Optimal/Near-Optimal Dimensionality Reduction for Distributed Estimation in Homogeneous and Certain Inhomogeneous Scenarios

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

***Abstract*—We consider distributed estimation of a deterministic vector parameter from noisy sensor observations in a wireless sensor network (WSN). The observation noise is assumed uncorrelated across sensors. To meet stringent power and bandwidth budgets inherent in WSNs, local data dimensionality reduction is performed at each sensor to reduce the number of messages sent to a fusion center (FC). The problem of interest is to jointly design the compression matrices associated with those sensors, aiming at minimizing the estimation error at the FC. Such a dimensionality reduction problem is investigated in this paper. Specifically, we study a homogeneous environment where all sensors have identical noise covariance matrices and an inhomogeneous environment where the noise covariance matrices across the sensors have the same correlation structure but with different scaling factors. Given a total number of messages sent to the FC, theoretical lower bounds on the estimation error of any compression strategy are derived for both cases. Compression strategies are developed to approach or even attain the corresponding theoretical lower bounds. Performance analysis and simulations are carried out to illustrate the optimality and effectiveness of the proposed compression strategies.**

***Index Terms*—Distributed estimation, dimensionality reduction, wireless sensor network (WSN).**

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 [1], [2]. Power is a primary issue in sensor networks as the sensors constructing the network are powered by small batteries that are often irreplaceable in practice. Also, 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 bandwidth- and energy-efficient strategies for various sensor network processing tasks. A multitude of studies along this line have appeared recently in the context of distributed detection (e.g., [3]–[5]), low-rate quantization-based distributed estimation (e.g., [6]–[12]), distributed estimation using reduced dimensionality sensor observations (e.g., [13]–[18]), and others.

In this paper, we consider distributed estimation of a deterministic vector parameter, where the unknown vector parameter is observed by multiple sensors whose observations are processed and sent to a fusion center (FC) to reconstruct the unknown parameter. Vector parameters/observations arise from a variety of scenarios. For example, the sensor observation at each sensor can be signals collected from different time instances of a dynamic process (e.g., temperature measurements for different time points), or can be multi-modal signals regarding a target state (e.g., measurements of the location, speed, and bearing of a vehicle at a certain time). To meet the stringent bandwidth and power constraints inherent in WSNs, the high-dimensional sensor observation should be converted into low-dimensional data by carrying out local data dimensionality reduction. The problem of interest is to jointly design the compression matrices associated with those sensors such that the estimation error at the FC is minimized. Such a problem has been extensively investigated in a number of studies [13]–[22]. The first study possibly appeared in [13], where the authors considered dimension reduction and data fusion for a two-sensor case. The multi-sensor distributed compression-estimation problem was addressed later in [14], followed by [15]–[17]. In these works, the communication links between sensors and the FC are assumed ideal. This assumption was relaxed in [18], [22], where the link noise and the transmit power constraint were taken into account in designing the compression matrices. Despite all these efforts, optimum dimensionality reduction and the best achievable performance for the multi-sensor case are still an open problem. The most attractive solution [14], [16]–[19] so far is to employ a Gauss-Seidel iterative technique to reduce the number of optimization variables, which yields an iterative algorithm. An asymptotic distortion analysis of this iterative algorithm was provided in [21] in an infinite dimensional regime. Nevertheless, this algorithm is not guaranteed to converge to a global minimum, and it is unclear how close the stationary point to which the algorithm converges is to the global minimum. Another problem with these algorithms [14], [16]–[18] is that they require *a priori* knowledge of the compression dimension associated with each sensor. As different compression dimension assignments usually lead to different estimation performance, it is desirable to jointly consider the compression dimension assignment and the compression matrix design. Note that an iterative approach that can adaptively determine the compression dimensions was proposed in [19]. Nevertheless, this algorithm still suffers from local maxima

Manuscript received November 02, 2009; accepted April 02, 2010. Date of publication April 15, 2010; date of current version July 14, 2010. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Aleksander Dogandzic. This work was supported in part by the National Science Foundation under Grant CCF-0514938 and ECCS-0901066, and by the Air Force Office of Scientific Research under Grant FA9550-09-1-0310.

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

Color versions of one or more figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TSP.2010.2048213

and stationary points. In addition, due to its iterative nature, a performance evaluation is difficult to carry out, which, to some extent, may hinder its practical applications.

In this paper, we continue the efforts to investigate dimensionality reduction for distributed estimation. Unlike most existing works [13]–[22] modeling the unknown parameter as a random parameter, our study focuses on deterministic parameters. The extension of our results to random parameters will also be briefly discussed in this paper. We first develop an efficient iterative algorithm for a general noise scenario. We then focus on two specific but important noise scenarios: a homogeneous environment where all sensors have identical noise covariance matrices and an inhomogeneous environment where the noise covariance matrices across the sensors have the same correlation structure but with different scaling factors. The following questions are considered: given a specified number of messages sent to the FC, what is the minimum achievable estimation error, and how to jointly assign the compression dimension and design the compression matrices to approach or even attain this theoretical lower bound? These questions will be addressed to provide a fundamental understanding of dimensionality reduction for distributed estimation. Specifically, for a homogeneous environment, our results reveal that the rows of each compression matrix should be chosen from the eigenvectors of the noise covariance matrix, and the number of messages corresponding to a certain eigenvector should be proportional to the square root of the corresponding eigenvalue. Our performance analysis shows that the proposed compression strategy is very effective. In particular, when the noise covariance matrix has one or only a few dominant eigenvalues, it is even possible to achieve almost the same estimation performance as that of a centralized estimator using all original observations, while transmitting only $1/p$ (p denotes the dimension of the vector parameter) times the total number of messages required by the centralized estimator. For the inhomogeneous scenario, two compression strategies are developed: the first strategy is effective when the eigenvalues of the noise covariance matrix are diverse, while the second strategy achieves optimum/near-optimum performance for the case of identical or roughly identical eigenvalues. We note that a similar dimensionality reduction problem was studied within a vector quantization context in [23]. Nevertheless, the study [23] was confined to the case where each sensor compresses its vector observation into only a one-dimensional message, while the current work considers dimensionality reduction under a more general compression dimension assignment framework.

The following notations are adopted throughout this paper, where $[\cdot]^T$ stands for transpose, $\mathrm{tr}(\mathbf{A})$ denotes the trace of $\mathbf{A}$, and $\mathbf{A} \succeq \mathbf{0}$ means that the matrix is positive semidefinite. We let $[\mathbf{A}]_{ij}$ denote the (i, j)th entry of $\mathbf{A}, \mathbf{I}_m$ denote an $m \times m$ identity matrix. The symbols $\mathbb{R}^{n \times m}$ and $\mathbb{R}^n$ stand for the set of $n \times m$ matrices and the set of n-dimensional column vectors with real entries, respectively.

The rest of the paper is organized as follows. In Section II, we introduce the data model, basic assumptions, and the distributed compression-estimation problem. Section III presents an iterative algorithm for compression design for general noise scenarios. The compression design for the special cases (including a homogeneous case and an inhomogeneous case) is studied in Section IV. The generalization of the optimization and the extension to the random parameter case are discussed in V. Simulation results are provided in Section VI, followed by concluding remarks in Section VII.

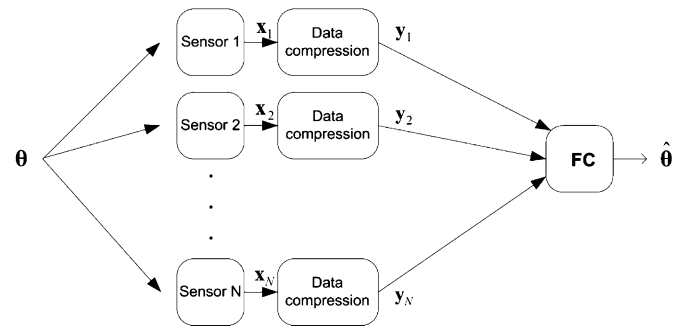

Fig. 1. Dimensionality reduction for distributed estimation: each sensor makes a noisy observation of the unknown parameter $\boldsymbol{\theta}$ and then convert its noisy observation $\mathbf{x}_n$ into some low dimensional data $\mathbf{y}_n$.

II. Problem Formulation

Consider a WSN consisting of N spatially distributed sensors. Each sensor makes a noisy observation of an unknown deterministic vector parameter $\boldsymbol{\theta} \in \mathbb{R}^p$

$$\mathbf{x}_n = \boldsymbol{\theta} + \mathbf{w}_n, \qquad n = 1, \ldots, N \tag{1}$$

where $\mathbf{w}_n \in \mathbb{R}^p$ denotes the additive noise with zero mean and full rank auto-covariance matrix $\mathbf{R}_{w_n}$. The noise is assumed uncorrelated across sensors and the knowledge of the noise covariance matrices is available at the FC. In the above model, the observation matrix $\mathbf{H}_n$ defining the input/output relation: $\mathbf{x}_n = \mathbf{H}_n\boldsymbol{\theta} + \mathbf{w}_n$, is assumed to be an identity matrix. Such a simplified model occurs in many practical applications, for example, the sensor observation is a collection of multiple snapshots of a dynamic process (temperature or humidity) of interest. The extension to the general linear model will be discussed in Section V.B.

To meet the stringent bandwidth/power budgets inherent in WSNs, dimensionality reduction is carried out at each sensor to convert the observation vector $\mathbf{x}_n$ into low-dimensional data $\mathbf{y}_n$ by using a linear compression matrix $\mathbf{B}_n \in \mathbb{R}^{q_n \times p} (q_n \leq p)$, i.e.

$$\mathbf{y}_n = \mathbf{B}_n\mathbf{x}_n \quad n = 1, \ldots, N. \tag{2}$$

These compressed data $\{\mathbf{y}_n\}_{n=1}^N$ are then sent to the FC to reconstruct the unknown parameter $\boldsymbol{\theta}$ (see Fig. 1). We adopt the following assumptions for the dimensionality reduction problem:

[A1] there is no intersensor communication;

[A2] the compressed messages $\{\mathbf{y}_n\}_{n=1}^N$ are sent to the FC without distortion;

[A3] the total average transmit power is proportional to the total number of compressed messages sent to the FC.

Remarks: In assumption (A2), we assume that the compressed messages can be reliably transmitted to the FC without any distortion. This assumption was also adopted in many other decentralized compression-estimation works, e.g., [14]–[19]. Assumption (A3) is a simple but reasonable assumption. In many practical scenarios, a centralized FC could be located far away from the deployed sensor field, and the distances between the sensors and the FC are roughly identical. The average energy spent by each sensor in transmitting a message reliably to the FC are roughly identical in a statistical sense.

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

$$\mathbf{y} = \mathbf{B}\mathbf{x} = \mathbf{B}\mathbf{J}\boldsymbol{\theta} + \mathbf{B}\mathbf{w} \tag{3}$$

where $\mathbf{B} \triangleq \mathrm{diag}\{\mathbf{B}_1, \dots, \mathbf{B}_N\}$ is a block diagonal matrix with its nth block-diagonal element equal to $\mathbf{B}_n, \mathbf{x} \triangleq [\mathbf{x}_1^T \ \mathbf{x}_2^T \ \dots \ \mathbf{x}_N^T]^T, \mathbf{w} \triangleq [\mathbf{w}_1^T \ \mathbf{w}_2^T \ \dots \ \mathbf{w}_N^T]^T, \mathbf{J} \triangleq \mathbf{1}_{N\times 1} \otimes \mathbf{I}_p$, in which $\mathbf{1}_{N\times 1}$ is an N-dimensional column vector with its all entries equal to one, $\otimes$ denotes the Kronecker product, and $\mathbf{I}_p$ denotes an $p \times p$ identity matrix. Using the received data $\mathbf{y}$, the best linear unbiased estimator (BLUE) is given by [24]

$$\begin{aligned}\hat{\boldsymbol{\theta}} &= [\mathbf{J}^T\mathbf{B}^T(\mathbf{B}\mathbf{R}_{\bar{\mathbf{w}}}\mathbf{B}^T)^{-1}\mathbf{B}\mathbf{J}]^{-1}\mathbf{J}^T\mathbf{B}^T(\mathbf{B}\mathbf{R}_{\bar{\mathbf{w}}}\mathbf{B}^T)^{-1}\mathbf{y} \\ &= \left[\sum_{n=1}^{N} \mathbf{B}_n^T \left(\mathbf{B}_n\mathbf{R}_{w_n}\mathbf{B}_n^T\right)^{-1}\mathbf{B}_n\right]^{-1} \\ &\quad \times \left[\sum_{n=1}^{N} \mathbf{B}_n^T \left(\mathbf{B}_n\mathbf{R}_{w_n}\mathbf{B}_n^T\right)^{-1}\mathbf{y}_n\right]\end{aligned} \tag{4}$$

where $\mathbf{R}_{\bar{\mathbf{w}}} \triangleq E[\mathbf{w}\mathbf{w}^T]$. The covariance matrix of $\hat{\boldsymbol{\theta}}$ is

$$\begin{aligned}\mathbf{R}_{\hat{\theta}} &\triangleq E[(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})^T] \\ &= [\mathbf{J}^T\mathbf{B}^T(\mathbf{B}\mathbf{R}_{\bar{\mathbf{w}}}\mathbf{B}^T)^{-1}\mathbf{B}\mathbf{J}]^{-1} \\ &= \left[\sum_{n=1}^{N} \mathbf{B}_n^T(\mathbf{B}_n\mathbf{R}_{w_n}\mathbf{B}_n^T)^{-1}\mathbf{B}_n\right]^{-1}\end{aligned} \tag{5}$$

with the variance of each component given by the corresponding diagonal element of $\mathbf{R}_{\hat{\theta}}$, i.e., $\mathrm{var}(\hat{\theta}_i) = [\mathbf{R}_{\hat{\theta}}]_{ii}$. For general linear models, if the observation noise is Gaussian, the BLUE is also the minimum variance unbiased (MVU) estimator and attains the Cramér-Rao bound. A natural question arising from the above scenario is to find out an overall optimum compression matrix $\mathbf{B}$, or equivalently, a set of individual compression matrices $\{\mathbf{B}_n\}_{n=1}^N$, to achieve a minimum overall estimation distortion at the FC. The optimization therefore can be formulated as follows:

$$\min_{\{\mathbf{B}_n\}} \mathrm{tr}\{\mathbf{R}_{\hat{\theta}}\} = \mathrm{tr}\left\{\left[\sum_{n=1}^{N} \mathbf{B}_n^T \left(\mathbf{B}_n\mathbf{R}_{w_n}\mathbf{B}_n^T\right)^{-1}\mathbf{B}_n\right]^{-1}\right\}. \tag{6}$$

Note that in the above optimization, we assume that the compression dimensions $\{q_n\}$ are specified *a priori*. This assumption will be relaxed later in Section IV and we will consider compression matrix design along with the compression dimension assignment. In the following, we will first develop an efficient iterative algorithm for compression design under a general noise scenario.

III. Compression Design: A General Case

The optimization (6) involves determining N compression matrices. To simplify the problem, we employ a Gauss-Seidel approach [25] to develop an iterative algorithm, where, at every iteration, we optimize the compression matrix for each sensor given the other compression matrices fixed. This iterative procedure has also been adopted in other works, e.g., [14], [18], to determine the compression matrices in a random parameter framework. As we will show later, an efficient iterative algorithm for the deterministic case can be readily developed based on previous results [14], [18].

Let $\mathbf{Q}_k \triangleq \sum_{n\neq k} \mathbf{B}_n^T(\mathbf{B}_n\mathbf{R}_{w_n}\mathbf{B}_n^T)^{-1}\mathbf{B}_n$, the optimization of $\mathbf{B}_k$ given fixed $\{\mathbf{B}_n\}_{n\neq k}$ is formulated as

$$\arg\min_{\mathbf{B}_k} \quad \mathrm{tr}\left\{\left(\mathbf{Q}_k + \mathbf{B}_k^T\left(\mathbf{B}_k\mathbf{R}_{w_k}\mathbf{B}_k^T\right)^{-1}\mathbf{B}_k\right)^{-1}\right\}. \tag{7}$$

Utilizing the Woodbury identity, the objective function of (7) can be rewritten as

$$\begin{aligned}&\mathrm{tr}\left\{\left(\mathbf{Q}_k + \mathbf{B}_k^T\left(\mathbf{B}_k\mathbf{R}_{w_k}\mathbf{B}_k^T\right)^{-1}\mathbf{B}_k\right)^{-1}\right\} \\ &\quad = \mathrm{tr}\left\{\mathbf{Q}_k^{-1} - \mathbf{Q}_k^{-1}\mathbf{B}_k^T \right. \\ &\qquad \left. \times \left[\mathbf{B}_k\left(\mathbf{R}_{w_k} + \mathbf{Q}_k^{-1}\right)\mathbf{B}_k^T\right]^{-1}\mathbf{B}_k\mathbf{Q}_k^{-1}\right\}.\end{aligned} \tag{8}$$

Since $\mathbf{Q}_k$ is fixed, (7) becomes

$$\max_{\mathbf{B}_k} \quad \mathrm{tr}\left\{\mathbf{Q}_k^{-1}\mathbf{B}_k^T\left[\mathbf{B}_k\left(\mathbf{R}_{w_k} + \mathbf{Q}_k^{-1}\right)\mathbf{B}_k^T\right]^{-1}\mathbf{B}_k\mathbf{Q}_k^{-1}\right\}. \tag{9}$$

The above optimization has been solved by previous works, e.g., [14], [18]. Its solution is summarized as follows. Let $\mathbf{G}_k \triangleq \mathbf{R}_{w,k} + \mathbf{Q}_k^{-1}$. The optimum solution to (9) is given by

$$\mathbf{B}_k = \mathbf{V}^T\mathbf{G}_k^{-\frac{1}{2}} \tag{10}$$

where $\mathbf{V} \in \mathbb{R}^{p\times q_k}$ is obtained as the eigenvectors corresponding to the q_k largest eigenvalues of $\mathbf{G}_k^{-(1/2)}\mathbf{Q}_k^{-1}\mathbf{Q}_k^{-1}\mathbf{G}_k^{-(1/2)}$.

Based on the above results, we can establish an iterative algorithm by successively optimizing and replacing each compression matrix $\mathbf{B}_k$. The algorithm is summarized as follows.

1) Randomly generate a set of compression matrices $\{\mathbf{B}_n^{(0)}\}$ as an initialization.
2) At iteration $i+1$ ($i = 0, 1, \dots$): determine $\mathbf{B}_1^{(i+1)}$ given: $\{\mathbf{B}_2^{(i)}, \dots, \mathbf{B}_N^{(i)}\}$; determine $\mathbf{B}_k^{(i+1)}$ given: $\{\mathbf{B}_1^{(i+1)}, \dots, \mathbf{B}_{k-1}^{(i+1)}, \mathbf{B}_{k+1}^{(i)}, \dots, \mathbf{B}_N^{(i)}\}$ for $k = 2, \dots, N$.
3) Go to Step 2) if $|f(\{\mathbf{B}_n^{(i+1)}\}) - f(\{\mathbf{B}_n^{(i)}\})| > \epsilon$, where $f(\cdot)$ denotes the objective function defined in (6), and ϵ is a prescribed tolerance value; otherwise stop.

Clearly, in this algorithm, every iteration results in a nonincreasing objective function value. In this manner, the iterative algorithm converges to a stationary point and finds an effective set of compression matrices. Nevertheless, this algorithm is not guaranteed to converge to the global minimum, and it is unclear how close the achieved stationary point is to the global minimum. Moreover, for this iterative algorithm, the compression dimension, q_k, associated with each sensor needs to be known *a priori*. Given the total number of compressed messages, finding the compression dimension for each sensor is in fact a resource allocation problem. It is desirable to integrate this resource allocation problem into the compression design framework.

IV. Compression Design: Special Cases

In this section, we study two important specific scenarios, namely, a homogeneous environment where the sensors have identical observation noise covariance matrices and an inhomogeneous environment where the noise covariance matrices have the same correlation structure with different scaling factors. These two scenarios can be modeled as $\mathbf{R}_{w_n} = \sigma_{w,n}^2\mathbf{R}_w, \forall n$ with identical or nonidentical scaling factors. Many applications can be characterized by this specific noise model. For example, in many cases, due to the underlying physical sensing mechanism, the sensor observation is a linear or nonlinear function of the signal of interest, i.e., $\tilde{\mathbf{x}}_n = \mathbf{f}(\boldsymbol{\theta}) + \tilde{\mathbf{w}}_n$, where $\mathbf{f}(\cdot)$ is a known linear or nonlinear vector function (for example, $\mathbf{f}(\cdot)$ can be a linear function with each component of $\boldsymbol{\theta}$ amplified with a different amplification factor), and $\tilde{\mathbf{w}}_n$ is usually additive measurement noise with zero mean and covariance matrix $\sigma_{w_n}^2\mathbf{I}$. By carrying out the inverse of $\mathbf{f}(\cdot)$ or resorting to a local estimator of $\boldsymbol{\theta}$, the general observation model can be reduced to a simplified model: $\mathbf{x}_n = \boldsymbol{\theta} + \mathbf{w}_n$, with the noise having the same noise correlation structure (which is introduced by the function inverse) but identical or different scaling factors. The optimization can be formulated as

$$\begin{aligned} \min_{\{\mathbf{B}_n\},\{q_n\}} \quad & \mathrm{tr}\left\{\left[\sum_{n=1}^{N}\frac{1}{\sigma_{w,n}^2}\mathbf{B}_n^T\left(\mathbf{B}_n\mathbf{R}_w\mathbf{B}_n^T\right)^{-1}\mathbf{B}_n\right]^{-1}\right\} \\ \text{s.t.} \quad & \sum_{n=1}^{N} q_n = K \\ \text{where} \quad & \mathbf{B}_n \in \mathbb{R}^{q_n\times p} \quad \forall n. \end{aligned} \tag{11}$$

Note that unlike (6) assuming *a priori* specified compression dimension assignment, in the above optimization, the compression dimensions $\{q_n\}$ associated with the compression matrices $\{\mathbf{B}_n\}$ are also variables to be optimized. The constraint on the total number of compressed messages is equivalent to placing a total transmit power constraint since we assume that the total average transmit power is proportional to the total number of compressed messages sent to the FC (see (A3)). To deal with (11), we first carry out the following simplifications. Let $\mathbf{R}_w = \mathbf{U}_w\mathbf{D}_w\mathbf{U}_w^T$ denote the eigenvalue decomposition (EVD) and $\bar{\mathbf{C}}_n \triangleq \mathbf{B}_n\mathbf{U}_w\mathbf{D}_w^{(1/2)}$. We have

$$\begin{aligned} &\mathbf{B}_n^T\left(\mathbf{B}_n\mathbf{R}_w\mathbf{B}_n^T\right)^{-1}\mathbf{B}_n \\ &\qquad = \mathbf{U}_w\mathbf{D}_w^{-\frac{1}{2}}\bar{\mathbf{C}}_n^T\left(\bar{\mathbf{C}}_n\bar{\mathbf{C}}_n^T\right)^{-1}\bar{\mathbf{C}}_n\mathbf{D}_w^{-\frac{1}{2}}\mathbf{U}_w^T. \end{aligned} \tag{12}$$

Furthermore, we can write $\bar{\mathbf{C}}_n = \mathbf{P}_n\mathbf{C}_n$, where $\mathbf{P}_n \in \mathbb{R}^{q_n\times q_n}$ is a full rank matrix and $\mathbf{C}_n \in \mathbb{R}^{q_n\times p}$ consists of q_n orthonormal rows, i.e., $\mathbf{C}_n\mathbf{C}_n^T = \mathbf{I}_{q_n}$ (note that without loss of generality, we assume $\bar{\mathbf{C}}_n$ is a full row rank matrix, i.e., the compression matrix has q_n independent rows; otherwise the messages sent to the FC contain redundant information, which wastes the communication resource and is obviously undesirable). Substituting $\bar{\mathbf{C}}_n = \mathbf{P}_n\mathbf{C}_n$ into (12), we arrive at

$$\begin{aligned} &\mathbf{B}_n^T\left(\mathbf{B}_n\mathbf{R}_w\mathbf{B}_n^T\right)^{-1}\mathbf{B}_n \\ &\quad = \mathbf{U}_w\mathbf{D}_w^{-\frac{1}{2}}\mathbf{C}_n^T\left(\mathbf{C}_n\mathbf{C}_n^T\right)^{-1}\mathbf{C}_n\mathbf{D}_w^{-\frac{1}{2}}\mathbf{U}_w^T \\ &\quad = \mathbf{U}_w\mathbf{D}_w^{-\frac{1}{2}}\mathbf{C}_n^T\mathbf{C}_n\mathbf{D}_w^{-\frac{1}{2}}\mathbf{U}_w^T. \end{aligned} \tag{13}$$

As we can see, the expression on the right-hand side of (13) circumvents the inverse of a variable matrix and is much easier for our following development. Therefore we shall focus our study on the design of these newly constructed compression matrices $\{\mathbf{C}_n\}$. The original compression matrices $\{\mathbf{B}_n\}$ can be easily recovered from the relationship

$$\mathbf{B}_n = \mathbf{P}_n\mathbf{C}_n\mathbf{D}_w^{-\frac{1}{2}}\mathbf{U}_w^T, \tag{14}$$

where $\mathbf{P}_n$ can be any arbitrary full rank matrix as it can be readily verified that the objective function value is independent of the choice of $\mathbf{P}_n$. By substituting (13) into (11), and using the trace identity $\mathrm{tr}(\mathbf{A}_1\mathbf{A}_2) = \mathrm{tr}(\mathbf{A}_2\mathbf{A}_1)$, the optimization (11) can be reformulated as

$$\begin{aligned} \min_{\{\mathbf{C}_n\},\{q_n\}} \quad & \mathrm{tr}\left\{\left[\sum_{n=1}^{N}\frac{1}{\sigma_{w,n}^2}\mathbf{C}_n^T\mathbf{C}_n\right]^{-1}\mathbf{D}_w\right\} \\ \text{s.t.} \quad & \mathbf{C}_n\mathbf{C}_n^T = \mathbf{I}_{q_n} \quad \forall n \in \{1,\ldots,N\} \\ & \sum_{n=1}^{N} q_n = K. \end{aligned} \tag{15}$$

The optimization (15) involves a joint search over compression dimensions and compression matrices, which is complicated for analysis. To gain an insight into (15), we first develop a theoretical lower bound on the estimation error of any compression strategy associated with a specific choice of $\{q_n\}$, that is, a lower bound on the minimum achievable objective function value of the following nonconvex optimization for a given $\{q_n\}$ ($\{q_n\}$ satisfies the constraint $\sum_{n=1}^{N} q_n = K$):

$$\begin{aligned} \min_{\{\mathbf{C}_n\}} \quad & \mathrm{tr}\left\{\left[\sum_{n=1}^{N}\frac{1}{\sigma_{w,n}^2}\mathbf{C}_n^T\mathbf{C}_n\right]^{-1}\mathbf{D}_w\right\} \\ \text{s.t.} \quad & \mathbf{C}_n\mathbf{C}_n^T = \mathbf{I}_{q_n} \quad \forall n \in \{1,\ldots,N\}. \end{aligned} \tag{16}$$

A. Lower Bound on the Estimation Error

The results regarding the lower bound on the estimation error of any compression strategy for a specified $\{q_n\}$ are summarized as follows.

Lemma 1: Consider the case where the noise covariance matrices across the sensors have the same correlation structure, i.e., $\mathbf{R}_{w_n} = \sigma_{w,n}^2\mathbf{R}_w$. For each specific compression dimension assignment $\{q_n\}_{n=1}^N$, the estimation error of any compression strategy is lower bounded by

$$\mathrm{tr}(\mathbf{R}_{\hat{\theta}}) \geq \frac{1}{L}\left[\sum_{i=1}^{p}\sqrt{\lambda_{w,i}}\right]^2 \tag{17}$$

where $\lambda_{w,i}$ denotes the ith eigenvalue of $\mathbf{R}_w$, i.e., the ith diagonal element of $\mathbf{D}_w$, and L is defined as

$$L \triangleq \sum_{n=1}^{N} \frac{1}{\sigma_{w,n}^2} q_n. \tag{18}$$

This lower bound can be attained if the compression matrices $\{\mathbf{C}_n\}$ satisfy the constraints $\mathbf{C}_n\mathbf{C}_n^T = \mathbf{I}_{q_n}, \forall n$ and also the following condition

$$\mathbf{C}^T\mathbf{D}_\sigma\mathbf{C} = \mathbf{D}^* \triangleq \text{diag}(d_1^*, \dots, d_p^*), \tag{19}$$

in which $\mathbf{C} \in \mathbb{R}^{K\times p}$ is a matrix formed by stacking the N compression matrices

$$\mathbf{C} \triangleq [\,\mathbf{C}_1^T \quad \mathbf{C}_2^T \quad \dots \quad \mathbf{C}_N^T\,]^T \tag{20}$$

$\mathbf{D}_\sigma$ denotes a block diagonal matrix with its nth block equal to $(1/\sigma_{w,n}^2)\mathbf{I}_{q_n}$

$$\mathbf{D}_\sigma \triangleq \text{diag}\left(\frac{1}{\sigma_{w,1}^2}\mathbf{I}_{q_1}, \frac{1}{\sigma_{w,2}^2}\mathbf{I}_{q_2}, \dots, \frac{1}{\sigma_{w,N}^2}\mathbf{I}_{q_N}\right) \tag{21}$$

and d_i^* is defined as

$$d_i^* \triangleq \frac{L\sqrt{\lambda_{w,i}}}{\sum_{i=1}^{p}\sqrt{\lambda_{w,i}}}. \tag{22}$$

Proof: See Appendix A. ■

For each compression dimension assignment $\{q_n\}$, Lemma 1 not only reveals the best achievable performance of any compression strategy, but also provides the conditions under which the compression design can approach or even attain the corresponding theoretical lower bound. We note that the derived lower bound is inversely proportional to L, which is a value generally dependent on the compression dimension assignment $\{q_n\}$ (exception occurs for the homogeneous case due to identical observation qualities). Hence the derived lower bound (17) is a lower bound for a specific compression dimension assignment, but in general not a universal lower bound that applies to all feasible compression dimension assignments. Nevertheless, as will be shown later, the results are still helpful and shed light on how to choose effective compression dimension assignment and design the compression matrices. In the following, for presentation clarity, a homogeneous case is first considered, then followed by an inhomogeneous case with nonidentical scaling factors.

B. A Homogeneous Case

Clearly, for the homogeneous case $\mathbf{R}_{w_n} = \mathbf{R}_w, \forall n$ (assuming $\sigma_{w,n}^2 = 1$ without loss of generality), the parameter L is equivalent to $L = \sum_{n=1}^{N} q_n = K$, which is a constant independent of the compression dimension assignment. Therefore the theoretical lower bound (17) derived specifically for a given compression dimension assignment in fact becomes a universal lower bound. The results are summarized as follows.

Theorem 1: Consider a homogeneous environment where all sensors have identical noise covariance matrix $\mathbf{R}_w$. Suppose that the number of total compressed messages sent to the FC is K. Then for any compression dimension assignment $\{q_n\}$, the estimation error of any compression strategy is lower bounded by

$$\text{tr}\left\{\mathbf{R}_{\hat{\theta}}\right\} \geq \frac{1}{K}\left[\sum_{i=1}^{p}\sqrt{\lambda_{w,i}}\right]^2. \tag{23}$$

Proof: The results come directly from Lemma 1 by noting that $\sigma_{w,n}^2 = 1, \forall n$. ■

To approach this universal theoretical lower bound, we propose the following compression strategy which uses any K of N sensors (implicitly we assume $N \geq K$), with each sensor compressing its observation into only one message. Let

$$l_i \triangleq \frac{K\sqrt{\lambda_{w,i}}}{\sum_{i=1}^{p}\sqrt{\lambda_{w,i}}} \qquad \forall i \in \{1, \dots, p\}. \tag{24}$$

The details of the compression strategy are as follows.

- *Proposed compression strategy:* We divide the K sensors into p groups, with the ith group consisting of f_i sensors, where f_i is obtained by rounding l_i to its nearest integer while still preserving their summation, i.e., $\sum_{i=1}^{p} f_i = \sum_{i=1}^{p} l_i = K$. All sensors in the ith group choosing $\mathbf{e}_i$ to be their compression vectors, i.e.

$$\mathbf{C}_n = \mathbf{e}_i^T \quad \forall n = i_1, i_2, \dots, i_{f_i} \tag{25}$$

  where $\mathbf{e}_i \in \mathbb{R}^p$ is a unit column vector with its ith entry equal to one, whereas other entries equal to zero; $\{i_1, i_2, \dots, i_{f_i}\}$ are the indices of the sensors in group i.

Since each compression vector $\mathbf{C}_n$ is selected to be a unit row vector, the constraints $\mathbf{C}_n\mathbf{C}_n^T = \mathbf{I}_{q_n}, \forall n$ imposed on the compression matrices $\{\mathbf{C}_n\}$ in (15) are automatically satisfied. We have the following easily verified result regarding the proposed compression strategy.

Proposition 1: The estimation error achieved by the above proposed compression strategy is given by

$$\text{tr}\{\mathbf{R}_{\hat{\theta}}\} = \sum_{i=1}^{p}\frac{\lambda_{w,i}}{f_i}. \tag{26}$$

If $\{l_i\}$ have integer values, i.e., $f_i = l_i$, then the achieved estimation error attains the universal lower bound (23).

Remark 1: The assumption that $\{l_i\}$ are integers may not be satisfied in practice. In this case, the universal lower bound cannot be attained. Nevertheless, the performance degradation caused by the rounding operation is usually very mild due to the smoothness of the estimation error cost function (26), which is also corroborated by our simulation results. In particular, since l_i is proportional to K, as K increases, the estimation error (26) can be arbitrarily close to the universal lower bound (23).

Remark 2: With (25), the original compression matrix $\mathbf{B}_n$ can be recovered from (14), i.e., $\mathbf{B}_n = \mathbf{P}_n\mathbf{C}_n\mathbf{D}_w^{-(1/2)}\mathbf{U}_w^T$, where $\mathbf{P}_n$ is a scalar as $q_n = 1$. Since $\mathbf{C}_n$ is a unit row vector, we see that $\mathbf{B}_n$ is a scaled eigenvector of the noise covariance matrix $\mathbf{R}_w$. Also, we observe that the number of messages/sensors corresponding to a certain eigenvector should be proportional to the square root of the corresponding eigenvalue (see (24)). This result has an intuitive explanation. Note that a larger eigenvalue indicates that the corresponding eigenvector direction has a larger amount of noise power. Intuitively, to minimize the estimation error over all directions, more resources should be assigned to the directions with larger amounts of noise power because this can lead to a more significant overall estimation error decrease as compared with putting the same amount of resources to the directions with less noise power. This judicious resource allocation renders a power-distortion efficiency, as we will discuss later. Moreover, we see that this resource allocation is closely tied to the criterion of minimizing the overall estimation error. Other criteria may lead to other resource allocation schemes.

Remark 3: Since sensors have identical observation qualities, messages corresponding to a same eigenvector but coming from different sensors make no difference. Also, bear in mind that each sensor can compress its observation into multiple compressed messages, as long as these messages correspond to different eigenvectors. Therefore we can find other compression strategies which need fewer number of sensors (less than K) but have the same number of messages corresponding to each eigenvector as that of the current strategy.

Comparison With the Non-Compression Estimator: It is also interesting to examine the performance of the proposed compression strategy as compared with a BLUE estimator using $T(T \leq N)$ sensors' original observations $\{\mathbf{x}_n\}_{n=1}^{T}$. Clearly, if $T = N$, the BLUE estimator is the centralized estimator which provides a benchmark on the performance of all rate constrained methods. It can be easily verified that the estimation error of the BLUE estimator is given by

$$\mathrm{tr}\{\mathbf{R}_{\hat{\theta},\mathrm{NC}}\} = \mathrm{tr}\left\{\left[\sum_{n=1}^{T}\mathbf{R}_w^{-1}\right]^{-1}\right\} = \frac{1}{T}\sum_{i=1}^{p}\lambda_{w,i}, \tag{27}$$

where the subscript 'NC' stands for a estimator with no data compression. Since the proposed compression strategy approaches the universal lower bound (23), the ratio of the estimation error of the proposed compression strategy to that of this noncompression BLUE estimator is given by

$$\gamma \triangleq \frac{\mathrm{tr}\{\mathbf{R}_{\hat{\theta}}\}}{\mathrm{tr}\{\mathbf{R}_{\hat{\theta},\mathrm{NC}}\}} \rightarrow \frac{T\left(\sum_{i=1}^{p}\sqrt{\lambda_{w,i}}\right)^2}{K\left(\sum_{i=1}^{p}\lambda_{w,i}\right)} = \frac{T}{K}\eta \tag{28}$$

where

$$\eta \triangleq \frac{\left(\sum_{i=1}^{p}\sqrt{\lambda_{w,i}}\right)^2}{\sum_{i=1}^{p}\lambda_{w,i}}. \tag{29}$$

Using the Cauchy-Schwarz inequality, we can show that η is a value lower and upper bounded by: $1 < \eta \leq p$. The upper bound p is reached when all the eigenvalues $\{\lambda_{w,i}\}$ are identical. On the other hand, if the eigenvalues are diverse, then η tends towards its lower bound 1. We see that in order to achieve the same estimation performance as that of the noncompression BLUE estimator, the proposed compression strategy requires to send a total number of $K = T\eta$ messages, which is η/p times the total number of messages needed by the noncompression estimator. Since $(\eta/p) \leq 1$, our compression scheme is generally more efficient than the noncompression estimator as it requires fewer messages to meet a distortion target. In particular, if the noise covariance matrix have one or only a few dominant eigenvalues, η approaches its lower bound 1, meaning that we can attain almost the same estimation performance as that of a noncompression estimator by sending only $1/p$ times the number of messages required by the noncompression estimator. The above result can also be perceived from a power perspective. From Assumption A3), we can say that our compression scheme is more efficient in a power-distortion sense.

C. An Inhomogeneous Case

From Lemma 1, we know that for the inhomogeneous case, the derived theoretical lower bound (17) is inversely proportional to L, which is a value dependent on the compression dimension assignment $\{q_n\}$. This is different from the homogeneous case where the lower bound is only dependent on the total number of compressed messages K. Naturally one may wish to increase L (consequently decrease the lower bound (17)) by assigning more compression resources, i.e., larger q_n, to those sensors with better observation qualities. However, it is generally difficult for us to find a compression strategy to approach the corresponding lower bound for such an intuitive compression dimension assignment.

In the following, we consider the compression design under two different compression dimension assignment schemes. In the first scheme, each sensor compresses its observation into only one message, irrespective of the disparity of the sensors' observation qualities. For the second scheme, the observation qualities are taken into account and only a small number of sensors are selected to transmit their data. Two different universal lower bounds are developed to evaluate these two proposed compression strategies, respectively. Our analysis reveals that the first compression strategy is effective when the eigenvalues of the noise covariance matrix are diverse, while the second compression strategy is able to achieve optimum/near optimum performance for the case of identical or roughly identical eigenvalues.

1) Proposed Compression Strategy I: Without loss of generality, we assume that the scaling factors $\{\sigma_{w,n}^2\}$ are in ascending order. We first introduce the following theorem which provides a universal lower bound on the estimation error of any compression strategy.

Theorem 2: Suppose that the number of total compressed messages sent to the FC is K, where $K \leq N$. Then for any compression dimension assignment $\{q_n\}$, the estimation error of any compression strategy is lower bounded by

$$\mathrm{tr}(\mathbf{R}_{\hat{\theta}}) \geq \frac{1}{J_1}\sum_{i=1}^{p}\lambda_{w,i} \triangleq \mathrm{ULB_I} \tag{30}$$

where

$$J_1 \triangleq \sum_{n=1}^{K}\frac{1}{\sigma_{w,n}^2}. \tag{31}$$

Proof: To prove (30), we construct a BLUE estimator which has access to the first K sensors' original observations $\{\mathbf{x}_n\}_{n=1}^{K}$ (note that the first K sensors have the best K observation qualities since $\{\sigma_{w,n}^2\}$ are in ascending order). The estimation error of this estimator is given by [24]

$$\mathrm{tr}\{\mathbf{R}_{\hat{\theta},\mathrm{NC}}\} = \mathrm{tr}\left\{\left[\sum_{n=1}^{K}\frac{1}{\sigma_{w,n}^2}\mathbf{R}_w^{-1}\right]^{-1}\right\} = \frac{1}{J_1}\sum_{i=1}^{p}\lambda_{w,i}. \tag{32}$$

Now consider any compression strategy which sends K compressed messages to the FC. Since these K messages come from at most K sensors, the above constructed estimator provides a benchmark (lower bound) on the performance of any K-message constrained method. ■

We now introduce the following compression strategy which can approach the universal lower bound (30) under certain circumstances.

- *Proposed strategy*: Suppose that the total number of compressed messages sent to the FC is K. We select the first K sensors with the best observation qualities and divide these K sensors into p groups (the partition principle will be discussed below). Let the sensors in the ith group choose $\mathbf{e}_i$ to be their compression vector, i.e.

$$\mathbf{C}_n = \mathbf{e}_i^T \quad \forall n = i_1, i_2, \ldots, i_{K_i} \tag{33}$$

where $\{i_1, i_2, \ldots, i_{K_i}\}$ are the indices of sensors in group i, and K_i denotes the number of sensors of group i.

For notational convenience, define

$$\chi_i \triangleq \sum_{n=i_1}^{i_{K_i}} \frac{1}{\sigma_{w,n}^2} \quad \forall i \in \{1, \ldots, p\}. \tag{34}$$

It can be easily verified that the estimation error of the proposed compression strategy is given by

$$\text{tr}(\mathbf{R}_{\hat{\theta}}) = \text{tr}\left\{\left[\sum_{n=1}^{K} \frac{1}{\sigma_{w,n}^2} \mathbf{C}_n^T \mathbf{C}_n\right]^{-1} \mathbf{D}_w\right\} = \sum_{i=1}^{p} \frac{\lambda_{w,i}}{\chi_i} \tag{35}$$

from which we can see that the estimation error is a function of $\{\chi_i\}$, and consequently dependent on the sensor partition. We have the following results (which can be easily proved) regarding the proposed compression strategy.

Proposition 2: The proposed compression strategy achieves its minimum estimation error when

$$\chi_i = \frac{J_1 \sqrt{\lambda_{w,i}}}{\sum_{i=1}^{p} \sqrt{\lambda_{w,i}}} \quad \forall i \in \{1, \ldots, p\} \tag{36}$$

and the minimum estimation error is

$$\text{tr}(\mathbf{R}_{\hat{\theta}}) = \frac{1}{J_1} \left(\sum_{i=1}^{p} \sqrt{\lambda_{w,i}}\right)^2 \tag{37}$$

which is exactly the theoretical lower bound (17) specific to the compression dimension assignment $q_n = 1, \forall n \in \{1, \ldots, K\}$. Also, the ratio of the minimum estimation error of the compression strategy to the universal lower bound (30) is given by

$$\gamma_1 \triangleq \frac{\text{tr}\{\mathbf{R}_{\hat{\theta}}\}}{\text{ULB}_\text{I}} = \eta \tag{38}$$

where η is defined in (29).

Remark 1: Clearly, in order to achieve a smaller estimation error, the K sensors should be properly partitioned into p groups. Specifically, if we can find a partition such that $\{\chi_i\}$ satisfy (36), the proposed compression strategy achieves the minimum estimation error (37). Of course, finding a partition exactly satisfying (36) may not be possible in practice. However, finding a set $\{\chi_i\}$ roughly equivalent to their optimal values is not that difficult when $K \gg p$, in which case the proposed strategy approaches the minimum estimation error (37).

Remark 2: Recalling that η is a value lower and upper bounded by: $1 < \eta \leq p$, and tends towards its lower bound 1 if the eigenvalues are diverse. Therefore the proposed compression strategy is effective and should be used for the case of diverse eigenvalues $\{\lambda_{w,i}\}$. In particular, when the noise covariance matrix have one or only a few dominant eigenvalues, it is even possible to achieve almost the same estimation performance as the universal lower bound, i.e., the performance of a noncompression estimator with access to all the best K sensors' original observations.

2) Proposed Compression Strategy II: The proposed compression strategy I is effective when the eigenvalues of the noise covariance matrix are diverse. For the case where the eigenvalues are identical or roughly identical, we propose the following strategy. Before proceeding, we first introduce another universal lower bound which is obtained by maximizing L in (17).

Theorem 3: Suppose that the number of total compressed messages sent to the FC is K, where we write $K = Mp + k, M = \text{floor}(K/p)$ is an integer and $k < p$. Then for any compression dimension assignment $\{q_n\}$ and compression strategy, the estimation error is lower bounded by

$$\text{tr}(\mathbf{R}_{\hat{\theta}}) \geq \frac{1}{L_{\max}} \left(\sum_{i=1}^{p} \sqrt{\lambda_{w,i}}\right)^2 \triangleq \text{ULB}_\text{II} \tag{39}$$

where

$$L_{\max} = \sum_{n=1}^{M} \frac{p}{\sigma_{w,n}^2} + \frac{k}{\sigma_{w,M+1}^2}. \tag{40}$$

Proof: From (18), L is maximized when the compression resource is allocated as follows: $q_n = p$ for $n \in \{1, \ldots, M\}, q_n = k$ for $n = M + 1$, and $q_n = 0$ otherwise (note that q_n cannot be greater than p). The maximum L is then given by (40). Substituting (40) into (17), the estimation error of any compression strategy therefore is lower bounded by (39). ■

The universal lower bound (39) is in fact the lower bound (17) specific to the following particular compression dimension assignment: $q_n = p$ for $n \in \{1, \ldots, M\}, q_n = k$ for $n = M + 1$, and $q_n = 0$ otherwise. For such an intuitive compression dimension assignment, it is generally difficult to find a compression strategy to approach the corresponding lower bound. Nevertheless, when the eigenvalues of the noise covariance matrix are identical, the following compression strategy is able to approach or attain the universal lower bound (39).

- *Proposed strategy:* Suppose $K = Mp + k$. The first $M + 1$ sensors are selected to transmit their data. Specifically, the compression matrices $\{\mathbf{C}_n\}$ are designed as follows:

$$\mathbf{C}_n = \begin{cases} \mathbf{I}_p, & \text{if } n \in \{1, \ldots, M\} \\ \mathbf{I}_p[1:k, :], & \text{if } n = M + 1 \\ \mathbf{0}, & \text{otherwise} \end{cases} \tag{41}$$

where $\mathbf{I}_p[1:k, :]$ denotes a submatrix consisting of rows 1 thru k of $\mathbf{I}_p$.

Clearly, the constraints $\mathbf{C}_n\mathbf{C}_n^T = \mathbf{I}_{q_n}, \forall n$ imposed on the compression matrices $\{\mathbf{C}_n\}$ in (15) are satisfied with the above choice (note that only those sensors which transmit need to be considered).

Proposition 3: The estimation error achieved by the proposed compression strategy is approximately given by

$$\mathrm{tr}(\mathbf{R}_{\hat{\theta}}) \approx \mathrm{tr}\left\{\left[\sum_{n=1}^{M}\frac{1}{\sigma_{w,n}^2}\mathbf{C}_n^T\mathbf{C}_n\right]^{-1}\mathbf{D}_w\right\} = \frac{1}{J_2}\left(\sum_{i=1}^{p}\lambda_{w,i}\right) \quad (42)$$

where the approximation becomes an equality when $K = Mp$, i.e., $k = 0$, and J_2 is defined as (note that $M = \mathrm{floor}(K/p)$)

$$J_2 \triangleq \sum_{n=1}^{M}\frac{1}{\sigma_{w,n}^2}. \quad (43)$$

The ratio of the estimation error of the proposed compression strategy to the universal lower bound (39) is

$$\gamma_2 \approx \frac{L_{\max}}{J_2}\frac{\sum_{i=1}^{p}\lambda_{w,i}}{\left(\sum_{i=1}^{p}\sqrt{\lambda_{w,i}}\right)^2} = \frac{L_{\max}}{J_2}\frac{1}{\eta} \approx \frac{p}{\eta} \quad (44)$$

where the above approximations become an identity when $K = Mp$.

Remark 1: When all eigenvalues $\{\lambda_{w,i}\}$ are identical, $(1/\eta) = (1/p)$, therefore we have $\gamma_2 \approx 1$ (or $\gamma_2 = 1$), suggesting that the estimation error is approximately (or exactly) identical to the universal lower bound (39). The original compression matrix $\mathbf{B}_n$ can be easily recovered from (14) and (41). It can be easily verified that $\mathbf{B}_n, \forall n \in \{1,\ldots,M\}$ can in fact be any full rank matrix.

3) Discussions of Two Proposed Compression Strategies: Note that both lower bounds (30) and (39) provide a benchmark (lower bound) on the achievable performance of any compression strategy. The reason for us to propose two different universal lower bounds is that both universal lower bounds are not always tight. In fact, it can be readily verified that the latter lower bound (39) is tighter than the former one when the eigenvalues are identical, i.e., $\mathrm{ULB}_{\mathrm{II}} \geq \mathrm{ULB}_{\mathrm{I}}$. On the other hand, for the case where the eigenvalues are diverse, the former lower bound (30) could be tighter than the latter one. In general, the compression Strategy I should be used for the case of diverse eigenvalues, while Strategy II is near-optimal/optimal for the case of identical or roughly identical eigenvalues. Of course, with the information of the eigenvalues and the scaling factors, a more accurate decision can be made from (37) and (42) to determine which compression strategy yields a lower estimation error.

From (38) and (44), we can show that the ratio of the lower estimation error of the two strategies to the tighter universal lower bound is smaller than or equal to the minimum of η and p/η, i.e.

$$\frac{\min(\mathrm{tr}(\mathbf{R}_{\hat{\theta},\mathrm{I}}), \mathrm{tr}(\mathbf{R}_{\hat{\theta},\mathrm{II}}))}{\max(\mathrm{ULB}_{\mathrm{I}}, \mathrm{ULB}_{\mathrm{II}})} \leq \min\left(\eta, \frac{p}{\eta}\right) \quad (45)$$

where for clarity, we use $\mathrm{tr}(\mathbf{R}_{\hat{\theta},\mathrm{I}})$ and $\mathrm{tr}(\mathbf{R}_{\hat{\theta},\mathrm{II}})$ to denote the estimation error of Strategy I and II, respectively. Since $1 < \eta \leq p$, the maximum value of the function $\min(\eta, (p/\eta))$ is $\sqrt{p}$, i.e., $\max_\eta \min(\eta, (p/\eta)) = \sqrt{p}$. Therefore we can guarantee that, for any set of eigenvalues, the best strategy out of the proposed two strategies can achieve an estimation error within a factor $\sqrt{p}$ of the tighter universal lower bound.

We provide an intuitive explanation for the effectiveness of the proposed compression strategies. If the eigenvalues are diverse, most noise power is concentrated on a small number of eigenvector directions. Assume a limiting case where there are $p-1$ zero eigenvalues and one nonzero eigenvalue. When $\boldsymbol{\theta}$ is projected on the $p-1$ eigenvectors associated with the null eigenvalues, we have a perfect estimate for these $p-1$ projections. Therefore, the best strategy is to let each sensor transmit one message, with $p-1$ sensors to provide perfect reconstructions of these $p-1$ projections, while other sensors are used to reconstruct the projection onto the noisy eigenvector direction. Transmitting multiple messages for any sensor may not be a good idea since it may entail more than necessary messages to reconstruct these $p-1$ projections onto the $p-1$ noiseless eigenvectors (note that multiple messages coming from each sensor have to correspond to different eigenvectors). For the case where the eigenvalues are identical, the noise power is evenly distributed among all eigenvector directions. Therefore the best strategy is to let sensors with the best observation qualities report the projections along all directions.

V. Discussions

A. Generalization of the Optimization (6)

Note that in the optimization (6), the objective is to minimize the overall estimation distortion of the vector parameter. In certain practical applications, we may wish to place more emphasis on some components of $\boldsymbol{\theta}$. With this in mind, we can generalize the optimization criterion (6) to the following:

$$\min_{\{\mathbf{B}_n\}} \mathrm{tr}\{\mathbf{R}_{\hat{\theta}}\boldsymbol{\Phi}\} = \mathrm{tr}\left\{\left[\sum_{n=1}^{N}\mathbf{B}_n^T\left(\mathbf{B}_n\mathbf{R}_{w_n}\mathbf{B}_n^T\right)^{-1}\mathbf{B}_n\right]^{-1}\boldsymbol{\Phi}\right\} \quad (46)$$

where $\boldsymbol{\Phi} \triangleq \mathrm{diag}(\phi_1, \phi_2, \ldots, \phi_p)$ is a diagonal matrix and ϕ_i is a positive weighting factor of user choice. The optimization (46) can be rewritten as

$$\min_{\{\mathbf{B}_n\}} \mathrm{tr}\left\{\left[\sum_{n=1}^{N}\boldsymbol{\Phi}^{-\frac{1}{2}}\mathbf{B}_n^T\left(\mathbf{B}_n\mathbf{R}_{w_n}\mathbf{B}_n^T\right)^{-1}\mathbf{B}_n\boldsymbol{\Phi}^{-\frac{1}{2}}\right]^{-1}\right\}. \quad (47)$$

Let $\dot{\mathbf{B}}_n \triangleq \mathbf{B}_n\boldsymbol{\Phi}^{-(1/2)}$. We can reformulate (47) as

$$\min_{\{\dot{\mathbf{B}}_n\}} \mathrm{tr}\left\{\left[\sum_{n=1}^{N}\dot{\mathbf{B}}_n^T\left(\dot{\mathbf{B}}_n\boldsymbol{\Phi}^{\frac{1}{2}}\mathbf{R}_{w_n}\boldsymbol{\Phi}^{\frac{1}{2}}\dot{\mathbf{B}}_n^T\right)^{-1}\dot{\mathbf{B}}_n\right]^{-1}\right\} \quad (48)$$

which has a same formulation as (6) and hence can be solved by the proposed compression strategies. In fact, the optimization (6) can be further generalized by considering an objective function in the form of $\text{trace}(\mathbf{S}\mathbf{R}_{\hat{\theta}}\mathbf{S}^T)$, where $\mathbf{S}$ can be arbitrary matrix of appropriate size. A similar approach can be used to solve this problem.

B. Extension of the Data Model

Throughout this paper, a simplified linear data model: $\mathbf{x}_n = \boldsymbol{\theta} + \mathbf{w}_n$ is considered, in which the observation matrix defining the input/output relation is assumed to be an identity matrix. We now discuss the extension to a general linear model

$$\mathbf{x}_n = \mathbf{H}_n\boldsymbol{\theta} + \mathbf{w}_n \qquad n = 1, \dots, N, \tag{49}$$

where $\mathbf{H}_n \in \mathbb{R}^{r_n \times p}$ is assumed to be a full column rank (this condition is usually met in practice) matrix known at the FC. When the observation matrices across the sensors are identical, i.e., $\mathbf{H}_n = \mathbf{H}, \forall n$, the compression design problem for the case $(\mathbf{R}_{w_n} = \sigma_{w,n}^2\mathbf{R}_w)$ can be reduced to the optimization problem (15) (see Appendix C for a detailed derivation). On the other hand, if the observation matrices in (49) are different across sensors, then we have to resort to the iterative algorithm proposed in Section III to search for an effective compression design. Note that by following the same derivation as we did in Section III, the proposed iterative algorithm can be readily adapted to accommodate this general linear model (the details are omitted here).

C. Extension to the Random Parameter Case

The unknown parameter $\boldsymbol{\theta}$ is modeled as a deterministic parameter in our paper. In many other studies, $\boldsymbol{\theta}$ is treated as a random parameter with zero mean and covariance matrix $\mathbf{R}_\theta$. In that case, a linear minimum mean-square error (LMMSE) estimator can be used and the corresponding covariance matrix of the estimate error $\boldsymbol{\epsilon} = \boldsymbol{\theta} - \hat{\boldsymbol{\theta}}$ is given by [24]

$$\begin{aligned}\mathbf{R}_\epsilon &= E[\boldsymbol{\epsilon}\boldsymbol{\epsilon}^T] \\ &= \left[\mathbf{R}_\theta^{-1} + \mathbf{J}^T\mathbf{B}^T\left(\mathbf{B}\mathbf{R}_{\tilde{\mathbf{w}}}\mathbf{B}^T\right)^{-1}\mathbf{B}\mathbf{J}\right]^{-1} \\ &= \left[\mathbf{R}_\theta^{-1} + \sum_{n=1}^N \mathbf{B}_n^T\left(\mathbf{B}_n\mathbf{R}_{w_n}\mathbf{B}_n^T\right)^{-1}\mathbf{B}_n\right]^{-1}. \end{aligned} \tag{50}$$

Note that a different formulation of $\mathbf{R}_\epsilon$ was adopted in some other works, e.g., [18], [19]. Nevertheless, as shown in ([24, Th. 12.1]), they are in fact identical for general linear models. The compression design problem therefore can be formulated as

$$\min_{\{\mathbf{B}_n\}} \text{tr}\left\{\left[\mathbf{R}_\theta^{-1} + \sum_{n=1}^N \mathbf{B}_n^T\left(\mathbf{B}_n\mathbf{R}_{w_n}\mathbf{B}_n^T\right)^{-1}\mathbf{B}_n\right]^{-1}\right\}. \tag{51}$$

Although such a compression design problem (51) has been extensively studied in [14]–[19], as we mentioned earlier, all these algorithms are iterative and suffer from local maxima and stationary points. In the following, we show that the compression strategies proposed in this paper could be effective solutions to (51). We take the homogeneous case as an example. The extension to the inhomogeneous case with $\mathbf{R}_{w_n} = \sigma_{w,n}^2\mathbf{R}_w$ follows a similar derivation.

Let $\lambda_{R_\theta,\min}$ denote the smallest eigenvalue of $\mathbf{R}_\theta$. The estimation error of the LMMSE estimator is upper and lower bounded by

$$\begin{aligned} &f_{\text{ub}}(\{\mathbf{B}_n\}) \\ &\quad\triangleq \text{tr}\left\{\left[\sum_{n=1}^N \mathbf{B}_n^T\left(\mathbf{B}_n\mathbf{R}_w\mathbf{B}_n^T\right)^{-1}\mathbf{B}_n\right]^{-1}\right\} > \text{tr}\{\mathbf{R}_\epsilon\} \\ &\quad\geq \text{tr}\left\{\left[\lambda_{R_\theta,\min}^{-1}\mathbf{I} + \sum_{n=1}^N \mathbf{B}_n^T\left(\mathbf{B}_n\mathbf{R}_w\mathbf{B}_n^T\right)^{-1}\mathbf{B}_n\right]^{-1}\right\} \\ &\quad\triangleq f_{\text{lb}}(\{\mathbf{B}_n\}) \end{aligned} \tag{52}$$

since $\text{tr}(\mathbf{A}^{-1})$ is convex over the set of positive definite matrix and $\lambda_{R_\theta,\min}^{-1}\mathbf{I} - \mathbf{R}_\theta^{-1} \succeq \mathbf{0}$. The upper bound $f_{\text{ub}}(\{\mathbf{B}_n\})$ is exactly the estimation error of the BLUE estimator for the deterministic case. Recalling that the compression strategy proposed in Section IV-B approaches its universal lower bound (23), therefore the proposed compression strategy is able to attain an estimation error upper bounded by

$$f_{\text{ub}}(\{\mathbf{B}_n\}) \to \frac{1}{K}\left[\sum_{i=1}^p \sqrt{\lambda_{w,i}}\right]^2 > \text{tr}\{\mathbf{R}_\epsilon\}. \tag{53}$$

Also, for any compression dimension assignment, the estimation error of the LMMSE estimator achieved by any compression strategy is lower bounded by (see Appendix D for a detailed derivation)

$$\text{tr}\{\mathbf{R}_\epsilon\} \geq f_{\text{lb}}(\{\mathbf{B}_n\}) \geq \frac{1}{\tilde{K}}\left[\sum_{i=1}^p \sqrt{\lambda_{w,i}}\right]^2 \tag{54}$$

where

$$\tilde{K} \triangleq \lambda_{R_\theta,\min}^{-1}\sum_{i=1}^p \lambda_{w,i} + K. \tag{55}$$

Combining (53) and (54), we can assure that the estimation error of the LMMSE estimator achieved by our compression strategy is within a factor $\tilde{K}/K$ of the minimum achievable estimation error. Suppose $\mathbf{R}_\theta = \mathbf{I}$ and $\mathbf{R}_w = \mathbf{I}$, we have $\tilde{K}/K = 1 + (p/K)$, which approaches one when K is relatively larger than p.

D. Discussions of an Existing Suboptimal Solution

In [15], to circumvent the difficulty in solving the optimization (6), the authors proposed to consider a more tractable criterion

$$\max_{\{\mathbf{B}_n\}} \text{tr}\left\{\sum_{n=1}^N \mathbf{B}_n^T\left(\mathbf{B}_n\mathbf{R}_{w_n}\mathbf{B}_n^T\right)^{-1}\mathbf{B}_n\right\}. \tag{56}$$

Although the above optimization (56) can be analytically solved, it does not always guarantee to yield an effective solution. To see this, let us consider a simple example where $q_n = q$, and $\mathbf{R}_{w_n} = \mathbf{\Lambda}, \forall n, \mathbf{\Lambda}$ is a diagonal matrix with ascending diagonal elements. The optimization (56) can be decoupled into a set of identical subtasks with each compression matrix $\mathbf{B}_n$ determined from the following optimization:

$$\max_{\mathbf{B}_n} \operatorname{tr}\left\{\mathbf{B}_n^T(\mathbf{B}_n\mathbf{\Lambda}\mathbf{B}_n^T)^{-1}\mathbf{B}_n\right\}. \tag{57}$$

It can be readily verified that the optimal solution $\mathbf{B}_n$ is given by the first q rows of the identity matrix $\mathbf{I}_p$ and all compression matrices $\{\mathbf{B}_n\}$ are identical. Clearly this is not a meaningful solution because with such a compression choice, all sensors transmit only the noisy observations of the first q parameters, whereas the information concerning the last $p-q$ parameters is not reported to the FC. Mathematically, this solution leads to a rank-deficient matrix $\sum_{n=1}^{N}\mathbf{B}_n^T(\mathbf{B}_n\mathbf{\Lambda}\mathbf{B}_n^T)^{-1}\mathbf{B}_n$. In this case, the overall estimation error can be considered going to infinity.

VI. Simulation Results

In this section, we carry out experiments to corroborate our previous analysis and to illustrate the performance of our compression strategies.

A. Homogeneous Case

We first consider a homogeneous environment where sensors have identical noise covariance matrices. In this case, the compression matrices can be determined by the iterative algorithm proposed in Section III or by the near-optimal compression strategy proposed in Section IV-B. We investigate the performance of our compression strategies and compare them with a noncompression BLUE estimator using a certain number of sensors' original observations. All these schemes send the same number of messages to the FC (note that different schemes may need different number of sensors). The universal lower bound for the estimation error of any compression strategy, which is given in (23), is also included for comparison. In our simulations, the parameter dimension p is set to 5 and the noise covariance matrix is chosen as $\mathbf{R}_w = \operatorname{diag}(1, 0.1, 0.1, 0.1, 0.001)$. Fig. 2 depicts the total number of messages K, versus the estimation distortion. From Fig. 2, we see that the compression strategy proposed in Section IV-B attains performance that is very close to (in fact indistinguishable from) the universal lower bound, which corroborates our claim that the rounding operation incurs a very mild performance loss. It is also interesting to examine the performance of the iterative algorithm proposed in Section III. We consider three different compression dimension assignments, namely, each sensor compresses its local observation into q messages, where $q = 1, 2, 3$, respectively. We observe that different compression dimension assignments affects the performance of the proposed iterative algorithm: the performance of the iterative algorithm degrades as q increases. The reason, as we mentioned in Section IV-C.3, is that transmitting multiple messages for each sensor may result in an insufficient number of messages to reconstruct the projections onto the noisier eigenvectors (that is, eigenvectors associated with large eigenvalues). We also notice that when $q = 1$ (the same compression dimension assignment as that of the proposed near-optimum strategy), the proposed iterative algorithm achieves almost the same performance as that of the proposed near-optimum strategy. This implies that the iterative algorithm converges to a stationary point that is close to the global minimum, although theoretically this convergence is not guaranteed. Moreover, it can be seen that all compression strategies present an advantage over the noncompression estimator in a K-distortion sense.

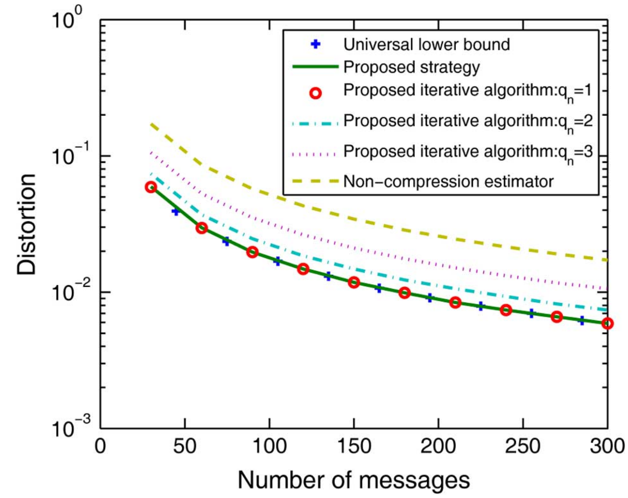

Fig. 2. Homogeneous case: Overall estimation error versus the total number of messages, K, sent to the FC for respective schemes.

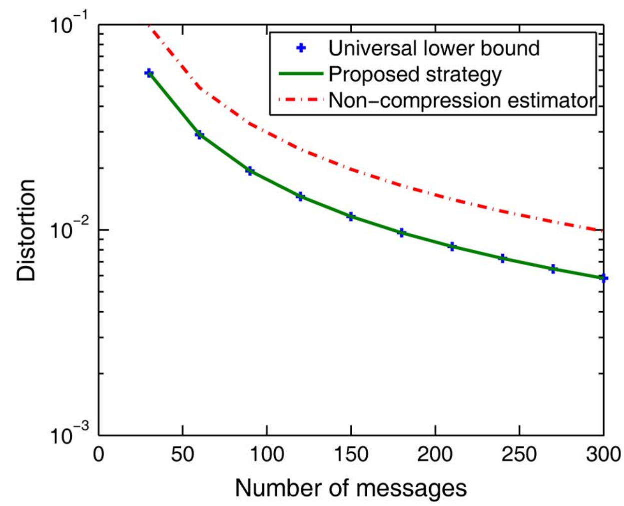

Fig. 3. Homogeneous case: Overall estimation error versus the total number of messages, K, sent to the FC for respective schemes.

We also study an example where the eigenvalues are randomly generated according to $\lambda_{w,i} = \alpha\nu_i \; \forall i \in \{1,\ldots,p\}$, where $\alpha = 0.1$ and $\nu_i \sim \mathcal{X}_1^2$ is a central chi-square distributed random variable with one degree-of-freedom. Note that the chi-square distribution has been used to model the sensor noise variance, e.g., [26]. Since the noise variances are closely related to the eigenvalues of the noise covariance matrix (in particular, the noise variances are exactly the eigenvalues when the noise covariance matrix is diagonal), we adopt the same statistical model to characterize the distribution of the eigenvalues. Fig. 3 shows the performance of the proposed near-optimum

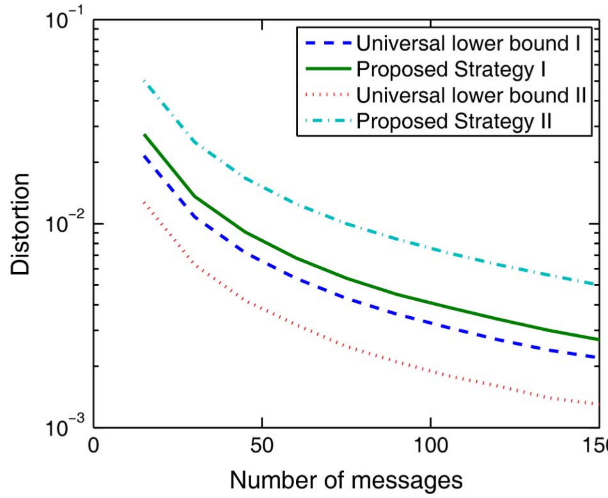

Fig. 4. Inhomogeneous case with diverse eigenvalues: Overall estimation error versus the total number of messages, K, for our proposed strategies.

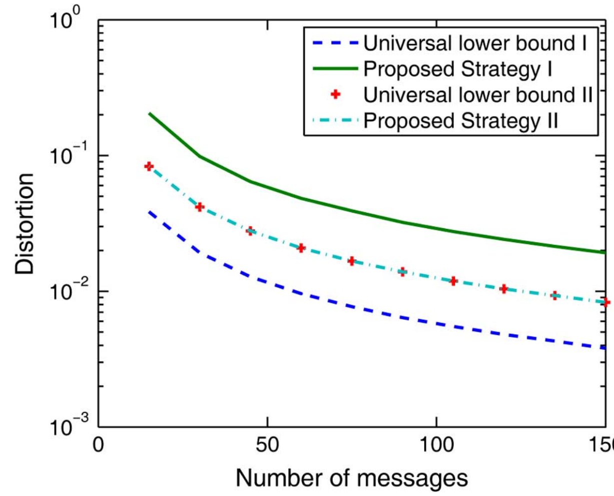

Fig. 5. Inhomogeneous case with identical eigenvalues: Overall estimation error versus the total number of messages, K, for our proposed strategies.

strategy and the noncompression estimator. The results are averaged over 500 Monte Carlo runs. We see that our compression strategy presents a clear advantage over the noncompression estimator. To meet the same distortion target, say, 0.01, the number of messages required by our compression scheme is about $1/2$ of that required by the noncompression estimator.

B. Inhomogeneous Case

We conduct experiments to examine the performance of the two compression strategies proposed in Section IV-C. As we discussed before, these two compression strategies are different in nature: the first strategy exploits the diversity of the eigenvalues of the noise covariance matrix, while the second one takes advantage of the disparity of the observation qualities. We assume that the sensors are equally divided into three clusters, with sensors in each cluster having the same observation quality, i.e., $\sigma_{w_n}^2 = \sigma_{c_m}^2, \forall n \in C_m$, where C_m denotes cluster m. In our simulations, the number of messages sent to the FC is set to $K = 3N_c$, where N_c denotes the number of sensors in each cluster (N_c varies from 5 to 50 in the simulations). In our first example, we set $p = 5$ and $\mathbf{R}_w$ has one unit (dominant) eigenvalue and four small eigenvalues 0.001. The factors $\sigma_{c_m}^2$ for the three clusters are set to be 0.15, 0.6, and 1. The Strategy I involves partitioning sensors into p groups, with the objective that the variable χ_i associated with each group approaches their optimum values (36). This partition can be easily accomplished in most cases. For example, suppose $N_c = 50$, it can be calculated that the optimum values of $\{\chi_i\}$ are $\{13.1002, 13.1002, 13.1002, 13.1002, 414.2657\}$. We can easily find a partition whose corresponding values are $\{\chi_i\} = \{13, 13, 13, 12.6667, 415\}$ which are close to the optimum values. Fig. 4 shows the estimation distortions of the two proposed compression strategies as a function of the total number of messages sent to the FC. The two universal lower bounds (30) and (39), referred to as universal lower bound I and II respectively, are also included to evaluate the proposed strategies. From Fig. 4, we observe that (30) provides a tighter lower bound, whereas the lower bound (39) is a loose one. Also, it can be seen that the compression Strategy I is very effective and achieves performance that is close to the universal lower bound I, which corroborates our analysis in Section IV-C1. In contrast, the proposed Strategy II yields inferior performance in this case.

We now study a different scenario where the noise covariance matrix has identical eigenvalues. We set $\mathbf{R}_w = \mathbf{I}$ and the factors $\sigma_{c_m}^2$ for the three clusters are set to be 0.1, 0.5, and 1. The performance is depicted in Fig. 5. We see that, in this case, the universal lower bound II turns out to be a tighter lower bound. Also, as we analyzed in Section IV-C2, the proposed Strategy II approaches/attains the universal lower bound II. On the other hand, unlike the previous example, the Strategy I becomes apart from both universal lower bounds since the eigenvalues of the noise covariance matrix are identical and no diversity can be exploited.

VII. Conclusion

We considered the problem of distributed estimation of a deterministic vector parameter in wireless sensor networks, where due to the stringent power/bandwidth constraints, each sensor carries out local data dimensionality reduction to reduce the transmission requirement. The problem of interest is to jointly design the compression matrices associated with those sensors in order to achieve a minimal estimation error at the FC. Such a compression design problem was investigated in this paper. We first developed an efficient iterative algorithm for a general noise scenario. We then examined two specific but important noise scenarios: a homogeneous environment where all sensors have identical noise covariance matrices and an inhomogeneous environment where the noise covariance matrices across the sensors have the same correlation structure but with different scaling factors. Compression strategies were proposed for both noise scenarios, respectively. Universal lower bounds were derived to evaluate the performance of the proposed compression strategies. Simulation results were presented to corroborate our analysis and to illustrate the effectiveness of the proposed compression strategies.

An interesting topic of future study is to work beyond the homogeneous and the specific inhomogeneous cases considered in this paper. Currently an optimum compression strategy for the general noise case still remains an open problem. The difficulty in dealing with the general noise case lies in that the reduction of the optimization into a tractable form is very difficult, due to the fact that different noise covariance matrices have different eigenvalue decompositions, which makes the summation term inside the inverse operator irreducible. The generalization of our approach or results to the general noise case requires to find a way to overcome the above mentioned difficulty.

Appendix A
Proof of Theorem 1

To derive a lower bound on the minimum achievable objective function value of (16), we construct a new optimization that has the same objective function as (16) while with a relaxed constraint. From the constraints $\mathbf{C}_n\mathbf{C}_n^T = \mathbf{I}_{q_n}, \forall n \in \{1,\ldots,N\}$, we can deduce that

$$\begin{aligned}\mathrm{tr}(\mathbf{C}^T\mathbf{D}_\sigma\mathbf{C}) &= \sum_{n=1}^{N} \frac{1}{\sigma_{w,n}^2}\mathrm{tr}\left(\mathbf{C}_n^T\mathbf{C}_n\right)\\ &= \sum_{n=1}^{N} \frac{1}{\sigma_{w,n}^2}\mathrm{tr}\left(\mathbf{C}_n\mathbf{C}_n^T\right)\\ &= \sum_{n=1}^{N} \frac{1}{\sigma_{w,n}^2} q_n \triangleq L \end{aligned}\tag{58}$$

where $\mathbf{C}$ and $\mathbf{D}_\sigma$ are defined in (20) and (21), respectively. Clearly the identity (58) is more relaxed than the constraints $\mathbf{C}_n\mathbf{C}_n^T = \mathbf{I}_{q_n}, \forall n$ because we can derive (58) from these constraints, but the converse is not true. Therefore a lower bound on the minimum achievable function value of (16) can be obtained by solving the following optimization

$$\begin{aligned}\min_{\mathbf{C}} \quad & \mathrm{tr}\left\{\left[\sum_{n=1}^{N}\frac{1}{\sigma_{w,n}^2}\mathbf{C}_n^T\mathbf{C}_n\right]^{-1}\mathbf{D}_w\right\}\\ & = \mathrm{tr}\left\{[\mathbf{C}^T\mathbf{D}_\sigma\mathbf{C}]^{-1}\mathbf{D}_w\right\}\\ \text{s.t.} \quad & \mathrm{tr}(\mathbf{C}^T\mathbf{D}_\sigma\mathbf{C}) = L. \end{aligned}\tag{59}$$

This optimization, unlike (16), can be analytically solved and the results are summarized as follows.

Lemma 2: Let $\lambda_{w,i}$ denote the ith eigenvalue of $\mathbf{R}_w$, i.e., the ith diagonal element of $\mathbf{D}_w$. If any matrix $\mathbf{C}$ satisfies the following condition:

$$\mathbf{C}^T\mathbf{D}_\sigma\mathbf{C} = \mathbf{D}^* \triangleq \mathrm{diag}(d_1^*,\ldots,d_p^*), \tag{60}$$

then it is an optimum solution to (59), where

$$d_i^* \triangleq \frac{L\sqrt{\lambda_{w,i}}}{\sum_{i=1}^{p}\sqrt{\lambda_{w,i}}}. \tag{61}$$

The minimum objective function value achieved by this optimum solution is given by

$$f_{\min} = \frac{1}{L}\left[\sum_{i=1}^{p}\sqrt{\lambda_{w,i}}\right]^2. \tag{62}$$

Proof: See Appendix B. ■

From Lemma 2, it is clear to see that the estimation error of any compression design strategy is lower bounded by

$$\mathrm{tr}\left\{\mathbf{R}_{\hat{\theta}}\right\} \geq \frac{1}{L}\left[\sum_{i=1}^{p}\sqrt{\lambda_{w,i}}\right]^2. \tag{63}$$

The lower bound can be attained if the compression matrices $\{\mathbf{C}_n\}$ satisfy the constraints defined in (16) and the condition (60). The proof is completed here.

Appendix B
Proof of Lemma 2

Let $\mathbf{C}^T\mathbf{D}_\sigma\mathbf{C} = \mathbf{U}\mathbf{D}\mathbf{U}^T$ denote its eigenvalue decomposition (EVD), where $\mathbf{U} \in \mathbb{R}^{p\times p}$ and $\mathbf{D} \in \mathbb{R}^{p\times p}$. By replacing $\mathbf{C}^T\mathbf{D}_\sigma\mathbf{C}$ with its EVD, the optimization (59) is reduced to determining the orthonormal matrix $\mathbf{U}$ and the diagonal matrix $\mathbf{D}$

$$\begin{aligned}\min_{\mathbf{U},\mathbf{D}} \quad & \mathrm{tr}\{\mathbf{U}\mathbf{D}^{-1}\mathbf{U}^T\mathbf{D}_w\}\\ \text{s.t.} \quad & \mathrm{tr}(\mathbf{D}) = L\\ & \mathbf{D} = \mathrm{diag}(d_1,\ldots,d_p) \quad d_i > 0 \quad \forall i\\ & \mathbf{U}\mathbf{U}^T = \mathbf{I}. \end{aligned}\tag{64}$$

Although the problem (64) involves searching for multiple optimization variable matrices, a close examination shows that it can be decoupled into two sequential subproblems. We can, firstly, find an optimal $\mathbf{U}$ by fixing the variable $\mathbf{D}$ (its diagonal elements $\{d_i\}$, without loss of generality, are assumed in descending order). By using the following matrix inequality [27]:

$$\mathrm{tr}(\mathbf{A}\mathbf{B}) \geq \sum_{k=1}^{p}\lambda_k(\mathbf{A})\lambda_{p+1-k}(\mathbf{B}) \tag{65}$$

where $\mathbf{A}$ and $\mathbf{B}$ are any positive semi-definite Hermitian matrices with eigenvalues $\lambda_k(\mathbf{A})$ and $\lambda_k(\mathbf{B})$ arranged in a descending order, we can reach that the objective function is lower bounded by

$$\mathrm{tr}\{\mathbf{U}\mathbf{D}^{-1}\mathbf{U}^T\mathbf{D}_w\} \geq \sum_{i=1}^{p}\frac{1}{d_i}\lambda_{w,i}. \tag{66}$$

It is easy to verify that this lower bound is attained when $\mathbf{U} = \mathbf{I}$. Therefore $\mathbf{U} = \mathbf{I}$ is an optimum solution to (64). With this result, the optimum diagonal matrix $\mathbf{D}$ can be determined via the following optimization:

$$\begin{aligned}\min_{\mathbf{D}} \quad & \mathrm{tr}\{\mathbf{D}^{-1}\mathbf{D}_w\}\\ \text{s.t.} \quad & \mathrm{tr}(\mathbf{D}) = L\\ & \mathbf{D} = \mathrm{diag}(d_1,\ldots,d_p) \quad d_i > 0 \quad \forall i \end{aligned}\tag{67}$$

which is equivalent to finding its diagonal elements by

$$\min_{\{d_i\}} \quad \sum_{i=1}^{p} \frac{\lambda_{w,i}}{d_i}$$
$$\text{s.t.} \quad \sum_{i=1}^{p} d_i = L \quad d_i > 0 \quad \forall i. \tag{68}$$

The Lagrangian function f_L associated with (68) is given by

$$f_L(d_i;\lambda;\nu_i) = \sum_{i=1}^{p} \frac{\lambda_{w,i}}{d_i} - \lambda\left(L - \sum_{i=1}^{p} d_i\right) - \sum_{i=1}^{p} \nu_i d_i, \tag{69}$$

which gives the following KKT conditions:

$$-\frac{\lambda_{w,i}}{d_i^2} + \lambda - \nu_i = 0 \quad \forall i$$
$$L - \sum_{i=1}^{p} d_i = 0$$
$$\nu_i d_i = 0 \quad \forall i$$
$$\nu_i \geq 0 \quad \forall i$$
$$d_i > 0 \quad \forall i.$$

The last three KKT conditions imply that $\nu_i = 0, \forall i$. Substituting this result into the first equation, we obtain

$$d_i = \sqrt{\frac{\lambda_{w,i}}{\lambda}}. \tag{70}$$

The Lagrangian multiplier λ can be determined from the second KKT condition, from which λ is given by

$$\sqrt{\lambda} = \frac{\sum_{i=1}^{p} \sqrt{\lambda_{w,i}}}{L}. \tag{71}$$

By combining (70) and (71), the optimum solution is solved as

$$d_i^* = \frac{L\sqrt{\lambda_{w,i}}}{\sum_{i=1}^{p} \sqrt{\lambda_{w,i}}}. \tag{72}$$

The proof is completed here.

APPENDIX C
SIMPLIFICATION OF THE COMPRESSION DESIGN OPTIMIZATION

For the data model: $\mathbf{x}_n = \mathbf{H}\boldsymbol{\theta} + \mathbf{w}$, the compression design problem for the special case $\mathbf{R}_{w_n} = \sigma_{w,n}^2 \mathbf{R}_w$ is formulated into an optimization as follows (for simplicity, the constraint on the total number of compressed messages is omitted)

$$\min_{\{\mathbf{B}_n\}} \quad \text{tr}\left\{\left[\sum_{n=1}^{N} \frac{1}{\sigma_{w,n}^2} \mathbf{H}^T\mathbf{B}_n^T\left(\mathbf{B}_n\mathbf{R}_w\mathbf{B}_n^T\right)^{-1}\mathbf{B}_n\mathbf{H}\right]^{-1}\right\} \tag{73}$$

where $\mathbf{H} \in \mathbb{R}^{r\times p}$ $(r \geq p)$, and $\mathbf{B}_n \in \mathbb{R}^{q_n\times r}$. Define $\bar{\mathbf{C}}_n \triangleq \mathbf{B}_n\mathbf{U}_w\mathbf{D}_w^{(1/2)}$. We have

$$\begin{aligned}\mathbf{H}^T\mathbf{B}_n^T&\left(\mathbf{B}_n\mathbf{R}_w\mathbf{B}_n^T\right)^{-1}\mathbf{B}_n\mathbf{H}\\ &= \mathbf{H}^T\mathbf{U}_w\mathbf{D}_w^{-\frac{1}{2}}\bar{\mathbf{C}}_n^T\left(\bar{\mathbf{C}}_n\bar{\mathbf{C}}_n^T\right)^{-1}\bar{\mathbf{C}}_n\mathbf{D}_w^{-\frac{1}{2}}\mathbf{U}_w^T\mathbf{H}\\ &\triangleq \mathbf{S}\bar{\mathbf{C}}_n^T\left(\bar{\mathbf{C}}_n\bar{\mathbf{C}}_n^T\right)^{-1}\bar{\mathbf{C}}_n\mathbf{S}^T\end{aligned} \tag{74}$$

in which $\mathbf{S} \triangleq \mathbf{H}^T\mathbf{U}_w\mathbf{D}_w^{-(1/2)}$. Let $\mathbf{S} = \mathbf{U}_s\mathbf{D}_s\mathbf{V}_s^T$ denote the reduced singular value decomposition (SVD) of $\mathbf{S}$, where $\mathbf{U}_s \in \mathbb{R}^{p\times p}, \mathbf{D}_s \in \mathbb{R}^{p\times p}$, and $\mathbf{V}_s \in \mathbb{R}^{r\times p}$. Also, we write $\bar{\mathbf{C}}_n = \mathbf{P}_n\mathbf{C}_n$, where $\mathbf{P}_n \in \mathbb{R}^{q_n\times q_n}$ is a full rank matrix and $\mathbf{C}_n \in \mathbb{R}^{q_n\times r}$ consists of q_n orthonormal rows, i.e., $\mathbf{C}_n\mathbf{C}_n^T = \mathbf{I}_{q_n}$. Using these results, (74) can be rewritten as

$$\mathbf{H}^T\mathbf{B}_n^T\left(\mathbf{B}_n\mathbf{R}_w\mathbf{B}_n^T\right)^{-1}\mathbf{B}_n\mathbf{H} = \mathbf{U}_s\mathbf{D}_s\mathbf{V}_s^T\mathbf{C}_n^T\mathbf{C}_n\mathbf{V}_s\mathbf{D}_s\mathbf{U}_s^T. \tag{75}$$

Substituting (75) into (73), the optimization (73) becomes

$$\min_{\{\mathbf{C}_n\}} \quad \text{tr}\left\{\left[\mathbf{V}_s^T\left(\sum_{n=1}^{N}\frac{1}{\sigma_{w,n}^2}\mathbf{C}_n^T\mathbf{C}_n\right)\mathbf{V}_s\right]^{-1}\mathbf{D}_s^{-2}\right\}$$
$$\text{s.t.} \quad \mathbf{C}_n\mathbf{C}_n^T = \mathbf{I}_{q_n} \quad \forall n. \tag{76}$$

Let $\mathbf{V}_s^{\perp} \in \mathbb{R}^{r\times(r-p)}$ denote the orthogonal complement of the subspace $\mathbf{V}_s$, i.e., $(\mathbf{V}_s^{\perp})^T\mathbf{V}_s = \mathbf{0}$ and $(\mathbf{V}_s^{\perp})^T\mathbf{V}_s^{\perp} = \mathbf{I}$; and define $\mathbf{V} \triangleq [\mathbf{V}_s \;\; \mathbf{V}_s^{\perp}]$. Since the columns of $\mathbf{V}$ form a complete orthonormal basis, each compression matrix $\mathbf{C}_n$ can be expressed as $\mathbf{C}_n = \tilde{\mathbf{C}}_n\mathbf{V}^T$, where $\tilde{\mathbf{C}}_n \in \mathbb{R}^{q_n\times r}$. Using this new representation, we can rewrite (76) as

$$\min_{\{\tilde{\mathbf{C}}_n\}} \quad \text{tr}\left\{\left[\mathbf{T}\left(\sum_{n=1}^{N}\frac{1}{\sigma_{w,n}^2}\tilde{\mathbf{C}}_n^T\tilde{\mathbf{C}}_n\right)\mathbf{T}^T\right]^{-1}\mathbf{D}_s^{-2}\right\}$$
$$\text{s.t.} \quad \tilde{\mathbf{C}}_n\tilde{\mathbf{C}}_n^T = \mathbf{I}_{q_n} \quad \forall n \tag{77}$$

where $\mathbf{T} \triangleq \mathbf{V}_s^T\mathbf{V} = [\mathbf{I}_p \quad \mathbf{0}]$. To gain insight into (77), we decompose each compression matrix $\tilde{\mathbf{C}}_n$ into two parts: $\tilde{\mathbf{C}}_n = [\tilde{\mathbf{C}}_{n,1} \quad \tilde{\mathbf{C}}_{n,2}]$, where $\tilde{\mathbf{C}}_{n,1} \in \mathbb{R}^{q_n\times p}$ and $\tilde{\mathbf{C}}_{n,2} \in \mathbb{R}^{q_n\times(r-p)}$. It is easy to verify that

$$\mathbf{T}\left(\sum_{n=1}^{N}\frac{1}{\sigma_{w,n}^2}\tilde{\mathbf{C}}_n^T\tilde{\mathbf{C}}_n\right)\mathbf{T}^T = \sum_{n=1}^{N}\frac{1}{\sigma_{w,n}^2}\tilde{\mathbf{C}}_{n,1}^T\tilde{\mathbf{C}}_{n,1}. \tag{78}$$

Therefore the optimization (77) is rewritten as

$$\min_{\{\tilde{\mathbf{C}}_{n,1},\tilde{\mathbf{C}}_{n,2}\}} \quad \text{tr}\left\{\left[\sum_{n=1}^{N}\frac{1}{\sigma_{w,n}^2}\tilde{\mathbf{C}}_{n,1}^T\tilde{\mathbf{C}}_{n,1}\right]^{-1}\mathbf{D}_s^{-2}\right\}$$
$$\text{s.t.} \quad \tilde{\mathbf{C}}_{n,1}\tilde{\mathbf{C}}_{n,1}^T + \tilde{\mathbf{C}}_{n,2}\tilde{\mathbf{C}}_{n,2}^T = \mathbf{I}_{q_n} \quad \forall n. \tag{79}$$

Since $\mathbf{I}_{q_n} \succeq \tilde{\mathbf{C}}_{n,1}\tilde{\mathbf{C}}_{n,1}^T$ and $\text{tr}(\mathbf{A}^{-1})$ is convex over the set of positive definite matrices, it is clear that we should have $\tilde{\mathbf{C}}_{n,2} =$

$\mathbf{0}, \forall n$ in order to minimize (79) and eventually we arrive at the following:

$$\min_{\{\tilde{\mathbf{C}}_{n,1}\}} \quad \operatorname{tr}\left\{\left[\sum_{n=1}^{N}\frac{1}{\sigma_{w,n}^2}\tilde{\mathbf{C}}_{n,1}^T\tilde{\mathbf{C}}_{n,1}\right]^{-1}\mathbf{D}_s^{-2}\right\}$$
$$\text{s.t.} \quad \tilde{\mathbf{C}}_{n,1}\tilde{\mathbf{C}}_{n,1}^T = \mathbf{I}_{q_n} \quad \forall n \tag{80}$$

which has the same formulation as the problem (16) we studied in this paper.

Appendix D
Derivation of (54)

We only need to prove

$$f_{\text{lb}}(\{\mathbf{B}_n\}) \geq \frac{1}{\tilde{K}}\left[\sum_{i=1}^{p}\sqrt{\lambda_{w,i}}\right]^2. \tag{81}$$

By resorting to (12)–(13) and writing $\lambda_{R_\theta,\min}^{-1}\mathbf{I} = \lambda_{R_\theta,\min}^{-1}\mathbf{U}_w\mathbf{D}_w^{-(1/2)}\mathbf{D}_w\mathbf{D}_w^{-(1/2)}\mathbf{U}_w^T$, the minimal $f_{\text{lb}}(\{\mathbf{B}_n\})$ can be solved as

$$\min_{\{\mathbf{C}_n\}} \quad \operatorname{tr}\left\{\left[\lambda_{R_\theta,\min}^{-1}\mathbf{D}_w + \sum_{n=1}^{N}\mathbf{C}_n^T\mathbf{C}_n\right]^{-1}\mathbf{D}_w\right\}$$
$$\text{s.t.} \quad \mathbf{C}_n\mathbf{C}_n^T = \mathbf{I}_{q_n} \quad \forall n \in \{1,\ldots,N\}. \tag{82}$$

Following a similar derivation in Appendix A, we know that the minimum achievable function value of the above optimization is lower bounded by the minimum achievable function value of the following optimization:

$$\min_{\{\mathbf{C}_n\}} \quad \operatorname{tr}\left\{[\lambda_{R_\theta,\min}^{-1}\mathbf{D}_w + \mathbf{C}^T\mathbf{C}]^{-1}\mathbf{D}_w\right\}$$
$$\text{s.t.} \quad \operatorname{tr}(\lambda_{R_\theta,\min}^{-1}\mathbf{D}_w + \mathbf{C}^T\mathbf{C}) = \lambda_{R_\theta,\min}^{-1}\sum_{i=1}^{p}\lambda_{w,i} + K \triangleq \tilde{K}. \tag{83}$$

The above optimization is equivalent to (59). From Lemma 2, its minimum achievable function value is given by

$$f_{\min} = \frac{1}{\tilde{K}}\left[\sum_{i=1}^{p}\sqrt{\lambda_{w,i}}\right]^2. \tag{84}$$

Acknowledgment

The authors would like to extend their appreciation to the reviewers for their valuable comments and suggestions.

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] J.-J. Xiao and Z.-Q. Luo, "Universal decentralized detection in a bandwidth-constrained sensor network," *IEEE Trans. Signal Processing*, vol. 53, no. 8, pp. 2617–2624, Aug. 2005.

[4] R. Viswanathan and P. K. Varshney, "Distributed detection with multiple sensors—Part I: Fundamentals," *Proc. IEEE*, vol. 85, no. 1, pp. 54–63, Jan. 1997.

[5] R. S. Blum, S. A. Kassam, and H. V. Poor, "Distributed detection with multiple sensors: Part II—Advance topics," *Proc. IEEE*, vol. 85, no. 1, pp. 64–79, Jan. 1997.

[6] 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.

[7] Z. Luo and J. Xiao, "Decentralized estimation in an inhomogeneous sensing environment," *IEEE Trans. Inf. Theory*, vol. 51, no. 10, pp. 3564–3575, Oct. 2005.

[8] 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.

[9] P. Venkitasubramaniam, L. Tong, and A. Swami, "Quantization for maximin ARE in distributed estimation," *IEEE Trans. Signal Processing*, vol. 55, no. 7, pp. 3596–3605, Jul. 2007.

[10] 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.

[11] A. Dogandžiá and K. Qiu, "Decentralized random-field estimation for sensor networks using quantized spatially correlated data and fusion-center feedback," *IEEE Trans. Signal Processing*, no. 12, pp. 6069–6085, Dec. 2008.

[12] 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.

[13] Z.-Q. Luo and J. N. Tsitsiklis, "Data fusion with minimal communication," *IEEE Trans. Inf. Theorey*, no. 5, pp. 1551–1563, Sep. 1994.

[14] K. Zhang, X. R. Li, P. Zhang, and H. Li, "Optimal linear estimation fusion—Part VI: Sensor data compression," in *Proc. Int. Conf. Inf. Fusion*, Queensland, Australia, 2003.

[15] 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.

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

[17] M. Gastpar, P. L. Dragotti, and M. Vetterli, "The distributed Karhunen-Loève transform," *IEEE Trans. Inf. Theory*, no. 12, pp. 5177–5196, Dec. 2006.

[18] 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.

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

[20] J.-J. Xiao, S. Cui, Z.-Q. Luo, and A. J. Goldsmith, "Linear coherent decentralized estimation," *IEEE Trans. Signal Processing*, vol. 56, no. 2, pp. 757–770, Feb. 2008.

[21] O. Roy and M. Vetterli, "Dimensionality reduction for distributed estimation in the infinite dimensional regime," *IEEE Trans. Inf. Theorey*, no. 4, pp. 1655–1669, Apr. 2008.

[22] J. Fang and H. Li, "Power constrained distributed estimation with cluster-based sensor collaboration," *IEEE Trans. Wireless Commun.*, vol. 8, no. 7, pp. 3833–3832, July 2009.

[23] J. Fang and H. Li, "Hyperplane-based vector quantization for distributed estimation in wireless sensor networks," *IEEE Trans. Inf. Theory*, vol. 55, no. 12, pp. 5682–5699, Dec. 2009.

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

[25] G. H. Golub and C. F. Van Loan, *Matrix Computations*, 3rd ed. Baltimore, MD: Johns Hopkins University Press, 1996.

[26] 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.

[27] R. A. Horn and C. R. Johnson, *Matrix Analysis*. Cambridge, U.K.: Cambridge University Press, 1985.

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 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 (SM'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, 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 in summers 2003, 2004, and 2009. Since July 1999, he has been with the Department of Electrical and Computer Engineering, Stevens Institute of Technology, Hoboken, NJ, where he is a 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.