicating with its nearby nodes. One major advantage of distributed solutions is that they save the cost of transmitting all data to a fusion center, and each sensor can operate as a data sink to achieve robustness.

1.2 Related Work

The problem of sensor selection has been investigated for various purposes. A recent work in [4] proposed an information driven sensor query (IDSQ) algorithm where only a *single* sensor is active at any given time and it passes its measurement to the most informative sensor in the network, which will be next active node. This work was extended in [5], which attempts to select the sensor measurement with maximum mutual information. In [6], the authors used a local greedy strategy to select the next most informative sensor node to reduce information entropy for target location. The problem of selecting sensors to minimize error in estimating the target position was investigated in [7]. However, none of these works consider energy efficiency and system lifetime [8].

There are several works addressing the issue of energy efficiency for sensor selection. In [9], a power scheduling scheme was proposed to minimize the total transmit power while satisfying a given estimation requirement. The effect of fading channels on estimation errors with power constraints was studied in [10]. The results implied that the sensors with bad channels and bad observation qualities should be turned off to conserve energy. In [11], a sensor selection algorithm was proposed to reconstruct the data image of spatially bandlimited physical phenomenon based on blue noise masking. With the assumption that all sensors are identical and the quality-of-service (QoS) can be expressed as a function of the sensor number, the paper [12] gave an adaptive sensor control scheme to minimize the number of active sensors. In [13], an innovations-based framework was proposed to select sensor measurements to estimate some unknown parameter under a desired distortion constraint. All these works considered a centralized network, in which there is a fusion center responsible for the network management and *global* computation.

Distributed algorithms are attractive in large-scale networks where a centralized solution is infeasible, non-scalable, or too costly. A number of distributed solutions have been proposed for detection, estimation, and inference [14]-[19]. In this context, each node does not need the global information and can perform local computation by communicating with nearby nodes.

1.3 Paper Outline

In this paper, we develop a distributed sensor selection strategy. We consider a wireless sensor network with many redundant nodes, each of which can observe a physical phenomenon in the field. For each estimation period, the network chooses a subset of sensor measurements $\mathcal{A}_n$ to achieve the desired fidelity. We refer to the procedure of selecting sensor measurements for estimation as *sensor sampling*. Intuitively, if the noise level is low, a small number of sensors is sufficient to achieve the desired fidelity; however, if the noise condition is severe, more sensors should be activated for accurate estimation. We propose a sampling and estimation framework that exploits a strategy called *innovations diffusion*. Innovation refers to the new information that a sensor measurement contributes to the reduction of the estimation error, and diffusion means the process by which the innovation is communicated across the network. Within this framework, the sampling procedure is accomplished in a distributed and automated way, and each active sensor locally computes the m.v.u.e. $\hat{\boldsymbol{\theta}}$ based on the selected measurements. The proposed algorithm activates a nearly minimum number of sensors to ensure the desired fidelity D_0 at the beginning of each estimation period.

This paper is organized as follows. Section 2 describes the system model for innovations diffusion sampling in sensor networks. The proposed framework for sampling and estimation is presented in section 3. Simulation results illustrating the effectiveness of the proposed algorithms are given in section 4. Section 5 concludes the paper with discussion on future research.

2. SYSTEM MODEL

Consider a wireless network with N sensor nodes spatially distributed in the field, as depicted in Fig. 1. The network wishes to select a group of sensor measurements to estimate the unknown parameter $\boldsymbol{\theta}$ in a *distributed* manner by relying on local computation and inter-sensor communication.

2.1 Network Graph

The sensor network is represented as an undirected graph $G(\mathcal{V}, \mathcal{E})$, where $\mathcal{V} = \{1, 2, \ldots, N\}$ denotes the set of vertices (sensor nodes) and $\mathcal{E} \subset \{(i,j) \mid i,j \in \mathcal{V}\}$ is the edge set. A graph is connected if there exists a path in $\mathcal{E}$ for any two vertices i and j. In this paper, we assume that each sensor transmits at a constant power level P and the receiver has an ambient noise power level N_0. Let $X(i)$ represent the position of sensor i. Then, the transmission from sensor i is successfully received by a sensor j if

$$\frac{P}{N_0 d_{ij}^{\alpha}} \geq \beta \quad (5)$$

where $d_{ij} = |X(i) - X(j)|$ and α $(2 < \alpha < 6)$ is the known pathloss coefficient. Equation (5) models a situation where a minimum signal-to-noise ratio (SNR) is necessary for successful receptions and signal power decays with distance r

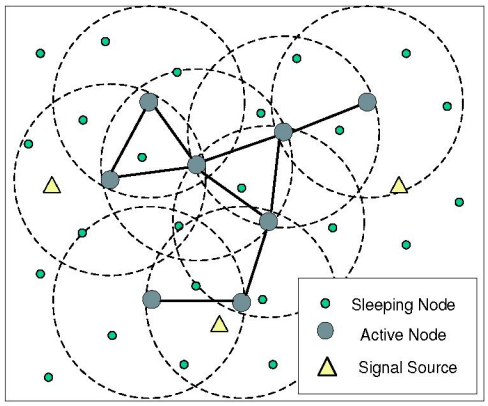

Figure 1: A schematic representation of distributed sampling in wireless sensor networks.

as $1/r^{\alpha}$. Thus, the receiver should be within a distance of

$$r_\beta = \sqrt[\alpha]{\frac{P}{N_0\beta}} \tag{6}$$

from the transmitter such that a reliable wireless link can be ensured. Therefore, the edge set is defined by

$$\mathcal{E} \triangleq \{(i,j) \mid d_{ij} \leq r_\beta\} \tag{7}$$

and the neighbor set of sensor i is defined as

$$\mathcal{N}(i) \triangleq \{j \mid (i,j) \in \mathcal{E}\} \tag{8}$$

Sensors i and j are termed *neighbors* if their distance is less than or equal to r_β. As illustrated in Fig. 1, sensors within the distance of the connectivity radius r_β are directly connected and are neighbors of each other.

For a sampling decision $\mathcal{A}_n = \{i_1, i_2, \ldots, i_n\}$, the active and inactive neighbors of sensor i_k ($i_k \in \mathcal{A}_n$) can be respectively represented as

$$\mathcal{N}_A(i_k) = \mathcal{N}(i_k) \cap \mathcal{A}_n \tag{9}$$

and

$$\mathcal{N}_I(i_k) = \mathcal{N}(i_k) \setminus \mathcal{A}_n \tag{10}$$

The local knowledge of sensor i contains not only its own distortion matrix and noise covariance but also its neighbors' distortion matrices, noise covariances, and correlations. Let $\mathcal{N}'(i) = \{\mathcal{N}(i), i\}$. Then, each sensor i has the knowledge of $\{\mathbf{H}_{\mathcal{N}'(i)}, \mathbf{C}_{\mathcal{N}'(i)}\}$ at the stage of deployment.

2.2 Spatial Correlation Models

The computation of the linear m.v.u.e. of $\boldsymbol{\theta}$ from the noisy measurements can be facilitated if the network has *a priori* knowledge of the covariance structure $\mathbf{C}$. In practice, the matrix $\mathbf{C}$ can be estimated from repeated measurements at all sensors. Nevertheless, this approach is not efficient if the number of sensors N is large. Therefore, a model that characterizes the correlation between sensors will be useful. In this paper, we consider three different correlation models as follows.

2.2.1 Uncorrelated Noise (UC)

The measurement noise at each sensor is uncorrelated with other sensors. If the noises are Gaussian, then they are independent with each other as well.

2.2.2 Near Correlation (NC)

In this model, each sensor is only correlated with its neighboring nodes and is independent of other sensors beyond the distance r_β, i.e., $\mathbf{C}_{ij} = \mathbf{0}$ if $j \notin \mathcal{N}(i)$. This is motivated by the fact that the correlation decays with the distance between sensors and approaches zero if the corresponding sensors are far apart [20].

2.2.3 Far Correlation (FC)

Each sensor is not only correlated with its own neighbors, but also with other sensors. The correlation is assumed to be a decreasing function of the Euclidean distance between the corresponding nodes [20, 21], i.e.,

$$\mathbf{C}_{ij} = \mathbf{F}\left(\mathbf{C}_{ii}, \mathbf{C}_{jj}, d_{ij}\right) \tag{11}$$

where $\mathbf{F} : \mathcal{R}^{n_i \times n_i} \times R^{n_j \times n_j} \times \mathcal{R} \mapsto \mathcal{R}^{n_i \times n_j}$.

These correlation models will be used to test our distributed sampling and estimation algorithms. Although the design of distributed algorithms considerably depends on the correlation model, we show that these models can be unified in the proposed framework.

3. INNOVATIONS DIFFUSION SAMPLING

In this section, we present a distributed framework for sampling and estimation under a certain distortion constraint. Starting with an initial sensor, the set of active sensors collaboratively activates one sleeping sensor at each step. The procedure continues until the set of active sensors achieves the desired estimation fidelity. Although there is no direct relation between the system lifetime and the number of active sensors, reducing the number of active sensors generally leads to energy efficiency. In general, finding the exact minimum number of measurements to achieve a desired fidelity is an NP-hard problem. Thus, the proposed approach will pursue a greedy solution that provides useful sub-optimal approximations at polynomial complexity. Assuming that each inactive sensor can listen to the control channel for the awakening message, at each step the sampling algorithm will select the most informative sensor with respect to the previous ones in order for the fidelity to be satisfied with the nearly minimum number of active sensors. The results will exhibit good performance at reasonable cost and will be tractable for both analysis and implementation.

3.1 Uncorrelated Noise

If the noises are spatially uncorrelated, i.e., $\mathbf{C}_{kl} = \mathbf{0}$ for $k \neq l$, then (3) can be written as

$$\mathbf{D}\left(\mathcal{A}_n\right)^{-1} = \sum_{k=1}^{n} \mathbf{H}_{i_k}^T \mathbf{C}_{i_k i_k}^{-1} \mathbf{H}_{i_k} \tag{12}$$

This expression decouples the contribution of each sensor to the total MMSE value $\mathbf{D}(\mathcal{A}_n)$. Each term $\mathbf{H}_{i_k}^T \mathbf{C}_{i_k i_k}^{-1} \mathbf{H}_{i_k}$ has the essential properties of an information measure in that it is

1. non-negative definite,
2. and additive for independent observations.

Intuitively, the more information sensor i_k has (i.e., the larger $\mathbf{H}_{i_k}^T \mathbf{C}_{i_k i_k}^{-1} \mathbf{H}_{i_k}$ is), the more its contribution to reducing the MMSE. This suggests that at each step the network

should choose the most informative sensor in order to maximally reduce the MMSE. Specifically, assume that the nodes $\mathcal{A}_{n-1}$ have been selected. Then, at step n we would choose from the coverage of $\mathcal{A}_{n-1}$ a sensor i_n whose information measure $\mathbf{H}_{i_n}^T \mathbf{C}_{i_n}^{-1} \mathbf{H}_{i_n}$ is the largest. In this way, the resulting $\mathbf{D}(\mathcal{A}_n)$ will be the smallest compared with other choices.

3.2 Correlated Noise

However, the noises are generally spatially correlated in practice. In this case, the contributions of the individual sensors are coupled in the MMSE expression $\mathbf{D}(\mathcal{A}_n)$ in (3). We thus need to develop a procedure to find the most informative sensor with respect to the previous selected ones. To achieve this goal, we start by whitening the observation data subject to the order dictated by the choice of sensors, and then obtain a set of transformed measurements with uncorrelated noises.

Suppose that we have already selected $n-1$ sensors, i.e., $\mathcal{A}_{n-1} = \{i_1, i_2, \ldots, i_{n-1}\}$. For $i_n \notin \mathcal{A}_{n-1}$, we define its innovation [3, 22] as

$$\mathbf{e}_{i_n} \triangleq \mathbf{y}_{i_n} - \hat{\mathbf{y}}_{i_n|\mathcal{A}_{n-1}} \tag{13}$$

where

$$\hat{\mathbf{y}}_{i_n|\mathcal{A}_{n-1}} = \mathbf{H}_{i_n}\boldsymbol{\theta} + \hat{\mathbf{v}}_{i_n|\mathcal{A}_{n-1}} \tag{14}$$

with $\hat{\mathbf{v}}_{i_n|\mathcal{A}_{n-1}}$ representing the projection of $\mathbf{v}_i$ onto the affine space of the previous selected measurements, denoted by $\mathcal{L}\{\mathbf{y}_{i_1}, \mathbf{y}_{i_2}, \ldots, \mathbf{y}_{i_{n-1}}\}$. The quantity $\mathbf{e}_{i_n}$ in (13) possesses the new information contained in sensor i_n and not in any of the previous measurements $\{\mathbf{y}_{i_1}, \mathbf{y}_{i_2}, \ldots, \mathbf{y}_{i_{n-1}}\}$. Now note that $\hat{\mathbf{v}}_{i_n|\mathcal{A}_{n-1}}$ is given by

$$\hat{\mathbf{v}}_{i_n|\mathcal{A}_{n-1}} = \mathbf{B}_{i_n|\mathcal{A}_{n-1}} \mathbf{C}_{\mathcal{A}_{n-1}}^{-1} \mathbf{v}_{\mathcal{A}_{n-1}} \tag{15}$$

where

$$\mathbf{B}_{i_n|\mathcal{A}_{n-1}} = \begin{pmatrix} \mathbf{C}_{i_n i_1} & \mathbf{C}_{i_n i_2} & \ldots & \mathbf{C}_{i_n i_{n-1}} \end{pmatrix} \tag{16}$$

Combining (1), (13), (14), and (15) gives

$$\begin{aligned} \mathbf{e}_{i_n} &= \mathbf{v}_{i_n} - \hat{\mathbf{v}}_{i_n|\mathcal{A}_{n-1}} \\ &= \begin{pmatrix} -\mathbf{B}_{i_n|\mathcal{A}_{n-1}} \mathbf{C}_{\mathcal{A}_{n-1}}^{-1} & \mathbf{I} \end{pmatrix} \begin{pmatrix} \mathbf{v}_{\mathcal{A}_{n-1}} \\ \mathbf{v}_{i_n} \end{pmatrix} \end{aligned} \tag{17}$$

It can be verified that $\mathbf{e}_{i_n} \perp \mathbf{v}_j$ for any $j \in \mathcal{A}_{n-1}$. Fig. 2 illustrates a geometrical interpretation of the projection relationship between $\mathbf{e}_i$, $\hat{\mathbf{v}}_{i_n|\mathcal{A}_{n-1}}$, and $\mathbf{v}_i$. Thus, the matrix

$$\mathbf{P}_{i_n|\mathcal{A}_{n-1}} \triangleq \begin{pmatrix} -\mathbf{B}_{i_n|\mathcal{A}_{n-1}} \mathbf{C}_{\mathcal{A}_{n-1}}^{-1} & \mathbf{I} \end{pmatrix} \tag{18}$$

projects $\mathbf{v}_{i_n}$ onto a space orthogonal to the previous noises $\mathcal{L}\{\mathbf{v}_{i_1}, \mathbf{v}_{i_2}, \ldots, \mathbf{v}_{i_{n-1}}\}$.

LEMMA 1 (INNOVATIONS). *Given a sampling decision $\mathcal{A}_n$, the corresponding innovation process $\{\mathbf{e}_{i_k}\}_{k=1}^n$ has the important property*

$$\mathrm{E}(\mathbf{e}_{i_k} \mathbf{e}_{i_l}^T) = \begin{cases} \mathbf{0} & i_k \neq i_l \\ \mathbf{Q}_{i_k} & \textit{otherwise} \end{cases} \tag{19}$$

where $\mathbf{Q}_{i_k}$ is the covariance matrix of $\mathbf{e}_{i_k}$ and has the form

$$\mathbf{Q}_{i_k} = \mathbf{C}_{i_k i_k} - \mathbf{B}_{i_k|\mathcal{A}_{k-1}} \mathbf{C}_{\mathcal{A}_{k-1}}^{-1} \mathbf{B}_{i_k|\mathcal{A}_{k-1}}^T \tag{20}$$

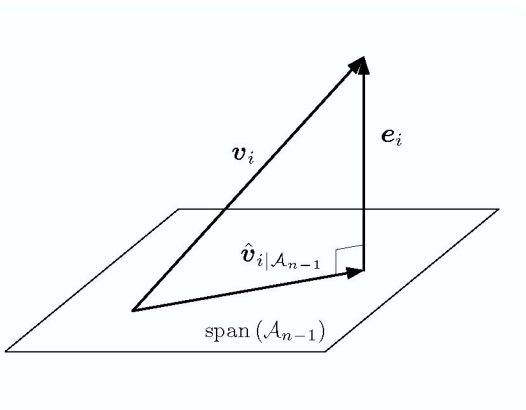

Figure 2: A geometrical interpretation of the relations between vectors $\mathbf{v}_i$, $\mathbf{e}_i$, and $\hat{\mathbf{v}}_{i|\mathcal{A}_{n-1}}$ with respect to the space spanned by $\{\mathbf{v}_{i_1}, \mathbf{v}_{i_2}, \ldots, \mathbf{v}_{i_{n-1}}\}$, denoted by span $(\mathcal{A}_{n-1})$.

We now introduce a transformed form of $\mathbf{y}_{i_n}$ with respect to $\{\mathbf{y}_{i_1}, \mathbf{y}_{i_2}, \ldots, \mathbf{y}_{i_{n-1}}\}$ as

$$\mathbf{z}_{i_n} \triangleq \mathbf{P}_{i_n|\mathcal{A}_{n-1}} \begin{pmatrix} \mathbf{y}_{\mathcal{A}_{n-1}} \\ \mathbf{y}_{i_n} \end{pmatrix} \tag{21}$$

i.e.,

$$\mathbf{z}_{i_n} = \mathbf{G}_{i_n|\mathcal{A}_{n-1}} \boldsymbol{\theta} + \mathbf{e}_{i_n} \tag{22}$$

where

$$\mathbf{G}_{i_n|\mathcal{A}_{n-1}} = \mathbf{P}_{i_n|\mathcal{A}_{n-1}} \begin{pmatrix} \mathbf{H}_{\mathcal{A}_{n-1}} \\ \mathbf{H}_{i_n} \end{pmatrix} \tag{23}$$

The contribution of $\mathbf{z}_{i_n}$ to the MMSE is additive respect to the previous measurements in $\mathcal{A}_{n-1}$ since[1]

$$\mathbf{z}_{i_n} \perp \mathcal{L}\{\mathbf{y}_{i_1}, \mathbf{y}_{i_2}, \ldots, \mathbf{y}_{i_{n-1}}\} \tag{24}$$

Moreover, the linear m.v. u.e. of $\boldsymbol{\theta}$ given $\{\mathbf{z}_{i_1}, \mathbf{z}_{i_2}, \ldots, \mathbf{z}_{i_n}\}$ coincides with the m.v.u.e. using $\{\mathbf{y}_{i_1}, \mathbf{y}_{i_2}, \ldots, \mathbf{y}_{i_n}\}$. This result is stated as follows.

THEOREM 1 (EQUIVALENT ESTIMATION). *Consider the linear model in (1). For a given set of sensor measurements $\{\mathbf{y}_{i_1}, \mathbf{y}_{i_2}, \ldots, \mathbf{y}_{i_n}\}$, the minimum-variance-unbiased linear estimator of $\boldsymbol{\theta}$ is equivalent to that estimated using the transformed measurements as in (22), i.e.,*

$$\begin{aligned} &\hat{\boldsymbol{\theta}}(\mathbf{y}_{\mathcal{A}_n}) = \hat{\boldsymbol{\theta}}(\mathbf{z}_{\mathcal{A}_n}) \\ &= \left(\sum_{k=1}^n \mathbf{G}_{i_k|A_{k-1}}^T \mathbf{Q}_{i_k}^{-1} \mathbf{G}_{i_k|A_{k-1}} \right)^{-1} \sum_{k=1}^n \mathbf{G}_{i_k|A_{k-1}}^T \mathbf{Q}_{i_k}^{-1} \mathbf{z}_{i_k} \end{aligned} \tag{25}$$

and the resulting MMSEs are identical, i.e.,

$$\begin{aligned} \mathbf{D}(\mathcal{A}_n) &= \left(\mathbf{H}_{\mathcal{A}_n}^T \mathbf{C}_{\mathcal{A}_n}^{-1} \mathbf{H}_{\mathcal{A}_n} \right)^{-1} \\ &= \left(\sum_{k=1}^n \mathbf{G}_{i_k|\mathcal{A}_{k-1}}^T \mathbf{Q}_{i_k}^{-1} \mathbf{G}_{i_k|\mathcal{A}_{k-1}} \right)^{-1} \end{aligned} \tag{26}$$

[1]Specifically, for $i_n \notin \mathcal{A}_{n-1}$ and all $i_k \in \mathcal{A}_{n-1}$, it has $\mathrm{E}\left(\mathbf{z}_{i_n} - \mathrm{E}\mathbf{z}_{i_n}\right)^T \left(\mathbf{y}_{i_k} - \mathrm{E}\mathbf{y}_{i_k}\right) = 0$.

The advantage of working with the transformed quantities $\{\mathbf{z}_{i_k}\}_{k=1}^n$ is that the noises $\{\mathbf{e}_{i_k}\}_{k=1}^n$ in the model (22) are now uncorrelated and the contributions of the individual sensors to the MMSE can be decoupled.

3.3 Iterative Diffusion Sampling

Diffusion is the means by which innovations become useful by spreading throughout the current active sensors. The sampling algorithm assumes an active node at the beginning, i.e., $\mathcal{A}_1 = \{i_1\}$, which serves as the seed to activate other sensors. Sensor i_1 chooses within its inactive neighbors a sensor and then activates it. At the n-th step, the current active sensors $\mathcal{A}_{n-1} = \{i_1, i_2, \ldots, i_{n-1}\}$ make a connected network and collaboratively activate one sleeping sensor that is within the coverage of $\mathcal{A}_{n-1}$. To meet the fidelity requirement with a nearly minimum number of measurements, the sampling algorithm should choose the sensor that is most informative with respect to the previous sensors $\mathcal{A}_{n-1}$. Before we proceed to present the algorithm, we need to define a utility function that measures the innovation of the sensor measurement.

According to (26), we define a *utility matrix* as

$$\mathbf{U}_{i_n|\mathcal{A}_{n-1}} \triangleq \mathbf{G}_{i_n|\mathcal{A}_{n-1}}^T \mathbf{Q}_{i_n}^{-1} \mathbf{G}_{i_n|\mathcal{A}_{n-1}} \tag{27}$$

which has the essential properties as an information measure. Recall that the set of sensor observations $\mathcal{A}_{n-1}$ has an MMSE $D_{\mathcal{A}_{n-1}} = \mathrm{Tr}\left[\mathbf{D}\left(\mathcal{A}_{n-1}\right)\right]$. From (26), the MMSE with measurements $\{\mathcal{A}_{n-1}, i_n\}$ can be written as

$$D_{\mathcal{A}_n} = \mathrm{Tr}\left[\left(\mathbf{D}^{-1}\left(\mathcal{A}_{n-1}\right) + \mathbf{U}_{i_n|\mathcal{A}_{n-1}}\right)^{-1}\right] \tag{28}$$

Then, we can define a utility function as the difference between $D_{\mathcal{A}_{n-1}}$ and $D_{\mathcal{A}_n}$, i.e.,

$$\begin{aligned} & u(i_n|\mathcal{A}_{n-1}) \triangleq D_{\mathcal{A}_{n-1}} - D_{\mathcal{A}_n} \\ =& \mathrm{Tr}\left[\mathbf{D}\left(\mathcal{A}_{n-1}\right) - \left(\mathbf{D}^{-1}\left(\mathcal{A}_{n-1}\right) + \mathbf{U}_{i_n|\mathcal{A}_{n-1}}\right)^{-1}\right] \\ =& \mathrm{Tr}\left[\mathbf{D}\left(\mathcal{A}_{n-1}\right)\left(\mathbf{D}\left(\mathcal{A}_{n-1}\right) + \mathbf{U}_{i|\mathcal{A}_{n-1}}^{-1}\right)^{-1}\mathbf{D}\left(\mathcal{A}_{n-1}\right)\right] \end{aligned} \tag{29}$$

where the last equality follows the matrix inversion lemma [3]:

$$(\mathbf{A}+\mathbf{BCD})^{-1} = \mathbf{A}^{-1} - \mathbf{A}^{-1}\mathbf{B}\left(\mathbf{C}^{-1} + \mathbf{DA}^{-1}\mathbf{B}\right)^{-1}\mathbf{DA}^{-1}$$

Although the definition of the utility function is not unique, it can be seen that $u(i_n|\mathcal{A}_{n-1})$ in (29) indicates a good measure of the new information provided by sensor i_n.

The basic strategy of diffusion sampling is to successively choose one sensor with the maximum utility from the sleeping nodes within the coverage of $\mathcal{A}_{n-1}$, i.e.,

$$i_n = \arg \max_{j \in \bigcup_{k=1}^{n-1} \mathcal{N}_I(i_k)} u\left(j|\mathcal{A}_{n-1}\right) \tag{32}$$

by local computation and message exchange between neighboring sensors. The procedure continues until the desired estimation fidelity D_0 is achieved. The details of the sampling algorithm are presented in Algorithm 1.

Consider the case that each node has full knowledge of the distortion matrices $\{\mathbf{H}_i\}_{i=1}^N$ and the covariance matrix $\mathbf{C}$ at the stage of deployment. Then each active sensor can locally compute (30) without communicating with its neighboring sensors. The messages exchanged between sensors only need to contain the index and associated value of the utility function, i.e.,

Algorithm 1 Innovations Diffusion Sampling

0: Start with $n = 1$ and $\mathcal{A}_1 = \{i_1\}$.
1: **while** $D_{\mathcal{A}_n} > D_0$ **do**
2: For each $i_k \in \mathcal{A}_{n-1}$, compute

$$i_k^{\max} = \arg \max_{j \in \mathcal{N}_I(i_k)} u(j|\mathcal{A}_{n-1}) \tag{30}$$

3: **repeat**
4: Each active node i_k sends a message M_{i_k} including the index $i_k^{\max}$ and the associated utility $u(i_k^{\max}|\mathcal{A}_{n-1})$ to its active neighbors $\mathcal{N}_A(i_k)$.
5: Upon receiving the messages from neighboring nodes, each sensor i_k in $\mathcal{A}_{n-1}$ updates its value of $i_k^{\max}$ by

$$i_k^{\max} = \arg \max_{j \in \mathcal{N}_A(i_k)} u(i_j^{\max}|\mathcal{A}_{n-1}) \tag{31}$$

and stores the associated utility.
6: **until** all the active sensors reach the consensus.
7: Let $i_n = i_k^{\max}$.
8: Sensor i_n is activated by its closet active neighbor and inherits the necessary information related to $\mathcal{A}_{n-1}$.
9: Each active sensor adds i_n into the set of active sensors, i.e., $\mathcal{A}_n = \{i_1, i_2, \ldots, i_n\}$ and $n = n + 1$.
10: **end while**

$$M_{i_k} = \{i_k^{\max}, u(i_k^{\max}|\mathcal{A}_{n-1})\} \tag{33}$$

However, the sensors might not have global neighborhood knowledge of the whole network in some circumstances. In such situations, the diffusion sampling algorithm can be implemented by adding necessary information into the messages that are exchanged between nodes. Then the sensors can collaborate with each other to activate the next sensor. To achieve this goal, the message sent by sensor i_k in line 4 should contain the distortion matrix and noise covariance of sensor $i_k^{\max}$ in addition to $\{i_k^{\max}, u(i_k^{\max}|\mathcal{A}_{n-1})\}$, i.e.,

$$M_{i_k} = \{i_k^{\max}, u(i_k^{\max}|\mathcal{A}_{n-1}), \mathbf{H}_{i_k^{\max}}, \mathbf{C}_{i_k^{\max} i_k^{\max}}\} \tag{34}$$

To avoid transmitting redundant information, in line 4 of the algorithm each active sensor sends the message M_{i_k} to its neighbors only if the message content has been updated; otherwise, it turns off radio to save energy.

Since the number of active nodes is finite, the consensus can be reached in line 6 within a finite number of iterations. Then, each sensor can locally construct the matrices $\mathbf{H}_{\mathcal{A}_n}$ and $\mathbf{C}_{\mathcal{A}_n}$, which will be used for activating the next sensor if $D_{\mathcal{A}_n} > D_0$ and for data fusion otherwise. For the UC and NC models, constructing $\mathbf{C}_{\mathcal{A}_n}$ is straightforward. If sensor noises are characterized by the FC model, then each active sensor can locally construct $\mathbf{C}_{\mathcal{A}_n}$ using (11) with the assumption that sensor localization has been performed, i.e., sensor i_k knows the position of sensor i_n. Afterward, when sensor i_n is activated in line 8, it inherits the information of $\{\mathbf{H}_{\mathcal{A}_{n-1}}, \mathbf{C}_{\mathcal{A}_{n-1}}\}$ from its closest active neighbor and locally constructs its own $\{\mathbf{H}_{\mathcal{A}_n}, \mathbf{C}_{\mathcal{A}_n}\}$.

In brief, the algorithm successively adds the sensor that has the maximum utility into the set of selected sensors until the MMSE is less than or equal to the desired estimation fidelity level. This scheme ensures that the set of selected nodes make a connected graph at any time and the number of active sensor is no more than necessary.

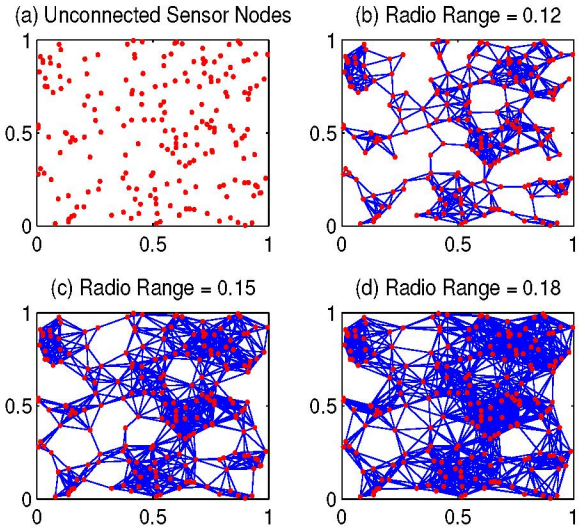

Figure 3: The connected network of $N = 200$ randomly distributed sensors with different radio ranges. (a) $r_\beta = 0$; (b) $r_\beta = 0.12$; (c) $r_\beta = 0.15$; (d) $r_\beta = 0.18$.

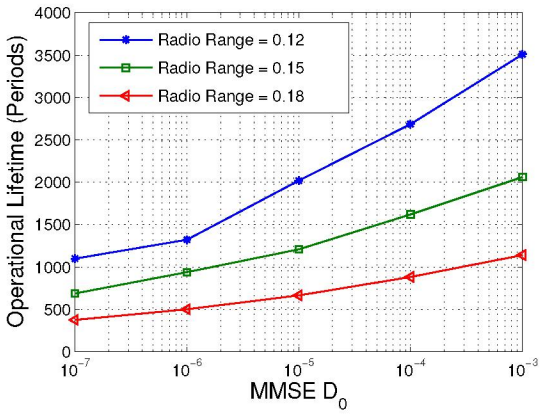

Figure 4: The operational lifetime versus the estimation fidelity constraint in a wireless sensor network with constant transmit power. For each radio range, the results are averaged over 100 simulations.

4. NUMERICAL EXAMPLES

This section presents numerical results that illustrate the effectiveness of our proposed sampling scheme. We are in particular interested in the operational system lifetime of the networked sensors.

4.1 Simulation Setup

We randomly generate $N = 200$ sensor nodes in a unit square $[0, 1]$ by $[0, 1]$. The connectivity of the randomly distributed sensors is shown in Fig. 3. Consider a simple linear model $y_i = \theta + v_i$, $i = 1, 2, \ldots, N$. The noise covariance matrix $\mathbf{C}$ is randomly generated according to the spatial correlation model

$$C_{ij} = \begin{cases} \sigma_i^2 & i = j \\ \sigma_i \sigma_j \exp\left(-a d_{ij}^2\right) & i \neq j \end{cases} \tag{35}$$

where σ_i^2, $i = 1, 2, \ldots, N$, are randomly generated with a uniform distribution in $(0, 1]$. The correlation is characterized as an exponential function of the distance between nodes.

The sensors are assumed to have a unit of initial energy after deployment. Each work period starts with a sampling stage and ends with an estimation stage, where the duration of the estimation stage is assumed to be much longer than that of the sampling stage such that the cost caused by sampling does not result in much overhead. In the sampling stage and starting with a seed node that has maximum left-over energy, a group of sensors are activated for estimating the unknown parameter $\boldsymbol{\theta}$ later in the estimation stage, and then are powered off at the end of the period. When the next period starts, a new group of sensors will be activated and the procedure repeats until the sensors deplete their energy. We neglect the energy consumed by sensors during the inactive duration, assuming the amount of energy required for listening to the awakening signal is much less than the amount required for computing and communication. It is assumed that each sensor has an omnidirectional radio antennas at its transmitter, with which the sensor can transmit a common message to all of its neighbors simultaneously with a single broadcast.

4.2 Constant Transmit Power

In this scenario, the transmitter at each sensor has a constant transmit power level $P(r_\beta)$ for a given radio range r_β. We assume that during the sampling stage the computation energy of (30) expended by sensor i is $E_i^{\text{comp}} = K_1$ and the inter-node communication energy expended by sensor i in each iteration is proportional to

$$E_i^{\text{comm}}(r_\beta) = \frac{K_2 P(r_\beta)}{\log_2(1+\beta)} \tag{36}$$

where $P(r_\beta) = r_\beta^\alpha N_0 \beta$, K_1 and K_2 are certain constants. During the estimation stage, the total energy expended by sensor i is assumed to be

$$E_i^{\text{est}}(r_\beta) = \frac{K_3 P(r_\beta)}{\log_2(1+\beta)} \tag{37}$$

In our simulations, we set $E_i^{\text{comp}} = 0.00001$, $E_i^{\text{comm}}(0.12) = 0.0001$, $E_i^{\text{est}}(0.12) = 0.01$, and $\alpha = 3.5$.

Figure 4 illustrates the operational lifetime of the network with different radio ranges against the fidelity constraint D_0. As seen from the observation, the network lifetime increases when the fidelity constraint becomes relaxed. The reason behind this is that during each work period, a slack fidelity requirement requires only a small number of measurements while a stringent fidelity constraint needs a large number of active sensors. The number of sensors activated during each work period has a direct influence on the network lifetime.

The radio ranges of sensors are also of importance for the operational lifetime. Provided that the sensor positions are fixed, the number of neighboring nodes of each sensor is determined by the radio range r_β. More specifically, if the sensor has a larger radio range, i.e., the transmitter has a larger transmit power, then it can communicate with more neighboring sensors. Consequently, a large radio range will be able to keep good connectivity when some sensor nodes

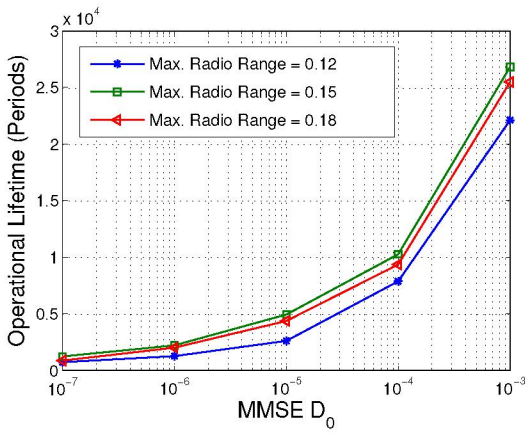

Figure 5: The operational lifetime versus the estimation fidelity constraint in a wireless sensor network with adaptive transmit power. For each constraint on the radio range, i.e., $r_{\max} = 0.12, 0.15$, and 0.18, the results are averaged over 100 simulations.

run out of energy. However, a larger radio range requires a larger transmit power level at the transmitter, which may drain out the energy or sensors more quickly. Given a constant transmit power level $P(r_\beta)$ at each sensor, Fig. 4 depicts the system operational lifetimes with respect to different radio ranges. It can be observed that the sensors run out of energy more quickly with a larger radio range.

4.3 Adaptive Transmit Power

We consider that each sensor can adjust its transmit power such that the desired SNR β at the receiver can be achieved with a minimum transmit power level at the transmitter. Thus, the transmit power of active sensor i depends on the distance from its current active neighbors, i.e.,

$$P_i = \max_{j \in \mathcal{N}_A(i)} d_{ij}^{\alpha} N_0 \beta \tag{38}$$

In the simulation setup, the amounts of energy expended during the sampling and estimation stages are respectively given by (36) and (37) as well, except that the transmit power now uses (38). Specifically, we set $E_i^{\text{comp}} = 0.00001$, $E_i^{\text{comm}} = 0.0001 \times P_i/P(0.12)$, $E_i^{\text{est}} = 0.01 \times P_i/P(0.12)$, and $\alpha = 3.5$.

When choosing the radio range, there is a trade-off between energy expenditure and maintaining the network connectivity. Using large radio ranges will draw off the energy quickly. On the other hand, larger radio ranges may keep the network connected and may activate a more informative sensor with a large probability at each iteration. Therefore, the number of active sensors required to satisfy the desired fidelity may be reduced, resulting in a longer operational lifetime. The simulation results are presented in Fig. 5 with different radio power constraints, i.e., $r_{\max} = 0.12, 0.15$, and 0.18. It shows that a mediate radio range $r_{\max} = 0.15$ has the best performance in terms of the system lifetime.

Compared to the constant transmit power, the adaptive scheme can achieve a significant improvement in the system lifetime because the desired SNR can be obtained with a minimum transmit power level and the sensor energy can be more efficiently used.

5. CONCLUSION

In this paper, we proposed an innovation diffusion sampling scheme for distributed estimation in sensor networks. The framework suggests that selecting the most informative sensor measurements for estimation will reduce the number of active sensors and prolong the operational system lifetime. The main advantages of the proposed scheme lie in that it can be implemented efficiently in an asynchronous and scalable way. The performance of the sampling scheme is evaluated through numerical simulations. Some interesting extensions are worth further investigation for important practical issues such as finite-bit communication, transmission errors, and reliable communication protocols.

6. REFERENCES

[1] I. F. Akyildiz, W. Su, Y. Sankarsubramaniam, and E. Cayirci. Wireless sensor networks: A survey. *Computer Networks*, 38:393–422, Mar. 2002.

[2] F. Gustafsson, F. Gunnarsson, N. Bergman, U. Forssell, J. Jansson, R. Karlsson, and P.-J. Nordlund. Particle filters for positioning, navigation, and tracking. *IEEE Trans. Signal Processing*, 50(2):425–437, Feb. 2002.

[3] A. H. Sayed. *Fundamentals of Adaptive Filtering*. John Wiley & Sons, NJ, 2003.

[4] F. Zhao, J. Shin, and J. Reich. Information-driven dynamic sensor collaboration. *IEEE Signal Processing Magazine*, 19(1):61–72, 2002.

[5] E. Ertin, J. Fisher, and L. Potter. Maximum mutual information principle for dynamic sensor query problems. In *Proc. IPSN*, pages 405–416, Palo Alto, CA, Apr. 2003.

[6] H. Wang, G. Pottie, K. Yao, and D. Estrin. Entropy-based sensor selection heuristic for target localization. In *Proc. IPSN*, pages 36–45, Berkeley, CA, Apr. 2004.

[7] V. Isler and R. Bajcsy. The sensor selection problem for bounded uncertainty sensing models. In *Proc. IPSN*, pages 151–158, Los Angeles, CA, Apr. 2005.

[8] Q. Dong. Maximizing system lifetime in wireless sensor networks. In *Proc. IPSN*, pages 13–19, Los Angeles, CA, Apr. 2005.

[9] J. Xiao, S. Cui, Z. Q. Luo, and A. J. Goldsmith. Power scheduling of universal decentralized estimation in sensor networks. *IEEE Trans. Signal Processing*, 54(2):413–422, Feb. 2006.

[10] S. Cui, J. Xiao, A. J. Goldsmith, Z.-Q. Luo, and H. V. Poor. Estimation diversity and energy efficiency in distributed sensing. *IEEE Trans. Signal Processing*, 2007.

[11] M. Perillo, Z. Ignjatovic, and W. Heinzelman. An energy conservation method for wireless sensor networks employing a blue noise spatial sampling technique. In *Proc. IPSN*, pages 116–123, Berkeley, CA, Apr. 2004.

[12] Z. Hu, J. Zhang, and L. Tong. Adaptive sensor activity control in many-to-one sensor networks. *IEEE J. Select. Areas Commun.*, 24(8):1525–1534, Aug. 2006.

[13] Z. Quan and A. H. Sayed. Innovations-based sampling over spatially-correlated sensors. In *Proc. IEEE ICASSP*, Honolulu, Hawaii, Apr. 2007.

[14] V. Saligrama, M. Alanyali, and O. Savas. Distributed detection in sensor networks with packet losses and finite capacity links. *IEEE Trans. Signal Processing*, 2007.

[15] V. Delouille, R. Neelamani, and R. Baraniuk. Robust distributed estimation in sensor networks using the embedded polygons algorithm. In *Proc. IPSN*, pages 405–413, Berkeley, CA, Apr. 2004.

[16] L. Xiao, S. Boyd, and S. Lall. A scheme for robust distributed sensor fusion based on average consensus. In *Proc. IPSN*, pages 63–70, Los Angeles, CA, Apr. 2005.

[17] R. Nowak, U. Mitra, and R. Willett. Estimating inhomogeneous fields using wireless sensor networks. *IEEE J. Select. Areas Commun.*, 22(6):999–1006, Aug. 2004.

[18] M. Rabbat and R. Nowak. Distributed optimization in sensor networks. In *Proc. IPSN*, pages 20–27, Berkeley, CA, Apr. 2004.

[19] C. Lopes and A. H. Sayed. Distributed processing over adaptive networks. In *Proc. Adaptive Sensor Array Processing Workshop*, MIT Lincoln Laboratory, MA, June 2006.

[20] N. A. C. Cressie. *Statistics for Spatial Data*. John Wiley & Sons, NJ, 1993.

[21] M. C. Vuran, O. B. Akan, and I. F. Akyildiz. Spatio-temporal correlation: theory and applications for wireless sensor networks. *Computer Networks*, 45:245–259, 2004.

[22] T. Kailath, A. H. Sayed, and B. Hassibi. *Linear Estimation*. Prentice-Hall, NJ, 2000.