Power Constrained Distributed Estimation with Cluster-Based Sensor Collaboration

Jun Fang, *Member, IEEE*, and Hongbin Li, *Senior Member, IEEE*

***Abstract*—We consider the problem of distributed estimation in a power constrained collaborative wireless sensor network (WSN), where the network is divided into a set of sensor clusters, with collaboration allowed among sensors within the same cluster but not across clusters. Specifically, each cluster forms one or multiple local messages via sensor collaboration (in particular, linear operation is considered) and transmits the messages over noisy channels to a fusion center (FC). The final estimate is constructed at the FC based on the noisy data received from all clusters. In this collaborative setup, we study the following fundamental problems. Given a total transmit power constraint, shall we transmit the raw data or some low-dimensional local messages for each cluster? What is the optimal collaboration scheme for each cluster? How to optimally allocate the power among different clusters? These questions are addressed in this paper. We will show that the optimum collaboration strategy is to compress the data into one local message which, depending on the channel characteristics, is transmitted using one or multiple available channels to the FC. The optimal power allocation among the clusters is also investigated, which yields a water-filling type of scheme.**

***Index Terms*—Distributed estimation, sensor collaboration, power allocation, wireless sensor networks (WSNs).**

I. Introduction

Wireless sensor networks (WSNs) have been of significant interest over the past few years due to their potential applications in environment monitoring, battlefield surveillance, target localization and tracking, and many more [1], [2]. The sensors constructing the network are often powered by small batteries that are often irreplaceable. Hence limited energy resource is a key challenge one has to overcome before such applications become practical. In a sensor network, communication consumes a significant portion of the total energy as compared with the sensing and computation related energy cost. It is therefore important to develop transmission energy-efficient strategies for various sensor network processing tasks such as estimation, detection and tracking. Specifically, in this paper, we consider distributed parameter estimation, which is a fundamental problem in sensor network research.

Distributed estimation has attracted much attention recently. One of the network architectures for distributed estimation involves a set of spatially distributed sensors linked with a fusion center (FC). Each sensor makes a noisy observation of the phenomena of interest and transmits its processed information to the FC, where a final estimate is formed. Many works (e.g. [3]–[10]) were carried out in this setup. Among them, some [3]–[6] addressed the power constraint issue by resorting to aggressive quantization schemes which quantize the original observations into one or a few bits of information. In this case, quantization becomes an integral part of the estimation process and is critical to the estimation performance. Other works [7]–[10] studied the problem of optimal power allocation among sensors given a total transmit power constraint, aiming at minimizing the estimation distortion at the FC. For all of these works except [6], the inter-sensor communication is not considered. Inter-sensor collaboration can instead be exploited to enhance the transmission energy efficiency and improve the system performance. For example, [11] introduced a passive sensor cooperation scheme (sequential transmission) which achieves an estimation accuracy comparable to those schemes without sensor cooperation at a reduced energy cost. Also, in [12]–[16], a class of fully distributed schemes (no FC is required) were proposed, where sensors , through local data exchange (i.e., collaboration) and computation, can eventually reach a consensus estimate. Since the algorithms [12]–[16] involve communication only among neighboring sensors, the energy consumption can be considerably reduced. A problem associated with these schemes lies in that they are sensitive to communication errors. It was shown [17] that in the presence of link noise, the estimate diverges and has an asymptotic unbounded mean-square error.

In this paper, we consider distributed estimation in a hierarchical network architecture with localized collaboration. Specifically, we assume that the network is divided into a number of sensor clusters linked with a FC. The sensors within the same cluster have the communication resources to locally collaborate, whereas no collaboration is allowed across clusters. This might be the case for scenarios where multiple sets of sensors are spatially distributed, with each set of sensors within a small neighborhood. In other situations, a sensor node may consist of a sensor-array with multiple modalities (acoustic, seismic, infrared, etc.) [18]. This scenario can also be captured by the above model with the individual components of the sensor-array corresponding to the collaborating sensors in a cluster. Each cluster then transmits one or multiple one-dimensional messages, which could be the raw data or obtained via sensor collaboration, over noisy channels to the FC, where a final estimate is formed based on the data received from all clusters. In this context, the following natural questions arise: given a fixed amount of

Manuscript received November 5, 2008; revised March 1, 2009; accepted March 29, 2009. The associate editor coordinating the review of this paper and approving it for publication was E. Hossain.

J. Fang and H. Li are with the Department of Electrical and Computer Engineering, Stevens Institute of Technology, Hoboken, NJ 07030, USA (e-mail: {Jun.Fang, Hongbin.Li}@stevens.edu).

This work was supported in part by the National Science Foundation under Grant CCF-0514938.

Digital Object Identifier 10.1109/TWC.2009.081438

total transmit power, how should each cluster process their local measurements such that a minimum estimation distortion can be achieved at the FC? how should we allocate the power among different clusters in an optimal power-distortion fashion? These questions will be addressed in this paper and we will develop a fundamental understanding of the above important hierarchical collaborative strategy for distributed estimation. We notice that this hierarchical collaborative strategy was also studied in [18], where the authors examined the encoding of the local measurements under a communication rate constraint. A similar hierarchical collaborative strategy was recently proposed in [19] (referred to as branch and tree network topologies), where the information of the sensors is processed/forwarded to the FC by multiple relay nodes. Nevertheless, the authors focused on the power allocation problem with some pre-determined collaboration schemes (i.e., they are not part of the optimization problem) at the relay nodes. Our work is also closely related to the distributed compression-estimation approaches [20]–[25] whose objective is to reduce the transmission requirement via dimensionality reduction. While most of these works assume ideal communication links (between sensors and the FC) and no transmit power constraint is imposed, our work takes into account the transmit power constraint and the link noise (between clusters and the FC) in formulating our problem.

The rest of the paper is organized as follows. In Section II, we introduce the sensor network collaboration model, some basic assumptions, and our objective in the collaborative setting. Next, in Section III, we investigate the optimal collaboration for one sensor cluster case. A simple performance analysis and some numerical results are also included. The optimal collaboration and power allocation problem for multiple sensor clusters is studied in Section IV, followed by concluding remarks in Section V.

II. System Model and Problem Formulation

We consider a WSN consisting of N spatially distributed sensors, with each sensor making a noisy observation of an unknown random parameter θ: $x_n = h_n\theta + w_n$, where h_n denotes the observation gain and w_n denotes the additive observation noise. The sensors in the network are divided into M sensor clusters (see Fig. 1). Each cluster, say cluster m, consists of N_m geographically closely located sensors. The sensors in each cluster are able to collaborate to form local messages which are sent to the FC, whereas no communication is allowed across different clusters. The objective is to obtain an estimate of the unknown parameter at the FC based on the information received from the clusters. In practice, the sensor collaboration can be easily implemented. For each cluster, we choose one sensor to be the cluster head whose task is to collect the data from other sensors within the same cluster and carry out the collaborative processing. The resultant local messages are then transmitted by the cluster head to the FC. We adopt the following assumptions for this collaborative setting:

A1 The links between sensors and the cluster head within each cluster are ideal. Sensor collaboration is confined to be linear operations.

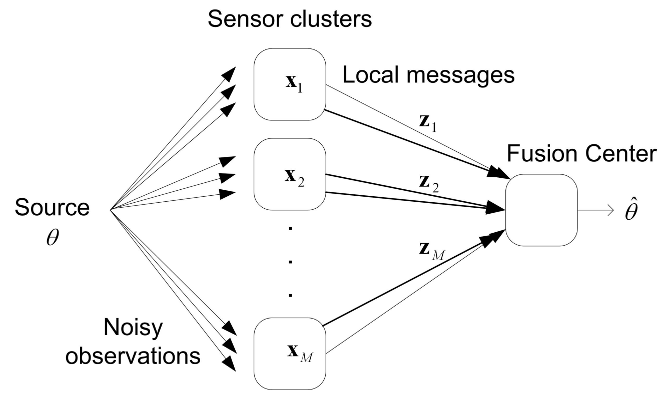


Fig. 1. Collaborative setting: the network is divided into a number of sensor clusters. Sensors within each cluster can collaborate to convert their noisy observations $\{\mathbf{x}_m\}$ into some local estimates $\{\mathbf{z}_m\}$.

A2 An uncoded analog amplify-and-forward scheme is employed to transmit the local messages from the cluster heads to the FC over noisy wireless channels.

Remarks: Assumption (A1) can be justified in the sense that sensors within a same cluster are closely located and their communication links can be made reliable with negligible communication errors at an affordable energy cost. This assumption was also adopted in [18]. In contrast, in many practical scenarios, a centralized FC can be located far away from the deployed sensor field. Hence assuming an ideal wireless link between the sensor node and the FC may not be practical (note that received power of a radio signal decays exponentially with the distance). We confine our study on linear collaborative processing due to its simplicity in implementation and development. The uncoded analog transmission scheme has been widely adopted in WSNs, e.g. [8], [19], [24]. Its optimality and advantage over the separate source-channel coding have been proved in a point-to-point transmission case [26] and in scenarios where sensors adopt a synchronized (non-orthogonal) interference multiple access channel to the FC [27]. The analog transmission scheme is, however, found inferior to the optimal separate source-channel coding strategy for orthogonal/independent multiple access channels (from sensors to the FC) [28]. Nevertheless, an optimal, practical separate source-channel coding strategy is usually difficult to obtain. Therefore the analog transmission scheme is still a reasonable choice in designing wireless sensor network for various applications.

For notational convenience, we use $x_{m,n}$ to denote the sensor measurement of sensor n in cluster m, where $n \in \{1,\dots,N_m\}$, $m \in \{1,\dots,M\}$ and

$$x_{m,n} = h_{m,n}\theta + w_{m,n} \tag{1}$$

in which $h_{m,n}$ and $w_{m,n}$ denote the corresponding observation gain and additive observation noise, respectively. To capture the cluster-based collaborative scenario, we write the measurements within a cluster in a vector form: $\mathbf{x}_m \triangleq [x_{m,1}\ x_{m,2}\ \dots\ x_{m,N_m}]^T$, which is given by

$$\mathbf{x}_m = \mathbf{h}_m\theta + \mathbf{w}_m \tag{2}$$

with $\mathbf{h}_m \triangleq [h_{m,1}\ h_{m,2}\ \dots\ h_{m,N_m}]^T$ and $\mathbf{w}_m \triangleq [w_{m,1}\ w_{m,2}\ \dots\ w_{m,N_m}]^T$. The local messages via sensor collaboration within each cluster can therefore be expressed as

$$\mathbf{z}_m = \mathbf{C}_m \mathbf{x}_m \quad (3)$$

where $\mathbf{C}_m \in \mathbb{R}^{p_m \times N_m}$ denotes the collaboration matrix for cluster m, $p_m \leq N_m$ is the dimensionality of the message vector $\mathbf{z}_m$ whose choice will become evident later. The signal at the FC received from the mth cluster is given by

$$\mathbf{y}_m = \mathbf{G}_m \mathbf{A}_m \mathbf{C}_m \mathbf{x}_m + \mathbf{v}_m \quad (4)$$

where $\mathbf{G}_m \in \mathbb{R}^{p_m \times p_m}$ denotes a fading multiplicative channel matrix, which can be diagonal or non-diagonal, depending on the transmission scheme (e.g., orthogonal vs. non-orthogonal channel access); $\mathbf{A}_m \triangleq \text{diag}\{a_1, \ldots, a_{p_m}\}$ is an amplification matrix with a_i denoting the amplification factor used in transmitting the ith message of $\mathbf{z}_m$; $\mathbf{v}_m \in \mathbb{R}^{p_m}$ denotes the additive channel noise vector. Without loss of generality, we assume $\mathbf{G}_m = \mathbf{I}$ and $\mathbf{A}_m = \mathbf{I}$, where $\mathbf{I}$ denotes the identity matrix, as the multiplicative effect of the channel matrix can be removed by carrying out a matrix inverse using an estimate of the channel matrix $\mathbf{G}_m$ at the receiver [21], and the amplification matrix $\mathbf{A}_m$ can be absorbed into $\mathbf{C}_m$. We have the following assumption regarding the observation noise $\{\mathbf{w}_m\}$ and the channel noise $\{\mathbf{v}_m\}$.

A3 The noise $\{\mathbf{w}_m\}$ and $\{\mathbf{v}_m\}$ are zero mean with positive definite auto-covariance $\{\mathbf{R}_{w,m}\}$ and $\{\mathbf{R}_{v,m}\}$, respectively, which are available at the FC. The noise across different clusters are mutually uncorrelated, i.e. $E[\mathbf{w}_i \mathbf{w}_j^T] = \mathbf{0} \ \forall i \neq j$ and $E[\mathbf{v}_i \mathbf{v}_j^T] = \mathbf{0} \ \forall i \neq j$.

Remarks: It is reasonable to assume independence across different clusters since these clusters are usually geographically sufficiently apart. On the other hand, within a cluster, the noise, $\mathbf{w}_m$ and $\mathbf{v}_m$, are allowed to have arbitrary spatial correlation. This assumption is not trivial as the observation noise $\mathbf{w}_m$ could be correlated when sensors are densely deployed, due to the fact that the phenomena to be measured is subject to similar disturbance with a high probability. Also, correlation is introduced to the additive channel noise $\mathbf{v}_m$ when the inverse of a non-diagonal channel matrix $\mathbf{G}_m$ is carried out to compensate the channel effect [21]. The priori knowledge of the auto-correlation matrices, $\{\mathbf{R}_{w,m}\}$ and $\{\mathbf{R}_{v,m}\}$, may come either from specific data models or from sample estimation after a training phase [8], [19], [24].

Let $\mathbf{y} \triangleq [\mathbf{y}_1 \ \mathbf{y}_2 \ \ldots \ \mathbf{y}_M]^T$ denote a column vector formed by stacking the data received from all clusters. We have

$$\begin{aligned} \mathbf{y} =& \mathbf{C}\mathbf{x} + \mathbf{v} \\ =& \mathbf{C}(\mathbf{h}\theta + \mathbf{w}) + \mathbf{v} \end{aligned} \quad (5)$$

where $\mathbf{C} \triangleq \text{diag}\{\mathbf{C}_1, \ldots, \mathbf{C}_M\}$ is a block diagonal matrix with its mth block-diagonal element equal to $\mathbf{C}_m$, $\mathbf{x} \triangleq [\mathbf{x}_1 \ \mathbf{x}_2 \ \ldots \ \mathbf{x}_M]^T$, $\mathbf{v} \triangleq [\mathbf{v}_1 \ \mathbf{v}_2 \ \ldots \ \mathbf{v}_M]^T$, $\mathbf{h} \triangleq [\mathbf{h}_1 \ \mathbf{h}_2 \ \ldots \ \mathbf{h}_M]^T$, and $\mathbf{w} \triangleq [\mathbf{w}_1 \ \mathbf{w}_2 \ \ldots \ \mathbf{w}_M]^T$. A natural question arising from the above scenario is to find out an overall optimal collaboration matrix $\mathbf{C}$, or equivalently, a set of individual collaboration matrices $\{\mathbf{C}_m\}_{m=1}^M$, to achieve a minimum estimation distortion at the FC. Also, since the amplification factors $\{\mathbf{A}_m\}$ are incorporated into the collaboration matrices $\{\mathbf{C}_m\}$, the overall collaboration matrix $\mathbf{C}$ has to satisfy a total transmit power constraint. Specifically, using a linear minimum mean-square error (LMMSE) estimator [29], it can be readily verified that we are faced with the following optimization problem

$$\begin{aligned} \min_{\mathbf{C}} \quad & E[(\theta - \hat{\theta})^2] = \sigma_\theta^2 - \sigma_\theta^4 \mathbf{h}^T \mathbf{C}^T (\mathbf{C}\mathbf{R}_x \mathbf{C}^T + \mathbf{R}_v)^{-1} \mathbf{C}\mathbf{h} \\ \text{s.t.} \quad & \text{tr}\left(\mathbf{C}\mathbf{R}_x\mathbf{C}^T\right) \leq P \end{aligned} \quad (6)$$

where σ_θ^2 denotes the signal variance, $\mathbf{R}_x \triangleq E[\mathbf{x}\mathbf{x}^T]$, $\text{tr}\left(\mathbf{C}\mathbf{R}_x\mathbf{C}^T\right)$ is the average transmit power required to send the local messages from all clusters to the FC, P is a pre-specified power budget for transmission. We note that a similar problem was examined in [24] in the context of distributed compression-estimation. However, the work, studying the problem in a framework of a canonical correlation analysis (CCA), has a formulation and an approach different from ours. Also, instead of using a sum power constraint as we did here, [24] imposed power constraints on individual sensors (corresponding to clusters in this paper). This requires assigning power to the clusters *a priori*, which can be a tricky problem for inhomogeneous environments, where clusters at different locations have dissimilar observation and link qualities, and an equal power allocation scheme could be far away from optimum. In contrast, our formulation implicitly involves automatic power allocation among different clusters. More importantly, the proposed solution of [24] is iterative and sub-optimal, while as will be shown later, an optimum solution can be obtained for our case.

III. Optimal Collaboration and Power Allocation: Single Cluster Case

In the section, we will examine the optimization problem (6) for the case of one cluster, i.e. $M = 1$. A general scenario with multiple clusters will be studied later by using the theoretical results developed for the single cluster case.

A. Proposed Approach

When $M = 1$, the collaboration matrix $\mathbf{C} \in \mathbb{R}^{p \times N}$, where $p \leq N$, can be any arbitrary matrix as long as it satisfies the transmit power constraint. To solve (6), we firstly carry out the following simplifications. Let $\mathbf{R}_v = \mathbf{U}_v \mathbf{D}_v \mathbf{U}_v^T$ denote the eigenvalue decomposition (EVD) and $\dot{\mathbf{C}} \triangleq \mathbf{U}_v^T \mathbf{C} \mathbf{R}_x^{\frac{1}{2}}$, where $\mathbf{D}_v \in \mathbb{R}^{p \times p}$ is a positive definite diagonal matrix. We have

$$\begin{aligned} &\mathbf{h}^T \mathbf{C}^T \left(\mathbf{C}\mathbf{R}_x\mathbf{C}^T + \mathbf{R}_v\right)^{-1} \mathbf{C}\mathbf{h} \\ =&\mathbf{h}^T \mathbf{R}_x^{-\frac{1}{2}} \dot{\mathbf{C}}^T (\dot{\mathbf{C}}\dot{\mathbf{C}}^T + \mathbf{D}_v)^{-1} \dot{\mathbf{C}} \mathbf{R}_x^{-\frac{1}{2}} \mathbf{h} \\ \stackrel{(a)}{=}&\text{tr}\left((\dot{\mathbf{C}}\dot{\mathbf{C}}^T + \mathbf{D}_v)^{-1} \dot{\mathbf{C}}\mathbf{G}\dot{\mathbf{C}}^T\right) \\ \stackrel{(b)}{=}&\text{tr}\left((\tilde{\mathbf{C}}\tilde{\mathbf{C}}^T + \mathbf{D}_v)^{-1} \tilde{\mathbf{C}}\mathbf{D}_g\tilde{\mathbf{C}}^T\right) \\ =&\text{tr}\left(\mathbf{D}_g^{\frac{1}{2}} \tilde{\mathbf{C}}^T (\tilde{\mathbf{C}}\tilde{\mathbf{C}}^T + \mathbf{D}_v)^{-1} \tilde{\mathbf{C}} \mathbf{D}_g^{\frac{1}{2}}\right) \end{aligned} \quad (7)$$

where (a) comes from the trace identity $\text{tr}(\mathbf{AB}) = \text{tr}(\mathbf{BA})$, in which $\mathbf{G} \triangleq \mathbf{R}_x^{-\frac{1}{2}} \mathbf{h}\mathbf{h}^T \mathbf{R}_x^{-\frac{1}{2}}$, and (b) follows by defining $\tilde{\mathbf{C}} \triangleq \dot{\mathbf{C}}\mathbf{U}_g$, in which $\tilde{\mathbf{C}} \in \mathbb{R}^{p \times N}$, $\mathbf{U}_g \in \mathbb{R}^{N \times N}$ is an orthonormal matrix and $\mathbf{D}_g \in \mathbb{R}^{N \times N}$ is a positive semi-definite diagonal matrix obtained by carrying out the EVD of $\mathbf{G}$, i.e. $\mathbf{G} = \mathbf{U}_g \mathbf{D}_g \mathbf{U}_g^T$. Note that $\mathbf{D}_g$ has only one nonzero diagonal element because $\mathbf{G}$ is rank-one. It is easy to verify that the relationship between $\mathbf{C}$ and $\tilde{\mathbf{C}}$ is given by:

$$\mathbf{C} = \mathbf{U}_v \tilde{\mathbf{C}} \mathbf{U}_g^T \mathbf{R}_x^{-\frac{1}{2}} \tag{8}$$

Consequently the power constraint becomes

$$\begin{aligned} \text{tr}\left(\mathbf{C}\mathbf{R}_x\mathbf{C}^T\right) =& \text{tr}\left(\mathbf{U}_v\tilde{\mathbf{C}}\mathbf{U}_g^T\mathbf{R}_x^{-\frac{1}{2}}\mathbf{R}_x\mathbf{R}_x^{-\frac{1}{2}}\mathbf{U}_g\tilde{\mathbf{C}}^T\mathbf{U}_v^T\right) \\ =& \text{tr}\left(\mathbf{U}_v\tilde{\mathbf{C}}\tilde{\mathbf{C}}^T\mathbf{U}_v^T\right) \\ =& \text{tr}\left(\tilde{\mathbf{C}}\tilde{\mathbf{C}}^T\right) \leq P \end{aligned} \tag{9}$$

Combining (6)–(9), our problem, therefore, is reformulated as

$$\begin{aligned} \max_{\tilde{\mathbf{C}}} \quad & \text{tr}\left(\mathbf{D}_g^{\frac{1}{2}}\tilde{\mathbf{C}}^T(\tilde{\mathbf{C}}\tilde{\mathbf{C}}^T + \mathbf{D}_v)^{-1}\tilde{\mathbf{C}}\mathbf{D}_g^{\frac{1}{2}}\right) \\ \text{s.t.} \quad & \text{tr}\left(\tilde{\mathbf{C}}\tilde{\mathbf{C}}^T\right) \leq P \end{aligned} \tag{10}$$

This formulation (10), however, is still hard to deal with due to the inverse of a variable matrix. To overcome this difficulty, we define $\mathbf{T} \triangleq \tilde{\mathbf{C}}^T\mathbf{D}_v^{-1}\tilde{\mathbf{C}}$ and let $\mathbf{T} = \mathbf{U}\mathbf{D}\mathbf{U}^T$ denote the reduced EVD, where $\mathbf{U} \in \mathbb{R}^{N\times p}$, $\mathbf{D} \in \mathbb{R}^{p\times p}$. Since $\tilde{\mathbf{C}}^T\mathbf{D}_v^{-1}\tilde{\mathbf{C}} = \mathbf{U}\mathbf{D}\mathbf{U}^T$, we have

$$\tilde{\mathbf{C}}^T\mathbf{D}_v^{-\frac{1}{2}}\mathbf{Q} = \mathbf{U}\mathbf{D}^{\frac{1}{2}} \Rightarrow \tilde{\mathbf{C}}^T = \mathbf{U}\mathbf{D}^{\frac{1}{2}}\mathbf{Q}^T\mathbf{D}_v^{\frac{1}{2}} \tag{11}$$

where $\mathbf{Q} \in \mathbb{R}^{p\times p}$ can be any orthonormal matrix as we can write $\mathbf{U}\mathbf{D}\mathbf{U}^T = \mathbf{U}\mathbf{D}^{\frac{1}{2}}(\mathbf{U}\mathbf{D}^{\frac{1}{2}})^T = \tilde{\mathbf{C}}^T\mathbf{D}_v^{-\frac{1}{2}}\mathbf{Q}\mathbf{Q}^T\mathbf{D}_v^{-\frac{1}{2}}\tilde{\mathbf{C}}^T = \tilde{\mathbf{C}}^T\mathbf{D}_v^{-1}\tilde{\mathbf{C}}$, and the diagonal elements of $\mathbf{D}$ must be non-negative because $\tilde{\mathbf{C}}^T\mathbf{D}_v^{-1}\tilde{\mathbf{C}}$ is positive semi-definite. The motivation to do so is justified by the following relationship (easily established by utilizing (11))

$$\tilde{\mathbf{C}}^T(\tilde{\mathbf{C}}\tilde{\mathbf{C}}^T + \mathbf{D}_v)^{-1}\tilde{\mathbf{C}} = \mathbf{U}\mathbf{D}^{\frac{1}{2}}(\mathbf{I} + \mathbf{D})^{-1}\mathbf{D}^{\frac{1}{2}}\mathbf{U}^T \tag{12}$$

which shows that the non-tractable matrix inverse is reduced to an inverse of a diagonal matrix (albeit still variable) that is much easier for analysis and computation. From (11), we see that $\tilde{\mathbf{C}}$ is determined by $\mathbf{U}$, $\mathbf{Q}$ and a diagonal matrix $\mathbf{D}$. Our objective therefore is to find the optimal solutions of $\{\mathbf{U}, \mathbf{Q}, \mathbf{D}\}$. Based on the previous discussions and combining (10)–(12), we eventually reach the following optimization problem

$$\begin{aligned} \max_{\{\mathbf{U},\mathbf{D},\mathbf{Q}\}} \quad & \text{tr}\left(\mathbf{D}_g^{\frac{1}{2}}\mathbf{U}\mathbf{D}^{\frac{1}{2}}(\mathbf{I} + \mathbf{D})^{-1}\mathbf{D}^{\frac{1}{2}}\mathbf{U}^T\mathbf{D}_g^{\frac{1}{2}}\right) \\ \text{s.t.} \quad & \text{tr}\left(\mathbf{D}\mathbf{Q}^T\mathbf{D}_v\mathbf{Q}\right) \leq P \\ & \mathbf{U}^T\mathbf{U} = \mathbf{I} \\ & \mathbf{D} = \text{diag}(d_1, \ldots, d_p) \quad d_i \geq 0 \quad \forall i \\ & \mathbf{Q}\mathbf{Q}^T = \mathbf{I} \end{aligned} \tag{13}$$

Although the problem (13) involves searching for multiple optimization variable matrices, a close examination shows that it can be decoupled into two sequential sub-problems. We can, firstly, find an optimal $\mathbf{U}$ by fixing the variables $\{\mathbf{Q}, \mathbf{D}\}$ (in fact, as we will show in our proof, the optimal $\mathbf{U}$ is independent of the optimization variables $\{\mathbf{Q}, \mathbf{D}\}$). Then using the previous results, we search for the optimal matrices $\{\mathbf{Q}, \mathbf{D}\}$. Both optimization problems can be solved by exploiting the diagonal structures and related matrix properties. We summarize our results as follows.

Lemma 1: Without loss of generality, we assume that the diagonal elements of $\mathbf{D}_g$ are sorted in a descending order; whereas the diagonal elements of $\mathbf{D}_v$ are in an ascending order. The optimal solution to (13) is then given by $\mathbf{U}^* = \mathbf{\Gamma}\mathbf{P}^T$, $\mathbf{Q}^* = \mathbf{P}^T$, and $\mathbf{D}^* = \mathbf{P}\mathbf{\Lambda}\mathbf{P}^T$, where $\mathbf{P} \in \mathbb{R}^{p\times p}$ can be any permutation matrix, $\mathbf{\Gamma} \in \mathbb{R}^{p\times N}$ and $\mathbf{\Lambda} \in \mathbb{R}^{p\times p}$ are respectively given by

$$\mathbf{\Gamma} \triangleq \begin{bmatrix} 1 & \mathbf{0}_{1\times(p-1)} \\ \mathbf{0}_{(N-1)\times 1} & \mathbf{E} \end{bmatrix} \tag{14}$$

$$\mathbf{\Lambda} \triangleq \text{diag}\left(\frac{P}{\min(\mathbf{d}_v)}, 0, \ldots, 0\right) \tag{15}$$

in which $\mathbf{E} \in \mathbb{R}^{(N-1)\times(p-1)}$ can be any matrix satisfying $\mathbf{E}^T\mathbf{E} = \mathbf{I}$, $\mathbf{d}_v \triangleq \text{diag}(\mathbf{D}_v)$, and $\min(\mathbf{d}_v) = \mathbf{D}_v(1,1)$ since the diagonal elements of $\mathbf{D}_v$ are in an ascending order, where $\mathbf{D}_v(1,1)$ denotes the $(1,1)$th entry of $\mathbf{D}_v$. The maximum value of the objective function achieved by the optimal matrices is

$$f(\mathbf{U}^*, \mathbf{D}^*, \mathbf{Q}^*) = \frac{P}{P + \min(\mathbf{d}_v)}\mathbf{h}^T\mathbf{R}_x^{-1}\mathbf{h} \tag{16}$$

where $f(\cdot)$ denotes the objective function of (13). ■

Proof: See Appendix A.

By utilizing Lemma 1, we can trace back to find out the optimal collaboration matrix and the associated estimation mean-square error (MSE). We have the following result.

Theorem 1: Consider the optimal collaboration design problem formulated in (6) and described in Fig. 1, where the sensor measurements $\mathbf{x}_m$, the local messages $\mathbf{z}_m$, and the received messages at the FC $\mathbf{y}_m$ are given by (2), (3) and (4), respectively. When $M = 1$, the optimal solution to (6) is

$$\mathbf{C}^* = \gamma\sqrt{P}\mathbf{U}_v[:,1]\mathbf{h}^T\mathbf{R}_x^{-1} \tag{17}$$

where $\mathbf{U}_v[:,1]$ denotes the first column of $\mathbf{U}_v$, and $\gamma \triangleq \frac{1}{\sqrt{\mathbf{h}^T\mathbf{R}_x^{-1}\mathbf{h}}}$. The associated estimation MSE, i.e. the value of the minimum objective function of (6), is given by

$$E[(\theta - \hat{\theta}(\mathbf{C}^*))^2] = \sigma_\theta^2 - \sigma_\theta^4\frac{P}{P + \min(\mathbf{d}_v)}\mathbf{h}^T\mathbf{R}_x^{-1}\mathbf{h} \tag{18}$$

■

Proof: See Appendix B.

The optimal solution (17) has very important implications which we shall explore in the following. Considering the scenario of independent or uncorrelated channels, i.e. $\mathbf{R}_v$ is diagonal, we have $\mathbf{U}_v = \mathbf{I}$ and $\mathbf{U}_v[:,1] = \mathbf{e}_1$, where $\mathbf{e}_i$ denotes the unit column vector with its ith entry equal to one, and its other entries equal to zero. Therefore the optimal collaboration matrix becomes

$$\mathbf{C}^* = \begin{bmatrix} \gamma\sqrt{P}\mathbf{h}^T\mathbf{R}_x^{-1} \\ \mathbf{0}_{(p-1)\times N} \end{bmatrix} \tag{19}$$

which is a matrix with its first row equal to $\gamma\sqrt{P}\mathbf{h}^T\mathbf{R}_x^{-1}$ and all other rows equal to zero. The solution suggests that we should compress the measurements into only one local message and transmit it via the best-quality channel (note that the first row corresponds to the first channel which has the smallest noise variance since the diagonal elements of $\mathbf{R}_v$ are assumed in an ascending order) to the FC. If the channels have identical qualities, then we can use any of them to send out the local message. Also, by rewriting the collaboration weighting vector $\gamma\sqrt{P}\mathbf{h}^T\mathbf{R}_x^{-1}$ as

$$\begin{aligned}\gamma\sqrt{P}\mathbf{h}^T\mathbf{R}_x^{-1} &= \gamma\sqrt{P}\sigma_\theta^{-2}\sigma_\theta^2\mathbf{h}^T\mathbf{R}_x^{-1} \\ &= \gamma\sqrt{P}\sigma_\theta^{-2}\mathbf{R}_{\theta x}\mathbf{R}_x^{-1}\end{aligned} \tag{20}$$

where $\mathbf{R}_{\theta x} \triangleq E[\theta\mathbf{x}^T]$, we can immediately see that the local message is exactly the LMMSE estimate $\mathbf{R}_{\theta x}\mathbf{R}_x^{-1}\mathbf{x}$ multiplied by a scalar $\gamma\sqrt{P}\sigma_\theta^{-2}$. This means that when channels are independent, LMMSE estimation followed by an amplification factor is optimal in a power-distortion sense. Interestingly, as a counterpart, we noticed that in [30], [31], it is shown that in a rate-constrained scenario, when there is only one sensor cluster, minimum mean-square error (MMSE) estimation followed by vector-quantization is optimal in a rate distortion sense.

We now investigate the case where the channels are correlated, i.e. $\mathbf{R}_v$ is non-diagonal. Each row of the optimal collaboration matrix can be readily expressed as follows by combining (17) and (20)

$$\mathbf{C}^*[i,:] = \mathbf{U}_v[i,1]\gamma\sqrt{P}\sigma_\theta^{-2}\mathbf{R}_{\theta x}\mathbf{R}_x^{-1} \tag{21}$$

where $\mathbf{U}_v[i,1]$ denotes the $(i,1)$th entry of $\mathbf{U}_v$. Therefore the LMMSE estimate is transmitted by multiple channels with different amplification gains that are proportional to $\{\mathbf{U}_v[i,1]\}_{i=1}^p$. We see that for the correlated case, unlike the independent case assigning all power on one channel, power should be distributed to all available channels. This is even true for strongly correlated channels because the correlation across channels can be exploited to cancel the channel noise and hence to enhance the estimation performance.

B. Performance Analysis and Simulation Results

We carry out a simple performance analysis to corroborate our theoretical results. We compare our optimal collaboration strategy with the scheme proposed in [8], where there is no inter-sensor collaboration and each sensor transmits its observation to the FC with optimally assigned power. For simplicity, we consider a homogeneous environment with identical observation and channel qualities, where σ_w^2 denotes the observation noise variance and σ_v^2 represents the channel noise variance. The observation and channel noise are assumed to be independent across sensors and links, respectively. All observation and channel gains are unit throughout all examples in the paper. Clearly, an equal power allocation is optimum for [8] and the corresponding estimation MSE can be shown to be

$$\text{MSE}_{\text{NC}} = \frac{P\sigma_w^2\sigma_\theta^2 + N\sigma_v^2\sigma_\theta^4 + N\sigma_v^2\sigma_\theta^2\sigma_w^2}{PN\sigma_\theta^2 + P\sigma_w^2 + N\sigma_v^2\sigma_\theta^2 + N\sigma_v^2\sigma_w^2} \tag{22}$$

where the subscript NC denotes non-collaboration. For our collaboration strategy, the estimation MSE can be computed by using (18), which reduces to

$$\text{MSE}_{\text{OC}} = \frac{P\sigma_w^2\sigma_\theta^2 + N\sigma_v^2\sigma_\theta^4 + \sigma_v^2\sigma_\theta^2\sigma_w^2}{PN\sigma_\theta^2 + P\sigma_w^2 + N\sigma_v^2\sigma_\theta^2 + \sigma_v^2\sigma_w^2} \tag{23}$$

where the subscript OC denotes optimal collaboration. For notational convenience, let $a \triangleq P\sigma_w^2\sigma_\theta^2 + N\sigma_v^2\sigma_\theta^4$ and $b \triangleq PN\sigma_\theta^2 + P\sigma_w^2 + N\sigma_v^2\sigma_\theta^2$. It can be easily verified that

$$(a + N\sigma_v^2\sigma_\theta^2\sigma_w^2)(b + \sigma_v^2\sigma_w^2) \geq (a + \sigma_v^2\sigma_\theta^2\sigma_w^2)(b + N\sigma_v^2\sigma_w^2) \tag{24}$$

where (24) becomes an equality only when $N = 1$. Hence as expected, the following relationship $\text{MSE}_{\text{NC}} \geq \text{MSE}_{\text{OC}}$ holds, which means that the optimal collaboration scheme should always outperform the non-collaboration scheme.

Fig. 2 depicts the estimation MSEs of the two schemes as a function of N under a total transmit power constraint, with $\sigma_w^2 = 0.2$ and $\sigma_w^2 = 1$, respectively. From Fig. 2, we see that both schemes benefit from an increasing number of sensors; as N increases, the estimation MSEs will asymptotically approach certain values that, however, are nonzero. Also, it can be seen that the non-collaborative scheme is sensitive to the value of σ_w^2; as the observation quality deteriorates, its performance degrades considerably. In contrast, the collaborative strategy demonstrates a certain degree of robustness against the observation quality deterioration. In Fig. 3, we plot the estimation MSE vs. the total transmit power. We see that the performance gap between the two strategies shrinks as the transmit power increases. In fact, from (22)–(23) we can observe that as the transmit power goes to infinity, these two strategies approach identical performance. This suggests that the collaborative strategy should be preferred especially when the sensor observation qualities are bad and the transmit power is severely constrained.

We provide a numerical example for an inhomogeneous environment with varying observation and link qualities, where $\sigma_{w,i}^2 \sim \text{Unif}(0,1)$ for each sensor and $\sigma_{v,i}^2 \sim \text{Unif}(0,1)$ for each link. The observation/channel noise are independent across sensors/links. To assure fairness, for both collaborative and non-collaborative schemes, we assume they have identical auto-covariance matrices $\mathbf{R}_w$ and $\mathbf{R}_v$ for each Monte Carlo run (note that since channels are independent, only the best-quality channel is used for the collaborative strategy). For the non-collaborative strategy, we consider an optimal power allocation (OPA) scheme (i.e., [8]) and an equal power allocation (EPA) scheme. Fig. 4 shows the estimation MSEs of three schemes as a function of the total transmit power. We see that in the inhomogeneous setting, the collaborative strategy presents a superior performance advantage over the non-collaborative schemes. To meet a same distortion target, the collaborative scheme needs much less power than the non-collaborative schemes and hence considerable power savings can be achieved.

IV. Optimal Collaboration and Power Allocation: Multiple Cluster Case

A. Proposed Approach

We now examine a general scenario where the network consists of multiple sensor clusters. In this case, the collaboration matrix $\mathbf{C}$ has a block diagonal structure since the inter-cluster collaboration is not allowed. The approach described in previous subsection, therefore, cannot be directly applied here. To solve (6), we hope to decouple the optimization problem into a set of tractable subtasks. To this goal, we rewrite the estimation MSE as follows

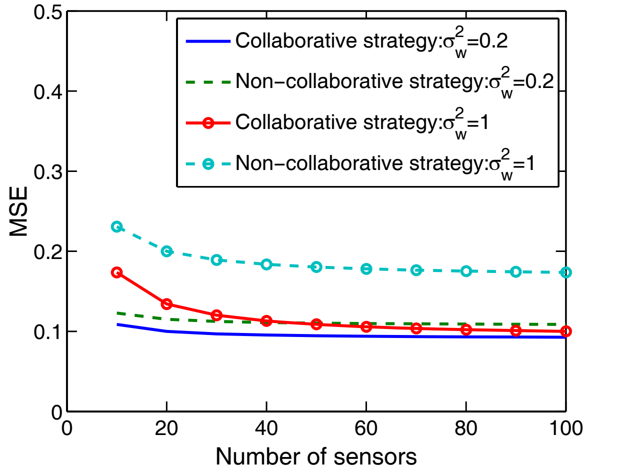


Fig. 2. MSEs of collaborative and non-collaborative strategies vs. number of sensors. $\sigma_v^2 = 0.1$, $\sigma_\theta^2 = 1$, $P = 1$.

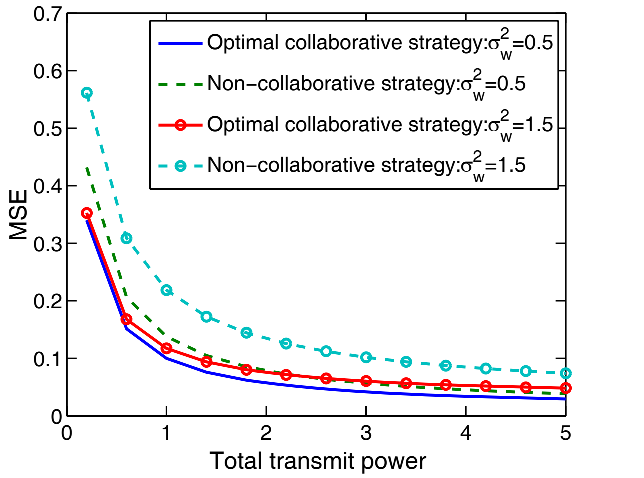


Fig. 3. MSEs of collaborative and non-collaborative strategies vs. total transmit power. $\sigma_v^2 = 0.1$, $\sigma_\theta^2 = 1$, $N = 50$.

$$
\begin{aligned}
E[(\theta - \hat{\theta})^2] &= \sigma_\theta^2 - \sigma_\theta^4 \mathbf{h}^T \mathbf{C}^T (\mathbf{C}\mathbf{R}_x \mathbf{C}^T + \mathbf{R}_v)^{-1} \mathbf{C}\mathbf{h} \\
&\overset{(a)}{=} \left(\sigma_\theta^{-2} + \mathbf{h}^T \mathbf{C}^T (\mathbf{C}\mathbf{R}_w \mathbf{C}^T + \mathbf{R}_v)^{-1} \mathbf{C}\mathbf{h} \right)^{-1} \\
&\overset{(b)}{=} \left(\sigma_\theta^{-2} + \sum_{i=1}^{M} \mathbf{h}_i^T \mathbf{C}_i^T (\mathbf{C}_i \mathbf{R}_{w,i} \mathbf{C}_i^T + \mathbf{R}_{v,i})^{-1} \mathbf{C}_i \mathbf{h}_i \right)^{-1}
\end{aligned} \tag{25}
$$

where (a) is obtained by using the Woodbury identity, along with the fact that $\mathbf{R}_x = \sigma_\theta^2 \mathbf{h}\mathbf{h}^T + \mathbf{R}_w$, and (b) comes by exploiting the block diagonal structures of $\mathbf{C}$, $\mathbf{R}_w$, and $\mathbf{R}_v$. Therefore the optimization problem (6) becomes

$$
\begin{aligned}
\max_{\{\mathbf{C}_i\}} \quad & \sum_{i=1}^{M} \mathbf{h}_i^T \mathbf{C}_i^T (\mathbf{C}_i \mathbf{R}_{w,i} \mathbf{C}_i^T + \mathbf{R}_{v,i})^{-1} \mathbf{C}_i \mathbf{h}_i \\
\text{s.t.} \quad & \sum_{i=1}^{M} \text{tr}\left(\mathbf{C}_i \mathbf{R}_{x,i} \mathbf{C}_i^T \right) \le P
\end{aligned} \tag{26}
$$

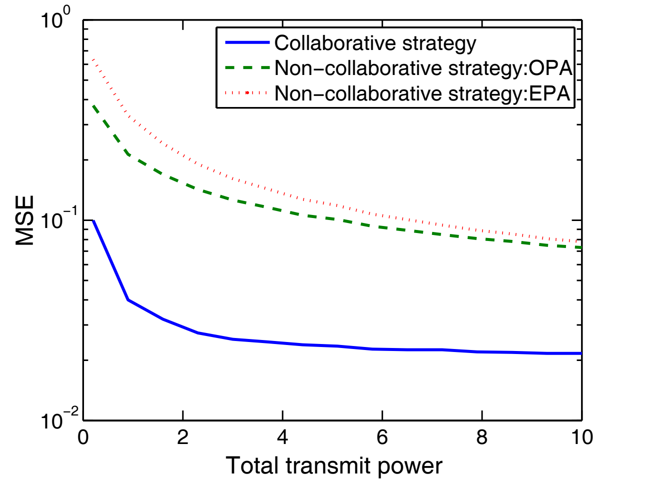


Fig. 4. MSEs of collaborative and non-collaborative schemes vs. total transmit power in inhomogeneous environments. $\sigma_\theta^2 = 1$, $N = 50$.

in which the power constraint follows from the following identity

$$
\text{tr}\left(\mathbf{C}\mathbf{R}_x \mathbf{C}^T \right) = \sum_{i=1}^{M} \text{tr}\left(\mathbf{C}_i \mathbf{R}_{x,i} \mathbf{C}_i^T \right)
$$

In order to utilize the theoretical results obtained for $M = 1$, we express the component $\mathbf{h}_i^T \mathbf{C}_i^T (\mathbf{C}_i \mathbf{R}_{w,i} \mathbf{C}_i^T + \mathbf{R}_{v,i})^{-1} \mathbf{C}_i \mathbf{h}_i$ in (26) as a function of $\mathbf{h}_i^T \mathbf{C}_i^T (\mathbf{C}_i \mathbf{R}_{x,i} \mathbf{C}_i^T + \mathbf{R}_{v,i})^{-1} \mathbf{C}_i \mathbf{h}_i$. Again, this can be done by resorting to the Woodbury identity. We have

$$
\begin{aligned}
&\sigma_\theta^2 - \sigma_\theta^4 \mathbf{h}_i^T \mathbf{C}_i^T (\mathbf{C}_i \mathbf{R}_{x,i} \mathbf{C}_i^T + \mathbf{R}_{v,i})^{-1} \mathbf{C}_i \mathbf{h}_i \\
&= \left(\sigma_\theta^{-2} + \mathbf{h}_i^T \mathbf{C}_i^T (\mathbf{C}_i \mathbf{R}_{w,i} \mathbf{C}_i^T + \mathbf{R}_{v,i})^{-1} \mathbf{C}_i \mathbf{h}_i \right)^{-1}
\end{aligned} \tag{27}
$$

For notational convenience, let

$$
\begin{aligned}
\mu_i(\mathbf{C}_i) &\triangleq \mathbf{h}_i^T \mathbf{C}_i^T (\mathbf{C}_i \mathbf{R}_{w,i} \mathbf{C}_i^T + \mathbf{R}_{v,i})^{-1} \mathbf{C}_i \mathbf{h}_i \\
\eta_i(\mathbf{C}_i) &\triangleq \mathbf{h}_i^T \mathbf{C}_i^T (\mathbf{C}_i \mathbf{R}_{x,i} \mathbf{C}_i^T + \mathbf{R}_{v,i})^{-1} \mathbf{C}_i \mathbf{h}_i
\end{aligned}
$$

Therefore (27) can be rewritten as

$$
\mu_i(\mathbf{C}_i) = \frac{1}{\sigma_\theta^2} \left(\frac{1}{1 - \sigma_\theta^2 \eta_i(\mathbf{C}_i)} - 1 \right) \tag{28}
$$

Substituting (28) into (26), we arrive at the following optimization

$$
\begin{aligned}
\max_{\{\mathbf{C}_i\}} \quad & \sum_{i=1}^{M} \frac{1}{\sigma_\theta^2} \left(\frac{1}{1 - \sigma_\theta^2 \eta_i(\mathbf{C}_i)} - 1 \right) \\
\text{s.t.} \quad & \sum_{i=1}^{M} \text{tr}\left(\mathbf{C}_i \mathbf{R}_{x,i} \mathbf{C}_i^T \right) \le P
\end{aligned} \tag{29}
$$

Clearly, (29) can be decoupled into two sequential subtasks, i.e. a power allocation (among clusters) problem and a set of collaboration matrix design problems that can be solved using the previous results. To see this, suppose $\{P_1^*, P_2^*, \ldots, P_M^*\}$ is an optimum power assignment with

$$
\begin{aligned}
&\text{tr}\left(\mathbf{C}_i \mathbf{R}_{x,i} \mathbf{C}_i^T \right) \le P_i^* \quad \forall i \in \{1, \ldots, M\} \\
&\sum_{i=1}^{M} P_i^* \le P
\end{aligned}
$$

then (29) is simplified into a set of identical problems as

$$\max_{\mathbf{C}_i} \quad \mu_i(\mathbf{C}_i) = \frac{1}{\sigma_\theta^2}\left(\frac{1}{1-\sigma_\theta^2\eta_i(\mathbf{C}_i)} - 1\right)$$
$$\text{s.t.} \quad \text{tr}\left(\mathbf{C}_i\mathbf{R}_{x,i}\mathbf{C}_i^T\right) \le P_i^* \tag{30}$$

Note that $\sigma_\theta^2\eta_i(\mathbf{C}_i)$ must lie within the interval $(0, 1)$ because we have $\eta_i(\mathbf{C}_i) > 0$ and $\mu_i(\mathbf{C}_i) > 0$ from their definitions. Hence (30) is equivalent to

$$\max_{\mathbf{C}_i} \quad \eta_i(\mathbf{C}_i)$$
$$\text{s.t.} \quad \text{tr}\left(\mathbf{C}_i\mathbf{R}_{x,i}\mathbf{C}_i^T\right) \le P_i^* \tag{31}$$

which is exactly the optimization problem discussed in the previous section. The optimal solution to (31) is given in Theorem 1. The key problem, therefore, is to determine the optimum power assignment $\{P_1^*, P_2^*, \ldots, P_M^*\}$. To this goal, we need to find out the relationship between the maximum objective function value $\eta_i(\mathbf{C}_i^*)$ and P_i^*. Recalling Theorem 1, more precisely, (18), we have

$$\eta_i(\mathbf{C}_i^*) = \frac{P_i^*}{P_i^* + \min(\mathbf{d}_{v,i})}\mathbf{h}_i^T\mathbf{R}_{x,i}^{-1}\mathbf{h}_i \triangleq \frac{\alpha_i P_i^*}{\beta_i + P_i^*} \tag{32}$$

where we define $\alpha_i \triangleq \mathbf{h}_i^T\mathbf{R}_{x,i}^{-1}\mathbf{h}_i$, $\beta_i \triangleq \min(\mathbf{d}_{v,i})$, and $\mathbf{d}_{v,i}$ is a column vector consisting of the eigenvalues of $\mathbf{R}_{v,i}$ (note that $\mathbf{R}_{v,i}$ can be non-diagonal). Substituting (32) into the objective function of (29), we get

$$\sum_{i=1}^{M}\frac{1}{\sigma_\theta^2}\left(\frac{1}{1-\sigma_\theta^2\eta_i(\mathbf{C}_i^*)} - 1\right) = \sum_{i=1}^{M}\frac{\alpha_i P_i^*}{(1-\sigma_\theta^2\alpha_i)P_i^* + \beta_i} \tag{33}$$

Clearly, the optimal power allocation $\{P_1^*, P_2^*, \ldots, P_M^*\}$ must be the one, among all feasible power assignments, which maximizes (33). Therefore, it can be found out by

$$\min_{\{P_1,\ldots,P_M\}} \quad -\sum_{i=1}^{M}\frac{\alpha_i P_i}{(1-\sigma_\theta^2\alpha_i)P_i + \beta_i}$$
$$\text{s.t.} \quad \sum_{i=1}^{M} P_i \le P$$
$$P_i \ge 0 \quad \forall i \in \{1, \ldots, M\} \tag{34}$$

It is easy to verify that the optimization problem (34) is convex because its Hessian matrix, which is a diagonal matrix in this case, is positive semidefinite on the convex set defined by the linear constraints. Although (34) is efficiently solvable by numerical methods, it can also be solved analytically by resorting to the Lagrangian function and Karush-Kuhn-Tucker (KKT) conditions, which leads to a water-filling type power allocation scheme. The details are elaborated in Appendix C. Briefly speaking, for a threshold λ that is uniquely determined by a procedure described in Appendix C, we have

$$P_i = \begin{cases} \frac{1}{\varphi_i}\left(\sqrt{\frac{\phi_i}{\lambda}} - 1\right) & \phi_i \ge \lambda \\ 0 & \text{otherwise} \end{cases} \tag{35}$$

where $\phi_i \triangleq \alpha_i/\beta_i$ and $\varphi_i \triangleq (1-\sigma_\theta^2\alpha_i)/\beta_i$. It is easy to see that each cluster can decide whether to transmit or keep silent by the criterion $\phi_i \ge \lambda$. Note that ϕ_i is the ratio of $\mathbf{h}_i^T\mathbf{R}_{x,i}^{-1}\mathbf{h}_i$ to $\min(\mathbf{d}_{v,i})$, with the former a measure of the cluster's estimation quality (a larger value indicates a better estimation accuracy) and the latter a measure of the cluster's channel quality (a smaller value indicates a better channel quality).

So far we have developed an analytical approach which leads to an optimal solution to (6). For clarity, we now summarize the steps of our proposed method.

1) Given the prior knowledge of the auto-correlation matrices $\{\mathbf{R}_{v,i}\}_{i=1}^M$, $\{\mathbf{R}_{w,i}\}_{i=1}^M$ and the observation gain vectors $\{\mathbf{h}_i\}_{i=1}^M$, compute $\{\alpha_i\}_{i=1}^M$ and $\{\beta_i\}_{i=1}^M$, where $\alpha_i = \mathbf{h}_i^T\mathbf{R}_{x,i}^{-1}\mathbf{h}_i$ and $\beta_i = \min(\mathbf{d}_{v,i})$.
2) Given the total power constraint P, find the optimal power allocation among clusters via (34). The solution of (34) is elaborated in Appendix C.
3) With the optimal power assignment $\{P_1^*, P_2^*, \ldots, P_M^*\}$ derived in the previous step, determine the optimal collaboration matrices $\{\mathbf{C}_i\}_{i=1}^M$ via (31), whose solution is detailed in Theorem 1.

Briefly speaking, our collaborative scheme involves power allocation (among clusters) and collaboration matrix design. Both steps admit simple distributed implementations. For the power allocation step, note that the power assigned to each cluster, P_i, is computed by (35), where α_i and β_i can be calculated using each cluster's local information, and λ is common to all clusters, thus can be computed at the FC and then broadcasted to all clusters. Based on the computed power P_i, each cluster can determine the corresponding collaboration matrix locally without requiring information from other clusters (c.f. (17)). In particular, since the optimal collaboration is equivalent to constructing a local LMMSE estimate, the sensor collaboration can be implemented in a fully distributed manner without the aid of a cluster head. The distributed implementation of a LMMSE estimator has been studied in many works, e.g. [14] and the references therein.

B. Simulation Results

We now provide some numerical examples to illustrate our proposed method. We firstly investigate the relationship between the estimation accuracy and the number of clusters, given a fixed amount of total transmit power. Consider a homogeneous environment with identical observation qualities across all sensors and identical link qualities across all channels from the cluster heads to the FC. Each cluster has $N_i = 10$ sensors for $i \in \{1, \ldots, M\}$. Fig. 5 shows the estimation MSE of our collaborative scheme as a function of the number of clusters, M, with the total transmit $P = 1$, $P = 5$, and $P = 10$, respectively. We see that with an increasing number of clusters, the estimation MSE is monotonically decreasing and approaches to a nonzero limit. This observation can be readily verified by our theoretical results because an equal power allocation is optimum for a homogeneous environment. By using (25) and (33), the estimation MSE of the collaborative scheme is given by

$$\mathrm{MSE_{OC}} = \left[\frac{1}{\sigma_\theta^2} + \frac{\alpha PM}{(1-\sigma_\theta^2\alpha)P + \beta M}\right]^{-1} \quad (36)$$

with $\alpha = \mathbf{h}_i^T \mathbf{R}_{x,i}^{-1}\mathbf{h}_i\ \forall i$, and $\beta = \sigma_v^2$. In this setting, we also compare our collaborative scheme with the non-collaborative scheme in Fig. 6. As expected, the collaborative scheme surpasses the non-collaborative scheme in accuracy/power efficiency.

We now consider an inhomogeneous environment with varying observation and channel qualities across clusters. For each cluster, say cluster i, we have $\mathbf{R}_{w,i} = \sigma_{w,i}^2\mathbf{I}$, where $\sigma_{w,i}^2 \sim \mathrm{Unif}(0,1)$, and there is one channel from the cluster head to the FC with $\sigma_{v,i}^2 \sim \mathrm{Unif}(0,1)$. To illustrate the effectiveness of the optimal power allocation among clusters, we compare our optimal solution with an collaborative scheme with an equal power allocation among the cluster. Note that for the latter strategy, the collaboration matrices can be determined from (31) with P_i^* replaced by the equally assigned power. Fig. 7 depicts the estimation MSE of the two schemes as a function of the total transmit power. The performance gain is due to the fact that the optimal scheme is able to make more efficient use of the total transmit power by taking into account cluster disparity.

It is also interesting to consider the following question that would give us an useful hint for the network design: is it better to place all sensors into one cluster or to divide them into a few clusters? Our theoretical results for one cluster case implicitly provide an answer to this question: the best choice is to form one huge cluster. This is because Theorem 1 points out that when sensors have the communication resources to collaborate, the optimal collaboration scheme is to pool all data from all sensors to construct a LMMSE estimate, instead of partitioning those sensors into a few clusters. Nevertheless, when sensors are geographically far apart, it may be impractical to form and maintain one huge cluster because the energy consumed by sensor collaboration increases exponentially with the inter-sensor distance and will be no longer negligible. Anyway, our results suggest that, with an acceptable collaboration cost, collaboration across sensors should be encouraged as much as possible.

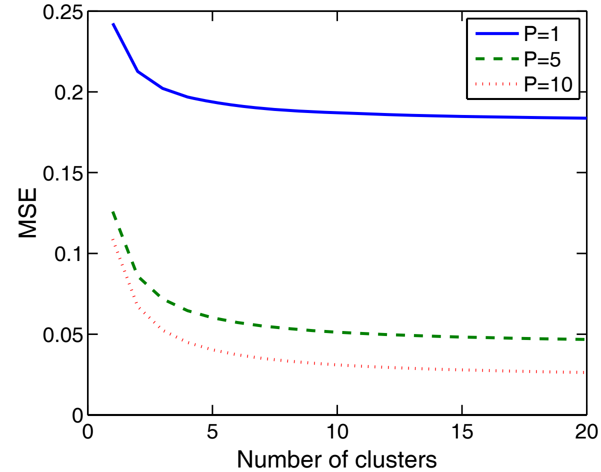


Fig. 5. MSE of the collaborative scheme vs. number of clusters. $\sigma_v^2 = 0.2$, $\sigma_w^2 = 1$, $\sigma_\theta^2 = 1$.

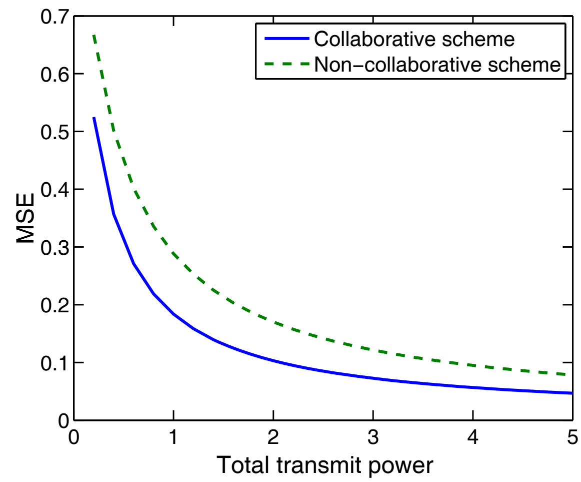


Fig. 6. MSEs of the collaborative and non-collaborative schemes. $\sigma_v^2 = 0.2$, $\sigma_w^2 = 1$, $\sigma_\theta^2 = 1$, $M = 15$.

V. Conclusions and Outlooks

We studied an optimal collaboration and power allocation problem for distributed estimation in a power-constrained collaborative sensor network, where the network consists of a number of sensor clusters, and collaboration is allowed within the same cluster but not across clusters. Our theoretical results showed that, given a specified total transmit power, the power should be assigned among the clusters in a water-filling manner, with each cluster deciding whether to transmit or keep silent by comparing with a threshold, namely, the ratio of a measure of the cluster's estimation quality to a measure of the cluster's channel quality. Also, for each cluster, if the channels from this cluster to the FC are uncorrelated, then an optimal collaboration yields only one local message which is sent from the best channel within the cluster to the FC; otherwise the local message has to be sent across all channels within the cluster at different power levels matched to their channel quality. Specifically, in either case, the compressed local message is exactly the local LMMSE estimate multiplied by an amplification factor. Simulation results have been presented to corroborate our theoretical analysis.

Our future investigation will relax the ideal collaboration assumption and take into account the noise arising from communication within each cluster (between sensors and the cluster head) and the energy consumption caused by sensor collaboration. It would be interesting to examine the optimal power allocation between collaboration and message transmission (from cluster heads to the FC). Another interesting direction is the extension of our proposed collaboration scheme to the vector parameter case, i.e. the parameter to be estimated is a vector instead of a scalar, and the elements of the vector parameter could be mutually independent or correlated. In this case, we can treat the scalar parameters individually and place individual power constraint on each scalar parameter. Hence the problem becomes the same as that in this paper except that the observation model may no longer be a linear signal-

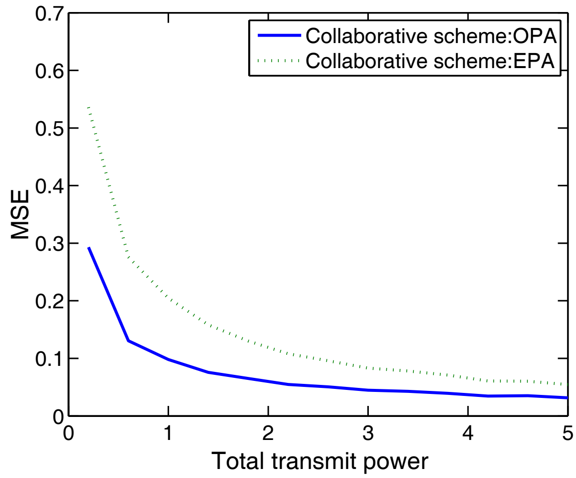


Fig. 7. MSEs of the collaborative schemes with optimal power allocation and equal power allocation. $\sigma_\theta^2 = 1$, $M = 10$.

plus-noise model due to the correlation among different scalar parameters. Nevertheless, the optimal collaboration matrix and the optimal power allocation among the clusters can still be determined by following the same derivation as we did in this paper. Another formulation, instead of treating the parameters individually, is to consider minimizing the overall estimation distortion under a total power constraint. This formulation has the disadvantage that the estimation distortion target can only be set for overall parameters, but not for individual parameters. Also, an optimal solution, in this case, may not be found and hence an effective sub-optimal solution is desirable.

Appendix A
Proof of Lemma 1

We firstly fix the optimization variables $\{\mathbf{D}, \mathbf{Q}\}$ and search for an optimal $\mathbf{U}$. The optimization (13) is reduced to

$$\begin{aligned} \max_{\mathbf{U}} \quad & \text{tr}\left(\mathbf{D}_g^{\frac{1}{2}}\mathbf{U}\mathbf{D}^{\frac{1}{2}}(\mathbf{I}+\mathbf{D})^{-1}\mathbf{D}^{\frac{1}{2}}\mathbf{U}^T\mathbf{D}_g^{\frac{1}{2}}\right) \\ \text{s.t.} \quad & \mathbf{U}^T\mathbf{U} = \mathbf{I} \end{aligned} \tag{37}$$

Note that for any matrix $\mathbf{U}$ satisfying the orthogonality constraint: $\mathbf{U}^T\mathbf{U} = \mathbf{I}$, the objective function in (37) is upper-bounded by

$$\begin{aligned} &\text{tr}\left(\mathbf{D}_g^{\frac{1}{2}}\mathbf{U}\mathbf{D}^{\frac{1}{2}}(\mathbf{I}+\mathbf{D})^{-1}\mathbf{D}^{\frac{1}{2}}\mathbf{U}^T\mathbf{D}_g^{\frac{1}{2}}\right) \\ =&\text{tr}\left(\mathbf{D}^{\frac{1}{2}}(\mathbf{I}+\mathbf{D})^{-1}\mathbf{D}^{\frac{1}{2}}\mathbf{U}^T\mathbf{D}_g\mathbf{U}\right) \\ \leq&\frac{d_{\max}}{1+d_{\max}}\text{tr}\left(\mathbf{U}^T\mathbf{D}_g\mathbf{U}\right) \\ \overset{(a)}{\leq}&\frac{d_{\max}}{1+d_{\max}}\text{tr}\left(\mathbf{D}_g\right) \\ \overset{(b)}{=}&\frac{d_{\max}}{1+d_{\max}}\mathbf{D}_g(1,1) \end{aligned} \tag{38}$$

where $d_{\max}$ denotes the largest diagonal element of $\mathbf{D}$, (a) follows from an inequality [32, Corollary 4.3.18] which is a generalized result of the Rayleigh-Ritz theorem (note that $\mathbf{U}$ is not an orthonormal matrix, but only a portion of it), and (b) follows by recognizing that $\mathbf{D}_g(1,1)$ is the only one nonzero diagonal element of $\mathbf{D}_g$ (note that $\mathbf{D}_g$ has only one nonzero diagonal element since $\mathbf{G}$ is rank-one). Without loss of generality, we assume that the diagonal elements of $\mathbf{D}$ are sorted in a descending order, i.e. $d_1 \geq d_2 \geq \ldots \geq d_p$. We can easily verify that this upper bound is achieved only when $\mathbf{U} = \mathbf{\Gamma}$, where $\mathbf{\Gamma}$ is defined in (14). Consequently we conclude that $\mathbf{U} = \mathbf{\Gamma}$ is an optimal solution to (37), which is independent of the optimization variables $\{\mathbf{D}, \mathbf{Q}\}$.

We now search for optimal $\{\mathbf{D}, \mathbf{Q}\}$ by utilizing the above results. The optimization is simplified into

$$\begin{aligned} \max_{\mathbf{D},\mathbf{Q}} \quad & \frac{d_1}{1+d_1}\mathbf{D}_g(1,1) \\ \text{s.t.} \quad & \text{tr}\left(\mathbf{D}\mathbf{Q}^T\mathbf{D}_v\mathbf{Q}\right) \leq P \\ & \mathbf{D} = \text{diag}(d_1,\ldots,d_p) \quad d_1 \geq d_2 \geq \ldots \geq d_p \geq 0 \\ & \mathbf{Q}\mathbf{Q}^T = \mathbf{I} \end{aligned} \tag{39}$$

Note that we replace $d_{\max}$ with d_1 because we assume that the diagonal elements of $\mathbf{D}$ are in a descending order. It is clear to see that (39) is equivalent to maximizing d_1, which is upper bounded by

$$\begin{aligned} d_1 &\overset{(a)}{\leq} \frac{P - \sum_{i=2}^p d_i\xi_i}{\xi_1} \\ &\overset{(b)}{\leq} \frac{P}{\xi_1} \overset{(c)}{\leq} \frac{P}{\min(\mathbf{d}_v)} \end{aligned} \tag{40}$$

where the inequality (a) comes from the constraint $\text{tr}\left(\mathbf{D}\mathbf{Q}^T\mathbf{D}_v\mathbf{Q}\right) \leq P$ and ξ_i denotes the ith diagonal element of $\mathbf{\Xi} \triangleq \mathbf{Q}^T\mathbf{D}_v\mathbf{Q}$ (note that $\mathbf{\Xi}$ is not necessarily diagonal); the inequality (b) comes by recognizing that both $\{d_i\}$ and $\{\xi_i\}$ are non-negative (it can be easily proved that $\xi_i > 0$ since $\mathbf{D}_v$ has positive diagonal elements); the inequality (c) follows, again, from $\xi_1 \geq \min(\mathbf{d}_v)$, which is a result readily derived from the inequality [32, Corollary 4.3.18]. As mentioned before, we have $\min(\mathbf{d}_v) = \mathbf{D}_v(1,1)$ because $\mathbf{D}_v$ is in an ascending order. Clearly, the upper bound is attainable only for matrices $\mathbf{Q} = \mathbf{I}$ and $\mathbf{D} = \mathbf{\Lambda}$, where $\mathbf{\Lambda}$ is defined in (15). They also satisfy the constraints specified in (39). Hence $\mathbf{Q} = \mathbf{I}$ and $\mathbf{D} = \mathbf{\Lambda}$ are the optimal solution of (39).

Note that, in the previous derivation, the diagonal elements of $\mathbf{D}$ are assumed in a descending order. This assumption can be relaxed by setting $\mathbf{D}^* = \mathbf{P}\mathbf{\Lambda}\mathbf{P}^T$ for any permutation matrix $\mathbf{P}$. The optimal solution of $\mathbf{U}$ and $\mathbf{Q}$, correspondingly becomes $\mathbf{U}^* = \mathbf{\Gamma}\mathbf{P}^T$, and $\mathbf{Q}^* = \mathbf{P}^T$, which can be readily verified.

Substituting the derived optimal matrices $\{\mathbf{U}^*, \mathbf{D}^*, \mathbf{Q}^*\}$ into (13), the objective function is computed as follows

$$\begin{aligned} f(\mathbf{U}^*, \mathbf{D}^*, \mathbf{Q}^*) =& \frac{P}{P+\min(\mathbf{d}_v)}\mathbf{D}_g(1,1) \\ \overset{(a)}{=}& \frac{P}{P+\min(\mathbf{d}_v)}\mathbf{h}^T\mathbf{R}_x^{-1}\mathbf{h} \end{aligned} \tag{41}$$

where (a) follows directly by noting that $\mathbf{D}_g(1,1)$ is the only nonzero eigenvalue of $\mathbf{G} = \mathbf{R}_x^{-\frac{1}{2}}\mathbf{h}\mathbf{h}^T\mathbf{R}_x^{-\frac{1}{2}}$.

APPENDIX B
PROOF OF THEOREM 1

Combining (8) and (11), we can quickly establish the following relationship

$$\mathbf{C} = \mathbf{U}_v \mathbf{D}_v^{\frac{1}{2}} \mathbf{Q} \mathbf{D}^{\frac{1}{2}} \mathbf{U}^T \mathbf{U}_g^T \mathbf{R}_x^{-\frac{1}{2}} \quad (42)$$

By substituting the derived optimal matrices $\{\mathbf{U}, \mathbf{Q}, \mathbf{D}\}$ into (42), and utilizing $(\mathbf{D}^*)^{\frac{1}{2}} = \mathbf{P}\mathbf{\Lambda}^{\frac{1}{2}}\mathbf{P}^T$ and $\mathbf{P}\mathbf{P}^T = \mathbf{P}^T\mathbf{P} = \mathbf{I}$, the optimal collaboration matrix is therefore given by

$$\begin{aligned} \mathbf{C}^* =& \mathbf{U}_v \mathbf{D}_v^{\frac{1}{2}} \mathbf{\Lambda}^{\frac{1}{2}} \mathbf{\Gamma}^T \mathbf{U}_g^T \mathbf{R}_x^{-\frac{1}{2}} \\ \overset{(a)}{=}& \mathbf{U}_v \mathbf{\Psi} \mathbf{U}_g^T \mathbf{R}_x^{-\frac{1}{2}} \\ \overset{(b)}{=}& \sqrt{P} \mathbf{U}_v[:,1] \mathbf{e}_1^T \mathbf{U}_g^T \mathbf{R}_x^{-\frac{1}{2}} \\ \overset{(c)}{=}& \sqrt{P} \mathbf{U}_v[:,1] \gamma \mathbf{h}^T \mathbf{R}_x^{-\frac{1}{2}} \mathbf{R}_x^{-\frac{1}{2}} \\ =& \sqrt{P} \mathbf{U}_v[:,1] \gamma \mathbf{h}^T \mathbf{R}_x^{-1} \end{aligned} \quad (43)$$

where in (a), we define $\mathbf{\Psi} \triangleq \mathbf{D}_v^{\frac{1}{2}} \mathbf{\Lambda}^{\frac{1}{2}} \mathbf{\Gamma}^T$, and $\mathbf{\Psi}$ can be easily computed, which is a matrix with its first row equal to $\sqrt{P}\mathbf{e}_1^T$ and all other elements equal to zero, i.e.

$$\mathbf{\Psi} = \begin{bmatrix} \sqrt{P}\mathbf{e}_1^T \\ \mathbf{0}_{(p-1)\times N} \end{bmatrix} \quad (44)$$

$\mathbf{e}_i$ denotes the unit column vector with its ith entry equal to one, and its other entries equal to zero; (b) follows directly by noting the sparse structure of $\mathbf{\Psi}$ and $\mathbf{U}_v[:,1]$ denotes the first column of $\mathbf{U}_v$; (c) comes from the fact that $\mathbf{G} \triangleq \mathbf{R}_x^{-\frac{1}{2}} \mathbf{h}\mathbf{h}^T \mathbf{R}_x^{-\frac{1}{2}} = \mathbf{U}_g \mathbf{D}_g \mathbf{U}_g^T$ is a rank-one matrix with only one null eigenvalue associated with the eigenvector $\gamma \mathbf{R}_x^{-\frac{1}{2}} \mathbf{h}$ (note that this eigenvector is exactly the first column of $\mathbf{U}_g$), where $\gamma \triangleq \frac{1}{\sqrt{\mathbf{h}^T \mathbf{R}_x^{-1} \mathbf{h}}}$ is the normalization scalar. The associated estimation mean-square error can be readily derived by using (16).

APPENDIX C
AN ANALYTICAL SOLUTION TO (34)

For notational convenience, let $\phi_i \triangleq \alpha_i/\beta_i$ and $\varphi_i \triangleq (1 - \sigma_\theta^2 \alpha_i)/\beta_i$. The Lagrangian function L associated with (34) is given by

$$L(P_i; \lambda; \nu_i) = -\sum_{i=1}^{M} \frac{\phi_i P_i}{\varphi_i P_i + 1} - \lambda \left(P - \sum_{i=1}^{M} P_i \right) - \sum_{i=1}^{M} \nu_i P_i \quad (45)$$

which gives the following KKT conditions [33]:

$$\begin{aligned} -\frac{\phi_i}{(\varphi_i P_i + 1)^2} + \lambda - \nu_i =& 0 \quad \forall i \\ P - \sum_{i=1}^{M} P_i =& 0 \\ \nu_i P_i =& 0 \quad \forall i \\ \nu_i \geq& 0 \quad \forall i \\ P_i \geq& 0 \quad \forall i \end{aligned}$$

By solving the first equation of the above KKT conditions, we obtain

$$P_i = \frac{1}{\varphi_i} \left[\sqrt{\frac{\phi_i}{\lambda - \nu_i}} - 1 \right] \quad \forall i \quad (46)$$

Also, the KKT conditions: $\nu_i P_i = 0$, $\nu_i \geq 0$, and $P_i \geq 0$ imply that we have either $\{\nu_i = 0, P_i > 0\}$ or $\{\nu_i > 0, P_i = 0\}$. Therefore (46) becomes

$$P_i = \frac{1}{\varphi_i} \left[\sqrt{\frac{\phi_i}{\lambda}} - 1 \right]^+ \quad \forall i \quad (47)$$

where $[x]^+$ is equal to x if $x > 0$, otherwise it is zero. The Lagrangian multiplier λ and the number of active clusters (those are assigned nonzero power) can be uniquely determined from the power constraint.

Suppose we have $K \in \{1, \ldots, M\}$ active clusters, according to (47), these K clusters must be $\{k_1, k_2, \ldots, k_K\}$, where $\{k_i\}$ is a set of indices such that $\phi_{k_1} \geq \phi_{k_2} \geq \ldots \geq \phi_{k_M}$. Therefore λ can be solved by substituting $\{P_{k_1}, P_{k_2}, \ldots, P_{k_K}\}$ into the second KKT condition, where P_{k_i} is given by

$$P_{k_i} = \frac{1}{\varphi_{k_i}} \left[\sqrt{\frac{\phi_{k_i}}{\lambda}} - 1 \right] \quad (48)$$

Now we substitute λ back to (47). We will get a new solution $\{P'_{k_1}, P'_{k_2}, \ldots, P'_{k_K}, P'_{k_{K+1}}, \ldots, P'_{k_M}\}$. If this new solution is exactly identical to the one we assumed before, i.e. $\{P_{k_1}, P_{k_2}, \ldots, P_{k_K}, 0, \ldots, 0\}$. Then it is the true solution we are looking for; otherwise we have to choose another K to repeat the above procedure. Also, it has been proved that such a solution is unique and always exists [7].

REFERENCES

[1] D. Li, K. D. Wong, Y. H. Hu, and A. M. Sayeed, "Detection, classification, and tracking of targets," *IEEE Signal Processing Mag.*, vol. 19, no. 2, pp. 17-29, Mar. 2002.

[2] I. F. Akyildiz, W. Su, Y. Sankarasubramaniam, and E. Cayirci, "A survey on sensor networks," *IEEE Commun. Mag.*, pp. 102-114, Aug. 2002.

[3] H. Papadopoulos, G. Wornell, and A. Oppenheim, "Sequential signal encoding from noisy measurements using quantizers with dynamic bias control," *IEEE Trans. Inform. Theory*, vol. 47, no. 2, pp. 978-1002, Mar. 2001.

[4] A. Ribeiro and G. B. Giannakis, "Bandwidth-constrained distributed estimation for wireless sensor networks—part I: Gaussian PDF," *IEEE Trans. Signal Processing*, vol. 54, no. 3, pp. 1131-1143, Mar. 2006.

[5] Z. Luo, "An isotropic universal decentralized estimation scheme for a bandwidth constrained ad hoc sensor network," *IEEE J. Select. Areas Commun.*, vol. 23, no. 4, pp. 735-744, Apr. 2005.

[6] J. Fang and H. Li, "Distributed adaptive quantization for wireless sensor networks: from Delta modulation to maximum likelihood," *IEEE Trans. Signal Processing*, vol. 56, no. 10, pp. 5246-5257, Oct. 2008.

[7] J.-J. Xiao, S. Cui, Z.-Q. Luo, and A. J. Goldsmith, "Power scheduling of universal decentralized estimation in sensor networks," *IEEE Trans. Signal Processing*, vol. 54, no. 2, pp. 413-422, Feb. 2006.

[8] S. Cui, J.-J. Xiao, A. J. Goldsmith, Z.-Q. Luo, and H. V. Poor, "Estimation diversity and energy efficiency in distributed sensing," *IEEE Trans. Signal Processing*, no. 9, pp. 4683-4695, Sept. 2007.

[9] J. Li and G. AlRegib, "Rate-constrained distributed estimation in wireless sensor networks," *IEEE Trans. Signal Processing*, vol. 55, no. 5, pp. 1634-1643, May 2007.

[10] J.-Y. Wu, Q.-Z. Huang, and T.-S. Lee, "Minimal energy decentralized estimation via exploiting the statistical knowledge of sensor noise variance," *IEEE Trans. Signal Processing*, vol. 56, no. 5, pp. 2171-2176, May 2008.

[11] Y. Huang and Y. Hua, "Multi-hop progressive decentralized estimation in wireless sensor networks," *IEEE Signal Processing Lett.*, vol. 14, no. 12, pp. 1004-1007, Dec. 2007.

[12] L. Xiao and S. Boyd, "Fast linear iterations for distributed averaging," *Systems Control Lett.*, no. 53, pp. 65-78, Feb. 2004.

[13] D. S. Scherber and H. C. Papadopoulos, "Distributed computation of averages over ad hoc networks," *IEEE J. Select. Areas Commun.*, vol. 23, pp. 776-787, Apr. 2005.

[14] V. Delouille, R. N. Neelamani, and R. G. Baraniuk, "Robust distributed estimation using the embedded subgraphs algorithm," *IEEE Trans. Signal Processing*, no. 8, pp. 2998-3010, Aug. 2006.

[15] E. Kokiopoulou and P. Frossard, "Accelerating distributed consensus using extrapolation," *IEEE Signal Processing Lett.*, vol. 14, pp. 665-668, Oct. 2007.

[16] S. Barbarossa and G. Scutari, "Decentralized maximum-likelihood estimation for sensor networks composed of nonlinearly coupled dynamical systems," *IEEE Trans. Signal Processing*, no. 7, pp. 3456-3470, July 2007.

[17] L. Xiao, S. Boyd, and S.-J. Kim, "Distributed average consensus with least-mean-square deviation," in *Proc. 17th International Symp. Mathematical Theory Networks Systems*, 2006.

[18] P. Ishwar, R. Puri, K. Ramchandran, and S. S. Pradhan, "On rate-constrained distributed estimation in unreliable sensor networks," *IEEE J. Select. Areas Commun.*, vol. 23, no. 4, pp. 765-775, Apr. 2005.

[19] G. Thatte and U. Mitra, "Sensor selection and power allocation for distributed estimation in sensor networks: beyond the star topology," *IEEE Trans. Signal Processing*, vol. 56, no. 7, pp. 2649-2661, July 2008.

[20] K. Zhang, X. R. Li, P. Zhang, and H. Li, "Optimal linear estimation fusion—part VI: sensor data compression," in *Peoc. International Conf. Inform. Fusion*, Queensland, Australia, 2003.

[21] Z.-Q. Luo, G. B. Giannakis, and S. Zhang, "Optimal linear decentralized estimation in a bandwidth constrained sensor network," in *Proc. 2005 IEEE International Symp. Information*, Sept. 2005.

[22] Y. Zhu, E. Song, J. Zhou, and Z. You, "Optimal dimensionality reduction of sensor data in multisensor estimation fusion," *IEEE Trans. Signal Processing*, vol. 53, no. 5, pp. 1631-1639, May 2005.

[23] E. Song, Y. Zhu, and J. Zhou, "Sensors' optimal dimensionality compression matrix in estimation fusion," *Automatica*, vol. 41, pp. 2131-2139, Nov. 2005.

[24] I. D. Schizas, G. B. Giannakis, and Z.-Q. Luo, "Distributed estimation using reduced dimensionality sensor observations," *IEEE Trans. Signal Processing*, no. 8, pp. 4284-4299, Aug. 2007.

[25] J. Fang and H. Li, "Joint dimension assignment and compression for distributed multisensor estimation," *IEEE Signal Processing Lett.*, pp. 174-177, Jan. 2008.

[26] M. Gastpar, B. Rimoldi, and M. Vetterli, "To code, or not to code: lossy source channel communication revisited," *IEEE Trans. Inform. Theory*, vol. 49, no. 5, pp. 1147-1158, May 2003.

[27] M. Gastpar and M. Vetterli, "Source-channel communication in sensor networks," *Lecture Notes Computer Science*, pp. 162-177, Apr. 2003.

[28] J.-J. Xiao and Z.-Q. Luo, "Multiterminal source-channel communication over an orthogonal multiple-access channel," *IEEE Trans. Inform. Theory*, vol. 53, no. 9, pp. 3255-3264, Sept. 2007.

[29] S. M. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory*. Upper Saddle River, NJ: Prentice Hall, 1993.

[30] D. J. Sakrison, "Source encoding in the presence of random disturbance," *IEEE Trans. Inform. Theory*, vol. 14, no. 1, pp. 165-167, Jan. 1968.

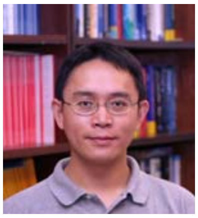
[31] J. Wolf and J. Ziv, "Transmission of noisy information to a noisy receiver with minimum distortion," *IEEE Trans. Inform. Theory*, vol. 16, no. 4, pp. 406-411, July 1970.

[32] R. A. Horn and C. R. Johnson, *Matrix Analysis*. Cambridge University Press, 1985.

[33] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge, U.K.: Cambridge University Press, 2003.

Jun Fang (M'08) received the B.Sc. degree and M.Sc. degrees in electrical engineering from Xidian University, Xi'an, China in 1998 and 2001, respectively, and the Ph.D. degree in electrical engineering from National University of Singapore, Singapore, in 2006.

During 2006, he was with the Department of Electrical and Computer Engineering, Duke University as a postdoctoral research associate. Currently he is a postdoctoral research associate with the Department of Electrical and Computer Engineering, Stevens Institute of Technology. His research interests include statistical signal processing, wireless communications, and distributed estimation and detection with their applications on wireless sensor networks.



Hongbin Li (M'99) received the B.S. and M.S. degrees from the University of Electronic Science and Technology of China, Chengdu, in 1991 and 1994, respectively, and the Ph.D. degree from the University of Florida, Gainesville, FL, in 1999, all in electrical engineering.

From July 1996 to May 1999, he was a Research Assistant in the Department of Electrical and Computer Engineering at the University of Florida. He was a Summer Visiting Faculty Member at the Air Force Research Laboratory, Rome, NY, in the summers of 2003 and 2004. Since July 1999, he has been with the Department of Electrical and Computer Engineering, Stevens Institute of Technology, Hoboken, NJ, where he is an Associate Professor. His current research interests include statistical signal processing, wireless communications, and radars.

Dr. Li is a member of Tau Beta Pi and Phi Kappa Phi. He received the Harvey N. Davis Teaching Award in 2003 and the Jess H. Davis Memorial Award for excellence in research in 2001 from Stevens Institute of Technology, and the Sigma Xi Graduate Research Award from the University of Florida in 1999. He is a member of the Sensor Array and Multichannel (SAM) Technical Committee of the IEEE Signal Processing Society. He is/has been an editor or associate editor for the IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS, IEEE SIGNAL PROCESSING LETTERS, and IEEE TRANSACTIONS ON SIGNAL PROCESSING, and served as a Guest Editor for EURASIP JOURNAL ON APPLIED SIGNAL PROCESSING, Special Issue on Distributed Signal Processing Techniques for Wireless Sensor Networks.