SEQUENTIAL ESTIMATION OVER NOISY CHANNELS WITH DISTRIBUTED NODE SELECTION

Thakshila Wimalajeewa and Sudharman K. Jayaweera

Department of Electrical and Computer Engineering
University of New Mexico
Albuquerque, NM 87131-0001, USA
Email: {thakshilaw,jayaweera}@ece.unm.edu

ABSTRACT

In an autonomous sensor network without a central fusion center, it is desired that any node has the ability to make the final decision once it has enough information about the Phenomenon of Interest (PoI) within a certain confidence level. In this paper, we propose a distributed sequential methodology which updates the current node's estimator based on its own observation and noise corrupted decision from the previous node. We show that sequential processing is useful only when the channel quality of inter-sensor communication links satisfies a certain condition. We develop a distributed node selection algorithm to select the order of processing nodes based on information utilities and the communication cost. In the proposed scheme, each node only needs to keep track of its neighbor nodes leading to reduced complexity. Simulation results show that a significant reduction in the required number of processing nodes to achieve a desired performance level is obtained compared to that with nearest node selection method.

1. INTRODUCTION

In a typical wireless sensor network (WSN), each node communicates wirelessly with other nodes or the fusion center in its radio communication range. A particular challenge in WSN's is the need for distributed estimation algorithms that efficiently allocate limited energy resources at a node for communication and sensing. In most sensor network applications considered in literature it is assumed that the spatially separated sensor nodes send their locally processed information to a fusion center that makes the final decision about the state of the nature [1, 2]. There are, however, sensor network applications in which the system needs to have the ability to make a final decision at any given distributed node. Doing this, only the nodes that contribute to the final decision are in the decision process while other nodes can remain idle preserving their transmit energy. Since sensor networks are usually intended to last for long periods of time, it is important for the nodes to process and communicate local data only if necessary.

The distributed sequential estimation problem was formulated in [3–5]. According to [3, 4], a leader node sequentially queries sensor nodes and updates its estimator (based on the posterior distribution of the state of the PoI) until a desired performance level is reached. In these schemes, the leader node has to keep track of all the nodes which have been participated in the decision process at each step. Several information utility measures based on entropy and the geometry were also proposed in [4] to select the next best node to be participated in the decision process. In [5], the posterior distribution (belief) at the current node is transmitted to the next node where it updates the state of belief based on current node's belief and the new measurement at the next node. In [6], a node selection algorithm for target tracking based on the posterior Cramer-Rao Lower Bound (CRLB) was presented. However, neither of these considered the noise in inter-sensor communications. In [7], the sequential estimation of a non-random parameter over noisy correlated channels was considered. However, it did not consider the best ordering of the nodes in the decision processing.

In this paper, on the other hand, we consider the distributed sequential estimation of a random parameter in which the updated estimate of a node is sent to the next node through a noisy channel. The ordering of nodes for the estimation process is selected based on a given criteria. We consider the next node selection as a trade-off between an information utility measure and the communication cost between nodes. In [8], a distributed architecture is presented in solving inference problems in which nodes in the network assemble themselves into a network junction tree where the presented algorithms minimize the computation and communication cost required by inference. Contrast to that in [8], in our distributed algorithms, we formulate the cost function as a trade-off between information gains and communication costs. Also, our formulation combines the information in the current node and the previous node in a sequential manner contrast to work in [8]. In fact, we assume that each node has a set of neighbors that it can communicate with an affordable communication cost. The candidate next nodes at each node are allowed to be selected only from its neighbors. In the proposed scheme, each node has to keep track of only its neighbors to determine which nodes have been participated in the decision process. The information utility measures of the possible candidate nodes are computed according to the current node's information and the knowledge of positions of neighbor nodes and target positions. From simulation results we see that when the proposed node selection criteria is used, an improved performance with a less number of nodes can be achieved compared to the nearest node selection method.

The remainder of this paper is organized as follows: Section 2 formulates the distributed sequential estimation problem over noisy channels and derives the MMSE performance. The distributed node selection schemes based on information utility measures and the inter-node communication cost are discussed in Section 3. Section 4 shows the performance results and concluding remarks are given in Section 5.

2. SENSOR NETWORK MODEL AND THE SEQUENTIAL ESTIMATION ALGORITHM

We consider a spatially distributed, n-node sensor network in which there is no designated fusion center. Denote by s_k the k-th node, for $k = 1, \cdots, n$. Note that, when there is no ambiguity, we use s_k and k to denote the k-th processing node interchangeably. The network is deployed to estimate a target amplitude based on the following observation model at node s_k: $w'_k = \frac{\theta}{\|\mathbf{x}_k - \mathbf{x}_t\|^{\alpha/2}} + v'_k$, for $k = 1, \cdots, n$, where θ is the parameter to be estimated (target amplitude) that is assumed to be Gaussian with zero mean and variance σ_θ^2, $\mathbf{x}_k$ and $\mathbf{x}_t$ denote the position of sensor node s_k and the target, respectively, v_k is the measurement noise that is assumed to be white Gaussian with zero mean and variance σ_0^2 and α is the path loss exponent that is determined by the propagation environment. This model can be used, for example, in applications in which acoustic sensors are used to estimate the amplitude of sound signals emitted by a target [4,9]. By rearranging, we can re-write the observation at the node s_k in the equivalent form of

$$w_k = \theta + v_k, \text{ for } k = 1, \cdots, n, \tag{1}$$

where now v_k is assumed to be independent but not identically distributed. In particular, v_k is Gaussian with mean zero and variance $\sigma_k^2 \propto \|\mathbf{x}_k - \mathbf{x}_t\|^\alpha$.

The first sensor has the observation $z_1 = w_1 = \theta + v_1$. Assuming that the parameter θ is independent of noise v_1, the optimal Minimum Mean Squared Error (MMSE) estimate at node s_1 based on z_1 is given by

$$\hat{\theta}_1(w_1) = \frac{\sigma_\theta^2}{\sigma_\theta^2 + \sigma_1^2} w_1, \tag{2}$$

and the corresponding MMSE, denoted by M_1, of the estimator (2) is

$$M_1 = \frac{\sigma_1^2 \sigma_\theta^2}{\sigma_1^2 + \sigma_\theta^2} = \left(\frac{1}{\sigma_1^2} + \frac{1}{\sigma_\theta^2} \right)^{-1}. \tag{3}$$

(2) can also be expressed as $\hat{\theta}_1(w_1) = \frac{M_1}{\sigma_1^2} w_1$.

If the MMSE M_1 does not meet the desired performance, the estimator $\hat{\theta}_1$ is transmitted to the next node over a noisy channel. The criteria for selection of next node is discussed later in this paper. For $k > 1$, the node s_k estimates the parameter θ based on its own observation and the received estimator from the node s_{k-1}. The observation vector at node s_k is

$$\mathbf{z_k} = \begin{bmatrix} w_k \\ y_k \end{bmatrix} = \begin{bmatrix} \theta + v_k \\ \hat{\theta}_{k-1} + n_k \end{bmatrix}, \text{ for } k = 2, \cdots, n,$$

where y_k is the noise corrupted decision from node s_{k-1}. The channel noise n_k, from node s_{k-1} to node s_k is assumed to be independent Gaussian with mean zero and variance $\sigma_{c(k-1,k)}^2 \propto \|\mathbf{x}_k - \mathbf{x}_{k-1}\|^\alpha$ for $k = 2 \cdots, n$. The MMSE estimator at node s_k can thus be written as [10],

$$\hat{\theta}_k(w_k, y_k) = \frac{M_k}{\sigma_k^2} w_k + \frac{M_k(\sigma_\theta^2 - M_{k-1})}{M_{k-1}(\sigma_\theta^2 - M_{k-1}) + \sigma_\theta^2 \sigma_{c(k-1,k)}^2} y_k, \tag{4}$$

where M_k is the MMSE at the node s_k that can be shown to be

$$M_k = \frac{\sigma_\theta^2}{\sigma_\theta^2 d_k^2 + 1}, \tag{5}$$

where $d_k^2 = \frac{1}{\sigma_k^2} + \frac{(\sigma_\theta^2 - M_{k-1})^2}{\sigma_\theta^2 \left[M_{k-1}(\sigma_\theta^2 - M_{k-1}) + \sigma_\theta^2 \sigma_{c(k-1,k)}^2 \right]}$ and M_{k-1} is the MMSE at the node s_{k-1} that is assumed to be available at the node s_k. Note that the MMSE at the node s_k is determined only by statistics of observations and it is reasonable to assume that they can be made available at neighbors [7]. From (4), it can be seen that the MMSE estimator at the node s_k is determined by its own observation, information from the node s_{k-1} and the channel noise quality.

It is interesting to examine the dependence of M_k in (5) on the channel quality of the inter-sensor communication links. When channel quality is good

$$\lim_{\sigma_{c(k-1,k)}^2 \to 0} M_k = \frac{M_{k-1}}{1 + \frac{M_{k-1}}{\sigma_k^2}}, \text{ for } k = 2 = 1, \cdots, n.$$

Therefore, it is seen that when $\sigma_{c(k-1,k)}^2 \to 0$, $M_k \leq M_{k-1}$ for all k. That is, by sending the decision at node s_{k-1} to the node s_k always improves the MMSE performance at node s_k. On the other hand, if the inter-sensor communication channel quality is poor, we have

$$\lim_{\sigma_{c(k-1,k)}^2 \to \infty} M_k = \frac{\sigma_\theta^2 \sigma_k^2}{\sigma_\theta^2 + \sigma_k^2}, \text{ for } k = 2, \cdots, n. \tag{6}$$

That is, when the quality of inter-sensor communication link is poor, the performance at the node s_k does not depend on the decision at node s_{k-1}, but is entirely determined by the observation quality at the node s_k. It can be shown that if for all k, $\sigma_{c(k-1,k)}^2$ satisfies the inequality

$$\sigma_{c(k-1,k)}^2 \leq \frac{M_{k-1}^2(\sigma_\theta^2 - M_{k-1})}{\sigma_k^2(\sigma_\theta^2 - M_{k-1}) - M_{k-1}\sigma_\theta^2}, \tag{7}$$

then $M_k \leq M_{k-1}$; i.e. sending the decision at node s_{k-1} to the node s_k improves the MMSE performance at the node s_k. This is further illustrated in Section 4 via simulation results.

If we assume that the node observations are i.i.d and the inter-node communication is noiseless such that $\sigma_k^2 = \sigma_0^2$ and $\sigma_{c(k-1,k)}^2 = 0$ for $k = 1, \cdots, n$, it can be shown that the MMSE at the node s_k (5) reduces to $M_k = \frac{\sigma_\theta^2 \sigma_0^2}{\sigma_v^2 + k\sigma_\theta^2}$, which is a monotonically decreasing function of k. It is also interesting to see that in this case, the minimum number of nodes n_{min} required to achieve a required MMSE performance level ϵ_d is given by, $n_{min} = \sigma_0^2 \left(\frac{1}{\epsilon_d} - \frac{1}{\sigma_\theta^2} \right)$.

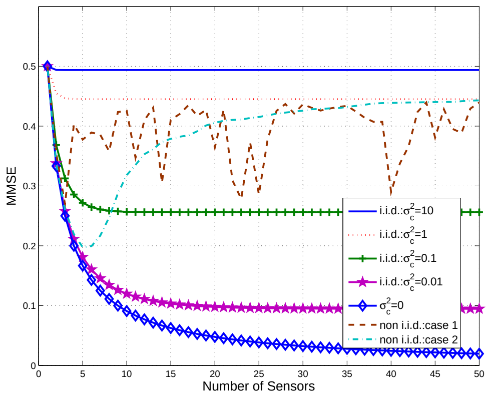


Fig. 1. MMSE vs. number of sensor when observation noise is i.i.d.

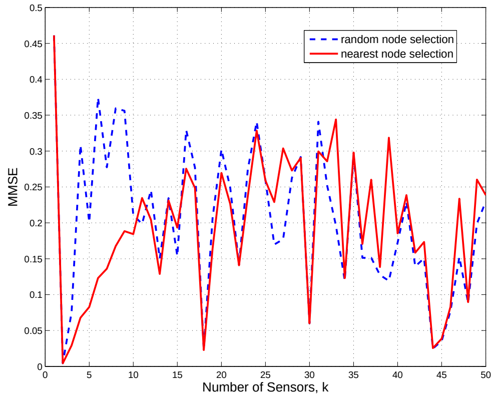


Fig. 2. MMSE vs number of sensors when observation noise is non-i.i.d.

In the case of i.i.d. observation noise such that $\sigma_k^2 = \sigma_0^2$ for all k, Fig. 1 shows the MMSE performance of the sequential estimation process with different channel noise qualities on inter-sensor communication links. In Fig. 1, we have let $\sigma_0^2 = 1$ and $\sigma_\theta^2 = 1$. In the special case when channel noise is also i.i.d. such that $\sigma_{c(k-1,k)}^2 = \sigma_c^2$ for all k, from Fig. 1 it can be seen that $M_k \leq M_{k-1}$ holds for all k. Moreover, as expected from (6) the MMSE performance converges to 0.5 as σ_c^2 increases. It is expected that when both observations and channel noise are i.i.d., the performance of the MMSE estimator is independent of the order of the processing nodes. Figure 1 also shows the performance of the MMSE estimator when channel noise is not identical (still the observation noise is i.i.d.). We have considered two cases: In the first, $\sigma_{c(k-1,k)}^2$'s are drawn randomly from a uniform distribution in $[0, 1]$ without any order. In the second case, these random $\sigma_{c(k-1,k)}^2$'s are arranged in an ascending order. From Fig. 1, it can be seen that whenever the condition (7) is satisfied at node s_k, $M_k \leq M_{k-1}$. In this case, to find the node where the minimum MMSE is achieved, the process should be continued for all nodes. On the other hand, in case 2, where nodes are selected with minimum distance from the current node, we can observe that after a certain node, the MMSE is monotonically increasing. Therefore, it is enough to continue the sequential estimation process only until this specific node, thereby, saving the network power.

Figure 2 shows the MMSE performance of the sequential distributed estimation process with non identical observations and channel noise. Dashed line corresponds to channel noise variance drawn from a uniform distribution without any order while the solid line corresponds to channel noise variance in ascending order with k. In both cases, the observation noise variances are drawn from a uniform distribution on $[0, 1]$. As can be observed from Fig. 2, when observations are not i.i.d., just selecting the nearest node as the next node does not always improve the performance. Therefore, when observations are not identical, it is required to have an information driven approach to select the nodes with higher information gain as well as lower communication cost.

3. SENSOR NODE SELECTION

3.1. Distributed node selection: global approach

In a sequential estimation process, selecting nodes which provide best information gain with affordable communication cost would lead to an improved performance by minimizing the network resource consumption. Essentially, it is important to determine the best ordering of nodes that would complete the estimation process by reaching at the desired performance level with a minimum number of processing nodes as a trade-off between the information gain and the communication cost. Let us denote by $\mathcal{V} = \{1, 2, \cdots, n\}$ the set of nodes in the network. Let $\mathcal{V}_j$ denote the set of nodes that have been participated in the sequential estimation process up to step j. Let $s_j \in \mathcal{V}$ be the selected processing node at step j. Then the next node s_{j+1} at step $(j+1)$ is chosen as,

$$s_{j+1} = \underset{s_k \in \mathcal{V}_j^c}{\operatorname{argmax}}\ R(s_j, s_k) \tag{8}$$

where $\mathcal{V}_j^c$ denotes the set complement of $\mathcal{V}_j$ with respect to $\mathcal{V}$ and the objective function $R(s_j, s_k)$ is defined as,

$$R(s_j, s_k) = \beta R_I(\theta, w_{s_k}, \hat{\theta}_j) - (1-\beta) R_c(s_j, s_k), \tag{9}$$

where $R_I(.)$ and R_c are the information utility function and a measure of communications cost and $\beta \in [0, 1]$ is a trade-off parameter that balances the contributions from the two terms in (9). The choice of β will depend on the required information gain and the tolerable communications cost. There are several possible information utility measures that can be used to quantify the information gain provided by a sensor measurement. For example, [4], [5] provided a detailed description of entropy- and geometry-based information utility measures. In this paper, we consider the MMSE at the node s_{j+1} when the current node is s_j, denoted $M_{j+1|j}$, for $s_j, s_{j+1} \in \mathcal{V}$, as the information utility measure. Then we have

$$R_I(\theta, w_{s_k}, \hat{\theta}_j) = -\frac{\sigma_\theta^2}{\sigma_\theta^2 d_{j,k}^2 + 1} \tag{10}$$

where $d_{j,k}^2 = \frac{1}{\sigma_k^2} + \frac{(\sigma_\theta^2 - M_j)^2}{\sigma_\theta^2 \left[M_j(\sigma_\theta^2 - M_j) + \sigma_\theta^2 \sigma_{c(j,k)}^2 \right]}$.

The communication cost function between current node s_j and the possible next node s_k is taken as, $R_c(s_j, s_k) = \frac{1}{d_{max}}(\mathbf{x}_{s_j} - \mathbf{x}_{s_k})^T(\mathbf{x}_{s_j} - \mathbf{x}_{s_k})$ where d_{max} is the maximum distance between any two sensors in the network. Then the composite objective function (9) can be written as,

$$\begin{aligned} R(s_j, s_k) &= \beta R_I(\theta, w_{s_k}, \hat{\theta}_j) \\ &- \frac{(1-\beta)}{d_{max}}(\mathbf{x}_{s_j} - \mathbf{x}_{s_k})^T(\mathbf{x}_{s_j} - \mathbf{x}_{s_k}), \quad (11) \end{aligned}$$

where $R_I(\theta, w_{s_k}, \hat{\theta}_j)$ is as given in (10). To find the next best processing node, the node s_j has to compute the reward function (11) for all candidate sensors in $\mathcal{V}_j^c$. Moreover, each node s_j has to keep track of the nodes that have already been participated in the estimation process up to step j. Due to these, implementing this scheme distributively would be difficult. Thus, in the following, we propose a distributed algorithm for sensor node selection with reduced computational complexity that still improves the performance of the sequential estimation process significantly.

3.2. Distributed node selection: local approach

Assume that the node s_k in the network has a set of neighbors $\mathcal{N}_k$ for $k = 1, \cdots, n$ where the neighbors are determined based on a node's effective communication range and the affordable communication cost. We assume that each node has the same effective communication range so that the criteria for selection of neighbors is the same for all nodes. Moreover, if node s_k is a neighbor of node s_i, for $i \neq k$, then node s_i is also a neighbor of node s_k as well. Let s_j be the current processing node at step j. The node s_j selects the next node based on the objective function (10) from the set of candidate sensors $\mathcal{C}_j^{s_j}$ that is its neighbor nodes who have have not been participated in the estimation process previously. Note that each node $s_j \in \mathcal{V}$ updates its set of candidate sensors $\mathcal{C}_j^{s_j}$ based on the information received from its neighbors. Each node has to keep track of the nodes participated in the estimation process only in its neighborhood. The next node s_{j+1} at step $(j+1)$ is chosen as,

$$s_{j+1} = \underset{s_k \in \mathcal{C}_j^{s_j}}{\operatorname{argmax}} R(s_j, s_k). \quad (12)$$

To find the next best node, the node s_j has to compute the reward function (11) only for all candidate nodes in $\mathcal{C}_j^{s_j}$. The proposed distributed sequential estimation process is summarized in Algorithm 1.

Note that since node s_j selects the next node from the candidate set $\mathcal{C}_j^{s_j}$, node s_j only needs to perform $|\mathcal{C}_j^{s_j}|$ number of computations. Also, node s_j needs to keep track of the nodes which are not participated in the decision process in its neighborhood only. Once the final decision is made, a signal is broadcast implying the final decision is made. Then all unprocessed nodes go to sleep mode, until the next event occurs.

Algorithm 1 Sequential estimation process at step j at node s_j

while ($j \geq$)1 **do**
  Compute estimate $\hat{\theta}_j$
  Compute MMSE M_j
  if ($M_j <$ Desired performance or $\mathcal{C}_j^{s_j} = \emptyset$) **then**
    Make final decision
    Go to sleep mode
  else
    1. Select next node from $\mathcal{C}_j^{s_j}$
    2. Send estimate to the node selected
    3. Broadcast signal to nodes in $\mathcal{C}_j^{s_j}$ implying node s_j has been participated in decision process
    4. Go to sleep mode
  end if
end while

Updating candidate set at k-th node

Denote $\mathcal{C}_j^{s_k}$ to be the candidate set of node s_k at the step j. Algorithm for updating the candidate set at node s_k is explained in Algorithm 2.

Algorithm 2 Updating candidate set at k-th node

NOTATION
s_k: k-th node, s_j: processing node at step j, $\mathcal{C}_j^{s_j}$: candidate set of the processing node s_j at step j, $\mathcal{C}_j^{s_k}$: candidate set of the node s_k at step j
INITIALIZATION
$\mathcal{C}_0^{s_k} = \mathcal{N}_k$
UPDATING
while ($j \geq 1$) **do**
  $\mathcal{C}_j^{s_j} = \mathcal{C}_{j-1}^{s_j}$
  if $s_k = s_j$ (i.e. node s_k becomes the current processing node at step j) **then**
    $\mathcal{C}_j^{s_k} = \mathcal{C}_j^{s_j}$
  else $\{s_k \in \mathcal{C}_j^{s_j}$ (i.e. node s_k belongs to the candidate set of the current processing node at step j)$\}$
    $\mathcal{C}_j^{s_k} = \mathcal{C}_{j-1}^{s_k} \setminus s_j$
  else
    $\mathcal{C}_j^{s_k} = \mathcal{C}_{j-1}^{s_k}$
  end if
end while

Note that, node s_j is not going to be a neighboring node for any node in the network except those that are in $\mathcal{N}_j$. Thus it is not necessary for nodes that are not in $\mathcal{N}_j$ to keep track of node s_j. According to this scheme each node is required to communicate with and keep track of only its neighbors. However, this process will be terminated when the current node does not have any candidate neighboring nodes (i.e. $\mathcal{C}_j^{s_j} = \emptyset$), where $\emptyset$ is the null set, irrespective of whether the desired performance level is reached or not, eventhough there might be remaining nodes in other neighborhoods of the network. However, as observed from simulations, this does not seem to cause a significant performance loss.

3.3. Remark on the optimality of node selection schemes

In both schemes discussed above in subsections 3.1 and 3.2, a global minimum is not guaranteed in general since in both schemes current node selects the next best node from all unvisited nodes in the network (in scheme discussed in 3.1), or in neighborhood (in scheme discussed in 3.2). If the starting node has the information regarding all the nodes in the

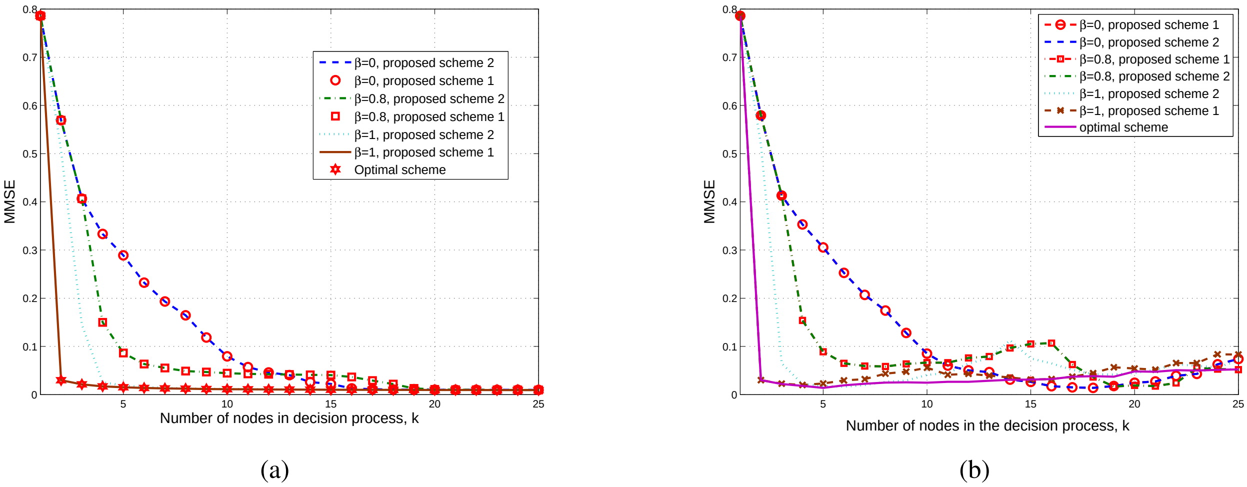


Fig. 3. MMSE at the k-th node with exact target location: (a). No channel noise (b). With channel noise: $\sigma^2_{c(k-1,k)}$'s are as given in (14),$n = 40$, $\sigma_c^2 = 0.001$, $\alpha' = 2$, $r_0 = 1$

network, the optimal node ordering which yields the global minimum over all possible nodes can be computed via dynamic programming by formulating the problem as a shortest path problem as described in [11] with a worst-case complexity of order $O(n^3)$ where n is the number of nodes. When there is a large number of nodes, this computation might be complex. However, we observe (see Section 4) that when there is no channel noise, the scheme discussed in subsection 3.1 coincides with the optimal scheme which yield the global minimum (computed based on dynamic programming) and the scheme proposed in subsection 3.2 is getting close to the optimal scheme after processing relatively a small number of sensors. Even when there is channel noise, we can see that both schemes perform close to the optimal scheme which yields the global minimum.

4. PERFORMANCE ANALYSIS

We consider a 2D square sensor network of area A on $X \times Y$ plane. The locations of the node s_k and the target are denoted by $\mathbf{x}_k = (x_k, y_k)$, for $k = 1, \cdots, n$, and $\mathbf{x}_t = (x_t, y_t)$, respectively. In the following we analyze the performance of a fixed 2D network when the target location is known exactly as well as statistically.

4.1. Exact target location is known at each node

First, we assume that the node s_j has knowledge of its own position, target location and the positions of its neighbors $\mathcal{N}_j$. Then the observation noise variance at the node s_k according to the model (1) can be expressed as,

$$\sigma_k^2 = \left(\frac{r_{kt}}{r_0}\right)^{\alpha} \sigma_0^2, \tag{13}$$

where $r_{kt} = \sqrt{(x_k - x_t)^2 + (y_k - y_t)^2}$ is the distance between the node s_k and the target, α is the path loss index and r_0 and σ_0^2 are constants. The noise variance of the channel between nodes s_{k-1} and s_k is given by,

$$\sigma^2_{c(k-1,k)} = \left(\frac{r_{k-1,k}}{r'_0}\right)^{\alpha'} \sigma_c^2, \tag{14}$$

where $r_{k-1,k} = \sqrt{(x_k - x_{k-1})^2 + (y_k - y_{k-1})^2}$ is the distance between the node s_k and the node s_{k-1}, α' is the path loss index and r'_0 and σ_c^2 are constants.

Node s_j computes the estimator and the MMSE according to (4) and (5). If the desired MMSE threshold is not met, node s_j sends its information to the node s_{j+1}, where the node s_{j+1} is selected from the candidate set $\mathcal{C}_j^{s_j}$ according to (12).

Figure 3 shows the performance of MMSE at the k-th node with no channel noise such that $\sigma^2_{c(k-1,k)} = \sigma_c^2 = 0$ (in Fig. 3 (a)) and with channel noise with $\sigma^2_{c(k-1,k)}$ as given in (14) (in Fig. 3 (b)). We assume that there is a total of 40 sensors deployed in a square region of area 10×10 square units. The target is assumed to be at the origin and the initial node is selected randomly and assumed same for both plots in Fig. 3. Neighbors at each node are selected as the set of nodes located within a disk of radius $r_c = 3$ units. In Figures, we refer scheme 1 and scheme 2 as the schemes presented in subsections 3.1 and 3.2. We refer optimal scheme as the scheme which results the global minimum over all nodes, computed based on dynamic programming algorithm.

With no channel noise, it can be seen that node ordering based on $\beta = 1$ in scheme 1 (based on global search) coincides with the optimal scheme which results the global minimum. In that case, from Fig. 3 (a), it can be seen that the proposed scheme 2 (based on local search) converges to the scheme 1 as with a relatively small number of nodes in the decision process with $\beta = 1$. For $\beta = 0.8$ and $\beta = 0$ (nearest node selection scheme), it is seen that proposed scheme 1 and scheme 2 give similar performance. To achieve a required performance level, for example, to achieve a MMSE of 0.05, scheme 1 with $\beta = 1$ requires 2 nodes, while scheme 2 with $\beta = 1$ requires 4 nodes. Both scheme 1 and scheme 2 require 8 and 12 nodes with $\beta = 0.8$ and $\beta = 0$, respectively to achieve the same performance level.

It is noted that the proposed scheme 2 is terminated at node 25, 32 and 35 with $\beta = 1$, $\beta = 0.8$ and $\beta = 0$, respectively, due to the reason explained in subsection 3.2. However, it is seen that when such a number of nodes are processed node ordering does not affect the overall performance level. This implies that when the sequential estimation process is continued among a large number of sensors, the performance converges to the same value irrespective of how the nodes are selected, which of course is not desirable in many resource constrained sensor networks.

On the other hand, when there is channel noise, it is seen that continuing the sequential processing after some point does not yield improved performance irrespective of which scheme is used for node ordering. This essentially is due to the fact observed in (7). However, in this case, from Fig. 3 (b), it can be seen that the proposed schemes 1 and 2 with $\beta = 1$ gives closer performance to the optimal scheme which yields the global minimum. The results in Fig. 3 are averaged over 10 trials with each based on 5×10^4 Monte-Carlo runs. The standard deviation is very small compared with the means.

From the performance results, we can see that in the proposed sequential estimation process, greedy-type algorithm essentially results a near optimal solution in finding the best ordering of nodes. In the proposed distributed algorithms it was assumed that the parameter σ_θ^2 is available at sensor nodes and the variance of the channel noise of the link between node s_k and s_{k+1}, $\sigma^2_{c(k,k+1)}$, can be computed based on (14) having the knowledge of the distance $r_{k,k+1}$. Figure 4 shows the robustness of the proposed algorithms to errors in the knowledge of these parameters. Let $\hat{\sigma}_\theta^2 = \sigma_\theta^2 + \epsilon_\theta$ be the available value of σ_θ^2 at each node where ϵ_θ is the error term which is assumed to be distributed uniformly in the interval $[-\delta, \delta]$ with $|\delta| \leq \sigma_\theta^2$. The top plot in Fig. 4 shows the robustness of the proposed scheme 2 for error in the knowledge of σ_θ^2 for different δ values assuming no channel noise. It can be seen that the performance of the proposed node selection scheme is robust against errors in the knowledge of σ_θ^2. The bottom plot of Fig. 4 shows the performance difference when the channel noise variance is deviating from its actual value. We obtain the best node ordering according to scheme 2 for $\sigma_c^2 = 0$ and plot the performance curves when σ_c^2 is deviating from 0. It can be seen from the bottom plot of Fig. 4 that when σ_c^2 is more deviating from 0, after a certain number of processing nodes, the performance is degraded. However, it is noted that this happens after achieving the minimum MMSE that can be achieved with $\sigma_c^2 = 0$. Since, processing more nodes after achieving the minimum MMSE that can be achieved with $\sigma_c^2 = 0$, is not desirable, it can be seen that the proposed scheme is also robust for error in the knowledge of channel noise variance as well.

4.2. Statistics of the target location is known at each node

In this section we assume that the the two coordinates of target location x_t and y_t (with $(0,0)$ point is at the center of the square) are distributed as marginal Gaussian with mean zero and the equal variance $\sigma_t^2 = 5$ units. Then it can be verified that $\frac{r_{kt}^2}{\sigma_t^2}$ has a non-central chi-squared distribution with the pdf $f_X(x) = \frac{1}{2} e^{-(x+\lambda_k)/2} I_0(\sqrt{\lambda_k x})$ where $\lambda_k = \frac{x_k^2 + y_k^2}{\sigma_t^2}$

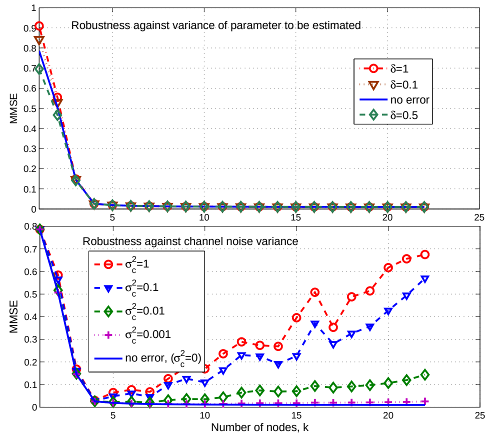


Fig. 4. Robustness of the proposed sequential estimation process for errors in parameters σ_θ^2 and channel noise variance

and $I_a(x)$ is the modified bessel function of the first kind given by $I_a(x) := (x/2)^a \sum_{i=0}^{\infty} \frac{(x^2/4)^i}{i!\Gamma(a+i+1)}$ where $\Gamma(z) = \int_0^\infty t^{z-1} e^{-t} dt$ is the Gamma function. Using (13), the average MMSE at the k-th node is given by,

$$\bar{M}_k = \frac{\sigma_t^2 \sigma_0^2 Q_{k-1}}{r_0^2} \mathbb{E}_x \left\{ \frac{x}{x + Q_{k-1}} \right\} \quad (15)$$

$$= \frac{\sigma_\theta^2 \sigma_0^2 Q_{k-1} e^{-\frac{\lambda_k}{2}}}{2 r_0^2} \left[J_0(\sqrt{\lambda_k Q_{k-1}}) Q_{k-1} e^{\frac{Q_{k-1}}{2}} Ei\left(-\frac{Q_{k-1}}{2}\right) \right. \quad (16)$$

$$\left. + \sum_{i=0}^{\infty} \frac{(\frac{\lambda_k}{4})^i}{i!\Gamma(i+1)} \sum_{l=1}^{i+1} (l-1)!(-Q_{k-1})^{i+1-l} (\frac{1}{2})^{-l} \right], \quad (17)$$

where $Q_{k-1} = \frac{r_0^2 \sigma_\theta^2 B_{k-1}}{\sigma_t^2 \sigma_0^2 (B_{k-1} + \sigma_\theta^2 A_{k-1})}$, $A_{k-1} = (\sigma_\theta^2 - \bar{M}_{k-1})^2$, $B_{k-1} = \sigma_\theta^2 \left[\bar{M}_{k-1}(\sigma_\theta^2 - \bar{M}_{k-1}) + \sigma_\theta^2 \sigma^2_{c(k-1,k)} \right]$ and $J_0(.)$ is the zero-th order Bessel function of first kind.

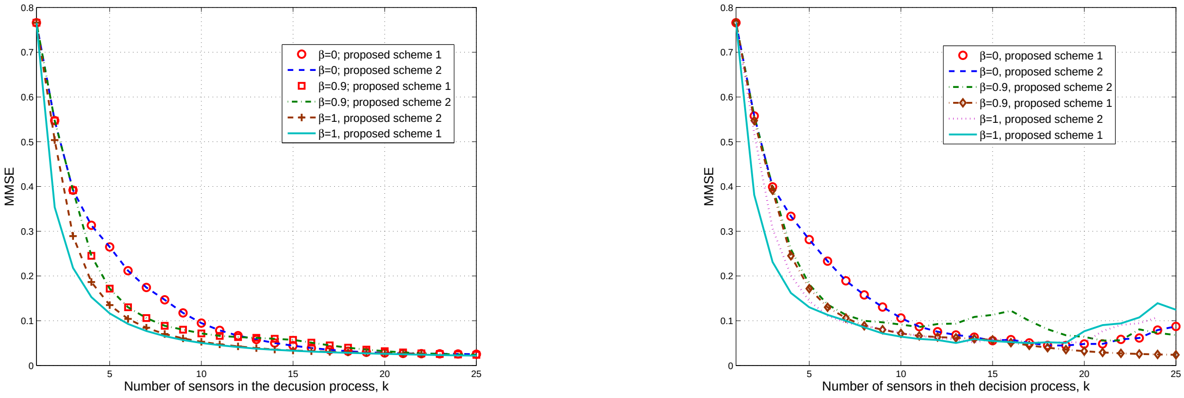
When the current node is s_j, to determine the next node s_{j+1} node s_j has to compute the reward function (11) for its candidate set $\mathcal{C}_j^{s_j}$, where now the information utility function, $\bar{M}_k$ is as given in (17).

Figure 5 shows the MMSE performance of the 2D sensor network in this case where network parameters are the same as that in Fig. 3. It can be seen that the performance of the proposed scheme 2 (local search) is getting closer to the scheme 1 (global scheme) with $\beta = 1$ with perfect as well as noisy inter-node communication. For other values of β considered, the proposed scheme almost coincides with the scheme 1.

It is also noted that (although figures are not included) when σ_t^2 is increasing, that is the uncertainty of the target location is high, the performance of the sequential estimation process does not depend much on the ordering of nodes.

5. CONCLUSIONS

In this paper we have proposed a distributed scheme for sequential estimation of a Gaussian parameter with distributed node selection. Each node makes a local estimate by combining its own observation with the decision from the previous



(a) (b)

Fig. 5. MMSE at the k-th node when the statistics of target location are available: (a). No channel noise (b). With channel noise: $\sigma^2_{c(k-1,k)}$'s are as given in (14), $\sigma_c^2 = 0.005$, $\alpha' = 2$, $r_0 = 1$, $\sigma_0^2 = 0.1$

node. The current node's decision is sent to the next node through a noisy channel. It was shown that such a sequential estimation scheme is useful only if channel noise satisfies a certain threshold condition. The criteria for the next node selection is considered based on an information utility measure and the inter-node communication cost. To perform the distributed sequential estimation process, each node has to keep track of only its neighboring nodes. We derived the MMSE performance for 2-D sensor network models when exact as well as only the statistics of target position information are available at each node. The proposed sequential estimation can be performed distributively having only the information regarding neighbor nodes at each node.

ACKNOWLEDGEMENT

This research was supported in part by the National Science Foundation under the grant CCF-0830545.

6. REFERENCES

[1] Z.-Q. Luo and J.-J. Xiao, "Universal decentralized estimation in a bandwidth constrained sensor network," *IEEE Trans. Inf. Theory*, vol. 51, no. 6, pp. 2210–2219, June 2005.

[2] V. Aravinthan, S. K. Jayaweera, and K. Al Tarazi, "Distributed estimation in a power constrained sensor network," in *IEEE 63rd Vehicular Technology Conference (VTC) 2006 Spring*, Melbourne, Australia, May 2006.

[3] F. Zhao, J. Shin, and J. Reich, "Information-driven dynamic sensor collaboration," *IEEE Signal Processing Magazine*, vol. 19, no. 2, pp. 61–72, Mar. 2002.

[4] 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. 293–313, Fall, 2002.

[5] T. Zhao and A. Nehorai, "Distributed sequential bayesian estimation of a diffusive source in wireless sensor networks," *IEEE Trans. Signal Processing.*, vol. 55, no. 4, pp. 1511–1524, Apr. 2007.

[6] L. Zuo, R. Niu, and P. K. Varshney, "Posterior CRLB based sensor selection for target tracking in sensor networks," in *IEEE Int. Conf. on Acoustics, Speech and Signal Processing (ICASSP)*, Honolulu, Hawaii, USA, April 2007, vol. 2, pp. II–1041–II–1044.

[7] S. K. Jayaweera and C. Mosquera, "Distributed sequential estimation with noisy, correlated observations," *IEEE Signal Processing Letters*, vol. 15, pp. 741–744, 2008.

[8] M. Paskin, C. Guestrin, and J. McFadden, "A robust architecture for distributed inference in sensor networks," in *Proc. Information Processing in Sensor Networks, (IPSN)*, 2005, pp. 55–62.

[9] R. Niu and P. K. Varshney, "Performance analysis of distributed detection in a random sensor field," *IEEE Trans. on Signal Processing*, vol. 56, no. 1, pp. 339 – 349, Jan. 2008.

[10] H. V. Poor, *An Introduction to Signal Detection and Estimation*, Springer-Verlag, New York, 1994.

[11] T. Wimalajeewa and S. K. Jayaweera, "Distributed node selection for sequential estimation over noisy communication channels," *IEEE Trans. Wireless Communications*, 2009, submitted.