

Hyperplane-Based Vector Quantization for Distributed Estimation in Wireless Sensor Networks

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

Abstract—This paper considers distributed estimation of a vector parameter in the presence of zero-mean additive multivariate Gaussian noise in wireless sensor networks. Due to stringent power and bandwidth constraints, vector quantization is performed at each sensor to convert its local noisy vector observation into one bit of information, which is then forwarded to a fusion center where a final estimate of the vector parameter is obtained. Within such a context, this paper focuses on a class of hyperplane-based vector quantizers which linearly convert the observation vector into a scalar by using a compression vector and then carry out a scalar quantization. It is shown that the key of the vector quantization design is to find a compression vector for each sensor. Under the framework of the Cramér–Rao bound (CRB) analysis, the compression vector design problem is formulated as an optimization problem that minimizes the trace of the CRB matrix. Such an optimization problem is extensively studied. In particular, an efficient iterative algorithm is developed for the general case, along with optimal and near-optimal solutions for some specific but important noise scenarios. Performance analysis and simulation results are carried out to illustrate the effectiveness of the proposed scheme.

Index Terms—Cramér–Rao bound, distributed estimation, hyperplane-based vector quantization, optimization, wireless sensor networks.

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 efficiency is a primary issue in sensor networks as the sensors constructing the network are powered by small batteries that are often irreplaceable. Also, in a sensor network, communication, relative to sensing and computation, consumes a significant portion of the total energy. It is therefore important to develop bandwidth- and energy-efficient strategies for various sensor network processing tasks.

Distributed parameter estimation is a fundamental problem arising from sensor network applications. One of the most commonly used network settings for distributed estimation involves a set of spatially distributed sensors linked with a fusion center

(FC). Each sensor makes a noisy observation of the phenomenon of interest and transmits its processed information to the FC, where a final estimate is formed. To address the stringent power and bandwidth constraints inherent in WSNs, the noisy observation at each sensor has to be quantized into one or a few bits of information. In this setup, quantization becomes an integral part of the estimation process and is critical to the estimation performance. The quantization design in such a distributed estimation context has been extensively studied in many works, e.g., [3]–[12]. Specifically, by modeling the unknown parameter as a random parameter, Bayesian techniques were proposed in, e.g., [3]–[5]. These methods require knowledge of the joint distribution of the unknown parameter and the observed signals for quantizer design. Another category of methods [6], [7], [11] treated the unknown parameter as a deterministic unknown parameter and carried out a Cramér–Rao bound (CRB) analysis in examining the impact of the choice of the quantization threshold on the estimation performance. It was found [6], [7] that when a common quantization threshold is applied at all sensors, the estimation performance degrades exponentially as the quantization threshold deviates from the unknown parameter. To overcome this difficulty, a multi-thresholding approach [6], [7] employing a set of non-identical thresholds and an adaptive quantization approach [11] adaptively adjusting the thresholds from sensor to sensor were proposed.

So far most of previous studies focused on the scalar parameter case. For the general vector parameter scenario, the problem becomes much more complicated because, unlike the scalar case which is concerned about only the choice of the quantization threshold, vector quantization involves partitioning of a high dimensional space. Although important, such a problem has not received adequate attention as its scalar counterpart. In [12], the vector quantization problem was briefly discussed. The authors proposed a hyperplane-based approach, where each sensor employs p (p denotes the dimension of the vector observation) hyperplanes perpendicular to the eigenvectors of the noise covariance matrix to quantize the vector observation into p bits of information, i.e., one bit per sensor per dimension. Although effective, its choice of the hyperplanes is heuristic and the purpose is to generate p independent binary data for each sensor, in which case the vector quantization problem can be treated as p scalar quantization problems. Another work [13] studied the vector quantization design in the context of best linear unbiased estimation fusion, where the unknown parameter is treated random and the joint distribution of the parameter and the observed signals is required for partition design.

In this paper, we consider the problem of distributed quantization and estimation of a deterministic vector parameter, where the noisy vector observation of each sensor is quantized into

Manuscript received May 25, 2009; revised August 24, 2009. Current version published November 20, 2009. 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).

Communicated by E. Serpedin, Associate Editor for Signal Processing.

Color versions of Figures 1–7 in this paper are available online at <http://ieeexplore.ieee.org>.

Digital Object Identifier 10.1109/TIT.2009.2032856

only one bit of information. We study how to design the vector quantizer for each sensor, aimed at achieving the best estimation performance at the FC. Specifically, as in [12], we consider hyperplane-based vector quantization which utilizes a compression vector to convert the high-dimensional observation vector into a one-dimensional scalar, and then compares the resultant scalar with a quantization threshold to generate one bit quantized data. Our CRB analysis shows that the estimation performance is dependent on the quantization thresholds as well as the compression vectors. The optimal choice of the quantization thresholds, as in the scalar case, is dependent on the unknown parameter. In contrast, the design of the compression vectors is independent of the unknown parameter and hence is the focus of this paper. We develop an efficient iterative algorithm for the compression vector design for a general case and propose optimal/near-optimal solutions for some specific but important noise scenarios. Specifically, for a homogeneous environment where all sensors have identical noise covariance matrix, our results reveal that the compression vectors should be chosen from the eigenvectors of the noise covariance matrix, and the number of sensors selecting the same eigenvector as their compression vectors should be matched to their corresponding eigenvalue. Our performance analysis shows that our proposed one bit quantization scheme generally outperforms [12] in a rate distortion sense. In particular, when the eigenvalues of the noise covariance matrix are sharply diverse, it is even possible to achieve almost the same estimation performance as that of [12], while transmitting only $1/p$ times the total number bits required by [12]. Simulation results are presented to corroborate our theoretical results and to illustrate the performance of our proposed scheme.

The following notations are adopted throughout this paper, where $[\cdot]^T$ stands for transpose, $\text{tr}(\mathbf{A})$ 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} , $\mathbb{R}^{n \times m}$ and \mathbb{R}^n denote 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 estimation problem. The ML estimator is developed and the corresponding CRB is analyzed in Section III. Based on the CRB analysis, we investigate the vector quantization problem for distributed estimation. Section IV studies the optimum choice of the quantization threshold, and the design of the compression vectors is examined in Section V. Some discussions are provided in VI. Performance analysis and simulation results are provided in Section VII, followed by concluding remarks in Section VIII.

II. PROBLEM FORMULATION

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

$$\mathbf{x}_n = \boldsymbol{\theta} + \mathbf{w}_n, \quad n = 1, \dots, N \quad (1)$$

where $\mathbf{w}_n \in \mathbb{R}^p$ denotes the additive multivariate Gaussian noise with zero mean and autocovariance matrix $\mathbf{R}_{w,n}$, and the

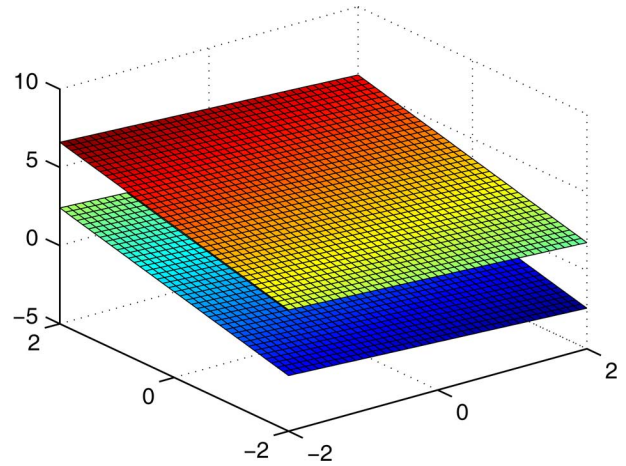


Fig. 1. Two parallel hyperplanes generated by the same compression vector but different thresholds, from which we can see that the compression vector determines the orientation of a hyperplane and the threshold controls its altitude.

noise is assumed independent (but not necessarily identically distributed) across sensors. The observation matrix \mathbf{H}_n defining the input/output relation: $\mathbf{x}_n = \mathbf{H}_n \boldsymbol{\theta} + \mathbf{w}_n$ is assumed equal to \mathbf{I} in the above model because the multiplicative effect of the observation matrix can be removed by carrying out a matrix inverse as long as \mathbf{H}_n is full column rank, which is usually met in practice. To meet stringent bandwidth/power budgets in WSNs, we consider the case where each sensor quantizes its vector observation into one bit binary data b_n which is sent to the FC to form an estimate of $\boldsymbol{\theta}$. The problem of interest is to determine the vector quantizers for each sensor, and to develop a maximum likelihood (ML) estimator to estimate $\boldsymbol{\theta}$ given $\{b_n\}_{n=1}^N$ for the FC.

Basically, vector quantization can be viewed as a space partitioning problem. For each sensor, the binary data b_n is given by

$$b_n = \mathbf{1}\{\mathbf{x}_n \in B_n\} \quad (2)$$

where b_n takes the value 1 when \mathbf{x}_n belongs to the region $B_n \subset \mathbb{R}^p$, and 0 otherwise. In this paper, to simplify the problem, the region B_n is confined to be a half-space whose border is a hyperplane defined by a compression vector \mathbf{c}_n and a quantization threshold τ_n , i.e., (see Fig. 1)

$$B_n = \{\mathbf{x} \in \mathbb{R}^p \mid \mathbf{c}_n^T \mathbf{x} > \tau_n\}. \quad (3)$$

The vector quantization problem therefore reduces to finding a set of compression vectors $\{\mathbf{c}_n\}$ and thresholds $\{\tau_n\}$. Another important reason for us to consider half-space regions, as we will discuss later, is that the likelihood function of $\boldsymbol{\theta}$ given $\{b_n\}_{n=1}^N$ is concave when a hyperplane-based quantizer (3) is adopted. Thus the ML estimation of $\boldsymbol{\theta}$ is a well-behaved numerical problem: any gradient-based search starting from a random initial estimate is guaranteed to converge to the global maximum, and many efficient routines exist for this type of work (e.g., [14]).

In the following, we will firstly develop the ML estimator and carry out a corresponding CRB analysis. The vector quantization design is then studied based on the CRB matrix of the unknown vector parameter $\boldsymbol{\theta}$.

III. MLE AND CRB ANALYSIS

A. MLE

By combining (2) and (3), we have

$$b_n = \text{sgn}(\mathbf{c}_n^T \mathbf{x}_n - \tau_n), \quad n = 1, 2, \dots, N \quad (4)$$

where

$$\begin{cases} \text{sgn}\{x\} = 1, & \text{if } x > 0 \\ \text{sgn}\{x\} = 0, & \text{otherwise.} \end{cases}$$

It can be readily shown that the probability mass function (PMF) of b_n is given by

$$P(b_n; \boldsymbol{\theta}) = [F_{v_n}(\tau_n - \mathbf{c}_n^T \boldsymbol{\theta})]^{b_n} [1 - F_{v_n}(\tau_n - \mathbf{c}_n^T \boldsymbol{\theta})]^{1-b_n} \quad (5)$$

where $F_{v_n}(x)$ denotes the complementary cumulative density function (CCDF) of $v_n \triangleq \mathbf{c}_n^T \mathbf{w}_n$, and v_n is a Gaussian random variable with zero mean and variance $\sigma_{v_n}^2 \triangleq \mathbf{c}_n^T \mathbf{R}_{w,n} \mathbf{c}_n$. Since $\{b_n\}$ are independent, the log-PMF or log-likelihood function is

$$\begin{aligned} L(\boldsymbol{\theta}) &\triangleq \log P(b_1, \dots, b_N; \boldsymbol{\theta}) \\ &= \sum_{n=1}^N \{b_n \log [F_{v_n}(\tau_n - \mathbf{c}_n^T \boldsymbol{\theta})] \\ &\quad + (1 - b_n) \log [1 - F_{v_n}(\tau_n - \mathbf{c}_n^T \boldsymbol{\theta})]\}. \end{aligned} \quad (6)$$

The ML estimate of $\boldsymbol{\theta}$, therefore, is given as

$$\hat{\boldsymbol{\theta}} = \arg \max_{\boldsymbol{\theta}} L(\boldsymbol{\theta}). \quad (7)$$

Although the ML estimation often suffers from drawbacks of high computational complexity and local maxima, this is not true for our case. In fact, it can be proved that the log-likelihood function $L(\boldsymbol{\theta})$ is a concave function. Hence computationally efficient search algorithms can be used to find the global maximum. The proof of the concavity is given in Appendix A. One can also refer to [12] for the concavity proof in a more general context.

B. CRB

We carry out a CRB analysis of our proposed scheme. Through the CRB analysis, we can gain insight into the vector quantizer design and examine the impact of the choice of $\{\mathbf{c}_n\}$ and $\{\tau_n\}$ on the estimation performance.

Let us define a new variable $z_n \triangleq \mathbf{c}_n^T \boldsymbol{\theta}$ and define

$$l(z_n) \triangleq b_n \log [F_{v_n}(\tau_n - z_n)] + (1 - b_n) \log [1 - F_{v_n}(\tau_n - z_n)]. \quad (8)$$

The first and second-order derivative of $L(\boldsymbol{\theta})$ are given by

$$\frac{\partial L(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = \sum_{n=1}^N \frac{\partial l(z_n)}{\partial z_n} \frac{\partial z_n}{\partial \boldsymbol{\theta}} = \sum_{n=1}^N \frac{\partial l(z_n)}{\partial z_n} \mathbf{c}_n \quad (9)$$

and

$$\begin{aligned} \frac{\partial^2 L(\boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^T} &= \sum_{n=1}^N \mathbf{c}_n \frac{\partial^2 l(z_n)}{\partial z_n^2} \frac{\partial z_n}{\partial \boldsymbol{\theta}^T} \\ &= \sum_{n=1}^N \frac{\partial^2 l(z_n)}{\partial z_n^2} \mathbf{c}_n \mathbf{c}_n^T \end{aligned} \quad (10)$$

where $\frac{\partial^2 l(z_n)}{\partial z_n^2}$ has been derived in many studies on distributed quantization of a scalar parameter, e.g., [11], and is given by

$$\begin{aligned} \frac{\partial^2 l(z_n)}{\partial z_n^2} &= b_n \left[-\frac{p'_{v_n}(\tau_n - z_n)}{F_{v_n}(\tau_n - z_n)} - \frac{p_{v_n}^2(\tau_n - z_n)}{F_{v_n}^2(\tau_n - z_n)} \right] \\ &\quad + (1 - b_n) \left[\frac{p'_{v_n}(\tau_n - z_n)}{[1 - F_{v_n}(\tau_n - z_n)]} \right. \\ &\quad \left. - \frac{p_{v_n}^2(\tau_n - z_n)}{[1 - F_{v_n}(\tau_n - z_n)]^2} \right] \end{aligned} \quad (11)$$

in which $p_{v_n}(x)$ denotes the probability density function (PDF) of v_n , and $p'_{v_n}(x) \triangleq \frac{\partial p_{v_n}(x)}{\partial x}$. The Fisher information matrix (FIM) of the estimation problem, therefore, is given as [15]

$$\begin{aligned} \mathbf{J}(\boldsymbol{\theta}) &= -E \left[\frac{\partial^2 L(\boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^T} \right] \\ &= -\sum_{n=1}^N E_{b_n} \left[\frac{\partial^2 l(z_n)}{\partial z_n^2} \right] \mathbf{c}_n \mathbf{c}_n^T \\ &\stackrel{(a)}{=} \sum_{n=1}^N \frac{p_{v_n}^2(\tau_n - \mathbf{c}_n^T \boldsymbol{\theta})}{F_{v_n}(\tau_n - \mathbf{c}_n^T \boldsymbol{\theta}) [1 - F_{v_n}(\tau_n - \mathbf{c}_n^T \boldsymbol{\theta})]} \mathbf{c}_n \mathbf{c}_n^T \end{aligned} \quad (12)$$

where E_{b_n} denotes the expectation with respect to the distribution of b_n , (a) follows from the fact that b_n is a binary random variable with $P(b_n = 1 | \tau_n, z_n) = F_{v_n}(\tau_n - z_n)$ and $P(b_n = 0 | \tau_n, z_n) = 1 - F_{v_n}(\tau_n - z_n)$.

By defining

$$g_n(\tau_n, \mathbf{c}_n) \triangleq \frac{p_{v_n}^2(\tau_n - \mathbf{c}_n^T \boldsymbol{\theta})}{F_{v_n}(\tau_n - \mathbf{c}_n^T \boldsymbol{\theta}) [1 - F_{v_n}(\tau_n - \mathbf{c}_n^T \boldsymbol{\theta})]} \quad (13)$$

the FIM (12) can be re-expressed as

$$\mathbf{J}(\boldsymbol{\theta}) = \sum_{n=1}^N g_n(\tau_n, \mathbf{c}_n) \mathbf{c}_n \mathbf{c}_n^T \quad (14)$$

and consequently the CRB matrix is

$$\text{CRB}(\boldsymbol{\theta}) = \mathbf{J}^{-1}(\boldsymbol{\theta}) = \left(\sum_{n=1}^N g_n(\tau_n, \mathbf{c}_n) \mathbf{c}_n \mathbf{c}_n^T \right)^{-1} \quad (15)$$

The CRB places a lower bound on the estimation error of any unbiased estimator and is asymptotically attained by the ML estimator [15]. Specifically, the covariance matrix of any unbiased estimate $\hat{\boldsymbol{\theta}}$ satisfies: $\text{cov}(\hat{\boldsymbol{\theta}}) - \text{CRB}(\boldsymbol{\theta}) \succeq \mathbf{0}$. Also, the variance of each component is bounded by the corresponding diagonal element of $\text{CRB}(\boldsymbol{\theta})$, i.e., $\text{var}(\hat{\theta}_i) \geq [\text{CRB}(\boldsymbol{\theta})]_{ii}$. It is observed

from (15) that the CRB matrix of $\boldsymbol{\theta}$ depends on the compression vectors $\{\mathbf{c}_n\}$ and the quantization thresholds $\{\tau_n\}$. Naturally, we may wish to optimize $\{\mathbf{c}_n\}$ and $\{\tau_n\}$ by minimizing the trace of the CRB matrix, i.e., the overall estimation error asymptotically achieved by the ML estimator. The optimization therefore is formulated as follows (note that minimizing the overall estimation error is a natural objective which has been widely adopted in distributed vector parameter estimation, e.g., [13], [16]–[19]. A generalization of this optimization by placing different weights on parameter components will be discussed in Section VI-B)

$$\min_{\{\mathbf{c}_n\}, \{\tau_n\}} \text{tr}\{\text{CRB}(\boldsymbol{\theta})\} = \text{tr}\left\{\left(\sum_{n=1}^N g_n(\tau_n, \mathbf{c}_n) \mathbf{c}_n \mathbf{c}_n^T\right)^{-1}\right\}. \quad (16)$$

Such an optimization is examined in the following section, where it is shown that the optimization of the compression vectors $\{\mathbf{c}_n\}$ can be decoupled from the choice of the thresholds $\{\tau_n\}$ and is the key to the vector quantization design.

IV. VECTOR QUANTIZATION DESIGN: THRESHOLD DETERMINATION

Note that for the Gaussian random variable v_n , $g_n(\tau_n, \mathbf{c}_n)$ defined in (13) is a unimodal, positive and symmetric function attaining its maximum when $\tau_n = \mathbf{c}_n^T \boldsymbol{\theta}$. Hence, given a fixed set of compression vectors $\{\mathbf{c}_n\}$, the optimal quantization thresholds conditional on $\{\mathbf{c}_n\}$ can be readily solved from (16) and are given by

$$\tau_n^* = \mathbf{c}_n^T \boldsymbol{\theta}, \quad \forall n \in \{1, \dots, N\}. \quad (17)$$

The result (17) comes directly by noting that

$$\sum_{n=1}^N g_n(\tau_n^*, \mathbf{c}_n) \mathbf{c}_n \mathbf{c}_n^T - \sum_{n=1}^N g_n(\tau_n, \mathbf{c}_n) \mathbf{c}_n \mathbf{c}_n^T \succeq \mathbf{0} \quad (18)$$

and resorting to the convexity of $\text{tr}(\mathbf{A}^{-1})$ over the set of positive definite matrix, i.e., for any $\mathbf{A} \succ \mathbf{0}$, $\mathbf{B} \succ \mathbf{0}$, and $\mathbf{A} - \mathbf{B} \succeq \mathbf{0}$, the following inequality $\text{tr}(\mathbf{A}^{-1}) \leq \text{tr}(\mathbf{B}^{-1})$ holds (see [20]).

We see that, as in the scalar parameter case, the optimal choice of the quantization threshold τ_n is dependent on the parameter $\boldsymbol{\theta}$. If $\boldsymbol{\theta}$ is perfectly known, then the optimization (16) simply reduces to finding a set of compression vectors $\{\mathbf{c}_n\}$, with $\tau_n = \mathbf{c}_n^T \boldsymbol{\theta}$, i.e.

$$\begin{aligned} \min_{\{\mathbf{c}_n\}} \quad & \text{tr}\left\{\left(\sum_{n=1}^N g_n(\tau_n, \mathbf{c}_n) \mathbf{c}_n \mathbf{c}_n^T\right)^{-1}\right\} \\ \text{s.t.} \quad & \tau_n = \mathbf{c}_n^T \boldsymbol{\theta}, \quad \forall n. \end{aligned} \quad (19)$$

By noting that

$$\begin{aligned} g_n(\tau_n^*, \mathbf{c}_n) &= \frac{p_{v_n}^2(0)}{F_{v_n}(0)[1 - F_{v_n}(0)]} = \frac{2}{\pi \sigma_{v_n}^2} \\ &= \frac{2}{\pi \mathbf{c}_n^T \mathbf{R}_{w_n} \mathbf{c}_n} \end{aligned} \quad (20)$$

(16) (or (19)) becomes an optimization independent of $\boldsymbol{\theta}$

$$\min_{\{\mathbf{c}_n\}} \text{tr}\left\{\left(\sum_{n=1}^N \frac{\mathbf{c}_n \mathbf{c}_n^T}{\mathbf{c}_n^T \mathbf{R}_{w_n} \mathbf{c}_n}\right)^{-1}\right\}. \quad (21)$$

The problem lies in that the vector parameter $\boldsymbol{\theta}$, to be estimated, is unknown and unusable in practice. Hence the choice of the thresholds $\{\tau_n\}$ is tricky. There are two strategies to address this difficulty, which are described in the following subsections. We also derive the corresponding optimization formulation of the compression vector design under these two strategies.

A. FC Feedback-Based Iterative Algorithm

One strategy is to use a FC feedback-based iterative algorithm [7], [11], [21] in which the thresholds are iteratively refined by the FC based on the previous estimate. Specifically, at iteration i , the FC assigns the quantization thresholds $\{\tau_n^{(i)}\}_{n=1}^N$ to the sensors. With this assigned quantization threshold, each sensor generates its quantized data $b_n^{(i)}$ and reports back to the FC. Upon receiving the quantized data $\{b_n^{(i)}\}_{n=1}^N$, the FC can compute an estimate $\hat{\boldsymbol{\theta}}^{(i)}$ from the ML estimator (7). This ML estimate is then plugged in (17) to obtain updated quantization thresholds, i.e., $\tau_n^{(i+1)} = \mathbf{c}_n^T \hat{\boldsymbol{\theta}}^{(i)}$, which are assigned to the sensors for subsequent iteration. Note that when computing an ML estimate $\hat{\boldsymbol{\theta}}^{(i)}$, not only the quantized data from the current but also from all previous iterations can be used (the ML estimator (7) can be easily adapted to accommodate these quantized data since the data are independent across different iterations). Due to the consistency of the ML estimator for large data records, this iterative process will asymptotically lead to conditional optimal quantization thresholds, i.e., $\tau_n^{(i)} \xrightarrow{t \rightarrow \infty} \mathbf{c}_n^T \boldsymbol{\theta}$. A rigorous proof of this asymptotic optimality was provided in [11], where a similar feedback-based iterative algorithm was proposed to adjust the quantization thresholds. Given this asymptotic optimality, the problem of interest, therefore, is to determine the set of compression vectors by assuming the quantization threshold attaining their conditional optimal values $\{\mathbf{c}_n^T \boldsymbol{\theta}\}$. Hence we are faced with the optimization (21).

Note that the proposed feedback-based iterative algorithm is based on the assumption that the parameter to be estimated is constant over time. In practice, the observed parameter, for example, the intensity of a physical phenomenon like temperature or humidity, can be time-varying. For time-varying environments, the algorithm still works if the parameter undergoes a negligible variation throughout the iterative process. This is the case when the variation of the environments is slow relative to the iterative process. In view of this, we can use some means such as the multiple access technique to accelerate the iterative process. Another solution to cope with this issue is to collect and store a set of independent observation samples over a short duration (the parameter is considered static over this period) at each sensor. The iterative algorithm can work offline by using these stored data without collecting the current observation samples.

B. Heuristic Approach for Threshold Determination

Another strategy, without resorting to the iterative algorithm, determines the thresholds based on some prior knowledge of $\boldsymbol{\theta}$, e.g., [6], [7]. Suppose that the unknown vector parameter $\boldsymbol{\theta}$ is bounded between $\boldsymbol{\theta}_{\text{ub}} > \boldsymbol{\theta} > -\boldsymbol{\theta}_{\text{ub}}$ (note that we do not assume the prior distribution of $\boldsymbol{\theta}$ since we treat $\boldsymbol{\theta}$ as a deterministic vector parameter). It can be readily verified that

$$\|\mathbf{c}_n\|_1 \max(|\boldsymbol{\theta}_{\text{ub}}|) > \mathbf{c}_n^T \boldsymbol{\theta} > -\|\mathbf{c}_n\|_1 \max(|\boldsymbol{\theta}_{\text{ub}}|) \quad (22)$$

where $\|\cdot\|_1$ is the vector-1 norm, $\max(|\boldsymbol{\theta}_{\text{ub}}|)$ denotes the maximum element of the absolute value of the vector $\boldsymbol{\theta}_{\text{ub}}$. As in [6], the threshold τ_n can be chosen within the range defined in (22) according to a certain heuristic rule. For example, τ_n can be a value selected uniformly at random from the range, in the hope that some thresholds are close to its optimal value $\mathbf{c}_n^T \boldsymbol{\theta}$. Obviously the thresholds determined based on the bound of $\boldsymbol{\theta}$ are not optimal, and the function $g_n(\tau_n, \mathbf{c}_n)$ does not lead to a simple analytical form since generally $\tau_n \neq \mathbf{c}_n^T \boldsymbol{\theta}$. In this case, optimizing the compression vectors $\{\mathbf{c}_n\}$ is a tricky problem due to the irreducible form of the function $g_n(\tau_n, \mathbf{c}_n)$. Nevertheless, we can simplify the problem by considering minimizing an upper bound of the objective function of (16). It can be shown that the function $g_n(\tau_n, \mathbf{c}_n)$ is lower bounded by (see Appendix B for a detailed derivation)

$$g_n(\tau_n, \mathbf{c}_n) > \frac{t_n}{\mathbf{c}_n^T \mathbf{R}_{w_n} \mathbf{c}_n} \quad (23)$$

where

$$t_n \triangleq \frac{2}{\pi} \exp\left(-\frac{\delta^2}{2\lambda_{\min}(\mathbf{R}_{w_n})}\right) \quad (24)$$

is a coefficient independent of the vector parameter $\boldsymbol{\theta}$ and the optimization variables \mathbf{c}_n and τ_n , in which $\lambda_{\min}(\mathbf{R}_{w_n})$ denotes the smallest eigenvalue of \mathbf{R}_{w_n} , and $\delta \triangleq 2\sqrt{p} \max(|\boldsymbol{\theta}_{\text{ub}}|)$. Since $\text{tr}(\mathbf{A}^{-1})$ is convex over the set of positive definite matrix and

$$\sum_{n=1}^N g_n(\tau_n, \mathbf{c}_n) \mathbf{c}_n \mathbf{c}_n^T - \sum_{n=1}^N \frac{t_n \mathbf{c}_n \mathbf{c}_n^T}{\mathbf{c}_n^T \mathbf{R}_{w_n} \mathbf{c}_n} \succ \mathbf{0} \quad (25)$$

the objective function of (16) is upper bounded by

$$\text{tr} \left\{ \left(\sum_{n=1}^N g_n(\tau_n, \mathbf{c}_n) \mathbf{c}_n \mathbf{c}_n^T \right)^{-1} \right\} < \text{tr} \left\{ \left(\sum_{n=1}^N \frac{t_n \mathbf{c}_n \mathbf{c}_n^T}{\mathbf{c}_n^T \mathbf{R}_{w_n} \mathbf{c}_n} \right)^{-1} \right\}. \quad (26)$$

Finding the compression vectors $\{\mathbf{c}_n\}$ by minimizing the upper bound (26), therefore, leads to the following optimization

$$\min_{\{\mathbf{c}_n\}} \text{tr} \left\{ \left(\sum_{n=1}^N \frac{t_n \mathbf{c}_n \mathbf{c}_n^T}{\mathbf{c}_n^T \mathbf{R}_{w_n} \mathbf{c}_n} \right)^{-1} \right\}. \quad (27)$$

Clearly (27) has a same formulation as (21) as the coefficient t_n can be absorbed into the covariance matrix \mathbf{R}_{w_n} .

Through the above analysis, we see that for both strategies where the thresholds are chosen by using an iterative algorithm

or by a heuristic approach based on a prior knowledge of the unknown vector parameter, (16) is reduced to a same optimization problem (21), where the design of the compression vectors is independent of the unknown vector parameter. We will study the compression vector optimization problem (21) in the following section.

V. VECTOR QUANTIZATION DESIGN: COMPRESSION VECTOR DESIGN

In this section, we first develop an efficient iterative algorithm for compression vector design under a general case where the observation noise is independent but not necessarily identically distributed across sensors. We then examine some specific but important homogeneous and inhomogeneous scenarios where sensors have identical noise covariance matrices or the noise covariance matrices have the same correlation structure with different scaling factors. For these scenarios, analytical optimal/near-optimal solutions can be found and these solutions provide an insight into the compression vector design problem.

A. An Iterative Algorithm

The optimization (21) involves determining N compression vectors. Clearly, joint searching over the N compression vectors is practically infeasible since it has a complexity that grows exponential with N . To simplify the problem, we employ a Gauss-Seidel iterative technique [22] to reduce the number of optimization variables. Specifically, we study how to determine the k th compression vector \mathbf{c}_k when the remaining $(N-1)$ compression vectors are fixed, through which we can develop an efficient iterative algorithm to search for an effective, albeit suboptimal, solution. Let

$$\mathbf{Q}_k \triangleq \sum_{n \neq k} \frac{\mathbf{c}_n \mathbf{c}_n^T}{\mathbf{c}_n^T \mathbf{R}_{w_n} \mathbf{c}_n}. \quad (28)$$

The optimization of \mathbf{c}_k given fixed $\{\mathbf{c}_n\}_{n \neq k}$ is formulated as

$$\min_{\mathbf{c}_k} \text{tr} \left\{ \left(\mathbf{Q}_k + \frac{\mathbf{c}_k \mathbf{c}_k^T}{\mathbf{c}_k^T \mathbf{R}_{w_k} \mathbf{c}_k} \right)^{-1} \right\}. \quad (29)$$

Recalling the Woodbury identity, the objective function of (29) can be rewritten as

$$\begin{aligned} & \text{tr} \left\{ \left(\mathbf{Q}_k + \frac{\mathbf{c}_k \mathbf{c}_k^T}{\mathbf{c}_k^T \mathbf{R}_{w_k} \mathbf{c}_k} \right)^{-1} \right\} \\ &= \text{tr} \left\{ \mathbf{Q}_k^{-1} - \mathbf{Q}_k^{-1} \mathbf{c}_k [\mathbf{c}_k^T \mathbf{R}_{w_k} \mathbf{c}_k + \mathbf{c}_k^T \mathbf{Q}_k^{-1} \mathbf{c}_k]^{-1} \right. \\ & \quad \left. \times \mathbf{c}_k^T \mathbf{Q}_k^{-1} \right\} \\ &= \text{tr} \left\{ \mathbf{Q}_k^{-1} \right\} - \text{tr} \left\{ \frac{\mathbf{Q}_k^{-1} \mathbf{c}_k \mathbf{c}_k^T \mathbf{Q}_k^{-1}}{\mathbf{c}_k^T (\mathbf{R}_{w_k} + \mathbf{Q}_k^{-1}) \mathbf{c}_k} \right\} \\ &\stackrel{(a)}{=} \text{tr} \left\{ \mathbf{Q}_k^{-1} \right\} - \frac{\mathbf{c}_k^T \mathbf{Q}_k^{-1} \mathbf{Q}_k^{-1} \mathbf{c}_k}{\mathbf{c}_k^T (\mathbf{R}_{w_k} + \mathbf{Q}_k^{-1}) \mathbf{c}_k} \end{aligned} \quad (30)$$

where (a) comes from the trace identity $\text{tr}(\mathbf{AB}) = \text{tr}(\mathbf{BA})$. Since \mathbf{Q}_k is fixed, (29) becomes

$$\max_{\mathbf{c}_k} \frac{\mathbf{c}_k^T \mathbf{Q}_k^{-1} \mathbf{Q}_k^{-1} \mathbf{c}_k}{\mathbf{c}_k^T (\mathbf{R}_{w_k} + \mathbf{Q}_k^{-1}) \mathbf{c}_k}. \quad (31)$$

The above optimization is a generalized Rayleigh quotient problem and has a closed form solution which can be obtained as the eigenvector of $\mathbf{\Gamma}$ associated with the largest eigenvalue, where

$$\mathbf{\Gamma} \triangleq (\mathbf{R}_{w_k} + \mathbf{Q}_k^{-1})^{-\frac{1}{2}} \mathbf{Q}_k^{-1} \mathbf{Q}_k^{-1} (\mathbf{R}_{w_k} + \mathbf{Q}_k^{-1})^{-\frac{1}{2}}. \quad (32)$$

By using the above results, we can establish an iterative algorithm to solve (21) by successively optimizing and replacing each compression vector \mathbf{c}_k . The algorithm is summarized as follows.

- 1) Randomly generate a set of compression vectors $\{\mathbf{c}_n^{(0)}\}$ as an initialization.
- 2) At iteration $i + 1$ ($i = 0, 1, \dots$): determine $\mathbf{c}_1^{(i+1)}$ given: $\{\mathbf{c}_2^{(i)}, \dots, \mathbf{c}_N^{(i)}\}$; determine $\mathbf{c}_k^{(i+1)}$ given: $\{\mathbf{c}_1^{(i+1)}, \dots, \mathbf{c}_{k-1}^{(i+1)}, \mathbf{c}_{k+1}^{(i)}, \dots, \mathbf{c}_N^{(i)}\}$ for $k = 2, \dots, N$.
- 3) Go to Step 2 if $|f(\{\mathbf{c}_n^{(i+1)}\}) - f(\{\mathbf{c}_n^{(i)}\})| > \epsilon$, where $f(\cdot)$ denotes the objective function defined in (21), ϵ is a prescribed tolerance value; otherwise stop.

Clearly, in this algorithm, every iteration results in a non-increasing objective function value. In this manner, the iterative algorithm converges to a stationary point and finds an effective set of compression vectors. 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. In the following subsections, we will show that near-optimal or even optimal solutions can be obtained for some specific but important noise models that characterize most practical scenarios. These optimal/near-optimal solutions render an insight into the compression vector design and the theoretical analysis provides a fundamental understanding of the performance of our proposed one-bit quantization scheme.

Below are some other comments on the proposed iterative algorithm.

Remarks: Due to the nonlinearity of the objective function, a theoretical analysis of the convergence rate of the proposed iterative algorithm is difficult. Nevertheless, our numerical results reported are very encouraging: even with hundreds of sensors, the proposed iterative algorithm can usually achieve an acceptable performance within only a few iterations. A lower bound on the minimum achievable value of (21) for the general case can be derived as follows

$$\begin{aligned} & \text{tr} \left\{ \left(\sum_{n=1}^N \frac{\mathbf{c}_n \mathbf{c}_n^T}{\mathbf{c}_n^T \mathbf{R}_{w_n} \mathbf{c}_n} \right)^{-1} \right\} \\ & \stackrel{(a)}{\geq} \text{tr} \left\{ \left(\sum_{n=1}^N \frac{1}{\lambda_{\min}(\mathbf{R}_{w_n})} \frac{\mathbf{c}_n \mathbf{c}_n^T}{\mathbf{c}_n^T \mathbf{c}_n} \right)^{-1} \right\} \\ & \stackrel{(b)}{\geq} \frac{p^2}{\chi} \end{aligned} \quad (33)$$

where (a) comes by noting that

$$\sum_{n=1}^N \frac{1}{\lambda_{\min}(\mathbf{R}_{w_n})} \frac{\mathbf{c}_n \mathbf{c}_n^T}{\mathbf{c}_n^T \mathbf{c}_n} - \sum_{n=1}^N \frac{\mathbf{c}_n \mathbf{c}_n^T}{\mathbf{c}_n^T \mathbf{R}_{w_n} \mathbf{c}_n} \succeq \mathbf{0}. \quad (34)$$

(b) follows from Theorem 2 presented in Section V-C, in which

$$\chi \triangleq \sum_{n=1}^N \frac{1}{\lambda_{\min}(\mathbf{R}_{w_n})}. \quad (35)$$

Although generally non-achievable, this lower bound may still serve as a benchmark to evaluate the performance of the proposed iterative algorithm.

Since the proposed iterative algorithm successively optimizes the compression vector associated with each sensor, it is well suited for implementation in a distributed sequential process, in which sensors sequentially update their compression vectors and forward information to their next sensors. Specifically, during this process, each sensor, say sensor n , forwards a global information matrix $\mathbf{Q} \triangleq \sum_{n=1}^N \frac{\mathbf{c}_n \mathbf{c}_n^T}{\mathbf{c}_n^T \mathbf{R}_{w_n} \mathbf{c}_n}$ to its next sensor, say sensor $n + 1$. Based on this received global information matrix and its local information, the sensor $n + 1$ computes a new compression vector by solving (31), updates the global information matrix with this new compression vector, and forwards this updated global information matrix to its next sensor, and so on and so forth.

B. Optimal/Near-Optimal Solution: $\mathbf{R}_{w_n} = \sigma_w^2 \mathbf{I}$

We firstly consider the identical noise covariance matrix case with $\mathbf{R}_{w_n} = \sigma_w^2 \mathbf{I}$. In this case, it can be readily shown that (21) is equivalent to

$$\begin{aligned} & \min_{\{\mathbf{c}_n\}} \text{tr} \left\{ \left(\sum_{n=1}^N \mathbf{c}_n \mathbf{c}_n^T \right)^{-1} \right\} \\ & \text{s.t.} \quad \mathbf{c}_n^T \mathbf{c}_n = 1, \quad \forall n \end{aligned} \quad (36)$$

where all compression vectors are confined to unit norm vectors. Although the optimization (36) is generally nonconvex, a near-optimal solution or even an optimal solution can still be found. The results are summarized as follows.

Theorem 1: Suppose $N \geq p$. Let $f_1(\{\mathbf{c}_n\})$ denote the objective function (36). Then for any feasible solution of (36), we have

$$f_1(\{\mathbf{c}_n\}) \geq \frac{p^2}{N} \quad (37)$$

which places a lower bound on the minimum achievable value of (36). This lower bound can be approached or even reached by the compression vector design strategy proposed below, which attains an objective function value that is within a factor κ_1 of the lower bound p^2/N , where

$$\kappa_1 = \frac{N/p}{\text{floor}(N/p)} \quad (38)$$

in which $\text{floor}(x)$ stands for the floor operator rounding x to the nearest integer toward the negative infinity.

Proof: See Appendix C. \square

Proposed Strategy: We firstly randomly choose M sensors from these N sensors, where $M = \text{floor}(N/p)p$ is a multiple of p . The selected M sensors are then equally divided into p groups, with the sensors in the i th group using the i th column

of an orthonormal matrix $\mathbf{U} \in \mathbb{R}^{p \times p}$, i.e., $\mathbf{U}[:, i]$, as their compression vectors, where \mathbf{U} can be any orthonormal matrix. The compression vectors of the remaining $N - M$ sensors can be chosen to be any $N - M$ columns of the orthonormal matrix \mathbf{U} (note that $N - M < p$).

Remarks: When the number of sensors is a multiple of the vector parameter dimension, i.e., N/p is an integer, then $\kappa_1 = 1$. Clearly, in this case, the proposed strategy is an optimal solution to (36) since it achieves the lower bound. If N is not a multiple of p , κ_1 approaches to one as $N \gg p$. In this case, the proposed strategy provides near-optimal performance. Note that the condition $N \gg p$ is mild and can usually be met in practical applications.

To gain an insight into the proposed strategy, we choose the orthonormal matrix \mathbf{U} to be an identity matrix \mathbf{I} . In this case, the proposed strategy becomes an intuitive solution where the vector parameter estimation is decoupled into p scalar estimation tasks, with each task accomplished by each group of sensors independently. Our theorem shows that such an intuitive solution achieves near-optimal or even optimal performance.

C. Optimal/Near-Optimal Solution: $\mathbf{R}_{w_n} = \sigma_{w_n}^2 \mathbf{I}$

When the noise covariance matrices are nonidentical with $\mathbf{R}_{w_n} = \sigma_{w_n}^2 \mathbf{I}$, (21) becomes

$$\begin{aligned} \min_{\{\mathbf{c}_n\}} \quad & \text{tr} \left\{ \left(\sum_{n=1}^N \frac{1}{\sigma_{w_n}^2} \mathbf{c}_n \mathbf{c}_n^T \right)^{-1} \right\} \\ \text{s.t.} \quad & \mathbf{c}_n^T \mathbf{c}_n = 1, \quad \forall n. \end{aligned} \quad (39)$$

The optimization (39) can be analyzed by following a similar procedure as that of (36). We have the following results regarding (39).

Theorem 2: Suppose that $N \geq p$. Let $f_2(\{\mathbf{c}_n\})$ denote the objective function (39). Then for any feasible solution of (39), we have

$$f_2(\{\mathbf{c}_n\}) \geq \frac{p^2}{K} \quad (40)$$

where

$$K \triangleq \sum_{n=1}^N \frac{1}{\sigma_{w_n}^2} \quad (41)$$

(40) provides a lower bound on the minimum achievable value of (39). This lower bound can be approached or even reached by the compression vector design strategy described below, which attains an objective function value within a factor κ_2 of the lower bound p^2/K , where

$$\kappa_2 = \frac{K}{p \min(\{K_i\})} \quad (42)$$

where $\min(\{K_i\})$ denotes the minimum value of the set $\{K_i\}$, and K_i is defined in (43).

Proof: See Appendix D. \square

Proposed Strategy: We divide the sensors into p groups (not necessarily equally divided), with

$$\sum_{n=i_1}^{i_{N_i}} \frac{1}{\sigma_{w_n}^2} = K_i, \quad \forall i \in \{1, \dots, p\} \quad (43)$$

where $\{i_1, i_2, \dots, i_{N_i}\}$ are the indices of the sensors in group i , N_i denotes the number of sensors of group i . The sensors in group i choose the i th column of an orthonormal matrix $\mathbf{U} \in \mathbb{R}^{p \times p}$ to be their compression vectors, where \mathbf{U} can be any orthonormal matrix.

Remarks: Theorem 2 implies that a near-optimal solution can be obtained by properly partitioning the sensors such that $\min(\{K_i\})$ is close to the value K/p . In particular, when $K_1 = K_2 = \dots = K_p$, the proposed solution is optimum because $\kappa_2 = 1$ in this case. When $N \gg p$ and $1/\sigma_{w_n}^2 < K/p, \forall n \in \{1, \dots, N\}$, finding an equally or quasi-equally weighted partition is not difficult (here the term ‘‘quasi-equally weighted partition’’ means that the sensors are divided into p groups such that the weights $\{K_i\}_{i=1}^p$ associated with these p groups are approximately equal, i.e., $K_1 \approx K_2 \approx \dots \approx K_p$). There, of course, may be cases where an equally or quasi-equally weighted partition is not possible. In this case, Theorem 2 suggests to find a partition to maximize the minimum K_i over the set $\{K_i\}$ in order to minimize κ_2 .

Similarly as in last subsection, when $\mathbf{U} = \mathbf{I}$, sensors are divided into p groups and each group undertakes the task of estimating a scalar parameter. Nevertheless, sensors are no longer equally partitioned, instead, each sensor is weighted with a factor $1/\sigma_{w_n}^2$ and the partition of the sensors should ensure that all groups have an identical or roughly identical weighted summation. If we consider the factor $1/\sigma_{w_n}^2$ as a quantity to measure the sensor observation quality (the larger the value, the better the observation quality), this equally weighted summation partition is to seek p groups with identical observation qualities.

D. Optimal/Near-Optimal Solution: $\mathbf{R}_{w_n} = \mathbf{R}_w$

We now consider the scenario where sensors have identical but arbitrary noise covariance matrices. This scenario represents a general homogeneous environment allowing correlation among the noise multivariate components. The optimization (21) can be rewritten as

$$\begin{aligned} \min_{\{\mathbf{c}_n\}} \quad & \text{tr} \left\{ \left(\sum_{n=1}^N \frac{\mathbf{c}_n \mathbf{c}_n^T}{\mathbf{c}_n^T \mathbf{R}_w \mathbf{c}_n} \right)^{-1} \right\} \\ \text{s.t.} \quad & \mathbf{c}_n^T \mathbf{c}_n = 1, \quad \forall n. \end{aligned} \quad (44)$$

Clearly, (44) is more complicated than (36) and (39) since the denominator of the fraction in (44), i.e., $\mathbf{c}_n^T \mathbf{R}_w \mathbf{c}_n$, is no longer a constant independent of the choice of \mathbf{c}_n . Nevertheless, a near-optimal or even an optimal solution can still be obtained. In the following, to facilitate our presentation, we will first introduce our proposed compression vector design strategy.

Proposed Strategy: Let $\mathbf{R}_w = \mathbf{U}_w \mathbf{D}_w \mathbf{U}_w^T$ denote the eigenvalue decomposition (EVD) of \mathbf{R}_w , where $\mathbf{U}_w \in \mathbb{R}^{p \times p}$, $\mathbf{D}_w \in \mathbb{R}^{p \times p}$, and $d_{w,i}$ denotes the i th diagonal element of \mathbf{D}_w . Let

$$\chi_i^* \triangleq \frac{N \sqrt{d_{w,i}}}{\sum_{n=1}^p \sqrt{d_{w,n}}}, \quad \forall i \in \{1, \dots, p\}. \quad (45)$$

We divide the sensors into p groups, with each group consisting of N_i sensors and each sensor in group i using the i th column

of \mathbf{U}_w , i.e., $\mathbf{U}_w[:, i]$, as its compression vector, where N_i is obtained by rounding χ_i^* to its nearest integers (towards the positive or negative infinity) while preserving the summation of the set $\{\chi_i^*\}$, i.e., $\sum_{i=1}^p N_i = N$.

Clearly, the objective function value (44) achieved by the above proposed strategy is

$$\text{tr} \left\{ \left(\sum_{i=1}^p \frac{N_i}{d_{w,i}} \mathbf{U}_w[:, i] \mathbf{U}_w[:, i]^T \right)^{-1} \right\} = \sum_{i=1}^p \frac{d_{w,i}}{N_i}. \quad (46)$$

Some important results regarding the near-optimality of the proposed strategy are provided in the following theorem.

Theorem 3: Suppose $N \geq p$. Let $f_3(\{\mathbf{c}_n\})$ denote the objective function of (44). Then for any feasible solution of (44), we have

$$f_3(\{\mathbf{c}_n\}) \geq \frac{1}{N} \left[\sum_{i=1}^p \sqrt{d_{w,i}} \right]^2 \quad (47)$$

which places a lower bound on the minimum achievable function value of (44). This lower bound can be achieved by the above proposed strategy when $N_i = \chi_i^*$, i.e., when $\{\chi_i^*\}$ defined in (45) are exactly integers. If $\{\chi_i^*\}$ are not integers, then the strategy attains an objective function value within a factor κ_3 of the lower bound, where

$$\kappa_3 = 1 + \frac{1}{\min(\{\chi_i^*\})} \quad (48)$$

in which $\min(\{\chi_i^*\})$ denotes the minimum value among the set $\{\chi_i^*\}$.

Proof: See Appendix E. \square

Remarks: We can see that when $\{\chi_i^*\}$ defined in (45) are exactly integers, then $N_i = \chi_i^* \forall i$, and the proposed strategy is optimum. If $\{\chi_i^*\}$ are not integers, the proposed strategy attains an objective function value within the factor κ_3 of the lower bound, where the factor κ_3 approaches to one when the number of sensors, N , is sufficiently large. In this case, near-optimal performance can be achieved.

Unlike the previously proposed strategies that allow the compression vectors to be chosen from any arbitrary orthonormal matrix, the proposed strategy to (44) selects compression vectors from the eigenvectors of the noise covariance matrix \mathbf{R}_w , i.e., the hyperplanes are perpendicular to the eigenvectors of the noise covariance matrix. Note that although the same principle was adopted in [12] in choosing the hyperplanes, the rationales are different: our theoretical analysis reveals that selecting such perpendicular hyperplanes leads to optimal or near-optimal performance for our one-bit quantization scheme; whereas the proposed approach [12] employs perpendicular hyperplanes in order to obtain independent binary data generated by each sensor and its optimality for the approach [12] may not hold. Also, for our strategy, the number of sensors selecting a certain eigenvector of \mathbf{R}_w is proportional to the square root of the corresponding eigenvalue. Considering a diagonal \mathbf{R}_w where the eigenvalues are equivalent to the noise variances, our strategy suggests that more sensors should be used to estimate the scalar parameters with bad observation qualities.

E. Optimal/Near-Optimal Solution: $\mathbf{R}_{w_n} = \sigma_{w_n}^2 \mathbf{R}_w$

The scenario considered herein represents an inhomogeneous environment where the noise covariance matrices across the sensors have the same correlation structure but with different scaling factors. This modeling is a generalization of the previous case and can be used to characterize a broader range of practical applications. The optimization (21) is formulated as follows:

$$\begin{aligned} \min_{\{\mathbf{c}_n\}} \quad & \text{tr} \left\{ \left(\sum_{n=1}^N \frac{1}{\sigma_{w_n}^2} \frac{\mathbf{c}_n \mathbf{c}_n^T}{\mathbf{c}_n^T \mathbf{R}_w \mathbf{c}_n} \right)^{-1} \right\} \\ \text{s.t.} \quad & \mathbf{c}_n^T \mathbf{c}_n = 1, \quad \forall n \end{aligned} \quad (49)$$

which is a generalized problem of (44). By utilizing the results of Theorem 3, we can reach a similar strategy for the compression vector design.

Proposed Strategy: We divide sensors into p groups and let the sensors of group i select $\mathbf{U}_w[:, i]$ to be their compression vectors.

Let

$$K_i \triangleq \sum_{n=i_1}^{i_{N_i}} \frac{1}{\sigma_{w_n}^2}, \quad \forall i \in \{1, \dots, p\} \quad (50)$$

where $\{i_1, i_2, \dots, i_{N_i}\}$ are the indices of the sensors of group i , and N_i denotes the number of sensors in group i . It can be easily verified that the objective function value achieved by our proposed strategy is

$$\text{tr} \left\{ \left(\sum_{i=1}^p \frac{K_i}{d_{w,i}} \mathbf{U}_w[:, i] \mathbf{U}_w[:, i]^T \right)^{-1} \right\} = \sum_{i=1}^p \frac{d_{w,i}}{K_i} \quad (51)$$

which is a function of $\{K_i\}$. As we will show in the following theorem, this proposed strategy is able to achieve near-optimal even optimal performance by properly partitioning the sensors.

Theorem 4: Suppose $N \geq p$. Let $f_4(\{\mathbf{c}_n\})$ denote the objective function (49). For any feasible solution of (49), we have

$$f_4(\{\mathbf{c}_n\}) \geq \frac{1}{K} \left[\sum_{i=1}^p \sqrt{d_{w,i}} \right]^2 \quad (52)$$

which provides a lower bound on the minimum achievable function value of (49). This lower bound can be attained by the above proposed strategy when the following equality holds:

$$K_i = \frac{K \sqrt{d_{w,i}}}{\sum_{n=1}^p \sqrt{d_{w,n}}}, \quad \forall i \in \{1, \dots, p\} \quad (53)$$

where

$$K \triangleq \sum_{n=1}^N \frac{1}{\sigma_{w_n}^2} = \sum_{i=1}^p K_i \quad (54)$$

Proof: See Appendix F. \square

Remarks: We see that if sensors can be partitioned such that each group satisfies the equality (53), then the proposed strategy is an optimum solution of (49). Of course, finding a partition to

exactly satisfy (53) may not be possible in practice. Nevertheless, finding a set $\{K_i\}$ roughly equal to their optimal values is not difficult when $N \gg p$, in which case the proposed strategy achieves near-optimal performance.

As in our previous case, the proposed strategy selects the compression vector from the eigenvectors of the noise covariance matrix. Selection from any other orthonormal matrix does not lead to optimal or near-optimal performance. The difference between the current strategy and the previous one is that the current strategy requires a careful sensor partition (into p groups) to ensure that the value K_i associated with each group is equal to or close to its optimal value defined in (53), whereas the previous strategy can easily determine the sensor partition based on (45).

VI. DISCUSSIONS

A. Insight Into Hyperplane Design

We have examined the near-optimum design of the compression vectors $\{\mathbf{c}_n\}$ under different noise cases. In fact, it can be easily seen that the three scenarios discussed in Sections V-B–V-D are special cases of the scenario in Section V-E. Our theoretical results suggest that for those scenarios, the sensors should be properly partitioned into p groups, with the sensors in each group sharing a common compression vector which is selected from the eigenvectors of the noise covariance matrix. Note that for the case $\mathbf{R}_{w_n} = \sigma_w^2 \mathbf{I}$, \mathbf{U}_w obtained from the EVD of \mathbf{R}_{w_n} can be any orthonormal matrix. Hence their compression vectors can be chosen from any orthonormal matrix.

We now consider the correspondingly constructed hyperplanes. Recalling that the optimum choice of the threshold is given by (17), the hyperplane associated with each group therefore is

$$\{\mathbf{x} \in \mathbb{R}^p | \mathbf{U}_w[:, i]^T \mathbf{x} = \mathbf{U}_w[:, i]^T \boldsymbol{\theta}\}, \quad \forall i \in \{1, \dots, p\}. \quad (55)$$

To gain better insight into (55), let us consider a three-dimensional case where $\mathbf{U}_w = \mathbf{I}$ and $\boldsymbol{\theta} = [1 \ 1 \ 1]^T$. Fig. 2 depicts the three hyperplanes corresponding to three different groups of sensors. We see that the hyperplanes are mutually perpendicular and these hyperplanes intersect at exactly the point $\mathbf{x} = \boldsymbol{\theta}$. It can be readily verified that for any other \mathbf{U}_w (\mathbf{U}_w has to be an orthonormal matrix), the perpendicular property and the intersection point remain unaltered. For our one-bit quantization scheme, the above defined hyperplanes (55) for vector quantization achieves near-optimal or even optimal estimation performance, given that the sensors are properly partitioned and associated with the corresponding hyperplanes as described in our proposed strategies. This is an important result of this paper. As we mentioned earlier, although [12] adopts the same principle in choosing the hyperplanes, its purpose is to obtain independent binary data for each sensor and the near-optimality or optimality of the hyperplane selection rule was not provided and may not hold for the p -bit quantization scheme [12].

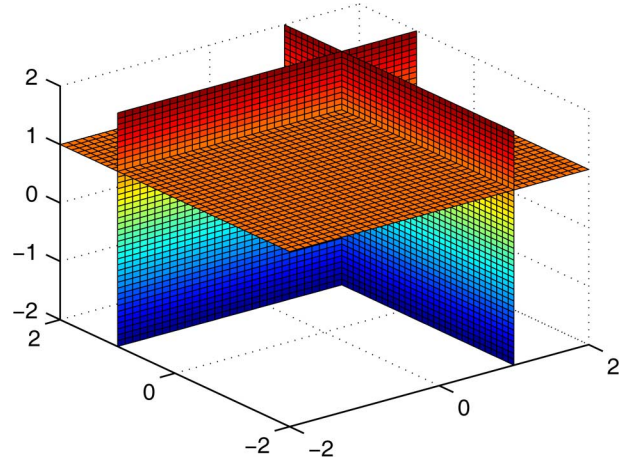


Fig. 2. An example of optimal/near-optimal design of the hyperplanes for our proposed one-bit quantization scheme. Each sensor, according certain strategy, selects one of these hyperplanes to quantize its local vector data. The hyperplanes are mutually perpendicular and these hyperplanes intersect at exactly the point $\mathbf{x} = \boldsymbol{\theta}$.

B. Generalization of the Optimization Criterion (16)

Note that 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 (16) to the following:

$$\begin{aligned} \min_{\{\mathbf{c}_n\}, \{\tau_n\}} \quad & \text{tr} \{ \text{CRB}(\boldsymbol{\theta}) \boldsymbol{\Phi} \} \\ & = \text{tr} \left\{ \left(\sum_{n=1}^N g_n(\tau_n, \mathbf{c}_n) \mathbf{c}_n \mathbf{c}_n^T \right)^{-1} \boldsymbol{\Phi} \right\} \end{aligned} \quad (56)$$

where $\boldsymbol{\Phi} \triangleq \text{diag}(\phi_1, \phi_2, \dots, \phi_p)$ is a diagonal matrix and ϕ_i is a positive weighting factor of user choice. The optimization (56) can be rewritten as

$$\min_{\{\mathbf{c}_n\}, \{\tau_n\}} \quad \text{tr} \left\{ \left(\sum_{n=1}^N g_n(\tau_n, \mathbf{c}_n) \boldsymbol{\Phi}^{-\frac{1}{2}} \mathbf{c}_n \mathbf{c}_n^T \boldsymbol{\Phi}^{-\frac{1}{2}} \right)^{-1} \right\}. \quad (57)$$

By following the same derivation, it can be easily verified that the optimal quantization thresholds conditional on $\{\mathbf{c}_n\}$ are still given by (17). Substituting (17) into (57), the design of the compression vectors therefore reduces to the following optimization:

$$\min_{\{\mathbf{c}_n\}} \quad \text{tr} \left\{ \left(\sum_{n=1}^N \frac{\boldsymbol{\Phi}^{-\frac{1}{2}} \mathbf{c}_n \mathbf{c}_n^T \boldsymbol{\Phi}^{-\frac{1}{2}}}{\mathbf{c}_n^T \mathbf{R}_{w_n} \mathbf{c}_n} \right)^{-1} \right\}. \quad (58)$$

Let $\tilde{\mathbf{c}}_n \triangleq \boldsymbol{\Phi}^{-\frac{1}{2}} \mathbf{c}_n$. We can reformulate (58) as

$$\min_{\{\tilde{\mathbf{c}}_n\}} \quad \text{tr} \left\{ \left(\sum_{n=1}^N \frac{\tilde{\mathbf{c}}_n \tilde{\mathbf{c}}_n^T}{\tilde{\mathbf{c}}_n^T \boldsymbol{\Phi}^{\frac{1}{2}} \mathbf{R}_{w_n} \boldsymbol{\Phi}^{\frac{1}{2}} \tilde{\mathbf{c}}_n} \right)^{-1} \right\} \quad (59)$$

which has a same formulation as (21) and hence can be solved by our proposed iterative algorithm and optimal/near-optimal strategies.

VII. PERFORMANCE ANALYSIS AND SIMULATION RESULTS

A. Performance Analysis

In this section, we evaluate the performance of our proposed hyperplane-based vector quantization scheme. We compare our scheme with the ML estimator using unquantized vector observations $\{\mathbf{x}_n\}$ (also referred to as the clairvoyant estimator) and the approach [12] where each sensor employs p hyperplanes to quantize the vector observation \mathbf{x}_n into p bits (instead of one bit in our scheme) of information, i.e., *one bit per sensor per dimension*. In [12], the p hyperplanes are chosen to be perpendicular to the eigenvectors of the noise covariance matrix in order to ensure that the resultant p binary data are independent. It is shown that such a choice of the hyperplanes is effective and given that the knowledge of $\boldsymbol{\theta}$ is available in selecting the thresholds,¹ the approach [12] suffers from a mild performance loss relative to the clairvoyant estimator using unquantized data, with the CRB increasing by only a factor of $\frac{\pi}{2}$, i.e.

$$\text{CRB}_{\text{Q-p}}(\boldsymbol{\theta}) = \frac{\pi}{2} \text{CRB}_{\text{CE}}(\boldsymbol{\theta}) \quad (60)$$

where we use the subscripts Q-p and CE to stand for the approach [12] and the clairvoyant estimator, respectively. Also, it can be readily verified that the CRB of the clairvoyant estimator using unquantized vector observations $\{\mathbf{x}_n\}$ is given by

$$\text{CRB}_{\text{CE}}(\boldsymbol{\theta}) = \left[\sum_{n=1}^N \mathbf{R}_{w_n}^{-1} \right]^{-1}. \quad (61)$$

Suppose that the noise covariance matrices across sensors are identical, i.e., $\mathbf{R}_{w_n} = \mathbf{R}_w \forall n$. Then the overall estimation errors asymptotically achieved by the clairvoyant estimator and the approach [12] are given as follows, respectively

$$\begin{aligned} \text{tr} \{ \text{CRB}_{\text{CE}}(\boldsymbol{\theta}) \} &= \text{tr} \left\{ \left[\sum_{n=1}^N \mathbf{R}_w^{-1} \right]^{-1} \right\} \\ &= \text{tr} \left\{ \left[\sum_{n=1}^N \mathbf{D}_w^{-1} \right]^{-1} \right\} \\ &= \frac{1}{N} \sum_{i=1}^p d_{w,i} \end{aligned} \quad (62)$$

$$\text{tr} \{ \text{CRB}_{\text{Q-p}}(\boldsymbol{\theta}) \} = \frac{\pi}{2N} \sum_{i=1}^p d_{w,i}. \quad (63)$$

We now consider the estimation performance of our proposed scheme. When the thresholds are optimum, we have the following by combining (16) and (20)

$$\text{tr} \{ \text{CRB}_{\text{Q-1}}(\boldsymbol{\theta}) \} = \frac{\pi}{2} \text{tr} \left\{ \left(\sum_{n=1}^N \frac{\mathbf{c}_n \mathbf{c}_n^T}{\mathbf{c}_n^T \mathbf{R}_w \mathbf{c}_n} \right)^{-1} \right\}$$

¹Note that both [12] and our scheme involve determining the thresholds whose optimal values are dependent on the unknown parameter. To focus our study on the evaluation of the compression vector design, we assume the knowledge of the parameter is available in choosing the optimal thresholds for both schemes.

$$\stackrel{(a)}{\geq} \frac{\pi}{2N} \left[\sum_{i=1}^p \sqrt{d_{w,i}} \right]^2 \quad (64)$$

where for clarity, we use the subscript $Q-1$ to represent our proposed scheme, and (a) comes by noting that the trace term on the right hand side of (64) is lower bounded by (47) (see Theorem 3). Also, this lower bound can be attained or approached with a negligible gap by the proposed compression vector design strategy, as we analyzed in Section V-D and will show later in our simulation results. Therefore the ratio of the overall estimation error of our scheme to that of [12] approaches to

$$\gamma_1 \triangleq \frac{\text{tr} \{ \text{CRB}_{\text{Q-1}}(\boldsymbol{\theta}) \}}{\text{tr} \{ \text{CRB}_{\text{Q-p}}(\boldsymbol{\theta}) \}} \rightarrow \frac{(\sum_{i=1}^p \sqrt{d_{w,i}})^2}{\sum_{i=1}^p d_{w,i}}. \quad (65)$$

By using the Cauchy-Schwarz inequality, it can be easily verified that for any set $\{d_{w,i}\}$ with $d_{w,i} > 0 \forall i$, we have $1 < \gamma_1 \leq p$, where γ_1 reaches its upper bound when all the eigenvalues of \mathbf{R}_w are identical, i.e., $d_{w,1} = \dots = d_{w,p}$. If the eigenvalues $\{d_{w,i}\}$ have diverse values, then the ratio γ_1 tends towards its lower bound one.

We see that as compared with [12], our proposed scheme results in an overall estimation error increase by a factor of γ_1 , where $1 < \gamma_1 \leq p$. Nevertheless, we note that the approach [12] requires each sensor to transmit p bits of information to the FC, whereas only one bit per sensor is sent to the FC for our scheme. Therefore, to meet a same overall estimation distortion target, the total number of bits required by our scheme is γ_1/p times the total number of bits needed by the approach [12]. Hence our scheme is generally more efficient than [12] in a rate distortion sense, especially when the eigenvalues of the noise covariance matrix vary in a sharp manner. For example, if there is a single dominant eigenvalue, then γ_1 approaches to one.

We now compare our scheme with the clairvoyant estimator. The ratio of the overall estimation error of our scheme to that of the clairvoyant estimator using raw data is given by

$$\gamma_2 \triangleq \frac{\text{tr} \{ \text{CRB}_{\text{Q-1}}(\boldsymbol{\theta}) \}}{\text{tr} \{ \text{CRB}_{\text{CE}}(\boldsymbol{\theta}) \}} = \frac{\pi \gamma_1}{2}. \quad (66)$$

Note that the clairvoyant estimator requires sending $(N \times p)$ real-valued messages while our scheme needs to transmit only N total bits of information. Suppose each real-valued message can be represented by q -bit binary data. To meet a same distortion target, the total number of bits sent by our scheme is $\pi \gamma_1 / 2pq$ times that needed by the clairvoyant estimator. In particular, when γ_1 is close to one, our analysis shows that sending one bit per sensor results in an overall estimation error increase by only a factor of about $\pi/2$.

In our above analysis, we considered the case of identical noise covariance matrices. It can be easily verified that for the inhomogeneous environments where $\mathbf{R}_{w_n} = \sigma_{w_n}^2 \mathbf{R}_w$, we have

$$\begin{aligned} \text{tr} \{ \text{CRB}_{\text{CE}}(\boldsymbol{\theta}) \} &= \text{tr} \left\{ \left[\sum_{n=1}^N \frac{1}{\sigma_{w_n}^2} \mathbf{R}_w^{-1} \right]^{-1} \right\} \\ &= \frac{1}{K} \sum_{i=1}^p d_{w,i} \end{aligned} \quad (67)$$

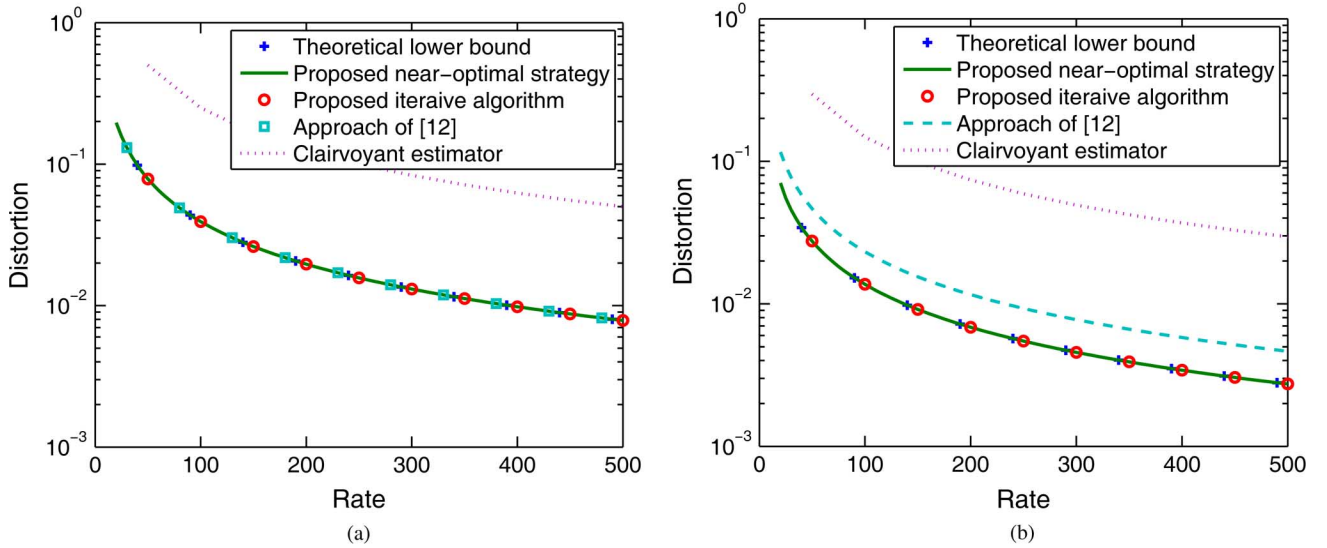


Fig. 3. Homogeneous environments: Rate-distortion performance of respective schemes. (a) Case of identical eigenvalues and (b) case of nonidentical eigenvalues.

$$\text{tr}\{\text{CRB}_{\text{Q-p}}(\boldsymbol{\theta})\} = \frac{\pi}{2K} \sum_{i=1}^p d_{w,i} \quad (68)$$

with $K \triangleq \sum_{n=1}^N \frac{1}{\sigma_{w_n}^2}$, and

$$\begin{aligned} \text{tr}\{\text{CRB}_{\text{Q-1}}(\boldsymbol{\theta})\} &= \frac{\pi}{2} \text{tr} \left\{ \left(\sum_{n=1}^N \frac{1}{\sigma_{w_n}^2} \frac{\mathbf{c}_n \mathbf{c}_n^T}{\mathbf{c}_n^T \mathbf{R}_w \mathbf{c}_n} \right)^{-1} \right\} \\ &\stackrel{(a)}{\rightarrow} \frac{\pi}{2K} \left[\sum_{i=1}^p \sqrt{d_{w,i}} \right]^2 \end{aligned} \quad (69)$$

where (a) comes from the result (52) derived in Theorem 4, and recognizing that this lower bound can be attained or approached with a negligible gap by the proposed compression vector design strategy. Based on (67)–(69), we see that our results (65)–(66) remain the same. For a more general case where \mathbf{R}_w cannot be expressed as $\sigma_{w_n}^2 \mathbf{R}_w$, a theoretical analysis is difficult to carry out because we do not have an analytical near-optimal solution. Nevertheless, numerical experiments can be conducted to compare the rate-distortion performance of our proposed iterative algorithm and the other two schemes.

B. Simulation Results

We now carry out experiments to corroborate our previous analysis and to illustrate the performance of our one-bit quantization scheme. We compare our scheme with the approach [12] and the clairvoyant estimator using unquantized data.

We firstly examine the rate-distortion performance of the three schemes in a homogeneous environment, i.e., identical noise covariance matrices across sensors. To investigate the rate-distortion performance of the clairvoyant estimator, we assume that each real-valued message can be represented by $q = 10$ bits binary data. Hence the clairvoyant estimator needs to send a total number of $10Np$ bits of information to the FC. To focus our study on the evaluation of the compression vector design, we assume the thresholds for our scheme and the approach

[12] are optimally selected. For our scheme, the compression vectors can be determined by the iterative algorithm proposed in Section V-A or by the optimal/near-optimal strategy proposed in Section V-D. We include the performance of both compression vector design strategies in our results. The theoretical lower bound for our scheme, which is given in (64), is also included for comparison. In our simulations, two different cases are considered: the former one corresponds to a case of equivalent eigenvalues ($\mathbf{R}_w = 0.1\mathbf{I}$) and the latter one corresponds to a case of diverse eigenvalues with the eigenvalues randomly generated according to $d_{w,i} = \alpha \nu_i \forall i \in \{1, \dots, p\}$ (the results are averaged over 500 Monte Carlo runs), where we set $p = 5$, $\alpha = 0.1$ and $\nu_i \sim \chi_1^2$ 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., [23]. 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 rate-distortion performance of the three schemes, where the x-coordinate represents the total number of bits used by each scheme, y-coordinate represents the asymptotically achieved overall estimation error, i.e., the trace of the CRB matrix. From Fig. 3, we see that our proposed optimal/near-optimal strategy, as analyzed, approaches the theoretical lower bound with a negligible gap, which corroborates our near-optimality of the proposed strategy. Also, it is interesting to observe that the proposed iterative algorithm achieves almost the same performance as that of the proposed optimal/near-optimal strategy. This implies that the iterative algorithm oftentimes converges to a stationary point that is close to the global minimum, although theoretically this convergence is not guaranteed. We also see that, when $\mathbf{R}_w = 0.1\mathbf{I}$, our scheme has a same rate-distortion performance as that of [12]. The reason is that the ratio γ_1/p is equal to one in the case of identical eigenvalues, as we discussed in the previous

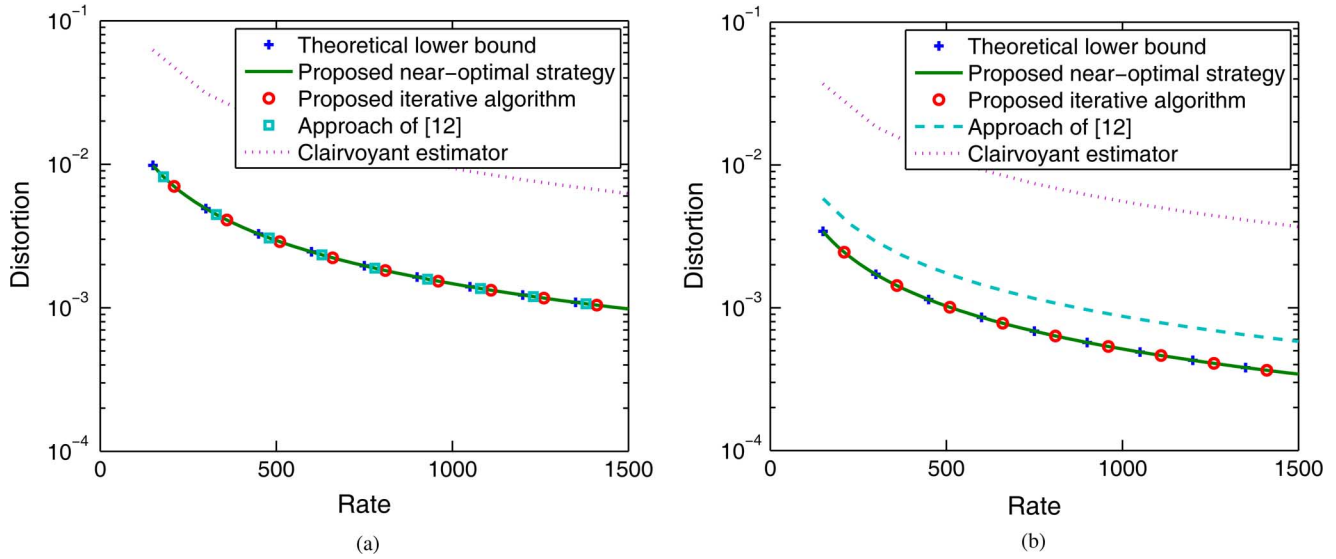


Fig. 4. Nonhomogeneous environments: Rate-distortion performance of respective schemes. (a) Case of identical eigenvalues and (b) case of nonidentical eigenvalues.

subsection. Nevertheless, when the eigenvalues becomes diverse, the ratio γ_1/p is strictly less than one and our scheme presents a performance advantage over [12]. To meet the same distortion target, say, 0.01, the rate required by our scheme is about 1/2 of that required by [12]. It can also be observed that our scheme presents a superior performance advantage over the clairvoyant estimator: our scheme needs transmitting much fewer bits of information than the clairvoyant estimator to attain a same distortion target. Hence considerable power/bandwidth savings can be achieved. The above rate-distortion results have twofold physical interpretations. On one hand, it tells us how many bits are needed by each scheme in order to achieve an estimation distortion target. On the other hand, another interesting perspective to look at these rate-distortion results is that given a total number of bits we can collect, say N bits, how should we distribute them among sensors? The three schemes provide three different strategies: our proposed scheme collects data from N sensors with one bit per sensor, the approach [12] collects data from N/p sensors with p bits per sensor, and the clairvoyant estimator requires the collection from $N/(pq)$ sensors with pq bits per sensor, assuming each raw message is represented by q bits. Our theoretical and numerical results point out the first allocation strategy can provide the best estimation performance.

To further corroborate our analysis, we conduct experiments to examine the rate-distortion performance in inhomogeneous environments with $\mathbf{R}_{w_n} = \sigma_{w_n}^2 \mathbf{R}_w$. 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 . The factor $\sigma_{c_m}^2$ for the three clusters is set to be 0.2, 0.5, and 1, respectively. Fig. 4 depicts the rate-distortion performance of the three schemes for a case of equivalent eigenvalues ($\mathbf{R}_w = 0.1\mathbf{I}$) and a case of diverse eigenvalues (\mathbf{R}_w is generated in a same way as described in the previous example). From Fig. 4, we see that, similar to the homogeneous environments, our scheme has a same rate-distortion performance as that of [12] when the eigenvalues are iden-

tical and strictly outperforms [12] for the case of diverse eigenvalues, which corroborates our theoretical analysis (67)–(69) in Section VII-A.

We now consider the scenario of generally nonidentical noise covariance matrices. As we mentioned earlier, we do not have an analytical optimal/near-optimal solution in this case, instead, the iterative algorithm proposed in Section V-A can be employed to find a sub-optimal set of compression vectors. In our simulations, the noise covariance matrix $\mathbf{R}_{w_n} \in \mathbb{R}^{p \times p}$ for each sensor is diagonal with its diagonal elements $[\mathbf{R}_{w_n}]_{ii} = \alpha v_i$, where $\alpha = 1$ and $v_i \sim \chi_1^2$, and p is set to be 5. The results are averaged over 500 Monte Carlo runs. Fig. 5 shows the overall estimation distortion (trace of the CRB matrix) of the three schemes as a function of the number of sensors, N . From Fig. 5, we see that our proposed iterative algorithm is very effective. It achieves almost the same estimation performance as that of [12], while transmitting only $1/p$ times the total number of bits required by [12]. Also, both our proposed scheme and the scheme [12] incur a mild performance loss compared with the clairvoyant estimator using raw data (note that as in previous example, we assume the thresholds for our scheme and [12] are optimally chosen).

We are also interested in examining the mean-square error (MSE) of the ML estimator for our scheme. We consider a homogeneous scenario where all sensors have identical noise covariance matrices which are given by $\mathbf{R}_w = 0.1\mathbf{I}$ (referred to as Case I) and $\mathbf{R}_w = \text{diag}\{0.8, 0.5, 0.2, 0.1, 0.1\}$ (referred to as Case II), respectively. The thresholds are optimally determined by assuming the knowledge of the unknown parameter and the compression vectors are designed by our proposed optimal/near-optimal strategy. Fig. 6 depicts the MSEs of the ML estimator for our scheme. It is observed that the MSEs approach the corresponding CRBs asymptotically with an increasing N for both cases.

So far we have thoroughly studied the performance of our proposed scheme by assuming the thresholds $\{\tau_n\}$ are optimally selected. In this case, the estimation performance only depends

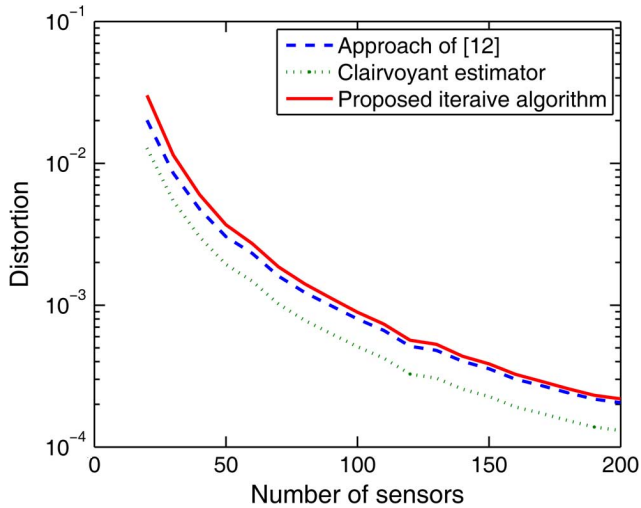


Fig. 5. Overall estimation error versus the number of sensors for three schemes.

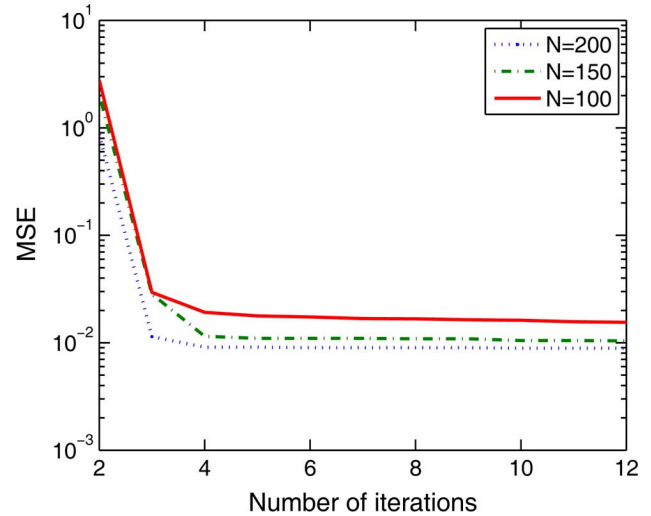


Fig. 7. MSE of FC-based feedback iterative algorithm for threshold determination.

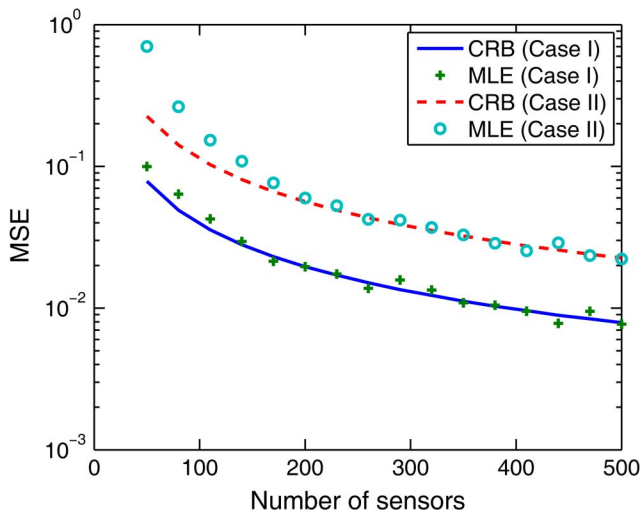


Fig. 6. MSEs of the MLE for our scheme under two different cases.

on the choice of the compression vectors $\{\mathbf{c}_n\}$. Since $\boldsymbol{\theta}$ to be estimated is unknown in practice, it is interesting to examine the performance of our proposed threshold determination strategy. We consider the FC feedback-based iterative algorithm introduced in Section IV-A which iteratively adjusts the threshold based on the ML estimate. In our simulations, we assume a homogeneous scenario with $\mathbf{R}_w = 0.1\mathbf{I}$. The dimension of the unknown parameter, p , is set to 3 and the number of sensors varies from 100 to 200. The compression vectors are determined by our optimal/near-optimal strategy proposed in Section V-D (note that since the design of the compression vectors is independent of the unknown parameter, they can be determined in advance). Fig. 7 shows the estimation MSE of the ML estimator versus the number of iterations. From Fig. 7, we see that the FC feedback-based iterative algorithm is very effective and can achieve an accurate estimate of the unknown parameter within several iterations.

VIII. CONCLUSION

In this paper, we studied the problem of vector quantization for distributed estimation in wireless sensor networks, where

each sensor quantizes its local vector observation into one bit of information which is sent to the FC to form a final estimate of the vector parameter. Specifically, we confine our studies to the hyperplane-based vector quantization whose design involves threshold determination and compression vector design. Our analysis reveals that the optimal choice of the thresholds, as in the scalar case, is dependent on the unknown parameter. In contrast, the compression vector design is independent of the vector parameter and can be determined prior to each estimation task. We developed an efficient iterative algorithm for the compression vector design for a general case and proposed optimal/near-optimal strategies for some specific but important noise scenarios. Our performance analysis shows that our proposed scheme generally outperforms existing schemes in a rate distortion sense, especially when the eigenvalues of the noise covariance matrix are sharply diverse. Simulation results were presented to corroborate our theoretical analysis and to illustrate the effectiveness of our proposed scheme. These results show that our proposed scheme is able to achieve almost the same estimation performance as some existing schemes with considerable rate/bandwidth savings.

APPENDIX A

PROOF OF CONCAVITY OF THE LOG-LIKELIHOOD FUNCTION (6)

It can be easily verified that $p_{v_n}(\tau_n - \mathbf{c}_n^T \boldsymbol{\theta})$ is log-concave in $\boldsymbol{\theta}$ since the Hessian matrix of $\log p_{v_n}(\tau_n - \mathbf{c}_n^T \boldsymbol{\theta})$, which is given by

$$\frac{\partial^2 \log p_{v_n}(\tau_n - \mathbf{c}_n^T \boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^T} = -\frac{\mathbf{c}_n \mathbf{c}_n^T}{\sigma_{v_n}^2} \quad (70)$$

is negative semidefinite. Consequently the corresponding cumulative density function (CDF) and complementary CDF (CCDF), which are integrals of the log-concave function $p_{v_n}(\tau_n - \mathbf{c}_n^T \boldsymbol{\theta})$ over convex sets $(-\infty, \tau_n)$ and (τ_n, ∞) respectively, are also log-concave, and their logarithms are concave. Since summation preserves concavity, $L(\boldsymbol{\theta})$ is a concave function of $\boldsymbol{\theta}$.

APPENDIX B
DERIVATION OF THE LOWER BOUND (23)

We have

$$\begin{aligned}
 g_n(\tau_n, \mathbf{c}_n) &= \frac{p_{v_n}^2 (\tau_n - \mathbf{c}_n^T \boldsymbol{\theta})}{F_{v_n}(\tau_n - \mathbf{c}_n^T \boldsymbol{\theta}) [1 - F_{v_n}(\tau_n - \mathbf{c}_n^T \boldsymbol{\theta})]} \\
 &\stackrel{(a)}{\geq} \frac{2}{\pi \sigma_{v_n}^2} \exp\left(-\frac{(\tau_n - \mathbf{c}_n^T \boldsymbol{\theta})^2}{2\sigma_{v_n}^2}\right) \\
 &\stackrel{(b)}{>} \frac{2}{\pi \mathbf{c}_n^T \mathbf{R}_{w_n} \mathbf{c}_n} \exp\left(-\frac{\delta^2}{2\lambda_{\min}(\mathbf{R}_{w_n})}\right) \\
 &\triangleq t_n \frac{1}{\mathbf{c}_n^T \mathbf{R}_{w_n} \mathbf{c}_n} \quad (71)
 \end{aligned}$$

where (a) comes by using the Chernoff bound for the CCDF

$$F_{v_n}(x) [1 - F_{v_n}(x)] \leq \frac{1}{4} e^{-\frac{x^2}{2\sigma_{v_n}^2}} \quad (72)$$

(b) follows from (i): $\sigma_{v_n}^2 = \mathbf{c}_n^T \mathbf{R}_{w_n} \mathbf{c}_n \geq \|\mathbf{c}_n\|_2^2 \lambda_{\min}(\mathbf{R}_{w_n})$, in which $\lambda_{\min}(\mathbf{R}_{w_n})$ denotes the smallest eigenvalue of \mathbf{R}_{w_n} , and (ii):

$$\begin{aligned}
 |\tau_n - \mathbf{c}_n^T \boldsymbol{\theta}| &< 2\|\mathbf{c}_n\|_1 \max(|\boldsymbol{\theta}_{\text{ub}}|) \\
 &\leq 2\sqrt{p}\|\mathbf{c}_n\|_2 \max(|\boldsymbol{\theta}_{\text{ub}}|) \triangleq \|\mathbf{c}_n\|_2 \delta \quad (73)
 \end{aligned}$$

in which the inequality comes by noting that τ_n is within the dynamic range of $\mathbf{c}_n^T \boldsymbol{\theta}$.

APPENDIX C
PROOF OF THEOREM 1

Let

$$\mathbf{C} \triangleq [\mathbf{c}_1 \quad \mathbf{c}_2 \quad \dots \quad \mathbf{c}_N] \quad (74)$$

where $\mathbf{C} \in \mathbb{R}^{p \times N}$. We can re-express (36) as

$$\begin{aligned}
 \min_{\mathbf{C}} \quad &\text{tr} \left\{ (\mathbf{C}\mathbf{C}^T)^{-1} \right\} \\
 \text{s.t.} \quad &[\mathbf{C}^T \mathbf{C}]_{ii} = 1, \quad \forall i \in \{1, \dots, N\}. \quad (75)
 \end{aligned}$$

To prove Theorem 1, let us first consider a new optimization that has the same objective function as (75) while with a relaxed constraint:

$$\begin{aligned}
 \min_{\mathbf{C}} \quad &\text{tr} \left\{ (\mathbf{C}\mathbf{C}^T)^{-1} \right\} \\
 \text{s.t.} \quad &\text{tr}(\mathbf{C}^T \mathbf{C}) = \text{tr}(\mathbf{C}\mathbf{C}^T) = N. \quad (76)
 \end{aligned}$$

Apparently, the feasible region defined by the constraint of (75) is a subset of that defined by (76). Hence the global minimum of (76) places a lower bound on the minimum achievable objective function value of (75). Also, since $\text{tr}(\mathbf{A}^{-1})$ is convex over the set of positive definite matrix and $\text{tr}(\mathbf{A}) = N$ is a linear constraint, the optimization (76) is convex whose optimum solution is given as follows.

Lemma 1: Consider the following optimization problem

$$\min_{\mathbf{A}} \text{tr}(\mathbf{A}^{-1}) \quad \text{s.t.} \quad \text{tr}(\mathbf{A}) = P \quad (77)$$

where $\mathbf{A} \in \mathbb{R}^{p \times p}$ is positive definite. The optimum solution to (77) is given by $\mathbf{A} = \frac{P}{p} \mathbf{I}$ and the minimum objective function value is p^2/P .

Proof: See Appendix G. □

From Lemma 1, we know that any \mathbf{C} satisfying $\mathbf{C}\mathbf{C}^T = \frac{N}{p} \mathbf{I}$ is an optimum solution of (76) and the attained minimum value is p^2/N , which is a lower bound on the minimum achievable objective function value of (75). When $N = kp$, where $k > 0$ is an integer, it can be easily verified that the proposed strategy attains the lower bound p^2/N . Therefore the strategy is an optimal solution to (75), i.e., (36). On the other hand, if $N \neq kp$, the objective function value achieved by the proposed strategy is upper bounded as

$$\begin{aligned}
 \text{tr} \left\{ \left(\sum_{n=1}^N \mathbf{c}_n \mathbf{c}_n^T \right)^{-1} \right\} &\stackrel{(a)}{<} \text{tr} \left\{ \left(\sum_{n=1}^M \mathbf{c}_n \mathbf{c}_n^T \right)^{-1} \right\} \\
 &\stackrel{(b)}{=} \text{tr} \left\{ (\text{floor}(N/p) \mathbf{U}\mathbf{U}^T)^{-1} \right\} \\
 &= \frac{p}{\text{floor}(N/p)} \quad (78)
 \end{aligned}$$

where (a) comes from the fact that $\text{tr}(\mathbf{A}^{-1})$ is convex over the set of positive definite matrix, (b) follows from $M = \text{floor}(N/p)p$ and the compression vector assignment strategy for these M sensors. The ratio of the objective function value attained by the proposed strategy to the lower bound p^2/N , γ , is therefore upper bounded by

$$\gamma < \frac{N/p}{\text{floor}(N/p)}. \quad (79)$$

The proof is completed here.

APPENDIX D
PROOF OF THEOREM 2

The proof of Theorem 2 follows a same procedure as that of Theorem 1. Let

$$\mathbf{\Lambda} \triangleq \text{diag} \left(\frac{1}{\sigma_{w_1}^2}, \dots, \frac{1}{\sigma_{w_N}^2} \right). \quad (80)$$

We can re-express (39) as

$$\begin{aligned}
 \min_{\mathbf{C}} \quad &\text{tr} \left\{ (\mathbf{C}\mathbf{\Lambda}\mathbf{C}^T)^{-1} \right\} \\
 \text{s.t.} \quad &[\mathbf{C}^T \mathbf{C}]_{ii} = 1, \quad \forall i \in \{1, \dots, N\}. \quad (81)
 \end{aligned}$$

Similarly as in the proof of Theorem 1, we construct a new optimization with a relaxed constraint:

$$\begin{aligned}
 \min_{\mathbf{C}} \quad &\text{tr} \left\{ (\mathbf{C}\mathbf{\Lambda}\mathbf{C}^T)^{-1} \right\} \\
 \text{s.t.} \quad &\text{tr}(\mathbf{C}\mathbf{\Lambda}\mathbf{C}^T) = \sum_{n=1}^N \frac{1}{\sigma_{w_n}^2} \triangleq K. \quad (82)
 \end{aligned}$$

It can be readily verified that the feasible region of (81) is a subset of that of (82). Hence the global minimum of (82) provides a lower bound on the minimum achievable objective function value of (81). Recalling Lemma 1, we know that the

global minimum of (82) is p^2/K , which is achieved when $\mathbf{C}\mathbf{A}\mathbf{C}^T = \frac{K}{p}\mathbf{I}$.

Now considering the proposed strategy, we have

$$\begin{aligned} \mathbf{C}\mathbf{A}\mathbf{C}^T &= \sum_{n=1}^N \frac{1}{\sigma_{w_n}^2} \mathbf{c}_n \mathbf{c}_n^T = \sum_{i=1}^p K_i \mathbf{U}[:, i] \mathbf{U}[:, i]^T \\ &\succeq \min(\{K_i\}) \mathbf{U}\mathbf{U}^T = \min(\{K_i\}) \mathbf{I}. \end{aligned} \quad (83)$$

Since $\text{tr}\{\mathbf{A}^{-1}\}$ is convex over the set of positive definite matrix, the objective function value attained by the proposed strategy is upper bounded by

$$\begin{aligned} \text{tr}\left\{(\mathbf{C}\mathbf{A}\mathbf{C}^T)^{-1}\right\} &\leq \text{tr}\left\{(\min(\{K_i\})\mathbf{I})^{-1}\right\} \\ &= \frac{p}{\min(\{K_i\})}. \end{aligned} \quad (84)$$

The ratio of the objective function value achieved by the proposed strategy to the lower bound p^2/K is therefore upper bounded by

$$\gamma \leq \frac{K}{p \min(\{K_i\})}. \quad (85)$$

The proof is completed here.

APPENDIX E PROOF OF THEOREM 3

As we mentioned earlier, due to the nonirreducible form of the denominator, the optimization (44) is more challenging as compared with (36). In this case, constructing a new optimization as we did before does not lead to an insight into solving the problem. Here we take a different approach. We first find a transitional lower bound on the objective function (44) for any specific set of compression vectors $\{\mathbf{c}_n\}$ by using the following lemma.

Lemma 2: Suppose that we have two positive definite matrices $\mathbf{A} \in \mathbb{R}^{p \times p}$ and $\mathbf{B} \in \mathbb{R}^{p \times p}$, where \mathbf{A} is a diagonal matrix with its diagonal elements equivalent to those of \mathbf{B} , i.e., $[\mathbf{A}]_{ii} = [\mathbf{B}]_{ii}$, then we have

$$\text{tr}(\mathbf{A}^{-1}) \leq \text{tr}(\mathbf{B}^{-1}). \quad (86)$$

Proof: See Appendix H. \square

We write $\mathbf{c}_n = \mathbf{U}_w \boldsymbol{\beta}_n$, where \mathbf{U}_w is obtained by carrying out the EVD of \mathbf{R}_w , i.e., $\mathbf{R}_w = \mathbf{U}_w \mathbf{D}_w \mathbf{U}_w^T$, and $\boldsymbol{\beta}_n \triangleq [\beta_{n,1} \ \beta_{n,2} \ \dots \ \beta_{n,p}]^T$ is a normalized column vector with $\boldsymbol{\beta}_n^T \boldsymbol{\beta}_n = 1$. Then we have

$$\begin{aligned} \sum_{n=1}^N \frac{\mathbf{c}_n \mathbf{c}_n^T}{\mathbf{c}_n^T \mathbf{R}_w \mathbf{c}_n} &= \mathbf{U}_w \left(\sum_{n=1}^N \frac{\boldsymbol{\beta}_n \boldsymbol{\beta}_n^T}{\sum_{i=1}^p \beta_{n,i}^2 d_{w,i}} \right) \mathbf{U}_w^T \\ &\triangleq \mathbf{U}_w \left(\sum_{n=1}^N \frac{\boldsymbol{\beta}_n \boldsymbol{\beta}_n^T}{\gamma_n} \right) \mathbf{U}_w^T \triangleq \mathbf{U}_w \boldsymbol{\Omega} \mathbf{U}_w^T \end{aligned} \quad (87)$$

where $d_{w,i}$ denotes the i th diagonal entry of \mathbf{D}_w , $\gamma_n \triangleq \sum_{i=1}^p \beta_{n,i}^2 d_{w,i}$, and

$$\boldsymbol{\Omega} \triangleq \sum_{n=1}^N \frac{\boldsymbol{\beta}_n \boldsymbol{\beta}_n^T}{\gamma_n}. \quad (88)$$

Let $g_{n,i} \triangleq \beta_{n,i}^2 / \gamma_n$ and

$$\boldsymbol{\Gamma} \triangleq \text{diag} \left(\sum_{n=1}^N g_{n,1}, \dots, \sum_{n=1}^N g_{n,p} \right). \quad (89)$$

Note that the corresponding diagonal elements of the two matrices $\boldsymbol{\Omega}$ and $\boldsymbol{\Gamma}$ are identical, i.e., $[\boldsymbol{\Omega}]_{ii} = [\boldsymbol{\Gamma}]_{ii}$. Recalling Lemma 2, the objective function (44) therefore is lower bounded by

$$\begin{aligned} \text{tr} \left\{ \left(\sum_{n=1}^N \frac{\mathbf{c}_n \mathbf{c}_n^T}{\mathbf{c}_n^T \mathbf{R}_w \mathbf{c}_n} \right)^{-1} \right\} &= \text{tr} \left\{ (\mathbf{U}_w \boldsymbol{\Omega} \mathbf{U}_w^T)^{-1} \right\} \\ &\geq \text{tr} \left\{ (\mathbf{U}_w \boldsymbol{\Gamma} \mathbf{U}_w^T)^{-1} \right\}. \end{aligned} \quad (90)$$

We note that the above lower bound is a transitional result because it depends on the vectors $\{\boldsymbol{\beta}_n\}$ which are specific for each set of compression vectors $\{\mathbf{c}_n\}$. We now proceed to find a universal lower bound which is independent of the choice of the compression vectors. For notational convenience, let

$$\chi_i \triangleq \sum_{n=1}^N g_{n,i} d_{w,i}, \quad \forall i \in \{1, \dots, p\}. \quad (91)$$

Since we have

$$\sum_{i=1}^p g_{n,i} d_{w,i} = \sum_{i=1}^p \frac{\beta_{n,i}^2 d_{w,i}}{\gamma_n} = 1 \quad (92)$$

it can be easily verified that $\{\chi_i\}$ satisfies the following constraint:

$$\sum_{i=1}^p \chi_i = N. \quad (93)$$

Noting that

$$\mathbf{U}_w \boldsymbol{\Gamma} \mathbf{U}_w^T = \sum_{i=1}^p \frac{\chi_i}{d_{w,i}} \mathbf{U}_w[:, i] \mathbf{U}_w[:, i]^T \quad (94)$$

the universal lower bound can therefore be obtained by solving the following optimization

$$\begin{aligned} \min_{\{\chi_i\}} \quad &\text{tr} \left\{ \left(\sum_{i=1}^p \frac{\chi_i}{d_{w,i}} \mathbf{U}_w[:, i] \mathbf{U}_w[:, i]^T \right)^{-1} \right\} \\ \text{s.t.} \quad &\sum_{i=1}^p \chi_i = N, \quad \chi_i > 0, \quad \forall i \end{aligned} \quad (95)$$

which is equivalent to

$$\begin{aligned} \min_{\{\chi_i\}} \quad &\sum_{i=1}^p \frac{d_{w,i}}{\chi_i} \\ \text{s.t.} \quad &\sum_{i=1}^p \chi_i = N, \quad \chi_i > 0, \quad \forall i. \end{aligned} \quad (96)$$

The above optimization (96) can be solved analytically by resorting to the Lagrangian function and KKT conditions (the development is similar as that in Appendix G and thus omitted

here). Its optimum solution and minimum objective function value are given as follows, respectively

$$\chi_i^* = \frac{N\sqrt{d_{w,i}}}{\sum_{n=1}^p \sqrt{d_{w,n}}}, \quad \forall i \in \{1, \dots, p\} \quad (97)$$

$$\sum_{i=1}^p \frac{d_{w,i}}{\chi_i^*} = \frac{1}{N} \left[\sum_{i=1}^p \sqrt{d_{w,i}} \right]^2. \quad (98)$$

Hence (98) is a lower bound on the minimum achievable objective function value of (44). Apparently this lower bound can be achieved by the proposed strategy when $N_i = \chi_i^* \forall i$. If $\{\chi_i^*\}$ are not integers, near-optimal performance can be attained by the proposed strategy. A rough analysis is provided in the following to show that the achieved objective function value is within a small factor of the lower bound. Let $N_i = \chi_i^* + \delta_i$ denote the rounded integer. Using Taylor expansion, we can write

$$\frac{1}{N_i} = \frac{1}{\chi_i^*} \sum_{n=0}^{\infty} \left(\frac{\delta_i}{\chi_i^*} \right)^n \approx \frac{1}{\chi_i^*} + \frac{\delta_i}{(\chi_i^*)^2}. \quad (99)$$

The deviation between the achieved objective function value and the lower bound is upper bounded by

$$\begin{aligned} \left| \sum_{i=1}^p \frac{d_{w,i}}{\chi_i^*} - \sum_{i=1}^p \frac{d_{w,i}}{N_i} \right| &\approx \left| \sum_{i=1}^p \frac{d_{w,i} \delta_i}{(\chi_i^*)^2} \right| \\ &\stackrel{(a)}{\leq} \sum_{i=1}^p \frac{d_{w,i}}{(\chi_i^*)^2} \\ &\leq \frac{1}{\min(\{\chi_i^*\})} \sum_{i=1}^p \frac{d_{w,i}}{\chi_i^*} \end{aligned} \quad (100)$$

where (a) comes from the fact that $|\delta_i| < 1 \forall i$, $\min(\{\chi_i^*\})$ denotes the minimum value among the set $\{\chi_i^*\}$. Therefore the normalized deviation of the proposed solution from the lower bound is

$$\frac{\left| \sum_{i=1}^p \frac{d_{w,i}}{\chi_i^*} - \sum_{i=1}^p \frac{d_{w,i}}{N_i} \right|}{\sum_{i=1}^p \frac{d_{w,i}}{\chi_i^*}} \leq \frac{1}{\min(\{\chi_i^*\})} \quad (101)$$

which indicates that the objective function value associated with the proposed strategy is within a factor $1 + \frac{1}{\min(\{\chi_i^*\})}$ of the lower bound.

APPENDIX F
PROOF OF THEOREM 4

We introduce a sufficiently small value $\epsilon > 0$ such that for any $n \in \{1, \dots, N\}$, $\phi_n \triangleq \frac{1}{\epsilon \sigma_{w_n}^2}$ is an integer or is sufficiently large such that it can be rounded to its nearest integer with a negligible error. We therefore can write

$$\sum_{n=1}^N \frac{1}{\sigma_{w_n}^2} \frac{\mathbf{c}_n \mathbf{c}_n^T}{\mathbf{c}_n^T \mathbf{R}_w \mathbf{c}_n} = \epsilon \sum_{n=1}^N \frac{\phi_n \mathbf{c}_n \mathbf{c}_n^T}{\mathbf{c}_n^T \mathbf{R}_w \mathbf{c}_n}. \quad (102)$$

Since $\{\phi_n\}$ are integers, the term on the right hand side of (102) can be considered as a particular case of the following form:

$$\epsilon \sum_{n=1}^L \frac{\mathbf{c}_n \mathbf{c}_n^T}{\mathbf{c}_n^T \mathbf{R}_w \mathbf{c}_n} \quad (103)$$

with L sensors divided into N groups, group n consisting of ϕ_n sensors and sharing the same compression vector, where $L \triangleq \sum_{n=1}^N \phi_n$. Hence the global minimum of the following optimization provides a lower bound on the minimum achievable value of the optimization (49)

$$\begin{aligned} \min_{\{\mathbf{c}_n\}} \quad &\text{tr} \left\{ \left(\epsilon \sum_{n=1}^L \frac{\mathbf{c}_n \mathbf{c}_n^T}{\mathbf{c}_n^T \mathbf{R}_w \mathbf{c}_n} \right)^{-1} \right\} \\ \text{s.t.} \quad &\mathbf{c}_n^T \mathbf{c}_n = 1, \quad \forall n \in \{1, \dots, L\}. \end{aligned} \quad (104)$$

The above optimization (104) is exactly the problem we studied in Section V-D. From Theorem 3, we know that the minimum achievable value of (104) is lower bounded by

$$\begin{aligned} \text{tr} \left\{ \left(\epsilon \sum_{n=1}^L \frac{\mathbf{c}_n \mathbf{c}_n^T}{\mathbf{c}_n^T \mathbf{R}_w \mathbf{c}_n} \right)^{-1} \right\} &\geq \frac{1}{\epsilon L} \left[\sum_{i=1}^p \sqrt{d_{w,i}} \right]^2 \\ &\stackrel{(a)}{=} \frac{1}{K} \left[\sum_{i=1}^p \sqrt{d_{w,i}} \right]^2 \end{aligned} \quad (105)$$

where (a) comes from

$$\epsilon L = \epsilon \sum_{n=1}^N \phi_n = \sum_{n=1}^N \frac{1}{\sigma_{w_n}^2} = K. \quad (106)$$

Consequently, the minimum achievable value of (49) is lower bounded by (105). Also, by substituting (53) into (51), we see that this lower bound can be attained by the proposed strategy. The proof is completed here.

APPENDIX G
PROOF OF LEMMA 1

Let $\mathbf{A} = \mathbf{U}_a \mathbf{D}_a \mathbf{U}_a^T$ denote the EVD of \mathbf{A} , where $\mathbf{U}_a \in \mathbb{R}^{p \times p}$ and $\mathbf{D}_a \in \mathbb{R}^{p \times p}$. By replacing \mathbf{A} with $\mathbf{U}_a \mathbf{D}_a \mathbf{U}_a^T$, the optimization (77) is reduced to determining the diagonal matrix $\mathbf{D}_a \triangleq \text{diag}(d_1, \dots, d_p)$

$$\begin{aligned} \min_{\{d_i\}} \quad &\sum_{i=1}^p \frac{1}{d_i} \\ \text{s.t.} \quad &\sum_{i=1}^p d_i = P, \\ &d_i > 0, \quad \forall i \in \{1, \dots, p\}. \end{aligned} \quad (107)$$

The Lagrangian function L associated with (107) is given by

$$L(d_i; \lambda; \nu_i) = \sum_{i=1}^p \frac{1}{d_i} - \lambda \left(P - \sum_{i=1}^p d_i \right) - \sum_{i=1}^p \nu_i d_i \quad (108)$$

which gives the following KKT conditions [20]:

$$\begin{aligned} -\frac{1}{d_i^2} + \lambda - \nu_i &= 0, \quad \forall i \\ P - \sum_{i=1}^p d_i &= 0 \\ \nu_i d_i &= 0, \quad \forall i \\ \nu_i &\geq 0, \quad \forall i \\ d_i &> 0, \quad \forall i. \end{aligned}$$

From the last three equations of the above KKT conditions, we have $\nu_i = 0, \forall i$. With this result, d_i can be readily solved using the first two equations as $d_i = \frac{P}{p}, \forall i$, i.e., the optimal \mathbf{D}_a is given by $\mathbf{D}_a^* = \frac{P}{p}\mathbf{I}$. Consequently we have $\mathbf{A}^* = \frac{P}{p}\mathbf{I}$ and the minimum objective function value is $\frac{P^2}{p}$. The proof is completed here.

APPENDIX H PROOF OF LEMMA 2

By carrying out the EVD: $\mathbf{B} = \mathbf{U}_b \mathbf{D}_b \mathbf{U}_b^T$, we can write

$$\text{tr}(\mathbf{B}^{-1}) = \text{tr}(\mathbf{D}_b^{-1}). \quad (109)$$

Hence we only need to prove

$$\text{tr}(\mathbf{A}^{-1}) \leq \text{tr}(\mathbf{D}_b^{-1}) \quad (110)$$

which is equivalent to establishing the following:

$$\sum_{i=1}^p \frac{1}{a_i} \leq \sum_{i=1}^p \frac{1}{b_i} \quad (111)$$

where a_i and b_i denote the i th diagonal elements of \mathbf{A} and \mathbf{D}_b , respectively. Without loss of generality, we assume $0 < a_1 \leq a_2 \leq \dots \leq a_p$ and $0 < b_1 \leq b_2 \leq \dots \leq b_p$.

Since $[\mathbf{A}]_{ii} = [\mathbf{B}]_{ii} \forall i$, we have the following two important properties regarding the diagonal elements of $\{a_i\}$ and $\{b_i\}$. First, we have

$$\sum_{i=1}^p a_i = \sum_{i=1}^p b_i \quad (112)$$

which can be readily verified from $\text{tr}(\mathbf{A}) = \text{tr}(\mathbf{B})$. Second, for any $p > r \geq 1$, we have the following inequalities:

$$\begin{aligned} \sum_{i=1}^r b_i &\leq \sum_{i=1}^r a_i \\ \sum_{i=1}^r b_{p-i+1} &\geq \sum_{i=1}^r a_{p-i+1} \end{aligned} \quad (113)$$

which are generalized results of the Rayleigh–Ritz theorem [24, Corollary 4.3.18]. We now prove (111) by utilizing the properties (112)–(113).

From (113), we know that $b_1 \leq a_1$ and $b_p \geq a_p$. Without loss of generality, we can find a set of indices $(k_1, k_2, \dots, k_{2m-1}, p)$ such that $b_i \leq a_i$ for $i = 1, \dots, k_1$, and $b_i \geq a_i$ for $i = k_1 + 1, \dots, k_2$; $b_i \leq a_i$ for $i = k_2 + 1, \dots, k_3$, and $b_i \geq a_i$ for $i = k_3 + 1, \dots, k_4$; and so on and so forth. Considering the first k_2 elements, we have

$$\begin{aligned} \sum_{i=1}^{k_2} \left(\frac{1}{b_i} - \frac{1}{a_i} \right) &= \sum_{i=1}^{k_1} \frac{a_i - b_i}{a_i b_i} + \sum_{i=k_1+1}^{k_2} \frac{a_i - b_i}{a_i b_i} \\ &\stackrel{(a)}{\geq} \sum_{i=1}^{k_1} \frac{a_i - b_i}{a_{k_1} b_{k_1}} + \sum_{i=k_1+1}^{k_2} \frac{a_i - b_i}{a_{k_1} b_{k_1}} \\ &= \frac{\sum_{i=1}^{k_2} (a_i - b_i)}{a_{k_1} b_{k_1}} \stackrel{(b)}{\geq} 0 \end{aligned} \quad (114)$$

where (a) comes by noting that the elements $\{a_i\}$ and $\{b_i\}$ are in ascending orders, $a_i - b_i \geq 0$ for $i = 1, \dots, k_1$, and $a_i - b_i \leq 0$ for $i = k_1 + 1, \dots, k_2$; and (b) follows from the inequality (113). Now considering the first k_4 elements, we have

$$\begin{aligned} \sum_{i=1}^{k_4} \left(\frac{1}{b_i} - \frac{1}{a_i} \right) &\stackrel{(a)}{\geq} \frac{\sum_{i=1}^{k_2} (a_i - b_i)}{a_{k_1} b_{k_1}} + \sum_{i=k_2+1}^{k_3} \frac{a_i - b_i}{a_i b_i} + \sum_{i=k_3+1}^{k_4} \frac{a_i - b_i}{a_i b_i} \\ &\stackrel{(b)}{\geq} \frac{\sum_{i=1}^{k_2} (a_i - b_i)}{a_{k_3} b_{k_3}} + \sum_{i=k_2+1}^{k_3} \frac{a_i - b_i}{a_{k_3} b_{k_3}} + \sum_{i=k_3+1}^{k_4} \frac{a_i - b_i}{a_{k_3} b_{k_3}} \\ &= \frac{\sum_{i=1}^{k_4} (a_i - b_i)}{a_{k_3} b_{k_3}} \stackrel{(c)}{\geq} 0 \end{aligned} \quad (115)$$

where (a) comes by using the previous result (114); (b) comes by following the same derivation as in (114); and (c) follows from the inequality (113). By repeating this procedure, we can eventually prove that such an inequality holds for all p elements. The proof is completed here.

REFERENCES

- [1] D. Li, K. D. Wong, Y. H. Hu, and A. M. Sayeed, "Detection, classification, and tracking of targets," *IEEE Signal Process. 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. Gubner, "Distributed estimation and quantization," *IEEE Trans. Inf. Theory*, vol. 39, no. 4, pp. 1456–1459, Jul. 1993.
- [4] V. Megalooikonomou and Y. Yesha, "Quantizer design for distributed estimation with communication constraints and unknown observation statistics," *IEEE Trans. Commun.*, vol. 48, no. 2, pp. 181–184, Feb. 2000.
- [5] W. M. Lam and A. R. Reibman, "Design of quantizers for decentralized estimation systems," *IEEE Trans. Commun.*, vol. 41, pp. 1602–1605, Nov. 1993.
- [6] H. Papadopoulos, G. Wornell, and A. Oppenheim, "Sequential signal encoding from noisy measurements using quantizers with dynamic bias control," *IEEE Trans. Inf. Theory*, vol. 47, no. 2, pp. 978–1002, Mar. 2001.
- [7] 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.

- [8] Z. Luo, "Universal decentralized estimation in a bandwidth constrained sensor network," *IEEE Trans. Inf. Theory*, vol. 51, no. 6, pp. 2210–2219, Jun. 2005.
- [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] P. Venkatasubramanian, 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.
- [11] 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.
- [12] A. Ribeiro and G. B. Giannakis, "Bandwidth-constrained distributed estimation for wireless sensor networks—Part II: Unknown probability density function," *IEEE Trans. Signal Processing*, vol. 54, no. 7, pp. 2784–2796, July 2006.
- [13] K. Zhang and X. R. Li, "Optimal sensor data quantization for best linear unbiased estimation fusion," in *Proc. 43rd IEEE Conf. Decision Contr.*, Atlantis, Paradise Island, Bahamas, 2004.
- [14] D. A. Pierre, *Optimization Theory With Applications*. New York: Wiley, 1969.
- [15] S. M. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory*. Upper Saddle River, NJ: Prentice Hall, 1993.
- [16] 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.
- [17] 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.
- [18] I. D. Schizas, G. B. Giannakis, and Z.-Q. Luo, "Distributed estimation using reduced dimensionality sensor observations," *IEEE Trans. Signal Processing*, vol. 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] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge, U.K.: Cambridge University Press, 2003.
- [21] A. Dogandžić and K. Liu, "Decentralized random-field estimation for sensor networks using quantized spatially correlated data and fusion-center feedback," *IEEE Trans. Signal Processing*, vol. 12, pp. 6069–6085, Dec. 2008.
- [22] G. H. Golub and C. F. Van Loan, *Matrix Computations*, 3rd ed. Baltimore, MD: Johns Hopkins University Press, 1996.
- [23] 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.
- [24] R. A. Horn and C. R. Johnson, *Matrix Analysis*. Cambridge, U.K.: Cambridge University Press, 1985.

Jun Fang (M'08) received the B.Sc. 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, Raleigh, NC, as a Postdoctoral Research Associate. Currently, he is a Postdoctoral Research Associate with the Department of Electrical and Computer Engineering, Stevens Institute of Technology, Hoboken, NJ. 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–SM'03) 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 the summers of 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 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.